

Aplicación del enfoque de minería de datos en la variabilidad del clima en el municipio de ciudad Serdán, Puebla, México

María de los Ángeles Velasco-Hernández
Benemérita Universidad Autónoma de Puebla
ambientalbuap@yahoo.com.mx

Tomás Morales-Acoltzi
Universidad Nacional Autónoma de México
acoltzi@atmosfera.unam.mx

Néstor Gabriel Estrella Chulim
Colegio de Postgraduados
nestrella@colpos.mx

Resumen

¿Existe alguna recomendación para este nuevo milenio, en relación a la investigación? En la “*Conferencia Mundial Sobre la Ciencia*” con sede en Budapest, Hungría, julio 26 de 1999 plantearon: “Declaración Sobre la Ciencia y el Uso del Saber Científico”. Destacando “La ciencia constituye un poderoso instrumento, para comprender los **sistemas naturales y sociales**, que desempeñarán probablemente un papel aún más importante en un futuro inmediato, a medida que se conozca mejor la **complejidad** de las relaciones entre la sociedad y el medio natural”. La Minería de Datos (**MD**) se aplica cuando la complejidad de los sistemas es relativamente grande. Hacemos énfasis en que la **MD** es un proceso que invierte la dinámica del método científico (**MC**) así: En el **MC**, primero se formula la hipótesis y luego se diseña el experimento para coleccionar los datos que confirmen o refuten la hipótesis. Si se hace con la formalidad adecuada, cuidando cuáles son las variables controladas y cuáles experimentales, se obtiene un nuevo conocimiento. Por el contrario, en la **MD**, se coleccionan los datos y esperamos extraer hipótesis de ellos. Queremos que los datos nos describan o indiquen por qué son como son. Validar esa hipótesis inspirada por los datos en los datos mismos, sería numéricamente significativa, pero experimentalmente

inválida. De ahí que la **MD** debe presentar un enfoque exploratorio, y no confirmador. Aplicamos la **MD** para identificar la variabilidad del clima en Ciudad Serdán.

Palabras Clave: complejidad, cambio climático

Introducción

En Serdán, los productores acostumbran sembrar variedades criollas de maíz de diferente coloración y precocidad, de acuerdo a la humedad residual existente en el suelo o al inicio de la temporada de lluvias; así maneja en forma escalonada maíces de grano blanco (siembras de humedad residual en marzo y abril), amarillos, azules y rojos (siembras de mayo a primera quincena de junio). El maíz de grano amarillo es precoz, considerando su ciclo a partir de la siembra presenta su floración masculina (espiga) aproximadamente a los 89 días, y su madurez fisiológica a los 155 días. En Serdán, casi en la totalidad de la superficie se aplicaron fertilizantes químicos (18,296 ha), mientras que, en sólo 622, se sembró con semillas mejoradas (3.4%). En este mismo municipio, se reportó que a 2,241 ha se les incorporó abonos naturales, y que se aplicaron herbicidas químicos en 4,646 ha, con sólo 107 ha con algún tipo de control orgánico de malezas.

La problemática compleja que enfrentan los Agricultores están compuesta por: Variación en el clima (**VC**), Cambio climático (**CC**), son Vulnerables a eventos extremos como: sequía, heladas, canícula, inundaciones, ondas de calor, etc. Degradación de suelos, plagas, rendimientos bajos, disminución de agua, daños irreversibles. La superficie agrícola que recibió la aplicación de algún insecticida químico fue relativamente poca, con sólo 1,727 ha. Se reporta que sólo 37 ha recibieron control orgánico de insectos (1997). Pero resalta que son **Cultivos de temporal**: maíz, cebada y frijol.

Como la **VC** y el **CC** quedan registrados simultáneamente al realizar el monitoreo diario del tiempo en las estaciones hidrometeorológicas, quedando registros de Precipitación acumulada en 24 horas, Evaporación, Temperaturas Máxima y Mínima, entre otras. Así que en esta investigación, consideraremos ambas componentes de **VC** y **CC**, además, ver si es

posible identificar ya una firma del **CC** en las **ST** de Serdán, siguiendo las recomendaciones del **IPCC**. Esto significa, por ejemplo, que de tener datos desde 1961 a la fecha 2012, pueden analizarse dos periodos, el primero de 1961 al 1990 y de estarse dando ya el **CC**, debe analizarse el segundo periodo 1991 al 2012.

MATERIALES Y MÉTODOS

La base de datos (**BD**) está conformada por series de tiempo (**ST**) diarias de las variables: precipitación (**Pcp**) y temperaturas máximas (**Tmáx**) y mínimas (**Tmím**) de la estación de Ciudad Serdán que cubre periodos variables de observación.



Foto1 Principales forzantes orográficos

Dentro de los forzantes orográficos, destacan:

- 1) **Llanura escalonada**. Constituye la parte central del Valle, de norte a sur y es la zona más productiva desde el punto de vista agrícola.
- 2) **Lomeríos**. Este tipo de orografía se localiza en las faldas de las sierras, principalmente, en la parte oriental y occidental, siendo las zonas más notables las faldas del pico de Orizaba.

Esta región, aunque también se dedica la producción de cultivos, está mayormente dedicada a la explotación forestal y al pastoreo.

3) **Sierras**. El Pico de Orizaba, 5,545 msnm, forma parte de las sierras del Valle de Serdán, y del cual se desprende la Sierra Negra, 3,000 a 4,000 msnm. Incluye la Sierra de Soltepec, 2,200 a 3,000 msnm.

De los impactos hidrometeorológicos, sobresalen:

La incidencia de **heladas** es una condición problemática para la producción agrícola en el Valle de Serdán. Éstas se presentan en las zonas de lomerío en promedio 70 veces al año, concentrándose entre los meses de octubre a marzo; sin embargo, también existe riesgo de heladas entre abril y septiembre.

Destacando: El centro de Tlachichuca, el extremo oeste de San Nicolás Buenos Aires, el norte y el extremo sur de San Salvador El Seco, el extremo sur de Soltepec y Mazapiltepec, el extremo suroeste de San Juan Atenco, **el oeste de Serdán**, el extremo suroeste de Atzitzintla, el centro y oeste de Esperanza y Cañada Morelos.

El periodo de crecimiento del cultivo resultante del periodo libre de heladas, en el Valle de Serdán va desde 0 hasta 180 o 200 días al año.

Las **granizadas** es otra condición que dificulta la óptima producción agrícola en el Valle de Serdán. Entre marzo y junio la probabilidad de que caiga una granizada es de un 40 y 80%, siendo mayor en abril y mayo.

Los **vientos**, principalmente, antes del inicio de la temporada de lluvias, contribuyen a erosionar los suelos, pues están considerados de fuertes a moderados con una rapidez promedio de 40 km/hr, con direcciones noroeste y suroeste barlovento al Pico de Orizaba.

Los productores acostumbran sembrar variedades criollas de maíz de diferente coloración y precocidad, de acuerdo a la **humedad residual** existente en el suelo o al **inicio de la temporada de lluvias**; así maneja en forma escalonada maíces de grano blanco (siembras de humedad residual en marzo y abril), amarillos, azules y rojos (siembras de mayo a primera quincena de junio). El maíz de grano amarillo es precoz, considerando su ciclo a partir de la siembra presenta su floración masculina (espiga) aproximadamente a los 89 días,

y su madurez fisiológica a los 155 días. Cabe destacar que la agricultura predominante en el Valle de Serdán es la de temporal aplicando su conocimiento empírico.

Se aplicó primero un control de calidad de datos, con el Software **RClimDex** de acuerdo a las recomendaciones del **IPCC**. La **MD** es un proceso que invierte la dinámica del método científico en el siguiente sentido: En éste, primero formulamos la hipótesis, luego diseñamos el experimento para generar los datos que confirmen o rechacen la hipótesis. Si esto se hace con la formalidad adecuada, obtenemos un nuevo conocimiento. En la **MD**, se coleccionan los datos y esperamos extraer hipótesis de ellos. Queremos que los datos nos describan o indiquen por qué son como son. Validar esa hipótesis inspirada por los datos en los datos mismos, será numéricamente significativa, pero experimentalmente inválida. De ahí que la minería de datos debe presentar un enfoque exploratorio, y no confirmador.

Métodos de Minería de Datos

La **MD** cuenta con tres grandes componentes: Clustering o Clasificación, Reglas de Asociación y Análisis de Secuencias. Sin embargo, no depende la **MD** de un método en particular. En esta investigación aplicamos conceptos de asociados a los índices que calcula **RClimDex**.

A continuación presentamos resultados de los índices climáticos calculados buscando una posible firma de la Variabilidad natural del clima y del cambio climático.

RESULTADOS

Los experimentos numéricos con el software **RCLimDex**, para las **ST** de **Pcp**, **Tmáx** y **Tmím**, se presentan en ese orden.

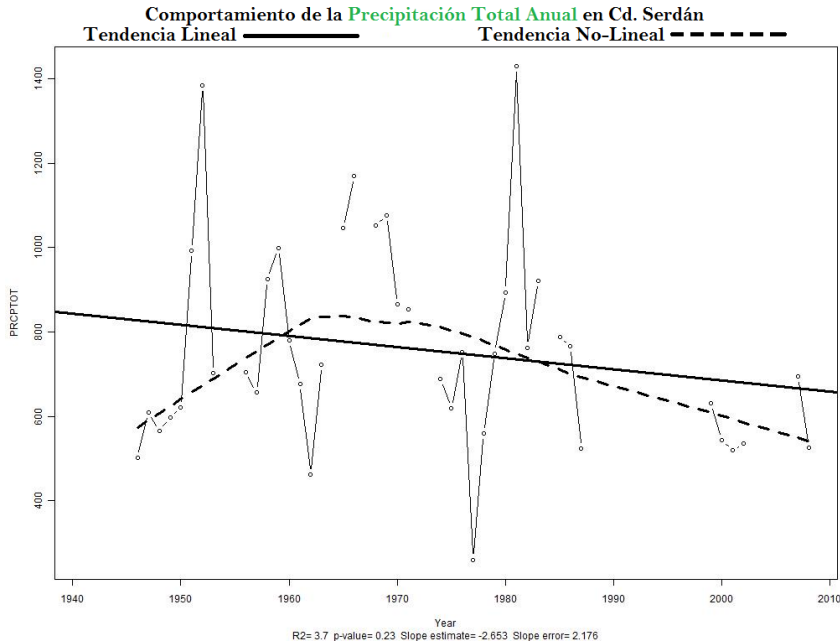


Fig. 1 Marcha anual de la precipitación

En la figura 1, destaca la alta variabilidad natural del total anual de Pcp, para el periodo 1945 a casi 1990. También resalta una variabilidad, de casi cuarenta años, con precipitación acumulada anual mayor con respecto a todo el periodo: entre 1950 y 1990. Ambas líneas de tendencia: lineal y no-lineal, muestran un descenso de la **Pcp** total anual, a pesar de que hay relativamente pocos datos, en los recientes veinte años.

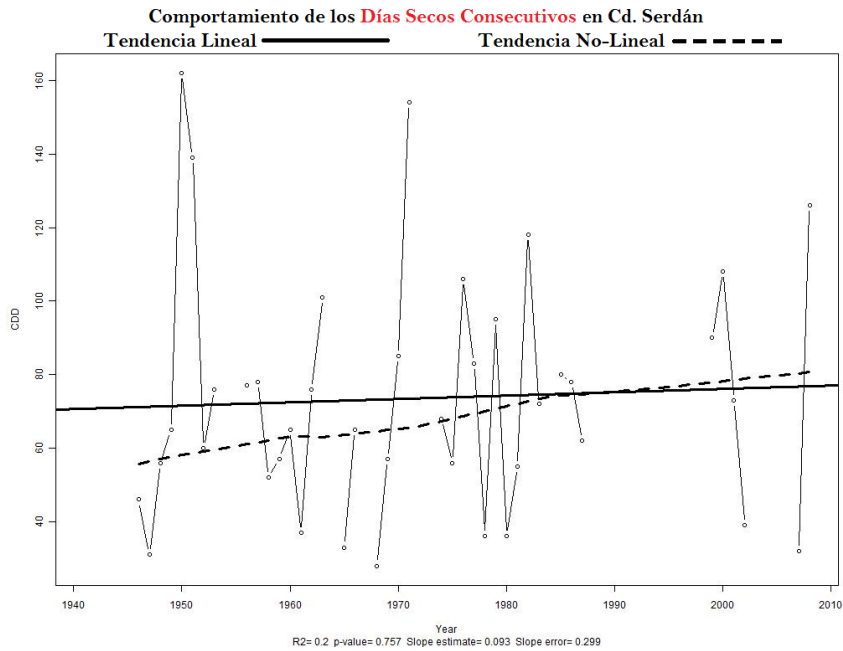


Fig. 2 Marcha anual de los días secos consecutivos

Complementariamente al comportamiento de la Pcp acumulada anual, hay una ligera tendencia positiva y con una alta variabilidad en el número de días secos consecutivos.

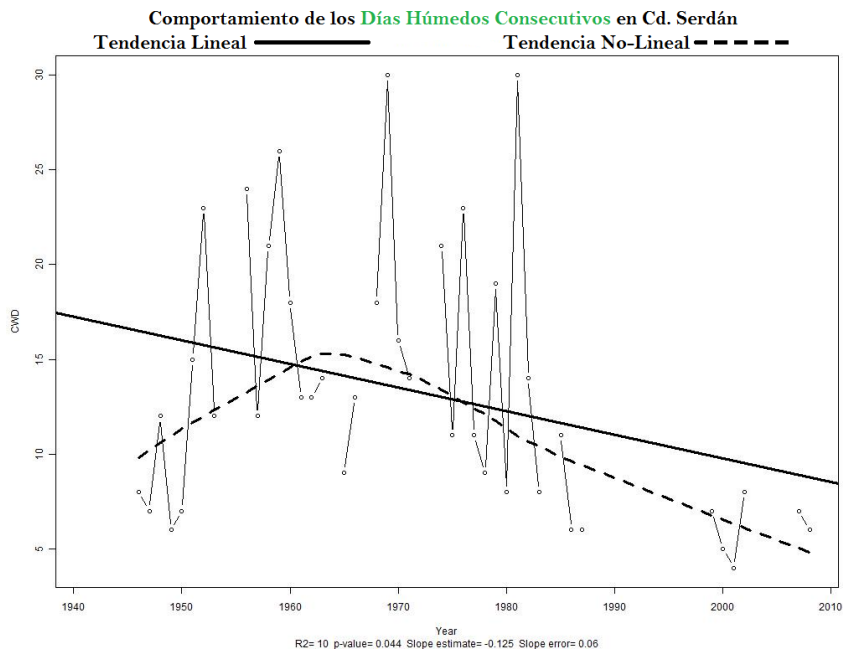


Fig. 3 Marcha anual de los días húmedos consecutivos

Así mismo el comportamiento de los días húmedos consecutivos, presentan una tendencia a la baja, inclusive con respecto a todo el periodo. En este índice se identifica mejor aún la variabilidad de cuarenta años, entre 1950 y 1990. De igual forma induce a una firma del **CC**.

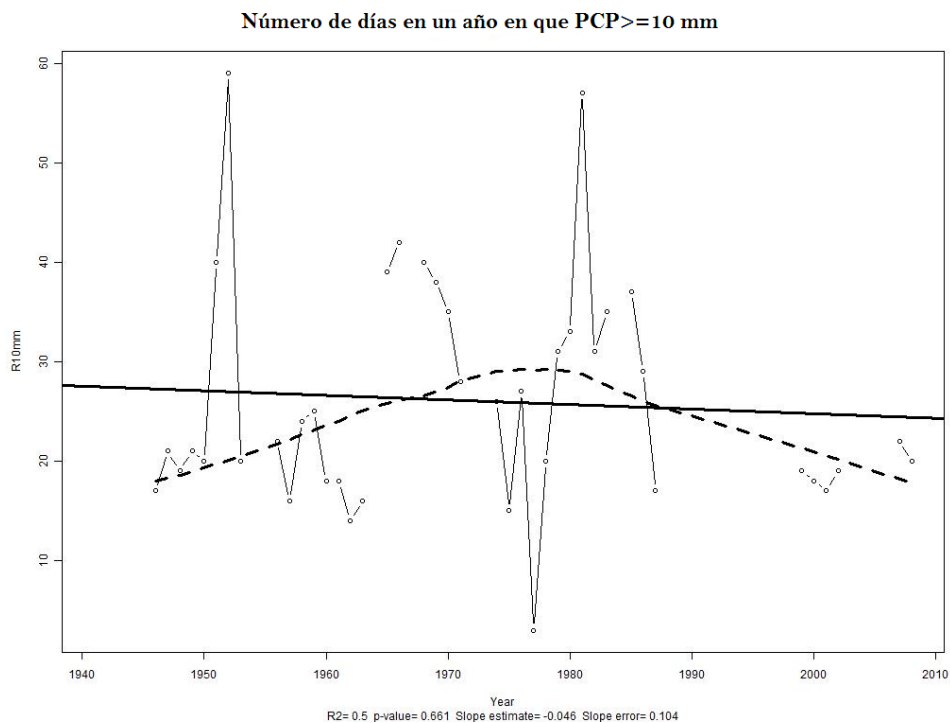


Fig. 4 Marcha anual del número de días con precipitación \geq 10mm

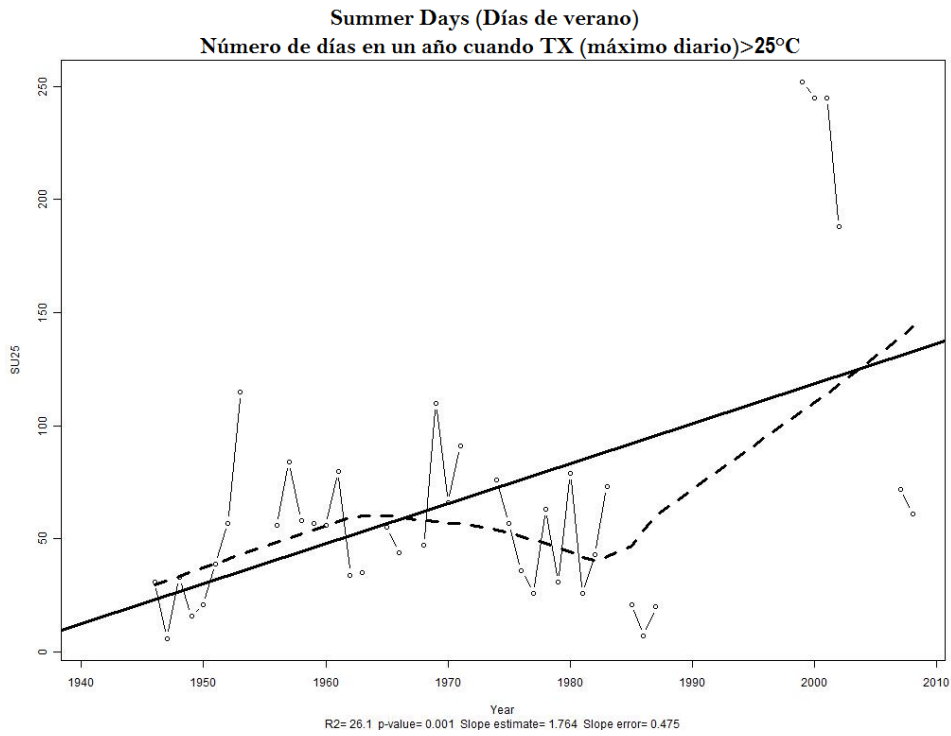


Fig. 5 Marcha anual del número de días de verano con $T_{\text{máx}} > 25^{\circ}\text{C}$

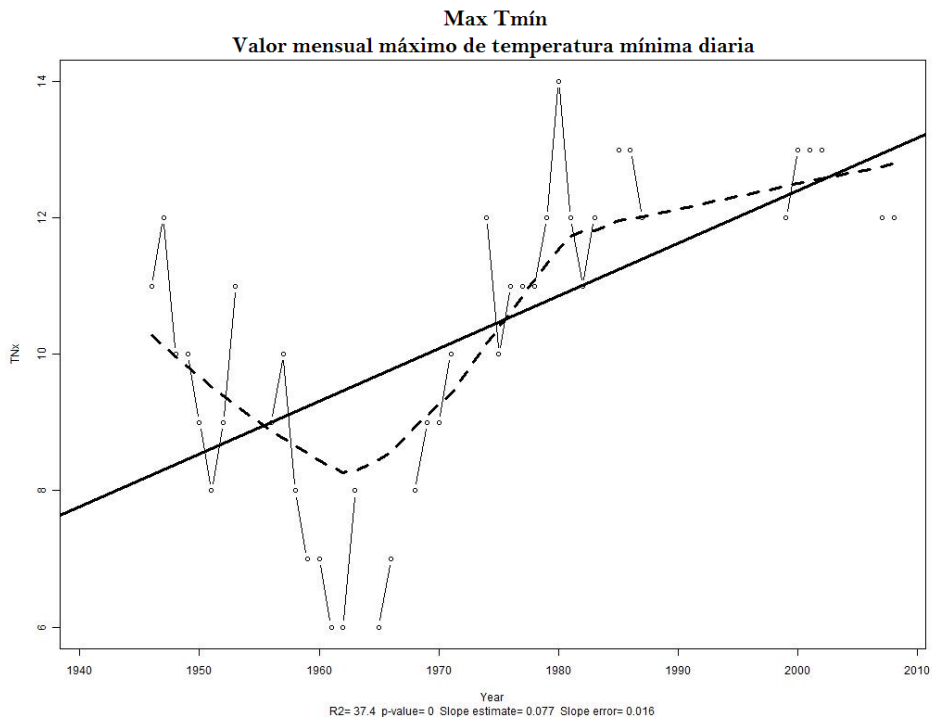


Fig. 6 Marcha del valor mensual máximo de temperatura mínima diaria $T_{\text{mín}}$

Del análisis de la Fig. 4, el índice de valores de Pcp rebasando el umbral no indica gran cosa, Pero el índice donde la $T_{máx}$ fue mayor que un valor umbral, 25°C , o días de verano Fig.5, así como en máximo de $T_{mín}$ muestran tanto la señal de cuarenta años como la firma del **CC**.

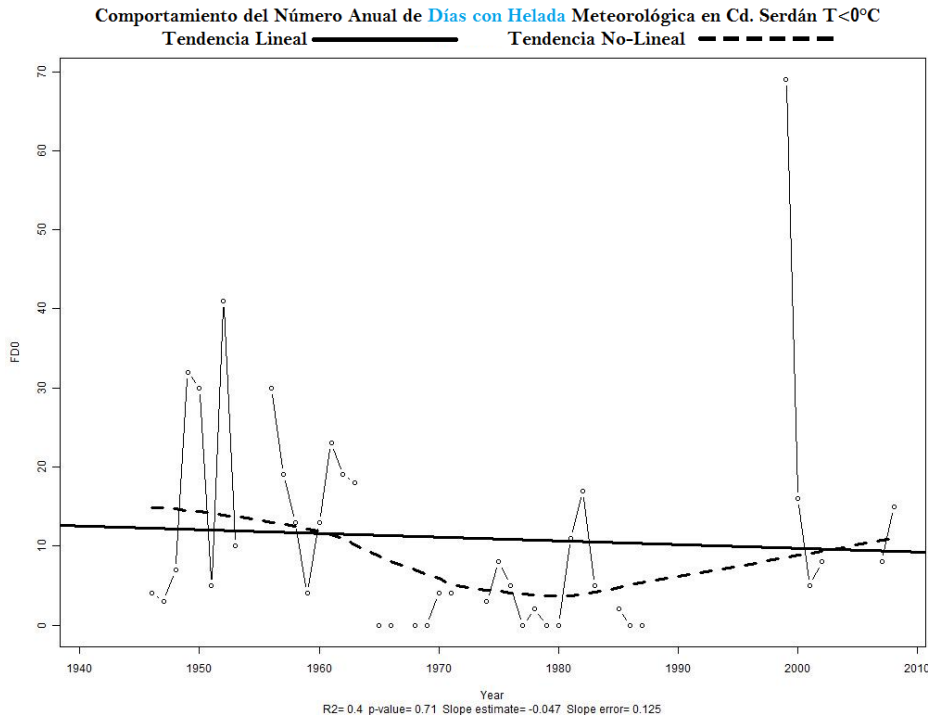


Fig. 7 Marcha anual del número de días con helada, $T_{mín} < 0^{\circ}\text{C}$

La información en la Fig. 7, sólo destacan la variación de cuarenta años y la mayor amplitud de la variabilidad en los recientes veinte años.

DISCUSIÓN

Se identificó una oscilación, en la variabilidad del clima, del orden de cuarenta años, durante el periodo 1950 al 1990. Hubo índices que sugieren ya una señal de **CC**, al menos en los últimos veinte años. Destacando el índice dado por comportamiento del valor mensual máximo de temperatura mínima diaria $T_{mín}$. Logramos mostrar las fases involucradas en la aplicación del enfoque de minería de datos: Datos-Información-Conocimiento-Toma de Decisiones al analizar series de tiempo de datos diarios de precipitación, temperatura máxima y temperatura mínima de una sola estación de Serdán, presentando gran complejidad al registrar simultáneamente variabilidad natural del clima y el cambio

climático, además con alta falta de continuidad en los registros al menos en los últimos veinte años.

Conclusión

Por lo tanto, se deberá aplicar la misma metodología a otras series de tiempo de las mismas variables, Pcp, T_{máx}, T_{mím}, para obtener resultados más robustos, debido a que en las series de datos analizadas en este estudio, presentan del 1990 al 2012 una gran ausencia de datos, inclusive buscando con un enfoque regional. También, aplicar un análisis con Wavelets para identificar las frecuencias activadas que más contribuyen a la varianza e indique cuando estuvieron activadas, ya que es un método que trabaja dos dominios, tiempo-frecuencia y permitirá identificar la oscilación de casi treinta años.

AGRADECIMIENTOS

Los autores agradecen al Ing. Rogelio Bernal Morales, Académico de la Facultad de Agrobiología, Universidad Autónoma de Tlaxcala, Campus Tlaxco, el apoyo con los experimentos numéricos con el Software **RCLimDex**.

Bibliografía

Equipo técnico del Plan Llanos de Serdán y del Cader 03, Cd. Serdán. **1987**. Zonificación Regional del Centro de Apoyo al Desarrollo Rural 03, Cd. Serdán, del Distrito de Desarrollo Rural IV de Libres, Puebla. Secretaría de Agricultura y Recursos Hidráulicos. Colegio de Postgraduados. Cd. Serdán, Pue. 455 pp.

Broomhead D.S. and Gregory P. King. 1986 Extracting Qualitative Dynamics from Experimental Data, *Physica D*,20,217-236,.

Elsner J.B. and A.A. Tsonis. 1992, Nonlinear Prediction, Chaos, and Noise. *Bulletin American Meteorological Society*, Vol. 73, No. 1, January.

Hargrove, W. W., F. M. Hoffman, and T. Sterling. 2001, The Do-It-Yourself Supercomputer. *Scientific American*, 265(2):72–79.

Hoffman Forrest M., William W. Hargrove, Richard T. Millsa, Salil Mahajanc, David J. Ericksona, and Robert J. Oglesbyd. 2008. Multivariate Spatio-Temporal Clustering (MSTC) as a Data Mining Tool for Environmental Applications. *iEMSs*,. International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making. 4th Biennial Meeting of *iEMSs*, <http://www.iemss.org/iemss2008/index.php?n=Main>. Proceedings M. Sánchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.) International Environmental Modelling and Software Society (*iEMSs*), 1774-1781.

Kadioglu M. and Sen Z. 1998, Power-law Relationship in Describing Temporal and Spatial Precipitation Pattern in Turkey *Theor. Appl. Climatol.* 59, 93-106

Klonowski W. 2002, Chaotic dynamics applied to signal complexity in phase space and in time domain. *Chaos, solitons and fractals*, 14, 1379-1387.

Kalkstein, L.S., G. Tan, and J.A. Skindlov, 1987, A comparison of three clustering procedures for use in synoptic climatological classification. *Journal of Climatology & Applied Meteorology*, 19, 717-730.

Hegger Rainer, Holger Kantz, Thomas Schreiber. 1999, Practical implementation of nonlinear time series methods: The **TISEAN** package. *Chaos* Vol. 9, No. 2, 413-435.

Morales-Acoltzi T, Medel-Rojas A, Morales-Aceves F. J, Cortés-Pérez E, Morales-Calva X. 2008, Data Mining vs. Mathematical Modelling: Nonlinear Dynamics and Chaos Theory. *iEMSs* . International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making. 4th Biennial Meeting of *iEMSs*, <http://www.iemss.org/iemss2008/index.php?n=Main>. Proceedings M. Sánchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.) International Environmental Modelling and Software Society (*iEMSs*). 1963-1964.

Selvam, A.M., and Fadnavis, S., Signatures of a universal spectrum for atmospheric interannual variability in some disparate climatic regimes, *Meteor. Atmos. Phys.*, 66(1- 2), 87-112, 1998.

- Sivertsen**, T.H. 2005. Discussing the scientific method and a documentation system of meteorological and biological parameters. *Phys. Chem. Earth*. 30:35-43,
- Sfetsos** A., Vlachogiannis D. A data mining approach to discover weather patterns contributing to **PM10** exceedances. **iEMSs** 2008, 1727-1733. International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making. 4th Biennial Meeting of **iEMSs**, <http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings> M. Sánchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.) International Environmental Modelling and Software Society (**iEMSs**).
- Turcotte**, D.L. 2003 “The Relationship of fractals in Geophysics to “New Science”. *Chaos, Solutions & Fractals*, Vol. 27, Issue 1, 210-217.
- van der Vaart**, H.R., A comparative investigation of certain difference equations and related differential equations: Implications for model building. *Bull. of Math. Biol.*, 35, 195-211, 1973.
- Zupanski**, M., and Navon, I.M. Predictability, observations, and uncertainties in the geosciences. *Bull. Amer. Meteor. Soc.*, 88, 1431–1433, 2007.