

Big Data Analytics:

una perspectiva multidisciplinaria
para la mejora del proceso de toma
de decisiones en las organizaciones

José Alberto Hernández Aguilar
José Crispín Zavala Díaz
Federico Alonso Pecina
Martín G. Martínez Rangel
(coordinadores)



BONILLA
ARTIGAS
EDITORES

Universidad Autónoma
del Estado de Morelos

En este libro se recopilan un conjunto de aplicaciones que ilustran el funcionamiento del análisis de datos grandes (Big Data Analytics), éstas se encuentran en áreas de conocimiento diversas entre las que se incluyen: el sector eléctrico, el sector educativo, el sector salud, el gobierno abierto, la detección de faltantes y anomalías en las organizaciones, ciudades inteligentes, entre otras, todas ellas de relevancia para mejorar la toma de decisiones en las organizaciones. Así mismo se discuten los procesos fundamentales para el análisis de datos (Data Analytics) y el descubrimiento de conocimiento (Knowledge Discovery in DataBase) desde conceptos como: datos grandes (Big Data), la minería de datos (Data Mining), y la minería de datos sociales (Social data Mining). El libro está orientado a alumnos de pregrado y posgrado, así como investigadores y practicantes, que desean introducirse al conocimiento y aplicación del análisis de datos grandes en diversas áreas de las organizaciones. Big data es un tema de discusión actual dada la variedad de datos que se tienen, la velocidad con la que se producen, la veracidad y el volumen creciente de éstos generado con el uso de las TICs.

Big data/ Toma de decisiones



BONILLA
ARTIGAS
EDITORES



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Big Data Analytics

Big Data Analytics: una perspectiva multidisciplinaria para la mejora del proceso de toma de decisiones en las organizaciones

José Alberto Hernández Aguilar

José Crispín Zavala Díaz

Federico Alonso Pecina

Martín G. Martínez Rangel

(coordinadores)



BONILLA
ARTIGAS
EDITORES



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Esta publicación fue financiada con recursos del Programa de Fortalecimiento de la Calidad Educativa (PFCE) 2017.

Big Data Analytics: una perspectiva multidisciplinaria para la mejora del proceso de toma de decisiones en las organizaciones / José Alberto Hernández Aguilar, José Crispín Zavala Díaz, Federico Alonso Pecina, Martín G. Martínez Rangel, coordinadores. Ciudad de México : Bonilla Artigas Editores : Universidad Autónoma del Estado de Morelos, 2019

312 p. ; 17 x 23 cm.
(Colección Pública investigación : No. 3)

ISBN: 978-607-8560-32-5 (Bonilla Artigas Editores)
ISBN: 978-607-8519-81-1 (UAEM)

1. Toma de decisiones. 2. Big data.

I. Hernández Aguilar, José Alberto, coord. II. Zavala Díaz, José Crispín, coord. III. Alonso Pecina, Federico, coord. IV. Martínez Rangel, Martín G. coord.

LC HD30.23

DEWEY 658.4

Esta publicación fue dictaminada por pares académicos bajo la modalidad doble ciego.

Los derechos exclusivos de la presente edición quedan reservados para todos los países de habla hispana. Queda prohibida su reproducción, parcial o total, por cualquier medio conocido o por conocerse sin el consentimiento por escrito de los legítimos poseedores de derechos.

Big Data Analytics: una perspectiva multidisciplinaria para la mejora del proceso de toma de decisiones en las organizaciones
D.R. © José Alberto Hernández Aguilar
D.R. © José Crispín Zavala Díaz
D.R. © Federico Alonso Pecina
D.R. © Martín G. Martínez Rangel

Primera edición: marzo 2018

© 2018, Bonilla Distribución y Edición S.A. de C. V.
Hermenegildo Galeana #111
Col. Barrio del Niño Jesús, C.P. 14080
Ciudad de México
procesoseditoriales@bonillaartigaseditores.com.mx
www.bonillaartigaseditores.com

D.R. © 2018, Universidad Autónoma del Estado de Morelos
Av. Universidad 1001
Col. Chamilpa, C.P. 62209
Cuernavaca, Morelos
publicaciones@uaem.mx
libros.uaem.mx

Coordinación editorial: Bonilla Artigas Editores
Formación de interiores: Saúl Marcos Castillejos
Diseño de portada: Mariana Guerrero del Cueto

ISBN: 978-607-8560-32-5 (Bonilla Artigas Editores)
ISBN: 978-607-8519-81-1 (UAEM)

Impreso y hecho en México

Contenido

Prólogo.....	9
I. APLICACIONES A LA ENERGÍA ELÉCTRICA Y RECURSOS NATURALES	
<i>Big Data y Analytics: infraestructura esencial para la red eléctrica inteligente. Aplicación a sistemas de generación fotovoltaicos</i> <i>Guillermo Flavio Escobedo Briones, José Alberto Incera Diéguez, Gustavo Arroyo Figueroa y Norma Elena Jácome Grajales.....</i>	15
Visualización y análisis de precipitaciones pluviales en la República mexicana desde 1981 al 2016 <i>María Concepción Salvador-González y Juana Canul-Reich.....</i>	43
Desarrollo de un sistema inteligente mediante redes bayesianas y procesamiento de imágenes aéreas para analizar el potencial eólico en Querétaro y sus implicaciones legales <i>Daniel Cantón Enriquez, Alberto Ochoa Zezzatti, Sandra Bustillos Durán, Saul Gonzalez Campos, Rodolfo Rincones Delgado, Martín Montes-Rivera y Ana Marcela Herrera Navarro.....</i>	59
II. BASES DE DATOS EN LA EDUCACIÓN	
<i>Big Data en el análisis del rendimiento académico a nivel licenciatura</i> <i>Julio Cesar Ponce Gallegos, Beatriz Angélica Toscano de la Torre, Aurora Torres Soto, Ma. Dolores Torres Soto, Francisco Javier Ornelas Zapata y María Lourdes Montes Torres.....</i>	87
Identificación visual del compromiso estudiantil universitario en licenciaturas en tecnologías a través de técnicas de análisis multivariado en Lenguaje R <i>Erika Yunuen Morales Mateos, María Arely López Garrido, Oscar Alberto González González, Rubén Jerónimo Yedra y Julián Javier Francisco León.....</i>	123
III. ACCESO A DATOS PÚBLICOS	
Modelo de investigación para el análisis del gobierno abierto en México <i>José Alonso Fernández López, Maribel Leyva Gaxiola y Rubén Antonio García Mendoza.....</i>	145

El uso de analítica de datos aplicada en las cuentas por cobrar de los empleados de un organismo público
Enrique Montoya Flores, José Alberto Hernández Aguilar, José Crispín Zavala Díaz, Carlos Alberto Ochoa Ortiz, Ana Karen Castañeda Escobar y Jacqueline López Calderón183

IV. ANÁLISIS SOCIAL

Estudio geoespacial de las expectativas de logro social y espacios de diversión mediante predicción numérica en una Smart City
Sandra Bustillos Durán, Rodolfo Rincones Delgado, Saúl Gonzalez Campos, Alberto Ochoa Zezzatti, Luis Cervera Gómez, Rocío Maciel Arellano, Víctor Larios Rosillo203

Las ciencias sociales y el Big Data social como propiedad emergente para el análisis del comportamiento de los sujetos en las organizaciones
Augusto Renato Pérez Mayo, Nohemí Roque Nieto y Fidji Danaé Pérez Alvarez227

V. BASES DE DATOS EN EL SECTOR SALUD

Análisis de bases de datos en el sector salud para la detección de factores de riesgo
Aurora Torres Soto, Ma. Dolores Torres Soto, Julio César Ponce Gallegos, Beatriz Angélica Toscano de la Torre y María de la Luz Torres Soto247

Detección de anomalías y tendencias fuera de la norma
Federico Alonso-Pecina y Rocío Diego-Celis261

VI. OTRAS APLICACIONES

Identificador de semáforos mediante un algoritmo PSO como clasificador de color en conjunto con detección de área de interés con algoritmo K-means
Martín Montes Rivera, Alejandro Padilla Díaz, Juana Canul Reich, Julio César Ponce Gallegos y Alberto Ochoa Zezzatti275

Explorando el uso de la lógica dinámica como soporte para analítica en la toma de decisiones
Miguel Pérez Ramírez, José Alberto Hernández Aguilar y Norma J. Ontiveros Hernández291

En este libro se recopilan un conjunto de aplicaciones que ilustran el funcionamiento del análisis de datos grandes (*Big Data Analytics*), éstas se encuentran en áreas de conocimiento diversas entre las que se incluyen: el sector eléctrico, el sector educativo, el sector salud, el gobierno abierto, la detección de faltantes y anomalías en las organizaciones, ciudades inteligentes, entre otras; todas ellas de relevancia para mejorar la toma de decisiones en las organizaciones. Así mismo, se discuten los procesos fundamentales para el análisis de datos (*data analytics*) y el descubrimiento de conocimiento (*knowledge discovery in database*) desde conceptos como: datos grandes (*Big Data*), la minería de datos (*data mining*), y la minería de datos sociales (*social data mining*).

Este libro ha sido creado con capítulos escritos por alumnos e investigadores de los programas de posgrado de la Facultad de Contaduría, Administración e Informática de la Universidad Autónoma del Estado de Morelos, así como de sus colaboradores en distintos centros de investigación y universidades reconocidas a nivel nacional, entre los que se encuentran el Instituto Nacional de Electricidad y Energías Limpias, la Universidad Juárez Autónoma de Tabasco, la Universidad Autónoma de Ciudad Juárez, la Universidad Autónoma de Querétaro, la Universidad Tecnológica de Corregidora, la Universidad Autónoma de Aguascalientes, la Universidad Autónoma de Nayarit, el Centro de Competitividad en Ciudades Inteligentes de la Universidad de Guadalajara, la Universidad Politécnica de Aguascalientes y el Instituto Tecnológico de Zacatepec.

El libro está orientado a alumnos de pregrado y posgrado, así como investigadores y practicantes inquietos, que desean introducirse al conocimiento y aplicación del análisis de datos grandes en las organizaciones, tema de discusión actual dada la variedad de datos que se tienen, la velocidad con la que se producen y su creciente volumen, derivado del uso de las nuevas tecnologías.

Este libro es una compilación de trece capítulos sobre esta temática, los cuales nos permiten conocer las bondades del análisis de datos, mostrando también las metodologías existentes y los procesos necesarios como el preprocesamiento de los datos mediante la selección, transformación, fusión e integración, así como de herramientas (*software*) existentes que permiten el análisis de los datos, la implementación de algoritmos, y la interpretación o visualización de los resultados.

Los capítulos fueron agrupados por su temática en las siguientes secciones: La sección uno –“Aplicaciones a la energía eléctrica y recursos naturales”– agrupa los capítulos primero al tercero, en los cuales se discuten aplicaciones a la energía y recursos naturales: energía eléctrica, eólica y la precipitación pluvial. El primer

- CONAGUA. (2017). Precipitación promedio. Sitio oficial de datos abiertos del Gobierno de la República mexicana. Recuperado de <https://datos.gob.mx/busca/dataset/precipitacion>
- COOK, K. A., y Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics (No. PNNL-SA-45230). Pacific Northwest National Lab (PNNL), Richland, WA, Estados Unidos.
- GÓMEZ, F. A. (2010). Archivos KML estados y municipios de México. Recuperado de <https://sites.google.com/site/fernandoagomezf/Home/noticias-del-sitio/archivoskml estadosymunicipiosdemexico>
- LUDWIG, M. (2015). QgisMarkerCluster Plugin. Recuperado de <https://plugins.qgis.org/plugins/QgisMarkerCluster/>
- MORALES, A. (2017). Publica tus mapas en la web con qgis2web. Recuperado de <https://mappinggis.com/2016/03/crea-aplicaciones-webmapping-con-qgis/>
- QGIS. (2018). QGIS - El SIG líder de código abierto para escritorio. Recuperado de <https://qgis.org/es/site/about/index.html>
- ROCHA, L.A. (2007). Guidelines for designing WEB maps. An Academic Experience. *6th FIG Regional Conference, San José, Costa Rica.*
- TABLEAU. (2017a). Seis años de liderazgo en el cuadrante mágico de Gartner en análisis e inteligencia de negocios. Tableau Software. Recuperado de <https://www.tableau.com/es-es/reports/gartner>
- _____. (2017b). Tableau Desktop. Recuperado de <https://www.tableau.com/es-es/products/desktop>

Desarrollo de un sistema inteligente mediante redes bayesianas y procesamiento de imágenes aéreas para analizar el potencial eólico en Querétaro y sus implicaciones legales

Daniel Cantón Enriquez¹
 Alberto Ochoa Zezzatti²
 Sandra Bustillos Durán²
 Saul Gonzalez Campos²
 Rodolfo Rincones Delgado²
 Martín Montes-Rivera³
 Ana Marcela Herrera Navarro⁴

Introducción

El uso de las energías renovables se ha popularizado gradualmente desde finales del siglo xx, debido a la fuerte dependencia de los combustibles fósiles y el daño que estos causan al medio ambiente, por lo que a nivel mundial existe un aumento en la demanda de energías eléctricas que provengan de fuentes renovables, caracterizadas por reponerse a un ritmo igual o superior al que son consumidas (González, 2009). Desde tiempos inmemoriales el hombre ha utilizado la fuerza del viento para distintos propósitos desde el bombeo de agua hasta para moler granos en molinos. La creciente demanda de energía eléctrica a nivel mundial ha hecho que la energía eólica tenga un rol muy importante, ya que es una fuente de energía renovable que es inagotable y reduce el uso de combustibles fósiles, los cuales originan las emisiones de efecto invernadero causando el calentamiento global, de esta manera el uso eficiente de la energía eólica contribuye al desarrollo sostenible (AMDEE, 2014). El propósito de esta investigación consiste en el desarrollo de un sistema inteligente mediante redes bayesianas y procesamiento de imágenes aéreas para analizar el potencial eólico en el estado de Querétaro. Actualmente el uso de sistemas inteligentes se convertido en una herramienta popular para la solución de diversos

¹ Maestría en Ciencias Computacionales, Universidad Autónoma de Querétaro.

² Universidad Autónoma de Ciudad Juárez.

³ Universidad Politécnica de Aguascalientes.

⁴ Maestría en Ciencias Computacionales, Universidad Autónoma de Querétaro.

problemas aplicados empleando diferentes técnicas de inteligencia artificial (IA), algunas de estas aplicaciones se han enfocado en sistemas robóticos y automatizados, sistemas de salud, sistemas de asistencia y colaboración, así como al pronóstico de las ciencias atmosféricas. Hasta el momento son pocos los sistemas inteligentes relacionados al análisis del potencial eólico, algunos de éstos están relacionados con empresas y otros gestionados por instituciones de gobierno. Éstos funcionan con diferentes técnicas de IA como las redes bayesianas pero, hasta el momento no emplean un sistema de información geográfico (SIG) como fuente de base de datos para el análisis del potencial eólico y como apoyo a la visualización de los resultados obtenidos en un área específica. En este capítulo se presenta el análisis de intervención de sistemas inteligentes como lo son las redes bayesianas, el procesamiento de imágenes, los sistemas de información geográficos, el potencial eólico en el estado de Querétaro, y los diferentes trabajos relacionados en el uso de las redes bayesianas para el análisis del potencial eólico, así como el uso de un SIG y la metodología que se propone para la elaboración de esta investigación. Este capítulo está estructurado de la siguiente forma: en la primera sección se cuenta con el planteamiento del problema, la justificación, preguntas de investigación con su respectiva hipótesis, el objetivo general y los objetivos específicos, alcances, limitaciones y el esquema general del sistema. En la segunda sección se hace una recopilación del marco de referencia necesario para el desarrollo de este proyecto. En la tercera sección se describen los trabajos relacionados que mencionan los resultados recientes acerca de los diferentes sistemas inteligentes que emplean las redes bayesianas para el análisis del potencial eólico. En la cuarta sección se presentan los mapas y modelos considerados. En la quinta sección se muestra el diseño y la implementación de la herramienta inteligente, en la sexta sección se presentan los resultados generados por el sistema inteligente propuesto. Finalmente, en la séptima sección se muestran las conclusiones y el trabajo futuro.

Definición de la problemática

La energía eólica ha tenido un crecimiento relevante a nivel mundial, los países líderes en cuanto a capacidad instalada son China, EE.UU., Alemania, España e India (Wagner y Jyotirmay, 2013), esto ha propiciado la instalación de parques eólicos. En los últimos años se ha dado un gran impulso al sector eólico en México, por lo que en la actualidad se cuentan con 37 parques eólicos en estados como Baja California,

Nuevo León, San Luis Potosí, Zacatecas, Tamaulipas, Jalisco, Puebla, Chiapas y Oaxaca;⁵ con una potencia de 3 527 MW a nivel nacional (AMDEE, 2016).

El incremento del sector eólico en México y en el mundo, ha motivado la necesidad de investigar y promover el desarrollo de herramientas que analicen el potencial eólico disponible en aquellas zonas donde se pretende instalar un parque eólico. Por lo anterior es necesario evaluar los posibles emplazamientos con recursos eólicos sobresalientes, de acuerdo a los rangos de viento de referencia establecidos en la clasificación del NREL⁶ (Elliot, Holladay, Barchet, Foote y Sandusky, 2009). Por lo regular para analizar el potencial eólico, se utilizan diferentes modelos como los estadísticos y de análisis multivariable, que desde hace tiempo son los encargados de pronosticar las condiciones ideales para la explotación del potencial eólico, el desarrollo de estas herramientas de análisis se ha impulsado derivado del interés de analizar las condiciones meteorológicas y predicciones del viento, ya que en numerosos aspectos de la vida diaria, es de suma importancia, como se describe en ANEMOS (2011). Hoy en día ante la creciente disponibilidad de información meteorológica y climática, es necesario el uso de nuevas técnicas más eficientes y automatizadas, que permitan obtener un mejor análisis del recurso eólico. La IA, a través de las diferentes áreas que la componen, como las redes neuronales, la programación genética y las redes bayesianas, han sido utilizadas como alternativa a los métodos tradicionales para evaluar y analizar el potencial eólico (Lazarevska, 2016), esto con la finalidad de desarrollar herramientas distintas en las áreas del conocimiento de las ciencias computacionales. A estas implementaciones se les denomina sistemas inteligentes ya que permiten solucionar un problema en forma con mecanismos automáticos que permiten obtener una conclusión, e incluso tomar decisiones sin la intervención del ser humano (Gutiérrez, Cano, Cofiño y Sordo, 2004).

Generalmente, las redes bayesianas tienden a formar modelos con respecto a procesos ambientales relacionados con las ciencias atmosféricas, debido a esto es natural que exista una conexión entre las redes bayesianas y un SIG, igualmente se sabe que las redes bayesianas son en gran medida eficiente para realizar clasificación de imágenes de manera remota (Ames y Anselmo, 2015). Las aplicaciones *software* y trabajos de investigación que existen actualmente para el análisis del potencial eólico se limitan a realizar la predicción y/o pronóstico del potencial eólico solo mediante redes bayesianas, sin utilizar un SIG y en caso de utilizarlo se emplean

⁵ En esta entidad, en la región del Istmo de Tehuantepec, se cuenta con 27 parques eólicos con una capacidad de 2 359.97 MW contribuyendo con más de 65% de la energía producida por el sector eólico.

⁶ Laboratorio Nacional de Energía Eólica de Estados Unidos por sus siglas en inglés NREL (National Renewable Energy Laboratory).

modelos estadísticos y probabilísticos. Es necesario analizar de una manera más eficiente y eficaz el potencial eólico de una región de interés (estado de Querétaro) y de las zonas de mayor potencial eólico (micrositios de la entidad).

Por lo tanto, se puede optimizar el proceso asociado con un sistema inteligente que permita a los usuarios conocer, comprender y manipular la información del potencial eólico dentro del estado de Querétaro, utilizando redes bayesianas y el procesamiento de imágenes aéreas a través de un SIG.

Explicación del funcionamiento de la Investigación realizada

El empleo de diferentes técnicas de IA como las redes bayesianas, se está convirtiendo en un medio fiable que complementa a los métodos tradicionales que analizan y evalúan el potencial eólico de una manera más eficiente y eficaz. Además, en el caso de la energía eólica existen varias razones que justifican el uso de un sistema inteligente:

- Ante la creciente disponibilidad de información meteorológica y climática es permitiría procesar de forma rápida y fiable los datos disponibles, de manera que puedan resolver el problema dado como se describe en (Gutiérrez et al. 2004).
- Se optimizarían los métodos que analizan el potencial eólico, de manera que se obtenga la clasificación de la potencia del viento para considerar si la zona de interés es apta para proyectos a escala comercial como se describe en Elliot et al. (2009).

Por lo tanto, el análisis del potencial eólico en el estado de Querétaro mediante redes bayesianas y procesamiento de imágenes aéreas ofrecerá a los usuarios la oportunidad de analizar el potencial eólico en un área de interés determinada por el usuario, así como por los datos del viento recibidos de las distintas fuentes de recolección. A través del Internet se recolectarán los datos del viento y las consultas *ad hoc* por parte de los usuarios a través del SIG; el sistema inteligente analizará en un primer paso las imágenes del aérea acotada por el usuario en el SIG, utilizando para ello el procesamiento de imágenes aéreas y por medio de las redes bayesianas se identifican los datos del viento y aquellos proporcionados en las consultas *ad hoc*. Los resultados del sistema inteligente se enviarán por Internet para ser visualizados como capas en el mapa cartográfico del SIG, así mismo se podrá añadir información por medio de tablas, gráficos y texto plano. De acuerdo con AMDEE se tiene previsto que para el año 2020 en el estado de Querétaro se encuentren instalados 60 MW

de potencia en energía eólica en la entidad (AMDEE, 2015). Uno de los beneficios esperados, es que mediante el análisis del potencial eólico se pueda obtener la clasificación de vientos de la entidad, lo que permitiría saber que zonas son aptas para proyectos a escala comercial y cuales para proyectos en zonas rurales o fuera de la red eléctrica nacional. De esta manera se podría:

- Obtener energía eléctrica renovable y limpia para el sector empresarial, hogares e instituciones públicas, evitando así la contaminación por medio de gases de efecto invernadero.
- Promover la investigación y el desarrollo del sector eólico en la entidad y en la región del Bajío, actualmente hay instituciones públicas como la Universidad Autónoma de Querétaro y el Ciateq que forman parte del Centro Mexicano de Innovación en Energía Eólica (Cemie-Eólico), por lo que se comienza a tener personal capacitado en las diferentes áreas del sector eólico.

Implicaciones del desarrollo de un proyecto eólico

- Desarrollar un SIG con consultas *ad-hoc* en web, en donde el usuario podrá interactuar con el sistema inteligente para identificar los cambios ocurridos en el análisis del viento.
- Examinar y evaluar a través de pruebas de usabilidad SUS con escala de Likert y el Test de esfuerzo NASA-TLX la funcionalidad del sistema inteligente con los usuarios, el grado de usabilidad y esfuerzo requerido para usar el SIG a través de consultas *ad hoc*.
- Comparar los resultados obtenidos del sistema inteligente con el módulo de MCP del *software* WindPRO, así como con otros sistemas de predicción y/o pronóstico del viento a fin de comprobar que el sistema inteligente sea confiable, con un porcentaje de dispersión bajo. (opcional, se tendría que conseguir el WindPRO).

Alcances

- Se utilizará el procesamiento de imágenes aéreas en las áreas de interés, en las cuales hayan sido seleccionadas a través del SIG por los usuarios.
- Mediante redes bayesianas se analizará el potencial eólico con los datos proporcionados por las estaciones meteorológicas del estado de Querétaro.

- Los usuarios podrán interactuar con el SIG por medio de consultas *ad hoc* desde cualquier equipo o dispositivo móvil que cuente con una conexión a Internet.
- El análisis realizado por el sistema inteligente y el hospedaje del SIG se encontrarán en servidores. Conocer las principales características del recurso eólico en el estado de Querétaro, obteniendo así un primer mapa de los recursos eólicos de la entidad.

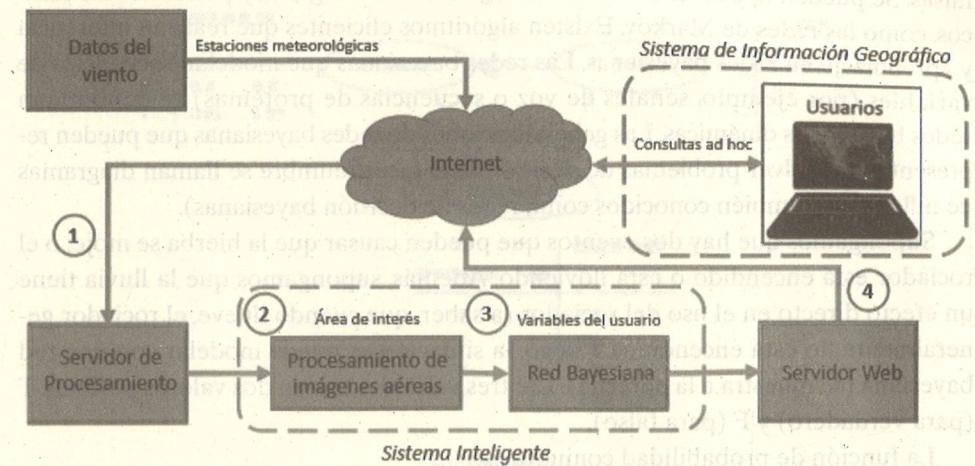
Limitaciones de la herramienta inteligente propuesta

- El sistema inteligente se limitará sólo al espacio geográfico del estado de Querétaro, con la única posibilidad de poder determinar microsítios para aquellas zonas de amplio potencial eólico en la entidad.
- Para el análisis del área de estudio por parte del sistema inteligente y la interacción del usuario con el SIG se necesitará de una conexión a Internet.
- Se requerirán mediciones de más de un ciclo anual del recurso eólico de la entidad, ya que un año es insuficiente para resolver dudas sobre la variabilidad de un año a otro.
- Se tomará únicamente en cuenta los datos del viento proporcionados por estaciones meteorológicas que cuenten con torres por arriba de los 30 metros de altura, ya que en alturas inferiores se ve reducido de manera considerable las medidas del viento, por obstrucción de edificios y árboles.

Esquema general del sistema propuesto

El esquema de desarrollo propuesto muestra los diferentes elementos que se necesitan para el desarrollo del sistema inteligente, así como los datos del viento y el área de interés (estado de Querétaro). En la figura 1 se hace una representación entre la interacción del sistema inteligente con el SIG por medio de consultas *ad hoc* y los diferentes servidores que se utilizarán para retroalimentar ambos sistemas por medio de una conexión de Internet.

Figura 1. Esquema general del sistema propuesto



Fuente: Elaboración propia.

Marco de referencia

Una red bayesiana, red de Bayes, red de creencias, modelo Bayes(iano) o modelo gráfico acíclico dirigido probabilístico es un modelo gráfico probabilístico (un tipo de modelo estadístico) que representa un conjunto de variables y sus dependencias condicionales a través de un gráfico acíclico dirigido (DAG). Por ejemplo, una red bayesiana podría representar las relaciones probabilísticas entre enfermedades y síntomas. Los síntomas dados, la red se puede utilizar para calcular las probabilidades de la presencia de diversas enfermedades.

Formalmente, las redes bayesianas son DAG cuyos nodos representan variables en el sentido bayesiano: pueden ser cantidades observables, variables latentes, parámetros desconocidos o hipótesis. Los bordes representan dependencias condicionales; los nodos que no están conectados (no hay una ruta de una de las variables a la otra en la red bayesiana) representan variables que son condicionalmente independientes entre sí. Cada nodo está asociado con una función de probabilidad que toma, como entrada, un conjunto particular de valores para las variables primarias del nodo, y da (como salida) la probabilidad (o distribución de probabilidad, si corresponde) de la variable representada por el nodo. Por ejemplo, si $\{m\}$ m nodos principales representan $\{m\}$ m variables booleanas, entonces la función de probabilidad podría representarse mediante una tabla de entradas $\{2^{\wedge} \{m\}\} 2^{\wedge} \{m\}$, una entrada para cada una

de las $\{2^m\} 2^m$ combinaciones posibles de sus padres que sean verdaderas o falsas. Se pueden aplicar ideas similares a gráficos no dirigidos, y posiblemente cíclicos, como las redes de Markov. Existen algoritmos eficientes que realizan inferencia y aprendizaje en redes bayesianas. Las redes bayesianas que modelan secuencias de variables (por ejemplo, señales de voz o secuencias de proteínas) se denominan redes bayesianas dinámicas. Las generalizaciones de redes bayesianas que pueden representar y resolver problemas de decisión bajo incertidumbre se llaman diagramas de influencia (también conocidos como redes de decisión bayesianas).

Supongamos que hay dos eventos que pueden causar que la hierba se moje: o el rociador está encendido o está lloviendo. Además, supongamos que la lluvia tiene un efecto directo en el uso del rociador (a saber, que cuando llueve, el rociador generalmente no está encendido). Luego, la situación se puede modelar con una red bayesiana (se muestra a la derecha). Las tres variables tienen dos valores posibles, T (para verdadero) y F (para falso).

La función de probabilidad conjunta es:

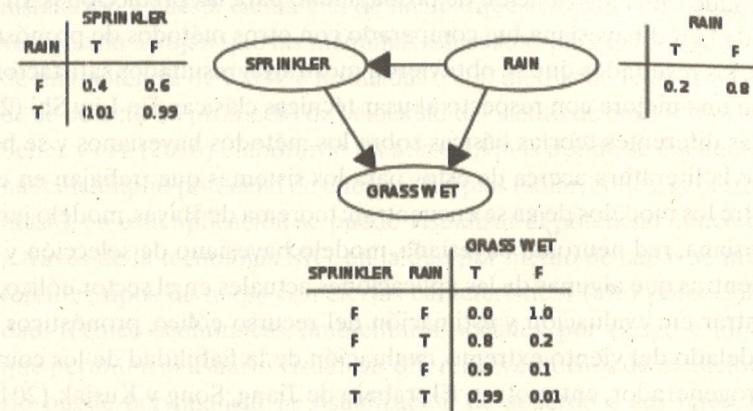
$$\Pr(G, S, R) = \Pr(G | S, R) \Pr(S | R) \Pr(R) \quad \{ \Pr(G, S, R) = \Pr(G | S, R) \Pr(S | R) \Pr(R) \}$$

Donde los nombres de las variables se han abreviado como G = hierba húmeda (verdadero / falso), S = rociador encendido (verdadero / falso), y R = lluvia (verdadero / falso).

El modelo puede responder preguntas como: “¿cuál es la probabilidad de que llueva, dado que el césped está mojado?” usando la fórmula de probabilidad condicional y sumando todas las variables afectando dicho modelo, como se muestra en la figura 2:

Los sistemas inteligentes que utilizan redes bayesianas han sido abordados en múltiples sectores, como en el análisis del potencial eólico. Así mismo el uso de la tecnología SIG ha sido relevante como una herramienta de análisis geoespacial, ésta ha sido utilizada en diversos sectores, como lo es en la estimación del potencial de los recursos energéticos de una región, en este caso, el recurso eólico.

Figura 2. Una red bayesiana simple con tablas de probabilidad condicionales



Fuente: INEEL.

Trabajos relacionados

En Borunda, Jaramillo, Reyes y Ibarguengoytia (2016) se hace una revisión del estado del arte en el campo de las energías renovables utilizando las redes bayesianas (RB), ya que éstas por sus características, pueden resolver cuestiones con un cierto grado de incertidumbre, lo que dificulta el análisis mediante otras técnicas de IA. En el apartado correspondiente a la energía eólica se mencionan numerosos trabajos en cuanto a predicción del recurso eólico, operación, diagnóstico de fallas, mantenimiento y la gestión de los aerogeneradores en un parque eólico. Algunos de los trabajos descritos en Borunda, Jaramillo, Reyes y Ibarguengoytia (2016) y los cuales ameritan mencionarse por el tema propuesto son: De la Torre, Soto, Guevara y Rico (2011), se utilizaron RB para determinar las dependencias espaciales y temporales entre las variables climáticas. El conjunto de datos que utilizaron fue: temperatura, punto de rocío, humedad, presión y velocidad del viento. Dentro de un periodo de tres años y en tres sitios del estado de Querétaro, los datos estudiados fueron proporcionados por la comisión estatal de aguas. Para la estructura de la red bayesiana se ocupó el método de aprendizaje del algoritmo K2, esto con la finalidad de cuantificar la probabilidad condicional, lo que permite observar las relaciones causales entre las variables que se analizaron. Ibarguengoytia et al. (2014), desarrollaron un modelo de red bayesiana dinámica para el pronóstico de la predicción del viento basados en datos históricos en la estación experimental de la Ventosa en Oa-

xaca. Se propusieron en el modelo todas las variables relevantes en el dominio eólico para producir una distribución de probabilidad para las predicciones. El modelo dinámico de la red bayesiana fue comparado con otros métodos de pronósticos establecidos, los resultados que se obtuvieron mostraron resultados satisfactorios, observándose una mejora con respecto al usar técnicas clásicas. En Li y Shi (2012), se resumen las diferentes teorías básicas sobre los métodos bayesianos y se hace una revisión de la literatura acerca de estos para los sistemas que trabajan en el sector eólico. Entre los modelos de RB se encuentran: teorema de Bayes, modelo jerárquico y red bayesiana, red neuronal bayesiana, modelo bayesiano de selección y de promedio. Mientras que algunas de las aplicaciones actuales en el sector eólico, se pueden encontrar en: evaluación y estimación del recurso eólico, pronósticos a corto plazo, modelado del viento extremo, evaluación de la fiabilidad de los componentes del aerogenerador, entre otros. El trabajo de Jiang, Song y Kusiak (2012), presentan un nuevo método de series de tiempo para pronósticos de la velocidad del viento en un corto plazo, el cual está centrado en el pronóstico durante minutos, horas o días adelantados. Este modelo se basa en la teoría bayesiana y en el modelo de ruptura estructural, dicho modelo predice no sólo un valor, sino un conjunto de valores posibles, lo cual lo hace diferente a los modelos clásicos de series temporales. Como prueba del modelo de pronósticos desarrollado por medio de las RB se comprueba con series temporales de velocidad del viento en el mundo real, el modelo propuesto supera a todos los demás modelos probados con excepción de la predicción que está basada en redes neuronales NN.

De igual manera existen trabajos donde emplea la tecnología SIG, por ejemplo:

- Dlamini (2011) se realizó un estudio para estimar el riesgo de incendio en Suazilandia utilizando un SIG. Las áreas de estudio de incendios activos, así como áreas quemadas con anterioridad se obtuvieron del MODIS6 (imágenes de satélite de alta resolución) con las que se evaluaron un total de 13 variables explicativas biofísicas y socioeconómicas, a través de una red bayesiana la cual utilizó el algoritmo de aprendizaje basado en maximización de expectativas. La evaluación de exactitud y comparación con los mapas de riesgo de incendio activo y área quemada fueron de 93.14% y 96.16%, lo que permite un mayor control de incendios, al poder dividir el país en zonas de acuerdo a las probabilidades de que surja un incendio.
- De igual manera, Abudureyimu, Hayashi, Litifu y Nagasaka (2012) utilizaron la tecnología SIG para analizar el potencial eólico que rodea la costa de Japón, la razón de usar GIS es que cuando se decide el emplazamiento de un parque eólico, es necesario contar con los mapas de: distribución de la velocidad del viento, las condiciones sociales y las condiciones naturales marinas. Esto

permite determinar de forma óptima los sitios adecuados en alta mar para la instalación y puesta en marcha de una parque eólico. Los resultados de la investigación arrojan que las máquinas adecuadas para las costas de Japón son de una potencia de 2MW e instaladas a 100 metros de la línea de costa, donde se detectó un promedio de velocidad del viento de 6m/s a 70m de altura.

- Sener y CFE (2016) elaboraron un atlas del país donde se encuentran las zonas con amplio potencial de energías limpias (solar, eólica, geotérmica y biomasa), en esta aplicación se puede visualizar el potencial eólico de México a través de la tecnología SIG, en la que por medio de capas se muestran diferentes tipos de mapa con ciertas características (alto potencial, alta calidad, técnico-económicas, ambientales, sociales, por riesgo e informativas) que permiten al usuario visualizar diferentes entornos de medición. El usuario puede personalizar la visualización de acuerdo a sus intereses ya que cuenta con mapas bases como imagen satelital, por calles, satelital/calles, topográficos y batimétrico. Cuenta con un manual de usuario para aquellos nuevos usuarios que quieran interactuar con el sig. Por último, Bassamzadeh y Ghanem (2017) proponen un modelo de predicción de carga probabilística utilizando RB para la predicción del consumo energético en edificios residenciales, este modelo aprende las dependencias entre variables sin hacer suposiciones previas (*machine learning*). Se estudia el impacto de los precios en tiempo real sobre el comportamiento del consumo de los clientes. Las pruebas se realizan tomando datos que son proporcionados por Pacific Northwest National Lab (PNNL) para investigar el rendimiento del modelo en diferentes niveles de agregación espacio temporal.

Comparación con trabajos relacionados

La presente propuesta de investigación, ha sido comparada con diversos trabajos. En la tabla 1 se muestra las claves de referencia de los trabajos relacionados.

Tabla 1. Claves de referencia de los trabajos relacionados

Clave	Nombre del Trabajo
T1	“Bayesian networks in renewable energy systems: A bibliographical survey” (Borunda, Jaramillo, Reyes y Ibarguemgoytia, 2016).
T2	“Bayesian networks for defining relationships among climate factors” (De la Torre, Soto, Guevara y Rico, 2011).

(continuación)

Clave	Nombre del Trabajo
T3	“Wind power forecasting using dynamic Bayesian models” (Ibargüengoytia et al., 2014).
T4	“Applications of Bayesian methods in wind energy conversion systems” (Li y Shi, 2012).
T5	“Very short-term wind speed forecasting with Bayesian structural break model” (Jiang, Song y Kusiak, 2012).
T6	“Application of Bayesian networks for fire risk mapping using GIS and remote sensing data” (Dlamini, 2011).
T7	“Off-shore wind power potential evaluation and economy analysis of entire Japan using GIS technology” (Abudureyimu, Hayashi, Litifu y Nagasaka, 2012).
T8	“Atlas de zonas con energías limpias” (Sener y CFE, 2016).
T9	“Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks” (Bassamzadeh y Ghanem, 2017).
P1	“Protocolo de tesis maestría Universidad Autónoma de Querétaro” (Cantón, 2018).

Fuente: Elaboración propia.

Para realizar la comparación se consideraron cinco características: técnica de IA empleada, procesamiento de imágenes, si describe o no el área de interés, si menciona por quien fue proporcionado los datos del viento, y si se analiza o no el potencial eólico. En la tabla 2 se muestra la comparación de los trabajos relacionados.

Tabla 2. Comparación de trabajos relacionados

Trabajo	Técnica de IA	Procesamiento de imágenes	Área de interés	Datos del viento	Analiza el potencial
T1	Sí, redes bayesianas	No	No	No	Sí
T2	Sí, redes bayesianas	No	Sí, estado de Querétaro	Sí, Comisión Estatal de Aguas	Sí
T3	Sí, redes bayesianas	No	Sí, La Ventosa, Oaxaca	Sí, CERTE	Sí
T4	Sí, redes bayesianas	No	No	No	Sí

(continuación)

Trabajo	Técnica de IA	Procesamiento de imágenes	Área de interés	Datos del viento	Analiza el potencial
T5	Sí, redes bayesianas	No	Sí, Jiangsu, China	Sí, anemómetros de los aerogeneradores	Sí
T6	Sí, redes bayesianas y <i>machine learning</i>	Sí, GIS e imágenes satelitales	v, Suazilandia, África del Sur	No aplica	No estima el riesgo de los incendios
T7	No, distribución Weibull	Sí, GIS	Costas de Japón	Departamento Hidrográfico de la Guardia Costera de Japón	Sí
T8	No, método no indicado	Sí, GIS e imágenes satelitales	Sí, República mexicana	Sí, Cener y CFE	Sí
T9	Sí, redes bayesianas y <i>machine learning</i>	No	Sí, Olympic, Península Washington	No aplica	No, predice la demanda de energía eléctrica
P1	Sí, redes bayesianas y <i>machine learning</i>	Sí, GIS e imágenes satelitales	Sí, estado de Querétaro y micrositos del mismo	Sí, estaciones meteorológicas	Sí

Fuente: Elaboración propia.

Se puede concluir de acuerdo al análisis de los trabajos relacionados lo siguiente:

1. Las RB han sido utilizadas para analizar el potencial eólico por diversos métodos de las mismas, en algunos casos apoyado con otra área de la IA como las redes neuronales y *machine learning*.
2. De acuerdo al estado del arte encontrado hasta el momento, el trabajo propuesto es el único que utilizaría las RB junto con el procesamiento de imágenes (SIG e imágenes aéreas) para analizar el potencial eólico.
3. La mayoría de los casos cuanta con un área de interés en donde se planea hacer uso de las diferentes áreas de la IA, en este caso las redes bayesianas.

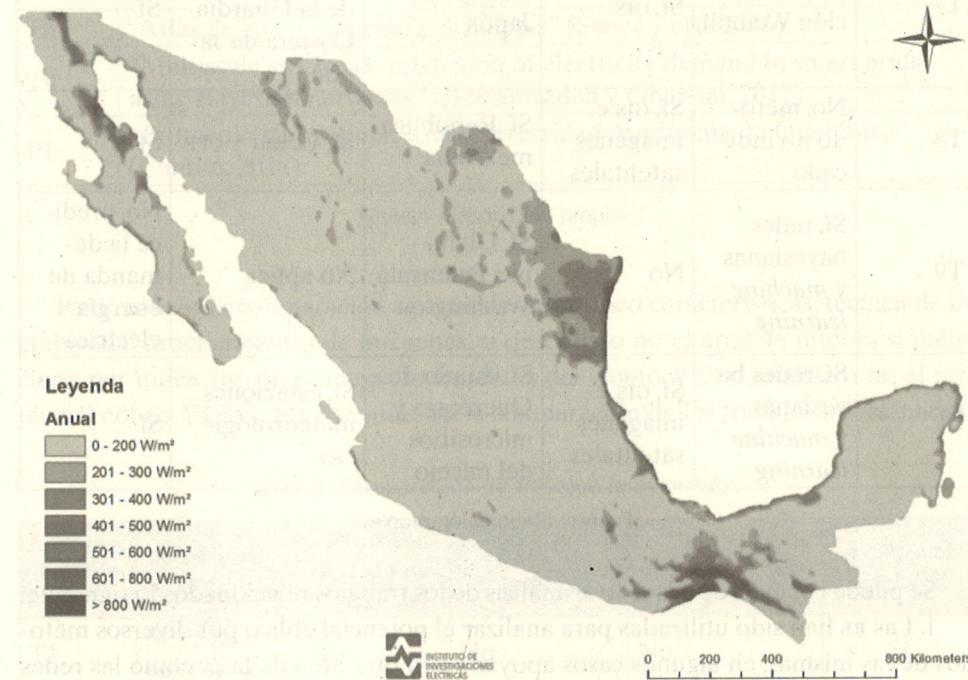
4. En algunos casos se cuenta con datos de viento de por los menos dos ciclos anuales del recurso eólico y generalmente es del área de interés.

5. Dos tercios de los trabajos tiene como finalidad analizar el potencial eólico.

Mapas y modelos considerados

Los mapas eólicos se crearon a partir de datos generados por el Centro de Ciencias Atmosféricas de la UNAM utilizando el modelo de mesoescala MM. Los datos se generaron cada hora con una resolución de 4,5 km, a 50 m y 80 m de altura, para todo el país.

Figura 3. Potencia eólica anual promedio del mapa a 50m de altura



Fuente: Elaboración propia con datos del INEEL.

Modelos numéricos de predicción de tiempo (NWP)

Los modelos numéricos del pronóstico del tiempo son modelos matemáticos que resuelven las ecuaciones que describen el comportamiento de la atmósfera. Estos modelos proporcionan el estado de las variables meteorológicas en un momento posterior, tomando los valores actuales de esas variables llamadas condiciones iniciales. Lezaun (2002) describe muy bien la evolución de los modelos numéricos de la predicción meteorológica. Menciona que, en 1903, Vilhelm Bjerknes propuso tratar el comportamiento de la atmósfera basándose en las leyes físicas de la termodinámica y la mecánica de fluidos, sin embargo, en ese momento, como el mismo Bjerknes dijo, su modelo requería resolver un sistema de ecuaciones en derivadas parciales no lineales que representaban una gran dificultad. A partir de este trabajo, durante los siguientes años se realizaron trabajos que respaldaron y mejoraron lo propuesto por Bjerknes. En una revisión de la literatura enfocada a modelos básicos de obtención de datos meteorológicos, Richardson el autor publica los resultados de su método numérico basado en las ecuaciones propuestas por Bjerknes, con las cuales tuvo un pronóstico (sin buenos resultados) durante seis horas, lo que le llevó seis semanas realizar los cálculos necesarios. Hasta 1950, cuando se probó el modelo barotrópico propuesto por Rossby (1940) en una computadora, la contribución de Rossby fue agregar la ecuación del equilibrio del componente vertical de la vorticidad a partir de varias aproximaciones en las ecuaciones de Richardson. La computadora utilizada fue ENIAC (Analizador Numérico Integrador Electrónico y Computadora), la primera computadora electrónica que se desarrolló en 1946 en la Universidad de Pensilvania. John von Neumann, impulsor del desarrollo informático, decidió probar la máquina con un gran problema científico y junto con Jule Charney y Ragnar Fjörtoft ejecutaron el modelo de Rossby para predecir el tiempo hasta 24 horas con resultados muy alentadores.

Ecuaciones primitivas

Para el pronóstico de la PNT, intentan describir el estado y el comportamiento de la atmósfera terrestre resolviendo un conjunto de ecuaciones diferenciales en derivadas parciales que traducen las leyes generales de la física que rigen esta atmósfera utilizando un modelo matemático. Desde un estado inicial en el que se conocen los valores de las variables atmosféricas, este conjunto de ecuaciones se puede resolver para calcular la evolución de estas mismas variables. Una representación simple es el siguiente modelo de muestra, dada en la ecuación 1:

$$\frac{\partial A}{\partial t} = F(A), A(0) = A_0 \quad (1)$$

Donde A es una variable meteorológica que cambia las horas extraordinarias debido a la acción de F. La variable F se llama forzado, combinando todos los tipos de acciones de cambio sobre A. La solución de Eq. (1) tiene el siguiente formato:

$$A_p = A_0 + \int AF(A) dt \quad (2)$$

Este proceso se sigue para configurar las ecuaciones para PNT (Lighezzolo, 2014).

Los modelos atmosféricos se desarrollan sobre la base de 7 ecuaciones fundamentales: ecuación de conservación del movimiento (una ecuación para cada componente de velocidad):

$$\frac{dv}{dt} = -\frac{1}{\rho p} + g^* + F - 2\Omega \times V - \Omega \times (\Omega \times r) \quad (3)$$

$$V = \frac{Hk}{Rt} Bw \quad (4)$$

La ecuación de continuidad:

$$\frac{\partial \rho}{\partial t} = \rho v \quad (5)$$

La primera ley de:

$$Q = C_p \frac{dT}{dt} - \alpha \frac{dp}{dt} \quad (6)$$

La ecuación de estado:

$$P\alpha = RT \quad (7)$$

La ecuación de conservación del vapor de agua:

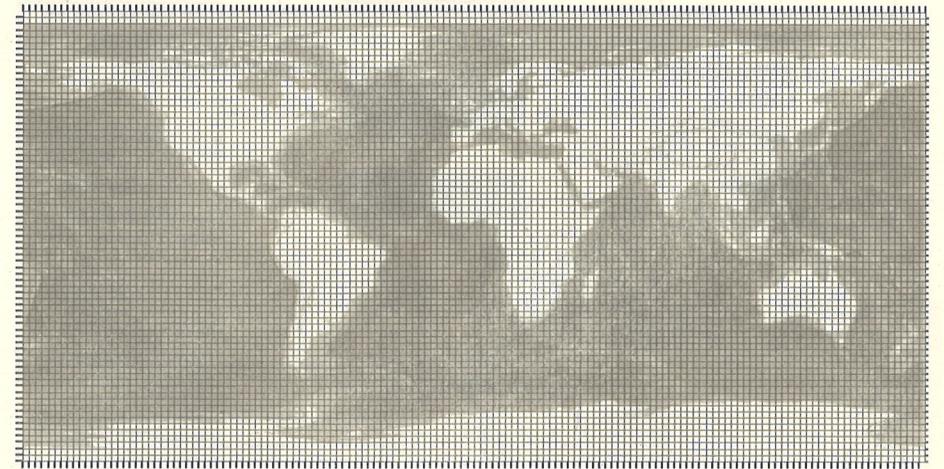
$$\frac{dq}{dt} = E - C \quad (8)$$

Según Kalnay (2003) el resultado son siete ecuaciones con siete incógnitas $v = (u, v, w)$, T, p, ρ en algunos α en otros, y q .

Tipos de modelos de tiempo numéricos

Los PNT se pueden clasificar en términos de su resolución espacial, temporal y el rango del área que se modelará, por ejemplo, modelando la atmosférica (véase figura 4).

Figura 4. Grilla global de 2.5 grados de resolución horizontal utilizada por modelos de circulación general en todo el mundo; el tamaño de la cuadrícula es $144 \times 73 = 10512$ puntos



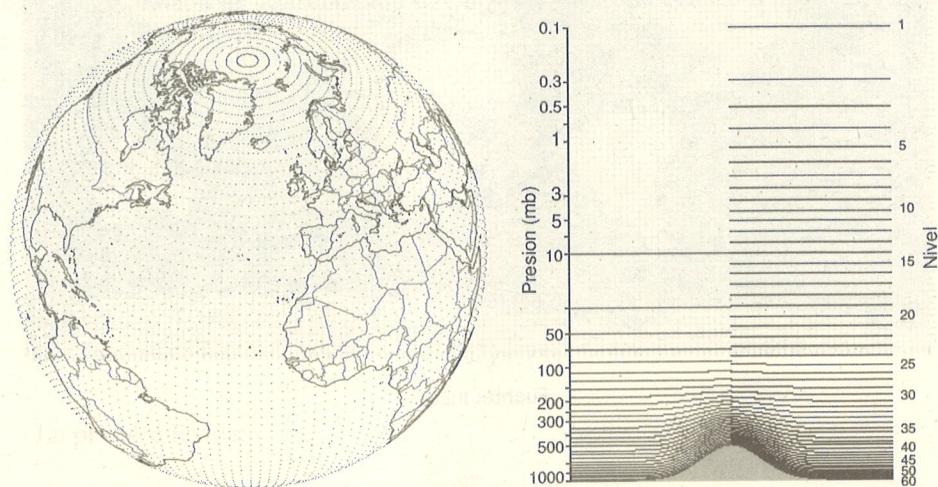
Fuente: INEEL.

Modelos globales

Los modelos globales calculan el estado de la atmósfera para todo el planeta, por lo que su representación se da en coordenadas geográficas. Estos modelos necesitan condiciones iniciales generalmente dadas por un modelo de asimilación de datos, que toma los valores de las variables atmosféricas disponibles alrededor del planeta. Estos valores se obtienen por estaciones meteorológicas, radio sondeo, boyas, entre otros. El desarrollo de estos modelos empleando una resolución espacial muy baja, emplea la aproximación hidrostática de la atmósfera, es decir, no resuelven la ecuación de movimiento vertical completo, pero los movimientos verticales se parametrizan de acuerdo con el equilibrio *supuesto* entre el peso de la atmósfera y el gradiente de presión vertical. Estos modelos, además de las condiciones iniciales, también utilizan una parametrización del terreno y el uso de la tierra, donde los

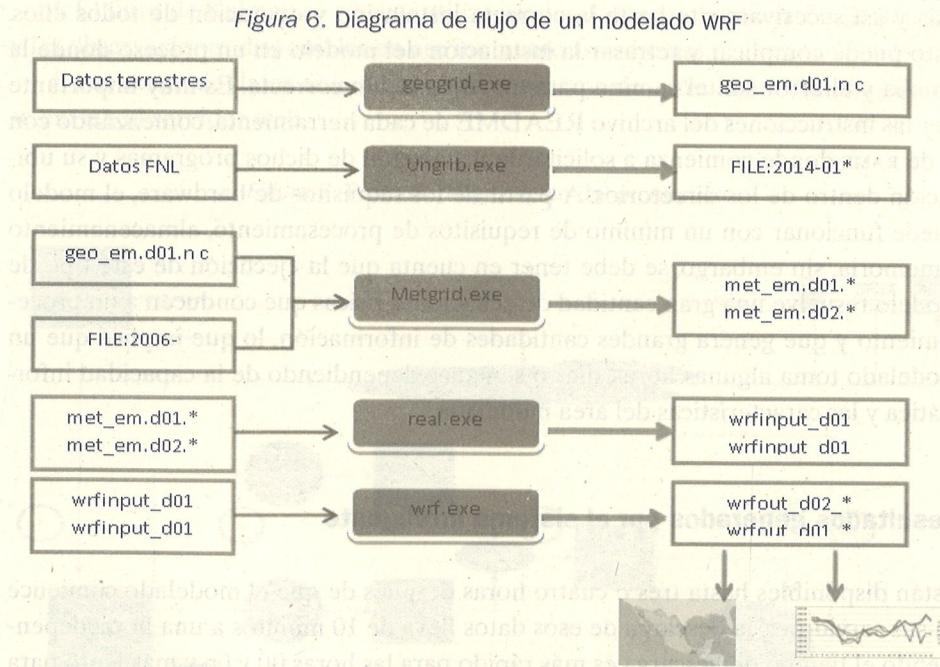
valores se extrapolan al nivel de resolución de la grilla del modelo, por lo que los valores de estos son muy suaves. Por otro lado, los modelos globales son considerados el núcleo del pronóstico del clima ya que estos proporcionan las condiciones iniciales y de frontera para los modelos de área limitada. En la actualidad, varias instituciones han desarrollado sus propios modelos, que se caracterizan por su resolución espacial horizontal (véase figura 5).

Figura 5. Cuadrícula general de 2.5 grados de resolución horizontal utilizada por los modelos de circulación general en todo el mundo. El tamaño de la cuadrícula es $144 \times 73 = 10512$ puntos (Derecha). Dos perfiles verticales con 31 y 60 niveles de altura geopotencial (expresados en milibares mb, y en los niveles de modelo, respectivamente). La altura máxima mostrada (0.1 mb) es de aproximadamente 64 km



El módulo WRF contiene las ecuaciones de pronóstico físico y de diagnóstico que permiten hacer una predicción con un horizonte de tiempo preestablecido. Este módulo también incluye las parametrizaciones que permiten describir los fenómenos físicos cuya escala espacial / temporal es menor que la resolución fija en las simulaciones (Ver figura 6.).

Diagrama de flujo de WRF:



Fuente: Elaboración propia.

Diseño e implementación de la herramienta inteligente

El modelado adecuado de una herramienta inteligente para la toma de decisiones asociada con la predicción de potencial eólico requiere de diversos tipos de análisis multivariable, al utilizar redes bayesianas, se pueden considerar diversos aspectos asociados con la disminución de incertidumbre, es por ello, que un sistema de medición de calidad del viento para granjas de viento, basado en su confiabilidad, disponibilidad, sustentabilidad y capacidad de servicio es *Reliability, Availability, Maintainability* y *Serviceability* (RAMS, por sus siglas en inglés,) funciona sólo en sistemas operativos UNIX y Linux. La mayoría del código del modelo está escrito en Fortran y requiere la instalación del compilador Fortran 90. Algunas partes están escritas en código C para facilitar los procesos de entrada / salida, requiriendo la instalación del compilador de C también. Además de estos dos programas es necesario instalar varios más, que son necesarios ya que cada una de las herramientas requeridas para operar el modelo está instalada, esto genera un proceso en cadena donde una herramienta solicita la instalación de otra y esta solicita la instalación de otros

más y así sucesivamente hasta la correcta instalación y operación de todos ellos. Esto puede complicar y retrasar la instalación del modelo en un proceso donde la prueba y el error son el camino para una instalación correcta. Es muy importante leer las instrucciones del archivo README de cada herramienta, comenzando con la de RAMS donde comienza a solicitar la instalación de dichos programas y su ubicación dentro de los directorios. A partir de los requisitos de hardware, el modelo puede funcionar con un mínimo de requisitos de procesamiento, almacenamiento y memoria, sin embargo, se debe tener en cuenta que la ejecución de este tipo de modelo resuelve una gran cantidad de ecuaciones físicas que conducen a un procesamiento y que genera grandes cantidades de información, lo que implica que un modelado toma algunas horas, días o semanas dependiendo de la capacidad informática y las características del área modelada.

Resultados generados por el sistema inteligente

Están disponibles hasta tres o cuatro horas después de que el modelado comience en sus servidores, la descarga de esos datos lleva de 10 minutos a una hora, dependiendo el tiempo de descarga es más rápido para las horas 00 y 06 y más lento para 12 y 18 porque en la primera se utiliza la red INEEL que está menos ocupada. Los modelos WRF y RAMS (Regional Atmospheric Modeling System) tardan aproximadamente una hora para el pronóstico, lo que hace que un total de entre 5 y 6 horas tenga un pronóstico aproximadamente a las 24, 72 o 120 horas. Este desplazamiento hace que, cuando se hayan modelado los datos medidos en cinco o seis horas del período de pronóstico, los datos reales ya estén disponibles para validación, lo que permite que los datos se usen en el proceso de asimilación con el filtro de Kalman y para hacer un ajuste al pronóstico. Después de modelar e interpolar los datos a las alturas correspondientes, se exportan como conjuntos de datos en archivos ASCII al servidor web que contiene el sistema de consulta. El sistema de consulta reemplaza los nuevos datos de pronóstico para estar disponibles en las diferentes formas de consulta proporcionadas. Este proceso se muestra en la figura 7.

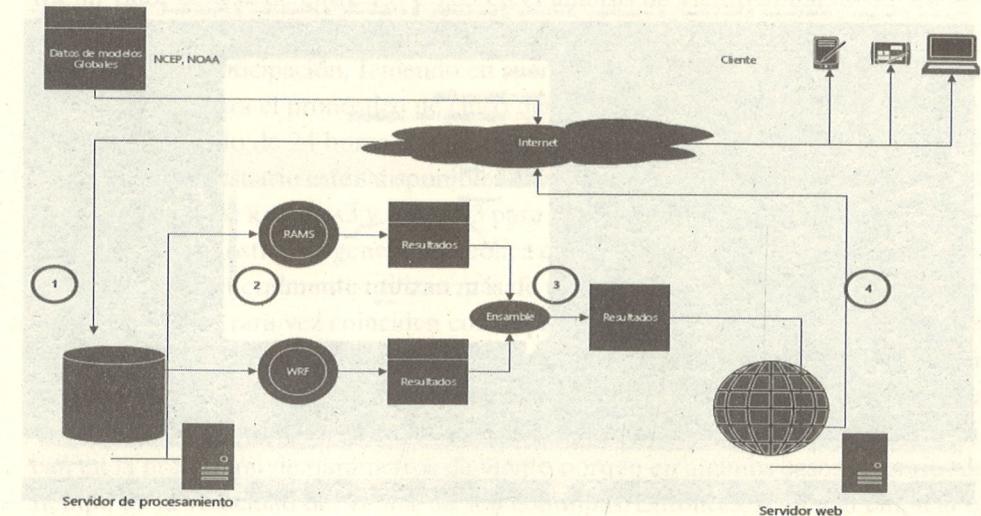
1. Una secuencia de comandos programada en el shell de Linux se ejecuta a través del administrador de tareas crontab de Linux para descargar de los servidores NCEP los datos meteorológicos globales utilizados como las condiciones iniciales de cada modelo de clima.

2. Ambos modelos se ejecutan en cada servidor usando procesamiento paralelo. Cada modelo ejecuta sus módulos de preprocesamiento, simulación y posprocesamiento.

3. El filtro de Kalman se ejecuta utilizando los datos predichos por cada modelo y las observaciones disponibles en ese momento (seis horas).

4. Las series de datos se generan para el sistema de consulta en formato ASCII y se envían al servidor web.

Figura 7. Proceso de ejecución del pronóstico



Fuente: Elaboración propia.

Desarrollo del sistema de consulta

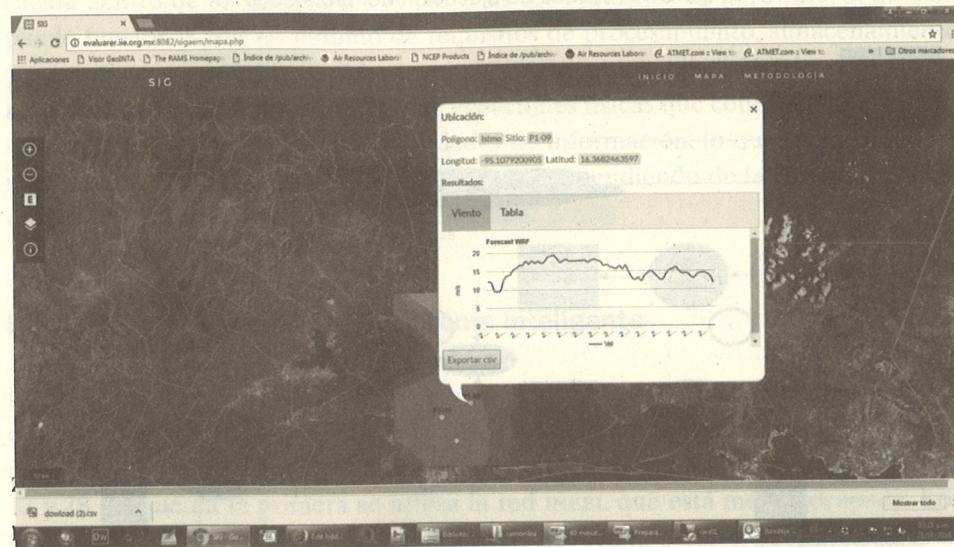
Se desarrolló un sistema de consulta basado en la web que permite la búsqueda y descarga de datos de una determinada ubicación geográfica dentro del área modelada. El sistema se implementó en la siguiente plataforma:

- Servidor web HP, con el sistema operativo Windows Server 2008.
- Servidor web Apache Server.
- Servidor de mapas GeoServer, para la publicación de servicios web de mapas y la georreferenciación de información.
- Código escrito en HTML, PHP y Java Script.

Como se puede observar en la figura 8 la aplicación web le muestra para acceder a ella, una ventana distribuida en un banner y barra de navegación en la parte superior, y un área de visualización en la parte inferior, esta área cambiará de acuer-

do con el botón seleccionado en la barra de navegación, puede mostrar la página de inicio, la descripción de la metodología utilizada en el pronóstico o puede visualizar el mapa de la ubicación de un sitio y consultar el pronóstico.

Figura 8. Consulta de pronósticos



Para realizar una consulta, el usuario debe seleccionar el icono de identificación en la barra de herramientas y hacer clic en cualquier ubicación del mapa, los resultados pueden ser tres: si hace clic en uno de los cuatro puntos de las estaciones, se mostrará una ventana con el nombre de la estación, la ubicación geográfica, el gráfico de pronóstico para esa zona y la tabla de datos, esta ventana contiene un botón para descargar los datos en formato csv. Si el usuario hace clic en la superficie de modelado, se mostrará la ubicación y el valor de la velocidad media anual del viento. Si el usuario hace clic en otra parte del mapa, se mostrará la ventana, pero con la leyenda de que no hay datos en ese punto. El desarrollo del sistema inteligente permite utilizar un Modelo RAMS (Regional Atmospheric Modeling System).

Conclusiones, recomendaciones y trabajo futuro

De acuerdo con los resultados, el mejor modelo de pronóstico para el período de cinco días es el WRF con un RMSE promedio de 2.48 y un MAE promedio de 1.7. Utili-

zando el Modelo RAMS (Regional Atmospheric Modeling System) para el pronóstico de 24 horas, el mejor modelo fue el RAMS-KA3, que es la aplicación del filtro Kalman a las salidas del modelo RAMS, este modelo muestra los valores promedio RMSE en 1.74 t y un MSE de 1.32. Estos valores son similares a los resultados mostrados por otros trabajos. Cprore (2015) presenta un valor promedio de RMSE de tres para los pronósticos de siete días por adelantado, C. Chancham et al. (2014) muestra un valor de MSE de entre 1,6 y 5,83 en el análisis de viento anual medio en seis estaciones Louka et al. (2008) muestra valores RMSE de 3.36 y 2.25 para pronósticos de 48 horas de anticipación. Teniendo en cuenta que según los resultados el segundo mejor modelo para el pronóstico de cinco días es el RAMS y el segundo mejor modelo para el período de 24 horas es el WRF-KA3, el sistema de consulta, hará que los pronósticos del usuario estén disponibles WRF y RAMS para el período de cinco días y las previsiones de RAMS-KA3 y WRF-KA3 para el período de 24 horas. Muchos de los modelos de pronóstico de generación eólica que utilizan los resultados de los modelos numéricos generalmente utilizan más de un pronóstico de estos, porque los modelos numéricos rara vez coinciden con el pronóstico (Galanis, Louka, Katsafados, Kallos y Pytharoulis, 2006) argumentan que, aunque el método del filtro de Kalman se ha utilizado ampliamente con el fin de mejorar el pronóstico del tiempo, la forma lineal del algoritmo puede afectar significativamente los resultados cuando se aplican en la predicción de parámetros de viento porque en algunos casos las series de tiempo de la velocidad del viento no son continuas. Entonces, si bien al emplear el filtro Kalman para la mejora de la predicción de la temperatura del aire es exitosa, puede que no sea bueno para la predicción de la velocidad del viento, esto puede verificarse con los resultados de este trabajo porque, aunque los resultados del pronóstico a 24 horas mejoran con el Kalman Filtro utilizado en las salidas RAMS, este no es el caso para las salidas WRF. Uno de los aspectos más relevantes es poder realizar, una adecuada predicción numérica del potencial eólico, lo que permitirá determinar la competitividad de una región, en nuestra investigación determinamos que Queretano, y en especial la Sierra Gorda de Querétaro, tienen un adecuado potencial para ello.

Referencias

ABUDUREYIMU, A., Hayashi, Y., Litifu, Z., y Nagasaka, K. (2012). Off-shore wind power potential evaluation and economy analysis of entire Japan using GIS technology. *The 2011 International Conference on Advanced Mechatronic Sys-*

- tems (pp. 86-93). Zhengzhou, China: IEEE. Recuperado de doi:10.1504/IJA-MECHS.2012.048397
- AMDEE. (2015). Parques eólicos de México 2015. Sitio oficial de la Asociación Mexicana de Energía Eólica. Recuperado el 8 de noviembre de 2015 de <http://www.amdee.org/parques-eolicos-mexico-2015>.
- AMES, D. P., y Anselmo, A. (2015). Bayesian network integration with GIS. En S. Shekhar, H. Xiong y X. Zhou, *Encyclopedia of GIS* (pp. 1-8). Cham, Suiza: Springer International Publishing. Recuperado de doi:10.1007/978-0-387-35973-1_95
- ANEMOS. (2011, 28 de enero). *The state of the art in short-term prediction of wind power. A literature overview*. Roskilde, Dinamarca.
- BASSAMZADEH, N., y Ghanem, R. (2017). Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. *Applied Energy*, 193, 369-380. Recuperado de doi:10.1016/j.apenergy.2017.01.017
- BERNHARDSEN, T. (2002). *Geographic information systems: An introduction*. New Jersey: John Wiley & Sons.
- BORUNDA, M., Jaramillo, O., Reyes, A., y Ibarguengoytia, P. (2016). Bayesian networks in renewable energy systems: A bibliographical survey. *Renewable and Sustainable Energy Reviews*. Recuperado de doi:10.1016/j.rser.2016.04.030
- CHAI, T., y Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. Recuperado de <http://dx.doi.org/10.5194/gmd-7-1247-2014>.
- CHANCHAM, C., Waewsak, J., Chaichana, T., Landry, M., y Gagnon, Y. (2014). Assessment of onshore wind energy potential using regional atmospheric modeling system (RAMS) for Thailand. *Energy Procedia*, 52, 487-496. Recuperado de <http://dx.doi.org/10.1016/j.egypro.2014.07.102>
- CONAGUA. (2010). *Manual teórico práctico del observador de superficie*. México: Secretaría de Medio Ambiente y Recursos Naturales.
- CPRORÉ. (2015). Servicio de predicción de la velocidad de viento y producción de energía a corto plazo. Recuperado de www.cproré.com.
- DE LA TORRE, G., Soto, G., Guevara, R., y Rico, E. (2011). Bayesian networks for defining relationships among climate factors. *International Journal of the Physical Sciences*, 6(18), 4412-4418. Recuperado de doi:10.5897/IJPS11.631
- DLAMINI, W. (2011). Application of bayesian networks for fire risk mapping using GIS and remote sensing data. *GeoJournal*, 76(3), 282-296. doi:10.1007/s10708-010-9362-x
- DORADO, J., y Ruíz, J. F. (2013a). Manual de usuario ApplyKalmanFilter_Linux. Grupo de Modelamiento de Tiempo y Clima. Subdirección de Meteorología. IDEAM.
- _____. (2013b). Implementación de filtros de Kalman como método de ajuste a los modelos de pronóstico (GFS) de temperaturas máximas y mínima para algunas ciudades de Colombia. Grupo de Modelamiento de Tiempo y Clima. Subdirección de Meteorología. IDEAM.
- ELLIOT, D., Holladay, C., Barchet, W., Foote, H., y Sandusky, W. (2009). Wind energy resource of the United States. NREL (National Renewable Energy Laboratory). Obtenido de <http://rredc.nrel.gov/wind/pubs/atlas/tables/1-1T.html>
- FECYT. (2004). Meteorología y climatología. Madrid, España: Villena Artes Gráficas.
- FERNÁNDEZ JIMÉNEZ, L. (2008). Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos [Tesis doctoral]. Universidad de La Rioja.
- GALANIS, G., Louka, P., Katsafados, P., Kallos, G. y Pytharoulis, I. (2006). Applications of Kalman filters based on non-linear functions to numerical weather predictions. *Annales Geophysicae*, 24, 1-10.
- GARCIA-MOYA, J. A., Callado, A., Santos, C., Santos-Muñoz, D., y Simarro, J. (2007). El sistema de predicción por conjuntos para el corto plazo del INM. *Asociación Meteorológica Española*, 5(16), 22-27.
- GONZÁLEZ, J. (2009). *Energías renovables*. Barcelona: Reverté.
- GUTIÉRREZ PULIDO, H., y Vara Salazar, R. (2012). *Análisis y diseño de experimentos* (3rd ed.). México: McGraw-Hill Interamericana.
- GUTIÉRREZ, J., Cano, R., Cofiño, A., y Sordo, C. (2004). Técnicas de aprendizaje automático. En *Redes probabilísticas y neuronales en las ciencias atmosféricas* (pp. 127-130). Madrid, España: Ministerio de Medio Ambiente.
- HERNÁNDEZ DEL VALLE, A., y Hernández-Lerma, O. (2003). *Elementos de probabilidad y estadística*. México: Sociedad Matemática Mexicana.
- IBARGÜENGOYTIA, P., Reyes, A., Romero, I., Pech, D., García, U., Sucar, L., y Morales, E. (2014). Wind power forecasting using dynamic bayesian models. *Nature-Inspired Computation and Machine Learning. MICA 2014* (pp. 184-197). Cham, Suiza: Springer. Recuperado de doi:10.1007/978-3-319-13650-9_17
- JIANG, Y., Song, Z., y Kusiak, A. (2012). Very short-term wind speed forecasting with bayesian structural break model. *Renewable Energy*, 637-647. Recuperado de doi:10.1016/j.renene.2012.07.041
- JORBA, C. O. (2005). Simulación de campos de viento de la península ibérica y el área geográfica de Catalunya con alta resolución espacial para distintas situaciones

- meteorológicas típicas [Tesis doctoral]. Programa de Doctorado en Ingeniería Ambiental, Universidad Politécnica de Catalunya. Barcelona, España.
- KALNAY, E. (2003). Atmospheric modeling, data assimilation and predictability. Harvard: Cambridge University Press.
- LAZAREVSKA, E. (2016). Wind speed prediction with extreme learning machine. *2016 IEEE 8th International Conference on Intelligent Systems* (pp. 154-159). Sofia, Bulgaria: IEEE.
- LEZAUN, M. (2002). Predicciones del tiempo y matemáticas departamento de matemática aplicada, estadística e i. o. Universidad del País Vasco-Euskal Herriko Unibertsitatea.
- LI, G., y Shi, J. (2012). Applications of bayesian methods in wind energy conversion systems. *Renewable Energy*, 43, 1-8. Recuperado de doi:10.1016/j.renene.2011.12.006
- LIGHEZZOLO, R. A. (2014). Integración de modelos numéricos de predicción meteorológica en sistemas de alerta temprana a emergencias [Tesis de maestría]. Universidad Nacional de Córdoba, Córdoba, Argentina.
- LOUKA, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., y Kallos, G. (2008). Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 2348-2362.
- MONTEIRO, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., y Conzelmann, G. (2009). Wind power forecasting: State-of-the-Art 2009 [Reporte técnico]. U.S. Department of Energy Office of Scientific and Technical Information.
- OBANDO-MONTAÑO, A. (2014). An approach to determine the Weibull parameters for wind energy analysis: The case of Galicia (España). *Energies*, 7, 2676-2700.
- WAGNER, H.-J., y Jyotirmay, M. (2013). Introduction to wind energy systems. Berlin: Springer. Recuperado de doi:10.1007/978-3-642-32976-0.

II Bases de datos en la educación

En este libro se recopilan un conjunto de aplicaciones que ilustran el funcionamiento del análisis de datos grandes (Big Data Analytics), éstas se encuentran en áreas de conocimiento diversas entre las que se incluyen: el sector eléctrico, el sector educativo, el sector salud, el gobierno abierto, la detección de faltantes y anomalías en las organizaciones, ciudades inteligentes, entre otras, todas ellas de relevancia para mejorar la toma de decisiones en las organizaciones. Así mismo se discuten los procesos fundamentales para el análisis de datos (Data Analytics) y el descubrimiento de conocimiento (Knowledge Discovery in DataBase) desde conceptos como: datos grandes (Big Data), la minería de datos (Data Mining), y la minería de datos sociales (Social data Mining). El libro está orientado a alumnos de pregrado y posgrado, así como investigadores y practicantes, que desean introducirse al conocimiento y aplicación del análisis de datos grandes en diversas áreas de las organizaciones. Big data es un tema de discusión actual dada la variedad de datos que se tienen, la velocidad con la que se producen, la veracidad y el volumen creciente de éstos generado con el uso de las TICs.

Big data/ Toma de decisiones



BONILLA
ARTIGAS
EDITORES



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

