

Dossier de Auditoría — MIRHE

MONITOR INTEGRAL DE RIESGO HÍDRICO-ENERGÉTICO

Ministerio de Minas y Energía · Grupo de Datos Estratégicos

Hoja de vida completa del sistema · 29 de July de 2026

Documento para revisión y evaluación técnica **previa a publicación**. Auto-generado desde el sistema real (inventario, hashes de integridad y métricas en vivo).

0. CÓMO LEER ESTE DOCUMENTO (PARA EL AUDITOR)

Este dossier está pensado para que usted pueda **verificar, no solo creer**. Cada afirmación cuantitativa proviene de un artefacto reproducible cuyo hash de integridad se lista en el §11. Donde el sistema tiene un límite, se declara explícitamente (§10). El principio rector del proyecto es **"sin humo"**: se reportan los hallazgos negativos (modelos que no funcionan) con el mismo énfasis que los positivos.

Índice: 1) Qué es y qué no es · 2) Arquitectura · 3) Fuentes de datos · 4) Los modelos · 5) El índice de riesgo · 6) Capas de gas · 7) Modelos del grupo · 8) Controles de calidad · 9) Reproducibilidad · 10) Límites declarados · 11) Inventario técnico con hashes.

1. QUÉ ES Y QUÉ NO ES

MIRHE ES un sistema analítico que integra, a diario y de forma automática, ~32844 líneas de código sobre **11 fuentes oficiales**, para producir una lectura integrada del riesgo de abastecimiento eléctrico ante el Fenómeno de El Niño. Su propósito es **dar soporte a la decisión ministerial y contrastar la información de XM** con modelo propio.

MIRHE NO ES un predictor de apagones. No existe —ni aquí ni en ningún lado— un modelo que "prediga apagones" en Colombia con validez estadística, porque **solo ha habido un racionamiento en la historia (1992)**: una sola observación no permite entrenar ni validar un clasificador. Cualquier herramienta que afirme lo contrario estaría sobreajustando o inventando. MIRHE mide **ESTRÉS del sistema** (tensión de suficiencia), no probabilidad calibrada de apagón, y lo declara en toda su documentación.

Estado actual del índice: 72/100 — nivel Naranja (Riesgo alto).

2. ARQUITECTURA DEL SISTEMA

Flujo de datos de extremo a extremo, todo automatizado en un pipeline diario (11:00 UTC):

FUENTES OFICIALES → INGESTA (bronze) → CURADO (gold) → MODELOS → SUPERFICIES

XM / SIMEM / NOAA → 120 scripts Python → 34 capas JSON/CSV → El Niño (logística) → Tablero web (13 pág)

IDEAM / BEC / PARATEC → con trazabilidad de → validadas y con → Precio (GB) → Boletines/informes PDF

IPSE / MME / UPME / CREG → fuente por dato → hash de integridad → Índice (scorecard) → Correo diario + vistas

- **Bronze:** datos crudos tal como llegan de la fuente (evidencia).
- **Gold:** datos curados, validados y versionados, cada uno con su fuente declarada.
- **Reproducible:** todo se regenera con un comando; nada se edita a mano.

3. FUENTES DE DATOS (TODAS OFICIALES Y PÚBLICAS)

Fuente	Qué aporta	Acceso
XM / Sinergox	Embalses, aportes, demanda, generación, ENFICC, restricciones, precio	API pública <code>pydataxm</code>
SIMEM	Precio de bolsa ponderado (histórico validado)	API pública
NOAA / CPC	Índice ONI, discusión ENSO, probabilidades de intensidad	Web pública
IDEAM	Contexto hidrometeorológico nacional	Boletines públicos
BEC / BMC	Gas: inyección, nominaciones, gas a térmicas, CDP, mercado secundario, BI completo	Portal del gestor (sesión del Ministerio)
MME	Balance de producción de gas, DPGN, GLP	Publicaciones oficiales
PARATEC / UPME	Inventario y coordenadas del parque generador, proyección demanda	Públicas
IPSE / SUI	Telemetría de Zonas No Interconectadas	Públicas

Fuente	Qué aporta	Acceso
CREG	Precio de escasez de activación, Cargo por Confiabilidad	Resoluciones

Cada variable del dataset maestro tiene su fuente y métrica de origen documentadas en el **diccionario de datos** (publicado en el tablero, descargable).

4. LOS MODELOS (CON SU VALIDACIÓN HONESTA)

4.1 Modelo de El Niño — VALIDADO (el sustento más fuerte del índice)

- **Tipo:** regresión logística sobre episodios históricos de El Niño.
- **Validación:** LOOCV (leave-one-out) — cada episodio se predice con un modelo que no lo vio.
- **Verificación de calibración:** punto de inflexión ONI ≈ 1.52 (coherente con la frontera Niño moderado→fuerte).
- **Salida hoy:** P(estrés) del episodio en Advisory = **rango 81.3% - 97.0%** (sigmoide vs no-paramétrico; se reporta el rango, no un punto falso-preciso).

4.2 Modelo de precio de bolsa — reportado por su ERROR, no por su R²

- **Métrica principal:** MAE 36 COP/kWh (error absoluto medio fuera de muestra).
- **Nota de honestidad clave:** el R² del modelo es 0.94, pero la línea base trivial "mañana = hoy" ya da R² 0.94. El precio de bolsa es tan persistente que ese R² alto viene de la autocorrelación, **no del modelo**. Por eso el titular es el MAE y no el R²: evita dar una falsa impresión de precisión. El valor del modelo es identificar QUÉ fundamentales mueven el precio, no predecirlo.

4.3 Estrés operativo — NO PREDECIBLE (hallazgo declarado, no error)

- Se intentó predecir el estrés operativo (generación de seguridad + fuera de mérito + demanda no atendida) con fundamentales.
- **Resultado:** el modelo NO supera ni al promedio histórico (R² -1.72, negativo). Esto **no es un modelo roto**: es la evidencia de que el estrés lo determinan la topología de transmisión y el despacho de seguridad, no las series energéticas diarias. Se probaron incluso los mantenimientos (indisponibilidad XM): no mejora.
- **Se publica el fracaso** en vez de esconderlo o maquillar un R² positivo. Esto es lo que da credibilidad al resto.

5. EL ÍNDICE DE RIESGO (COMPOSICIÓN Y SUSTENTO DE PESOS)

El índice es un **marcador compuesto transparente (scorecard)**, no un modelo con pesos entrenados contra apagones (imposible, §1). Cada factor declara su valor, su peso máximo, su aporte y su fuente:

A. Clima-hidrología — techo 40 pts

Factor	Peso máx	Valor hoy	Fuente
Probabilidad de estrés por El Niño (modelo validado)	22	81% · El Niño Advisory	Modelo estadístico episodios El Niño (logística, validado LOOCV)
Aportes hídricos vs media (30d)	10	81%	XM PorcApor
Embalse agregado del SIN	8	80%	XM VoluUtilDiarEner/CapaUtilDiarEner (Sistema)

B. Demanda — techo 12 pts

Factor	Peso máx	Valor hoy	Fuente
Crecimiento interanual de demanda	8	+6.0%	XM DemaReal
Cobertura de generación	4	102%	XM Gene/DemaReal

C. Gas-respaldo — techo 38 pts

Factor	Peso máx	Valor hoy	Fuente
Cobertura de gas firme de térmicas	10	41%	XM consumo vs BEC contratación firme
Firmeza de oferta de gas en pico El Niño	8	2%	PTDVF/CIDVF SEGAS
Balance comercial de gas en firme	8	deficitario 35/35 TEE	Balance TEN SEGAS
Dependencia de importación de gas (GNL)	6	30% (media 30d)	BEC Información Operativa — Energía Inyectada (importado/total)
Balance físico observado oferta–demanda de gas	6	superávit 83,423 MBTUD (media 30d)	BEC Información Operativa — inyectada vs tomada

D. Eventos operativos — techo 24 pts

Factor	Peso máx	Valor hoy	Fuente
Indisponibilidad de capacidad (mantenimientos y fallas)	5	4,352 MW fuera (20.0%)	XM CapEfecNeta vs DispoReal
Generación de seguridad (restricciones de confiabilidad)	5	34.4 GWh/día	XM GeneSeguridad
Generación fuera de mérito (restricciones de red)	4	15.7 GWh/día	XM GeneFueraMerito
Cercanía al precio de escasez	6	bolsa al 64.2% del techo de escasez	XM PrecEscaAct 1,221 COP/kWh
Demanda no atendida (30 días)	4	4,309 MWh · foco AREA CARIBE	XM DemaNoAten por área

E. Mitigación (renovables) — techo 8 pts

Factor	Peso máx	Valor hoy	Fuente
FN CER cubriendo demanda hoy	-5	9.5%	XM Gene solar/eólica
Pipeline renovable en pruebas	-3	1231 MW	PARATEC

Sustento de los pesos: el único "peso de modelo" es El Niño (salida de la logística validada). El resto son **cortes de severidad anclados al percentil del histórico observado** (2023-2026) y verificados contra el Niño 2023-24. La jerarquía entre grupos (clima 40 > gas 26 > eventos 18 > demanda 12) es juicio experto informado por la estructura del sistema (70% hidroeléctrico), declarado como tal. Detalle completo en el documento **Sustento de los pesos del índice** (anexo).

Verificación independiente: en paralelo opera un cálculo de **suficiencia física determinista** (ENFICC vs demanda) que NO depende de pesos. Hoy: energía firme 245.0 vs demanda 248.1 GWh/día.

6. CAPAS DE GAS (LA MAYOR PROFUNDIDAD DEL SISTEMA)

MIRHE integra ~20 archivos operativos del BEC + acceso completo al BI del gestor (mapa de gasoductos oficial con % de uso por tramo, balance Costa/Interior, demanda por sector, oferta por campo). Responde preguntas que el boletín de XM no cubre: cuánto gas nacional vs importado va a térmicas, de qué campos, por qué tramos, y el declive estructural de la oferta nacional (−40% en la década observada). Todo con fuente oficial y refresco diario.

7. INTEGRACIÓN CON LOS MODELOS DEL GRUPO

MIRHE se alimenta de y contrasta con los otros modelos de Datos Estratégicos:

- **ENOS + CVaR** (capa económica del riesgo) — auto-hospedado y sincronizado.
- **Prophet + XGBoost de volumen de embalses** — corriendo en la misma máquina; su senda proyecta el mínimo del embalse.

El cruce modelo↔observado detectó, por ejemplo, que el modelo ENOS subestima la demanda actual (+27 GWh/día) — un aporte concreto del contraste.

8. CONTROLES DE CALIDAD (LO QUE HACE AUDITABLE AL SISTEMA)

1. **Conciliación diaria XM↔MIRHE:** 2 indicadores OK · 5 con diferencia explicada (causa documentada) · **0 divergencias sin explicar**. Cuando MIRHE y XM difieren, la causa está pre-calculada.
2. **Scorecard vivo:** cada estimación publicada se verifica contra lo observado cuando llega el dato (registro auditable que nunca se reescribe); alerta automática si el modelo se degrada.
3. **Diccionario de datos + descargas:** cada capa con definición, fuente, rezago y CSV descargable, para que un especialista recalcule por su cuenta.
4. **Dos versiones del índice** (vigente vs propuesto con ENFICC) con puntos de recuperación git.

8 BIS. ASEGURAMIENTO DE LA CALIDAD DE LOS DATOS

La calidad de los datos se verifica **a diario y de forma medible**, en las 6 dimensiones estándar de calidad de datos + frescura. Veredicto actual: **APROBADO CON AVISOS** (6 OK · 1 avisos · 0 fallas).

Dimensión	Estado	Métrica (medida hoy)
✓ Completitud	OK	100.0% en la variable menos completa (de 5 clave)
✓ Validez (rangos físicos)	OK	0 valores fuera de rango plausible (de 6400 evaluados)
✓ Unicidad	OK	0 fechas duplicadas (debe ser 0: una fila por día)
✓ Consistencia temporal	OK	0 saltos > 7 días en la serie (2023-01-15 → 2026-07-23)

Dimensión	Estado	Métrica (medida hoy)
✓ Exactitud (vs fuentes externas)	OK	Conciliación XM: 0 divergencias sin explicar · 3 OK, 4 explicadas. Nowcast precio MAE 30d: 53.37 (vs persistencia 36.83)
✓ Integridad (SHA-256)	OK	6/6 capas clave con hash de integridad verificable
⚠ Frescura (actualización diaria)	AVISO	24 capas actualizadas en las últimas 30h · 2 atrasadas

Controles permanentes (de diseño):

- Serie de precio validada contra SIMEM (0 discrepancias en 29.712 filas verificadas).
- Embalse unificado de 4 fuentes a una sola serie observada (elimina el desfase base-masa vs energía).
- Guardas de plausibilidad: valores implausibles se marcan 'requiere validación', no se publican.
- Dataset incremental con deduplicación por fecha (la versión revisada gana en el traslape).
- Compuerta de sintaxis JS (node --check) antes de publicar cada página.
- Puntos de recuperación git antes de cualquier cambio al índice.
- Principio declarado: lo que no se puede verificar se marca como faltante, nunca se inventa.

Principio rector de calidad: un dato que no pasa validación **no se publica como cierto** — se declara su limitación ("requiere validación"), nunca se inventa. Este reporte se regenera cada día y es descargable.

9. REPRODUCIBILIDAD

- **Un comando regenera todo:** `python3 scripts/tuxilo_render_dashboard.py` ejecuta el pipeline completo (ingesta → curado → modelos → índice → superficies).
- **Versionado total:** 129 commits desde 2026-06-10 hasta 2026-07-09; 3 puntos de recuperación (git tags) antes de cambios sensibles al índice.
- **Integridad:** cada capa gold verificable por hash SHA-256 (§11).
- **Entorno declarado:** Python 3.12 · scikit-learn 1.9 · pandas 2.3 · numpy 2.4 · pydataxm.

10. LÍMITES DECLARADOS (LO QUE MIRHE NO HACE)

1. **No predice apagones** (1 evento histórico — §1). Mide estrés.
2. **El estrés operativo no es predecible** con fundamentales (§4.3) — declarado.
3. **Los pesos entre grupos son juicio experto**, no un óptimo estadístico (no existe la muestra).
4. **El estudio estocástico multi-serie** (VEREC/series críticas) es metodología propietaria de XM — no se replica; se usa la suficiencia física como equivalente honesto.
5. **Inventario de GNL en tanques** sin fuente pública — se usa el margen de regasificación como proxy, declarado.
6. **73 recursos térmicos de XM** sin municipio en la fuente oficial — se muestran en tabla, no se ubican artificialmente en el mapa.

11. INVENTARIO TÉCNICO (VERIFICABLE)

- **Código:** 120 scripts Python · ~32844 líneas · 21 módulos temáticos.
- **Capas de datos:** 34 artefactos gold (JSON/CSV) + dataset maestro diario.
- **Superficies:** 13 páginas web + boletines/informes PDF + correo diario + 3 vistas por dirección.
- **Historia:** 129 commits (2026-06-10 → 2026-07-09).

Hashes de integridad de las capas clave (SHA-256, primeros 16 hex — verificables con `sha256sum`):

Capa	Tamaño	SHA-256 (16)
<code>dataset_riesgo_diario.csv</code>	149 KB	<code>c3874d21c1194fd1</code>
<code>enficc_summary.json</code>	2 KB	<code>7dc1fe9a9846a088</code>
<code>modelo_entrenado.json</code>	4 KB	<code>853d6a88a4730024</code>
<code>gas_operativo_summary.json</code>	2 KB	<code>87f66004c0207173</code>
<code>conciliacion_summary.json</code>	3 KB	<code>d56eba1a2906c5ba</code>
<code>scorecard_registro.csv</code>	15 KB	<code>9c1f3d30e5474255</code>
<code>gas_esquema_summary.json</code>	3 KB	<code>73eac706c5b8ddf8</code>
<code>boletin_vars_summary.json</code>	4 KB	<code>761b39dbff9b15ab</code>

Capa	Tamaño	SHA-256 (16)
indice_historico.csv	1 KB	7da8af9d59139bcc

Puntos de recuperación (git tags): punto-recuperacion-2026-06-21-pre-gas-termico, punto-recuperacion-2026-07-02-pre-enficc, punto-recuperacion-2026-07-08-pre-conciliacion-xm

CIERRE

Este sistema fue construido bajo un estándar explícito de honestidad metodológica: **preferir un hallazgo negativo declarado a un número positivo maquillado**. Los tres ejemplos que mejor lo ilustran —el estrés no predecible (R^2 negativo publicado), el R^2 del precio reportado con su contexto de persistencia, y la negativa a fingir un "predictor de apagones"— no son debilidades del sistema: son la razón por la que sus cifras positivas son creíbles. Queda a disposición del equipo auditor para su análisis, con todos los artefactos abiertos y reproducibles.

Documento auto-generado por MIRHE desde el estado real del sistema. Se regenera con cada corrida; su contenido refleja el sistema al momento de la fecha del encabezado.