

VISUALIZACIÓN DEL POTENCIAL DE RADIACIÓN SOLAR EN
EL DEPARTAMENTO DE NARIÑO UTILIZANDO SERIES DE
TIEMPO A PARTIR DE IMÁGENES SATELITALES

BAYRON ANTONIO CHAMPUTIZ MARTINEZ



UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE SISTEMAS
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2016

VISUALIZACIÓN DEL POTENCIAL DE RADIACIÓN SOLAR EN
EL DEPARTAMENTO DE NARIÑO UTILIZANDO SERIES DE
TIEMPO A PARTIR DE IMÁGENES SATELITALES

BAYRON ANTONIO CHAMPUTIZ MARTINEZ

TRABAJO DE GRADO PRESENTADO COMO REQUISITO
PARCIAL PARA OPTAR AL TÍTULO DE INGENIERO DE
SISTEMAS

DIRECTOR: OMAR ERNESTO CABRERA ROSERO, ING.

UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE SISTEMAS
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2016

NOTA DE RESPONSABILIDAD

“Las ideas y conclusiones aportadas en la tesis de grado, son responsabilidad exclusiva de sus autores”.

Artículo 1º del acuerdo N° 324 del 11 de octubre de 1966, emanado del Honorable Consejo Directivo de la Universidad de Nariño

“La Universidad de Nariño no se hace responsable de las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor”

Artículo 13, Acuerdo N. 005 de 2010 emanado del Honorable Consejo Académico.

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

San Juan de Pasto, 2016



You are free:



to Share – to copy, distribute and transmit this work



to Remix – to adapt this work

Under the following conditions:



Attribution – You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work)



Noncommercial – You may not use this work for commercial purposes.

Subject to conditions outlined in the license.

This work is licensed under the *Creative Commons Attribution-NonCommercial 3.0 Unported License*. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc/3.0/>

or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Dedicado a:

A mi madre Maria Martinez y a mi padre José Champutis, por todo el apoyo que fué brindado durante el transcurso de mi carrera, gracias por inculcarme valores que me permiten tener armonía en mi entorno social, por aconsejarme en de la mejor manera y estar presente en todas las dificultades.

A mis compañeros de estudio que me brindaron ayuda incondicional, fortaleza en momentos de debilidad, consejo y orientación en momentos de incertidumbre.

A todos los profesores que contribuyeron con conocimientos, enseñanzas, consejos y experiencias en los momentos necesarios y oportunos, en especial a los docentes Andres Oswaldo Calderon y Jimmy Mateo Guerrero.

A mi asesor, amigo y compañero Ing. Omar Ernesto Cabrera por brindar las experiencias y conocimiento pertinentes en la etapa de investigación.

Al Grupo de Investigación en Ingeniería Eléctrica y Electrónica (GIEE) por acogerme y brindarme apoyo pertinente en el transcurso de la investigación.

“Bayron Antonio Champutiz Martinez”

AGRADECIMIENTOS

El presente trabajo de grado está dedicado a mis padres por el apoyo a mi formación personal y profesional, por confiar siempre en mis cualidades y entregarme su apoyo incondicional, gracias porque siempre están presentes para apoyarme en cualquier dificultad.

Agradezco a Dios, por darme la fe y la fuerza necesaria en todo momento y permitirme creer que es posible culminar con éxito cualquier meta a pesar que existen momentos que se pueden tornar difíciles.

También está dedicado a todos aquellos amigos incondicionales que compartieron un momento de sus vidas conmigo y siempre me seguirán apoyando sin importar las circunstancias.

RESUMEN

Actualmente el la revolución de fuentes energéticas y los cambios acelerados de las condiciones climáticas obliga a las naciones a plantear un cambio a los enfoques de generación de energía, para esta serie de cambios se requiere un análisis de todas y cada una de las fuentes de energía renovables y en varias situaciones es pertinente resaltar las condiciones necesarias para optimizar el aprovechamiento del potencial energético. La actual investigación se centró en la construcción de una serie de tiempo en radiación solar presente en el departamento de Nariño, con el objetivo principal de brindar apoyo en el estudio de indicadores para radiación solar en el proyecto “Análisis de oportunidades energéticas con fuentes alternativas en el departamento de Nariño”, proyecto que realiza un estudio de diversas fuentes de energía presentes en el departamento.

La investigación comprende descargar, procesar y analizar 11 años de imágenes satelitales brindadas por el sensor MODIS y 15 años de imágenes satelitales brindadas por el sensor Landsat, almacenar en una base de datos las bandas necesarias para la identificación de radiación solar, aplicar modelos para la construcción de mapas de radiación solar, series de tiempo de radiación solar y nubosidad, realizar un análisis enfocado en la detección de patrones, análisis del comportamiento de la radiación solar y nubosidad en el departamento de Nariño.

Adicionalmente se pretende establecer y compartir los diferentes tipos de información y metodologías aplicadas a los interesados en la temática enfocada en el procesamiento y análisis de imágenes satelitales MODIS y LandSat.

Palabras clave: MODIS, LandSat, detección de patrones, detección de nubes, análisis de regresión, series de tiempo, patrones secuenciales.

ABSTRACT

Currently the Revolution energy sources and the rapid changes of weather conditions forces nations to propose a change to the approaches of power generation, for this series of changes required an analysis of each and every one of the sources of renewable energy; in some situations it is relevant to note the conditions to optimize use of the energy potential. Current research focuses on the construction of a time series of solar radiation present in the department of Nariño, with the main objective of supporting indicators in the study of solar radiation in the “Análisis de oportunidades energéticas con fuentes alternativas en el departamento de Nariño” project; This project is conducting a study of various sources of energy present in the department.

The research includes download, process and analyze 11 years of satellite images provided by MODIS sensor and 15 years of satellite images provided by Landsat sensor, storing in a database the necessary bands for detecting solar radiation, making a focused pattern detection analysis, to visualize the behavior of solar radiation on the department of Nariño.

Additionally it is to establish and share different types of information and methodologies applied to those interested in the topic focused on the processing and analysis of MODIS satellite image.

Keywords: MODIS, LandSat,, pattern detection, detection of clouds, regression analysis, Time series, sequential patterns.

CONTENIDO

INTRODUCCIÓN	13
1 FUNDAMENTOS TEÓRICOS	17
1.1 SERIES DE TIEMPO	17
1.1.1 Componentes	18
1.1.2 Análisis Grafico de una Serie de Tiempo	20
1.1.3 Suavizar la Serie de Tiempo	22
1.2 Bases de Datos Espacio-Temporales	24
1.3 Técnicas de Regresión	25
1.4 Redes Neuronales	27
1.5 Interpolación Espacial	30
1.6 Minería de Datos	33
2 TRABAJOS RELACIONADOS	37
3 METODOLOGÍA	45
3.1 Identificar y Adquirir Imágenes Satelitales	45

3.2 Procesar y Almacenar Información	48
3.3 Selección, Aplicación del Modelo de Regresión y Construcción de Serie de Tiempo	52
3.4 Mapas de Radiación Aplicando Interpolación Espacial	53
3.5 Detección de Patrones Secuenciales	56
3.6 Comportamiento de Nubes	61
4 VALIDACIÓN DE LA INFORMACIÓN CON ESTACIONES ACTIVAS	64
5 ANÁLISIS DE RESULTADOS	67
5.1 Análisis de Patrón y Nubes en un punto Específico	69
6 CONCLUSIONES	72
7 TRABAJOS FUTUROS	75
REFERENCIAS	80
ANEXOS	81

ACRÓNIMOS

MCO Minimos Cuadrados Ordinarios

CAD Computer Aided Design

DBMS Database Management System

GIS Geographic Information System

LCM Linear time Closed itemset Miner

MOD Moving Objects Databases

MOD09GA The MODIS Surface Reflectance products

MODIS Moderate-Resolution Imaging Spectroradiometer

MPS Minería de Patrones Secuenciales

RN Redes Neuronales

KDD Knowledge Discovery in Databases

INTRODUCCIÓN

Países como Alemania, España, Japón, Estado Unidos, Italia son las potencias mundiales en la generación de energía renovable mediante el uso de energía solar, este tipo de fuente de energía limpia ha adquirido gran importancia en el entorno investigativo en los últimos años y han surgido grandes expectativas para el futuro.

Cada año el sol arroja 4000 veces más energía que la que se consume, lo que demuestra que esta fuente energética está aún infravalorada y desaprovechada en relación a sus potenciales energéticos; la implementación para esta fuente energética es posible si un territorio cuenta una buena posición geográfica, economía sostenible y fenómenos climáticos regulados, estas condiciones favorables para la energía a base de paneles solares o energía térmica se encuentran presentes en muchas regiones de América Central y América del Sur, como ventaja adicional la energía solar puede aprovecharse directamente o almacenarse para un consumo posterior y no tiene riesgo de agotarse.

A pesar de que el Colombia cuenta con las condiciones ideales para producir energía solar, aún existen regiones del territorio colombiano donde la sociedad aún está destinada a vivir sin energía eléctrica, un caso particular es el territorio nariñense donde varias poblaciones costeras cuentan con la radiación solar necesaria para abastecer el consumo requerido en esa región y adicionalmente vender la energía sobrante a poblaciones cercanas, y sin embargo no se ha realizado un estudio enfocado en el aprovechamiento de la energía a base de radiación solar presente en el departamento de Nariño; es necesario un estudio para el comportamiento de la radiación solar en lugares con potencial energético, un análisis del comportamiento de fenómenos climáticos en zonas del departamento de Nariño, esta y más información que es de gran importancia para la implementación de generadores a base de energía solar.

Planteamiento del problema

El clima colombiano posee grandes características para implementación de plantas de energía renovable debido a la ubicación geográfica en la zona ecuatorial, características geográficas que favorecen la presencia del clima tropical y una temperatura uniforme la mayor parte del año. La posición estratégica de Colombia en la zona tropical genera

muchos beneficios a la gran mayoría de departamentos, ocasionando que los territorios reciban una mayor proporción de energía que el sol le transfiere a la Tierra.

El departamento de Nariño hereda la diversidad de climas presentes en el territorio colombiano, en esta zona el clima es más estable y menos variable respecto a cambios en los fenómenos climáticos que obedecen a la posición de una zona o altitud donde se encuentra un lugar determinado, la información referente al comportamiento de estas características son de gran beneficio para aprovechar de forma óptima las diferentes fuentes de energía renovable. El proyecto “Análisis de oportunidades energéticas con fuentes alternativas en el departamento de Nariño” busca aprovechar los recursos naturales disponibles en este departamento con el objetivo inicial de estudiar y analizar las fuentes de energías biomasa, hidráulica, viento y solar que se encuentran presentes dentro del departamento; el presente proyecto de investigación busca realizar el estudio previo para identificar las zonas con posible potencial solar para generar energía, la radiación solar es una fuente renovable utilizada mucho tiempo atrás por países desarrollados, y actualmente el manejo de esta fuente energética es una tendencia de las grandes potencias mundiales para mejorar la economía y mitigar los efectos de agentes contaminantes.

La información que provee el IDEAM no permite tener gran certeza en los datos de radiación para una zona determinada, esto se debe a que cuenta con un número reducido de estaciones y es imposible tratar de generalizar datos aproximados para un territorio con grandes variaciones en su terreno, por este motivo se puede decir que en la actualidad el departamento de Nariño no cuenta con un estudio preciso en información de radiación solar, información que podría ser usada y aprovechada para plantas energéticas a base de las fuentes de energía renovable. Tener un registro histórico de la información climatológica es muy importante debido a que se puede hacer una predicción del comportamiento de fenómenos climáticos como nubes, viento, características de la vegetación, radiación solar, precipitaciones o datos que permitan la identificación de zonas con potencial energético en una zona específica del territorio nariñense. El actual proyecto de investigación se enfoca en el estudio de radiación solar, mediante la construcción de una serie de tiempo de 11 años de información referente a radiación solar, comportamiento de la radiación en zonas determinadas y el comportamiento de nubes sobre el departamento de Nariño; actualmente la información referente a la cantidad de energía radiante que llega a la superficie del departamento en una época o tiempo determinado es difícil establecer y muy complejo de predecir; por este motivo

el presente estudio resuelve esta carencia de información, la investigación radica en el procesamiento y análisis de imágenes satelitales de libre acceso proveídas por el sensor MODIS y Landsat, imágenes que a partir de las características y propiedades del sensor presenta determinado número de bandas con la capacidad de representar la medida de la gama de frecuencias capturada por el sensor, esta información representada en un formato determinado permite establecer características geográficas, hidrográficas, climáticas, vegetación, gases atmosféricos, radiación solar, precipitación y otro tipo información adicional del planeta que se encuentra presentes en los últimos sensores activos.

El estudio de energía solar mediante procesamiento y análisis de imágenes satelitales actualmente está apoyado por el uso de herramientas de libre acceso especializadas para realizar análisis de la información, aplicación de modelos de regresión, construcción de series de tiempo, visualización del potencial energético a través de mapas de radiación solar y detección de patrones secuenciales en una serie de tiempo.

Objetivo general

Detectar patrones secuenciales que permitan predecir el comportamiento de fenómenos climáticos y su impacto en áreas con potencial de radiación solar dentro del departamento de Nariño.

Objetivos Específicos

- Construir series de tiempo de radiación solar dentro del área de estudio a partir del procesamiento de imágenes satelitales de libre acceso.
- Detectar la presencia de nubosidad en las imágenes satelitales obtenidas utilizando los algoritmos más pertinentes.
- Seleccionar y aplicar herramientas para la detección de patrones secuenciales.
- Generar un conjunto de datos que relacione la presencia de nubosidad en regiones con potencial de radiación solar.
- Validar la relación de resultados obtenido mediante la comparación datos producto del estudio con datos reales capturados con algunos sensores activos.
- Generar mapas y reportes que permitan visualizar la radiación solar presente en el departamento de Nariño.

Organización del documento

Este documento está organizado en 5 secciones, la primera sección abarca los fundamentos teóricos empleados en la investigación, la sección 2 presenta la metodología utilizada para el desarrollo de la investigación, esta sección se desarrolla en las siguientes fases: a)la identificación y adquisición de imágenes satelitales, b)procesamiento y almacenamiento de la información, c)selección, aplicación del modelo de regresión y construcción de series de tiempo, d)mapas de radiación solar usando interpolación espacial, e)detección de patrones secuenciales, f)comportamiento de las nubes; la sección 3 contiene la información referente a la validación de la información, en la sección 4 se presentan las conclusiones y trabajos futuros y para finalizar se presenta referencias bibliográficas y los anexos correspondientes.

1 FUNDAMENTOS TEÓRICOS

1.1 SERIES DE TIEMPO

Son una secuencia de observaciones refiriéndose a datos que se recopilan y registran en intervalos de tiempo regulares, la información se encuentra ordenada cronológicamente por un único identificador de secuencia tanto cronológico como posicional, el principal objetivo de la series de tiempo, es permitir el análisis de la información procesada para hacer un pronóstico de los futuros eventos.

Cuando una serie de tiempo enfoca una unidad observable en diferentes momentos se denomina *Serie Temporal Univariante* y matemáticamente se representan de la forma: $y_1, y_2, \dots, y_N; (yt)^N t = 1; (yt : t = 1, \dots, N)$, Donde yt es la observación $n^0 t (1 \leq t \leq N)$ de la serie y N es el número de observaciones de que consta la serie completa. Las N observaciones y_1, y_2, \dots, y_N pueden recogerse en un vector columna $Y \equiv [y_1, y_2, \dots, y_N]$ de orden $Nx1$

Sí en una serie de tiempo enfoca varias unidades observables en diferentes momentos se denomina *Serie Temporal Multivariante* y frecuentemente se la puede encontrar representada de la forma:

$Y_1, Y_2, \dots, Y_N; (Yt)^N t = 1; (Yt : t = 1, \dots, N)$, donde $Y \equiv [Y_1, Y_2, \dots, Y_N] (M \geq 2)$ es la observación $n^0 t (1 \leq t \leq N)$ de la serie y N es el número de observaciones de que consta la serie completa, las N observaciones Y_1, Y_2, \dots, Y_N pueden recogerse en una matriz Y de orden MxN

$$Y \equiv \begin{bmatrix} Y'_1 \\ Y'_2 \\ \vdots \\ Y'_N \end{bmatrix} \equiv \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1M} \\ y_{21} & y_{22} & \dots & y_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N1} & y_{N2} & \dots & y_{NM} \end{bmatrix}$$

Donde Y_{tj} es la observación n^0 tal que $t (1 \leq t \leq N)$ sobre la característica o variable N^0 tal que $j (1 \leq t \leq N)$, que es la misma en todo momento t .

1.1.1 Componentes los componentes más comunes en las series de tiempo están directamente relacionados con la variación de las observaciones, observaciones que al sobreponerse o actuar en concierto, contribuyen a los cambios observados en un período de tiempo y dan a la serie su tendencia, aspecto errático, entre los componentes mas comunes encontramos:

- **Tendencia secular:** también denominada tendencia a largo plazo de una serie, es por lo común el resultado de factores a largo plazo. En términos intuitivos, la tendencia de una serie de tiempo caracteriza el patrón gradual y consistente de las variaciones de la propia serie, que se consideran consecuencias de fuerzas persistentes que afectan el crecimiento o la reducción de la misma. Las tendencias a largo plazo se ajustan a diversos esquemas. Algunas se mueven continuamente hacia arriba, otras declinan, y otras más permanecen igual en un cierto período o intervalo de tiempo.

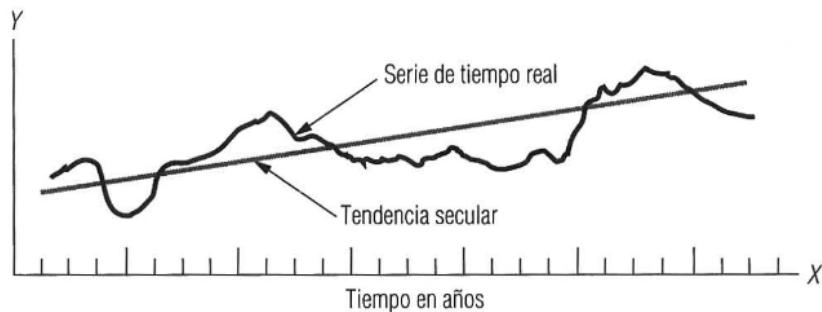


Figura 1.1: Representación de Tendencia secular en una serie de tiempo

- **Variación estacional:** comprende la representación de la variabilidad en los datos debida a influencias de las estaciones. Esta variación corresponde a los movimientos de la serie que recurren año tras año en los mismos meses del año poco más o menos con la misma intensidad. Este tipo de variación implica patrones de cambio en el lapso de un año que tienden a repetirse anualmente. Por ejemplo, un médico puede esperar un aumento sustancial en el número de casos de gripe cada invierno y de afectados de tifoidea cada verano. Como se trata de patrones regulares son útiles al pronosticar el futuro. La Figura 1.2 muestra la variación estacional, se puede notar como la serie de tiempo alcanza un pico cada cuarto trimestre del año.



Figura 1.2: Representación de Variación estacional en una serie de tiempo

- **Variación cíclica:** con frecuencia las series de tiempo presentan secuencias alternas de puntos abajo y arriba de la línea de tendencia que duran más de un año, esta variación se mantiene después de que se han eliminado las variaciones o tendencias estacional e irregular. Un ejemplo de este tipo de variación son los ciclos comerciales cuyos períodos recurrentes dependen de la prosperidad, recepción, depresión y recuperación, las cuales no dependen de factores como el clima o las costumbres sociales. La Figura 1.3 ilustra un patrón típico de fluctuación cíclica arriba y abajo de la linea de tendencia secular; se puede observar que los movimientos cíclicos no siguen ningún patrón regular sino que se mueve de forma un tanto impredecible.

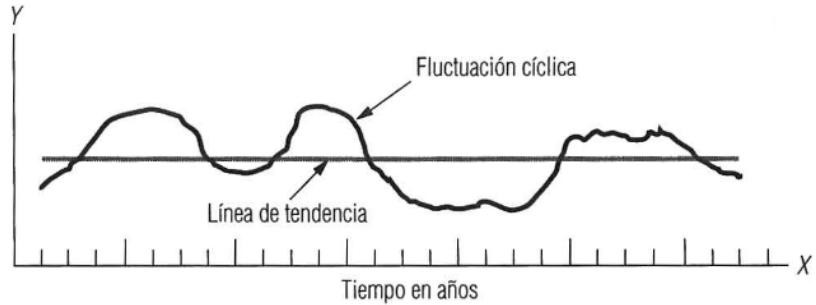


Figura 1.3: Representación de Variación cíclica en una serie de tiempo

- **Variación Irregular:** se debe a factores a corto plazo, imprevisibles y no recurrentes que afectan a la serie de tiempo. Como este componente explica la variabilidad aleatoria de la serie, es impredecible; no se puede esperar predecir su impacto sobre la serie de tiempo. Existen dos tipos de variación irregular: a) Las variaciones que son provocadas por acontecimientos especiales, fácilmente identificables, como las elecciones, inundaciones, huelgas, terremotos. b) Variaciones

aleatorias o por casualidad, cuyas causas no se pueden señalar en forma exacta, pero que tienden a equilibrarse a la larga.



Figura 1.4: Representación de Variación Irregular en una serie de tiempo

1.1.2 Análisis Grafico de una Serie de Tiempo Por muy simple que parezca, el paso más importante en el análisis de series de tiempo consiste en graficar la serie, esto debe hacerse siempre, independiente de cuán simples o complejos sean los procedimientos que se emplean posteriormente. El análisis del gráfico de la serie permitirá detectar los siguientes elementos:

- **Outliers:** representa cambios abruptos en los puntos de la serie que se escapan de lo normal; si una serie temporal tiene observaciones anormalmente elevadas (o reducidas), y no se tratan de modo especial, podrían dominar los resultados obtenidos de la estimación. A modo de intuición recuérdese que MCO minimiza la suma de cuadrados de los residuos, y el cuadrado del residuo de estos valores será anormalmente alto. Por lo tanto, el estimador tenderá a sobreponer estos valores anómalos, afectando a los parámetros estimados y reduciendo el ajuste total del modelo. La aparición de un outlier puede tener un motivo identificable, en algunas ocasiones no es sencillo encontrar la causa que provoca el outlier. Se trata de casos en los que es probable que el outlier se deba a algún error en el proceso de tratamiento de la información.

Un adecuado tratamiento para contrarestar los efectos del outlier es sustituirlo por una media móvil ponderada ad-hoc de los valores cercanos, aunque esta solución no es eficiente debido a que puede descartar información relevante. Otra medida muy utilizada es la de incorporar el outlier en el modelo por medio de variables ficticias, para controlar su efecto.

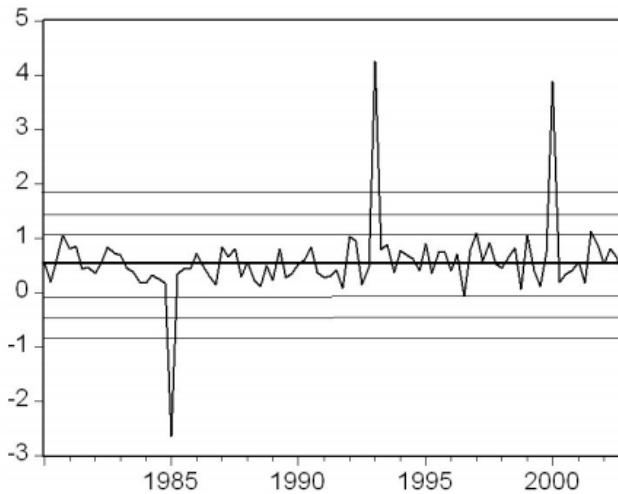


Figura 1.5: Representación de Outliers en una serie de tiempo

- **Tendencias:** La tendencia representa el comportamiento predominante de la serie, se puede definir como un cambio a largo plazo que se produce en relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo. Una forma de visualizar la tendencia es mediante suavizamiento de la serie, la idea central es definir a partir de la serie observada una nueva serie que filtra o suaviza los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios), de manera que podamos visualizar la tendencia.

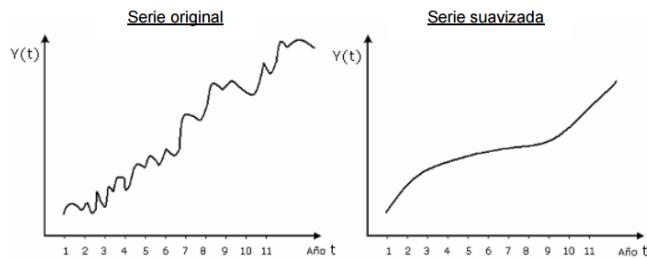


Figura 1.6: Tendencia luego de suavizamiento de una serie de timepo

El gráfico de una serie de tiempo permite visualizar de mejor forma los componentes de una serie de tiempo, facilitando un análisis mediante la percepción visual y resumiendo el comportamiento de un evento a lo largo del registro de sus observaciones.

1.1.3 Suavizar la Serie de Tiempo Los métodos de suavizamiento contrarrestan a las fluctuaciones aleatorias causadas por el componente irregular de la serie. Estos métodos resultan apropiados para series estables, es decir, aquellas que no exhiban ningún comportamiento de tendencia, ni variaciones cíclicas ni estacionales, además es conveniente suavizar cuando existen cambios bruscos o movimientos irregulares en la serie. A continuación se detalla los métodos más comunes utilizados para suavizar una serie de tiempo.

- **Método de Medias Móviles:** El objetivo es eliminar de la serie los componentes estacionales y accidentales; utiliza como pronóstico para el siguiente período, el promedio de los n valores de los datos más recientes de la serie de tiempo, matemáticamente puede expresarse como:

$$PromedioMóvil = \frac{\sum(n \text{ valores de datos más recientes})}{n} \quad (1.1)$$

El término móvil indica que conforme se tenga disponible una nueva observación de la serie de tiempo, se reemplaza la observación más antigua en la ecuación y se calcula un nuevo pronóstico. Como resultado el promedio se modificará, a medida que se agreguen nuevas observaciones.

La variable n es una indicación de cuantos períodos habrán que tomarse para calcular el promedio, generalmente suele variar entre 3 a 5, dependiendo de cuantos elementos tiene la serie de datos. Para eliminar el componente irregular o el componente estacional en la construcción de una serie de variables a partir de los datos originales se aplica las variaciones del método listadas a continuación.

- Media móvil anual o de orden 12: Consiste en sustituir el dato de cada mes por el promedio de los datos de los últimos 12 meses, es decir:

$$M_t^{12} = \frac{\sum_{t=1}^{12} y_{t-1}}{12} \quad (1.2)$$

- Para una Serie Mensual con Estacionalidad Anual ($s = 12$)

$$y(k) = \frac{\frac{1}{2}y(k-6) + y(k-5) + \dots + y(k+5) + \frac{1}{2}y(k+6)}{12}, \text{ tal que } 7 \leq k \leq n-6 \quad (1.3)$$

- Para una serie trimestral, con estacionalidad anual($s = 4$)

$$y(k) = \frac{\frac{1}{2}y(k-2) + y(k-1) + y(k) + y(k+1) + \frac{1}{2}y(k+2)}{4}, \text{ tal que } 3 \leq k \leq n-2 \quad (1.4)$$

A este procedimiento se le llama: filtro simétrico finito.

- Media Móvil de Orden s : Podemos construir medias móviles de orden inferior a 12, es decir, que no sean anuales. Así, la Media Móvil de Orden 3 se construye promediando los tres últimos datos. De manera general, la Media Móvil de Orden s se construye de la siguiente forma:

$$M_t^s = \frac{y_1 + y_{t-1} + y_{t-2+\dots+y_{t-s+1}}}{s} \quad (1.5)$$

La principal ventaja que presenta esta media móvil será que no retrasará tanto la evolución de la variable como lo hacía la media móvil anual, ya que al no utilizar tantas observaciones, su comportamiento se asemejará más al de la serie original. Sin embargo, el principal inconveniente de esta media móvil viene precisamente de esa menor utilización de observaciones. Al no construirse con 12 observaciones, no se elimina completamente el componente estacional, llegando en ciertos casos a multiplicarse este efecto.

- Media Móvil Centrada: Esta media móvil no se construye promediando los datos anteriores al dato original, sino que se utilizan simétricamente los datos adyacentes. En general, la Media Móvil Centrada de orden s va a estar definida por la siguiente expresión:

$$MC_1^s = \frac{Y_{t+\frac{s-1}{2}} + \dots + y_{t+1} + y_1 + y_{t-1} + \dots + y_{t-\frac{s-1}{2}}}{s} \quad (1.6)$$

La principal ventaja de esta media móvil es que al utilizar los datos adyacentes, tanto anteriores como posteriores, no vamos a tener el problema de retrasar la evolución de la variable, es decir, tanto la serie original como la media móvil centrada van a tener en las mismas fechas los valores máximos y mínimos. El principal inconveniente será que al utilizar las observaciones posteriores a cada dato original, para su construcción, no tendremos observaciones de la MC al final de la muestra disponible, siendo ésta precisamente la parte de la muestra más interesante para el análisis. Esta dificultad se resuelve si disponemos de predicciones fiables para los próximos meses.

- **Método de Suavización Exponencial Simple:** Esta técnica se basa en la atenuación de los valores de la serie de tiempo, obteniendo el promedio de estos de manera exponencial; es decir, los datos se ponderan dando un mayor peso a las observaciones más recientes y uno menor a las más antiguas. Al peso para ponderar la observación más reciente se le da el valor α , la observación inmediata anterior se pondera con un peso de $1 - \alpha$, a la siguiente observación inmediata

anterior se le da un peso de ponderación de $\alpha(1 - \alpha)^2$ y así sucesivamente hasta completar el número de valores observados en la serie de tiempo a tomar en cuenta para realizar la atenuación, es decir, para calcular el promedio ponderado. La estimación o pronóstico será el valor obtenido del cálculo del promedio. Por lo que la expresión para realizar el cálculo de la suavización exponencial simple es:

$$P_{t+1} = \alpha Y_t + \alpha(\alpha - 1)Y_{t-1} + \alpha(\alpha - 1)^2Y_{t-2} + \dots + \alpha(\alpha - 1)^{n-1}Y_{t-(n-1)} \quad (1.7)$$

Donde

Y_t : Valor de la serie en el período t .

P_{t+1} : Pronóstico o predicción para el período $t + 1$

α : Factor de suavización, $(0 \leq \alpha \leq 1)$

Es decir el valor de la serie suavizada en el período $t + 1$ es igual a α veces el valor de la serie en el período t , más $1 - \alpha$ veces el valor predicho en el período t . Es así que para determinar los valores de la serie suavizada se necesita un valor inicial P_0 , el cual puede ser un promedio de los datos anteriores o simplemente el primer valor de la serie.

1.2 Bases de Datos Espacio-Temporales

Dos campos que han surgido a mediados y finales de la década de 1980 son las bases de datos espaciales y bases de datos temporales las cuales se han convertido, en cierto modo como las precursoras de las bases de datos espacio-temporales.

Base de datos espaciales

Las bases de datos espaciales tratan del eficiente almacenamiento y la recuperación de objetos en el espacio que tienen identidad y extensión bien definidas, ubicaciones, así como ciertas relaciones geométricas y / o topológica entre ellas, debido al desarrollo en campos de aplicación (GIS, diseño VLSI, CAD) que necesita tratar con grandes cantidades de entidades geométricas, geográficas o espaciales. Además de algunos estables y maduros prototipos basado en fundamentos algebraicos de tipo sólido, proveedores de sistemas gestores de bases de datos comercial (DBMS) han ofrecido extensiones de sus productos, soportando tipos espaciales y operaciones (Oracle Spatial, DB2 Spatial Extender, PostgresGIS , Microsoft SQL Server, MySQL). Sin lugar a dudas, los resultados

en bases de datos espaciales han impulsado varias líneas de investigación importantes en bases de datos de objetos en movimiento (MOD), como por ejemplo: consultas populares de MOD (rango, vecino mas cercano), propiedades topológicas y relaciones entre tipos espaciales. [1]

Base de datos temporales

Existen muchas aplicaciones de bases de datos, como por ejemplo, contabilidad, gestión de cartera, registros médicos y gestión de inventario; información que varía en el tiempo. El centro de las bases de datos temporales se diferencia por:

Tiempo válido de un hecho, es el tiempo en el cual un dato particular es recogido y se convierte en verdad hasta donde el mundo representado por la base de datos es referida, posiblemente abarque el pasado, el presente y el futuro. Sin embargo, el tiempo válido no puede ser conocido, o los registros no pueden ser relevantes para las aplicaciones soportadas por la base de datos o en el caso que los modelos de base de datos del mundo real puedan variar entre los diferentes mundos reales.

Tiempo de transacción de un hecho, es el tiempo que un hecho dado es el actual en la base de datos. El tiempo de transacción puede asociarse con diferentes entidades de bases de datos, como, por ejemplo, objetos y valores que no son hechos, porque no pueden ser verdaderos o falsos en forma aislada. Por lo tanto, todas las entidades de la base de datos tiene un aspecto de tiempo de transacción, que tiene una duración: desde inserción (lógico) a la eliminación de una determinada entidad.

Capturar el tiempo variable utilizando los modelos de datos tradicionales y lenguajes de consulta puede ser una actividad engorrosa y, como consecuencia, se necesitan construcciones que permitirán capturar los tiempos válidos y operación de los hechos, lo que conduce a relaciones temporales. Además, en los lenguajes de consulta se necesitan extensiones sintácticas que permiten las operaciones de base de datos en modelos temporales. [1]

1.3 Técnicas de Regresión

Permiten investigar y modelar la relación entre variables, la principal ventaja de la regresión es la descripción de los datos con el objetivo de resumir y ajustar un conjunto de datos. Adicionalmente es posible realizar una estimación de parámetros. Una vez seleccionado los modelos de regresión es necesario realizar un análisis y comparación de los diferentes modelos de regresión y chequear que tan bueno es un modelo ajustado respecto a otro. El resultado de este chequeo puede indicar si el modelo es razonable o si el ajuste original debe ser modificado.

Selección de las variables del modelo

Existen varios métodos para construir el modelo de regresión, es decir, para seleccionar de entre todas las variables que introducimos en el modelo, cuáles son las que necesitamos para explicarlo. El modelo de regresión se puede construir utilizando las siguientes técnicas [2]:

- **Técnica de pasos hacia adelante (Forward)** Consiste en ir introduciendo las variables en el modelo únicamente si cumplen una serie de condiciones hasta que no se pueda introducir ninguna más, hasta que ninguna cumpla la condición impuesta
- **Técnica de pasos hacia atrás (Backward)** Se introducen en el modelo todas las variables y se van suprimiendo si cumplen una serie de condiciones definidas a priori hasta que no se pueden eliminar más, es decir ninguna variable cumpla la condición impuesta.
- **Técnica por pasos (Stepwise)** Combina los dos métodos anteriores, adelante y atrás introduciendo o eliminando variables del modelo si cumplen una serie de condiciones definidas a priori hasta que ninguna variable satisfaga ninguna de las condiciones expuestas de entrada o salida del modelo.
- **Técnica de introducir todas las variables obligatoriamente (Enter)** Esta última técnica de selección de variables para construir el modelo de regresión, produce que el proceso de selección de las variables sea manual, partiendo de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se va evaluando qué variable es la que menos participa en él y se elimina, volviendo

a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que los componen.

1.4 Redes Neuronales

Se define redes neuronales a una nueva forma de computación inspirada en modelos biológicos, un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles, también se define redes neuronales a un sistema de computación compuesto por un gran número de elementos simples, elementos de procesos muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas o se puede definir las redes neuronales como redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico 1.7.

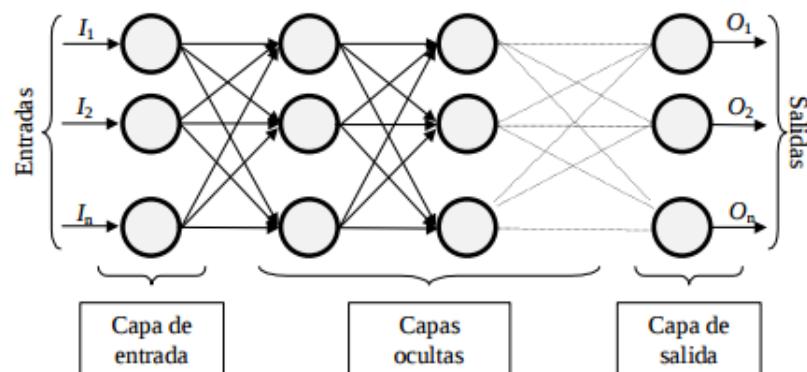


Figura 1.7: un esquema de una red neuronal.

Las redes neuronales presentan características como:

- **Aprendizaje adaptativo**

Las redes neuronales pueden aprender a diferenciar patrones mediante ejemplos y entrenamientos, no es necesario elaborar modelos a priori ni necesidad de especificar funciones de distribución de probabilidad; las RN son sistemas dinámicos auto-adaptativos con la capacidad de auto-ajuste de los elementos procesales (neuronas) que componen el sistema, son capaces de estar constantemente cambiando para adaptarse a las nuevas condiciones; los enlaces ponderados de las neuronas se ajustan de manera que se obtengan ciertos resultados específicos. Una red neuronal no necesita un algoritmo para resolver un problema, ya que ella puede generar su propia distribución de pesos en los enlaces mediante el aprendizaje. También existen redes que continúan aprendiendo a lo largo de su vida, después de completado su período de entrenamiento.

- **Auto-organización**

Las redes neuronales emplean su capacidad de aprendizaje adaptativo para auto-organizar la información que reciben durante el aprendizaje y/o la operación. Mientras que el aprendizaje es la modificación de cada elemento procesal, la auto-organización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico. Cuando las redes neuronales se usan para reconocer ciertas clases de patrones, ellas auto-organizan la información usada. Por ejemplo, la red llamada backpropagation, creará su propia representación característica, mediante la cual puede reconocer ciertos patrones. Esta auto-organización provoca la generalización: facultad de las redes neuronales de responder apropiadamente cuando se les presentan datos o situaciones a las que no había sido expuesta anteriormente. El sistema puede generalizar la entrada para obtener una respuesta.

Esta característica es muy importante cuando se tiene que solucionar problemas en los cuales la información de entrada no es muy clara; además permite que el sistema dé una solución, incluso cuando la información de entrada está especificada de forma incompleta.

- **Tolerancia a fallos**

Las redes neuronales fueron los primeros métodos computacionales con la capacidad inherente de tolerancia a fallos. Comparados con los sistemas computacionales

tradicionales, los cuales pierden su funcionalidad cuando sufren un pequeño error de memoria, en las redes neuronales, si se produce un fallo en un número no muy grande de neuronas y aunque el comportamiento del sistema se ve influenciado, no sufre una caída repentina. Hay dos aspectos distintos respecto a la tolerancia a fallos:

- Las redes pueden aprender a reconocer patrones con ruido, distorsionados o incompletos. Esta es una tolerancia a fallos respecto a los datos.
- Las redes pueden seguir realizando su función (con cierta degradación) aunque se destruya parte de la red.

La razón por la que las redes neuronales son tolerantes a los fallos es que tienen su información distribuida en las conexiones entre neuronas, existiendo cierto grado de redundancia en este tipo de almacenamiento. La mayoría de los ordenadores algorítmicos y sistemas de recuperación de datos almacenan cada pieza de información en un espacio único, localizado y direccionable. En cambio, las redes neuronales almacenan información no localizada. Por lo tanto, la mayoría de las interconexiones entre los nodos de la red tendrán sus valores en función de los estímulos recibidos, y se generará un patrón de salida que represente la información almacenada.

- **Operación en tiempo real**

Una de las mayores prioridades, casi en la totalidad de las áreas de aplicación, es la necesidad de realizar procesos con datos de forma muy rápida. Las redes neuronales se adaptan bien a esto debido a su implementación paralela. Para que la mayoría de las redes puedan operar en un entorno de tiempo real, la necesidad de cambio en los pesos de las conexiones o entrenamiento es mínimo.

- **Fácil inserción dentro de la tecnología existente.**

Una red individual puede ser entrenada para desarrollar una única y bien definida tarea (tareas complejas, que hagan múltiples selecciones de patrones, requerirán sistemas de redes interconectadas). Con las herramientas computacionales existentes (no del tipo PC), una red puede ser rápidamente entrenada, comprobada, verificada y trasladada a una implementación hardware de bajo coste. Por lo tanto, no se presentan dificultades para la inserción de redes neuronales en aplicaciones específicas, por ejemplo de control, dentro de los sistemas existentes. De

esta manera, las redes neuronales se pueden utilizar para mejorar sistemas en forma incremental y cada paso puede ser evaluado antes de acometer un desarrollo más amplio [3].

Entre las propiedades de las RN que han llamado la atención de los estadísticos destacan las relativas a su buen rendimiento ante problemas no lineales o datos con mucho «ruido», y el poderse utilizar independientemente del cumplimiento de los supuestos teóricos relativos a las técnicas estadísticas (y de ahí que se haya hablado de ellas como de «técnicas de distribución libre o no paramétricas»). Por ello las RN han sido aplicadas a problemas de tradición estadística como predicción y clasificación (a través de las llamadas redes hetero-asociativas: perceptrón multicapa y redes de función base radial), reducción de la dimensionalidad (a través de las llamadas redes auto-asociativas: Hopfield, Kohonen), series temporales, etc.[4][5][6].

1.5 Interpolación Espacial

La interpolación espacial es el proceso de utilizar puntos con valores conocidos para estimar valores desconocidos en otros puntos. Por ejemplo, para realizar un mapa de precipitación (lluvia) para el país no se encontrarán suficientes estaciones meteorológicas distribuidas uniformemente para cubrir toda la región. La interpolación espacial puede estimar las temperaturas en lugares que no tienen ese dato utilizando lecturas de temperatura conocida en estaciones meteorológicas cercanas 1.8. A este tipo de superficie interpolada con frecuencia se le llama una superficie estadística. Datos de elevación, precipitación, acumulación de nieve, tabla de agua y densidad de población son otros tipos de datos que pueden ser calculados utilizando la interpolación [7].

Debido al alto costo y a los recursos limitados la recolección de los datos usualmente es llevada a cabo sólo en un número limitado de ubicaciones de puntos seleccionados. En un SIG, la interpolación de esos puntos puede ser aplicada para crear una superficie raster con estimaciones realizadas para todas las celdas del raster. Con el fin de generar un mapa continuo, por ejemplo, un mapa de elevaciones digitales de los puntos de elevación medidos con un dispositivo GPS, se debe utilizar un método de interpolación adecuado para estimar de manera óptima los valores en aquellas ubicaciones en donde no fueron tomadas muestras o mediciones. Los resultados del análisis de interpolación pueden entonces ser utilizados para análisis que cubran el área completa y para el modelado.

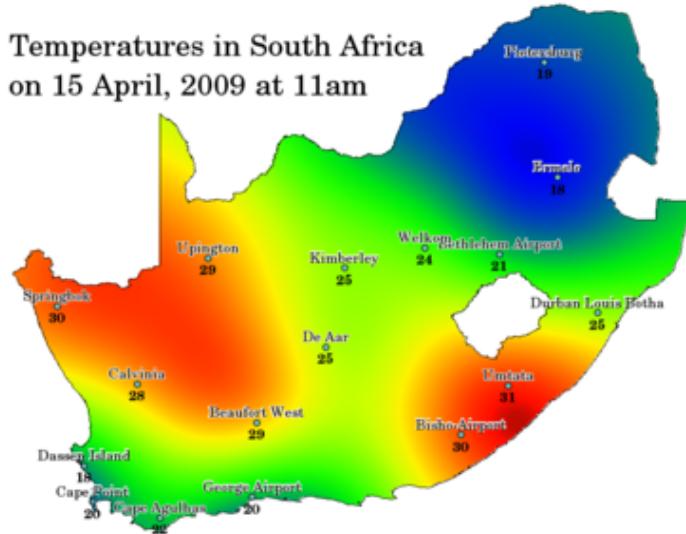


Figura 1.8: Mapa de temperaturas interpolado de estaciones.

Metodos de Interpolación

Método Trend Surface: es un método analítico, global e inexacto a partir de puntos. Este método se utiliza para separar y describir determinados componentes de variación presentes en los datos, facilitando su interpretación, puesto que cada una de las observaciones puede ser considerada como resultado de la adición de un componente regional o de tendencia y un componente local. Asimismo, considera la autocorelación de la variable. Ajustándose la variable Z a una ecuación de regresión cuyas variables explicativas son X e Y de los puntos muestrales. Proporciona una descripción sintética de la superficie ondulada que se está tratando y de la variación espacial de la variable temática, facilitando un método para estimar el valor de Z en un punto no muestral cuyas coordenadas X e Y sean conocidas Figura 1.9.

Método Moving average: es un método directo, local, a partir de puntos, puede ser exacto o no según el factor de ponderación. Se aplica para un gran conjunto de datos. Extrae tendencias intermedias de un número mínimo de puntos definidos dentro de una Search Elipse, asociada a cada uno de los puntos del grid. El valor final de cada uno de los puntos del grid es igual a la media aritmética de todos los puntos vecinos identificados. Si dentro de la Search ellipse no hay un número mínimo de puntos definidos para el cálculo, el área estará en blanco Figura 1.10.

Método Kriging: es un método analítico, donde la función de interpolación depende

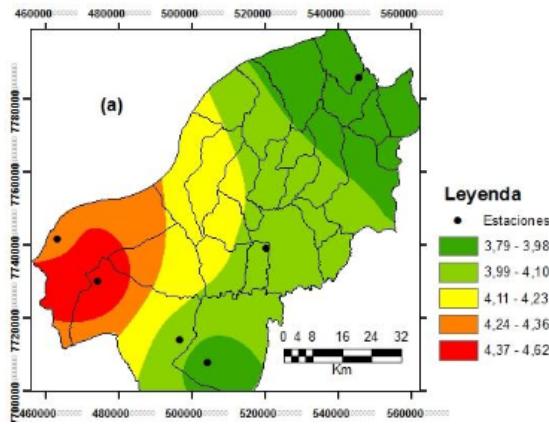


Figura 1.9: Interpolación con el método Trend Surface[8]

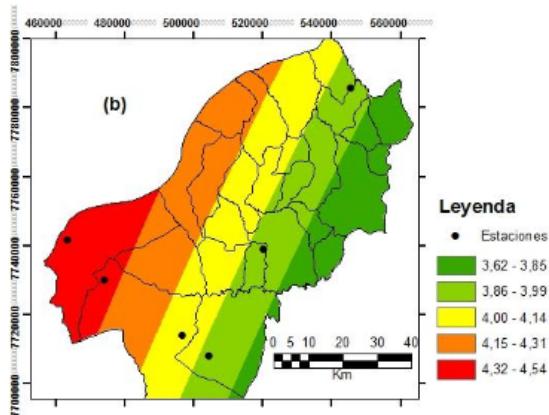


Figura 1.10: Interpolación con el método Moving Average[8]

de la autocorelación espacial de la variable, que se representa en variogramas. Utiliza datos tabulares y su posición geográfica para el cálculo de las interpolaciones. Utilizando el principio de la primera ley geográfica de Tobler, que dice que las unidades de análisis más próximas entre si son mas similares que las unidades más lejanas, el kriging utiliza funciones matemáticas para añadir más peso en las posiciones más cercanas a los puntos de muestreo y menores pesos en posiciones más distantes, y así crear nuevos puntos interpolados basados en estas combinaciones lineares de datos. Además se está basado en optimizar funciones usando autocorelación espacial Figura 1.11.

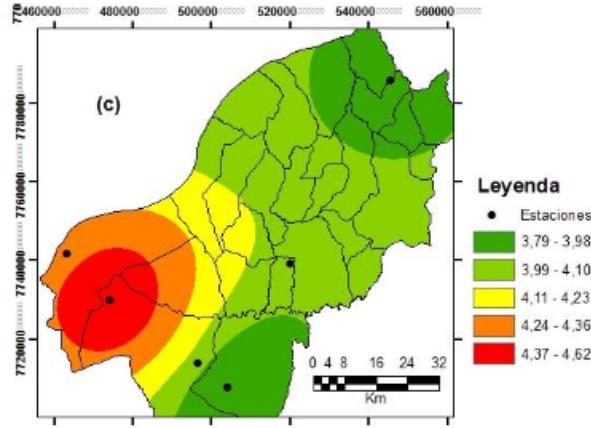


Figura 1.11: Interpolación con el método Kriging [8]

1.6 Minería de Datos

El descubrimiento de conocimiento en bases de datos (KDD) se define como el proceso de identificar patrones significativos en los datos que sean válidos, novedosos, potencialmente útiles y comprensibles para un usuario. El proceso global consiste en transformar información de bajo nivel en conocimiento de alto nivel [9]. El proceso KDD es interactivo e iterativo conteniendo los siguientes pasos:

1. **Comprender el dominio de aplicación:** este paso incluye el conocimiento relevante previo y las metas de la aplicación.
2. **Extraer la base de datos objetivo:** recogida de los datos, evaluar la calidad de los datos y utilizar análisis exploratorio de los datos para familiarizarse con ellos.
3. **Preparar los datos:** incluye limpieza, transformación, integración y reducción de datos. Se intenta mejorar la calidad de los datos a la vez que disminuir el tiempo requerido por el algoritmo de aprendizaje aplicado posteriormente.
4. **Minería de datos:** como se ha señalado anteriormente, este es la fase fundamental del proceso. Está constituido por una o más de las siguientes funciones, clasificación, regresión, clustering, resumen, recuperación de imágenes, extracción de reglas, etc.

5. **Interpretación:** explicar los patrones descubiertos, así como la posibilidad de visualizarlos.

A continuación se comentan brevemente las tareas más comunes en cualquier proceso de la minería de datos.

- Clasificación: clasifica un dato dentro de una de las clases categóricas predefinidas. Responde a preguntas tales como, ¿Cuál es el riesgo de conceder un crédito a este cliente? ¿Dado este nuevo paciente qué estado de la enfermedad indican sus análisis?
- Regresión: el propósito de este modelo es hacer corresponder un dato con un valor real de una variable. Responde a cuestiones como ¿Cuál es la previsión de ventas para el mes que viene? ¿De qué depende?
- Clustering: se refiere a la agrupación de registros, observaciones, o casos en clases de objetos similares. Un cluster es una colección de registros que son similares entre sí, y distintos a los registros de otro cluster. ¿Cuántos tipos de clientes vienen a mi negocio? ¿Qué perfiles de necesidades se dan en un cierto grupo de pacientes?
- Generación de reglas: aquí se extraen o generan reglas de los datos. Estas reglas hacen referencia al descubrimiento de relaciones de asociación y dependencias funcionales entre los diferentes atributos.
- Resumen o summarización: estos modelos proporcionan una descripción compacta de un subconjunto de datos. ¿Cuáles son las principales características de los datos?
- Análisis de secuencias: se modelan patrones secuenciales, como análisis de series temporales, secuencias de genes, etc. El objetivo es modelar los estados del proceso, o extraer e informar de la desviación y tendencias en el tiempo. ¿El consumo de energía eléctrica de este mes es similar al del año pasado? Dados los niveles de contaminación atmosférica de la última semana cuál es la previsión para las próximas 24 horas.

Patrones Frecuentes en Bases de Datos Tradicionales

Patrones frecuentes son conjuntos de elementos, subsecuencias, o subestructuras que aparecen en un conjunto de datos con una frecuencia no inferior a un umbral especificado por el usuario [10]. La cuestión de las modalidades interesantes desvelando en las bases de datos en diferentes contextos ha sido un tema de investigación recurrente durante los últimos 15 años. Minería de datos general ha sido ampliamente reconocido como un crítico campo por empresas de todo tipo. Como parte de los métodos de minería de datos, la tarea de aprendizaje de reglas de asociación han estudiado diferentes algoritmos de minería de patrones frecuente de identificar tendencias relevantes en los conjuntos de datos en diferentes disciplinas [11].

Una de las áreas en las que las técnicas de aprendizaje de reglas de asociación y el patrón frecuente algoritmos de minería se han aplicado con más frecuencia en el análisis de datos y tendencias del mercado en transacciones de clientes de grandes supermercados y tiendas [12]. Por lo general, esta técnica tiene dado el nombre del problema de la cesta de compras a pesar de que los métodos derivados de resolverlo puede ser aplicado en diferentes contextos [13].

Minería de Patrones Secuenciales

Sea DB un conjunto de registros (objetos), donde cada registro R consiste de tres elementos de información:

- un identificador de registro object-id
- un registro de tiempo timestamp
- un conjunto A de atributos binarios

Un atributo a puede estar presente '1', o ausente '0'. Si bien los algoritmos intentan encontrar secuencias de atributos presentes, esta claro que su ausencia también brinda significativa información. Un atributo a presente es llamado ítem i, y dentro del contexto de números, un ítem es una tupla (atributo, valor). Un conjunto de ítems i₁, i₂, . . . , i_k es denotado por I, donde I es un subconjunto de A. Esto es una representación no ordenada. Una secuencia s es una lista ordenada no vacía de los conjuntos de ítems s, denotado por s₁, s₂, . . . , s_i. Una secuencia es una secuencia de n ítems o de tamaño n. Decimos entonces que R es un conjunto de registros que contienen los atributo presente (o ítems) i. Gráficamente R(I) podría verse así como en la figura 1.12.

Object-id	time-stamp	a_1 ausente	a_2 ítem-i 2 presente	...	a_2 ítem-i n presente
-----------	------------	------------------	----------------------------	-----	----------------------------

Figura 1.12: Algunos Atributos Presentes o items.

Secuencia de datos Una secuencia de datos hs_1, s_2, \dots, s_i es una agrupación ordenada de registros con atributos presentes iguales, similares o pertenecientes. La agrupación se ordena según los datos de timestamp.

Frecuencia de secuencias Los grados de pertenencia se obtienen al evaluar la frecuencia de ocurrencia de las secuencias $\text{freq}(s)$ con distintos niveles. Algunos autores establecen un valor de referencia o parámetro, por ejemplo minFreq , con el fin de decidir si una frecuencia es más o menos frecuente.

LCM (Linear Time Closed Itemset Miner)

El problema de LCM propuesto por [14] se define de la siguiente manera. Sea I un conjunto de elementos. Sea D una base de datos transaccional de tal manera que cada registro (llamada transacción) es un conjunto de elementos. La frecuencia de un conjunto de elementos es el número de transacciones, incluyendo el conjunto de elementos. Para un número dado t (llamado soporte), un conjunto de elementos se dice que es frecuente si su frecuencia es no menos de t . Un conjunto de elementos frecuente se llama máxima si está incluido en ningún otro conjunto de elementos frecuentes, y se llama cerrada si está incluido en ningún otro conjunto de elementos de la misma frecuencia. La tarea de LCM, es enumerar (sacar, o contar) todos los conjuntos de elementos frecuentes, todos los conjuntos de elementos frecuentes máximos, o todos los conjuntos de elementos frecuentes cerrados en una base de datos transaccional para un soporte dado.

2 TRABAJOS RELACIONADOS

Dada la creciente demanda por la generación y apropiamiento de las energías limpias, el estudio de alternativas energéticas basadas en paneles solares también ha venido en aumento. Sin embargo, un componente clave para la implementación de una solución de este tipo es analizar de antemano las posibles ubicaciones con mayor potencial de generación eléctrica a base de radiación solar. Diversos estudios han consolidado el uso de imágenes satelitales de libre acceso como herramientas fundamentales para este propósito. [15] realizó la construcción de mapas solares de alta resolución comparando metodologías basadas en imágenes satelitales y modelamiento climático. Para estos modelos de radiación se contemplan condiciones propias del terreno como la latitud , el terreno, estación, hora del día y condición atmosférica(nubes, polvo, polución, vapor de agua y efectos de la montaña)

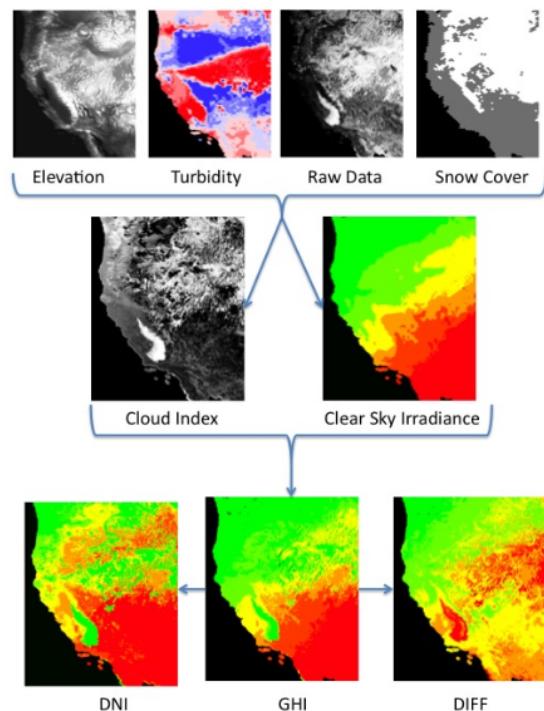


Figura 2.1: Diagrama de flujo que representa el cálculo de los valores de irradiancia (Global Horizontal, Directa normal , y Difusa horizontal) de un área para un área, la fecha dada y ángulo cenital solar.

[16] desarrolló el método HELIOSAT con el objetivo de estimar los niveles de radiación solar a partir de imágenes satelitales geoestacionarias, el método se encuentra implementado en un algoritmo que permite separar la irradiación de los componentes atmosférico de la irradiación en nubes para finalmente obtener la irradiación superficial.

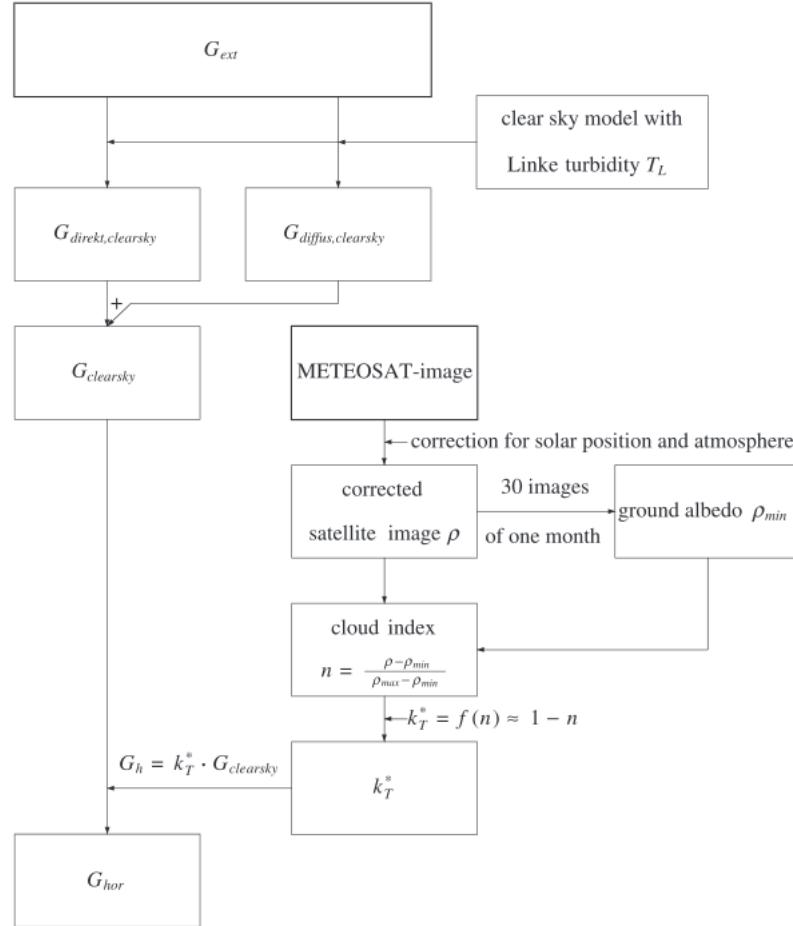


Figura 2.2: Descripción general del método HELIOSAT.

[17] realiza una recopilación de métodos para la predicción de radiación solar también basados en imágenes por satélite o modelos climáticos, en este estudio se contempla la variación de la resolución espacial de la imagen satelital, tipos de sensores utilizados y la escala temporal.

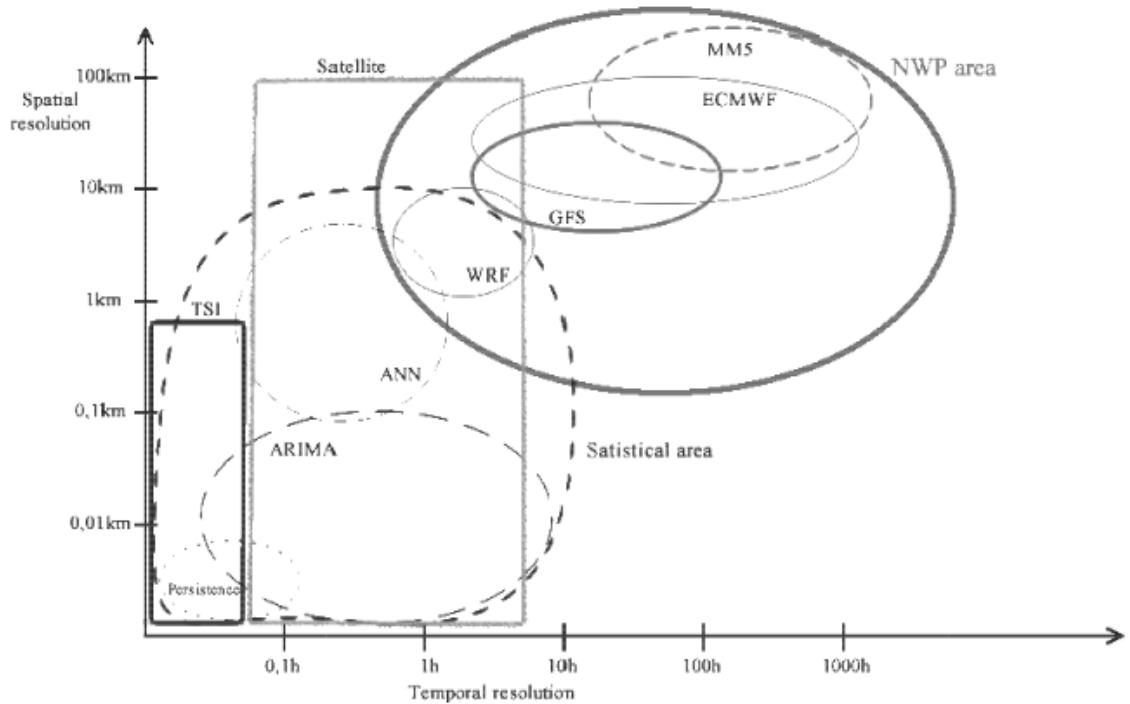


Figura 2.3: Clasificación de modelos de predicción basados en resolución espacial de datos y resolución temporal.

[18] obtuvo gran éxito al proponer un nuevo modelo basado en redes neuronales apropiado para pronosticar el potencial solar a corto plazo bajo condiciones meteorológicas en constante cambio.

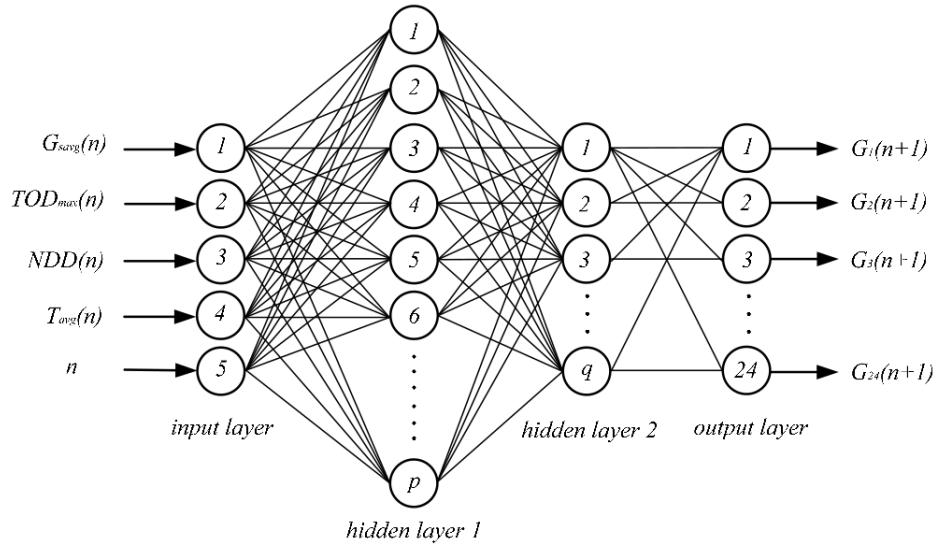


Figura 2.4: Modelo de pronóstico ANN utilizando parámetros de características estadísticas.

[19] señala dos enfoques para obtener radiación neta diaria utilizando los productos de Kalpana VHRR y Oceansat OCM2 mediante el manejo de la radiación de onda larga resultante del flujo de ondas entrantes y salientes, este resultado es computarizado usando la ecuación de Stefan Boltzmann para corregir la humedad y la nubosidad, también se realiza un enfoque basado en la estimación mediante la computación de un cielo despejado con el flujo de ondas cortas de radiación que permiten obtener una radiación neta

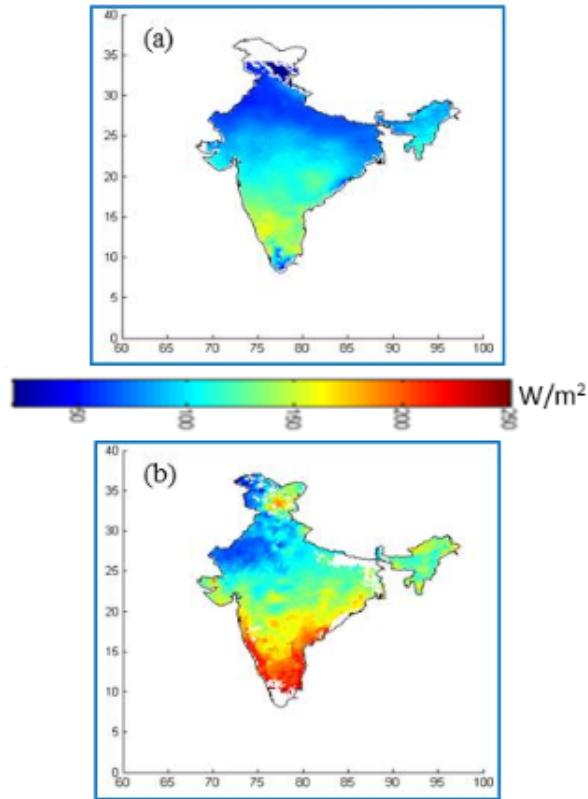


Figura 2.5: Radiación superficial diaria neta de productos de Kalpana VHRR y Ocean-sat OCM2.

[20],[21] generan información de cobertura de suelo, y los métodos que permiten obtener más detalle conservando una fiabilidad aceptable sobre la región del Tancítaro, Michoacán y comprende bosques templados y tropicales secos, pastizales y áreas de cultivos. Enfocan el estudio en índices de vegetación mediante compuestos espectrales de 8 días e imágenes de reflectancia diarias que fueron evaluados por medio de dos metodologías; la máxima verosimilitud y redes neuronales, en cada una de estas se incorporaron dos tipos de datos auxiliares. Los resultados muestran que es posible obtener mapas confiables a partir de estos datos de baja resolución si se usan categorías generales.

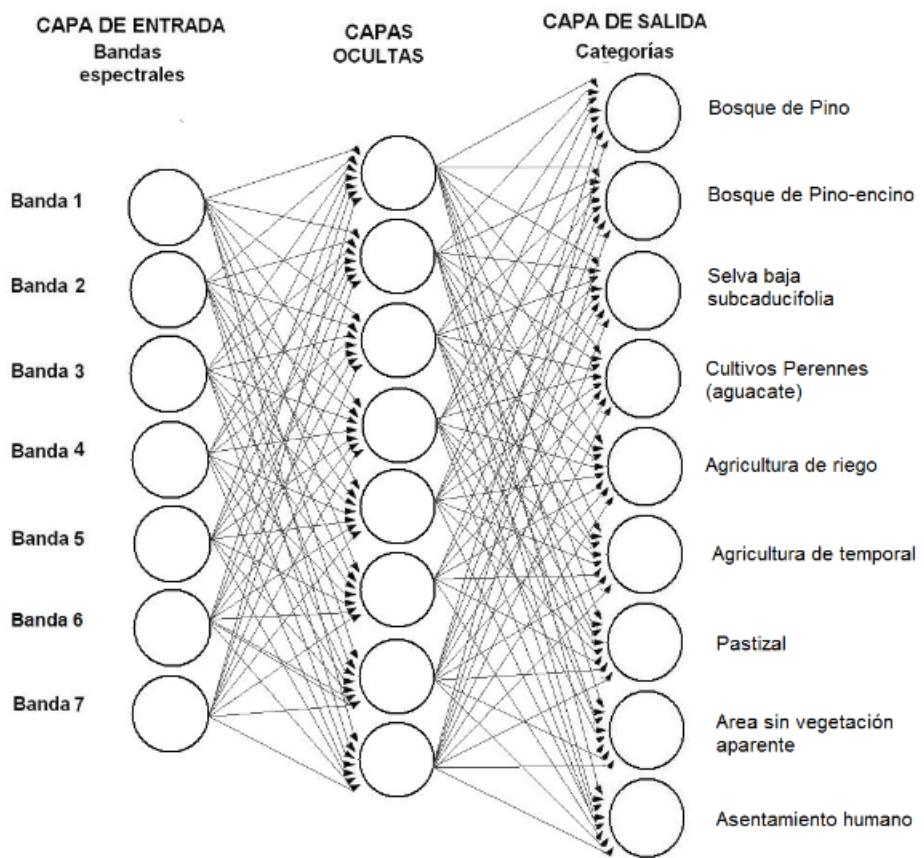


Figura 2.6: Red Neuronal perceptrón multicapa para clasificar una imagen de siete bandas en nueve categorías

[22] destaca la correlación existente entre la radiación solar incidente en la superficie terrestre y la radiación solar reflejada al espacio; Adicionalmente realiza un estudio comparativo en el análisis de imágenes satelitales MODIS respecto a imágenes satelitales provenientes de SRB, Earth Observing System (EOS) y Tropical Rainfall Measuring Mission (TRMM), en este estudio se destaca los métodos para estimar la radiación neta obtenida de ondas cortas medidas en la superficie.

[23] prueban diferentes algoritmos para la detección de aerosoles sobre la superficie terrestre contempladas dentro de imágenes satelitales MODIS, implementación del algoritmo Miller sobre las bandas de infrarrojo y adicionalmente se realiza una clasificación de los aerosoles detectado.

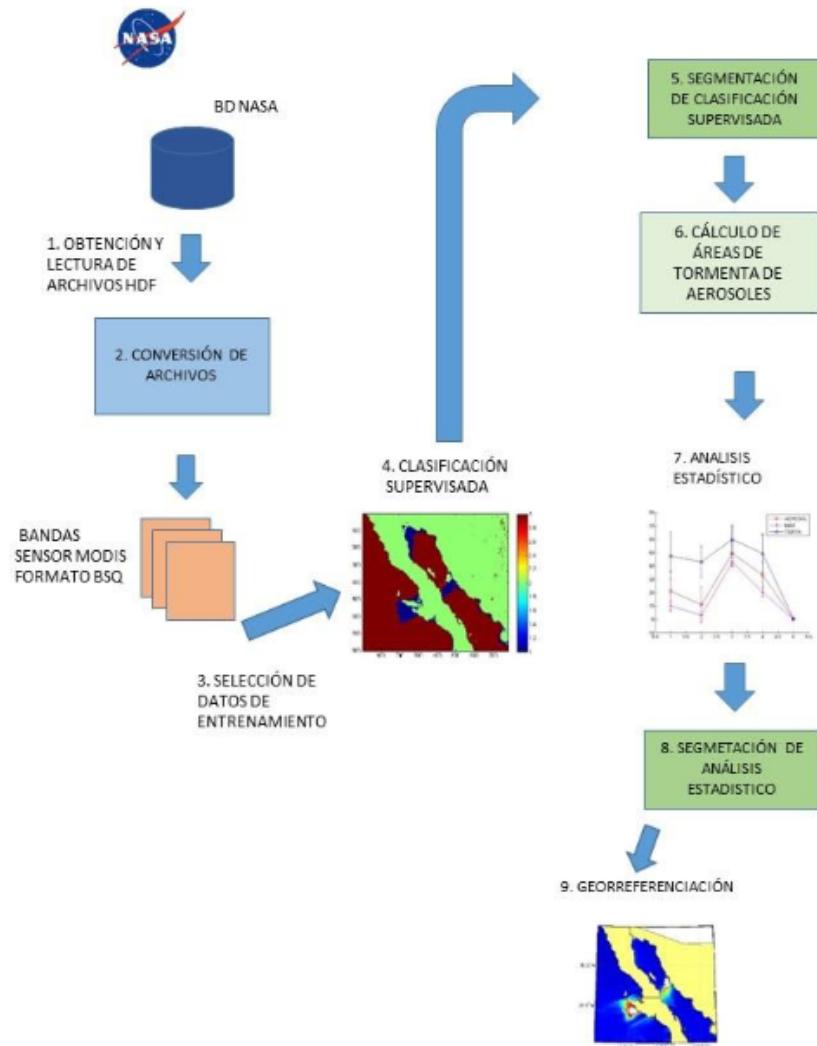
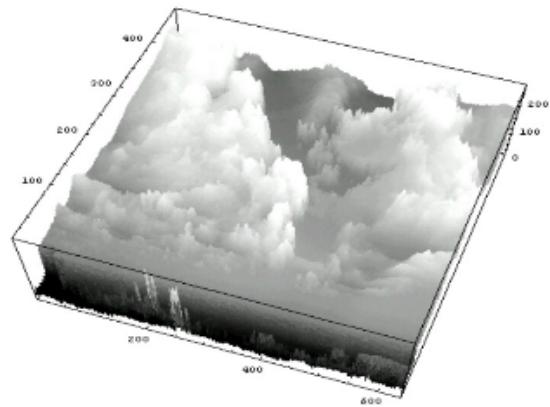


Figura 2.7: Detección de Aerosoles a partir de imágenes satelitales MODIS

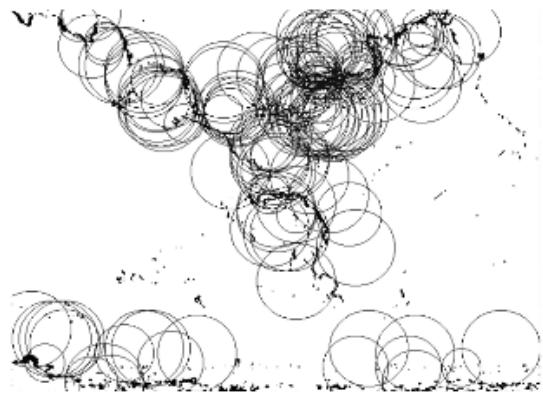
[24] calcula la caída de potencial y fluctuaciones en la generación de energía a partir del seguimiento tridimensional de nubes y su impacto sobre paneles solares.



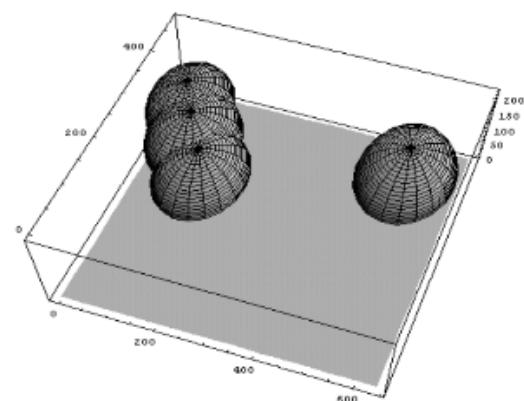
(a) Original image (observed image)



(a) Volumetric representation of gray level



(b) Processed image by the Hough transformation
(many small circles were detected)



(b) Detected sphere

3 METODOLOGÍA

Para poder cumplir todos los objetivos de la investigación, inicialmente se realizó la apropiación de conocimiento correspondiente con la temática y se definió el marco conceptual sobre el cual se iba a realizar el proyecto. Como primer paso fué importante entender la problemática que se pretende solucionar, identificar los recursos necesarios para el estudio y fijar las metas que se pretende alcanzar, las actividades mas relevantes empleadas en la investigación son detalladas a continuación:

3.1 Identificar y Adquirir Imágenes Satelitales

Se realiza el estudio de las imágenes satelitales más pertinentes para obtener la información, para la elección se tiene en cuenta 4 factores importantes *resolución espacial*, *resolución espectral*, *resolución radiométrica* y *resolución temporal* de la imagen satelital; Las imágenes de LandSat 7 son generadas cada 16 días y permiten adquirir datos de 8 bandas espectrales con una resolución de 15m por pixel para la banda 6 y 30m por pixel para las 7 bandas restantes, estas bandas tienen la capacidad de detectar diferentes indicadores para vegetación, reflectancia, temperatura, precipitación, nubosidad, etc. Por otra parte, las imágenes satelitales MODIS adquieren datos de 36 bandas espectrales con las cuales ofrece el producto MOD09GA con una resolución espacial aproximada de 500x500m en cada pixel, estos productos son generados diariamente y están diseñado para medir la reflectancia de la superficie terrestre[25][26], MOD09GA presenta 7 bandas que tienen relación directa con la delimitación de territorios, tipos de vegetación, incidencia de aerosoles, temperatura y reflectancia la cual está relacionada con la propiedad reflectiva de la vegetación y los aerosoles; mediante la propiedad reflectiva de la vegetación se pretende realizar una estimación para la radiación solar que se irradiana en una superficie.

Descarga de imágenes Satelitales

Las imágenes satelitales de MODIS están ubicadas en una grilla senoidal de aproximadamente 10x10 grados cada sección como lo muestra la figura 3.1. Esta grilla cuenta

con un índice vertical y un índice horizontal que permiten ubicar la imagen satelital en un área determinada; el territorio colombiano se encuentra ubicado en la columna 10 (h10) en la fila 8 (v8) como lo muestra la figura 3.2, estos 2 índices son necesarios para realizar la descarga de todas las imágenes satelitales comprendidas entre el año 2005 y 2015 para posteriormente construir la serie de tiempo de radiación solar. El gestor de descargas utilizado en esta ocasión es PyModis[27], herramienta que permite la creación de scripts para gestionar las descargas, cambio de formato, procesamiento y re-proyección de una imagen satelital a un sistema de coordenadas espaciales determinado.

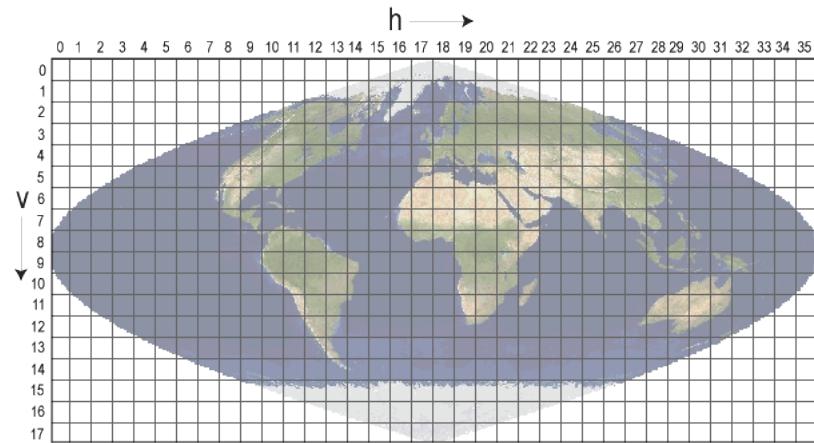


Figura 3.1: Grilla Senoidal de MODIS

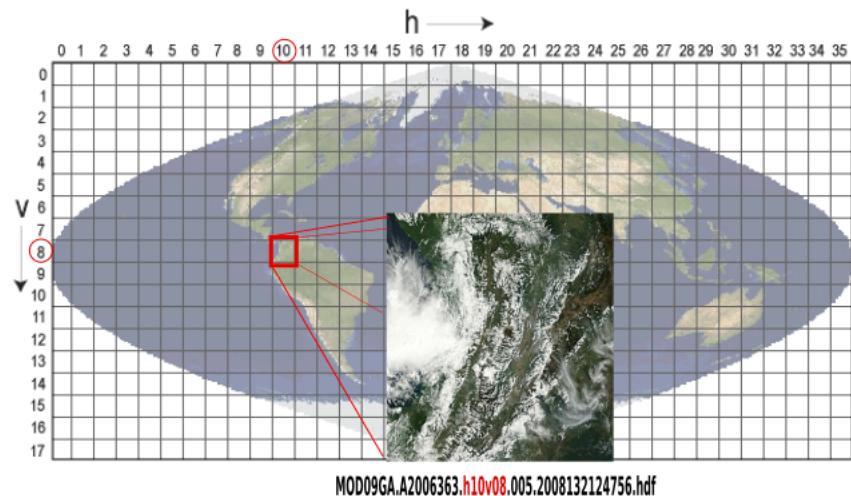


Figura 3.2: Ubicación del territorio colombiano en grilla senoidal de MODIS

Para el caso del sensor MODIS el territorio colombiano y sus 32 departamentos se

encuentra ubicado dentro de una sola imagen satelital, por este motivo solo se realizó la descarga diaria de una imagen durante los 11 años establecidos.

Por otra parte, las imágenes satelitales de LandSat 7 están ordenadas en una grilla de menor tamaño en comparación con la grilla de MODIS, para ubicar una imagen satelital en LandSat 7 se debe tener en cuenta PATH y ROW de las imágenes que logren la cobertura de una región definida como lo muestra la figura 3.3, EarthExplorer¹ permite identificar de manera sencilla las imágenes satelitales necesarias para la cobertura de un territorio.

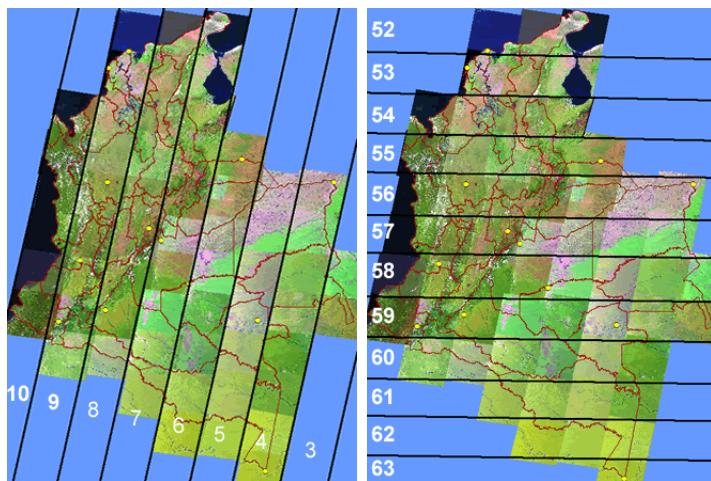


Figura 3.3: Path y Row de la Grilla sensor LandSat en Colombia - Fuente UNODC Colombia

Para el caso particular del departamento de Nariño, se logra una cobertura del territorio con 5 imágenes satelitales de LandSat 7 correspondientes a la lista de paths:rows (9:59), (9:60), (10:58), (10:59), (11:59) como lo muestra la figura 3.4.

Actualmente existe software especializado en manejo de imágenes satelitales; gsutil² dedicado explícitamente para la descarga de imágenes satelitales LandSat y pyModis³ que facilita la descarga y procesamiento a gran cantidad de imágenes satelitales proveniente del sensor MODIS. Para almacenar la información necesaria para la estimación de radiación solar sobre el departamento de Nariño se trabajó de forma independiente la manipulación de las imágenes LandSat y las imágenes de MODIS.

¹<http://earthexplorer.usgs.gov/>

²<https://cloud.google.com/storage/docs/gsutil>

³<http://pymodis.fem-environment.eu/>

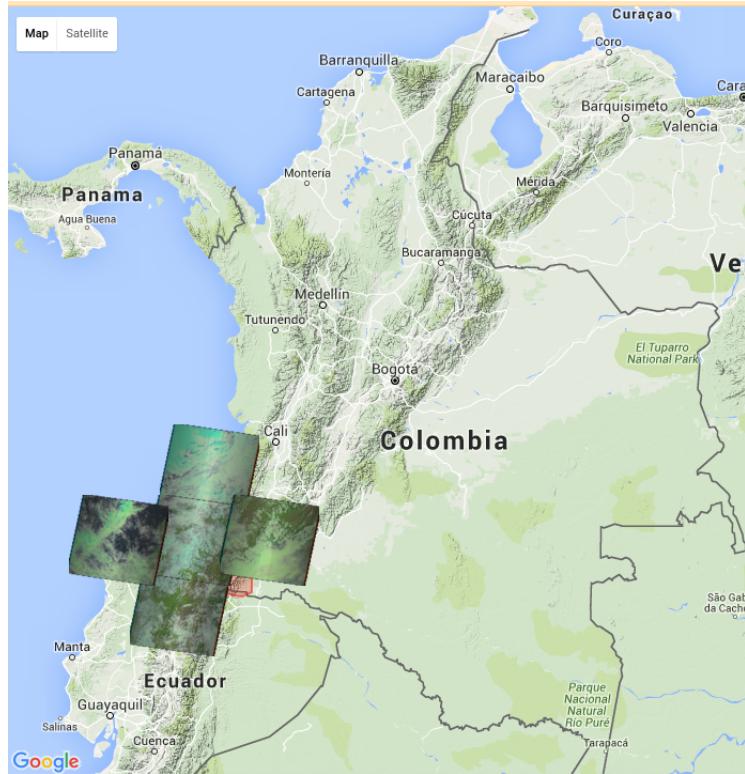


Figura 3.4: Paths y Rows necesarios para la cobertura del territorio Nariñense.

3.2 Procesar y Almacenar Información

Para el caso de MODIS se desarrolló un script dedicado al manejo de aproximadamente 3.500 imágenes satelitales MOD09GA, mediante el script se convierte el formato científico original de la imagen satelital(HDF) en un formato de imágenes de mapa de bits(TIFF) reproyectando el sistema de coordenadas espaciales original a EPSG:3857, se utiliza EPSG:3857 por la simplicidad en cálculos, interpolación, aproximación y manejo en BD. Posterior a la conversión de formato y reproyección, se realizó un recorte a la imagen satelital en el que se abarque el territorio de Nariño y se proceda a recorrer las 7 bandas en formato TIFF obtenidas del producto MOD09GA, a continuación se almacenan muestras de reflectancia con referencia geográfica a cada 450 metros como lo muestra la figura 3.5 y se aplica el filtro de nubes según las especificaciones de las bandas[28].

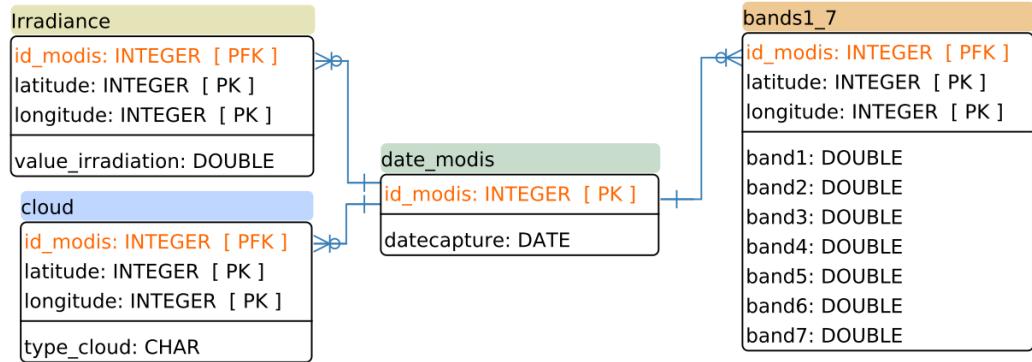


Figura 3.5: Esquema BD para almacenamiento de imágenes satelitales MOD09GA.

Las imágenes provenientes del sensor LandSat 7 ofrecen 8 bandas que se encuentran compresas y en formato TIFF, para manipular este tipo de imágenes se tiene en cuenta la reproyección del sistema de coordenadas espaciales original a EPSG:3857 al igual que se lo realizo con las bandas de MOD09GA, para lograr la cobertura del departamento de Nariño es necesario la unión de 5 imágenes satelitales ver figura 3.6 y 3.4, luego se procede a realizar el recorte de la imagen satelital usando el croquis del departamento de Nariño ver figura 3.7.

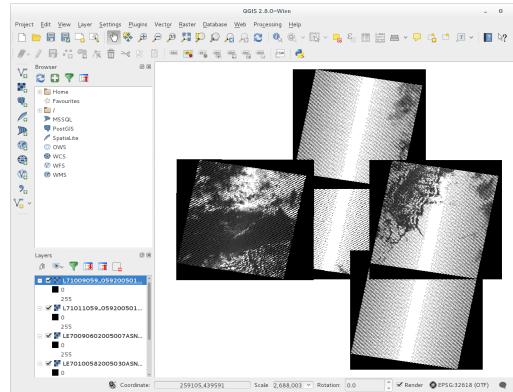


Figura 3.6: Imágenes satelitales LandSat para lograr la cobertura del departamento de Nariño.

Una vez terminado la reproyección y recorte de aproximadamente 1.300 imágenes satelitales, se procede a tomar muestras con referencia geográfica del Número Digital(DN) la imagen satelital, este número digital indica la escala de gris en la que se encuentra un pixel en determinada banda, esta escala de grises es interpretada de diferente forma en cada banda; una vez capturado el DN se procede a calcular la reflectancia, esto se

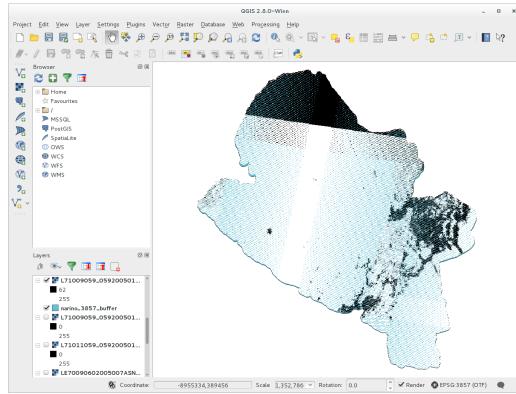


Figura 3.7: Recorte de imágenes satelitales LandSat ajustado al territorio Nariñense.

realiza calibrando la muestra mediante correcciones atmosferica, estimación la ganancia de las bandas, corrección de ángulo de insidencia del sol, determinación de la distancia de la tierra al sol y calculos o correcciones recomendados por LandSat[29].

La reflectancia obtenida es almacenada en BD ver figura 3.8 de igual forma se almacena información de datos descartados al aplicar un filtro para detectar información de nubes, gases, lagos, ciudades, etc.[30].

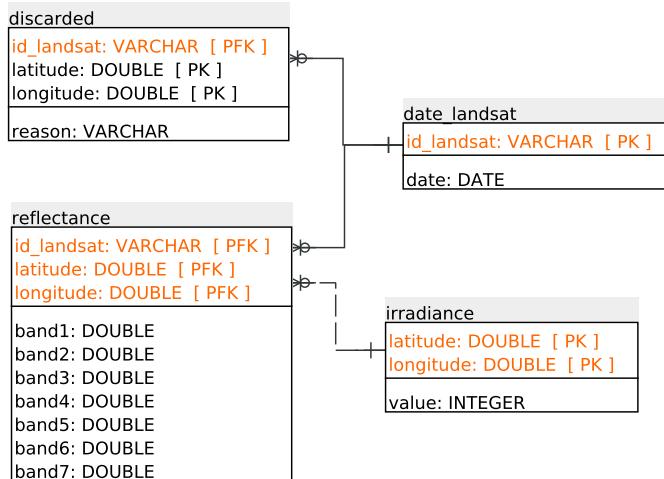
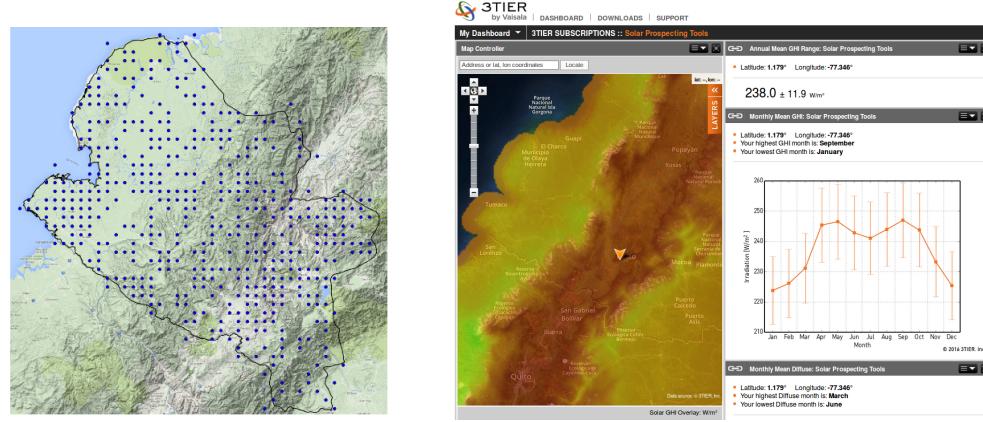


Figura 3.8: Esquema BD para almacenamiento de imagen satelital LandSat.

Finalizado el almacenamiento de las bandas presentes en la imagen satelital, se procede a realizar un muestreo de radiación solar georeferenciado proveniente de una fuente confiable de información, en este caso se seleccionó 500 puntos distribuidos uniformemente

sobre el territorio de Nariño y se procedió a consultar la información de 3TIER Figura 3.9 para construir modelos de regresión y extrapolar los datos. Para la información referente a nubes se discriminó tres tipos de nubosidad correspondientes a la escala de reflectancia[31] según la altura del aerosol.



(a) Muestras Uniformemente distribuidas en el Departamento de Nariño

(b) Muestras de radiación tomadas manualmente de 3TIER

Figura 3.9: Muestreo de radiación solar.

3.3 Selección, Aplicación del Modelo de Regresión y Construcción de Serie de Tiempo

En base a la información de las muestras recolectadas de 3TIER y la información recolectada de las bandas 1-7 del producto MOD09GA, se desarrolló un script en R para construir diferentes modelos de regresión, adicionalmente se evalúa cual es el modelo ideal para ser aplicado a la serie de tiempo y proceder a extrapolar los datos de reflectancia a datos de radiación solar, se realizó un análisis con 13 modelos para extrapolar los datos de reflectancia a datos de radiación.

Para la selección del modelo más adecuado se analizó las muestras de reflectancia del producto de MOD09GA y las muestras tomadas de 3TIER, mediante el análisis se obtuvo el resultado mostrado en la siguiente tabla 3.1.

Modelo	SAE	MAE	RAE	RMSE	COR	R ²
crtree	1360.88975	9.13349	60.07076	13.06545	0.64264	0.41299
rpart	1373.21922	9.21624	60.61499	14.14182	0.60074	0.36089
kknn	920.48535	6.17775	40.63096	10.31934	0.79280	0.62854
mlp	485.60361	3.25908	21.43493	4.51435	0.96284	0.92705
mlpe	443.11836	2.97395	19.55960	3.97730	0.97157	0.94394
ksvm	823.63424	5.52775	36.35587	8.12712	0.87861	0.77195
randomForest	1159.07988	7.77906	51.16271	11.04659	0.75129	0.56444
mr	597.58282	4.01062	26.37778	5.36391	0.94694	0.89670
mars	618.82261	4.15317	27.31532	5.47634	0.94475	0.89255
cubist	597.29188	4.00867	26.36494	5.36508	0.94688	0.89658
pcr	597.58282	4.01062	26.37778	5.36391	0.94694	0.89670
plsr	597.58282	4.01062	26.37778	5.36391	0.94694	0.89670
cppls	597.58282	4.01062	26.37778	5.36391	0.94694	0.89670

Tabla 3.1: Análisis de modelos de Regresión aplicados a datos de reflectancia de 7 bandas de MOD09GA.

Para encontrar el modelo más adecuado se realizó un análisis a los datos de reflectancia de LandSat y las muestras tomadas de 3TIER, los resultados se observan en la siguiente tabla 3.2.

Modelo	SAE	MAE	RAE	RMSE	COR	R^2
crtree	771.66185	5.32181	33.37397	8.89544	0.85991	0.73945
rpart	819.97501	5.65500	35.46349	9.23174	0.84774	0.71865
kknn	583.36151	4.02318	25.23008	6.19161	0.93584	0.87580
mlp	558.43603	3.85128	24.15206	5.49114	0.94968	0.90189
mlpe	461.93253	3.18574	19.97834	4.73616	0.96292	0.92721
ksvm	574.76656	3.96391	24.85835	5.71528	0.94664	0.89613
randomForest	663.70528	4.57728	28.70490	6.89480	0.92117	0.84856
mr	752.19550	5.18756	32.53206	6.75745	0.92222	0.85049
mars	680.67053	4.69428	29.43864	6.34212	0.93186	0.86837
cubist	538.20590	3.71176	23.27712	6.34056	0.93141	0.86752
pcr	748.89239	5.16478	32.38920	6.76538	0.92208	0.85023
plsr	748.89239	5.16478	32.38920	6.76538	0.92208	0.85023
cppls	748.89239	5.16478	32.38920	6.76538	0.92208	0.85023

Tabla 3.2: Análisis de modelos de Regresión aplicados a datos de reflectancia de 8 bandas de LandSat.

Según las métricas SAE, MAE, RAE, RMSE, COR y R^2 evaluadas en los 13 modelos [32], se logra concluir que el modelo más óptimo para extrapolar los datos de radiación es “mlpe”—multilayer perceptron ensemble, mlpe es un modelo implementado en R que se basa en una red neuronal artificial con la capacidad de detectar características en los datos de la bandas de las imágenes satelitales LandSat y el producto MOD09GA, posteriormente el modelo se encarga de asociarlos a un valor objetivo y determinar las métricas de relación más comunes. Al realizar el análisis de los modelos se logra construir los todos modelos de regresión que son utilizados para extrapolar los datos, en este caso se procede a construir la serie de tiempo de radiación solar mediante la aplicación del modelo más óptimo; mlpe se encarga de evaluar los datos de reflectancia contenidos en las 7 bandas de MOD09GA y asocia un resultado de radiación que es almacenado en BD, de la misma forma analiza las 8 bandas de LandSat almacenados en la relación Radiance de la BD ver figura 3.10 y entrega como resultado un valor de radiación que es almacenado en BD.

3.4 Mapas de Radiación Aplicando Interpolación Espacial

La manera más adecuada de visualizar la información es mediante la creación de mapas de promedios mensuales, anuales y un promedio general de radiación solar sobre el departamento de Nariño. Una vez construida la serie de tiempo, se procede obtener agregados mensuales con el objetivo de establecer cual es el mes con más alta o baja radiación solar, el promedio anual es necesario para establecer como es la variación de la

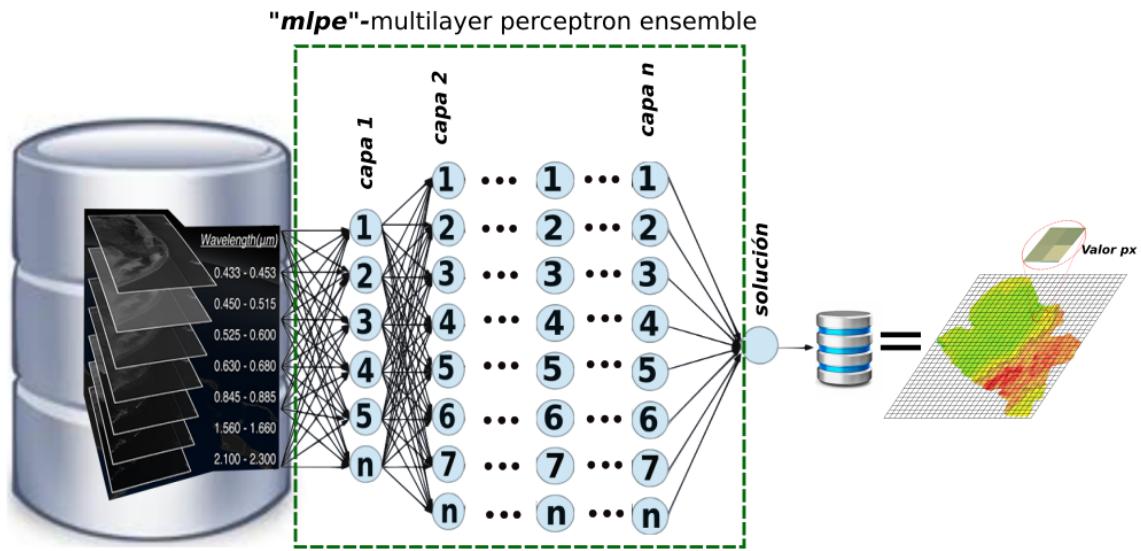


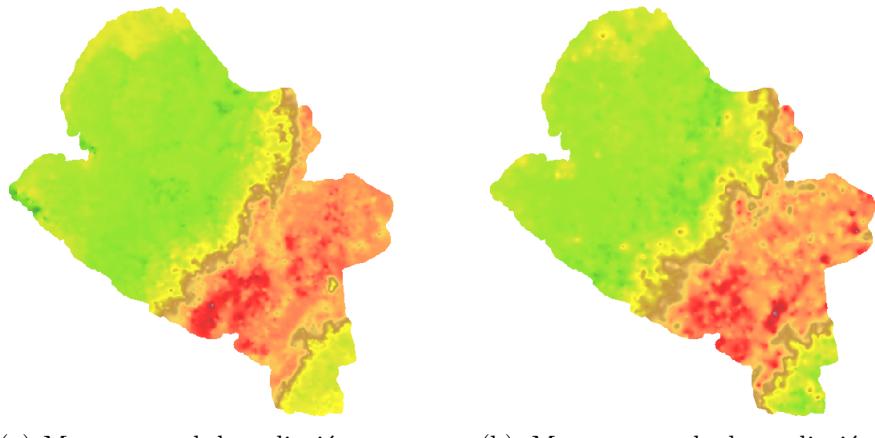
Figura 3.10: Proceso construcción serie de tiempo radiación solar a partir de imágenes satelitales LandSat y MODIS

radiación solar en el transcurso del tiempo y el promedio general permite establecer la radiación solar promedio en Nariño para los 11 años de MODIS o 15 años de LandSat. Una vez obtenidos los agregados mensuales y anuales se desarrolló un script en R que permite implementar la interpolación espacial aplicando el método kriging mediante el uso de una grilla de 450x450m que indica la resolución del pixel para el mapa de radiación generado, adicionalmente se debe parametrizar el sistema de referencia de coordenadas estándar EPSG:3857 y formato GeoTiff para los mapas ver figuras 3.4, 3.4y 3.4; estos parámetros sirven para cualquier manejo cartográfico o manipulación de los mapas en un SIG.

En términos generales el mapa de radiación de MODIS tiene mayor número de muestras en comparación con los mapas de LandSat, pero se puede observar similaridad en la imágenes al representar la estimación de radiación solar sobre el departamento de Nariño.

Los mapas en formato GeoTiff son almacenados en BD para mayor simplicidad en el manejo de información y reportes, de forma complementaria son presentados en la plataforma GEOAlternar⁴ ver figura 3.14 permitiendo visualizar mapas, datos de promedios mensuales, datos de promedios anuales, descargar datos, generar reportes y

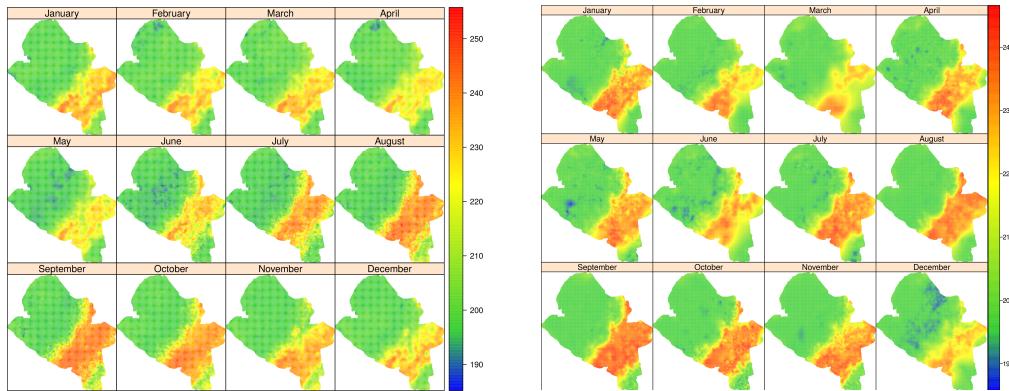
⁴<http://geoalternar.udenar.edu.co>



(a) Mapa general de radiación solar MODIS

(b) Mapa general de radiación LandSat

Figura 3.11: Promedios de radiación solar



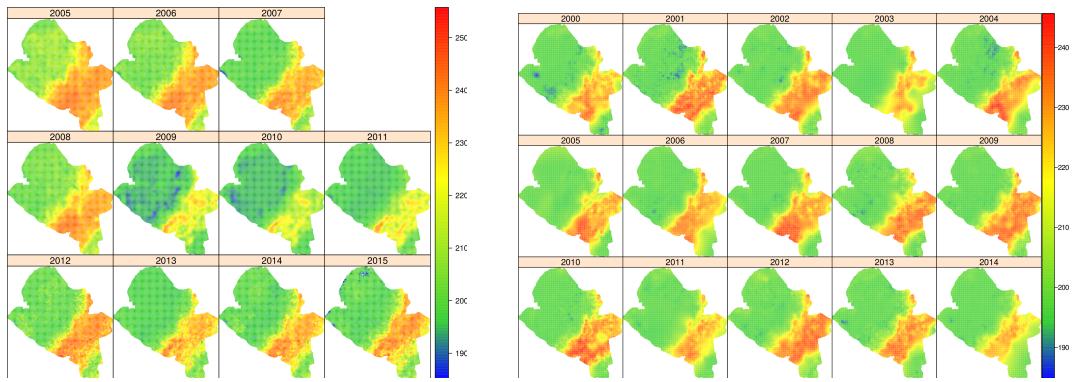
(a) Mapas mensuales de radiación solar MODIS

(b) Mapa mensuales de radiación solar LandSat

Figura 3.12: Promedios Mensuales

buscar información de radiación solar, eólica, biomasa e hídrica presente en cualquier punto dentro del territorio Nariñense.

La gran ventaja de crear los mapas de radiación se ve reflejada en la fácil identificación de zonas con potencial de radiación solar, esta información es importante para futuras instalaciones de plantas energéticas a base de paneles solares o energía térmica, adicionalmente se puede contemplar el comportamiento mes a mes o año tras año de la radiación solar en una área específica.



(a) Mapas anuales de radiación solar MODIS (2005-2015)

(b) Mapas anuales de radiación solar LandSat (2000-2014)

Figura 3.13: Promedios Anuales

3.5 Detección de Patrones Secuenciales

La detección de patrones secuenciales se realizó utilizando la serie de tiempo construida anteriormente mediante los datos del sensor MODIS y el modelo mlpe. Se aplicó un proceso de minería de datos a la serie de tiempo de radiación solar con el objetivo de extraer información de eventos frecuentes mediante la identificación de patrones secuenciales. Para este proceso se cuenta con las relaciones Irradiance y Datemodis ver figura 3.15 con un historial de 11 años de datos que abarcan aproximadamente 340 millones de registros, también se realizó la selección de los 300 mejores puntos ver figura 3.15 con mayor radiación solar dentro del departamento de Nariño, luego se sometió los registro histórico de radiación al algoritmo de LCM secuencial.

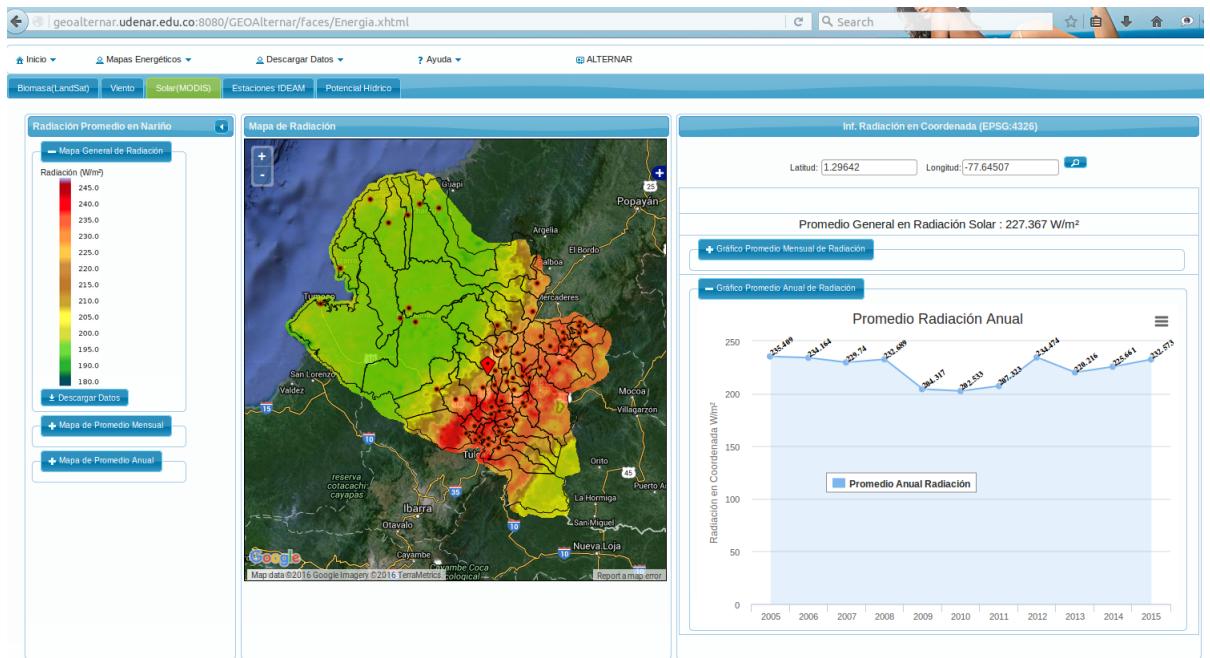
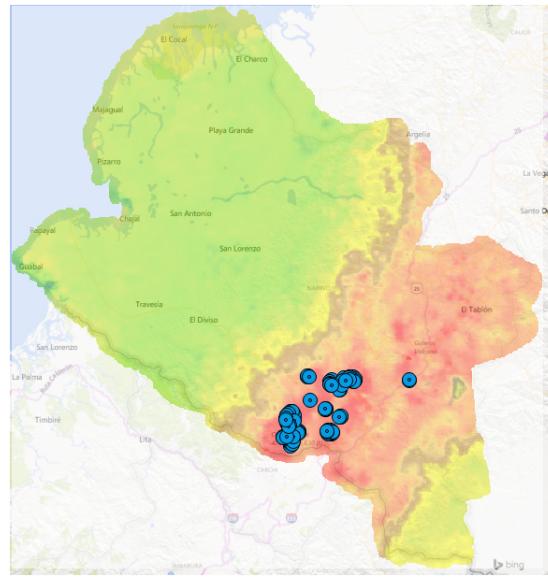


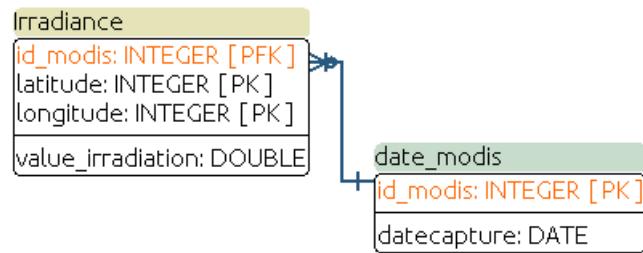
Figura 3.14: Plataforma GEOAlternar- mapas energéticos del departamento de Nariño

La detección de patrones se realizó mediante el uso del algoritmo LCM secuencial[14] el cual permite enumerar la frecuencia de acontecimientos presentes en una serie de tiempo; el algoritmo LCM secuencial cuenta con las funcionalidades necesarias para detección de patrones secuenciales por lo que es necesario establecer los parámetros para realizar la extracción de la información, para el proceso de minería se desarrolló un script en Python que permite ejecutar LCM secuencial y aplicarlo a la serie de tiempo de radiación solar utilizando los siguientes parámetros:

- Serie de tiempo de muestra de radiación solar.
- Generación de 8 bins los cuales permiten discriminar la radiación como se muestra en la tabla 3.3.
- Ventana de 30 días para evaluar la ocurrencia de acontecimientos.
- Soporte de 90 % para las transacciones.



(a) Mejores 300 puntos de radiación solar dentro del departamento de Nariño



(b) Relaciones para identificación de patrones secuenciales

Figura 3.15: Insumos para identificación de patrones secuenciales

Bin	Rango radiación solar	Representación
1	(x<195.985) - 195.985	
2	195.985 - 196.496	●
3	196.496 - 197.047	●
4	197.047 - 199.124	●
5	199.124 - 205.954	●
6	205.954 - 228.426	●
7	228.426 - 233.534	●
8	233.534 - (x>233.534)	●

Tabla 3.3: Bines utilizados en la detección y representación de patrones secuenciales.

Se implementó el algoritmo incluyendo los parámetros correspondientes, el pseudo-código base propuesto se contempla en Algorítmo 1.

Algoritmo 1 Identificación de Patrones Secuenciales

Input: BD Seie de Tiempo Radiación

Output: Patrones de Radiación Solar

```

1: ListaPatrones  $\leftarrow \emptyset$ 
2: for SerieTiempoMuestraReflectancialat,lon  $\in$  BD Serie de Tiempo Radiación do
3:   Soporte  $\leftarrow 90\%$ 
4:   Bines  $\leftarrow$  [Tabla 3.3]
5:   for Ventana  $\in$  SerieTiempoMReflectancialat,lon do
6:     Patron  $\leftarrow$  Aplicar LCM_seq(MReflectancia(lat,lon),Bines,Soporte) LCM Algorithm [14]
7:     FP  $\leftarrow$  PatronFrecuente
8:     DP  $\leftarrow$  PatronDiferente
9:     TP  $\leftarrow$  PatronTamaño
10:    end for
11:    ListaPatrones  $\leftarrow$  FP,DP,TP
12:    AlmacenarPatronDB
13: end for

```

Los resultados obtenidos son almacenados en BD ver figura 3.16 para posteriormente proceder al análisis y visualización. Tal y como muestra la figura 3.16 se creó la relación “patrones por día” que presenta 8 atributos donde los más relevantes para identificar patrones son detallados a continuación.

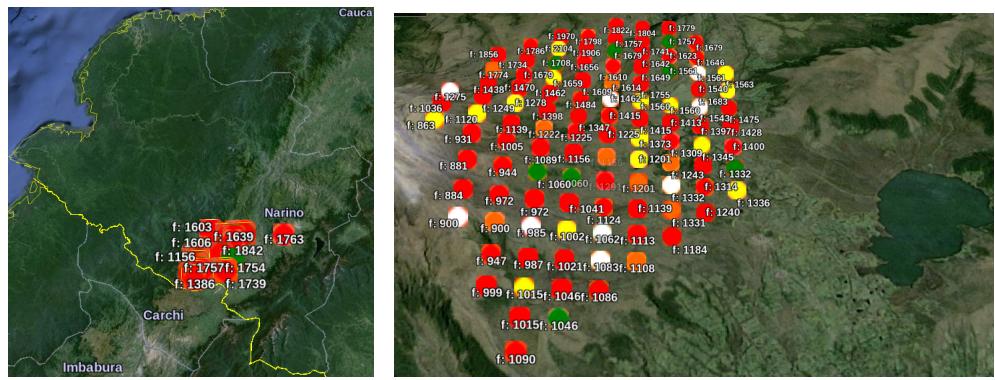
latitude-longitude \rightarrow Ubicación geográfica en el sistema de coordenadas EPSG:3857
pattern \rightarrow Representa la detección de un conjunto de bines del la tabla 3.3.
len_pattern \rightarrow Indica el número consecutivo de días de la ocurrencia de un evento.
frecuency \rightarrow Indica la frecuencia del patrón detectado.
solar \rightarrow Representa el conjunto de bines en el rango de la escala solar 3.3.
diff_solar \rightarrow Registra los tipos de bines que hay en el patrón.

Para la representación visual de los patrones detectados en los 300 mejores puntos de radiación, se desarrolló un script para generar un archivo kml que permite visualizar la ubicación geográfica, días presentes y frecuencia del patrón como lo muestra la figura 3.17, el archivo es desplegado sobre la plataforma google-earth.

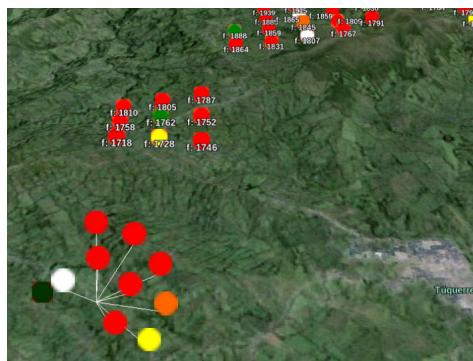
Los patrones detectados anteriormente permiten establecer un periodo de tiempo con presencia de radiación solar constante, en este periodo se puede aprovechar al máximo las plantas solares debido a que se puede garantizar una óptima radiación solar por un numero determinado de días.

id integer	latitude integer	longitude integer	pattern character varying[]	len_pattern integer	frequency integer	solar character varying[]	diff_solar integer
511442 -8670600	98100	{'1','8','8','8','8','8','8','7','8','8','8','8','8'}	12	1306	{(0,195.985),(233.534,inf),(233.534	3	
511448 -8670600	98100	{'1','8','8','8','8','8','8','6','8','8','8','8'}	13	1303	{(0,195.985),(233.534,inf),(233.534	3	
511438 -8670600	98100	{'1','8','8','8','6','8','6','8','8','8','8','8'}	12	1166	{(0,195.985),(233.534,inf),(233.534	3	
511441 -8670600	98100	{'1','8','8','8','6','1','8','8','8','8','8'}	10	1390	{(0,195.985),(233.534,inf),(233.534	3	
511445 -8670600	98100	{'1','8','8','8','8','8','8','8','8','8','8','8'}	14	1082	{(0,195.985),(233.534,inf),(233.534	3	
511447 -8670600	98100	{'1','8','8','8','8','8','8','8','8','8','8','8'}	13	1390	{(0,195.985),(233.534,inf),(233.534	2	
511436 -8670600	98100	{'1','8','8','8','8','6','8','8','8','8','8','8'}	13	1351	{(0,195.985),(233.534,inf),(233.534	3	
511437 -8670600	98100	{'1','8','8','8','8','6','8','8','8','8','8','8'}	11	1137	{(0,195.985),(233.534,inf),(233.534	3	
511439 -8670600	98100	{'1','8','8','8','6','8','1','8','8','8','8','8'}	11	1157	{(0,195.985),(233.534,inf),(233.534	3	
511440 -8670600	98100	{'1','8','8','8','6','8','8','8','8','8','8','8'}	11	1427	{(0,195.985),(233.534,inf),(233.534	3	
511443 -8670600	98100	{'1','8','8','8','8','8','8','8','8','8','8','8'}	13	1278	{(0,195.985),(233.534,inf),(233.534	3	
511444 -8670600	98100	{'1','8','8','8','8','8','8','8','8','8','8','8'}	16	1251	{(0,195.985),(233.534,inf),(233.534	2	
511446 -8670600	98100	{'1','8','8','8','8','8','8','8','8','8','8','8'}	13	1418	{(0,195.985),(233.534,inf),(233.534	3	
511450 -8670600	98100	{'1','8','8','8','8','8','7','8','8','8','8','8'}	12	1258	{(0,195.985),(233.534,inf),(233.534	3	
511449 -8670600	98100	{'1','8','8','8','8','8','1','8','8','8','8','8'}	13	1303	{(0,195.985),(233.534,inf),(233.534	2	

Figura 3.16: Base de datos resultado de aplicación LCM secuencial



(a) Ubicación geográfica de los patrones de radiación en departamento de Nariño
(b) Acercamiento a zona con patrones de radiación solar



(c) Detalle de patrón de radiación solar

Figura 3.17: Visualización de patrones de radiación solar

3.6 Comportamiento de Nubes

El comportamiento de las nubes está estrechamente ligado a la producción de energía a base de paneles de radiación solar o energía térmica, por este motivo fué necesario construir un registro histórico referente a la presencia de nubes, este registro permite evidenciar la presencia de aerosoles sobre el departamento de Nariño, haciendo posible la identificación de la presencia de de nubosidad en una zona con potencial de radiación solar. La identificación de nubes se realizó mediante filtros aplicados a las muestras de reflectancia, el filtro se implemento teniendo en cuenta el tratamiento y las recomendaciones necesarias que se deben dar a las bandas de la imagen satelital[26][28][30].

Para identificar el comportamiento de las nubes se utilizó el registro histórico con el objetivo de establecer la probabilidad para la presencia de nubes en una zona determinada. Como primera medida se construyó una BD donde se almacena la probabilidad de la presencia de nubes en un punto determinado para construir la BD fué necesario realizar un conteo de los días del mes en los que se presentó nubes en cada punto dentro del departamento de Nariño, posteriormente se agrupo mensual y anualmente para calcular la probabilidad y establecer un promedio para la presencia de nubes en un punto determinado dentro del departamento de Nariño.

Los promedios mensuales permiten establecer como es el comportamiento de las nubes en el tiempo para un punto específico dentro del territorio Nariñense, la información de nubosidad tiene varios campos de aplicación, con estos datos se hace posible prever la presencia nubosidad sobre plantas de energía solar en un periodo determinado y tomar medidas adicionales para mitigar los efectos, por ejemplo, si se logra determinar en que zonas hay periodos de abundante radiación pero también existe una época donde se presenta gran nubosidad; esta información permite contrarrestar los periodos de baja radiación mediante el uso de generadores de energía con fuentes alternativas como eólica, hídrica o biomasa.

A continuación se realiza un análisis del comportamiento de nubes en un punto específico dentro del departamento de Nariñofiguras 3.18, este análisis se lo puede realizar en cualquier zona que se encuentre dentro del territorio nariñense debido a que el las muestras de nubes detectadas también son interpoladas aplicando el método de kriging para el todo el departamento.

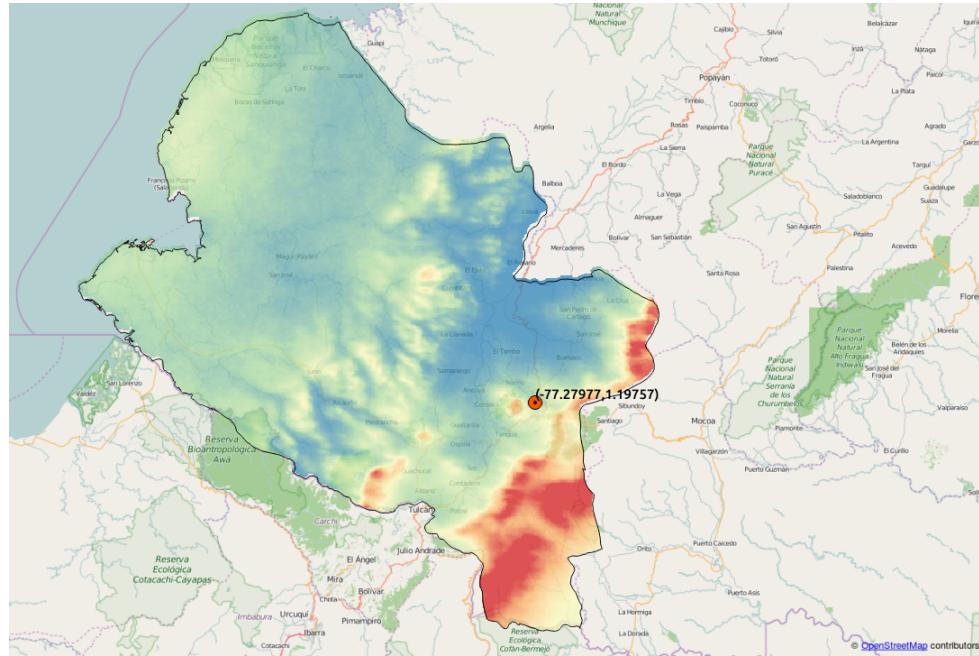


Figura 3.18: Comportamiento mensual de nubes en un punto específico

En las figuras 3.19 y 3.20 se logra observar los registros en la coordenada (-77.27977,1.19757); como lo muestra el el gráfico 3.19 se observa que en 5 meses el porcentaje de nubosidad supera el 20 % permitiendo establecer que 7 dias del mes el territorio se encuentra nublado por lo que se hace necesario tener una alternativa energética diferente para solventar la posible reducción de energía a base de paneles solares, por otro lado el comportamiento de las nubes en el transcurso de los 11 años de información es regulado debido a que se encuentra alrededor del 20 % lo que indica que en promedio 75 dias del año esta zona se encuentra cubierta de nubes.

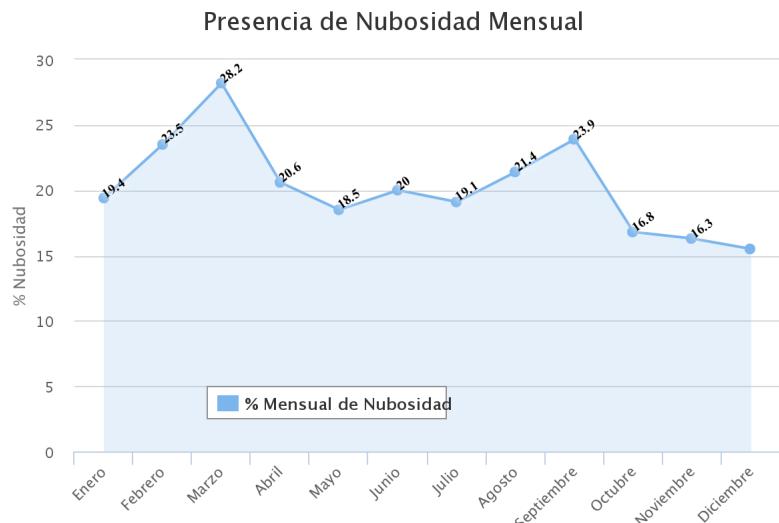


Figura 3.19: Comportamiento mensual de nubes en un punto específico

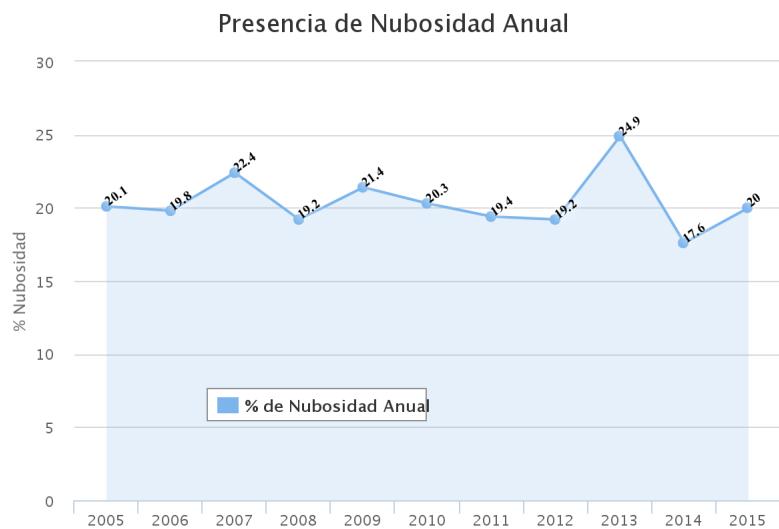


Figura 3.20: Comportamiento Anual de nubes en un punto específico

4 VALIDACIÓN DE LA INFORMACIÓN CON ESTACIONES ACTIVAS

Actualmente en el departamento de Nariño se han instalado alrededor de 25 estaciones con la capacidad de reportar datos radiación solar, viento, precipitación, temperatura y humedad del aire Figura 4.1; de las 25 estaciones instaladas se realizó la selección de 11 estaciones Figura 4.2, esta decisión fué aplicada por inconvenientes en las estaciones como tiempos cortos de reporte de estaciones, debido a que muchas estaciones fueron instaladas a finales del año 2015, este corto tiempo de reporte no permite medir con gran certeza la relación entre la serie de tiempo construida y los reportes de las estaciones que fueron instaladas hace pocos meses, otros factores que provocaron que se descartaran varias estaciones son los errores en el reporte de datos ocasionados por fallas en la calibración de los dispositivos, interrupciones en reporte de información, diferencia en tipo de dispositivo y factores exógenos que inciden de forma importante en el reporte de datos. En la validación es necesario gran cantidad de muestras que per-

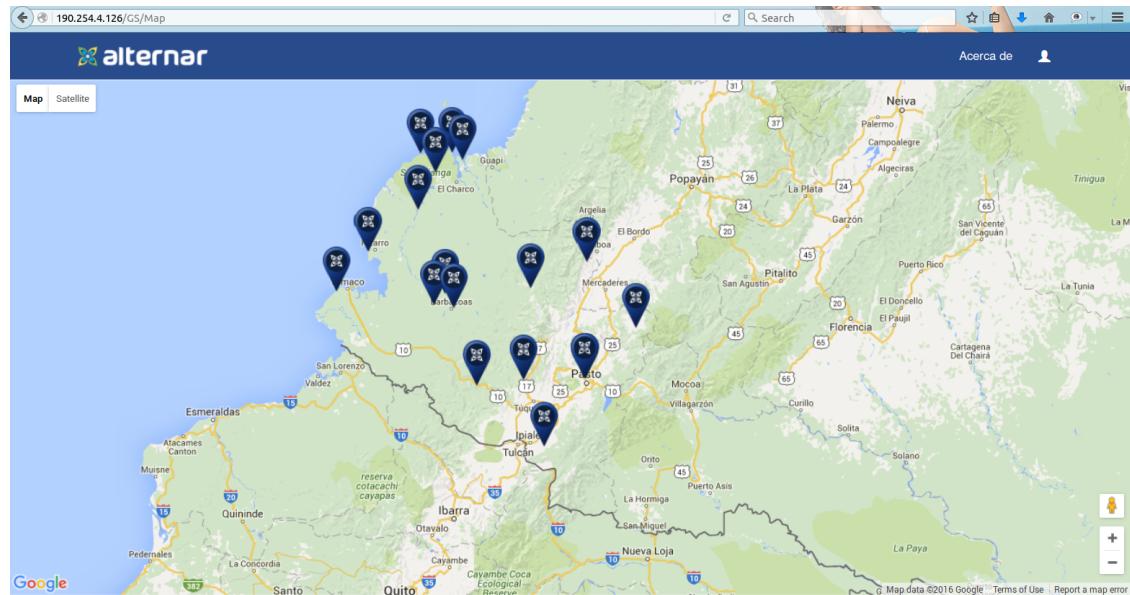


Figura 4.1: Plataforma de reporte y ubicación de estaciones del proyecto Alternar

mitan establecer una mayor certeza en el análisis de los datos reales con los datos de las imágenes satelitales, por este motivo es importante destacar que las imágenes satelita-

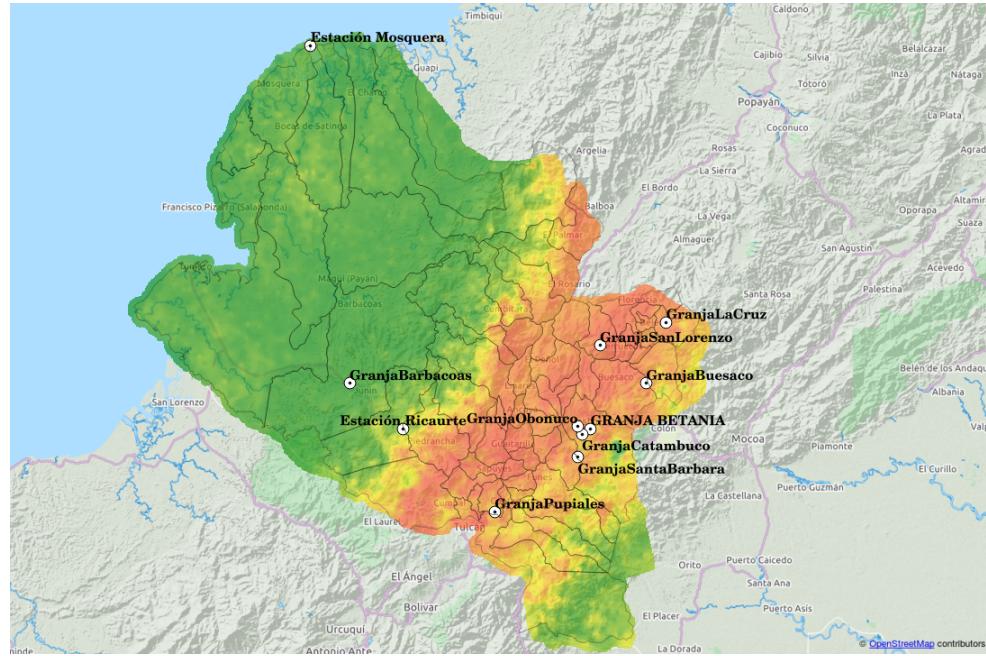


Figura 4.2: Estaciones seleccionadas para proceder a validación de datos de radiación solar

les MODIS tienen una resolución temporal diaria mientras que las imágenes satelitales LandSat tiene una resolución temporal cada 16 días, por este motivo se descartó para la validación los datos de LandSat.

Al disponer del registro histórico de las 11 estaciones, se procede a realizar la validación de los datos de radiación solar; para el proceso de validación es necesario obtener los promedios mensuales de las 11 estaciones y los promedios mensuales de radiación de la serie de tiempo construida ubicando la zona correspondiente, los promedios obtenidos permiten establecer una correlación buena entre los datos de radiación solar de las estaciones y los datos de radiación solar de MODIS. A continuación se puede observar los resultados obtenidos del análisis realizado a las 11 estaciones con reportes superiores a 5 meses y la serie de tiempo de radiación solar obtenida del sensor MODIS.

Estación	Ubicación	R^2
Estación Mosquera	(-78.333333,2.655)	86,57 %
Estación Ricaurte	(-77.975833,1.186944)	90,33 %
Granja Barbacoas	(-78.18173,1.361483333)	74,09 %
Granja Betania	(-77.257028,1.186317)	84,64 %
Granja Buesaco	(-77.04079,1.35932)	76,55 %
Granja Catambuco	(-77.290397,1.166761)	60,82 %
Granja La Cruz	(-76.9640701,1.5955594)	91,26 %
Granja Obonuco	(-77.305556,1.193458)	71,52 %
Granja Pupiales	(-77.624667,0.867483)	90,47 %
Granja San Lorenzo	(-77.219125,1.503858)	65,11 %
Granja Santa Barbara	(-77.305544,1.078422)	81,27 %

Tabla 4.1: Análisis datos de radiación solar Reporte Estaciones Vs Serie de Tiempo MODIS.

5 ANÁLISIS DE RESULTADOS

Al comparar los resultados obtenidos al construir los mapas de radiación solar de Landsat con lo de MODIS se puede observar gran similitud, la manera de establecer relación entre los datos es mediante un análisis de correlación y coeficiente de determinación entre los mapas resultantes con los datos de Landsat7 y MODIS. Se tomo los mapas por años del 2005 al 2014 como lo muestra la tabla 5, por meses y el mapa general ver tabla 5

Adicionalmente se muestra resultados de datos reales reportados por las estaciones y se los compara con la serie de tiempo construida a partir de imágenes satelitales MODIS, los resultados obtenidos permiten concluir que la serie de tiempo de radiación solar construida para todo el departamento de Nariño es fiable y puede servir como punto de referencia para varios estudios que contemplen la temática.

Estación	R^2
Estación Mosquera	86,57 %
Estación Ricaurte	90,33 %
Granja Barbacoas	74,09 %
Granja Betania	84,64 %
Granja Buesaco	76,55 %
Granja Catambuco	60,82 %
Granja La Cruz	91,26 %
Granja Obonuco	71,52 %
Granja Pupiales	90,47 %
Granja San Lorenzo	65,11 %
Granja Santa Barbara	81,27 %

Tabla 5.1: Resultado Análisis de Datos para Radiación Solar Reporte Estaciones Vs Serie de Tiempo MODIS.

Map	COR	R2
2005	0.85594	0.73263
2006	0.90754	0.82363
2007	0.92884	0.86275
2008	0.90594	0.82073
2009	0.89443	0.80000
2010	0.88594	0.78488
2011	0.88233	0.77851
2012	0.92056	0.84743
2013	0.92982	0.86456
2014	0.93167	0.86800

Tabla 5.2: Análisis Datos Anuales de MODIS vs Landsat 7

Map	COR	R2
January	0.89182	0.79535
February	0.89131	0.79444
March	0.90943	0.82706
April	0.90930	0.82683
May	0.91770	0.84218
June	0.90244	0.81439
July	0.90434	0.81783
August	0.91527	0.83772
September	0.91810	0.84290
October	0.92340	0.85266
November	0.87242	0.76111
December	0.86934	0.75576
General	0.94179	0.88696

Tabla 5.3: Análisis Datos Mensuales de MODIS vs Landsat 7

5.1 Análisis de Patrón y Nubes en un punto Específico

Los patrones identificados en la serie de tiempo de los mejores 300 puntos de radiación dentro departamento de Nariño permiten determinar acontecimiento frecuentes respecto a los días en que se presenta mayor o menor radiación solar, por ejemplo para el punto (0.97067,-77.84763) Figura 5.1, se identificó el patrón [‘8’,’1’,’7’,’8’,’6’,’8’,’8’,’8’,’8’,’8’] el significado de este patrón esta representado en la tabla 5.1, este patrón se presentó 1320 veces y permite concluir que de los 10 días de radiación hay 7 días en los cuales se presenta radiación superior a 233.5 W/m² y se puede garantizar que este punto es una zona con buena radiación solar, esta información es necesaria para instalar plantas energéticas a base de paneles solares siempre que la frecuencia de nubes sea baja.

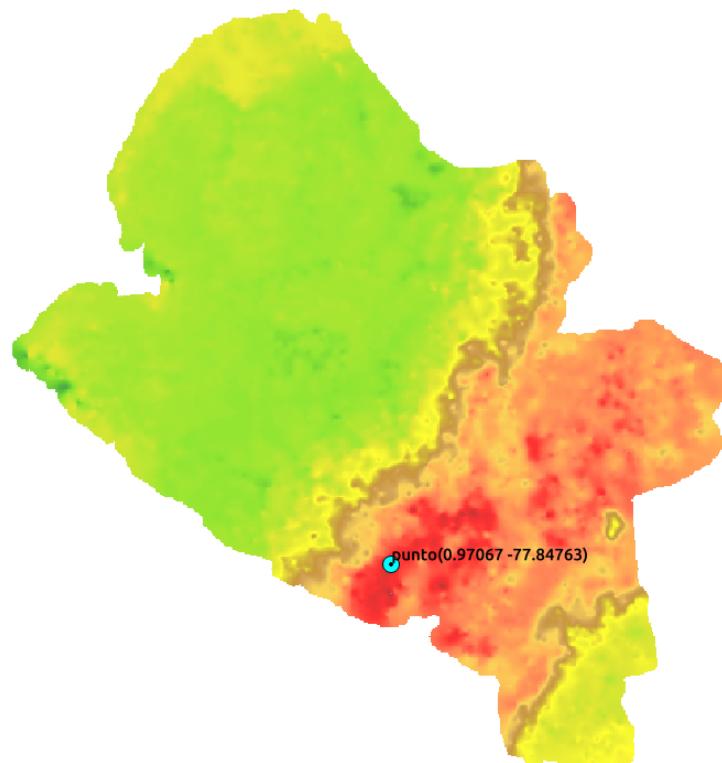


Figura 5.1: Plataforma GEOAlternar- mapas energéticos del departamento de Nariño

N. Dias	Bin BD(Patrón)	Color	Rango
1	8	●	233.534 - x>233.534
2	1		0 - 195.985
3	7	●	228.426 - 233.534
4	8	●	233.534 - x>233.534
5	6	●	205.954 - 228.426
6	8	●	233.534 - x>233.534
7	8	●	233.534 - x>233.534
8	8	●	233.534 - x>233.534
9	8	●	233.534 - x>233.534
10	8	●	233.534 - x>233.534

Tabla 5.4: Análisis de Patrón en el punto (0.97067,-77.84763)

La probabilidad de que en el punto (0.97067,-77.84763) se presenten nubes es de 28 % en cuanto al cambio en el tiempo se puede observar en la tabla 5.1 y la tabla 5.1, esta información es muy importante debido a que la presencia constante de nubes termina afectando la producción energética a base de paneles solares y así poder plantear una solución alternativa para generar energía.

Año	Probabilidad
2005	26.1475 %
2006	26.8242 %
2007	31.5918 %
2008	31.1175 %
2009	29.7392 %
2010	26.9092 %
2011	26.0167 %
2012	28.0983 %
2013	28.2908 %
2014	26.67 %
2015	26.9967 %

Tabla 5.5: Nubes en el punto (0.97067,-77.84763)

Mes	Probabilidad
January	29.676 %
February	31.9309 %
March	35.7782 %
April	25.7582 %
May	27.5673 %
June	31.5164 %
July	28.4464 %
August	26.3918 %
September	27.2718 %
October	23.225 %
November	21.667 %
December	26.129 %

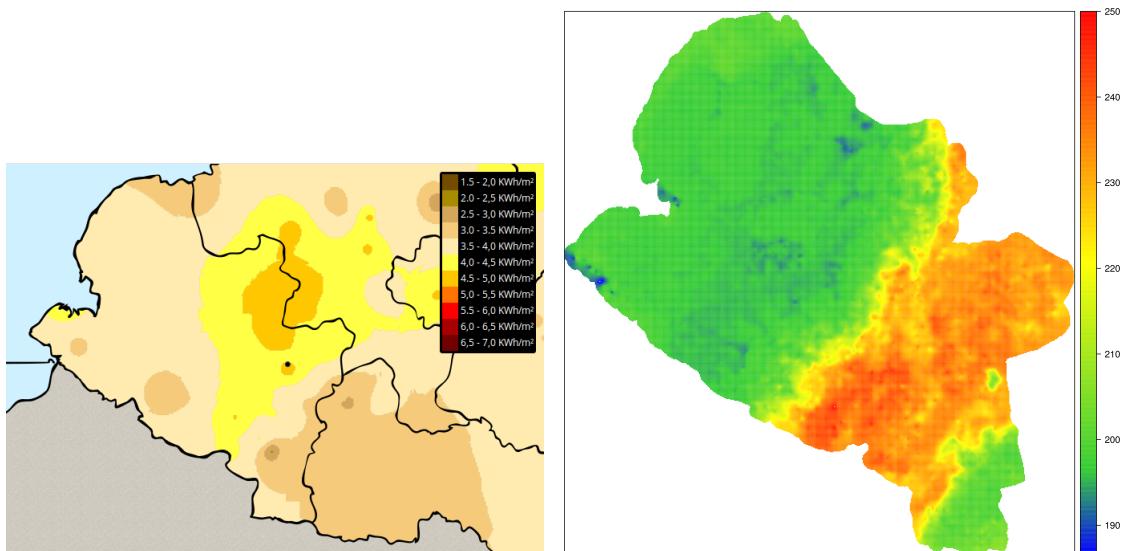
Tabla 5.6: Nubes en el punto (0.97067,-77.84763)

6 CONCLUSIONES

- * Se detectó patrones secuenciales que permiten predecir el comportamiento de fenómenos climáticos y su impacto en áreas con potencial de radiación solar dentro del departamento de Nariño.
- * Se construyeron mapas energéticos con el componente solar en el departamento de Nariño, usando los sensores Landsat 7 y MODIS, además se construyeron series de tiempo para radiación solar y nubosidad.
- * Las imágenes satelitales son una gran fuente de información debido a la capacidad de almacenar gran cantidad de registros históricos para diferentes tipos de datos, estos datos poco a poco están siendo utilizados por organizaciones para determinar características terrestres, fenómenos naturales, condiciones de los mares, características de la vegetación, etc. Por esta razón el uso de imágenes satelitales en la investigación da resultados aproximados y a bajo costo, teniendo en cuenta el costo que puede implicar hacer muestreo en campo.
- * Como se observó en la validación de los datos resumida en la Tabla 4.1 hay un buen porcentaje de fiabilidad para la estimación del radiación solar en las estaciones activas; mediante este hecho se puede plantear que el registro histórico construido para todo el departamento de Nariño, permite tener gran certeza para la toma de decisiones al momento de instalar plantas a base de energía solar en una zona determinada; un resultado agregado al estudio es la identificación de la presencia de nubosidad la cual es fundamental para una óptima generación de energía a base de paneles solares o energía térmica, adicionalmente se puede utilizar los datos de radiación para contemplar los efecto en el tiempo que presenta el sol sobre la vegetación.
- * La detección de patrones en la serie de tiempo permite establecer un periodo de tiempo para un óptimo aprovechamiento de las plantas solares debido a que se garantiza un número determinado de días en los cuales se presentará altas tasas de radiación solar y baja probabilidad de nubes. La metodología para la

detección de patrones ya se encuentra establecida para poder ser aplicada a todo el departamento.

- * Las herramientas de software libre son adaptable a las necesidades de los usuarios y productos como LandSat y MODIS de libre descarga presenta estanadres de calidad para los cualquier tipo de estudio.
- * Se realizó una mejora a las estimaciones de radiación presentes en la actualidad, debido a que el reporte generado por el IDEAM no permite observar la diferencia ni cambios de radiación sobre el territorio nariñense Figura 6.



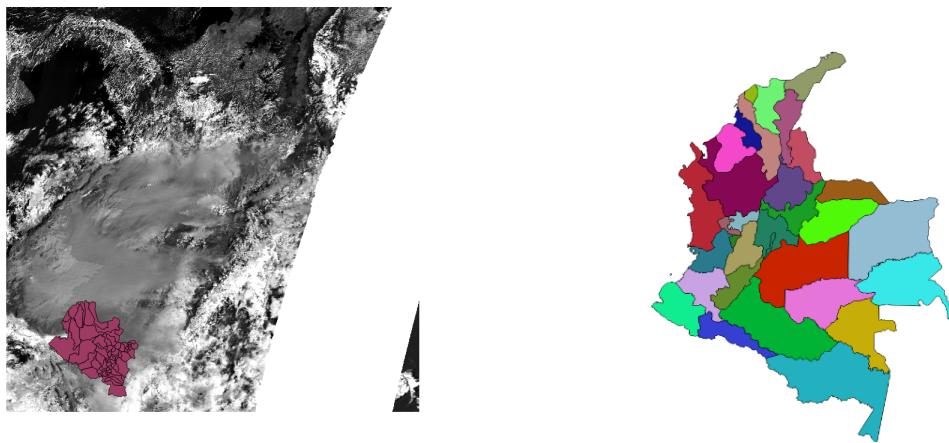
(a) Información de radiación solar presentada por el IDEAM para el departamento de Nariño (b) Información de radiación solar producto de la investigación para el departamento de Nariño

- * Se construyó una metodología para la construcción de mapas energéticos con el potencial de radiación solar, el cual se lo puede aplicar en zonas pantropicas de Colombia utilizando imágenes satelitales de los sensores Landsat o MODIS.
- * Además se obtuvieron los siguientes productos:
 - 11 mapas solares por año con el sensor MODIS.
 - 12 mapas solares por mes con imágenes satelitales del sensor MODIS.
 - 16 mapas solares por año con el sensor LandSat.
 - 12 mapas solares por mes con imágenes satelitales del sensor LandSat.

- Serie de tiempo de radiación solar diaria para el departamento de Nariño.
- Serie de tiempo nubosidad diaria presente en el departamento de Nariño.
- Algoritmo para detección de patrones secuenciales solares.
- Algoritmo de detección de patrones secuenciales para nubosidad.
- “GEOAlternar“ Plataforma para la visualización del potencial energético dentro del departamento de Nariño.
- Articulo para revista, el cual esta en proceso de evaluacion denominado: Análisis de regresión para el calculo de irradiación solar en el departamento de Nariño (Colombia) utilizando imágenes satelitales Landsat y MODIS.

7 TRABAJOS FUTUROS

La construcción de la serie de tiempo puede ser replicada para todo el territorio colombiano, debido a que se planteó una metodología y actualmente no se cuenta con una fuente que permita ver a detalle estos tipo de variables climáticas. En la Figura 7 se puede observar el producto MOD09GA y la dimensión con respecto al territorio de Nariño, con esta observación se puede concluir que hay facilidad para replicar la metodología hacia los 31 departamentos que componen a Colombia.



(c) Tamaño de la imagen satelital respecto al departamento de Nariño (d) Dimension de territorio colombiano

Las imágenes satelitales permiten versatilidad de estudios debido a las propiedades de los sensores con los que cuenta cada satélite, por estas características se puede enfocar estudios en:

- Comportamiento y tipos de vegetación en una región.
- Temperatura sobre una región.
- Precipitaciones de un territorio.
- Estudios de aerosoles sobre una región.
- Incendios y efectos sobre las regiones.

- Comportamiento de los mares en el tiempo.
- Identificación de Biomasa.
- Determinar propiedades de la cuencas hídrica.

Modelar un sistema de micro-redes que contemplen una planta de energía a base de paneles solares, el sistema contará con la serie de tiempo de radiación solar la probabilidad de presencia de nubes, patrones secuenciales aplicados a la serie de tiempo de radiación solar en una zona determinada y contemplando la implementación de fuentes alternas de energía utilizando la información recolectada hasta ahora por el proyecto Alternar.

REFERENCIAS

- [1] Y. Zheng and X. Zhou, *Computing with Spatial Trajectories*. Springer Science and Business Media, LLC, 2011.
- [2] I. M. Peláez, “Modelos de regresión: lineal simple y regresión logística.” revista-seden, 2006, pp. 195–214.
- [3] A. Pitarque, J. C. Ruiz, and J. F. Roy, “Las redes neuronales como herramientas estadísticas no paramétricas de clasificación,” *Psicothema*, vol. 12, no. Suplemento, pp. 459–463, 2000.
- [4] A. Pitarque, J. C. Ruiz, and J. Roy, “Redes neurales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación,” *Psicológica: Revista de metodología y psicología experimental*, vol. 19, no. 3, pp. 387–398, 1998.
- [5] M. Maneta and S. Schnabel, “Aplicación de redes neuronales artificiales para determinar la distribución espacial de la humedad del suelo en una pequeña cuenca de drenaje: estudios preliminares,” *Estudios de la Zona no Saturada del Suelo*, vol. 6, pp. 295–304, 2003.
- [6] D. J. Matich, “Redes neuronales: Conceptos básicos y aplicaciones.” 2001.
- [7] M. S. L. N. S. M. Tim Sutton, Otto Dassau, “Interpolación espacial en detalle,” url`https://docs.qgis.org/`, 2000.
- [8] F. Coaguila, Tangerino and Gonçalves, “Análisis comparativo de métodos de interpolación espacial utilizando datos de evapotranspiración,” in *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE*, 2013. [Online]. Available: <http://www.dsr.inpe.br/sbsr2013/files/p1616.pdf>
- [9] J. C. Riquelme, R. Ruiz, and K. Gilbert, “Mineria de datos: Conceptos y tendencias,” *Revista Iberoamericana de Inteligencia Artificial*, vol. 10, no. 29, pp. 11–18, 2006.

- [10] J. Han, H. Cheng, D. Xin, and X. Yan, “Frequent pattern mining: current status and future directions,” *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 55–86, 2007.
- [11] C. Creighton and S. Hanash, “Mining gene expression databases for association rules,” *Bioinformatics*, vol. 19, no. 1, pp. 79–86, 2003.
- [12] R. Agrawal, R. Srikant *et al.*, “Fast algorithms for mining association rules,” in *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, 1994, pp. 487–499.
- [13] J. Han and M. Kamber, *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.
- [14] T. Uno, M. Kiyomi, and H. Arimura, “Lcm ver.3: Collaboration of array, bitmap and prefix tree for frequent itemset mining,” in *Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations*, ser. OSDM ’05. New York, NY, USA: ACM, 2005, pp. 77–86. [Online]. Available: <http://doi.acm.org/10.1145/1133905.1133916>
- [15] K. Kaku and C. W. Potter, “Creating high-resolution solar information from satellite imagery and numerical weather prediction modelling,” in *Proceedings of the 47th ANZSES Annual Conference, Townsville, Queensland Australia, 29th Sept–2nd October 2009*, 2009. [Online]. Available: http://c0402442.cdn.cloudfiles.rackspacecloud.com/static/ttcms/1.0.0.26/us/documents/publications/ANZSES_Kaku_and_Potter.pdf
- [16] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. G. Beyer, “Solar energy assessment using remote sensing technologies,” *Remote Sensing of Environment*, vol. 86, no. 3, pp. 423–432, Aug. 2003. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S003442570300083X>
- [17] H. M. Diagne, M. David, P. Lauret, and J. Boland, “Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids,” Ph.D. dissertation, American Solar Energy Society, 2012. [Online]. Available: https://ases.conference-services.net/resources/252/2859/pdf/SOLAR2012_0617_full%20paper.pdf
- [18] F. Wang, Z. Mi, S. Su, and H. Zhao, “Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters,”

- Energies*, vol. 5, no. 12, pp. 1355–1370, May 2012. [Online]. Available: <http://www.mdpi.com/1996-1073/5/5/1355/>
- [19] S. Sai Krishna, P. Manavalan, and P. Rao, “Estimation of net radiation using satellite based data inputs,” *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, pp. 307–313, 2014.
- [20] J.-F. Mas, “Aplicaciones del sensor modis para el monitoreo del territorio,” *Méjico, ISBN*, pp. 978–607, 2011.
- [21] T. J. García-Mora and J.-F. Mas, “Evaluación de imágenes del sensor modis para la cartografía de la cobertura del suelo en una región altamente diversa de méxico,” *Boletín de la Sociedad Geológica Mexicana*, vol. 63, no. 1, pp. 83–94, 2011.
- [22] H.-Y. Kim, “Estimation of land surface radiation budget from modis data,” 2008.
- [23] L. Y. G. ÁLVAREZ, “Procesamiento de imágenes para la teledetección de aerosoles en el noroeste de méxico.”
- [24] T. Hashimoto and Y. Nagakura, “Prediction of output power variation of solar power plant by image measurement of cloud movement,” *Journal of Advanced Research in Physics*, vol. 2, no. 2, 2012. [Online]. Available: <http://85.122.26.3/jarp/index.php/jarp/article/viewArticle/86>
- [25] C. f. E. R. O. LP DAAC User Services, U.S. Geological Survey (USGS) and S. (EROS), “Mod09ga,” url<https://lpdaac.usgs.gov>, 2000.
- [26] N. W. P. Policy and I. Notices, “Modis,” url<http://modis.gsfc.nasa.gov>, 2000.
- [27] F. E. Mach, “pymodis-open source python based library to work with modis data,” url<http://pymodis.fem-environment.eu>, 2010.
- [28] N. W. P. Policy and I. Notices, “Modis band specifications,” url<http://modis.gsfc.nasa.gov/about/specifications.php>, 2010.
- [29] N. I. Gasparri, M. G. Parmuchi, J. Bono, H. Karszenbaum, and C. L. Montenegro, “Utilidad de imágenes landsat 7 etm+ de diferentes fechas para la estimación de biomasa aérea en bosques subtropicales secos de argentina,” in *XII Congreso de la Asociación Española de Teledetección*, 2007, pp. 19–21.

- [30] C. Cea, J. Cristóbal, P. Serra, and X. Pons, “Mejoras en la detección semiautomática de nubes y sombras en imágenes landsat,” in *XI Congreso Nacional de Teledetección, Puerto de la Cruz. Tenerife*, 2005.
- [31] J. Li, W. P. Menzel, Z. Yang, R. A. Frey, and S. A. Ackerman, “High-spatial-resolution surface and cloud-type classification from modis multispectral band measurements,” *Journal of Applied Meteorology*, vol. 42, no. 2, pp. 204–226, 2003.
- [32] P. Cortez, “Compute classification or regression error metrics.” url`http://www.inside-r.org/packages/cran/rminer/docs/mmetric`, 2005.

ANEXOS

- Anexo1: Manuales para al descarga, procesamiento de imágenes satelitales MODIS y construcción de series de tiempo.
- Anexo2: Articulo Análisis de regresión para el calculo de radiación solar en el departamento de Nariño (Colombia) utilizando imágenes satelitales Landsat y MODIS finalizado para ser evaluado y publicado en eventos de SIG.
- Anexo3: Archivo patrones.mp4 donde se puede visualizar con mayor detalle los patrones aplicados a los 300 mejores puntos de radiación solar dentro del departamento de Nariño.
- Anexo4: Repositorio <https://github.com/edixred/Solar.git> donde se encuentran scripts para procesamiento,Modelo mlpe, Mapas de radiación, kml, listas de 300 mejores puntos, herramientas, etc.