

## Modelado de las Temperaturas del Aire a 850 milibares: un Potencial Indicador de las Ondas Cálidas en el Noroeste de México

Elvia Contreras-Navarro<sup>(\*)</sup>, O. Rafael García-Cueto<sup>(1)</sup>, Félix F. González-Navarro<sup>(1)</sup> y Ernesto A. Valenzuela-Palacios<sup>(2)</sup>

Universidad Autónoma de Baja California, (1) Instituto de Ingeniería, Blvd. Benito Juárez y Calle de la Normal s/n Col. Insurgentes Este, C.P. 21280, Mexicali, B.C., México, (2) Centro de Investigación Científica y Educación Superior de Ensenada. (e-mail: elvia.contreras@uabc.edu.mx; rafaelcueto@uabc.edu.mx; fernando.gonzalez@uabc.edu.mx; evalenzu@cicese.mx)

\*Autor a quien debe ser dirigida la correspondencia

*Recibido Ago. 18, 2015; Aceptado Oct. 16, 2015; Versión final Dic. 18, 2015, Publicado Abr. 2016*

---

### Resumen

Se modela la temperatura del aire a 850 milibares para determinar su potencial como indicador de las ondas cálidas en el noroeste de México. El análisis se realizó a nivel de meso-escala pero se enfocó para la ciudad de Mexicali, México. Se identificaron las principales variables que causan la formación de una onda cálida, y se propuso la temperatura del aire a 850 milibares como un indicador de su desarrollo. Esto se hace en función de ocho variables climáticas, utilizando seis algoritmos de inteligencia artificial. La técnica numérica de redes neuronales mostró un mejor desempeño, obteniéndose un coeficiente de regresión de 0.76 con un valor-p de 0.0019. Se concluye que este modelo no lineal es una herramienta prometedora que podría utilizarse en un sistema de alerta de este peligroso fenómeno atmosférico.

*Palabras clave: onda cálida; temperatura del aire; inteligencia artificial; redes neuronales.*

## Modeling Air Temperatures at 850 millibars: An Indicator Potential of Heatwaves in Northwest Mexico

### Abstract

The temperatures of atmospheric air at 850 millibars are modeled to determine their potential as an indicator of heat waves in northwest Mexico. The analysis was performed at meso-scale level but focused on the city of Mexicali, Mexico. The main variables that cause the formation of a heat wave are identified, and the modeling of air temperature at 850 millibars as an indicator of its development. This is done considering eight climate variables, using six artificial intelligence algorithms. The numerical technique of artificial neural networks showed a better performance, obtaining a regression coefficient of 0.76 with a p-value of 0.0019. It is concluded that this non-linear model is a promising tool that could be used in a warning system of this dangerous atmospheric phenomenon.

*Keywords: heat wave; air temperature; artificial intelligence; neural network*

## INTRODUCCIÓN

Las ondas cálidas (OC) son períodos de tiempo inusualmente cálido que afectan la salud humana mediante el estrés térmico y son capaces de exacerbar condiciones subyacentes, tales como enfermedades cardiovasculares, cerebro vasculares y respiratorias (Ebi y Meehl, 2007), lo que puede incrementar la incidencia de morbilidad y mortalidad en las poblaciones afectadas (Laxe et al., 2013). Aunque este tipo de eventos son de riesgo para toda la población, varios factores específicos como edad, ingreso, nivel de aislamiento social, trabajar sin aire acondicionado y vivir en apartamentos en pisos superiores, pueden incrementar el riesgo de ser afectado por el calor extremo (McMichael et al., 2008). La Organización Meteorológica Mundial (OMM) no ha podido definir este término completamente, ya que varían en carácter e impacto, aún en la misma localidad, y muchos países no tienen definiciones operacionales para alertar a la población; sin embargo, algunos autores concuerdan que tales eventos deberían incluir la excedencia de umbrales de temperatura local en el área de interés (Robinson, 2001; Tan et al., 2007), lo que posibilitaría identificar cambios en la frecuencia y duración de las ondas cálidas.

Estudios recientes (Gil y Jiménez, 2012; Sunyer, 2010; Blázquez, 2010; Tobías et al., 2010; Haines et al., 2005; Kysel, 2004) han analizado los efectos en la salud derivados de la exposición durante largo tiempo a altas temperaturas. Tal es el caso de los denominados golpes de calor, que se trata de la alteración más grave de la regulación térmica, que ocurre cuando la temperatura corporal rebasa límites fisiológicos asociándose con el aumento de mortalidad en todo el mundo (Morabito et al., 2012, Robin et al., 2008, Laidi, 2012, Chen et al. 2014). Los eventos cálidos de 2003 y 2010 en el este de Europa y gran parte de Rusia, clasificadas como "Mega- ondas de calor", probablemente rompieron records de 500 años de registros de temperaturas estacionales en más de la mitad de Europa. De acuerdo con experimentos regionales multi-modelo, la probabilidad de experimentar un verano similar se incrementará en un factor de 5 a 10 en los próximos 40 años (Barriopedro et al., 2011). En el 2003, Shanghai registró el verano más caliente en más de 50 años, esta onda de calor tuvo un efecto sustancial sobre la mortalidad, siendo las personas de edad avanzada (mayores de 65 años) las más vulnerables (Huang et al., 2010, Hongyun et al., 2014). Aunado a las consecuencias descritas, el principal efecto del cambio climático actual es el incremento de la temperatura del aire, y específicamente las ondas cálidas (Ashtiani et al., 2014; Raynal, 2011; Amador y Alfaro, 2009), por lo que es de suma importancia analizar con detalle los factores regionales que producen las ondas de calor para modelar su comportamiento, y en un futuro proponer un sistema de alerta temprana.

En particular el noroeste de México es especialmente vulnerable a los cambios del clima mundial y regional, pues de acuerdo a los escenarios que presenta el Panel Intergubernamental de Cambio Climático (IPCC, 2013) esta región tendrá un incremento en la temperatura media anual entre 1.5 y 2.5°C, y una disminución del 10 al 20% en su precipitación total anual para los próximos 50 años. Particularmente, en el estado de Baja California, México, hay estudios que evidencian alteraciones a eventos extremos (García-Cueto et al., 2013; Casillas et al., 2014); se ha observado un aumento en las ondas cálidas (García-Cueto et al., 2010) con un fuerte impacto en el sector salud ya que en el período 2004 a 2007 ocurrieron 43 defunciones por el golpe de calor en el municipio de Mexicali, México. En el caso de la ciudad de Mexicali, B.C., México, los umbrales de temperatura determinados por su afectación a la salud humana son que la temperatura máxima sea  $\geq 44^{\circ}\text{C}$ , y la temperatura mínima  $\geq 29^{\circ}\text{C}$ , al menos durante dos días consecutivos (Contreras-Navarro, 2010).

Por otra parte, el primer factor para que se desarrolle una onda cálida está relacionado a la situación sinóptica. Los eventos extremos de calor están generalmente asociados con sistemas de alta presión en niveles atmosféricos superiores que generan aire descendente a gran escala y calentamiento del mismo por compresión adiabática. Asimismo se ha observado en episodios de ondas cálidas en el noroeste de México y suroeste de Estados Unidos, que el centro de esta cuña de alta presión atmosférica se encuentra muy bien definido al nivel de 850 mb ( $\cong 1500$  msnm), por lo que la masa de aire caliente puede ser localizada también a este nivel atmosférico. Tratándose de un evento de características sinópticas la homogeneidad de temperaturas asociadas a esa masa de aire muy posiblemente podría indicar la fortaleza de la onda cálida a nivel de superficie. Gershunov et al., (2009), estudiaron la onda cálida del 2006 que impactó la costa de California y mencionan los potenciales factores causales que han ocasionado estos fenómenos con base en los eventos registrados en la historia de las ciudades de California y Nevada, considerando algunas variables climatológicas.

En cuanto a los algoritmos de inteligencia artificial se han utilizado en diversos estudios para modelar o predecir eventos ambientales; por ejemplo Shanket et al., (2008), desarrollaron modelos generales para la predicción de la temperatura del punto de rocío en 20 ubicaciones de Georgia, Estados Unidos, aplicando redes neuronales de retropropagación de tres capas y datos meteorológicos; estos modelos predijeron adecuadamente la temperatura del punto de rocío. Moustriset et al. (2013) realizaron pronóstico a

24 horas de los valores máximos y mínimos de la Temperatura Fisiológicamente Equivalente (PET) para el área metropolitana de Atenas, Grecia, aplicando redes neuronales artificiales (RNA). Los resultados del análisis mostraron que el estrés por calor extremo son 10,0% de las horas examinadas en el período cálido del año. En cuanto a la prognosis mediante el PET, las RNA's tienen una notable capacidad de predicción para predecir la PET diaria extrema de un día por adelantado, así como la persistencia de condiciones extremas durante el día. De igual manera Patil et al. (2013), usaron las redes neuronales para predecir los valores mensuales de la temperatura media de superficie, con una base de datos de 61 años de seis lugares diferentes de India obteniendo predicciones satisfactorias en todas las proyecciones. Así también, las RNA's son algunos de los modelos que utilizan los organismos encargados de los servicios de pronóstico meteorológico en Estados Unidos (Bonilla, 2006). Se ha utilizado también la combinación de algoritmos neuronales y de lógica difusa (modelos Neuro-difusos) para aumentar la certidumbre de la predicción (López-Cruz y Hernández-Larragoiti, 2010).

Así que siguiendo el enfoque de métodos de inteligencia artificial, en este artículo se utiliza el aprendizaje de máquina en el que se entrenan seis algoritmos de aprendizaje, como son las redes neuronales (NN), mínimos cuadrados parciales (PLS), modelo lineal generalizado (GLM), regresión RIDGE, modelo LASSO y árbol de clasificación y regresión (CART, por sus siglas en inglés), a fin de generar modelos de predicción de la temperatura del aire a 850 mb ( $\approx 1500$  msnm), lo que permitiría realizar una aproximación al conocimiento de las ondas cálidas, en función de las variables que propician su formación. La hipótesis es que la temperatura a 850 mb, es indicadora de la formación y evolución de las ondas cálidas. Una vez logrado lo anterior se perfilaría la predicción de este evento extremo y su información al sector salud con la finalidad de poner en marcha un programa de alerta a la población, y de esta manera disminuir el impacto en la morbilidad y mortalidad en la ciudad de Mexicali, México.

## METODOLOGÍA

### *Bases de datos*

Las ondas cálidas son fenómenos de escala sinóptica a meso-escala con impactos regionales como ha sido claramente observado, por ejemplo, en la extensión espacial de la onda de calor europea en 2003 (Beniston y Díaz, 2004; Stott et al., 2004), y la onda de calor del 2006 que afectó al noroeste de México y suroeste de Estados Unidos (García-Cueto et al., 2010). Por lo anterior, y dado que el estudio de Gershunov op. cit. (2009) se realizó en la frontera de los estados de California, Estados Unidos, y Baja California, México, que es donde se encuentra la ciudad de Mexicali (figura 1), se decidió utilizar algunos de los predictores utilizados en su estudio, añadiéndole otros predictores que parecen estar relacionados, particularmente con las ondas de calor 2006 y 2009 (NOAA, 2006). Las variables que se seleccionaron en este estudio son: 1) las temperaturas del aire ( $^{\circ}\text{C}$ ) a 850 milibares ( $T_{aire850}$ ), 2) Altura geopotencial (m) a 500 milibares ( $Alt_{Geop500}$ ), 3) Componente u del viento (m/s) a 500 milibares ( $V_{to} U_{500}$ ), 4) Componente v del viento (m/s) a 500 milibares ( $V_{to} V_{500}$ ), 5) Presión atmosférica (mb) al nivel del mar ( $Patm$ ), 6) Radiación Solar ( $\text{W}/\text{m}^2$ ) en superficie ( $RadSol$ ), 7) Humedad Relativa de superficie en % ( $HR$ ), 8) Temperatura Mínima de superficie en  $^{\circ}\text{C}$  ( $Tmin$ ) y 9) Temperatura Máxima de superficie ( $Tmax$ ). En este caso la modelación de la temperatura del aire a 850 milibares se puso en función de las otras ocho variables mencionadas anteriormente.

Los datos diarios para cada variable fueron descargados de la página de la Administración Nacional de la Atmósfera y el Océano (NOAA, por sus siglas en inglés) (<http://www.esrl.noaa.gov/>) eligiéndose el período del 1 al 31 de Julio de los años 2006, 2009 y 2012, ya quede acuerdo a García-Cueto op. cit. (2010) la mayor frecuencia de ondas cálidas (110 en total) en la ciudad de Mexicali, durante el período 1951-2006, han ocurrido durante el mes de julio, y en los meses de junio, agosto y septiembre, un total de 76, 73 y 25 ondas cálidas, respectivamente. Por añadidura, el mayor número de afectados por este evento cálido extremo ha sido en el mes de julio, ya que tan solo en la onda cálida del 2006, hubo 28 decesos. Se utilizó una malla de  $30^{\circ}$  a  $42^{\circ}$  de latitud norte y  $120^{\circ}$  a  $104^{\circ}$  de longitud oeste, que se consideró conveniente para obtener una imagen espacial adecuada del fenómeno obteniendo así una malla de  $12 \times 7$  puntos, por lo que, considerando las 9 variables a nivel diario del mes de julio, se dispuso de un conjunto de 61,752 observaciones. Para procesar los datos se utilizó el lenguaje de programación Python con las librerías Numpy y NetCDF.

### *Métodos utilizados*

Los métodos usados para realizar el estudio fueron los siguientes: a) Regresión con redes neuronales, b) algoritmo de mínimos cuadrados parciales, c) Modelo lineal generalizado, d) Regresión Ridge, e) Modelo Lasso, f) árbol de clasificación y regresión (CART).

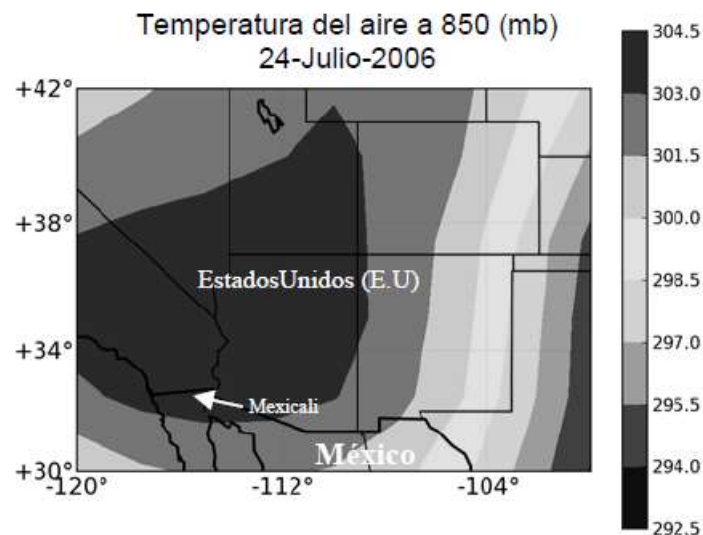


Fig 1: Malla de la temperatura del aire a 850 milibares (°K) del 24 de julio del 2006

### Regresión con redes neuronales (NN)

Los modelos neuronales se han empleado como modelos flexibles de regresión no lineal y de clasificación. Las redes neuronales pueden ser entendidas como modelos de regresión. Por lo tanto se emplean en muchas ocasiones como herramienta para predecir valores futuros de una o múltiples variables objetivo, que en estadística son variables de respuesta. Muchos métodos estadísticos clásicos y otros más recientes han sido reescritos, no siempre de forma consciente, como redes neuronales. Esto nos da idea de lo generales que pueden llegar a ser las estructuras representadas a través de un esquema de redes neuronales, y de su clara relación con la estadística (Castellano, 2009; Vázquez et al., 2012). El modelo de Red Neuronal utilizado en este estudio es del tipo *Feedforward-Backpropagation*, con una topología de dos capas internas de cinco neuronas cada una de ellas. La estrategia de aprendizaje de la red es el algoritmo de *Levenberg-Marquardt* que emplea una aproximación a la matriz Hessiana mediante el cuadrado de la matriz Jacobiana. Este modelo tiene la ventaja de no tener parámetros de ajuste, pues el gradiente de actualización de pesos es derivado del Jacobiano de la función de error y el vector de errores de predicción. Una vez entrenados los modelos, se utilizaron para hacer pruebas de predicción en el conjunto independiente como se detalla más adelante.

### Algoritmo de mínimos cuadrados parciales (PLS)

El método de mínimos cuadrados parciales (PLS) es una extensión del modelo de regresión lineal múltiple y se encuentra relacionado con el análisis de componentes principales. En su forma más simple un modelo lineal especifica la relación lineal entre la variable dependiente  $Y$ , y un grupo de variables predictivas  $X$  (Gonzalez y Alciaturi, 2012). El PLS se ha utilizado para predicción, p.e., Ibrahim et al. (2013), lo utilizó para pronosticar el nivel de agua en el río Galas mediante la adopción de la regresión lineal ordinaria y PLS.

### Modelo Lineal Generalizado (GLM)

Un modelo lineal generalizado (GLM) expresa de manera cualitativa la relación que existe en un conjunto de variables, donde existe solo una variable dependiente o de respuesta y el resto son variables independientes o explicativas (Dukić et al., 2012, Limet et al., 2012). El MLG consiste en lo siguiente: sea  $Y$  una variable aleatoria cuya función de distribución de probabilidad pertenece a una familia de distribuciones de probabilidades  $H$ , y es explicada por un conjunto de variables  $X$ , las cuales se fijan antes de conocer  $Y$ . La expresión matemática de este modelo se describe en la ecuación 1:

$$E\left(\frac{Y}{X_1, X_2, \dots, X_k}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k = \mu \quad (1)$$

### Regresión Ridge

Es un algoritmo de máxima probabilidad para modelos mixtos con un único componente de varianza, además del error residual, lo que permite la predicción eficiente con datos de entrenamiento no replicados

(Endelman, 2011); la expresión matemática de este algoritmo está dada por la ecuación (2). Piepho et al., (2012), revisaron varias opciones para el cálculo de regresión ridge la selección genómica y compararon su eficiencia computacional cuando se utiliza un paquete de modelo mixto. La regresión ridge disminuye los coeficientes de regresión. Los coeficientes más altos se minimizan con una suma de residual de cuadrados penalizado.

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^P \beta_j^2 \right\}. \quad (2)$$

Aquí  $\lambda \geq 0$  es un parámetro de control: cuanto mayor es el valor de  $\lambda$ , mayor es la cantidad de contracción de los coeficientes. La idea de penalizar con los cuadrados la suma de los parámetros también se utiliza en las redes neuronales, donde se conoce como la descomposición de peso.

#### Modelo Lasso

Hebiri et al. (2013), argumentaron que la predicción Lasso funciona bien para cualquier grado de correlación si se eligen los parámetros de ajuste adecuados. Lasso es un método de contracción con diferencias sutiles pero importantes. La estimación Lasso está definida por la ecuación (3):

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2 \text{ subject to } \lambda \sum_{j=1}^P |\beta_j| \leq t \right\}. \quad (3)$$

Al igual que en la regresión Ridge, se puede volver a parametrizar el coeficiente  $\beta_0$  constante mediante la estandarización de los predictores, la solución para  $\beta_0$  es  $\bar{Y}$ , por lo que se tiene un modelo sin intercepto.

#### Modelo árbol de clasificación y regresión (CART)

Este modelo consiste en un conjunto de reglas determinadas por un procedimiento de ajuste mediante particiones binarias recursivas, en el que un conjunto de datos es sucesivamente particionado. Esta técnica está relacionada con los conglomerados divisivos, en los que inicialmente todos los objetivos son considerados como pertenecientes al mismo grupo. En cada instancia de separación el algoritmo analiza todas las variables predictoras y selecciona, para realizar la partición binaria de los datos, lo que permite crear subgrupos o nodos más homogéneos dentro y más heterogéneos entre ellos (Soleana et al, 2010).

## MÉTODO DE EVALUACIÓN E INTERCOMPARACION DE MODELOS

#### Medida de validación y normalización de los datos

El método de retención es el tipo más simple de validación cruzada. El conjunto de datos se divide en dos grupos, llamado el conjunto de entrenamiento y el conjunto de pruebas. Los errores que hace se acumulan como antes para dar el error medio de prueba absoluta, que se utiliza para evaluar el modelo. La ventaja de este método es que por lo general es preferible el método residual. Sin embargo, su evaluación puede tener una alta varianza. La evaluación puede depender en gran medida de qué puntos de datos terminan en el conjunto de entrenamiento y qué terminan en la prueba, y por lo tanto la evaluación puede ser muy diferente en función de cómo se hace la división. En el caso bajo estudio se seleccionó el 70% de los datos de entrada a cada uno de los seis algoritmos para entrenar a los modelos mediante validación cruzada en su modalidad de 10 repeticiones por 10 divisiones. Por lo consiguiente se dejó un 30% de las observaciones para tener un conjunto de datos, independiente de los utilizados para la modelación, la cual se considera suficientemente grande para verificar el desempeño de los modelos de predicción.

La medida de evaluación de la validación cruzada fue la Raíz del Error Cuadrado Medio Normalizado (NRMSE, por sus siglas en inglés), que se define como:

$$NRMSE = \sqrt{\frac{1/n \sum_{i=1}^n (\text{observado} - \text{predicho})^2}{\frac{1}{n} \sum_{i=1}^n (\text{observado} - \overline{\text{predicho}})^2}} \quad (4)$$

Para introducir los datos se aplicó la normalización de los datos requerida según las reglas de cada algoritmo. Para los algoritmos de Mínimos cuadrados parciales (PLS), RIDGE, Modelo lineal generalizado (GLM) Y LASSO, se utilizaron la media igual a 0 y desviación estándar igual a 1 como entradas para la predicción. En el caso del algoritmo CART los datos se introdujeron directamente. Y para el caso de la red neuronal (NN) se escalan en un intervalo de [0,1]. Posteriormente se extrajo de forma aleatoria el 30% de las observaciones, a fin de generar un conjunto de datos de prueba independiente y poder verificar el rendimiento de los modelos de predicción.

#### *Prueba de los rangos con signo de Wilcoxon*

La prueba de los rangos con signo de Wilcoxon es una prueba no paramétrica para comparar la mediana de dos muestras relacionadas y determinar si existen diferencias entre ellas (Wilcoxon, 1945).

El estadístico de la prueba de los signos de Wilcoxon,  $W^+$ , es

$$W^+ = \sum R_i, \approx i > 0 \quad (5)$$

La suma de los rangos  $R_i$  corresponden a los valores positivos de  $Z_i$ . La distribución del estadístico  $W^+$  puede consultarse en tablas para determinar si se acepta o no la hipótesis nula.

## RESULTADOS

En la figura 2 se observa la distribución de los datos mediante diagramas de dispersión destacando el modelo NN por tener la menor dispersión. La calidad del modelo se verificó mediante el coeficiente de correlación (R), el coeficiente de determinación ( $R^2$ ) y la varianza (VAR) de cada modelo, cuyos valores se presentan en la tabla 2. Destaca el modelo NN como el óptimo de los 6 presentados.

De acuerdo a la figura 2 se observa que conforme la temperatura es más alta la dispersión va siendo menor. Si se analizan valores superiores a 300°K (26.85°C) la diferencia entre los datos observados y predichos son del orden de  $\pm 3^\circ\text{K}$ , mientras que valores menores a este límite (300°K) son del orden de  $\pm 5^\circ\text{K}$ ; lo anterior indica que la identificación de potenciales ondas cálidas mediante la modelación propuesta parece trabajar en la dirección correcta, ya que los valores más altos de temperatura están asociados con eventos cálidos extremos.

La tabla 1 muestra que las Redes Neuronales, en comparación con los otros métodos, explican mejor a la temperatura del aire a 850 mb con una  $R^2$  de 0.76, y una menor variabilidad, estimada por la varianza, con un valor de 4.95. En la figura 3 se observa el diagrama de dispersión del análisis aplicando regresión con redes neuronales. El modelo lineal generado por el método de Mínimos Cuadrados Parciales, tiene un desempeño un tanto similar que la Red pero con una  $R^2$  menor en .09.

Tabla 1: Comparación de los modelos con  $R^2$  y Varianza

MODELO	$R^2$	VAR
CART	0.667	7,11749003
PLS	0.675	6,76779143
RIDGE	0.675	6,91148692
GLM	0.675	6,7598003
LASSO	0.677	6,77608221
NN	0.766	4,9524449

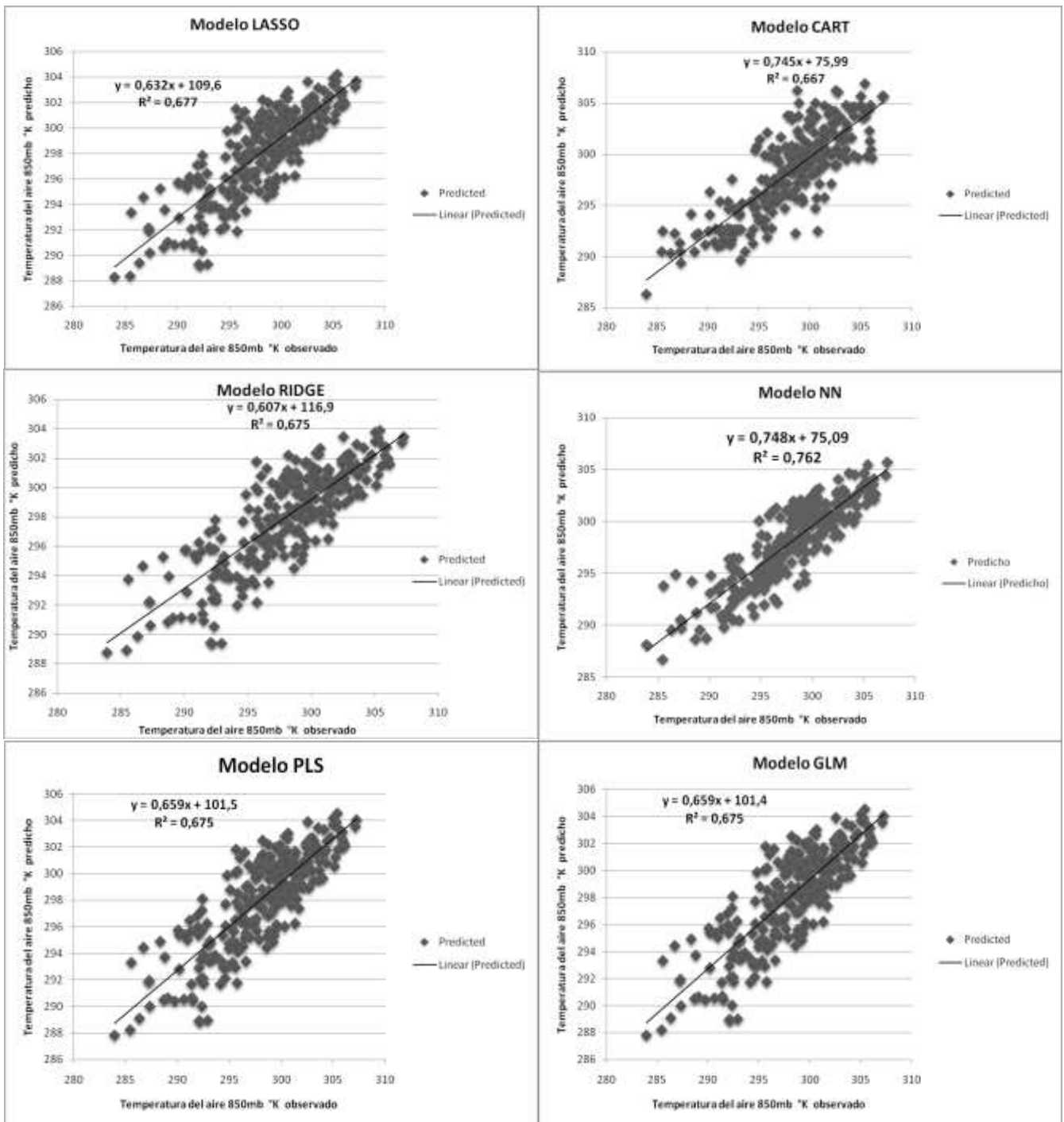


Fig 2: Diagramas de dispersion de modelos utilizados

**Validación de los modelos**

Los errores en la validación (NRMSE) son bajos y una variabilidad, estimada por el error estándar, también consistentemente pequeños, i.e. un error estándar menor a 0.01 para los modelos (tabla 2).

Tabla 2: Resultados del proceso de entrenamiento de modelos por validación cruzada. NRMSE=Raíz del Error Cuadrado Medio Normalizado.

Medidas estadística	Mínimos Cuadrados Parciales(PLS)	Red Neuronal (NN)	Modelo lineal generalizado(GLM)	LASSO	RIDGE
NRMSE	0,574	0,508	0,574	0,576	0,578
Error estándar	0,001	0,009	0,001	0,001	0,001
Límite inferior	0,572	0,497	0,572	0,574	0,577
Límite superior	0,577	0,529	0,577	0,579	0,581

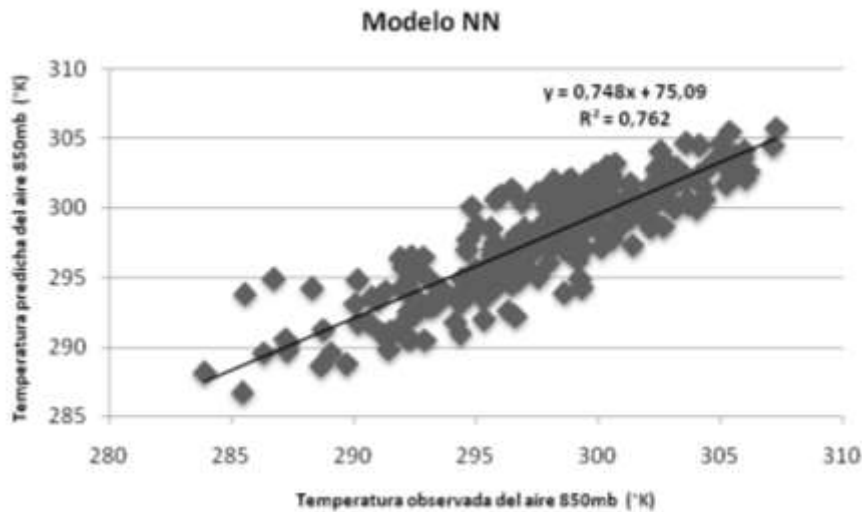


Fig. 3: Análisis de regresión simple de la predicción generada por la red neuronal vs los datos observados en la muestra independiente de prueba.

El modelo lineal generado, por ejemplo, con el método de Mínimos Cuadrados Parciales (PLS), tiene un desempeño un tanto similar que la Red Neuronal (NN) pero con una  $R^2$  menor en 0.09. Ante esto, el test de suma de ranqueo de Wilcoxon fue aplicado para fijar esta diferencia de desempeño (Tabla 3). El p-value obtenido fue de 0.0019 por lo que se puede afirmar que para el problema en particular, un modelo no lineal como la Red Neuronal tiene un mejor desempeño en la predicción de valores independientes que un modelo lineal.

La figura 4 corresponde a una extracción aleatoria del 30% de las observaciones que se dejaron para validar los datos, así que es un conjunto de datos independiente de los que se utilizaron para ajustar cada uno de los modelos. Aun cuando las ondas cálidas no se pueden determinar directamente de la figura 4, la hipótesis propuesta de la asociación entre las temperaturas a 850 mb y las variables independientes elegidas parece ser aceptada con los escenarios históricos de ondas cálidas seleccionadas. Por tanto la técnica de redes neuronales parece ser una técnica satisfactoria que podría ser utilizada para la predicción de las ondas cálidas a nivel local.

Tabla 3-: Resultados del test de ranqueo de Wilcoxon aplicado a cada modelo

TEST WILCOXON	PLS VS NN	GLM VS NN	LASSO VS NN	CART VS NN	RIDGE VS NN
P	0.8568	0.8559	0.7151	0.9445	0.5838
H	0	0	0	0	0

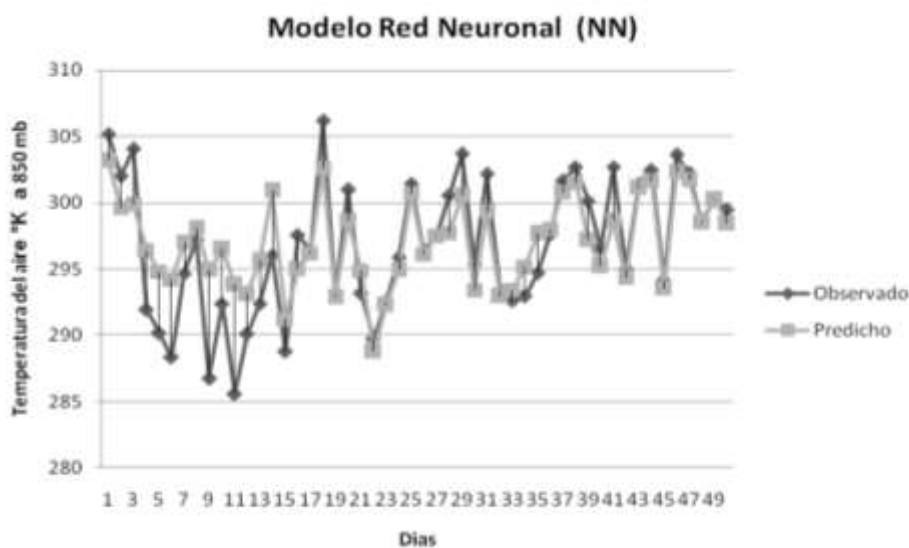


Fig. 4: Ventana de 50 predicciones de la red neuronal vs datos observados en la muestra independiente de prueba.



## DISCUSIÓN

La predicción de eventos climáticos extremos, como las ondas cálidas, es uno de los desafíos actuales que enfrenta la comunidad científica. La dinámica de esos fenómenos climáticos no es fácilmente incluida en los modelos actuales de predicción numérica del tiempo debido a su pequeña escala y al entendimiento limitado que aún tenemos de ellos. Las técnicas de inteligencia artificial, y particularmente las redes neuronales, se han utilizado en una amplia variedad de tareas, que se pueden categorizar en predicción, aproximación de funciones y clasificación de patrones. En nuestro caso se está utilizando en el área de predicción, en conjunto con otros cinco algoritmos. En base a los resultados obtenidos se observa que los modelos PLS, GLM, RIDGE y LASSO lo hacen de una manera aceptable con una  $R^2$  de 0.67 y una varianza de 6.7; el modelo CART es el de peor desempeño con un  $R^2$  de 0.66 y varianza de 7.1, mientras que las NN con una  $R^2$  de 0.76 y una varianza de 4.9, parecen ser las de mejor implementación en base a estos estadísticos. Los datos predichos, mostrados en la figura 4, que se ajustan a los datos observados dan confianza en las Redes Neuronales como el mejor modelador de los algoritmos seleccionados. Se debe comentar, aunque es obvio de la figura 4, que los primeros días de la simulación, del día 3 al día 14, no sigue el comportamiento de una manera satisfactoria lo que lleva a pensar que es necesario recolectar más datos de las mismas variables para disminuir esas discrepancias considerando no solamente el mes de julio, sino los tres meses de verano, con la finalidad de tener un antecedente del comportamiento durante toda esa estación, al menos de los últimos 15 años, para lograr una visión más amplia del comportamiento del fenómeno.

Para reforzar el párrafo anterior, y mediante la prueba de rangos de Wilcoxon se comparó cada modelo contra las NN y se observa que el modelo RIDGE mostró menor diferencia con  $p=0.58$ , y el modelo CART la mayor diferencia con  $p=0.94$ , confirmando ser el modelo de menor desempeño del resto. Una de las limitantes de estos modelos inteligentes es que no son fáciles de implementar y aplicar; sin embargo, hay información disponible que puede asistir en el proceso, y que permite que errores comunes puedan evitarse. Existe software comercialmente disponible para darle solución a estos algoritmos, sin embargo, un conocimiento básico del problema por resolver es útil, aunque el trabajo interdisciplinario, como el realizado en este estudio ha sido fundamental para su solución.

Respecto a la ponderación de variables predictoras se observa, partir del análisis de correlaciones parciales, que la altura geopotencial a 500 mb y la presión atmosférica de superficie son las más importantes en la explicación de la temperatura del aire a 850 mb, en menor medida la radiación solar superficial, y se podría prescindir de las temperaturas del aire, máxima y mínima, y humedad relativa, por el valor tan pequeño del coeficiente de correlación parcial

Tabla 4: Correlación parcial de variables climatológicas

<i>Variable Climatológica</i>	<i>Correlación Parcial</i>
Altura Geopotencial 500mb	0,65
Temperatura Maxima	0,03
Humedad Relativa %	-0,01
Temperatura Mínima	-0,02
Radiación Solar	-0,11
Presión atmosférica	-0,43

Hasta el momento se tiene un notable avance en el estudio de algunos indicadores que propician el desarrollo de las ondas cálidas en la ciudad de Mexicali, el siguiente paso en este sentido será la combinación de RN y lógica difusa o máquinas de soporte vectorial, con el fin de lograr un sistema de alerta temprana para la población logrando así disminuir la morbilidad y mortalidad por las ondas cálidas en Mexicali B.C.

## CONCLUSIONES

En este estudio se identificaron las principales variables que causan el desarrollo de una onda cálida, y se propuso la modelación de la temperatura del aire a 850 mb ( $T_{aire_{850}}$ ) en función de varias variables climáticas, mediante varias técnicas de inteligencia artificial, entre ellas las Redes Neuronales que mostraron un mejor desempeño obteniéndose un coeficiente de determinación de 0.76, con una significancia estadística de 0.0019, por lo que se concluye que este modelo tiene un mejor desempeño en comparación con las demás técnicas. Las variables predictoras más importantes en la modelación propuesta fueron la altura geopotencial a 500 mb y la presión atmosférica de superficie. El siguiente paso en este esquema de modelación será mejorar la predicción de las ondas cálidas, mediante la combinación de algoritmos inteligentes de las Redes neuronales y máquinas de soporte vectorial o lógica difusa, para que en un futuro se logre un sistema de alerta temprana.

## REFERENCIAS

- Ashtiani, A., Mirzaei, P. A., y Haghghat, F., *Indoor Thermal Condition In Urban Heat Island: Comparison Of The Artificial Neural Network And Regression Methods Prediction*. Energy And Buildings, 76, 597-604 (2014)
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., y García-Herrera, R., *The hot summer of 2010: redrawing the temperature record map of Europe*. Science, 332(6026), 220-224. (2011)
- Beniston, M., y H. F. Diaz, *The 2003 heat wave as an example of summers in a greenhouse climate? Observations and climate model simulations for Basel, Switzerland*. Global Planet. Change, 44, 73–81. (2004)
- Bonilla J. E., Ramirez J., Ramirez O., *Metodología Para El Diseño De Un Modelo Univariado De Red Neuronal Para El Pronóstico De La Temperatura Mínima En La Zona De Mosquera*. Cundinamarca, Colombia. 10 pp 111-120., (2006)
- Casillas H. A., García-Cueto R., Camacho O., González F. F. *Detección de la Isla Urbana de Calor mediante Modelado Dinámico en Mexicali, BC, México*. Información Tecnológica, 25(1), 139-150. (2014)
- Chen, D., Wang, X., Thatcher, M., Barnett, G., Kachenko, A., y Prince, R., *Urban vegetation for reducing heat related mortality*. Environmental pollution, 192, 275-284, (2014)
- Contreras-Navarro E. "Las ondas cálidas y su impacto en la salud: análisis y evolución histórica en Mexicali B.C. México". Instituto de Ingeniería, UABC, 91 páginas (2010)
- Dukić, V., Hayden, M., Forgor, A. A., Hopson, T., Akweongo, P., Hodgson, A., y Trzaska, S., *The Role Of Weather In Meningitis Outbreaks In Navrongo, Ghana: A Generalized Additive Modeling Approach*. Journal Of Agricultural, Biological, And Environmental Statistics, 17(3), 442-460. (2012)
- Ebi LK, Meehl AG. *Heatwaves & global climate change. The heat is on: climate change & heatwaves in the Midwest*. Pew Center on Global Climate Change, Arlington, p 14. (2007)
- Endelman, J. B., *Ridge Regression And Other Kernels For Genomic Selection With R Package Rrblup*. The Plant Genome, 4(3), 250-255. (2011)
- García Cueto R O. , Tejeda M. A. T., y Ostos, E. J. ., *Heatwaves And Heatdays In An Arid City In The Northwest Of Mexico: Current Trends And In Climate Change Scenarios*. International J. Biometeorol., (54) Pp 36-46. (2010)
- García-Cueto, O. R., Cavazos, M. T., de Grau, P., y Santillán-Soto, N., *Analysis and modeling of extreme temperatures in several cities in northwestern Mexico under climate change conditions*. Theoretical Applied Climatology 112(3-4).2013
- García-Cueto, O.R., Tejeda, M.A., Jáuregui, E., *Heat waves and heat days in an arid city in the northwest of México: current trends and in climate change scenarios*. Int J Biometeorol 54:335–345. (2010)
- Gershunov, A., Cayan, D. R., y Iacobellis, S. F., *The great 2006 heat wave over California and Nevada: signal of an increasing trend*. Journal of Climate, 22(23), 6181-6203. (2009)
- Gil, Cristina L., y Julio D.J., *Cómo Se Analiza El Impacto De Las Olas De Calor Sobre La Mortalidad A Partir De Series De Datos Estadísticos* 25. Índice: Revista De Estadística Y Sociedad, (50), 25-26. (2012)
- González, P., y Alciaturi, C., *Desarrollo De Un Programa Para Estudiar El Comportamiento De Una Columna De Fraccionamiento Etano/Etileno De Una Planta De Olefinas*. Development Of A Program For Studying The Behavior Of An Ethane - Ethylene Splitter Of An Olefins Plant, 11 89-99pp, (2012)
- Haines, A., Kovats, R. S., Campbell-Lendrum, D., y Corvalán, C., *Climate Change And Human Health: Impacts, Vulnerability And Public Health*. Public Health, Vol 120, 585–596pp. (2006)
- Hebiri, M., y Lederer, J, *How Correlations Influence Lasso Prediction*. Information Theory, IEEE Transactions On, 59(3), 1846-1854. (2013)
- Huang, W., Kan, H., y Kovats, S., *The impact of the 2003 heat wave on mortality in Shanghai, China*. Science of the total environment, 408(11), 2418-2420. (2010)

- Ibrahim, N., & Wibowo, A., *Partial Least Squares Regression Based Variables Selection For Water Level Predictions*. American Journal Of Applied Sciences, 10(4), 322. (2013)
- Jorge A. A., y Erick J. A., *Métodos De Reducción De Escala: Aplicaciones Al Tiempo, Clima, Variabilidad Climática Y Cambio Climático*. Revista Iberoamericana De Economía Ecológica (Revibec), (11), 39-52. (2009)
- Kysely, J., *Mortality And Displaced Mortality During Heat Waves In The Czech Republic*. Int J Biometeorol 49:91–97 Pp. (2004)
- Laidi, M., Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E., y Beaudou, P., *The impact of heat islands on mortality in Paris during the August 2003 heat wave*. Environmental health perspectives, 120(2), 254. (2012)
- Laxe, S., Zúniga-Inestroza, L., y Bernabeu-Guitart, M., *Manifestaciones Neurológicas Y Su Impacto Funcional En Sujetos Que Han Padecido Un Golpe De Calor*. Revista De Neurología, 56(1), 19-24. (2013)
- Lim Y. H., Park A, K., Kim H., *Modifiers Of Diurnal Temperature Range And Mortality Association In Six Korean Cities*. International Journal Of Biometeorology, 56(1), 33-42. (2012)
- López C. I. L., y Hernández L. L., *Modelos Neuro-Difusos Para Temperatura Y Humedad Del Aire En Invernaderos Tipo Cenital Y Capilla En El Centro De México*. Agrobiencia, 44(7), 791-805. (2010)
- Ma, H., Shao, H., y Song, J., *Modeling the relative roles of the foehn wind and urban expansion in the 2002 Beijing heat wave and possible mitigation by high reflective roofs*. Meteorology and Atmospheric Physics, 123(3-4), 105-114. (2014)
- McMichael A.J., y otros 16 autores, *International study of temperature, heat and urban mortality: the 'ISOTHURM' project*. Int J Epidemiol 1–11. (2008)
- Méndez M. C., *Modelización Estadística Con Redes Neuronales. Aplicaciones A La Hidrología, Aerobiología Y Modelización De Procesos*. Tesis Doctoral, Universidad De Da Coruña. (2009)
- Morabito, M., Profili, F., Crisci, A., Francesconi, P., Gensini, G. F., y Orlandini, S., *Heat-related mortality in the Florentine area (Italy) before and after the exceptional 2003 heat wave in Europe: an improved public health response?*. International journal of biometeorology, 56(5), 801-810. (2012)
- Moustris, K. P., Nastos, P. T., y Paliatsos, A. G. *One-Day Prediction Of Biometeorological Conditions In A Mediterranean Urban Environment Using Artificial Neural Networks Modeling*. Advances in Meteorology Article ID 538508, 15 pages Volume 2013
- NOAA Earth System Research Laboratory / Physical Sciences Division, 2006: Daily Mean Composites. <http://www.cdc.noaa.gov/Composites/Day/>, último acceso Marzo 20, 2015
- Patil K., Deo M. C., Ghosh S., y Ravichandran M., *Predicting Sea Surface Temperatures In The North Indian Ocean With Nonlinear Autoregressive Neural Networks*. International Journal Of Oceanography, vol.2013 1-10pp (2013)
- Piepho, H. P., Ogotu, J. O., Schulz-Streeck, T., Estaghirou, B., Gordillo, A., y Technow, F., *Efficient Computation Of Ridge-Regression Best Linear Unbiased Prediction In Genomic Selection In Plant Breeding*. Crop Science, 52(3), 1093-1104. (2012)
- Raynal Villaseñor, J. A., *Cambio Climático Global: Una Realidad Inequívoca Global Climate Change: An Unequivocal Reality*. Departamento De Ingeniería Civil Y Ambiental Escuela De Ingeniería Universidad De Las Américas, Puebla. Ing. Invest. Y Tecnol. Vol. 12(4), (2011)
- Robin, J. M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J. P., y Herrmann, F. R., *Death toll exceeded 70,000 in Europe during the summer of 2003*. Comptesrendusbiologies, 331(2), 171-178. (2008)
- Robinson PJ., *On the definition of a Heat Wave*. J Appl Meteor 40:762–775. (2001)
- School Of Computer Science <Http://Www.Cs.Cmu.Edu/~Schneide/Tut5/Node42.Html>. Acceso en 18 de diciembre , (2014)

- Shank, D. B., Hoogenboom, G., y McClendon, R. W., *Dewpoint Temperature Prediction Using Artificial Neural Networks*. Journal Of Applied Meteorology And Climatology, 47(6), 1757-1769. (2008)
- Soleana R. H., Cecilia B., Mónica B., *Identificación De Relaciones Entre Rendimientos Y Variables Ambientales Vía Árboles De Clasificación Y Regresión (Cart)* Interciencia , 35 (12) 2010
- Stott, P. A., D. A. Stone, y M. R. Allen. *Human contribution to the European heatwave of 2003*. Nature, 432, 610–613. (2004)
- Sunyer, J. *Promoción De La Salud Frente Al Cambio Climático*. Gac Sanit, 24(2), 101-102. (2010)
- Tan J, Youfei Z, Song G, Kalkstein L.S., Kalkstein AJ, Tang X., *Heat wave impacts on mortality in Shanghai, 1998 and 2003*. Int J Biometeorol 51:193–200. (2007)
- Tobías, A., de Olalla, P. G., Linares, C., Bleda, M. J., Caylà, J. A., y Díaz, J *Short-Term Effects Of Extreme Hot Summer Temperatures On Total Daily Mortality In Barcelona, Spain*. Int J Biometeorol. (2010)
- Vázquez, C., Beristain, F., y Audelo, E., *Caracterización del Patrón de Comportamiento de las Partículas en Suspensión en el Valle de México*. Información tecnológica, 23(3), 3-10, (2012)
- Wilcoxon, F. *"Individual Comparisons by Ranking Methods."* Biometrics 1, 80-83. (1945)