

Escuela de Negocios

Tipo de documento: Tesis de maestría



Master in Management + Analytics

Optimización del Abastecimiento de Gas Natural en una Empresa Midstream mediante Programación Lineal con Parámetros Estimados por Regresión

Autoría: Goldman, Nicolás

Año: 2025

¿Cómo citar este trabajo?

Goldman, N. (2025) "*Optimización del Abastecimiento de Gas Natural en una Empresa Midstream mediante Programación Lineal con Parámetros Estimados por Regresión*". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/13676>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

Master in Management + Analytics

**Optimización del Abastecimiento de Gas Natural en una Empresa
Midstream mediante Programación Lineal con Parámetros Estimados
por Regresión**

ARGENTINA

Nicolas

Goldman

Mayo

2025

Índice General

Resumen	4
Abstract.....	5
1. Introducción	6
1. Contexto.....	6
2. Problema.....	10
3. Objetivo	12
4. Estado del Arte.....	13
5. Contenido de la tesis	14
2. Datos y condiciones del problema	16
1. Datos.....	16
a) Fuentes de Datos, Parámetros y Supuestos	16
b) Datos Operativos Históricos de Planta.....	16
c) Datos Comerciales y de Suministro	18
d) Parámetros Operativos y Restricciones Adicionales	18
e) Uso y Preparación de Datos	18
2. Condiciones del problema	19
a) Disponibilidad de Materia Prima.....	19
b) Capacidades de Procesamiento de las Plantas	19
c) Requerimientos Comerciales y Flexibilidad Operativa	20
3. Análisis de los datos	21
1. Caudales de Producción	22
2. Consumo de Energía.....	23
3. Análisis de Outliers.....	23
4. Correlaciones	27
a) Análisis de multicolinealidad	28
4. Metodología.....	29
1. Introducción y Enfoque Metodológico.....	29
2. Modelo de Predicción de Consumo Energético	29
a) Introducción al modelo predictivo.....	29
b) Objetivos del análisis	29
c) Contexto y relevancia	30
d) Razones para dos niveles de análisis	31
e) Metodología de regresión	31
3. Modelo de Optimización	33
a) Definición de términos	33
b) Consideraciones y Supuestos del Modelo	38

c) Sensibilidades.....	38
4. Herramientas Computacionales	38
5. Resultados y análisis de Sensibilidad	40
1. Resultados de Modelo predictivo.....	40
a) Consumo de gas combustible total vs caudal de cada componente	40
b) Regresión Lineal Múltiple.....	43
c) Modelo de Boosting con XGBoost	44
d) Discusión de Resultados del modelo predictivo	46
2. Resultados del modelo de optimización.....	49
a) Caso base	49
b) Aumento de precios del 20%	49
c) Expansión con limitaciones de proveedores	50
d) Expansión sin limitaciones de proveedores.....	50
e) Aumento de riqueza del 15%	51
f) Disminución de riqueza del 15%	51
g) Sin límites de capacidad de proveedores.....	52
h) Nominación de Etano DOW	52
i) Discusión de resultados.....	54
6. Conclusiones	57
1. Implicaciones estratégicas para Compañía Mega	58
7. Trabajos Futuros.....	59
1. Incorporación del Modelo como Herramienta de Decisión Estratégica y Táctica	59
8. Bibliografía.....	60

Resumen

Esta tesis presenta el desarrollo de un modelo híbrido que combina técnicas de optimización lineal y regresión lineal para apoyar la toma de decisiones en la gestión del abastecimiento de materia prima en el sector midstream del Oil & Gas. Se analiza el caso de Compañía Mega, una empresa conjunta entre YPF, DOW y Petrobras, que opera plantas en Neuquén y Bahía Blanca conectadas por un poliducto. El objetivo principal es maximizar la rentabilidad mediante la optimización de la compra de gas crudo y la predicción del consumo energético a partir de datos históricos, considerando las restricciones operativas y comerciales.

Para ello, primero se realizó un análisis exploratorio de datos históricos (enero 2022–octubre 2024) de consumo de gas combustible, potencia eléctrica y caudales de productos. Se construyeron modelos de regresión (lineal, polinómico y log-log) para estimar el costo energético a partir de los caudales de cada componente. Como ninguno de los modelos utilizados se ajustaba a los datos, se procedió a identificar si las variaciones en el consumo específico de gas combustible tienen un impacto significativo, y se observó que este es relativamente bajo (5–10 % del valor de venta por tonelada) y pueden tratarse como constantes en el modelo de optimización. A continuación, se formuló un problema de programación lineal que determina los volúmenes óptimos de compra a seis proveedores de la Cuenca Neuquina, considerando composiciones de alimentación, restricciones operativas como límites de disponibilidad, capacidades de procesamiento en ambas plantas y compromisos comerciales.

Los resultados revelan que las principales restricciones se encuentran en la capacidad combinada de separación de propano, butano y gasolina natural en Bahía Blanca, operando muy cerca de sus límites. El alivio de estas incrementaría la facturación en 30%, y si además se eleva la restricción de disponibilidad de materia prima, la rentabilidad puede triplicarse. Asimismo, se confirma que el etano no constituye un cuello de botella relevante y que variaciones moderadas en precios internacionales no alteran la combinación óptima de mezcla, aunque impactan directamente en la facturación. A partir de estos hallazgos se deducen las implicaciones estratégicas de inversión: priorizar proyectos de “debottlenecking” en Bahía Blanca, diseñar estrategias de gestión de riesgo de precios y renegociar volúmenes con proveedores clave para sostener la resiliencia de Compañía Mega frente a la volatilidad del mercado.

Abstract

This thesis develops and applies a hybrid analytical framework that integrates multiple linear regression and linear programming to optimize natural gas procurement in the midstream segment of the Oil & Gas industry. Using Compañía Mega—a joint venture between YPF, DOW and Petrobras with separation plants in Neuquén and Bahía Blanca—as a case study, we first model plant energy consumption based on historical data (January 2022–October 2024), finding that variations in fuel-gas usage per tonne contribute only 5–10 % of product value and can be treated as effectively constant in the optimization.

We then formulate a daily linear program that determines optimal purchase volumes from six Neuquén suppliers, subject to composition fractions, supply limits, processing capacities, and commercial commitments (e.g., a 1 600 Tn/d ethane sale requirement) . Base-case results highlight bottlenecks in Bahía Blanca’s combined butane plus natural gasoline and propane plus butane plus natural gasoline capacities—operating at 1 450/1 450 and 3 073/3 100 Tn/d—which, if relaxed, increase daily revenue from 1.6 to 2.1 MM USD; further relief of supplier constraints can raise it to 4.9 MM USD.

Sensitivity analyses confirm that moderate price fluctuations do not alter the optimal blend, though they directly affect revenue, and that ethane is not a binding constraint. Strategic implications include prioritizing debottlenecking investments in Bahía Blanca, enhancing market-risk management, and renegotiating supplier volumes to bolster operational resilience.

1. Introducción

1. Contexto

El sector de midstream de Oil & Gas es una etapa fundamental de la cadena de valor de los combustibles fósiles y sus derivados. Esta fase abarca el procesamiento y transporte del gas natural rico, definido así aquel que contiene altas concentraciones de etano (C₂), propano (C₃), butano (C₄), y pentano y superiores (C₅+). Al propano y butano se los denomina GLP, y a los pentanos y superiores se los denomina gasolina natural. La finalidad del sector midstream es, separar sus componentes y comercializarlos tanto como producto puro para consumo directo, como para la transformación en subproductos industriales, tales como plásticos, lubricantes y fertilizantes. Una característica distintiva de este sector es que la planificación de la producción en las plantas depende estrechamente de la calidad y cantidad de materia prima (gas natural rico) suministrada por los productores, así como de las demandas de consumo de los clientes.

El caso de estudio de esta tesis es Compañía Mega, un joint venture conformado por YPF, DOW y Petrobras. Esta empresa opera una planta separadora en Loma La Lata, Neuquén (Figura 2), una planta fraccionadora en Bahía Blanca (Figura 3) y un poliducto de 600 km que conecta ambas instalaciones con una estación de bombeo intermedia (Figura 4). La planta separadora en Neuquén recibe gas rico proveniente de la Cuenca Neuquina y retiene etano, propano, butano y gasolina natural. Esta mezcla es luego transportada a través del poliducto hacia la planta en Bahía Blanca, donde se realiza la separación final de los componentes puros.

Actualmente, Compañía Mega se encuentra en un contexto dinámico, con importantes oportunidades y desafíos. La planta separadora está ubicada en el corazón de Vaca Muerta, una formación geológica con hidrocarburos no convencionales que abarca varias provincias argentinas. Vaca Muerta es la segunda reserva mundial de gas no convencional y la cuarta en petróleo convencional [1]. Su relevancia estratégica a nivel global radica en sus vastas reservas y en el crecimiento sostenido de su producción, que en 2025 superó el 50 % del gas producido en Argentina, alcanzando los 70 millones de metros cúbicos diarios [2]. Esto posiciona a Argentina como un actor relevante en el mercado energético internacional.

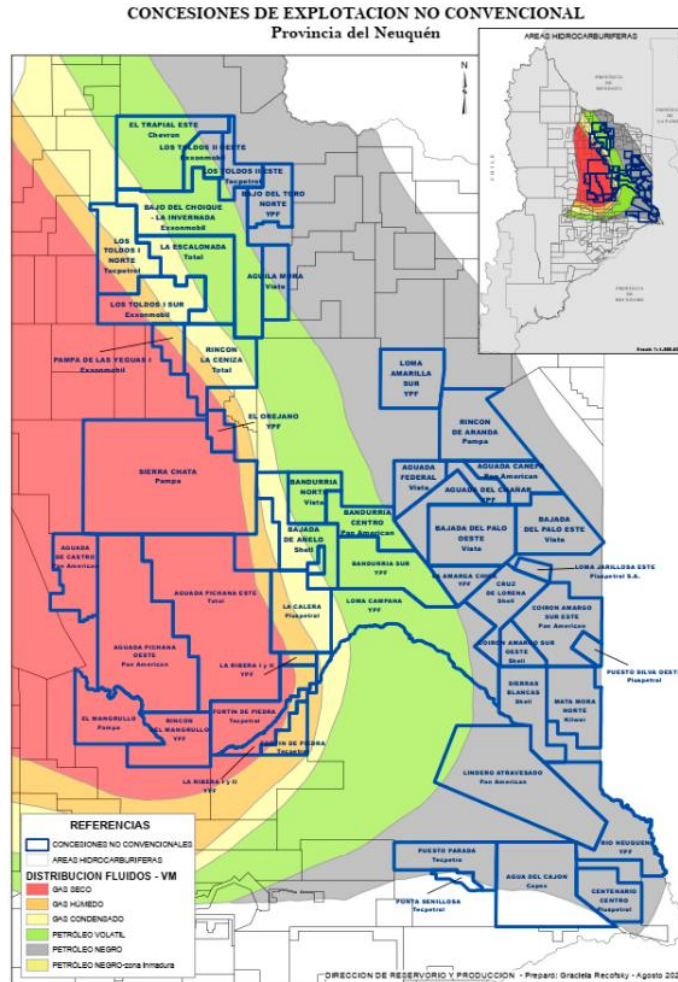


Figura 1: Concesiones de explotación no convencional en la Provincia de Neuquén

Dentro del mapa de Vaca Muerta en la Figura 1 se pueden identificar distintas zonas de producción. Zonas de Gas Seco, Gas Húmedo y Gas Condensado. Las zonas de Gas Húmedo y Gas Condensado son las principales generadoras de gas natural rico, que es de interés para el procesamiento en las plantas de Compañía Mega.

Entre 2019 y 2023, Compañía Mega invirtió aproximadamente 100 millones de dólares destinados a aumentar la capacidad de captación de gas, eliminar cuellos de botella y actualizar tecnologías en las plantas de Neuquén y Bahía Blanca. Entre estas inversiones se destaca la construcción del gasoducto Tratayén-Loma La Lata, con una inversión de 50 millones de dólares, que diversificó las fuentes de gas no convencional procesado por la compañía.

En este contexto, Mega emprendió un proyecto de expansión en dos etapas:

1. Construcción de un nuevo tren de fraccionamiento que se inaugurará en febrero de 2026, incrementando la capacidad de producción en 900 toneladas por día, pasando de 4.700 a 5.600 toneladas diarias.
2. Construcción de dos nuevas estaciones de rebombeo en el poliducto, ubicadas en General Roca (Río Negro) y La Adela (La Pampa), con una finalización prevista para 2028. Estas estaciones permitirán aumentar la capacidad hasta 7.000 toneladas diarias.

La relación con los proveedores (productores) es otro componente clave para Compañía Mega. Actualmente cuenta con una posibilidad de suministro de seis proveedores distintos, cada uno con la posibilidad de aportar gas natural rico procedente de diferentes yacimientos, con variaciones en capacidades de producción y composición del gas. Además, la empresa debe atender los requerimientos comerciales de su accionista DOW, que posee una planta productora de polietileno en Bahía Blanca cuya principal materia prima es el etano suministrado por Compañía Mega. La demanda de DOW es variable y está sujeta a fluctuaciones derivadas de condiciones operativas y factores externos económicos. Además, está expuesta a eventos externos, como las tormentas que afectaron la ciudad de Bahía Blanca en diciembre de 2023 y marzo de 2025 [3], las cuales causaron importantes daños materiales y humanos, impactando negativamente en las operaciones.

Por otra parte, en esta industria de commodities, los ingresos de la empresa están fuertemente condicionados por los precios internacionales. En general, los componentes más pesados de la mezcla tienen un mayor valor de mercado; sin embargo, debido a sus mayores puntos de ebullición, la producción de estos componentes implica costos operativos variables superiores, principalmente vinculados al consumo de gas combustible.

En este marco, surgen dos preguntas clave para el análisis de esta tesis: ¿Es conveniente estudiar de forma analítica el “mix óptimo” de producción para maximizar las ganancias de Compañía Mega? y, en caso afirmativo, ¿existe un compromiso entre maximizar la facturación y minimizar los costos operativos?



Figura 2: Planta de Compañía Mega en Neuquén, cortesía de Compañía Mega



Figura 3: Planta de Compañía Mega en Bahía Blanca, cortesía de Compañía Mega

En Figura 2 y Figura 3, cortesía de Compañía Mega, se observan vistas aéreas de ambas plantas. En la Figura 3, se ven los tanques de propano, butano y gasolina, donde se almacenan hasta su venta, en barcos o camiones.

TRAZA DEL POLIDUCTO MEGA



Figura 4: Ubicación de las plantas y el poliducto en el mapa, cortesía de Compañía Mega

Como se puede observar en la Figura 4, una planta de rebombéo a mitad de camino incrementa la presión del fluido para transportarlo hasta Bahía Blanca.

La Figura 5 muestra un esquema simplificado del flujo de materia en la empresa, desde su ingreso en la planta de Loma La Lata en Neuquén, hasta su entrega a los clientes finales en Bahía Blanca.

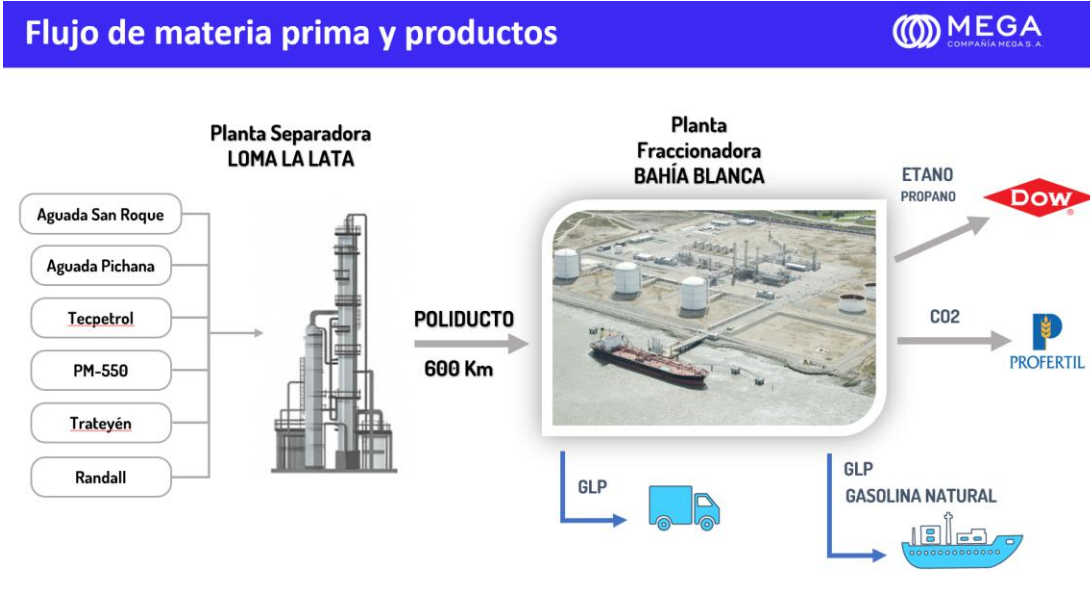


Figura 5: Esquema de flujo simplificado de la empresa, con la alimentación, plantas, poliducto y entrega del producto final. Cortesía de Compañía Mega.

2. Problema

Compañía Mega se encuentra en plena fase de expansión y enfrenta el reto de mantener su rentabilidad en un entorno donde sus ingresos por la venta de productos derivados del gas natural y sus costos operativos están fuertemente interrelacionados. La rentabilidad depende de la diferencia entre la facturación (producto de volúmenes y precios de etano, propano, butano y gasolina natural) y los costos variables, especialmente el consumo de gas combustible en las torres de fraccionamiento.

La empresa procesa gas natural de varios proveedores de la Cuenca Neuquina, a partir del cual produce cuatro hidrocarburos distintos: etano (C2), propano (C3), butano (C4), y gasolina natural (C5+). Estos compuestos están formados por moléculas con 2, 3, 4 y 5 carbonos respectivamente, y a más carbonos, se dice que el compuesto es más "pesado". Este gas se adquiere a proveedores que tienen corrientes con distintas composiciones y caudales disponibles.

A su vez, las composiciones de los proveedores evolucionan con el tiempo según la evolución de la composición de la cuenca. En los últimos años, se observó un incremento en la riqueza de la cuenca (mayor composición de productos más pesados). La empresa no conoce los motivos de esta evolución, y sigue esta variable dado su impacto en los resultados.

La composición molar y el caudal del gas de alimentación determinan la cantidad de producto que se obtiene y se supone que este impacta en el nivel de demanda energética de las torres de separación. Estas torres son etapas sucesivas donde se aíslan los compuestos, aprovechando sus diferentes temperaturas de ebullición para separarlos en corrientes puras. Los componentes más pesados tienen mayor punto de ebullición y requieren más calor para separarse. A pesar de contar con información histórica detallada, aún no se ha realizado un análisis exhaustivo que vincule de forma cuantitativa la mezcla producida con el consumo energético, ni que incorpore este vínculo en la planificación de compras. Por otra parte, el precio de cada producto depende de su cotización internacional, y en general los productos más pesados tienen mayor precio.

Adicionalmente, la compañía dispone de flexibilidad para negociar volúmenes con cada proveedor de la Cuenca Neuquina (el precio es uniforme). Sin embargo, existen restricciones operativas —como los máximos de capacidad de procesamiento y separación en Neuquén y Bahía Blanca— que debemos tener presente para el análisis de optimización. Respecto a la venta de etano, en Argentina, actualmente DOW es el único cliente con una capacidad de utilización limitada por sus plantas de procesos. Para el GLP y la gasolina natural, si bien existe la necesidad de abastecer el mercado interno argentino de GLP, el principal destino es el mercado de exportación, donde existe una mayor amplitud de clientes. Sin embargo, este mercado está fuertemente influenciado por factores externos políticos y económicos.

En este contexto, **el problema a resolver** consiste en determinar los volúmenes óptimos de gas a adquirir de cada productor, de manera que la mezcla resultante alimente las plantas con la composición que maximice la diferencia entre facturación y costos energéticos, respetando simultáneamente todos los límites operativos y comerciales.

En la Figura 5 podemos observar las principales corrientes de abastecimiento a la Planta Separadora Loma La Lata en cuenca neuquina. Dichas corrientes contienen aportantes de los principales bloques de producción de gas no convencional (NOC) del área Vaca Muerta:

- Aguada Pichana: Aguada Pichana Este (yacimiento NOC). Proveedores: YPF, Total, WIAR, PAE.

- Tecpetrol (Proveedor): Fortin de Piedra (yacimiento NOC).
- PM 555-0, Tratayen y Randall: son los puntos principales de inyección a Planta Separadora Loma La Lata, entre los yacimientos NOC aportantes podemos mencionar:
 - Loma Campana
 - La Amarga Chica
 - Bandurrias Sur
 - La Calera
 - El Orejano
 - La Ribera
 - Bajada de Añelo
 - Aguada de la Arena

Teniendo en cuenta el detalle de puntos de entrega mencionados, se realiza el análisis de los volúmenes óptimos de inyección para maximizar los beneficios de Compañía Mega.

3. Objetivo

El propósito de este trabajo es diseñar e implementar un **modelo híbrido** —que combine un módulo predictivo de consumo energético y un programa de optimización lineal— para definir la **estrategia óptima de compra de gas natural** en Compañía Mega, maximizando la rentabilidad operativa. Concretamente, el modelo debe:

1. **Analizar** la relación entre la composición de la mezcla de gas de entrada y el consumo de gas combustible y electricidad en las torres de fraccionamiento, a partir de datos históricos de alta frecuencia.
2. **Estimar** los ingresos potenciales por la venta de etano, propano, butano y pentanos según precios internacionales de mercado.
3. **Formular** un programa lineal que integre estos ingresos y los costos energéticos, sujeto a:
 - Capacidades de procesamiento en Neuquén y Bahía Blanca.
 - Requerimientos comerciales (p. ej., nominaciones contractuales de etano con DOW).
 - Límites de suministro de cada proveedor de la Cuenca Neuquina .

Desafíos principales

- **Calidad y consistencia de los datos:** Los registros de caudales, composiciones y consumos energéticos presentan vacíos, valores atípicos y cambios de calibración, por lo que es esencial aplicar técnicas de limpieza, filtrado y validación rigurosas.
- **Modelado de no linealidades y multicolinealidad:** Capturar interacciones físicas complejas (presión, temperatura, reflujo) requiere explorar transformaciones, regresiones robustas o incluso algoritmos de machine learning avanzados.
- **Escalabilidad y simplicidad del modelo de optimización:** Integrar todas las restricciones operativas y comerciales en un programa lineal interpretable y eficiente.
- **Integración predictivo-prescriptiva:** Traducir los coeficientes del modelo de consumo energético en parámetros de la función objetivo del modelo Lineal, garantizando coherencia y estabilidad.
- **Análisis de sensibilidad y escenarios:** Evaluar la robustez de la solución frente a variaciones de precios, nominaciones de etano y disponibilidad de proveedores, de modo de ofrecer recomendaciones resilientes ante la volatilidad del mercado.

Con este enfoque, la tesis responderá a la pregunta:

¿Qué volúmenes y composiciones de gas natural debe adquirir Compañía Mega a cada proveedor para maximizar su rentabilidad, considerando el impacto de la mezcla resultante en los costos energéticos y las restricciones operativas y comerciales?

4. Estado del Arte

Este capítulo presenta una revisión de la literatura académica y técnica relevante para el desarrollo de esta tesis. Se divide en dos secciones que abordan los pilares metodológicos del trabajo: la aplicación de la programación lineal para la optimización en la industria de *Oil & Gas* y el uso de modelos de regresión para la predicción del consumo energético.

a. Optimización con Programación Lineal en Oil & Gas

La optimización en el sector de petróleo y gas busca mejorar la eficiencia y maximizar las ganancias. La programación lineal (PL) es una de las herramientas matemáticas más consolidadas para resolver problemas de planificación, logística y producción bajo restricciones lineales.

Un caso clásico es la mezcla de crudos en refinerías. Como cada tipo de petróleo difiere en calidad, precio y disponibilidad, el desafío es determinar la proporción que optimiza el margen, cumpliendo las especificaciones del producto final. Por ejemplo, Hegazy et al. [4] diseñaron un modelo de PL que considera restricciones de calidad y producción, así como exigencias de especificaciones en las gasolinas, y considerando limitaciones químicas la estabilidad de la mezcla y la precipitación de impurezas. Con este enfoque, evaluaron distintos escenarios de abastecimiento y demanda, demostrando cómo la PL puede guiar decisiones de compra y producción en un entorno cambiante.

La versatilidad de la PL va más allá del blending. Enekwe C.B. et al. [5] aplicaron un enfoque similar para reducir la importación de crudo en la producción de aceites lubricantes en Nigeria. En este trabajo, trasladamos esos principios al problema de abastecer gas natural desde diversos yacimientos de la Cuenca Neuquina, buscando el “mix” más rentable para las plantas de separación y fraccionamiento, dentro de los compromisos comerciales existentes.

b. Predicción del Consumo Energético mediante Regresión

En la industria de procesos, el gasto energético supone una parte sustancial de los costos operativos. Poder anticipar ese consumo en función de las variables de producción ayuda a planificar, controlar presupuestos y optimizar procesos.

Los modelos de regresión lineal son una herramienta sencilla pero potente para establecer la relación entre el consumo energético (variable dependiente) y factores como el caudal de producción o las condiciones ambientales (variables independientes). Por ejemplo, Dao et al. aplicaron regresión múltiple para estimar el consumo eléctrico y de gas en un supermercado del Reino Unido, incorporando temperatura y humedad, y lograron explicar buena parte de la variabilidad con un nivel de predicción aceptable. Además, exploraron casos que incluyen sistemas de distribución eléctrica en Filipinas y datos de IoT, confirmando la utilidad de estos modelos para distintos contextos.

En esta tesis, se adopta esa misma lógica para predecir el consumo de energía de las plantas midstream. Al relacionar el gasto energético con variables operativas clave, se pueden incorporar las predicciones al modelo de optimización.

5. Contenido de la tesis

Datos y condiciones del problema

En esta sección se describen las fuentes de datos operativos e comerciales, los parámetros clave y los supuestos que definen el problema de abastecimiento de gas natural. Se detallan las características de los seis proveedores de la Cuenca Neuquina, los límites de procesamiento de las plantas de Neuquén y Bahía Blanca, así como las restricciones comerciales y operativas que se incorporarán al modelo de optimización.

Análisis de los datos

Aquí se realiza un análisis exploratorio de los datos históricos, incluyendo limpieza de valores atípicos, tratamiento de datos faltantes y evaluación de la normalidad. Se presenta el estudio de correlaciones y el diagnóstico de multicolinealidad para asegurar la solidez de los modelos estadísticos.

Metodología

Se introduce el enfoque híbrido que combina machine learning y programación lineal. Se explica la estrategia para desarrollar el modelo predictivo de consumo energético y su integración en la formulación del modelo de optimización lineal, detallando el flujo de trabajo y las herramientas computacionales usadas.

Modelo predictivo

En este capítulo se construye y valida el modelo de regresión —lineal, polinómico, log y log-log— que estima el consumo de gas combustible y vapor en función de los caudales de etano, propano, butano y pentanos. Se presentan métricas de ajuste y criterios de selección de la especificación final.

Modelo de optimización

Se formula el problema de programación lineal completo: definición de variables de decisión, parámetros (fracciones máxicas, precios, factores de conversión), función objetivo (maximización de la rentabilidad) y restricciones (capacidades, requerimientos de etano, límites de suministro). Se explica cómo incorporar los coeficientes del modelo predictivo al costo energético.

Resultados y análisis de sensibilidad

Se muestran los resultados de la solución óptima bajo el caso base y diversos escenarios: variaciones de precios internacionales, relajación de cuellos de botella y cambios en la disponibilidad de proveedores y demandas de los clientes. Se discute la robustez de la solución y su comportamiento ante cambios en las condiciones de mercado y operativas.

Conclusiones

Se resumen los hallazgos principales, destacando las implicaciones estratégicas para Compañía Mega, recomendaciones de inversión y gestión de riesgos.

Trabajos futuros

Se proponen posibles extensiones del modelo, como la incorporación de dinámicas de inventario, escenarios estocásticos de precios y mejoras en la modelización de transporte en el poliducto, para ampliar y profundizar los resultados obtenidos.

2. Datos y condiciones del problema

Esta tesis combina optimización matemática y modelos de regresión. La empresa cuenta con varios parámetros para la optimización:

- Composición y caudales potenciales del gas de cada proveedor
- Costos de materia prima por proveedor
- Precios internacionales actuales y proyectados
- Requerimientos operativos de las plantas
- Requerimiento mínimo de producción por producto
- Límites de compra máxima y mínima a proveedores
- Costos unitarios de electricidad y gas combustible

1. Datos

a) Fuentes de Datos, Parámetros y Supuestos

El desarrollo de los modelos predictivos y de optimización, así como el análisis exploratorio de datos (EDA), se basa en un conjunto diverso de información proporcionada por Compañía Mega y fuentes externas especializadas. A continuación, se detallan las principales fuentes de datos, parámetros clave y supuestos relevantes.

b) Datos Operativos Históricos de Planta

Se dispone de datos operativos detallados y registrados en alta frecuencia de las dos plantas de procesamiento principales: Loma La Lata (Neuquén) y Bahía Blanca. Estos datos históricos son la base fundamental para el EDA y para el entrenamiento y validación del modelo de Machine Learning de predicción de consumo energético. El período principal cubierto es **enero de 2022 a octubre de 2024**, salvo que se especifique lo contrario.

i. **Planta Loma La Lata (Neuquén):**

- **Consumo de Gas Combustible:** Registros en m³/h cada 15 minutos (Ene 2022 – Oct 2024), detallando el consumo de turbinas (210, 220, 230) y hornos (360, 050).
- **Potencia Eléctrica Consumida:** Datos en KW cada 15 minutos, disponibles para un período anterior y más acotado (Abr 2021 – Sep 2022).
- **Caudal Instantáneo y Composición del Gas:** Mediciones en Sm³/d y % molar respectivamente, registradas cada 5 minutos (Ene 2022 – Oct 2024). Estos datos corresponden específicamente a la corriente de *salida* de la planta.

ii. Planta Bahía Blanca:

- **Disponibilidad:** Datos disponibles consistentemente de Ene 2022 a Oct 2024 para las variables principales.
- **Consumo de Gas Combustible:** Registros en m³/h cada 15 minutos.
- **Potencia Eléctrica Consumida, Caudal Instantáneo y Composición del Gas:** Mediciones en KW, Sm³/d y % molar respectivamente, registradas con alta frecuencia (cada 5 minutos).

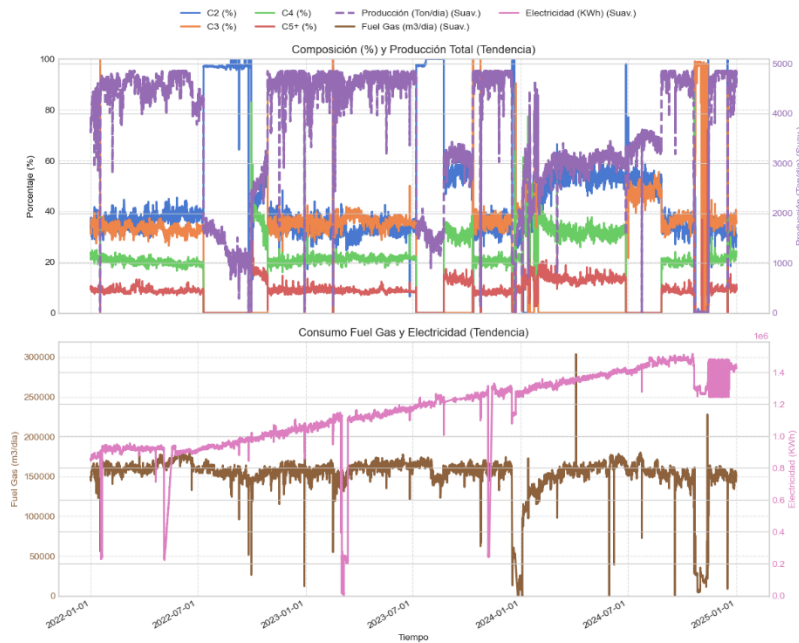


Figura 6: Composición, producción total y consumo de fuel gas y electricidad en Bahía Blanca en el período de tiempo estudiado

iii. Interconexión y Supuestos Asociados:

Las plantas están interconectadas mediante un poliducto que permite el transporte de productos entre ellas, con un tiempo estimado de tránsito de aproximadamente una semana. Cada una de las plantas dispone de cromatógrafos que permiten realizar mediciones independientes de los productos transportados.

Para simplificar el análisis y evitar la complejidad asociada al modelado del flujo dinámico en el poliducto, se asumirá un transporte instantáneo entre las plantas. Este supuesto simplificador busca facilitar la construcción y resolución del modelo, sin comprometer la validez de los resultados.

A diferencia de Loma La Lata, en Bahía Blanca los datos de gas combustible, composición y caudal están disponibles para la corriente de salida de cada etapa de producción, permitiendo un análisis potencialmente más granular. Existen también registros adicionales de temperaturas y presiones que no se consideran primordiales para el alcance inicial de este análisis.

Aunque ambas plantas (Neuquén y Bahía Blanca) participan en el proceso de separación, la tesis se centra en los datos de Bahía Blanca debido a que la composición de la mezcla tiene un

impacto significativamente mayor en los costos energéticos y las restricciones operativas de esta planta, dado su enfoque en el fraccionamiento final de componentes más pesados (C_3 , C_4 , C_5^+).

c) Datos Comerciales y de Suministro

Estos datos son parámetros fundamentales para el modelo de optimización lineal (PL).

- **Características de Proveedores:** Se cuenta con la composición másica característica (y_{ij}) y el caudal máximo disponible (Q_i^{\max} en Tn/d) para cada uno de los 6 proveedores de gas natural (San Roque, Pichana, Tecpetrol, PM 550-0, Tratayen, Randall). Estos valores se detallan en la Tabla 1.
- **Precios de Productos (P_j):** Los precios internacionales de los productos finales (C_2 , C_3 , C_4 , C_5^+). utilizados en el modelo provienen de **proyecciones proporcionadas por la consultora especializada S&P Global**. Se utilizarán valores promedio o representativos de estas proyecciones para el período de optimización. Los valores específicos se detallan en la Tabla 2.
- **Costo de Materia Prima (Gas Crudo):** Se considera un costo uniforme de **3.2 USD/MMBTU** para el gas crudo procesado, independientemente del proveedor. Debido a esta uniformidad, este costo no se incluye en la función objetivo del PL, aunque sí influye en la rentabilidad global y se considera implícitamente al limitar la disponibilidad (Q_i^{\max}).
- **Precio de Gas Combustible:** Se dispone del precio de referencia del gas utilizado como combustible en planta, fijado en **3 USD/MMBTU**. Este valor es relevante para la valoración económica del consumo de gas combustible, ya sea directamente o como input para la estimación de los coeficientes del costo energético (C_E).

d) Parámetros Operativos y Restricciones Adicionales

Complementando los datos anteriores, el modelo de Programación Lineal requiere parámetros específicos que definen los límites y reglas operativas:

- **Límites de Capacidad de Planta:** Capacidades máximas de procesamiento de Neuquén (L_{Neuquen}) y límites de producción/separación (individuales y combinados) de Bahía Blanca ($L_{\text{Prod...}}$, $L_{\text{Comb...}}$).
- **Factores de Conversión ($r_c(i)$):** Utilizados para estimar el C1 asociado a cada corriente de entrada en la restricción de Neuquén, Tabla 6.
- **Requerimientos Específicos:** Obligaciones comerciales como la venta de una cantidad fija de Etano a DOW (Req_{C2}).

Los valores numéricos específicos de estos parámetros se presentan y detallan junto con la formulación matemática del modelo en la Sección 3.

e) Uso y Preparación de Datos

Los datos operativos históricos serán sometidos a un proceso de limpieza y preparación estándar (tratamiento de outliers, valores faltantes si aplica) antes de ser utilizados en el Análisis Exploratorio de Datos y para entrenar el modelo de Machine Learning. Los datos comerciales y parámetros operativos se utilizarán directamente para configurar el modelo de Programación Lineal.

2. Condiciones del problema

a) Disponibilidad de Materia Prima

Llamaremos *proveedor* a cada cuenca de donde proviene el gas, y *corriente* al caudal proveniente de las seis corrientes (San Roque, Pichana, Tecpetrol, PM 550-0, Tratayen y Randall). En la práctica, algunas cuencas son operadas por una misma empresa.

Cada una de las seis corrientes de alimentación de gas natural tiene un caudal máximo diario disponible, medido en toneladas por día. La cantidad de gas procesada desde cada fuente no debe superar este límite máximo. Esta información se expone en la Tabla 1.

Asumiremos que hay flexibilidad comercial para exigir a los proveedores el caudal deseado, desde 0 Tn/d hasta el caudal máximo que puede entregar cada uno. Para simplificar la notación, asignamos un número a cada corriente.

Corriente	Nº de corriente	Q^{max} (Tn/d)	Fracción C2	Fracción C3	Fracción C4	Fracción C5+
SAN ROQUE	1	341	0,44	0,30	0,18	0,08
PICHANA	2	1968	0,56	0,25	0,14	0,05
TECPETROL	3	4837	0,57	0,25	0,14	0,04
PM 550-0	4	3547	0,50	0,27	0,15	0,07
TRATAYEN	5	1376	0,46	0,25	0,17	0,12
RANDALL	6	1539	0,45	0,31	0,19	0,05

Tabla 1: Fracciones y caudales máximos de las corrientes que llegan a la Planta de Neuquén

b) Capacidades de Procesamiento de las Plantas

Las instalaciones presentan restricciones operativas clave:

i. Límite de Procesamiento en Neuquén

En la planta de Neuquén, el tope diario de procesamiento se sitúa en 31.000 toneladas equivalentes de C1+, un valor que integra todos los componentes desde C2 en adelante más una estimación de metano asociado. Para convertir los flujos de C2+ a su equivalente en metano, se emplean factores $r_c(i)$ que reflejan la relación entre el caudal total entrante y el volumen de C2+ retenido. Se suman primero los aportes de cada proveedor excluyendo C1, N2 y CO2; a continuación, ese caudal retenido de C2+ se divide entre el flujo completo de C1+ enviado, resultando en el factor $r_c(i)$ de conversión. Multiplicando cada flujo de C2+ por su $r_c(i)$ y sumando los productos se obtiene el total diario de C1+ procesado en Neuquén, asegurando que nunca supere las 31.000 toneladas.

ii. Restricciones de Producción/Separación en Bahía Blanca

En Bahía Blanca, por su parte, las limitaciones diarias responden a la capacidad de separación, almacenamiento y transporte. Allí se fija un techo independiente para la producción de etano (C2), propano (C3) y butano (C4), y además se imponen dos restricciones combinadas: una para la suma de butano y gasolina natural (C4 + C5+) y otra que abarca propano, butano y gasolina natural (C3 + C4 + C5+). De este modo, la planta garantiza un equilibrio operativo eficiente en todos sus procesos.

c) Requerimientos Comerciales y Flexibilidad Operativa

Algunos compromisos comerciales y condiciones operativas específicas también deben cumplirse:

- **Compromiso de Venta de Etano**

Es obligatorio producir exactamente 1600 toneladas diarias de Etano (C2) para cumplir con un contrato específico con DOW, lo que impone una restricción de igualdad.

- **Posibilidad de Venteo de Etano**

A diferencia de otros productos (C3, C4, C5+), Mega tiene flexibilidad para “rechazar” o ventear etano en la planta de Neuquén, devolviéndolo junto con el metano a través de la cañería troncal. DOW aprovecha esta flexibilidad operativa para regular sus requerimientos de etano según sus propias necesidades.

3. Análisis de los datos

Se realizará un análisis descriptivo de las variables consideradas más relevantes para la investigación indicadas en la Figura 7, utilizando los conjuntos de datos disponibles. Este análisis inicial permitirá obtener una comprensión general de las características principales de las variables de interés.

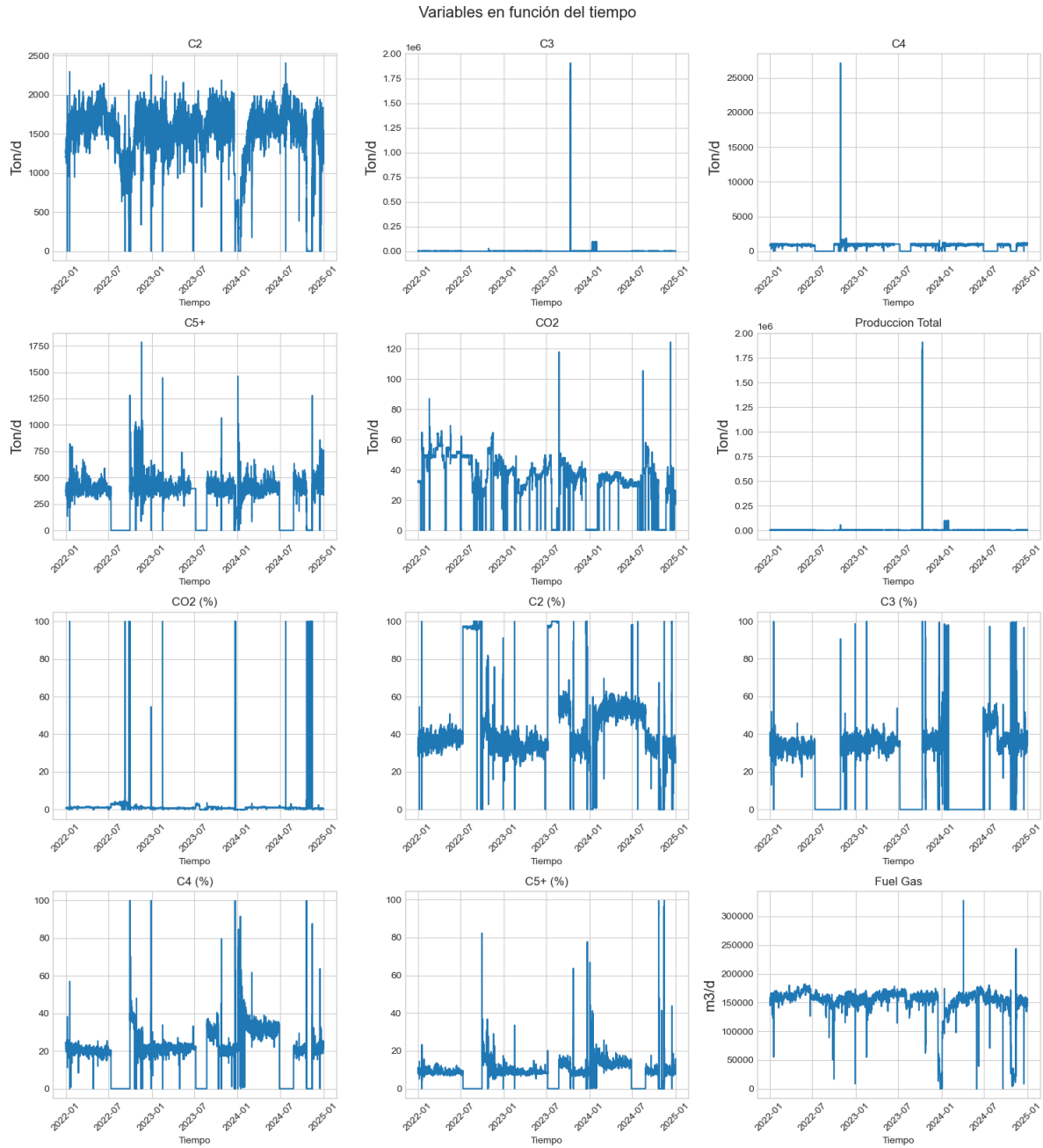


Figura 7: Variables en la Planta de Bahia Blanca (datos crudos)

1. Caudales de Producción

Las condiciones operativas de la planta se caracterizan por una relativa estabilidad, lo que se traduce en un volumen de producción de cada componente que generalmente se mantiene dentro de rangos definidos a lo largo del tiempo. Los caudales de producción promedio se presentan en la Tabla 2.

Componente	Caudal Promedio (Tn/d)
C2 (Etano)	1600
C3 (Propano)	1580
C4 (Butano)	1000
C5+ (Pentano y +)	450
CO2 (Dióxido Carbónico)	2

Tabla 2: Producción promedio de Compañía Mega en Bahía Blanca, por producto

No obstante, la observación detallada de los gráficos en la Figura 7 revela la presencia de fluctuaciones significativas respecto a estos valores medios, incluyendo tanto aumentos como disminuciones abruptas. Esta variabilidad en la producción a lo largo del tiempo es el resultado de la interacción de diversos factores operativos y comerciales.

Entre los factores operativos, se destaca la susceptibilidad de la planta a interrupciones por causas tanto internas como externas. Eventos climáticos adversos, como temporales en Bahía Blanca, pueden ocasionar la detención de las operaciones, ya sea por daños directos a la infraestructura de la planta o a las instalaciones de sus clientes. Adicionalmente, las paradas programadas y no programadas de la planta generan interrupciones totales o parciales en los caudales de producción.

Desde una perspectiva comercial, la producción de etano (C2) se ve influenciada por las demandas de DOW, lo que ocasiona variaciones en su caudal a lo largo del tiempo. Asimismo, la composición de la mezcla producida está directamente relacionada con la proporción en la que se adquiere la materia prima de los proveedores.

En el caso del etano (C2), se aprecia una estabilidad general, donde las caídas pronunciadas pueden asociarse a paradas de la planta. Las fluctuaciones menores alrededor del promedio probablemente responden a ajustes en el caudal requerido por DOW.

Los caudales de propano (C3) y butano (C4) presentan valores atípicos (outliers) de magnitud considerable, lo que podría indicar errores de medición y dificulta la interpretación de la variabilidad en el resto de los datos representados gráficamente.

El pentano y superiores (C5+) y el dióxido de carbono (CO2) muestran patrones de comportamiento más similares al etano (C2), aunque sus variaciones se atribuyen principalmente a factores operativos, dado que no están sujetos a requerimientos comerciales directos.

La producción total se ve afectada significativamente por las distorsiones observadas en la producción de propano (C3) y butano (C4), lo que impide una interpretación clara de la tendencia general con los datos en su estado actual.

En contraste, el análisis de los porcentajes de producción permite una interpretación más sencilla: el etano (C2) y el propano (C3) representan aproximadamente el 35% cada uno, el butano (C4) alrededor del 20%, y el pentano y superiores (C5+) el 10% restante, con una proporción mínima correspondiente al dióxido de carbono (CO2).

2. Consumo de Energía

El consumo energético de la planta se manifiesta tanto en forma de energía eléctrica como de gas combustible.

La energía eléctrica se obtiene de la red externa, medida a través del contador A+B, y se complementa con la producción de un generador interno. El consumo total de electricidad se calcula como la suma de ambas fuentes.

El consumo de gas combustible se mide a nivel global. Sin embargo, para facilitar la modelización del consumo en función del caudal de cada producto procesado, se decidió analizar el consumo específico en cada una de las torres de separación.

Se observa una tendencia al aumento en el consumo de energía eléctrica a lo largo del tiempo, mientras que el consumo de gas combustible se ha mantenido relativamente constante. Adicionalmente, el indicador de consumo de gas combustible por tonelada producida resulta relevante para evaluar el impacto de esta variable en los costos operativos.

3. Análisis de Outliers

Al inspeccionar los histogramas de la Figura 7, surgieron valores atípicos muy marcados —principalmente ceros o cifras cercanas a cero— que no obedecen a la operación normal de la planta. Estas anomalías suelen corresponder a periodos de inactividad o a posibles fallos en los instrumentos de medida, cuando la producción cae a cero o los sensores registran lecturas erróneas. Gracias a la claridad que ofrecen los histogramas, podemos identificar de forma rápida qué variables presentan estas desviaciones y, al mismo tiempo, señalar cuáles se beneficiarían

de una normalización. Al transformar adecuadamente esas distribuciones, ganamos en interpretabilidad y obtenemos una imagen más precisa del comportamiento habitual de la planta.

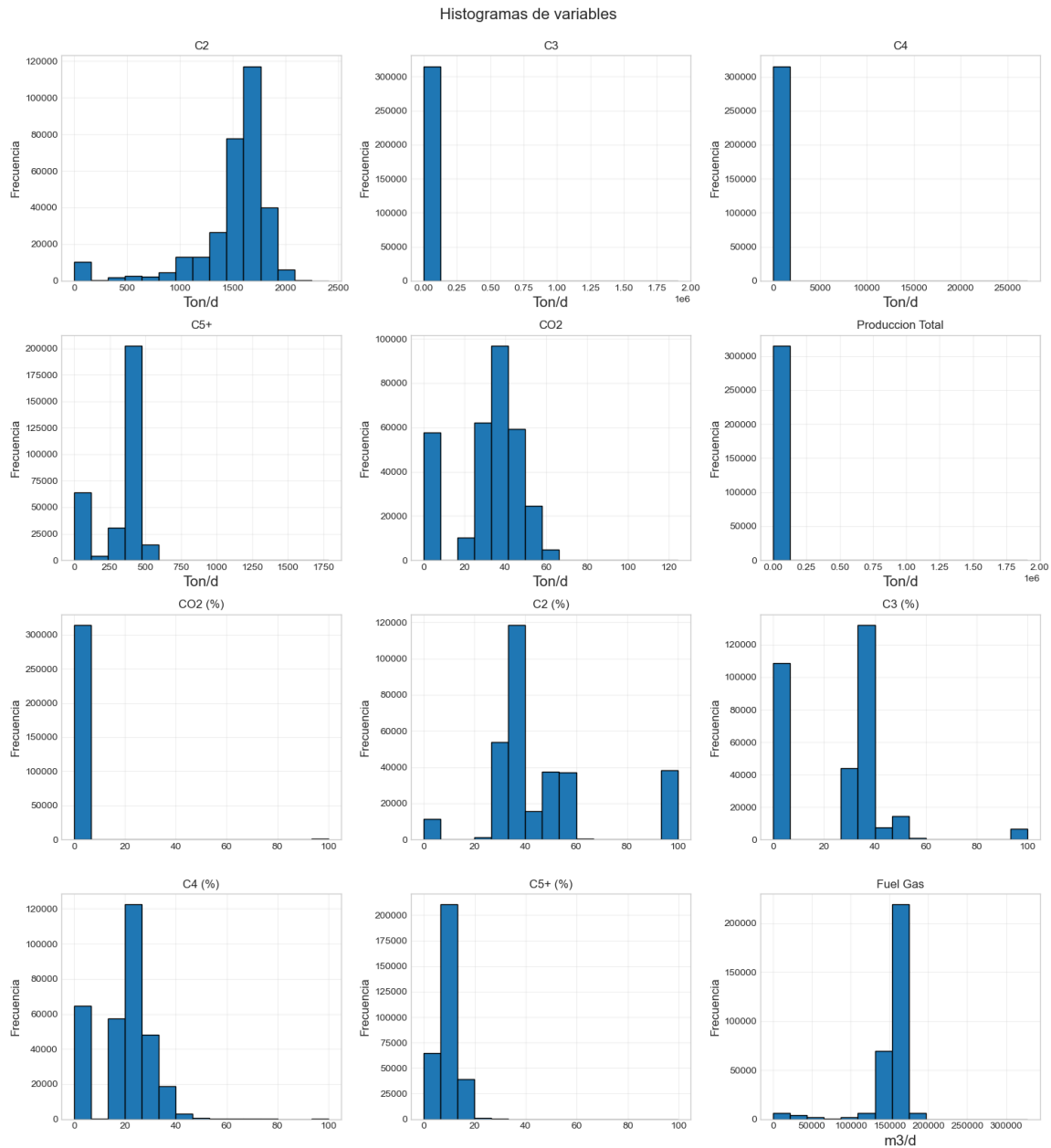


Figura 8: Histogramas de las variables analizadas (datos crudos)

Posteriormente, se aplicó un preprocesamiento a los datos, restringiéndolos mediante los siguientes criterios:

- Producción total ≥ 500 Tn/d
- Producción total < 6000 Tn/d
- C3, C4, C5+ > 100 Tn/d

Estos umbrales se establecieron con el objetivo de asegurar que los datos considerados correspondan a períodos de operación normal de la planta, pudiendo incluso aplicarse criterios más restrictivos. Adicionalmente, se eliminaron los datos correspondientes a los quintiles extremos de la distribución de cada variable para suavizar las series temporales.

En la Figura 9 se observan los gráficos de la evolución de las variables relevantes con estos filtros aplicados. Las líneas indican períodos de tiempo donde se eliminaron los datos, por no cumplir con las condiciones indicadas antes.

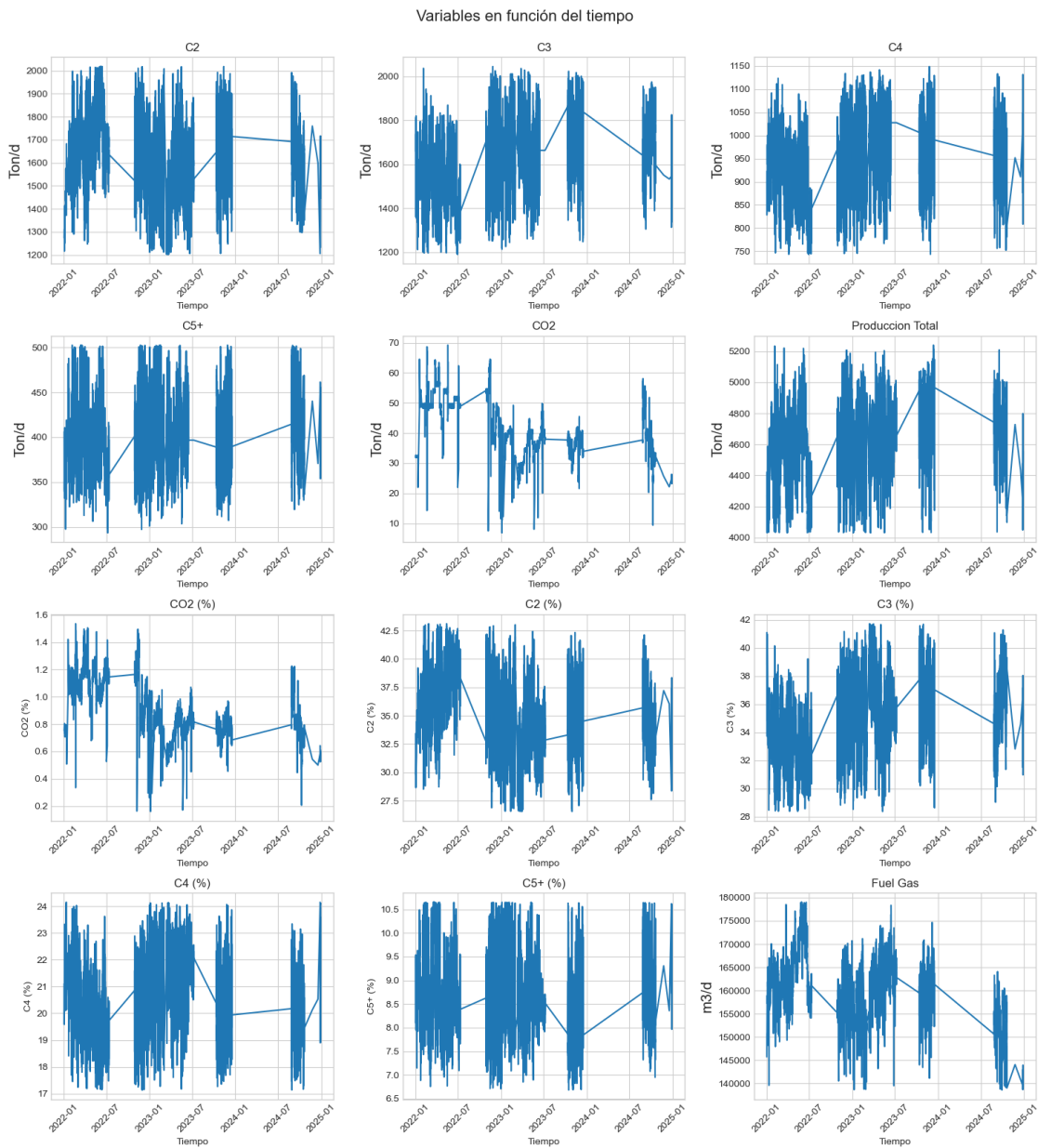


Figura 9: Variables en la Planta de Bahia Blanca (sin outliers)

Esta técnica de suavización permite una mejor visualización de los datos y facilita la identificación de aquellas paradas de planta que tuvieron un impacto prolongado en la producción.

Histogramas de variables

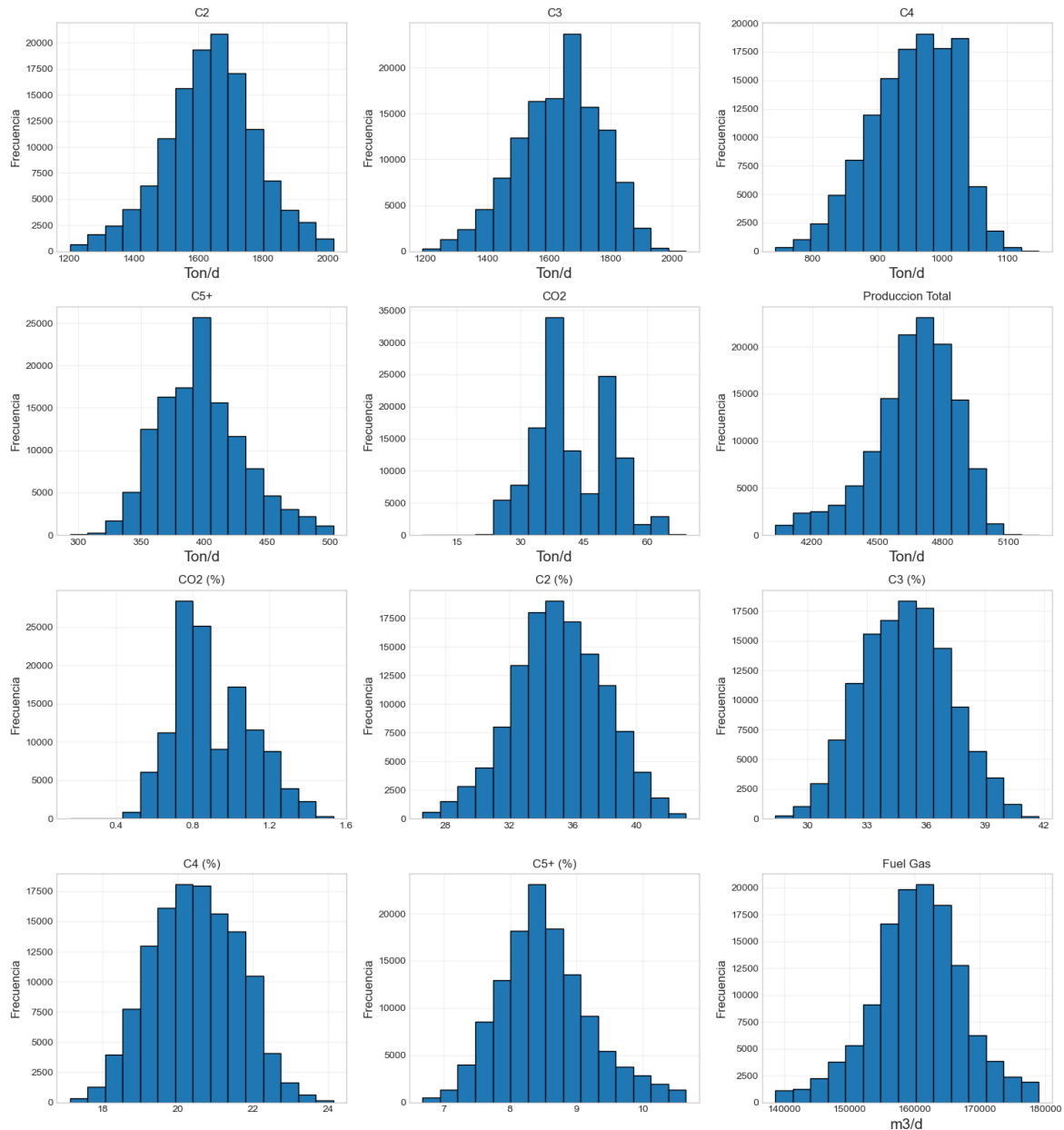


Figura 10: Histogramas de las variables analizadas (sin outliers)

Además, los histogramas resultantes (Figura 10) indican que la mayoría de las variables analizadas presentan distribuciones aproximadamente normales después de la aplicación de este preprocesamiento.

Dado que el objetivo principal es desarrollar un modelo para relacionar la producción de cada componente y el consumo energético, los coeficientes del modelo se aplicarán a la variable "caudal" de cada uno de los componentes. A continuación, se analizarán las correlaciones.

4. Correlaciones

Se elaboró una matriz de correlaciones (Figura 11), con el fin de identificar la posible existencia de multicolinealidad entre las variables consideradas en el estudio.

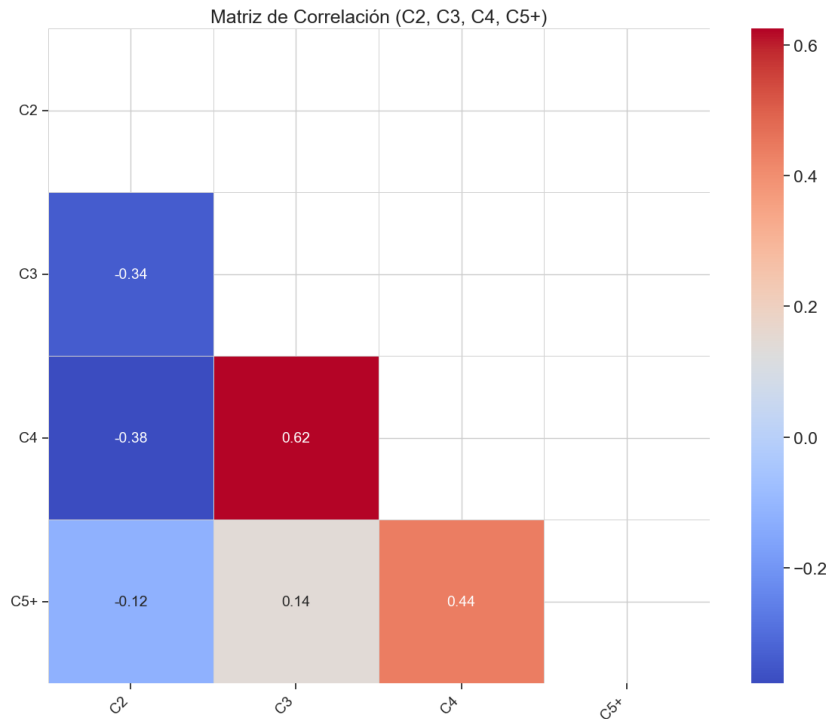


Figura 11: Matriz de correlaciones para las variables de caudales de cada producto

Entre las correlaciones más destacadas se encuentran las siguientes:

- **C3 y C4:** Correlación positiva fuerte (0.62), lo que sugiere una tendencia a que estas variables aumenten o disminuyan de manera conjunta.
- **C2 y C4:** Correlación negativa moderada (-0.38), indicando una tendencia a que cuando una de estas variables aumenta, la otra disminuya.
- **C3 y C5+:** Correlación negativa significativa (-0.57).
- **C5+ y C2:** Correlación casi nula (0.05), lo que sugiere una escasa o nula relación lineal entre estas variables.

a) Análisis de multicolinealidad

La multicolinealidad ocurre cuando dos o más variables independientes están altamente correlacionadas, lo que dificulta la interpretación de los coeficientes y puede reducir la precisión de las predicciones.

Variable	VIF
const	637
C2	1.19
C3	1.74
C4	2.17
C5+	1.29

Tabla 3: Resultado del análisis de multicolinealidad

Los resultados obtenidos (Tabla 3) indican que todos los valores de VIF se encuentran por debajo del umbral de 5, lo cual sugiere que la multicolinealidad es mínima y no representa un problema significativo para los modelos desarrollados. Este hallazgo permite asegurar que los coeficientes estimados son confiables y que las conclusiones derivadas del modelo de regresión serán más robustas y útiles para los análisis posteriores.

En la matriz de correlación se observó una relación significativa entre las variables C3 y C4. Esta correlación se justifica desde una perspectiva física, ya que las corrientes de hidrocarburos más pesados ("ricas") tienden a contener proporciones elevadas de ambos componentes. Por lo tanto, la concurrencia de valores altos o bajos para estas variables en los datos no resulta inesperada.

A pesar de esta correlación, ambas variables se incluyeron en el modelo debido a su relevancia individual para explicar el fenómeno estudiado y porque el análisis de VIF no reveló problemas de multicolinealidad significativos (ambos valores se sitúan por debajo del umbral crítico de 5). Además, cada variable aporta información única y relevante para comprender el comportamiento del sistema en relación con el consumo energético.

En resumen, si bien existe una correlación entre C3 y C4, su inclusión conjunta en el modelo se considera justificada tanto por fundamentos físicos como estadísticos, ya que no distorsionan los resultados y contribuyen a una explicación más completa del consumo energético.

4. Metodología

1. Introducción y Enfoque Metodológico

El objetivo central de esta tesis es determinar la estrategia óptima de compra de gas natural a los distintos proveedores de la Cuenca Neuquina para maximizar la rentabilidad de Compañía Mega. Dada la complejidad del problema, que involucra la interacción entre la composición de la materia prima, los costos operativos y los ingresos por venta de productos finales, se propone un enfoque metodológico híbrido que combina técnicas de Machine Learning (ML) y Optimización Matemática (Programación Lineal).

Este enfoque se estructura en dos componentes principales:

1. Un **modelo predictivo (ML)** para estimar el consumo energético de la planta en función de las características del gas procesado.
2. Un **modelo de optimización (PL)** que utiliza la predicción de costos energéticos y considera las restricciones operativas y comerciales para determinar la mezcla óptima de compra de materia prima.

El flujo de trabajo consistirá en desarrollar y validar primero el modelo predictivo de consumo energético utilizando datos históricos. Luego, la función o los parámetros de costo energético derivados de este modelo se integrarán en el modelo de optimización lineal para encontrar la estrategia de compra que maximice la rentabilidad neta.

2. Modelo de Predicción de Consumo Energético

a) Introducción al modelo predictivo

Reducir el consumo de energía es un objetivo clave en la industria del procesamiento de gas y petróleo. Entre las distintas secciones de una planta, la de fraccionamiento —donde se separan las corrientes de hidrocarburos en sus componentes básicos— destaca por su elevada demanda energética. Cada grado de pureza y cada punto de ebullición diferente requieren calor y frío controlados, lo que se traduce en un uso intensivo de *Gas Combustible* (Fuel Gas en inglés) en los rehervidores y de energía para el condensador.

En este capítulo desarrollaremos un modelo para predecir cuánta energía se consume en la sección de fraccionamiento, partiendo de la composición de la alimentación. A través de un enfoque riguroso pero accesible, mostraremos cómo la proporción de hidrocarburos más livianos influye en la necesidad de energía, y ofreceremos un análisis sistemático de los resultados.

b) Objetivos del análisis

En este análisis se propone comprender cómo variaciones en la composición de la corriente de alimentación afectan el consumo de gas combustible. Para ello, se comenzará estableciendo una relación directa entre el consumo de gas combustible y el volumen diario de etano (C2), propano (C3), butano (C4) y gasolina (C5+), de modo de cuantificar el aporte de cada fracción a la demanda energética global. A continuación, se focalizará en las torres de Deetanización, Depropanización y Debutanización para entender cómo el caudal específico que alimenta cada unidad influye en su propio consumo de vapor (que se genera quemando gas combustible), permitiéndonos entender si hay alguna relación entre el caudal producido y alguna variable de

consumo energético.

Con este fin es que evaluamos distintos modelos matemáticos: en particular probaremos con regresión lineal, polinomio de segundo grado, transformaciones logarítmicas y análisis log–log. Al comparar la calidad de ajuste y la capacidad predictiva de cada modelo, podremos identificar cuál resulta más adecuado para anticipar el comportamiento energético en función de la composición de la alimentación.

El producto final será una función que relacione directamente los caudales finales de los productos principales (F_{C2} , F_{C3} , F_{C4} y F_{C5+}) con C_E , reconociendo que otros indicadores —como el caudal total o composiciones intermedias— podrán añadirse si demuestran poder predictivo durante el análisis exploratorio.

El punto de partida será una regresión lineal múltiple de la forma

$$C_E = a + b F_{C2} + c F_{C3} + d F_{C4} + e F_{C5+}$$

Si se confirma la relación lineal, la función resultante será evaluada con métricas estándar (R^2 , MAE) y validación cruzada. Si el desempeño sugiere no linealidad, se explorarán alternativas para modelar el sistema y aproximarlos a un modelo lineal. El modelo entrenado sobre datos depurados (sin outliers) proporcionará un conjunto de coeficientes que se integrarán directamente en la función objetivo del modelo de optimización lineal para actualizar de forma precisa el costo energético.

c) Contexto y relevancia

Las torres de destilación separan los componentes aprovechando sus diferentes puntos de ebullición:

- **Etano (C2)**: punto de ebullición alrededor de $-89\text{ }^\circ\text{C}$.
- **Propano (C3)**: $-42\text{ }^\circ\text{C}$.
- **Butanos (C4)**: entre $-12\text{ }^\circ\text{C}$ y $0\text{ }^\circ\text{C}$.
- **Pentanos y superiores (C5+)**: desde $36\text{ }^\circ\text{C}$ en adelante.

En el rehervidor (reboiler) se suministra calor, quemando gas combustible, para vaporizar la fracción más volátil; en el condensador se remueve calor para recuperar el producto más liviano. La **facilidad de separación** depende de la diferencia en volatilidad: cuantos más componentes con puntos de ebullición cercanos, más reflujo y más energía se necesitan.

En una configuración de torres en serie, **en la primera torre se separa por el tope el etano**, aprovechando su bajo punto de ebullición. La corriente de fondo de esta torre, que contiene los componentes más pesados, pasa a la siguiente. **En la segunda torre se separa el propano por el tope**, siendo este el siguiente componente más ligero. Finalmente, **en la tercera torre se obtiene el butano por el tope**, mientras que por el fondo se recupera la gasolina (pentanos y componentes superiores, C5+), que son los de mayor punto de ebullición.

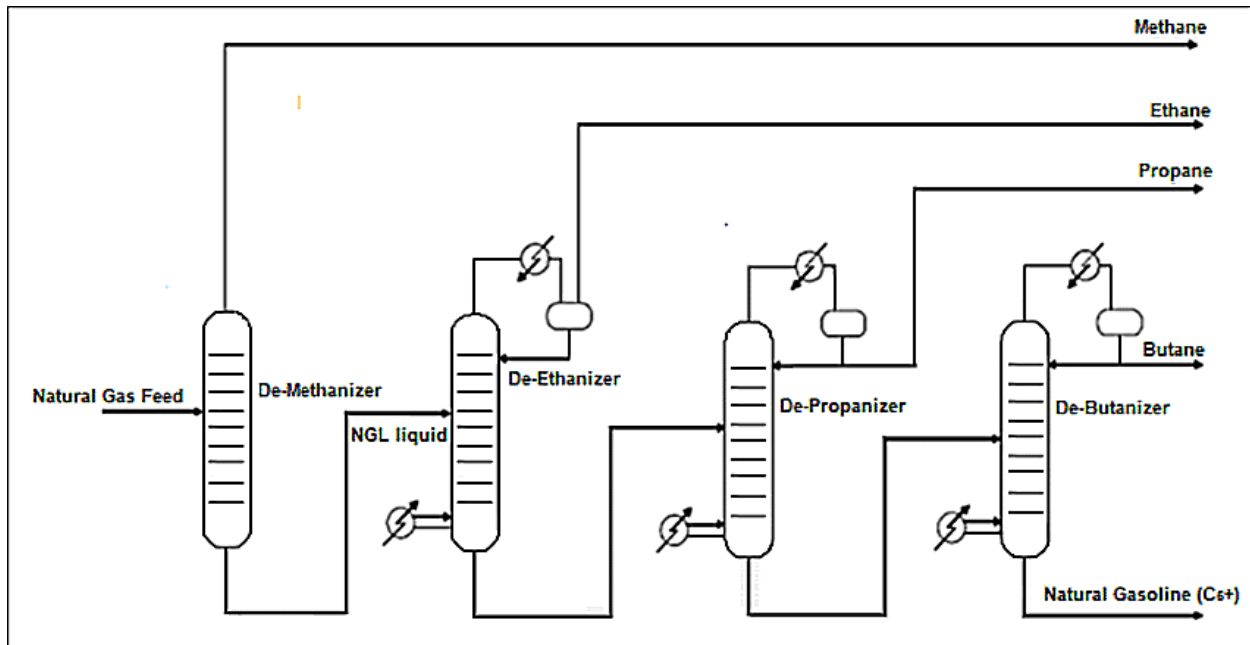


Figura 12: Esquema de un proceso de fraccionamiento de gas natural [6]

La Figura 12 muestra un modelo esquemático de una planta de separación. La separación del metano en Mega ocurre en Neuquén, mientras que los demás componentes se separan en Bahía Blanca.

d) Razones para dos niveles de análisis

Para obtener una visión completa, realizaremos dos estudios paralelos:

1. **Gas Combustible total vs. caudal de cada componente**

En este análisis se examina la relación entre el consumo total de gas combustible y los caudales de cada componente, de modo de captar el comportamiento energético de la planta en su conjunto y comprender la influencia del consumo de unidades auxiliares o de otras variables externas.

2. **Consumo de vapor por torre vs. caudal de su componente**

Este enfoque se centra en el consumo de vapor de cada torre de fraccionamiento en función del caudal del producto específico que procesa (C2, C3 o C4), lo que permite aislar el impacto de la composición de otras variables que afectan el consumo total de gas combustible.

Al comparar ambos resultados, podremos elegir el modelo que mejor prediga el consumo de gas combustible.

e) Metodología de regresión

Para desarrollar un modelo predictivo preciso del consumo de gas combustible en las torres de destilación, considerando la inherente complejidad y naturaleza a menudo no lineal de estos

procesos de separación, se exploró la incorporación de **transformaciones** a las variables predictoras (caudales de hidrocarburos) y a la variable respuesta ('Gas Combustible').

Estas transformaciones permiten que la estructura del modelo lineal capture patrones más complejos, ofreciendo una representación más fiel de la dinámica del proceso:

1. **Polinomios de Segundo Grado:** La inclusión de términos cuadráticos (X^2) permite modelar relaciones con curvatura. Esto es particularmente relevante ya que la eficiencia energética de una torre (y, por lo tanto, su consumo de Gas Combustible) no necesariamente varía de forma lineal con los caudales de alimentación; su comportamiento puede acelerarse o desacelerarse en diferentes regímenes operativos.
2. **Transformaciones Logarítmicas (Log):** La aplicación de la función logaritmo a las variables (predictoras o respuesta) sirve para modelar efectos proporcionales (cuando un cambio porcentual en una variable de entrada tiene un impacto consistente, absoluto o porcentual, en el consumo de gas combustible), y desde el punto de vista estadístico, para estabilizar la varianza de los residuos.
3. **Especificación Log-Log:** Al aplicar transformaciones logarítmicas tanto a las variables predictoras como a la variable respuesta, el modelo resultante permite estimar directamente la **elasticidad**. Esta métrica cuantifica el cambio porcentual en el consumo de gas combustible ante un cambio del uno por ciento en el caudal de un componente específico, y proporciona una medida directa y muy útil de la sensibilidad energética del proceso a las variaciones en los caudales de alimentación.

Se evaluarán comparativamente los modelos que incorporan estas transformaciones para seleccionar aquel que no solo logre un ajuste estadístico óptimo a los datos observados, sino que también ofrezca una **interpretación físicamente plausible** del comportamiento del sistema y cuyos parámetros sean **estadísticamente robustos**.

Para cada relación (total o torre–componente) aplicaremos:

- **Regresión lineal (grado 1):** tendencia directa.
- **Regresión polinómica (grado 2):** posibles curvaturas.
- **Regresión logarítmica** ($\log(y)$ vs x): efecto de escalas de energía.
- **Regresión log-log** ($\log(x)$ vs $\log(y)$): elasticidad porcentual.

Calcularemos los coeficientes, el intercepto y R^2 el de cada modelo, identificando cuál (si alguno) ofrece un poder predictivo razonable.

Adicionalmente, posteriormente ampliaremos el análisis a un **modelo de regresión lineal múltiple**, considerando simultáneamente los caudales de C2, C3, C4 y C5+ como variables predictoras del consumo de gas combustible. En caso de que esta especificación múltiple no alcance un nivel de ajuste adecuado, implementaremos **XGBoost** como técnica de **boosting**, reconociendo que, aunque su naturaleza no paramétrica impide su inclusión directa en un modelo de optimización lineal, proporcionará una estimación de cuánto del consumo de gas combustible puede ser explicado por el conjunto de caudales producidos. De esta forma, ganaremos una visión complementaria de la contribución relativa de cada componente al consumo energético total.

3. Modelo de Optimización

El modelo de optimización busca determinar los volúmenes de gas más convenientes a adquirir de cada proveedor, con el objetivo de maximizar la rentabilidad del sistema. Para ello, se plantea como un problema de programación lineal que integra:

1. **Costo energético:** los costos se calculan a partir de los caudales de productos obtenidos, empleando los resultados del modelo de regresión desarrollado.
2. **Disponibilidad y límites de suministro:** cada corriente de gas incorpora restricciones de capacidad y especificaciones de composición, garantizando que no se superen los volúmenes máximos disponibles.
3. **Capacidades operativas:** se modelan para las plantas de procesamiento (Neuquén y Bahía Blanca) sus límites operativos.
4. **Requisitos comerciales:** se incluyen compromisos como la venta de etano a DOW.
5. **Parámetros confiables:** fracciones de composición, costos unitarios y límites de producción se extraen de fuentes históricas, comerciales y técnicas.

Con esta estructura, el modelo brindará recomendaciones de compra que combinan las capacidades de los proveedores, las restricciones operativas y los requerimientos del mercado, maximizando la rentabilidad global.

a) Definición de términos

iii. Conjuntos e Índices:

- $I = \{1, \dots, 6\}$: Conjunto de proveedores (corrientes de alimentación), indexado por i .
- $J = \{C2, C2vent, C3, C4, C5+\}$: Conjunto de componentes relevantes de hidrocarburos, indexado por j , incluyendo la fracción de etano que se vendea.

iv. Parámetros:

- y_{ij} : Fracción másica (dato conocido) del componente j en la corriente del proveedor i [-]. Ver Tabla 4.
- Q_i^{max} : Caudal máximo disponible (límite de suministro) del proveedor i [Tn/d]. Ver Tabla 4.
- P_j : Precio de venta estimado para el producto final asociado al componente j [USD/Tn]. Ver Tabla 5.
- a, b, c, d, e : Coeficientes de la función lineal de costo energético, obtenidos del modelo de regresión [a: USD/d; b, c, d, e: USD/Tn].
- $r_c(i)$: Factor de conversión para estimar el C1 asociado al flujo de C2+ del proveedor i [-]. Ver Tabla 6.
- $L_{Neuquen}$: Límite máximo de procesamiento en Neuquén, considerando C1 estimado (= 31,000 Tn/d).
- $L_{Prod_C2}, L_{Prod_C3}, L_{Prod_C4}$: Límites máximos de producción/separación para C2, C3, C4 en Bahía Blanca (= 2100, 1900, 1100 Tn/d, respectivamente).
- $L_{Comb_C4C5}, L_{Comb_C3C4C5}$: Límites máximos combinados de producción/separación en Bahía Blanca (= 1450 y 3100 Tn/d, respectivamente).
- Req_{c2} : Requerimiento obligatorio de venta de Etano a DOW (= 1600 Tn/d).

Corriente	Nº de corriente	Q_i^{max} (Tn/d)	Fracción C2	Fracción C3	Fracción C4	Fracción C5+
SAN ROQUE	1	341	0,44	0,30	0,18	0,08
PICHANA	2	1968	0,56	0,25	0,14	0,05
TECPETROL	3	4837	0,57	0,25	0,14	0,04
PM 550-0	4	3547	0,50	0,27	0,15	0,07
TRATAYEN	5	1376	0,46	0,25	0,17	0,12
RANDALL	6	1539	0,45	0,31	0,19	0,05

Tabla 4: Características de las Corrientes de Alimentación

La Tabla 4 contiene las corrientes de alimentación disponibles ($i=1\dots 6$), su caudal máximo diario (Q_i^{max}) y la fracción másica de cada componente (y_{ij}) en cada corriente. Esta información es clave para determinar cuánto material podemos comprar de cada proveedor y qué proporción de C2, C3, C4 y C5+ aporta al modelo.

Producto (j)	P_j (USD/Tn)
C2	225
C3	380
C4	420
C5+	560

Tabla 5: Precios estimados de los productos

La Tabla 5 presenta el precio de venta estimado (P_j) de cada producto final (C2, C3, C4 y C5+). Estos valores son aproximaciones, porque los precios varían constantemente, pero permiten calcular los ingresos totales en la función objetivo, determinando la contribución económica de cada tonelada producida. En la práctica, los precios de todos los hidrocarburos están correlacionados, aunque no siempre en la misma proporción.

	SAN ROQUE	PICHANA	TECPETROL	PM-550	TRATAYEN	RANDALL
$r_c(i)$	3.53	3.26	2.54	3.35	3.01	1.06

Tabla 6: Factores de Conversión $r_c(i)$

La Tabla 6 muestra los coeficientes para estimar el caudal de C1 asociado al flujo de C2+ de cada proveedor. Se utilizan en la restricción de límite de procesamiento en Neuquén para convertir los flujos de alimentación en un equivalente de C1+.

Restricción	Parámetro	Capacidad
Capacidad máxima para C ₂	$L_{Prod_C2} L_{Prod_C2}$	2100
Capacidad máxima para C ₃	$L_{Prod_C3} L_{Prod_C3}$	1900
Capacidad máxima para C ₄	$L_{Prod_C4} L_{Prod_C4}$	1100
Capacidad combinada C ₄ y C ₅₊	$L_{Comb_C4C5} L_{Comb_C4C5}$	1450
Capacidad combinada C ₃ , C ₄ y C ₅₊	$L_{Comb_C3C4C5} L_{Comb_C3C4C5}$	3100

Tabla 7: Restricciones de capacidad

En la Tabla 7 se definen las capacidades máximas de producción y separación en Bahía Blanca:

- L_{Prod_C2} , L_{Prod_C3} , L_{Prod_C4} : máximos diarios para cada fracción.
- L_{Comb_C4C5} , L_{Comb_C3C4C5} : límites combinados que reflejan la capacidad conjunta de separación de varias fracciones en la misma planta.

Estas limitaciones responden a la capacidad de diseño del cuello de botella del proceso, que son las torres de destilación

v. Variables de Decisión:

- Q_i : Caudal de gas a comprar al proveedor i [Tn/d].

vi. Variables Calculadas:

- z_{ij} : Caudal del componente j aportado por el proveedor i [Tn/d]. permite calcular la distribución de los distintos componentes de hidrocarburos provenientes de las corrientes de alimentación.
- F_j : Caudal total del producto final j obtenido [Tn/d].
- C_E : Costo energético total estimado [USD/d]. Esta variable refleja el volumen total de cada componente (C2, C3, C4, C5+) producido, sumando las contribuciones de todos los proveedores. Esta variable se calcula utilizando el modelo de regresión que tiene en cuenta el consumo energético asociado a la producción de los distintos productos.
- F_{C2vent} : Representa la cantidad de etano que se ventea (no se vende, se devuelve al troncal). Esta variable es útil para entender el caudal del etano no aprovechado.

vii. Función Objetivo:

Maximizar el Beneficio Neto

$$\max \left(F_{C_2} \cdot P_{C_2} + F_{C_3} \cdot P_{C_3} + F_{C_4} \cdot P_{C_4} + F_{C_{5+}} \cdot P_{C_{5+}} - C_E \right) \quad (4.1)$$

Versión simplificada de la función objetivo

$$\max \left(\sum_{j \in \{C_2, C_3, C_4, C_{5+}\}} (F_j \cdot P_j) - C_E \right) \quad (4.2)$$

viii. Restricciones:

$$z_{ij} = Q_i \cdot y_{ij} \quad \forall i, j \quad (4.3)$$

$$F_j = \sum_{i=1}^6 z_{ij} \quad \forall j \neq C_2 \quad (4.4)$$

$$C_E = a + b \cdot F_{C_2} + c \cdot F_{C_3} + d \cdot F_{C_4} + e \cdot F_{C_{5+}} \quad (4.5)$$

$$Q_i = \sum_{j \in \{C_2, C_3, C_4, C_{5+}\}} z_{ij} \quad \forall i \quad (4.6)$$

$$F_{C_2vent} = \sum_{i=1}^6 z_{C_2} - F_{C_2} \quad (4.7)$$

$$Q_i \leq Q_i^{max} \quad \forall i \in I \quad (4.8)$$

Es restricción asegura que no se compre más materia prima de la disponible en cada proveedor.

$$z_{ij} \geq 0 \quad \forall i \in I \quad (4.9)$$

Esta restricción garantiza que los flujos de componentes sean no negativos.

Límite de Procesamiento Neuquén:

$$\sum_{j \in \{C_1, C_2, C_3, C_4, C_{5+}\}} \sum_{i=1}^6 z_{ij} (r_c(i) + 1) \leq L_{Neuquen} \quad (4.10)$$

Convierte cada flujo de alimentación a su equivalente en C1 y limita la carga total de procesamiento diaria en la región de Neuquén.

Límites Producción/Separación Bahía Blanca:

$$F_{C_2} \leq L_{Prod_C2} \quad (4.11)$$

$$F_{C_3} \leq L_{Prod_C3} \quad (4.12)$$

$$F_{C_4} \leq L_{Prod_C4} \quad (4.13)$$

$$F_{C_4} + F_{C_{5+}} \leq L_{Comb_C4C5} \quad (4.14)$$

$$F_{C_3} + F_{C_4} + F_{C_{5+}} \leq L_{Comb_C3C4C5} \quad (4.15)$$

Estas restricciones aseguran que la producción o separación de cada fracción no supere la capacidad nominal de la planta en Bahía Blanca.

Requerimiento Comercial Etano:

$$F_{C_2} = \text{Req}_{C_2} \quad (4.16)$$

Garantiza el suministro mínimo obligatorio de etano a DOW, cumpliendo con el contrato de venta.

b) Consideraciones y Supuestos del Modelo

La validez de los resultados de esta metodología está sujeta a ciertas consideraciones y supuestos clave:

- **Linealidad:** Se asume una relación lineal entre el costo energético y los caudales de producto (o se logra una linealización/aproximación adecuada). El modelo de optimización es lineal.
- **Parámetros Constantes:** Se asume que las composiciones de entrada (y_{ij}), los precios de venta (P_j), los límites de disponibilidad (Q_i^{max}) y las capacidades de planta son constantes durante el período de optimización considerado (ej. diario, semanal).
- **Transporte:** Se asume transporte instantáneo o se considera que los tiempos de residencia no afectan las decisiones de compra optimizadas en la escala de tiempo del modelo (como se discutió previamente).
- **Precisión de Datos:** La calidad de los resultados depende de la precisión de los datos de entrada (composiciones, precios, límites, datos históricos para ML).
- **Agregación:** El modelo opera a un nivel agregado (ej. diario), simplificando fluctuaciones intra-diarias.

c) Sensibilidades

Para poner a prueba la solidez del modelo, se realizarán varios análisis de sensibilidad:

- Variación de precios internacionales de los productos (+20 %), para evaluar su impacto en la facturación y en la mezcla óptima.
- Relajación de cuellos de botella en Bahía Blanca, analizando cómo cambian los resultados al aumentar las capacidades de separación.
- Modificación de los límites de suministro de cada proveedor, identificando qué corrientes ejercen mayor influencia sobre la solución óptima.
- Variación de riqueza de la cuenca, para entender el impacto de cambios en la cromatografía en la solución óptima y la facturación.
- Ajustes en la nominación obligatoria de etano, para medir la sensibilidad de la rentabilidad ante distintos niveles de compromiso contractual.

4. Herramientas Computacionales

El análisis de datos, el desarrollo del modelo de ML y la implementación y solución del modelo de programación lineal se realizarán utilizando el lenguaje de programación Python (versión 3.9.21) y sus librerías científicas estándar, tales como:

- Pandas: Para manipulación y análisis de datos.
- NumPy: Para operaciones numéricas.
- Matplotlib / Seaborn: Para visualización de datos y resultados.

- Scikit-learn: Para el desarrollo y evaluación de modelos de Machine Learning.
- CPLEX: Para la formulación y solución del modelo de Programación Lineal.
- Openpyxl: Para exportar los resultados a planillas de Excel

El código se correrá en una computadora Lenovo Yoga 7 16IAP7, que posee un procesador Intel Core i5-1240P, de 12va generación (16 núcleos de 1,7 GHz), 8 GB de RAM y tarjeta gráfica Intel Iris Xe Graphics. El tiempo de ejecución de la regresión lineal múltiple es de 36,8 milisegundos, mientras que el de el método de boosting es de 758 milisegundos.

5. Resultados y análisis de Sensibilidad

1. Resultados de Modelo predictivo

En las siguientes secciones presentaremos los resultados gráficos y tabulados de cada regresión, seguidos de su interpretación. Asumiremos como aceptable un R^2 superior a 0,7, que indica que el 70% de la variación en la variable dependiente se explica por variaciones en la variable independiente. Finalmente, extraeremos conclusiones sobre la viabilidad de predecir el consumo energético a partir de la composición de alimentación.

a) Consumo de gas combustible total vs caudal de cada componente

En esta sección, se analizará la relación entre el consumo de Gas Combustible y la producción de cada producto de manera individual.

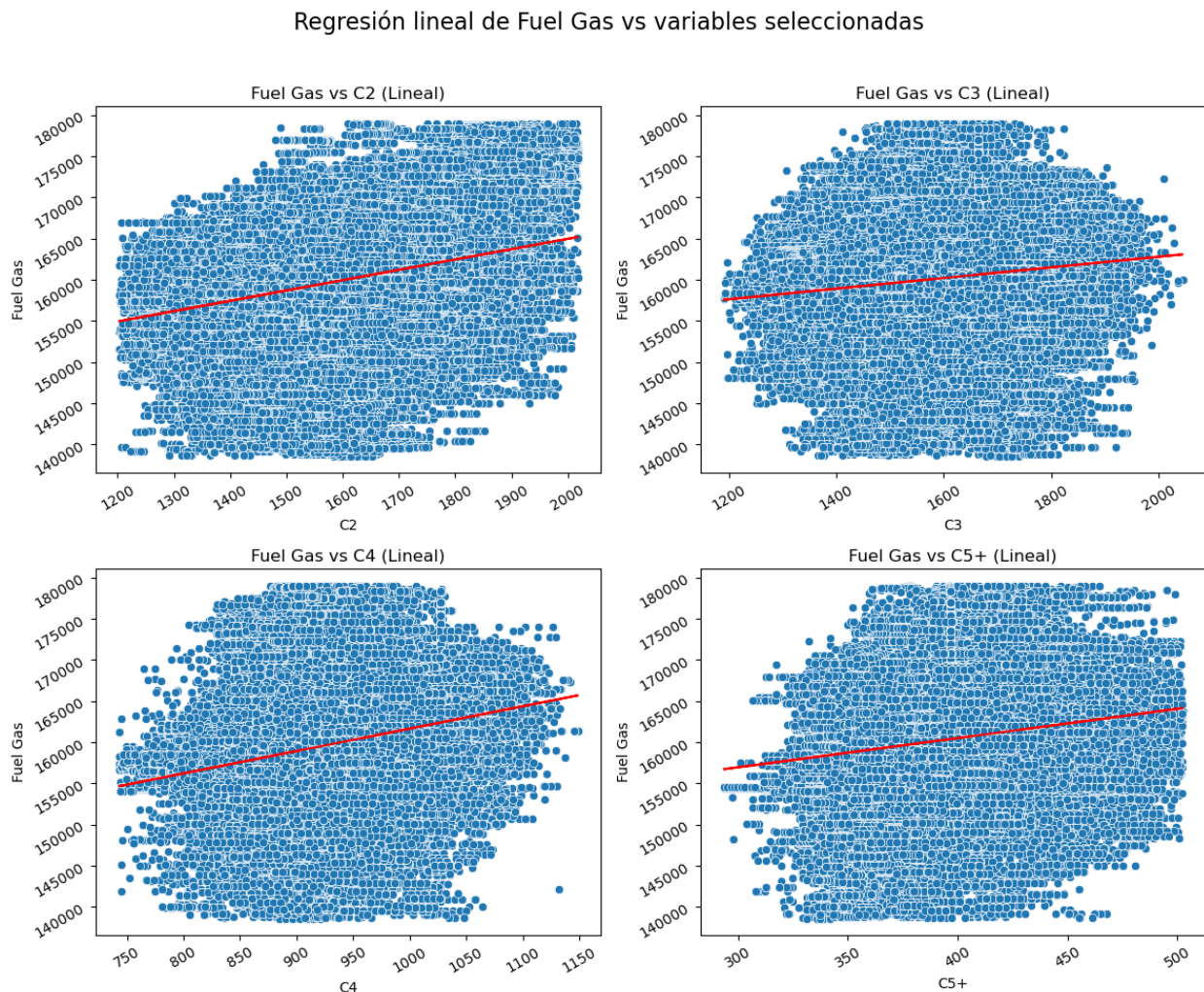


Figura 13: Regresiones lineales gas combustible en Bahía Blanca en función del caudal producido de cada producto, en m³/d y Tn/d respectivamente

En la Figura 13, se grafican las regresiones lineales de cada componente vs el Gas Combustible. Se realizaron también regresión polinómica de segundo grado, logarítmica y log-log. Los resultados de todas las regresiones se exponen en la Tabla 8.

Producto	Modelo	Intercepto	Coef x ¹	Coef x ²	R ²
Etano	Lineal (grado 1)	139.874	12,54	N/A	0,06
	Polinomial (grado 2)	224.678	-92,43	0,03	0,08
	Log(y) vs x	11,86	0,00	N/A	0,06
	Log-Log	11,09	0,12	N/A	0,06
Propano	Lineal (grado 1)	149.875	6,46	N/A	0,02
	Polinomial (grado 2)	119.429	44,47	-0,01	0,02
	Log(y) vs x	11,91	0,00	N/A	0,02
	Log-Log	11,49	0,07	N/A	0,02
Butano	Lineal (grado 1)	134.462	27,18	N/A	0,06
	Polinomial (grado 2)	156.423	-19,64	0,02	0,06
	Log(y) vs x	11,82	0,00	N/A	0,07
	Log-Log	10,87	0,16	N/A	0,07
Gasolina	Lineal (grado 1)	146.286	35,52	N/A	0,03
	Polinomial (grado 2)	28.920	619,05	-0,72	0,06
	Log(y) vs x	11,89	0,00	N/A	0,03
	Log-Log	11,42	0,094	N/A	0,03

Tabla 8: Resultados de las regresiones lineales de grado 1, grado 2, con transformación log y log-log del consumo de gas combustible en Bahía Blanca en función del caudal de etano

Interpretación

Etano

La regresión del etano no modela el comportamiento de la variables gas combustible frente a variaciones en el caudal de etano, ya que ninguna regresión presenta un R² aceptable. A simple vista hay una dispersión muy grande en los resultados, los cuales no parecen responder a ningún patrón fácilmente reconocible.

Propano

Las regresiones para el propano tienen aún menor poder explicativo que las del etano, con valores de R² muy inferiores. Una característica de este gráfico que lo diferencia del anterior es que los valores más bajos y más altos de consumo de gas combustible se encuentran más cerca de la media que de los extremos de producción. Por ejemplo, hay más casos de consumo de gas combustible superior a 170.000 kg/d cerca de las 1600 ton/d de C3, que cerca de 2000 ton/d.

Butano

En el caso del butano, el coeficiente de correlación R² sigue siendo bajo. Análogamente al caso del propano, los valores más bajos y más altos de consumo de gas combustible se encuentran más cerca de la media que de los extremos de producción, aunque en este caso si hay valores de producción de butano bajos que se corresponden con un consumo de gas combustible también bajo.

Gasolina

La gasolina también muestra una correlación baja, con valores de R^2 sin poder explicativo.

ix. Gas combustible vs Producción Total

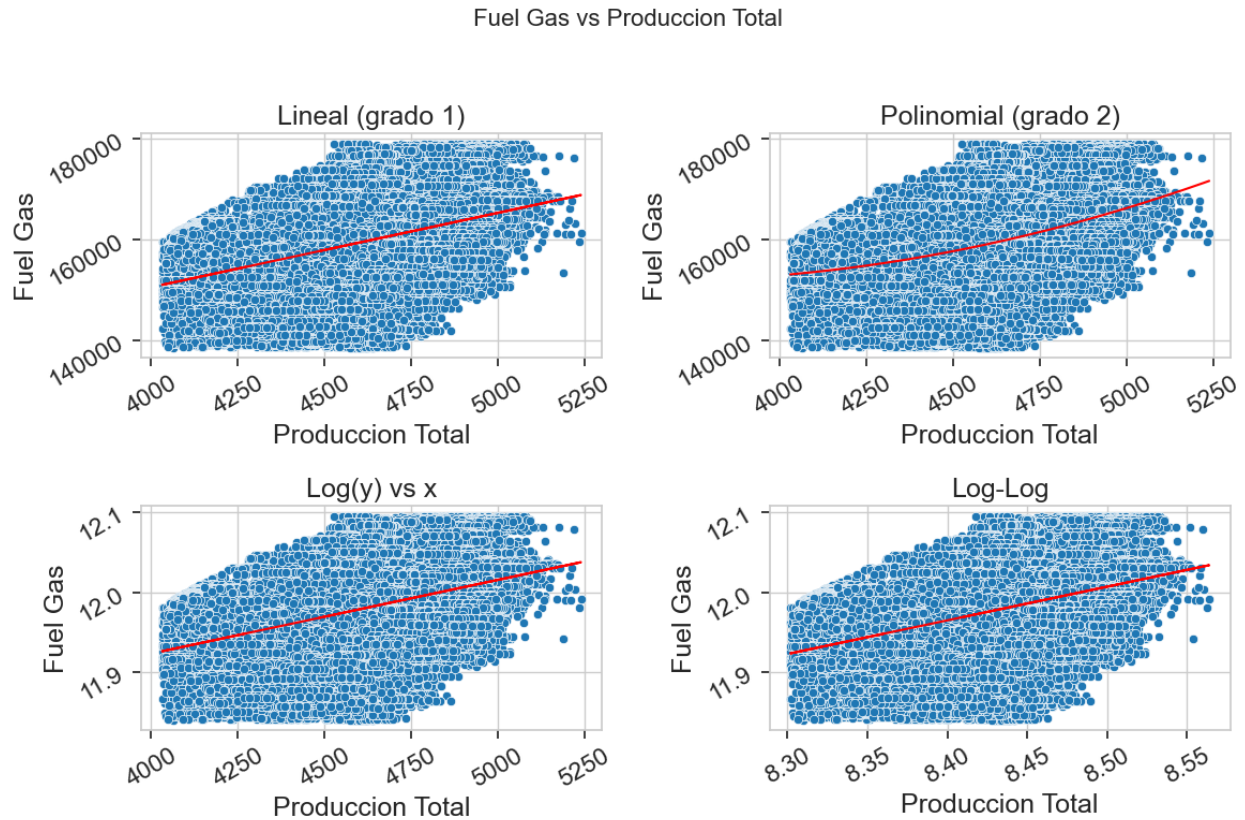


Figura 14: Regresiones lineales de grado 1, grado 2, con transformación log y log-log del consumo de gas combustible en Bahía Blanca en función del caudal total producido, en kg/d y Tn/d respectivamente

Modelo	Intercepto	Coef x^1	Coef x^2	R^2
Lineal (grado 1)	91.699	14,73	N/A	0,16
Polinomial (grado 2)	249.761	-54,20	0,00	0,16
Log(y) vs x	11,55	0,00	N/A	0,16
Log-Log	8,42	0,42	N/A	0,16

Tabla 9: Resultados de las regresiones lineales de grado 1, grado 2, con transformación log y log-log del consumo de gas combustible en Bahía Blanca en función del caudal total producido, en kg/d y Tn/d respectivamente

Interpretación

Por último, en la Figura 14 se ve un patrón más claro que en las regresiones del consumo de gas combustible de cada producto por separado. Al analizar los parámetros en la Tabla 9, se observa un R^2 de 0,16, el cual tiene un poder explicativo bajo pero indica que hay una correlación entre la producción y el consumo de gas combustible.

Discusión

Todos los modelos (lineales, polinomiales, logarítmicos y log-log) arrojan R^2 muy bajos ($< 17\%$), lo que confirma que ninguna de estas variables explica de manera significativa las variaciones de **gas combustible** en el conjunto de datos.

Es decir, individualmente, no parece haber relación entre el consumo total de Gas Combustible y ninguno de estos productos. Los datos no muestran patrones claros ni funciones que ajusten estas relaciones. Sin embargo, a nivel global sí parece haber una relación, aunque probablemente hay muchas más variables involucradas y por eso el poder explicativo de la producción total es bajo también.

b) Regresión Lineal Múltiple

Aquí se desarrollará un modelo lineal, con coeficientes para cada producto: C2, C3, C4 y C5+. Si el modelo logra predecir correctamente el comportamiento del consumo de Gas Combustible, se implementará la función dentro de la optimización lineal.

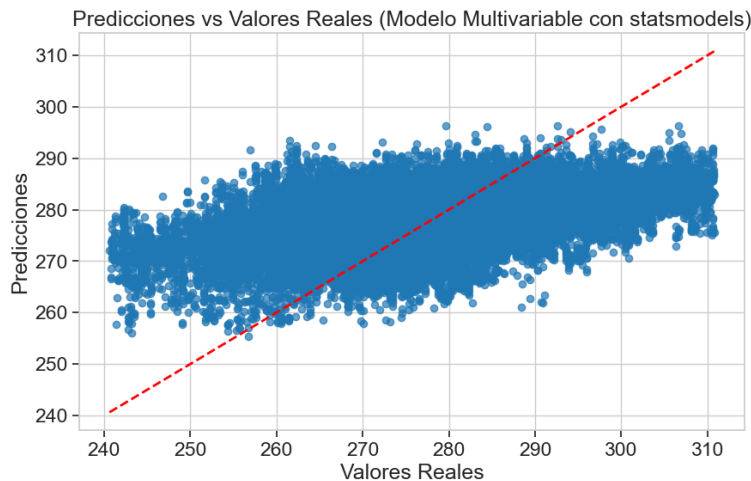


Figura 15: Predicciones del modelo de regresión lineal múltiple vs valores reales.

Parámetro	Valor
RMSE	10,8
R^2	0,21
No. Observaciones	87629

Tabla 10: Parámetros de error en la regresión lineal múltiple

Variable	Coef.	Std. Err.	t	P> t	[0.025	0.975]
const	144	0,923	155	0	141,84	145,45
C2	0,035	0,0003	124	0	0,034	0,035
C3	0,0026	0,0003	7,5	0	0,002	0,003
C4	0,068	0,001	81,8	0	0,067	0,07
C5+	0,021	0,0012	16,7	0	0,018	0,023

Tabla 10: Coeficientes de la regresión lineal múltiple

Intepretación

RMSE (10,8): En promedio, las predicciones del modelo difieren del valor real del consumo de gas por unos 10,8 m3/d, lo cual da una idea de la magnitud del error.

R² (0,21): El modelo explica aproximadamente el 21 % de la variabilidad observada en el consumo de gas. Esto sugiere que, aunque las variables C₂–C₅⁺ tienen un efecto significativo (p-valores << 0,001), gran parte de la variación del consumo queda sin explicar.

Coeficientes:

- **const (144):** Nivel de consumo base cuando todos los caudales son cero (más bien un intercepto matemático).
- **C2 (0,035):** Por cada unidad adicional de C₂, el consumo de gas aumenta en 0,035 unidades, manteniendo constantes los demás.
- **C3 (0,0026):** Su contribución es pequeña pero estadísticamente significativa.
- **C4 (0,068):** El coeficiente más alto; C₄ es el producto cuyo caudal tiene mayor impacto unitario en el consumo de gas.
- **C5+ (0,021):** Incrementa el consumo en ~0,02 unidades por cada unidad de C₅⁺.

En conjunto, aunque todas las variables son estadísticamente significativas y tienen signo positivo (mayor caudal implica mayor consumo de gas), el bajo R² indica que conviene explorar variables adicionales o transformaciones más complejas si se busca mejorar el poder predictivo.

c) Modelo de Boosting con XGBoost

Para complementar el análisis, se implementará un modelo de boosting mediante **XGBoost**, con el objetivo de capturar posibles interacciones y no linealidades entre el Consumo de Gas Combustible y todas las variables. A diferencia de los modelos anteriores, en este estudio se incluirán también los porcentajes de cada producto en la mezcla. Si bien se dispone de más variables (como el consumo de electricidad), **no se incluirán otras ya que el objetivo es predecir el consumo energético en base a la producción.**

Este algoritmo construye árboles de decisión de forma secuencial corrigiendo en cada paso los errores del modelo anterior. En este caso, se implementará una **búsqueda de hiperparámetros** básica, ajustando el número de árboles, la profundidad máxima y la tasa de aprendizaje; junto con una **validación cruzada de 5 particiones** (5 folds), entrenando cinco veces con diferentes divisiones de los datos y promediando el RMSE resultante. Esto garantiza

que la evaluación del modelo sea estable y no dependa de un único corte entre entrenamiento y prueba.

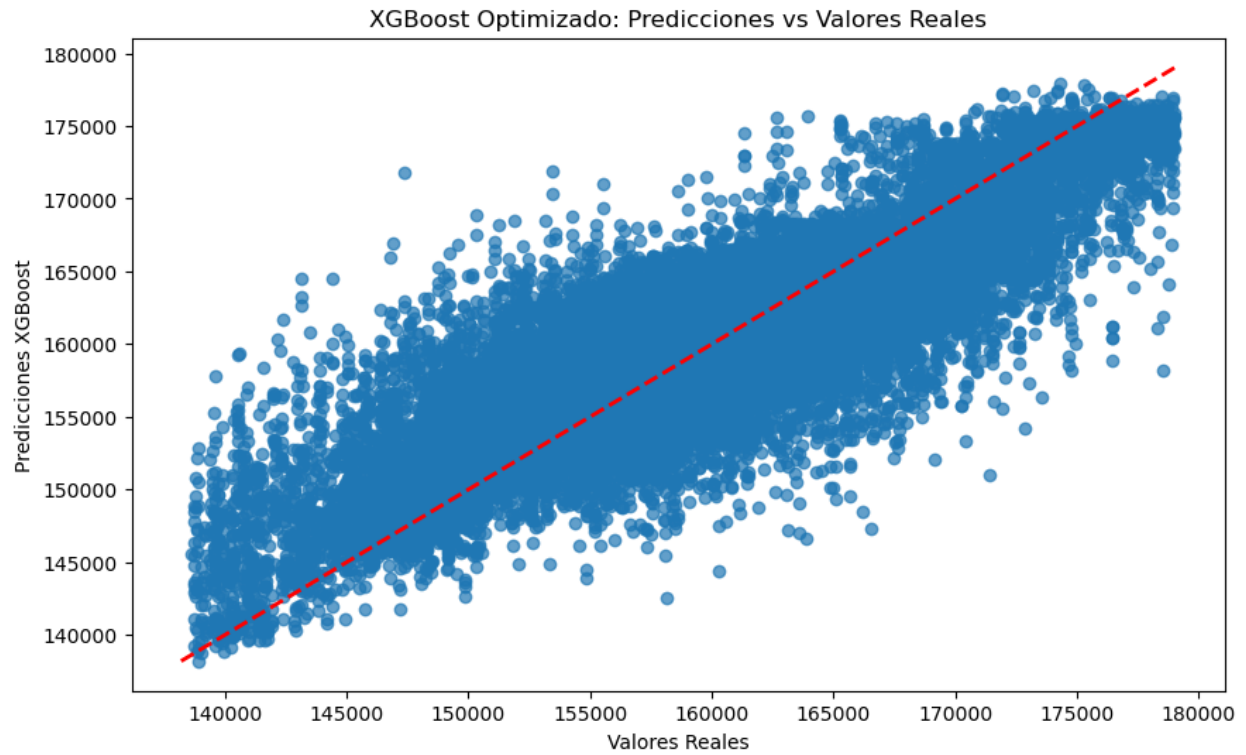


Figura 16: Predicciones algoritmo de boosting XGBoost.

Parámetro	Valor
RMSE	3672
R ²	72%
MAE	2701

Tabla 11: Parámetros de error del algoritmo XGBoost.

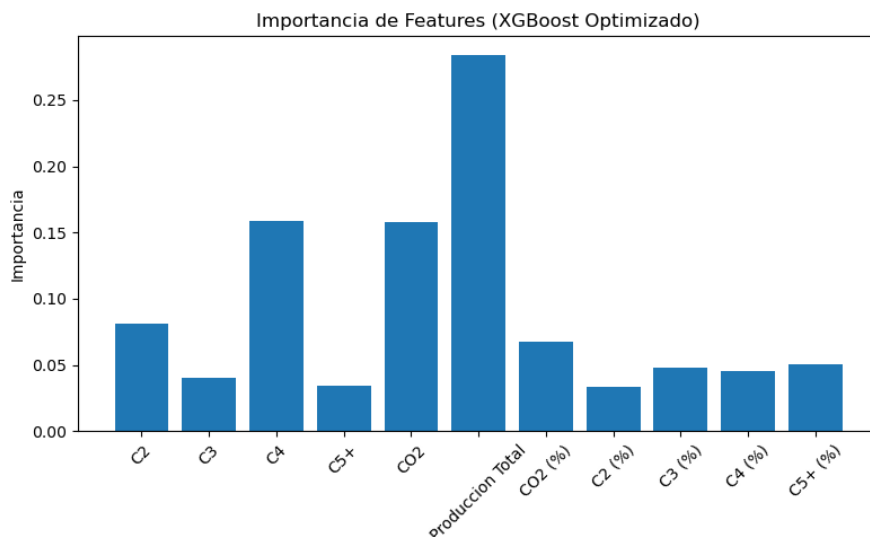


Figura 17: Importancia de features según XGBoost.

Los resultados demuestran que el algoritmo de XGBoosting con búsqueda de hiperparámetros y validación cruzada alcanza un R^2 mucho mayor, de 0,72. Esta mejora respecto de la regresión lineal múltiple implica que las variables analizadas sí se relacionan con el Gas Combustible, describiendo un 72% de las variaciones en el valor del mismo. Este valor supera el umbral que se estableció de 0,7, y se lo puede considerar razonablemente bueno.

Al observar la importancia de las variables en la Figura 17, se observa que la variable más relevante es la producción total, seguida de C4 y CO2. Las demás variables tienen importancia relativamente baja.

d) Discusión de Resultados del modelo predictivo

Se evaluaron diversos modelos para predecir el consumo de gas combustible en las torres de destilación a partir de los caudales de C2, C3, C4 y C5+. A pesar de probar distintas formulaciones, desde regresiones simples (lineales, polinómicas de segundo grado, logarítmicas y log-log) hasta una **regresión múltiple** que integraba simultáneamente todos los caudales, todos los modelos lineales arrojaron valores de R^2 bajos o moderados. Esto demostró que **una simple combinación lineal de los caudales no es suficiente para explicar la variabilidad del consumo energético**.

Sin embargo, al incluir variables adicionales de composición (porcentajes) y emplear un modelo no lineal como **XGBoost**, se logró un **R^2 de 0,72**. Este resultado es significativamente superior y demuestra que **es posible alcanzar una estimación mucho más precisa** cuando el modelo es capaz de capturar las complejas interacciones y no linealidades inherentes al proceso.

Es importante aclarar que el modelo de XGBoost, al estar basado en árboles de decisión, no produce coeficientes lineales explícitos y no puede usarse dentro de un modelo de optimización, por lo cual no se halló un modelo sencillo con poder de predicción que pueda incorporarse en un modelo de optimización (ya sea lineal o no lineal). Esto implica

que se deberá tomar una decisión sobre la inclusión del consumo de Gas Combustible en el modelo, lo cual se desarrollará en la próxima sección.

A continuación, se describen algunas explicaciones posibles para los resultados encontrados:

1. Complejidad y No Linealidad del Proceso Físico

- Las separaciones fraccionadas responden a balances de masa y energía intrínsecamente no lineales. La relación entre los caudales y el consumo de gas combustible no es directa, sino que involucra interacciones complejas entre múltiples flujos, dinámicas de transferencia de calor y pérdidas de carga. Las regresiones simples o múltiples no pueden capturar estas relaciones.
- Fenómenos como la formación de espumas o el arrastre de líquido introducen discontinuidades y efectos no lineales que solo algoritmos avanzados como los basados en árboles de decisión (XGBoost) pueden modelar eficazmente.

2. Variables Clave Ausentes y sus Interacciones

- El consumo energético de una torre de destilación depende de múltiples condiciones operativas además de los caudales: temperatura y presión de alimentación, razón de reflujo, eficiencia térmica del rehervidor y composición detallada de la alimentación.
- Un modelo lineal asume que el efecto de cada variable es independiente y aditivo. Sin embargo, en la realidad, estas variables interactúan. XGBoost es capaz de identificar y modelar interacciones más complejas.

3. Limitaciones de los Modelos Estadísticos Tradicionales

- Las regresiones lineales y polinómicas asumen una estructura rígida en la relación entre variables. El polinomio de segundo grado, aunque introduce una curvatura, es insuficiente para describir un proceso de destilación.
- Cuando las variables explicativas están correlacionadas (un mayor caudal de C2 suele acompañarse de cambios en C3), los coeficientes de una regresión múltiple pueden volverse inestables y difíciles de interpretar. Modelos como XGBoost son más robustos frente a la multicolinealidad.

4. Ruido y Variabilidad en los Datos

- Los modelos lineales son particularmente sensibles a los datos atípicos, comunes en las mediciones industriales. El bajo rendimiento inicial (R2 bajo) se debe en parte a esta dispersión.
- El éxito del modelo XGBoost sugiere que gran parte de lo que un modelo lineal considera "ruido" es en realidad "información no lineal". El algoritmo fue capaz de identificar patrones dentro de los datos que los modelos más simples no pudieron capturar.

5. Dinámica Temporal y Efectos de Memoria

- El consumo de gas puede depender del historial operativo (inercia). Estos fenómenos no son modelados correctamente por los modelos más simples ni por XGBoost, y podrían explicar parte de la variación que XGBoost no logra predecir.

x. Simplificaciones

Dado que no se encontró un modelo lineal que estimara con precisión el consumo de combustible en función de la producción, se estudiará prescindir de este término en el modelo de optimización. Para evaluar si podemos simplificar el modelo obviando las pequeñas variaciones en el consumo de gas combustible, se analizó la distribución del **gas combustible consumido por tonelada** mediante un gráfico tipo violín (Figura 18).

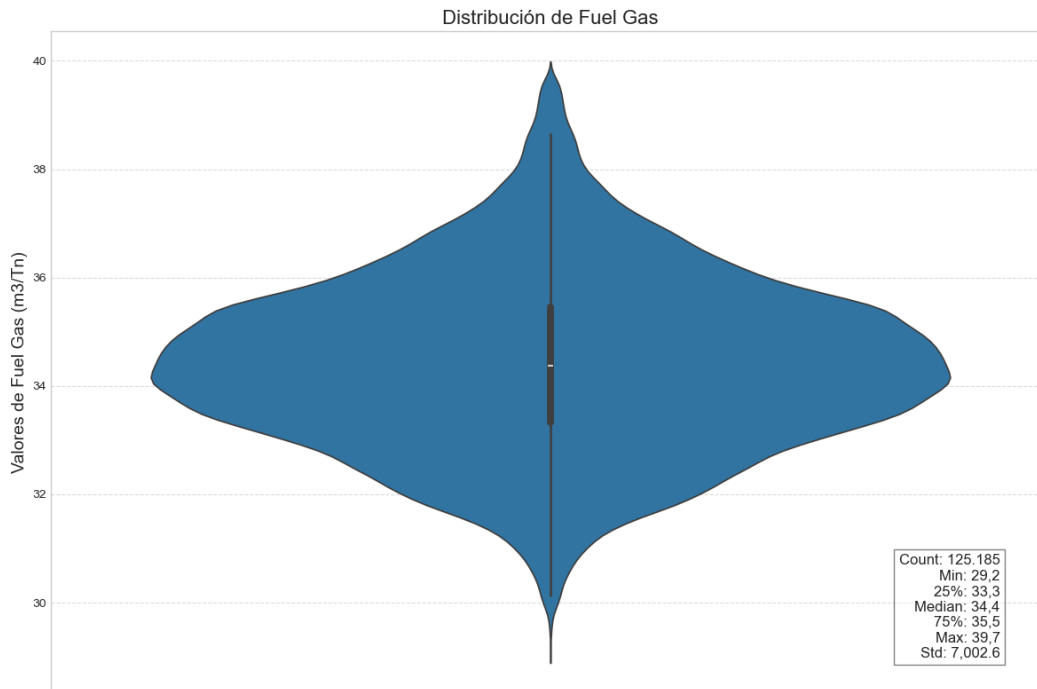


Figura 18: Gráfico de violín de la distribución del consumo de gas combustible (m3/Tn) en el dataset

Como se aprecia en la Figura 18 y en las estadísticas descriptivas de la variable (con outliers previamente eliminados), el consumo de gas combustible por tonelada sigue una distribución aproximadamente normal, centrada en torno a su media y con una dispersión moderada. La diferencia entre el valor máximo y el mínimo observados es de **10 m3/Tn**.

Si consideramos un precio de gas combustible de **3 USD/MMBTU**, que con un factor de conversión de 27 m3/MMBTU queda en **0,1 USD/m3**. De esta manera, dicha variación representa alrededor de **1 USD/Tn** de costo energético. Frente a precios de venta de los productos finales que oscilan entre **300 y 600 USD/Tn**, este rango es significativamente menor y desestimable. Por tanto, aunque el costo de gas combustible no es completamente despreciable, su incidencia en la **ganancia total** resulta generalmente baja respecto de la variación de ingresos por oscilaciones en los precios internacionales.

En consecuencia, y con el fin de concentrar el análisis en los factores que más afectan la rentabilidad global, consideraremos que las fluctuaciones en el consumo de gas combustible

pueden tratarse como un término relativamente constante dentro de nuestro modelo de optimización, y por lo tanto la desestimaremos en el modelo.

2. Resultados del modelo de optimización

En esta sección se exponen los resultados de las simulaciones del modelo de optimización, incluyendo el caso base, y las siguientes sensibilidades:

- Sensibilidad de precios (aumento del 20%)
- Expansión ilimitada de la capacidad de producción en Bahía Blanca
 - Manteniendo la limitación de capacidad máxima de cada proveedor
 - Eliminando las limitaciones de los proveedores
- Aumento y disminución de riqueza del 15%
- Eliminación de las limitaciones de los proveedores
- Escenarios en la nominación de etano por parte de DOW: 0, 1500 y 1700 Tn/d

Todos los resultados se expresan en toneladas por día y en valor óptimo de facturación (MM USD).

a) Caso base

Producto	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F _{C2}	1600	Q ₁	341	168	z ₁	74	50	30	13
F _{C2_vent}	1544	Q ₂	1968	1153	z ₂	646	288	161	58
F _{C3}	1650	Q ₃	4837	0	z ₃	0	0	0	0
F _{C4}	963	Q ₄	3547	3547	z ₄	1791	967	537	251
F _{C5+}	487	Q ₅	1376	1376	z ₅	633	344	234	165
		Q ₆	1539	0	z ₆	0	0	0	0

Tabla 12: Distribución de corrientes y aporte de cada producto de cada una y total para el caso base

Valor Óptimo (MMUSD): 1,652

Comentarios

En el caso base, se consumen completamente las corrientes 4 y 5, mientras que la 3 y la 6 no se consumen y la 1 y 2 sólo parcialmente.

Por otra parte, la suma de las corrientes de C4 y C5+ da 1450 (exactamente el límite), mientras que la suma de C3, C4 y C5+ da 3073, sobre un máximo de 3100 Tn/d.

b) Aumento de precios del 20%

Producto	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F _{C2}	1600	Q ₁	341	168	z ₁	74	50	30	13
F _{C2_vent}	1544	Q ₂	1968	1153	z ₂	646	288	161	58
F _{C3}	1650	Q ₃	4837	0	z ₃	0	0	0	0
F _{C4}	963	Q ₄	3547	3547	z ₄	1791	967	537	251

F_{C5+}	487	Q₅	1376	1376	z₅	633	344	234	165
		Q₆	1539	0	z₆	0	0	0	0

Tabla 13: : Distribución de corrientes y aporte de cada producto de cada una y total para el caso de aumento de precios del 20%

Valor Óptimo (MMUSD): 2,011

Comentarios

El caso de aumento de precios en 20% no muestra diferencias en cuanto a las decisiones con respecto al caso, pero el impacto en la facturación es casi directamente proporcional.

c) Expansión con limitaciones de proveedores

Producto	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F_{C2}	4685	Q₁	341	0	z₁	0	0	0	0
F_{C2_vent}	0	Q₂	1968	0	z₂	0	0	0	0
F_{C3}	2355	Q₃	4837	4837	z₃	2757	1209	677	193
F_{C4}	1384	Q₄	3547	1193	z₄	602	325	181	84
F_{C5+}	520	Q₅	1376	1376	z₅	633	344	234	165
		Q₆	1539	1539	z₆	693	477	292	77

Tabla 14: Distribución de corrientes y aporte de cada producto de cada una y total para el caso de capacidad de producción ilimitada pero manteniendo las restricciones de los proveedores

Valor óptimo (MMUSD): 2,118

Comentarios

En este caso, al no haber límite en la capacidad de procesamiento, se procesa todo el etano. Además, las corrientes 3, 4 y 6 se exigen al máximo, mientras que la 5 sólo parcialmente, y la 1 y la 2 no se eligen. La facturación resulta un 30% superior al caso base.

Por otro lado, la suma de las corrientes de C4 y C5+ da 1904, sobre el límite original de 1450, mientras que la suma de C3, C4 y C5+ da 4259, sobre un máximo original de 3100 Tn/d.

d) Expansión sin limitaciones de proveedores

	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F_{C2}	6772	Q₁	-	0	z₁	0	0	0	0
F_{C2_vent}	0	Q₂	-	0	z₂	0	0	0	0
F_{C3}	4665	Q₃	-	0	z₃	0	0	0	0
F_{C4}	2859	Q₄	-	0	z₄	0	0	0	0
F_{C5+}	752	Q₅	-	0	z₅	0	0	0	0
		Q₆	-	15049	z₆	6772	4665	2859	752

Tabla 15: Distribución de corrientes y aporte de cada producto de cada una y total para el caso de capacidad de producción y abastecimiento ilimitadas

Valor óptimo (MMUSD): 4,879

Comentarios

En este caso, donde todos los proveedores tienen una capacidad ilimitada y la única restricción es la (4.10), correspondiente al límite de capacidad en Neuquén, se eligió comprarle exclusivamente al proveedor 6. El impacto en la facturación de levantar estas restricciones es significativo, un 220% superior al caso base.

e) Aumento de riqueza del 15%

	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F_{C2}	1600	Q₁	341	0	z₁	0	0	0	0
F_{C2_vent}	516	Q₂	1968	0	z₂	0	0	0	0
F_{C3}	1552	Q₃	4837	4837	z₃	2032	1451	919	435
F_{C4}	987	Q₄	3547	0	z₄	0	0	0	0
F_{C5+}	463	Q₅	1376	0	z₅	0	0	0	0
		Q₆	1539	281	z₆	84	101	68	28

Tabla 16: Distribución de corrientes y aporte de cada producto de cada una y total para el caso de aumento de riqueza, con 5 puntos porcentuales más de C3, C4 y C5+, y una disminución de C2 en 15 puntos porcentuales para compensar

Valor Óptimo (MMUSD): 1,652

Comentarios

En este escenario, se modificó la composición de las corrientes, agregándole 5 puntos porcentuales a cada componente más pesado que el etano, y restando 15 puntos porcentuales al etano para compensar. Aquí se compra todo el caudal de la corriente 3, y un poco de la corriente 6. La facturación no varió significativamente respecto del caso base.

Mirando las restricciones, la suma de las corrientes de C4 y C5+ da 1450 (exactamente el límite), mientras que la suma de C3, C4 y C5+ da 3002, sobre un máximo de 3100 Tn/d.

f) Disminución de riqueza del 15%

	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F_{C2}	1600	Q₁	341	341	z₁	201	85	44	10
F_{C2_vent}	3577	Q₂	1968	1253	z₂	890	251	113	0
F_{C3}	1801	Q₃	4837	0	z₃	0	0	0	0
F_{C4}	898	Q₄	3547	3547	z₄	2323	790	360	73
F_{C5+}	180	Q₅	1376	1376	z₅	839	275	165	96
		Q₆	1539	1539	z₆	923	400	215	0

Tabla 17: Distribución de corrientes y aporte de cada producto de cada una y total para el caso de disminución de riqueza, con 5 puntos porcentuales menos de C3, C4 y C5+, y un aumento de C2 en 15 puntos porcentuales para compensar

Valor óptimo (MMUSD): 1,647

Comentarios

Este caso es opuesto al anterior: se quitaron 5 puntos porcentuales a cada componente más pesado que el etano, y se agregaron 15 puntos porcentuales al etano para compensar. De forma opuesta al caso anterior, se eligieron todas las corrientes menos la 3. La facturación tampoco varió significativamente respecto del caso base.

g) Sin límites de capacidad de proveedores

Producto	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F _{C2}	1600	Q ₁	-	0	z ₁	0	0	0	0
F _{C2_vent}	1452	Q ₂	-	0	z ₂	0	0	0	0
F _{C3}	1650	Q ₃	-	4929	z ₃	2489	1344	747	349
F _{C4}	955	Q ₄	-	1223	z ₄	563	306	208	147
F _{C5+}	495	Q ₅	-	0	z ₅	0	0	0	0
		Q ₆	-	0	z ₆	0	0	0	0

Tabla 18: Distribución de corrientes y aporte de cada producto de cada una y total para el caso de capacidad de abastecimiento ilimitada de parte de proveedores

Valor óptimo (MMUSD): 1,653

Comentarios

En el caso sin límites de capacidad de proveedores, se exigió al la corriente 3 un poco por encima de la capacidad máxima (4929 vs 4837 ton/día), pero el caudal total no fue muy distinto a los casos con restricción en capacidad de proveedores, y la facturación tampoco.

h) Nominación de Etano DOW

	Caudal	Corriente	Q^{max}	Caudal	Corriente	C2	C3	C4	C5+
F _{C2}	1500/1600/1700	Q ₁	341	168	z ₁	74	50	30	13
F _{C2_vent}	1644/1544/1444	Q ₂	1968	1153	z ₂	646	288	161	58
F _{C3}	1650	Q ₃	4837	0	z ₃	0	0	0	0
F _{C4}	963	Q ₄	3547	3547	z ₄	1791	967	537	251
F _{C5+}	487	Q ₅	1376	1376	z ₅	633	344	234	165
		Q ₆	1539	0	z ₆	0	0	0	0

Tabla 19: Distribución de corrientes y aporte de cada producto de cada una y total para tres casos de nominación de etano por parte de DOW: 1500 Tn/d, 1600 Tn/d y 1700 Tn/d

Comentarios

Se indican en el gráfico los diferentes valores que admiten C2 y C2_vent en los casos de caudal de etano exigido por DOW de 1500, 1600 y 1700 Tn/d respectivamente.

Las decisiones no varían respecto del caso base. Sólo se modifica la facturación. Los valores de facturación calculados son:

Caudal de etano de 1500 Tn/d: Valor óptimo de 1,629 MMUSD

Caudal de etano de 1600 Tn/d: Caso base (1,652 MMUSD)

Caudal de etano de 1700 Tn/d: Valor óptimo de 1,653 MMUSD

Caudal de etano de 0 Tn/d (Paro de planta): Valor óptimo de 1,295 MMUSD

i) Discusión de resultados

i. Resultado caso base

El análisis realizado revela patrones interesantes en la selección de proveedores y en las decisiones de compra. Por ejemplo, las corrientes 4 y 5 son las preferidas, pues se usan al máximo de su producción. Esto sugiere que su composición, rica en C3, C4 y C5+, encaja bien con las restricciones del modelo.

En contraste, las corrientes 1 y 2, a pesar de ser consideradas, no se utilizan al máximo de su capacidad. Esta decisión probablemente se debe a las restricciones de los límites para componentes más pesados. Por otro lado, las corrientes 3 y 6, que presentan una menor riqueza en componentes pesados, tampoco fueron seleccionadas, probablemente para maximizar el caudal de productos más caros.

Un caso particular es el de la corriente 1, cuya alta riqueza podría haber justificado un mayor uso, pero las restricciones de componentes más pesados parecen haber limitado su caudal.

En resumen, la selección de corrientes responde a un equilibrio entre objetivos de producción y límites en la proporción de componentes, asegurando el cumplimiento de las metas del proceso.

ii. Sensibilidad a los Precios Internacionales de Productos (P_j)

El mercado de hidrocarburos experimenta constantes oscilaciones en los precios de C2, C3, C4 y C5+, por lo que resulta crucial comprender cómo un aumento del 20% en cada uno de estos productos afecta la rentabilidad de la planta y las decisiones de compra. Para ello se resolvió el modelo en cinco escenarios: cuatro en los que cada precio subió de manera independiente (excepto para el C2, ya que el modelo tiene un requisito específico de producción), y uno en el que todos incrementaron simultáneamente.

Discusión

Dado que los precios aumentan proporcionalmente en todos los productos, la composición óptima no se modifica. El dato relevante aquí es que la facturación aumenta en 20%. Para una empresa que no tiene mucha flexibilidad para reducir costos y que no tiene influencia en el precio de su producto, esta variabilidad demuestra la exposición de Mega a las fluctuaciones de precios internacionales.

Los resultados modificando independientemente los precios de un solo producto son los mismos en todos los casos, hasta que el incremento llega a un valor específico. Para los precios establecidos, el incremento para C3 es 45%, para C4 35% y para C5+ 220%. Por debajo de estos incrementos de precios, el modelo es insensible a los aumentos. Las facturaciones respectivas para estos casos fueron 1,912 MMUSD, 1,771 MMUSD y 2,229 MMUSD.

iii. Impacto de la Disponibilidad Ilimitada de Materia Prima

Para explorar el potencial de negociación con los proveedores, eliminamos las restricciones $Q_i \leq Q_i^{\max}$ sin alterar las demás limitaciones de mezcla y proceso. De este modo, el modelo identifica qué volúmenes demanda idealmente de cada proveedor. Algunos proveedores tienen limitaciones dadas por acuerdos comerciales, y es interesante estudiar cuál es el volumen óptimo que se le debería exigir en condiciones ideales a cada proveedor, para considerar en futuras negociaciones comerciales.

Discusión

La facturación se mantiene constante respecto del caso original. Se nota que se prefiere la corriente 3 porque al ser pobre, no se puede acceder a más C3, considerando la restricción fuerte de C4 y C5+.

iv. Impacto de las Restricciones de Producción en Bahía Blanca (Simulación de Expansión)

El proyecto de expansión de Mega contempla la eliminación progresiva de las restricciones de la planta en Bahía Blanca, por lo que diseñamos dos casos para simular esta evolución. En el **Caso I**, desactivamos tanto los límites de producción individuales de C2, C3, C4 y C5+ como las restricciones combinadas, manteniendo únicamente la limitación de origen desde Neuquén. En el **Caso II**, añadimos la remoción de los techos de suministro de los proveedores.

Discusión

Caso 1

En este caso se aprovechó el levantamiento de la restricción de los proveedores y se llegó a la restricción de la capacidad de la planta fraccionadora.

La corriente 2 volvió a ser obviada, probablemente por ser pobre.

Caso 2

Cuando no hay restricciones en la planta fraccionadora de Bahía Blanca ni en los caudales de los clientes, el modelo elige comprarle al proveedor de la corriente 6. Esta no es la corriente más rica, pero la medida a optimizar es la combinación de las facturaciones para todos los componentes. Además, dado que la restricción de procesamiento de Neuquén sigue vigente, también influye el hecho de tener un coeficiente de restricción bajo.

v. Sensibilidad a la Composición de las Corrientes de Alimentación (y_{ij})

Como se mencionó en la sección **Problema**, la evolución geológica de los yacimientos ha enriquecido progresivamente la composición de las corrientes, por lo que resulta realista evaluar variaciones de ± 5 puntos porcentuales en los componentes C3, C4 y C5+, ajustando la fracción de C2 de manera compensatoria.

Discusión

Cuando las corrientes son más ricas, tienen mucho más impacto las restricciones. Es interesante ver cómo no mejora la facturación, pero se alcanza el techo en una restricción y prácticamente también en la otra.

vi. Sensibilidad al Requerimiento Comercial de Etano (Req_{c2})

DOW establece requisitos variables de etano —que en ocasiones puede reducirse a cero debido a paradas por condiciones climáticas— por lo que analizamos los escenarios Req_{c2} = 0, 1500, 1600 (caso base) y 1700 Tn/d, manteniendo el resto de las restricciones.

Discusión

Esta variable no tiene impacto operativo, ya que hay libertad para el venteo, con lo cual mientras se mantenga este requerimiento fijo, su modificación no va a impactar en el modelo, al menos hasta alcanzar el caudal máximo del resultado, en cuyo caso deberá revisarse el planteo del modelo.

6. Conclusiones

Este trabajo se propuso desarrollar y aplicar un modelo de optimización lineal para determinar la estrategia óptima de compra de gas natural de Compañía Mega, maximizando la rentabilidad operativa al equilibrar los ingresos por venta de productos con los costos energéticos, sujeto a un conjunto complejo de restricciones operativas, comerciales y de suministro. Además, se propuso modelar la variable de consumo energético a partir de los datos disponibles. De esta manera, esta investigación se distingue por buscar la combinación de análisis prescriptivo y predictivo en la resolución de un problema operativo comercial en la industria.

El análisis del caso base y los estudios de sensibilidad han permitido obtener las siguientes conclusiones:

Los parámetros operativos en una instalación industrial, como el consumo energético, involucran muchas variables, las cuales a su vez probablemente implican no linealidades o fenómenos estocásticos, que hacen difícil pensar un modelo de predicción sencillo basado en datos históricos. En particular, las torres de destilación se ven afectadas por variables como la composición y el caudal de la alimentación, las condiciones de presión y temperatura, y el reflujo aplicado, entre otras variables técnicas de los procesos industriales. No obstante, es posible predecir los consumos energéticos con relativa precisión a partir de datos históricos y aplicando modelos no lineales. Además, el análisis descriptivo de datos históricos permite acotar el impacto de variables clave y proponer modelos simplificados basados en criterios claros.

Por otro lado, la optimización está fuertemente condicionada por las restricciones del proceso, tanto operativas como comerciales. En particular, las restricciones de procesamiento en Bahía Blanca (producción combinada de C4+C5+ y de C3+C4+C5+) operan muy cerca de sus límites (1450/1.450 Tn/d y 3073/3.100 Tn/d), y son el factor que más afecta la rentabilidad. Levantar únicamente esta restricción aumenta la facturación de 1,6 a 2,1 MM USD/día. Por su parte, la disponibilidad de materia prima de proveedores constituye un segundo cuello de botella: solo al aliviar ambas restricciones la facturación alcanza 4,9 MM USD/día.

El modelo presentó poca sensibilidad a los precios, e incluso variaciones moderadas (< 35 %) no alteran la combinación óptima de productos, aunque sí impactan significativamente la facturación. Esto indica que la solución óptima es estable frente a fluctuaciones razonables, pero que la rentabilidad financiera de Mega está expuesta a la volatilidad del mercado internacional.

También es notable observar que un enriquecimiento (+5% en C3, C4, C5+ en fuentes clave) **no mejoró la rentabilidad** (se mantuvo en 1,652 MMUSD), ya que las restricciones de Bahía Blanca se volvieron aún más limitantes, forzando la selección de proveedores más pobres para compensar. Por otra parte, un **empobrecimiento** generalizado **redujo ligeramente la rentabilidad**.

Una observación, intuitiva pero confirmada mediante el análisis, es que el etano (C2) no resulta crítico en el modelo: su producción actual está muy por debajo del límite operativo y no comparte restricciones con otros productos, por lo que no representa un cuello de botella relevante. Por supuesto, todas las instalaciones industriales tienen limitaciones, pero probablemente el cuello de botella de C2 sea tan grande relativamente que no es un parámetro considerado relevante.

1. Implicaciones estratégicas para Compañía Mega

Prioridad de Inversión: El análisis identifica claramente las restricciones de procesamiento/logística de Bahía Blanca como el principal factor limitante de la rentabilidad. Proyectos de expansión o debottlenecking en esta área tendrían un alto retorno potencial. La capacidad de Neuquén es relevante, pero secundaria respecto a Bahía Blanca bajo las condiciones actuales de suministro.

Gestión de Suministros: Si bien la disponibilidad actual no parece ser el principal limitante hoy, la estrategia óptima cambiaría si se alivian las restricciones de proceso. Asegurar flexibilidad o mayor volumen de proveedores que resulten óptimos en escenarios de expansión, podría ser valioso a futuro. Esta decisión no es trivial: corrientes de alimentación más ricas pueden alcanzar cuellos de botella de producción que limiten la facturación, y la solución óptima dependerá de cómo evolucionen las composiciones de los proveedores y los cuellos de botella a medida que la empresa invierte en expandir su capacidad de producción.

Gestión de Riesgo de Mercado: Dada la exposición de la rentabilidad a los precios, es importante explorar estrategias de gestión de riesgo de precios.

Negociación Comercial (Etano): Existe flexibilidad operativa para manejar variaciones en la nominación de etano e impactar en la facturación, lo cual es un punto a considerar en la relación con DOW. Sin embargo, abrir la posibilidad a vender el excedente puede tener un impacto significativo en la producción, ya que según el análisis del caso base, hoy se estaría venteando la mitad del caudal procesado.

En definitiva, esta investigación ha puesto de manifiesto tanto el potencial como las limitaciones de emplear modelos de regresión y programación lineal en la planificación industrial. Gracias al desarrollo y aplicación del modelo, no solo se logró definir la combinación óptima de abastecimiento de gas natural, sino que, a partir de los análisis de sensibilidad, se identificaron los principales cuellos de botella en la planta. Esto ha permitido trazar un plan de acción claro, priorizando las inversiones y decisiones estratégicas que Compañía Mega debe tomar para maximizar su facturación a largo plazo y fortalecer la resiliencia de sus operaciones frente a las fluctuaciones del mercado.

7. Trabajos Futuros

Los modelos matemáticos siempre son representaciones simplificadas de la realidad para simular el comportamiento de fenómenos complejos. En el caso desarrollado, los modelos son sencillos y fácilmente interpretables. Sin embargo, estas simplificaciones le quitan poder explicativo al modelo, y pueden pensarse alternativas para proponer nuevos modelos que representen más fielmente la realidad. Algunas opciones son:

Incorporación de Incertidumbre (Optimización Estocástica/Robusta): Esto implicaría modelar los parámetros inciertos como variables aleatorias con distribuciones de probabilidad (basadas en datos históricos o pronósticos) o dentro de rangos de incertidumbre. Permitiría encontrar una estrategia de compra que sea óptima en promedio o que tenga un buen desempeño bajo diferentes escenarios posibles.

Modelos dinámicos: Extender el modelo a un **horizonte multi-período** (semanal, mensual). Estos modelos pueden incluir retardos y variables lag para reflejar la dinámica de la planta. Aquí se incluiría explícitamente variaciones estacionales, y facilitaría la toma de decisiones comerciales de largo plazo, como la firma de contratos para abastecimiento. Así se podría realizar recomendaciones con perspectiva más estratégica, considerando la evolución de las variables clave a lo largo del tiempo, de acuerdo a los escenarios más probables.

Modelos más avanzados: se pueden hacer regresiones múltiples con más variables, como presión de operación, temperatura de alimentación, razón de reflujo, eficiencia del rehervidor, etc. También se pueden incorporar modelos no paramétricos de Machine Learning como Gradient Boosting, Redes neuronales, Análisis de componentes principales. Estos modelos pierden el poder explicativo y la flexibilidad de las regresiones para incorporar sus resultados a modelos de optimización, pero pueden lograr predicciones más precisas.

Integración con simulación de procesos: podría buscarse una integración con herramientas de simulación de procesos químicos especializadas como Aspen HYSYS para calcular los rendimientos y consumos energéticos, optimizando el resultado mediante métodos iterativos. De esta forma se lograría combinar principios físicos (balances de energía y masa) con métodos estadísticos para mejorar la interpretabilidad y el ajuste.

1. Incorporación del Modelo como Herramienta de Decisión Estratégica y Táctica

La herramienta de optimización desarrollada es un recurso útil para Compañía Mega. Su integración en la planificación estratégica y táctica permitirá respaldar decisiones con análisis basados en datos.

El modelo permite simular escenarios “what-if”, como evaluar nuevas corrientes de gas, ajustar contratos con proveedores o clientes, analizar inversiones en expansión o preparar respuestas a cambios en precios de mercado. Esto facilita la comparación de alternativas en todo tipo de decisiones: desde la evaluación de un proyecto hasta la planificación operativa diaria.

Es fundamental actualizar regularmente el modelo optimización lineal con datos recientes y revisar sus supuestos para mantener su relevancia ante cambios en las condiciones operativas de Compañía Mega, sus instalaciones, o el mismo mercado.

8. Bibliografía

- [1] U.S. Energy Information Administration, «World Shale Resource Assessments,» 24 Septiembre 2015. [En línea]. Available: <https://www.eia.gov/analysis/studies/worldshalegas/>.
- [2] El Litoral, «Vaca Muerta impulsa un año récord en la producción de hidrocarburos en Argentina,» 6 Febrero 2025. [En línea]. Available: https://www.ellitoral.com/economia/vaca-muerta-impulsa-ano-record-produccion-hidrocarburos-argentina-hidrocarburos-ahora-hoy_0_1CymUe2A98.html.
- [3] «Confirman trece muertos por las inundaciones en Bahía Blanca,» 8 Marzo 2025. [En línea]. Available: <https://www.infobae.com/america/agencias/2025/03/08/confirman-trece-muertos-por-las-inundaciones-en-bahia-blanca/>.
- [4] M. M. H. A. M. S. Hesham A. Hegazy, "Optimum Blending of Crude Oils Using Linear Programming," *Egyptian Journal of Chemistry*, vol. 66, no. Applied Chemistry for Greener Life and Sustainability, pp. 1341-1353, 2023.
- [5] I. E. O. E. Enekwe C.B., "Optimization of Blends of Selected Foreign and Nigerian Crudes," *International Journal of Science and Engineering Investigations*, vol. 1, no. 9, 2012.
- [6] R. & A. M. & S. M. Zin, «Energy Analysis and Remixing Effect of Thermal Coupling Petlyuk Column for Natural Gas Liquid (NGL) Fractionation Train,» *International Journal of Renewable Energy Development.*, vol. 10, pp. 459-469, 2021.
- [7] W. A. P. J. G. S. Saeid Mokhatab, *Handbook of Natural Gas Transmission and Processing*, Elsevier, 2012.
- [8] E. A. P. G. G. V. Douglas C. Montgomery, *Regression Analysis*, Fifth ed., Wiley, 2012.
- [9] H. P. Williams, *Model Building in Mathematical Programming*, Fifth ed., Wiley, 2013.
- [10] P. Jirutitjaroen, «An optimization model for natural gas supply portfolios of a power generation company,» *Applied Energy*, vol. 107, pp. 1-9, 2013.
- [11] S. C. Bhattacharyya, *Energy Economics*, 1st ed., Springer, 2011.
- [12] G. L. Frederick Hillier, *Introduction to Operations Research*, McGraw Hill, 2024.
- [13] N. M. A. N. B. H. Phan Thanh Dao, «Using Linear Regression Analysis to Predict Energy Consumption,» *Applied Energy*, vol. 130, pp. 305-313, 2024.