UNIVERSITY OF SÃO PAULO

INSTITUTE OF GEOSCIENCES

COMPARATIVE GEOCHEMICAL ANALYSIS OF ROCKS WITH POTENTIAL OIL AND GAS GENERATION IN THE LLANOS, EASTERN CORDILLERA AND MIDDLE MAGDALENA VALLEY BASINS, COLOMBIA: INTEGRATED APPROACH TO CHARACTERIZATION OF DEPOSITIONAL ENVIRONMENTS, SOURCE ROCKS AND GENERATION POTENTIAL.

Yasmin Pelayo Serrano

Advisor: Professor Ph.D. Colombo Celso Gaeta Tassinari

Ph.D. THESIS

Postgraduate Program in Geochemistry and Geotectonic

São Paulo

2023

Serviço de Biblioteca e Documentação do IGc/USP
Ficha catalográfica gerada automaticamente com dados fornecidos pelo(a) autor(a)
via programa desenvolvido pela Seção Técnica de Informática do ICMC/USP

Bibliotecários responsáveis pela estrutura de catalogação da publicação:
Sonia Regina Yole Guerra - CRB-8/4208 | Anderson de Santana - CRB-8/6658

Pelayo Serrano, Yasmin
Comparative Geochemical Analysis of Rocks with Potential Oil and Gas Generation in the Llanos, Eastern Cordillera and Middle Magdalena Valley Basins, Colombia: Integrated Approach to Characterization of Depositional Environments, Source Rocks And Generation Potential / Yasmin Pelayo Serrano; orientador Colombo Celso Gaeta Tassinari. -- São Paulo, 2023.
155 p.

Tese (Doutorado - Programa de Pós-Graduação em Geoquímica e Geotectônica) -- Instituto de Geociências, Universidade de São Paulo, 2023.

1. Geochemistry. 2. Source rocks. 3. Generation Potencial. 4. Colombia. 5. Depositional Environments. I. Gaeta Tassinari, Colombo Celso, orient. II. Título.

UNIVERSITY OF SÃO PAULO

INSTITUTE OF GEOSCIENCES

COMPARATIVE GEOCHEMICAL ANALYSIS OF ROCKS WITH POTENTIAL OIL AND GAS GENERATION IN THE LLANOS, EASTERN CORDILLERA AND MIDDLE MAGDALENA VALLEY BASINS, COLOMBIA: INTEGRATED APPROACH TO CHARACTERIZATION OF DEPOSITIONAL ENVIRONMENTS, SOURCE ROCKS AND GENERATION POTENTIAL.

Yasmin Pelayo Serrano

Advisor: Professor Ph.D. Colombo Celso Gaeta Tassinari

Ph.D. THESIS

N° 666

Examining Committee

Ph.D. Colombo Celso Gaeta Tassinari

Ph.D. Mauricio Parra Amézquita

Ph.D. Veridiana Teixeira de Souza Martins

Ph.D. Haline Vasconcellos Rocha

Ph.D. Luis Carlos Mantilla Figueroa

São Paulo

2023

ACKNOWLEDGMENT

Firstly, I thank God for allowing me to live and finish this experience.

I would like my enormous gratitude to Professor PhD Colombo Tassinari, for their guidance, collaboration and patience during this long process.

To the Universidad de São Paulo, CAPES and ANH, entities that supported the development of my studies and project, provided financial support and/or geological information.

I want to thank the professor PhD Antonio Nunes of the Department of Geology of the Faculty of Sciences of the University of Lisbon, for their contributions and accompaniment during the field phase.

I want to thank the IGC professors for their teachings and their help especially Professor PhD. Mauricio Parra, who gave me knowledge and his beautiful friendship.

To the CPGeo professionals, Professor PhD Maria Helena Bezerra Maia de Hollanda, Liliana, Giselle, Izabel and Rodrigo and for all their collaboration and kindness during my stay in the laboratory.

To Erika, Bia, Silvana and Don Vasco, thank you for your help in the preparation of my samples as well as their friendship.

I am very grateful to Juan Carlos Ramirez for his love, support and sharing his knowledge during every moment of this experience. You are my happiness and strength.

To my family and friends for their unconditional love and words of support.

For my two angels, Caramelo and Clarabella.

RESUMEN PORTUGUES

O estudo das potenciais rochas geradoras de hidrocarbonetos numa bacia sedimentar tem sido feito integrando múltiplas técnicas, normalmente reservadas para diferentes tipos de estudos geológicos, incluindo ferramentas emergentes de mineração de dados. Os componentes inorgânicos da rocha estudada auxiliaram na definição das condições paleoclimáticas e paleorredox durante sua deposição, além de complementarem a proveniência e configuração tectônica da bacia em diferentes épocas. O estudo de proveniência foi acompanhado por composições isotópicas de Pb, Sr e Nd, bem como datação detrítica U/Pb em zircão. Essas técnicas mostraram procedência variável durante a evolução das bacias nas diferentes localidades estudadas. A geoquímica orgânica tradicional de rochas geradoras e a petrografia orgânica de macerais, auxiliaram na avaliação da quantidade, qualidade e maturidade de diferentes intervalos com potencial como rochas geradoras de hidrocarbonetos, que se localizam principalmente na Cordilheira Oriental e no Vale Médio do Magdalena. As rochas do Cretáceo na bacia dos Llanos têm menor potencial por sua qualidade, quantidade e maturidade. Intervalos do Cenozóico também foram identificados como potenciais rochas geradoras, mas sua maturidade é menor e só pode ocorrer em alguns locais ao longo do sopé da Cordilheira Oriental. O estudo geral beneficiou-se de dados de geoquímica orgânica de hidrocarburetos em vários locais, permitindo identificar as principais populações, e suas características que poderiam apontar para possíveis rochas geradoras dentro das bacias. Estas características permitiram identificar feições comuns em alguns intervalos das rochas geradoras, e propor uma correlação óleo-rocha geradora com base nos parâmetros estudados, incluindo a localização geográfica. Este último parâmetro mostrou controlar vários parâmetros e ajudou nas correlações e classificação das amostras como um dos melhores parâmetros em vários conjuntos de dados.

Palavras-chave: Geoquímica, Rochas Fontes, Potencial de Geração, Colômbia, Ambientes Deposicionais

ABSTRACT

The study of the potential hydrocarbon source rocks in a sedimentary basin has been made integrating multiple techniques, usually reserved for different types of geological studies, including emerging data mining tools. The inorganic components of the studied rock helped in the definition of paleoclimatic and paleoredox conditions during their deposition, as well as complementing the provenance and tectonic setting of the basin at different times. The provenance study was accompanied by isotopic compositions of Pb, Sr and Nd, as well as detrital zircon U/Pb dating. These techniques showed varying provenance during the evolution of the basins at the different locations studied. Traditional organic geochemistry of source rocks and organic petrography of macerals, assisted on the evaluation of the quantity, quality and maturity of different intervals with potential as hydrocarbon source rocks, which are mostly located in the Eastern Cordillera and Middle Magdalena Valley. The Cretaceous rocks of the Llanos basin have lower potential given their quality, quantity and maturity. Cenozoic intervals also were identified as potential source rocks, but their maturity is lower and could only occur at some locations along the foothills of the Eastern Cordillera. The overall study was benefited from organic geochemistry data from crude oil at several locations, allowing to identify the main populations, their characteristics that could point towards possible source rocks within the basins. These characteristics allowed to identify common features in some intervals of the source rocks, and propose an oil-source rock correlation based on the parameters studied, including the geographical location. These las parameter showed to control several parameters and helped in the correlations and classification of the samples as one of the best parameters in several datasets.

Keywords: Geochemistry, Source Rocks, Generation Potential, Colombia, Depositional Environments

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST ABREVIATURES AND SYMBOLS

ANH: Agencia Nacional de Hidrocarburos - Colombia

ANOVA: Analysis of Variance

API: API Gravity

CB: Choco Block

CIA: Chemical index of Alteration

CIW: Chemical index of Weathering

CPGeo Centro de Pesquisas em Geocronologia e Geoquímica Isotópica at University of São Paulo

E: Eastern

E-EC: Eastern - Eastern Cordillera

Fms: Formations

GCMS: Gas Chromatography-Mass Spectrometry

HC: Hydrocarbon

HCA: Hierarchical Clustering Analysis

HI: Hydrogen Index

ICV: Index of Compositional Variation

IEE – USP: Institute of Energy and Environment at University of São Paulo

IGc-USP: Institute of GeoScience at University of São Paulo

LA –HR-ICPMS: Laser Ablation High Resolution Inductively Coupled Plasma Mass Spectrometry

LLA: Llanos Basin

LS: Lithic Sandstone

MB: Maracaibo Block

MMV: Middle Magdalena Valley Basin

N: North

NE: Northeast

NW: North West

OI: Oxygen Index

PAAS: Post-Archean Average Australian Shale

PC: Principal Component

PCA: Principal Component Analysis

PI: Production Index

Qz-Ss: Quartz Sandstone

R: Rock

Ro: Vitrinite Reflectance

SE: Southeast

SLS: Sublithic Sandstone

SMB: Bucaramanga - Santa Marta Fault

SW: Southwest

Tmax: Rock Eval Pyrolysis temperature at transformation peak S2

TOC: Total Organic Carbon

UIS: Industrial University of Santander

V: Vein

W- EC: West - Eastern Cordillera

WKS: Wackestone;

1. INTRODUCTION

The present research project, entitled “Comparative Geochemical Analysis of Rocks with Potential Oil and Gas Generation in the Llanos, Eastern Cordillera and Middle Magdalena Valley Basins: Integrated Approach to Characterization of Depositional Environments, Source Rocks and Generation Potential”, aims to improve the understanding of the petroleum systems of Colombian sedimentary basins by a multiple approach geochemical study applying data mining techniques to existing and new data sets. The project will be developed through a research program associated with the National Institute of Science and Technology Project on Analytical Techniques Applied to Oil and Gas Exploration, coordinated by the advisor.

In the last two decades, in the area of Earth Sciences, the analysis of data sets using recent development in data mining, has improved and optimized the results for exploration and assessment of resources across different areas. The oil and gas exploration industry has seen the same developments in these techniques, with increased interest in reviewing existing data sets to unravel relationships that could be previously unseen using traditional data analysis. These tools could even be applied to multiple data sets simultaneously to compare results from different techniques and find unidentified associations.

The knowledge of hydrocarbons source rock is crucial in the petroleum systems definition, traditionally focusing in the quality, quantity and maturity of the organic matter through organic geochemistry analyses like Rock Eval pyrolysis, total organic carbon (TOC), and vitrinite reflectance (Ro). Additional techniques can be incorporated to enhance the characterization of source rocks, including elemental composition of source rocks and organic petrography, which add information about paleotectonic settings, provenance, depositional and redox conditions for organic matter preservation. The geographic and stratigraphic distribution of the source rocks properties are essential to evaluate the extent of a petroleum system, associating these properties to those of the known hydrocarbon appearances, in order to stablish an oil-rock correlation and define the extent of the petroleum systems.

1.1. Study Area

The study area is located in Colombia, focused in the Eastern Cordillera (EC) and adjacent Llanos (LLA) and the Middle Magdalena Valley basins (MMV, Figure 1). The geological history of Colombia during the Pre-Mesozoic was dominated by collisions between tectonic plates, generating multiple tectonic episodes recorded as metamorphic and igneous events. Due to its peripheral position on the South American plate, allochthonous terrain accretion has been proposed based on global paleogeographic reconstructions, geological observations, variations in ages and composition. of the basement (Cediel et al., 2003; Kay et al., 2009; Mann, 1995). Terrains accretion has generated regional discontinuities across the country that later controlled or influenced various stages of geological evolution since the Mesozoic.

Geological records suggest an active Mesozoic rift phase in part of the Triassic-Jurassic and Cretaceous across the country (Cooper et al., 1995; Pindell and Erikson, 1994; Pindell and Kennan, 2009; Sarmiento, 2019, 2011a). The initial phase of continental rift deposits corresponds to reddish layers including conglomerates, sandstones, claystones and volcanoclastic rocks in several parts of the country, mainly in the western margin of the EC and the MMV (Mojica and Dorado, 1987). Associated with these rift events, there are also intrusions of granite bodies in various parts of the country that also include the Central Cordillera (Pindell and Kennan, 2009; Restrepo and Toussaint, 1988).

This rift phase extends into the Early Cretaceous, where subsidence is mostly pronounced in the EC, basin deepening has deposited proximal facies on the eastern flank of the EC, which shift to shelf marine facies and distal facies to the west, in the MMV. Some middle uplift and subsidence events controlled the distribution of sedimentary facies in the current EC region. These conditions continued through the Cretaceous and reached the maximum level of inundation in the Cenomanian-Turonian (Etayo et al., 1976; Fabre, 1985; Sarmiento, 2019, 2011a). The rocks deposited during maximum flooding are rich in organic matter and have been identified as the main source rock in the country (Cooper et al., 1995)

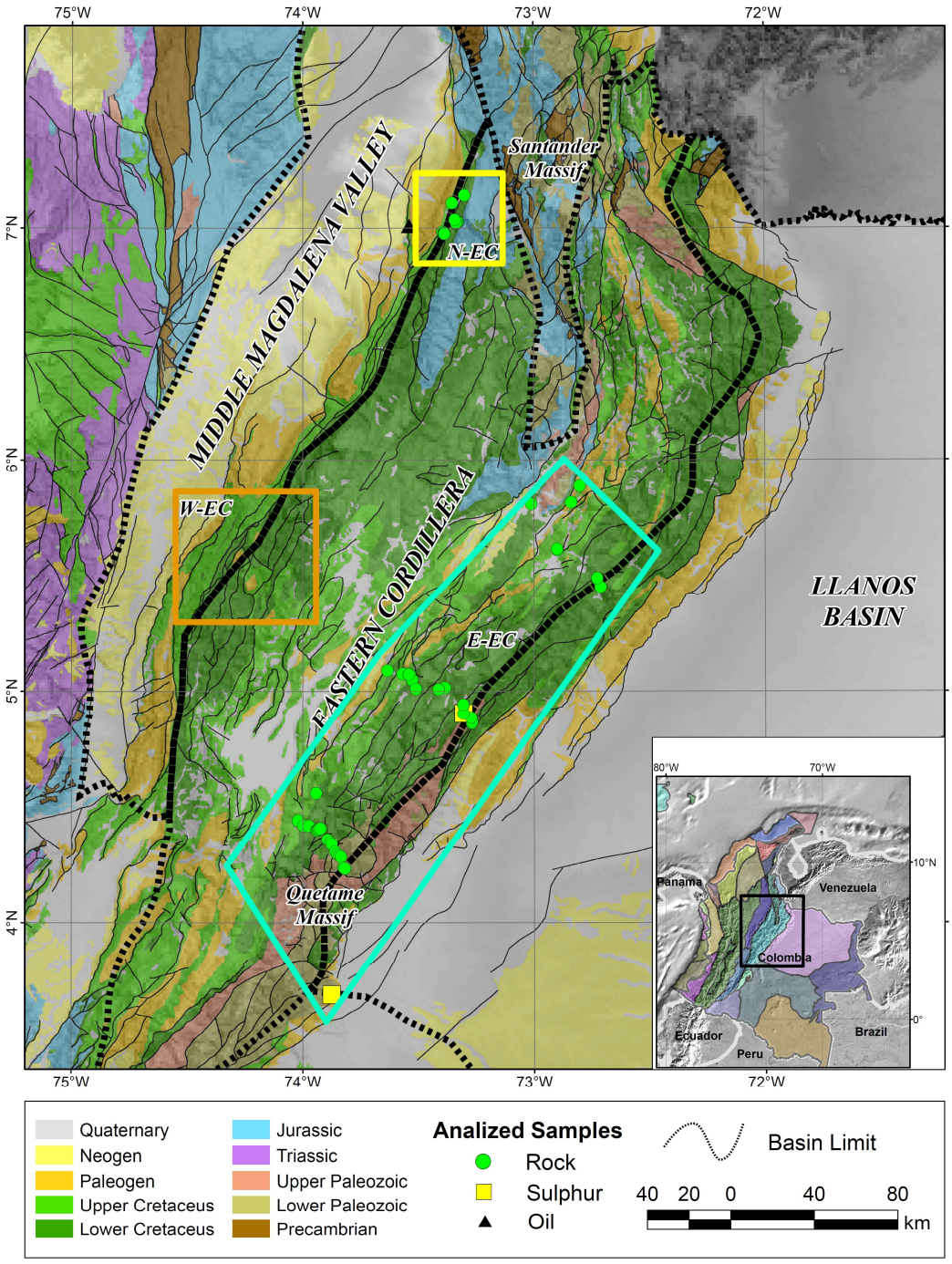

Figure 1. Geological map and location of the samples analyzed in this project. N – EC: North – Eastern Cordillera. W- EC: West – Eastern Cordillera. E -EC: Eastern – Eastern Cordillera. Geology after Gomez et al. (2015).

The end of the Cretaceous marks the limit to a transitional deposition, mainly continental in most of the country, with the exception of marine incursions in the LLA, the eastern margin of the Eastern Cordillera and northern Colombia. This mainly

continental deposition is associated with the shift in the tectonic regime to a compression environment, with the accretion of the Western Cordillera, the uplift of the Central Cordillera and the incipient inversion of some structures in the Eastern Cordillera, whose total inversion starts from the Oligocene, reaching the climax since the late Miocene to the present (Mora et al., 2010; Reyes-Harker et al., 2015). During this final inversion phase, the formation and change of many of the structural traps occurs, as well as the activation of subthrust generating pods in parts of the eastern and western foothills of the Eastern Cordillera.

1.2. **Previous Studies**

Regional studies are based on actual geographical basin limits (Barrero et al., 2007), which roughly coincides with the administrative structure from Ecopetrol and was adopted by the Agencia Nacional de Hidrocarburos (ANH). These studies can be biased by these geographical limits and often overlook the regional distribution of the elements and processes of the petroleum systems in Colombia (Figure 2), and even adjacent countries. Next are summarized some basin studies on the source rock and oil families characterization.

1.2.1. Llanos basin (LLA)

- Source rocks

Dark gray to black shales, and calcareous shales from the Gachetá Formation of the Cenomanian-Santonian, constitute the main source rock of the basin (Table 1). The unit contains type II and III kerogen capable of generating gas and oil. On the other hand, the shales of the Carbonera Formation contain some intervals with good characteristics of possible source rocks (Table 1). In the central part of the Eastern Cordillera foothills (Figure 2), the Paleocene shales in the Cuervos and Barco Formations also show good source rock characteristics (Table 1).

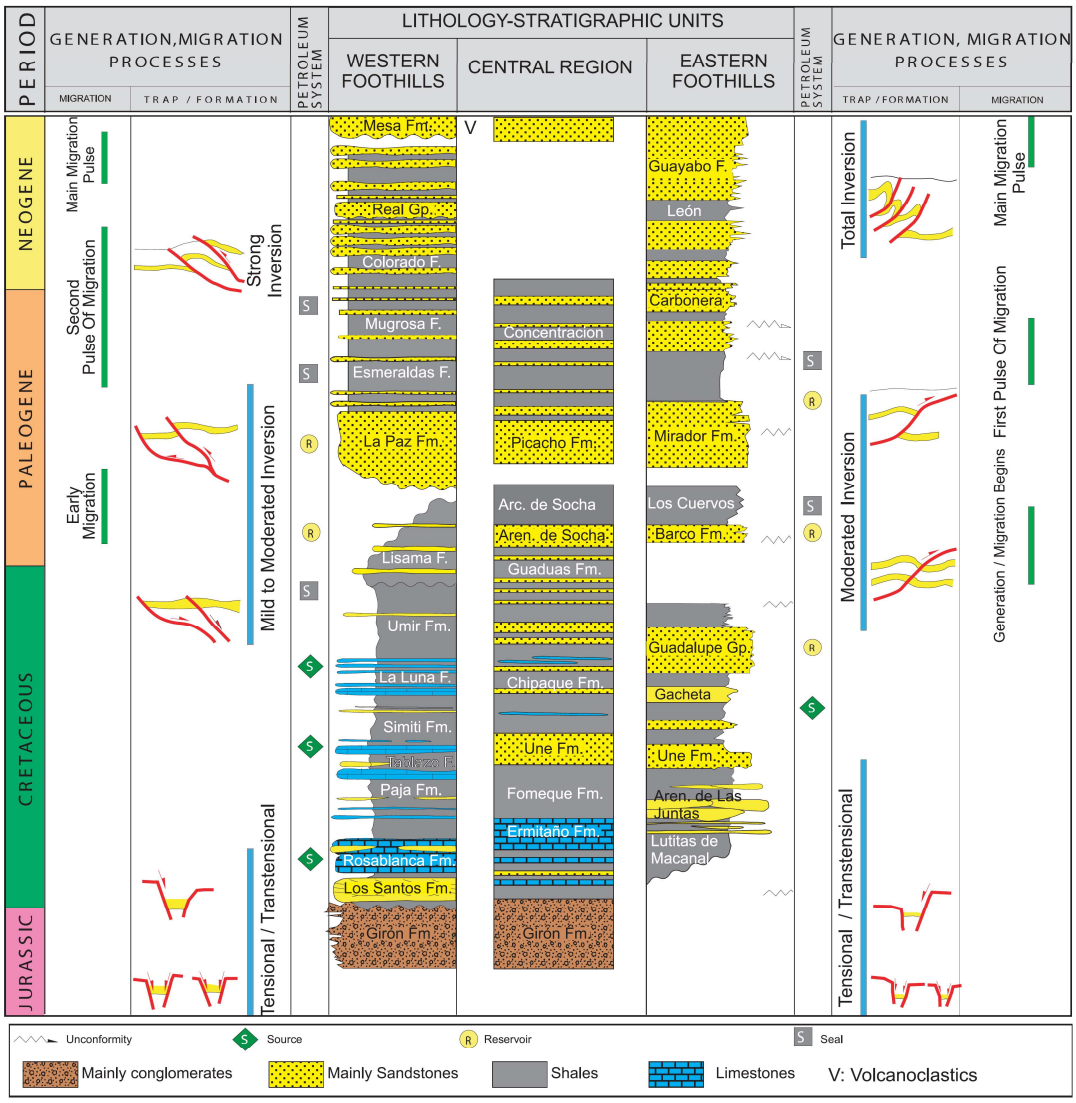

Figure 2. Schematic tectono-stratigraphic column including the nomenclature, petroleum system elements and processes of the Eastern Cordillera (Barrero et al., 2007). Eastern and Western foothills share geologic history and stratigraphic nomenclature with LLA and MMV basins respectively.

- Oil Families

Studies carried out in the Llanos basin have shown different types of oils. Garcia, (2008) presents a classification of oils based on their genetic aspects, separating them into three large families (Figure 3): two families with marine affinity, but generated by different organic precursors deposited under different oxygenation conditions, as suggested by the differences in signal isotope (δ13C) and other organic parameters (Garcia, 2008). One family representing source rocks deposited in anoxic and restricted environments, probably associated with carbonate formation (Group M), and another family (Group A), derived from siliciclastic marine rocks deposited in less

restricted environments. The third family (Group B) was generated in rocks deposited in transitional environments (deltas and/or estuaries) where the accumulation of continental organic matter predominated.

Table 1. Summary of characteristics of the main source rocks in the Cordillera Oriental basin. Extracted from (Sarmiento, 2011b)

FORMATION	ORGANIC MATTER QUANTITY EVALUATION	ORGANIC MATTER QUALITY EVALUATION	MATURITY EVALUATION	GENETIC POTENCIAL EVALUATION
Carbonera	TOC Main:<1%. Less: Until 6.5%	Kerogen type: III	Inmadure to Mature	Fair to Good
Paleocene	TOC Main:<1%. Less: Until 8%	Kerogen type: III	Inmadure to Mature	Fair to Good
Gachetá	TOC Main:<2%. Less: Until 14%	Kerogen type: II y III	Foreland: Inmadure. Foothills: Mature	Good to Excellent

Marine oils occupy the reservoirs in the domain of “foreland” basins, with the M group dominant to the south, while the A group clustered to the north. The geographic location of the petroleum genetic groups was controlled by the position of the source rocks, the dominant migration mechanism, the available fluid volumes and the pre-Andean geometry of the reservoir units (Garcia, 2008).

The M group was generated south of the Eastern Cordillera and the migration occurred through the sandy layers of the Cretaceous formations (Guadalupe and Une). This genetic group dominates the accumulations south of the basin, and is recorded in the southwest corner of the study area. The oil sources migrated northward within the Eastern Cordillera, and the more siliciclastic source rock facies produced the group A oils. The migration mechanism is also lateral, filling first the structures of the western part (current foothill) and then the foreland-like structures (Garcia, 2008).

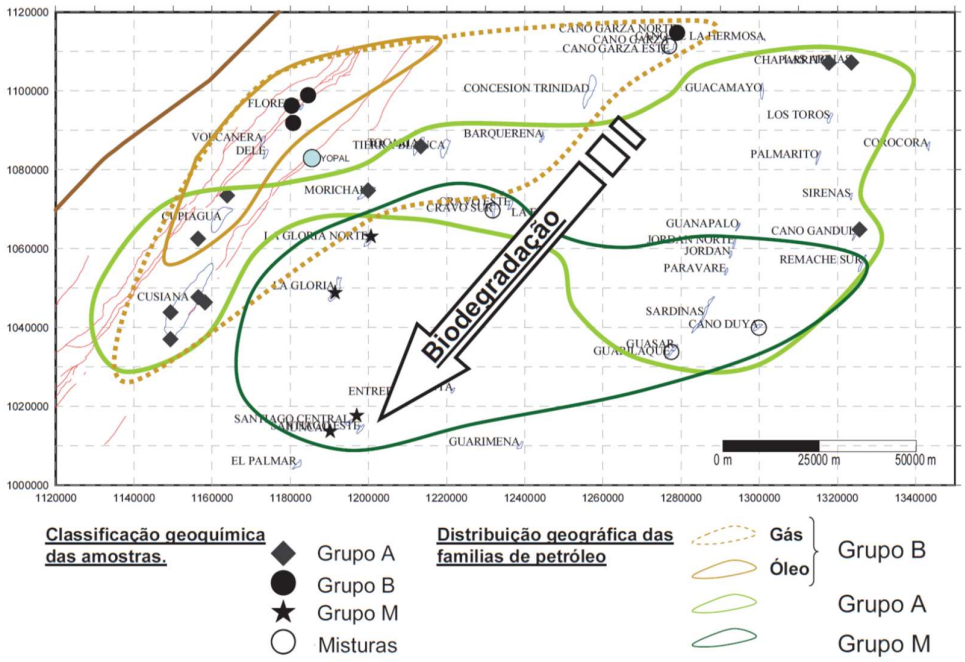

Figure 3. Map of the geographic distribution of the genetic groups of oils. There are areas of mixing of oil of different origin, and an increase in the intensity of biodegradation in the fields to the south of the region. Taken from Garcia, (2008).

Oil group B was generated by organic material derived in part from higher plant remains and accumulated in transitional environments, based on the interpretation of the ratios of various biomarkers (Garcia, 2008). The generation zones of group B communicate vertically with the reservoirs through the system of fractures and faults generated by the Andean deformation. The petroleum of this group is dominated by light oils that were concentrated in the northern portion of the studied area. These fluids mixed with biodegraded oils that previously filled the southern fields of the region (Group A oils), forming fluid systems that mixed with intensely altered oils (Group M) previously accumulated in the Cretaceous reservoirs of the western part of the “type” basin. foreland”, in whose fields layers of gas were formed above layers of heavy oils (Garcia, 2008).

More recent works such as Rangel et al. (2017), study the oils in the basins of Colombia, except for the Llanos basin. In this study, high-resolution biomarker analyzes were carried out in the main oil fields in the country. The main conclusions of the work are presented below:

1.2.2. Middle Magdalena Valley

- Source rocks

In the basin of the Middle Valley of Magdalena, the main rock that generates hydrocarbons is La Luna Fm. with more than 1% TOC and contains type II kerogen. The Paja and Tablazo Formations, made up of limestone rocks and limestone shales (Table 2), also act as generators due to their high TOC. The Rosablanca Formation has been reported as a possible source rock, however these source features have not been seen regionally throughout the basin.

The TOC percentage for the La Luna Formation is between 0.3-12.25% with an average of 2.6%, which indicates it as a good to excellent unit in terms of source rock. The Simití Formation has values of 0.55-12.08% with an average of 2.6% and is therefore a very good source rock (Table 2). The La Luna Formation contains mainly type II kerogen and produces oil. The Simití and Tablazo Formations contain lipid material corresponding to Type II kerogen as well.

The values of organic matter maturation in the Luna and Simití Formations indicate that they have reached the final phase of the oil generation window, while the Tablazo Formation has reached the over-maturation phase (Table 2).

- Oil families in the Middle Magdalena Valley

The Oil Family I have a characteristic that suggests an origin in a rock source, mostly limestone deposited on the marine platform, probably related to the La Luna Formation. The Oil Family II originated from Cretaceous marine shelf rocks, with some more siliciclastic suboxic conditions. The oils in these families are likely associated with the facies intervals of the Tablazo Fm. and/or shales of La Luna Fm. In this basin there are no oils derived from Cenozoic source rocks.

Table 2. Summary of characteristics of the main source rocks in the VMM basin. Extracted from (Sarmiento, 2011c)

FORMATION	ORGANIC MATTER QUANTITY EVALUATION	ORGANIC MATTER QUALITY EVALUATION	MATURITY EVALUATION	GENETIC POTENCIAL EVALUATION
La Luna	TOC Wt. %: > 4	Kerogen type: II	Mature	Excellent (Oil Prone)
Simití	TOC Wt. %: Averange 2.6	Kerogen type: II	Inmadure to Mature	Good to Excellent (Oil Prone)
Basal Limestone Group	TOC Wt. %: > 4	Kerogen type: II	Mature to postmadure	Good to Excellent (Oil Prone)

1.2.3. Eastern Cordillera

- Source rocks

The Cretaceous sequence of Chipaque, Une, Tibasosa and Fómeque formations constitute the main source rocks of the basin with potential for the generation of liquid hydrocarbons. On the other hand, Cenozoic formations such as Guaduas and Upper Socha with potential to generate gas are also considered generators.

The Chipaque Formation reaches a TOC level between 2-8%, which makes it an excellent hydrocarbon generator (Table 3). In the Chipaque and Tibasosa Formations, the organic matter is undifferentiated from the amorphous corresponding to type II kerogen. The Cenozoic formations present type III and IV kerogen. Hydrogen indices are strongly affected by rock maturity. Most of the Cretaceous sequence was considered with Ro values above 4% and a Tmax between 430-490 degrees.

The pyrolysis data and the visual analysis of the kerogen indicate that the organic matter of the Guaduas Formation is favorable to the generation of gas without ruling out that this formation can generate liquids if it presents type II kerogen (Table 3). In summary, the Chipaque, Tibasosa and Guaduas formations are the source rocks of the basin.

Table 3. Summary of characteristics of the main source rocks in the Cordillera Oriental basin. Extracted from (Sarmiento, 2011a)

FORMATION	ORGANIC MATTER QUANTITY EVALUATION	ORGANIC MATTER QUALITY EVALUATION	MATURITY EVALUATION	GENETIC POTENCIAL EVALUATION
Guaduas	TOC Wt. %: 1 to 4	Kerogen type: II/III	Inmadure to Early Mature	Fair to Good (Gas Prone)
Chipaque	TOC Wt. %: > 2	Kerogen type: II	Mature	Very Good (Oil Prone)
Tibasosa	TOC Wt. %: 0.1 to 4 Average 1	Kerogen type: II	Mature	Good (Oil Prone)

- Oil families in the Eastern Cordillera

A first family of oils originated from rocks from the Cretaceous marine platform, with a moderate calcareous nature. The La Luna Fm. exhibits good source rock conditions in this region. In the foothills region, the characteristics of the oils suggest an origin in siliciclastic source rocks, deposited in a proximal marine environment, probably deltaic, with a significant contribution from the Cenozoic source rock.

The geochemical data, in all Colombian basins, show that at least one set of oils was derived from source rocks deposited on a marine anoxic carbonate platform. This oil would be correlated with the La Luna formation in the MMV, and some works suggest the presence of carbonate horizons in Chipaque Formation. This is supported by studies that include biomarker data for oil source correlation, such as Rangel et al. (2000) in MMV; Rangel et al. (1991) in LLA.

On the other hand, the Hauterivian-Aptian includes marine sedimentation along the Eastern Cordillera and the MMV, where this sequence mainly corresponds to the shallow marine-inner shelf and marine-external shelf shales that have been suggested as a source rock in these basins. These rocks would be correlated with the Paja and Tablazo formations at the MMV, and the Fómeque and Une formations in the Eastern Cordillera from these rocks probably derived part of the siliciclastic marine oils, or the shale facies of the La Luna and Chipaque formations. This is suggested by studies that include biomarker data for oil source correlation, such as Rangel et al. (2002).

In addition, in the Eastern foothills of the Eastern Cordillera, there are oils that are related to very proximal facies deposited under suboxic conditions and greater

contribution of terrestrial organic matter, indicating the contribution of Cenozoic source rocks. These oils contain indicators, confirming different contributions in addition to the Cretaceous source rocks, there are previous works that also demonstrated the possibility of generating Cenozoic age intervals in these basins (Rangel et al., 1999, 1991; Tegelaar et al., 1995).

Regarding the processes after accumulation, in all basins, the effects of biodegradation were observed in varying degrees. These processes are dominant towards more stable regions, in addition to areas with more tectonic activity, away from the pedestal of the Eastern Cordillera. Instead, near the Eastern Cordillera, paleobiodegradation processes and mixing with younger, lighter oils are common.

1.3. **Justification**

Crudes from the MMV basin are shown to be related to intervals deposited in offshore platforms of Cretaceous age, with crudes not derived from Tertiary source rocks, unlike the crudes of the eastern foothills of the Eastern Cordillera associated to proximal Cretaceous source rocks (Rangel et al., 2017).

So, in the MMV was identified two families of crude oil, the first family associated with a calcareous marine platform and the second with a siliciclastic marine platform. the oil families recognized in the Eastern Cordillera basin were one associate with calcareous Cretaceous marine platform and second oil family from source rocks of Tertiary age (Rangel et al., 2017).

Previous studies carried out on organic matter and petroleum are not precise in determining the origin of oils in the Llanos basin, and the proposed correlation with source rocks is still not well understood. According to (Garcia, 2008), it is difficult to prove whether the geochemical characteristics observed are due to the mixture of oils in the reservoir or, on the contrary, they obey a lateral variation of the generating facies, probably due to the dynamics of sedimentation during the Cretaceous; in addition to different histories of burial and thermal evolution across structural blocks adjacent to the basin (Cortes et al., 2010; Garcia, 2008; Garcia et al., 2015; Rangel et al., 2017; Trindade et al., 1997; Vayssaire et al., 2014)

These previous studies were not able to clarify some doubts, as the processes of formation of hydrocarbons in the Eastern Cordillera, MMV and Llanos basins correspond to a single source rock; or whether there are multiple source rocks, in addition to the question of a possible source rock with a facial variation across the basin. Furthermore, the characteristics of the oils analyzed in previous studies indicate possible different generation pulses (Garcia, 2008; Rangel et al., 2017).

The in-depth study of petroleum systems allows, among other benefits: to reduce exploratory risks, to increase geological knowledge, to know the history of the generation and migration of hydrocarbons in the basins of NE Colombia.

Thus, the questions to be investigated in this project are:

- What are the source rocks for the different oils in the Eastern Cordillera, MMV and Llanos basins?
- Which was the generation potential of the possible source rocks in the studied basins?
- Which were the depositional environments and sediments provenance of possible source rocks in the studied basins?
- Are oil differences the product of facial variations from the same source rock level, or from different source rock levels?

These questions are the basis for the hypotheses:

- The possible generation of oils in the Eastern Cordillera, Middle Magdalena Valley (MMV) and Llanos basins from different levels of the source rocks
- The possible generation of oils in the study basins, from the same level as the source rocks, but with a change of facies

1.4. **Main Objective**

Characterize the depositional environments of the source rocks and their hydrocarbon generation potential in the Llanos, Eastern Cordillera (C.O) and Middle Magdalena Valley (MMV) Colombia, through a comparative integration of elemental, organic and isotopic geochemical analyses.

In order to achieve the main objective, the following research focuses were established:

1.4.1. Specific Objectives

- Distinguish the depositional environments and sediments provenance through elemental composition of possible source rocks.
- Characterize the composition and properties of possible source rocks of hydrothermal fluids and/or hydrocarbons, through Pb, Sr and Nd isotopic compositions.
- Analyze the sediments provenance by means of detrital zircon U-Pb Geochronology
- Compare the organic geochemical characteristics (TOC, Rock eval Pyrolysis and vitrinite reflectance) of the possible source rocks of the analyzed basins.
- Determine the depositional environment and generation potential from its organic geochemical characteristics.
- Correlate the crude families determined in previous works with the geochemical characteristics and depositional environments of the source rocks, determined in this project.

2. GEOLOGICAL SETTING

In South America, the Andes Mountain Range, from northern Colombia to Cabo de Hornos in the south, originated by the subduction of the Nazca plate and other oceanic plates below the South American plate, with varying rates of convergence since the Paleocene. (Figure 4, Taboada et al., 2000) Convergence over most of the Andes generated EW compression as well as northward transpression behavior, with NW-SE compression and NE-SW compression southward (Taboada et al., 2000).

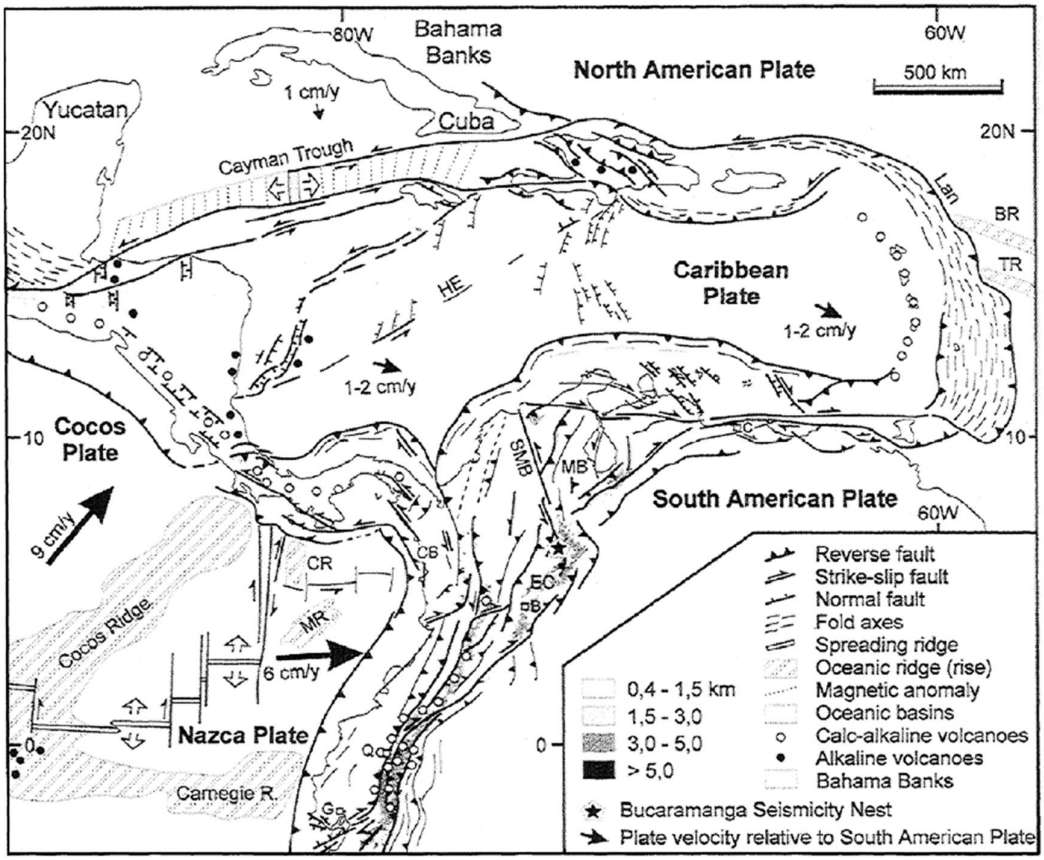

Figure 4. Colombia's tectonic map with the main tectonic boundaries. SMB: Bucaramanga - Santa Marta Fault, CE: Eastern Cordillera, MB: Maracaibo Block, CB: Choco Block. Extracted to Taboada et al., (2000).

In Colombia, the Cordillera de los Andes is divided into three mountain ranges (Eastern, Central and Occidental), which are separated by intermontane basins (Valley

of Magdalena and Cauca -Patía) and represent the three main geological provinces, the Guiana Shield, the central zone and the western zone. The Guyana shield, corresponding to the Eastern Llanos, is the most stable area in Colombia, formed by Paleozoic, Mesozoic, Cenozoic and Quaternary sediments sustained in the igneous-metamorphic basement of the Proterozoic.

The Central zone, formed by the Cordillera Central, Eastern Cordillera and the intermontane basin of the Magdalena Valley. The central zone is formed by sedimentary rocks from Paleozoic to recent, which show Cenozoic deformation events, covering the Proterozoic crystalline basement of continental affinity. The western zone, formed by the Western Cordillera and the Cauca-Patia intermontane basin, is constituted by the oceanic basement added as an allochthonous terrain, associated with the Andean deformation in the Colombian territory (Cooper et al., 1995).

The crustal extension and thinning events affected the Central zone during the Mesozoic and were mainly explained by two deformation models, the first associated with the separation between North and South America (Jaillard et al., 1990) and the second associated with extension in a back-arc environment, with a transpression component in the basin due to the oblique orientation of the plate subduction in relation to the continental margin (Cooper et al., 1995; L F Sarmiento et al., 2006; Sarmiento, 2001). This latter model indicates that the transform component decreased as the extension component increased, and was active from the Triassic to the Cretaceous in discrete events (L F Sarmiento et al., 2006; Sarmiento, 2001).

Narrow, asymmetric and small (<150 km) graben basins were developed in the region bounded to the east by the Servitá fault system, associated with a Paleozoic rift system, and to the west by the Magdalena - A Salina fault system. During the Cretaceous the basin evolved, reaching its greatest amplitude (> 180 km) during the Hauterivian - Berriasian period. Lithosphere cooling during the Late Cretaceous caused thermal subsidence, associated in turn with water loading and land accretion efforts to the West (L F Sarmiento et al., 2006; Sarmiento, 2001).

During the Maastrichtian-Paleocene, the predominant regime of marine deposition in the Cretaceous changed to a regime of transitional environments associated with the elevation of the proto-Central Cordillera, due to the addition of terrains to the West (Cooper et al., 1995; Sarmiento, 2001). This boundary has been designated as the first

phase of pre-Andean deformation, which occurred in several events during the Cenozoic, with intermediate phases of deposition in areas of the Magdalena Valley, the Eastern Cordillera and the Eastern Llanos.

Due to changes in the convergence velocities of the South American, Nazca and Caribbean plates, two later stages of pre-Andean deformation were developed during the Middle Eocene and Late Oligocene - early Miocene (Cooper et al., 1995; Sarmiento, 2001; Villamil, 1999). The elevation of the Cordillera Central, followed by deformation on the western flank of the Eastern Cordillera and ending with the deformation of the rest of the ridge, was the product of the aforementioned deformation steps.

Due to deformation events, a foreland basin was developed that encompassed the entire Magdalena Valley, the Eastern Cordillera and the Eastern Llanos during the Paleogene. The Andean deformation caused the main elevation of the Eastern Cordillera from the Late Miocene to the present, in addition to being responsible for the formation of the main physiographic features, such as the three mountain ranges, intermontane valleys and the plains of the Llanos (Cooper et al., 1995; Gómez et al., 2005; Mora et al., 2008; Parra et al., 2009; Sarmiento, 2001; Villamil, 1999). This is how the Eastern Cordillera became a barrier that transformed the Magdalena Valley into an intermontane basin and the Llanos region into the main foreland basin (Gómez et al., 2005; Parra et al., 2009).

2.1. **Stratigraphy**

2.1.1. Cretaceous Sedimentation.

The following paragraphs are based on the work of Sarmiento (2019), who presents a compilation of Cretaceous sedimentation in NW South America:

In the Jurassic, red layers from the La Quinta, Jordan-Girón, Saldaña and other equivalent formations were deposited in grabens in the Guajira Peninsula, Perijá Range, San Lucas Range, Eastern Cordillera, Magdalena Valley and possibly in the Arauca Graben in the Llanos Basin.

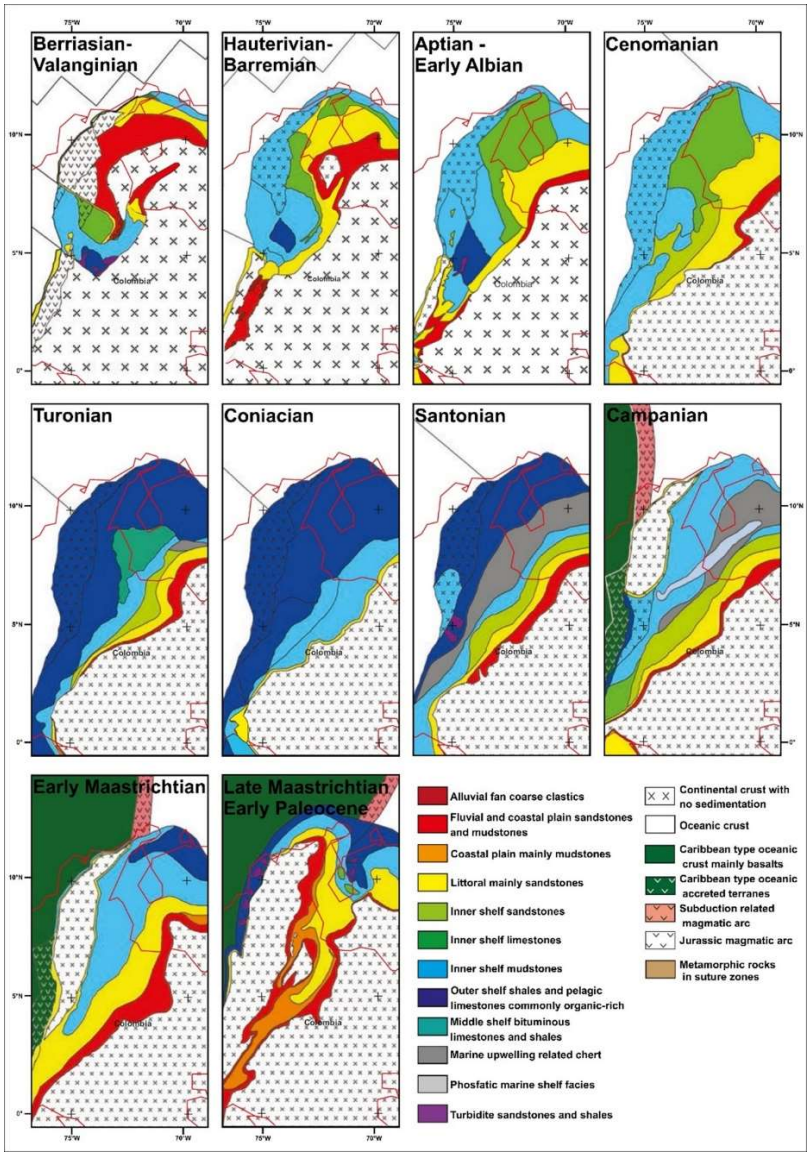

Figure 5. Paleogeographic maps for the Cretaceous, after Sarmiento (2019).

In the Cundinamarca sub-basin, in the center of the present-day Eastern Cordillera , some evidence suggests a marine incursion during the Berriasian-Valanginian that flooded an arid continental area with evaporite deposition during the early stages of the marine transgression (Figure 5, Bürgl, 1967, 1964, 1960; McLaughlin, 1972). Then, the sea left the Cundinamarca sub-basin to the north, filling two sub-basins, Tablazo-Magdalena to the west, and Cocuy to the east, while the paleomassif of Santander-Floresta remained emergent or at a shallower depth (Figure 6, Cooper et al., 1995; Etayo et al., 1976; Fabre, 1987, 1985).

In the Tablazo-Magdalena sub-basin, Valanginian sedimentation started with sandstone facies deposited in fluvial environments (Etayo and Rodriguez, 1985), in some regions without tectonic angular unconformity such as the locality of the Lebrija River. However, in other regions, Lower Cretaceous sedimentary rocks cover along an angular unconformity, Paleozoic and even Precambrian rocks. In this sub-basin, marine deposits from marginal environments recorded a marine transgression that continued with shallow water marine carbonate deposits during the Valanginian-Hauterivian (Rosablanca Fm), followed by black shales from Paja Fm. during the Hauterivian-Barremian (Etayo et al., 1976). This transgression, although continuous, presented some periods of relative retreat from the sea, as for example during the Aptian, with changes from facies rich in organic matter from the marine platform to sabka facies and algal limestones. This event is also represented by condensed sections to the east of the basin and abrupt changes to the north from organic-rich shale facies to fossiliferous and bioclastic sandstone and limestone facies.

In the Cocuy sub-basin to the east, marine transgression started in the south, as recorded by Breccias of the Buenavista Fm. and the Limestones of Guavio Fm. to the north, facies changes record the transition from continental to shallow marine sedimentation (Macanal Shale Fm., Figure 5 and Figure 6), during the Berriasian to the Valanginian period (Fabre, 1985). During the Hauterivian to Barremian, sandstones dominated by prograding deltaic environments are recorded (Las Juntas Sandstone Fm.).

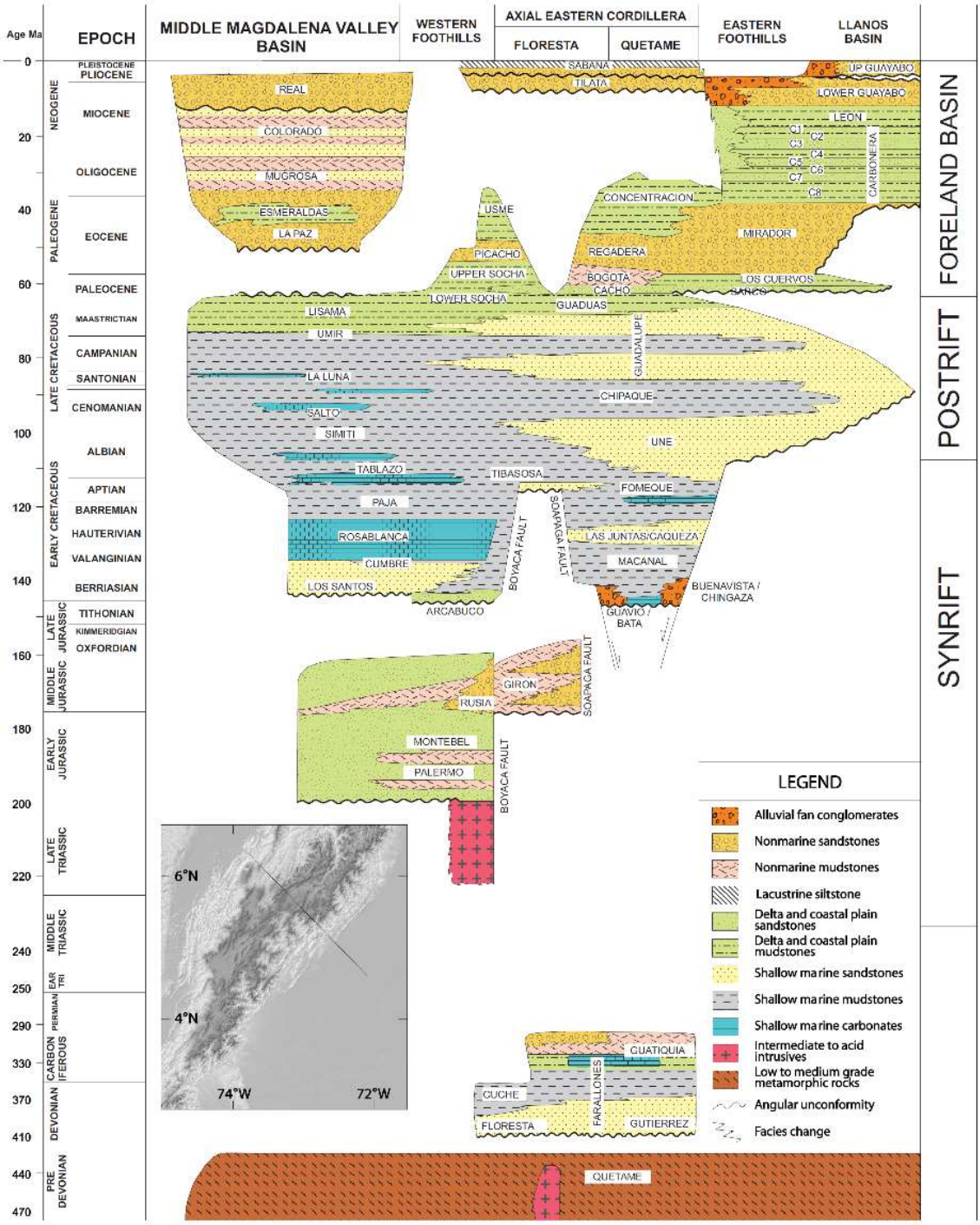

Figure 6. Stratigraphic chart in the central region of the Eastern Cordillera of Colombia. Based on works by (Mora et al., 2013, 2010, 2006) and references therein.

The Paleo-Massif of Santander-Floresta remained emergent until the Hauterivian, when the two sub-basins coalesced into a single basin. The deposits of equivalent facies of the Tablazo-Magdalena sub-basin are younger in the east and north direction,

over the Santander paleo massif -Floresta and in the Cocuy sub-basin, due to marine transgression during the Valanginian to the Aptian (Figure 6).

In the Cundinamarca Sub-basin, the subsidence exceeded the supply of sediments, in the same way as in the south of the Tablazo-Magdalena and Cocuy sub-basins, where there is an increase in shale deposits rich in organic matter up to the Aptian. This increase in subsidence gave rise to turbidite deposits on the eastern and southern flanks of the Cundinamarca sub-basin (Sarmiento, 2019). After the Berriasian, an increase in sediment supply outpaced the basin's subsidence, resulting in progradation of turbidite and delta systems to the Aptian (Figure 5).

From the Albian onward, basin subsidence was largely controlled by the cooling of the elongated crust during the early Cretaceous, and the relative change in sea level. During the Albian, a relative fall in sea level allowed the progradation of deltaic and coastal sandstones from the south and along the eastern flank of the basin in the current region of the Eastern Foothill (Fms. Une and Caballos, Figure 5). During the Middle Albian a relative sea level rise was recorded by the trend of deepening upwards in the basin with changes in the basin edges to deeper environments and deposition of distal platform facies, rich in organic matter (Fms. Tetuan and Hiló), but in the region of the Middle Magdalena Valley, a condensed section did not favor the deposition of organic matter. A relative fall in sea level caused a general trend towards shallower facies and a progradation of the Fm sandstones. It unites during the Late Albian to the beginning of the Cenomanian (Sarmiento, 2019).

The sea level later reached its maximum level during the Late Cenomanian, Turonian and Coniacian. At this time, the sea flooded the entire northwestern region of South America, and dark gray shale rich in organic matter was deposited from Venezuela to northern Peru, including the edge of the Eastern Llanos. The maximum flood surface located at the Cenomanian-Turonian boundary is characterized by a highly fossiliferous concretion horizon within the Frontera and La Luna Fms. (Figure 5).

At the end of the Cretaceous, a general regression occurred during the Santonian, Campanian, Maastrichtian and Paleocene. This regression and progradation were recorded in littoral to transitional coastal plain facies (Guadalupe Gp., Guaduas Fm.). The sandstones of Guadalupe Gp. represent two cycles of progradation, silting and westward retrogradation of the coastline, dominated by quartz sandstones derived

from the Guyana Shield (Figure 5 and Figure 6, Cooper et al., 1995; Sarmiento, 2019). Regression did not occur continuously, but with small transgressive events recorded by fine-grained phosphatic and siliceous facies (Plaeners Fm., Olini Gp., and upper part of La Luna Fm.).

This marine regression and shift to shallower environments was contemporaneous with the compressive deformation recorded in NW South America, due to the edge transformation with the obduction of oceanic terrains along the edge of the Pacific that occurred during the Campanian-Maastrichtian, (Sarmiento, 2019). At this time, the collision and accretion of the Caribbean plate fragments on the continental margin generated the Cordillera Occidental and caused uplift in the Cordillera Central and in some regions of the current Eastern Cordillera. These events induced a transition from marine to continental environments, extending from Venezuela to Ecuador (Figure 5).

2.1.2. The Cenozoic of the Eastern Cordillera and Adjacent Basins.

The deformation events in the Eastern Cordillera and its deposits are presented in several works. In the following paragraphs, the ideas of some of them are presented, such as Cooper et al. (1995); Mora et al. (2010); Reyes-Harker et al. (2015) and references therein.

According to Cooper 1995, four important deformation occurrences were recognized in the Cenozoic of central Colombia: Late Cretaceous - early Paleocene, Middle Eocene, Late Oligocene - early Miocene and Late Miocene - Pliocene. Each of these events has associated deposits and nonconformities in each region.

The first deformation event was restricted to the Central and Western Cordilleras, and occurred during the Late Cretaceous - early Paleocene and were related to the final accretion of the Western Cordillera. This deformation marks a significant shift in depositional environments along the Eastern Cordillera , Magdalena basin, and Llanos basin, from marine to continental (Figure 6), and development of an incipient pre-Andean foreland basin divided into two mega-sequences by Middle Eocene deformation in the Magdalena Valley, results in changes in direction and rate of subduction.

In the Llanos basin, both megasequences of pre-Andean basins contain mature quartz-rich river sands derived from the Guyana shield (Fms. Barco and Mirador). In contrast, in the MMV, feldspathic and lithic fluvial sands derived from the Cordillera Central dominate (Figure 5). In both regions, these megasequences are important oil reservoir rocks of different types. Middle Eocene deformation created folding and thrusting in the Middle Magdalena Valley. These folds are truncated in a nonconformity and covered by Late Eocene clastics (Ward et al., 1973). This deformation may be related to an increase in the convergence rate at that time.

Changes in tectonic plate movements documented in the late Oligocene and early Miocene did not cause any deformation in the Eastern Cordillera or the Llanos (Cooper et al., 1995). However, more recent works present thermochronological data that support the theory of a deformation in the Eastern Cordillera during the late Oligocene (Mora et al., 2010; Reyes-Harker et al., 2015). Deformations of this age were in the Magdalena Valley, where the reactivation of Middle Eocene structures created a Late Oligocene nonconformity. Collision of the Chocó terrain with the northwest margin of South America also occurred during the Middle Miocene, which may have contributed to the initial deformation in the Eastern Cordillera.

In the middle Miocene, a global sea level rise coincided with the first significant deformation and elevation in the Eastern Cordillera and initiated the Andean Mega-sequence of the foreland basin. Marine shales were deposited further east covering part of the Guyana Shield. Rising sand content to the west documents the initial partial emergence of the Eastern Cordillera and the resulting isolation of the Llanos and MMV

The main deformation of the Eastern Cordillera and Llanos Foothills began at approximately 10.5 Ma and resulted from the collision of Panama with South America. During this deformation phase, the Eastern Cordillera was uplifted and eroded. Old extension faults were reversed and new compression structures were developed. On the western flank of the Eastern Cordillera and in the Magdalena Valley, the Middle Eocene folds were reactivated. The final depositional episode is the deposition of a thick and thick continental clastic sequence (Figure 6), derived from the Eastern Cordillera during deformation and uplift, which caused the source rocks to generate hydrocarbons. Erosion deposits from the Eastern Cordillera are preserved in Guayabo Fm., in the Llanos basin and in the Real and Mesa Fms. from the MMV.

Deformation and elevation are still active, periodically causing earthquakes in Llanos foothills. Studies of the Pliocene Tilatá Fm. suggest that 1000-2000 m of uplift occurred at approximately 3.5 Ma, however, the Tilatá Fm. rests with a pronounced angular unconformity in a variety of ancient layers, indicating that some deformation preceded the deposition of the Tilatá Fm.

More recent work focused on studies of provenance, thermochronology, facies analysis and thickness of Cenozoic layers in the MMV, Eastern Cordillera and Llanos basins, resulted in a Cenozoic evolution of the northern Andes (Mora et al., 2010; Reyes-Harker et al., 2015). In these works, three main deformation events are proposed, similar to those described by Cooper et al., (1995):

- A period of deformation focused on the Central Cordillera and MMV that may have started in the Late Cretaceous, although thermochronological data point to maximum shortening and exhumation during the late Paleocene.

- A period of mostly slow deformation rates or even tectonic quiescence during the middle Eocene.

- A renewed deformation phase occurred in successive events in the different regions of the Eastern Cordillera from the center to the flanks during the Late Eocene, the late Oligocene, and the Late Miocene, until today (Pleistocene / Holocene) expressed in provenance, thermochronology and increased subsidence rates in the Llanos basin and the MMV. The sedimentary response in the Llanos basin is controlled by proximity to the area of origin, exhumation and shortening rates, relationships between accommodation and sediment supply, as well as potential paleoclimatic forcing.

3. METHODOLOGY

3.1. Sampling

3.1.1. Rock sampling

The first phase of the project corresponded to the sampling of the different rocks of the proposed petroleum systems. Three field campaigns were carried out in the Eastern Cordillera, MMV and Llanos basins, in addition to a sample located in San Vicente do Caguan. The first sampling campaign was carried out in support of professors Ph.D. Colombo Tassinari and Ph.D. Antonio Mateus, in order to guide and verify the geology of the study areas, as well as the sampling methodology of the horizons of interest. The following sampling campaigns were carried out by the student, as a complement to the initial sampling.

Four transects were studied (Location in Figure 8):

- The first through the MMV, across the formations from the basal Cretaceous (Rosablanca Fm.) to the Oligocene (Mugrosa Fm.).
- The second, close to the Machetá town, across the formations from the Paleozoic to the Guadalupe group.
- The third route was between Sogamoso and Yopal, since the Macanal Shale Fm., to the base of Picacho Fm. (Eocene).
- The fourth route was Bogotá-Villavicencio, sampling from the Paleozoic to rocks of the Chipaque Fm.

Figure 7. YP-14 outcrop of Paja Fm. in the MMV basin.

Based on the information on the rocks that are part of the different petroleum systems in the basins, a sampling of possible source rocks was carried out, corresponding mainly to black or gray shales (Figure 7, Figure 8), and of the reservoir units, seeking the best representation of the facies of the formations involved in the study. The collection of samples in the field was carried out with due care to avoid any type of contamination, but could be limited to the degree of alteration of the outcrops.

The sampling of possible source rocks was carried out in the outcrops that allowed it, from several stratigraphic levels and laterally separated points, to verify lateral and vertical changes in their isotopic characteristics.

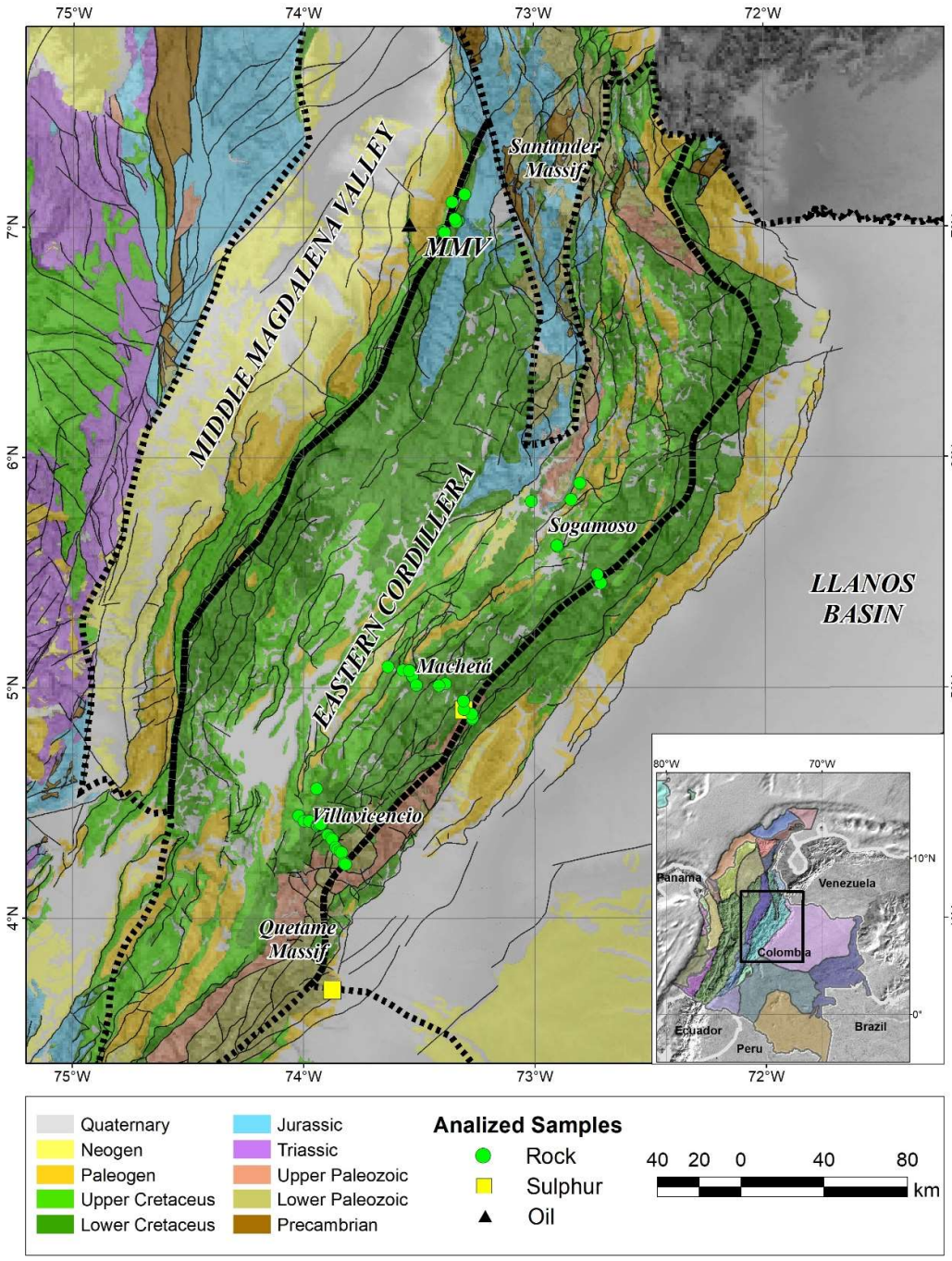

Figure 8. Location of transects and samples analyzed in the project. Rock and Sulphur were analyzed for elemental geochemistry, isotopic ratios and petrography according to Table 9. Geologic map after Gomez et al. (2015).

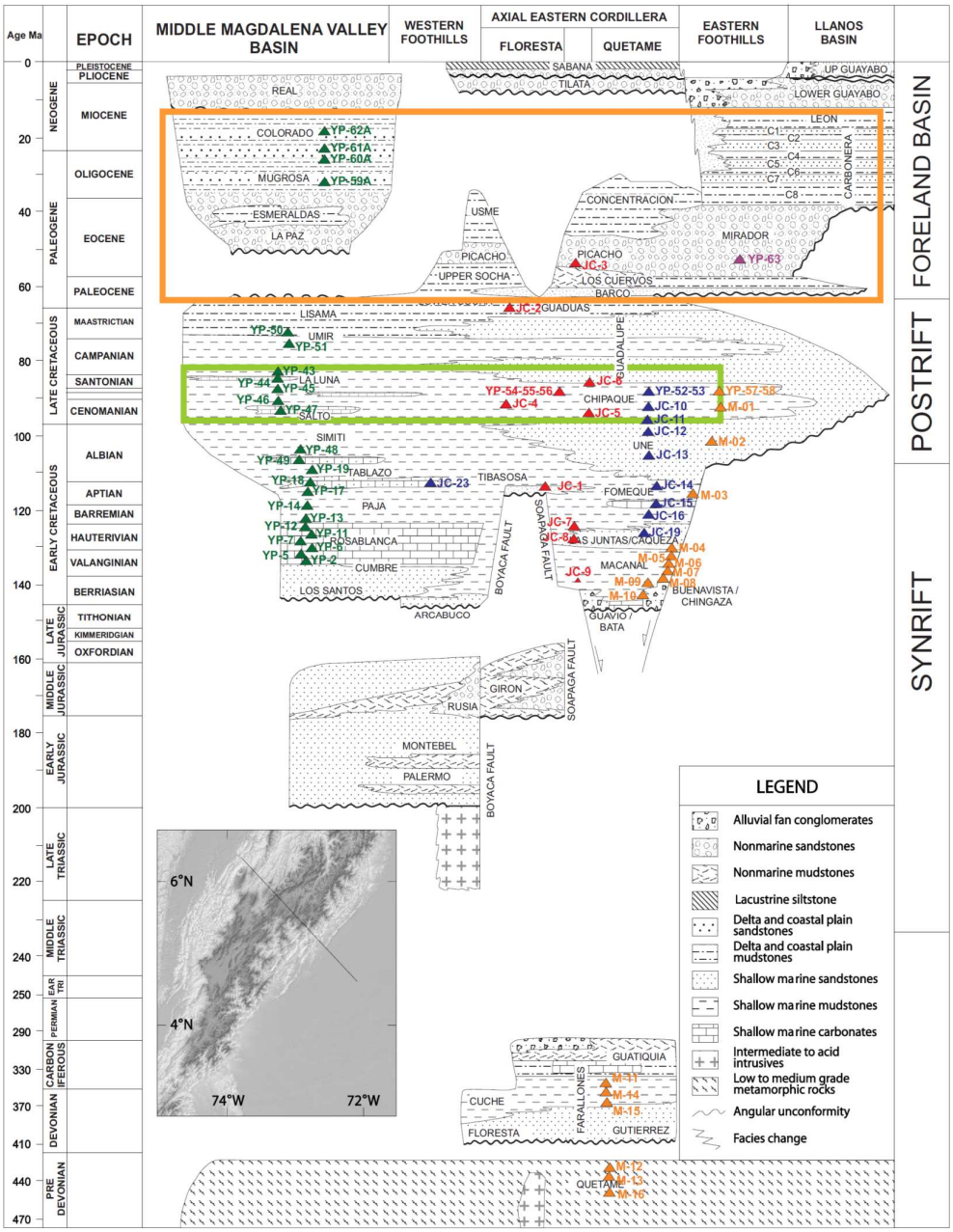

Figure 9. Location of samples on the regional stratigraphic chart. The colors of the samples correspond to each of the transects that were visited: VMM-green; Sogamoso-blue; red -Macheta; Orange -Villavicencio. The green box corresponds to the main generator rock recognized in the basins of Colombia. In the orange box are the main Cenozoic reservoir rocks in the MMV, Eastern Cordillera and Llanos basins. After Mora et al. (2010).

3.1.2. Oil sampling

The oil samples collected in this project belong to some of the main producing fields of hydrocarbons in Colombia, and were obtained during the characterization project of

the oil provinces of Colombia, carried out by ANH during the years 2014 to 2015, and whose most relevant results are shown in Rangel et al. (2017). These samples were properly preserved in approximate amounts of 15 to 20 ml, in the IEE laboratories of USP. These samples were collected for a Pb - Sr and Nd isotopic characterization, in order to make a correlation with the rock samples analyzed for the same systems.

3.2. **Elemental Geochemistry Analysis in Whole Rock**

The samples selected for multi-element geochemistry analyzes were prepared in the laboratories of receiving, washing, cutting and spraying samples at IGc-USP. These were fragmented avoiding contamination using white paper and plastic bags to wrap them and avoid contact with any metallic or contaminant surface, then the fresh fragments were manually selected, eliminating contaminated or weathered surfaces. The selected fragments were reduced in a steel pestle following the white paper method, until reaching an approximate size of 0.5 cm, and an approximate amount of 100 g, depending on the samples.

Then, the samples continued for a pulverizing process in an agate mill in the Sample Preparation Laboratory IGC-USP (Figure 10), and sent to external laboratories for analysis. The laboratory that performed the analyzes was SGS Geosol Laboratórios Ltda in MG, Brazil, with the analytical package ICM90A corresponding to determination by Sodium Peroxide Fusion - ICP OES / ICP MS (Table 4).

Table 4. Determination by Sodium Peroxide Fusion - ICP OES / ICP MS.

Ag	Al	As	B
Ba	Be	Bi	Ca
Cd	Ce	Co	Cr
Cs	Cu	Dy	Er
Eu	Fe	Ga	Gd
Ge	Hf	Ho	In
K	La	Li	Lu
Mg	Mn	Mo	Nb
Nd	Ni	P	Pb
Pr	Rb	Sb	Sc
Sm	Sn	Sr	Ta
Tb	Th	Ti	Tl
Tm	U	V	W
Y	Yb	Zn	Zr

Figure 10. Agate mill. LDA Laboratory IGC-USP

3.3. **Isotopic Analysis Pb, Sr and Nd**

The isotopic analyzes of Pb, Sr, Nd and Rb were carried out in the CPGeo of the IGc-USP, Brazil, being carried out on the whole rock samples, pyrites and oils according to criteria and possibilities.

3.3.1. Sample preparation:

Preparation of samples for isotopic analysis in whole rock were prepared in the laboratories of receiving, washing, cutting and spraying samples at IGc-USP. The samples used to carry out these analyzes are part of the samples prepared for the multi-element geochemical analysis, following the preparation procedures described earlier in this chapter with the difference that the samples were pulverized in a tungsten ring mill.

Concentration of sulfides for isotopic analysis: the samples selected for their higher pyrite content were manually reduced and concentrated, until visually obtaining the best result. Then the sample was washed and dried with a heating lamp and the magnetic material was removed. The material concentration procedure using dense liquids will be carried out following the protocols of the IGc-USP geological sample preparation laboratory, and will end with the manual selection of the material to be analyzed.

Chemical attacks on whole rock samples, sulphides and oils; the chemical separation of elements in columns by ion exchange (Figure 11), and the determination of isotopic

ratios, were carried out according to the procedures established by the CPGeo of the IGc-USP, summarized in Appendix 1, and described in detail in Estremina, (2017).

Figure 11. chemical separation of Nd element in column.

3.4. **Detrital Zircon U-Pb Geochronology**

The U-Pb method is a method that allows obtaining reliable crystallization ages, being applied to minerals that retain radiogenic lead and are resistant to alterations, such as zircon, xenothymium or titanite. Zircon is a mineral used for its advantages such as its high parent isotope content, low daughter isotope content, high system shutdown temperature, and possible detection of daughter isotope loss.

In the case of the analysis of detrital zircons, it is possible to carry out qualitative and quantitative analyzes of the populations of zircons, allowing, in the qualitative analysis, a description of the origin of the studied unit and, in the quantitative analysis, a comparison of the proportions of populations of present ages (Figure 12).

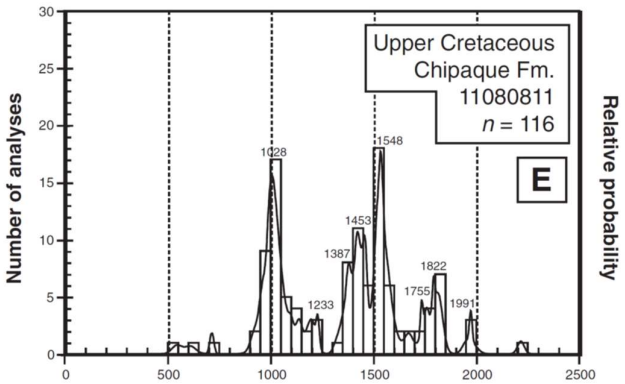

Figure 12. Distribution curve and histogram of U-Pb ages of detrital zircons. Extracted from (Horton et al., 2010)

Quantitative analyzes, to characterize the origin of sediments by U-Pb analyzes in detrital zircons, by LA –HR-ICPMS are recommended for studies of the origin of sediments and maximum ages of deposition, being carried out in this work, a selected sample of a continuous unimpregnated sandstone from an HC impregnation zone.

The sample selected for analysis of provenance corresponds to a sample of a continuous non-impregnated sandstone from an HC impregnation zone. It was not cemented as it was easy to separate and sift. The fraction <500 microns was selected to concentrate heavy minerals using a pan until an adequate concentration of zircons was guaranteed. Later it was washed with hydrogen peroxide, acetic acid to eliminate organic matter and carbonates that could keep the particles together. The sample was dried at a temperature below 60°C, and the magnetic separation was performed using a Franz-type isodynamic magnetic separator. The selected fraction was used in the separation of dense liquids (SPT and di-iodomethane) to perform the separation of zircons and apatites, mainly to finish drying. The sample was manually purified, removing contaminating materials, in order to perform the assembly for detrital zircons.

The assembly, the cathode-luminescence images (Figure 13) and the U-Pb measurements by LA –HR-ICPMS Neptune, were performed following the procedures established by the CPGeo of the IGc-USP, (Dickinson and Gehrels, 2003; Gehrels, 2000).

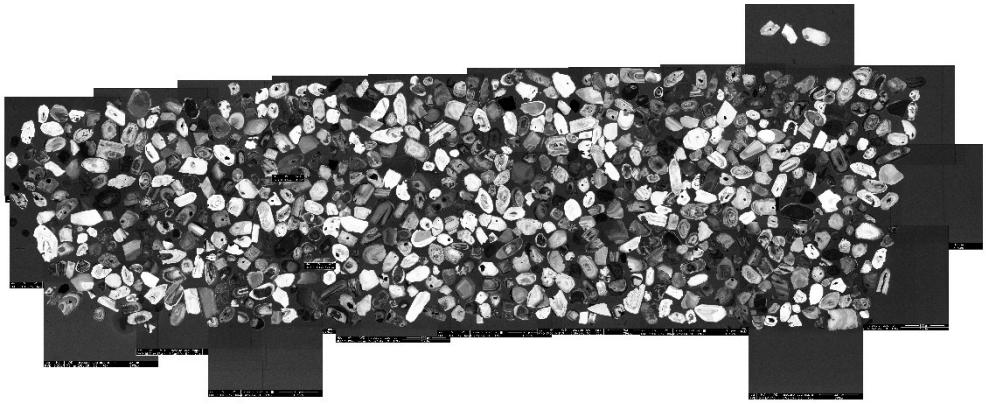

Figure 13. Mounting cathode-luminescence images of zircons for U-Pb measurements by LA –HR-ICPMS Neptune

3.5. **Organic Geochemistry**

Shale and limestone samples were analyzed by the LECO TOC method to determine their content of organic matter. The samples were fragmented and reduced in a steel pestle to the smallest size of 0.5 mm. These samples were then sent to the laboratory SGS Geosol Laboratórios Ltda in MG, Brazil, to make the analytical package CVA02B corresponding to determination of Total Carbon and Sulfur by Infrared in LECO.

3.6. **Data Mining**

Increasing amounts of data from different sources are being collected and available to us, whether from daily life style, social, business and many other areas of knowledge. Earth sciences is one of the fields with higher growth of data available from multiple sources, including field data, descriptive information, laboratory results, logging of geophysical data, and many other sources. The analysis of large datasets has been growing since the advent of the computational technology, with the creation of databases and management systems, which have evolved into advanced database systems and data analysis (Han et al., 2012).

The data analysis techniques have evolved together with the computing technology, as a need to analyze the large amounts of data that has surpassed the human capabilities. To process and analyze large volumes of data and extract valuable knowledge, beyond the human observation capabilities, or optimizing human processing of these data sets, new tools have emerged as part of the Data Mining Process to discover patterns and knowledge from large amounts of data (Han et al., 2012). The Data Mining Process was performed using the free software Orange 3.34.0 following several steps (Figure 14), from which, Data Mining itself is one of the steps (Han et al., 2012). The previous steps organize the data, and later steps analyze the results of the Data Mining in search for new information (Figure 14).

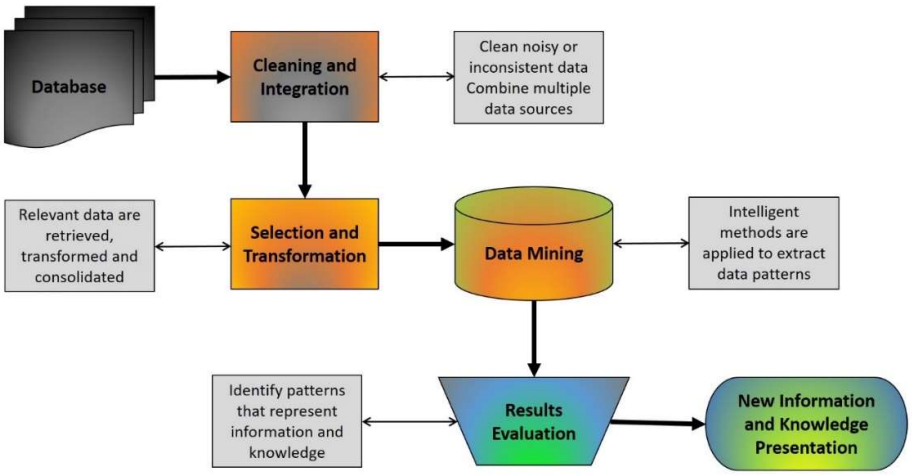

Figure 14. Data Mining Process which includes the preparation of data, the data mining, and evaluation and presentation of results. Modified after Han et al. (2012).

3.6.1. Preparation of data

The used data in the project included in-project generated data and published data from different sources for each type of data. The samples were georeferenced and classified according to their stratigraphic age and basin location following the Table 6.

- Elemental geochemistry data

The data used originated from the sources presented in Table 5. The samples from Ingeominas - UIS (2007) were filtered by formation of interest and outliers analysis was performed to remove anomalous data possibly associated to local mineralization, and to reduce the number of samples from Muzo and Rosablanca Fms. Samples from in-project were not filtered due to sparsity of the samples for each formation of interest.

Table 5. Elemental geochemistry data sources and number of samples used from each source.

Data Source	Available Samples	Used Samples
In-Project	20	20
Ingeominas - UIS (2007)	252	221

Table 6. Formation groups according to their stratigraphic age for basin studied in this Project.

Epoch	Formation	Basin			Epoch	Formation	Basin		
		VMM	EC	LLA			VMM	EC	LLA
1- Late Miocene	Guayabo Fm.			X	7- Cenomanian - Coniacian	Chipaque Fm.		X	X
	Mesa Fm.	X				Churuvita Fm.		X	
	Tilata Fm.		X			Conejo Fm.		X	
	Honda Gr.	X				La Frontera Fm.	X	X	
	Real Gr.	X				La Luna Fm.	X	X	X
2- Middle Miocene	Leon Fm.		X	X		Villeta Gr.	X	X	
	Honda Gr.	X			8- Albian - Cenomanian	Pacho Fm.		X	
	Real Gr.	X				Otanche Fm.		X	
3- Oligocene - Early Miocene	Carbonera Fm.		X	X		Hilo Fm.	X	X	
	Colorado Fm.	X				Capotes Fm.		X	
	Concentracion Fm.		X			Simiti Fm.	X	X	
	Mugrosa Fm.	X				Tablazo Fm.	X	X	
4- Middle to Late Eocene	Esmeraldas Fm.	X			9- Barremian - Aptian	Une Fm.		X	X
	La Paz Fm.	X				Fomeque Fm.		X	X
	Mirador Fm.			X		Paja Fm.	X	X	
	Picacho Fm.		X			Muzo Fm.		X	
5- Paleocene	Barco Fm.		X	X		Tibasosa Fm.		X	
	Bogota Fm.		X		10- Berriasian - Hauterivian	Buenavista Fm.			X
	Cacho Fm.		X			La Naveta Fm.		X	
	Lisama Fm.	X	X			Las Juntas Fm.		X	
	Los Cuervos Fm.		X	X		Los Santos Fm.	X	X	
	Socha Fm.		X			Macanal Fm.		X	
6- Santonian - Maastrichtian	Basal Cretaceous			X		Furatena Fm.		X	
	Cimarrona Fm.	X				Rosablanca Fm.	X	X	
	Guaduas Fm.		X			Socota Fm.		X	
	Monserrate Fm.		X			Trincheras Fm.		X	
	Umir Fm.	X	X		11- Triassic - Jurassic	Giron Fm.	X	X	
	Guadalupe Gr.		X	X	12- Paleozoic	Farallones Gr.		X	X
	Olini Gr.		X			Paleozoic			X

- Organic geochemistry data

The organic geochemistry information used comprises different types of analysis (Table 7), including Rock Eval pyrolysis, total organic carbon (TOC), organic petrography, vitrinite reflectance (Ro), and hydrocarbon geochemistry (ANH and GEMS, 2006).

Table 7. Source rock organic geochemistry data basis and number of samples used.

Data Source	Available Samples	Used Samples	Integration
Rock Eval-TOC	11772	3435	526
Organic Petrography	2993	1526	526
Vitrinite Reflectance	6022	3823	526
Crude Oil	204	132	132

These data were gathered, georeferenced, filtered from error or no values, stratigraphically located, and completed in cases as Rock Eval pyrolysis indexes, in order to have data sets as complete as possible for their use during interpretation and for the final interpretation and comparison.

- Isotopic data

Data included consists in the published data of whole rock (29 samples) and veins (23 samples) in Ingeominas - UIS, (2007) and the isotopic data included in-project. The analyzes completed consist of Pb-Pb, Nd-Nd and Sr-Sr isotopic ratios for whole rock samples of black shale (20 samples), and sulfide veins (3 samples) and one hydrocarbon sample (Table 8). These data were not filtered for outliers in order to see the possible extreme differences between formations and types of samples.

Table 8. Isotopic data included in the Project with total samples per analysis and amount of analyses per type of sample, R: rock, V: vein, H: hydrocarbon.

Source	**Pb-Pb** **R-V**	**Nd-Nd** **R-V**	**Sr-Sr** **R-V-H**
In-project	23	23	24
	20-3	20-3	20-3-1
Ingeominas - UIS (2007)			37
			14-23
Ingeominas - UIS (2008)	15	10	
	15	10	

3.6.2. Data Mining

The use of data mining tools will help in the process to know the data, their relationships, and valuable knowledge that can be obtained from these data sets.

- Characterization

The first part comprises the description and discrimination of the data sets by summarizing the general characteristics of the samples using statistical measures and plots (Han et al., 2012), also used for interpretation of the general properties of the source rocks characteristics related to provenance, depositional conditions, and hydrocarbon generation potential.

- Correlations

Knowing the relationships between the multiple features of the samples is necessary to understand the data, and find possibly unknown patterns that could correlate different datasets. The Spearman correlation defines if two variables have a monotonic connection, which is represented by simultaneous change (Barham et al., 2021), even at varying rates (non-linear correlation). This has the potential to unravel non-linear correlated processes, common during the reactions of transformation of different components. On the other hand, Pearson correlation is best suited at grading linear correlation between variables. Using the comparison between these two correlation coefficients, it is possible to determine if there is any correlation between two variables and if it is best described by a linear or non-linear correlation, based on which of the coefficients. Pearson or Spearman; is best.

To analyze these two coefficients, non-relevant coefficients were filtered for values below 0.4. The remaining values (≥ 0.4), were compared to each other using their absolute values, to finally obtain one matrix with color code ranking according to their values, indicating positive or negative correlations. Text color code was used to indicate if the best value was the Pearson or Spearman coefficient for each pair of variables.

- Sensitivity Analysis

A sensitivity analysis was performed for each data set available in order to evaluate the best combination of process and parameters that could increase the explained variance using the principal component analysis (PCA) for data reduction, as well as maximizing the silhouette scores of the K-means result. These included the ranking of variables, exploring different methods of ranking (Gain Ratio, Analysis of Variance -

ANOVA, and Chi Square - X^2), and the number of selected attributes. Also, the effect of outliers exclusion and normalization was included.

- Principal Component Analysis

The available data sets contain large amounts of variables or dimensions, which is difficult to analyze and track across different data sets. In order to simplify their interpretation, the high dimensionality must be transformed into few meaningful dimensions that can explain the data sets, and their relationships (Barham et al., 2021; Han et al., 2012; Vermeesch, 2013). The PCA is one of the most commonly used techniques for this purpose, and was applied to each of the data sets previously described in order to reduce the amount of variables and obtain significant knowledge of the data (O'Sullivan et al., 2020; Vermeesch, 2013).

- Cluster Analysis

The cluster analysis is used to group the samples based on the connection of the samples within a group, and the variations between different groups (Han et al., 2012). This is useful to find patterns and relationships, aside from their formation or age grouping previously known. The characteristics of each cluster can be used as directions to describe geologic processes of interest, and relations between different data sets.

Different clustering algorithms can yield different results, and some can best suited for different tasks. However, given the main goal of knowing the groups of samples with similar characteristics, two algorithms were selected, the K-means and the Hierarchical Clustering Analysis (HCA). The K-means allows to explore large data sets and quantify the best suitable number of groups (k) that can explain the data set. The HCA is best suited for smaller groups of datasets, due to the dendrogram visualization, and has been traditionally used to classify hydrocarbon families, as well as taxonomic applications. Even though, both algorithms have the capacity of recognizing similar clusters, given the pre-processing of the data is equal before applying them. These two algorithms were applied to the datasets and the results were compared to explore the capabilities of each one.

3.6.3. Evaluation and Presentation of Results

The results for each of the previous step is shown using scatter plots, lines plots, correlation matrices and maps to characterize the results of each step of the data mining process of the data sets.

4. RESULTS AND DISCUSION

4.1. Elemental Geochemistry Analysis

Provenance studies and determination of depositional conditions in fine-grained sedimentary rocks are possible due to the specific contribution of major, minor, and trace elements, contained in the material that makes them up, such as terrigenous materials, organic components, and mineral phases. of precipitates from the water column, among others (Goodarzi et al., 2021; Han et al., 2020; Spalletti et al., 2014). The contribution of this material allows us to know the average composition of the source rocks of the sediments, the origin of the organic components and the paleoredox conditions during deposition.

The mobility of elements such as Al, Ti, Cr, Nb, Th, Zr, Cr, Sc and rare earth elements (REE) in the provenance studies, shows how due to its constant nature, the abundance of these elements in the sediment is proportional to their source rocks. Due to the sensitivity to oxidation-reduction conditions in environments, many trace elements such as U, V, Ni, Cr, Co and their U/Th, V/Cr, Ni/Co and V/(V+Ni) ratios are used as indicators of paleoenvironmental conditions, establishing three categories as anoxic, disoxic and oxic. This is possible since the values of these ratios are inversely proportional to the oxygen content in the water table (Han et al., 2020).

During early diagenesis under oxic conditions, the content of U, Mo and V increases, while in highly productive environments with anoxic conditions, as a result of the decomposition of organic matter, Co and Ni increase (McLennan, 2001). Elements such as Fe, Mn, V, Cr, Co and Ni have shown in previous studies that they allow us to infer paleoenvironmental conditions, given that when the deposition conditions are humid, they are enriched, as well as an increase in the alkalinity of the water. Due to increased evaporation and precipitation of mineral salts and Ca, Mg, Na, K and Sr (Han et al., 2020; McLennan, 2001) the relationships of elements such as Ga/Rb and Sr/Cu allow us to infer paleoclimatic conditions at the time of deposition of fine-grained sedimentary rocks. This is due to the fact that the content of trace elements such as Ga are associated with clay minerals such as kaolinite deposited in warm and humid environments, while minerals such as illite with an enrichment in Rb content, are

associated with colder and arid paleoenvironmental conditions. Therefore, fine-grained sedimentary rocks with a low Ga/Rb ratio infer cold and dry conditions, while high ratios are associated with warmer and more humid environments at the time of deposition. In the case of the Sr/Cu content, high contents with values greater than 5.0 are associated with hot and arid environments (Han et al., 2020)

In the case of paleoproductivity, it is possible to use the TOC content of the rocks, as well as some trace elements sensitive to environmental oxidation reduction conditions such as Cr, Mo, Ni and V. The relationships between them, as well as the information obtained by organic petrography indicates the redox relationships at the time of deposition and its stratigraphic location (McLennan, 2001; Spalletti et al., 2014)

Elemental geochemical analysis was carried out on 20 samples of possible source rocks of lithologies rich in organic matter from the MMV and Eastern Cordillera basins Table 9. The elements analyzed are reported in Table 4. The results are included in Appendix .

Table 9. Samples for elemental composition for each basin

Basin	Sample	Formation
Eastern Cordillera	JC5A	Chipaque
	YP-54-COMP	Chipaque
	YP-58-COMP	Chipaque
	JC12C	Une
	JC7B	Fomeque
	JC7E	Fomeque
	JC15D	Fomeque
	YP33A	Fomeque
	YP28A	Macanal
	YP31B	Macanal
	YP27A	Paleozoic
Middle Magdalena Valley	YP43A	La Luna
	YP44A	La Luna
	YP45B	La Luna
	YP46A	La Luna
	YP47B	La Luna
	YP14A	Paja
	YP14B	Paja
	YP13A	Paja
	YP19B	Tablazo

Additional data were gathered from two studies on the flanks of the Eastern Cordillera focused on Cretaceous rocks (Ingeominas - UIS, 2008, 2007). Some of them are the hosts for the best emeralds of the world formed in deposits classified as hydrothermal origin (Branquet et al., 2002; Ingeominas - UIS, 2008, 2007). Each of these regions are known as the Western and Eastern Emerald Belts (WEB and EEB respectively). Here, they are included as the West – Eastern Cordillera (W-EC); and as part of the East – Eastern Cordillera (E-EC), as shown in Figure 15.

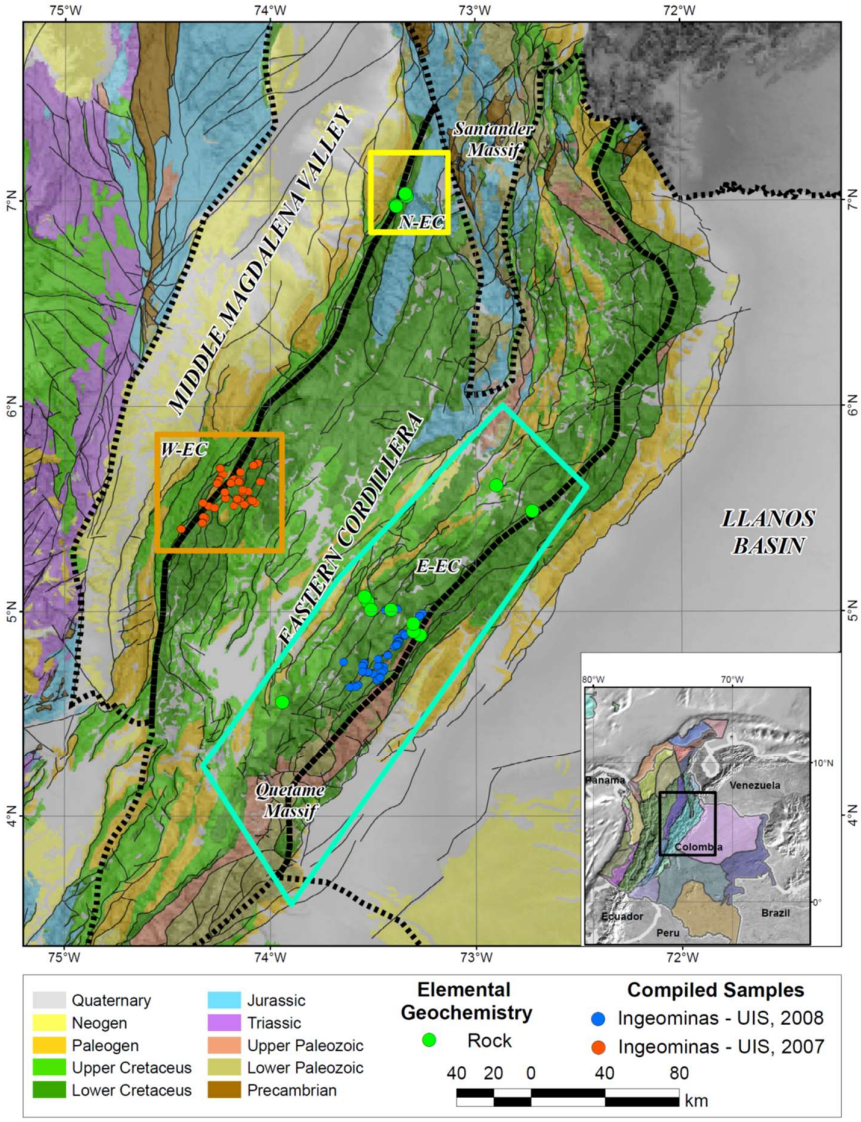

Figure 15. Location map of the samples analyzed for elemental composition in this project, and samples compiled from Ingeominas - UIS (2007 and 2008).

4.1.1. Results and Discussion

The results of the analyzed samples and gathered data were treated using normalizing procedures with respect to the aluminum content of the samples, and relative to the Post-Archean Average Australian Shale (Wedepohl, 1971); as well as calculating the enrichment factors. REE results where normalized using the values from Taylor and McLennan (1985). The analysis has been divided according to the geographic location of the samples, and their stratigraphic age. This structure allows to track changes of the observed parameters over time and across the region of study. The relative abundance of the major oxides, minor elements and REE were plotted for each of three regions, and for each stratigraphic interval available (Figure 16, Figure 17 and Figure 18).

Major oxides patterns present high CaO content on the Northern and Western EC at every interval analyzed (Figure 16 and Table 10). In contrast, the samples of the Eastern EC only show high CaO content for ages older than Aptian age (Table 10). The general patterns of the Northern and Western EC are similar between age intervals and have higher similarity for each location; whilst on the Eastern EC show higher dispersion for each age interval older than Aptian age (Figure 16 and Table 10).

The minor elements patterns show ubiquitous peaks for Cd at every interval and for the three regions of the EC (Figure 17 and Table 11). High content of Mo is also common, except for the samples of the Fomeque, Une and Chipaque Fms. of the E-EC (Figure 17 and Table 11). Apart from these remarkable features, most of the minor elements do not show noticeable characteristics that can be applied to every stratigraphic interval of any basin or any stratigraphic interval of every basin. This lack of noticeable characteristics is the basis for a later analysis using data mining tools.

The rare earth elements show smooth patterns subparallel to the base line of the PAAS (Figure 18 and Table 12), with concentrations below or close to the PAAS for most of the samples at every interval of the N-EC and W-EC, and for the intervals older than Hauterivian for the E-EC. The intervals younger than Barremian of the E-EC show concentrations above the PAAS for most of its samples (Figure 18 and Table 12).

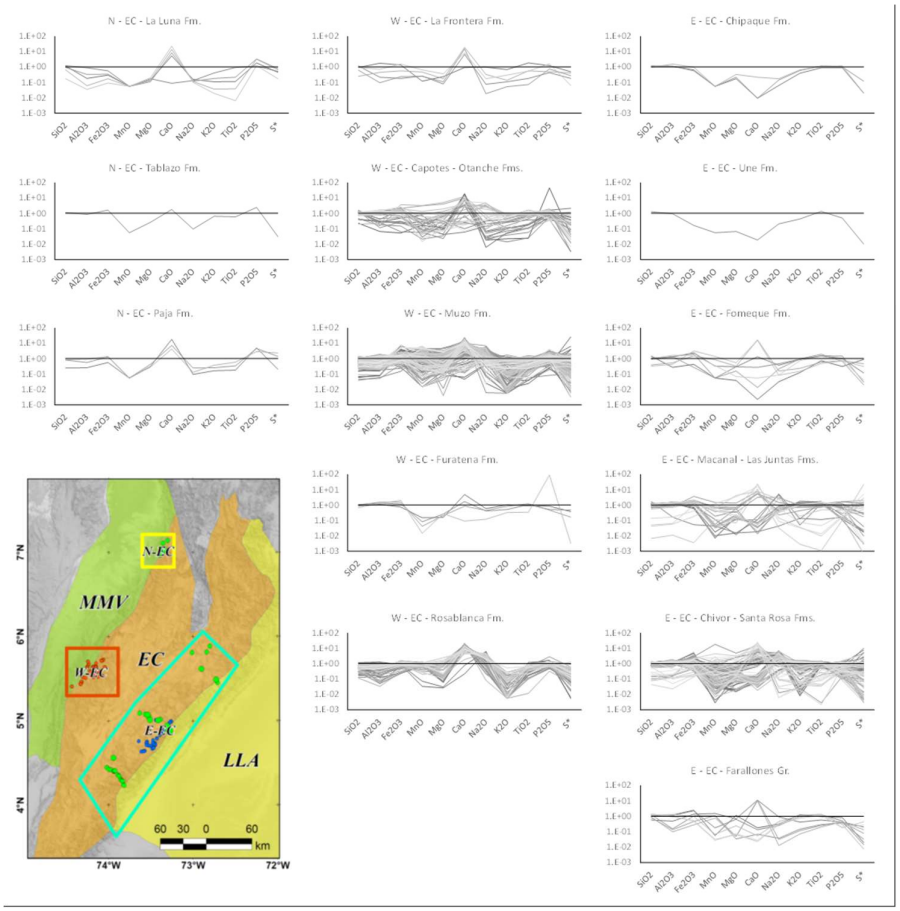

Figure 16. Major oxides normalized to the Post-Archean Average Australian Shale (PAAS, Wedepohl, 1971).

Table 10. Main characteristics of the distribution of the major oxides normalized to PAAS, according to the diagrams of the Figure 16.

Age	Northern - EC	Western - EC	Eastern - EC
Cenomanian - Coniacian	CaO >		CaO <
	P2O5 trend towards standard		
	Na2O, K2O, TiO2 < standard		Na2O, K2O, TiO2 close to standard
Albian - Cenomanian	MnO <		
	CaO >	CaO > dispersed	CaO <
	P2O5 >	P2O5 < > trend towards standard	P2O5 <
	Na2O, K2O, TiO2 < standard	Na2O < > , K2O - TiO2 < standard	Na2O, K2O, TiO2 < standard
Barremian - Aptian	MnO <	MnO < > dispersed	MnO <
	CaO >	CaO < > dispersed	CaO <
	Na2O, K2O, TiO2 <	Na2O, K2O, TiO2< > dispersed	Na2O, K2O, TiO2 < standard
	P2O5 >	P2O5 < >	P2O5 < close to standard
Valanginian - Hauterivian		MnO <	MnO < > dispersed
		CaO < >	CaO < >
		Na2O, K2O, TiO2 < standard	Na2O, K2O, TiO2 < > dispersed
Berriasian - Valanginian		Al2O3 <	
		MnO MgO < > dispersed	MnO < > dispersed
		CaO < >	CaO < > dispersed
		Na2O, K2O, TiO2 < >	Na2O, K2O, TiO2 < > dispersed
		P2O5 < close to standard	P2O5 < dispersed
Paleozoic			MnO MgO < close to standard
			CaO < > dispersed
			Na2O, K2O, TiO2 <
			P2O5 < close to standard

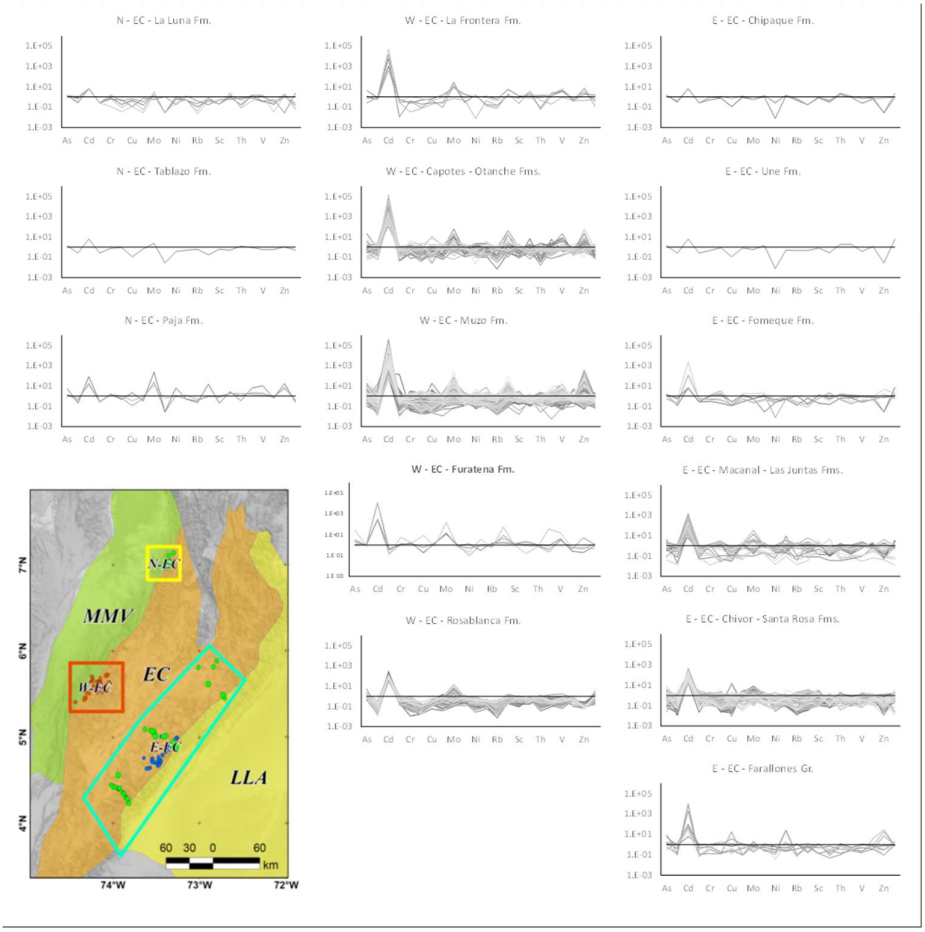

Figure 17. Minor elements normalized to the Post-Archean Average Australian Shale (PAAS, Wedepohl, 1971).

Table 11. Main characteristics of the distribution of the minor elements normalized to PAAS, according to the diagrams of the Figure 17Figure 16

Age	Relative to PAAS	Northern - EC	Western - EC	Eastern - EC
Cenomanian - Coniacian	Higher	Cd-Mo-Sr-Zn	Cd-Mo-Sb-Sr-Zn-V	Cd
	Close to	Cr-Sb-Sr-U-V	Ba-Y	Cr-Cs-Ga-Nb-Pb-Sc-U-Y-Zr
	Lower	Co-Cs-Ga-Nb-Rb-Th-Y	Co-Cr-Ga-Ni-Rb-Th	Ba-Co-Cu-Ni-Sb-Sr-V-Zn
Albian - Cenomanian	Higher	Cd-Mo	Cd-Mo-Sb-V-Zn	Cd-Th-U-Zr
	Close to	Cr-Cs-Ga-Rb-Th-Zn	Cr-Cu-Mo-Ni-Pb-Sc-Sr-Y	Cs-Ga-Mo-Nb-Pb-Rb-Sb-Sc-Y
	Lower	Ba-Co-Cu-Nb-Sb-V-Y	Ba-Co-Cs-Cu-Rb-Th	Co-Cu-Ni-Sr-V-Zn
Barremian - Aptian	Higher	Cd-Mo-Ni-Sb-Sr-U-V-Zn	Cd-Cu-Mo-Pb-Sb-V-Zn	Cd-Cs-Ga-Nb-Zr
	Close to	Cr-Cs-Ga-Rb-Th	Cr-Cs-Cu-Mo-Ni-Pb-Sc-Sr-Th-U-Y	Ga-Mo-Nb-Pb-Sc-Sr-Th-U-V-Y
	Lower	Ba-Co-Cu-Nb-Sc-Th	Ba-Co-Cs-Ga-Ni-Rb-V-Zn	Ba-Co-Cu-Nb-Ni-Rb-Sb-Sr-U-Zn
Valanginian - Hauterivian	Higher		Cd-Cr-Cu-Mo-Pb-Sb-Sr-U-V-Zn	Cd,Cs-Mo-Pb-Sb-Sr-V-Zn
	Close to		Ba-Cr-Cs-Ga-Nb-Ni-Rb-Sc-Th-U-Y	Ba-Co-Cu-Ni-Pb-Sb-Sr-Th-V-Y
	Lower		Co-Cu-Ni-Pb-Sr-U-Y-Zn	Ba-Co-Cs-Cu-Mo-Ni-Rb-Sr-U-Zn
Berriasian - Valanginian	Higher		Cd-Mo-Zr	Cd-Cu-Mo-Sb-Sc-Th
	Close to		Cr-Ga-Nb-Ni-Pb-Rb-Sb-Sc-Sr-Th-U-V	As-Ba-Co-Cr-Cs-Cu-Mo-Ni-Pb-Rb-Sb-Sc-Th-U-Y-Zr
	Lower		Ba-Cr-Ga-Ni-Rb-Sc-Th-V-Zn	Ba-Co-Cs-Cu-Ga-Mo-Ni-Rb-Sr-Th-V-Zn
Paleozoic	Higher			Cd-Cu-Pb-Zn
	Close to			Ba-Cr-Ga-Mo-Nb-Pb-Rb-Sb-Sc-Th-U-V-Y
	Lower			Ba-Co-Cs-Ga-Mo-Ni-Rb-Sb-Sr-U-Zn

Figure 18. Rare earth elements normalized to the Post-Archean Average Australian Shale (PAAS, Taylor and McLennan, 1985).

Table 12. Main characteristics of the distribution of the rare earth elements normalized to PAAS, according to the diagrams of the Figure 18.

Age	N - EC	W - EC	E - EC
Cenomanian - Coniacian	Below standard	Below standard	Above standard
Albian - Cenomanian	Close to standard	Below standard, Eu above standard	Above standard
Barremian - Aptian	Below standard	Variable, Eu above standard, Er below standard	Above standard
Valanginian - Hauterivian		Variable- below standard	Variable- below standard
Berriasian - Valanginian		Variable- below standard	Variable- Eu above standard
Paleozoic			Variable- below standard

The interpretation of this data aids in the understanding of the provenance, tectonic setting, and paleoclimatic – paleoenvironmental conditions of deposition of the units of interest in this project. Agreeing with the main goal of the project, the analysis of the data has been divided according to the nature of the information that can be extracted.

4.1.2. Lithology, Provenance and Tectonic Setting

The lithology of the samples selected for elemental geochemistry analyses was fine-grained samples, mostly shales to siltstones, with variable content of calcareous materials which can be abundant in some intervals (Figure 16 and Figure 19 left). According to the geochemical analyzes, most of the samples are classified as mudstones, ferrous mudstones and wackestone.

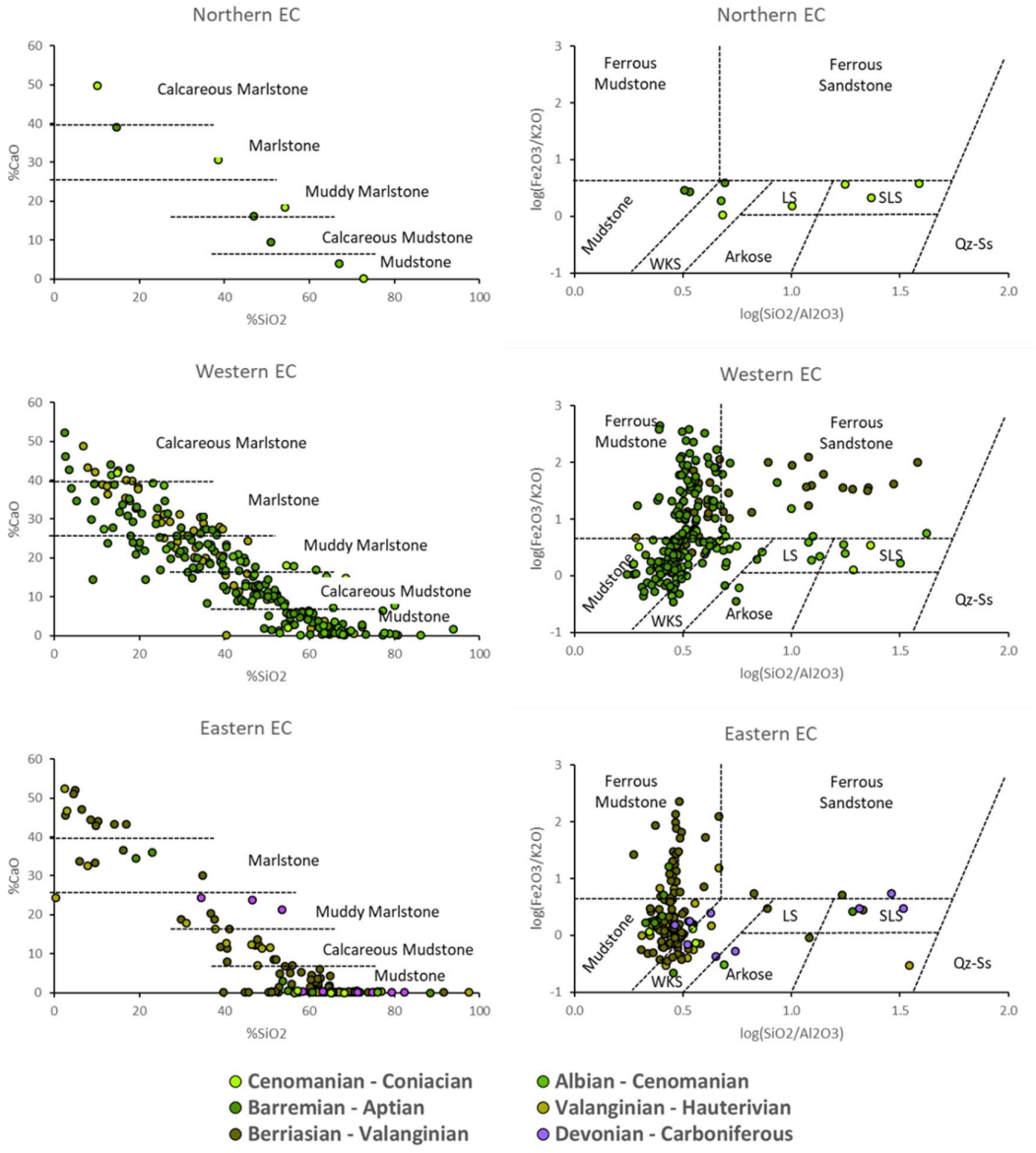

Figure 19. Left column: CaO vs SiO_2 diagram showing a lithological classification based on the content of CaO used as a proxy for carbonates content, and SiO_2 as the terrigenous fraction of the samples. Right column: Geochemical classification diagram of terrigenous sands and shales (Herron, 1988). WKS: Wackestone; LS: Lithic Sandstone; SLS: Sublithic Sandstone; Qz-Ss: Quartz Sandstone.

The calcareous content is best shown in Figure 19, where samples are spalled between mudstone and calcareous marlstone, predominantly on the N-EC and W-EC. For the E-EC, a high calcareous content is seen in samples older than Barremian - Aptian age (Figure 19 left). This variability pushes some of the limestone samples from the W-EC and N-EC to the fields of sandstones (Figure 19 right). Siltstones and sandstones from the E-EC can be seen in the arkose, lithic and sublithic sandstones fields.

Multiple approaches were used to estimate the parent rocks during the Cretaceous (Figure 20), mostly over binary plots of minor elements (Th Vs. Sc, Figure 20 left), or combined with major oxide (TiO_2 Vs. Zr Figure 20 center), and ratios of minor elements (Co/Th Vs. La/Sc Figure 20 right). The results agree for most of the samples in the three types of diagrams, showing felsic to intermediate source rocks, similar to andesitic and felsic volcanic rocks. Differing results are shown in the Th Vs. Sc diagram, where some of the samples are plotted in the mafic source domain (Figure 20 left), including samples of the Barremian-Aptian in the W-EC, and Berriasian-Valanginian in the E-EC, and one sample of the Cenomanian-Coniacian of the N-EC. These results can be properly fitted to the basement rocks of the surrounding areas, such as the Llanos basin, Santander Massif and Central Cordillera, where mostly granitic to andesitic igneous rocks are present, as well as gneisses, schists and amphibolites. This last could partially account for mafic sources, added to Mid-Cretaceous mafic extension-related magmatism on some regions of the EC (Vásquez and Altenberger, 2005).

The tectonic setting field of acid and andesitic arc fits most of the samples, with the remaining on the old sediment to passive margin setting (Figure 21 left). Just a few samples were projected in the tholeiitic arc field, corresponding with those plotted as mafic source in Figure 20-left. In the W-EC is visible a pattern of older samples located in the old sediment to passive margin, with younger samples moving towards acid and andesitic arc, and even onto the tholeiitic field. In contrast, the samples of the E-EC are dispersed over the entire area, with younger samples located onto the acid arc and old sediments fields (Figure 21 left).

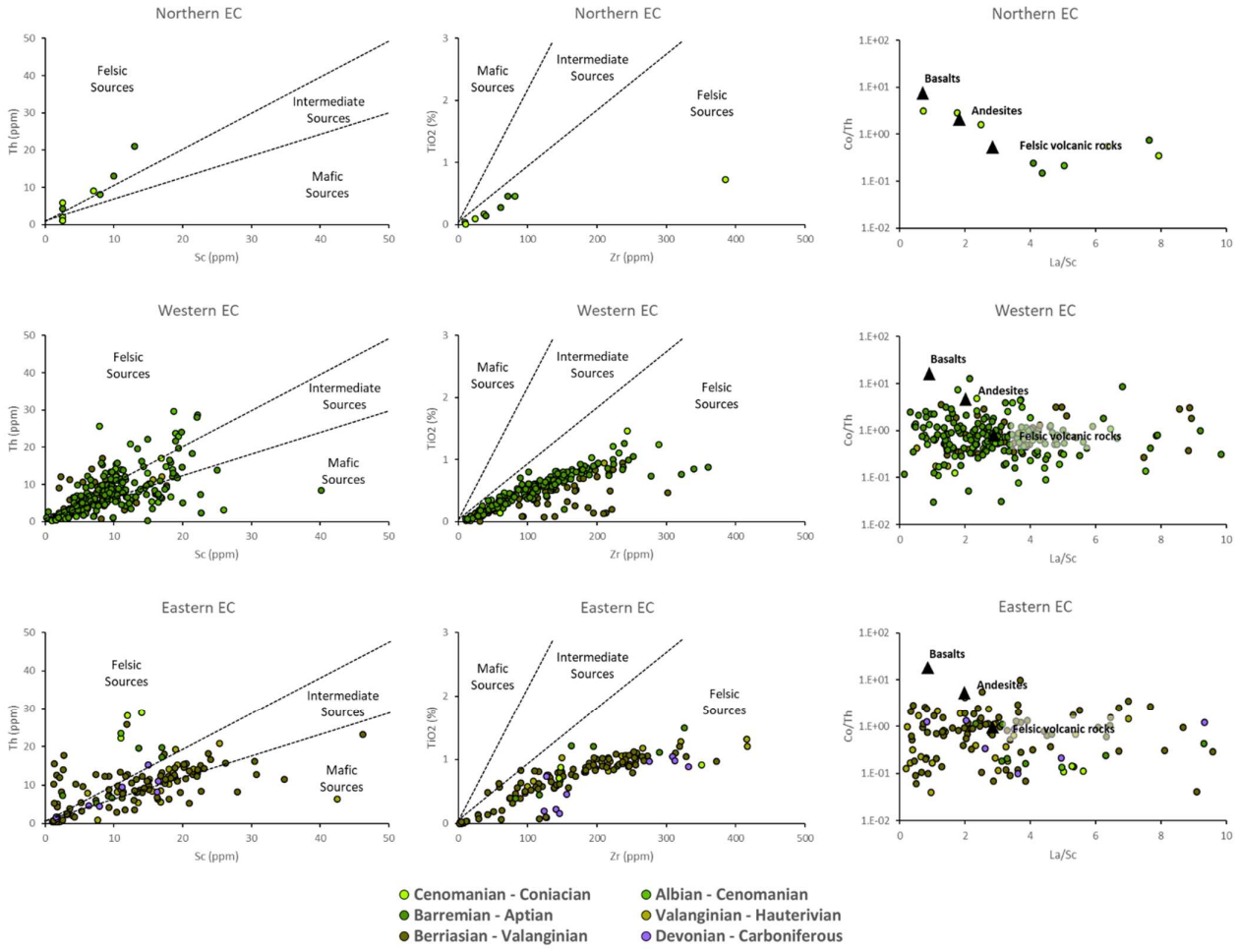

Figure 20. Left: Th Vs. Sc diagram showing provenance fields from felsic, intermediate and mafic sources (Han et al., 2020). Center: TiO2 Vs. Zr diagram showing provenance fields from mafic, intermediate and felsic sources (Hayashi et al., 1997). Right: Co/Th Vs. La/Sc diagram showing the main types of igneous rocks as a reference (Han et al., 2020).

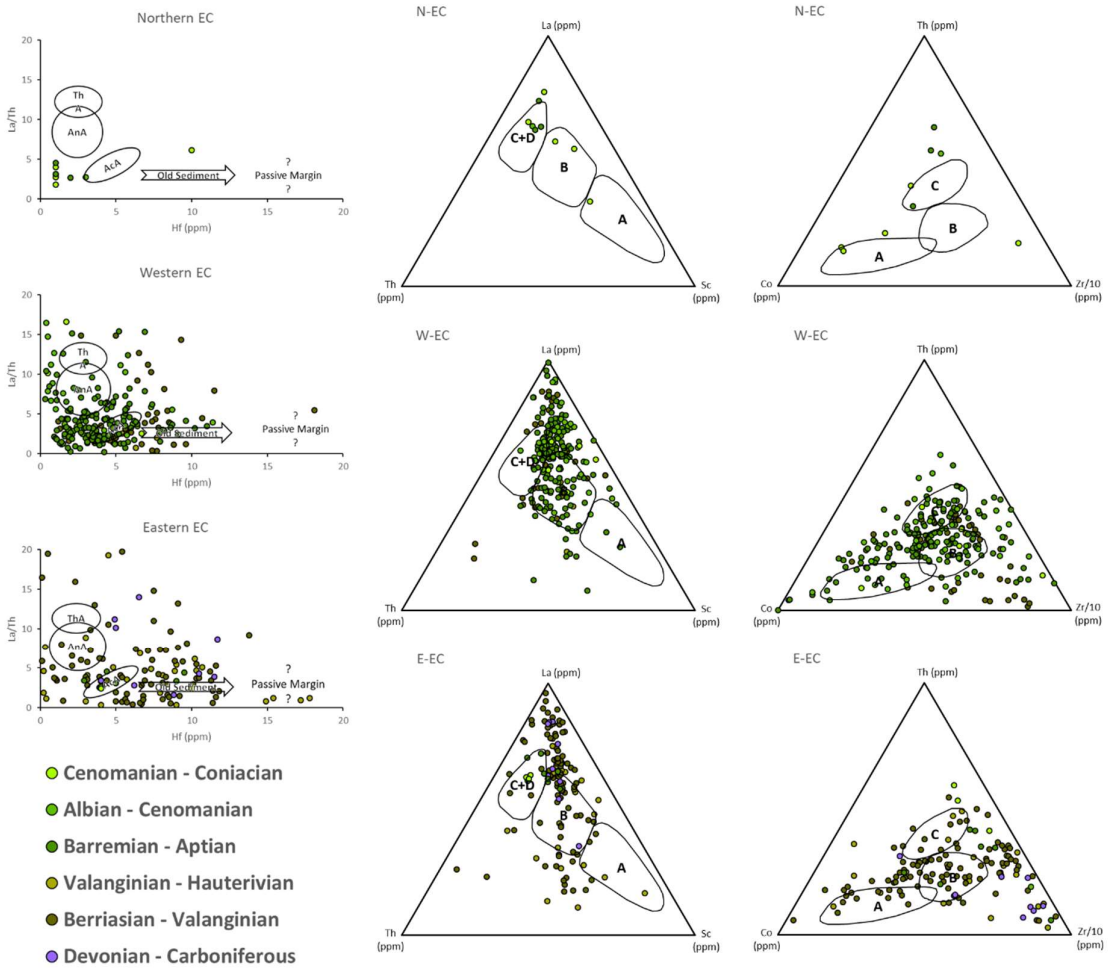

Figure 21. Left: La/Th Vs. Hf diagram for tectonic discrimination. ThA: Tholeitic Oceanic Island Arc; AnA: Andesitic Arc; AcA: Acid Arc (Floyd and Leveridge, 1987). Ternary plots of La-Th-Sc (center) and Th-Co-Zr/10 (right) for tectonic setting discrimination (Bhatia and Crook, 1986). A: Oceanic Island Arc; B: Continental Island Arc; C: Active Continental Margin; D: Passive Continental Margin

The ternary diagrams show most of the samples fit the tectonic setting fields of continental island arc, active margin and passive margin (B-C-D in Figure 21 center-right); while fewer samples show a tendency towards Oceanic island arc (A in Figure 21 center-right). These diagrams also show that older samples of the W-EC and E-EC have higher dispersion, and younger samples tend to agglomerate between the continental island arc and active margin (Figure 21 center-right). The controlling tectonic setting proposed during the Cretaceous in this region corresponds to continental active margin and island arc to the west, and continental passive margin to the east (Cooper et al., 1995; Sarmiento, 2019; Sarmiento et al., 2006; Taboada et al., 2000, among others). However, intrabasinal highs have been proposed too by the same authors, and the lack of observations of large magmatic bodies in these area (Vásquez and Altenberger, 2005), only conforms to the nonexistence of outcrops and knowledge of the deeper parts of the basin.

4.1.3. Paleoweathering, Sediment Maturity and Paleoclimatic Regime

In order to study paleoweathering conditions and sediment maturity, the chemical index of alteration (CIA), chemical index of weathering (CIW), and index of compositional variation (ICV) were calculated for these samples as follows:

$$CIA = [Al_2O_3 / (Al_2O_3 + CaO^* + Na_2O + K_2O)] \times 100$$

$$CIW = [Al_2O_3 / (Al_2O_3 + CaO^* + Na_2O)] \times 100$$

Where $CaO^* = CaO - P2O5 \times 10/3$; if $CaO^* < Na_2O$, then use CaO^* (Han et al., 2020). If not, $CaO^* = Na_2O$. The CIA shows the ratio between original minerals and secondary products (i.e. clay minerals); showing higher values up to 100 as the weathering increases. The CIW is similar to the CIA calculation, and is used to evaluate the weathering intensity of the source areas of the sediments (Han et al., 2020). Another index used is the index of compositional variation (ICV), being highest in highly weathered minerals and decreases for more stable minerals. It's calculation is shown next:

$$ICV = (Fe_2O_3 + K_2O + Na_2O + CaO^* + MgO + MnO) / Al_2O_3$$

CaO* is used with the purpose of reduce the effect of calcareous samples with high CaO content. Samples with ICV values lower than 1, are compositionally mature and were deposited under stable conditions with active sediment recycling. On the other hand, ICV > 1 are classified as compositionally immature, and deposited in active tectonic settings (Han et al., 2020).

The paleoclimatic conditions were studied using the ratios of Ga/Rb and Sr/Cu, which can reflect these conditions during deposition of fine-grained rocks (Roy and Roser, 2013). Clay minerals indicators of high weathering under warm and humid conditions are enriched in Ga (i.e. kaolinite); whilst cold and arid climatic conditions can be related to illite, which is enriched in Rb. Lower Ga/Rb ratios indicate drier and colder paleoclimatic conditions. Drier conditions are also reflected in the Sr content, due to increase of water alkalinity and precipitation of salt minerals (Han et al., 2020). Lower Sr/Cu (<5.0), and high Ga/Rb (>0.3), indicate warm and humid paleoclimatic conditions. Values higher than 5.0 for the Sr/Cu ratio indicate arid and hot conditions (Han et al., 2020).

Given the high CaO content from carbonates, and even with the correction using P_2O_5 and Na_2O as a reference shown previously (Han et al., 2020); differences still arise between the high and low CaO content samples, not only because the W-EC shows more samples with high CaO content (Figure 16). A limit of 12% was used as a cutoff, which is close to the maximum Na_2O content of all the samples. These groups are shown separately in Figure 22 and Figure 23. The main difference between the groups is that samples with higher CaO content plot on the fields of Non Altered/Non Weathered – Immature sediments (Figure 23); whereas the samples with lower CaO content plot on the fields of Weak to Strong Altered/Weathered – Immature to Mature sediments (Figure 22). About the paleoclimatic conditions, the least calcareous rocks suggest that were deposited under Hot Humid to Arid conditions (Figure 22), especially for the E-EC; with some samples spread on the Warm Humid to Arid conditions. In any of the groups, the younger samples (Albian to Coniacian) have a tendency to Hot Arid paleoclimatic conditions; and older samples present variable conditions during deposition. With respect to weathering, the samples with low CaO content, show a tendency to high Alteration/Weathering/Mature for younger samples and Paleozoic samples of the E-EC.

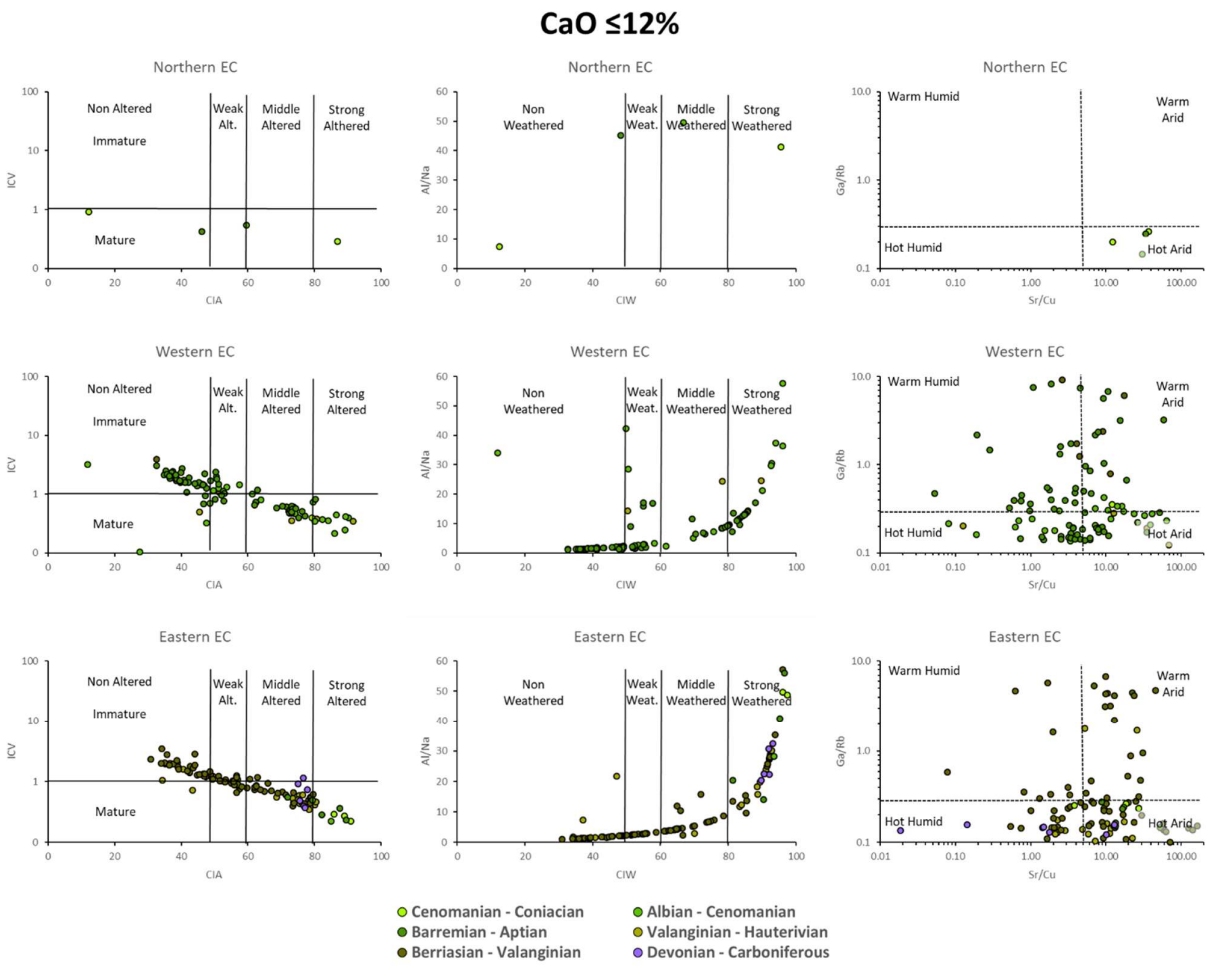

Figure 22. Samples with CaO content under 12%. Left: diagram of ICV Vs. CIA indexes to show the maturity and chemical alteration of the source of the sediments (Baiyegunhi et al., 2017; Han et al., 2020). Center: diagram of Al/Na Vs. CIW to indicate the weathering of the source of the sediments (Han et al., 2020). Right: diagram of Ga/Rb Vs. Sr/Cu to show the paleoclimatic conditions (Roy and Roser, 2013).

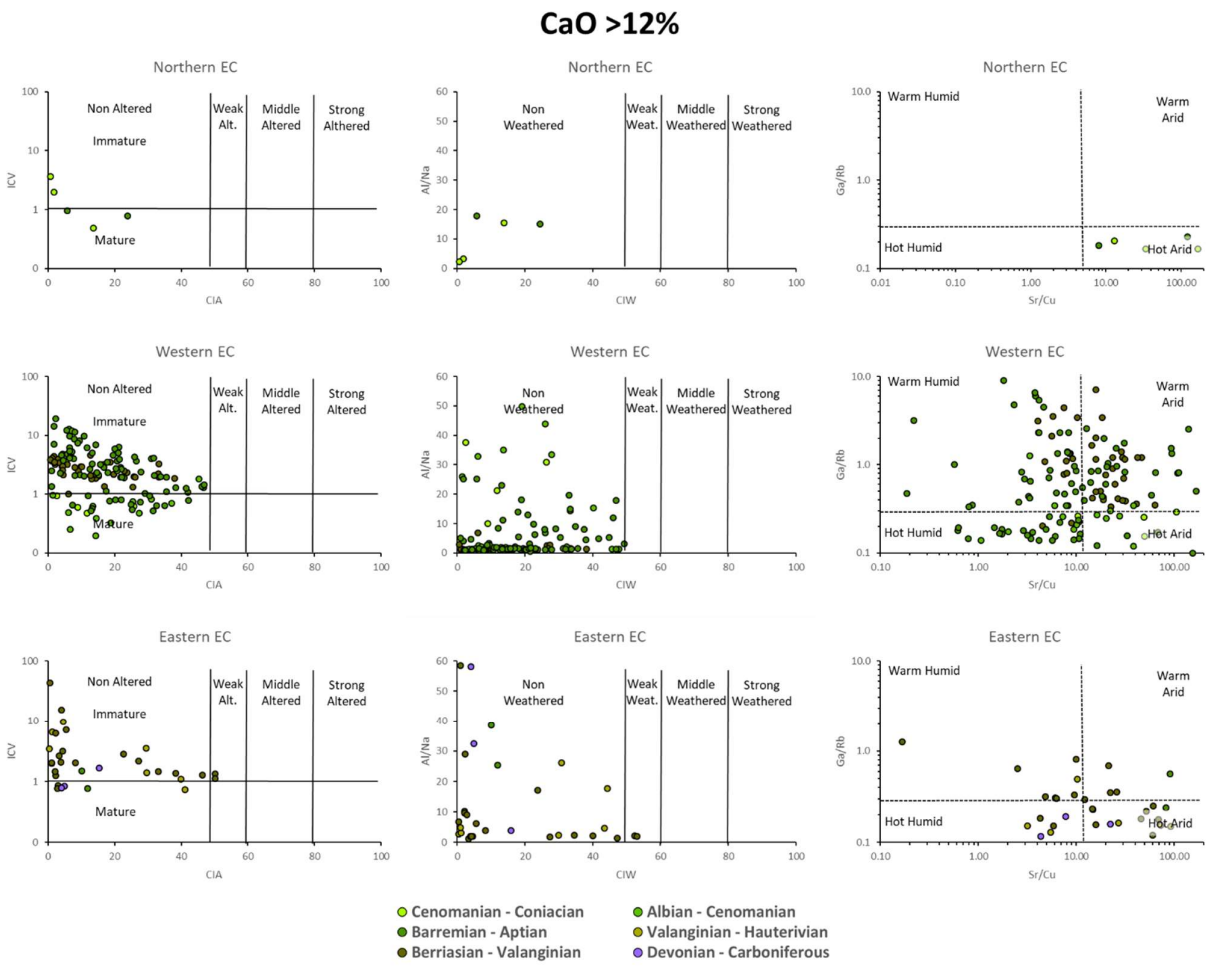

Figure 23. Samples with CaO content above 12%. Left: diagram of ICV Vs. CIA indexes to show the maturity and chemical alteration of the source of the sediments (Baiyegunhi et al., 2017; Han et al., 2020). Center: diagram of Al/Na Vs. CIW to indicate the weathering of the source of the sediments (Han et al., 2020). Right: diagram of Ga/Rb Vs. Sr/Cu to show the paleoclimatic conditions (Roy and Roser, 2013).

4.1.4. Paleoredox Conditions

Due to the sensitivity to oxidation-reduction conditions in depositional environments, many trace elements such as U, V, Ni, Cr, Co, Mo; and their U/Th, V/Cr, Ni/Co and V/(V+Ni) ratios are used as indicators of paleoenvironmental conditions, establishing three categories as anoxic, disoxic and oxic. This is possible since the values of these ratios are inversely proportional to the oxygen content in the water table (Hatch and Leventhal, 1992). During early diagenesis under oxic conditions, the content of U, Mo and V increases, while in highly productive environments with anoxic conditions, as a result of the decomposition of organic matter, Co and Ni increase (Jones and Manning, 1994; Tribovillard et al., 2012).

The ratios used for interpretation of paleoredox conditions show that the samples of the N-EC and W-EC were mostly deposited under disoxic to anoxic conditions (Figure 24). However, the samples of the E-EC show different response depending on the ratios used, with the U/Th showing oxic conditions, the V/Cr and Ni/Co show oxic to disoxic conditions, and V/(V+Ni) show disoxic to anoxic conditions (Figure 24). The sulfur content of the samples is in average below 1%, except for the Barremian-Aptian of the N-EC and the Valanginian of the E-EC (Figure 24). The average total organic carbon content (TOC), is higher than 1% for most of the intervals, except for the Paleozoic samples that is lower. The average TOC of the Barremian-Aptian and Cenomanian-Coniacian samples of the N-EC have values of 7-8% (Figure 24).

The combination of some of these parameters can improve the knowledge of the evolution of the basin. Using the ratios of Mo Vs TOC, it is possible to determine the restriction conditions of the basin over time. The N-EC show increasing restriction over time, reaching conditions of strong restriction during the Cenomanian-Coniacian (Figure 25 left). On the other hand, The W-EC shows a pattern of evolution from strong restriction during most of the Early Cretaceous, to weak restriction since the Barremian-Aptian up to the Cenomanian-Coniacian (Figure 25 left). The E-EC shows a few samples of the Berriasian-Valanginian-Hauterivian going from strong restriction to weak restriction. For the remaining samples, the clustering of the data does not allow for identification of a pattern of restriction conditions (Figure 25 left).

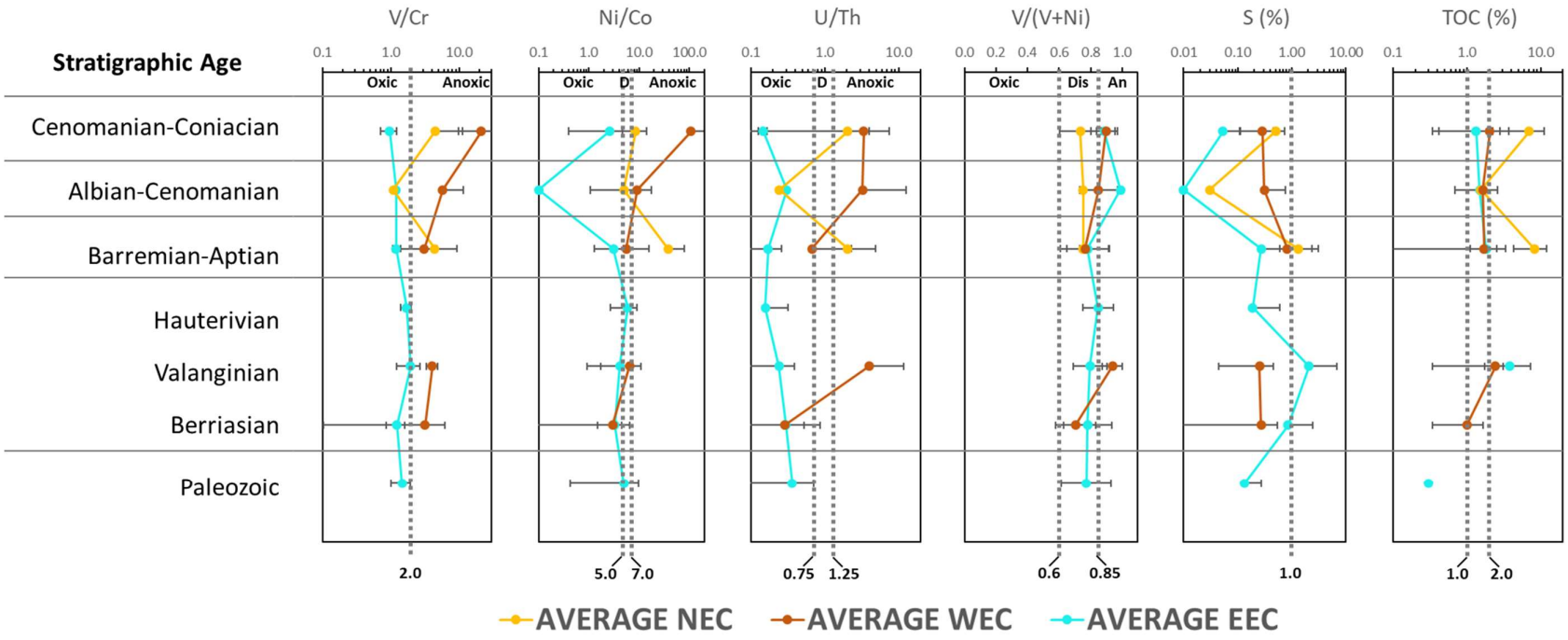

Figure 24.Avergae profiles for multiple proxies commonly used paleoredox conditions interpretation. The thresholds used for V/Cr, Ni/Co, U/Th and V/(V+Ni) after Goodarzi et al. (2021). Sulfur threshold used for crude oil depositional environment (Lewan, 1984). TOC threshold used for source rocks quantity assessment.

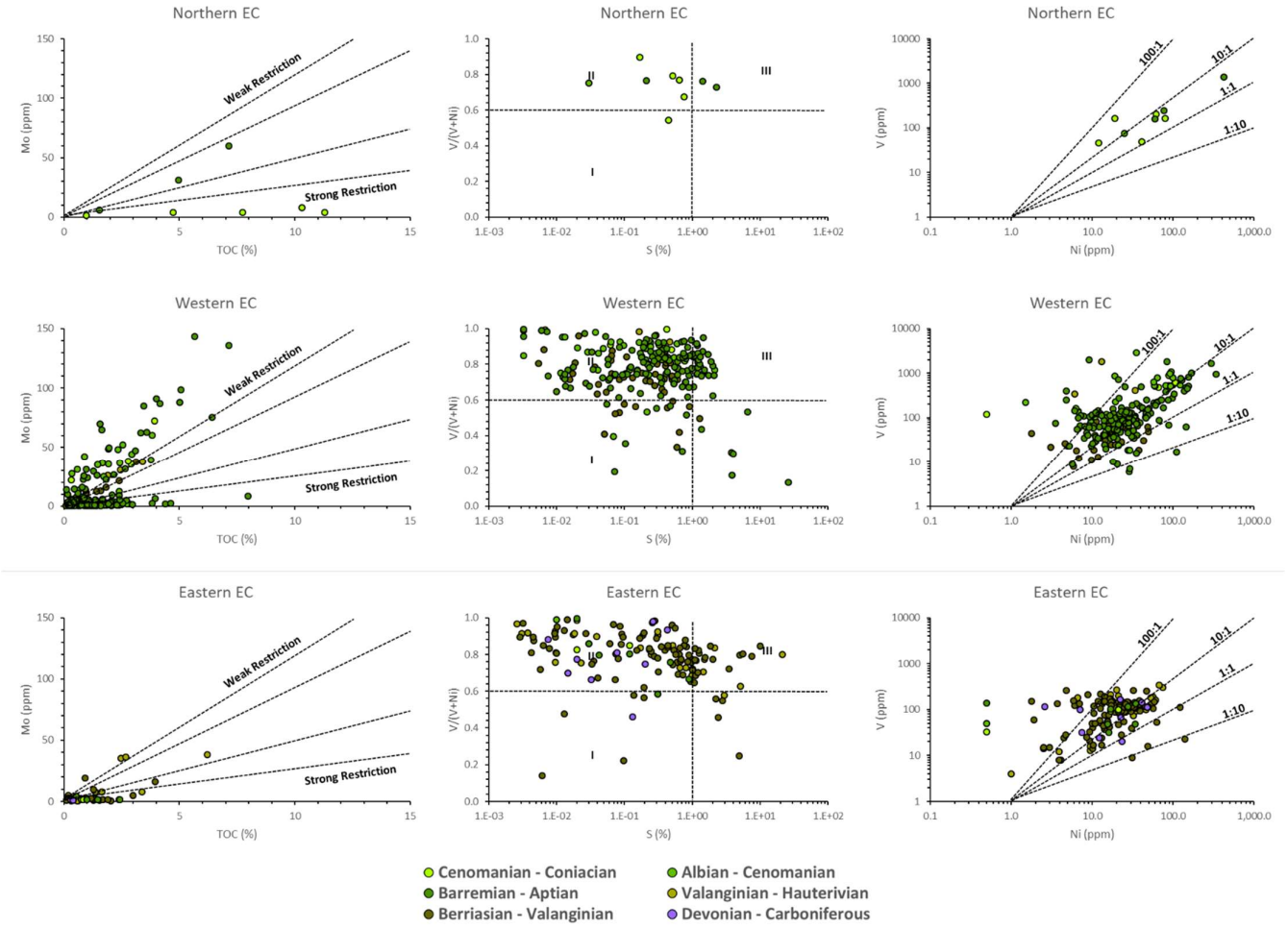

Figure 25. Left: diagram of Mo Vs. TOC used to assess the degree of restriction in oxygen-limited marine basins (Tribovillard et al., 2012). Center: diagram of V/(V+Ni) Vs. Sulfur used for interpretation of depositional environment of oil source rocks. Right: V Vs. Ni diagram showing the proportionality between these elements for all the samples, frequently used for oil to source rock correlation (Lewan, 1984; López and Lo Mónaco, 2017).

The diagrams of the V/(V+Ni) Vs. Sulfur, and V Vs. Ni (Figure 25 center and right), are routinely used for the characterization and interpretation of the depositional environment of oil source rocks (Lewan, 1984; López and Lo Mónaco, 2017; Rangel et al., 2017). Here they are used with the possible source rocks to learn their distribution for later comparison with oil geochemistry data. These diagrams show that most of the samples correspond to the areas II-III, where the interpretation suggest siliciclastic and carbonatic marine source rocks (Figure 25 center). However, the use of this diagram with source rocks data and comparison with oil data, can be affected by multiple processes like biodegradation, mixing, thermal cracking, among others. The V Vs. Ni diagram, shows most of the samples located at ratios between 1:1 and 100:1. The samples of the E-EC show lower content of V and Ni in comparison with the samples of the N-EC and W-EC (Figure 25 right). The younger samples of the N-EC and W-EC are among those with higher concentration of V and Ni (Figure 25 right).

The paleoceanographic evolution of the basin can be studied using the ratios between molybdenum and uranium. The information obtained can assess the basin restriction, bottom water redox conditions, and the process of metal-oxyhydroxide particulate shuttle - PS (Tribovillard et al., 2012). The particulate shuttle process enhances the Mo extraction from the water column, through adsorption of molybdate oxyanions over particles of Mn-oxides, creating a pattern of strong Mo enrichment of the sediments in relation to U, which is unaffected by this process (Tribovillard et al., 2012). When studying the relation of enrichment factors of Mo and U (EF-Mo and EF-U respectively), three main patterns have been identified (Figure 26):

- Unrestricted basins show progressive increase of the Mo:U ratios of 0.1 to 0.3xSW (Sea Water), reaching values greater than the actual sea water (>1XSW).
- Particulate Shuttle show strong enrichment of Mo relative to U, reaching Mo:U ratios between 3 to 10xSW.
- Restricted basin is characterized by decreasing Mo:U ratios with increasing EFs, representing the changes over depth of water chemistry, which requires an extended interval of deepwater isolation to allow chemical evolution of the water mass.

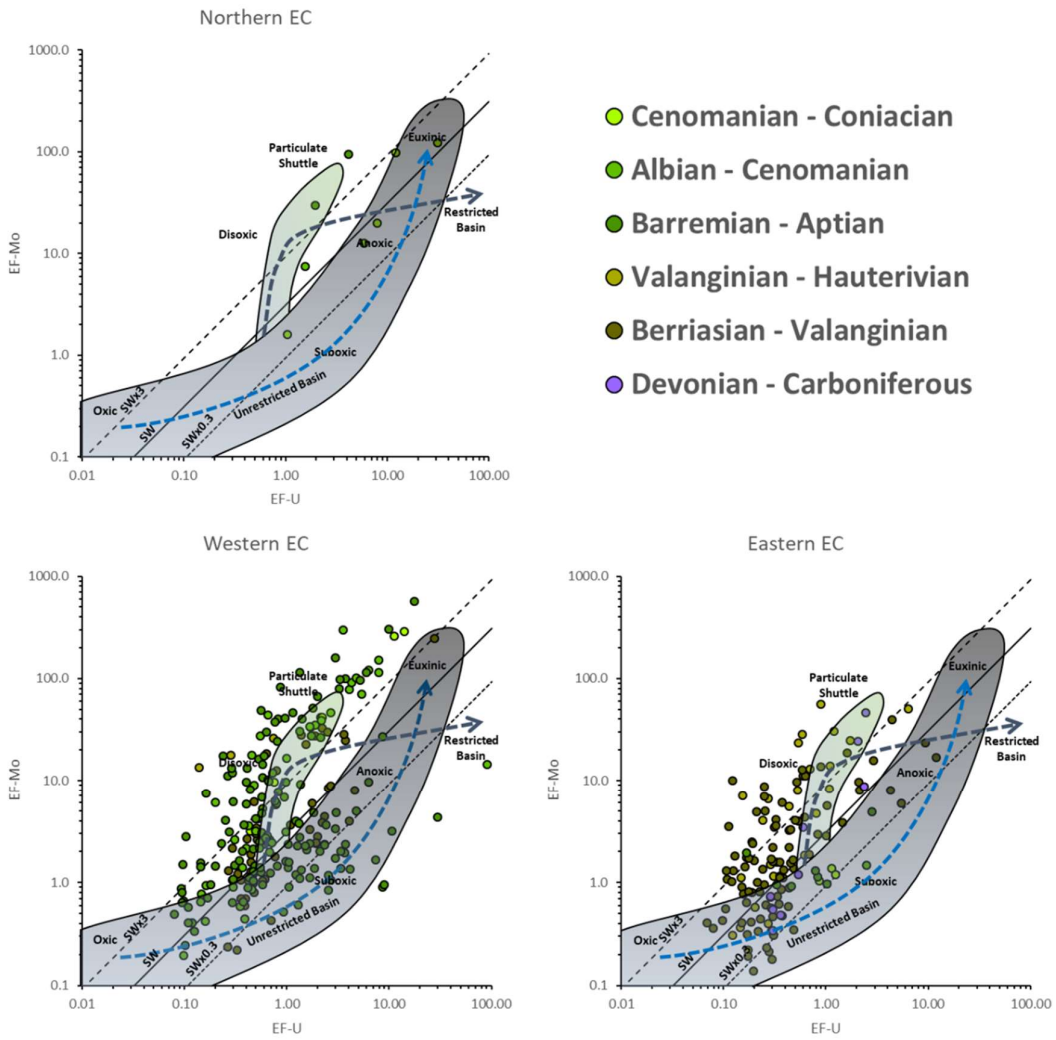

Figure 26. Diagrams of the Molybdenum and Uranium enrichment factors (EF-Mo Vs. EF-U). Basin patterns evolution are identified as Unrestricted Basin (blue arrow); Restricted Basin (gray arrow) and Particulate Shuttle (green polygon). Mo:U ratio of present-day seawater is represented with diagonal lines at multiples of 0.3, 1 and 3 (SWx0.3, SW and SWx3 respectively)

The N-EC presents Barremian-Aptian Samples over the PS field, but younger samples of the Albian-Coniacian show a pattern of unrestricted basin reaching euxinic conditions and Mo:U ratios between 0.3 to 3xSW (Figure 26). The W-EC shows a Berriasian-Aptian trend of unrestricted basin that reaches anoxic conditions; however, samples of the Barremian-Coniacian follow the PS pattern reaching euxinic conditions and Mo:U ratios higher than 3xSW (Figure 26). On the contrary, the E-EC shows older samples over the PS pattern reaching Mo:U ratios higher thas 3xSW, while younger samples are located over the unrestricted basin with suboxic conditions and Mo:U ratios close to 0.3xSW (Figure 26).

4.1.5. Data Mining of Elemental Geochemistry

Tool from data mining were used to find unknown correlations between variables, reduce the number of variables, and find groups with common characteristics and find the variability of the rocks for each stratigraphic interval. These groups could easily be described in their lithology, provenance, paleoclimatic and depositional conditions, in order to simplify the correlation with other datasets (organic geochemistry).

- Correlations

The correlation coefficients comparison was produced using the original data without any normalization. East and north coordinates were included to explore spatial correlation with any of the parameters obtained from the elemental analyses (Table 13). Normalization to PAAS does not improve correlation coefficients and not all the elements determined in the analyses, are available in the PAAS for normalization (Wedepohl, 1971).

Some of the variables show fewer correlations with any other one (32 variables out of 61, Table 13). Such variables include the geographical coordinates (East and North); the major oxides Fe2O3, MgO, Na2O, P2O5; the TOC and S; the minor elements As, Ba, Cd, Co, Cu, In, Li, Mo, Ni, Pb, Sb, Sr, U, Y, Zn; and the REE Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, and Lu.

Out of the 29 variables left, MnO and CaO show preferentially negative Spearman correlations coefficients (Table 13). The remaining oxides and elements show mostly positive Spearman correlations. It's important to note the lack of correlations from variables like TOC, S and Ni, which are some of the variables that applied for source rock characterization and rock to oil correlation. Also, the heavy REE show little correlation with most of the major oxides and minor elements, except with Y (Table 13).

The correlation coefficients on Table 13 correspond to the entire set of samples, however, these vary depending on the selected discrimination parameter, such as formation, stratigraphic age or basin. In Figure 27 and Figure 28 are show bivariate diagrams of CaO and SiO2 respectively, discriminated for each basin. These diagrams show the degree of dispersion of some of the variables, which difficult the direct interpretation of the data.

Table 13. Pearson and Spearman coefficients comparison. Green color are positive correlations and blue values are negative correlations. Black font are Pearson coefficient and red font are Spearman Coefficient. Cutoff value was 0.4, as an indicator of strong correlation.

	EAST	NORTH	SiO2	Al2O3	Fe2O3	MnO	MgO	CaO	Na2O	K2O	TiO2	P2O5	TOC	S	B	Be	Bi	Ge	Hf	In	Li	Sn	Ta	Tl	W	As	Ba	Cd	Co
EAST	0																												
NORTH	-0.6	0																											
SiO2	0	0	0																										
Al2O3	0	0	0.7	0																									
Fe2O3	0	0	0	0	0																								
MnO	0	0	-0.7	-0.6	0	0																							
MgO	0	0	-0.4	0	0	0.6	0																						
CaO	0	0.4	-0.9	-0.8	0	0.8	0	0																					
Na2O	0	0	0	0	0	0	0	0	0																				
K2O	0	0	0.5	0.7	0	-0.4	0	-0.6	0	0																			
TiO2	0	0	0.7	0.9	0	-0.6	0	-0.8	0	0.7	0																		
P2O5	0	0.4	0	0	0	0	0	0	0	0	0	0																	
TOC	0	0	0	0	0	0	0	0	0	0	0	0	0																
S	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0															
B	0	0	0.5	0.7	0	0	0	-0.5	0	0.6	0.7	0	0	0	0														
Be	0	0	0.6	0.7	0	-0.5	0	-0.6	0	0.8	0.7	0	0	0	0.6	0													
Bi	0	0	0.4	0.6	0	-0.4	0	-0.5	0	0.5	0.5	0	0	0	0.4	0.6	0												
Ge	0	0	0.7	0.8	0	-0.6	0	-0.8	0	0.5	0.7	0	0	0	0.6	0.6	0.5	0											
Hf	0	0	0.6	0.6	0	-0.4	0	-0.6	0	0.4	0.8	0	0	0	0.5	0.4	0	0.5	0										
In	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0.6	0	0	0	0									
Li	0	0	0	0	0	0	0	0	-0.7	0.5	0	0	0	0	0	0.4	0	0	0	0.5	0								
Sn	0	0	0.6	0.7	0	-0.4	0	-0.6	0	0.7	0.7	0	0	0	0.6	0.7	0.4	0.6	0.5	0.5	0	0							
Ta	0	0	0.7	0.8	0	-0.6	0	-0.7	0	0.7	0.9	0	0	0	0.8	0.8	0.6	0.7	0.7	0	0	0.7	0						
Tl	0	0	0	0.4	0	0	0	0	-0.4	0.7	0	0	0	0	0.4	0.7	0.4	0.4	0	0.6	0.6	0.5	0.5	0					
W	0	0	0.5	0.6	0	0	0	-0.5	0	0.5	0.6	0	0	0	0.7	0.6	0.5	0.5	0.5	0	0	0.6	0.8	0	0				
As	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0			
Ba	0	0	0.4	0.5	0	0	0	0	-0.4	0.8	0.4	0	0	0	0.5	0.7	0	0	0	0.5	0.6	0.6	0.5	0.8	0	0	0		
Cd	-0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.6	0	0	0	0	
Co	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 13. Continuation

	EAST	NORTH	SiO2	Al2O3	Fe2O3	MnO	MgO	CaO	Na2O	K2O	TiO2	P2O5	TOC	S	B	Be	Bi	Ge	Hf	In	Li	Sn	Ta	Tl	W	As	Ba	Cd	Co	Cr	Cs	Cu	Ga	Mo	Nb	Ni	Pb	Rb	Sb	Sc	Sr	Th	U	V	Y	Zn	Zr	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Cr	0	0	0.6	0.8	0	-0.7	0	-0.7	0	0.7	0.7	0	0.4	0	0.5	0.8	0.6		0.4	0.4	0	0.7	0.8	0.7	0.6	0	0.6	0	0	0																																					
Cs	0	0	0.5	0.7	0	-0.4	0	-0.6	0	0.9	0.6	0	0	0	0.6	0.8	0.5	0.6	0	0.6	0.6	0.7	0.7	0.9	0.5	0	0.8	0	0	0.8	0																																		
Cu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0.4	0	0	0	0																																		
Ga	0	0	0.7	0.9	0	-0.6	0	-0.8	0	0.8	0.9	0	0	0	0.7	0.8	0.6	0.8	0.6	0	0	0.7	0.9	0.5	0.7	0	0.5	0	0	0.8	0.7	0	0																																	
Mo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0.7	0	0.7	0.4	0	0	0.5	0.5	0	0	0																																	
Nb	0	0	0.7	0.9	0	-0.6	0	-0.8	0	0.6	0.9	0	0	0	0.7	0.7	0.6	0.7	0.8	0	0	0.7	0.9	0	0.7	0	0	0	0	0.7	0.6	0	0.9	0	0																																
Ni	0	0	0	0	0.4	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0.5	0	0.5	0	0	0.4	0	0	0.5	0	0.7	0	0																																
Pb	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0.4	0.5	0	0	0.5	0.5	0	0	0.7	0	0.5	0.5	0	0	0	0.6	0.5	0	0.5	0	0	0																																
Rb	0	0	0.6	0.7	0	-0.4	0	-0.6	0	1.0	0.7	0	0	0	0.6	0.8	0.5	0.6	0.4	0.6	0.5	0.7	0.7	0.8	0.5	0	0.8	0	0	0.8	0.9	0	0.8	0.4	0.7	0	0.5	0																																
Sb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0.6	0	0.8	0	0.5	0	0.5	0	0.6	0	0.8	0	0.6	0.6	0	0																																
Sc	0	0	0.4	0.7	0.4	0	0	-0.5	0	0.7	0.6	0	0	0	0.6	0.7	0.5	0.5	0.4	0.5	0	0.6	0.7	0.5	0.5	0	0.5	0	0	0.6	0.7	0	0.8	0	0.7	0	0	0.8	0	0																																
Sr	0	0	-0.4	-0.4	0	0	0	0.5	-0.6	0	-0.4	0	0	0	0	0	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.4	0	0	0	0	0	0																																
Th	0	0	0.5	0.7	0	0	0	-0.6	0	0.6	0.7	0	0	0	0.5	0.7	0.6	0.6	0.5	0	0	0.6	0.7	0	0.6	0	0	0	0	0.6	0.5	0	0.8	0	0.7	0	0	0.6	0	0.6	0	0																																
U	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0.4	0	0	0	0.5	0	0	0.6	0	0.4	0	0	0.6	0	0	0	0																																
V	0	0	0.5	0.6	0	-0.5	0	-0.5	0	0.7	0.5	0	0	0	0.6	0.7	0.4	0.6	0	0.5	0.4	0.6	0.6	0.8	0.5	0	0.7	0	0	0.8	0.7	0	0.6	0.7	0.5	0.5	0.4	0.7	0.6	0.6	0	0.4	0.5	0																																
Y	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																
Zn	-0.5	0	0	0	0	0	0	0	-0.4	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0.5	0	0	0	0.9	0	0	0	0.4	0	0	0	0	0.4	0	0.5	0	0	0	0.5	0	0	0																																
Zr	0	0	0.7	0.7	0	-0.5	0	-0.7	0	0.5	0.9	0	0	0	0.5	0.5	0	0.5	0.9	0	0	0.6	0.7	0	0.5	0	0	0	0	0.5	0.4	0	0.7	0	0.8	0	0	0.5	0	0.5	0	0.5	0	0	0	0	0																																
La	0	0	0	0.5	0	0	0	-0.4	0	0.4	0.5	0	0	0	0	0.5	0.4	0.4	0	0	0	0.4	0.5	0	0.4	0	0	0	0	0.4	0.4	0	0.5	0	0.5	0	0	0.4	0	0.5	0	0.5	0	0	0	0	0.4	0																																
Ce	0	0	0.4	0.6	0	0	0	-0.5	0	0.4	0.6	0	0	0	0.4	0.5	0.4	0.4	0.4	0	0	0.5	0.6	0	0.4	0	0	0	0	0.5	0.5	0	0.6	0	0.5	0	0	0.4	0	0.5	0	0.5	0	0	0	0	0.5	1.0	0																																
Pr	0	0	0.4	0.6	0	0	0	-0.5	0	0.4	0.6	0	0	0	0.4	0.5	0.4	0.4	0.4	0	0	0.5	0.6	0	0.4	0	0	0	0	0.5	0.5	0	0.6	0	0.6	0	0	0.5	0	0.5	0	0.5	0	0	0	0	0.5	1.0	1.0	0																																
Nd	0	0	0	0.5	0	0	0	-0.5	0	0.4	0.6	0	0	0	0.4	0.5	0.4	0.4	0.4	0	0	0.5	0.6	0	0.4	0	0	0	0	0.5	0.5	0	0.6	0	0.5	0	0	0.5	0	0.5	0	0.5	0	0	0	0	0.4	1.0	1.0	1.0	0																																
Sm	0	0	0	0.5	0	0	0	0	0	0	0.5	0	0	0	0	0.4	0	0	0	0	0	0.4	0.5	0	0	0	0	0	0	0.4	0.4	0	0.5	0	0.5	0	0	0.4	0	0.5	0	0.5	0	0	0.4	0	0	0.9	1.0	1.0	1.0	0																																
Eu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.7	0.7	0.8	0.9	0																																
Gd	0	0	0	0.4	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0.5	0	0.5	0	0	0	0	0.5	0	0.5	0	0	0.6	0	0	0.7	0.8	0.8	0.8	0.9	0.9	0																																
Tb	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0.4	0	0.5	0	0	0	0	0.5	0	0.4	0	0	0.8	0	0	0.5	0.5	0.6	0.6	0.8	0.7	0.9	0																																
Dy	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0.4	0	0.4	0	0	0.8	0	0	0	0.4	0.5	0.5	0.6	0.6	0.9	1.0	0																																
Ho	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0.4	0	0.4	0	0	0.9	0	0	0	0	0	0	0.4	0.5	0.5	0.8	0.9	1.0	0																																
Er	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0.4	0.4	0.7	0.8	1.0	1.0	0																																
Tm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0.4	0.7	0.8	1.0	1.0	1.0	0																																
Yb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0.4	0.4	0.7	0.8	0.9	1.0	1.0	0	0																																
Lu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0.4	0.4	0.6	0.8	0.9	0.9	0	1.0	1.0	0																																

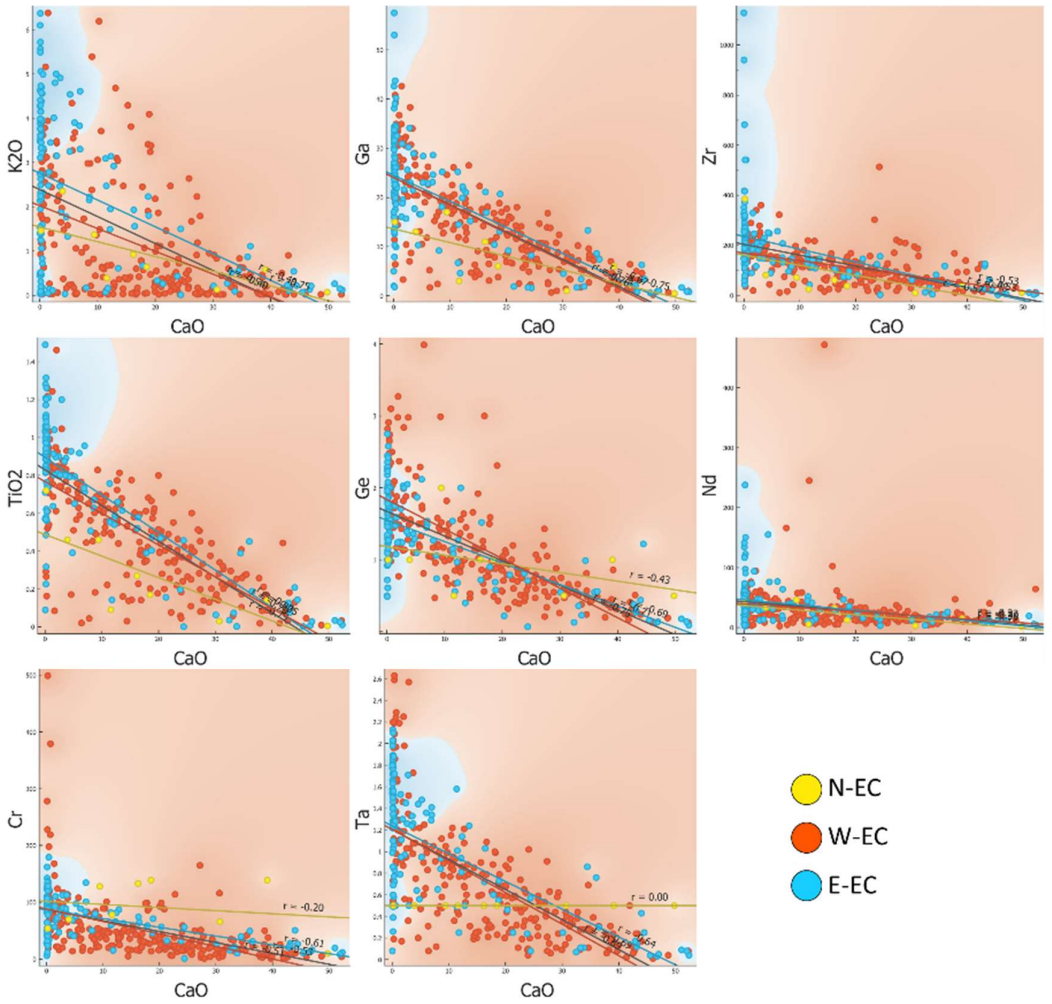

Figure 27. Bivariate diagrams showing the correlations of CaO with some major oxides (K2O, TiO2), minor elements (Cr, Ga, Ge, Ta, Zr), and REE (Nd).

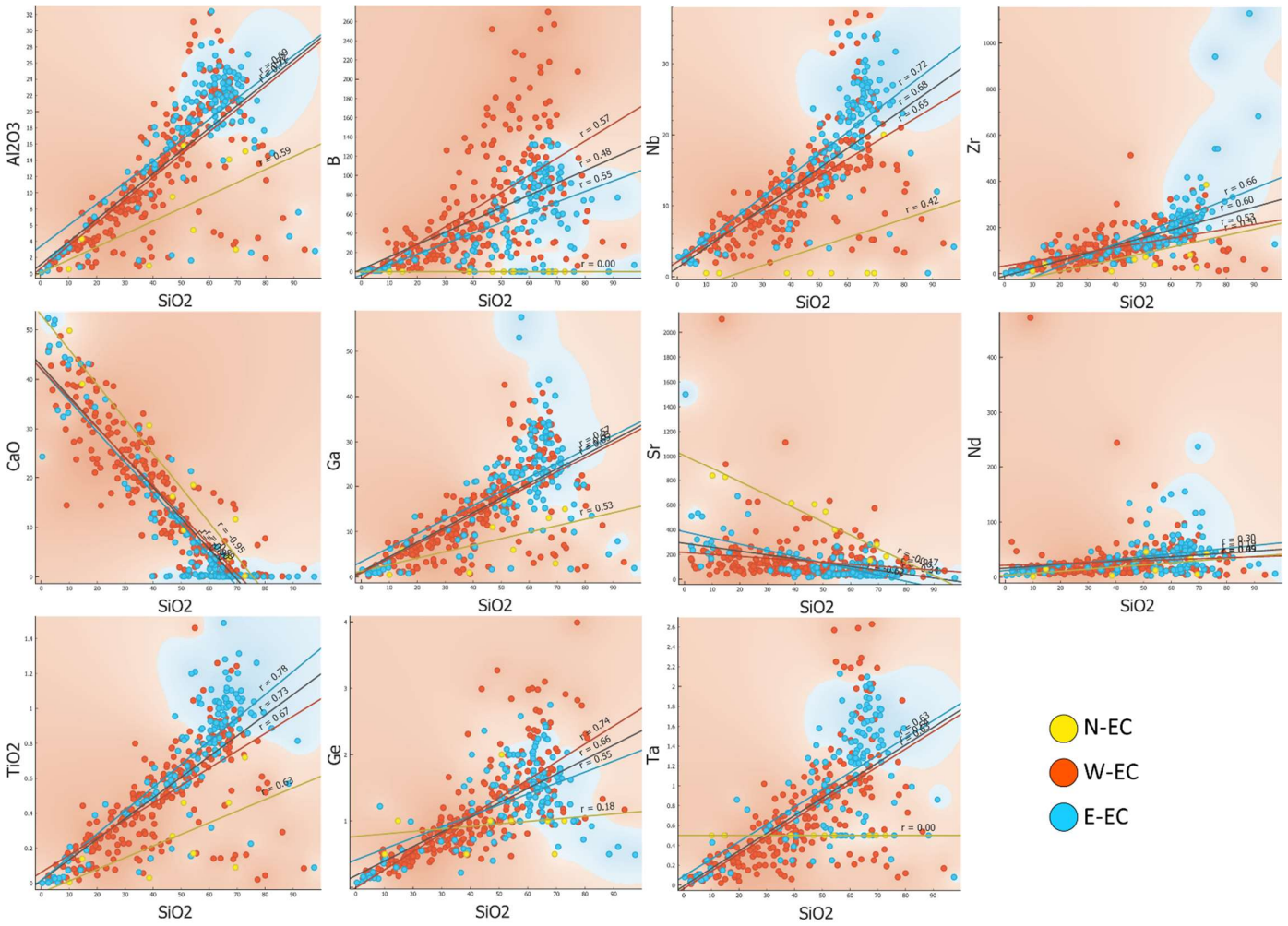

Figure 28 Bivariate diagrams showing the correlations of CaO with some major oxides (Al2O3, CaO, TiO2), minor elements (B, Ga, Nb, Sr, Ta, Zr), and REE (Nd).

- Sensitivity Test

It was performed with different combinations using original and PAAS normalized data, removal of outliers (14 out of 387), and ranking (Table 14). The number of PCA was set to 5 in all cases for comparison, and number of clusters (K-Means) was between 2 and 5. Also the effect of the order of the processes was included, whether PCA first and the K-means clustering later, or vice versa. The use of PAAS normalized data, removal of outliers and ranking, improved the percent of variance explained during the PCA, and the silhouette scores of the K-means clusters. These was an unexpected result, since the correlations do not vary between original or PAAS normalized data.

Table 14. Results of the sensitivity analysis which included original and PAAS normalized data, removal of outliers, and ranking of variables using ANOVA with 30 best ranked; and the results of the explained variance and silhouette scores of K-means clustering. Green color for higher values and red color for lower values

TEST	PCA VARIANCE	KMEANS-PCA	PCA-KMEANS	N-K-PCA	N-PCA-K
ORIGINAL	0.583	0.233	0.231	5	3
ORIGINAL INLIERS	0.586	0.231	0.223	5	3
ORIGINAL INLIERS ANOVA 30	0.63	0.257	0.254	3	3
PAAS	0.625	0.228	0.228	4	4
PAAS INLIERS	0.616	0.211	0.215	4	3
PAAS RANK ANOVA 30	0.627	0.238	0.238	3	3
PAAS INLIERS ANOVA 30	0.628	0.235	0.235	3	3

The method of variables ranking and the number of selected variables was explored too (Table 15). The scenarios used included three methods for ranking variables (ANOVA, X^2, and Gain Ratio), and different number of variables selected (10 to 50). The original data included 61 variables and the PAAS normalized data included 57 variables.

The results indicate that the ANOVA method produces the best results of explained variance (PCA) and silhouette scores (K-means). Additionally, PAAS normalized data improve the explained variance at some values of selected variables during ranking, as well as for the silhouette score (Table 15 and Figure 29). Nevertheless, reducing the number of variables can over-simplify the characteristics of the PCA and K-means clustering, and produce an unsupported interpretation. The correlations analysis showed that nearly half of the variables did not have strong correlations, leaving approximately 20 to 30 variables with strong correlations to each other (Table 13). From the PCA and K-means statistics results (Figure 29), it can be seen that using whether original or PAAS normalized data and removing outliers, combined with top

20 to 30 variables ranked using ANOVA, generates good results with 3 to 4 clusters, while being consistent with the correlations result. The PCA explained variance can be improved by increasing the number of components, without affecting the clustering results.

Table 15. Results of the sensitivity analysis using different ranking methods and increasing number of selected variables (N-Variables), and the results of the explained variance and silhouette scores of K-means clustering. Green color for higher values and red color for lower values

TEST	N-VARIABLES	PCA VARIANCE	KMEANS-PCA	PCA-KMEANS	N-K-PCA	N-PCA-K
ORIGINAL INLIERS ANOVA	10	0.832	0.29	0.29	3	3
	15	0.688	0.232	0.228	3	3
	20	0.686	0.257	0.256	3	3
	30	0.63	0.257	0.254	3	3
	40	0.613	0.245	0.241	4	3
	50	0.595	0.243	0.242	4	3
PAAS INLIERS ANOVA	10	0.747	0.303	0.325	3	5
	15	0.694	0.267	0.267	3	3
	20	0.657	0.249	0.249	3	3
	30	0.617	0.237	0.231	5	3
	40	0.631	0.234	0.24	4	3
	50	0.597	0.216	0.214	4	3
ORIGINAL INLIERS GAIN RATIO	10	0.799	0.372	0.379	4	3
	15	0.619	0.254	0.245	4	4
	20	0.549	0.22	0.215	4	4
	30	0.543	0.212	0.209	4	4
	40	0.568	0.227	0.226	3	3
	50	0.546	0.224	0.224	3	3
ORIGINAL INLIERS CHI2	10	0.778	0.366	0.369	4	3
	15	0.644	0.241	0.236	4	4
	20	0.598	0.22	0.216	4	3
	30	0.57	0.21	0.218	3	3
	40	0.589	0.228	0.233	3	3
	50	0.571	0.234	0.24	4	3

These observations allowed to reduce the number of initial variables and to add new calculated variables used during the interpretation of provenance, tectonic setting, paleoclimatic and paleoredox conditions. These variables were combined with original and PAAS normalized data for the sensitivity analysis (Figure 30). The results show that the ANOVA ranking method yield the best effects on the explained variance from PCA and K-means silhouette scores. These parameters have a sharp increase when less than 20 variables are selected from the ranking, which are about the 25% of the total 78 variables included in these analyses. Even though selecting less variables has the potential to increase the quality of the statistical results, the analysis will be made using between 20 and 30 variables so that the variability of data is accounted.

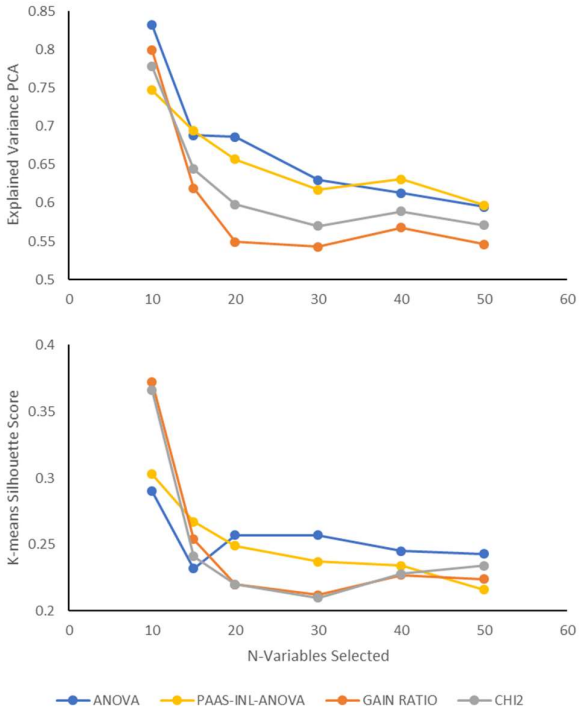

Figure 29. Results of the effect of the number of variables selected and ranking method on the explained variance (PCA) and K-means silhouette scores for Original and PAAS normalized data. The total variables were 61, and 57 for PAAS normalized data.

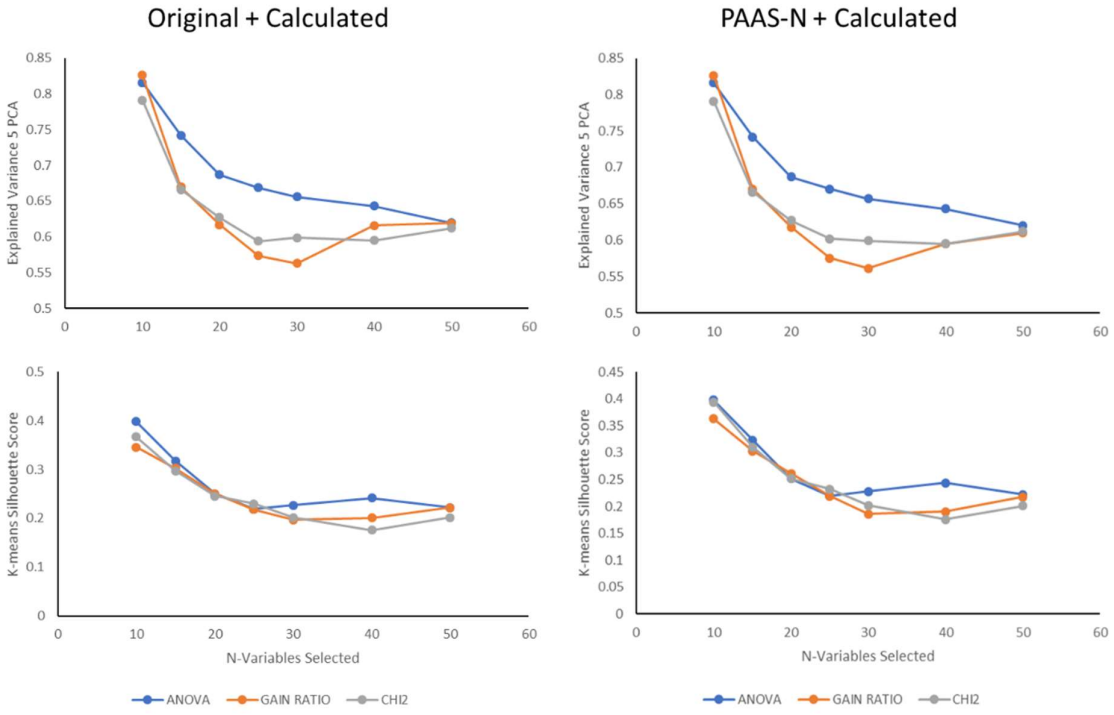

Figure 30. Results of the effect of the number of variables selected and ranking method on the explained variance (PCA) and K-means silhouette scores for Original and PAAS normalized data, together with calculated parameters. The total variables were 78.

A comparison of the ANOVA ranking using different data sets show that, at higher variables selected, the original data can yield better results, but, at about 20 variables, where all the data sets have similar behavior, the data sets with added parameters (ANOVA+), show better statistics for the clustering and slightly better explained variance from PCA (Figure 31).

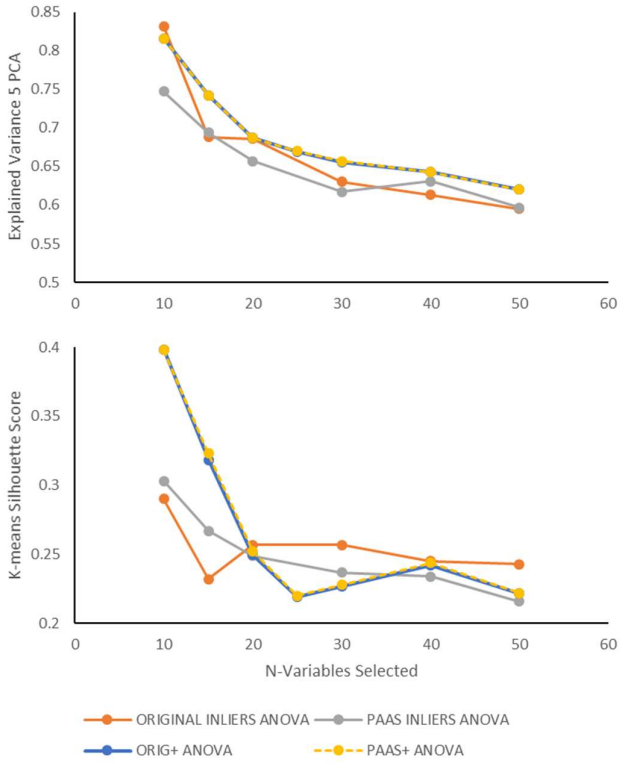

Figure 31. Comparison of the ANOVA ranking methods with different data sets, including the calculated parameters (ANOVA+) to the selected original and PAAS normalized data from previous runs.

- Principal Component Analysis

Multiple runs of the PCA where made, using different data sets, ranking methods and number of variables selected, as shown in the previous section. The PCA was run using the best 30 variables ranked with ANOVA method and 5 components that explain 61% to 66% of variance. Additional components can increase the explained variance, but with lower correlations (Spearman Coefficient < 0.3). The correlations of the PC and the respective variables show correlations with similar variables for PC1 and PC3, followed by PC5. On the other hand, PC2 and PC4 showed correlations similar

between the two main groups of data set, those with original and PAAS normalized data set (1 and 2 in Table 16); and those with added calculated variables (3 and 4 in Table 16).

Table 16. Comparison of correlations between principal components and variables for different data sets. Ranking was set to top 30 variables using ANOVA, and PCA was set to 5 components. Removal of inliers was included in all the runs. Runs 1 and 2 included original and PAAS normalized data sets. Runs 3 and 4 included best variables from runs 1 and 2, and added calculated variables. Negative correlations in blue and positive correlations in green. Underlined variables have 2 appearances within the same component. Bold variables have 3-4 appearances within the same component.

PC	1- ORIG-INLIERS-ANOVA30		2- PAAS-INLIERS-ANOVA30		3- ORIG+ -INLIERS-ANOVA30		4- PAAS+ -INLIERS-ANOVA30		INDICATOR
	VARIABLE	SPEARMAN	VARIABLE	SPEARMAN	VARIABLE	SPEARMAN	VARIABLE	SPEARMAN	
PC1	**TiO2**	0.903	**TiO2_PAAS**	0.89	Log(CaO/Al2O3)	0.892	Log(CaO/Al2O3)	0.894	Lithology: SiO2, Log(CaO/Al2O3), Log(Fe2O3/K2O) Provenance: Sc, Th, TiO2, Zr, Hf Paleoclimatic: CIW, Rb Paleoredox: U-EF, Ni-EF, U/(U+Th)
	Nb	0.878	**Nb_PAAS**	0.86	**Nb**	-0.89	**Nb_PAAS**	-0.891	
	CaO	-0.868	**Cr_PAAS**	0.856	CIW	-0.847	CIW	-0.849	
	Cr	0.85	**Rb_PAAS**	0.821	**Zr**	-0.777	**Zr_PAAS**	-0.777	
	SiO2	0.825	**K2O_PAAS**	0.808	U_EF	0.738	U_EF	0.735	
	Rb	0.816	**SiO2_PAAS**	0.787	**SiO2**	-0.724	**SiO2_PAAS**	-0.726	
	K2O	0.797	**Cs_PAAS**	0.784	Ni_EF	0.714	Ni_EF	0.713	
	Ge	0.794	**Zr_PAAS**	0.754	**Rb**	-0.694	**Rb_PAAS**	-0.698	
	Cs	0.783	**Sc_PAAS**	0.722	U/(U+Th)	0.671	U/(U+Th)	0.668	
	Zr	0.78			Log(Fe2O3/K2O)	0.617	Log(Fe2O3/K2O)	0.62	
	Hf	0.709			**Cs**	-0.609	**Cs_PAAS**	-0.612	
	Sc	0.708							
PC2	Sr	0.633	Mo_PAAS	0.64	Cs	0.811	Rb_EF	0.81	Lithology: Na2O, Log(Fe2O3/K2O) Provenance: Hf, Zr, Paleoclimatic: Al/Na, Rb Paleoredox: U, V, Ni/(Ni+Co)
	Cd	0.589	Ce/Ce*_PAAS	-0.625	Rb_EF	0.81	Cs_PAAS	0.809	
	Na2O	-0.571	Sb_PAAS	0.574	V	0.803	V_PAAS	0.799	
	Tl	0.562	Cd_PAAS	0.574	Al/Na	0.759	Al/Na	0.76	
	Hf	-0.559	U_PAAS	0.529	Rb	0.751	Rb_PAAS	0.749	
	U	0.545			Log(Fe2O3/K2O)	-0.746	Log(Fe2O3/K2O)	-0.743	
	Zr	-0.512			Ni/(Ni+Co)	0.653	Ni/(Ni+Co)	0.654	
	Sb	0.502							
PC3	Na2O	-0.543	ErN_PAAS	0.752	**LONG**	0.696	**LONG**	0.7	Provenance: Zr Paleoclimatic: K2O, Na2O, Rb, Sr Paleoredox: V/(V+Cr), TOC/(TOC+S) Geographical: Latitude, Longitude
	In	0.383	TmN_PAAS	0.746	TOC/(TOC+S)	-0.656	C/(C+S)	-0.656	
	Silhouette	0.367	**Y_PAAS**	0.722	**LAT**	-0.533	**LAT**	-0.532	
	Sc	0.352	LaN/SmN_PAAS	-0.478	V/(V+Cr)	-0.515	V/(V+Cr)	-0.52	
	K2O	0.351	**LONG**	0.359	**Y**	0.483	**Y_PAAS**	0.482	
	LONG	-0.344	**LAT**	-0.324					
	Rb	0.344	Na2O_PAAS	-0.321					
	Cs	0.343							
	MnO	0.312							
	Sr	0.301							
PC4	LAT	0.693	Fe2O3_PAAS	0.643	Log(SiO2/Al2O3)	0.553	Log(SiO2/Al2O3)	0.55	Lithology: SiO2, Log(SiO2/Al2O3) Provenance: Ti/Zr, Paleoclimatic: K2O, Rb
	MgO	-0.521	MgO_PAAS	0.537	Ti/Zr	-0.446	Ti/Zr	-0.442	
	Cd	-0.486	Sc_PAAS	0.461	SiO2	0.432	SiO2_PAAS	0.431	
	B	-0.479	Rb_PAAS	0.424	MnO	-0.402	MnO_PAAS	-0.402	
	Y	0.47	K2O_PAAS	0.422					
PC5	**Sr**	-0.521	Na2O_PAAS	0.719	CaO*	0.591	CaO*	0.594	Lithology: CaO* Provenance: Ti/Zr, Paleoclimatic: Na2O,Sr, Rb, Al/Na
	LONG	-0.438	**Sr_PAAS**	-0.657	Al/Na	-0.353	Al/Na	-0.358	
	PC2	-0.434	PC2	-0.386	Ti/Zr	0.346	Ti/Zr	0.353	
	Y	0.367	Cs_PAAS	-0.365	TOC	0.341	Rb_EF	-0.338	
	Na2O	0.324	Pb_PAAS	-0.331	Rb_EF	-0.329	**Sr_PAAS**	-0.333	
			Fe2O3_PAAS	0.306	**Sr**	-0.327	C_PAAS	0.312	

Each of the principal components have multiple variables associated, with relation to different aspects of the rocks characterization (Table 16). The most common category is the lithology (PC1-PC2-PC4-PC5), with parameters regarding its composition, however, a variable as CaO, associated with carbonates, was not included in the top 30 variables, except for run 1 (Table 16). Instead, the geographical coordinates made into those 30 variables for every run, and have representative correlations in PC3 (Table 16).

- Cluster Analysis

Cluster analysis was implemented using the K-means algorithm, with best cluster numbers (K) selected according to its silhouette score. For most of the tests, the number of clusters was 3, with about 35% of the results with 4-5 clusters. However, additional cluster did not show meaning separation and only less than 2% of the samples were included in those additional clusters (Figure 32).

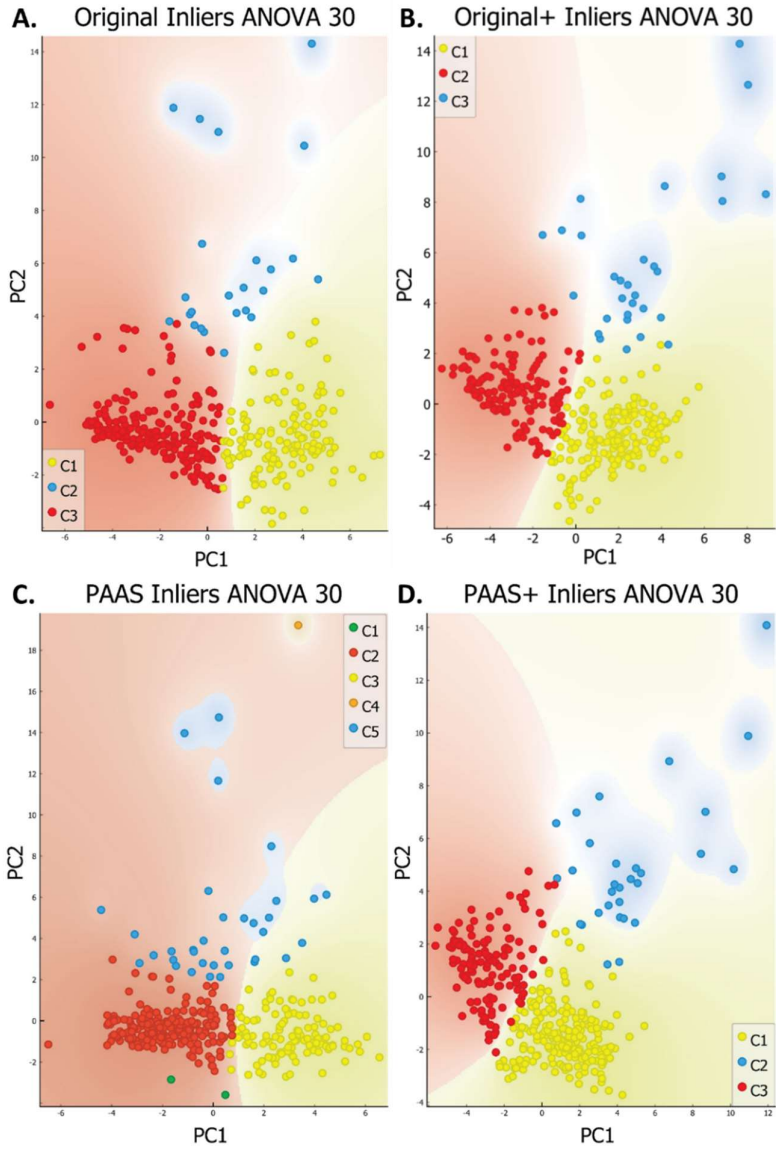

Figure 32. Comparison of the clusters obtained for each of the runs shown in Table 16and Figure 31.

The obtained clusters show good separation with the principal components calculated, with three clusters being the most prevalent number (Figure 32). Additional cluster of

run C shown in Figure 32, only represented 3 samples, out of 372; however, those samples were included in other clusters with different data sets (Figure 32). Despite the differences between part of the variables in run B and D, which included some original or PAAS normalized data respectively, their clusters and principal components are mostly the same, except for 1 sample deleted from the diagram 7 in Figure 32.

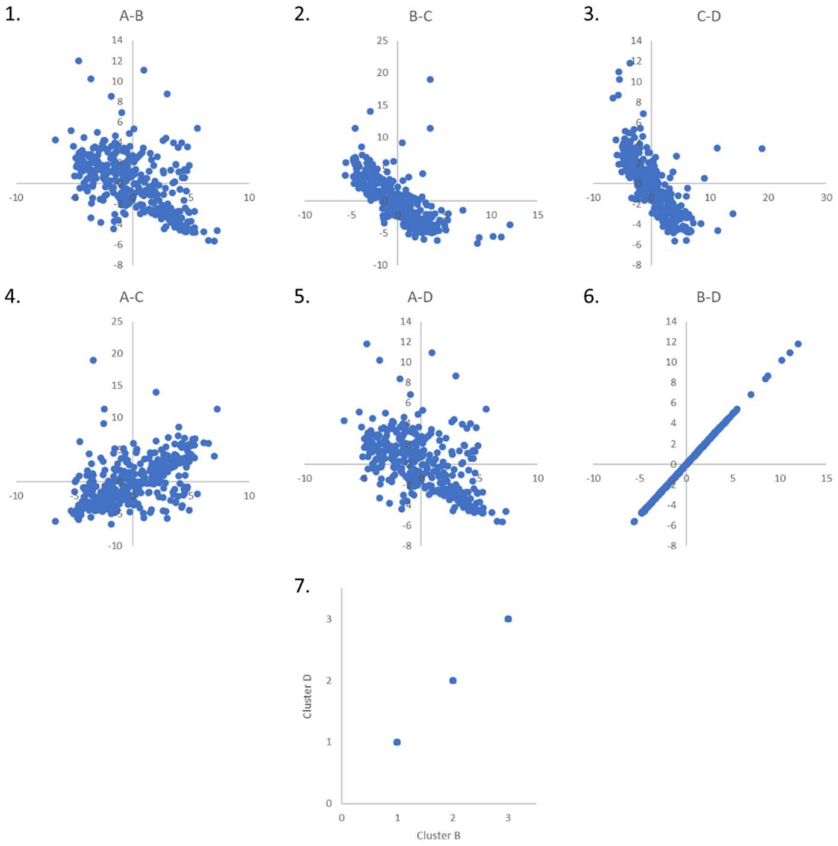

Figure 33. 1-6: Comparison of the first principal component (PC1) for the runs shown in Figure 32. 7: assigned cluster between run B and D; numbers correspond to clusters in Figure 32.

The quality of the clusters obtained with the K-means algorithm can be addressed using the Silhouette values. The higher Silhouette values indicate that the cohesion or similarity of the samples within a cluster is higher. Also, the higher values of this parameter indicate a better distinction between different clusters (Figure 34). The cluster C3 shows lower values of Silhouette, indicating that it has lower intra-cluster cohesion and lower inter-cluster distinction, as can be seen from overlapping points of C2 and C1 (Figure 34). When comparing Silhouette with PC2, a better distinction of the C3 is clear, explaining the separation of this cluster form C1 and C2 (Figure 35).

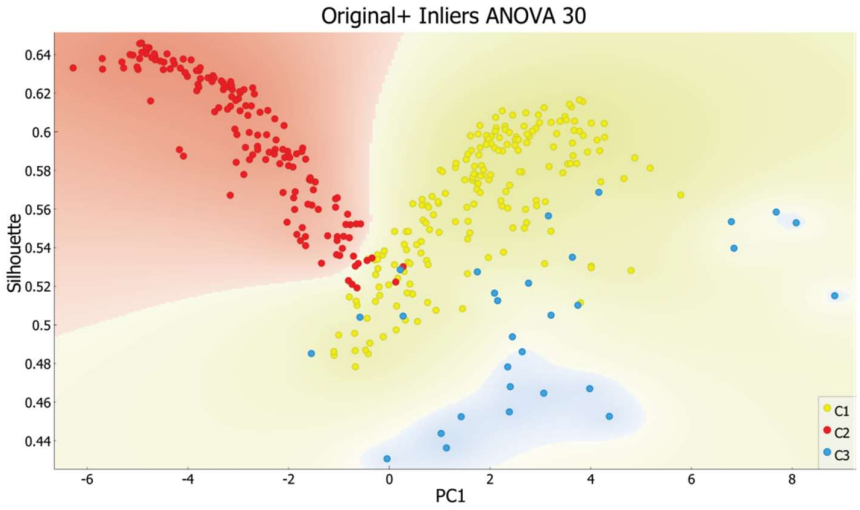

Figure 34. Diagram of Silhouette Vs. PC1 showing the quality of the clusters. Higher Silhouette values indicate better intra-cluster similarity and inter-cluster distinction.

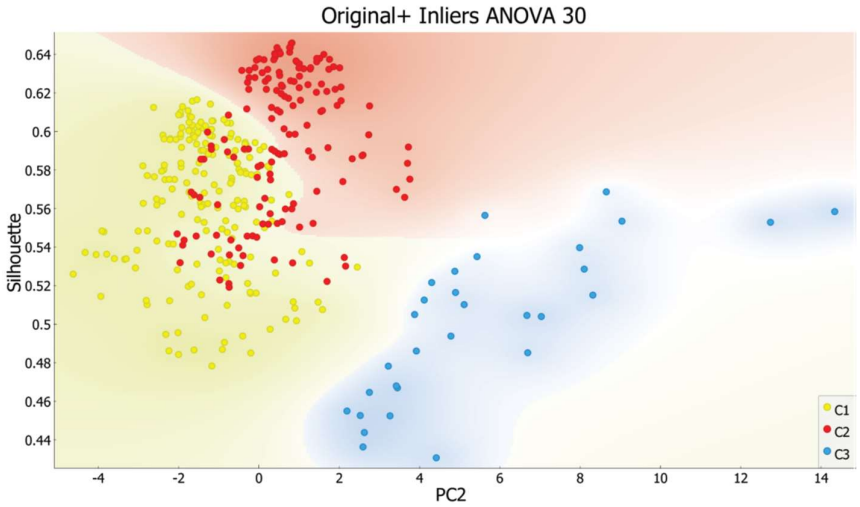

Figure 35. Diagram of Silhouette Vs. PC2 showing the quality of the clusters. Higher Silhouette values indicate better intra-cluster similarity and inter-cluster distinction.

As previously seen in the interpretation of the data, each stratigraphic horizons exhibit differences due to varying conditions at each geographic location, and within each horizon. This can be seen from the distribution of the clusters across the regions of the study and for each stratigraphic interval (Figure 36). These distribution shows preferences for samples of each cluster located at specific region and/or stratigraphic interval. Samples from cluster 1 (C1), are preferentially located at the WEC on the stratigraphic intervals of the Berriasian-Valanginian and Barremian-Aptian. On the EEC they appear within older stratigraphic intervals, and are distributed in similar proportion to those of the C2. The samples of C2 are at every stratigraphic interval of

the EEC, and appear at WEC in intervals younger than Valanginian-Hauterivian. Samples of the C3 are mostly present at the NEC and WEC, at intervals younger than Valanginian-Hauterivian (Figure 36). From the geographic distribution, it is also clear that the properties of the rocks vary across the basin, being the processes distinct at each location (Figure 36).

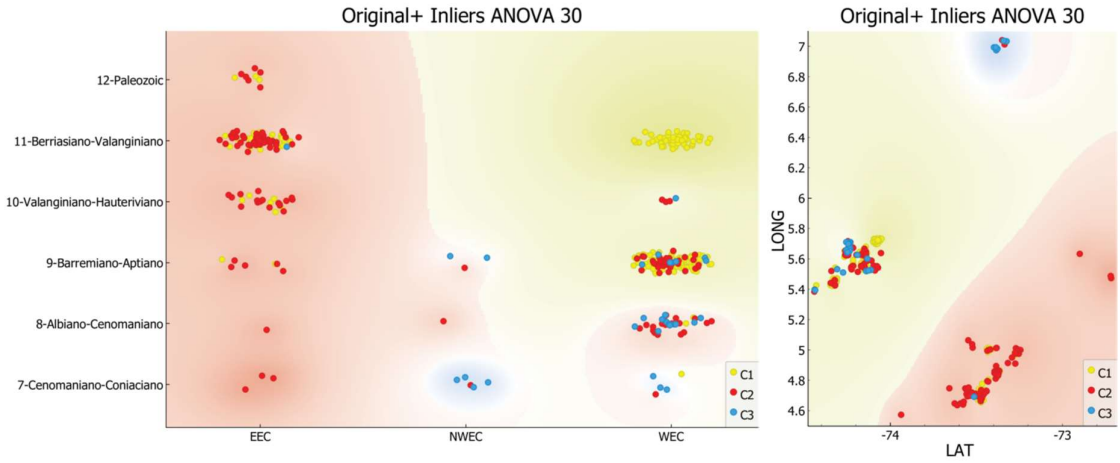

Figure 36. Left: Clusters of samples classified for each stratigraphic horizon and region studied. Right: Geographical distribution of the clusters across the study area.

Each of these clusters show characteristics with respect to the defining parameters of lithology, provenance, paleoclimatic and paleoredox conditions (Table 17 and Figure 37).

Table 17. Main characteristic of the clusters of samples based on run B (Figure 32). Indicative diagrams shown in Figure 37.

Characteristic	Cluster-1	Cluster-2	Cluster-3
Lithology	Muddy Marlstone to Calcareous Marlstone	Mudstone to Muddy Marlstone	Mudstone to Sublithic Sandstone
Provenance	Felsic	Felsic	Felsic
Paleoclimate	Non-Weathering Arid	Weak to Strong Weathering Arid	Strong Weathering Humid
Paleoredox	Disoxic to Anoxic	Suboxic to Disoxic	Particulate Shuttle

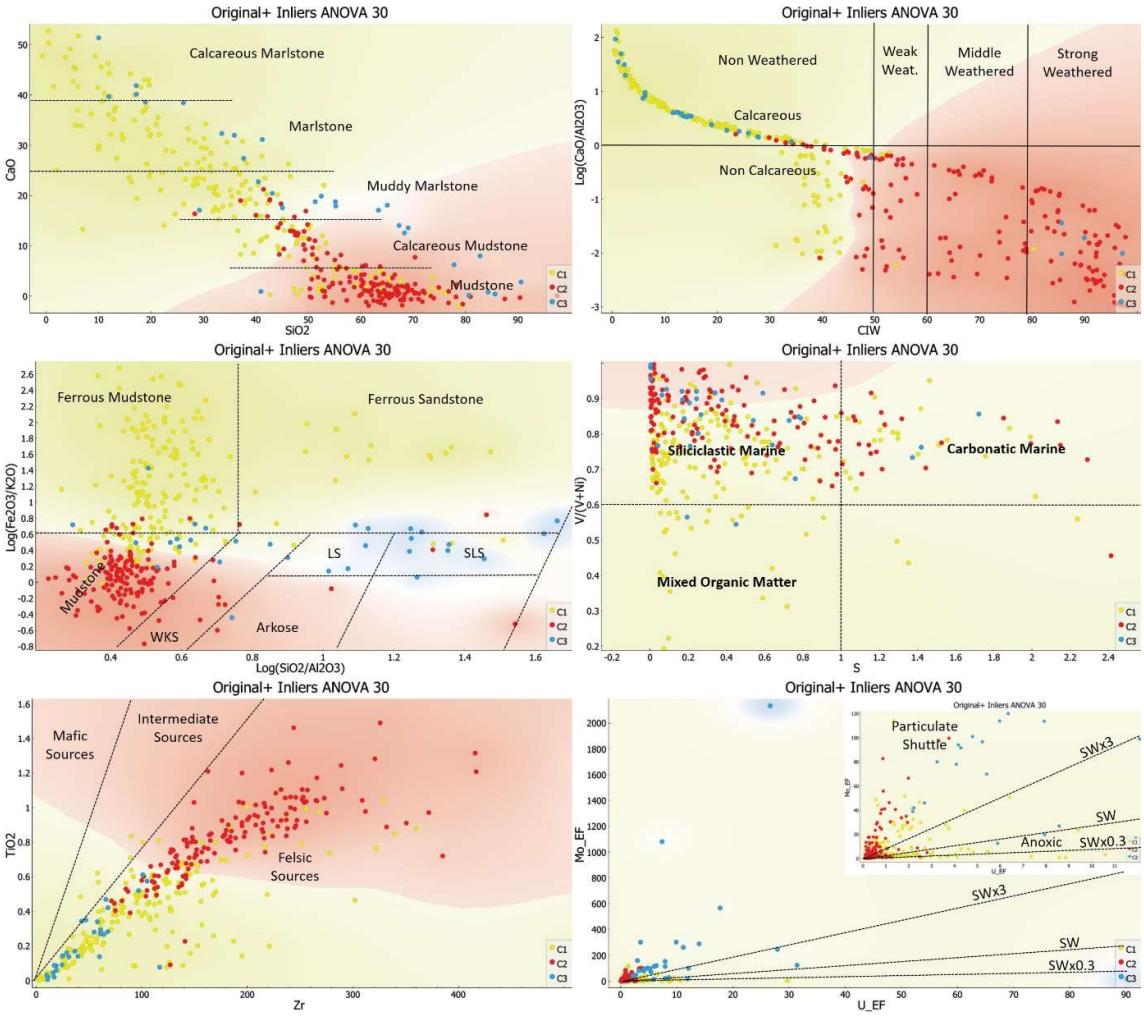

Figure 37. Bivariate diagrams that show multiple parameters of the samples and its relation to the clusters built

4.2. **Isotopic Analysis**

Isotope analyzes of Pb, Sr, and Nd were performed in the whole rock of 23 selected shale samples of the MMV and Eastern Cordillera basins, of 3 pyrites like vein filling from Paja, Macanal and Paleozoic formations (Figure 38), as well as of Rb in 5 samples of shales the La Luna formation of the MMV basin. Whole rock samples were also analyzed for multielemental geochemistry, additionally, the analysis of 1 crude of the same was performed. basin for Sr from the Mugrosa formation.

Figure 38. Pyrites analyzed in this project arranged in veinlets embedded in A) Paleozoic rocks, B) Paja Formation and C) Macanal formation.

4.2.1. Pb Isotopes

The highest values in the content of U, Th and Pb in the Middle Magdalena Valley (MMV) correspond to the Paja formation, with values between 5.74 to 22.03 ppm for U, values between 4.2 to 21 ppm for Th, and values between < 20 to 23 ppm for Pb. The isotopic ratios for this formation are 20.588 to 20.423 for the $^{206}Pb/^{204}Pb$, 15.844 to 15.847 for $^{207}Pb/^{204}Pb$,40.012 to 40.109 for $^{208}Pb/^{204}Pb$.

In the Luna formation of MMV the values in the content of U, Th and Pb vary between 1.39 to 5.84 ppm for U, 1 to 9 for Th and < 20 to 26 ppm for Pb. The isotopic ratios of $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{208}Pb/^{204}Pb$ in the MMV in the Luna formation varying from 20.822 to 21.259 for $^{206}Pb/^{204}Pb$, 15.831 to 15.926 for $^{207}Pb/^{204}Pb$, and 39.325 to 40.359 for $^{208}Pb/^{204}Pb$.respectively.

The content of U in the shales of the Eastern Cordillera have higher values for the Une formation (7.16 ppm), while for Th the highest content is reported for the Chipaque

formation (28.9 ppm to 22.2 ppm) and the Macanal formation for the Pb (69 ppm to < 20).

In the Eastern Cordillera, the isotopic ratios of $^{206}Pb/^{204}Pb$ are the highest for the rocks of the Paleozoic, with values that vary between 21.634 to 17.688. In the case of the isotopic ratios of $^{207}Pb/^{204}Pb$, the maximum values are from the Chipaque formation and vary between 15.935 to 15.904, while for the $^{208}Pb/^{204}Pb$ the maximum values are from the rocks of the Paleozoic (45.346 to 37.663) and Chipaque formation (40.635 to 40.183) as seen in the Table 18.

Table 18. Pb isotopic results for the samples from the MMV and Eastern Cordillera basins.

Basin	Formation	Sample	U (ppm)	Th (ppm)	Pb (ppm)	$^{206}Pb/^{204}Pb$	2SD	$^{207}Pb/^{204}Pb$	2SD	$^{208}Pb/^{204}Pb$	2SD
MMV	La Luna	YP43A	4.56	2	<20	21.529	0.010	15.832	0.008	39.370	0.017
	La Luna	YP44A	5.34	5.8	<20	21.163	0.004	15.895	0.004	40.000	0.009
	La Luna	YP45B	5.84	1.1	<20	21.385	0.012	15.854	0.009	39.325	0.026
	La Luna	YP46A	1.39	1	<20	21.079	0.011	15.851	0.010	39.599	0.026
	La Luna	YP47B	2.77	9	26	20.822	0.004	15.927	0.003	40.360	0.010
	Tablazo	YP19B	3.59	14.7	<20	20.209	0.005	15.832	0.004	39.984	0.013
	Paja	YP14B	7.57	13.1	<20	20.589	0.004	15.845	0.003	40.013	0.008
	Paja	YP14A	5.74	21	23	20.423	0.007	15.848	0.007	40.109	0.019
	Paja	YP14C				20.333	0.012	15.837	0.010	40.226	0.026
Easter Cordillera	Une	JC12 C	7.16	23.4	<20	21.083	0.005	15.906	0.004	40.427	0.008
	Fómeque	JC7 B	4.54	19.8	<20	19.671	0.006	15.736	0.005	39.708	0.016
	Fómeque	JC15 D	3.12	17.3	<20	19.354	0.006	15.729	0.006	39.539	0.019
	Macanal	YP28A	3.85	19.3	<20	19.372	0.005	15.711	0.004	39.484	0.011
	Paleozóico	YP27A	1.95	15.3	<20	21.635	0.036	15.851	0.027	45.347	0.076
	Chipaque	YP54-COMP	3.54	22.2	27	20.821	0.005	15.904	0.004	40.563	0.012
	Chipaque	JC5A	4.43	28.1	32	20.632	0.006	15.905	0.006	40.184	0.018
	Chipaque	YP58-COMP	3.54	28.9	28	20.987	0.005	15.935	0.004	40.636	0.010
	Une	JC13 A	22.03	4.2	<20	20.278	0.009	15.838	0.010	40.099	0.031
	Fómeque	JC7 E	2.43	7.3	<20	21.533	0.004	15.891	0.003	39.877	0.008
	Fómeque	YP33A	2.57	15.2	<20	19.687	0.009	15.744	0.008	39.804	0.027
	Macanal	YP31B	4.28	14.2	69	19.091	0.006	15.713	0.006	39.208	0.019
	Macanal	YP28A				19.088	0.010	15.706	0.009	39.337	0.025
	PZ	PS17-07A				17.689	0.011	15.568	0.014	37.663	0.044

The analysis of the lead isotopes was divided for each region where samples were studied (Table 18). The rock samples were plotted according to their stratigraphic age, together with the pyrite samples available for comparison as a mineral formed in veins during fluid migration. The plumbotectonic model suggests that the source of the lead isotopes in the analyzed samples is the upper crust, as the samples are closer to this curve of the model (Figure 39). The samples of the same age plot close to each other, and the vein samples plot close to their corresponding wall rock, suggesting that the Pb may have been derived from the wall rock. However, such is not the case for the Paleozoic samples, given that the pyrite vein and the rock sample were collected at different locations. The sample of this vein, collected in the southern region of the Eastern Cordillera, plots closer to the orogen curve, which sets it apart from the other samples analyzed (Figure 39). Calculating the initial ratios for the samples, bring them closer to each other suggesting a common source for the Pb. Even though, the

Paleozoic pyrite sample plots along the orogeny curve and apart from the other samples (Figure 40).

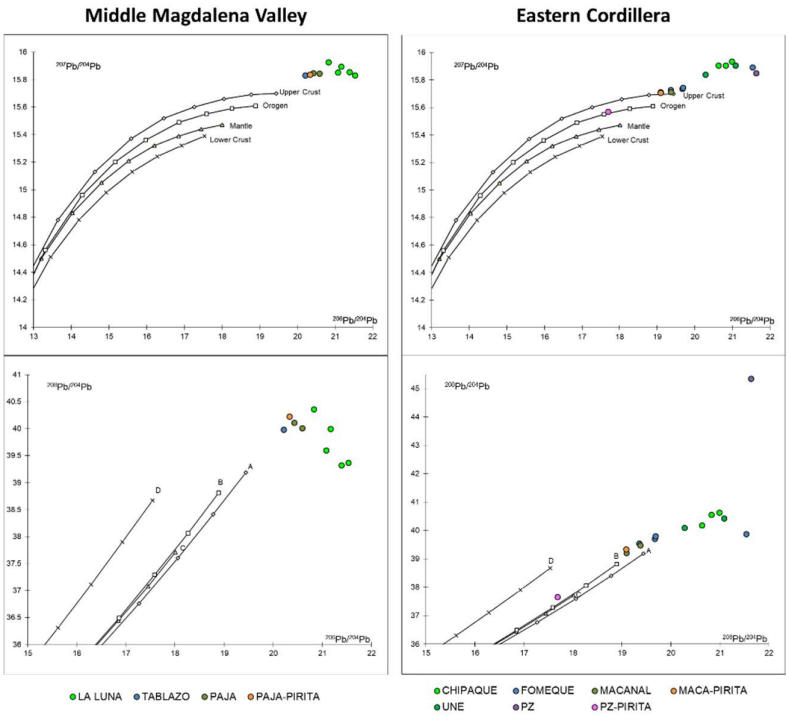

Figure 39. Plumbotectonic model for the samples analyzed in the Middle Magdalena Valley and the Eastern Cordillera. Model from Zartman and Doe (1981).

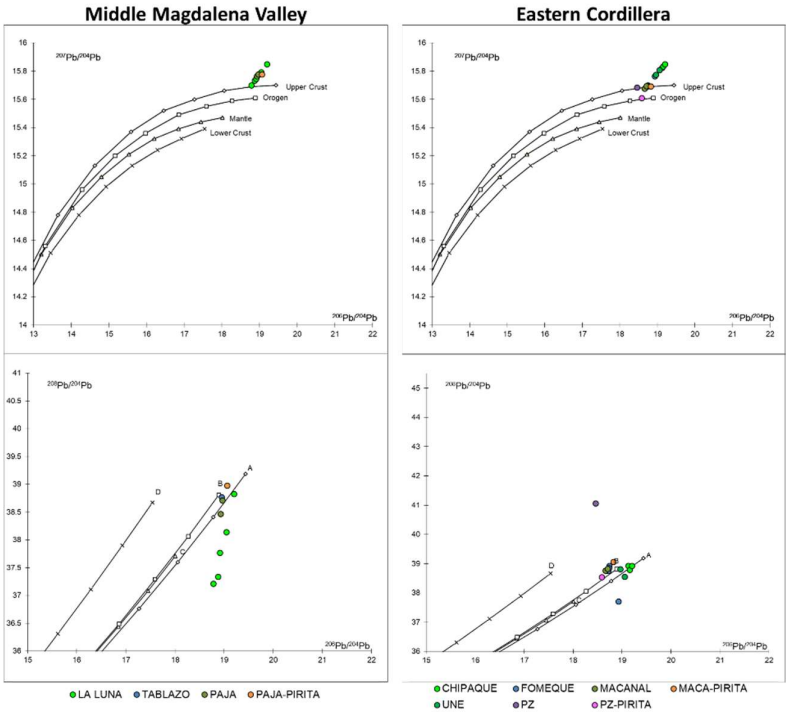

Figure 40. Plumbotectonic model for the samples with Pb ratios calculated for their age of deposition. in the Middle Magdalena Valley and the Eastern Cordillera. Model from Zartman and Doe (1981).

4.2.2. Rb – Sr System

The rocks of the La Luna formation of the MMV contain Rb concentrations that vary between 59.41 ppm and 4.06 ppm, while the Sr content varies between 883.99 and 183.34, with the $^{87}Rb/^{86}Sr$ ratio being 0.9389 to 0.0133 and $^{87}Sr/^{86}Sr$ varying between 0.7075 to 0.718 (Table 19). The isotopic ratio of $^{87}Sr/^{86}Sr$ in the MMV corresponds to the Tablazo formation is 0.7189 and for the Paja formation, the values vary between 0.708 to 0.711.

Table 19.Rb isotopic results for the samples from the MMV Basin

	Formation	Sample	Rb (ppm)	Sr (ppm)	$^{87}Rb/^{86}Sr$	erro (1σ)	$^{87}Sr/^{86}Sr$	erro (2σ)
Middle Magdalena Valley	La Luna	YP43A	13.37	283.18	0.1366	0.0011	0.708241	0.000040
		YP44A	30.49	471.82	0.1870	0.0016	0.708980	0.000041
		YP45B	5.41	786.32	0.0199	0.0002	0.707791	0.000043
		YP46A	4.06	883.99	0.0133	0.0003	0.707568	0.000038
		YP47B	59.41	183.34	0.9389	0.0078	0.718840	0.000041

In the case of Eastern Cordillera, the highest values for the isotopic ratio of $^{87}Sr/^{86}Sr$ corresponds to the Une formation with values between 0.736 to 0.725 and the Chipaque formation with values between 0.733 to 0.729 (Table 20).

Table 20. Sr isotopic results for the samples from the MMV and Eastern Cordillera basins. the concentrations of Rb and Sr expressed are results of the measurements of minor elements (lithogeochemistry)

	Formation	Sample	Rb (ppm)	Sr (ppm)	$^{87}Sr/^{86}Sr$	erro (2s)
MMV	La Luna	YP43A	15	281	0.708241	0.000040
	La Luna	YP44A	29	401	0.708980	0.000041
	La Luna	YP45B	6	615	0.707791	0.000043
	La Luna	YP46A	3	837	0.707568	0.000038
	La Luna	YP47B	57	185	0.718840	0.000041
	Paja	YP14B	48	606	0.709646	0.000043
	Paja	YP14A	69	546	0.711222	0.000015
	Paja	YP14C Pyrite			0.708906	0.000007
	Tablazo	YP19B	89	153	0.718914	0.000043
	Mugrosa	C5 - Oil			0.708775	0.000072
Eastern Cordillera	Chipaque	JC5A	89	136	0.729340	0.000040
	Chipaque	YP58-Comp	102	111	0.733581	0.000043
	Chipaque	YP54-Comp	64	93	0.729579	0.000048
	Fómeque	JC7 B	158	149	0.726842	0.000049
	Fómeque	JC15 D	151	176	0.716862	0.000038
	Fómeque	JC7 E	18	44	0.718959	0.000024
	Fómeque	YP33A	152	266	0.714358	0.000004
	Macanal	YP28A	241	181	0.723976	0.000044
	Macanal	YP31B	163	207	0.723138	0.000022
	Macanal	YP28A Pyrite			0.721012	0.000011
	Paleozoic	YP27A	163	43	0.721351	0.000043
	Paleozoic	PS17-07A Pyrite			0.733320	0.000017
	Une	JC12 C	66	86	0.736747	0.000043
	Une	JC13 A	33	826	0.725171	0.000005

The relation between the samples was compared with curves of the evolution of the $^{87}Sr/^{86}Sr$ ratio since the age of deposition for each unit. For veins, an age of 50 Ma was selected, given that at this age, a regional unconformity was formed across the territory of Colombia, with greater expression in the MMV (Cooper et al., 1995; Mora et al., 2010; Restrepo-Pace et al., 2004; Reyes-Harker et al., 2015). For the crude oil sample, an age of 25 Ma was used, since the reservoir is of Oligocene age (Mugrosa Fm. - 23-34 Ma), and the oil has to be younger than the reservoir, if the hydrocarbon reached this unit during primary migration. Given that the Rb/Sr ratio is necessary to calculate the isotopic ratio at any age, the Rb and Sr content for veins and crude oil had to be selected from adjacent samples or averaged over closer samples. For the crude oil, three Rb/Sr ratios were used, the minimum and maximum observed in the rock samples (0.0035842 and 0.581699 respectively). A calculated value from a regression using rock data for Rb/Sr versus V/Ni, the last one was available for the crude oil and is was possible to estimate a Rb/Sr ratio of 0.121062. Additional samples on the western flank of the Eastern Cordillera (W-EC) were included to provide a wider comparison (Ingeominas - UIS, 2008, 2007).

The samples of the Northern MMV (N-MMV), show closer cluster between most of the samples of La Luna and Paja Fms, as well as the crude oil and Pyrite vein sample from Paja Fm. (Figure 41). One sample from La Luna Fm. and the Tablazo Fm. separate with similar measured ratios, but due to different Rb/Sr ratios, their initial ratios are different. This dispersion of the samples of La Luna Fm. shows the greater variability that could be expected, even within the same stratigraphic unit. The crude oil sample and the vein from Paja Fm. plot closer to the samples of La Luna Fm., however, knowing the difference of the vein with the wall rocks analyzed from Paja Fm., shows that the characterization of these units require large data sets to address the natural variability and confirm or discard possible correlations. The crude oil could have been generated from La Luna Fm. given their similarity, but also other rocks not analyzed could give a positive correlation as the possible origin for this crude oil.

The group of samples from the W-EC were plot in two groups, including the carbonate veins and their respective wall rock into each group (Figure 42). Each group of samples show closer correspondence between the results of the veins and their corresponding wall rocks, with lower or similar dispersion than the samples from the N-MMV (Figure 41).

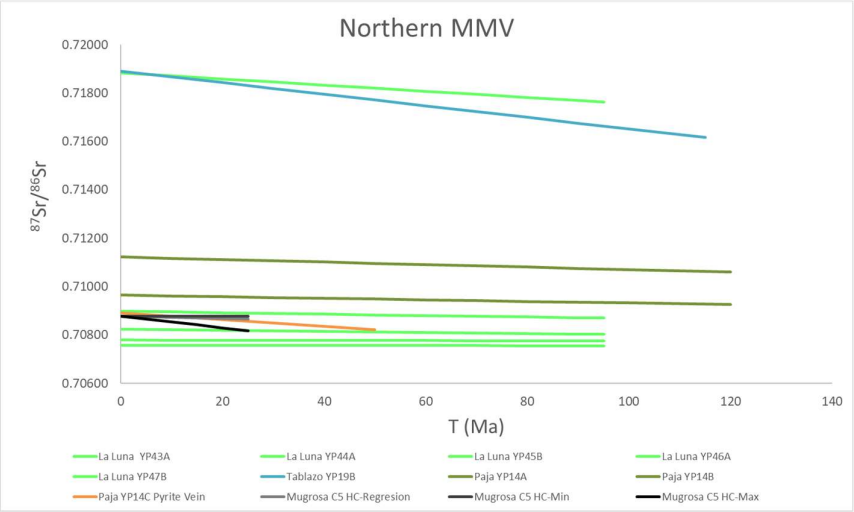

Figure 41. Curves of evolution for the $^{87}Sr/^{86}Sr$ ratio from the actual measured ratio, and calculated for the stratigraphic ages of each sample. Crude oil and Pyrite vein were calculated for the oldest possible age.

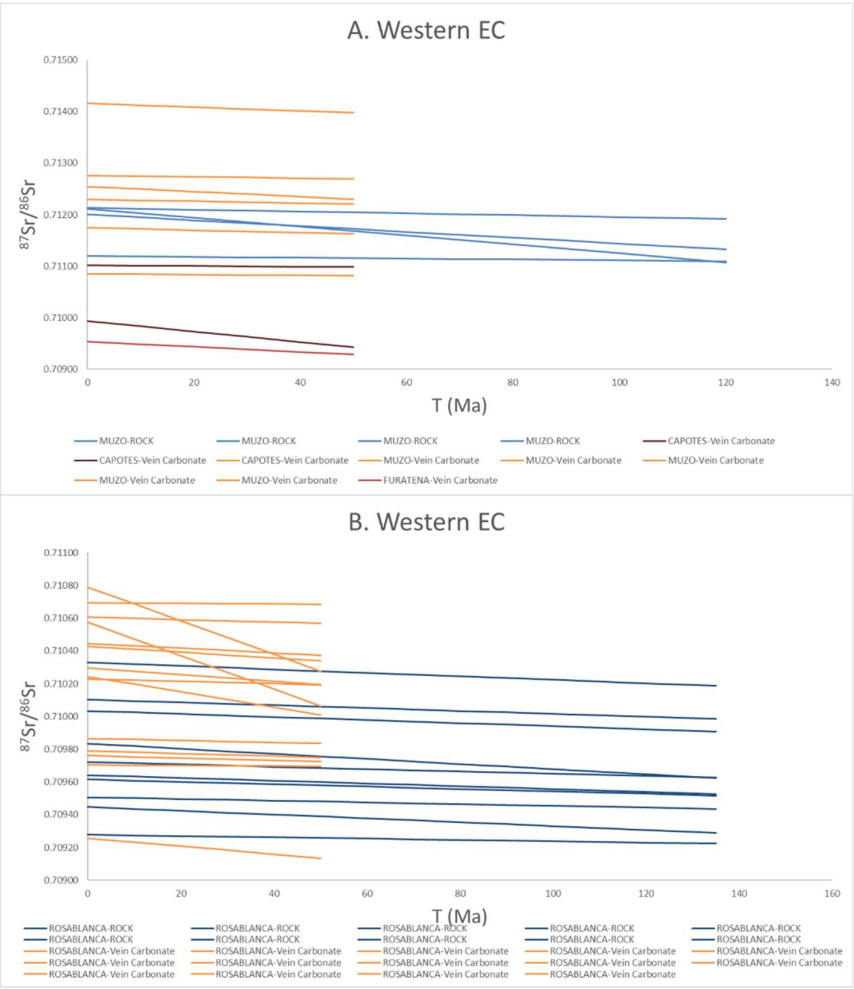

Figure 42. Curves of evolution for the $^{87}Sr/^{86}Sr$ ratio from the actual measured ratio, and calculated for the stratigraphic ages of each sample. Carbonate veins were calculated for the oldest possible age.

The samples of the eastern flank of the Eastern Cordillera (E-EC) analyzed in this project included one Paleozoic sample, which, given its Rb/Sr ratio of 3.79069, shows an evolution different from any of the other samples (Figure 43). The measured $^{87}Sr/^{86}Sr$ ratios of these samples have wider dispersion, when compared to the groups from the W-EC and N-MMV. The Pyrite vein from Macanal Fm. has lower $^{87}Sr/^{86}Sr$ ratio, but using a Rb/Sr ratio 0.72101, it could intersect the evolution of the wall rocks at around 60 Ma, suggesting a good correlation (Figure 43). The pyrite vein from Paleozoic age plots closer to the Chipaque and Une Fm., possibly due to the long-distance geographical location or similar facies with the original wall rocks.

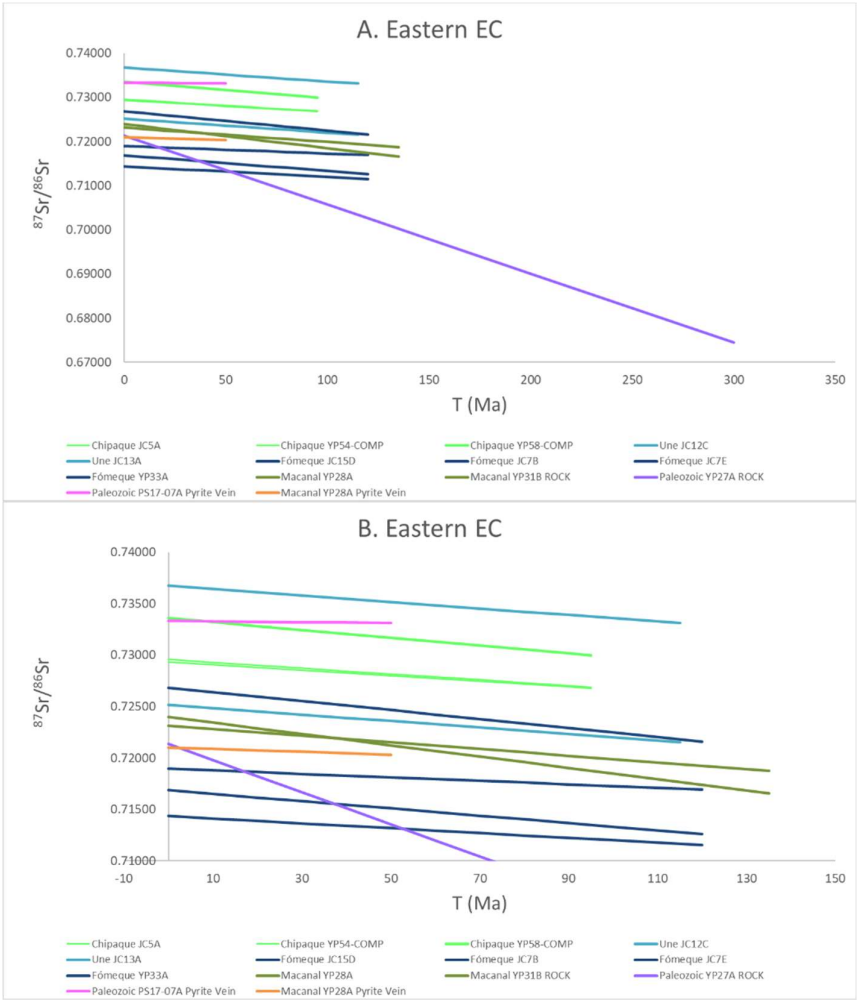

Figure 43. Curves of evolution for the $^{87}Sr/^{86}Sr$ ratio from the actual measured ratio, and calculated for the stratigraphic ages of each sample. Carbonate veins were calculated for the oldest possible age.

The main problem for the comparison between the data of the rocks, and the fluid migration products (veins and crude oil), is the lack of measured Rb/Sr ratio, which will

control the evolution trajectory of the $^{87}Sr/^{86}Sr$ ratio, increasing uncertainties in the possible correlation.

4.2.3. Nd Isotopes

The highest values of the Sm content of the shales of the MMV basin belong to the Paja formation variaty between 8.7 to 6, while the Nd content is between 20 ppm to values less than 10 ppm for Luna formation. In the case of the Eastern Cordillera basin, the highest values of the Sm are from Fómeque formation between 11.1 ppm to 9.3 ppm, and the highest values of the Nd are from Une formation with 27 ppm and the Chipaque formation with 26 ppm to 18 ppm.

The $^{143}Nd/^{144}Nd$ isotopic ratio for the MMV shows highest values (0.512 to 0.5116) for the Luna formation with the (0) 13.31, while for the Eastern Cordillera basin, the Macanal formation has highest values 0.5121 to 0.5120 with the **ε**Nd(0) -10.08 (Table 21).

Table 21. Nd isotopic results for the samples from the MMV and Eastern Cordillera basins

	Formation	Sample	Sm (ppm)	Nd (ppm)	$^{143}Nd/^{144}Nd$	erro (2σ)	$\varepsilon_{Nd(0)}$
MMV	La Luna	YP43A	1	<10	0.512211	0.000005	-8.33
	La Luna	YP44A	2.2	<10	0.511769	0.000004	-16.95
	La Luna	YP45B	0.5	<10	0.511956	0.000007	-13.31
	La Luna	YP46A	0.3	<10	0.511790	0.000010	-16.54
	La Luna	YP47B	6.2	20	0.511627	0.000005	-19.72
	Tablazo	YP19B	6.1	<10	0.511784	0.000004	-16.66
	Paja	YP14B	6	<10	0.511827	0.000004	-15.82
	Paja	YP14A	8.7	11	0.511761	0.000004	-17.10
	Paja	YP14C Pyrite			0.51181	0.000007	-16.15
Eastern Cordillera	Une	JC12 C	8.3	27	0.511694	0.000005	-18.42
	Fómeque	JC7 B	11.1	25	0.511988	0.000005	-12.68
	Fómeque	JC15 D	9.3	18	0.512005	0.000004	-12.35
	Macanal	YP28A	1.9	18	0.512121	0.000005	-10.08
	Paleozóico	YP27A	6.1	14	0.511969	0.000005	-13.04
	Chipaque	YP54-Comp	8.8	18	0.511675	0.000004	-18.78
	Chipaque	JC5A	9.8	26	0.511656	0.000005	-19.16
	Chipaque	YP58-Comp	9.9	23	0.511644	0.000004	-19.39
	Une	JC13 A	3	<10	0.511758	0.000004	-17.17
	Fómeque	JC7 E	4.3	<10	0.511923	0.000004	-13.94
	Fómeque	YP33A	7.7	17	0.511950	0.000004	-13.43
	Macanal	YP31B	4	14	0.512071	0.000003	-11.06
	Macanal	YP28A Pyrite			0.512065	0.000007	-11.18
	PZ	PS17-07A Pyrite			0.51188	0.000008	-14.79

The isotopic $^{143}Nd/^{144}Nd$ ratio in combination with the $^{87}Sr/^{86}Sr$ ratio shows the possible reservoir that gave origin to the analyzed rocks. The regions include the N-MMV and E-EC, showing an important difference in the reservoir from which the rocks were derived (Figure 44 and Figure 45). The samples of the N-MMV show an origin closer to the Lower Continental Crust, and one sample of the La Luna Fm. is located in the field of continental basalts and close to the actual seawater (Figure 44). The vein from the Paja Fm. is located between the samples of the same formation and the samples of La Luna Fm.

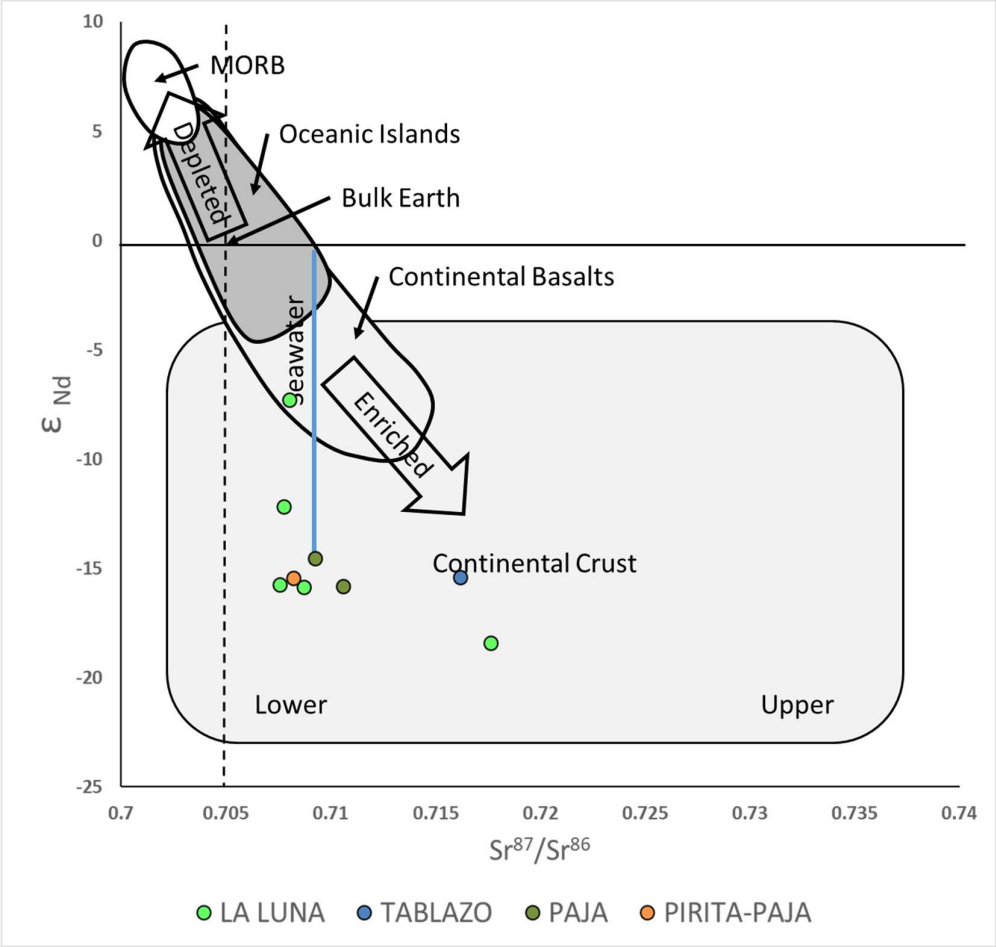

Figure 44. Sr and Nd isotope ratios of the rocks analyzed in the N-MMV, calculated for their respective stratigraphic ages

The samples of the E-EC show an origin from the Upper Continental Crust, with younger samples more enriched with respect to the older ones. The vein of the Macanal Fm. is close to the samples of the same formation, and to the samples of the Fomeque and Paleozoic formations. The vein of the Paleozoic rocks is located close to the younger samples of the Chipaque and Une Fms (Figure 45).

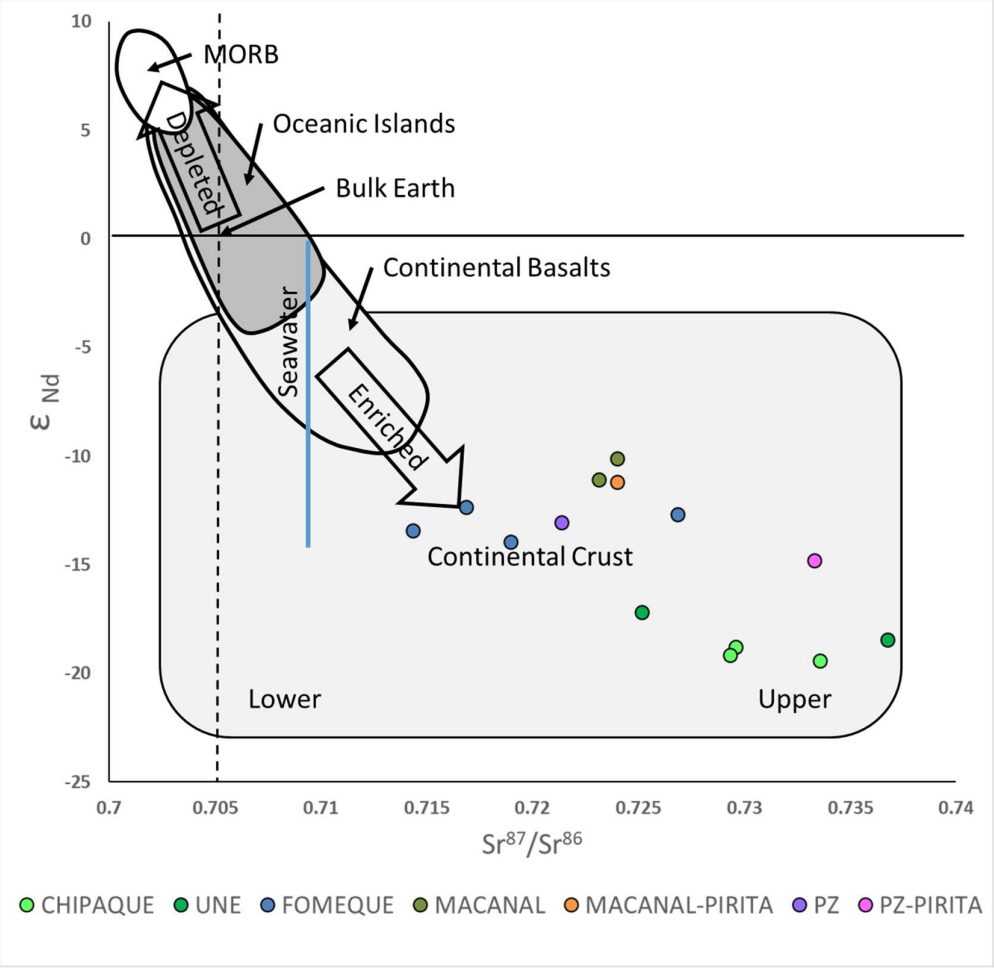

Figure 45. Sr and Nd isotope ratios of the rocks analyzed in the E-EC, calculated for their respective stratigraphic ages.

4.3. **Sediments Provenance**

The Mugrosa Fm. is one of the main hydrocarbon reservoirs in the MMV, corresponding to deposits derived from the initial stages of the Andean Orogeny during the Oligocene (Reyes-Harker et al., 2015). Sampling was carried out in the La Cira - Infantes oil field, with lithologies mainly of sandstones in the outcrop there is presence of hydrocarbons (Figure 46). In this formation, a sample is chosen to perform U/Pb analysis on detrital zircons to study the provenance of the sediments, together with published samples from the study area (Reyes-Harker et al., 2015; Silva et al., 2013).

The results of the detrital zircon U/Pb ages area shown in Figure 47. The distribution of the ages shows peaks at 2 Ga, 0.8 to 1.6 Ga, and between 300 and 100 Ma (Figure 47). The youngest single grain age is 117±2 Ma and the youngest age peak with 4 grains is at 162 Ma (Dickinson and Gehrels, 2009). These ages are older than the depositional age of the sample (Oligocene 33-23 Ma), being all the ages inherited from the rocks in the source area of the sediments at the depositional age.

Figure 46. A) Outcrop of the Mugrosa Formation in the MMV, close to the Cira-Infantas oil field (Station YP-62). B) Detail of the presence of oil in this outcrop.

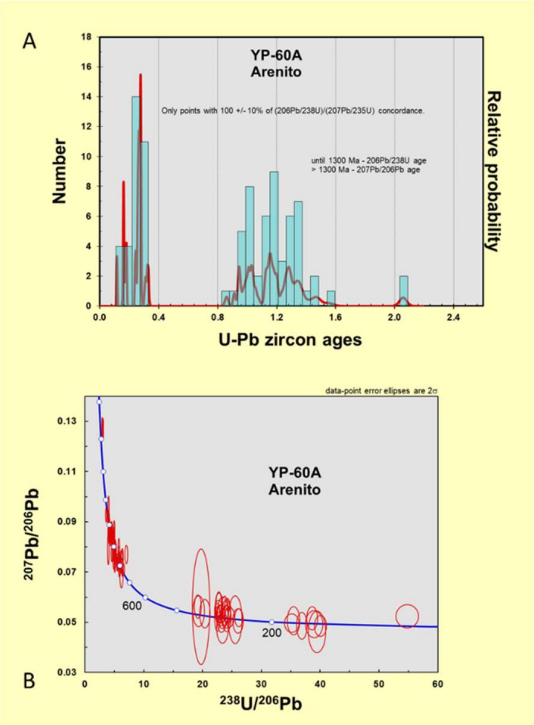

Figure 47. A) Distribution of the ages measured for the analysis of provenance. b) Concordia plots of the U – Pb measured ages.

To study the provenance of any sample, regardless the method, it is necessary to evaluate the possible source areas for the sediments. To stablish a base for comparison, basement ages were gathered from the Radiometric Dating Catalog from Colombia (Gomez et al., 2015a) and were divided into different regions (1887 samples divided in 7 regions; Figure 48).

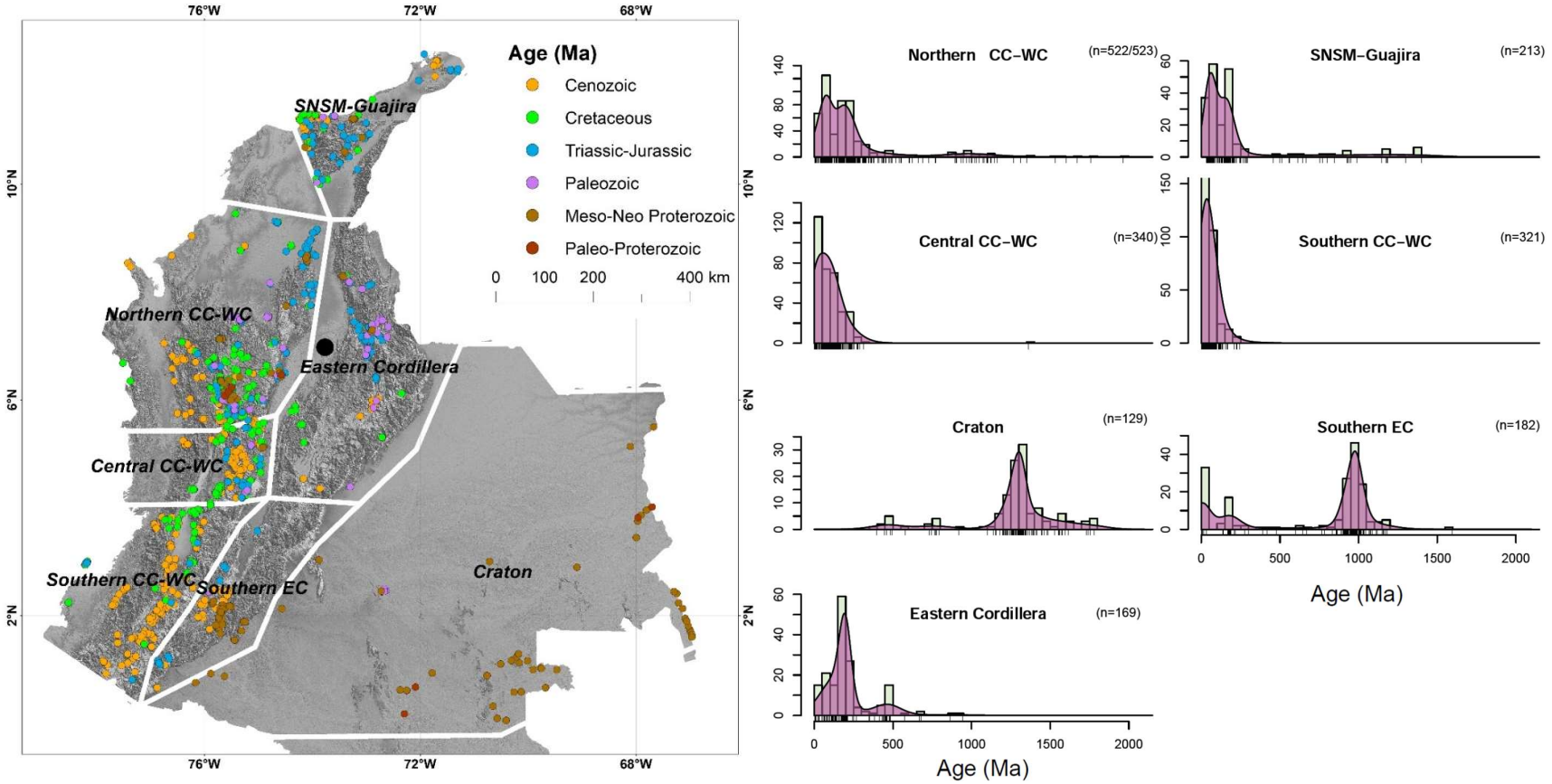

Figure 48. Left: Map of the samples gathered from the Radiometric Dating Catalog from Colombia (Gomez et al., 2015a). Black dot corresponds to the sample analyzed in this project. White lines are the limits for the division of the source areas. Right: Distribution of the ages for each domain selected for comparison.

For better comparison of the source areas, MDS plot was made including the different regions and detrital Cretaceous-Cenozoic samples (Reyes-Harker et al., 2015). The comparison shows that the Craton region sets apart from the other regions (Red circle in Figure 49), while Cretaceous samples are gathered closer to the Craton region (Con-Alb-Cenom-Maas samples in Figure 49). Eocene Samples Are closer to the domains of the Northern CC-WC, Southern EC and Eastern Cordillera (Paz-Esm). Oligocene-Miocene samples are spread between the Cretaceous-Craton samples and the Southern EC, as well as some closer to the Eocene samples (Figure 49). This high dispersion shows variability in the source of the sediments and important reworking of older Cretaceous sediments incorporated to the sediments since the early stages of the Andean Orogeny.

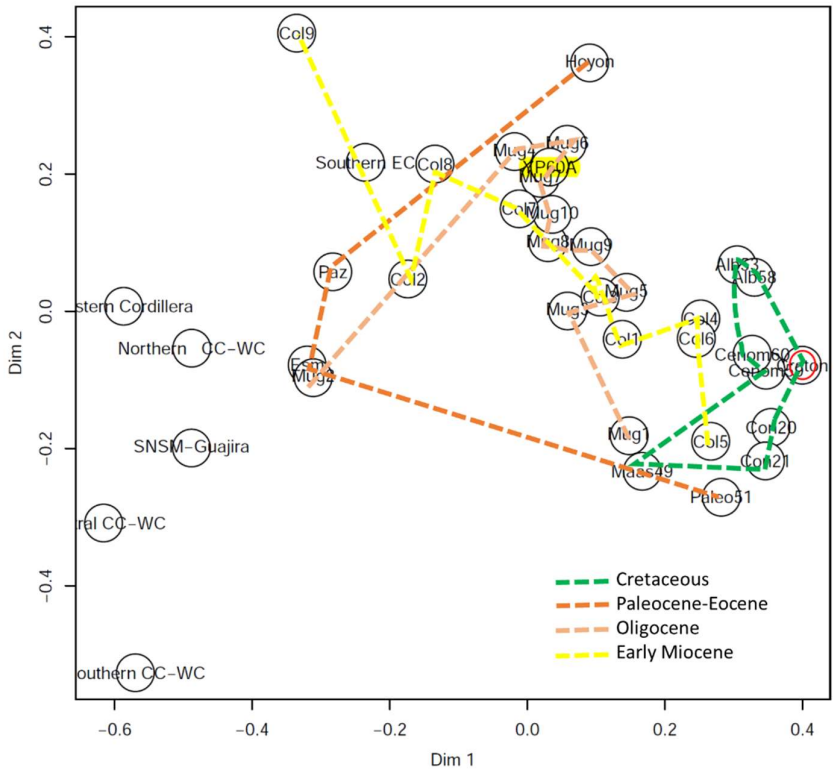

Figure 49. MDS diagram showing the comparison of the basement regions and Cretaceous and Cenozoic detrital samples of the MMV and Eastern Cordillera.

4.4. **Organic Geochemistry and Petrography**

Source rock assessment has been focused on the main aspects that define a good source rock: quantity, quality and maturity. Quantity is evaluated with the total organic carbon content (%TOC) and hydrogen index (HI). HI also allows to evaluate the quality of the source rock, as it dictates the amount of hydrocarbons an immature source rock can expel during its maturation process. Maturity is evaluated with the vitrinite reflectance (%Ro) and maximum temperature of Rock Eval Pyrolysis (Tmax). Data on the proportions of maceral components through organic petrography were available for a subset of samples, with partial results of Rock Eval Pyrolysis, TOC and vitrinite reflectance. The combination of these data sets allows to compare each other and associate the organic geochemistry data with the maceral composition to refine the information regarding the depositional conditions. The sedimentary record was divided in intervals, following the lithostratigraphic nomenclature equivalent across the area of study, and grouping units of related tectonostratigraphic events between major unconformities and with similar sedimentological characteristics.

4.4.1. Rock Eval Pyrolysis – %TOC - %Ro

The data used from these techniques were selected for each studied basin (ANH and GEMS, 2006), and filtered to guarantee their stratigraphic and geographic location (Figure 50), as well as the quality of the data. The data where analyzed for each basin and filtered by stratigraphic age in order to improve visualization, due to the large quantity of data.

The geographic distribution of the samples includes well samples for the MMV and LLA basins; and mostly outcrop samples for the EC. The sampling is not homogenously spread across the study area, but it includes rocks of the Cretaceous, as the main source rocks, as well as Cenozoic, and older rocks of the Paleozoic in the LLA.

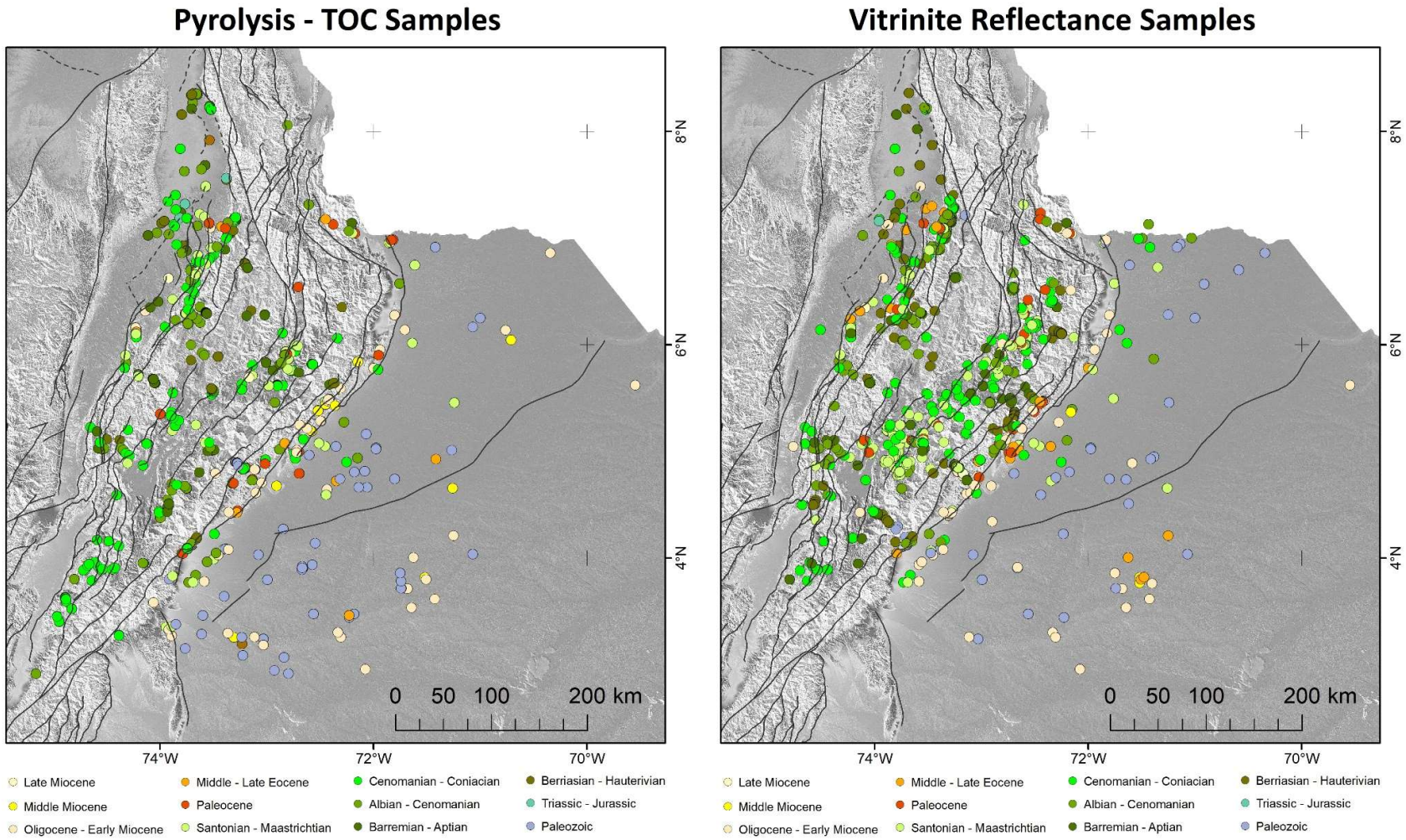

Figure 50. Location map of the selected samples from Rock Eval pyrolysis and total organic carbon (left); and vitrinite reflectance (right) compiled for the study area (ANH and GEMS, 2006). Colors according to the stratigraphic age.

Average and standard deviation profiles of the TOC and HI allow to screen the best source rock intervals, according to their quantity and quality (Figure 51). The TOC profiles show that in the MMV and EC, the Cretaceous and Paleocene rocks have good to excellent quantity of organic matter; while in the LLA, most of the rocks can be considered to have good quantity of organic matter.

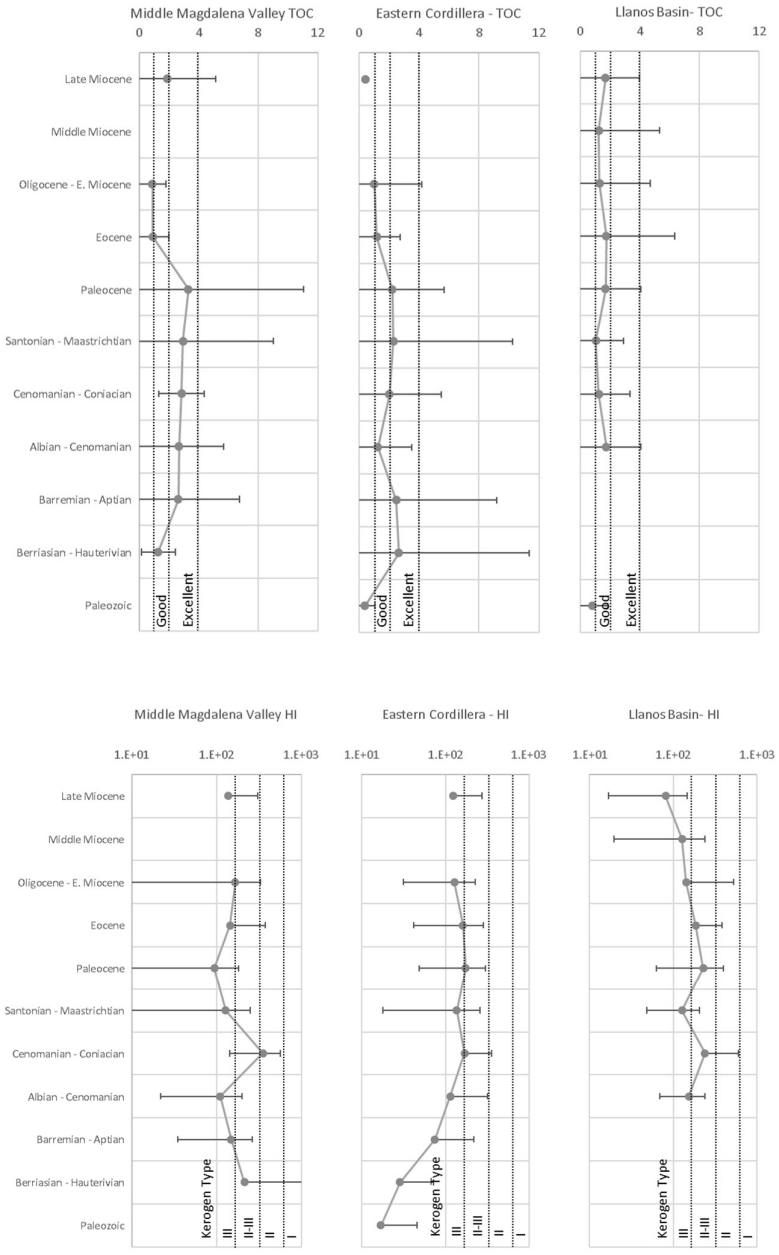

Figure 51. Profiles of the average values and standard deviation for TOC and Hydrogen Index (HI), for each basin and stratigraphic interval.

The hydrogen index also shows good quality for some of the intervals (Figure 51), with values corresponding to kerogen type II and II-III for some of the intervals of the MMV (Berriasian-Hauterivian and Cenomanian-Coniacian), and the LLA (Cenomanian-Coniacian, Paleocene and Eocene). Most of the remaining intervals could be classified as kerogen type III and IV, including those of the EC. However, the Ro profiles in Figure 52 show that for the MMV and EC, the rocks older than Cenomanian-Coniacian are in the oil window or at higher maturity levels, which will cause a severe decrease in the HI values, causing an erroneous assessment if kerogen type III and IV is assigned.

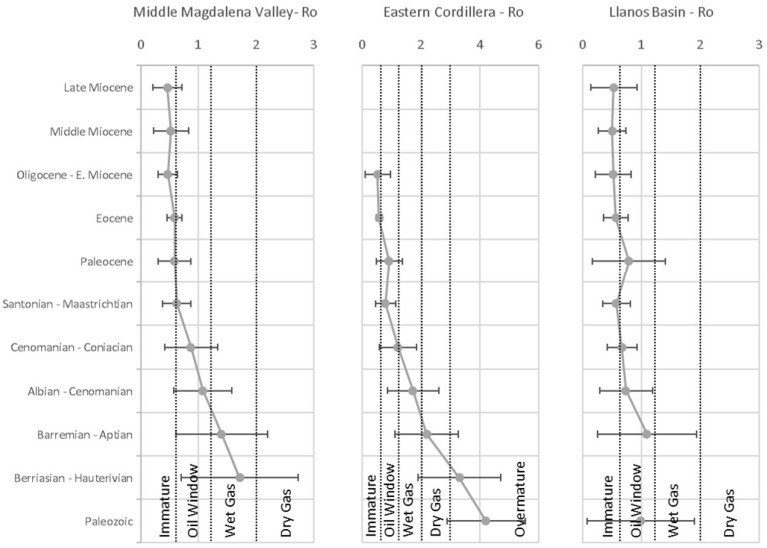

Figure 52. Profiles of the average values and standard deviation for the vitrinite reflectance (Ro), for each basin and stratigraphic interval.

The cross plots of the Genetic Potential Vs. TOC for each basin, show a wide distribution of the samples, with the highest remaining potential for rocks of the Cretaceous in the MMV, followed by the EC (Figure 54); both with data distributed between fair to excellent. For the LLA, where Paleozoic samples are included, most of the samples can be assigned as poor to fair, with a minor group with good to excellent potential. The Cenozoic samples for the three basins have varying potential, with most of the samples in the fields of poor to fair, and fewer samples in the good to excellent fields (Figure 54). The distribution of some of the parameters for each basin is clear at pointing the key differences in the characteristics of their possible source rocks (Figure 53).

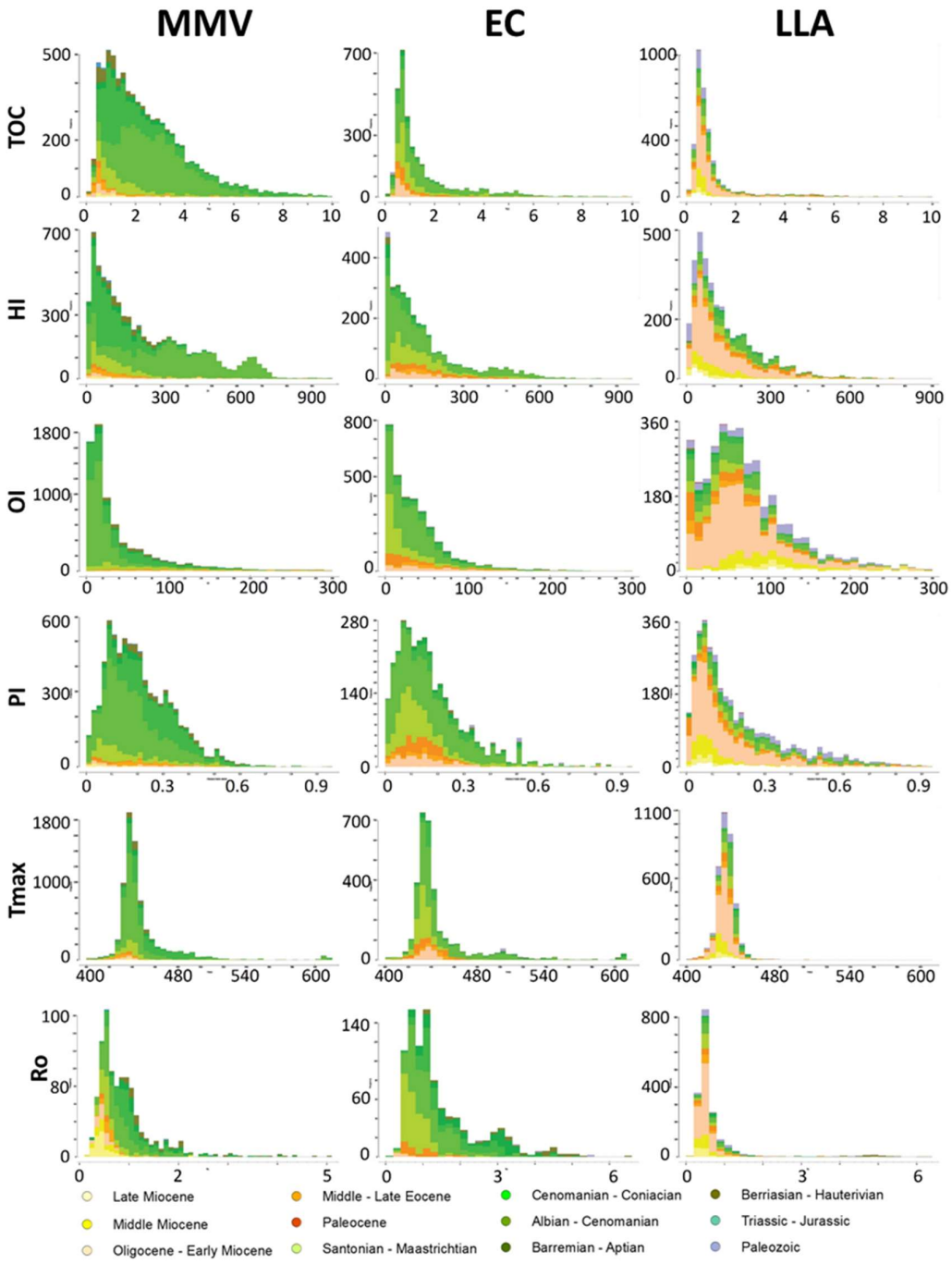

Figure 53. Distribution for each basin of multiple parameters used to characterize the sources rocks. TOC: Total Organic Carbon; HI: Hydrogen Index; OI: Oxygen Index; PI: Production Index; Tmax: Rock Eval Pyrolysis temperature at transformation peak S2; Ro: Vitrinite Reflectance. Numbers indicate the maximum value for each parameter. Vertical axis is frequency in all diagrams.

The samples of Cenozoic age are more abundant in the LLA with respect to EC and MMV. TOC and HI have wider distribution for the MMV than EC and LLA. OI is higher on the LLA, while the EC has intermediate values and the MMV has the lower values

of OI. PI is higher on the MMV, followed by EC and LLA has the lowest values. Tmax has wider distribution for MMV and EC, and even bimodal for EC; while for LLA, it has a narrower distribution. On the EC, Ro has a bimodal pattern similar to that of Tmax, with higher values reaching 3.5%Ro, indicating higher thermal maturity. The MMV is next on the maturity data, reaching 2%Ro for the main group. The LLA has the lowest thermal maturity, with most of the samples below 0.6%Ro, and fewer samples showing values up to 5 to 6%Ro.

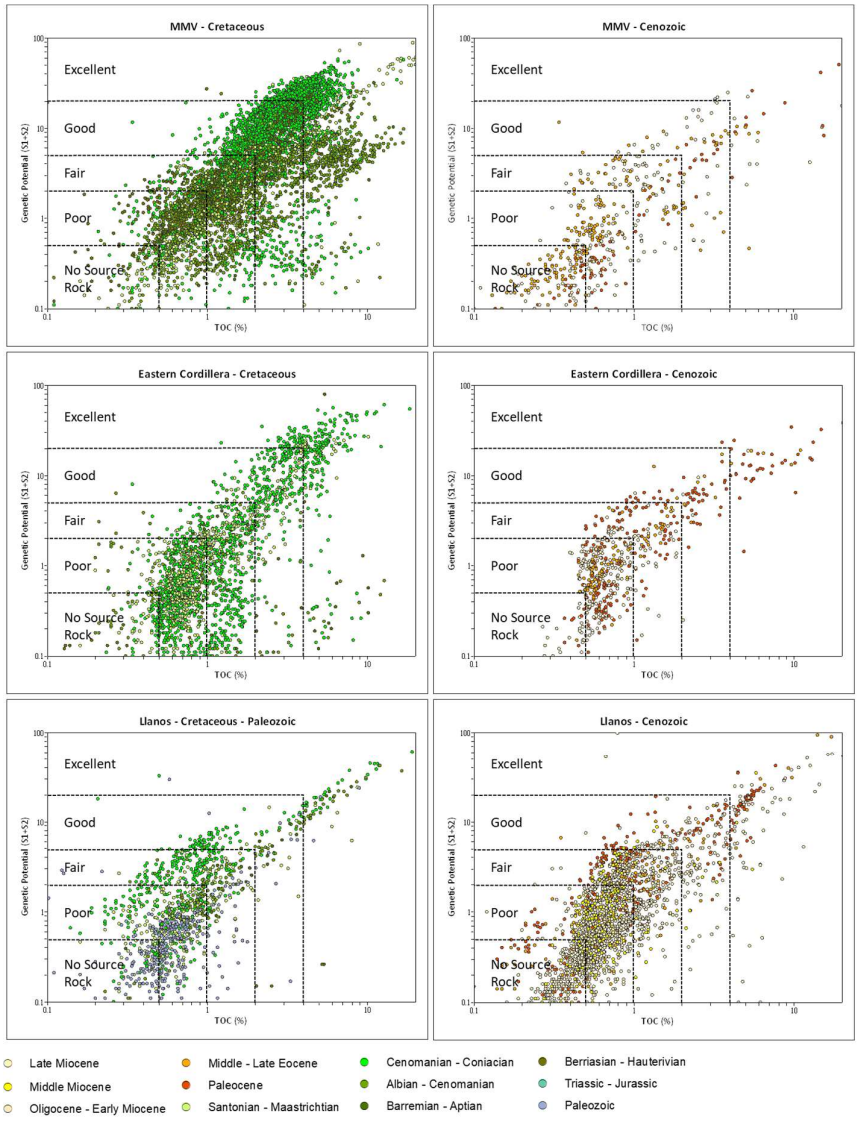

Figure 54. Source rock assessment according to the quantity (TOC) and quality (Genetic Potential) of the organic matter. Samples were plotted in different diagrams for Cretaceous and Cenozoic samples for each basin.

The analysis of the kerogen type using the cross plot of HI Vs. Tmax, better explains the distributions of the quality of the source rocks for each basin (Figure 55). The

comparison of the basins shows that for the Cretaceous samples, que quality diminishes from the MMV (HI> 600 mgHC/g rock), to the EC (HI< 600 mgHC/g rock), and finally at the LLA, where the quality is lower with HI values under to 400 mgHC/g rock. Conversely, the samples of the Cenozoic age show increasing quality from west to east (MMV<EC<LLA), reaching values of HI close to 500 mgHC/g rock in the LLA. The Tmax values are disperse, with values higher than 430 along the maturation trajectories of different kerogen types.

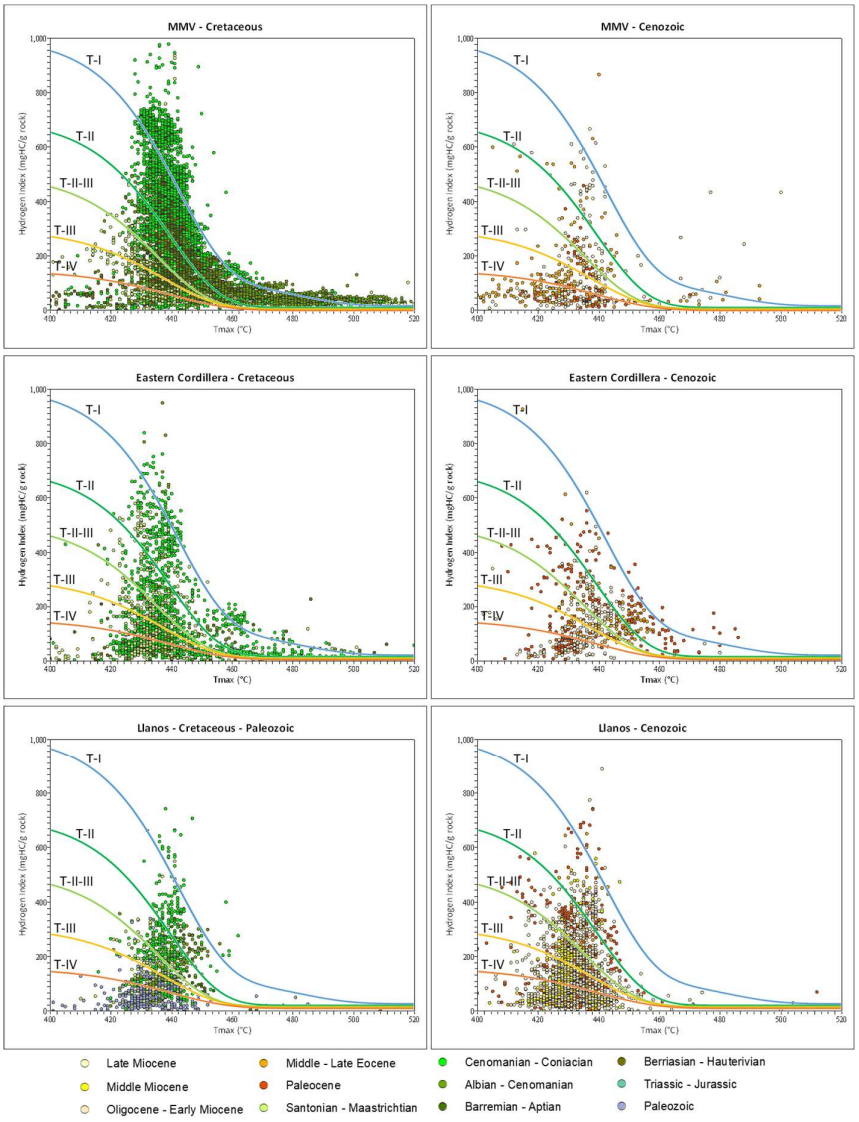

Figure 55. Source rock assessment according to the quality of the organic matter (HI), and maturity (Tmax). Samples were plotted in different diagrams for Cretaceous and Cenozoic samples for each basin.

The assessment of the kerogen type including the modified Van Krevelen diagram (Figure 56), show that for the MMV and EC, most of the Cretaceous samples are

divided between type I-II and type III, with lower oxygen index (OI), when compared with Cretaceous samples of the LLA. The Cenozoic samples show a similar pattern for the three basins, with higher OI values, implying increasing input of organic matter originating from land plants. A small group of Paleocene samples in LLA and EC show OI values below 20, corresponding to a kerogen type II-III.

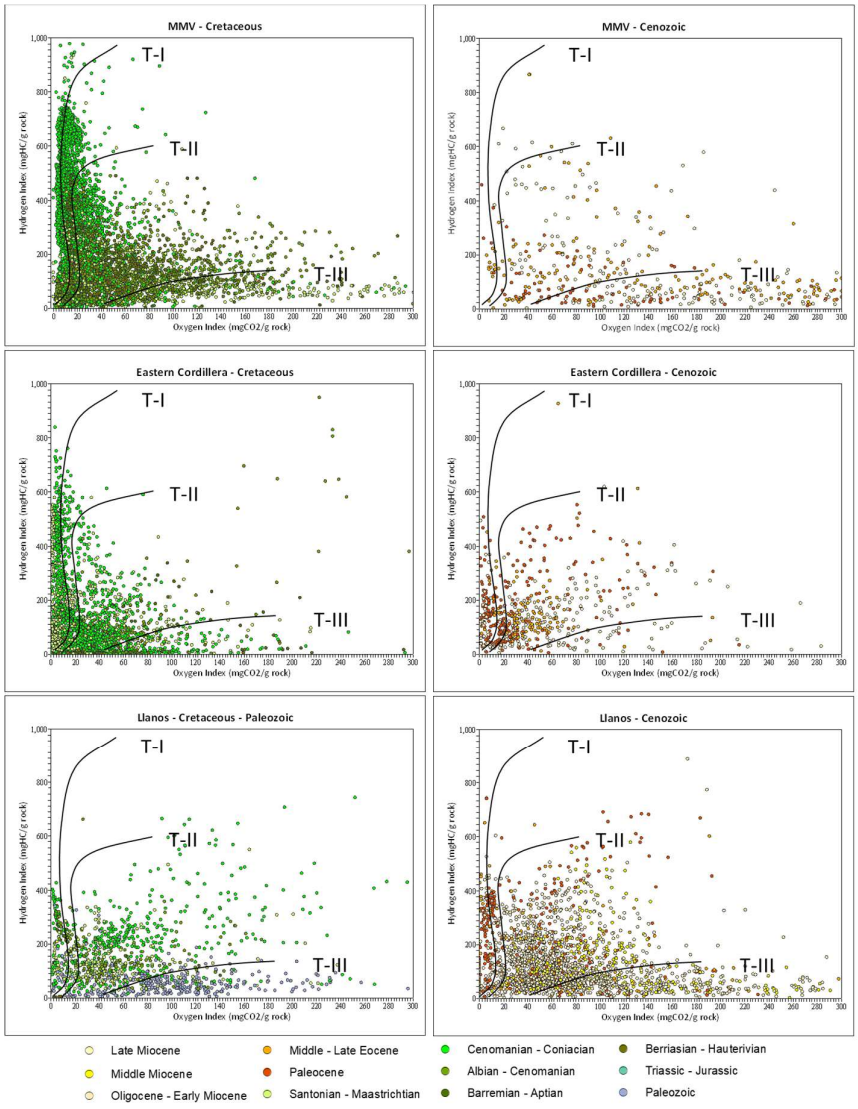

Figure 56. Modified Van Krevelen diagram (HI Vs. OI), showing the types of kerogen of the samples. Samples were plotted in different diagrams for Cretaceous and Cenozoic samples for each basin.

The thermal maturity indicators available for these samples show that most of the Cretaceous samples in the MMV and EC are within the oil window, or overmature; while the Cenozoic samples are mostly below the oil window, including the Cretaceous and Cenozoic samples in the LLA (Figure 57). The relationship between Ro and Tmax

is not as clear as expected, and given the kinetic variability of the organic matter in the source rocks (Chen et al., 2018, 2017b, 2017a, 2016), it is possible to have misleading interpretations when relying solely on Tmax values.

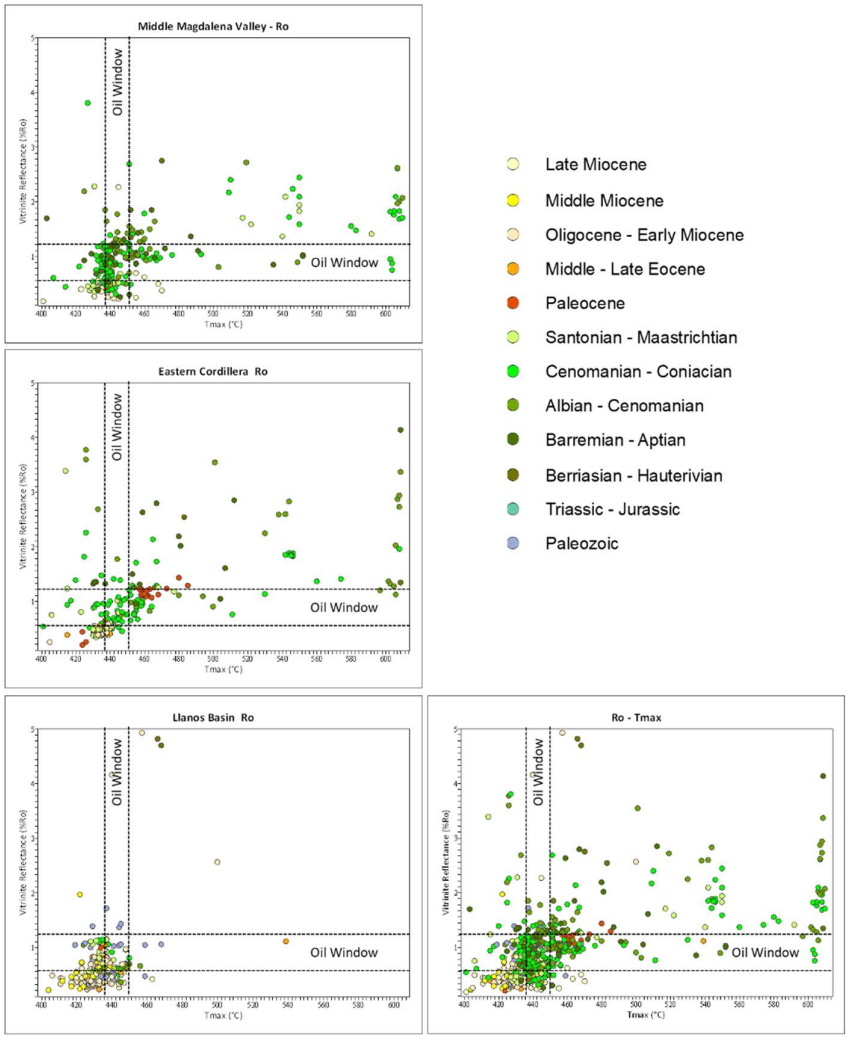

Figure 57. Cross plot of Ro Vs. Tmax as a comparison of the maturity of the samples that had both data available.

The cross plot previously shown gives an idea of the range of data we have for the stratigraphic intervals and each region studied; however, the spatial distribution needs to be addressed through maps that illustrate the distribution of some of the main variables. Maps for TOC, HI and Ro were produced for each stratigraphic interval across the region of study (Figure 58 to Figure 62). These maps were created for the area of known deposit for each stratigraphic interval and in some cases, were restricted to the limits of the studied basins.

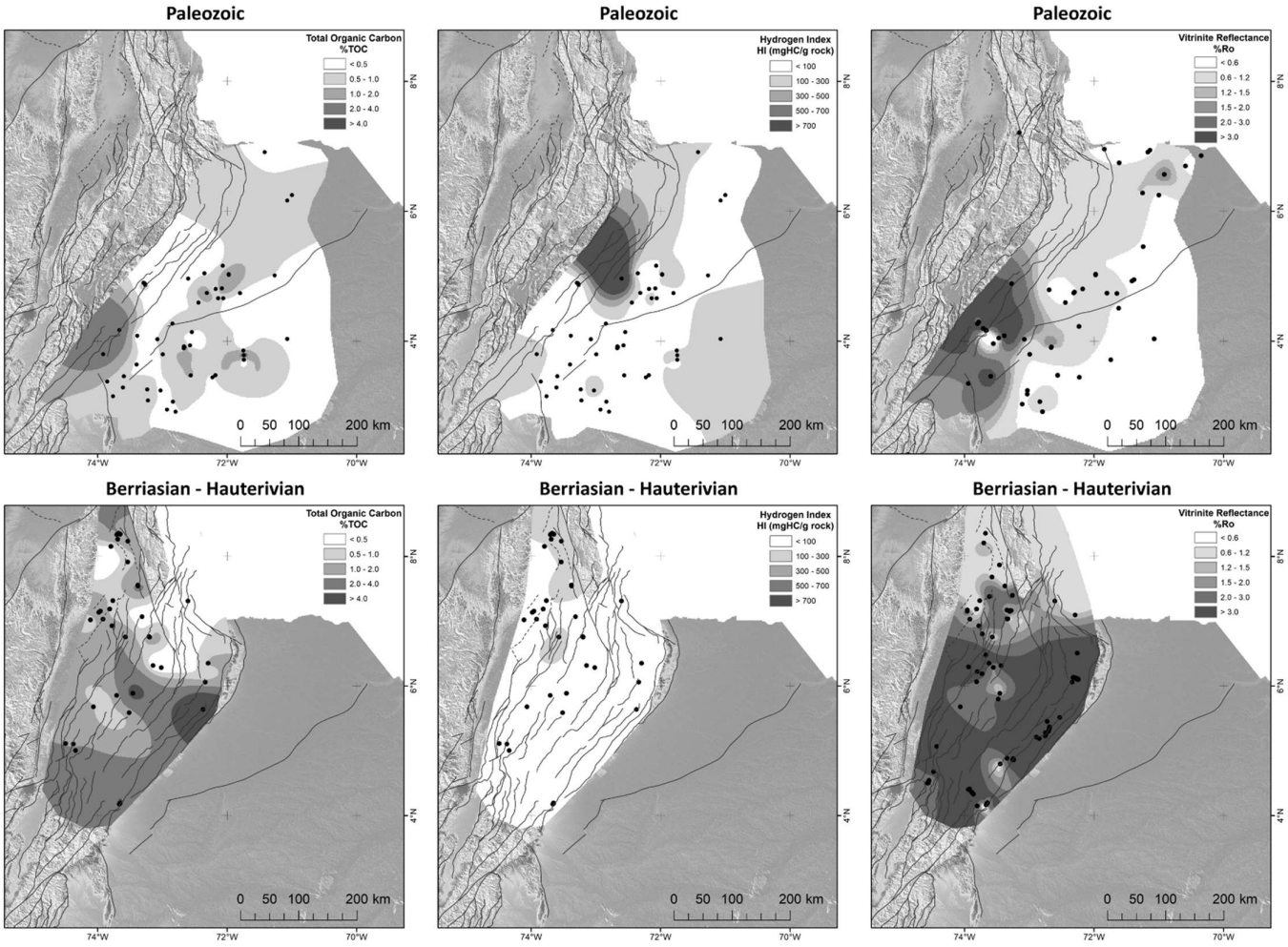

Figure 58. Map of distribution for Total Organic Carbon (%TOC), Hydrogen Index (HI – mgHC/g rock) and Vitrinite Reflectance (%Ro), for the Paleozoic and Berriasian – Hauterivian intervals. Extent for Paleozoic after Reyes-Harker et al. (2015), and for the Berriasian – Hauterivian after Sarmiento (2019).

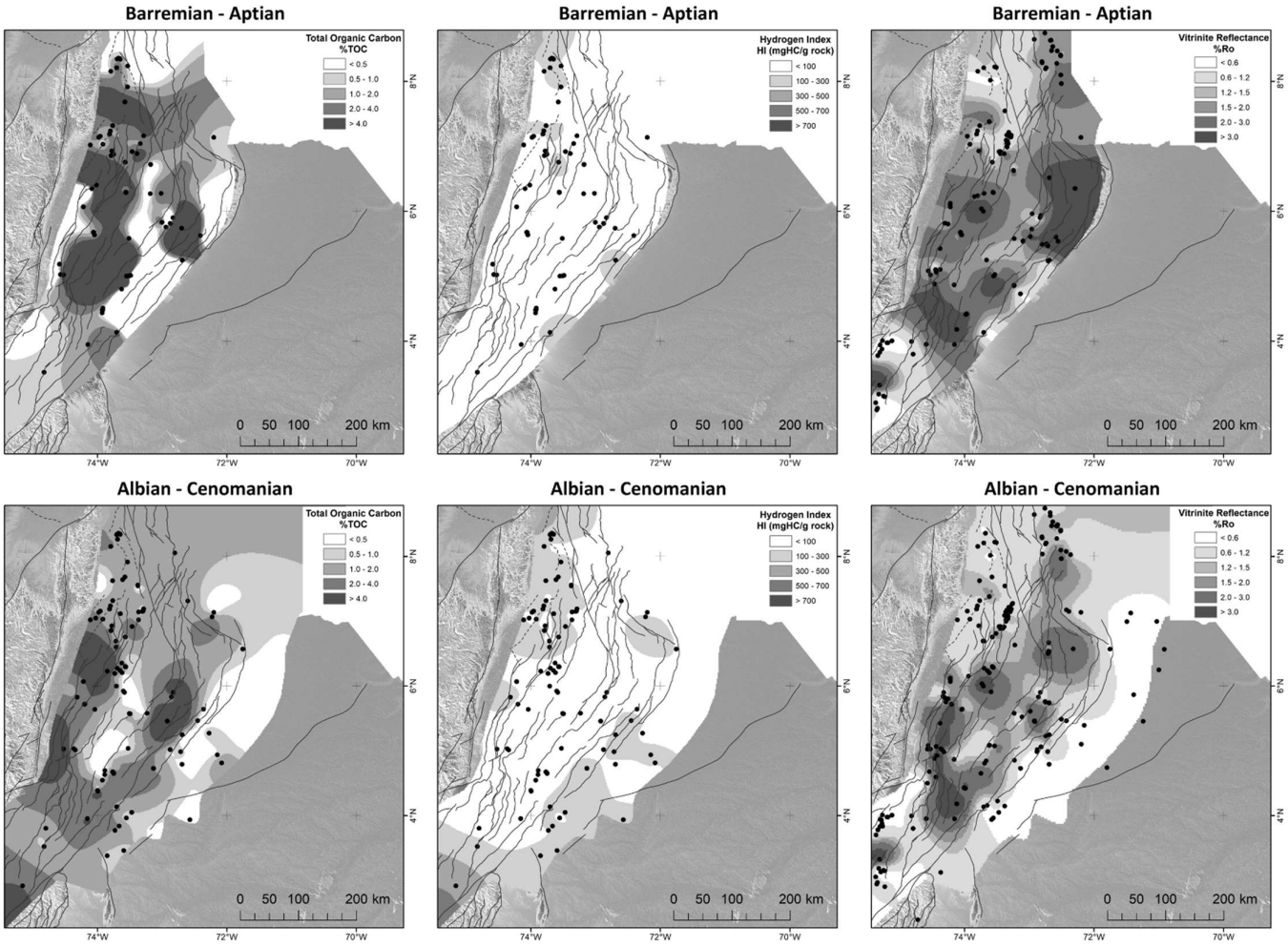

Figure 59. Map of distribution for Total Organic Carbon (%TOC), Hydrogen Index (HI – mgHC/g rock) and Vitrinite Reflectance (%Ro), for the Barremian - Aptian and Albian – Cenomanian intervals. Extent of maps after Reyes-Harker et al. (2015) and Sarmiento (2019).

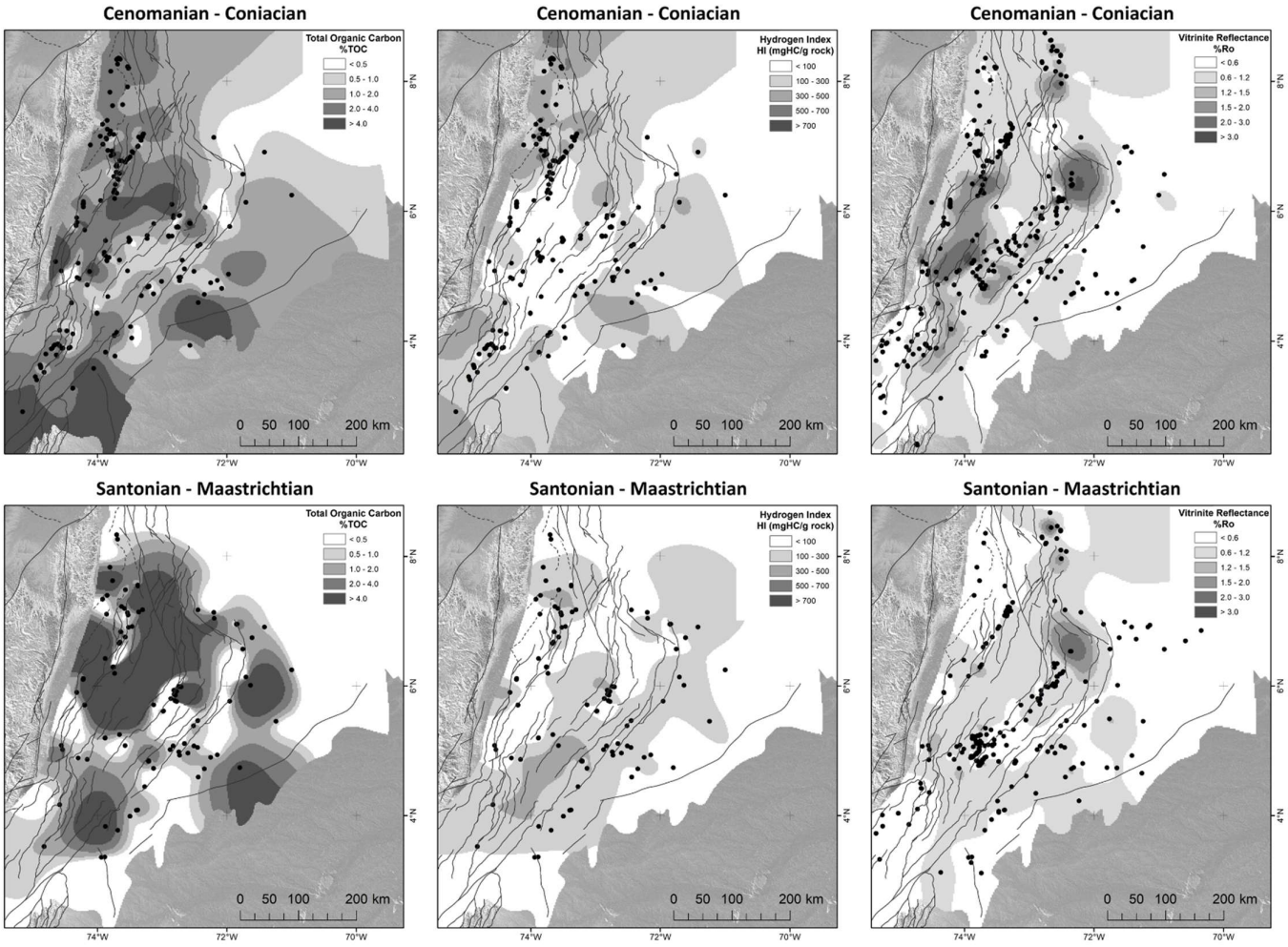

Figure 60. Map of distribution for Total Organic Carbon (%TOC), Hydrogen Index (HI – mgHC/g rock) and Vitrinite Reflectance (%Ro), for the Cenomanian - Coniacian and Santonian – Maastrichtian intervals. Extent of maps after Reyes-Harker et al. (2015) and Sarmiento (2019).

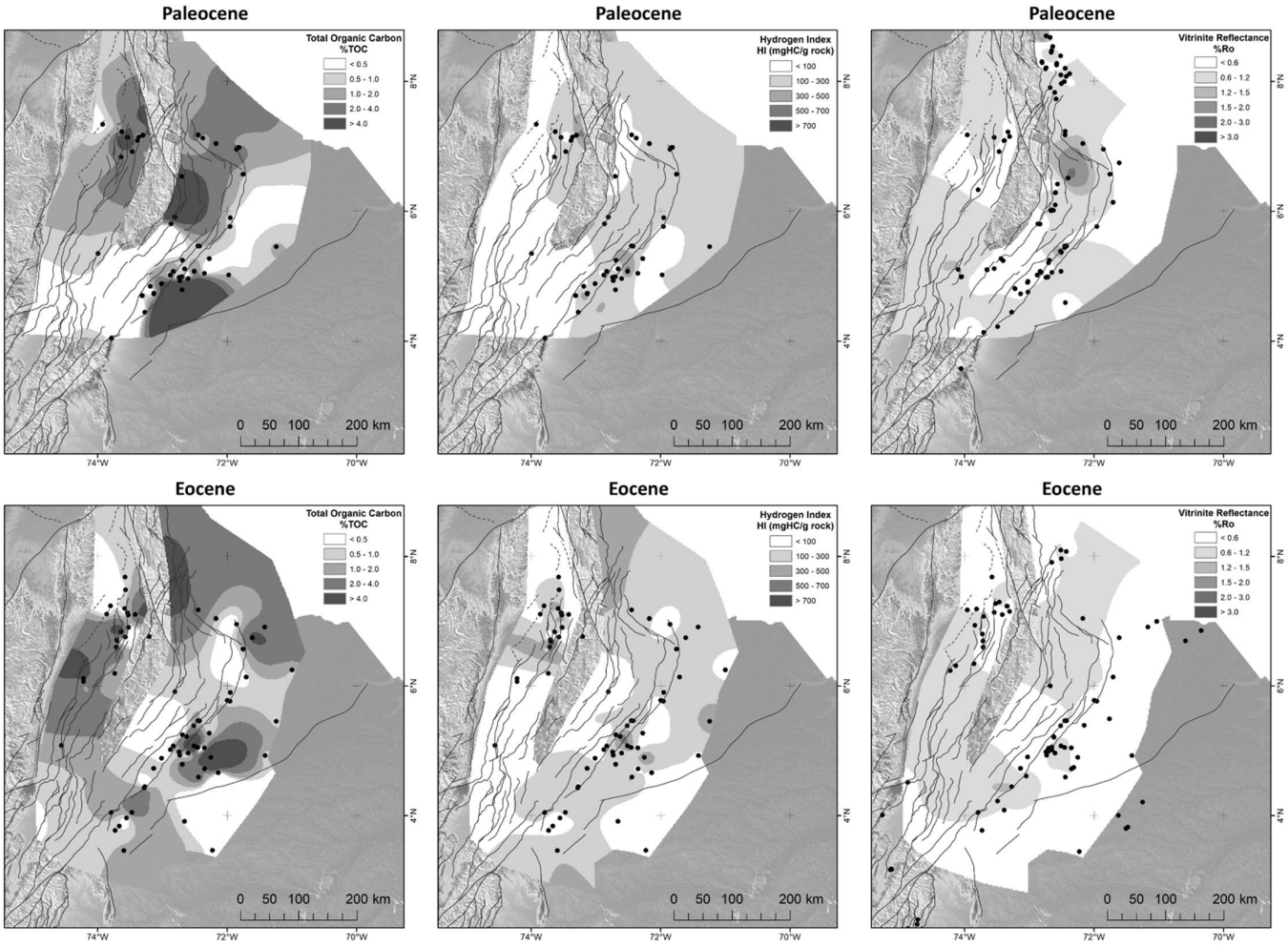

Figure 61. Map of distribution for Total Organic Carbon (%TOC), Hydrogen Index (HI – mgHC/g rock) and Vitrinite Reflectance (%Ro), for the Paleocene and Eocene intervals. Extent of maps after Reyes-Harker et al. (2015).

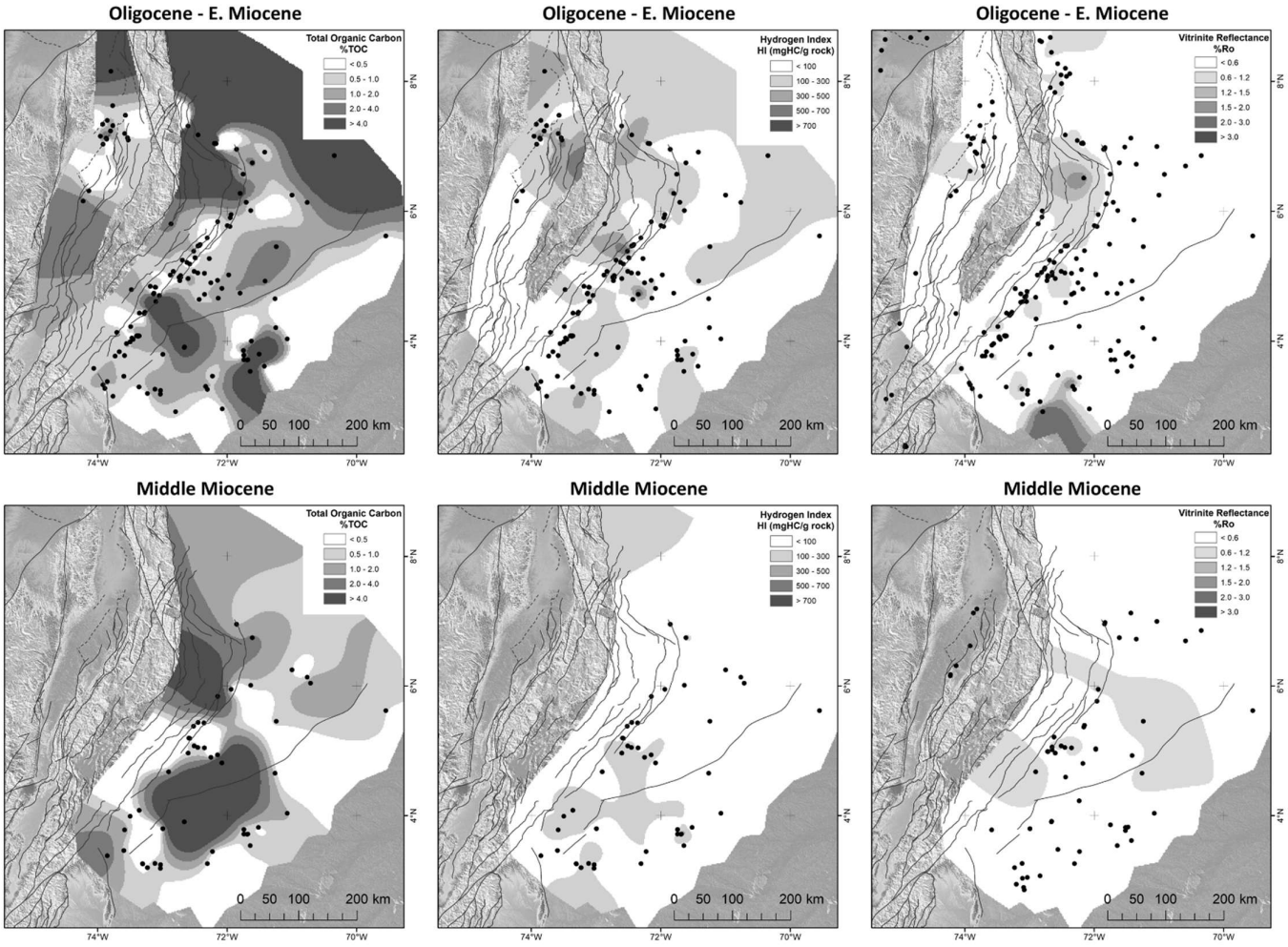

Figure 62. Map of distribution for Total Organic Carbon (%TOC), Hydrogen Index (HI – mgHC/g rock) and Vitrinite Reflectance (%Ro), for the Oligocene – Early Miocene and Middle Miocene intervals. Extent of maps after Reyes-Harker et al. (2015).

Figure 58 to Figure 62 show maps with the interpolation and location of the samples for each interval. It's particularly true with this type of analyses that they are usually done with preference in organic rich samples, instead of regular distribution of sampling. Due to these complications, interpolation of some of the variables were produced in maps with the location of the samples to evaluate the distribution of data. The results of these interpolations should be taken with care around the samples, but could be meaningless where data were extrapolated. Some of the values were not used for average calculation to avoid confusing results (less than 1% of the samples), including HI>2000 and TOC>20%.

Comparing the resulting maps with the distribution of the same parameters for each basin (Figure 53), the higher values are not represented on the maps, due to the effect of averaging the values that are closer or at the same location, such as in wells, where the data are distributed vertically, a dimension not considered in the maps. Furthermore, those high values comprise a less than 10% of the total samples, or even lower fraction for most of the stratigraphic intervals.

The thermal maturity of the samples is crucial at determining their generation potential, since, as can be seen for most of the Cretaceous rocks (Figure 52, Figure 53, and Figure 58 to Figure 62), they have undergone partial generation to total exhaustion of their generation potential, at values above 1.5%Ro. This reduction in their actual generation potential, has reduced the importance of the Lower Cretaceous formations during the hydrocarbon exploratory phases and generally are disregarded as source rocks.

4.4.2. Organic Petrography

The proportion of the main maceral groups in the organic matter of any source rock, will reflect their paleoenvironmental conditions, and will aid in the characterization of the source rocks, and their potential. This information can be used as a guide for indirect oil-source rock correlation, helping in the categorization of the source rocks. Here is studied a large data set of maceral composition for every stratigraphic interval (Figure 63), and part of this data set with Rock Eval pyrolysis data is used to further analyze the relations between these parameters.

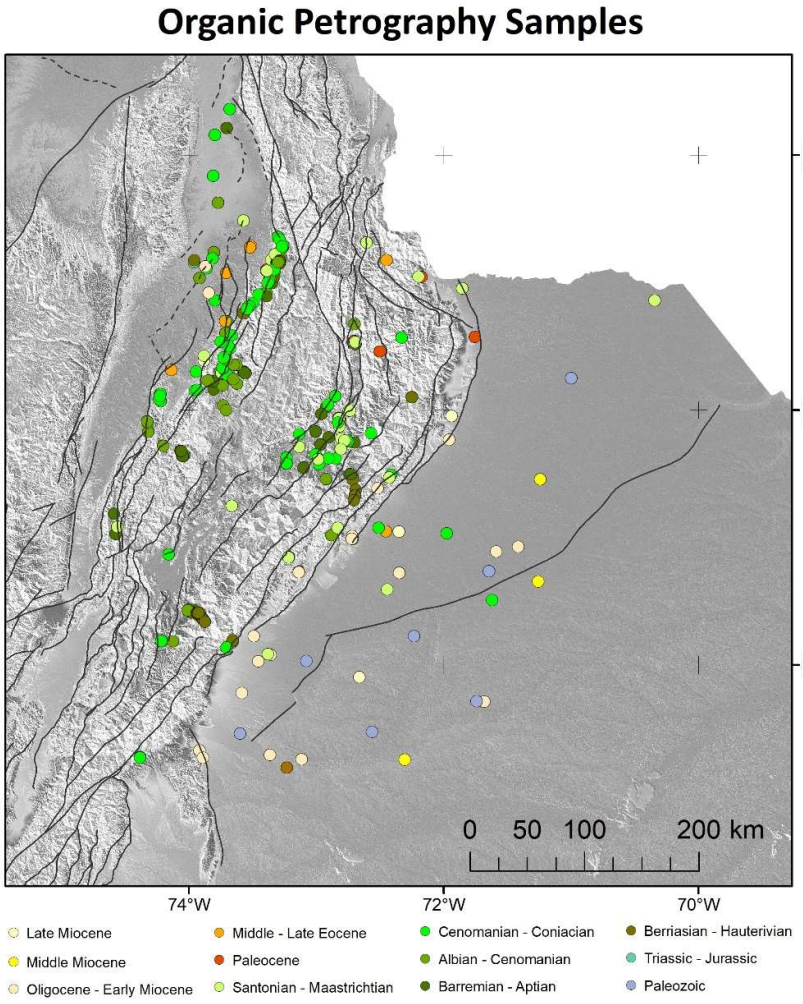

Figure 63. Location map of the selected samples from organic petrography compiled for the study area (ANH and GEMS, 2006). Colors according to the stratigraphic age.

The maceral composition data shows that for the MMV, most of the samples correspond to kerogen types I, II and II-III, with potential to generate mostly Oil and Oil-Gas mixtures. Some Cretaceous samples show a trend towards kerogen type IV, more likely to generate Gas (Figure 64). The EC samples show potential to generate Oil and Oil-Gas mixtures. This region has higher proportion of Cenozoic samples with varying composition between kerogen type I, II, II-III and IV, while most of the Cretaceous samples correspond to kerogen type I and II, with some samples in the field of kerogen type IV (Figure 64). The LLA region presents two groupings, the first with higher proportion of liptinite with most of the Cretaceous samples and Cenozoic samples as young as Middle Miocene. This group shows higher potential to generate Oil and Oil-Gas mixtures (Figure 64). The second group, with higher proportion of inertinite, includes fewer Cretaceous samples, and mostly Oligocene and Late Miocene samples. This group has potential for generate Oil and Oil-Gas mixtures, but also the younger samples could generate mostly Gas (Figure 64).

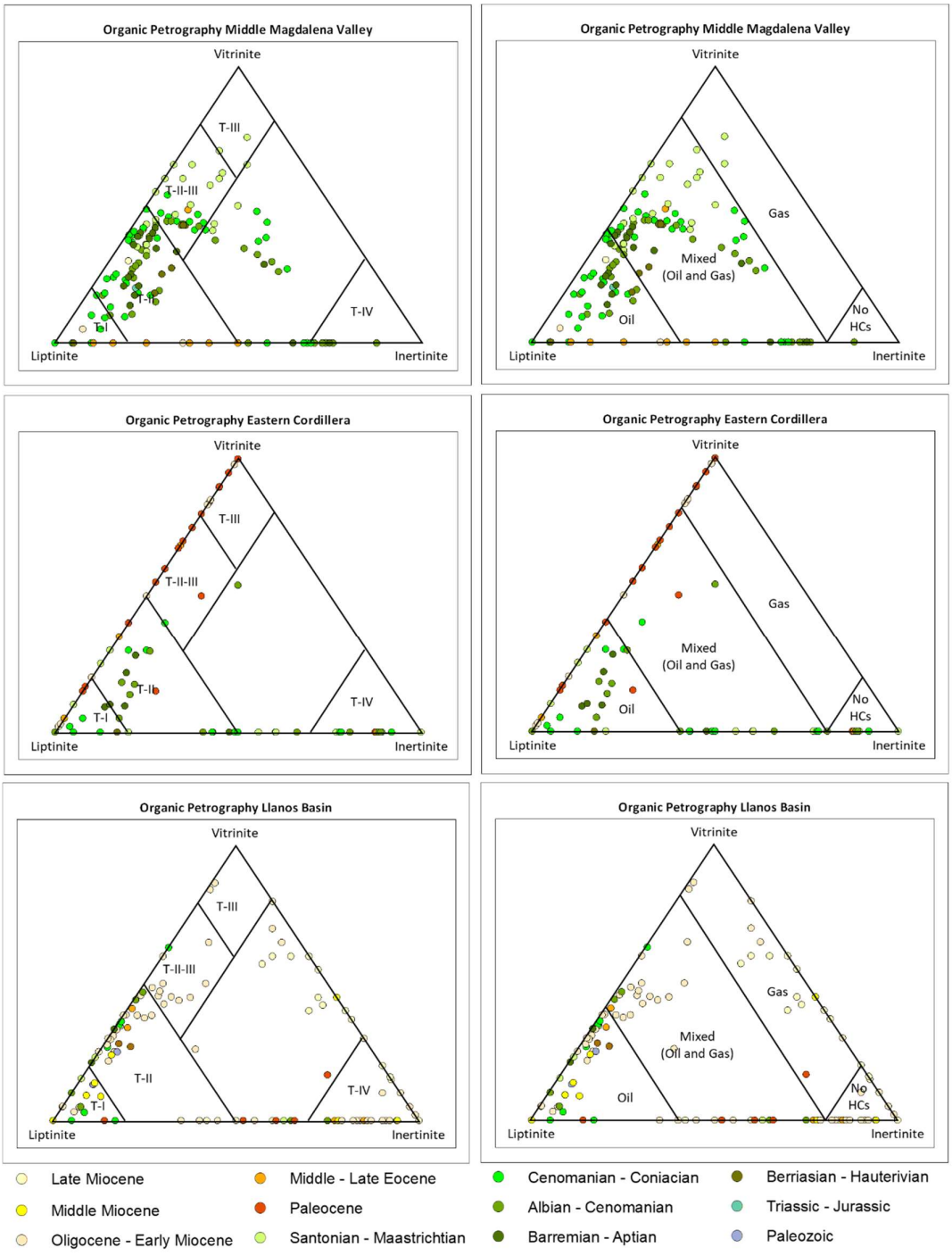

Figure 64. Left: ternary diagrams on the kerogen type according to maceral composition. Right: ternary diagrams on the type of hydrocarbon to be generated derived from maceral composition (Hakimi et al., 2018).

The maceral components allow to classify the samples according to their depositional environment, which can be complementary to the organic geochemistry characterization. The environmental classification shows that the rocks of the MMV were mostly deposited at environments of proximal suboxic-anoxic shelf, suboxic-anoxic lakes and distal suboxic-anoxic basin (Figure 65). A minor portion of the

samples indicates deposition at distal dysoxic-anoxic to oxic-dysoxic shelf. Latest Cretaceous samples of Santonian-Maastrichtian age show an increase in vitrinite, and were deposited in a heterolithic oxic proximal shelf and shelf to basin transition (Figure 65 and Table 22).

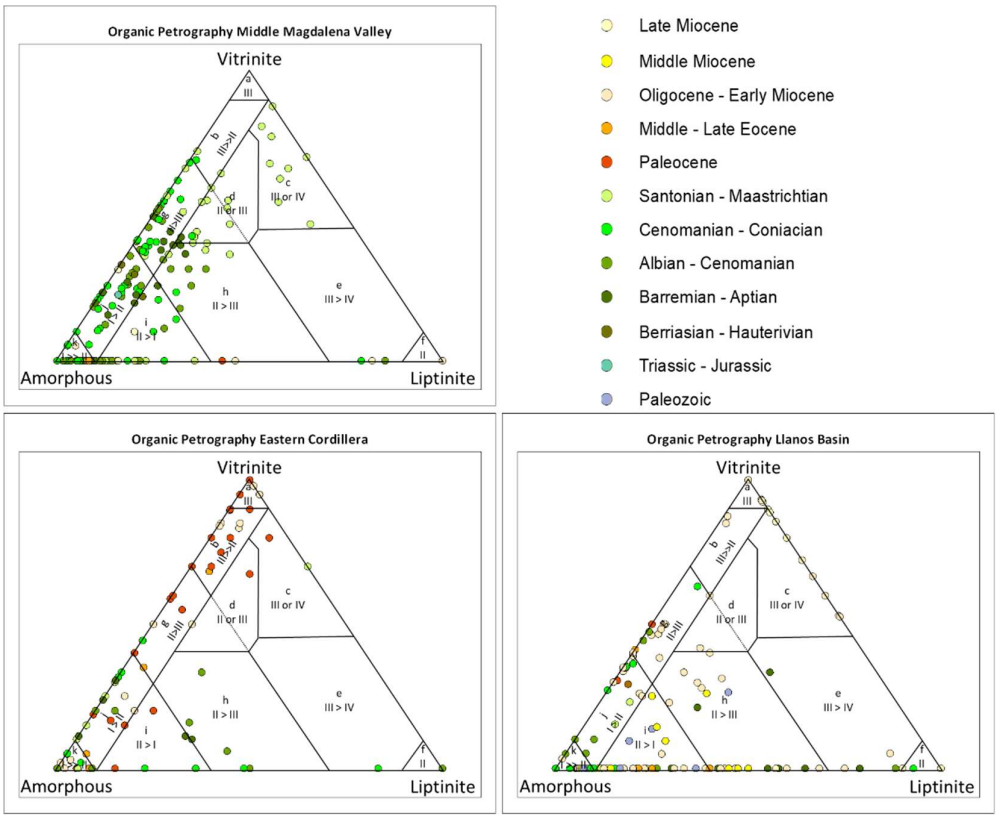

Figure 65. Ternary diagrams on the kerogen type and palynofacies field for environmental interpretation based on maceral composition (Biswas et al., 2020). Palynofacies fields and depositional environment on Table 22.

Table 22. Palynofacies assemblages and environmental interpretation based on maceral composition (Biswas et al., 2020).

Palynofacies field	Depositional Environment	Palynofacies field	Depositional Environment	Palynofacies field	Depositional Environment
a	Highly proximal Shelf or basin	e	Mud-dominated oxic shelf (distal shelf)	i	Distal dysoxic-oxic shelf
b	Marginal dysoxic-anoxic basin	f	Floating spore accumulations transported by wind and water and deposited with vegetable mud	j	Suboxic-anoxic lake environmental conditions
c	Heterolithic oxic Shelf (proximal shelf)	g	Proximal suboxic-anoxic shelf	k	Distal suboxic-anoxic basin
d	Shelf to basin transition	h	Distal dysoxic-anoxic 'shelf		

On the EC, Cretaceous samples indicate deposition under similar conditions to those of the MMV, mainly suboxic-anoxic lakes and distal suboxic-anoxic basin (Figure 65 and Figure 66). The Cenozoic samples show varying conditions, from distal suboxic-

anoxic basin and lakes, to proximal – highly proximal shelf (Figure 65 and Figure 66). The LLA has greater variability, with the Paleozoic samples deposited under conditions of oxic-dysoxic-anoxic shelf to distal suboxic-anoxic basin. The Cretaceous samples were deposited at proximal suboxic-anoxic shelf and lakes and distal basin; while other were deposited at distal dysoxic-anoxic shelf and mud dominated oxic shelf. The Cenozoic samples show contrasting conditions of deposition, from proximal oxic shelf for some Oligocene and Late Miocene samples, to suboxic-dysoxic-anoxic shelf and distal suboxic-anoxic basin for samples of the Eocene, Oligocene and Middle Miocene, showing intermittent conditions that could lead to deposition of intervals of source rocks (Figure 65 and Figure 66).

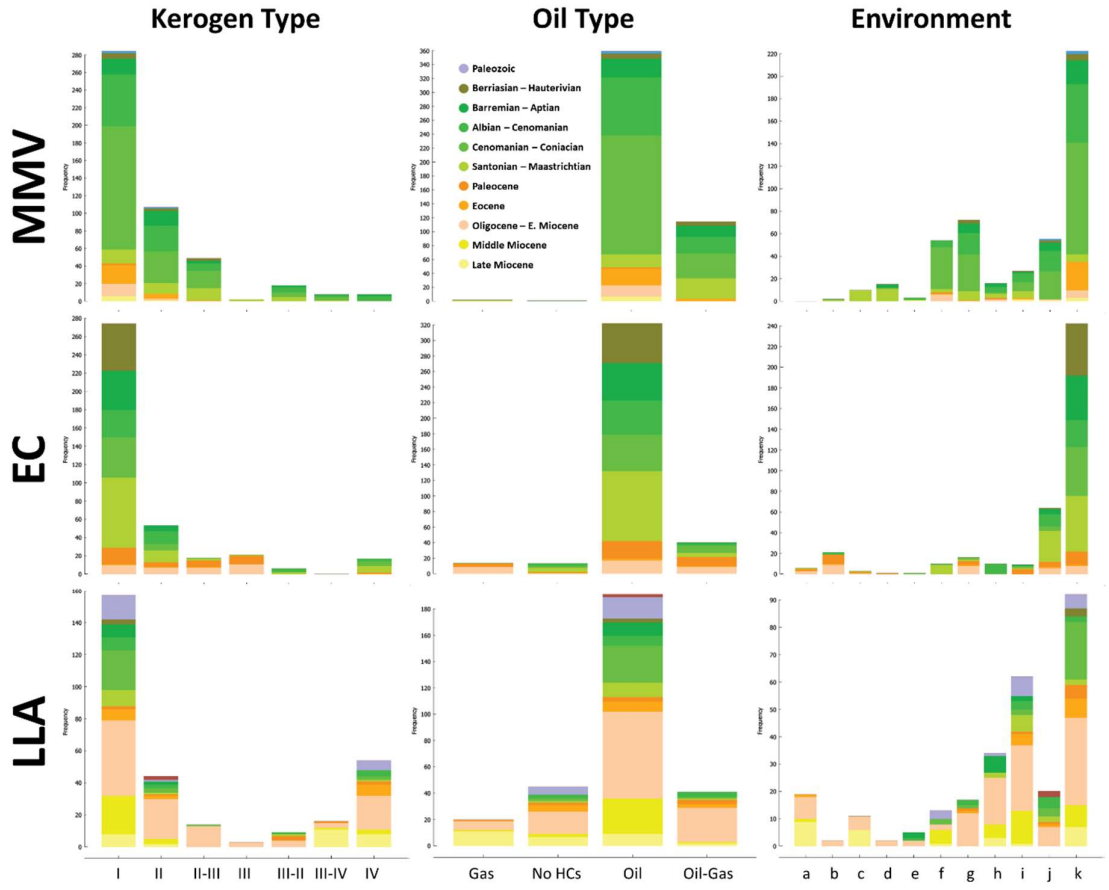

Figure 66. Distribution of samples classified using maceral composition for Kerogen Type, Oil Type and Depositional Environment (Table 22).

4.4.3. Data Mining in Organic Petrography and Geochemistry

The organic geochemistry and maceral composition data for source rocks provide the basis for their characterization. However, the geochemical information can be biased by the effect of thermal maturity, where the quality of the source rocks and the type of kerogen can be mistakenly determined. The maceral composition could be altered in their color, which could difficult identification under high thermal maturation conditions

(i.e. dry gas zone). Exploring the relations between these data sets can improve the knowledge necessary for oil-source rock correlations, such as the depositional environment and kerogen type.

Previously show data mining methodology was simplified to analyze these data sets, due to the lower amount of variables used here (13 in total). Some variables like the peaks of Rock Eval Pyrolysis and Tmax were eliminated due to the reduction in statistical parameters obtained during the PCA and K-Means analysis. The correlations obtained for these variables and the principal components (PC) are shown in Table 23, where the Spearman correlations are low between the parameters from the organic geochemistry and those from the maceral composition. However, each of these groups of parameters are highly correlated to specific PC. The organic geochemistry parameters have higher correlation coefficient with the PC3, while maceral composition parameters have better correlation with PC1, PC2 and PC5 (Table 23).

Table 23. Spearman coefficients comparison. Green color are positive correlations and blue color are negative correlations. Values above 0.3 are bold.

	TOC	HI	OI	GENETIC POTENTIAL	HC TYPE INDEX	MIGRATION INDEX	PRODUCTION INDEX	AMORPHPOUS	LIPTINITE	INERTINITE	VITRINITE	LIPT-AMORPH	TERRESTRIAL_INDEX	PC1	PC2	PC3	PC4	PC5	Silhouette
TOC	1.0																		
HI	0.2	1.0																	
OI	**-0.3**	0.0	1.0																
GENETIC POTENTIAL	**0.6**	**0.7**	-0.1	1.0															
HC TYPE INDEX	**0.4**	**0.7**	**-0.3**	**0.8**	1.0														
MIGRATION INDEX	0.1	**0.5**	0.1	**0.5**	**0.4**	1.0													
PRODUCTION INDEX	0.0	**-0.4**	0.1	**-0.2**	**-0.3**	**0.3**	1.0												
AMORPHPOUS	-0.1	0.0	0.1	0.0	0.0	0.0	0.1	1.0											
LIPTINITE	0.0	0.1	0.1	0.0	0.0	-0.1	-0.2	**-0.5**	1.0										
INERTINITE	0.0	0.0	-0.1	0.0	-0.1	0.1	0.1	**-0.6**	0.0	1.0									
VITRINITE	-0.1	0.0	-0.2	-0.1	-0.1	0.0	-0.1	**-0.5**	0.1	**0.4**	1.0								
LIPT-AMORPH	0.0	0.0	0.1	0.1	0.1	0.0	0.0	**0.8**	0.0	**-0.8**	**-0.7**	1.0							
TERRESTRIAL_INDEX	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	**-0.8**	0.0	**0.8**	**0.7**	**-1.0**	1.0						
PC1	0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.0	**-0.9**	**0.3**	**0.7**	**0.7**	**-0.9**	**0.9**	1.0					
PC2	0.1	-0.2	0.2	-0.1	-0.1	-0.1	0.1	**0.4**	**-0.4**	**-0.1**	**-0.8**	**0.3**	**-0.3**	**-0.4**	1.0				
PC3	**0.4**	**0.7**	-0.2	**0.7**	**0.7**	**0.4**	**-0.4**	-0.2	0.2	0.2	0.0	-0.1	0.1	0.2	-0.2	1.0			
PC4	0.2	0.1	-0.1	0.1	0.2	-0.2	**-0.3**	**-0.3**	**0.5**	-0.1	0.0	0.1	-0.1	0.1	-0.1	0.2	1.0		
PC5	0.2	0.2	-0.1	**0.3**	**0.3**	0.1	-0.1	**0.6**	**-0.6**	**-0.3**	-0.2	**0.3**	**-0.3**	**-0.5**	**0.3**	0.0	0.0	1.0	
Silhouette	0.1	0.1	0.2	0.2	0.2	0.2	0.1	**0.4**	-0.2	**-0.5**	**-0.7**	**0.6**	**-0.6**	**-0.6**	**0.5**	-0.1	0.0	**0.3**	1.0

Despite the lack of numerical correlation, some degree of correlation can be observed between the two groups of variables (Figure 67). The modified Van Krevelen diagram shows that kerogens type III and IV correspond with lower values of HI, and intermediate values of HI correspond with mixtures of kerogens type II and III (Figure 67-A). Higher values in the proportion of Liptinite + Amorphous maceral components correspond with higher values of HI and were classified as kerogens deposited under dysoxic to anoxic conditions (Figure 67-B). The input of Terrestrial maceral components (Vitrinite + Inertinite) show an inverse correlation with HI, and the

classification of depositional environments correspond as expected with these two parameters (Figure 67). Depositional environments from palynofacies field, show good correlation with PC1; and the type of hydrocarbon to be generated also shows good separation between the different environments (Figure 67).

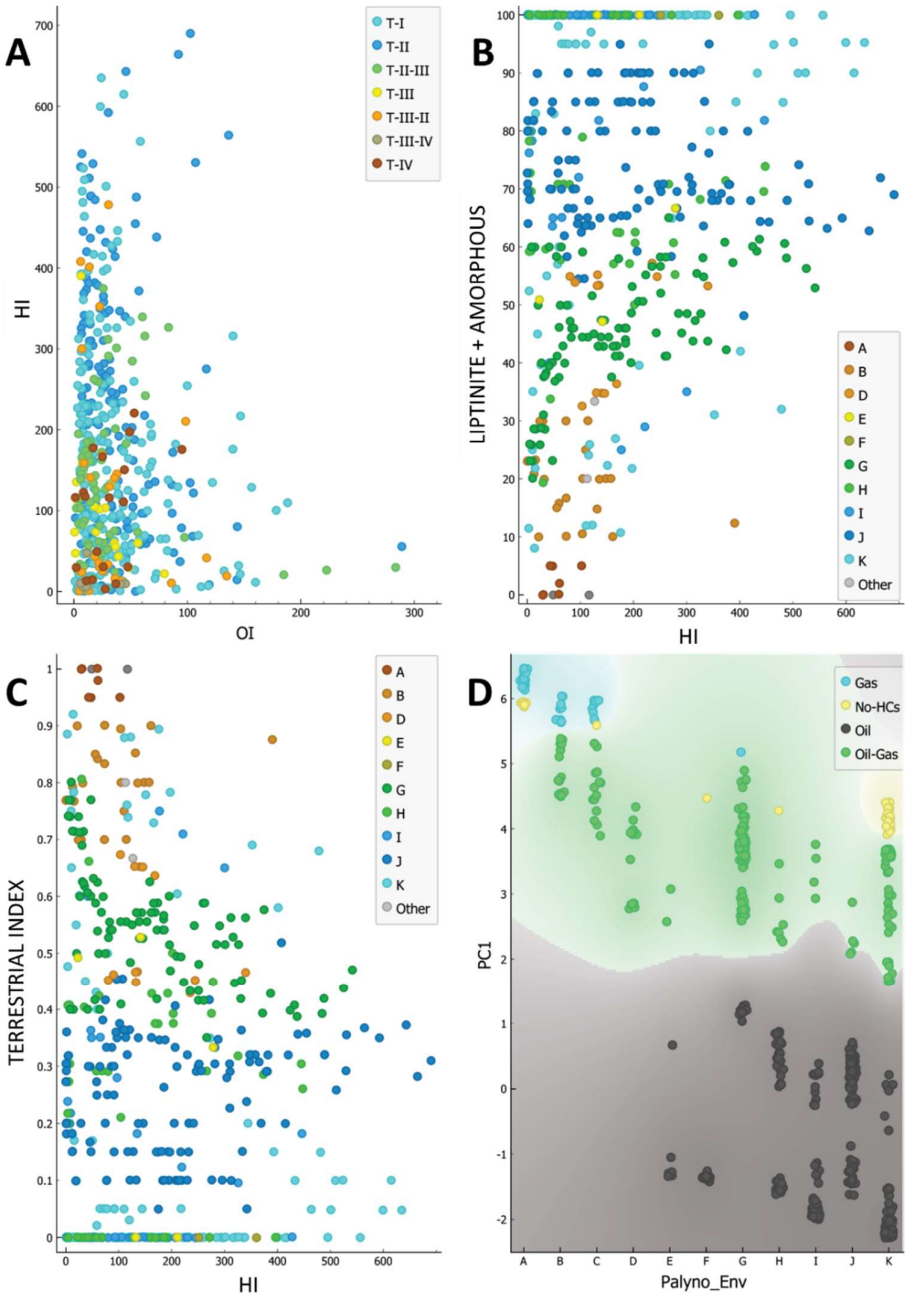

Figure 67. Bivariate diagrams showing correlations between organic geochemistry parameters and maceral components. Colors correspond to categories obtained from ternary diagrams using maceral components (Figure 64 and Figure 65).

4.5. Organic Geochemistry of Crude Oil

Hydrocarbons will get their characteristics from the source rocks and the conditions under which they were generated. These characteristics will be further modified by processes like migration, biodegradation, thermal cracking, refreshing, among others (Rangel et al., 2017). Multiple laboratory analyses are completed on hydrocarbon in order to characterize mostly organic compounds to identify the features that help to understand the origin and processes hydrocarbons have gone through geological history.

A data set of 209 samples located in the area of this study was used to describe the main characteristics of hydrocarbons in the area. The data set included basic laboratory results (API gravity, liquid chromatography fractions, sulfur, nickel, vanadium), isoprenoid ratios from whole oil gas chromatography, stable carbon isotopes of saturate and aromatic fractions, and biomarker ratios from gas chromatography-mass spectrometry (GCMS) data of saturate and aromatic fractions (ANH and GEMS, 2006). The characterization of oil families was made following the procedures of data mining previously described, including 25 parameters (Table 24). The procedure comprised the identification and elimination of outliers (15/209), multivariate correlations comparison using Spearman coefficient, the principal component analysis (PCA), and clustering using K-means algorithm and hierarchical clustering.

The matrix of correlation coefficients indicates which parameters have a relationship to each other, allowing to explore the behavior of the parameters. The results of this correlation matrix show an important geographical component to the distribution of the data, especially for the longitude, which shows important correlations with many variables (Table 24). The east-west change in the variables is expected, given that changes in the geology also occur in the same direction across the different regions of this study. Most of the basic parameters have good correlations to each other, in addition to some of the biomarker parameters (Table 24)., such as the ratio phytane/C18 ratio (Ph/C18), C29/C30 Hopanes ratio (C29/C30Hop), and 28,30-bisnorhopane/hopane (28B/H) (Table 24). Among the biomarkers group, there are some with greater correlations, like 17a-trisnorhopane/18a-trisnorhopane ratio (Ts/Tm), 25-norhopane/hopane ratio (25Nor/Hop), diasterane/sterane ratio (Dias/Ster), and oleanane/C30-hopane ratio (Olea/C30Hop).

The PCA correlations are also shown in Table 24, with PC1 as the one with highest correlations with basic parameters and some biomarkers, while PC2 and PC3 have better correlations with biomarkers. PC4 and PC5 have correlations with fewer parameters in comparison with the previous principal components.

Table 24. Spearman coefficients comparison. Green color are positive correlations and blue color are negative correlations. Values above 0.3 are bold.

	Latitude	Longitude	Depth	API	Aromatic	Saturate	Sat/Arom	NSO	Sulfur	Ni	V	V/V+Ni	Ph/C18	Pr/C17	Pr/Ph	C13aro	C13Sat	Ts/Tm	28B/H	25Nor/Hop	35/34-Hop	ABB/ABB+AAA	C29/C30Hop	Dias/Ster	Gam/C30Hop	Olea/C30Hop	PC1	PC2	PC3	PC4	PC5	Silhouette
Latitude	1.0																															
Longitude	-0.3	1.0																														
Depth	-0.2	0.2	1.0																													
API	-0.1	0.3	0.3	1.0																												
Aromatic	0.1	-0.2	-0.1	-0.3	1.0																											
Saturate	-0.1	0.4	0.2	0.6	-0.6	1.0																										
Sat/Arom	-0.1	0.2	0.2	0.5	-0.8	0.8	1.0																									
NSO	0.1	-0.3	-0.2	-0.5	0.2	-0.8	-0.6	1.0																								
Sulfur	0.3	-0.5	-0.3	-0.7	0.3	-0.7	-0.5	0.6	1.0																							
Ni	0.0	0.0	-0.2	-0.3	0.1	-0.5	-0.3	0.5	0.4	1.0																						
V	0.1	-0.4	-0.1	-0.6	0.2	-0.6	-0.5	0.6	0.6	0.7	1.0																					
V/V+Ni	0.1	-0.3	-0.1	-0.5	0.3	-0.5	-0.4	0.4	0.5	0.1	0.7	1.0																				
Ph/C18	0.6	-0.5	-0.4	-0.4	0.3	-0.5	-0.3	0.4	0.5	0.2	0.3	0.2	1.0																			
Pr/C17	0.1	-0.3	-0.3	-0.3	0.1	-0.3	-0.2	0.3	0.2	0.2	0.2	0.1	0.6	1.0																		
Pr/Ph	-0.7	0.4	0.3	0.2	-0.1	0.3	0.2	-0.2	-0.5	-0.1	-0.2	-0.1	-0.8	-0.1	1.0																	
C13aro	-0.4	0.1	0.2	0.0	0.0	0.3	0.2	-0.4	-0.3	-0.3	-0.2	-0.1	-0.4	-0.2	0.5	1.0																
C13Sat	-0.3	-0.1	0.1	-0.1	0.1	0.0	0.0	-0.1	-0.1	-0.2	0.0	0.1	-0.2	-0.1	0.3	0.8	1.0															
Ts/Tm	-0.3	0.3	0.1	0.2	0.0	0.2	0.2	-0.2	-0.4	-0.2	-0.4	-0.3	-0.2	0.0	0.2	0.4	0.3	1.0														
28B/H	-0.1	0.1	0.0	-0.4	0.2	-0.5	-0.4	0.5	0.3	0.5	0.4	0.3	0.1	0.1	-0.1	-0.2	-0.1	0.0	1.0													
25Nor/Hop	-0.6	0.0	0.2	-0.2	0.1	-0.1	0.0	0.1	0.0	0.1	0.1	0.0	-0.2	0.0	0.4	0.5	0.5	0.3	0.4	1.0												
35/34-Hop	-0.2	0.2	0.0	-0.2	0.0	-0.1	-0.1	0.2	0.3	0.3	0.2	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.3	0.4	0.2	1.0											
ABB/ABB+AAA	0.2	-0.5	0.0	-0.2	0.1	-0.2	-0.1	0.0	0.2	-0.1	0.2	0.2	0.3	0.1	-0.3	0.1	0.1	0.0	-0.1	0.0	-0.2	1.0										
C29/C30Hop	0.0	-0.4	0.0	-0.5	0.1	-0.4	-0.2	0.3	0.4	0.2	0.5	0.5	0.2	0.1	0.0	0.1	0.1	-0.5	0.3	0.3	-0.1	0.3	1.0									
Dias/Ster	-0.6	0.3	0.2	0.1	-0.1	0.3	0.2	-0.2	-0.4	-0.3	-0.3	-0.1	-0.5	-0.1	0.6	0.7	0.5	0.5	-0.2	0.4	0.1	-0.1	-0.1	1.0								
Gam/C30Hop	-0.1	0.3	0.1	0.1	-0.2	0.1	0.2	0.0	0.0	0.2	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.3	0.3	0.0	0.6	-0.2	-0.2	-0.1	1.0							
Olea/C30Hop	-0.7	0.6	0.3	0.2	-0.3	0.3	0.3	-0.2	-0.4	0.0	-0.2	-0.2	-0.6	-0.1	0.7	0.3	0.2	0.4	0.1	0.5	0.2	-0.2	-0.1	0.6	0.2	1.0						
PC1	0.2	-0.5	-0.3	-0.8	0.4	-0.9	-0.7	0.8	0.9	0.6	0.8	0.6	0.6	0.3	-0.4	-0.4	-0.1	-0.4	0.5	0.0	0.2	0.2	0.5	-0.4	0.0	-0.4	1.0					
PC2	-0.7	0.1	0.2	-0.3	0.2	-0.1	-0.2	0.0	-0.1	0.0	0.1	0.1	-0.4	-0.1	0.6	0.7	0.6	0.2	0.2	0.8	0.1	0.0	0.4	0.6	-0.1	0.5	0.0	1.0				
PC3	-0.4	0.6	0.1	0.1	-0.2	0.1	0.1	0.1	-0.1	0.4	0.0	-0.2	-0.4	-0.2	0.3	-0.2	-0.3	0.2	0.5	0.2	0.6	-0.4	-0.2	0.1	0.5	0.5	-0.1	0.1	1.0			
PC4	0.0	0.3	0.0	0.0	0.5	-0.2	-0.4	0.0	-0.2	0.0	-0.3	-0.3	0.1	0.2	0.0	0.1	0.1	0.5	0.1	0.0	0.2	-0.1	-0.5	0.2	0.2	0.0	-0.1	0.0	0.1	1.0		
PC5	-0.2	0.0	-0.1	-0.1	-0.6	0.3	0.5	-0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.0	0.0	0.0	-0.1	0.1	0.2	0.0	0.0	0.3	0.0	0.2	0.2	0.0	0.1	0.2	-0.5	1.0	
Silhouette	0.3	-0.1	0.0	0.1	0.0	0.0	-0.1	0.0	0.1	0.1	0.0	-0.1	0.1	-0.1	-0.2	-0.3	-0.3	-0.4	-0.1	-0.3	-0.2	-0.2	0.0	-0.4	-0.1	-0.3	0.1	-0.3	-0.1	-0.1	-0.2	1.0

The K-means results, showed the highest silhouette score for 8 clusters, followed by 4 clusters. The last option was preferred due to the reduction in complexity and the fact that additional clusters had between 2 and 5 samples. Including coordinates as parameters within the algorithms has minor effect on the clusters, and some of them are spread across the basins, meaning that their origin/processes could be similar despite the distance that separates them (Figure 68). The characteristics of the possible source rocks for the hydrocarbons were obtained using bivariate diagrams, where samples were categorized according to the cluster and basin to illustrate their features (Figure 68, Figure 69, Figure 70 and Table 25).

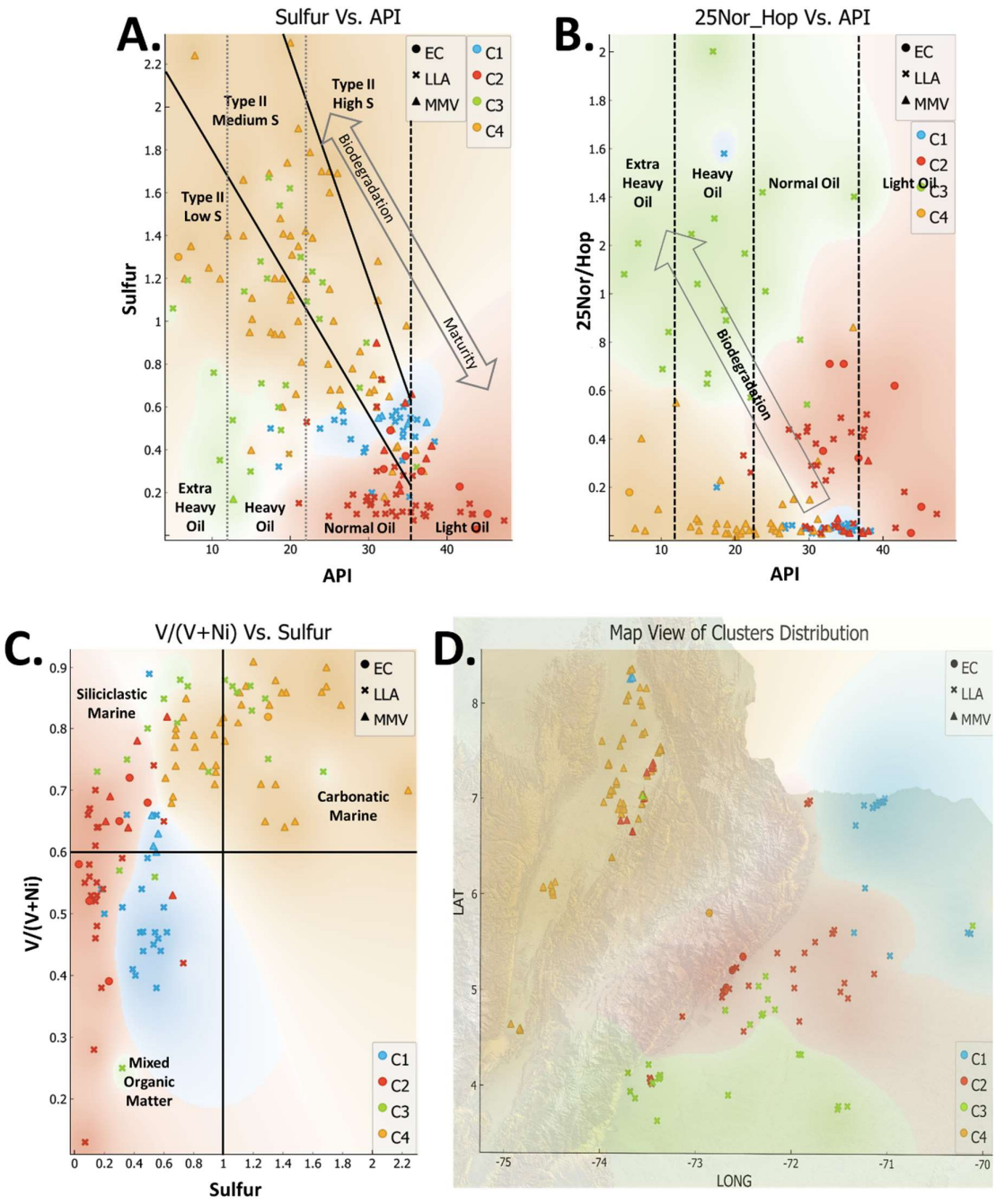

Figure 68. A-B: Bivariate diagrams showing the interpretation of the basic parameters of the hydrocarbons studied here, showing their quality and effect of biodegradation. C: Interpretation of the source rock characteristics from the V/(V+Ni) Vs. Sulfur diagram. D: Map view showing the distribution of the clusters.

The quality of the hydrocarbons is evaluated using their API and sulfur content, and in combination with the 25Nor/Hop, the advance of their biodegradation is assessed too (Figure 68). Here we can see that the hydrocarbon of the clusters C1 and C2 have lower sulfur content and higher API, while samples from clusters C3 and C4 are more

dispersed with higher sulfur and lower API (Figure 68 A-B). The lower quality of these two clusters can be explained by higher biodegradation, but also the natural effect of carbonatic source rocks that will imprint these characteristics from the origin, and biodegradation will increase them. The diagram of V/(V+Ni) Vs sulfur presents an interpretation of the characteristics of the source rocks, showing that clusters C3 and C4 have a marine carbonatic source rock, while clusters C1 and C2 where originated from siliciclastic rocks with marine to marine-terrestrial organic matter (Figure 68 C-D). The geographic distribution shows that clusters C1, C2 and C3 are mostly present at the LLA and Eastern Foothill of the EC, while cluster C4 is restricted to the MMV, and a couple of samples on the central region of the EC. Clusters C1-C2-C3 have some samples in the northern of the MMV (Figure 68 D).

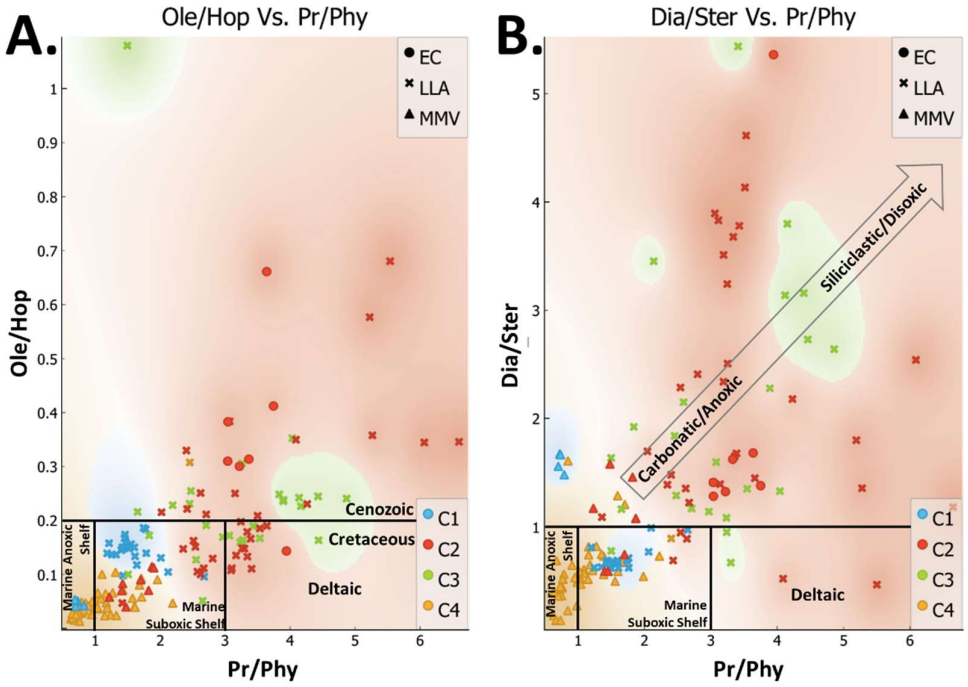

Figure 69. A: Oleanane/Hopane Vs Pristane/Phytane diagram showing the possible age of the source rocks and redox conditions. B: Diasteranes/Steranes Vs Pristane/Phytane showing the trend of composition of the sourse rocks and redox conditions.

The Ole/Hop ratio indicates that some samples of the LLA and EC belonging to C2-C3 were derived from Cenozoic source rocks, while for the majority of the samples, the source rock should have been from Cretaceous age (Figure 69 A). According to the Pr/Phy ratio, samples of the C1 and C4 where originated from source rocks deposited at anoxic-suboxic marine shelf, as well as some samples from C2 (Figure 69 A-B). The

Dia/Ster ratio gives partially contrary indications about the lithology of the source rock for cluster C1, suggesting a carbonatic anoxic source rock, while the Figure 68-C indicates siliciclastic origin with mixed marine-terrestrial organic matter (Figure 69 B).

Furthermore, the Gam/Hop ratio indicates high salinity conditions for the source rocks of the C1 samples (Figure 70-A). The origin for most of the samples of clusters C2, C3 and C4 is similar to the interpretation obtained from Figure 69 B. The Figure 70-B shows important maturity increase for some of the samples from clusters C1-C2-C3 located at LLA and EC with input of terrestrial organic matter. On the other hand, a group of samples from the MMV and clusters C1-C4 indicate algal organic matter input, and increasing maturity also.

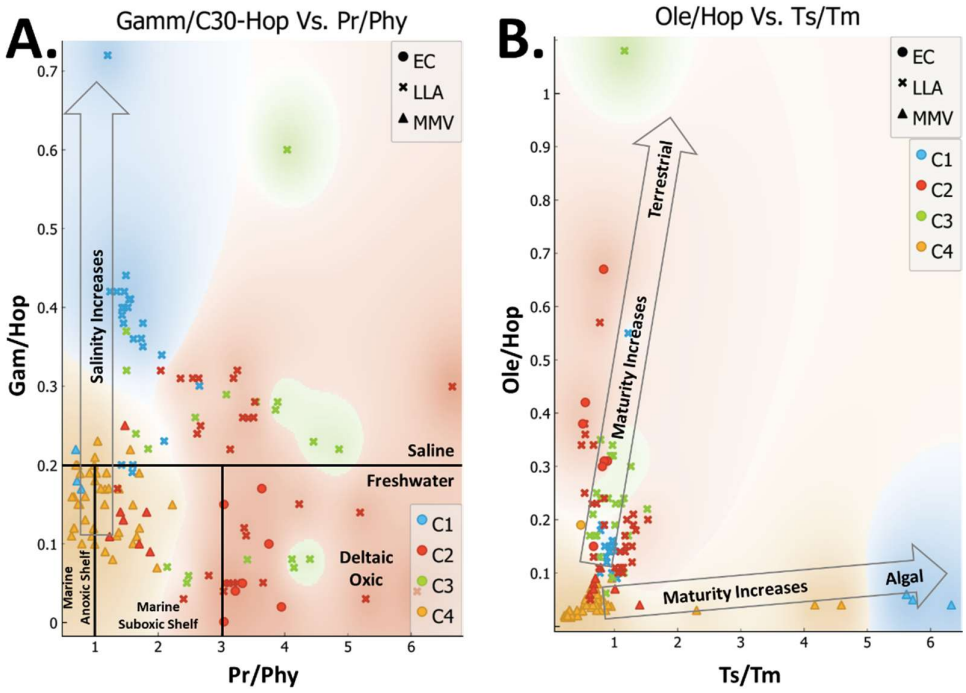

Figure 70. A: Gammacerane/Hopane Vs Pristane/Phytane diagram showing the salinity conditions and redox of the source rocks. B: Oleanane/Hopane Vs Ts/Tm showing the input of terrestrial and algal organic matter.

Table 25. Principal characteristics for each cluster at the basin where they were identified

CLUSTER	MMV	EC	LLA
C1	High API Low Sulfur Mixed to marine siliciclastic source rock Algal organic matter Increasing maturity of source rock		High API Low Sulfur Mixed to marine siliciclastic source rock Terrestrial organic matter Increasing maturity of source rock Salinity increase
C2	High API Low Sulfur Mixed to marine siliciclastic source rock	High API Low Sulfur Mixed to marine siliciclastic source rock Cenozoic source rock	High API Low Sulfur Mixed to marine siliciclastic source rock Cenozoic source rock Terrestrial organic matter Increasing maturity of source rock Variable Salinity
C3	Low API Low Sulfur		Low to High API High to Low Sulfur High Biodegradation Cenozoic source rock Carbonatic to siliciclastic marine source rock Terrestrial organic matter Increasing maturity of source rock Variable Salinity
C4	Low to High API High to Low Sulfur Carbonatic to siliciclastic marine source rock Algal organic matter Increasing maturity of source rock Low Salinity	Low to High API High to Low Sulfur Carbonatic to siliciclastic marine source rock	

5. CONCLUSIONS

The elemental composition study of the selected samples, and those gathered in nearby regions, allowed to classify them through multivariate interpretation and data mining tools according to their characteristics inherited from lithology, provenance, paleoclimatic and paleoredox conditions during their deposition. The data mining tools allowed to improve the classification into fewer clusters with the most meaningful parameters, reducing them into principal components for easier visualization. even though the reduction of variables and clustering, it is clear that lithological variability within each stratigraphic interval, and geographically at the same time, is the most prevalent rule to follow. Each cluster incorporates a series of samples from different ages and locations (Figure 71), but with common characteristics, which are summarized in Table 26, and represented as diagrams of probability distributions for each stratigraphic interval and location.

Table 26. Main characteristic of the clusters of samples based on run B (Figure 32). Distribution of the clusters by stratigraphic interval and for each location in Figure 71

Characteristic	Cluster-1	Cluster-2	Cluster-3
Lithology	Muddy Marlstone to Calcareous Marlstone	Mudstone to Muddy Marlstone	Mudstone to Sublithic Sandstone
Provenance	Felsic	Felsic	Felsic
Paleoclimate	Non-Weathering Arid	Weak to Strong Weathering Arid	Strong Weathering Humid
Paleoredox	Disoxic to Anoxic	Suboxic to Disoxic	Particulate Shuttle

The isotopic characterization of samples showed that the composition changes geographically and stratigraphically for each of the systems (Pb-Pb, Rb-Sr and Nd-Nd). The Pb-Pb indicated that the isotopic composition of the studied rocks was derived from the Upper Crust, and for one sample (Pyrite vein from Paleozoic age), it was derived from the orogen system. The pyrite veins and hosting rocks show clear correlation with this system, but different stratigraphic intervals can have similar composition, reducing the effectiveness to identify the source of some hydrothermal materials. The $^{87}Sr/^{86}Sr$ isotopic ratio and Rb/Sr composition ratio, showed little

efficiency at differentiate some of the samples, or to identify the source of any hydrothermal material, given that samples from different formation can show similar compositions, and samples from one formation, can have widely variable compositions. However, when combined with $^{143}Nd/^{144}Nd$ and ε_{Nd}, it was useful to differentiate the source of each area, to the north with composition derived from the Lower Continental Crust (Figure 44), while to the south, the composition was similar to that of the Upper Continental Crust (Figure 45).

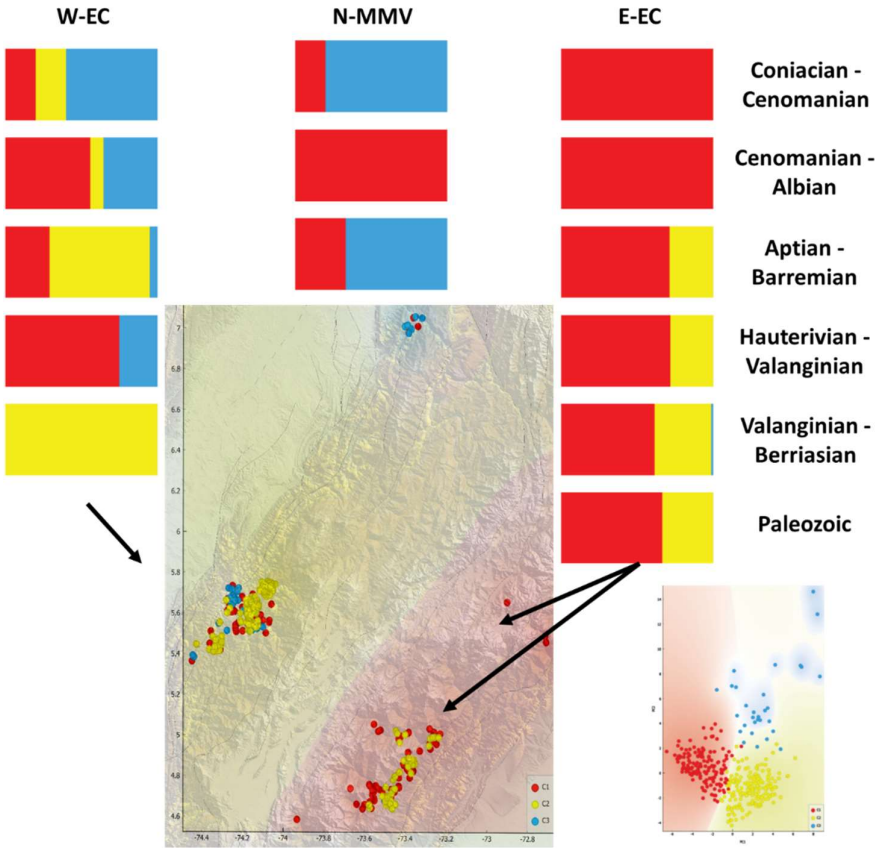

Figure 71. Distribution of clusters for each location and stratigraphic interval. Features for each cluster in Table 26

The provenance analysis of the hydrocarbon reservoirs and source rocks was accompanied with detrital zircon U/Pb dating of one sample from the Mugrosa Fm. in the MMV, along with published data from different stratigraphic intervals, and multiple dating from basement rocks at each tectonic domain from Colombia (Gomez et al., 2015a). MDS analysis from each basement domain and for different samples, showed a strong correlation of the Cretaceous rocks with the Craton, which is consistent with the paleogeographic evolution and facies trend of proximal facies to the east, and distal

facies to the west. During the Cenozoic, these same rocks seem to be the source for the new developing basins and new areas to the south of Colombia began to contribute to the sediment composition, as well as some areas of the Central Cordillera.

The organic geochemical characteristics for Paleozoic, Cretaceous and Cenozoic intervals, showed how the quantity and quality of the organic matter of the Cretaceous intervals, are reduced from west to east, with the most potential in the MMV, followed by the Eastern Cordillera, and lower potential in the Llanos Basin. The most common kerogen is type II, followed by some intervals with Type I and III. The thermal maturity is the third component to fully characterize the potential of any source rock, so that an interesting interval has fully expelled their potential hydrocarbons, or it is actually an active source rock. This parameter is highest in the Eastern Cordillera, as the area where the main depocenter developed during the Cretaceous. The Magdalena Valley also reached good maturity values for a good prospective area. The Llanos basin has lower thermal maturity and only small regions with good source rocks are within the oil generation window.

The hydrocarbons present in the Colombian basins can be divided in at least four groups, each one with dissimilar characteristics, but that partially overlap in some others, which makes difficult to discern them by simple inspection. The use of data mining tools allowed to classify the hydrocarbon according to their salinity, redox conditions, type of organic matter, and basic parameters that have been traditionally used as a first proxy for crude oil families.

Using categorical characteristics assigned to rocks and crude oil samples based on their environment, redox conditions, lithology, and possible age of source rock for crude oil samples; a proposed correlation obtained from Principal Component Analysis (PCA) and K-means clustering is presented here (Figure 72). This correlation shows that across basins, similar hydrocarbons are present, but they correlate in each basin with different source rocks intervals, and even for the LLA, crude oil samples from clusters C2 and C4 do not correlate with local source rocks, but most likely can be correlated to rocks of the EC (Figure 72-A). For the MMV-EC samples, the environment is marine to marine proximal with anoxic to suboxic conditions; while for the LLA-EC marine proximal with anoxic, suboxic conditions, and deltaic with oxic conditions are the source rocks environment (Figure 72-B and C). The correlations show that

Cenozoic samples have similar parameters, however, as shown in Figure 53, most of the Cenozoic samples are undermature, with local pods in the eastern flank of the EC and proximal LLA basins. The variability found at every formation, does not allow to correlate any crude oil family to a single stratigraphic interval, but most likely, each crude oil family could be generated from different intervals with similar characteristics across the basins.

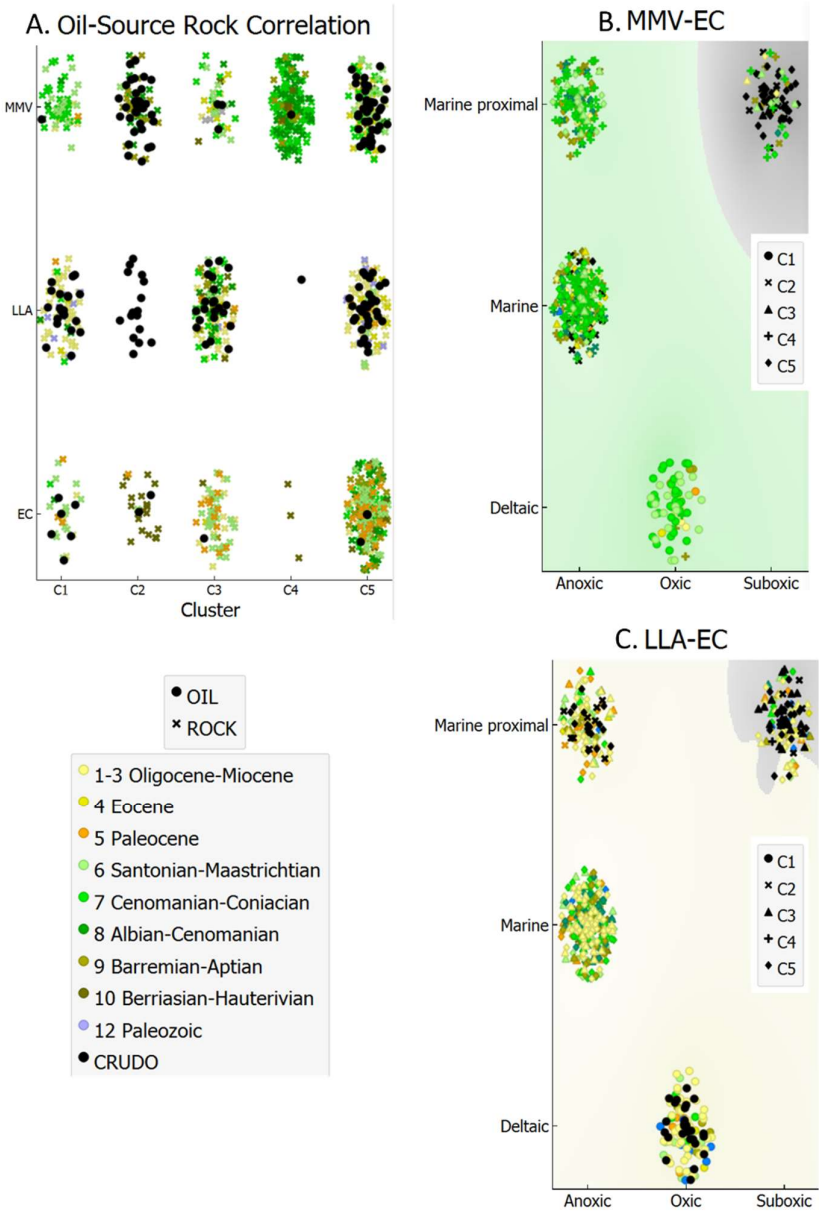

Figure 72. A. Clustering results for oil to source rock correlation using categorical variables of their environment, redox conditions, lithology and age of source rock for oil samples based on the Oleanane/Hopane ratio. B. and C. show the correlation results according to the environment and redox conditions.

The distribution of the source rock and crude oil samples for each cluster can be observed in Figure 73, Figure 74 y Figure 75.

Clusters 1, 2, 3 and 5 contain most of the crude oil samples. Cluster 1 includes most of the crude oil samples located in the central region of the LLA and one sample on the northern of the MMV (Figure 73). The rock samples of this cluster were deposited in deltaic environment and under oxic conditions, while the crude oil samples were originated from source rocks with these characteristics, and, their Oleanane/Hopane ratio indicates that the source rocks of the samples could be from Cenozoic age (Figure 69).

Cluster 2 includes crude oil samples from the northern MMV, the central region of the EC and southern LLA. These crude oil samples indicate an origin from source rocks deposited in marine and proximal marine environments and under suboxic to anoxic conditions (Figure 72). The associated source rocks samples include Berirasian to Cenomanian from the EC and MMV. None of the samples of the LLA were included in this cluster, mostly due to the indicators of calcareous source rocks, which are not present in this region (Figure 73).

Cluster 3 includes crude oil samples from the northern LLA and MMV, as well as central LLA (Figure 74). Source rocks samples incorporated in this cluster mostly present ages from Cenomanian to Maastrichtian, and some Paleocene to Eocene. Oligocene to Miocene samples of the LLA in this cluster cannot be accounted as probable source rocks, since samples from this interval are immature (Figure 53 and Figure 74).

Cluster 4 only include two crude oil samples, one in the northern MMV and another in the southern LLA (Figure 74). Source rock samples are mostly Albian to Coniacian age, with minor samples from Berriasian to Hauterivian age. These samples were deposited at marine to proximal marine environments and under anoxic conditions (Figure 72 and Figure 74).

Cluster 5 contains crude oil samples from the northern MMV and central LLA. Samples from this cluster were deposited in marine proximal environment under anoxic conditions (Figure 72 and Figure 75). Rock samples from this cluster are mostly Albian to Maastrichtian, with some Cenozoic and Paleozoic samples from the LLA.

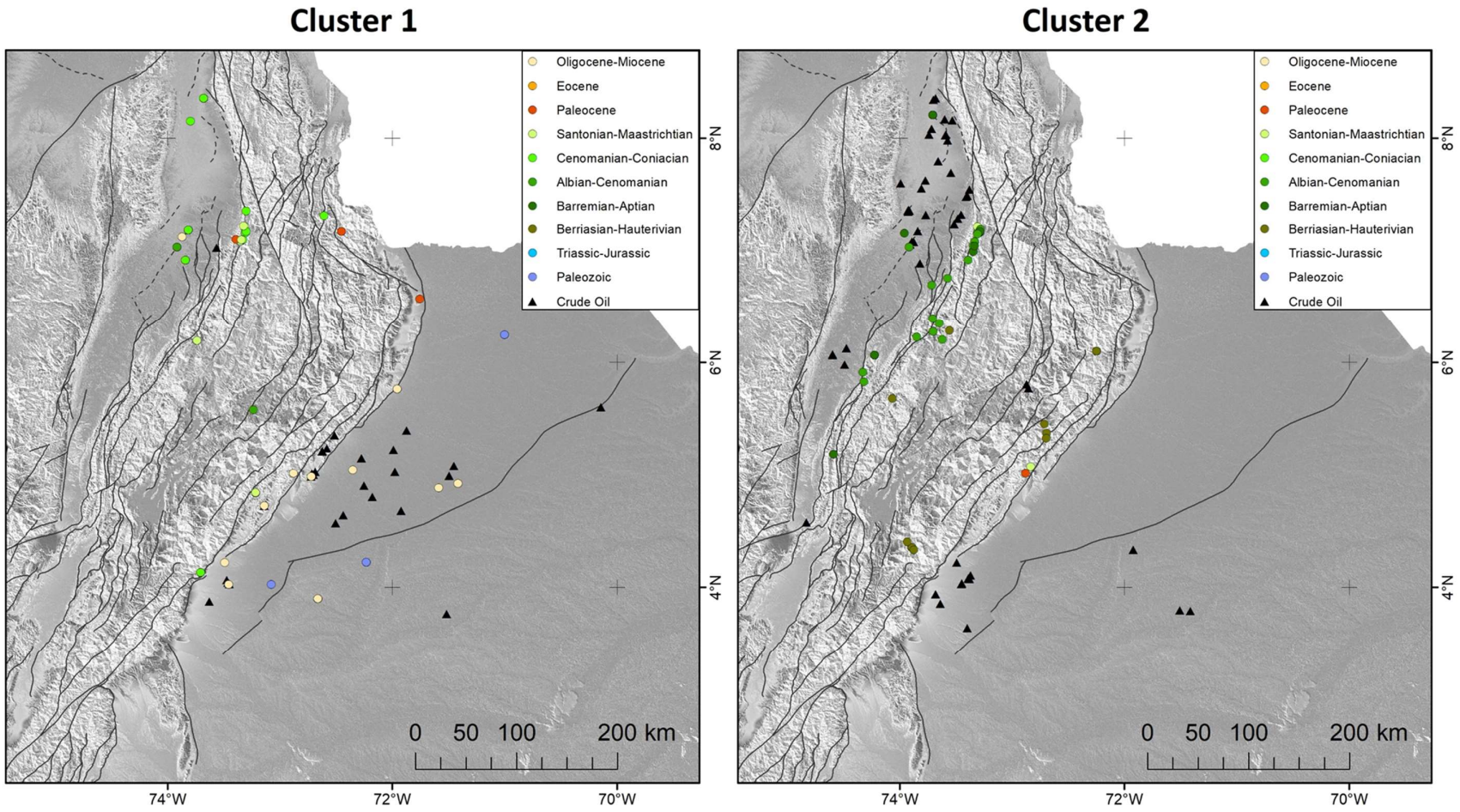

Figure 73. Distribution of source rock and crude oil samples belonging to Cluster 1 and 2.

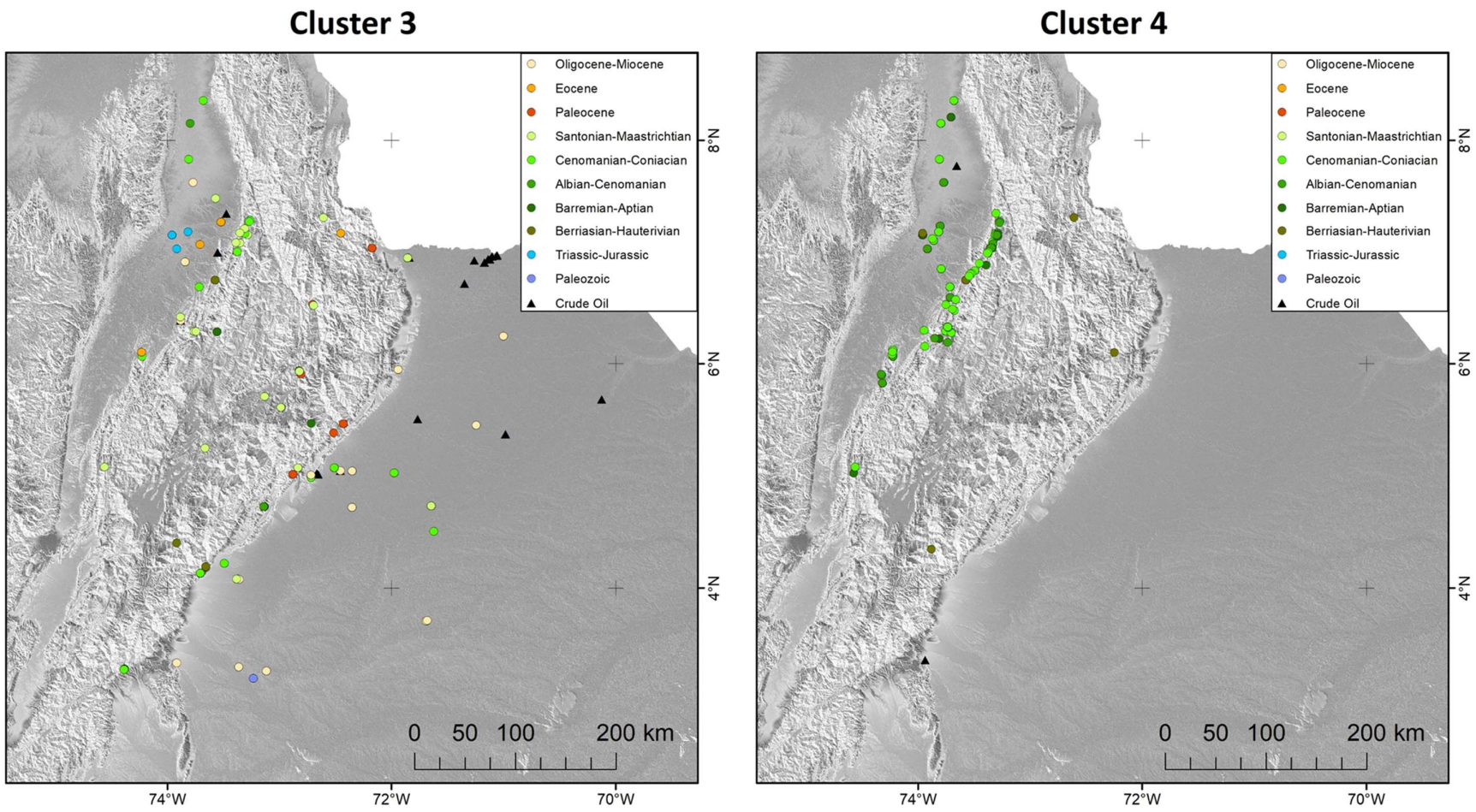

Figure 74. Distribution of source rock and crude oil samples belonging to Cluster 3 and 4.

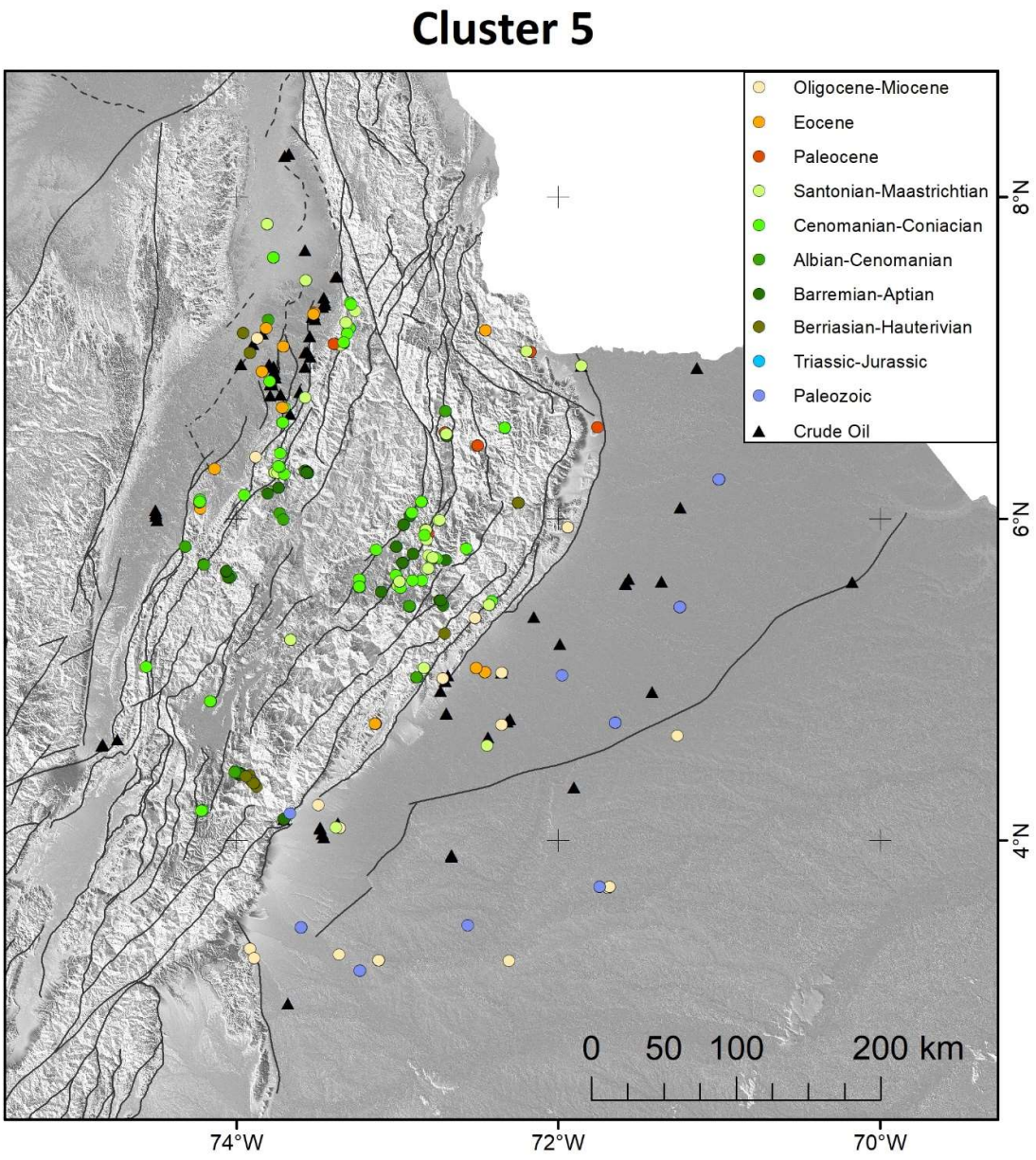

Figure 75. Distribution of source rock and crude oil samples belonging to Cluster 5.

6. REFERENCES

ANH, GEMS, 2006. Atlas Geoquimico de Colombia - 2006. Bogota.

Baiyegunhi, C., Liu, K., Gwavava, O., 2017. Geochemistry of sandstones and shales from the Ecca Group, Karoo Supergroup, in the Eastern Cape Province of South Africa: Implications for provenance, weathering and tectonic setting. Open Geosci. 9, 340–360. https://doi.org/10.1515/GEO-2017-0028/ASSET/GRAPHIC/J_GEO-2017-0028_FIG_020.JPG

Barham, A., Ismail, M.S., Hermana, M., 2021. Statistical Modelling for the Source Rock Parameters of the Montney Formation, NE British Columbia, Canada. Appl. Sci. 2022, Vol. 12, Page 267 12, 267. https://doi.org/10.3390/APP12010267

Barrero, D., Pardo-Trujillo, A., Vargas, C., Martinez, J.F., 2007. Colombian Sedimentary Basins. ANH Colombia, Bogota.

Bhatia, M.R., Crook, K.A.W., 1986. Trace element characteristics of graywackes and tectonic setting discrimination of sedimentary basins. Contrib. to Mineral. Petrol. 92, 181–193. https://doi.org/10.1007/BF00375292

Biswas, S., Varma, A.K., Kumar, M., Mani, D., Saxena, V.K., Mishra, V., 2020. Influence of geochemical, organo-petrographical and palynofacies assemblages on hydrocarbon generation: A study from upper Oligocene coal and shale of the Makum Coal Basin, Assam, India. Mar. Pet. Geol. 114, 104206. https://doi.org/10.1016/J.MARPETGEO.2019.104206

Branquet, Y., Cheilletz, A., Cobbold, P.R., Baby, P., Laumonier, B., Giuliani, G., 2002. Andean deformation and rift inversion, eastern edge of Cordillera Oriental (Guateque–Medina area), Colombia. J. South Am. Earth Sci. 15, 391–407.

Bürgl, H., 1967. The orogenesis in the andean system of colombia. Tectonophysics 4, 429–443.

Bürgl, H., 1964. El Jura-Triásico de Colombia. Boletín Geológico Serv. Geológico Nac. 12, 5–31.

Bürgl, H., 1960. El Jurásico e Infracretáceo del Río Batá, Boyacá: Informe 1319. Boletín Geológico Serv. Geológico Nac. 1, 169–211.

Cediel, F., Shaw, R., Cáceres, C., 2003. Tectonic Assembly of the Northern Andean Block. eds., Circum-Gulf Mex. Caribb. Hydrocarb. habitats, basin Form. plate tectonics 79, 815–848.

Chen, Z., Jiang, C., Lavoie, D., Reyes, J., 2016. Model-assisted Rock-Eval data interpretation for source rock evaluation: Examples from producing and potential shale gas resource plays. Int. J. Coal Geol. 165, 290–302. https://doi.org/10.1016/j.coal.2016.08.026

Chen, Z., Liu, C., X., J., 2018. Forward and inverse modeling of kerogen generation kinetics based on routine Rock-Eval pyrolysis with application examples from Canadian sedimentary basins and elsewhere, in: Geoconvention. Calgary.

Chen, Z., Liu, X., Guo, Q., Jiang, C., Mort, A., 2017a. Inversion of source rock hydrocarbon generation kinetics from Rock-Eval data. Fuel 194, 91–101.

Chen, Z., Liu, X., Jiang, C., 2017b. Quick Evaluation of Source Rock Kerogen Kinetics Using Hydrocarbon Pyrograms from Regular Rock-Eval Analysis. Energy and Fuels 31, 1832–1841. https://doi.org/10.1021/acs.energyfuels.6b01569

Cooper, M.A., Addison, F.T., Alvarez, R., Hayward, A.B., Howe, S., Pulham, A.J., Taborda, A., 1995. Basin Development and Tectonic History of the Llanos Basin, Colombia. Pet. Basins South Am. https://doi.org/10.1306/M62593C35

Cortes, J., Rincon, J., Jaramillo, J.M., Philp, P., Allen, J., 2010. Biomarkers and compound-specific stable carbon isotope of n-alkanes in crude oils from eastern Llanos Basin, Colombia. J. South Am. Earth Sci. 29, 198–213. https://doi.org/10.1016/j.jsames.2009.03.010.

Dickinson, W.R., Gehrels, G.E., 2009. Use of U-Pb ages of detrital zircons to infer maximum depositional ages of strata: A test against a Colorado Plateau Mesozoic database. Earth Planet. Sci. Lett. 288, 115–125. https://doi.org/10.1016/j.epsl.2009.09.013

Dickinson, W.R., Gehrels, G.E., 2003. U–Pb ages of detrital zircons from Permian and

Jurassic eolian sandstones of the Colorado Plateau, USA: paleogeographic implications. Sediment. Geol. 163, 29–66. https://doi.org/10.1016/S0037-0738(03)00158-1

Estremina, E., 2017. Caracterização de fluidos hidrotermais e suas fontes em bacias sedimentares com potencial petrolifero: Aplicação inovadora de análises isotópicas na exploração de petróleo e gás. Universidad de São Paulo.

Etayo, F., Renzoni, G., Barrero, D., 1976. Contornos sucesivos del mar Cretácico en Colombia, in: Primer Congreso Colombiano de Geología. pp. 217–252.

Etayo, F., Rodriguez, G., 1985. Edad de la Formación Los Santos, in: Etayo, F., Laverde, F. (Eds.), Proyecto Crétacico, Contribuciones. Bogota, p. 13.

Fabre, A., 1987. Tectonique et génération d'hydrocarbures : un modèle de l'évolution de la cordillère orientale de Colombie et du bassin des Llanos pendant le Crétacé et le Tertiaire. Arch. des Sci.

Fabre, A., 1985. Dinámica de la sedimentación Cretácica en la región de la Sierra Nevada del Cocuy (Cordillera Oriental de Colombia), in: Etayo, F., Laverde, F. (Eds.), Proyecto Crétacico, Contribuciones. Bogota, p. 20.

Floyd, P.A., Leveridge, B.E., 1987. Tectonic environment of the Devonian Gramscatho basin, south Cornwall: framework mode and geochemical evidence from turbiditic sandstones (England). J. - Geol. Soc. 144, 531–542. https://doi.org/10.1144/GSJGS.144.4.0531

Garcia, D., 2008. Estudo dos Sistemas Petrolíferos no Setor Central da Bacia dos "Llanos Orientales", Colômbia. Um Modelo para Explicar as Mudanças na Qualidade do Petróleo. Universidade Federal do Rio de Janeiro.

Garcia, D., Vaz dos Santos Neto, E., Penteado, H.L. d. B., 2015. Controls on petroleum composition in the Llanos basin, Colombia: Implications for exploration. Am. Assoc. Pet. Geol. Bull. 99, 1503–1535. https://doi.org/10.1306/10231411111

Gehrels, G.E., 2000. Introduction to detrital zircon studies of Paleozoic and Triassic strata in western Nevada and northern California. Spec. Pap. Geol. Soc. Am. 347, 1–17. https://doi.org/10.1130/0-8137-2347-7.1

Gómez, E., Jordan, T.E., Allmendinger, R.W., Cardozo, N., 2005. Development of the Colombian foreland-basin system as a consequence of diachronous exhumation of the northern Andes. Geol. Soc. Am. Bull. 117, 1272. https://doi.org/10.1130/B25456.1

Gomez, J., Montes, N., Alcarcel, F., Ceballos, J., 2015a. Catálogo de dataciones radiométricas de Colombia en ArcGIS y Google Earth, in: Gomez, J., Almanza, M. (Eds.), Compilando La Geología de Colombia: Una Visión a 2015. Servicio Geológico Colombiano, Bogota, pp. 63–420.

Gomez, J., Montes, N., Nivia, A., Diederix, H., 2015b. Mapa Geológico de Colombia 2015.

Goodarzi, F., Goodarzi, N.N., Malachowska, A., 2021. Elemental composition, environment of deposition of the lower Carboniferous Emma Fiord formation oil shale in Arctic Canada. Int. J. Coal Geol. 244, 103715. https://doi.org/10.1016/J.COAL.2021.103715

Hakimi, M.H., Al-Matary, A.M., Salad Hersi, O., 2018. Late Jurassic bituminous shales from Marib oilfields in the Sabatayn Basin (NW Yemen): Geochemical and petrological analyses reveal oil-shale resource. Fuel 232, 530–542. https://doi.org/10.1016/J.FUEL.2018.05.138

Han, J., Kamber, M., Pei, J., 2012. Data Mining: Concepts and Techniques. Data Min. Concepts Tech. https://doi.org/10.1016/C2009-0-61819-5

Han, S., Zhang, Y., Huang, J., Rui, Y., Tang, Z., 2020. Elemental Geochemical Characterization of Sedimentary Conditions and Organic Matter Enrichment for Lower Cambrian Shale Formations in Northern Guizhou, South China, Minerals 2020, Vol. 10, Page 793. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/MIN10090793

Hatch, J.R., Leventhal, J.S., 1992. Relationship between inferred redox potential of the depositional environment and geochemistry of the Upper Pennsylvanian (Missourian) Stark Shale Member of the Dennis Limestone, Wabaunsee County, Kansas, U.S.A. Chem. Geol. 99, 65–82. https://doi.org/10.1016/0009-2541(92)90031-Y

Hayashi, K.I., Fujisawa, H., Holland, H.D., Ohmoto, H., 1997. Geochemistry of ~1.9 Ga sedimentary rocks from northeastern Labrador, Canada. Geochim. Cosmochim. Acta 61, 4115–4137. https://doi.org/10.1016/S0016-7037(97)00214-7

Herron, M.M., 1988. Geochemical classification of terrigenous sands and shales from core or log data. J. Sediment. Petrol. 58, 820–829. https://doi.org/10.1306/212F8E77-2B24-11D7-8648000102C1865D

Horton, B.K., Saylor, J.E., Nie, J., Mora, A., Parra, M., Reyes-Harker, A., Stockli, D.F., 2010. Linking sedimentation in the northern Andes to basement configuration, Mesozoic extension, and Cenozoic shortening: Evidence from detrital zircon U-Pb ages, Eastern Cordillera, Colombia. Bull. Geol. Soc. Am. 122, 1423–1442. https://doi.org/10.1130/B30118.1

Ingeominas - UIS, 2008. ESTUDIO DE LOS PROCESOS DE INTERACCIÓN FLUIDO-ROCA EN EL CINTURÓN ESMERALDÍFERO ORIENTAL (CORDILLERA ORIENTAL, COLOMBIA) Y SU IMPORTANCIA EN LA EXPLORACIÓN DE NUEVOS YACIMIENTOS HIDROTERMALES".

Ingeominas - UIS, 2007. Investigación petrográfica y geoquímica de las sedimentitas del Cretácico Inferior (K1) y sus manifestaciones hidrotermales asociadas; planchas 169, 170, 189, 190 (Cordillera Oriental): implicaciones en la búsqueda de esmeraldas. Bogota.

Jaillard, E., Solar, P., Carlier, G., Mourier, T., 1990. Geodynamic evolution of the northern and central Andes during early to middle Mesozoic times: a Tethyan model. J. Geol. Soc. London 147, 1009–1022.

Jones, B., Manning, D.A.C., 1994. Comparison of geochemical indices used for the interpretation of palaeoredox conditions in ancient mudstones. Chem. Geol. 111, 111–129. https://doi.org/10.1016/0009-2541(94)90085-X

Kay, S.M., Ramos, V.A., Dickinson, W.R., 2009. Backbone of the Americas: Shallow Subduction, Plateau Uplift, and Ridge and Terrane Collision. Geological Society of America. https://doi.org/10.1130/MEM204

Lewan, M.D., 1984. Factors controlling the proportionality of vanadium to nickel in

crude oils. Geochim. Cosmochim. Acta 48, 2231–2238. https://doi.org/10.1016/0016-7037(84)90219-9

López, L., Lo Mónaco, S., 2017. Vanadium, nickel and sulfur in crude oils and source rocks and their relationship with biomarkers: Implications for the origin of crude oils in Venezuelan basins. Org. Geochem. 104, 53–68. https://doi.org/10.1016/J.ORGGEOCHEM.2016.11.007

Mann, P., 1995. Geologic and tectonic development of the Caribbean Plate boundary in southern Central America. Geological Society of America.

McLaughlin, D.H.J., 1972. Evaporite deposits of Bogotá area, Cordillera Oriental, Colombia. Am. Assoc. Pet. Geol. Bull. 56, 2240–2259.

McLennan, S.M., 2001. Relationships between the trace element composition of sedimentary rocks and upper continental crust. Geochemistry, Geophys. Geosystems 2. https://doi.org/10.1029/2000GC000109

Mojica, J., Dorado, J., 1987. El Jurasico anterior a los movimientos intermalmicos en las Andes Colombianos:, in: Volkheimer, W. (Ed.), Bioestratigrafía de Los Sistemas Regionales Del Jurasico y Cretácico de América Del Sur. Mendoza, pp. 49–110.

Mora, A., Horton, B.K., Mesa, A., Rubiano, J., Ketcham, R.A., Parra, M., Blanco, V., Garcia, D., Stockli, D.F., 2010. Migration of cenozoic deformation in the eastern cordillera of colombia interpreted from fission track results and structural relationships: Implications for petroleum systems. Am. Assoc. Pet. Geol. Bull. 94, 1543–1580. https://doi.org/10.1306/01051009111

Mora, A., Parra, M., Strecker, M.R., Kammer, A., Dimate, C., Rodriguez, F., 2006. Cenozoic contractional reactivation of Mesozoic extensional structures in the Eastern Cordillera of Colombia. Tectonics 25, TC2010. https://doi.org/10.1029/2005TC001854

Mora, A., Parra, M., Strecker, M.R., Sobel, E.R., Hooghiemstra, H., Torres, V., Jaramillo, J. V, 2008. Climatic forcing of asymmetric orogenic evolution in the Eastern Cordillera of Colombia. Geol. Soc. Am. Bull. 120, 930–949. https://doi.org/10.1130/B26186.1

Mora, A., Reyes-Harker, A., Rodriguez, G., Tesón, E., Ramirez-Arias, J.C., Parra, M., Caballero, V., Mora, J.P., Quintero, I., Valencia, V., Ibañez, M., Horton, B.K., Stockli, D.F., 2013. Inversion tectonics under increasing rates of shortening and sedimentation: Cenozoic example from the Eastern Cordillera of Colombia, Geological Society Special Publication. https://doi.org/10.1144/SP377.6

O'Sullivan, G., Chew, D., Kenny, G., Henrichs, I., Mulligan, D., 2020. The trace element composition of apatite and its application to detrital provenance studies. Earth-Science Rev. https://doi.org/10.1016/j.earscirev.2019.103044

Parra, M., Mora, A., Sobel, E.R., Strecker, M.R., González, R., 2009. Episodic orogenic front migration in the northern Andes: Constraints from low-temperature thermochronology in the Eastern Cordillera, Colombia. Tectonics 28. https://doi.org/10.1029/2008TC002423

Pindell, J.L., Erikson, J.P., 1994. The Mesozoic Passive Margin of Northern South America, in: Cretaceous Tectonics of the Andes. Vieweg+Teubner Verlag, Wiesbaden, pp. 1–60. https://doi.org/10.1007/978-3-322-85472-8_1

Pindell, J.L., Kennan, L., 2009. Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: an update. Geol. Soc. London, Spec. Publ. 328, 1.1-55. https://doi.org/10.1144/SP328.1

Rangel, A., Blandon, A., Giraldo, B., Ramon, J.C., Cordoba, F., 1991. Evaluacion Geoquimica de la Cuenca Llanos Orientales., in: Memorias Del Simposio Bolivariano de Cuencas Subandinas. Bogota.

Rangel, A., Brooks, P.W., Giraldo, B.N., 1999. Petroleum geochemistry of the oils in the foothills of the Llanos basin, Colombia. Rev. Latinoam. Geoquímica Org_anica 5, 5–24.

Rangel, A., Moldowan, J.M., Nino, C., Parra, P., Giraldo, B.N., 2002. Umir Formation: organic geochemical and stratigraphic assessment as cosource for Middle Magdalena basin oil, Colombia. Am. Assoc. Pet. Geol. Bull. 85, 2069–2087.

Rangel, A., Osorno, J.F., Ramirez, J.C., De Bedout, J., González, J.L., Pabón, J.M., 2017. Geochemical assessment of the Colombian oils based on bulk petroleum properties and biomarker parameters. Mar. Pet. Geol. 86, 1291–1309.

https://doi.org/10.1016/j.marpetgeo.2017.07.010

Rangel, A., Parra, P., Nino, C., 2000. The La Luna formation: chemostratigraphy and organic facies in the middle Magdalena basin. Org. Geochem. 31, 1267–1284.

Restrepo-Pace, P.A., Colmenares, F., Higuera, C., Mayorga, M., 2004. A Fold-and-thrust belt along the western flank of the Eastern Cordillera of Colombia— Style, kinematics, and timing constraints derived from seismic data and detailed surface mapping. AAPG Mem. 598–613. https://doi.org/10.1306/M82813C31

Restrepo, J.J., Toussaint, J.F., 1988. Terranes and continental accretion in the Colombian Andes. Episodes 11, 189–193.

Reyes-Harker, A., Ruiz-Valdivieso, C.F., Mora, A., Ramírez-Arias, J.C., Rodriguez, G., De La Parra, F., Caballero, V., Parra, M., Moreno, N., Horton, B.K., Saylor, J.E., Silva, A., Valencia, V., Stockli, D., Blanco, V., 2015. Cenozoic paleogeography of the Andean foreland and retroarc hinterland of Colombia. Am. Assoc. Pet. Geol. Bull. 99, 1407–1453. https://doi.org/10.1306/06181411110

Roy, D.K., Roser, B.P., 2013. Climatic control on the composition of Carboniferous–Permian Gondwana sediments, Khalaspir basin, Bangladesh. Gondwana Res. 23, 1163–1171. https://doi.org/10.1016/J.GR.2012.07.006

Sarmiento, L.F., 2019. Cretaceous Stratigraphy and Paleo-Facies Maps of Northwestern South America. pp. 673–747. https://doi.org/10.1007/978-3-319-76132-9_10

Sarmiento, L.F., 2011a. Eastern Cordillera Basin, in: Cediel, F., Ojeda, G. (Eds.), Petroleum Geology of Colombia. EAFIT, Medellin, p. 141.

Sarmiento, L.F., 2011b. Llanos Basin, in: Cediel, F., Ojeda, G. (Eds.), Petroleum Geology of Colombia. EAFIT, Medellin, p. 186.

Sarmiento, L.F., 2011c. Middle Magdalena Basin, in: Cediel, F., Ojeda, G. (Eds.), Petroleum Geology of Colombia. EAFIT, Medellin, p. 195.

Sarmiento, L.F., 2001. Mesozoic Rifting and Cenozoic Basin Inversion History of the Eastern Cordillera, Colombian Andes Inferences from Tectonic Models. Vrije Universiteit Amsterdam.

Sarmiento, L F, Van Wess, J.D., Cloetingh, S., 2006. Mesozoic transtensional basin history of the Eastern Cordillera, Colombian Andes: Inferences from tectonic models. J. South Am. Earth Sci. 21, 383–411.

Sarmiento, L. F., Van Wess, J.D., Cloetingh, S., 2006. Mesozoic transtensional basin history of the Eastern Cordillera, Colombian Andes: Inferences from tectonic models. J. South Am. Earth Sci. 21, 383–411.

Silva, A., Mora, A., Caballero, V., Rodriguez, G., Ruiz, C., Moreno, N., Parra, M., Ramirez-Arias, J.C., Ibáñez, M., Quintero, I., 2013. Basin compartmentalization and drainage evolution during rift inversion: evidence from the Eastern Cordillera of Colombia. Geol. Soc. London, Spec. Publ. 377, 369–409. https://doi.org/10.1144/SP377.15

Spalletti, L.A., Schwarz, E., Veiga, G.D., 2014. Geoquímica inorgánica como indicador de procedencia y ambiente sedimentario en sucesiones de lutitas negras: Los depósitos transgresivos titonianos (Formación Vaca Muerta) de la Cuenca Neuquina, Argentina. Andean Geol. 41, 401–435. https://doi.org/10.5027/ANDGEOV41N2-A07

Taboada, A., Rivera, L.A., Fuenzalida, A., Cisternas, A., Philip, H., Bijwaard, H., Olaya, J., Rivera, C., 2000. Geodynamics of the northern Andes: Subductions and intracontinental deformation (Colombia). Tectonics 19, 787–813. https://doi.org/10.1029/2000TC900004

Taylor, J., McLennan, S., 1985. The Continental Crust: Its Composition and Evolution. xvi + 312 pp. Oxford, London, Edinburgh, Boston, Palo Alto, Melbourne: Blackwell Scientific. Price £16.80 (paperback). ISBN 0 632 01148 3. Cambridge University Press. https://doi.org/10.1017/S0016756800032167

Tegelaar, E.W., Zaugg, P., Rangel, A., 1995. Petroleum Systems of the Foothills of the Suthern Llanos and Putumayo Basins., in: Congreso Colombiano Del Petroleo.

Tribovillard, N., Algeo, T.J., Baudin, F., Riboulleau, A., 2012. Analysis of marine environmental conditions based onmolybdenum–uranium covariation—Applications to Mesozoic paleoceanography. Chem. Geol. 324–325, 46–58. https://doi.org/10.1016/J.CHEMGEO.2011.09.009

Trindade, L.A., Mellos, M.R., Luna, O., Velandia, J., 1997. Petroleum Systems of the Piedemonte Oriental, in: Exploracion Petrolera En Cuancas Subandinas. Cartagena, pp. 303–323.

Vásquez, M., Altenberger, U., 2005. Mid-Cretaceous extension-related magmatism in the eastern Colombian Andes. J. South Am. Earth Sci. 20, 193–210. https://doi.org/10.1016/j.jsames.2005.05.010

Vayssaire, A., Addallah, H., Hermoza, W., Figari, E., 2014. Regional Study and Petroleum System Modeling of the Eastern Llanos Basin. Search Discov. Artic. 9.

Vermeesch, P., 2013. Multi-sample comparison of detrital age distributions. Chem. Geol. 341, 140–146. https://doi.org/10.1016/j.chemgeo.2013.01.010

Villamil, T., 1999. Campanian–Miocene tectonostratigraphy, depocenter evolution and basin development of Colombia and western Venezuela. Palaeogeogr. Palaeoclimatol. Palaeoecol. 153, 239–275. https://doi.org/10.1016/S0031-0182(99)00075-9

Ward, D., Goldsmith, R., Cruz, J., Restrepo, H., 1973. Geología de los Cuadrangulos H12-Bucaramanga y H13-Pamplona, Departamento de Santander. Boletín Geológico del Serv. Geológico Nac. XXI, 144.

Wedepohl, K.H., 1971. Environmental influences on the chemical composition of shales and clays. Phys. Chem. Earth 8, 307–333. https://doi.org/10.1016/0079-1946(71)90020-6

Zartman, R.E., Doe, B.R., 1981. Plumbotectonics—the model. Tectonophysics 75, 135–162. https://doi.org/10.1016/0040-1951(81)90213-4

7. COMPLEMENTARY MATERIAL

7.1. Appendix 1

The samples were analyzed for whole rock by natural chemical attack and corresponded to mudstones with a high content of organic matter. Values from the literature on the concentrations of Rb, Sr and Nd for those lithologies were taken as a guide for the determination of the initial aliquots.

The acid digestion was carried out in Parr-type pump, due to the lack of digestion observer when using CEM MARS-5 microwaves with Easy Prep tubes. Tests with H_2O_2 were carried without improvement in the total digestion of the sample. The procedure for natural chemical attack for whole rocks began with the weighing of approximately 100 mg of pulverized sample and located in a Teflon savillex, to which concentrated and distilled acids (1 ml of HNO3 and 3 ml of HF) were added. and totaled to 5 ml with H2O2 and then placed on the heat plate inside the Clean Box for 12 hours. The samples were placed in the PARR bombs in the oven at 220℃ for 5 days for their attack, to then be transferred to savillex and placed to dry inside the clean box at 60°C, then 0.5ml of concentrated HNO3 was added and placed again in the clean box to evaporate. It was dissolved with 10 ml of 6.0 M HCL and the closed savillex was left heating overnight in the clean box. Subsequently, it was placed to evaporate and 1 ml of concentrated HNO3 was added to the final residue to place it to evaporate before directing it to the elements separation columns.

Some samples were not completely dissolved in HBr 0.7 M despite being attacked in PARR bombs. The samples were deposited in the columns without the undissolved part. The solid part remained in the matrix to be made a tentative solution with 2 M HNO3 before the Sr column. In the Sr column, the Sr Spec resin (Eichrom) was used, which was conditioned twice with 0.4 ml of HNO3 2 M. Subsequently, the sample was added and 0.2 ml of 2 M HNO3 was added to direct the matrix for rare earths. Then I discard the matrix of adding 5 times 0.2 ml of 7M HNO3 and three times 0.2 ml of 2M HNO3 due to the barium. The Sr collection was completed with the 6-fold addition of 0.5 ml of 0.05 M HNO3. The previously collected matrix was directed to the columns with LN Spec resin (Eichroom) for the separation of Nd and Sm.

7.2. **Appendix 2**

Geochemistry results of major, minor and trace elements from this study.

Sample ID	FORMATION	C	S	Al2O3	BaO	CaO	Cr2O3	Fe2O3	K2O	MgO	MnO	Na2O	P2O5	SiO2	SrO	TiO2	V2O5	LOI
		%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
YP13A	Paja	12.6	0.21	4.31	0.03	39.1	0.02	1.56	0.58	0.8	<0,01	0.15	0.81	14.7	0.11	0.14	0.28	38.18
JC7E	Fomeque	2.4	0.89	4.62	<0,01	<0,01	<0,01	1.19	0.45	0.1	<0,01	<0,1	0.05	88.4	<0,01	0.57	<0,01	3.96
YP43A	La Luna	4.74	0.45	2.98	0.04	11.6	<0,01	0.84	0.39	0.36	<0,01	0.25	0.29	69.4	0.03	0.09	0.01	12.88
YP14B	Paja	7.15	1.42	9.48	0.05	16.2	0.01	3.55	0.92	1.12	<0,01	0.39	0.73	46.9	0.07	0.27	0.04	19.32
JC15D	Fomeque	1	0.12	21.2	0.08	2.93	<0,01	6.87	3.15	0.76	0.03	0.6	0.16	53.7	0.02	1.2	0.03	7.27
YP47B	La Luna	0.98	0.51	15.1	0.04	0.19	<0,01	1.55	1.46	0.44	<0,01	0.21	0.19	72.8	0.03	0.72	0.01	7.29
YP33A	Fomeque	1.45	0.03	23.5	0.08	0.59	0.02	5.58	3.32	0.77	<0,01	0.75	0.24	55	0.06	1.21	0.02	8.09
YP54COMP	Chipaque	2.4	0.02	18.5	0.04	0.02	<0,01	1.85	1.4	0.44	<0,01	0.1	0.14	65	0.01	0.71	0.02	10.2
YP14A	Paja	4.96	2.29	15.8	0.03	9.42	0.02	3.98	1.37	0.75	<0,01	0.21	0.47	50.9	0.07	0.46	0.03	16.65
YP44A	La Luna	7.73	0.76	5.4	0.07	18.5	0.02	0.96	0.64	0.28	<0,01	0.19	0.53	54.3	0.07	0.17	0.05	19.99
YP27A	Paleozoico	0.3	0.02	20.1	0.08	0.38	0.02	6.91	4.56	2.97	<0,01	0.52	0.13	58.5	<0,01	0.74	0.03	4.33
YP31B	Macanal	6.22	0.25	18.1	0.06	12.8	0.01	4.95	2.92	1.86	0.07	0.65	0.2	40.5	0.03	0.66	0.07	16.49
YP19B	Tablazo	1.54	0.03	14.1	0.04	3.93	<0,01	4.45	2.35	0.67	<0,01	0.15	0.39	67	0.02	0.46	0.02	7.76
YP45B	La Luna	10.3	0.64	1	0.03	30.7	0.01	0.49	0.13	0.37	<0,01	0.18	0.52	38.6	0.1	0.03	0.04	27.6
JC5A	Chipaque	0.76	0.12	18.8	0.05	0.02	<0,01	1.57	2.14	0.56	<0,01	0.23	0.16	68.2	0.01	0.91	<0,01	6.6
YP28A	Macanal	1.35	0.31	25.3	0.07	0.03	0.02	4.65	4.63	1.51	<0,01	0.98	0.1	56.5	0.03	1.03	0.04	6.26
YP58COMP	Chipaque	0.85	0.02	25.8	0.04	0.48	0.02	2.68	2.32	0.83	<0,01	0.27	0.19	57.2	0.02	0.88	0.03	9.1
JC12C	Une	1.48	0.01	15.5	0.04	0.04	<0,01	0.47	1.54	0.17	<0,01	0.33	0.08	76	0.02	1.04	<0,01	6.67
JC7B	Fomeque	2.41	0.02	22.8	0.07	0.03	<0,01	0.77	3.48	0.4	<0,01	0.35	0.11	65.2	0.02	1.49	0.03	6.65
YP46A	La Luna	11.3	0.17	0.57	0.1	49.8	<0,01	0.26	0.07	0.52	<0,01	0.16	0.2	10.1	0.11	<0,01	0.02	39.51

Sample ID	FORMATION	Ag	Al	As	B	Ba	Be	Bi	Ca	Cd	Ce	Co	Cr	Cs	Cu	Dy	Er	Eu	Fe
		ppm	ppm	ppm	%	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
YP13A	Paja	<1	19811	52	<0,1	114	<5	<0,5	242603	64	24.6	<10	139	2.6	103	3.04	2	0.67	10076
JC7E	Fomeque	2	20799	<30	<0,1	63	<5	<0,5	1153	<10	46.4	<10	28	1.1	<10	3.73	2.51	0.87	7313
YP43A	La Luna	6	13758	<30	<0,1	415	<5	<0,5	76122	<10	9.8	<10	78	1	23	1.12	0.67	0.22	5381
YP14B	Paja	6	43464	<30	<0,1	213	<5	<0,5	103454	15	73	<10	133	3.4	<10	4.59	2.89	1.11	22029
JC15D	Fomeque	<1	90404	<30	<0,1	326	<5	<0,5	17195	<10	107.3	<10	90	7.5	13	6.88	3.71	1.76	34491
YP47B	La Luna	5	64285	<30	<0,1	157	<5	<0,5	1862	<10	113.8	<10	54	2.9	<10	4.04	2.99	0.89	8808
YP33A	Fomeque	6	78145	<30	<0,1	292	<5	<0,5	3475	<10	98.3	<10	89	9.9	<10	6.12	3.44	1.46	24664
YP54COMP	Chipaque	3	74004	<30	<0,1	181	<5	<0,5	<1000	<10	116.8	<10	50	4.1	<10	5.79	3.31	1.33	10054
YP14A	Paja	1	70534	<30	<0,1	350	<5	<0,5	58233	10	105.4	<10	128	5.4	16	6.11	3.28	1.47	23705
YP44A	La Luna	4	21713	<30	<0,1	486	<5	<0,5	105100	<10	27.5	<10	139	2.2	31	1.72	1.13	0.39	5305
YP27A	Paleozoico	2	86040	55	<0,1	470	<5	<0,5	2813	<10	87.2	21	88	7.1	29	5.78	3.32	0.86	37199
YP31B	Macanal	<1	85272	<30	<0,1	352	<5	<0,5	82278	<10	23.1	<10	87	11	38	6.06	3.55	1.16	30346
YP19B	Tablazo	2	55127	<30	<0,1	158	<5	<0,5	22177	<10	75.9	<10	70	5.1	<10	4.42	2.44	1.12	22584
YP45B	La Luna	6	4452	<30	<0,1	193	<5	<0,5	166128	<10	5.5	<10	66	0.5	18	0.61	0.47	0.1	2789
JC5A	Chipaque	6	84827	<30	<0,1	231	<5	0.5	<1000	<10	132.6	<10	83	4.6	<10	7.57	4.27	1.61	9709
YP28A	Macanal	<1	114422	<30	<0,1	473	<5	<0,5	1112	<10	11.4	<10	92	13.7	31	4.77	3.13	0.53	27083
YP58COMP	Chipaque	2	97507	<30	<0,1	214	<5	<0,5	3579	<10	139.7	<10	92	6.6	29	6.03	3.46	1.42	14771
JC12C	Une	3	68964	<30	<0,1	147	<5	<0,5	1080	<10	107.4	<10	42	4.9	<10	8.76	5.69	1.4	2853
JC7B	Fomeque	<1	106022	<30	<0,1	382	<5	<0,5	1217	<10	130.5	<10	110	17.6	11	8.98	5	2.22	5041
YP46A	La Luna	<1	2754	<30	<0,1	755	<5	<0,5	318715	<10	2.9	<10	<20	0.2	<10	0.23	0.13	<0,05	1807

Sample ID	FORMATION	Ga	Gd	Ge	Hf	Ho	In	K	La	Li	Lu	Mg	Mn	Mo	Nb	Nd	Ni	P	Pb	Pr
		ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
YP13A	Paja	6	3.2	1	<2	0.64	<0,2	5764	19.1	10	0.29	4472	<100	621	<10	15	428	3023	<20	3.69
JC7E	Fomeque	5	3.82	<1	27	0.83	<0,2	5029	23.3	<10	0.47	375	<100	<3	<10	20.9	16	<100	<20	5.41
YP43A	La Luna	3	1.06	<1	<2	0.23	<0,2	4092	6.2	22	0.12	2131	<100	4	<10	5.4	41	958	<20	1.35
YP14B	Paja	11	5.6	1	<2	0.98	<0,2	9179	40.8	53	0.45	5402	<100	60	<10	32.5	77	2994	<20	8.39
JC15D	Fomeque	23	8.04	2	5	1.34	<0,2	25719	53	109	0.5	3910	251	<3	18	47	28	573	<20	12.57
YP47B	La Luna	15	4.31	1	10	0.91	<0,2	12176	55.5	64	0.67	2383	<100	<3	20	43	12	822	26	12.72
YP33A	Fomeque	22	6.42	2	4	1.21	<0,2	23263	49.3	110	0.51	3080	<100	<3	17	43.3	17	754	<20	11.32
YP54COMP	Chipaque	17	6.94	1	4	1.09	<0,2	11425	57.9	52	0.49	2299	<100	<3	18	48.6	<10	383	27	13.66
YP14A	Paja	17	7.44	2	2	1.18	<0,2	11713	56.7	101	0.43	3307	<100	31	11	46.2	60	1859	23	12.32
YP44A	La Luna	6	2.03	1	<2	0.35	<0,2	5131	15.9	15	0.15	1730	<100	4	<10	12.6	80	1806	<20	3.21
YP27A	Paleozoico	24	5.62	2	4	1.11	<0,2	36260	52.5	68	0.42	14761	<100	<3	14	35.4	41	711	<20	9.49
YP31B	Macanal	21	5.71	2	3	1.18	<0,2	25061	11.6	78	0.48	9773	572	38	14	14.2	75	606	69	3.11
YP19B	Tablazo	13	5.23	1	3	0.85	<0,2	17981	40.3	46	0.32	3161	<100	6	<10	33.5	25	1163	<20	8.9
YP45B	La Luna	1	0.57	<1	<2	0.14	<0,2	1440	4.4	21	0.07	1368	<100	8	<10	3	62	1791	<20	0.8
JC5A	Chipaque	21	8.38	1	10	1.47	<0,2	18084	67.5	66	0.69	2678	<100	<3	26	55.7	17	782	32	15.21
YP28A	Macanal	30	3.28	2	4	1.05	<0,2	37291	6	44	0.46	7667	<100	8	18	6.1	21	200	<20	1.49
YP58COMP	Chipaque	26	6.88	1	4	1.13	<0,2	17769	70.3	83	0.56	3643	<100	<3	23	57.6	21	879	28	16.1
JC12C	Une	16	7.65	1	23	1.86	<0,2	11800	54.5	41	1.02	696	<100	<3	27	45.7	<10	659	<20	12.21
JC7B	Fomeque	26	10.06	1	9	1.74	<0,2	26981	65.6	53	0.72	1905	<100	<3	25	59	<10	359	<20	15.19
YP46A	La Luna	<1	0.22	<1	<2	<0,05	<0,2	<1000	1.8	16	<0,05	3125	<100	4	<10	1.4	19	1045	<20	0.38

Sample ID	FORMATION	Rb	Sb	Sc	Sm	Sn	Sr	Ta	Tb	Th	Ti	Tl	Tm	U	V	W	Y	Yb	Zn	Zr
		ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
YP13A	Paja	33	23	<5	3	<5	826	<10	0.44	4.2	798	8.7	0.26	22.03	1399	<100	25.22	1.8	1683	39.5
JC7E	Fomeque	18	0.9	<5	4.3	<5	44	<10	0.59	7.3	3191	1	0.42	2.43	32	<100	20.69	3	<50	1128.4
YP43A	La Luna	15	0.7	<5	1	<5	281	<10	0.16	2	675	0.6	0.1	4.56	49	<100	8.22	0.7	100	24.5
YP14B	Paja	48	1.5	10	6	7	606	<10	0.78	13.1	1630	2.2	0.4	7.57	247	<100	31.26	2.8	692	60.9
JC15D	Fomeque	151	<0,5	17	9.3	<5	176	<10	1.17	17.3	6638	1	0.55	3.12	114	<100	32.91	3.7	69	195
YP47B	La Luna	57	<0,5	7	6.2	<5	185	<10	0.67	9	3906	0.5	0.54	2.77	46	<100	23.42	4	<50	385.2
YP33A	Fomeque	152	<0,5	15	7.7	<5	266	<10	0.96	15.2	5246	0.7	0.53	2.57	104	<100	29.46	3.3	70	162.7
YP54COMP	Chipaque	64	<0,5	11	8.8	<5	93	<10	1	22.2	3812	<0,5	0.49	3.54	33	<100	26.12	3.3	<50	145
YP14A	Paja	69	1.2	13	8.7	<5	546	<10	1.06	21	2682	1.6	0.48	5.74	160	<100	33.66	2.9	477	71.3
YP44A	La Luna	29	1.1	<5	2.2	<5	401	<10	0.29	5.8	901	0.5	0.16	5.34	165	<100	12.29	1	182	36.7
YP27A	Paleozoico	163	<0,5	15	6.1	<5	43	<10	0.87	15.3	3979	0.9	0.46	1.95	141	<100	27.69	3.1	72	127.2
YP31B	Macanal	163	2.5	18	4	<5	207	<10	0.9	14.2	4026	1.9	0.52	4.28	302	<100	30.8	3.3	176	108.5
YP19B	Tablazo	89	<0,5	8	6.1	<5	153	<10	0.76	14.7	2431	1.1	0.35	3.59	76	<100	23.14	2.2	101	81.2
YP45B	La Luna	6	1.3	<5	0.5	<5	615	<10	0.1	1.1	179	<0,5	0.07	5.84	206	<100	6.47	0.4	156	9.8
JC5A	Chipaque	89	<0,5	12	9.8	<5	136	<10	1.24	28.1	5301	0.5	0.64	4.43	95	<100	36.15	4.3	<50	350.8
YP28A	Macanal	241	1.5	19	1.9	<5	181	<10	0.68	19.3	5799	2.1	0.48	3.85	196	<100	26.8	3.2	60	149.6
YP58COMP	Chipaque	102	<0,5	14	9.9	<5	111	<10	1.04	28.9	4358	0.6	0.56	3.54	100	<100	26.6	3.7	73	147.8
JC12C	Une	66	0.7	11	8.3	<5	86	<10	1.4	23.4	5797	0.5	0.91	7.16	50	<100	49.57	6.1	<50	940.2
JC7B	Fomeque	158	<0,5	17	11.1	18	149	<10	1.47	19.8	8429	1	0.75	4.54	139	<100	42.41	5	<50	326
YP46A	La Luna	3	0.6	<5	0.3	<5	837	<10	<0,05	1	<100	<0,5	<0,05	1.39	165	<100	1.35	0.1	127	10.8