1 INTRODUCTION

To establish a geochemical background, the first decision that must be made is of a conceptual nature, and concerns to the use and application of the background to be calculated. In its original proposition, the geochemical background was conceived as a value which separates the lower contents, considered as "normal", from the higher ones that could reflect the presence of mineral deposits.

Geochemistry Brasiliensis 34(2):161-175, 2020

Otavio Augusto Boni Licht
Programa de Pós-Graduação em Geologia
Universidade Federal do Paraná
Centro Politécnico - UFPR
Av. Cel. Francisco H dos Santos, 100
81530-000 - Jardim das Américas,
Curitiba, State of Paraná, Brazil
otavio.licht@gmail.com

RESUMO
O termo background foi introduzido na geoquímica aplicada à prospecção de depósitos minerais para designar um valor, abaixo do qual as amostras representariam teores normais e acima do qual seriam desvios da normalidade e poderiam representar anomalias relacionadas a depósitos minerais. Quando na década de 1980, as técnicas aplicadas na prospecção mineral, foram absorvidas pelas nascentes ciências ambientais, o termo background foi também incorporado com significado bastante semelhante. No entanto, desde sua primeira aplicação, esse termo tem sido adotado arbitrariamente, sem levar em consideração a enorme diversidade de variáveis subjacentes e que podem alterar substancialmente seu significado numérico. No presente artigo, são apresentados e discutidos os diversos componentes que participam desse “sistema complexo” conhecido como background geoquímico que englobam praticamente todos os conceitos da prospecção geoquímica, seja ela aplicada a atividades geológico-mineiras quanto ambientais, quais sejam o tipo de amostra coletada, os procedimentos de preparação da amostra, a fração que será analisada, a técnica analítica, o critério para a escolha do valor, a densidade amostral, a variabilidade espacial e as variações temporais.

ABSTRACT
The term background was introduced in geochemistry applied to the prospecting of mineral deposits to designate a value, below which the samples would represent normal contents and above which they would be deviations from normality and could represent anomalies related to mineral deposits. In the 1980’s, when the techniques applied in mineral exploration were absorbed by the nascent environmental sciences, the term background was also incorporated with a very similar meaning. However, since its first application, this term has been adopted arbitrarily, without considering the enormous diversity of underlying variables that can substantially alter its numerical meaning. In this article, the large number of components that make up this "complex system" known as geochemical background will be presented and discussed, and which cover virtually all concepts of geochemical exploration whether applied in geological-mining or environmental actions, such as: the type of sample collected, the sample preparation procedures, the fraction to be analyzed, the analytical technique used, the criterion for choosing the value, the sample density, the spatial variability and the variations over time.
levels of lack or excess of macro and micronutrients in soils and correcting them to increase the health of crops and their productivity in food supply. Thus, in agriculture, the background has a very particular conception, because it directly refers to the levels of tolerance of plant species to the excess and lack of nutrients.

At the end of the 20th century, a fundamental and emerging interdisciplinary branch of science received great interest from medical doctors, geochemists and geologists, namely Medical Geology (Komatina, 2004; Plumlee et al., 2006; Sahai et al., 2006; De Capitani et al., 2010; Selinus et al., 2010). Due to its typically inter- and transdisciplinary conception, immediately the Medical Geology adopted the technology of geochemical surveys to identify risk areas to human beings, given the abundance and lack of macronutrients, micronutrients and toxic chemical elements and compounds. In this very sensitive type of application, the background has a very special concept that is controlled by the dose-response curve of each chemical in humans, which establishes the needs for a good health and also, the levels of toxicity of the chemical elements and compounds. Each substance will have a reference level above which it will become toxic that depends on the substance, the exposure route, and the individual (Sullivan et al. 2001 apud Plumlee et al. 2006).

2 SAMPLING MEDIA

Geochemical surveys use some well-known sampling media to describe the distribution of chemical elements or compounds in an area. Each sampling medium (Figure 1) has properties, accordingly to its constitution, and capacities in terms of its representativeness and the influence it suffers from abrupt changes in local climatic conditions (Demetriades et al., 2018).

Soils are composed by mineral and organic compounds strongly influenced by the original rock. The deeper the soil and the closer to the parent rock, the more its chemical composition will mirror that of the original rock. As sampling media, soils are relatively stable and represent almost the location of the sampling point, except in those cases of highly grading topography. Due to their exposure and closer contact with atmospheric agents, the upper horizon suffers intensely from the deposition of solid particles that may have originated hundreds or thousands of kilometers away.

Stream sediments are composed of a mineral and organic load comprising a very diverse range of diameters and densities, whose transport is strongly controlled by the energy of the stream. Thus, climatic events such as droughts and intense rains can significantly alter the mineralogical and, subsequently, their chemical composition. Stream sediment samples do not represent the sampling point, but the hydrographic basin, comprised from the sampling point to the headwaters, and delimited by the water dividers that separate the basin from its neighbors.
Stream water samples, in turn, are extremely dependent as their composition significantly varies with sudden changes in weather conditions and with the influence of human activities. Like stream sediments, stream water samples represent a catchment basin and not just the sample collection point.

Solid particles suspended on stream water comprise extremely fine and light grains, mainly colloidal, made of organic matter, clay minerals and Fe-Mn hydroxides or diverse types of fine particles produced by human activity. As they almost float and are transported by stream water, they are extremely mobile and greatly influenced by sudden weather events. They also represent the catchment basin and not the sample collection point.

Solid particles floating in the air are composed of varied materials of natural origin or produced by human activity. Due to their very small grain and low density, they are extremely susceptible to changes in the direction and intensity of the wind. Under high-energy wind regimes, these particulates can be transported and deposited hundreds or thousands of kilometers from the source.

Less common sample media have been used in geochemical surveys directed to examine specific regions or situations, such as vegetation, moss, particulate organic matter, lake sediments, soil gas, ice cores (COKER et al., 1979; MILLER, 1979; NATIONAL GEOCHEMICAL RECONNAISSANCE, 1981; DUNN, 1989; BØLVIKEN et al., 1989; HORNBUCK, 1989; ROGERS et al., 1990; STEENFELT, 1990; BERNARDO et al., 2003; HAMMES et al., 2009; XUEQIU WANG et al., 1997, 2015; DUNN et al., 2020).

3 FRACTION

Samples collected from these different media are composed of materials with multiple grain sizes, with river sediments comprising a huge granulometric range from gravel to colloids, to river water and atmospheric air with solid particulate (Figure 2). It is well known that the finer the fraction, the greater the contact surface and the greater the availability of free sites for ion exchange and sorption. Thus, in general, the finer the fraction (e.g., clay minerals and colloids), the higher the metal content should be.

4 SAMPLE PREPARATION

The sample preparation that precedes the chemical analysis is a factor of great importance and control over the content of the elements that will be obtained in the laboratory. Crushing and grinding will mechanically destroy larger mineral grains into small, dust-like particles. Therefore, the elemental content will not only represent the ions absorbed by the exchange sites related to fine and colloidal fractions, but also the elemental composition of stable and almost inert minerals, such as oxides and silicates.

Figure 1

Sampling media commonly used in ordinary geochemical surveys, its surface representativeness and its susceptibility to sudden weather events. After BURENKOV et al. (1999)
In order to avoid this confusion between naturally fine components generated by rock weathering or by human activities, and those that have been made fine by crushing and milling coarse minerals, the simple sieving of soil and river sediments samples can separate the finer fractions, that is, those that are really important. This also applies to the separation of solid particles transported by running water and atmospheric air, directly on field station, using portable equipment with a manual vacuum pump and equipped with ultrafine filtration membranes.

![Diagram showing various types of materials and fractions that compose a soil or stream sediment sample and in which the chemical elements may be fractionated. The content of an element obtained when a "total" chemical analysis is applied will not make it possible to identify its fractioning. After JOHN; LEVENTHAL (2004).](image)

**Figure 2**

5 DIGESTION

The choice of the type of digestion to be applied in the samples for it chemical analysis is also a determining factor in the release of the chemical elements present in a sample. If it is necessary to obtain only the levels of elements associated with mineral species and colloids produced during the weathering process, very mild chemical attacks such as bi-distilled and deionized water, NH₄ acetate could be applied. If the interest is the elemental content adsorbed to colloidal organic matter, it should be digested by hydrogen peroxide. If the interest is in the elements composing secondary carbonates, NH₄ + acetic acid is an option. If the interest is in elements absorbed by Fe and Mn amorphous hydroxides, an alternative is to attack the sample with diluted NH₄OH, or EDTA + HCl. However, if Fe oxides are crystallized, concentrated NH₄OH, diluted HCl or aqua regia are valid options (Figure 3).

It is understandable that to release elements linked to a mineral species or a crystalline form will require specific attacks. If the content of the element was obtained with a very strong attack, the notion of this fractioning will not be possible. Nevertheless, for a first knowledge of the distribution of the elements and the characterization of promising areas and geochemical structures at different scales, mineral prospecting projects as well as the Global Geochemical Baselines project (DARNLEY et al., 1995) and similar projects that follow its methodology, very strong attacks have been increasingly used, like those known as tetracid or the fire-assay. After the interpretation of these results, the following investigations, aimed at the most diverse
applications, such as environmental diagnosis or monitoring, soil fertility, medical geology or ecotoxicology, may adopt analytical techniques designed to release only specific fractions of the “total” content of an element contained in a sampling medium.

When dealing with water samples, it is almost a routine to collect at least two vials in each sampling station, all of them containing filtered water. The vial containing water for cation analysis must be acidified and the other, for anions, will be preserved intact.

Solid particles suspended on stream water or on atmospheric air can be examined and analyzed after being released from the filtering membranes in which they were retained.

![Diagram](image)

**Figure 3**

Relationship between diverse geochemical extractions techniques and their ability to release mineral components from a geochemical sample (COHEN et al., 2010).

6 ANALYTICAL TECHNOLOGY

Another component of this complex system, which is the geochemical background, is the analytical technique to be adopted for the production of the elemental content in each sample.

It is important to remember that in the past, large areas were covered with geochemical surveys, with the production of huge databases. However, most of these analyzes were performed using techniques with considerably high detection limits, preventing the knowledge of the distribution of the lowest levels of many relevant elements. This is the case the optical emission spectrography, which has a lower detection limit of 200 ppm As. Thus, it is now clear that these data cannot be used for environmental diagnostic purposes or applied to medical geology researches.

Modern analytical technologies such as ICP OES, ICP-MS and ion chromatography are extremely suitable to all applications of geochemical surveys as they have the ability to detect a large spectrum of elements and compounds with very high sensitivity, resulting in very low detection limits (e.g., 0.2 ppm As).
7 STATISTICAL APPROACH

Another point of greatest relevance is the selection of the correct statistical technique to define the background. In the literature, the geochemical background is estimated from traditional statistical estimates like the arithmetic or the geometric mean, or else the median a robust statistical estimate, whether using numerical or graphical approaches (Reinmann; Filzmoser 2000; Salomão et al., 2019).

It is very common to find the background established as the arithmetic mean. This is based on the theoretical concept that the mean divides a data set represented by a unimodal normal distribution curve, into two equal parts. The statistical concept for the adoption of the arithmetic mean estimator is probabilistic, that is, given the configuration of the unimodal and symmetrical distribution curve, there is a probability of 50% that any value in a data set is greater or less than the arithmetic mean of that data set. Thus, the data set must obey the assumption of representing a population that follows a “normal” distribution (Figure 4 center).

But the most frequent situation in geochemical datasets is that of the histogram and the data distribution curve have a positive asymmetry, or a right tail curve (Figure 4 right). This means that there is a higher frequency or an abundance of low contents and less frequency or a shortage of higher ones. These high values may reflect natural phenomena such as different lithogeochemical types, the presence of mineral deposits, or be due to human action as scattered or point sources of pollution. When this happens, and it happens very often, the average is shifted to the right by the influence of the high values. As a result, the average it no longer divides the curve into two identical parts giving a wrong or biased estimate. In this case, the best estimator of the background is the median, considered robust, since it is not influenced by the high values that cause the asymmetry of the distribution curve.

Very rarely, the data may have a negative asymmetric distribution, i.e., a tail to the left, meaning an abundance of high contents and a shortage of low ones (Figure 4 left). Also in this case, the median will give an unbiased estimate of the background. Thus the decision between which technique should be adopted will depend on the results of the exploratory statistical analysis, especially the shape of the histogram and the distribution curve of the dataset (Figure 4). Each technique has its specific application and meaning, especially controlled by the statistical distribution of data and the presence of outliers, which may have been generated by a mineral deposit, specific lithotypes or even localized or dispersed sources of anthropic pollution.

However, even considering the confidence provided by the application of statistical techniques, it is important to remember they were built on assumptions of unimodality, that is, that the data are represented by a single population. It is necessary take into account that the geochemical signal identified by the analysis of the samples of a survey, can be a mixed value, representing the sum of several geochemical signals generated by different sources, ranging...
from a complex geology, to the contamination due to human activities.

In a situation like the one we see in Figure 5, what would be the meaning of the average or median calculated for the set of the three mixed subpopulations, since each of them must have a statistical behavior and, therefore, will have its own average and median?

Thus, under the strict statistical point of view, each geochemical signal that composes the geochemical framework of a surveyed area will probably constitute a subpopulation (Figure 6). When properly isolated each subpopulation will have its own statistical estimates and, subsequently, its specific background.

The isolation of subpopulations mixed on a geochemical dataset, would be achieved with a graphical technique like the probability plots modelling (SINCLAIR, 1974, 1976, 1991; SINCLAIR; BLACKWELL, 2004), which is based on finding inflection points which separate the subpopulations on a cumulated frequency curve. At the end of the process, the subpopulations will be isolated, and may be unimodal symmetrical or asymmetrical. Thus, they can be analyzed separately to obtain their own and most appropriate statistical estimates.

The isolation of subpopulations could also be achieved by numerical analysis like Jenks' natural breaks that involves multiple computational iterations seeking to separate
groups of data based on the optimization of the variance within the groups, by maximizing the variance between them (JENKS, 1967).

It is important to stress that, as emphasized by Darnley et al., (1995) “a geological map is not a substitute for a geochemical map. Lithological information on a geological map generally indicates the probable distribution of major elements, but inferences concerning minor and (especially) trace elements may be erroneous or unknown, with important consequences”.

8 NEW VARIABLES DERIVED FROM THE COMBINATION OF THE ANALYZED ELEMENTS

The improvement of analytical techniques, e.g., ICP OES and ICP-MS, led to the increase of the amount of elements and the increase of the sensitivity with very low detection limits. This have led the production of very accurate multielemental geochemical databases composed of dozens of elements, allowing a much more complete assessment of the geochemical characteristics of the surveyed areas. At the same time, the advance of multivariate statistical techniques, for example, factor and principal component analyses, allowed the data processing of these large databases, producing new variables resulting from the combination of highly correlated original variables (elements and / or oxides). In applied geochemistry, these factors or components, have received the generic name of geochemical signature (MCQUEEN, 2008, WANG KUN, et al., 2017, CRISIGIOVANNI et al., 2018, LARIZZATTI et al., 2018). Even considering the complexity of its constitution, when observed from a strictly numerical or statistical perspective, these new variables can and should be treated to characterize it specific background. This treatment must follow the steps and care established for the treatment of any original variable, be it an element or an oxide.

9 WOULD THE GEOCHEMICAL BACKGROUND BE A HORIZONTAL PLANE OR AN ONDULATED SURFACE?

In fact, considering the spatial variation in the distribution of chemical elements, the background cannot be represented by a horizontal plane. Instead, it should be considered as a wavy surface, which is geographically controlled by the location of the sampling stations and by the elemental content determined in the samples (Figure 7).

A good metaphor to be adopted is that of an Indian fakir’s bed, in which all nails have the same height, equally distributing the weight of the fakir that is lying. If a bed sheet is placed in this fakir bed to cover all the nails, this will result in a well-adjusted plane that will touch all the nail’s sharp ends. On the other hand, if each nail has a different height, it is logical that the body of the fakir will be hurt by those nails whose height exceeds that of the neighbors. If we apply this metaphor to a soil sampling grid, each nail will coincide with a sampling station and will have a specific height accordingly to the element content measured at that location, as for the elements shown in Figure 7. If we mathematically adjust a surface to the element contents in the samples of the grid, the surface should be calculated in such a way that the largest as possible number of “nails” touches the surface, not being too lower or crossing it. Thus, to calculate this wavy surface, it is mandatory to find the best algorithm that produces the minimum residue at each sampling station, which means the smallest difference between the original value and that calculated by the algorithm. If one uses an algorithm that does not respect these local inequalities and particularities, the calculated surface will incorrectly represent the natural fluctuations of the contents of an element, by producing residues in excess. The extreme situation would be to adjust a horizontal plane without considering the fluctuations and regionalization of the element's contents.

In this way, it is possible to visualize the two extreme points of a continuous series: at one end, a horizontal plane that would represent the average or the median of the contents, and not taking the regional fluctuations into account, and at the other end, a wavy and well-adjusted surface that considers the regionalization of the contents (Figure 8).
Two concepts of the geochemical background for the State of Paraná, with 200,000 km² and a huge geological and environmental diversity: (left) single and horizontal planes representing the average, median or other statistical estimate for six elements, which were calculated using all data generated by a regional geochemical survey; (right) well-adjusted surfaces, each one calculated using the best fit of the algorithm to the regional distribution of the six geochemical variables.

10 SAMPLE AMOUNT AND DISTRIBUTION

Another point of major relevance in any research planned to establish the background of an area or a given environment, concerns the quantity of samples collected and their spatial distribution. Even though there are statistical techniques to determine these parameters when designing a survey, this is not a strictly numerical issue, because it is important to take into account that geological, pedological or land use maps, which serve as the basis for geochemical surveys, are drawn from a limited number of field data and therefore cannot be considered as absolute truths.

Thus, within the operational and financial limitations, it is recommended to collect as many samples as possible so that the area is covered systematically, with regularly spaced samples, and without any preconceptions. Planned in this way, the survey will produce a necessary and sufficient geochemical database for the calculation of statistical estimators, or for reliable surfaces to be estimated, representing the background with the necessary reliability. As the base maps are being detailed and improved, the background can be recalculated for the new geological, pedological or land use units (DEMETRIADES et al., 2018).

The Low Density Geochemical Survey of the State of Paraná (LICH'T, 2001a) and the Soil Geochemical Survey of the State of Paraná (LICH'T; PLAWIAK, 2007), both designed to represent the elements distribution on the 200,000 km² territory of the State of Paraná, can be used as examples for this systematic spacing of sampling stations of two different sampling media (Figure 9).
The simplified geological map of the State of Paraná (above) and the spatial distribution of barium, analyzed in two different sampling media. It is perceived that the high levels of Ba$^{2+}$ (mg/L) in stream water, are mostly concentrated in three regions: (a) on the sandstones of the Caiuá and Bauru groups to the northwest, (b) a curved strip coinciding with the Passa Dois Group Paleozoic sediments in the center, and (c) on the rocks of the Paraná Precambrian shield to the east (left). On the other hand, the contents of Ba (ppm) in stream sediments coincide, in the central portion and following the same curved strip, with the Paleozoic sediments of the Paraná Basin, but are mainly and strongly concentrated on the region of the Paraná Precambrian shield to the east (right). These discrepancies in the distribution reflect the speciation and the form of occurrence of barium in the different geological environments and detected by the two and so diverse sampling media. Note: The coordinates are projected to UTM SAD69.

As for geochemical surveys on a detailed scale, the Geochemical Atlas of Curitiba (LICHT, 2001) designed to cover a 1: 100,000 topographic sheet, with approximately 2,500 km², comprising a large part of Curitiba, the capital of the State of Paraná, and some neighboring municipalities, as well as the Geochemical Mapping of the city of Rio Grande, in the state of Rio Grande do Sul (MIRLEAN; GRIPP, 2018) covering 12 km², well exemplify these concepts of number of sampling stations and the regularity of their spacing (Figure 10).
11 TIME SERIES

Under the view of mineral exploration, it is possible to say that if sampling, preparation, analysis and statistical techniques were properly applied, the background calculated or one population or one subpopulation will never change. If that is the case, a background value calculated today would not undergo significant changes within 10 years, if compared with the data of a new survey carried out with the same technology and covering the same area.

However, when the application of geochemical data is for environmental sciences and medical geology, the situation may be strongly different. The changes caused by sources of pollution, which did not exist at the time of the first survey, will substantially affect the geochemical framework of an area and consequently its background. Surface soils may have received significant loads of solid particles emitted by industrial sources or even distant natural foci such as volcanic centers.

12 FINAL REMARKS

The author hopes that the reflections made in this article may contribute to clarifying the concept of geochemical background. This reference level should not be adopted as a single and absolute truth, since it is the final product of a very complex system, derived and controlled by the combination of a large number of variables, such as amount of samples, density and spacing between samples, the sample medium, the type of sample preparation for analysis, the chemical attack or digestion, the instrumental analytical techniques and the statistical techniques for data processing. Thus, it is clear that in a given area, depending on the characteristics of the variables that make up this system, a geochemical background will be defined. If these variables are modified, even if it is just one of them, the final results will be different and another geochemical background will be established. This systemic approach to the background concept has an impact on various applications of geochemistry.

The integration of data produced by geochemical surveys aimed at mineral exploration is a feasible task only when the different surveys have adopted identical procedures in the selection of the sample medium, particle size, sample preparation technique and analytical technique. When these different surveys adopted different parameters, it is clear that the respective analytical results...
will not be comparable and will be practically impossible to be integrated.

The direct application of the analytical results produced by geochemical surveys aimed at mineral exploration to environmental sciences should be seen only as a general reference. In areas defined as representing possible or probable environmental risk, new geochemical surveys planned to adopt sampling media, particle size fractions, chemical attacks and specific analytical techniques will be necessary for a proper characterization of the impacts arising from anthropic activities.

The problem is even more complex when the final application is medical geology, because in this case, the background is no longer statistically defined, but due to the subtle limits between health and disease that are well established in the dose-response curves and the intensity and duration of exposure (the dose). If the dose is lower than required, this deficiency will cause disease, and conversely if the dose is higher than the tolerable amount, it will cause intoxication. Thus, the use of data generated by geochemical surveys aimed at mineral exploration only serves to delimit areas of potential health risk. In this case too, new geochemical surveys will have to be designed and performed by interdisciplinary teams composed at least by geochemists and toxicologists, and capable of detecting the chemical species and forms that can be absorbed by humans. In this case, it is essential to adopt analytical techniques of high analytical sensitivity, that is, that are able to reach very low detection limits to delimit and characterize health risk areas with the greatest possible accuracy. These surveys must be accompanied by epidemiological surveys to establish cause and effect relationships, and the extent to which the levels of the elements detected in the environment are related to the etiology of known and geographically well-defined diseases. On the other hand, a geographical alteration in the toxicological reference value for a potentially toxic element, will be a good indicator to investigate diseases still unknown in the researched region.

Lastly, even though they are currently used in geochemical and environmental literature, from an etymological point of view, the application of the terms geogenic and anthropogenic is a mistake. Ethimologically, anthropogenic means the genesis of man. Thus, its use in the sense of contamination produced by human activities is a severe mistake. It is preferable to use anthropic origin, or anthropic sources, or human-related sources. The same applies to the term geogenic that means genesis of the Earth. The use of this term in the sense of high levels related to lithologies or mineralizations, is also a mistake, being preferable to adopt natural origin or natural sources or geological origin or geological source.

13 ACKNOWLEDGEMENTS

The author acknowledge the colleagues Katia Norma Siedleki (Instituto Águas e Terras do Estado do Paraná, Brazil), Eduardo Salamuni (Universidade Federal do Paraná, Brazil), Allan da Silva Gomes (University of Queensland, Australia) and two anonymous reviewers of Geochimica Brasiliensis for their criticisms, comments and suggestions for the improvement of this article.

14 REFERENCES


LICHT, O. A. B. (2018) Multiscale, multimedia and multielementar geochemical mapping of...


WANG KUN, XIAO KEYAN, LI NAN, CONG YUAN, LI SHENGMIAO. (2017). Application of partial least squares regression for identifying
