

EVALUACIÓN DE LA DISPONIBILIDAD DE ÁGUA EN LA CUENCA VENTO APLICANDO MINERÍA DE DATOS

EVALUATION OF THE WATER AVAILABILITY IN THE VENTO BASIN APPLYING DATA MINING

Odalys MÉNDEZ VALDÉS^{1,2}, Raisa Socorro LLANES³, Haydee Llanusa RUÍZ³, Daniel Enrique Valdes SALOMÓN³

¹Aguas de La Habana. 4J29+46P, Fomento y Recreo, La Habana, Cuba

²Universidad Tecnológica de La Habana. 3H3M+XJ6, La Habana, Cuba. E-mail: omendez@ahabana.co.cu

³Universidad Tecnológica de La Habana José A. Echeverría (Cujae). Calle 114 entre Ciclovía y Rotonda. Marianao. La Habana. Cuba. E-mails: raisa@ceis.cujae.edu.cu; hllanusa@tesla.cujae.edu.cu; danielvs@tesla.cujae.edu.cu

Introducción
Materiales y métodos
Fase I (Comprensión de los objetivos de proyecto)
Fase II (Comprensión de los datos)
Fase III (Preparación de los datos)
Fase IV (Modelado)
Fase V (Evaluación del modelo)
Fase VI (Implementación)
Resultados y discusión
Conclusiones
Recomendaciones
Referencias

Resumo - A disponibilidade de água é cada vez menos condicionada por fatores climáticos, ambientais e econômicos, enquanto a demanda cresce conforme o desenvolvimento e a elevação do padrão de vida. A gestão integral da água é a chave para um equilíbrio adequado. Havana não escapa desse problema. Abastecida fundamentalmente por águas subterrâneas, está exposta a variações sazonais de abastecimento, com um déficit que ultrapassa os 1.000 l/s, afetando direta e indiretamente o abastecimento a mais de 250.000 habitantes. O estudo propõe um modelo para classificar a disponibilidade no aquífero para abastecimento urbano, aplicando ferramentas de mineração de dados aos registros em um período de cinco anos de 2016 a 2020, com frequência diária, cujas principais variáveis são: precipitação, níveis do aquífero, extrações e volume de descarga, entre outros. O modelo de classificação obtido tem uma taxa de sucesso de 98%, constituindo uma ferramenta de avaliação da disponibilidade de água disponível para especialistas e tomadores de decisão, servindo de base para identificar a ocorrência de seca operacional e estruturar um plano de redução de riscos de desabastecimento urbano na região. em caso de escassez e/ou seca.

Palavras-chave: Abastecimento de água. Águas subterrâneas. Aquífero. Mineração de dados. Previsão. Seca operacional.

ABSTRACT - The water availability is decreasing mainly due to climate, environmental and economic factors, while the demand increases with the development and life standard improvement. The water integral management is key to an adequate balance. Havana isn't outside this problematic. Supplied mainly by groundwater, it is exposed to seasonal variations in water supply, with a deficit that has exceed 1,000 l/s, directly and indirectly affecting the supply to more than 250,000 inhabitants. The study proposes a model to classify the water availability in the aquifer for urban supply, applying tools of data mining to the records taken with a daily frequency in a period o five years from 2016 to 2020, which main variables are: precipitation, aquifer levels, withdrawals, and discharge volume, among others. The classification model obtained has a success percent of 98%, constituting a tool for evaluating the resources available to specialists and decision makers, being the basis to identify the occurrence of operational drought and to structure a plan to reduce the risk of urban deficiencies in the event of shortage and/or drought.

Keywords: Water supply. Groundwater. Aquifer. Data mining. Prediction. Operational drought.

INTRODUCCIÓN

El crecimiento de la población mundial, el desarrollo urbanístico y socioeconómico inciden de manera directa en el incremento de la demanda de agua, así como en el cambio en los modelos de consumo; incluso contingencias sanitarias de carácter mundial como la COVID 19, marcan cambios significativos en la demanda, por lo que el desarrollo sostenible dependerá cada vez más de la gestión adecuada y la adaptación a los cambios.

Actualmente, la disponibilidad y la calidad de

agua a nivel internacional están en riesgo, tanto por razones antrópicas vinculadas al uso y cuidado del recurso como por el impacto del cambio climático sobre el mismo. Tal es así que entre los Objetivos de Desarrollo Sostenible (ODS), el número 6, Agua Limpia y Saneamiento, propone “Garantizar la disponibilidad y la gestión sostenible del agua y el saneamiento para todos”. El Instituto Nacional de Recursos Hidráulicos (INRH) como parte de la estrategia para dar cumplimiento a lo previsto en la Agenda para el Desarrollo

Sostenible 2030 del Sistema de la Naciones Unidas, se propone optimizar el aprovechamiento integrado del agua en función de su uso racional y productivo, la protección de su calidad y del medio ambiente.

En Cuba, el potencial hídrico asciende a 38,1 km³, de los cuales 31,6 km³ (83 %) corresponden a las aguas superficiales y 6,5 km³ (17 %) a las aguas subterráneas. De este potencial solo son aprovechables 24 km³, siendo el 67 % aguas superficiales (Díaz, 2018).

Específicamente en la provincia de La Habana, el potencial hídrico es básicamente subterráneo (solamente el 3 % del agua suministrada es agua superficial) y el uso es fundamentalmente urbano.

Según los informes de explotación desde la creación de Aguas de La Habana, entre el 45 % y 48 % del volumen distribuido en la ciudad es aportado por la cuenca Vento, dependiendo del comportamiento climático. La variabilidad está marcada fundamentalmente por las aportaciones de los manantiales del mismo nombre, que son captados y conducidos a través “Canal de Albear” (Monumento Nacional), hasta el centro histórico de la ciudad. (Aguas de La Habana, 2021).

Una problemática actual en la gestión de operación de la actividad de acueducto en la capital es la previsión de la disponibilidad de agua en la cuenca Vento y su relación con la ocurrencia de eventos de sequía. El objetivo propuesto es identificar las variables principales en la caracterización/clasificación de la disponibilidad de agua en el acuífero, así como las condiciones que desencadenan la sequía operacional, evento complejo por las indefiniciones en su momento de inicio, duración y severidad.

Por una parte, el INRH realiza el monitoreo de los pozos de observación enclavados en la cuenca para evaluar la disponibilidad de agua en el acuífero, actualizando el Gráfico de Control de Balance de Agua Subterránea (GCBAS). En esta base de datos se trabaja con la media de los niveles medidos en todos los pozos de observación. El nivel de agua se clasifica como favorable, normal, desfavorable y crítico con tendencia al descenso, ascenso y estable, sin que esa clasificación refleje realmente la disponibilidad de agua para el abasto.

Por otro lado, Cutié & Lapinel (2013) plantean que el Instituto de Meteorología (INSMET) cuenta con un Sistema de Alerta Temprana para el diagnóstico de sequía, implementado desde el año 2000, lo que unido a acciones gubernamen-

tales contribuye progresivamente a su enfrentamiento. Esta es una herramienta sólida para evaluar sequía meteorológica, relacionada con el déficit de precipitación que es el principal desencadenante del resto de las sequías, sin embargo, no establece una relación inmediata con la escasez de agua en los abastecimientos urbanos desde fuentes subterráneas.

Méndez et al. (2018), reconoce que una carencia relacionada con la gestión de sequía es la inexistencia de un indicador desencadenante y/o de métodos de cálculo que permitan definir la situación de escasez o sequía para las diferentes captaciones, planteando que el enfrentamiento a la sequía operacional sigue siendo reactivo y no preventivo. Para transformar esa situación se hace necesario establecer indicadores de referencia que sirvan como base para establecer un plan de gestión explícito, debidamente estructurado, que constituya la guía de trabajo oportuna.

Por ejemplo, Cutié & Lapinel (2013), refieren que en el período del 1961-2010, La Habana presentó un déficit de lluvia severo en el período húmedo del año 2009 y un déficit extremo en los acumulados de lluvia en el año hidrológico 2004/05, reconociéndose como eventos de sequía meteorológica.

Tomando como referencia los años naturales 2004-2005 y 2009-2010, se presenta una valoración del impacto en el abastecimiento, expresado en habitantes afectados (Figura 1). El gráfico muestra que los efectos de la sequía son más severos al final del período húmedo (mayo-octubre) y en el año siguiente, lo que evidencia la existencia de un desfase entre la sequía meteorológica y la operacional. En el propio gráfico se aprecia un descenso en la afectación a la población en el año 2005, que no se relaciona con la lluvia (porque el año fue seco), sino con el efecto de medidas en la gestión de abastecimiento.

En el caso del período 2009-2010 se observan los efectos de la sequía, a pesar de que el 2010 no se presenta como un año seco, de ahí la importancia de realizar análisis de la sequía operacional, que es la que se produce cuando la oferta de agua es menor que la demanda, afectándose el abasto.

Otro aspecto de interés es que el índice de sequía meteorológica y la lluvia que se maneja en el INSMET se realiza por provincias y regiones; sin embargo, para el abasto es preciso el análisis por cuencas ya que la distribución de la lluvia no es homogénea. Por ejemplo, en el estudio del Grupo

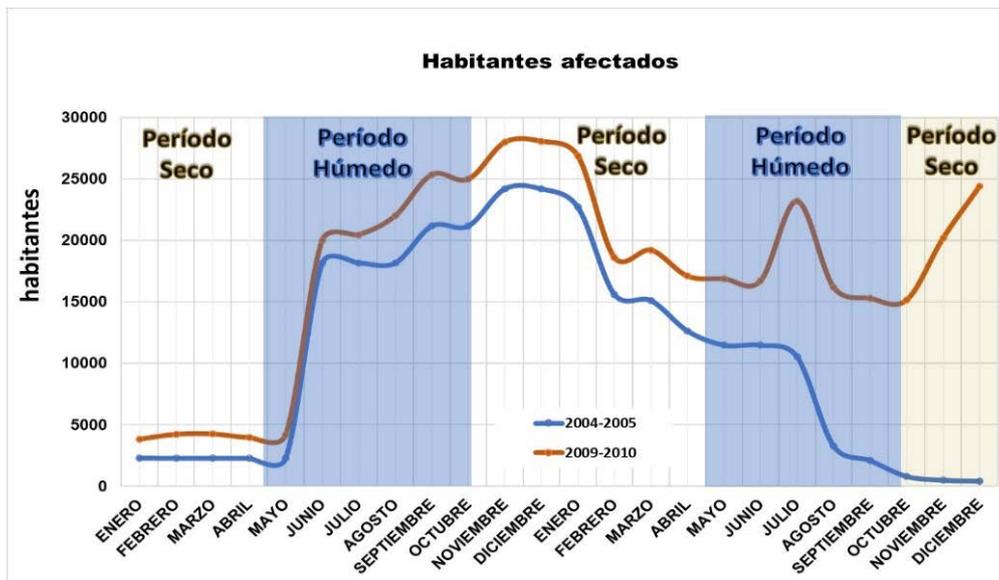


Figura 1 - Población con afectación en servicio de agua por redes 2004/2005 y 2009/2010.

de la sequía, del centro de clima del INSMET, Cutié & Lapinel (2013), muestran como en el período húmedo (mayo-octubre) desde el 1971 al 2000, los mayores acumulados de lluvia en la provincia La Habana se produjeron hacia el norte y centro; sin embargo, en el área que coincide con la zona de alimentación del acuífero, que es la lluvia aprovechable para este fin, los acumulados no alcanzaron los 100 mm mensuales.

Desde el año 2000 al 2020, se reconocen cuatro eventos que afectaron la gestión de abastecimiento con déficit que sobrepasan los 2.000 l/s para la operación, lo que unido al elevado volumen de pérdidas (estimadas en más del 50%), evidencian la existencia de sequía operacional.

Esta afectó a más del 20% de la población de La Habana, tanto por reducción del tiempo medio de servicio como por afectación total, según consta en informes y registros de la empresa Aguas de La Habana (Aguas de la Habana, 2021).

Los afectados por insatisfacción de la demanda, atribuido de manera genérica a la sequía en la cuenca Vento, han llegado a 30.000 habitantes afectados directamente, situación ante la cual, la empresa se ve obligada a tomar soluciones emergentes, como plantean Cutié & Lapinel (2013). Esta respuesta es lo que se puede definir como gestión de crisis, referida a la actuación en medio del evento, una vez que ya resulta impostergable tomar una acción, definiéndose en oportunidades medidas estructurales que no garantizan la inmediatez requerida, quedando operativas para futuros eventos de sequía.

Este tema es de interés tanto para investigadores como para gestores ya que se encuentra estrechamente relacionado con los planes de

Reducción de Riesgos implementados en el país, que abarcan procedimientos metodológicos para el Sistema de Alerta Temprana por Sequía (SATs), las Medidas de respuesta, así como Evaluación de impactos.

Los principales trabajos para evaluar la disponibilidad de agua subterránea en la cuenca dentro del país parten del modelo conceptual del acuífero, aplicando un modelo numérico para el cálculo del flujo subterráneo y los niveles del acuífero, que resuelve las ecuaciones de flujo por el Método de Diferencias Finitas (MDF) o Métodos de Elementos Finitos (MEF).

En Cuba, se destaca la aplicación del paquete de programas de AÑIMPE, desarrollado en el Centro de Investigaciones Hidráulicas para estudiar la compleja dinámica del acuífero, combinado con el Modelo de Administración de Acuíferos (MADA), en un algoritmo de programación lineal, vinculándose al Sistema de Información Geográfica (SIG) a través de la herramienta AQTRIGEO para la gestión de toda la base informativa de carácter espacial y de atributos, aplicándose a trabajos relacionados con el plan de uso del agua y la evaluación del comportamiento del acuífero bajo diferentes regímenes de explotación (Cabrera & Dilla, 2011).

Para evaluar el impacto de la operación sobre los manantiales de Vento aplicando AÑIMPE Cabrera (2014, según Hernández, 2022) modeló un sector del acuífero Vento y posteriormente Chirinos (2015, según Hernández, 2022) realizó modificaciones en el modelo conceptual, logrando este último una buena correspondencia entre los niveles observados y los simulados durante los años 2004-2005. Posteriormente, Duvergel (2019)

realiza la modelación del periodo del 2014-2016 obteniendo un error inferior al 5% en los niveles del acuífero, y una correlación entre caudales observados y simulados de 0.895, lo que se considera aceptable para este tipo de estudio. Sin embargo, se realizaron comparaciones puntuales en condiciones de escasez dentro de ese mismo período observándose que predominaba un error entre los caudales modelados y observados mayor de 20% lo que a juicio de la autora no satisface el enfoque a estudios de escasez y/o sequía que se requiere para este proyecto.

Otro modelo de apoyo a la toma de decisiones y manejo integral de los recursos hídricos, es el Sistema de Evaluación y Planificación de Agua (*WEAP* por sus siglas en inglés), el que también ha sido utilizado para estudios dentro de Cuba en diferentes sectores. Sánchez et al. (2018), desarrollaron un modelo aplicando *WEAP* a nivel de esquema para la predicción de la demanda y generación de escenarios en la cuenca Vento, evaluando la relación oferta-demanda ante cambios futuros en la disponibilidad de agua y/o en la operación y gestión del sistema de acueducto. Entre las principales limitaciones para la aplicación del software estuvieron: la dificultad para disponer de la respuesta hidrogeológica de la cuenca, la ausencia de estaciones de aforo en el río Almendares, las escasas mediciones de demandas y el desconocimiento de los patrones de consumo. Además, este software se comercializa y activa bajo licencia de tiempo limitado, por lo que se dificulta su libre uso, especialmente en Cuba, para dónde existen restricciones (Sánchez et al, 2018).

Aunque los modelos numéricos que resuelven las ecuaciones del flujo resultan de gran utilidad para la gestión de acuíferos, la existencia de gran cantidad de datos disponibles, vinculados a los parámetros de las cuencas permite aplicar la minería de datos para la exploración de variables, análisis, identificación de tendencias y relaciones no conocidas, obtener información relevante para la gestión y mejor comprensión de los fenómenos que se producen, lo cual constituye un instrumento valioso en el proceso de toma de decisiones, resultando factible su aplicación.

El progreso del 2008 a la actualidad en el uso de modelos de Minería de Datos (MD), Inteligencia Artificial (IA) y Aprendizaje Automático (*ML*, por sus siglas en inglés), aplicados en problemas relacionados con el agua subterránea, superficial y contaminación de acuíferos, es

reconocido por diferentes investigadores. Tao et al. (2022) realizaron una revisión del tema a través del análisis de 136 publicaciones en las que los principales países con aportaciones en este campo son Irán (36%), India (17%), China (9%) y Estados Unidos de América (9%).

Algunos problemas vinculados específicamente a la valoración del funcionamiento de los acuíferos han sido resueltos con la aplicación de la MD como son:

- Behzad et al. (2010) presentaron un estudio de predicción de niveles del acuífero de Towako en New Jersey, en relación con el bombeo y condiciones climáticas, aplicando máquina de vector soporte (*SVM* por sus siglas en inglés) conjunto de algoritmos de aprendizaje supervisado para resolver problemas de clasificación y regresión considerándolo como un métodos rápido y preciso en comparación con los modelos numéricos.
- En China, Huang et al. (2017) propusieron un modelo no lineal con *SVM*, aplicando optimización por enjambre de partículas (*PSO*, por sus siglas en inglés) conformando un modelo *PSO-SVM*, aplicado en tiempo real, para predecir los niveles de agua subterránea por filtración del embalse de las Tres Gargantas a partir de los datos de entrada y salida en el embalse, como parte de los estudios de deslizamientos de tierra en las zonas de Huayuan y Baijiabao.
- En Irán, en la cuenca ubicada al noroeste de la Llanura de Karaj et al. (2019) emplean el método de Algoritmo Genético de Regresión de Soporte Vectorial (*GA-SVR* por sus siglas en inglés), comparando los resultados con diferentes métodos de partición de los datos de entrenamiento y validación, siendo la opción aleatoria la de mejores resultados para la predicción de los niveles del acuífero a partir de la relación entre diferentes pozos de observación.
- En Cuba, en la cuenca Vento, Valcarce et. al. (2021) presentaron un estudio para evaluar la vulnerabilidad a la contaminación del agua subterránea. En el mismo se aplica la MD como un método alternativo a la evaluación de la vulnerabilidad, destacando como ventaja que se reducen el número de consideraciones que se realizan a *priori* para el procesamiento de parámetros. El método aplicado fue el agrupamiento, y según referencia de los propios autores se logra una mejor discriminación espacial que la obtenida por el método del RISK (roca, infiltración, suelo y karstificación) demostrando un mayor poder resolutivo.

Considerando entonces la tendencia internacional en la aplicación de distintos modelos de MD a la problemática del agua, y específicamente a la predicción de los niveles del acuífero, como herramienta que permite explorar en la búsqueda de patrones no visibles, realizar análisis multivariados y trabajar a varios niveles de abstracción a partir de la información disponible, en este trabajo se propone específicamente, evaluar la disponibilidad de agua en la cuenca Vento y obtener un modelo predictivo a partir de patrones y tendencias de los parámetros hidrometeorológicos,

como instrumento de apoyo a la gestión de abastecimiento de agua en la capital desde la relación disponibilidad-demanda.

En abril del 2020 se creó un Grupo Temporal de Trabajo (GTT) de la Cuenca Ariguanabo-Vento, rectorado por el INRH, que reúne a expertos que participan en la recopilación, análisis y diagnóstico de la situación de ambas cuencas, trazando un plan de acciones prioritarias para la conservación y gestión de los recursos, constituyendo este un marco propicio para la integración y validación de los resultados de esta investigación.

MATERIALES Y MÉTODOS

La cuenca subterránea Vento forma parte de la cuenca Almendares-Vento, reconocida por el Consejo Nacional de Cuencas Hidrográficas (CNCH) como cuenca de Interés Nacional. Se ubica en la región occidental de Cuba, con una extensión de 422 km² y sus recursos explotables están calculados en 287 millones de m³/año (Suárez et al., 2021). Esta cuenca es de gran importancia en el abastecimiento de agua potable a La Habana ya que beneficia a una población de 1.479.750 habitantes, según los registros de Operaciones de la empresa Aguas de La Habana.

La herramienta de MD utilizada fue desarrollada en la Universidad de Constanza en Alemania, conocido como *Konstanz Information Miner (KNIME)*, es un software para el análisis de datos, de código abierto y fácil uso para no programadores. Tiene como objetivo desarrollar procesos de descubrimiento de conocimiento en Base de Datos (BD) en un entorno visual. Este software admite varios métodos de procesamiento de datos, selección de columnas, suma de columnas, filtrado, matriz, normalización, desnormalización, entre otros procesamientos, constituyendo una herramienta novedosa, apropiadas para la gestión, que podrán utilizar los gestores y operadores, en la que se pueden integrar las distintas bases de datos (Pupo et al., 2018). El software también permite realizar análisis estadístico de la base de datos, lo que resulta muy útil para el proceso de preparación de los mismos y análisis de calidad dentro del sistema de trabajo.

El estudio se apoya en la Metodología para el Desarrollo de Proyectos de Minería de Datos (*CRISP-DM* por sus siglas en inglés) surgida en 1999 y aplicada en múltiples proyectos de MD, tanto empresariales como académicos. Según Espinosa (2020), esta proporciona una descripción normalizada del ciclo de vida de un proyecto

estándar de análisis de datos descrito en 6 fases.

Fase I (Comprensión de los objetivos del proyecto)

Se enfoca en la comprensión de los objetivos del proyecto, llevando a la definición de un problema de minería de datos y un plan preliminar para alcanzar los objetivos, restricciones y beneficios. En este caso, el objetivo específico es clasificar la disponibilidad de agua en la cuenca Vento y obtener un modelo para la predicción de los eventos de sequía operacional, con un enfoque hacia la mejora del servicio y a la reducción de riesgos de desabastecimiento, ahora empleando un método de Ciencia de Datos, cuyo resultado es aplicable en la gestión de la empresa Aguas de La Habana, en la empresa de Aprovechamiento Hidráulico de La Habana y en el INRH.

La hipótesis de trabajo plantea que la aplicación de técnicas de minería de datos permitirá identificar tendencia y patrones del acuífero Vento, en su interrelación con las distintas variables que intervienen en el balance de agua en la cuenca, constituyendo una herramienta para realizar pronósticos de disponibilidad de agua aplicable a la gestión de la sequía operacional en el caso de estudio. Ello contribuirá en la reducción del riesgo de desabastecimiento al que están expuestos los sistemas de abasto, determinado por la relación disponibilidad-demanda a nivel local.

Como resultado se obtendrá una propuesta para alertar sobre la disponibilidad de agua ante la variación climática y del régimen de explotación en condiciones normales y/o de sequía, para predicción a corto o medio plazo, y abre otras potencialidades para incorporar patrones y reglas que obedezcan a diferentes intereses. Por ejemplo, se puede vincular el resultado de la clasificación de la disponibilidad de agua con la

ocurrencia de sequía operacional aplicados a los planes de gestión.

Fase II (Comprensión de los datos)

Comienza con la identificación de los datos vinculados al problema de estudio: tipo, formato, fuente y significado. Igualmente, es necesario identificar problemas de calidad, establecer relaciones e hipótesis preliminares.

La BD para el estudio propuesto en la cuenca Vento se vienen recopilando y analizando desde hace varios años, obteniéndose de la red de monitoreo y observación del INRH y de la empresa Aguas de La Habana, seleccionándose el período de enero del 2016 a diciembre del 2020, conteniendo 1.826 registros diarios por cada uno de los parámetros que influyen en la disponibilidad de agua en el acuífero como son:

1. precipitación (14 estaciones pluviométricas),
2. temperatura (estación de Casa Blanca),
3. nivel piezométrico del acuífero (12 pozos de observación enclavados en la cuenca que cuentan con registros mensuales),
4. gasto extraído (42/58 de captaciones de agua subterránea para abastecimiento urbano),
5. descarga del acuífero a través de los manantiales de Vento (obtenido a través de los registros históricos de caudal del canal de Albear).

Los datos se almacenaron en dos archivos: uno contiene parámetros tales como nivel y caudal del canal del Albear, niveles estáticos de los pozos de observación, lámina de lluvia de los pluviómetros y la temperatura, mientras que el otro documento contiene los registros de los volúmenes extraídos por bombeo de las captaciones enclavadas en la cuenca. Se cuenta con un total de 70 columnas de 1.826 filas, las cuales se procesan y seleccionan convenientemente según los objetivos de cada procesamiento.

En esta investigación el parámetro “Precipitación” se trabaja con la Red Pluviométrica de la cuenca Vento (Figura 2), seleccionándose 9 de los 14 pluviómetros descartando los que tienen información incompleta. Los mayores acumulados de lluvia se registran de mayo a octubre, en el que ocurren el 70% de las precipitaciones, comportándose acorde con la marcada estacionalidad que presentan las precipitaciones en Cuba, acorde con el estudio de Cutié & Lapinel (2013). De los 1.826 días que comprende el estudio, en 1.271 no ocurrieron precipitaciones en la cuenca, y en 8 de 9 pluviómetros se registraron lluvias intensas por encima de los 100 mm/d, con un registro máximo de 136 mm/d en el pluviómetro de La Güinera, el 14 de octubre del 2017.

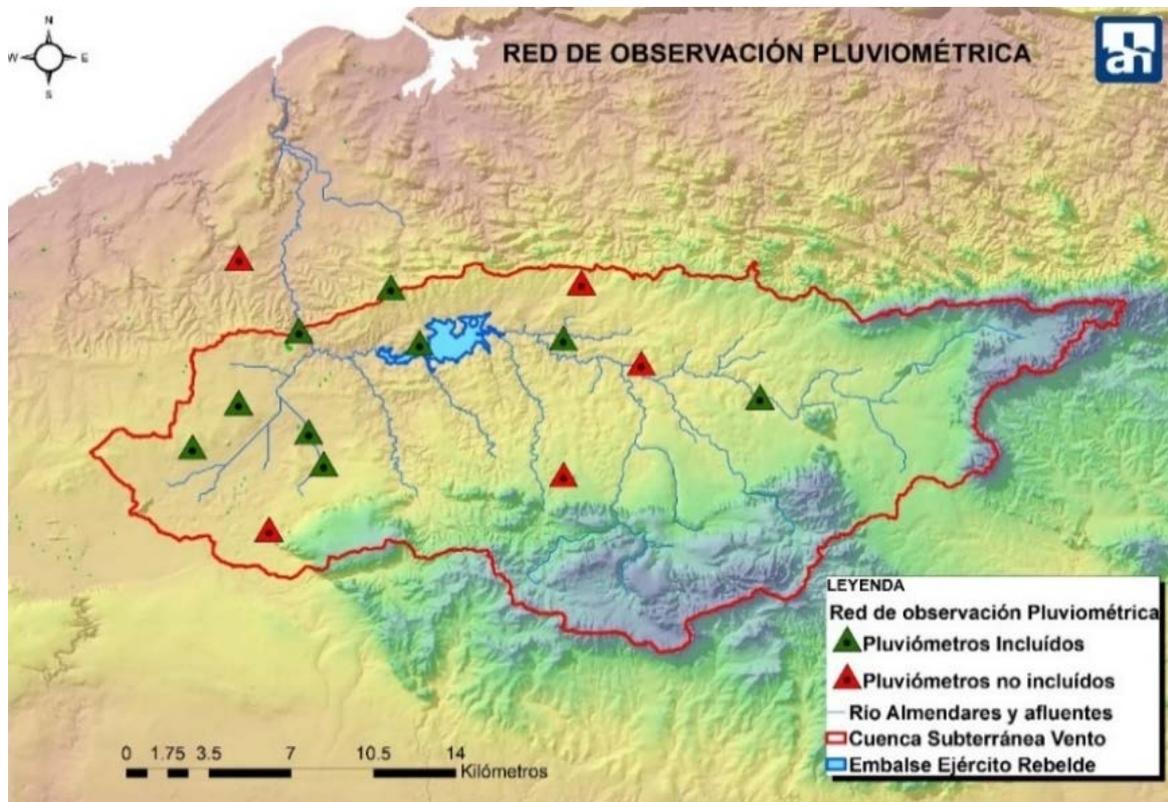


Figura 2 - Cuenca Vento y sus puntos de monitoreo. Red Pluviométrica.

La Red de Observación Hidrogeológica en la figura 3, seleccionándose para el modelo los cuenta con 12 pozos de observación mostrados que no tenían datos faltantes. Las mediciones

disponibles hasta el 2020 tienen frecuencia mensual, transformándose a niveles diarios a partir de la ecuación lineal de la recta formada entre el nivel a inicio de cada mes expresado en metros. La cota mínima del agua registrada fue

en el pozo de observación de Mazorra, con 36.21 msnm en abril del 2016, y para ese mismo pozo el mayor nivel de agua registrado fue de 39.97 msnm en octubre del propio año, con el menor diferencial entre nivel mínimo y máximo.

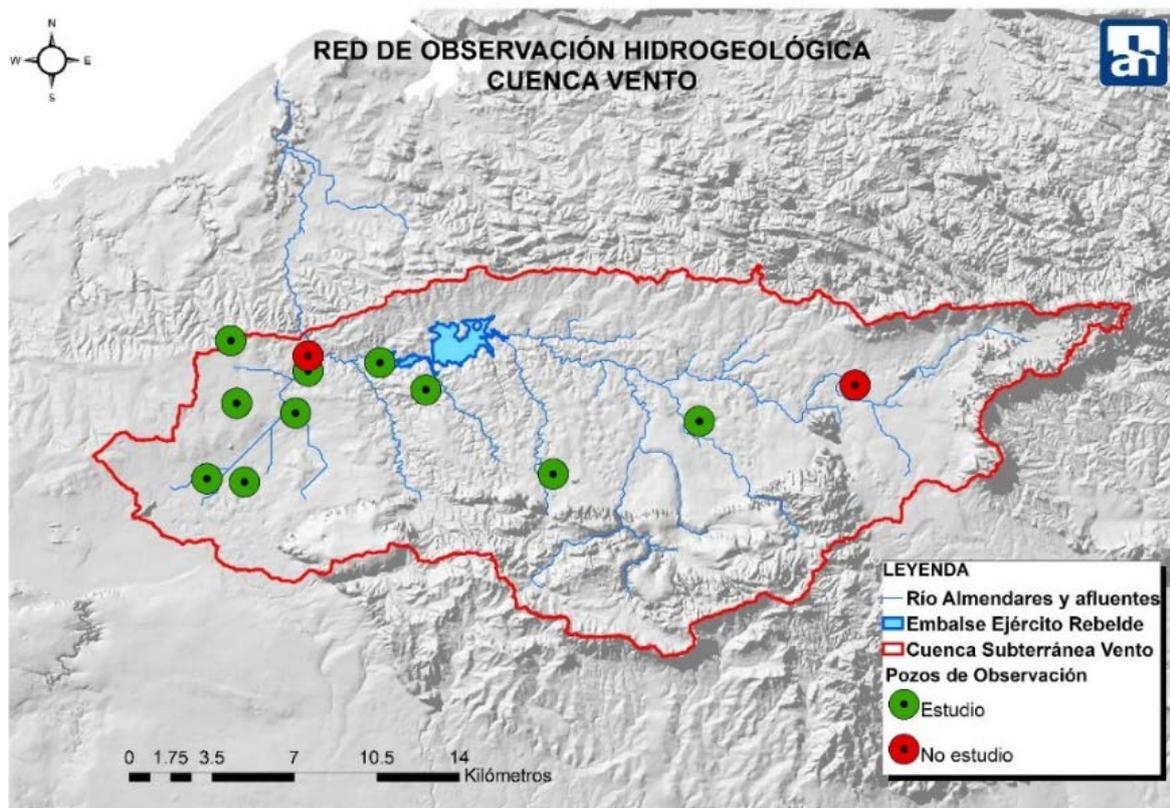


Figura 3 - Red de Observación Hidrogeológica valorada para el estudio en la cuenca Vento. Fuente: GIS Aguas de La Habana, valorada para el estudio.

La Red de Extracción (captaciones) está formada por las captaciones por bombeo enclavadas en la cuenca. Alrededor del 90 % del consumo corresponde al sector residencial y los objetivos socioeconómicos enclavados en la misma, contando actualmente con 58 pozos de explotación.

La figura 4 muestra los pozos de extracción seleccionados para llevar a cabo el análisis y extraer información relevante de su comportamiento, así como con los demás parámetros que intervienen en la cuenca Vento. La información referida a estos, es el volumen diario extraído en el período de enero del 2016 a diciembre del 2020 y se expresa en m³/día. Entre las principales captaciones se encuentran Cosculluela con 105.000 m³/día como mínimo, Paso Seco con 103.000 m³/día, el Rincón con 31.580 m³/día y Cuatro Caminos 31.620 m³/día y Aguada del Cura con grandes fluctuaciones estacionales registra máximos en el orden de los 146.700 m³/día.

La descarga del acuífero se cuantifica a partir de los niveles del canal de Albear (magistral obra, considerado como una de las maravillas de la

ingeniería en Cuba y monumento Nacional). Los datos de nivel son adquiridos por el sistema SCADA (Supervisión, Control y Adquisición de Datos), con registros diarios en la empresa Aguas de La Habana. Los valores están expresados en metros (m) y a través de la curva de Q vs H del canal de Albear se obtuvieron los caudales en litros por segundo (l/s) aportados por los manantiales para cada nivel conocido.

Las aportaciones de los manantiales son conducidas por el canal de Albear totalmente por gravedad, abasteciendo a la población del casco histórico de La Habana. El mínimo caudal registrado en esta etapa fue de 156 l/s en mayo del 2020 y el máximo de 1481 l/s en noviembre del 2016. La descarga de los manantiales en el período de estudio tiene una tendencia descendente (Figura 5), con un ciclo de 4 años, observándose una reducción de las reservas. Ello refleja la vulnerabilidad a la que están expuestos los manantiales, corriendo el riesgo hasta de desaparecer, con el consiguiente daño para el abasto a la ciudad, el patrimonio y la economía

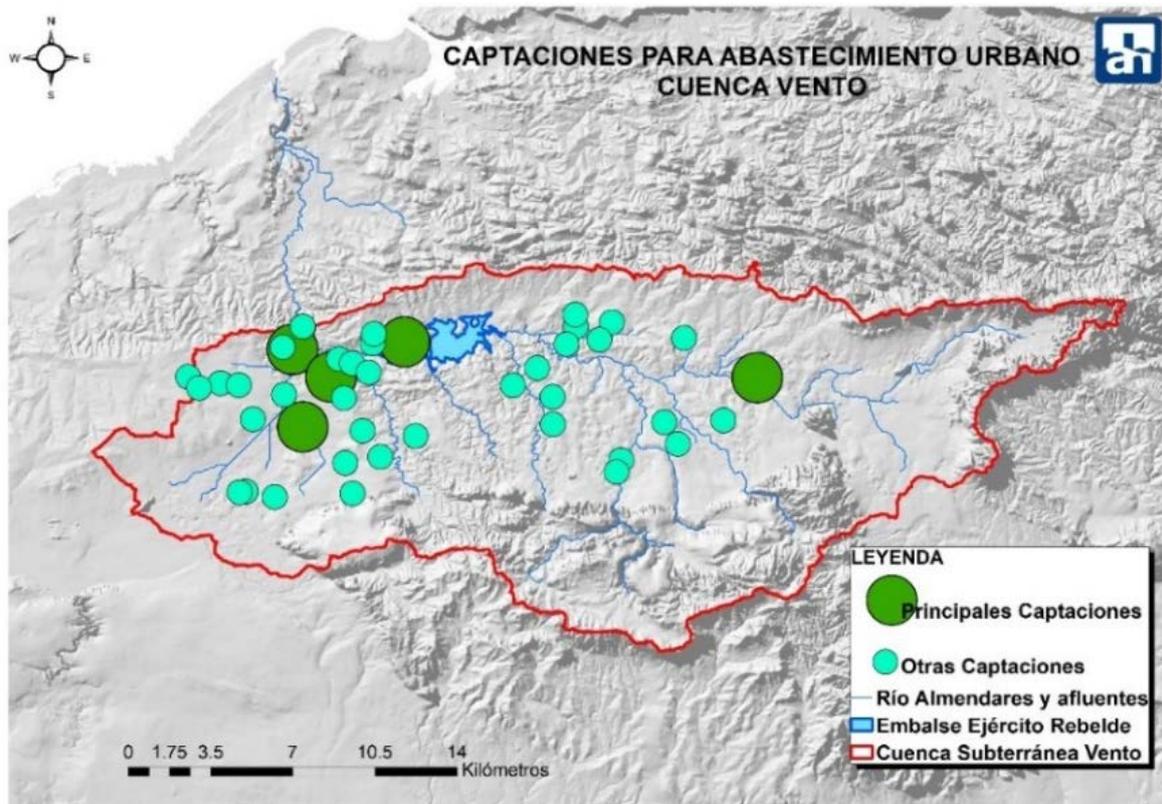


Figura 4 - Captaciones para abastecimiento urbano. Cuenca Vento. Fuente: GIS Aguas de La Habana.

La descarga del acuífero se cuantifica a partir de los niveles del canal de Albear (magistral obra, considerado como una de las maravillas de la ingeniería en Cuba y monumento Nacional). Los datos de nivel son adquiridos por el sistema SCADA (Supervisión, Control y Adquisición de Datos), con registros diarios en la empresa Aguas de La Habana. Los valores están expresados en metros (m) y a través de la curva de Q vs H del canal de Albear se obtuvieron los caudales en litros por segundo (l/s) aportados por los manantiales para cada nivel conocido.

Las aportaciones de los manantiales son

conducidas por el canal de Albear totalmente por gravedad, abasteciendo a la población del casco histórico de La Habana. El mínimo caudal registrado en esta etapa fue de 156 l/s en mayo del 2020 y el máximo de 1481 l/s en noviembre del 2016. La descarga de los manantiales en el período de estudio tiene una tendencia descendente (Figura 5), con un ciclo de 4 años, observándose una reducción de las reservas. Ello refleja la vulnerabilidad a la que están expuestos los manantiales, corriendo el riesgo hasta de desaparecer, con el consiguiente daño para el abasto a la ciudad, el patrimonio y la economía.



Figura 5 - Descarga de los manantiales de Vento (l/s).

Otro aspecto de interés es que los valores mínimos de descarga de los manantiales se alcanzan a finales de junio y los máximos se alcanzan a finales de noviembre, lo que evidencia un desfase del tiempo de recarga del acuífero.

Fase III (Preparación de los Datos)

Se realizan un conjunto de procesamientos seleccionando cuáles y cómo serán utilizados en la modelación.

En esta etapa se incluye la limpieza de datos, selección de tablas, registros y atributos. Otro procesamiento realizado fue la determinación de grupos de pozos de observación de agua subterránea, empleando para ello el método de Correlación Lineal (*Linear Correlation*), que calcula para cada par de columnas seleccionadas un coeficiente de correlación, es decir, una medida de la variación conjunta de dos variables. El valor de esta medida varía de -1 (correlación negativa fuerte) a 1 (correlación positiva fuerte), un valor de “cero” no representa una correlación lineal (aunque las variables aún pueden depender mucho unas de otras).

Mediante el análisis de correlación del nivel piezométrico entre los pozos de observación, en el período de estudio, se obtiene la matriz de correlación que sirve de base clasificar los pozos de observación en 4 grupos, lo que tributa a una mejor comprensión del funcionamiento del acuífero y las

interrelaciones que se dan en el mismo. Los grupos quedan establecidos como sigue:

- Grupo 1: HV-19 Cabañas, HV-29A Figueredo, HV- 645 Aeropuerto, HV-39 Mazorra y Wajay;
- Grupo 2: AL-3 Paso Seco, AL-4 El Globo, AL-5 Santa Amelia;
- Grupo 3: HV- 39 Mazorra Limnógrafo, AL-6 Guayabal;
- Grupo 4: HV-649 Vento y el AL-7 Managua.

Los valores de correlación que muestran los pozos del Grupo 3 indican un comportamiento contrario a los niveles del resto de los pozos (correlación negativa). Ante este resultado, la primera acción fue revisar cuidadosamente la validez de la información, no detectándose anomalías en las mismas. Para una mejor comprensión de ese comportamiento se requiere realizar estudios locales que permitan identificar qué elemento del acuífero y/o del entorno podría condicionar ese comportamiento. Durante este estudio los mismos serán excluidos de los análisis. Los pozos de Grupo 4 presentan una correlación lineal débil con el resto de los pozos.

Otro análisis estadístico de interés que afianza los resultados obtenidos y la relación entre las variables es la correlación de cada pozo de observación de agua subterránea con la descarga del acuífero (Tabla 1).

Tabla 1 - Correlación Descarga de Manantiales vs Pozos de observación.

Nivel Pozos de Observación	Correlación entre las variables “Descarga de Manantiales” y “Pozos de Observación”	Grupo según la correlación entre pozos de observación
Pozo AL-3 Paso Seco	0.9	G-2
Pozo HV-19 Cabañas	0.88	G-1
Pozo HV-29A Figueredo	0.88	G-1
Pozo HV-645 Aeropuerto	0.87	G-1
Pozo HV-648 Mazorra	0.87	G-1
Pozo AL-4 El Globo	0.85	G-2
Pozo AL-5 Santa Amelia	0.63	G-2
Pozo HV- 28 Wajay	0.58	G-1
Pozo AL-7 Managua	0.36	G-4
Pozo HV-649 Vento	0.22	G-4
Pozo HV-39 Mazorra Lim	-0.46	G-3
Pozo AL-6 Guayabal	-0.56	G-3

De la tabla 1, se comprueba la correlación de los grupos con la descarga del acuífero:

- Los pozos del Grupo 1 tienen una correlación superior al 0,85 excepto Wajay.
- Los pozos del Grupo 2 presentan una correlación entre 0,50 y 0,85, excepto AL-3 Paso Seco que es el que mejor correlaciona de todos.
- Los pozos que mejor correlacionan con la

descarga del acuífero son AL-3, HV-19 y HV-29A.

- El Grupo 3 presentan una correlación negativa con la descarga del acuífero.
- El grupo 4 mantienen una correlación lineal débil con la descarga, menor que 0,50.

Actualmente, en la práctica diaria para la gestión de abastecimiento en la empresa Aguas de La Habana, se consideraba el caudal de descarga

a través de los manantiales de Vento la variable más representativa del estado del acuífero, y ahora el análisis estadístico ampara este criterio, ya que el mismo correlaciona por encima de 0.85 con el 50% de las estaciones piezométricas.

Igualmente, teniendo en cuenta los resultados de la correlación se excluirán de los análisis posteriores de los pozos de observación de Managua y el Guayabal, que presentan una correlación lineal débil tanto con la descarga del acuífero como con el resto de los pozos. Igualmente serán excluidos los que presentan una correlación negativa.

Otro elemento a considerar para el estudio de predicción es la estimación del tiempo de recarga natural del acuífero a partir de la relación del hidrograma de descarga de los manantiales con la precipitación. Para su obtención, López & Sanz (1997) plantean considerar los datos de aportaciones de los manantiales y de precipitación como series temporales, verificándose el tiempo para el cual se obtiene el mayor coeficiente de correlación e interpretándose este como el desfase temporal máximo entre ambas variables.

Para determinar el desfase en el caso de estudio se realizaron varios experimentos desplazando la serie de lluvia con respecto a la descarga de los manantiales, en intervalos de 5 días aplicando un nodo Correlación Lineal (*Linear Correlation*). La figura 6 muestra el tiempo en el que se alcanza la mayor correlación (100 a 145 días), considerándose que el tiempo de recarga natural del acuífero se refleja en los manantiales a los 100 días. Este resultado se puede incorporar a los análisis para la operación del sistema de abastecimiento, que hasta ahora por la experiencia práctica en la empresa Aguas de La Habana se consideraba igual o mayor a 60 días en dependencia de las características del período húmedo. El tiempo de recarga natural del acuífero estimado tiene implícito el comportamiento de otras variables como la temperatura, las extracciones, la geometría y geología de la cuenca, incluso la repercusión del embalse Ejército Rebelde en la recarga del acuífero. Se recomienda continuar los estudios precedentes de la influencia de este último aspecto en la curva de recarga del acuífero.

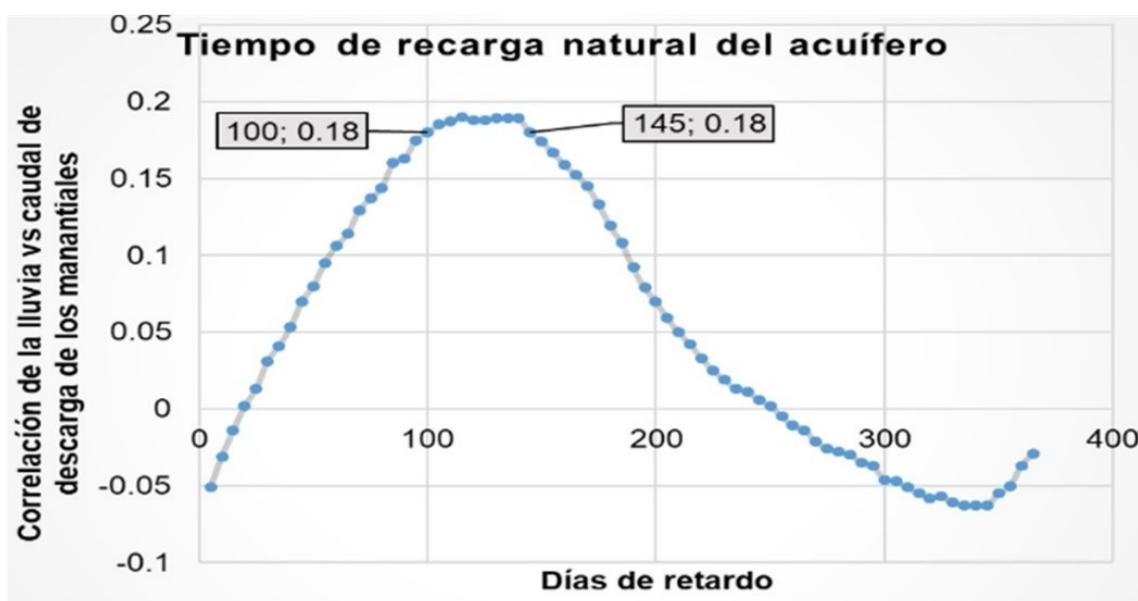


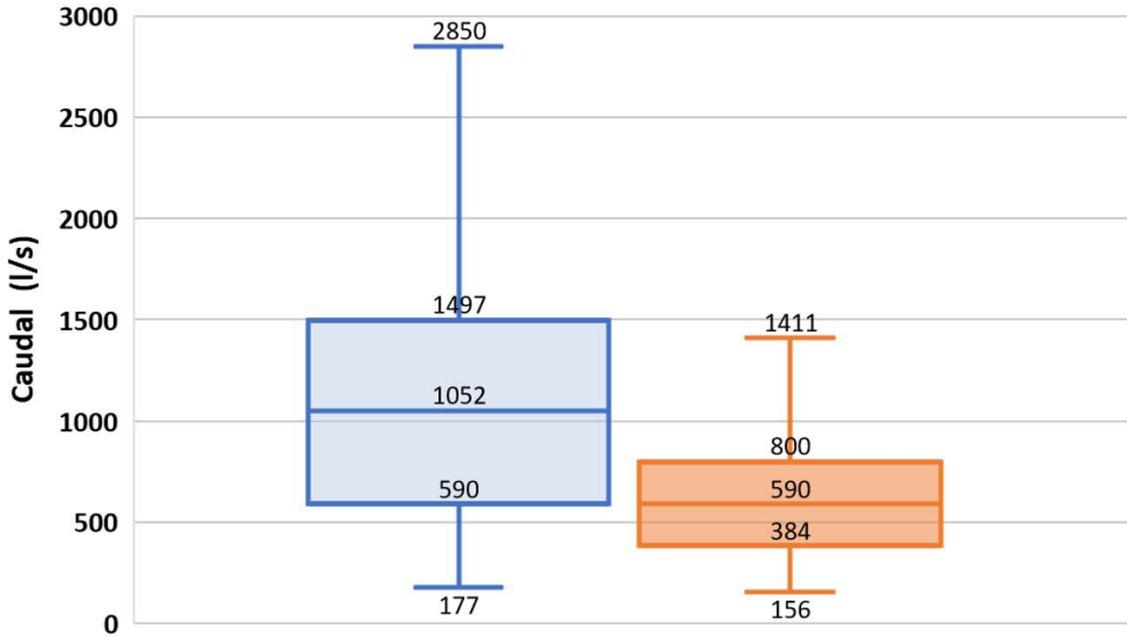
Figura 6 - Estimación del tiempo de recarga natural del acuífero.

En la preparación de los datos también se establecieron los límites del modelo. Para ello, se partió del análisis previo del diagrama de Caja y bigotes, obtenido para dos series: una serie promedio mensual de la descarga de los manantiales del 2000 al 2020 con 360 registros y la serie promedio diaria de los caudales 2016 al 2020, con 1.826 registros.

El análisis de las 5 medidas principales resultantes en cada una de las series permite establecer

la comparación del comportamiento a través del tiempo y comparar la tendencia. En la figura 7, se evidencia que, en los últimos 5 años (imagen derecha) hay un descenso notable en el comportamiento de la descarga del acuífero, con una reducción prácticamente del 50 % para cada uno de los cuartiles (25, 50 y 75 %). Igualmente se aprecia que el promedio mínimo mensual fue de 177 l/s, con un mínimo diario de 156 l/s, que coincide en los últimos 5 años de estudio.

EVOLUCION DEL CAUDAL DE DESCARGA



■ Caudal promedio mensual (2000-2020) ■ Caudal promedio diario (2016-2020)

Figura 7 - Comparación de la evolución de caudales en el tiempo.

El comportamiento anterior, referido a la tendencia de la descarga del acuífero sirvió de base para establecer los intervalos (clases) a aplicar en el modelo de clasificación, referidos a la disponibilidad de agua, proponiéndose los siguientes criterios:

- Clase 1 - Caudal bajo: Menor o igual que 590 l/s.
- Clase 2 - Caudal medio: Entre 590 l/s y 800 l/s.
- Clase 3 - Caudal óptimo: Más de 800 l/s.

Fase IV (Modelado)

En la misma se definen las técnicas a aplicar seleccionando aquellas que se ajusten al problema de estudio. Mientras mayor número de análisis y

técnicas se realicen, mejor conocimiento y comprensión se tendrá del problema. Como resultado se han de obtener modelos calibrados con parámetros óptimos. En ocasiones esta etapa requiere regresar a la fase de preparación de datos para adecuarlos a las exigencias de los modelos seleccionados.

El flujo de trabajo *KNIME*, que se observa en la figura 8, se inicia con la lectura de los datos y su manipulación realizándose análisis estadísticos, luego se realiza exploración visual para el apoyo a la comprensión de los datos hasta llegar a la aplicación del modelo de trabajo y la evaluación de los resultados obtenidos.

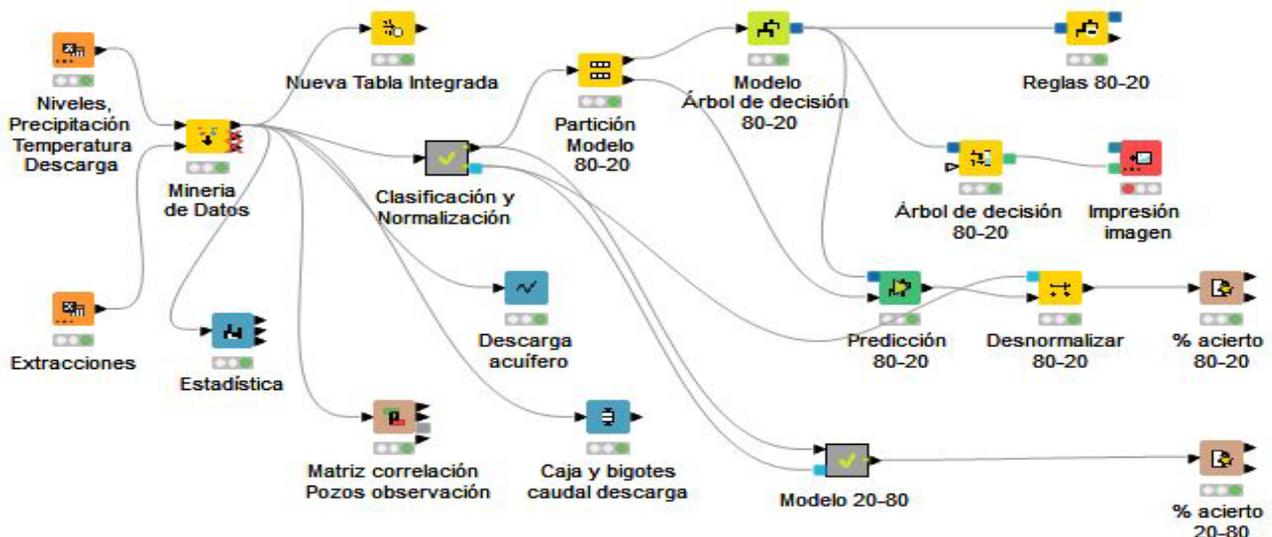


Figura 8 - Flujo de trabajo. Modelo Árbol de decisión realizado en *KNIME*.

En el caso de estudio se propone utilizar la serie diaria, aplicando modelos predictivos, los que permiten realizar inferencias a partir de los datos, con el fin de realizar proyecciones a futuro. El modelo de aprendizaje automático aplicado fue el árbol de decisión con un algoritmo de clasificación y regresión para la predicción de la descarga del acuífero. El árbol de decisión trabaja con dos nodos fundamentales, árbol de decisión para entrenamiento (*Decision Tree Learner*) y un árbol de decisión de predicción (*Decision Tree Predictor*). En el proceso de entrenamiento se obtienen las reglas que luego son aplicadas para la predicción. En el estudio se realizan dos variantes del mismo modelo de árbol de decisión, cambiando la relación de partición, en la Variante

1 (80% de los datos para entrenamiento y 20% para validar), y en la Variante 2 con el 20% de la serie para aprender y el 80 % para validar.

La variable que se empleó para establecer el criterio de clasificación es el caudal de descarga de los manantiales, expresado en litros por segundo (Q (l/s)). La figura 9 muestra los intervalos de clasificación definidos en orden ascendente y los límites de los intervalos, los que fueron definidos a partir del análisis del gráfico de caja y bigotes, siendo consistentes con los objetivos del estudio.

Una vez clasificado el caudal, el flujo de trabajo pasa por el nodo Administrador de Color (*Color Manager*) que permite asignar colores a esta columna (la cual es nominal) para una mejor distinción visual de la información (Figura 10).

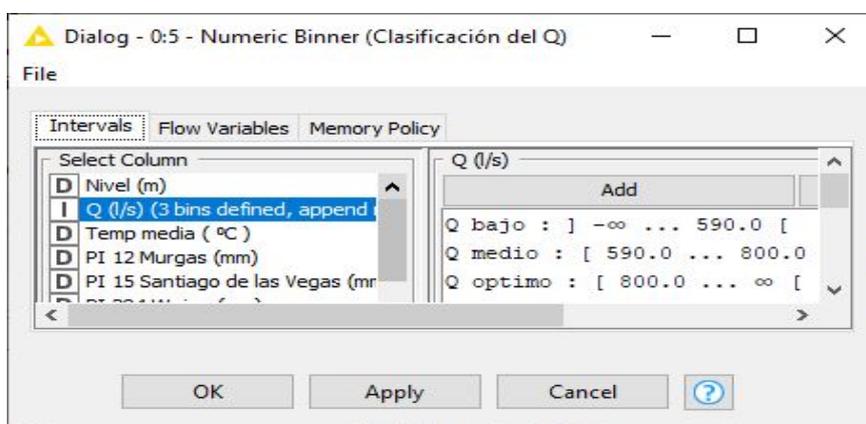


Figura 9 - Clasificación del caudal de descarga, nodo Clasificador Numérico (*Numeric Binner*) realizado en KNIME.

Row ID	Días	Nivel (m)	Q (l/s)
Row582_Row...	2017-08-05	0.81	590
Row583_Row...	2017-08-06	0.8	570
Row584_Row...	2017-08-07	0.8	570
Row585_Row...	2017-08-08	0.79	557
Row586_Row...	2017-08-09	0.79	557
Row587_Row...	2017-08-10	0.81	590

Figura 10 - Distinción visual de la clasificación aplicando nodo Administrador de Color (*Color Manager*) realizado en KNIME.

A continuación, el nodo Filtro de Columnas (*Column Filter*) es empleado eliminando la columna numérica del caudal, que fue sustituida por una columna nominal, simplificando el aprendizaje y la comprensión de la realización del modelo, y la de nivel porque el caudal es una variable dependiente de este, y provocaría un sobreajuste. Los campos que entrarán en el proceso se transforman mediante el nodo Normalizador (*Normalizer*) encargado de normalizar los datos

en los rangos de 0 a 1 para homogenizarlos, evitando que el modelo sobrevalore los campos, o sea el modelo trabaja con los valores normalizados de todas las variables.

Llegado este punto el nodo Particionador (*Partitioning*) permite dividir la minería de datos ya transformada en dos particiones, una para entrenar y otra para validar. Por la parte superior, el modelo se enlaza con el nodo Aprendiz del Árbol de Decisiones (*Decision Tree Learner*) y

por la salida inferior con el nodo Predictor del Árbol de Decisiones (*Decision Tree Predictor*) al cual le entra el flujo del nodo Aprendiz del Árbol de Decisiones. El nodo de aprendizaje (*Decision Tree Learner*) posibilita la inducción de un árbol de decisiones de clasificación en la memoria principal. El atributo de destino debe ser nominal y los otros atributos utilizados para la toma de decisiones pueden ser nominales o numéricos. Las divisiones numéricas son siempre binarias (dos resultados), dividiendo el dominio en dos particiones en un punto de división dado.

El nodo Predictor del Árbol de Decisiones (*Decision Tree Predictor*), por su parte, utiliza un árbol de decisión existente (que se pasa a través del puerto del modelo) para predecir el valor de clase de los nuevos patrones. Dicho modelo provee una representación visual en forma de árbol, que contiene la dependencia del comportamiento de una variable con otra y a su vez una clasificación del caudal de descarga según el criterio de partición, en el mismo el gasto desfavorable se representa en color rojo, el medio en amarillo y uno óptimo en verde, el árbol se va desplegando hacia abajo, con diferentes niveles de decisión, en diferentes ramas, en función de las reglas del modelo de clasificación.

Fase V (Evaluación del modelo)

Se analizan los resultados de los modelos construidos en base al análisis de ciertas métricas. Esta etapa exige una evaluación a fondo, revisar los pasos seguidos, comparar los resultados del modelo con el problema a resolver.

De este análisis se decide la aplicación de los resultados o regresar a una etapa anterior e incluso desechar el proyecto.

El nodo Calificador (*Scorer*) permite validar el modelo comparando la columna de predicción con la original obteniendo una matriz de confusión con el número de coincidencias en cada celda y una serie de estadísticas de precisión, como verdaderos positivos, falsos positivos, verdaderos negativos, falsos negativos, recuperación, precisión, sensibilidad, especificidad, así como la precisión general y el coeficiente kappa de Cohen que se aplica a modelos de clasificación, siendo la medida estadística que ajusta el efecto del azar, mientras más cerca de 1 esté, mayor será la concordancia del modelo con la observación más allá del azar.

La figura 11 muestra la matriz de confusión obtenida en el modelo 80-20, en el que se observa que fueron clasificados correctamente 359 caudales de descarga (170 Q bajo+95 Q medio+94 Q óptimo), con 5 errores de clasificación (3+3+1), lo cual resulta un alto nivel de coincidencia.

En la tabla 2 se observa que la precisión en ambos modelos es elevada, la variante 1 (80-20) con 98,087 % y 96,922 % para la variante 2 (20-80), igualmente el coeficiente kappa de Cohen, también es alto, con un 0,97 y 0,96 para cada modelo.

Los resultados obtenidos, con una precisión por encima del 90%, pudieran hacer pensar en un modelo sobre ajustado, de ahí que sea necesario realizar un análisis detallado de la matriz de confusión.

Row ID	Q bajo	Q medio	Q optimo
Q bajo	170	3	0
Q medio	3	95	0
Q optimo	0	1	94

Figura 11 - Matriz de confusión Variante 1 (80-20) obtenida con KNIME.

Tabla 2 - Comparación de la métrica a dos variantes de clasificación.

Métrica	Clasificación Variante 1 (80-20)	Clasificación Variante 2 (20-80)
Clasificación correcta	359	1417
Clasificación errónea	7	45
Precisión (%)	98.087	96.922
Error (%)	1.913	3.078
Coeficiente Kappa de Cohen (%)	0.97	0.952
Modelo	Estratificado	Estratificado

Esta muestra el rendimiento de un modelo de clasificación, es decir, cuántos eventos positivos y negativos se predicen correcta o incorrectamente en cada clase, asociado a la métrica de *recall* (exhaustividad) y de sensibilidad del modelo equipo. La exhaustividad indica cuán bueno es el modelo para detectar eventos positivos. En este caso específico, se observa que estas métricas coinciden (exhaustividad y sensibilidad) con valores por encima de 0.9 para cada clase, por lo que el modelo puede considerarse apropiado para clasificar en cada una de las clases, no considerándose sobre ajustado.

Fase VI (Implementación)

Comprende la etapa de puesta en marcha del proyecto, una vez definido el mismo. Se requiere organizar el conocimiento adquirido y prepararlo para ponerlo a disposición del o los destinatarios finales.

RESULTADOS Y DISCUSIÓN

En el estudio se presentan los resultados de la primera etapa de investigación, en que se desarrolla un modelo predictivo de Clasificación, donde los atributos inmersos pueden tomar valores entre un grupo de clases ya conocidas, el mismo se aplicó para evaluar la disponibilidad de agua, considerando tres rangos dentro de la serie de descarga del acuífero.

La selección del conjunto de datos para entrenamiento y prueba se realizó en dos variantes de proporción de datos entrenamiento-validación, la variante 1 (80-20) y la variante 2 (20-80), con la muestra estratificada, obteniéndose en ambos modelos de clasificación una elevada precisión para la predicción del caudal de descarga del acuífero (98 y 97 % respectivamente).

Se obtiene un modelo de predicción de la descarga del acuífero mediante herramientas de clasificación, empleando para este modelo el árbol de decisión (*Decision Tree*). Las variables empleadas fueron precipitación, cota del nivel estático de los pozos de observación, temperatura y volumen de extracción de las captaciones.

En el proceso de clasificación se crearon reglas en las que se ampara el modelo de clasificación, para la Variante 1 se generan 22 Reglas y para la Variante 2 se generan 14 Reglas. Al analizarse las condicionantes que se imponen en las reglas con los valores normalizados, se obtiene información útil de las componentes principales que determina el comportamiento del acuífero. Ejemplo de las reglas obtenidas en el árbol de decisión:

El resultado a presentar al cliente puede ir desde la generación de un informe de manera periódica, hasta la realización de análisis sistémicos de los datos y actualización del modelo.

Una vez definida la validez del modelo para el estudio se propone su implementación, cuyo resultado se vincula a la gestión de sequía operacional en la actividad de acueducto, aplicado en la cuenca Vento. La clasificación obtenida se asocia a los umbrales de sequía operacional, proponiéndose tres fases, que coinciden con el modelo de clasificación:

- Clase 1 - Sequía Operacional: Disponibilidad baja ($Q < 590$ l/s).
- Clase 2 - Escasez: Disponibilidad media ($590 \text{ l/s} \leq Q \leq 800$ l/s).
- Clase 3 - Normalidad: Disponibilidad Óptimo ($Q > 800$ l/s).

En la Variante 1 (80-20), la Regla 1 (con parámetros normalizados) para Clase Caudal bajo plantea: Si la cota del Nivel Estático del pozo de observación AL-4 “El Globo” es menor o igual a 0,283 y el caudal en la captación “Paso Seco” es menor que 0,615 y la cota del Nivel Estático del pozo de observación AL-5 “Santa Amelia” es menor o igual que 0,728 y la cota del Nivel Estático del pozo de observación HV-29 A “Figueredo” es menor o igual que 0,556 y la cota del Nivel Estático del pozo de observación Al-3 Paso Seco es menor o igual que 0,545.

En la Variante 2 (20-80), la Regla 1 (con parámetros normalizados), para Caudal Bajo plantea; si la lluvia del pluviómetro 341 de “Cuatro Caminos” es menor o igual que 0,662 y la cota del Nivel Estático del pozo de observación AL-5 “Santa Amelia” es menor o igual que 0,718 y la cota del Nivel Estático del pozo de observación HV-29A “Figueredo” es menor o igual que 0,558 y la cota del Nivel Estático del pozo de observación Al-3 “Paso Seco” es menor o igual que 0,547.

Como se aprecia, en las reglas observadas en las dos variantes, la clasificación es función fundamentalmente del comportamiento de los pozos de observación. Las relaciones más significativas se dan vinculadas al pozo de observación AL-3 “Paso Seco”, que se había clasificado en el grupo 2 de los pozos de observación y tiene una correlación con la descarga de los manantiales del 0,90 y el pozo HV-29A “Figueredo”, clasificado dentro del

Grupo 1 de los pozos de observación, teniendo ambos parámetros una alta presencia en la clasificación en ambos modelos.

Se corrobora a partir de las reglas de clasificación que el caudal de descarga está determinado por el nivel del acuífero, identificándose los pozos de observación AL-3 Paso Seco y el pozo HV 29 A Figueredo los de mayor incidencia para la clasificación en cualquiera de los rangos (bajo, medio y alto), lo que resulta una información útil para priorizar el acondicionamiento de las estaciones de medición y la adquisición de datos en el acuífero.

Específicamente, de acuerdo con el resultado de las reglas expresado en niveles reales, cuando la cota del acuífero está por debajo de 40,858 msnm en el pozo de observación AL-3 “Paso Seco” y de 42,126 msnm en el pozo HV-29A “Figueredo”, se está en una condición de baja

disponibilidad de agua, lo que resulta un aporte ya que en la actualidad la clasificación del estado del acuífero se hace a partir de los valores medios de los niveles observados.

Otra contribución es la obtención del tiempo de recarga natural del acuífero, observándose un desfase de la descarga de los manantiales con respecto a la lluvia entre 100 y 145 días, aspecto de interés para el estudio de la sequía operacional.

Los resultados de clasificación de disponibilidad de agua obtenidos se emplean para definir los umbrales de sequía operacional, a partir de relacionar la condición de baja disponibilidad (caudal de descarga de los manantiales por debajo de 590 l/s) con la fase de sequía operacional, la disponibilidad media con la condición de escasez y la normalidad con disponibilidad por encima de 800 l/s lo que se considera óptimo a los efectos operacionales.

CONCLUSIONES

Con el empleo de la minería de datos al caso de estudio se demuestra la validez de su aplicación para la gestión integral del agua, lo que permitió a partir de la base de datos disponible descubrir relaciones entre las variables y determinar las componentes principales que actúan en el sistema. El estudio del modelo predictivo de clasificación de la disponibilidad de agua en la cuenca Vento se aplica por vez primera y los resultados obtenidos son muy alentadores, demostrando en la práctica las posibilidades que brinda la minería de datos y la herramienta *KNIME* para la realización de varios modelos del acuífero.

Los resultados obtenidos aportan información sobre el comportamiento de la descarga de los manantiales en su interacción con los principales parámetros y tiene aplicación práctica inmediata en la gestión integral de la cuenca en su

interrelación con el abastecimiento de agua.

En el estudio se establece una interrelación disponibilidad-sequía operacional, definiendo los umbrales para un pronóstico diferenciando en la gestión de abastecimiento de agua en la cuenca Vento, dando la posibilidad de trazar un plan de acción oportuno que minimice el impacto de la sequía operacional y mantener el mayor número de usuarios abastecidos por la red de suministro.

A partir de los resultados obtenidos, se puede elaborar una metodología aplicable a otras cuencas y sectores hidráulicos, más si se tiene en cuenta, que el INRH cuenta con un Sistema Integral de Gestión de las Aguas (SGIA), lo que constituye una amplia base de datos con información de los principales parámetros hidro meteorológicos en todo el país, condición necesaria para implementar un modelo basado en la minería de datos.

RECOMENDACIONES

El empleo de la herramienta *KNIME* y los resultados obtenidos, así como otras prestaciones se pueden vincular como un módulo en la Plataforma Modular Integral para la gestión de

Recursos Hídricos (PMI-RH), desarrollada para la gestión del agua subterránea por Gómez (2008), lo que permitirá validar resultados y complementar la toma de decisiones.

REFERENCIAS

AGUAS DE LA HABANA. **Informe anual de la dirección de Acueducto**. La Habana. Cuba, 2021.
BEHZAD, M.; ASGHARI, K.; COPPOLA E.A. JR. Comparative Study of SVMs and ANNs in Aquifer Water Level prediction. **Journal of Computing in Civil Engineering**, p. 408-413, 2010. Disponible en: [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000043](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000043) Fecha acceso: 22feb2022.

CABRERA, E. “**Modelación matemática del acuífero de Vento**”, Trabajo de Diploma. Instituto Superior Politécnico “José A. Echeverría”, La Habana, Cuba, 2014.
CABRERA, E.E y DILLA S.F. Modelo de administración de acuíferos: MADA. **Tecnología y ciencias del agua**, v. 2, n.4 Jiutepec oct./dic. 2011. versión On-line ISSN 2007-2422. Disponible en: <https://www.redalyc.org/pdf/3535/35353>

- 1975001.pdf Fecha de acceso: 15mar2022
- CHIRINOS D. “**Modelación matemática de un sector del acuífero Vento**”. Trabajo de Diploma. Instituto Superior Politécnico “José A. Echeverría”, La Habana, Cuba, 2015.
- CUTIÉ, V. & LAPINEL, B. Capítulo IV. La Habana. En “**La sequía en Cuba, un texto de referencia**”. (p. 120-125). ISBN: 978-959-300-053-6. Sello editorial AMA. 2013.
- DÍAZ, J.A. El agua en Cuba: un desafío a la sostenibilidad. **RIHA**, v. 39, n. 2, p. 46–59, 2018. Disponible en: <https://riha.cujae.edu.cu/index.php/riha/article/view/447> Fecha acceso: 20dic2020
- DUVERGEL P.M.L. **Evaluación de un sector de la cuenca Vento con la herramienta META-Aquíimpe**. Trabajo de diploma. Cujae. La Habana. 2019.
- ESPINOSA, J.J. Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. **Ingeniería, investigación y tecnología**, v. 21, n. 1, 2020. <https://doi.org/10.22201/fi.25940732e.2020.21n1.008> Fecha acceso: 21oct2021
- GÓMEZ, M. **Plataforma Modular Integrada para la gestión del agua subterránea como recurso**. Tesis en opción de grado de Doctor. Instituto Superior Politécnico “José A. Echeverría”, La Habana. Cuba, 2008.
- HERNÁNDEZ, A.O. El aeropuerto José Martí y peligros asociados a su sistema de protección contra inundaciones. **Ingeniería Hidráulica Y Ambiental**, v. 43, n. 2, p. 71–83, 2022. <https://riha.cujae.edu.cu/index.php/riha/article/view/587> Fecha acceso: 31may2022
- HUANG, F.; HUANG, J.; JIANG, S. H.; ZHOU, C. Prediction of groundwater levels using evidence of chaos and support vector machine. **Journal of Hydroinformatics**, v. 19, n. 4, p. 586–606, 2017. <https://iwaponline.com/jh/article-abstract/19/4/586/3639> Fecha acceso: 14abr2022.
- KNIME. **Herramienta de analítica de datos**. Licencia GNU-GPL. V3. 2020. <https://www.KNIME.com>
- LÓPEZ, J.J. & SANZ, E. Cálculo de la recarga natural de un acuífero a partir del hidrograma del manantial. Aplicación al manantial de Vozmediano. **Revista Estudios Geol.**, v. 53, p. 25-32. España. 1997. Disponible en: <https://doi.org/10.3989/egol.97531-2244> Fecha acceso: 10dic2022.
- MÉNDEZ, O.; RIVERA, E.; LLANUSA, H.; HERNÁNDEZ, A. Enfrentamiento a la sequía operacional en la empresa Aguas de La Habana. **RIHA**, v. 39, n. 2, p. 112-123, 2018. Disponible en: <https://riha.cujae.edu.cu/index.php/riha/article/view/448> Fecha acceso: 10ene2019
- PUPO, K.G.; MENA, O.L.; RODRÍGUEZ, D.M. PAR-KNIME: conjunto de plugins para extraer reglas de asociación cuantitativas en KNIME. **Revista Cubana de Ingeniería**, v. 9, n. 2, p. 18-28. 2018. Disponible en: <https://rci.cujae.edu.cu/index.php/rci/article/view/629> Fecha acceso: 20dic2021
- SÁNCHEZ, Y.; MÉNDEZ, O.; HERNÁNDEZ, A. Evaluación de la Cuenca subterránea Vento con el modelo Water Evaluation and Planning system (WEAP). **Revista Hidrolatinoamericana de Jóvenes Investigadores y Profesionales**, v. 2, p. 29-32, 2018. Disponible en: <http://www.revistahidrolatinoamericana.com> Fecha acceso: 15mar2019
- SUÁREZ, O.; VALCARCE, R.M.; VEGA, M.; RODRÍGUEZ, W.R. Riesgo de contaminación de las aguas subterráneas en la cuenca Almendares–Vento, Cuba. **RIHA**, v. 42, n. 3, p. 95–109, 2021. Disponible en: <https://riha.cujae.edu.cu/index.php/riha/article/view/555> Fecha acceso: 8abr2022
- TAO, H.; MAJEED, M.; ABDULAMEER, H. Groundwater level prediction using machine learning models: A comprehensive review. **Revista Neorocomputing**, v. 489, p. 271-378, 2022. Disponible en: <http://www.elsevier.com/locate/neucom> Fecha acceso: 20may2022
- VALCARCE, R.M.; SUÁREZ G.O.; RODRÍGUEZ M., W.; VEGA, C.M. Aplicación de la minería de datos a la evaluación de la vulnerabilidad de acuíferos. **Revista Cubana de Ciencias Informáticas**, v 15, n. 2, p. 1-23, ab-jn 2021. Disponible en: <http://rcci.uci.cu> Fecha acceso: 10 mar2022.

Submetido em 18 de agosto de 2022

Aceito para publicação em 22 de fevereiro de 2023