

**Rodrigo Guedes Borba**

**UTILIZAÇÃO DE ANÁLISE COMPOSICIONAL EM UM ESTUDO  
GEOQUÍMICO NA REGIÃO DE AMAPARI**

**Trabalho Final de Curso  
(Geologia)**

UFRJ  
Rio de Janeiro  
2016



UFRJ

**Rodrigo Guedes Borba**

## **UTILIZAÇÃO DE ANÁLISE COMPOSICIONAL EM UM ESTUDO GEOQUÍMICO NA REGIÃO DE AMAPARI**

Trabalho Final de Curso de Graduação em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, apresentado como requisito necessário para obtenção do grau de Geólogo.

Orientador(es):

Luis Paulo Vieira Braga  
Claudio Gerheim Porto

Rio de Janeiro

Março de 2016

Borba, Rodrigo Guedes

**Utilização de análise composicional em um estudo geoquímico na região de Amapari / Rodrigo Guedes**

Borba - - Rio de Janeiro: UFRJ / IGeo, 2016

xi, 20 p. : il.; 30cm

Trabalho Final de Curso (Geologia) – Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia, 2016.

Orientador(es): Luis Paulo Vieira Braga, Claudio Gerheim Porto

1. Geologia. 2. Geologia Aplicada – Trabalho de Conclusão de Curso. I. Luis Paulo Vieira Braga, Claudio Gerheim Porto. II. Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia. III. Utilização de análise composicional em um estudo geoquímico na região de Amapari.

Rodrigo Guedes Borba

UTILIZAÇÃO DE ANÁLISE COMPOSICIONAL EM UM ESTUDO  
GEOQUÍMICO NA REGIÃO DE AMAPARI

Trabalho Final de Curso de Graduação em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, apresentado como requisito necessário para obtenção do grau de Geólogo.

Orientador(es):

Dr. Luís Paulo Vieira Braga

Dr. Claudio Gerheim Porto

Aprovada em:

Por:

---

Orientador: Luís Paulo Vieira Braga (UFRJ)

---

Orientador: Claudio Gerheim Porto (UFRJ)

---

Dra. Cícera Neysi de Almeida (UFRJ)

---

Dr. Francisco José da Silva (UFRRJ)

UFRJ  
Rio de Janeiro  
2016

“A cada um de nós, só é dada uma pequena dose de loucura. E você não deve perde-la”

Robin Williams

## **Agradecimentos**

Agradeço ao orientador Prof. Dr. Luis Paulo Vieira Braga, pelas dicas, conselhos, conhecimento e muita paciência.

Ao orientador Prof. Dr. Claudio Gerheim Porto, por todo o conhecimento compartilhado na Universidade Federal do Rio de Janeiro e na Companhia de Pesquisa de Recursos Minerais.

Ao Prof. Dr. Francisco José da Silva da Universidade Federal Rural do Rio de Janeiro pela orientação na preparação dos dados.

Ao colega, Ronald Armando, pelos anos de amizade e pela ajuda tanto na monografia quanto no decorrer de todo o curso.

Aos colegas, Camila Leão, Camilla Bazzarella, Daniel Coré, Ariadne Ázaro, Erika Martins e Patricia Helena, pelas dicas e companheirismo em todo o percurso da graduação.

À minha família por todo amor, suporte e paciência.

## Resumo

BORBA, Rodrigo Guedes. **Utilização de análise composicional em um estudo geoquímico na região de Amapari**. 2016. xi, 22 p. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

O uso de ferramentas estatísticas na área de geoquímica vem se tornando mais popular desde o século XX, a facilidade de acesso à informação e avanço de tecnologia favorece a utilização dessas ferramentas. Terrenos muito alterados sofrem com a modificação de sinais geoquímicos na superfície, dificultando a descoberta de áreas mineralizadas a partir de amostras de solo. O uso da análise composicional associada à análise de componentes principais é recomendado para análise de dados multivariados expressos por proporções relativas que somam 1 ou 100%.

Neste estudo realizado com dados provenientes da região do município de Amapari, estado do Amapá, foram utilizadas 808 amostras de solo, analisadas para 51 elementos utilizando ICP-AES/MS. Estas amostras foram separadas em amostras de treinamento, para gerar um modelo estatístico, e amostras de validação, para a aplicação e comprovação do modelo, para posteriormente o modelo ser aplicado em regiões não exploradas e possivelmente detectar outras áreas mineralizadas.

Palavras-chave: análise composicional; análise de componentes principais; geoestatística; geoquímica.

## Abstract

BORBA, Rodrigo Guedes. **Compositional analysis usage in a geochemical study in the Amapari region**. 2016. xi, 22 p. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

The use of statistical tools in geochemistry is becoming more popular since the twentieth century, the easy access to information and technology advancement benefit the use of these tools. Very weathered terrains are often faced with the problem of strong modifications of the geochemical signal at surface, making it difficult to discover mineralized areas from soil samples. The use of compositional analysis associated with the principal components analysis is advised to analyze multivariate data given by relative proportions summing up to 1 or 100%.

In this study with data from Amapari region, state of Amapá, 808 soil samples were selected to be analyzed for 51 elements using ICP-AES / MS technology. These samples were separated into training samples to generate a statistical model, and validation samples for the application and verification of the model, that may be applied in areas not exploited and possibly detect other mineralized areas.

Key-Words: Compositional Analysis, Principal Components Analysis; Geostatistics; Geochemistry.

## LISTA DE FIGURAS

- Figura 3.1 - Geologia e localização da área de estudo. (Faraco et al., 2004). **3**
- Figura 4.1 – Mapa de pontos da área, com as subdivisões Green, contornada em verde, e Target, contornada em vermelho. **6**
- Figura 4.2 – Como funciona o sistema de treinamento e validação, onde a figura acima representa os pontos do *subset* de treinamento e a figura de baixo representa os pontos do *subset* de validação com o modelo criado no treinamento. **7**
- Figura 4.3 – Análise de componentes principais utilizando os mesmos elementos usados na análise composicional. **9**
- Figura 5.1 – Histograma do elemento B (boro), indicando muitos valores abaixo do limite de detecção. **11**
- Figura 5.2 – *Boxplot* padronizado do elemento P (Fósforo), indicando elementos com medianas próximas, que não ajudariam a discriminação das áreas. **12**
- Figura 5.3 – *Biplot* indicando a separação das amostras em 3 *clusters*, sendo o B caracterizado pela forte influência de K (potássio), o C pela alta influência de Tl (tálio) e o A pela baixa influência de K (potássio). **13**
- Figura 5.4 – Mapa de pontos de treinamento da região de Amapari com as litologias. **14**
- Figura 5.5 – *Biplot* do segundo treinamento, indicando dois *clusters*, sendo o *cluster* deslocado a parte inferior direita do gráfico sendo principalmente representado por Tl (tálio). **15**
- Figura 5.6 – *Biplot* separado, a esquerda os *loadings* das variáveis, e a direita os *scores* das amostras. **16**

Figura 5.7 – Mapa de pontos do Segundo treinamento. Pontos verdes são relativos ao *cluster* inferior direito e pontos vermelhos do *cluster* superior esquerdo. **17**

Figura 6.1 – Mapa de pontos da validação. Pontos verdes relativos ao *cluster* inferior direito da etapa de treinamento e pontos vermelhos relativos ao *cluster* superior esquerdo. **18**

Figura 6.2 – Matriz de confusão da validação. Destacando os valores de Falso Positivo e Falso Negativo. **19**

## Sumário

Agradecimentos .....	vi
Resumo .....	vii
<i>Abstract</i> .....	viii
Lista de figuras .....	ix
1. INTRODUÇÃO .....	1
2. OBJETIVOS .....	2
3. GEOLOGIA REGIONAL .....	3
4. MATERIAIS E MÉTODOS .....	5
4.1 Dados .....	5
4.2 Treinamento e Validação .....	7
4.3 Análise de Componentes Principais (PCA) .....	8
4.4 Análise Composicional .....	9
5. DESENVOLVIMENTO .....	10
5.1 Eliminação de variáveis .....	10
5.2 Primeiro Treinamento .....	12
5.3 Segundo Treinamento .....	13
6. VALIDAÇÃO .....	17
7. CONCLUSÕES.....	19
Referências bibliográficas .....	20

## 1. INTRODUÇÃO

A análise composicional é uma metodologia relativamente nova na Geoquímica, embora seus princípios teóricos datem de 1960 (Aitchison, 2005), somente no ano 2000 foi formulado um conjunto de transformações que permite utilizar a análise multivariada em dados composicionais. Nesse trabalho, utiliza-se o pacote Compositions, implementado no ambiente R.

A área de estudo está localizada em uma região próxima ao município de Amapari, que se encontra no estado do Amapá, e vem sendo estudada desde os anos 70 por diversas empresas, o estudo utiliza os dados coletados por empresas através de malhas prospectivas de geoquímica de solo para investigar alvos com alto potencial para mineralização de ouro. Estas amostras foram analisadas no contexto do projeto LATAM (Porto, 2007) para 51 elementos utilizando os métodos analíticos ICP-AES/MS.

Nesta área foram coletadas 14263 amostras das quais 808 foram utilizadas, pois foi delimitada uma região mineralizada que chamamos de área Target e contem 329 pontos e outra área sem mineralização conhecida que chamamos de Green e contem 479 pontos. Destas 808 amostras, separamos 707 para serem as amostras de treinamento e selecionamos 101 para serem as amostras de validação, mantendo a proporção.

## **2. OBJETIVOS**

O principal objetivo desse trabalho é utilizar a análise composicional para a criação de um modelo capaz de identificar os elementos farejadores de ouro na região do estudo, além disso, utilizar os dados para confeccionar um mapa de pontos com os resultados, e gerar uma matriz de confusão dos dados de validação para uma representação numérica dos resultados obtidos.

### 3. GEOLOGIA REGIONAL

O Cráton Amazonas, situado no norte da América do Sul e um dos maiores crátons do mundo, está estável desde o início do Neoproterozóico, (Cordani e Sato, 1999). É formado por dois escudos, o Escudo das Guianas (norte) e o Escudo Brasil Central (sul), com sua separação sendo marcada pela Bacia do Amazonas de idade Paleozoica. (Melo *et al*, 2003). A área de estudo está localizada na região centro-sudeste do estado do Amapá e suas principais unidades correspondem são: o Embasamento cristalino, o Grupo Vila Nova e o Complexo Máfico-Ultramáfico Bacuri.

O embasamento regional, conhecido como Complexo das Guianas (Lima *et al*, 1974), ocorre como núcleos escassos nos terrenos Proterozóicos. Consiste em rochas polimetamorfizadas, parcialmente retrabalhadas durante o evento Transamazônico e representado por granulitos, gnaisses, anfibolitos e migmatitos, e rochas ígneas associadas. A idade de datação U-Pb do embasamento é de 3.4Ga, (Montgomery, 1979).

O Complexo Máfico-Ultramáfico Bacuri intrude o embasamento por um *trend* leste-oeste e é composto de anfibolitos, serpentinitas, tremolititas e cromitas. As rochas são muito deformadas, da fácies anfibolito, no entanto é possível identificar algumas estruturas primárias, (Spier, 1999).

A Suíte Metamórfica Vila Nova ocorre ao longo de *trends* NW-SE e apresenta *dips* subverticais, geralmente para NE. Zonas de cisalhamento são comuns devido à deformação rúptil-dúctil que afetou as rochas vulcano-sedimentares.

A Suíte Vila Nova é de idade proterozóica e é caracterizada pela sequência vulcano-sedimentar metamorfizadas sob fácies xisto verde – anfibolito. McReath e Faraco, 1997) dizem que, de acordo com a idade de  $2.264 \pm 34$  Ma dos anfibolitos, a Suíte Vila Nova é

cronologicamente equivalente aos *greenstone belts* que ocorrem ao norte do Escudo das Guianas. A figura 3.1 mostra o mapa da geologia regional.

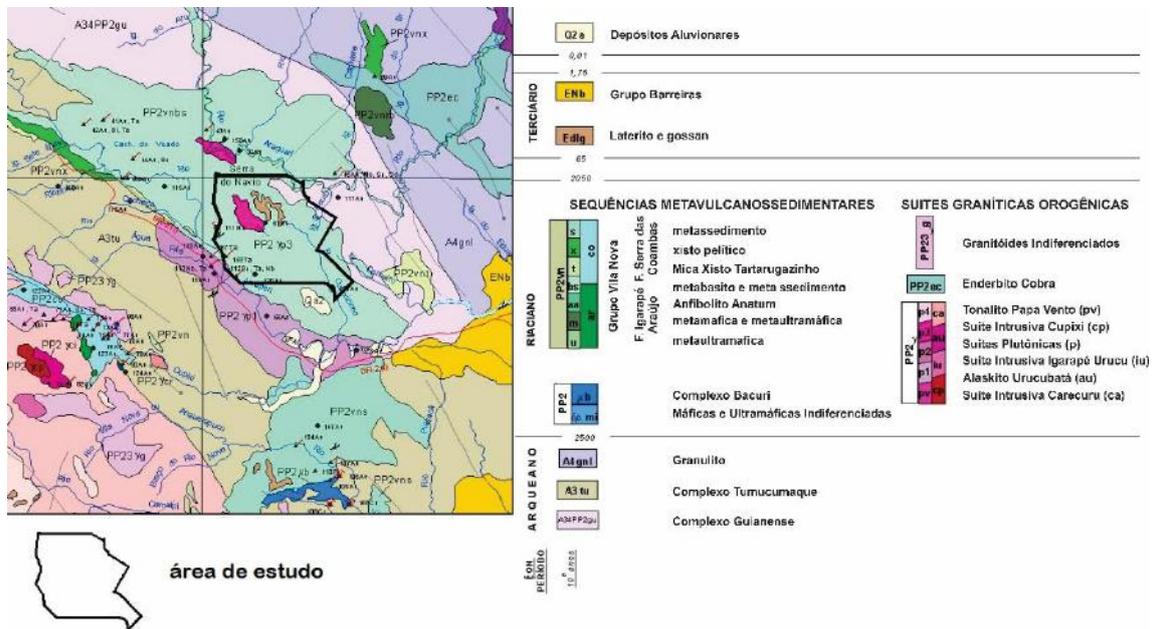


Figura 3.1 - Geologia e localização da área de estudo. (Faraco et al., 2004).

## 4. MATERIAIS E MÉTODOS

### 4.1 Dados

A fonte de dados é proveniente do Projeto LATAM (Porto, 2007). As amostras foram analisadas para ICP/AES-MS em abertura de água régia. Foram identificados 51 (cinquenta e um) elementos em um total de 14263 (quatorze mil, duzentas e sessenta e três) amostras.

Os 51 elementos analisados foram: Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Ti, Tl, U, V, W, Y, Zn e Zr.

Quanto ao tratamento utilizado para limites de detecção, foi utilizado o recomendado nas instruções do pacote *compositions* no software R-project (Boggart and Tolosana-Delgado, 2013). Limite de detecção são os valores máximos e mínimos que o espectrômetro de massa consegue identificar, os valores alteram de acordo com cada elemento. O pacote *compositions* especifica que valores menores que o limite de detecção inferior do elemento devem ser transformados no valor negativo do limite de detecção do próprio elemento, e que os valores acima do limite de detecção superior devem ser transformados no próprio limite de detecção superior.

Após a análise dos elementos, foi feito um tratamento para eliminar as amostras provenientes de um mesmo ponto, resultando em 10235 (dez mil, duzentas e trinta e cinco) amostras.

Para maior objetividade no estudo, foram criadas duas áreas menores, uma com mineralização conhecida e outra sem mineralização, denominadas Target e Green, respectivamente (Braga et al, 2014).

Com estas subdivisões, os pontos contidos nelas somam um total de 808 (oitocentos e oito), dos quais 479 (quatrocentos e setenta e nove) na área Green e 329 (trezentos e vinte e nove) na área Target. A figura 4.1 apresenta as 10235 amostras e as duas áreas criadas para o estudo.

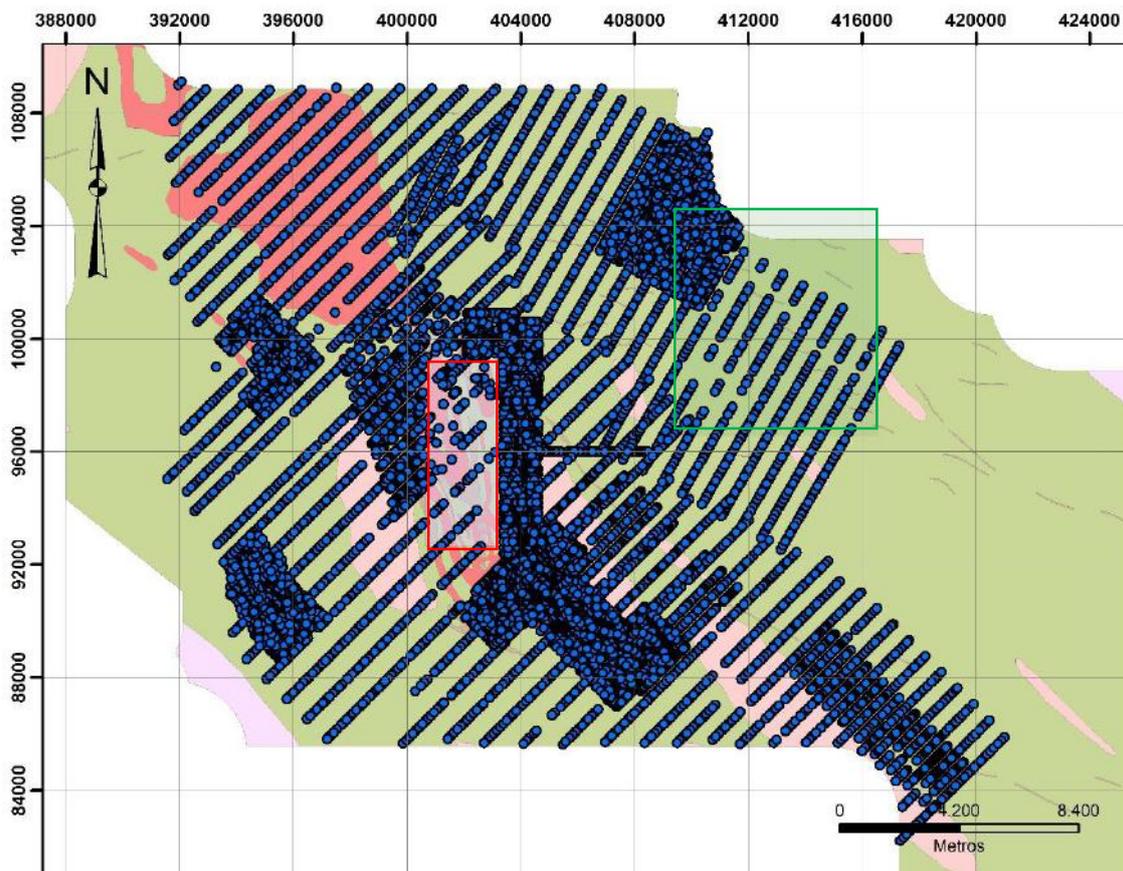


Figura 4.1 – Mapa de pontos da área, com as subdivisões Green, contornada em verde, e Target, contornada em vermelho.

## 4.2 Treinamento e Validação

Em uma abordagem genérica (Braga, 2005) para um problema de análise de dados podemos considerar as seguintes etapas: a) Definição do problema; b) Aquisição e avaliação dos dados; c) Extração de características e realce; d) Treinamento; e) Validação; f) Implementação; g) Prospecção. Nesta seção serão detalhadas as etapas d) e e).

Para o desenvolvimento do estudo foi necessário a criação de dois *subsets* utilizando os 808 (oitocentos e oito) pontos. Um *subset* para o treinamento e outro para a validação.

O *subset* de treinamento é utilizado para o ajuste de um modelo que se atenda ao objetivo do estudo. O *subset* de validação confirma se o modelo é válido ou não, caso não seja, volta-se ao *subset* de treinamento até um ajuste do modelo que se adeque aos objetivos. A figura 4.2 demonstra o ajuste e a validação de um modelo genérico.

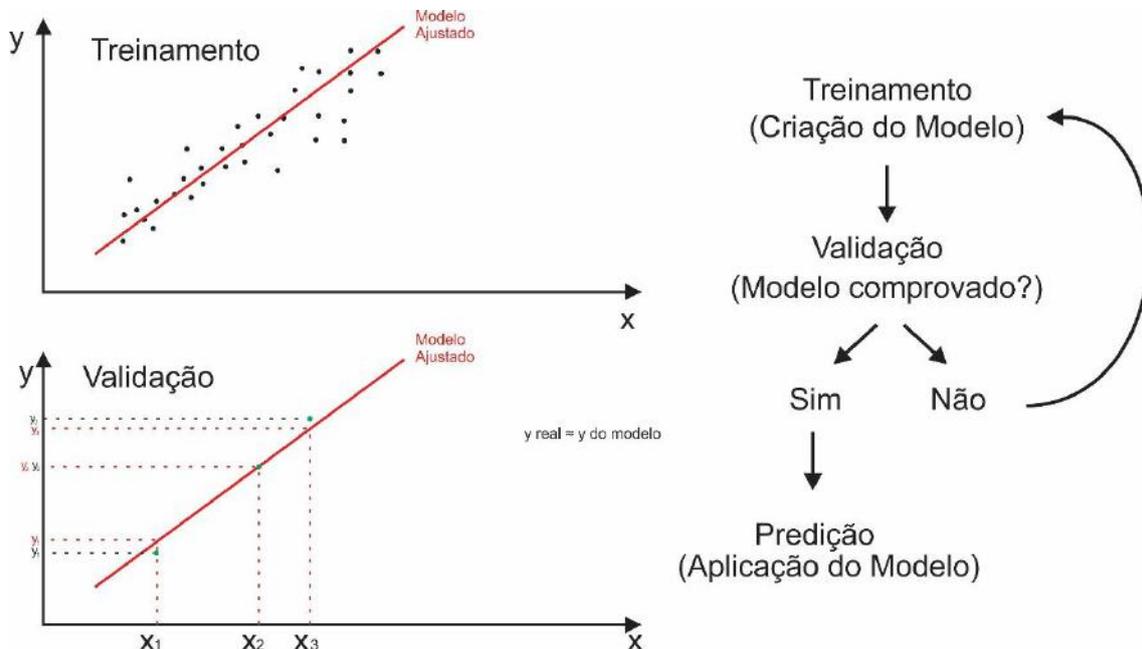


Figura 4.2 – Como funciona o sistema de treinamento e validação, onde a figura acima representa os pontos do *subset* de treinamento e a figura de baixo representa os pontos do *subset* de validação com o modelo criado no treinamento.

Em ambos os casos, tanto no treinamento quanto na validação, todas as variáveis são conhecidas, só assim é possível a verificação do modelo, diferente da predição, quando não se conhece todos os dados e o modelo é aplicado para inferir valores das variáveis com base nas que já se tem.

Utilizando o programa Microsoft Excel, separam-se as 808 (oitocentos e oito) amostras para os dois *subsets*, mantendo as proporções de pontos das áreas Target e Green, e tendo em mente que quanto maior o número de amostras no *subset* de treinamento, mais preciso será o modelo. Com isso o *subset* de treinamento tem 707 (setecentos e sete) amostras e o *subset* de validação 101 (cento e uma) amostras, ambos com a proporção de aproximadamente 59% (cinquenta e nove por cento) de amostras da área Green e 41% (quarenta e um por cento) de amostras da área Target.

### **4.3 Análise de Componentes Principais (PCA)**

A análise de componentes principais é um método para extrair variáveis significativas de um grande conjunto de variáveis disponíveis em um *dataset*. Com menos variáveis a visualização é mais compreensível. Uma componente principal é uma combinação linear das variáveis originais no *dataset*.

A representação gráfica de uma análise de componentes principais é um *biplot*, que é um gráfico composto por Componente 1 x Componente 2, que são os dois componentes com a maior variância e sem correlação entre eles. O *biplot* apresenta duas informações, *loadings* e *scores*, os *loadings* são os vetores que representam as variáveis, onde sua direção e magnitude representam a importância na separação de *clusters*, e os *scores* são pontos que representam a amostra em relação ao componente 1 e componente 2.

Os *scores* ajudam a interpretar a ocorrência de *clusters* entre as amostras, o que pode indicar similaridade entre grupos de observações.

O uso da PCA para dados geoquímicos tem sido útil em caracterizar processos geológicos como diferenciação de magma, alteração, mineralização e processos intempéricos (Birks, 1987).

#### 4.4 Análise Composicional

Amostras geoquímicas são comumente analisadas para dezenas de elementos e amostras de solo ou sedimentos são extraídas em grande número, o que gera uma grande quantidade de dados. Para obter o máximo de informação existem várias técnicas de análise multivariada que podem ajudar. Em muitos casos estas técnicas reduzem essa grande quantidade de dados para simples diagramas que frequentemente esboçam os principais *trends* geoquímicos e ajudam na interpretação, em alguns casos, de zonas de mineralização (Grunsky, 2007).

Dados geoquímicos são apresentados em proporção (% , ppm, ppb, etc.). Os elementos maiores de cada amostra somam 100%, por definição. De acordo com a transformação composicional são pontos de um simplex de n arestas sendo n o número de elementos.

Dados composicionais devem ser ajustados utilizando log-ratios (proporções logarítmicas). O log-centered ratio é útil porque preserva todas as variáveis de uma composição (Grunsky, 2007), esta transformação é feita pela formula:

$$\text{clr}(x) = [\log(x_1/g(x))] [\log(x_2/g(x))] \dots [\log(x_n/g(x))]$$

Onde  $x_i$  é um valor medido de um elemento da amostra (x) e  $g(x)$  é a média geométrica dos elementos de x.

## 5. DESENVOLVIMENTO

### 5.1 Eliminação de variáveis

O início deste processo passa pela eliminação da variável Au (ouro), já que o objetivo é identificar farejadores deste elemento na região, e pela confecção de gráficos básicos utilizados na estatística, *boxplots* e histogramas. Os *boxplots*, conhecidos também como diagramas em caixa, são a representação de uma variável e seus quartis, é composto por 4 ou 5 partes, *whisker* inferior (a linha que se liga a parte inferior da caixa) que corresponde os valores abaixo do primeiro quartil (25%), a caixa que corresponde os valores entre o primeiro e terceiro quartil (entre 25% e 75%), a mediana (50%) que é a linha dentro da caixa, o *whisker* superior que é a linha que se liga ao terceiro quartil e os *outliers*, que existem devido a uma limitação dos *whiskers* que só podem ter o tamanho máximo igual a 1,5x o tamanho da caixa, são representados por pontos acima ou abaixo dos *whiskers*. Os histogramas são gráficos em coluna de uma variável onde a largura da coluna simboliza um intervalo de valores e a altura da coluna representa a frequência (relativa ou absoluta) do intervalo.

Após a geração de *boxplots* e histogramas foi necessária a eliminação de algumas variáveis (elementos), pelo grande número de amostras com valores abaixo do limite de detecção ou por variáveis que não ajudariam a discriminação das áreas. A figura 5.1 mostra o histograma do elemento B (Boro) e nela podemos observar que a sua frequência absoluta no intervalo que representa os valores abaixo do limite de detecção é alta, assim o boro foi eliminado por ter muitas amostras abaixo do limite de detecção. A figura 5.2 é o *boxplot* do elemento P (Fósforo), como observado, os valores das medianas nas áreas Target e Green são próximas, assim dificultando a diferenciação entre as duas áreas, sendo este elemento também eliminado.

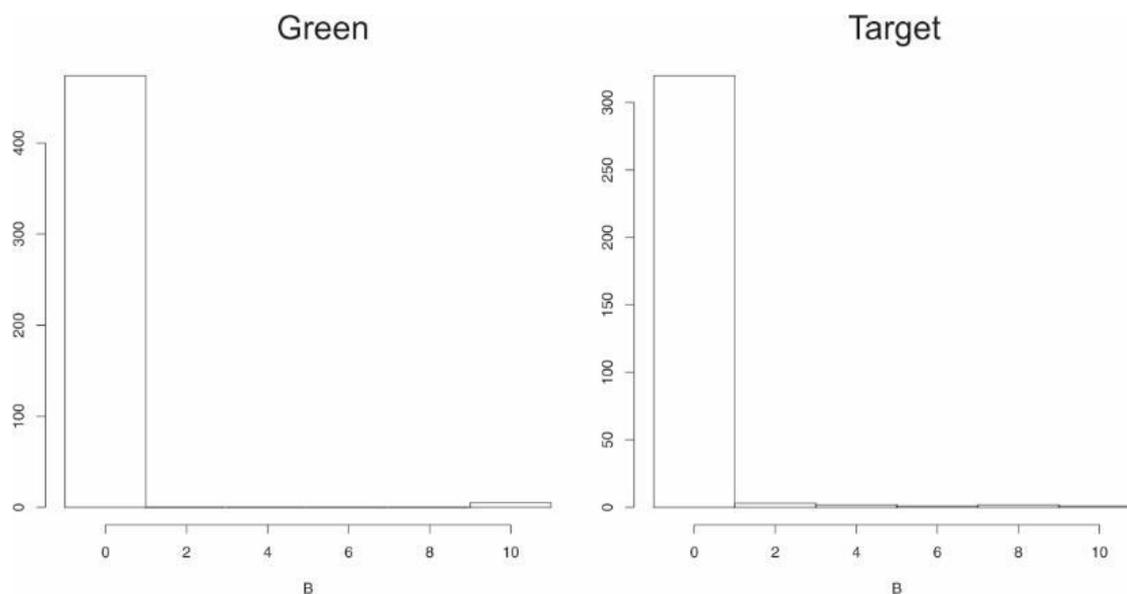


Figura 5.1 – Histograma do elemento B (Boro), indicando muitos valores abaixo do limite de detecção.

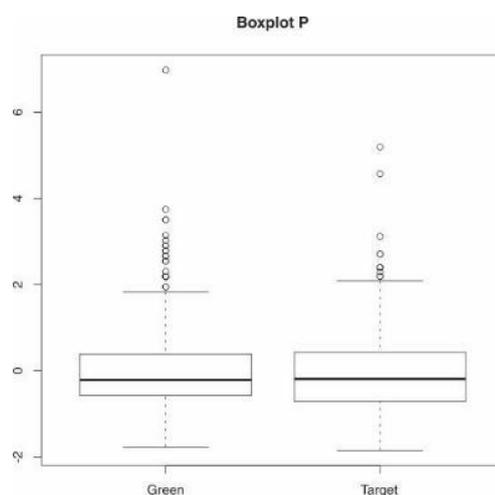


Figura 5.2 – *Boxplot* padronizado do elemento P (Fósforo), indicando elementos com medianas próximas, que não ajudariam a discriminação das áreas.

O processo de eliminação utilizando estes métodos retirou 19 variáveis (B, Ba, Ca, Cd, Ce, Co, Mg, Mn, Na, P, Re, Sb, Se, Sn, Sr, Ta, Te, Y, Zn), resultando em 31 elementos restantes.

## 5.2 Primeiro Treinamento

Aplicando a transformação em clr (centered log-ratio) às 707 amostras de treinamento para 31 elementos (Ag, Al, As, Be, Bi, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mo, Nb, Ni, Pb, Rb, S, Sc, Th, Ti, Tl, U, V, W, Zr) e posteriormente utilizando a análise de componentes principais, o *biplot* gerado pela análise pode ser observado na figura 5.3.

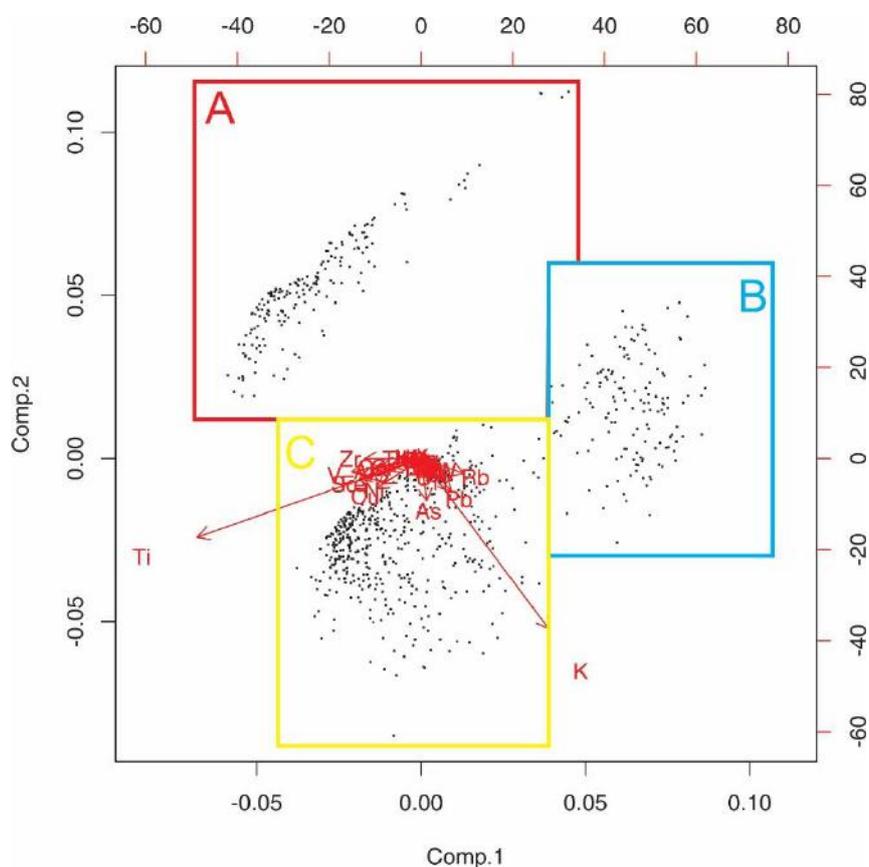


Figura 5.3 – *Biplot* indicando a separação das amostras em 3 *clusters*, sendo o B caracterizado pela forte influência de K (potássio), o C pela alta influência de Ti (tálio) e o A pela baixa influência de K (potássio).

Após ver que a análise composicional ajudou a separação das amostras em 3 *clusters* diferentes foi criado o mapa mostrado na figura 5.4 para uma visualização dos pontos em relação as áreas Target e Green e seus respectivos *clusters*.

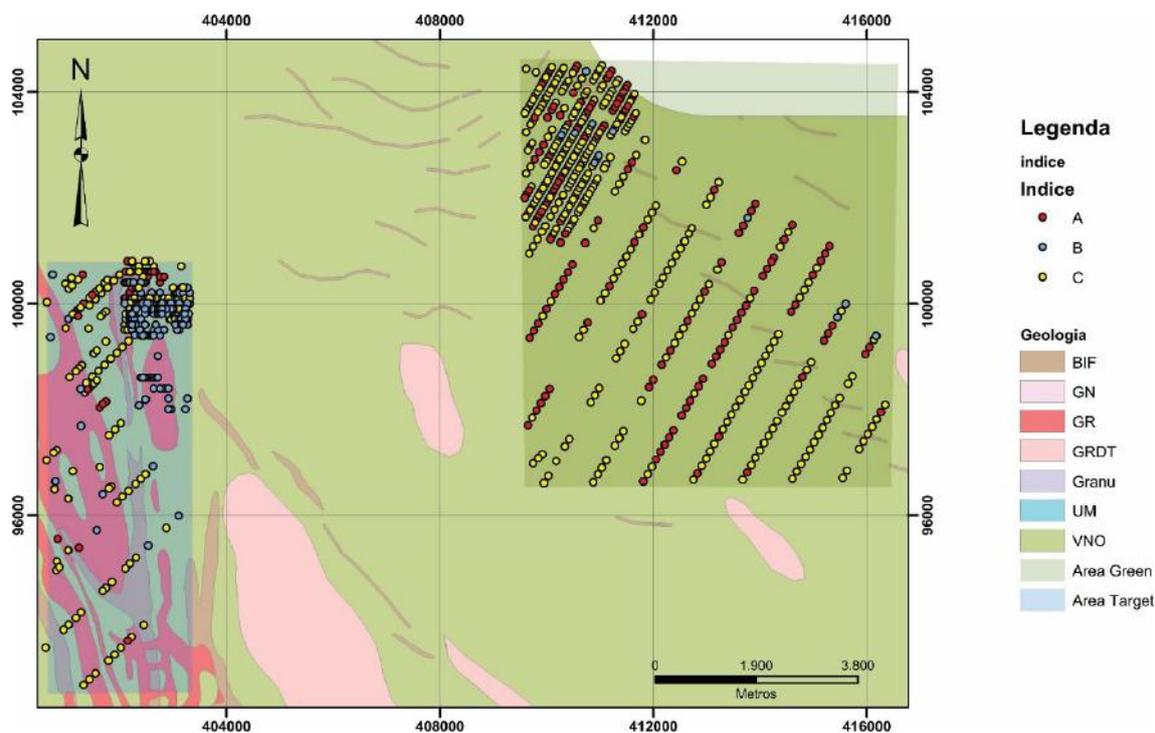


Figura 5.4 – Mapa de pontos de treinamento da região de Amapari com as litologias.

É importante notar que a área Target compreende granitóides e segundo Melo (2003) os granitoides são pós-mineralização, e já que o estudo trata de amostras de solo, não podemos garantir que o K (potássio) é elemento indicador de ouro e não de granitoides. Se fosse indicador de ouro, a maioria dos pontos deveria ser azul na área e não apenas na região NE da área Target.

### 5.3 Segundo Treinamento

O segundo treinamento consiste na eliminação de elementos que causaram interferência no primeiro treinamento, elementos indicadores de granitoides. Foram excluídas mais cinco variáveis, sobrando apenas 26 para o segundo treinamento, os elementos retirados foram: Be, Cs, K, Li e Rb.

Com os elementos indicadores de granitoides fora do modelo, foi utilizada novamente a transformação em clr e usada a análise de componentes principais, que resultou em um *biplot* diferente como mostrado na figura 5.5.

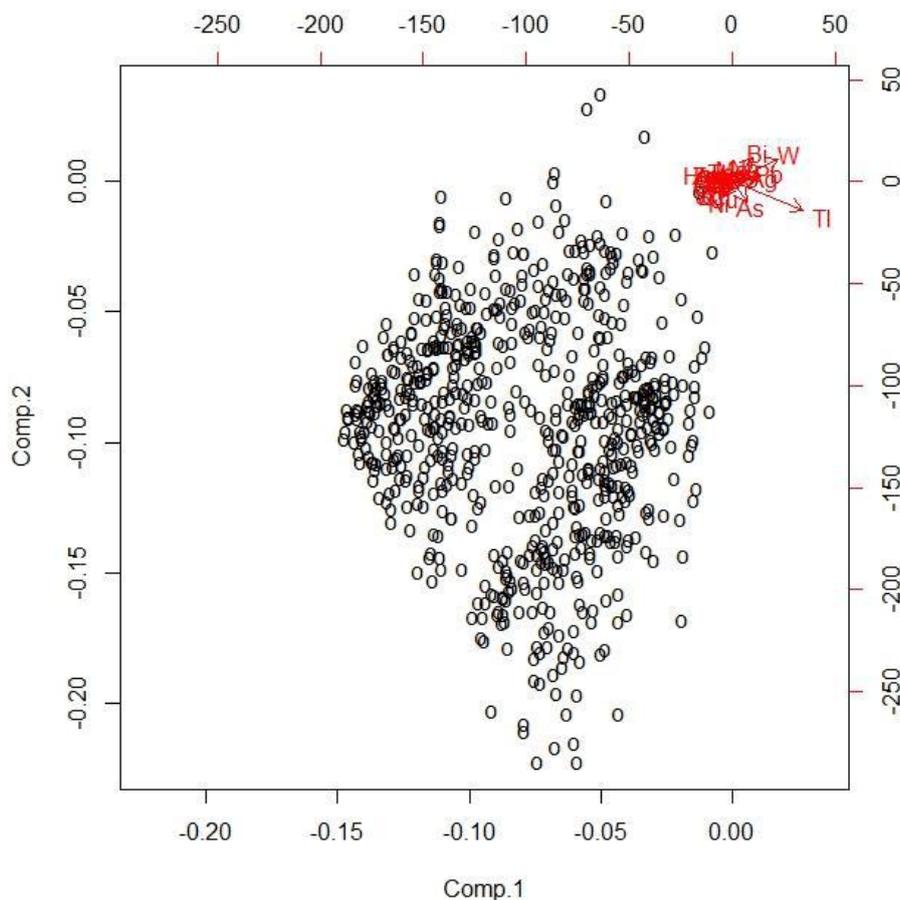


Figura 5.5 – *Biplot* do segundo treinamento, indicando dois *clusters*, sendo o cluster deslocado a parte inferior direita do gráfico sendo principalmente representado por TI (tálio).

Este novo gráfico mostrou uma separação menos clara entre os *clusters* da figura 5.3, porém é possível identificar uma separação em dois *clusters*. Como o gráfico da figura 5.5 não identifica com clareza os elementos (*loadings*), foram criados dois novos gráficos, um com os *scores* (amostras), outro com os *loadings* (variáveis). A figura 5.6 representa o *biplot* da figura 5.5 separando os *loadings* e os *scores*.

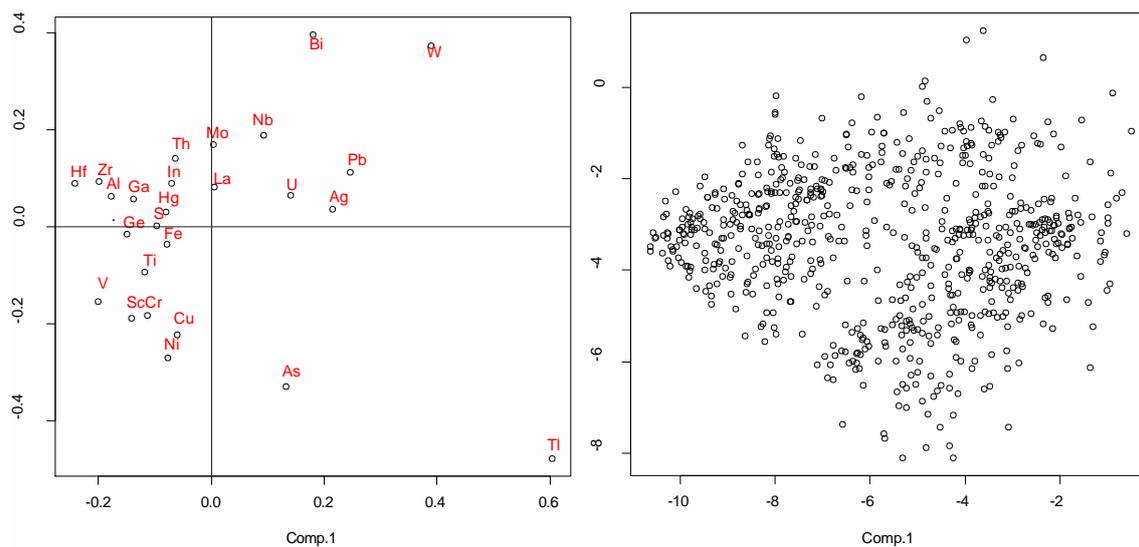


Figura 5.6 – *Biplot* separado, a esquerda os *loadings* das variáveis, e a direita os *scores* das amostras

Agora é possível identificar que os vetores relacionados ao Tl (Tálio) e As (Arsênio) são os responsáveis pela separação do cluster inferior direito. A figura 5.7 mostra o novo mapa de pontos gerados a partir dos dados gerados no segundo treinamento.

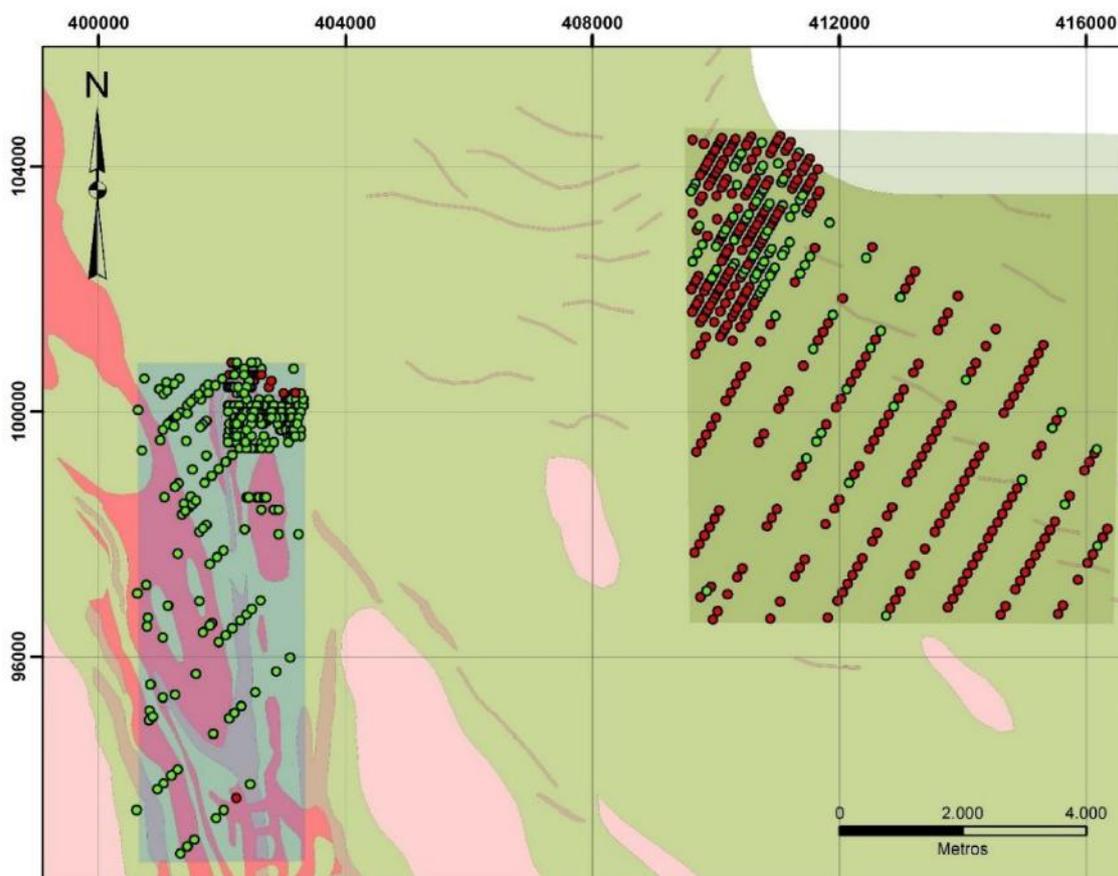


Figura 5.7 – Mapa de pontos do Segundo treinamento. Pontos verdes são relativos ao cluster inferior direito e pontos vermelhos do cluster superior esquerdo.

Como é possível observar no mapa, a separação dos *clusters* foi bem executada já que a maioria dos pontos verdes está localizada na área Target e a maioria dos pontos vermelhos na área Green. Isto também quer dizer que os melhores indicadores de ouro desta região são Tl e As. Na validação será visto se o modelo está ajustado corretamente.

## 6. VALIDAÇÃO

A validação é a comprovação do modelo criado a partir do treinamento. Pegamos os 101 pontos separados anteriormente para esta etapa e aplicamos o modelo criado, esta etapa não gera um gráfico, apenas uma previsão de qual cluster o ponto faz parte. Sabendo que a área Target representa as amostras mineralizadas e a área Green as amostras não mineralizadas, através do mapa de pontos e da matriz de confusão podemos comprovar o modelo criado no treinamento. A figura 6.1 é o mapa de pontos da validação.

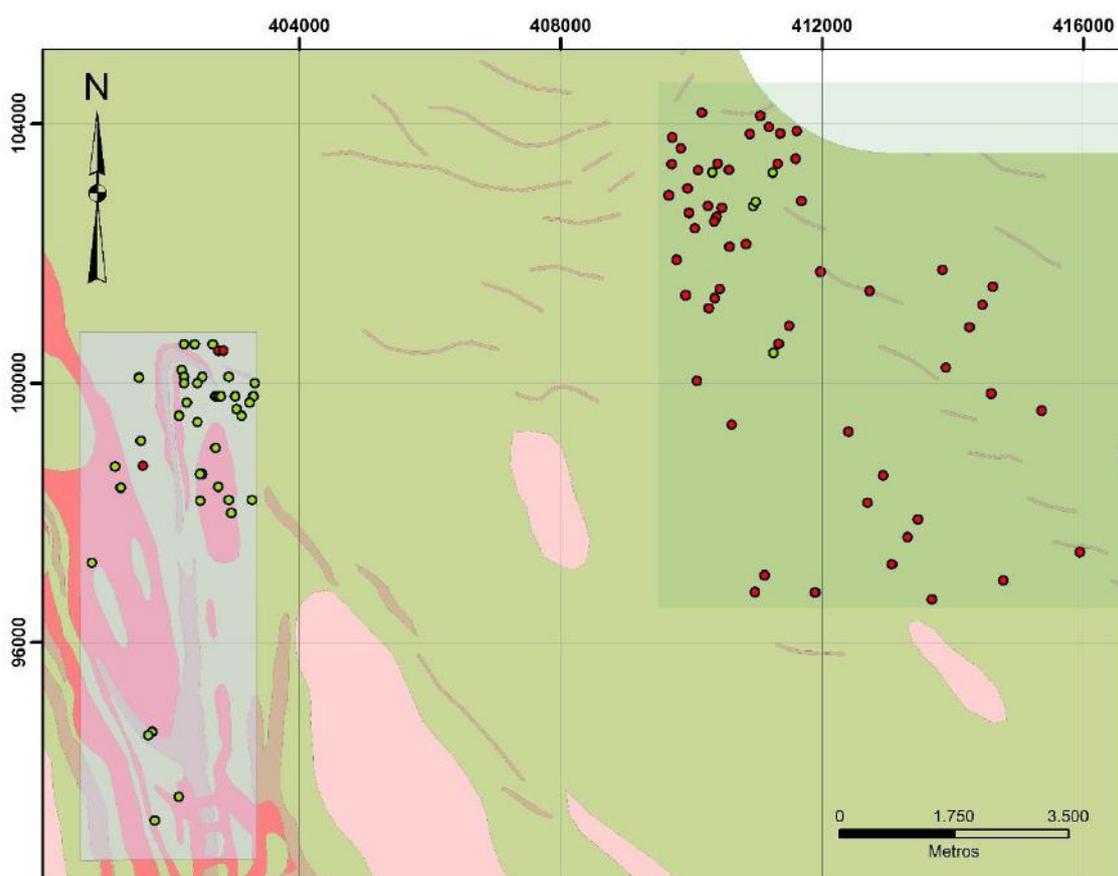


Figura 6.1 – Mapa de pontos da validação. Pontos verdes relativos ao cluster inferior direito da etapa de treinamento e pontos vermelhos relativos ao cluster superior esquerdo.

Como observado na figura 6.1, assim como no mapa de pontos do segundo treinamento (Figura 5.7), a maioria dos pontos verdes estão localizados na área mineralizada e a maioria dos pontos vermelhos, localizados na área não mineralizada.

O próximo passo é criar a matriz de confusão dos dados provenientes do processo de validação, uma matriz de confusão é criada classificando-se todos os casos do modelo em categorias (Verdadeiro Positivo, Verdadeiro Negativo, Falso Positivo e Falso Negativo), determinando se o valor previsto correspondeu ao valor real (<https://msdn.microsoft.com/pt-br/library/ms174811%28v=sql.120%29.aspx>). A figura 6.2 mostra a matriz de confusão, seus valores observados (real x modelo) e com destaque aos erros do modelo (Falso Positivo e Falso Negativo).

REAL \ MODELO	Target	Green	
Target	38	5	→ Falso Positivo
Green	3	55	↓ Falso Negativo

Figura 6.2 – Matriz de confusão da validação. Destacando os valores de Falso Positivo e Falso Negativo.

Na matriz de confusão foram observados os resultados de maneira mais expressiva, os valores de falso positivo indicam valores que o modelo identificou com mineralizados quando na realidade não são, e o falso negativo é quando o modelo identifica como não mineralizado uma amostra mineralizada. Colocando em porcentagem o erro do modelo foi menor que 8%, e o pior tipo de erro que se pode ter tratando de dados de mineralização, é o falso negativo, o qual o erro foi inferior a 3%.

## 7. CONCLUSÕES

Neste trabalho foi observado que a análise composicional é uma ferramenta a ser usada na área de geociências e que mostrou ótimos resultados, conseguindo identificar farejadores para este tipo de mineralização em uma dada região, como observado na seção 5.2, os elementos farejadores de Au (Ouro) nesta região são Tl (Tálio) e As (Arsênio).

Foi observado que somente o uso da ferramenta não é o suficiente, pois sem o conhecimento geológico da área e sem a percepção que os elementos indicadores de granitóides mascararam o resultado do primeiro treinamento, não seria possível chegar ao resultado do segundo treinamento e nem a comprovação do modelo na validação.

Também foi observado que a seleção prévia de variáveis com o uso de ferramentas estatísticas básicas contribuiu para o ajuste de um modelo de componentes principais mais conciso.

A análise composicional pode ser utilizada em outras áreas de geociências, não só de exploração, mas também para identificar associações geoquímicas em uma região, ou análise de sedimentos.

## 8. BIBLIOGRAFIA

BIRKS, H. J. B. 1987. **Multivariate analysis in geology and geochemistry: an introduction.** *Chemometrics and Intelligent Laboratory Systems*, v.2, n.1, p.15-28.

EVERITT, B. & HOTHORN, T. 2011. **An introduction to applied multivariate analysis with R.** New York, Springer, p.1-23.

GRUNSKY, E. C. 2007, **Statistical Analysis in the Geosciences**, Eolss Publishers, Oxford, UK, [<http://www.eolss.net>].

R Development Core Team. 2013. **R: A Language and Environment for Statistical Computing.** Vienna, Austria: the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Disponível online em: <http://www.R-project.org/>.

MELO, L. V. 2003. **Geological setting and mineralizing fluids of the Amapari gold deposit, Amapá State, Brazil.** *Géologie de la France*, 2003, n 2-3-4, 234-255.

LEEK, J. **The Elements of Data Analytic Style: A guide for people who want to analyze data.** Lean pub, 2015, p. 16-22.

PORTO, C.G. 2007. **Caracterização do regolito para a exploração mineral em terrenos lateríticos na Amazônia - Projeto Latam.** Relatório final de pesquisa, CNPq proc. 506292/2003-7. Dep. Geologia, Instituto de Geociências, UFRJ, 318 p.

VAN DEN BOGGART, K.G., TOLOSANA-DELGADO, R. **Analyzing compositional data with R.** Springer, Berlin, 2013.

CORDANI, U.G., SATO, K. 1999. **Crustal evolution of the South American Platform, based on Nd isotopic systematics on granitoid rocks.** *Episodes*, **22** (3), 167-173.

MONTGOMERY, C.W. 1979. **Uranium-lead geochronology of the Archean Imataca Series, Venezuelan Guyana Shield**. Contributions to Mineralogy and Petrology, 69, 167-176.

SPIER, C.A. 1999. **Petrologia e metalogenia dos depósitos de cromita associados ao complexo máfico-ultramáfico Bacuri, AP**. Brasília. Universidade de Brasília. Instituto de Geociências. 93 p. (Dissertação de Mestrado).

McREATH, I., FARACO M.T.L. 1997. **Sm-Nd and Rb-Sr systems in part of the Vila Nova metamorphic suite, northern Brazil**. In: South American Symposium on Isotope Geology, 1. Campos do Jordão. Extended Abstracts. SAS. p. 194-196.

LIMA, M.I.C., MONTALVÃO, R.M.G., ISSLER, R.S., OLIVEIRA, A., BASEI, M.A.S., ARAUJO, J.F.V., SILVA, G.G. 1974. **Geologia: Folha NA/NB. 22. Macapá**. Rio de Janeiro, Projeto RADAM, DNPM. p. 1-120. (Levantamento de Recursos Naturais, 6).

BRAGA, L. P., SILVA, F.J., PORTO, C.G. **Statistical Techniques Applied to Gold Exploration in the Amapari Area, Amazon Region, Brazil**. In: 15th Annual Conference of the International Association for Mathematical Geosciences, 2014, Madrid. Mathematics of Planet Earth - Proceedings of the 15th Annual Conference of the IAMG, 2014.

BRAGA, L. P. **Introdução à Mineração de Dados**, 2a edição. 2. ed. Rio de Janeiro: E-papers, 2005. v. 1. 212p.

**SQL Server 2014. Matriz de classificação (Analysis Services - Mineração de dados)**  
<https://msdn.microsoft.com/pt-br/library/ms174811%28v=sql.120%29.aspx>. Acesso em: 16.03.2016.

AITCHISON, J., EGOZCUE, J.J. **Compositional Data Analysis: Where Are We and Where Should We Be Heading?**. Kluwer Academic Publishers-Plenum. 2005.