



Procedimiento para captura de datos y transformación de valores atípicos

Estimación de la Línea Base de Consumo (LBC)

Gerencia del Mercado de Energía Mayorista
Dirección Analítica y Desarrollo del Mercado
24 de febrero 2023

Procedimiento para captura de datos, transformación de valores atípicos e iguales a cero y transformación de valores cuando haya desconexión

Procedimiento:

El procedimiento para captura de datos, transformación de valores atípicos e iguales a cero y transformación de valores cuando haya desconexión utilizados para la estimación de la Línea Base de Consumo -LBC- se realizará en el siguiente orden:

1. Captura de datos

La obtención de la LBC se realizará por tipo de día, considerando que los mismos corresponderán a cuatro tipos así:

- **Laboral:** Los cuales abarcan los días de la semana comprendidos entre el lunes y viernes que no sean festivos.
- **Sábado:** Corresponde a los días sábado no festivos.
- **Domingo:** Corresponde a todos los días domingos.
- **Festivos —No Domingo—:** Abarca cualquier día de la semana que es festivo excepto el día domingo.

La metodología se aplicará sobre una ventana de tiempo de sesenta (60) datos históricos de consumo diario de energía tomados como la suma de las veinticuatro (24) horas del día. Los datos históricos se tomarán a partir de la mejor información disponible que disponga el ASIC al momento del cálculo de la LBC, considerando para ello la más reciente versión de la liquidación publicada para los respectivos datos históricos según corresponda¹, ya sea:

- Segunda versión de la liquidación (versión TX2).
- Publicación de la información resumen mensual de la liquidación (versión TXR).
- Publicación de la facturación (versión TXF).
- Publicación de Ajustes a la facturación mensual (versiones TXn, n a partir de 3).

En caso de que un usuario tenga registrada más de una frontera comercial en un mismo predio o inmueble, el cálculo de la LBC se hará considerando la sumatoria de los consumos de cada una de las fronteras asociadas al predio o inmueble, el cual tendrá una única LBC.

Si para al menos una de las fronteras de comercialización para agentes y usuarios del predio, al momento del cálculo no se dispone de la muestra de sesenta (60)

¹ De acuerdo con la Resolución CREG 157 de 2011 y aquellas que la adicionan, modifiquen o sustituyan.

datos históricos de consumo diario, no se podrá efectuar el cálculo de la LBC y por tanto no podrá participar en el mecanismo de la DDV hasta que se disponga de ese mínimo número de datos.

2. Depuración de datos

Se clasificarán los sesenta (60) datos históricos de consumo diario (muestra inicial) por tipo de día (laboral, sábado, domingo y festivo —no domingo—).

Para el caso del tipo de día festivo (no domingo) y en caso de que la muestra sea inferior a cinco (5) datos históricos de consumo diario, se aplicarán los siguientes pasos en el siguiente orden:

- i. La muestra para este tipo de día se completará con los festivos (no domingo) que se encuentren en una muestra de ciento veinte (120) datos históricos de consumo (sesenta -60- datos históricos de consumo diario adicionales a la muestra inicial).
- ii. En caso de no encontrar los festivos (no domingo) suficientes para completar los cinco (5) datos, estos se completarán con los días de más bajo consumo distintos de cero que se tengan disponibles en la muestra inicial para el tipo de día domingo y que no correspondan a días en los cuales se hayan presentado desconexiones del mecanismo DDV y/o cualquier otro programa de respuesta de la demanda o de desconexión programada que defina la regulación.
- iii. Si con el paso anterior aún no se completa la muestra de cinco (5) datos para este tipo de día, se aplicará nuevamente el paso anterior completando la muestra con los días de más bajo consumo distintos de cero de la muestra inicial primero para el tipo de día sábado, y luego, en caso de ser necesario para el tipo de día laboral. Los días no pueden corresponder a días en los cuales se hayan presentado desconexiones del mecanismo DDV y/o cualquier otro programa de respuesta de la demanda o de desconexión programada que defina la regulación.
- iv. Si con los pasos anteriores no se completa la muestra de cinco (5) datos para este tipo de día, se considerarán los datos disponibles de la muestra inicial para ese tipo de día y con esto se calculará la LBC, sin aplicar ninguna de las transformaciones establecidas en este procedimiento.

Si todos los datos depurados de la muestra para un tipo de día son iguales a cero, no se aplicará ninguna de las transformaciones establecidas en este

procedimiento, y se considerarán los datos disponibles para ese tipo de día para el cálculo de la respectiva LBC.

2.1. Transformación de valores iguales a cero

Una vez se ha completado la depuración de datos para el tipo de día festivo (no domingo), y si existe al menos un valor distinto de cero dentro de la muestra para cada tipo de día, se procede a realizar la transformación de los valores iguales a cero de la siguiente manera:

- a) Para cada tipo de día, se ordenan los datos de la muestra desde la fecha más antigua a la más reciente, es decir, de manera ascendente, siendo la posición $n=1$ la fecha más antigua, la posición $n=2$ la siguiente fecha y así sucesivamente hasta la posición $n=m$ (fecha más reciente). Esta será la **muestra resultante**, por tipo de día.
- b) Los valores iguales a cero ($X_{n,TD}^0$) en la muestra resultante ordenada desde la posición $n=1$ hasta $n=m$, para cada tipo de día, serán reemplazados por el promedio simple móvil central con amplitud de cinco (5) datos de la muestra resultante, ubicando en el centro la posición cuyo valor es igual a cero y es el que se está reemplazando ($X_{n,TD}^0$).

El promedio será realizado con los valores de la amplitud que se encuentren disponibles que sean distintos de cero y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente, es decir, se utilizará los valores de las dos posiciones anteriores ($X_{n-2,TD}$, $X_{n-1,TD}$) y los valores de las dos posiciones posteriores ($X_{n+2,TD}$, $X_{n+1,TD}$) a la posición que se está transformando ($X_{n,TD}^0$). Para realizar dicho promedio se considerarán los valores que hayan sido transformados ($X_{n,TD}^{T_0}$) aplicando el presente procedimiento.

Si el primer valor de la muestra resultante (posición $n=1$) es igual a cero ($X_{1,TD}^0$), este será transformado por el promedio de los dos valores posteriores ($X_{n+2,TD}$, $X_{n+1,TD}$) siempre y cuando estos sean distintos de cero y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente. De manera análoga, si el último valor de la muestra resultante (posición $n=m$) es igual a cero ($X_{m,TD}^0$), este será transformado por el promedio de los dos valores anteriores ($X_{n-2,TD}$, $X_{n-1,TD}$) siempre y cuando estos sean distintos de cero y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente. Para realizar dicho promedio se

considerarán los valores que hayan sido transformados ($X_{n,TD}^{T_0}$) aplicando el presente procedimiento.

En todo caso, de no lograr aplicar este procedimiento a algún valor igual a cero ($X_{n,TD}^0$) en la muestra resultante, se descartará dicho dato ($X_{n,TD}^0$) de la muestra para ese tipo de día.

Lo anterior se explica de la siguiente forma: De manera iterativa y ascendente (desde posición $n=1$ hasta $n=m$), por tipo de día y para cada posición n donde el valor de muestra resultante sea igual a cero ($X_{n,TD}^0$), se realizará el siguiente reemplazo:

$$\forall_n | X_{n,TD} = X_{n,TD}^0 \wedge C = \{X_{n-2,TD}, X_{n-1,TD}, X_{n+1,TD}, X_{n+2,TD}\} :$$

$$X_{n,TD}^{T_0} = \frac{X_{n-2,TD} + X_{n-1,TD} + X_{n,TD}^0 + X_{n+1,TD} + X_{n+2,TD}}{N_{TD}} \Leftrightarrow c \in C \mid c > 0 \wedge c \neq X_{n,TD}^{desconexión}$$

Luego

$$X_{n,TD}^0 = X_{n,TD}^{T_0}$$

Donde,

$X_{n,TD}^{T_0}$: valor de energía transformado para la posición n para el tipo de día TD

$X_{n,TD}^0$: valor de energía de la posición n que es igual a cero para el tipo de día TD

$X_{n,TD}^{desconexión}$: valor donde se presentó una desconexión en la posición n para el tipo de día TD

$X_{n,TD}$: valor de energía de la **muestra resultante** para la posición n para el tipo de día TD

$X_{n-2,TD}$: valor de energía de la **muestra resultante** para la posición $n - 2$ diferente de cero para el tipo de día TD

$X_{n-1,TD}$: valor de energía de la **muestra resultante** para la posición $n - 1$ diferente de cero para el tipo de día TD

$X_{n+1,TD}$: valor de energía de la **muestra resultante** para la posición $n + 1$ diferente de cero o transformado

$X_{n+2,TD}$: valor de energía de la **muestra resultante** para la posición $n + 2$ diferente de cero para el tipo de día TD

N_{TD} : Cantidad total de datos disponibles para el promedio móvil diferentes a cero para el tipo de día TD.

TD: tipo de día (laboral, sábado, domingo y festivo no domingo)

n : valor que indica la posición de la muestra de 1 ... hasta m , según el tipo de día TD.

Finalmente, si luego de aplicar la transformación de valores iguales a cero para un tipo de día se encuentra que todos los datos de la muestra resultan con el mismo valor, entonces, no se continuará con la aplicación de los pasos siguientes de este procedimiento para ese tipo de día, y con los datos resultantes hasta ese momento se calculará la respectiva LBC.

2.2. Eliminación de mayor y menor consumo

Una vez transformados o eliminados los valores de la muestra resultante iguales a cero, de conformidad con lo establecido en el paso 2.1; para los tipos de día laboral, sábado y domingo (se excluye el tipo de día festivo —no domingo—), se identificarán los valores del consumo diario máximo y mínimo de la muestra resultante excluyendo los valores de los días que se hayan presentado desconexiones o reducciones de energía de acuerdo con la regulación vigente. Estos dos datos identificados serán eliminados de la respectiva muestra.

En caso de presentarse para un mismo tipo de día un consumo igual al valor máximo o al valor mínimo para varios días de la muestra resultante, entonces se eliminará el valor correspondiente a la fecha más antigua. Lo anterior, se aplicará tanto para el valor máximo como para el mínimo.

2.3. Transformación de valores atípicos

- a) La identificación y transformación de valores atípicos, de todos los tipos de día, se realizará siempre y cuando la cantidad de datos de la muestra que resulte por tipo de día luego de aplicar la *Eliminación de mayor y menor consumo* y excluyendo los valores de los días que se hayan presentado desconexiones o reducciones de energía de acuerdo con la regulación vigente, sea mayor o igual a cinco (5) datos. En caso de que la muestra no cumpla dicha desigualdad, no se realizará la *Transformación de valores atípicos* para el respectivo tipo de día y se continuará con el paso 2.4.

Para la identificación de valores atípicos de los tipos de días que cumpla la desigualdad antes mencionada, se utilizará una medida de dispersión estadística denominada rango intercuartílico que utiliza la mediana como medida central. La aplicación de este procedimiento se realizará con la muestra resultante que se tenga luego de aplicar los pasos anteriores excluyendo los valores de los días que se hayan presentado desconexiones o reducciones de energía de acuerdo con la regulación vigente. Esta será la muestra resultante que se utilizará en adelante.

El proceso para aplicar esta medida de dispersión es el siguiente para cada tipo de día (laboral, sábado, domingo, festivo —no domingo—):

- i. Se ordena el conjunto de datos de la muestra resultante de forma ascendente por el valor de energía independientemente del orden de las fechas.

- ii. Se identifican los cuartiles Q1, Q2 y Q3 que determinan los valores correspondientes al 25%, 50% y 75% de los datos, el Q2 corresponde a la mediana. Para la identificación de los cuartiles Q1 y Q3 se debe utilizar el método de Moore y McCaben ²
- iii. Se halla el rango intercuartil (IQR) que corresponde a Q3 menos Q1.
- iv. Los valores atípicos serán aquellos $X_{n,TD}$ que cumplan con alguna de las siguientes condiciones:

$$X_{n,TD}^{atípico} < Q_{1,TD} - 1.5 * IQR_{TD} \quad o \quad X_{n,TD}^{atípico} > Q_{3,TD} + 1.5 * IQR_{TD}$$

Donde:

$X_{n,TD}^{atípico}$ = Valor de energía atípico en la posición n de la muestra resultante para el tipo de día TD

$Q_{1,TD}$ = Cuartil 1 del tipo de día TD

$Q_{3,TD}$ = Cuartil 3 del tipo de día TD

IQR_{TD} = Rango Intercuartil del tipo de día TD

- b) Para la transformación de valores atípicos $X_{n,TD}^{atípico}$ identificados en el paso anterior, se empleará el procedimiento establecido para la transformación de valores en cero, es decir:

- i. Para cada tipo de día, se ordenan los datos de la muestra resultante que se disponga desde la fecha más antigua a la más reciente, es decir, de manera ascendente, siendo la posición n=1 la fecha más antigua, la posición n=2 la siguiente fecha y así sucesivamente hasta la posición n=m (fecha más reciente). Esta será la **muestra resultante**, por tipo de día.
- ii. Los valores identificados como atípicos ($X_{n,TD}^{atípico}$) en la muestra resultante ordenada desde la posición n=1 hasta n=m, para cada tipo de día, serán reemplazados por el promedio simple móvil central con amplitud de cinco (5) datos de la muestra resultante, ubicando en el centro la posición cuyo valor sea atípico y sea el que se está reemplazando ($X_{n,TD}^{atípico}$).

El promedio será realizado con los valores de la amplitud que se encuentren disponibles que no sean valores atípicos y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente, es decir, se utilizará los valores de las dos posiciones anteriores ($X_{n-2,TD}, X_{n-1,TD}$) y los valores de las dos

² Método utilizado para la identificación de cuartiles por medianas.

posiciones posteriores ($X_{n+2,TD}$, $X_{n+1,TD}$) a la posición que se está transformando ($X_{n,TD}^{atípico}$). Para realizar dicho promedio se considerarán los valores que hayan sido transformados ($X_{n,TD}^{T_{atípico}}$) aplicando el presente procedimiento.

Si el primer valor de la muestra resultante (posición $n=1$) es atípico ($X_{1,TD}^{atípico}$), este será transformado por el promedio de los dos valores posteriores ($X_{n+2,TD}$, $X_{n+1,TD}$) siempre y cuando estos no sean valores atípicos y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente. De manera análoga, si el último valor de la muestra resultante (posición $n=m$) es atípico ($X_{m,TD}^{atípico}$), este será transformado por el promedio de los dos valores anteriores ($X_{n-2,TD}$, $X_{n-1,TD}$) siempre y cuando estos no sean valores atípicos y no hayan sido días en los cuales se hayan presentado desconexiones de acuerdo con la regulación vigente. Para realizar dicho promedio se considerarán los valores que hayan sido transformados ($X_{n,TD}^{T_{atípico}}$) aplicando el presente procedimiento.

En todo caso, de no lograr aplicar este procedimiento, se descartará el valor identificado como atípico ($X_{n,TD}^{atípico}$) de la muestra para ese tipo de día.

Lo anterior se explica de la siguiente forma: De manera iterativa y ascendente (desde la posición $n=1$ hasta $n=m$), para cada posición n donde el valor de muestra sea identificado como atípico ($X_{n,TD}^{atípico}$), por tipo de día, se realizará el siguiente reemplazo:

$$\forall_n \mid X_{n,TD} = X_{n,TD}^{atípico} \wedge C = \{X_{n-2,TD}, X_{n-1,TD}, X_{n+1,TD}, X_{n+2,TD}\}$$

$$X_{n,TD}^{T_{atípico}} = \frac{X_{n-2,TD} + X_{n-1,TD} + X_{n,TD}^{atípico} + X_{n+1,TD} + X_{n+2,TD}}{N_{TD}} \Leftrightarrow c \in C \mid c \neq X_{n,TD}^{atípico} \wedge c \neq X_{n,TD}^{desconexión}$$

Luego

$$X_{n,TD}^{atípico} = X_{n,TD}^{T_{atípico}}$$

Donde,

$X_{n,TD}^{T_{atípico}}$: valor transformado para la posición n para el tipo de día TD

$X_{n,TD}^{atípico}$: valor atípico presentado en la posición n para el tipo de día TD

$X_{n,TD}^{desconexión}$: valor donde se presentó una desconexión en la posición n para el tipo de día TD

$X_{n,TD}$: valor de energía de la **muestra resultante** para la posición n para el tipo de día TD

- $X_{n-2,TD}$: valor de energía de la **muestra resultante** para la posición $n - 2$ diferente a atípico para el tipo de día TD
- $X_{n-1,TD}$: valor de energía de la **muestra resultante** para la posición $n - 1$ diferente a atípico para el tipo de día TD
- $X_{n+1,TD}$: valor de energía de la **muestra resultante** para la posición $n + 1$ diferente a atípico para el tipo de día TD
- $X_{n+2,TD}$: valor de energía de la **muestra resultante** para la posición $n + 2$ diferente a atípico para el tipo de día TD
- N_{TD} : Cantidad total de datos disponibles para el promedio móvil diferentes a valores atípicos para el tipo de día TD

2.4. Transformación de valores para días con desconexiones

Partiendo de la muestra resultante que se tenga luego de aplicar los pasos anteriores, si para un tipo de día, todos o ninguno de los días de la muestra hubo desconexiones o reducciones de energía en cumplimiento de la regulación vigente, no se aplicará la transformación de valores para días con desconexiones. En este caso, para el cálculo de la LBC se considerarán los datos de la muestra resultante disponibles para ese tipo de día. De lo contrario, se aplicará lo siguiente:

Para los días en que se hayan presentado desconexiones o reducciones de energía en cumplimiento del mecanismo de la DDV o sus pruebas de disponibilidad, y/o cualquier otro programa de respuesta de la demanda o de desconexión programada que defina el Regulador, el valor de desconexión se remplazará por el promedio de las cuatro (4) posiciones anteriores del mismo tipo de día, siempre y cuando estos no sean días en los que se hayan presentado desconexiones o correspondan a valores ajustados con este procedimiento.

En caso de no disponer de datos para todas las cuatro (4) posiciones anteriores del mismo tipo de día, se calculará el promedio con los datos de las posiciones anteriores del mismo tipo de día disponibles dentro de la muestra resultante que se disponga, siempre y cuando no sean valores de días donde se presentaron desconexiones o correspondan a valores ajustados con este procedimiento, según lo establecido en el literal c del numeral 1.1 del Anexo 1 de la Resolución CREG 101 019 de 2022.

Si no se dispone de ningún valor anterior para efectuar el promedio, se calculará el promedio de las cuatro posiciones posteriores disponibles para cada tipo de día, siempre y cuando no sean valores de días donde se presentaron desconexiones. En caso de que no existan valores disponibles de aplicar la condición anteriormente descrita y si para el valor de la posición del día que se presentó la desconexión corresponde a un valor que se ha transformado por la metodología de transformación de valores atípicos, entonces se tomará el valor

resultante de dicha transformación. Si no es posible aplicar este paso se descartará el valor del día en que se presentó una desconexión de la muestra para ese tipo de día.

Lo anterior se explica de la siguiente forma: De manera iterativa y ascendente (desde la posición $n=1$ —fecha más antigua— hasta $n=m$ —fecha más reciente—), para cada posición n donde el valor de muestra sea un día donde se haya presentado una desconexión ($X_{n,TD}^{desconexión}$), por tipo de día, se realizará el siguiente reemplazo:

$$\forall_n \mid X_{n,TD} = X_{n,TD}^{desconexión} \wedge C = \{X_{n-4,TD}, X_{n-3,TD}, X_{n-2,TD}, X_{n-1,TD}\}$$

$$X_{n,TD}^{T_{desconexión}} = \frac{X_{n-4,TD} + X_{n-3,TD} + X_{n-2,TD} + X_{n-1,TD}}{N_{TD}} \Leftrightarrow c \in C \mid c \neq X_{n,TD}^{desconexión}$$

Luego

$$X_{n,TD}^{desconexión} = X_{n,TD}^{T_{desconexión}}$$

Donde,

$X_{n,TD}^{T_{desconexión}}$: valor transformado para la posición n para el tipo de día TD

$X_{n,TD}^{desconexión}$: valor donde se presentó una desconexión en la posición n para el tipo de día TD

$X_{n-4,TD}$: valor de energía de la **muestra resultante** para la posición n – 4 diferentes a valores de desconexión para el tipo de día TD

$X_{n-3,TD}$: valor de energía de la **muestra resultante** para la posición n – 3 diferentes a valores de desconexión para el tipo de día TD

$X_{n-2,TD}$: valor de energía de la **muestra resultante** para la posición n – 2 diferentes a valores de desconexión para el tipo de día TD

$X_{n-1,TD}$: valor de energía de la **muestra resultante** para la posición n – 1 diferentes a valores de desconexión para el tipo de día TD

N_{TD} : Cantidad total de datos disponibles para el cálculo del promedio diferentes a valores de desconexión para el tipo de día TD

Si, $C = \emptyset \rightarrow B = \{X_{n+1,TD}, X_{n+2,TD}, X_{n+3,TD}, X_{n+4,TD}\}$:

$$X_{n,TD}^{T_{desconexión}} = \frac{X_{n+1,TD} + X_{n+2,TD} + X_{n+3,TD} + X_{n+4,TD}}{N_{TD}} \Leftrightarrow b \in B \mid b \neq X_{n,TD}^{desconexión}$$

Donde,

$X_{n+4,TD}$: valor de energía de la **muestra resultante** para la posición n + 4 diferentes a valores de desconexión para el tipo de día TD

$X_{n+3,TD}$: valor de energía de la **muestra resultante** para la posición n + 3 diferentes a valores de desconexión para el tipo de día TD

$X_{n+2,TD}$: valor de energía de la **muestra resultante** para la posición n + 2 diferentes a valores de desconexión para el tipo de día TD

- $X_{n+1,TD}$: valor de energía de la **muestra resultante** para la posición $n + 1$ diferente a valores de desconexión para el tipo de día TD
- N_{TD} : Cantidad total de datos disponibles para el cálculo del promedio diferentes a valores de desconexión para el tipo de día TD

En todo caso, de no lograr aplicar este procedimiento, se descartará el valor del día de la desconexión o reducción de energía ($X_{n,TD}^{desconexión}$) de la muestra para ese tipo de día.

3. Definición de la LBC

Una vez finalizado este procedimiento, se tendrá en cuenta la muestra que resulte para la estimación de la LBC según lo establecido en el numeral 1 del Anexo 1 de la Resolución CREG 101 019 de 2022.