

UNIVERSIDAD AUTÓNOMA DE ZACATECAS
“Francisco García Salinas”



**Clasificación de la temperatura del agua
subterránea, utilizando Machine Learning**

Tesis para obtener el grado de:
Maestro en Ciencias del Procesamiento de la Información

Presenta
Ing. Carlos Regis Robles

Director:
Dr. Carlos Francisco Bautista Capetillo

Co-Directores:
Dr. Julián González Trinidad
Dr. Efrén González Ramírez

Asesores:
Dr. Hugo Enrique Júnez Ferreira
Dr. Cruz Octavio Robles Roveló

Zacatecas, Zac., noviembre 2023



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL



Zacatecas Zac a 14 de noviembre de 2023

Carlos Regis Robles

Estudiante de la MCPI

Presente.

Dr. Huizilopoztli Luna García

Responsable de la MCPI

Nos es grato comunicarle que después de haber sometido a revisión académica la propuesta de tesis intitulada **“Clasificación de la temperatura del agua subterránea, utilizando Machine Learning”**, presentada por el estudiante Ing. Carlos Regis Robles y habiendo efectuada todas las correcciones indicadas por este comité académico, se **AUTORIZA** el documento de tesis para su impresión.

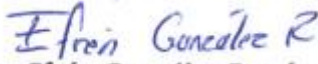
Sin más por el momento reciba un cordial saludo.

COMITÉ TUTORIAL

PROCESAMIENTO Y ANÁLISIS DE DATOS


Dr. Julián González Trinidad


Dr. Carlos Francisco Bautista Capetillo


Dr. Efrén González Ramírez


Dr. Hugo Enrique Júnez Ferreira


Dr. Cruz Octavio Robles Rovelo

C.c.p. Interesado

C.c.p. Responsable de la Maestría en Ciencias en Procesamiento de la Información.



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL

**COORDINACIÓN DE
INVESTIGACIÓN Y POSGRADO**

Carta de similitud núm. 486/ IyP
Zacatecas, Zacatecas 17/noviembre/2023

Dr. Huizilopoztli Luna García
Responsable de la MCPI – UAZ
Presente

Estimado Dr. Huizilopoztli,

Después de saludarlo, sirva el presente oficio para notificar que el documento

"Clasificación de la temperatura del agua subterránea, utilizando Machine Learning"
de Carlos Regis Robles

Fue analizado con el software Copyleaks, con la intención de detectar similitudes; el resultado en cuestión fue

18.4 % de similitud

De acuerdo a lo anterior, el porcentaje se considera **ACEPTABLE** de acuerdo a los estándares internacionales.

Atentamente
"Somos Arte, Ciencia y Desarrollo Cultural"

Dr. Carlos Francisco Bautista Capetillo
Coordinador de Investigación y Posgrado
Universidad Autónoma de Zacatecas



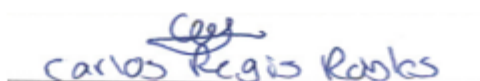
Zacatecas, Zac., 23 de noviembre de 2023

Carta Cesión de Derechos

A QUIEN CORRESPONDA

El que suscribe C. Carlos Regis Robles alumno del Programa **Maestría en Ciencias del Procesamiento de la Información** con numero de matricula **35163249**, adscrito a la Unidad Académica de Ingeniería Eléctrica de la Universidad Autónoma de Zacatecas, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Carlos Francisco Bautista Capetillo y cede los derechos del trabajo titulado: **"Clasificación de la temperatura del agua subterránea, utilizando Machine Learning"** a la Universidad Autónoma de Zacatecas para su difusión, con fines académicos y de investigación.

ATENTAMENTE:


Carlos Regis Robles



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL



AGRADECIMIENTO ESPECIAL

Al Programa Nacional de Posgrados de Calidad (PNPC) del Consejo Nacional de Ciencia y Tecnología (CONACYT) por su apoyo económico a través de la convocatoria “Becas Nacionales (Tradicional) 2021-2023”

Agradecimientos

A mi familia y seres queridos, quienes siempre me han brindado su apoyo para cumplir cualquiera de mis metas.

A mis directores y asesores de tesis; por su constancia, tiempo, dedicación a este trabajo y por compartir sus conocimientos.

A la casa de estudios, la Universidad Autónoma de Zacatecas, por sus posgrados de calidad.

Dedicatoria

A mi familia y seres queridos, les dedico a ustedes esta meta alcanzada. Gracias por creer en mí y estar junto a mí en todo momento. A mi pareja, por su constante apoyo incondicional, por impulsarme y motivarme a superarme y vencer mis miedos.

Índice	Pág
I. Introducción.....	1
II. Antecedentes.....	4
2.1.- Generalidades	4
2.2.- Definiciones.....	6
Hidrología	6
Acuífero	7
La Recarga de los Acuíferos	8
Agua subterránea	8
Química del Agua	9
Pozo.....	11
Movimiento de las aguas subterráneas.....	12
Monitoreo de temperatura con leveloggers.....	13
Teoría de sistemas de flujo.....	14
Inteligencia Artificial	15
Machine Learning	15
Tipos de Machine Learning	16
Aprendizaje Supervisado	16
Aprendizaje No Supervisado	17
Clustering.....	17
Clustering por Distancia.....	17
Clustering por Probabilidad	18
K-means	18
Elección del número de clústeres.....	21
Escalado de los datos	22
Z-score	22
Limitaciones k-means	23
Correlación.....	23
Análisis de correlación.....	24
Métodos de Correlación	24
Método Pearson	25
Limitaciones.....	25
2.3.- Principales estudios relacionados	26

2.4.- Contribuciones y limitaciones de estudios previos.....	27
2.5.- Comparación entre los trabajos relacionados y la propuesta de investigación.....	29
III. Metodología	32
3.1.- Localización de la zona de estudio	32
3.2.- Modelo o esquema general de investigación	34
3.3.- Modelo de investigación.....	34
3.4.- Descripción de la propuesta.....	35
3.5.- Desarrollo del modelo de clasificación	37
IV. Resultados y Discusiones	39
4.1.- Registro de temperaturas diarias con leveloggers	39
4.2.- Modelo K-means con la Química del agua.....	42
V. Conclusiones.....	45
VI. Literatura citada.....	46
VII. Trabajos publicados	48

Índice de Figuras

Figura 1. Comportamiento de los sistemas de flujo. Tomado de Woessner, William W. and Eileen P. Poeter 2020	5
Figura 2. Esquema de los tipos de acuíferos. (Tomado de Ordoñez Gáldez, 2011).	8
Figura 3. Esquema del agua subterránea (Tomado de Woessner, William W. and Eileen P. Poeter 2020).	9
Tabla 1. Elementos Químicos del agua subterránea, tomado de Freeze & Cherry, 1979.	10
Tabla 2. Muestras de del agua subterránea, tomado de Todd & Mays, 2005.	11
Figura 4. Pozo de la Red COTAS (Tomada de COTAS-UAZ).	11
Figura 5. Sensor levellogger y lector óptico marca Solinist. Tomado de la página web de Solinist www.solinist.com	14
Figura 6. Funcionamiento del algortimo k-means. Tomado de https://www.unioviado.es/compnum/laboratorios_py/kmeans/kmeans.html	21
Tabla 3. Contribuciones de estudios relacionados	28
Figura 7. Localización del Acuífero.	33
Figura 8. Modelo General de investigación	35
Figura 9 Levelloggers y Química del agua.	36
Figura 10 Agrupación K-means con R studio	37
Figura 11 Código R studio	38
Figura 12 Matriz de correlación de los levelloggers	39
Figura 13 Mapa de Levelogeers relacionados	40
Figura 14 Matriz de correlación de la química del agua	41
Figura 15 Mapa con el modelo K-means	42
Figura 16 Mapa Espacial orográfico de los grupos con k-means.	43

I. Introducción

El agua subterránea es el gran reservorio que tiene la humanidad para abastecer de agua a los diferentes usos (riego, urbano, industrial, pecuario y ambiental), por lo cual el estudio de esta es prioritario, debido a que, durante las últimas décadas, la demanda del agua ha crecido al igual que la población y la agricultura de riego. A nivel mundial la accesibilidad al agua ha disminuido a la vez que su contaminación ha aumentado por lo que la gestión que garantice sustentabilidad requiere de trabajos científicos que permitan entender y comprender los procesos que determinan la calidad y cantidad del agua subterránea, la estimación clásica utiliza la ecuación de balance como referencia. En las últimas décadas se han desarrollado una serie de técnicas que permiten un mejor análisis del comportamiento del agua, como la teoría de sistemas de flujo (TSF), inteligencia artificial (IA) y dentro de redes neuronales (NR) entre otras (Graf and Aghelpour, 2021).

Se estima que anualmente, México recibe aproximadamente 1, 449, 471 millones de metros cúbicos de agua en forma de precipitación pluvial, de esta agua, el 72.1% se evapotranspira y regresa a la atmósfera, el 21.4% escurre por los ríos o arroyos, y el 6.4% restante se infiltra al subsuelo de forma natural y recarga los acuíferos. Tomando en cuenta los flujos de salida (exportaciones) y de entrada (importaciones) de agua con los países vecinos, el país anualmente cuenta con 451, 584.7 millones de metros cúbicos de agua dulce renovable (CONAGUA, 2019).

El agua sin calidad potable no puede ser útil para ningún fin y la evaluación de las variaciones espaciales de la vulnerabilidad de las aguas subterráneas es un gran problema debido a la participación de diversos factores como la distribución de las precipitaciones, la escorrentía, el tamaño de grano del suelo, las características topográficas, el tipo de accidentes geográficos, las condiciones de drenaje y las características litológicas, prácticas de uso de la tierra, profundidad de las aguas subterráneas y restricciones ambientales y añadiéndole que estas no son uniformes en ningún aspecto (Subba Rao et al., 2013).

Para entender el comportamiento del agua subterránea la teoría de sistemas de flujo utiliza algunas variables como piezometría, geología, hidrogeoquímica, temperatura del agua y las interacciones entre ellas, para entender el movimiento del agua subterránea, identificando posibles zonas de recarga y descarga (Cardona, 2020). Investigaciones recientes reportan que el comportamiento de las condiciones térmicas del agua y disponer de esta información ayudan al

seguimiento del estado ecológico y el buen funcionamiento del ecosistema acuático, ya que la calidad del agua cuando en sus características físicas, químicas y biológicas se deteriora y se vuelve perjudicial para el hombre, entonces se dice que está contaminada (Graf & Aghelpour, 2021; Hossain et al., 2013; Subba Rao et al., 2013).

Los modelos de redes neuronales artificiales (ANN) en el sector de los recursos hídricos han demostrado ser una herramienta valiosa para el modelado de los procesos hidrológicos mediante técnicas de modelado y análisis de reconocimiento de patrones. La utilidad de las ANN en el estudio de la hidrología del agua subterránea se ha utilizado ampliamente para la predicción del nivel del agua subterránea, estimación de parámetros de los acuíferos, entre otros. (Roshni et al., 2019)

Las técnicas de para el diseño de redes de aguas subterráneas están impulsadas por datos hasta esencialmente impulsadas por procesos. Los enfoques basados en datos incluyen las técnicas estadísticas y de inteligencia artificial (IA), estas nos ayudan a identificar puntos en común y patrones en los datos como objetivo principal, y de esto inferir los procesos que los causan. Por el otro lado, las técnicas que se sana en procesos primero intentan simular procesos mecánicos que se cree que controlan la calidad del agua subterránea, como las reacciones geoquímicas entre el agua y la roca, y, en segundo lugar, comparan los resultados modelados con las observaciones. Las ventajas de las técnicas basadas en datos para la las investigaciones de aguas subterráneas es que no se basan en el conocimiento de la variación espacial de propiedades como la mineralogía, la porosidad, el área superficial reactiva, entre otras, que se necesitan como entrada para los enfoques basados en procesos. La principal ventaja de las técnicas basadas en procesos es que brindan información sobre los factores de control que pueden ser difíciles de inferir utilizando las técnicas basadas en datos (Daughney et al., 2012).

En esta investigación se utilizará, la teoría de sistemas de flujo, para entender el movimiento del agua en un acuífero, empleando la red de leveloggers (CONAGUA; COTAS 2015), que se tiene instalados en el acuífero “Calera”, la red contiene 17 aprovechamientos donde se tienen instalados los equipos para monitorear la profundidad y temperatura del agua subterránea. El monitoreo del comportamiento de la temperatura del agua subterránea, a través de los leveloggers, permite generar información de la evolución del agua en un acuífero, por otro lado, ayuda a identificar los sistemas de flujo. Sin embargo, uno de los inconvenientes que se tiene es la obtención en campo de los datos debido al costo elevado de los equipos y del personal que

realiza el monitoreo.

Empleando para ello el uso de técnicas de inteligencia artificial (IA), como lo son las redes neuronales artificiales, ayudaran a generar información de manera más rápida, fácil y sin uso de equipos y personal costoso para realizar estudios en acuíferos. Con base en lo anterior se tiene la pregunta de investigación, ¿Cuáles han sido las relaciones de las variaciones, de temperatura registradas con el leveloggers en el acuífero Calera, en los diferentes flujos?, para dar respuesta a esta se genera la Hipótesis El uso de un modelo k-means, permite pronosticar la variabilidad espacio-temporal de la temperatura registrada por levelogger, en los diferentes flujos del acuífero Calera. Con el Objetivo general, estimar el comportamiento espacio-temporal de la temperatura, de los leveloggers del acuífero Calera, Zacatecas y su clasificación con un algoritmo k-means, para los flujos. Desglosado en los objetivos específicos a) Obtener la información de temperatura de la red de instrumentación instalada en el acuífero Calera, b) analizar la variabilidad de la temperatura registrada con el leveloggers a través del tiempo y el espacio y c) elaborar el modelo k-means para la clasificación de los flujos subterránea.

II. Antecedentes

2.1.- Generalidades

Las aguas subterráneas desempeñan un papel de gran importancia en el crecimiento socioeconómico del país, gracias a sus características físicas que les permiten extraer agua en cualquier época del año de prácticamente cualquier punto de la superficie del acuífero y además funcionan como filtros purificadores, preservando la calidad del agua. A partir del año 2001 el número de acuíferos sobreexplotados ha variado anualmente entre 100 y 115. Al 31 de diciembre de 2018 se reportan 115 acuíferos sobreexplotados (CONAGUA, 2019)

La aplicación de la ecuación de balance permite realizar algunas estimaciones de la cantidad de agua disponible en un acuífero, con base en los criterios de registros de lluvia e infiltración uniforme en toda el área del acuífero, además de la experiencia del que realiza el cálculo, lo cual, ha generado que se tiene poca disponibilidad de agua, ya que generalmente estos balances son negativos (Júnez y González, 2020). Además, se asume que la infiltración es uniforme en todo el espesor, y por otro lado no existe una interconexión entre acuíferos.

Para contribuir a entender mejor el sistema se ha aplicado la teoría de sistemas de flujo, que indica que existe una interconexión entre las diferentes formaciones geológicas por donde circula el agua subterránea (Figura 1). Una forma de demostrar esto es el monitoreo de la evolución de la temperatura del agua, bajo el principio de que bajas temperaturas predomina el flujo local, temperaturas del orden de 18-28 °C flujo intermedios y mayores de 30 °C flujos regionales (Woessner & Poeter, 2020).

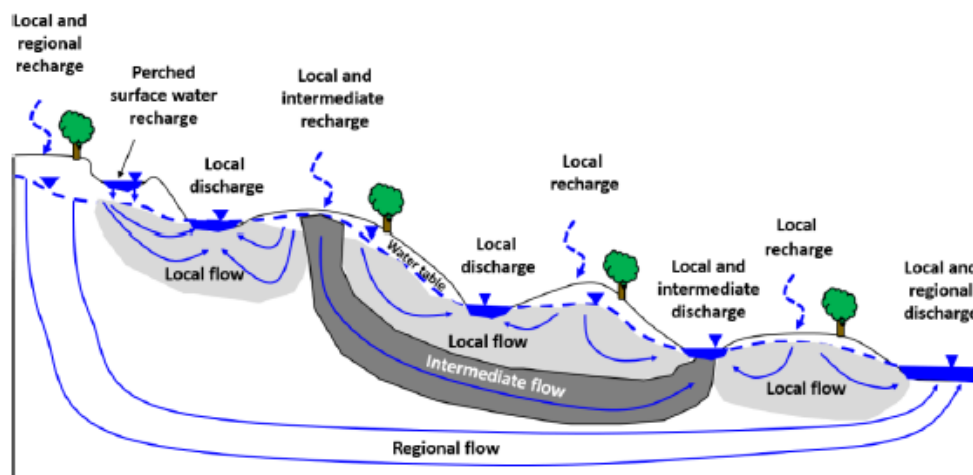


Figura 1. Comportamiento de los sistemas de flujo. Tomado de Woessner, William W. and Eileen P. Poeter 2020

El agua subterránea es un recurso vital para la sociedad, la industria, la agricultura y la vida silvestre, especialmente en ambientes áridos y semiáridos, ya que en estas regiones la precipitación pluvial, es variable y considerablemente inferior a la tasa de evapotranspiración; por lo tanto, el agua subterránea puede ser el componente principal del recurso hídrico, además el agua obtenida de fuentes subterráneas debe ser químicamente pura y razonablemente libre de microorganismos (Alizamir et al., 2018; Hossain et al., 2013; Tissen et al., 2019).

En el estudio de la evolución del agua subterránea es común encontrar que los procesos químicos de la zona de agua subterránea son los que mayor influyen al cambio de los paisajes. Al analizar los componentes disueltos en el agua subterránea ayuda a generar un panorama de historia geológica y su influencia en el suelo por donde ha transitado (depósitos de minerales, tipos de roca, entre otros). Los componentes químicos y bioquímicos del agua subterránea establecen el uso que se le puede dar al agua en las ramas de la industria, agricultura y el hogar (Freeze & Cherry, 1979).

Es importante establecer planes para la gestión del abastecimiento de recursos hídricos y es necesario contar con el conocimiento del ciclo anual de circulación de las aguas subterráneas, ya que, para el análisis del balance anual de agua, el paso más fundamental es evaluar con precisión la cantidad total de precipitación anual, la escorrentía directa y la descarga de agua subterránea (T. Kim et al., 2007).

En la literatura se reportan algunas investigaciones que utilizan el uso de datos de temperatura de agua subterránea por ejemplo (Danis, 2014) quien hace uso de este tipo de información en los reportes de exploración geotérmica para la calibración de modelos térmicos 3D y la revisión del potencial geotérmico de la cuenca, ya que estos son ventajosos, considerando algunas pautas como el equilibrio térmico.

El estudio de la calidad del agua subterránea proporciona información sobre las evaluaciones de las reacciones que producen la química natural del agua y la recarga, movimiento, mezcla y descarga del agua subterránea. Además, que con la química de los sistemas de flujo de agua subterránea se puede utilizar para determinar los cambios en la química del agua subterránea, las trayectorias de flujo del agua subterránea, la mezcla del agua subterránea y las edades y tasas de flujo del agua subterránea (Todd, David; Mays, 2005).

Según las investigaciones de (Agudelo-Vera et al., 2020; Maldonado et al., 2022), las variaciones de temperatura del agua subterránea tienen un impacto importante en los procesos físicos, químicos y biológicos, incluida la solubilidad de los gases en el agua subterránea, las tasas de reacción geoquímica, las tasas de respiración microbiana, la propagación de patógenos, entre otros; esto llega a ser determinante en la calidad del agua.

2.2.- Definiciones

Hidrología

Existen varias definiciones de hidrología:

La hidrología es la ciencia del estudio del agua. Sin embargo, la hidrología contemporánea no estudia todas las propiedades del agua, en cambio la hidrología moderna se ocupa de la distribución del agua en la superficie de la tierra y su movimiento, además también considera el agua debajo de la superficie y a través de la atmósfera. De acuerdo a la definición anterior sugiere que un hidrólogo es el encargado de estudiar el agua pero en realidad solo estudia el agua dulce ya que es la de principal preocupación (Davie, 2008).

El Consejo Nacional de Investigación de EE. UU. (1991) definió que la hidrología es la ciencia que trata las aguas de la Tierra, su ocurrencia, circulación y distribución, sus propiedades químicas y físicas, y su reacción con el medio ambiente, incluyendo la relación con los seres vivos. El dominio de la hidrología abarca la historia de vida completa del agua en la Tierra.

De acuerdo con Aparicio (1989), la hidrología es la ciencia natural que estudia el agua, su ocurrencia, circulación y distribución en la superficie terrestre, sus propiedades químicas y físicas y su relación con el medio ambiente, incluyendo los seres vivos.

Sin embargo, para esta investigación se centrará en la parte de la hidrología subterránea y todos sus elementos que la comprenden. Esta puede ser definida de la siguiente manera:

La hidrología de aguas subterráneas puede definirse como la ciencia de la ocurrencia, distribución y movimiento del agua debajo de la superficie de la tierra (Todd, David; Mays, 2005).

Acuífero

De todas las palabras del vocabulario hidrológico, probablemente no haya ninguna con más significados que el término acuífero, ya que significa diferentes cosas para diferentes personas, y tal vez diferentes cosas para la misma persona en diferentes momentos. Generalmente esta se utiliza para referirse a capas geológicas individuales, a formaciones geológicas completas e incluso a grupos de formaciones geológicas.

Es necesario definir el significado de acuífero, ya que la zona donde se realiza esta investigación es sobre un acuífero.

Un acuífero es una unidad geológica saturada y permeable que puede transmitir y almacenar cantidades significativas de agua bajo gradientes hidráulicos ordinarios, y dan agua a los pozos (Freeze & Cherry, 1979).

Otra definición de acuífero por (Todd, David; Mays, 2005) es una formación que contiene suficientes materiales permeables que permite ceder cantidades de agua a pozos y manantiales.

Los acuíferos se clasifican en (Ordoñez 2011).

Acuíferos libres: Son aquellos en los que el nivel de agua se encuentra por debajo del techo de la formación permeable. Liberan agua por desaturación, es decir, el agua que ceden es la procedente del drenaje de sus poros.

Acuíferos confinados: Son aquellos cubiertos por una capa impermeable confinante. El nivel de agua en los acuíferos cautivos está por encima del techo de la formación acuífera. El agua que ceden procede de la expansión del agua y de la descompresión de la estructura permeable vertical, cuando se produce la depresión en el acuífero. También se les denomina acuíferos

cautivos.

Acuíferos semiconfinados: Se pueden considerar un caso particular de los acuíferos cautivos, en los que muro, techo o ambos no son totalmente impermeables, sino que permiten una circulación vertical del agua (Figura 2).



Figura 2. Esquema de los tipos de acuíferos. (Tomado de Ordoñez Gáldez, 2011).

La Recarga de los Acuíferos

La recarga de un acuífero puede darse naturalmente o artificialmente. De manera natural es debido a la precipitación y a las aguas superficiales (ríos y lagos), o por medio de transferencias desde otras unidades hidrogeológicas o acuíferos. Y de manera artificial es producto de actividades como la irrigación, fugas de redes de abastecimiento o por infiltraciones de embalses y depósitos, actualmente el método más utilizado para medir esta recarga en los acuíferos es mediante un balance hídrico (Vélez & Vásquez, 2004).

Agua subterránea

El agua subterránea es agua que se encuentra debajo de la superficie del suelo en zonas porosas de los materiales donde todas las grietas y espacios porosos están llenos de agua. Dicha agua tiene una presión superior a la presión atmosférica, esta zona es conocida por varios nombres

como zona de saturación, zona saturada, zona de agua freática o zona de aguas subterráneas(Woessner & Poeter, 2020).

Para ilustrar lo anterior Woessner & Poeter utilizan el siguiente esquema (Figura 3) donde las zonas de aguas poco profundas en el subsuelo (incluida la zona saturada) está representado por el área punteada de color marrón violeta, el nivel freático es el que está debajo de la línea discontinua gruesa con un triángulo azul invertido, la zona vadosa es la que se encuentra arriba de marrón claro y por último la franja capilar esta entre ellos.

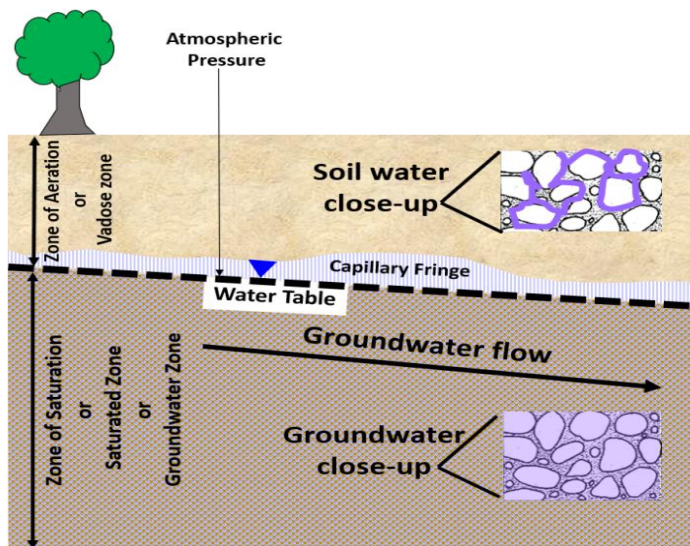


Figura 3. Esquema del agua subterránea (Tomado de Woessner, William W. and Eileen P. Poeter 2020).

El agua en los sistemas de agua subterránea siempre está en movimiento, físicamente, fluye desde las áreas de recarga (áreas de carga hidráulica alta) hacia las áreas de descarga (áreas de carga hidráulica baja), (Woessner, 2020).

Química del Agua

La mayoría del agua que se infiltra a través del suelo en los sistemas de flujo proviene de la lluvia o nieve derretida. Las zonas del suelo tienen características únicas que alteran la química del agua en el momento que ocurre la infiltración, además esta no es la única manera de que el agua subterránea se modifique, a medida que el agua subterránea se mueve por las líneas de flujo de las áreas de recarga hasta las de descarga la química se va alterando por los diferentes procesos geoquímicos. Con esto se sabe que la zona del suelo tiene una influencia en la química

del agua que se infiltra a través de él. Esto debido a que el suelo tiene la capacidad de generar cantidades relativamente grandes de ácido y consumir gran parte o todo el oxígeno disuelto disponible en el agua que se infiltra (Freeze & Cherry, 1979).

Los elementos químicos que pueden contener en las aguas subterráneas se muestran en la Tabla 1.

Constituyentes mayores de 5 mg/l	
Bicarbonato	Sodio
Calcio	Magnesio
Constituyentes menores entre 0.01-10 mg/l	
Boro	Nitrato
Hierro	Potasio
Constituyentes menores de 0.1 mg/l	
Aluminio	Selenio
Arsénico	Plata
Bromo	Zinc

Tabla 1. Elementos Químicos del agua subterránea, tomado de Freeze & Cherry, 1979.

Según (Todd, David; Mays, 2005) gracias a la evaluación de las composiciones químicas e isotópicas naturales del agua subterránea puede servir para obtener inferencias de las reacciones que producen la química natural del agua y la recarga, movimiento, mezcla y descarga del agua subterránea, la química de los sistemas de flujo de agua subterránea natural se puede utilizar para determinar las reacciones geoquímicas, las trayectorias de flujo del agua subterránea, la mezcla del agua subterránea y las edades y tasas de flujo del agua subterránea.

A continuación, se observa en la Tabla 2, unos ejemplos de algunas muestras tomadas de agua subterránea y su registro.

Sitio	Temperatura	pH	Ca	Mg	Na	K	HCO ₃	SO ₄	Cl
<i>Acuífero de roca carbonatada</i>									
Manantiales Fríos	10	7.6	1.72	0.64	0.07	0.01	4.79	0.09	0.03
Manantiales Indios	26	7.4	1.15	0.92	0.18	0.03	3.97	0.16	0.10
<i>Acuífero de roca carbonatada con Mineral</i>									
Gran manantial de Lewiston	10.6	7.58	1.87	1.15	0.10	0.02	3.31	1.46	0.05
Manantial Caliente de Vanek	19.6	7.40	3.25	1.65	0.16	0.03	3.53	3.44	0.07
Manantial de Landusky	20.4	7.24	6.50	4.08	1.79	0.24	3.81	10.11	0.54
<i>Acuífero de roca granítica</i>									
Manantial Efímero	-	6.2	0.08	0.03	0.13	0.03	0.33	0.01	0.01
Manantial Perenne	-	6.8	0.26	0.08	0.26	0.04	0.90	0.03	0.03
Pozo poco Profundo	8.0	7.89	0.85	0.19	0.54	0.04	2.34	0.09	0.10

Tabla 2. Muestras de del agua subterránea, tomado de Todd & Mays, 2005.

Pozo

Un pozo es una excavación o túnel vertical que perfora la tierra, hasta una cierta profundidad (Figura 4), suficiente para alcanzar lo que se busca, sea una reserva de agua subterránea del nivel freático o fluidos como el petróleo (Ordoñez Gáldez, 2011).



Figura 4. Pozo de la Red COTAS (Tomada de COTAS-UAZ).

Es necesario entender la definición anterior ya que en los pozos es donde se recaba la información que se utilizará en esta investigación y así para realizar el experimento.

Movimiento de las aguas subterráneas

Para comprender el flujo de agua subterránea, es de ayuda hacerse la siguiente pregunta:

“¿De dónde viene el agua y hacia dónde va?”

El agua subterránea siempre se encuentra en movimiento, esta se mueve a través de todos los materiales saturados entre un área de recarga y descarga, esto puede incluir acuíferos de todo tipo. Esta fluye desde las áreas de recarga (áreas de carga hidráulica alta) hacia las áreas de descarga (áreas de carga hidráulica baja). El agua ingresa al sistema de aguas subterráneas en las áreas de recarga y sale del sistema en las áreas de descarga (Woessner & Poeter, 2020).

Existen propiedades y principios que influyen con el movimiento el agua, estos pueden ser geológicos, geofísicos e hidráulicos. Un ejemplo de lo anterior puede ser la conductividad hidráulica, que es una constante de proporcionalidad entre la tasa de flujo y el gradiente de energía que causa ese flujo de acuerdo con la ley de Darcy (Gupta, 2016; Todd, David; Mays, 2005).

La ley fundamental del movimiento de las aguas subterráneas fue descubierta por Henri Darcy en 1856. Hizo un experimento en una tubería llena de arena en ciertas condiciones simuladas por la figura 15. Llegó a la conclusión de que el caudal (Q) era proporcional al área de la sección transversal A, inversamente proporcional a la longitud L de la ruta de flujo del filtro de arena, y proporcional a la caída de carga ($h_1 - h_2$) (Gupta, 2016; Todd, David; Mays, 2005; Woessner & Poeter, 2020).

Esto proporcionó la famosa ecuación de Darcy:

$$Q = -\frac{\Delta h}{\Delta L} * A \dots \dots (1)$$

Donde:

$Q =$ caudal volumétrico (L^3/T)

$\Delta h =$ diferencia de carga hidráulica entre dos puntos de medición, $h_2 - h_1$, donde h_2 se dirige a una ubicación más allá de la ubicación de h_1 en la dirección de caudal (L).

$\Delta L =$ longitud a lo largo de la ruta de flujo entre ubicaciones (L)

A = *área de la sección transversal perpendicular a la dirección del flujo (L^2)*

Para la construcción de mapas de las vistas de un acuífero se consideran diferentes elementos como los gradientes, las direcciones de flujo, los flujos, las velocidades del agua subterránea y la elevación del nivel del agua en un pozo (carga hidráulica), pero en esta investigación se realizó solamente considerando las temperaturas de los acuíferos en distintos pozos y además de la química del agua.

Monitoreo de temperatura con levelloggers

Según estudios de Freeze & Cherry, 1979 para lograr el aprovechamiento adecuado de los recursos de aguas subterráneas, se base en tres fases. La primera es la etapa de exploración donde se utilizan técnicas geológicas y geofísicas del subsuelo para la búsqueda de acuíferos, la segunda de evaluación que abarca la medición de parámetros hidrogeológicos y el cálculo de los rendimientos de los acuíferos, y por último se encuentra la explotación o gestión, que debe abarca la selección de estrategias óptimas de desarrollo y una evaluación de las interacciones entre la explotación de las aguas subterráneas y el sistema hidrológico regional.

Sin embargo, casi todos los principales acuíferos en ya han sido localizados y están siendo aprovechados en cierta medida, por lo que la época de exploración ya ha concluido y estas tres fases se pueden colocar en una perspectiva histórica. En la actualidad estamos en una etapa en la que la evaluación y la gestión de los acuíferos conocidos es de mayor importancia.

El calor transportado por el agua subterránea puede servir como un indicador natural del flujo de agua subterránea. Al construir modelos de aguas subterráneas de cuencas a gran escala con datos hidrológicos limitados, las observaciones de temperatura pueden ser valiosas para la calibración del modelo. Una comprensión clara del patrón de distribución de la temperatura del agua subterránea a escala de cuenca es la clave para usar la temperatura como indicador (An et al., 2015).

Para esta investigación se recabo la información un instrumento llamado Levellogger 5 modelo 3001 de la marca Solinst este registra mediciones de temperatura y nivel de agua subterránea o superficial con una alta precisión. Combina un sensor de presión, un sensor de temperatura, una batería de litio de 10 años y un registrador de datos, todo sellado dentro de una carcasa de acero inoxidable de 22 mm x 160 mm con un recubrimiento resistente a la corrosión a base de una tecnología de polimerización al horno. Este cuenta con un sensor detector de platino para

temperatura de resistencia (RTD), el rango de presión del sensor es de $\pm 0.05^{\circ}\text{C}$, la vida útil de la pila es de 10 años, opera entre una temperatura de -20°C a 80°C , la precisión del reloj es de ± 1 minuto / año (-20°C a 80°C). (Solinst Canada Ltd. 35 Todd Road, Georgetown, Ontario Canadá L7G 4R8 www.solinst.com).

En la Figura 5 se muestra el levellogger y el lector óptico de este mismo.



Figura 5. Sensor levellogger y lector óptico marca Solinst. Tomado de la página web de Solinst www.solinst.com.

Teoría de sistemas de flujo

Toth (1963) proporcionó alguna base para conceptualizar los sistemas de flujo de agua subterránea. Su trabajo mostró que múltiples sistemas de flujo no confinado pueden ocurrir en paisajes con suficiente recarga. Empareja las áreas de recarga y descarga para definir los sistemas de flujo.

Describió tres niveles de sistemas de flujo: local, intermedio y regional, como se muestra en la Figura 1.

Los sistemas de flujo local se forman entre áreas adyacentes de recarga y descarga, y las rutas de flujo del sistema local son relativamente cortas. Los sistemas de flujo intermedio se originan en áreas de recarga y descargan gradiente abajo, pero no al lugar de descarga más cercano. Los sistemas de flujo intermedio abarcan al menos un sistema de flujo local. Los sistemas de flujo regional se originan en áreas de recarga regionales y fluyen hacia ubicaciones de descarga distantes, con trayectorias de flujo relativamente largas. El sistema regional a menudo rodea uno o más sistemas de flujo locales y/o intermedios (Woessner & Poeter, 2020).

La clasificación para los flujos por temperatura esta dado de la siguiente manera:

- Menores de 18°C flujos locales.
- Temperaturas del orden de 18-30 °C flujo intermedios.
- Mayores de 30 °C flujos regionales.

Inteligencia Artificial

De acuerdo con Ertel (2017) han definido la inteligencia artificial (IA) de distintas maneras; por ejemplo, en el año de 1955 uno de los pioneros de la inteligencia artificial (IA) conocido como John McCarthy definía por primera vez el término de la inteligencia artificial como “El objetivo de la IA es desarrollar maquinas que se comporten como si fueran inteligentes”. La enciclopedia británica define la IA como: “La IA es la capacidad de las computadoras digitales o robots controlados por computadora para resolver problemas que normalmente se asocian con las capacidades de procesamiento intelectual más altas de los humanos. Este dilema fue resuelto por Elaine Rich en 1986 quien describió de la siguiente manera “La IA es el estudio de cómo hacer que las computadoras hagan cosas en las que, al momento, la gente es mejor.

Desde otra perspectiva los expertos en el campo de la IA usan el termino Artificial General Intelligence (AGI) para referirse a maquinas con nivel humano o superior de inteligencia. Aunque IA se refiere a un conjunto más amplio de técnicas computacionales, la mayoría de las soluciones modernas exitosas están impulsadas por el uso del aprendizaje automático y sus algoritmos(Yao et al., 2018).

Machine Learning

El aprendizaje, como la inteligencia, cubre una gama tan amplia de procesos que es difícil de definir con precisión. Una definición de diccionario incluye frases como “obtener conocimiento, comprensión o habilidad mediante el estudio, la instrucción o la experiencia”. Las técnicas en el aprendizaje automático (ML) se derivan de los estudios de los psicólogos para hacer más precisas sus teorías del aprendizaje en animales y humano a través de modelos computacionales y así pueden iluminar ciertos aspectos del aprendizaje biológico (Nilsson, 2005).

Si se considera que la capacidad de aprendizaje de la computadora es especialmente inferior a la de los humanos se infiere que los mecanismos de aprendizaje y el desarrollo de algoritmos de aprendizaje automático (ML) son una de las ramas más importantes de la IA(Ertel, 2017).

Según Kim (2017) una definición breve de Machine Learning (ML) es que es una técnica de modelado que involucra datos. ML es una técnica que descubre el “modelo” a partir de “datos”, los datos son información como documentos, imágenes, entre otros y el modelo es el producto final de ML.

El aprendizaje automático (ML) es la práctica de programar computadoras para que aprendan a través de los datos que se le proporcionan, a estos datos se les conoce como conjuntos de entrenamiento, ejemplos o experiencias pasadas (Alpaydin, 2014; Russell, 2018).

Según Nilsson (2005) se podría decir que una máquina aprende cada vez que cambia su estructura, programa o datos (basado en sus entradas o en respuesta a información externa) de tal manera que su el rendimiento mejora. Algunos de estos cambios, como la adición de un registro a una base de datos, caen cómodamente dentro de la provincia de otras disciplinas y son no necesariamente mejor entendido por ser llamado aprendizaje.

Tipos de Machine Learning

Se han desarrollado muchas técnicas distintas de aprendizaje automático para resolver diferentes problemas en diversos campos. Estas técnicas de ML pueden ser clasificados en dos tipos según el método de entrenamiento que se vaya a realizar y estos son:

- Aprendizaje supervisado
- Aprendizaje no supervisado

Aprendizaje Supervisado

El aprendizaje supervisado es muy similar al proceso en el que los humanos aprendemos cosas, se considera que los humanos obtenemos los conocimientos nuevos a medida que resolvemos problemas o ejercicios. En el aprendizaje supervisado cada conjunto de datos de entrenamiento debe contener pares de entradas y salidas, la salida es lo que debe obtener el modelo. El aprendizaje para el modelo supervisado consta de una serie de revisiones de un modelo para reducir la diferencia entre la salida correcta y la salida del modelo por la misma entrada (P. Kim, 2017).

Aprendizaje No Supervisado

Por el contrario, los datos de entrenamiento del aprendizaje no supervisado contienen solo entradas sin salidas correctas, el objetivo es encontrar las regularidades en la entrada. Este tipo de aprendizaje son aquellos en los que se interesan aumentar el conocimiento estructural de los datos disponibles, esto para observar que ciertos patrones ocurren con más frecuencia y que otros no, en general queremos ver qué sucede y qué no. En estadística, esto se conoce como estimación de densidad (Alpaydin, 2014; P. Kim, 2017).

Clustering

Un método para la estimación de la densidad es el agrupamiento o clustering, donde el objetivo es encontrar grupos o agrupaciones de entrada. Es decir que el conjunto de información contenga atributos similares aun grupo.(Alpaydin, 2014).

Una aplicación de la agrupación o clustering es la cuantización vectorial, que también se utiliza para para el preprocesamiento antes de una etapa posterior de clasificación o regresión(Alpaydin, 2014).

La mayoría de los métodos de aprendizaje no supervisados utilizan una medida de similitud entre patrones para agruparlos en grupos y otros utilizan una partición de grupos por probabilidades (Nilsson, 2005).

Dicho lo anterior el caso de este estudio se pretende clasificar los datos con los que se cuentan, por lo que indica la literatura se deberá aplicar un aprendizaje no supervisado ya que el conjunto de datos de entrenamiento consta de entradas solamente entradas y no hay salidas correctas.

Clustering por Distancia

Uno de los métodos de aprendizaje no supervisados más simple es definir una medida distancia entre patrones, estos patrones son de característica numéricas, la medida de la distancia puede ser la distancia euclidiana ordinaria entre dos puntos en un espacio n-dimensional. El método de agrupamiento basado en la distancia es iterativo, puede describirse como se muestra a continuación: se supone que se tiene R buscadores de grupos elegidos al azar, estos son puntos en un espacio n-dimensional que queremos ajustar para que cada uno se mueva hacia el centro de uno de los grupos de patrones. Los patrones se mostrarán en el conjunto de entrenamiento y encontramos que los buscadores de clústeres van acercarlo hacia él (Nilsson, 2005).

Clustering por Probabilidad

Como su nombre lo dice es un método de agrupación por probabilidad este consiste en tener una partición del conjunto de entrenamiento, en R mutuamente excluyentes y grupos exhaustivos, estos se agruparán a la probabilidad o a un umbral, se puede decidir a cuál de estos clústeres debe asignarse algún patrón arbitrario, para el cual la probabilidad es mayor que algún valor fijo en un umbral. Para utilizar este método se puede utilizar la regla de Bayes(Nilsson, 2005).

K-means

El algoritmo k-means son algoritmos de clasificación no supervisada que agrupa objetos en k-grupos basándose en sus características, es decir, para encontrar grupos que tengan datos similares, donde los grupos están representados por sus centros, que son los representantes típicos de los grupos. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo(Alpaydin, 2014).

La aplicación del algoritmo k-means según Bishop (2006) y Vishwanathan (2014) parte del problema de identificar grupos, o conglomerados, de puntos de datos en un espacio multidimensional. Se supone que se tiene un conjunto de datos $\{x_1, \dots, x_N\}$ que consiste de N observaciones de una variable euclidiana aleatoria x de D -dimensional. La meta es dividir el conjunto de datos en algún número K de grupos, donde supondremos por el momento en que se da el valor de K .

Se piensa en un grupo que comprende un grupo de puntos de datos cuyas distancias entre puntos son pequeñas en comparación con las distancias a puntos fuera del clúster. Podemos formalizar esta noción introduciendo primero un conjunto de vectores D -dimensionales μ_k , donde $k = 1, \dots, K$, en los que μ_k es un prototipo asociado con el grupo k -ésimo.

Podemos decir que μ_k es la representación de los centros de los grupos. Nuestro objetivo es entonces encontrar una asignación de puntos de datos a grupos, así como un conjunto de vectores $\{\mu_k\}$, tal que la suma de los cuadrados de las distancias de cada punto de datos a su vector μ_k más cercano, es un mínimo. En este punto se define la notación para describir la asignación de puntos de datos a clústeres. Para cada punto de datos x_n , introducimos un conjunto correspondiente de variables indicadoras binarias $r_{nk} \in \{0, 1\}$, donde $k = 1, \dots, K$ que describe cuál de los K clústeres a los que se asigna el punto de datos x_n , de modo que si se asigna el punto de datos x_n a cluster k entonces $r_{nk} = 1$, y $r_{nj} = 0$ para $j \neq k$.

Esto se conoce como un esquema de codificación 1 de K. Entonces se puede definir una función objetivo dada por:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \dots \dots \dots (2)$$

Esto representa la suma de los cuadrados de las distancias de cada punto de datos a su vector asignado μ_k .

El objetivo es encontrar valores para $\{r_{nk}\}$ y $\{\mu_k\}$ para minimizar J. Esto se logra a través de un procedimiento iterativo en el que cada iteración implica dos pasos sucesivos correspondientes a optimizaciones sucesivas con respecto al r_{nk} y al μ_k .

Primero elegimos algunos valores iniciales para el μ_k . Luego en la primera fase minimizamos J con respecto al r_{nk} , manteniendo el μ_k fijo. En la segunda fase minimizamos J con respecto a μ_k , manteniendo r_{nk} fijo.

Estas dos etapas de optimización después se van a ir repitiendo hasta la convergencia.

Considere primero la determinación del rango. Como J en la ecuación 1 es una función lineal de r_{nk} , esta optimización se puede realizar fácilmente para dar una solución de forma cerrada. Los términos que involucran diferentes n son independientes, por lo que podemos optimizar para cada uno n por separado eligiendo que r_{nk} sea 1 para cualquier valor de k que dé el valor mínimo de $\|x_n - \mu_k\|^2$. En otras palabras, simplemente asignamos el punto de datos n al centro de conglomerado más cercano. Más formalmente, esto se puede expresar como:

$$r_{nk} = \begin{cases} 1 & \text{si } k = \arg \min_j \|x_n - \mu_k\|^2 \\ 0 & \text{de otra manera} \end{cases} \dots \dots \dots (3)$$

Ahora considere la optimización de μ_k con r_{nk} fijo. El objetivo la función J es una función cuadrática de μ_k , y se puede minimizar estableciendo su derivada con respecto a μ_k a cero dando:

$$2 \sum_{n=1}^N r_{nk} (x_n - \mu_k) = 0 \dots \dots \dots (4)$$

que podemos resolver fácilmente para μ_k para dar:

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \dots \dots \dots (5)$$

El denominador en esta expresión es igual al número de puntos asignados al grupo k, por lo que este resultado tiene una interpretación simple, a saber, establecer μ_k igual a la media de todos los puntos de datos x_n asignados al conglomerado k. Por esta razón, el procedimiento se conoce como el algoritmo de K-means.

Las dos fases de reasignación de puntos de datos a los clusters y volver a calcular las medias de los clusters se repiten sucesivamente hasta que no haya más cambios en las asignaciones (o hasta que se exceda un número máximo de iteraciones). Porque cada fase reduce el valor de la función objetivo J, la convergencia del algoritmo está asegurada. Sin embargo, puede converger a un mínimo local en lugar de global de J. (Christopher M. Bishop, 2006; Shalev-Shwartz & Ben-David, 2013; Vishwanathan, 2014)

Resumiendo lo anterior y explicado en palabras más sencillas el algoritmo consta básicamente de tres pasos:

1. **Inicialización:** se establecen k-centroides en el espacio de los datos, esto se puede realizar de manera aleatoria.
2. **Asignación de los objetos a los centroides:** cada objeto de la información es asignado o agrupado a su centroide más cercano.
3. **Actualización de los centroides:** Se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se muevan, o se muevan por debajo de una distancia con cierto umbral.

El algoritmo k-means resuelve un problema de optimización, siendo la función a optimizar (minimizar) la suma de las distancias cuadráticas de cada objeto al centroide de su grupo o cluster. Es común que para este algoritmo se use la distancia cuadrática. (Wishart, 2014)

En la figura 4 se muestra como es el funcionamiento del algoritmo k-means, desde su inicialización hasta como se llega a la separación final de los grupos.

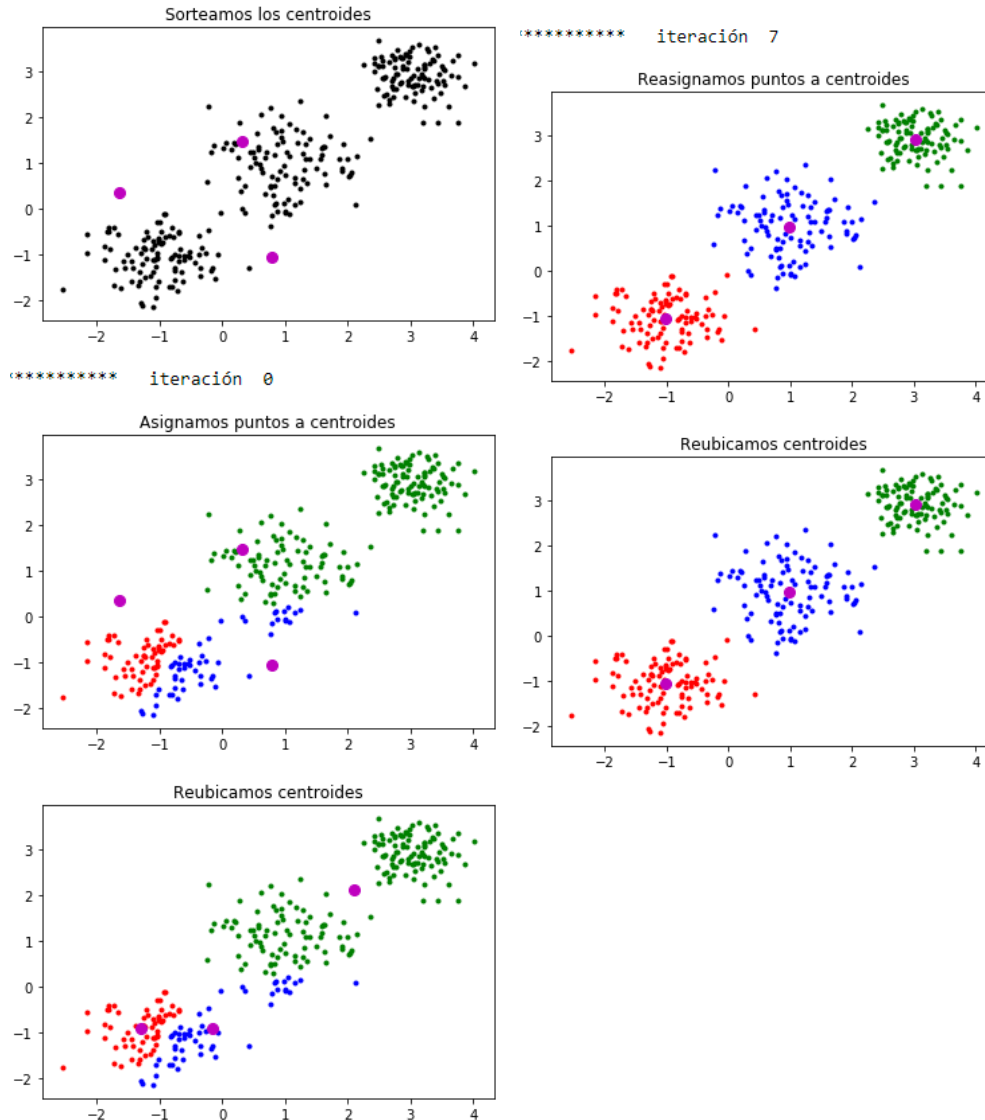


Figura 6. Funcionamiento del algoritmo k-means. Tomado de https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html

Elección del número de clústeres

Como cualquier método de aprendizaje, el agrupamiento también tiene su manera para ajustar la complejidad; y esta es k , el número de grupos. Dado cualquier k , el agrupamiento siempre encontrará k centros, ya sea que realmente sean grupos significativos o que sean impuestos por el método que usamos (Alpaydin, 2014).

Hay varias formas que podemos usar para ajustar k :

- En algunas aplicaciones, como la cuantificación del color, la aplicación define k .
- La representación gráfica de los datos en dos dimensiones mediante PCA se puede

utilizar para descubrir la estructura de los datos y el número de grupos en los datos.

- Un enfoque incremental también puede ayudar: establecer una distancia máxima permitida es equivalente a establecer un error de reconstrucción máximo permitido por instancia.
- En algunas aplicaciones, la validación de los grupos se puede realizar manualmente comprobando si los clústeres realmente codifican grupos de datos significativos.

Según el tipo de método de agrupación que utilicemos, podemos trazar el error de reconstrucción o la probabilidad de registro como una función de k y buscar el "codo". Después de una k lo suficientemente grande, el algoritmo comenzará a dividir grupos, en cuyo caso no habrá una gran disminución en el error de reconstrucción ni un gran aumento en la probabilidad logarítmica.

Escalado de los datos

En los métodos basados en la distancia, es importante escalar los componentes de los vectores de patrón. La variación de valores a lo largo de algunas dimensiones del patrón del vector puede ser muy diferente al de otras dimensiones y generar que la clasificación de los datos se agrupe en base a solo una característica y no todo el conjunto. Una técnica comúnmente utilizada es la normalización de los valores de los componentes y otra técnica sería la llamada z-score(Nilsson, 2005).

Z-score

Las distribuciones estandarizadas son de una sola unidad de medida y las puntuaciones brutas se transforman a dicha unidad de medida estandarizada para ser así comparada entre sí, a pesar de la transformación estas deberían de observarse como la distribución original y la única diferencia es que las puntuaciones se han colocado en una unidad diferente de medición (Lehmann, 2020).

Los Z-score son los puntajes estandarizados más comunes y estos son usados para la descripción de la ubicación de la puntuación en una distribución (estadísticas descriptivas) y así podemos observar la posición relativa de una puntuación en una muestra o una muestra en una población. La ecuación está dada por:

$$Z = \frac{(X - \mu)}{\sigma} \dots \dots \dots (6)$$

Donde:

$Z = z - score$

$X = Variable\ que\ se\ va\ a\ convertir$

$\mu = Media\ de\ la\ distribución\ original,$

$\sigma = Desviación\ estándar\ de\ la\ distribución\ original.$

En un z-score, la media se coloca en 0 y cada número por debajo o por encima es una representación de cuántas desviaciones estándar se alejan de una puntuación. Para los propósitos de su clase, todas las distribuciones de z-score son distribuciones normales. No se utilizan z-score en otros tipos de distribuciones porque las gráficas y proporciones están diseñados para describir distribuciones normales (Lehmann, 2020).

Limitaciones k-means

Según Bishop (2006) y Vishwanathan (2014) el algoritmo k-means presenta algunas limitaciones y son las siguientes:

- Es necesario definir el número de clústeres y esta decisión puede afectar seriamente los resultados.
- La ubicación de los centroides iniciales es aleatoria, los resultados pueden no ser comparables y mostrar una falta de consistencia.
- Es sensible a outliers y esto hace que los casos extremos hacen que el clúster se vea afectado. Aunque esto puede ser algo positivo a la hora de detectar anomalías.

Correlación

La palabra “correlación” es de origen latino tardío y significa “asociación”, “conexión”, “correspondencia”, “interdependencia”, “relación”, pero relación no en la forma funcional determinista convencional para ese momento. (Shevlyakov & Oja, 2016)

Francis Galton (1822-1911), antropólogo, biólogo, psicólogo y meteorólogo, entendió que correlación es la interrelación en promedio entre cualquier variable aleatoria, él menciona que:

“Dos variables están correlacionadas cuando la variación de una es acompañada en promedio por más o menos variación de la otra, y en la misma dirección”

Se puede llegar a la conclusión de que la correlación debe ser la consecuencia de que las variaciones de las dos variables se deban en parte a una causa común, si no, la correlación sería nula. (Shevlyakov & Oja, 2016)

Análisis de correlación

El análisis de correlación también fue recalcado por Galton, este se ocupa de la estimación del valor de correlación por índices numéricos o coeficientes. (Reimann et al., 2008; Shevlyakov & Oja, 2016)

El análisis de correlación estima el alcance de la relación entre cualquier par de variables. El la covarianza es una medida de esta relación y depende de la variabilidad de cada uno de los dos variables Debido a que las covarianzas pueden tomar cualquier número, solo el signo (+ o -) es informativo; sin embargo, no se puede interpretar la fuerza de la relación entre las dos variables. Para obtener números estandarizados, los coeficientes de correlación, es necesario eliminar la dependencia de la variabilidad de cada variable (Reimann et al., 2008)

Métodos de Correlación

Existen tres métodos ampliamente utilizados para calcular un coeficiente de correlación son Pearson, Spearman y correlación de Kendall. Estos métodos resultan en un número entre -1 y $+1$ que expresa cuán estrechamente están relacionadas las dos variables, ± 1 muestra una relación perfecta 1:1 (positiva o negativa) y 0 indica que no existe una relación sistemática entre las dos variables. Una correlación de ± 0.5 generalmente indica una significativa relación, pero esto depende del número de muestras. Con el aumento del número de muestras, los valores absolutos decrecientes de los coeficientes de correlación se vuelven significativos. (Mackridge & Rowe, 2018; Reimann et al., 2008)

El término “correlación” puede referirse a una amplia variedad de coeficientes de correlación;

sin embargo, el más popular es, el coeficiente de correlación producto-momento de Pearson o la r de Pearson para abreviar. Sin embargo, tenga cuidado al asociar automáticamente la palabra "correlación" con la r de Pearson, ya que otros coeficientes de correlación pueden reportar diferentes fuerzas de relación. La r de Pearson es la más apropiada para datos en los que ambas variables son continuas y desea medir la relación lineal entre ellos. (Denis, 2020)

Método Pearson

Karl Pearson (1857–1936), un matemático, estadístico, biólogo y filósofo británico, había escrito las fórmulas explícitas para el coeficiente de correlación producto-momento de la población (Pearson 1895):

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{[\text{var}(X) \text{var}(Y)]^{1/2}} \dots \dots \dots (7)$$

y su versión simplificada:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}} \dots \dots \dots (8)$$

La correlación producto-momento de Pearson, o r de Pearson para abreviar, es una medida de la relación lineal entre dos variables continuas (Denis, 2020)

La estandarización también tiene el efecto de colocar límites de -1 a $+1$ en esta nueva medida de

R de Pearson Una r de -1 significa que a medida que aumentan los valores de x , los valores de y disminuyen en una forma lineal perfecta. Por el contrario, una r de $+1$ significa que a medida que aumentan los valores de x , también lo hacen los valores de y en una forma lineal perfecta. Una r de Pearson igual a 0 indica la ausencia de una relación lineal entre las variables. Tenga mucho cuidado en notar que $r = 0$ no implica necesariamente la ausencia de una relación entre x e y , pero

sólo de uno lineal. Es totalmente posible obtener $r = 0$, pero todavía existe una relación que es algo más que lineal. (Denis, 2020; Reimann et al., 2008)

Limitaciones

El método “clásico” de Pearson es ideal si los datos siguen un modelo bivariado. distribución

normal (la función de densidad resultante tiene la forma de una campana tridimensional), una realidad improbable con datos ambientales o geoquímicos (Mackridge & Rowe, 2018). El método de Pearson es muy vulnerable a datos atípicos de cualquier tipo, no solo muy valores altos (o bajos) sino también desviaciones de la estructura principal de los datos. Por lo tanto, cuando se utiliza el método de Pearson, los datos geoquímicos generalmente deberán transformarse para aproximarse a la distribución normal bivariada requerida. (Reimann et al., 2008)

2.3.- Principales estudios relacionados

Los principales estudios relacionados con la investigación que han aplicado distintos algoritmos de inteligencia artificial (IA) son:

- Ran An, Xiao-Wei Jiang, Jun-Zhi Wang, Li Wan, Xu-Sheng Wang y Hailong Li en el año 2014, publicaron una investigación “A theoretical analysis of basin-scale groundwater temperature distribution” en el cual modelaban una cuenca con un sistema simple y complejo para demostrar como la temperatura del agua subterránea se redistribuye en estos dos tipos antes mencionados debido a la convección del agua.
- Daughney, C. J., Raiber, M., Moreau-Fournier, M., Morgenstern, U., & van der Raaij, R. en el año 2012, realizaron una investigación llamada “Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: Comparison of New Zealand’s national and regional groundwater monitoring programs” en el cual haciendo uso de HCA y la información de que contaba de diversas instituciones de Nueva Zelanda se podría monitorear la calidad del agua en las regiones de estudio.
- Renata Graf y Pouya Aghelpour en el año 2021, hicieron una investigación llamada “Daily River Water Temperature Prediction: A Comparison between Neural Network and Stochastic Techniques” en este estudio se compara dos tipos diferentes de modelos numéricos para predecir la TRW diaria en la cuenca del río Warta en Polonia.
- Ana D. Maldonado, María Morales, Francisco Navarro, Francisco Sánchez-Martos and Pedro A. Aguilera en el año 2021, publicaron una investigación “Modeling Semiarid River–Aquifer Systems with Bayesian Networks and Artificial Neural Networks” en el cual comparan los algoritmos de redes neuronales y redes bayesianas para determinar

cuál de estos son mejores para clasificar y predecir utilizando valores de temperatura en un acuífero.

- Subba Rao, N., Surya Rao, P., & Deva Varma, D. en el año del 2013 publicaron un artículo llamado “Spatial variations of groundwater vulnerability using cluster analysis” en el cual clasificaban la calidad del agua subterránea utilizando algoritmos HCA.
- Woessner, W. W., & Poeter, E. P. publicaron un libro en el año 2020, llamado “Hydrogeologic Properties of Earth Materials and Principles of Groundwater Flow” en el cual describen como funciona la teoría de Toth (1963), además de todos los principios básicos hidrológicos para entender e interpretar esta teoría.

2.4.- Contribuciones y limitaciones de estudios previos

La Tabla 3 Presenta el resumen de las investigaciones realizadas, con el enfoque de IA y el agua subterránea para entender su comportamiento.

Artículo	Autores y año	Contribuciones	Algoritmo	Limitaciones
A theoretical analysis of basin-scale groundwater temperature distribution	An, R., Jiang, X. W., Wang, J. Z., Wan, L., Wang, X. S., & Li, H. (2015)	El patrón de distribución de la temperatura del agua subterránea y el fuerte efecto del transporte de calor por convección revelado en este estudio también podrían ser valiosos para comprender las compensaciones entre el flujo de agua subterránea y el transporte de calor, y para usar el calor como indicador para identificar el patrón del flujo de agua subterránea en las cuencas.	Modela numéricamente resolviendo las ecuaciones bidimensionales (2D) acopladas de flujo de agua subterránea y transporte de calor en estado estacionario utilizando un software basado en el método de elementos finitos.	Uso de métodos tradicionales para el modelado de la cuenca.

Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: comparison of New Zealand's national and regional groundwater monitoring programs	Christopher J. Daughney, Matthias Raiber, Magali Moreau-Fournier, Uwe Morgenstern y Rob van der Raaij (2012)	El análisis de Hierarchical cluster (HCA) permite que las redes de monitoreo se comparen en términos de la cantidad de categorías de calidad del agua identificadas en cada red, la hidroquímica de la calidad del agua, las proporciones de los sitios de monitoreo asignados a cada categoría de calidad del agua y el rango de concentraciones dentro de cada categoría de calidad del agua.	Hierarchical cluster (HCA)	No se utilizó otro algoritmo de comparación.
Daily River Water Temperature Prediction: A Comparison between Neural Network and Stochastic Techniques	Renata Graf y Pouya Aghelpour (2021)	Una predicción para la temperatura diaria en el río se obtuvieron diferentes resultados al utilizar los modelos estocásticos y de IA, las comparaciones de los resultados de los modelos utilizados no han sido concluyentes.	Modelos estocásticos, sistema de inferencia adaptativo neuro-fuzzy (ANFIS), función de base radial (RBFNN).	El nivel de resolución de datos y la duración de la serie de observación.

Modeling Semiarid River–Aquifer Systems with Bayesian Networks and Artificial Neural Networks	Ana D. Maldonado, María Morales, Francisco Navarro, Francisco Sánchez-Martos and Pedro A. Aguilera (2021)	Los experimentos llevados a cabo en este artículo muestran que, en general, tanto las ANN como las BN se desempeñan bien como clasificadores, y las BN muestran mejores habilidades de desempeño para clasificar la clase positiva. El costo computacional asociado con la estimación de los modelos BN es significativamente menor y el modelo resultante es más versátil.	Redes Neuronales Artificiales (RNA) y Redes Bayesianas (BN).	Solo se consideran variables consideradas que cambian con el tiempo y además de esto solo se usaron 2 años de registro.
Spatial Variations of Groundwater Vulnerability using Cluster Analysis	N. Subba Rao, P. Surya Rao y D. Deva Varma (2013)	La calidad del agua subterránea se puede clasificar en dos grupos para observar la calidad de esta y a su vez se pueden clasificar en más subgrupos con ayuda de la química del agua para observar los componentes del agua subterránea con ayuda del análisis por cluster.	Hierarchical cluster (HCA)	Solo se utilizaron treinta muestras para la clasificación

Tabla 3. Contribuciones de estudios relacionados

2.5.- Comparación entre los trabajos relacionados y la propuesta de investigación

A continuación, en la Tabla 4, se presenta un comparativo de los estudios relacionados y la propuesta de investigación en donde se muestra los enfoques que esta investigación aportara.

Artículo	Variable temperatura	Química del agua	Algoritmo cluster	Años de registro
Pronóstico de la temperatura del agua subterránea utilizando machine learning				
A theoretical analysis of basin-scale groundwater temperature distribution				
Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: comparison of New Zealand's national and regional groundwater monitoring programs				
Daily River Water Temperature Prediction: A Comparison between Neural Network and Stochastic Techniques				
Modeling Semiarid River–Aquifer Systems with Bayesian Networks and Artificial Neural Networks				
Spatial Variations of Groundwater Vulnerability using Cluster Analysis				

Tabla 4. Contribuciones de estudios relacionados

En base a la tabla anterior se puede inferir que el desarrollo de la presente investigación difiere con los trabajos relacionados, estas variaciones se deben a los distintos usos de algoritmos de inteligencia artificial (IA), también a que ciertas investigaciones no utilizan la química del agua en conjunto con la temperatura y por último los registros tiempo usandos; estos son importantes ya que definen la muestra de aprendizaje para el modelo.

Además, también se está comparando con diferentes investigaciones que se enfocan en el modelado de las fluctuaciones espacio-temporales de las aguas subterráneas de un acuífero, aunque estos reportes no son igual al que se plantean en esta investigación de temperatura y química del agua son de gran ayuda ya que los modelos que se usan pueden modificarse para usarlos en esta investigación desarrollada.

III. Metodología

3.1.- Localización de la zona de estudio

El acuífero Calera, definido con la clave 3225 en el Sistema de Información Geográfica para el Manejo de Aguas Subterráneas (SIGMAS) de la CONAGUA, se localiza en la porción central del estado de Zacatecas; entre los paralelos 22°41' y 23° 24' de latitud norte y entre los meridianos 102°33' y 103° 01' de longitud oeste, cubriendo una superficie aproximada de 2,226 km².

Limita al norte con los acuíferos Aguanaval y Guadalupe de Las Corrientes, al noreste y este con Chupaderos, al sur con Benito Juárez, al suroeste con Jerez y al oeste también con Aguanaval; todos ellos pertenecientes al estado de Zacatecas (Figura 5). Geopolíticamente el área que cubre el acuífero comprende la totalidad de los municipios General Enrique Estrada y Morelos, la mayor parte de Calera, Frenillo y Zacatecas, así como porciones menores y pequeñas de Pánuco, Vetagrande, Jerez y Villa de Cos.

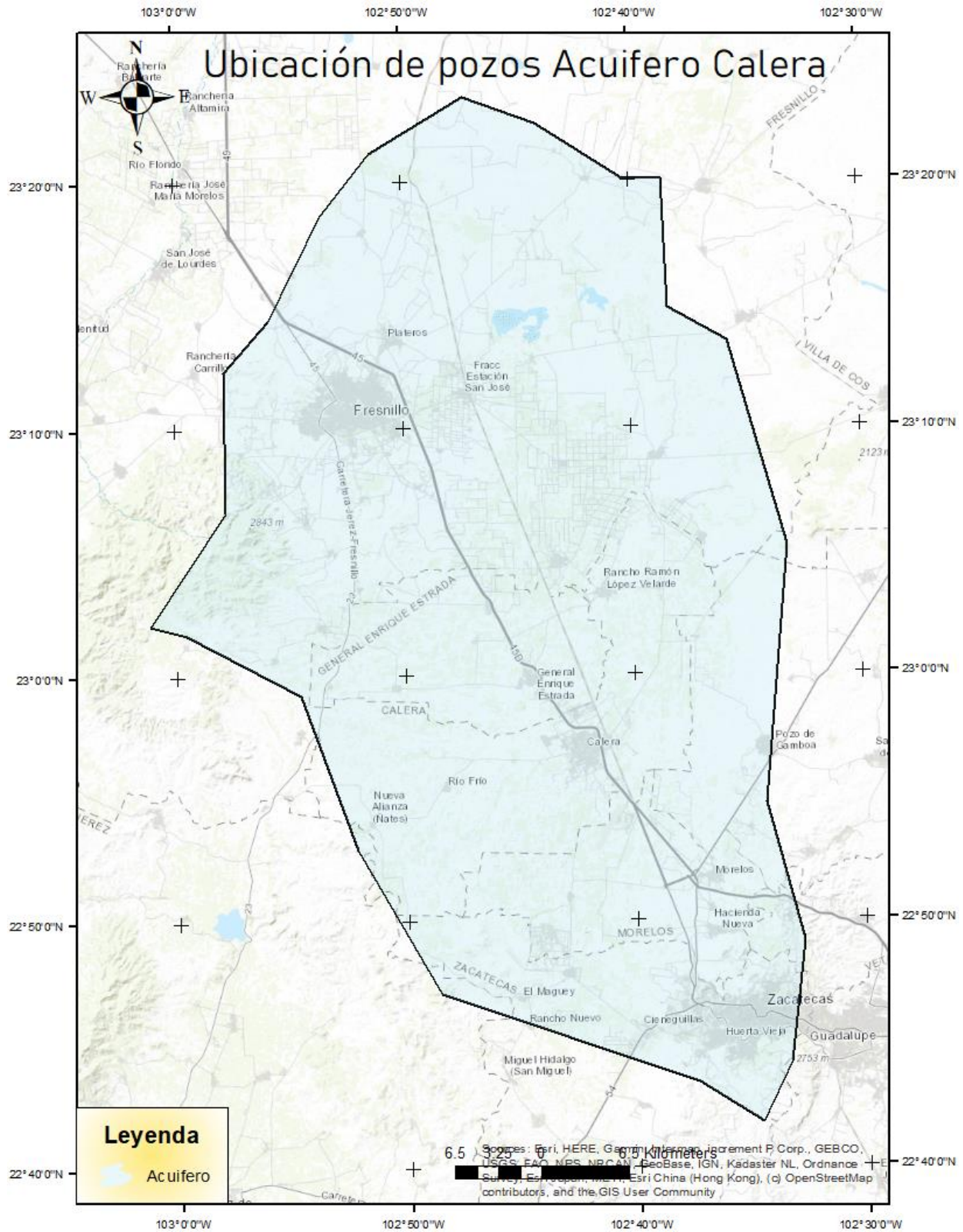


Figura 7. Localización del Acuífero.

3.2.- Modelo o esquema general de investigación

Los datos disponibles son de origen cuantitativo debido a que son mediciones específicas de una característica de un acuífero, siendo esta la temperatura y la química del agua, y se utilizaron con el fin de probar si se pueden clasificar los diferentes flujos antes mencionados en la sección 2.2 mediante Machine Learning (ML).

Los datos de carga química del agua medidos en el acuífero “Calera” se utilizaron en esta investigación para realizar el modelo y comparar con los registros de temperatura el desempeño del modelo

Con los datos disponibles de temperatura, se construirá una matriz de correlación para ver las posibles relaciones existentes entre los diferentes pozos de muestra que se utilizaron y si coinciden con la clasificación del modelo k-means.

De acuerdo a lo que se describe en Sampieri, Valencia, Torres, & Romo (2017) esta investigación tiene un diseño de tipo cuantitativa no experimental porque en este estudio no pueden ser manipuladas las variables dependientes, siendo esta la temperatura y las muestras de la química del agua subterránea que se tienen, solo es posible observar y analizar los registros de temperatura y química del agua proporcionados por los instrumentos Levellogger y las muestras tomadas en los pozos respectivamente, además este reporte es de tipo longitudinal ya que los datos que se estudiarán son de varios años, no solo de uno específico porque se busca ver los cambios en el acuífero y los datos de temperatura con el paso del tiempo.

3.3.- Modelo de investigación

La muestra consta de 17 aprovechamientos donde se tiene instalados los levelloggers y además de 15 estaciones climatológicas también, distribuidas en zonas cercanas del acuífero. La selección de la muestra es de tipo no probabilística ya que la selección de información está limitada a la disponibilidad de datos que se tengan de cada una de las mediciones de cada pozo y que se puedan acceder a ellas dentro del acuífero.

Para la recolección de datos, se visitaron los puntos de muestreo o bien la información que sea otorgada por el Comité Técnico de Aguas Subterráneas (COTAS), el levelloggers registra diariamente a las 5 A.M. los datos para calcular niveles piezométricos y temperatura. Este registro se puede obtener haciendo una visita de campo a cada uno de estos puntos de interés y por medio de un software se puede exportar dicha información.

La Figura (8) muestra los pasos que se seguirán para cumplir con los objetivos de esta investigación:

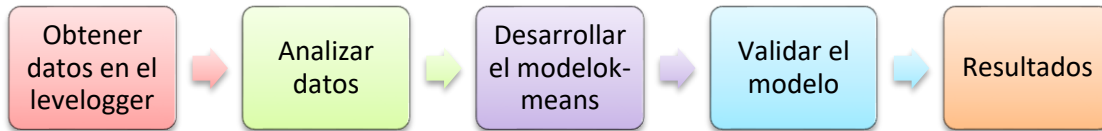


Figura 8. Modelo General de investigación

3.4.- Descripción de la propuesta

La información que se utilizaron para esta investigación fue tomado a través de la red de levelloggers que se tiene instalada en el acuífero “Calera” por COTAS; CONAGUA (2015), la red contiene 17 aprovechamientos con registros (2017-2021). Además, se cuenta con un segundo registro que contiene información de otros 99 pozos que tienen una medición única, es de la química del agua. La ubicación de los aprovechamientos monitoreados (Figura 9), además del comportamiento de estos.

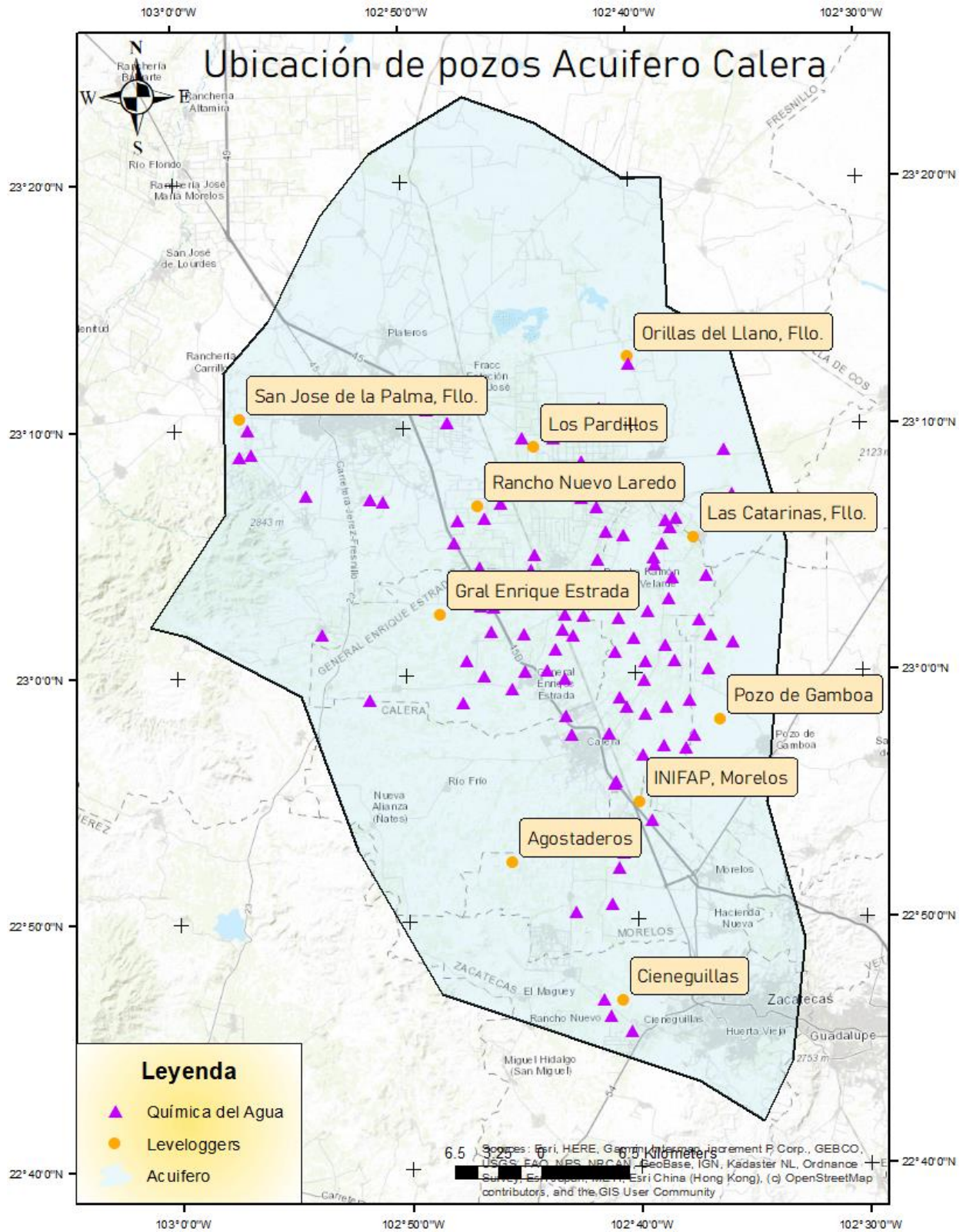


Figura 9 Levelloggers y Química del agua

3.5.- Desarrollo del modelo de clasificación

Se realizó un algoritmo k-means con $k=3$, para así poder clasificar los 3 flujos subterráneos (locales, intermedios y regionales), esta clasificación se realizó solo con los datos de las características del agua, además, se escalaron los datos, debido a la variabilidad de los datos, esto permite una agrupación considerando el mayor peso en la clasificación, la manera en la que se escalaron los datos fue usando un z-score. Para la inicialización de los centroides se hizo de manera aleatoria y se utilizó las distancias cuadráticas (Figura 10).

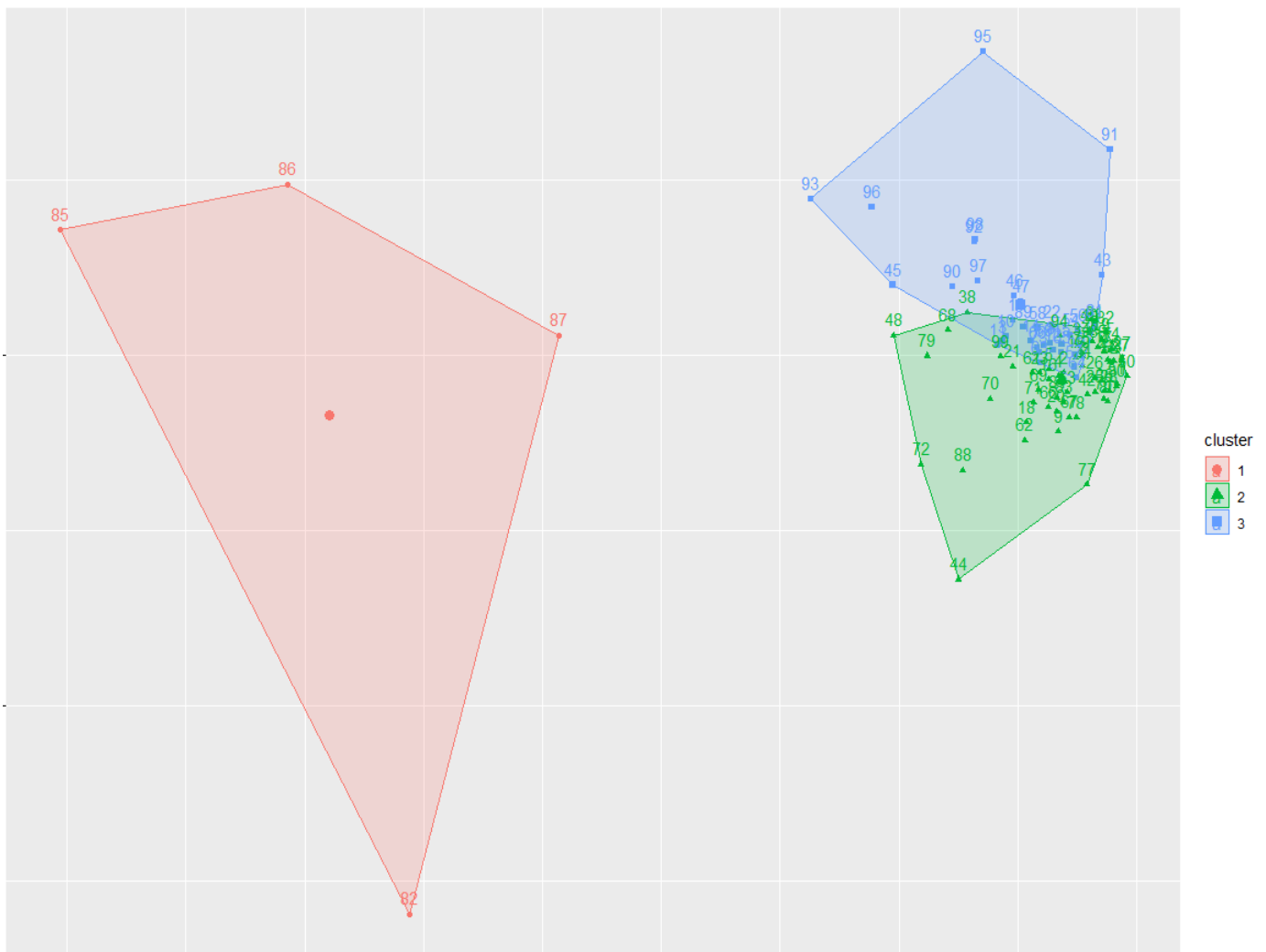


Figura 10 Agrupación K-means con R studio

Utilizando el software de R studio y su interfaz con los siguientes paquetes, funciones y comandos:

```

1 library(devtools)
2 devtools::install_github("kassambara/factoextra")
3 library(factoextra)
4 library(corrplot)
5 library(tidyverse)
6
7
8 #####
9 ##### Correlaciones entre pozos #####
10 #####
11
12 Estaciones <- read.csv("~/Maestría Procesamiento de la Información/Tesis/Base de datos/ESTACIONESNUEVAS.csv")
13 data<-Estaciones
14 data<-Estaciones[1:1640,2:11]
15 #data=na.omit(data)
16 data.pearson <- cor(data,method="pearson",use = "complete.obs")
17 corrplot(data.pearson, shade.col = NA, tl.col="black",tl.srt=45)
18 corrplot(data.pearson,method = "number" ,shade.col = NA, tl.col="black",tl.srt=45)
19
20 #####
21 ##### K-means #####
22 #####
23
24 set.seed(2015)
25
26 NuevasTemperaturas <- read.csv("~/Maestría Procesamiento de la Información/Tesis/Base_de_datos/NuevasTemperaturas.csv")
27 Basedatos=NuevasTemperaturas[,~c(1,2,3,9,10,18,52,53,54)]
28 CoordenasExtras=NuevasTemperaturas[,c(1,2,3)]
29 datos.scaled=as.data.frame(scale(Basedatos))
30
31 km=kmeans(datos.scaled,centers=3) #nstart = 10)
32
33 Grupos=km$cluster
34 fviz_cluster(km, data=datos.scaled)
35 km$totss
36 km$withinss
37 km$tot.withinss
38 km$betweenss
39
40 Basedatos[,46]=Grupos
41 CoordenasExtras[,4]=Grupos
42
43 data<-Basedatos
44 data.pearson <- cor(data,method="pearson",use = "complete.obs")
45 corrplot(data.pearson, shade.col = NA, tl.col="black",
46         tl.srt=45)
47
48 corrplot(data.pearson,method = "number" ,shade.col = NA, tl.col="black",
49         tl.srt=45)
50
51 #####
52 ##### Guardar Base de datos para arcgis #####
53 #####
54
55 setwd("D:/Procesamiento de datos maestria/Tesis/Base de datos")
56 write.csv(CoordenasExtras, file = "Basedatos.csv", row.names = FALSE)
57 write.csv(mc, file = "Pruebas.csv", row.names = FALSE)
58 getwd()

```

Figura 11 Código R studio

IV. Resultados y Discusiones

4.1.- Registro de temperaturas diarias con levelloggers

La matriz de correlación entre los datos de temperatura registrados en los levelloggers instalados en los 10 pozos del acuífero Calera (Figura 12).

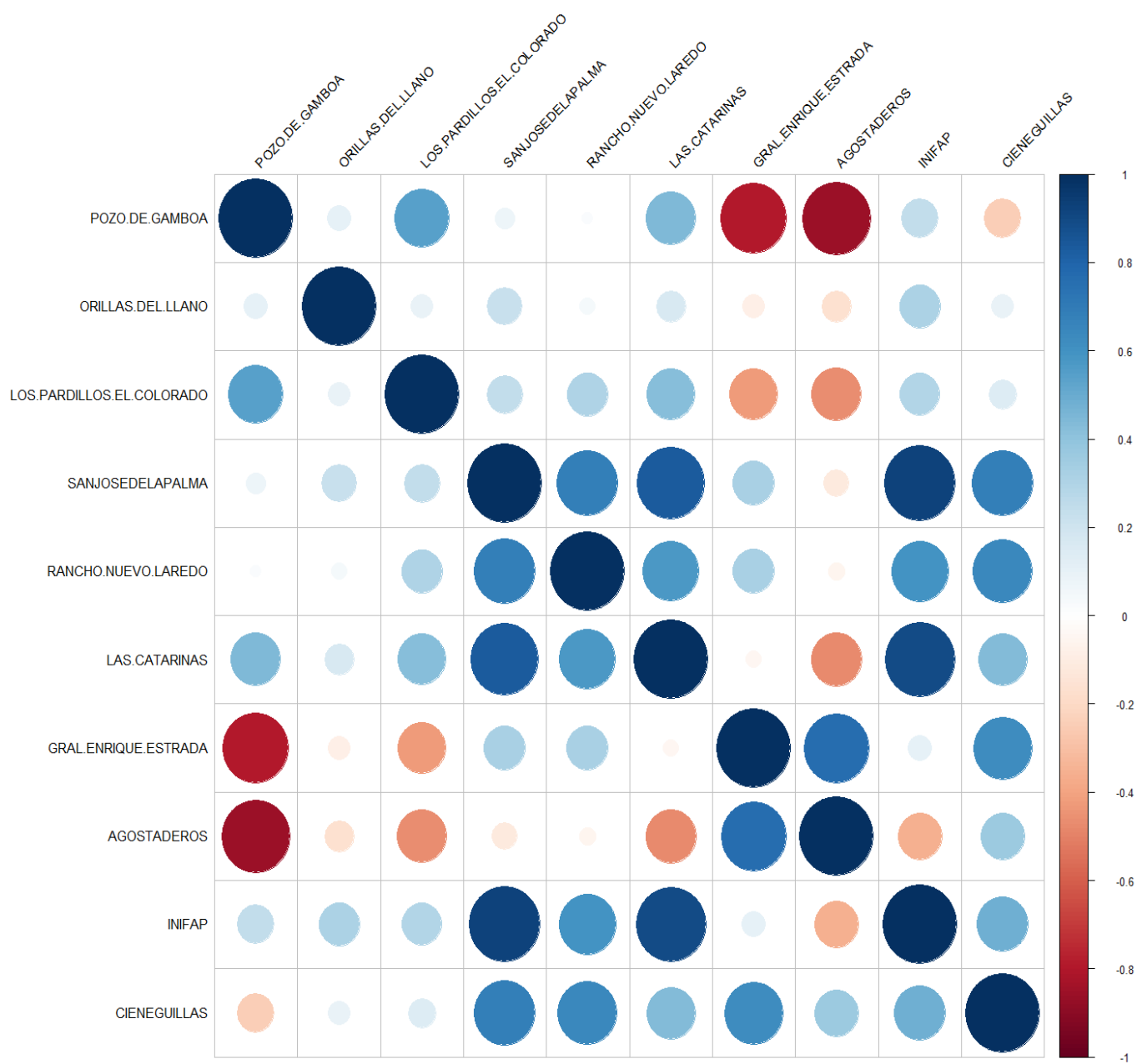


Figura 12 Matriz de correlación de los levelloggers

En la matriz de correlación se puede observar que existen una relación entre en los pozos de San José de la Palma, INIFAP y Las Catarinas, visualizando el mapa del acuífero (Figura 13), se tiene además que entre ellos existe una distancia aproximada de 41 km y 21 km respectivamente.

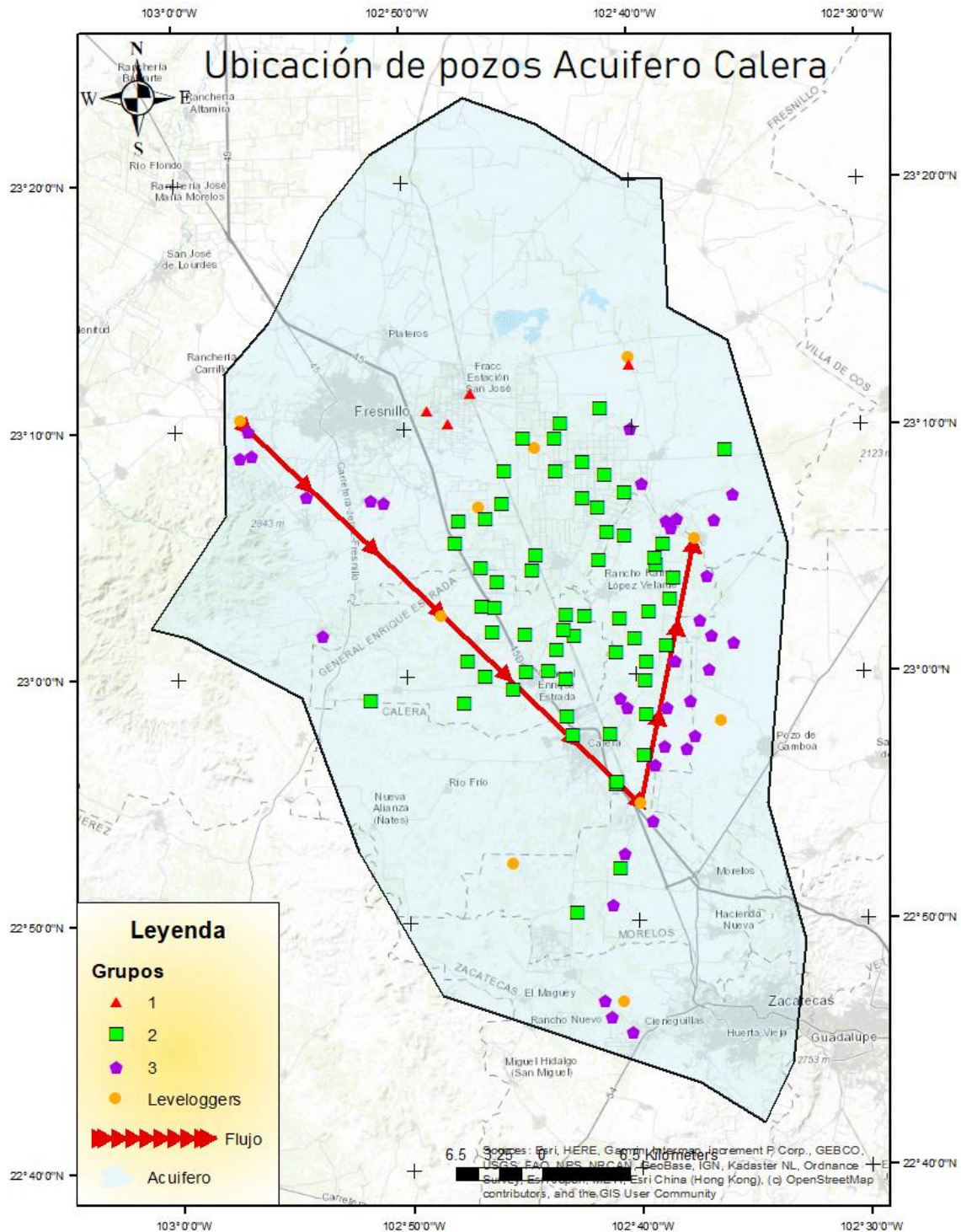


Figura 13 Mapa de Leveloggers relacionados

Por otro lado, también se realizó una matriz de correlación, entre los aprovechamientos que tienen las características físico-químicas del agua (Figura 13).

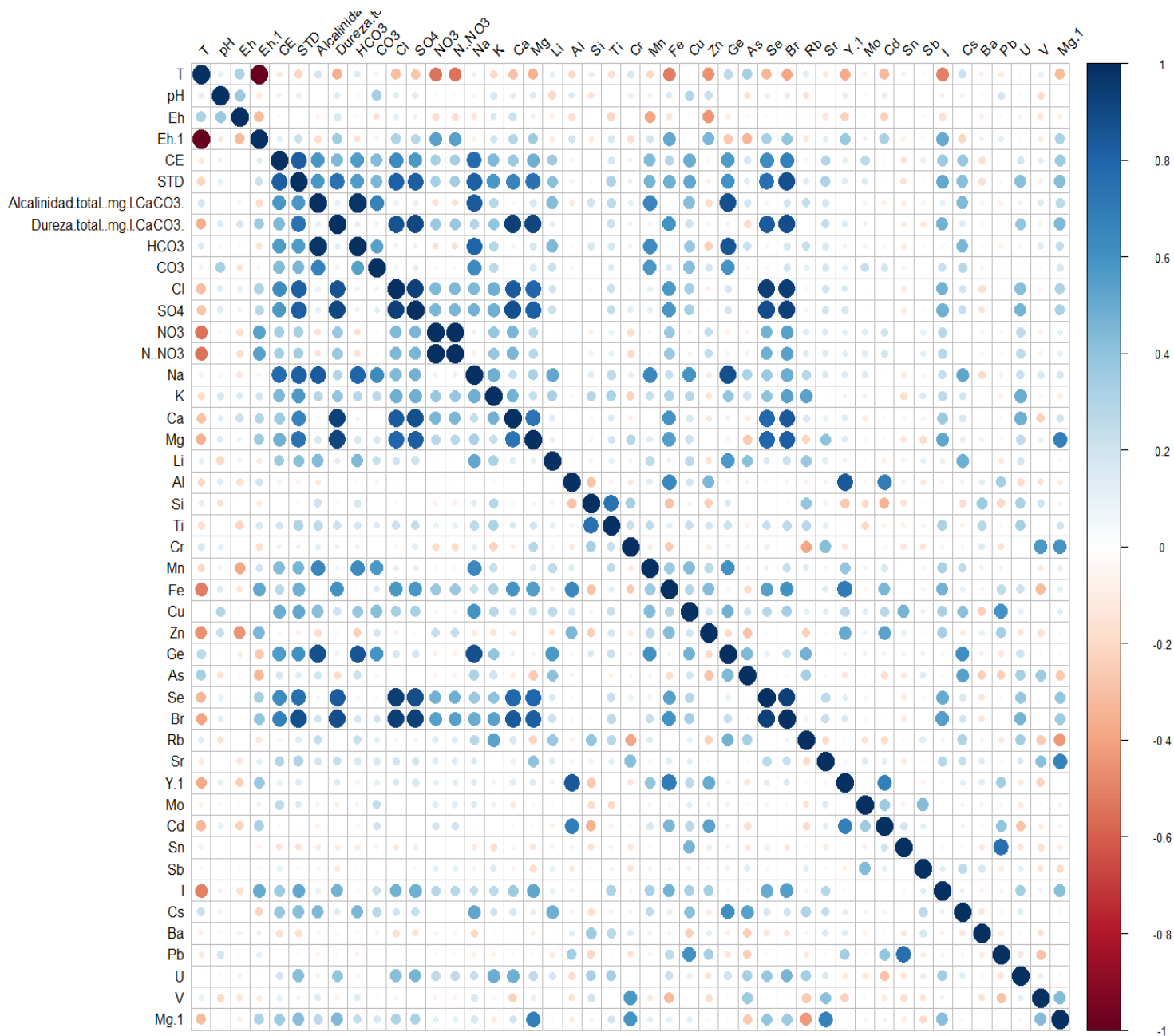


Figura 14 Matriz de correlación de la química del agua

En la matriz se puede observar que existen algunas características que son dependientes entre sí, como por ejemplo los parámetros químicos del Selenio (Se) y Bromo (Br).

4.2.- Modelo K-means con la Química del agua.

El modelo k-means logro agrupar la información en la Figura 10 mencionada en el apartado 3.5. Desarrollo del modelo de clasificación y ya con el modelo realizado se transfirió esta clasificación de manera espacial utilizando la herramienta ArcMap para realizar el análisis espacial (Figura 15).

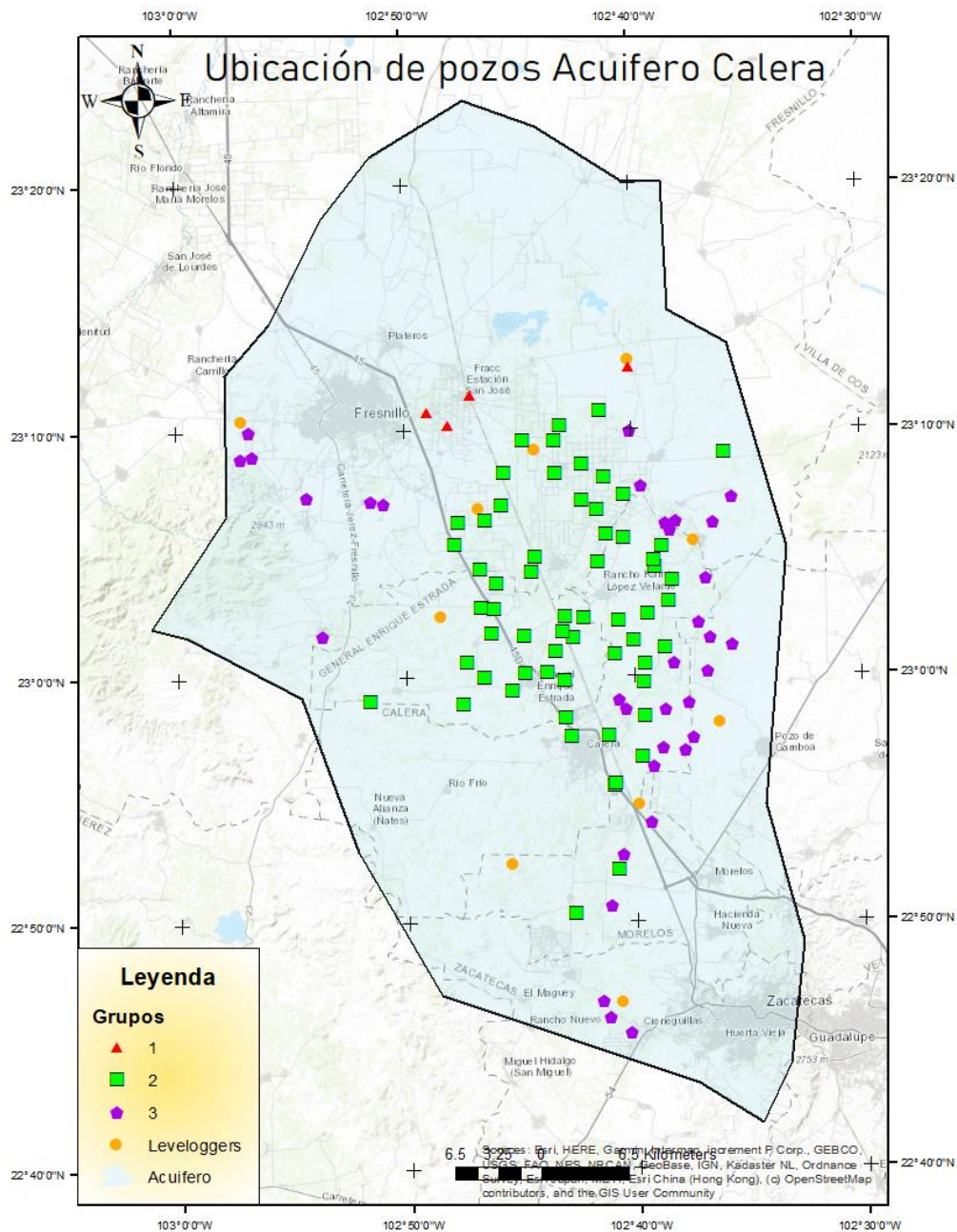


Figura 15 Mapa con el modelo K-means

De acuerdo a la Figura 15 se tiene un poco dispersión en los grupos de clasificación, con el apoyo de la herramienta de Google Earth Pro se visualizó la orografía de la región para evaluar posible efecto en la agrupación.

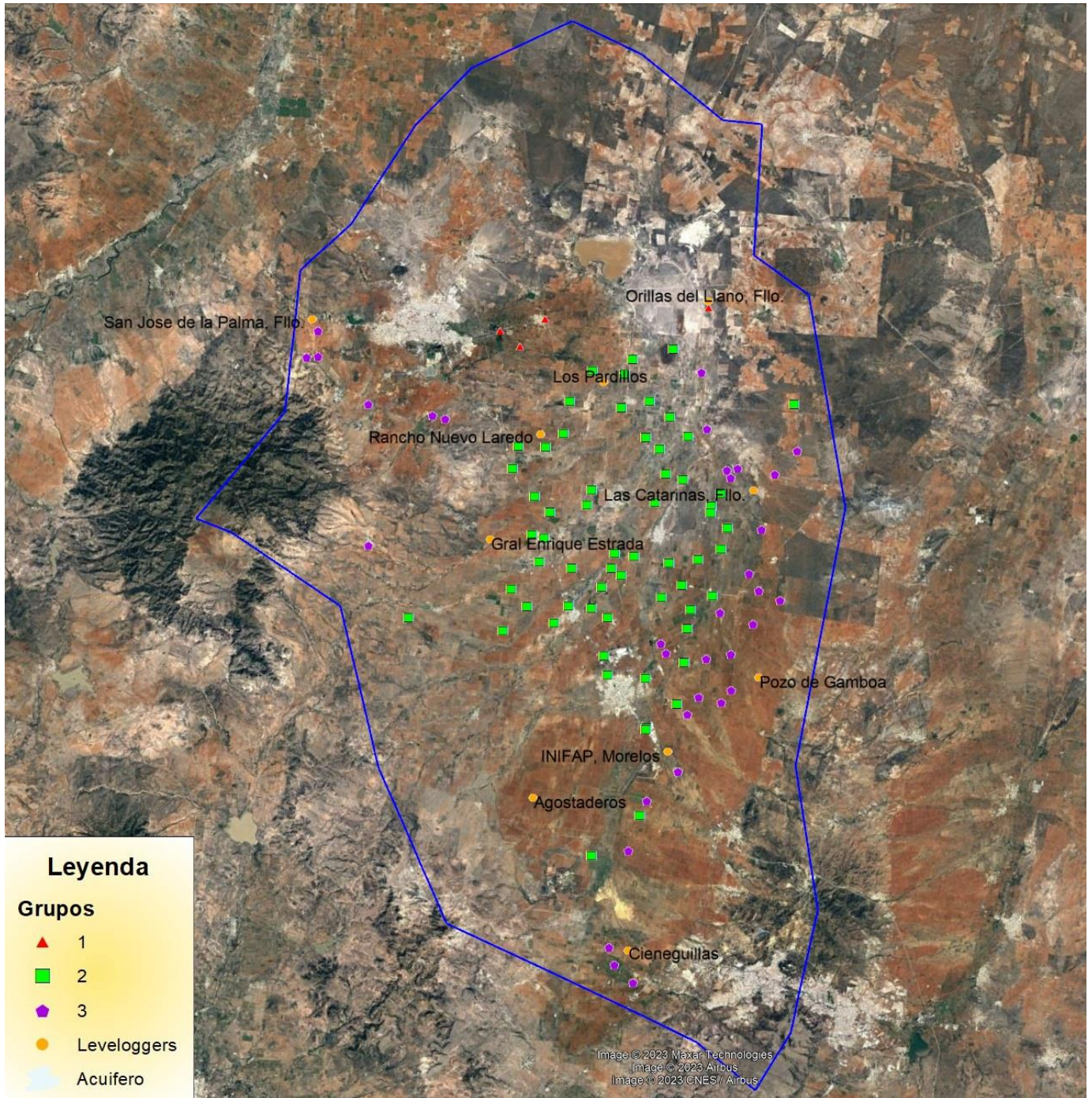


Figura 16 Mapa Espacial orográfico de los grupos con k-means.

Se puede observar en la Figura 16 que los pozos pertenecientes al grupo 3 (Pentágonos morados) se encuentran sobre el mismo tipo de suelo, que se usa para la agricultura, con respecto a los pozos del grupo 2 (Cuadrados verdes) se tiene que es otro tipo de suelo al igual que grupo 1.

En comparación con el estudio realizado por **An, R., Jiang, X. W., Wang, J. Z., Wan, L., Wang, X. S., & Li, H. (2015)** y el de esta investigación se puede observar que al utilizar datos similares como la temperatura y la química del agua y métodos de modelado diferentes se obtuvieron de resultados un producto similar que es una figura en el cual se representan los flujos del agua subterránea. Sin embargo, en la investigación comparada utilizan ciertos parámetros físicos de la cuenca de estudio y que en esta investigación no se consideraron.

El estudio de Christopher J. Daughney, Matthias Raiber, Magali Moreau-Fournier, Uwe Morgenstern y Rob van der Raaij (2012) aunque los resultados son desde otro enfoque en comparación con el estudio realizado se puede notar que la aplicación de algoritmos de Machine Learning como los clustering ayudan a clasificar o agrupar las características del agua subterránea para diversos fines y dan buenos resultados en la obtención de información.

Así como en la investigación de N. Subba Rao, P. Surya Rao y D. Deva Varma (2013) para tener una noción de la relación espacial de los procesos geoquímicos a lo largo de las trayectorias de flujo se apoya de la química del agua en diferentes puntos en la zona de estudio.

V. Conclusiones

Los valores de los coeficientes de correlación entre las temperaturas de los aprovechamientos monitoreados con leveloger fluctuó entre 0.6 a 0.8 predominando la de Rancho Nuevo-INIFAP-Catarinas, comportamiento similar tuvo la relación entre la temperatura y algunos parámetros físico químicos del agua medidos en aprovechamientos de la zona de estudio.

Con base a los análisis realizados se concluye que existe una relación en las variaciones de la temperatura de tipo espacial y una posible asociación con el tipo de suelo en la zona de estudio, lo cual se refuerza con el modelo de k-means.

Se encontró además la evolución del agua en los aprovechamientos Rancho Nuevo, INIFAP y las Catarinas, considerando que la temperatura va incrementada entre este trayecto, lo que puede indicar que la zona de Rancho Nuevo podría ser una zona de recarga importante para el acuífero de Calera.

Se puede concluir que la metodología del uso de un modelo k-means si permite pronosticar la variabilidad espacio-temporal de la temperatura registrada por leveloger y además clasificar con estas los diferentes sistemas de flujos que pueden ocurrir en el acuífero Calera.

Los resultados aquí presentados son parciales, se recomienda seguir investigando la relación entre la temperatura y algunos parámetros como conductividad eléctrica del agua, cloruros y sulfatos. Además del comportamiento de la lluvia en la zona identificada como de recarga y el trayecto hacia la evolución del sistema de flujo de local a intermedio o posiblemente regional.

VI. Literatura citada

- Agudelo-Vera, C., Avvedimento, S., Boxall, J., Creaco, E., de Kater, H., Nardo, A. Di, Djukic, A., Douterelo, I., Fish, K. E., Rey, P. L. G., Jacimovic, N., Jacobs, H. E., Kapelan, Z., Solano, J. M., Pachongo, C. M., Piller, O., Quintiliani, C., Ručka, J., Tuhovčák, L., & Blokker, M. (2020). Drinking water temperature around the globe: Understanding, policies, challenges and opportunities. *Water (Switzerland)*, 12(4). <https://doi.org/10.3390/W12041049>
- Alizamir, M., Kisi, O., & Zounemat-Kermani, M. (2018). Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological Sciences Journal*, 63(1), 63–73. <https://doi.org/10.1080/02626667.2017.1410891>
- Alpaydin, E. (2014). Introducing to Machine Learning Third Edition. In *The MIT Press* (Vol. 3).
- An, R., Jiang, X. W., Wang, J. Z., Wan, L., Wang, X. S., & Li, H. (2015). A theoretical analysis of basin-scale groundwater temperature distribution. *Hydrogeology Journal*, 23(2), 397–404. <https://doi.org/10.1007/s10040-014-1197-y>
- Christopher M. Bishop. (2006). Pattern Recognition and Machine Learning. In *EAI/Springer Innovations in Communication and Computing*. https://doi.org/10.1007/978-3-030-57077-4_11
- CONAGUA. (2019). *Estadísticas del Agua en México 2019*.
- Danis, C. (2014). Use of groundwater temperature data in geothermal exploration: The example of Sydney Basin, Australia. *Hydrogeology Journal*, 22(1), 87–106. <https://doi.org/10.1007/s10040-013-1070-4>
- Daughney, C. J., Raiber, M., Moreau-Fournier, M., Morgenstern, U., & van der Raaij, R. (2012). Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: Comparison of New Zealand's national and regional groundwater monitoring programs. *Hydrogeology Journal*, 20(1), 185–200. <https://doi.org/10.1007/s10040-011-0786-2>
- Davie, T. (2008). Fundamentals of hidrology. In *Routledge Fundamentals of Physical Geography*.
- Denis, D. J. (2020). Univariate, Bivariate, and Multivariate Statistics Using R. In *Univariate, Bivariate, and Multivariate Statistics Using R* (Vol. 185, Issue 2). Wiley. <https://doi.org/10.1002/9781119549963>
- Ertel, W. (2017). Introduction to Artificial Intelligence (Undergraduate Topics in Computer Science). In *Springer*.
- Freeze, R. A., & Cherry, J. A. (1979). Groundwater. *Groundwater*.
- Graf, R., & Aghelpour, P. (2021). Daily river water temperature prediction: A comparison between neural network and stochastic techniques. *Atmosphere*, 12(9), 1–22. <https://doi.org/10.3390/atmos12091154>
- Gupta, R. S. (2016). Hydrology and Hydraulic Systems | Ram S. Gupta. In *Waveland Press*. <https://en.my1lib.org/book/3583279/083c24>
- Hossain, M. G., Selim Reza, A. H. M., Lutfun-Nessa, M., & Ahmed, S. S. (2013). Factor and cluster analysis of water quality data of the groundwater wells of Kushtia, Bangladesh: Implication for arsenic enrichment and mobilization. *Journal of the Geological Society of India*, 81(3), 377–384. <https://doi.org/10.1007/s12594-013-0048-0>
- Kim, P. (2017). Matlab deep learning: With machine learning, neural networks and artificial intelligence. In *Apress, New York*.
- Kim, T., Moon, D.-C., Park, W.-B., Park, K.-H., & Ko, G.-W. (2007). Classification of springs of Jeju Island using cluster analysis of annual fluctuations in discharge variables: investigation of the regional groundwater system. *Geosciences Journal*, 11(4), 397–413. <https://doi.org/10.1007/bf02857055>
- Lehmann, J. (2020). Statistics and Excel: Evaluating Normality. In *Math and Statistics Guides from UB's Math & Statistics Center* (pp. 175–182). <https://ubalt.pressbooks.pub/mathstatsguides/chapter/statistics-and-excel-evaluating-normality/>
- Mackridge, A., & Rowe, P. (2018). Pearson Correlation. *A Practical Approach to Using Statistics in Health Research*, 165–172. <https://doi.org/10.1002/9781119383628.ch17>
- Maldonado, A. D., Morales, M., Navarro, F., Sánchez-Martos, F., & Aguilera, P. A. (2022). Modeling

- semiarid river–aquifer systems with bayesian networks and artificial neural networks. *Mathematics*, 10(1), 1–17. <https://doi.org/10.3390/math10010107>
- Nilsson, N. J. (2005). INTRODUCTION TO MACHINE LEARNING. *Machine Learning*, 56(2), 387–399. <http://www.ncbi.nlm.nih.gov/pubmed/21172442>
- Ordoñez Gáldez, J. J. (2011). Aguas Subterráneas - Acuíferos. *Sociedad Geográfica de Lima*, 2–44, 10. https://www.gwp.org/globalassets/global/gwp-sam_files/publicaciones/varios/aguas_subterraneas.pdf
- Reimann, C., Filzmoser, P., Garrett, R. G., & Dutter, R. (2008). Statistical Data Analysis Explained: Applied Environmental Statistics with R. *Statistical Data Analysis Explained: Applied Environmental Statistics with R*, 1–343. <https://doi.org/10.1002/9780470987605>
- Roshni, T., Jha, M. K., Deo, R. C., & Vandana, A. (2019). Development and Evaluation of Hybrid Artificial Neural Network Architectures for Modeling Spatio-Temporal Groundwater Fluctuations in a Complex Aquifer System. *Water Resources Management*, 33(7), 2381–2397. <https://doi.org/10.1007/s11269-019-02253-4>
- Russell, R. (2018). *Machine learning step-by-step guide to implement machine learning algorithms with Python*. 106. <http://booksdescr.org/item/index.php?md5=D161EE832B8007A058CD006DD67E388E>
- Shalev-Shwartz, S., & Ben-David, S. (2013). Understanding machine learning: From theory to algorithms. In *Understanding Machine Learning: From Theory to Algorithms* (Vol. 9781107057). <https://doi.org/10.1017/CBO9781107298019>
- Shevlyakov, G. L., & Oja, H. (2016). Robust Correlation. In *Mathematical Gnostics*. Wiley. <https://doi.org/10.1002/9781119264507>
- Subba Rao, N., Surya Rao, P., & Deva Varma, D. (2013). Spatial variations of groundwater vulnerability using cluster analysis. *Journal of the Geological Society of India*, 81(5), 685–697. <https://doi.org/10.1007/s12594-013-0090-y>
- Tissen, C., Benz, S. A., Menberg, K., Bayer, P., & Blum, P. (2019). Groundwater temperature anomalies in central Europe. *Environmental Research Letters*, 14(10). <https://doi.org/10.1088/1748-9326/ab4240>
- Todd, David; Mays, L. (2005). *Groundwater Hydrology* (pp. 1–652). http://water.usgs.gov/pubs/circ/circ1186/html/gw_effect.html
- Vélez, M. V., & Vásquez, L. M. (2004). Métodos para determinar la recarga en acuíferos. *Primer Congreso Colombiano de Hidrogeología*, 20.
- Vishwanathan, A. S. and S. V. N. (2014). Introduction to Machine Learning. *Cambridge University*, 1–13. https://doi.org/10.1007/978-94-007-6964-9_1
- Wishart, D. (2014). k -means Analysis. In *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.stat06492>
- Woessner, W. W. (2020). Groundwater-Surface Water Exchange. In *The Groundwater Project*.
- Woessner, W. W., & Poeter, E. P. (2020). Hydrogeologic Properties of Earth Materials and Principles of Groundwater Flow. In *Groundwater* (Vol. 59, Issue 3). <https://doi.org/10.1111/gwat.13085>
- Yao, M., Jia, M., & Zhou, A. (2018). Applied Artificial Intelligence - A Handbook for Business Leaders. In *Statistical Field Theor* (Vol. 53, Issue 9).

VII. Trabajos publicados

El trabajo titulado “**Análisis de la Temperatura del Agua Subterránea con Machine Learning (ML)**” publicado en el Congreso Internacional de Investigación de Academia Journals Puebla TecNM 2023, donde se presentó una metodología propuesta del análisis de la temperatura del agua subterránea.

