

Evaluación del impacto de las variables bioclimáticas en el consumo: Caso estudio Sector de Aguas de La Habana.

MARIO RAMOS JOSEPH

E-MAIL: mramos@ahabana.co.cu

Empresa Aguas de La Habana, Fomento y Recreo, Palatino, Cerro, La Habana, Cuba

RENIEL CARVAJAL ALFONSO

E-MAIL: rcarvajal@ahabana.co.cu

Empresa Aguas de La Habana, Fomento y Recreo, Palatino, Cerro, La Habana, Cuba

ALCIDES J. LEÓN MÉNDEZ

E-MAIL: aleonm@civil.cujae.edu.cu

Centro de Investigaciones Hidráulicas (CIH), Universidad Tecnológica de La Habana José Antonio Echeverría (Cujae), La Habana, Cuba.

RAISA SOCORRO LLANES

E-MAIL: aleonm@civil.cujae.edu.cu

Centro de Investigaciones Hidráulicas (CIH), Universidad Tecnológica de La Habana José Antonio Echeverría (Cujae), La Habana, Cuba.

RESUMEN

Cada día aumenta la necesidad de gestionar de forma más eficiente los recursos naturales, entre estos el agua se mantiene como elemento primordial, por su importancia para la salud, actividad y desarrollo humanos. Garantizar la demanda a corto, mediano y largo plazo se ha convertido en un gran desafío para investigadores y gestores de agua en todo el planeta. En este trabajo se utiliza un modelo de series temporales para evaluar el impacto de las variables bioclimáticas en el consumo. Este modelo se aplica a diferentes horizontes de consumo histórico en un sector operado por Aguas de La Habana. Se comprueba el efecto novedoso que tienen estas variables bioclimáticas en la predicción de la demanda doméstica. Los resultados obtenidos aportan valores de r^2 ajustada que oscilan entre 0,69 y 0,79.

PALABRAS CLAVES: demanda de agua, ensemble, horizonte de predicción, series temporales, variables bioclimáticas.

Evaluation of the impact of bioclimatic variables on consumption: Case study of the Aguas de La Habana sector.

ABSTRACT

Every day it increases the necessity to negotiate in a more efficient way the natural resources, among these the water stays as primordial element, for its importance for the health, activity and development humans. To guarantee the demand to short, medium and I release term he/she has become a great challenge for investigators and agents of water in the whole planet. In this work a model of temporary series is used to evaluate the impact of the variable bioclimatic in the consumption. This model is applied to different horizons of historical consumption in a sector operated by Waters of Havana. It is proven the novel effect that you/they have these variable bioclimatic in the prediction of the domestic demand. The obtained results contribute securities of r^2 squared that oscillate between 0,69 and 0,79.

KEYWORDS: water demand, ensemble, prediction horizon, time series, bioclimatic variables

01 INTRODUCCIÓN

Entre los principales desafíos que enfrenta la humanidad se encuentra el de la explotación racional del agua, un recurso indispensable y cuya desigual distribución en el planeta conjuntamente con la aparición simultánea de eventos extremos ha provocado la aparición de prolongadas sequías, inundaciones y ciclones.

Ante este escenario, aquellos estudios con la perspectiva de un uso racional y eficiente del agua, así como la predicción de la demanda, son una herramienta necesaria para las empresas de abasto. Una de las necesidades más perentorias en toda empresa de abasto es conocer la demanda de agua de sus clientes y los principales factores que afectan a la misma (Hao, 2022).

De Souza et al. (2019) y Hao et al. (2022) destacan en sus trabajos, que en la actualidad coexisten innumerables métodos para predecir la demanda de agua utilizados por los expertos tales como: Bosque Aleatorio (Random Forest, RF por sus siglas en inglés), Máquina de Vectores de Soportes (Support Vector Machine, SVM por sus siglas en inglés), Redes Neuronales Artificiales (Artificial Neural Network, ANN por sus siglas en inglés) y los Sistemas de Inferencia Neuro Difuso Adaptativo (Adaptative Neuro Fuzzy Inferential System, ANFIS por sus siglas en inglés).

Avalado por los recientes desarrollos realizados, en minería de datos e inteligencia artificial, se han logrado excepcionales avances en su aplicación a prácticamente todas las esferas del conocimiento y dentro de esta la aplicación al mundo de la gestión de los recursos hídricos. A esto se une el cambio climático y sus efectos (Hao et al., 2022) y (De Souza et al., 2019).

Entre los principales métodos se encuentran los que aplican la regresión. La regresión es un método estadístico utilizado para analizar la relación entre variables (DDmáx, 2022) y (Alshaikhli et al., 2021). Es el método más sencillo entre los utilizados para predecir el consumo de agua potable.

Las técnicas de regresión han sido muy utilizadas en estudios para analizar la relación entre el consumo de agua potable y otros factores como la calidad del agua, la infraestructura de distribución y el precio del agua, (Arellano y Peña, 2020). En la regresión lineal simple solo se utiliza una variable independiente para predecir la variable dependiente. Se puede definir la velocidad del viento como variable independiente para con ella predecir el consumo de agua potable (variable dependiente). La regresión lineal múltiple utiliza varias variables independientes, para predecir una variable dependiente. En este caso, las variables independientes pueden ser la población, el clima y el uso del agua, o la densidad poblacional y se reconocen como variables predictoras (Alshaikhli et al., 2021).

Alshaikhli et al. (2021) investigaron los determinantes del consumo de agua en el estado de Qatar, centrándose en los factores vinculados a los meses estacionales mediante el uso de un modelo de regresión lineal múltiple. Los resultados del modelo desarrollado se expresan en una ecuación de regresión y demuestran que un incremento en la temperatura está asociado con un incremento en el consumo doméstico de agua per cápita de 6,523 litros diarios. Y contrariamente a lo anterior, un incremento en la densidad de la población se correlaciona con una disminución del consumo per cápita de agua.

El estudio en Qatar se llevó a cabo sobre una variable dependiente (consumo de agua per cápita), en relación con varias variables independientes (densidad de población, temperatura, humedad, precipitaciones, luz diurna y luz solar) con el objetivo de utilizar la regresión lineal múltiple para describir el impacto del efecto de las variables climáticas y poblacionales en el consumo de agua per cápita en Qatar. Este estudio permitió comprender el efecto combinado de varias influencias en el uso doméstico del agua en Qatar. A pesar de la inclusión de variables socio-demográficas, como

resultado de este modelo de regresión se muestra que el factor climático de la temperatura influye significativamente en la demanda urbana del agua (Alshaikhli et al., 2021).

Mientras tanto en Ecuador un reporte incluyó factores socio-demográficos, socio-económicos, gestión y calidad del agua para definir los modelos matemáticos que predicen el consumo de agua potable residencial. Se aplicaron dos modelos de regresión múltiples, el primero implica 19 variables y el segundo tan solo 6. Se evaluaron los efectos de las variables que influyen en el consumo del agua potable, obteniéndose como resultado, que solamente las variables "número de personas por vivienda" y "nivel de ingresos económicos" de las 19 presentadas, inciden significativamente en el consumo semestral de agua potable. Mientras tanto en el segundo modelo las variables climatológicas "humedad atmosférica máxima" y "temperatura máxima" tienen más significancia estadística que las otras 4 variables que intervienen en el consumo mensual de agua potable (Arellano y Peña, 2020).

En Cuba la empresa Aguas de La Habana gestiona el abasto de agua a la capital. Desde el año 2010, se creó un grupo ad hoc para el estudio de la demanda de agua del sector poblacional que atiende la empresa. El grupo liderado por el primer autor de este trabajo, ha tenido entre sus miembros a especialistas de la propia empresa que han realizado sus estudios de maestría en este tema, además de especialistas del Centro de Investigaciones Hidráulicas, el Instituto de Meteorología, y la Facultad de Informática, en el caso de estos últimos, su participación permitió un vuelco favorable en la evolución de esta investigación. Este grupo de profesionales, conjuntamente con numerosos alumnos del último año de la carrera de Ingeniería Hidráulica de la Universidad Tecnológica de La Habana, Cujae, han trabajado coordinadamente para establecer indicadores, herramientas y llegar a conclusiones que faciliten la toma de decisión. En el caso de los estudiantes, el trabajo realizado les ha servido para escribir sus trabajos de culminación de estudios. A continuación, se relacionan los principales trabajos del grupo ad hoc referenciados en Pomares (2022).

Dentro de los resultados obtenidos está el de Senovia Iglesias Granados en el año 2013, que tuvo como objetivo determinar las curvas de los patrones de consumo de un grupo de clientes domésticos con servicio diario de 24 horas (24/7), estableciendo una metodología para su determinación. Además, se determinaron las curvas de consumos clasificados para tres clientes domésticos con servicio continuo, donde se aprecian elevados consumos a bajos caudales; y, sin consumos, por encima de 750 L/h. También se determinó el tipo de contador que mejor se adecuaba a las curvas obtenidas y finalmente se determinó el volumen de pérdidas aparentes, que implicó obtener, para el sector, la curva promedio de consumo.

En el año 2015, Rosaura Socarrás Ordaz, cumple con el objetivo propuesto de determinar las curvas de consumo clasificadas de una muestra de clientes en un sector de la empresa Aguas de La Habana con servicio 24/7 durante un período de 6 meses, utilizando contadores inteligentes. Además, se determina el perfil de consumo de una muestra de 14 clientes estableciendo los horarios de mayor y menor consumo. También se determina el volumen de pérdidas aparentes producto de submedición durante el tiempo que duró el experimento y se definió el tipo de contador más idóneo para los clientes de acuerdo a su forma de consumir agua.

En el año 2017, Edelino Guilherme Foquiço, tuvo como objetivo en su trabajo determinar los patrones de consumo de una muestra de clientes de la empresa Aguas de La Habana. Para lograr el objetivo se siguen las pautas de trabajo establecidas por Socarrás Ordaz aplicadas a una muestra de 14 clientes durante un período de 25 meses y utilizando el programa de estadística multivariada Table Curve 3D, se obtienen a resultados similares al trabajo de Socarrás. Aunado a esto, se realiza un estudio pionero en Aguas de La Habana donde se determina la relación entre el consumo per cápita

doméstico y las variables climáticas y bioclimáticas en una muestra de clientes durante un período de 35 meses, determinando el efecto individual y combinado de las variables climatológicas y bioclimáticas contra el consumo. De aquí se destaca que el efecto simultáneo de las variables humedad relativa máxima y temperatura máxima provoca aumentos significativos del consumo per cápita con una correspondencia de 28,6 %. Se determina que, de las variables climatológicas, las variables de precipitación media, humedad relativa media y temperatura media son las que menos variabilidad del consumo inducen. Por último, se obtiene que, de las variables bioclimáticas, la temperatura efectiva equivalente es la variable de mayor incidencia en el consumo.

En el mismo año, en el trabajo de Jessica Romaguera Palomino, se determinan los patrones de consumo de una muestra de clientes de la empresa y su relación con variables climáticas y bioclimáticas utilizando el programa Table Curve 3D. Con este resultado se logra insertar a la empresa Aguas de La Habana, en la carrera por ganar conocimiento acerca de la influencia de la variabilidad climática en el consumo residencial y se concluyó que, al contrastar los datos de las estaciones meteorológicas de Santiago de las Vegas y Casablanca, la que mejor expresa el comportamiento bioclimático de los consumidores es esta última. Los resultados que se obtuvieron constituyen una herramienta inmediata para áreas como la operación, el planeamiento y la creación de planes de contingencia que garanticen resiliencia ante el cambio climático.

En el año 2020, David Sosa Izquierdo se trazó como objetivo caracterizar el grado de influencia de los factores climáticos y bioclimáticos en el consumo de agua residencial. En el estudio se pudo concluir que la temperatura es la variable que más afecta al consumo per cápita, siendo la precipitación la variable que mostró menos influencia.

Es importante reconocer que hasta ese momento en los estudios acerca de la influencia de las variables climáticas y bioclimáticas en el consumo de agua, utilizando herramientas estadísticas, se lograron importantes resultados de inmediata aplicación para la toma de decisiones de la empresa.

En el año 2022 se aplican por vez primera herramientas de minería de datos para el posprocesamiento de los datos. En ese año Josué Pomares Palermo logra, aplicando estas herramientas, definir la relación entre el consumo doméstico y las principales variables climatológicas y cuál de las variables bioclimáticas es la de mayor influencia. A partir de una muestra de 34 clientes para un período de 13 años, nunca antes utilizado en un análisis de este tipo, se logró determinar las variables de mayor impacto sobre el consumo. Estas variables fueron: la amplitud térmica, la humedad relativa máxima y la lámina de lluvia y para el caso de las variables bioclimáticas, la de mayor impacto sobre el consumo fue la temperatura efectiva equivalente. Con la colaboración de expertos del Instituto de Meteorología de Cuba (INSMET), se incluyó la amplitud térmica ($T_{\text{máx}}-T_{\text{mín}}$), como una nueva variable bioclimática (Pomares, 2022).

En una reciente contribución Ramos Joseph et al. (2023), apoyado en los resultados anteriores y utilizando técnicas de minería de datos específicamente, aplica el método de selección de variables y se logra determinar aquellas variables climáticas y bioclimáticas que impactan en el consumo, distinguiendo las que más lo afectan y se destacan como predictoras en una muestra del sector AB de Aguas de La Habana. Los experimentos fueron realizados para 6 técnicas de regresión, y se logró identificar que las variables que mejor explican el comportamiento de la demanda son: intensidad de la precipitación, velocidad máxima del viento, amplitud térmica, humedad relativa máxima y humedad relativa mínima y de acuerdo al modelo son las variables predictoras.

Relevante resultó ser, el caso de las variables bioclimáticas, de las cuales, la que más aparece es la amplitud térmica, por lo que definitivamente es la variable que se impone en este acápite, seguida

por la temperatura efectiva y la temperatura efectiva equivalente. Es de destacar el hecho de que la velocidad máxima del viento tiene en cuenta el comportamiento sensorial lo que demuestra la importancia de las variables bioclimáticas en la predicción del consumo.

Hasta este momento se ha logrado reconocer el impacto que tiene la variabilidad climática sobre el consumo en clientes de Aguas de La Habana expresado en el conocimiento adquirido en los estudios realizados. El carácter multifactorial del consumo ha obligado a incluir un mayor número de variables independientes climáticas y bioclimáticas en los estudios. Este panorama muestra la necesidad de conocer cómo evoluciona la demanda en el tiempo, que finalmente se traduce en predecir el consumo a muy corto, corto, mediano y largo plazo utilizando la minería de datos.

El objetivo de este trabajo es evaluar el impacto que tienen las variables bioclimáticas en la demanda de agua doméstica mediante un modelo de series temporales aplicado a un sector de Aguas de La Habana.

02 ESTRATEGIA DE TRABAJO UTILIZADA

En esta sección se explican las principales características del sector hidráulico seleccionado, las de los clientes de la muestra, el proceso de obtención de las variables climáticas, la herramienta de minería de datos seleccionada y el procesamiento de los datos.

Para lograr el objetivo propuesto y teniendo en cuenta los antecedentes expresados anteriormente se seleccionó un sector de la empresa que cumpliera con los requisitos necesarios (servicio 24/7, información precisa y veraz de los clientes) y al mismo tiempo estuviera claramente dentro del área de influencia de la estación meteorológica de Casablanca con los datos de las principales variables climáticas. El sector seleccionado pertenece a la empresa Aguas de La Habana y se denominara a partir de aquí sector AB.

El sector de estudio tiene como características fundamentales estar ubicado en uno de los municipios más importantes de la ciudad. Está compuesto por viviendas de la década de los años 50 del siglo pasado, las redes hidráulicas en su totalidad tienen una longitud total de algo más de 1500 metros, con más de 340 clientes fundamentalmente domésticos, distribuidos en viviendas unifamiliares y multifamiliares con servicio de agua continuo (24/7). Como a cada sector de la empresa, se le mantiene un seguimiento en la evolución de sus principales parámetros en el tiempo desde principio de la década del año 2000.

De este sector se tomaron los datos de consumo de agua de las viviendas unifamiliares, que son un total de 46 del total de clientes del sector. La información fue aportada por la Dirección Comercial de la empresa a partir del historial de consumo de los clientes del sector. Después de un procesamiento de datos estadístico que abarcó la información del 2000 hasta el 2020, se seleccionaron 34 clientes que cumplieran con estabilidad en las mediciones y con mínimo de roturas de los contadores. A estos 34 clientes se les determinó el consumo promedio mensual durante el periodo de análisis desde el año 2006 al 2020 expresado en litros por persona por mes (lppm) para su contraste con las variables meteorológicas.

Para acceder a las variables climáticas la información fue gestionada por el INSMET en sus registros diarios y se realizó un procesamiento de la información para llevarla a valores mensuales. Se incluyó un grupo de variables de las cuales algunas se mencionaron anteriormente, que intervienen en el comportamiento del consumo doméstico del agua y en el caso de otras, su inclusión es realmente novedosa, por ejemplo variables bioclimáticas como la temperatura efectiva, la temperatura efectiva equivalente y la amplitud térmica. Las variables meteorológicas se expresan en valores mensuales

para su contraste con el consumo residencial mensual y fueron conceptualizadas en Ramos Joseph et al. (2023).

Para procesar la información recolectada se define como variable dependiente el consumo doméstico expresado en (lppm), a partir de los consumos domésticos promedios de los clientes de la muestra en el periodo comprendido desde el 1 de enero del 2006 hasta el 31 de diciembre del 2020. El resto de las variables meteorológicas se definen como variables independientes. Con esto se conforma la vista minable con un total de 180 registros. Es relevante destacar que con este estudio se realiza por vez primera la predicción de consumos de agua domésticos en Aguas de La Habana a partir de la variabilidad climática. Se incluyen, además, las variables bioclimáticas las cuales incorporan el efecto sensorial sobre los seres vivos, lo cual es novedoso. Para este ejercicio de predicción se utiliza un horizonte de predicción de 3, 6, 9 y 12 meses seleccionados de acuerdo a la necesidad de la empresa. Las conclusiones anteriores se resumen como novedosas para el país.

Las variables independientes empleadas fueron precipitación (LI), ocurrencia de lluvia ($OcLI$), intensidad de las precipitaciones (r_{24h}), temperatura mínima, media y máxima ($T_{mín}$, T_{med} , $T_{máx}$), humedad relativa mínima, media y máxima ($Hr_{mín}$, Hr_{med} , $Hr_{máx}$), velocidad promedio del viento (FF_{med}), dirección del viento ($DD_{máx}$, amplitud de la humedad relativa (ΔHr), temperatura efectiva (TE), temperatura efectiva equivalente (TEE), y amplitud térmica (ΔT). El significado de cada variable aparece explicado en Ramos Joseph et al. (2023).

En función de todo lo anterior se confecciona una vista minable compuesta por 180 filas y 19 columnas de las que se tienen 16 variables independientes, dentro de las que sobresalen algunas consideradas como bioclimáticas, una variable dependiente y dos variables temporales.

Como analizador y procesador de los datos se seleccionó el software KNIME, el cual es una herramienta utilizada para el procesamiento de variables y lograr obtener el conocimiento oculto que tienen los datos, sobre la base de criterios matemáticos y estadísticos.

Para confirmar la validez de los resultados que se obtengan se trabajara con el r^2 ajustado, como métrica fundamental, la misma tiene como característica que se utiliza para comparar modelos con diferentes cantidades de variables predictoras y en función del mayor valor reconocer el mejor de los modelos comparados a partir de tener una menor cantidad de predictores.

En la literatura consultada se detecta que la mayoría de los investigadores que trabajan predicción del consumo urbano utilizan estas métricas independientemente del modelo utilizado ya sea regresión lineal simple, regresión múltiple o redes neuronales.

Las variables predictoras obtenidas se utilizan para obtener el modelo de regresión múltiple con el objetivo de estimar el consumo en lppm. Para lograr esto se definieron varios experimentos que se describen a continuación en varios experimentos

CARACTERÍSTICAS DE LOS TRES EXPERIMENTOS REALIZADOS.

A continuación, se exponen los objetivos y características de tres experimentos realizados para evaluar las relaciones funcionales entre las variables climáticas y bioclimáticas con el consumo de agua en el sector en estudio.

Primer experimento

Su objetivo es el de evaluar la influencia del consumo histórico en la estimación del valor del consumo real. En este experimento se plantea el uso de una serie temporal a partir de varios horizontes de predicción: a corto y mediano (3, 6, 9 y 12 meses).

Este Modelo Base de Serie Temporal se implementó en el software KNIME, buscando las variantes acordes a los 4 horizontes propuestos. La figura 1 muestra los nodos que se utilizan en la conformación del Modelo Base. Como se observa, el primer nodo es donde se carga la vista minable obtenida a partir del proceso anteriormente descrito. El segundo nodo construye la serie temporal a partir de un metanodo de aprendizaje compuesto por una acción de desfasado (lag column) y una de fraccionamiento (partitioning) y el nodo de aprendizaje tal y como se muestra en la figura 2.

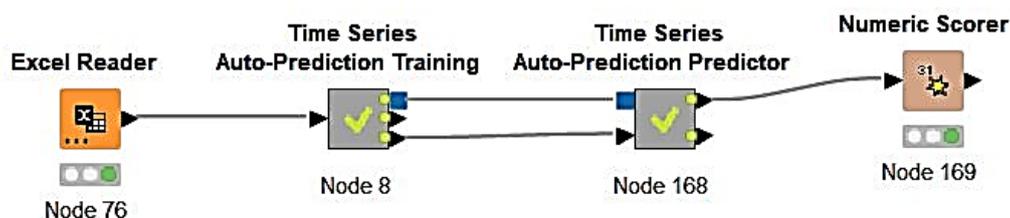


Figura 1 Modelo Base de Serie temporal empleado. Fuente KNIME.

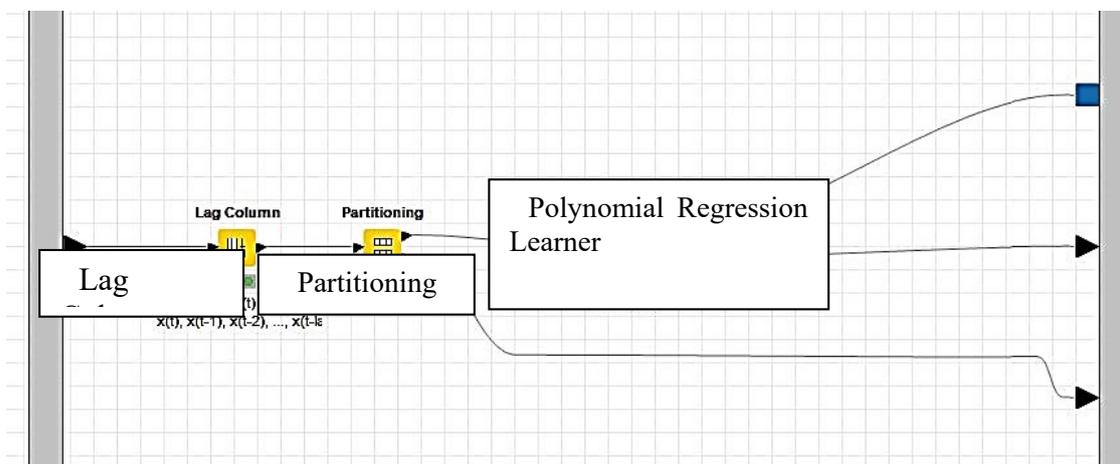


Figura 2. Metanodo de aprendizaje. Fuente KNIME

El Modelo Base de la figura 1 se aplica para cada uno de los horizontes de predicción seleccionados.

Segundo experimento

El objetivo de este experimento es eliminar aquellas variables que no tienen un impacto directo en el nivel sensorial explicado por las variables bioclimáticas. En este caso de las 2 variables temporales incluidas se decide excluir el año.

A continuación, se repite la metodología desarrollada en Ramos et al. (2023) eliminando entre las variables a analizar el año. Esta metodología consiste en utilizar el método de selección de variables, en este caso aplicando el forward o la introducción sucesiva de variables y el backward o la eliminación progresiva de variables. Contrastando ambos métodos se puede determinar las variables que tienen mayor influencia en la estimación del consumo.

Tercer experimento

Analizar diferentes escenarios con presencia de combinaciones de variables climáticas tradicionales y variables bioclimáticas.

El estudio termina con un ejercicio de comparación de las mejores combinaciones de variables predictoras que se obtienen.

03 ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

En esta sección se muestran los resultados de los tres experimentos definidos en la sección anterior.

El primer experimento se realizó siguiendo la misma estrategia de partición del conjunto de datos descrito en Ramos et al. 2023 y aplicando para los 4 horizontes el Modelo Base Temporal presentado en la figura 1. En cada horizonte se contrastan dos casos el primero relacionando el consumo con todas las variables climáticas y bioclimáticas (16 en total) y en el segundo las variables predictoras seleccionadas en que se reconocerá como la V2023 (Ramos et al., 2023). Los resultados de este experimento se muestran en la figura 4. En este primer resultado se puede comprobar en ambos escenarios sin tener en cuenta el consumo histórico que el valor de la métrica r^2 ajustado es mínimo (0,33) y por lo tanto no permite predecir de forma satisfactoria.

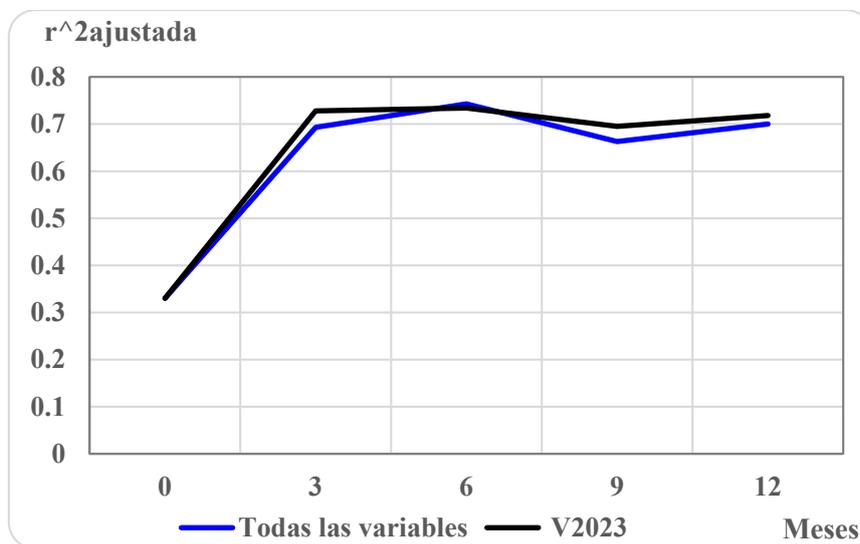


Figura 4. Predicción del consumo en función del r^2 ajustada.

Otro aspecto a destacar del experimento, es que en los dos conjuntos de datos utilizados los valores de r^2 ajustada, para los 4 horizontes analizados, mejoran los resultados iniciales que no tenían en cuenta el consumo histórico para realizar la predicción. Se puede apreciar que los mejores resultados se obtienen al tener en cuenta el consumo de los últimos 6 meses. También se puede apreciar que cuando se utilizan variables predictoras (V2023) se pueden obtener resultados similares o mejores a cuando se utiliza todas las variables lo cual implica una notable reducción de tiempo y recursos en la selección y búsqueda de datos para el proceso de predicción.

Con el segundo experimento se evalúa el criterio inicial de los autores, de que las variables que tienen un impacto sensorial pueden ser utilizadas para estimar el consumo de agua. Este experimento se realiza utilizando el mismo conjunto de datos de la V2023, pero en esa ocasión eliminando la variable año, lo llamaremos V2023A a partir de ahora. Los resultados obtenidos se pueden observar en la tabla 1 y se comparan con los obtenidos en el experimento anterior para V2023.

Tabla 1 Efecto de la variable año en la predicción de consumos.

| Escenario | Horizonte de predicción (mes) | r ² ajustada con año | r ² ajustada sin año |
|-----------|-------------------------------|---------------------------------|---------------------------------|
| V 2023 | 3 | 0,728 | 0,724 |
| | 6 | 0,734 | 0,723 |
| | 9 | 0,695 | 0,701 |
| | 12 | 0,718 | 0,708 |

Como se puede observar la presencia del año no marca una diferencia significativa en los resultados de predicción para la métrica utilizada, esto permite afirmar que se pueden realizar ejercicios de predicción del consumo sin incluir el año y obtener resultados satisfactorios. Es relevante comentar que la no utilización del año aumenta el valor de los resultados obtenidos al no estar sujetos a la variable anual, lo cual los hace más universales.

En el tercer experimento, a partir del resumen de los resultados anteriores se decidió repetir el método de selección de variables para un grupo de algoritmos y proceder nuevamente al proceso de conformación de variables predictoras en este caso sin el año.

Los resultados para obtener las variables predictoras en estos nuevos escenarios se muestran en la tabla 2 con la utilización del método de selección de variables Forward. Cada algoritmo presenta entre paréntesis el número de escenarios con un r² ajustado mayor igual a 54,7 %.

Tabla 2. Frecuencia de aparición de las variables aplicando el método de selección de variables Forward sin año para diferentes algoritmos.

| Variables/Escenarios | Isotonic Regression (9) | Linear Regression (4) | Pace Regression (4) | RBF Regressor (4) | Additive Regression (4) | Total (25) | % Aparición |
|------------------------|-------------------------|-----------------------|---------------------|-------------------|-------------------------|------------|-------------|
| <i>LPPM</i> | 9 | 4 | 4 | 4 | 4 | 25 | 100 |
| <i>Tmín</i> | 9 | 4 | 4 | 1 | 4 | 22 | 88 |
| <i>Tmed</i> | 9 | 1 | 2 | 4 | 1 | 17 | 68 |
| <i>Tmáx</i> | 9 | 2 | 4 | 4 | 4 | 23 | 92 |
| ΔT | 9 | 3 | 1 | 4 | 4 | 21 | 84 |
| <i>Hrmín</i> | 7 | 4 | 4 | 4 | 4 | 23 | 92 |
| <i>Hrmed</i> | 1 | 1 | 2 | 1 | 4 | 9 | 36 |
| <i>Hrmáx</i> | 9 | 4 | 4 | 4 | 1 | 22 | 88 |
| ΔHr | 4 | 4 | 2 | 3 | 4 | 17 | 68 |
| <i>OclI</i> | 9 | 4 | 4 | 0 | 3 | 20 | 80 |
| <i>TE</i> | 2 | 4 | 4 | 4 | 2 | 16 | 64 |
| <i>TEE</i> | 6 | 3 | 0 | 0 | 4 | 13 | 52 |
| <i>r²4h</i> | 9 | 4 | 4 | 0 | 1 | 18 | 72 |
| <i>Lluvia</i> | 9 | 0 | 4 | 2 | 4 | 19 | 76 |
| <i>Ffmed</i> | 3 | 2 | 2 | 0 | 3 | 10 | 40 |
| <i>Ffmáx</i> | 5 | 4 | 4 | 0 | 4 | 17 | 68 |
| <i>Ddmáx</i> | 8 | 4 | 3 | 1 | 4 | 20 | 80 |

A partir del análisis de la tabla 2 se puede resumir que todas las variables influyen en el consumo doméstico con una frecuencia de aparición que abarca desde 36 % hasta 92 %. Las variables con mayor peso son $T_{m\acute{a}x}$, $T_{m\acute{i}n}$, $H_{r_{m\acute{a}x}}$, $H_{r_{m\acute{i}n}}$ y ΔT .

Mientras tanto para las variables bioclimáticas la de mayor relevancia es la ΔT , seguida por la ΔH_r , la TE y la TEE en un rango que abarca desde 52 % hasta 84 % de frecuencia de aparición.

En el mismo ejercicio los resultados de la selección de variables para el backward se aprecian en la tabla 3. En la misma cada algoritmo presenta entre paréntesis el número de escenarios con un r^2 ajustado mayor igual a 50 %.

Tabla 3. Frecuencia de aparición de las variables aplicando el método de selección de variables Backward sin año para diferentes algoritmos.

| Variables/ Escenarios | Isotonic Regresión (8) | Linear Regresión (5) | Pace Regresión (6) | RBF Regressor (4) | Ensemble Regressor (5) | M5P (7) | Total (35) | % Aparición |
|---|---------------------------|-------------------------|-----------------------|----------------------|---------------------------|---------|------------|-------------|
| LPPM | 8 | 5 | 6 | 4 | 5 | 7 | 35 | 100,00 |
| T mín | 1 | 0 | 6 | 3 | 5 | 2 | 17 | 48,57 |
| T med | 2 | 1 | 6 | 4 | 2 | 7 | 22 | 62,86 |
| T máx | 8 | 5 | 6 | 4 | 5 | 7 | 35 | 100,00 |
| ΔT | 8 | 4 | 4 | 2 | 1 | 4 | 23 | 65,71 |
| $H_{r_{m\acute{i}n}}$ | 6 | 5 | 1 | 4 | 5 | 1 | 22 | 62,86 |
| $H_{r_{med}}$ | 8 | 5 | 2 | 0 | 5 | 7 | 27 | 77,14 |
| $H_{r_{m\acute{a}x}}$ | 4 | 2 | 6 | 4 | 5 | 7 | 28 | 80,00 |
| ΔH_r | 8 | 5 | 0 | 2 | 4 | 7 | 26 | 74,29 |
| OcLI | 8 | 5 | 5 | 4 | 5 | 7 | 34 | 97,14 |
| TE | 5 | 5 | 3 | 4 | 5 | 7 | 29 | 82,86 |
| TEE | 7 | 3 | 6 | 1 | 3 | 7 | 27 | 77,14 |
| $r^2 4h$ | 8 | 5 | 6 | 4 | 5 | 5 | 33 | 94,29 |
| Lluvia | 8 | 1 | 6 | 1 | 3 | 7 | 26 | 74,29 |
| FFmed | 8 | 3 | 6 | 4 | 3 | 7 | 31 | 88,57 |
| FFmáx | 8 | 5 | 6 | 0 | 5 | 6 | 30 | 85,71 |
| DDmáx | 8 | 5 | 6 | 1 | 5 | 4 | 29 | 82,86 |

Como se puede apreciar en la tabla 3, todas las variables influyen en el consumo con una frecuencia de aparición que abarca desde 48,57 % hasta 100 %. Las variables de mayor peso son $T_{m\acute{a}x}$, $OcLI$, y la $r^2 4h$.

En tanto para las variables bioclimáticas el análisis muestra que la de mayor frecuencia de aparición es la TE , seguida por la TEE , la ΔH_r y la ΔT , en valores que van desde 65,71 % hasta 82,86 %.

El análisis en conjunto de las tablas 2 y 3 permite reconocer que en ambos métodos las variables a tener en cuenta para confeccionar un modelo de predicción por su mayor frecuencia de aparición son la $T_{m\acute{a}x}$, $H_{r_{m\acute{i}n}}$, $OcLI$, $T_{m\acute{i}n}$, $r^2 4h$, $H_{r_{m\acute{a}x}}$, $FF_{m\acute{a}x}$, ΔT y FF_{med} . Estas 9 variables conformarían el modelo resultante de variables predictoras, es interesante apreciar dentro del mismo la ΔT que es la única variable bioclimática incluida entre las de mayor frecuencia de aparición. Se aprecia también la presencia de la $FF_{m\acute{a}x}$ que sin ser bioclimática es una variable que implica un efecto sensorial.

A partir de las tablas 2 y 3 si se realiza el análisis integral de las variables bioclimáticas, las que se deben tener en cuenta para confeccionar un modelo de predicción son la ΔT seguida de TE , ΔHr y finalmente la TEE . En ese orden se pueden combinar con otras variables predictoras y también se pueden utilizar como un bloque único lo cual es un enfoque totalmente novedoso.

Teniendo en cuenta las variables predictoras obtenidas, se realizan varias combinaciones que aportan los siguientes escenarios tal y como se observa en la tabla 4.

Tabla 4. Combinaciones de predicción obtenidas

| Escenario | Variables incluidas |
|-----------|---|
| V2023A | r^24h , $FFmáx$, AT , $Hrmáx$, $Hrmín$ |
| VA1 | ΔT , ΔH , TEE , r^24h |
| VA3 | $OcLI$, TE , ΔT , $DDmáx$, $Hrmáx$ |
| VA4 | $Tmáx$, $Hrmín$, ΔT , $Ocll$, $Hrmáx$ |
| VA5 | ΔT , ΔH , TE , Ocl , r^24h |
| VB1 | $Tmáx$, $Hrmín$, $OcLI$, $Tmín$, $Hrmáx$, r^24h , ΔT , $FFmáx$, $FFmed$ |
| VB2 | ΔT , ΔHr , TE |

Una vez conformados los escenarios compuestos por las variables, tal y como se muestra en la tabla anterior, se procede a realizar el modelo de predicción para cada combinación propuesta anteriormente. Se utilizarán horizontes que incluyen los 12 meses de predicción en cada variante. La métrica de seguimiento y valoración de los resultados es el r^2 ajustada. Los resultados se pueden apreciar en la figura 5.

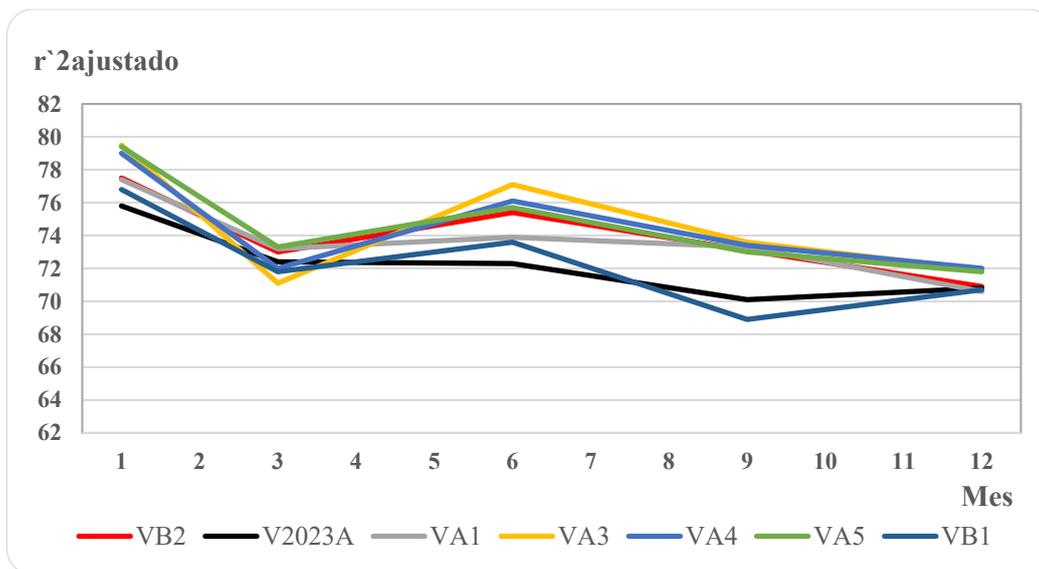


Figura 5. Comportamiento del r^2 ajustado para cada variante

De los resultados de la figura 5 se puede comentar lo siguiente, todas las variantes transitan en valores de un r^2 ajustada que van desde 68,9 % hasta 79,5 %. Para horizontes de 6 meses y 1 mes en prácticamente todas las variantes se obtienen mejores niveles de predicción. Se demuestra de forma práctica que la utilización de un modelo de predicción sin incluir el año permite obtener resultados satisfactorios en la predicción del consumo doméstico.

Es relevante reconocer que las variantes VA1 y VA5 conformadas por más de una variable bioclimática en conjunto con variables tradicionales, presentan altos valores de ajuste lo que permite

fundamentar la necesidad de la inclusión de este tipo de variables en los modelos de predicción para mejorar los niveles de ajuste y comprensión del consumo doméstico.

Sobresale entre los resultados de la figura 5 la variante VB2, la misma compuesta únicamente por 3 variables bioclimáticas presenta valores de ajustes similares al resto de las variantes, que están conformadas total o parcialmente por variables tradicionales. Los valores de ajuste permiten concluir que la sensación térmica tiene un impacto directo en el consumo doméstico de agua. Este resultado significa que solo con las variables bioclimáticas se puede explicar el comportamiento del consumo a partir de la variabilidad climática de forma satisfactoria, siendo este un resultado del cual no se ha encontrado similar en la bibliografía consultada por lo que es un resultado novedoso.

A partir de lo que se muestra en la figura 5 y para lograr un mejor entendimiento de lo que implica la utilización de las variables bioclimáticas se decidió incluir en la variante VB2 el escenario de análisis de la ecuación de regresión polinomial de primer orden para un horizonte de predicción de 6 meses el cual se muestra a continuación:

$$Lppm = -a * MES + b * AT - c * AmpHumedad + d * TE + e * LPPM(-1) + f * LPPM(-2) - g * LPPM(-3) - h * LPPM(-4) + i * LPPM(-5) - j * LPPM(-6) \tag{1}$$

Para esta ecuación se presentan en la tabla 5 los coeficientes asociados a cada variable.

Tabla 5. Coeficientes de la ecuación de regresión

| Coeficientes de la Ecuación | Coeficientes de la Ecuación | Coeficientes de la Ecuación |
|-----------------------------|-----------------------------|-----------------------------|
| a= 54,221 | e= 0,8581 | i= 0,2912 |
| b= 34,2303 | f= 0,1445 | j= 0,0909 |
| c= 27,7235 | g= 0,0178 | |
| d= 1,1722 | h= 0,3236 | |

De la ecuación de regresión (1) se puede concluir que todas las variables tienen una relación directa con el consumo, de acuerdo a los coeficientes, la que más peso tiene para el consumo es la ΔT seguida por la ΔHr con signo contrario, ambas son significativas en este esquema, terminando con la TE. De esta ecuación se puede interpretar que los cambios bruscos de temperatura y humedad relativa inciden de forma directa en la demanda

04 CONCLUSIONES Y RECOMENDACIONES

Los resultados de la investigación permiten afirmar que las variables bioclimáticas tienen una gran influencia en el consumo, proponiéndose un modelo de predicción de la demanda utilizando tan solo variables bioclimáticas que explican el 73 % del comportamiento del consumo, novedoso aporte, sin precedentes en las investigaciones consultadas en el estado del arte que sirvieron de base a la conformación del marco teórico.

Igualmente, los resultados obtenidos demuestran que la sensación térmica tiene un efecto directo en la predicción de la demanda doméstica de agua, con modelos con buen ajuste de r^2 ajustado como se demuestra en los escenarios que combinan bioclimáticas con otras variables tradicionales tales como el VA1, VA3, VA4 y VA5 respectivamente, pudiéndose realizar modelos satisfactorios de predicción de la demanda sin incluir el año como variable.

Por último, se pudo comprobar que la utilización de los valores históricos de consumo en el modelo de predicción, permite obtener escenarios con mejores ajustes del mismo.

RECONOCIMIENTO

La investigación que da origen a los resultados presentados en esta contribución recibió fondos del Instituto Nacional de Recursos Hidráulicos (INRH) de la República de Cuba.

Se agradece al destacado meteorólogo Carlos M González Ramírez, fallecido sin cuyo apoyo y experticia no hubiera sido posible obtener estos resultados.

05 REFERENCIAS

- Alshaihlí M., Aqeel S., Valdeolmillos N., Fathima F. and Choe P.** (2021). "A Multi-Linear Regression Model to Predict the Factors Affecting Water Consumption in Qatar." IOP Conf. Series Earth Environ. Sci 691 012004. Doi 10.1088/1755-1315/691/1/012004.
- Arellano A. y Peña D.** (2020). "Modelos de regresión lineal para predecir el consumo de agua potable". NOVASINERGIA, ISSN 2631-2654, 3(1), 27-36. <https://doi.org/10.37135/ns.01.05.03>.
- De Souza G., Azevedo G., Marcelo C. and Marcelo L.** (2019) "Predicting water demand: a review of the methods employed and future possibilities". IWA Publishing Water Supply | 19.8.
- Hao W., Cominola A., Castelletti A.** (2022). "Comparing predictive machine learning Models for Short- and Long-Term Urban Water Demand Forecasting in Milan, Italy." IFAC On Line: 55-33 (2022) 92-98.
- Pomares J.** (2022) "Principales factores climáticos que afectan la demanda doméstica". Tesis de diploma, Instituto Superior Politécnico José Antonio Echeverría, La Habana. Cuba.
- Ramos J.M., León A.J., Socorro R., Carvajal R. and González C. M.** (2023) "Climatic factors that impact the consumption patterns tame of water. A case of study". Progress in Artificial Intelligence and Pattern Recognition. 8th International Congress on Artificial Intelligence and Pattern Recognition, pages 35-47. IWAIPR 2023, Varadero, Cuba, September 27–29, 2023, Proceedings. <https://link.springer.com/book/10.1007/978-3-031-49552-6-4>

CONFLICTO DE INTERESES

Los autores declaran que no existen conflictos de intereses.

CONTRIBUCIÓN DE LOS AUTORES

- Mario Ramos Joseph** <https://orcid.org/0000-0002-2784-206X>
Realizó contribuciones en conceptualización, metodología, validación, análisis formal, escritura y discusión de los resultados
- Reniel Carvajal Alfonso** <https://orcid.org/0000-0002-8656-4709>
Realizó contribuciones en software, validación, investigación y discusión de los resultados
- Raisa Socorro Llanes** <https://orcid.org/0000-0002-2627-1912>
Realizó contribuciones en el diseño de la investigación, corrección del documento y análisis y discusión de los resultados.
- Alcides J. León Méndez** <https://orcid.org/0000-0002-5072-6438>
Realizó contribuciones en el diseño de la investigación, corrección del documento y análisis y discusión de los resultados.