

**DESARROLLO DE UN FRAMEWORK PARA LA CONSTRUCCIÓN DE
MODELOS PREDICTIVOS PARA LA CORRECCIÓN DEL ERROR
IONOSFÉRICO EN SISTEMAS GNSS**

**DIEGO SEBASTIÁN ROMÁN BUENO
LAURA CAROLINA PORRAS SUÁREZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2016

**DESARROLLO DE UN FRAMEWORK PARA LA CONSTRUCCIÓN DE
MODELOS PREDICTIVOS PARA LA CORRECCIÓN DEL ERROR
IONOSFÉRICO EN SISTEMAS GNSS**

**DIEGO SEBASTIÁN ROMÁN BUENO
LAURA CAROLINA PORRAS SUÁREZ**

Trabajo de grado para optar al título de Ingeniero de Sistemas

**DIRECTOR: RAÚL RAMOS POLLÁN
DOCTOR EN INGENIERÍA INFORMÁTICA**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2016

DEDICATORIA

Dedico este trabajo a Dios y mi familia. A mi padre por ser el ejemplo de ahínco, entrega y dedicación cada día de mi vida.

A mi madre por su paciencia, apoyo incondicional y sabiduría en cada momento que lo necesito, y a lo largo de estos años.

A mis hermanos Mónica y Juan David, por apoyarme en mis proyectos y metas, y por darme motivos para darles el mejor ejemplo.

A Diego, mi novio, amigo y compañero de carrera por su apoyo, alegría, fortaleza y ser mi complemento en cada momento.

Laura Carolina Porras.

DEDICATORIA

Agradezco a mi madre Gladys por su entrega, amor, paciencia y por impulsarme a ser una mejor persona día a día.

A mi padre Julio Cesar por su sacrificio, sabiduría, por enseñarme a ser un hombre correcto y responsable.

A mis tías Luz Marina y Azucena por su dedicación, guía y apoyo para lograr cumplir mis metas.

A Laura Carolina una persona con grandes valores, que me acompaña en este largo camino, me motiva a continuar con nuevos proyectos y ser esa espectacular compañera, amiga y novia, agradezco su esfuerzo.

A todas las personas que me apoyaron en momentos difíciles desde el inicio de este camino, especialmente a mis compañeros de la Universidad Industrial de Santander.

Diego Sebastián Román Bueno.

AGRADECIMIENTOS

Al nuestro director del proyecto Raúl Ramos por su dedicación, guía, paciencia y enseñanza con el conocimiento necesario para aprovechar esta oportunidad de aprender.

Al grupo de Supercomputación y Cálculo Científico por brindarnos las herramientas necesarias para culminación de nuestro proceso de formación.

Al Ingeniero Duván Sanabria por permitirnos aprender y vivir una experiencia diferente en el diplomado en la línea de Investigación en Nuevas Tecnologías en Visión y Visualización Computacional y Servicio de Cómputo , gracias por su conocimiento y ser guía en el proceso.

CONTENIDO

	Pág.
INTRODUCCIÓN	18
1. PLANTEAMIENTO Y JUSTIFICACION DEL PROBLEMA	20
2. OBJETIVOS.....	21
2.1 OBJETIVO GENERAL	21
2.2. OBJETIVOS ESPECÍFICOS.....	21
3. MARCO DE REFERENCIA.....	22
3.1. SISTEMAS DE NAVEGACIÓN GLOBAL POR SATÉLITE	22
3.1.1 Principios de posicionamiento.....	22
3.1.2 Fuentes de error.	24
3.1.3 Rinex.:.....	27
3.1.4 IGS (International GNSS Service).....	31
3.2 MODELO IONOSFÉRICO	33
3.2.1 Caracterización de la ionósfera.....	33
3.2.2 TEC (Total Electron Content):.....	35
3.2.3 Mapas TEC.....	36
3.2.4 Modelado..	37
3.2.5 Esquema de una sola capa.....	39
3.3 MACHINE LEARNING	40
3.3.1 Flujo de trabajo de <i>machine learning</i> . o.	41
3.3.2 Dataset.....	42
3.3.3 Scikit-Learn.....	42
3.3.4 Sklearn pipeline..	42

3.4 FRAMEWORK	43
4. METODOLOGÍA	44
4.1 ESTRATEGIA GENERAL	44
4.2 ORGANIZACIÓN DEL PROYECTO	44
4.2.1 Bloque de Trabajo [BT] 1: Módulo Obtención de datos	44
4.2.2. BT2: Módulo Preprocesado de Datos	45
4.2.3. BT3: Módulo de generación de modelos predictivos	45
5. DESARROLLO DEL PROYECTO	47
5.1 ADQUISICIÓN DE DATOS DE ESTACIONES.....	48
5.1.1 Identificación y selección de herramientas necesarias para la construcción del framework.	48
5.1.2 Descarga de ficheros de observación y navegación.....	48
5.2 PROCESAMIENTO DE DATOS PARA LA CONSTRUCCION DE DATASET.....	51
5.3 CONSTRUCCION DE MAPAS TEC CON DIFERENTES ALGORITMOS DE INTERPOLACION.....	57
5.3.1 Implementación de mapas TEC.....	57
5.3.2. Comparación de diferentes mapas TEC.....	66
5.4 MÓDULO PREDICTIVO DEL ERROR IONOSFÉRICO.....	70
5.4.1 Conjunto de Entrada.....	71
5.4.2 Formato de Especificación.....	74
5.4.3 Entrenamiento y prueba del modelo.	74
5.4.4 Experimentos con algoritmos de <i>machine learning</i>	74
5.4.5 Evaluación de modelos.....	84
6. CONCLUSIONES	89
7. RECOMENDACIONES.....	91

REFERENCIAS BIBLIOGRÁFICAS.....92

BIBLIOGRAFÍA.....98

LISTA DE FIGURAS

	Pág.
Figura 1. Funcionamiento del posicionamiento global.	23
Figura 2. Origen de error para GPS.....	25
Figura 3. Ejemplo de un fichero de navegación.....	28
Figura 4. Ejemplo de un fichero de observación.....	29
Figura 5. Estaciones de referencia de todo el mundo.....	32
Figura 6. Contenido total de electrones.	35
Figura 7. Mapa TEC para el 3 de Junio del 2.015 a las 2:05 UTC.....	37
Figura 8. Modelo Klobuchar.....	38
Figura 9. Modelo NeQuick.	39
Figura 10. Modelo de una sola capa de la ionósfera.	40
Figura 11. Flujo de trabajo para el módulo de <i>machine learning</i>	41
Figura 12. Flujo de trabajo del proyecto.....	47
Figura 13. Tabla de estaciones de la red IGS.....	49
Figura 14. Estaciones donde no existen de ficheros de navegación.	50
Figura 15. Página del servidor FTP del IGS.....	51
Figura 16. Estaciones sin fichero de observación para una hora determinada.....	53
Figura 17. Mensaje de confirmación, los datos han sido procesados.....	56
Figura 18. Extracto del dataset para el día 061 de año 2016.	56
Figura 19. Diagrama del flujo de trabajo para el procesamiento de ficheros GNSS.....	57
Figura 20. Imagen de fondo de cada mapa TEC.	58
Figura 21. Mapa TEC realizado con la función RBF.	59
Figura 22. Mapa TEC realizado con la función Griddata.	60
Figura 23. Mapa TEC realizado con la función de interpolación Griddata y la función de graficas countour.....	61
Figura 24. Mapa TEC realizado con la función de interpolación Bisplrep.	62

Figura 25. Mapa TEC de la zona Ecuatorial.	63
Figura 26. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 2:00 UTC.	64
Figura 27. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 8:00 UTC.	64
Figura 28. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 14:00 UTC.	64
Figura 29. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 20:00 UTC.	65
Figura 30. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 23:00 UTC.	65
Figura 31. Mapa TEC del círculo polar Ártico del planeta (de la zona templada del norte).	66
Figura 32. Mapa TEC de la región de Suramérica.	66
Figura 33. Mapa TEC de Caltech, NASA (Jet propulsión Laboratory).	67
Figura 34. Mapa TEC de oficina de meteorología del Gobierno de Australia.	68
Figura 35. Mapa TEC del Instituto nacional de geofísica, geodesia y geografía de Bulgaria.	69
Figura 36. Mapa TEC desarrollado con el módulo.	70
Figura 37. Extracto de conjunto de entrada para la predicción del día <i>test</i> , 051 del año 201.	72
Figura 38. Comportamiento de la ionósfera para zona tropical con las estaciones AREQ y SALU.	73
Figura 39. Comportamiento de la ionósfera para zona templada con las estaciones QUIN y GOLD	73
Figura 40. Pasos para el algoritmo <i>linearregression</i> de <i>Scikit learn</i>	75
Figura 41. Predicción con <i>linearregression</i> para la estación SOLO de Islas Salomón en Oceanía.	76
Figura 42. Predicción con <i>linearregression</i> para la estación ADIS.	76
Figura 43. Predicción con <i>linearregression</i> para la estación ANKR.	77

Figura 44. Pasos para el algoritmo <i>DecisionTreeRegressor</i> de <i>scikit learn</i>	78
Figura 45. Predicción con <i>DecisionTreeRegressor</i> para la estación SOLO de Islas Salomón en Oceanía.	78
Figura 46. Predicción con <i>DecisionTreeRegressor</i> para la estación ADIS.	79
Figura 47. Predicción con <i>DecisionTreeRegressor</i> para la estación ANKR.....	79
Figura 48. Pasos para el algoritmo <i>Support Vector Regression</i> de <i>scikit learn</i>	80
Figura 49. Predicción con <i>Support Vector Regression</i> para la estación SOLO de Islas Salomón en Oceanía.	81
Figura 50. Predicción con <i>Support Vector Regression</i> para la estación ADIS.	81
Figura 51. Predicción con <i>Support Vector Regression</i> para la estación ANKR.	82
Figura 52. Pasos para el algoritmo <i>Random Forest Regressor</i> de <i>scikit learn</i>	82
Figura 53. Predicción con <i>Random Forest Regressor</i> para la estación SOLO, Islas Salomón en Oceanía.	83
Figura 54. Predicción con <i>Random Forest Regressor</i> para la estación ADIS.....	84
Figura 55. Predicción con <i>Random Forest Regressor</i> para la estación ANKR.	84
Figura 56. Métricas de regresión para estación SOLO de Islas Salomón en Oceanía.	85
Figura 57. Métricas de regresión para estación ASPA de Pago Pago en Estados Unidos.....	86
Figura 58. Cálculo del RMSE con conjunto de datos Train y test para estaciones de la zona tropical.....	87
Figura 59. Cálculo del RMSE con conjunto de datos Train y test para estaciones de la zona templada.....	87

RESUMEN

TÍTULO: DESARROLLO DE UN FRAMEWORK PARA LA CONSTRUCCIÓN DE MODELOS PREDICTIVOS PARA LA CORRECCIÓN DEL ERROR IONOSFÉRICO EN SISTEMAS GNSS.

AUTORES: PORRAS SUAREZ, Laura Carolina. ROMAN BUENO, Diego Sebastián. **

PALABRAS CLAVE: GNSS, VTEC, TEC Maps, machine learning, Framework.

DESCRIPCIÓN:

Los sistemas GNSS se refieren a las constelaciones GPS, GLONASS, Beidu y Galileo, cada una integrada por un conjunto de satélites que transmiten señales recibidas por antenas en la tierra cuyo tiempo de transmisión es utilizado para estimar la distancia al satélite y, de esta forma, calcular la posición por trilateración.

Pero las señales se ven afectadas por distintas causas durante su recorrido e inducen errores en el posicionamiento obtenido. Entre el 30% y 70% del error típico en posicionamiento viene dado por el retardo de las señales causado por la ionización de la radiación solar en la atmósfera superior de la tierra generando electrones libres que modifican el grado de ionización de la ionósfera y perturban las señales emitidas por los satélites. El 99% del retardo ionosférico puede ser eliminado por un receptor de doble frecuencia, usado para calcular el contenido total de electrones TEC que se encuentran desde el satélite hasta el receptor. En cambio los receptores de una sola frecuencia de consumo masivo no tienen esta posibilidad, debiendo utilizar modelos que predicen el TEC y estiman su retardo producido en la señal reduciendo un 50% el error.

Este trabajo tiene dos objetivos: (1) construir mapas globales TEC usando redes de receptores GNSS de doble frecuencia; y (2) producir modelos basados en técnicas de machine learning para predecir el VTEC para su uso con receptores de una frecuencia.

Los mapas TEC obtenidos muestran una gran similitud con los generados a nivel global por los servicios de la NASA1, SWS2 y NIGGG3; y las predicciones de nuestros modelos aproximan el valor real reduciendo el error usando modelos locales. De esta manera, el framework facilita el estudio y predicción del comportamiento de la ionósfera especialmente en las zonas tropicales en donde la ionósfera presenta mayor actividad y comportamiento más desconocido.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas. Director: Raúl Ramos Pollán.

ABSTRACT

TITLE: DEVELOPING OF A FRAMEWORK FOR THE CONSTRUCTION OF PREDICTIVE MODELS TO CORRECT THE IONOSPHERIC ERROR IN GNSS SYSTEMS. *

AUTHORS: PORRAS SUAREZ, Laura Carolina. ROMAN BUENO, Diego Sebastián. **

KEYWORDS: GNSS, VTEC, TEC Maps, machine learning, Framework.

DESCRIPTION:

GNSS systems refers to the GPS constellations, GLONASS, Beidou and Galile, integrated by a set of satellites that transmits signals received by ground antennas, which time of transmission is used to estimate the distance to every satellite and in this way, calculate the position of the antenna by trilateration.

But signs are affected by different causes during their circuit inducing errors in the positioning obtained. Between 30% and 70% of the typical positioning error is caused by the delay of signs caused by the ionization of solar radiation in the upper atmosphere of land, which generates free electrons that modify the ionosphere's degree of ionization and disturb signs from satellites. The 99% of ionospheric delay can be eliminated by a dual frequency receptor that calculates the total electrons content TEC, which are from satellite to the receptor. By contrast, receptors of an only frequency of massive consumption do not have this possibility. Therefore, they must use ionospheric models, which predict TEC and estimate the signal delay reducing 50% of error.

This work aims to (1) build TEC global maps using network of GNSS dual frequency receptors, as well as to (2) produce models based on techniques of machine learning to predict TEC for its use with receptors of one frequency.

TEC maps obtained show great similarities with maps generated by NASA1, SWS2 and NIGGG3. TEC predictions of our models approaches real TEC, reducing the error by using local models. Thus, the Framework facilitates the study and prediction of ionosphere, which is especially important in Tropical areas where ionosphere has higher activity and its behavior is more unknown.

* Bachelor Thesis

** Faculty of Physico-Mechanical Engineering. School of Systems Engineering. Director Raúl Ramos

INTRODUCCIÓN

Los sistemas de navegación globales satelital se han convertido en una herramienta muy útil en los últimos años ya que nos permite conocer la ubicación de cualquier objeto que cuente con un sistema GPS, como automóviles o teléfonos inteligentes. Las señales transmitidas por cualquier satélite GPS atraviesan la ionósfera, una capa de la atmósfera que tiene 430 kilómetros de altitud media y donde se generan procesos de ionización que originan grandes cantidades de electrones, estos cambian según su densidad y dependiendo de la latitud y longitud ubicada además del tiempo en el que sea medido. Este cambio de ionización constante no permite crear un modelo que corrija el error generado por la ionósfera al calcular la posición de un dispositivo GPS. La necesidad de realizar la corrección de este error se hace imprescindible ya que permitirá a los usuarios de cualquier sistema de navegación obtener una mayor precisión en la posición de su dispositivo móvil con GPS.

En este proyecto procesamos los datos mediante GPSTK, una herramienta que nos facilita el procesamiento de los ficheros RINEX de navegación y observación de cada estación para luego realizar el cálculo de posición el IPP (del inglés, ionospheric pierce point) y el VTEC (del inglés, vertical total electron content). De esta manera generamos mapas TEC de cualquier región de la tierra y predecimos el valor del VTEC mediante los algoritmos: Linear Regression, Support Vector Regression (SVR), Decision Tree Regressor y Random Forest de la librería scikit learn de Python, con los datos seleccionados para train y test.

El alcance de este proyecto viene dado por el número de estaciones que se utilicen para la adquisición y procesado de datos, nuestro framework se adapta a las estaciones que encuentren disponibles para obtener mayor desempeño al

crear mapa TEC y predecir el valor del VTEC.

Este framework facilitará a investigadores que trabajen este tema para regiones tropicales, donde la intensidad solar es mayor generando una dificultad para crear modelos ionosféricos y también cuando los receptores de la señal no tengan doble frecuencia debido a su costo.

1. PLANTEAMIENTO Y JUSTIFICACION DEL PROBLEMA

¿Cómo se podría predecir la corrección del error ionosférico en zonas tropicales para sistemas GNSS?

Los sistemas de posicionamiento y localización global GNSS permiten determinar las coordenadas geográficas por medio de señales, estas son vulnerables debido a las irregularidades en la densidad de los electrones presentes en la ionósfera, produciendo un error que disminuye la precisión en estas señales. En este momento existen los modelos de corrección *Klobuchar*[1] y *NeQuick*[2] adoptados por GPS y Galileo respectivamente, y cuyas correcciones se transmiten en los mensajes de los satélites para su disponibilidad en los receptores. Estos modelos están tradicionalmente calibrados con datos del hemisferio norte, con lo cual el nivel de corrección que ofrecen en zonas tropicales es limitado.

La ionósfera en las regiones tropicales es particularmente muy activa por ello es más difícil el diseño de modelos ionosféricos para sistemas GNSS debido a una mayor interacción de electrones por la radiación de luz solar, con este propósito analizaremos la funcionalidad de cada algoritmo predictivo buscando un desempeño adecuado en este proceso.

Este trabajo pretende generar un framework que facilita la investigación y la creación de modelos ionosféricos integrando datos de observación y generando mapas de contenido de electrones de la ionósfera. Esto permitirá conocer mejor el comportamiento de la ionósfera y permitirá la experimentación con modelos predictivos del VTEC, especialmente para zonas tropicales, ahorrando tiempo y esfuerzo a investigadores en el área

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Construcción de un framework que implementa un flujo de trabajo para la adquisición, pre-procesado, entrenamiento y evaluación de distintos modelos predictivos para la corrección del error ionosférico para sistemas GNSS.

2.2. OBJETIVOS ESPECÍFICOS

- Desarrollar un módulo para la construcción de datasets para el modelado del error ionosférico a partir de los datos crudos disponibles en la red IGS (International GNSS Service).
- Desarrollar un módulo que, a partir de los datasets creados en el objetivo anterior, genere mapas de contenido TEC (Total Electron Content) de la ionósfera, usando distintos algoritmos de interpolación.
- Diseñar un formato para la especificación de experimentos de modelado y predicción de la ionósfera.
- Desarrollar un módulo que realice los experimentos definidos según el formato anterior. Esto incluye la integración de los datos obtenidos en los experimentos anteriores con el API de predicción y evaluación de modelos predictivos de *Scikit learn*.
- Evaluar el desempeño de distintos algoritmos predictivos implementados en *Scikit learn* y comparar su funcionamiento en la corrección del error ionosférico

3. MARCO DE REFERENCIA

3.1. SISTEMAS DE NAVEGACIÓN GLOBAL POR SATÉLITE

Un sistema global de navegación por satélite (GNSS) es una constelación de satélites que transmite señales utilizadas para el posicionamiento y localización en cualquier parte del globo terrestre, ya sea en tierra, mar o aire. Estos permiten determinar las coordenadas geográficas y la altitud de un punto dado como resultado de la recepción de señales provenientes de constelaciones de satélites artificiales de la Tierra para fines de navegación, transporte, geodésicos, hidrográficos, agrícolas, y otras actividades afines.

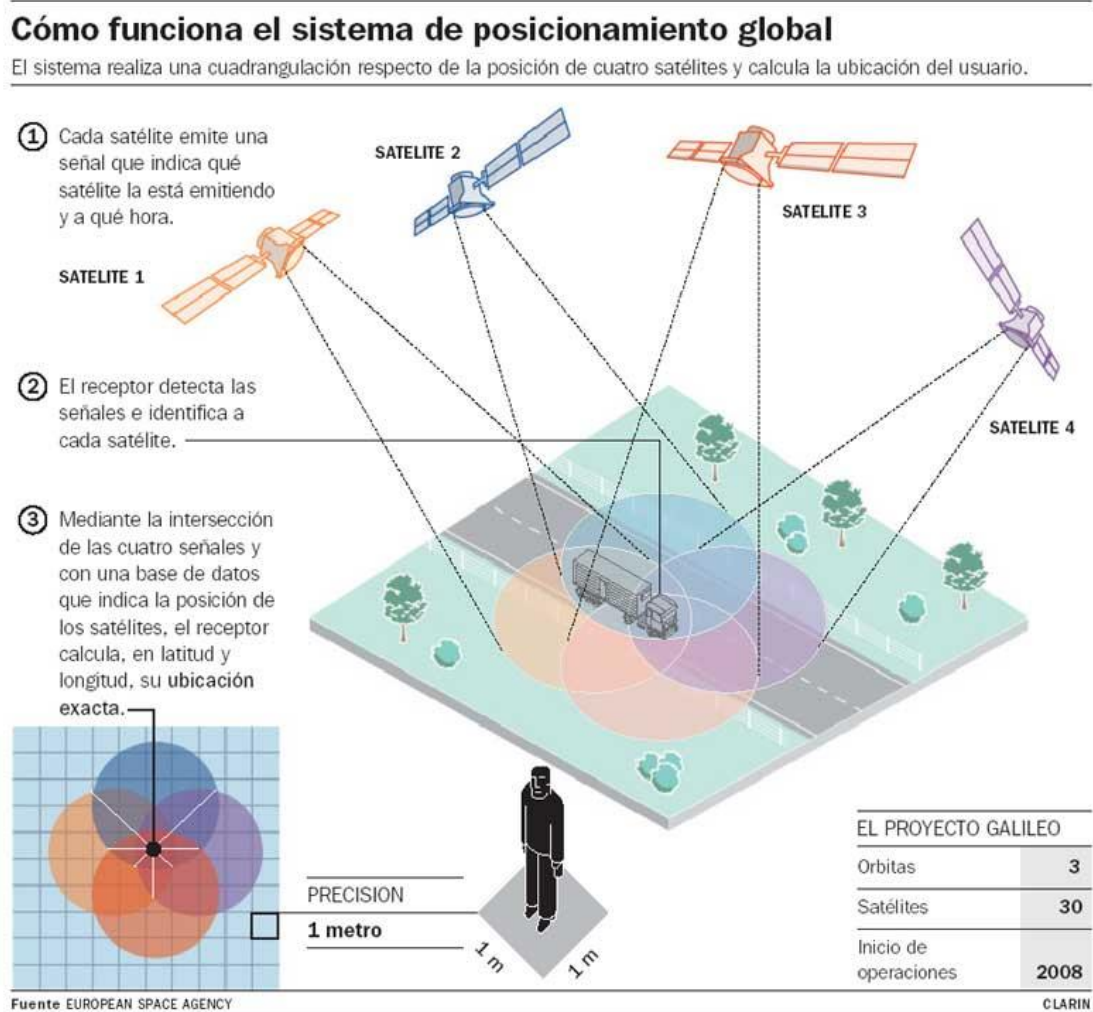
Un sistema GNSS puede proporcionar a los usuarios información sobre la posición y la hora (cuatro dimensiones) con distintos grados de precisión, en cualquier parte del mundo, las 24 horas del día y en todas las condiciones climatológicas. [3] Cada sistema GNSS cuenta con satélites no geoestacionarios, diseñados para que por medio de sus orbitas y el número de satélites se pueda ver al menos cuatro satélites desde cualquier punto de la Tierra. [4]

3.1.1 Principios de posicionamiento

3.1.1.1 Triangulación: Principio elemental del funcionamiento del sistema GPS utilizando diferentes satélites como punto de referencia para determinar nuestra posición en la superficie de la Tierra obteniendo la distancia precisa de al menos tres satélites de la constelación para obtener la posición en el espacio. El receptor

calcula el tiempo que le ha tardado en llegar la señal al satélite, se calcula la distancia que es igual a la velocidad por el tiempo que se tarda, luego se halla una ecuación de una esfera alrededor de cada satélite, la intersección de todas ellas determina la posición real del receptor. El receptor calcula la distancia por medio de las órbitas del satélite, cuanto mayor sea la cantidad de satélites será mayor la precisión de la distancia. [5]

Figura 1. Funcionamiento del posicionamiento global.



Fuente: [6]

3.1.1.2 Medición del tiempo: La medición del tiempo de la señal no es la adecuada debido a que los relojes de los satélites (con precisión de nanosegundos) utilizados no están sincronizados con el dispositivo receptor GPS. Para obtener la precisión requerida es necesario un satélite más para el cálculo de la distancia exacta. Por tanto se necesitan un mínimo de cuatro satélites visibles para el posicionamiento.

Existe una secuencia que permite estimar el tiempo de recepción e información sobre la localización del satélite, cuya diferencia es hallada con el tiempo en que fue enviada la señal y el tiempo de recepción de la señal. [7]

3.1.1.3 Efectos relativistas: La teoría de la relatividad muestra que el tiempo va más despacio durante movimientos más rápidos, esto va a afectar a los satélites en órbita, ya que su velocidad alrededor de la tierra es de 3874 m/s, entonces estos relojes se atrasan respecto a los que están en la superficie del planeta. También conocemos que el tiempo corre más despacio cuanto mayor es el campo gravitatorio, por tanto para un observador en la Tierra el reloj del satélite irá más deprisa que el suyo, ya que estos se encuentran a una altitud de 20.000 km y el potencial gravitatorio es entonces más débil.

Cuando el que recibe la señal está en un potencial gravitatorio mayor que el emisor, las señales serán recibidas más lentamente que cuando fueron emitidas (y viceversa).

3.1.2 Fuentes de error. Las fuentes de error se especifican en la siguiente tabla por su respectivo potencial de error y su error típico de acuerdo a su origen:

Figura 2. Origen de error para GPS.

ORIGEN DEL ERROR	ERROR POTENCIAL	ERROR TÍPICO
Ionósfera	5.0 metros	0.4 metros
Tropósfera	0.5 metros	0.2 metros
Datos Efemérides	2.5 metros	0 metros
Reloj de Satélite	1.5 metros	0 metros
Multipath	0.6 metros	0.3 metros
Medición del ruido	0.3 metros	0.3 metros
Total	~15 metros	~10 metros

Fuente: [8]

Las fuentes del error en el sistema de navegación satelital están dadas por varias causas entre ellas:

El efecto cambiante de la tropósfera, surge cuando la señal pasa por esta zona ya que es un medio no dispersivo para las ondas de radio, además sus efectos refractivos son independientes de la frecuencia y depende de los parámetros meteorológicos y de la longitud de su recorrido.

Otra de las causas del error proviene del medio de propagación, ya que la refracción depende de la frecuencia (hora y del sitio de medición) en la ionósfera, en donde el error es estimado haciendo comparaciones en dos frecuencias diferentes. Igualmente es un medio dispersivo en donde el índice de refracción depende de la frecuencia afectando de la señales GPS que sufren retardo o adelanto. [9]

Un error considerable se da por la cercanía que existe entre las antenas y las estructuras que se reflejan, cuando una señal llega por diferentes trayectorias y ocurre una interferencia a esto se le conoce como multipath. [10]

3.1.2.1 Estrategias de mitigación. Las mitigaciones de las diferentes fuentes de error se pueden utilizar de las siguientes maneras: para la tropósfera, la componente del retardo está altamente correlacionada con las componentes de posición de la estación, por lo que un error en la obtención del retraso, degrada la posición obtenida para el receptor.

Para la ionósfera existe el modelo ionosférico Klobuchar (oficial de GPS) [1] y modelo ionosférico NeQuick (oficial de Galileo) [2].

Para multipath se conocen las frecuencias en las antenas, su corrección se realiza utilizando receptores de mapeo de grado, receptores de mapeo, receptores de ocio y de grado general para recibir correcciones diferenciales en tiempo real.

Para el modelo "Global Pressure and Temperature" GPT, permite obtener valores atmosféricos más realistas. [11]

3.1.2.2 Corrección diferencial: La corrección diferencial es un método que reduce los efectos de error atmosférico y otras fuentes de posicionamiento (referencia y receptores itinerantes), la base es una estación terrestre que calcula los errores de sincronización; El diferencial de sistema de posicionamiento global (DGPS) es el encargado de mejorar la precisión, mediante una red de estaciones de referencia en Tierra así medir la diferencia entre los satélites del GPS y las posiciones fijas, en el post-procesado se recogen los datos del receptor pero en el campo donde se recogen no están siendo corregidos tiene que ir a procesar los datos con software de corrección diferencial.

Para configurar la estación base local se debe establecer enlaces de radio y su área de cobertura. El área está limitada por la fuerza de las transmisiones de radio, además el gobierno proporciona archivos de corrección diferencial de post-procesado. Si se compra una estación diferente al software de corrección tiene que convertirse a RINEX [12].

3.1.3 Rinex. Receiver Independent Exchange (RINEX), es un formato de ficheros de texto orientado a almacenar, de manera estandarizada medidas proporcionadas por receptores de sistemas de navegación por satélite. La salida final de un receptor de navegación suele ser su posición y velocidad, estas se calculan en tiempo real. Este formato estandarizado permite la gestión y almacenamiento de las medidas generadas por un receptor, existen dos tipos de ficheros RINEX:

- Fichero de navegación: contiene información de las órbitas, los parámetros del reloj y la precisión de las medidas de pseudodistancia de los satélites observados.
- Fichero de observación: contiene la pseudodistancia, la fase y el tiempo de cada una de las observaciones realizadas.

3.1.3.1 Fichero de navegación: La siguiente figura muestra un extracto de un fichero de navegación que tiene un encabezado y la sección de datos.

Figura 3. Ejemplo de un fichero de navegación.

```

2.11          N: GPS NAV. MESSAGE          RINEX VERSION / TYPE
gLAB          gAGE / UPC                   PGM / RUN BY / DATE
EXAMPLE OF A NAVIGATION RINEX FILE       COMMENT
THIS FILE IS PART OF THE gLAB TOOL SUITE  COMMENT
FILE PREPARED BY: ADRIA ROVIRA GARCIA     COMMENT
PLEASE EMAIL ANY COMMENT OR REQUEST TO:   glab @ gage.es  COMMENT
0.1676D-07   0.2235D-07   -0.1192D-07   -0.1192D-07 ION ALPHA
0.1208D+06   0.1310D+06   -0.1310D+06   -0.1966D+06 ION BETA
0.133179128170D-06  0.107469588780D-12  552960   1025 DELTA-UTC: A0,A1,T,W
15                                               LEAP SECONDS
                                               END OF HEADER

6 99 9 2 19 0 0.0 -.839701388031D-03 -.165982783074D-10 .000000000000D+00
.910000000000D+02 .934062500000D+02 .116040547840D-08 .162092304801D+00
.484101474285D-05 .626740418375D-02 .652112066746D-05 .515365489006D+04
.409904000000D+06 -.242143869400D-07 .329237003460D+00 -.596046447754D-07
.942817490922D+00 .326593750000D+03 .206958726335D+01 -.638312302555D-08
.307155651409D-09 .000000000000D+00 .102500000000D+04 .000000000000D+00
.000000000000D+00 .000000000000D+00 .000000000000D+00 .910000000000D+02
.406800000000D+06 .000000000000D+00

13 99 9 2 19 0 0.0 .490025617182D-03 .204636307899D-11 .000000000000D+00
.133000000000D+03 -.963125000000D+02 .146970407622D-08 .292961152146D+01
-.498816370964D-05 .200239347760D-02 .928156077862D-05 .515328476143D+04
.414000000000D+06 -.279396772385D-07 .243031939942D+01 -.558793544769D-07
.986307770581D+00 .271187500000D+03 -.232757915425D+01 -.619632953057D-08
-.785747015231D-11 .000000000000D+00 .102500000000D+04 .000000000000D+00
.000000000000D+00 .000000000000D+00 .000000000000D+00 .389000000000D+03
.410400000000D+06 .000000000000D+00

```

Fuente: [13]

Este es el formato RINEX para navegación GPS, que contiene dos partes:

El encabezado donde podemos evidenciar que corresponde a la versión de RINEX 2.11 y para navegación GPS, con nombre de archivo gLAB, fecha de la creación del archivo de Marzo 17 del 2.010 a las 12:14, los parámetros de A0 a A3 de la ionósfera son de 0.1676×10^{-7} , 0.2235×10^{-7} , -0.1192×10^{-7} y -0.1129×10^{-7} y los parámetros de B0 y B3 para la ionósfera corresponden a 0.1208×10^6 , 0.1310×10^6 , -0.1192×10^6 y -0.1966×10^6 . Los parámetros de almacenamiento del tiempo del computador con respecto UTC (Tiempo Universal Controlado), son: la condición de polinomio A0 es de $0.13317912870 \times 10^{-6}$ y A1 con $0.1074695887808 \times 10^{12}$, un tiempo de referencia para UTC de 552960, con un número de semana de referencia de 1025, y un tiempo delta de 15 segundos. Y en la otra parte del formato RINEX, tenemos los datos propios donde observamos

el satélite 6 con código PRN, en el año 99, el mes 9, el día 2, a la hora 19, al minuto 0 y al segundo 0 tenía un espacio de sesgo de reloj de -0.0083970138801 segundos, una sincronización del reloj de -0.000000000165982783074 segundos, una tasa de sincronización del reloj de 0 segundos, el número de efemérides de datos de 091, el radio componente sinusal corrección de 093.40625 metros y el delta de 0.000000011604054740 radianes/segundos y con ángulo de 0.162092304801 radianes. [13]

3.1.3.2 Fichero de observación: La siguiente figura muestra un extracto de un fichero de observación que tiene un encabezado y sus datos propios.

Figura 4. Ejemplo de un fichero de observación.

```

2.11 OBSERVATION DATA M (MIXED) RINEX VERSION / TYPE
BLANK OR G = GPS, R = GLONASS, E = GALILEO, M = MIXED COMMENT
gLAB gAGE 17-MAR-10 12:14 PGM / RUN BY / DATE
EXAMPLE OF A MIXED RINEX FILE COMMENT
MRKR MARKER NAME
9080.1.34 MARKER NUMBER
gAGE UPC: Technical University of Catalonia OBSERVER / AGENCY
THIS FILE IS PART OF THE gLAB TOOL SUITE COMMENT
FILE PREPARED BY: ADRIA ROVIRA GARCIA COMMENT
PLEASE EMAIL ANY COMMENT OR REQUEST TO: glab @ gage.es COMMENT
IR2200716006 ASHTECH UZ-12 CQ00 REC # / TYPE / VERS
482 AOAD/M_T NONE ANT # / TYPE
4789028.4701 176610.0133 4195017.0310 APPROX POSITION XYZ
0.9030 0.0000 0.0000 ANTENNA: DELTA H/E/N
1 1 WAVELENGTH FACT L1/2
1 2 3 G14 G18 G19 WAVELENGTH FACT L1/2
7 L1 L2 P1 P2 C1 S1 S2 # / TYPES OF OBSERV
30.000 INTERVAL
2010 3 5 0 0 0.0000000 GPS TIME OF FIRST OBS
2010 3 5 23 59 30.0000000 GPS TIME OF LAST OBS
1 RCV CLOCK OFFS APPL
15 LEAP SECONDS
14 # OF SATELLITES
G07 815 815 815 815 815 815 815 PRN / # OF OBS
G09 246 246 246 246 246 246 246 PRN / # OF OBS
G12 687 687 687 687 687 687 687 PRN / # OF OBS
G13 762 762 762 762 762 762 762 PRN / # OF OBS
G15 454 454 454 454 454 454 454 PRN / # OF OBS
G20 599 599 599 599 599 599 599 PRN / # OF OBS
G21 636 636 636 636 636 636 636 PRN / # OF OBS
G26 210 210 210 210 210 210 210 PRN / # OF OBS
G31 874 874 874 874 874 874 874 PRN / # OF OBS
G32 457 457 457 457 457 457 457 PRN / # OF OBS
R11 907 907 907 907 907 907 907 PRN / # OF OBS
R19 348 348 348 348 348 348 348 PRN / # OF OBS
R23 936 936 936 936 936 936 936 PRN / # OF OBS
S24 198 198 198 198 198 198 198 PRN / # OF OBS
END OF HEADER
10 3 5 0 0 0.0000000 0 14G13R19G32G 7R23G31G20R11G12G26G 9G21 -0.12345
G15S24
121367582.20508 94572134.49208 23095489.677 9 23095481.949 9 23095483.463 7
42.000 40.000
134357446.85408 23095483.463 7
51.000
34357446.85408 104694102.10708 25567381.585 9 25567371.841 9 25567379.659 7
76.000 84.000
118767195.32608 91018570.22508 22600658.277 9 22600648.232 9 22227666.760 7
57.000 32.000

```

Fuente: [14].

Este es el formato RINEX para observación GPS, contiene dos partes: En el encabezado se establece la versión del formato RINEX 2.11 y el tipo de archivo RINEX en este caso fichero de observación, se muestra el nombre del programa que creó el archivo gLAB y la agencia que la creó gAGE, también la fecha de los datos obtenidos.

Se observa el nombre del marcador de la antena MRKR y el número 9080.1.34, luego en la siguiente línea se aprecia el número, tipo y versión del receptor (IR2200716006, ASHTECH UZ-12, CQ00), también se ve el número de la antena 482 y el tipo de antena, se aprecia la posición aproximada en (x, 4789028.4701 - y, 176610.0133 - z, 4195017.0310) y el factor por defecto Wavelength (1,1) L1, L2.

El número de observaciones almacenadas en el tipo de archivo RINEX es 7, con un intervalo de observación de 30 segundos, con un formato para la fecha de la primera observación, (Año, mes, día, hora, minuto, segundo, tipo de sistema) (2010-3-5-0-0-0.00000000 GPS), y con el mismo formato de la primera fecha para el de la última observación (2010-3-5-23-59-30.00000000 GPS), también se establece el número de segundos desde la transmisión de los diferentes tiempos de sistemas(15) y el número de satélites(14), los cuales son observados del archivo

El primer satélite G07 se observa que para cada una de las siete establecidas es valor observado es el mismo 815, así será para cada uno de los diferentes satélites de los 14.

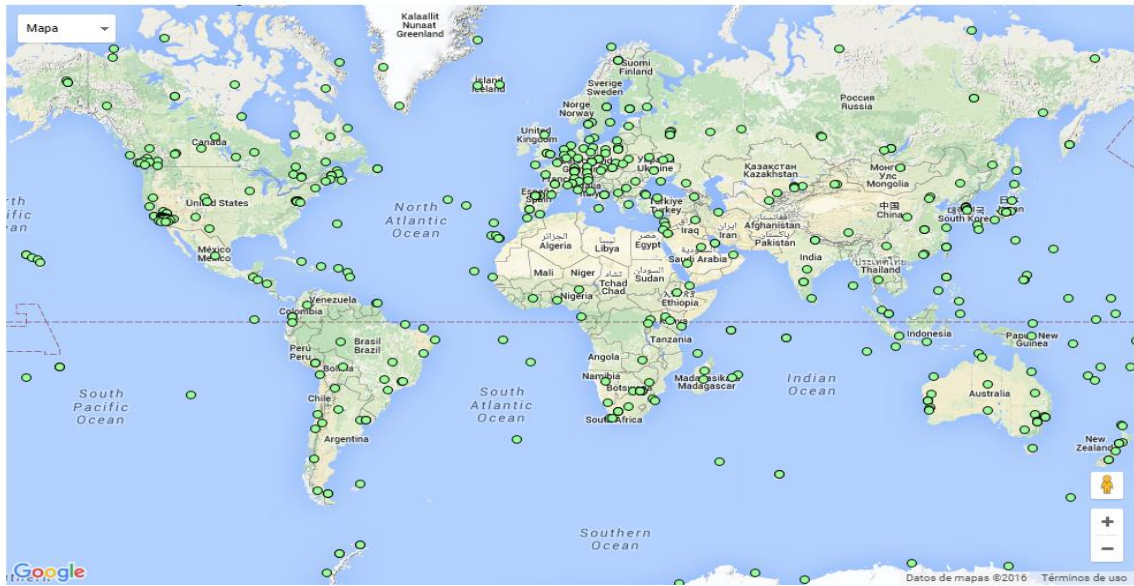
En este formato observamos los ficheros RINEX de observación para los siguientes 14 satélites: G13, R19, G32, G7, R23, G31, G20, R11, G12, G26, G9, G21, G15, S24 para el año 2010, el mes 3(Marzo), el día 5, la hora 0, el minuto 0 y el segundo 0 con -0.12345 segundos de tiempo real de receptor.

En la siguiente parte del formato RINEX, para el satélite G13 vemos los diferentes tipos de observables para el formato RINEX: para el L1 observable la fase de la portadora es de 121367582.205, la pérdida de bloqueo del indicador es 0 lo cual significa que es aceptado, la fuerza de señal del indicador es 8, una buena señal de intensidad de la señal; para el L2 observable la fase de la portadora es de 94572134.492, con 0 para la pérdida de bloqueo del indicador haciendo aceptado, la fuerza de señal del indicador es de 8; para el P1 observable, el código de pseudodistancia es 2309548.677 y con 9 para la fuerza de señal de indicador. [14]

3.1.4 IGS (International GNSS Service). El sistema global IGS comprende una red de estaciones satelitales de seguimiento, centros de datos y centros de análisis para poner los datos GPS de alta calidad y productos de datos en línea dentro de un día de observaciones.

Es un servicio internacional con más de 200 agencias en todo el mundo para formar una red global que apoya investigaciones científicas basadas en precisión y exactitud en observaciones terrestres usando tecnología de sistema de navegación por satélite mundial(GNSS).

Figura 5. Estaciones de referencia de todo el mundo



Fuente: [15].

IGS provee datos GNSS y productos en el soporte de referencia terrestre, navegación precisa, aplicación de máquinas, inspección y cartografía. Estos datos permiten la determinación de órbitas de los satélites y el seguimiento de la tropósfera y la ionósfera.

Las aproximaciones de posiciones relativas requieren la combinación mínima de dos receptores con al menos una estación, con aproximaciones, diferentes observaciones es un camino común para eliminar el error en el GPS satélite y el error en el tiempo del receptor. [16]

3.1.4.1 Productos IGS: Los tipos de los productos más representativos IGS que corresponden a efemérides del satélite GPS son:

- Difusión: Maneja una exactitud aproximadamente de 100 cm para las órbitas y para los relojes del satélite una desviación estándar de 2,5 ns. Así mismo tienen una latencia en tiempo real y un intervalo de muestra diario.
- Ultra-rápido: Posee una exactitud aproximadamente 3 cm para las órbitas, para los relojes de los satélites una desviación estándar de 50 picosegundos y el estado latente está entre 3 y 9 horas, con actualizaciones 03, 09, 15, 21 UTC al día con un intervalo de muestra diario.
- Rápido: Se observa una exactitud aproximadamente de 2,5 cm para las órbitas y hay intervalos de muestras de 15 minutos. Para los relojes del satélite existe una desviación aproximadamente de 25 picosegundos y un intervalo de muestras de 5 minutos y una latencia entre 3 y 9 horas.
- Final: Este tipo tiene una exactitud en las órbitas de 2.5 cm y un intervalo de muestras de 15 minutos. Para los relojes de los satélites tiene una desviación estándar de 20 picosegundos, el estado latente es de 12 y 18 días, y las actualizaciones cada jueves.

Para los anteriores tipos de productos la precisión de los relojes se expresan en relación a la escala de tiempo IGS, alineado con el tiempo GPS. [17]

3.2 MODELO IONOSFÉRICO

3.2.1 Caracterización de la ionósfera. La ionósfera es la parte de la atmósfera que está ionizada de radiación solar en donde existe una gran densidad libre de electrones, que se ionizan por el sol, los rayos ultravioleta y parcialmente por rayos cósmicos. Por encima de unos 60 kilómetros el número de estos electrones libres es suficiente para afectar a la propagación de las ondas electromagnéticas.

Risbeth y Garriott en 1969, dividen la ionósfera en varias capas:

Capa D: Es la capa más cercana a la Tierra con una altitud de 60 a 90 km, una presión de 2 Pascales, una temperatura de -76°C y una concentración de electrones de $10^7 - 10^{10} \text{ e}^- / \text{m}^3$.

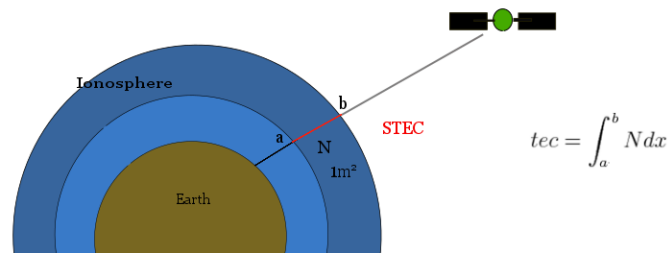
Es una capa de absorción, debido a que las ondas electromagnéticas que la atraviesan sufren una atenuación. En el día su altitud y densidad de ionización son estables pero durante la noche, esta capa se desaparece debido a la combinación de los iones y electrones.

Capa E: Es la capa de gas ionizado con una altitud de 90 a 150 km, una presión de 0.01 Pascal, una temperatura de -50°C y concentración de electrones de $10^{11} \text{ e}^- / \text{m}^3$. Corresponde a una capa con un comportamiento regular donde su densidad de ionización depende de su ángulo de incidencia de la radiación solar, alcanzando la incidencia perpendicular en lo máximo. Es la capa capacitada para reflejar ondas de frecuencia media, lo que contribuye a que se permite una propagación más allá del horizonte.

Capa F: Esta capa, alcanza una altitud de 120 a 800 km, una presión de 10^{-4} Pascales, una temperatura de 1000°C y una concentración de electrones de $10^{11} \text{ e}^- / \text{m}^3$. Para que se produzca reflexión las ondas deben tener una frecuencia inferior a 10 MHz o superiores a 3 MHz. De esta forma se pueden observar propagaciones de hasta 4000 km por cada salto alcanzando más distancias. Es la región de interés para el estudio de las radiocomunicaciones. Así mismo en esta capa se encuentran las subcapas: F1 con una altitud de 180 a 300 km, donde se evidencia una mayor fluctuación de la capa F2 y la F2 de altitud 300 a 600 km, siendo así la capa más alta de la ionósfera. [18]

3.2.2 TEC (Total Electron Content). Contenido total de electrones (TEC) es una cantidad descriptiva para la ionósfera. Se puede representar por la siguiente ecuación para los valores entre A y B:

Figura 6. Contenido total de electrones.



El valor del contenido total de electrones está dada por N siendo la densidad electrónica dada en e/m^3 y, dh el incremento en altura dada en metros, la integral se da a lo largo de los puntos A y B.

Los valores TEC varían entre 10^{16} y 10^{19} e/metro cuadrado, esto depende de la hora del día, la estación, la actividad solar, la actividad magnética, la latitud geomagnética, entre otras.

TEC es significativo en la determinación de los centelleos, retardos de grupo y de fase de una onda de radio a través de un medio. Se caracteriza por la observación de portadores de retardos de fase de las señales de radio recibidas y transmitidas desde satélites situados por encima de la ionósfera, a menudo utilizando sistema de posicionamiento global satelital.

Se describe una técnica para la recuperación de la distribución mundial de contenido vertical total de electrones (TEC) a partir de mediciones basadas en

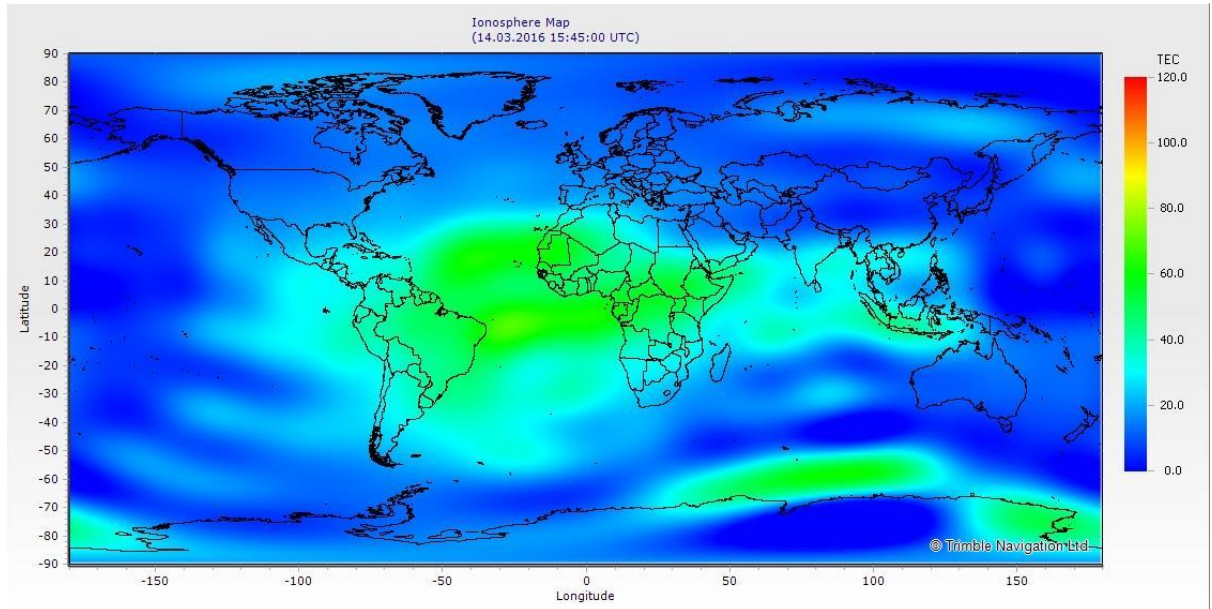
GPS. La alta resolución espacial de los métodos basados en pixel, donde las regiones ampliamente separadas pueden ser recuperadas independientemente una de la otra, se combina con la recuperación eficiente de los gradientes característicos de ajuste polinómico. TEC realiza predicciones de los modelos climatológicos se incorporan como datos simulados para cerrar brechas significativas entre las mediciones. Las secuencias de tiempo de mapas globales TEC se forman para incrementar y actualizar la más reciente recuperación de los datos a medida que estén disponibles. [19]

3.2.3 Mapas TEC. Los mapas globales del contenido total de electrones en la ionósfera (TEC) son producidos en tiempo real gracias a la recolección de estaciones terrestres, estos mapas prueban en tiempo real la adquisición de datos; facilita la coordinación y las técnicas de mapeo. Proporcionan precisión a los sistemas de navegación; producen mediante la estimación de los valores TEC en una cuadrícula densa utilizada ampliamente por un algoritmo de interpolación geoestadístico.

Cuanto mayor sea la cantidad TEC es más largo el retraso en una señal GNSS recibida, y cuando se conoce el valor TEC es posible determinar el retardo de las señales GPS causadas por la refracción ionosférica. Algunos modelos TEC están diseñados para regiones específicas y otros para mitigar esto globalmente.

Dependiendo del servicio, existen mapas TEC implementados por: la NASA y Caltech [20] que ofrecen una actualización cada cinco minutos; el Space Weather Service [21] que actualiza los mapas TEC cada hora; el centro nacional de geoinformación de Bulgaria [22] que proporciona los últimos veinticuatro mapas creados para las últimas veinticuatro horas; por último Trimble [23] actualiza el mapa cada cinco minutos según UTC.

Figura 7. Mapa TEC para el 3 de Junio del 2015 a las 2:05 UTC



Fuente: [23].

3.2.4 Modelado. Utilizaremos una herramienta llamada GPSTK, que se compone de un conjunto de aplicaciones precompiladas que permite el procesamiento de datos GNSS y un conjunto de ficheros cabecera (*.hpp), que sirven para desarrollar aplicaciones embebidas en la librería.

Esta librería contiene procesamiento avanzado de GPS y proporciona el cálculo de acontecimientos importantes ocurridos en la misma fecha, modelos de refracción atmosférica y los algoritmos de posicionamiento, también el cálculo del contenido total de electrones(TEC) de la ionósfera y manipulación de archivos RINEX.[24]

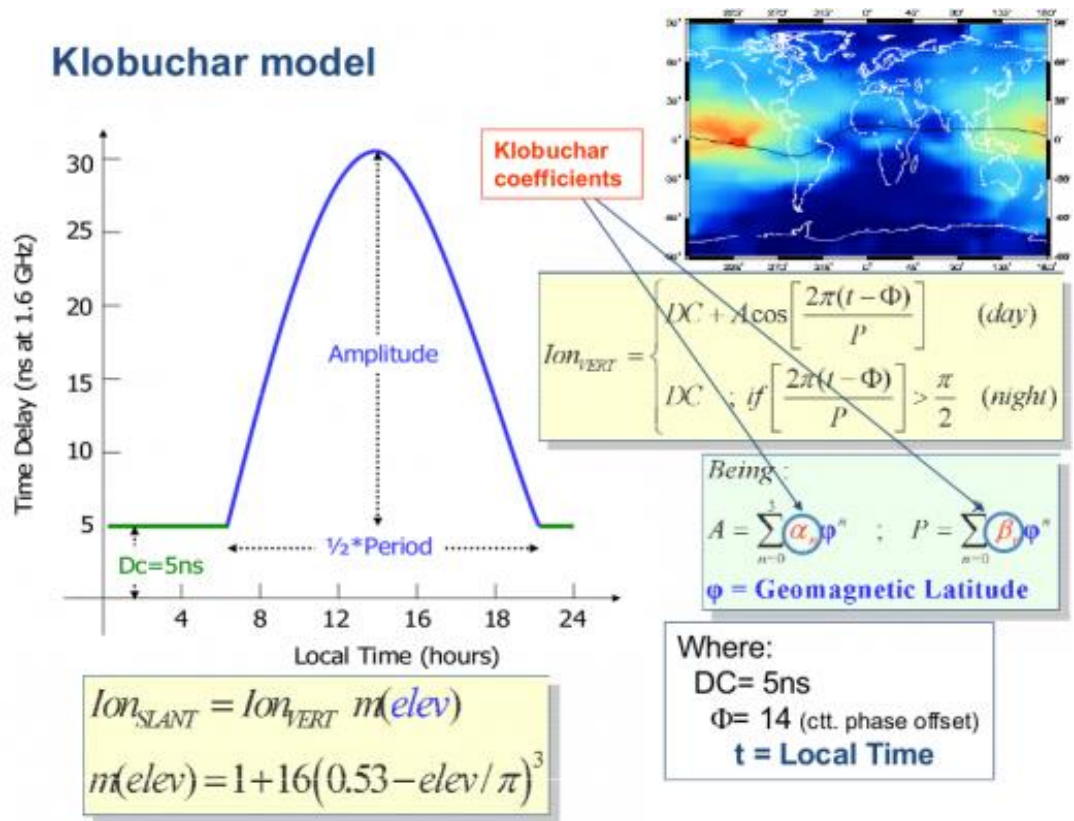
Los modelos ya diseñados para zona del hemisferio norte son:

3.2.4.1 Modelo ionosférico Klobuchar: El modelo Klobuchar fue diseñado para minimizar la complejidad computacional y almacenamiento del ordenador del usuario en cuanto a tener un número mínimo de coeficientes para transmitir en el satélite; Los satélites GPS transmiten los parámetros del modelo para usuarios individuales de frecuencia. Su estructuración es muy simple porque utiliza el

mismo modelo a todas las partes del mundo, y además los retrasos verticales tienen un valor constante en el tiempo de la noche y una función de la mitad de coseno en el día, donde la amplitud y el periodo se dan como función de ocho parámetros en el mensaje de navegación GPS.

Este modelo tiene un enfoque empírico y se estima que reduce por la RMS, el rango 50 % del error ionosférico a nivel mundial. Se supone que el contenido de electrones se concentra en una capa delgada en kilómetros de altura. [1]

Figura 8. Modelo Klobuchar.

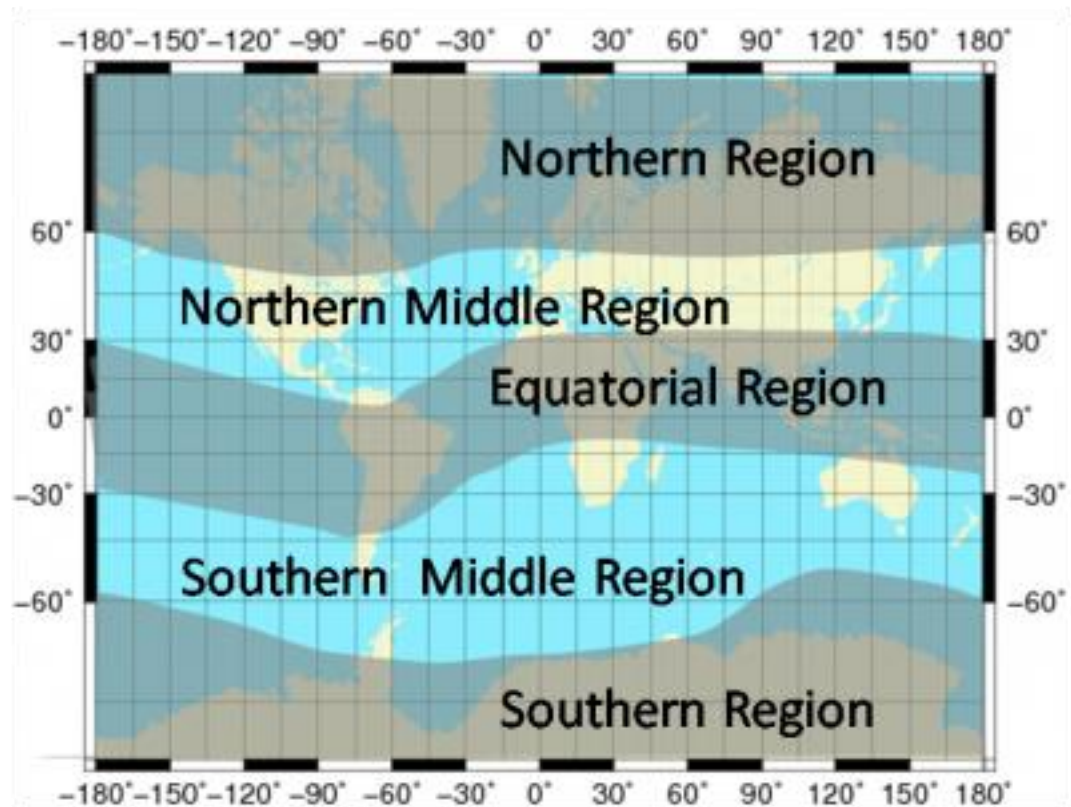


Fuente:[1].

3.2.4.2 Modelo ionosférico NeQuick: NeQuick es el modelo ionosférico propuesto para ser utilizado por el receptor de frecuencia única de Galileo, para calcular las correcciones ionosféricas, para ser se ajustado según la ubicación

terrestre, ya que para cada región definida por el modelo existe un comportamiento diferente de la ionósfera. Es un modelo dependiente del tiempo, que proporciona la densidad de electrones en la ionósfera como una función de la posición y el tiempo. De allí, se calcula los retrasos ionosféricos, ya que esta la densidad de electrones integrada a lo largo de cualquier trayectoria de rayos. [2]

Figura 9. Modelo NeQuick.



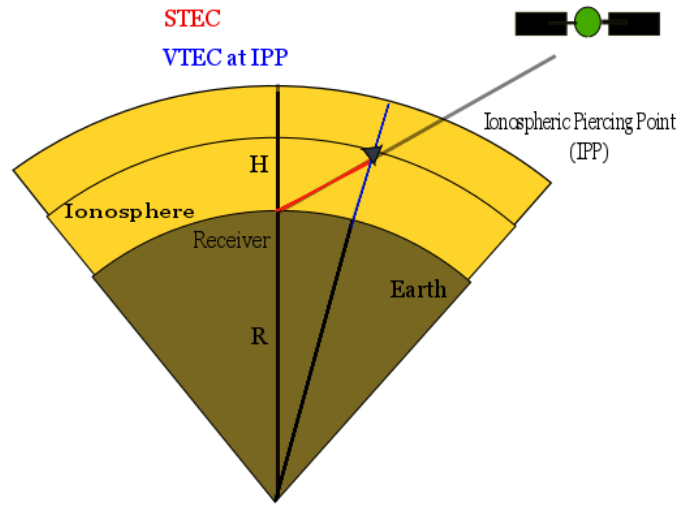
Fuente: [2].

3.2.5 Esquema de una sola capa. El esquema ionosférico de una sola capa (MSLM) asume que todos los electrones libres están contenidos en una capa de espesor infinitesimal de 450 kilómetros de altura, que corresponde a la altitud máxima de la densidad de electrones.

Este esquema es utilizado para corregir el efecto de la longitud de la trayectoria de

la señal transmitida por el satélite que atraviesa la ionósfera, por ello es necesario realizar una estimación del valor del VTEC por encima de un punto dado en la superficie de la Tierra.

Figura 10. Modelo de una sola capa de la ionósfera.



Single-layer model for the Ionosphere

3.3 MACHINE LEARNING

Machine learning se basa en construir un sistema de cómputo que pueda adaptarse y aprender de su experiencia. Se implementan algoritmos informáticos para que la máquina pueda aprender, predecir y analizar automáticamente dependiente de los datos para poder realizar la predicción.

Se puede clasificar en categorías a cada elemento de un conjunto de datos, realizando un acercamiento a un valor real mediante una regresión, ordenación o segmentación de datos de acuerdo con algún criterio.

También se puede ver como un método para enseñar a la máquina para que reaccione ante nuestro comportamiento y seguir mejorando luego de hacer su trabajo.

Por eso se describe como el proceso de conversión de datos en conocimiento, desarrollo de herramientas aplicadas en la interpretación e integración de datos. [25]

3.3.1 Flujo de trabajo de *machine learning*. El flujo de trabajo que se utilizó está basado en machine learning, su implementación se basa por los datos recolectados, para luego crear un modelo predictivo basado en un algoritmo de regression de *scikit learn* y de este modo contribuir el análisis y corrección del error que se presenta en los modelos de error ionosférico.

Figura 11. Flujo de trabajo para el módulo de *machine learning*.



3.3.2 Dataset. Un dataset es una representación residente en memoria o en disco de datos que proporciona un modelo de programación relacional coherente independientemente del origen de datos. El dataset representa un conjunto completo de datos que incluye tablas relacionadas y restricciones, así como relaciones entre las tablas.

3.3.3 Scikit-Learn. Es una librería de código abierto en el lenguaje de programación Python para machine learning, minería de datos y análisis de datos, su objetivo principal es proporcionar un conjunto de algoritmos comunes a sus usuarios por medio de una interfaz consistente.

Esta herramienta tiene una buena documentación que se apoya con sus grandes colaboradores capaces de dedicar tiempo a este proyecto, tiene grandes implementaciones de algoritmos populares, proporciona algoritmos a los usuarios en Python a través de una interfaz. Algunos proyectos son manejados independientemente por la gran cantidad de datos y dependencias necesarias, aunque *Scikit* escala la mayoría de problemas de datos.

3.3.4 Sklearn pipeline. Es un módulo se construye un estimador compuesto para solicitar una lista de transformaciones y un estimador final en métodos. El uso de un Pipeline permite simplificar el proceso de ejecución manual de una serie de pasos repetitivos y tediosos, en cambio se obtiene una interfaz declarativa donde se aprecia el modelo. Este modelo extrae los documentos de texto, tokenizers, conteo de fichas y realiza una transformación TF-IDF antes de pasar estas características a un clasificador de Bayes.

Los dos propósitos de Pipeline son:

- Utilidad al momento de tener los datos y lograr adaptarse a la sucesión de estimadores por medio del ajuste y predicción.
 - Selección de parámetros buscando los estimadores en una sola búsqueda.
 - Para algunos parámetros la búsqueda no está directamente dentro de los estimadores es posible configurarla por medio de un espacio de parámetros para la validación cruzada de acuerdo al desempeño estimador de puntuación.
- [26]

3.4 FRAMEWORK

Un framework de software es una plataforma que integra distintos componentes para facilitar el uso de los mismos por parte del usuario final, proporcionando un software de aplicación. También se puede ver como un entorno de software universal que ofrece una funcionalidad concreta como parte de una mejor plataforma facilitando el desarrollo software de aplicaciones y productos, incluyendo compiladores, bibliotecas de código, herramientas e interfaces de programación para permitir el desarrollo de un proyecto. [27]

En nuestro caso, nuestro framework integrará:

- GPSTK
- *Scikit Learn*.
- Procesado de ficheros RINEX.
- Construcción y validación de modelos de *machine learning*
- Generación de mapas Ionosféricos.

4. METODOLOGÍA

4.1 ESTRATEGIA GENERAL

Se obtendrán datos para el entrenamiento y el pre-procesado de distintos modelos predictivos para la corrección del error ionosférico para sistemas GNSS.

- Metodología de desarrollo ágil: Estrategia de desarrollo incremental, con productos entregables a lo largo de cada iteración, definir para cada una de ellas los requisitos y comunicación constante.
- Calidad del producto, con iteraciones cortas y módulos funcionando, además de aceptar cambios e incorporarlos en el proyecto.

4.2 ORGANIZACIÓN DEL PROYECTO

El proyecto se organiza en bloques de trabajo y actividades dentro de cada uno de ellos. De cada bloque de trabajo se obtiene un conjunto de entregables

4.2.1 Bloque de Trabajo [BT] 1: Módulo Obtención de datos

- Actividad 1.1: Identificación de herramientas y diseño del framework
- Actividad 1.2: Definir la configuración de datos crudos IGS
- Actividad 1.3: Seleccionar las estaciones que van a componer el dataset
- Actividad 1.4: Descargar datos crudos IGS
- Actividad 1.5: Analizar el error ionosférico
- Entregable 1.1: Reporte de las características del framework
- Entregable 1.2: Modelo de la estructura del framework

- Entregable 1.3: Listado de estaciones
- Entregable 1.4: Datos RINEX descargados
- Entregable 1.5: Calculo del error ionosférico

4.2.2. BT2: Módulo Preprocesado de Datos

- Actividad 2.1: Generar la estructura para los ficheros descargados
- Actividad 2.2: Procesar datos crudos IGS
- Actividad 2.3: Creación de Dataset para almacenamiento de datos procesados
- Actividad 2.4: Estructurar el dataset para la generación de Mapas TEC
- Actividad 2.5: Implementación de algoritmos de interpolación para la generación de Mapas TEC
- Entregable 2.1: Formato para el Preprocesado de los ficheros descargados
- Entregable 2.2: Datos organizados para la creación del dataset
- Entregable 2.3: Dataset a partir de los datos procesados
- Entregable 2.4: Datos organizados para crear Mapas TEC
- Entregable 2.5: Mapas TEC de la Ionósfera

4.2.3. BT3: Módulo de generación de modelos predictivos

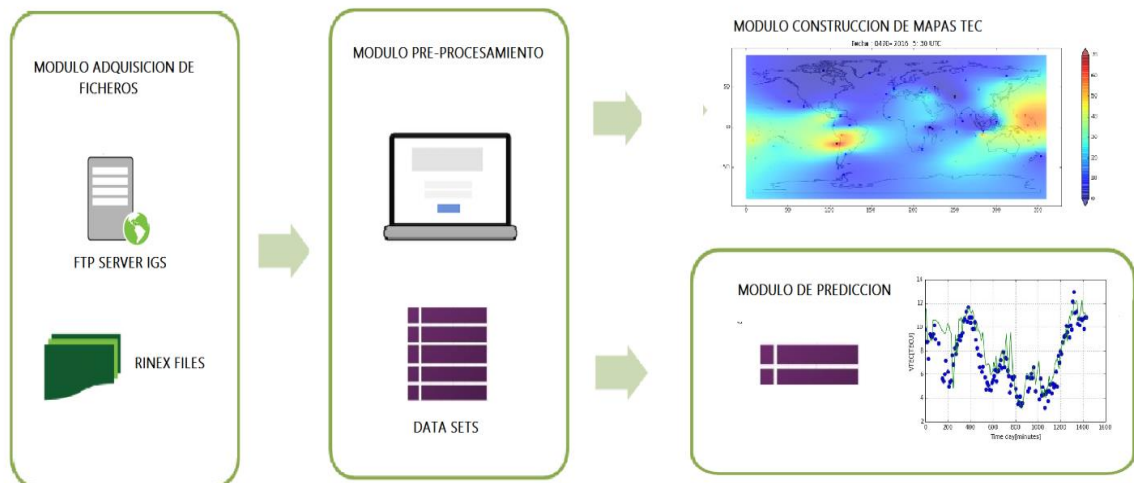
- Actividad 3.1: Crear modulo para experimentos con el formato para la predicción de la ionósfera
- Actividad 3.2: Integración de datos con *Scikit-Learn*
- Actividad 3.3: Evaluar algoritmos con *Scikit-Learn*
- Actividad 3.4: Comparar funcionamiento de algoritmos
- Entregable 3.1: Módulo para predicción con los datos integrados
- Entregable 3.2: Datos transformados con *Scikit-Learn*
- Entregable 3.3: Desempeño de algoritmos predictivos.

- Entregable 3.4: Reporte del funcionamiento en la corrección del error ionosférico por medio de algoritmos predictivos en el *Scikit-Learn*

5. DESARROLLO DEL PROYECTO

En este capítulo presentamos los módulos del framework implementados en jupyter notebook, tres de ellos dependen de los dos primeros, el módulo de adquisición de ficheros y el módulo de procesamiento de datos. De allí obtenemos los dataset que contienen los valores que se cargan en los siguientes módulos para funcionar adecuadamente, el primero es el módulo para generar mapas TEC de acuerdo a la posición del Ionospheric Pierce Point (IPP) y el valor del Vertical Total Electron Content (VTEC), el segundo es el módulo predictivo del VTEC, por medio de los algoritmos *Linear Regression*, *Support Vector Regression (SVR)*, *Decision Tree Regressor* y *Random Forest* de *Scikit learn*. Para realizar un flujo de trabajo, como se muestra en la figura 12.

Figura 12. Flujo de trabajo del proyecto.



5.1 ADQUISICIÓN DE DATOS DE ESTACIONES.

5.1.1 Identificación y selección de herramientas necesarias para la construcción del framework. Observando que el error ionosférico es el mayor contribuyente del error total en los sistemas de posicionamiento global utilizando dos frecuencias GPS, el retardo ionosférico calculado puede ser corregido, pero los receptores de dos frecuencias son demasiado costosos.

Por este motivo desarrollamos el Framework, puesto que utilizamos el modelo de predicción para estimar el valor del VTEC en cualquier momento del día considerando la hora del día y el Ionospheric Pierce Point que fue calculado mediante la fórmula del retardo ionosférico que utiliza los observables C1 y P2 del formato RINEX.

Pero los receptores de doble frecuencia son los que generan estos observables, por medio de nuestro modelo de predicción generamos mapas TEC sin la necesidad de estos receptores.

Para crear el flujo de trabajo correcto utilizamos la librería WGET que nos permitió realizar la descarga de los ficheros y luego mediante la herramienta GPSTK se realizó el procesamiento de los ficheros de observación y navegación de la red IGS, en la creación de los dataset se utilizó la librería PANDAS para facilitar la estructuración y el análisis de datos. Finalmente para la interpolación de los valores VTEC en el mapa se utilizó la librería SciPy y en la predicción la herramienta de minería de y análisis de datos de *scikit learn*.

5.1.2 Descarga de ficheros de observación y navegación. La selección de las estaciones que conformaran nuestro dataset se estableció verificando la existencia de cada uno de los ficheros para el procesamiento. En lo posible escogimos estaciones que tuvieran una fecha de actualización de datos reciente, para

generar diariamente los dataset según la hora establecida.

Consideramos que la calibración de la antena fuera óptima además que el sistema satelital establecido fuera GPS, como se aprecia en la figura 13. La estación SHAO ubicada en Sheshan, China cuenta con una disponibilidad frecuente y GPS es su único sistema de satelital.

Figura 13. Tabla de estaciones de la red IGS

Lat	Long	Site	Calibration	Last Data Avail	Sat System
31.5930556	35.3919444	DRAG	ROBOT	20-Mar-16	GPS+GLO
31.0994444	121.2002778	SHAO	n/a	20-Mar-16	GPS
30.6802778	-104.0147222	MDO1	n/a	n/a	GPS
30.5977778	34.7630556	RAMO	COPIED	20-Mar-16	GPS+GLO
30.5563889	131.0155556	GMSD	ROBOT	20-Mar-16	GPS+GLO+GAL+BDS+QZSS
30.5316667	114.3572222	WUH2	ROBOT	n/a	GPS+GLO+GAL+QZSS+SBAS
30.5313889	114.3572222	WUHN	n/a	20-Mar-16	GPS+GLO
30.5155556	114.4908333	JFNG	ROBOT	19-Mar-16	GPS+GLO+GAL+BDS+QZSS+SBAS
29.6572222	91.1038889	LHAZ	ROBOT	19-Mar-16	GPS+GLO+GAL+BDS
29.1386111	36.0997222	HALY	FIELD	n/a	GPS

Fuente: [15].

Teniendo en cuenta principalmente la ubicación de cada estación se eligieron estratégicamente 100 estaciones de 491 disponibles en la red IGS, que cumplieran con la necesidad de lograr gran cobertura en toda la superficie terrestre, excluyendo un número de estaciones relativamente cercanas a las elegidas que generarían información excesiva para interpolar el valor del VTEC según fecha y hora establecida de cada estación generando así el respectivo mapa TEC.

Los ficheros son descargados mediante el módulo de adquisición de datos, como

se aprecia en la figura 14. Algunos ficheros como el de navegación en RECF en Brasil, no existen para el 17 de febrero de 2016, luego de la descarga de cada uno de los ficheros serán descomprimidos y almacenados en su respectivo directorio, Para las estaciones como la de RECF sin fichero de navegación se realiza la búsqueda de la estación más cercana que tenga fichero de navegación para ese día, cuando se encuentra se copia el contenido del fichero a otro con el nombre de la estación que no tenía fichero de navegación inicialmente.




























Figura 14. Estaciones donde no existen de ficheros de navegación.

```
nril doesn't exist
ntus doesn't exist
pbri doesn't exist
pove doesn't exist
recf doesn't exist
rio2 doesn't exist
salu doesn't exist
savo doesn't exist
scor doesn't exist
sey1 doesn't exist
sgoc doesn't exist
tash doesn't exist
tixi doesn't exist
toka doesn't exist
ufpr doesn't exist
yakt doesn't exist
66 stations have been downloaded
```

La fecha de descarga se establece al inicio del notebook con los ficheros de navegación y observación que se van a procesar y estén disponibles en el servidor ftp de la página web de la red IGS. En la figura 15 se muestran los ficheros de observación que están disponibles para el día 059 (28 de febrero) del año 2016 del servidor ftp de la red IGS.

Figura 15. Página del servidor FTP del IGS.

Índice de /pub/gps/data/daily/2016/059/16o/

Nombre	Tamaño	Fecha de modificación
 [directorio principal]		
 MD5SUMS	23.3 kB	18/03/16 11:10:00
 SHA512SUMS	68.8 kB	18/03/16 11:10:00
 WTZZ00DEU_R_20160590000_01D_30S_MO.mx.gz	5.5 MB	28/02/16 20:20:00
 aber0590.16o.Z	2.8 MB	04/03/16 9:44:00
 abmf0590.16o.Z	3.2 MB	01/03/16 2:36:00
 abpo0590.16o.Z	1006 kB	29/02/16 2:10:00
 adis0590.16o.Z	1.7 MB	29/02/16 3:09:00
 aira0590.16o.Z	1.5 MB	01/03/16 17:51:00
 albh0590.16o.Z	1.9 MB	29/02/16 1:21:00
 alg20590.16o.Z	2.1 MB	29/02/16 1:21:00
 alg30590.16o.Z	2.1 MB	29/02/16 1:21:00
 algo0590.16o.Z	2.1 MB	29/02/16 1:21:00
 alic0590.16o.Z	1.6 MB	29/02/16 1:21:00
 alrt0590.16o.Z	1.1 MB	29/02/16 1:21:00
 amc20590.16o.Z	999 kB	29/02/16 2:10:00
 amu20590.16o.Z	2.2 MB	29/02/16 3:31:00
 ankr0590.16o.Z	2.4 MB	18/03/16 5:54:00
 areg0590.16o.Z	3.7 MB	18/03/16 5:54:00
 areq0590.16o.Z	1.0 MB	29/02/16 2:10:00
 arev0590.16o.Z	2.3 MB	29/02/16 2:10:00
 artu0590.16o.Z	1.0 MB	29/02/16 2:54:00
 aruc0590.16o.Z	972 kB	29/02/16 2:10:00
 ascg0590.16o.Z	3.2 MB	01/03/16 10:12:00
 aspa0590.16o.Z	1.9 MB	29/02/16 1:21:00
 auck0590.16o.Z	1.8 MB	18/03/16 5:54:00
 aukt0590.16o.Z	790 kB	18/03/16 5:54:00

Fuente: [28].

5.2 PROCESAMIENTO DE DATOS PARA LA CONSTRUCCION DE DATASET.

Considerando la metodología elegida para el segundo bloque de trabajo, se desarrolló el módulo de procesamiento cuyo fin era exportar los dataset que

contenían los valores que permiten realizar el mapa TEC y predecir el valor del VTEC.

Para ello realizamos las siguientes iteraciones cada una con su respectivo entregable:

- Cargar el nombre de las 100 estaciones seleccionadas.
- Seleccionar las estaciones con ficheros de observación
- Crear una lista que contenga latitud, longitud y nombre de cada estación
- Crear los ficheros de navegación no existentes
- Computar el retardo ionosférico, IPP Latitude, IPP Longitude, VTEC y el Id satélite.
- Crear el dataset

Para realizar la segunda iteración se tomaron los datos crudos disponibles en la red IGS de cada una de las estaciones elegidas, en algunos casos estas estaciones no tenían disponible el fichero de observación para la fecha establecida en el inicio del notebook, la estación que no posee este fichero es eliminada de la lista que conformara el dataset porque es imposible reconocer los satélites que registraron alguna observación para esa estación.

En la cuarta iteración se realizó una búsqueda de la estación más cercana tomando en cuenta la latitud y la longitud que se aproximaba más a las coordenadas de la estación sin fichero de navegación. Una vez encontrada la estación más cercana se copia el contenido del fichero de navegación con el nombre de la estación sin fichero de navegación dada inicialmente, este proceso se puede realizar ya que los satélites de una estación cercana a otra puede percibir los mismos satélites en la hora que se establezca, en la figura 16 se observa la estación BOGT ubicada en Colombia que no posee fichero de navegación, pero tomando las estaciones con fichero de navegación se encuentra,

que la más cercana es la estación RIOP ubicada en Ecuador, por lo tanto se procede a realizar la copia del fichero.

Figura 16. Estaciones sin fichero de observación para una hora determinada.

```
auck does n't have Navigation file. The nearest station is: kouc
bako does n't have Navigation file. The nearest station is: xmis
bjnm does n't have Navigation file. The nearest station is: shao
bogt does n't have Navigation file. The nearest station is: riop
braz does n't have Navigation file. The nearest station is: scrz
btng does n't have Navigation file. The nearest station is: darw
cusv does n't have Navigation file. The nearest station is: xmis
gras does n't have Navigation file. The nearest station is: zimm
guam does n't have Navigation file. The nearest station is: mcil
guat does n't have Navigation file. The nearest station is: mana
hrao does n't have Navigation file. The nearest station is: suth
lpgs does n't have Navigation file. The nearest station is: cord
maui does n't have Navigation file. The nearest station is: aspa
nril does n't have Navigation file. The nearest station is: irkj
pbri does n't have Navigation file. The nearest station is: coco
pove does n't have Navigation file. The nearest station is: scrz
recf does n't have Navigation file. The nearest station is: kour
rio2 does n't have Navigation file. The nearest station is: cord
salu does n't have Navigation file. The nearest station is: kour
savo does n't have Navigation file. The nearest station is: kour
scor does n't have Navigation file. The nearest station is: reyk
sgoc does n't have Navigation file. The nearest station is: dgav
tash does n't have Navigation file. The nearest station is: zwe2
tixi does n't have Navigation file. The nearest station is: irkj
ufpr does n't have Navigation file. The nearest station is: scrz
yakt does n't have Navigation file. The nearest station is: irkj
Stations without Navigation file: 26
```

La selección de estaciones y creación de nuevos ficheros de navegación a partir de otros ficheros generan la mejor predicción del valor VTEC, por que se toman en cuenta una gran cantidad de estaciones y se obtiene interpolación precisa para crear el mapa TEC.

En la quinta iteración se realizaron algunas lecturas sobre el análisis matemático

del VTEC con doble frecuencia además teniendo en cuenta el modelo de una sola capa para la ionósfera, se realizó los cálculos necesarios del valor VTEC. Teniendo en cuenta esto, se llevó acabo el siguiente procedimiento:

- Crear una lista para almacenar todas las observaciones existentes, llamada bcestore[29]
- Seleccionar los satélites que conforme el sistema satelital GPS.
- Verificar que el fichero de navegación contenga por lo menos una observación, con el fin de descartar ficheros que no almacenaron en su bcestore.
- Comprobar que la hora de la observación sea la misma a la hora establecida para realizar el procesamiento.
- Examinar que exista una órbita (ephemeris) para el satélite que se encuentre buscando en el momento indicado.
- Buscar una órbita para el satélite indicado en el tiempo respectivo, tomando en cuenta su bcestore.
- Calcular la posición del satélite y la posición del receptor (antena).
- Calcular la elevación y el azimut de la antena con respecto a la posición del satélite.
- Calcular el IPP.
- Calcular la latitud y la longitud de acuerdo al IPP.
- Comprobar que la diferencia entre el tiempo de recepción y el tiempo de transmisión de una señal de satélite distinta (C1 y P2) en cada frecuencia (F1 y F2) sea diferente de cero.
- Computar el índice de cada pseudodistancia (C1 y P2).
- Obtener el RinexDatum proporcionado por el fichero de observación, teniendo en cuenta el id del satélite y el índice de cada pseudodistancia. [30]
- Calcular el valor del VTEC en el tiempo respectivo por las siguientes ecuaciones:

Ecuación 1. Slant factor: factor de inclinación del contenido total de electrones en la ionósfera.

$$\text{slant factor} = 1 + 16 * (0.53 - \text{elev} * \frac{1}{180})^3$$

Ecuación 2. Ionospheric delay.

$$\text{Ionospheric delay} = \frac{1}{(1 - \text{gpstk.gamma_GPS}) * (C1 - P2)}$$

Ecuación 3. Vertical Total Electron Content.

$$V_{tec} = \frac{\text{Ionospheric delay}}{\text{Slant factor}}$$

La refracción ionosférica causa el retardo de la propagación de la señal GPS para cada una de las frecuencias, por ello la ecuación del retardo ionosférico depende de la señal de las frecuencias, aproximadamente el retraso es de 16 cm, proporcionado por los observables (C1 y P2) de cada una de las frecuencias.

El VTEC se mide en una unidad llamada TECU (total electron content units), donde 1 TECU = 10^{16} electrones por m^2 de acuerdo a la cantidad total de electrones que se encuentran dentro de la ionósfera en el ángulo de elevación proporcionado. Este valor se encuentra establecido dentro del rango desde 0 hasta 75 TECU, por esto que algunas estaciones no son procesadas. En la figura 17 se aprecia la cantidad de estaciones que no tiene datos para el 18 de febrero de 2016 que es la fecha establecida en este caso.

Figura 17. Mensaje de confirmación, los datos han sido procesados.

```
86 Stations has been processed
1 Stations without observation in that time
```

El retardo en las señales GNSS, la dispersión de la señal GNSS ocurre por irregularidades ionosféricas, estos son fenómenos llamados centelleos que ocurren de noche frecuentemente en la zona tropical.

En la iteración final para generar el dataset en la hora establecida, los valores calculados en la anterior iteración se fueron almacenando en la respectiva columna que finalmente conformaría el dataset de esa hora. Como se muestra en la figura 18, cada estación tendrá 8 parámetros que serán necesarios para los próximos módulos.

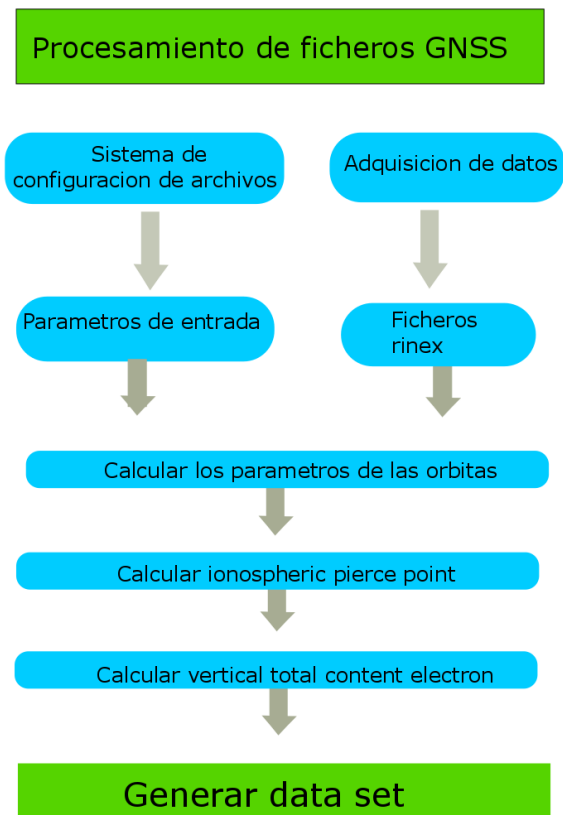
Figura 18. Extracto del dataset para el día 061 de año 2016.

	station	date	hour	prn	ipp lat	ipp lon	iono delay	vtec
0	ADIS	0610-16	23:00	22	5.968494	212.419814	8.535825	8.163438
1	ANKR	0610-16	23:00	21	30.318010	201.622412	6.041012	2.269591
2	AREQ	0610-16	23:00	31	-17.156179	115.456148	19.637744	12.959147
3	AUCK	0610-16	23:00	19	-34.935358	358.255255	5.748253	5.605470
4	bako	0610-16	23:00	21	-3.060427	284.091757	-1.323778	1.307527
5	BJCO	0610-16	23:00	23	-2.261637	183.151950	38.544864	18.508279
6	bjnm	0610-16	23:00	22	42.005890	304.628148	13.357434	6.514965
7	BOGT	0610-16	23:00	38	1.341822	114.904530	26.389014	10.709315

El flujo de trabajo del procesamiento de todos los ficheros que se descargan en el módulo de adquisición de datos, se puede observar en la figura 19, de allí se toman los dataset establecidos para las fechas y horas necesarias, se podrá continuar con los siguientes módulos y apreciaremos en

comportamiento del VTEC a lo largo el tiempo.

Figura 19. Diagrama del flujo de trabajo para el procesamiento de ficheros GNSS.



5.3 CONSTRUCCION DE MAPAS TEC CON DIFERENTES ALGORITMOS DE INTERPOLACION

5.3.1 Implementación de mapas TEC. Una vez creado el dataset de la hora y fecha establecida con anterioridad, se genera el mapa TEC extrayendo tres valores, las coordenadas geodésicas (latitud y longitud de la ionósfera) y el VTEC que permitió establecer el color en esa coordenada para este momento, de hecho

se puede apreciar la secuencia de coloración en el mapa a lo largo del día en cada estación, que va cambiando de acuerdo al valor que tome el VTEC en cada hora.

Las dimensiones del mapa se fijaron de acuerdo a la imagen de fondo, que se aprecia en la figura 20, inicialmente en la esquina inferior izquierda se fija la coordenada (0,-90) y en la esquina superior derecha (360, 90) en coordenadas angulares teniendo en cuenta que es la tierra es un plano horizontal.

Figura 20. Imagen de fondo de cada mapa TEC.



Fuente: [31]

Cada imagen se realizó mediante la función `imshow` de la librería `matplotlib`. [32] y se almaceno en un directorio con el nombre de la fecha y hora establecida previamente, además de la función de interpolación que se utilizó. La interpolación se realizó con tres algoritmos diferentes basados en `scipy.interpolate` [33], que es un sub paquete de funciones spline y de clases, que permiten generar una interpolación sobre los valores instaurados entre 0 y 255 que permiten tomar una tonalidad diferente a cada punto.

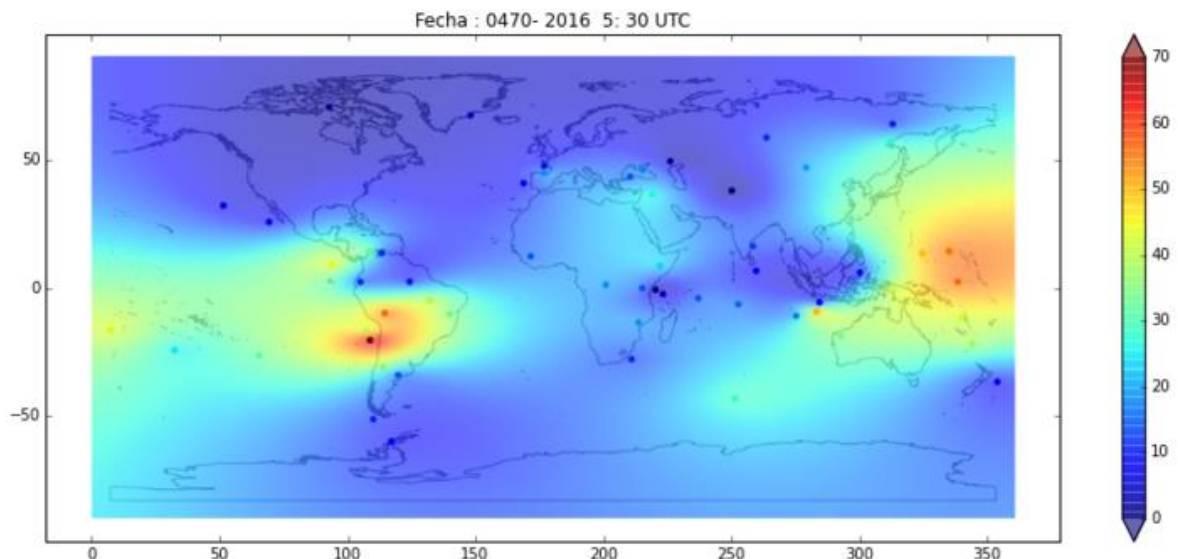
La primera interpolación se realizó mediante `Rbf` [34], una función de aproximación

e interpolación de datos dispersos en n dimensiones, es con una combinación lineal de todos los puntos calculando la norma de la distancia euclidiana.

Esta función realiza la interpolación de los puntos dados mediante una base radial tomando su norma. El parámetro función puede tomar alguno de las siguientes opciones `cubic`, `gaussian`, `inverse_multiquadric`, `linear`, `multiquadric`, `quintic`, `thin_plate`. Para esto utilizamos el método `linear`, ya que toma como parámetro la distancia entre cada estación y permite observar el comportamiento de la ionósfera de una mejor manera.

En la figura 21 se aprecia que la zona costera de Chile y el norte de Papua Nueva Guinea el VTEC tiene un valor muy alto registrado para el día 047 del año 2016 que corresponde al 16 de febrero a las 5: 30 UTC.

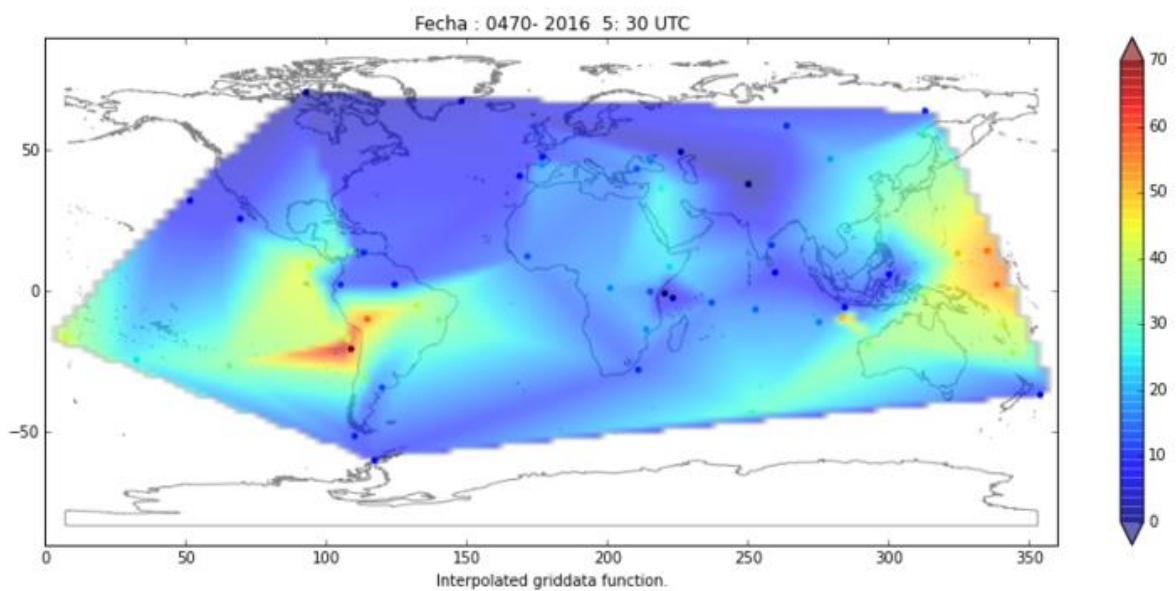
Figura 21. Mapa TEC realizado con la función RBF.



La segunda interpolación se realizó mediante `griddata` [35], una función para

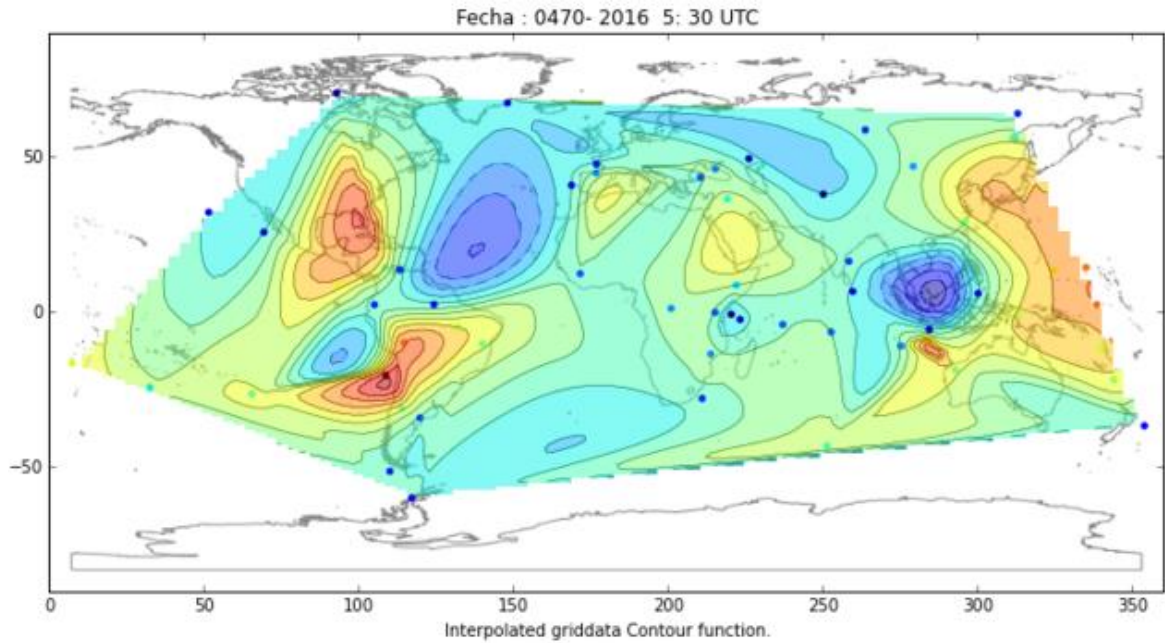
puntos no estructurados en cada una de las estaciones. Para realizar una mejor interpolación se ajustan los datos para una cuadrícula, además utilizamos el método lineal para dispersar consistentemente los puntos, en la figura 22 se percibe que la interpolación solamente se realiza para el espacio que se encuentra dentro de los límites del total de las estaciones.

Figura 22. Mapa TEC realizado con la función Griddata.



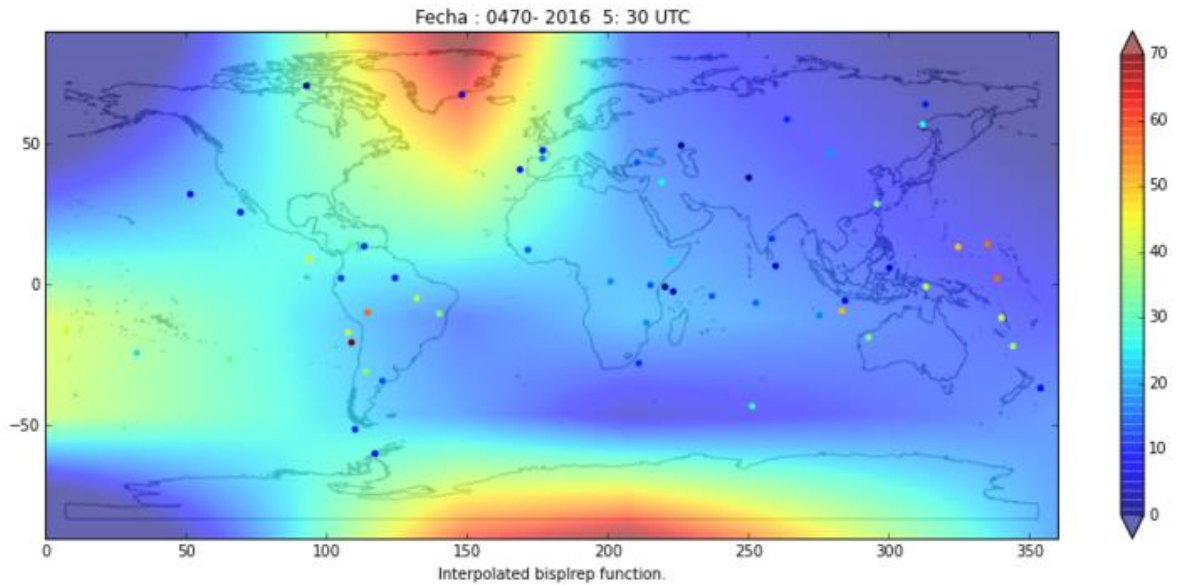
Además se realizó un gráfico adicional dibujando curvas de nivel, contornos rellenos y añadiendo bordes. Esta grafica fue implementada mediante una función countour de la librería matplotlib de Python, cada curva de nivel corresponde a un valor diferente del VTEC como se puede observar en la figura 23.

Figura 23. Mapa TEC realizado con la función de interpolación Griddata y la función de graficas countour.



La tercera interpolación se implementó por medio de bisprep [36], una función para encontrar variantes para representar una superficie de acuerdo a dos variables dentro de un conjunto de puntos, y aplicada a un grado de suavidad para cada punto, como se muestra en la figura 24.

Figura 24. Mapa TEC realizado con la función de interpolación Bispirep.



Observando cada mapa TEC generado, se aprecia que la mejor interpolación se realiza con la función RBF, sin importar el número de estaciones proporcionadas inicialmente.

El análisis de las zonas donde la ionósfera es más susceptible se realizó mediante esta función y se observa que en la zona tropical específicamente en la zona ecuatorial, el comportamiento registra un pico en la ionización por su gran contenido de electrones alrededor de la línea del ecuador en unos 10 a 20 grados.

[20]

La configuración que se realiza para determinar la interpolación de las estaciones se realiza por medio de los parámetros y_{max} , y_{min} , x_{max} y x_{min} que indican el límite de los valores que permiten graficar el mapa TEC, en la figura 25 se establecieron la máxima y la mínima latitud, 15 y -5 respectivamente y para la longitud 360 y 0 que permiten observar toda la superficie terrestre desde oriente hasta occidente, en ella se observa que el mayor valor del VTEC se encuentra en

la estación BJCO en Nigeria y la de menor valor se ubica en Filipinas es la estación PGEN.

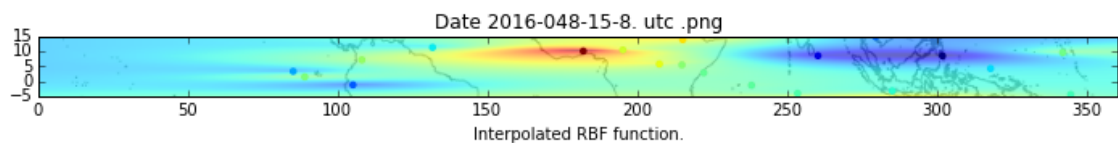
Figura 25. Mapa TEC de la zona Ecuatorial.

1. Input Parameters

```
#Name of each file: data_stations-year-day-hour.txt

xmin,xmax = 0,360      #Dimension of the interpolation
ymin,ymax = -5,15
imageBackground = 'dataImg/world2.png' #Directory of the image background of TEC map
directoryDatasets = "data_stations"
directoryImages = "dataImg/"
function = "rbf"      #function = ["rbf", "griddata", "griddataContour", "bisplrep"]

create_image(function,directoryDatasets,directoryImages,xmin,xmax,ymin,ymax,imageBackground)
```



En la zona ecuatorial se concentran los valores más altos del VTEC, ya que en esa zona el contenido total de electrones tiene mayor movimiento en la ionósfera.

El comportamiento del VTEC varía durante el día por el grado de ionización que perciba la ionósfera en ese momento, ese comportamiento se aprecia en las siguientes figuras.

Figura 26. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 2:00 UTC.

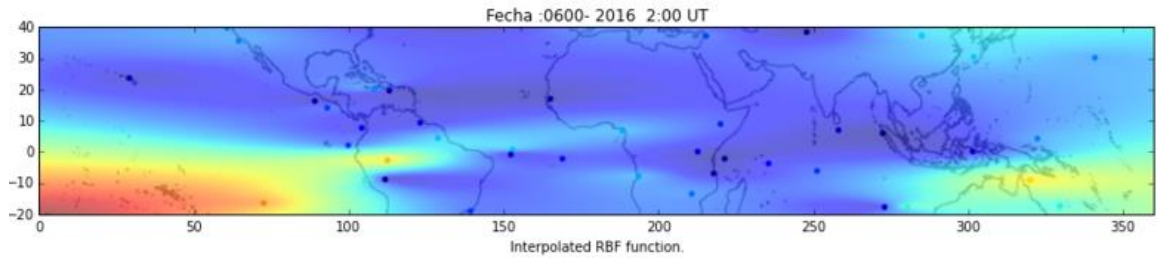


Figura 27. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 8:00 UTC.

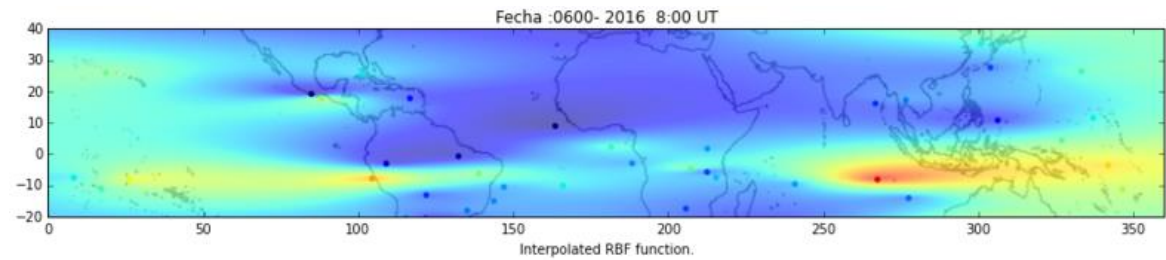


Figura 28. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 14:00 UTC.

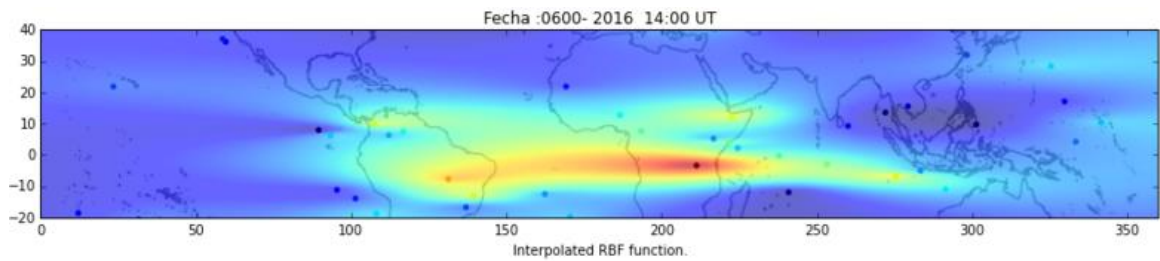


Figura 29. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 20:00 UTC.

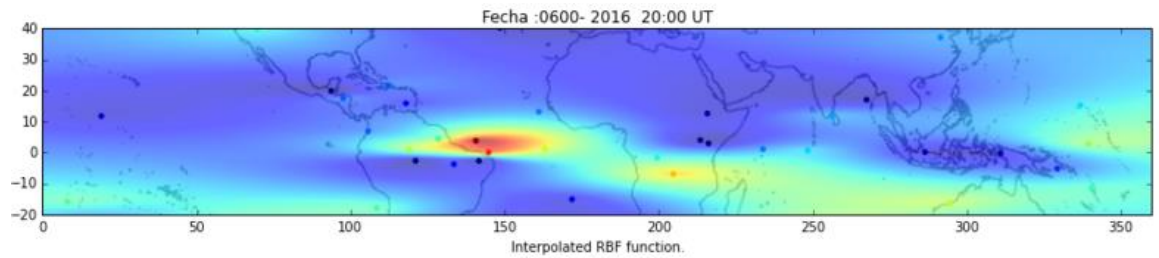
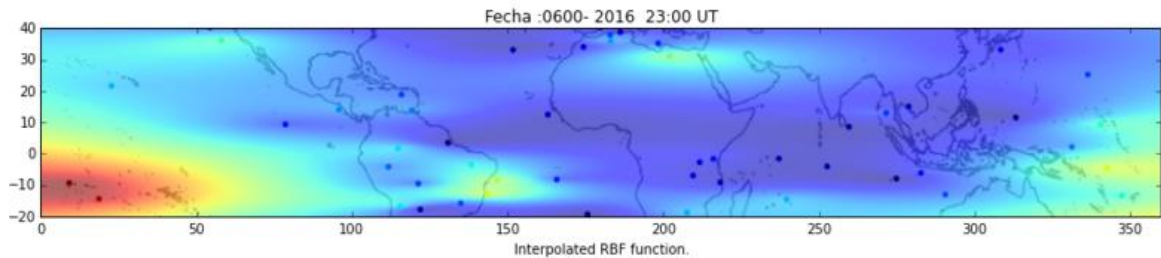


Figura 30. Mapa TEC de la zona tropical del planeta del día 29 de febrero del año 2016 a las 23:00 UTC.



Los grandes picos en los valores se perciben durante las horas del día, en la horas más tempranas de la mañana se observa ligeramente el inicio de la ionización, a lo largo del día se aprecia la secuencia de la ionización.

Mediante nuestro módulo de creación de mapas TEC, se puede analizar el comportamiento de la ionósfera en cualquier zona del planeta y cualquier momento del día, teniendo como entrada una dataset con los parámetros anteriormente mencionados, esto se aprecia en la figura 31 y figura 32.

Figura 31. Mapa TEC del círculo polar Ártico del planeta (de la zona templada del norte).

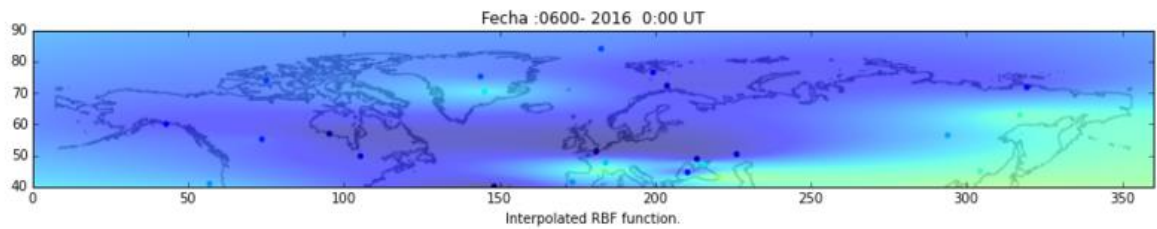
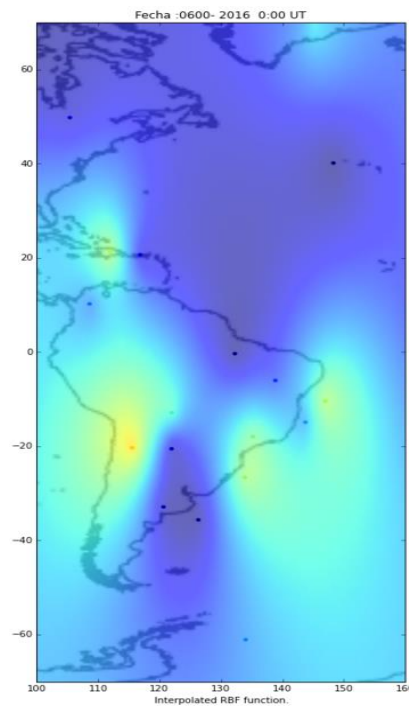


Figura 32. Mapa TEC de la región de Suramérica.



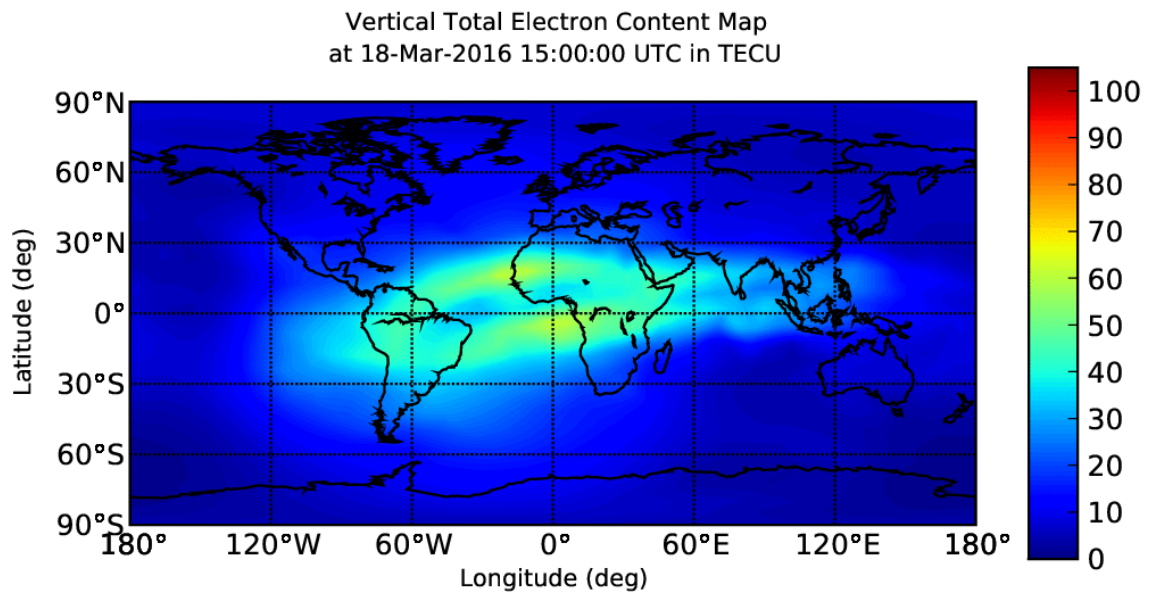
5.3.2. Comparación de diferentes mapas TEC. La siguiente comparación va a tener en cuenta tres diferentes mapas generados hasta con 250 estaciones GNSS distribuidas globalmente usando interpolación TEC, además de un procesado de datos.

Para el modelamiento de nuestros mapas TEC fueron utilizadas hasta 100

estaciones IGS, además se realizó la interpolación mediante la función RBF, por ello apreciamos que nuestros mapas generalmente tienen un comportamiento similar a cada uno de los mapas mostrados a continuación:

El primer mapa fue implementado por la NASA y Caltech, con la fecha del 18 de marzo de 2016 a las 15: 00 UTC, de ahí se pueden analizar los mapas cada cinco minutos.

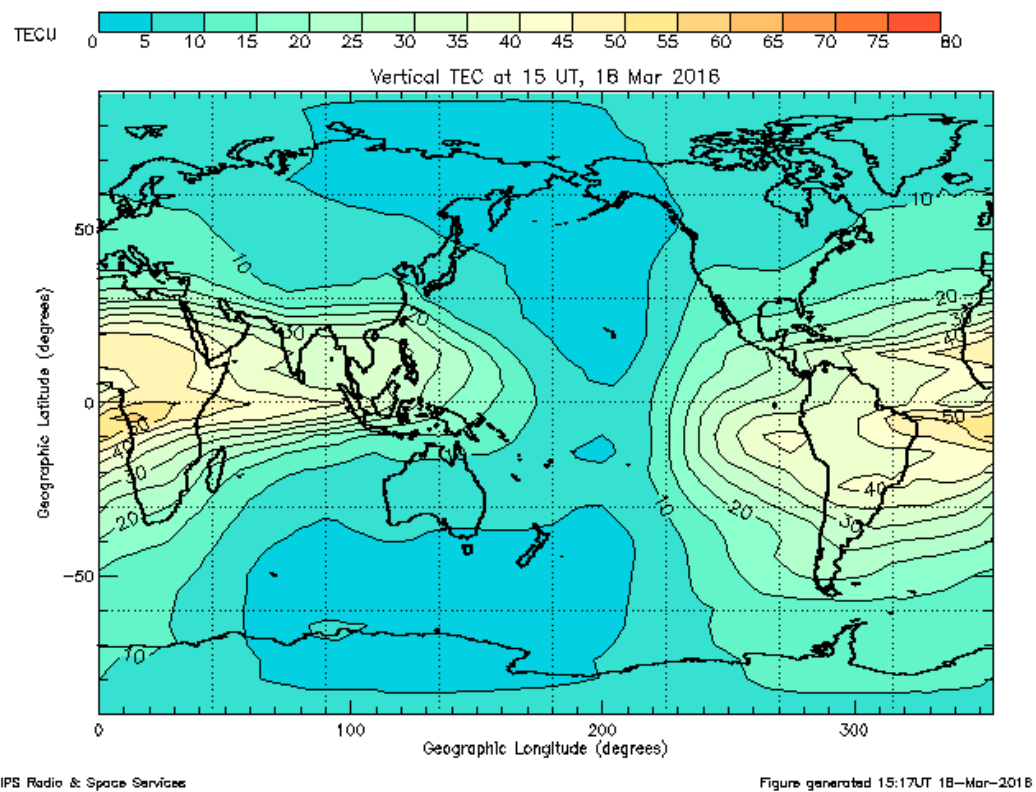
Figura 33. Mapa TEC de Caltech, NASA (Jet propulsión Laboratory).



Fuente: [20]

El segundo mapa fue implementado por la Space Weather Services del gobierno de Australia, con la fecha del 18 de marzo de 2016 a las 15: 00 UTC, de ahí se pueden observar los mapas cada hora.

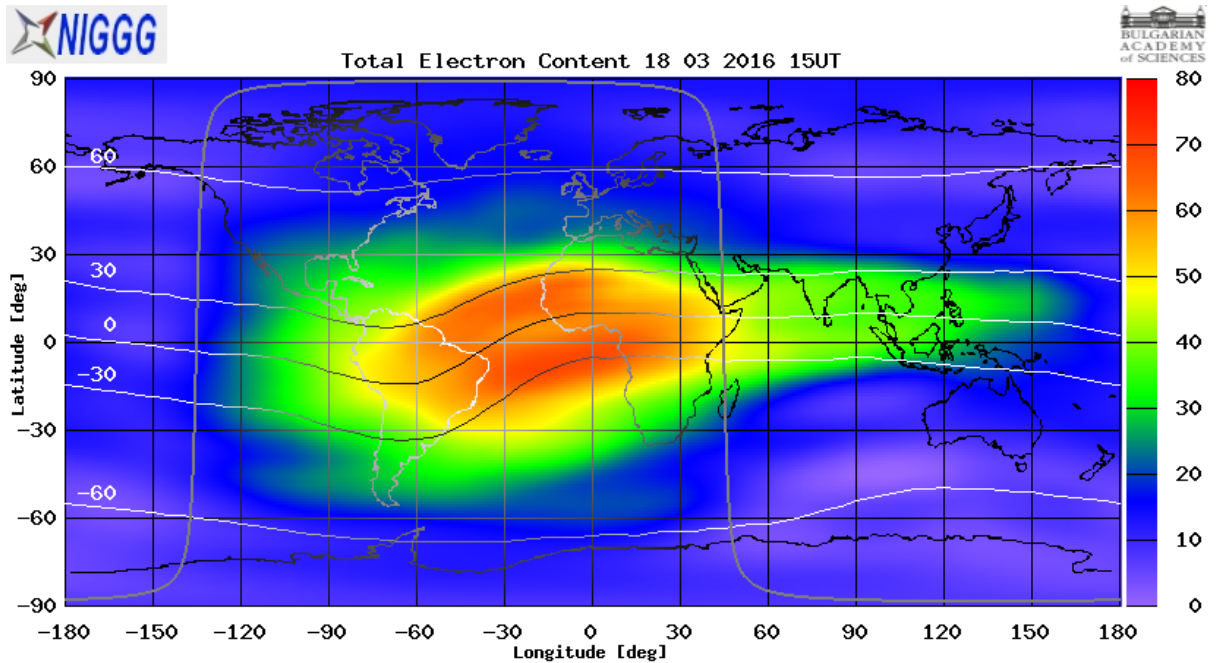
Figura 34. Mapa TEC de oficina de meteorología del Gobierno de Australia.



Fuente: [21]

El tercer mapa fue implementado por el centro nacional de geoinformación de Bulgaria, con la fecha del 18 de marzo de 2016 a las 15: 00 UTC, de ahí se pueden observar los mapas cada hora.

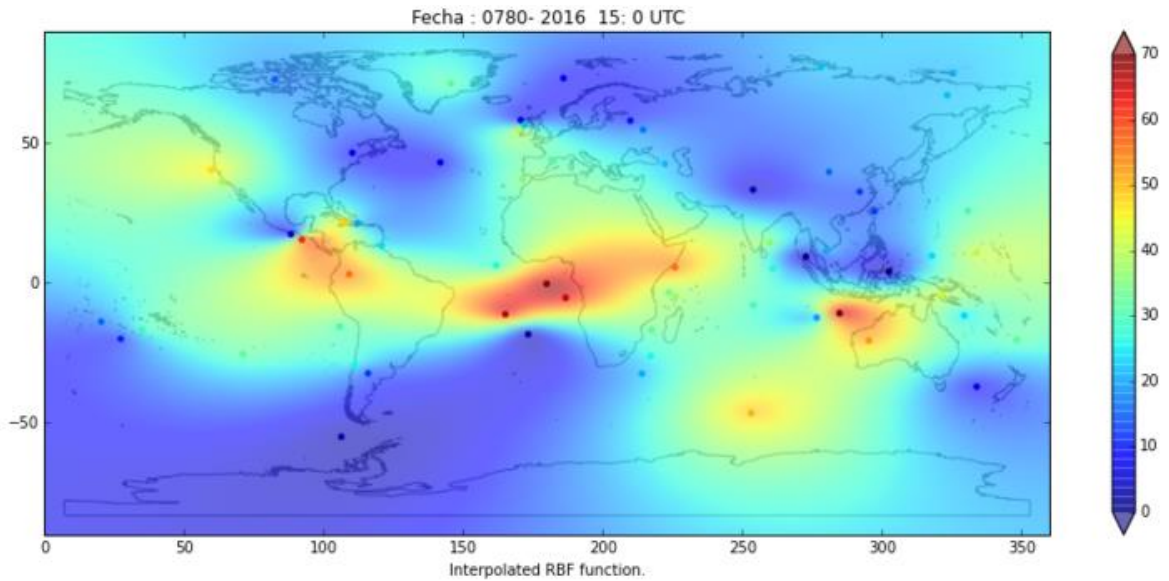
Figura 35. Mapa TEC del Instituto nacional de geofísica, geodesia y geografía de Bulgaria.



Fuente: [22]

El mapa que realizamos mediante el módulo fue generado por 64 estaciones, establecidas a lo largo de toda la superficie terrestre, aunque inicialmente la cantidad de estaciones son 100 algunas de ellas no son procesadas debido a que no tienen fichero de observación para el día establecido, en otros casos ocurre un error al calcular el valor de VTEC y tiene que realizarse un filtro en el script para que este valor siempre sea positivo, como se observa en la figura 36, en esta imagen podemos apreciar que en algunas estaciones como en Nicaragua e Indonesia los valores del VTEC no concuerdan con el comportamiento de la ionosfera para ese momento, esto ocurre por el ruido que tiene el pseudorange de los ficheros RINEX que es utilizado para calcular el VTEC.

Figura 36. Mapa TEC desarrollado con el módulo.



5.4 MÓDULO PREDICTIVO DEL ERROR IONOSFÉRICO

Este módulo se desarrolló para predecir el error ionosférico y contribuir en su corrección, para esto recopilamos un conjunto de datos de algunos días del año para posteriormente crear un modelo por medio de un algoritmo *Scikit-learn* y realizar predicciones para otro día. Para la recopilación de datos descargamos los ficheros de observación y navegación de seis días (046, 047, 048, 048, 050,051) del año 2016, el día 046 corresponde al 15 de febrero de ese año.

En el módulo de adquisición de datos se realizó su procesamiento con intervalos de tiempo de cada diez minutos durante el día, consiguiendo el comportamiento de la ionósfera a través el día en cada estación y continuamente generar su respectivo dataset. Los intervalos del tiempo utilizados de un día se dan desde el minuto 0 hasta el 1430, que corresponde a las 23:50.

Para este trabajo el valor de la predicción es continuo y por esto se utilizó el tipo de *machine learning* regresión, entonces entrenamos un modelo con muestras de datos y así pudimos clasificar nuevos datos en valores continuos. Para este módulo los modelos predictivos se crearon de acuerdo a nuestro objetivo de la predicción del error ionosférico.

5.4.1 Conjunto de Entrada. Ya recolectado los datos realizamos la selección del conjunto de entrada para entrenar al sistema por los siguientes criterios: filtrar cinco días para *train* y un día para *test* por medio de indexación de la columna día, eliminar la columnas del PRN y retardo ionosférico, y las líneas del dataset con los valores del VTEC negativo por medio de un procesado como el filtro de datos que proporciona DataFrame, logrando unos datos bien estructurados para tener mejores predicciones.

Y así pudiendo descartar características que no se ajustan exactamente a nuestro objetivo, para lograr que cada línea corresponde a una estación con sus mejores características y con la última columna que indica el valor del VTEC, como se muestra en la figura 37.

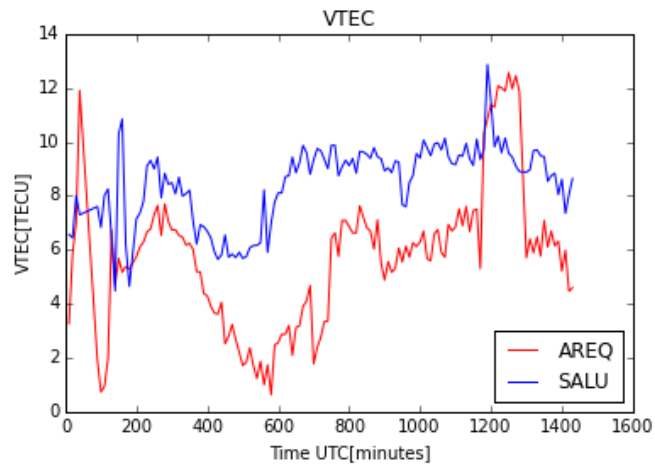
Figura 37. Extracto de conjunto de entrada para la predicción del día *test*, 051 del año 201.

	stations	day	minute	ipplon	ipplat	vtec
47776	ADIS	51	0	224.391026162	8.23346442717	3.202871
47777	ANKR	51	0	224.586788463	35.1105842187	4.380238
47779	ASCG	51	0	165.814042191	-8.82058596841	6.532043
47780	ASPA	51	0	9.49118986243	-7.92931782706	11.739835
47781	AUCK	51	0	358.650445652	-34.2745873391	1.850460
47782	bako	51	0	284.314462183	-3.83554028384	0.122747
47783	BJCO	51	0	181.33309909	4.13630583408	8.845927
47784	BOGT	51	0	115.601921148	-0.159869301391	3.941822
47787	CKIS	51	0	19.0009082508	-12.3368655806	16.122606
47788	COCO	51	0	275.142199519	-9.09579312255	0.520964
47789	CORD	51	0	121.035182879	-31.3479690897	1.398419

Después realizamos un análisis de los datos para visualizar de forma general cada característica, buscando tener las mejores para la predicción, así observamos un buen rango de valores para IPP longitud y latitud respectivamente.

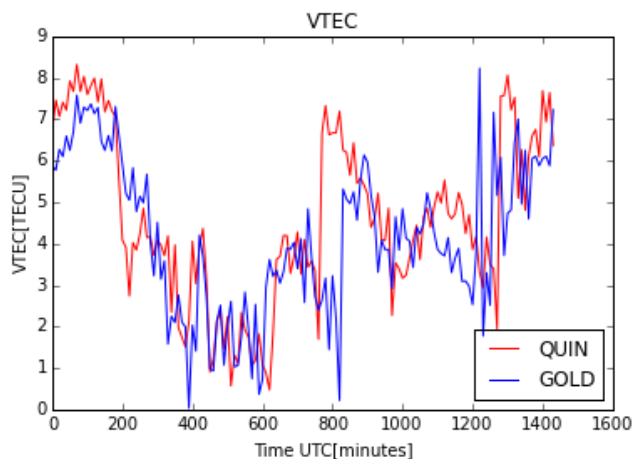
En la figura 38 se observa el comportamiento de la ionósfera de un parte de la zona tropical, exactamente en las estaciones AREQ situada en Arequipa en Perú y SALU en São Luis Brasil, del día 51 del año 2016, el comportamiento es bueno pero la diferencia que existen entre los valores de las dos estaciones es debido a la distancia entre ellas.

Figura 38. Comportamiento de la ionósfera para zona tropical con las estaciones AREQ y SALU.



Y así mismo para la zona templada con las estaciones QUIN de Quincy en Estados Unidos y Gold Stone en Estados Unidos, el comportamiento de la ionosfera como se muestra en la figura 39.

Figura 39. Comportamiento de la ionósfera para zona templada con las estaciones QUIN y GOLD .



Con este análisis observamos que tenemos buenos datos para realizar las predicciones con su respectivo modelo.

5.4.2 Formato de Especificación. Una de las fases de este módulo es el formato para la especificación de experimentos de modelado y predicción de la ionósfera, conformado por:

- Nombre de la estación para los días *train*
- Nombre de la estación para los días *test*
- Conjunto de datos *train*
- Conjunto de datos *test*
- Algoritmo de predicción como: *linear regression*, *random forest*, *decision tree* o *support vector regression*.

En este formato los usuarios podrán diligenciar las cinco entradas de acuerdo a su objetivo, volviéndose útil para los beneficiarios de este framework, ya que con las entrada de datos para su investigación pueden crear un modelo determinado para la estación con el algoritmo de predicción que deseen.

5.4.3 Entrenamiento y prueba del modelo. En la fase de entrenamiento y prueba del modelo, se divide el dataset en dos partes, *train* y *test* donde el *train* corresponde a cinco días de anteriores y el *test* al día siguiente, según el conjunto de entrada realizado en el formato anterior. Dividimos por las características de: VTEC, día, IPP longitud y latitud para *train* y *test*.

5.4.4 Experimentos con algoritmos de *machine learning*. Para los experimentos de *machine learning* se implementaron cuatro algoritmos: *Linear Regression*, *Support Vector Regression (SVR)*, *Decision Tree Regressor* y *Random Forest* de *Scikit Learn*, una biblioteca de análisis y minería de datos en Python.

El primer modelo se realizó mediante *linear regression*[37] buscando predecir un conjunto de datos mediante combinaciones lineales de las variable y utilizando los

parámetros:

fit_intercept, normalize, copy_X y n_jobs.

Para este modelo se realizó la implementación de *linearregression* de *Scikit learn*, para posteriormente inicializar la clase con los parámetros escogidos porque tenían un buen desempeño en las variables predichas. Después debemos ajustar este modelo con los datos de train recolectados y realizar las predicciones para el test, como se muestra en la figura 40.

Figura 40. Pasos para el algoritmo *linearregression* de *Scikit learn*.

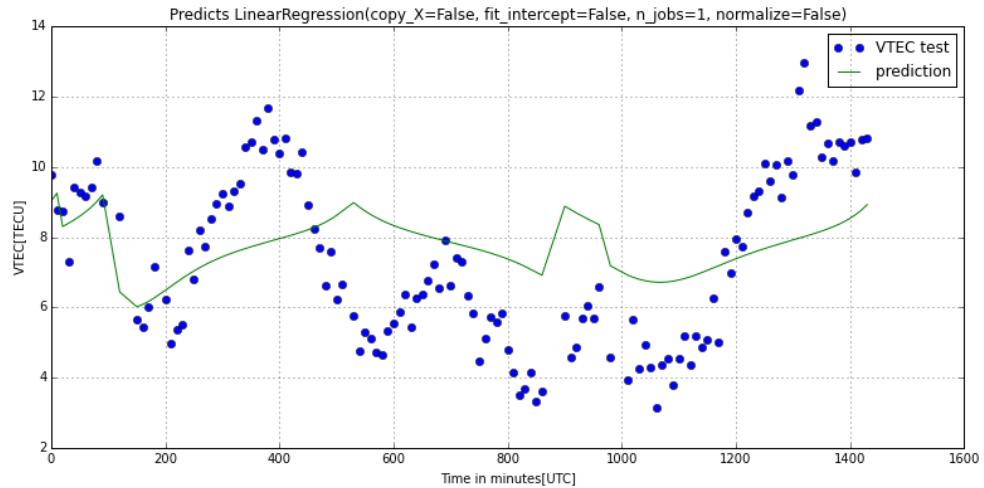
```
# Importar el modelo regresion lineal
from sklearn.linear_model import LinearRegression

# Iniciar la clase del modelo
model = LinearRegression(copy_X=False, fit_intercept=False, n_jobs=1, normalize=False)
# Ajustar el modelo con los datos train
model.fit(X_train, Y_train)
#Realizar las predicciones
model.predict(x_test)
```

Posteriormente realizamos una comparación del comportamiento de las estaciones: SOLO de Islas Salomón en Oceanía, ADIS de Addis Ababa en Etiopia y ANKR de Ankara en Turquía para predecir con *linearregression* y comparar con el conjunto de datos *test* del VTEC.

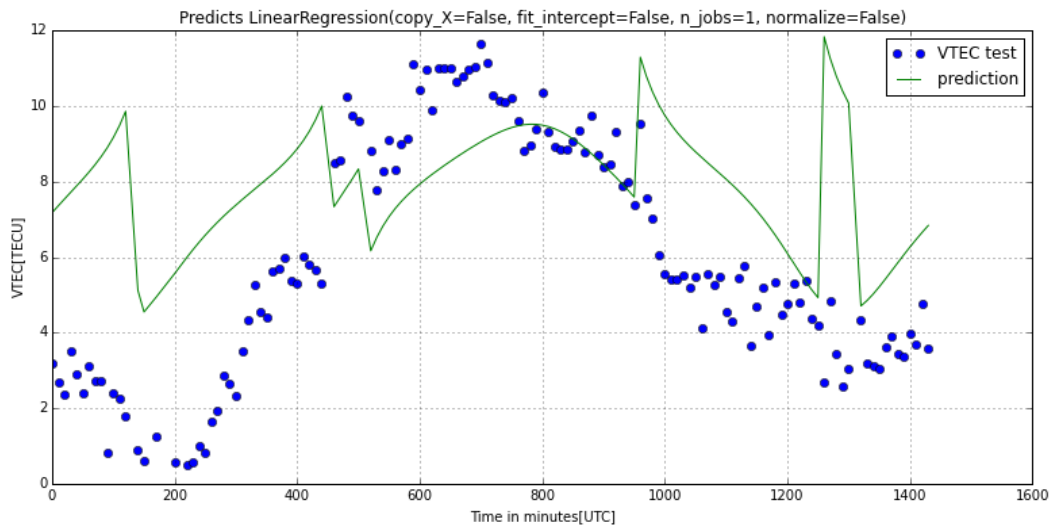
Para la estación SOLO el resultado de la predicción es regular porque no existe relación lineal entre la predicción y nuestros datos *test*, y su valor de RMSE(*root-mean-square error*) de 2.27814321439, como se muestra en la figura 41.

Figura 41. Predicción con *linearregression* para la estación SOLO de Islas Salomón en Oceanía.



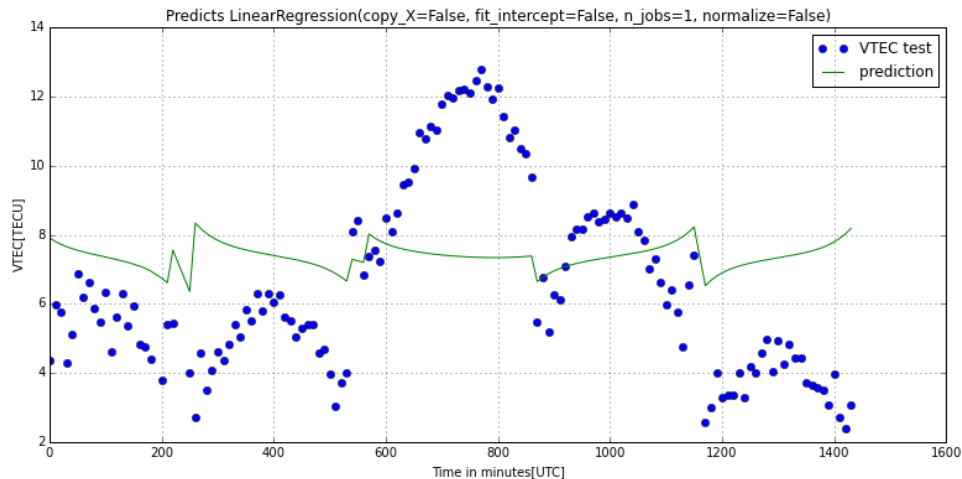
En la figura 42 se observa que la predicción para la estación ADIS no se asemeja a un comportamiento lineal con los datos *test* y teniendo como RMSE un alto valor de 3.33968345585.

Figura 42. Predicción con *linearregression* para la estación ADIS.



Para la estación IRKJ el valor de RMSE es 1.93143932388, observamos la poca relación que existe entre la predicción y el valor de test, como se muestra en la figura 43.

Figura 43. Predicción con *linearregression* para la estación ANKR.



El segundo modelo se realiza mediante *DecisionTreeRegressor* [38], un estimador que construye una predicción en base a las entradas con un conjunto de atributos de los datos generando una salida en base al árbol de decisión, sus parámetros son: *criterion*, *splitter*, *max_features*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *min_weight_fraction_leaf*, *max_leaf_nodes*, *random_state* y *presort*.

Este modelo de *DecisionTreeRegressor* se ejecuta su predicción mediante los siguientes pasos: importación del modelo para luego inicializar la clase con los parámetros seleccionados por el rendimiento de las predicciones para posterior ajustar el modelo con datos *train* para realizar las predicciones con *test*, como se muestra en la figura 44.

Figura 44. Pasos para el algoritmo DecisionTreeRegressor de scikit learn.

```
# Importar el modelo decision tree regresor
from sklearn.tree import DecisionTreeRegressor

# Iniciar La clase del modelo
model = DecisionTreeRegressor(random_state=None,min_samples_split=5,
                              max_depth=None,criterion="mse",splitter="best")

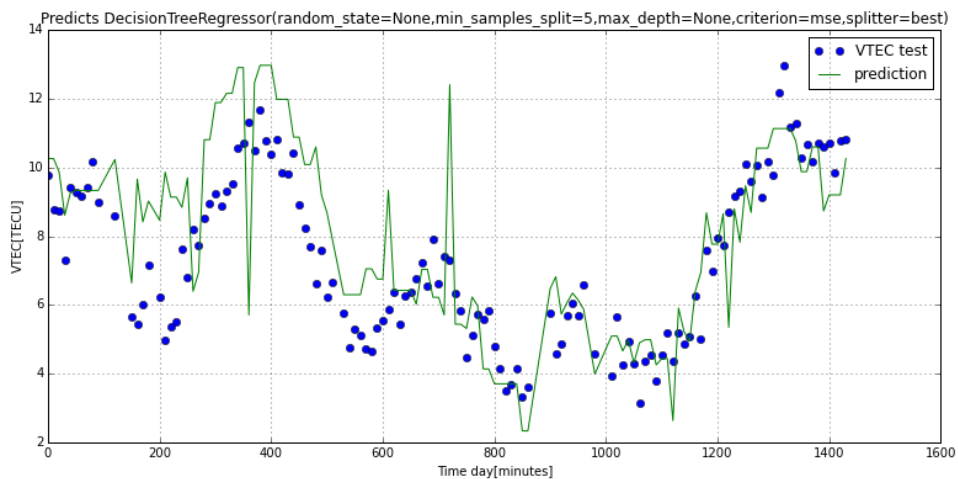
# Ajustar el modelo con Los datos train
model.fit(X_train, Y_train)

#Realizar Las predicciones
model.predict(x_test)
```

Realizamos una comparación del comportamiento de estaciones: SOLO de Islas Salomón en Oceanía, ADIS de Addis Ababa en Etiopia y ANKR de Ankara en Turquía para predecir con *DecisionTreeRegressor* y comparar con el conjunto de datos *test* del VTEC.

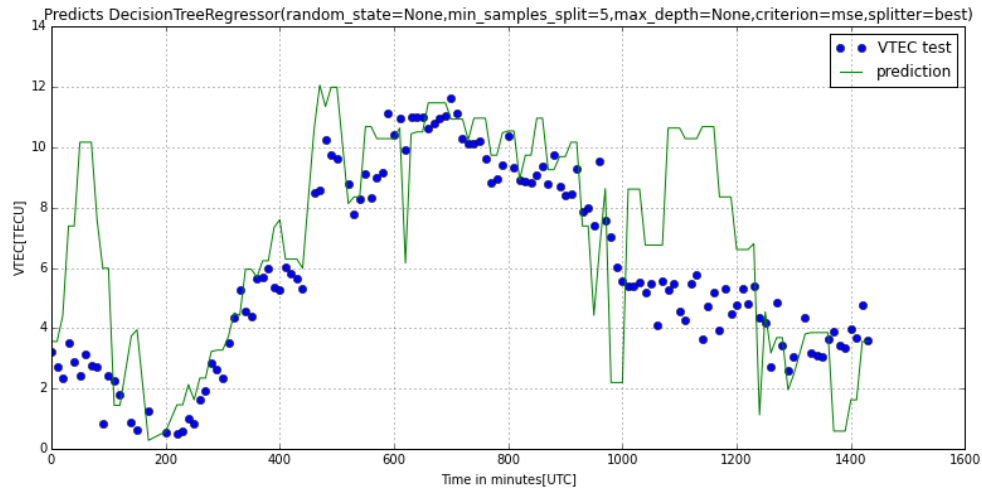
En la figura 45 se muestra la predicción de la estación SOLO con *DecisionTreeRegressor* y el conjunto de datos *test* del VTEC, evidenciamos como el estimador tiene la capacidad de predecir valores pero se presentan uno picos en donde la intensidad solar no deja predecir correctamente, esto se refleja en su RMSE de 1.716309.

Figura 45. Predicción con *DecisionTreeRegressor* para la estación SOLO de Islas Salomón en Oceanía.



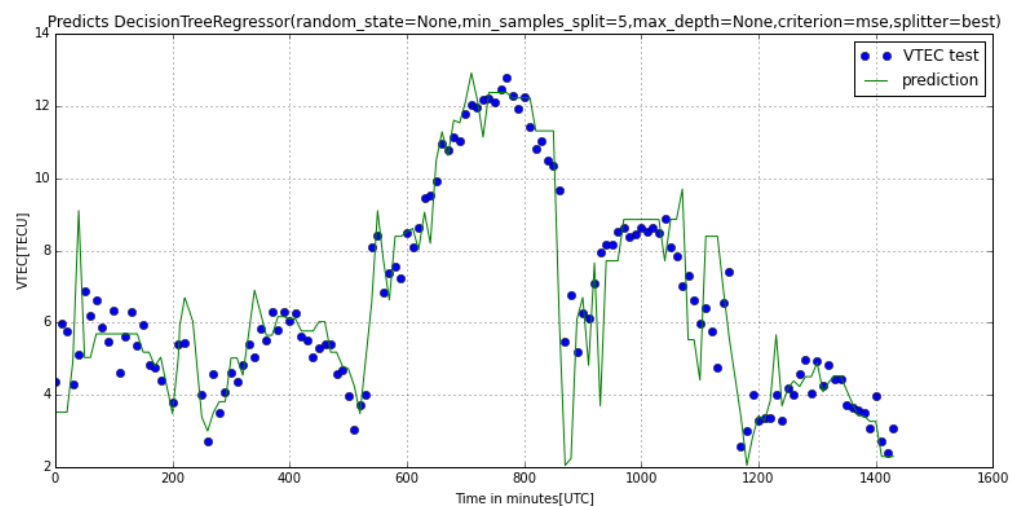
Para la predicción de la estación ADIS el valor de RSME 2.56736169396, nos indica que el comportamiento no se ajusta adecuadamente con el valor test, como se muestra en la figura 46.

Figura 46. Predicción con *DecisionTreeRegressor* para la estación ADIS.



Para la figura 47 se observa una mejor predicción de los valores del VTEC con respecto a los datos *test*, con un valor de RMSE de 1.17211596366.

Figura 47. Predicción con *DecisionTreeRegressor* para la estación ANKR



El tercer modelo se realizó mediante *Support Vector Regression* [40] para representar los puntos de los datos con uso de funciones lineales. Los parámetros de este modelo son: *C*, *epsilon*, *kernel*, *degree*, *gamma*, *coef0*, *shrinking*, *tol*, *cache_size*, *verbose* y *max_iter*.

La predicción de los datos mediante el modelo de *Support Vector Regression*[39]se realiza por la importación de su modelo para así inicializar su clase con los parámetros escogidos por su ajuste lineal y después poder ajustar el modelo con datos *train* para realizar las predicciones para el test, como se muestra en la figura 48.

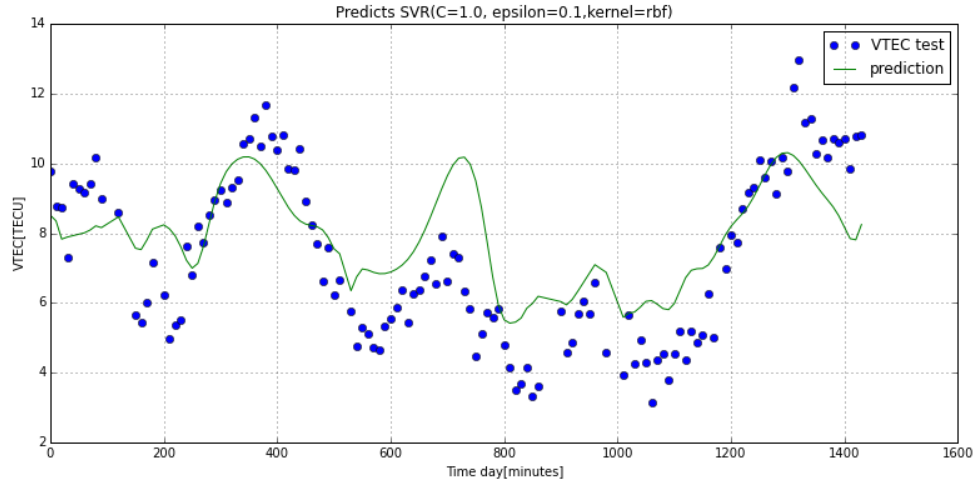
Figura 48. Pasos para el algoritmo Support Vector Regression de scikit learn.

```
# Importar el modelo de support vector regression
from sklearn.svm import SVR
from sklearn import svm

# Iniciar la clase del modelo
model = svm.SVR(C=1.0, epsilon=0.1, kernel='rbf')
# Ajustar el modelo con los datos train
model.fit(X_train, Y_train)
#Realizar las predicciones
model.predict(x_test)
```

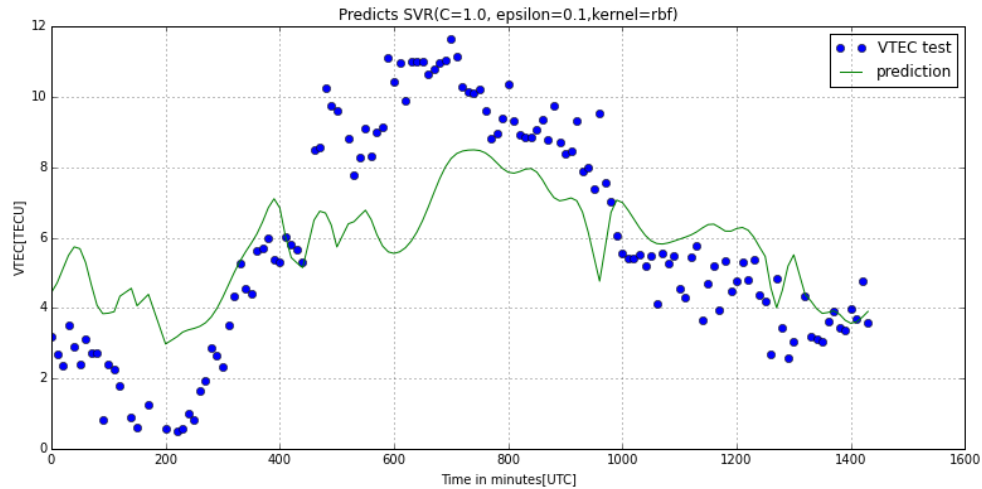
Entonces evidenciamos en la figura 49 que la predicción para el algoritmo SVR, no es precisa porque da un resultado lineal y los valores de los datos reales no se asemeja linealmente. Y su RMSE es de 1.67246212996

Figura 49. Predicción con *Support Vector Regression* para la estación SOLO de Islas Salomón en Oceanía.



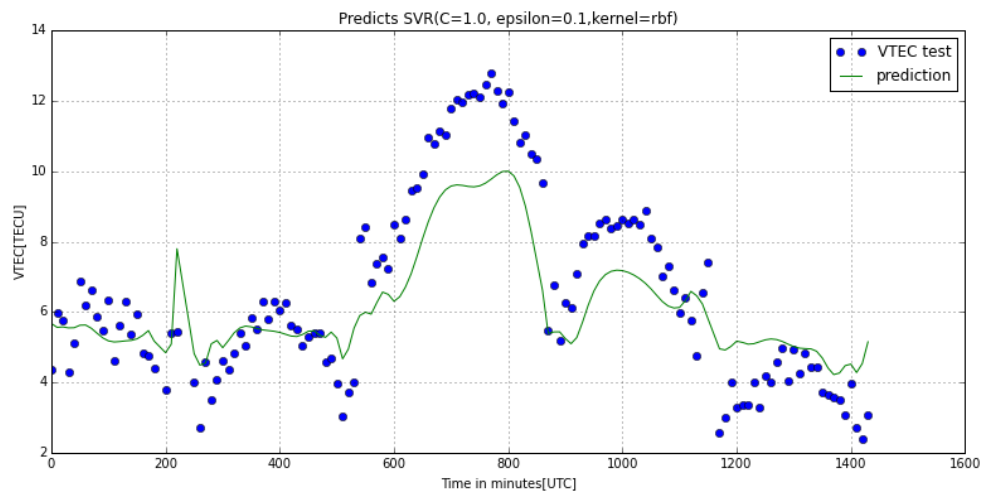
Para la figura 50 se observa el comportamiento de la predicción no se ajusta a los datos test, para lo cual su RMSE de 2.6120488213.

Figura 50. Predicción con *Support Vector Regression* para la estación ADIS.



La predicción para la estación ANKR se asemeja para algunos datos del conjunto de datos *test* y su RMSE es equivalente a 1.43251097127, como se muestra en la figura 51.

Figura 51. Predicción con *Support Vector Regression* para la estación ANKR.



El cuarto modelo se realizó mediante Random Forest Regressor[40], un estimador que ajusta el número de clasificación de decision tree en varias pequeñas muestras del conjunto de datos, sus parámetros son: *criterion*, *splitter*, *max_features*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *min_weight_fraction_leaf*, *max_leaf_nodes* y *presort random_state*.

Este modelo realiza la predicción de los datos mediante la importación de random forest regressor, para así inicializar su clase con los parámetros correctos poder lograr el mejor resultado en la predicción y luego poder ajustar el modelo con datos *train* para realizar las predicciones para el test, como se muestra en la figura 52.

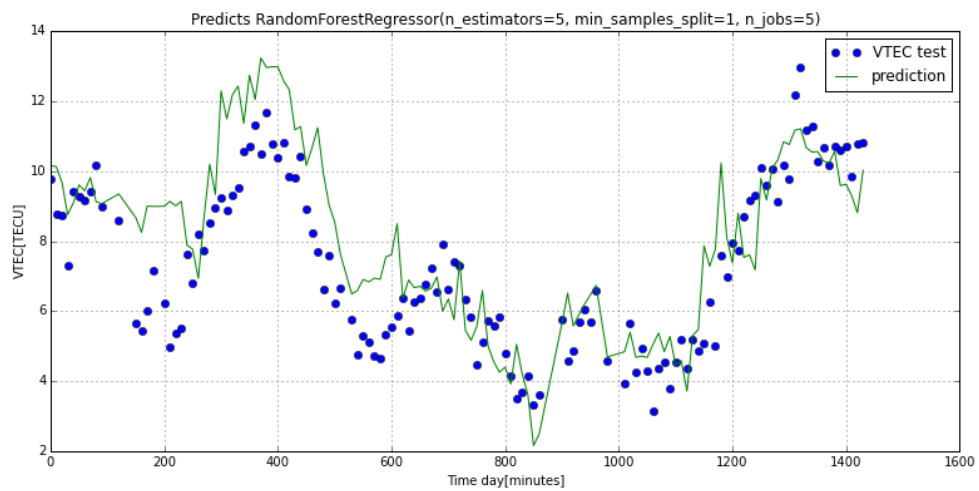
Figura 52. Pasos para el algoritmo Random Forest Regressor de *scikit learn*.

```
# Importar el modelo RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor

# Iniciar la clase del modelo
model = RandomForestRegressor(n_estimators=5, min_samples_split=1, n_jobs=5)
# Ajustar el modelo con los datos train
model.fit(X_train, Y_train)
# Realizar las predicciones
model.predict(x_test)
```

Para la figura 53 se observa que el estimador *Random Forest Regressor*, está en capacidad de predecir el error ionosférico porque su resultado se asemeja al valor *test* del VTEC, dando así un RSME de 1.57636784193, un valor considerable por el rango del VTEC.

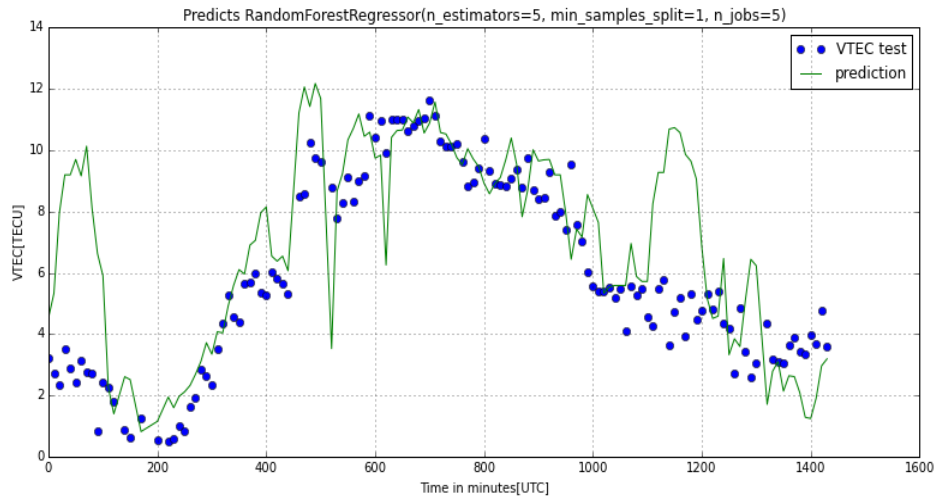
Figura 53. Predicción con Random Forest Regressor para la estación SOLO, Islas Salomón en Oceanía.



El comportamiento de la predicción para la estación ADIS es similar con respecto a los datos *test*

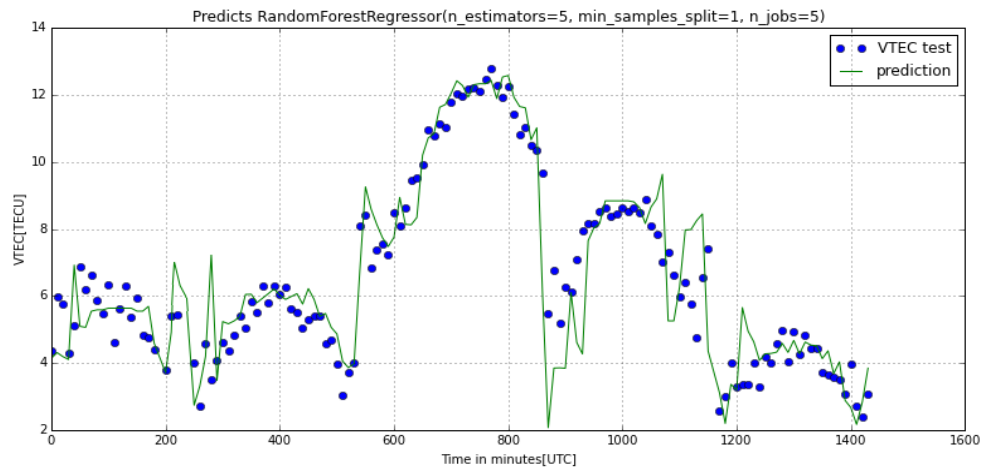
pero existen picos que indican que el modelo para estos datos no se ajustó adecuadamente, dando así como RMSE de 2.45182919051, como se muestra en la figura 54.

Figura 54. Predicción con Random Forest Regressor para la estación ADIS.



Para la predicción de la estación ANKR su comportamiento es bastante similar a los datos *test* del VTEC, pero existen algunos picos donde el modelo no se ajustó correctamente, se muestra en la figura 55. El RMSE es de 1.07737497168, un valor pequeño para demostrar la veracidad de la predicción.

Figura 55. Predicción con Random Forest Regressor para la estación ANKR.



5.4.5 Evaluación de modelos. Una vez realizadas las predicciones, podemos calcular el error entre el valor de las predicciones y los valores reales. Esta evaluación se realizó por las métricas de regresión y RMSE(*root-mean-square*

error).

5.4.5.1 Métricas de regresión: Las métricas de regresión de *scikit learn* utilizadas son: varianza explicada, coeficiente de determinación y *RMSE*

La primera métrica utilizada es varianza explicada, [41] mide desviación de los valores reales y la predicción. La segunda métrica es *RMSE* (*root-mean-square error*) [42] corresponde a la raíz de los valor de las medidas de la media de los cuadrados de los errores. Y la tercera métrica es coeficiente de determinación [43] que proporciona la medida que indica lo bien que encajan los datos al modelo, por medio de a una línea de regresión.

Por acuerdo a las tres métricas de regresión realizamos la predicción con cinco días para train y un día para test, para la estación SOLO de Islas Salomón en Oceanía situada en zona tropical con cuatro algoritmos, y lograr comparar estas métricas. Como resultado obtuvimos que para el algoritmo de *random forest* hay un mejor rendimiento debido a tiene un menor valor de error y también para las métricas de varianza explicada y coeficiente de determinación sus valores son los mejores debido a su rendimiento, como se muestra en la figura 56.

Figura 56. Métricas de regresión para estación SOLO de Islas Salomón en Oceanía.

ALGORITMOS	<i>RMSE</i>	Coeficiente de determinación	Varianza
<i>Linear regression</i>	2.285063	0.082397	0.159224
<i>decision tree</i>	1.721245	0.479352	0.549241
<i>SVR</i>	1.672462	0.508446	0.567455
<i>Random forest</i>	1.542411	0.576482	0.664746

También realizamos otra predicción para la estación ASPA de Pago Pago en

Estados Unidos, situada en zona templada del norte y el resultado fue que el mejor rendimiento lo tiene a random forest obteniendo los valores más significativos para las métricas de regresión, como se muestra en la figura 57.

Figura 57. Métricas de regresión para estación ASPA de Pago Pago en Estados Unidos.

ALGORITMOS	RMSE	Coeficiente de determinación	Varianza
<i>Linear regression</i>	1.957456	0.075909	0.076374
<i>decision tree</i>	1.337161	0.568781	0.569435
<i>SVR</i>	1.478852	0.472551	0.500088
<i>Random forest</i>	1.171749	0.668869	0.670798

Podemos concluir que la zona tropical tiene el RMSE es mayor y no predice con mayor exactitud los mejores valores porque su intensidad solar es mayor.

5.4.5.2 Reutilización de modelos locales: Realizamos una comparación por RMSE para cinco estaciones del trópico como: ADIS de Etiopia, MBAR de Uganda, XMIS de Australia, PNGM de Papúa Nueva Guinea , SOLO de Oceanía y otras cinco de zona templada con su valor de latitud: QUIN en Estados Unidos, ANKR en Turquía, SASS en Alemania, IRKJ en Rusia, GRAS en Francia con un conjunto de datos *train* de cinco días y *test* para un día, usando el algoritmo de *random forest*. El resultado es, que si se entrena y prueba con la misma estación, el RMSE es pequeño porque se está utilizando el modelo entrenado y probado con las mismas características, como se evidencia en las figuras 58 y 59.

Figura 58. Cálculo del RMSE con conjunto de datos Train y test para estaciones de la zona tropical.

<i>TRAIN/ TEST</i>	ADIS Latitud(9.035)	MBAR Latitud (-0.60138)	XMIS Latitud(-10.449722)	PNGM Latitud (-2.04305)	SOLO Latitud (-9.43472)
ADIS Latitud (9.035)	2.5710	4.2983	6.576348	4.609478	4.51318
MBAR Latitud (-0.60138)	3.5621	1.2619	5.484347	4.663816	3.04982
XMIS Latitud(-10.449722)	3.1459	3.1913	1.533279	1.931020	3.01249
PNGM Latitud(-2.0430556)	5.7602	5.4973	3.521999	1.369138	4.53025
SOLO Latitud (-9.43472)	3.6109	2.7883	5.875741	3.296531	1.626613

Figura 59. Cálculo del RMSE con conjunto de datos Train y test para estaciones de la zona templada.

<i>TRAIN/TEST</i>	QUIN Latitud(39.9744)	ANKR Latitud(39.88750)	SASS Latitud(54.51333)	IRKJ Latitud(52.21888)	GRAS Latitud(43.754722)
QUIN Latitud (39.9744)	1.06095	5.41800	2.44891	2.11644	5.10021
ANKR Latitud (39.88750)	2.82611	1.06234	2.96765	1.89758	3.42009
SASS Latitud (54.51333)	2.16492	5.43495	1.07847	2.40992	4.94940
IRKJ Latitud (52.21888)	2.01948	3.46119	2.44551	1.06058	4.34262
GRAS Latitud (43.754722)	3.10760	3.38653	3.04290	1.94167	0.58400

De acuerdo a esta evaluación se observa que para la zona templada existe menor error que en la zona tropical debido a que su intensidad solar es menor.

6. CONCLUSIONES

- Se construyó un framework que integro módulos de adquisición y procesamiento de datos, para generar mapas TEC y buscar predecir el valor del VTEC.
- Para la producción de dataset evidenciamos que en algunas ocasiones los ficheros de navegación no existían para determinada estación en la red IGS, además de la dependencia del VTEC de la ubicación del Primer punto de perforación de la ionósfera (IPP) y el tiempo en el que es calculado.
- En el módulo de creación de mapas TEC mostramos que la función que mejor genero la imagen sobre la interpolación de los puntos (VTEC) fue la RBF, ya que solamente tiene en cuenta la distancia euclidiana entre los puntos, también observamos que el comportamiento de la interpolación mejora con una mayor cantidad de estaciones que conformen nuestro dataset. El generar mapas TEC no depende del valor del pseudorange, ya que se pueden tomar dataset con valores del pseudorange sin ruido para construir mejores mapas.
- A partir de las predicciones llevadas a cabo en la fase final del proyecto, se observa la utilidad del formato para la especificación de experimentos de modelado y predicción de la ionósfera para la realizar predicciones con más exactitud de una zona específica.
- El modelo ionosférico con mayor desempeño se implementó, con el algoritmo *random forest* de *Scikit learn*, mostrando predicciones semejantes a valores reales del VTEC pero también mostrando picos en donde existe gran intensidad solar.

- Para la corrección del error ionosférico podemos dar uso de este framework, pudiendo ahorrar el uso de receptores para la señal con dos frecuencias.
- Nuestro proyecto de grado nos motivó a realizar investigación sobre GNSS, además la herramienta desarrollada facilitara el comportamiento de la ionósfera por medio de la predicción, generara mapas TEC, pre-procesado de ficheros RINEX y apoyara a próximos trabajos e investigaciones relacionados con GNSS, VTEC, *machine learning*.

7. RECOMENDACIONES

- Para un flujo de trabajo óptimo se recomienda seleccionar una mayor cantidad de estaciones, por lo tanto un mayor número de ficheros, además para obtener un mejor mapa TEC se aconseja utilizar la función RBF para interpolar el valor VTEC y el mejor algoritmo de predicción es *random forest*
- Para reducir el tiempo de procesamiento de un día de los ficheros RINEX, se recomienda trabajar con modelos de arquitectura de alto rendimiento.
- De acuerdo al enfoque del proyecto se recomienda buscar y trabajar con una red de estaciones GNSS que se ajuste adecuadamente.
- Por último es recomendable para el módulo de predicción del VTEC tener un mejor procesamiento de datos, para poder aprovechar el modelo ionosférico creado, ya que un día podían haber intervalos de tiempo en las que el valor no se encontraba porque era negativo.

REFERENCIAS BIBLIOGRÁFICAS

[1] Modelo ionosférico Klobuchar – Navipedia. Citado Julio 7 del 2.015, from http://www.navipedia.net/index.php/Klobuchar_Ionospheric_Model.

[2] Modelo ionosférico NeQuick, – Navipedia. Citado Julio 7 del 2.015, from http://www.navipedia.net/index.php/NeQuick_Ionospheric_Model.

[3] Sistema global de navegación por satélite - Gage. Citado Julio 9 del 2.015, from <http://www.gage.upc.edu/drupal6/tutorials>.

[4] Sistema global de navegación por satélite - Wikipedia. Citado Junio 9 del 2.015, from http://es.wikipedia.org/wiki/Sistema_global_de_navegaci%C3%B3n_por_sat%C3%A9lite

[5] STS ha acabado con los dispositivos de seguimiento - MD Press. Citado en Agosto 28 del 2.015, from <http://mihaeladanpress.ro/wp/2014/01/braila-doar-sts-aparatura-de-localizare/>.

[6] Proceso de triangulación - Taringa. Citado en Agosto 24 del 2.015, from <http://www.taringa.net/post/info/7134490/Como-funciona-un-gps.html>.

[7] Navegación por posicionamiento GNSS: Navegación por satélite GPS - Ingeniería Aeroespacial Universidad de Sevilla. Citado en Julio 7 del 2.105, from <http://aero.us.es/na/files1112/T5NA.pdf>.

[8] GPS error presupuesto - Universidad Estatal Montana. Citado en Julio 9 del

2.015, from <http://www.montana.edu/gps/understd.html>.

[9]De la Tierra ionósfera - Centro Solar Stanford. Citado en Julio 25 del 2.015, from <http://solar-center.stanford.edu/SID/activities/ionosphere.html>.

[10]Multipath- Navipedia. Citado en Junio 18 del 2.015, from <http://www.navipedia.net/index.php/Multipath>.

[11] Perdiguier, R, J. Zurutuza'v, M. C. Ruiz y M. J. Sevilla. Estrategias de cálculo del retardo troposférico y su influencia en procesamiento GPS de alta precisión .Asamblea Hispano Portuguesa de Geodesia y Geofísica. [En línea]. Citado en Junio 18 del 2.015, from http://digital.csic.es/bitstream/10261/23301/1/N203_2008_6.pdf.

[12] Sistema de posicionamiento global - Tamara Giménez Rodríguez María Elena Ros Bernabeu, Gravitación y Astrofísica. Citado en Julio 10 del 2.015, from http://webs.um.es/bussons/GPSresumen_TamaraElena.pdf.

[13] GNSS format descriptions - Grupo GaGe, Universidad Politécnica de Catalunya. Citado en Junio 2 del 2.015, from http://gage14.upc.es/gLAB/HTML/GPS_Navigation_Rinex_v2.11.html.

[14] GNSS format descriptions - Grupo GaGe, Universidad Politecnica de Catalunya. Citado en Junio 2 del 2.015, from http://gage14.upc.es/gLAB/HTML/Observation_Rinex_v2.11.html.

[15]Redes de todo el mundo de IGS – IGS. Citado en Agosto 11 del 2.015, from <http://igs.org/network>

[16] IGS - Servicio Internacional GNSS. Citado en Julio 10 del 2.015, from

<https://igscb.jpl.nasa.gov/overview/viewindex.html>

[17] Productos IGS para GPS satélites efemérides - IGS. Citado en Agosto 13 del 2.015, from <http://igs.org/products>.

[18] Capas de la ionósfera - XE2PNA. Citado Julio 3 del 2.015, from http://qsl.net/xe2pna/html_files/capas_de_la_inosfera.html

[19] TEC - Digital C.SIC Ciencia en Abierto. Citado en Julio 25 del 2.015, from <http://digital.csic.es/bitstream/10261/6637/1/ionósfera.pdf>,

[20] Real Time Ionospheric Map - NASA Jet Propulsión Laboratory. Citado en Diciembre 1 de 2.016, from <http://iono.jpl.nasa.gov/>.

[21] Tec global map, Space Weather Services - Australian Government. Citado en Febrero 8 de 2.016, from <http://www.sws.bom.gov.au/Satellite/2/1/5>.

[22] Mapas Tec - Instituto Nacional de Geofísica, Geodesia y Geografía. Citado en Febrero 8 de 2.016, from http://www.geophys.bas.bg/index_bg.php.

[23] Transforming the Way the World Works – Trimble. Citado en Julio 26 del 2.015, from http://www.trimble.com/positioning-services/ionosphere_map.aspx.

[24] GPSTK - Aplicado a la Universidad de Texas a Austin. Citado en Julio 30 del 2.015, from <http://www.gpstk.org/bin/view/Documentation/WebHome>.

[25] Rob, Schapire. (2008). *Teoría e Aprendizaje Automático*, Departamento de Ciencia de Computación, Universidad de Princeton. Citado en Julio 30 del 2.015, from

http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf.

[26] Sklearn pipeline - *Scikit Learn*. Citado en Julio 27 del 2.015, from <http://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>.

[27] Framework de Software - Wikipedia. Citado en Agosto 24 del 2.015, from https://en.wikipedia.org/wiki/Software_framework.

[28] Descarga de RINEX – CDDIS citado en Agosto 3 del 2.015, from <ftp://cddis.gsfc.nasa.gov/pub/gps/data/daily/>.

[29] The GPS Toolkit documentation – GPSTK. Citado en Enero 27 de 2.016, from <http://www.gpstk.org/pythondoc/gpstk.GPSEphemerisStore.html>.

[30] The GPS Toolkit documentation. – GPSTK. Citado en Enero 27 de 2.016, from <http://www.gpstk.org/pythondoc/gpstk.Rinex3ObsData.html#gpstk.Rinex3ObsData.getObs>

[31] Newcastle University Student Publishing. Citado en Febrero 4 de 2.016, from <https://www.students.ncl.ac.uk/keith.newman/r/maps-in-r.c>

[32] Interpolation, Reference guide - SciPy. Citado en Febrero 16 de 2.016, from <http://docs.scipy.org/doc/scipy/reference/interpolate.html>.

[33] Rbf, Interpolation, Reference guide - SciPy. Citado en Febrero 16 de 2.016, from <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.Rbf.html>.

[34] Griddata, Interpolation, Reference guide - SciPy. Citado en Febrero 16 de 2.016, from <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.griddata.html>

[35] Bisplrep, Interpolation, Reference guide - SciPy. Citado en Febrero 16 de 2.016, from <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.bisplrep.html>.

[36] Marcel H. De Canck, (2006), Ionosphere Properties and Behaviors, antennex Issue No. 113. Citado en Enero 10 de 2016 <https://www.antennex.com/prop/prop0906/prop0906-guests.pdf> .

[37] Algoritmo *linear regression* - *Scikit learn*. Citado Febrero 10 del 2.016, from http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression.

[38] Algoritmo *decision tree regressor* - *Scikit learn*. Citado Febrero 10 del 2.016, from <http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>

[39] Algoritmo *support vector regression* - *Scikit learn*. Citado Febrero 10 del 2.016, from <http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>.

[40] Algoritmo random forest - *Scikit learn*. Citado Febrero 11 del 2.016, from <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor>.

[41] Métricas de regresión *explained variance score* - *Scikit learn*. Citado Febrero 12 del 2.016, from http://scikit-learn.org/stable/modules/generated/sklearn.metrics.explained_variance_score.html#sklearn.metrics.explained_variance_score.

[42] Métrica de regresión *mean squared error* - *Scikit learn*. Citado Febrero 12 del 2.016, from http://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html#sklearn.metrics.mean_squared_error.

[43] Métrica de regresión *coefficient of determination* - *Scikit learn*. Citado Febrero 11 del 2.016, from http://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html#sklearn.metrics.r2_score.

BIBLIOGRAFÍA

BLEWITT, G. Basics of the GPS Technique: Observation Equations. In B. Johnson, *Geodetic Applications of GPS* (pp. 10-54). Sweden: Nordic Geodetic Commission.1997

CDDIS Descarga de RINEX - [en línea] [citado en Agosto 3 del 2.015]
Disponibile en: <ftp://cddis.gsfc.nasa.gov/pub/gps/data/daily/>.

CENTRO SOLAR STANFORD De la Tierra ionósfera -. [en línea] [Citado en Julio 25 del 2.015], Disponible en: <http://solar-center.stanford.edu/SID/activities/ionosphere.html>.

DE CANCK, Marcel H. Ionosphere Properties and Behaviors, *antenneX* Issue No. 113. 2006 [en línea] [Citado en enero 10 de 2.016] Disponible en: <https://www.antennex.com/prop/prop0906/prop0906-guests.pdf> .

ESI UNIVERSIDAD DE SEVILLA. Navegación por posicionamiento GNSS: Navegación por satélite GPS - Ingeniería Aeroespacial Universidad de Sevilla. [en línea] [Citado en Julio 17 de 2.016] Disponible en: <http://aero.us.es/na/files1112/T5NA.pdf>.

GAGE. Sistema global de navegación por satélite - Citado Julio 9 del 2.015, from <http://www.gage.upc.edu/drupal6/tutorials>.

GAUSSIRAN, T. An Open Source Toolkit for GPS Processing, Total Electron Content Effects, Measurements and Modeling. 2004

GEOPHYS. Mapas Tec - Instituto Nacional de Geofísica, Geodesia y Geografía. [en línea] [Citado en Febrero 08 de 2.016] Disponible en: http://www.geophys.bas.bg/index_bg.php.

GIFFARD, R. Estimation of gps ionospheric delay using LI code and carrier phase observables. 31 st Annual Precise Time and Time Interval p. 14. 1999

GIMÉNEZ RODRÍGUEZ Tamara ROS BERNABEU María Elena, Sistema de posicionamiento global - Gravitación y Astrofísica. [en línea] [Citado en Julio 10 del 2.015] Disponible en: http://webs.um.es/bussons/GPSresumen_TamaraElena.pdf.

GNSS format descriptions - Grupo GaGe, Universidad Politecnica de Catalunya. [en línea] [Citado en Junio 2 del 2.015], Disponible en: http://gage14.upc.es/gLAB/HTML/Observation_Rinex_v2.11.html.

GNSS format descriptions - Grupo GaGe, Universidad Politécnica de Catalunya. [en línea] [Citado en Junio 2 del 2.015], Disponible en: http://gage14.upc.es/gLAB/HTML/GPS_Navigation_Rinex_v2.11.html.

GPS error presupuesto - Universidad Estatal Montana. [en línea] [Citado en Junio 2 del 2.015] Disponible en: <http://www.montana.edu/gps/understd.html>.

GPSTK - Aplicado a la Universidad de Texas a Austin. [en línea] [Citado en Junio 2 del 2.015], Disponible en: <http://www.gpstk.org/bin/view/Documentation/WebHome>.

GPSTK The GPS Toolkit documentation. -. [en línea] [Citado en Enero 27 del 2.015], Disponible en:

<http://www.gpstk.org/pythondoc/gpstk.Rinex3ObsData.html#gpstk.Rinex3ObsData.getObs>

GPSTK. The GPS Toolkit documentation – [en línea] [Citado en Enero 27 del 2.015], Disponible en: <http://www.gpstk.org/pythondoc/gpstk.GPSEphemerisStore.html>.

HERN, M. JUAN, J M. SANZ, J. Application of ionospheric tomography to real-time GPS carrier-phase ambiguities resolution, at scales of 400-1000 km and with high geomagnetic activity. 2012

IGS - Servicio Internacional GNSS. [en línea] [Citado en Julio 10 del 2.015] Disponible en: <https://igscb.jpl.nasa.gov/overview/viewindex.html>

IGS. Productos IGS para GPS satélites efemérides. [en línea] [Citado en Agosto 13 del 2.015], Disponible en: <http://igs.org/products>.

IGS. Redes de todo el mundo de IGS – [en línea] [Citado en Agosto 11 del 2.015] Disponible en: <http://igs.org/network>

NAVIPEDIA. Modelo ionosférico NeQuick, [en línea] [Citado en Julio 7 de 2.016] Disponible en: http://www.navipedia.net/index.php/NeQuick_Ionospheric_Model.

NAVIPEDIA. Modelo ionosférico Klobuchar [en línea] [Citado en Julio 7 de 2.016] Disponible en: http://www.navipedia.net/index.php/Klobuchar_Ionospheric_Model.

NAVIPEDIA. Multipath- [en línea] [Citado en Julio 18 de 2.016] Disponible en: <http://www.navipedia.net/index.php/Multipath>.

NEWCASTLE University Student Publishing. [en línea] [Citado en febrero 04 de 2.016] Disponible en: <https://www.students.ncl.ac.uk/keith.newman/r/maps-in-r> .

NORSUZILA, Y. ABDULLAH, M. ISMAIL, M. IBRAHIM, M. ZAKARIA, Z. Total Electron Content (TEC) and Estimation of Positioning Error Using Malaysia Data. Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London

PERDIGUER, R, J. ZURUTUZA'V, M. C. RUIZ Y M. J. SEVILLA. Estrategias de cálculo del retardo troposférico y su influencia en procesamiento GPS de alta precisión .Asamblea Hispano Portuguesa de Geodesia y Geofísica. [en línea]. [Citado en Junio 18 del 2.015], Disponible en: http://digital.csic.es/bitstream/10261/23301/1/N203_2008_6.pdf.

RBF, Interpolation, Reference guide - SciPy. [en línea] [Citado en Febrero 16 de 2.016], Disponible en: <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.Rbf.html>.

REAL TIME Ionospheric Map - NASA Jet Propulsión Laboratory. [en línea] [Citado en Diciembre 1 de 2.016], Disponible en: <http://iono.jpl.nasa.gov/>.

ROB, SCHAPIRE. Teoría e Aprendizaje Automático, Departamento de Ciencia de Computación, Universidad de Princeton. 2008 [en línea] [Citado en Julio 30 del 2.015], Disponible en: http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf.

SCIKIT LEARN Algoritmo random forest. [en línea] [Citado Febrero 11 del 2.016] Disponible en: <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor>.

[html#sklearn.ensemble.RandomForestRegressor](#).

SCIKIT LEARN Algoritmo support vector regression -. [en línea] [Citado Febrero 10 del 2.016] Disponible en: <http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>.

SCIKIT LEARN. Algoritmo decision tree regressor – [en línea] [Citado Febrero 10 del 2.016] Disponible en:, <http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>

SCIKIT LEARN. Algoritmo linear regression - [en línea] [Citado Febrero 10 del 2.016] Disponible en:, http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression.

SCIKIT LEARN. Métrica de regresión coefficient of determination. [en línea] [Citado en febrero 11 de 2.016] Disponible en: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html#sklearn.metrics.r2_score.

SCIKIT LEARN. Métrica de regresión mean squared error - [en línea] [Citado en Febrero 12 de 2.016] Disponible en: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html#sklearn.metrics.mean_squared_error.

SCIKIT LEARN. Métricas de regresión explained variance score [en línea] [Citado en Febrero 12 de 2.016] Disponible en: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.explained_variance_score.html#sklearn.metrics.explained_variance_score.

SCIKIT LEARN. Sklearn pipeline - [en línea] [Citado en Julio 10 del 2.015]
Disponible en: <http://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>.

SCIPY. Bsplrep, Interpolation, Reference guide - – [en línea] [Citado Febrero 16 del 2.016] Disponible en: <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.bsplrep.html>.

SCIPY. Griddata, Interpolation, Reference guide - [en línea] [Citado en Febrero 16 de 2.016, from <http://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.griddata.html>

SCIPY. Interpolation, Reference guide -. [en línea] [Citado en Febrero 16 de 2.016] Disponible en: <http://docs.scipy.org/doc/scipy/reference/interpolate.html>.

SHAGIMURATOV, I .CHERNYAK, Y. ZAKHARENKOVA, I. YAKIMOVA, G. .Use of Total Electron Content Maps for Analysis of Spatial – Temporal Structures of the Ionosphere. (pp. 656-662).2013

STS ha acabado con los dispositivos de seguimiento - MD Press. [en línea] [Citado en Julio 10 del 2.015] Disponible en: <http://mihaeladanpress.ro/wp/2014/01/braila-doar-sts-aparatura-de-localizare/>.

TARINGA Proceso de triangulación -. [en línea] [Citado en Agosto 24 del 2.015] Dispñible en: <http://www.taringa.net/post/info/7134490/Como-funciona-un-gps.html>.

TEC - Digital C.SIC Ciencia en Abierto. [en línea] [Citado en Julio 25 del 2.015],

Disponible en: <http://digital.csic.es/bitstream/10261/6637/1/ionósfera.pdf>,

TEC dual frequency technique and TEC map at ionosphere layer using Malaysian data. Proceedings - 2009 International Conference on Future Computer and Communication, ICFCC 2009. (pp 589-593).

TEC GLOBAL MAP, Space Weather Services - Australian Government. Citado en Febrero 8 de 2.016, from

TRIMBLE. Transforming the Way the World Works – [en línea] [Citado en Julio 26 del 2.015], Disponible en: http://www.trimble.com/positioning-services/ionosphere_map.aspx.

WIENIA, R J. Use of Global Ionospheric Maps for Precise Point Positioning. 2008

WIKIPEDIA. Framework de Software - [en línea] [Citado en Agosto 24 del 2.015], Disponible en: https://en.wikipedia.org/wiki/Software_framework.

WIKIPEDIA.. Sistema global de navegación por satélite - Citado [en línea] [Citado en Julio 09 del 2.015] Disponible en: http://es.wikipedia.org/wiki/Sistema_global_de_navegaci%C3%B3n_por_sat%C3%A9lite

XE2PNA. Capas de la ionósfera - [en línea] [Citado Julio 3 del 2.015] Disponible, http://qsl.net/xe2pna/html_files/capas_de_la_ionosfera.html

YA'ACOB, N. ABDULLAH, M. ISMAIL, M. GPS vertical total electron content 2009