IEE CONICET INSTITUTO DE ENERGÍA ELÉCTRICA | SICEL 2025 | UNIVERSIDAD NACIONAL DE COLOMBIA

Metodología de Muestreo basada en Entropía de Shannon para la Caracterización de la Demanda Energética en Regiones de Colombia

Óscar Alberto Bustos[1], Julián David Osorio[1], Javier Rosero García[1], Cristian Camilo Marín Cano[2], Luis Alirio Bolaños[2], Pedro Jose Toro[2], Óscar Andrés Ocampo[2]

[1] EM&D Research Group, Universidad Nacional de Colombia, sede Bogotá, Colombia; jaroserog@unal.edu.co

[2] CHEC-Grupo EPM, Medellín, Colombia

***Abstract*—Este trabajo propone una metodología de muestreo basada en entropía de Shannon para caracterizar eficientemente la demanda energética en regiones de Colombia con infraestructura eléctrica de desarrollo moderado. Ante la ausencia de tecnologías de medición avanzada (AMI) y el uso de registros convencionales de consumo, se diseña una estrategia de muestreo que permita representar adecuadamente la demanda sin requerir sensores automatizados ni monitoreo complejo. El objetivo es obtener una muestra de usuarios que represente adecuadamente a la población para facilitar la planificación y gestión del sistema eléctrico. Se utilizaron datos históricos de consumo mensual, ubicación geográfica, tipo de cliente y población. Se dividió el estudio en cuatro subregiones (Norte, Sur, Centro y Oriente) y se establecieron criterios para el tamaño de muestra, el balance entre categorías de usuarios y la selección representativa de usuarios. La muestra fue evaluada según su eficiencia logística y representatividad frente a las distribuciones originales de consumo y categorías auxiliares. Se concluye que el enfoque por entropía genera una muestra densa y estratégicamente ubicada, con excelente ajuste a la distribución de consumo, siendo una opción robusta y práctica para el operador de red.**

Keywords*—*Entropía de Shannon, caracterización de la demanda, muestreo energético, planificación energética

I. Introducción

Este estudio aborda el problema de un operador de red eléctrica en Colombia que necesita caracterizar la demanda eléctrica en su zona de cobertura. La región cuenta con una red de distribución de desarrollo moderado debido a limitaciones económicas y geográficas. Se trata de un territorio con extensas áreas montañosas en el centro del país. Como no se dispone de tecnologías avanzadas, se requiere una estrategia de muestreo que funcione sin sensores automatizados ni herramientas tecnológicas robustas. Contar con una estrategia flexible y costo-eficiente es clave ante el reto que representan zonas rurales o con bajo desarrollo tecnológico, como ocurre en varias regiones del país [1]. Esta estrategia debe mantener la representatividad y al mismo tiempo ser fácil de implementar. Su objetivo es apoyar al operador en la toma de decisiones sobre planificación energética y en la mejora de la eficiencia del suministro y la gestión de la demanda. Se propone una metodología de muestreo basada en la entropía de Shannon y se analizan sus resultados en términos de representatividad frente a la población original, flexibilidad y viabilidad logística. Esta metodología aplica conceptos de teoría de la información poco comunes en el sector eléctrico, donde predominan enfoques basados en muestreo aleatorio. Con este enfoque se busca una estrategia más determinista [2]. La propuesta está diseñada para operar con datos de medición convencional y no depende de redes AMI. Esto responde a la necesidad de metodologías más eficientes y sostenibles para la gestión energética en regiones en desarrollo.

II. Estado del Arte

Las metodologías de muestreo han evolucionado con técnicas que mejoran la representatividad y reducen sesgos. Un ejemplo es el muestreo activo asistido por aprendizaje automático que alterna entre estimación y recolección de datos según predicciones sobre datos no observados [2]. En el monitoreo ambiental se propone una estrategia para seleccionar sitios de muestreo de microcontaminantes en ríos que combina interpolación Spline, análisis jerárquico y sistemas de información geográfica para optimizar la distribución espacial [3]. En estudios longitudinales se han evaluado técnicas de imputación mediante el error cuadrático medio para diferentes tipos de datos faltantes (MCAR, MAR y NMAR). Esto mejora la calidad analítica con muestras incompletas [4]. En inteligencia artificial, el muestreo cumple una función clave en el entrenamiento de modelos. En deep metric learning se ha propuesto un enfoque basado en agrupamiento de representaciones en espacios de características para seleccionar muestras informativas. Esto mejora la convergencia evitando mínimos locales [5]. El muestreo también resulta esencial en contextos con datos desbalanceados. Aumentar el tamaño de muestra y equilibrar clases mejora la precisión y la capacidad predictiva. La elección de semillas aleatorias también influye en la estabilidad y generalización del modelo [6]. Además, el muestreo de Gibbs junto con redes neuronales GRNN permite generar muestras sintéticas que preservan la estructura original de los datos. Esto mejora la representación de la distribución

http://doi.org/10.15446/sicel.v12.121277

real y el rendimiento predictivo en contextos con datos limitados [7]. Estos avances muestran la variedad de metodologías y destacan la importancia de adaptar cada técnica a su contexto. La combinación de métodos clásicos con herramientas computacionales y modelos estadísticos modernos ha mejorado la calidad de las muestras y ha reducido la incertidumbre en decisiones basadas en datos. En este estudio sobre caracterización de la demanda se requiere una muestra que represente con precisión los hábitos de consumo y las características de la población. Tras una búsqueda bibliográfica en distintas bases académicas, se logró clasificar los métodos usados en la caracterización de la demanda en cuatro grandes grupos. La Fig. 1 presenta un resumen de estos hallazgos.

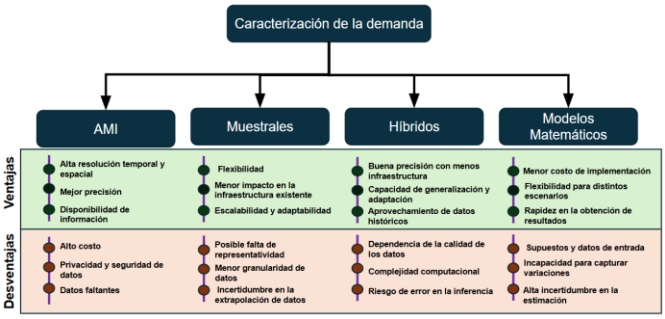

Fig. 1. Categorización de los métodos para obtener datos para efectuar la caracterización de la demanda.

El método muestral no depende de tecnologías de medición inteligente ausentes en la región de Colombia analizada. Por esta razón, este enfoque es más pertinente para el caso de uso de este estudio. A continuación se profundizará en el estado del arte de este enfoque.

A. *Métodos Muestrales*

Este enfoque entiende el muestreo como un proceso de investigación en el que un subconjunto de la población permite hacer inferencias válidas sobre el total. Para lograr representatividad es necesario definir objetivos, diseñar el estudio y establecer criterios claros de inclusión y exclusión. En sistemas eléctricos se ha propuesto tomar múltiples muestras de usuarios para construir curvas de demanda representativas siempre que el muestreo permita extrapolar los datos con precisión [9]. Aunque esta estrategia requiere muchas mediciones, logra reflejar bien la diversidad de usuarios y permite estimaciones precisas [10]. Una técnica eficaz para representar distintos tipos de usuarios es el muestreo estratificado. Este método divide la población en subgrupos homogéneos según criterios como tipo de usuario, ubicación o nivel de consumo. Esto mejora la precisión de las inferencias y reduce los sesgos. Por ejemplo, el estudio TURKSTAT 2019 Household Budget Survey aplicó muestreo estratificado por conglomerados para estimar el consumo eléctrico mensual en hogares de Turquía. Esto mejoró la extrapolación de resultados [11]. Para definir el tamaño de la muestra se consideran el nivel de confianza que suele ser del noventa y cinco por ciento y el margen de error aceptado [12]. También es importante validar la muestra mediante encuestas a los usuarios seleccionados. Esto aporta más información y mejora la representación del consumo real [10]. La caracterización de la demanda ha mejorado gracias al modelado de carga basado en datos. Estas técnicas capturan la diversidad y variabilidad de las cargas en redes de distribución y mejoran la simulación y planificación [13]. También se aplican técnicas de muestreo inteligente que usan análisis de datos y métodos avanzados para seleccionar escenarios representativos. Esto optimiza el estudio del sistema eléctrico [14]. En la gestión de la demanda, algoritmos como el A-DPC permiten priorizar clientes en programas de respuesta según sus patrones de consumo. Esto aumenta la efectividad de las estrategias [15]. Una caracterización precisa de la demanda es clave para la planificación energética. Se han creado metodologías que generan series temporales representativas. Estas metodologías reducen el costo computacional sin perder precisión [16]. También se han desarrollado enfoques que combinan análisis exploratorios con técnicas avanzadas para construir subconjuntos útiles en estudios eléctricos [14]. Además, existen modelos de carga basados en datos que capturan la variabilidad diaria del consumo [17].

En conjunto, estas contribuciones subrayan la importancia de seleccionar muestras representativas y aplicar técnicas adecuadas para mejorar la estimación y caracterización de la demanda eléctrica.

III. Marco Teórico

A. *Entropía de Shannon*

Es una medida de la incertidumbre o cantidad de información contenida en una variable aleatoria. En el contexto del muestreo energético, se puede utilizar para identificar los transformadores de distribución cuyos usuarios aportan mayor información. Por ejemplo, esto ocurre cuando hay patrones de demanda más diversos en función del tipo o población del usuario. La entropía de una variable discreta X con probabilidades pi para cada posible estado xi se calcula como [18]:

$$H(X) = -\sum_i \; p_i \log_2(p_i) \tag{1}$$

B. *Muestra de Cochran*

Usada para el cálculo del tamaño de muestra en poblaciones finitas, donde se determina el tamaño de muestra necesario para estimar parámetros poblacionales con un nivel de confianza y margen de error específicos.

$$n_0 = \frac{Z^2 \cdot p \cdot (1-p)}{e^2} \tag{2}$$

Z es el puntaje para el nivel de confianza seleccionado e indica el margen de error. La proporción p representa la fracción esperada de la población con la característica de interés. Por ejemplo, si se sabe que el 40% de la población usa energías renovables, se puede usar $p = 0.4$ como probabilidad de seleccionar un usuario que las consuma. Sin datos previos se recomienda mantener $p = 0.5$ [19].

IV. Metodológica de la Estrategia de Muestreo

En esta sección se describen las estrategias propuestas y adicionalmente se detalla el proceso de análisis preliminar ejecutado sobre los datos para comprender mejor el caso de estudio y los datos de la población.

A. *Análisis preliminar de datos*

En esta sección se plantean brevemente los pasos a realizar para el análisis preliminar de los datos junto con una explicación general de los datos disponibles. El análisis se realiza siguiendo los siguientes pasos:

1. Descripción detallada de Fuentes de Datos.
2. Procesamiento y Limpieza de Datos.

En general se cuenta con información del consumo mensual de los usuarios de la región a lo largo de 5 años. A cada usuario se le relaciona con un transformador eléctrico el cual suple su demanda. Se cuenta con información de caracterización general de los usuarios según su tipo de cliente y tipo de población:

- **Tipo de Cliente:** Comerciales, Industriales, Oficiales, Provisionales, Residenciales (Estrato 1 a 6), Otros.
- **Tipo de Población:** Urbano, Rural, Centro Poblado.

Se cuenta con las coordenadas geográficas (latitud y longitud) de cada usuario. Los transformadores —y por extensión los usuarios— están clasificados en 4 subregiones definidas por el operador de red: Norte/Nor-occidente, Sur/Sur-occidente, Centro y Oriente. Esta división permite descomponer el problema de muestreo en 4 subproblemas independientes. Es posible aplicar la estrategia de muestreo por separado en cada subregión para lograr una muestra con mayor representatividad geográfica. La subregión Norte/Nor-occidente es principalmente rural, con relieve montañoso y limitada infraestructura vial. La subregión Sur/Sur-occidente combina zonas rurales y urbanas en crecimiento, con actividad industrial y portuaria concentrada cerca de un río principal. La subregión Centro concentra las principales zonas metropolitanas y la mayor parte de la población, con diversidad de usuarios residenciales, comerciales e industriales. La subregión Oriente tiene baja elevación y un río de importancia comercial. Los pisos térmicos templados y cálidos dominan en toda la región.

B. *Muestreo por entropía de Shannon*

Se presenta la metodología de muestreo basada en la entropía de Shannon. Esta se orienta a identificar los transformadores que aportan más información y mayor variedad de perfiles de usuarios.

El proceso comienza con el cálculo de la entropía de Shannon para cada transformador según los usuarios asociados. Se consideran las categorías de tipo de cliente y tipo de población. Para cada transformador se determina la proporción de usuarios por categoría. Estas proporciones se usan como probabilidades para calcular la entropía con la fórmula (1). Una categoría con probabilidad 0 no aporta a la entropía. Para evitar que transformadores pequeños pero diversos tengan un valor inflado, se normaliza la entropía multiplicándose por el número de usuarios. Esto permite priorizar transformadores que tienen diversidad y volumen al mismo tiempo. Así se logra un muestreo representativo y eficiente sin tener que medir muchos transformadores pequeños de forma innecesaria. Luego se construye un ranking de usuarios basado en la entropía. Se calculan las entropías por tipo de cliente y por tipo de población. Ambos rankings se combinan con un esquema de ponderación ajustable que equilibra los criterios de selección. Se da mayor peso al tipo de cliente por su mayor diversidad, pero se asegura la inclusión de usuarios rurales. Para esto se prueban combinaciones de pesos como 90/10 y 10/90 entre tipo de cliente y tipo de población. Se aplica la fórmula de Cochran (2) con niveles de confianza n_c y márgenes de error $1 - n_c$ para definir el tamaño óptimo de la muestra. A partir del ranking se seleccionan transformadores en orden descendente y se extraen sus usuarios hasta alcanzar el tamaño determinado. Debido a la variabilidad en usuarios por transformador, puede haber una ligera diferencia entre el tamaño ideal y el obtenido.

Cada subpoblación se trata por separado y las cuatro submuestras se integran en una muestra final de toda la región. Se visualiza el proceso general en la Fig. 2.

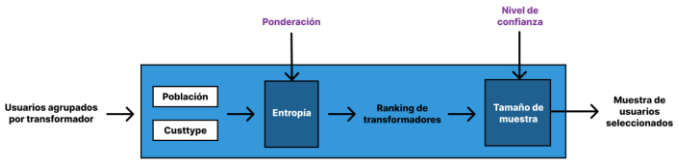

Fig. 2. La metodología recibe los usuarios agrupados y utiliza algunas variables para generar un ranking de transformadores del más informativo al menos informativo. A partir de este ranking se genera una muestra. Los parámetros de la metodología se resaltan en morado.

V. Resultados de la Estrategia de Muestreo

A. *Resultados del Análisis preliminar de Datos*

Esta sección presenta brevemente los principales hallazgos derivados del estudio exploratorio de las bases de datos de consumo energético y detalla en la preparación de los datos para su posterior muestreo.

1) Fuentes de datos analizadas: Para el estudio del consumo energético por usuario se utilizaron archivos que contienen el consumo mensual total de cada usuario. Los registros cubren un total de 58 meses. Cada usuario tiene un registro por mes con su consumo. Algunos usuarios tienen meses con registros nulos. El consumo mensual está registrado en KWH. Cada registro incluye también el transformador que suministra energía al usuario. En total se identificaron 568.405 usuarios activos distribuidos en 20.102 transformadores. Además, se contó con un archivo de asignación geográfica de transformadores. Este archivo ubica cada transformador en una de las subregiones del área de influencia de la empresa. Las subregiones son Norte/Nor-occidente, Sur/Sur-occidente, Centro y Oriente. Esta división permite realizar el análisis por subregión y mejora la representatividad del muestreo en áreas específicas. También se identificaron los usuarios que no cumplían con el criterio de tener al menos 48 meses de registros. Este umbral equivale al 82% de cobertura mínima. En total se encontraron 89.015 usuarios con información insuficiente. Esto representa cerca del 15% del total. Esta cantidad se puede explicar por el uso de métodos convencionales de recolección de datos en la región. La zona no cuenta con sistemas robustos como los sensores AMI.

2) Procesamiento y limpieza de Datos: Una vez identificadas las fuentes de información se realizó el procesamiento y limpieza de los datos. Se filtraron los usuarios con información insuficiente y se calculó una columna

adicional con el promedio anual de consumo en KWH. Para ello se sumó el consumo mensual por cada año y se halló el promedio anual. Esta medida resume el consumo de los usuarios y permite considerar usuarios que aún tienen datos faltantes en algunos meses. El promedio anual de consumo representa la distribución de consumo de la población. Esta distribución se usará como base para aplicar las estrategias de muestreo. Cada usuario tiene un transformador asociado y cada transformador está clasificado en una de las 4 subregiones. Entonces, se agruparon los usuarios en los 4 subconjuntos independientes.

B. Resultados del Tamaño de la Muestra

En esta breve sección se presenta la Tabla I con los resultados para los tamaños de la muestra teóricos empleando la fórmula de Cochran (2) segregados por subregión.

TABLA I. TAMAÑO DE MUESTRA TEÓRICO PARA DIFERENTES NIVELES DE CONFIANZA.

Nivel	Centro	Norte	Sur	Oriente	Total
0.90	68	68	68	68	272
0.95	384	383	384	383	1534
0.98	3329	3288	3316	3223	13156
0.99	15394	14540	15112	13358	58404

A partir de los resultados de la Tabla I se seleccionó un nivel de confianza del 95%. Este nivel ofrece un tamaño de la muestra moderado. El tamaño de la muestra se hace demasiado grande para un nivel de confianza superior. Con esto se pierde el propósito inicial de lograr una muestra conveniente de implementar.

C. Análisis de mapas

En esta sección se describe y compara la distribución espacial resultante. Se presentan mapas anonimizados con el único propósito de ilustrar cómo se posicionan los usuarios de la muestra respecto al resto.

La Fig. 3 muestra la distribución espacial de la muestra basada en entropía de Shannon. Corresponde a una muestra con nivel de confianza del 95% y una ponderación tipo de cliente/tipo de población de 90/10 (la ponderación se explica en la sección V.D.). La Fig. 4 presenta la distribución espacial de una muestra aleatoria de control basada en la curva de consumo de los usuarios, también al 95% de confianza. Esta muestra es mucho más dispersa a nivel geográfico que la de la Fig. 3. Se observa que los usuarios tienden a agruparse en zonas de alta actividad. Esta concentración se vuelve menos evidente con niveles de confianza menores. Aunque el número de usuarios parece bajo, su agrupación genera puntos pequeños en el mapa general. Para lograr una distribución más uniforme sería necesario subir el nivel de confianza al 98% y aumentar el tamaño de muestra 7.7 veces. Dividir la población por subgrupos evita que toda la muestra se concentre en la subregión Centro. Lo anterior llegaría a ocurrir si se aplicara el muestreo directamente sobre toda la población, debido a que los transformadores urbanos son los más diversos. El esquema de ranking ponderado garantiza presencia tanto en centros urbanos como rurales. Sin este ajuste, la muestra tiende a formar cúmulos en barrios residenciales de ciudades principales. Estos barrios concentran la mayoría de usuarios residenciales y comerciales. Por subregión, se identifican áreas densas cerca de centros poblados. Esto refleja la distribución natural del consumo en zonas urbanas.

En síntesis, la muestra por entropía tiende a concentrarse en zonas densamente pobladas. Esto facilita la implementación de sistemas de monitoreo y puede ajustarse mediante ponderaciones que favorezcan usuarios en áreas rurales más dispersas.

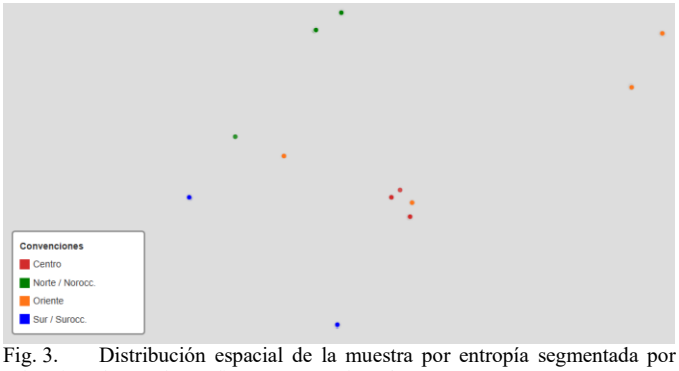

Fig. 3. Distribución espacial de la muestra por entropía segmentada por subregión al 95% de confianza y pesos de 90/10.

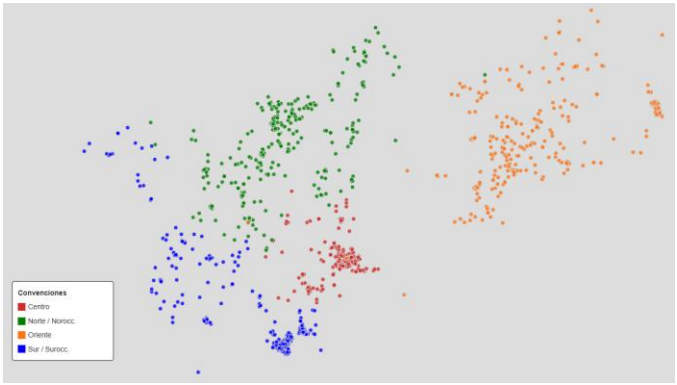

Fig. 4. Distribución espacial usando una muestra aleatoria de control segmentada por subregión al 95% de confianza.

D. Resultados del muestreo por entropía

El primer paso fue seleccionar los pesos entre tipo de cliente y tipo de población para construir el ranking de transformadores de distribución según su entropía. Para la tabla 2 se generaron muestras con nivel de confianza del 95% y diferentes combinaciones de pesos cliente/población. Luego se calcularon las proporciones de cada muestra según tipo de cliente y tipo de población. Estas proporciones se compararon con las de la población general. A partir de las diferencias se estimó una desviación total acumulada y se seleccionó la muestra con menor desviación. La mejor configuración fue la de 90/10.

La muestra seleccionada tiene 1,750 usuarios asociados a 8 transformadores de distribución (2 transformadores por subregión). Estos transformadores presentan una entropía de Shannon promedio normalizada de 186.41. En comparación, la población general incluye 569,423 usuarios, 20,222

transformadores y su entropía promedio ponderada con pesos 90/10 es 12.03. Se determinó que el tamaño de muestra para un nivel de confianza del 95% es 1,525. No obstante, la estrategia por entropía genera tamaños ligeramente distintos. Esto se debe a que se seleccionan transformadores hasta cubrir o superar el tamaño teórico y todos los usuarios asociados a cada transformador se incluyen para no perder información. Cada transformador tiene una cantidad de usuarios distinta. Entonces, las variaciones en tamaño dependen de la ponderación que afecta el orden del ranking de transformadores.

Las Tablas III y IV muestran la distribución de usuarios por subregión con confianza 95% y pesos 90/10, además de la distancia promedio entre usuarios. La subregión Centro es la zona más urbanizada de la región y se observa una distancia promedio menor con respecto a la población original. Esto se debe a la formación de cúmulos densos en estas zonas urbanas. La distancia promedio general aumenta 11.48 km respecto a la población. Esto se explica por la menor cantidad de usuarios y la formación de cúmulos que deja vacíos entre subregiones, especialmente en la montañosa subregión Oriente.

La Fig. 5 destaca un ajuste excelente con respecto al consumo anual. Esto se logra a pesar de que la muestra es 325 veces más pequeña que la población y no usa el consumo como criterio de selección. Esto indica que las categorías de tipo de cliente y tipo de población contienen información relevante sobre los patrones de consumo. Este resultado es importante porque muestra que la muestra puede describir la demanda de la población original. En la Fig. 6 se observa que la muestra sub-representa a los usuarios de estrato 3 y no incluye a los estratos 4, 5 y 6. Se hace evidente un sesgo que excluye a los usuarios de estrato alto. También se observa una fuerte sobre-representación de usuarios de estrato 1. Corregir este sesgo requeriría aumentar el nivel de confianza. Esto implicaría un incremento importante en el tamaño de la muestra. Las metodologías basadas en consumo y divergencia KL exploran un enfoque basado en consumo con el que se busca aliviar este sesgo hacia usuarios más atípicos. La Fig. 7 muestra que la muestra tiene una proporción ligeramente menor de usuarios rurales y una mayor de usuarios urbanos. Aun así, el ajuste general para la categoría de tipo de población es bueno. Esto confirma que el balance 90/10 es adecuado para esta dimensión.

En síntesis, la estrategia de muestreo por entropía permite construir muestras con alta diversidad informativa sin usar muestreo aleatorio sobre el consumo como criterio directo. Gracias a esto, el enfoque brinda resultados deterministas. Se identifican agrupaciones en áreas densamente pobladas. Estas reflejan excelentemente la distribución de consumo y demanda. La ponderación entre tipo de cliente y población facilita conservar representatividad en grupos clave como zonas rurales y usuarios de bajo estrato. Sin embargo, la muestra tiende a sub-representar usuarios con consumos atípicos, como los usuarios industriales o de alto estrato. En conjunto, es una alternativa sólida para obtener muestras compactas, ricas en información y logísticamente manejables.

TABLA II. DIFERENCIA DE LAS PROPORCIONES DE LAS MUESTRAS CON LAS DE LA POBLACIÓN GENERAL EN FUNCIÓN DE LAS CATEGORÍAS SEGÚN PROPORCIÓN DE MEZCLA.

Diferencias en las proporciones de tipo de cliente									
	100	90	80	70	50	30	20	10	0
Total	13.5	44.7	35.2	34.9	51.5	60.6	60.6	65.8	76.5
Diferencias en las proporciones de tipo de población									
Total	48.2	6.2	18.2	16.5	56.5	65.2	65.2	70.8	69.4
Acc.	**61.7**	**50.9**	**53.3**	**51.4**	**108.0**	**125.8**	**125.8**	**136.6**	**145.9**

a. El nombre de cada columna representa el peso para el tipo de cliente.

TABLA III. DIFERENCIA EN LAS PROPORCIONES ENTRE POBLACIÓN Y MUESTRA BASADA EN ENTROPÍA DE SHANNON POR TIPO DE CLIENTE Y REGIÓN.

Diferencias en las proporciones según tipo de cliente												
Región	A.P	Com	Ind	Ofc	Otr	Prov	E1	E2	E3	E4	E5	E6
Nor.	0.6	0.8	0.3	0.4	0.0	0.4	19.1	7.6	27.7	0.7	0.1	0.1
Sur	0.4	9.3	0.3	0.3	0.0	0.4	29.0	19.9	8.4	8.1	0.4	0.1
Cen.	0.7	6.9	0.3	0.3	0.0	0.4	46.5	8.3	24.9	11.6	4.0	6.4
Ori.	0.6	6.9	0.1	1.0	0.0	0.4	7.0	11.9	9.5	0.8	0.2	0.1
Gen.	0.6	1.6	0.3	0.3	0.0	0.0	20.2	2.1	8.6	7.0	1.7	2.4

TABLA IV. DIFERENCIA DE LAS DISTANCIAS PROMEDIO ENTRE USUARIOS Y LAS PROPORCIONES ENTRE POBLACIÓN Y MUESTRA POR TIPO DE POBLACIÓN Y REGIÓN.

Diferencias en las proporciones según tipo de Población			**Distancia**
Región	C. Poblado	Rural	Urbano	AVG [km]
Nor.	18.32	26.72	45.04	06.04
Sur	3.41	20.00	23.41	6.78
Cen.	19.54	4.87	24.41	0.43
Ori.	14.31	20.57	34.88	2.00
Gen.	0.41	03.09	2.68	11.48

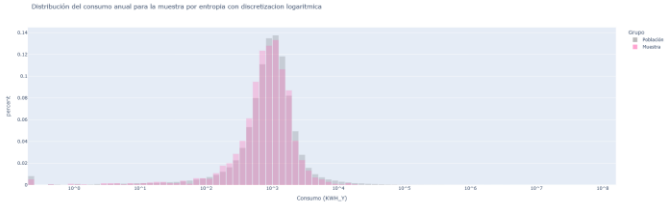

Fig. 5. Comparativa de distribución en función del consumo anual.

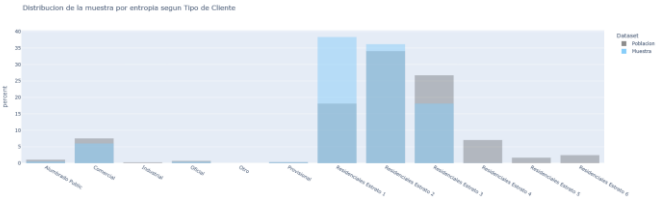

Fig. 6. Comparativa de distribución en función del tipo de cliente.

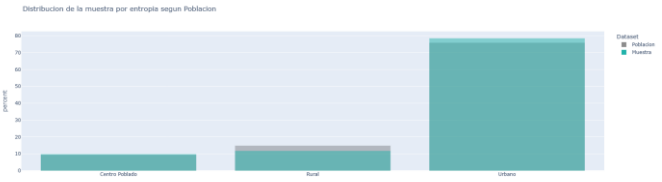

Fig. 7. Comparativa de distribución en función del tipo de población.

VI. Análisis y Discusión de Resultados

En esta sección se analizan las ventajas, limitaciones y aplicaciones recomendadas de la estrategia propuesta. Se consideran diversos criterios de evaluación. Adicionalmente, se consideran especialmente contextos como los de una región en desarrollo, donde se requiere un balance entre representatividad y eficiencia logística.

Uno de los criterios de evaluación más importantes para este estudio es el de **facilidad logística y representatividad** frente a la población. El análisis comparativo entre estos criterios revela características contrastantes pero complementarias. Desde el punto de vista logístico, la estrategia basada en entropía ofrece una ventaja clara en regiones con baja infraestructura. Esta metodología tiende a generar cúmulos de usuarios ubicados en zonas específicas. Esta concentración facilita la implementación de proyectos de monitoreo o instalación de infraestructura, como sensores de medición. La agrupación geográfica reduce los desplazamientos necesarios. Esto disminuye los costos de transporte, mantenimiento y operación. La posibilidad de monitorear segmentos compactos y representativos permite una implementación más eficiente y viable en territorios de difícil acceso o con recursos operativos limitados. En términos de representatividad, la estrategia por entropía logra capturar con precisión la estructura real de la demanda, aunque no use el consumo como criterio directo. El excelente ajuste a la distribución de consumo indica que las variables categóricas seleccionadas contienen información latente relevante sobre los patrones de uso energético. Esto permite construir una muestra significativamente más pequeña sin perder diversidad socioeconómica ni distorsionar los patrones reales de consumo. Esta característica es especialmente útil en contextos con presupuestos restringidos donde se necesita una muestra válida sin comprometer la calidad del análisis.

La **robustez metodológica** de la estrategia es fundamental porque permite determinar qué tan consistentes, fiables y replicables son los resultados ante variaciones en los datos, parámetros o condiciones iniciales. Una estrategia robusta garantiza que las conclusiones obtenidas no dependen de circunstancias particulares o aleatorias. Esto fortalece la validez científica del estudio y su aplicabilidad en contextos reales con incertidumbre o limitaciones operativas. La estrategia por entropía destaca por su solidez metodológica. Se fundamenta en el concepto de entropía de Shannon de la teoría de la información y permite identificar subconjuntos representativos a partir de variables categóricas como el tipo de cliente y el tipo de población. Esta característica la hace especialmente adecuada en contextos con baja disponibilidad o calidad de datos cuantitativos, como ocurre en regiones con infraestructura limitada o registros incompletos. Esta estrategia es menos sensible a fluctuaciones estacionales o errores de medición al no depender de valores continuos o series temporales como el consumo eléctrico.

Por otro lado, la **interpretabilidad de los resultados** es fundamental porque permite comprender y justificar las decisiones tomadas durante el proceso de muestreo. Esto facilita la validación, la comunicación y la aceptación de los hallazgos por parte de tomadores de decisiones y perfiles técnicos. Una estrategia con resultados interpretables mejora la transparencia metodológica y permite detectar sesgos potenciales y refuerza la confianza en que la muestra represente adecuadamente a la población objetivo. También facilita la replicabilidad del estudio y su uso como base para decisiones de política o planificación. La estrategia basada en entropía ofrece una ventaja clara en términos de interpretabilidad frente al enfoque basado solo en consumo. Esto se debe a que cuenta con una métrica cuantitativa clara como la entropía de Shannon (1). La lógica de maximizar la diversidad informativa en función de variables conocidas permite justificar de forma técnica la selección de la muestra. Esta claridad facilita la comunicación del proceso y promueve la apropiación de los resultados por parte de actores no técnicos, especialmente en el sector público. El tomador de decisiones puede desconocer como se recolecta la información de consumo. En cambio, categorías como tipo de cliente o tipo de población resultan familiares y comprensibles.

La **flexibilidad y adaptabilidad de la estrategia** determinan su capacidad para ajustarse a distintos objetivos, restricciones o contextos operativos sin perder efectividad. En escenarios reales, los requerimientos pueden cambiar según las metas del estudio o las condiciones del entorno. Una estrategia adaptable facilita su aplicación en múltiples casos de uso. Esto aumenta su utilidad práctica y su potencial de escalabilidad. En el enfoque basado en entropía es posible ajustar la ponderación de las variables que componen el ranking de transformadores. Esto permite adaptar el criterio de selección según las prioridades del estudio. Por ejemplo, se puede favorecer la diversidad geográfica, ampliar la variedad de tipos de cliente o incluir nuevas dimensiones como estrato socioeconómico o indicadores de vulnerabilidad. Esta flexibilidad permite aplicar la estrategia en diferentes escalas territoriales y sectores sin perder capacidad para generar muestras informativas y equilibradas. A continuación, se explorarán algunas aplicaciones alternativas potenciales de la estrategia de muestreo propuesta. Esto para resaltar la adaptabilidad que esta estrategia pueda tener.

La estrategia de muestreo basada en entropía tiende a formar agrupaciones de usuarios en zonas específicas, especialmente en áreas urbanas. En lugar de cubrir toda la ciudad de forma homogénea, la muestra se concentra en sectores como barrios completos. Este patrón se ha observado en estudios recientes

que usan análisis de entropía para identificar patrones de consumo en entornos urbanos. Por ejemplo, un estudio sobre demanda de agua aplicó entropía y agrupamiento de series temporales para identificar patrones residenciales distintos. Este trabajo mostró cómo algunas zonas presentan comportamientos de consumo más homogéneos que otras [20]. En otro estudio sobre demanda eléctrica se usaron métricas basadas en entropía para maximizar la diversidad informativa con muestras pequeñas. Esto permitió optimizar la caracterización y gestión del consumo residencial [21]. La propiedad de este enfoque de generar agrupaciones es útil más allá del sector eléctrico. En contextos con restricciones logísticas o presupuestales permite seleccionar pocos usuarios agrupados que representan bien a la población. Esto optimiza el uso de recursos. En estudios ambientales, los métodos basados en entropía han sido efectivos para elegir puntos de observación que maximizan la información recolectada, incluso en escenarios urbanos complejos. Se ha usado modelado de máxima entropía para seleccionar sitios de muestreo en redes ambientales. Esto permitió capturar la heterogeneidad con pocos puntos de observación [22]. También, se han usado criterios de entropía para optimizar redes de sensores en ciudades. Estos criterios permitieron detectar fuentes de emisiones peligrosas con pocos sensores [23]. En redes de agua, métricas de entropía ayudaron a ubicar un número limitado de sensores que detectan pérdidas con alta eficiencia [24]. En contextos industriales se usó una estrategia basada en entropía para estimar las fuentes de emisiones en plantas químicas y optimizar la ubicación de sensores para captar la variabilidad de las condiciones operativas [25]. Estos enfoques evidencian la utilidad de la entropía como herramienta para diseñar redes de monitoreo eficientes y representativas en contextos con limitaciones operativas. La capacidad de la estrategia basada en entropía para formar agrupaciones que representan patrones complejos la convierte en una herramienta eficaz para apoyar decisiones en entornos heterogéneos.

VII. CONCLUSIONES

La estrategia basada en entropía para el muestro se destaca por su implementación sencilla que permite ofrecer una herramientas de caracterización de demanda eléctrica en contextos con infraestructura limitada. Permite focalizar en zonas densamente pobladas con alta heterogeneidad informativa y ofrece un enfoque determinista libre de aleatoriedad. Esto es útil en regiones con restricciones geográficas o económicas, ya que optimiza el uso de recursos al enfocarse en áreas accesibles e informativamente relevantes. Este aporte es relevante siempre que no se requiera una muestra muy dispersa geográficamente. Su principal limitación es la baja representatividad de usuarios con consumos atípicos. Esto reduce su aplicabilidad para caracterizaciones más integrales. Sin embargo, puede ser efectiva si se complementa con categorías que incluyan información de zonas menos representadas en la ponderación. Adicionalmente, su aplicabilidad se extiende a potencial utilidad en tareas como la ubicación óptima de sensores, el diseño de redes de monitoreo, la planificación en entornos urbanos, o la selección de beneficiarios para programas focalizados en distribución eficiente de recursos.

En términos de robustez metodológica y adaptabilidad, la muestra por entropía se apoya en variables categóricas y no depende de datos continuos. Esto la hace menos sensible a errores de medición o vacíos en los registros, especialmente en contextos con infraestructura limitada. La posibilidad de ajustar el criterio de ponderación permite adaptar la estrategia a distintos objetivos analíticos.

La estrategia basada en entropía destaca por su alto nivel de interpretabilidad. Esto facilita su adopción por parte de tomadores de decisiones no técnicos. Al utilizar variables categóricas conocidas como tipo de cliente o tipo de población y operar con una métrica clara como la entropía de Shannon. Permite comunicar de forma más transparente las decisiones de selección de muestra. Esta cualidad fortalece la confianza en los resultados y facilita la apropiación institucional de los hallazgos, especialmente en sectores públicos que requieren justificar técnicamente la focalización territorial de recursos.

El propósito del trabajo es justamente ofrecer una herramienta flexible según el contexto de la región y los objetivos del operador para implementar metodologías de muestreo en medición inteligente. Aunque la estrategia no es óptima en todos los casos, puede adaptarse a distintos equilibrios entre cobertura, viabilidad logística y representatividad.

REFERENCIAS

[1] Carrillo Romero, J y Perdomo Arias, A (2017). *Caracterización y análisis del consumo energético en zonas rurales para los municipios de Arauca* [Trabajo de grado- Pregrado, Universidad de los llanos, Villavicencio- Colombia]. https://repositorio.unillanos.edu.co/entities/publication/74e07b53-1086-4f9f-9ec0-7a5eddec8374.

[2] Cardot, H and De Moliner, A. (2018). Conditional bias robust estimation of the total of curve data by sampling in a finite population: an illustration on electricity load curves. arXiv preprint arXiv:1806.09949.

[3] Imberg, H, Yang, X, Flannagan, C y Bärgman, J (2024) Active Sampling: A Machine-Learning-Assisted Framework for Finite Population Inference with Optimal Subsamples. *Technometrics, 67 (1),* 46–57,. doi: 10.1080/00401706.2024.2374554.

[4] Reina García, J (2017). *Diseño metodológico para la selección de sitios de muestreo en una red de monitoreo de micro–contaminantes en ríos de Valle: caso de estudio río Cauca* [Trabajo de grado- Maestría, Universidad del Valle, Santiago de Cali- Colombia]. https://bibliotecadigital.univalle.edu.co/server/api/core/bitstreams/1d278333-a5f3-4dff-8851-e8ae3e7aa47a/content

[5] Viloria Rodríguez, A (2024). *Comparación de metodologías utilizadas para abordar el problema de datos faltantes en estudios longitudinales*. [Trabajo de grado- Maestría, Universidad Nacional de Colombia, Medellín- Colombia]. https://repositorio.unal.edu.co/items/8279675a-b486-451a-9287-bdeda6cbe6e9

[6] Rafiee, H, Abin, A and Majd, S. (2024) Cluster Sampling: A Cluster-Driven Sampling Strategy for Deep Metric Learning, *14th International Conference on Computer and Knowledge Engineering (ICCKE),* Mashhad, Iran, Islamic Republic of, 2024, pp. 460-465, doi: 10.1109/ICCKE65377.2024.10874603.

[7] Chen, S, Zheng, J and Li, J. (2024). The Impact of Sample Size after Sampling on the Accuracy of Machine Learning Models. *International Conference on Computers, Information Processing and Advanced Education (CIPAE), Ottawa, ON, Canada*, pp. 61-66, doi: 10.1109/CIPAE64326.2024.00017.

[8] Q. -X. Zhu, Q. -Q. Zhao, Y. Xu and Y. -L. He, (2023). Novel virtual sample generation using Gibbs Sampling integrated with GRNN for

handling small data in soft sensing. *IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China*, 2023, pp. 89-94, doi: 10.1109/DDCLS58216.2023.10166679.

[9] Swan, L. G. and Ugursal, V. I. (2009) Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviewst, 13(8)*,1819-1835 https://doi.org/10.1016/j.rser.2008.09.033.

[10] Valverde Mora, G, Marín, L, Chacón Vásquez, M (2019). *Metodología para la Determinación de Curvas de Carga y Consumo Eléctrico Residencial por Uso.* [Informe Final, Universidad de Costa Rica, San Pedro de Montes de Oca, San José- Costa Rica], Olade, https://biblioteca.olade.org/opac-tmpl/Documentos/cg00912.pdf

[11] İ. Y. Yarbaşı and A. K. Çelik, (2023) The determinants of household electricity demand in Turkey: An implementation of the Heckman Sample Selection model, *Energy,* vol. 283, doi: 10.1016/j.energy.2023.128431.

[12] L. Provincia, S. Elena, C. Pavón, and J. Barzola-Monteses (2015) *Estimación de la demanda energética mensual mediante encuesta aplicada en en la Provincia de Santa Elena*.[Universidad Laica Vicente Rocafuerte de Guayaquil].Available: https://www.researchgate.net/publication/309286132

[13] X. Zhu and B. Mather, (2019) Data-Driven Load Diversity and Variability Modeling for Quasi-Static Time-Series Simulation on Distribution Feeders, *IEEE Power & Energy Society General Meeting (PESGM), Atlanta, GA, USA,* 2, pp. 1-5, doi: 10.1109/PESGM40551.2019.8973929.

[14] X. Sun, X. Li, S. Datta, X. Ke, Q. Huang, R. Huang y Z. J. Hou (2021) Smart Sampling for Reduced and Representative Power System Scenario Selection. *IEEE Open Access Journal of Power and Energy*, *vol. 8*, pp. 241–251, 2021. doi: 10.1109/OAJPE.2021.3093278.

[15] Asghari, P., Zakariazadeh, A., Siano, P.(2022) Selecting and prioritizing the electricity customers for participating in demand response programs. *IET Gener. Transm. Distrib. 16*, 2086–2096. doi: 10.1049/gtd2.12417

[16] J. Henze, S. Kutzner y B. Sick (2018). Sampling Strategies for Representative Time Series in Load Flow Calculations. *Data Analytics for Renewable Energy Integration. Technologies, Systems and Society (DARE 2018)*, W. Woon, Z. Aung, A. C. Feliú y S. Madnick, Eds., vol. 11325, Lecture Notes in Computer Science. Cham: Springer pp. 33–47. doi: 10.1007/978-3-030-04303-2_3.

[17] X. Zhu and B. Mather (2020) Data-Driven Distribution System Load Modeling for Quasi-Static Time-Series Simulation. *IEEE Transactions on Smart Grid, vol. 11(2)*, pp. 1556-1565. doi: 10.1109/TSG.2019.2940084.

[18] Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27 (3),* 379–423, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[19] S. K. Ahmed (2024) How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. *Oral Oncology Reports, vol. 12, p.* 100662. doi: 10.1016/j.oor.2024.100662.

[20] R. Wang, X. Zhao, H. Qiu, X. Cheng, y X. Liu (2024) Uncovering urban water consumption patterns through time series clustering and entropy analysis, *Water Research, vol. 262*, p. 122085. doi: 10.1016/j.watres.2024.122085

[21] T. J. Stohlgren, S. Kumar, D. T. Barnett, y P. H. Evangelista (2011) Using Maximum Entropy Modeling for Optimal Selection of Sampling Sites for Monitoring Networks. *Diversity, vol. 3 (2)*, pp. 252–261. doi: 10.3390/d3020252

[22] P. Ngae, H. Kouichi, P. Kumar, A.-A. Feiz, y A. Chpoun (2019) Optimization of an urban monitoring network for emergency response applications: An approach for characterizing the source of hazardous releases *Quarterly Journal of the Royal Meteorological Society, 16*. doi: 10.1002/qj.3471

[23] M. S. Khorshidi, M. R. Nikoo y M. Sadegh (2018) Optimal and objective placement of sensors in water distribution systems using information theory. *Water Research, vol. 143,* pp. 218–228. doi: 10.1016/j.watres.2018.06.050

[24] H. Tian, Z. Lang, C. Cao y B. Wang (2025) Optimizing Sensor Placement for Enhanced Source Term Estimation in Chemical Plants. *Processes, vol. 13 (3)*, art. 825. doi: 10.3390/pr13030825.

[25] D. Hock, M. Kappes y B. Ghita (2020) Entropy-Based Metrics for Occupancy Detection Using Energy Demand *Entropy, vol. 22, (7)*, art. 731. doi: 10.3390/e2207073