Pronóstico de Demanda e Inventario estratégico MRO con Series de Tiempo y Variables Exógenas en una Operación Minera



Luis Carlos Lozano Sanabria

Tutor: Vladimir Moreno G., MSc
Cotutor: David Arango L., MSc

Pontificia Universidad Javeriana
Facultad de Ciencias
Pregrado en Ciencia de Datos
Bogotá D.C.
2026

Resumen

La gestión de inventarios MRO en operaciones mineras intensivas en activos enfrenta dos retos simultáneos: la heterogeneidad del patrón de demanda entre artículos y la ausencia de modelos cuantitativos que traduzcan la incertidumbre en reglas operativas concretas. Este trabajo desarrolla un sistema de pronóstico de demanda e integración de política de inventario para seis insumos estratégicos de C.I. Milpa S.A., empresa colombiana de minería de carbón coquizable y producción de coque metalúrgico.

La metodología combina clasificación de demanda mediante el esquema Syntetos-Boylan, validación fuera de muestra por rolling-origin backtesting y familias de modelos diferenciadas: SARIMAX con variables exógenas climáticas y macroeconómicas, y modelos ETS para series de consumo continuo; métodos Croston-SBA, TSB y un modelo hurdle con bosque aleatorio para series intermitentes. Los parámetros de inventario se derivaron mediante simulación Monte Carlo sin suponer distribución normal de la demanda.

Los resultados muestran que la selección óptima del modelo depende del patrón individual de cada serie. ETS dominó en dos de los tres artículos de consumo continuo; SARIMAX fue superior en el único caso donde las variables exógenas presentaron correlación estadísticamente relevante con la demanda. En series intermitentes, el modelo hurdle superó a los métodos clásicos en dos de tres artículos. La simulación Monte Carlo produjo puntos de reorden y stocks de seguridad para un nivel de servicio del 95%.

El backtesting de la política propuesta sobre datos reales (agosto 2025 – marzo 2026) cuantificó una liberación potencial de $76,5 millones de COP en capital inmovilizado, equivalente al 34,3% del inventario medio observado para los seis SKUs evaluados, sin comprometer el nivel de servicio en las series de consumo continuo. El trabajo demuestra que un flujo analítico reproducible, construido sobre registros históricos de consumo y variables exógenas de acceso público, puede institucionalizar criterios cuantitativos de reabastecimiento en entornos donde las decisiones se toman actualmente por intuición.

Palabras clave: gestión de inventarios MRO, clasificación Syntetos-Boylan, demanda intermitente, SARIMAX, ETS, Hurdle-RF, rolling-origin backtesting, simulación Monte Carlo, punto de reorden, minería de carbón.

Abstract

MRO inventory management in asset-intensive mining operations faces two simultaneous challenges: heterogeneous demand patterns across items and the absence of quantitative models capable of translating uncertainty into actionable replenishment rules. This work develops a demand forecasting and inventory policy integration system for six strategic items at C.I. Milpa S.A., a Colombian

company engaged in coking coal mining and metallurgical coke production.
The methodology combines demand classification using the Syntetos-Boylan scheme, out-of-sample validation through rolling-origin backtesting, and separate model families: SARIMAX with climatic and macroeconomic exogenous variables and ETS models for continuous demand series; Croston-SBA, TSB, and a hurdle model with random forest for intermittent series. Inventory parameters were derived via Monte Carlo simulation without assuming a normal demand distribution.
Results show that optimal model selection depends on each series' individual demand pattern. ETS dominated in two of the three continuous-consumption items; SARIMAX outperformed in the only case where exogenous variables showed statistically relevant correlation with demand. For intermittent series, the hurdle model outperformed classical methods in two of three items. Monte Carlo simulation produced reorder points and safety stocks targeting a 95% service level.
The backtesting of the proposed policy on real data (August 2025 – March 2026) quantified a potential release of $76.5 million COP in immobilized capital, equivalent to 34.3% of the average observed inventory for the six evaluated SKUs, without compromising the service level in the continuous-consumption series. The work demonstrates that a reproducible analytical pipeline, built on historical consumption records and publicly available exogenous variables, can institutionalize quantitative replenishment criteria in environments where decisions currently rely on operational intuition.
Keywords: MRO inventory management, Syntetos-Boylan classification, intermittent demand forecasting, SARIMAX, ETS, Hurdle-RF, rolling-origin backtesting, Monte Carlo simulation, reorder point, coking coal mining.

Índice

Resumen **i**

Abstract **i**

1 Introducción **1**
1.1 Planteamiento del problema 1
1.2 Justificación 2

2 Objetivos **4**
2.1 Objetivo general 4
2.2 Objetivos específicos 4

3 Marco teórico **4**

3.1 Antecedentes 4
3.2 Revisión de la literatura 6
3.2.1 Pronóstico de series continuas: SARIMA, SARIMAX y suavizamiento exponencial 6
3.2.2 Pronóstico de demanda intermitente: métodos clásicos y modelos de dos etapas 7
3.2.3 Política de inventario, stock de seguridad y simulación Monte Carlo 8
3.3 Definición de conceptos 9
3.3.1 C. I. Milpa S. A. y aliadas 9
3.3.2 Gestión de Inventarios MRO (Mantenimiento, Reparación y Operaciones) . 10
3.3.3 Matriz de Kraljic y artículos estratégicos 10
3.3.4 Clasificación de Syntetos-Boylan 12
3.3.4.1 ADI (Average Demand Interval / Intervalo promedio entre demandas) 13
3.3.4.2 CV^2 (Coeficiente de variación al cuadrado) 13
3.3.5 Modelos de Series de Tiempo para Series Estables 14
3.3.5.1 Seasonal Naive 14
3.3.5.2 Modelo ETS 14
3.3.5.3 Modelo SARIMAX 15
3.3.6 Modelos de series de tiempo para series intermitentes 15
3.3.6.1 Método Croston-SBA 16
3.3.6.2 Método TSB 16
3.3.6.3 Modelo Hurdle con bosque aleatorio (Hurdle-RF) 16
3.3.7 Métricas de evaluación del pronóstico 17
3.3.7.1 Error absoluto medio (MAE) 17
3.3.7.2 Raíz del error cuadrático medio (RMSE) 18
3.3.7.3 Error porcentual absoluto medio simétrico (sMAPE) 18
3.3.7.4 Error porcentual absoluto medio ponderado (WAPE) 18
3.3.7.5 Error absoluto medio escalado (MASE) 18
3.3.7.6 Sesgo relativo (Bias) 19
3.3.7.7 Precisión de actividad (*Activity accuracy*) 19
3.3.8 Diagnóstico estadístico de residuos 19
3.3.8.1 Prueba de Ljung-Box 20
3.3.8.2 Prueba de Jarque-Bera 20
3.3.8.3 Prueba ARCH-LM 20
3.3.9 Diagnóstico de covariables 21

3.3.9.1 Factor de inflación de la varianza (VIF) . . . 21
3.3.9.2 Correlación de Spearman con rezagos . . . 21
3.3.10 Validación fuera de muestra mediante *rolling-origin* (ventana rodante) . . . 22
3.3.11 Políticas de Reabastecimiento . . . 22
3.3.11.1 Punto de Reorden (ROP) . . . 22
3.3.11.2 Stock de Seguridad y simulación Monte Carlo . . . 23

4 Metodología **24**
4.1 Enfoque de investigación . . . 24
4.2 Diseño y tipo de investigación . . . 24
4.3 Población y muestra . . . 24
4.4 Técnicas e instrumentos de recolección de datos . . . 25
4.5 Procedimiento y análisis de datos . . . 26

5 Resultados **28**
5.1 Clasificación de la demanda y auditoría de la base de datos . . . 28
5.2 Diagnóstico exploratorio y econométrico . . . 30
5.2.1 Estacionariedad y estacionalidad . . . 30
5.2.2 Diagnóstico de residuos del modelo ganador (series estables) . . . 33
5.2.3 Variables exógenas: colinealidad y correlación con la demanda . . . 34
5.3 Backtesting rolling-origin y selección de modelos . . . 36
5.4 Pronóstico a 12 meses y política de inventario . . . 40
5.4.1 Trayectorias de pronóstico . . . 41
5.4.2 Parámetros de política de inventario . . . 41
5.5 Backtesting de política de inventario: gestión real frente al sistema propuesto . . . 44
5.5.1 Supuestos de la simulación . . . 45
5.5.2 Resultados de la simulación . . . 45
5.5.3 Valorización financiera . . . 47

6 Discusión y análisis de resultados **48**

7 Conclusiones **51**

8 Trabajos futuros **53**

Índice de tablas

1 *Clasificación de la demanda según los umbrales de Syntetos-Boylan* . . . 12

2 SKUs seleccionados para el modelado de demanda 25
3 Resumen estructural del dataset 28
4 Auditoría de demanda por SKU en vida activa 28
5 Diagnóstico preliminar de estacionariedad y estacionalidad 33
6 Diagnóstico de residuos del modelo ganador por SKU (series estables) 34
7 Colinealidad entre variables exógenas 35
8 Mayor correlación absoluta por serie frente a variables exógenas con rezagos cortos 35
9 Especificación SARIMAX utilizada por serie estable 37
10 Métricas de backtesting rolling-origin: series estables 38
11 Métricas de backtesting rolling-origin: series intermitentes 38
12 Modelo seleccionado por serie según ranking compuesto 39
13 Desempeño promedio por familia de serie y modelo 40
14 Escenario base de variables exógenas para el horizonte de pronóstico (2026) 41
15 Parámetros de política de inventario por SKU 43
16 Backtesting de política de inventario: indicadores por SKU 45
17 Valorización del capital promedio inmovilizado por SKU (COP) 47

Índice de figuras

1 Matriz de Kraljic aplicada a los insumos de mantenimiento en Milpa S.A. 11
2 Patrones de comportamiento de la demanda según variabilidad de frecuencia y cantidad 13
3 Disponibilidad temporal y presencia de ceros en la vida activa 29
4 Comportamiento histórico mensual por SKU 31
5 Descomposición STL por serie 32
6 Matriz de correlación de Spearman entre consumos de SKU y variables exógenas . 36
7 Comparación visual del WAPE y ranking compuesto por serie y modelo 39
8 Pronóstico final a 12 meses por SKU con bandas de incertidumbre 42
9 Puntos de reorden y stocks de seguridad por familia de demanda 44
10 Evolución diaria del inventario: gestión real vs. sistema propuesto 46
11 Valorización financiera del backtesting: capital promedio inmovilizado por escenario 48

1. Introducción

1.1 Planteamiento del problema

C.I. Milpa S.A. y sus compañías aliadas (en adelante, Milpa) operan en Colombia una cadena integrada de minería de carbón coquizable y producción de coque metalúrgico, insumo esencial para industrias como la siderurgia, la fundición y las ferroaleaciones. La continuidad de esta operación, caracterizada por activos intensivos en capital, altos requerimientos de mantenimiento y múltiples frentes productivos, depende de manera crítica de la disponibilidad oportuna de repuestos y materiales MRO (Mantenimiento, Reparación y Operaciones), tales como rodamientos, bandas, componentes eléctricos, ferretería industrial y elementos asociados al mantenimiento preventivo y correctivo. Visto así, la bodega además de cumplir con su función logística, se convierte también en un habilitador directo de la confiabilidad operativa.

Sin embargo, Milpa enfrenta un sobrecosto relevante asociado al sobrestock. Actualmente, la compañía mantiene inventarios MRO valuados en aproximadamente $40.000 millones COP con niveles significativos de baja rotación o inmovilización de existencias. Este capital detenido tiene un impacto financiero material: además del costo de oportunidad, el inventario genera costos de conservación (almacenamiento, seguros, deterioro, pérdidas, obsolescencia y gestión) que, en términos referenciales, suelen representar una fracción no menor del valor inventariado a lo largo del año. En consecuencia, el exceso de inventario no solo ocupa capacidad física y administrativa, sino que también incrementa el gasto operativo sin traducirse necesariamente en mayor productividad.

El problema central no se limita a la presencia de quiebres de stock (que tienden a ser poco frecuentes por la política implícita de "pedir de más"), sino a la gestión empírica de la incertidumbre. Las decisiones de reposición y compra se realizan de forma reactiva, con un alto peso de la intuición y del temor al desabastecimiento, y con un uso limitado de la información histórica. En particular, la empresa no cuenta con modelos estadísticos o analíticos formales para estimar la demanda, ni con un marco cuantitativo que permita balancear explícitamente el compromiso entre costo y servicio. Esta ausencia de modelación impide responder, con sustento técnico, a preguntas operativas clave: cuánto comprar, cuándo comprar y qué nivel de inventario es razonable para sostener el servicio sin sobredimensionar existencias.

Desde la perspectiva de la Ciencia de Datos aplicada a operaciones, el reto se intensifica porque la demanda de repuestos en minería y coque no siempre sigue patrones estables, presentando a veces intermitencia, y está condicionada por factores que van más allá del historial de consumos. Variables como el clima, el ritmo de producción, las horas de uso de maquinaria, la influencia de la tasa de cambio en importaciones, la ocurrencia de mantenimientos no programados y los cambios en las condiciones operativas pueden alterar la probabilidad y magnitud del consumo. Al no incorporar estas variables exógenas en la toma de decisiones, la incertidumbre se gestiona de

manera conservadora mediante compras excesivas, trasladando el riesgo del desabastecimiento al costo financiero y operacional del sobreinventario.

Para este trabajo se propone abordar el problema mediante una prueba de concepto: no se busca optimizar la totalidad del inventario, sino desarrollar un enfoque sólido sobre una muestra estratégica de artículos priorizados por su criticidad y riesgo de suministro. La necesidad es avanzar desde decisiones basadas en percepción hacia decisiones reproducibles y auditables, soportadas por modelos de pronóstico y criterios de reposición. Por ello, se plantea la implementación de modelos de series de tiempo que integren factores externos para estimar la demanda y apoyar políticas de reposición más profesionales. Este primer resultado funcionará como hoja de ruta para una adopción gradual de analítica en la bodega y, a mediano plazo, para impulsar una cultura de datos en la cadena de suministro de Milpa.

1.2 Justificación

La gestión de inventarios MRO en Milpa afecta tanto a la misma bodega como a la rentabilidad y el flujo de caja de la operación. En una empresa minera y de coque, donde la continuidad depende de la disponibilidad de repuestos, es natural que el sistema tienda a ser conservador. No obstante, cuando la incertidumbre se gestiona principalmente con "colchones" permanentes, el resultado es un uso ineficiente de los recursos disponibles. En este sentido, la Ciencia de Datos se justifica como una herramienta de eficiencia financiera, porque permite convertir información dispersa (consumos, producción, clima, uso de activos) en reglas de reposición que reduzcan costos sin comprometer el servicio interno.

Desde el punto de vista económico, intervenir inventarios representa una palanca especialmente relevante para Milpa por su margen real de maniobra. De acuerdo con la jefatura de almacenes, cerca del 80% de los costos se concentra en componentes externos y poco controlables (principalmente transporte y compra de carbón), mientras que el 20% restante corresponde a gastos en los que la empresa sí puede incidir de forma recurrente, como operación, mantenimiento y MRO. En ese 20%, la bodega se convierte en un punto estratégico: liberar capital inmovilizado en inventario más allá de "ahorrar" implica, a su vez, optimizar el capital de trabajo. En la práctica, los recursos que hoy están "enterrados" en estanterías podrían redirigirse a mantenimiento mayor, inversiones en confiabilidad, mejoras de capacidad o iniciativas de expansión. Por ello, el aporte financiero del proyecto reduce el valor del inventario y, sobre todo, mejora el uso del efectivo y el costo de oportunidad asociado.

La justificación organizacional del proyecto va más allá de una reducción puntual de costos: busca migrar de la intuición a la gobernanza. En entornos industriales, la sobrecompra suele estar alimentada por un incentivo claro: "mejor tener de más que parar la operación". El problema es que, sin un marco cuantitativo, esa lógica se traduce en decisiones difíciles de auditar, replicar y mejorar

con el tiempo. Implementar un enfoque basado en datos permite institucionalizar conocimiento operativo en forma de reglas explícitas (por ejemplo, puntos de reorden y stocks de seguridad) que puedan ser revisadas, ajustadas y defendidas técnicamente. En otras palabras, la Ciencia de Datos aquí actúa como un mecanismo para transformar decisiones individuales, altamente dependientes de la experiencia y percepción, en un sistema consistente de reposición alineado con objetivos de costo y servicio.

En términos académicos y técnicos, el valor del trabajo reside en integrar el pronóstico y la política de inventario como un solo problema operativo. En la práctica, medir únicamente la precisión de un modelo es insuficiente si el resultado no se traduce en decisiones concretas de compra. Por ello, el proyecto se justifica en desarrollar un enfoque donde el pronóstico de demanda no sea un fin, sino un insumo para definir cuánto comprar y cuándo comprar, incorporando explícitamente la incertidumbre y el desempeño esperado en el servicio. Esta integración es clave para que el análisis tenga un impacto real: conecta el componente predictivo con la toma de decisiones operativas y permite evaluar el beneficio no solo en métricas estadísticas, sino también en indicadores de inventario (nivel de servicio, quiebres, rotación, costo total).

Adicionalmente, la naturaleza del consumo MRO en minería exige un tratamiento metodológico que la gestión tradicional difícilmente captura. La demanda puede ser intermitente, con periodos largos sin consumo y picos asociados a eventos operativos; además, suele estar influenciada por variables externas, como condiciones climáticas, ritmo de producción u horas de uso de maquinaria. En este contexto, incorporar variables exógenas en modelos de series de tiempo es una respuesta adecuada a un sistema donde la demanda no se comporta de forma aislada. Un modelo que aprende relaciones entre la operación y el consumo puede estimar mejor el riesgo de agotamiento y reducir la necesidad de "comprar por miedo", habilitando una reducción del inventario con menor impacto en el servicio.

Finalmente, el proyecto se justifica por su escalabilidad. Aunque el alcance se limita a una muestra estratégica de artículos críticos y de alto riesgo de suministro, el objetivo es construir una metodología replicable: un "template" de datos, modelación y reglas de reposición que pueda extenderse progresivamente a otras familias de materiales. Demostrar que el enfoque funciona en ítems donde el costo de equivocarse es alto (ya sea por importancia en la producción o riesgo de suministro) genera evidencia interna, acelera la adopción y deja una hoja de ruta para una transformación más amplia en la cadena de suministro. Así, el aporte del trabajo trasciende el resultado obtenido sobre un conjunto de repuestos, sentando bases técnicas y operativas para una gestión de inventarios gobernada por datos.

2. Objetivos

2.1 Objetivo general

Desarrollar y evaluar un sistema de pronóstico de demanda basado en modelos de series de tiempo e integración de variables exógenas para C.I. Milpa S.A., con el fin de institucionalizar una toma de decisiones basada en datos que reemplace la gestión empírica actual, permitiendo optimizar los niveles de inventario sin comprometer la continuidad de la operación minera.

2.2 Objetivos específicos

1. Caracterizar el comportamiento histórico de la demanda de artículos estratégicos mediante técnicas de análisis exploratorio y descomposición de series, para identificar patrones, estacionalidades y la influencia potencial de variables operativas externas.

2. Desarrollar y entrenar modelos de series de tiempo con esquemas de validación cruzada para series temporales, seleccionando las arquitecturas con mayor capacidad predictiva.

3. Diseñar reglas de reabastecimiento (puntos de reorden y stocks de seguridad) que integren los pronósticos obtenidos con las restricciones logísticas de la empresa, traduciendo los resultados técnicos en parámetros operativos ejecutables.

4. Validar el impacto del sistema propuesto mediante un análisis backtesting, comparando el desempeño del modelo frente a la gestión histórica de la empresa para estimar la reducción potencial de sobrestock y el nivel de servicio alcanzado.

3. Marco teórico

3.1 Antecedentes

La gestión de inventarios de repuestos en industrias intensivas en activos ha sido objeto de estudio académico y empresarial durante décadas, aunque la intersección específica entre la minería de carbón, los inventarios MRO y el uso de modelos cuantitativos de pronóstico sigue siendo un área con espacio notable para la investigación aplicada. Los antecedentes más relevantes para el presente trabajo se ubican en tres frentes: los estudios sobre pronóstico de demanda intermitente en repuestos, las iniciativas de analítica aplicada a inventarios en empresas mineras y de manufactura intensiva, y los trabajos que integran variables exógenas (como condiciones climáticas o macroeconómicas) dentro de modelos de series de tiempo para la gestión de materiales.

En el ámbito del pronóstico de demanda intermitente, el trabajo seminal de Croston (1972) estableció las bases para tratar series de tiempo con muchos periodos de demanda nula, descomponiendo el proceso en la frecuencia de ocurrencia de demanda y la magnitud de cada evento. Décadas

después, Syntetos et al. (2005) demostraron que el estimador de Croston es sesgado y propusieron una corrección aproximadamente insesgada, conocida en la literatura como SBA (*Syntetos–Boylan Approximation*), que reduce el error paramétrico sin incrementar la complejidad computacional. Esta corrección fue validada empíricamente sobre más de 18.750 SKUs de la Royal Air Force del Reino Unido y se convirtió en referencia obligada para el pronóstico de repuestos con demanda esporádica. Posteriormente, Teunter et al. (2011) identificaron una limitación estructural tanto de Croston como de SBA: al actualizar el intervalo entre demandas únicamente en periodos con consumo positivo, los métodos no ajustan los pronósticos a la baja en presencia de obsolescencia o reducción estructural de la demanda. Como respuesta, propusieron el método TSB (*Teunter–Syntetos–Babai*), que actualiza en cada periodo la probabilidad de ocurrencia de demanda, resultando más robusto ante cambios de nivel y no estacionariedad descendente (Teunter et al., 2011).

En paralelo, la literatura sobre pronóstico de repuestos en minería ha evidenciado que las características de esta demanda (alta intermitencia, alta variabilidad en el tamaño cuando ocurre consumo, y dependencia de condiciones operativas) requieren métodos específicos y que el desempeño de los enfoques clásicos puede deteriorarse significativamente si no se considera la naturaleza del contexto industrial (Rego & Mesquita, 2011; Zhang et al., 2021). Romeijnders et al. (2012) fueron de los primeros en documentar sistemáticamente el comportamiento de distintos métodos de pronóstico en entornos de repuestos y señalaron que la elección del estimador tiene un impacto directo sobre las políticas de inventario resultantes. Más recientemente, Driessen et al. (2015) desarrollaron un método híbrido basado en heurísticas y *bootstrapping* específicamente diseñado para empresas mineras del sector del hierro en Brasil, destacando que la cantidad de ítems analizada en ese estudio no tenía precedentes en la literatura que combinara minería y pronóstico de repuestos, y exponiendo características de clasificación de la demanda que difieren de las encontradas en otros sectores industriales. Este trabajo es particularmente relevante como antecedente directo del presente, dado que aborda el mismo sector de industria intensiva en activos y propone un método empíricamente evaluado sobre datos reales de operaciones mineras.

En lo que respecta al uso de variables exógenas para enriquecer el pronóstico de consumo, Barabadi et al. (2014) demostraron que ignorar las condiciones operativas, incluyendo factores climáticos y de entorno, puede generar diferencias de hasta el 25% en la estimación del número de repuestos requeridos en equipos mineros. Esta evidencia sustenta la pertinencia de incorporar variables como temperatura y precipitación en modelos de pronóstico orientados a operaciones a cielo abierto, donde las condiciones meteorológicas afectan directamente el ritmo de desgaste de los activos. Desde una perspectiva macroeconómica, la tasa de cambio ha sido identificada como una variable de contexto relevante para el pronóstico de consumo de materiales importados, en la medida en que sus variaciones modifican el costo de reposición y pueden alterar el comportamiento de compra anticipada (Wagner & Bode, 2009).

Finalmente, en cuanto a la integración entre pronóstico y política de inventario, la revisión crítica de Syntetos et al. (2016) ha puesto de relieve la brecha existente entre la literatura de pronóstico, que optimiza métricas de error estadístico, y la literatura de control de inventarios, que evalúa el desempeño operativo (nivel de servicio, costos de ruptura y de mantenimiento). El uso de simulación Monte Carlo para derivar parámetros de inventario (stock de seguridad y punto de reorden) a partir de la distribución empírica de la demanda pronosticada representa una respuesta práctica a esta brecha, al no imponer supuestos de normalidad sobre una distribución que, en repuestos estratégicos, puede ser asimétrica o multimodal (Dolgui & Pashkevich, 2008).

3.2 Revisión de la literatura

La revisión de la literatura se organiza en torno a los tres ejes metodológicos que articulan el presente trabajo: (i) el pronóstico de demanda continua con variables exógenas mediante modelos SARIMA/SARIMAX y suavizamiento exponencial; (ii) el pronóstico de demanda intermitente con métodos clásicos y enfoques basados en aprendizaje automático de dos etapas; y (iii) la derivación de políticas de inventario mediante simulación, con énfasis en entornos de demanda no normal y repuestos estratégicos en industrias intensivas en activos.

3.2.1 Pronóstico de series continuas: SARIMA, SARIMAX y suavizamiento exponencial

Los modelos ARIMA y su extensión estacional SARIMA constituyen el punto de partida estándar para el pronóstico de series de tiempo con componentes autorregresivos, de medias móviles y diferenciación (Box et al., 2016). La extensión a SARIMAX incorpora variables exógenas observables o regresores externos que permiten al modelo condicionar el pronóstico a información adicional más allá del pasado de la propia serie. Esta capacidad es especialmente valiosa cuando existen factores del entorno ya sean climáticos, macroeconómicos u operativos, que se sabe que influyen sobre la variable de interés, pero que no quedan capturados en los rezagos del consumo histórico.

En el contexto del pronóstico de demanda industrial, Nikolopoulos et al. (2011) destacaron la utilidad de enfoques multivariados para series de insumos de planta cuando las condiciones del entorno tienen correlación medible con el consumo. Kourentzes (2013) comparó modelos ARIMA univariados contra especificaciones que incorporan indicadores externos y mostraron que la ganancia de precisión depende del grado de correlación entre los regresores y la serie objetivo: cuando dicha correlación es estadísticamente significativa, incluir la variable exógena mejora sistemáticamente el error fuera de muestra. Resultados análogos han sido reportados para series industriales con componente climático, donde la temperatura y la precipitación han sido empleadas como predictores de demanda de insumos en operaciones de manufactura y extracción (Barabadi et al., 2014; Wagner & Bode, 2009).

Dentro de la familia de modelos de suavizamiento exponencial, los modelos ETS (*Error–*

Trend–Seasonality), también denominados modelos de espacio de estados, ofrecen una alternativa parsimoniosa para series sin regresores exógenos. Hyndman et al. (2008) sistematizaron esta familia y demostraron que, para series con tendencia amortiguada, la variante ETS con amortiguamiento ofrece un buen equilibrio entre flexibilidad y riesgo de sobreajuste, lo que los convierte en un *benchmark* robusto para evaluar la ganancia marginal de un modelo más complejo como SARIMAX. En la práctica, la comparación entre ETS y SARIMAX bajo un esquema de validación *rolling-origin* permite establecer si la incorporación de variables exógenas justifica la mayor complejidad paramétrica para cada serie particular (Makridakis et al., 2020).

La validación fuera de muestra mediante *rolling-origin backtesting* ha sido identificada como el esquema de evaluación más adecuado para series de tiempo en contextos de pronóstico operacional, ya que preserva la causalidad temporal y evita la filtración de información futura al conjunto de entrenamiento (Tashman, 2000). En entornos con históricos cortos, como los que presenta una operación minera con 7 años de datos mensuales por SKU (Stock Keeping Unit o Unidad de Mantenimiento de Existencias, código único asignado a cada producto de una empresa), la elección del tamaño de ventana y el número de iteraciones del *rolling-origin* tiene un impacto directo sobre la estabilidad de los estimadores de error y, en consecuencia, sobre la selección del modelo final (Makridakis et al., 2020).

3.2.2 Pronóstico de demanda intermitente: métodos clásicos y modelos de dos etapas

La demanda intermitente, caracterizada por periodos de consumo nulo alternados con eventos de demanda positiva, representa uno de los desafíos más documentados en la gestión de inventarios de repuestos (Boylan & Syntetos, 2010). El esquema de clasificación de Syntetos et al. (2005) a partir del intervalo promedio entre demandas (ADI) y el coeficiente de variación al cuadrado (CV^2) provee un criterio operacional para asignar ítems a familias de modelos: estable, intermitente, errática e irregular. Este esquema es ampliamente utilizado en la literatura aplicada y constituye el punto de partida para la selección de métodos en estudios de inventario MRO (Kaya et al., 2020).

En este marco, el método de Croston y su variante SBA siguen siendo los modelos de referencia para demanda intermitente pura (alta frecuencia de ceros, baja variabilidad en el tamaño de la demanda cuando ocurre). Su popularidad radica en la simplicidad computacional, la transparencia del mecanismo de actualización y la disponibilidad de resultados teóricos sobre sesgo y varianza de los estimadores (Johnston & Boylan, 1996; Syntetos et al., 2005). El método TSB extiende esta familia al actualizar también en periodos de demanda nula, lo que resulta ventajoso cuando el nivel de actividad de la pieza está declinando o cuando existe riesgo de obsolescencia Teunter et al. (2011).

La limitación fundamental de estos métodos clásicos es que modelan la serie como un proceso estacionario en el largo plazo y no incorporan información contextual sobre condiciones operati-

vas, macroeconómicas o de mantenimiento que puedan modificar la probabilidad de ocurrencia de demanda o su magnitud. Para superar esta restricción, la literatura reciente ha explorado enfoques de aprendizaje automático de dos etapas (modelos *hurdle*), en los que una primera etapa modela la probabilidad de que la demanda sea positiva (clasificación binaria) y una segunda etapa modela el tamaño de la demanda condicional a que sea no nula (regresión) (Kourentzes, 2013; Pennings et al., 2017). Este enfoque permite incorporar variables exógenas, incluyendo rezagos del consumo, indicadores climáticos o macroeconómicos, de forma directa y natural en ambas etapas.

En el contexto específico de repuestos con demanda intermitente en industrias de activos intensivos, Bacchetti y Saccani (2012) revisaron más de 100 contribuciones académicas y concluyeron que los modelos de aprendizaje automático comienzan a superar a los métodos clásicos cuando el número de periodos históricos es suficiente para entrenamiento y cuando se dispone de covariables relevantes. Sin embargo, también advirtieron que con series cortas la capacidad de aprendizaje de modelos complejos puede estar limitada y el riesgo de sobreajuste aumenta, lo que justifica utilizar los modelos de dos etapas como una alternativa complementaria, más no sustituta, a los métodos clásicos (Bacchetti & Saccani, 2012).

Un trabajo particularmente próximo al presente es el de Sillanpää y Liesiö (2018), que propone un modelo de dos etapas con bosque aleatorio (*Random Forest*) para el pronóstico de repuestos industriales con demanda esporádica. Los autores muestran que el enfoque de dos etapas supera consistentemente a Croston-SBA cuando se dispone de variables operativas adicionales, y destacan que el mecanismo de *feature importance* del bosque aleatorio permite identificar cuáles covariables contribuyen más a explicar la ocurrencia y la magnitud de la demanda, lo que facilita la comunicación de los resultados a equipos de mantenimiento y logística. En una línea similar, Boylan et al. (2022) señalaron que los enfoques de conjunto (*ensemble*) que combinan modelos clásicos de demanda intermitente con componentes de aprendizaje automático muestran ganancias consistentes en precisión fuera de muestra, especialmente en contextos con alta variabilidad en el tamaño de los eventos de demanda.

3.2.3 Política de inventario, stock de seguridad y simulación Monte Carlo

La derivación de parámetros de inventario a partir de pronósticos de demanda es el eslabón que conecta la modelación estadística con las decisiones operativas. En el modelo clásico de punto de reorden (ROP) con stock de seguridad, el cálculo se basa en suponer que la demanda durante el *lead time* (tiempo de reposición) sigue una distribución normal, lo que permite derivar analíticamente el stock de seguridad como múltiplo de la desviación estándar de la demanda (Rego & Mesquita, 2011). Sin embargo, esta suposición es cuestionable para repuestos con demanda intermitente o asimétrica: en presencia de muchos ceros y colas pesadas, la distribución real de la demanda durante el plazo de reposición puede estar muy alejada de la normalidad, lo que hace

que las fórmulas analíticas subestimen el stock necesario para alcanzar un determinado nivel de servicio (Dolgui & Pashkevich, 2008).

La simulación Monte Carlo ofrece una alternativa robusta que no requiere supuestos de forma distribucional: a partir de la distribución empírica de los errores de pronóstico o de la serie histórica de consumos, se generan miles de trayectorias posibles de demanda durante el *lead time* y se calculan directamente los cuantiles de la distribución simulada. El stock de seguridad se fija de modo que el cuantil correspondiente al nivel de servicio deseado cubra la demanda durante el plazo de reposición (Dolgui & Pashkevich, 2008). Este enfoque es especialmente pertinente para ítems estratégicos en minería, donde la asimetría de la distribución de consumo es la norma y donde el costo de ruptura es desproporcionadamente alto respecto al costo de mantenimiento de inventario (Rego & Mesquita, 2011; Zhang et al., 2021).

Strijbosch et al. (2000) compararon el enfoque analítico y el de simulación para la determinación del stock de seguridad en entornos de demanda no normal y concluyeron que la simulación produce niveles de servicio más cercanos al objetivo cuando la demanda presenta asimetría positiva, lo que es precisamente el caso de repuestos con predominio de ceros seguidos de eventos de demanda ocasionalmente grandes. Adicionalmente, trabajos como el de Babai et al. (2013) han señalado que la variabilidad del lead time amplifica los efectos de la no normalidad, haciendo aún más recomendable el uso de simulación cuando los tiempos de reposición de los ítems varían entre pedidos, como ocurre típicamente en importaciones de repuestos estratégicos.

En concordancia con lo anterior, se confirma que el problema abordado en el presente trabajo: pronóstico de demanda MRO en minería con variables exógenas e integración con política de inventario, se ubica en la intersección de varias corrientes de investigación activas, y que la combinación de SARIMAX con exógenas para series continuas, métodos clásicos de demanda intermitente (Croston-SBA y TSB) y un modelo de dos etapas con bosque aleatorio (Hurdle-RF) para series esporádicas, validados mediante *rolling-origin backtesting* y traducidos a parámetros de inventario por simulación Monte Carlo, representa un aporte metodológico coherente con el estado del arte y adaptado a las características concretas del entorno operativo analizado.

3.3 Definición de conceptos

3.3.1 C. I. Milpa S. A. y aliadas

C.I. Milpa S.A. es una empresa colombiana del sector minero–industrial dedicada a la minería de carbón coquizable (metalúrgico), beneficio (lavado), coquización y comercialización del coque resultante, insumo utilizado en industrias como la siderurgia, la fundición y las ferroaleaciones (Unidad de Planeación Minero Energética (UPME), 2024). En su comunicación institucional, la compañía se presenta como productor líder de coque metalúrgico en Colombia (C.I. Milpa S.A., n.d.).

Para efectos de este trabajo, cuando se menciona “Milpa y aliadas”, o simplemente ”Milpa”, se adopta una convención operativa, que se refiere a un grupo de empresas (C.I. Milpa S.A., Incarsa S.A.S., Carboinsa S.A.S. y Uniminas S.A.S.) que trabajan de forma integrada como una sola operación, compartiendo recursos como mano de obra, procesos, materiales y soporte logístico. Esta definición se utiliza únicamente para delimitar la unidad de análisis y evitar ambigüedades organizacionales.

3.3.2 Gestión de Inventarios MRO (Mantenimiento, Reparación y Operaciones)

La gestión de inventarios MRO se refiere al conjunto de prácticas para planificar, abastecer, almacenar y reponer los materiales y repuestos necesarios para el mantenimiento y la operación de los activos, con la particularidad de que estos ítems no se incorporan al producto final, pero sí condicionan la continuidad de la operación (Association for Supply Chain Management (ASCM), n.d. IBM, n.d.). En industrias intensivas en activos, como la minería y el coque, el inventario MRO suele incluir repuestos, consumibles y componentes críticos cuya no disponibilidad puede traducirse en demoras de mantenimiento o paradas; por lo que el objetivo no es “minimizar inventario” en abstracto, sino garantizar la disponibilidad con un costo total controlado (Rego & Mesquita, 2011; Zhang et al., 2021).

Desde el punto de vista técnico, una forma estándar de operacionalizar esta gestión es mediante políticas de control, como el punto de reorden (ROP) y el stock de seguridad, que definen cuándo y cuánto reponer en condiciones de incertidumbre, explicadas más adelante.

3.3.3 Matriz de Kraljic y artículos estratégicos

La matriz de Kraljic es un modelo de segmentación del portafolio de compras que clasifica materiales y servicios según dos dimensiones: impacto en el negocio y riesgo de suministro. Su propósito es diferenciar las estrategias de abastecimiento, pasando de una gestión homogénea de compras a un enfoque selectivo según el tipo de ítem. El resultado es una matriz 2×2 con cuatro cuadrantes (Gelderman & Van Weele, 2005; Kraljic, 1983):

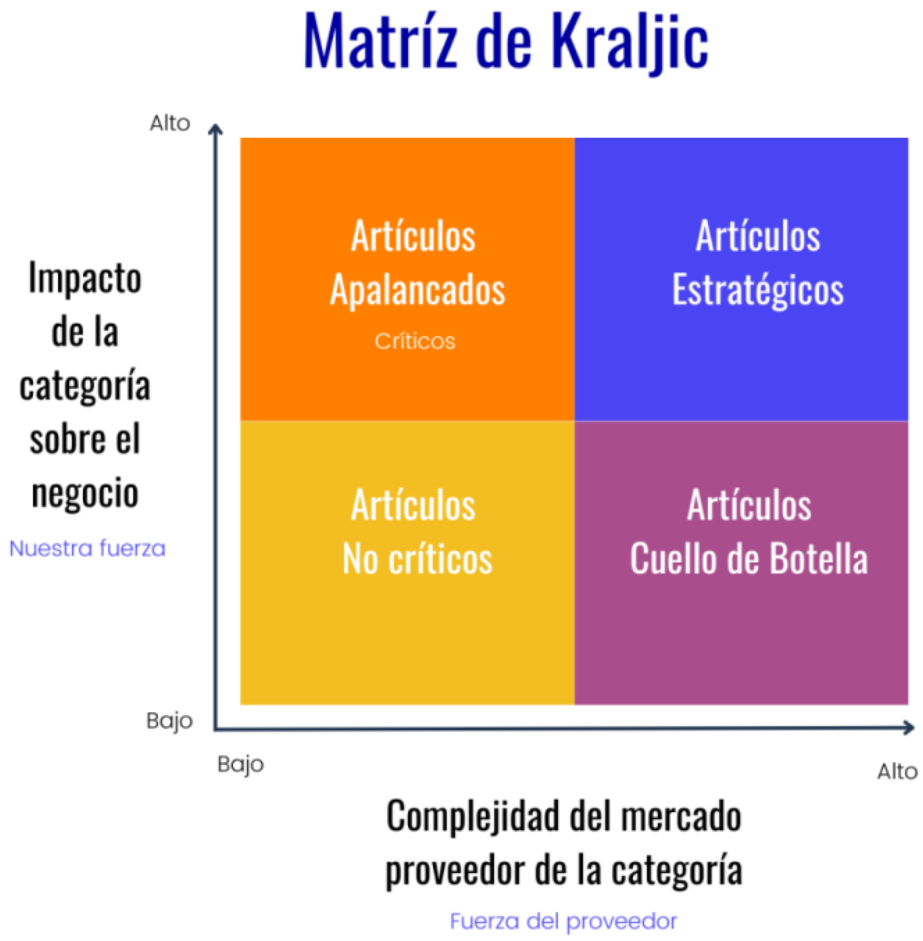
1. No críticos: bajo impacto y bajo riesgo.
2. Apalancables: alto impacto y bajo riesgo.
3. Cuello de botella: bajo impacto y alto riesgo.
4. Estratégicos: alto impacto y alto riesgo.

Los artículos estratégicos (alto impacto y alto riesgo) se caracterizan por combinar importancia financiera/operativa con vulnerabilidad de abastecimiento: suelen tener mercados de proveedores limitados, tiempos de reposición sensibles o alta dependencia tecnológica, y cualquier falla de suministro puede afectar significativamente la continuidad del negocio (Chartered Institute of

Procurement & Supply (CIPS), n.d.). Por eso, Kraljic (1983) sugiere que estos ítems requieren una gestión más cercana, con estrategias orientadas a asegurar continuidad (por ejemplo, el desarrollo de proveedores, acuerdos a largo plazo, estandarización o planes de contingencia), más que únicamente negociaciones de precio, tal como ilustra la Figura 1 para el caso de Milpa.

Figura 1

Matriz de Kraljic aplicada a los insumos de mantenimiento en Milpa S.A.



Nota. Adaptado de Esolbay (2024), material divulgativo sobre segmentación de portafolio de compras.

En este trabajo, el enfoque en artículos estratégicos se justifica porque, en inventarios MRO para minería, este cuadrante concentra repuestos cuya no disponibilidad puede generar impactos operativos desproporcionados y cuya reposición está expuesta a incertidumbre de suministro (p. ej., importaciones, pocos fabricantes, tiempos de entrega largos). En consecuencia, priorizarlos permite diseñar políticas de reposición con mayor rigor y obtener beneficios prácticos sin intentar abarcar

todo el inventario desde el inicio.

3.3.4 Clasificación de Syntetos-Boylan

La clasificación de Syntetos–Boylan es un esquema para caracterizar series de demanda (frecuente en repuestos y materiales MRO) a partir de dos métricas: el intervalo promedio entre demandas no nulas (Average Demand Interval - ADI) y la variabilidad del tamaño de la demanda (Coeficiente de Variación al Cuadrado - CV^2). Con base en estos dos indicadores, Syntetos et al. (2005) proponen separar los ítems en cuatro patrones a partir de los umbrales definidos en la Tabla 1:

1. Estable: demanda regular en frecuencia y cantidad.
2. Intermitente: demanda con periodos con demanda nula (0)
3. Errática: demanda constante en frecuencia pero con grandes variaciones en la cantidad.
4. Irregular: demanda con alta variabilidad tanto en tiempo como en cantidad.

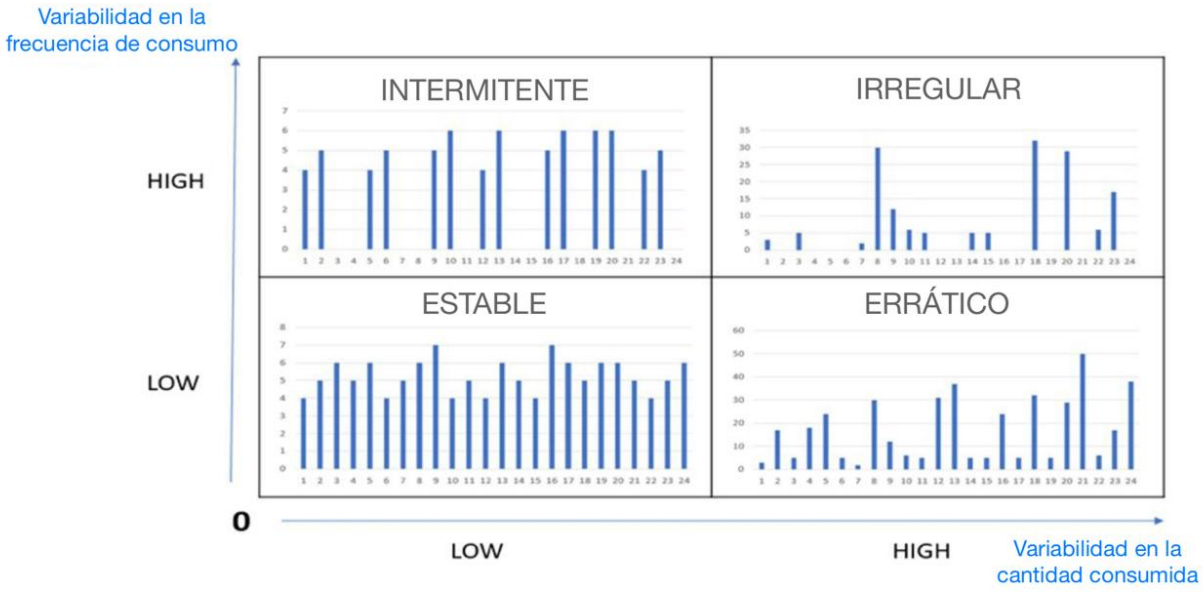
Tabla 1
Clasificación de la demanda según los umbrales de Syntetos-Boylan

Categoría	**Intervalo (ADI)**	**Variabilidad (CV^2)**
Estable	$ADI < 1.32$	$CV^2 < 0.49$
Intermitente	$ADI \geq 1.32$	$CV^2 < 0.49$
Errática	$ADI < 1.32$	$CV^2 \geq 0.49$
Irregular	$ADI \geq 1.32$	$CV^2 \geq 0.49$

La utilidad de esta clasificación, en el contexto de inventarios MRO (por ejemplo, repuestos en minería), es que permite entender de forma rápida el tipo de incertidumbre dominante: si el reto principal es la intermitencia (muchos ceros) o la variabilidad del tamaño cuando ocurre el consumo. En consecuencia, sirve como base para priorizar métodos de pronóstico y decisiones de inventario coherentes con el comportamiento observado, en lugar de tratar todos los artículos con el mismo enfoque (Kaya et al., 2020; Syntetos et al., 2005). La Figura 2 permite entender de manera visual cómo se distribuyen cada una de estas clasificaciones según su patrón de demanda.

Figura 2

Patrones de comportamiento de la demanda según variabilidad de frecuencia y cantidad



Nota. Representación visual de los cuadrantes de demanda (Estable, Intermitente, Errática e Irregular) basada en la fluctuación del consumo y su periodicidad. Adaptado de https://frepple.com/blog/demand-classification/

3.3.4.1 ADI **(Average Demand Interval / Intervalo promedio entre demandas).** Mide la *intermitencia*: cuántos periodos, en promedio, transcurren entre ocurrencias de demanda no nula. Una forma común de calcularlo es:

$$ADI = \frac{T}{N} \tag{1}$$

donde T es el número total de periodos observados y N es el número de periodos con demanda distinta de cero (Johnston & Boylan, 1996; Syntetos et al., 2005).

3.3.4.2 CV^2 **(Coeficiente de variación al cuadrado).** Mide la *variabilidad del tamaño de la demanda* cuando esta ocurre (normalmente calculado sobre las demandas no nulas):

$$CV^2 = \left(\frac{s}{\bar{D}}\right)^2 \tag{2}$$

Donde $\bar{D}$ es la media de los tamaños de demanda no nulos y s es su desviación estándar (Kaya et al., 2020; Syntetos et al., 2005).

3.3.5 Modelos de Series de Tiempo para Series Estables

Las series con demanda continua admiten una familia de modelos cuyo supuesto de partida es que el consumo no nulo en todos los periodos, o en la gran mayoría de ellos, aporta suficiente información como para estimar tendencia, estacionalidad y, eventualmente, la relación con variables externas. En este trabajo se evaluaron tres alternativas para las series clasificadas como estables: un modelo de referencia estacional pasivo, un modelo de suavizamiento exponencial con espacio de estados y un modelo autorregresivo con variables exógenas.

3.3.5.1 Seasonal Naive. El modelo Seasonal Naive constituye la línea base mínima para series con componente estacional. Su pronóstico para el periodo $t + h$ es simplemente el valor observado en el mismo mes del año anterior:

$$\hat{y}_{t+h} = y_{t+h-m} \tag{3}$$

donde m es el periodo estacional, igual a 12 para series mensuales (Hyndman & Athanasopoulos, 2021). Por construcción, el modelo no aprende de la dinámica reciente ni del contexto externo; su utilidad dentro de la comparación es establecer un umbral mínimo de desempeño: cualquier modelo que no supere al Seasonal Naive en el backtesting no justifica su mayor complejidad.

3.3.5.2 Modelo ETS. Los modelos ETS (Error, Trend, Seasonality) son una familia de suavizamiento exponencial formulada en espacio de estados (Hyndman et al., 2008). Cada variante combina una especificación para el error (aditivo o multiplicativo), una para la tendencia (ninguna, aditiva, amortiguada) y una para la estacionalidad (ninguna, aditiva, multiplicativa), lo que da lugar a una taxonomía de hasta 30 modelos distintos. La selección automática entre ellos se realiza minimizando el criterio de información de Akaike corregido (AICc) sobre los datos de entrenamiento.

En la variante con tendencia amortiguada, el nivel ℓ_t, la pendiente b_t y el componente estacional s_t se actualizan período a período mediante:

$$\ell_t = \alpha\,(y_t - s_{t-m}) + (1-\alpha)\,(\ell_{t-1} + \phi\, b_{t-1}) \tag{4}$$

$$b_t = \beta^*\,(\ell_t - \ell_{t-1}) + (1-\beta^*)\,\phi\, b_{t-1} \tag{5}$$

$$s_t = \gamma\,(y_t - \ell_{t-1} - \phi\, b_{t-1}) + (1-\gamma)\, s_{t-m} \tag{6}$$

donde α, β^* y γ son parámetros de suavizamiento y $\phi \in (0,1)$ controla el amortiguamiento de la tendencia. El pronóstico a h pasos es:

$$\hat{y}_{t+h} = \ell_t + \left(\sum_{i=1}^{h} \phi^i\right) b_t + s_{t+h-m} \tag{7}$$

Hyndman et al. (2008) demostraron que la versión con amortiguamiento tiende a ofrecer un buen equilibrio entre flexibilidad y riesgo de sobreajuste en horizontes medianos, lo que la convierte en un referente robusto para evaluar si la incorporación de variables exógenas agrega valor real sobre un modelo univariado bien especificado (Makridakis et al., 2020).

3.3.5.3 Modelo SARIMAX. El modelo SARIMAX extiende la familia ARIMA estacional para incorporar variables exógenas observables (Box et al., 2016). Su estructura combina un componente autorregresivo no estacional de orden p, diferenciación de orden d, medias móviles de orden q, y sus análogos estacionales P, D, Q con periodo m, más un vector de regresores externos $\mathbf{X}_t$:

$$\Phi_P(B^m)\,\phi_p(B)\,(1-B)^d(1-B^m)^D\,y_t = \boldsymbol{\beta}^\top \mathbf{X}_t + \Theta_Q(B^m)\,\theta_q(B)\,\varepsilon_t \tag{8}$$

donde B es el operador de rezago, ϕ_p, θ_q, Φ_P, Θ_Q son los polinomios autorregresivos y de medias móviles, y $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. En este trabajo, el vector $\mathbf{X}_t$ contiene la temperatura media mensual, la precipitación acumulada y la tasa representativa del mercado (TRM) promedio.

La utilidad de incluir variables exógenas depende del grado de correlación que estas sostengan con la serie objetivo. Kourentzes (2013) mostraron que la ganancia de precisión fuera de muestra es sistemática cuando dicha correlación es estadísticamente significativa, pero no justificada cuando las covariables son débilmente informativas. Esta evidencia sustenta la decisión de comparar SARIMAX contra ETS bajo el mismo esquema de backtesting: si el modelo con exógenas no supera al univariado, la complejidad adicional no está respaldada por los datos. En el contexto del presente trabajo, esa comparación también actúa como validación empírica de la pertinencia de las variables climáticas y macroeconómicas incorporadas, argumento que Barabadi et al. (2014) respaldaron al mostrar que ignorar condiciones operativas puede generar diferencias de hasta el 25% en la estimación de repuestos requeridos en equipos mineros.

La especificación de los órdenes del modelo se determinó a partir del diagnóstico estadístico previo (pruebas ADF, KPSS, descomposición STL, ACF y PACF), sin recurrir a búsqueda automática de parámetros dentro del bucle de backtesting. Esta decisión fue consciente: permitir que el orden cambie en cada paso del rolling-origin introduce una forma sutil de fuga de información de especificación que compromete la comparabilidad entre iteraciones (Tashman, 2000).

3.3.6 Modelos de series de tiempo para series intermitentes

Las series con demanda esporádica presentan dos fuentes de incertidumbre que los modelos para series continuas no están diseñados para capturar: la probabilidad de que haya demanda en un periodo dado y la magnitud de esa demanda cuando ocurre (Nikolopoulos et al., 2011). En este trabajo se evaluaron tres alternativas para las series clasificadas como intermitentes o irregulares: dos métodos clásicos desarrollados específicamente para este tipo de patrón y un modelo de dos etapas con aprendizaje automático.

3.3.6.1 Método Croston-SBA. El método de Croston (1972) fue el primero en proponer tratar la demanda intermitente descomponiendo el proceso en dos subprocesos separados: el tamaño de la demanda cuando ocurre (z_i) y el intervalo entre demandas positivas consecutivas (p_i). Cada subproceso se actualiza mediante suavizamiento exponencial simple, y el pronóstico de la demanda esperada por periodo se calcula como:

$$\hat{y} = \frac{\hat{z}}{\hat{p}} \tag{9}$$

Syntetos et al. (2005) demostraron que este estimador es sesgado hacia arriba y propusieron una corrección aproximadamente insesgada, conocida como SBA (Syntetos-Boylan Approximation):

$$\hat{y}_{\text{SBA}} = \left(1 - \frac{\alpha}{2}\right) \frac{\hat{z}}{\hat{p}} \tag{10}$$

donde α es el parámetro de suavizamiento del intervalo. La corrección reduce el error paramétrico sin incrementar la complejidad computacional (Syntetos et al., 2005). Dentro de la comparación del presente trabajo, Croston-SBA actúa como referencia analítica especializada: es el método de mayor respaldo teórico para demanda intermitente pura y su desempeño relativo indica si los modelos más complejos justifican su mayor carga computacional.

3.3.6.2 Método TSB. Teunter et al. (2011) identificaron una limitación estructural de Croston y SBA: al actualizar el intervalo entre demandas únicamente en los periodos con consumo positivo, ambos métodos no ajustan el pronóstico a la baja cuando la demanda de un ítem declina de forma sostenida o cuando un artículo comienza a caer en obsolescencia. El método TSB (Teunter-Syntetos-Babai) resuelve esta limitación reemplazando el intervalo $\hat{p}$ por una probabilidad de ocurrencia de demanda $\hat{q}_t$, que se actualiza en *todos* los periodos, con y sin consumo:

$$\hat{q}_t = \begin{cases} (1-\beta)\,\hat{q}_{t-1} + \beta & \text{si } y_t > 0 \\ (1-\beta)\,\hat{q}_{t-1} & \text{si } y_t = 0 \end{cases} \tag{11}$$

$$\hat{z}_t = \begin{cases} (1-\alpha)\,\hat{z}_{t-1} + \alpha\, y_t & \text{si } y_t > 0 \\ \hat{z}_{t-1} & \text{si } y_t = 0 \end{cases} \tag{12}$$

y el pronóstico es $\hat{y} = \hat{q}_t \cdot \hat{z}_t$. Esta formulación resulta más robusta ante cambios de nivel y no estacionariedad descendente, dos condiciones que pueden presentarse en repuestos cuyo ritmo de uso varía con la operación (Teunter et al., 2011).

3.3.6.3 Modelo Hurdle con bosque aleatorio (Hurdle-RF). Los métodos clásicos de demanda intermitente comparten una restricción: no incorporan información contextual externa que pueda modificar tanto la probabilidad de ocurrencia como la magnitud del consumo. Para supe-

rar esta limitación, la literatura reciente ha explorado enfoques de aprendizaje automático de dos etapas, denominados modelos *hurdle* (Boylan et al., 2022; Sillanpää & Liesiö, 2018).

En la primera etapa, un clasificador modela la probabilidad de que la demanda sea positiva en el periodo t:

$$P(y_t > 0 \mid \mathbf{x}_t) = f_{\text{clase}}(\mathbf{x}_t) \tag{13}$$

En la segunda etapa, un regresor modela la magnitud del consumo condicional a que sea no nulo:

$$E[y_t \mid y_t > 0,\, \mathbf{x}_t] = f_{\text{reg}}(\mathbf{x}_t) \tag{14}$$

El pronóstico final combina ambas etapas:

$$\hat{y}_t = P(y_t > 0 \mid \mathbf{x}_t) \cdot E[y_t \mid y_t > 0,\, \mathbf{x}_t] \tag{15}$$

En este trabajo, ambas etapas se implementan mediante bosques aleatorios (Random Forest). El vector de covariables $\mathbf{x}_t$ incluye rezagos del consumo, temperatura media, precipitación y TRM, lo que permite al modelo capturar relaciones entre el contexto operativo y climático y la demanda esporádica de repuestos. Sillanpää y Liesiö (2018) mostraron que este enfoque supera consistentemente a Croston-SBA cuando se dispone de variables operativas adicionales, y que el mecanismo de importancia de variables del bosque aleatorio facilita identificar cuáles covariables explican mejor la ocurrencia y la magnitud de la demanda, aspecto relevante para la comunicación de resultados a equipos de mantenimiento y logística.

3.3.7 Métricas de evaluación del pronóstico

La comparación entre modelos de pronóstico requiere indicadores que capturen distintas dimensiones del error: magnitud absoluta, sensibilidad a valores extremos, escala relativa y dirección sistemática del sesgo. En este trabajo se emplearon seis métricas complementarias para construir el ranking compuesto que determina el modelo ganador por serie.

3.3.7.1 Error absoluto medio (MAE). El MAE mide el promedio de las diferencias absolutas entre los valores observados y los pronosticados a lo largo de los H periodos de prueba:

$$\text{MAE} = \frac{1}{H} \sum_{h=1}^{H} |y_{t+h} - \hat{y}_{t+h}| \tag{16}$$

Su interpretación es directa: expresa el error típico en las mismas unidades de la serie (toneladas, galones, unidades). Por no elevar al cuadrado las desviaciones, no penaliza de forma desproporcionada los errores grandes, lo que lo convierte en una medida robusta ante valores atípicos ocasionales (Hyndman & Athanasopoulos, 2021).

3.3.7.2 Raíz del error cuadrático medio (RMSE).

$$\text{RMSE} = \sqrt{\frac{1}{H}\sum_{h=1}^{H}\left(y_{t+h} - \hat{y}_{t+h}\right)^2} \tag{17}$$

Al elevar al cuadrado cada error antes de promediar, el RMSE penaliza los desvíos grandes más que los pequeños, por lo que resulta especialmente sensible a pronósticos muy alejados del valor real. En series industriales con consumos ocasionalmente extremos, el RMSE complementa al MAE al revelar si los errores grandes son sistemáticos o aislados (Hyndman & Athanasopoulos, 2021).

3.3.7.3 Error porcentual absoluto medio simétrico (sMAPE). El sMAPE normaliza el error respecto al promedio del valor observado y el pronosticado, lo que lo hace simétrico ante sobreestimaciones y subestimaciones:

$$\text{sMAPE} = \frac{1}{H}\sum_{h=1}^{H}\frac{2|y_{t+h} - \hat{y}_{t+h}|}{|y_{t+h}| + |\hat{y}_{t+h}|} \times 100 \tag{18}$$

A diferencia del MAPE clásico, el sMAPE no diverge cuando el valor observado se acerca a cero, lo que lo hace aplicable en series con demanda esporádica, aunque valores cercanos a 200% indican que al menos una de las dos cantidades es cercana a cero (Makridakis et al., 2020).

3.3.7.4 Error porcentual absoluto medio ponderado (WAPE). El WAPE (también denominado MAD/Mean ratio en algunas referencias) agrega los errores absolutos y los normaliza por el total observado, en lugar de calcular un porcentaje por periodo y luego promediarlo:

$$\text{WAPE} = \frac{\sum_{h=1}^{H}|y_{t+h} - \hat{y}_{t+h}|}{\sum_{h=1}^{H}|y_{t+h}|} \times 100 \tag{19}$$

Esta formulación le otorga mayor peso a los periodos con mayor consumo y es menos sensible a los periodos con demanda nula o muy baja que otras métricas porcentuales (Kolassa & Schütz, 2011). En series intermitentes, donde la mayoría de los periodos tienen demanda nula, valores de WAPE superiores al 80% son esperables y no implican necesariamente un modelo deficiente: reflejan la dificultad intrínseca de pronosticar series con alta proporción de ceros. Por esta razón, el WAPE se interpretó siempre en conjunto con la precisión de actividad.

3.3.7.5 Error absoluto medio escalado (MASE). El MASE escala el MAE del modelo respecto al MAE del pronóstico naive estacional en la muestra de entrenamiento, lo que permite comparaciones entre series con unidades o escalas distintas (Hyndman & Koehler, 2006):

$$\text{MASE} = \frac{\text{MAE}}{\frac{1}{n-m}\sum_{t=m+1}^{n}|y_t - y_{t-m}|} \tag{20}$$

donde n es la longitud de la muestra de entrenamiento y m es el periodo estacional (12 para series mensuales). Un valor MASE < 1 indica que el modelo supera al naive estacional; un valor > 1 señala que el naive sería preferible. Su ausencia de supuestos distribucionales y su estabilidad ante series con ceros lo hacen adecuado para comparar modelos sobre familias de demanda heterogéneas (Hyndman & Koehler, 2006).

3.3.7.6 Sesgo relativo (Bias). El sesgo mide la dirección sistemática del error: si el modelo tiende a sobreestimar o subestimar la demanda de forma consistente a lo largo del horizonte de prueba.

$$\text{Bias} = \frac{\sum_{h=1}^{H}(\hat{y}_{t+h} - y_{t+h})}{\sum_{h=1}^{H}|y_{t+h}|} \times 100 \tag{21}$$

Un sesgo positivo indica sobreestimación sistemática (el modelo predice más de lo que se consume), y un sesgo negativo indica subestimación. Ambas direcciones tienen consecuencias operativas distintas en inventarios: la sobreestimación genera exceso de stock, mientras que la subestimación aumenta el riesgo de quiebre (Syntetos et al., 2005).

3.3.7.7 Precisión de actividad (*Activity accuracy*). La precisión de actividad mide la proporción de periodos en los que el modelo clasifica correctamente si habrá o no demanda, con independencia de la magnitud pronosticada:

$$\text{Activity accuracy} = \frac{1}{H}\sum_{h=1}^{H}\mathbf{1}[(y_{t+h} > 0) = (\hat{y}_{t+h} > 0)] \tag{22}$$

Esta métrica es especialmente informativa para series intermitentes, donde anticipar correctamente si un periodo tendrá consumo o no es tan relevante como la magnitud del pronóstico. Un modelo con WAPE alto pero con precisión de actividad elevada identifica correctamente los meses de demanda; sus errores de magnitud tienen menor impacto sobre las decisiones de reabastecimiento que un sesgo sistemático en la detección de ocurrencia (Kourentzes, 2013).

3.3.8 Diagnóstico estadístico de residuos

Una vez ajustado un modelo sobre la muestra de entrenamiento completa, la revisión de sus residuos permite verificar si quedan patrones sistemáticos no capturados. En este trabajo se aplicaron tres pruebas sobre los residuos de los modelos ganadores para las series estables, las únicas familias para las que los residuos de un proceso continuo son estadísticamente interpretables

con estas herramientas.

3.3.8.1 Prueba de Ljung-Box. La prueba de Ljung-Box evalúa la hipótesis nula de que los primeros k rezagos de la función de autocorrelación de los residuos son conjuntamente iguales a cero, es decir, que no hay autocorrelación serial remanente (Ljung & Box, 1978). El estadístico es:

$$Q = n(n+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{n-j} \tag{23}$$

donde n es el número de observaciones, $\hat{\rho}_j$ es la autocorrelación muestral en el rezago j y k es el número de rezagos evaluados. Bajo la hipótesis nula, Q sigue asintóticamente una distribución χ^2 con k grados de libertad. Un p-valor superior a $0{,}05$ indica que los residuos no presentan estructura temporal identificable, lo que confirma que el modelo capturó adecuadamente la dinámica de la serie.

3.3.8.2 Prueba de Jarque-Bera. La prueba de Jarque-Bera contrasta si los residuos provienen de una distribución normal, evaluando conjuntamente su asimetría y curtosis (Jarque & Bera, 1980):

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{24}$$

donde S es la asimetría muestral y K la curtosis muestral de los residuos. Bajo normalidad, JB sigue asintóticamente una distribución χ^2 con 2 grados de libertad. En el contexto de este trabajo, el rechazo de la hipótesis nula no invalida los pronósticos puntuales, cuya calidad descansa en el backtesting fuera de muestra. Su implicación práctica se limita a la construcción de intervalos de predicción paramétricos, aspecto que no tiene consecuencias operativas aquí porque el stock de seguridad y el punto de reorden se derivaron mediante simulación Monte Carlo, sin asumir normalidad en la distribución del error (Strijbosch et al., 2000).

3.3.8.3 Prueba ARCH-LM. La prueba ARCH-LM (Autoregressive Conditional Heteroskedasticity, Lagrange Multiplier) verifica si la varianza de los residuos es constante en el tiempo o si, por el contrario, presenta agrupamiento: periodos de alta volatilidad seguidos de periodos de baja volatilidad (Engle, 1982). El procedimiento consiste en regresar los residuos al cuadrado sobre sus propios rezagos y evaluar la significancia conjunta de los coeficientes mediante un estadístico LM:

$$LM = n \cdot R^2 \tag{25}$$

donde R^2 proviene de la regresión auxiliar de $\hat{\varepsilon}_t^2$ sobre $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-p}^2$. Bajo la hipótesis nula

de homocedasticidad condicional, $LM \sim \chi^2(p)$. La ausencia de efectos ARCH en los residuos confirma que la estructura de volatilidad de la serie fue correctamente recogida por el modelo, lo que refuerza la validez de los intervalos de predicción generados.

3.3.9 Diagnóstico de covariables

Antes de incorporar las variables exógenas en los modelos, se verificaron dos condiciones: que no existiera redundancia lineal entre ellas (lo que distorsionaría la estimación de coeficientes) y que presentaran asociación estadística con las series de demanda. Para esto se emplearon el Factor de Inflación de la Varianza y la correlación de Spearman con rezagos cortos.

3.3.9.1 Factor de inflación de la varianza (VIF). El VIF cuantifica en qué medida la varianza del coeficiente estimado de una variable explicativa se infla debido a su correlación lineal con las demás variables del modelo (Montgomery et al., 2021). Para una variable X_j, el VIF se calcula como:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \tag{26}$$

donde R_j^2 es el coeficiente de determinación de la regresión de X_j sobre las restantes variables explicativas. Un $\text{VIF}_j = 1$ indica ausencia total de colinealidad; valores superiores a 5 suelen considerarse problemáticos y valores superiores a 10 señalan multicolinealidad severa (Hair et al., 2019). En este trabajo, los VIF entre las tres variables exógenas (temperatura, precipitación y TRM) se mantuvieron entre $1{,}02$ y $1{,}04$, descartando redundancia lineal y confirmando que las tres pueden incorporarse simultáneamente en los modelos sin distorsión en la estimación de sus efectos.

3.3.9.2 Correlación de Spearman con rezagos. La correlación de Spearman es una medida no paramétrica de asociación monótona entre dos variables. A diferencia del coeficiente de Pearson, no supone linealidad ni distribución normal, lo que la hace más adecuada para series de demanda industrial que pueden presentar distribuciones asimétricas o con valores extremos (Spearman, 1904):

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{27}$$

donde d_i es la diferencia entre los rangos asignados a cada observación en las dos variables. En este trabajo se calculó r_s entre cada serie de consumo y cada variable exógena con rezagos de 0 a 3 meses, con el objetivo de identificar si el efecto de las covariables sobre la demanda es contemporáneo o diferido. Este análisis informó la decisión de incorporar las variables exógenas en los modelos SARIMAX y en el componente de ocurrencia del Hurdle-RF, sin pretender establecer causalidad sino caracterizar la asociación estadística observable en los datos disponibles.

3.3.10 Validación fuera de muestra mediante rolling-origin (ventana rodante)

Evaluar un modelo de series de tiempo con los mismos datos usados para entrenarlo produce estimaciones de error optimistas que no reflejan el desempeño real en producción. La alternativa estándar para contextos de pronóstico operacional es el *rolling-origin backtesting*, un esquema de validación fuera de muestra que preserva la causalidad temporal al nunca utilizar información futura durante el entrenamiento (Tashman, 2000).

El procedimiento funciona así: dado un histórico de T periodos, se reservan los últimos H como ventana de prueba. En cada iteración, el origen de pronóstico τ avanza un periodo, el conjunto de entrenamiento se expande con la información disponible hasta τ, y se genera una predicción a un paso adelante:

$$\hat{y}_{\tau+1|\tau} = f(y_{1:\tau},\ \mathbf{X}_{1:\tau},\ \mathbf{X}_{\tau+1}) \tag{28}$$

Este proceso se repite secuencialmente para $\tau = T - H,\ T - H + 1,\ \ldots,\ T - 1$, acumulando H errores de pronóstico fuera de muestra sobre los cuales se calculan las métricas de desempeño. En este trabajo se utilizaron los últimos 12 meses como ventana de prueba, lo que produce 12 errores de un paso para cada SKU y modelo evaluado.

La variante expansiva del rolling-origin, adoptada aquí, incorpora todos los datos históricos disponibles en cada reentrenamiento en lugar de mantener una ventana deslizante de tamaño fijo. Esta elección es apropiada cuando el histórico es relativamente corto, como ocurre con series mensuales de 7 años, porque descarta la menor cantidad posible de información en cada iteración (Makridakis et al., 2020). Tashman (2000) identificó el rolling-origin como el esquema de evaluación más adecuado para series de tiempo en contextos de pronóstico operacional, precisamente porque evita validaciones aleatorias que mezclarían información futura con el entrenamiento y producirían resultados no replicables en la práctica.

3.3.11 Políticas de Reabastecimiento

3.3.11.1 Punto de Reorden (ROP). El punto de reorden (ROP, por sus siglas en inglés *Reorder Point*) es el nivel de inventario que, al alcanzarse, dispara una orden de compra. Bajo un sistema de revisión continua, su función es garantizar que el inventario disponible cubra la demanda esperada durante el tiempo que transcurre entre la emisión del pedido y la recepción del material, es decir, durante el lead time L (Rego & Mesquita, 2011). Formalmente:

$$\text{ROP} = \mu_L + SS \tag{29}$$

donde μ_L es la demanda esperada acumulada durante el lead time y SS es el stock de seguridad que absorbe la variabilidad de esa demanda. Cuando el lead time está medido en los mismos periodos que la frecuencia de la serie (meses, en este caso), la demanda esperada durante el lead time se

obtiene sumando los pronósticos del modelo para los L periodos siguientes:

$$\mu_L = \sum_{h=1}^{L} \hat{y}_{t+h} \tag{30}$$

En términos operativos, el ROP convierte el pronóstico estadístico en una regla de acción concreta: cuando el nivel de inventario cae por debajo de este umbral, se emite la orden. Cuando el lead time es inferior a un mes (por ejemplo, 20 días equivalen a 0,66 meses), la demanda durante el lead time se calcula como la fracción correspondiente del pronóstico del primer mes simulado. Para lead times mayores a un mes con componente fraccional, se suman los meses completos del horizonte y se agrega la fracción correspondiente del siguiente mes. Su correcta calibración depende tanto de la precisión del pronóstico de demanda como de una estimación adecuada del stock de seguridad, aspecto que se aborda en la sección siguiente.

3.3.11.2 Stock de Seguridad y simulación Monte Carlo. El stock de seguridad (SS) es el colchón de inventario que protege contra la incertidumbre de la demanda durante el lead time. En el enfoque clásico, se calcula como un múltiplo del factor de servicio z sobre la desviación estándar de la demanda en ese horizonte:

$$SS_{\text{clásico}} = z_\alpha \cdot \sigma_L \tag{31}$$

donde z_α es el cuantil de la distribución normal estándar asociado al nivel de servicio objetivo α y σ_L es la desviación estándar de la demanda durante el lead time (Rego & Mesquita, 2011). Este cálculo asume implícitamente que la distribución de la demanda en ese horizonte es normal, supuesto que puede sostenerse para series con consumo continuo y simétrico, pero que resulta cuestionable para repuestos con demanda intermitente o asimétrica: en presencia de muchos ceros y colas pesadas, la fórmula analítica tiende a subestimar el inventario necesario para alcanzar el nivel de servicio deseado (Dolgui & Pashkevich, 2008).

La simulación Monte Carlo ofrece una alternativa que no impone ningún supuesto sobre la forma de la distribución. A partir de la distribución empírica de los errores de pronóstico obtenidos en el backtesting, se generan miles de trayectorias posibles de demanda durante el lead time y se calcula directamente la distribución simulada de la demanda acumulada D_L. El stock de seguridad y el punto de reorden se obtienen entonces como cuantiles de esa distribución:

$$SS = Q_\alpha(D_L) - \mathbb{E}[D_L] \tag{32}$$

$$ROP = Q_\alpha(D_L) \tag{33}$$

donde $Q_\alpha(D_L)$ es el cuantil de la distribución simulada de la demanda en lead time asociado al nivel de servicio α.

En este trabajo se adoptó el enfoque de simulación Monte Carlo para todos los SKUs analizados, tanto estables como intermitentes. Esta decisión es especialmente pertinente para las series ICP, donde la asimetría de la distribución de consumo es la norma, pero también para las series ISM, donde la variabilidad residual del backtesting no garantiza normalidad. El nivel de servicio objetivo se fijó en 95% para todos los artículos, en línea con la criticidad operativa de los repuestos seleccionados.

4. Metodología

4.1 Enfoque de investigación

La investigación se desarrolla bajo un enfoque cuantitativo, ya que analiza datos numéricos del consumo histórico de insumos MRO, variables exógenas y métricas estadísticas de desempeño predictivo. El propósito no es explorar percepciones ni construir teoría, sino estimar el consumo futuro de artículos estratégicos y convertir esos resultados en parámetros operativos de inventario; por eso, el trabajo se centra en la medición de patrones de demanda, la comparación rigurosa de modelos y la generación de reglas cuantificables para apoyar decisiones de reabastecimiento.

4.2 Diseño y tipo de investigación

El diseño es no experimental, porque las variables se analizan tal como fueron registradas, sin intervención del investigador sobre el proceso operativo. Los consumos históricos, la temperatura, la precipitación y la TRM se toman de sus fuentes originales y se examinan en su estado natural.

Por su alcance, la investigación combina cuatro características. Es descriptiva porque caracteriza el comportamiento de cada SKU mediante frecuencia de consumo, porcentaje de ceros, indicadores ADI y CV^2, estacionalidad y tendencia. Es explicativa porque examina la relación entre la demanda y variables externas como el clima y la tasa de cambio. Es documental porque trabaja con registros históricos consolidados. Y es aplicada porque los pronósticos obtenidos se traducen directamente en parámetros de gestión: demanda esperada durante el lead time, stock de seguridad y punto de reorden.

4.3 Población y muestra

La población está conformada por los insumos, repuestos y materiales MRO utilizados en la operación minera y de producción de coque de C.I. Milpa S.A. y sus compañías aliadas. Estos artículos constituyen unidades de análisis cuantitativas, ya que su comportamiento puede observarse mediante series mensuales de consumo, variabilidad, frecuencia de demanda y relación con variables exógenas.

La muestra corresponde a seis insumos estratégicos. La selección no fue aleatoria, sino intencional, orientada por criterios de criticidad, relevancia operativa y riesgo de suministro. Concretamente, el análisis se limitó a las categorías ISM (Insumos de Planta Lavadora) e ICP (Importaciones para Consumo Propio), identificadas por la Gerencia de Producción como grupos de artículos estratégicos dentro de la matriz de Kraljic, por su alto impacto sobre la continuidad productiva y su mayor exposición al riesgo de abastecimiento (Kraljic, 1983).

Dentro de la categoría ISM se priorizaron artículos con comportamiento estable, que corresponden a insumos de consumo relativamente continuo en la operación de planta; de este grupo se seleccionaron los tres SKUs con mayor costo anual acumulado. En la categoría ICP se buscó trabajar con artículos de comportamiento intermitente, más representativos de la lógica de reposición de repuestos estratégicos; para su inclusión se exigió una antigüedad mínima de 48 meses dentro de la ventana de análisis y una proporción de meses con consumo no nulo de al menos 30%, y entre los artículos que cumplieron estas condiciones se eligieron igualmente los tres de mayor costo anual acumulado. Es pertinente aclarar que, aunque se intentó apuntar a insumos intermitentes, se evidencia la presencia de un SKU irregular (ICP10095), cuyas implicaciones se abordan en la Sección 5.1. La Tabla 2 presenta los seis artículos resultantes.

Tabla 2
SKUs seleccionados para el modelado de demanda

SKU	Nombre	Unidad	Categoría	ADI	CV^2
ISM10001	Magnetita	Tn	ISM	1,000	0,319
ISM10003	Espumante S3124	Gal	ISM	1,000	0,115
ISM10006	ACPM	Gal	ISM	1,000	0,165
ICP10107	Zuper Block	Ud	ICP	2,542	0,262
ICP10095	Malla Ine Flow	Ud	ICP	3,250	0,635
ICP10360	Bobina MU800	Ud	ICP	3,267	0,450

4.4 Técnicas e instrumentos de recolección de datos

La recolección de datos se realizó mediante un análisis documental de registros internos y fuentes secundarias externas. El instrumento principal fue una base de datos consolidada con frecuencia mensual, compuesta por el histórico de consumo de los seis SKUs seleccionados y tres variables exógenas: temperatura media, precipitación acumulada y TRM.

Para las variables climáticas, se descargaron datos diarios a través de la API de Open-Meteo para los municipios de Samacá, Socha y Guachetá, zonas asociadas a la operación analizada. A partir de esta información se calculó, para cada ubicación, la precipitación mensual acumulada y la temperatura mensual promedio; luego se promediaron los valores de las tres zonas para construir un indicador climático regional representativo. Para la TRM se extrajo la serie diaria desde Yahoo

Finance mediante el *ticker* COP=X, y se calculó el promedio mensual para reducir el ruido de corto plazo y hacerla consistente con la periodicidad mensual de la demanda.

La base consolidada resultante cubre el periodo enero de 2019 a diciembre de 2025, con 84 observaciones mensuales continuas, y constituyó el insumo principal para las etapas de exploración, modelado y evaluación de políticas de inventario.

La justificación de las tres variables exógenas descansa en argumentos operativos concretos: la TRM incide directamente en el costo de importación de los artículos ICP, cuyo precio en pesos depende del tipo de cambio vigente al momento de la compra; la temperatura afecta el desgaste de componentes, dado que condiciones de mayor calor aceleran la degradación de aceites y lubricantes; y la precipitación condiciona el ritmo operativo en minería a cielo abierto, con efectos directos sobre el consumo de insumos de planta. Adicionalmente, los lead times reales por SKU fueron informados directamente por la jefatura de almacenes de Milpa: 20 días para ISM10001 e ISM10003 y 10 días para ISM10006 (proveedores nacionales con reposición rápida), y 90 días para las tres series ICP por tratarse de importaciones. Estos valores se utilizan tanto en el cálculo de los parámetros de inventario (Sección 5.4) como en la simulación de la política propuesta (Sección 5.5).

4.5 Procedimiento y análisis de datos

El procedimiento se organizó en cinco fases secuenciales.

Fase 1: Auditoría y clasificación. El análisis comenzó con la carga y revisión del conjunto de datos: se verificó la continuidad del calendario mensual, la existencia de las variables obligatorias y la vida activa de cada SKU, con el fin de diferenciar los valores faltantes previos al inicio real de la serie de los ceros observados durante su operación. Los ceros previos a la vida activa se trataron como ausencia estructural, no como demanda nula, para evitar inflar artificialmente la intermitencia de los artículos. A continuación se calcularon los indicadores ADI y CV^2 para clasificar los patrones de demanda según la regla de Syntetos-Boylan (Syntetos et al., 2005), separando las series en familias: estable, intermitente e irregular.

Fase 2: Análisis exploratorio y diagnóstico. Sobre cada serie se realizó un diagnóstico estadístico que incluyó la revisión visual de trayectorias históricas, la descomposición STL para separar tendencia, componente estacional y residuo, pruebas de raíz unitaria ADF y KPSS para evaluar estacionariedad, análisis de autocorrelación y autocorrelación parcial (ACF y PACF), correlaciones de Spearman entre consumos y variables exógenas con rezagos cortos, y cálculo de VIF para detectar colinealidad entre las exógenas. Esta fase no buscó seleccionar modelos todavía, sino documentar las propiedades estadísticas de cada serie para informar la especificación posterior.

Fase 3: Backtesting rolling-origin y selección de modelos. Para cada serie se implementó un esquema de validación fuera de muestra mediante *rolling-origin* de un paso adelante, entrenando el modelo con información disponible hasta el periodo τ y pronosticando el mes $\tau+1$ sin incorporar

datos futuros. Este procedimiento se repitió secuencialmente sobre los últimos *H* periodos del histórico, preservando la causalidad temporal. Para las series estables se compararon Seasonal Naive, ETS y SARIMAX con variables exógenas; para las series intermitentes se compararon Croston-SBA, TSB y Hurdle-RF. La especificación del SARIMAX se definió con base en el diagnóstico previo, sin reoptimizar el orden en cada partición, para evitar filtraciones sutiles de información futura en la estructura del modelo y para evitar complejidad computacional extra al tener que evaluar para cada periodo de la prueba. La selección del mejor modelo por serie se apoyó en un ranking compuesto por WAPE, RMSE y MAE, priorizando el modelo con menor posición promedio en las tres métricas (Kourentzes et al., 2020).

Fase 4: Pronóstico final y política de inventario. El modelo seleccionado para cada SKU se reentrenó con toda la vida activa disponible y se utilizó para generar un pronóstico a 12 meses. Las variables exógenas futuras se proyectaron mediante un escenario base parsimonioso: temperatura y precipitación a partir del promedio histórico del mismo mes calendario, y TRM a partir del promedio de los últimos meses observados. A partir de la distribución simulada de la demanda en lead time mediante simulación Monte Carlo (generada con 900 trayectorias para los modelos estables, 800 para los modelos esporádicos de tasa (Croston-SBA y TSB) y 450 para Hurdle-RF, con semilla aleatoria fija (RANDOM_STATE = 42) para garantizar reproducibilidad), se calcularon el stock de seguridad y el punto de reorden como el cuantil asociado al nivel de servicio objetivo menos y equivalente a la demanda esperada, respectivamente.

Fase 5: Backtesting de política y valorización financiera. Para cerrar el ciclo entre pronóstico y decisión operativa, se diseñó una simulación de inventario que contrasta la gestión real de Milpa con el comportamiento hipotético del sistema propuesto sobre el periodo agosto de 2025 a marzo de 2026. Con el propósito de evitar solapamiento entre entrenamiento y evaluación, los parámetros de política se recalcularon con información disponible hasta julio de 2025. La simulación operó bajo un esquema de revisión continua con posición de inventario: cada día del horizonte evaluado, la posición disponible más pedidos en tránsito se comparó con el punto de reorden, y cuando cayó por debajo de ese umbral se emitió una orden cuya cantidad cubre la demanda esperada durante el lead time más el stock de seguridad. Los indicadores de nivel de servicio y stock promedio diario se compararon entre ambos escenarios para cada SKU, y la diferencia en capital inmovilizado se valorizó en pesos colombianos utilizando los costos unitarios reportados por la empresa, con el fin de cuantificar el impacto financiero potencial del sistema propuesto.

5. Resultados

5.1 Clasificación de la demanda y auditoría de la base de datos

La auditoría inicial confirmó que el dataset cuenta con 84 registros mensuales continuos entre enero de 2019 y diciembre de 2025, sin periodos faltantes en el calendario. Las seis series objetivo cubren dos familias: tres series ISM de consumo estable y tres series ICP de consumo esporádico (intermitente e irregular). La Tabla 3 resume la estructura general de la base de datos consolidada.

Tabla 3
Resumen estructural del dataset

Filas	Fecha mín.	Fecha máx.	Series objetivo	ISM (estable)	ICP (esporádico)	Exógenas	Horizonte (meses)	Periodos faltantes
84	2019-01	2025-12	6	3	3	3	12	0

La clasificación mediante los indicadores ADI y CV² confirmó que las tres series ISM corresponden a la categoría estable (ADI $= 1{,}000$ en los tres casos, sin ningún mes con consumo cero durante la vida activa), mientras que las series ICP presentaron comportamientos diferenciados: ICP10107 e ICP10360 quedaron clasificadas como intermitentes y ICP10095 como irregular, con porcentajes de meses sin consumo que oscilaron entre 60,7% y 69,4%, valores contenidos en la Tabla 4. La clasificación de ICP10095 como irregular (no intermitente, como se creía incialmente) es metodológicamente relevante: el criterio ADI/CV² anticipa que esta serie será la más difícil de pronosticar dentro del grupo esporádico, hipótesis que el backtesting confirmó posteriormente cuando Croston-SBA superó a Hurdle-RF, contrario a lo observado en las otras dos series ICP.

Tabla 4
Auditoría de demanda por SKU en vida activa

Serie	Inicio	Fin	n obs.	% ceros	ADI	CV²	Categoría	Media	σ	Máx.
ISM10001_smooth	2019-01	2025-12	84	0,0%	1,000	0,319	Estable	62,23	35,15	177,22
ISM10003_smooth	2019-01	2025-12	84	0,0%	1,000	0,115	Estable	383,68	130,08	741,52
ISM10006_smooth	2019-07	2025-12	78	0,0%	1,000	0,199	Estable	2.538,81	1.030,11	4.706,28
ICP10107_intermittent	2020-12	2025-12	61	60,7%	2,542	0,262	Intermitente	47,30	70,41	250,00
ICP10095_intermittent	2021-09	2025-12	52	69,2%	3,250	0,635	Irregular	1,13	2,34	12,00
ICP10360_intermittent	2021-12	2025-12	49	69,4%	3,267	0,450	Intermitente	0,63	1,22	6,00

Nota. La coherencia entre el sufijo declarado en el nombre de cada serie y la categoría empírica obtenida mediante ADI/CV^2 fue total en los seis casos. Las estadísticas de media, desviación estándar y máximo excluyen los periodos con consumo igual a cero.

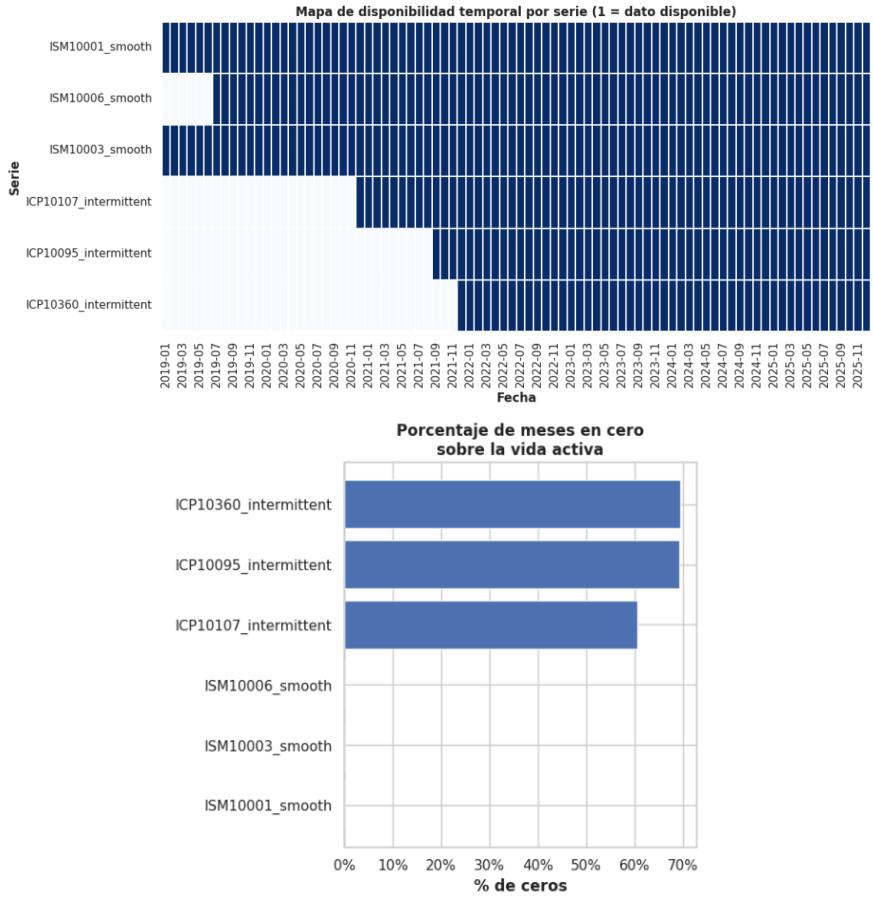
Un aspecto relevante de esta etapa fue verificar que el sufijo declarado en el nombre de cada

serie coincidiera con su categoría empírica. En todos los casos la coherencia fue total, lo que indica que la etiqueta asignada originalmente por el sistema de la empresa refleja correctamente el patrón estadístico de cada artículo.

La Figura 3 permite apreciar con claridad que las series estables concentran toda su vida activa sin interrupciones, mientras que las series intermitentes tienen vidas activas más cortas y recientes, con largos tramos de consumo nulo. Esta heterogeneidad estructural justifica la separación en familias de modelado: aplicar un único modelo sobre el conjunto completo habría producido resultados sistemáticamente sesgados para alguno de los dos grupos.

Figura 3

Disponibilidad temporal y presencia de ceros en la vida activa



Nota. El mapa de disponibilidad (parte superior) muestra los periodos mensuales con datos observados para cada serie, donde el valor 1 (azul oscuro) indica presencia de información disponible. El valor 0 (azul claro) indica que el SKU no existía aún en el entorno operativo de la empresa, no representa que la demanda haya sido nula. El gráfico de barras (parte inferior) presenta el porcentaje de meses con consumo igual a cero dentro de la vida activa de cada SKU.

5.2 Diagnóstico exploratorio y econométrico

El análisis exploratorio se realizó sobre cada serie de forma individual, combinando revisión visual, descomposición STL, pruebas de raíz unitaria y análisis de correlación.

La Figura 4 muestra dos comportamientos claramente distintos. Las series ISM presentan un consumo continuo a lo largo del periodo, con variaciones de nivel y picos puntuales, pero sin meses en cero. ISM10006 registra los mayores valores absolutos y algunos eventos extremos; en particular, el registro de diciembre de 2021 (7.262,85 galones) fue identificado como un valor atípico aislado, cuya magnitud duplica con creces el máximo del resto de la serie y no tiene correspondencia en los meses adyacentes. Dado que su inclusión distorsionaría la estimación de parámetros del modelo, ese registro fue reemplazado por la mediana de los cuatro meses vecinos (±2 meses), resultando en un valor de sustitución de 2.514,96 galones, coherente con el nivel de consumo circundante. Es importante aclarar que la Tabla 4 en la Sección anterior presenta los valores posterior a esta modificación. ISM10001 e ISM10003 muestran oscilaciones más moderadas, con una tendencia creciente hacia el tramo final. Las series ICP, en cambio, alternan largos periodos sin consumo con apariciones puntuales de demanda positiva, confirmando visualmente la clasificación obtenida en la fase anterior.

5.2.1 Estacionariedad y estacionalidad

Las pruebas ADF y KPSS y la descomposición STL revelaron que la estacionalidad anual no es un componente fuerte en ninguna de las seis series: los valores de fuerza estacional y las autocorrelaciones en el rezago 12 fueron bajos en todos los casos. Respecto a la estacionariedad en nivel, dos de las tres series ISM (ISM10006 e ISM10003) requirieron diferenciación de primer orden según la evidencia conjunta de ambas pruebas, mientras que ISM10001 resultó estacionaria. Las series ICP, al tener muchos ceros, moderan la interpretación clásica de estas pruebas. La Tabla 5 sintetiza los resultados del diagnóstico de estacionariedad y estacionalidad por serie.

Las descomposiciones STL individuales mostraron patrones consistentes con la clasificación general. ISM10001 evidenció un ciclo de crecimiento, caída y recuperación sin una estacionalidad mensual claramente definida; ISM10006 mostró una tendencia creciente en varios tramos con oscilaciones en el componente estacional que no se repiten de forma regular; ISM10003 presentó un crecimiento gradual con mayor aceleración hacia el final del periodo. En las series ICP, la tendencia capturada por el STL sugiere una mayor aparición reciente de consumo, especialmente en ICP10107 e ICP10360, aunque sin consolidarse como demanda continua, como muestra la Figura 5.

Figura 4

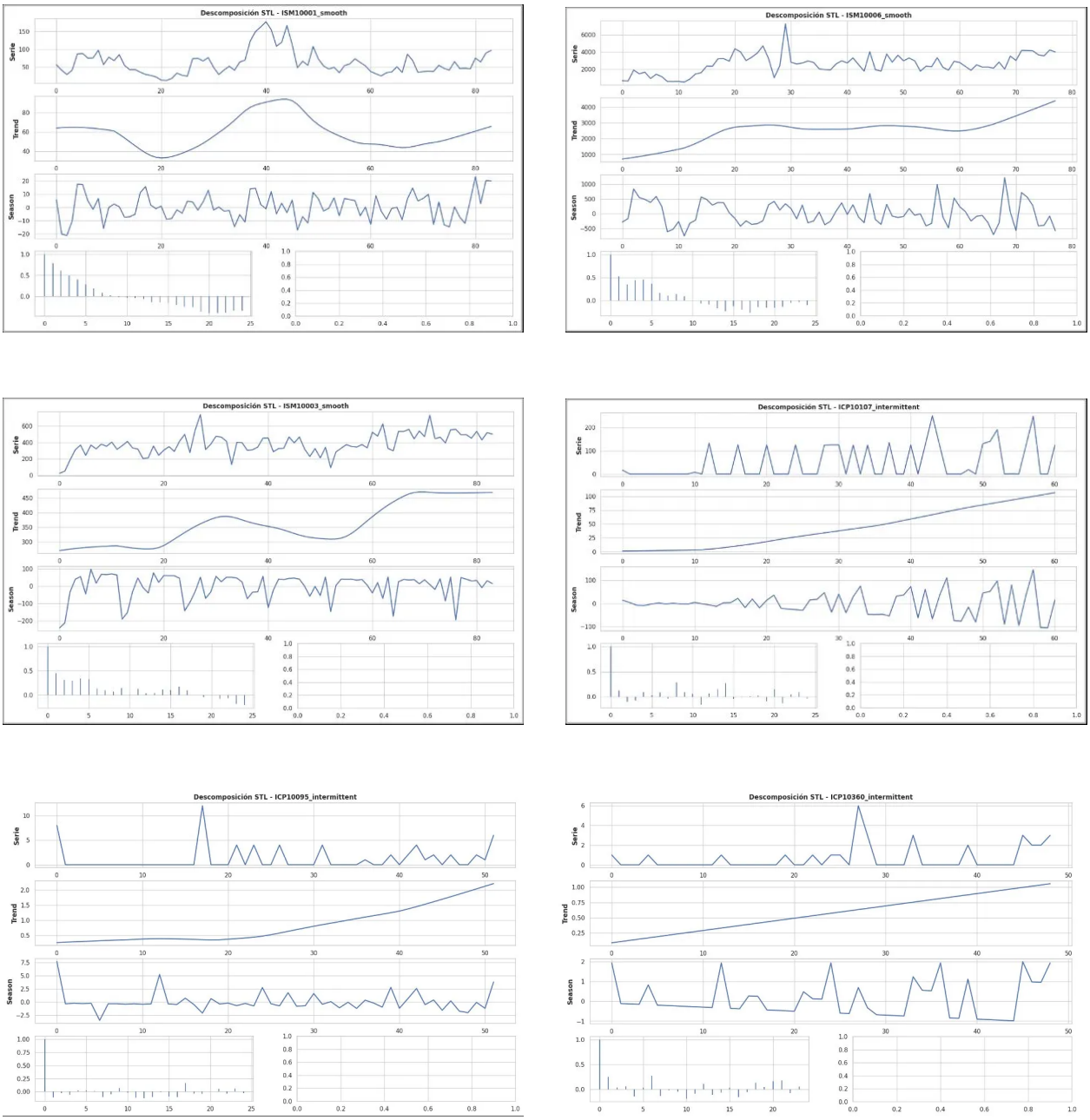
Comportamiento histórico mensual por SKU



Nota. Cada panel presenta la evolución mensual del consumo para una serie específica durante su vida activa. La línea punteada representa la media histórica de consumo de cada SKU y permite comparar los valores observados frente a su nivel promedio.

Figura 5

Descomposición STL por serie



Nota. Cada panel separa el consumo observado en tres componentes: tendencia de mediano plazo, variación estacional y residuo. Las series estables (ISM10001, ISM10006, ISM10003) presentan mayor continuidad temporal y componentes de tendencia más pronunciados; las series intermitentes (ICP10107, ICP10095, ICP10360) muestran alta concentración de ceros y picos aislados. El subpanel inferior de cada serie presenta la función de autocorrelación (ACF) de los residuos del componente STL.

Tabla 5
Diagnóstico preliminar de estacionariedad y estacionalidad

Serie	Categoría	n obs.	Fuerza estacional	ADF p-valor	KPSS p-valor	ACF lag 12	d sugerido	D sugerido
ISM10001_smooth	Estable	84	0,242	0,039	0,100	−0,061	0	0
ISM10006_smooth	Estable	78	0,000	0,003	0,0218	−0,170	1	0
ISM10003_smooth	Estable	84	0,000	0,288	0,010	0,032	1	0
ICP10107_intermittent	Intermitente	61	0,000	0,000	0,016	0,069	1	0
ICP10095_intermittent	Irregular	52	0,000	0,000	0,100	−0,124	0	0
ICP10360_intermittent	Intermitente	49	0,228	0,000	0,096	0,104	0	0

Nota. La fuerza estacional se calculó como $F_S = \max\left(0,\, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)}\right)$. Para la prueba ADF la hipótesis nula es presencia de raíz unitaria; para KPSS la hipótesis nula es estacionariedad. El valor p del KPSS truncado en 0,100 indica que el estadístico cae fuera del rango tabular por arriba o por abajo. d y D corresponden a los órdenes de diferenciación sugeridos en nivel y estacional respectivamente.

5.2.2 Diagnóstico de residuos del modelo ganador (series estables)

Como complemento al backtesting rolling-origin, se realizó un diagnóstico de los residuos del modelo ganador sobre cada una de las tres series estables, entrenado con la vida activa completa. Este análisis aplica exclusivamente a las series estables porque sus modelos (SARIMAX_EXOG y ETS) generan residuos de un proceso continuo cuya estructura temporal puede evaluarse formalmente. En las series intermitentes (ICP), la alta proporción de ceros en la serie observada hace que los residuos no sigan un proceso estacionario interpretable con las mismas pruebas, por lo que su validación descansa en las métricas de backtesting y en la precisión de actividad reportadas en la sección 5.3.

Las tres pruebas aplicadas fueron: Ljung-Box para detectar autocorrelación residual, Jarque-Bera para evaluar normalidad de los residuos, y ARCH-LM para verificar la presencia de heterocedasticidad condicional. Los resultados se presentan en la Tabla 6.

Las tres series muestran residuos sin autocorrelación serial (Ljung-Box $p > 0{,}05$ en todos los casos) y sin heterocedasticidad condicional (ARCH-LM $p > 0{,}05$ en todos los casos), lo que confirma que los modelos capturan adecuadamente la dinámica temporal de cada serie y que no quedan patrones sistemáticos explotables en los errores. ISM10003 pasa las tres pruebas sin observaciones, con residuos compatibles con normalidad ($p = 0{,}537$).

Para ISM10001 e ISM10006, la prueba Jarque-Bera rechaza la normalidad de los residuos ($p < 0{,}001$ y $p = 0{,}001$, respectivamente). La no normalidad de los residuos no afecta la validez de los pronósticos puntuales, cuya evaluación descansa en el backtesting fuera de muestra. Su efecto potencial se limita a la construcción de intervalos de predicción paramétricos, punto que

Tabla 6
Diagnóstico de residuos del modelo ganador por SKU (series estables)

Serie	Modelo	n	Ljung-Box		Jarque-Bera	
			p-valor	Resultado	p-valor	Resultado
ISM10001_smooth	SARIMAX_EXOG	84	0,873	✓	0,000	×
ISM10006_smooth	ETS	78	0,502	✓	0,001	×
ISM10003_smooth	ETS	84	0,212	✓	0,537	✓

Serie	Modelo	n	ARCH-LM	
			p-valor	Resultado
ISM10001_smooth	SARIMAX_EXOG	84	0,309	✓
ISM10006_smooth	ETS	78	0,812	✓
ISM10003_smooth	ETS	84	0,551	✓

Nota. ✓ indica $p > 0{,}05$ (no se rechaza la hipótesis nula). × indica $p \leq 0{,}05$ (se rechaza la hipótesis nula). Para Ljung-Box la hipótesis nula es ausencia de autocorrelación residual; para Jarque-Bera es normalidad de los residuos; para ARCH-LM es ausencia de heterocedasticidad condicional.

no tiene consecuencias prácticas aquí porque los parámetros de inventario (stock de seguridad y punto de reorden) se calcularon mediante simulación Monte Carlo sobre la distribución empírica de los errores, precisamente por anticipar esta posibilidad en series con consumos ocasionalmente extremos.

5.2.3 Variables exógenas: colinealidad y correlación con la demanda

El análisis de colinealidad entre las variables exógenas arrojó valores VIF entre 1,02 y 1,04, descartando redundancia entre ellas y confirmando que las tres pueden incorporarse simultáneamente en modelos condicionales sin distorsión por multicolinealidad, según lo observado en la Tabla 7.

La correlación de Spearman (Figura 6) entre cada serie y las exógenas con rezagos cortos mostró relaciones moderadas, sin que ninguna variable explique por sí sola el comportamiento de la demanda. La asociación más marcada correspondió a ICP10360 con la TRM en rezago contemporáneo (Spearman $= -0{,}452$), lo que es consistente con que esta serie agrupa artículos importados cuyo consumo puede ajustarse ante variaciones del tipo de cambio. Para ISM10001, la temperatura media con un rezago de un mes presentó la correlación más alta entre las series estables (Spearman $= -0{,}349$), sugiriendo una posible relación entre condiciones climáticas y el ritmo de consumo de magnetita en planta. La Tabla 8 reporta, para cada serie, la variable exógena con mayor correlación absoluta de Spearman tras explorar rezagos de 0 a 3 meses.

Tabla 7
Colinealidad entre variables exógenas

Variable	VIF
Temperatura media (°C)	1,020
Precipitación (mm)	1,036
TRM promedio (COP/USD)	1,034

Nota. Valores VIF cercanos a 1 indican ausencia de redundancia lineal entre las variables. El umbral convencional de preocupación se ubica en VIF ≥ 5; los valores obtenidos descartan multicolinealidad en cualquier modelo condicional que las incorpore simultáneamente.

Tabla 8
Mayor correlación absoluta por serie frente a variables exógenas con rezagos cortos

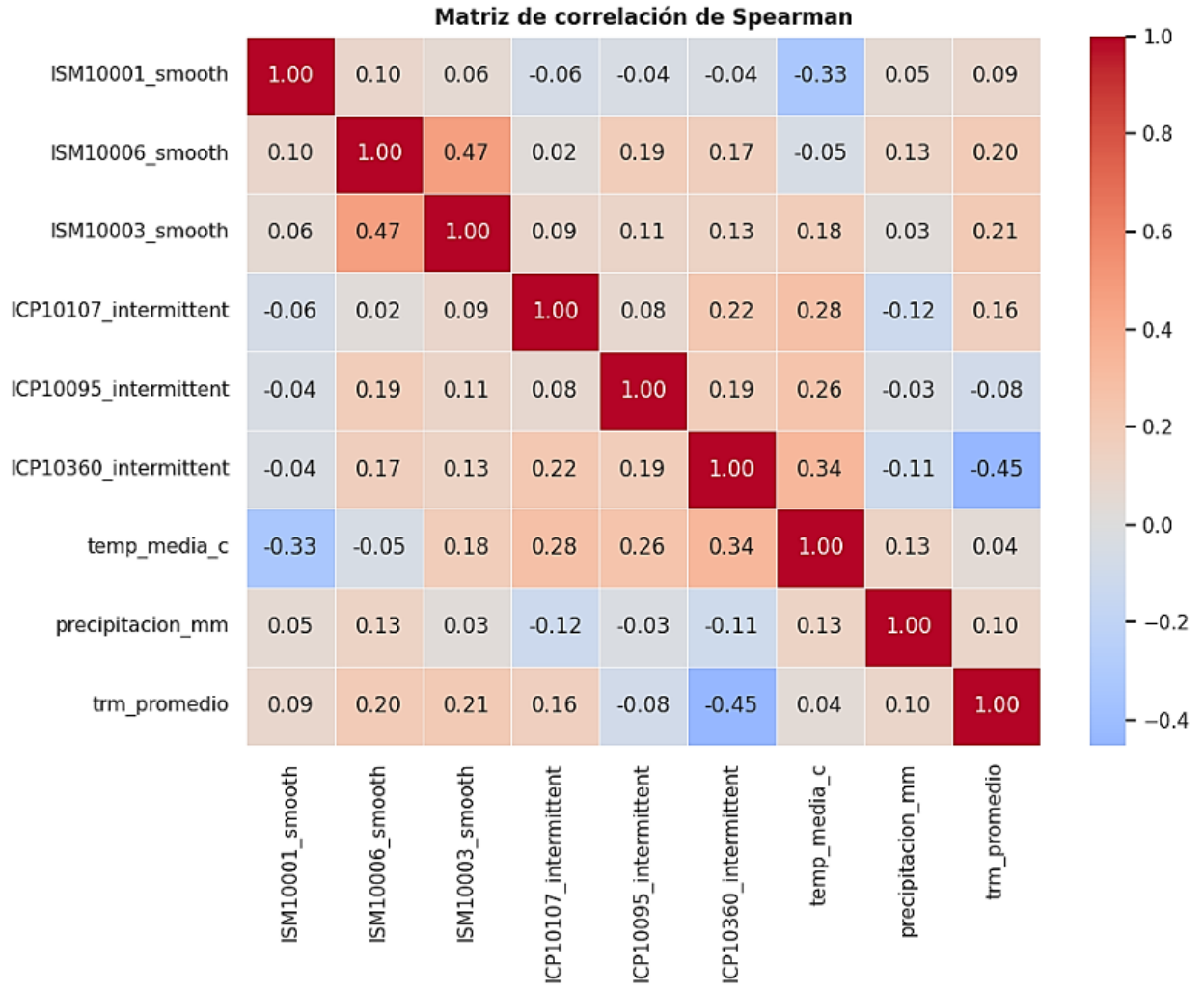
Serie	Variable exógena	Rezago	Spearman	$\lvert\rho\rvert$
ICP10095_intermittent	Temperatura media	1	0,285	0,285
ICP10107_intermittent	Temperatura media	0	0,276	0,276
ICP10360_intermittent	TRM promedio	0	−0,452	0,452
ISM10001_smooth	Temperatura media	1	−0,349	0,349
ISM10003_smooth	Temperatura media	2	0,289	0,289
ISM10006_smooth	TRM promedio	3	0,277	0,277

Nota. Se reporta únicamente la variable exógena con mayor correlación absoluta de Spearman para cada serie, considerando rezagos de 0 a 3 meses. La correlación negativa de ICP10360 con la TRM es consistente con el efecto del tipo de cambio sobre el costo de importación y la decisión de compra.

Estos resultados sugieren que las variables exógenas aportan información complementaria para el modelado, particularmente en los modelos SARIMAX para series estables y en el componente de ocurrencia del Hurdle-RF para series intermitentes, aunque no reemplazan la dinámica propia de cada serie.

Figura 6

Matriz de correlación de Spearman entre consumos de SKU y variables exógenas



Nota. La matriz presenta el grado de asociación monotónica contemporánea (rezago cero) entre las series de consumo y las variables exógenas consideradas en el estudio. Los valores cercanos a 1 indican relación positiva fuerte, los valores cercanos a -1 indican relación negativa fuerte y los valores próximos a 0 sugieren baja asociación directa. Las correlaciones con rezagos cortos se reportan en la Tabla 8.

5.3 Backtesting rolling-origin y selección de modelos

La validación fuera de muestra se realizó sobre los últimos 12 periodos de cada serie, comparando tres modelos por familia. Para las series estables se evaluaron Seasonal Naive, ETS y SARIMAX con variables exógenas; para las series intermitentes se evaluaron Croston-SBA, TSB y Hurdle-RF.

La especificación SARIMAX adoptada para las series estables (Tabla 9) priorizó una estructura parsimoniosa definida a partir del diagnóstico previo, sin reoptimizar el orden en cada partición. Las tres series quedaron con componente ARIMA no estacional y sin diferenciación estacional, lo que es consistente con la ausencia de estacionalidad anual identificada en la fase exploratoria.

Tabla 9
Especificación SARIMAX utilizada por serie estable

Serie	Orden (p, d, q)	Orden estacional (P, D, Q, s)	Tendencia
ISM10001_smooth	(1, 0, 1)	(0, 0, 0, 12)	Constante
ISM10006_smooth	(1, 1, 1)	(0, 0, 0, 12)	Constante
ISM10003_smooth	(1, 1, 1)	(0, 0, 0, 12)	Constante

Nota. La especificación se definió a partir del diagnóstico previo (pruebas ADF/KPSS y análisis STL), priorizando una estructura parsimoniosa consistente a lo largo de los 12 pasos del rolling-origin. No se aplicó auto-ARIMA dentro del bucle de backtesting para evitar filtraciones sutiles de especificación entre particiones. Las tres variables exógenas (temperatura media, precipitación y TRM) se incluyeron en todas las especificaciones.

Los resultados del backtesting muestran diferencias claras entre modelos y entre familias.

En las series estables (Tabla 10), ETS superó a los demás modelos en ISM10003 e ISM10006 con WAPE de 12,4% y 20,6% respectivamente, mientras que SARIMAX_EXOG resultó el mejor para ISM10001 con un WAPE de 19,5%. El desempeño de SARIMAX fue notoriamente inferior en ISM10003 e ISM10006 (WAPE superiores al 95%), lo que indica que la incorporación de variables exógenas no aportó capacidad predictiva adicional en esas series durante la ventana de validación. El Seasonal Naive actuó correctamente como referencia mínima, con errores consistentemente más altos que ETS en todas las series estables.

En las series intermitentes (Tabla 11), Hurdle-RF obtuvo el mejor desempeño en ICP10107 e ICP10360 (WAPE de 90,5% y 76,0% respectivamente), mientras que Croston-SBA resultó superior para ICP10095 (WAPE de 83,2%). Es importante contextualizar estos niveles de error: valores de WAPE cercanos o superiores al 80% son esperables en series con alta proporción de ceros, donde la mayoría de los periodos tienen demanda nula y cualquier predicción de consumo positivo genera un error grande en términos relativos. La métrica de precisión de actividad (*activity accuracy*) complementa esta lectura: Hurdle-RF logró el mejor balance entre identificar correctamente los meses con consumo y los meses sin consumo.

El ranking compuesto, construido como la posición promedio en WAPE, RMSE y MAE, confirmó la selección diferenciada que muestra la Tabla 12, y que la Figura 7 resume visualmente.

Tabla 10
Métricas de backtesting rolling-origin: series estables

Serie	Modelo	MAE	RMSE	sMAPE (%)	WAPE (%)	MASE	Sesgo (%)	Act. acc.
ISM10001	SARIMAX_EXOG	11,51	13,80	19,12	19,48	0,249	−5,88	100%
	ETS	13,44	16,11	24,39	22,75	0,291	−9,00	100%
	Seasonal Naive	23,88	30,01	44,36	40,44	0,517	−27,37	100%
ISM10003	ETS	60,79	80,06	12,35	12,43	0,420	4,19	100%
	Seasonal Naive	123,74	150,59	25,19	25,31	0,855	3,74	100%
	SARIMAX_EXOG	467,93	470,22	185,15	95,71	3,234	−95,71	67%
ISM10006	ETS	713,07	857,23	23,75	20,64	0,554	−12,12	100%
	Seasonal Naive	1.290,22	1.383,85	43,07	37,34	1,003	−29,90	100%
	SARIMAX_EXOG	3.368,41	3.456,69	189,86	97,49	2,617	−97,49	42%

Nota. Los modelos se ordenan de mejor a peor dentro de cada serie según el ranking compuesto (WAPE, RMSE, MAE). Act. acc. = precisión de actividad, proporción de meses en que el modelo clasificó correctamente la presencia o ausencia de consumo. $n = 12$ periodos de prueba en todas las series.

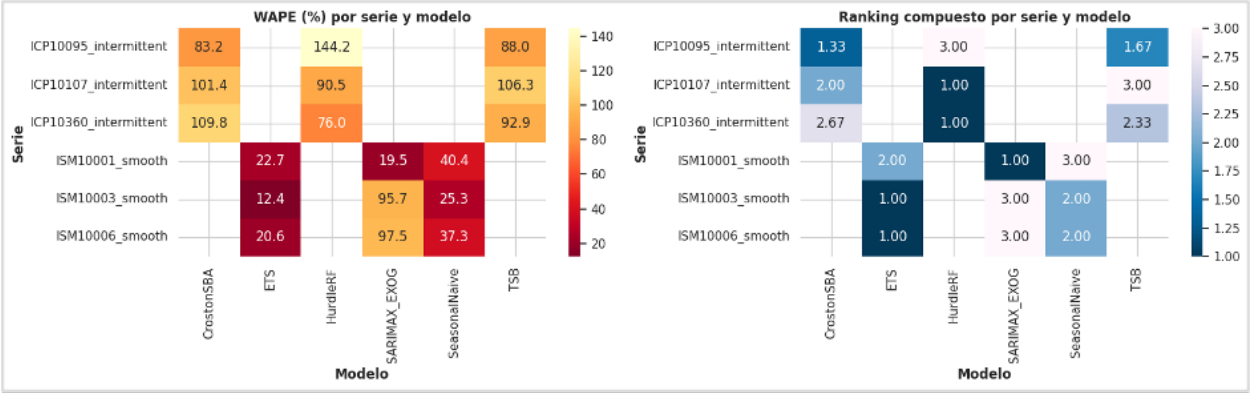
Tabla 11
Métricas de backtesting rolling-origin: series intermitentes

Serie	Modelo	MAE	RMSE	sMAPE (%)	WAPE (%)	MASE	Sesgo (%)	Act. acc.
ICP10095	Croston-SBA	1,39	1,95	115,72	83,16	0,795	−43,71	67%
	TSB	1,47	1,89	108,43	88,05	0,842	−12,61	67%
	Hurdle-RF	2,40	2,70	109,21	144,25	1,379	96,95	67%
ICP10107	Hurdle-RF	72,15	92,83	111,72	90,47	1,225	−11,37	50%
	Croston-SBA	80,88	94,65	156,18	101,41	1,373	−46,34	58%
	TSB	84,79	95,17	156,19	106,32	1,440	−39,77	58%
ICP10360	Hurdle-RF	0,760	1,239	57,28	76,02	0,944	−15,08	75%
	TSB	0,929	1,289	172,03	92,91	1,153	−63,63	42%
	Croston-SBA	1,098	1,277	164,36	109,80	1,363	−37,11	42%

Nota. Los niveles de WAPE superiores al 80% son esperables en series con alta proporción de ceros: cualquier predicción positiva en un mes de demanda nula genera un error relativo alto. La precisión de actividad complementa la lectura al medir la capacidad del modelo para clasificar correctamente meses con y sin consumo.

Figura 7

Comparación visual del WAPE y ranking compuesto por serie y modelo



Nota. La figura presenta dos mapas de calor empleados para comparar el desempeño de los modelos evaluados durante el proceso de backtesting. El panel izquierdo muestra el WAPE porcentual por combinación de serie y modelo, mientras que el panel derecho expone el ranking compuesto calculado a partir de las métricas principales de error.

Tabla 12

Modelo seleccionado por serie según ranking compuesto

Serie	Familia	Modelo seleccionado	WAPE (%)	Rank compuesto
ISM10001_smooth	Estable	SARIMAX_EXOG	19,48	1,00
ISM10003_smooth	Estable	ETS	12,43	1,00
ISM10006_smooth	Estable	ETS	20,64	1,00
ICP10095_intermittent	Esporádica	Croston-SBA	83,16	1,33
ICP10107_intermittent	Esporádica	Hurdle-RF	90,47	1,00
ICP10360_intermittent	Esporádica	Hurdle-RF	76,02	1,00

Nota. El ranking compuesto corresponde a la posición promedio de cada modelo en las métricas WAPE, RMSE y MAE dentro de su serie. Valores más bajos indican mejor desempeño relativo. En los tres casos estables el ganador obtuvo posición 1 en las tres métricas simultáneamente.

Un resultado relevante de esta fase es que ningún modelo resultó óptimo para todas las series: ETS dominó en dos de los tres ISM pero fue superado por SARIMAX en el tercero; en las series ICP, Hurdle-RF fue preferido en dos casos pero Croston-SBA resultó más adecuado para la serie con mayor irregularidad. Esto confirma que la selección del modelo debe realizarse por serie y no por familia, al menos cuando las series dentro de una misma categoría presentan diferencias de escala, vida activa y comportamiento tan marcadas como las observadas aquí. Aún así, la Tabla 13

muestra un consolidado de las métricas por familia, aportando una visión más generalizada de los resultados de los modelos.

Tabla 13
Desempeño promedio por familia de serie y modelo

Familia	Modelo	Series eval.	WAPE prom. (%)	RMSE prom.	MAE prom.	sMAPE prom. (%)	Rank compuesto
Estable	ETS	3	18,61	317,80	262,43	20,16	1,33
Estable	Seasonal Naive	3	34,36	521,48	479,28	37,54	2,33
Estable	SARIMAX_EXOG	3	70,90	1.313,57	1.282,62	131,38	2,33
Esporádica	Hurdle-RF	3	103,58	32,26	25,10	92,74	1,67
Esporádica	Croston-SBA	3	98,12	32,62	27,79	145,42	2,00
Esporádica	TSB	3	95,76	32,78	29,06	145,55	2,33

Nota. El promedio de WAPE en la familia esporádica no refleja el desempeño individual de la selección por serie, ya que el modelo con menor WAPE promedio (TSB) no fue seleccionado en ninguna serie. El ranking compuesto por serie, que integra WAPE, RMSE y MAE simultáneamente, resulta más informativo que cualquier métrica promedio aislada.

5.4 Pronóstico a 12 meses y política de inventario

El modelo seleccionado para cada SKU se reentrenó con toda la vida activa disponible para generar el pronóstico operativo a 12 meses. En todos los casos el modelo ganador del backtesting se conservó para la proyección final, sin modificaciones de arquitectura.

Las variables exógenas se proyectaron bajo un escenario base: temperatura y precipitación según el promedio histórico del mismo mes calendario, y TRM constante al promedio reciente de los últimos meses observados ($3.908 COP/USD). La Tabla 14 documenta los valores utilizados.

Tabla 14

Escenario base de variables exógenas para el horizonte de pronóstico (2026)

Mes	Temperatura media (°C)	Precipitación (mm)	TRM (COP/USD)
Enero	12,63	33,85	3.908,42
Febrero	13,24	71,37	3.908,42
Marzo	13,30	126,16	3.908,42
Abril	13,24	131,76	3.908,42
Mayo	13,25	97,35	3.908,42
Junio	12,67	73,85	3.908,42
Julio	12,52	38,21	3.908,42
Agosto	12,57	43,96	3.908,42
Septiembre	12,79	53,39	3.908,42
Octubre	12,83	92,17	3.908,42
Noviembre	12,77	123,56	3.908,42
Diciembre	12,90	49,18	3.908,42

Nota. Temperatura y precipitación se proyectaron mediante el promedio histórico del mismo mes calendario sobre toda la vida activa disponible, preservando la climatología mensual típica de la región. La TRM se mantuvo constante al promedio de los últimos meses observados para evitar imponer una deriva arbitraria sobre el horizonte de pronóstico.

5.4.1 Trayectorias de pronóstico

Las series estables proyectan trayectorias continuas, coherentes con su comportamiento histórico, como se observa en la Figura 8. ISM10001 muestra una reducción gradual respecto al nivel reciente según el modelo SARIMAX_EXOG, con bandas de incertidumbre que se amplían conforme avanza el horizonte. ISM10006 e ISM10003 conservan una dinámica relativamente estable bajo ETS, con intervalos de predicción crecientes que reflejan mayor incertidumbre acumulada en el horizonte de 12 meses.

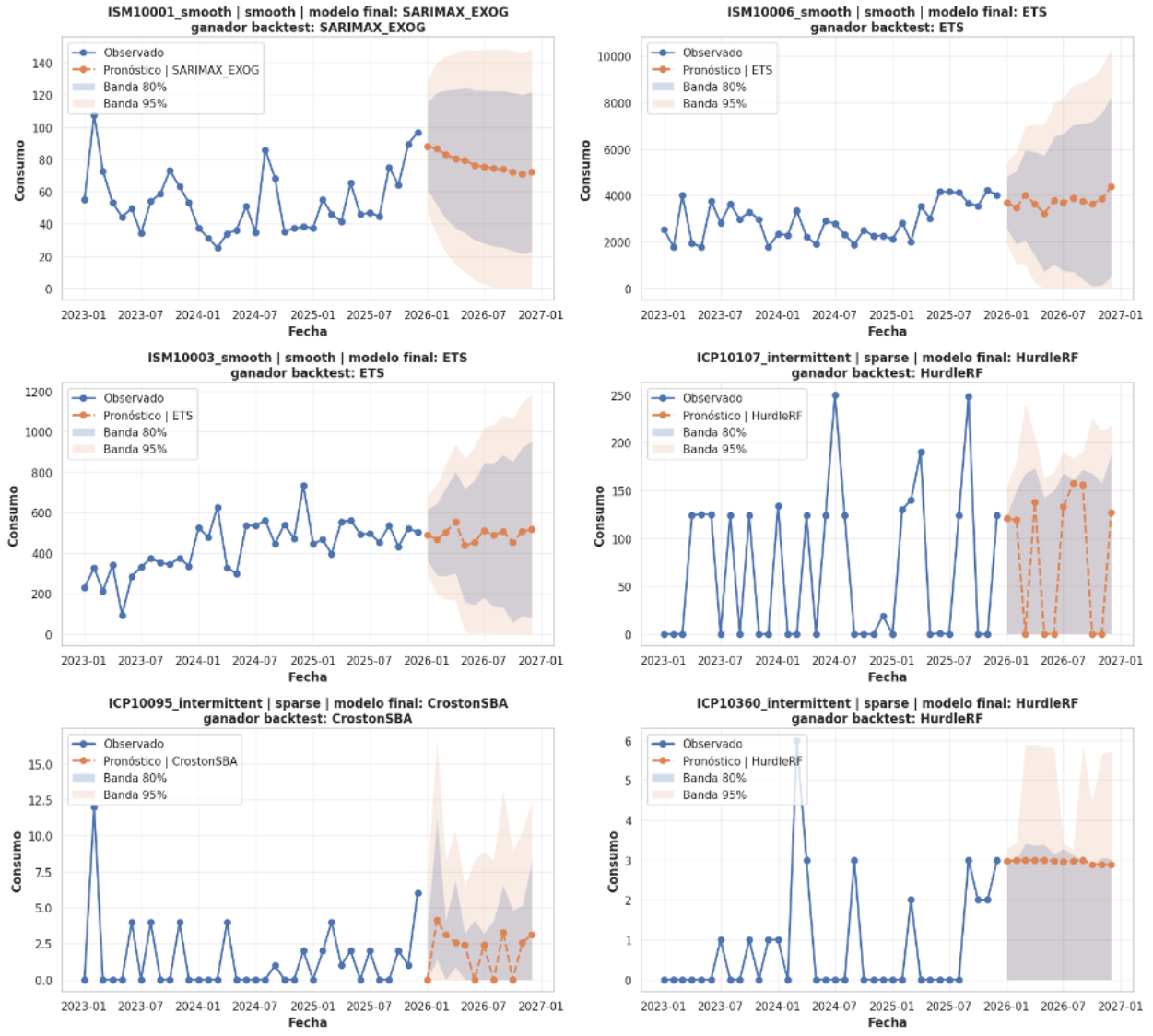
Las series intermitentes, por su parte, mantienen la lógica esporádica observada en el histórico. ICP10107 alterna meses con consumo positivo proyectado y meses con pronóstico nulo a través del Hurdle-RF, reproduciendo la alternancia característica de su demanda. ICP10095 proyecta consumos de baja magnitud con varios meses en cero, consistente con la baja frecuencia histórica de la serie. ICP10360 presenta valores bajos y dispersos, con bandas que reflejan la variabilidad asociada a demanda poco frecuente y de magnitud reducida.

5.4.2 Parámetros de política de inventario

A partir de la distribución simulada de la demanda acumulada en el lead time, calculada mediante simulación Monte Carlo, se obtuvieron el stock de seguridad y el punto de reorden para cada SKU, asumiendo los lead times reales por SKU informados por la empresa (0,66 meses para ISM10001 e ISM10003, 0,33 meses para ISM10006 y 3 meses para las series ICP) y un nivel de

Figura 8

Pronóstico final a 12 meses por SKU con bandas de incertidumbre



Nota. La figura presenta el empalme entre los consumos históricos recientes y los pronósticos finales generados para cada SKU. La línea azul representa los valores observados, la línea naranja corresponde al pronóstico mensual, y las áreas sombreadas muestran las bandas de incertidumbre al 80% y 95%. Cada panel identifica la familia de demanda, el modelo final utilizado y el modelo ganador durante el proceso de backtesting.

servicio objetivo del 95%.

Los mayores requerimientos de inventario se concentran en las series ISM, como era esperable dado su mayor volumen de consumo (Tabla 15). ISM10006 registra el pronóstico anual más alto (48.768 unidades), el stock de seguridad más elevado (382 Gal) y el punto de reorden más alto (1.717 Gal). ISM10003 presenta valores intermedios dentro de la misma familia, mientras que ISM10001 mantiene requerimientos más moderados.

Tabla 15

Parámetros de política de inventario por SKU

Serie	Familia	*L* (meses)	SL	Pronóstico anual	Prom. mensual	Stock de seguridad	Punto de reorden
ISM10001_smooth	Estable	0,66	95%	934,31	77,86	21,55	80,07
ISM10003_smooth	Estable	0,66	95%	5.894,42	491,20	110,26	434,28
ISM10006_smooth	Estable	0,33	95%	48.767,87	4.063,99	381,99	1.716,89
ICP10095_intermittent	Esporádica	3	95%	23,63	1,97	10,64	18,47
ICP10107_intermittent	Esporádica	3	95%	812,80	67,73	161,63	341,70
ICP10360_intermittent	Esporádica	3	95%	36,05	3,00	4,43	9,84

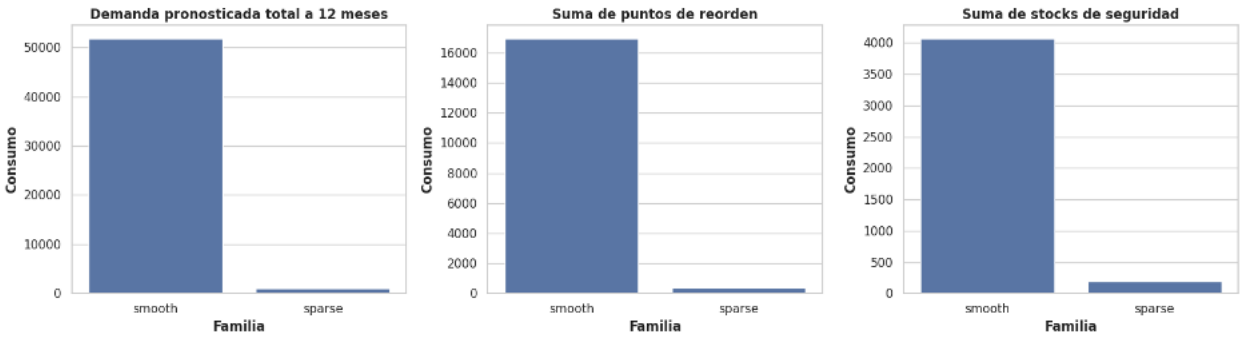
Nota. L = lead time en meses. SL = nivel de servicio objetivo. El stock de seguridad y el punto de reorden se calcularon mediante simulación Monte Carlo sobre la distribución de demanda acumulada durante el lead time, sin asumir distribución paramétrica. Los valores de pronóstico anual y promedio mensual corresponden a la suma e media de los 12 meses pronosticados (enero a diciembre de 2026). Las unidades de medida corresponden a las declaradas en la Tabla 2.

Entre las series ICP, ICP10107 concentra los mayores requerimientos de la familia intermitente: un pronóstico anual de 813 unidades, un stock de seguridad de 162 unidades y un punto de reorden de 342 unidades. ICP10095 e ICP10360 muestran niveles de demanda proyectada considerablemente más bajos, con puntos de reorden inferiores a 19 y 10 unidades respectivamente, reflejando la baja frecuencia y magnitud de su consumo histórico.

La Figura 9 permite apreciar que la concentración de requerimientos en la familia estable no implica que la gestión de las series esporádicas sea trivial. Aunque los valores agregados son menores, la incertidumbre en la ocurrencia de demanda hace que el stock de seguridad represente una proporción más alta respecto a la demanda esperada: para ICP10095, el stock de seguridad (10,64 unidades) supera en valor absoluto a la demanda media mensual proyectada (2,0 unidades), lo que refleja la necesidad de mantener reservas ante consumos puntuales impredecibles.

Figura 9

Puntos de reorden y stocks de seguridad por familia de demanda



Nota. La figura presenta tres indicadores agregados por familia de serie: demanda total pronosticada a 12 meses, suma de puntos de reorden y suma de stocks de seguridad. La familia estable agrupa las series con consumo continuo, mientras que la familia esporádica reúne las series con demanda intermitente o irregular.

Es importante destacar que, aunque por la naturaleza de la unidad de medida de ciertos insumos sus valores deberían ser enteros, no se realiza el redondeo y se enuncian valores continuos por simplicidad. El sistema puede ajustarlo dependiendo la medición de cada SKU.

Asimismo, los parámetros calculados constituyen una base cuantitativa inicial. Su adopción operativa requiere contrastarlos con los lead times reales de cada proveedor, la criticidad específica de cada artículo, los costos de quiebre, las restricciones presupuestales y la frecuencia efectiva de revisión del inventario en Milpa.

5.5 Backtesting de política de inventario: gestión real frente al sistema propuesto

Para cerrar el ciclo entre pronóstico y decisión operativa, se diseñó una simulación de inventario que contrasta la gestión real observada entre agosto de 2025 y marzo de 2026 con el comportamiento hipotético del sistema de reposición propuesto. Para evitar solapamiento entre entrenamiento y evaluación, los parámetros aplicados en esta sección se recalcularon con la información disponible hasta julio de 2025, lo que produce ROP y SS diferentes a los reportados en la Tabla 15 (entrenada con todo el histórico hasta diciembre de 2025). El ejercicio opera bajo un esquema de revisión continua con posición de inventario: en cada día del periodo evaluado, la posición (stock físico más pedidos en tránsito) se compara con el punto de reorden; cuando cae por debajo de este umbral, se emite una orden cuya cantidad cubre la demanda esperada durante el lead time más el stock de seguridad. Los lead times se expresan en días (20 días para ISM10001 e ISM10003, 10 días para ISM10006 y 90 días para las series ICP), lo que permite capturar correctamente la dinámica

intra-mes de los SKUs con reposición rápida.

5.5.1 Supuestos de la simulación

La simulación parte de cuatro supuestos documentados. Primero, el stock inicial de cada SKU al 31 de julio de 2025 se fijó en el 120 % del punto de reorden calculado, como aproximación conservadora al nivel que mantendría una operación con continuidad mínima garantizada. Segundo, los parámetros de política (ROP y SS) se mantienen fijos durante todo el horizonte, derivados del modelo entrenado con datos hasta julio de 2025 para evitar solapamiento con el periodo de evaluación. Tercero, la cantidad de orden se calcula como el máximo entre el pronóstico promedio mensual y la brecha entre el ROP, el stock de seguridad y la posición actual. Cuarto, la simulación opera sobre cantidades continuas; en una implementación operativa real, cada cantidad debería redondearse al múltiplo de empaque o unidad mínima de compra del proveedor.

5.5.2 Resultados de la simulación

La Tabla 16 consolida los indicadores de nivel de servicio y stock promedio diario para ambos escenarios. La Figura 10 presenta la evolución temporal del inventario físico día a día para los seis SKUs.

Tabla 16

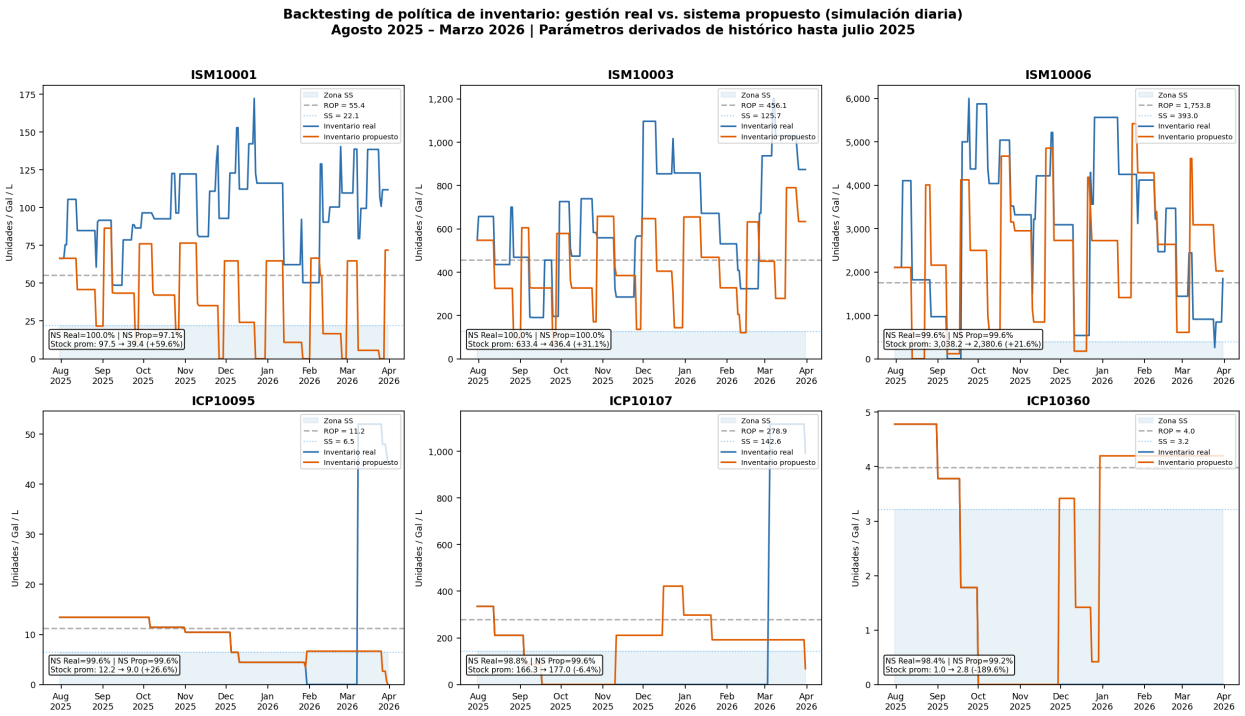
Backtesting de política de inventario: indicadores por SKU

SKU	NS Real	NS Prop.	Stock prom. real	Stock prom. prop.	Reducción	Órdenes prop.
ISM10001	100,0 %	100,0 %	97,5	39,4	+59,6 %	8
ISM10003	100,0 %	100,0 %	633,4	436,4	+31,1 %	7
ISM10006	99,6 %	99,6 %	3.038	2.381	+21,6 %	7
ICP10095	99,6 %	99,6 %	12,2	9,0	+26,6 %	3
ICP10107	98,8 %	99,6 %	166,3	177,0	−6,4 %	3
ICP10360	98,4 %	99,2 %	1,0	2,8	−190 %	2

Nota. NS = nivel de servicio diario (proporción de días sin quiebre de stock). Stock promedio expresado en las unidades de cada SKU (Tabla 2). El nivel de servicio de la gestión propuesta se calcula sobre la política de revisión continua con posición de inventario; las órdenes reales corresponden a las recepciones efectivas registradas en el sistema ERP de la empresa durante el periodo agosto 2025 – marzo 2026.

Figura 10

Evolución diaria del inventario: gestión real vs. sistema propuesto



Nota. Cada panel presenta el stock físico diario para un SKU entre agosto de 2025 y marzo de 2026. La línea azul corresponde al inventario real registrado en el ERP; la línea naranja representa el inventario que habría resultado de aplicar la política de punto de reorden con los parámetros derivados del modelo. La banda azul claro delimita la zona del stock de seguridad y la línea punteada gris indica el punto de reorden. El stock inicial (cierre de julio de 2025) es idéntico en ambos escenarios.

Para las tres series estables, el sistema propuesto mantiene niveles de servicio equivalentes a la gestión real (entre 99,6 % y 100 %) con stocks promedio entre un 22 % y un 60 % menores. ISM10001 es el caso más marcado: la política activa de reposición con lead time corto (20 días) permite operar con un promedio de 39 unidades frente a las 98 que mantuvo la empresa, sin ningún día adicional de quiebre. ISM10003 e ISM10006 muestran reducciones moderadas pero consistentes, de 31 % y 22 % respectivamente.

En las series ICP el resultado refleja el *trade-off* teórico entre nivel de servicio y capital inmovilizado en contextos de demanda intermitente con lead time largo. ICP10107 e ICP10360 elevan el stock promedio respecto a la realidad, pero en ambos casos la gestión real operó con quiebres efectivos que el sistema propuesto habría evitado. ICP10360 es el caso más ilustrativo: con un stock promedio real de apenas una unidad y un nivel de servicio del 98,4 %, pero eso obedece a una demanda esporádica que en otro ciclo podría haber generado desabastecimiento. El sistema propuesto, al activar el stock de seguridad estructural (3,2 unidades), eleva el capital inmovilizado

pero garantiza cobertura ante ocurrencias imprevistas.

5.5.3 Valorización financiera

Para cuantificar el impacto en términos de capital de trabajo, la Tabla 17 reporta la valorización financiera obtenida al multiplicar el stock promedio diario de cada escenario por el costo unitario promedio reportado por la empresa. La diferencia entre ambos representa la liberación media de capital que el sistema propuesto habría logrado durante el periodo evaluado.

Tabla 17

Valorización del capital promedio inmovilizado por SKU (COP)

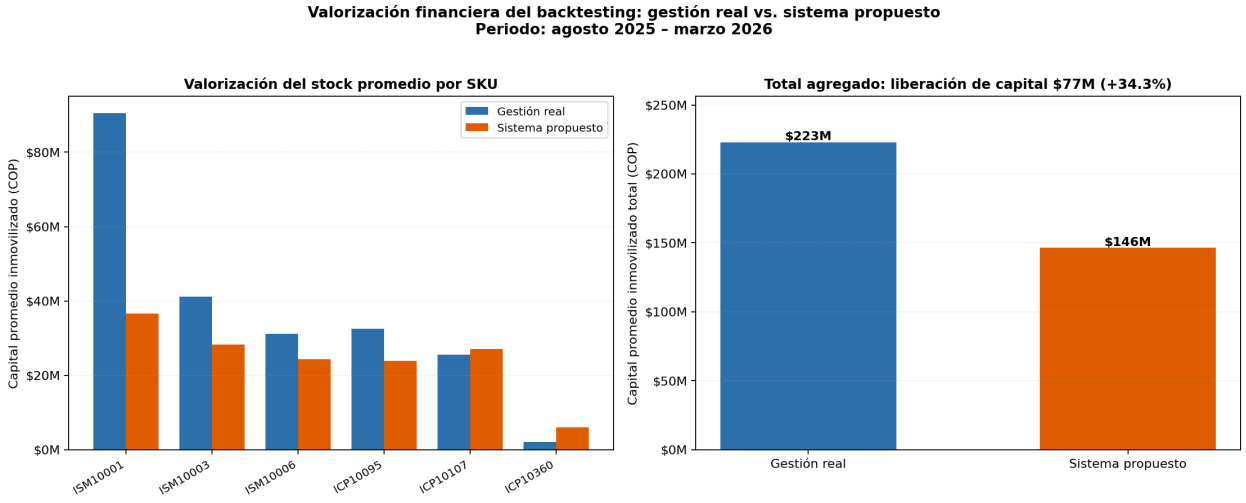
SKU	Costo unitario	Valor real	Valor prop.	Liberación	Var.
ISM10001	$ 928.602	$ 90.503.708	$ 36.583.893	$ 53.919.816	+59,6 %
ISM10003	$ 64.939	$ 41.131.679	$ 28.336.479	$ 12.795.200	+31,1 %
ISM10006	$ 10.233	$ 31.090.109	$ 24.361.062	$ 6.729.046	+21,6 %
ICP10095	$2.666.788	$ 32.555.222	$ 23.897.707	$ 8.657.514	+26,6 %
ICP10107	$ 153.279	$ 25.488.918	$ 27.132.685	$ -1.643.767	−6,4 %
ICP10360	$2.146.022	$ 2.080.924	$ 6.026.178	$ -3.945.254	−190 %
Total		**$222.850.560**	**$146.338.004**	**$76.512.556**	**+34,3 %**

Nota. El valor de cada escenario resulta de multiplicar el stock promedio diario (en unidades) por el costo unitario promedio reportado por la empresa para el periodo agosto 2025 – marzo 2026. La liberación de capital representa la diferencia media sostenida en inventario valorizado y no un flujo de caja puntual; su conversión a costo financiero requiere aplicar la tasa de oportunidad del capital de trabajo de la empresa.

Sobre el conjunto de los seis SKUs, el sistema propuesto habría reducido el capital promedio inmovilizado en $76,5 millones de COP, equivalente al 34,3 % del valor total observado en la gestión real durante el mismo periodo (Figura 11). Esta cifra corresponde únicamente a seis artículos de una muestra estratégica; extendida al portafolio completo de materiales críticos, la magnitud potencial del impacto financiero es considerablemente mayor. Vale precisar que esta reducción no es un ahorro contable directo sino una liberación de capital de trabajo: recursos que hoy permanecen inmovilizados en estanterías y que, bajo el sistema propuesto, estarían disponibles para mantenimiento mayor, inversión en confiabilidad o mejora de capacidad operativa.

Figura 11

Valorización financiera del backtesting: capital promedio inmovilizado por escenario



Nota. El panel izquierdo compara el capital promedio inmovilizado por SKU en cada escenario. El panel derecho presenta el total agregado de los seis SKUs. La diferencia de $76,5 millones de COP representa el capital de trabajo que habría sido liberado de la bodega bajo la política propuesta, disponible para otros usos productivos de la operación.

6. Discusión y análisis de resultados

Los resultados obtenidos permiten examinar la validez de las decisiones metodológicas, contrastar los hallazgos con las expectativas teóricas y reconocer los supuestos que condicionan el alcance de las conclusiones.

Sobre la clasificación y la separación de familias. La clasificación ADI/CV² cumplió su función de separar rutas de modelado antes de ajustar cualquier modelo. La coherencia total entre el sufijo declarado en el nombre de cada serie y su categoría empírica no es un resultado trivial: confirma que la etiqueta asignada originalmente por el sistema refleja un patrón estadístico real y no una convención arbitraria. Esto le da solidez al punto de partida del análisis y permite confiar en que la separación entre familias estables e intermitentes no fue una decisión subjetiva.

Un aspecto que merece atención es el caso de ICP10095, clasificada como irregular (ADI $\geq 1{,}32$ y CV² $\geq 0{,}49$) aunque su sufijo la etiqueta como intermittent. La diferencia entre ambas categorías no es menor: una serie irregular combina baja frecuencia de consumo con alta variabilidad en la magnitud, lo que la hace más difícil de pronosticar que una serie simplemente intermitente. El hecho de que Croston-SBA haya resultado el mejor modelo para esta serie, por encima de Hurdle-RF, sugiere que la mayor flexibilidad del modelo de dos etapas no compensa necesariamente la escasez de datos positivos disponibles para entrenar el componente de magnitud.

Sobre la utilidad de las variables exógenas. El análisis de correlación y el desempeño comparativo en el backtesting ofrecen una lectura consistente sobre el papel de las variables exógenas. En las series estables, el único caso donde SARIMAX_EXOG superó a ETS fue ISM10001, que a su vez registró la mayor correlación absoluta con una variable exógena dentro de esa familia: temperatura media con un rezago de un mes (Spearman $= -0{,}349$). En cambio, ISM10003 e ISM10006, cuyas correlaciones con las exógenas fueron más bajas, terminaron con ETS como modelo ganador y SARIMAX obtuvo los peores resultados de la comparación, con WAPE superiores al 95%. Esta correspondencia entre nivel de correlación y utilidad del modelo condicional no es una coincidencia: cuando las variables exógenas no tienen una relación suficientemente estable con la demanda, incorporarlas añade ruido en lugar de señal, especialmente en muestras cortas de 84 observaciones donde cada parámetro adicional consume grados de libertad.

Para las series intermitentes, la situación es distinta. Hurdle-RF incorpora las variables exógenas en el componente de ocurrencia (clasificación binaria) y en el componente de magnitud (regresión sobre demanda positiva), lo que le permite explotar relaciones no lineales y rezagadas sin imponer una estructura paramétrica rígida. El mejor desempeño de Hurdle-RF en ICP10107 e ICP10360 frente a los métodos clásicos sugiere que esa flexibilidad fue útil, aunque no es posible aislar cuánto de la ganancia proviene de las exógenas y cuánto de la arquitectura de dos etapas en sí misma.

Un supuesto relevante en esta etapa es que las correlaciones de Spearman observadas sobre el histórico se mantienen estables en el horizonte futuro. Si las condiciones operativas o macroeconómicas cambian de forma estructural, como una variación sostenida en la TRM o un cambio en el proceso productivo, las relaciones estimadas perderían vigencia y los modelos condicionales se verían más afectados que ETS o Croston-SBA, que dependen únicamente del comportamiento pasado de la propia serie.

Sobre la especificación fija del SARIMAX. La decisión de no reoptimizar el orden ARIMA en cada partición del rolling-origin fue una elección consciente orientada a evitar filtraciones sutiles de especificación entre ventanas de entrenamiento. La consecuencia es que la estructura fija puede no ser óptima para todos los puntos de validación, lo que puede inflar el error reportado para SARIMAX respecto a lo que obtendría con reoptimización. Sin embargo, el beneficio es que el desempeño reportado refleja condiciones más cercanas a las de uso real, donde el analista no recalibra la estructura del modelo cada mes.

Sobre los parámetros de inventario. Como se argumentó en el marco teórico 3.3.11.2, la simulación Monte Carlo es la alternativa más adecuada para este contexto precisamente porque no impone supuestos de normalidad sobre la distribución de la demanda en lead time. No obstante, los parámetros obtenidos descansan sobre dos supuestos que la empresa debe contrastar con su

operación real: que los tiempos de entrega dados por la misma se cumplan en la realidad por parte de los proveedores y que el nivel de servicio objetivo del 95% es uniforme. En la práctica, los tiempos de entrega varían por proveedor y los niveles de servicio deberían diferenciarse según la criticidad del artículo y el costo de quiebre asociado.

Un sesgo adicional a reconocer es que los pronósticos que alimentan la simulación Monte Carlo provienen de modelos entrenados sobre un periodo que incluye años atípicos (2020–2021 por efectos de la pandemia y ajustes operativos); aunque la operación durante esos años se mantuvo relativamente estable, insumos como ISM10001 presentan tendencia negativa en ese periodo. Si esos eventos distorsionaron el patrón de consumo de algún SKU de forma no representativa del comportamiento futuro esperado, el stock de seguridad calculado puede estar sobreestimado o subestimado. La actualización periódica del modelo con datos recientes es la forma más directa de reducir este riesgo.

El diagnóstico de residuos de los modelos estables (Tabla 6) refuerza esta decisión metodológica. Aunque los residuos de ISM10001 e ISM10006 no superan la prueba de normalidad de Jarque-Bera, las tres series muestran ausencia de autocorrelación serial y de heterocedasticidad condicional, lo que valida la especificación dinámica de los modelos. La desviación de la normalidad, atribuible a eventos atípicos puntuales en el historial, es precisamente el argumento que justifica el uso de simulación Monte Carlo en lugar de la fórmula paramétrica $SS = z \cdot \sigma$: al derivar los cuantiles directamente sobre la distribución empírica simulada, el cálculo de inventario es robusto frente a la asimetría y las colas pesadas de los residuos reales.

Sobre el backtesting de política y la valorización financiera. El ejercicio de simulación cierra el argumento central del trabajo al traducir los parámetros estadísticos en consecuencias operativas concretas. Que las series estables reduzcan el stock promedio entre un 22 % y un 60 % sin sacrificar nivel de servicio no es un resultado trivial: la gestión real de la empresa durante ese mismo periodo operó con compras reactivas y sin ningún criterio cuantitativo explícito de reposición, lo que produjo stocks significativamente mayores al necesario para garantizar continuidad. El sistema propuesto compra en el momento correcto y en la cantidad que el pronóstico justifica, lo que naturalmente resulta en menor acumulación. No obstante, este resultado está condicionado a un supuesto que, aunque reconocido dentro de las limitaciones del estudio, merece señalarse con claridad: los parámetros de política se calcularon asumiendo que los tiempos de entrega informados por la empresa se cumplen de forma exacta y constante por parte de los proveedores. Bajo ese supuesto, el stock de seguridad derivado de la simulación Monte Carlo cubre únicamente la incertidumbre de la demanda, no la variabilidad del lead time. Si en la práctica los proveedores incurren en retrasos, el punto de reorden calculado puede resultar insuficiente para evitar quiebres, y la reducción de capital inmovilizado reportada sería parcialmente ilusoria, pues el sistema compraría menos, pero en algunos casos compraría tarde. La magnitud del impacto financiero positivo debe interpretarse en-

tonces como una estimación válida bajo condiciones de suministro estables, no como un resultado garantizado en contextos con alta variabilidad de entrega.

El comportamiento de las series ICP merece una lectura más cuidadosa. El incremento en capital inmovilizado para ICP10107 e ICP10360 no es una falla del modelo sino su consecuencia directa en presencia de lead times largos e incertidumbre alta: un stock de seguridad de 143 unidades para ICP10107 es costoso, pero es el precio de evitar un quiebre en un artículo cuya reposición tarda tres meses. La gestión real, que operó con un solo pedido durante todo el periodo, incurrió efectivamente en días de quiebre. La comparación pone de manifiesto que la empresa estaba asumiendo un riesgo operativo implícito que no aparecía en el inventario valorizado pero que sí representaba una exposición real a la detención de equipos.

Un aspecto metodológico que conviene señalar es que los parámetros de política se calcularon con datos hasta julio de 2025 y se aplicaron a un periodo que arranca en agosto del mismo año, garantizando separación temporal entre entrenamiento y evaluación. Esta decisión, más rigurosa que usar los parámetros del modelo final entrenado hasta diciembre de 2025, implica que el sistema opera con pronósticos derivados de un histórico más corto, lo que en algunos SKUs produce umbrales más conservadores. Los resultados obtenidos bajo esta restricción son, por tanto, una cota inferior del potencial del sistema con información completa.

Limitaciones del estudio. Los resultados deben interpretarse considerando cinco limitaciones. Primera, la ventana de evaluación de la política operativa cubre solo ocho meses, lo que limita la generalización del impacto financiero a contextos con mayor variabilidad anual. Segunda, el alcance se concentra en seis SKUs sobre un portafolio MRO mucho más amplio, por lo que las cifras agregadas no son extrapolables sin un análisis específico por familia. Tercera, las variables exógenas climáticas se construyeron como promedio simple de tres municipios sin ponderación por contribución productiva. Cuarta, los lead times se asumen deterministas, lo que subestima el riesgo de quiebre cuando un proveedor se retrasa. Quinta, el periodo de entrenamiento incluye los años 2020 y 2021 con patrones de consumo atípicos asociados a la pandemia, cuya representatividad respecto al comportamiento futuro esperado es discutible.

7. Conclusiones

El trabajo desarrolló un sistema de pronóstico de demanda y política de inventario para seis insumos MRO estratégicos de C.I. Milpa S.A., cubriendo los cuatro objetivos planteados.

Primer objetivo específico. el análisis exploratorio permitió establecer que el dataset no puede tratarse como un conjunto homogéneo. Las tres series ISM presentaron consumo continuo sin ningún mes en cero, con variaciones asociadas principalmente a cambios de nivel y oscilaciones de corto plazo, y sin estacionalidad anual marcada en ninguna de ellas. Las tres series ICP mostraron

una dinámica opuesta: porcentajes de ceros entre 60,7% y 69,4%, vidas activas más cortas y picos puntuales de demanda que no siguen un patrón mensual repetitivo. Esta caracterización justificó la separación en familias de modelado y descartó la aplicación de un único método sobre todas las series.

Segundo objetivo específico. la validación mediante backtesting rolling-origin confirmó que la selección óptima depende del patrón de cada serie y no puede definirse a nivel de familia. En las series estables, ETS resultó superior en dos de los tres casos (ISM10003 e ISM10006) mientras que SARIMAX_EXOG fue preferido en ISM10001, la única serie estable con una correlación moderada y estable con una variable exógena. En las series intermitentes, Hurdle-RF dominó en ICP10107 e ICP10360, mientras que Croston-SBA fue más adecuado para ICP10095, la serie con mayor irregularidad. Esta distribución de resultados es coherente con la literatura: ningún modelo garantiza desempeño uniforme sobre series con estructuras estadísticas distintas (Ulrich et al., 2022), y la selección basada en evidencia fuera de muestra es más confiable que la elección a priori por categoría.

Un hallazgo que merece destacarse es la correspondencia observada entre el nivel de correlación con variables exógenas y la utilidad del modelo condicional. ISM10001, la única serie estable donde SARIMAX superó a ETS, fue también la que mostró la mayor correlación absoluta con una exógena dentro de esa familia. En las series donde esa correlación fue baja, SARIMAX produjo los errores más altos de la comparación. Esta evidencia sugiere que la incorporación de variables externas tiene sentido metodológico cuando existe una relación empírica suficientemente estable, y que su inclusión indiscriminada puede perjudicar el pronóstico en muestras cortas.

Tercer objetivo específico. los pronósticos a 12 meses se tradujeron en parámetros operativos de inventario mediante simulación Monte Carlo, con lead times de 0,3; 0,6 y 3 meses y nivel de servicio del 95%. Los resultados muestran una concentración de requerimientos en la familia estable, donde ISM10006 registra el mayor pronóstico anual (48.768 galones), el stock de seguridad más alto (382 Gal) y el punto de reorden más elevado (1.716 Gal). En las series esporádicas, ICP10107 concentra los mayores requerimientos de su familia, con un punto de reorden de 342 unidades, mientras que ICP10095 e ICP10360 presentan requerimientos considerablemente más bajos, aunque con stocks de seguridad que representan una proporción alta respecto a la demanda media mensual proyectada, lo que refleja la incertidumbre inherente a la ocurrencia esporádica del consumo.

Desde una perspectiva más amplia, el trabajo demuestra que es posible construir un flujo analítico reproducible para la gestión de inventarios MRO en un contexto minero, partiendo de registros históricos de consumo y variables exógenas de acceso público. La secuencia de auditoría, clasificación, diagnóstico, comparación de modelos y traducción a política de inventario genera

criterios cuantificables y trazables que pueden reemplazar progresivamente las decisiones basadas únicamente en experiencia operativa. Esta institucionalización del conocimiento es, en sí misma, un aporte organizacional independiente de los valores numéricos obtenidos para cada SKU.

La metodología desarrollada es escalable: aunque el alcance se limitó a seis artículos estratégicos, la estructura de datos, los criterios de clasificación y las rutinas de modelado pueden extenderse a otras familias del portafolio MRO con ajustes menores. Replicar el enfoque sobre artículos de cuello de botella o sobre referencias con lead times más largos permitiría ampliar la cobertura del sistema y generar evidencia adicional sobre su desempeño en condiciones distintas.

Cuarto objetivo específico. La simulación de política de inventario sobre datos reales del periodo agosto 2025 – marzo 2026 permite cuantificar el impacto operativo del sistema propuesto con evidencia fuera de muestra. En las tres series de consumo continuo (ISM), la política de punto de reorden con stock de seguridad derivada del modelo habría mantenido niveles de servicio equivalentes a la gestión real, entre el 99,6 % y el 100 % de días sin quiebre, con un capital promedio inmovilizado entre un 22 % y un 60 % menor. Valorizado al costo unitario de cada artículo, el diferencial agrega $76,5 millones de COP sobre los seis SKUs evaluados, equivalente a una reducción del 34,3 % en el inventario medio. Para las series intermitentes (ICP), el sistema eleva el stock promedio en dos de los tres SKUs pero mejora el nivel de servicio frente a una gestión real que operó con pedidos esporádicos e incurrió en días efectivos de quiebre. El ejercicio confirma que los parámetros obtenidos en la Sección 5.4 traducidos a una política operativa concreta, producen resultados defendibles y cuantificables en términos de continuidad y eficiencia de capital.

Objetivo general. El sistema desarrollado cumple con los compromisos planteados al inicio del trabajo. A partir de registros históricos de consumo y variables exógenas de acceso público, se construyó un flujo analítico reproducible que cubre desde la caracterización de la demanda hasta la derivación de parámetros operativos de reabastecimiento, pasando por la comparación rigurosa de modelos mediante validación fuera de muestra. El backtesting de política sobre datos reales cuantificó una reducción del 34,3% en el capital promedio inmovilizado para los seis SKUs evaluados, sin deterioro del nivel de servicio en las series de consumo continuo. Más allá del resultado numérico, el trabajo establece criterios de reposición auditables, trazables y actualizables que pueden reemplazar progresivamente las decisiones basadas en percepción operativa. En ese sentido, el método es replicable sobre otras familias del portafolio MRO, y su aplicación sistemática podría ampliar de forma considerable el impacto financiero y organizacional documentado aquí.

8. Trabajos futuros

Los resultados obtenidos abren varias líneas de extensión metodológica y operativa que no pudieron abordarse dentro del alcance del presente trabajo.

Incorporación del plan de minado como variable exógena determinante. El sistema de pronóstico asume que el futuro replicará los patrones estadísticos del pasado, lo que lo expone a cambios operativos disruptivos como modificaciones en el plan de minado, renovación de flota o cambios en los frentes de extracción. Una extensión natural consiste en incorporar el tonelaje proyectado por zona como variable exógena en los modelos, de modo que variaciones planificadas en el ritmo productivo se traduzcan directamente en ajustes del pronóstico de consumo. Esta incorporación requiere, sin embargo, que los jefes de proceso comuniquen oportunamente al sistema cualquier información de contexto operativo que no quede capturada en el histórico: el conocimiento empírico sobre mantenimientos mayores programados, cambios de proceso o eventos operativos previsibles es un insumo que el modelo no puede inferir por sí solo y cuya omisión puede afectar la calidad del pronóstico con independencia de su complejidad estadística.

Actualización periódica de los modelos. Los modelos entrenados sobre el periodo 2019–2025 incluyen años con patrones de consumo atípicos asociados a la pandemia y a ajustes operativos de ese período. A medida que se acumulen nuevos datos, el reentrenamiento periódico de los modelos con ventanas de historia más recientes reducirá el peso de esos eventos sobre los parámetros estimados y mejorará la representatividad del pronóstico respecto al comportamiento futuro esperado. Esta actualización no requiere infraestructura especializada: basta con ejecutar nuevamente el flujo de auditoría, diagnóstico y backtesting sobre el dataset ampliado, conservando la misma estructura metodológica desarrollada en este trabajo.

Modelado estocástico del lead time y coordinación con proveedores. El cálculo del punto de reorden asume tiempos de entrega fijos informados por la empresa, lo que subestima el riesgo de quiebre cuando un proveedor se retrasa. Incorporar el lead time como una variable estocástica dentro de la simulación Monte Carlo, estimada a partir del historial real de recepciones, permitiría derivar parámetros de inventario que capturen explícitamente esa fuente de incertidumbre. En paralelo, la disponibilidad de pronósticos cuantitativos de consumo a 12 meses abre la posibilidad de negociar con proveedores acuerdos marco con compromisos de tiempos de entrega, convirtiendo el sistema de pronóstico en un instrumento de gestión de la relación comercial.

Escala del modelo a otros segmentos del portafolio MRO. El presente trabajo se limitó intencionalmente a seis artículos estratégicos de las categorías ISM e ICP, donde el costo de equivocarse es más alto. La metodología desarrollada, incluyendo el flujo de auditoría, la clasificación Syntetos-Boylan y el esquema de backtesting diferenciado por familia de demanda, está diseñada para ser replicable. Una extensión lógica consiste en aplicar el mismo enfoque a artículos de cuello de botella y, progresivamente, a segmentos más amplios del inventario, priorizando las referencias con mayor rotación de capital o mayor frecuencia de quiebres históricos. Esta escalabilidad era un objetivo implícito del diseño metodológico desde el inicio.

Ponderación geográfica de las variables climáticas. El indicador climático regional utilizado en este trabajo se construyó promediando en partes iguales los datos de temperatura y precipitación de los municipios de Samacá, Socha y Guachetá. Este promedio simple asume una contribución productiva equivalente de las tres zonas, lo que no necesariamente refleja la distribución real de la actividad minera. Una extensión metodológica consiste en ponderar los datos climáticos de cada zona por su porcentaje de producción sobre el total, construyendo un indicador regional más representativo del clima al que efectivamente están expuestos los activos. Este refinamiento podría mejorar la correlación entre la variable climática y el consumo de insumos, especialmente en los modelos SARIMAX donde esa relación es determinante para la utilidad de la exógena.

Estabilización de la varianza en las series de consumo continuo. Los diagnósticos de residuos de los modelos seleccionados para ISM10001 e ISM10006 muestran no-normalidad estadísticamente significativa. Aunque esta condición no invalida los pronósticos puntuales, sí afecta la precisión de los intervalos de confianza. Una extensión metodológica consiste en aplicar transformaciones de estabilización de la varianza previo al ajuste del modelo, y comparar mediante backtesting si la normalidad de los residuos mejora sin deteriorar el error de pronóstico fuera de muestra. De confirmarse una mejora, esta evidencia podría justificar la sustitución de los modelos actualmente seleccionados por versiones ajustadas sobre la serie transformada, consolidando así una selección más robusta antes de dar por definitiva la especificación SARIMAX o ETS para cada serie.

Ampliación del horizonte histórico y exploración de modelos adicionales. Las series ICP cuentan con entre 49 y 61 observaciones de vida activa, lo que restringe la capacidad de aprendizaje de modelos con mayor número de parámetros. A medida que se acumulen datos, será viable evaluar variantes como XGBoost o LightGBM para las series estables, cuyo potencial en series suaves con covariables está documentado en la literatura pero que con 84 observaciones presenta riesgo elevado de sobreajuste. Asimismo, series actualmente clasificadas como erráticas o irregulares con históricos cortos podrían beneficiarse de metodologías especializadas como ADIDA o iETS cuando dispongan de suficiente historia para una estimación robusta. En el mismo sentido, construir escenarios de sensibilidad sobre la TRM y las condiciones climáticas permitiría cuantificar el riesgo de inventario ante variaciones macroeconómicas o climáticas extremas, convirtiendo el sistema en una herramienta de análisis de riesgo además de pronóstico.

Referencias

Association for Supply Chain Management (ASCM). (n.d.). What Is Inventory Management? [Consultado en 2026]. https://www.ascm.org/topics/inventory-management/

Babai, M. Z., Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2013). Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1)/order-up-to replenishment systems. *International Journal of Production Economics*, *143*(2), 463-471. https://doi.org/10.1016/j.ijpe.2011.09.004

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *40*(6), 722-737. https://doi.org/10.1016/j.omega.2011.06.008

Barabadi, A., Barabady, J., & Markeset, T. (2014). Application of reliability models with covariates in spare part prediction and optimization – A case study. *Reliability Engineering & System Safety*, *123*, 1-7. https://doi.org/10.1016/j.ress.2013.09.012

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control* (5th). Wiley.

Boylan, J. E., Petropoulos, F., & Syntetos, A. A. (2022). Forecasting for inventory management of service parts. *International Journal of Production Economics*, *248*, 108481. https://doi.org/10.1016/j.ijpe.2022.108481

Boylan, J. E., & Syntetos, A. A. (2010). Intermittent demand forecasting: A review. *Journal of the Operational Research Society*, *61*(6), 925-937. https://doi.org/10.1057/jors.2009.129

Chartered Institute of Procurement & Supply (CIPS). (n.d.). Kraljic Matrix [Consultado en 2026]. https://www.cips.org/intelligence-hub/supplier-relationship-management/kraljic-matrix

C.I. Milpa S.A. (n.d.). En el corazón de la metalurgia [Consultado en 2026]. https://www.milpa.com.co/en_corazon.html

Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly*, *23*(3), 289-303. https://doi.org/10.1057/jors.1972.50

Dolgui, A., & Pashkevich, A. (2008). On the performance of binomial and beta-binomial models of demand forecasting for multiple slow-moving inventory items. *Computers & Operations Research*, *35*(3), 893-905. https://doi.org/10.1016/j.cor.2006.06.012

Driessen, M., Arts, J., Van Houtum, G. J., Rustenburg, J. W., & Huisman, B. (2015). Maintenance spare parts planning and control: A framework for control and agenda for future research. *Production Planning & Control*, *26*(5), 407-426. https://doi.org/10.1080/09537287.2014.907583

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987-1007. https://doi.org/10.2307/1912773

Esolbay. (2024). ¿Qué es la Matriz de Kraljic? [Consultado en abril de 2026]. https://www.esolbay.com

Gelderman, C. J., & Van Weele, A. J. (2005). Purchasing Strategies in the Kraljic Matrix—A Power and Dependence Perspective. *Journal of Purchasing and Supply Management*, *11*(2–3), 141-155. https://doi.org/10.1016/j.pursup.2005.10.004

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8ª ed.). Cengage Learning.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd) [Consultado en 2026]. OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer. https://doi.org/10.1007/978-3-540-71918-2

IBM. (n.d.). What Is Maintenance, Repair and Operations (MRO)? [Consultado en 2026]. https://www.ibm.com/think/topics/maintenance-repair-and-operations

Jarque, C. M., & Bera, A. K. (1980). Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters*, *6*(3), 255-259. https://doi.org/10.1016/0165-1765(80)90024-5

Johnston, F. R., & Boylan, J. E. (1996). Forecasting for Items with Intermittent Demand. *Journal of the Operational Research Society*, *47*(1), 113-121. https://doi.org/10.1057/jors.1996.10

Kaya, G. O., Sahin, M., & Demirel, O. F. (2020). Intermittent demand forecasting: a guideline for method selection [PDF]. *Sādhanā*, *45*(51). https://www.ias.ac.in/public/Volumes/sadh/045/00/0051.pdf

Kolassa, S., & Schütz, W. (2011). Why the MAD/Mean Ratio Is Not a Good Measure of Forecast Accuracy. *Foresight: The International Journal of Applied Forecasting*, *24*, 19-26.

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, *143*(1), 198-206. https://doi.org/10.1016/j.ijpe.2013.01.009

Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, *225*, 107597. https://doi.org/10.1016/j.ijpe.2019.107597

Kraljic, P. (1983). Purchasing Must Become Supply Management [Artículo clásico que introduce la matriz de Kraljic]. *Harvard Business Review*, *61*(5), 109-117. https://hbr.org/1983/09/purchasing-must-become-supply-management

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297-303. https://doi.org/10.1093/biomet/65.2.297

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54-74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6ª ed.). Wiley.

Nikolopoulos, K. I., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society*, *62*(3), 544-554. https://doi.org/10.1057/jors.2010.57

Pennings, C. L. P., van Dalen, J., & van der Laan, E. A. (2017). Exploiting returns in spare parts logistics: The value of knowing return volumes upfront. *International Journal of Production Economics*, *183*, 421-431. https://doi.org/10.1016/j.ijpe.2016.05.008

Rego, J. R., & Mesquita, M. A. (2011). Spare parts inventory control: a literature review [PDF]. *Production*. https://www.scielo.br/j/prod/a/kxpq3WLx4ycx5hN5TMpKnRD/?format=pdf&lang=en

Romeijnders, W., Teunter, R., & Van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, *220*(2), 386-393. https://doi.org/10.1016/j.ejor.2012.01.019

Sillanpää, V., & Liesiö, J. (2018). Forecasting spare part demand in a case company: Challenges and solutions. *International Journal of Production Economics*, *200*, 219-229. https://doi.org/10.1016/j.ijpe.2018.03.025

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72-101. https://doi.org/10.2307/1412159

Strijbosch, L. W. G., Heuts, R. M. J., & van der Schoot, E. H. M. (2000). A combined forecast-inventory control procedure for spare parts. *Journal of the Operational Research Society*, *51*(10), 1184-1192. https://doi.org/10.1057/palgrave.jors.2601013

Syntetos, A. A., Babai, M. Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1-26. https://doi.org/10.1016/j.ejor.2015.11.010

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495-503. https://doi.org/10.1057/palgrave.jors.2601841

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, *16*(4), 437-450. https://doi.org/10.1016/S0169-2070(00)00065-0

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*(3), 606-615. https://doi.org/10.1016/j.ejor.2011.05.018

Ulrich, M., Jahnke, H., Langrock, R., Pesch, R., & Senge, R. (2022). Classification-based model selection in retail demand forecasting. *International Journal of Forecasting*, *38*(1), 209-223. https://doi.org/10.1016/j.ijforecast.2021.05.010

Unidad de Planeación Minero Energética (UPME). (2024). *Anexo 5. Visitas estratégicas (Plan Nacional de Desarrollo Minero) — Visita a MILPA S.A.* (Documento en PDF). UPME. https://docs.upme.gov.co/SIMEC/simco/PlaneacionSector/Documents/Anexo_5_Visitas_estrategicas_PNDM.pdf

Wagner, S. M., & Bode, C. (2009). Supplier default dependencies: Empirical evidence from the automotive industry. *European Journal of Operational Research*, *199*(1), 150-161. https://doi.org/10.1016/j.ejor.2008.11.012

Zhang, S., Huang, K., & Yuan, Y. (2021). Spare Parts Inventory Management: A Literature Review. *Sustainability*, *13*(5), 2460. https://doi.org/10.3390/su13052460