



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



Artículo: COMEII-21026

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

ESTIMACIÓN DE LA CONDUCTIVIDAD HIDRÁULICA SATURADA EMPLEANDO REDES NEURONALES ARTIFICIALES

Josué Trejo Alonso¹, Carlos Fuentes², Antonio Quevedo², Carlos Chávez^{1*}

¹Centro de Investigaciones del Agua. Departamento de Ingeniería de Riego y Drenaje. Universidad Autónoma de Querétaro. Cerro de las Campas S.N., Col. Las Campanas, C.P. 76010.
chagcarlos@uaq.mx – 442 192 12 00, Ext. 6036 (*Autor de correspondencia)

²Coordinación de Riego y Drenaje. Instituto Mexicano de Tecnología del Agua. Paseo Cuauhnáhuac Núm. 8532, Jiutepec, Morelos 62550, México.

Resumen

En el presente trabajo se construyeron varias redes neuronales artificiales (variando los datos de entrada) para calcular la conductividad hidráulica saturada (K_s) usando una base de datos construida a partir de 900 muestras de suelo obtenidas del Distrito de Riego 023 en San Juan del Río, Querétaro, México. Dichas redes usaron dos capas ocultas, un algoritmo de back-propagation para el proceso de aprendizaje y una función logística como función de transferencia no lineal. Para explorar las diversas configuraciones de neuronas dentro de la red, se hizo uso de la técnica de bootstrap para cada una de las redes creadas y se eligió aquella que tuviera el menor valor de RMSE (Root Mean Square Error). Nuestros resultados muestran un rango de RMSE de 0.0459 a 0.0413 y valores de R^2 de 0.9725 a 0.9780 que son similares a los encontrados en otros trabajos. Se encontró también que reduciendo la cantidad de datos de entrada se mejora la precisión de las predicciones, es decir, que considerando sólo tres datos de entrada tenemos un mejor valor del RMSE.

Palabras claves: diseño eficiente, riego por gravedad, infiltración, funciones de pedotransferencia.



Introducción

El movimiento del agua del suelo es importante en varias áreas, como el riego, el drenaje, la hidrología y la agricultura (Fuentes *et al.*, 2020). Entre todas las cantidades medibles de los suelos, una de las más importantes es la conductividad hidráulica saturada (K_s), definida como la capacidad de transmitir agua a través de la zona saturada (Wu *et al.*, 2017; Chávez y Fuentes, 2019;), que está altamente correlacionada con la optimización del caudal. aplicado al borde o surco en el riego por gravedad (Saucedo *et al.*, 2011; Gillies y Smith, 2015; Weibo *et al.*, 2017; Wu *et al.*, 2017; Chávez y Fuentes, 2019). Aunque esta propiedad se mide fácilmente en el laboratorio o en campo, debe aplicarse a pequeña escala, y la mayoría de las veces, se requiere su uso a gran escala (Saucedo *et al.*, 2011; Weibo *et al.*, 2017). Esto es un inconveniente debido al hecho de que todas estas pruebas y mediciones requieren mucho tiempo, son poco prácticas y no son rentables (Abdelbaki, 2018; Zhang y Schaap, 2019).

Con el fin de solucionar los inconvenientes mencionados anteriormente, se publicaron un gran número de estudios sobre funciones de pedotransferencia (PTF, por sus siglas en inglés; Brakensiek *et al.*, 1984; Rasoulzadeh, 2011; Abdelbaki, 2018; Zhang y Schaap, 2019). Estos modelos matemáticos nos permiten estimar el K_s a partir de algunas características del suelo, como textura, capacidad de campo, punto de marchitez permanente, densidad aparente, porosidad y materia orgánica, entre otras (Brakensiek *et al.*, 1984; Rasoulzadeh, 2011; Abdelbaki, 2018; Zhang y Schaap, 2019). La robustez del modelo está vinculada al número de parámetros físicos utilizados para calcular la conductividad hidráulica saturada; cuantos más parámetros, más precisa será la predicción. Sin embargo, como se mencionó anteriormente, dependiendo de las mediciones, los PTF son difíciles de obtener, debido a los recursos económicos y al tiempo que se tarda en medir todas las variables, lo que se presenta como una limitación para este tipo de función y la capacidad predictiva. Además, algunos trabajos han sido cuestionados porque el suelo en el que se quieren aplicar es diferente al suelo utilizado para su desarrollo, como en Moreira *et al.* (2004).

En los últimos años, se ha explorado otra alternativa: las redes neuronales artificiales (ANN, por sus siglas en inglés), que se han convertido en una herramienta común utilizada como una clase especial de PTF, por ejemplo, en los trabajos de Agyare *et al.* (2007) y Erzin *et al.* (2009). Las ANN son un tipo de inteligencia artificial que simula el comportamiento del cerebro humano, y su estructura consiste en una serie de elementos interconectados llamados neuronas que están organizados lógicamente en capas, conocidas como entrada, salida y ocultas (ver, por ejemplo, Erzin *et al.*, 2009 y referencias en el mismo). Cada neurona se conecta a todas las neuronas de la siguiente capa a través de conexiones ponderadas. En la Figura 1, mostramos una estructura típica de una ANN.

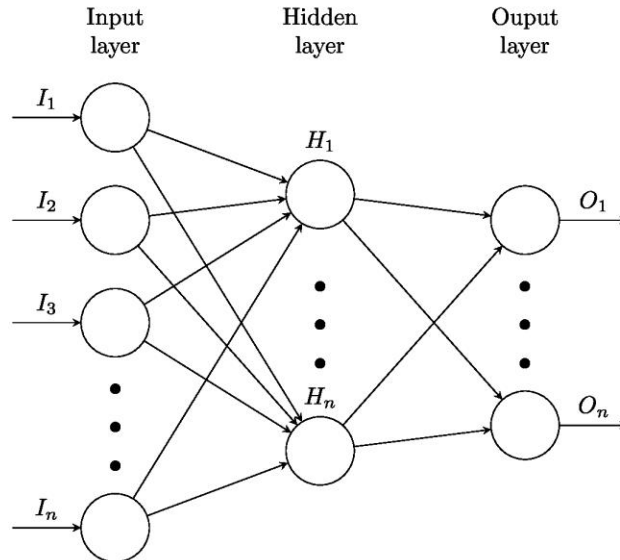


Figura 1. Representación esquemática de una ANN. Cada círculo representa una neurona y se muestran las capas de entrada, oculta y salida. Imagen tomada de Trejo-Alonso *et al.*, (2021).

Hasta ahora, no ha habido una forma analítica de obtener la estructura de red ideal (número de capas ocultas y neuronas dentro de ellas) en función de la complejidad del problema. La estructura debe seleccionarse mediante un proceso de prueba y error. Las ANN con una o dos capas ocultas y un número adecuado de neuronas ocultas resultan suficientes para la mayoría de los problemas (e. g. Sonmez *et al.*, 2006; Erzin *et al.*, 2009). También hay varios trabajos que estudian el número ideal de neuronas en las capas ocultas (Grima y Babuska, 1999; Haque y Sudhakar, 2002), pero estos métodos presentan, únicamente, pautas generales para la selección en el número de neuronas.

En este trabajo se prueban diferentes configuraciones de ANN para obtener valores de la K_s utilizando datos de 900 parcelas en el Distrito de Riego 023 San Juan del Río, Querétaro, explicando la mejor configuración de capas que se utilizó, así como la capacidad predictiva que tiene con respecto a los datos medidos. Estas predicciones, como ya se ha mencionado antes, nos ofrecen una alternativa para el cálculo de K_s . Este trabajo se divide como sigue: en el capítulo de Materiales y Métodos se describe de manera detallada la muestra, los parámetros medidos y el proceso que se realizó para generar la red neuronal. Seguido a esto tenemos los Resultados y Discusión, donde se muestra el proceso de selección en la configuración de la ANN, los valores predichos comparados con los resultados medidos y un análisis de estos mismos. Finalmente, en la parte de Discusión, tendremos una descripción de las implicaciones y resultados de este trabajo.

Materiales y Métodos

El Distrito de Riego 023 está ubicado en Querétaro, México, a $20^{\circ}18'$ a $20^{\circ}34'$ N, $99^{\circ}56'$ a $100^{\circ}12'$ W con una altitud de 1892 m.s.n.m. y tiene un área de 11 048 ha. Incluye los municipios de San Juan del Río y Pedro Escobedo (figura 2). Su clima predominante es semiárido con lluvias de verano, con una precipitación media anual de 599 mm y una

temperatura media anual de 20°C (CONAGUA, 2018). El agua se conduce a través de canales abiertos. Los canales principales están revestidos de hormigón, pero todos los canales laterales que llevan el agua a las parcelas están sin revestir. La separación de las parcelas en algunos casos es por árboles, canales sin revestimiento, desagües o caminos (Chávez *et al.*, 2020).

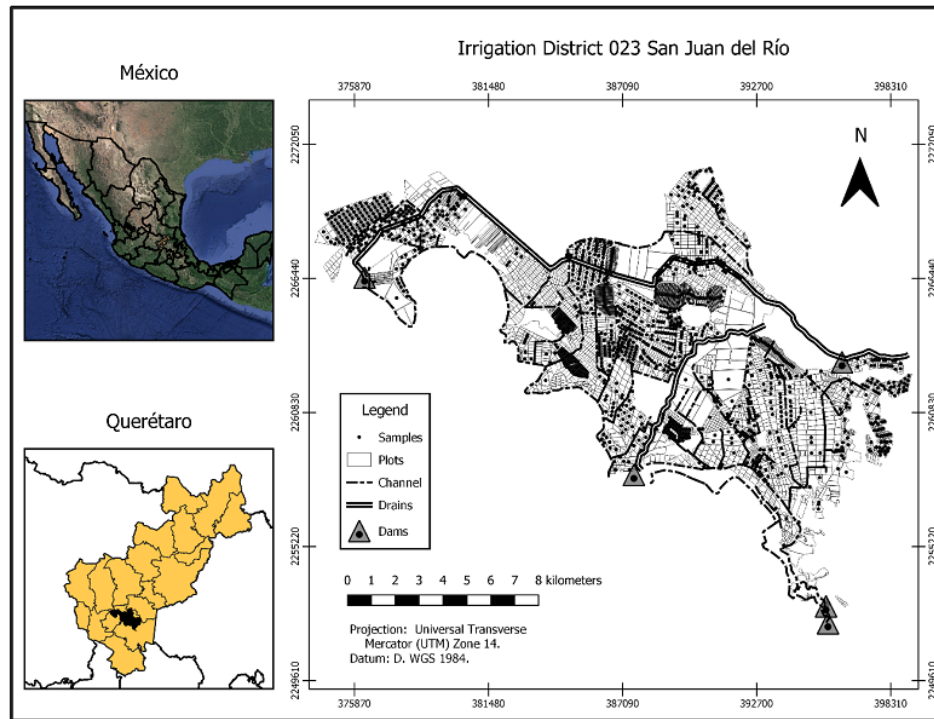


Figura 2. Distrito de Riego 023, San Juan del Río Querétaro, México.

La descripción extensa y detallada de la base de datos utilizada en este trabajo se puede ver en Trejo-Alonso *et al.*, 2020. En resumen, fue desarrollada a partir de muestreos en 900 parcelas en el Distrito de Riego 023, San Juan del Río Querétaro. Estas muestras fueron enviadas al laboratorio para obtener los siguientes parámetros: textura del suelo por el hidrómetro de Bouyucos, densidad aparente (ρ_s) por el método del cilindro de volumen conocido, contenido de humedad en saturación (θ_s), capacidad de campo (CC) y punto de marchitamiento permanente (PMP) por el método de la olla de membrana de presión, y la conductividad hidráulica saturada por el método de permeámetro de carga variable. De estas mediciones, tomamos siete y las consideramos como datos de entrada: porcentaje de arcilla, arena y limo, densidad aparente, punto de marchitamiento permanente, contenido de humedad en saturación y capacidad de campo (Cuadro 1).

Como se mencionó anteriormente, no existe una manera analítica para definir la estructura de la ANN, pero basados en los trabajos de Haque y Sudhakar (2002), Sonmez *et al.*, (2006) y Erzin *et al.*, (2009), se puede notar que las ANN tienen una o dos capas ocultas solamente. Con esto en mente, se probaron diversas estructuras para las ANN en este trabajo con dos capas ocultas.

Cuadro 1. Propiedades estadísticas de los datos medidos en laboratorio

Variable	Min	Max	Mediana	Media	S.D.	Q1	Q3
Arena (%)	0.07	77.83	28.35	31.14	20.22	13.75	52.00
Arcilla (%)	2.12	59.46	21.74	21.95	12.06	13.44	30.00
Limo (%)	0.80	92.00	45.27	46.91	23.48	27.30	59.79
Densidad (g/cm ³)	1.18	1.70	1.40	1.41	0.11	1.32	1.47
θ_s (cm ³ /cm ³)	0.35	0.56	0.47	0.47	0.04	0.45	0.50
CC (cm ³ /cm ³)	0.17	0.47	0.29	0.30	0.06	0.25	0.32
PMP (cm ³ /cm ³)	0.07	0.35	0.13	0.15	0.05	0.10	0.17
K_s (cm/h)	0.05	5.15	0.78	1.42	1.42	0.40	1.80

Se comenzó con todos los datos de entrada (siete) y se probaron varias configuraciones de la siguiente manera: se comenzó con dos neuronas en la primera capa y dos neuronas en la segunda. Luego se continuó variando el número de neuronas en la primera capa de dos a diez dejando la segunda capa con un número de neuronas constante. A continuación, el número de neuronas en la segunda capa se incrementa a tres y variamos el número de neuronas en la primera, nuevamente, de dos a diez, y así sucesivamente hasta que el número de la segunda capa varía de dos a diez, al igual que el primero. Esta capa de entrada tiene los siete datos de entrada mencionados anteriormente. Finalmente, la capa de salida contiene el valor predicho de K_s .

Posteriormente, se cambió el número de datos en la capa de entrada (disminuyéndolo en uno) y se repite el proceso como se explicó anteriormente. La elección de qué parámetro debe eliminarse se basa en la gráfica de importancia, excepto en la última configuración, donde se eliminó θ_s en lugar del porcentaje de limo, porque θ_s está estrechamente relacionado con el porcentaje de arcilla. Se siguió eliminando los datos de entrada hasta que sólo tuvimos se tienen tres variables. Para cada configuración, se varió el número de neuronas en las capas ocultas de dos a diez.

La ANN se programó utilizando el paquete neuralnet (Fritsch *et al.*, 2020), y el paquete caret (Kuhn, 2020), ambos proporcionados por el software R (R Core Team, 2020). Utilizando la medida RMSE se seleccionó la estructura óptima de la ANN en cada caso, y finalmente, se mantuvo esta última estructura ANN como ideal. La medición del error absoluto medio (MAE) se calculó como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (1)$$

donde y_i son los valores medidos, x_i son los valores predichos y n es el número de medidas totales. También se calculó la RMSE como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (2)$$

Para validar los resultados de las ANN, se hizo un análisis de cross-validation usando la función train del paquete caret. Esta función configura una cuadrícula de parámetros de ajuste para una serie de rutinas de clasificación y regresión, se ajusta a cada modelo y calcula una medida de rendimiento basada en el remuestreo. En este caso, usamos los parámetros optimizados para neuralnet y generamos 25 replicaciones de bootstrap para cada configuración de ANN. Finalmente, para la predicción de nuevas muestras, usamos la función de predicción.

La función train arroja varios resultados: la configuración ideal de ANN, los mejores valores de RMSE, MAE y R^2 para cada configuración probada, una matriz de RMSE y una gráfica de importancia. Este último gráfico indica qué parámetro contribuyó más a la aproximación final de K_s . Basándonos en este último gráfico, decidimos qué datos de entrada se eliminarían en la siguiente ejecución, con la excepción ya mencionada anteriormente.

Resultados y Discusión

Se generaron diversas ANN que contienen diferentes configuraciones del número de neuronas y de la ubicación de éstas en las capas ocultas. En la figura 3 se muestra una de estas pruebas y de cómo varía el RMSE de acuerdo a la configuración de la ANN. Una vez que seleccionamos la estructura con el menor valor del RMSE, se generan las predicciones de K_s con 4 datos de entrada (que fue el mejor resultado obtenido).

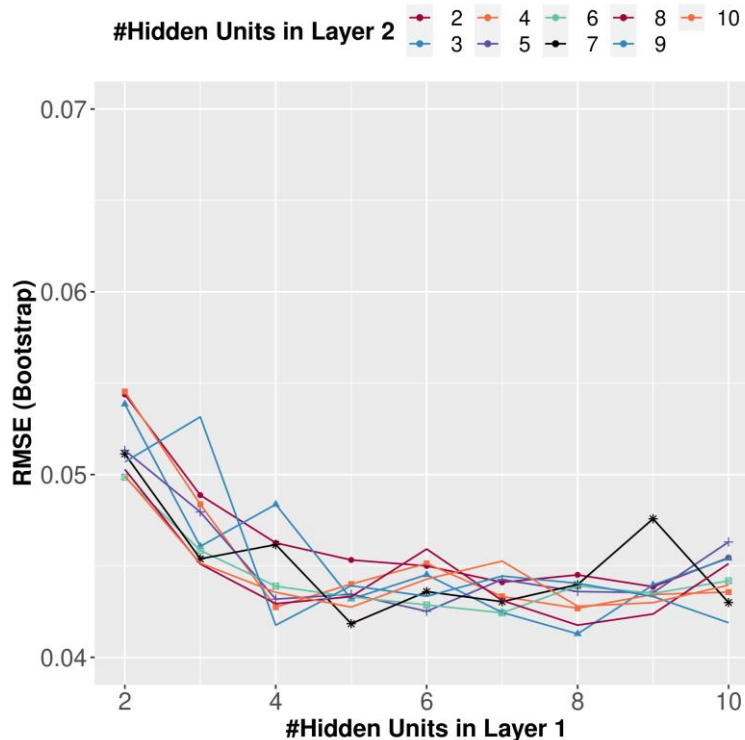


Figura 3. Ejemplo de la variación del RMSE contra la estructura de la ANN. Imagen tomada de Trejo-Alonso *et al.* (2021).

La figura 4 muestra estos resultados, donde también se puede observar la distribución de los residuos (la resta entre el valor predicho con el valor medido), la línea punteada representa una relación 1:1.

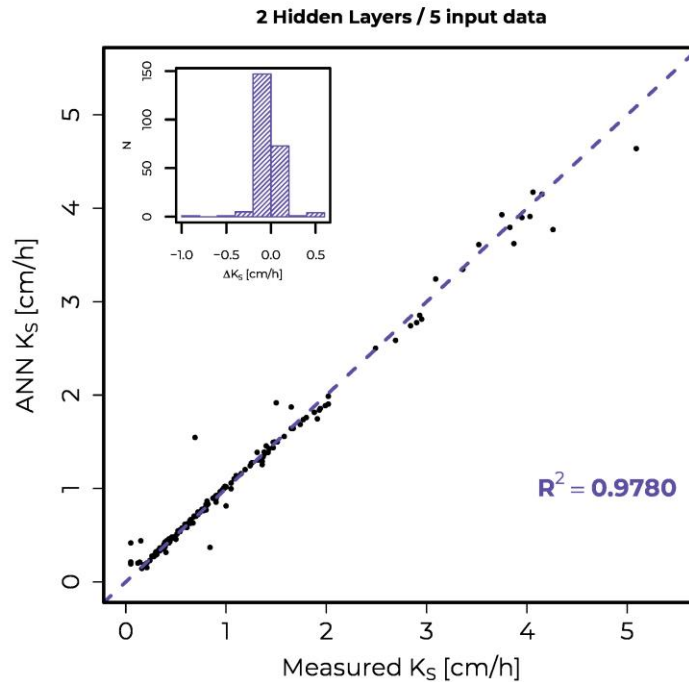


Figura 4. Comparación entre los datos medidos y los datos predichos de K_s . Imagen tomada de Trejo-Alonso *et al.* (2021).

Conclusiones

En este trabajo se muestra el uso de las Redes Neuronales Artificiales para calcular la conductividad hidráulica saturada a partir de una base de datos de 900 muestras de suelos en el Distrito de Riego 023. Recordemos que, como se ha mencionado anteriormente, la conductividad hidráulica saturada tiene una aplicación directa en los riegos por gravedad, por lo que es primordial calcularlo o predecirlo de la manera más eficiente posible.

Los resultados muestran que, en comparación con otros trabajos, obtenemos resultados mejores o similares para las mediciones de RMSE y R^2 y configuraciones similares para nuestra ANN. Finalmente, podemos decir que si se dispone de los recursos necesarios para obtener una gran cantidad de datos en campo, es necesario desarrollar un estudio de PTF así como ANN para comparar los resultados de cada proceso y poder elegir la mejor opción entre ambos. Este último no solo se basará en las mediciones RMSE o R^2 , sino también en la aplicación deseada (se propone un estudio estadístico de propiedades del suelo o predicción para riego). Además, queda abierto a debate si teniendo una muestra más grande o más pequeña se podrían reproducir los resultados aquí mostrados, para explorar esta opción se debe realizar un estudio para definir un rango exacto de la cantidad de datos necesarios de una red neuronal artificial confiable, pero este último está más allá del objetivo de este trabajo. Además, tenemos que ser más cuidadosos en las características de la muestra donde salen los modelos. En nuestro

caso, analizamos 10 de los 12 tipos de suelos donde la densidad aparente y el porcentaje de arcilla se volvieron parámetros más importantes en comparación con otros. Esto hizo que nuestros modelos fueran más confiables para casi cualquier tipo de suelo.

Finalmente, se ha demostrado que la aplicación de redes neuronales artificiales reproducen (o mejoran) los resultados conseguidos con modelos de pedotransferencia, y esto probablemente se deba a la relación no lineal entre las variables. El cálculo de la conductividad hidráulica saturada en este trabajo demuestra que solo necesitamos tres variables para predecir nuevos valores, pero las propiedades del suelo son cruciales para la correcta aplicación de estos modelos en contraste con la configuración ANN, que se ha demostrado juega un papel menor en los resultados finales.

REFERENCIAS

- Abdelbaki, A.M. (2018). Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils. *Ain Shams Engineering Journal*, 9, 1611–1619.
- Agyare, W.A.; Park, S.J.; Vlek, P.L.G. (2007). Artificial Neural Network Estimation of Saturated Hydraulic Conductivity. *Vadose Zone Journal*, 6(2), 423–431.
- Brakensiek, D.; Rawls, W.J.; Stephenson, G.R. (1984). Modifying SCS Hydrologic Soil Groups and Curve Numbers for Rangeland Soils; Paper No. PNR-84203; ASAE: St. Joseph, MN, USA.
- Chávez, C.; Fuentes, C. (2019). Design and evaluation of surface irrigation systems applying an analytical formula in the irrigation district 085, La Begoña, Mexico. *Agricultural Water Management*. 221, 279–285.
- Chávez, C.; Limón-Jiménez, I.; Espinoza-Alcántara, B.; López-Hernández, J.A.; Bárcenas-Ferruzca, E.; Trejo-Alonso, J. (2020) Water-Use Efficiency and Productivity Improvements in Surface Irrigation Systems. *Agronomy*, 10(11), 1759-1771.
- Comisión Nacional del Agua (CONAGUA). (2018). Estadísticas del Agua en México; CONAGUA: Coyoacán, México, p. 306.
- Erzin, Y.; Gumaste, S.D.; Gupta, A.K.; Singh, D.N. (2009). Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. *Canadian Geotechnical Journal*, 46(8), 955–968.
- Fritsch, S.; Guenther, F; Wright, M.N. (2020). Neuralnet: Training of Neural Networks. R Package Version 1.44.2. Disponible en línea: <https://CRAN.R-project.org/package=neuralnet> (acceso el 19 octubre 2020).
- Fuentes, S.; Trejo-Alonso, J.; Quevedo, A.; Fuentes, C.; Chávez, C. (2020). Modeling Soil Water Redistribution under Gravity Irrigation with the Richards Equation. *Mathematics* 8(9), 1581-1594.
- Gillies, M.H.; Smith, R.J. (2015). SISCO: Surface irrigation simulation, calibration and optimization. *Irrigation Science*, 33, 339–355.
- Grima, M.A.; Babuska, R. (1999). Fuzzy model for the prediction of unconfined compressive strength of rock samples. *International Journal of Rock Mechanics and Mining Sciences*, 36(3), 339–349.
- Haque, M.E.; Sudhakar, K.V. (2002). ANN back-propagation prediction model for fracture toughness in microalloy steel. *International Journal Fatigue*, 24(9), 1003–1010.
- Kuhn, M. (2020). caret: Classification and Regression Training. R Package Version 6.0-86. 2020. Disponible en línea: <https://CRAN.R-project.org/package=caret> (acceso el 22 enero 2021).



- Moreira, L.; Righetto, A.M.; Medeiros, V.M. (2004). Soil hydraulics properties estimation by using pedotransfer functions in a northeastern semiarid zone catchment, Brazil. International Environmental Modelling and Software Society, Osnabrueck. Complexity and Integrated Resources Management. In Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs, International Environmental Modelling and Software Society, Volume 2, pp. 990–995, Manno, Switzerland.
- Rasoulzadeh, A. (2011) Estimating Hydraulic Conductivity Using Pedotransfer Functions. Hydraulic Conductivity-Issues, Determination and Applications, Elango, L., pp. 145–164, Rijeka, Croatia, InTech.
- R Core Team. (2020). R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. Disponible en línea: <https://www.R-project.org/> (acceso el 12 enero 2021).
- Saucedo, H.; Zavala, M.; Fuentes, C. (2011). Complete hydrodynamic model for border irrigation. Water Technology Science, 2, 23–38.
- Sonmez, H.; Gokceoglu, C.; Nefeslioglu, H.A.; Kayabasi, A. (2006) Estimation of rock modulus: For intact rocks with an artificial neural network and for rock masses with a new empirical equation. International Journal Rock Mechanics and Mining Sciences, 43(2), 224–235.
- Trejo-Alonso, J.; Fuentes, C.; Chávez, C.; Quevedo, A.; Gutierrez-Lopez, A.; González-Correa, B. 2021. Saturated Hydraulic Conductivity Estimation Using Artificial Neural Networks. Water, 13, 705
- Trejo-Alonso, J.; Quevedo, A.; Fuentes, C.; Chávez, C. (2020). Evaluation and Development of Pedotransfer Functions for Predicting Saturated Hydraulic Conductivity for Mexican Soils. Agronomy, 10(10), 1516-1526.
- Weibo, N.; Ma, X.; Fei, L. (2017). Evaluation of infiltration models and variability of soil infiltration properties at multiple scales. Irrigation and Drainage, 66, 589–599.
- Wu, D.; Xue, J.; Bo, X.; Meng, W.; Wu, Y.; Du, T. (2017). Simulation of irrigation uniformity and optimization of irrigation technical parameters based on the SIRMOD model under alternate furrow irrigation. Irrigation and Drainage, 66, 478–491.
- Zhang, Y.; Schaap, M.G. (2019). Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. Journal of Hydrology, 575, 1011–1030.