

Propuesta de reforma del Cargo por Confiabilidad

PREPARADO PARA ANDEG

Junio 30 de 2026

Fernando Barrera Rey
fernando.barrera@barrera-rey.com
Barrera | Rey Economic Advisory
www.barrera-rey.com

Mauro González Sierra
mauro.gonzalez@ggcinv.com
GGC Consulting
gonzalezsierra.co

Propuesta de reforma del Cargo por Confiabilidad

Introducción	1
1 Necesidad de una reforma del cargo por confiabilidad	3
1.1 Problemas del modelo actual	3
1.2 Problemas del cargo sin solucionar	16
2 Propuestas de reforma del cargo	25
2.1 Volver a un solo precio de escasez.....	25
2.2 Incorporar el producto potencia.....	26
2.3 Acortar las vigencias para nuevas plantas.....	31
2.4 Subastas para sustituir el mecanismo de la prorrata	32
2.5 Simplificar la liquidación	33
2.6 OEF horaria	34
2.7 Adoptar un Margen de Reserva Técnico y Económico	35
3 Medidas marginales de mejora	41
3.1 Algoritmo de casación de la subasta.....	41
3.2 Participación de las GPPS en la subasta primaria.....	42
3.3 Modificación de las garantías.....	44
Anexo 1: Remuneración de confiabilidad	46
Recuperación de costos en un mercado	48
Remuneración de confiabilidad.....	52

Propuesta de reforma del Cargo por Confiabilidad

Figura 1: Ingresos con OEFs	9
Figura 2: La distorsión del precio de Bolsa por dos precios de escasez	10
Figura 3: Curvas de demanda objetivo de dos productos	14
Figura 4: Ratio del precio de Bolsa entre las horas de punta y las horas fuera de punta — marzo–abril de 2026	23
Figura 5: Demanda máxima de potencia mensual (MW) 2021-mar 2026	28
Figura 6: Demanda máxima de potencia del SIN	28
Figura 7: Curva de optimización del costo total de confiabilidad	37
Figura 8: Sensibilidad del margen de reserva frente a la indisponibilidad del parque existente	39
Figura 9: Pérdidas por indivisibilidad de ofertas de la planta marginal	41
Figura 10: Curva monótona de carga	46
Figura 11: Tecnologías disponibles	46
Figura 12: La tecnología de ENS	47
Figura 13: Capacidad instalada óptima	48
Figura 14. Recuperación de costos en un mercado de pura energía	50
Figura 15: Remuneración de confiabilidad	53
Tabla 1: Beneficios de participar en tres mercados	7
Tabla 2: Beneficios de participar en tres mercados 2	8
Tabla 3: Vigencias de productos subastados	17
Tabla 4: Vigencia de las asignaciones de plantas	19
Tabla 5: Expectativa de pérdida de carga (EPC)	20

Introducción

El Cargo por Confiabilidad—Resolución CREG 071 de 2006—cumple casi 20 años de vigencia y aunque ha requerido modificaciones necesarias para hacerlo coherente con el modelo original propuesto por Cramton y Stoft (2006)—como la utilización de una planta real del sistema para el precio de escasez—también ha sido modificado para incluir objetivos adicionales—por ejemplo, la utilización de dos precios de escasez—para tener mayor cobertura de precios.¹

Por la visibilidad del cargo—ha sido protagonista de debates en el Congreso de Colombia y utilizado con fines políticos—su diseño suele tener presiones de modificación y múltiples quejas en torno a su funcionamiento en los eventos críticos. Sin embargo, dicho cargo ha permitido pasar por estos períodos en tres ocasiones (2009-10, 2015-6 y 2023-4) que se añaden, con la de 1997-98, a los éxitos del mercado eléctrico en evitar un racionamiento como los sufridos en otros eventos de Niño (1991-92 y 1980-81).

A pesar de la inconformidad con aspectos del cargo la expansión no había sido mayor problema desde la ley eléctrica. Aunque muchas de las plantas asignadas en las subastas no fueron construidas (p.e. TermoCol, Porce 4, Miel), el sistema pudo brindar confiabilidad en momentos críticos. A la fecha, los retrasos en la puesta en marcha de proyectos—de acuerdo con XM, en 2025 sólo se construyeron 22% de la energía esperada y en 2026 llevamos 0,6% de la capacidad proyectadas—parecen augurar un déficit de energía firme de 2,3% de la demanda esperada para 2026 y 4,4% para 2027. Pero, además, el sistema está mostrando un verdadero problema en la punta de demanda (19:00-21:00) porque la tecnología con mayor crecimiento—la energía solar fotovoltaica—no puede prestar ese servicio.

En esa coyuntura, se ha organizado la primera subasta con dos precios de escasez. Esta subasta tiene serios problemas de diseño porque se estarán adjudicando proyectos con diferentes precios de escasez—dos productos—con el mismo mecanismo. La Sección 1 de este documento presenta los principales problemas del cargo por confiabilidad actual que pueden llevar a una asignación errónea ineficiente y por ende costosa.

¹ Los dos precios de escasez—en su conjunto—no remuneran confiabilidad. La confiabilidad tiene que ver con remunerar a las plantas que contribuyen a la satisfacción de la demanda en el momento de punta térmica del sistema (cuando se resta de la demanda la generación no controlable)—Ver el Anexo. En Colombia se aproxima la demanda térmica del sistema a la demanda en un evento de hidrología crítico. En ese momento el precio debe subir por encima de la planta térmica más costosa y como pueden ocurrir muchas cosas para que ese precio no pase de ese nivel, se hace que el cargo sea la diferencia entre los precios de mercado de una situación crítica y el precio de la última térmica necesaria para abastecer la demanda. Esto equivale a un único precio de escasez, precios de escasez por tecnología implican cobertura de precios, algo asociado a los contratos y no a la confiabilidad.

En la Sección 1 motivamos las razones por las cuales se debe volver al esquema de un precio de escasez único, que implica cambios en el mecanismo de liquidación y en la remuneración de las plantas existentes. Pero la Sección también analiza las modificaciones pertinentes a problemas del cargo que se han quedado sin solucionar con las mejoras más recientes. Nos referimos a temas como:

- La ausencia de un mecanismo de remuneración de potencia que ya se había previsto por los autores (Cramton y Stoft) que diseñaron el cargo por confiabilidad y cuya ausencia—más la alta entrada de plantas que no contribuyen a la punta de demanda del sistema—han hecho que la restricción de potencia se vuelva vinculante.
- El problema de la prorrata: que pasó de ser un mecanismo de excepción a un mecanismo habitual y que, a pesar de múltiples intentos de modificación por parte de la Comisión, permanece sin modificar satisfactoriamente;
- La definición de una metodología para calcular el margen de reserva del sistema necesario porque el cargo debe cumplir con contingencias esperadas de la oferta y la demanda (no solo de la demanda); y
- El retorno a una liquidación simplificada en la que no se distorsione la contratación bilateral como instrumento de cobertura de precios.

La segunda y tercera sección del documento se centran en las propuestas de resolución de estos problemas propendiendo por la búsqueda de un cargo por confiabilidad más eficiente y que, por su funcionamiento, genere menos presión para su modificación.

1 Necesidad de una reforma incremental del cargo por confiabilidad

En esta Sección del informe motivamos la reforma del cargo actual. Analizamos las razones por las cuales las reformas más recientes—en especial los dos precios de escasez—no representan una remuneración por confiabilidad y como la manera en que se está encajando esta reforma tampoco es la apropiada. En el segundo apartado analizamos problemas que el cargo previo a estas reformas ya arrastraba y que sería conveniente resolver. De hecho, nuestras propuestas se basan en resolver estos problemas más que en emprender una gran reforma del cargo por confiabilidad.

1.1 PROBLEMAS DEL MODELO ACTUAL

En este apartado abordamos los problemas esenciales del modelo actual. Nos referimos a la existencia de dos precios de escasez que no corresponde al producto confiabilidad, sino que se asemeja a una cobertura de precios. Esto genera tres distorsiones; una distorsión de precios spot al tener un salto en la función de ingresos-precio de Bolsa, una liquidación muy compleja y una subasta ineficiente por tener dos productos casados con un solo algoritmo.

A continuación, los analizamos.

1.1.1 El producto confiabilidad se desvirtúa con dos precios de escasez

La confiabilidad—definida eléctricamente—consiste en la habilidad del sistema para entregar energía en la cantidad y calidad demandada en los puntos de consumo. Esto es un producto doble:

- Seguridad: la habilidad de un sistema eléctrico para soportar cambios repentinos.
- Suficiencia: capacidad del sistema de entregar energía en todo momento, tomando en cuenta contingencias programadas o cambios factibles.

La seguridad se presta con servicios auxiliares de frecuencia (primaria, secundaria, terciaria) y la suficiencia contando con la capacidad de producción que pueda suplir la demanda de energía en las situaciones previstas. La suficiencia implica variar la capacidad; un problema de mediano plazo y se asemeja más al concepto de confiabilidad utilizado en Colombia.

Confiabilidad en el sentido colombiano de la palabra es asegurarse que hay suficiente capacidad y energía para abastecer la demanda en las condiciones previstas. En un sistema eléctrico lo importante es suministrar la punta de demanda del sistema porque este es el momento más complejo y porque los

generadores no tienen otro uso para su capacidad, si se cuenta con suficiencia para atender la punta se cuenta para cualquier otro momento.

Cuánto y a quién se remunera la confiabilidad

Scheweppe, Caramanis, Tabors y Bohn (1988)² demuestran que en un mercado en equilibrio la fijación de precios spot lleva a que todas las centrales recuperen sus costos fijos y variables cuando se raciona el sistema en las horas en que el costo de racionamiento es más económico que otra tecnología. Es decir que un mercado de pura energía—sin remuneración explícita de capacidad—permita cubrir la demanda a mínimo costo, pero requiere unas horas de racionamiento porque es más barato racionar que construir una central de punta para esas pocas horas (ver el apéndice). Sin embargo, esta solución first-best suele no alcanzarse por diversos motivos:

- 1- La reticencia de las autoridades a racionar. Como el costo de racionamiento de las autoridades puede ser mayor que el costo de racionamiento de los consumidores las autoridades pueden querer tener reservas en exceso de lo que un mercado haría. Estas reservas deprimen el precio de energía con lo cual hace falta dinero para recuperar todas las inversiones y se requiere un pago explícito.
- 2- La posibilidad de oportunismo regulatorio. Como las inversiones en generación de energía no tienen uso en otra actividad, un regulador puede evitar remunerar la inversión y sólo pagar el costo de funcionamiento. En este caso se genera “*missing money*” que es equivalente a la anualidad de una central de punta.
- 3- El hecho que la confiabilidad es un bien público. Todos los usuarios se benefician de tener un sistema confiable, pero no es factible excluir a los usuarios que no paguen por esa confiabilidad. Por la imposibilidad de cobrar la confiabilidad a todo el que la provea se generan incentivos a no pagarla y el resultado es que se construye menor confiabilidad de la óptima. Esto requiere que el planificador central determine la cantidad de confiabilidad y sus beneficiarios.
- 4- Los topes a los precios. Cuando las autoridades ponen topes a los precios de mercado para evitar mayores tarifas se produce la ausencia de recuperación de las inversiones. En anticipación de estos topes las empresas no invierten y el mercado no brinda la confiabilidad necesaria. El regulador debe brindar garantía creíble de no intervenir el mercado.

² Fred Scheweppe, Michael C. Caramanis, Richard D. Tabors y Roger E. Bohn (1988) Spot pricing of electricity. Kluwer Academic Publishers, Boston, MA, 1988

- 5- La posibilidad de ciclos económicos. El resultado de Schweppe, et al (1988) surge cuando el sistema está en equilibrio, pero uno de los aspectos habituales del mercado son los ciclos de inversión, especialmente en mercados de menor tamaño. Por efecto de la incertidumbre se construye mucho o muy poco, cuando se construye poco los precios aumentan estimulando inversión y viceversa. Esto implica presiones de los inversionistas y de los usuarios que generan incertidumbre.

Por todas estas razones es necesario remunerar la confiabilidad de manera explícita. Y la respuesta de cuánto y a quién surge del mismo modelo que es el que inspira todos los pagos por potencia en el mundo:³

- La anualidad de una central de punta; y
- A todas las centrales que estén disponibles/generando en la punta del sistema.

¿Los dos precios de escasez remuneran confiabilidad?

Colombia escogió como método de remuneración de la confiabilidad una opción call (compra) americana (ejercida en cualquier momento) a precio de ejercicio igual al precio de escasez. En equilibrio la prima de esta opción es la planta más eficiente en generar únicamente durante los períodos de escasez, es decir que la prima de la opción es equivalente a la anualidad de una central de punta. Este valor se determina por medio de una subasta, pero la curva de demanda del producto está inspirada en el Costo del Nuevo Entrante, CONE, que es la anualidad de esta central.

Como el precio de escasez es equivalente al costo de funcionamiento de esta central, todos los modelos de remuneración de la confiabilidad por medio de opciones que se discutieron para el caso colombiano tenían esta característica. Nos referimos a los estudios de Cramton y Stoft, el de la Universidad de Comillas o el diseñado por Oren.⁴ La utilización de más de un precio de escasez surgió en Colombia a raíz del estudio que realizó TERA para la CREG del año 1999-2000, pero con la diferencia de la propuesta CREG que los generadores proponían el precio de escasez y la prima y competían en las dos dimensiones.

El modelo de TERA surge de que el objetivo de la consultoría era el de estandarizar contratos de energía y concibió las opciones como una manera de hacer cobertura

³ Ver, por ejemplo, los pagos por potencia en mercados de costos de América Latina como Panamá, Chile, Perú o Colombia con la resolución CREG-116 de 1996 donde el pago era de 5,25 USD/kw-mes que era la anualidad de una central de punta (turbina con gas natural) y que es también el pago por potencia en todos estos mercados.

⁴ Ignacio Pérez Arriaga, et al (2000) Estudio Cargo por Capacidad en Colombia realizado por la universidad de Comillas para Acolgén y presentado a los miembros de la CREG. Peter Cramton y Steven Stoft (2007) "Colombia Firm Energy Market" 11 de febrero de 2007. Shmuel Oren (2003) Ensuring Generation Adequacy in Competitive Electricity Markets" mimeografiado Universidad de Berkeley.

de precios. De la búsqueda de una cobertura de precios de Bolsa surge la idea de los dos precios de escasez y surge porque el cargo por confiabilidad de Cramton y Stoft buscaba cubrir toda la demanda de energía con el cargo, algo que no ocurre en nueva Inglaterra donde cada generador debe entregar la energía que le asigne la subasta y no una proporción de la demanda, como en el caso colombiano.

Con un precio de escasez superior y otro inferior lo que se hace es vender dos contratos de opciones de energía firme; uno a precio de escasez inferior—con una mayor prima—y otro a precio de escasez superior—con una menor prima. Es decir, la CREG ya no remunera confiabilidad por medio del cargo por confiabilidad, sino que el cargo ahora remunera cobertura de precios para toda la demanda, parte de ella a precio de escasez inferior y parte a precio de escasez superior.

Por medio de esta decisión se ha cambiado el papel del mercado spot por el de un mercado de contratos organizado por la CREG. Esto tiene efectos claros sobre el mercado de contratos porque el cargo por confiabilidad cubre de precios de Bolsa entre el precio de escasez respectivo y el costo de racionamiento cuando antes cubría por la diferencia entre costo de racionamiento y precio de escasez superior.

1.1.2 La liquidación es compleja y distorsiona

La existencia de varios precios de escasez y el efecto sobre el mercado de contratos es la razón por la cual la liquidación se vuelve muy compleja y la decisión que toma la CREG no es liquidar cada opción cuando se ejerce, sino liquidarla cuando el precio de Bolsa supera el precio de escasez más alto, pero liquidar por la diferencia entre el precio de bolsa y el precio de escasez respectivo.

Esto introduce una discontinuidad muy pronunciada en la función de ingresos del generador y el precio de Bolsa y distorsiones por la interacción de los contratos y el cargo. El efecto de esta discontinuidad sobre la formación de precios fue descubierto por Wolak y McRae (2016):⁵

“we show that the incentive for behavior caused by a supplier’s Firm Energy obligation and the net fixed price forward contract position can interact in perverse ways during scarcity conditions and near-scarcity conditions in a manner that works against the RPM mechanism providing a reliable supply of energy at reasonable price during El Niño Events” [pp 69].

A riesgo de ser un poco técnico es relevante explicar como el cargo, el precio de Bolsa y el contrato interactúan en el MEM.

⁵ Shaun D. McRae y Frank A. Wolak (2016) “Diagnosing the Causes of the Recent El Niño Event and Recommendations for Reform”. Ver Sección 6.1 del informe.

Distorsión de la OEF

Los incentivos de los generadores que pueden influir en los precios son complejos. Es importante analizar los incentivos que tiene un generador con posición de dominio, que puede generar en el ideal (Q_i), que vende contratos a precio fijo y que tiene una asignación de OEF (Q_f).

Para hacerlo debemos entender los ingresos y egresos por cada concepto. La tabla lo ilustra en el caso de una central hidráulica.

Tabla 1: Beneficios de participar en tres mercados

Mercado	Ingresos	Egresos
Bolsa	pQ_i	0
Contratos	$p_c Q_c$	$-p Q_c$
OEF	$p_f Q_f$	$Max[0, p - p_e]Q_f$

Fuente. Ejemplo

Donde p es el precio de bolsa, p_c de contratos, p_f de OEF y p_e el precio de escasez. Q_i es cantidad en el ideal, Q_c en contratos y Q_f en OEFs.

O una función de beneficios del estilo:

$$\pi = p(Q_i - Q_c) + p_c Q_c + p_f Q_f - Max[p - p_e, 0]Q_f$$

Para ver los efectos de la OEF veamos las dos situaciones:

1. Cuando no se ejerce la OEF:

$$\pi = p(Q_i - Q_c) + p_c Q_c + p_f Q_f$$

2. Cuando se ejerce la OEF

$$\pi = p(Q_i - Q_c - Q_f) + p_c Q_c + p_f Q_f + p_e Q_f$$

Un generador con posición de dominio puede fijar algunos de los tres precios, p_i , p_c o p , pero en la operación lo importante es el precio de Bolsa, p . Así es fácil ver que los incentivos del generador dependen de las dos cantidades: la vendida en Bolsa y la vendida en contratos—cuando los niveles de precios están por debajo del precio de escasez—y de las tres cantidades cuando los precios están por *encima* del precio de escasez.

- Cuando la OEF no se ejerce su decisión depende de las cantidades vendidas en contratos. Si las cantidades contratadas son bajas (altas) el

incentivo es el de subir (bajar) el precio de bolsa. Un resultado famoso desde el trabajo de Green y Newbery (1992).⁶

- Cuando se ejerce la OEF, los incentivos cambian, porque ahora la OEF es como un contrato que está activo. Una subida del precio de Bolsa tiene un impacto de reducción del beneficio de $Q_c + Q_f$ y el ingreso de sólo Q_i .

Entonces tenemos que los incentivos del generador pueden cambiar cerca del precio de escasez. La relación se vuelve muy compleja porque depende de tres cantidades variables en la cual los incentivos pueden ser subir, bajar precios o los dos. Podemos poner un par de ejemplos numéricos para entenderlo con mayor claridad.

Supongamos una situación con los siguientes precios:

$$p = 300 \text{ \$/kwh};$$

$$p_e = 250 \text{ \$/kwh};$$

El precio del contrato o de la OEF no es relevante porque no afecta los incentivos a manipular el precio de bolsa y porque es el mismo en los dos casos y para evitar información innecesaria no lo incluimos en la función de beneficios.

La siguiente tabla calcula los beneficios para una situación en la cual:

$$Q_c = 70 \text{ kwh};$$

$$Q_i = 100 \text{ kwh}; \text{ y}$$

$$Q_f = 20 \text{ kwh o } Q_f = 40 \text{ kwh}.$$

Tabla 2: Beneficios de participar en tres mercados 2

OEF	Liquidación
$Q_f = 20$	$\$300 \times (100 - 70 - 20) + 250 \times 20 = 3.000 + 500 = \mathbf{3.500}$
$Q_f = 40$	$\$300 \times (100 - 70 - 40) + 250 \times 40 = -3.000 + 1000 = \mathbf{-2.000}$

Fuente. Ejemplo

Nótese como la activación de la OEF puede llevar a un cambio de los incentivos. Si el generador se encuentra largo en contratos y OEF quiere deprimir los precios. Si es un generador hidráulico esto significa que desembalsa. Esto es más probable en cuanto tenga mucha OEF asignada o mucha cantidad comprometida en contratos.

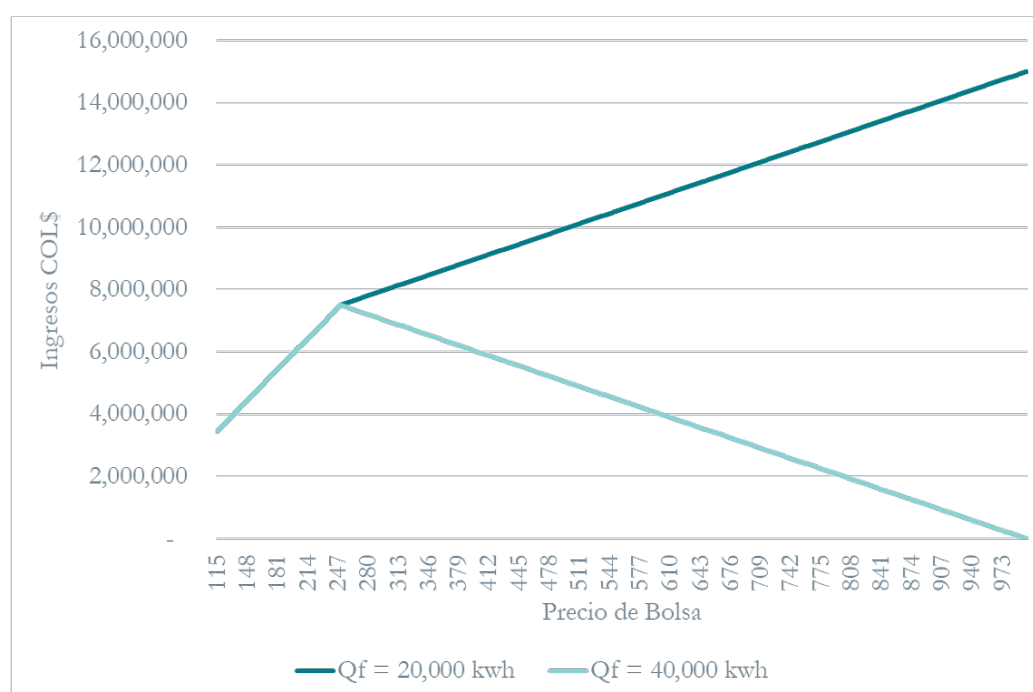
Si, por el contrario, tiene mucha agua que vender al ideal es probable que quiera subir los precios porque gana el precio de Bolsa y, probablemente, pueda vender

⁶ Green y Newbery (1992) "Competition in the British Electricity Spot Market" Journal of Political Economy, febrero de 1992, Vol 100(5):929-53

desviaciones a los generadores que no cumplen con su OEF. Ese parece haber sido el caso en el Niño 2015-6.

En conclusión, la OEF funciona como un contrato a precio de escasez que suele juntarse con los contratos que suscriben las centrales hidráulicas generando incentivos perversos cuando se ejerce el precio de escasez dependiendo de la contratación total. Así, puede ocurrir que la curva de ingreso marginal cambie en este punto haciendo que la decisión de oferta a la Bolsa sea compleja. Esto ocurre con mayor probabilidad cuando el precio de escasez es bajo, como en el caso del precio de escasez inferior. El siguiente gráfico muestra cómo el precio de escasez cambia los incentivos de las plantas.

Figura 1: Ingresos con OEFs



Fuente: Ejemplo

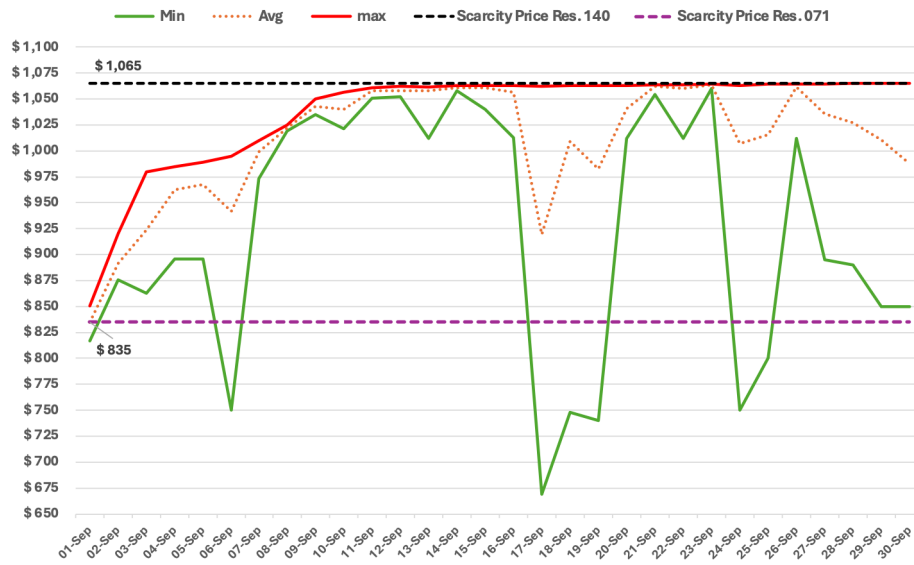
La discontinuidad en el precio de escasez de activación fue ampliamente debatida en el pasado. El efecto que esta discontinuidad genera en la operación del cargo por confiabilidad en la operación consiste en dos efectos:

- Evitar la activación de la OEF para las plantas con bajo precio de escasez: para de esta manera evitar la pérdida de ingresos de manera escalonada (perder la diferencia entre el precio de escasez de activación y el precio de escasez de la planta);
- La subida temprana de las pujas hidráulicas para evitar llegar a la condición de escasez.

La siguiente figura enseña cómo el precio spot deja de estar determinado por los fundamentales del mercado y depende del precio de escasez. Uno de los objetivos

de diseño del cargo fue que su introducción no distorsionara la señal de precios spot. Claramente la existencia de dos precios de escasez es una fuente de distorsión del precio de Bolsa, como lo enseña la siguiente Figura.

Figura 2: La distorsión del precio de Bolsa por dos precios de escasez



Fuente: XM

La interacción entre el precio de Bolsa, el precio de escasez de activación y los incentivos contractuales descritos en el apartado anterior se manifiesta empíricamente en una marcada concentración del precio de Bolsa alrededor del precio de escasez de activación durante las condiciones críticas.

La revisión de los tres últimos ciclos de Niño con la metodología expuesta por la propia CREG en el Documento CREG 901 338 de 2026 muestra que—a lo largo de períodos de 47 a 63 meses—el precio de Bolsa supera al precio de escasez inferior en aproximadamente el 40% de los días, mientras que rara vez se aleja sostenidamente por encima del precio de escasez superior. En la práctica, el precio de Bolsa se “pega” al gatillo de activación: una vez que el spot cruza el umbral, el ejercicio de la OEF deprime los ingresos marginales del generador con OEF y se pierde el incentivo a empujar el precio por encima del nivel de activación. La coincidencia entre el techo efectivo del mercado spot y el precio de escasez de activación, lejos de ser una rareza estadística, es la manifestación natural de los incentivos derivados del cargo, de los contratos y de la liquidación de la OEF.

Cuanto más bajo es el precio de escasez—como ocurre con el precio de escasez inferior—mayor es la duración de la activación y, por ende, mayor es el período en que el precio de Bolsa opera adherido al gatillo, lo que reduce la dispersión de precios y la calidad de la señal económica que el mercado spot debería entregar a los agentes en condiciones de estrés del sistema. Esto puede generar un exceso de utilización del embalse y la presión a que las autoridades utilicen medidas de

excepción—como las generaciones forzadas de las centrales térmicas—para proteger la confiabilidad.

Complejidad de la liquidación

La liquidación del cargo por confiabilidad se diseñó originalmente como un mecanismo simple. El Anexo 7 de la Resolución CREG 071 de 2006 establecía dos pasos secuenciales: (i) verificar, hora a hora, si el precio de Bolsa superaba el precio de escasez y, en ese caso, (ii) liquidar la diferencia entre el precio de Bolsa y el precio de escasez sobre la cantidad de OEF asignada a cada planta. Es decir, para cada hora h en la que el precio de Bolsa superaba el precio de escasez, el valor liquidado a cada planta i tomaba la forma:

$$VC_{i,h} = (PB_h - PE) \times OEF_{i,h}$$

Donde $VC_{i,h}$ es la compensación liquidada a la planta i en la hora h ; PB_h es el precio de Bolsa de esa hora; PE es el precio de escasez vigente; y $OEF_{i,h}$ es la Obligación de Energía Firme horaria asignada a la planta. La liquidación opera únicamente en las horas en que $PB_h > PE$.

Las transacciones de Bolsa se reliquidaban al precio de escasez (Artículo 55) y el ASIC distribuía los saldos remanentes a los comercializadores a prorrata de su demanda (Artículo 54). Aún en su versión original, la mecánica involucraba tres niveles de obligación—Mensual (OMEFR), Diaria (ODEFR) y Horaria (OEFT)—y reconciliaciones posteriores; pero el principio rector era único: un precio de escasez, un gatillo, una diferencia.

La introducción de dos precios de escasez—precio de escasez superior (PES) y precio de escasez inferior (PEI)—mediante la Resolución CREG 101 066 de 2024 rompe esa simplicidad. Tras dos rondas adicionales de modificaciones (Resolución CREG 101 069 de 2025 y, finalmente, Resolución CREG 101 097 de 2026), la liquidación del cargo se transformó en un procedimiento de nueve etapas secuenciales que requiere mantener simultáneamente tres precios de escasez de referencia—el PEI, el precio de escasez del Anexo 1 (PE) y el PES—reconciliados contra un único precio de Bolsa.

El nuevo Anexo 7 redefine, en primer lugar, el gatillo. La activación de las OEF se produce cuando el precio de Bolsa supera el Precio de Escasez de Activación (PEa), definido formalmente como:

$$PE_a = \max\{PEI, PE, PES\}$$

Es decir, el máximo entre el precio de escasez inferior (PEI), el precio de escasez del Anexo 1 de la Resolución CREG 071 de 2006 (PE) y el precio de escasez superior (PES).

Activado el cargo, la liquidación se desdobra por planta y por β —el subíndice que identifica cada precio de escasez al que la planta tiene OEF asignadas. Una misma central con OEF al PES y al PEI se trata, para efectos de liquidación, como si

fuesen dos plantas independientes: la Obligación Diaria de Energía Firme Ajustada (ODEFA) se prorratea por β según las cantidades adjudicadas en cada subasta:

$$ODEFA_{i,j,d,m,\beta} = ODEFA_{i,j,d,m} \times \frac{QOEF_{i,j,\beta}}{\sum_{\beta} QOEF_{i,j,\beta}}$$

Donde $ODEFA_{i,j,d,m,\beta}$ es la obligación diaria ajustada de la planta i del agente j en el día d del mes m , asociada al precio de escasez β ; $ODEFA_{i,j,d,m}$ es la obligación diaria ajustada total de la planta antes del prorrateo por β ; y $QOEF_{i,j,\beta}$ es la cantidad de OEF asignada a la planta al precio de escasez β en las subastas. El sumatorio del denominador recorre todos los precios de escasez β a los que la planta tenga OEF asignadas.

La generación ideal se distribuye entre los β proporcionalmente a la ODEFA por categoría, las desviaciones se calculan separadamente y la valoración financiera se ejecuta en paralelo para cada precio de escasez de la planta.

A esta segregación se añade una segunda capa de liquidación. La regulación reemplaza el precio único al que se reliquidaban las transacciones de Bolsa por un Precio de Transacción en Bolsa (PTB), calculado como el promedio ponderado de los precios de escasez por la OMEFR asignada a cada uno:

$$PTB_m = \frac{\sum_{\beta} PE_{\beta} \times OMEFR_{\beta}}{\sum_{\beta} OMEFR_{\beta}}$$

Donde PTB_m es el Precio de Transacción en Bolsa para el mes m ; PE_{β} es cada uno de los precios de escasez vigentes (precio de escasez inferior, precio de escasez del Anexo 1 y precio de escasez superior); y $OMEFR_{\beta}$ es la Obligación Mensual de Energía Firme Real asignada al precio de escasez β . El sumatorio recorre todos los precios de escasez de las OEF vigentes en el mercado.

Como el PTB es un único valor, pero cada planta tiene derecho a recibir el PE específico al que se le asignó la OEF, la regulación introduce ajustes residuales sobre las cantidades contratadas que comparan el PE de cada β contra el PTB para preservar la neutralidad del mecanismo. Estos ajustes operan en ambas direcciones—pagos cuando la planta tiene desviación positiva, cobros cuando la tiene negativa—y constituyen la última etapa del proceso. La propia CREG reconoce que el resultado son dos capas superpuestas de liquidación: en la primera se liquida la energía en bolsa al precio de Bolsa, o al PTB si se activa la condición crítica; en la segunda—la del Anexo 7—se concilian las desviaciones de OEF, los ajustes frente al PTB y la demanda sin cubrimiento de OEF si la hubiere.

La complejidad se completa con la incorporación de la Demanda Sin Cubrimiento (DSC). Como la suma de las OEF puede ser inferior a la demanda comercial en condiciones críticas, el regulador asigna proporcionalmente los cobros entre las plantas con desviación negativa y la demanda expuesta. El Precio de Reconciliación Negativa (PRN), por su parte, se redefine con casos diferenciados según si el precio de Bolsa supera o no el PEa, y con subcasos adicionales por β y por la relación

entre la generación real y la generación ideal de la planta. El resultado es un cuadro de liquidación en el cual cada hora de condición crítica genera, para cada planta, hasta nueve cálculos paralelos por cada precio de escasez al que tenga OEF asignadas.

Las consecuencias prácticas de esta arquitectura son visibles. En el proceso de consulta del Documento CREG 901 338 de 2026, los agentes—ANDEG, CELSIA, ANDESCO, ENEL, AES, ENFRAGEN, ERCO, NITRO, Grupo Scatec y Óptima, entre otros—coinciden en señalar que la fórmula no es replicable por los agentes, que el PTB puede tomar valores indeterminados o incluso superiores al precio de Bolsa, y que la única solución eficiente sería derogar las resoluciones que introdujeron los dos precios de escasez.

La comparación numérica respalda la inquietud: con los valores de referencia de diciembre de 2025—PEI de 329 \$/kWh, PE de 625 \$/kWh y PES de 865 \$/kWh—el PTB que resulta de la fórmula vigente es de 611 \$/kWh, mientras que los generadores con OEF al PES esperan ser remunerados a 865 \$/kWh. La brecha de 254 \$/kWh entre el PTB y el PES debe luego corregirse a través de los ajustes diferenciales del numeral 11 del Anexo 7, multiplicando los flujos contables sin alterar la sustancia de la remuneración.

En síntesis, la liquidación del cargo por confiabilidad pasó de ser un cálculo razonablemente sencillo—diferencial entre el precio de Bolsa y un único precio de escasez—a un sistema de doble capa, cuádruple precio de referencia (PB, PEa, PE asociado a cada β y PTB), nueve etapas y desdoblamiento por β en cada planta. Esta arquitectura no responde a una mayor precisión económica del mecanismo, sino a la necesidad de reconciliar contablemente la coexistencia de productos heterogéneos—las OEF al PES y las OEF al PEI—dentro de un mismo procedimiento de liquidación. Como se discute en la Sección 3, la simplificación de la liquidación es una de las reformas estructurales que el Cargo por Confiabilidad requiere.

1.1.3 La subasta es ineficiente

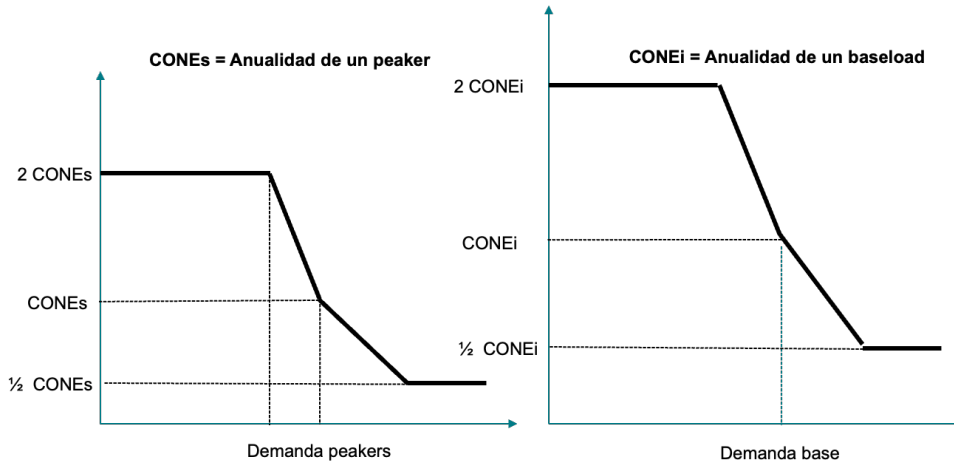
La CREG sale a subastar dos productos, el del precio de escasez inferior y el del precio de escasez superior. Dos productos significan dos mercados, dos curvas de oferta y dos curvas de demanda que pueden estar vinculadas—dependiendo de si los productos son complementarios o sustitutos del lado de la oferta o de la demanda—pero la CREG diseña el mercado con una sola curva de oferta y una sola curva de demanda.

Problemas de definición de la demanda objetivo

Es así como la curva de demanda objetivo está diseñada para comprar el producto de precio de escasez superior porque, como se ha dicho, en equilibrio este es el precio de la confiabilidad. Si el regulador quisiese comprar el producto—la anualidad—con precio de escasez inferior, es probable que esta anualidad

correspondiese a la de una planta de carbón. Es decir que las dos curvas de demanda fuesen las que se dibujan a continuación.

Figura 3: Curvas de demanda objetivo de dos productos



Fuente: ejemplos

Aunque podemos dibujar, con las dos demandas objetivos, la ordenada de las curvas de demanda es más difícil dibujar la abscisa. La razón es que las dos tecnologías tienen elementos de sustitución, pero también elementos de complementariedad. Un sistema óptimamente diseñado necesita de las dos tecnologías y un mercado sin fallas de mercado permite que este mix emerja, pero con los dos precios de escasez y la obligación de satisfacer toda la demanda el regulador sesgará dicho mix.

A la simplificación de la demanda objetivo el regulador añade dos problemas, uno por la oferta y el otro por el lado del algoritmo de casación.

Problema de oferta

La oferta de OEFs con dos precios de escasez surge de la mecánica de la subasta y de las ofertas de primas de las plantas. Pero las ofertas, en término de primas, no son comparables. No es más barato para el sistema una puja de prima X a precio de escasez superior que una prima $X + \varepsilon$ a precio de escasez inferior porque el sistema necesita saber el número de veces que se va a ejercer cada OEF y la OEF con precio de escasez inferior se ejercerá más veces o compensará un mayor monto. Con un solo precio de escasez se puede hacer dicha comparación porque el número de veces que se ejerce—y el monto que se compensa—es el mismo para las dos plantas.

Así que comparar las dos pujas como si entregasen el mismo servicio es un error importante y puede llevar a el resultado de la subasta no sea el de mínimo costo.

Problema del algoritmo de la subasta

El algoritmo de cierre diseñado por la CREG consiste en que cuando la subasta se cierre (última oferta casada cruce la demanda objetivo) la prima de cada grupo de plantas será la que haya fijado la marginal de su grupo. Esta lógica lleva a que se determinen dos primas; las de precio de escasez superior y las de precio de escasez inferior.

Sin embargo, esto no significa que los precios relativos de las primas sean correctos, es decir:

$$\Pi_{pes} \nless \Pi_{pei}$$

Que puede ocurrir que la prima de las plantas de precio de escasez superior no sea inferior a la prima del precio de escasez superior. Puede ocurrir, con una muy alta probabilidad, que una planta de precio de escasez superior sea la marginal. La razón es que la razón energía firme-capacidad:

$$REFC = ENFICC / \text{Capacidad}$$

de una planta de precio de escasez superior es muy superior a la una planta fotovoltaica de precio de escasez inferior que son las que se espera que acudan en mayor número a la subasta de cargo.

Esto ocurre por una razón muy sencilla, en un mercado los precios salen de la intersección de las curvas de oferta y de demanda. Aquí se pretende que dos precios salgan de una sola curva de oferta y una curva de demanda. Un error que esperamos no se presente en la próxima subasta.

1.1.4 El cargo para existentes no debería ser diferente

El cuarto error del cargo actual es la discriminación a las plantas existentes. La razón que se presenta es que las plantas pueden ya haber amortizado su inversión por su antigüedad. Ese argumento corresponde a un mercado regulado donde existe la certeza de la recuperación de todos los costos eficientes y se basa en las inversiones realizadas por las empresas reguladas.

Pero en Colombia el segmento de generación no es regulado, desde la ley 143 de 1994 se estableció que era un mercado competitivo con libertad de entrada (y de salida) y con precios fijados por la interacción de la oferta y la demanda. Es así como la Bolsa de Energía paga el mismo precio de Bolsa a todas las plantas que estén en mérito y todas ellas, menos la marginal, están recibiendo una compensación a sus costos fijos al haber pujado su costo variable. Ni la Bolsa de Energía, ni el mercado de contratos paga precios diferenciados por antigüedad a las plantas, sino que paga los precios que salgan del libre juego de la oferta y de la demanda.

Con respecto al cargo el argumento de la Bolsa de Energía aplica con fuerza también. El cargo está construido sobre la base de la renuncia que hace una planta

a los beneficios que obtendrá de los precios de bolsa por encima del precio de escasez:

$$B_i = (p - p_e)Q_i; Q_i > 0$$

Donde $Q_i > 0$ porque el precio de Bolsa excede el costo marginal c_i .

Nótese en esta formulación que el índice de la planta sólo aplica a la cantidad y al beneficio porque éste es función de los costos marginales. Pero ni el precio de Bolsa, ni el precio de escasez son específicos a la empresa, ni mucho menos a su antigüedad.

Con este ajuste la Comisión introduce remuneraciones diferenciales sin un sustento económico.

1.2 PROBLEMAS DEL CARGO SIN SOLUCIONAR

A pesar de que el cargo ha sufrido muchas modificaciones existen problemas que se han dejado sin solucionar. Aquí nos centramos en tres aspectos, en la discriminación de precios que genera problemas en las pujas de la subasta, el problema de la prorrata que surge del mismo problema de diferentes vigencias y la ausencia de un margen de reserva en la compra de OEFs que ha llevado a compras determinísticas

1.2.1 Mismo precio para diferentes vigencias

En las subastas de cargo realizadas hasta la fecha se han comprado productos muy diferentes. En la actualidad en Colombia se realiza una sola subasta para la entrega del producto OEF en una fecha de entrada. Sin embargo, no se subasta el mismo producto ya que en la subasta se asignan OEFs para distintas vigencias como puede apreciarse en la siguiente tabla.

Puede apreciarse como se subastan una gran cantidad de productos (vigencias de $t + j$ ($j = 2$ a 4) a $t + i$ con $i = 5, 14, 9, 24$ y hasta 30) pero todos los productos tienen el mismo precio de venta. Esta decisión se ha tomado debido a las largas vigencias de nuevas centrales que son difícilmente extensibles a agentes existentes, pero hace que los agentes pujen por cosas diferentes y salga un precio único. Esto hace que no se valore bien el producto.

La diferencia entre los períodos de vigencia introduce problemas. La central existente vende cobertura para el próximo año mientras que la central nueva la vende para 20 años. Como los periodos no coinciden puede ocurrir que el precio de la cobertura sea más alto, o más bajo, para el primer año que el precio de la subasta es decir el próximo año puede ser más seco o húmedo que la media de los próximos 20. Si el precio es más bajo (alto) la central no (sí) querría vender la cobertura, pero estaría limitada a hacerlo para prevenir el ejercicio de poder de mercado.

Necesidad de una reforma incremental del cargo por confiabilidad

Tabla 3: Vigencias de productos subastados⁷

Nombre	Descripción	Casos	Inicio	Fin
Existentes	Operación Comercial	Todos	T+4	T+5
Especiales	En Construcción	ENFICC ≤ 2TWha Inc ≥ 40%	T+4	T+14
	Existentes con adiciones de capacidad	ENFICC > 2 TWha Inc ≥ 0.8 TWha	T+4	T+14
Existentes con obras	Existentes y adiciones de capacidad (Incremental)	ENFICC ≤ 2TWha Inc ≥ 40%	T+4	T+9
		ENFICC > 2 TWha Inc ≥ 0.8 TWha	T+4	T+9
Nuevas	Todos		T+4...T+10	T+24,...T+30
Entrada Temprana	Todos		Antes de T + 4	Depende

Fuente: Regulación

Como los productos subastados en Colombia son diferentes y se utiliza un solo precio en la subasta, no es claro que los precios de las OEFs sean correctos o que sirvan para asignar, por ejemplo, GPPSs. Por ejemplo, en la subasta primaria participan:

- Recursos nuevos: para los cuales hay una vigencia de 20 años;
- Recursos especiales: para los cuales hay una vigencia de 10 años;
- Recursos existentes: para los cuales hay una vigencia de un único año y pueden retirar cantidades a partir de un precio fijado por la CREG.

Como el precio de la OEF es el mismo para los tres productos algunos lo venderán por debajo de costo de oportunidad y otros por encima. Por ejemplo, supongamos hay una alta probabilidad que los precios de mercado sean altos en los próximos 5 años y bajos después. Las centrales nuevas creerán en 5 años (33%) de precios altos y 15 de precios bajos mientras que las especiales creerán 50% de precios altos y las existentes 100% de período de precios altos. Claramente las nuevas—de salir asignadas—tendrán una renta y las existentes y las especiales una pérdida. Si el perfil de precios es el opuesto lo contrario ocurrirá.

Por eso es importante unificar vigencias para los recursos activos en la subasta.

1.2.2 Largas vigencias para nuevas plantas

Uno de los aspectos notorios del cargo por confiabilidad es el período de vigencia de la obligación. El cargo por capacidad tenía duración de un año, cada noviembre se realizaba la asignación para el año siguiente y solo se ajustaba de acuerdo con la

⁷ En la última subasta se han añadido productos como el de transición energética.

declaración comercial de la planta. Las plantas podían perder cargo en cada asignación con base en el despacho simulado lo cual era función de la competitividad de las plantas, de la evolución de los embalses, de la evolución de la demanda, etc.

El cargo por confiabilidad introduce diferencias dependiendo de si la planta es existente, nueva, existente con obras, pero el período de 20 años aplica a inversión de novo. Como a esto se le debe agregar el período de planeación (hasta 10 años para GPPS), la opción horaria debe valorarse durante un lapso de 30 años lo cual parecería a priori un período muy prolongado y que en las discusiones del cargo generaron múltiples argumentos respecto a la dificultad de valorar una opción tan compleja – el precio de la electricidad tiene propiedades estadísticas complejas – en un período tan largo de tiempo y ejercitable hora a hora.

El modelo colombiano se parece al de Nueva Inglaterra, pero en este tema hay una diferencia importante. La duración de la vigencia en ese mercado se fijó en 5 años. En Colombia se justificó de la siguiente manera:

“For new resources, the commitment period is up to twenty years. At qualification, the supplier specifies the desired commitment period from one to twenty years for each of its new projects. A winning supplier locks in the firm energy clearing price for the period it specified.”

The price is adjusted for inflation. The long commitment period lets new resources lock-in a firm energy price, reducing risk and encouraging investment.”

La duración de período era el instrumento para incentivar inversión. Sin embargo, no es correcto dar seguridad de inversión—lo cual sería el caso para una planta pico con un cargo a 20 años—sino que lo relevante es brindar una perspectiva razonable de recuperación de la inversión. Si se dan demasiadas seguridades se generan barreras a la entrada para plantas que quieran competir con los entrantes, es mejor contar con un cargo bien hecho que brinde señales razonables de recuperación de inversiones.

Este caso es claro si hay cambio tecnológico. En el caso extremo la duración de la vigencia debería ser igual a la vida útil económica de la planta, pero con innovación tecnológica que se está viendo en el caso de las renovables 20 años es un período muy largo hasta en este caso extremo.

En definitiva, para dar una perspectiva razonable de recuperación de costos de inversión lo que se requiere es tener un cargo que no se quiera cambiar sistemáticamente y que no represente barreras a la entrada. El número 20 se escogió sin mucho fundamento, aunque es largo si se contrasta con la experiencia internacional. La siguiente tabla muestra la duración máxima de los contratos para las nuevas instalaciones en diversas jurisdicciones.

Puede verse que la duración promedio de las jurisdicciones mostradas se sitúa alrededor de los 9 años, mientras que en ningún caso se sobrepasan los 15 años de duración. En el caso de las jurisdicciones que tienen implementado un mecanismo

de opciones, como Irlanda e Italia, ofrecen contratos de hasta 10 y 15 años de duración, respectivamente. Por tanto, Colombia se sitúa a nivel internacional claramente por encima tanto de la media como del rango alto en la duración de los contratos de OEF otorgados a las nuevas instalaciones.⁸

Tabla 4: Vigencia de las asignaciones de plantas

Jurisdicción	Duración (años)	Mecanismo
Alemania	2	Reserva estratégica
Bélgica	1	Reserva estratégica
España	10	Pago por capacidad
Francia	7	Obligaciones comercializadores
Irlanda	10	Opciones de confiabilidad
Italia	15	Opciones de confiabilidad
PJM	3	Mercado de capacidad
Polonia	15	Mercado de capacidad
Reino Unido	15	Mercado de capacidad

Fuente: regulación de los países.

1.2.3 El problema de la prorrata

El esquema de prorrata surge cuando no hay subasta y opera sobre el último valor salido de la subasta. Es un sistema imperfecto porque es administrado y se concibió como un mecanismo ocasional ya que se esperaba que, por el crecimiento de la demanda y el uso del período de planeación se realizaran subastas cada año.

Pero las subastas han sido poco frecuentes, hemos tenido subastas primarias en 2008, 2011, 2019, 2024 y 2026. Es decir, 5 en 20 años con lo cual la prorrata se ha vuelto un mecanismo habitual en lugar de uno de excepción. La prorrata ajusta por cantidades (que es lo que sobra) y no por precio que es el mecanismo elegido.

La prorrata, si se utiliza en exceso, hace que una planta ineficiente siga recibiendo cargo por confiabilidad. La razón es que recibirá una prima alta por parte de su energía firme, aunque la planta es probable que no saliese despachada en el mercado. Una planta sigue disponible si los ingresos de mercado cubren su costo fijo de mantenimiento. En la medida en que estos ingresos sean superiores a este costo la planta preferirá seguir disponible a pesar de que para el sistema puede ser mejor que cierre.

La CREG es consciente, como lo demuestran sus múltiples intentos por modificarla, que debe cambiar el mecanismo de prorrata, aunque no es explícita en sus razones para cambiarlo. Por eso ha recurrido a múltiples iniciativas administradas, algunas basadas en costo de funcionamiento (pero el valor del agua

⁸ Los pagos de capacidad en los países de la región son de 1 año de duración. En Brasil hay contratos de 20 años, como en Colombia, pero son contratos de energía, no son asignaciones de capacidad.

depende de la escasez), otras en términos históricos, y a iniciativas de subastas, algunas a un año, otras escalonadas, sin llegar a una solución. No es exagerado decir que la asignación más habitual, la prorratea, era la pensada como excepcional.

1.2.4 La ausencia de un margen de reserva

La confiabilidad, entendida como suficiencia, implica cubrir la demanda de energía ante las situaciones previstas. Un planificador prevé:

- Que algunas plantas no entren a tiempo;
- Que algunas plantas tengan mantenimientos programados;
- Que algunas plantas tengan fallas no programadas;⁹
- Que la previsión de demanda tenga imprecisiones.

Pero la metodología colombiana sólo reconoce la última y, en 2019, reconoció algo de la primera. Por eso es vulnerable a indisponibilidad de plantas. Otros países realizan estudios de confiabilidad que haríamos bien en analizar.

La experiencia internacional muestra que existen otras formas de determinar la necesidad de confiabilidad. Una práctica generalizada consiste en la aplicación de métodos probabilísticos. Este es el caso de jurisdicciones como Bélgica, Francia, Irlanda, Países Bajos, PJM, Polonia o el Reino Unido.

En estos mercados, la demanda de energía necesaria a cubrir mediante contratos de capacidad se calcula a partir de un estándar de confiabilidad denominado Expectativa de Pérdida de Carga (“LOLE” en sus siglas en inglés). El LOLE representa la cantidad de horas por año en que, a largo plazo, se espera estadísticamente que la oferta no cubra toda la demanda.

A continuación, se muestra la Expectativa de Pérdida de Carga de algunos de los países anteriormente mencionados y añadimos la colombiana

Tabla 5: Expectativa de pérdida de carga (EPC)

País	EPC
Bélgica	3 horas
Francia	3 horas
Colombia	23 horas
Irlanda	8 horas
Países Bajos	4 horas
Polonia	3 horas
Reino Unido	3 horas

Fuente. Ejemplo. Calculado como CONE/VOLL (el CONE de la subasta de cargo de 2024 y el Cuarto Escalón del Costo de Racionamiento)

⁹ La salida de Guatapé en el Niño 1997-98 y la del 2015-16 no fueron incluidas en el margen de reserva.

Para su cálculo se considera el valor que los consumidores asignan a la seguridad de suministro (“VoLL” en sus siglas en inglés o nuestro costo de racionamiento) y el costo del nuevo entrante (“CONE”). En concreto, $LOLE = (VoLL / CONE)$. De esta manera, se cumple la condición de que el costo incremental de asegurar la seguridad de suministro sea igual al costo incremental de un evento de energía no suministrada. En la tabla puede verse que el CR está muy subvalorado en Colombia (unos 4.000 USD/MWh, cuando los países de la muestra tienen 10.000-20.000 USD/MWh).

1.2.5 La no inclusión del producto potencia en la punta

Para poder definir el producto que remunera la confiabilidad lo que se necesita es determinar la punta de demanda térmica del sistema. Punta de demanda térmica consiste en sustraer de la demanda de energía del SIN los recursos que no son gestionables para atender la demanda. Se dice térmica, por tradición, pero en un entorno con energías renovables lo más correcto sería decir demanda flexible del sistema (DFS) expresada como:

$$DFS = \text{Demanda de energía hora } h - \text{Generación inflexible hora } h$$

La generación inflexible es toda la energía intermitente fotovoltaica, eólica y la hidro filo de agua; es decir energía que no puede gestionarse para abastecer la demanda de una hora específica. Una vez determinada la DFS se ordena de mayor a menor y se identifican las horas de mayor DFS.

Como puede verse hay dos factores que explican la DFS o la demanda del producto confiabilidad, la oferta flexible y la demanda de energía. En Colombia se reconoció, en el año 2006, que debido a que en un Niño mucha de la capacidad hidráulica no puede generar en períodos largos, los puntos más altos de la DFM se concentran en el tiempo. Esto hizo que fuese evidente que el producto a contratar fuese energía firme, simulada como la capacidad de cubrir la demanda durante el período crítico.

Pero, a pesar de definir un único producto, siempre se consideró que podía necesitarse potencia para cubrir la demanda punta de energía. De hecho, Cramton y Stoft lo dijeron en su informe sobre las OEFs:¹⁰

“Todo mercado eléctrico requiere suficiente capacidad y energía para satisfacer la demanda de forma confiable. Nos centramos en la energía firme —la capacidad de generar electricidad en períodos secos— porque esa es la restricción de confiabilidad

¹⁰ Peter Cramton y Steven Stoft (2007) “Colombia Firm Energy Market” 11 de febrero de 2007: Traducción de: “Ever[y] electricity market requires sufficient capacity and energy to reliably satisfy load. We focus on firm energy—the ability to generate electricity in dry periods—because that is the reliability constraint that is currently binding in Colombia. Today there is roughly 13 GW of capacity to service the annual peak of about 8 GW. There is a 5 GW or 62% surplus of capacity. Given this surplus, it is clear that the market price for capacity is zero. Firm energy is the scarce resource, and that is what consumers should pay for. In contrast, in the electricity markets in the thermal-dominated U.S. markets, the reliability constraint is having enough capacity to handle the annual peak, so the focus there is on capacity.”

actualmente vinculante en Colombia. En la actualidad, hay aproximadamente 13 GW de capacidad para cubrir un pico anual de alrededor de 8 GW. Esto implica un excedente de 5 GW, es decir, un 62% de capacidad adicional.

Dado este excedente, es evidente que el precio de mercado de la capacidad es cero. La energía firme es el recurso escaso, y es el producto que los consumidores deberían pagar. En cambio, en los mercados eléctricos de Estados Unidos dominados por generación térmica, la restricción de confiabilidad es disponer de suficiente capacidad para cubrir el pico anual, por lo que allí el enfoque se centra en la capacidad.” [pp 19]

Para ponerlo en los términos de Cramton y Stoft la restricción de confiabilidad era energía firme, pero podía llegar a ser capacidad para atender la punta de demanda anual si el parque cambiaba. Así lo ponen los autores:¹¹

“A lo largo de varias décadas, es posible que la combinación óptima de tecnologías de generación en Colombia cambie de tal manera que la capacidad se convierta en una restricción vinculante. Sería deseable que el mercado pudiera adaptarse a dicha evolución. En la práctica, acomodar la escasez tanto de energía firme como de capacidad puede lograrse fácilmente dentro del marco del diseño actual. Para completar el diseño, esbozamos el enfoque. No creemos que, en este momento, sea necesario introducir un mercado de dos productos, aunque hacerlo no implicaría costes de transacción significativos ni modificaría los resultados del mercado hasta que la capacidad se vuelva escasa. Por lo tanto, el mercado de dos productos podría introducirse de inmediato o más adelante, si y cuando resulte plausible que la restricción de capacidad pueda empezar a ser vinculante”. [resaltado fuera de texto]

Consideramos que este momento ha llegado. En Colombia, con la entrada de energía fotovoltaica y la incorporación de plantas hidráulicas con poca holgura la demanda punta puede verse desatendida. XM lo ponía de manifiesto en el informe del pasado junio de 2025:¹²

“En los últimos 5 años han ingresado al SIN 4.145 MW nuevos. En 2023 ingresó el 17% de la capacidad que se esperaba fuera puesta en servicio, mientras que en el 2024 ingresó el 25%. Con lo anterior se observa que la tasa de entrada real de nuevos proyectos de generación, respecto a las expectativas por sus FPO declaradas, no ha superado el 28% en los últimos 5 años, lo que indica una baja incorporación de nuevos recursos al sistema, que podría comprometer el abastecimiento seguro y confiable de la demanda;

¹¹ Traducción de: “Over a period of decades, it may be that the optimal mix of plants in Colombia will change in such a way that capacity becomes a binding constraint. It would be nice if the market could handle such an evolution. As it turns out accommodating scarcity of both firm energy and capacity is easily accomplished within the framework of the current design. For completeness, we sketch the approach. We do not believe at this time that it is necessary to introduce the two-product market, although doing so would not add any significant transaction costs, nor change market outcomes until capacity becomes scarce. Thus, the two-product market could be introduced immediately, or later, if and when it becomes plausible that the capacity constraint could begin binding”

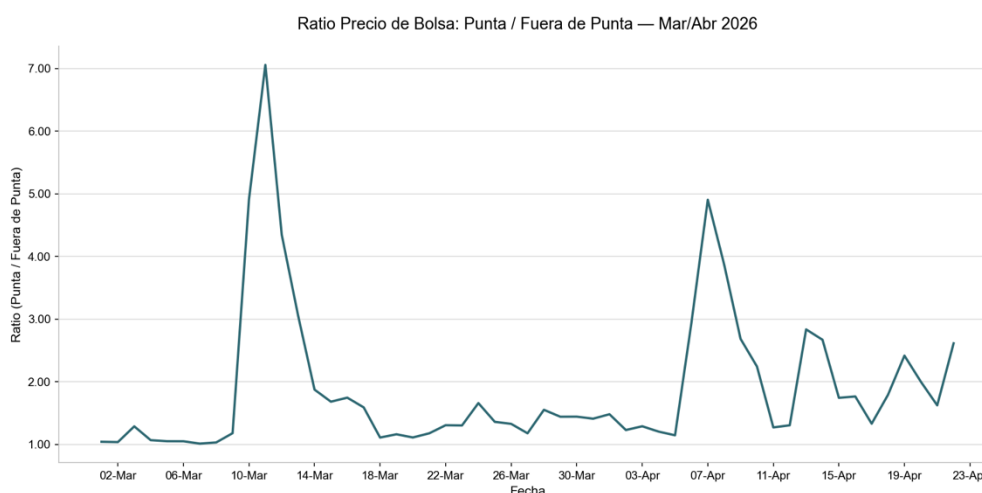
¹² XM (2025) “Propuestas del Centro Nacional de Despacho sobre riesgos operativos del SIN” Documento XM-CND-014 de junio de 2025

más aún, un volumen importante de los recursos que se integran corresponden a plantas solares, las cuales no aportan potencia para atender periodos de punta, lo que podría generar estrechez para atender la demanda pico del sistema, generalmente materializada entre los periodos 19 al 21 en los cuales no es posible contar con la generación basada en tecnología solar.” [pp 19, resaltado fuera de texto]

Evidencia de un déficit de potencia también se puede apreciar en los precios en la punta de demanda diaria del mercado spot que en este invierno prolongado muestran precios muy altos en la punta de demanda comparados con precios a lo largo del día.

Para cuantificar este fenómeno, calculamos para cada día del bimestre marzo–abril de 2026 el cociente entre el precio promedio de Bolsa observado en las horas de punta del sistema (19:00–21:00) y el precio promedio de Bolsa observado en el resto del día. Si la oferta de capacidad fuese holgada y el sistema acomodase la demanda sin tensiones, el cociente debería oscilar de manera estable en torno a la unidad; cuando la capacidad disponible en la punta deja de ser suficiente, el cociente se separa de uno y refleja el sobreprecio que el mercado spot le asigna a la potencia disponible en esas horas.

Figura 4: Ratio del precio de Bolsa entre las horas de punta y las horas fuera de punta — marzo–abril de 2026



Fuente: cálculos propios con base en información de XM.

La serie evidencia un patrón claro. La mayor parte del tiempo el cociente se sitúa entre 1,0 y 1,5—un rango compatible con un sistema en el que la energía firme es la restricción de confiabilidad vinculante pero la capacidad para atender la punta aún no lo es—pero exhibe episodios puntuales en los que el precio promedio de la punta supera entre cuatro y siete veces al precio promedio del resto del día. El pico del 10–12 de marzo, con un cociente cercano a 7, y el del 7 de abril, próximo a 5, muestran que el mercado spot está señalizando episodios de escasez que se concentran exclusivamente en las horas en que la generación solar fotovoltaica deja de aportar potencia. Adicionalmente, en las dos últimas semanas reportadas el

cociente se desplaza al alza—oscilando entre 1,5 y 2,8—, lo que sugiere que la estrechez de potencia ha dejado de ser un fenómeno episódico para acercarse a una condición estructural del sistema. Esta separación entre el precio en la punta y el precio fuera de la punta es, precisamente, la señal de mercado que justifica la incorporación de un producto potencia al cargo por confiabilidad: un mercado de pura energía firme no diferencia, ni remunera de manera explícita, la capacidad disponible en las horas críticas del día.

Aunque parece haber evidencia de que la restricción de confiabilidad de la potencia puede activarse la verdad es que añadir el producto potencia sólo tiene beneficios:

- a) Si el producto se necesita, habrá incentivos para proporcionarlo;
- b) Si el producto no se necesita (por exceso de oferta) su precio tenderá a cero.

En la siguiente Sección discutimos las propuestas de reforma del cargo.

2 Propuestas de reforma del cargo

En esta Sección del informe hacemos las propuestas generales de modificación del cargo por confiabilidad actual que resuelvan los problemas identificados. Las medidas son 7, que se complementan con medidas menores en la siguiente sección. Las principales medidas son:

- Volver a un único precio de escasez
- Incluir el producto potencia en punta;
- Reducir el número de vigencias a solo dos:
 - 15 años para nuevas y las especiales;
 - 5 años para el resto.
- Hacer subastas con vigencias a 5 años para sustituir el mecanismo de prorrata;
- Simplificar la liquidación de la OEF;
- Hacer que la OEF sea una obligación horaria (y no de seguimiento de la demanda); y
- Utilizar un mecanismo probabilístico para determinar el margen de reserva;

A continuación se presentan.

2.1 VOLVER A UN SOLO PRECIO DE ESCASEZ

La necesidad de volver a un único precio de escasez—bien fijado—es muy importante. A lo largo de su existencia el cargo por confiabilidad ha sido utilizado con diferentes fines y tener más de un precio de escasez significa que el fin del cargo ya no es remunerar la confiabilidad—expresada como suficiente capacidad/energía para atender la demanda más un margen de reserva. Este pago, como demuestra la teoría y práctica de los mercados, es equivalente a la anualidad de una central de punta y—como una central de punta es una opción call con precio de ejercicio igual al costo variable de dicha central—la equivalencia entre un pago por capacidad y la prima de dicha opción es clara.

La utilización de dos precios de escasez, inferior y superior, lleva a que el cargo ya no remunere confiabilidad porque el instrumento se convierte en una cobertura de precios para toda la demanda que compra energía en Bolsa. Esta cobertura existe al utilizar el precio superior de escasez, pero es resultado de comprar confiabilidad a un precio fijo (precio de escasez y prima) en el que la penalización por no entrega

es fácil de comprobar y del diseño colombiano que hace que toda la demanda se cubra.¹³

El argumento de un precio de escasez único es parte esencial del diseño, los otros precios de escasez surgieron de errores en la concepción del cargo porque, el de la Resolución CREG-071 de 2006, no se basó en una planta real del mercado colombiano. Como algunas plantas tenían asignaciones con ese precio de escasez, la corrección del precio de escasez para acercarlo a valores reales de una planta de punta llevó a la existencia de dos precios de escasez. Finalmente, la utilización del precio de escasez inferior es resultado de cierta desconfianza con el sistema marginalista, pero es claro que no hay manera de valorar un producto estandarizado de mejor manera.

Proponemos, dado el mejor desempeño con este precio utilizar el precio de escasez de la Resolución CREG-140 de 2017 y aprovechar que las plantas a las que les aplica el precio de escasez de las CREG-071 de 2006 están a punto de finalizar su asignación. Dar a las plantas con dos precios de escasez el incentivo a pasarse a un único precio de escasez.

2.2 INCORPORAR EL PRODUCTO POTENCIA

Como hemos discutido en la Sección anterior, la adición del producto potencia fue una asignatura pendiente del cargo por confiabilidad. Se dijo, pero no se implementó porque la potencia sobraba en ese entonces. Así, introducirlo no tiene un efecto sobre la subasta de energía firme—se compran las mismas cantidades—pero permite asegurar la cobertura en la punta de demanda del sistema. Si la energía firme permite cubrir la punta, el precio de potencia será cero, con lo cual es inocuo.

Es más, existe cierta evidencia, documentada anteriormente, que nos hace creer que se pueda estar necesitando el producto potencia en la punta (19:00-21:00). Por eso proponemos dos productos para asegurar confiabilidad:

- Potencia en punta: Disponibilidad comercial en los días de verano anual entre las 19:00-21:00.
- Energía firme: ENFICC.

Los siguientes apartados lo desarrollan.

2.2.1 El producto potencia estaba contemplado en el diseño

Aunque parece haber evidencia de que la restricción de confiabilidad de la potencia puede activarse la verdad es que añadir el producto potencia sólo tiene beneficios:

¹³ De hecho, otras opciones de confiabilidad no obligan a los generadores a seguir la demanda, sino a entregar el producto adjudicado hora a hora (pe. Italia, Nueva Inglaterra o Irlanda).

- a) Si el producto se necesita, habrá incentivos para proporcionarlo;
- b) Si el producto no se necesita (por exceso de oferta) su precio tenderá a cero.

Es decir, que el modelo original pretendía incluir el producto potencia, pero no lo incluyó porque la restricción no estaba activa. En palabras de Cramton y Stoft:

“In today’s market, the auctioneer would know that capacity will be in surplus, even at a zero price, and so the auctioneer would quickly let the capacity price fall to zero. This is why the two-product market outcome would be the same as in the single-product market. Once the capacity price is zero, all that matters to the supplier is the price of firm energy.” [pp 20]

2.2.2 Demanda de los dos productos

Los productos objeto de subasta son los siguientes;

1. Energía firme (OEF): opción americana en la cual la prima se fija en subasta con entrega horaria en la cual el generador recibe el Mín [Precio de Bolsa, Precio de escasez]; y
2. Potencia en la punta (PEP): en la cual el generador recibe un pago USD/MW determinado en subasta por su disponibilidad comercial en la punta de demanda anual del sistema [19:00-21:00] del verano anual.

Producto Potencia en la Punta (PEP)

El primer producto es el de la Resolución CREG-071 de 2006, con un único precio de escasez (resolución CREG-140 de 2017) y liquidación a cada precio de escasez. Es decir, una opción americana al precio de escasez de cada recurso.

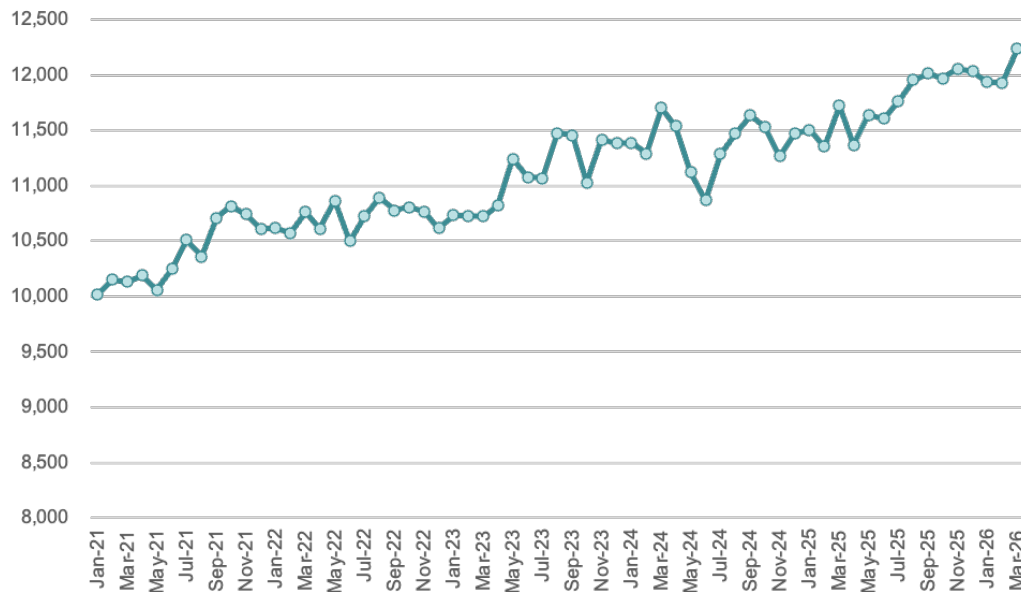
El producto potencia en la punta es el nuevo producto, intenta cubrir la demanda punta anual. En Colombia la demanda máxima de potencia viene subiendo de manera importante como puede apreciarse en la siguiente figura.

La demanda punta viene creciendo, pero no es estacional, como puede apreciarse en la siguiente gráfica en la que tomamos datos diarios y calculamos la demanda máxima del mes (a partir de la demanda máxima de potencia del día) desde 2021 hasta la fecha.

Como no hay una demanda máxima de potencia estacional y sí una estacionalidad de la oferta hidráulica con la consiguiente reducción de la Capacidad Efectiva Neta del parque hidráulico durante los meses de verano, lo más razonable es exigir entrega entre las 19:00-21:00 de los días del verano anual.

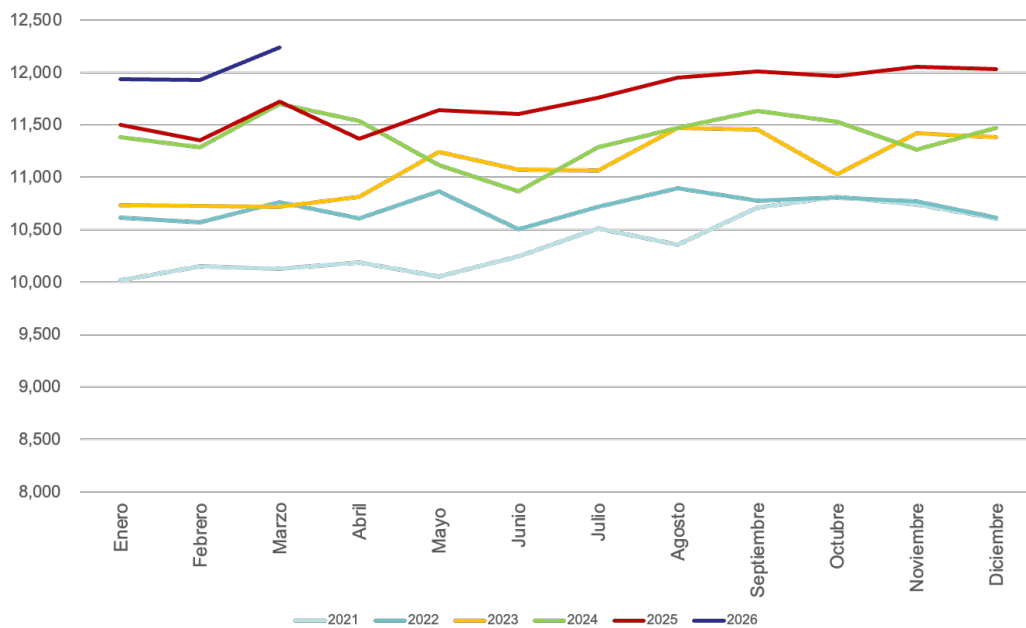
Para decidir las cantidades a comprar de cada producto, lo razonable son las proyecciones de demanda de la UPME, tanto de demanda de energía como de demanda de potencia. Pero a estos valores hay que agregarles un margen de reserva.

Figura 5: Demanda máxima de potencia mensual (MW) 2021-mar 2026



Fuente: XM Sinergox.

Figura 6: Demanda máxima de potencia del SIN



Fuente: XM Sinergox

Un margen de reserva para los dos productos deber ser probabilístico, basado en procesos estocásticos de la oferta como de la demanda. Estos se desarrollan más en detalle en el apartado adelante, pero aquí es suficiente decir que debe haber demandas objetivos basadas en anualidades de las centrales flexibles para cubrir la demanda esperada más un margen de reserva.

Propuestas de reforma del cargo

Una vez discutamos el margen de reserva también hablaremos de la curva de demanda del producto PEP.

Demanda de energía firme

La demanda de energía firme es la utilizada en la actualidad y su forma está relacionada, como en la actualidad, con la anualidad de una central de punta.

2.2.3 Oferta de los dos productos

Para definir la oferta se cuenta con generadores existentes y nuevos. Al igual que en el caso de la energía firme el regulador debe determinar las restricciones a las cantidades (ENFICC y Capacidad Efectiva Neta en la Punta CENP). Para las diferentes tecnologías se utiliza:

Energía Firme

Energía Firme. Se debe utilizar la metodología actual de cálculo de la ENFICC para no incurrir en debates complejos que pueden solucionarse con el producto potencia.

PEP

En este caso, también, depende de la tecnología:

Térmicas: CEN afectada por el IHF;

- Hidráulicas existentes: Min(Disponibilidad Comercial en la punta, generación en la punta).
- Hidráulicas nuevas: basadas en diseño y aportes.
- Fotovoltaicas sin baterías: Cero.
- Eólicas: basadas en min [generación en la punta, mapas de viento horarios]
- Fotovoltaicas con baterías: producción promedio continua durante las tres horas de la punta. Dentro de la demanda objetivo de energía debe incluirse la demanda de la batería.
- Baterías: cero, debe incluirse en la demanda objetivo y su carga debe pagar el cargo por energía firme y potencia.
- Respuesta de la demanda: de consumidores que tengan generación in situ.

Estas restricciones deben aplicarse de manera ex ante para la participación de los recursos en la subasta de los dos productos.

En la medida en que entren más renovables al sistema es posible que el producto tenga condiciones de rampas para los participantes.

2.2.4 Subasta de los dos productos

Los productos tienen relación entre sí para las diferentes tecnologías. Por ejemplo:

- Los productos son complementarios para las centrales flexibles porque se necesita estar disponible para poder entregar energía firme.¹⁴
- Para algunas plantas parcialmente flexibles es posible que los productos sean sustitutos (si se entrega energía en una hora puede que no se tenga potencia disponible para la punta);
- Para otros no hay relación, por ejemplo, una central fotovoltaica califica para OEF—con las reglas actuales—pero no tiene PEP.

En una subasta, para que sea eficiente, estas relaciones deben reconocerse por medio de una *subasta combinatoria* en la cual los agentes ofrecen paquetes de ofertas de cada producto. Estas subastas suelen ser dinámicas en las cuales los agentes van reaccionando a los precios que vean de uno u otro recurso. En cada ronda el subastador informa de los precios de la OEF y del PEP (P_{OEF} , P_{PEP}) y los postores ofertan cantidades de los dos productos. En ese proceso los precios descienden a velocidades distintas (función del exceso de oferta) y los agentes sólo pueden reducir sus ofertas de acuerdo con la reducción del precio.

La ventaja de que la subasta sea dinámica es función del valor del mercado de los productos. Es decir, cuando un postor no conoce cuánto vale su producto se beneficia de conocer la opinión del mercado. Sin embargo, las subastas dinámicas en mercados muy estrechos (donde la oferta potencial es un poco más que la demanda objetivo) son proclives al ejercicio de poder de mercado.

De hecho, el producto energía firme se vendía en subasta dinámica (mayo de 2008), pero desde que el escalón de precios de la segunda subasta (noviembre de 2011) se hizo muy alto—convirtiendo la demanda, en resumidas cuentas, a una subasta estática y tras las recomendaciones de Harbord y Pagnozzi¹⁵ la CREG ha utilizado subastas de sobre cerrado.

Así Cramton y Stoft proponían una subasta dinámica de package bidding basada en el modelo de subasta de Ausubel et al (2004)¹⁶ en el que la subasta es de reloj descendente con una ronda final. En el caso de la subasta de los dos productos colombiana en el que los postores conocen sus costos de inversión y la información

¹⁴ Diferente sería el caso si la OEF tuviera una fecha de entrega que no coincidiera con la entrega del producto PEP. Pero la OEF se debe entregar en todas las horas de la vigencia y esto hace que para las centrales que tienen OEF deban contar con PEP.

¹⁵ David Harbord y Marco Pagnozzi (2012) Second Review of Firm Energy Auctions in Colombia, 18 December 2012 estudio para la CREG

¹⁶ Ausubel, Lawrence M., Peter Cramton, y Paul Milgrom (2006), “The Clock-Proxy Auction: A Practical Combinatorial Auction Design” publicado en Cramton, Shoham, y Steinberg (eds.), Combinatorial Auctions, MIT Press, Chapter 5, 115-138, 2006.

sobre las variables de mercado es pública (precios de combustible, hidrologías, etc.) no hay mucho beneficio de proponer una subasta dinámica y las pérdidas de eficiencia de hacerla de sobre cerrado son pocas.

Así que podemos proponer una subasta de sobre cerrado, pero que reconozca las cantidades máximas (probablemente cantidades fijas para prevenir la reducción estratégica de oferta) que cada tecnología puede entregar y las restricciones que se introduzcan para plantas nuevas (p.e. entrega de energía firme si asignación de potencia). En este sentido se puede hacer una subasta a sobre cerrado en la cual la función objetivo sea:

$$\text{Min} \sum_{i,j}^{I,J} \pi_{OEF}^i OEF^i + \pi_{PEP}^j PEP^j$$

Donde:

π_{OEF}^i : la prima por OEF de la planta i de los I plantas adjudicadas;

π_{PEP}^j : la prima por PEP de la planta j de los J plantas adjudicadas;

OEF^i : asignación de OEF de la planta i de los I plantas adjudicadas

PEP^j : asignación de PEP de la planta j de los J plantas adjudicadas

La función objetivo está sujeta a varias restricciones:

- Se debe cubrir la demanda objetivo de energía firme;
- Se debe cubrir la demanda objetivo de la potencia en punta.
- Se deben respetar las condiciones técnicas de las plantas y las restricciones de los paquetes (x OEF si PEP > 0)

2.3 ACORTAR LAS VIGENCIAS PARA NUEVAS PLANTAS

La vigencia actual de 20 años es bastante larga, comparada con la vigencia en otros países o hasta con, dados los cambios tecnológicos, la vida útil de algunas tecnologías.

Con 15 años de vigencia (unos 3 Niños) nos acercaríamos a las vigencias de las plantas especiales y se acercaría a las vigencias de otros países de la OCDE. Así podríamos permitir que todas las plantas que fijan precios en la subasta estén activas en la subasta y no se genere el problema que tenemos hoy:

- Una planta con vigencia de 10 años si sale asignada es porque la de 20 era menos competitiva (los precios en 10 años iban a ser más altos);
- Una planta con vigencia de 10 años que no salga asignada es porque la de 20 era más competitiva (los precios de los primeros 10 años eran más altos).

Con 15 años también se gestiona un problema que el cargo tiene hoy día y es que está pensando sólo para incrementos en la demanda. Sabiendo que se vienen cambios en las tecnologías renovables y cambios tecnológicos en la gestión de la demanda es recomendable reducir las vigencias.¹⁷

Esta vigencia aplicaría a las plantas a las que hoy día les aplica vigencias de 20 y 10 años.

Para las demás plantas proponemos 5 años lo cual nos permite hacer subastas cuando aplique el mecanismo de prorrata.

2.4 SUBASTAS PARA SUSTITUIR EL MECANISMO DE LA PRORRATA

La manera de resolver el problema de la prorrata es llevar a cabo subastas cuando se requiera de nueva energía firme. Realizar subastas es lo más relevante para que se acaben las barreras a la entrada, pero es necesario que en esas subastas se valore el precio de la confiabilidad que, en Colombia, implica la posibilidad que exista un Niño.

Esto implica que la vigencia no podría ser inferior a los 3-5 años. Como la CREG fija tarifas por 5 años, asignaciones de 5 años para las demás plantas es razonable.

Ahora bien, esto significa que las plantas existentes compiten en los dos productos, potencia y energía en casos en los que hay exceso de demanda (no se requiere nueva entrada), pero esto significa que en estas subastas las pujas se hacen por prima y como el precio de escasez actual representa poco riesgo de recuperación de los costos variables, la competencia sesga a plantas que tienen costos fijos bajos y costos variables altos. La razón es que una planta de costo variable alto tiene un costo fijo más bajo con lo cual si la competencia se hace con base en el costo fijo la planta que tenga costo fijo más bajo es más competitiva.

Es cierto que éstas son las mejores plantas para contingencias, pero cuando hay exceso de energía firme lo importante es el costo total y por eso es mejor que las plantas existentes compitan en estos términos.

La propuesta es que en estos casos se compita en prima, pero que se incluya la posibilidad de ejercicio de la opción de energía firme, de tal manera que el algoritmo de casación sea de costo total:

$$\text{Min } \pi_i \text{ENFICC}_i + \text{ENFICC}_{ih}(p_h - CV_{ih})$$

s.a.

¹⁷ La principal razón para tener un cargo es que no hay respuesta de la demanda. Una duración muy larga quita oportunidades a la respuesta de la demanda,

$$\sum_i^I ENFICC_i \geq DO$$

En el que p_b es el precio de bolsa en el período crítico y DO la demanda objetivo. Los costos de combustible se simulan con base en proyecciones de terceros (p.e. AIE o precios de mercados de futuros, o Departamento de Energía de los EEUU, etc).

A esto se le añade la puja por el producto Potencia en Punta para seleccionar a los ganadores. Las empresas deben pujar toda su ENFICC y no hacer pujas parciales de ENFICC. Las plantas que no ganen pueden quedarse en hibernación.

2.5 SIMPLIFICAR LA LIQUIDACIÓN

La liquidación se ha vuelto muy compleja como resultado de varios precios de escasez y de los problemas de contratos. Ilustremos el caso de, primero, un único precio de escasez y múltiples precios de escasez.

2.5.1 Múltiples precios de escasez

En el caso de múltiples precios de escasez se debe exigir la OEF cada vez que el precio de Bolsa excede el precio de escasez respectivo. Esto significa que no toda la demanda está cubierta, pero se asigna el diferencial a prorrata de todas las compras en Bolsa.

En el caso de que haya vigencias con más de un precio de escasez, lo razonable es hacer una liquidación como la que se contempla a continuación.

2.5.2 Un único precio de escasez

La lógica de la liquidación es asegurarse que la OEF no distorsione su contrato. Esto implica que el generador debe pujar la compra de los contratos que sean despachados por medio de la puja de una cantidad de compra a la Bolsa.

El generador presenta al ASIC:

- Su disponibilidad comercial horaria;
- Su precio de oferta de generación; y
- La oferta de recompra de su contrato.

De esta manera, la curva de beneficios del generador se basa en esta ecuación:

$$\pi = p(Q_i - Q_c) + p_c Q_c + p_f Q_f - \text{Max}[p - p_e, 0] Q_f$$

Donde las cantidades relevantes son;

Q_i : disponibilidad en el ideal;

Q_c : Cantidad en contratos (nótese que se puja a la Bolsa como oferta de compra de energía)

Q_f : OEF horaria;

Y las demás variables son p , para precio de Bolsa, p_e para precio de escasez y p_f como cargo por confiabilidad.

La liquidación del cargo es sólo función de p y p_e como tiene que ser. La liquidación de contrato es sólo función de p y p_c , como tiene que ser.

Cuando se ejerce la OEF:

$$\pi = p(Q_i) - [p - p_e]Q_f;$$

Que es función del precio spot, únicamente, porque p_e y Q_f están fijas.

Cuando se ejerce el contrato:

$$\pi = p(Q_i - Q_c) + p_c Q_c$$

Con lo cual los contratos son solo función del precio de Bolsa, de la generación ideal (función del costo variable) y del precio de los contratos. La OEF no entra en la ecuación.

Cuando se ejerce el contrato y la OEF:

$$\pi = p(Q_i - Q_c) - \text{Max}[p - p_e, 0]Q_f$$

Que implica que la cantidad contratada y la OEF no se condicionan entre sí.

2.5.3 Dos precios de escasez

Hemos dicho que lo ideal es que haya un solo precio de escasez, pero existe la posibilidad que en la transición convivan dos precios de escasez. Aunque lo natural es que las opciones de cada uno se ejerzan cuando el precio de Bolsa supere su precio de escasez, la CREG ha permitido que se use un solo precio de escasez de activación. Esto cambia la función de beneficios a la siguiente:

$$\pi = p(Q_i - Q_c) + p_c Q_c + p_f Q_f - [p - p_e]Q_f \text{ [si } p \geq p_{ea}]$$

$$\pi = p(Q_i - Q_c) + p_c Q_c + p_f Q_f - [si } p < p_{ea}]$$

Lo cual introduce la no linealidad de la que hablamos en la Sección 1, pero que permite que la liquidación de la OEF y la de contratos no se solapen.

Sin embargo, toda esta simplificación es más sencilla si la OEF se vuelve horaria e igual a la asignada en subasta que es nuestra propuesta a continuación.

2.6 OEF HORARIA

En el diseño actual la entrega de OEF es diaria, con lo cual es exigible cuando se excede el precio de escasez en una hora dada.

Propuestas de reforma del cargo

La asignación de OEF es horaria y exigible cada vez que $P_b > \text{Pescasez}$. Esto puede llevar a que se cubra más o menos que la demanda horaria, es decir que sobre dinero o que falte para que la energía comprada en Bolsa tenga el precio de escasez como techo. Pero permite reducir el costo de hacer seguimiento de la demanda, simplificar la liquidación y brindar incentivos claros para ofertas de producción.

2.7 ADOPTAR UN MARGEN DE RESERVA TÉCNICO Y ECONÓMICO

El cargo por confiabilidad vigente carece de una señal explícita de margen de reserva en potencia para la hora crítica. Tal como está construido, la demanda objetivo se expresa como un bloque de energía firme prácticamente plano—la ENFICC—que aproxima la confiabilidad a la suficiencia energética del año hidrológico, pero que es ciego al momento en que el sistema enfrenta realmente la probabilidad de déficit. Como discutimos en la Sección 2.2, la penetración creciente de generación solar fotovoltaica ha desplazado el riesgo sistémico desde la suficiencia energética diaria—el problema hidrológico tradicional del Niño—hacia la suficiencia de potencia horaria en la rampa de la tarde-noche (19:00–21:00), momento en el que el aporte solar cae a cero con independencia de la ENFICC diaria asignada. Un diseño de dos productos—energía firme y potencia en la punta—exige, por tanto, un criterio explícito y defendible para dimensionar la demanda objetivo de cada producto y el nivel de aseguramiento requerido.

Esta sección formaliza ese criterio. Su propósito es servir de puente regulador y matemático entre el estándar de confiabilidad—que debe fijarse con fundamento económico—y la restricción de demanda vinculante del Producto Potencia en la subasta co-optimizada de la Sección 2.2. El margen de reserva no es un parámetro arbitrario ni un colchón administrativo: es el resultado de una optimización del costo total del sistema bajo incertidumbre, que internaliza de manera explícita las indisponibilidades del parque generador existente. Seguimos, en lo conceptual, la metodología de evaluación de adecuación de recursos de la Unión Europea (ERAA), adoptada mediante la Decisión ACER 24/2020, por ser el estándar internacional más riguroso en el tratamiento probabilístico de la suficiencia.

2.7.1 Determinación Económica del Estándar de Confiabilidad Óptimo

El nivel óptimo de confiabilidad de un sistema eléctrico no es una decisión técnica arbitraria, sino un equilibrio económico. Añadir capacidad reduce la energía esperada no suministrada, pero a un costo de inversión creciente; el nivel óptimo es aquel que minimiza el costo total del sistema, entendido como la suma del costo de invertir en capacidad nueva y el costo esperado de la energía que el sistema no logra abastecer. Formalmente, si denotamos por K la capacidad firme instalada, el problema del planificador es minimizar:

$$CT(K) = \text{CONE} \cdot K + \text{VOLL} \cdot \text{EENS}(K)$$

donde $CT(K)$ es el costo total del sistema; CONE (Cost of New Entry) es el costo anualizado de la tecnología de punta marginal—típicamente una turbina de gas de ciclo abierto o, crecientemente, un sistema de almacenamiento en baterías (BESS)—expresado en USD/MW-año; VOLL (Value of Lost Load) es el valor de la energía no suministrada para la economía, equivalente al costo de racionamiento, expresado en USD/MWh; y $\text{EENS}(K)$ es la energía esperada no suministrada (Expected Energy Not Served), función decreciente y convexa de la capacidad instalada.

La condición de primer orden de este problema iguala el costo marginal de añadir un megavatio de capacidad con el beneficio marginal de la reducción del racionamiento esperado:

$$\frac{dCT}{dK} = \text{CONE} + \text{VOLL} \cdot \frac{d\text{EENS}}{dK} = 0$$

Como la derivada de la energía esperada no suministrada respecto de la capacidad, con signo cambiado y agregada a lo largo del año, es precisamente el número esperado de horas en que la unidad marginal resulta indispensable—es decir, la Expectativa de Pérdida de Carga (LOLE, Loss of Load Expectation)—la condición de óptimo se reescribe como:

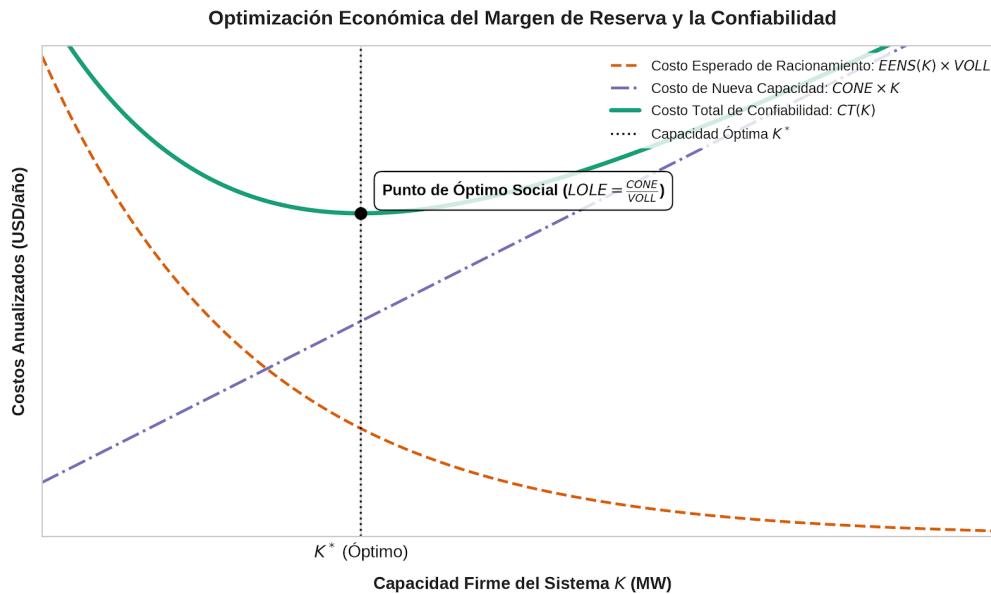
$$\text{CONE} = \text{VOLL} \cdot \text{LOLE}^*$$

de donde se obtiene la regla de oro del dimensionamiento de la confiabilidad:

$$\text{LOLE}^* = \frac{\text{CONE}}{\text{VOLL}}$$

Esta es la formulación económicamente correcta del estándar: el LOLE^* óptimo, expresado en horas/año, es igual al cociente entre el costo anualizado de la nueva entrada (USD/MW-año) y el valor de la energía no suministrada (USD/MWh), cuyas unidades se cancelan en horas/año. Conviene advertir que la versión que circula en algunos documentos regulatorios, que invierte el cociente, es dimensional y conceptualmente incorrecta: a mayor costo de racionamiento (VOLL), el sistema debe tolerar menos horas de déficit, no más, lo que solo es consistente con el LOLE en el numerador del CONE . La virtud de este método es que evita por construcción tanto el sobredimensionamiento—que encarece la tarifa sin beneficio de confiabilidad proporcional—como el subdimensionamiento de las Obligaciones de Energía Firme y de Potencia. La Figura 7 ilustra gráficamente esta lógica de optimización: el óptimo social se alcanza en el punto en que la curva de costo total de confiabilidad—suma del costo de nueva capacidad y del costo esperado de racionamiento—es mínima, condición que coincide exactamente con el criterio $\text{LOLE} = \text{CONE}/\text{VOLL}$.

Figura 7: Curva de optimización del costo total de confiabilidad



Fuente: elaboración propia.

2.7.2 Traducción del Estándar Probabilístico a Margen de Reserva Operativo

El estándar $LOLE = CONE/VOLL$ es de naturaleza probabilística y se expresa como una expectativa estadística. Para que sea operativo en una subasta de sobre cerrado se requiere traducirlo a un indicador determinístico: el Margen de Reserva Objetivo (Target Reserve Margin, RM). El margen de reserva es la capacidad firme que el sistema debe contratar por encima de la demanda máxima proyectada para que, considerando la incertidumbre de la oferta y de la demanda, el LOLE resultante no supere el estándar óptimo. Si $K^*(LOLE^*)$ es la capacidad firme que mantiene el LOLE en su valor objetivo, el margen de reserva se define como:

$$RM^* = \frac{K^*(LOLE^*)}{D_{\text{punta}}^{\text{máx}}} - 1$$

La determinación rigurosa de $K^*(LOLE^*)$ exige un modelo probabilístico de simulación, en línea con la metodología ERAA de ACER. Dicho modelo construye, mediante simulación de Monte Carlo, miles de estados posibles del sistema combinando: (i) años climáticos históricos—que capturan la correlación entre hidrología, irradiancia solar, velocidad de viento y temperatura, y por tanto los eventos de Niño; (ii) patrones aleatorios de salidas forzadas de las unidades de generación; y (iii) los factores de carga horarios de las fuentes variables, que determinan su aporte efectivo en la punta. La adecuación se evalúa sobre el despacho económico de cada estado simulado, y el criterio de convergencia se controla sobre el coeficiente de variación de la energía esperada no suministrada.

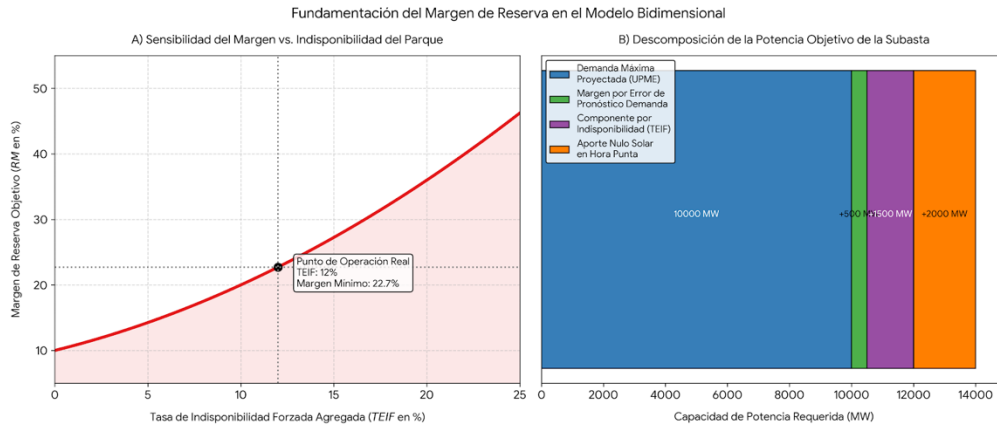
El punto crítico—y el que la metodología colombiana ha omitido sistemáticamente—es que el cálculo del margen de reserva debe internalizar de forma explícita las indisponibilidades forzadas y técnicas del parque generador existente. La capacidad nominal instalada no es capacidad firme: una fracción de ella está indisponible en cualquier momento dado por mantenimientos y fallas. La capacidad firme con la que cada planta i contribuye a la punta debe deratearse por su tasa de indisponibilidad forzada de salida (TEIF, o su equivalente IHF en la regulación vigente):

$$q_{\text{punta},i} = \text{CENP}_i \cdot (1 - \text{TEIF}_i)$$

donde $q_{\text{punta},i}$ es la capacidad firme aportada por la planta i en la punta; CENP_i es su Capacidad Efectiva Neta en la Punta; y TEIF_i es su tasa de indisponibilidad forzada de salida. Para las plantas térmicas, la TEIF refleja las fallas no programadas históricas; para las hidráulicas, la disponibilidad comercial efectiva en la punta durante el verano; y para las fuentes variables, el aporte se determina por sus factores de carga horarios—nulo para la fotovoltaica sin almacenamiento en la franja 19:00–21:00—y no por su capacidad nominal.

La consecuencia es directa: un margen de reserva que ignore la tasa de falla del parque existente subestima el riesgo real de desatención en la punta nocturna. A medida que la indisponibilidad agregada del parque aumenta—por el envejecimiento del parque térmico, por la concentración de mantenimientos o por la salida simultánea de unidades durante un Niño—la capacidad K^* necesaria para mantener constante el LOLE objetivo debe ensancharse, y con ella el margen de reserva nominal. Dicho de otro modo, el margen de reserva no es una cifra fija: es una función creciente de la indisponibilidad esperada del parque instalado. Este es el argumento central que sustenta la adopción de un margen probabilístico. La Figura 8 cuantifica este argumento. El panel A muestra cómo el margen de reserva objetivo se ensancha de forma convexa a medida que aumenta la tasa de indisponibilidad forzada agregada del parque—alcanzando, por ejemplo, un margen mínimo del 22,7% para una TEIF del 12%—; el panel B descompone la potencia objetivo de la subasta en la demanda máxima proyectada por la UPME y los componentes incrementales por error de pronóstico de demanda, por indisponibilidad forzada del parque y por el aporte nulo de la generación solar en la hora punta.

Figura 8: Sensibilidad del margen de reserva frente a la indisponibilidad del parque existente



Fuente: elaboración propia.

2.7.3 Formalización Matemática en la Subasta Bidimensional

El margen de reserva se integra al diseño de mercado como la restricción de demanda del Producto Potencia en la subasta co-optimizada de la Sección 2.2. Recordemos que la subasta de sobre cerrado minimiza el costo total de adquirir los dos productos:

$$\text{Mín} \sum_{i,j} \pi_{\text{OEF}}^i \text{OEF}^i + \pi_{\text{PEP}}^j \text{PEP}^j$$

sujeta, entre otras, a dos restricciones de suficiencia. La primera garantiza la suficiencia energética—el problema hidrológico tradicional—exigiendo que la energía firme contratada cubra la demanda objetivo de energía:

$$\sum_i \text{ENFICC}_i \geq \text{DO}_{\text{energía}}$$

La segunda—la innovación que esta propuesta formaliza—garantiza la suficiencia de potencia en la hora crítica. La demanda máxima de potencia proyectada se escala mediante el margen de reserva para establecer el requerimiento vinculante de capacidad firme despachable en la punta:

$$\sum_i (x_i \cdot q_{\text{punta},i}) \geq D_{\text{punta}} \cdot (1 + \text{RM})$$

donde x_i es la asignación de la planta i en la subasta; $q_{\text{punta},i}$ es su capacidad firme en la punta, ya derateada por indisponibilidad según la ecuación anterior; D_{punta} es la demanda máxima de potencia proyectada por la UPME; y RM es el margen de reserva objetivo. La restricción obliga al mercado a contratar capacidad firme despachable suficiente para cubrir la punta y, además, las desviaciones por

indisponibilidad simultánea de unidades, que es justamente lo que el margen de reserva internaliza.

La interacción de esta restricción con la de suficiencia energética es la que confiere valor económico al atributo de la despachabilidad. En el problema co-optimizado, la restricción de potencia tiene asociado un precio sombra—el multiplicador de Lagrange $\lambda(\text{punta})$ —que representa el costo marginal de aumentar en un megavatio el requerimiento de capacidad firme en la punta. Este precio sombra es el precio del Producto Potencia. Su comportamiento es coherente con el diseño original de Cramton y Stoft: cuando la energía firme contratada basta para cubrir la punta—es decir, cuando la restricción de potencia no es activa— $\lambda(\text{punta})$ es igual a cero y la incorporación del producto resulta inocua; cuando la punta nocturna se vuelve la restricción vinculante, $\lambda(\text{punta})$ se torna positivo y remunera explícitamente la capacidad despachable disponible en las horas críticas, enviando la señal de inversión que el bloque plano de energía firme no es capaz de transmitir.

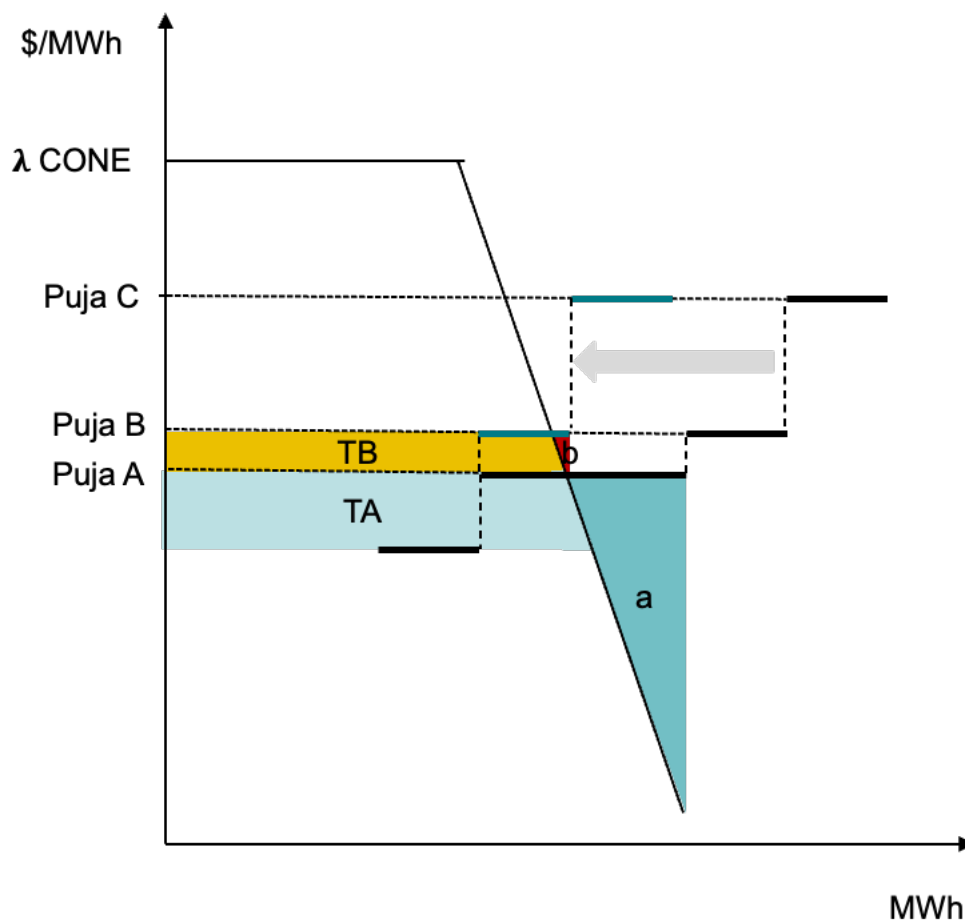
3 Medidas marginales de mejora

En esta Sección proponemos medidas que pueden tener un efecto más marginal, pero que se pueden considerar en una mejora posterior del cargo.

3.1 ALGORITMO DE CASACIÓN DE LA SUBASTA

El precio de mercado es aquel punto en el que la oferta y la demanda son iguales. Pero en la OEF el precio de mercado es aquel en el cual la planta marginal cubre la demanda, pero siempre ocurre que se compra mayor cantidad. El siguiente gráfico lo ilustra.

Figura 9: Pérdidas por indivisibilidad de ofertas de la planta marginal



La curva de demanda es la fijada por el regulador y la curva de oferta son las pujas de las plantas (A, B y C). Con el algoritmo de cierre actual la central A es la que fija el precio de la OEF en un valor de Puja A \$/MWh. Pero aceptar esta central implica comprar más MWh de los considerados necesarios generando una pérdida de bienestar del valor del triángulo a. Reemplazar la planta A por la planta B (de

menor tamaño) genera una pérdida de bienestar del triángulo **b** que es inferior a pesar de que los usuarios pagan un menor precio por confiabilidad.

Ahora bien, es cierto que el precio es más alto, pero esto es una transferencia de consumidores a productores con lo cual el regulador puede poner diferentes pesos a las transferencias. Este es una mejora marginal, pero busca eliminar el exceso de compra de ENFICCs que surge por la indivisibilidad de la planta marginal.

3.2 PARTICIPACIÓN DE LAS GPPS EN LA SUBASTA PRIMARIA

El modelo colombiano, inspirado por Stoft y Cramton, diverge del modelo de los mismos autores para Nueva Inglaterra en la existencia de las subastas GPPS que se consideran necesarias por las consideraciones asociadas al período de planificación. El período de planificación es fundamental para el desempeño de la subasta, Cramton y Stoft proponen 4 años:

“(...) a planning period of four years. This is far enough ahead of the commitment that potential new projects can compete in the auction before substantial costs have been sunk in the project. In this way, the bids from new resources can reflect the cost of new entry. It makes the firm energy market contestable and allows new entry to set the price. This long planning period is essential. Existing resources would set the wrong price because of sunk costs and market power. We must rely on new entry to set the price.”

El diseño de estas subastas requiere que la nueva entrada pueda disciplinar a los generadores existentes porque de otra manera – dado el equilibrio entre oferta y demanda – muchos de los generadores existentes son pivotaes. Además, el diseño buscado en el que por medio de subastas se determine el precio necesita necesariamente de nuevos entrantes porque el costo de la capacidad para generadores existentes puede ser bajo debido a los altos costos hundidos de la actividad. Es decir, el período de planificación debe ser igual al tiempo de construcción de una central de punta que es de 3-4 años que sería la central que podría disciplinar el mercado si se vieran precios de confiabilidad muy altos.

Las subastas confían en que un participante construya una central con miras a cumplir únicamente con la opción de energía firme y no tener que confiar en otros ingresos del mercado para financiar la inversión. El valor de esa inversión es el valor de la opción – siempre y cuando exista la posibilidad de escasez – y no un costo de inversión diferente.

Un mayor período de planeación hace que:

- Aumente la incertidumbre respecto a la energía firme que se quiere comprar;
- Aumente la incertidumbre de los precios de energía;
- Aumente el riesgo de no-construcción (retraso) porque aumenta el tiempo entre el resultado de la subasta y la puesta en marcha de la central.

Medidas marginales de mejora

En Colombia en lugar de unificar el período se decidió dividir la subasta en dos lo cual es una diferenciación del precio de un mismo producto. La diferenciación no es tan grave en la medida en que los períodos de vigencia no coinciden, pero el diseño de la subasta hace que el precio resultante de la subasta primaria sea un tope para las GPPS lo cual no tiene por qué ser así.

Las subastas de GPPS fueron introducidas porque se creía que de otra manera se discriminaría contra inversiones con plazos de duración superiores al período de planificación. La justificación de dicha subasta la presenta Cramton en su documento para la CREG.

“In light of the importance of longer-lead-time hydro projects in Colombia, a four year planning period is probably best. However, a shorter planning period can be used in the first two auctions as follows:

- 1. First auction, November 2007 (2011 commitment): 3 years*
- 2. Second auction, May 2008 (2012 commitment): 3.5 years*

All later auctions (2013 commitment and beyond) would have a four-year planning period.

Certainly, some projects, such as a large hydro project cannot be completed in four years. These projects may take 6-8 years to complete. For large hydro projects, we allow the investor to lock-in the auction price from the 4-year ahead auction up to seven years ahead. The large hydro project is a price taker in the auction, since it is not selling firm energy four years [sic] ahead, but up to seven years ahead. At the conclusion of the auction, the investor specifies the fraction of the firm energy from the project it desires to lock-in at 4-year ahead auction price.”

Harbord y Pagnozzi plantearon unificar el período de planificación por medio de su alaruge es decir hacer el período de planeación igual al máximo en lugar de igual al mínimo. Esta propuesta tiene como fin aumentar la presión competitiva en la subasta, pero con la premisa que, al igual que Cramton y Stoft, que los recursos necesitan la seguridad de ganarse la opción de energía firme para desarrollarse. Seguridad de recuperación de parte de la inversión no es la característica necesaria para un mercado competitivo aún con los problemas exclusivos del sector eléctrico.

Pero no es evidente que una central hidroeléctrica se construya porque el cargo por confiabilidad sea alto o tenga un período de planificación superior a los 4 años. De hecho, la mayor parte de los ingresos de una central de éstas proviene de contratos a plazo y ventas en la Bolsa con lo cual el riesgo primordial de la central es el riesgo de mercado y no el riesgo de un pago como el cargo por confiabilidad el cual es un pago durante un plazo muy inferior a la vida útil de estas centrales y a sus costos de inversión.

La parte en riesgo en un mercado de energía es la anualidad de una central de punta que solo es una parte del costo de inversión y el monto que el regulador debe dar la perspectiva de recuperar a las centrales. Pero esto no significa seguridad de recuperación porque tal seguridad hace que el mercado sea muy susceptible a las

decisiones regulatorias y a las equivocaciones en el valor de la energía firme a comprar en el mercado. El mercado colombiano – por medio del cargo por confiabilidad – ha pasado de ser un mercado con libre entrada a un mercado con entrada asistida por lo cual es muy susceptible al tamaño de tal asistencia.

Esta incertidumbre aumenta con el período de vigencia y con el período de planeación. Por lo tanto, lo sensato es unificar el período de planeación, pero por la parte baja y eliminar las subastas GPPS. La única motivación de las GPPS es la creencia que las centrales hidráulicas no se constituirían sin la seguridad de este pago, pero en Colombia lo han hecho sin asegurarse toda su energía firme en la subasta, lo han hecho con la perspectiva de poder acceder al cargo en el futuro.

Por eso lo que sería recomendable es hacer una sola subasta y permitir la participación en ella a centrales en obras. Así una central hidráulica con 10 años de tiempo de construcción participaría en las subastas a los 6 años de haber empezado su construcción y así poder cumplir con la entrega en 4 años.

Eso permitiría:

- Tener mayor presión competitiva en la subasta;
- Tener el mismo precio para opciones que cubren el mismo período (hoy en día se tiene el mismo precio para opciones con períodos diferentes);
- Reducir la incertidumbre respecto a la cantidad de energía firme que se quiere comprar; y
- Tener mayor certeza de la probabilidad de cumplir de los recursos comprometidos.

Aunque la CREG no la ha utilizado desde el 2011, sería conveniente derogar la parte pertinente de la Resolución CREG-071 de 2006 y la Resolución CREG-040 de 2008.

3.3 MODIFICACIÓN DE LAS GARANTÍAS

El documento de la CREG 027, de 10 de mayo de 2007, planteó una metodología de valoración de garantías con el objetivo de:

- Asegurar el cumplimiento del compromiso adquirido;
- Vincular su monto al perjuicio que causaría al sistema su incumplimiento;
- Asegurar un margen temporal suficiente para que el sistema pueda reaccionar en caso de incumplimiento.

En su análisis, la CREG determinó una fórmula que relacionaba el valor de la garantía al costo, la probabilidad y la duración de un evento de Niño, tal y como se muestra a continuación:

$$\begin{aligned} & \text{Valor garantía anual} \\ &= OEF \times \text{Precio de cobertura} \\ &\times \text{Probabilidad de ocurrencia del Niño en el próximo año} \\ &\times \text{Duración histórica del Niño} \times 30\% \end{aligned}$$

Donde:

OEF corresponde al valor de las obligaciones de energía firme anuales asignadas mediante subasta;

Precio de cobertura corresponde al precio (estimado) por encima del precio de escasez que debería enfrentar el sistema por compras en período de escasez. La CREG proponía diferentes métodos para estimar el precio de cobertura: valoración de opciones, valoración de opciones con componentes de saltos, simulación de Montecarlo o mediante un modelo de despacho óptimo. Finalmente, la CREG decide emplear la diferencia entre el costo de racionamiento y el precio de escasez.

De acuerdo con el análisis inicial de la CREG, la garantía se aplicaría durante un período de tres años. Para plantas existentes, se depositarían durante los tres años antes del período de vigencia de las OEF, mientras que para las nuevas instalaciones tendrían efecto durante los tres años a partir del período de vigencia de las OEF.

Sin embargo y una vez efectuado el análisis previo, la CREG tomó la decisión de desvincularse de la metodología de valoración, y recomendó que el valor de la garantía fuera el equivalente al 100% de los ingresos estimados por las OEF durante un período de tres años. En caso de incumplimiento de fechas que generaran un retraso más allá de la fecha de inicio del Período de Vigencia de la Obligación, el valor de la garantía podría llegar a triplicar la cantidad inicial.

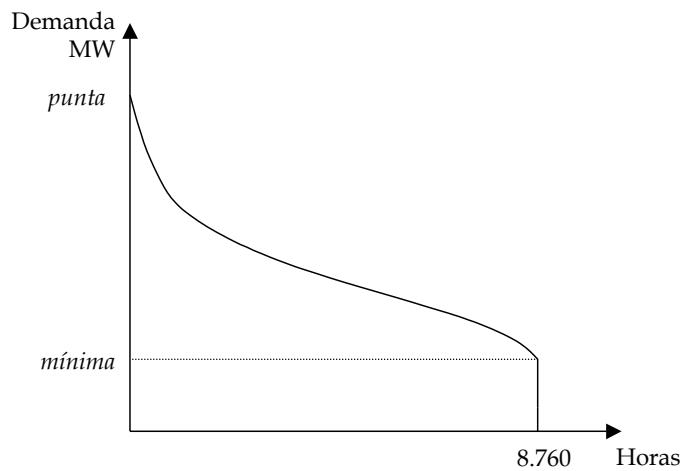
Esta recomendación ha venido siendo aplicada por el ASIC hasta la actualidad. A diferencia de la metodología previa, que vinculaba el valor de la garantía al perjuicio económico y la probabilidad de un evento de incumplimiento del compromiso adquirido con la OEF, el mecanismo actual no responde a ningún criterio económico.

La CREG había establecido la metodología correcta para fijar unas garantías que reflejaron el daño que podría hacer la planta en cada momento. Esta metodología es la que debe utilizarse y no una basada en el ingreso por OEF.

Anexo 1: Remuneración de confiabilidad

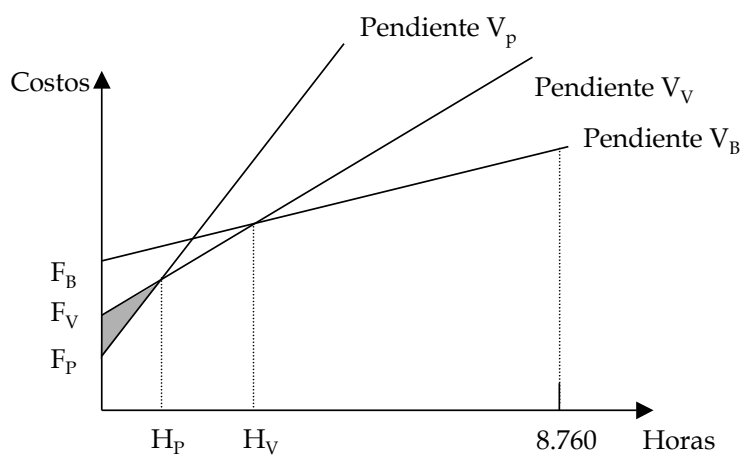
Para ilustrar la manera en que se recuperan costos fijos y variables en el mercado eléctrico utilizamos un modelo tradicional en la industria eléctrica. Al ordenar la demanda horaria de energía para las diferentes 8.760 horas del año de mayor demanda (la demanda punta del año) hasta la hora con menor demanda (la demanda mínima), obtenemos la curva monótona de carga que se presenta en el gráfico a continuación.

Figura 10: Curva monótona de carga



Para abastecer esta demanda suponemos el sistema cuenta con tres tecnologías; la B (central de base), la V (central de valle) y la P (central de punta), cada una con costos fijos y costos variables (F_B , F_V , F_P y V_B , V_V , V_P). El siguiente gráfico ilustra los costos de producción de las tres tecnologías para producir durante 8.760 horas del año.

Figura 11: Tecnologías disponibles

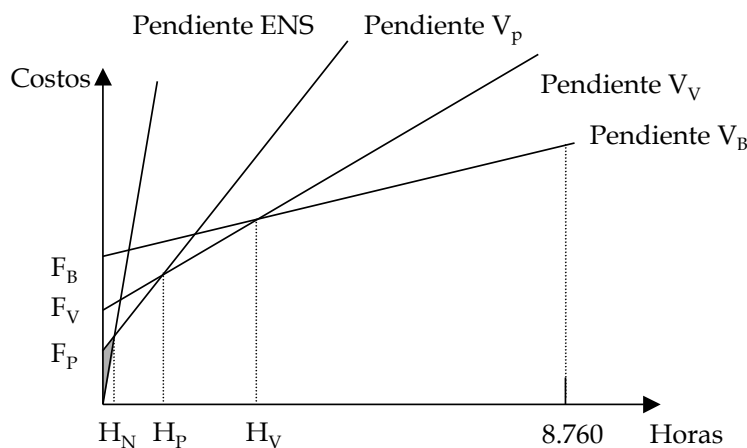


Como ocurre en la práctica, la tecnología de punta tiene el costo variable más alto pero el costo fijo más bajo, la tecnología de base tiene el costo fijo más alto pero el variable más bajo, mientras que la central de valle es la central intermedia en los dos costos. Para explicar en qué momentos es más competitiva la central de punta que la central de valle, nótese que el costo total de funcionar H_P horas usando la tecnología de punta es el área bajo la recta de costos de las dos tecnologías. Es sencillo notar que los costos totales de funcionar H_P horas para la tecnología de valle son superiores en una cantidad igual al triángulo sombreado. Por este motivo no se construyen centrales de base para cubrir las puntas de demanda sino que se construyen centrales de punta las cuales tienen un costo de operación más elevado pero un costo total menor.

Con el mismo razonamiento se puede apreciar que la central de valle es más competitiva que cualquier otra central en las horas $H_V - H_P$, y la central de base es más competitiva en las $8.760 - H_V$ horas restantes.

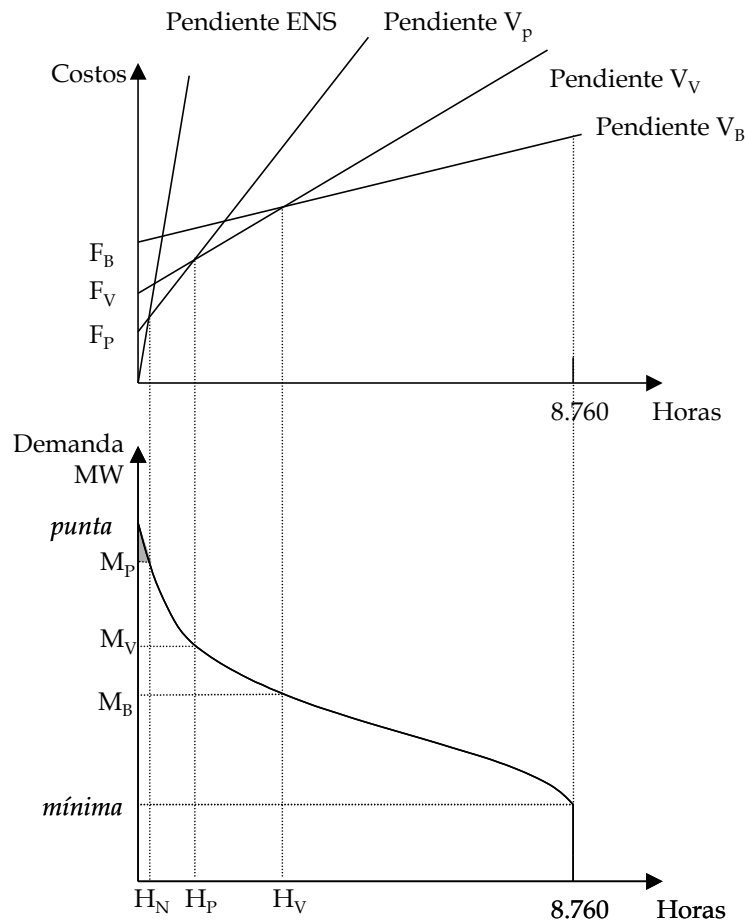
Sin embargo, para un número de horas existe una “tecnología” más competitiva que la central de punta: la energía no suministrada (ENS o Costo de Racionamiento, como se denomina en Colombia). La ENS no tiene asociado un costo fijo, pero sí un costo variable muy elevado; este costo se interpreta como el perjuicio que ocasiona a los consumidores no suministrados el no recibir la energía que han demandado, y se mide como el valor unitario que dicha energía tiene para ellos.

Figura 12: La tecnología de ENS



Si el planificador central pudiera decidir qué tecnología utilizar para abastecer la demanda del sistema su decisión sería la de minimizar el costo total de suministro. Esto lo haría juntando la oferta potencial de electricidad con la demanda de electricidad. El resultado sería como lo ilustra el gráfico abajo.

Figura 13: Capacidad instalada óptima



Una expansión óptima haría que se construyeran M_B megavatios de base, $M_V - M_B$ megavatios de valle, $M_P - M_V$ megavatios de punta y que se racionara el consumo por una cantidad (en MW) de (*punta* - M_P) o una cantidad de energía representada por la figura sombreada en el diagrama inferior de la Figura.

RECUPERACIÓN DE COSTOS EN UN MERCADO

En un sistema dimensionado de manera óptima, los ingresos de cada generador son iguales a sus costos totales, es decir, a la suma de costos fijos y variables. Si el tipo de central que se construye no es el óptimo, los inversionistas no podrán recuperar el coste de su inversión. Por lo tanto, en un sistema competitivo los inversores construirán el mismo tipo de capacidad que la que construiría un planificador central.

Si se construyeran estos megavatios, los precios en el mercado spot vendrían dados por el costo variable unitario de la central marginal, excepto en los casos en que la demanda exceda a la capacidad instalada (en esas horas el precio subirá hasta el

valor de la ENS). Así desde la hora H_B hasta la hora 8.760, el precio del mercado será igual al coste variable, V_B , de generación de las centrales de base.¹⁸

Central de punta

Con estos precios de mercado los ingresos de las centrales de punta serían:

De la hora 0 a la hora H_N el valor de la ENS, es decir:

$$H_N \times ENS \times M_P; \text{ y}$$

De la hora H_N a la hora H_P el costo variable de las centrales de punta:

$$(H_P - H_N) \times V_P \times M_P.$$

Sus ingresos totales son la suma de estos dos $M_P \times [(H_N \times ENS) + (H_P - H_N) \times V_P]$.

Para demostrar que este ingreso es igual al costo total de la central de punta puede apreciarse que la curva de costo total de la energía no suministrada se intersecta con la curva de costo total de la central de punta en la hora H_P . Así:

$$H_N \times ENS = F_P + H_N \times V_P;$$

O, sustituyendo el primer término, los ingresos de la central de punta son equivalentes a:

$$M_P \times [F_P + H_N \times V_P + (H_P - H_N) \times V_P];$$

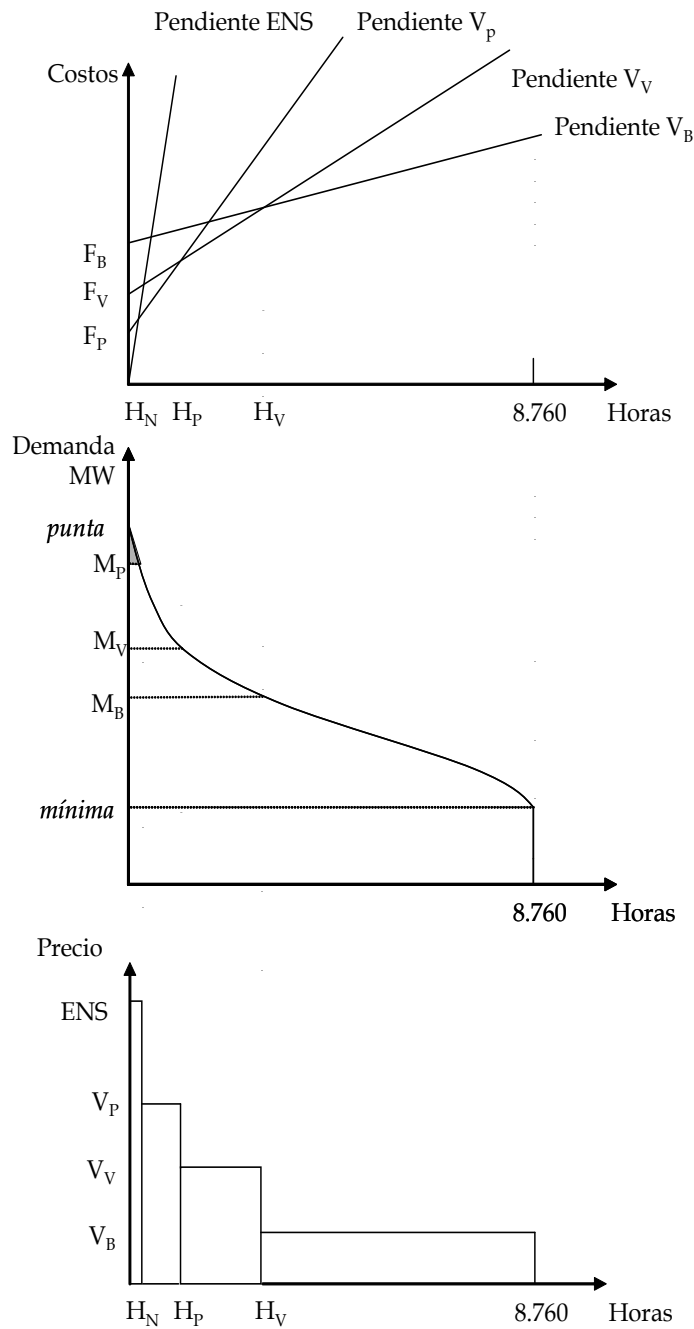
lo cual es igual a:

$$M_P \times [F_P + H_P \times V_P].$$

Este es el costo total de todas las centrales de punta. Es decir, si el precio de mercado puede subir hasta el valor de la energía no suministrada (costo de racionamiento de energía) en las horas en que hay racionamiento, todos los megavatios de punta en este sistema recuperan sus costos fijos y variables de producción.

¹⁸ De igual manera, en las horas 0 a H_N el precio sería ENS, en las horas H_N a H_P el precio sería V_P , y en las horas H_P hasta la hora H_V el precio sería V_V .

Figura 14. Recuperación de costos en un mercado de pura energía



La derivación del resultado que en un mercado competitivo se recuperan los costos de operación y de inversión es similar para las otras dos tecnologías.

Central de valle

Las centrales de valle reciben los siguientes ingresos:

De la hora 0 a la hora H_N el valor de la ENS, es decir:

Anexo 1: Remuneración de confiabilidad

$$H_N \times ENS \times M_V;$$

De la hora H_N a la hora H_P el costo variable de la central de punta:

$$(H_P - H_N) \times V_P \times M_V; \text{ y}$$

De la hora H_P a la hora H_V el costo variable de la central de valle:

$$(H_V - H_P) \times V_V \times M_V.$$

Sus ingresos totales son la suma de estos tres:

$$M_V \times [(H_N \times ENS) + (H_P - H_N) \times V_P + (H_V - H_P) \times V_V].$$

Debe recordarse que la curva de costos de las centrales de punta se intersecta con la ENS en el punto:

$$H_N \times ENS = F_P + H_N \times V_P;$$

y que la curva de costos de la central de valle se intersecta con la curva de costos de la central de punta en el punto:

$$F_P + H_P \times V_P = F_V + H_P \times V_V.$$

Si reemplazamos la primera intersección en la función de ingresos de la central de valle, obtenemos:

$$M_V \times [(F_P + H_N \times V_P) + (H_P - H_N) \times V_P + (H_V - H_P) \times V_V].$$

Y simplificando:

$$M_V \times [(F_P + H_P \times V_P) + (H_V - H_P) \times V_V].$$

Reemplazando la segunda intersección, obtenemos:

$$M_V \times [(F_V + H_P \times V_V) + (H_V - H_P) \times V_V].$$

Que simplificando es igual a:

$$M_V \times [(F_V + H_V \times V_V)].$$

Los costos de la central de valle. Es decir, los ingresos de la central de valle son iguales a sus costos en un mercado competitivo.

Central de base

Finalmente, para las centrales de base sus ingresos son:

De la hora 0 a la hora H_N el valor de la ENS, es decir:

$$H_N \times ENS \times M_B;$$

De la hora H_N a la hora H_P el costo variable de la central de punta:

$$(H_P - H_N) \times V_P \times M_B;$$

De la hora H_p a la hora H_v el costo variable de la central de base:

$$(H_v - H_p) \times V_v \times M_B;$$

De la hora H_v hasta la hora 8.760, el costo variable de la central de base:

$$(8.760 - H_v) \times V_B \times M_B;$$

Sus ingresos totales son la suma de estos cuatro:

$$M_B \times [(H_N \times ENS) + (H_p - H_N) \times V_p + (H_v - H_p) \times V_v + (8.760 - H_v) \times V_B].$$

Al igual que en el caso de la central de valle hay tres intersecciones relevantes para las curvas; las dos listadas anteriormente más:

$$F_B + H_v \times V_B = F_v + H_v \times V_v.$$

Con los mismos argumentos utilizados para el caso de la central de valle, hemos llegado a la ecuación:

$$M_v \times [F_v + H_v \times V_v + (8.760 - H_v) \times V_B]$$

Remplazando la tercera intersección en esta ecuación, se obtiene:

$$M_v \times [F_v + H_v \times V_v + 8.760 \times V_B - (-F_B + F_v + H_v \times V_v)];$$

Es decir:

$$M_v \times [F_B + 8.760 \times V_B];$$

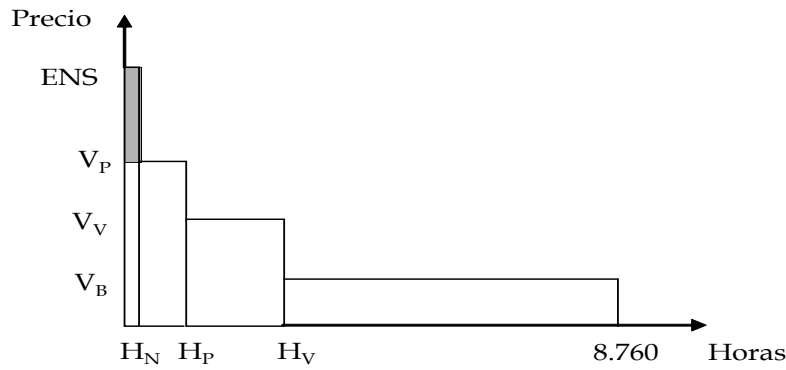
El costo total de las centrales de base.

Queda así demostrado que las centrales recuperan todos sus costos cuando el precio de mercado sube al nivel de la energía no suministrada en las horas en que no se cubre toda la demanda.

REMUNERACIÓN DE CONFIABILIDAD

En caso que no se permitiera que el precio de la electricidad alcanzara el valor de la ENS en las horas en que hay racionamiento, se tendría que buscar una forma para pagar los costos fijos a las centrales de punta. Este pago es el pago por capacidad que se debe pagar a las centrales que están generando cuando el sistema tiene energía no suministrada. El valor de este pago es equivalente a la zona sombreada del gráfico que reproducimos a continuación.

Figura 15: Remuneración de confiabilidad



Este pago es equivalente al área del rectángulo (base por altura):

$$H_N \times (ENS - V_P);$$

Y como conocemos que las curvas de costo total de la ENS y de la central de punta se intersectan en H_N horas:

$$H_N \times ENS = F_P + H_N \times V_P;$$

O, lo que es igual:

$$H_N \times (ENS - V_P) = F_P.$$

Es decir que el pago por confiabilidad es igual a la anualidad de una central de punta.

Barrera I Rey Economic Advisory

Madrid | España

www.barrera-rey.com

GGC | GLOBAL GENERAL CONSULTING

Barranquilla | Colombia

<https://ggc.aluez.com/>