International Journal of Geosciences, 2025, 16(5), 302-337
https://www.scirp.org/journal/ijg
ISSN Online: 2156-8367
ISSN Print: 2156-8359

Correlation of High Iron Content in Groundwater of the State of Bahia, Brazil with Climate, Lithology, Soil and Vegetation, Using Multivariable Analyses

José Carlos Cruz do Carmo[1], Deize Elle Ribeiro Moitinho[1], Iara Brandão de Oliveira[2]*

[1]Bahia Water and Sanitation Engineering Company (CERB), Salvador, Brazil
[2]Polytechnic School, Federal University of Bahia, Salvador, Brazil
Email: jose.carmo@cerb.ba.gov.br, deize.moitinho@cerb.ba.gov.br, *oliveira@ufba.br

How to cite this paper: Carmo, J.C.C., Moitinho, D.E.R. and Oliveira, I.B. (2025) Correlation of High Iron Content in Groundwater of the State of Bahia, Brazil with Climate, Lithology, Soil and Vegetation, Using Multivariable Analyses. *International Journal of Geosciences*, **16**, 302-337.
https://doi.org/10.4236/ijg.2025.165016

Received: August 2, 2024
Accepted: May 26, 2025
Published: May 29, 2025

Abstract

This work developed a statistical correlation between groundwater's high iron content in the four hydrogeological domains of the State of Bahia, Brazil, and the environmental attributes of climate, lithology, soil, and vegetation. From 3539 wells, flow test ≥ 1 $m^3 \cdot h^{-1}$, drilling period 2003-2013, 940 wells with high iron content (>0.3 mg/L) were used in this study. All groundwater samples came from new wells soon after the drilling, well construction, and a long pumping time for their development: 24 hours for sedimentary aquifers and 12 hours for karstic, crystalline, and metasedimentary aquifers. The application of Pearson and Spearman linear regression to seventeen physicochemical parameters (SPSS V.12) resulted in no correlations between iron and fourteen parameters, indicating no common origin between those parameters and iron. Only color and turbidity presented correlations > 0.20 with iron. After spatializing the 940 values of iron concentration (ArcGIS V.9) on the maps of each environmental attribute, grades 1 - 5 were given to the variables of each attribute based on the largest iron concentration value. The grades allowed the application of multivariable methods PCA and FA (SPSS V.12). The PCA indicated two factors explaining 59.52% of the total variance, closely attending the recommended minimum of 60%. The significant factor weights from the application of FA were: in Factor 1, soil, −0.71; vegetation, −0.68; and lithology, −0.52; and in Factor 2, climate, +0.74. Indeed, in the crystalline and metasedimentary domains with mafic-ultramafic rocks rich in iron, percentages of wells, 53.3% - 66.7%, occurred in iron-rich soils; of 49.8% - 59.8% in humid to dry forest and of 55.3% - 86.8% in humid to sub-humid climate. While, for the sedimentary domain (primarily sandstones) and karstic domain (carbonate rocks) poor in iron content percentages of wells, 80.9% - 100% occurred in iron-rich soils, 57.0% - 61.8% in humid to dry forest, and 58.6% - 62.4% in sub-humid to dry and semi-arid climate. These results indicated that, although lithology is a determinant for high dissolved iron content in the state of Bahia groundwater, this attribute alone (factor weight −0.52) cannot explain the whole phenomenon. The present work, using multivariable analysis with geospatial mapping of high iron content on top of environmental attributes, revealed the role of each environmental attribute in groundwater's high iron content. For the governmental drilling well company and its groundwater managers, this knowledge will result in better well locations and a reduction of both well and economic losses, as the long-term maintenance cost for the treatment process due to high iron content is prohibitive for rural municipalities.

Keywords

Groundwater High Iron Content, Multivariate Statistical Correlation, Environmental Attributes, State of Bahia, Brazil

1. Introduction

A variety of recent works applying statistical techniques to groundwater research involve groundwater geochemistry, water quality for drinking, domestic, and irrigation purposes, aquifer vulnerability, water quality evaluation, and more. The statistical techniques used were parametric and non-parametric statistical methods, multiple linear regressions, correlation analysis, hierarchical cluster analysis (HCA), principal component analysis (PCA), factor analysis (FA), geographical information system (GIS), geostatistics, analysis of variance and more. Thomas's work (2023) [1] evaluated the groundwater quality in Ibadan, Nigeria, using parametric and non-parametric statistics. The author considered it very successful in analyzing groundwater hydrochemistry using the quantitative methodologies: multivariate statistical analysis (MSA), hierarchical cluster analysis (HCA), and principal component analysis (PCA); besides the application of Kruskal-Wallis test, Pearson correlation, and the independent sample test to assess water quality. Ismail *et al.* (2023) [2] analyzed hydrogeochemistry and the groundwater quality for drinking, domestic, and irrigation purposes in West El Minia, Egypt. They evaluated the groundwater quality based on 49 samples, using principal component analysis (PCA), hierarchical cluster analysis (HCA), geostatistics, and spatial mapping to highlight the areas of health risks. Mehdi *et al.* (2023) [3] developed a multivariate statistical analysis to define the groundwater hydrogeological and physicochemical characteristics, the degree of water potability, and the irrigation suitability in a region of Hassi R'mel, Algeria. They applied the multivariate statistical techniques of (PCA), (HCA) and Diagram Analysis to a dataset of 17 wells and 12 chemical variables over the entire study area.

Patel *et al.* (2023) [4] published a comprehensive review paper for more than a hundred works (1980-2023), applying multivariate statistical techniques to groundwater research involving water quality for drinking and irrigation purposes, geochemical mobility, aquifer vulnerability, and more. Among the works the authors reviewed, one from Barkat *et al.* (2021) [5] assessed the hydrogeochemical evolution of the groundwater in Oued Souf Valley for drinking and irrigation purposes, applying multivariate statistical methods, geostatistical modeling, and water quality index. For physicochemical parameters, the Q mode clustering analysis detected the four major water groups. Applying Kriging for groundwater ionic constituents, the exponential semivariogram model fitted seven, while the rational quadratic semivariogram model best fitted nine other groundwater ionic constituents. Other work from Gaikwad *et al.* (2020) [6] addressed the geochemical mobility of ions in groundwater and the implication on groundwater quality in Maharashtra, India. They applied multiple linear regressions, correlation analysis, hierarchical cluster analysis (HCA), and principal component analysis (PCA) to 65 groundwater samples. In addition, the work from Machiwal *et al.* (2018) [7] demonstrated how statistical methods, from simple descriptive statistics to multivariate statistical analyses, incorporating explanatory variable, multiple linear regression, discriminant analysis, analysis of variance, and more in combination with groundwater quality data, hydrogeological data, and land use, are valuable to study aquifer vulnerability.

The application of multivariate statistical analysis by Gaikwad *et al.* (2020) [8] helped to elucidate aspects of the groundwater geochemistry and drinking water suitability in the Kudal region of India. In China, Wu *et al.* (2020) [9] used MSA to trace the sources and factors for groundwater pollution in a shallow urban aquifer of Yan'an City, while Li *et al.* (2019) [10] used to understand the hydrogeochemical processes occurring in the water of the Guohua phosphorite mine. Cloutier *et al.* (2008) [11] analyzed geochemical data to identify the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system in Quebec, Canada.

For groundwater studies in Brazil using multivariable statistical analysis, the literature presents the work of Barbosa Filho and Oliveira (2021) [12], which developed a groundwater quality index (GWQI) for the hydrogeological domains of the state of Bahia, Brazil. The GWQI development used principal component analysis (PCA), factor analysis (FA), and hierarchical cluster analysis (HCA) to fully define the five parameters that participate in the index and their importance or parameter weight. The GWQI was a successful index due to its ability to represent the groundwater quality of the entire state of Bahia, using a single mathematical formulation, the same parameters, and a single weight for each parameter. Gomes *et al.* (2020) [13] applied MSA for groundwater quality evaluation in the central-southern portion of the state of Bahia, Brazil, while Gomes and Cavalcante (2017) [14] used to explain the processes associated with the groundwater quality of Fortaleza, state of Ceará aquifers.

Due to the ability of the multivariable statistical analysis to elucidate aspects of the hydrogeochemical processes occurring in the aquifers, this work applied MSA to find the correlations between high iron content in groundwater with natural environmental attributes: climate, lithology, soil, and vegetation. The goal was to correlate the pattern of the spatial distribution of groundwater high iron content in the state of Bahia with the environmental attributes and their variables to elucidate the issue of which compartment comes from the high iron content in groundwater of a specific location.

Iron in groundwater is considered a minor constituent among the dissolved inorganics, with a concentration range of 0.01 - 0.1 mg·L^{-1}, controlled by the presence of elements in the soil and rocks, water flow, recharge, climate, depth and time (Freeze and Cherry, 1979) [15]. However, many regions of Brazil have reported high dissolved iron content affecting the groundwater quality (Oliveira *et al.*, 2004; Nascimento, 2008) [16] [17]. In the state of Bahia, groundwater with high iron content occurs in the four hydrogeological domains: crystalline, karstic, metasedimentary, and sedimentary. Carmo (2016) [18] stated that many wells perforated by CERB, the governmental well drilling company, were never completed and installed for public supply due to high iron content, with significant financial loss due to well abandonment since the long-term maintenance cost and treatment process is unfeasible for small communities or municipalities. Indeed, dissolved iron in groundwater is a challenge for the treatment operation. It causes scale and the development of colonies of rust bacteria in the supply networks, affecting human health, giving taste and odor to the water, producing stains on clothing and utensils, and interfering with industrial processes (Siqueira *et al.*, 2011; Ityel, 2011; Zimbres, 2002) [19]-[21].

Szikszay (2012) [22] developed a comprehensive study involving many research works (Schoeller, 1962; Carroll, 1962; Matthess, 1982; Perel'man, 1963; Davis, 1967) [23]-[27] on the geochemistry of water and the acquisition of its chemical composition, starting from rainwater until reaching the aquifers, as groundwater, while going through the geochemical spheres (atmosphere, biosphere, lithosphere). Rainwater from the natural environment has dissolved CO_2 in very small concentrations, 0.5 - 0.8 mg·L^{-1}, as atmospheric CO_2 corresponds to only 0.03% by volume of the air mass, and the small amounts of dissolved substances result in average pH = 5 (Schoeller, 1962) [23]. However, depending on place and temperature, the rainwater may dissolve more gases and substances from the atmosphere, produced by anthropogenic activities such as the combustion of gasoline, biochemical processes in soil and water, volcanic and geothermal activities; and the pH can vary (3.0 - 9.8) (Carroll, 1962) [24]. Large amounts of dissolved CO_2 in rainwater increase the water dissolving power, which is necessary to transform solid rocks in soils of different types. Among the most important compounds found in rainwater predominate Cl^-, SO_4^{2-} and Na^+; however, as the rainwater moves through the soil, the chemical composition change, predominating, HCO_3^-, H_2CO_3, CO_3^{2-} and Ca^{2+} (Schoeller, 1962; Matthess, 1982) [23] [25].

Two important factors for the groundwater chemical composition are the solubility of elements and compounds and the geochemical mobility of the elements (Perel'man, 1963) [26]. Soluble salts in rocks are sulfates, chlorides, and carbonates, and the most important ions in water are Ca^{2+}, Mg^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}, and HCO_3^{-}. The classification of ion geochemical mobility is very mobile (Cl^{-}, SO_4^{2-}); mobile (Si, P, K^{+}); moderately mobile (Ca^{2+}, Mg^{2+}, Na^{+}); little mobile and inert (Fe^{3+}, Cr^{3+}, Al^{3+}, Pb^{2+}).

Biological processes can directly or indirectly influence the water quality as the water moves through the soil (Davis, 1967) [27]. Direct effects come from cellular organisms using dissolved and suspended solids for their metabolism and releasing their waste into the water. Microbial metabolism is of great importance in groundwater quality because oxidation by organic processes proceeds much more quickly under optimal conditions of humidity and temperature. In this case, physical and chemical factors are relatively unimportant. Indirect effects come from increased CO_2 content in the soil atmosphere through root respiration, temporary withdrawal of nutrients by higher plants and microbial activity, and changes in soluble salts through microbial breakdown of insoluble substances. The biological transformations of materials usually accelerate geochemical processes. Thus, more effective transformations should occur in hot, humid climates.

When the groundwater with dissolved iron is pumped and exposed to the atmosphere, iron precipitates in the oxidized form (Fe^{3+}), producing color and turbidity in the water. Color and turbidity must be limited, respectively, 15 $mgPt \cdot L^{-1}$; 5 UT), according to the Health Minister Resolution 888/2021 [28], as they are parameters that indicate aesthetic quality for public supply water.

The present work, though academic, serves to address the lack of understanding of the governmental well drilling company and its managers about the geochemistry compartment of the state of Bahia from which the groundwater's high iron content originated (Carmo, 2016) [18].

2. Study Area: State of Bahia, Brazil

The study area is the entire state of Bahia, Brazil, located between the coordinates 38˚E to 46˚W of longitude and 9˚N to 17˚S of latitude, with an area of 567.295 km^2, being the largest northeastern state in terms of land area, and fifth in the national ranking (Bahia, 2003) [29]. The focus of the present work is the area encompassing the four hydrogeological domains: crystalline, karstic, metasedimentary, and sedimentary.

The iron distribution in groundwater of the state of Bahia is described below, as well as the impact of climate, lithology, soil, and vegetation on the dissolved iron.

2.1. Iron Distribution in Groundwater of the State of Bahia

The hydrogeologic database CERB (2014) [30] registered 5583 wells, drilling period 2003-2013, with flow tests either below or above 1 $m^3 \cdot h^{-1}$. From those, 978 (17.5%) presented iron content > 0.3 $mg \cdot L^{-1}$, the maximum recommended limit for human consumption from the Ministry of Health Resolutions 2914/2011 and 888/2021 [28] [31]. **Table 1** presents the distribution of wells with high iron content per domain.

Table 1. Number of drilled wells, 2003-2013, by hydrogeological domain, flow test below and above 1 $m^3 \cdot h^{-1}$, and percentages with high iron content.

Hydrogeologic Domains	N˚ of Wells/Percentage with Iron (>0.3 $mg \cdot L^{-1}$)	N˚ of Wells with Iron (>0.3 $mg \cdot L^{-1}$)/Percentage
Sedimentary	514/(29.6)	152 (15.54)
Karstic	941/(9.9)	93 (9.51)
Sedimentary/Crystalline	81/(46.9)	38 (3.89)
Crystalline	3372/(13.9)	468 (47.85)
Metasedimentary	675/(33.6)	227 (23.21)
Totals	**5583**	**978; (978/5583 = 17.52%)**

Table 1 shows the ratio of two numbers (wells with high iron content/total number of drilled wells) with the resulting percentage: for crystalline plus metasedimentary domains (695/4047 = 17.2%) and for sedimentary plus karstic domains (245/1455 = 16.8%); percentages quite similar to the high iron content for the entire state of Bahia, 17.5%. In the sample of 978 wells with high iron content, the crystalline domain has the largest percentage at 47.85%, and the karstic domain is the smallest at 9.51%.

2.2. Climate Types and Impact on Dissolved Iron in Groundwater

According to Freeze and Cherry (1979) [15], among the climatic factors for iron in groundwater, rainfall is predominant since it recharges the aquifers while carrying the chemical elements dissolved by rock weathering. The greater the rainfall, the greater the mantle of weathering, which favors the conditions for aquifer recharge, influencing the well production and the chemical characteristic of groundwater (Lima *et al.*, 2009) [32].

Figure 1 presents the map of the climate types of the state of Bahia, according to the Thornthwhite classification, and **Table 2** briefly presents the location and description of each climate.

Table 2. Climate types of the state of Bahia, average rainfall, and location.

Climate	Average rainfall (mm/year)	Location
Humid	1400 to 2600	Coast, range of 18 to 65 km wide
	1300 to 1600	Western region, range of 20 to 80 km wide
Humid to sub-humid	1000 to 1400	Parallel to the cost and the western region
Sub-humid to dry	800 to 1200; decreases towards the center of the state	Parallel to the coast, also at the center, in regions of high topography around the karstic terrains
Semi-arid	In some regions ≤ 600 or ≤800	Occurs in almost 70% of the State of Bahia
Arid	300 to 500 in just three months	State far north

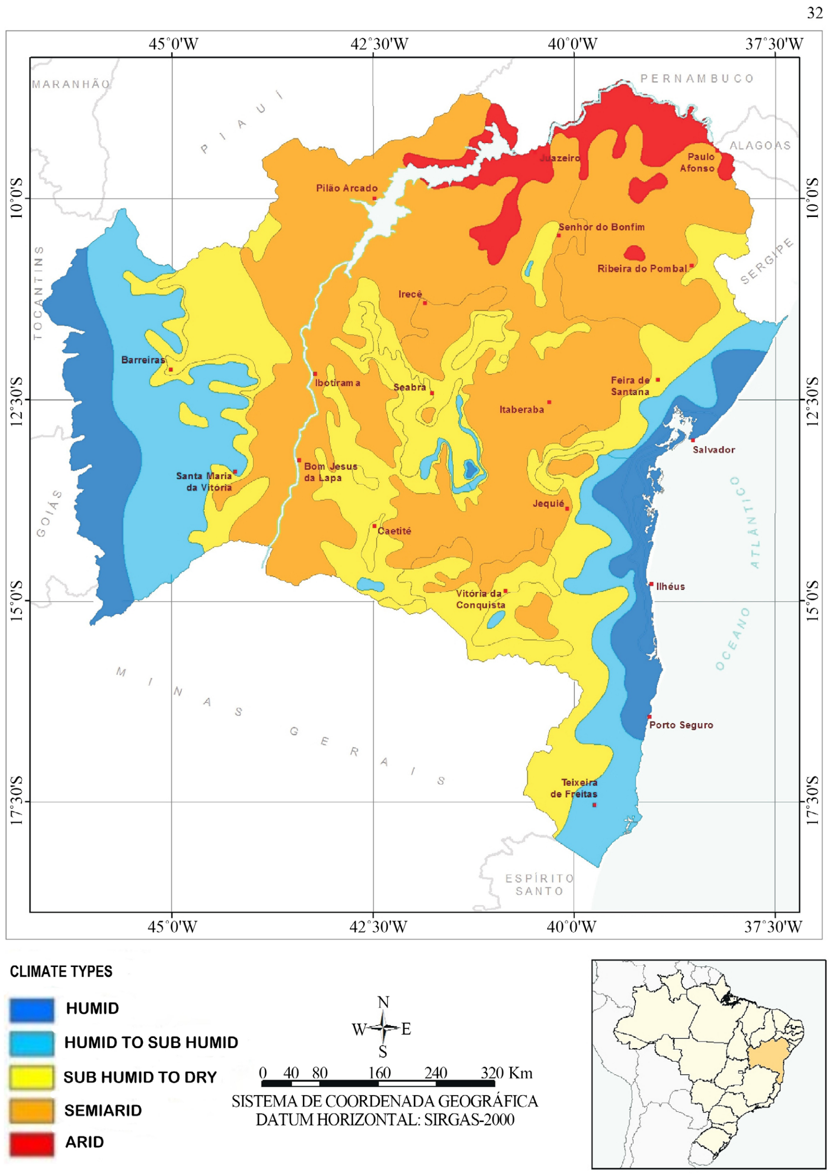

Figure 1. Map of the State of Bahia climate types. Source: modified from [29]. (http://www.inema.ba.gov.br)

Figure 1 shows that the humid climate is found in the western limit of the state, as well as in the coastal region, with average annual rainfall rates of 1300 - 2600 mm. The humid to sub-humid climate is found contiguous to the humid climate, with average annual rainfall rates of 1000 - 1400 mm. The sub-humid to dry climate runs parallel to the Atlantic strip, surrounding the Chapada Diamantina and at the edges of the Western Chapadões, with average annual rainfall rates of 800 - 1200 mm, decreasing towards the center of the state. The semi-arid climate occurs in approximately 70% of the state of Bahia, with annual rainfall rates of either <800 mm and <600 mm. The arid climate occurs in the state of Bahia's extreme north, with the highest temperatures and annual rainfall rates between 300 - 500 mm, concentrated in just three months of the year. Based on the climate impact on rock and soil weathering processes, climates with larger annual rainfall rates will have a significant correlation with groundwater's high iron content in the state of Bahia. On the other hand, climates with lower annual rainfall rates may have a weak correlation with high iron content in groundwater.

2.3. Lithology, Hydrogeologic Domains, and Impact on Dissolved Iron in Groundwater

Figure 2 presents eleven geological domains and respective lithology for the state of Bahia (Bahia, 2003; Guerra and Negrão, 1996) [29] [33]; and **Table 3** briefly presents the description of the hydrogeologic domains from Barbosa and Dominguez (1996) [34], based on **Figure 2**.

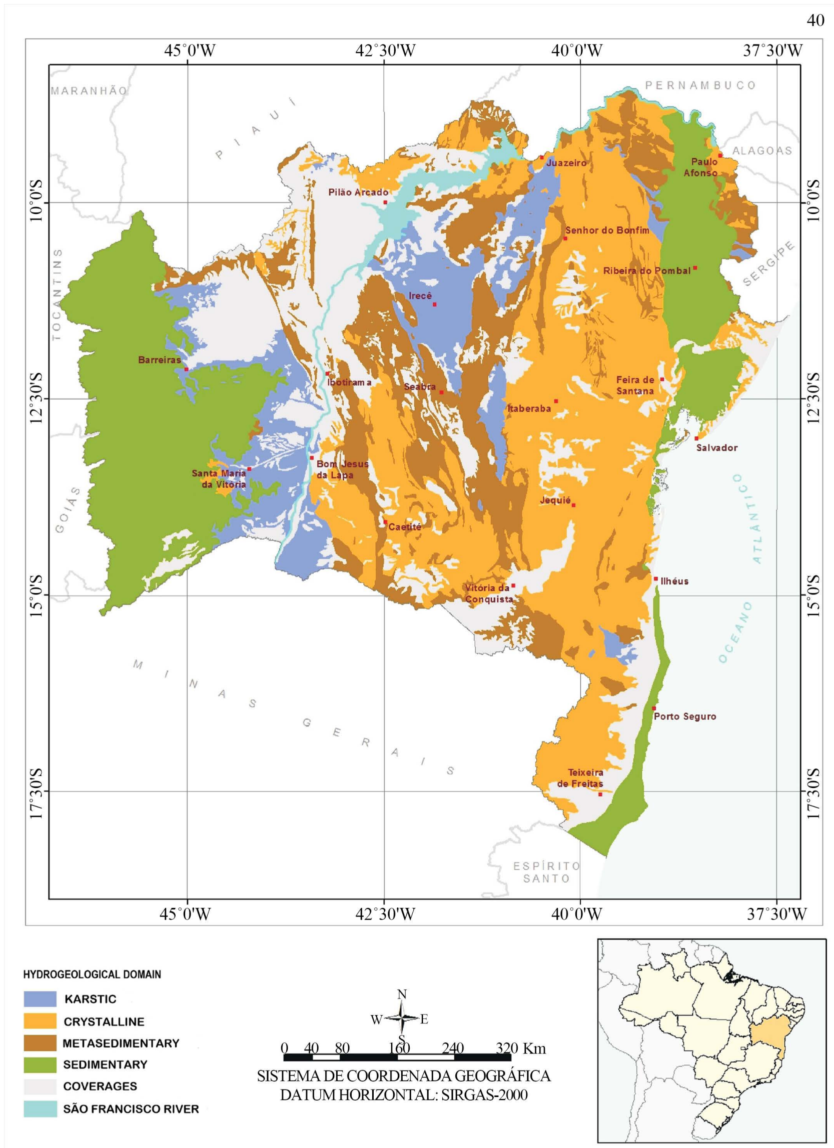

Figure 2. Map of the hydrogeologic domains of the State of Bahia. Source: Bahia (2003) [29], modified from Guerra and Negrão (1996) [33].
(http://www.inema.ba.gov.br, https://aguassubterraneas.abas.org)

Table 3. Hydrogeologic domains of the State of Bahia and description.

Domain	Description	State area %	Lithology	Aquifer characteristic
Sedimentary basins	State eastern region Reconcavo and Tucano (South, Central, North)	6.9	Predominance of Sandstone	Upper aquifer system with large capacity. Wells 450 m deep, flow rates ≥ 350 m^3/h
	State western region: Urucuia	16.3	Predominance of Sandstone	State largest groundwater reserve; high potentiality; excellent groundwater quality
Detrital covers	Shallow cover (dune and alluvial sands); Deep cover (Barreiras Formation)	15.0	Dune and alluvial sands; clayey sand sediments	Barreiras Formation: wells with 150 m deep, flow rates greater than 50 m^3/h
Crystalline	State central-eastern region from north to south	34.3	Large blocks of basic granulites, basalt, and gabbro, plutonic bodies	Shallow, fissural, or fractured free aquifers, low storage capacity, low permeability, heterogeneous and anisotropic characteristics. If rainfall ≤ 800 mm/year, the average flow rate is 3.41 m^3/h. If rainfall > 800 mm/year, the average flow rate is 4.0 m^3/h
	State north central region		Granodioritic and granitic rocks intercalated with amphibolite, mafic-ultramafic and calcium-silicate rocks	
Karstic	State central region: Una Group (area of ≤800 mm/year)	13.2	Base: glacial diamictites, top: carbonate lithofacies deposited in shallow marine and tidal plains	Free aquifers of high heterogeneity and anisotropy can store large volumes of water depending on rainfall
	State western region: Bambuí Group (area of ≤800 mm/year)		Up to 1000 m of carbonate rocks, diamictites, and other glaciogenic sediments	
	State north-northeast region: Vasa-Barris group and Miaba group		Up to 1,300 m of limestones and dolomites	
	State southern region: Rio Pardo group		Dolomitic marbles and meta dolomite, micaceous limestones, and pure quartzites	
Metasedimentary	Serra de Jacobina, 200 Km in a north-south direction	14.3	Quartzites, metaphellites; Greenstone Belt; mafic-ultramafic intrusions	Free aquifers of fissural and fractured nature occur in a region with rainfall from 690 to 1200 mm/year
	Chapada Diamantina, 100 Km north-south direction		Greenstone Belts; ultramafic, mafic to felsic spills; basalts; layers of ferrous formations	
	Northern Serra Espinhaço, 70 Km in a north-south direction.		Eastern flank with a metavolcanic sedimentary sequence; iron ore deposits economically exploitable	
	Southern end: Araçuaí Belt		Micaceous schists, calcissilicatic rocks, metarenites, and graphite shales	

The chemical composition of groundwater depends on lithology or geological causes. Primarily, depends on the soluble substances present in the rocks; secondarily, from reactions (concentrations, exchange of bases, reduction of sulfates, etc.). In the earth's crust, iron is the fourth most abundant element, slightly less than oxygen, silicon, and aluminum, and the second most common metal, slightly less than aluminum. A comprehensive literature review about this issue (Szikszay, 2012; Moruzzi, 2012; Lee, 1999) [22] [35] [36] indicates the following: 1) the most important minerals containing iron are: magnetite, hematite, limonite, siderite, ilmenite, and pyrite. 2) iron is present in minerals, mainly as insoluble ferric oxide, insoluble carbonate, biotite, pyroxenes, and amphiboles. 3) in crystalline rocks composed of silicates, the solubilization of minerals (quartz, olivine, pyroxene, amphibole, feldspar, mica, etc.) is quite difficult, occurring by hydrolysis and dependent on CO_2. 4) iron solubilization is related to redox reactions, with pH being crucial to forming different iron materials. 5) the ferrous iron released from the mineral with a carbonic acid attack dissolves, forming mainly hydroxides, which are soluble compounds. 6) When reaching oxidizing environments, ferrous goes to a ferric state, favoring insoluble precipitation and strongly coloring the water.

Below is a description of the hydrogeological domains of the state of Bahia according to Guerra and Negrão (1996) and Barbosa and Dominguez (1996) [33] [34].

2.4. Sedimentary Basins and Detrital Covers

The sedimentary basins and detrital covers occupy approximately 38.2% of the state of Bahia territory with the following distribution: eastern region, 6.9%; western region, 16.3%; shallow and deep detrital covers, approximately 15%. In general, these domains are composed of permeable sands and sandstones, in which the porosity greatly facilitates the chemical attack of the terrain, and the water has significant flow rates. The constituent minerals are preponderantly quartz, kaolinite, halloysite, illites, and smectites. Due to the lack of iron in its chemical composition, this lithology must have a less significant correlation with high iron content in the groundwater of the state of Bahia.

2.5. Crystalline Domain

The crystalline domain occupies approximately 34.3% of the state of Bahia territory with the following distribution: central-eastern region (north to south), north-central region, and south-southwest regions. The domain is predominantly composed of shallow, fissural, or fractured free aquifers with low storage capacity, low permeability, and heterogeneous and anisotropic characteristics. The overall description for the composition of the crystalline rock is the following: tectonic blocks composed of basic and acid granulites; iron formations and quartzites; intrusive plutonic bodies; granitic rocks; mafic-ultramafic and calcium-silicate rocks; marbles, formation iron bands, and graffiti; gabronorite, peridotite, and pyroxenite bodies; diorite and amphibolites; orthogneisses and granodiorites. Thus, the crystalline rocks of the state of Bahia, composed of iron-producing rocks, mafic-ultramafic rocks, and formation iron bands, can have a significant correlation with high iron content in groundwater. The comprehensive reviews on the leaching of crystalline rocks (Szikszay, 2012; Matthess, 1982; Schoeller, 1962) [22] [23] [25] indicate that the decomposition of diorite produces quartz, plagioclase, and amphibole, with the waters rich in iron and alkali metals, especially Na^+.

2.6. Karstic Domain

The Karstic domain occupies approximately 13.2% of the state of Bahia territory, with the following distribution: central region, western region, north-northeast region, and southern region. It forms free aquifers of high heterogeneity and anisotropy, capable of storing considerable volumes of water depending on the rainfall annual rates. The rocks of the karstic domains are glacial diamictites; carbonate lithofacies from shallow marine and tidal plains; carbonate rocks, ritmitites, oolitic limestones, calcarenites, and dark carbonaceous phyllites; dolomitic marbles and meta dolomite, micaceous limestones, and pure quartzites. Preponderantly, the constituent minerals of these rocks are calcite, dolomite, gypsum, and anhydrite. The review from Szikszay (2012) [22] indicates that dissolution of calcite and dolomite is more difficult, resulting in weak solid residues, most of which are (HCO_3^- and Ca_2^+). Thus, the karstic domain, rich in carbonate and associated rocks, typically a non-iron-producing lithology, may have a weak correlation with high iron content in groundwater.

2.7. Metasedimentary Domain

The metasediments occur in approximately 14.3% of the state of Bahia territory with the following distribution in the north-south direction: Serra de Jacobina, Chapada Diamantina, and Serra Espinhaço, in addition to the southern end. The metasediments form free aquifers of fissural and fractured nature, with lower salinization than the crystalline, in the presence of quartz-rich composition and higher flow rates in regions of high topography and rainfall rates (Carmo, 2016) [18]. In general, the rock composition is quartzites and metamorphic rock, greenstone belts, mafic-ultramafic intrusions, basalts, and layers of banded ferrous formations. In the eastern flank, there is a metavolcanic sedimentary sequence, as well as micaceous schists with calcissilicatic rocks, metarenites, and graphite shales. According to Szikszay (2012) [22], the leaching of mica-schist and schist rocks results in alkali metals and iron. Thus, the metasedimentary domain presenting iron-producing rocks (mafic-ultramafic intrusions; layers of banded ferrous formations and micaceous schists) can have a significant correlation with high iron content in groundwater.

2.8. Soil Types and Impact on Dissolved Iron in Groundwater

The rock weathering process is the starting point for soil formation (Brady, 1989) [37], with water being responsible for the rock dissolution by hydration and hydrolysis (Szikszay, 2012) [22]. Oxidation is another important phenomenon for soil formation, occurring while the water percolates the infiltration zone. The oxidation of sulfides, such as pyrite, results in iron oxide and sulfuric acid, which can attack limestone with the formation of sulfate. Iron oxides have the same basic composition (Fe, O, OH), with a variety of valence and crystal structures, producing different minerals: goethite, akaganeite, lepidocrocite, magnetite, and hematite (Schwertmann and Cornell, 2018) [38]. The review by Szikszay (2012) [22] indicates that during the rainy season, the water's downward movement predominates, and soil leaching occurs; during the dry season, when the upward movement predominates, the dissolved substances rise with the possibility of precipitation in the soil. These factors influence the formation of soils of different types and the chemical composition of the waters. Organic matter is another important soil component, with the dead matter in the upper part contributing to soil pedology and geochemistry. Aerobic microorganisms, such as bacteria, more abundant between 10 - 20 cm of soil depth, are essentially CO_2 producers and organic matter destroyers. The decomposition of organic matter by chemical or biological reactions, such as the respiration of roots and microorganisms, is the major source of CO_2 in soil water. Thus, the presence of iron in the soil is related to organic deposits of plant debris and can be associated with colloids or humus.

Figure 3 presents the map of soil types of the state of Bahia, modified from Bahia (2003) [29]. Based on **Figure 3**, **Table 4** presents the soil classes, their main characteristics, and the percentage of each soil distribution in decreasing order of participation in the state of Bahia.

Table 4. Soil types in the state of Bahia, characteristics, and percentage of participation.

Soil Class	Characteristic	Participation %
Latosols	Deep, medium-drained, clayey texture, low natural fertility. Occur in flat to wavy topography due to prolonged weathering, with strong disintegration of the rock matrix and the decomposition of the primary minerals. The predominance of iron and aluminum sesquioxides and 1:1 clays, all with a reduced specific area, cause low CEC, less than 17 cmol/kg.	39.40
Neosols	Soil poorly evolved without a diagnostic B-horizon, rich in mineral material, with organic material less than 20 cm thick. Excessively drained, low natural fertility.	22.96
Argisols	Deep soils in the moderate weathering stage, with clear differentiation between horizons. The superficial horizon presents a sandy texture, is well-drained, and has low natural fertility. The textural B-horizon presents at least 50% more clay. They occur in flat to gently undulating topography.	18.17
Planosols	They have shallow, medium to deep horizons A, B, and C. Occur in flat to smooth wavy topography. The subsurface B-horizon presents a high clay content, dispersed, and dense clay, with a massive or well-developed structure.	7.98
Cambisols	Soils consisting of mineral material with a shallow A-horizon followed by an incipient B-horizon. Present clayey texture with high natural fertility. They occur in soft wavy to wavy topography.	6.89
Chernosols	They present a texture medium to clayey, with gravel texture in some soils. It occurs in soft to wavy topography with natural grassland vegetation and is susceptible to soil darkening by the action of organic matter. The dark horizon, rich in organic matter, is very thick and with a high calcium content.	1.18

Continued

Luvissols	Present a textural B-horizon with clays of high activity. Its pedogenesis occurs concomitantly with the formation of iron oxides and the translocation of clay. Shallow, moderately acidic to neutral, poorly drained. They occur in soft wavy to wavy topography.	1.01
Gleissols	Presenting clay with a very clay texture, poorly drained, and formed under prolonged waterlogging promotes the reduction and removal of iron. Occur in flat topography, associated with depressions, river terraces, plains, and floodplains. It is a high content of organic matter.	0.93
Vertisols	Usually, it has a clayey to very clayey texture, with high plasticity and stickiness when soaked. Presently, there is high fertility, but with strong agricultural restrictions. Occur in very flat topography.	0.71
Spodosols	Present a light horizon with a very high percentage of sand on another extremely hardened dark horizon due to the alluvial transport to the B-horizon of colloidal complexes (organic matter + aluminum), which may not contain iron. The deeper layers have low natural fertility and acidic reactions.	0.47

Source: Bahia (2003).

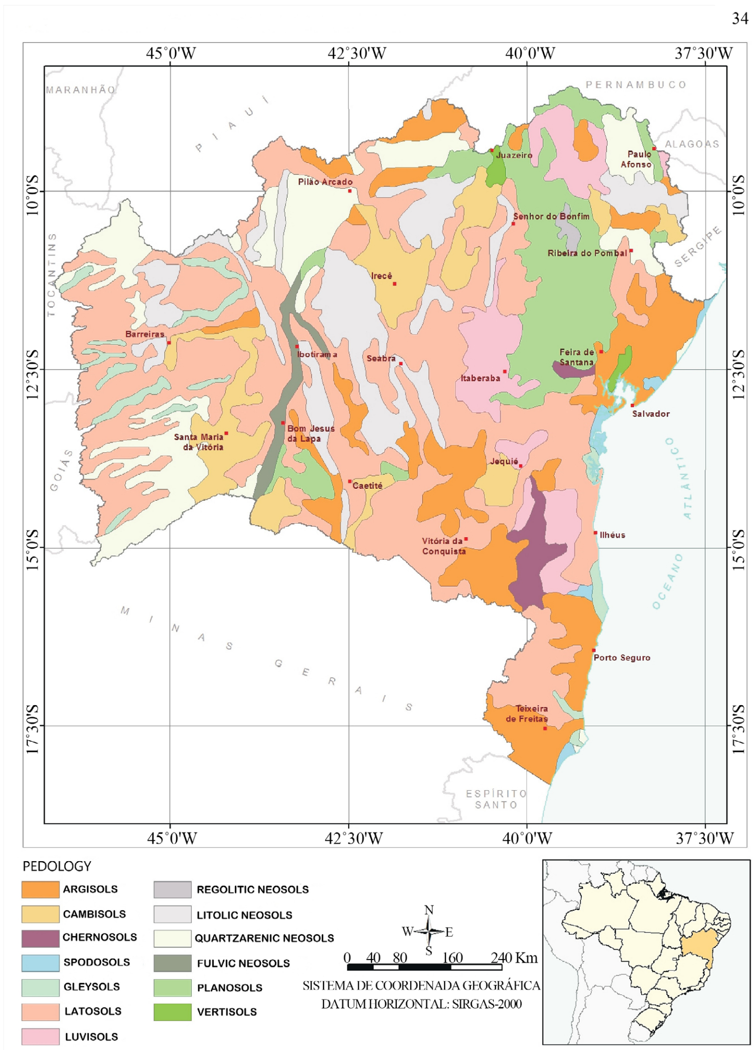

Figure 3. Map of soil types of the state of Bahia, modified from Bahia (2003) [29]. (http://www.inema.ba.gov.br)

Table 4 presents the soil classes in the state of Bahia and their main characteristics, as described in Bahia (2003) [29]. Latosols: deep soils, medium drained, clayey texture, with low fertility. Occur in flat to wavy topography, resulting from prolonged weathering, strong rock disintegration, and decomposition of the primary minerals. Have a predominance of iron and aluminum in their composition. Argisols: deep soils, sandy texture, well-drained with low fertility on top, and clayey B-horizon. Occur in flat to gently undulating topography from the moderate weathering stage. The combination of Latosols and Argisols encompasses 57.57% of the total soils occurring in the state of Bahia. The presence of iron in the Latosol indicates a significant correlation with high iron content in the groundwater of the state of Bahia.

Neosols: without a diagnostic B-horizon, excessively drained, low natural fertility, and low content of organic matter, and correspond to 22.96% of total soils in the state of Bahia. The characteristic of low natural fertility and low content of organic matter seems unfavorable for the production of iron, as presented above. Thus, this soil alone may have a weak correlation with high iron content in the groundwater of the state of Bahia.

Planosols: B horizon with a high clay content. Occur in flat to smooth wavy topography, corresponding to 7.98% of the total soils occurring in the state of Bahia. There is no indication of the presence of iron in the soil matrix, with a probable weak correlation with high iron content in the groundwater of the state of Bahia.

Cambisols: horizons A and B incipient, clayey texture with high natural fertility. Occur in soft wavy to wavy topography, corresponding to 6.89% of total soils in the state of Bahia. The characteristics of high natural fertility seem favorable for the presence of iron, with a probable significant correlation with high iron content in the groundwater of the state of Bahia.

Luvissols: shallow, poorly drained, with B textural and clay of high activity. Occur in soft wavy to wavy topography. In soil, pedogenesis is described as the formation of iron oxides (IO), which are insoluble substances resulting from aqueous reactions under various redox and pH conditions. Chernosols: medium to clayey texture rich in organic matter. It occurs in soft to wavy topography with natural grassland vegetation and is susceptible to soil darkening by the action of organic matter. The combination of Luvissols and Chernosols encompasses 2.19% of the total soils occurring in the state of Bahia. Because iron oxides are insoluble, these two soils together may have a weak correlation with high iron content in the groundwater of the state of Bahia.

The remaining soils are Gleissols (clayey texture, formed under prolonged waterlogging), Vertisols (clayey texture, with high plasticity and stickiness when soaked), and Spodosols (very high percentage of sand, B-horizon high in organic matter and aluminum, which may not contain iron); altogether, answer for only 2.11% of the total soils occurring in the state of Bahia and were not considered in this study.

2.9. Vegetation and Impact on Dissolved Iron in Groundwater

According to Brady (1989) [37], the weathering process contributes at the same time to the soil formation and the establishment of the local vegetation. The large amount of rainfall in the humid and sub-humid regions is responsible for the development of dense forests, with greater generation of organic matter and water retention in the soil. According to Schoeller (1962) [23], organic matter creates a reducing environment that is very important for groundwater quality. In water, the redox balance is influenced either by the inflow of O_2; by the ionic loads (Fe^{3+}, Mn^{3+}, SO_4^{2-}, and H^+); or by the consumption of O_2 by the reducing substances such as organic matter and ions (Fe^{2+}, Mn^{2+}, NH_4^+, H_2 and OH^-).

Figure 4 shows the map of vegetation of the State of Bahia and the six ecoregions, from east to west of the map: Coastal Forest, Inland Forest, Caatinga, Northeast Dry Forest, Chapada Diamantina, and Savannah. **Table 5** describes the vegetation characteristics and occurrence based on the description from Bahia (2003) [23].

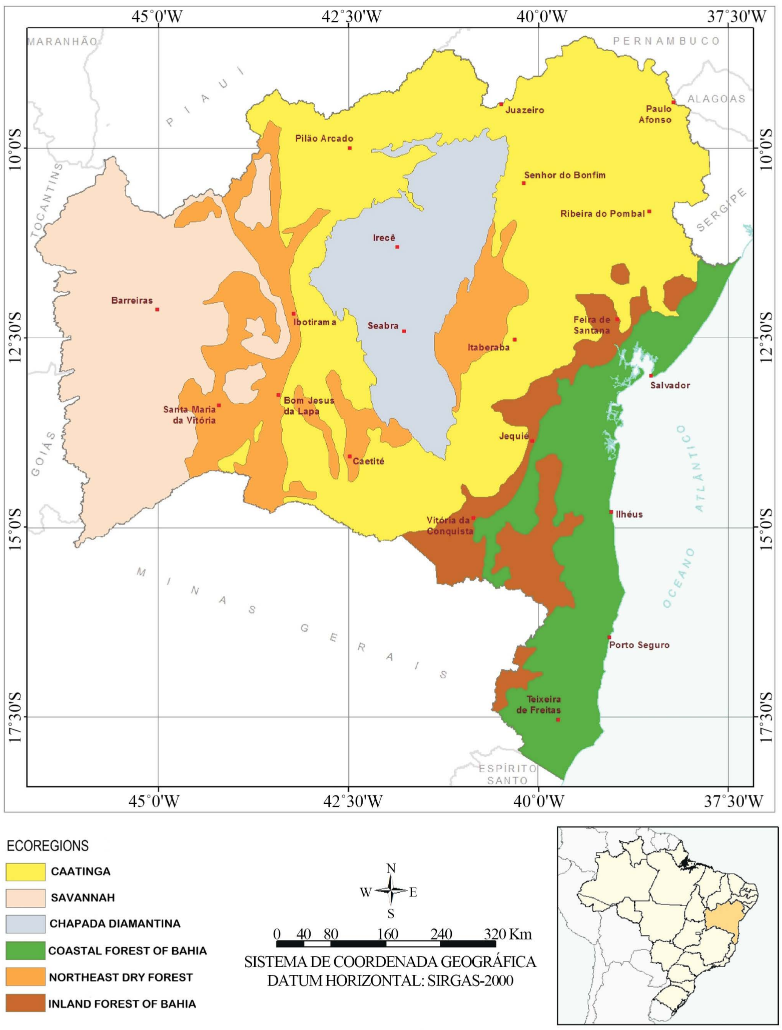

Figure 4. Map of the ecoregions of the State of Bahia, modified from Bahia (2003) [29]. (http://www.inema.ba.gov.br)

Table 5. Ecoregions of the State of Bahia, vegetation characteristics, and occurrence.

Ecoregions	Vegetation Characteristics	Occurrence
Coastal Forests	Dense Ombrophylous Forest with vegetation determined by tropical factors: high average temperature (25˚C), high rainfall rates (900 - 2000 mm/year), well distributed during the year. This evergreen forest has trees up to 40 meters high, because it likes much wetter and higher temperatures. However, nowadays, in many areas, the natural vegetation is strongly altered for urban development, agriculture, and livestock.	At the entire coastal region of the State of Bahia.
Inland Forests	Seasonal Semideciduous Forest with vegetation determined by dual climatic seasonality and precipitation that falls below 900 mm/year. The tropical climate has intense summer rainfall, followed by a severe drought, and the subtropical climate is without a dry period, but with physiological drought caused by intense cold winter, with average temperatures below 15˚C.	In a strip parallel to the coast, just after the domain of the Coastal Forests of Bahia.
Caatinga	Usually, very dry vegetation with thorns and very few leaves is formed by plants adapted to the dry climate, which exceeds five months of the year with little rainfall. However, the natural vegetation is strongly altered for agriculture and livestock, with the riparian areas suffering the greatest impacts, as well as swamps and floodplains, whose natural vegetation has been reduced for the expansion of agriculture.	In the semi-arid region of Bahia, it occupies about 50% of the territory.
Northeast Dry Forests	Deciduous Forest with vegetation determined by two well-demarcated climatic seasons, one rainy and one dry. Deciduous tree stratum predominates, is protected from drought by scales (cataphy or hair), and has adult sclerophyllous leaves or deciduous membranes.	At the border of the central part of the State and to the left of the São Francisco River.
Chapada Diamantina	The vegetation is exuberant, with species from the Caatinga and mountain flora species, especially bromeliads, orchids, and evergreens, and a mosaic of high biological diversity.	In the State's central area.
Savannah	The predominant vegetation is grassy plants, with sparse trees and shrubs isolated or in small groups, tortuous aspects, irregular branching, and thick bark trees. It is characteristic of leached sandstone areas with deep soils, occurring in seasonal tropical climates.	In the State's western region.

Source: Carmo (2016).

Based on **Figure 4** and **Table 5**, the main characteristics of the ecoregions of the state of Bahia are as follows:

Coastal Forests: dense ombrophylous forest, evergreen, that occurs in practically the entire coastal region of the State of Bahia, with vegetation determined by tropical factors: average temperature 25˚C, and rainfall rates 900 - 2000 mm/year, well distributed during the year. In dense forests, iron in ionic form tends to be bound to organic matter, favoring its enrichment in the soil (Richter and Azevedo Neto, 1991) [39]. Inland Forests: seasonal semi-deciduous forests that occur parallel to the coastal forests, with rainfall rates < 900 mm/year, and vegetation determined by tropical and subtropical seasonality, with the last climate displaying intense cold winter, average temperatures < 15˚C and physiological drought. These two ecoregions (Coastal Forests and inland Forests) have a tendency to concentrate iron in an ionic form bound to the organic matter in the soil, with a probable significant correlation with high iron concentration in the groundwater of the State of Bahia.

Northeast Dry Forests: occur surrounding the central part of the State and to the left of the São Francisco River, with deciduous forest and vegetation determined by two climatic seasons, rainy and dry. The deciduous trees are protected from drought by scales and deciduous membranes. This ecoregion may also have a significant correlation with high iron concentration in the groundwater of the State of Bahia.

Chapada Diamantina: This geographical region occurs in the central area of the State of Bahia and is characterized by dense forests, wetlands, Caatinga, and transition areas. This region has a predominance of creeping plants, exhibiting exuberant vegetation (aroeiras, barrigudas, carnaúba), species from the Caatinga (cactus, catingueiras, and medium-sized, twisted, and thorny trees), besides mountain flora. This ecoregion may have an important correlation with high iron concentration in the groundwater of the State of Bahia.

Caatinga: occurs in the central area of the state, in the semi-arid region of Bahia. The semi-arid climate of Brazil dominates a unique Brazilian biome, the Caatinga, a vegetation adapted to the dry climate, corresponding to approximately 50% of the territory. This region is marked by high temperatures, little rainfall in more than five months of the year, periods of drought, and a small hydrographic network, which is formed in general by intermittent rivers. The Caatinga is the richest and most diverse biome in Brazil, with a large number of endemic species. The soil is extremely fertile, but very shallow and stony. Savannah: Occurs in the state western region, marked by grassy plants, sparse trees, and shrubs isolated or in small groups. The characteristics of these two ecoregions (Caatinga and Savannah) indicate a weak correlation with a high iron concentration in the groundwater of the State of Bahia.

3. Materials and Methods

3.1. Selection of Wells and Groundwater Samples for Statistical Analysis

A total of 3539 wells were selected from a list of 5583 from the Hydrogeological Database (CERB, 2014) [30] of the Bahia Water and Sanitation Engineering Company (CERB), the governmental well drilling company. The wells were perforated in the period 2003-2013 in the four hydrogeological domains (crystalline, karstic, metasedimentary, and sedimentary) of the State of Bahia, Brazil. The 3539 wells have flow test $\geq 1\ m^3 \cdot h^{-1}$, satisfying the condition required for simplified groundwater supply systems. The final selection for this work was 940 wells out of 3539, with high iron content > $0.3\ mg \cdot L^{-1}$.

CERB data bank provided a comprehensive amount of hydrogeological data. Two spreadsheets with raw data were submitted as Supplementary Material: CERB_3539Wells_Iron_Data and CERB_940Wells_Data&Factors-Weight. The spreadsheet with 940 wells presents the groundwater hydrochemical data, well code and number, municipality, locality, coordinates, depth, static and dynamic levels, test flow, rock type, and physicochemical results. Calculations from this work were introduced in the spreadsheet.

The LABDEA/UFBA-Laboratory of the Environmental Engineering Department of the Polytechnic School and the Chemical Laboratory of SENAI-CETIND developed the groundwater physicochemical analysis. The analysis and detection limits follow the Standard Methods for the Examination of Water and Wastewater (Clesceri *et al.*, 1998) [40] and from ISO (International Standardization Organization). The analysis is performed at room temperature within a maximum of 15 days after sample collection.

The procedure for sampling and conditioning of the groundwater samples is as follows. During the well construction and after a long pumping process, water samples will be collected directly from the wells in 5000 ml plastic containers. The uninterrupted pumping time is 12 hours for wells in (crystalline, karstic, and metasedimentary rocks), and 24 hours in (sedimentary rocks). The physicochemical analysis encompasses total acidity, alkalinity (bicarbonate, carbonate, and hydroxide), total calcium, chlorides, electrical conductivity, apparent color, total hardness, total iron, fluorides, total magnesium, nitrates, nitrites, pH, potassium, reactive silica, sodium, total solids, sulfates, and turbidity. The physicochemical analysis is performed in triplicates, and the average results are evaluated.

3.2. Choice of Environmental Attributes for Multivariate Statistical Analyses

The choices of environmental attributes were climate, lithology (represented by hydrogeological domains), soils, and vegetation (represented by ecoregions). The choices followed the classical literature on water geochemistry, which states that the change in water composition from rain to groundwater comes from interactions with the geochemical spheres (atmosphere, biosphere, lithosphere). This work used maps (shapes) for the environmental attributes and the georeferenced wells from the Hydrogeological Database (CERB 2014) [30]. **Table 6** summarizes the description of the four environmental attributes for each hydrogeological domain. The description follows the direction (east-center-west) of the map of the State of Bahia (**Figure 2**).

Table 6. Summary of the environmental attributes by hydrogeological domains of the State of Bahia.

Hydrogeological Domains	Rainfall (mm/year)	Climate	Ecoregions	Soil
Sedimentary Basins: Southernmost	1100 - 2600	Humid; Humid to Sub-Humid	Coastal Forest	Latosols, Argisols
Reconcavo	1100 - 2600	Humid; Humid to Sub-Humid; Sub-Humid to Dry	Coastal Forest; Inland Forest	Argisols
Crystalline Domain: South	500 - 1600	Humid to Sub-Humid; Sub-Humid to Dry	Coastal Forest; Inland Forest	Latosols, Argisols and Chernosols
Sedimentary Basin: Tucano	400 - 500	Semiarid; Arid	Caatinga	Luvissols, Neosols and Planosols

Continued

Crystalline Domain: Center-North	400 - 500	Semiarid; Arid	Caatinga	Planosols
Crystalline Domain: Southwest	500 - 800	Semi-Arid	Caatinga; Chapada Diamantina; Inland Forest	Latosols, Argisols and Planosols
Karstic Domain: Irece	500 - 800	Sub-Humid to Dry; Semi-Arid	Chapada Diamantina; Northeast Dry Forest	Cambisols
Metasedimentary Domain	600 - 1100	Sub-Humid to Dry	Chapada Diamantina; Caatinga; Northeast Dry Forest	Neosols, Latosols
Karstic Domain: West	800 - 1100	Sub-Humid to Dry; Semi-Arid	Inland Forest	Cambisols
Sedimentary Basin: Urucuia	1100 - 2000	Humid; Humid to Sub-Humid	Savannah	Latosols, Neosols

Source: Carmo (2016).

Table 6 presents the following: Eastern Region: 1) sedimentary basins: Southernmost and Reconcavo, with maximum rainfall rate 2600 mm/year, humid climate, coastal forest, argisols and latosols, in which predominate iron and aluminum sesquioxides. 2) crystalline domain: South, with a maximum rainfall rate of 1600 mm/year, humid climate, coastal forest, and latosols. 3) sedimentary basin: Tucano, in the northeast, with a maximum rainfall rate of 500 mm/year with an arid climate. Caatinga ecoregion and luvissols are where iron oxides and translocation of clay occur. Overall, in the eastern region, the presence of large rainfall rates, humid climates, coastal forests, latosols, and mafic-ultramafic iron-forming rocks in the crystalline domain is relevant to explain the high iron content in the groundwater.

Central Region: 1) crystalline domain: Center-North, with maximum rainfall rate of 500 mm/year, arid climate, caatinga ecoregion, and planosols, with high clay content, dispersed, and dense clay. 2) crystalline domain: Southwest, with a maximum rainfall rate of 800 mm/year, semiarid climate, inland forest, and latosols. 3) karstic domain: Irecê, with a maximum rainfall rate of 800 mm/year, sub-humid to dry climate, Chapada Diamantina ecoregion, and cambisols, with clayey texture and high natural fertility. 4) metasedimentary domain: with a maximum rainfall rate of 1100 mm/year, sub-humid to dry climate, northeast dry forest, and latosols. In the central region, the significant rainfall rates, sub-humid to dry climate, inland forests, latosols, and mafic-ultramafic iron-forming rocks in the crystalline domain are relevant to explaining high iron content values in the groundwater.

Western Region: 1) karstic domain: West, with a maximum rainfall rate of 1100 mm/year, sub-humid to dry climate, inland forest, and cambisols. 2) sedimentary basin: Urucuia, with a maximum rainfall rate of 2000 mm/year, humid climate, savannah ecoregion, and latosols, in which predominate iron and aluminum sesquioxides. In the western region, the significant rainfall rates, humid and sub-humid to dry climate, inland forests, and latosols are relevant to explain the high iron content values in the groundwater.

3.3. Procedure to Define the Grades for the Environmental Attributes

In order to evaluate the influence of each environmental attribute, the values of the dissolved iron content for the 940 georeferenced wells were spatialized on top of the map of each environmental attribute, using ArcGIS Version 9 (ESRI 2004) [41]. The values of iron content served to assign grades 1 - 5 for the variable of each attribute. The lowest value of iron content represents the lowest influence (grade 1), and the largest value represents the largest influence (grade 5).

Figure 5(a) to Figure 5(d) present the variables of each environmental attribute, the graphs for iron content versus the corresponding assigned grades. The data are presented in (CERB_940Wells_Data&Factors-Weight) and submitted as Supplementary Material.

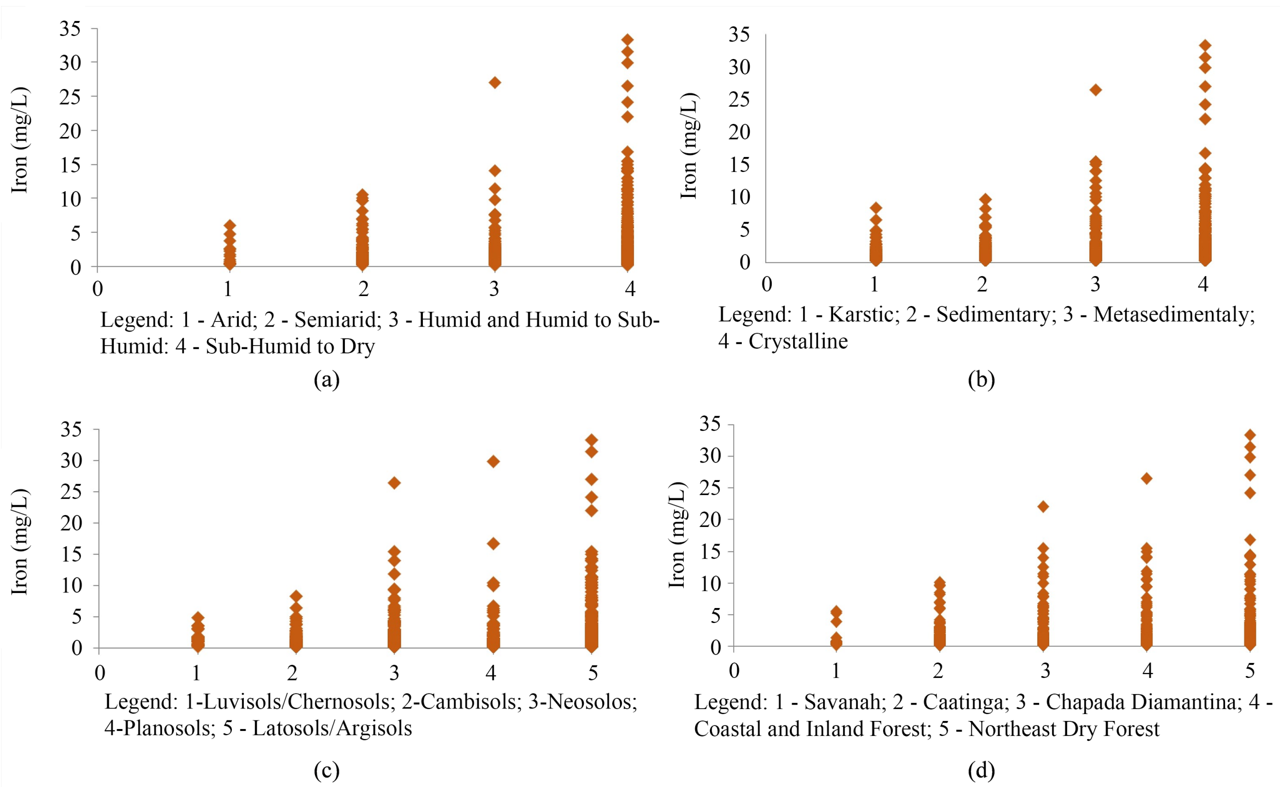

Figure 5. (a) High iron content in the 940 wells versus the variables of climate. (b) High iron content in the 940 wells versus the variables of lithology. (c) High iron content in the 940 wells versus the variables of soil. (d) High iron content in the 940 wells versus the variables of vegetation.

Highlighting the influential variable for the largest values of iron content: Figure 5(a) for climate shows the sub-humid to dry climate with grade 5. The reason should be that higher temperatures accelerate the chemical reactions and iron liberation. Figure 5(b) shows the crystalline domain with grade 4 for lithology. The reason is the predominant mafic-ultramafic rocks and the formation of iron bands. Figure 5(c) shows, for soil, the latosols/argisols with grade 5. The reason is the predominance of iron and aluminum sesquioxides in the latosols. Figure 5(d) shows vegetation in the northeast dry forest with grade 5. The reason should be the abundant organic matter in the soil and less aerial organic matter, favoring the decomposition and leaching of iron into groundwater.

Highlighting the influential variable for smaller values of iron content: Figure 5(a) for climate shows the arid climate with grade 1. The reason is the low average rainfall of 300 - 500 mm/year during only three months of the year, disfavoring chemical reactions and iron liberation. Figure 5(b) shows the karstic domain with grade 1 for lithology. The reason is the predominance of carbonate rocks, typically non-iron-producing rocks. Figure 5(c) for soil shows the Luvissols/Chernosols with grade 1. The reason is that iron oxides are insoluble, B textural is poorly drained due to high activity clay, and medium to clayey texture is typical for natural grassland vegetation. Figure 5(d) for vegetation shows the Savannah with grade 1. The reason is that grassy plants, sparse trees, and shrubs isolated or in small groups have less organic matter in the soil and less decomposition and leaching of iron to groundwater.

The grades 1 - 5 given to the environmental attributes variables allowed the application of multivariable methods PCA and FA (SPSS V.12) [42].

3.4. Multivariable Statistical Analysis

The Multivariable statistical analysis (MSA) encompasses the methods of principal component analysis (PCA) and factorial analysis (FA). The PCA condenses the information from the total number of variables into a smaller set of statistical variables, with a minimal loss of information, using a linear transformation from a p-dimensional space to a k-dimensional space, with $k < p$ (Lever *et al.*, 2017; Pérez-Arribas *et al.*, 2017) [43] [44]. The PCA has the objective to calculate the eigenvalues, the percentage of variance explained by each variable, and the cumulative variance indicating the optimal number of factors for extraction. The eigenvalues represent the variability of each component and the percentage of variance explained by each one. The cumulative variance identifies the optimal number of factors for extraction. According to Zanella *et al.* (2007) [45], the minimal number of factors to be extracted should correspond to a minimum explanation of 60% of the total cumulative variance.

The FA defines the structure of the variables' correlations, calculates the correlation matrix between variables, extracts initial factors, and rotates the matrix (Hair *et al.*, 2005) [46]. The FA extracts the factors and calculates the factorial loads to explain the similarities and dissimilarities of the matrix variables. The highest factorial load indicates greater importance and greater influence on the factor label, which has practical significance if the values are above (±0.50) (Hair *et al.*, 2005) [46].

4. Results and Discussion

4.1. Groundwater Quality

Table 7 presents the chemical analysis averaged values for 940 wells and seventeen physicochemical parameters. The analysis was performed in triplicates. The groundwater quality is discussed comparatively with water quality standards for human consumption, Health Minister Resolution 888/2021 (Brazil, 2021) [28], and comparatively between the domains.

Table 7. Averaged values for seventeen physicochemical parameters of 940 wells, iron content > 0.3 $mg \cdot L^{-1}$, drilling period 2003-2013.

Chemical Parameters	MPV-HM 888/2021	Karstic	Sedimentary	Metasedimentary	Crystalline
		mg/L			
Total Residue, mg/L	1000.0	743.20	553.49	304.73	4499.24
Chloride, mg/L Cl	250.0	143.59	232.10	55.98	1776.31
Total Hardness, mg/L $CaCO_3$	300.0	424.11	118.18	118.01	1681.70
Nitrate, mg/L $N-NO_3$	10.0	5.20	0.60	0.63	1.70
Alkali-HCO, mg/L $CaCO_3$	NA	247.52	53.53	86.44	217.65
Alkali-CO, mg/L $CaCO_3$	NA	4.46	1.16	1.25	4.43
Alkali-OH, mg/L CaCO3	NA	0.47	0.40	0.42	0.40
Calcium, mg/L $CaCO_3$	NA	286.11	45.18	67.39	569.61
Electrical Conductivity, µS/cm	NA	1028.40	795.50	382.36	5335.02
Color, mg/L Pt-Co	15.0	86.20	95.33	397.41	99.85
Iron, mg/L Fe	0.3	1.36	1.52	2.25	2.45
Fluoride, mg/L F	1.5	0.45	0.15	0.12	0.61
Magnesium, mg/L Mg	NA	32.44	18.08	12.32	271.13
Nitrite, mg/L $N-NO_2$	1.0	0.08	0.06	0.02	0.02
pH	6.0 - 9.0	7.53	6.72	7.20	7.37
Silica, mg/L SiO_2		23.48	36.94	22.44	59.42
Sulphate, mg/L SO_4	250.0	60.59	29.09	18.43	207.23
Turbidity, NTU	5.0	27.51	19.57	56.00	33.13

Note: NA = Not Available; MPV-HM = Most Probable Value from Health Minister.

The data in Table 7 show that the crystalline and metasedimentary domains present the largest iron averaged values, respectively: 2.45 and 2.25 $mg \cdot L^{-1}$, while the sedimentary and karstic domains have the smallest iron averaged values, respectively: 1.52 and 1.36 $mg \cdot L^{-1}$.

The crystalline domain presents for total residue, chloride, and total hardness the averaged values, 4499.24, 1776.31, and 1681.7 $mg \cdot L^{-1}$, far above the respective values of the quality standards: 1000, 250, and 300 $mg \cdot L^{-1}$. In addition, the values for calcium, magnesium, and sulfate are larger than in other domains. The crystalline domain presents high ionic concentration in groundwater, as indicated by the largest value of electrical conductivity among the domains, 5335.02 $mg \cdot L^{-1}$.

The karstic domain presents the averaged values for total hardness and total residue, 424.1 and 743.2 mg·L^{-1}, above the respective values for the sedimentary, 118.18 and 553.49 mg·L^{-1}; also for the metasedimentary domain, 118.01 and 304.73 mg·L^{-1}.

The concentration value for nitrate, 5.2 mg·L^{-1}, is below the limit, 10.0 mg·L^{-1}, but far above the averaged values for the other domain: 0.6, 0.63, and 1.7 mg·L^{-1}. Nitrate is not a regionalized variable for the groundwater of the State of Bahia (Negrão, 2008) [47], because it was found, based on the semivariogram, that a = 4.95 Km is a small distance, after which the nitrate values no longer correlate, characterizing a local variable. This author verified that large nitrate concentrations were associated with vectors of pollution (irrigated agriculture and domestic wastewater effluents) in the most vulnerable areas of the karstic and crystalline aquifers: shallow aquifers, karstic, and fractured structures.

4.2. Statistical Analysis of Physicochemical Parameters

The linear correlation between iron and the other sixteen physicochemical parameters, using Pearson and Spearman matrices and SPSS V.12 (SPSS 2003) [42] in the sample of 940 wells, indicated that for n = 940, and p-Value 0.05, only correlation > 0.20 is significant. **Table 8** and **Table 9** present the results.

Table 8. Pearson matrix of linear correlation for 940 samples and seventeen physicochemical parameters.

n = 17	TDS	Chloride	Hardness	Nitrate	Alkali-HCO	Alkali-CO	Alkali-OH	Calcium	EC	Color	Iron	Fluoride	Magnesium	Nitrite	pH	Sulphate	Turbidity
TDS	1.00																
Chloride	0.98	1.00															
Hardness	0.96	0.95	1.00														
Nitrate	0.10	0.11	0.13	1.00													
Alkali-HCO	0.41	0.37	0.41	0.26	1.00												
Alkali-CO	0.09	0.11	0.07	−0.01	0.10	1.00											
Alkali-OH	0.06	0.09	0.07	0.03	0.04	0.28	1.00										
Calcium	0.88	0.87	0.93	0.16	0.42	0.05	0.09	1.00									
EC	0.97	0.99	0.93	0.13	0.42	0.12	0.10	0.86	1.00								
Color	−0.01	−0.02	−0.02	−0.05	−0.07	−0.06	−0.15	−0.02	−0.03	1.00							
Iron	0.16	0.15	0.16	−0.11	−0.05	−0.06	−0.05	0.17	0.13	0.32	1.00						
Fluoride	0.34	0.38	0.36	0.19	0.40	−0.06	−0.03	0.40	0.40	−0.04	−0.03	1.00					
Magnesium	0.94	0.93	0.97	0.10	0.38	0.08	0.06	0.84	0.91	−0.02	0.15	0.31	1.00				
Nitrite	0.09	0.14	0.12	0.29	0.10	0.02	0.11	0.11	0.15	−0.02	−0.02	0.24	0.12	1.00			
pH	0.00	−0.01	0.01	0.02	0.33	0.21	−0.08	0.02	0.02	−0.01	−0.17	0.19	0.00	0.01	1.00		
Sulphate	0.72	0.68	0.71	0.08	0.36	0.06	0.10	0.72	0.71	−0.04	0.11	0.35	0.66	0.07	0.03	1.00	
Turbidity	0.05	0.03	0.03	−0.04	−0.07	−0.05	−0.03	0.04	0.02	0.75	0.49	−0.05	0.02	−0.03	−0.06	0.01	1.00

Table 9. Spearman matrix of linear correlation for 940 samples and seventeen physicochemical parameters.

n = 17	TDS	Chloride	Hardness	Nitrate	Alkali-HCO	Alkali-CO	Alkali-OH	Calcium	EC	Color	Iron	Fluoride	Magnesium	Nitrite	pH	Sulphate	Turbidity
TDS	1.00																
Chloride	0.91	1.00															
Hardness	0.93	0.85	1.00														
Nitrate	0.25	0.27	0.23	1.00													
Alkali-HCO	0.75	0.60	0.81	0.23	1.00												
Alkali-CO	0.03	−0.04	0.04	0.00	0.08	1.00											
Alkali-OH	0.01	−0.02	0.03	0.01	0.03	0.84	1.00										
Calcium	0.90	0.79	0.98	0.23	0.81	0.06	0.05	1.00									
EC	0.97	0.93	0.96	0.24	0.78	0.04	0.01	0.92	1.00								
Color	−0.01	0.02	−0.03	−0.08	−0.09	−0.55	−0.57	−0.03	−0.03	1.00							
Iron	0.08	0.09	0.05	−0.19	−0.06	−0.10	−0.05	0.04	0.04	0.42	1.00						
Fluoride	0.59	0.55	0.58	0.17	0.56	−0.02	−0.07	0.57	0.61	0.04	−0.02	1.00					
Magnesium	0.92	0.88	0.96	0.22	0.75	0.02	0.02	0.90	0.94	−0.03	0.05	0.56	1.00				
Nitrite	0.21	0.19	0.23	0.43	0.23	0.11	0.13	0.24	0.21	−0.10	−0.08	0.17	0.19	1.00			
pH	0.14	0.02	0.20	−0.04	0.36	0.12	−0.10	0.22	0.17	−0.05	−0.19	0.24	0.15	0.06	1.00		
Sulphate	0.81	0.80	0.78	0.27	0.66	−0.01	−0.02	0.76	0.82	0.01	0.02	0.59	0.79	0.18	0.11	1.00	
Turbidity	0.07	0.05	0.00	−0.04	−0.10	−0.18	−0.15	0.00	−0.01	0.54	0.63	−0.08	0.00	−0.06	−0.20	−0.01	1.00

From Table 8 and Table 9, only two parameters, color and turbidity, with correlation values > 0.20, have significant correlation with iron. The correlation with color is associated with the degree of light intensity reduction when light passes through the water sample, due to the presence of dissolved solids, mainly organic and inorganic colloidal material (iron and manganese oxides), abundant in different types of soil (CETESB, 2015) [48]. The correlation with turbidity is also associated with the degree of light intensity reduction, due to the presence of suspended solids in the water (silt, clay, colloids, organic matter, etc.). Groundwater rich in iron may present high turbidity when coming in contact with atmospheric oxygen. The oxidation of the soluble iron (II) to the insoluble species iron (III) produces precipitates, brown coloring the water. Color and turbidity are highly correlated to each other, with values of 0.75 and 0.54 from Table 9 and Table 10, respectively. Thus, color and turbidity are consequences and not a cause of iron concentration. On the other hand, the absence of a linear correlation between iron and the other fourteen chemical parameters indicates that iron has its specific chemical origin.

4.3. Iron Distribution in Groundwater of the State of Bahia

The distribution of iron content in groundwater in the state of Bahia is presented in three maps below. Figure 6 presents the spatialization of 3539 wells over the map of the State of Bahia. **Figure 7** presents the distribution of 546 wells on top of the Karstic domain. **Figure 8** presents the distribution of 486 wells on top of the metasedimentary domain. The data are from (CERB, 2014) [30], drilling period 2003-2013, iron content, either below or above 0.3 mg/L, and flow test ≥ 1.0 $m^3 \cdot h^{-1}$, on top of five geological charts published by CPRM (millionth scale): SC24 Aracaju, SD24 Salvador, SE24 Rio Doce, SC23 São Francisco, and SD23 Brasilia. The software ArcGIS 9.3 (ESRI, 2004) [41] was used to edit shape files, to do data manipulation, data interpretation, analysis, and presentation.

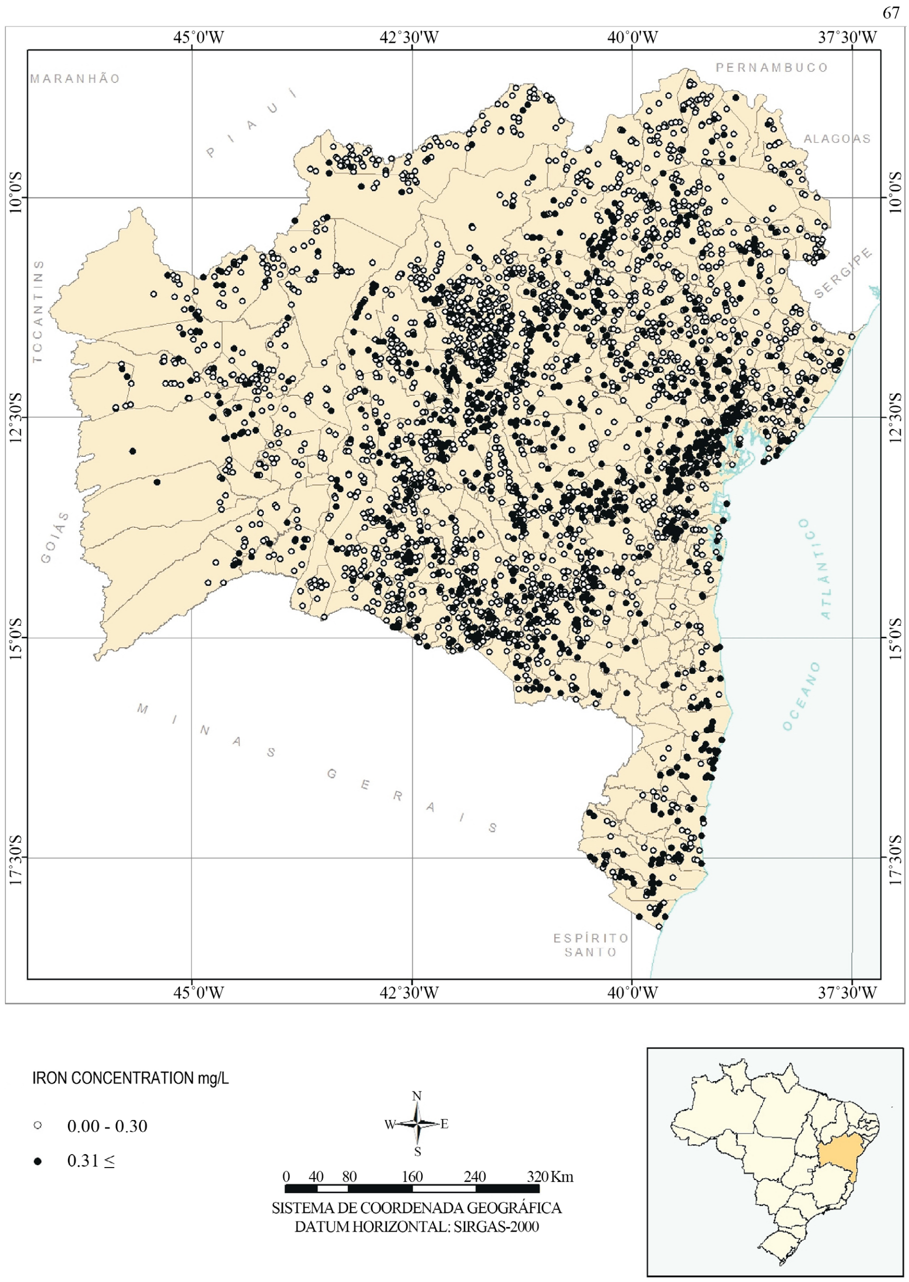

Figure 6. Spatialization of the 3539 wells, drilling period 2003-2013, flow test ≥ 1.0 $m^3 \cdot h^{-1}$ and iron content below or above 0.3 $mg \cdot L^{-1}$ in the map of the State of Bahia.

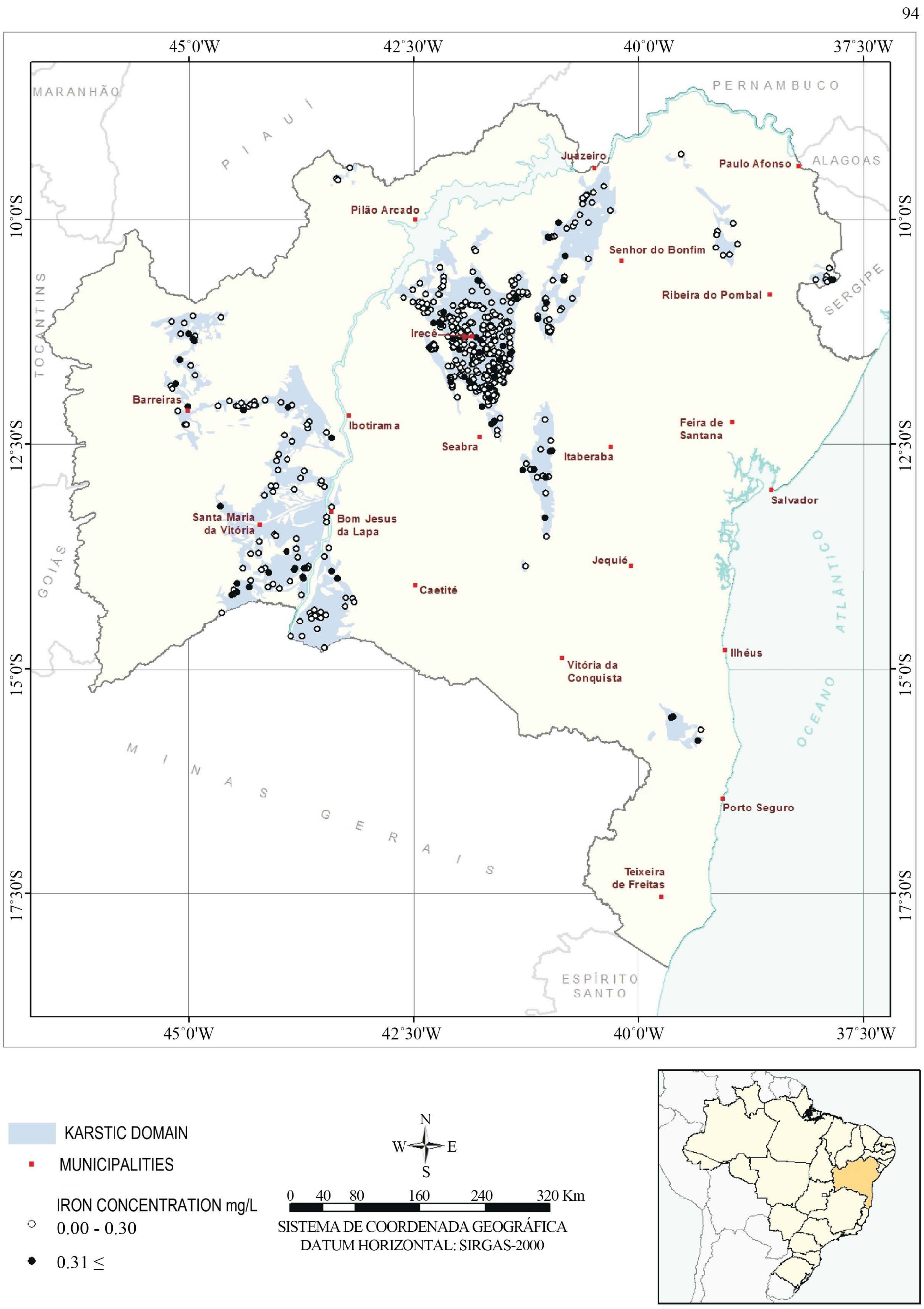

Figure 7. Spatialization of 546 wells, drilling period 2003-2013, flow test ≥ 1.0 $m^3 \cdot h^{-1}$ and iron content below or above 0.3 $mg \cdot L^{-1}$ on top of the Karstic Domain.

Figure 6 shows that high iron content (black dots) is distributed all over the map of the State of Bahia. However, it displays different percentages in different regions, due to local environmental conditions, which is the subject of the present study. To exemplify these findings, **Figure 7** shows that for the karstic domain, a percentage of 11.72% of wells (64/546) had high iron content. However, locally, for different karstic areas, the percentages for wells with high iron content are 10.1, 11.1, and 35.5%, which reflects local environmental conditions. In addition, **Figure 8** shows a percentage of 38.1% wells (185/486) with high iron content for the metasedimentary domain. Again, for different metasedimentary areas, the percentages are 29.9%, 30.0%, 38.1%, 46.5%, and 57.1%, which reflects local environmental conditions. It can be seen that larger percentages of high iron content occur for the metasedimentary domain than for the karstic domain.

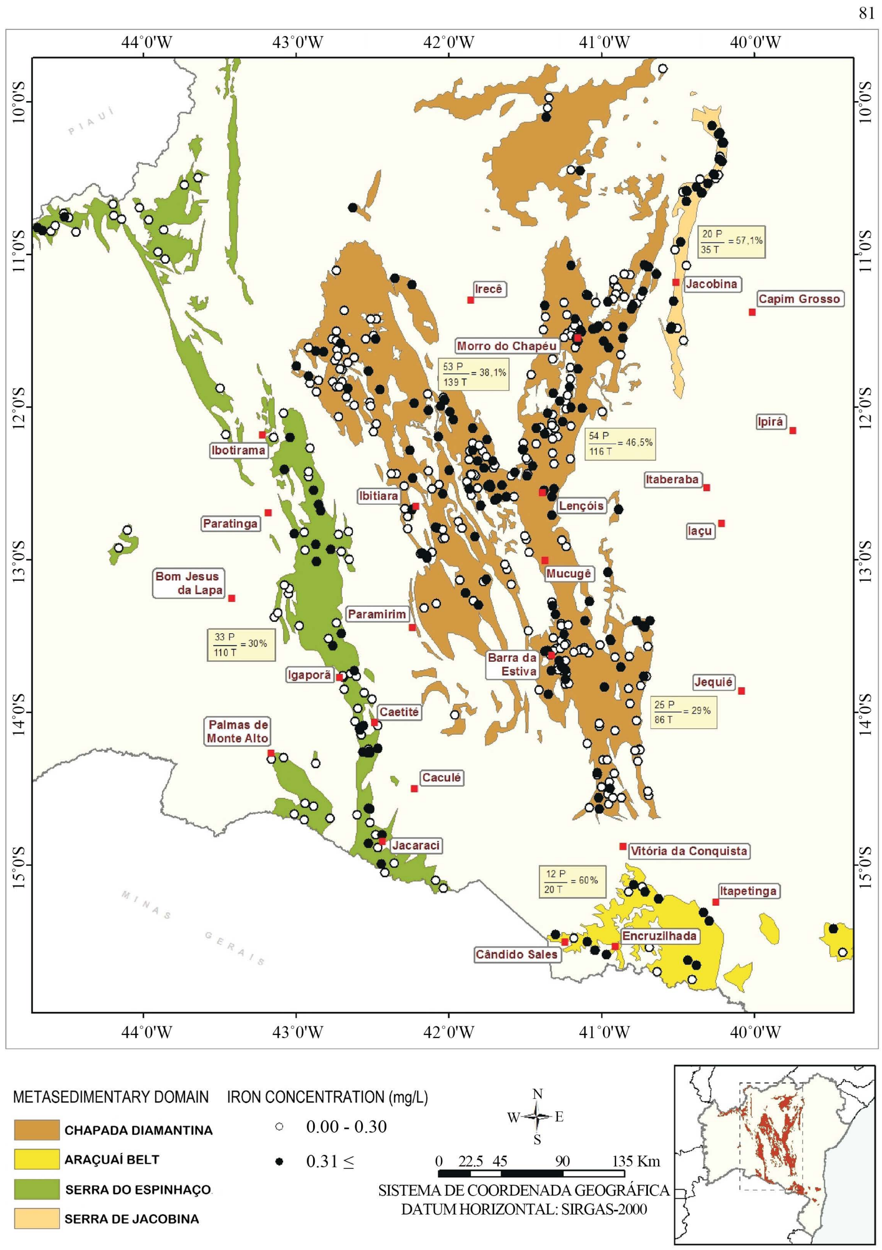

Figure 8. Spatialization of 486 wells, drilling period 2003-2013, flow test ≥ 1.0 $m^3 \cdot h^{-1}$ and iron content below or above 0.3 $mg \cdot L^{-1}$ on top of the Metasedimentary Domain.

Table 10 presents the statistics for the distribution of wells in the hydrogeologic domains of the State of Bahia and the respective high iron content percentages for the two sets of wells: 3539 and 940.

Table 10. Number of drilled wells, 2003-2013, by hydrogeological domains: 3539 with flow test ≥ 1 $m^3 \cdot h^{-1}$; 940 with iron > 0.3 $mg \cdot L^{-1}$ and respective percentages.

Hydrogeologic Domains	N˚ of Wells/Percentage with Iron > 0.3 $mg \cdot L^{-1}$	N˚ of Wells with Iron > 0.3 $mg \cdot L^{-1}$)/Percentage
Sedimentary	494/(30.8)	152 (16.17)
Crystalline	1688/(27.7)	468 (49.79)
Karstic	660/(14.1)	93 (9.89)
Metasedimentary	697/(32.6)	227 (24.15)
Totals	**3539/(26.56%)**	**940 (100%)**

Table 10 presents the four hydrogeological domains with 26.56% (940/3539) wells with iron > 0.3 $mg \cdot L^{-1}$. In the sample of 940 wells, the percentages of wells with iron > 0.3 $mg \cdot L^{-1}$ for the crystalline and metasedimentary domains are, respectively, 49.79 and 24.15%, while the percentages of wells for the sedimentary and karstic domains are, respectively, 16.17 and 9.89%. Consequently, in domains with mafic-ultramafic rocks or iron-forming rocks, the percentage of wells with high iron content is larger, with smaller percentages for domains with sandstone or carbonate rocks.

4.4. Grades for the Environmental Attributes

Table 11 presents the summary of the variables of the four environmental attributes, the corresponding grades (1 - 5), and sample percentages based on 940 wells.

Table 11. Summary of the grades for the environmental attributes and sample percentages based on 940 wells.

Soils & Grades	5 (Latosols; Argisols)	4 (Planosols)	3 (Neosols)	2 (Cambisols)	1 (Luvissols; Chernosols)
Number of Samples/Percentages	554/58.9.%	90/9.6%	169/18.0%	93/9.9%	34/3.6%
Maximum/Average Iron (mg/L)	33.3/2.49	29.9/1.67	26.5/2.43	8.32/1.56	4.85/1.18
Vegetation & Grades	5 (Northeast Dry Forest)	4 (Coastal Forest; Inland Forest)	3 (Chapada Diamantina)	2 (Caatinga)	1 (Savanah)
Number of Samples/Percentages	386/41.1%	215/22.9%	161/17.1%	163/17.3%	15/16.0%
Maximum/Average Iron (mg/L)	33.3/2.23	27.0/2.52	26.5/2.39	16.8/1.68	8.32/1.14
Lithology & Grades	4 (Crystalline)	3 (Meta Sedimentary)	2 (Sedimentary)	1 (Karstic)	
Number of Samples/Percentages	468/49.8%	227/24.1%	152/16.2%	93/9.9%	
Maximum/Average Iron (mg/L)	33.3/2.42	26.5/2.15	9.69/0.90	8.32/1.36	
Climate & Grades	4 (Sub-Humid to Dray)	3 (Humid; Humid to Sub-Humid)	2 (Semiarid)	1 (Arid)	
Number of Samples/Percentages	228/24.3%	504/53.6%	185/19.7%	23/24.5%	
Maximum/Average Iron (mg/L)	33.3/2.5	27.0/2.44	10.6/1.73	6.0/1.66	

Based on data from **Table 11**, for 940 wells, the environmental attributes with the most significant impact on high iron content are: Latosols/Argisols, which occupies a large percentage of the State of Bahia territory, 57.57% (**Table 5**), has grade 5, and present a significant percentage of wells, 58.9%. The Northeast dry forest, located at the border of the state's central region and left of the São Francisco river, has grade 5, also a significant percentage of wells, 41.1%. The Crystalline domain occupying the State of Bahia territory by 34.3% (**Table 4**) has grade 4, and the percentage of wells is 49.8%. The climate Sub-Humid to Dry presents a low percentage of 24.3% wells, but grade 5 indicates the largest values of iron content. The reason is the combination of sub-humid conditions and higher temperatures accelerating the chemical reactions and iron liberation. On the other hand, the climate Humid and Humid to Sub-Humid presents the larger percentage of wells, 53.6%, but grade 4 indicates smaller values of iron content. This climate is associated with dense forests with abundant aerial organic matter and less organic matter in the soil, consequently less decomposition and leaching of iron from the debris to groundwater.

The grades 1 - 5 for the variables of each attribute allowed the application of multivariable methods PCA and FA (SPSS V.12).

Multivariable statistical analysis The method of principal component analysis (PCA) calculated the variance for the first component (PC1 = 34.36%) and the second component (PC2 = 25.16%), with a total variance of 59.52%. **Figure 9** presents the graph of PC1 x PC2.

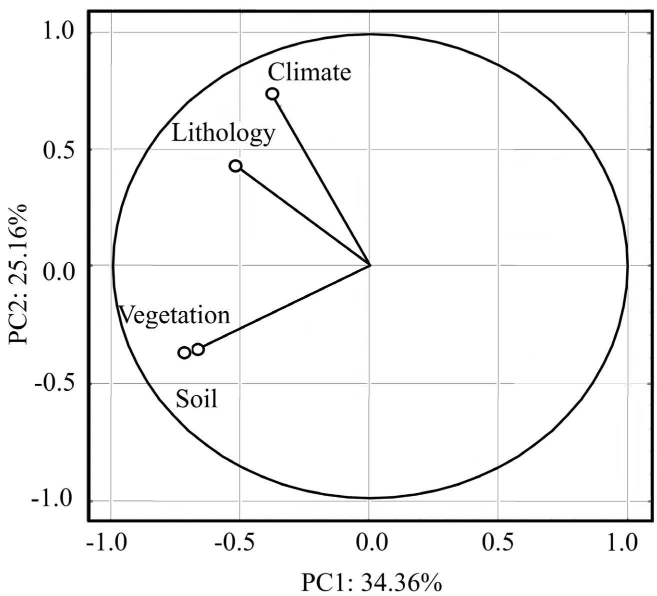

Figure 9. Principal component analysis (PC1 × PC2).

The graph (PC1 × PC2) in **Figure 9** shows a clear separation between the lines for climate and lithology, while the lines for vegetation and soil have a strong approximation, revealing their coupled influence upon high iron content in the groundwater. The components PC1 and PC2 are sufficient to define the correlation and the degree of significance of each environmental attribute on the dissolved high iron content in groundwater of the State of Bahia, as the value PC1 + PC2 = 59.52% closely attended the minimum explanation of 60% of the total accumulated variance (Zanella *et al*., 2007) [45].

The application of factorial analysis (FA) presented the following environmental attributes and factorial loads. In Factor 1, the soil is −0.71, vegetation is −0.68, and lithology is −0.52, while for Factor 2, the climate is +0.74.

Soil and vegetation appear with the largest factorial loads in Factor 1, respectively: −0.71 and −0.68. Both attributes are able to liberate iron in the groundwater from dissolution and biological processes. Latosols/Argisols appear to be the most important soil variable for iron content in groundwater and occupy 57.57% of the state territory (**Table 4**). For vegetation, the most important environmental variables are Northeast Dry Forests, Coastal Forests, and Inland Forests, which occupy less than 50% of the state territory (**Table 5**).

Lithology has a smaller factorial load in Factor 1, −0.52. This smallest factorial load can be justified by the following: in **Table 4**, it is observed that the crystalline and metasedimentary domains, rich in mafic-ultramafic rocks and iron forming bands, occupy 48.6% of the state territory, and the sedimentary and karstic lithology, composed by non-iron forming rocks, occupy 51.4%, a quite similar percentage. In addition, in **Table 1**, the crystalline and metasedimentary domains present 17.2% of wells with high iron content, while the sedimentary plus karstic domains present 16.8%, again, quite similar percentages. As the percentages are quite similar, lithology is not the most important parameter, and alone, it cannot explain the groundwater's high iron content for the entire State of Bahia.

Climate has a significant factor load in Factor 2, +0.74. Climate, with emphasis in rainfall and temperature, does not produce iron. However, iron dissolution and transportation to the aquifer are very important.

Table 12 presents, for each hydrogeological domain and the variables of each environmental attribute: climate, lithology, soil, and vegetation, the percentages of wells drilled, organized from the smaller to the larger grades.

Table 12. For 940 wells, the percentages of wells for each hydrogeological domain and each variable of the environmental attributes.

Statistics		Percentages			
		1 (Karstic Domain)	2 (Sedimentary Domain)	3 (Metasedimentary Domain)	4 (Crystalline Domain)
Number of Wells		93	152	227	468
Climate					
1	Arid	6.5	0.7	0.9	3.0
2	Semiarid	62.4	27.6	8.8	13.9
3	Humid; Humid to Sub-Humid	30.1	13.2	86.8	55.3
4	Sub-Humid to Dry	1.1	58.6	3.5	27.8

Continued

Vegetation					
1	Savanah	0.0	9.9	0.0	0.0
2	Caatinga	2.2	26.3	11.9	20.1
3	Chapada Diamantina	37.6	0.7	34.4	10.0
4	Coastal Forest; Inland Forest	57.0	1.3	49.8	10.0
5	Northeast Dry Forest	3.2	61.8	4.0	59.8
Soil					
1	Luvissols; Chernosols	0.0	0.7	0.0	6.6
2	Cambisols	100.0	0.0	0.0	0.0
3	Neosols	0.0	14.5	46.7	8.8
4	Planosols	0.0	3.9	0.0	17.9
5	Latosols; Argisols	0.0	80.9	53.3	66.7

The discussion of data from Table 12 systematically references the environmental variables with higher grades: climate (grade 4, sub-humid to dry), vegetation (grade 5, northeast dry forest), and soil (grade 5, Latosols and Argisols).

The Karstic domain (grade 1), with 93 wells and unfavorable lithology for dissolved iron in groundwater, presents low percentages of wells drilled in environmental variables with higher grades, respectively: 1.1%, 3.2%, and 0.0%. The coastal forest and inland forest (grade 4), with a percentage of 57% wells, preponderantly answer for the groundwater high iron content in the Karstic domain. On the other hand, the semiarid climate, with a percentage of 62.4%, and the cambisols, with 100% of the wells, help to explain the low interval for the averaged values of high iron content, 1.36 - 8.32 mg·L^{-1} (Table 12).

The Sedimentary domain (grade 2), with 152 wells and unfavorable lithology to dissolved iron in groundwater, presents significant percentages of wells drilled on environmental variables with higher grades, 58.6%, 61.8%, and 80.9%, respectively. On the other hand, the semiarid climate, with a percentage of 27.6%, and the Caatinga, with 26.3% of the wells, help to explain the low interval for the averaged values of high iron content, 0.90 - 9.69 mg·L^{-1} (Table 12).

The Metasedimentary domain (grade 3), with 227 wells and favorable lithology to dissolved iron in groundwater, presents the following percentages of wells drilled on environmental variables with higher grades: 3.5%, 4.0%, and 53.3%, respectively. On the other hand, the humid to sub-humid climate (grade 3), with a percentage of 86.8%, and the coastal forest and inland forest (Grade 4), with 49.8% of the wells, help to explain the large interval for the averaged values of high iron content, 2.15 - 26.5 mg·L^{-1} (Table 12).

The Crystalline domain (grade 4), with 468 wells and favorable lithology to dissolved iron in groundwater, presents the following percentages of wells drilled on environmental variables with higher grades: 27.8%, 59.8%, and 66.7%, respectively. On the other hand, the humid to sub-humid climate (grade 3), with a percentage of 55.3%, helps to explain the larger interval for the averaged values of high iron content, 2.42 - 33.3 mg·L^{-1} (**Table 12**).

5. Conclusions

The method PCA of the multivariable statistical analysis calculated the total accumulated variance of 59.52% for the environmental attributes, climate, lithology, soil, and vegetation to explain the high iron content in the state of Bahia groundwater. This result closely attended the recommendation of 60% for a minimum explanation of the total data variance.

The method FA of the multivariable statistical analysis defined the following factorial loads: in Factor 1: soil, -0.71; vegetation, −0.68; and lithology, −0.52; and in Factor 2: climate, +0.74. Based on the factorial loads, soil and vegetation are the attributes that are more significant in explaining the groundwater's high iron content for the entire State of Bahia, and lithology, alone, cannot explain such behavior, as its factorial load is the smallest.

In agreement with the values of the factorial loads, it was observed for the sample of 940 wells in the four hydrogeological domains of the state of Bahia (crystalline, karstic, metasedimentary, and sedimentary) that 59% of the wells were drilled on top of soils with the highest grade or largest value for iron content (latosols; argisols); 41% on top of vegetation with the highest grade (northeast dry forest); and 24% in climate regions with the highest grade (sub-humid to dry). These findings confirm that climate is very important but has a secondary influence, as rainfall and temperature are not iron producers but are important for iron dissolution and transportation to the aquifer.

It was also found that for the sedimentary domain (primarily sandstones) and karstic domain (carbonate rocks), poor in iron content: percentages of wells, 80.9% - 100% occurred in iron-rich soils; of 57.0% - 61.8% in humid to the dry forest; and of 58.6% - 62.4% in sub-humid to dry and semiarid climate. These results indicated that, although lithology is a determinant for high dissolved iron content in the state of Bahia groundwater, this attribute alone (factor weight −0.52) cannot explain the phenomenon in the entire state of Bahia.

In conclusion, the present work uses multivariable analysis with geospatial mapping of high iron content on top of maps of the environmental attributes, which reveals the role of each attribute variable on the groundwater high iron content of a specific location.

This is useful information for the state well drilling company (CERB) and its groundwater managers to avoid inadequate well locations, as they had reported that many wells perforated by CERB were never completed and installed for public supply, due to high iron content, with significant financial loss due to well abandonment, since the long term maintenance cost and treatment process is unfeasible for small communities or municipalities.

Acknowledgements

The authors thank the governmental well drilling company from the State of Bahia, Brazil, Bahia Water and Sanitation Engineering Company (CERB), for making available the hydrogeological database for this research.

Availability of Data and Materials

It was submitted to the journal two spreadsheets as Supplementary Material (CERB_3539Wells_Iron_Data and CERB_940Wells_Data&FactorsWeight). The 940 wells were submitted to multivariable analyses to define the importance of each environmental attribute upon the high iron content in the groundwater of the state of Bahia.

Authors' Contributions

The first author, José Carlos Cruz do Carmo, a Master's degree student at the Polytechnic School, Federal University of Bahia, did the research work (data search and selection, statistical studies, maps development, and report writing). He also wrote the first version of this paper. The second author, Deize Elle Ribeiro Moitinho, a hydrogeologist for the Bahia Water and Sanitation Engineering Company (CERB), reviewed the paper and made corrections; The third author, Prof. Iara Brandão de Oliveira, was the Master's degree advisor and the paper reviser for submission.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Thomas, E.O. (2023) Evaluation of Groundwater Quality Using Multivariate, Parametric and Non-Parametric Statistics, and GWQI in Ibadan, Nigeria. *Water Science*, **37**, 117-130. https://doi.org/10.1080/23570008.2023.2221493

[2] Ismail, E., Snousy, M.G., Alexakis, D.E., Gamvroula, D.E., Howard, G., El Sayed, E., *et al*. (2023) Multivariate Statistical Analysis and Geospatial Mapping for Assessing Groundwater Quality in West El Minia District, Egypt. *Water*, **15**, Article 2909. https://doi.org/10.3390/w15162909

[3] Mehdi, M., Hakim, D., Amina, A. and Nourredine, G. (2023) Multivariate Statistical Analysis of Groundwater Quality in the Region of Hassi R'mel, Algeria. *Journal of Ecological Engineering*, **24**, 22-31. https://doi.org/10.12911/22998993/161140

[4] Patel, P.S., Pandya, D.M. and Shah, M. (2023) A Holistic Review on the Assessment of Groundwater Quality Using Multivariate Statistical Techniques. *Environmental Science and Pollution Research*, **30**, 85046-85070. https://doi.org/10.1007/s11356-023-27605-x

[5] Barkat, A., Bouaicha, F., Bouteraa, O., Mester, T., Ata, B., Balla, D., *et al*. (2021) Assessment of Complex Terminal Groundwater Aquifer for Different Use of Oued Souf Valley (Algeria) Using Multivariate Statistical Methods, Geostatistical Modeling, and Water Quality Index. *Water*, **13**, Article 1609. https://doi.org/10.3390/w13111609

[6] Gaikwad, S., Gaikwad, S., Meshram, D., Wagh, V., Kandekar, A. and Kadam, A. (2019) Geochemical Mobility of Ions in Groundwater from the Tropical Western Coast of Maharashtra, India: Implication to Groundwater Quality. *Environment, Development and Sustainability*, **22**, 2591-2624. https://doi.org/10.1007/s10668-019-00312-9

[7] Machiwal, D., Jha, M.K., Singh, V.P. and Mohan, C. (2018) Assessment and Mapping of Groundwater Vulnerability to Pollution: Current Status and Challenges. *Earth-Science Reviews*, **185**, 901-927. https://doi.org/10.1016/j.earscirev.2018.08.009

[8] Gaikwad, S.K., Kadam, A.K., Ramgir, R.R., Kashikar, A.S., Wagh, V.M., Kandekar, A.M., *et al*. (2020) Assessment of the Groundwater Geochemistry from a Part of West Coast of India Using Statistical Methods and Water Quality Index. *HydroResearch*, **3**, 48-60. https://doi.org/10.1016/j.hydres.2020.04.001

[9] Wu, J., Li, P., Wang, D., Ren, X. and Wei, M. (2019) Statistical and Multivariate Statistical Techniques to Trace the Sources and Affecting Factors of Groundwater Pollution in a Rapidly Growing City on the Chinese Loess Plateau. *Human and Ecological Risk Assessment*: *An International Journal*, **26**, 1603-1621. https://doi.org/10.1080/10807039.2019.1594156

[10] Li, P., Tian, R. and Liu, R. (2018) Solute Geochemistry and Multivariate Analysis of Water Quality in the Guohua Phosphorite Mine, Guizhou Province, China. *Exposure and Health*, **11**, 81-94. https://doi.org/10.1007/s12403-018-0277-y

[11] Cloutier, V., Lefebvre, R., Therrien, R. and Savard, M.M. (2008) Multivariate Statistical Analysis of Geochemical Data as Indicative of the Hydrogeochemical Evolution of Groundwater in a Sedimentary Rock Aquifer System. *Journal of Hydrology*, **353**, 294-313. https://doi.org/10.1016/j.jhydrol.2008.02.015

[12] Barbosa Filho, J. and de Oliveira, I.B. (2021) Development of a Groundwater Quality Index: GWQI, for the Aquifers of the State of Bahia, Brazil Using Multivariable Analyses. *Scientific Reports*, **11**, Article No. 16520. https://doi.org/10.1038/s41598-021-95912-9

[13] Gomes, M.D.C.R., Anjos, J.Â.S.A.D. and Daltro, R.R. (2020) Multivariate Statistical Analysis Applied to the Evaluation of Groundwater Quality in the Central-Southern Portion of the State of Bahia-Brazil. *Ambiente e Agua—An Interdisciplinary Journal of Applied Science*, **15**, e2408. https://doi.org/10.4136/ambi-agua.2408

[14] Gomes, M.d.C.R. and Cavalcante, I.N. (2017) Aplicação da análise estatística multivariada no estudo da qualidade da água subterrânea. *Águas Subterrâneas*, **31**, 134-149. https://doi.org/10.14295/ras.v31i1.28617

[15] Freeze, A.R. and Cherry, J.A. (1979) Groundwater. Prentice Hall Inc, 604. http://www.sciepub.com

[16] Oliveira, D.A., Schmidt, G. and Macedo de Freitas, D. (2004) Avaliação do teor de ferro nas águas subterrâneas de alguns poços tubulares no plano diretor de Palmas. *Anais do XXXI Congreso Interamericano de Ingeniería Sanitaria y Ambiental*, AIDIS, San Juan, 20-22 Novembro 2004, 22-27. https://aidisnet.org

[17] Nascimento, S.A.M. (2008) Diagnóstico hidrogeológico, hidroquímico e da qualidade da água do aquífero freático do alto cristalino de Salvador-Bahia. Ph.D. Thesis, Instituto de Geociências da Universidade Federal da Bahia. https://repositorio.ufba.br

[18] Carmo, J.C.C. (2016) Ferro na água subterrânea do Estado da Bahia: Uma avaliação preliminar dos fatores influentes na concentração. Master's Thesis, Universidade Federal da Bahia. https://repositorio.ufba.br

[19] Siqueira, L.F.S., Rojas, M.O.A.I. and Neto, J.D.J.G.D.C. (2011) Determinação espectrométrica de ferro (II) pelo sistema fe (II)/ kscn em água do mar da praia do calhau, são luis, maranhão (Brasil). *Acta Tecnológica*, **6**, 27-34. https://doi.org/10.35818/acta.v6i1.40

[20] Ityel, D. (2011) Ground Water: Dealing with Iron Contamination. Water & Wastewater. https://www.filtsep.com

[21] Zimbres, E. (2002) Química das Águas. Universidade Estadual do Rio de Janeiro/RJ.

https://www.meioambiente.pro.br/agua/quimica.htm

[22] Szikszay, M. (2012) Geoquímica das águas. Boletim IG/USP, 167. https://www.revistas.usp.br

[23] Schoeller, N. (1962) Les eaux souterraines. Mason, 642.

[24] Carroll, D. (1962) Rainwater as a Chemical Agent of Geologic Processes: A Review. U.S. Geological Survey Water Supply Paper, 1535-G, 1-18.

[25] Matthess, G. (1982) The Properties of Groundwater. John Wiley & Sons, 406.

[26] Perel'man, A.L. (1963) The Aquatic Transport and Biologic Absorption of Elements in the Soil. In: Schoeller, M., Ed., *Recherches sur l aquisition de la composition chimique des eau souterraines*, 179-187.

[27] Davis, J.B. (1967) Petroleum Microbiology. Elsevier, 604.

[28] BRASIL (2021) Portaria. No. 888, de 4 de maio de 2021. Altera o Anexo XX da Portaria de Consolidação GM/MS nº 5, de 28 de setembro de 2017, para dispor sobre os procedimentos de controle e de vigilância da qualidade da água para consumo humano e seu padrão de potabilidade. https://www.gov.br

[29] Bahia (2003) Plano Estadual de Recursos Hídricos. Secretaria de Infraestrutura, Superintendência de Recursos Hídricos. Consortium MAGNA-BRL. http://www.inema.ba.gov.br

[30] CERB (2014) Dados hidrogeológicos de poços. Salvador: Companhia de Engenharia Ambiental e Recursos Hídricos da Bahia; Divisão de Cadastro Hidrogeológico. Relatório de Gestão 2007/2014. http://www.cerb.ba.gov.br

[31] Brasil (2011) Portaria nº 2.914, de 12 de dezembro de 2011. Dispõe sobre os procedimentos de controle e de vigilância da qualidade da água para consumo humano e seu padrão de potabilidade. https://www.gov.br

[32] Lima, E.A., Beraldo, V.J., Guilera, S.C., Brandão, L.C.R. and Costa, E.A. (2009) Mapa de variação da qualidade das águas subterrâneas da região do nordeste do Brasil para uso na irrigação. *Revista Águas Subterrâneas*, 1-13. https://aguassubterraneas.abas.org

[33] Guerra, A.M. and Negrão, F.I. (1996) Domínios hidrogeológicos do estado da Bahia. Anais do IX Congresso Brasileiro de Águas Subterrâneas, Salvador: ABAS. https://aguassubterraneas.abas.org

[34] Barbosa, J.S.F. and Dominguez, J.M.L. (1996) Texto explicativo do Mapa Geológico da Bahia, ao milionésimo. SICM/SGM, 400. http://www.ppegeo.igc.usp.br

[35] Moruzzi, R.B. and Reali, M.A.P. (2012) Oxidação e remoção de ferro e manganês em águas para fins de abastecimento público ou industrial-uma abordagem geral. *Revista de Engenharia e Tecnologia*, **4**, 29-43. https://repositorio.unesp.br

[36] Lee, J.D. (1999) Química Inorgânica Não Tão Concisa. Edgard Blucher, 527. https://www.blucher.com.br

[37] Brady, N.C. (1989) Natureza e Propriedade dos Solos. 7ª Edição, Atena Editora, 878. https://www.atenaeditora.com.br

[38] Schwertmann, U. and Cornell, M. (2018) Iron Oxides in the Laboratory: Preparation and Characterization. John Wiley & Sons.

[39] Richter, C.A. and Azevedo Netto, J.M. (1991) Tratamento de água: Tecnologia atualizada. Edgard Blücher, 332. http://www.cesa.eng.uerj.br

[40] Clesceri, L.S., Greenberg, A.E. and Eaton, A.D. (1998) Standard Methods for the Examination of Water. 20th Edition, American Public Health Association, American Water Works Association, Water Environment Federation, 1325.

[41] ESRI (2004) ArcGIS, Version 9; ArcView, Version 9.3. http://www.esri.com

[42] SPSS (2003) Statistical Package for Social Sciences for Windows Version 12, Chicago, USA. https://www.stata-brasil.com

[43] Lever, J., Krzywinski, M. and Altman, N. (2017) Points of Significance: Principal component analysis. *Nature Methods*, **14**, 641-642. https://doi.org/10.1038/nmeth.4346

[44] Pérez-Arribas, L.V., León-González, M.E. and Rosales-Conrado, N. (2017) Learning Principal Component Analysis by Using Data from Air Quality Networks. *Journal of Chemical Education*, **94**, 458-464. https://doi.org/10.1021/acs.jchemed.6b00550

[45] Zanella, A., Lima, L.S. and Lopes, L.F.D. (2007) Identificação de fatores que influenciam na satisfação dos clientes de um Clube Recreativo por meio da Análise Fatorial. *Revista Gestão Da Produção Operações E Sistemas*, **3**, 167-176. https://revista.feb.unesp.br

[46] Hair Jr., J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (2005) Análise Multivariada de Dados. 5th Edition, Bookman, 593. https://pt.slideshare.net

[47] Negrão, F.I. (2008) Hidrogeologia do Estado da Bahia: Qualidade, potencialidade, disponibilidade, vulnerabilidade e grau de poluição. Ph.D. Thesis, Universidade da Coruña. https://ruc.udc.es

[48] CETESB (2015) Companhia Ambiental do Estado de São Paulo. https://www.cetesb.sp.gov.br