

Modelación de la expansión **energética nacional y regional del PEN**

2025 - 2055

Agosto 2026



UPME
Av Calle 26 # 69D – 91 Torre 1 – Piso 9
Bogotá – Colombia
Tel.: +57 6012220601
upme.gov.co

INDIRA CRISTINA PORTOCARRERO OSPINA

Directora General

JOHANNA CASTELLANOS ARIAS

Subdirectora de Demanda

Equipo Autor y Líder PEN

David Bedoya Rodríguez

Ingrid G. Quiroga Mojica

John A. Sánchez Cardozo

Equipo Autor del Documento

Carlos Espitia

Cristian Blanco

Daniel Galindo

José Chávez

Lorenzo Cardona

Natalia Plata

Colaboradores Externos

Juan Dominguez

Diseño y diagramación

Olga Lucía Rojas Solorzano

República de Colombia - Ministerio de Minas y Energía

Bogotá, mes 2026

CONTENIDO

- 1. ¿Qué es OSeMOSYS? 5
- 2. Planteamiento del modelo de optimización para la regionalización del PEN 5
 - 1.1 Modelo base de optimización y modificaciones para el PEN5
 - 1.2 Lógica de optimización del modelo base6
 - 1.3 Estructura del problema de optimización6
 - 1.4 Reformulación del balance energético13
 - 1.5 Extensiones incorporadas al modelo estándar16
 - 1.6 Valor agregado de la herramienta desarrollada16
 - 1.7 Evolución de la capacidad analítica17
 - 1.8 Gobernanza colaborativa del escenario18
 - 1.9 Análisis, visualización y entregables institucionales19
 - 1.10 Modelación de la regionalización20
 - 1.11 Principio metodológico de la desagregación espacial20
 - 1.12 Convenciones de nomenclatura y cobertura por región21
 - 1.13 Tecnologías de interconexión y transporte de energía entre regiones23
 - 1.14 Arquitectura de la plataforma24
 - 1.15 Alcance y Limitaciones27

INTRODUCCIÓN

El Plan Energético Nacional (PEN) 2025-2055 desarrolla, en su Tomo III, una metodología integral para regionalizar las proyecciones nacionales de demanda y oferta de energía en las siete regiones definidas para este ejercicio. Esa metodología responde a una pregunta específica, que es cómo distribuir territorialmente la información energética del país. El presente documento responde a una pregunta distinta y complementaria, que es cómo esa información regionalizada se convierte en escenarios cuantitativos capaces de orientar decisiones de inversión y operación en el sistema energético colombiano.

Para resolver esa conversión, la Unidad de Planeación Minero Energética (UPME) desarrolló OSeMOSYS-Colombia, una plataforma institucional construida alrededor de OSeMOSYS (Open Source Energy Modelling System), un modelo de optimización de código abierto ampliamente utilizado en la planeación energética de largo plazo alrededor del mundo. En términos sencillos, un modelo de este tipo no se limita a proyectar cómo evolucionará la demanda de energía. Su función es determinar, entre todas las combinaciones posibles de tecnologías, inversiones y modos de operación que podrían satisfacer esa demanda, cuál de ellas logra hacerlo al menor costo total posible para el país, respetando al mismo tiempo las restricciones técnicas, económicas y de política pública que se le impongan. Dicho de otra manera, el modelo no pregunta solamente cuánta energía necesitará Colombia en un año determinado, sino que responde qué combinación de plantas, redes y tecnologías conviene construir y operar para atender esa necesidad de la forma más eficiente.

Este documento describe a alto nivel, con ese propósito, la arquitectura de la plataforma, la lógica de optimización del modelo que la sostiene, el valor que aporta al proceso de planeación institucional, la forma en que su capacidad analítica ha evolucionado, su esquema de gobernanza colaborativa, su capa de análisis y visualización de resultados, la manera concreta en que se implementó la desagregación regional dentro del modelo, y los alcances y limitaciones que hoy tiene el ejercicio.

1. ¿Qué es OSeMOSYS?

OSeMOSYS es la sigla de Open Source Energy Modelling System, un marco de modelación energética de código abierto que distintos países y entidades de planeación usan alrededor del mundo para pensar el futuro de sus sistemas de energía a largo plazo (Howells et al., 2011). Fue desarrollado por un consorcio internacional de universidades y organismos de cooperación, entre ellos el KTH Royal Institute of Technology de Suecia y la International Energy Agency, con el propósito explícito de que cualquier gobierno o institución pudiera usarlo, revisarlo y adaptarlo sin depender de un proveedor comercial ni de una licencia cerrada.

Su función puede entenderse con una analogía sencilla. Si a un sistema energético se le conociera de antemano toda la demanda futura de electricidad, combustibles y demás energéticos, quedaría pendiente una pregunta nada trivial, que es con qué combinación de plantas, tecnologías e infraestructura conviene atender esa demanda, año tras año, gastando lo menos posible y sin violar restricciones como los límites de emisiones o la disponibilidad real de cada recurso. OSeMOSYS es la herramienta que resuelve exactamente esa pregunta. No pronostica la demanda por sí mismo, sino que toma esa demanda como un dato de entrada y calcula, entre miles de combinaciones posibles de inversión y operación, cuál es la más económica para el conjunto del sistema a lo largo de todo el horizonte de planeación.

Para lograrlo, el modelo traduce el sistema energético en un problema matemático de optimización. Cada tecnología, cada energético y cada restricción técnica o de política se convierte en una ecuación, y un algoritmo de programación lineal explora el espacio de soluciones posibles hasta encontrar la de menor costo total. Esta naturaleza matemática explícita es lo que permite que cualquier supuesto o resultado del modelo pueda auditarse y reproducirse, a diferencia de herramientas cuya lógica interna no es visible para quien las usa. En el caso colombiano, OSeMOSYS es el núcleo de optimización sobre el cual la UPME construyó OSeMOSYS-Colombia, la plataforma institucional que se describe en las siguientes secciones de este documento.

2. Planteamiento del modelo de optimización para la regionalización del PEN

1.1 Modelo base de optimización y modificaciones para el PEN

Una vez definidas las proyecciones regionales de demanda por sector y los parámetros de oferta energética descritos en las secciones anteriores, el paso siguiente consiste en integrar esa información dentro de un modelo que permita evaluar de forma conjunta las decisiones de inversión y operación del sistema energético a lo largo del horizonte de planeación. Esta sección describe la herramienta de modelación empleada para ese propósito,

las adaptaciones que se le incorporaron para atender las necesidades del PEN y la forma en que se implementó la desagregación regional dentro de su estructura.

1.2 Lógica de optimización del modelo base

El modelo de optimización empleado para el PEN es un marco de modelación energética de código abierto, diseñado para analizar sistemas energéticos en horizontes de largo plazo. Su función dentro del PEN es representar la cadena energética completa, desde la extracción o importación de recursos primarios hasta la satisfacción de la demanda final, y a partir de esa representación determinar simultáneamente qué capacidad conviene instalar en cada año del horizonte y cómo operar el parque resultante en cada fracción del tiempo modelado. A diferencia de una simulación con reglas operativas predefinidas, el modelo resuelve un problema de optimización en el que todas las decisiones de inversión y operación se toman de manera conjunta, buscando minimizar el costo total del sistema a lo largo de todo el período analizado (Barnes et al, 2022).

La estructura del modelo de decisión se compone principalmente de cinco componentes: conjuntos, parámetros, variables, restricciones, y la función objetivo. Los conjuntos definen el alcance del ejercicio: los años del horizonte, las tecnologías consideradas, los energéticos que circulan por el sistema, las emisiones contabilizadas, las fracciones temporales en que se divide cada año, los modos en que una tecnología puede operar y las regiones geográficas representadas. Los parámetros aportan los valores numéricos que caracterizan técnica, económica y ambientalmente cada elemento, y son el mecanismo principal por el cual se construyen escenarios alternativos sin modificar la formulación matemática. Las variables de decisión recogen aquello que el modelo determina: la nueva capacidad instalada por tecnología y año, y el nivel de operación de cada tecnología en cada fracción temporal. Las restricciones establecen las relaciones que toda solución admisible debe respetar, desde la coherencia física entre capacidad y producción hasta los límites de política pública. La función objetivo, por su parte, minimiza el costo total descontado del sistema energético a lo largo de todo el horizonte, de modo que la solución seleccionada es aquella que satisface la demanda y cumple todas las restricciones al menor costo posible.

La adopción de este marco de modelación para el PEN respondió a consideraciones que van más allá de lo estrictamente técnico. Al tratarse de un modelo de código abierto y formulación explícita, cada supuesto, cada ecuación y cada dato empleado en la planeación queda disponible para inspección, lo que permite auditar y reproducir cualquier resultado publicado. Esta condición otorga soberanía técnica sobre la metodología, dado que la UPME puede incorporar progresivamente nuevas restricciones, sectores u objetivos de interés nacional sin depender de herramientas propietarias cuya lógica interna no sea accesible.

1.3 Estructura del problema de optimización

El modelo implementado se fundamenta en un problema de programación lineal cuyo criterio es la minimización del costo total descontado del sistema, agregado sobre todos los años del horizonte. Este costo se compone del costo de operación descontado, la inversión de capital descontada y las penalizaciones económicas por emisiones, de los cuales se descuenta el valor de salvamento de las tecnologías cuya vida útil excede el horizonte de planeación. A esos términos convencionales el modelo OSeMOSYS-Colombia añade dos conceptos adicionales incluyendo el costo de desmantelamiento al final de la vida útil y el valor de recuperación de materiales.

Dos variables concentran las decisiones fundamentales y todas las demás se derivan de ellas. La primera es la nueva capacidad instalada de cada tecnología en cada año, que representa la decisión de inversión. La segunda es la tasa de actividad de cada tecnología en cada fracción temporal, modo de operación y año, que representa la decisión de operación y constituye el vínculo entre la capacidad disponible, la producción de energéticos y el consumo de insumos. La capacidad total disponible en un año no es una variable independiente sino la acumulación de las inversiones realizadas en los años previos que aún se encuentran dentro de su vida útil, sumada a la capacidad remanente heredada del parque existente al inicio del horizonte.

Sobre esas decisiones operan varios bloques de restricciones. El bloque de adecuación de capacidad impide que una tecnología opere por encima de la capacidad efectivamente disponible, considerando su factor de disponibilidad y el factor de conversión entre capacidad y actividad. El bloque de balance energético garantiza que, para cada energético, región, fracción temporal y año, la producción total sea suficiente para cubrir tanto la demanda final como el consumo intermedio de otras tecnologías. Los bloques de contabilidad económica traducen las decisiones físicas en flujos de costo y los actualizan a valor presente. Los bloques de límites acotan la capacidad y la actividad, tanto anualmente como en el acumulado del período modelado. El bloque de emisiones contabiliza la producción de contaminantes y aplica límites anuales y de período. El bloque de margen de reserva exige que la capacidad disponible supere la demanda que requiere respaldo. Existe además un bloque de almacenamiento, de activación condicional.

1.3.1 Conjuntos del modelo

La formulación se expresa sobre un conjunto de índices que delimitan el dominio del problema. Cada conjunto agrupa elementos del mismo tipo y sirve como dimensión de indexación para los parámetros, las variables y las restricciones. Los conjuntos principales son los siguientes:

- Y (años): cada período del horizonte de planeación. Si el ejercicio abarca de 2025 a 2055, el conjunto Y contiene los 31 años sobre los cuales el modelo toma decisiones de inversión y operación.
- T (tecnologías): el universo de alternativas de producción, transformación y transporte de energía representadas en el modelo. Incluye plantas de generación eléctrica (solar, eólica, térmica, hidroeléctrica), refinerías, infraestructura de transmisión, entre otras.
- F (combustibles o energéticos): los portadores de energía que circulan por el sistema. Electricidad, gas natural, diésel, hidrógeno, biomasa y demás vectores energéticos relevantes.
- L (fracciones temporales): subdivisiones dentro de cada año que permiten representar diferencias entre períodos de alta y baja demanda, o entre estaciones. Cuando el modelo emplea una sola fracción, cada año se trata como un bloque uniforme.
- M (modos de operación): determinadas tecnologías pueden operar de más de una manera. Por ejemplo, una planta de cogeneración puede producir exclusivamente electricidad o electricidad junto con calor útil.
- R (regiones): las unidades geográficas del ejercicio. En el escenario nacional existe una sola región; en el escenario regional se representan siete (Antioquia, Caribe, Interior, Nordeste, Oriente, Sureste y Suroccidente).

· E (emisiones): los contaminantes cuya producción se contabiliza y, cuando corresponde, se acota mediante límites de política. Incluye CO₂, CH₄, N₂O, entre otros.

La cardinalidad de cada conjunto determina el tamaño del problema de programación lineal resultante. Al pasar del escenario nacional al regional, el número de tecnologías se multiplica porque cada región incorpora su propia representación de las tecnologías pertinentes a su territorio, lo que incrementa de forma significativa el número de ecuaciones y variables que el solver debe procesar.

1.3.2 Parámetro de entrada

Los parámetros constituyen los datos numéricos que caracterizan el sistema energético y que el modelo recibe como información fija. Representan las condiciones técnicas, económicas y de política bajo las cuales se evalúan las decisiones. Se organizan en cinco familias.

Familia 1 – Demanda

Estos parámetros establecen la cantidad de energía que el sistema debe suministrar:

- SpecifiedAnnualDemand(r, f, y): demanda anual total del energético f en la región r para el año y.
- SpecifiedDemandProfile(r, f, l, y): distribución de esa demanda anual entre las fracciones temporales.

El modelo calcula internamente la demanda por fracción temporal como el producto de ambos parámetros:

$$\text{Demand}(r, l, f, y) = \text{SpecifiedAnnualDemand}(r, f, y) \times \text{SpecifiedDemandProfile}(r, f, l, y)$$

Familia 2 – Desempeño tecnológico

Definen la capacidad productiva y la eficiencia de cada tecnología:

- OutputActivityRatio(r, t, f, m, y): unidades de energético f producidas por la tecnología t por cada unidad de actividad, operando en modo m.
- InputActivityRatio(r, t, f, m, y): unidades de energético f consumidas por la tecnología t por cada unidad de actividad.
- CapacityFactor(r, t, l, y): fracción de la capacidad instalada efectivamente disponible en la fracción temporal l (valor entre 0 y 1).
- AvailabilityFactor(r, t, y): disponibilidad anual promedio de la tecnología (valor entre 0 y 1).
- CapacityToActivityUnit(r, t): factor de conversión entre unidades de capacidad (GW) y unidades de actividad (PJ).
- OperationalLife(r, t): vida útil de la tecnología a partir de su instalación, expresada en años.

· $\text{ResidualCapacity}(r, t, y)$: capacidad heredada del parque existente al inicio del horizonte, que se retira progresivamente según su vida útil remanente.

Familia 3 – Costos

Cuantifican el esfuerzo económico asociado a la inversión y operación de cada tecnología:

- $\text{CapitalCost}(r, t, y)$: costo unitario de instalación de capacidad nueva.
- $\text{FixedCost}(r, t, y)$: costo fijo anual por unidad de capacidad instalada.
- $\text{VariableCost}(r, t, m, y)$: costo variable por unidad de actividad.
- $\text{DiscountRate}(r)$: tasa de descuento empleada para actualizar flujos futuros a valor presente.

Familia 4 – Emisiones

Relacionan la actividad de cada tecnología con la generación de contaminantes:

- $\text{EmissionActivityRatio}(r, t, e, m, y)$: cantidad del contaminante e generada por unidad de actividad de la tecnología t .
- $\text{AnnualEmissionLimit}(r, e, y)$: límite máximo de emisión del contaminante e en el año y .
- $\text{EmissionsPenalty}(r, e, y)$: costo económico impuesto por cada unidad de emisión (cuando corresponde).

Familia 5 – Límites de política y operación

Acotan las decisiones del modelo conforme a restricciones de política pública o factibilidad física:

- $\text{TotalAnnualMaxCapacity} / \text{MinCapacity}$: capacidad máxima o mínima admisible de una tecnología en un año.
- $\text{TotalAnnualMaxCapacityInvestment} / \text{MinCapacityInvestment}$: inversión máxima o mínima permitida por año.
- $\text{ReserveMargin}(r, y)$: margen de reserva exigido sobre la demanda que requiere respaldo.

1.3.3 Variables de decisión

Las variables de decisión constituyen las incógnitas del problema. El solver les asigna valores numéricos que satisfacen simultáneamente todas las restricciones y minimizan el costo total del sistema. Las dos variables fundamentales son las siguientes:

- 1) $\text{NewCapacity}(r, t, y) \geq 0$ — representa la decisión de inversión: la cantidad de capacidad nueva de la tecnología t que se instala en la región r durante el año y .

- 2) $\text{RateOfActivity}(r, l, t, m, y) \geq 0$ — representa la decisión de operación: el nivel al que opera la tecnología t en la región r , fracción temporal l , modo m y año y . Constituye el vínculo entre la capacidad instalada y la energía efectivamente producida o consumida.

Las demás variables del modelo se derivan de estas dos. La capacidad total disponible en un año, por ejemplo, no es una decisión independiente sino la acumulación de las inversiones realizadas en los años previos que permanecen dentro de su vida útil, sumada a la capacidad heredada del parque preexistente:

$$\text{TotalCapacity}(r, t, y) = \sum \text{NewCapacity}(r, t, yy) \quad [\text{para todo } yy \text{ tal que } 0 \leq y - yy < \text{OperationalLife}(r, t)] + \text{ResidualCapacity}(r, t, y)$$

De manera análoga, los costos anuales, las emisiones y la energía producida se determinan como consecuencia directa de los valores que adoptan NewCapacity y RateOfActivity .

1.3.4 Función objetivo

El modelo determina la combinación de inversiones y niveles de operación que minimiza el costo total del sistema energético a lo largo de todo el horizonte de planeación. La función objetivo se expresa como:

$$\min Z = \sum_{r \in R} \sum_{y \in Y} \text{TotalDiscountedCost}(r, y)$$

La suma recorre todas las regiones y todos los años del horizonte. El término $\text{TotalDiscountedCost}$ de cada región y año consolida los costos de todas las tecnologías traídos a valor presente. Su descomposición es la siguiente:

$$\begin{aligned} \text{TotalDiscountedCost}(r, y) = \sum_{t \in T} [& \\ & \text{DiscountedCapitalInvestment}(r, t, y) \\ & + \text{DiscountedOperatingCost}(r, t, y) \\ & + \text{DiscountedEmissionsPenalty}(r, t, y) \\ & - \text{DiscountedSalvageValue}(r, t, y) \\ & + \text{DiscountedDisposalCost}(r, t, y) \\ & - \text{DiscountedRecoveryValue}(r, t, y) \\ &] \end{aligned}$$

Cada componente posee un significado económico preciso:

- Inversión de capital descontada: corresponde al costo de instalar capacidad nueva, anualizado mediante un factor de recuperación de capital y actualizado a valor presente con la tasa de descuento.

$$\text{CapitalInvestment}(r, t, y) = \text{CapitalCost}(r, t, y) \times \text{NewCapacity}(r, t, y) \times \text{CRF}(r, t) \times \text{PvAnnuity}(r, t)$$

$$\text{DiscountedCapitalInvestment}(r, t, y) = \text{CapitalInvestment}(r, t, y) / (1 + dr)^{(y - y_0)}$$

- Costo de operación descontado: suma del costo variable (proporcional a la actividad) y el costo fijo (proporcional a la capacidad instalada), ambos traídos a valor presente.

$$\text{CostoVariableAnual}(r, t, y) = \sum_m [\sum_l \text{RateOfActivity}(r, l, t, m, y) \times \text{YearSplit}(l, y)] \times \text{VariableCost}(r, t, m, y)$$

$$\text{CostoFijoAnual}(r, t, y) = \text{TotalCapacity}(r, t, y) \times \text{FixedCost}(r, t, y)$$

$$\text{DiscountedOperatingCost}(r, t, y) = (\text{CostoVariable} + \text{CostoFijo}) / (1 + dr)^{(y - y_0 + 0.5)}$$

- Penalización por emisiones: cuando el modelo incorpora un costo económico por emisiones, cada unidad emitida se valoriza y se suma al costo total del sistema.
- Valor de salvamento: las tecnologías cuya vida útil se extiende más allá del horizonte poseen un valor residual que se descuenta del costo total, en reconocimiento de la fracción de inversión aún no depreciada.
- Desmantelamiento y recuperación: extensiones incorporadas al modelo estándar que contabilizan el costo de retirar una tecnología al final de su vida útil y el valor de los materiales susceptibles de recuperación.

1.3.5 Restricciones del modelo

Las restricciones son las ecuaciones y desigualdades que garantizan la coherencia física de la solución y el cumplimiento de las políticas definidas. Se organizan en los siguientes bloques funcionales.

Bloque A — Adecuación de capacidad

Impide que una tecnología produzca más energía de la que su capacidad instalada le permite en cada fracción temporal. Se aplica para cada combinación de región, fracción temporal, tecnología y año:

$$\sum_m \text{RateOfActivity}(r, l, t, m, y) \leq \text{TotalCapacity}(r, t, y) \times \text{CapacityFactor}(r, t, l, y) \times \text{CapacityToActivityUnit}(r, t)$$

La suma del lado izquierdo agrupa la actividad de la tecnología en todos sus modos de operación. El lado derecho representa la capacidad disponible en esa fracción del tiempo, ajustada por el factor de capacidad (que refleja, por ejemplo, que una planta solar solo produce durante las horas con radiación) y el factor de conversión entre unidades de capacidad y actividad.

Bloque B — Balance energético por fracción temporal

Para cada energético, en cada región y fracción temporal, la producción total debe ser suficiente para cubrir la demanda final más el consumo intermedio de otras tecnologías que lo requieren como insumo:

$$\sum_{(t,m) \in \text{Productores}} \text{RateOfActivity}(r, l, t, m, y) \times \text{OutputActivityRatio}(r, t, f, m, y) \times \text{YearSplit}(l, y)$$

$$\geq \text{Demand}(r, l, f, y) + \sum_{(t,m) \in \text{Consumidores}} \text{RateOfActivity}(r, l, t, m, y) \times \text{InputActivityRatio}(r, t, f, m, y) \times \text{YearSplit}(l, y)$$

El lado izquierdo corresponde a la producción total del energético f: la suma sobre todas las tecnologías y modos que lo generan, ponderada por el coeficiente de producción y la fracción del año. El lado derecho acumula la demanda final y el consumo de las tecnologías que utilizan ese energético como insumo. La desigualdad exige que la oferta sea al menos igual a la necesidad total.

En esta formulación la demanda constituye una restricción estricta: no existe una variable de holgura que permita dejar demanda sin atender dentro del problema de optimización.

Bloque C — Balance energético anual

Para los energéticos cuya demanda se expresa de forma acumulada anual (sin distribución por fracción temporal), la ecuación de balance opera sobre la suma del año completo:

$$\begin{aligned} \sum_l \sum_{(t,m) \in \text{Productores}} \text{RateOfActivity}(r, l, t, m, y) \times \text{OutputActivityRatio}(r, t, f, m, y) \times \text{YearSplit}(l, y) \\ \geq \text{AccumulatedAnnualDemand}(r, f, y) + \sum_l \sum_{(t,m) \in \text{Consumidores}} \text{RateOfActivity}(r, l, t, m, y) \times \\ \text{InputActivityRatio}(r, t, f, m, y) \times \text{YearSplit}(l, y) \end{aligned}$$

Bloque D — Límites de capacidad y actividad

Imponen cotas superiores e inferiores a la capacidad instalada y a la actividad de cada tecnología, tanto anuales como acumuladas en el horizonte de planeación:

$$\begin{aligned} \text{TotalAnnualMinCapacity}(r, t, y) &\leq \text{TotalCapacity}(r, t, y) \leq \text{TotalAnnualMaxCapacity}(r, t, y) \\ \text{NewCapacity}(r, t, y) &\leq \text{TotalAnnualMaxCapacityInvestment}(r, t, y) \\ \text{NewCapacity}(r, t, y) &\geq \text{TotalAnnualMinCapacityInvestment}(r, t, y) \end{aligned}$$

Cuando el valor de un límite superior es igual al centinela convencional (9.999.999), la restricción se desactiva y la tecnología puede crecer sin cota.

Bloque E — Contabilidad de emisiones y límites

Las emisiones de cada tecnología se determinan como el producto del ratio de emisión por la actividad anual:

$$\text{EmisiónAnual}(r, t, e, m, y) = \text{EmissionActivityRatio}(r, t, e, m, y) \times \sum_l \text{RateOfActivity}(r, l, t, m, y) \times \text{YearSplit}(l, y)$$

La emisión total del sistema en un año corresponde a la suma sobre todas las tecnologías y modos, más las emisiones exógenas. El límite anual impone:

$$\sum_t \sum_m \text{EmisiónAnual}(r, t, e, m, y) + \text{AnnualExogenousEmission}(r, e, y) \leq \text{AnnualEmissionLimit}(r, e, y)$$

De manera complementaria, existe un límite acumulado sobre la totalidad del horizonte modelado (ModelPeriodEmissionLimit).

Bloque F — Margen de reserva

Garantiza que la capacidad total de las tecnologías clasificadas como respaldo supere la demanda que requiere cobertura, multiplicada por el factor de seguridad:

$$\text{DemandNeedingReserveMargin}(r, l, y) \times \text{ReserveMargin}(r, y) \leq \text{TotalCapacityInReserveMargin}(r, y)$$

Esta restricción asegura que el sistema disponga de capacidad suficiente para responder ante incrementos de demanda o salidas imprevistas de unidades de generación.

1.4 Reformulación del balance energético

La modificación de mayor alcance respecto del planteamiento de referencia no altera qué representa el modelo sino cómo lo escribe. En la formulación estándar, el paso desde la tasa de actividad de una tecnología hasta la producción de un energético atraviesa una cadena de ecuaciones intermedias. La actividad se convierte primero en producción por tecnología y modo, luego se agrega por tecnología, después se agrega en producción total del energético, y en paralelo se recorre una cadena equivalente para el consumo. Solo al final de esas cadenas se escribe la ecuación de balance que compara producción contra demanda más consumo intermedio. Cada eslabón introduce variables auxiliares y las ecuaciones que las definen.

Para maximizar la tratabilidad matemática del modelo, se aplican dos criterios (Kazempour et al, 2010): en primer lugar, el dominio del problema se construye recorriendo únicamente las combinaciones con coeficientes técnicos distintos de cero, evitando instanciar el producto cartesiano completo de tecnologías, energéticos y modos; en segundo lugar, se desactivan las variables y las restricciones de balance de aquellos energéticos que, para un periodo dado, no registran flujos de producción, comercio o demanda.

Esta reformulación responde a un requerimiento estricto de manejabilidad algorítmica computacional. Debido a la alta resolución espacial de tecnologías y energéticos en múltiples regiones, el modelo regionalizado alcanza una dimensión del orden de millones de restricciones. Mantener la representación nodal explícita con matrices densas vuelve la optimización numéricamente intratable dentro de los tiempos de ejecución razonables.

1.5 Motor de optimización

La plataforma emplea HiGHS como motor principal para resolver el problema de programación lineal. HiGHS es un solver de código abierto desarrollado en la Universidad de Edimburgo, reconocido por su eficiencia en problemas lineales y mixtos de gran escala (Huangfu et al, 2018). Se encuentra disponible sin licencia comercial, lo cual es consistente con el compromiso de la plataforma con herramientas de código abierto.

HiGHS implementa dos métodos numéricos principales para la resolución de problemas lineales:

- Método simplex: recorre los vértices del poliedro de soluciones factibles hasta identificar el óptimo. Resulta eficiente para la mayoría de los problemas y permite extraer información dual (precios sombra) de valor interpretativo para el análisis de resultados.
- Método de punto interior (IPM): se desplaza por el interior del poliedro factible aproximándose iterativamente a la solución óptima. Puede resultar más eficiente en problemas de gran dimensión o estructura dispersa, aunque requiere un paso adicional para obtener una solución de vértice exacta.

La plataforma permite seleccionar el método de resolución o delegar la elección a HiGHS, que determina automáticamente el método más adecuado según las características del problema. Además, para facilitar la operación y estandarizar las condiciones de ejecución, la plataforma define perfiles predefinidos que agrupan las opciones del solver de forma coherente:

- Perfil equilibrado: activa el preprocesamiento del problema, permite que HiGHS determine si conviene paralelizar, y desactiva el crossover cuando los valores duales exactos no son imprescindibles.
- Perfil rápido: emplea punto interior con preprocesamiento y paralelización activos. Prioriza la velocidad de resolución.
- Perfil de bajo consumo de memoria: emplea simplex sin paralelización. Reduce el consumo de recursos a costa de un mayor tiempo de ejecución.
- Para los escenarios regionales, la plataforma puede asignar automáticamente el método de punto interior, configuración que ofrece mayor estabilidad numérica cuando el número de restricciones es del orden de los millones.

1.6 Flujo de resolución

El proceso desde los datos de entrada hasta la obtención de la solución atraviesa las siguientes etapas:

Etapla 1 — Escritura del problema en formato LP. El modelo construido en Pyomo se exporta a un archivo de texto en formato LP estándar. Este archivo contiene la totalidad de las ecuaciones con etiquetas simbólicas legibles, lo que permite inspeccionar cualquier restricción de forma individual si resulta necesario.

Etapla 2 — Lectura del problema por HiGHS. El solver lee el archivo LP y construye su representación interna del problema: matrices de coeficientes, vectores de costos y cotas.

Etapla 3 — Preprocesamiento. Antes de resolver, HiGHS simplifica el problema eliminando variables fijas, detectando restricciones redundantes e identificando infactibilidades triviales. Este paso puede reducir de forma significativa el tamaño efectivo del problema.

Etapas 4 — Resolución. HiGHS ejecuta el método numérico seleccionado hasta alcanzar la solución óptima o determinar que el problema no admite solución factible.

Etapas 5 — Transferencia de la solución. Los valores óptimos de todas las variables se transfieren de vuelta al modelo Pyomo, donde quedan disponibles para consulta por nombre e índice.

Si el solver principal presenta una falla (por ejemplo, un error numérico), la plataforma dispone de un mecanismo de reintento con un backend alternativo para mantener la robustez operativa del sistema.

1.7 Escalabilidad y tiempos de ejecución

El tiempo total de una corrida de optimización depende de tres factores principales: el tamaño del problema (determinado por la cardinalidad de los conjuntos), la densidad de coeficientes técnicos (la proporción de combinaciones tecnología-energético-modo efectivamente activas) y los recursos computacionales disponibles. En la experiencia operativa con los escenarios del PEN, los casos nacionales se resuelven en tiempos del orden de minutos. Los casos regionales, al incorporar un mayor número de tecnologías y ecuaciones, requieren tiempos mayores, del orden de decenas de minutos, dependiendo de la configuración del solver y los recursos de cómputo asignados. La plataforma registra los tiempos de cada etapa (exportación de datos, construcción del modelo, escritura LP, resolución, transferencia de solución) como parte de la trazabilidad de cada corrida, lo que permite diagnosticar cuellos de botella y ajustar la configuración de forma informada.

1.8 Validación de datos previa a la resolución

Antes de construir el problema de optimización, los datos de entrada se someten a una serie de controles automáticos diseñados para detectar inconsistencias que, de otro modo, se manifestarían como infactibilidades de difícil interpretación:

- Detección de conflictos en cotas: si un parámetro de capacidad mínima excede al correspondiente máximo para la misma tecnología y año, el sistema identifica el conflicto antes de enviarlo al solver. Los conflictos cuya magnitud es inferior a 0,0001 se clasifican como diferencias de precisión numérica; los de mayor magnitud se reportan como conflictos reales que requieren revisión del modelador.
- Exclusión de años inactivos: si un año presenta todas sus fracciones temporales con peso nulo ($\text{YearSplit} = 0$ en todos los timeslices), ese año se excluye automáticamente del horizonte, pues no aporta restricciones ni variables al problema.
- Normalización de valores extremos: magnitudes inferiores a 10^{-300} que podrían desestabilizar la aritmética del solver se sustituyen por cero.
- Orden canónico de archivos: los archivos de datos se ordenan de forma determinista antes de alimentar al modelo. Esta medida garantiza que dos corridas con los mismos datos produzcan exactamente el mismo problema lineal, independientemente de la forma en que los datos se hayan cargado (desde archivo Excel, desde la base de datos o mediante importación de archivos externos).

Para reducir la complejidad matemática del modelo, se aplican dos criterios de filtrado estructural. En primer lugar, el dominio del problema se construye recorriendo únicamente las combinaciones con coeficientes técnicos

distintos de cero, evitando instanciar el producto cartesiano completo de tecnologías, energéticos y modos; en segundo lugar, se desactivan las variables y las restricciones de balance de aquellos energéticos que, para un periodo dado, no registran flujos de producción, comercio o demanda.

Esta reformulación responde a un requerimiento estricto de manejabilidad algorítmica computacional. Debido a la alta resolución espacial de tecnologías y energéticos en múltiples regiones, el modelo regionalizado alcanza

una dimensión del orden de millones de restricciones. Mantener la representación nodal explícita con matrices densas vuelve la optimización numéricamente intratable dentro de los tiempos de ejecución razonables.

1.9 Extensiones incorporadas al modelo estándar

La principal extensión incorporada al modelo estándar es la representación del costo de desmantelamiento y el valor de recuperación de materiales al final de la vida útil de cada tecnología. El planteamiento de referencia reconoce el valor de salvamento (la fracción de capital no depreciada cuando la vida útil excede el horizonte), pero no contempla ni el costo en que se incurre al retirar una instalación ni el valor de los materiales que pueden recuperarse en ese retiro. El nuevo modelo elaborado por la UPME incorpora ambos conceptos como proporciones del costo de inversión, descontados al año en que efectivamente ocurriría el retiro, e incluidos en la función objetivo con signo opuesto: el desmantelamiento incrementa el costo del sistema y la recuperación lo reduce. La razón detrás de esta incorporación es que las decisiones de inversión de largo plazo en el sector energético generan obligaciones de cierre que la evaluación económica convencional no suele considerar, y que el aprovechamiento de materiales al final del ciclo es un criterio de sostenibilidad que la planeación energética necesita poder valorar.

1.10 Valor agregado de la herramienta desarrollada

Construir los escenarios del PEN a partir de un modelo de optimización implica gestionar decenas de miles de parámetros que provienen de equipos distintos, de diversas fuentes, y que se revisan y modifican a lo largo de varios ciclos de discusión. Cuando esa información se administra mediante hojas de cálculo y cuadernos de análisis operados individualmente, el conocimiento sobre qué contiene exactamente un escenario y quién introdujo cada cambio tiende a quedar en la memoria de quienes participaron, lo que dificulta reproducir un resultado publicado meses después o responder ante una observación sobre un supuesto específico. Para atender este desafío se desarrolló una plataforma propia que no reemplaza el criterio de modelación, sino que convierte cada escenario en un objeto institucional persistente, consultable y reconstruible. A continuación se describen las principales capacidades que esta plataforma aporta al proceso de planeación.

La plataforma OSeMOSYS-Colombia garantiza, en primer lugar, que un mismo conjunto de datos produzca siempre el mismo resultado. Para ello organiza la información con un orden único antes de construir el problema de optimización y verifica que el resultado sea independiente de la forma en que los datos se hayan cargado. Esta reproducibilidad es fundamental para que los escenarios del PEN puedan ser auditados y replicados en cualquier momento.

En segundo lugar, toda modificación sobre un escenario queda registrada de manera permanente, identificando quién la realizó, cuándo ocurrió y qué valores cambió. Este registro permite deshacer cambios individuales,

rastrear la historia completa de un escenario y construir familias de escenarios derivados unos de otros sin perder de vista su genealogía. En la práctica, esto significa que explorar una variante de política pasó de ser un ejercicio que comprometía varios días de preparación a convertirse en algo que el equipo puede hacer dentro de una misma jornada de trabajo.

En tercer lugar, la plataforma valida automáticamente los datos de entrada antes de cada ejecución, identificando inconsistencias que de otro modo sólo aparecerían como errores difíciles de interpretar durante la ejecución del modelo. Además, permite que varios analistas trabajen simultáneamente en diferentes escenarios sin interferir entre sí, lo cual resulta particularmente importante dado el volumen de cálculo que exige el modelo regionalizado. En términos generales, el flujo de trabajo parte de los datos de entrada (provenientes de archivos estructurados o de la base de datos institucional), que pasan por una etapa de preprocesamiento donde se construyen los conjuntos del modelo y se corrigen las inconsistencias detectadas. Luego, se formula y resuelve el problema de optimización, y los resultados se almacenan para su consulta, la comparación entre escenarios y la generación de los análisis requeridos por el PEN.

1.11 Evolución de la capacidad analítica

La capacidad analítica con que hoy cuenta la UPME para la construcción del PEN es el resultado de un proceso progresivo que fue resolviendo limitaciones concretas del ciclo de planeación. La [Tabla 1](#) condensa el proceso anteriormente mencionado. Dicho proceso comenzó con la migración desde hojas de cálculo individuales hacia una base de datos institucional, pasó por la incorporación de fracciones temporales dentro del año para capturar la variabilidad de la demanda y los recursos renovables, y culminó con la desagregación regional del sistema en siete regiones y la implementación de un procedimiento de diagnóstico que permite identificar la causa cuando un escenario no tiene solución factible. En cada etapa se verificó que los resultados fueran consistentes con los obtenidos en la etapa anterior, de modo que cada avance se consolidó antes de abordar el siguiente.

Etapas	Situación previa	Capacidad incorporada	Criterio de aceptación
I. Modelación base	Cuadernos de análisis y libros de cálculo operados individualmente	Escenario persistente en base de datos, con ingreso y salida en formato estructurado y simulación desde la plataforma	Paridad numérica exacta con el flujo previo sobre el escenario nacional
II. Resolución temporal	Predominio de la agregación anual	Representación de fracciones temporales dentro del año en el despacho y en el análisis de resultados	Capacidad de representar la variabilidad de demanda y de recursos renovables

Etapa	Situación previa	Capacidad incorporada	Criterio de aceptación
III. Regionalización	Modelo nacional único	Siete regiones, tecnologías de transporte interregional y vistas agregada, filtrada y comparada	Solución óptima con cobertura completa y sin energía no suministrada
IV. Diagnóstico y operación	Fallo del solver sin explicación accionable	Diagnóstico progresivo por niveles, seguimiento operativo de la cola de simulación y exportación del problema para análisis externo	Clasificación explícita del fallo y distinción entre evidencia certificada y heurística

Tabla 1. Evolución de la capacidad analítica en el proceso de consolidación del modelo de planeación del PEN.

1.12 Gobernanza colaborativa del escenario

La construcción de un escenario del PEN es un ejercicio colectivo en el que participan equipos con responsabilidades sobre distintos segmentos de la cadena energética. Para que esa colaboración no comprometa la coherencia del escenario, la plataforma organiza el acceso mediante niveles diferenciados de permisos, es decir, quién puede modificar directamente, quién propone cambios para revisión y quién administra el conjunto. Como se detalla en la Tabla 2, Las modificaciones se registran de forma permanente y pueden revertirse, y el equipo ha adoptado una metodología de archivos de cambios que permite revisar y comparar cada intervención antes de incorporar al escenario vigente. De esta manera, el escenario deja de depender de la memoria de los participantes y pasa a ser el producto de una secuencia documentada de intervenciones que cualquier miembro del equipo puede examinar.

Mecanismo	Qué habilita en el PEN
Política de edición por escenario	Ajustar el régimen de acceso al momento del ciclo: abierto durante la construcción, restringido durante la revisión, exclusivo una vez publicado
Autorizaciones granulares diferenciadas	Que un equipo sectorial intervenga sobre su segmento sin capacidad de alterar el resto del escenario
Solicitudes de cambio con estados de revisión	Incorporar aportes de participantes sin edición directa conservando el control sobre lo que se acepta, con dictamen registrado
Etiquetado jerárquico con combinaciones excluyentes	Clasificar familias de escenarios impidiendo que uno quede rotulado bajo supuestos mutuamente contradictorios
Registro de auditoría con reversión por operación	Reconstruir la historia completa de un escenario y revertir intervenciones identificando cuándo la reversión no es segura
Archivos de cambios reducidos en repositorio compartido	Documentar cada iteración como artefacto independiente, revisable y aplicable por cualquier miembro del equipo

Revisión de diferencias previa a la aplicación	Verificar el efecto exacto de un conjunto de cambios antes de incorporarlo en el escenario
Verificación de integración de archivos acumulados	Comprobar que la acumulación de varios conjuntos de cambios produjo el escenario esperado
Linaje entre escenarios derivados	Rastrear de qué escenario proviene cada variante y qué parámetros la diferencian

Tabla 2. Mecanismos de gobernanza y su aporte al proceso de construcción del PEN.

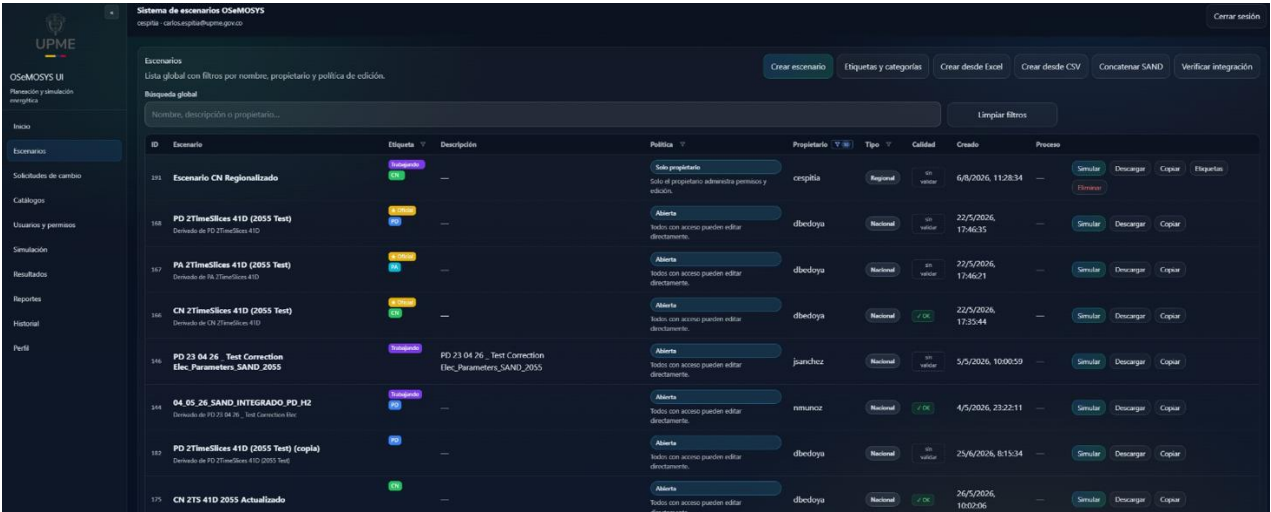


Figura 1. Interfaz del módulo de escenarios

1.13 Análisis, visualización y entregables institucionales

Un tomo del PEN necesita sostener lecturas dirigidas a audiencias distintas. La síntesis que requiere quien toma las decisiones, el detalle sectorial que examina el equipo técnico y el dato desagregado que exige la revisión externa. Para que esas lecturas sean consistentes entre sí, la plataforma estandariza la forma en que se presentan los resultados. El catálogo de análisis está alineado con la estructura sectorial del PEN y la configuración de las series (orden, nombres, colores) se define una sola vez y se aplica automáticamente a todas las visualizaciones (ver Figura 1). Esto evita que un mismo indicador se presente de formas distintas según quién lo elabore.

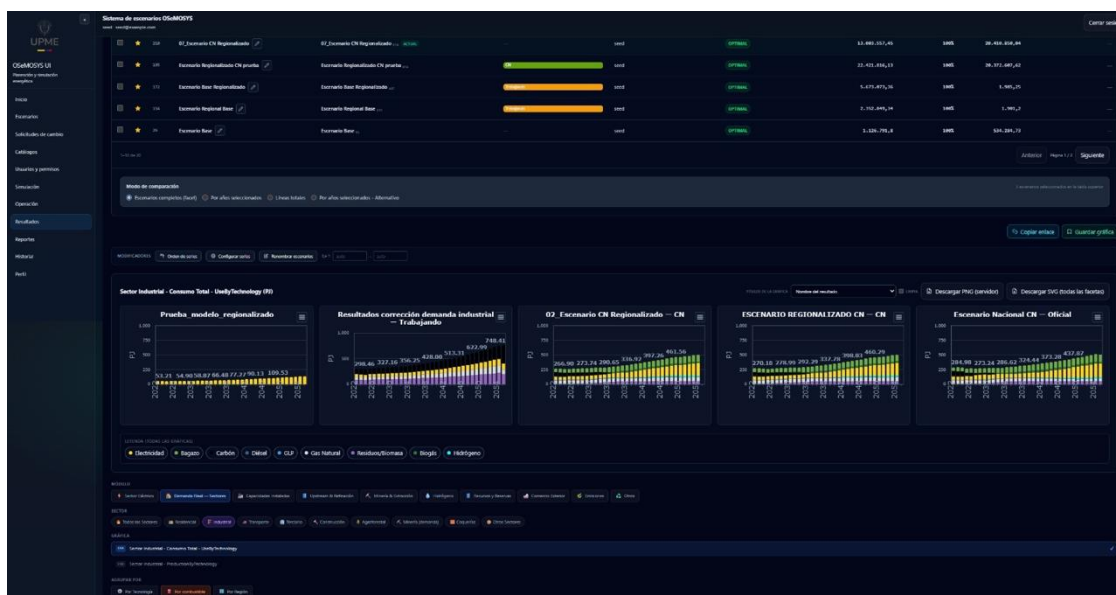


Figura 2. Interfaz del módulo de resultados: Herramienta de comparación de escenarios y evolución de iteraciones.

Para el escenario regionalizado, la capa de análisis resuelve además un aspecto técnico importante: al consolidar resultados nacionales excluye las tecnologías de transporte entre regiones, evitando un doble conteo que produciría contabilizar como generación la energía que solo se transfiere de una región a otra. La comparación entre escenarios permite analizar hasta diez escenarios simultáneamente bajo criterios homogéneos (ver Figura 2), y los resultados pueden exportarse en los formatos gráficos y tabulares que requiere el Tomo.

1.14 Modelación de la regionalización

1.15 Principio metodológico de la desagregación espacial

La decisión metodológica que estructura todo el ejercicio regionalizado es que la desagregación espacial no se implementó modificando la formulación matemática del modelo. La desagregación se resuelve, de hecho, creando versiones regionales de las tecnologías y los energéticos que lo requieren, distinguiendo cada versión mediante una etiqueta geográfica (un código de dos letras) antepuesta a su código original. No todas las tecnologías ni todos los energéticos aparecen en las siete regiones, dado que la presencia de cada elemento en una región determinada depende de si esa tecnología o energético existe actualmente o tiene potencial futuro en ese territorio. Así, una misma tecnología puede estar representada en una, dos, tres o más regiones según corresponda, cada una con sus propios parámetros técnicos y económicos. Adicionalmente, el modelo conserva tecnologías y energéticos sin prefijo regional, es decir, elementos de alcance nacional cuya naturaleza no requiere o no hace conveniente su regionalización, permitiendo que el modelo integre componentes globales junto con los componentes regionalizados. Lo anterior se puede encontrar resumido en la Tabla 3.

Esta elección tiene tres consecuencias favorables que justifican su elección sobre la alternativa de emplear la dimensión regional nativa. La primera es la paridad de formulación, puesto que al no introducirse ecuaciones específicas para el caso regional, la misma implementación matemática atiende tanto el escenario nacional como el regional, lo que reduce la posibilidad de que una diferencia de resultados provenga de una divergencia en el planteamiento en lugar de una diferencia en los datos. La segunda es la trazabilidad de la desagregación, como cada elemento regional conserva en su código la referencia al elemento nacional del que proviene, reconstruir el

agregado nacional a partir de los resultados regionales es una operación directa y verificable. La tercera es la previsibilidad del costo computacional ya que la desagregación incrementa el número de filas del problema proporcionalmente al número de elementos regionalizados, pero no introduce clases nuevas de ecuación, de modo que el crecimiento del tiempo de solución responde al tamaño del problema y no a un cambio en su naturaleza.

Dimensión	Modelo nacional	Modelo regional
Formulación matemática	OSeMOSYS con las extensiones descritas en 6.1	Idéntica, sin modificación alguna
Resolución espacial efectiva	Territorio nacional agregado	Siete regiones, mediante etiqueta geográfica en tecnologías y energéticos
Nomenclatura	Código de tecnología o energético sin etiqueta	Etiqueta regional de dos letras antepuesta al código
Tratamiento de la demanda	Demanda nacional por sector y uso final	Demanda desagregada por región, sector y uso final
Tratamiento de la oferta	Parque de conversión nacional	Parque asignado a cada región
Intercambio entre regiones	No aplica	Tecnologías explícitas de transporte interregional
Modos de operación	Uno	Dos; el segundo habilita la bidireccionalidad del transporte
Método de solución	Simplex	Punto interior con refinamiento, aplicado automáticamente
Obtención del agregado nacional	No aplica	Retirar la etiqueta regional y excluir el transporte interregional

Tabla 3. Comparativa estructural entre el modelo nacional y el modelo regional.

La contrapartida es que el intercambio de energía entre regiones, que en una formulación con dimensión regional nativa se representaría mediante variables de comercio, debe modelarse aquí como tecnologías explícitas de transporte que representan procesos que consumen el energético de una región y entregan el mismo energético en otra.

1.16 Convenciones de nomenclatura y cobertura por región

El modelo regionalizado representa siete regiones que cubren la totalidad del territorio nacional mediante la agrupación de sus treinta y tres entidades territoriales. Cada región se identifica con un código de dos letras que

antecede al código de toda tecnología y de todo energético que le pertenece, separado por un guión bajo. La convención se aplica de manera uniforme, una misma tecnología de generación existente en cinco regiones aparece como cinco elementos distintos que comparten el código base y difieren únicamente en la etiqueta. (ver Tabla 4) Los energéticos siguen la misma regla cuando su ámbito es regional, mientras que aquellos cuyo alcance es transversal al sistema conservan su código sin etiqueta.

Región	Código	Entidades territoriales	Nivel de representación	Intercambio con otras regiones
Caribe	CA	Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, Sucre	Completa	Sí
Oriente	OR	Bogotá D.C., Cundinamarca, Guaviare, Meta	Completa	Sí
Suroccidente	SO	Caldas, Caquetá, Cauca, Huila, Nariño, Putumayo, Quindío, Risaralda, Tolima, Valle del Cauca	Completa	Sí
Antioquia	AN	Antioquia, Chocó	Completa	Sí
Nordeste	NE	Arauca, Boyacá, Casanare, Norte de Santander, Santander	Completa	Sí
Este	SE	Amazonas, Guainía, Vaupés, Vichada	Reducida	Parcial
Insular	IN	San Andrés y Providencia	Reducida	No

Tabla 4. Cobertura territorial y nivel de representación por región.

La cobertura territorial, sin embargo, no implica que todas las regiones tengan el mismo nivel de detalle. Las regiones Caribe, Oriente, Suroccidente, Antioquia y Nordeste pertenecen al sistema interconectado nacional y cuentan con una representación completa de la cadena energética. Cada una incorpora alrededor de cuatrocientas tecnologías, noventa energéticos y un parque de generación diversificado que incluye recursos térmicos, hidráulicos y fuentes renovables no convencionales. Las regiones este e Insular, por su parte, se representan con un nivel de detalle menor. La región Insular se modela como un sistema aislado que no intercambia energía con las demás regiones, dada su condición geográfica. La región este, si bien cuenta con una

representación más reducida que las regiones plenamente interconectadas, no se encuentra completamente aislada del resto del sistema. En ambos casos, la escala de demanda y la diversidad tecnológica son menores que las de las regiones interconectadas, por lo que una representación con el mismo nivel de detalle habría incrementado el tamaño del problema sin aportar capacidad explicativa proporcional.

1.17 Tecnologías de interconexión y transporte de energía entre regiones

Como la desagregación regional no emplea la dimensión regional nativa del modelo, el intercambio de energía entre regiones no puede representarse mediante variables de comercio y se modela como un conjunto de tecnologías dedicadas. Cada una de estas tecnologías de transporte consume el energético correspondiente en la región de origen y entrega ese mismo energético en la región de destino, de modo que el flujo interregional queda sujeto a las mismas restricciones de adecuación de capacidad y a la misma contabilidad económica que cualquier otro proceso del sistema. Su nomenclatura identifica el energético transportado y el par de regiones que conecta, y en el caso del crudo incorpora además la calidad transportada.

La direccionalidad del intercambio se resuelve mediante el segundo modo de operación, cuya introducción constituye la única decisión de modelación genuinamente atribuible a la regionalización. Una tecnología de transporte bidireccional opera en su primer modo consumiendo el energético de la primera región y entregándolo en la segunda, y en su segundo modo invierte esa relación, consumiendo en la segunda y entregando en la primera. La totalidad de los enlaces eléctricos se modela de esta forma, reflejando que el sistema interconectado admite flujos en ambos sentidos según las condiciones de despacho (Loulou et al, 2005). En el caso del gas natural la bidireccionalidad se aplica de manera selectiva y responde a la topología efectiva de la infraestructura de transporte, de modo que sólo uno de los enlaces admite reversión mientras los restantes operan en un único sentido (Gago et al, 2013). El transporte de crudo se modela como un proceso unidireccional y distingue tres calidades de crudo: liviano, medio y pesado. Cada una se transporta por corredores independientes, ya que la calidad del crudo determina su destino de refinación y no puede sustituirse durante el proceso.

Un supuesto transversal a toda la representación del intercambio debe hacerse explícito, por su efecto sobre la interpretación de los resultados, dado que las tecnologías de transporte interregional no tienen límites físicos de capacidad. No se les asigna capacidad remanente que represente la infraestructura existente, ni capacidad máxima instalable, ni límite anual de actividad, ni factor de disponibilidad. El modelo decide endógenamente cuánta capacidad de interconexión construir, con el único condicionamiento del costo de inversión asociado. En consecuencia, los flujos interregionales que arroja el escenario deben leerse como el intercambio que resultaría económicamente conveniente en ausencia de restricciones de red, y no como un flujo verificado contra la capacidad de transporte efectivamente disponible o proyectada.

Energético	Enlaces modelados	Direccionalidad
Electricidad	Diez enlaces entre las cinco regiones interconectadas	Bidireccional en la totalidad de los enlaces, mediante el segundo modo de operación
Gas natural	Cuatro enlaces	Bidireccional en un enlace; unidireccional en los tres restantes, según la topología de la infraestructura

Energético	Enlaces modelados	Direccionalidad
Crudo	Quince enlaces, correspondientes a cinco corredores diferenciados en tres calidades	Unidireccional en la totalidad de los enlaces

Tabla 5. Representación del intercambio interregional por energético.

1.18 Arquitectura de la plataforma

Las secciones anteriores describieron qué es OSeMOSYS, qué valor aporta a la planeación del PEN, cómo ha evolucionado su capacidad analítica, bajo qué esquema de gobernanza se construyen los escenarios y de qué manera se implementó dentro de él la desagregación regional. Toda esa capacidad descansa, en la práctica, sobre una infraestructura técnica concreta. Antes de cerrar con los alcances y las limitaciones actuales del ejercicio, conviene entonces detenerse en esa infraestructura, es decir, en la arquitectura de la plataforma que hace posible todo lo descrito hasta aquí.

El desarrollo del PEN 2025-2055 se apoya en OSeMOSYS-Colombia, una nueva plataforma institucional construida alrededor del núcleo de optimización OSeMOSYS para sostener todo el ciclo de modelación. Los detalles técnicos de la plataforma, incluido su código fuente, se encuentran disponibles en https://github.com/UPME-SubDemanda/Osemosys_UPME. Como se ilustra en la Figura 3 esta plataforma integra siete componentes que operan de manera coordinada. Una base de datos relacional almacena cada escenario como un objeto institucional consultable asociado con un único identificador. Una capa de caché agiliza el acceso a resultados intermedios. Un sistema de gestión de flujos (o pipelines por su sinónimo en inglés) ejecuta las simulaciones. El motor de optimización resuelve el problema de programación lineal que sostiene el ejercicio de planeación. Un esquema de contenedores empaqueta y estandariza todo el entorno de ejecución. Un sistema de control de versiones garantiza la trazabilidad de los cambios sobre el código y los escenarios. Y una cola de tareas asíncronas permite ejecutar simulaciones de forma concurrente sin saturar la infraestructura computacional independientemente si es física o en la nube.



Figura 3. Descripción general de componentes de la plataforma OSeMOSYS-Colombia

Por otra parte, la Figura 4 muestra una representación de alto nivel de la implementación concreta de la mayoría de estos componentes, los cuales pueden ser observados con más detalle en la Tabla 6. El Analista o Planificador accede vía HTTPS a una interfaz desde donde crea escenarios, lanza simulaciones y explora resultados. La interfaz consume mediante una API REST la capa de backend, que expone los endpoints de escenarios, procesos, progreso y resultados y que cumple el rol de sistema de gestión de flujos descrito arriba. Esta API lee y escribe en la base de datos relacional, que almacena los esquemas de usuarios y del dominio energético, y encola cada trabajo de simulación en la capa de caché, que actúa como broker de mensajería y cumple la función de gestionar tareas asíncronas. Un proceso consume esos trabajos, ejecuta el pipeline completo de simulación y persiste resultados y eventos de vuelta en la base de datos. Para resolver el problema de optimización, un módulo invoca al solver de preferencia que actúa como motor de optimización. El esquema de contenedores y el sistema de control de versiones que completan los siete componentes operan como infraestructura de despliegue y desarrollo por debajo de esta capa aplicativa, y no se representan en el diagrama.

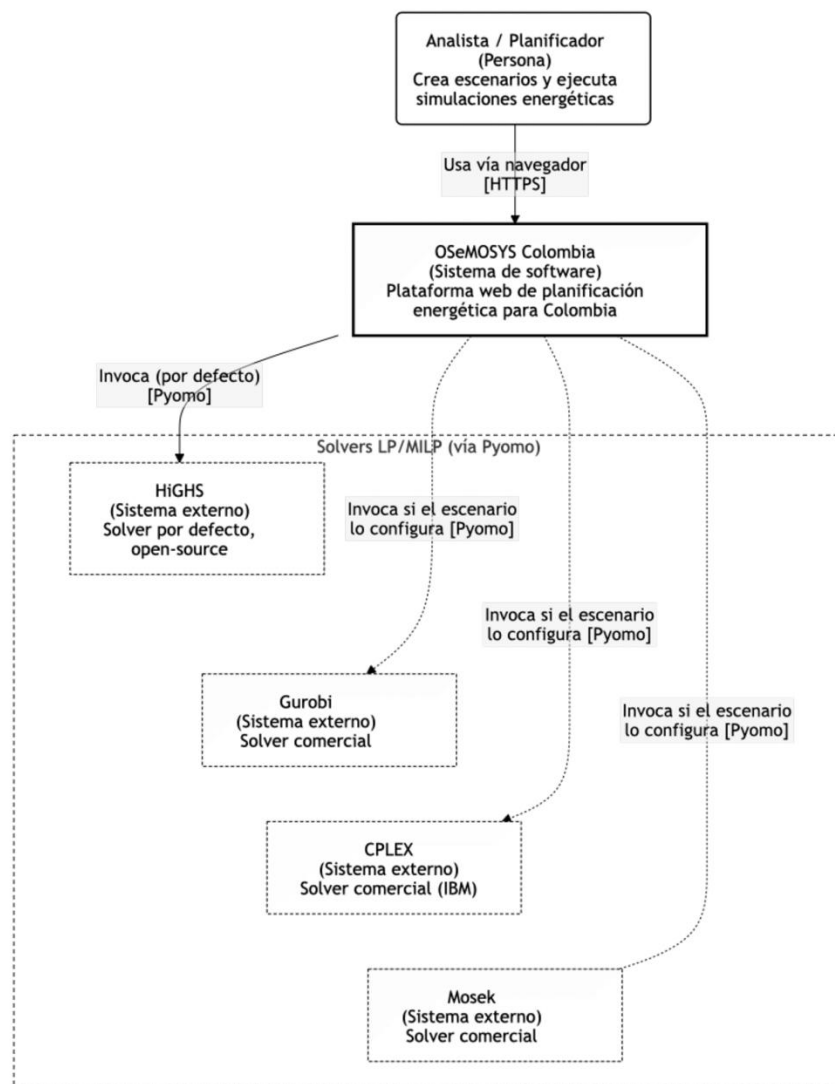


Figura 4. Representación C4 - Nivel contexto.

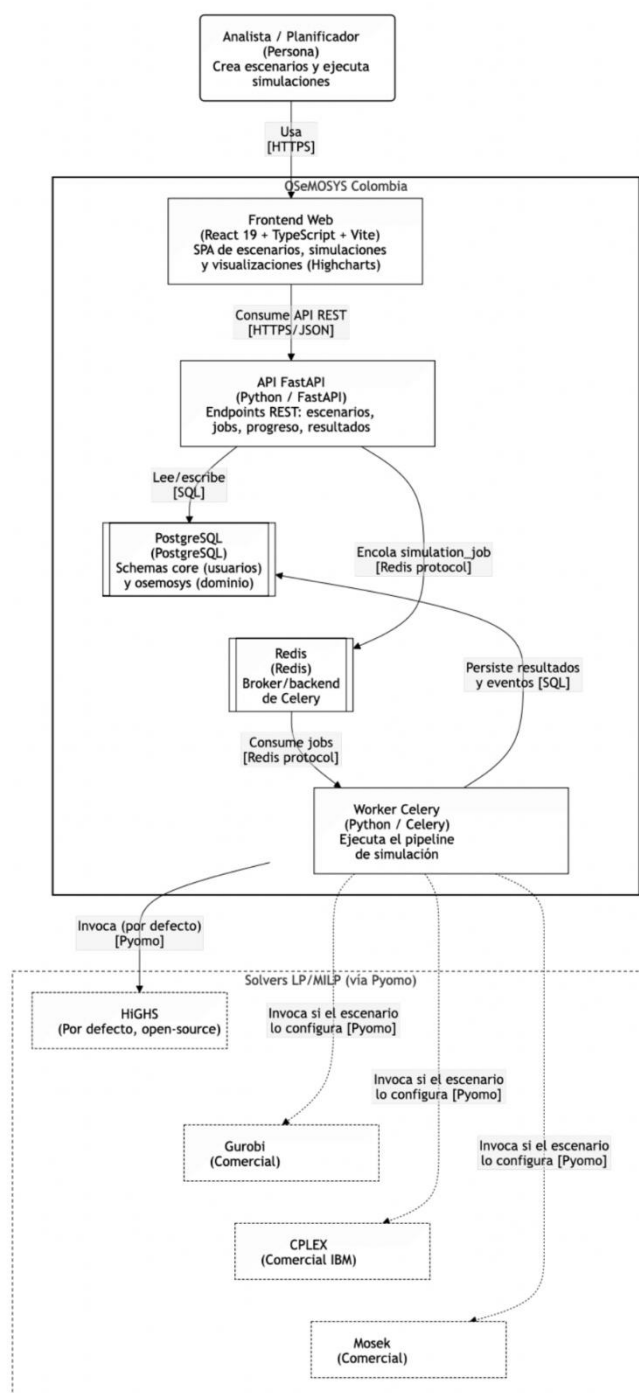


Figura 5. Representación C4 - Nivel contenedor

Componente funcional	Tecnología	Rol
Frontend Web	React 19, TypeScript, Vite	Interfaz para crear escenarios, lanzar simulaciones y visualizar resultados (Highcharts)

Componente funcional	Tecnología	Rol
API REST	FastAPI (Python)	Expone los endpoints de escenarios, jobs, progreso y resultados
Base de datos relacional	PostgreSQL	Almacena la información de los usuarios (analistas) y la información de los escenarios a simular
Capa de caché / broker	Redis	Broker y backend de mensajería para las tareas asíncronas
Gestión de pipelines / cola de tareas	Celery (worker)	Ejecuta el pipeline de simulación de forma concurrente
Motor de optimización	Pyomo, HiGHS (por defecto)	Resuelve el problema de programación lineal o entera mixta
Solvers alternativos	Gurobi, CPLEX, Mosek	Solvers comerciales, invocables si el escenario los configura
Esquema de contenedores	Docker	Empaqueta y estandariza el entorno de ejecución
Control de versiones	Git	Trazabilidad de cambios sobre código y escenarios

Tabla 6. Conjunto de herramientas que conforman OSeMOSYS-Colombia.

1.19 Alcance y Limitaciones

Hasta el momento, el ejercicio ha alcanzado dos avances que vale la pena señalar. El primero es contar con una primera implementación operativa del modelo energético nacional desagregado en siete regiones, resuelta hasta alcanzar una solución óptima y con cobertura de la demanda representada. Si bien esta desagregación constituye un paso importante, la representación de algunas regiones aún presenta limitaciones que deberán irse

resolviendo en futuras iteraciones del ejercicio. El segundo es disponer de una plataforma institucional que sostiene la construcción de los escenarios con gobernanza, registro de auditoría, análisis estandarizado y comparación entre escenarios, y que permite vincular el formato de trabajo del equipo de modelación con la base de datos institucional.

Las limitaciones deben entenderse en el mismo contexto. El escenario regional vigente constituye una primera versión. Aunque sus resultados son coherentes y comparables con los del escenario nacional, todavía pueden mejorar a medida que se refinan los datos utilizados para la desagregación regional. Asimismo, la representación del intercambio entre regiones no incorpora límites físicos de capacidad, por lo que los flujos interregionales reflejan la conveniencia económica, pero no la factibilidad de la red. El escenario regional también opera con una única fracción temporal anual, a diferencia del escenario nacional, lo que impide representar la variabilidad intradiaria y estacional de la demanda y de los recursos renovables (TaHERi et al, 2026). Esta simplificación responde al mayor esfuerzo computacional que exige el modelo regional y podrá superarse a medida que se disponga de una mayor capacidad de cómputo. Finalmente, el escenario regional no incorpora un requisito de margen de reserva y considera un conjunto reducido de límites anuales de emisión, lo que limita la comparación directa de sus resultados con los del escenario nacional en términos de seguridad del suministro y restricciones ambientales.

Finalmente, la calidad del escenario regionalizado depende en gran medida de los datos de entrada que se generan fuera de la plataforma. La asignación territorial de la demanda, la localización del parque de generación y la caracterización de la infraestructura de transporte provienen del proceso de modelación y de las fuentes de información del sector. La plataforma garantiza que estos datos se procesen de forma reproducible y trazable, pero no puede compensar posibles imprecisiones en la información de origen.

CONCLUSIÓN

OSeMOSYS-Colombia representa dos avances que conviene distinguir con claridad. El primero es metodológico, y corresponde a la primera implementación operativa del modelo energético nacional desagregado en siete regiones, construida desde el inicio con una visión transparente y alineada con los principios de la comunidad internacional de código abierto que sostiene a OSeMOSYS. El segundo es institucional, y consiste en contar con una plataforma propia que sostiene la construcción de los escenarios del PEN con reproducibilidad, trazabilidad, gobernanza colaborativa y análisis estandarizado, convirtiendo cada escenario en un objeto institucional consultable en lugar de un archivo disperso en la memoria de quienes participaron en su construcción.

Estos avances marcan un punto de partida sólido para la planeación energética territorial del país. Como toda primera versión de un ejercicio de esta escala, el escenario regional vigente seguirá madurando a medida que se refinan los datos de entrada y se fortalezca la capacidad de cómputo disponible, un proceso que la sección de alcances y limitaciones de este documento describe con el detalle técnico correspondiente.

Lo relevante de cara al futuro es que OSeMOSYS-Colombia fue concebido precisamente para crecer de esa manera. La misma arquitectura que hoy sostiene el ejercicio regional del PEN es la que permitirá incorporar, en próximas iteraciones, una resolución temporal más detallada, límites físicos de transporte interregional y una representación más profunda de las regiones que hoy cuentan con menor nivel de detalle. El código fuente de la plataforma permanece disponible para consulta pública en el repositorio institucional, lo cual refleja el compromiso de la UPME con una planeación energética auditable, reproducible y abierta al escrutinio técnico externo, y sienta las bases para que esa mejora continua ocurra a la vista de cualquier interesado en el futuro energético del país.

Bibliografía

Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolís, J., Bazilian, M., & Roehrl, A. (2011). OSeMOSYS: The Open Source Energy Modeling System: An introduction to its ethos, structure and development. *Energy Policy*, 39(10), 5850–5870. [1]

Barnes, T., Shivakumar, A., Brinkerink, M., & Niet, T. (2022). OSeMOSYS Global, an open-source, open data global electricity system model generator. *Scientific Data*, 9(1), 623.

Kazempour, S. J., Conejo, A. J., & Ruiz, C. (2010). Strategic generation investment using a complementarity approach. *IEEE transactions on power systems*, 26(2), 940-948.

Parallelizing the dual revised simplex method, Q. Huangfu and J. A. J. Hall, *Mathematical Programming Computation*, 10 (1), 119-142, 2018. DOI: 10.1007/s12532-017-0130-5

Loulou, R., Remne, U., Kanudia, A., Lehtila, A., & Goldstein, G. (2005). Documentation for the TIMES model PART I. Energy technology systems analysis programme, 384.

GAGO, D., Nijs, W., RUIZ, C. P., Sgobbi, A., Radu, D., Bolat, P., ... & PETEVES, E. (2013). The JRC-EU-TIMES model-Assessing the long-term role of the SET plan energy technologies.

Taheri, S. S., Rezakhanlou, M., Pour, A. M., & Fallahi, F. (2026). Evaluation of efficient policies on the renewable penetration in electricity sector for an integrated energy planning model in TIMES. *Environment, Development and Sustainability*, 28(2), 3563-3589.

