

MEJORAMIENTO DE LA CALIDAD DE LA INFORMACIÓN DE
SENSORES REMOTOS PARA PRODUCTOS DE VEGETACIÓN Y
EVAPOTRANSPIRACIÓN, ALTIPLANO DE MÉXICO.

Ing. Jaime Alberto Arteaga Vargas.



Universidad Autónoma de Querétaro

Facultad de Ingeniería

MEJORAMIENTO DE LA CALIDAD DE LA
INFORMACIÓN DE SENSORES REMOTOS PARA
PRODUCTOS DE VEGETACIÓN Y
EVAPOTRANSPIRACIÓN, ALTIPLANO DE MÉXICO.

TESIS

Que como parte de los requisitos para obtener el grado de

MAESTRO EN CIENCIAS
(HIDROLOGÍA AMBIENTAL)

Presenta:

Ing. Jaime Alberto Arteaga Vargas

Dirigida por:

Dr. Enrique González Sosa.



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría En Hidrología Ambiental

**MEJORAMIENTO DE LA CALIDAD DE LA INFORMACIÓN DE SENSORES
REMOTOS PARA PRODUCTOS DE VEGETACIÓN Y EVAPOTRANSPIRACIÓN,
ALTIPLANO DE MÉXICO.**

TESIS

Que como parte de los requisitos para obtener el Grado de
Maestría en Ciencias
(Hidrología Ambiental)

Presenta:

Ing. Jaime Alberto Arteaga Vargas.

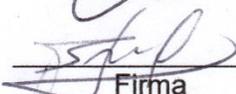
Dirigido por:

Dr. Enrique González Sosa.

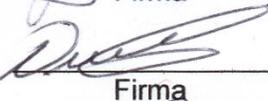
Dr. Enrique González Sosa
Presidente


Firma

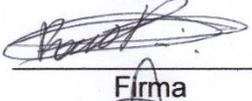
Dr. Pascal Breil
Secretario


Firma

Dr. Eusebio Jr. Ventura Ramos
Vocal


Firma

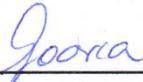
Dra. Rocío Becerril Piña
Suplente


Firma

M.C. María Alejandrina Leticia Montes León
Suplente


Firma


Dr. Manuel Toledano Ayala.
Director de la Facultad


Dra. Ma. Guadalupe Flavia Loarca Piña
Director de Investigación y Posgrado

Centro Universitario
Querétaro, Qro.
Enero 2019

RESUMEN

El campo de estudio de la teledetección ha crecido rápidamente en los últimos 15 años debido en a la necesidad de mantener un constante monitoreo la superficie terrestre, a través de estas técnicas nos es posible obtener mediciones de grandes extensiones de superficie para el monitoreo de los cambios en la cobertura vegetal, uso de suelo, temperatura, etc. La información obtenida con estos métodos nos ofrece una perspectiva única de la superficie de la tierra, sus recursos y el impacto que tienen las actividades humanas sobre esta. Al manipular esta información nos encontraremos con problemas a causa de la presencia de nubes en la atmosfera resultando en valores de reflectancia muy pequeños o anormales, así como valores faltantes debido a problemas de bandeo comúnmente ocasionados por una mala calibración de los sensores remotos. En los últimos años se han desarrollado una gran variedad de algoritmos computacionales empleados para la estimación de valores faltantes a través de información obtenida de imágenes de satélite, si bien es verdad que estos procesos se han visto mejorados, aún se encuentran limitados cuando se trata de analizar grandes cantidades de datos, debido a que año con año se agregan nuevas mediciones a los registros históricos. Este aumento en el tamaño de las bases de datos representa un gran problema por la exigencia en el poder de procesamiento y almacenamiento del equipo de cómputo. El método gapfill nos permitirá estimar valores faltantes a través de bases de datos de cuatro dimensiones por medio de métodos de interpolación espacial combinados con el análisis de series de tiempo además de la posibilidad de ser ejecutado en paralelo y de esta manera reducir los tiempos de procesamiento y obtener un mejor aprovechamiento de la capacidad de procesamiento de nuestros equipos.

Palabras clave: Teledetección, MODIS, Productos de vegetación, gapfill.

SUMMARY

Remote Sensing has improved over the last 15 years due to the necessity to keep a constant monitoring over the earth surface, by using these new techniques it possible to obtain measurements of huge areas over the surface for the monitoring of land coverage, temperature and soil classification. The information obtained by this method let us have a unique perspective of the earth surface, our resources and the impact of human activities over it. It's until we use satellite images when we face problems caused by a cloudy atmosphere, or a bad calibration or remote sensors, that will cause wrong values of reflectance or missing values over the image. On the las years it has been developed lots of methods for the estimation of missing values on satellite data, it is true that it represents a huge improvement, but it starts to complicate it when these methods face to the processing of huge amount of data that year by year keeps increasing the size of our data bases. Gapfill will let us estimate missing values by processing four dimensions array trough a spatial interpolation and time series analysis with the possibility to be executed in parallel to reduce processing time and take advantage of the whole processing capacity of our computers.

Key words: Remote sensing, MODIS, Vegetation products, gapfill.

DEDICATORIAS

A mi madre Elizabeth por todo su amor, apoyo y cada uno de los sacrificios que ha hecho los cuales me han llevado a ser quien soy y a estar en donde me encuentro, agradezco a dios por darme la mejor de las madres en el mundo.

A mi hermana Mary por estar siempre conmigo, por compartir esos momentos inolvidables y tantas horas de juego desde pequeños.

A Mariana, esa maravillosa e increíble mujer que tengo por novia y a quien amo más que a nadie, a ti por motivarme y estar ahí en todo momento junto a mí, contigo he compartido muchos de los momentos más felices de mi vida y he vivido experiencias que jamás olvidaré, te has convertido en una parte fundamental e irremplazable en mi vida.

A mis abuelos Antonio por todo por todo el apoyo a mis padres, los sacrificios que ha tenido que hacer, así como todo el amor y cuidados que ha tenido conmigo.

A mi tía Claudia quien ha estado junto a mi toda mi vida y que me ha dado tanto, esa maravillosa tía que trae diversión a mi vida y a quien quiero como a una madre.

A mi padre José por todo lo que me ha enseñado, por apoyarme en todo momento y brindarme su ayuda cuando lo necesito.

A mi tío Humberto por ser un ejemplo a seguir y por haberme apoyado y querido tanto desde pequeño, por todo lo que me ha enseñado y continúa enseñándome.

A Freya, esa perrita tan hermosa a quien amo con todo mi corazón, por ser una de mis más grandes motivaciones junto a mi familia y mi novia, siempre tendrás un lugar especial en mi corazón.

A mi bisabuela Lupita por todo el amor que me dio desde pequeño, por criarme junto a esa gran familia que tengo, desearía que estuvieras aquí con nosotros y que nada hubiera cambiado, te amo y siempre te llevé en mi corazón.

A cada uno de los miembros de mi increíble familia por sus enseñanzas y todo el amor pues a cada uno de ellos le debo lo que soy.

AGRADECIMIENTOS.

Al Dr. Enrique González Sosa por permitirme trabajar junto a él, por enseñarme tanto durante esta etapa en la maestría como profesor y director, por permitirme e impulsarme a realizar una estancia en el extranjero la cual sin lugar a duda ha colaborado de manera positiva en mi formación y que se convirtió en una de las mejores experiencias en mi vida, sin él esto nada de esto habría sido posible

Al Dr. Pascal Breil, un gran amigo e investigador que me apoyó y enseñó tanto durante mi estancia y a quien tengo el honor de incluir como parte de mi sínodo.

Al Dr. José Antonio Quevedo Tiznado por ayudarme a pulir mis conocimientos en programación, por ser un gran amigo y profesor, una de las personas con mayor dedicación y compromiso que conozco.

Al Dr. Israel Ruíz González por todas sus enseñanzas, un gran amigo que conozco desde mi formación en la licenciatura, a quien le debo todos mis conocimientos en programación y quien despertó en mí ese interés por la Hidrología.

Al Dr. Eusebio Ventura Jr. por ser un gran profesor desde la licenciatura y por colaborar con su experiencia y conocimientos en mi formación.

A mi gran amigo Samuel con quien compartí tantos momentos durante la maestría y quien me ha enseñado tantas cosas tanto de ingeniería como de anime y videojuegos.

A otro gran amigo, Edgar a quien yo sí le dedicaré un párrafo para agradecerle por momentos tan divertidos que compartimos durante la maestría, por tantas escapadas con el "Don" por el tiempo que fuimos al gym y esos sabios consejos que siempre me brinda, un ejemplo a seguir como estudiante.

A cada una de las personas que conocí durante la maestría por convertirse en grandes amigos, a Elvick, Aletz, Clara, Poyul, Paulina y los changs.

A mis mejores amigos Jorge, Ramón y Paco a quienes siempre tengo en mente y aunque ahora estén lejos siempre formarán parte de mi vida.

Finalmente, un agradecimiento especial a Conacyt pues gracias al apoyo que brindan a los estudiantes es posible que el campo de la investigación siga avanzando.

ÍNDICE GENERAL

RESUMEN.....	i
SUMMARY.....	ii
DEDICATORIAS.....	iii
AGRADECIMIENTOS.....	iv
ÍNDICE GENERAL.....	v
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	x
1. INTRODUCCIÓN.....	14
2. HIPOTESIS.....	17
3. OBJETIVOS.....	17
3.1 OBJETIVO GENERAL:	17
3.2 OBJETIVOS PARTICULARES:	17
4. REVISIÓN DE LITERATURA.....	18
4.1 TELEDETECCIÓN.....	18
4.2 SENSORES REMOTOS.....	19
4.2.1 <i>Sensores Pasivos</i>	20
4.2.2 <i>Sensores Activos</i>	20
4.3 REFLECTANCIA.....	20
4.4 ENERGÍA ELECTROMAGNÉTICA.....	21
4.5 ESPECTRO ELECTROMAGNÉTICO.....	22
4.6 IMÁGENES SATELITALES.....	23
4.7 EJEMPLOS DE APLICACIONES DE IMÁGENES SATELITALES.....	25
4.7.1 <i>Modelo Digital de Elevaciones</i>	25

4.7.2	<i>Mapas de clasificación</i>	25
4.7.3	<i>Evapotranspiración</i>	26
4.8	CARACTERÍSTICAS DE LOS SENSORES REMOTOS	26
4.8.1	<i>Resolución Espacial</i>	27
4.8.2	<i>Resolución Temporal</i>	27
4.8.3	<i>Resolución Espectral</i>	27
4.8.4	<i>Tamaño de Escena</i>	27
4.9	PRODUCTOS DE IMÁGENES.....	28
4.9.1	<i>NDVI</i>	28
4.9.2	<i>EVI</i>	28
4.9.3	<i>LAI</i>	29
4.9.4	<i>Albedo</i>	29
4.10	CAUSAS PRINCIPALES DEL PROBLEMA DE VALORES FALTANTES	29
4.10.1	<i>Valores faltantes a causa de la nubosidad</i>	29
4.10.2	<i>Valores faltantes por problema de bandeo</i>	30
4.11	MODIS.....	31
4.12	GAPFILL	33
4.13	INTERPOLACIÓN ESPACIAL.....	33
4.13.1	<i>Vecino Cercano</i>	34
4.14	SERIES DE TIEMPO	34
5.	METODOLOGÍA.	35
5.1	ÁREA DE ESTUDIO.....	35
5.2	RECOLECCIÓN DE INFORMACIÓN	35
5.3	TABLA COMPARATIVA ENTRE ALGUNOS DE LOS MÉTODOS EXISTENTES Y EL MÉTODO GAPFILL	36
5.4	EJECUCIÓN Y ADAPTACIÓN DEL ALGORITMO	37
5.4.1	<i>RStudio</i>	37
5.4.2	<i>Algoritmo gapfill en RStudio</i>	37
5.4.3	<i>Estimación de valores faltantes</i>	37
5.5	VALIDACIÓN DE LOS RESULTADOS	41

5.5.1	<i>Validación Cruzada</i>	41
5.5.2	<i>Estimación del error absoluto y error relativo</i>	41
5.5.3	<i>Selección de muestra de datos y validación de los resultados</i>	42
5.6	IMPLEMENTACIÓN DEL ALGORITMO EN PARALELO	43
5.7	TABLA COMPARATIVA DE TIEMPOS DE EJECUCIÓN DEL ALGORITMO . ¡ERROR! MARCADOR NO DEFINIDO.	
5.8	ESTIMACIÓN DE VALORES FALTANTES PARA LA BASE DE DATOS FINAL	44
6.	RESULTADOS	45
6.1	TABLA COMPARATIVA ENTRE ALGUNOS ALGORITMOS EXISTENTES Y EL MÉTODO GAPFILL.....	45
6.2	ADAPTACIÓN Y EJECUCIÓN DEL ALGORITMO PARA UNA MUESTRA DE DATOS	47
6.3	VALIDACIÓN DEL ALGORITMO	50
6.4	IMPLEMENTACIÓN DEL ALGORITMO EN PARALELO Y COMPARATIVA DE TIEMPOS DE PROCESAMIENTO .	60
7.	CONCLUSIONES	62
8.	LITERATURA CITADA	67
9.	ANEXOS	72
9.1	ANEXO 1.....	72
9.2	ANEXO 2.....	73
9.3	ANEXO 3.....	74
9.4	ANEXO 4.....	75
9.5	ANEXO 5.....	76
9.5.1	<i>Código de la herramienta en Matlab</i>	80

ÍNDICE DE TABLAS

Tabla 4-1	Aplicación de las imágenes en diferentes longitudes de onda.....	23
Tabla 4-2	Clasificación y uso principal de cada banda del MODIS con su respectivo ancho de banda.....	31
Tabla 5-1	Argumentos de la función gapfill en RStudio.....	39

Tabla 6-1 Comparativa entre ventajas y desventajas de cada método 45

Tabla 6-2 Tabla comparativa de tiempos de procesamiento en paralelo CPU..... 61

ÍNDICE DE FIGURAS

Figura 4-1 Elementos de la etapa de adquisición de información por teledetección (Fuente: Lima, 2011). 19

Figura 4-2 Esquema del funcionamiento de un sensor pasivo (Fuente: Sanderson, 1991). 20

Figura 4-3 Esquema del funcionamiento de un sensor activo (Fuente: Sanderson, 1991). 20

Figura 4-4 Comportamiento de la energía electromagnética en un objeto (Fuente: Fontal, 2005). 21

Figura 4-5 Diagrama representativo con algunos rangos del espectro electromagnético (Fuente: Fontal, 2005)..... 22

Figura 4-6 Esquema fundamental de la estructura de una Imagen Satelital (Fuente: Canada, 1976). 24

Figura 4-7 Modelo digital de terreno (Fuente: Felicísimo, 2007). 25

Figura 4-8 Mapa de clasificación y uso de suelo Michoacán 2011 (Fuente: Ruiz *et al.*, 2013). 26

Figura 4-9 Ejemplo de una imagen satelital con presencia de nubosidad (Fuente: Conabio MODIS, 2017). 30

Figura 4-10 Ejemplo de la presencia de bandeo en una imagen satelital (Fuente: Chen, Zhu, Vogelmann, Gao, & Jin, 2011). 30

Figura 4-11 Ejemplo de una imagen Satelital-Golfo de México (Fuente: Mapadesatelite, 2017).....	33
Figura 5-1 Mapa correspondiente al área de estudio (Fuente: Elaboración propia).	35
Figura 5-2 Ventana de búsqueda y descarga de información (Fuente: Earth Explorer,2018).....	36
Figura 5-3 Ráster con valores de NDVI provenientes del MODIS para la República Mexicana en RStudio (Fuente: Elaboración propia).	38
Figura 5-4 Arreglo de 11x11 con el valor faltante al centro de la cruz blanca (Fuente: Gerber <i>et al.</i> , 2017).	40
Figura 5-5 Diagrama de la metodología (Fuente: Elaboración propia).....	44
Figura 6-1 Arreglo de datos utilizado para la validación (Fuente: Elaboración propia).	48
Figura 6-2 Valores de NDVI diarios para cada uno de los años de registros. (Fuente: Elaboración propia).	49
Figura 6-3 Extracción de valores en diferentes zonas de la misma imagen (Fuente: Elaboración propia).	50
Figura 6-4 Intervalos correspondientes al error relativo para los valores resultantes para distintas estimaciones en función de la ubicación de los valores faltantes (Fuente: Elaboración propia).....	51
Figura 6-5 Dispersión de los datos para la estimación con 40% de valores faltantes (Fuente: Elaboración propia).....	51

Figura 6-6 Distribución de frecuencias de los valores estimados para el caso con 40% de valores faltantes (Fuente: Elaboración propia).....	52
Figura 6-7 Dispersión de los datos para la estimación con 30% de valores faltantes (Fuente: Elaboración propia).....	52
Figura 6-8 Distribución de frecuencias de los valores estimados para el caso con 30% de valores faltantes (Fuente: Elaboración propia).....	53
Figura 6-9 Dispersión de los datos para la estimación con 20% de valores faltantes (Fuente: Elaboración propia).....	53
Figura 6-10 Distribución de frecuencias de los valores estimados para el caso con 20% de valores faltantes (Fuente: Elaboración propia).....	54
Figura 6-11 Dispersión de los datos para la estimación con 10% de valores faltantes (Fuente: Elaboración propia).....	54
Figura 6-12 Distribución de frecuencias de los valores estimados para el caso con 10% de valores faltantes (Fuente: Elaboración propia).....	55
Figura 6-13 Distribución de frecuencias de los valores originales registrados por el sensor (Fuente: Elaboración propia).	55
Figura 6-14 Intervalos correspondientes al error relativo para los valores estimados tras la ejecución del algoritmo para distintos porcentajes de valores faltantes (Fuente: Elaboración propia).....	56
Figura 6-15 Valores de NDVI sin previo tratamiento para el estado de Querétaro (Fuente: Elaboración propia).....	57
Figura 6-16 Valores estimados de NDVI aplicando el algoritmo gapfill (Fuente: Elaboración propia).	58

Figura 6-17 Imágenes de radar correspondientes al día 1 del año 1985 y 1993...	58
Figura 6-18 Valores de precipitación resultantes para estimar valores faltantes en la escena correspondiente al día 1 de 1985.	59
Figura 6-19 Curva de desempeño del algoritmo.	61
Figura 9-1 Ventana para visualización de imágenes de radar (Fuente: Elaboración propia).	77
Figura 9-2 Campos de Precipitación intensa (Fuente: Elaboración propia).	78
Figura 9-3 Mapa en formato raster de frecuencias (Fuente: Elaboración propia).	78
Figura 9-4 Raster con valores de precipitación (Fuente: Elaboración propia).	79
Figura 9-5 Extracción de información dentro del área de un archivo shape (Fuente: Elaboración propia).	79

1. INTRODUCCIÓN.

La producción de vegetación es de vital importancia para el ser humano, no solo por ser una de las principales fuentes de obtención de materias primas como lo son fibras y madera, sino también porque el entorno en el que ésta se desarrolla tarde o temprano y debido al crecimiento poblacional, esta se convertirá en zonas que estaremos habitando los seres humanos mientras no se le dé la importancia debida a la preservación y monitoreo de las áreas naturales.

Esta situación ha captado la atención de recientes investigaciones por su relación directa con el ciclo del carbono en donde la concentración de CO₂ atmosférico se ha visto aumentada en un 71% desde 1750 a causa de nuestras actividades resultando en alteraciones en el balance energético de la superficie terrestre así como en lo que conocemos como cambio climático (IPCC, 1995).

En ese contexto, el uso de sensores remotos, que obtienen información a través de imágenes satelitales, representa un método de bajo costo que nos permite obtener mediciones a lo largo de años consecutivos con la capacidad de monitorear grandes extensiones de la superficie así como alcanzar zonas remotas, inaccesibles o en conflicto pudiendo abarcar países enteros (Wulder, 1998).

Uno de los principales usos de la información obtenida a través de estos sensores es el de ofrecer una perspectiva única de la superficie de la tierra, sus recursos y el impacto que tienen las actividades humanas sobre esta.

Esta información nos permite la estimación de parámetros de superficie como lo son el Índice de Área Foliar (LAI) ampliamente usado en el campo de la agricultura y ecología, el cual nos permite conocer la superficie del suelo que es ocupada por la vegetación.

Podemos mencionar un segundo parámetro como lo es la Fracción de Radiación Fotosintéticamente Absorbida (FPAR) que nos permite caracterizar la capacidad de absorción de energía de la vegetación cuyo rol es crítico en el ciclo del carbono y el albedo superficial, el cual nos da una idea de cómo se regula el intercambio de

energía entre la superficie terrestre y la atmosfera, factor crítico cuando se trata de representar o alimentar modelos de cambio climático.

Finalmente, un tercer ejemplo es el Índice de Vegetación de Diferencia Normalizado (NDVI, por sus siglas en inglés) el cual nos arroja como resultado valores estimados que nos permiten caracterizar el verdor y la condición de la vegetación.

Se parte de la premisa de que el NDVI es un indicador de la salud vegetal teniendo en cuenta que la degradación o disminución de la vegetación se vería reflejada en valores reducidos de NDVI, es por ello que mediante la determinación de la relación entre la magnitud de este índice de vegetación en varios ecosistemas forestales es posible hacer el seguimiento temporal de los procesos de degradación (FAO, 2011).

El campo de estudio de la teledetección ha crecido rápidamente en los últimos 15 años debido en a la necesidad de mantener en constante monitoreo las alteraciones de la superficie terrestre tanto en su uso de suelo como el desarrollo y la evolución de la vegetación (Turner, 2005).

A través de técnicas de teledetección, el registro y análisis de alteraciones en el comportamiento espectral de la cobertura vegetal, nos es posible caracterizar y evaluar el estado general de la vegetación teniendo la posibilidad de detectar plagas y enfermedades o posibles carencias nutricionales (Sá, Antonio, & Almorox, 1999).

Los estudios realizados acerca de los cambios progresivos en la cubierta vegetal a menudo emplean indicadores que permiten su caracterización y cuantificación a través del tiempo, traduciendo los registros históricos en series de tiempo (*Linke et al.*, 2009).

En los últimos años se han desarrollado una gran variedad de algoritmos computacionales empleados para la estimación de productos de vegetación a través de información obtenida de imágenes de satélite mediante regresiones de datos así como métodos físicos o combinaciones de los anteriores que nos permitan hacer más eficiente esta tarea, como ejemplo podemos mencionar los clúster y el manejo y procesamiento de información en la nube (*Verrelst et al.*, 2015).

Si bien es cierto que los procesos en la teledetección se han visto mejorados, aun se muestran limitados cuando se trata de analizar grandes cantidades de datos, esto debido a que año con año se agregan nuevas mediciones a los registros históricos, es por eso que utilizamos técnicas como lo es la segmentación y el procesamiento parcial de las imágenes hasta lograr analizar la imagen en su totalidad (Li, Chen, & Liu, 2010).

Además del aumento en el tamaño de las bases de datos día con día otro gran problema es el que se presenta cuando debido a la presencia de nubes en la atmosfera las mediciones arrojan valores de reflectancia muy pequeños, o presentan problemas de bandeo comúnmente ocasionados por una mala calibración de los sensores.

Un problema especial es el de valores faltantes, que se presenta cuando trabajamos en el análisis de la vegetación en la superficie empleando imágenes provenientes del espectrorradiómetro de imágenes de media resolución (MODIS) y estos datos faltantes o píxeles dañados presentan perturbaciones atmosféricas anteriormente mencionadas.

Son varias las aproximaciones desarrolladas para tratar de dar solución a este problema y es posible dividir en dos categorías los métodos utilizados, el primero a través de un análisis estadístico robusto de los datos y un segundo que nos permite estimar faltantes en función de los valores observados con métodos geoestadísticos fuera de los convencionales.

El objetivo del método gapfill es el de estimar los valores faltantes utilizando los valores medidos, este método pertenece a la segunda clasificación.

Este método se adapta a los datos con una estructura típica de matriz de cuatro dimensiones, las primeras dos dimensiones utilizadas para describir la ubicación espacial de los valores a través de longitud y latitud. Siendo las últimas dos elementos para describir su ubicación temporal a través de un índice correspondiente al día y el año de la medición (Gerber, Furrer, Schaepman-strub, Jong, & Schaepman, 2017).

Con ayuda de nuevas técnicas de procesamiento como ésta, podemos tener mejores aproximaciones de las condiciones ecológicas y sustentabilidad del medio que estamos analizando, si a esto le sumamos una calibración con mediciones realizadas in-situ y bases de datos con un mayor número de registros, nos es posible construir mejores modelos matemáticos y empíricos que nos permitan describir el comportamiento del clima sobre la superficie terrestre (Unninayar & Olsen, 2008).

2. HIPOTESIS.

La calidad de la información proveniente de sensores remotos como la plataforma MODIS se verá mejorada con la implementación de un algoritmo para la estimación de valores faltantes por efecto de la nubosidad y problemas de calibración de los sensores, este aumento en la calidad de la información nos permitirá tener un mejor entendimiento de los fenómenos de evapotranspiración, el balance hídrico y cambio climático.

3. OBJETIVOS.

3.1 Objetivo General:

Mejorar la calidad de los datos provenientes de imágenes adquiridas a través de sensores remotos, implementando un algoritmo computacional para la estimación de valores de píxel faltantes.

3.2 Objetivos Particulares:

- Mejorar la calidad de los datos en imágenes provenientes de sensores remotos implementando el algoritmo gapfill para la estimación de valores de píxel faltantes.
- Realizar una comparativa sobre las ventajas y desventajas entre algunos de los métodos existentes para la estimación de datos faltantes.
- Realizar un análisis de los tiempos de procesamiento utilizando el algoritmo gapfill para diferentes casos de valores faltantes.

- Realizar un análisis de sensibilidad del algoritmo para su empleo en diferentes secciones de la imagen, así como en imágenes con distintos porcentajes de valores faltantes
- Aplicar el algoritmo en nuestra base de datos para la estimación de valores faltantes en productos de vegetación para la zona del Altiplano de México.

4. MARCO DE REFERENCIA.

En este capítulo se hace una revisión de los conceptos y principios fundamentales para el entendimiento del trabajo presentado en esta investigación teniendo en cuenta que la teledetección y los nuevos métodos para estimación de valores faltantes continúan mejorando día a día.

4.1 Teledetección.

La percepción remota o teledetección puede definirse como la ciencia que se encarga de la obtención de información de un objeto analizando los datos adquiridos mediante algún dispositivo que no se encuentra en contacto físico con dicho objeto (Martinez Muñoz & Díaz Ponce, 2005).

La energía que es emanada por los cuerpos se registra con un valor de reflectancia por medio de sensores, ésta es almacenada asignando un valor a cada uno de los pixeles hasta tener el registro de toda el área de medición, posteriormente esta información es transmitida a la tierra en donde es procesada y convertida en una escala de colores o matices de gris para crear una imagen satelital (Oprisescu & Dumitrescu, 2005).

Esta ciencia nos provee una perspectiva única sobre las observaciones ya sea la parte de la biosfera que ocupa la superficie terrestre, las zonas costeras o aquellas áreas cubiertas por nieve como las montañas y los polos así también nos provee una vista global detallada de los cambios naturales y antropogénicos que suceden en la superficie de la tierra.

En un proceso de teledetección Figura 4-1 es posible identificar 7 elementos principales: La fuente de emisión de energía, interacción de la energía con la superficie, registro por el sensor, transmisión, recepción y procesamiento, interpretación y análisis de la información hasta llegar finalmente a su aplicación.

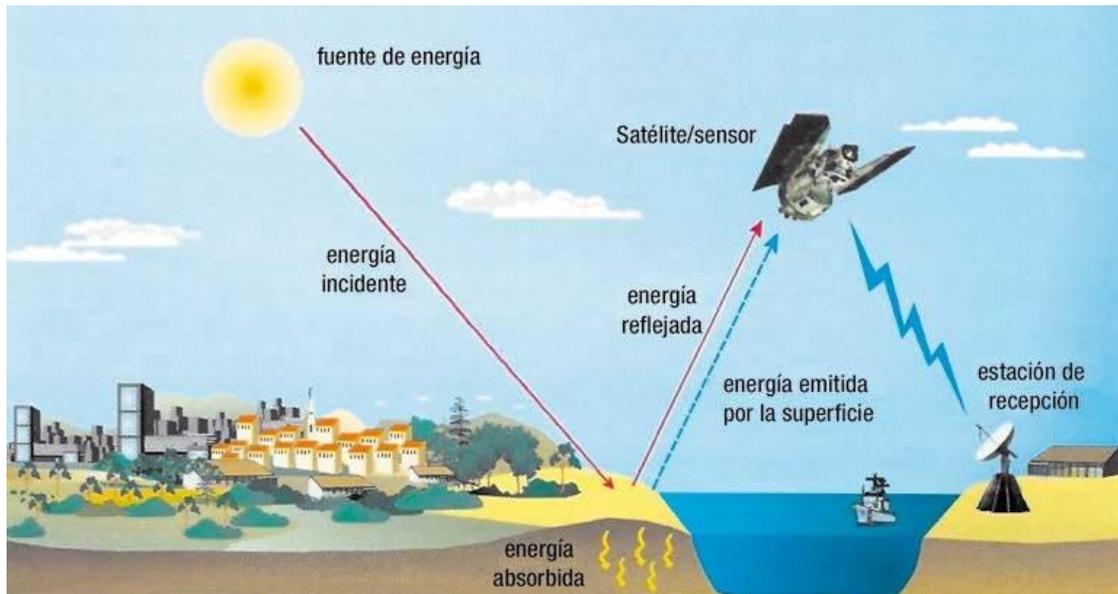


Figura 4-1 Elementos de la etapa de adquisición de información por teledetección (Fuente: Lima, 2011).

4.2 Sensores Remotos

Se trata de aquellos dispositivos que se encargan de la obtención de información de un objeto a distancia, estos nos permiten cualificar y cuantificar el flujo de energía que proviene de la superficie de la tierra y los objetos sobre está.

Estos dispositivos cuentan con miles de sensores diminutos que miden la cantidad de radiación electromagnética que es reflejada y se encuentran montados en satélites y más recientemente en drones.

La obtención de información se realiza a través de la medición de los 4 fenómenos que ocurren sobre la superficie, estos son: emisión, absorción, reflexión y transmisión cada uno en un rango característico dentro del espectro electromagnético (Sá *et al.*, 1999).

4.2.1 Sensores Pasivos

Son aquellos sensores que depende de una fuente externa de energía electromagnética, usualmente el sol, para la elaboración de una imagen a partir de las mediciones espectrales, estos miden la energía que es reflejada por la superficie u objeto sobre esta y se llaman pasivos pues no disponen de una fuente de energía propia Figura 4-2.

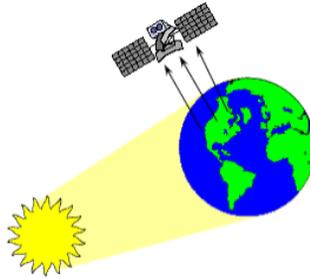


Figura 4-2 Esquema del funcionamiento de un sensor pasivo (Fuente: Sanderson, 1991).

4.2.2 Sensores Activos

A diferencia de los sensores pasivos, este tipo de sensores cuenta con su propia fuente de emisión de energía electromagnética por lo que, para la recopilación de información, este tipo de sensores mide la cantidad de energía reflejada que fue emitida por el mismo sensor Figura 4.3.

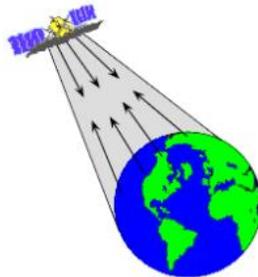


Figura 4-3 Esquema del funcionamiento de un sensor activo (Fuente: Sanderson, 1991).

4.3 Reflectancia

La radiación electromagnética al tener contacto con el blanco puede ser absorbida, reflejada y transmitida Figura 4-4. En el caso específico de la vegetación la interacción con la radiación es diferente para cada longitud de onda por lo que la

absorción de energía solar dependerá principalmente de la concentración de clorofila.

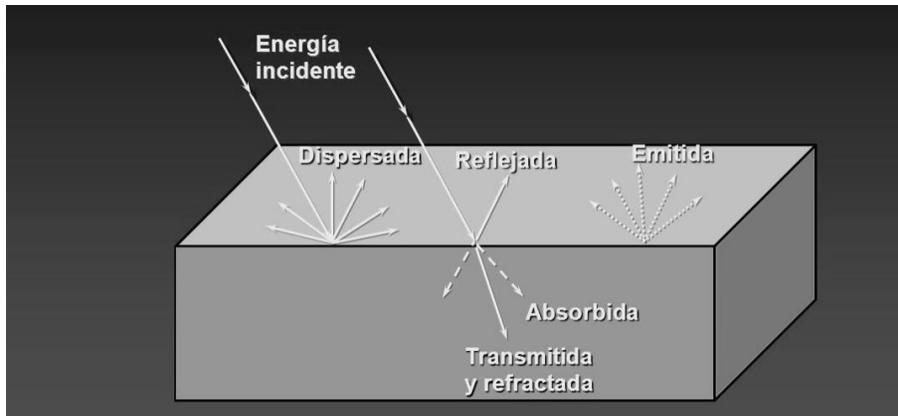


Figura 4-4 Comportamiento de la energía electromagnética en un objeto (Fuente: Fontal, 2005).

El análisis del comportamiento espectral de las superficies en los distintos niveles del espectro electromagnético (bandas) lo que nos permite a través de algunas operaciones algebraicas, estimar algunos índices de variables sintéticas asociadas a procesos biofísicos (Paruelo, 2008).

La densidad, el contenido de agua, la composición química y otros factores y características no visibles de un objeto específico de la superficie, influyen globalmente en cómo interactúa la energía con dicho objeto en diversas longitudes de onda del espectro y en cómo se refleja en él. El sensor digital mide esta interacción espectral, que a su vez proporciona la comprensión de tales estados y características invisibles.

Las medidas de reflectancia pueden revelarnos el contenido mineral de las rocas, la humedad del suelo, la salud de la vegetación, la composición física de los edificios y miles de otros detalles invisibles al ojo humano.

4.4 Energía Electromagnética

La energía electromagnética está compuesta por ondas de campos eléctricos y magnéticos que se propagan a través del espacio en forma de voltaje a través de

un material conductor, como ondas de radio o como ondas de luminosas a través de un intervalo casi infinito de frecuencias (Tomasi, 2003).

4.5 Espectro electromagnético

El sol nos provee directa o indirectamente de la mayor parte de la energía que requerimos en el planeta, de manera directa esta energía nos llega en forma de luz o radiación y de manera indirecta a través de fenómenos climáticos y reacciones químicas.

Esta luz y radiación mencionada anteriormente forma parte de un rango muy amplio cuyo límite inferior y superior es inexistente, este rango es denominado el espectro electromagnético Figura 4-5 y para referirse a este es común utilizar dos parámetros muy conocidos: la frecuencia y longitud de onda (Fontal, 2005).

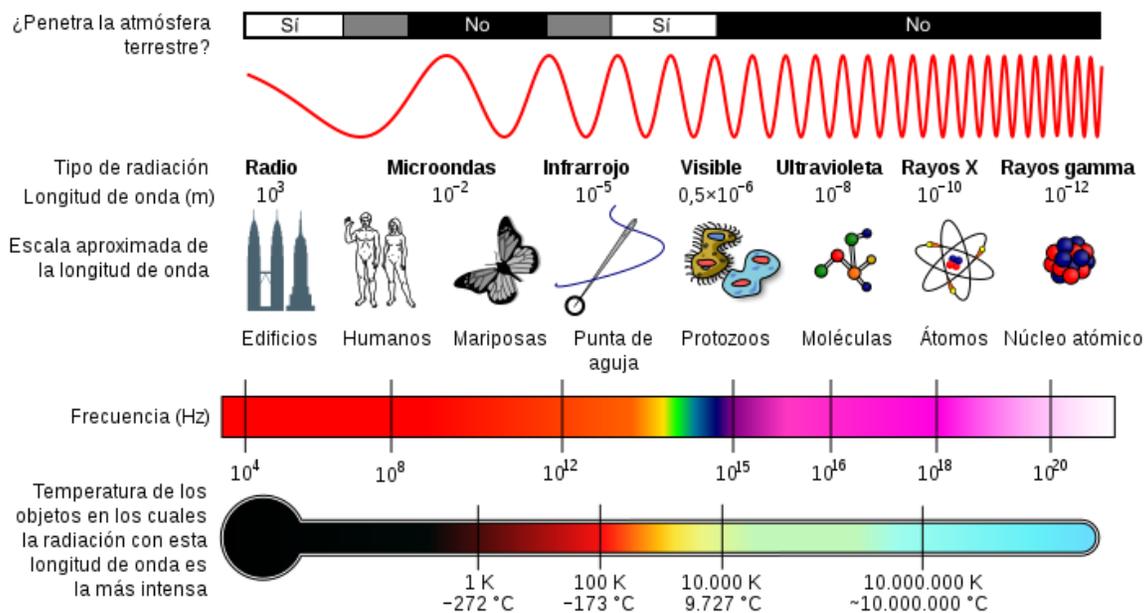


Figura 4-5 Diagrama representativo con algunos rangos del espectro electromagnético (Fuente: Fontal, 2005).

Tratándose del comportamiento espectral de la cobertura vegetal intervienen diversos factores como lo son la fisiología, morfología y fenología de la planta, así

como factores ambientales y estacionales, siendo las hojas el elemento más importante para definir este comportamiento espectral (Sá *et al.*, 1999).

En la Tabla 4-1 se muestran algunas de las aplicaciones para imágenes en función de la longitud de onda.

Tabla 4-1 Aplicación de las imágenes en diferentes longitudes de onda.

Longitud de Onda	Aplicación
Azul Visible	Cartografía en aguas someras y clasificación de Suelo y Vegetación
Verde Visible	Características de la Vegetación
Rojo Visible	Clasificación de la Vegetación
Infrarrojo Cercano	Cartografía y clasificación de Vegetación
Infrarrojo Medio	Clasificación de suelo y estructuras geológicas

Dependiendo de la sensibilidad para la que fue concebido el sensor, la mayoría de los satélites de teledetección miden la energía que es reflejada por la superficie terrestre en intervalos de longitudes de onda del espectro electromagnético muy específicas llamados bandas.

4.6 Imágenes Satelitales

Una imagen satelital es una matriz digital de elementos llamados pixeles a los cuales se les asigna un valor numérico en función de la cantidad de luz que el objeto en cuestión refleja al sensor (fenómeno mejor conocido como reflectancia), una vez asignados estos valores, al transmitir la información a tierra estos serán convertidos a colores o matices de gris para crear una imagen parecida a una fotografía.

Cada uno de estos pixeles tendrá un valor para cada banda espectral así como información espacial, la intensidad del color de la imagen nos indicará la magnitud de la energía que fue reflejada a través del brillo de cada pixel y la interpretación de

cada una de las imágenes dependerá del usuario podrá realizar desde una simple inspección visual hasta la implementación de programas de tratamiento de imágenes y algoritmos de procesamiento para la obtención de información para la utilización de modelos complejos.

La representación de una imagen digital en la mayoría de los programas destinados para el procesamiento de estas se basará en su manipulación y tratamiento a partir de un ráster mediante el cual le será posible representar la imagen por medio de una matriz con celdas llamadas pixeles Figura 4-6 con un valor de reflectancia asignado de manera individual.

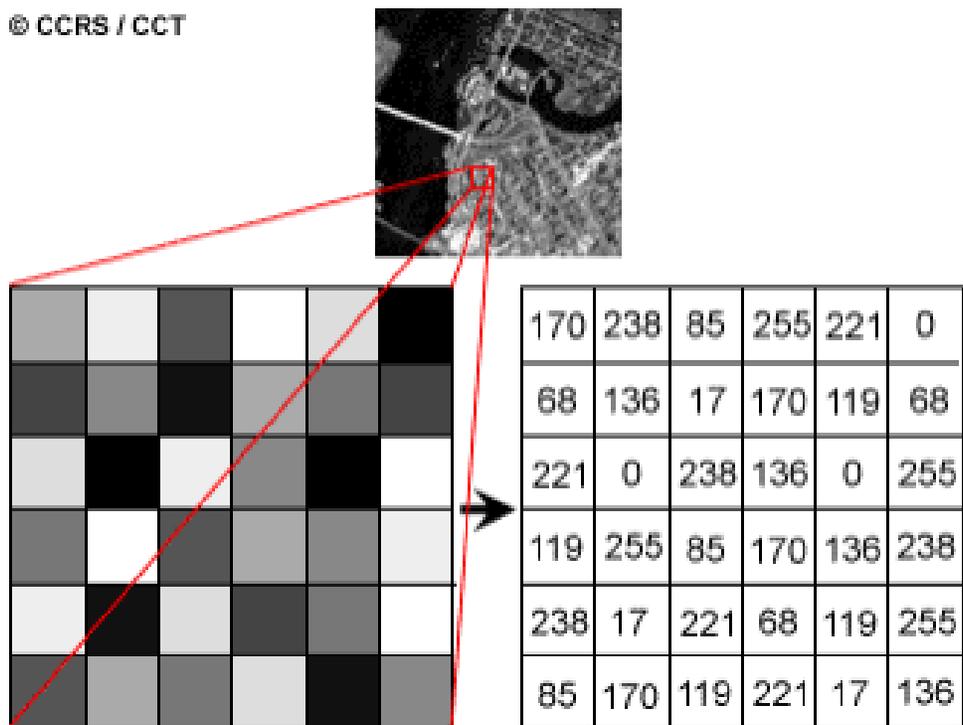


Figura 4-6 Esquema fundamental de la estructura de una Imagen Satelital (Fuente: Canada, 1976).

Este formato se trata de una estructura de datos simple que está conformada por una matriz de celdas o pixeles organizados en filas y columnas en la que cada uno de estos elementos contiene datos almacenados.

Una de sus principales ventajas además de su simplicidad, es el que las operaciones de superposición para la estimación de productos de vegetación se

realizan de forma más rápida y eficiente, además de la posibilidad de realizar un re-escalado evitando la pérdida de resolución.

4.7 Ejemplos de aplicaciones de imágenes satelitales

4.7.1 Modelo Digital de Elevaciones

Llamados también modelos digitales del terreno que contienen conjuntos de datos con medidas de elevación del terreno Figura 4-7, obtenidos tras aplicar procedimientos fotogramétricos y cuyo uso principal es el de crear modelos tridimensionales con la topografía de la superficie.

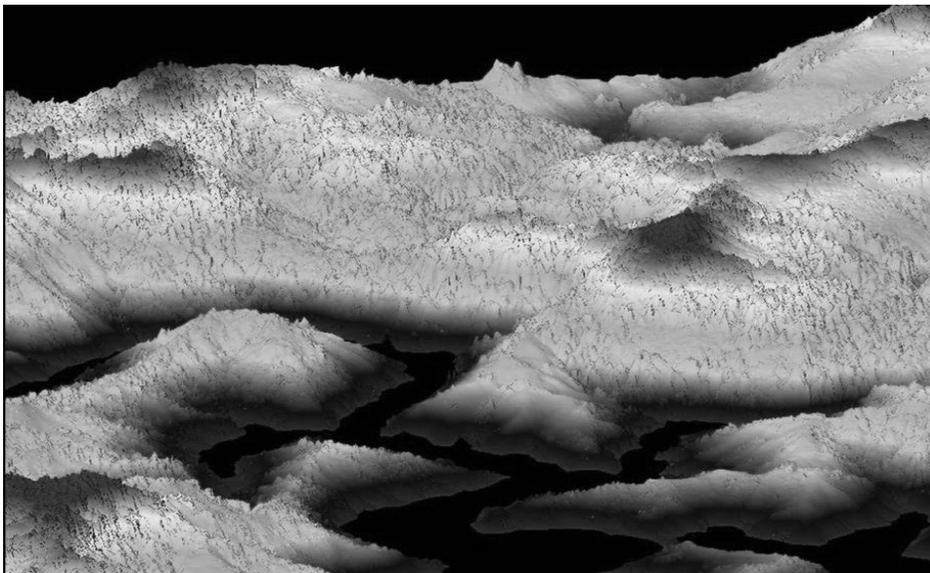


Figura 4-7 Modelo digital de terreno (Fuente: Felicísimo, 2007).

4.7.2 Mapas de clasificación

Probablemente la aplicación más común creada a partir de imágenes de satélite, cuyo principio se basa en realizar mapas de clasificación a través del análisis y comparación de la firma espectral de cada píxel, agrupando en distintas clases cada uno de los rangos identificados Figura 4-8, los cuales se utilizan para delimitar y monitorear zonas de interés (Ruiz, Savé, & Herrera, 2013).

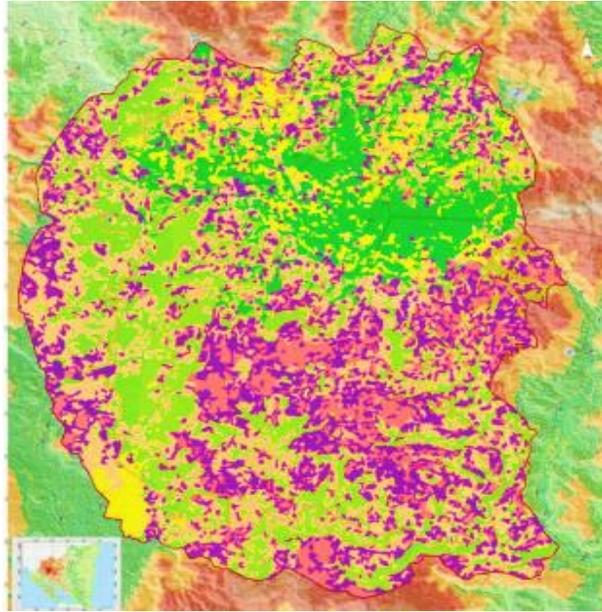


Figura 4-8 Mapa de clasificación y uso de suelo Michoacán 2011 (Fuente: Ruiz *et al.*, 2013).

A través de estos mapas nos es posible monitorear el crecimiento de la mancha urbana a través del tiempo, así como los cambios de uso de suelo y cobertura vegetal en la superficie.

4.7.3 Evapotranspiración

La estimación de la evapotranspiración mediante el uso de sensores remotos se basa en el balance energético de la superficie terrestre (Paruelo, 2008) y se estima mediante la ecuación (4.1).

$$LET = Rn + H + G + S + P \quad (4.1)$$

donde Rn es la radiación neta (W/m^2), LET es el flujo de calor latente resultado del producto del calor latente por vaporización L y la evapotranspiración ET (mm), H es el flujo de calor sensible (W/m^2), G es el flujo calor en el suelo, S la energía acumulada en la cobertura vegetal y P la energía utilizada en el proceso de fotosíntesis.

4.8 Características de los sensores remotos

Se trata de aquellas características propias de cada sensor por medio de las cuales nos será posible decidir los propósitos de cada uno de ellos, habrá ocasiones en las

que se necesitará de una mayor resolución espacial que nos permita identificar zonas urbanas, otras en que se requiera de una mayor resolución espectral que nos permita identificar detalles más pequeños sobre el terreno o simplemente deberemos utilizar imágenes con una menor resolución espacial que nos permite tener mejor perspectiva de una gran extensión de superficie sin que esto implique mayores tiempos de procesamiento de la información.

4.8.1 Resolución Espacial

La resolución espacial está dada por el tamaño de cada píxel de la imagen proyectada sobre la superficie terrestre y se refiere al tamaño más pequeño en el que es posible realizar mediciones.

En la mayoría de los casos esta característica es la responsable de la variación en el tamaño de los archivos digitales pues al aumentar la resolución espacial de una imagen, de igual forma estaremos aumentando el tamaño de esta pudiendo pasar de unos cuantos megabytes a una decena de gigabytes.

4.8.2 Resolución Temporal

Para definir esta característica es necesario entrar en contexto, supóngase un satélite que se encuentra orbitando la tierra mientras que a su vez la tierra se encuentra girando sobre su propio eje, durante este movimiento transcurrirá un tiempo hasta que el satélite vuelva a pasar sobre un mismo punto, este tiempo que tarda el satélite en volver a pasar sobre el mismo punto se denomina resolución temporal.

4.8.3 Resolución Espectral

La resolución espectral de un sistema de teledetección se refiere al número de bandas espectrales que el sensor es capaz de detectar, comúnmente un satélite cuenta con 4 bandas sin embargo existen algunos que pueden llegar a las 10 o 15 bandas.

4.8.4 Tamaño de Escena

Se refiere al tamaño total o área que abarca la imagen tomando en cuenta que una malla compuesta por filas y columnas es la que conforma la imagen proyectada sobre la superficie de la tierra, lo que conocemos como una escena.

4.9 Productos de Imágenes

Una de las principales ventajas cuando utilizamos imágenes satelitales es la de tener la posibilidad de analizar, procesar, manipular y realzar ciertas características de estas a través una gran variedad de procedimientos entre ellos algunas operaciones algebraicas para la obtención de diversos productos con usos específicos.

4.9.1 NDVI

El Índice de Vegetación Normalizado o NDVI es el índice más usado como estimador de la presencia y condición de la vegetación, fue Compton Tucker del Goddard Space Fligth Center de la NASA el primero en instaurar este índice como un descriptor de la vegetación (Tucker, Maroulis, & Bandy, 1985).

Este índice integra dos aspectos del comportamiento vegetal, valores bajos de niveles de reflectancia a causa de la absorción de la clorofila en el Rojo cercano y la alta reflectancia de la planta y sus hojas en el Infra rojo, estimando el índice a través de la ecuación (4.2).

$$NDVI = \frac{IR-R}{IR+R} \quad (4.2)$$

Donde IR corresponde a la reflectancia en la porción del infrarroja del espectro mientras que R corresponde a la porción correspondiente al rojo.

4.9.2 EVI

El Índice de Vegetación Mejorado o EVI es similar al NDVI, pero incorpora otras bandas espectrales con el fin de optimizar la señal que emite la vegetación y de esta manera resaltar la firma espectral del sustrato y vegetación restándole influencia al efecto de la atmósfera, esta modificación se utiliza particularmente para niveles altos de biomasa.

$$EVI = 2.5 \frac{IR-R}{IR+C1*R-C2*A+L} \quad (4.3)$$

Donde IR y R son las bandas correspondientes al infrarrojo y rojo, C1 y C2 coeficientes por la presencia de aerosoles y L un ajuste en función del sustrato presente en la escena.

4.9.3 LAI

El índice de área foliar o LAI por sus siglas en inglés, es un parámetro ampliamente usado en el campo de la agricultura y ecología, mediante el cual nos podemos conocer la superficie del suelo que es ocupada la vegetación.

4.9.4 Albedo

Es el porcentaje de radiación que cualquier superficie refleja respecto a la radiación incidente. En superficies claras tendrá un valor superior al de superficies oscuras, es un índice adimensional con una escala de cero a uno y a nivel global cambiará en función de la estación del año. El albedo es uno de los atributos más importantes en la caracterización del balance de energía en la superficie debido a que cambios en su magnitud tienen consecuencias en procesos climáticos y biofísicos (Dickinson, 1995).

Se trata también de un componente necesario para la estimación de la evapotranspiración a partir de modelos de balance de energía cuya variable principal es el flujo de radiación neta (Dickinson, 1995).

4.10 Causas principales del problema de valores faltantes

4.10.1 Valores faltantes a causa de la nubosidad

Tal y como se mencionó anteriormente en función de las condiciones atmosféricas, nuestras imágenes tendrán valores distorsionados por el efecto de la nubosidad lo que nos dará como resultado valores de reflectancia que no corresponden a los de la superficie del terreno, como se muestra en la Figura 4-9.



Figura 4-9 Ejemplo de una imagen satelital con presencia de nubosidad (Fuente: Conabio MODIS, 2017).

4.10.2 Valores faltantes por problema de bandeado

Un segundo problema al que nos enfrentaremos será el del bandeado Figura 4-10, en una imagen el cual puede ser ocasionado por una falta de calibración en el sensor o debido a la forma en que el satélite realiza las mediciones al orbitar la tierra, tratándose de satélites de órbita polar.



Figura 4-10 Ejemplo de la presencia de bandeado en una imagen satelital (Fuente: Chen, Zhu, Vogelmann, Gao, & Jin, 2011).

4.11 MODIS

MODIS o “Moderate Resolution Imaging Spectroradiometer” es un instrumento dentro de los satélites Terra (EOS AM-1) y AQUA (EOS PM-1). El satélite Terra órbita alrededor de la tierra sincronizado para pasar de norte a sur a través del ecuador por la mañana mientras que AQUA pasa de sur a norte sobre el ecuador por la tarde.

El instrumento MODIS fue diseñado para proveer información diaria a escala global en 36 bandas espectrales, la selección de bandas espectrales y sus respectivos intervalos presentados en la Tabla 4-2 se realizó cuidadosamente para minimizar el impacto de los gases en la atmosfera especialmente el vapor de agua principal limitante en instrumentos lanzados con anterioridad (Vermote, El Saleous, & Justice, 2002).

Tabla 4-2 Clasificación y uso principal de cada banda del MODIS con su respectivo ancho de banda.

Uso Principal	Banda	Intervalo espectral	Uso Principal	Banda	Intervalo espectral
Superficie/Nubes/ Aerosoles	1	620-670		19	915-965
	2	841-876	Superficie/Nubes/ Aerosoles	20	3660-3840
Superficie/Nubes/ Aerosoles	3	459-479		21	3929-3989
	4	545-565		22	3929-3989
	5	1230-1250		23	4020-4080
	6	1628-1652	Temperatura Atmosférica	24	4433-4498

	7	2105-2155		25	4482-4549
Océano/Phytoplankton	8	405-420	Vapor de agua y nubes cirrus	26	1360-1390
	9	438-448		27	6535-6895
	10	483-493		28	7157-7475
	11	526-536	Propiedades de las nubes	29	8400-8700
	12	546-556		30	9580-9880
	13	662-672	Superficie y temperatura de nubes	31	10780-11280
	14	673-683		32	1170-12270
	15	743-753	Superficie Terrestre y nubes	33	13185-13485
	16	862-877		34	13485-13785
Vapor de agua	17	890-920		35	13785-14085
	18	931-941		36	14085-14385

El objetivo de este satélite es el de mejorar nuestro entendimiento del comportamiento de la dinámica global en la superficie terrestre y océanos, así como de la parte baja de la atmósfera, que nos permita realizar predicciones más precisas del fenómeno de cambio climático lo que nos permitirá elegir las medidas más adecuadas para la protección del medio ambiente (“MODIS about,” 2017).

Este instrumento obtiene imágenes con un ancho de escena de 2,330 km de ancho Figura 4-11, en tres resoluciones espaciales, de 250, 500 y 1000 m por pixel (“MODIS data,” 2017).

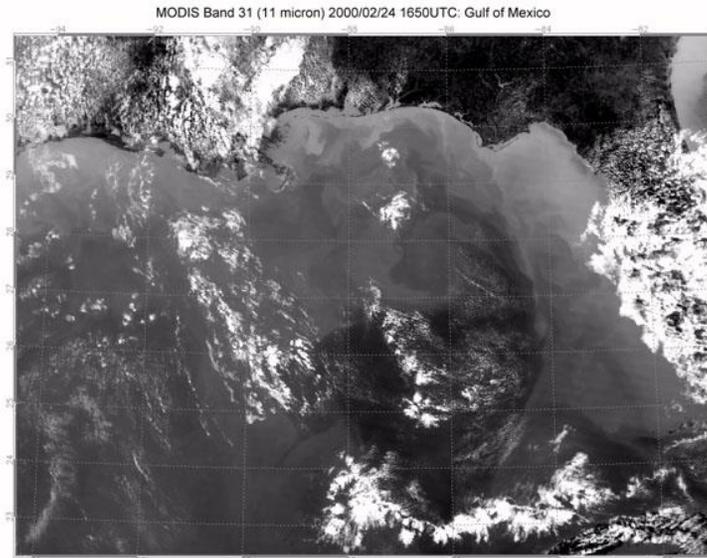


Figura 4-11 Ejemplo de una imagen Satelital-Golfo de México (Fuente: Mapadesatelite, 2017).

4.12 Gapfill

Se trata de un algoritmo de estimación de valores de píxel faltantes que puede ser aplicado en imágenes satelitales provenientes de la plataforma MODIS que cuenten con una estructura de datos de 4 dimensiones.

El proceso para la estimación de valores faltantes con este algoritmo se basa en la selección de subconjunto de datos teniendo como elemento central el valor a estimar, se selecciona un subconjunto de “n” elementos que forman una matriz cuadrada cuyo valor central es el elemento desconocido.

Una vez hecho esto se procede a la estimación de cada valor de manera independiente en función de los valores que lo rodean empleando el método de interpolación del vecino más cercano realizando una serie de ajustes a través del análisis histórico de las series de tiempo para cada píxel.

4.13 Interpolación Espacial

La interpolación espacial es un procedimiento mediante el cual podemos estimar el valor de una variable en determinada posición del espacio conociendo los valores de esa misma variable en otras posiciones (Sendra & García, 2000).

4.13.1 Vecino Cercano

Este método de interpolación asigna el valor del punto faltante en función de los valores de los elementos más cercanos a cada uno de estos elementos faltantes (nodos), el usuario puede definir los intervalos de búsqueda así como el número de elementos a tomar en cuenta para la estimación de los valores faltantes (Fallas, 2007).

En el caso de un arreglo bidimensional se trata de los elementos Y_{n-1} y Y_{n+1} ampliando esta búsqueda por criterio del usuario, en el caso de arreglos tridimensionales o en este caso de matrices, se trata de estimar el valor faltante a partir de los valores de las celdas cercanas a la posición del nodo.

4.14 Series de Tiempo

Una serie de tiempo es un conjunto ordenado de valores (registros) que corresponden a un muestro equiespaciado de una función en el tiempo (Gallego Gómez, 2008), algunos de los principales para su análisis son:

Descripción: En donde el principal objetivo es el de graficar cada una de las observaciones para obtener medidas simples descriptivas de las características principales de los datos.

Explicación: Cuando las observaciones son realizadas en dos o más variables, es posible utilizar cada una de sus características para explicar la variación una de otra.

Predicción: Dados los valores de una serie es posible que se desee predecir valores futuros en función del comportamiento pasado de la serie.

5. METODOLOGÍA.

5.1 Área de estudio.

Se trata de la región correspondiente al altiplano central de México la cual está integrada por parte del norte y centro de Jalisco, el sur de Zacatecas, el centro de San Luis Potosí; Aguascalientes; Guanajuato, el centro y sur de Querétaro e Hidalgo, el norte y centro de Michoacán, el centro de Guerrero y el Distrito Federal. Figura 5-1 , en términos generales se trata de la región delimitada al este por la Sierra Madre Oriental, al poniente por la Sierra Madre Occidental y al sur por la depresión del Balsas.



Figura 5-1 Mapa correspondiente al área de estudio (Fuente: Elaboración propia).

5.2 Recolección de Información

El principal insumo de esta investigación serán las imágenes satelitales provenientes del satélite MODIS las cuales serán descargadas para conformar nuestra base de datos para la zona de estudio una vez que se haya definido el periodo de tiempo que delimitará la extensión de nuestra base de datos, las fechas de inicio y final dependerán de la disponibilidad de los datos en la página o en su defecto de la capacidad de procesamiento del equipo utilizado.

Para descargar cada una de estas imágenes será necesario ingresar al enlace “https://earthexplorer.usgs.gov/”, una vez dentro se deberán llenar los campos correspondientes a los parámetros de búsqueda, en primer lugar, se tendrá que seleccionar la zona de interés a través del ingreso de algún archivo en formato shape, punto o región para posteriormente elegir el periodo de búsqueda de imágenes ingresando el día y mes del inicio y término del periodo Figura 5-2.

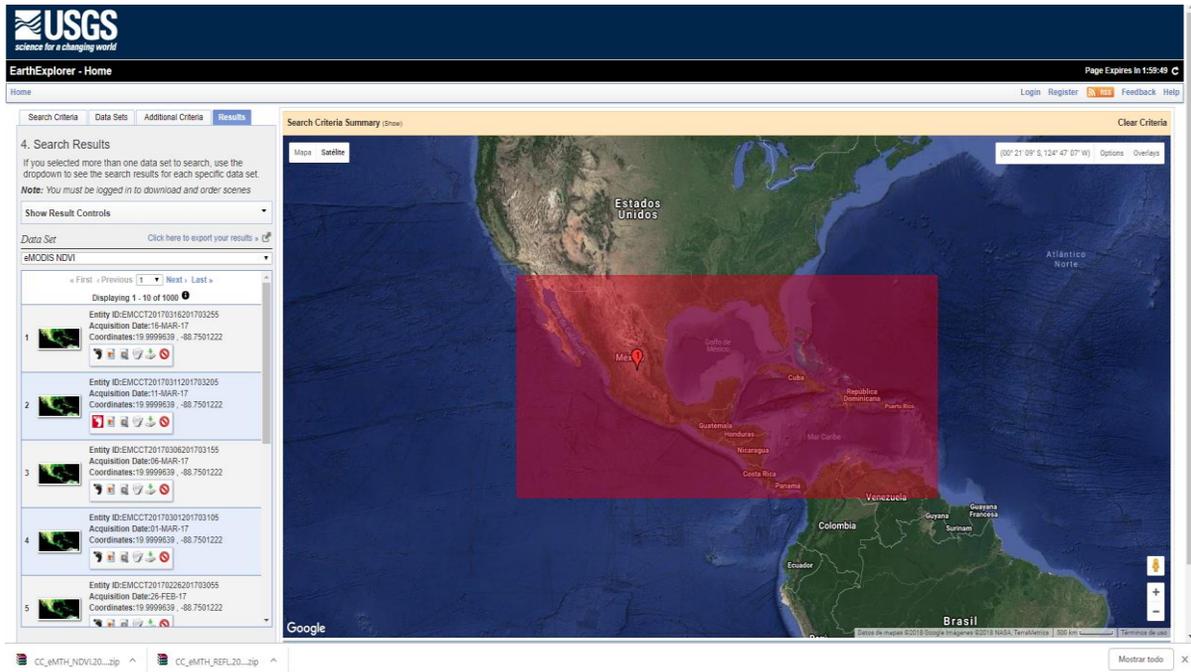


Figura 5-2 Ventana de búsqueda y descarga de información (Fuente: Earth Explorer,2018).

En este sitio es posible seleccionar entre distintas opciones para la descarga de información, se tiene la opción de descargar cada una de las bandas del MODIS de manera independiente, ciertos productos de vegetación como el LAI, FPAR y cobertura vegetal entre otros.

5.3 Tabla comparativa entre algunos de los métodos existentes y el método gapfill

Como parte del objetivo de esta investigación se realizará una tabla comparativa de las ventajas y desventajas entre algunos de los métodos existentes para estimación de valores faltantes y el procesamiento de información proveniente de sensores

remotos, esto se realizará entre otros fines para justificar la elección del método gapfill como el método principal para la estimación de valores faltantes

5.4 Ejecución y adaptación del algoritmo

De entre la amplia variedad de algoritmos desarrollados para la estimación de valores de píxel faltantes en imágenes satelitales, el método analizado será el utilizado en el algoritmo gapfill implementado en la plataforma RStudio.

5.4.1 RStudio

RStudio es un entorno de desarrollo integrado (IDE) para R que funciona con la versión estándar disponible de R, capaz de soportar los procedimientos y técnicas requeridas para realizar análisis de alta calidad y confianza. Fue diseñada para ser una herramienta con una interfaz sencilla e intuitiva tanto para los usuarios ya experimentados como para los nuevos.

R es un lenguaje de programación especialmente indicado para el análisis estadístico, inicialmente diseñado por Robert Gentleman y Ross Ihaka. El código R está disponible como software libre bajo las condiciones GNU-GPL.

5.4.2 Algoritmo gapfill en RStudio

El método gapfill fue seleccionado como el objeto de estudio de esta investigación por ser un método ya incorporado en RStudio el cual es un software libre que además de las ventajas que ofrece al ser un software de código abierto que nos permite adaptarlo a nuestras necesidades y compartir el conocimiento desarrollado con la comunidad científica, finalmente pero no menos importante se eligió por ser una plataforma que ofrece una amplia gama de herramientas estadísticas y técnicas para el procesamiento de datos entre ellas la posibilidad de la implementación del algoritmo en paralelo.

5.4.3 Estimación de valores faltantes

Antes de poder proceder a la ejecución del algoritmo fue necesaria la extracción de la información correspondiente a nuestra zona de estudio pues en el caso de algunos productos la descarga de datos se realiza para áreas que llegan a comprender todo el país Figura 5-3, tarea para la cual contamos con numerosas

opciones en función de lo que se deseaba analizar, una de ellas fue la implementación de una rutina para la selección y extracción de cierto número de filas y columnas correspondientes a nuestra zona de estudio.

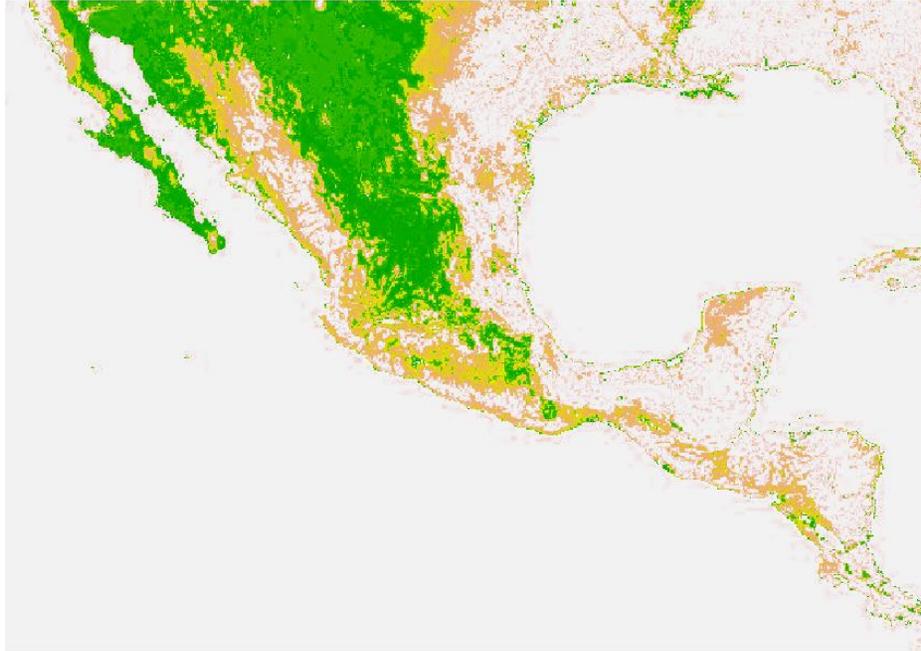


Figura 5-3 Ráster con valores de NDVI provenientes del MODIS para la República Mexicana en RStudio (Fuente: Elaboración propia).

Una segunda opción y la más sencilla, utilizada para el recorte y extracción de información para una mayor cantidad de imágenes fue la implementación del comando clip utilizando el programa IDRISI Selva, el cual nos brindó la posibilidad de recortar cada una de las imágenes de nuestra base de datos utilizando una misma mascaró con la única condición de que nuestras imágenes a recortar fuesen del mismo tamaño.

El proceso de estimación de valores faltantes con este algoritmo se basa en la selección de subconjunto de datos teniendo como elemento central el valor que se desea estimar, es decir de la matriz que conforma la imagen se selecciona un subconjunto de “n” elementos que forman una matriz cuadrada cuyo valor central es el que se desea estimar, una vez hecho esto se realiza la estimación de cada

valor faltante de manera independiente en función de los valores que lo rodean empleando el método de interpolación del vecino más cercano.

La función Gapfill contenida en RStudio requiere de una serie de argumentos de entrada mostrados en la Tabla 5-1 como lo son la matriz de datos, el método de interpolación que se utilizará, el número de iteraciones para cada estimación, entre otros.

Gapfill(data, fnSubset = Subset, fnPredict = Predict, iMax = Inf, nPredict = 1L, subset = "missing", clipRange = c(-Inf, Inf), dopar = FALSE, verbose = TRUE, ...)

Tabla 5-1 Argumentos de la función gapfill en RStudio.

Argumento	Descripción
Data	Se refiere al arreglo de 4 dimensiones mencionado anteriormente.
fnPredict	Función de interpolación seleccionada como parte del proceso de estimación, por default se encuentra seleccionada la función vecino más cercano.
iMax	Corresponde al número máximo de iteraciones para la estimación de los valores faltantes.
dopar	Nos permite ejecutar el algoritmo en paralelo modificando el uso de la capacidad de procesamiento del equipo.

El algoritmo requiere como entrada de datos un arreglo de 4 dimensiones, latitud, longitud, el intervalo de tiempo en que se está midiendo y el año correspondiente a la medición, esta información será utilizada para la estimación del valor faltante a través de un método de interpolación con las últimas dos dimensiones como parámetros condicionales de la rutina considerando que el algoritmo se detiene cuando se alcanza alguno de los siguientes dos criterios:

- La función de estimación “fnPredict” regresa un valor NA.
- El número de iteraciones máximo “iMax” se ha alcanzado.

Por la naturaleza de la operación algebraica que da origen al producto NDVI los valores que puede adquirir cada píxel rondan entre el 0 y 1, como parte de un pretratamiento de la información se aplicará un filtro a los valores anormales en la imagen identificados con un 69999 asignándoles en la categoría de NA.

A partir de la observación de la gráfica de distribución de puntos es posible realizar una primera inspección visual de los datos y así saber en cuál de los años o días según sea el caso, hay una menor cantidad de datos medidos con lo que será posible identificar si disponemos de suficientes datos para realizar una interpolación o poder excluir ese día de nuestras estimaciones.

Para la estimación de una serie de datos será necesario repetir el procedimiento descrito a continuación pues este es aplicable a cada uno de los píxeles con valores faltantes; como primera etapa el algoritmo identifica la celda con valor faltante “NA” y forma un subconjunto de los píxeles vecinos al píxel a estimar creando un arreglo cuyas dimensiones son $\lambda_1 = \lambda_2 = 5$ por lo que tendremos una matriz de 11x11 píxeles Figura 5-4.

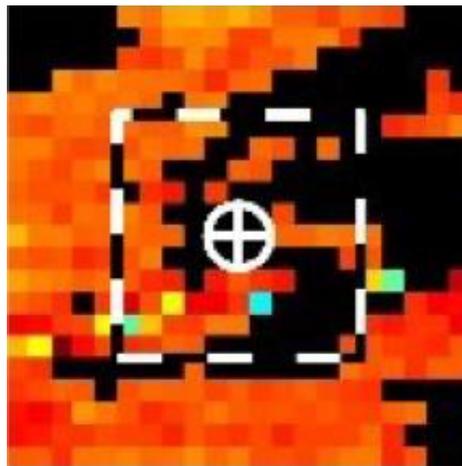


Figura 5-4 Arreglo de 11x11 con el valor faltante al centro de la cruz blanca (Fuente: Gerber *et al.*, 2017).

La tercera dimensión llamada índice temporal $\lambda_3 = 1$ nos indicará que las imágenes satelitales utilizadas para la estimación no deberán exceder un día consecutivo antes y después para poder realizar el cálculo, la cuarta dimensión $\lambda_4 = 5$ nos

permitirá limitar nuestra base de datos estableciendo la condicionante de contar con mediciones para un periodo de por lo menos 5 años consecutivos.

El algoritmo validará dos criterios para asegurar que tendrá los suficientes datos para la estimación del valor faltante:

C1) el conjunto deberá contener al menos $\theta_1=5$ imágenes que no estén vacías

C2) el subconjunto deberá tener al menos $\theta_2=25$ valores no faltantes.

En caso de que alguno de estos dos criterios no se cumpliera, las dimensiones $\lambda_1=$
 λ_2 serán incrementadas hasta satisfacerlos para posteriormente estimar el valor faltante.

5.5 Validación de los resultados

5.5.1 Validación Cruzada

Existen diversas técnicas para la validación de los resultados obtenidos a partir de simulaciones y modelos físico-teóricos, uno de ellos es el que encontramos al comparar cada uno de los valores obtenidos tras la aplicación de nuestro modelo; con el conjunto de valores conocido (Kozak & Kozak, 2003).

El método hold-out es uno de los más sencillos y conocidos de los distintos métodos de validación que existen, este separa nuestro conjunto de datos en dos subconjuntos, el primero permanece intacto y el segundo se trata de un subconjunto modificado en el cual extraeremos valores conocidos y con este subconjunto alimentaremos nuestro modelo (Kozak & Kozak, 2003).

5.5.2 Estimación del error absoluto y error relativo

El error absoluto se trata de la diferencia en valor absoluto, entre el valor estimado y el valor medido (Alfaro & Soley, 2009), tal y como aparece en la ecuación (5.1).

$$\varepsilon_a = |V \text{ estimado} - V \text{ medido}| \quad (5.1)$$

Posteriormente se realizará la estimación del error relativo el cual se define como el cociente entre el valor absoluto y el valor exacto ecuación (5.2).

$$\varepsilon_r = \frac{\varepsilon_a}{V_{\text{estimado}}} \quad (5.2)$$

5.5.3 Selección de muestra de datos y validación de los resultados

Para el proceso de validación será utilizada una pequeña muestra de datos la cual tendrá como característica principal la ausencia del fenómeno de nubosidad y sin la presencia de datos faltantes, posteriormente ciertas zonas de la imagen serán extraídas para realizar algunas pruebas y análisis que nos permitan conocer el comportamiento del algoritmo y sus resultados.

En esta base de datos de muestra serán realizadas las pruebas necesarias para la validación de los resultados y el análisis del comportamiento del algoritmo en cada una de las pruebas que serán realizadas y que se describen con detalle en los párrafos siguientes.

La validación de los valores estimados, así como la estimación de la precisión del algoritmo será cuantificada con ayuda de los métodos descritos anteriormente empleando una base de datos de muestra sin valores faltantes y sin presencia de nubosidad para posteriormente extraer valores en cierta región de la imagen para la estimación del error entre el valor medido y el valor estimado.

En una primera etapa se realizarán diferentes corridas del algoritmo para la estimación de valores faltantes en diferentes zonas de la imagen, es decir se extraerá un cierto porcentaje de valores haciendo un recorrido desde el centro a las esquinas para conocer la influencia que tiene la falta de datos en las distintas zonas.

A través de este análisis nos será posible conocer la influencia que tiene la falta de datos en ciertas zonas y tomar precauciones cuando se quieran estimar valores esto por la influencia que tiene el método de vecino más cercano en el algoritmo.

Como etapa posterior se realizará un segundo análisis, pero esta vez variando la cantidad de imágenes que conforma nuestra base de datos con lo que podremos tener una mejor perspectiva del efecto que tiene la extensión de la base de datos, es decir la cantidad de valores que integra la serie de tiempo de cada píxel cuando se realiza la estimación de valores faltantes.

Con ayuda de los resultados anteriores podremos conocer la precisión del algoritmo en función de la posición en la que se encuentre el valor faltante, así como de la influencia que tiene la cantidad de escenas disponibles para el análisis recordando que la estimación se realiza en función de los valores provenientes de la serie de tiempo para cada píxel, así de su ubicación espacial y valores de los píxeles más cercanos por medio del método del vecino más cercano.

5.6 Implementación del algoritmo en paralelo

Es importante recordar que el verdadero potencial de esta función reside en su posible implementación en paralelo con el fin de reducir los tiempos de ejecución del algoritmo.

Se realizaron las modificaciones necesarias para la ejecución de la rutina computacional utilizando en su totalidad la capacidad de procesamiento de nuestro equipo y sus unidades de procesamiento, esta capacidad puede ser medida en función de los núcleos disponibles y utilizados durante el procedimiento pues el mismo RStudio nos da la posibilidad identificar y seleccionar el número de núcleos con el que se desea ejecutar el algoritmo.

La biblioteca para la realización de esta tarea tiene por nombre `doParallel` creada por Rich Gallaway, ésta nos permite la ejecución de algoritmos en paralelo utilizando la función `registerDoParallel(cores=n)` donde `n` es el número de núcleos con el cual se realizará la corrida.

5.7 Optimización de tiempos de ejecución

Una vez adaptado el algoritmo se realizarán algunas pruebas con una muestra de datos controlada para cuantificar los tiempos de ejecución del algoritmo bajo distintas configuraciones, realizando corridas del algoritmo modificando la cantidad de núcleos que el equipo utilizado en cada ocasión.

A través de esas pruebas podremos tener una perspectiva de la mejora en la eficiencia de ejecución del algoritmo aprovechando nuevas tecnologías como lo es

el procesamiento en paralelo recientemente aplicado para mejorar la eficiencia de procesos de renderizado de polígonos en 3D.

5.8 Estimación de valores faltantes para la base de datos final

Finalmente se realizó una corrida final para la estimación de los valores faltantes en todas las imágenes que conformen nuestra base de datos de muestra, ejecutando la rutina en función de las condiciones identificadas como óptimas para la estimación de una mayor cantidad y calidad de datos.

Teniendo en cuenta que la única restricción del algoritmo es la existencia de una relación temporal entre las mediciones, podemos aplicar esta metodología sin importar que tipo de imágenes utilicemos, la resolución temporal, así como la extensión de nuestra base de datos.

A continuación, se muestra el diagrama correspondiente a la metodología de esta investigación Figura 5-5.

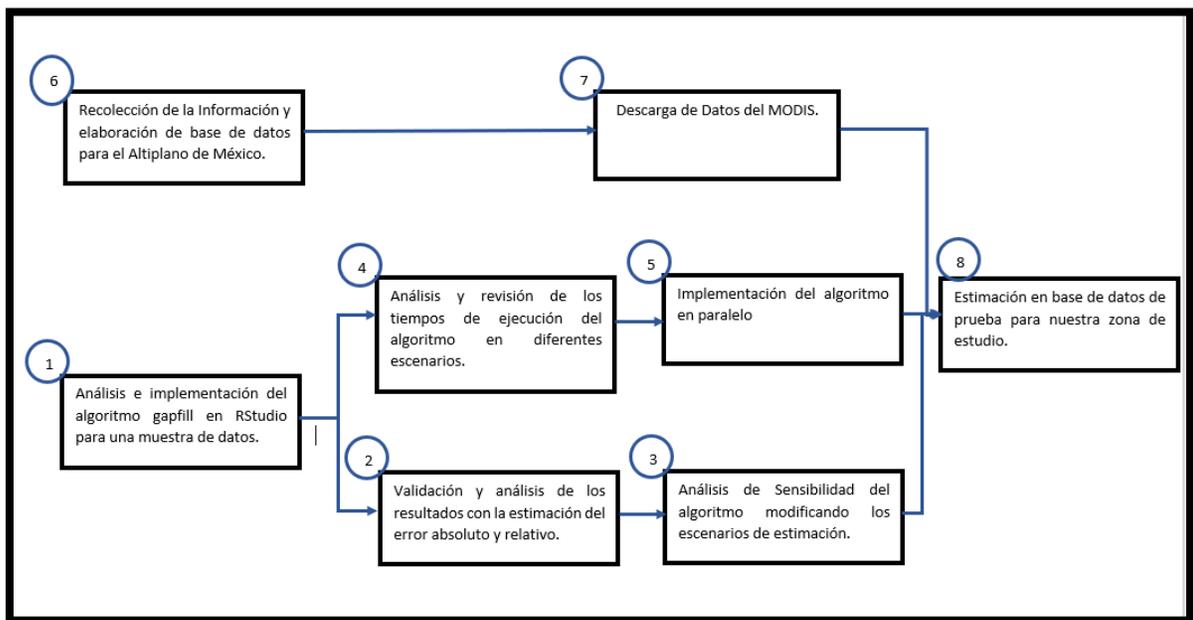


Figura 5-5 Diagrama de la metodología (Fuente: Elaboración propia).

6. RESULTADOS.

6.1 Comparación entre algunos algoritmos existentes y el método gapfill

A continuación, se presenta en la Tabla 6.1 una comparativa entre las ventajas y desventajas de algunos de los métodos disponibles para la estimación de valores faltantes en imágenes satelitales contra el algoritmo gapfill, el cual será sujeto de estudio de esta investigación.

Tabla 6-1 Comparativa entre ventajas y desventajas de cada método

Método	Ventajas	Desventajas
Superficies Reflejantes (Myneni <i>et al.</i> , 2002)	El método y su desempeño han sido ampliamente estudiados. Buenos resultados para la estimación de LAI y FPAR con calibración en tierra.	Requiere de una calibración mediante puntos en tierra georreferenciados. Se requiere de correcciones atmosférica pues la presencia de nubes afecta los resultados. Se requiere una subclasificación previa del área de estudio en alguno de los 6 tipos de biomas establecidos en la metodología.
Segmentación (Li <i>et al.</i> , 2010)	Método orientado a la respuesta y visualización de información en tiempo real. Requiere de una menor capacidad de poder de procesamiento	Requiere de una mayor cantidad de etapas intermedias al procesar en diferentes etapas la imagen hasta cubrirla en su totalidad.
Hadoop (Codella, Hua, Natsev, & Smith, 2011)	Método mayormente orientado a la clasificación de uso de suelo cuando no se cuenta con información geográfica.	Se requiere conocimientos del framework Hadoop para la clasificación semántica de imágenes. No está del todo optimizado el método. Se requiere una preclasificación de uso de suelo del área de estudio.

<p>Procesamiento en la nube (Evangelidis, Ntouros, Makridis, & Papatheodorou, 2014)</p>	<p>No se requiere de grandes cantidades de espacio de almacenamiento.</p> <p>Posibilidad de ejecutar múltiples herramientas de geoprocésamiento de imágenes ejecutables en la nube.</p>	<p>Necesidad de una permanente conexión a internet.</p> <p>Vulnerabilidad de la información en la nube.</p> <p>Requiere de conocimientos en el framework en aplicaciones web para procesamiento geoespacial.</p>
<p>Cluster (Adhianto <i>et al.</i>, 2010)</p>	<p>De igual forma da solución al problema de grandes requerimientos de espacio de almacenamiento distribuyendo la información en múltiples equipos.</p> <p>Requiere de una preclasificación de la información</p>	<p>Retrasos a causa de la constante verificación de protocolos web por la verificación del firewall con cada ejecución.</p>
<p>Gapfill (Gerber <i>et al.</i>, 2017)</p>	<p>Algoritmo ampliamente estudiado y documentado en la plataforma RStudio</p> <p>No se requiere de una preclasificación de datos</p> <p>Estimación por medio de métodos de interpolación espacial con correcciones por medio de series de tiempo</p> <p>Potencial para implementación del algoritmo en paralelo</p>	<p>El tamaño de la base de datos aun se ve limitada por la capacidad de almacenamiento del equipo</p> <p>Aplicable únicamente en bases de datos con la misma separación temporal entre imágenes</p> <p>La precisión del algoritmo dependerá de la extensión de la base de datos y el porcentaje de datos faltantes</p>

Tal y como se mencionó anteriormente los procesos de teledetección en la actualidad se ven limitados cuando se trata de analizar grandes cantidades de datos

por lo que en esta primer metodología a comparar, los autores proponen una segmentación y procesamiento parcial de las imágenes hasta lograr analizar la imagen en su totalidad (Li *et al.*, 2010).

Una segunda técnica analizada es la de la clasificación de la imagen satelital en diferentes categorías de cobertura de suelo usando una clasificación basada en un entorno “Hadoop” que nos permite trabajar con miles de nodos y una gran cantidad de datos programando rutinas con java (Codella *et al.*, 2011).

El tercer método analizado se vale del aprovechamiento de los servicios en la nube en lo que el autor llama una transmisión de geo-espacial a la nube en diferentes etapas del procedimiento (Evangelidis *et al.*, 2014).

El cuarto método y el cual se presume es el más utilizado en la actualidad es el de crear una infraestructura de clúster en donde se gestiona y procesa una gran cantidad de información que es distribuida en diversos centros de información (Adhianto *et al.*, 2010).

Finalmente se realiza la comparación con un método basado en la física de la transferencia de radiación en áreas de vegetación para la recuperación del “Leaf Area Index” (LAI) y la Fraction of Absorbed Photosynthetically Active Radiatio” (FPAR) de superficies reflejantes desarrollado e implementado para el procesamiento de imágenes antes del lanzamiento del MODIS por sus siglas en inglés (Myneni *et al.*, 2002).

6.2 Adaptación y ejecución del algoritmo para una muestra de datos

El primer paso para la adaptación del algoritmo consistió en la selección de una base de datos de muestra generada a partir de imágenes de NDVI provenientes del MODIS para los días 145,161,177 y 193 en días julianos que corresponden al 24 de mayo, 24 de junio y 11 de julio con una resolución de 21x21 pixeles Figura 6-1.

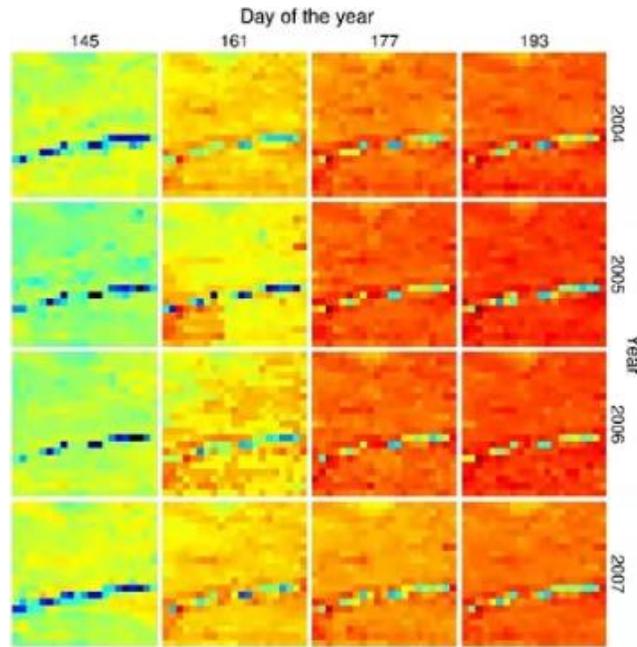


Figura 6-1 Arreglo de datos utilizado para la validación (Fuente: Elaboración propia).

Podemos decir que nuestra base de datos está compuesta por 441 valores diarios, 1764 valores anuales, un equivalente de 7056 valores para los 4 años con una misma separación de un año entre cada registro, esta base de datos fue seleccionada pues su estudio y análisis ha sido ampliamente documentados.

Una vez que ingresamos nuestro arreglo de datos de 4 dimensiones, Rstudio habrá almacenado los datos de NDVI en una estructura de datos con las dimensiones antes mencionadas y los valores correspondientes para cada campo, ya sea el valor correspondiente para cada año, día, coordenadas, así como los valores correspondientes a cada píxel teniendo un arreglo como el presentado a continuación:

```
> str(ndvi)
num [1:21, 1:21, 1:4, 1:4] 0.522 0.526 0.459 0.492 0.476 ...
- attr(*, "dimnames")=List of 4
..$ : chr [1:21] "-153.032" "-153.012" "-152.992" "-152.973" ...
..$ : chr [1:21] "69.1" "69.12" "69.14" "69.16" ...
..$ : chr [1:4] "145" "161" "177" "193"
..$ : chr [1:4] "2004" "2005" "2006" "2007"
```

A continuación, se presentan algunos resultados de la distribución de los datos Figura 6-2 de los datos, tal y como se mencionó anteriormente, mediante su análisis nos será posible realizar una primera inspección de los datos para identificar los años en los que se tiene una menor cantidad de datos, identificar segmentos con grandes carencias de información y de esta manera decidir si se excluirán registros o se trabajara con un periodo en común diferente al seleccionado previamente.

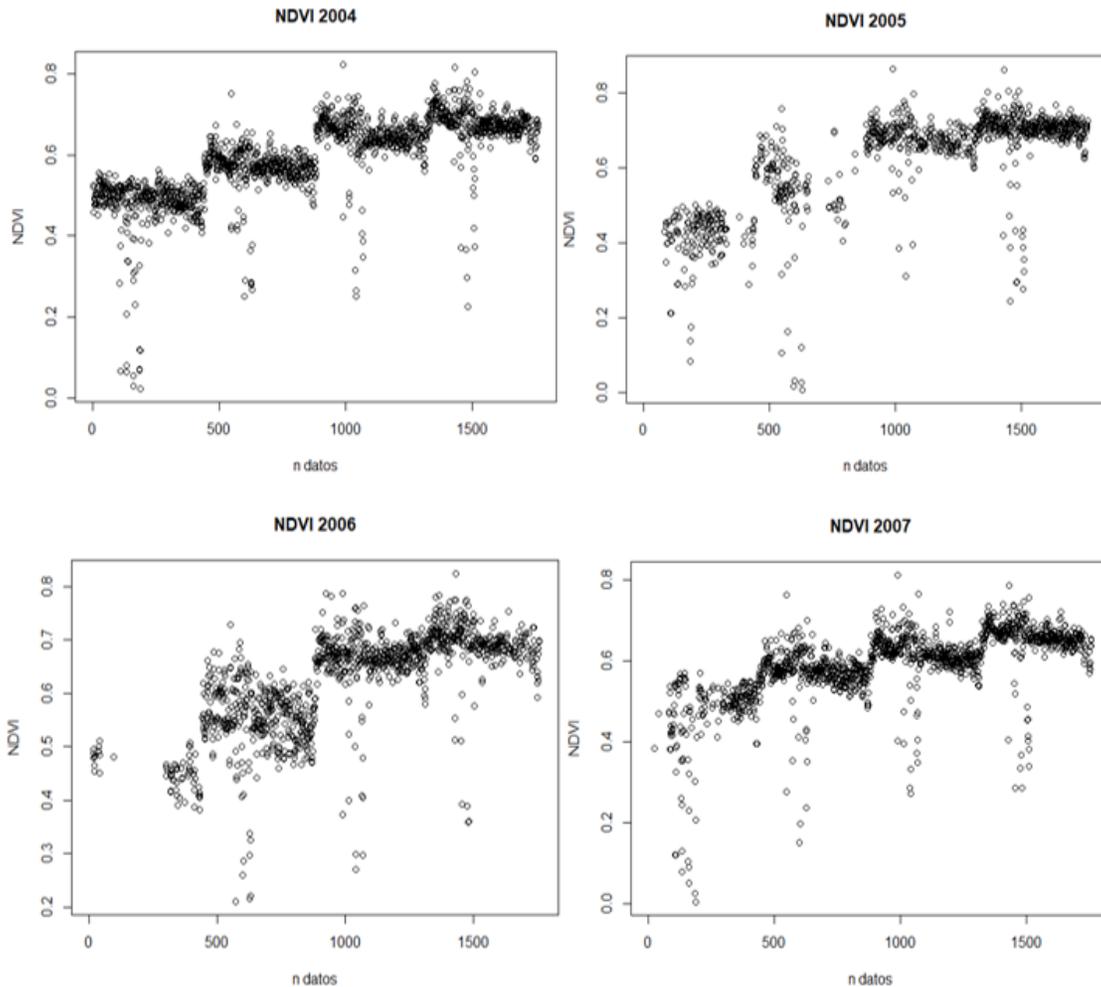


Figura 6-2 Valores de NDVI diarios para cada uno de los años de registros. (Fuente: Elaboración propia).

Este algoritmo se basa en la estimación de valores faltantes a partir del supuesto que los valores en nuestra base de datos tienen una continuidad espacial con una alta resolución temporal es decir los valores conocidos tienen un comportamiento

similar espacialmente que nos permite estimar el valor de sus vecinos faltantes haciendo las correcciones correspondientes con una mayor precisión en función de nuestra resolución temporal.

Una vez ingresada nuestra base de datos y tras una primera inspección se procedió a la modificación de nuestra base de datos mediante la extracción de pequeñas secciones en diversas zonas de las imágenes con ayuda del algoritmo presentado en el anexo 4 y posteriormente proceder a la estimación de estos valores ya conocidos.

6.3 Validación del algoritmo

La validación de los resultados obtenidos se realizó a través de una validación cruzada, reemplazando de la matriz de valores conocidos, ciertas regiones de la imagen por valores de pixeles por NA Figura 6-3 y así estimarlos ejecutando el algoritmo calculando el error absoluto y relativo para cada uno de estos pixeles.

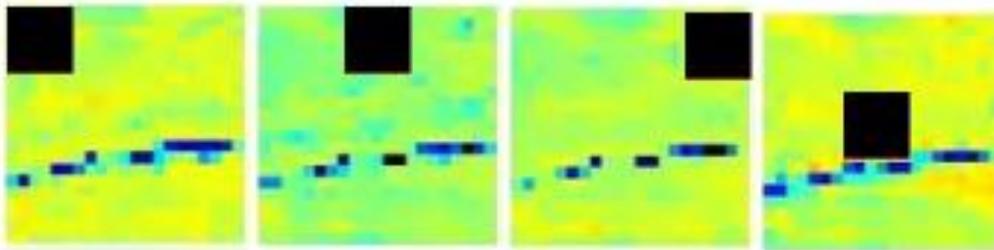


Figura 6-3 Extracción de valores en diferentes zonas de la misma imagen (Fuente: Elaboración propia).

Como parte de la validación anteriormente mencionada se analizaron múltiples casos retirando una matriz de 10x10 pixeles en el centro de la imagen y acercándonos a las esquinas de esta en cada una de las corridas del algoritmo con lo que se obtuvieron los siguientes resultados, para el primer caso de estudio se realizó la extracción de una matriz en el centro reemplazando los valores para posteriormente volver a estimarlos y de esta manera comparar los valores medidos con los estimados, es así que se obtuvo un error promedio en un rango de entre el 8 y 16%, a medida que nos acercábamos a las esquinas el error promedio fue de entre un 23% y 42% tal y como se aprecia en la Figura 6-4.

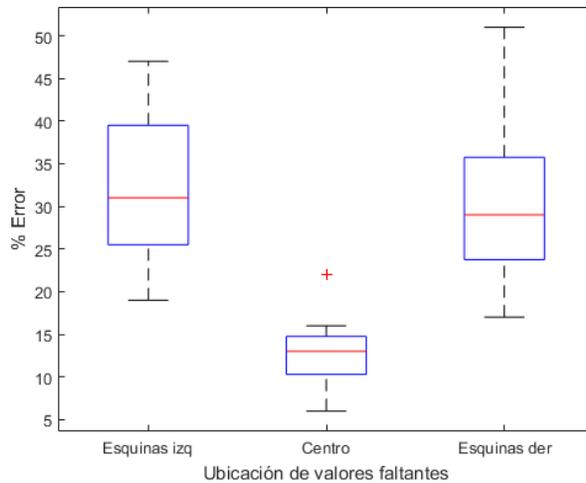


Figura 6-4 Intervalos correspondientes al error relativo para los valores resultantes para distintas estimaciones en función de la ubicación de los valores faltantes (Fuente: Elaboración propia).

Los resultados obtenidos una vez realizada la validación cruzada fueron presentados como diferentes gráficas de dispersión, cada una de ellas para diferentes porcentajes de valores faltantes pero esta vez sin importar la zona en que los valores fuesen extraídos.

En la Figura 6-5 tenemos la gráfica de dispersión correspondiente a la ejecución del algoritmo para la estimación de valores faltantes teniendo en cuenta que el 40% de los datos era valores faltantes, en ella tenemos presente la mayor de las dispersiones tomando como referencia los valores originales de las imágenes.

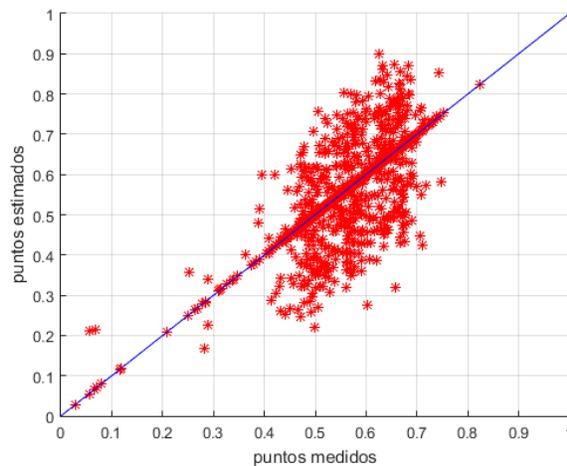


Figura 6-5 Dispersión de los datos para la estimación con 40% de valores faltantes (Fuente: Elaboración propia).

Adicional a esto se realizó una gráfica de frecuencias para conocer la distribución de los valores Figura 6-6 para el caso anterior.

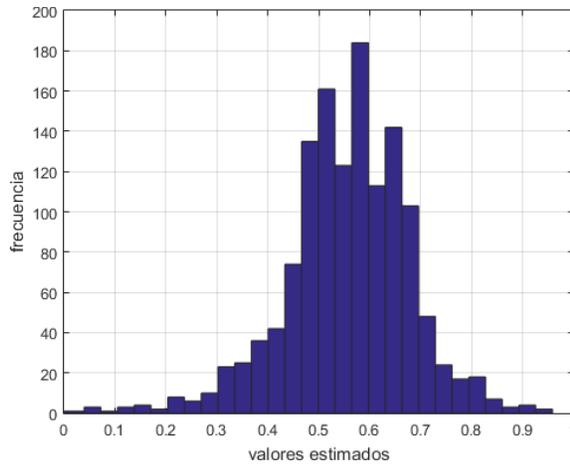


Figura 6-6 Distribución de frecuencias de los valores estimados para el caso con 40% de valores faltantes (Fuente: Elaboración propia).

El siguiente caso analizado fue la estimación de valores para un escenario con 30% de valores faltantes, en la Figura 6-7 podemos ver la gráfica de dispersión correspondiente a la ejecución del algoritmo para la estimación de valores faltantes teniendo en cuenta que el 30% de los datos era valores faltantes, en ella tenemos presente la mayor de las dispersiones tomando como referencia los valores originales de las imágenes.

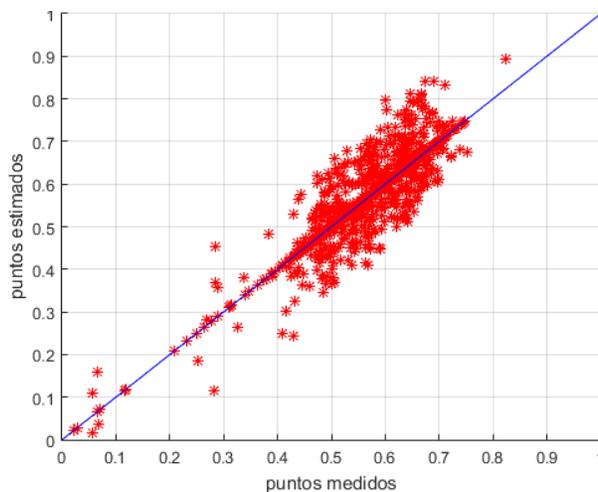


Figura 6-7 Dispersión de los datos para la estimación con 30% de valores faltantes (Fuente: Elaboración propia).

En la Figura 6-8 podemos observar la distribución de frecuencias correspondiente a la estimación con un 30 % de valores faltantes.

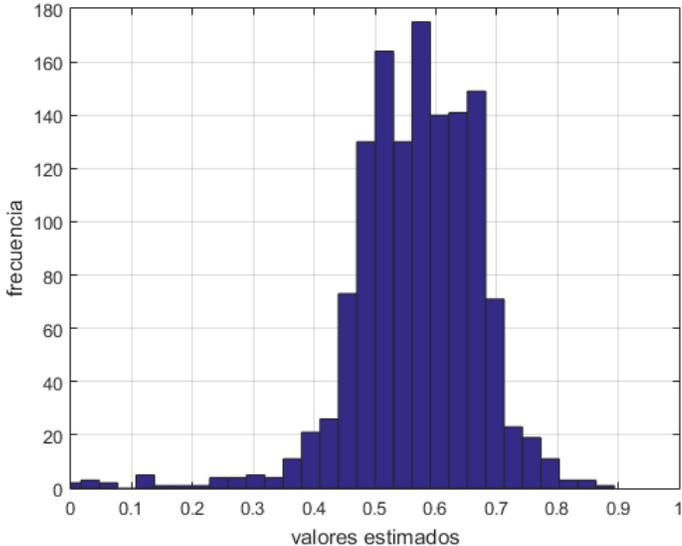


Figura 6-8 Distribución de frecuencias de los valores estimados para el caso con 30% de valores faltantes (Fuente: Elaboración propia).

El penúltimo caso analizado fue el de la ejecución del algoritmo para un escenario con el 20% de valores faltantes, en la Figura 6-9 se presenta su gráfica de dispersión correspondiente.

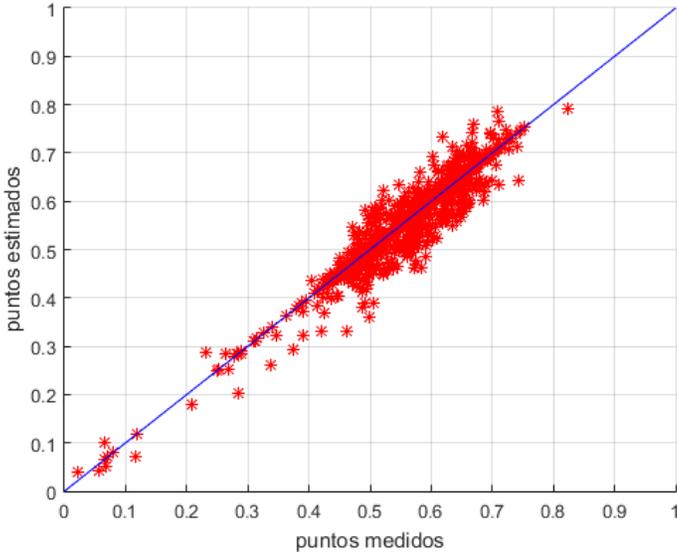


Figura 6-9 Dispersión de los datos para la estimación con 20% de valores faltantes (Fuente: Elaboración propia).

A continuación, se presenta la distribución de frecuencias Figura 6-10 correspondiente a la estimación con un 20% de valores faltantes.

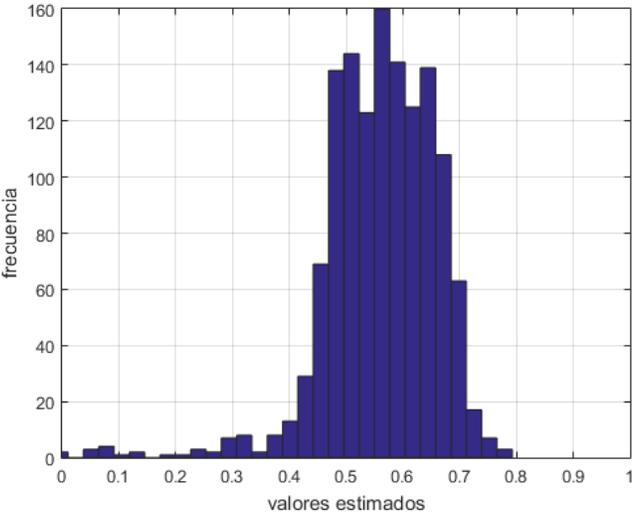


Figura 6-10 Distribución de frecuencias de los valores estimados para el caso con 20% de valores faltantes (Fuente: Elaboración propia).

Finalmente se presenta la gráfica de dispersión para la estimación con un 10% de valores faltantes Figura 6-11 con la cual tal y como era previsto al contar con una mayor cantidad de datos medidos se reduce la incertidumbre del algoritmo obteniendo una mejor estimación con una menor distribución de los datos es decir se obtuvo la mejor de las estimaciones de los valores faltantes.

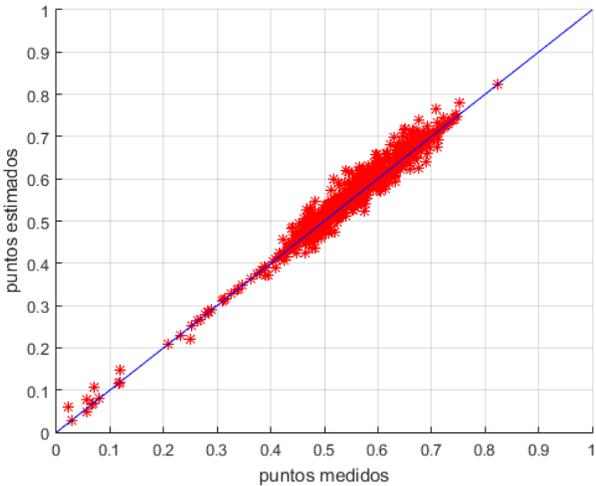


Figura 6-11 Dispersión de los datos para la estimación con 10% de valores faltantes (Fuente: Elaboración propia).

Adicionalmente se muestra su gráfica de barras con la frecuencia Figura 6-12 de los datos estimados para un 10% de valores faltantes.

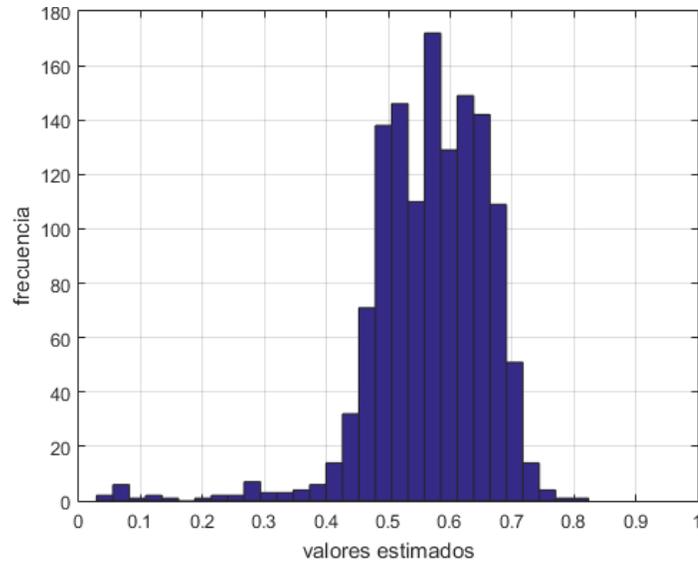


Figura 6-12 Distribución de frecuencias de los valores estimados para el caso con 10% de valores faltantes (Fuente: Elaboración propia).

En la Figura 6-13 se muestra la distribución de frecuencias para los valores originales por lo que por simple inspección vemos que la gráfica que más se asemeja a esta es la correspondiente al 10% de valores faltantes misma con la que obtuvimos la mejor aproximación de los resultados.

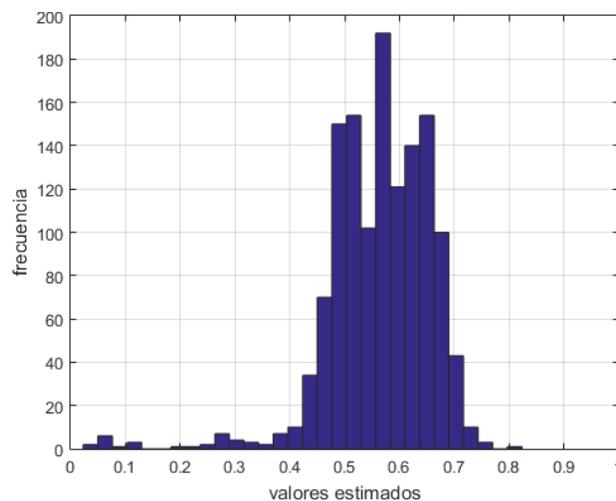


Figura 6-13 Distribución de frecuencias de los valores originales registrados por el sensor (Fuente: Elaboración propia).

En las gráficas anteriores podemos apreciar con bastante claridad el aumento en la dispersión de los datos a medida que la cantidad de valores faltantes aumentaba y es gracias a este análisis que podemos realizar estimación del error Figura 6-14 en cada una de las corridas y para cada uno de los casos analizados anteriormente.

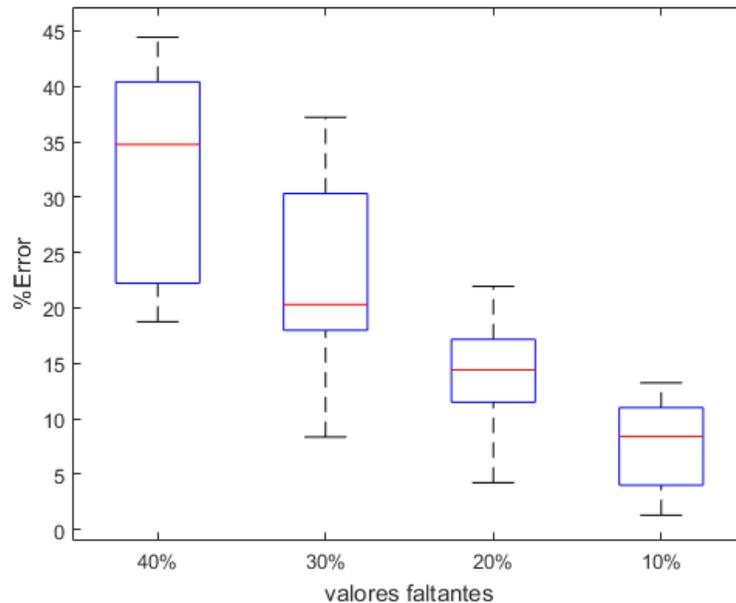


Figura 6-14 Intervalos correspondientes al error relativo para los valores estimados tras la ejecución del algoritmo para distintos porcentajes de valores faltantes (Fuente: Elaboración propia).

Mediante el análisis estadístico de la información anterior se obtuvo la desviación estándar para cada uno de los 4 casos anteriores obteniendo los siguientes valores, 0.1215 para el 40% de valores faltantes, 0.1053 para el 30%, 0.0977 para el 20% y 0.09456 para el 10% mostrando como era de esperarse valores superiores para las primeras pruebas.

Un segundo parámetro calculado fue el coeficiente de asimetría en donde se obtuvieron los siguientes valores de -0.0112 para el 40% de valores faltantes, -0.0113 para el 30%, -0.13113 para el 20% y -0.01335 para el 10% confirmando una clara asimetría negativa desplazada a la izquierda como bien lo podemos apreciar en las gráficas de frecuencias de los datos.

Una vez conocido el comportamiento del algoritmo se realizó la estimación de valores faltantes en una base de datos correspondiente al estado de Querétaro

mismo que se encuentra dentro de nuestra zona de estudio, estas fueron descargadas del sitio: “<https://earthexplorer.usgs.gov/>”, se analizaron las imágenes necesarias para la estimación de valores faltantes de NDVI para el periodo comprendido entre el 01/09/2018 al 30/09/2018.

Cada una de estas imágenes con una resolución espacial de 1 kilómetro por píxel con valores correspondientes al promedio diario del valor medido con una longitud de escena de 2332 kilómetros.

A continuación, se muestra una de las 30 escenas cuyos valores faltantes ocasionados por la interferencia de la nubosidad fueron estimados con la implementación del algoritmo Figura 6-15.

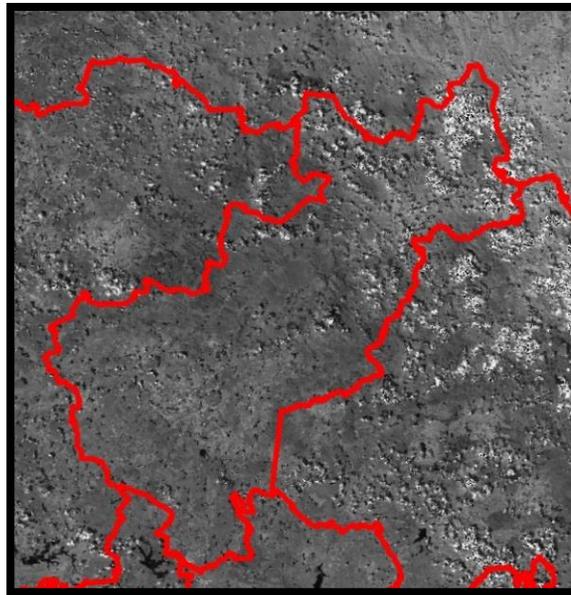


Figura 6-15 Valores de NDVI sin previo tratamiento para el estado de Querétaro (Fuente: Elaboración propia).

Esta escena corresponde al día 26 de septiembre y fue elegida al ser el caso ideal en el que el algoritmo debería de ser aplicado basándonos en los análisis anteriormente mencionados al contar con un porcentaje de valores faltantes menor al 20% con lo que se obtendría un error en las estimaciones menor al 15%, en la Figura 6-16 se muestra el resultado final tras la aplicación del algoritmo en la escena anteriormente descrita.

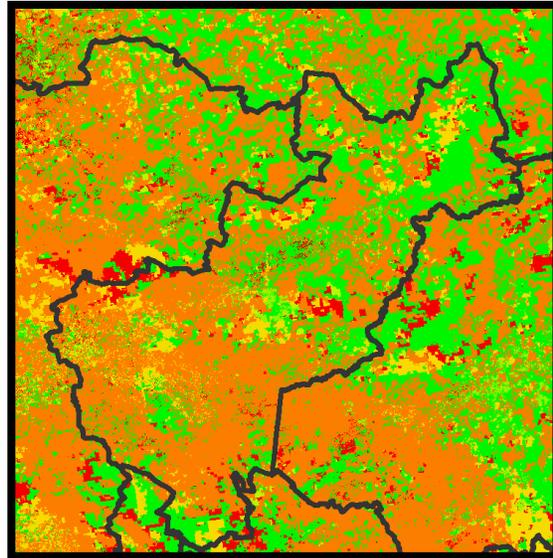


Figura 6-16 Valores estimados de NDVI aplicando el algoritmo gapfill (Fuente: Elaboración propia).

Como parte de la búsqueda para hallar otras aplicaciones de este algoritmo en la hidrología, el algoritmo fue implementado para la estimación de valores faltantes en información correspondiente a imágenes de radar con dos diferentes resoluciones temporales.

La primera de ellas aplicando el algoritmo a una serie de datos con una resolución temporal diaria conformada por 123 imágenes de radar cada una con una resolución de 636x365 píxeles, es decir un mes de mediciones para 4 años: 1985, 1993, 1994 y 2003 como el ejemplo presentado en la Figura 6-17.

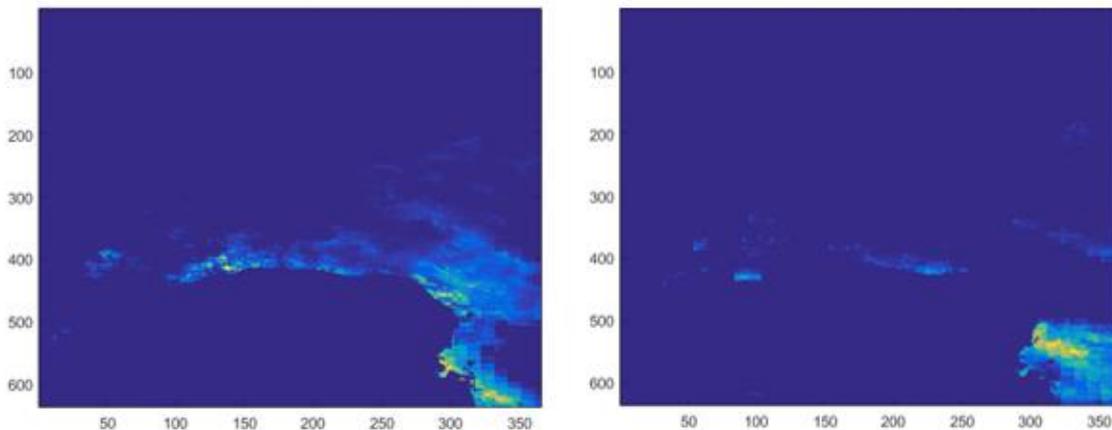


Figura 6-17 Imágenes de radar correspondientes al día 1 del año 1985 y 1993

En la Figura 6-18 se presentan los valores resultantes tras la implementación del algoritmo para la estimación de valores faltantes de precipitación en una base de datos de imágenes de radar.

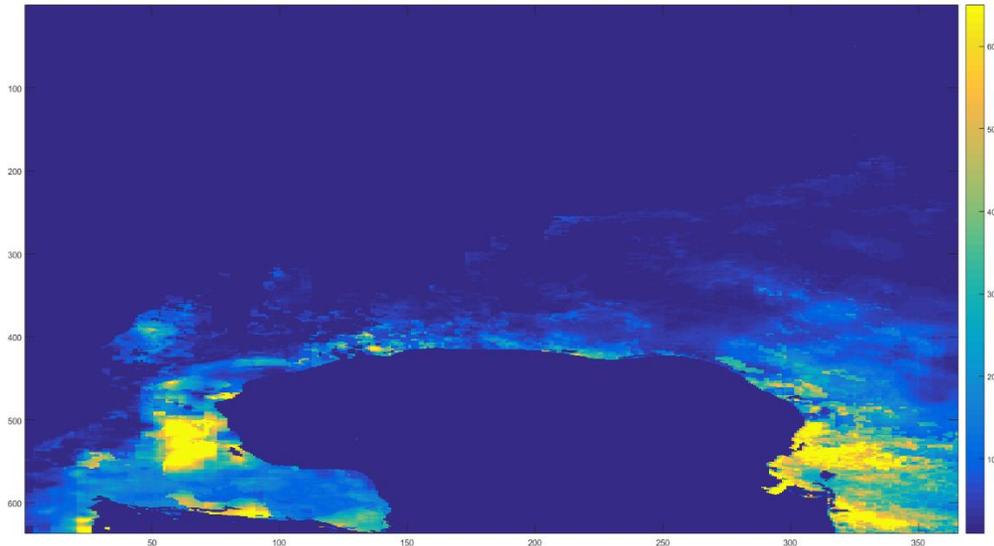


Figura 6-18 Valores de precipitación resultantes para estimar valores faltantes en la escena correspondiente al día 1 de 1985.

Si bien es posible distinguir parte de la forma de los campos de lluvias presentes el día 1 de 1985 también podemos notar una clara perturbación en los valores de los píxeles en los que no debería existir precipitación, esto debido al proceso de estimación el cual se auxilia de los valores para el píxel en la misma posición en escenas anteriores y posteriores.

En cuanto a los valores estimados, al tratarse de una interpolación es correcto afirmar que los valores estimados siempre estarán dentro de los límites que establece el rango de valores medidos sin embargo los resultados de la validación cruzada muestran errores entre el 78% y el 91% tan solo para el 20% de datos faltantes sin contar la aparición de valores en donde no debería haber registros originalmente.

Los resultados obtenidos pueden atribuirse a la variación espacial y temporal de la lluvia pues el cambio entre mediciones es más notorio a medida que se incrementa el intervalo entre cada escena.

6.4 Implementación del algoritmo en paralelo y comparativa de tiempos de procesamiento

Tal y como se mencionó anteriormente, el atractivo de la implementación del algoritmo gapfill reside en su posible implementación en Paralelo con el fin de realizar de forma más eficiente la estimación de valores faltantes en grandes volúmenes de datos, es por esto por lo que la plataforma RStudio ya contiene un paquete con una serie de funciones que nos permiten realizar esta tarea.

La biblioteca tiene por nombre doParallel creada por Rich Gallaway la cual nos permite la ejecución de algoritmos en paralelo utilizando la función “registerDoParallel(cores=n)” donde n es el número de núcleos con el cual se realizará la corrida, en el caso de que se pusieran más núcleos de los que se cuenta solamente aparecerá un mensaje de error.

Es importante destacar que la implementación de esta librería solamente fue para correr el algoritmo en paralelo utilizando los núcleos disponibles de la CPU para lo cual fue utilizada la función “detectCores” que como resultado nos arroja el número de núcleos disponibles para su utilización.

Es importante tener en cuenta que el equipo utilizado para esta investigación fue un ordenador Asus 752Vm con un procesador Intel core i7-6700HQ a 2.6GHz con una tarjeta de video NVIDIA GTX1060 de 6GB GDDR5 por lo que los resultados estarán en la función de la capacidad de procesamiento de cada equipo.

Realizando una tabla comparativa entre los tiempos de ejecución del algoritmo para nuestra base de datos de NDVI, se obtuvieron los resultados mostrados en la Tabla 6-3, en ella podemos observar que si realizamos una corrida del algoritmo bajo las condiciones estándar y sin haber modificado nada, el tiempo de procesamiento es de aproximadamente 30 segundos pero si utilizamos las funciones mencionadas anteriormente, este tiempo de procesamiento pasa de 0.01815 segundo por estimación de valor faltante a 0.00375 segundos con lo que estaríamos logrando una mayor eficiencia del algoritmo, reduciendo el tiempo de procesamiento hasta en un 70%.

Tabla 6-2 Tabla comparativa de tiempos de procesamiento en paralelo CPU.

Capacidad		
(%)	minutos	seg/NA
default	0.485	0.01815
20	0.13	0.00488
40	0.107	0.00399
100	0.1	0.00375

La información presentada anteriormente se encuentra representada en la Figura 6-17 a través de una curva de desempeño en la que podemos observar el comportamiento del algoritmo, en esta se compara el tiempo de procesamiento contra el numero de núcleos utilizados para el cálculo de valores faltantes.

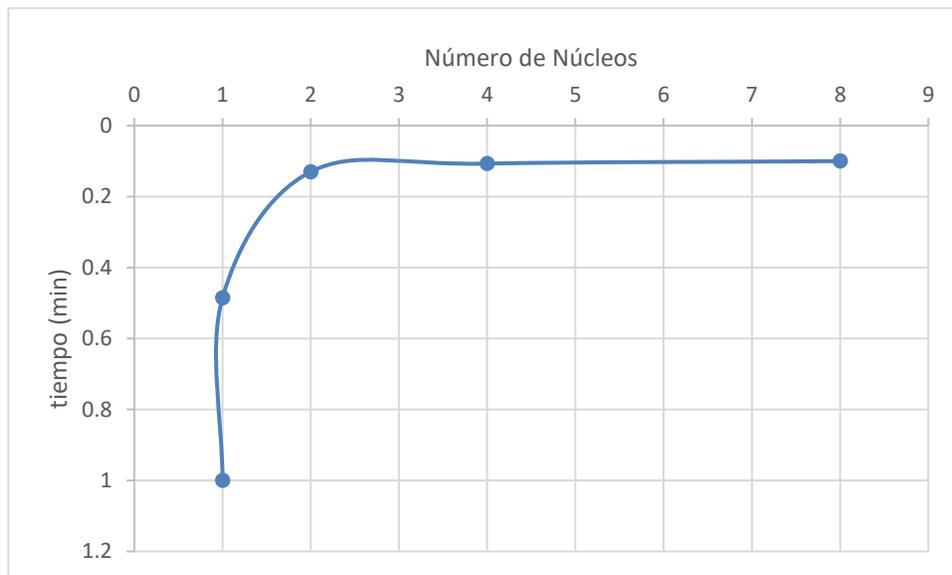


Figura 6-19 Curva de desempeño del algoritmo.

En la gráfica anterior podemos observar que a medida que aumentamos el número de núcleos para la estimación de valores faltantes, el tiempo de procesamiento disminuye hasta alcanzar una reducción del 70% antes mencionada.

7. CONCLUSIONES.

A partir del análisis de los resultados podemos concluir que la implementación del algoritmo gapfill en imágenes provenientes de sensores remotos nos permite tener una mejora tanto en la cantidad de datos como en la calidad de estos, la calidad dependerá de diversos factores, tres de los más importantes mencionados a continuación:

- Extensión de la base de datos
- Ubicación de los valores a estimar
- Porcentaje de valores faltantes

Es posible concluir que la variable más importante es el porcentaje de valores faltantes con el que se desea realizar la estimación pues el error aumenta de manera considerable a medida que aumenta la ausencia de valores.

La segunda variable de mayor importancia para obtener una mejora en los resultados es la extensión de la base de datos, al formar parte de las correcciones realizadas por el método gapfill con la lectura de los valores de cada píxel interpretándolo como una serie de tiempo, esta variable tiene una mayor trascendencia.

Es necesario tener cautela con la recomendación anterior pues si bien el aumento en la cantidad de registros puede beneficiar y traer una mejora para el entendimiento de fenómenos globales, es posible que el manejo de estas grandes cantidades de datos nos exija una mayor inversión en equipo de cómputo y poder de procesamiento para poder llevar a cabo esta tarea además de que día con día nuestra base de datos irá en aumento con nuevos registros.

En tercer lugar, pero no menos importante encontramos la ubicación de los valores faltantes que serán estimados, al realizar una primera estimación a través de interpolación espacial, el algoritmo tendrá una mejor aproximación entre mayor sea el número de datos, situación que se complica cuando el valor se acerca a las esquinas y solo se puede trabajar con la sección inferior o superior de la matriz de vecinos reduciendo la cantidad de datos al 50 o inclusive el 25%.

El objetivo general de esta tesis fue el de mejorar la calidad de la información en base de datos con la que se cuenta para el área correspondiente al altiplano de México, sin embargo, el algoritmo no está limitado a nuestra base de datos, realizando las modificaciones pertinentes la rutina podría ser ejecutada para el procesamiento de cualquier tipo de imágenes satelitales con la reserva de que quien la implemente haga la validación de los resultados correspondiente.

Este algoritmo se basa en la estimación de valores faltantes a partir del supuesto que los valores en nuestra base de datos tienen una continuidad espacial con una alta resolución temporal es decir los valores conocidos tienen un comportamiento similar espacialmente lo que nos permite estimar el valor de sus vecinos faltantes haciendo las correcciones correspondientes con una mayor precisión.

Es importante recordar que al tratarse de un método cuya base es la interpolación, no podremos estimar valores extremos que no hayan sido registrados dentro de las mediciones, es decir los valores estimados tendrán como límites los valores mínimos y máximos que hayan sido registrados por el sensor.

Respecto a los tiempos de procesamiento de los datos, como era de esperarse mediante las bondades del procesamiento en paralelo se logró obtener una mejora de 23.1 segundos, una mejor correspondiente a la reducción de tiempos en un 70%.

Si bien a través del análisis de la curva de desempeño del algoritmo podemos notar un límite en lo que supone la capacidad de procesamiento del algoritmo, puede ser interesante realizar esta misma revisión de los tiempos de ejecución, pero en esta ocasión aumento la capacidad en cuanto a memoria RAM y descartar si esta limitante se debe a la memoria y que se esté generando un cuello de botella; situación muy común en el mundo de la informática.

Hoy día dado que nos encontramos en una etapa donde el flujo de la información crece rápidamente en los diferentes campos de la informática, ya sea en seguridad, educación, videojuegos, medicina, etc. El manejo de información ha llegado a un punto en el que llamamos a estos grandes volúmenes de información BigData (Nina *et al.*, 2015).

Trasladando esto a una escala mayor resto representaría un aumento en la eficiencia que nos permitirá reducir los tiempos en horas tratándose del análisis de bases de datos con una mayor extensión.

Para el caso de esta investigación se mantuvo como método de interpolación por default el vecino más cercano sin embargo cada usuario podría seleccionar o ingresar su propia rutina de acuerdo con la experiencia para la obtención de mejores resultados en cuanto a precisión que pudiesen mejorarla en los valores de frontera por efectos de borde.

Si bien es cierto nos limitamos al análisis de una serie de escenas con las mismas dimensiones y resolución espacial es posible reducir el error en las estimaciones cuando los valores se encuentran en las esquinas integrando una nueva imagen a partir de un mosaico con las imágenes vecinas logrando dejar como nuevo centro de la escena la esquina en donde se presentaron los valores faltantes.

En este caso se trabajó con imágenes con un ancho de escena de 2332 kilómetros y una resolución espacial de 1000 metros, el algoritmo no se limita a este tipo de imágenes dejándonos la posibilidad de realizar estimaciones en imágenes obtenidas a partir de imágenes provenientes de sensores espectrales montados en drones con resoluciones de centímetros.

La información obtenida dependerá de la escala con la que se esté trabajando pues evidentemente no se obtendrán los mismos resultados estimando el valor para un píxel de 1km que para un píxel de 9 centímetros.

Los valores de NDVI estimados nos permitirán tener un mejor entendimiento de los procesos en la superficie terrestre con la posibilidad de realizar una mejor estimación de la evapotranspiración como parte del balance hídrico debido a la urgencia de la situación para realizar mejores diseños que se adapten a las condiciones presentes que acontecen en la atmósfera que afectan la intensidad y duración de las lluvias así como los periodos en que estas se presentaban, de igual forma nos será posible tener un mejor entendimiento de la relación entre la calidad

de la vegetación y la época del año en la que nos encontremos así como su relación con la concentración de CO₂ en la atmósfera.

Es bien sabido que las actividades humanas están directamente relacionadas con la producción de CO₂ y este aumento a su vez se cree está relacionado con el cambio climático por lo que nos corresponde además de regular nuestras actividades, también mantener un constante monitoreo de la superficie terrestre para la obtención de algunos parámetros que nos permitan la caracterización de esta.

Adicionalmente al entendimiento que nos brinda una mayor cantidad y calidad de la información obtenida de sensores remotos, a través de las estimaciones a nivel de cada píxel tendremos un beneficio cuando se trate del uso de esta información para alimentar modelos de parámetros distribuidos en los que cada una de las estimaciones se realiza a nivel local, es decir de cada una de las estimaciones se realiza de manera independiente y simultánea ligando la salida del n-ésimo valor con la entrada del n+1-ésimo valor.

En el primer caso nos enfocamos en la utilización de esta metodología en la información de productos de vegetación obtenidos a partir de sensores remotos, información que nos permite caracterizar la vegetación y con esto el monitoreo de uno de los principales componentes del ciclo hidrológico, la evapotranspiración.

A través de esta información nos será posible realizar nuevas investigaciones para relacionar la calidad de la vegetación medida a partir de sensores remotos y relacionarla con fenómenos como la erosión, sequías, monitoreo de plagas, entre otros.

Para el segundo caso analizado, el de la estimación de valores de precipitación a partir de imágenes de radar no fue posible obtener buenos resultados con la información disponible, como recomendación podría hacerse un segundo análisis en futura investigación con una base de datos de por lo menos 100 imágenes para cada escena así como también limitando el número de estas antes y después que serán utilizadas para la interpolación y de esta manera minimizar la aparición de

valores en donde originalmente no debería existir, por último sería prudente obtener imágenes con la mayor resolución temporal posible para de esta manera contribuir a reducir el efecto antes mencionado.

8. LITERATURA CITADA.

- Adhianto, L., Banerjee, S., Fagan, M., Krentel, M., Marin, G., Mellor-Crummey, J., & Tallent, N. R. (2010). HPCTOOLKIT: Tools for performance analysis of optimized parallel programs. *Concurrency Computation Practice and Experience*, 22(6), 685–701. <https://doi.org/10.1002/cpe>
- Alfaro, E. J., & Soley, F. J. (2009). Descripción de dos Métodos de Rellenado De Datos Ausentes En Series De Tiempo Meteorológicas, 16(1), 60–75.
- Canada, N. R. (1976). *Fundamentals of remote sensing. Resources Policy* (Vol. 2). [https://doi.org/10.1016/0301-4207\(76\)90065-9](https://doi.org/10.1016/0301-4207(76)90065-9)
- Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., & Jin, S. (2011). A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sensing of Environment*, 115(4), 1053–1064. <https://doi.org/10.1016/j.rse.2010.12.010>
- Codella, N. C. F., Hua, G., Natsev, A., & Smith, J. R. (2011). Towards large scale land-cover recognition of satellite images. *ICICS 2011 - 8th International Conference on Information, Communications and Signal Processing*, 6–10. <https://doi.org/10.1109/ICICS.2011.6174291>
- Conabio MODIS. (2017). Retrieved from <http://speck.conabio.gob.mx/modis/2010/1002/1002.html>
- Dickinson, R. E. (1995). Land processes in climate models. *Remote Sensing of Environment*, 51(1), 27–38. [https://doi.org/10.1016/0034-4257\(94\)00062-R](https://doi.org/10.1016/0034-4257(94)00062-R)
- Evangelidis, K., Ntouros, K., Makridis, S., & Papatheodorou, C. (2014). Geospatial services in the Cloud. *Computers and Geosciences*, 63, 116–122. <https://doi.org/10.1016/j.cageo.2013.10.007>
- Fallas, J. (2007). Modelos digitales de elevación: Teoría, métodos de interpolación y aplicaciones, 83. Retrieved from <https://revistas.ucm.es/index.php/AGUC/article/viewFile/AGUC0000110049A/31281>

- FAO. (2011). Medir la degradación del bosque. *Unasylva*, 62(238), 72. <https://doi.org/ISSN 0251-1584>
- Felicísimo, A. (2007). Modelos digitales del terreno: introducción y aplicaciones a las ciencias ambientales. *Oviedo: Universidad de Oviedo*, 118. Retrieved from <http://www.etsimo.uniovi.es/~feli>
- Fontal, B. (2005). El Espectro Electromagnético y sus Aplicaciones. *Escuela de La Ingeniería*, 1(August), 24.
- Gallego Gómez, J. L. (2008). Análisis de series temporales. *LADE y LE*, 1(Análisis temporal), 111–128. <https://doi.org/10.3989/mc.2004.v54.i274.233>
- Gerber, F., Furrer, R., Schaepman-strub, G., Jong, R. De, & Schaepman, M. E. (2017). Predicting missing values in spatio-temporal satellite data. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1605/1605.01038.pdf>
- Ippc. (1995). IPCC Second Assessment Synthesis of Scientific-Technical Information relevant to interpreting Article 2 of the UN Framework Convention on Climate Change. *Framework*, 3(July 1992), 21. Retrieved from <http://www.ipcc.ch/pdf/climate-changes-1995/2nd-assessment-synthesis.pdf>.
- Kozak, A., & Kozak, R. (2003). Does cross validation provide additional information in the evaluation of regression models? *Canadian Journal of Forest Research*, 33(6), 976–987. <https://doi.org/10.1139/x03-022>
- Li, L., Chen, F., & Liu, J. (2010). Technology research of the ground real-time data processing for long range image and information broadcast of remote sensing satellite. *2010 2nd IITA International Conference on Geoscience and Remote Sensing, IITA-GRS 2010*, 1, 1–5. <https://doi.org/10.1109/IITA-GRS.2010.5603083>
- Lima, S. F. S. (2011). Iniciação em Sensoriamento, 225–227.
- Linke, J., McDermid, G. J., Pape, A. D., McLane, A. J., Laskin, D. N., Hall-Beyer, M., & Franklin, S. E. (2009). The influence of patch-delineation mismatches on

multi-temporal landscape pattern analysis. *Landscape Ecology*, 24(2), 157–170. <https://doi.org/10.1007/s10980-008-9290-z>

Mapadesatelite. (2017). Retrieved from http://www.mapadesatelite.com/main.php/america/mexico/MODIS1000040_md.jpg-es.html

Martinez Muñoz, J., & Díaz Ponce, A. (2005). Fundamentos de Teledetección Espacial. *Comisión Nacional Del Agua*, 62. <https://doi.org/10.1017/CBO9781107415324.004>

MODIS about. (2017). Retrieved from <https://modis.gsfc.nasa.gov/about/>

MODIS data. (2017). Retrieved from <https://modis.gsfc.nasa.gov/data/>

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., ... Running, S. W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83(1–2), 214–231. [https://doi.org/10.1016/S0034-4257\(02\)00074-3](https://doi.org/10.1016/S0034-4257(02)00074-3)

Nina, W., Cruz, R., Serrano, J., Cuba, J., Huaynacho, Y., Mamani-Aliaga, A., ... Yanyachi, P. (2015). A new approach to the massive processing of satellite images. *Proceedings - 2015 41st Latin American Computing Conference, CLEI 2015*. <https://doi.org/10.1109/CLEI.2015.7359991>

Oprisco, S., & Dumitrescu, M. (2005). On the regularization of segmented satellite images. *ISSCS 2005: International Symposium on Signals, Circuits and Systems - Proceedings*, 1, 83–86. <https://doi.org/10.1109/ISSCS.2005.1509856>

Paruelo, J. M. (2008). La caracterización funcional de ecosistemas mediante sensores remotos. *Ecosistemas*, 17(3), 4–22.

Ruiz, V., Savé, R., & Herrera, A. (2013). Análisis multitemporal del cambio de uso del suelo, en el Paisaje Terrestre Protegido Miraflores Moropotente Nicaragua, 1993 – 2011. *Ecosistemas*, 22(3), 117–123. <https://doi.org/10.7818/re.2014.22->

3.00

- Sá, I., Antonio, R., & Almorox, J. (1999). Aplicación de sensores remotos en la detección y evaluación de plagas y enfermedades en la vegetación. *Teledetección. Avances y Aplicaciones.*, 64–67. Retrieved from <http://www.aet.org.es/congresos/viii/alb16.pdf>
- Sanderson, R. (1991). Introduction to Remote Sensing Introduction to Remote Sensing, 717. <https://doi.org/10.1002/mds.23834>
- Sendra, J. B., & García, R. C. (2000). El uso de los sistemas de información geográfica en la planificación territorial. *Anales de Geografía de La Universidad Complutense*, 20(ISSN: 0211-9803), 49–67. Retrieved from <https://revistas.ucm.es/index.php/AGUC/article/viewFile/AGUC0000110049A/31281>
- Tomasi, W. (2003). *Sistemas de Comunicaciones Electrónicas*. (P. Education, Ed.) (4 Edición).
- Tucker, B. J., Maroulis, P. J., & Bandy, A. R. (1985). Brian J. Tucker, 1 Peter J. Maroulis 2 and Alan R. Bandy, 12(1), 9–11.
- Turner, M. G. (2005). Landscape Ecology: What Is the State of the Science? *Annual Review of Ecology, Evolution, and Systematics*, 36(1), 319–344. <https://doi.org/10.1146/annurev.ecolsys.36.102003.152614>
- Unninayar, S., & Olsen, L. (2008). Monitoring, Observations, and Remote Sensing – Global Dimensions. *Encyclopedia of Ecology*, 2425–2446. <https://doi.org/10.1016/B978-008045405-4.00749-7>
- Vermote, E. F., El Saleous, N. Z., & Justice, C. O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: First results. *Remote Sensing of Environment*, 83(1–2), 97–111. [https://doi.org/10.1016/S0034-4257\(02\)00089-5](https://doi.org/10.1016/S0034-4257(02)00089-5)
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F.,

Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>

Wulder, M. (1998). Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*, 22(4), 449–476. <https://doi.org/10.1191/030913398675385488>

9. ANEXOS.

9.1 ANEXO 1.

Código utilizado para la estructuración de la base de datos de imágenes de NDVI como un arreglo en 4D para RStudio.

```
#crear arreglo 4D de 10x10x4x4

a<-array(1:100,dim=c(10,10,4,4))

View(a)

Image(a)

#cargar imagen en Rstudio

setwd('C:/Gap/Imágenes/CC_eMTH_NDVI.2017.075-
084.QKM.COMPRES.005.2017090194320') #cambiar slash por slash invertido

getwd()

a<-("C:/Gap/Imágenes/CC_eMTH_NDVI.2017.075-
084.QKM.COMPRES.005.2017090194320/CC_eMTH_NDVI.2017.075-
084.QKM.VI_ACQI.005.2017090181743.TIF")

d<-raster(a)

plot(d)

[x y días años]=dim(d)

#Crear arreglo 4D de 10x10x4x4

a<-array(ndvi,dim=c(size(x),size(y),días,años))

View(a)

Image(a)
```

9.2 ANEXO 2.

Código que se empleó para determinar los valores faltantes de NDVI utilizando la plataforma RStudio.

```
## Definir función de predicción
##-----mean -----
PredictMean <- function (a, i) mean(a, na.rm = TRUE)
out1 <- Gapfill(data = ndvi, fnPredict = PredictMean)
Image(out1$fill)
## Selección de subconjunto para la estimación
args(Subset)
out2 <- Gapfill(data = ndvi, fnPredict = PredictMean,
               initialSize = c(0, 0, 1, 6))
Image(out2$fill)
## Numero de elementos minimo "nNotNA", valores diferentes a NA
## Numero de iteraciones i máximas para la estimación del valor faltante
PredictMean2 <- function (a, i, nNotNA) {
  if (sum(!is.na(a)) < nNotNA)
    return (c(NA, NA))
  c(mean(a, na.rm = TRUE), i)
}
out3 <- Gapfill(data = ndvi, fnPredict = PredictMean2, nPredict = 2,
               initialSize = c(0, 0, 1, 6), nNotNA = 0)
stopifnot(identical(c(out2$fill), c(out3$fill[,,,1])))
Image(out3$fill[,,,2]) # Numero de iteraciones i requeridas
out4 <- Gapfill(data = ndvi, fnPredict = PredictMean2, nPredict = 2,
               initialSize = c(0, 0, 1, 6), nNotNA = 50)
Image(out4$fill[,,,1]) # Valores Rellenados
Image(out4$fill[,,,2]) # Numero de iteraciones i
```

9.3 ANEXO 3.

Código utilizado para cálculo del error en las estimaciones de los valores faltantes.

```
clear all
close all
clc

%algoritmo para la estimación del error promedio en valores estimados

%selección de archivo de imagen original
[archivoO,rutaO]=uigetfile('*.tiff','Choose tiff file');
if archivoT==0
    return
else
end

%selección de archivo de imagen con valores estimados
[archivoE,rutaE]=uigetfile('*.tiff','Choose tiff file');
if archivoT==0
    return
else
end

archiN=strcat(rutaE,archivoE);
archiE=strcat(rutaO,archivoO);
%leer imagen
ImE=imread(archiN);
[x,y]=size(Im); %x filas y columnas

ImO=imread(archiE);

%Estimación de error absoluto
Ea=zeros(x,y);

for i=1:x
    for j=1:y
        Ea(i,j)=abs(ImE(i,j)-ImO(i,j));
    end
end

%estimación del error relativo

Er=Ea./ImO;

Figure 1
Image (Ea)
```

Figure 2
Image (Er)

9.4 ANEXO 4.

Código utilizado para el análisis de sensibilidad del algoritmo.

```
clear all
close all
clc

%algoritmo para la estimación del error promedio en valores estimados

%selección de archivo de imagen original
[archivo0,ruta0]=uigetfile('*.tiff','Choose tiff file');
if archivoT==0
    return
else
end

archiE=strcat(ruta0,archivo0);
%leer imagen
ImE=imread(archiN);
[x,y]=size(Im); %x filas y columnas

ImO=imread(archivE);

%Reemplazo de una matriz de nxn en determinada zona de la imagen por
valores Nan
n=10; %dimensiones por default de vecino más cercano
E=nan(n,n);

%ubicación de donde se realizará el reemplazo de los valores conocidos
%centro, 2/3,1/4 1/8 de la esquina de la imagen
px=1/2;
py=1/2;

%si la imagen es cuadrada posx=psoy
posx=abs(x*px);
posy=abs(y*py);

%reemplazo de valores en la imagen original
for i=posx:x
    for j=posy:y
        Im(i,j)='Nan';
    end
end

%visualización de imagen final
Image(Im)
```

9.5 ANEXO 5.

Caracterización de campos de precipitación intensa con ayuda de imágenes de radar en Francia.

A continuación se presenta una breve descripción del proyecto de realizado durante la estancia de investigación en el “Institut national de recherche en sciences et technologies pour l’environnement et l’agriculture” en Lyon, Francia derivado de los conocimientos adquiridos con el desarrollo de esta de tesis.

En los siguientes párrafos se presenta el desarrollo de una herramienta computacional para el procesamiento de información proveniente de imágenes de radar para la caracterización de campos de lluvia intensos.

Con la información obtenida podremos caracterizar la distribución espacial y frecuencia de los campos de precipitación intensa para cualquier región de interés en Francia con la posibilidad de adaptar el algoritmo para su uso y aplicación en México.

Para la realización de la tarea descrita anteriormente se realizó el desarrollo de un algoritmo en Matlab que permitiese trabajar con los archivos de Imágenes de radar proporcionadas por Météo-France gracias al proyecto PANTHERE (Projet ARAMIS Nouvelles Technologies en Hydrométéorologie Extensión et Renouvellement).

El algoritmo desarrollado se dividió a grandes rasgos en cuatro etapas:

- Lectura de datos
- Visualización
- Procesamiento
- Filtrado de resultados

Para la primera etapa del desarrollo de esta herramienta fue necesario realizar una rutina para descomprimir y dar lectura a archivos de tipo binario de 16 bits los cuales contenían grupos de imágenes a cada 15 minutos para cada uno de los días que conformaban nuestra base de datos desde el 2006 al 2016, cada uno de estos

archivos se encontraba comprimido pues de no ser así la base de datos en su totalidad ocuparía 3.1 terabytes y aproximadamente 1 giga por archivo.

El segundo paso consistió en la lectura de datos y visualización en pantalla de cada una estas imágenes como parte de una primera inspección para ubicar zonas con mayor concentración de precipitación, así como la revisión del promedio de precipitación por imagen, lo cual nos daría una idea de en cuál de estas se registró una mayor cantidad de precipitación promedio Figura 9-1.

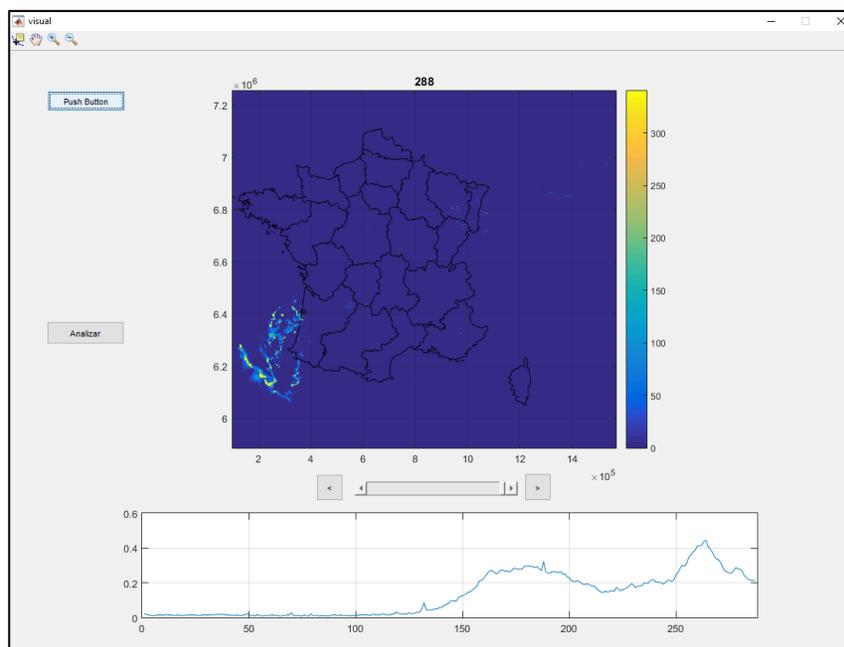


Figura 9-1 Ventana para visualización de imágenes de radar (Fuente: Elaboración propia).

Parte de la programación de esta segunda etapa consistió en un filtro para eliminar pixeles aislados espacialmente, así como valores presentes una sola ocasión dentro de la serie de tiempo es decir no recurrentes.

La tercera etapa y la más importante consistió en primer lugar en la utilización de un filtro para extraer un mapa con los valores de píxel mayores a un umbral seleccionado, en nuestro caso todos aquellos pixeles con valores mayores a 0.2 mm de precipitación resultando en una imagen con los campos de precipitación intensa Figura 9-2.

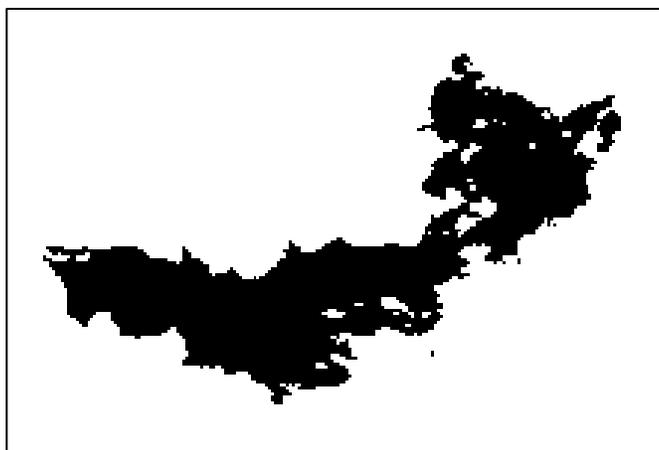


Figura 9-2 Campos de Precipitación intensa (Fuente: Elaboración propia).

Un segundo paso en el procesamiento de la información consistió en la obtención de un raster de frecuencias Figura 9-3 que nos permitiese identificar zonas de riesgo en función de la frecuencia con la que se presentó un valor de píxel por encima de nuestro umbral previamente definido.

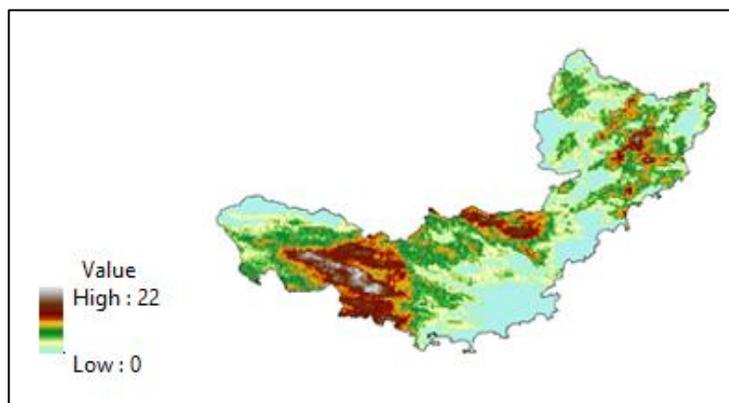


Figura 9-3 Mapa en formato raster de frecuencias (Fuente: Elaboración propia).

El último resultado obtenido de esta herramienta es un mapa con los valores de precipitación Figura 9-4 acumulados en función del intervalo analizado es decir 15 minutos, 30, 45, etc.

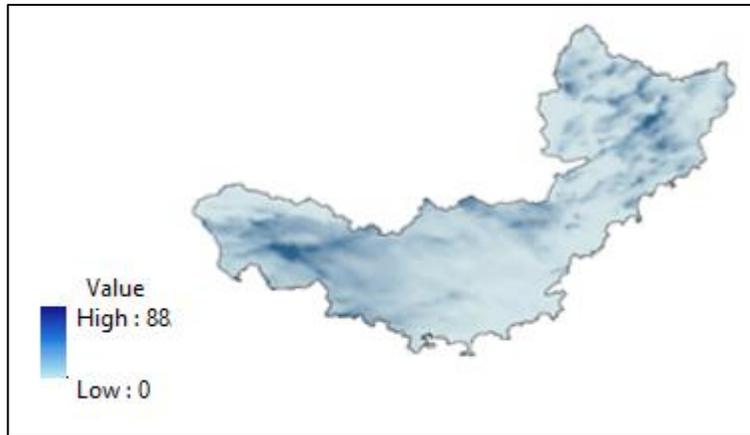


Figura 9-4 Raster con valores de precipitación (Fuente: Elaboración propia).

Es importante mencionar que esta herramienta te permite hacer los cálculos para toda Francia o filtrar ingresando un formato shape y así delimitar nuestra zona de interés Figura 9-5.

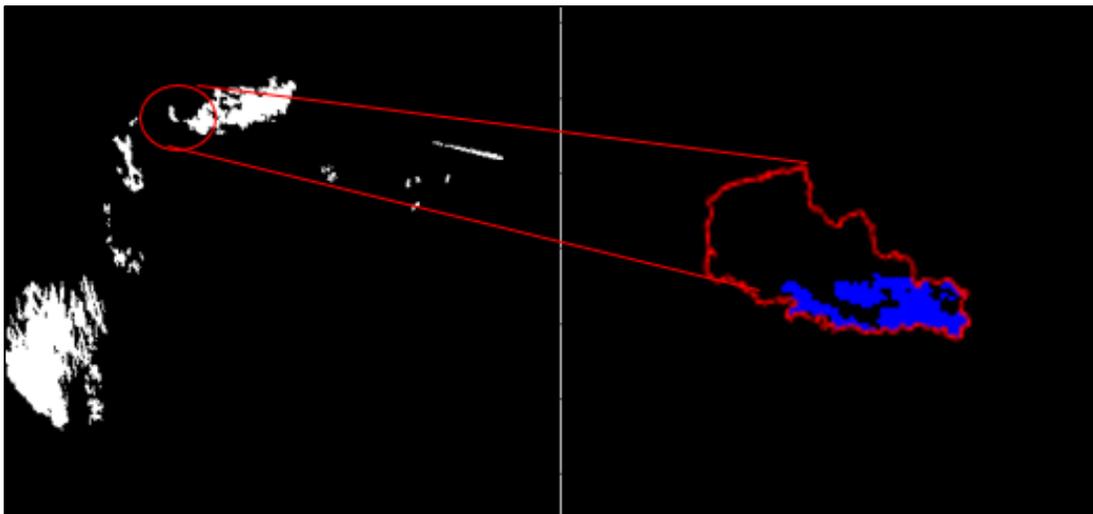


Figura 9-5 Extracción de información dentro del área de un archivo shape (Fuente: Elaboración propia).

Cada uno de los productos resultantes es guardado en un archivo independiente en formato ASCII compatible con la mayoría de los Sistemas de Información Geográfica de uso cotidiano como lo es ArcGis o Qgis.

Los resultados de esta investigación nos permitirán la caracterización de campos de precipitación intensos y de esta manera identificar zonas de riesgo que nos permitan destinar recursos para prevención de manera efectiva.

El estudio de los campos de precipitación intensa ha cobrado una gran importancia en los últimos años debido a su estrecha relación con la generación de flujos súbitos presentes en cuencas con pendientes muy pronunciadas, altamente urbanizadas y con cortos tiempos de concentración.

En este caso nuestro objeto de interés fueron las nubes sin embargo si por alguna razón como una mala calibración existiese la ausencia de valores es posible utilizar el algoritmo gapfill para la estimación de valores faltantes dentro de estos llamados campos de precipitación intensa sin embargo se debe tener precaución con la componente espacial del algoritmo pues deberá utilizarse imágenes con una resolución temporal de por lo menos 15 minutos y así evitar alterar los resultados por el movimiento de los campos de precipitación en periodos mayores al intervalo mencionado.

9.5.1 Código de la herramienta en Matlab.

```
function visual_OpeningFcn(hObject, eventdata, handles, varargin)
clc
clear l Base title
cla

set(handles.slider1, 'min', 0);
set(handles.slider1, 'max', 288);

set(handles.slider1, 'SliderStep', [1/288 , (2*1/288) ]);

axes(handles.axes1)
axis off
axes(handles.axes2)
axis off
% handles    structure with handles and user data (see GUIDATA)
global BD l S
set(handles.figure1, 'pointer', 'watch')
S = shaperead('C:\irstea\shapes\Francia');

[archivoT,ruta]=uigetfile('*.zip','Choose zip file');
if archivoT==0
    return
else
```

```

end

arcun=strcat(ruta,archivoT);
unzip(arcun)
separ=strsplit(archivoT, '.');
arcbin=strcat(separ(1), '.bin');
A=char(arcbin);
%para abrir el archivo utilizamos la siguiente rutina
archivo=fopen(A, 'r');
da=fread(archivo, '*ubit16');

BD=zeros(1351,1452,288);
l=0;
for l=0:287
    %Cargar una imagen con datos
    x=((1452*1351)*l)+1;
    y=(1452*1351)*(l+1);
    dda=da(x:y);

    %Reconstrucción de la imagen de radar
    Matr=zeros(1351,1452);
    z=0;

    for i=1:1351
        Ind=1452*(z)+1;
        for j=1:1452
            if dda(Ind)~=65535
                Matr(i,j)=dda(Ind);
                BD(i,j,l+1)=dda(Ind);
                Ind=Ind+1;
            else
                Matr(i,j)=0;
                BD(i,j,l+1)=0;
                Ind=Ind+1;
            end
        end
        z=z+1;
    end

end

axes(handles.axes1)
X=[98433+(1013.37/2), ((98433+(1452*1013.37))-(1013.37/2))];
Y=[5885335+(1013.37/2), ((5885335+(1351*1013.37))-(1013.37/2))];

image(X,Y,Matr)
title(l+1)
set(gca, 'YDir', 'normal');

mapshow(S, 'LineStyle', '-', 'FaceColor', 'None');
colorbar
l=l+1;
set(handles.slider1, 'Value', l); % Somewhere between max and min.
set(handles.slider1, 'Visible', 'on')
set(handles.pushbutton2, 'Visible', 'on')
set(handles.pushbutton3, 'Visible', 'on')
set(handles.pushbutton4, 'Visible', 'on')

```

```

grid on
for i=1:287

XXX(i)=mean(mean(BD(:,:,i)));
YYY=linspace(1,287,287);
end

axes(handles.axes2)
plot(YYY,XXX)
xlim([0 288])
grid on
set(handles.figure1, 'pointer', 'arrow')

image(X,Y,BD(:,:,l));
title(l)
colorbar
set(gca, 'YDir', 'normal');
mapshow(S, 'LineStyle', '-', 'FaceColor', 'None');

l=l+1;
image(X,Y,BD(:,:,l));
title(l)
colorbar
set(gca, 'YDir', 'normal');
mapshow(S, 'LineStyle', '-', 'FaceColor', 'None');

global BD l S
clear axes1
axes(handles.axes1)
sliderMin = get(handles.slider1, 'Min');
sliderMax = get(handles.slider1, 'Max');
sliderValue = round(get(handles.slider1, 'Value'));

image(X,Y,BD(:,:,sliderValue))
title(sliderValue)
set(gca, 'YDir', 'normal');
colorbar
l=sliderValue;
%mapshow(Z, 'FaceColor', 'black');
mapshow(S, 'LineStyle', '-', 'FaceColor', 'None');

```