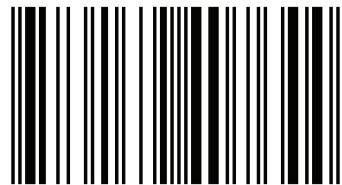


## Algoritmos, estrategias, modelos y sus aplicaciones

La serie Métodos cuantitativos aplicados, en su tomo 2, bajo el título Algoritmos, estrategias, modelos y sus aplicaciones, consta de 21 trabajos de investigación elaborados por investigadores de 26 universidades y centros de investigación de Cuba, México, España, Brasil, Colombia y Ecuador, agrupadas en RIDECA. El objetivo consiste en propiciar el desarrollo y aplicación de nuevos métodos cuantitativos para conocer las realidades locales, nacionales e internacionales con las que convivimos; por ello, se abordan temas tales como diseños secuenciales para ensayos clínicos, sistema informático para el análisis de los peligros naturales, análisis estadístico de datos turísticos y certificación de playas, modelo de regresión logística en rehabilitación, condicionantes de la alfabetización científica, relación entre variables, innovación y rendimiento en el sector comercial colombiano, diseño de criptosistemas, minería de textos, estadísticos de orden, simulación de dispersión de enfermedades, la motivación y el rendimiento escolar, evaluación diagnóstica de la gestión y, validez y confiabilidad en la calidad de cuestionarios, entre otros.

Carlos N. Bouza, Doctor en Ciencias, Universidad de La Habana, Cuba y Prof. de la Facultad de Matemática y Computación. Agustín Santiago, Doctor en Estadística e IO por la Ugr y Prof. en la Facultad de Matemáticas de la UAGro. México. José Félix García R., Doctor en Finanzas Públicas por la UV, México y Profesor de la Universidad Juárez de Tabasco.



978-613-9-40764-4

editorial académica española

ead  
editorial académica española



Carlos N. Bouza Herrera · Agustín Santiago M. · José Félix García R.

## Algoritmos, estrategias, modelos y sus aplicaciones

Serie Métodos Cuantitativos aplicados  
Tomo 2

Métodos Cuantitativos Aplicados Tomo 2

Bouza Herrera, Santiago M., García R.

**Carlos N. Bouza Herrera**  
**Agustín Santiago M.**  
**José Félix García R.**

**Algoritmos, estrategias, modelos y sus aplicaciones**

**Carlos N. Bouza Herrera  
Agustín Santiago M.  
José Félix García R.**

# **Algoritmos, estrategias, modelos y sus aplicaciones**

**Serie Métodos Cuantitativos aplicados Tomo 2**

**Editorial Académica Española**

**Imprint**

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: [www.ingimage.com](http://www.ingimage.com)

Publisher:

Editorial Académica Española

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

**ISBN: 978-613-9-40764-4**

Copyright © Carlos N. Bouza Herrera, Agustín Santiago M., José Félix García R.

Copyright © 2019 International Book Market Service Ltd., member of OmniScriptum Publishing Group

Algoritmos, estrategias y  
modelos en las aplicaciones de  
métodos cuantitativos

Tomo 2

**Serie Métodos cuantitativos aplicados**

## Índice de capítulos

CAPITULO	TÍTULO	AUTORES	PÁGINAS
1	REVISIÓN DE DISEÑOS SECUENCIALES PARA ENSAYO CLÍNICOS FASE II	Carmen Viada González, Martha Fors López, Maytec Robaina García, Mabel Álvarez Cardona, Aliuska Frías Blanco, Lázara García Fernández, Bárbara Wilkinson Brito, Liana Martínez Pérez, Mayelin Troche de la Concepción	1-16
2	METODOLOGÍA Y SISTEMAS PARA META-ANÁLISIS DE ENSAYOS CLÍNICOS	Carmen Viada Gonzalez, Martha Fors López, Javier Ballesteros, Maytec Robaina García, Mabel Álvarez Cardona, Aliuska Frías Blanco, Lázara García Fernández, Bárbara Wilkinson Brito, Liana Martínez Pérez, Mayelin Troche de la Concepción.	17-35
3	SISTEMA INFORMÁTICO PARA EL ANÁLISIS DE LOS PELIGROS NATURALES	Amaury Machín Armenteros, Yasser Vázquez Alfonso, Elaine Pérez Monteagudo, Efraín Velastegui López	37-48
4	ANÁLISIS ESTADÍSTICO DE LA OCUPACIÓN HOTELERA EN ACAPULCO Y LA CERTIFICACIÓN DE PLAYAS PARA LA COMPETITIVIDAD TURÍSTICA	Miguel Ángel Cruz Vicente, Darbelio Agatón Lorenzo, Arturo de León Chapa	49-69
5	MODELO DE REGRESIÓN LOGÍSTICA EN PACIENTES CON GONARTROSIS QUE REALIZAN TRATAMIENTO REHABILITADOR	G. Bayolo Soler, V. Sistachs Vega, Y. Coronados Valladares	71-88
6	GÉNERO Y ENTORNO DE NACIMIENTO EN EL NIVEL DE ALFABETIZACIÓN CIENTÍFICA DE ESTUDIANTES DE UNA UNIVERSIDAD PÚBLICA MEXICANA	Carlos David Zetina Pérez, Deneb Elí Magaña Medina, Pedro Antonio Sánchez Escobedo, José Félix García Rodríguez	89-102
7	RELACIÓN ENTRE EL RECONOCIMIENTO DE MARCA Y EL MODELO DE AJUSTE A LA TAREA TECNOLÓGICA (TTF)	Andrés Machado Daza	103-117
8	MIPYMES, INNOVACIÓN Y RENDIMIENTO EN EL SECTOR INDUSTRIAL COLOMBIANO	Paula Viviana Robayo-Acuña, Henry Laverde Rojas	119-136

<b>9</b>	CRIPOTOSISTEMA BASADO EN AUTÓMATAS CELULARES Y FUNCIONES CAÓTICAS	Luis René Marcial Castillo, Marcela Rivera Martínez, Lourdes Sandoval Solís, Juan Jesús Luna Padilla	137-149
<b>10</b>	MINERÍA DE TEXTO APLICADA AL MARKETING VERDE (1985-2018): UN ANÁLISIS BASADO EN SCOPUS	Jenny-Paola Lis-Gutiérrez, Mercedes Gaitán-Angulo, Cristian Beltrán-Oicatal, Melissa Lis-Gutiérrez, Andreina Moros, Carolina Henao	151-178
<b>11</b>	GENERALIDADES SOBRE LOS ESTADÍSTICOS DE ORDEN	Carlos N. Bouza-Herrera	179-200
<b>12</b>	LA TEORÍA DE EPIDEMIAS Y LA SIMULACIÓN DE LA DISPERSIÓN DE ENFERMEDADES RESPIRATORIAS	José Betancourt Bethencourt, Juan Jesús Llambias Peláez, Luis Acao Francoise.	201-209
<b>13</b>	INFLUENCIA DE LA MOTIVACIÓN EN EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIDAD ACADÉMICA DE ECONOMÍA DE LA UAGRO, PERIODO FEBRERO-JULIO 2013	Beatriz Sánchez-Estrada, Agustín Santiago-Moreno, Alfredo Pineda-Gómez, Susana Lasso-Abarca	211-236
<b>14</b>	VALIDEZ Y CONFIABILIDAD DE UN CUESTIONARIO SOBRE LA CALIDAD DE VIDA EN MUJERES CON LUPUS	Gladys Linares Fleites, María de Lourdes Sandoval Solís, María Lilia Cedillo Ramírez y Socorro Méndez Martínez	237-252
<b>15</b>	EVALUACIÓN DIAGNÓSTICA DE LA GESTIÓN DE LOS CONSEJOS TÉCNICOS ESCOLARES DEL NIVEL DE EDUCACIÓN BÁSICA EN ACAPULCO	Cinthia Corey Bustamante Aguirre, Tonantzin Martínez Bello, Elizabeth Ávila Cortes, Agustín Santiago Moreno.	253-263
<b>16</b>	CRITERIOS DE CALIDAD DEL CUESTIONARIO DE EVALUACIÓN DE LOS CTE DEL NIVEL BÁSICO DE EDUCACIÓN EN ACAPULCO	Elizabeth Ávila Cortes, Tonantzin Martínez Bello, Cinthia Corey Bustamante Aguirre y José Maclovio Sautto-Vallejo	265-274
<b>17</b>	PROPUESTA DE ANÁLISIS Y SELECCIÓN DE INICIATIVAS DE INNOVACIÓN SOCIAL	Jenny-Paola Lis-Gutiérrez, Jenny Cubillos-Díaz, Mercedes Gaitán-Angulo, Manuel Ignacio Balaguera, Carlos Bouza	275-291
<b>18</b>	TASAS DE CRECIMIENTO E ÍNDICES DE COMPETITIVIDAD DEL COMERCIO EXTERIOR DEL MANGO MEXICANO	Ignacio Caamal Cauch, Verna Gricel Pat Fernández, José Félix García Rodríguez, Zulia Helena Caamal Pat	293-310

## Afiliación

<b>Nombre</b>	<b>Afiliación</b>
Agustín Santiago-Moreno	Universidad Autónoma de Guerrero, Facultad de Matemáticas-Acapulco.
Aida Armenta Ramírez	Universidad Juárez Autónoma de Tabasco, México
Alfredo Pineda-Gómez	Universidad Autónoma de Guerrero, Facultad de Ciencias de la Educación-Acapulco.
Aliuska Frías Blanco	Centro de Inmunología Molecular, La Habana, Cuba
Amaury Machín Armenteros	Facultad de Ciencias Técnicas, Universidad Agraria de la Habana, Cuba
Andreina Moros	Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
Andrés Machado Daza	Fundación Universitaria Konrad Lorenz
Arturo de León Chapa	Universidad Autónoma de Guerrero
Bárbara Wilkinson Brito	Centro de Inmunología Molecular, La Habana, Cuba
Beatriz Sánchez-Estrada	Universidad Autónoma de Guerrero, Facultad de Ciencias de la Educación-Acapulco.
Boris Espinoza	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)
Carlos Bouza Herrera	Facultad de Matemática y Computación, Universidad de la Habana
Carlos David Zetina Pérez	Universidad Juárez Autónoma de Tabasco, México
Carmen Viada González	Centro de Inmunología Molecular, La Habana, Cuba
Carolina Henao	Universidad de Ciencias Aplicadas, Bogotá, Colombia
Cynthia Corey Bustamante Aguirre	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
Cristian Beltrán-Oicata	Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
Darbelio Agatón Lorenzo	Universidad Autónoma de Guerrero
Deneb Elí Magaña Medina	Universidad Juárez Autónoma de Tabasco, México
Diego Andrés Guevara-Fletcher	Facultad de Posgrados ESAP, Colombia.
Diego Ochoa-Jiménez	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)
Efraín Velastegui López	Universidad de Cotopaxi, Ecuador.
Elaine Pérez Monteagudo	Dirección Provincial de Planificación Física, Mayabeque, Cuba.
Elizabeth Ávila Cortes	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
G. Bayolo Soler	Facultad de Matemática y Computación, Universidad de La Habana, Cuba
Gladys Linares Fleites	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla



Henry Laverde Rojas	CIEN, Escuela de Negocios, Fundación Universitaria Konrad Lorenz Bogotá D.C., Colombia
Ignacio Caamal Cauich	Universidad Autónoma Chapingo, México.
Javier Ballesteros	Universidad del País Vasco, Bilbao, España
Jenny Cubillos-Díaz	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
Jenny-Paola Lis-Gutiérrez	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
José Betancourt Bethencourt	Universidad de Ciencias Médicas Carlos J. Finlay/Centro de Inmunología y Productos Biológicos/Camagüey, Cuba.
José Félix García Rodríguez	Universidad Juárez Autónoma de Tabasco, México
José Maclovio Sautto-Vallejo	Universidad Autónoma de Guerrero, Campus Costa Chica.
Juan Jesús Llambias Peláez	Dirección Provincial de Salud. Departamento de Vigilancia en Salud /Camagüey, Cuba
Juan Jesús Luna Padilla	Benemérita Universidad Autónoma de Puebla, México
Lázara García Fernández	Centro de Inmunología Molecular, La Habana, Cuba
Liana Martínez Pérez	Centro de Inmunología Molecular, La Habana, Cuba
Luis Acao Francoise	Universidad de Ciencias Médicas Carlos J. Finlay/Departamento de Salud, Cuba.
Luis René Marcial Castillo,	Benemérita Universidad Autónoma de Puebla, México
Mabel Álvarez Cardona	Centro de Inmunología Molecular, La Habana, Cuba
Manuel Ignacio Balaguera	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Facultad de Matemáticas e Ingeniería (CIMI)
Marcela Rivera Martínez	Benemérita Universidad Autónoma de Puebla, México
María de Lourdes Sandoval Solís	Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla
María Lilia Cedillo Ramírez	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla
Marisol de la Cruz León	Universidad Juárez Autónoma de Tabasco, México
Martha Fors López	Universidad de Las Américas, Quito, Ecuador.
Mayelin Troche de la Concepción	Centro de Inmunología Molecular, La Habana, Cuba
Maytee Robaina García	Universidad de Las Américas, Quito, Ecuador.
Maytee Robaina García	Centro Nacional Coordinador de Ensayos Clínicos, La Habana, Cuba
Melissa Lis-Gutiérrez	Corporación Universitaria Minuto de Dios
Mercedes Gaitán-Angulo	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
Miguel Ángel Cruz Vicente	Universidad Autónoma de Guerrero
Pascual García Macías	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)

Paula Viviana Robayo-Acuña	CIEN, Escuela de Negocios, Fundación Universitaria Konrad Lorenz Bogotá D.C., Colombia
Pedro Antonio Sánchez Escobedo	Universidad Autónoma de Yucatán, México
Rilton G. Goncalvez Primo	Universidad Federal de Bahía, Brasil
Socorro Méndez Martínez	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla
Susana Lasso-Abarca	Universidad Autónoma de Guerrero, Coordinación zona sur-Acapulco.
Tonantzin Martínez Bello	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
Verna Grisel Pat Fernández	Universidad Autónoma Chapingo, México.
Vivian Sistachs Vega	Facultad de Matemática y Computación, Universidad de La Habana, Cuba
Y. Coronados Valladares	Hospital Julio Díaz, Cuba.
Yasser Vázquez Alfonso	Facultad de Turismo, Universidad de la Habana, Cuba
Zulia Helena Caamal Pat	Colegio de Posgraduados, Texcoco, México.

# Capítulo 9

## CRITOSISTEMA BASADO EN AUTÓMATAS CELULARES Y FUNCIONES CAÓTICAS

Luis René Marcial Castillo, Marcela Rivera Martínez, Lourdes Sandoval Solís, Juan Jesús Luna Padilla

Benemérita Universidad Autónoma de Puebla, Puebla, Puebla, México

### RESUMEN

En este trabajo, se diseñó e implementó computacionalmente un algoritmo para criptografía simétrica, el cual se basa en el uso de autómatas celulares, funciones caóticas y la operación lógica xor. El sistema criptográfico implementado en el lenguaje de programación OCTAVE permite cifrar y descifrar en pocos minutos imágenes digitales a color de dimensiones menores o iguales a 800 x 800. Las pruebas realizadas muestran que la implementación computacional cifra y descifra de forma correcta y son una alternativa 6 veces más rápida al estándar AES de la criptografía simétrica. Se demuestra que el cripto sistema presentado es seguro ante los ataques de fuerza bruta y frecuencia de los datos.

**PALABRAS CLAVES:** Criptografía simétrica, AES, autómata celular, función caótica, operación xor.

### ABSTRACT

In this work, an algorithm for symmetric cryptography was designed and implemented computationally, which is based on the use of cellular automata, chaotic functions and the logical operation xor. The cryptographic system implemented in the programming language OCTAVE allows to encrypt and decipher in a few minutes' digital color images of dimensions less than or equal to 800 x 800. The tests carried out show that the computational implementation correctly encrypts and deciphers and is a 6 times faster alternative to the AES standard of symmetric cryptography. It is shown that the crypto system presented is safe against brute force attacks and frequency of data.

**KEYWORDS:** Symmetric cryptography, AES, cellular automata, chaotic function and xor operation.

## 1. INTRODUCCIÓN

El estándar de la criptografía simétrica AES (Advanced Encryption Standard), es un cripto sistema muy seguro, la desventaja de este método es su lentitud en el proceso de cifrado y descifrado, siendo esto un problema cuando se tienen grandes volúmenes de datos como sucede en imágenes digitales. Por ejemplo, los servicios de telemedicina requieren cifrar y descifrar de forma rápida imágenes digitales, con el objetivo de proporcionar confidencialidad sobre alguna enfermedad en un paciente; también se podría aplicar a imágenes del campo de la biología en donde se utilizaría para ocultar ciertas plagas de modo que solo los investigadores conozcan el contenido.

La propuesta de este trabajo, es desarrollar una implementación computacional bajo el ambiente de programación OCTAVE de un cripto sistema que se base en el uso de autómatas celulares, funciones caóticas y la función lógica xor. La finalidad de usar este esquema es cifrar y descifrar imágenes digitales mucho más rápido que el estándar AES de la criptografía simétrica y que proporcione suficiente seguridad ante ataques a fuerza bruta.

El trabajo se organiza de la manera siguiente: la sección 2, describe conceptos de criptografía, autómatas celulares, la operación lógica xor y métodos caóticos, la sección 3, presenta los antecedentes del proyecto en lo que se refiere al uso de autómatas celulares y funciones caóticas en el campo de la criptografía simétrica, en la sección 4, se muestran los pasos de los algoritmos que realizan el cifrado y descifrado de imágenes digitales, la sección 5, proporciona las pruebas del cripto sistema usando imágenes a color. En la sección 6, se presenta el análisis del cripto sistema ante un ataque a fuerza bruta, la sección 7, presenta las conclusiones del trabajo y finalmente se listan las referencias.

## 2. CONCEPTOS

La confidencialidad es de gran importancia en la sociedad moderna. La industria, gobierno y particulares confían en que la tecnología les garantice que el intercambio de datos sea seguro de modo que no se permita a terceros acceder al contenido de tal comunicación, la cuestión de la confidencialidad ha sido dejada al campo de la criptografía [2]. Goldreich transcribe el planteamiento de la criptografía como "*el problema de proveer comunicación secreta sobre medios inseguros*" [6]. La criptografía es una herramienta muy útil cuando se desea seguridad informática, es decir, cuando se cuenta con un medio para garantizar las propiedades de confidencialidad, para lograrlo, se crean mecanismos que garanticen en cierta medida las propiedades de disponibilidad, integridad y confidencialidad. La disponibilidad se refiere a que la información siempre este presente, la integridad es que no se pierda información y la confidencialidad es que, aunque sea robada la información, no se pueda acceder al contenido o se garantice de alguna forma que no se pueda llegar a ella, hasta que pierda su valor. El criptoanálisis, es la disciplina contraria a la criptografía, se encarga de analizar la información cifrada para

revelar la información original, el criptoanálisis lo usan los investigadores como una forma de probar las fortalezas o debilidades de los cripto sistemas [4].

La criptografía se divide en dos grandes ramas: la criptografía de clave privada o simétrica y la criptografía de clave pública o asimétrica. La primera, se refiere al conjunto de métodos que permiten una comunicación segura entre las partes siempre que con anterioridad se intercambie la clave correspondiente que se denomina clave o llave simétrica. La criptografía simétrica utiliza algún método matemático para cifrar o descifrar un mensaje utilizando únicamente la llave secreta, esta llave se debe compartir con las personas que se desea vean los mensajes. El problema de la criptografía simétrica es compartir la clave de forma confidencial, pero esto se resuelve con el uso de la criptografía asimétrica. Uno de los sistemas criptográficos simétricos más conocidos es el algoritmo DES (Data Encryption Standard), la U.S. National Institute of Standards and Technology (NIST) lo estandarizó en los años 70, el mejor ataque conocido al algoritmo DES es la búsqueda exhaustiva de la llave, los especialistas en cómputo diseñaron una supercomputadora usando las PCs conectadas a la red y usando el poder de 100000 PCs fueron capaces de encontrar la llave usada por el algoritmo DES después de 22 horas y 15 minutos, después de este resultado, la opción que se tomó fue sustituir el algoritmo DES por el cifrado múltiple, que consiste en aplicar varias veces el mismo algoritmo con el fin de fortalecer la longitud de la clave, esto ha tomado forma como un nuevo sistema para el cifrado, conocido actualmente como el algoritmo TDES, aunque actualmente el estándar de la criptografía simétrica es el algoritmo AES, el cual fue adoptado desde 1997 por la U.S. NIST y el algoritmo usado es el propuesto por Rijndael en el año 2002 [2]. La criptografía de clave pública o asimétrica, es aquella que utiliza dos llaves o claves diferentes para cada usuario: una llave se usa para cifrar y la otra se usa para descifrar (conocida como clave o llave privada). El estándar de la criptografía asimétrica es el algoritmo RSA (en honor a sus inventores Rivest, Shamir y Adleman) publicado en 1978 [2].

## 2.1 AUTÓMATAS CELULARES

Los Automatas Celulares (AC) son un modelo de cálculo que puede usarse para la búsqueda de inteligencia y desarrollo de vida artificial, John Von Neumann los presentó en su libro "Theory of Self-reproducing Automaton", su principal intención era desarrollar una máquina que fuera capaz de construir a otras máquinas. Wolfram de los años 1992 a 2002 hace un intenso estudio de los AC sobre su aplicación a la vida real, donde fundamenta la posibilidad de modelar y comprender cualquier fenómeno natural, Wolfram también menciona que los AC pueden aplicarse en diversos campos de la ciencia tales como el de la Criptografía. Los AC se basan en un sistema de vecindades. En una dimensión tomando únicamente al vecino izquierdo y derecho (una vecindad de  $1 \times 3$ ) se tiene un total de  $2^3 = 256$  reglas. Para una vecindad bidimensional de  $3 \times 3$  se tendría un

total  $2^{2^9} = 2512$  reglas. Por ejemplo, en la tabla 1, se presenta la regla 30, el valor de 1 significa que el vecino está presente y 0 significa que el vecino está ausente.

Tabla 1: Regla 30 del autómata celular de una dimensión.

Vecino izquierdo	Centro	Vecino derecho	Salida
1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	1
0	0	1	1
0	0	0	0

La figura 1, muestra 32 iteraciones (una en cada renglón de la imagen) de la ejecución del AC cuando se usa la regla 30, cada fila tiene una dimensión de 32, se inicia la fila 1 con un valor diferente de 0 (pixel negro) en su centro y 0 en todas las demás posiciones.



Figura 1:32 iteraciones de la regla 30 del AC unidimensional.

## 2.2 OPERACIÓN LÓGICA XOR

La operación booleana en las dos variables  $x, y$  se muestra en la tabla 2.

Tabla 2: Operación lógica xor.

$x$	$y$	xor
0	0	0
0	1	1
1	0	1
1	1	0

La operación booleana es muy útil en el campo de la criptografía simétrica ya que constituye la base de los estándares como lo fue el algoritmo DES y ahora el algoritmo AES. Por ejemplo, si se desea cifrar el dato 173 (cuya representación binaria es: 10101101), con la llave 211 (cuya representación binaria es: 11010011), al usar la

operación binaria xor se obtiene el cifrado **c** igual a **01111110**, que corresponde al dato 126 en decimal. La operación de cifrado se muestra en la tabla 3.

Tabla 3: Cifrado del dato 173 con la llave 211.

<b>x</b>	1	0	1	0	1	1	0	1
llave	1	1	0	1	0	0	1	1
<b>c = x xor llave</b>	0	1	1	1	1	1	1	0

Para obtener el descifrado se utiliza la misma llave, obteniendo el descifrado **d** igual a 10101101 que corresponde en decimal al dato original 173. La operación se ilustra en la tabla 4.

Tabla 4: Descifrado del dato 126 con la llave 211.

<b>c</b>	0	1	1	1	1	1	1	0
llave	1	1	0	1	0	0	1	1
<b>d = c xor llave</b>	1	0	1	0	1	1	0	1

Como lo ilustran las tablas 3 y 4, la operación xor permite cifrar y descifrar datos con la misma llave sin pérdida de información.

## 2.3 MÉTODOS CAÓTICOS

Puede decirse que la dinámica caótica inició con el trabajo del matemático Francés Henri Poincaré a finales del siglo XIX. La motivación de Poincaré fue promovida por el problema de las orbitas de tres cuerpos celestes experimentando atracción gravitacional mutua (por ejemplo, una estrella y dos planetas). Poincaré fue capaz de mostrar que orbitas muy complicadas eran posibles (ahora llamadas caóticas). No obstante, la posibilidad de caos en sistemas físicos reales no fue ampliamente apreciada sino hasta la actualidad, mucho del crédito por este cambio es atribuido a la extensa solución numérica de sistemas dinámicos en computadoras digitales [9]. En los últimos años, los hilos del caos y la dinámica no lineal se han esparcido a través de disciplinas científicas como una intrincada red araña. Caos y dinámica no lineal han provisto de nuevas herramientas teóricas y conceptuales que permiten capturar, entender y enlazar los compartimientos complejos de sistemas simples (el tipo de comportamiento llamado caos en la ciencia contemporánea) [7]. Se puede decir, que una función caótica es una función matemática que describe un sistema dinámico no lineal complejo cuya evolución en el tiempo hace imposible la predicción a largo plazo, luciendo errático y casi aleatorio. Los ejemplos más notorios de las características del caos son el llamado *efecto mariposa* y la

*impredecibilidad de órbitas pseudo-aleatorias*, generadas por ecuaciones deterministas. Estos fenómenos, así como otros relacionados con el caos han sido tradicionalmente asociados a mecanismos de *confusión* y *difusión* los cuales son la base principal de un buen sistema criptográfico.

### 3. ANTECEDENTES DEL PROYECTO

Desde la década de los 90's muchos investigadores han notado que existe una importante relación entre el caos y la criptografía: muchas propiedades de los sistemas caóticos tienen sus correspondientes contrapartes en los cripto sistemas tradicionales. Los sistemas caóticos tienen sus ecuaciones y sus condiciones iniciales, en donde la más mínima variación en sus condiciones iniciales provoca una evolución muy diferente en su comportamiento [2].

La teoría de funciones y mapas caóticos se presentan en varios trabajos: Fuan, Mengb, Zhanb, Zhuc, Laud, Tsed y Mae en el 2013 [3] proponen un esquema de protección de imágenes médicas basado en mapas caóticos; Gao y Chen en el 2008 [5] proponen un algoritmo nuevo de permutación de píxeles; Huang y Nien en el 2009 [8] proponen un sistema multi-caótico basado en el mismo principio de permutación de Gao y Chen; Patidar, Pareek y Sud en el 2009 [10] proponen un cifrado tipo sustitución-difusión basado en mapas logísticos y caóticos; Rhouma, Meherzi y Belghith en el 2009 [11] proponen el cifrado de imágenes a color basado en mapas caóticos; Sun, Liu y Li en el 2008 [15] proponen un esquema de encriptación basado en mapas caóticos espaciales; Tong y Cui en el 2009 [16] proponen un generador de cifrado de secuencias caóticas con componentes dinámicos; Wong, Kwok y Law en el 2008 [18] proponen un esquema de encriptación basado en el mapa caótico estándar; Xiao y Xia en el 2009 [19] proponen un esquema de encriptación usando mapas de permutación; Xu, Wang y Yang en el 2008 [20] proponen una mejora en el algoritmo de encriptación de imágenes que usan mapas caóticos; Ye en el 2009 [21] propone un cripto-sistema basado en las matrices de Töplitz y Hankel; Chong, Gao-Yuan, Mai, Zhe, y Wei-Min en el 2018 [1] proponen un cifrado para imágenes a color basado en caos donde realizan una sustitución de flujo de llave; Ranvir, Baihua, Alastair y Yan en el 2018 [11], proponen un algoritmo que combina mapas caóticos con el estándar AES; Xingyuan, Xiaoqiang y Yingqian en el 2018 [20], proponen un esquema de encriptación de imágenes, el cual emplea mapas caóticos mezclados y desplazamiento de Josephus.

El uso de autómatas celulares se utiliza en varios trabajos: Said y Saida en el 2014 [12], usan un autómata celular reversible para cifrar y descifrar texto; Shuiping y Huijune en 2012 [14], proponen un autómata celular caótico para cifrar imágenes digitales; Warakorn y Wimol en 2106 [17], usan reglas de autómata celular y funciones caóticas para cifrar imágenes digitales.

Como lo muestran los trabajos mencionados de los años recientes, los autómatas celulares y las funciones caóticas se han aplicado al campo de la criptografía, razón por la cual en



este trabajo se realiza una implementación computacional que permita cifrar y descifrar imágenes digitales a color.

#### 4. ALGORITMO

El algoritmo de cifrado consta de los pasos siguientes:

1. Se ingresa la imagen digital  $I$  de dimensión  $m \times n$  y se convierten los valores decimales de la imagen  $I$  a valores binario obteniendo  $I_b$ .
2. Se construye la primera llave de la forma siguiente: Se calculan  $k_1$  y  $k_2$  usando las siguientes ecuaciones:

$$k_1 = \frac{1}{126} \bmod \left( \sum_{i=1}^{m/2} \sum_{j=1}^n I_{ij}, 256 \right), \quad k_2 = \frac{1}{126} \bmod \left( \sum_{i=m/2+1}^m \sum_{j=1}^n I_{ij}, 256 \right).$$

Se obtienen los vectores:  $X, Y$  de la manera siguiente:

Se generan de forma aleatoria dos valores:  $x_0, y_0$  sobre el intervalo real  $[0, 1]$ . Se aplica el proceso siguiente para  $x_0$  (se repite el mismo proceso, pero usando  $y_0$  en lugar de  $x_0$ ) (Tere, Vaida, Alboae y Chiorea, 2011):

$$x_0 = x_1 + k_1$$

Si  $x_0 > 1$  entonces  $x_0 = \text{mod}(x_0, 1)$

Si no  $x_0 = x_0$ .

Se genera  $X = (x_1, \dots, x_{8m})$  como un vector columna y  $Y = (y_1, \dots, y_{8n})$  como un vector fila usando la siguiente función caótica conocida como *cross chaotic map* (Tere, Vaida, Alboae y Chiorea, 2011):

$$x_{i+1} = 1 - ny_i^2, \quad y_{i+1} = \cos(k \cos^{-1}(x_i)), \text{ donde } k \text{ es una constante.}$$

Se multiplica el vector columna  $X$  por el vector fila  $Y$ , obteniendo una matriz  $M$  de dimensión  $m \times 8n$ .

3. Se convierte la matriz  $M$  a binario donde se pone el valor a 0 si  $M(i, j) < 0$  y a valor 1 si  $M(i, j) > 0$ , el resultado se deja en la matriz  $M_b$ .
  4. Se realiza la operación lógica xor de  $M_b$  con  $I_b$  dejando el resultado en  $Ic1$ .
  5. Se procede a la generación de la segunda llave de la forma siguiente: Se genera de forma aleatoria el vector  $Is$  de dimensión  $1 \times 8n$  con valores binarios. Se selecciona de forma aleatoria alguna de las reglas del AC y se itera  $m-1$  veces dejando el resultado en los renglones 2 a  $m$  de  $Is$ . Por lo tanto,  $Is$  es una matriz de dimensión  $m \times 8n$ .
  6. Se realiza la operación lógica xor de  $Is$  con  $Ic1$  dejando el resultado en  $Ic2$ .
  7. Se pasa  $Ic2$  a valores decimales obteniendo  $C$ .
- El resultado final es la matriz  $C$ , la cual es la imagen cifrada.

Para el algoritmo de descifrado se siguen los siguientes pasos:

1. Pasa  $Ic2$  a binario obteniendo  $Ic2b$ .
2. Realiza la operación xor de  $Ic2b$  con la segunda llave  $Is$  obteniendo  $D1$ .
3. Realiza la operación xor de  $D1$  con la primera llave  $M_b$ , dejando el resultado en  $D2$ .

4. Pasa **D2** a valores decimales obteniendo **D**.

Al finalizar el algoritmo de descifrado, **D** debe contener la imagen original.

## 5. PRUEBAS

La implementación computacional se desarrolló bajo el lenguaje de programación OCTAVE [23] y las pruebas se realizaron en una computadora *hp* con 16Gb de memoria RAM y un procesador Intel core i7 a 2.6 GHz.

La figura 2a, muestra la imagen de la enfermedad *actinomicosis bovina* [24], la imagen tiene una dimensión 480 x 640, es decir, 480 filas y 640 columnas. La actinomicosis bovina es una enfermedad infecciosa de los bovinos que se caracteriza por la afección de las estructuras óseas, en particular de la mandíbula o quijada. La actinomicosis es una enfermedad que se conoce como el cáncer de mandíbula en bovinos, aunque afecta también a ovinos, caprinos y otras especies incluido el humano. La figura 2b ilustra el resultado del algoritmo de cifrado desarrollado en este trabajo, al cual se le denominará AACFC, en la figura 2c está el resultado del cifrado cuando se utiliza el estándar de criptografía simétrica conocido como algoritmo AES y la figura 2d muestra el descifrado de ambos algoritmos el cual siempre va a coincidir con la imagen original ya que el descifrado de ambos algoritmos siempre recuperan la imagen original. En este caso, el algoritmo AACFC reporta un tiempo de 140.3215 segundos en cifrar y descifrar mientras que el algoritmo AES lo realizó en 884.7273 segundos, es decir, AACFC fue 6.3 veces más rápido que AES.

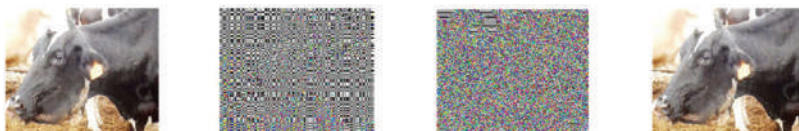


Figura 2: a) Imagen original de la enfermedad actinomicosis bovina, b) Imagen cifrada por la propuesta, c) Imagen cifrada por el algoritmo AES, d) Imagen descifrada por ambos algoritmos.

La figura 3a, muestra la imagen de la enfermedad *antracnosis* en la fruta de mango [25], la imagen tiene una dimensión 299 x 400. La antracnosis conocida también como cancro o chancro es un síntoma de enfermedad de las plantas de zonas calurosas y húmedas, causada por un hongo que puede ser generalmente de los géneros *Colletotrichum*, *Gloeosporium*, o la especie *Coniothyriumfuckelii*. La figura 3b ilustra el resultado del algoritmo de cifrado AACFC, en la figura 3c está el resultado del cifrado del algoritmo AES y la figura 3d muestra el descifrado de ambos algoritmos. En este caso, el algoritmo AACFC reporta un tiempo de 55.0867 segundos en cifrar y descifrar mientras que el

algoritmo AES lo realizó en 345.8781 segundos, es decir, AACFC fue 6.28 veces más rápido que AES.

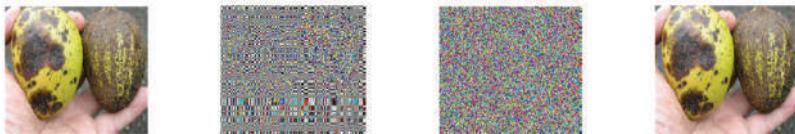


Figura 3: a) Imagen original de la enfermedad antracnosis en la fruta de mango, b) Imagen cifrada por la propuesta, c) Imagen cifrada por el algoritmo AES, d) Imagen descifrada por ambos algoritmos.

La figura 4a, muestra la imagen de la enfermedad *brucellosis caprina* [26], la imagen tiene una dimensión 532 x 800. Esta enfermedad es causada por la bacteria *Brucella Melitensis* provocando infección en las ubres de las cabras y en algunos casos abortos. Esta enfermedad la adquieren las cabras a través del agua y por alimentos contaminados debido a desechos de cabras recién paridas. La figura 4b ilustra el resultado del algoritmo de cifrado AACFC, en la figura 4c está el resultado del cifrado del algoritmo AES y la figura 4d muestra el descifrado de ambos algoritmos. En este caso, el algoritmo AACFC reporta un tiempo de 194.3886 segundos en cifrar y descifrar mientras que el algoritmo AES lo realizó en 1246.7 segundos, es decir, AACFC fue 6.4 veces más rápido que AES.

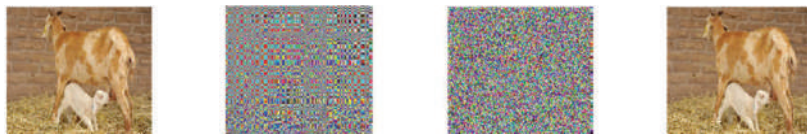


Figura 4: a) Imagen original de la enfermedad brucelosis, b) Imagen cifrada por la propuesta, c) Imagen cifrada por el algoritmo AES, d) Imagen descifrada por ambos algoritmos.

La figura 5a, muestra la imagen de la enfermedad *saprolegnia parasítica* en peces [27], la imagen tiene una dimensión 266 x 350. La *saprolegnia parasítica* se le reconoce como el mayor eucariota patógeno de peces de agua dulce. La figura 5b ilustra el resultado del algoritmo de cifrado AACFC, en la figura 5c está el resultado del cifrado del algoritmo AES y la figura 5d muestra el descifrado de ambos algoritmos. En este caso, el algoritmo AACFC reporta un tiempo de 58.0307 segundos en cifrar y descifrar mientras que el algoritmo AES lo realizó en 357.6577 segundos, es decir, AACFC fue 6.2 veces más rápido que AES.

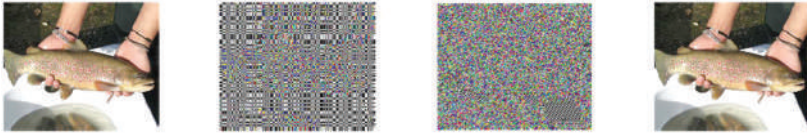


Figura 5: a) Imagen original de la enfermedad saprolegnia parasítica, b) Imagen cifrada por la propuesta, c) Imagen cifrada por el algoritmo AES, d) Imagen descifrada por ambos algoritmos.

La figura 6a, muestra la imagen de un paciente que tiene la enfermedad de sarna [28]. La imagen tiene dimensión de 407 x 700 píxeles. La sarna es una enfermedad de la piel causada por pequeños parásitos. Se transmite por contacto de piel a piel, usualmente durante el sexo. La sarna no es peligrosa y se puede curar. La figura 6b ilustra el resultado del algoritmo de cifrado AACFC, en la figura 6c está el resultado del cifrado del algoritmo AES y la figura 6d muestra el descifrado de ambos algoritmos. En este caso, el algoritmo AACFC reporta un tiempo de 170.7866 segundos en cifrar y descifrar mientras que el algoritmo AES lo realizó en 1086.7 segundos, es decir, AACFC fue 6.36 veces más rápido que AES.



Figura 6: a) Imagen original de la enfermedad sarna, b) Imagen cifrada por la propuesta, c) Imagen cifrada por el algoritmo AES, d) Imagen descifrada por ambos algoritmos.

## 6. SEGURIDAD DEL CRIPTOSISTEMA

### 6.1 ATAQUE A FUERZA BRUTA

El ataque a fuerza bruta, consiste en probar todas las posibles llaves hasta encontrar la llave con la cual se pueda recuperar el mensaje original. Para considerar que el espacio de llaves es adecuado para resistir un ataque de fuerza bruta, este espacio debe ser superior a  $2^{100} = 1.2677 \times 10^{30}$  [12]. El sistema criptográfico que aquí se presenta, depende de 2 llaves, la primera generada a partir de la función caótica, y la segunda a partir del autómata celular. Por lo tanto, el total de posibles llaves para una imagen de tamaño  $m \times n$  es:  $2^m \times 2^{8n} + 2^{8n} \times 2^8 = 2^{8n}(2^m + 2^8)$ . Si  $m = n = 32$  se tiene que el total de posibles llaves es igual a  $8.7502^{55}$ . De esto se concluye que el cripto sistema propuesto es seguro ante ataques a

fuerza bruta siempre y cuando la imagen sea al menos de tamaño 32x32. Por ejemplo, si  $m = n = 16$  el total de posibles llaves es  $9.3908^{27}$  y el cripto sistema no dará seguridad suficiente ante el ataque a fuerza bruta.

## 6.2 ANÁLISIS ESTADÍSTICO

El análisis estadístico que se presenta es el de la distribución de valores de los píxeles. Un buen cripto sistema deberá enmascarar de modo que el análisis sobre la frecuencia de los valores de los píxeles sea infactible [1]. La Figura 8, muestra en la línea 1, los histogramas de los canales rojo verde y azul de la imagen de la figura 3a, la línea 2 muestra los canales del cifrado por la propuesta y la línea 3 muestra los canales del cifrado AES. Se puede apreciar que ambos métodos distribuyen de forma uniforme los cifrados, por lo que se puede concluir que el cripto sistema propuesto da seguridad suficiente bajo un ataque sobre la frecuencia de los datos.

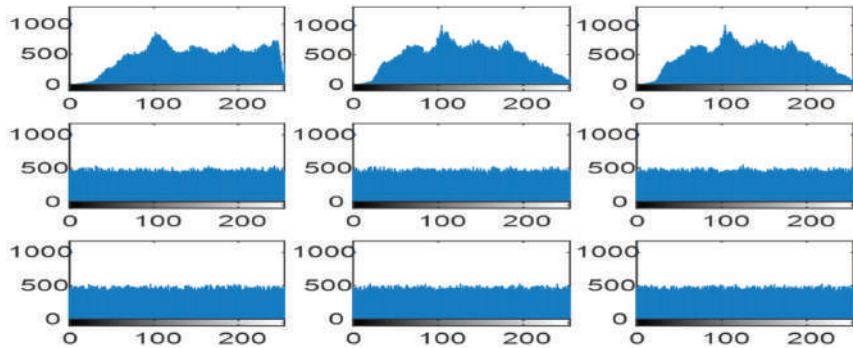


Figura 8: Se muestran los histogramas de los canales RGB. Línea 1, histogramas de la imagen de la figura 3a, línea 2 histogramas del cifrado propuesto y línea 3 histogramas del estándar de la criptografía simétrica.

## 7. CONCLUSIONES

Como se puede apreciar en las imágenes utilizadas en este trabajo, los resultados de la implementación computacional del algoritmo de cifrado dejan irreconocible el tipo de imagen que se trata y también se observa que el descifrado obtiene la imagen original, por supuesto que lo mismo pasa con el algoritmo estándar AES.

El criptosistema presentado es seguro ante ataques a fuerza bruta siempre y cuando el tamaño de la imagen sea al menos de dimensión 32x32, también es seguro ante ataques sobre la frecuencia de los datos. Además, como lo muestran las pruebas, la propuesta es una alternativa 6 veces más rápida que el estándar AES de la criptografía simétrica, por

lo que es aplicable para grandes volúmenes de datos como lo es el caso de las imágenes digitales.


La implementación computacional cifra y descifra imágenes digitales a color en pocos minutos para imágenes de dimensiones menores o iguales a 800x800.

Como trabajo futuro se analizará más la seguridad de la propuesta ante otros tipos de ataques.

## REFERENCIAS

- [1]. CHONG, F., GAO-YUAN, Z., MAI, Z., ZHE, CH., and WEI-MIN, L., (2018): A New Chaos-Based Color Image Encryption Scheme with an Efficient Substitution Keystream Generation Strategy, **Security and Communication Networks**, vol. 1, 1-13.
- [2]. DELFS, H., MAURER, U. and KNEBL, H., (2007): **Introduction to Cryptography Principles and Applications, Second Edition**, Springer-Verlag Berlin Heidelberg.
- [3]. FUAN, Ch., MENG, W., ZHAN, Y., ZHUC, Z., LAUD, F., TSED, Ch. And MAE, H. (2013) An efficient and secure medical image protection scheme based on chaotic maps. **Computers in Biology and Medicine, Elsevier**, vol. 43, 1000-1011.
- [4]. GALENDE, J. (1995): **Criptografía, historia de la escritura cifrada, 1ra edición**, editorial complutense.
- [5]. GAO, T.G., CHEN, Z.Q., (2008): Image encryption based on a new total shuffling algorithm, **Chaos Solutions & Fractals**, 38 (1), 213–220.
- [6]. GOLDREICH, O., (1999): **Modern Cryptography, theory and practice, discrete mathematics and its applications 3th edition**, Chapman & Hall.
- [7]. HILBORN, R.C. (2000): **Chaos and nonlinear dynamics 2da edición**, Oxford University Press.
- [8]. HUANG, C.K. and NIEN, H.H., (2009): Multi chaotic systems based pixel shuffle for image encryption, **Opt. Commun.**, 282 (11), 2123–2127.
- [9]. OTT, E. (1993): **Chaos in dynamical systems**, Cambridge University Press.
- [10]. PATIDAR, V., PAREEK, N.K. and SUD, K.K. (2009): A new substitution–diffusion based image cipher using chaotic standard and logistic maps, **Commun. Nonlinear Sci. Numer. Simulation**, 14 (7) 3056–3075.
- [11]. RANVIR, S., BAIHUA, L., ALASTAIR, G. and YAN, CH. (2018): Medical image encryption using chaotic map improved advanced encryption standard, **International Journal of Information Technology and Computer Science (IJITCS)**, 10 (8), pp.1-10. (REFERENCIAR)
- [12]. RHOUMA, R., MEHERZI, S. and BELGHITH, S. (2009): OCML-based colour image encryption, **Chaos Solitons & Fractals**, 40 (1), 309–318.

- [13]. SAID BOUCHKAREN and SAIDA LAZAAR (2014): A Fast Cryptosystem Using Reversible Cellular Automata, **International Journal of Advanced Computer Science and Applications**, 5 (5) 207-210.
- [14]. SHUIPING ZHANG and HUIJUNE LUO (2012): The Research of Image Encryption Algorithm Based on Chaos Cellular Automata, **Journal of Multimedia**, 7 (1) 66-73.
- [15]. SUN, F.Y., LIU, S.T. and LI, Z.Q. (2008): A novel image encryption scheme based on spatial chaos map, **Chaos Solitons & Fractals**, 38 (3), 631–640.
- [16]. TONG, X.J. and CUI, M.G., (2009): Image encryption scheme based on 3D baker with dynamical compound chaotic sequence cipher generator, **Signal Processing**, 89 (4) 480–49 1.
- [17]. WARAKORN SRICHAVENTSUP and WIMOL SAN-UM (2106): Data Encryption Scheme Based on Rules of Cellular Automata and Chaotic Map Function for Information Security, **International Journal of Network Security**, 18 (6) 1130-1142.
- [18]. WONG, K.W., KWOK, B.S. and LAW, W.S. (2008): A fast image encryption scheme based on chaotic standard map, **Phys. Lett. A.**, 372 (15), 2645–2652.
- [19]. XIAO Y.L. and XIA L.M. (2009): An Image Encryption Approach Using a Shuffling Map, **Commun. Theor. Phys**, 52 (5) 876–880.
- [20]. XINGYUAN, W., XIAOQIANG, Z. and YINGQIAN, Z., (2018): An Image Encryption Algorithm Based on Josephus Traversing and Mixed Chaotic Map, IEEE Access, vol 6, 23723-23746. (FALTA REF)
- [21]. XU, S.J., WANG, J.Z. and YANG, S.X. (2008): An improved image encryption algorithm based on chaotic maps, **Chin. Phys. B**, 17 (11), 4027–4032.
- [22]. YE, G.D. (2009): A chaotic image cryptosystem based on Toeplitz and Hankel matrices, **Imaging Sci. J.**, 57 (5) 266–273.
- [23]. OCTAVE, Disponible en <https://www.gnu.org/software/octave/>, consultada 10-10, 2018.
- [24]. Enfermedad actinomicosis bovina. Disponible en <https://manuelantonioweb.wordpress.com/2017/04/26/actinomicosis/>, consultada 20-10, 2018.
- [25]. Enfermedad antracnosis en la fruta de mango. Disponible en <https://www.fps.org.mx/portal/index.php/notas/686-prevenga-antracnosis-con-poda-en-mango>, consultada 20-10, 2018.
- [26]. Enfermedad brucelosis caprina. Disponible en <http://www.elsemiarido.com/caida-historica-de-los-casos-humanos-de-brucelosis-en-todo-mendoza-y-san-juan/>, consultada 20-10, 2018.
- [27]. Enfermedad Saprolegnia parasítica. Disponible en <http://www.fis.com/fis/techno/newtechno.asp?l=s&id=62233&ndb=1>, consultada 20-10, 2018.
- [28]. Enfermedad Sarna en humanos. Disponible en <http://chicaplus.com/remedios-caseros-para-eliminar-la-sarna/>, consultada 20-10, 2018.



APORTES Y APLICACIONES  
EN LAS CIENCIAS  
COMPUTACIONALES

Gustavo Trinidad Rubín Linares



**APORTES Y  
APLICACIONES EN LAS  
CIENCIAS COMPUTACIONALES**

María del Carmen Santiago Díaz  
Gustavo Trinidad Rubín Linares  
María de Lourdes Sandoval Solís  
Yeiny Romero Hernández  
(Editores)

Gustavo Trinidad Rubín Linares  
(Coordinador)

*María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, María de Lourdes Sandoval Solís, Yenny Romero Hernández*  
(editores)

*Gustavo Trinidad Rubín Linares*  
(coordinador)

*María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Ana Cinadia Zenteno Vázquez, Judith Pérez Marcial, Yenny Romero Hernández, Nicolás Oubros Hernández, Alma Delia Ambrosio Vázquez, Héctor David Ramírez Hernández, Gregorio Trinidad García, María de Lourdes Sandoval Solís, Rogelio González Veldaquez, Guillermina Sánchez Román, José Luis Hernández Ameca, Armando Espindola Posas, María Blanca García, Melissa Contreras González, Luis Enrique Cobencares Guillén, Elsa Charvra Martínez, Pedro García Juárez, Nelva Betabel Espinoza Hernández, Alberto Román Flores, Roberto Contreras Juárez, Graciano Cruz Almanza, Maya Carrillo Ruiz, Ernest Cortez, Bárbara Emma Sánchez Ríza, Beatriz Beltrán Martínez, José Ítalo Cortez, Hermes Moreno Álvarez.*  
(revisores)

Primera edición: 2019

ISBN: 978-607-7512-96-7

Montiel & Soriano Editores S.A. de C.V.  
15 sur 1103-6 Col. Santiago Puebla, Pua.

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

*Rector:*

Dr. José Alfonso Espanza Ortiz

*Secretario General:*

Dr. José Jaime Vázquez López

*Vicerrector de Investigación y Estudios de Posgrado*

Dr. Ygnacio Martínez Laguna

*Directora de la Facultad de Ciencias de la Computación*

M.I. María del Consuelo Molina García

## Contenido

<b>Prefacio</b> .....	7
<b>Estudio y Análisis de Seguridad en un Sistema de Información Web</b> <i>Luz María Hernández Cruz</i> <i>Diana Concepción Méx Álvarez</i> <i>Margarita Castillo Téllez</i> <i>Oscar Esquivel De la Cruz Ek</i> .....	8
<b>Análisis de Fraudes en Transacciones Bancarias Aplicando Minería de Datos</b> <i>Pictor Giovanni Morales Murillo</i> <i>María Josefa Somodivilla García</i> <i>Magali Fong Juárez</i> <i>Ivo Humberto Pinada Torres</i> .....	12
<b>Determinación del Tamaño de una Mano en Imágenes Radiográficas Utilizando Landmarks</b> <i>José Israel Saenz Gtrón</i> <i>Jonathan Gabriel Salazar Rolán</i> <i>Salvador Eugenio Ayala Roggi</i> <i>Aldrin Barreto Flores</i> .....	20
<b>Detección de Desbalance Financiero Utilizando Modelos Ocultos de Markov</b> <i>María de Lourdes Sandoval Solís</i> <i>Marcela Rivera Martínez</i> <i>Luis René Marcela Castillo</i> <i>Glady Linares Fleites</i> .....	28
<b>Intercomunicador con Cámara y Conectividad WiFi</b> <i>Joel Fernando Acevedo Ruiz</i> <i>Aldrin Barreto Flores</i> <i>Salvador Eugenio Ayala Roggi</i> <i>Verónica Esthèr Bautista López</i> .....	40
<b>Solución al Problema de la Programación de Producción - Inventario Usando Código LINGO</b> <i>Rogelio González Velázquez</i> <i>María Beatriz Bernabé Lavanca</i> <i>Adriana Méndez Montero</i> <i>Abraham Sánchez López</i> .....	56
<b>Sistema Multi Sensor para Medición, Procesamiento y Envío de Parámetros Ambientales en Tiempo Real</b> <i>María del Carmen Santiago Díaz</i> <i>Gustavo Trinidad Rubén Lourea</i> <i>Hermes Moreno Álvarez</i> <i>Antonio Eduardo Álvarez Nolas</i> .....	63
<b>Desarrollo de un Robot Interactivo Como Asistente en Terapias para Niños con Autismo</b> <i>José Luis Hernández Ameca</i> <i>Gregorio Trinidad García</i> <i>José Isidro Cortés</i> <i>Juan Chávez Bravo</i> .....	68
<b>Diseño de un Shield con Microcontrolador ARM-Cortex-M0</b>	

## Detección de Desbalance Financiero Utilizando Modelos Ocultos de Márkov

María De Lourdes Sandoval Solís, Marcela Rivera Martínez, Luis René Marcial Castillo,  
Gladys Linares Fleites

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla,  
4 Sur 104, Colonia Centro Puebla, 72000

<sup>1</sup>maria.sandoval@correo.buap.mx, <sup>2</sup>lrcm, lmc}@cs.buap.mx, <sup>4</sup>gladyslinaresl@yaho.es

**Resumen.** En este trabajo se presenta la aplicación de los Modelos Ocultos de Márkov para detectar el desbalance financiero. Se muestra la convergencia de la matriz de transición al aumentar el tamaño de la secuencia de observaciones. Se ejemplifica el método con la base de datos Berka.

**Palabras Clave:** Modelos Ocultos de Márkov, Detección de Desbalance Financiero.

### 1 Introducción

Los Modelo Oculto de Márkov HMM (Hidden Márkov Model), tradicionalmente se han utilizado en el área de computación para detectar patrones y detectar los cambios de éstos. En las cadenas de Márkov, los estados son observables, por esto también se les conoce como Modelos de Márkov Observables, sin embargo, para varios problemas de interés el estado no siempre es observable, como puede ser en el reconocimiento de voz, en este únicamente se ve la señal de la voz, pero no el conjunto de palabras ó fonemas que lo conforman, a este tipo de estados se les llama ocultos. Un Modelo Oculto de Márkov (HMM) permite modelar este tipo de problemas.

El objetivo principal de este trabajo es determinar el Modelo Oculto de Márkov HMM que se adapte al desbalance de transacciones bancarias de un usuario. Mediante este modelo determinar el desbalance bancario. Para este objetivo se aplicó la técnica a dos cuentas bancarias de la base datos conocida como Berka [1].

### 2 Modelos Ocultos de Márkov

Un Modelo Oculto de Márkov (HMM) es un proceso estocástico doblemente embebido con un proceso aleatorio que no es observable (está oculto), pero que puede ser visto solamente a través de otro conjunto de procesos aleatorios que produzcan la secuencia de observaciones y satisfice la siguiente propiedad: Propiedad de Márkov: Esta propiedad hace referencia a la falta de memoria de una proceso estocástico, esto es, el estado actual de un proceso únicamente depende de su estado anterior y no de la secuencia de eventos que precedieron a éste último estado. Los HMM está definido por los siguientes componentes:

- $Q$ : Número de estados en el modelo, cada estado individual se etiqueta como  $q_1, q_2, \dots, q_N$ .
- $O$ : Secuencia de observaciones tomadas de un vocabulario  $V = \{v_1, v_2, \dots, v_M\}$ , denotadas como  $O = (O_1, O_2, \dots, O_M)$ .
- $A$ : la matriz de transición  $A = \{a_{ij}\}$ , donde cada  $a_{ij}$  representa la probabilidad del estado  $i$  al estado  $j$ :  $a_{ij} = P(q_{t+1} = j | q_t = i)$ .
- $B$ : Probabilidad de observación de cada símbolo del vocabulario, también llamada probabilidad emisión,  $b_j(k) = P(O_t = v_k | q_t = j)$ , es decir, la probabilidad de que la observación  $O_t$  sea generada por el estado  $j$ .

- $\pi$ : Distribución de probabilidad sobre el estado inicial  $\pi = \{\pi_i\}$ , definida como:  $\pi_i = P(q_{1=i})$ , es decir, la probabilidad de iniciar en alguno de los estados de  $q$ .

Es evidente que una especificación completa de un HMM requiere la estimación de los parámetros del modelo,  $A$ ,  $B$  y  $\pi$  por lo tanto un HMM se define como  $\lambda$ .

$$\lambda = (A, B, \pi) \quad (1)$$

Este modelo toma dos suposiciones importantes, la primera, al igual que en las cadenas de Markov de primer orden es la suposici3n de Markov, el estado actual unicamente depende de su predecesor, expresado como:

$$\text{Suposici3n de Markov } P(q_t | q_1, \dots, q_{t-1}) = P(q_t | q_{t-1}) \quad (2)$$

Como segunda suposici3n, la probabilidad de observaci3n  $O_t$  depende unicamente del estado que la produce  $q_t$  y no de cualquier otro estado u observaci3n, expresado matematicamente como:

$$\text{Independencia de salida: } P(O_t | q_1, \dots, q_{t-1}, q_t, o_1, \dots, o_{t-1}, o_t) = P(o_t | q_t) \quad (3)$$

### 3 Tres Problemas De Los HMM

Los HMM introducidos por Rabiner (1989) [2], se caracterizan por tres problemas esenciales, necesarios a resolver para hacer que los HMM sean aplicables a aplicaciones reales.

- Problema 1 (Probabilidad): El primer problema consiste en la evaluaci3n de una secuencia de observaciones  $O$  dado el modelo  $\lambda = (A, B, \pi)$ . La soluci3n a este problema nos brinda la probabilidad de que la secuencia fuera generada por ese modelo  $\lambda$ , es decir:  $P(O|\lambda)$ .
- Problema 2 (Descodificaci3n): Dada una secuencia de observaci3n  $O$  y el modelo  $\lambda = (A, B, \pi)$ , determinar cual es la secuencia de estados  $Q$  mas probable que haya generado la observaci3n  $O$ .
- Problema 3 (Aprendizaje): Ajustar los parametros  $A$  y  $B$  del modelo para maximizar la probabilidad de observaci3n  $P(O|\lambda)$ .

#### 3.1 Probabilidad

Para obtener la probabilidad total de la observaci3n, se tiene que sumar sobre todas las posibles secuencias de  $Q$  estados ocultos, en general:

$$P(O|\lambda) = \sum_Q P(O, Q) = \sum_Q P(O|Q) P(Q) \quad (4)$$

Dada esto podemos notar que si tenemos un numero de estados ocultos muy grande al igual que el tamano de la cadena de observaci3n es muy grande tenemos un problema computacional del orden exponencial, lo cual implica que el tiempo de c3mputo sera muy grande.

Para atacar el problema de tiempo computacional al que hemos llegado tendremos que hacer uso del algoritmo Forward que es del orden  $O(N^2T)$ , el cual evita hacer cálculos repetidos.

### 3.1.1 Algoritmo Forward:

Considerando la siguiente variable  $\alpha_t(j)$ , definida de la siguiente forma:

$$\alpha_t(j) = P(O_1, O_2, \dots, O_t, q_t = j | \lambda) \quad (5)$$

Esto representa la probabilidad de estar en el estado  $j$  después de  $t$  observaciones. Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\alpha_t(j) = \left( \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad (6)$$

#### Algoritmo Forward

Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

Recursividad:

$$\alpha_t(j) = \left( \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad 1 \leq j \leq N, 1 < t \leq T$$

Terminación:

$$P(O | \lambda) = \left( \sum_{i=1}^N \alpha_T(i) \right)$$

### 3.1.2 Algoritmo Backward

Ahora consideremos una variable  $\beta_t(i)$ , definida de la siguiente forma:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = i | \lambda) \quad (7)$$

Que es la probabilidad de ver las observaciones desde el tiempo  $t+1$  hasta el final, y estar en el estado  $i$  al tiempo  $t$ . Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (8)$$

Para más detalles de la demostración de las recurrencias de los algoritmos consulte [2.4].

Algoritmo Backward:

**Inicialización:**

$$\beta_T(i) = 1 \quad 1 \leq i \leq N$$

**Recursividad:**

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \beta_{t+1}(j) b_j(O_{t+1}) \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N$$

**Terminación:**

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(O_1) \beta_1(j)$$

### 3.2 Descodificación

El segundo problema del HMM es la descodificación, esto consiste en obtener la secuencia de estados ocultos  $Q$  que genere la observación  $O$ , es decir, de entre todas las secuencias de estados que puedan generar determinada secuencia de observaciones, se desea encontrar aquella que sea más representativa de la observación  $O$ . Para ello tenemos que encontrar cual es la secuencia de estados ocultos con mayor probabilidad. Podemos resolver este problema ejecutando el algoritmo Viterbi que es muy similar al algoritmo Forward.

Definimos la siguiente variable  $v_t(j)$ , de la siguiente forma:

$$v_t(j) = \max_{q_0, q_1, \dots, q_{t-1}} P(q_0, q_1, \dots, q_{t-1}, O_1, O_2, \dots, O_t = j | \lambda) \quad (9)$$

Como la probabilidad de estar en el estado  $j$  después de  $t$  observaciones y pasar sobre la secuencia de estados más probables,  $q_0, q_1, \dots, q_{t-1}$ . Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con el cálculo del valor máximo de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (10)$$

Ahora se define también la variable denominada *backtrace* ( $bt$ ) similar a la variable del algoritmo de Viterbi ( $v$ ), en lugar de mantener la probabilidad más alta, esta variable mantiene el número del estado con dicha probabilidad.

$$bt_t(j) = \arg \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (11)$$

El valor de cada celda se puede calcular mediante el siguiente algoritmo:

#### 3.2.1 Algoritmo Viterbi.

**Inicialización:**

$$v_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

$$b_1(i) = 1 \quad 1 \leq i \leq N$$

Recursividad:

$$v_j(j) = \max_{i=1}^N v_{i-1}(i) a_{ij} b_j(O_i) \quad 1 \leq j \leq N, \quad 1 < i \leq T$$

$$\Delta t_i(j) = \arg \max_{i=1}^N v_{i-1}(i) a_{ij} b_j(O_i) \quad 1 \leq j \leq N, \quad 1 < i \leq T$$

Terminación:

$$P^* = \max_{i=1}^N v_T(i)$$

$$q_T^* = \arg \max_{i=1}^N v_T(i)$$

Para las demostraciones de las relaciones utilizadas en los algoritmos revisar los documentos de las referencias [4,5]

## 4 Desbalance de cuentas bancarias

### 4.1 Información seleccionada de la base de datos.

La información contenida en la base de datos se separó por el número de cuenta con sus respectivas transacciones y balance de cuenta. Posteriormente se seleccionaron 2 cuentas bancarias que pudieran aportar diferentes puntos de vista sobre el comportamiento de los movimientos que se realizan. Para el experimento se tomó una cuenta con balance positivo y una cuenta con balance negativo. Las 2 cuentas seleccionadas se muestran en la tabla las cuentas y el estado de balance.

Tabla 1. Cuentas bancarias de la Base de Datos Berka [1]

Número de cuenta	Tipo de balance	Cantidad obtenida del balance	Estado de balance
1069	Balance positivo	25230.50	Positivo
2335	Balance negativo cercano al neutro	2765.50	Negativo

Mostramos a continuación el comportamiento en las transacciones de las cuentas seleccionadas. Nos enfocamos en los montos que representa cada transacción y el balance que tiene la cuenta en cada momento de la transacción.

#### Cuenta 1069

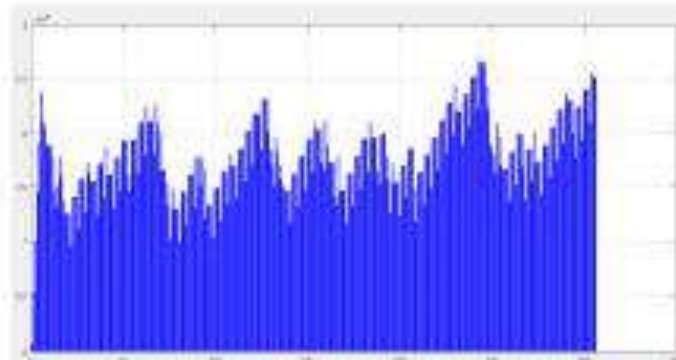


Fig. 1. Comportamiento del balance positivo en la cuenta 1069



**Cuenta 2335**

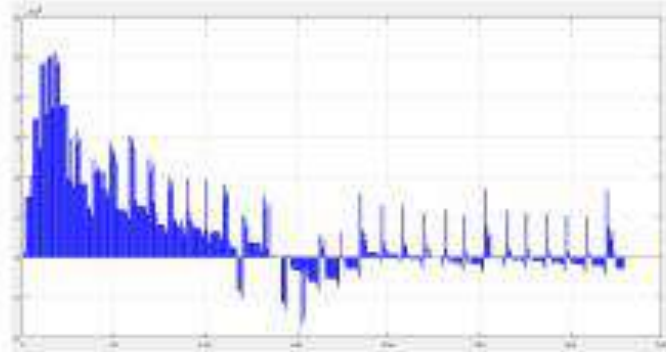


Fig. 2. Comportamiento del balance negativo en la cuenta 2335

**4.2 Determinación del Modelo Oculto de Márkov**

En el caso sobre el análisis de la Base de datos Berka, se usa la información de Transacciones [1], se toman como estados del modelo las variables que constaban en el campo Operation, se añadió una sexta variable dado que la información de la base de datos contenía en ese campo datos faltantes por lo cual se le asignó un estado nuevo a ese campo faltante. Para definir las observaciones del modelo se siguió lo propuesto en el artículo Srivastava (2008) [3], dividir las diferentes cantidades de las transacciones en tres particiones usando el algoritmo del vecino más cercano con tres centroides obteniendo la observación por estado en la forma {1,2,3}.

Se muestra en la figura la determinación del HMM para cada cuenta bancaria.

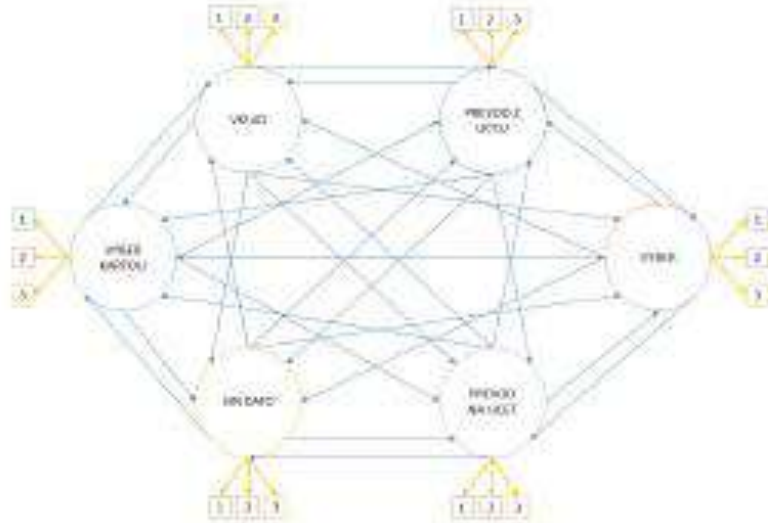


Fig. 3. Modelo Oculto de Márkov para las cuentas bancarias en la base de datos Berka.

### 4.3 Convergencia de las matrices de transición

Analizando el entrenamiento que se realiza en las matrices A y B del HMM, notamos que entre mayor sea el tamaño de muestra para entrenar estas matrices, se obtiene una convergencia para cada término de la matriz de transición. A continuación, se muestran los valores de la transición del estado  $i$  al estado  $j$  conforme aumenta el número de observaciones. Se muestra los resultados para una matriz de transición de tres estados.



Fig. 4. Convergencia de cada término de la matriz de transición.

### 4.4 Determinación del Desbalance Financiero

Para esta aplicación se utiliza el Toolbox HMM de MATLAB [6].

#### 4.4.1 Cuenta Bancaria 1069

En el caso de la cuenta 1069 no tenemos ninguna transacción bancaria en la cual se ocasione un balance negativo. Se diseñó el siguiente proceso para determinar si existe o no desbalance:

*Paso uno:* Se determina un número aleatorio, que en las secuencias de observaciones se considera como la observación  $O_{i-1}$  que se quiere determinar si es un desbalance o no.

*Paso dos:* Para entrenar el HMM se calculan las matrices de emisión y transición, se usa la secuencia de observaciones de  $O_1, \dots, O_i$  y sus respectivos estados.

```
[MatrizdeTransicion, MatrizdeEmision] = hmmestimate
(Observaciones, Estados);
```

Matriz de transición

0	0	0	1	0	0
0	0.23952632	0.15789474	0.10526316	0.42105263	0.10526316
0	0	0.08088889	0.33333333	0	0.57777778
0	0.08333333	0.05	0.275	0.4	0.19166667
0.01639344	0.03278689	0.52459016	0.19672131	0.08196721	0.14754098
0	0.03389331	0	0.98610169	0	0

Matriz de emisión

1	0	0
0.15789474	0.15789474	0.68421053
0	1	0
0.74166667	0.05	0.20833333
0.01639344	0.19672131	0.78688525
1	0	0

Paso tres: Determinar la secuencia de observaciones inicial de tamaño  $N$ ,  $O_1, \dots, O_N$ ,

siendo  $O_N = Q$

Secuencia Observaciones Inicial

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia Estados Inicial

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso cuatro: Determinar la secuencia de estados más probable

A partir de la rutina usada en Matlab (`hmmviterbi`) se calcula una secuencia de estados con la mayor probabilidad.

```
EstadosGenerados = hmmviterbi (SecuenciaInicial,
MatrizdeTransicion, MatrizdeEmision);
```

Secuencia de estados más probable obtenida

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Tomando esta secuencia de estados generada por la rutina se compara con la secuencia de estados de entrada.

```
Porcentaje = sum (Estados==EstadosGenerados) /NEJEMPLOS;
```

Porcentaje obtenido: 0.7000

Paso cinco: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados

Esta se obtiene usando la rutina en MATLAB `hmmdecode`

```
[ProfEstados, LogSecuencia] = hmmdecode (SecuenciaInicial,
MatrizdeTransicion, MatrizdeEmision);
```

Logaritmo de la probabilidad de la secuencia de estados Inicial: -18.5465

Paso seis: Determinar la secuencia de observaciones final de tamaño  $N+1$ ,  $O_1, \dots, O_{N+1}$

siendo  $O_{N+1} = Q$ , que incluye la observación que se desea determinar si es desbalance financiero

Secuencia Observaciones Final

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Para esta cadena de 21 observaciones se repite los Pasos Cuatro y Cinco para calcular la secuencia de estados más probable, usando `hmmdecode` y el logaritmo de la probabilidad usando `hmmdecodes`.

Logaritmo de la probabilidad de la secuencia de estados Final: -19.1147

*Paso siete:* Se calcula Delta definida como la diferencia de Logaritmo de la probabilidad de la secuencia de estados Final menos Logaritmo de la probabilidad de la secuencia de estados Inicial.

Delta: -0.4682

También se calcula la Delta proporcional respecto a Logaritmo de la probabilidad de la secuencia de estados Inicial.

Delta proporcional = 0.0231

Del paso tres al siete se realiza para diferentes tamaños de N. En cada una de las corridas se genera un número aleatorio diferente. En la siguiente tabla se reporta el promedio de cinco corridas para cada tamaño de la secuencia de observaciones inicial.

Tabla 2. Comportamiento de la Cuenta bancaria 1069 de la Base de Datos Beca [1], con balance positivo

Longitud	Delta	Delta Proporcional
20	0.56176	0.0354
17	0.56176	0.04016
14	0.56176	0.0544
11	0.56176	0.05748
8	0.56176	0.0671
5	0.56288	0.14836

#### 4.4.2 Cuenta Bancaria 2335

En el caso de la cuenta 2335 encontramos un número considerable de transacciones que generan en algún punto un balance negativo, que después se recupera a positivo sin embargo para efectos del experimento entrenamos el HMM un elemento antes del primer balance negativo.

*Paso uno:* Encontrar la primera observación donde se genera un desbalance negativo en la cuenta, en las secuencias de observaciones se considera como la observación  $O_{n-1}$ .

*Paso dos:* Para entrenar el HMM se calculan las matrices de emisión y transición, se usa la secuencia de observaciones de  $O_1 \dots O_n$  y sus respectivos estados.

```
[MatrizdeTransicion, MatrizdeEmision] = [estimate
(Observaciones, Estados);
```

Matriz de transición.

	0	0	0	1	0
0	0.11904762	0.0952381	0.21809524	0.35714286	0.19047619
0	0.08181818	0.13636364	0.31818182	0.13636364	0.22727273
0	0.08045977	0.03747126	0.2183908	0.14942529	0.49425287
0	0.2195122	0.09756098	0.16292683	0.45121951	0.04878049
0.0106383	0.07446809	0.0312796	0.38297872	0.13629787	0.37234043

Matriz de emisión.

0	0	1
0.47619048	0.45238093	0.07142857
0.95454545	0.04545455	0
0.71264368	0.10344828	0.18390805
0.75609756	0.09756098	0.14634146
1	0	0

Paso tres: Determinar la secuencia de observaciones inicial de tamaño  $N$ ,  $O_1, \dots, O_N$ ,

siendo  $O_N = O_1$

Secuencia Observaciones Inicial

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia Estados Inicial

4	3	1	4	3	2	4	3	4	3	3	3	3	3	2	4	4	4	4	4	4
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso cuatro: Determinar la secuencia de estados más probable

A partir de la rutina usada en Matlab (`hmmviterbi`) se calcula una secuencia de estados con la mayor probabilidad.

`EstadosGenerados = hmmviterbi (SecuenciaInicial, MatrizdeTransicion, MatrizdeEmision);`

Secuencia de estados más probable obtenida

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Teniendo esta secuencia de estados generada por la rutina se compara con la secuencia de estados de entrada,

`Porcentaje = sum (Estados==EstadosGenerados) /N EJEMPLOS;`

Porcentaje obtenido: 0.2500

Paso cinco: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados

Esta se obtiene usando la rutina en MATLAB `hmmdecode`

`[ProEstados, LogSecuencia] = hmmdecode (SecuenciaInicial, MatrizdeTransicion, MatrizdeEmision)`

Logaritmo de la probabilidad de la secuencia de estados Inicial: -12.5840

Paso seis: Determinar la secuencia de observaciones final de tamaño  $N+1$ ,  $O_1, \dots, O_{N+1}$

siendo  $O_{N+1} = O_1$  que incluye la observación que se desea determinar si es desbalance financiero

Secuencia Observaciones Final

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Para esta cadena de 21 observaciones se repite los Pasos Cuatro y Cinco para calcular la secuencia de estados más probable, usando  $\text{hmmViterbi}$  y el logaritmo de la probabilidad usando  $\text{hmmDecode}$ .

Logaritmo de la probabilidad de la secuencia de estados Final: -14.9128

**Paso siete:** Se calcula Delta definida como la diferencia de Logaritmo de la probabilidad de la secuencia de estados Final menos Logaritmo de la probabilidad de la secuencia de estados Inicial

Delta: -2.3268

También se calcula la Delta proporcional respecto a Logaritmo de la probabilidad de la secuencia de estados Inicial

Delta proporcional = 0.1851

Del paso tres al siete se realiza para diferentes tamaños de N. En la siguiente tabla se reporta Delta y la Delta proporcional para diferentes tamaños de la secuencia de observación inicial.

Tabla 3 Comportamiento de la Cuenta bancaria 2335 de la Base de Datos Berka [1], con balance negativo

Longitud	Delta	Delta Proporcional
20	1.4774	0.1412
17	1.4774	0.1726
14	1.4774	0.238
11	1.4773	0.2767
8	-1.477	0.3093
5	1.4739	0.5127

Se observa en la Tabla 2, que cuando no hay desbalance financiero el valor absoluto de Delta es menor que uno y el porcentaje de variación es menor a 10%. Mientras que la Tabla 3 muestra que cuando existe un desbalance financiero, el valor absoluto de la variable Delta es mayor que uno y el porcentaje de variación es mayor a 10%.

## 5 Conclusiones y trabajos a futuro

Los Modelos Ocultos de Markov nos permiten modelar la información de las transacciones bancarias de un usuario. Este modelo lo representan las matrices de Transición y Emisión. Se muestra con la base de datos Berka [1], la convergencia de la matriz de transición cuando el tamaño de la secuencia de estados aumenta. También el modelo HMM para una cuenta bancaria, permite determinar si existe un desbalance financiero, comparando dos secuencias de tamaño N y N+1. Cuando no hay desbalance financiero el valor absoluto de Delta es menor que uno y el porcentaje de variación es menor a 10%. Mientras que cuando existe un desbalance financiero, el valor absoluto de la variable Delta es mayor que uno y el porcentaje de variación es mayor a 10%.

Como trabajo futuro se desea aplicar este proceso a una base de datos de tarjetas de crédito para detectar fraude.

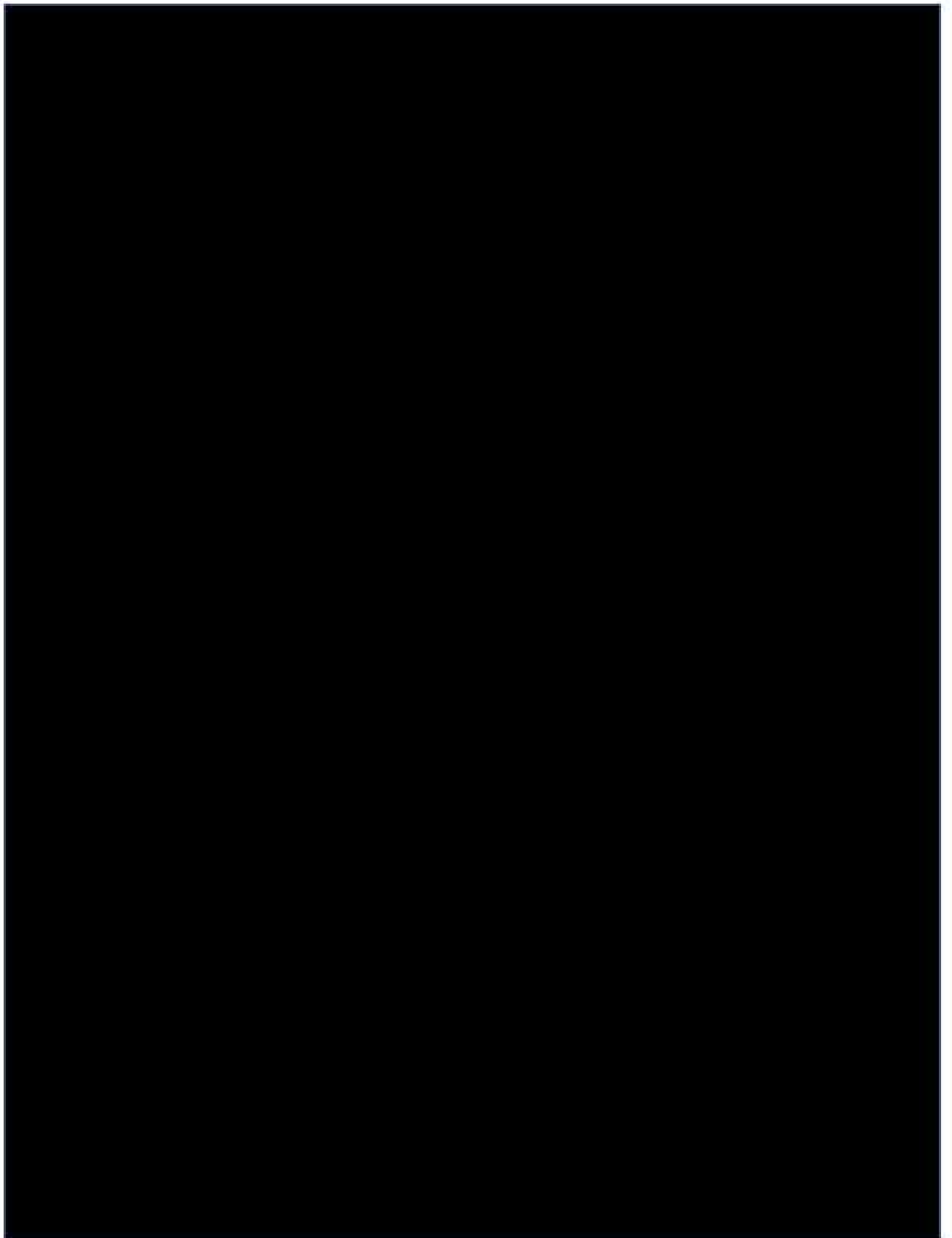
**Agradecimientos.** Este trabajo se realizó en la estancia de Prácticas profesionales en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla.

## Referencias

1. Base de datos Berka , recuperada de <http://isp.vot.co/>, fecha de acceso : marzo de 2019
2. Rabiner, Lawrence R. A tutorial on Hidden Markov Model and Selected Applications in Speech Recognition. *Proceeding of the IEEE*, Vol. 77 No. 2, pp. 257-286 (1989).
3. Pacheco Martínez, M. Redes Neuronales Recurrentes Aplicadas al Reconocimiento de Voz y Traducción Automática de Español a Japonés (Licenciatura). Benemérita Universidad Autónoma de Puebla, (2019).
4. Ángeles Millán, Y. Informe de Prácticas Profesionales, Modelos Ocultos de Markov, (2019).
5. Srivastava, A., Kanda, A., Sural, S., & Majumdar, A. Credit Card Fraud Detection Using Hidden Markov Model. *In Dependable and Secure Computing, IEEE Transactions on*, pp. 37-39 (2008).
6. Richard Durbin, Sean R. Eddy, Anders Krogh, Graeme Mitchison. Biological sequence analysis. The Pit. Building, Trumpington Street, Cambridge, United Kingdom. CAMBRIDGE UNIVERSITY PRESS. (1998)

*Aportes y Aplicaciones en las  
Ciencias Computacionales*  
se terminó de editar en Diciembre de 2019  
en la Facultad de Ciencias de la Computación Av. San Claudio y 14 Sur  
Jardines de San Manuel  
Ciudad Universitaria  
C.P. 72570





# Probabilidad, Estadística y sus Aplicaciones

**Editores:**

**Hugo Adán Cruz Suárez**

**Bulmaro Juárez Hernández**

**Francisco Solano Tajonar Sanabria**

**Hortensia Josefina Reyes Cervantes**

**Fernando Velasco Luna**

**Josè Dionicio Zacarías Flores**

**Víctor Hugo Vázquez Guevara (responsable)**



# Probabilidad, Estadística y sus Aplicaciones

# Probabilidad, Estadística y sus Aplicaciones

Editores:

Hugo Adán Cruz Suárez

Bulmaro Juárez Hernández

Francisco Solano Tajonar Sanabria

Hortensia Josefina Reyes Cervantes

Fernando Velasco Luna

José Dionicio Zacarías Flores

Víctor Hugo Vázquez Guevara (responsable)



Benemérita Universidad Autónoma de Puebla

2019

Primera edición: 2019

ISBN: 978-607-525-589-7

DR © Benemérita Universidad Autónoma de Puebla

4 sur 104, Col. Centro Histórico, Puebla, Pue. CP 72000

Teléfono: 01 (222) 229 55 00

[www.buap.mx](http://www.buap.mx)

Dirección General de Publicaciones

2 norte 1404, Col. Centro Histórico, Puebla, Pue. CP 72000

Teléfonos: 01 (222) 246 85 59 y 01 (222) 229 55 00 Ext. 5768 y 5764

[www.dgp.buap.mx](http://www.dgp.buap.mx)

[dgp@correo.buap.mx](mailto:dgp@correo.buap.mx)

Facultad de Ciencias Físico Matemáticas

Av. San Claudio y 18 sur, Colonia San Manuel, Puebla, Pue. CP 72570

Teléfonos: 01 (222) 229 55 00 Ext. 7552

[www.fcfm.buap.mx](http://www.fcfm.buap.mx)

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA • Rector: José Alfonso Esparza Ortiz

• Secretario General: José Jaime Vázquez López • Vicerrector de Extensión y Difusión de la Cultura: Fernando Santiesteban Llaguno • Director General de Publicaciones: Hugo Vargas Comsille • Director de la Facultad de Ciencias Físico Matemáticas: Martha Alicia Palomino Ovando

Impreso y hecho en México

Printed and made in Mexico

# Índice General

Capítulo 1 .....	1
Índice de la calidad de la Vivienda en México G. Olivera y O.V. Serrano	
Capítulo 2 .....	13
Uso del condón de los adolescentes según las características de la pareja sexual C. Menkes, I.A. Sosa y L. Núñez	
Capítulo 3 .....	25
Una caracterización de la satisfacción estudiantil mediante análisis de clases latentes R. Álvarez-Vaz y E. Vernazza	
Capítulo 4 .....	40
Modelación espacial de la plaga Sigatoka Negra (Mycosphaerella fijiensis) en cultivos de plátano del estado de Guerrero J.E. Solís, M. Guzmán, R. Reyes y D. Briones	
Capítulo 5.....	53
Aplicación del análisis de supervivencia en datos de pacientes sometidos a transplante renal J.A. Gil, B. Juárez y V.H. Vázquez	
Capítulo 6.....	64
Modelación espacial de la infestación del ácaro rojo De las palmas (Raoiella Indica) en el estado de Guerrero con un proceso espacial Poisson D.A. Ozuna, M. Guzmán, F. Godínez y R. Reyes	
Capítulo 7.....	76
Una aplicación de la metodología Biplot Logístico: Análisis de la sostenibilidad empresarial A. Urruticoechea y E. Vernazza	

Capítulo 8.....	92
Análisis de la deserción en las licenciaturas de la FCFM-BUAP mediante el modelo de riesgo proporcional semiparamétrico B.X. Muñoz, B. Juárez, L. Cervantes y H. Reyes	
Capítulo 9.....	108
Modelación estadística con imágenes Satelitales en Ciencias Ambientales A.A. Oroza, G. Linares, H. Reyes y M. L. Sandoval	
Capítulo 10.....	124
Riesgo de Crédito usando Redes Neuronales A. Herrera, H. Reyes, G. Linares y B. Juárez	
Capítulo 11.....	136
Embarazo adolescente, desigualdad social y salud sexual y reproductiva según condición de indigenismo en México I.A. Sosa e I. A. Quallenberg	
Capítulo 12.....	150
Uso de la distribución Bernoulli Multivariada en salud bucal R. Álvarez-Vaz y F. Massa	
Capítulo 13.....	166
Determinación de la distribución de Probabilidad de la demanda en un modelo de control de inventarios E. Hernández y R. Ilhuicatzí	
Capítulo 14.....	179
El Conteo en la Probabilidad F. Tajonar, E. Morales, F. Velasco, H. Cruz y J. Zacarías	
Capítulo 15.....	191
Optimalidad de políticas $(s,S)$ para un modelo de inventarios vía la teoría de los procesos de decisión de Markov R. Blancas, H. Cruz, F. Velasco y F. Tajonar	

# Capítulo 9

---

Modelación estadística con imágenes satelitales en Ciencias Ambientales

Ana Aleyda Oroza Hernández<sup>1</sup>, Gladys Linares Fleites<sup>2</sup>  
Hortensia Josefina Reyes Cervantes<sup>1</sup>, María de L. Sandoval Solís<sup>3</sup>

---

<sup>1</sup>Facultad de Ciencias Físico Matemáticas,

<sup>2</sup>Departamento de Investigación en Ciencias Agrícolas,

<sup>3</sup>Facultad de Ciencias de la Computación,

Benemérita Universidad Autónoma de Puebla,

Av. San Claudio y 18 Sur, Col. San Manuel,

C.P. 72570, Puebla, Puebla,

aleyda16188@hotmail.com, gladyslinares1@yahoo.com

hreyes@fcfm.buap.mx, malosala1753@gmail.com

**Resumen.** Actualmente la modelación medio ambiental puede considerarse un campo maduro con muchos artículos presentados en un amplio espectro de problemas ambientales y distintos estudios de caso que cubren diferentes modelos. El desarrollo de la teledetección o percepción remota ha incrementado aún más las posibilidades de la modelación estadística en este campo. En el presente capítulo se describe brevemente, a través de dos estudios de caso, la manera en que los modelos de regresión lineales de efectos- fijos y mixtos son construidos y aplicados a los problemas del medio ambiente. Un caso se refiere a la modelación de escenarios de las coberturas y usos de suelo a través de imágenes satelitales en el sureste de la Presa de Valsequillo, Puebla. El otro caso enfoca la modelación del secuestro de carbono por suelos forestales en la Región Terrestre Prioritaria 105, Puebla, para mostrar cómo integrar las herramientas estadísticas con otros enfoques y estrategias, como son las imágenes satelitales y los modelos mixtos, que combinados contribuyen a una mejor explicación del fenómeno que se estudia, debido a que se pueden incluir efectos fijos y efectos aleatorios. El software R es utilizado para la estimación, ajuste y selección de los modelos elaborados.

Palabras claves: Regresión Lineal y polinomial, Modelos mixtos, índices de vegetación.

**Abstract.** The development of remote sensing has further increased the possibilities of statistical modeling in this field. In this chapter we briefly describe, through two case studies, the way in which models are constructed and applied to environmental problems. One case refers to the modeling of land cover and land use scenarios through satellite images in the southeast of the Valsequillo Dam, Puebla. The other case focuses on the modeling of carbon sequestration by forest land in the Priority Land Area 105, Puebla, to show how to integrate statistical tools with other approaches and strategies, such as satellite images and mixed models, which combined contribute to a better explanation of the said phenomenon. The software R is used



for the estimation, adjustment and selection of the elaborated models.

*Keywords:* Linear and polynomial regression, mixed models, vegetation indexes.

## 9.1. Introducción

El uso de los modelos se ha incrementado en el manejo medioambiental, debido a que ellos son la única herramienta que permite relacionar cuantitativamente el impacto en un ecosistema con las consecuencias para el estado del ecosistema [13]. El campo de la modelación medio ambiental se ha desarrollado enormemente desde que emergieron los primeros modelos en 1970, no sólo debido al desarrollo de la tecnología computacional sino también al desarrollo de la teledetección o percepción remota.

En dependencia del nivel de complejidad, los modelos estadísticos de imágenes pueden tratar con modelos lineales o con modelos lineales generalizados de efectos mixtos o con modelos no lineales de efectos mixtos. El objetivo que se persigue en el presente capítulo es describir la manera en la que los esfuerzos de la modelación medioambiental han contribuido a esclarecer complejos problemas. Se presentan dos aplicaciones de la modelación medioambiental. La primera aplicación tiene como objetivo modelar tres Factores de cobertura de suelo, para conocer cuánto protege la vegetación la erosión del suelo en la presa de Valsequillo, Puebla. La segunda aplicación modela la cantidad de carbono orgánico para una zona terrestre prioritaria en Puebla.

## 9.2. Teledetección

La teledetección es una técnica que nos permite obtener información a distancia de objetos que se encuentran situados sobre la superficie terrestre. El fenómeno de la Teledetección es posible gracias a la interacción de la energía electromagnética con las cubiertas terrestres. Estas tienen un comportamiento reactivo variable, condicionado tanto por los factores externos (ambientales) como por sus propias características fisicoquímicas en el momento de la toma de la imagen.

La primera experiencia de teledetección se retoma en 1859, donde Gaspar Félix de Tournachon obtuvo las primeras fotografías aéreas desde un globo, después se hizo uso del avión hasta llegar en 1960 al uso de los satélites. Actualmente son numerosos los centros de producción, enseñanza e investigación que trabajan activamente en este campo con el uso de satélites (Teledetección espacial).

Para obtener información mediante teledetección espacial es importante que los objetos y el sensor tengan algún tipo de interacción. Para que la observación sea posible, se necesitan los siguientes elementos:

- Sensor: Instrumentos de grabación, instrumentos de escaneo, aviones, satélites, boyas o barcos.
- Objeto observado: arboles, suelos, etc.
- Flujo energético que permite poner a ambos en relación.

Dentro de la teledetección, se destacan las siguientes bandas espectrales, donde su unidad de medida es el Angstroms( $\mu m$ );

1. Espectro visible (**.4 a .7  $\mu m$** ): Es la radiación que puede percibir nuestros ojos, coincide con las longitudes de onda en donde es máxima la radiación solar.
2. Infrarrojo cercano (**.7 a 1.3  $\mu m$** ): Resulta importante por su capacidad para discriminar masas vegetales y concentraciones de humedad.
3. Infrarrojo medio (**1.3 a 8  $\mu m$** ): En esta banda se mezclan los procesos de reflexión de la luz solar y de emisión de la superficie terrestre.
4. Infrarrojo lejano o térmico (**8 a 14  $\mu m$** ).
5. Micro-ondas (**a partir de 1 metro**): Tipo de energía bastante transparente a la cubierta nubosa.

La constelación de satélites LANDSAT (LAND=tierra y SAT=satélite), que inicialmente se llamaron ERTS (por sus siglas en inglés, Earth Resources Technology Satellites), fue la primera misión de los Estados Unidos para el monitoreo de los recursos terrestres, el primer satélite se puso en órbita el 23 de julio de 1972.

### 9.2.1. Índices de Vegetación

Se han desarrollado muchos índices de vegetación basados en el hecho de que las plantas reflejan menos en luz roja visible, pero más en la radiación infrarroja cercana en comparación con la superficie sin vegetación [3] y [9].

Por lo tanto, los índices de vegetación pueden mejorar o extraer algunas características específicas que las bandas espectrales individuales no pueden. Los índices de vegetación que más se utilizan son: el índice de vegetación de diferencia normalizada (NDVI), el índice de vegetación ajustado al suelo (SAVI), el índice de vegetación de diferencia renormalizado (RDVI), el NDVI transformado (TNDVI), índice de vegetación simple (SVI) y la proporción simple (RVI).

El NDVI mide la cantidad de vegetación verde, tomando un cociente de las diferencias de reflectancia espectral entre infrarrojo cercano (NIR) y rojo (RED) para calcularlo, se ha utilizado ampliamente en estudios de teledetección cuyo rango de valores es de  $-1.0$  a  $1.0$ , donde los valores más altos son para vegetación verde y los valores bajos para otro tipo de superficie. Así, el suelo desnudo o rocas se representan con valores de NDVI más cercanos a 0 y los valores negativos corresponden principalmente a las nubes, el agua y la nieve.

La fórmula para calcular el NDVI queda de la siguiente forma:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}. \quad (9.1)$$

### 9.2.2. Índice de Vegetación Relativo (FVC)

El Índice de Vegetación Relativo (en inglés “Fractional Vegetation Cover”) es la estimación del porcentaje de vegetación [2]; se basa en el supuesto de que el NDVI

de la vegetación se distribuye gradualmente entre el NDVI del suelo desnudo y el NDVI de la vegetación totalmente verde. Para calcular el FVC se utiliza la ecuación:

$$FVC = \frac{NDVI - NDVI_{sd}}{(NDVI_{VV} - NDVI_{sd})^2} * 100. \quad (9.2)$$

donde  $NDVI_{sd}$  es el NDVI del suelo desnudo y  $NDVI_{VV}$  es el NDVI del pixel con la mayor cobertura de vegetación verde en el área.

### 9.3. Modelos estadísticos

El modelo de regresión para estos métodos puede escribirse como:

$$y = X\beta + \epsilon. \quad (9.3)$$

donde:

$y$  es un vector  $n \times 1$  de observaciones de la variable dependiente.

$X$  es una matriz  $n \times p$  que consiste de  $n$  observaciones de las  $p$  variables,

$\beta$  es un vector  $p \times 1$  de coeficientes de la regresión (parámetros desconocidos), y

$\epsilon$  es un vector  $n \times 1$  de errores independientes e idénticamente distribuidos con media cero y varianza  $\sigma^2$ .

El modelo lineal de efectos mixtos (LMM) es una poderosa herramienta de modelado que permite el análisis de conjuntos de datos correlacionados (clúster) [6]. El LMM fue presentado originalmente por Laird y Ware, donde presentaron diferentes aplicaciones en el estudio epidemiológico con datos longitudinales [15], la denominación de “efectos mixtos” se asigna para tener efectos tanto fijos como aleatorios. Actualmente se usan para modelar datos de muchos campos en diferentes áreas, como biología, bioestadística, ciencias sociales, salud y medicina [4, 6] y, particularmente, en el medio ambiente.

La formulación corresponde a la propuesta en el artículo clásico de Laird y Ware en [15], en forma matricial el modelo queda expresado como:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i \quad (9.4)$$

$$b_i \sim N_q(0, \psi)$$

$$\epsilon_i \sim N_{n_i}(0, \sigma^2 \Lambda_i)$$

donde

$y_i$  es el vector de respuesta  $n_i \times 1$  para observaciones en el grupo  $i$ -ésimo.

$X_i$  es la matriz del modelo  $n_i \times p$  para los efectos fijos y las observaciones en el grupo  $i$ .

$\beta$  es el vector  $p \times 1$  de los coeficientes de efectos fijos.

$Z_i$  es la matriz del modelo  $n_i \times q$  para los efectos aleatorios para las observaciones en el grupo  $i$ .

$b_i$  es el vector  $q \times 1$  de los coeficientes de efecto aleatorio para el grupo  $i$ .

$\epsilon_i$  es el vector de errores  $n_i \times 1$  para las observaciones en el grupo  $i$ .

$\psi$  es la matriz de covarianza  $q \times q$  para los efectos aleatorios.  
 $\sigma^2\Lambda_i$  es la matriz de covarianza  $n_i \times n_i$  para los errores en el grupo  $i$ .

Los errores residuales  $\varepsilon_i$  para el mismo grupo son independientes de los efectos aleatorios  $b_i$ .

## 9.4. Dos Casos de Estudio

### 9.4.1. Análisis de la cobertura edáfica en el sureste de la Presa Valsequillo, Puebla.

El uso y manejo no sostenible de la tierra está llevando a una mayor degradación del suelo y la pérdida de un recurso clave que es fundamental para la vida en el planeta [12]. En el Plan Estratégico Decenal 2008-2018 aprobado en 2007 por la 8va Conferencia de las Partes de la Convención de las Naciones Unidas de Lucha contra la Desertificación, se estableció el objetivo de proteger los suelos contra la erosión y la contaminación. En México, en el Programa Nacional Manejo Sustentable de Tierras a favor de un manejo y uso sostenible, publicado en 2008, se señaló que la erosión y el declive de la fertilidad del suelo afectan a la viabilidad de los terrenos agrícolas.

Desde hace décadas, en la región sur de la presa de Valsequillo, el manejo de los recursos naturales se ha llevado de manera inadecuada provocando la degradación del suelo y la pérdida de la vegetación. La escasa investigación en los diferentes ecosistemas, el empleo de las tecnologías de explotación sin tener en cuenta las condiciones del entorno y el desconocimiento de los principios físicos, químicos y biológicos que determinan la continuidad de los sistemas de producción, son los principales factores que han afectado gravemente la estabilidad de los sistemas naturales en la Región Sur de la Presa de Valsequillo. La Figura 9.1 muestra la localización de esta zona.

La zona de estudio corresponde a la parte norte del municipio de Tzicatlacoyan donde se sitúan las localidades San Miguel Acuexcomac, San Bernardino Tepenene, San José Texaluca y San Martín los Teteles (para más detalles ver [16]). Identificar mediante imágenes de satélite las afectaciones de la cobertura edáfica de esta zona es importante para la posterior toma de decisiones en el manejo de estos suelos y es el objetivo que se persigue en este estudio.

#### Factores de manejo de cobertura de suelo.

A continuación, se resumen algunos de los indicadores que se utilizan para estos estudios y que constituyen factores a tener en cuenta en el manejo de la cobertura del suelo.

- Los valores de este factor se calculan tomando el logaritmo del porcentaje de FVC usando la función de regresión

$$C\_FVC = 0.6508 - 0.343 \log(FVC). \quad (9.5)$$

- Es uno de los factores de la Ecuación Universal de Pérdida de Suelo, (USLE, por sus siglas en inglés). Este factor refleja el efecto de las prácticas de cultivo

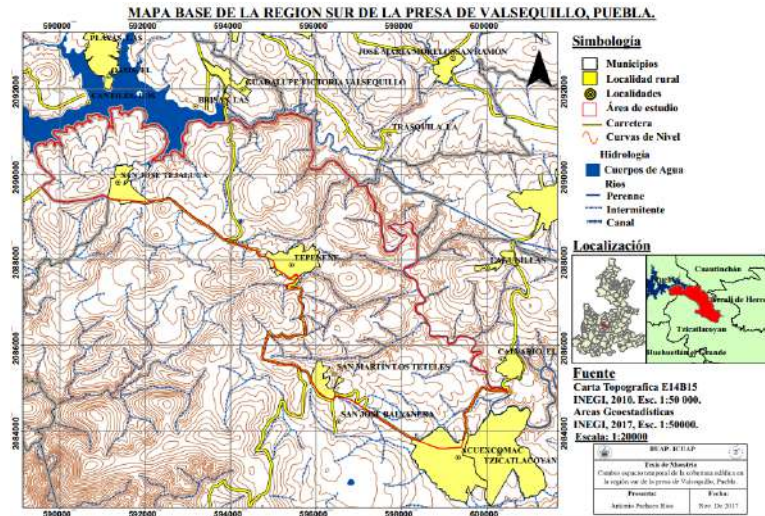


Figura 9.1: Localización de la zona de estudio. Fuente: Pacheco Ríos, A., 2018.

y manejo sobre las tasas de erosión. Sus valores indican cuánto protege la vegetación la erosión del suelo [19]. El factor  $C_{USLE}$  corresponde a la pérdida de suelo bajo condiciones específicas de cultivo, en relación con la que ocurre en un suelo desnudo. Toma valores de 0 (cobertura vegetal alta) a 1 (suelo desnudo). Esto puede ser representado como

$$C_{USLE} = \frac{A_{crop}}{A_{fallow}} * 100. \quad (9.6)$$

Este factor refleja el efecto de las prácticas de cultivo y manejo sobre las tasas de erosión. Los valores de  $C_{USLE}$  indican cuánto protege de la erosión la vegetación del suelo.

- El factor  $C_{exp}$  puede estimarse aplicando la relación utilizada por Van der Knijff en [14]:

$$C_{exp} = \exp\left(-\alpha * \frac{NDVI}{(\beta - NDVI)}\right). \quad (9.7)$$

Se considera como un factor de gestión de cobertura de cultivo calculado. En [14] se sugiere que al aplicar esta relación, se obtienen mejores resultados que usando una relación lineal.

### Obtención de la imagen satelital y análisis exploratorio de la tabla de datos

La imagen de satélite fue obtenida del Servicio Geológico de los Estados Unidos (USGS, por sus siglas en inglés) y corresponde al año 2015. Dicha imagen se obtuvo sin cobertura de nubes en la temporada de sequía, lo que permitió reducir el efecto espectral de la vegetación y de los cultivos en el procesamiento de la imagen. El

análisis e integración de la información se realizó con el Sistema de Información Geográfica de uso libre Quantum GIS[18]. Con la información obtenida se elaboró una tabla de 34552 filas (puntos georeferenciados) y 5 columnas con las variables cuantitativas NDVI, FVC, C\_FVC, C\_USLE y C\_exp, realizándose el análisis exploratorio de los datos. Se detectaron 5 puntos aberrantes o “outliers” que fueron eliminados, dado que el análisis gráfico realizado y la gran cantidad de datos disponibles indicaron que no se afectaría la modelación estadística posterior. La matriz de correlaciones, que se muestra en la Tabla 9.1, indica que las variables NDVI, FVC, C\_FVC y C\_exp se encuentra altamente correlacionadas entre ellas, por lo cual es importante considerar su relación para la construcción de modelos.

Tabla 9.1: Correlación entre las variables.

	FVC	C_FVC	C_USLE	C_exp
NDVI	0.98	-0.97	-0.32	-0.96
FVC		-0.93	-0.32	-0.90
C_FVC			0.30	0.99
C_USLE				0.29
C_exp				

### Modelos de regresión

Se ajustaron tres modelos lineales, uno para cada variable FVC, C\_FVC y C\_exp, tomadas como variables dependientes y como variable independiente el NDVI. Sus estimaciones fueron ajustadas usando la función `lm()` del software R. Las estimaciones de intercepto y pendientes de cada modelo se muestran en la Tabla 9.2.

Tabla 9.2: Estimaciones de los parámetros en cada modelo lineal.

MODELO		Estimate	Pr(> t )
Modelo 1	(Intercept)	-20.2621	0.0000
FVC	NDVI	95.7324	0.0000
Modelo 2	(Intercept)	0.5324	0.0000
C_FVC	NDVI	-0.6875	0.0000
Modelo 3	(Intercept)	0.7403	0.0000
C_exp	NDVI	-1.1064	0.0000

A pesar de que las estimaciones para cada modelo, tanto del intercepto como de la pendiente, son significativas, una parte importante dentro de la modelación que no debe ser olvidado, es la validez de los supuestos del modelo, para ello hicimos uso de los gráficos de residuales que proporciona el comando `plot()` en el software R. Los gráficos de valores ajustados y residuales mostraron un comportamiento particular que sugirieron el uso de otro tipo de modelo, por ejemplo, un modelo polinomial para ajustar los datos.

Las estimaciones de cada modelo usando modelos polinomiales se muestran en la Tabla 9.3. Estas estimaciones son de igual forma significativas, con excepción de la pendiente del modelo 1 no lineal, por lo cual sólo se realiza el ajuste del modelo sin intercepto para el primer modelo polinomial. El ajuste y las predicciones dentro

de cada modelo se muestran en los gráficos correspondientes en las Figuras 9.5, 9.6 y 9.7.

Las estimaciones de los parámetros de los modelos no lineales son significativas y también poseen un buen ajuste ya que el coeficiente de determinación  $R^2$  es igual 0.98.

La Figuras de los residuos que se muestran en 9.2, 9.3 y 9.4 son usados para el ajuste de modelos lineales, para verificar los supuestos de linealidad, normalidad y homogeneidad de varianzas, que no todos se consideran dentro de los modelos no lineales. Un modelo lineal adecuado muestra gráficas de residuos con valores ajustados distribuidos de manera homogénea sobre un intervalo, mientras que el supuesto de normalidad y linealidad se verifica con el gráfico de residuales estandarizados y cuantiles teóricos, en caso de cumplir el supuesto de normalidad los datos estos deben estar muy cercanos a la recta de dicho gráfico. Finalmente, estas ecuaciones que se estimaron se puede calcular la cantidad de cobertura vegetal, en función del índice de vegetación NDVI que es obtenido por medio de la imagen de satélite considerada.

Tabla 9.3: Estimaciones de los parámetros en cada modelo no lineal.

MODELOS		Estimate	Pr(> t )
Modelo 1. FVC	(Intercept)	0.0000	0.5800
	NDVI	-0.0000	0.2917
	I(NDVI^2)	100.0000	0.0000
Modelo 2. C_FVC	(Intercept)	0.8272616	0.0000
	NDVI	-2.4437965	0.0000
	I(NDVI^2)	3.0196644	0.0000
	I(NDVI^3)	-1.5291057	0.0000
Modelo 3. C_exp	(Intercept)	1.1108	0.0000
	NDVI	-2.8568	0.0000
	I(NDVI^2)	1.8285	0.0000

Finalmente los modelos estimados quedan de la siguiente forma:

$$FCV = 100 * NDVI^2 + \varepsilon \tag{9.8}$$

$$C\_FCV = .82 - 2.44 * NDVI + 3.01 * NDVI^2 - 1.52 * NDVI^3 + \varepsilon \tag{9.9}$$

$$C\_exp = 1.1 - 2.8 * NDVI + 1.8 * NDVI^2 + \varepsilon \tag{9.10}$$

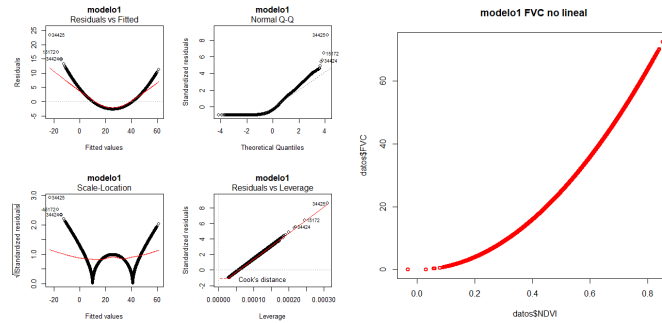


Figura 9.2: Gráfico del modelo ajustado: Modelo 1.

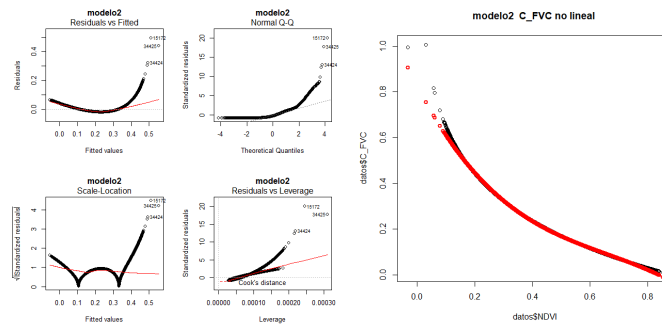


Figura 9.3: Gráfico del modelo ajustado: Modelo 2.

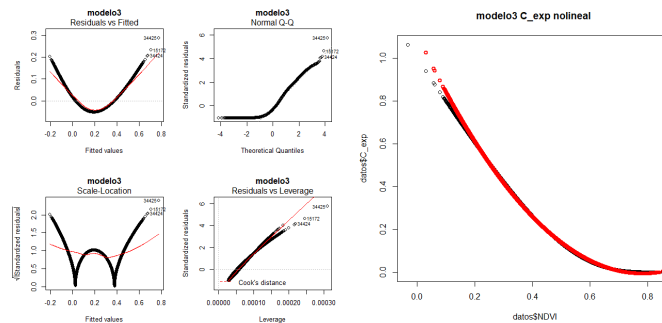


Figura 9.4: Gráfico del modelo ajustado: Modelo 3.

### 9.4.2. Secuestro de carbono en la RTP 105

Se presenta la localización de la Región Terrestre Prioritaria 105 (RPT 105) que se ubica en las coordenadas extremas:  $19^{\circ}46'23''$  y  $20^{\circ}11'55''$  de latitud norte y  $97^{\circ}09'17''$  a  $97^{\circ}38'36''$  de longitud oeste y está conformada por 28 municipios de los que 4 pertenecen al estado de Veracruz y 24 en la Sierra Norte de Puebla. La Figura 9.6 muestra en color rojo la localización de la zona de estudio.





Figura 9.5: Localización de la Región Terrestre Prioritaria 105.

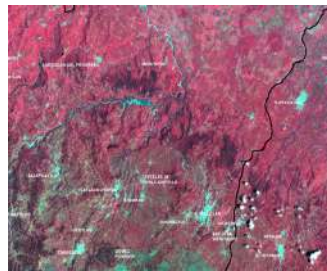


Figura 9.6: Imagen satelital de la zona de estudio.

Para el estudio del secuestro de carbono en suelos en la RTP-105, se buscaron imágenes con sensor ETM+ y TM. Las muestras de suelo (propiedades físico y químicas del suelo) habían sido tomadas en el año 2005, por lo que era de interés tomar imágenes de ese año y el satélite que se encontraba en órbita correspondía al satélite Landsat 7. Se descargaron las imágenes del satélite Landsat 7 (2 de noviembre de 2005), que se obtuvo a través de USGS (Earth Resource Observation Systems Data Center). Se consideraron también las condiciones atmosféricas para que la información no se viera afectada por las condiciones climáticas.

Ya obtenidas las imágenes de satélite que corresponden a la banda 3 y 4 del satélite Landsat 7, se hizo uso del software GIS (Geographic Information System), para poder obtener el índice de vegetación NDVI.

Se utilizó la metodología de Zuur que está enfocada a obtener el mejor modelo mixto, basado en los criterios de selección de modelos AIC y BIC. El procedimiento fue el siguiente:

**PASO 1** Ajustar un modelo donde la componente fija contenga todas las variables explicativas e interacciones posibles. Dentro de las variables de la zona RTP-105, se realizó una selección de variables que aportan más información al modelo. Los distintos modelos de regresión se realizaron usando la función

lm() del software R, llegando a que el mejor modelo de efectos fijos es el que se muestra en la tabla 9.4.

Tabla 9.4: Estimaciones del modelo `modelo1.lm`.

	Estimate	Std. Error	t value	Pr(>  t )
Ntot	3.9109	0.3168	12.35	1.25e-11 ***

Tomando como base el modelo anterior, se introduce la variable NDVI mostrando sus estimaciones en la Tabla 9.5.

Tabla 9.5: Estimaciones del modelo `modelo2.lm`.

	Estimate	Std. Error	t value	Pr(>  t )
Ntot	9.545	1.950	4.894	6.81e-05 ***
NDVI	5.382	2.204	2.442	0.0231 *

**PASO 2** Teniendo como base el mejor modelo con las componentes fijas, ahora continuamos con el segundo paso de la metodología de Zuur, sin embargo, en este estudio sólo se requiere ver las variaciones en F\_V (formación vegetal), como variable aleatoria que queremos introducir dentro del modelo, por lo cual el modelo en este paso de la metodología de de Zuur, para el caso de la Zona RTP-105 queda como: parte fija las variables del modelo `modelo2.lm` y parte aleatoria F\_V.

**PASO 3** Se busca la estructura óptima de la componente fija del modelo. Para ello podemos utilizar el estadístico F o el estadístico t obtenido mediante el estimador REML con la función `lme()` o comparar modelos anidados. Para comparar modelos que tienen la misma estructura en la componente aleatoria, pero difieren en la componente fija se debe de utilizar un estimador LM y no un estimador de REML.

Se construyen distintos modelos, donde se varía la componente aleatoria, introduciéndola como una intercepto, pendiente, que afecta a los distintos modelos, para buscar la mejor estructura de modelo con componente fija y componente aleatoria.

Para realizar los distintos modelos mixtos, en este trabajo se usan la librería `library(lme4)`, `library(Matrix)`, `library(Rcpp)` y `library(nlme)`. Estas librerías, contienen distintas funciones que nos ayudan a poder estimar los efectos de las variables fijas y aleatorias en modelos mixtos. Los siguientes modelos mixtos contienen la misma componente aleatoria, pero buscando la mejor componente fija. Para seleccionar el mejor modelo en esta parte, usamos los criterios de selección de modelos AIC y BIC.

En la Tabla 9.6, se observan los criterios de selección de modelos para seleccionar el mejor modelo. El modelo `mixed.model1` es seleccionado por tener los valores más pequeños de los criterios, con un valor de  $AIC = 38.34$  y  $BIC = 42.70463$ .

Tabla 9.6: Criterios de selección de modelos.

	df	AIC	BIC
mixed.model1	4.00	38.34	42.70463
mixed.model2	3.00	50.49	53.89353
mixed.model3	2.00	55.51	57.86206

**PASO 4** Finalmente el último paso consiste en presentar el modelo final utilizando un estimador REML y analizar las suposiciones establecidas en el mismo.

Modelo lineal Mixto seleccionado

```
mixed.model1 <- lme(log(Corg) ~ -1 + Ntot + NDVI, random = ~1 | F_V)
```

Como parte final de la metodología de Zuur se muestra la salida completa de las estimaciones del modelo, usando el metodo REML.

```
Linear mixed model fit by REML t-tests use Satterthwaite
approximations to degrees of
  freedom [lmerMod]
Formula: log(Corg) ~ -1 + Ntot + NDVI + (1 | F_V)
Random effects:
  Groups   Name                Variance Std.Dev.
  F_V      (Intercept)  0.1654   0.4067
  Residual                    0.1528   0.3909
Number of obs: 24, groups: F_V, 6
```

```
Fixed effects:
      Estimate Std. Error    df t value Pr(>|t|)
Ntot   1.9082     0.4780 21.7410   3.992 0.000627 ***
NDVI   2.2151     0.5194 21.9890   4.264 0.000317 ***
```

```
Correlation of Fixed Effects:
      Ntot
NDVI  -0.569
```

Con los siguientes comandos en R, realizamos los gráficos, para analizar las suposiciones del modelo mixed.model1.

```
Res <- residuals(mixed.model1, type="normalized")
Fit <- fitted(mixed.model1)
par(mfrow=c(2,2))
plot(Res ~ Fit, xlab="Fitted values", ylab="Residuals",
     main="Residuals vs. fitted")
abline(h=0)
plot(Res ~ datos$F_V, xlab="FORMACI\ 'ON VEGETAL", ylab="Residuals",
     main = "FORMACI\ 'ON VEGETAL")
abline(h=0)
```

```
hist(Res, main="Histogram of residuals", xlab="Residuals")
qqnorm(Res)
qqline(Res)
```

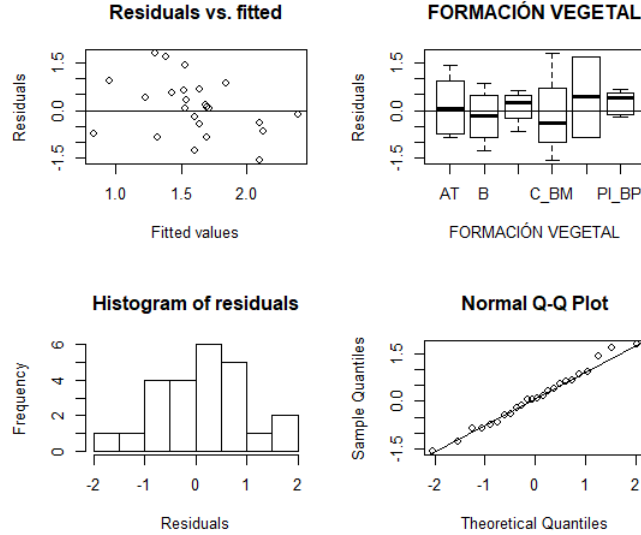


Figura 9.7: Suposiciones del modelo mixed.model1.

La forma del modelo mixto ajustado, se muestra a continuación:

$$\log(Corg_{ij}) = \beta_1 * Ntot_{ij} + \beta_2 * NDVI_{ij} + a_j + \epsilon_{ij}. \quad (9.11)$$

El índice  $j$  (representa a las distintas Formaciones vegetales  $F.V$ ) toma valores de 1 a 6, e  $i$  (representa las muestras dentro de cada formación vegetal).

La parte de los resultados que se refiere a los efectos aleatorios nos muestra que la varianza residual es  $\sigma^2 = .39^2$  y la varianza de la constante  $\sigma_a^2 = .40^2$ . Para la parte de los efectos fijos del modelo,  $\beta_1 * Ntot + \beta_2 * NDVI_{ij}$ , la constante se estima con el valor  $\beta_1 = 1.90$  y la segunda en  $\beta_2 = 2.21$ . Ambos parámetros son significativamente distintos de 0. Un observación importante es sobre la correlación entre las constantes es de  $-.50$ .

El Ntot y NDVI tienen efecto significativo sobre el logaritmo del carbono orgánico, observando en la salida del software R que hay en promedio más cantidad de Carbono orgánico por el índice de vegetación, mientras que el Ntot hay ligeramente menos. El modelo es bastante adecuado y se cumplen los supuestos (Figura 9.7).

## 9.5. Conclusiones

En los dos casos de estudios, donde el NDVI que se obtiene por medio de imágenes de satélite resultó ser significativo en ambos modelos, es decir permite predecir

los indicadores de suelo y/o la cantidad de carbono orgánico dependiendo el caso de estudio. En el primer caso se determinaron los modelos 9.8, 9.9, 9.10 de regresión polinómicos que mejoraron ajuste de los modelos, debido a que los datos tienen comportamientos curvilíneos. En el segundo caso los modelos lineales mixtos resultaron ser mejores para ajustar y estimar la cantidad de carbono orgánico dentro de la zona RTP-105, el modelo final de este caso corresponde a la fórmula 9.11.

Existe reconocimiento internacional del papel de la modelación estadística en la investigación medioambiental, la cuantificación de la incertidumbre presente en los problemas ambientales es uno de los retos actuales de la Ciencia Estadística, que junto al desarrollo de la Teledetección y la Ciencia de la Computación, se enfocan a brindar mejores explicaciones a tan complejos problemas, no viéndolos como un fin meramente técnico, sino como un medio necesario para mejorar las condiciones de vida de millones de personas.

# Bibliografía

- [1] ARRIAGA, L., J.M. ESPINOZA, C. AGUILAR, E. MARTÍNEZ, L. GÓMEZ y E. LOA (coordinadores). *Regiones terrestres prioritarias de México. Escala de trabajo 1:1 000 000. Comisión Nacional para el Conocimiento y uso de la Biodiversidad*. México, 2000.
- [2] BOETTINGER, J. L., RAMSEY, R.D. , BODILY, J.M., COLE, N.J., KIENAST-BROWN, S., NIELD, S.J., SAUNDERS, A.M., and STUM, A.K. *Landsat Spectral Data for Digital Soil Mapping. In Digital Soil Mapping with Limited Data*, A.E. 2008.
- [3] A. BANNARI and D. MORIN F. BONN A. R. HUETE . *A review of vegetation indices*, Remote Sensing Reviews. Vol.(13), p. (95-120), Taylor & Francis, 1995.
- [4] DEMIDENKO, E.. *Mixed models. Theory and applications with R*. 2nd edition. Wiley, New York, USA, 2013
- [5] DÍAZ G. y CERVIGÓN, J. J. *Estudio de índices de vegetación a partir de imágenes aéreas tomadas desde UAS/RPAS y aplicaciones de estos a la agricultura de precisión* , 2015.
- [6] GALECKI A. and BURZYKOWSKI T. *Linear mixed-effects models using R. A step by step approach*. 1st edition. Springer, New York, p.(542), 2013.
- [7] IENO N. E., SAVELIEV A. A., SMITH M. G., WALKER J. N. and ZUUR F. A *Mixed Effects Models and Extensions in Ecology with R*. Springer, 2009.
- [8] *Instituto Nacional de Estadística y Geografía (INEGI)*, Marco Geoestadístico Nacional, México. Disponible. <http://www.beta.inegi.org.mx/app/biblioteca/ficha>, 2017.
- [9] JENSEN, J.R. *Remote Sensing of the Environment: An Earth Resource Perspective*. Pearson Prentice Hall, New Jersey, 2000.
- [10] JENSEN, J.R. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Pearson Prentice Hall, New Jersey, 2005.
- [11] JOHN, R., JENSEN y DR. KALMESH LULLA, *Introductory digital image processing: A remote sensing perspective, Geocarto International*. Vol.(2), p.(65-65), 1987.

- [12] JONES, A. P., PANAGOS, S., BARCELO, F., BOURAOUI, C., BOSCO, O. DEWITTE, C., GARDI, *The State of Soil in Europe*. Joint Research Centre Reference Report. Publications Office of the European Union Luxembourg, p.(45), 2012.
- [13] JORGENSEN, S.E. . *Environmental Models and Simulations. In Environmental Systems*. [Ed. Achim Sydow], in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, <http://www.eolss.net>, 2004.
- [14] VAN DER KNIJFF J. M. , JONES R. J. A., MONTANARELLA L., *Soil Erosion Risk Assessment in Europe*, 2000.
- [15] NAN M. LAIRD and JAMES H. WARE, *Random-Effects Models for Longitudinal Data, Biometrics* , p.(963–974), Wiley, International Biometric Society, Vol.(38), 1982.
- [16] PACHECO R. A. . *Cambio espacio temporal de la cobertura edáfica en el sur-este de la Presa Valsequillo, Puebla. Causas y efectos*. Tesis de Maestría en Ciencias Ambientales. Posgrado en Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla, 2018.
- [17] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>, 2017.
- [18] Quantum GIS Development Team., *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>, 2017.
- [19] WISCHMEIER W. H., SMITH D. D.. *Predicting rainfall erosion losses R*, USDA Agricultural Handbook 537, 1978.

# MODELOS CUANTITATIVOS APLICADOS CON IMPACTO EN LA SOCIEDAD

Tomo 1

Carlos N. Bouza Herrera  
María del Mar Rueda García  
Agustín Santiago Moreno  
José Félix García Rodríguez





# Modelos Cuantitativos aplicados con impacto en la Sociedad

## Tomo 1

**Serie Métodos cuantitativos aplicados**

Editores:

Carlos N. Bouza Herrera

María del Mar Rueda García

Agustín Santiago Moreno

José Félix García Rodríguez

Primera edición. Marzo del 2020.

© Universidad de la Habana, Cuba.  
Universidad de Granada, España  
Universidad Autónoma de Guerrero, México  
Universidad Juárez Autónoma de Tabasco, México

**ISBN: 978-607-9440-91-6**

Editorial universitaria de la UAGro

El contenido de este libro es responsabilidad de los autores.  
La presentación y disposición en conjunto de este libro son propiedad de los editores.  
Ninguna parte de esta obra puede ser reproducida o transmitida, mediante ningún sistema o método, electrónico o mecánico, sin consentimiento de los autores.

Comité Científico:

**Marco Negreiros**, Universidad Estatal, Fortaleza, Brazil.

**Pedro Monterrey**, Universidad del Rosario, Colombia.

**Amer Ibrahim Al-Omari** / Ph.D. of Statistics Vice Dean of the Academic Research  
Department of Mathematics, Faculty of Science Al al-Bayt University, P.O. Box 130095,  
Mafrq 25113, Jordan. Mobile: 00962777906433

**Marie Cottrell**, Professeur émérite - Université Paris1. Tel et fax(prof): 33 1 44 07 89 22  
SAMM, Université Paris 1, 90, rue de Tolbiac-75634, PARIS CEDEX 13-FRANCE. E-  
mail: cottrell@univ-paris1.fr

**Jesús E. Sánchez García**, departamento de física aplicada de ICIMAF, la Havana,  
Cuba.

**Minerva Montero Díaz**, departamento de física aplicada de ICIMAF, la Havana, Cuba.

**Diseño de portada: Rosario Flores Sosa**

**Habana, Cuba, 5 marzo de 2020.**

## Índice de capítulos

CAPITULO	TÍTULO	AUTORES	PÁGINAS
1	PATRONES DE DISTRIBUCIÓN ECONÓMICA E INNOVACIÓN SOCIAL EN INICIATIVAS DE LAS VÍCTIMAS DE LA PALMA CUNDINAMARCA	Jenny Cubillos Díaz, Mercedes Gaitán Angulo, Jenny Paola Lis Gutiérrez	01-25
2	COSTOS Y RENTABILIDAD DE LA PRODUCCIÓN DEL LIMÓN PERSEA EN EL MUNICIPIO DE MISANTLA, VERACRUZ, MÉXICO	Ignacio Caamal Cauich, Verna Gricel Pat Fernández, José Félix García Rodríguez, Felipe Jerónimo Ascencio	27-48
3	DETERMINING PROBABILITY RECOMMENDATION OF HOTEL CUSTOMERS IN SIX LATIN AMERICAN CITIES	Carolina Heno, Andreina Moros, Yovanni Castro, Jenny Paola Lis-Gutiérrez, Manuel Ignacio Balaguera	49-60
4	SOBRE LA NECESIDAD DE IMPLEMENTAR INNOVACIÓN EN CURSOS INTRODUCTORIOS DE ESTADÍSTICA: UN ESTUDIO EXPLORATORIO EN LA EDUCACIÓN SUPERIOR MEXICANA	Cecilia Cruz López y Mario Miguel Ojeda Ramírez	61-91
5	INNOVACIÓN Y CAPITAL SOCIAL. ESTUDIO DE OPINIÓN UNIVERSITARIA EN EL SURESTE DE MÉXICO	José Félix García Rodríguez, Reyna Alejandra López Arias, Germán Martínez Prats, Ignacio Caamal Cauich, Aída Armenta Ramírez	93-102

6	RELACIÓN ENTRE EL DESEMPEÑO DOCENTE Y LA SATISFACCIÓN LABORAL DE PROFESORES DE EDUCACIÓN BÁSICA DE LA ZONA ESCOLAR 60, REGIÓN ACAPULCO-COYUCA	Tonantzin Martínez Bello, Elizabeth Ávila Cortes, Virginia García Cervantes y William de Jesús Merlán Avilés.	103-124
7	MODELOS OCULTOS DE MARKOV: APLICACIONES EN EL AMBIENTE Y EN LAS FINANZAS	María de Lourdes Sandoval Solís, Gladys Linares Fleites, Karla Mildred Cigarroa Alonso, Juan Badillo Iglecias	125-146
8	UNA REVISIÓN DE RECIENTES ENFOQUES EN LA IDENTIFICACIÓN DE CLÚSTERES EN EL ANÁLISIS DE GRANDES MASAS DE DATOS	Carlos N. Bouza Herrera, Agustín Santiago Moreno, Sira M. Allende Alonso, José M. Sautto-Vallejo	147-163

<b>Nombre</b>	<b>Afiliación</b>
<b>Agustín Santiago Moreno</b>	Facultad de Matemáticas, Universidad Autónoma de Guerrero, México.
<b>Aída Armenta Ramírez</b>	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
<b>Andreina Moros</b>	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
<b>Carlos N. Bouza Herrera</b>	Facultad de Matemática y computación, Universidad de la Habana
<b>Carolina Henao</b>	Corporación Universitaria Minuto de Dios (Bogotá, Colombia)
<b>Cecilia Cruz López</b>	Universidad Veracruzana
<b>Elizabeth Ávila Cortes</b>	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
<b>Felipe Jerónimo Ascencio</b>	Universidad Autónoma Chapingo (UACH).
<b>Germán Martínez Prats</b>	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
<b>Gladys Linares Fleites</b>	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México
<b>Ignacio Caamal Cauch</b>	Universidad Autónoma de Chapingo (UACH). Escuela de Economía
<b>Jenny Paola Lis Gutiérrez</b>	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
<b>José Félix García Rodríguez</b>	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
<b>José M. Sautto Vallejo</b>	Facultad de Matemáticas, Universidad Autónoma de Guerrero
<b>Juan Badillo Iglecias</b>	Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México
<b>Karla Mildred Cigarroa Alonso</b>	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México
<b>Manuel Ignacio Balaguera</b>	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
<b>María de Lourdes Sandoval Solís</b>	Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México
<b>Mario Miguel Ojeda Ramírez</b>	Universidad Veracruzana

<b>Reyna Alejandra López Arias</b>	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
<b>Sira M. Allende Alonso</b>	Facultad de Matemática y computación, Universidad de la Habana
<b>Tonantzin Martínez Bello</b>	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
<b>Verna Gricel Pat Fernández</b>	Universidad Autónoma Chapingo (UACH).
<b>Virginia García Cervantes</b>	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
<b>William Merlán Avilés</b>	Facultad de matemáticas de la Universidad Autónoma de Guerrero, México.
<b>Yovanni Castro</b>	Pontificia Universidad Javeriana (Bogotá, Colombia)

## PRÓLOGO

**El libro modelos cuantitativos aplicados con impacto en la sociedad**, está compuesto de ocho trabajos de investigación elaborados por investigadores de los centros de investigación de México, Cuba y Colombia que colaboran en la Red Iberoamericana de Estudios Cuantitativos Aplicados (RIDECA). La presente obra da cuenta de la importancia de las investigaciones basadas en métodos cuantitativos para la explicación de situaciones problema e identificación de los métodos y modelos apropiados que contribuyen en su resolución y que finalmente permiten explicar la complejidad, las estrategias de mejora, y la predicción sobre los fenómenos de estudio.

De manera específica, en la obra se plantean investigaciones sobre patrones de distribución económica e innovación social en iniciativas de víctimas, costos y rentabilidad de la producción del limón persa, determinación de la probabilidad de que un cliente recomiende un hotel en seis ciudades latinoamericanas, la necesidad de implementar innovación en cursos introductorios de estadística: un estudio exploratorio en la educación superior, innovación y capital social: estudio de opinión universitaria, relación entre el desempleo docente y la satisfacción laboral de profesores de educación básica, modelos ocultos de Markov: aplicaciones en el ambiente y en las finanzas, una revisión de recientes enfoques en la identificación de clústeres en el análisis de grandes masas de datos. Estos trabajos dan cuenta de las intervenciones en lo local, nacional e internacional. En RIDECA se ha desarrollado un escenario propicio para la contribución científica, la divulgación y socialización de los saberes.

Dr. Armando Morales Carballo  
Director de la Facultad de Matemáticas de la  
Universidad Autónoma de Guerrero, México.

Marzo de 2020





# CAPÍTULO 7

## MODELOS OCULTOS DE MARKOV: APLICACIONES EN EL AMBIENTE Y EN LAS FINANZAS

María de Lourdes Sandoval Solís<sup>1</sup>, Gladys Linares Fleites<sup>2</sup>, Karla Mildred Cigarroa  
Alonso<sup>3</sup>, Juan Badillo Iglecias<sup>4</sup>

<sup>1</sup>Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México. <sup>2</sup>Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México. <sup>3</sup>Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México. <sup>4</sup>Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México.

### RESUMEN.

A través de dos aplicaciones de los Modelos Ocultos de Markov se presenta la potencialidad de estos modelos para resolver problemas reales en dos diferentes ámbitos: cambio de uso de suelo y detección de desbalance financiero. En este trabajo, se presenta la distribución de los procesos de cambio cartografiados en la Barra San José, Chiapas, México, observándose que el proceso de cambio que predomina en el periodo de análisis en toda la zona de estudio es el incremento de los asentamientos humanos y el avance de la frontera agrícola, mientras que la deforestación de selvas, bosques y manglares se presenta en forma más acentuada. Por otra parte, ejemplificándose con la base de datos Berka, se presenta la detección de desbalance financiero a través de Modelos Ocultos de Markov, demostrándose numéricamente la convergencia de la matriz de transición y la relación del número de observaciones con la detección del desbalance financiero.

**PALABRAS CLAVE:** Cadenas de Márkov, Modelos Ocultos de Márkov, Cambio de Uso de Suelo, Detección de Desbalance Financiero.

### ABSTRACT

Through two applications of the Hidden Markov Models, the potential of these models to solve real problems in two different areas is presented: change of land use and detection of financial imbalance. In this work, the distribution of the change processes mapped in Barra San José, Chiapas, Mexico is presented, observing that the process of change that predominates in the period of analysis throughout the study area is the increase in human settlements and the advance of

the agricultural frontier, while the deforestation of forests, forests and mangroves is more pronounced. On the other hand, exemplifying with the Berka database, the detection of financial imbalance is presented through hidden Markov models, demonstrating numerically the convergence of the transition matrix and the relationship of the number of observations with the detection of financial imbalance

**KEY WORDS:** Markov Chain, Hidden Markov Model, land use change analysis, financial imbalance detection.

## 1. INTRODUCCIÓN

Un Modelo Oculto de Markov (HMM, por sus siglas en inglés) es un proceso doblemente estocástico con un proceso estocástico subyacente que no es directamente observable (está "oculto") pero se puede observar solo a través de otro proceso estocástico que produce la secuencia de observaciones. Desde 1989, que Rabiner [7] presenta los HMM para el reconocimiento de voz, se han aplicado en diversos problemas. Aquí se presentan dos aplicaciones concretas: una es el cambio de uso de suelo en manglares en la zona de San José Chiapa, México y, otra, la detección del desbalance financiero.

Desde la década de los años 90 del pasado siglo, muchos investigadores han usado modelos de Markov o matrices de probabilidad de transición para comprender características de la dinámica territorial, como la diversidad, los controles de cambio, o la dependencia de escala. Los HMM son técnicas de modelado secuencial de datos, que actualmente se consideran una herramienta casi indispensable en el análisis del cambio de uso del suelo [3] y [9]. Por otra parte, en el campo de las finanzas, los HMM también permite modelar la información de las transacciones bancarias de un usuario

A continuación, se presenta, en los epígrafes 2 y 3, una breve exposición de los HMM y sus algoritmos computacionales. Posteriormente, en el epígrafe 4, se desarrolla la aplicación del cambio de uso de suelo en manglares en San José Chiapas y, en el epígrafe 5, se expone la aplicación a la detección del desbalance financiero. Para finalizar, se brindan las conclusiones y las referencias.

## 2. MODELOS OCULTOS DE MÁRKOV.

Un Modelo Oculto de Márkov (HMM) es un proceso estocástico doblemente embebido con un proceso aleatorio que no es observable (está oculto), pero que puede ser visto solamente a través de otro conjunto de procesos aleatorios que produzcan la secuencia de observaciones y satisface la siguiente propiedad: Propiedad de Márkov: Esta propiedad hace referencia a la falta de memoria de una proceso estocástico, esto es, el estado actual de un proceso únicamente depende de su estado anterior y no de la secuencia de eventos

que precedieron a éste último estado. Los HMM está definido por los siguientes componentes:

Q: Número de estados en el modelo, cada estado individual se etiqueta como  $q_1, q_2, \dots, q_N$ .

O: Secuencia de observaciones tomadas de un vocabulario  $V = \{v_1, v_2, \dots, v_M\}$ , denotadas como  $O = (O_1, O_2, \dots, O_M)$ .

A: la matriz de transición  $A = \{a_{ij}\}$ , donde cada  $a_{ij}$  representa la probabilidad del estado  $i$  al estado  $j$ :

$$a_{ij} = P(q_{t+1}=j | q_t = i).$$

B: Probabilidad de observación de cada símbolo del vocabulario, también llamada probabilidad de

emisión,  $b_j(k) = P(O_t = v_k | q_t = j)$ , es decir, la probabilidad de que la observación  $O_t$  sea generada por el estado  $j$ .

$\pi$ : Distribución de probabilidad sobre el estado inicial  $\pi = \{\pi_i\}$ , definida como:  $\pi_i = P(q_1=i)$ , es decir, la probabilidad de iniciar en alguno de los estados de  $q$ .

Es evidente que una especificación completa de un HMM requiere la estimación de los parámetros del modelo,  $A, B$  y  $\pi$  por lo tanto un HMM se define como  $\lambda$ .

$$\lambda = (A, B, \pi) \tag{1}$$

Este modelo toma dos suposiciones importantes. La primera, al igual que en las cadenas de Markov de primer orden, es la suposición de Markov que establece que el estado actual únicamente depende de su predecesor, lo que puede ser expresado como:

$$\text{Suposición de Márkov } P(q_t | q_1, \dots, q_{t-1}) = P(q_t | q_{t-1}) \tag{2}$$

Como segunda suposición, la probabilidad de observación  $O_t$  depende únicamente del estado que la produce  $q_t$  y no de cualquier otro estado u observación, expresado matemáticamente como:

$$\text{Independencia de salida: } P(O_t | q_1 \dots q_t, \dots, q_T, o_1, \dots, o_t, \dots, o_T) = P(o_t | q_t) \tag{3}$$

### 3. ALGORITMOS DE LOS MODELOS OCULTOS DE MÁRKOV.

Los HMM, según Rabiner [7], se caracterizan por tres problemas esenciales, que son necesarios resolver para hacer que los HMM sean aplicables a problemas reales. Estos problemas son:

Problema 1 (Probabilidad): El primer problema consiste en la evaluación de una secuencia de observaciones  $O$  dado el modelo  $\lambda = (A, B, \pi)$ . La solución a este problema brinda la probabilidad de que la secuencia sea generada por ese modelo  $\lambda$ , es decir:  $P(O|\lambda)$ .

Problema 2 (Descodificación): Dada una secuencia de observaciones  $O$  y el modelo  $\lambda = (A, B, \pi)$ , es necesario determinar cuál es la secuencia de estados  $Q$  más probable que haya generado las observaciones  $O$ .

Problema 3 (Aprendizaje): Ajustar los parámetros  $A$  y  $B$  del modelo para maximizar las probabilidades de las observaciones  $P(O|\lambda)$ .

A continuación, se desarrollan aspectos fundamentales del problema.

### Probabilidad

Para obtener la probabilidad total de la observación, se tiene que sumar sobre todas las posibles secuencias de  $Q$  estados ocultos, en general:

$$P(O|\lambda) = \sum_Q P(O, Q) = \sum_Q P(O|Q) P(Q) \quad (4)$$

Se nota que si se tiene un número de estados ocultos muy grande al igual que el tamaño de la cadena de observación es muy grande existe un problema computacional del orden exponencial  $O(N^T)$  lo cual implica que el tiempo de cómputo es muy grande. Para atacar el problema de tiempo computacional se tendrá que hacer uso del algoritmo Forward que es del orden  $O(N^2T)$ , el cual también evita hacer cálculos repetidos

### Algoritmo Forward

La probabilidad de estar en el estado  $j$  después de  $t$  observaciones, se define con la variable  $\alpha_t(j)$

$$\alpha_t(j) = P(O_1, O_2 \dots O_t, q_t = j|\lambda) \quad (5)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\alpha_t(j) = \left( \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad (6)$$

### Algoritmo Forward

Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

Recursividad:

$$\alpha_t(j) = \left( \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad 1 \leq j \leq N, 1 < t \leq T$$

Terminación:

$$P(O|\lambda) = \left( \sum_{i=1}^N \alpha_T(i) \right)$$

### Algoritmo Backward

La probabilidad de las observaciones desde el tiempo t+1 hasta el final, y estar en el estado i al tiempo t, se define con la variable  $\beta_t(i)$

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = i | \lambda) \quad (7)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (8)$$

Para más detalles de la demostración matemática de las recurrencias de los algoritmos consulte [4],[5] y [6].

### Algoritmo Backward

Inicialización:

$$\beta_T(i) = 1 \quad 1 \leq i \leq N$$

Recursividad:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \beta_{t+1}(j) b_j(O_{t+1}) \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N$$

Terminación:

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(O_1) \beta_1(j)$$

## Descodificación

El segundo problema del HMM es la descodificación, que consiste en obtener la secuencia de estados ocultos Q que generen las observaciones O, es decir, de entre todas las secuencias de estados que puedan generar determinada secuencia de observaciones, se desea encontrar aquella que sea más representativa de la observación O. Para ello, se tiene que encontrar cual es la secuencia de estados ocultos con mayor probabilidad. Se puede resolver este problema ejecutando el algoritmo Viterbi que es muy similar al algoritmo Forward.

La probabilidad de estar en el estado j después de t observaciones y pasar sobre la secuencia de estados más probables,  $q_0, q_1, \dots, q_{T-1}$ , se define con la variable  $v_t(j)$ , de la siguiente forma:

$$v_t(j) = \max_{q_0, q_1, \dots, q_{T-1}} P(q_0, q_1, \dots, q_{t-1}, O_1, O_2, \dots, O_t = j | \lambda) \quad (9)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con el cálculo del valor máximo de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (10)$$

Ahora, se define también la variable denominada backtrace(bt) similar a la variable del algoritmo de Viterbi ( $v_t$ ), en lugar de mantener la probabilidad más alta, esta variable mantiene el número del estado con dicha probabilidad.

$$bt_t(j) = \arg \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (11)$$

El valor de cada celda se puede calcular mediante el siguiente algoritmo:

*Algoritmo Viterbi.*

Inicialización:

$$v_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

$$bt_1(i) = i \quad 1 \leq i \leq N$$

Recursividad:

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad 1 \leq j \leq N, \quad 1 < t \leq T$$

$$bt_t(j) = \arg \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad 1 \leq j \leq N, \quad 1 < t \leq T$$

Terminación:

$$P^* = \max_{i=1}^N v_T(i)$$

$$q_T^* = \arg \max_{i=1}^N v_T(i)$$

Para las demostraciones matemáticas detalladas de las relaciones utilizadas en los algoritmos revisar el informe de Millán (2019) [5]

### Algoritmo Baum-Welch.

Este es el problema crucial de los HMM y consiste en cómo aprenden los parámetros de un HMM, es decir, las matrices A y B. Por lo tanto, será necesario determinar las probabilidades que mejor se adaptan al modelo, a través de una secuencia de entrenamiento. Sin embargo, para un HMM, al tener estados ocultos, no se puede realizar directamente el cálculo de las transiciones, pues lo único que se conoce son las observaciones. Una forma de resolverlo es iniciar con una estimación para las probabilidades de transición y emisión e iterar para estimar mejores probabilidades, dicho algoritmo se conoce como Forward-Backward o Baum-Welch, basado en el algoritmo de Maximización de la Esperanza (EM, Expectation Maximization). Se inicia por estimar las probabilidades de transición  $\hat{a}_{ij}$ :

$$\hat{a}_{ij} = \frac{\text{estimación de transiciones del estado } i \text{ a } j}{\text{estimación de transiciones desde el estado } i} \quad (12)$$

Para estimar el numerador, se tienen que sumar las transiciones de i hacia j en cada tiempo t, la estimación de cada una de estas probabilidades en un tiempo t en particular se define como: La probabilidad de estar en el estado i en el tiempo t y en j en el tiempo t+1 dada la secuencia de observación y modelo, la denotaremos por  $\xi_t(i, j)$ , de esta manera:

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda) = \frac{P(q_t=i, q_{t+1}=j, O | \lambda)}{P(O | \lambda)} \quad (13)$$

Haciendo uso de las variables Forward y Backward llegamos al siguiente resultado:

$$\xi_t(i, j) = \frac{b_j(O_{t+1})\beta_{t+1}(j)a_{ij}\alpha_t(i)}{\sum_{i=1}^N \alpha_T(i)} \quad (14)$$

Este cálculo únicamente para un tiempo t en particular, así que para obtener la estimación de transiciones de i hacia j así como la estimación total de las transiciones de i a cualquier otro estado se tendrá que sumar sobre todos los tiempos, (hasta T-1, ya que el tiempo T no hay transición) de esta manera se tiene:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{k=1}^N \xi_t(i, k)} \quad (15)$$

Se procede de forma similar para la estimación de la probabilidad de emisión:

$$\hat{b}_j(k) = \frac{\text{estimación de estar en el estado } j \text{ y observar el símbolo } k}{\text{estimación de veces de estar en el estado } j} \quad (16)$$

Para estimar el numerador necesitamos conocer la probabilidad de estar en el estado j en el tiempo t y la denotaremos por  $\gamma_t(j)$ , formalmente como:

$$\gamma_t(j) = P(q_t = j | O, \lambda) = \frac{P(q_t=j, O | \lambda)}{P(O | \lambda)} \quad (17)$$

Nuevamente, usando las variables de Forward y Backward se llega a:

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{i=1}^N \alpha_T(i)} \quad (18)$$

De igual forma, este cálculo es únicamente para un tiempo t en particular donde se observa k, así, para obtener la estimación de la probabilidad de emisión de k, se tiene que sumar sobre todos los tiempos t en donde se observe el símbolo k, para la estimación de estar en el estado j, se suma sobre todos los tiempos, por lo tanto:

$$\hat{b}_j(k) = \frac{\sum_{t=1}^T O_{t=k} \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (19)$$

La demostración del algoritmo de Baum Welch la puede revisar en Richard (1998) [8].



#### **4. APLICACIÓN 1: CAMBIO DE USO DE SUELOS EN MANGLARES DE 1978 A 2017**

Un Modelo Oculto de Markov (HMM) se puede definir como una Cadena de Markov en un espacio producto. El análisis del Cambio de Uso de Suelo a través del método de cadenas de Márkov permite profundizar en proyecciones territoriales y realizar análisis probabilísticos y estadísticos de los problemas bajo estudio. Las cadenas de Márkov representan una clase de procesos estocásticos de gran interés para aplicaciones prácticas. En particular, las cadenas de Márkov de tiempo discreto (DTMC) permiten modelar las probabilidades de transición entre estados discretos mediante la ayuda de matrices. Para la aplicación del modelo de Márkov se han generado diversos paquetes de computación como lo refiere Bai, et al. (2017) [1].

El paquete Markovchain para R es un paquete específicamente dedicado al análisis DTMC, Spedicato (2017) [10] y [11]. Este tiene como objetivo lograr una mayor flexibilidad en el manejo de DTMC, que otras soluciones existentes no poseen, proporcionando clases S4 para Cadenas de Márkov homogéneas y no homogéneas. En aplicaciones de los HHM al estudio de cambio de cobertura y uso de suelos este paquete es de gran utilidad.

##### **Distribución de los procesos de Cambio de Uso de Suelo**

La distribución de los procesos de cambio cartografiados en la Barra San José, Chiapas se presenta en [8]. El proceso de cambio que predomina en el periodo de análisis en toda la zona de estudio es el incremento de los asentamientos humanos y el avance de la frontera agrícola, registrando un incremento de 5,941 ha en 39 años. Sin embargo, la deforestación de selvas, bosques y manglares se presentó en forma más acentuada, debido a que este proceso afectó el 49 % de la superficie

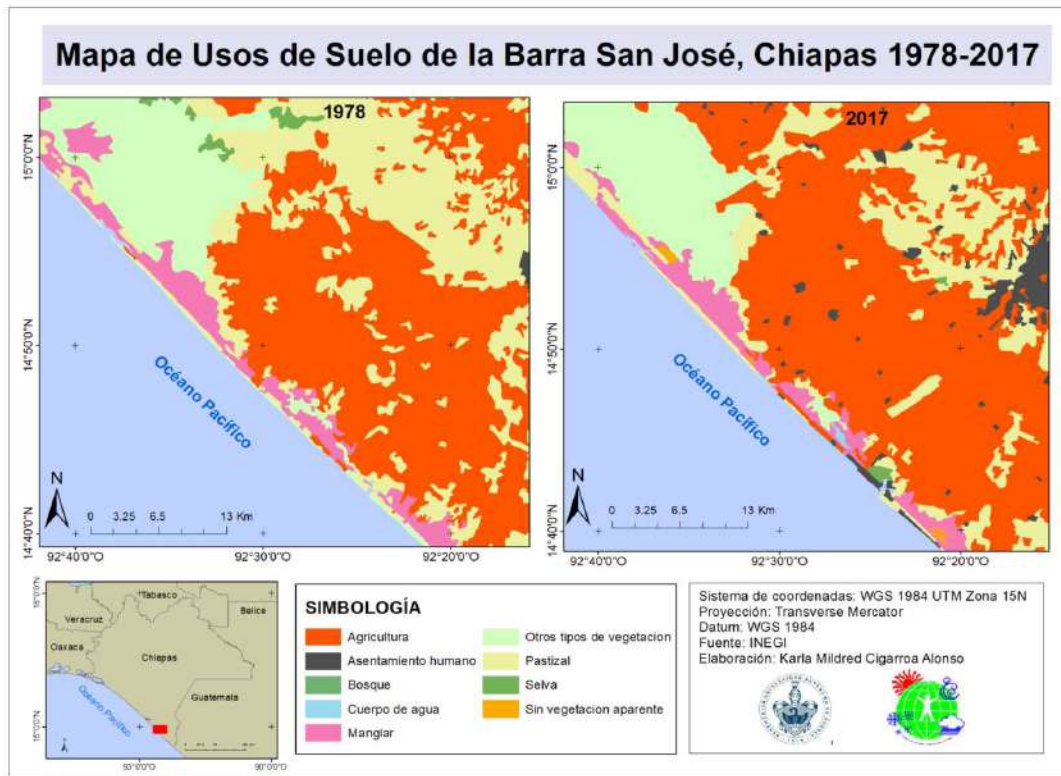
##### **Magnitud de los procesos de Cambio de Uso de Suelo**

La magnitud de los procesos de Cambio de Uso de Suelo en la Barra San José, Chiapas se muestra en la figura 1. En el mapa se observa que el proceso de cambio que predomina en el periodo de análisis es la deforestación de Selvas a Pastizales, Y la deforestación de Manglares a sitios de Asentamientos humanos.

Estos procesos de cambio también afectan las áreas forestales que se desarrollan en el interior de la reserva de la biosfera localizada en la Barra San José; sin embargo, los impactos más importantes se localizan en las periferias de los Asentamientos humanos.

Los procesos de Cambio aquí reportados se deben en gran medida a que la ganadería extensiva constituye en el sureste de México una alternativa con bajos costos y mejor beneficio económico para los campesinos y pequeños productores, ante la constante crisis de los precios de cultivos agroforestales con mayores beneficios ambientales como el café

y el cacao (*Theobroma cacao*); además del fuerte impacto de las olas de migrantes que recibe el estado de Chiapas.



**Figura 1.** Mapa de Usos de Suelo y tipos de vegetación de la Barra San José, Chiapas para el periodo 1978-2017

### Sistema de clasificación de Usos de Suelo

En la tabla 1 se relacionan las siglas utilizadas para cada una de las clases de Usos de suelo que se relacionan en las matrices de transición y probabilidad.

**Tabla 1.** Clases de las matrices de transición de Uso de Suelo y Probabilidad

Clase	Nombre	Siglas
1	Agricultura	AGR
2	Asentamiento humano	AH
3	Bosque	BO
4	Cuerpo de agua	CA
5	Manglar	M
6	Otros tipos de vegetación	O
7	Pastizal	PA
8	Selvas	SE
9	Sin vegetación aparente	SVA

A partir de las matrices de detección de cambio generadas de la sobre posición de los mapas 1978-2017 (Tabla 2) se estimaron los porcentajes de cambio en 1978 dominaban las Selvas y Pastizales inducidos o cultivados, esta condición cambio para 2017 aumentando la Agricultura y las Áreas sin vegetación aparente, indicativo de un alto grado de deforestación en la zona. Los Asentamientos humanos aumentaron drásticamente de 0.05 a 1.51% del total de la superficie. A lo largo de la franja de manglar se observaron trazos y delimitaciones de terrenos que fragmentan aún más a este ecosistema. Este patrón y su afectación pueden considerarse de gran magnitud, ya que la fragmentación del manglar afecta la integridad y funcionalidad ecológica, además de que limita aún más su regeneración.

**Tabla 2.** Matriz de transición de Uso de Suelo para el periodo 1978-2017

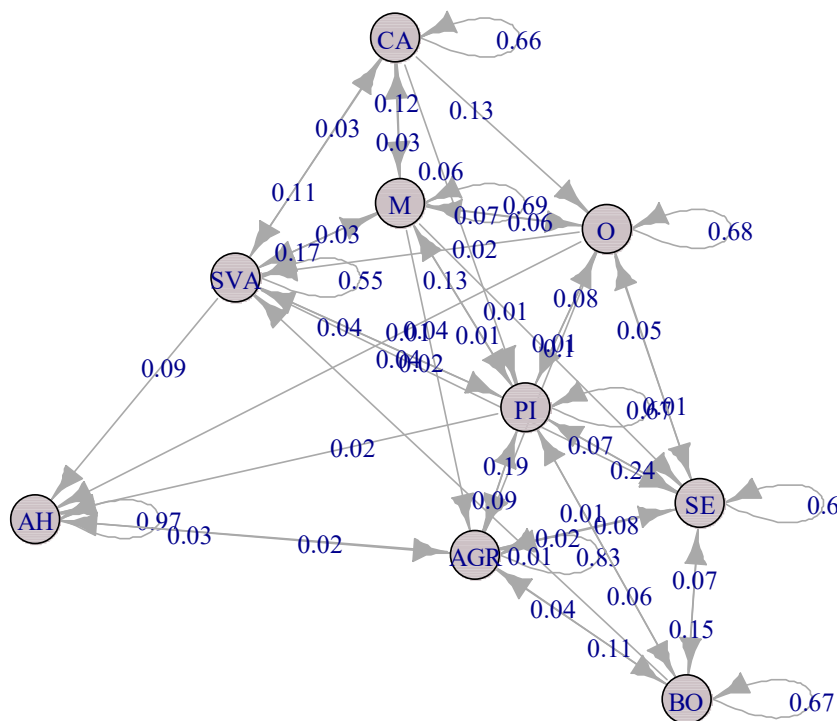
1978	2017									Total 1978
	AGR	A H	BO	C A	M	O	PA	SE	SV A	
Agricultura	9,816	27 7	482	6	22	18	1,027	191	5	11,844
Asentamiento humano	9	36	0	0	0	0	3	0	0	48
Bosque	673	10	4,20 5	0	0	15	400	964	0	6,267
Cuerpo de agua	4	4	0	32 5	59	68	32	2	16	510
Manglar	139	20	0	12 5	2,66 4	259	508	25	120	3,859
Otros tipos de vegetación	283	24	0	14	152	1,84 3	213	148	42	2,718
Pastizal	4,897	57 5	286	83	350	167	16,93 6	1,859	103	25,255
Selva	1,032	21	846	5	14	89	2,946	7,372	30	12,354
Sin vegetación aparente	0	5	0	9	9	0	5	2	32	62
Total 2017	16,85 2	97 0	5,81 9	56 7	3,26 9	2,45 9	22,07 0	10,56 2	347	62,916

La probabilidad de cambio para el periodo 1978-2017 se muestra en la Tabla 3. La categoría que tienen mayor probabilidad de permanecer en el tiempo es Asentamiento humano (0.97). Mientras que la clase que tuvo una menor probabilidad de permanecer en el mismo tipo de cobertura-uso, fue Sin vegetación aparente (0.56). El Manglar tuvo una probabilidad baja de mantenerse (0.69), registrando transiciones hacia la Agricultura, Cuerpos de agua, Otros tipos de vegetación y Pastizales.

**Tabla 3.** Matriz de probabilidad de transición durante 1978-2017

1978	2017									Total
	AGR	AH	BO	CA	M	O	PA	SE	SVA	
Agricultura	0.83	0.02	0.04	0.00	0.00	0.00	0.09	0.02	0.00	1.00
Asentamiento humano	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Bosque	0.11	0.00	0.67	0.00	0.00	0.00	0.06	0.15	0.00	1.00
Cuerpo de agua	0.00	0.00	0.00	0.66	0.12	0.13	0.06	0.00	0.03	1.00
Manglar	0.04	0.00	0.00	0.03	0.69	0.07	0.13	0.01	0.03	1.00
Otros tipos de vegetación	0.10	0.01	0.00	0.00	0.06	0.68	0.08	0.05	0.01	1.00
Pastizal	0.19	0.02	0.01	0.00	0.01	0.01	0.67	0.07	0.00	1.00
Selva	0.08	0.00	0.07	0.00	0.00	0.01	0.24	0.60	0.00	1.00
Sin vegetación aparente	0.00	0.09	0.00	0.11	0.17	0.00	0.04	0.04	0.56	1.00

En la Figura 2 se muestra la probabilidad de cambio de las clases entre el periodo 1978-2017. Se observa que todas las clases, a excepción de los Asentamiento humanos, indican probabilidades de transición hacia Pastizales. Es clara la dinámica de desplazo de los Bosques y Selvas hacia Agricultura y Pastizales. Se observa un intercambio importante entre el Manglar y los Cuerpos de agua, denotado por la proximidad entre las clases, marcando la fuerte dinámica entre los ecosistemas costeros de transición. Los sitios que ya no son productivos, por ser escasos en recursos forestales, ser parcelas de cultivos abandonadas o pertenecer a zonas con sobrepastoreo muestran altas transiciones a Asentamientos humanos, como reflejo de una evolución socio-cultural.



**Figura 2.** Diagrama de flujo de las probabilidades de transición de cada Uso de Suelo en el periodo 1978-2017.

Puede apreciarse que se ha realizado un análisis de la magnitud, dinámica y distribución de los procesos de Cambio de Uso de Suelo del manglar en los márgenes del vaso principal del sistema lagunar Barra San José, Chiapas, utilizando Sistemas de Información Geográfica y Cadenas de Markov en el periodo 1978 a 2017.

Se corroboran los factores que influyen en el Cambio de Uso de Suelo en esta zona, esto es, que la acelerada deforestación y la intensificación de actividades antrópicas, como la ganadería y la agricultura, están dificultando el buen manejo y la conservación de los manglares de la zona.

Se deben desarrollar buenas estrategias sobre el manejo de los manglares de la Barra San José, Chiapas, antes de implementar alguna política que pueda resultar en severas afectaciones en los remanentes de estos ecosistemas y su diversidad.

Debe destacarse que la importancia del estudio del Cambio en la Cobertura y Uso de Suelo está siendo ampliamente reconocida en la actualidad, planteándose que “una ciencia integrada del suelo” está emergiendo, uniendo las ciencias ambientales, humanas, los desarrollos tecnológicos en Sistemas de Información Geográfica junto a la modelación de procesos estocásticos, con el fin de resolver diferentes inquietudes sobre los Cambios y Uso de Suelo y los impactos que estos cambios tienen en la humanidad y en el medio ambiente.

Es necesario dar a conocer estos resultados a las diferentes autoridades encargadas de la planificación territorial, de manera que conozcan e implementen metodologías de planificación espacial para detectar y establecer las posibles modificaciones del crecimiento urbano que permitan reorientar y minimizar los impactos sobre los manglares de la Barra San José, Chiapas, bajo un contexto de sustentabilidad.

## **5. APLICACIÓN 2: DETECCIÓN DE DESBALANCE FINANCIEROS**

Cuando se tiene una cuenta bancaria el banco quiere detectar cuando hay una transacción bancaria que desbalancea la cuenta, es decir, está en números rojos. Se usan Modelos Ocultos de Márkov para detectar transacciones bancarias que conducen a un desbalance financiero.

### **Información seleccionada de la base de datos Berka.**

La información contenida en la base de datos Berka [2], contiene alrededor de un millón de transacciones bancarias. Se separó por el número de cuenta con sus respectivas transacciones y balance de cuenta. Posteriormente se seleccionaron dos cuentas bancarias que pudieran aportar diferentes puntos de vista sobre el comportamiento de los movimientos que se realizan. Para el experimento se tomó una cuenta con balance

positivo y una cuenta con algunos balances negativos. Las dos cuentas seleccionadas se muestran en la tabla las cuentas y el estado de balance.

**Tabla 4.** Cuentas bancarias de la Base de Datos Berka [2]

Número de cuenta	Tipo de balance	Cantidad obtenida del balance	Estado de balance
1069	Balance positivo	25220.50	Positivo
2335	Balance negativo cercano al neutro	2785.50	Negativo

Se muestra a continuación el comportamiento en las transacciones de las cuentas seleccionadas. Se enfoca en los montos que representa cada transacción y el balance que tiene la cuenta en cada momento de la transacción.

### Cuenta 1069

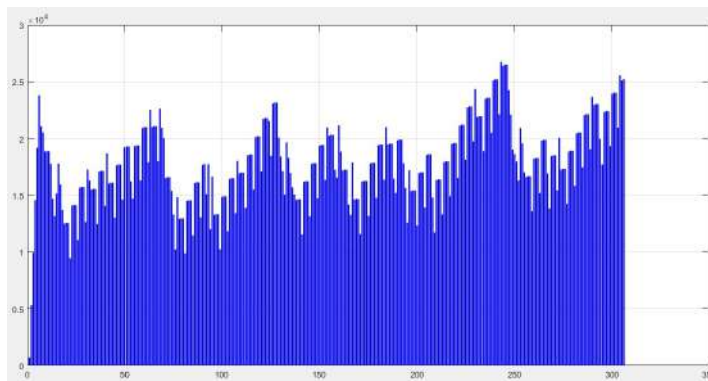


Figura 3. Comportamiento del balance positivo en la cuenta 1069

### Cuenta 2335

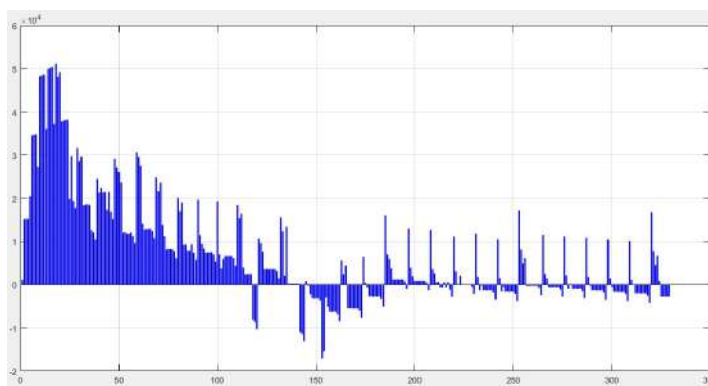


Figura 4 Comportamiento del balance negativo en la cuenta 2335

## Determinación del Modelo Oculto de Márkov.

En el caso sobre el análisis de la Base de datos Berka, se usa como estados del modelo la información de las variables del campo Operation: "VYBER KARTOU" retiro de tarjeta de crédito, "VKLAD " crédito en efectivo, "PREVOD Z UCTU" Deposito de otro banco, "VYBER" retiro en efectivo, "PREVOD NA UCET " remesas a otro banco, "YYYYY" Datos sin identificación.

Para definir las observaciones del modelo se siguió lo propuesto en el artículo Srivastava (2008) [12], dividir las diferentes cantidades de las transacciones en tres particiones usando el algoritmo del vecino más cercano con tres centroides obteniendo la observación por estado en la forma {1,2,3}.

Se muestra en la figura 3 la determinación del HMM para cada cuenta bancaria.

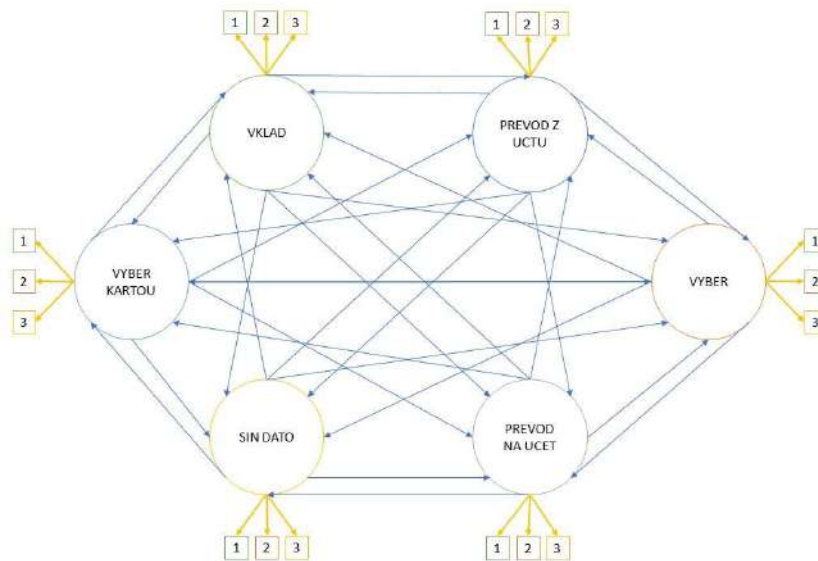


Figura 5. Modelo Oculto de Márkov para las cuentas bancarias de la base de datos Berka [2].

## Convergencia de las matrices de transición.

Se analiza el comportamiento de las matrices A y B del HMM, se nota que entre mayor sea el tamaño de muestra para calcular estas matrices, cada término de la matriz de transición converge. A continuación, se muestran los valores de la transición del estado i al estado j conforme aumenta el número de observaciones. Se muestra los resultados para una matriz de transición de tres estados, el elemento de la matriz  $A_{11}$  es cero

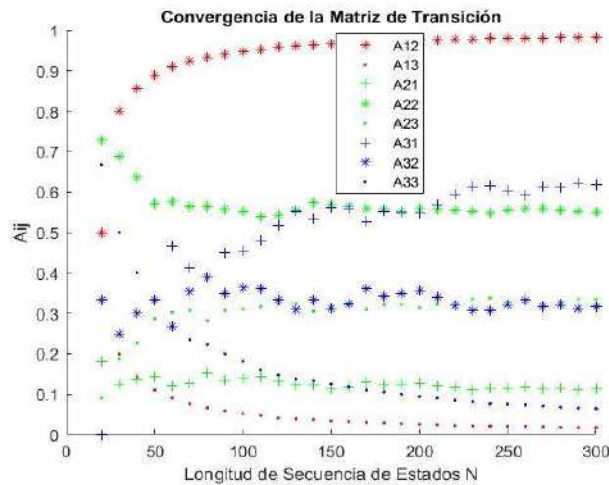


Figura 6. Convergencia de cada término de la matriz de transición

### Determinación del Desbalance Financiero

Para realizar las pruebas sobre las cuentas bancarias se compara el logaritmo de la probabilidad de la secuencia de estados más probable para la secuencia de observaciones  $O_1, O_2, \dots, O_t$  con el logaritmo de la probabilidad de la secuencia de estados para la secuencia de observaciones

$O_1, O_2, \dots, O_{t+1}$  que se obtienen al aplicar el algoritmo Viterbi . A continuación, se describe el procedimiento:

*Paso uno:* Se determina un número aleatorio, que en las secuencias de observaciones se considera como la

observación  $O_{t+1}$  que se quiere determinar si es un desbalance o no.

*Paso dos:* Para entrenar el HMM se calculan las matrices de emisión y transición, se usa la secuencia de

observaciones de  $O_1, \dots, O_t$  y sus respectivos estados.

*Paso tres:* Determinar la secuencia de observaciones inicial de tamaño  $N$ ,  $O_1, \dots, O_N$ ,

siendo  $O_N = O_t$ .

*Paso cuatro:* Determinar la secuencia de estados más probable y comparar el número de estados dado en el

paso dos.

*Paso cinco:* Calcular el Logaritmo de la Probabilidad total de la secuencia de estados obtenida en el paso 4.



*Paso seis:* Determinar la secuencia de observaciones final de tamaño  $N+1$ ,  $O_1, \dots, O_{N+1}$ , siendo

$O_{N+1} = O_{t+1}$ , que incluye la observación que se desea determinar si es desbalance financiero.

*Paso siete:* Determinar la secuencia de estados más probable para la secuencia de observaciones del paso 6

*Paso ocho:* Calcular el Logaritmo de la Probabilidad total de la secuencia de estados obtenida en el paso 7.

*Paso nueve:* Se calcula Delta definida como la diferencia de Logaritmo de la probabilidad de la secuencia de

estados Final (Paso 8) menos Logaritmo de la probabilidad de la secuencia de estados Inicial (Paso 5).

*Paso diez:* Se calcula la Delta proporcional respecto a Logaritmo de la probabilidad de la secuencia de estados Inicial

A continuación, se presenta los resultados para cada cuenta bancaria.

### Cuenta Bancaria 1069

En el caso de la cuenta 1069 no se tiene ninguna transacción bancaria en la cual se ocasione un balance negativo. Así que se elige aleatoriamente una observación  $O_{t+1}$ , las matrices de transición y emisión resultan:

Matriz de transición.

0	0	0	1	0	0
0	0.21052632	0.15789474	0.10526316	0.42105263	0.10526316
0	0	0.08888889	0.33333333	0	0.57777778
0	0.08333333	0.05	0.275	0.4	0.19166667
0.01639344	0.03278689	0.52459016	0.19672131	0.08196721	0.14754098
0	0.03389831	0	0.96610169	0	0

Matriz de emisión.

1	0	0
0.15789474	0.15789474	0.68421053
0	1	0
0.74166667	0.05	0.20833333
0.01639344	0.19672131	0.78688525
1	0	0

*Paso tres:* Determinar la secuencia de observaciones inicial de tamaño  $N$ ,  $O_1, \dots, O_N$ ,

siendo  $O_N = O_t$

Secuencia Observaciones Inicial

2	1	1	3	2	1	1	3	3	2	1	1	3	2	3	1	1	3	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia Estados Inicial

4	5	4	5	3	3	4	5	4	3	5	4	5	3	4	6	4	5	5	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

*Paso cuatro:*

Secuencia de estados más probable obtenida

4	6	4	5	3	6	4	5	5	3	6	4	5	3	4	6	4	5	3	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Teniendo esta secuencia de estados generada por el algoritmo de viterbi se compara con la secuencia de estados de entrada, se obtiene 0.7000

*Paso cinco:* Calcular el Logaritmo de la Probabilidad total de la secuencia de estados

Logaritmo de la probabilidad de la secuencia de estados Inicial: -18.6465

*Paso seis:* Secuencia Observaciones Final

2	1	1	3	2	1	1	3	3	2	1	1	3	2	3	1	1	3	2	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

*Paso ocho:* Logaritmo de la probabilidad de la secuencia de estados más probable Final:  
-19.1147

*Paso nueve:* Delta: -0.4682

*Paso diez:* Delta proporcional = 0.0251

Del paso tres al diez se realiza para diferentes tamaños de  $N$ . En cada una de las corridas se genera un número aleatorio diferente. En la siguiente tabla se reporta el promedio de cinco corridas para cada tamaño de la secuencia de observaciones inicial.

**Tabla 5.** Comportamiento de la Cuenta bancaria 1069 de la Base de Datos Berka [2], con balance positivo

Longitud de la Observación	Delta	Delta Proporcional
20	-0.56176	0.0354
17	-0.56176	0.04016
14	-0.56176	0.0544
11	-0.56176	0.05748
8	-0.56176	0.0871
5	-0.56288	0.14836

### Cuenta Bancaria 2335

En el caso de la cuenta 2335 se encuentra un número considerable de transacciones que generan en algún punto un balance negativo, que después se recupera a positivo sin embargo para efectos del experimento se define el HMM para un elemento antes del primer balance negativo. Para este caso se obtienen las matrices de transición y emisión:

Matriz de transición.

0	0	0	0	1	0
0	0.11904762	0.0952381	0.23809524	0.35714286	0.19047619
0	0.18181818	0.13636364	0.31818182	0.13636364	0.22727273
0	0.08045977	0.05747126	0.2183908	0.14942529	0.49425287
0	0.2195122	0.09756098	0.18292683	0.45121951	0.04878049
0.0106383	0.07446809	0.0212766	0.38297872	0.13829787	0.37234043

Matriz de emisión.

0	0	1
0.47619048	0.45238095	0.07142857
0.95454545	0.04545455	0
0.71264368	0.10344828	0.18390805
0.75609756	0.09756098	0.14634146
1	0	0

*Paso tres:*

Secuencia Observaciones Inicial

1	1	2	2	1	1	1	1	1	1	1	1	2	1	1	3	1	1	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia de estados Inicial

4	3	5	3	3	2	4	6	4	6	5	5	5	5	2	4	4	6	4	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

*Paso cuatro:* Secuencia de estados más probable obtenida

5	5	2	4	6	6	6	6	6	6	6	6	6	4	6	6	4	6	6	6	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Se compara la secuencia de estados generada por el algoritmo de viterbi con la secuencia de estados de entrada, en este caso se obtiene: 0.2500

*Paso cinco:* Calcular el Logaritmo de la Probabilidad total de la secuencia de estados más probable Inicial:

-12.5840

*Paso seis:* Secuencia Observaciones Final, que incluye la observación que se desea determinar si es desbalance financiero

1	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	3	1	1	1	1	3
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

*Paso ocho:* Logaritmo de la probabilidad de la secuencia de estados Final: -14.9128

*Paso nueve:* Delta: -2.3288

*Paso diez:* Delta proporcional = 0.1851

Del paso tres al diez se realiza para diferentes tamaños de N. En la siguiente tabla se reporta Delta y la Delta proporcional para diferentes tamaños de la secuencia de observación inicial.

**Tabla 6.** Comportamiento de la Cuenta bancaria 2335 de la Base de Datos Berka [2], con balance negativo

Longitud	Delta	Delta Proporcional
20	-1.4774	0.1412
17	-1.4774	0.1726
14	-1.4774	0.238
11	-1.4773	0.2767
8	-1.477	0.3033
5	-1.4739	0.5127

Se observa en la Tabla 2, que cuando no hay desbalance financiero el valor absoluto de Delta es menor que uno y el porcentaje de variación es menor a 15%. Mientras que la Tabla 3 se muestra que cuando existe un desbalance financiero, el valor absoluto de la variable Delta es mayor que uno y el porcentaje de variación es mayor a 14%.

Los Modelos Ocultos de Márkov permiten modelar la información de las transacciones bancarias de un usuario.

Este modelo lo representan las matrices de Transición y Emisión. Se demuestra numéricamente, con la base de datos Berka [2], la convergencia de la matriz de transición cuando el tamaño de la secuencia de estados aumenta.

## 6. CONCLUSIONES

Se han presentado dos aplicaciones de los HMM. Una sobre el Cambio de Uso de Suelo, donde se utiliza Cadenas de Markov para modelar y resolver el problema y otra, para detectar el desbalance de cuentas bancaria.

En la primera, los resultados muestran que, en la Barra San José, Chiapas se conjugan dinámicas regionales y locales de cambios, las cuales deben ser tomadas en cuenta en el proceso de planeación y ejecución de políticas de conservación para lograr una mejor conservación y uso de los recursos. Estos estudios ayudan a conocer el tipo de manejo y uso que el hombre hace de la naturaleza en un territorio, y a partir de esto, se puede predecir la dirección del cambio y, en consecuencia, se pueden medir las implicaciones ambientales o sociales de estos cambios probables, sus repercusiones económicas y establecer mecanismos para controlar las causas de origen, contribuyendo a generar un sistema de monitoreo periódico.

En la segunda, también estos modelos permiten determinar, para una cuenta bancaria, si existe un desbalance financiero, comparando dos secuencias de tamaño  $N$  y  $N+1$ . Como trabajo futuro resultaría de gran interés aplicar este proceso a una base de datos de tarjetas de crédito, lo que permitiría detectar la existencia de fraudes.

## REFERENCIAS

1. BAI, J., C. DEL CAMPO, L. R. KELLER. (2017): Modelos de cadenas de Márkov en la práctica: una revisión de opciones de software de bajo coste. *Revista Investigación Operacional*. Vol. 38, No. 1, pp 56-62.
2. BASE DE DATOS BERKA , recuperada de <http://lisp.vse.cz/>, consultado : marzo de 2019
3. CIGARROA-ALONSO K.M, LINARES-FLEITES G, SANDOVAL-SOLÍS M.L, VALERA-PÉREZ M.A. (2017). Análisis de Cambio de Uso de Suelo a través de Modelos Ocultos de Márkov. *Algoritmos, estrategias y modelos en las aplicaciones de métodos cuantitativos*, Editores Bouza Herrera C. N, García Rodríguez J. F, Santiago Moreno A, Rueda García M., RIDECA. Habana, Cuba. pp 166-177.

4. JURAFSKY, D., & MARTIN, J. (2000): *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd edition), pp. 122-139.
5. MILLÁN ÁNGELES, V. (2019): Modelos Ocultos de Márkov, Informe de Prácticas Profesionales, Benemérita Universidad Autónoma de Puebla.
6. PACHECO MARTÍNEZ, M. (2019): Redes Neuronales Recurrentes Aplicadas al Reconocimiento de Voz y Traducción Automática de español a japonés. Tesis de Licenciatura en Ingeniería en Ciencias Computacionales. Benemérita Universidad Autónoma de Puebla.
7. RABINER, LAWRENCE R. (1989): A tutorial on Hidden Márkov Model and Selected Applications in Speech Recognition. *Proceeding of the IEEE*. Vol. 77, No. 2, pp 257-286.
8. RICHARD D., SEAN R. E., ANDERS K., GRAEME M. (1998): Biological sequence analysis. The Pitt Building. Trumpington Street, Cambridge, United Kingdom: Cambridge University Press.
9. SANDOVAL-SOLÍS, M.L., LINARES-FLEITES, G., CIGARROA, K.M. Y VALERA, M.A. (2017): Cadenas de Márkov en la Práctica: Aplicación con Lenguaje R. *Desarrollo E Implementación De Las Ciencias Computacionales*. Benemérita Universidad Autónoma de Puebla, México. pp 9-17
10. SPEDICATO, G.A. (2016): Márkovchain: An R Package to Easily Handle Discrete Markov Chains. R package version 0.4.
11. SPEDICATO, G.A. (2017): Discrete Time Márkov Chains with R. The R Journal. IOP Web.  
<https://journal.r-project.org/archive/2017/RJ-2017-036/index.html>
12. SRIVASTAVA, A., KUNDU, A., SURAL, S., & MAJUMDAR, A. (2008): Credit Card Fraud Detection Using Hidden Markov Model. *In Dependable and Secure Computing, IEEE Transactions on*, pp. 37-39.



# INNOVACIONES TECNOLÓGICAS EN LAS CIENCIAS COMPUTACIONALES

María del Carmen Santiago Díaz





**INNOVACIONES  
TECNOLÓGICAS EN LAS  
CIENCIAS  
COMPUTACIONALES**

**INNOVACIONES  
TECNOLÓGICAS EN LAS  
CIENCIAS  
COMPUTACIONALES**

María del Carmen Santiago Díaz  
Gustavo Trinidad Rubín Linares  
Ana Claudia Zenteno Vázquez  
Judith Pérez Marcial  
(Editores)

María del Carmen Santiago Díaz  
(Coordinador)

## Innovaciones Tecnológicas en las Ciencias Computacionales

*María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Ana Claudia Zenteno Vázquez, Judith Pérez Marcial*  
(editores)

*María del Carmen Santiago Díaz*  
(coordinador)

*María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Ana Claudia Zenteno Vázquez, Judith Pérez Marcial, Yeiny Romero Hernandez, Nicolás Quiroz Hernández, Alma Delia Ambrosio Vázquez, Héctor David Ramírez Hernández, Gregorio Trinidad García, María De Lourdes Sandoval Solís, Rogelio González Velázquez, Guillermina Sánchez Román, Jose Luis Hernández Ameca, Armando Espíndola Pozos, María Blanca del Carmen Bermudez Juárez, José Martín Estrada Analco, Luz Del Carmen Reyes Garcés, Meliza Contreras González, Luis Enrique Colmenares Guillén, Elsa Chavira Martínez, Pedro García Juárez, Nelva Betzabel Espinoza Hernández, Alberto Román Flores, Roberto Contreras Juárez, Graciano Cruz Almanza, Maya Carrillo Ruiz, Ernest Cortez, Bárbara Emma Sánchez Rinza, Beatriz Beltrán Martínez, José Ítalo Cortez, Hermes Moreno Álvarez.*  
(revisores)

Primera edición: 2019

ISBN: 978-607-7512-97-4

**Montiel & Soriano Editores S.A. de C.V.**

15 sur 1103-6 Col. Santiago Puebla, Pue.

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

*Rector:*

Dr. José Alfonso Esparza Ortiz

*Secretario General:*

Dr. José Jaime Vázquez López

*Vicerrector de Investigación y Estudios de Posgrado*

Dr. Ygnacio Martínez Laguna

*Directora de la Facultad de Ciencias de la Computación*

M.I. María del Consuelo Molina García

## Contenido

<b>Prefacio</b> .....	7
<b>Minería de Datos Aplicada al Pensamiento Ambiental Universitario, en la Oferta Educativa de Posgrados 2017 de la BUAP</b> <i>Mercedes Monserrat Domínguez Navarro</i> <i>Gladys Linares Fleites</i> <i>J. Santos Hernandez Zepeda</i> <i>Ricardo Dario Peña Moreno</i> .....	8
<b>Hacia la Educación 4.0: Una experiencia en ESIME Ticomán Instituto Politécnico Nacional</b> <i>Jorge Sandoval Lezama</i> <i>Tiburcio Fernández Roque</i> <i>José Arturo Correa Arredondo</i> .....	19
<b>Diseño de un sistema multiagente basado en un entorno colaborativo para el testing de seguridad en sistemas web</b> <i>Jorge Alberto Mata Piña</i> <i>Ricardo Velázquez Bautista</i> <i>Niels Martínez, Guevara</i> <i>Javier Sánchez Acosta</i> .....	29
<b>Control de inventario usando RFID y Particle Photon</b> <i>Jesús Antonio Palomares Juárez</i> <i>Aldrin Barreto Flores</i> <i>Salvador Eugenio Ayala Raggi</i> <i>Verónica Edith Bautista López</i> .....	37
<b>Una estrategia para estimular habilidades de identificación de figuras geométricas, aplicando la tecnología lúdica en educación básica</b> <i>José Luis Hernández Ameca</i> <i>Lilia Mantilla Narvaez</i> <i>Carlos Armando Ríos Acevedo</i> <i>Daniel Arenas Romero</i> .....	47
<b>Metodología para migración entre gestores de bases de datos: caso MySQL® y MariaDB®</b> <i>Luis Enrique Velázquez Martínez</i> <i>Juan Carlos Varela Tecalero</i> <i>Ana Claudia Zenteno Vázquez</i> <i>José Luis Pérez Rendón</i> .....	56

**Diseño Thinking para resolución de problemas**

*María del Carmen Santiago Díaz*

*Gustavo Trinidad Rubín Linares5510548020*

*Judith Pérez Marcial*

*Ana Claudia Zenteno Vázquez.....59*

**Análisis de estabilidad en algoritmos de caminata de robots humanoides**

*Yeiny Romero Hernández*

*María del Carmen Santiago Díaz*

*Gustavo Trinidad Rubín Linares*

*Martín Otamendi Torres.....68*

**Minería de datos como herramienta de prevención de muertes que involucran violencia intrafamiliar en la sociedad mexicana**

*Gustavo Mendoza*

*Yadira Laureano de Jesús*

*María Josefa Somodevilla García*

*María Concepción Pérez de Celis Herrero.....75*

**Diseño de Programas Paralelos asistido con Diagramas de Flujo de Datos**

*Mario Rossainz López*

*Mireya Tovar Vidal*

*Ana Patricia Cervántes Márquez*

*Hilda Castillo Zacatelco.....87*

**Algorithm for collecting and sorting data from twitter through the use of dictionaries in Python**

*María Beatriz Bernábe Loranca*

*Enrique Espinoza Monroy*

*José Martín Estrada Analco*

*Abraham Sánchez López.....98*

**Monitoreo de un roedor usando un sensor térmico con Arduino**

*Jesús Manuel Roa Escalante*

*Aldrin Barreto Flores*

*María Isabel Martínez García*

*Salvador Eugenio Ayala Raggi.....105*

**Unity y FPGA Aplicados a Robótica**

*José Esteban Torres León*

*Ricardo Fabian Caballero Morales*

*Marbella Muñiz Sánchez*

*Carlos García Lucero.....113*

**SIG y Cadenas de Márkov en modelos ambientales, caso Atoyatempan, Puebla.**

*Luis Ignacio Juárez Ruanova*

<i>Gladys Linares Fleites</i> <i>María De Lourdes Sandoval Solís</i> <i>Karla Mildred Cigarroa Alonso</i> .....	121
<b>Análisis de sensibilidad mediante software en programación lineal</b> <i>María Beatriz Bernábe Loranca</i> <i>Luis Alejandro Cabrera Cruz</i> <i>Rogelio González Velázquez</i> <i>Gerardo Martínez Guzmán</i> .....	127
<b>Prueba de Shapiro –Wilk a una Muestra de Nanoreservorios de Taninos Extraídos de Mimosa Tenuiflora (Tepezcohuite) Encapsulados en Matrices de TiO2</b> <i>Marcos González Flores</i> <i>Albino Moreno Rodríguez</i> <i>José Ítalo Cortez</i> <i>Adrián Hernández Santiago</i> .....	135

## SIG y Cadenas de Márkov en modelos ambientales, caso Atoyatempan, Puebla.

Luis Ignacio Juárez Ruanova<sup>1</sup>, Gladys Linares Fleites<sup>1</sup>, María de Lourdes Sandoval Solis<sup>2</sup>,  
Karla Mildred Cigarroa Alonso<sup>1</sup>

<sup>1</sup> Posgrado en Ciencias Ambientales, Benemérita Universidad Autónoma de Puebla.  
4 Sur No.104, Colonia Centro, C.P. 72000, Puebla, México.

<sup>2</sup> Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla.  
4 Sur No.104, Colonia Centro, C.P. 72000, Puebla, México.

<sup>1</sup>bioluis1@hotmail.com <sup>1</sup>gladys.linares@correo.buap.mx

<sup>2</sup>maria.sandoval@correo.buap.mx <sup>1</sup>k.cigarroa@hotmail.com

**Resumen.** La problemática ambiental requiere cada vez más del uso de software para simular las dinámicas que se encuentran relacionados en el sistema ambiente-sociedad. Los SIG son de gran ayuda para procesar la cartografía en zonas determinadas en tiempos determinados, aunado a; los modelos estocásticos de probabilidad que permiten un análisis del comportamiento del sistema, que, de otra manera, no sería posible en los sistemas complejos. El cambio de uso de suelo es un sistema ambiente-sociedad, donde las diferentes actividades antropogénicas transforman el ambiente. Los SIG junto con las cadenas de Márkov, son técnicas que permiten entender las interacciones y procesos de cambio de uso de suelo. Este trabajo se realiza en el municipio de Atoyatempan, donde las problemáticas ambientales requieren de softwares específicos para comprender que los ocasiona.

**Palabras Clave:** SIG, Cadenas de Márkov, Uso de suelo, Problemática ambiental.

### 1 Introducción

Los Sistemas de información geográfica son útiles para analizar los patrones de crecimiento urbano y modelar subsecuentemente el crecimiento de las ciudades. También en modelos basados en transiciones espaciales, destaca la utilización de métodos estocásticos como las Cadenas de Márkov.

La planificación territorial supone, principalmente, el establecimiento de los usos más apropiados para cada porción del territorio. La decisión sobre cuáles son estos usos depende, entre otros factores, de razones y criterios derivados de la conservación del ambiente natural y debe tener en cuenta tanto la vocación "intrínseca" de cada punto del territorio, determinada por su aptitud para cada uso o actividad, como el impacto potencial que tendrá sobre el medio ambiente la realización de una determinada actividad en ese punto del territorio.

La identificación de la problemática ambiental, previa a la planificación propiamente dicha, los SIG son muy útiles por su capacidad de analizar los datos geográficos y estadísticos que definen el problema a resolver con la planificación territorial y ambiental. Facilitan la utilización de los diferentes aspectos que inciden en la cuestión a comprender; permiten, mediante la superposición y otras formas de manipulación de la información, adaptar datos y observar algunas de las interrelaciones [1].

El presente trabajo se ha realizado en el municipio de Atoyatempan, Puebla, México. Los principales problemas que enfrenta este municipio en la actualidad son: la falta de planificación para el crecimiento urbano y el crecimiento en la demanda del agua para uso doméstico y la agricultura.

## 2 Métodos

Se llevó a cabo un análisis de cambio de uso de suelo y vegetación utilizando la cartografía de las series II, III, IV y VI del INEGI a una escala de 1:250,000; este análisis permitió calcular la matriz de transición desde el año 1990 hasta el año 2015 y sus correspondientes matrices de probabilidades de transición y patrones y tasas de cambio. La probabilidad de transición de cada clase de la matriz es proporcional a la superficie remanente de la misma clase entre el año de inicio y el año final de análisis [2].

$$P_{ij} = S_{ij}(\text{fecha inicial}) / S_j(\text{fecha final}) \quad (1)$$

Donde  $S_{ij}$  es la superficie del elemento “ij” de la matriz de transición de cobertura/uso de suelo en fecha inicial y  $S_j$  es la superficie de la clase de cobertura/uso de suelo “j” en fecha final. Se cumple que:

$$\sum P_{ij} = 1 \quad (2)$$

Se obtuvieron las superficies para cada tipo de vegetación y se calcularon las tasas de deforestación [3].

La distribución de probabilidad de transición de un estado a otro puede representarse por la matriz de transición  $P = (p_{ij})$  donde cada elemento de la posición (i, j) representa la probabilidad  $p_{ij}$ .

La cuantificación de los procesos de cambio cartografiados se realizó a través de matrices markovianas de transición y las matrices de probabilidades de cambio calculadas. A partir de las matrices de probabilidades de transición se construyeron los modelos de flujos que muestran la dinámica entre los principales tipos de vegetación y usos del suelo para las series.

La cadena se mueve de un estado a otro y la probabilidad  $p_{ij}$  de moverse del estado si al estado sj en un paso es llamado probabilidad de transición y puede escribirse como

$$\Pr(X_1 = sj \mid X_0 = si) \quad (3)$$

## 3 Resultados y Discusión

El municipio de Atoyatempan a lo largo de tres décadas ha cambiado en su estructura paisajista, esto queda demostrado en el análisis de cambio de uso de suelo presentado en este trabajo. Estos cambios obedecen principalmente a las actividades agropecuarias y comerciales que han marcado el rumbo de su historia agrícola.



## Innovaciones Tecnológicas en las Ciencias Computacionales

**Tabla 1.** Matriz de transición de tipo de suelo y vegetación en hectáreas (ha) de la serie II a la serie VI del INEGI escala 1:250000 correspondiente al periodo 1990 a 2015. Corresponden a las coberturas de: Agricultura de riego, agricultura de temporal, pastizal inducido, selva baja caducifolia y urbano. Pérdida anual en superficie (Pa) e Índices de variación multitemporal (La Estabilidad de Localización (EL) y la Estabilidad de Residencia (ER)) de tipo de suelo y vegetación en hectáreas (ha).

Tipo de suelo	Agricultura R.	Agricultura T.	Pastizal	Selva	Urbano	Total	Tasa de camb	Pa	EL	ER
Agrícola R.	838.25	25.66	88.53	5.34	18.87	976.65	2.99	39.91	95.91	85.82
Agrícola T.	941.99	484.43	9.97	0.47	5.71	1442.57	-4.76	-44.05	103.05	-64.12
Pastizal	8.47	0.00	25.18	0.00	0.00	33.64	6.40	4.29	87.26	70.64
Selva	2.26	7.46	0.00	62.76	0.00	72.48	-0.26	-0.19	100.26	-5.40
Urbano	23.88	0.00	0.00	0.00	129.78	153.66	0.02	0.03	99.98	0.46
Total	1814.84	517.55	123.68	68.57	154.37	2679.01				

**Tabla 2.** Matriz de Transición de Probabilidades del cambio de uso de suelo y vegetación en el municipio de Atoyatempan, Puebla para el período 1990- 2015.

Tipo de suelo	Agricultura R.	Agricultura T.	Pastizal	Selva	Urbano	Total
Agrícola R.	0.86	0.02	0.09	0.01	0.02	1.00
Agrícola T.	0.65	0.34	0.01	0.00	0.00	1.00
Pastizal	0.25	0.00	0.75	0.00	0.00	1.00
Selva	0.03	0.10	0.00	0.87	0.00	1.00
Urbano	0.16	0.00	0.00	0.00	0.84	1.00

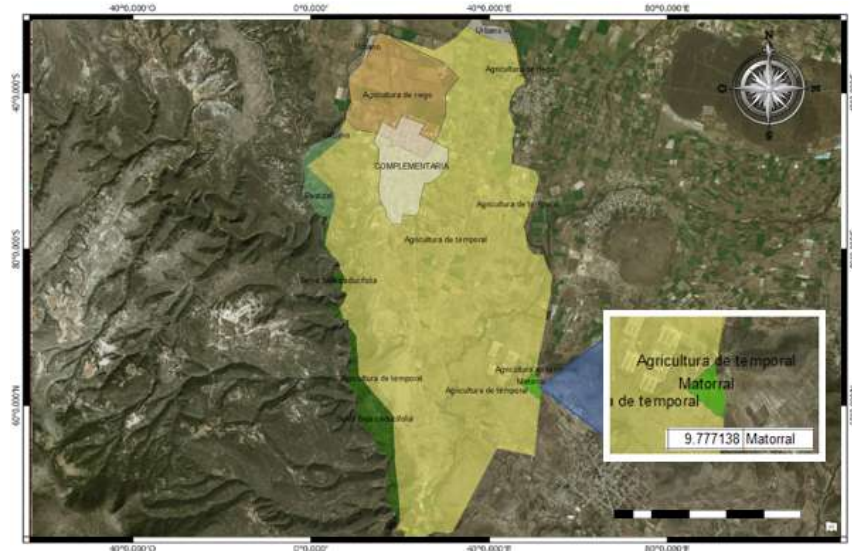
Los resultados mostraron que la agricultura de temporal se sustituyó por la agricultura de riego en este periodo, por lo que la disponibilidad de agua disminuyó en el municipio, debido también a que la zona tiene un balance hídrico no favorable con pendientes fuertes hacia la parte oeste, condiciones que no son idóneas para los asentamientos humanos. Mientras que el uso de suelo urbano ganó 0.71ha, la selva ha disminuido 3.91ha, en contraste con los pastizales inducidos, que aumentaron 90.04ha, por actividades antrópicas como la ganadería (Tabla 1).

El cambio de uso de suelo y vegetación se puede explicar en una sucesión de variables aleatorias que se desarrolla en función del tiempo. Considerando las categorías de uso y el periodo de tiempo para el análisis de la dinámica de cambio de uso de suelo. Las probabilidades se basan en periodo de tiempo del año inicial y el año final, considerándose como un proceso estocástico. (Tabla 2).

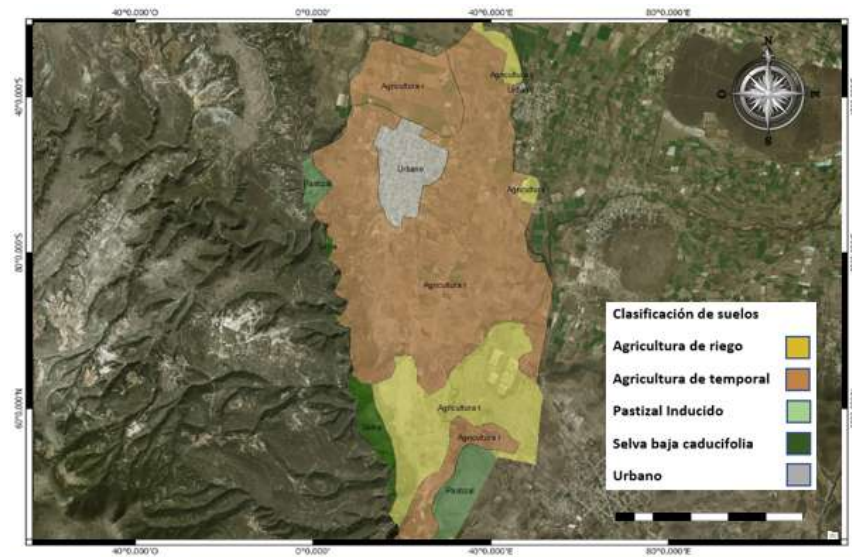
La reestructuración del sistema agroalimentario en el municipio, se relaciona con un nuevo modelo, modifica los procesos productivos en base a los requerimientos de consumo y de comercio de los municipios colindantes, siendo la actividad ganadera el principal agente transformador de la región en el año 1985 al año 2014 (Figura 1 y 2).

Las principales causas que amenazan la cobertura vegetal nativa son: la presión antropogénica sobre los recursos forestales, la práctica de actividades como la roza, tumba y quema, y la apertura de nuevos espacios para el establecimiento de cultivos en sistemas de lomeríos. Esto ha ocasionado un desequilibrio ambiental, poniendo en peligro la supervivencia de las poblaciones animales y vegetales que habitan este patrimonio ecológico. Sin embargo, los procesos asociados con la dinámica de los cambios ocurridos en la cobertura vegetal y uso del suelo, es uno de los principales factores que determina la permanencia, disminución y extinción de los ecosistemas.

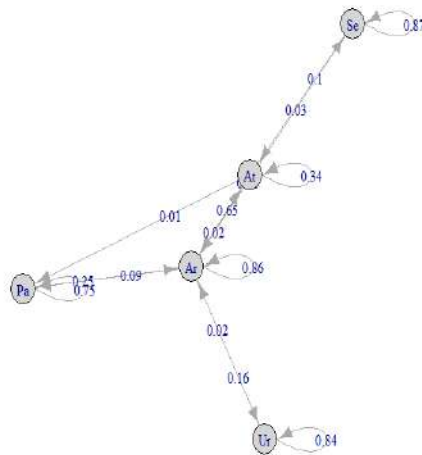
Los SIG son herramientas que permiten identificar y representar espacialmente las áreas más susceptibles a cambiar. Además, ayudan a comprender los procesos y la dinámica de cambio que experimentan las distintas coberturas terrestres y usos del suelo de un territorio en específico [4].



**Fig. 1.** Uso suelo y vegetación en hectáreas (ha) de la serie I del INEGI escala 1:250000 correspondiente al periodo 1985. Corresponden a las coberturas de: Agricultura de riego, agricultura de temporal, pastizal inducido, selva baja caducifolia, urbano y matorral xerófilo.



**Fig. 2.** Uso suelo y vegetación en hectáreas (ha) de la serie VI del INEGI escala 1:250000 correspondiente al periodo 2014. Corresponden a las coberturas de: Agricultura de riego, agricultura de temporal, pastizal inducido, selva baja caducifolia y urbano.



**Fig. 3.** Cadenas de Márkov. Corresponden a las coberturas de: Agricultura de riego, agricultura de temporal, pastizal inducido, selva baja caducifolia y urbano.

La cadena de Márkov establece las relaciones entre las diversas categorías o usos del suelo y las probabilidades de cambio de ese uso en porcentaje, lo cual es tradicionalmente modelado, sin tener en cuenta las variables explicativas y descriptivas, ya que se basa exclusivamente en el análisis de la dinámica interna del sistema (Figura 3). Muestra la dinámica de la transición probabilística de cada una de las coberturas en el sistema en porcentaje.

Los modelos estocásticos como las cadenas de Márkov permiten determinar un cambio de uso del suelo en 25 años, mostrando que la selva baja caducifolia pasa a agricultura de temporal y está a su vez se transforma a agricultura de riego, en zonas no idóneas para estas actividades antrópicas.

Los procesos de recuperación de zonas agrícolas incrementan la demanda de agua para la agricultura de riego y agricultura de temporal, lo que ocasiona disminución en los cuerpos de agua [5].

Esto es debido a la falta de planeación en los asentamientos humanos y políticas públicas, que requieren modelos social-ambientales para poder determinar un ordenamiento ecológico del territorio.

El crecimiento poblacional, las políticas públicas y las actividades de producción y de comercio, influyen de manera significativa en los procesos de transformación de las coberturas vegetales. Es de suma importancia revisar los planes de desarrollo municipal y estatal, junto con las políticas de gestión ambiental que ayuden a prevenir la disminución de coberturas de vegetación nativas.

La disponibilidad, calidad, cantidad y diversidad del recurso suelo se encuentra amenazada por un manejo inadecuado de ellos, cuyas consecuencias podrían afectar el desarrollo del municipio.

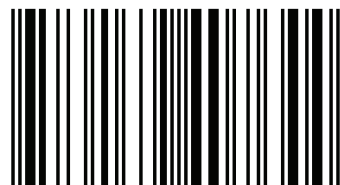
### Referencias

1. Bosque Sendra, J.: Técnicas de evaluación multicriterio y Sistemas de información geográfica en la ordenación del territorio, en Portugal-España. Ordenación territorial del suroeste comunitario. Coordinadores: A-J. Campesino Fernández y Carmen Velasco Bernardo. Universidad de Extremadura, Servicio de publicaciones, Cáceres, pp. 69-76 (1996)
2. Díaz Gallegos J.R.; Mas J.F.; Velásquez A.: Monitoreo de los patrones de deforestación en el Corredor Biológico Mesoamericano, México. *Interciencia* 33, 12: 882-890 (2008)
3. FAO.: Forest Resources Assessment 1990. Forestry Paper 130. FAO. Roma, Italia. 153 pp (1996)
4. Camacho Sanabria, J.M.; Juan Pérez, J.I.; Pineda Jaimes, N.B.; Cadena Vargas, E.G.; Bravo Peña, L.C.; & Sánchez López, M. (2015). Cambios de cobertura/uso del suelo en una porción de la Zona de Transición Mexicana de Montaña. *Madera y bosques*. Scielo. No 21, pp. 93-112. (2015).
5. López Vazquez, V.; Balderas Plata, M.; Chávez Mejía, M.; Juan Pérez, J., & Gutiérrez Cedillo, J.: Land Use Change and its Socio-Economic Implications in the Mazahua Area of the Mexican Highlands. *CIENCIA Ergo-Sum*, No. 22, pp. 136-144. (2015)

## Algoritmos, estrategias, modelos y sus aplicaciones

La serie Métodos cuantitativos aplicados, en su tomo 2, bajo el título Algoritmos, estrategias, modelos y sus aplicaciones, consta de 21 trabajos de investigación elaborados por investigadores de 26 universidades y centros de investigación de Cuba, México, España, Brasil, Colombia y Ecuador, agrupadas en RIDECA. El objetivo consiste en propiciar el desarrollo y aplicación de nuevos métodos cuantitativos para conocer las realidades locales, nacionales e internacionales con las que convivimos; por ello, se abordan temas tales como diseños secuenciales para ensayos clínicos, sistema informático para el análisis de los peligros naturales, análisis estadístico de datos turísticos y certificación de playas, modelo de regresión logística en rehabilitación, condicionantes de la alfabetización científica, relación entre variables, innovación y rendimiento en el sector comercial colombiano, diseño de criptosistemas, minería de textos, estadísticos de orden, simulación de dispersión de enfermedades, la motivación y el rendimiento escolar, evaluación diagnóstica de la gestión y, validez y confiabilidad en la calidad de cuestionarios, entre otros.

Carlos N. Bouza, Doctor en Ciencias, Universidad de La Habana, Cuba y Prof. de la Facultad de Matemática y Computación. Agustín Santiago, Doctor en Estadística e IO por la Ugr y Prof. en la Facultad de Matemáticas de la UAGro. México. José Félix García R., Doctor en Finanzas Públicas por la UV, México y Profesor de la Universidad Juárez de Tabasco.



978-613-9-40764-4

editorial académica española

ead  
editorial académica española



Carlos N. Bouza Herrera · Agustín Santiago M. · José Félix García R.

## Algoritmos, estrategias, modelos y sus aplicaciones

Serie Métodos Cuantitativos aplicados  
Tomo 2

Métodos Cuantitativos Aplicados Tomo 2

Bouza Herrera, Santiago M., García R.

**Carlos N. Bouza Herrera  
Agustín Santiago M.  
José Félix García R.**

**Algoritmos, estrategias, modelos y sus aplicaciones**

**Carlos N. Bouza Herrera  
Agustín Santiago M.  
José Félix García R.**

# **Algoritmos, estrategias, modelos y sus aplicaciones**

**Serie Métodos Cuantitativos aplicados Tomo 2**

**Editorial Académica Española**

## **Imprint**

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: [www.ingimage.com](http://www.ingimage.com)

Publisher:

Editorial Académica Española

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

**ISBN: 978-613-9-40764-4**

Copyright © Carlos N. Bouza Herrera, Agustín Santiago M., José Félix García R.

Copyright © 2019 International Book Market Service Ltd., member of OmniScriptum Publishing Group



Algoritmos, estrategias y  
modelos en las aplicaciones de  
métodos cuantitativos

Tomo 2

**Serie Métodos cuantitativos aplicados**

## Índice de capítulos

CAPITULO	TÍTULO	AUTORES	PÁGINAS
1	REVISIÓN DE DISEÑOS SECUENCIALES PARA ENSAYO CLÍNICOS FASE II	Carmen Viada González, Martha Fors López, Maytec Robaina García, Mabel Álvarez Cardona, Aliuska Frias Blanco, Lázara García Fernández, Bárbara Wilkinson Brito, Liana Martínez Pérez, Mayelin Troche de la Concepción	1-16
2	METODOLOGÍA Y SISTEMAS PARA META-ANÁLISIS DE ENSAYOS CLÍNICOS	Carmen Viada Gonzalez, Martha Fors López, Javier Ballesteros, Maytec Robaina García, Mabel Álvarez Cardona, Aliuska Frias Blanco, Lázara García Fernández, Bárbara Wilkinson Brito, Liana Martínez Pérez, Mayelin Troche de la Concepción.	17-35
3	SISTEMA INFORMÁTICO PARA EL ANÁLISIS DE LOS PELIGROS NATURALES	Amaury Machín Armenteros, Yasser Vázquez Alfonso, Elaine Pérez Monteagudo, Efraín Velastegui López	37-48
4	ANÁLISIS ESTADÍSTICO DE LA OCUPACIÓN HOTELERA EN ACAPULCO Y LA CERTIFICACIÓN DE PLAYAS PARA LA COMPETITIVIDAD TURÍSTICA	Miguel Ángel Cruz Vicente, Darbelio Agatón Lorenzo, Arturo de León Chapa	49-69
5	MODELO DE REGRESIÓN LOGÍSTICA EN PACIENTES CON GONARTROSIS QUE REALIZAN TRATAMIENTO REHABILITADOR	G. Bayolo Soler, V. Sistachs Vega, Y. Coronados Valladares	71-88
6	GÉNERO Y ENTORNO DE NACIMIENTO EN EL NIVEL DE ALFABETIZACIÓN CIENTÍFICA DE ESTUDIANTES DE UNA UNIVERSIDAD PÚBLICA MEXICANA	Carlos David Zetina Pérez, Deneb Elí Magaña Medina, Pedro Antonio Sánchez Escobedo, José Félix García Rodríguez	89-102
7	RELACIÓN ENTRE EL RECONOCIMIENTO DE MARCA Y EL MODELO DE AJUSTE A LA TAREA TECNOLÓGICA (TTF)	Andrés Machado Daza	103-117
8	MIPYMES, INNOVACIÓN Y RENDIMIENTO EN EL SECTOR INDUSTRIAL COLOMBIANO	Paula Viviana Robayo-Acuña, Henry Laverde Rojas	119-136

<b>9</b>	CRIPOTOSISTEMA BASADO EN AUTÓMATAS CELULARES Y FUNCIONES CAÓTICAS	Luis René Marcial Castillo, Marcela Rivera Martínez, Lourdes Sandoval Solís, Juan Jesús Luna Padilla	137-149
<b>10</b>	MINERÍA DE TEXTO APLICADA AL MARKETING VERDE (1985-2018): UN ANÁLISIS BASADO EN SCOPUS	Jenny-Paola Lis-Gutiérrez, Mercedes Gaitán-Angulo, Cristian Beltrán-Oicatal, Melissa Lis-Gutiérrez, Andreina Moros, Carolina Henao	151-178
<b>11</b>	GENERALIDADES SOBRE LOS ESTADÍSTICOS DE ORDEN	Carlos N. Bouza-Herrera	179-200
<b>12</b>	LA TEORÍA DE EPIDEMIAS Y LA SIMULACIÓN DE LA DISPERSIÓN DE ENFERMEDADES RESPIRATORIAS	José Betancourt Bethencourt, Juan Jesús Llambias Peláez, Luis Acao Francoise.	201-209
<b>13</b>	INFLUENCIA DE LA MOTIVACIÓN EN EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIDAD ACADÉMICA DE ECONOMÍA DE LA UAGRO, PERIODO FEBRERO-JULIO 2013	Beatriz Sánchez-Estrada, Agustín Santiago-Moreno, Alfredo Pineda-Gómez, Susana Lasso-Abarca	211-236
<b>14</b>	VALIDEZ Y CONFIABILIDAD DE UN CUESTIONARIO SOBRE LA CALIDAD DE VIDA EN MUJERES CON LUPUS	Gladys Linares Fleites, María de Lourdes Sandoval Solís, María Lilia Cedillo Ramírez y Socorro Méndez Martínez	237-252
<b>15</b>	EVALUACIÓN DIAGNÓSTICA DE LA GESTIÓN DE LOS CONSEJOS TÉCNICOS ESCOLARES DEL NIVEL DE EDUCACIÓN BÁSICA EN ACAPULCO	Cinthia Corey Bustamante Aguirre, Tonantzin Martínez Bello, Elizabeth Ávila Cortes, Agustín Santiago Moreno.	253-263
<b>16</b>	CRITERIOS DE CALIDAD DEL CUESTIONARIO DE EVALUACIÓN DE LOS CTE DEL NIVEL BÁSICO DE EDUCACIÓN EN ACAPULCO	Elizabeth Ávila Cortes, Tonantzin Martínez Bello, Cinthia Corey Bustamante Aguirre y José Maclovio Sautto-Vallejo	265-274
<b>17</b>	PROPUESTA DE ANÁLISIS Y SELECCIÓN DE INICIATIVAS DE INNOVACIÓN SOCIAL	Jenny-Paola Lis-Gutiérrez, Jenny Cubillos-Díaz, Mercedes Gaitán-Angulo, Manuel Ignacio Balaguera, Carlos Bouza	275-291
<b>18</b>	TASAS DE CRECIMIENTO E ÍNDICES DE COMPETITIVIDAD DEL COMERCIO EXTERIOR DEL MANGO MEXICANO	Ignacio Caamal Cauch, Verna Gricel Pat Fernández, José Félix García Rodríguez, Zulia Helena Caamal Pat	293-310

## Afiliación

<b>Nombre</b>	<b>Afiliación</b>
Agustín Santiago-Moreno	Universidad Autónoma de Guerrero, Facultad de Matemáticas-Acapulco.
Aida Armenta Ramírez	Universidad Juárez Autónoma de Tabasco, México
Alfredo Pineda-Gómez	Universidad Autónoma de Guerrero, Facultad de Ciencias de la Educación-Acapulco.
Aliuska Frías Blanco	Centro de Inmunología Molecular, La Habana, Cuba
Amaury Machín Armenteros	Facultad de Ciencias Técnicas, Universidad Agraria de la Habana, Cuba
Andreina Moros	Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
Andrés Machado Daza	Fundación Universitaria Konrad Lorenz
Arturo de León Chapa	Universidad Autónoma de Guerrero
Bárbara Wilkinson Brito	Centro de Inmunología Molecular, La Habana, Cuba
Beatriz Sánchez-Estrada	Universidad Autónoma de Guerrero, Facultad de Ciencias de la Educación-Acapulco.
Boris Espinoza	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)
Carlos Bouza Herrera	Facultad de Matemática y Computación, Universidad de la Habana
Carlos David Zetina Pérez	Universidad Juárez Autónoma de Tabasco, México
Carmen Viada González	Centro de Inmunología Molecular, La Habana, Cuba
Carolina Henao	Universidad de Ciencias Aplicadas, Bogotá, Colombia
Cynthia Corey Bustamante Aguirre	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
Cristian Beltrán-Oicata	Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
Darbelio Agatón Lorenzo	Universidad Autónoma de Guerrero
Deneb Elí Magaña Medina	Universidad Juárez Autónoma de Tabasco, México
Diego Andrés Guevara-Fletcher	Facultad de Posgrados ESAP, Colombia.
Diego Ochoa-Jiménez	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)
Efraín Velastegui López	Universidad de Cotopaxi, Ecuador.
Elaine Pérez Monteagudo	Dirección Provincial de Planificación Física, Mayabeque, Cuba.
Elizabeth Ávila Cortes	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
G. Bayolo Soler	Facultad de Matemática y Computación, Universidad de La Habana, Cuba
Gladys Linares Fleites	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla

Henry Laverde Rojas	CIEN, Escuela de Negocios, Fundación Universitaria Konrad Lorenz Bogotá D.C., Colombia
Ignacio Caamal Cauich	Universidad Autónoma Chapingo, México.
Javier Ballesteros	Universidad del País Vasco, Bilbao, España
Jenny Cubillos-Díaz	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
Jenny-Paola Lis-Gutiérrez	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
José Betancourt Bethencourt	Universidad de Ciencias Médicas Carlos J. Finlay/Centro de Inmunología y Productos Biológicos/Camagüey, Cuba.
José Félix García Rodríguez	Universidad Juárez Autónoma de Tabasco, México
José Maclovio Sautto-Vallejo	Universidad Autónoma de Guerrero, Campus Costa Chica.
Juan Jesús Llambias Peláez	Dirección Provincial de Salud. Departamento de Vigilancia en Salud /Camagüey, Cuba
Juan Jesús Luna Padilla	Benemérita Universidad Autónoma de Puebla, México
Lázara García Fernández	Centro de Inmunología Molecular, La Habana, Cuba
Liana Martínez Pérez	Centro de Inmunología Molecular, La Habana, Cuba
Luis Acao Francoise	Universidad de Ciencias Médicas Carlos J. Finlay/Departamento de Salud, Cuba.
Luis René Marcial Castillo,	Benemérita Universidad Autónoma de Puebla, México
Mabel Álvarez Cardona	Centro de Inmunología Molecular, La Habana, Cuba
Manuel Ignacio Balaguera	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Facultad de Matemáticas e Ingeniería (CIMI)
Marcela Rivera Martínez	Benemérita Universidad Autónoma de Puebla, México
María de Lourdes Sandoval Solís	Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla
María Lilia Cedillo Ramírez	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla
Marisol de la Cruz León	Universidad Juárez Autónoma de Tabasco, México
Martha Fors López	Universidad de Las Américas, Quito, Ecuador.
Mayelin Troche de la Concepción	Centro de Inmunología Molecular, La Habana, Cuba
Maytee Robaina García	Universidad de Las Américas, Quito, Ecuador.
Maytee Robaina García	Centro Nacional Coordinador de Ensayos Clínicos, La Habana, Cuba
Melissa Lis-Gutiérrez	Corporación Universitaria Minuto de Dios
Mercedes Gaitán-Angulo	Fundación Universitaria Konrad Lorenz, Centro de Investigaciones de la Escuela de Negocios (CIEN)
Miguel Ángel Cruz Vicente	Universidad Autónoma de Guerrero
Pascual García Macías	Instituto Nacional de Estadísticas y Censos de Ecuador y Universidad Técnica Particular de Loja (UTPL)

Paula Viviana Robayo-Acuña	CIEN, Escuela de Negocios, Fundación Universitaria Konrad Lorenz Bogotá D.C., Colombia
Pedro Antonio Sánchez Escobedo	Universidad Autónoma de Yucatán, México
Rilton G. Goncalvez Primo	Universidad Federal de Bahía, Brasil
Socorro Méndez Martínez	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla
Susana Lasso-Abarca	Universidad Autónoma de Guerrero, Coordinación zona sur-Acapulco.
Tonantzin Martínez Bello	Centro de Investigación y Desarrollo Educativo de Acapulco (CIDEA)
Verna Grisel Pat Fernández	Universidad Autónoma Chapingo, México.
Vivian Sistachs Vega	Facultad de Matemática y Computación, Universidad de La Habana, Cuba
Y. Coronados Valladares	Hospital Julio Díaz, Cuba.
Yasser Vázquez Alfonso	Facultad de Turismo, Universidad de la Habana, Cuba
Zulia Helena Caamal Pat	Colegio de Posgraduados, Texcoco, México.

# Capítulo 14

## VALIDEZ Y CONFIABILIDAD DE UN CUESTIONARIO SOBRE LA CALIDAD DE VIDA EN MUJERES CON LUPUS

Gladys Linares Fleites<sup>1</sup>, María de Lourdes Sandoval Solís<sup>2</sup>,  
María Lilia Cedillo Ramírez<sup>1</sup> y Socorro Méndez Martínez<sup>1</sup>

<sup>1</sup>Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla.

<sup>4</sup> Sur No.104, Colonia Centro, C.P. 72000, Puebla, México. [gladyslinares1@yahoo.es](mailto:gladyslinares1@yahoo.es)

<sup>2</sup> Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla. <sup>4</sup> Sur No.104, Colonia Centro, C.P. 72000, Puebla, México.

[malosa1753@gmail.com](mailto:malosa1753@gmail.com)

### RESUMEN

El propósito de este estudio es analizar la validez y la confiabilidad del Cuestionario SF-36 que evalúa la calidad de vida relacionada con la salud (CVRS), en mujeres con lupus eritematoso sistémico (LES) del estado de Puebla, México. Se encuestó a 92 mujeres mayores a 18 años, a quienes se les aplicó, de manera individual el cuestionario de Salud SF-36 compuesto por 36 preguntas (ítems) que valoran los estados tanto positivos como negativos de la salud. El cuestionario final cubre 8 escalas, que representan los conceptos de salud empleados con más frecuencia en los principales cuestionarios de salud, así como, los aspectos más relacionados con la enfermedad y el tratamiento. Los 36 ítems del instrumento cubren las siguientes escalas: Función física, Rol físico, Dolor corporal, Salud general, Vitalidad, Función social, Rol emocional y Salud mental. La confiabilidad del instrumento de 36 ítems presentó un alfa de Cronbach de 0.9329, lo que confiere a la escala una consistencia interna muy alta o elevada. El citado cuestionario, en su versión en español, es un instrumento confiable y válido para mujeres con LES del estado de Puebla, México.

**PALABRAS CLAVE:** Calidad de Vida relacionada con la salud (CVRS), Lupus eritematoso sistémico (LES).

### ABSTRACT

The purpose of this study is to analyze the validity and reliability of the SF-36 Questionnaire that assesses the quality of life related to health (HRQOL) in women with systemic lupus erythematosus (SLE) in the state of Puebla, Mexico. A total of 92 women older than 18 years were surveyed, to whom the SF-36 Health questionnaire was applied individually, consisting of 36 questions (items) that assess both positive and negative

health states. The final questionnaire covers 8 scales, which represent the health concepts most frequently used in the main health questionnaires, as well as the aspects most related to the disease and treatment. The 36 items of the instrument cover the following scales: physical function, physical role, body pain, general health, vitality, social function, emotional role and mental health. The reliability of the 36-item instrument showed a Cronbach's alpha of 0.9329, which gives the scale a very high or high internal consistency. The aforementioned questionnaire, in its Spanish version, is a reliable and valid instrument for women with SLE in the state of Puebla, Mexico.

**KEY WORDS:** Health related Quality of Life (HRQOL), Systemic Lupus Erythematosus (SLE),

## 1. INTRODUCCIÓN

El Lupus Eritematoso Sistémico (LES) es una enfermedad inflamatoria crónica de causa desconocida, aunque la herencia, el entorno y los cambios hormonales juegan un papel importante que afecta a todas las edades, pero con mayor frecuencia a adultos entre los 18 y 50 años con predominio del sexo femenino, en una proporción de un hombre por cada 10 a 12 mujeres.

La prevalencia del LES varía en los distintos grupos de población, oscilando entre 300 y 400 pacientes por cada 100.000 habitantes. Es más común en ciertos grupos étnicos, especialmente los afroamericanos. Alrededor del 70% de las personas que sufren lupus presentan la forma sistémica o LES (Molina *et al.*, 2005).

LES es un desorden del sistema inmunitario, que normalmente funciona protegiendo el organismo contra infecciones y virus. En LES, dicho sistema inmunitario es hiperactivo y se producen importantes cantidades de anticuerpos anormales que reaccionan con los tejidos del propio paciente, afecta a todos los órganos y sistemas, primordialmente las articulaciones, músculos y piel, pero también, aunque en menor grado a estructuras internas como los pulmones, corazón, riñones y cerebro (Camargo *et al.*, 2005). La inflamación se refiere a una reacción que provoca dolor, calor, enrojecimiento e hinchazón. En casi la mitad de las personas con LES se ven afectados los órganos internos principales, especialmente los riñones. En la otra mitad de los pacientes con LES están afectadas principalmente la piel y las articulaciones (Gómez *et al.*, 2005).

Aparte de los factores fisiológicos antes mencionados, LES puede afectar de diversas formas la Calidad de Vida Relacionada con la Salud (CVRS) de quienes padecen esta enfermedad, más concretamente a nivel de la actividad física, sexual, mental y social. Desde el punto de vista subjetivo, la calidad de vida relacionada con la salud es la valoración que realiza una persona, de acuerdo con sus propios criterios del estado físico, emocional y social en que se encuentra en un momento dado, y refleja el grado de satisfacción con una situación personal a nivel: fisiológico (síntomatología general, discapacidad funcional, situación analítica, sueño, respuesta sexual), emocional (sentimientos de tristeza, miedo, inseguridad, frustración) y social (situación laboral o



escolar, interacciones sociales en general, relaciones familiares, amistades, nivel económico, participación en la comunidad, actividades de ocio, entre otras) (Schwartzmann *et al.*, 2003).

Es necesario contar con un instrumento de medición para valorar la CVRS. Existen muchos requisitos que deben llenar los instrumentos de medición, ya que, si no los llenan, los datos recolectados tendrán limitaciones importantes. Entre estos requisitos o cualidades están la confiabilidad y la validez del instrumento.

La confiabilidad se refiere al grado en que su aplicación repetida al mismo sujeto, produce resultados iguales (consistentes y coherentes) (Hernández Sampieri *et al.*, 2006). La confiabilidad indica el grado de congruencia con que mide el atributo para el que está diseñado. También se señala que el instrumento es confiable si sus mediciones reflejan exactamente los valores verdaderos del atributo que se investiga. (Polit *et al.*, 2000). Entre las recomendaciones que se brindan para elevar la confiabilidad pueden citarse: aplicar las reglas generales de elaboración de instrumentos, de tal forma que se eliminen los errores de medición (por ejemplo, preguntas ambiguas), aumentar el número de preguntas sobre determinado tema o aspecto, elaborar instrucciones claras que orienten el llenado o utilización de los instrumentos y aplicar los instrumentos o realizar las mediciones en condiciones similares.

La validez constituye el segundo criterio de importancia para evaluar la idoneidad de un instrumento cuantitativo; denota el grado en que el instrumento mide lo que se supone que debe medir. Igual que la confiabilidad, la validez comprende diferentes aspectos y técnicas de evaluación, pero, a diferencia de la confiabilidad, es en extremo difícil determinar la validez de un instrumento.

La confiabilidad y la validez de un instrumento no son cualidades completamente independientes. Un dispositivo de medición que no sea confiable no puede ser válido, pues si es errático, incongruente e inexacto tampoco medirá con validez el atributo en cuestión.

Uno de los instrumentos genéricos más comunes para evaluar la calidad de vida en pacientes con enfermedades crónicas es el Cuestionario de Calidad de Vida SF-36 ya que es útil, fácil de contestar por los pacientes y permite obtener información sobre funcionalidad física, mental y percepción de la salud (Gemma, *et al.*, 2004). El cuestionario SF-36 ha probado su utilidad para evaluar la Calidad de Vida en pacientes con Lupus Eritematoso Sistémico (LES) en otros países, pero no en pacientes con LES en Puebla, México (Méndez, 2017).

El objetivo de este capítulo es utilizar las técnicas brindadas por la Estadística para evaluar la validez y confiabilidad del cuestionario SF-36, que se utiliza para explorar la Calidad de Vida Relacionada con la Salud (CVRS), en paciente con LES del Hospital Universitarios de la ciudad de Puebla, México.

## 1.1 INSTRUMENTO SF-36 PARA EVALUAR CALIDAD DE VIDA EN PACIENTES CON LUPUS ERITEMATOSO SISTÉMICO (LES)

La Calidad de Vida relacionada con la salud en pacientes con LES se ha venido evaluando a través de instrumentos tanto genéricos como específicos para enfermedades reumáticas y específicos para el mismo LES. En una revisión de literatura sobre instrumentos de CVRS utilizados en estudios con pacientes con LES (Thumboo *et al.*, 2007) se ha encontrado que las áreas de la CVRS más afectadas en estos pacientes han sido la percepción de salud, la fatiga, el dolor corporal, la actividad funcional a nivel laboral, la autonomía, las relaciones sociales, familiares y la desesperanza aprendida (respecto a la imprevisibilidad de la enfermedad del LES).

Como se planteó anteriormente, uno de los instrumentos genéricos más comunes para evaluar la calidad de vida en pacientes con enfermedades crónicas es el Cuestionario de Calidad de Vida SF-36 debido a su facilidad para obtener información sobre funcionalidad física, mental y percepción de la salud (Gemma, *et al.*, 2004).

El SF-36 consta de 36 preguntas (ítems) diseñadas para evaluar el estado físico y mental a través de varios aspectos distintos (vitalidad, molestias físicas, funcionalidad social y física, limitaciones por problemas emocionales, salud mental, percepción general de la salud y cambios en salud) en una escala de 0 a 100. Las puntuaciones de dominio están en una escala de 0 (peor) a 100 (mejor).

El cuestionario de salud SF-36 fue desarrollado a principios de los noventa, en Estados Unidos, para su uso en el Estudio de los Resultados Médicos (Medical Outcomes Study, MOS). Es una escala genérica que proporciona un perfil del estado de salud y es aplicable tanto a los pacientes como a la población general. Ha resultado útil para evaluar la calidad de vida relacionada con la salud (CVRS) en la población general y en subgrupos específicos, comparar la carga de muy diversas enfermedades, detectar los beneficios en la salud producidos por un amplio rango de tratamientos diferentes y valorar el estado de salud de pacientes individuales. Sus buenas propiedades psicométricas, que han sido evaluadas en más de 400 artículos, y la multitud de estudios ya realizados, que permiten la comparación de resultados, lo convierten en uno de los instrumentos con mayor potencial en el campo de la CVRS. Algunas de estas razones impulsaron a realizar la adaptación para su uso en España.

El SF-36 evalúa aspectos de la calidad de vida en poblaciones adultas (mayores de 16 años). El producto de su aplicación es la construcción de ocho conceptos o escalas de salud resultado del promedio de la suma de las preguntas contenidas en el cuestionario. Estos conceptos son: a) Función Física (FF), b) rol físico (RF), c) dolor corporal (DC), d) salud general (SG), e) vitalidad (VT), f) función social (FS), g) rol emocional (RE) y h) salud mental (SM). Además de los ocho conceptos de salud, el SF-36 incluye el concepto general de cambios en la percepción del estado de salud actual y en la del año anterior. La respuesta a esta pregunta describe la transición de la percepción respecto al mejoramiento o empeoramiento del estado de salud.

En la tabla 1 se presenta una descripción de las escalas de salud y sus respectivas interpretaciones de acuerdo con resultados bajos o altos por cada escala.

## 2. ESTUDIO DE LA CVRS EN UNA MUESTRA DE PACIENTES DEL HOSPITAL UNIVERSITARIO DE PUEBLA, MEXICO CON EL SF-36.

La población de estudio estuvo conformada por 92 pacientes de sexo femenino; todos los pacientes cumplían los criterios de clasificación de la American College of Rheumatology (ACR) para LES y acudían frecuentemente a la Clínica Ambulatoria, Hospital General Regional No. 36 del IMSS, en Puebla, México.

Al inicio del estudio, se registraron variables demográficas y clínicas. La actividad de la enfermedad se evaluó mediante pruebas y exámenes de laboratorio físicos. Se obtuvo consentimiento escrito de todos los pacientes y el comité de ética local aprobó el estudio (Méndez, 2017). El muestreo fue por conveniencia, se encuestaron a 92 mujeres mayores a 18 años, a quienes se les aplicó, de manera individual el cuestionario de Salud SF-36 compuesto por 36 preguntas (ítems) que valoran los estados tanto positivos como negativos de la Calidad de Vida Relacionada con la Salud (CVRS), en mujeres con el lupus eritematoso sistémico (LES) del estado de Puebla (México).

Con la aplicación del cuestionario se obtuvo una tabla de 92 individuos (filas) por 36 ítems (columnas) y con esta información se procedió a analizar la validez y confiabilidad del cuestionario SF-36 (Santos, 2017).

Tabla 1. Contenido de las escalas del SF-36			
Significado de las puntuaciones de 0 a 100			
Dimensión	N.º de ítems	-Peor- puntuación (0)	-Mejor- puntuación (100)
Función física	10	Muy limitado para llevar a cabo todas las actividades físicas, incluido bañarse o ducharse, debido a la salud	Lleva a cabo todo tipo de actividades físicas incluidas las más vigorosas sin ninguna limitación debido a la salud
Rol físico	4	Problemas con el trabajo u otras actividades diarias debido a la salud física	Ningún problema con el trabajo u otras actividades diarias debido a la salud física
Dolor corporal	2	Dolor muy intenso y extremadamente limitante	Ningún dolor ni limitaciones debidas a él
Salud general	5	Evalúa como mala la propia salud y cree posible que empeore	Evalúa la propia salud como excelente
Vitalidad	4	Se siente cansado y exhausto todo el tiempo	Se siente muy dinámico y lleno de energía todo el tiempo
Función social	2	Interferencia extrema y muy frecuente con las actividades sociales normales, debido a problemas físicos o emocionales	Lleva a cabo actividades sociales normales sin ninguna interferencia debido a problemas físicos o emocionales
Rol emocional	3	Problemas con el trabajo y otras actividades diarias debido a problemas emocionales	Ningún problema con el trabajo y otras actividades diarias debido a problemas emocionales
Salud mental	5	Sentimiento de angustia y depresión durante todo el tiempo	Sentimiento de felicidad, tranquilidad y calma durante todo el tiempo
Ítem de Transición de salud	1	Cree que su salud es mucho peor ahora que hace 1 año	Cree que su salud general es mucho mejor ahora que hace 1 año

Fuente: Gemma, *et al.*, 2004.

## 2.1 VALIDEZ DEL CUESTIONARIO SF-36

La validez de un instrumento, comprende diferentes aspectos y técnicas de evaluación. En este caso nos referiremos a la llamada “validez de constructo” que intenta responder la pregunta ¿hasta dónde un instrumento mide realmente una determinada característica de las personas y con cuánta eficiencia lo hace? Entre los procedimientos o técnicas estadísticas utilizadas para este tipo de validez destaca en mayor medida el Análisis Factorial (AF). En general, podemos decir que esta es la técnica por excelencia utilizada para analizar la validez de un instrumento (cuestionario o constructo).

El análisis factorial (AF) es una técnica utilizada para descubrir agrupaciones de variables de tal forma que las variables de cada grupo están altamente correlacionadas y los grupos están relativamente incorrelacionados. De este modo se consigue reducir un número de variables intercorrelacionadas a un número inferior de factores no correlacionados, que permiten explicar la mayor parte de variabilidad de cada una de las variables (Linares, 2006).

El modelo matemático del AF supone que cada una de las  $p$  variables observadas es función de un número  $m$  factores comunes ( $m < p$ ) más un factor específico o único. Tanto los factores comunes como los específicos no son observables y su determinación e interpretación es el resultado del AF. La forma matricial del modelo es

$$X = \Lambda F + U$$

donde

X: es el vector de las variables observadas.

$\Lambda$ : es la matriz factorial. (Recoge las cargas factoriales o saturaciones).

F: es el vector de factores comunes.

U: es el vector de factores únicos o factores específicos.

Todos los resultados, que se exponen a continuación, se obtuvieron con el paquete R-Commander, software libre en lenguaje R.

### 3.1.1 ANÁLISIS PRELIMINARES

Antes de aplicar el Análisis Factorial Exploratorio (AFE), se debe comprobar si éste es necesario, es decir, si la correlación entre las variables analizadas es lo suficientemente grande como para justificar la factorización de la matriz de coeficientes de correlación. Entre los muchos criterios expuestos en la literatura estadística (Santos, 2017) para hacer la comprobación se destacan:

- Que el índice de Kaiser-Meyer-Olkin (KMO) debe ser mayor a 0.60.

Este índice puede pensarse como un coeficiente de correlación parcial que mide la correlación existente entre dos variables una vez que se han descontado los efectos

lineales de otras variables. En un modelo factorial se pueden interpretar esos efectos de otras variables como los correspondientes a factores comunes.

- Que el resultado del test de esfericidad de Bartlett sea significativo.

El objetivo es comprobar que la matriz de correlaciones  $R$  es significativamente distinta de la matriz identidad  $I$ , o sea, se prueban las hipótesis  $H_0: R = I$  vs  $H_1: R \neq I$ . Si la matriz de correlaciones fuera la matriz identidad no habría correlación entre variables y no tendría sentido llevar a cabo un AF. El test de esfericidad de Bartlett nos permite hacer esta comparación.

En este estudio la comprobación se realizó mediante los criterios anteriores. Los resultados señalaron que la muestra es adecuada para la realización del AFE. El índice de KMO es de 0.82 lo que indica que se puede realizar el AFE. Por otro lado, la prueba de Bartlett es significativa, ya que el valor de  $p$  empírico es menor que el nivel de significación ( $\alpha = 0.05$ ) y, consecuentemente, se rechaza la hipótesis nula de que la matriz de coeficientes de correlación es igual a la matriz identidad.

Para determinar el número de factores a extraer se realizó un Análisis de Componentes Principales (ACP). Este método calcula tantas componentes principales como variables originales y, así, se reproduce la totalidad de la varianza. Las siguientes salidas en R-Commander (Tabla 2) muestran: (a) la varianza asociada a cada factor (el cuadrado de las desviaciones estándar) que viene expresada por el valor propio o raíz característica de la matriz de coeficientes de correlación. (b) la proporción de la varianza que explica cada componente principal; su suma es igual a 1 y, (c) proporción acumulada, que se calcula sumándolas progresivamente. Por ejemplo: puede observarse que las dos primeras componentes agrupan un 42.9% de la variación total, o lo que es lo mismo, hay un 57.1% de variación que no se explica. Los resultados mostraron que los primeros 8 factores o componentes principales, explicaban aproximadamente el 70% de la variancia total de los datos.

El gráfico de sedimentación (Figura 1) muestra la importancia de esos factores, pero destaca como el primer factor explica más de un tercio de la variabilidad total (34 %).

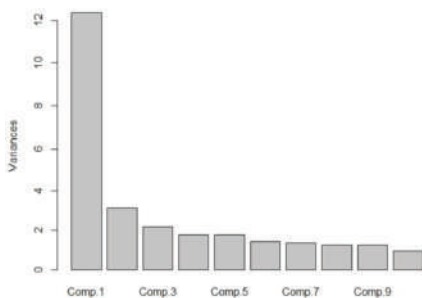


Figura 1: Gráfico de sedimentación.

Tabla 2: Salida de R-Commander sobre varianza total explicada del ACP

Importance of components:						
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	3.5174784	1.75919394	1.46618909	1.33036999	1.32316540	1.18682541
Proportion of Variance	0.3436849	0.08596565	0.05971418	0.04916345	0.04863241	0.03912652
Cumulative Proportion	0.3436849	0.42965050	0.48936468	0.53852813	0.58716054	0.62628705
	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
Standard deviation	1.16149383	1.12323919	1.1188882	0.97130315	0.90680453	0.87813254
Proportion of Variance	0.03747411	0.03504629	0.0347753	0.02620638	0.02284151	0.02141991
Cumulative Proportion	0.66376116	0.69880745	0.7335827	0.75978913	0.78263064	0.80405055
	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17	Comp.18
Standard deviation	0.82839633	0.80845687	0.7479307	0.71242162	0.69122278	0.66303788
Proportion of Variance	0.01906224	0.01815563	0.0155389	0.01409846	0.01327191	0.01221165
Cumulative Proportion	0.82311279	0.84126841	0.8568073	0.87090577	0.88417769	0.89638933
	Comp.19	Comp.20	Comp.21	Comp.22	Comp.23	Comp.24
Standard deviation	0.61443594	0.61085015	0.583060871	0.559647946	0.553629777	0.511406879
Proportion of Variance	0.01048699	0.01036494	0.009443333	0.008700162	0.008514054	0.007264917
Cumulative Proportion	0.90687632	0.91724126	0.926684594	0.935384755	0.943898809	0.951163726
	Comp.25	Comp.26	Comp.27	Comp.28	Comp.29	
Standard deviation	0.493915988	0.491651464	0.459529467	0.427059292	0.394103454	
Proportion of Variance	0.006776472	0.006714477	0.005865759	0.005066101	0.004314376	
Cumulative Proportion	0.957940198	0.964654675	0.970520434	0.975586535	0.979900911	
	Comp.30	Comp.31	Comp.32	Comp.33	Comp.34	
Standard deviation	0.386961293	0.361025951	0.340275198	0.328190096	0.303696786	
Proportion of Variance	0.004159418	0.003620548	0.003216311	0.002991909	0.002561993	
Cumulative Proportion	0.984060329	0.987680877	0.990897188	0.993889098	0.996451090	
	Comp.35	Comp.36				
Standard deviation	0.269280169	0.235050917				
Proportion of Variance	0.002014217	0.001534693				
Cumulative Proportion	0.998465307	1.000000000				

### 3.1.2 ANÁLISIS FACTORIAL EXPLORATORIO (AFE)

El objetivo del AFE es obtener la estructura factorial más simple desde el punto de vista de su interpretación. En el gráfico de sedimentación, se observó claramente que el primer componente podría ser extraído, por lo que en primer lugar consideramos el análisis monofactorial para analizar si todos los indicadores saturan en un único factor. Los resultados se muestran en la Tabla 3.

A continuación, se aplicó el modelo de ocho factores utilizando también el paquete R-Commander; se utilizó el método de máxima verosimilitud (ML, por sus siglas en inglés) para extraer los factores y la rotación oblicua (promax). En la Tabla 4 se muestra la salida correspondiente donde se proporcionan las saturaciones de cada ítem en cada factor, y como se trata de coeficientes de correlación se interpretan de la misma manera.

Los ítems importantes de cada factor son aquellos que tienen el peso mayor en un factor y mucho menor en los demás. Algunos autores indican que en ningún caso debe ser muy inferior a 0.40, sin embargo, otros autores consideran que una correlación ítem-factor de 0.35, e inclusive de 0.30, es suficiente para asumir la relación ítem factor e interpretarlo con claridad.

A continuación, se describe la interpretación de cada uno de los factores y los ítems, según la salida mencionada.

*Factor 1: Función Física (FF).* Grado en que la salud limita las actividades físicas tales como el autocuidado, caminar, subir escaleras, inclinarse, cargar o llevar pesos y los esfuerzos moderados e intensos. El factor está asociado a los ítems 3, 4, 5, 6, 7, 8, 9, 10, 11.

*Factor 2: Salud Mental (SM).* Salud mental general, incluyendo depresión, ansiedad, control de la conducta o bienestar general. El factor está asociado a las variables 19, 24, 25, 26, 28, 29, 30, 31, 32.

Tabla 3: Salida de R-Commander del modelo con un único factor

```

Uniquenesses:
item.1 item.2 item.3 item.4 item.5 item.6 item.7 item.8 item.9 item.10 item.11
0.871 0.805 0.588 0.647 0.568 0.405 0.523 0.552 0.593 0.550 0.635
item.12 item.13 item.14 item.15 item.16 item.17 item.18 item.19 item.20 item.21 item.22
0.622 0.691 0.699 0.746 0.774 0.769 0.811 0.995 0.762 0.658 0.743
item.23 item.24 item.25 item.26 item.27 item.28 item.29 item.30 item.31 item.32 item.33
0.619 0.687 0.528 0.659 0.513 0.597 0.534 0.715 0.511 0.612 0.725
item.34 item.35 item.36
0.892 0.884 0.808

Loadings:
Factor1
item.1 0.359
item.2 0.442
item.3 0.642
item.4 0.594
item.5 0.657
item.6 0.772
item.7 0.690
item.8 0.669
item.9 0.638
item.10 0.671
item.11 0.604
item.12 0.615
item.13 0.556
item.14 0.549
item.15 0.504
item.16 0.475
item.17 0.481
item.18 0.435
item.19
item.20 0.488
item.21 0.584
item.22 0.507
item.23 0.617
item.24 0.559
item.25 0.687
item.26 0.584
item.27 0.698
item.28 0.635
item.29 0.683
item.30 0.534
item.31 0.700
item.32 0.623
item.33 0.524
item.34 0.328
item.35 0.340
item.36 0.439

Factor1
SS loadings 11.709
Proportion Var 0.325

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 1252.37 on 594 degrees of freedom.
The p-value is 3.68e-49

```

Tabla 4: Salida de R-Commander del modelo con 8factores comunes.



Uniquenesses:

item.1	item.2	item.3	item.4	item.5	item.6	item.7	item.8	item.9	item.10	item.11
0.434	0.526	0.432	0.349	0.394	0.243	0.262	0.352	0.380	0.406	0.390
item.12	item.13	item.14	item.15	item.16	item.17	item.18	item.19	item.20	item.21	item.22
0.493	0.188	0.163	0.305	0.407	0.241	0.005	0.909	0.526	0.398	0.005
item.23	item.24	item.25	item.26	item.27	item.28	item.29	item.30	item.31	item.32	item.33
0.371	0.388	0.315	0.520	0.269	0.368	0.330	0.551	0.401	0.518	0.272
item.34	item.35	item.36								
0.565	0.450	0.512								

Loadings:

	FF	SM	RF	RE	SG	DC	FS	VT
item.3	0.65						-0.22	
item.4	0.87	-0.22						0.26
item.5	0.65		0.23					-0.25
item.6	0.68		0.21					
item.7	0.88							
item.8	0.73			-0.24				
item.9	0.78							
item.10	0.75							
item.11	0.81							
item.24		0.88					-0.21	
item.25		0.69						
item.26		0.51						0.27
item.28		0.78						
item.29		0.81						
item.31		0.68						
item.13			0.82					-0.22
item.14			0.87					-0.22
item.15			0.72					0.21
item.16			0.59				0.24	0.23
item.17				0.83				
item.18				1.04				
item.1					0.77			
item.34					0.64			
item.36					0.62			
item.21						0.61		
item.22						1.06		
item.33							0.79	
item.35							0.71	
item.23		0.25			0.26			0.55
item.2				0.24	0.45			
item.12	0.39			0.24				-0.25
item.19	-0.26	0.23		-0.22				
item.20				0.39	-0.21			0.32
item.27		0.42		0.41				0.44
item.32		0.35						0.24

	FF	DM	RF	RE	SG	DC	FS	VT
SS loadings	5.57	4.04	2.68	2.38	2.08	1.77	1.57	1.45
Proportion Var	0.15	0.11	0.07	0.07	0.06	0.05	0.04	0.04
Cumulative Var	0.15	0.27	0.34	0.41	0.47	0.51	0.56	0.60

Factor Correlations:

	FF	SM	RF	RE	SG	DC	FS	VT
Factor1	1.000	-0.345	-0.322	-0.247	-0.462	0.430	-0.231	-0.355
Factor2	-0.345	1.000	0.471	0.319	0.509	-0.293	0.232	0.382
Factor3	-0.322	0.471	1.000	0.407	0.600	-0.329	0.295	0.320
Factor4	-0.247	0.319	0.407	1.000	0.408	-0.277	0.367	0.310
Factor5	-0.462	0.509	0.600	0.408	1.000	-0.422	0.332	0.449
Factor6	0.430	-0.293	-0.329	-0.277	-0.422	1.000	-0.269	-0.309
Factor7	-0.231	0.232	0.295	0.367	0.332	-0.269	1.000	0.112
Factor8	-0.355	0.382	0.320	0.310	0.449	-0.309	0.112	1.000

Test of the hypothesis that 8 factors are sufficient.  
 The chi square statistic is 468.66 on 370 degrees of freedom.  
 The p-value is 0.000374.

*Factor 3: Rol Físico (RF).* Grado en que la salud física interese en el trabajo y otras actividades diarias, incluyendo rendimiento menor que el deseado, limitación en el tipo



de actividades realizadas o dificultad en la realización de actividades. El factor está asociado a las variables 13, 14, 15,16.

*Factor 4: Rol Emocional (RE).* Grado en que los problemas emocionales interesen en el trabajo u otras actividades diarias. El factor está asociado a las variables 17, 18, 20.

*Factor 5: Salud General (SG).* Valoración personal de la salud, que incluye la salud actual, las perspectivas de salud en el futuro y la resistencia a enfermar. El factor está asociado a las variables 1, 2, 34, 36.

*Factor 6: Dolor Corporal (DC).* Intensidad del dolor y su efecto en el trabajo habitual, tanto fuera de casa como en el hogar. El factor está asociado a las variables 21, 22.

*Factor 7: Función Social (FS).* Grado en que los problemas de salud física o emocional interesen en la vida social habitual. El factor está asociado a las variables 33, 35.

*Factor 8: Vitalidad (VT).* Sentimiento de energía y vitalidad, frente al sentimiento de cansancio y agotamiento. El factor está asociado a las variables 23, 27.

Este agrupamiento es coincidente con los 8 dominios encontrados por los autores del instrumento.

### **3.1.3 Análisis Factorial confirmatorio (AFC)**

El AFC tiene como objetivo determinar si un modelo de medida especificado por el investigador, basándose en hipótesis teóricas o en un AFE previo, es consistente con la realidad. Para llegar a obtener alguna conclusión al respecto, es preciso abordar siguientes fases: (1) la especificación del modelo, (2) la identificación del modelo, (3) la estimación de parámetros y (4) la evaluación del ajuste del modelo. A continuación, sólo hacemos, por razones de espacio, algunas referencias a la fase de evaluación del ajuste.

El software R-Comander) ofrece además del estadístico  $\chi^2$  como índice de ajuste global, un segundo estadístico que denominaremos modelo  $\chi^2$  independiente (también llamado modelo de referencia). Este estadístico se distribuye también como una  $\chi^2$  bajo la hipótesis nula de que existe una completa independencia entre las variables observadas (matriz de correlaciones es la identidad) y tendrá tantos grados de libertad como el número de datos menos el número de parámetros independientes (varianzas) que se han de estimar.

También brinda algunos índices que comparan el valor del modelo teórico que se evalúa, con el modelo independiente. Los índices más utilizados son:

- Error Cuadrático Medio de Aproximación (RMSEA).

Este índice ha sido desarrollado como una medida absoluta de la diferencia de la estructura de relaciones entre el modelo propuesto y los valores de covarianza en población medida. La importancia de este índice radica en que refleja una diferencia absoluta entre el modelo propuesto y los datos observados, tomando en cuenta el número de estimaciones y el tamaño de la muestra implicada por el modelo bajo prueba. El índice RMSEA y su intervalo de confianza, cuando toman valores menores a 0.05 es indicio de que el ajuste entre el modelo y los datos es muy bueno, pero si sus valores resultan entre 0.05 y 0.08 el ajuste del modelo a los datos es razonable, mientras que si sus valores están entre 0.08 y .10 indica un ajuste pobre o mediocre. Ahora bien, el modelo deberá de rechazarse si los valores del índice RMSEA resultan mayores a 0.10. No obstante lo anterior, por razones prácticas debe incorporarse evidencia que refuerce los resultados obtenidos. Para ello se recurre a interpretar otros índices de bondad de ajuste.

- Índice RMR y SRMR.

El índice RMR está basado en los residuos, que son un promedio de las diferencias entre las covarianzas muestrales y las estimadas que se derivan del modelo. Como los residuos sin estandarizar están afectados por la escala en que se mide la variable, se suelen utilizar los residuos estandarizados construyéndose el llamado SRMR que está acotado entre 0 y 1.

Para confirmar los resultados del Análisis Factorial Exploratorio, se realizó el AFC sobre la matriz de covarianzas de la muestra total. Primero se aplicó el modelo mono factorial, hallándose que la prueba  $\chi^2$  fue significativa, por lo tanto, rechazamos la hipótesis nula de que los indicadores saturan en un único factor. También el RMSEA es de 0.129 lo que indica un ajuste pobre y el SRMR es de 0.109, por lo cual este modelo no es aceptable.

Después se aplicó el AFC para los 8 factores, encontrándose que la prueba  $\chi^2$  (917.0971,  $p = 4.494937e-19.0001$ ) fue significativa y, por lo tanto, rechazamos la hipótesis nula; también observemos que el RMSEA = 0.08348538 con un 90% de confiabilidad, lo cual indica que el ajuste entre el modelo y los datos es muy bueno; y el SRMR = 0.08178055. Concluimos que los indicadores garantizaron que el modelo a los datos es aceptable.

De nuevo los resultados mostraron 8 factores produciéndose un agrupamiento similar a los 8 dominios encontrados por los autores del instrumento.

El segundo factor, Salud Mental (valor propio de 1.759) explica el 8.59% de la variancia común y agrupa los dos ítems (31, 29) referentes a la Vitalidad, un ítem (19) a RE y otro (32) a FS de la versión original. El cuarto factor, Rol Emocional (valor propio de 1.33), explica el 4.91% de la variancia común y agrupa un ítem (20) de FS de la versión original. El quinto factor, Salud General (valor propio de 1.323), explica el 4.86% de la variancia común y agrupa un ítem (2) de la evolución de la salud de la versión original. El séptimo factor, Función Social (valor propio de 1.1614), explica el 3.747% de la variancia común

y agrupa dos ítems, los dos ítems (33, 35) son de la Salud General de la versión original. Los restantes 4 factores conservaron los mismos ítems que la versión original.

### 3.2 CONFIABILIDAD DEL CUESTIONARIO SF-36

En este estudio se utilizó la prueba de consistencia interna o coeficiente alfa de Cronbach para determinar la confiabilidad del cuestionario SF-36. Este coeficiente alfa de Cronbach provee un estimado de confiabilidad basado en todas las posibles correlaciones entre dos grupos de preguntas en una prueba. Indica el grado en que los ítems de un test covarían. El cálculo del coeficiente se fundamenta en la correlación media de cada ítem en cada una de las escalas con el total de las mismas y el número de preguntas (ítems) que contiene.

El valor numérico del coeficiente de Cronbach se desplaza entre 0 (confiabilidad baja) y 1 (máximo de confiabilidad). Muchos autores señalan que el valor deseable del este coeficiente de confiabilidad debiera estar, al menos, entre 0.70 a 0.90.

Dado que SF-36 es un cuestionario evaluado a través de escalas por la suma de sus componentes (escalas tipo Likert), se considera que el cuestionario es confiable cuando los resultados observados tienen una correlación alta con los resultados verdaderos o que se establecieron en la hipótesis. Es común emplear el coeficiente alfa de Cronbach cuando se trata de alternativas de respuestas policotómicas, como las escalas tipo Likert.

El coeficiente alfa de Cronbach puede ser calculado de las dos formas siguientes:

1.-Mediante la varianza de los ítems y la varianza del puntaje total

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum s_i^2}{s_t^2} \right]$$

donde:

$\alpha$  : Coeficiente de confiabilidad de la prueba o cuestionario.

k: número de ítems del instrumento.

$\sum s_i^2$  : Suma de la varianza individual de los ítems.

$s_t^2$  : Varianza total del instrumento.

2.-Mediante la matriz de correlación de los ítems.

$$\alpha = \frac{k\bar{p}}{1 + \bar{p}(k-1)}$$

donde:

k : número de ítems.

$\bar{p}$ : Promedio de las correlaciones lineales entre cada uno de los ítems.

El análisis de la consistencia interna del instrumento se llevó a cabo, empleando el paquete R-Commander mediante el coeficiente Alfa de Cronbach. La tabla 5 muestra los resultados obtenidos. Puede observarse que la confiabilidad del instrumento de 36 ítems presentó un  $\alpha = 0.9329$ , lo que confiere a la escala una consistencia interna muy alta o elevada. El análisis ítem a ítem proporciona el  $\alpha$  de la escala si se elimina un ítem cada vez. En la última columna de esta salida se observan las correlaciones de cada ítem con la suma de los otros ítems lo que es un índice de homogeneidad de cada ítem.

De acuerdo a los resultados anteriores el cuestionario SF-36 puede considerarse confiable para las mujeres con Lupus en el estado de Puebla, México.

Tabla 5. Salida de R-Commander: Confiabilidad del cuestionario

```
Alpha reliability = 0.9329
Standardized alpha = 0.9398

reliability deleting each item in turn:

```

	Alpha	Std.Alpha	r(item, total)
sf..36.....1	0.9324	0.9396	0.3658
x2	0.9316	0.9386	0.4731
x3	0.9305	0.9379	0.5710
x4	0.9309	0.9383	0.5278
x5	0.9302	0.9377	0.5938
x6	0.9299	0.9366	0.7129
x7	0.9298	0.9373	0.6309
x8	0.9301	0.9376	0.6010
x9	0.9303	0.9376	0.5823
x10	0.9301	0.9375	0.6115
x11	0.9307	0.9380	0.5592
x12	0.9306	0.9376	0.6027
x13	0.9302	0.9379	0.6012
x14	0.9304	0.9380	0.5880
x15	0.9310	0.9384	0.5456
x16	0.9317	0.9387	0.5004
x17	0.9313	0.9383	0.5189
x18	0.9320	0.9388	0.4659
x19	0.9394	0.9435	-0.0761
x20	0.9316	0.9388	0.4608
x21	0.9309	0.9380	0.5350
x22	0.9313	0.9386	0.4996
x23	0.9307	0.9376	0.5798
x24	0.9313	0.9383	0.5086
x25	0.9301	0.9370	0.6425
x26	0.9307	0.9377	0.5857
x27	0.9300	0.9368	0.6748
x28	0.9307	0.9375	0.5929
x29	0.9303	0.9372	0.6281
x30	0.9313	0.9382	0.5150
x31	0.9302	0.9371	0.6428
x32	0.9305	0.9376	0.5984
x33	0.9310	0.9382	0.5217
x34	0.9326	0.9397	0.3637
x35	0.9324	0.9397	0.3583
x36	0.9319	0.9390	0.4306

### 3. CONCLUSIONES

Se realizaron los Análisis Factorial Exploratorio y Confirmatorio de la versión española del SF-36 con el fin de analizar la validez de constructo y comprobar la estructura del instrumento y sus dominios.

Los resultados mostraron 8 factores, los cuales explicaban el 73.358% de la variancia total de los datos, lo que produjo un agrupamiento de los 8 dominios encontrados por los autores del instrumento. El segundo factor, Salud Mental (valor propio de 1.759) explica el 8.59% de la variancia común y agrupa los dos ítems (31, 29) referentes a la Vitalidad, un ítem (19) a RE y otro (32) a FS de la versión original. El cuarto factor, Rol Emocional (valor propio de 1.33), explica el 4.91% de la variancia común y agrupa un ítem (20) de

FS de la versión original. El quinto factor, Salud General (valor propio de 1.323), explica el 4.86% de la variancia común y agrupa un ítem (2) de la evolución de la salud de la versión original. El séptimo factor, Función Social (valor propio de 1.1614), explica el 3.747% de la variancia común y agrupa dos ítems, los dos ítems (33, 35) son de la Salud General de la versión original. Los restantes 4 factores conservaron los mismos ítems que la versión original.

Los resultados encontrados en este estudio muestran que esta versión del SF-36 presenta propiedades psicométricas estables, muy útil tanto para utilizarse en investigación, ya que permite la obtención de datos empíricos, así como, para la práctica clínica de la valoración del estado de los pacientes y del curso de la evolución de la enfermedad.

## REFERENCIAS

- Camargo, J.F., Tobón, G.J., Fonseca, N., Díaz, J.L., Uribe, M., Molina, F. Autoimmune rheumatic diseases in the intensive care unit: experience from a tertiary referral hospital and review of the literature. *Lupus*, 4, 315-320. (2005).
- Gemma Vilaguta; Montse Ferrera; Luis Rajmil. El cuestionario de salud SF-36 español: una década de experiencia y nuevos desarrollos. (2004) Disponible en <http://www.scielo.org/pdf/gsv19n2/revision1.pdf>
- Gómez, L.M., Anaya, J.M., González, C.I., Pineda-Tamayo, R., Otero, W., Arango, A. Polymorphism in Colombian patients with autoimmune diseases. *Genes Immunology*, 6, 628-631. (2005).
- Hernández Sampieri, R. Fernández Collado, C. y Baptista Lucio, P. Metodología de la Investigación. Mc Graw Hill Interamericana. México. (2006).
- Linares Fleites, G. Análisis de Datos Multivariados. Editorial Benemérita Universidad Autónoma de Puebla. Facultad de Ciencias de la Computación. México. (2006).
- Méndez Martínez, S. Factores de la triada ecológica que influyen en la persistencia de la infección cervical por virus del papiloma humano en mujeres con lupus eritematoso sistémico. Tesis de Doctorado en Ciencias Ambientales. Posgrado en Ciencias Ambientales. Instituto de Ciencias Benemérita Universidad Autónoma de Puebla. (2017)

Molina, J.F., Anaya, J.M., & Molina, J. Lupus Eritematoso: Manual práctico para médicos y pacientes (2a. ed.). Medellín, Colombia: CIB. (2005).

Polit, D. y Hungler, B. P. (2000): Investigación Científica en Ciencias de la Salud. McGraw-Hill Interamericana. México. 725p.

RCommander una interfaz gráfica para R. Disponible en <http://www.rcommander.com/>

Santos Sánchez, G. Validez y confiabilidad del cuestionario de calidad de vida SF.36 en mujeres con LUPUS, Puebla. Tesis de Licenciatura en Matemáticas Aplicadas. BUAP, FCFM, (2017).

Schwartzmann, L. Calidad de vida relacionada con la salud: aspectos conceptuales. Ciencia y Enfermería, 2, 9-21. (2003).

Thumboo, J.,y Strand, V. Health-related quality of life in patients with systemic lupus erythematosus: an update. The Annals, Academy of Medicine, Singapore, 36, 115-122. . (2007).