

Predicción de la producción y rendimiento de frijol, con modelos de redes neuronales artificiales y datos climáticos

Prediction of bean production and yields, with artificial neural network models and climate data

Servín-Palestina M¹, Salazar-Moreno R^{1*}, López-Cruz IL¹, Medina-García G², Cid-Ríos JA²

¹ Universidad Autónoma Chapingo. Carretera México-Texcoco km 38.5, CP 56230, Chapingo, Estado de México, México.

² Campo Experimental Zacatecas. INIFAP. Carretera Zacatecas-Fresnillo km 24.5, CP 98600. Calera Zacatecas, México.

RESUMEN

El estado de Zacatecas ocupa el primer lugar en la producción de frijol de temporal en México, dada la importancia de las variables climatológicas, los objetivos de este trabajo fueron: 1) desarrollar modelos de redes neuronales artificiales RNA para la predicción de la superficie cosechada (SC), los rendimientos (Rto) y la producción (P) de frijol de temporal en el estado de Zacatecas, empleando datos del periodo 1988-2019, y 2) realizar el análisis de sensibilidad para determinar las variables de entrada que tienen mayor influencia en la producción y rendimiento de frijol. Se usó la librería Climatol del paquete estadístico R, para el llenado de datos faltantes. Los modelos de RNA son capaces de detectar la influencia del clima en la producción de frijol. La eficiencia global en los modelos RNA fue de 0.89 para Rto y 0.86 para SC. La producción se estimó con los modelos de RNA para Rto y SC y se obtuvo un $R^2 = 0.80$. De acuerdo al análisis de sensibilidad, la evaporación del ciclo del cultivo (Eva) fue la variable más importante en la predicción del rendimiento, mientras que la precipitación de agosto (Pp_Ago) y la temperatura mínima (Tmin) influyeron más en la producción.

Palabras clave: Inteligencia artificial, Zacatecas, temperatura, precipitación, cultivos de temporal, *Phaseolus vulgaris*, L.

ABSTRACT

The state of Zacatecas ranks first in the production of rainfed beans in Mexico. Due to the importance of climatological variables, the objectives of this work were 1) to develop artificial neural networks (ANN) models of for the harvested surface (SC), yields (Rto) and production (P) prediction of rainfed beans in the state of Zacatecas, using data from the 1988-2019 period, and 2) perform the sensitivity analysis to determine the input variables that have the greatest influence on bean production and yield. The Climatol library of the R statistical package was used to fill in missing data. The results show that the ANN models capture the influence of climate on bean production, with an overall efficiency of 0.89 for Rto and 0.86 for SC. The production was estimated using the outputs, Rto and SC, from ANN models and an $R^2 = 0.80$ was obtained. According to the sensitivity analysis, evaporation of the cycle (Eva) was the most important variable for predicting yield, while precipitation in August (Pp_Ago) and minimum temperature (Tmin) had a greater influence on production.

Keywords: Artificial intelligence, Zacatecas, temperature, rainfall, rainfed crops, *Phaseolus vulgaris*, L.

INTRODUCCIÓN

De acuerdo con Stacey *et al.* (2015) la producción de alimentos debe incrementarse en un 70 % para el año 2050, lo cual es difícil de alcanzarse en un escenario de cambio climático y de agotamiento de recursos naturales. Las zonas de temporal son las más afectadas debido a la variabilidad de la distribución temporal y espacial de la precipitación (Esquivel *et al.*, 2016; Trethowan *et al.*, 2010). Este fenómeno ha provocado reducción en la producción agrícola especialmente en el altiplano semiárido templado del centro-norte de México (Osuna *et al.*, 2019).

Actualmente, el estado de Zacatecas ocupa el primer lugar en la producción de frijol en México, con una superficie cosechada en el año 2019 de 395 mil ha y rendimiento promedio de 0.55 t ha⁻¹. Con respecto al año 2018, la superficie cosechada se redujo en 60 %, debido a que las condiciones climáticas ocasionaron el siniestro de 108 mil ha sembradas (SIAP, 2019). Por esta razón, es necesario contar con herramientas que permitan relacionar la producción agrícola con las condiciones climáticas. La modelación numérica con datos climáticos toma importancia para predecir eventos (Grageda *et al.*, 2012) y pronosticar requerimientos hídricos (Ruiz-Del-Angel *et al.*, 2019). Las redes neuronales artificiales son modelos de caja negra (estáticos o dinámicos) que permiten manejar relaciones no lineales entre variables de entrada y salida de un sistema, mediante elementos llamados nodos y sus interconexiones que imitan las neuronas de los sistemas biológicos (Hagan *et al.*, 1996). Las RNA son modelos capaces de encontrar relaciones de forma inductiva por medio de algoritmos de aprendizaje y datos existentes, utilizadas para clasificar, identificar, diagnosticar, optimizar y predecir, se basan en la analogía del comportamiento del cerebro humano (Fausett, 2006).

Los modelos de redes neuronales artificiales (RNA), han sido utilizados con éxito para tratar problemas en la agricultura moderna, en donde se integran variables climatológicas, del cultivo y manejo. Abraham *et al.* (2020) compararon los modelos RNA con métodos clásicos, con base en el error medio absoluto (MAE), error cuadrado medio (MSE) y coeficiente de determinación (R^2), los autores concluyeron que las RNA tuvieron mejor desempeño para predecir su-

perficie cosechada y producción de soya, mientras que los modelos clásicos fueron mejores para estimar el rendimiento. Khan *et al.* (2020) utilizaron tres algoritmos de aprendizaje (Levenberg-Marquardt (LM), scale conjugate gradient back propagation (SCG), Bayesian regularization back propagation (BR)) de RNA para predecir la producción de frutas en Pakistán, obtuvieron valores de R^2 de 65.6, 70.2 y 76.3 respectivamente y concluyeron que las redes neuronales predicen periódicamente las tendencias de la producción agrícola de frutas. Maimaitijiang *et al.* (2020) estimaron los rendimientos de tres variedades de soya en Columbia, Missouri, USA, con la implementación de tres técnicas de regresión de mínimos cuadrados parciales (PLSR), regresión aleatoria de bosques (RFR) y redes neuronales artificiales (RNA), utilizando las firmas espectrales, imágenes térmicas y textura del dosel como variables de entrada, concluyeron que el modelo de RNA tuvo el mejor desempeño ($R^2 = 0.720$ y $RMSE = 15.9\%$).

Una vez que el modelo de RNA ha sido generado, es necesario realizar el análisis de sensibilidad (AS) el cual permite particionar la incertidumbre en cada una de las entradas (Saltelli, 2002), a fin de identificar las variables más importantes que predicen la salida del modelo de RNA. Iriyoyen *et al.* (2012) realizaron un AS en RNA entrenadas para estimar la lámina de riego en maíz (Lr), con componentes del balance de agua (BHC) y estado de desarrollo del cultivo y concluyeron que los BHC son más sensibles para estimar la lámina de riego a los primeros 60 cm de profundidad del suelo.

Asimismo, García *et al.* (2020) estimaron el rendimiento de grano de maíz con una RNA multicapa, utilizando variables de entrada obtenidas de imágenes multiespectrales y datos colectados in-situ y obtuvieron un coeficiente de correlación entre los rendimientos observados y simulados mayores al 85 %, de igual manera, el análisis de sensibilidad mostró que el índice de vegetación y la densidad de plantas tiene mayor importancia relativa para la estimación de rendimiento.

Predecir la producción y rendimiento ayuda a: 1) dirigir los recursos para generar tecnologías que incrementen el rendimiento, 2) enfocar esfuerzos para mitigar el efecto del clima, 3) generar políticas gubernamentales para aumentar la productividad, 4) servir de base para los pagos de seguros y préstamos para la producción de frijol (Intell, 2013; Crane, 2018). Los modelos RNA son capaces de predecir fenómenos altamente no lineales y de allí su utilidad en la determinación de la disponibilidad de grano de frijol con anticipación. Los objetivos del presente trabajo fueron: 1) desarrollar modelos de RNA para la predicción de la superficie cosechada, los rendimientos y la producción de frijol de temporal en el estado de Zacatecas, empleando datos de temperatura máxima y mínima del aire, precipitación y evaporación durante el periodo 1988-2019; 2) realizar un análisis de sensibilidad para determinar cuáles variables de entrada tienen mayor influencia en la producción y rendimiento de frijol. La hipótesis es que los modelos de RNA son capaces de predecir de manera aceptable las variables producción, superficie cosechada y rendimiento de frijol utilizando como variables de entrada factores climáticos y de manejo.

MATERIALES Y MÉTODOS

Descripción del área de estudio

La "zona frijolera" en el estado de Zacatecas está ubicada al noroeste del estado, en las coordenadas $23^{\circ}53'02''$ latitud Norte y $103^{\circ}01'33''$ longitud Oeste, con una altura media de 1883 m.s.n.m. los suelos son tipo aluvión. La precipitación media anual es de 472.4 mm. La mayor precipitación se presenta en los meses de julio-septiembre, en donde se concentra el 63 % de la precipitación anual. La temperatura media anual y máxima mensual es de 16.8°C y 29.5°C , respectivamente. La evapotranspiración potencial media de 1 389 mm por año (Medina y Ruiz, 2004). Debido a las características climáticas anteriores la producción de frijol de temporal se lleva a cabo de julio a septiembre.

Obtención de datos

La base de datos climáticos (1988-2019) se obtuvo de la estación número 32021 del servicio meteorológico nacional (SMN, 2010) (Estación A), ubicada en la colonia Gonzáles Ortega (-103.44°O , 23.956°N) con datos diarios del 1988 a 2004, y de la estación agroclimática automatizada Col. González Ortega, de la red de monitoreo agroclimático del estado de Zacatecas, CEZAC-INIFAP (Medina *et al.*, 2007) (Estación B), ubicada a -103.45°O , 23.966°N y con datos diarios del 2005 al 2019. Ambas localizadas a 1214 msnm con una separación de 1.5 km, las cuales representan las condiciones climáticas para la producción de frijol de temporal en Zacatecas denominada como la "zona frijolera" (Galindo y Zandate, 2006).

Para el presente trabajo se utilizaron los datos climatológicos de junio a septiembre de: temperatura mínima (T_{\min}), temperatura máxima (T_{\max}), precipitación (P_p) y evaporación (E_{va}) de los meses de julio a septiembre, meses de producción de frijol de temporal en Zacatecas (Medina *et al.*, 2003). Los datos de radiación solar no se incluyen, debido a que no están disponibles durante el periodo 1988 a 2004. Sin embargo, la evaporación la incluye de manera indirecta (SMN, 2010).

Datos faltantes

Fue necesario un pre-procesamiento de ambas estaciones climáticas que consistió en: 1) Estimación de datos climáticos faltantes del año 1988 al 2004 de la estación A, 2) Estimación de evaporación de la estación B en función de evapotranspiración potencial, 3) Estadísticas descriptivas de las variables climáticas del ciclo de producción.

La estimación de datos faltantes para la estación A se llevó a cabo utilizando el paquete estadístico R mediante la librería de *Climatol* (Guijarro, 2018), usando dos estaciones aledañas 32028 y 32054 ubicadas en Juan Aldama y Sombrerete respectivamente. *Climatol* usa la homogenización de las estaciones para rellenar los datos faltantes a partir de las series más próximas. Se adaptó el método de Paulhus y Kohler (1952), para rellenar precipitaciones diarias mediante promedios de valores aledaños, normalizados mediante división por sus respectivas medias utilizando el método de estandarización de la ecuación 1.

$$x = \frac{(X - \bar{X})}{S_x} \quad (1)$$

Donde: x es normalización de la precipitación, \bar{X} y S_x es la media y desviación estándar de una serie X respectivamente.

Climatol obtiene los parámetros anteriores con los datos disponibles en cada serie, rellena los datos faltantes usando \bar{X}_i y S_{x_i} . Se calculan nuevamente las series rellenas con nuevos parámetros de \bar{X}_{i+1} y $S_{x_{i+1}}$, lo anterior se hace de manera recursiva hasta que las medias no cambian. Una vez estabilizadas las medias, se normalizan todos los datos y se procede a estimarlos (datos existentes y no existentes en todas las series) mediante la expresión (2):

$$\hat{y} = \frac{\sum_{j=1}^n w_j x_j}{\sum_{j=1}^n w_j} \quad (2)$$

Donde: \hat{y} es un dato estimado mediante los correspondientes $n=10$ datos x_j más próximos disponibles en paso temporal, y w_j es el peso asignado a cada uno de ellos en función de la distancia de la estación (Guijarro, 2002).

De la estación B se obtuvieron datos de evapotranspiración potencial (ET_o) estimados mediante la ecuación Penman-Monteith (Allen *et al.*, 2006). Raghuwanshi y Wallender (1998) mencionan que hay una relación entre la evapotranspiración potencial (ET_o) y la evaporación. Según Silva (1998) la evapotranspiración potencial (ET_o) se puede determinar a partir de las mediciones de evaporación (Eva) con base en unos coeficientes empíricos que dependen del suelo, tipo de cultivo, régimen de lluvias entre otros factores, como se expresa en la Ecuación 3:

$$ET_o = K * Eva \quad \therefore \quad Eva = \frac{ET_o}{K} \quad (3)$$

Donde: K es un coeficiente para el período de tiempo determinado, cuyo valor oscila entre 0.5 y 0.9, el cual depende de la intensidad y la distribución de las precipitaciones. Se utilizó el coeficiente K igual a 0.78 que se obtuvo al relacionar la ET_o de la estación Gonzales Ortega y Eva de la estación 32021 con los datos existentes del 2007 y 2008 ($n=555$).

Variables de entrada y salida

Las estadísticas descriptivas de las variables climáticas del ciclo de producción del mes de julio, agosto y septiembre se basaron en los promedios mensuales de las variables climáticas utilizadas, se obtuvieron mediante la siguiente expresión.

$$\bar{vc} = \frac{1}{n} \sum_{i=1}^n vc_i \quad (4)$$

Donde: \bar{vc} es el promedio mensual de la variable climática, n el número de días del mes (30 o 31 días) según corresponda y vc_i valor diario de las variables climáticas.

Se obtuvieron 12 datos por año, de un total de 32 años (1988-2019): T_{max_Jul} , T_{max_Ago} , T_{max_Sep} , T_{min_Jul} , T_{min_Ago} , T_{min_Sep} , Pp_Jul , Pp_Ago , Pp_Sep , Eva_Jul ,

Eva_Ago , Eva_Sep . También, se obtuvo el valor máximo de la T_{max} (TT_{max}) y el valor mínimo de T_{min} (TT_{min}), la precipitación total (SPp) y la evaporación total ($SEva$) del ciclo de producción, dando un total de 16 variables climáticas por año.

La información sobre rendimiento de frijol de temporal (Rto ; $t \text{ ha}^{-1}$), superficie sembrada (SS ; miles, ha), superficie cosechada (SC ; miles ha) y producción (P ; t), se obtuvo del Anuario Estadístico de la Producción Agrícola (SIAP, 2019) para los años 1988-2019, entidad federativa: Zacatecas, cultivo: frijol, ciclo: perenes y modalidad: temporal, obtenidos de los promedios anuales de las variedades y localidades del estado de Zacatecas.

Con el fin de conocer la fuerza de la relación lineal entre las 16 variables climáticas y superficie sembrada consideradas como variables de entrada, contra rendimiento de frijol, superficie cosechada y producción incluidas como variables de salida, se realizó un análisis de correlación (Gogtay y Thette, 2017), previamente al diseño de las RNA's.

Diseño y entrenamiento de la red

Se implementaron tres redes neuronales artificiales (RNA) con una estructura perceptrón multicapa. Con 16 variables climáticas y la superficie de siembra como las entradas y como variables de salida rendimiento de frijol, superficie cosechada y producción. El número de neuronas en la capa oculta (h) se varió utilizado como referencia los criterios: $h = \sqrt{mn}$, como límite inferior y $h = 2n + 1$ del teorema de Kolmogorov (Kolmogorov, 1957) como límite superior, en donde: n es número de entradas y m es el número de salidas. Las variables de entrada para cada una de las redes se seleccionaron con base a los resultados del análisis de correlación y un estudio exploratorio, combinando las variables de entrada y número de neuronas en la capa oculta. Se utilizó el 95 % de los datos (1988-2017) para el entrenamiento, validación y prueba con 50, 25 y 25 % de los datos respectivamente y el 5 % restante (2018-2019) para la evaluación. El algoritmo de aprendizaje utilizado fue gradiente descendiente con momentum (Garg *et al.*, 2016; Mohamed, 2019) y algoritmo de entrenamiento fue Levenberg-Marquardt backpropagation (LM) (Marquardt, 1963). Para generar los modelos de RNA se realizó un análisis exploratorio combinando las diferentes variables de entrada y número de neuronas en capa oculta hasta alcanzar un buen desempeño.

Todas las RNA se implementaron en MATLAB® con el Toolbox de "Neural Networks" (Demuth *et al.*, 2017; Moore, 2012). La función "newff" se utilizó para crear, entrenar y validar la red, el número de entradas y neuronas en la capa oculta se eligió con base al análisis exploratorio para cada modelo RNA. Se obtuvo el conjunto de pesos óptimos que minimizan el error de acuerdo con la metodología descrita por Chatterjee *et al.* (2018).

Medidas de desempeño

Para medir el desempeño de la RNA se calcularon: 1) el coeficiente de determinación (R^2), (Ecuación 5), (Demuth *et al.*, 2017; Pecar y Davis, 2018), 2) el error medio absoluto (MAE ;

Ecuación 6), 3) el error cuadrado medio (*MSE*; Ecuación 7); 4) la raíz cuadrada del error cuadrado medio (*RMSE*; Ecuación 8), (Heng *et al.*, 2009), y 5) la eficiencia propuesta por Nash y Sutcliffe (*E*; Ecuación 9) que caracteriza el comportamiento del modelo de simulación, el modelo perfecto debería tener una eficiencia cercana a 1 (Krause *et al.*, 2005).

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$E = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

Donde: y_i representa el valor observado, \hat{y}_i el valor predicho de la red, \bar{y} es el promedio de los valores observados y n el número de datos.

Análisis de sensibilidad

Con el fin de identificar las variables más importantes para predecir la salida, se implementó el método de Garson, que utiliza la matriz de pesos entre la capa de entrada y oculta, y entre la oculta y la capa de salida, se discrimina la importancia relativa de las variables predictoras para una sola variable de respuesta. La contribución relativa de cada entrada (RI_j) está dada por la Ecuación 10 (García *et al.*, 2020).

$$RI_j = \left[\frac{\sum_{j=1}^n \left[\frac{|W_{ij} \cdot W_{jo}|}{\sum_{k=1}^m |W_{ij} \cdot W_{jo}|} \right]}{\sum_{k=1}^m \left[\sum_{j=1}^n \left[\frac{|W_{ij} \cdot W_{jo}|}{\sum_{k=1}^m |W_{ij} \cdot W_{jo}|} \right] \right]} \right] \cdot 100 \quad (10)$$

Donde: W_{ij} -pesos sinápticos de la conexión de entrada i a la capa oculta, W_{jo} -pesos sinápticos de la conexión capa oculta j a la salida, m - número de neuronas en la capa oculta y n - número de capas ocultas.

RESULTADOS Y DISCUSIÓN

Datos faltantes

Las variables climáticas diarias de la estación (32021) González Ortega, tuvieron 3.8 % de datos faltantes para precipitación, temperatura máxima y mínima y 2.2 % de datos faltantes para evaporación (Figura 1). Dado que los datos existentes están por arriba del 80 % que recomienda la organización meteorológica mundial (WMO, 2011), se construyeron series completas a partir de cada subperiodo homogéneo, utilizando "CLIMATOL 3.1.1" (Guijarro, 2018).

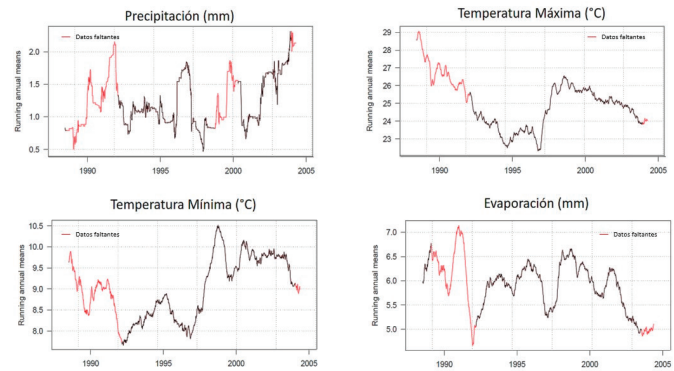


Figura 1. Estimación de datos faltantes de la Estación No. 32021 (1988 a 2005).

Figure 1. Estimation of missing data for Station No. 32021 (1988 to 2005).

Correlación de las variables de entrada y salida

Se consideraron las variables correlacionadas significativamente ($p < 0.05$), es decir variables con coeficientes absolutos mayores a 0.35 para ser utilizadas en los modelos de RNA para la predicción del Rendimiento (Rto), Superficie Cosechada (SC) y Producción (P). En Rto se observaron correlaciones positivas de 0.38, 0.44, 0.36, 0.62, 0.38, y 0.46 para Pp_Jul, Pp_Ago, Pp_Sep, SPp, Tmin_Jul, TTmin respectivamente y correlaciones negativas con valores -0.49 y -0.58 para Tmax_Sep y TTmax. Gourdji *et al.* (2015) en un trabajo sobre rendimiento de frijol y maíz en Nicaragua con datos históricos del clima encontraron una alta correlación entre rendimiento de frijol con la temperatura y precipitación. Wilson *et al.* (1995) mencionan que el rendimiento de los cultivos está afectado por precipitación, radiación solar y temperatura principalmente. Qian *et al.* (2006) obtuvo una correlación de 0.57 entre la radiación solar y evaporación con datos 1955-2000 en China. Por lo que se incluyó como entrada la SEva para generar el modelo RNA de Rto.

Para SC se observó una correlación negativa de -0.40, -0.59 y -0.55 para Eva_Ago, Eva_Sep y SEva y una alta correlación positiva de 0.71 con SS. Para P los valores absolutos fluctuaron entre 0.35, 0.36, 0.49, -0.40 -0.47 -0.42, -0.46, -0.5 y 0.39 para Pp_Ago, Pp_Sep, SPp, Tmax_Sep, Tmin_Jul, Eva_Jul, Eva_Sep, SEva y SS, respectivamente.

Con base en los resultados de correlación y el estudio exploratorio se definieron las variables de entrada de la siguiente manera: RTo → [Pp_Jul, Pp_Ago, Pp_Sep, TTmax, TTmin, SEva], SC → [SEva, SS] y P → [Pp_Jul, Pp_Ago, Pp_Sep, TTmax, TTmin, SEva, SS].

Modelos de Redes Neuronales Artificiales

RNA para predicción del rendimiento

Para el modelo de RNA rendimiento se utilizaron: Pp_Jul, Pp_Ago, Pp_Sep, TTmax, TTmin, SEva como variables de entradas y 13 neuronas en la capa oculta (Figura 2). La eficiencia fue de 0.74, la cual es un valor aceptable (> 0.73) de acuerdo con Demuth *et al.* (2017) y Pecar y Davis (2018). El conjunto de pesos óptimo que minimiza el error cuadrado medio, se obtuvo a las cinco iteraciones.

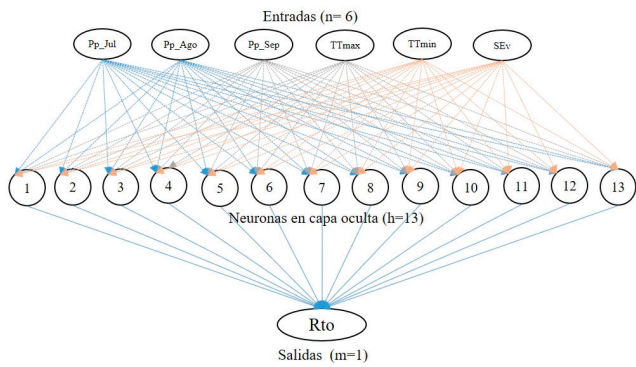


Figura 2. Arquitectura de la Red Neuronal Artificial para predecir rendimiento.

Figure 2. Artificial Neural Network architecture to predict bean yield.

En la Figura 3 se muestra una comparación entre el rendimiento de frijol observado y estimado por la red neuronal para la etapa de entrenamiento, validación y prueba (1988-2017) y evaluación (2018-2019). En la etapa de evaluación se obtuvo una eficiencia de 0.98 lo que significa que el modelo tiene una buena capacidad de simulación y predicción (Abraham *et al.*, 2020). La eficiencia global considerando la etapa de entrenamiento, validación y prueba y evaluación fue de 86 %.

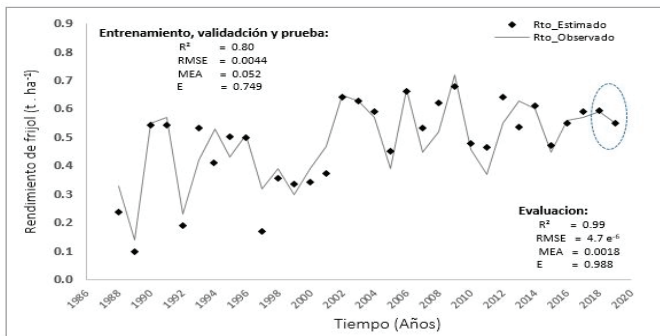


Figura 3. Rendimiento de frijol observado y predicho en la etapa de entrenamiento, validación, prueba y evaluación (círculo punteado) de la RNA.

Figure 3. Observed and predicted bean yield in the training, validation, testing and evaluation stage (dotted circle) of the ANN.

RNA para predicción de la superficie cosechada

En el modelo de RNA para la predicción de superficie cosechada se utilizaron 2 entradas (SEva, SS) y se varió el número de neuronas en la capa oculta (h) hasta obtener la máxima eficiencia con h=10, mostrando una eficiencia global de 0.85.

La eficiencia en la etapa de evaluación fue de 0.98, lo que significa que el modelo tiene una buena capacidad de simulación y predicción. La Figura 4 despliega el desempeño de la red para la variable de superficie cosechada del cultivo de frijol, la eficiencia global fue de 0.85 para las etapas de entrenamiento y evaluación. Un poco inferior a la presentada por Abraham *et al.* (2020) que utilizó redes autorregresivas no lineales con entrada externa-NARX para predecir rendimiento y superficie cosechada del cultivo de soja en Brasil.

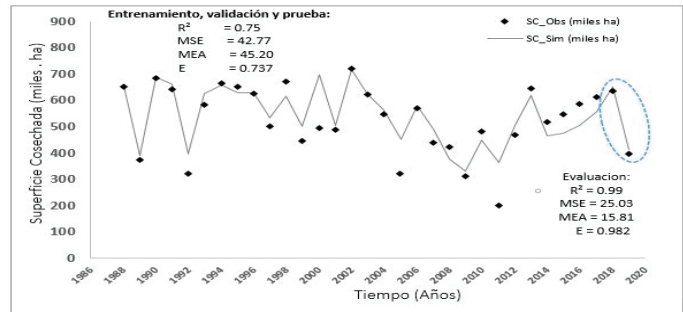


Figura 4. Superficie cosechada, observada y estimada para las etapas de entrenamiento, validación, prueba y evaluación (círculo punteado) de la RNA.

Figure 4. Harvested area, observed and estimated for the training, validation, testing and evaluation stages (dotted circle) of the ANN.

RNA para predicción de la producción

Para la producción de grano de frijol (Figura 5) se utilizaron: Pp_Jul, Pp_Ago, Pp_Sep, TTmax, TTmin, SEva, SS como variables de entrada y 15 neuronas en la capa oculta. La eficiencia encontrada fue de 0.64, en la etapa de entrenamiento, validación y prueba y en la etapa de evaluación la eficiencia fue de 0.65, valor menor a los aceptables (Demuth *et al.*, 2017; Pecar y Davis, 2018). El conjunto de pesos óptimo que minimiza el error cuadrado medio se obtuvo a las seis iteraciones. Debido al bajo desempeño de esta red la producción de frijol se estimó con las predicciones de superficie cosechada y rendimiento obtenidos con los modelos de RNA y se compararon con los datos obtenidos del SIAP (2019). La Figura 5 muestra la producción de 1988 a 2019 en miles de toneladas. Los coeficientes de correlación para la variable de producción fueron de 0.82 y 0.82 para rendimiento y superficie cosechada respectivamente.

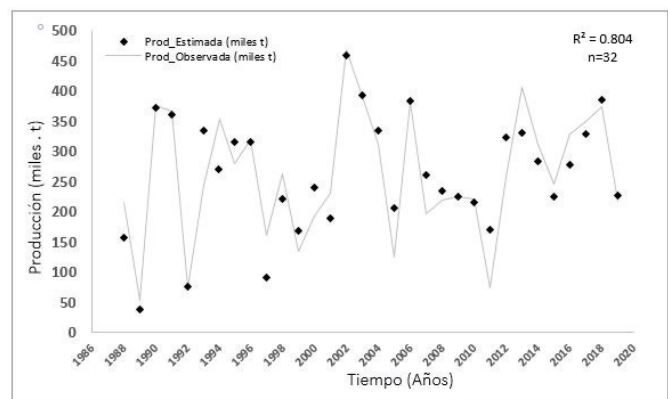


Figura 5. Producción de frijol estimada y observada en miles de toneladas para 32 años en el estado de Zacatecas.

Figure 5. Estimated and observed bean production in thousands of tons for 32 years in the state of Zacatecas.

Al estimar la producción con los modelos RNA se obtiene un R² = 80 %, y puede ser utilizada para simular la producción de frijol en el estado de Zacatecas con datos climáticos y superficie sembrada. Para las tres variables de salida simuladas, se observa un comportamiento estocástico

a través del tiempo al igual que las variables climáticas. Tal como lo afirma Geysler y Van (2001) acerca de la relación positiva entre la precipitación y la productividad de maíz. Otros trabajos citados por Gaytán *et al.* (2014) verifican el riesgo del clima en la producción de cultivos agrícolas para la cobertura de los seguros. De acuerdo con las medidas de desempeño de los modelos de RNA, el modelo para Rto es el que predice mejor en ambas etapas (entrenamiento, validación y prueba y evaluación).

Análisis de sensibilidad

La importancia de las variables predictoras para los modelos obtenidos se muestra en la Figura 6. En el modelo de RNA para simular el rendimiento de frijol, la variable más importante fue la sumatoria de la evaporación (SEva) del ciclo de producción seguida de la precipitación de agosto (Pp_Ago) y julio (Pp_Jul) con una importancia relativa de 21.59 %, 17.64 % y 17.41 % respectivamente. La variable con la mayor importancia relativa en la simulación de la superficie cosechada fue la superficie sembrada (SS) con un valor de 50.11 %, mientras que en el modelo de RNA para estimar la producción las variables más importantes fueron temperatura mínima del ciclo de producción (TTmin) y precipitación de agosto (Pp_Ago) con importancia relativa de 20.60 % y 20.21 %, respectivamente.

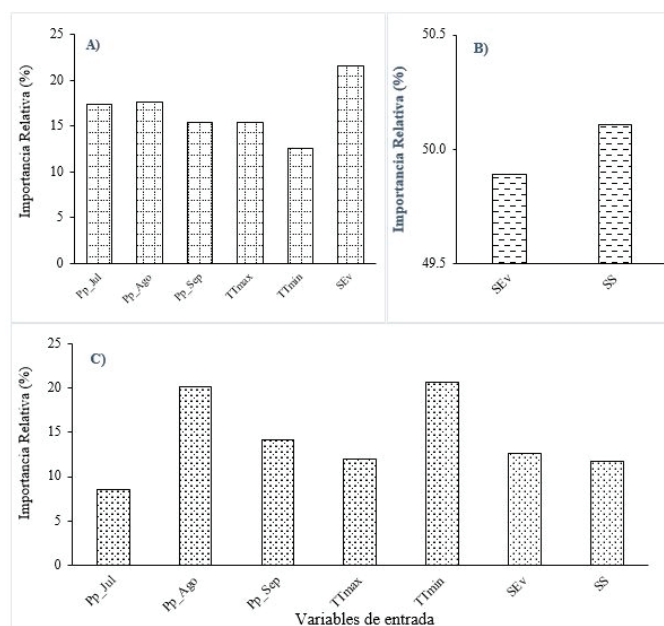


Figura 6. Análisis de sensibilidad para A) Rendimiento, B) Superficie cosechada y C) Producción de los modelos RNA para el cultivo de frijol en Zacatecas.

Figure 6. Sensitivity analysis for A) Yield, B) Harvested area and C) Production of RNA models for bean cultivation in Zacatecas.

Para los modelos de RTo y P la variable de mayor importancia fue la precipitación de agosto, cuando el cultivo está en la etapa de floración, esto coincide con lo reportado por Prieto *et al.* (2019) quienes hicieron un estudio del impacto del clima sobre el rendimiento de frijol en México y concluyeron que el agua es el factor limitante sobre todo en la etapa de floración.

CONCLUSIONES

Los modelos de redes neuronales artificiales generados para predicción de la superficie cosechada y rendimiento de frijol, utilizando variables climáticas y superficie de siembra, mostraron buen desempeño en la etapa de entrenamiento, validación y prueba para un conjunto de 32 años de datos, aunque se debe tener especial cuidado con las fuentes de información y contar con un mayor volumen de datos para que los modelos sean más confiables especialmente en la etapa de evaluación.

Los modelos de RNA generados podrán utilizarse con escenarios futuros de clima artificial para estimar la disponibilidad de grano y evaluación de riegos en los sistemas de producción.

La temperatura mínima del ciclo de producción y la precipitación observada en el mes de agosto, son las variables climatológicas con mayor influencia en la producción de frijol, el comportamiento de dichas variables puede ser un indicador para los productores acerca de la producción esperada. Mientras que la sumatoria de evaporación del ciclo de producción (SEva), fue la variable más importante en la predicción del rendimiento.

REFERENCIAS

- Abraham, E. R., Mendes dos Reis, J. G., Vendrametto, O., Oliveira Costa Neto, P. L. D., Carlo Tolo, R., Souza, A. E. D., y Oliveira Moraes, M. D. 2020. Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. *Agriculture*, 10(10), 475. <https://doi.org/10.3390/agriculture10100475>
- Allen, R. G., Pereira, L. S., Raes, D. y Smith, M. 2006. Evapotranspiración del cultivo: guía para la determinación de los requerimientos de agua de los cultivos. Estudio FAO Riego y Drenaje. FAO, Roma. 2006. 56-323 pp.
- Chatterjee, S., Dey, N., Sen, S. 2018. Soil moisture quantity prediction using optimized neural supported model for sustainable agricultural applications. *Sustain. Comput. Inform. Syst.* 100279. <https://doi.org/10.1016/j.suscom.2018.09.002>
- Crane, D. A. 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11), 114003.
- Demuth, H., Beale, M., Hagan, M. 2017. *Neural Network Toolbox User's Guide*; The MathWorks, Inc.: Natick, MA, USA, <http://www.mathworks.com/help/nnet/ug/dividedata-for-optimal-neural-network-training.html> (last review Mayo 2020)
- Esquivel, A. G., Sánchez, C. I., López, S. A., Velásquez, V. M. A., y Bueno, H. P. 2016. Enfoque metodológico para cuantificar el efecto del cambio en patrones de clima sobre el rendimiento del cultivo de frijol en el estado de Durango. *Revista Chapingo Serie Zonas Áridas*, 15(1), 17-28. <https://doi.org/10.5154/r.rchsa.2015.08.011>
- Fausett, L. V. 2006. *Fundamentals of neural networks: architectures, algorithms and applications*. Pearson Education India.
- Galindo, G. G. y Zandate, H. R. 2006. Adopción de variedades de frijol en el noroeste de zacatecas. *Terra Latinoamericana*, 24(1), 141-147.

- García, M. H.; Flores, M. H.; Ascencio, H., R.; Khalil, G. A.; Tijerina, C., L.; Mancilla, V. O.R.; Vázquez, P. M.A. 2020. Corn Grain Yield Estimation from Vegetation Indices, Canopy Cover, Plant Density, and a Neural Network Using Multispectral and RGB Images Acquired with Unmanned Aerial Vehicles. *Agriculture*, 10, 277. <https://doi.org/10.3390/agriculture10070277>
- Garg, B., Kirar, N.; Menon, S., Sah, T. 2016. A performance comparison of different back propagation neural networks methods for forecasting wheat production. *CSI Trans. ICT*, 4, 305–311. <https://doi.org/10.1007/s40012-016-0096-x>
- Gaytán, J. C. T., Acevedo, M. E. S., y Arias, J. Á. R. 2014. Modelación del clima bajo un proceso estocástico de reversión a la media estacional. *Estocástica: finanzas y riesgo*, 4(1), 9-32.
- Geyser, J. M., and van de Venter, T.W.G. 2001. "Hedging Maize Yield with Weather Derivatives". Working Paper 2001-13, University of Pretoria, Department of Agricultural Economics, Extension and Rural Development, 1-19 <http://purl.umn.edu/18067>
- Gogtay, N. J. y Thatte, U. M. 2017. Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3), 78-81. <https://pubmed.ncbi.nlm.nih.gov/29313581/>
- Gourdji, S., Läderach, P., Valle, A. M., Martínez, C. Z. y Lobell, D. B. 2015. Historical climate trends, deforestation, and maize and bean yields in Nicaragua. *Agricultural and Forest Meteorology*, 200, 270-281. <https://doi.org/10.1016/j.agrformet.2014.10.002>
- Guijarro, J.A. 2002. Tendencias de la precipitación en el litoral mediterráneo español. In *El Agua y el Clima*, Guijarro JA, Grimalt M, Laita M, Alonso S (ed). Asociación Española de Climatología: Palma de Mallorca; 237-246. <http://hdl.handle.net/20.500.11765/9143>
- Guijarro, J.A. 2018. Homogeneización de series climáticas con Climatol. Agencia Estatal de Meteorología (AEMET), DT en Islas Baleares, España. Versión, 3(1).
- Grageda, J. G., Lagunes, A. J., Rodríguez, J. C., Castillo, A. A. F., Chaparro, B. P., & Barrón, R. (2012). Uso de la información de una red de estaciones meteorológicas automatizadas en Sonora. *Biotecnia*, 14(1), 13-17. <https://doi.org/10.18633/bt.v14i1.115>
- Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. *Neural Network Design*. PWS Publishing, Boston, MA.
- Heng, L. K., T. Hsiao, S. Evett, T. Howell, and P. Steduto. 2009. Validating the FAO AquaCrop model for irrigated and water deficient field maize. *Agron. J.* 101(3): 488-498. <https://doi.org/10.2134/agronj2008.0029xs>.
- Intell, K. 2013. Special Issue on Artificial Intelligence in Agriculture. 78, Berlin. Springer-Verlag, 2013, Springer-Verlag Berlin Heidelberg. Vol. 43, pp 309-311. 0933-1875.
- Irigoyen, A. I., Della-Maggiara, A., & Angelocci, L. R. 2012. Análisis de los pesos sinápticos en redes neuronales artificiales entrenadas para estimar la lámina de agua en el suelo. XVII Congreso Brasileiro de MeteorologiaAt: Gramado, Brazil Volume: 1 September
- Khan, T., Qiu, J., Ali Qureshi, M.A., Iqbal, M.S., Mehmood, R., Hussain, W. 2020. Agricultural Fruit Prediction Using Deep Neural Networks. *Procedia Comput. Sci.* 174, 72–78. <https://doi.org/10.1016/j.procs.2020.06.058>
- Kolmogorov, A. N. 1957. On the representations of continuous functions of many variables by superpositions of continuous functions of one variable and addition. *Dokl. Akad. Nauk USSR* 114: 953 – 956. <http://mi.mathnet.ru/eng/dan/v114/i5/p953>
- Krause, P., D. P. Boyle, and F. Bäse. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5: 89-97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B. 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* 237, 111599. <https://doi.org/10.1016/j.rse.2019.111599>
- Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11(2): 431–441. <https://doi.org/10.1137/0111030>
- Medina, G., G., Rumayor, R. A. F., Cabañas, C. B., Luna, F. M.; Ruiz, C. J. A., Gallegos, V. C.; Madero, T. J., Gutiérrez, S., J. R.; Rubio, D. S. y Bravo, L. A. G. 2003. Potencial productivo de especies agrícolas en el estado de Zacatecas. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Centro de Investigación Regional Norte Centro. Calera de V. R., Zacatecas, México. (Libro Técnico Núm. 2). 15 7 p.
- Medina, G. G. y Ruiz, C. J. A. 2004. Estadísticas climatológicas básicas del estado de Zacatecas (Periodo 1961-2003). Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Centro de Investigación Regional Norte Centro. Campo Experimental Zacatecas. Calera de VR Zacatecas. (Libro Técnico No. 3).
- Medina, G. G., Báez, G. A. D. y Ramos, G. J. L. 2007. Red de monitoreo agroclimático del estado de Zacatecas. INIFAP, Campo Experimental Zacatecas, México. Desplegable informativa Núm. 15. 14 p. consultado mayo 2020 <http://www.zacatecas.inifap.gob.mx/historico.php?id=26786>
- Mohamed, Z.E. 2019. Using the artificial neural networks for prediction and validating solar radiation. *J. Egypt. Math. Soc.* 27, 47. <https://doi.org/10.1186/s42787-019-0043-8>
- Moore, H. 2012. *MATLAB®. MATLAB for Engineers, Third Edition*. Upper Saddle River, NJ: Pearson Education, Inc, 1-8.
- Stacey N., Friederike M., Hannes E., Naomi S. 2015. Economics of Land Degradation Initiative: Report for policy and decision makers_ Reaping economic and environmental benefits from sustainable land management. Bonn, Germany: ELD Initiative and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. <https://hdl.handle.net/20.500.11766/4881>
- Osuna, C. E. S., Figueroa, S. B., Martínez, G. M. Á., y Pimentel, L. J. 2019. Un sistema agroforestal de secano para el altiplano semiárido de México. *Revista mexicana de ciencias agrícolas*, 10(SPE22), 89-103. <https://doi.org/10.29312/remexca.v0i22.1861>
- Paulhus J. L. H, Kohler, M. A. 1952. Interpolation of missing precipitation records. *Month. Weath. Rev.*, 80:129-133. [https://doi.org/10.1175/1520-0493\(1952\)080<0129:IOMPR>2.0.CO2](https://doi.org/10.1175/1520-0493(1952)080<0129:IOMPR>2.0.CO2)
- Pecar, B., Davis, G. 2018. *Time Series Based Predictive Analytics Modelling: Using MS Excel, 3rd ed.*; Amazon Kindle: Seattle, WA, USA.
- Prieto, C. M. R., Matus, G. J. A., Gavi, R. F., Omaña, S. J. M., Brambila, P.J. J., Sánchez, E. J. y Martínez, D. M. A. 2019. Evolución de la superficie cultivada de frijol e impacto económico de la sequía sobre su rendimiento bajo condiciones de temporal en México. *Revista Fitotecnia Mexicana* 42:173-23 *Rev. Fitotec. Mex.* Vol. 43 (1) 2020 182, <https://doi.org/10.35196/rfm.2019.2.173-182>

- Qian, Y., Kaiser, D. P., Leung, L. R. y Xu, M. 2006. More frequent cloud-free sky and less surface solar radiation in China from 1955 to 2000. *Geophysical Research Letters*, 33(1). <https://doi.org/10.1029/2005GL024586>
- Raghuwanshi, N. S. y Wallender, W. W. 1998. Converting from pan evaporation to evapotranspiration. *Journal of irrigation and drainage engineering*, 124(5), 275-277.
- Ruiz-Del-Angel, E. O., Castilleja, R. D. C. V., Aguilar, J. C. R., & García, C. A. C. 2019. Modelo de requerimiento hídrico en un distrito de riego en México: incorporando escenarios de cambio climático. *Biotecnia*, 21(2), 129-136. <https://doi.org/10.18633/biotecnia.v21i2.946>
- Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk analysis*, 22(3), 579-590. <https://doi.org/10.1111/0272-4332.00040>
- Servicio Meteorológico Nacional (SMN). 2010. Mapa de climatología 1981-2010. <https://smn.conagua.gob.mx/es/climatologia/informacionclimatologica/informacion-estadistica-climatologica>
- SIAP. 2019. Servicio de Información y Estadística Agroalimentaria y Pesquera Información Agrícola. Avances mensuales por estado. <http://www.siap.sagarpa.gob.mx>.
- Silva M. G. 1998. Hidrología básica. Bogotá: Publicaciones Facultad de Ingeniería de la Universidad Nacional.
- Trethowan, R. M., Turner, M. A. y Chattha, T. M. 2010. Breeding strategies to adapt crops to a changing climate. In: Lobell, D. & Burke, M. (eds.), *Climate change and food security, advances in global change research*, 37, 175 - 154 https://doi.org/10.1007/978-90-481-2953-9_9
- Wilson, D. R.; Muchow, R. C. and Murgatoid, C. J. 1995. Model analysis of temperature and solar radiation limitations to maize potential productivity in cool climate. *Field Crops Res.* 43:1-18. [https://doi.org/10.1016/0378-4290\(95\)00037-Q](https://doi.org/10.1016/0378-4290(95)00037-Q)
- WMO. 2011. Guide to climatological practices. No. 100 https://www.wmo.int/pages/prog/wcp/ccl/guide/documents/WMO_100_en.pdf