



Sexto
Congreso Nacional de
Riego, Drenaje y Biosistemas
COMEII- 2021 / Hermosillo, Sonora



Artículo: COMEII-21017

Hermosillo, Son., del 9 al 11 de junio de 2021

DETECCIÓN DE ACUÍFEROS SOBREEXPLOTADOS MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO NO SUPERVISADO

Alberto González Sánchez^{1*}; Miguel Antonio Vega Castro²; Ronald Ernesto Ontiveros Capurata¹

¹Coordinación de Riego y Drenaje. Instituto Mexicano de Tecnología del Agua. Paseo Cuauhnáhuac 8532, Progreso, Jiutepec, Morelos, C.P. 62550. México.

²Estudiante de Posgrado. Maestría en Tecnologías de la Información. Universidad Politécnica del Estado de Morelos. C.P. 62550. México.

alberto_gonzalez@tlaloc.imta.mx – (777)3293600 ext. 115 (*Autor de correspondencia)

Resumen

Los acuíferos son sumamente importantes al aportar el 36.1% del volumen total anual destinado para la agricultura en México. Desde la primera publicación de la disponibilidad media anual de los acuíferos se ha observado un incremento de la cantidad de acuíferos en déficit, con un volumen extraído que supera la recarga. Esto representa un problema de desabastecimiento y mala distribución del recurso hídrico, por lo que se hace necesario el desarrollo de herramientas que ayuden a detectar los acuíferos propensos a caer en déficit en un corto plazo, aun cuando dicho riesgo no sea perceptible. Los modelos físicos pueden ayudar a este problema con simulaciones, pero requieren una gran cantidad de información, siendo costosos en tiempo y recursos. Las técnicas de aprendizaje no supervisado representan una alternativa viable en este tipo de problemas al identificar patrones derivados de las características de la información analizada, capturar tendencias y comportamientos de variables sin el conocimiento profundo requerido por los modelos físicos, basándose únicamente en métricas de similitud. En este contexto, el presente trabajo presenta el algoritmo *nAcuifDef*, el cual hace uso de la técnica de *clustering* difuso (*Fuzzy C-Means*) para detectar acuíferos propensos a caer en déficit. Las pruebas realizadas muestran que, para un número suficiente de grupos utilizados, el algoritmo puede inferir más algoritmos propensos al déficit que el ordenamiento simple basado en porcentaje de disponibilidad, siendo un ejemplo de que esta técnica permite resolver problemas de naturaleza compleja que serían difíciles de predecir.

Palabras claves: Machine learning, Fuzzy c-Means, déficit de agua subterránea

Introducción

En México, los acuíferos aportan 36.41% del volumen total de agua anual utilizado en la agricultura y 39.1% del volumen total concesionado para usos consuntivos (CONAGUA, 2018a). A partir del año 2001, la Comisión Nacional del Agua (CONAGUA) ha realizado estudios específicos por acuífero para determinar cantidad de agua disponible para su extracción a partir de la recarga y otras variables siguiendo la norma NOM-011-CONAGUA-2000 (CONAGUA, 2000). En base a estos estudios, la CONAGUA publica periódicamente la disponibilidad media anual de los acuíferos desde el año 2009, actualizando la cantidad de volumen disponible para su extracción (CONAGUA, 2009, 2010a, 2010b, 2010c, 2011a, 2011c, 2011b, 2013, 2015, 2018b, 2020). Este volumen debe ser considerado como base para el otorgamiento o limitación de concesiones, evitando la sobreexplotación de los acuíferos. No obstante, se observa un incremento en el número de acuíferos con déficit, donde el volumen concesionado supera la recarga (Cuadro 1), lo cual está asociado al incremento del uso del agua subterránea para usos consuntivos que se ha intensificado en los últimos años (Figura 1).

Cuadro 1. Cantidad de acuíferos en déficit acorde a publicaciones oficiales.

Año de publicación	Cantidad de acuíferos	
	En déficit	Con disponibilidad
2009-2011 (en parcialidades)	174	479
2013	193	460
2015	203	450
2018	245	408
2020	275	378

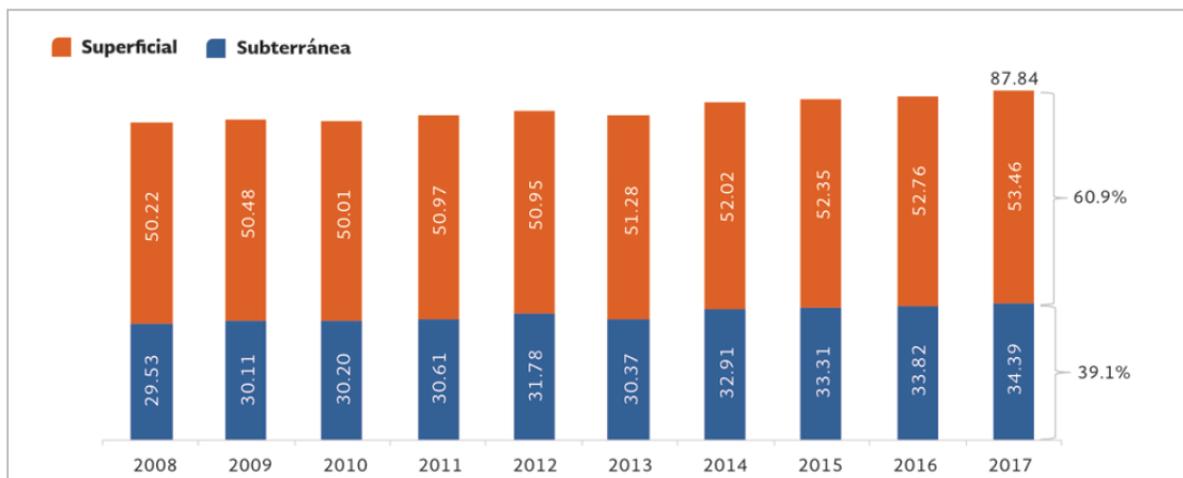


Figura 1. Volumen concesionado para usos consuntivos por tipo de fuente 2008-2017 (miles de hm^3) (CONAGUA, 2018a)

El aprovechamiento sustentable de los acuíferos requiere del análisis de una gran cantidad de variables involucradas, esto con el fin de contar con información para realizar un balance hídrico preciso. Sin embargo, los acuíferos presentan una naturaleza



dinámica difícil de modelar, ya que los sistemas de aguas subterráneas continuamente responden a los cambios en el uso de la tierra, el clima (incluida la temperatura y la precipitación), la recarga y las extracciones de agua subterránea (Wang et al., 2018). La recarga en particular resulta muy difícil de pronosticar, ya que no puede ser medida directamente (Crosbie, Davies, Harrington, & Lamontagne, 2015; Gao, Connor, & Dillon, 2014). Conocer con anticipación el estado en el que pueden caer los acuíferos a futuro es sumamente importante dentro del marco de dicha gestión. Una forma de detectar los acuíferos que pueden caer en déficit a corto plazo es por medio de los modelos basados en procesos físicos para simular los sistemas de agua subterránea (Kanyama, Ajoodha, Seyler, Makondo, & Tutu, 2020). Estos modelos son costosos por la gran demanda de información, que en muchas ocasiones requieren medición de variables directamente en campo para su calibración, lo que también demanda mucho tiempo (Coulibaly, Anctil, Aravena, & Bobée, 2001). Una alternativa se encuentra en el uso de modelos basados en los datos, que se construyen a partir de algoritmos de aprendizaje automático, para lo cual se requiere contar con suficiente información histórica. El aprendizaje automático puede identificar patrones directamente desde la información, capturar tendencias y comportamientos de variables sin el conocimiento profundo de los atributos físicos subyacentes de los modelos de flujo de agua subterránea (Steyn, 2018). Este tipo de técnicas se ha empleado cada vez con mayor frecuencia para abordar problemas de comportamiento complejo o impredecible, y pueden ser de gran valor para detectar los acuíferos próximos a caer en déficit.

Diversas técnicas de aprendizaje automático se han utilizado para resolver problemas relacionados con el agua subterránea (Solomatine & Ostfeld, 2008). Algunos ejemplos son las redes neuronales artificiales (*Artificial Neural Networks, ANN*) (Daliakopoulos, Coulibaly, & Tsanis, 2005), los bosques aleatorios (Random Forest, RF) y la regresión con máquinas de soporte vectorial (*Support Vector Regression, SVR*) (Kanyama et al., 2020). Estas técnicas pertenecen al tipo denominado aprendizaje supervisado, y requieren un conjunto de datos previamente clasificados para la generación de los modelos (Han & Kamber, 2006). Esta dependencia dificulta su aplicación generalizada para inferir los acuíferos que puedan caer en déficit en un corto plazo. Por el contrario, los métodos no supervisados no requieren un conjunto de entrenamiento que contenga información a priori de las etiquetas de clase de los objetos como entrada, y pueden inferir relaciones desconocidas a partir de las características de los objetos (Lopez, Tucker, Salameh, & Tucker, 2018). De entre este último tipo de técnicas, el algoritmo de formación de conglomerados *k-Means* (MacQueen, 1967) es uno de los más utilizados; no obstante, presenta algunas deficiencias al considerar que los grupos son exclusivos, asumiendo que no existe traslape entre ellos. Técnicas más recientes, como el algoritmo *cMeans* para *clustering* difuso, o *Fuzzy cMeans Clustering* (FCM) (Bezdek, Ehrlich, & Full, 1984), pueden lidiar mejor con este problema, generando grupos por grados de pertenencia. El presente trabajo propone que los grupos de acuíferos que se formen y su relación con el volumen disponible pueden proporcionar información útil para detectar los acuíferos que en un futuro cercano pueden caer en déficit, aún cuando dicha posibilidad no sea evidente. En este contexto, este artículo presenta el algoritmo denominado *nAcuifDef* que utiliza la técnica de aprendizaje no supervisado para conglomerados *Fuzzy cMeans* para detectar los acuíferos que en un futuro cercano puedan caer en un estado de déficit. La estimación realizada por *nAcuifDef* es comparada contra un pronóstico simple basado en



el ordenamiento ascendente de la disponibilidad media anual, mostrando que puede proporcionar un mayor acercamiento a la cantidad real de acuíferos que enfrentarán un agotamiento en corto plazo y representando un ejemplo de la forma que las técnicas de aprendizaje se aplican para lidiar con problemas de naturaleza compleja o impredecible.

Materiales y Métodos

Datos utilizados

El primer paso para la aplicación de la técnica de aprendizaje fue la conformación de una base de datos con las principales características de los acuíferos, siendo seleccionada la información que se muestra en el Cuadro 2. Este tipo de datos se han usado con anterioridad en estudios relacionados con la disponibilidad de agua en los acuíferos.

Cuadro 2. Información consultada para la conformación de la base de datos de entrenamiento

Tipo de información	Datos de interés	Fuente de la información	Período de análisis o fecha de publicación	Formato
Climatología	Temperaturas (mínima, máxima y promedio anual en °C), precipitación y evapotranspiración potencial anual (mm)	Global Climate Monitor (GCM) (Research Climate Group, s/f)	Datos anuales del 2005 al 2009	Shapefile
Uso del suelo y vegetación (serie IV)	Delimitación nacional de uso del suelo agrupada por agricultura, asentamientos humanos, bosque, cuerpos de agua, selva, vegetación y otros. Superficie en metros cuadrados (m ²)	Instituto Nacional de Estadística, Geografía e Informática (INEGI) a través de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) (CONABIO, 2018)	Publicado en el 2009	Shapefile
Permeabilidad	Tipo de permeabilidad de la superficie (tomado de la capa de hidrogeología, puede ser baja, baja a media, media a alta o alta)	CONABIO (CONABIO, 2018)	Estudio realizado en 1990	Shapefile
Concesiones de uso de agua subterránea	Volumen concesionado de agua subterránea (en miles de m ³)	CONAGUA (CONAGUA, 2017)	Actualizada a diciembre de 2019	CSV
Disponibilidad media anual de agua en los acuíferos años 2009-2011 (1ra act.), 2013 (2da act.), 2015 (3ra act.), 2018 (4ta act.) y 2020 (5ta act.)	Recarga total (R), volumen disponible de agua subterránea (DAS) o DÉFICIT (en hectómetros cúbicos). 5 series con todos los acuíferos	Publicaciones de CONAGUA en el Diario Oficial de la Federación (SEGOB, s/f)	Publicaciones en distintas fechas: 28/08/2009 08/07/2010, 16/08/2010, 25/01/2011, 14/12/2011, 20/12/2013, 20/04/2015, 04/01/2018, 17/09/2020	HTML

La información fue procesada para generar una colección organizada de manera tabular para la aplicación de un algoritmo de clasificación no supervisada. Para el caso de las fuentes de datos tipo *shapefile*, éstas fueron procesadas con el software para sistemas de información geográfica QGIS (QGIS, 2021) utilizando una capa vectorial con la delimitación de los acuíferos (CONABIO, 2018). Esto permitió obtener un valor promedio por acuífero generado de la intersección del polígono de cada acuífero con cada capa de información. De esta forma se obtuvieron los valores numéricos de los atributos de temperatura mínima, máxima y promedio anual, precipitación y evapotranspiración potencial. En el caso del uso del suelo, se obtuvo como resultado de la intersección el porcentaje de superficie del acuífero asociado a cada tipo de uso (agricultura, asentamientos humanos, bosque, cuerpos de agua, selva, vegetación y otros). De forma similar, en el caso de la permeabilidad, se asocia el porcentaje de superficie del acuífero que corresponda con cada tipo (baja, baja a media, media a alta o alta). Respecto a las concesiones de uso de agua subterránea, éstas fueron importadas a una base de datos MySQL (Oracle, 2020) con el fin de calcular el volumen concesionado de los títulos otorgados hasta el año 2009, agrupando dicho volumen por tipo de uso y expresado como porcentaje. De esta forma, se obtuvo la proporción de volumen concesionado para cada uso por acuífero, agrupando los usos por acuicultura, agrícola, agroindustrial, comercio, doméstico, industrial, pecuario, público, urbano, servicios, diferentes y otros usos.

En adición a los atributos que caracterizan los acuíferos, se determinó un atributo asociado a la disponibilidad volumétrica del acuífero, el cual fue calculado con la información de las publicaciones del DOF. Este atributo (denominado POR_DISP_DEF) se calcula dividiendo el volumen identificado como disponibilidad de agua subterránea (DAS) o su contraparte negativa (conocida como DÉFICIT) entre la recarga (R), tal y como lo muestra la ecuación 2. Las definiciones de estos términos son las contenidas en los numerales "3" y "4" de la Norma Oficial Mexicana NOM-011-CONAGUA-2000 (CONAGUA, 2000).

$$POR_DISP_DEF (\%) = \left(\frac{DAS + DÉFICIT}{R} \right) * 100 \quad (1)$$

Al finalizar este procesamiento, se conformó una base de datos con 30 atributos, representando el estado de los acuíferos en el año 2009 (Cuadro 3). Es importante señalar que el porcentaje de volumen disponible para años posteriores al 2009 (2013, 2015, 2018 y 2020) también fue calculado acorde a la Ecuación 1 para fines de validación, tal como se explica en la sección de resultados.

Cuadro 3. Atributos utilizados para la base de datos tabular de aprendizaje.

Grupo	Atributo	Unidades	Identificador
ID	Identificador oficial del acuífero (primero dos dígitos el estado, siguientes dos un valor secuencial)	No aplica	CLAVE_ACUIFERO
Climatología	Temperatura máxima promedio anual	°C	Tmin
	Temperatura mínima promedio anual	°C	Tmax
	Temperatura promedio anual	°C	Tprom
	Precipitación anual	mm	Pmedia



	Evapotranspiración potencial anual	mm	ETP
Uso del suelo	Porcentaje uso agricultura	%	US_Agricultura
	Porcentaje uso asentamientos humanos	%	US_Asentamiento
	Porcentaje uso suelo bosque	%	US_Bosque
	Porcentaje uso cuerpo de agua	%	US_CuerpoAgua
	Porcentaje uso suelo selva	%	US_Selva
	Porcentaje uso vegetación	%	US_Vegetacion
	Porcentaje uso otros usos	%	US_Otro
Permeabilidad	Porcentaje permeabilidad baja	%	PER_Baja
	Porcentaje baja a media	%	PER_Baja_Media
	Porcentaje baja a alta	%	PER_Baja_Alta
	Porcentaje media a alta	%	PER_Media_Alta
	Porcentaje alta	%	PER_Alta
Distribución volumen concesionado	Porcentaje acuicultura	%	R_Pacuicultura
	Porcentaje uso agrícola	%	R_Pagricula
	Porcentaje uso agroindustrial	%	R_Pagroindustrial
	Porcentaje comercio	%	R_Pcomercio
	Porcentaje diferentes usos	%	R_PDif_Usos
	Porcentaje uso doméstico	%	R_Pdomestico
	Porcentaje uso industrial	%	R_Pindustrial
	Porcentaje uso pecuario	%	R_Ppecuario
	Porcentaje uso público-urbano	%	R_Ppublico_urbano
	Porcentaje uso de servicios	%	R_Pservicios
	Porcentaje otros usos	%	R_Potros
Disponibilidad	Porcentaje de volumen disponible en el acuífero (2009)	%	POR_DISP_DEF

Descripción del algoritmo *nAcuifDef*

El algoritmo *nAcuifDef* (*datosAcuif*, *nA*, *nG*) recibe como entradas a *datosAcuif*, la base de datos tabular con los atributos que describen el estado de los acuíferos (Cuadro 3), la cantidad *nA*, que indica el número acuíferos solicitados próximos a caer en déficit y la variable *nG*, como el número de grupos a utilizar para realizar la estimación. En su ejecución, el algoritmo forma *nG* grupos basándose en la similitud de atributos y califica cada grupo en función de una métrica basada en el porcentaje de disponibilidad (Ecuación 1). Al formar los grupos, el algoritmo no utiliza el valor de *POR_DISP_DEF*, solo considera las variables que describen el acuífero. Posteriormente, el algoritmo ordena los grupos y selecciona los acuíferos más propensos de cada grupo a caer en déficit, siendo en este paso donde ya se utiliza el porcentaje de volumen disponible. El algoritmo realiza esta selección dentro de un ciclo iterativo, con el propósito de integrar la cantidad *nA* de acuíferos solicitados. La selección considera el tamaño de cada grupo (llamado en este contexto *clúster*) y la proporción de acuíferos que ya están en déficit. El algoritmo termina cuando la selección de los *nA* acuíferos es completada, o bien, que se agotan los candidatos, lo que puede suceder al solicitar una cantidad mayor de acuíferos que los

que se encuentran con disponibilidad. La Figura 2 muestra la función en R programada para hacer la selección de los próximos acuíferos en déficit.

```
# Regresa la información de los nA acuíferos próximos a caer en déficit
# agrupando en base a los atributos en datosAcuif y usando la cantidad
# nG de grupos para clasificación
nAcuifDef<-function(datosAcuif,nA,nG) {
  datosC=calificaAcuíferos(datosAcuif,nG,2)
  # damePropDef determina la proporción de acuíferos con déficit
  # en cada grupo
  propDefG=damePropDef(datosC,nG)
  nT=nA
  # difprop indica la proporción pendiente que no pudo
  # ser cubierta por los acuíferos con disponibilidad
  difprop=0
  # salida es una lista que almacena los acuíferos seleccionados
  salida=NULL
  g=1
  while (nA>0 && g<=nG) {
    # toma los acuíferos candidatos del grupo g
    # cuyo porcentaje de disponibilidad sea positivo
    candidatos=datosC[datosC$Grupo==g & datosC$POR_DISP_DEF>0,]
    candidatos=candidatos[order(candidatos$POR_DISP_DEF),]

    # El número de candidatos es proporcional a los acuíferos en déficit del cluster
    nc=round(nT*propDefG[g],0)
    if (nrow(candidatos)>=nc) { # hay suficientes candidatos, úsalos todos
      candidatos=candidatos[1:nc,]
    } else { # faltaron, calcula la diferencia de proporción pendiente
      difprop=propDefG[g]*(1-(nrow(candidatos)/nc))
      # distribuye la diferencia pendiente entre los grupos restantes
      if (difprop>0 && g<nG) {
        sumaprop=sum(propDefG[(g+1):nG])
        #actualiza la proporción para considerar la que no se cubrió pero de forma proporcional
        for (i in (g+1):nG) {
          propDefG[i]=propDefG[i]+(propDefG[i]/sumaprop)*difprop
        }
      }
    }
  }

  if (is.null(salida)) {
    salida=candidatos
  } else
  salida=rbind(salida,candidatos)
  nA=nA-nrow(candidatos)
  g=g+1
}
salida=salida[order(salida$POR_DISP_DEF),]
return (salida)
}
```

Figura 2. Código en R del algoritmo *nAcuifDef*

Es importante señalar, que el número *nG* de grupos en *nAcuifDef* no está predeterminado, y debe ser seleccionado de forma experimental. Sin embargo, el algoritmo propuesto considera que, al aumentar *nG*, la cantidad de acuíferos candidatos que se pueden tomar de los grupos puede disminuir, o incluso, pueden no existir, esto al tener grupos integrados únicamente por acuíferos con déficit. En este caso, se toma los grupos restantes el número de acuíferos faltantes para integrar los *nA* acuíferos solicitados, considerando siempre la proporción que debe otorgar el grupo.

Un elemento crítico en el *nAcuifDef* es la función *calificaAcuíferos(datosAcuif, nG, mcalif)*, responsable de formar los *nG* grupos y asignar un valor numérico en función de la métrica de calidad utilizada (representada por la variable *mcalif*). En la implementación propuesta, se utilizó como métrica la proporción de acuíferos que aún no están en déficit en el clúster, por lo que el valor de *mcalif* es 2. La alternativa es el promedio del porcentaje de todos los acuíferos del grupo, considerando también los que ya se encuentran en déficit. La

Figura 2 muestra la implementación en R de la función de evaluación para calificar a cada grupo de acuíferos.

```
library("e1071")

# califica los acuíferos en datosAcuif asignando a Grupo los valores de 1 a nG
# dependiendo de la métrica de calificación utilizada. El sentido es ascendente,
# 1=el grupo con la peor calificación, nG el mejor.
# mcalif==1 utiliza la disponibilidad promedio del grupo
# mcalif==2 utiliza la proporción de acuíferos que no están en déficit (default)
calificaAcuiferos<-function(datosAcuif,nG, mcalif=2) {
  # elimina atributos a no considerar en el agrupamiento (la variable Grupo
  # puede existir si no es el primer llamado a calificaAcuif)
  tabla=datosAcuif[,!(names(datosAcuif) %in% c("CLAVE_ACUIFERO", "POR_DISP_DEF", "Grupo"))]
  cmres=cmeans(tabla,nG)
  dpc=rep(0,length(cmres$clusters))
  for (i in 1:nG) {
    if (mcalif==1) {# la disponibilidad promedio del grupo
      dpc[i]=mean(datosAcuif[cmres$cluster==i,]$POR_DISP_DEF)
    }
    if (mcalif==2) # proporción de acuíferos no sobre-explotados
      dpc[i]=nrow(datosAcuif[cmres$cluster==i & datosAcuif$POR_DISP_DEF>0,])/
        nrow(datosAcuif[cmres$cluster==i,])
  }
  mi.dpc=data.frame("dpc"=dpc,"cluster"=(1:nG))
  mi.dpc=mi.dpc[order(mi.dpc$dpc),]
  # la siguiente línea crea la columna "Grupo"
  datosAcuif$Grupo=rep(-1,nrow(datos))
  for (i in 1:nG) {
    datosAcuif[cmres$cluster==mi.dpc$cluster[i],"Grupo"]=i
  }
  return(datosAcuif)
}
```

Figura 3. La función *calificaAcuíferos*

Como se observa, la función *calificaAcuíferos(datosAcuif, nG)* utiliza la función *cmeans(tabla, nG)* para formar los *nG* grupos utilizados en la clasificación. La variable *tabla* es igual a *datosAcuif*, una vez que se han eliminado las columnas que no deben ser consideradas para el agrupamiento, como lo son la identificación del acuífero y el número de grupo (que puede estar presente si no es el primer llamado a *calificaAcuíferos*). La función *cmeans* (librería *e1071* en R) es la implementación del algoritmo *Fuzzy cMeans* (FCM), un algoritmo de formación de conglomerados basado en distancias. Dentro del aprendizaje automático, la agrupación en clústeres es un ejemplo de aprendizaje no supervisado, que a diferencia de la clasificación, no se basa en clases predefinidas y ejemplos previamente etiquetados, por lo que es considerada una forma de aprendizaje mediante la observación en lugar de ejemplos (Han, Kamber, & Pei, 2012). El algoritmo FCM fue propuesto por Dunn en 1973 y mejorado por Bezdek en 1981 (Höppner, Klawonn, Kruse, & Runkler, 1999) y está sujeto al principio de que cada punto de datos pertenece a más de un clúster con diferentes valores de pertenencia, que van desde 0 a 1. La siguiente es una descripción del algoritmo tomada de (Kesemen, Tezel, Özkul, & Tiryaki, 2020) adaptada al contexto del algoritmo:

- 1) Especifica de antemano el número de grupos (conglomerados) requeridos (parámetro *nG* en la función *calificaAcuíferos*), de tal forma que $nG \in [2, n]$ donde *n* es el número máximo de acuíferos (653). Inicializa el parámetro de ponderación



$m > 0$ (usualmente 2), la cota de error $\varepsilon > 0$ y el número máximo de iteraciones $maxT$.

- 2) Proporciona pesos de forma aleatoria $\mu_{ij}^{(0)} \sim U(0,1)$. μ_{ij} es el valor de pertenencia del i -ésimo dato al j -ésimo grupo (clúster).
- 3) Inicializa el contador de iteraciones ($t = 1$)
- 4) Calcula los centroides de los nG clústers usando la Ecuación (2):

$$v_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m} \quad (j = 1, 2, \dots, nG) \quad (2)$$

- 5) Actualiza el valor de pertenencia μ_{ij} con v_j :

$$\mu_{ij} = \left(\sum_{k=1}^{nG} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, nG) \quad (3)$$

Donde x_i representa el vector de características del objeto i (acuífero i).

- 6) Calcula $e = \|\mu^{(t)} - \mu^{(t-1)}\|$
- 7) Si $e < \varepsilon$ o $t > maxT$ detén el algoritmo, en caso contrario $t = t + 1$ y regresa al paso 4

En el caso del llamado a *cmeans* utilizado en el algoritmo de la Figura 3, el número máximo de iteraciones utilizado fue el valor por definición $maxT=100$.

Resultados y Discusión

Pruebas de funcionamiento del algoritmo

Como se describió en la sección de datos utilizados, el algoritmo fue ejecutado tomando como base la información del estado de disponibilidad de los acuíferos asociada al año 2009 y sus características (Cuadro 3), que fueron uso del suelo (serie IV), temperatura máxima, mínima y promedio anual (promedios de los años 2005 al 2009 obtenidos del GCM), permeabilidad del suelo (CONABIO, 2018) y porcentaje de distribución de concesiones por tipo, con el fin de proyectar los próximos acuíferos a caer en déficit desde dicho año. Esto representa una simulación de lo que hubiera pasado si se hubiera utilizado el algoritmo con la información disponible en 2009. La clasificación realizada por el agrupamiento del algoritmo *nAcuifDef* fue comparada contra los resultados de una predicción simple basada en el ordenamiento ascendente del porcentaje de disponibilidad de los acuíferos que no estaban en déficit en dicho año, contrastando contra el porcentaje de disponibilidad de esos mismos acuíferos publicado en los años subsecuentes (CONAGUA, 2013, 2015, 2018b, 2020). La comparación fue realizada en grupos de cada 10 acuíferos, hasta alcanzar los 100 acuíferos. Para la determinación de este número, se consideró que, desde la primera publicación de disponibilidad de los acuíferos, 101 acuíferos han caído en déficit (Cuadro 1). Como un ejemplo de la comparación realizada,



considere la información de los primeros 20 acuíferos ordenados por porcentaje de disponibilidad de acuerdo con la información del año 2009 (Cuadro 3) y los 20 acuíferos seleccionados por el algoritmo *nAcuifDef* utilizando un número de *clústers* igual a 40 (Cuadro 4). En ambos casos, se muestra el porcentaje de disponibilidad que tendrán los acuíferos en las siguientes actualizaciones y el total de acuíferos que finalmente cayeron en déficit para cada año. Se puede observar que existen coincidencias en ambas listas; concretamente, los primeros cinco acuíferos seleccionados *nAcuifDef* están también en la selección del orden ascendente. Se observa también, que hay cinco acuíferos que dado este orden *nAcuifDef* no consideró; no obstante, en su lugar, el algoritmo selecciona nueve acuíferos que parecían muy alejados, incluyendo algunos con porcentajes de disponibilidad altos en 2009, como el acuífero 1017 (Valle del mezquital en Durango), que tenía una disponibilidad cercana al 80% y el acuífero 1103 (Ocampo en Guanajuato) con una disponibilidad del 97.46%. Esta selección es atribuible a su parecido con otros acuíferos que cayeron en déficit, inferido por el algoritmo FCM.

Cuadro 3. Estado del porcentaje de disponibilidad y cuantificación de acuíferos en déficit dado el orden ascendente del año 2009. Se resaltan en negritas los valores con porcentaje de disponibilidad negativa, esto acorde al cálculo de la Ecuación 1.

Clave	Estado	Nombre del acuífero	Porcentaje de disponibilidad (%)					En déficit
			2009	2013	2015	2018	2020	
0844	Chihuahua	Valle de Zaragoza	0.021	23.874	23.874	19.876	-14.383	1
0843	Chihuahua	Bocoyna	0.045	0.045	0.045	20.959	20.672	0
0846	Chihuahua	Carichi-Nonoava	0.070	0.358	0.358	23.751	22.961	0
0209	Baja California	Laguna Salada	0.178	-0.222	-8.811	8.252	5.887	1
2625	Sonora	Rio San Miguel	0.284	25.485	26.867	24.059	3.344	0
1505	Estado de México	Villa Victoria - Valle de Bravo	0.390	0.372	0.304	-0.436	-0.438	1
0306	Baja California Sur	Santo Domingo	0.445	1.023	0.979	-20.579	-20.746	1
2632	Sonora	Rio Frontera	0.859	41.871	41.993	16.862	11.251	0
2641	Sonora	Cocoraque	1.176	1.588	2.261	0.061	-0.085	1
2105	Puebla	Valle de Tehuacán	1.285	14.375	14.428	10.933	11.421	0
1804	Nayarit	Valle de Matatipac	1.499	23.908	17.091	17.272	13.439	0
0601	Colima	Colima	1.557	5.256	3.403	2.071	3.245	0
2412	San Luis Potosí	Jaral de Berrios-Villa de Reyes	1.623	-0.002	-0.002	0.728	-1.542	1
2624	Sonora	Rio Sonora	1.812	8.070	8.199	7.694	4.325	0
2025	Oaxaca	Valles Centrales	1.833	10.480	9.766	8.212	5.150	0
1917	Nuevo León	Sandia-La Unión	1.841	-0.930	-1.261	-17.044	-40.118	1
2205	Querétaro	Valle de Tequisquiapan	1.855	4.785	6.732	1.108	-2.373	1
0825	Chihuahua	El Sabinal	1.867	1.867	1.867	49.524	-54.962	1
1321	Hidalgo	Amajac	2.087	0.836	0.827	0.788	0.559	0
0815	Chihuahua	Laguna El Diablo	2.204	2.204	2.204	0.265	-0.128	1
Total de acuíferos en déficit			0	3	3	3	9	10

Cuadro 4. Estado del porcentaje de disponibilidad y cuantificación de acuíferos en déficit dada la selección del algoritmo nAcuifDef. Se resaltan en negritas los valores con porcentaje de disponibilidad negativa, esto acorde al cálculo de la Ecuación 1.

Clave	Estado	Nombre del acuífero	Porcentaje de disponibilidad (%)					En déficit
			2009	2013	2015	2018	2020	
0844	Chihuahua	Valle de Zaragoza	0.021	23.874	23.874	19.876	-14.383	1
0306	Baja California Sur	Santo Domingo	0.445	1.023	0.979	-20.579	-20.746	1
2412	San Luis Potosí	Jaral De Berrios-Villa De Reyes	1.623	-0.002	-0.002	0.728	-1.542	1
1917	Nuevo León	Sandia-La Unión	1.841	-0.930	-1.261	-17.044	-40.118	1
2205	Querétaro	Valle de Tequisquiapan	1.855	4.785	6.732	1.108	-2.373	1
0825	Chihuahua	El Sabinal	1.867	1.867	1.867	49.524	-54.962	1
0201	Baja California	Tijuana	2.753	45.618	45.633	1.659	1.608	0
0312	Baja California Sur	Cañada Honda	3.515	11.493	11.493	6.467	-2.944	1
0506	Coahuila	El Hundido	4.066	-7.815	-10.786	-21.629	-25.655	1
0501	Coahuila	Allende-Piedras Negras	4.080	3.988	14.847	2.882	-7.094	1
1622	Michoacán	Cotija	7.267	3.613	2.643	1.391	-0.013	1
1920	Nuevo León	Campo Papagayos	7.819	7.819	7.819	3.638	19.717	0
0248	Baja California	Real Del Castillo	8.481	13.385	9.630	7.081	5.071	0
1101	Guanajuato	Xichu-Atarjea	9.997	10.323	9.964	4.418	9.572	0
3216	Zacatecas	Sain Alto	10.954	10.409	8.804	24.356	7.138	0
0834	Chihuahua	Parral-Valle del Verano	11.855	13.261	13.154	-45.211	-127.909	1
2601	Sonora	Valle de San Luis Río Colorado	12.245	-3.179	-3.179	17.540	-39.424	1
3209	Zacatecas	Jalpa-Juchipila	23.603	13.777	6.260	6.875	0.055	0
1017	Durango	Valle del Mezquital	79.983	79.983	79.983	62.577	-257.440	1
1103	Guanajuato	Ocampo	97.469	71.497	71.497	-14.981	71.046	1
Total de acuíferos en déficit			0	4	4	5	13	14

En términos generales, se ordenaron los acuíferos de acuerdo con su porcentaje de disponibilidad del año 2009, siguiendo la suposición que los acuíferos caerían en déficit de forma asociada a este orden. Esto se validó para los siguientes 100 acuíferos, agrupando la predicción por números de 10, comparando el estado en el que dichos acuíferos caerían en los años 2013, 2015, 2018 y 2020 cuantificando las veces que efectivamente dichos acuíferos cayeron en estado de déficit. Esta estimación fue comparada contra la predicción del algoritmo variando la cantidad de grupos utilizados (nG), empezando en 10 y llegando hasta 60 grupos. Los resultados se muestran en el Cuadro 5.



Cuadro 5. Comparación de la estimación de la cantidad de acuíferos que caerán en déficit en los años 2013-2020 con los métodos de ordenamiento simple (columna 2) y el algoritmo *nAcuifDef* para diferentes cantidades de *clústers* (columnas 3 a 8).

Rango de acuíferos	Acuíferos en déficit en el futuro de acuerdo con el orden de 2009	Acuíferos en déficit correctamente pronosticados por <i>nAcuifDef</i> por cada <i>nG</i> (número de clústers) utilizado					
		<i>nG</i> =10	<i>nG</i> =20	<i>nG</i> =30	<i>nG</i> =40	<i>nG</i> =50	<i>nG</i> =60
10	5	5	5	6*	5	6*	6*
20	10	10	11	12	14	14	16*
30	15	14	14	17	17	19	21*
40	19	18	25	19	18	22	26*
50	22	22	25	26	28*	27	28*
60	28	25	29	30	33*	30	32
70	28	26	31	43*	41	38	32
80	33	32	35	39	44*	41	42
90	41	34	37	44	47	49*	47
100	46	37	40	50	54*	52	40

Como se puede observar, *nAcuifDef* puede tener una eficiencia menor en la estimación de los próximos acuíferos en déficit para valores de *nG* muy bajos, como 10 o 20, o altos, como 60. Sin embargo, existen valores para *nG* donde por lo regular se obtuvieron mejores estimaciones, que corresponden a los casos de *nG* igual a 30,40 o 50. Esto está acorde a un proceso de calibración presente en casi todas las técnicas de aprendizaje automático, donde al parecer existen una falta de entrenamiento con un número bajo de grupos y por otro lado un sobre-entrenamiento con números altos de *nG*. No obstante, en los parámetros donde *nAcuifDef* logra estimaciones más certeras que el orden simple, se observa que puede lograr pronosticar de manera acertada hasta 15 acuíferos más, como ocurre en el caso del rango de los 70 acuíferos y un *nG*=30. En esta combinación se observa que se predicen de manera acertada 41 acuíferos en déficit contra los 28 que realiza el ordenamiento simple. En el extremo para 100 acuíferos, un *nG*=40 logra 8 acuíferos más, con 54 acuíferos comparados contra 46 del ordenamiento simple. Es natural que el número de acuíferos “acertados” por *nAcuifDef* descienda conforme se aumenta el rango, ya que los acuíferos con disponibilidad se van agotando y sus resultados van coincidiendo con el orden ascendente. Por ejemplo, es claro visualizar que, al solicitarle a *nAcuifDef* un pronóstico de acuíferos a caer en déficit de cantidad igual a los que tenían un porcentaje disponible en el año 2009 (que eran 470), su resultado sería el mismo que el pronosticado por el orden ascendente.

Conclusiones

Este trabajo presenta un algoritmo basado en la técnica de clustering difuso (*Fuzzy CMeans, FCM*) para detectar los próximos acuíferos a caer en déficit basándose en la similitud con otros acuíferos que ya presentan dicho estado. Las pruebas realizadas indican que es necesario un proceso de calibración para obtener mejores resultados que realizar una selección basada en el ordenamiento del porcentaje de disponibilidad de los



acuíferos. Esto es de gran utilidad para monitorear acuíferos que de momento no parecen correr algún riesgo, pero que por sus características podrían caer en déficit muy pronto. Dada la gravedad del problema de la sobreexplotación de los recursos hídricos actuales y la velocidad con la que este crece, se requieren de técnicas que se puedan aplicar masivamente y que produzcan resultados rápidamente, situación en la que el aprendizaje no supervisado tiene mucho que aportar. Aún quedan parámetros por mejorar en el algoritmo, como la inclusión de otro tipo de atributos para clasificar los acuíferos y la determinación automática del número de grupos requerido para obtener buenos resultados. Sin embargo, a pesar de sus limitaciones, se considera que el algoritmo presentado es un ejemplo de la forma que las técnicas de aprendizaje automático se pueden aplicar para lidiar con problemas relacionados con el aprovechamiento de los recursos hídricos cuya disponibilidad es de naturaleza difícil o impredecible.

Referencias Bibliográficas

- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- CONABIO, I. N. de E. y G. (2018). Geoportal del Sistema Nacional de Información sobre Biodiversidad.
- CONAGUA. (2000). Norma Oficial Mexicana NOM-011-CONAGUA-2000 Conservación del recurso agua, Que establece las especificaciones y el método para determinar la disponibilidad media anual de las aguas nacionales. México, D. F.: Comisión Nacional del Agua.
- CONAGUA. (2009, agosto 28). ACUERDO por el que se da a conocer la ubicación geográfica de 371 acuíferos del territorio nacional, se actualiza la disponibilidad media anual de agua subterránea de 282 acuíferos, y se modifica, para su mejor precisión, la descripción geográfica de 202. *Diario Oficial de la Federación de México*. Recuperado de http://www.dof.gob.mx/nota_detalle.php?codigo=5107344&fecha=28/08/2009
- CONAGUA. (2010a, julio 8). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 36 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológicas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5151074&fecha=08/07/2010
- CONAGUA. (2010b, julio 8). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 44 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológicas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5151075&fecha=08/07/2010
- CONAGUA. (2010c, agosto 16). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 41 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológicas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5155455&fecha=16/08/2010
- CONAGUA. (2011a, enero 25). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 50 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones



- hidrológicas administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5175574&fecha=25/01/2011
- CONAGUA. (2011b, diciembre 14). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 142 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológico-administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5224246&fecha=14/12/2011
- CONAGUA. (2011c, diciembre 14). ACUERDO por el que se da a conocer el resultado de los estudios de disponibilidad media anual de las aguas subterráneas de 58 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológicas administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5224245&fecha=14/12/2011
- CONAGUA. (2013, diciembre 20). ACUERDO por el que se actualiza la disponibilidad media anual de agua subterránea de los 653 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológico-administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de http://dof.gob.mx/nota_detalle.php?codigo=5327360&fecha=20/12/2013
- CONAGUA. (2015, abril 20). ACUERDO por el que se actualiza la disponibilidad media anual de agua subterránea de los 653 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológico-administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de http://www.dof.gob.mx/nota_detalle.php?codigo=5389380&fecha=20/04/2015
- CONAGUA. (2017). Otorgamiento de concesiones, asignaciones y permisos para el uso explotación y/o aprovechamiento de las aguas nacionales y sus bienes públicos inherentes, así como el registro de obra situadas en zonas de libre alumbramiento. Recuperado el 1 de junio de 2020, de <https://datos.gob.mx/busca/dataset/concesiones-asignaciones-permisos-otorgados-y-registros-de-obras-situadas-en-zonas-de-libre-alu/resource/7cf416a3-32c8-4e5d-8a0d-0f184a772cf4>
- CONAGUA. (2018a). Estadísticas del Agua en México 2018, 303.
- CONAGUA. (2018b, enero 4). ACUERDO por el que se actualiza la disponibilidad media anual de agua subterránea de los 653 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las Regiones Hidrológico-Administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://dof.gob.mx/nota_detalle.php?codigo=5510042&fecha=04/01/2018
- CONAGUA. (2020, septiembre 17). ACUERDO por el que se actualiza la disponibilidad media anual de agua subterránea de los 653 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológico-administrativas que se indican. *Diario Oficial de la Federación de México*. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5600593&fecha=17/09/2020
- Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. *Water Resources Research*. <https://doi.org/10.1029/2000WR900368>
- Crosbie, R. S., Davies, P., Harrington, N., & Lamontagne, S. (2015). Ground truthing



- groundwater-recharge estimates derived from remotely sensed evapotranspiration: a case in South Australia. *Hydrogeology Journal*. <https://doi.org/10.1007/s10040-014-1200-7>
- Daliakopoulos, I. N., Coulibaly, P., & Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Gao, L., Connor, J. D., & Dillon, P. (2014). The economics of groundwater replenishment for reliable urban water supply. *Water (Switzerland)*. <https://doi.org/10.3390/w6061662>
- Han, J., & Kamber, M. (2006). Data Mining, Southeast Asia Edition: Concepts and Techniques. *Morgan kaufmann*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. (Wiley IBM PC Series, Ed.) (July 1999).
- Kanyama, Y., Ajoodha, R., Seyler, H., Makondo, N., & Tutu, H. (2020). Application of Machine Learning Techniques in Forecasting Groundwater Levels in the Grootfontein Aquifer. En *2020 2nd International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2020*. <https://doi.org/10.1109/IMITEC50163.2020.9334142>
- Kesemen, O., Tezel, Ö., Özkul, E., & Tiryaki, B. K. (2020). Fuzzy c-Means Directional Clustering (FCMDC) algorithm using trigonometric approximation. *Turkish Journal of Electrical Engineering and Computer Sciences*. <https://doi.org/10.3906/elk-1903-118>
- Lopez, C., Tucker, S., Salameh, T., & Tucker, C. (2018). An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *Journal of Biomedical Informatics*, 85, 30–39. <https://doi.org/https://doi.org/10.1016/j.jbi.2018.07.004>
- MacQueen. (1967). Some methods for classification and analysis of multivariate observations. *Berkeley Symp. Math. Statist. Prob.*
- Oracle. (2020). MySQL :: About MySQL.
- QGIS. (2021). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>. *Qgisorg*.
- Research Climate Group. (s/f). Global Climate Monitor. Recuperado el 15 de octubre de 2020, de <https://www.globalclimatemonitor.org/>
- SEGOB. (s/f). Diario Oficial de la Federación. Recuperado el 20 de julio de 2020, de <https://www.dof.gob.mx/>
- Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. En *Journal of Hydroinformatics*. <https://doi.org/10.2166/hydro.2008.015>
- Steyn, M. (2018). Short-term stream flow forecasting and downstream gap infilling using machine learning techniques.
- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., & Zhang, B. (2018). Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science*. <https://doi.org/10.1007/s13201-018-0742-6>