Analysis of the LPG industry in Colombia: use of Big Data and digital media analysis to forecast and analyze industry trends.

Carlos Yezid Rozo Alvarez[1], Astrid Ramírez Valencia[2], Lindsay Álvarez Pomar[3]

[1]Universidad Distrital Francisco José de Caldas, Bogotá -Colombia, http://orcid.org/0000-0001-7857-4600, mail: cyrozoa@udistrital.edu.co

[2]Universidad Distrital Francisco José de Caldas, Bogotá -Colombia, http://orcid.org/0000-0002-3025-5982, mail: aramirezv@udistrital.edu.co

[3]Universidad Distrital Francisco José de Caldas, Bogotá -Colombia, http://orcid.org/0000-0002-8818-0901, mail: lalvarez@udistrital.edu.co

Summary

This research, focused on the first link of the LPG Liquefied Petroleum Gas supply chain, seeks to expose the application of Big Data platforms in the sector, analyzed from the methodology of complex systems, since the data require to be analyzed if they present normality and if this requires the application of non-parametric or parametric tests.

In this research the use and application of Big Data ethics is relevant, since it is required to analyze each of the variables, to observe the existing correlation between the various variables that make up the first link of the supply chain; it is important to expose that the information and data is of a significant and considerable volume that requires such application, therefore the analysis of correlation and normality must be generated with a significant volume of data, therefore, with Big Data platforms we want to optimize the accuracy in identifying correlations, even those that might not be evident in smaller sets.

In this context, the data to be processed have been obtained from various sources that show market fluctuations, and therefore it is necessary to establish forecast scenarios that are closer to reality and that allow strategic decisions to be made in this sector. Correlation analysis is used as a statistical technique to determine the relationship between two quantitative variables, which, in this case, makes use of Pearson's correlation, or Pearson's product-moment correlation coefficient, which measures both the strength and the direction of the linear relationship between continuous variables.

Key words: GLP, Big Data, Correlation, scenarios.

1. Introduction

This paper investigates the Liquefied Petroleum Gas industry in Colombia over time, as it is based on 17 years of data. Data was obtained from sources such as the Unified Information System (SUI), Macro Data and CENIT [1][2][3]. Using these sources confirmed the reliability of the data provided and their interpretation through the complex systems methodology. On the other hand, the analysis was carried out with the help of Big Data ethics as it allows processing large amounts of unstructured data with a different view. As for the long-term effect, the behavior of the industry can be observed here, including hidden patterns, correlations and accurate predictions.

1.1 Background

The use of LPG as an energy source in Colombia dates back to the mid-twentieth century, although there are records indicating that its production and commercialization dates back to the 1930s. Initially, LPG was distributed in small volumes by a small number of agents to specific areas of consumption, but over time it became consolidated as an essential resource in sectors such as residential, industrial and agricultural.[4]

Considering the evolution of the LPG industry in Colombia has been marked by the need to satisfy a growing demand for fuels, the availability of the product and the regulation in terms of prices, subsidies and contributions, therefore, since the 1950s, the country has developed an infrastructure to produce, store and distribute LPG, especially through refineries such as Barrancabermeja and Cartagena [5]. However, domestic production has shown a decreasing trend in recent years, which has increased the dependence on imports to cover domestic demand.[6]

Currently, although LPG occupies an important place in the country's energy matrix, it faces significant challenges that affect its competitiveness. These include the volatility of international prices, the need to modernize the distribution infrastructure and competition with new emerging energy sources [6]. The price of LPG at the national level depends on multiple factors, such as the international costs of propane and butane, as well as the costs of production, distribution and government regulation [7]. In addition, the Energy and Gas Regulatory Commission (EGCR) establishes that the maximum price of LPG produced by Ecopetrol is based on the international price of the product in the Mont Belvieu market, located in the Gulf of Mexico [8]. In this sense, the use of Big Data ethics and digital media facilitate a more accurate and comprehensive examination of global changes in the industry, such as prices, consumption and distribution efficiency, becoming an important technique to analyze the supply chain and its competitiveness.

1.1.1 Origin of the LPG industry in Colombia

The production and commercialization of LPG in the country began in the 1930s and 1950s, although its use was limited to low availability and lack of adequate infrastructure for its distribution[9] , during the 1960s and 1970s, with the expansion of refineries such as the one in Barrancabermeja, LPG production increased significantly, allowing the creation of the first distribution networks aimed at supplying urban homes[10] , leading to the establishment of certain patterns and trends in the development and distribution of LPG, which is why through Big Data ethics and digital media analysis we seek to project future events and improve the management of the Colombian energy sector.

1.1.2 Growth and Consolidation

In the search to provide a solid basis for improving government policies and the efficiency of services, Big Data ethics and digital media are used to achieve an adequate collection, such as an exhaustive

examination of all data supporting consumption and distribution over time, which has made it possible to identify the main drivers of growth in the sector, Thus, based on the above, the Colombian government implemented regulations in the 1980s and 1990s regarding commercialization[11], which led to the growing demand in the late 1990s to consolidate LPG as an essential fuel for the residential sector, especially among the strata 1 and 2[12].

1.1.3 The LPG industry and its challenges

In 2007, the CREG issued Resolution 066, which allowed the liberalization of the market and fostered competition in the commercialization of LPG, thus encouraging private investment in distribution infrastructure [13]. This change generated an increase in domestic use and also boosted its use in the transportation sector, with initiatives such as Autogas [14]. Despite these advances, the industry faces constant challenges related to price volatility, competition with other energy sources and the need to modernize the distribution infrastructure [15]. Overcoming these difficulties requires not only improving existing infrastructure and ensuring product availability, but also adapting to changing international market conditions and emerging environmental policies [16]. Under these scenarios, the LPG sector must make intelligent decisions in a very unstable environment, so Big Data ethics and digital media as tools seek to model and forecast price fluctuations, monitor demand in real time and optimize the supply chain.

1.2 Scope of application

Liquefied Petroleum Gas is an essential energy resource with multiple applications in Colombia. Its versatility and efficiency position it as a fundamental alternative in the country's energy transition. Its use in the transportation sector and power generation reinforce its importance within the national energy matrix, contributing to the diversification and sustainability of the Colombian energy system, therefore by making an analysis of industry trends, seeking to optimize distribution and forecast future demand, Big Data and digital media analysis allow monitoring and managing existing information in this sector, which will facilitate strategic decision making to ensure its availability and efficiency at the national level.

This is why the analysis of data from various sources, with this tool, will support the identification of the various possibilities to improve the efficiency of the supply chain and the adequacy of supply to demand, reducing costs and optimizing distribution.

It cannot be ignored that, when performing an analysis of the LPG industry in Colombia, its main characteristics and its role within the national economy, as well as key regulatory aspects, must be considered. It is essential to understand that this industry is composed of various actors actively participating in the country's energy matrix, which allows combining data from various sources, including production, distribution and consumption data, which makes Big Data and digital media platforms enable a comprehensive and dynamic analysis of the sector, improving responsiveness to market concerns.

1.2.1 The Role of LPG in Colombia's Energy Matrix

According to Victor Melgarejo Arias, Country Manager of G+Energy and New Business Development Manager at UNIGAS, Colombia stands out worldwide for having one of the cleanest energy matrixes. Approximately 68% of its energy comes from hydroelectric sources, 2% from non-conventional

renewable sources and the remaining 30% from hydrocarbons, including thermoelectric plants that provide stability to the energy system [17]. In this scenario, LPG has remained as an imperative energy for energy services in regions where natural gas has not penetrated as a public service. Therefore, relying on Big Data and digital media analysis platforms in the analysis of the LPG industry ensures that the study of LPG within the energy matrix allows optimizing the industry and planning its supply chain.

1.3 Objective of the research

In essence, researching the Liquefied Petroleum Gas (LPG) industry in the national territory, especially since this fuel is framed as one of the clean fuels, aims to examine the strategic relevance of LPG in the Colombian energy basket. Thus, the research seeks to generate favorable scenarios that support the formulation of public policies and improve the efficiency of the sector. Within the research it is necessary to focus on the following questions:[18]

General research question (RQ):

Within Colombia's energy matrix, what is the role of LPG compared to other energy sources such as natural gas, coal and electricity?

Research-derived questions (SRQ):

SRQ1: How is it possible to guarantee that LPG will be able to support the demand for fuels in Colombia?

SRQ2: How much reliance is placed on imported fuels and how can LPG leverage this demand?

In consideration, to address the objectives of this research, the methodological approach proposed by Haugen is accepted, structured in four areas of interest: (1) exploration of the problem, (2) literature review, (3) conducting a case study and (4) synthesis of the findings, as shown in Figure 1[18].

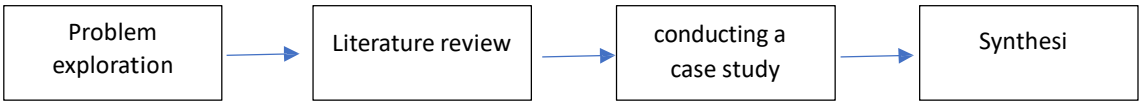

Figure 1. Research design.

2. Literature review

2.1 Conceptual framework

It is important to highlight that the analysis starts by studying the conceptualization of the Liquefied Petroleum Gas (LPG) industry and its supply chain as a complex system, understanding that there are innumerable actors and their respective interactions.
If the LPG sector is observed concretely, it can be analyzed as a complex system, as a result of the various interactions that occur between the constituent parts of the same industry. In this order of ideas, it is necessary to address the dynamics of feedback, uncertainty and adaptation of the actors, and how these affect efficiency, costs and strategic decision making within the sector.

Also, producers' decisions regarding what quantities of liquefied petroleum gas to use are largely determined by anticipated demand, which can be influenced by changes in the prices of other fuels or regulatory changes, such as subsidy policies or environmental regulations [19].

2.1.1 Complex systems

By conceptualizing the LPG industry as a system, holistic properties emerge that are derived from the interaction of its parts, known as emergent properties. This holistic approach allows us to identify how the interactions between the components of the system generate complexity [20]. In addition to this inherent complexity, it also recognizes the universal concatenation of phenomena, which establishes that each system is part of a larger one. This shows that some systems are more complex than others, depending on the number of elements that compose them and the variety of their interactions [21].

It is evident that any complex system presents a series of characteristics that will not be addressed in this section, but that the constituent parts of the system are mutually related, therefore, this connectivity can generate chain effects throughout the system [22].

2.1.1.1 The LPG Industry as a Complex System

The LPG industry in Colombia is a rather complex system, thanks to the way in which the different links in its supply chain interact in a dynamic and non-linear manner. These interactions give rise to emergent behaviors that are difficult to anticipate, all due to the interconnection between the various actors involved. Any change in one of these components can significantly impact the others, which in turn affects the overall performance of the system [23]. Supply chain relationships are neither linear nor follow predictable patterns; even small changes can trigger disproportionate effects, making it difficult to accurately anticipate outcomes [24].

In addition, the LPG industry shows a capacity for self-organization in the face of external changes, such as variations in the price of the dollar or population growth. This reorganization occurs without explicit centralized control, but through local interactions among its components. [25] Although there are entities such as the UPME, the CREG, the Ministry of Mines and Energy and the Mining and Energy Planning Unit, there is no external agent that absolutely controls the behavior of the system. Emerging patterns result from local interactions, showing the decentralized nature of the system [26].

2.1.1.2 Approaches to the LPG industry as a complex system

From the perspective of complex systems, the LPG industry in Colombia is characterized by the interaction of multiple agents that, when interacting with each other, generate non-linear, emergent behaviors that are difficult to predict from an individual analysis.

In this type of complex system, networks are formed from interactions between nodes or actors [27]. In the LPG industry, there are several actors that play a preponderant role such as producers, distributors, marketers, consumers and transporters. All of them interact in a complex network that affects the overall behavior of the sector. The interactions between these actors create emerging patterns that cannot be observed if a separate analysis is carried out [28].

2.1.2 System dynamics

System dynamics is a valuable tool for modeling and understanding how complex systems evolve over time. This approach takes into account feedback loops and flows of information, energy and resources [29]. It is especially useful in the study of the LPG industry, as it allows for the examination of causes and effects related to the availability, export, import, and variations in demand and supply of LPG over time.

2.1.2.1 Feedback model in system dynamics

System dynamics is based on feedback loops, both positive and negative, which are crucial for understanding the behavior of complex systems.

Positive feedback: If the price of LPG rises, demand may fall, which in turn reduces production and causes shortages, generating a further increase in prices in an amplifying cycle.

Negative feedback: Government policies that seek to lower LPG prices or increase supply can help stabilize the market, helping to balance prices.

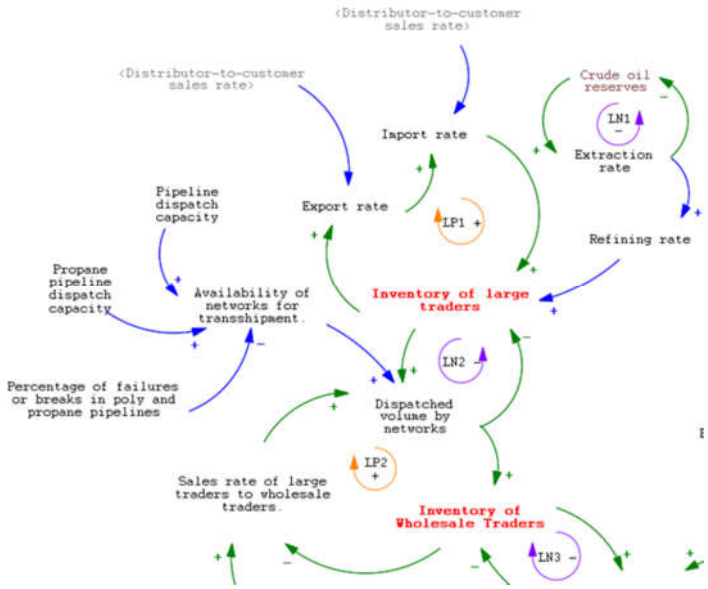

Figure 2. Feedback in the LPG industry. Source: author, research results

2.1.3 Modeling and simulation of complex systems

In order to understand the LPG industry as a complex system, computational modeling is used to help us represent the dynamics of this sector. In this case, the Vensim tool was used, which allows us to simulate and visualize how the different variables of the system interact over time. The systems that are part of this chain are composed of multiple agents that interact and adapt according to the experience they acquire. However, it is important to note that no single agent has total control over the overall behavior of the system [22]. Therefore, feedback processes are essential to sustain the internal diversity of the system, favoring fundamental characteristics such as self-organization, emergence and evolution [30].

3. Case study

This analysis is based on quantitative data processed using IBM SPSS (Statistical Package for the Social Sciences statistical software), a tool designed to analyze numerical or categorical data using functions such as descriptive statistics, significance testing, regression analysis, analysis of variance (ANOVA) and advanced techniques, all through a graphical interface or programmable commands [31].

The analysis focuses on the study of Large Marketers, the first link in the LPG supply chain, based on available big data on crude oil production sources and delivery points at the national level, noting that from the use of Big Data and digital media analysis, it is possible to perform a deep analysis, which allows greater precision in the identification of bottlenecks, to obtain greater efficiency at all levels of the supply chain.

3.1 Analysis.

This research focuses on analyzing the current situation of large marketers using data from different repositories to facilitate the necessary analysis.

3.1.1 Correlation analysis.

The inclusion of Big Data ethics and digital media analysis in this correlation study allows access to a much larger amount of data, which improves accuracy in identifying correlations, even those that may be less apparent in smaller data sets. Processing large volumes of data from various sources helps to adjust the study to changing market dynamics, generating more accurate and useful forecasts for strategic decision making in the LPG sector.

Therefore, correlation analysis is a statistical technique used to determine the relationship or association between two quantitative variables. According to Waldmann (2023), this study is carried out using Pearson's correlation method, also known as Pearson's product-moment correlation coefficient. This method measures both the strength and direction of the linear relationship between two continuous variables. The coefficient, which is represented as r, is calculated with the following formula: [33]

$$\bar{x} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \Sigma(y_i-\bar{y})^2}}$$

Where: Xi, Yi are the individual values of the variables X and Y.

$\bar{x}$, $\bar{y}$ are the means of X and Y, respectively.

The value of r varies between -1 and 1

r = 1 indicates a perfect positive linear correlation.

r = -1 indicates a perfect negative linear correlation.

r = 0 indicates no linear correlation

3.1.2 Scatter plot

Montgomery, Peck and Vining (2012) it emphasize that the scatter plot is a key tool for exploratory data analysis, describing it as the simplest and most effective way to assess the relationship between two variables. This graph allows one to visually identify whether there is a linear, nonlinear, or no discernible relationship before applying models.

Scatter plots are even more valuable in the world of Big Data and digital media analysis because they help us see connections between a large amount of data from different sources. The ability to create scatter plots that incorporate data from a variety of variables and time points is enhanced by the use of Big Data and digital media analysis, making it easier to identify patterns and trends that might not be evident in smaller data sets. Large- scale analysis also makes it possible to find intricate and subtle relationships, which can increase the accuracy of behavioral predictions in the LPG sector.

In a scatter plot, each point is defined by coordinates (x_i , y_i) , where:

x_i es the value of the independent variable (X-axis) for observation i,

y_i is the value of the dependent variable (Y-axis) for the same observation i. The bivariate data set is expressed as:

{(x_1, y_1), (x_2, y_2),…..,(x_n, y_n)}

Where n is the total number of observations. The graph plots these points without assuming a specific relationship beforehand. If the data show a linear relationship, a regression line can be fitted with the following formula:

y=β0 + β1x

where:

β0 is the intersection with the y-axis,

- β1 is the slope of the straight line, calculated as:

$$\beta 1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\beta 0 = \bar{y} - \beta_1\bar{x}$$

The use of the scatter plot in this research allows for

- Explore possible relationships between variables.
- Identify outliers that deviate from the general pattern.
- Visualize general trends or behaviors.

4. Results

The analysis conducted on the supply chain link associated with the Large Marketers considered several variables, such as: exploration contracts, wells drilled, crude oil reserves, oil production, LPG production, imports and exports. This analysis focused on the Large Marketers subsystem in order to understand their behavior and their impact on the overall LPG industry.

For this analysis, Big Data integration was carried out, which allowed the analysis of large volumes of data from various sources, including import, export, exploration and production records, which were housed in various databases, as well as historical data and facilitated the analysis of complex correlations between variables.

The first stage of this process consisted of applying goodness-of-fit tests to determine whether the data collected fit a specific distribution or probability model, such as the normal distribution[34] , This criterion was used in accordance with what was established by Romero (2016t: in consideration Big Data and digital media analysis allowed testing on large data sets, which improved the robustness and reliability of the statistical models used, by considering a greater diversity of information and possible interactions between variables.

4.1 Normality test

Hypothesis Statement

The null hypothesis (Ho): The sample comes from a normal distribution.

The alternative hypothesis (Ha): The data do not follow a normal probability model.

Significance level

Confidence 95%.

Significance (Alpha) 5%.

Data collected during the last 17 years are available, which constitutes a sample of less than 50 elements. Therefore, the Shapiro-Wilk test is used.

Table 1: Statistical test used

	Wilk- Shapiro		
	Statistical	gl	Sig.
Signed	.913	18	.097
Drilled Wells	.894	18	.044
Crude Oil Reserves	.947	18	.387
LPG Produced	.901	18	.061
LPG Imported	.663	18	.000
LPG Exported	.852	18	.009
Large Marketers Sales	.898	18	.054
Large Marketers	.914	18	.103

Source: the author, research results.

Decision
criteria

If p < 0.05 reject Ho and accept Ha

If p>= 0.05 accept Ho and reject Ha.

4.2 Decision and conclusion

Based on the applied test, since not all statistical significance (p) values are less than or equal to the alpha level, the null hypothesis (Hot is rejected and the alternative hypothesis (Hat is accepted. This means that the data do not follow a normal distribution, which is why we chose to use nonparametric methods for the correlation analysis.

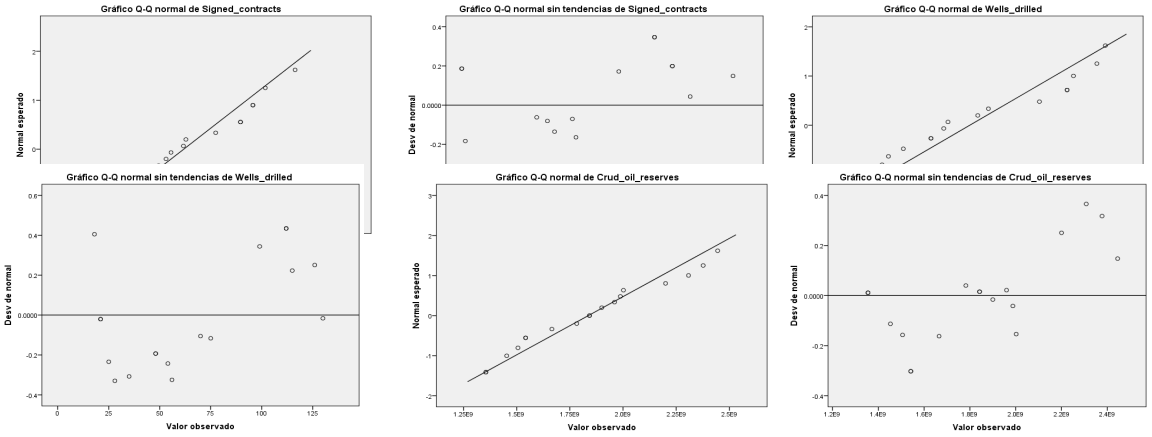

Figure 3. Normality test. Source: author, research results

Correlation analysis

Since the sample does not follow a normal distribution, nonparametric tests are applied to evaluate the correlation.

Statistical test

Spearman's Rho correlation test is used.

Hypothesis statement

Ho: p = 0 (No significant correlation).

Ha: p≠ 0 (There is a significant correlation).

Significance level

A significance level of 5% (α = 0.05t is established.

Decision rule

If p<0.05, the alternative hypothesis (Hat is accepted and the null hypothesis (Ho) is rejected.

If p>= 0.05, the alternative hypothesis (Ha) is rejected and the null hypothesis (Ho) is accepted.

Correlation Classification

0.00 < r ≤ 0.20: Very weak correlation.

0.20 < r ≤ 0.40: Low correlation.

0.40 < r ≤ 0.60: Moderate correlation.

0.60 < r ≤ 0.80: High correlation.

0.80 < r ≤ 1.00: Very high correlation.

Signed Contracts Drilled Wells Crude Oil Reserves LPG Produced LPG Imported LPG Exported LPG G.C Sales G.C Inventory

Table 2: Correlation between variables

			Signed Contracts	Drilled Wells	Crude Oil Reserves	LPG Produced	LPG Imported	LPG Exported	G.C Sales	G.C Inventory
Spearman's Rho	Signed Contracts Signed Contracts	Correlation Coefficient	1.000	.444	-.493*	.230	-.175	-.208	-.016	.291
		Sig. (bilateralt		.065	.038	.357	.487	.408	.951	.241
		N	18	18	18	18	18	18	18	18
	Drilled Wells	Correlation Coefficient	.444	1.000	-.146	-.395	-.238	.148	-.531*	-.288
		Sig. (bilateralt	.065		.563	.105	.341	.557	.023	.246
		N	18	18	18	18	18	18	18	18
	Crude Oil Reserves	Correlation Coefficient	-.493*	-.146	1.000	-.135	-.169	.585*	.007	-.450
		Sig. (bilateralt	.038	.563		.592	.503	.011	.977	.061
		N	18	18	18	18	18	18	18	18
	LPG Produced	Correlation Coefficient	.230	-.395	-.135	1.000	-.433	.126	.748**	.659**
		Sig. (bilateralt	.357	.105	.592		.073	.618	.000	.003
		N	18	18	18	18	18	18	18	18
	LPG Imported	Correlation Coefficient	-.175	-.238	-.169	-.433	1.000	-.382	-.082	-.112
		Sig. (bilateralt	.487	.341	.503	.073		.117	.746	.657
		N	18	18	18	18	18	18	18	18
	LPG Exported	Correlation Coefficient	-.208	.148	.585*	.126	-.382	1.000	-.036	-.268
		Sig. (bilateralt	.408	.557	.011	.618	.117		.886	.282
		N	18	18	18	18	18	18	18	18
	G.C. Sales	Correlation Coefficient	-.016	-.531*	.007	.748**	-.082	-.036	1.000	.294
		Sig. (bilateralt	.951	.023	.977	.000	.746	.886		.236
		N	18	18	18	18	18	18	18	18
	G.C Inventory	Correlation Coefficient	.291	-.288	-.450	.659**	-.112	-.268	.294	1.000
		Sig. (bilateralt	.241	.246	.061	.003	.657	.282	.236	
		N	18	18	18	18	18	18	18	18

*. Correlation is significant at the 0.05 level (bilateralt.

**. Correlation is significant at the 0.01 level (bilateralt.

Source: author, research results.

According to this test, the statistical significance level values (p-value) are lower than alpha in the relationships between the following variables: signed contracts and crude oil reserves; crude oil reserves and exported LPG; sales of Major Marketers and drilled wells. Although the correlation coefficient is moderate, as mentioned above, it is high in the cases of LPG produced with sales of Large Traders and LPG produced with inventory of Large Traders. There is no correlation between contracts signed and wells drilled, which suggests that, although exploration contracts are signed, no wells are drilled that would increase domestic production and, consequently, crude oil reserves. Likewise, there is no correlation between LPG produced, imported and exported.

4.3 Results of the modeling of the Large Marketer link

The analysis of the Large Traders link is reflected in its inventory level, which is influenced by flow variables such as: Volume of LPG produced, Volume of LPG imported, Volume of LPG exported and Sales of Large Traders to wholesale traders. Figure 4 shows its initial conditions.

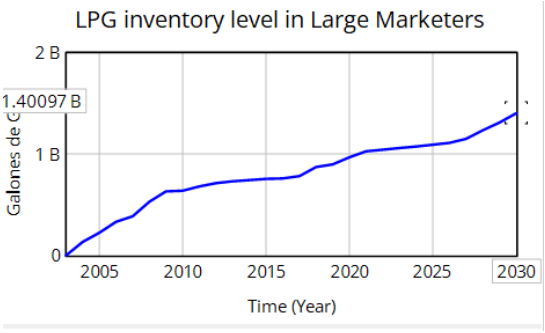

Figure 4. Inventory level of Large Marketers. Source: Unified Information System (SUI)

The first scenario, illustrated in Figure 5, considers a 10% reduction in crude oil reserves due to a decrease in the number of wells drilled. Given this decrease in reserves, it is expected that the extraction rate will also decrease.

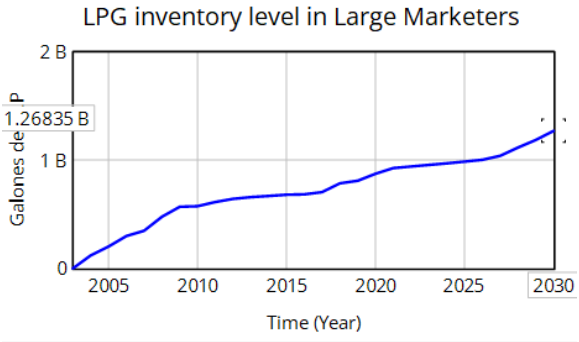

Figure 5. First simulation scenario. Source: author, research results.

Thus, in a second scenario, a 10% decrease in the extraction rate is projected, which generates a lower availability of oil and, therefore, a reduction in the inventory, as shown in Figure 6.

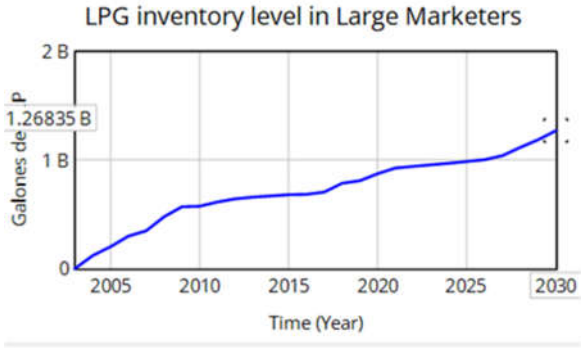

Figure 6. Second simulation scenario. Source: author, research results.

The third scenario evaluates a 20% decrease in the oil refining rate, which directly affects the volume of LPG produced, as shown in Figure 7.

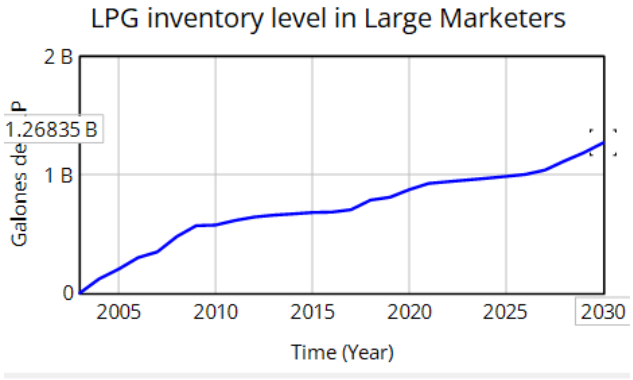

Figure 7. Third simulation scenario. Source: author, research results.

5 Conclusions

In short, Liquefied Petroleum Gas (LPG) plays an important role in Colombia's energy matrix. It is a key source of energy, especially in rural areas and in certain sectors that really depend on its availability and versatility. Although its share in the energy mix is lower compared to other alternatives such as natural gas, coal and electricity, its contribution to energy diversification and security is significant. LPG not only represents a cleaner option compared to traditional fuels, but also contributes to reducing dependence on imports, thus promoting greater energy self-sufficiency for the country.

Big Data analytics in the LPG sector has improved production, distribution and consumption decision making by enabling real-time collection and processing of large amounts of data, which can increase system efficiency and reduce operating expenses, as it more easily allows for tracking consumption patterns across multiple geographies, analyzing distribution infrastructure more effectively and predicting LPG demand, all of which can increase system effectiveness and reduce operating expenses.

The challenges facing the LPG industry, such as distribution infrastructure and competition with other energy sources, require a coordinated response to ensure its sustainability and relevance in the future. Improving infrastructure, developing appropriate policies and implementing strategies to optimize its competitive advantages are key steps to maximize its positive impact, as well as the fact that decision making will be easier with the use of technology based on Big Data and digital media analysis, which will make it possible to anticipate possible bottlenecks in the supply chain.

As Colombia moves towards a more sustainable energy model, Big Data and digital media analysis are essential, so as to more easily identify and recognize the role of LPG as a key transition element, in the same way the problems that demand immediate intervention, such as production, distribution and use, which will not only favor environmental sustainability, but also improve the quality of life of the most vulnerable communities, coming to establish a comprehensive and collaborative approach that promotes the development of LPG in all its aspects, thus ensuring its contribution to economic growth and social welfare of the country.

In general terms, LPG continues to be an important resource within Colombia's energy matrix, especially in sectors where other energy sources have limited access. Its use stands out in rural areas for cooking, water heating and minor industrial activities. In the area of electricity generation, its role is limited because electricity is generated mainly with natural gas and coal; however, in remote regions without access to natural gas, it is used in some thermoelectric plants.

LPG also has a relevant impact in other sectors; in transportation, although its use is lower compared to natural gas, it is used in public and private transportation vehicles as an economical and less polluting alternative to gasoline and diesel, especially in urban areas seeking to reduce costs and emissions; In industry and commerce, it is used in the manufacture of chemicals and plastics, and in commercial activities such as cooking food and heating water in restaurants and hotels, and in rural

areas without access to gas pipeline networks, LPG is essential to meet domestic energy needs such as cooking and heating.

In short, although LPG is not the main source of energy in Colombia, its contribution to the diversification and energy security of the country is crucial, especially in sectors where its accessibility and characteristics make it an indispensable option, establishing that Big Data and digital media analysis facilitate the visualization of production and consumption patterns, crucial to predict the future behavior of the industry.

In effect, energy diversification consists of using different energy sources to avoid excessive dependence on a single option. In Colombia, LPG plays an important role in this diversification in several ways; among which is as an alternative to natural gas, although natural gas is the main energy source in many areas of the country, LPG acts as an essential complement in rural and remote regions where the pipeline network does not reach; as cleaner option compared to liquid fuels, as it is a more sustainable and less polluting alternative to diesel and gasoline. Its use in transportation, industry and commerce helps diversify the energy matrix and reduce the national carbon footprint; from the point of view of domestic use in areas without access to natural gas, LPG is essential for cooking and heating in homes in rural areas, which reduces dependence on more polluting fuels such as coal or firewood.

Now, from the perspective of energy security, LPG strengthens the availability of reliable, affordable and sustainable sources to meet national needs by contributing to the reduction of external dependence, being a by-product of oil refining and natural gas processing, LPG strengthens energy autonomy by expanding the available domestic sources, reducing vulnerability to interruptions in the supply of other fuels, on the other hand to accessibility in remote areas, since in many rural areas and difficult to access, LPG is transported in cylinders or trucks, In addition to this, LPG becomes resilient to emergencies, as it is a key backup source in case of failures in the electricity grid or problems with the natural gas supply, especially during extreme weather phenomena such as droughts that affect hydroelectric generation. Therefore, the LPG industry in Colombia has great potential to reduce dependence on imported fuels and strengthen the country's energy self- sufficiency.

References

[1] Sistema Unico de Información de Servicios Públicos, "SUI," [Online]. Available: https://sui.superservicios.gov.co/. [Last access: 15 August 2022].

[2] Datosmacro, "Datosmacro.com," [Online]. Available: http://datosmacro.expansion.com/energia-y-medio-ambiente/petroleo/produccion/colombia. [Last access: 11 July 2022].

[3] CENIT, "Poliductos," [Online]. Available: http://cenit-transporte.com/poliductos/. [Last accessed: 14 November 2022].

[4] UPME, Cadena del Gas Licuado del Petróleo (GLPt, Bogotá: Unidad de Planeación Minero Energética, 2017.

[5] UPME, "Plan indicativo de abastecimiento de Gas Licuado del Petróleo GLP," Unidad de Planeación Minero Energética, 2019.

[6] GASNOVA, "LPG Annual Report," 2022.

[7] ECOPETROL, "Precios Vigentes," December 12, 2022. [Online]. Available: http://www.ecopetrol.com.co/wps/portal/Home/multisitios/comercial/es/precios/precios-vigentes/precios-vigentes/!ut/p/z1/pZHNDoIwEIShSfotkUsR4r81L8CAcFeDCfSRNGD8fnFxIMoKahz2-Sb3cwsUqhCqq1vuqmv-tzWx27eK-cgvZhhZoNkhDFI3cSf00DwMMGo7AORvwohjWlY5DsaZYQg9YefZ_.

[8] CREG, "How are LPG residential public service tariffs determined" 19 December 2022. [Online]. Available: http://creg.gov.co/publicaciones/15052/como-se-determinan-las-tarifas-del-servicio-publico-domiciliario-de-glp/tutm_source=chatgpt.com.

[9] CREG, "History in Colombia," 23 October 2013. [Online]. Available: http://creg.gov.co/publicaciones/7875/historia-en-colombia/tutm_source=chatgpt.com.

[10] A. Martinez Ortiz, Estudios sobre el impacto de la actividad petrolera en las regiones productoras de Colombia, Bogotá: Fedesarrollo, 2018.

[11] MINENERGÍA, "Liquefied Petroleum Gas," 18 June 2022. [Online]. Available: http://www.minenergia.gov.co/es/misional/hidrocarburos/funcionamiento-del-sector/gas-licuado-de-petr%C3%B3leo/?utm_source=chatgpt.com.

[12] GUIA DEL GAS, "Reality and perspectives of LPG in Colombia," 21 August 2019. [Online]. Available: http://guiadelgas.com/realidad-y-perspectivas-del-glp-en-colombia/tutm_source=chatgpt.com.

[13] CREG, "RESOLUTION 66 OF 2007," July 19, 2007. [Online]. Available: http://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0066_2007.htmtutm_source=chatgpt.com.

[14] COLGAS, "LPG Industry: Benefits of its use," January 18, 2023. [Online]. Available: http://colgas.com/blog/por-que-hacer-el-cambio-a-glp/tutm_source=chatgpt.com.

[15] AIGLP, "Challenges for the future of the LPG industry," 18 February 2022. [Online]. Available: http://aiglp.org/es/artigo/desafios-para-el-futuro-de-la-industria-del-glp/tutm_source=chatgpt.com.

[16] CONTRALORÍA, "Seguridad y confiabilidad energética en Colombia (2010-2022t," 12 September 2022. [Online]. Available: http://www.camara.gov.co/sites/default/files/2025-02/Doc-Seguridad-y-Confiabilidad-Energ%C3%A9tica-en-Colombia.pdf.

[17] UNIGAS, "Matriz energética en Colombia y el aporte del GLP," 2022. [Online]. Available: http://www.unigas.com.co/blog/aportes-del-glp-a-la-matriz-energetica-de-colombia/tutm_source=chatgpt.com.

[18] R. A. Haugen, N. O. Skeie, G. Muller, and E. Syverud, "Detecting emergence in engineered systems: A literature review and synthesis approach," Systems Engineering, vol. 26, no. 4, pp. 463-481, 7 February 2022.

[19] FEDESARROLLO, "Política de regulación de precios de combustibles líquidos para uso automotivo en Colombia," June 17, 2015. [Online]. Available: http://www.repository.fedesarrollo.org.co/bitstream/handle/11445/2432/Repor_Junio_2015_Martinez_Benavides_y_Ramirez.pdf;jsessionid=8CABFE0BD52DA0FB12270EDB8E5F6746?tsequence=3.

[20] N. Luhman, "Modern system theory as a form of complex social analysis," Sociological, vol. 1, no. 1, 2015.

[21] W. Wieser, "Organisms, structures, machines: for a theory of the organism," Organisms, structures, machines: for a theory of the organism, 1972.

[22] M. Madanchian, "The Role of Complex Systems in Predictive Analytics for E-Commerce Innovations in Business Management," Systems, vol. 12, no. 10, 2022.

[23] F. Battiston, E. Amico, and A. Barrat, "The physics of higher-order interactions in complex systems," Nature Physics, vol. 17, 2021.

[24] M. Newman, Networks, New York: Oxford university press, 2018.

[25] S. Camazine, J. L. Deneubourg, G. Theraula, J. Sneyd, and N. R. Franks, Self-Organization in Biological Systems, Princeton University Press, 2020.

[26] F. Heylighen, The science of self-organization and adaptivity, 1999.

[27] l. L. Barabási, How Everything is Connected to Everything Else and What It Means for Business, Science and Everyday Life, New York: 2003.

[28] M. M. Waldrop, Complexity: The emerging science at the edge of order and chaos, New York: Simon and Schuster, 1993.

[29] J. D. Sterman, "System Dynamics: systems thinking and modeling for a complex world," May 29-30,2002. [Online]. Available: http://dspace.mit.edu/bitstream/handle/1721.1/102741/esd-wp-2003-01.13.pdf.

[30] J. H. Holland, Hidden order. How Adaptation Butlds Complexity, Helix Books, 1995.

[31] J. Pallant, SPSS survival manual: A step by step guide to data analysis using IBM SPSS, Routledge, 2020. [32] ASA, "American Statistical Association," [Online]. Available: http://www.amstat.org/. [Last accessed: 05 03 2022].

[33] P. Waldmann, "On the use of the Pearson correlation coefficient for model evaluation in genome-wide prediction," Energy Reports, vol. 9, pp. 2881-2899, 2023.

[34] M. R. Saldaña, "Goodness-of-fit tests to a normal distribution," Occupational Nursing, vol. 6, no. 3, pp 105-114, 2016.