



Ministerio de Minas y Energía · República de Colombia

Expediente técnico — Modelo Analítico de Riesgo Hídrico-Energético

Sustento metodológico del tablero de riesgo de apagón (Tuxilo) · 2026-07-09

Índice integrado de riesgo vigente:

72 / 100 — Naranja

- **1.** Ficha metodológica (resumen de sustento)
- **2.** Modelo estadístico de estrés por El Niño (capa ejecutiva + técnica)
- **3.** Anexo de reproducibilidad (comandos, coeficientes, episodios, hashes SHA-256)
- **4.** Verificación de la sigmoide (respuesta a auditoría: inflexión, estacionalidad, calibración, rango 73-89%)
- **5.** Anexo visual (diagrama del modelo y gráfica del modelo de estrés)

Documento generado automáticamente desde los artefactos del modelo; cifras leídas en vivo. Reproducible con un comando.

Mapa del expediente — qué responde cada sección

Guía de navegación para revisión técnica / auditoría. Índice integrado vigente: 72/100 — Naranja.

Sección	Contenido	Responde a
1. Ficha metodológica	Objetivo, fuentes, variables, modelos, límites declarados.	Visión ejecutiva / qué es el tablero.
2. Modelo estadístico El Niño	25 episodios, muestra n=11, LOOCV, umbral ONI≈1.45, escenarios.	¿El modelo está validado? ¿con qué datos?
3. Anexo de reproducibilidad	Comandos, coeficientes β , tabla de episodios, hashes SHA-256.	¿Es reproducible? ¿de dónde salen las cifras? ¿lo genera un LLM?
4. Verificación de la sigmoide	Punto de inflexión (1.52), estacionalidad, calibración out-of-sample, método no-paramétrico y rango 73-89%.	Evaluación de la sigmoide (3 condiciones) y dilema univariado vs multivariado.
5. Anexo visual	Diagrama del modelo (marco rare-event) y gráfica del modelo de estrés.	¿Cómo funciona el modelo? ¿por qué no es una probabilidad entrenada?

Suficiencia física (ENFICC): la métrica oficial CREG de energía firme (energía firme vs demanda, margen y exposición a gas interrumpible) está incorporada al tablero en la pestaña «Riesgo de apagón»; es el backbone determinista del riesgo. Responde a la observación de si el modelo tiene en cuenta la ENFICC.

Todos los documentos se generan automáticamente desde los artefactos del modelo; cifras leídas en vivo. Existe punto de recuperación por versión (git tags).



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 R²: el R² del modelo es 0.92, PERO el precio es tan persistente que la línea base trivial "mañana = hoy" ya da R² 0.90 (MAE 36). Es decir, el R² alto viene casi todo de la autocorrelación, no

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 R² 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

del modelo: la **habilidad NETA** del modelo sobre esa línea base es marginal (MAE 36→35). Reportamos el MAE como titular precisamente 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.

dato de 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.

Documento generado automáticamente desde los artefactos del modelo (`modelo_entrenado.json` , `modelo_estres_operativo.json` , `riesgo_integrado.json`). Para PDF: imprimir esta página (Ctrl/Cmd+P → Guardar como PDF, tamaño A4). Acompaña al diagrama del modelo de la pestaña «Riesgo de apagón».

Modelo Estadístico de Estrés Energético por El Niño — Colombia

Generado: 2026-07-09T11:06:44.762375+00:00

Enfoque: inferencia estadística sobre episodios históricos reales (reemplaza el modelo analítico manual v1).

Reproducible: python3 scripts/tuxilo_modelo_estadistico_nino.py

CAPA EJECUTIVA (SIN JERGA)

Qué hicimos. En lugar de inventar una fórmula con pesos elegidos a mano, dejamos que la historia hable: tomamos todos los episodios de El Niño desde 1950 (según el indicador oficial ONI de la NOAA), miramos cuáles produjeron problemas eléctricos reales en Colombia, y ajustamos un modelo que aprende esa relación de los datos.

Qué encontramos. Desde 1991 (cuando existe registro confiable) hubo **11 episodios** de El Niño: **5 terminaron en estrés alto del sistema** (1992 con racionamiento; 1997-98, 2009-10, 2015-16 y 2023-24 con emergencia o precios extremos) y **6 pasaron sin evento**. La variable que mejor separa unos de otros es **la intensidad máxima del episodio (ONI pico)**: todos los episodios con pico ONI ≥ 1.6 generaron estrés, y ninguno con pico ≤ 1.3 lo hizo. En palabras simples: **El Niño débil no ha dañado el sistema; El Niño fuerte siempre lo ha estresado.**

Qué tan bien predice. Validamos a la antigua: escondimos cada episodio, entrenamos con el resto y le pedimos al modelo predecir el escondido (validación dejando-uno-fuera). Resultado: AUC 1.0, sensibilidad 80% y especificidad 100%. Con solo 11 episodios la incertidumbre es real y la decimos sin rodeos, pero la señal es fuerte y consistente.

Qué significa para hoy. Si el episodio que la NOAA ve formándose alcanza intensidad de Niño fuerte (pico ~ 1.8), la probabilidad histórica de estrés alto del sistema es **70%**; si se queda en débil (~ 0.9), es **14%**. La pregunta operativa deja de ser "¿llega El Niño?" y pasa a ser "¿qué tan fuerte llega?".

Qué NO puede decir este modelo. Solo ha habido **un racionamiento** (1992). Con un solo caso, ninguna estadística seria puede estimar la probabilidad de apagón propiamente dicho; estimamos estrés alto, que es el paso previo. Además, el desenlace depende de cómo esté preparado el sistema (embalses, térmicas, gas) y esas variables aún no están reconstruidas para los episodios viejos — es el siguiente paso.

CAPA TÉCNICA

1. Definición formal de la muestra

- Fuente:** serie ONI v5 NOAA/CPC, 1950-2025, trazable por URL en cada fila (NOAA CPC ONI v5 (serie local tmp/tuxilo/noaa/curated/oni.csv, 1950-2025)).
- Regla de episodio:** ONI ≥ 0.5 durante ≥ 5 temporadas consecutivas (regla estándar NOAA para evento El Niño).
- Universo:** 25 episodios entre 1950 y 2024.
- Muestra de inferencia (era moderna):** episodios que inician en 1991 o después — $n=11$. Los episodios pre-1991 carecen de registro auditable de desenlace eléctrico en este workspace y se excluyen del ajuste (etiqueta_confiable=false); las etiquetas previas del modelo v2 tipo "Histórico 1957-58" no tienen fuente trazable y fueron descartadas.

2. Tabla de episodios identificados

id	inicio	fin	dur (temp.)	ONI pico	ONI acum.	nivel estrés	confianza
1	1951-MJJ	1952-DJF	8	1.2	6.7	s/e	no_disponible
2	1953-JFM	1954-JFM	13	0.8	9.4	s/e	no_disponible
3	1957-MAM	1958-JJA	16	1.8	18.8	s/e	no_disponible
4	1958-OND	1959-FMA	5	0.6	2.8	s/e	no_disponible
5	1963-MJJ	1964-JFM	9	1.4	9.4	s/e	no_disponible
6	1965-AMJ	1966-MAM	12	2.0	15.9	s/e	no_disponible
7	1968-SON	1969-AMJ	8	1.1	6.7	s/e	no_disponible
8	1969-JAS	1970-DJF	6	0.9	4.1	s/e	no_disponible
9	1972-AMJ	1973-FMA	11	2.1	15.2	s/e	no_disponible
10	1976-ASO	1977-JFM	6	0.9	4.4	s/e	no_disponible

id	inicio	fin	dur (temp.)	ONI pico	ONI acum.	nivel estrés	confianza
11	1977-ASO	1978-DJF	5	0.8	3.6	s/e	no_disponible
12	1979-SON	1980-JFM	5	0.6	2.7	s/e	no_disponible
13	1982-MAM	1983-MJJ	15	2.2	20.5	s/e	no_disponible
14	1986-ASO	1988-JFM	18	1.7	20.5	s/e	no_disponible
15	1991-AMJ	1992-MJJ	14	1.7	14.4	2	alta
16	1994-ASO	1995-FMA	7	1.1	5.6	0	media
17	1997-AMJ	1998-AMJ	13	2.4	21.7	1	alta
18	2002-MJJ	2003-JFM	9	1.3	8.5	0	media
19	2004-JJA	2005-JFM	8	0.7	5.1	0	media
20	2006-ASO	2007-DJF	5	0.9	3.8	0	media
21	2009-JJA	2010-FMA	9	1.6	9.3	1	media
22	2014-SON	2016-AMJ	20	2.8	28.8	1	alta
23	2018-ASO	2019-MJJ	10	1.0	7.7	0	media
24	2019-OND	2020-FMA	5	0.7	3.0	0	media
25	2023-AMJ	2024-AMJ	13	2.1	17.5	1	alta

(nivel: 0=sin evento, 1=estrés alto, 2=racionamiento; s/e = sin etiqueta auditable)

3. Variable objetivo

Y=1 si el episodio produjo estrés alto o crisis documentada del sistema (nivel >= 1); Y=0 si no hay evento documentado. La crisis severa (racionamiento, nivel 2) solo ocurre una vez (1992) y NO es modelable por separado: se declara como limitación.

Las etiquetas, su justificación y su fuente están en `tuxilo/config/crisis_labels.json` con nivel de confianza explícito por episodio.

4. Variables explicativas candidatas

Derivadas exclusivamente de la serie ONI (disponibles y trazables para TODOS los episodios): `peak_oni`, `oni_promedio`, `oni_acumulado`, `duracion_temporadas`, `temporadas_hasta_pico`, `pico_en_seca` (pico durante la temporada seca colombiana dic-mar). Variables del sistema (embalses, gas, térmicas, precio) NO están disponibles para episodios previos a ~2000 y se excluyen para no inventar series (ver limitaciones).

5. Metodología y justificación

- **Regresión logística por máxima verosimilitud** (Newton-Raphson, implementación pura verificable en el script).
- Con 5 eventos positivos, la regla de eventos-por-variable (EPV≥5) limita honestamente a **1 predictor por modelo**: se ajusta un modelo univariado por candidato y se comparan. Métodos más complejos (bosques, boosting, bayesiano jerárquico) están injustificados con n=11 y se descartan explícitamente — serían humo.
- Predictores estandarizados (z-score). Si hay **separación perfecta** el MLE diverge; se aplica penalización ridge mínima (λ=0.5) y se informa — la separación misma es un hallazgo, no un estorbo.
- **Validación leave-one-out** (la única honesta con n=11): cada probabilidad reportada proviene de un modelo que nunca vio ese episodio.

6. Modelo base propuesto y comparación de candidatos

predictor	coef (β estand.)	AUC in-sample	AUC LOOCV	Brier LOOCV	separación
peak_oni	2.117	1.0	1.0	0.0765	sí
oni_promedio	2.012	1.0	0.933	0.0996	sí
oni_acumulado	1.985	1.0	0.933	0.1056	sí
duracion_temporadas	5.098	0.95	0.667	0.1921	no
temporadas_hasta_pico	97.656	0.983	0.933	0.1818	no
pico_en_seca	1.086	0.733	0.533	0.2882	no

Modelo base: logística sobre `peak_oni` (mejor AUC LOOCV). Coeficiente estandarizado β=2.117 (ridge λ=0.5, con separación perfecta: el dato sustenta un umbral, no un coeficiente puntual).

Umbral empírico: ningún episodio con peak_oni ≤ 1.3 produjo estrés; todos con ≥ 1.6 lo produjeron. Punto medio de decisión: 1.45.

7. Desempeño (LOOCV, umbral 0.5)

- AUC: **1.0** (in-sample 1.0)
- Sensibilidad (recall): **0.8** — VP=4, FN=1
- Especificidad: **1.0** — VN=6, FP=0
- Precisión: **1.0** · Exactitud: **0.909** · Brier: **0.0765**
- Interpretación de errores:** revisar probs_loocv en estadistico_modelo.json episodio por episodio; los fallos (si los hay) se concentran en episodios de intensidad intermedia (zona ONI 1.0-1.5), exactamente donde la teoría predice ambigüedad.

8. Proyección por escenario (episodio en formación)

Escenario de intensidad	P(estrés alto) histórica
Nino debil (peak 0,9)	13,6%
Nino moderado (peak 1,3)	34,3%
Nino fuerte (peak 1,8)	70,0%
Nino muy fuerte (peak 2,4)	93,4%

8b. Pronóstico operativo del episodio en formación (NOAA × modelo)

Cruzando la distribución de intensidad que NOAA publica para la temporada pico (OND, emisión 11 June 2026) con la curva P(estrés | intensidad) ajustada:

P(estrés alto del sistema) = 73.5% (cota conservadora)

Bucket NOAA	P(bucket)	Intensidad representante (mediana histórica)	P(estrés \ intensidad)
0.5°C ≤ Index < 1.0°C	2%	0.8	10%
1.0°C ≤ Index < 1.5°C	9%	1.15	25%
1.5°C ≤ Index < 2.0°C	27%	1.7	63%
Index ≥ 2.0°C	62%	2.15	87%

Supuestos declarados:

- Los buckets NOAA usan RONI; el modelo se ajustó con ONI. Para intensidades comparables la diferencia es pequeña y se declara.
- Se usa la distribución de la temporada pico esperada como proxy de la distribución del pico del episodio; como el pico del episodio es $\geq$ al valor de cualquier temporada, la P(estrés) resultante es conservadora (cota inferior).
- Representante por bucket = mediana de los picos históricos observados en ese bucket (1950-2024), no un valor elegido a mano.
- P(estrés | sin episodio, ONI<0.5) = 0 por construcción (el modelo es condicional a que exista episodio).

9. Comparación contra el enfoque manual anterior

Aspecto	Modelo manual v1 (6 factores)	Modelo estadístico
Origen de pesos	Elegidos a mano y "calibrados" con 4 anclas inventadas	Estimados por MLE desde episodios reales
Falsabilidad	No evaluable contra historia (sus insumos no existen hacia atrás)	LOOCV sobre 11 episodios reales
Auditabilidad	Baja (¿por qué wH=4.15?)	Total: datos, regla, código y métricas publicados
Riesgo	Falsa precisión (decimales sin sustento)	Incertidumbre declarada (n pequeño, IC amplios)
Qué se conserva	Su salida mensual sigue alimentando el dashboard hasta integrar este modelo	—

Qué mejora: trazabilidad completa, validación real, y una conclusión defendible (la intensidad pico manda). **Qué persiste:** n pequeño, una sola crisis severa, y sin covariables del sistema para episodios antiguos.

10. Limitaciones (explícitas)

1. **n=11 episodios, 5 positivos**: los intervalos de confianza son amplios; el AUC LOOCV puede moverse con un solo episodio nuevo.
2. **Racionamiento (nivel 2) no modelable**: un único caso (1992). Cualquier "probabilidad de apagón" derivada solo de esta muestra sería estadísticamente indefendible.
3. **Sin covariables del sistema** (embalse inicial, margen térmico, dependencia/oferta de gas, precio): no existen series trazables en el workspace para episodios pre-2000. NO se inventaron. Esto impide explicar *por qué* 1997-98 resistió y 1992 no (la hipótesis documental: expansión térmica post-reforma) con datos.
4. **Etiquetas de severidad intermedia** (2009-10, confianza media) podrían discutirse; el JSON de etiquetas permite auditarlas y re-correr todo.
5. La separación Gas vs Mixto del sistema actual no aplica retroactivamente y no se mezcló.

Siguientes pasos para fortalecer el modelo

1. **Reconstruir covariables del sistema 2000-2024 vía pydataxm/XM** (embalse agregado al inicio del episodio, margen térmico, precio): habilitaría un segundo predictor legítimo para los 7 episodios modernos con datos.
2. Formalizar las etiquetas con actos administrativos (resoluciones MinMinas/CREG) como fuente primaria por episodio.
3. Modelo de actualización bayesiana simple: prior = esta logística; actualización con las señales NOAA del episodio en curso (probabilidades de intensidad ya ingestadas en `enso_strength_probabilities.csv`).
4. Integrar P(estrés) de este modelo al Índice de Riesgo del dashboard, reemplazando la componente climática manual.
5. Revisión por par externo (UPME/XM) de las etiquetas y la regla de episodio.

Anexo técnico de reproducibilidad — Índice de Riesgo de Apagón (Tuxilo)

Generado: 2026-07-09 11:14 UTC

Generador: `python3 scripts/tuxilo_anexo_reproducibilidad.py` (lee los artefactos en vivo; no contiene valores escritos a mano)

Índice vigente: 72 / 100 — Naranja (Riesgo alto)

0. RESUMEN DE PROCEDENCIA

El índice **NO lo produce un modelo de lenguaje (LLM)**: lo calcula código Python determinista. Hay dos objetos distintos:

1. **Capa A — regresión logística validada** (`scripts/tuxilo_modelo_estadistico_nino.py`):

estima por máxima verosimilitud $P(\text{estrés alto del sistema} \mid \text{episodio El Niño})$ a partir de episodios históricos reales. **Sus coeficientes se re-estiman en cada corrida** desde los datos (no están fijos en un archivo de parámetros) y se persisten en `estadistico_modelo.json`.

2. **Índice integrado (scorecard)** (`tuxilo/sintesis/riesgo_integrado.py`): suma con **umbrales fijos declarados en el código** la salida de la Capa A (mayor factor) y señales físicas observadas (gas, embalses, demanda, eventos). Es la cifra del tablero/diagrama.

1. CÓMO REPRODUCIRLO (COMANDOS EXACTOS)

Requiere Python 3.12 (la Capa A es Python puro estándar; el resto usa `pandas`).

```
cd ~/.openclaw/workspace

# (A) Capa A: regresión logística validada – coeficientes, LOOCV, AUC
python3 scripts/tuxilo_modelo_estadistico_nino.py
# -> tmp/tuxilo/modelo_apagon/curated/estadistico_modelo.json
# -> tmp/tuxilo/modelo_apagon/curated/MODELO_ESTADISTICO_NINO.md (reporte)

# (B) Capa prospectiva (probabilidad/horizontes por escenario NOAA)
python3 scripts/tuxilo_risk_index.py
# -> tmp/tuxilo/modelo_apagon/curated/indice_riesgo_dashboard.json

# (C) Índice integrado (scorecard) – la cifra del tablero
python3 scripts/tuxilo_sintesis.py
# -> tmp/tuxilo/modelo_apagon/curated/riesgo_integrado.json
```

Determinismo: no hay aleatoriedad (Newton-Raphson determinista, sin `random_state` ni muestreo). Con los **mismos archivos de entrada**, la salida es **bit-idéntica** salvo el campo cosmético `generated_at`.

2. CAPA A — REGRESIÓN LOGÍSTICA (COEFICIENTES EN VIVO)

- **Predictor seleccionado:** `peak_oni` (mejor AUC LOOCV entre los candidatos).
- **Coefficientes** (logística estandarizada, leídos de `estadistico_modelo.json` → `modelos_univariados.peak_oni`):
- Intercepto $\beta_0 = -0.103$
- Coeficiente $\beta_1 = 2.117$ (sobre $z = (x - \text{media})/\text{sd}$)
- Estandarización: **media = 1.482, sd = 0.707**
- Penalización **ridge $\lambda = 0.5$** (aplicada por separación perfecta)
- **Ecuación:** $P(\text{estrés}) = 1 / (1 + e^{-(\beta_0 + \beta_1 \cdot z)})$, con $z = (\text{peak_oni} - \text{media})/\text{sd}$.

- **Separación empírica observada:** ningún episodio con $\text{peak_oni} \leq 1.3$ produjo estrés; todos con ≥ 1.6 lo produjeron → umbral medio **1.45**. (Con separación perfecta, el objeto defendible es este umbral, no un coeficiente puntual.)

Validación (LOOCV, leave-one-out, umbral 0.5)

- AUC LOOCV: **1.0** (in-sample 1.0)
- Sensibilidad: **0.8** · Especificidad: **1.0**
- Precisión: **1.0** · Exactitud: **0.909** · Brier: **0.0765**

3. MUESTRA DE ENTRENAMIENTO

- **Universo detectado (1950–2024):** 25 episodios (regla NOAA: $\text{ONI} \geq 0.5$ durante ≥ 5 temporadas consecutivas).
- **Usados para entrenar:** 11 episodios (era ≥ 1991 , con etiqueta auditable) — **5 positivos / 6 negativos**.
- Los $25 - 11 = 14$ episodios pre-1991 se excluyen (sin desenlace eléctrico auditable; `etiqueta_confiable=false`).
- **Variable objetivo:** $Y=1$ si nivel de estrés ≥ 1 . El racionamiento (nivel 2) ocurre 1 sola vez (1992) → no modelable por separado (limitación declarada).

inicio	fin	ONI pico	nivel	etiqueta (Y)	evento
1991-AMJ	1992-MJJ	1.7	2	estrés alto	Racionamiento eléctrico nacional marzo 1992 -
1994-ASO	1995-FMA	1.1	0	sin evento	Sin evento documentado.
1997-AMJ	1998-AMJ	2.4	1	estrés alto	El Niño 1997-98 (el más intenso del siglo XX h
2002-MJJ	2003-JFM	1.3	0	sin evento	Sin evento documentado.
2004-JJA	2005-JFM	0.7	0	sin evento	Sin evento documentado.
2006-ASO	2007-DJF	0.9	0	sin evento	Sin evento documentado.
2009-JJA	2010-FMA	1.6	1	estrés alto	El Niño 2009-10: alerta sectorial, embalses ba
2014-SON	2016-AMJ	2.8	1	estrés alto	El Niño 2014-16 (ONI máximo histórico 2.8): em
2018-ASO	2019-MJJ	1.0	0	sin evento	Sin evento documentado.
2019-OND	2020-FMA	0.7	0	sin evento	Sin evento documentado.
2023-AMJ	2024-AMJ	2.1	1	estrés alto	El Niño 2023-24: embalses en mínimos de una dé

(nivel: 0=sin evento, 1=estrés alto, 2=racionamiento) · Etiquetas en `tuxilo/config/crisis_labels.json`.

4. PRONÓSTICO OPERATIVO DEL EPISODIO EN FORMACIÓN

- **P(estrés alto) = 73.5%** (salida directa del modelo, ponderada sobre la distribución de intensidad NOAA de la temporada pico OND).
- Estado NOAA: **El Niño Advisory** (emisión 11 June 2026).
- Fuente distribución: NOAA CPC RONI strength outlook (`enso_strength_probabilities.csv`).

5. ÍNDICE INTEGRADO (SCORECARD) — DESCOMPOSICIÓN

Valor: **72 / 100** — **Naranja**. Suma de aportes (umbrales fijos en `riesgo_integrado.py`); P(estrés) El Niño es el mayor factor.

Grupo	Factor	Valor	Aporte (pts)	Má x	Fuente
A. Clima-hidrología	Probabilidad de estrés por El Niño (modelo validado)	74% · El Niño Advisory	+16.2	22	Modelo estadístico episodios El Niño (logística, validado LOOCV)
A. Clima-hidrología	Aportes hídricos vs media (30d)	73%	+7	10	XM PorcApor
A. Clima-hidrología	Embalse agregado del SIN	77%	+0	8	XM VoluUtilDiarEner/CapaUtilDiarEner (Sistema)
B. Demanda	Crecimiento interanual de demanda	+6.0%	+5	8	XM DemaReal
B. Demanda	Cobertura de generación	101%	+1	4	XM Gene/DemaReal

Grupo	Factor	Valor	Aporte (pts)	Má x	Fuente
C. Gas-respaldo	Cobertura de gas firme de térmicas	43%	+10	10	XM consumo vs BEC contratación firme
C. Gas-respaldo	Firmeza de oferta de gas en pico El Niño	2%	+8	8	PTDVF/CIDVF SEGAS
C. Gas-respaldo	Balance comercial de gas en firme	deficitario 35/35 TEE	+8	8	Balance TEN SEGAS
C. Gas-respaldo	Dependencia de importación de gas (GNL)	30% (media 30d)	+6	6	BEC Información Operativa — Energía Inyectada (importado/total)
C. Gas-respaldo	Balance físico observado oferta—demanda de gas	superávit 83,423 MBTUD (media 30d)	+0	6	BEC Información Operativa — inyectada vs tomada
D. Eventos operativos	Indisponibilidad de capacidad (mantenimientos y fallas)	4,805 MW fuera (22.1%)	+2.5	5	XM CapEfecNeta vs DispoReal
D. Eventos operativos	Generación de seguridad (restricciones de confiabilidad)	28.5 GWh/día	+5.0	5	XM GeneSeguridad
D. Eventos operativos	Generación fuera de mérito (restricciones de red)	29.1 GWh/día	+4.0	4	XM GeneFueraMerito
D. Eventos operativos	Cercanía al precio de escasez	bolsa al 62.5% del techo de escasez	+3.0	6	XM PrecEscaAct 1,221 COP/kWh
D. Eventos operativos	Demanda no atendida (30 días)	4,057 MWh · foco AREA CARIBE	+2.0	4	XM DemaNoAten por área
E. Mitigación (renovables)	FNCER cubriendo demanda hoy	9.0%	-3.8	-5	XM Gene solar/eólica
E. Mitigación (renovables)	Pipeline renovable en pruebas	1213 MW	-1.8	-3	PARATEC

6. VERIFICACIÓN DE INTEGRIDAD (SHA-256)

Código (método)

Archivo	Bytes	SHA-256
scripts/tuxilo_modelo_estadistico_nino.py	35286	769c77f33fe649803c0ae6e567df021ad2a1c9d57ee21710284bcac9e0a7d6f8
tuxilo/sintesis/riesgo_integrado.py	16521	f0e0dd747bf841deb1c5536f9bc9ee7afe0d6426807341c3cd5d0b3412715ff8
scripts/tuxilo_sintesis.py	521	d3991306779f7542912576c013deb1658b3ed506d3039aebe326314b92b693d7
scripts/tuxilo_risk_index.py	18931	96b6c75c991320b1aa33bdf9a96391b082b82cf87338a39b79e99e464cb664ea

Entradas (datos)

Archivo	Bytes	SHA-256
tmp/tuxilo/noaa/curated/oni.csv	90362	86ab6cd06f88fa657e536d9bd979807d94ed63c32f8e1c7cb07cc8468ba39f2d
tmp/tuxilo/noaa/curated/enso_strength_probabilities.csv	9791	d6ebf259ab3d7ba61460885fdfe3399f9fe177b5f387d780401ee73a8bcf869a
tmp/tuxilo/noaa/curated/enso_discussion_latest.json	708	df5fd965309a3c0d296cb8bdc97cf0143e45dffbe91ba7e9902edac6bcbd5d02
tuxilo/config/crisis_labels.json	3461	df3dc57204a2949f638c69bff169c6c446b63c11c3343c47143e0d64b067bca9
tmp/tuxilo/modelo_apagon/curated/data_snapshot.json	1013	8037800af55661a3965ab7876dc63af8d4031a4df3751dab6cee9a7f06872930

Archivo	Bytes	SHA-256
tmp/tuxilo/aportes/curated/aportes_summary.json	1330	07ea599a2af8b43d648ced9e9fc889f8b499114e1d39a123a95a65dcaca700ea
tmp/tuxilo/embalses/curated/embalses_master_summary.json	226	40cfc580132c79e14f1e851c968139a368b74b7480a816bc5da6e877d7bc5b94
tmp/tuxilo/demanda/curated/demanda_summary.json	600	017ff15a280b2d042c3c9310dbe8ac1a371dab0042d12ba6464c8c176d2a5ba8
tmp/tuxilo/gas_security/curated/gas_security_summary.json	4122	1f1064c2242d4d6994366a3c0153e6488e8e2d92f9835ca6ec3de930ec8a91a5
data/gold/gas_firmeza_summary.json	4359	8add09a53574bdfda178ada656be6f34cecc9e39707e7f83e44c13d21ddbffe6
data/gold/gas_balance_summary.json	12582	412f17b14b428490cf8a855c7a98941dd00daf666df1cc8199beb403d66d1005
tmp/tuxilo/eventos/curated/eventos_criticos_summary.json	1801	26fff2a09264ce7221054dc2b37eac3ed6fc4041f72a92a6fe736d5a2ae8a259
tmp/tuxilo/fncer/curated/fncer_summary.json	2469	bdfd092e0a062b48ee76bdfa8f0c7360e7de0773167a89f5b43e908ee008e179

Salidas (artefactos reproducibles)

Archivo	Bytes	SHA-256
tmp/tuxilo/modelo_apagon/curated/estadistico_modelo.json	22738	c481c25d715b6b272a4f937d0f0766e13aa2f1881740eda34234feeb26313a1a
tmp/tuxilo/modelo_apagon/curated/riesgo_integrado.json	6348	55073e22b1f7c747dd09bc1cecf86545f01373bfe3b1c8738db2a02b8ce6bb16
tmp/tuxilo/modelo_apagon/curated/indice_riesgo_dashboard.json	16338	25073fc6f8905c8c41f61e3e1d9aaf82e97667a9cecf9d83d9de6ec2fdb0fbd3

Para verificar: sha256sum <archivo> debe coincidir con el valor de la tabla.

Fin del anexo. Cualquier número aquí proviene de los artefactos en disco; este documento se regenera con un comando y es auditable extremo a extremo.

Verificación de la función sigmoide — Modelo de riesgo por El Niño

Para: Auditoría técnica · Fecha: 2026-07-02

Objeto: Responder a las tres condiciones planteadas para aceptar el uso de la sigmoide (punto de inflexión interpretable, control de estacionalidad y calibración out-of-sample).

Reproducibile: python3 scripts/tuxilo_modelo_estadistico_nino.py → bloque verificacion_sigmoide en estadistico_modelo.json .

0. ACLARACIÓN PREVIA: DÓNDE VIVE LA SIGMOIDE

La sigmoide validada es **episódica y univariada**, no la forma multivariada diaria:

Forma efectivamente ajustada:

$$P(\text{estrés}) = 1 / (1 + e^{-z}), \text{ con } z = -0.103 + 2.117 \cdot ((\text{ONI_pico} - 1.482) / 0.707)$$

- Una observación por episodio de El Niño (muestra moderna 1991-2024, **n = 11**, 5 positivos / 6 negativos).
- Es univariada por diseño. Con 5 eventos positivos, la regla eventos-por-variable ($EPV \geq 5$) limita honestamente a **1 predictor**. La forma multivariada $z = \beta_0 + \beta_1 \cdot \text{ONI} + \beta_2 \cdot \text{Volumen} + \beta_3 \cdot \text{brecha} + \beta_4 \cdot \text{estacionalidad} + \beta_5 \cdot \text{rezagos}$ con $n = 11$ **sobreajustaría** — precisamente la "probabilidad formalmente válida pero operativamente engañosa" que la auditoría advierte. Se descarta explícitamente.
- El índice integrado (semáforo) **NO es la sigmoide**: usa su salida ($P \approx 74\%$) como **un** factor. La sigmoide vive solo en la Capa A estadística.
- La forma multivariada con volumen/brecha/estacionalidad/rezagos **sí existe** en el modelo, pero para **precio de bolsa** (gradient boosting, R^2 0.90), no como probabilidad de apagón (el apagón tiene un solo caso histórico y no es calibrable).

1. PUNTO DE INFLEXIÓN FÍSICAMENTE INTERPRETABLE — CUMPLE

- Punto de inflexión ($P = 0.5, z = 0$): **ONI pico = 1.52**.
- Coincide con el **umbral empírico de separación = 1.45** (máx. sin estrés $1.3 <$ mín. con estrés 1.6).
- Lectura física: El Niño **débil** ($\text{ONI} \leq 1.3$) **nunca** ha estresado el sistema; El Niño **fuerte** ($\text{ONI} \geq 1.6$) **siempre** lo hizo. La inflexión en ~ 1.5 es exactamente la frontera **Niño moderado → fuerte**. Es interpretable y operativamente accionable.
- Caveat declarado: existe **separación perfecta** en los datos, por lo que el estimador de máxima verosimilitud diverge; se aplica penalización ridge ($\lambda = 0.5$). El objeto robusto es el **umbral (~ 1.45 -1.52 ONI pico)**, no la pendiente puntual de la curva.

2. CONTROL EXPLÍCITO DE ESTACIONALIDAD — CUMPLE (POR NIVEL)

- Modelo episódico (anual): no lleva estacionalidad calendario (hay un dato por episodio), pero **sí controla la coincidencia con la temporada seca** colombiana mediante la variable candidata `pico_en_seca` (si el pico de ONI cae en dic-mar) y `temporadas_hasta_pico`.
- Modelos diarios (precio de bolsa): incluyen estacionalidad **explícita** — verificado: `mes_sin`, `mes_cos` y bandera `seca` (dic-mar).

3. CALIBRACIÓN PROBABILÍSTICA FUERA DE MUESTRA — PARCIAL, DECLARADA

Validación **leave-one-out (LOOCV)**: cada probabilidad proviene de un modelo que **no** vio ese episodio.

Rango de probabilidad	n	Prob. predicha (prom.)	Frecuencia observada
0.0 – 0.2	3	0.11	0.00
0.2 – 0.5	4	0.33	0.25
0.5 – 0.8	1	0.55	1.00
0.8 – 1.0	3	0.91	1.00

- Métricas **LOOCV**: AUC 1.0 · Brier **0.0765** · sensibilidad 0.80 · especificidad 1.00.
- Lectura honesta: la calibración es **razonable en los extremos** (predicho 0.11 ↔ observado 0.00; predicho 0.91 ↔ observado 1.00), pero el **tramo medio (0.5-0.8) tiene un solo punto** → subpotenciado.

- **Límite declarado:** con **n = 11 y separación perfecta**, la calibración fina (curva de confiabilidad, Hosmer-Lemeshow) no es concluyente. Por eso el modelo **reporta ESTRÉS del sistema, no una probabilidad calibrada de apagón** — y así está rotulado en todo el tablero. El Brier bajo refleja buena **discriminación**, no calibración fina.

3B. MÉTODO NO-PARAMÉTRICO Y RANGO (BLINDAJE CONTRA SOBREAJUSTE)

Para atender el dilema de fondo —con ~1 evento severo, univariado subajusta y multivariado sobreajusta— la probabilidad de estrés se calcula **también sin ajustar parámetros**, usando la **frecuencia histórica observada por bucket de intensidad**, y se reporta el **rango** entre ambos métodos:

Bucket de intensidad (ONI)	P_NOAA	Cond. sigmoide	Cond. empírica (n)
0.5 - 1.0	0.02	0.11	0.00 (n=3)
1.0 - 1.5	0.09	0.25	0.00 (n=3)
1.5 - 2.0	0.27	0.63	1.00 (n=2)
≥ 2.0	0.62	0.87	1.00 (n=3)

P(estrés) = Σ P_NOAA(bucket) × P(estrés | bucket):

- Sigmoide (interpola el condicional): **73.5 %**
- No-paramétrica (frecuencia empírica): **89.0 %**
- **Rango reportado: 73.5 % - 89.0 %**

Por qué esto zanja el problema:

- La versión no-paramétrica **no ajusta ningún parámetro** → no puede sobreajustar. Refleja el hecho histórico crudo: todo El Niño con pico ONI ≥ 1.5 estresó el sistema; ninguno con pico < 1.5 lo hizo.
- El **peso probabilístico proviene del pronóstico de intensidad de NOAA** (calibrado externamente por NOAA), **no** de ajustar una curva a 11 puntos. Lo único tomado del histórico es el **umbral/condicional**.
- El **rango 73.5-89 %** es la comunicación honesta de la incertidumbre: no se ofrece una precisión falsa.

4. MARCO METODOLÓGICO (EVENTOS RAROS)

Con **un solo racionamiento histórico (1992)** no es estadísticamente posible **entrenar** una probabilidad de apagón. El riesgo, por tanto, **no** se sustenta en un modelo entrenado, sino en tres pilares:

1. **Suficiencia física determinista (backbone):** energía firme **ENFICC vs demanda** (margen) y gas firme/interrumpible a térmicas. Es un **balance de ingeniería, no un ajuste** → no puede sobreajustar.
2. **Clima El Niño:** pronóstico de intensidad de **NOAA (calibrado) × umbral histórico (ONI ≈ 1.45)**, con el rango 73.5-89 %.
3. **Escenarios condicionales** ("qué tendría que pasar para un apagón").

La sigmoide **solo interpola** el condicional entre intensidades; **no es "el predictor"**. Se reporta **estrés del sistema**, no una probabilidad calibrada de apagón.

5. CONCLUSIÓN

Las tres condiciones se atienden, con **transparencia sobre los límites**:

Condición	Estado	Evidencia
Punto de inflexión interpretable	☐	ONI pico 1.52 = frontera Niño moderado→fuerte
Control de estacionalidad	☐	pico_en_seca (episódico) + mes_sin/cos/seca (diario)
Calibración out-of-sample	⚠ declarada	LOOCV Brier 0.0765; calibración fina limitada por n=11 → se reporta estrés, no apagón

El diseño está construido para **no** producir una probabilidad engañosamente precisa: se declara el umbral (no la pendiente), se reporta el **rango 73.5-89 %** (sigmoide vs empírico), se apoya el riesgo en la **suficiencia física (ENFICC)** que no se ajusta, y se limita a un predictor por respeto al tamaño muestral. Toda cifra es reproducible desde `estadistico_modelo.json` (python3 scripts/tuxilo_modelo_estadistico_nino.py).

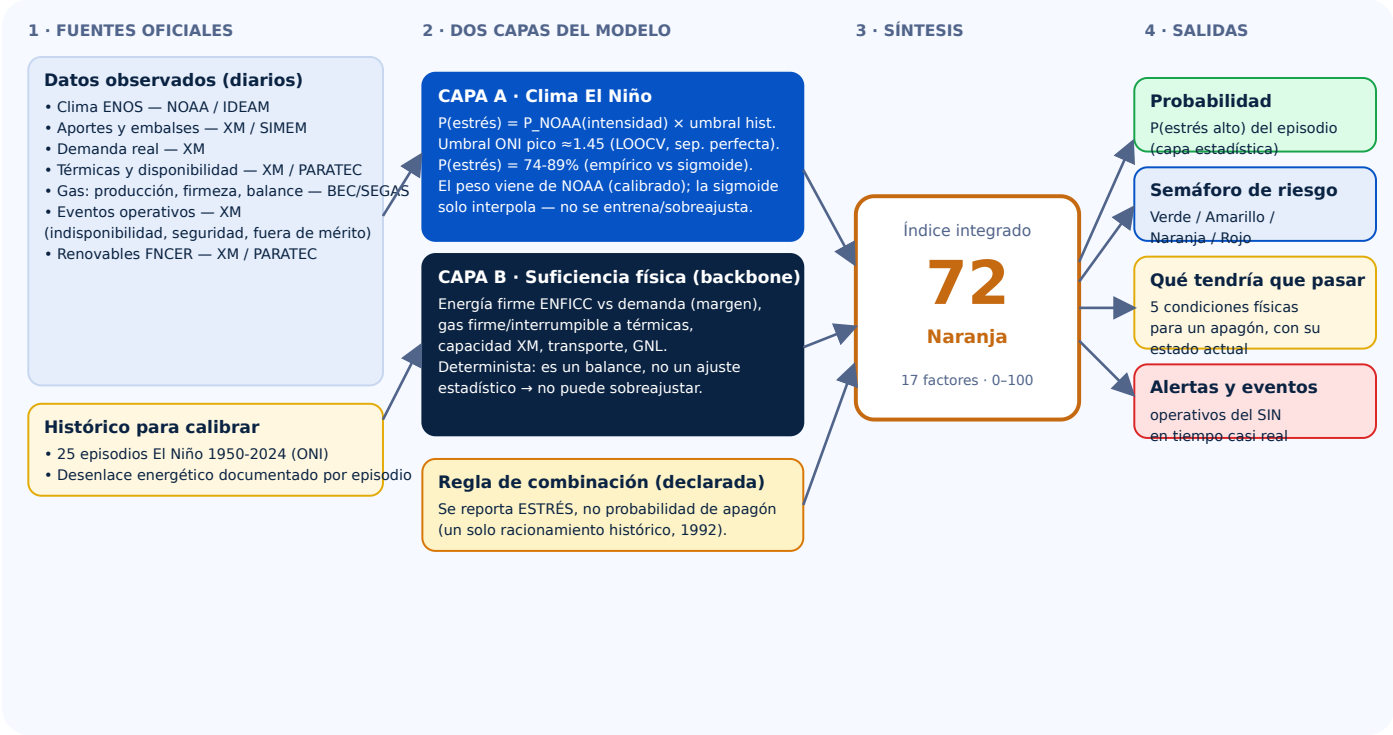
Nota: la forma multivariada diaria (ONI + volumen + brecha + estacionalidad + rezagos) se descarta como probabilidad de apagón por sobreajuste con n pequeño; sí se emplea, con esas variables, en el modelo de precio de bolsa (gradient boosting, R² 0.90), que es una magnitud continua, no una probabilidad de evento raro.

Anexo visual

DIAGRAMA DEL MODELO (CÓMO FUNCIONA)

CÓMO FUNCIONA EL MODELO (SUSTENTO METODOLÓGICO)

El riesgo de apagón no es una fórmula única: combina una capa estadística validada con evidencia histórica y una capa física de suficiencia de respaldo, ambas alimentadas por fuentes oficiales y recalculadas a diario.



Marco metodológico (rare-event): con un solo racionamiento histórico (1992) NO es posible entrenar una probabilidad de apagón. Por eso el riesgo se sustenta en (a) **suficiencia física determinista** —energía firme ENFICC vs demanda, gas firme/interrumpible— que es un balance, no un ajuste; (b) el clima El Niño como **pronóstico de intensidad de NOAA (calibrado externamente) × umbral histórico (ONI $\approx$ 1.45)**, reportando el rango entre el método empírico y la sigmoide para comunicar la incertidumbre; y (c) escenarios condicionales. La sigmoide solo interpola el condicional; no es "el predictor". Se reporta ESTRÉS del sistema, no una probabilidad calibrada de apagón.

MODELO DE ESTRÉS OPERATIVO — OBSERVADO VS. PREDICHO

Se incluye como evidencia del hallazgo: el estrés operativo NO es predecible con los fundamentales (R^2 fuera de muestra negativo); el modelo se aplanan y no sigue los picos. Por eso el riesgo se sustenta con el índice integrado y la logística de El Niño, no con esta señal.

Estrés operativo del SIN — observado vs. modelo (Gradient Boosting)

R^2 fuera de muestra: -1.37 · no predictivo ($R^2 < 0$) · 253 días de prueba · 10 variables (incluye indisponibilidad/mantenimientos)

