



Ficha metodológica — Modelo Analítico de Riesgo Hídrico-Energético

Sustento técnico del tablero de riesgo de apagón · MIRHE · Actualizado 2026-07-09 · Cifras leídas en vivo de los modelos entrenados

1 · OBJETIVO

Estimar, con datos oficiales y actualización diaria, el **riesgo de desabastecimiento eléctrico (apagón/raционamiento)** del Sistema Interconectado Nacional, integrando la cadena sequía → embalses → demanda → respaldo térmico → gas (nacional e importado) → renovables. Objetivo único; toda la información converge en una sola decisión.

2 · FUENTES DE DATOS (OFICIALES)

XM / Sinergox (API pydataxm) · SIMEM PublicData (XM) · PARATEC (mapa de generación XM) · NOAA CPC (ONI / probabilidad ENOS) · IDEAM (boletines hidrológicos) · IPSE / SIGMA (telemetría ZNI) · BEC / BMC (balance de gas) · DANE / DIVIPOLA (georreferenciación).

3 · VARIABLES DEL MODELO

Variable	Definición	Fuente
aportes_pct	Aportes hídricos como % de la media histórica del SIN	XM/Sinergox · PorcApor
embalse_pct	Nivel de embalse agregado (volumen útil / capacidad útil)	XM · VoluUtilDiarEner / CapaUtilDiarEner
demanda_gwh	Demanda real diaria del SIN (GWh/día)	XM · DemaReal
margen_gen_dem_gwh	Margen generación – demanda (GWh/día)	XM · Gene – DemaReal
oni	Oceanic Niño Index (intensidad de El Niño/La Niña)	NOAA CPC · ONI v5
precio_ponderado	Precio de bolsa ponderado diario (COP/kWh)	SIMEM/XM · PPBO (96D56E)
precio_lag7	Precio de bolsa rezagado 7 días (autoregresivo)	Derivado de PPBO
aportes_mm30 / embalse_mm30	Medias móviles 30 días de aportes y embalse (tendencia hidrológica)	Derivado de XM
indisponibilidad_pct	Capacidad efectiva neta – disponibilidad real (proxy de mantenimientos/salidas)	XM · CapEfecNeta – DispoReal

4 · MODELOS MATEMÁTICOS ENTRENADOS

Algoritmos: regresión Ridge (lineal, interpretable) y Gradient Boosting (no lineal). Dataset: 1247 días (2023–2026).

Precio de bolsa — error de ± 35 COP/kWh MAE 35

Métrica principal: error absoluto medio fuera de muestra (COP/kWh), interpretable y no inflable.

Sobre el R^2 : el R^2 del modelo es 0.92, PERO el precio es tan persistente que la línea base trivial "mañana = hoy" ya da R^2 0.90 (MAE 36). Es decir, el R^2 alto viene casi todo de la autocorrelación, no del modelo: la **habilidad NETA** del modelo sobre esa línea base es marginal (MAE 36→35). Reportamos el MAE como titular precisamente

Estrés operativo — NO PREDECIBLE (hallazgo)

No predecible

Un modelo entrenado con fundamentales **no logra superar ni siquiera a la media histórica** — esa es la definición de un R^2 negativo (-1.27), que aquí NO es un "mal puntaje" sino la **evidencia** de que el estrés operativo por restricciones no depende de las series energéticas diarias. Se probó además el dato de

para no dar una falsa impresión de precisión. El valor del modelo no es "predecir el precio" sino explicar QUÉ fundamentales (hidrología, El Niño, gas) mueven el régimen.

mantenimientos (indisponibilidad XM) y tampoco mejora: la causa es estructural (topología de transmisión y decisiones de despacho, no observables en el dataset). Publicar este resultado —en vez de esconderlo o maquillar un R^2 positivo— es lo que da credibilidad al resto del modelo.

5 · VALIDACIÓN

Partición **temporal 80/20 sin fugas de información** (997 días de entrenamiento / 250 de prueba): se entrena con el pasado y se evalúa con el futuro inmediato, replicando el uso real. Métricas reportadas fuera de muestra (R^2 , RMSE, MAE). Los coeficientes Ridge se publican estandarizados para lectura directa de la influencia de cada variable. Un R^2 negativo se reporta tal cual como hallazgo, no se oculta.

6 · ÍNDICE INTEGRADO DE RIESGO (INSTRUMENTO DE DECISIÓN)

Dado que el apagón no es entrenable como evento único, la decisión se sustenta en un **índice integrado 0-100**, analítico y transparente: la probabilidad estadística de El Niño (validada LOOCV) es el mayor factor y el resto son señales físicas observadas puntuadas contra umbrales documentados (no pesos ajustados a mano). Cada factor declara valor, aporte y fuente. **Es la misma cifra que muestra el diagrama y todo el tablero**, recalculada a diario.

Índice actual	Nivel	P(estrés) El Niño (Capa A)	N.º de factores
72 / 100	Naranja	73.5%	17

Principales impulsores del índice (factor · valor · aporte de puntos):

Factor	Valor	Aporte
Probabilidad de estrés por El Niño (modelo validado)	74% · El Niño Advisory	16.2
Cobertura de gas firme de térmicas	43%	10.0
Firmeza de oferta de gas en pico El Niño	2%	8.0
Balance comercial de gas en firme	deficitario 35/35 TEE	8.0
Aportes hídricos vs media (30d)	73%	7.0

No es probabilidad calibrada de apagón (un solo racionamiento histórico, 1992): estima estrés del sistema, no apagón directo.

7 · LÍMITES DECLARADOS (TRANSPARENCIA)

- **Apagón directo no entrenable.** Solo hay un episodio de racionamiento moderno (1992); no hay muestra estadística para entrenar un clasificador de apagón. Se modela por componentes y escenarios, no como evento único.
- **Estrés operativo = límite estructural.** El estrés por restricciones (generación de seguridad/fuera de mérito/demanda no atendida) NO es predecible con fundamentales hidrológicos/demanda. Se probó además el dato de mantenimientos (indisponibilidad XM) y no mejora: lo determinan topología de transmisión y decisiones de despacho, no observables en series diarias.
- **Proyección de precio por escenarios.** La heurística original de proyección por escenarios se conserva de la última corrida documentada; no se reconstruyó con fórmulas inventadas. El modelo entrenado (ridge + boosting) la complementa con coeficientes aprendidos.
- **Faltantes BEC declarados.** Inventario de GNL en tanques y días de autonomía no tienen fuente pública (tablero BEC autenticado). Se declara como pendiente; no se aproxima.
- **Sin coordenadas inventadas.** Activos sin ficha oficial quedan como 'pendiente de georreferenciación', nunca con lat/lon estimada arbitraria.