

**Universidad Autónoma de Querétaro**  
**Facultad de Ciencias Naturales**  
**Maestría en Ciencias Biológicas**

Evaluación de la fluorescencia inducida por el sol y productos  
satelitales de PPB para estimar la captura de carbono en un  
ambiente semiárido

**Tesis**

Que como parte de los requisitos para obtener el Grado de

**Maestra en Ciencias Biológicas**

Presenta

**Jaquelin Guadalupe Ramos García**

Dirigida por

**Mónica Cervantes Jiménez**

Querétaro, Querétaro

octubre de 2023



Dirección General de Bibliotecas y Servicios Digitales  
de Información



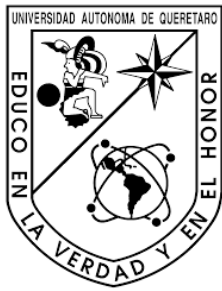
Evaluación de la fluorescencia inducida por el sol y  
productos satelitales de PPB para estimar la captura  
de carbono en un ambiente semiárido

**por**

Jaquelin Guadalupe Ramos García

se distribuye bajo una [Licencia Creative Commons  
Atribución-NoComercial-SinDerivadas 4.0 Internacional](#).

**Clave RI:** CNMAC-246060



Universidad Autónoma de Querétaro  
Facultad de Ciencias Naturales  
Maestría en Ciencias Biológicas

Evaluación de la fluorescencia inducida por el sol y productos satelitales  
de PPB para estimar la captura de carbono en un ambiente semiárido

Tesis

Que como parte de los requisitos para obtener el Grado de  
Maestra en Ciencias Biológicas

Presenta

Jaquelin Guadalupe Ramos García

Dirigida por

Mónica Cervantes Jiménez

Dra. Mónica Cervantes Jiménez  
Presidente

\_\_\_\_\_  
Firma

Dra. Mónica Elisa Queijeiro Bolaños  
Secretaria

\_\_\_\_\_  
Firma

Dr. Samuel Villarreal Rodríguez  
Vocal

\_\_\_\_\_  
Firma

Dr. Aurelio Guevara Escobar  
Vocal

\_\_\_\_\_  
Firma

Dr. Enrique González Sosa  
Vocal

\_\_\_\_\_  
Firma

Centro Universitario, Querétaro, Qro.

octubre de 2023

## Dedicatoria

*A mi familia,  
con profundo amor y agradecimiento*

## **Agradecimientos**

Agradezco a la Universidad Autónoma de Querétaro y a la Facultad de Ciencias Naturales por el invaluable apoyo y preparación necesaria que me brindaron durante todo el proceso de completar el programa de la Maestría en Ciencias Biológicas.

Al Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCyT) por la beca de maestría con número 804930, que me permitió realizar esta tesis.

Quiero expresar mi más sincero agradecimiento a la Dra. Mónica Cervantes Jiménez por su dedicación en la dirección de esta tesis, por la oportunidad de trabajar juntas y por su disposición e interés para resolver mis inquietudes. Gracias por la motivación constante y por confiar en mis capacidades y habilidades durante todo mi proceso de formación.

Al Dr. Aurelio Guevara Escobar, gracias por brindarme acceso al equipo de cómputo especializado para llevar a cabo la metodología de este proyecto y por compartir su conocimiento conmigo.

A la Dra. Mónica Elisa Queijeiro Bolaños, gracias por la constante disposición para escucharme, por aclarar mis dudas y por la gestión de apoyos para los cursos.

Al Dr. Samuel Villarreal Rodríguez, gracias por estar al pendiente de mis avances, por compartir su vasto conocimiento y perspectiva que contribuyeron a la mejora de este trabajo.

Al Dr. Enrique González Sosa, le agradezco por compartir equipo y datos clave, así como por sus recomendaciones que enriquecieron enormemente este proyecto. Su colaboración ha sido fundamental para alcanzar este logro.

Agradezco a mis padres y a mis hermanos, quienes siempre han apoyado con gran amor y dedicación mis metas y ganas de superarme.

A David, por su cariño y compañía invaluable y por darme todo el apoyo para iniciar y finalizar este proyecto en mi vida.

Agradezco a todos los profesores y miembros del Laboratorio de Ecología Vegetal de la Facultad de Ciencias Naturales, por su apoyo y por los momentos gratos.

A mis compañeros de maestría, por alentarnos mutuamente en cada paso de este proceso.

Índice	
Agradecimientos .....	ii
Resumen .....	ix
Abstract .....	x
I. Introducción .....	1
II. Marco teórico y conceptual .....	3
2.1 Dinámica del carbono en ecosistemas terrestres .....	3
2.2 Torres de flujo de Eddy Covariance.....	5
2.3 Productos satelitales para estimar la PPB .....	7
2.3.2 Producto FluxSat V2.0 .....	8
2.3.3 Fluorescencia Inducida por el Sol .....	8
2.4 Aprendizaje automático para la estimación de PPB .....	12
III. Antecedentes.....	12
IV. Justificación.....	15
V. Hipótesis .....	15
VI. Objetivo general .....	16
6.1 Objetivos particulares.....	16
VII. Materiales y Métodos .....	16
7.1 Sitio de estudio.....	16
7.2 Obtención de datos en campo.....	17
7.3 Obtención de datos por percepción remota .....	18
7.3.1 Productividad Primaria Bruta MODIS .....	18
7.3.2 Productividad Primaria Bruta FluxSat.....	19
7.3.3 Fluorescencia Inducida por el Sol .....	19
7.3.4 Índices de vegetación y variables meteorológicas.....	19
7.4 Modelación.....	21
7.4.1 Regresión simple .....	21
7.4.2 Modelos de aprendizaje automático para predecir la PPB .....	21
VIII. Resultados .....	24
8.1 Modelos ML para predecir la PPB .....	27
8.2 Análisis estacional.....	31
8.2.1 Temporada de lluvias.....	31
8.2.2 Temporada de sequía.....	34
IX. Discusión.....	37

9.1 Regresión simple.....	37
9.2 Modelos de aprendizaje automático para predecir la PPB.....	39
9.2 Análisis estacional.....	42
X. Conclusiones.....	43
Referencias .....	46
Anexos .....	56
Anexo I. Características de las reflectancias e índices de MODIS.....	56
Anexo II. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB de 2017 a 2018: a) Modelo XG Boost basado en PPB_MODIS; b) Best of Family Stacked Ensemble basado en PPB_FluxSat; c) GBM basado en la SIF_MODIS.....	57
Anexo III. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB en temporada de lluvias: a) XG Boost basado en PPB_MODIS; b) XG Boost basado en PPB_FluxSat; c) XG Boost basado en la SIF_MODIS.....	58
Anexo IV. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB en temporada de sequía: a) GBM basado en PPB_MODIS; b) DRF basado en PPB_FluxSat; c) XGBoost basado en la SIF_MODIS.....	59



## Índice de Cuadros

Cuadro I. Índices de vegetación diarios a partir de las bandas de reflectancias de MODIS.....	20
Cuadro II. Detalle de las variables meteorológicas. ....	20
Cuadro III. Temporada de lluvias y sequía en la zona de estudio. Amarillo: sequía, verde: lluvias. NA indica la ausencia de datos de PPB de EC. Se incluye la precipitación (P) mensual promedio de acuerdo a la base de datos climatológica nacional (sistema CLICOM).....	24
Cuadro IV. Medidas de bondad de ajuste calculadas de los modelos ML basados en los productos satelitales y mejor modelo durante el proceso para estimar la PPB en la localidad de Bernal de 2017 a 2018. ....	30
Cuadro V. Comparación de las medidas de bondad de ajuste de los modelos ML basados en los productos satelitales y mejor modelo para estimar la PPB en la localidad de Bernal en la temporada de lluvias de 2017 a 2018.....	33
Cuadro VI. Comparación de las medidas de bondad de ajuste de los modelos ML basados en los productos satelitales y mejor modelo durante el proceso de entrenamiento para estimar la PPB en la localidad de Bernal en la temporada de secas de 2017 a 2018.....	36

## Índice de figuras

Figura 1. Representación esquemática de los principales componentes del intercambio neto de carbono del ecosistema (NEE). Existen entradas relativamente menores, como carbono orgánico disuelto por la precipitación (lixiviados) y emisiones por disturbios. Créditos: Elaboración propia. ....	4
Figura 2. Conceptualización de la emisión de fluorescencia. La luz absorbida puede ser utilizada para la fotoquímica, disiparse térmicamente o reemitirse como fluorescencia de la clorofila. En la figura se presenta una figura conceptual de la emisión de fluorescencia en las hojas, con picos máximos en las regiones espectrales rojas y rojas lejanas, y provenientes de los fotosistemas PSII y PSI. PSII contribuye tanto a las emisiones rojas (~685 nm) como a las rojas lejanas (~740nm), mientras que PSI contribuye principalmente a la región roja lejana. Modificado de: Mohammed <i>et al.</i> , (2019).....	9
Figura 3. Localización geográfica de la zona de estudio. ....	17
Figura 4. Modelos de regresión simple entre la PPB registrada por la Torre de EC en el sitio de estudio y los productos satelitales: a) PPB_MODIS diaria versus las mediciones diarias de PPB de EC; b) PPB_FluxSat diaria en comparación con las mediciones diarias de PPB de EC; c) SIF diaria del producto MODIS en relación con las mediciones diarias de PPB de EC. En todos los casos, se empleó el conjunto de datos completo para llevar a cabo las comparaciones. ....	26
Figura 5. Comparación y medidas de tendencia central de las series temporales entre los valores de PPB de la torre de EC y los datos satelitales: a) PPB_MODIS y mediciones de PPB de EC; b) PPB_FluxSat en comparación con las mediciones de PPB de EC; c) SIF del producto MODIS en relación con las mediciones de PPB de EC. En cada figura se incluyen medidas de tendencia central: promedio $\bar{X}$ , mediana Me, y moda Mo, con el color correspondiente de cada producto. ....	27
Figura 6. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en PPB_MODIS y las observaciones de PPB derivadas de EC en el sitio Bernal, validadas con el 15% del conjunto de datos. ....	28
Figura 7. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en PPB_Fluxsat y las observaciones de PPB derivadas de EC en el sitio Bernal, validadas con el 15% del conjunto de datos. ....	29
Figura 8. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en SIF de MODIS y las observaciones de PPB derivadas de EC en el sitio Bernal, validado con el 15% del conjunto de datos. ....	30
Figura 9. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB de MODIS frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.....	31
Figura 10. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB FluxSat frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.....	32

Figura 11. Concordancia entre las predicciones de PPB del modelo ML basado en la SIF de MODIS frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.....	33
Figura 12. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB MODIS y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía.....	34
Figura 13. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB FluxSat y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía.....	35
Figura 14. Concordancia entre las predicciones de PPB del modelo ML basado en la SIF de MODIS y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía.....	36

## Resumen

La productividad primaria bruta (PPB) representa es el carbono fijado por la vegetación a través de la fotosíntesis en los ecosistemas. En entornos áridos y semiáridos, donde el agua es limitada, es esencial estimar la PPB para comprender su funcionamiento ante disturbios y cambio climático. Desafortunadamente, estos ecosistemas están subrepresentados en los modelos de dinámica de carbono. El método de Eddy Covariance (EC), mide directamente el intercambio de gases entre biosfera y la atmósfera y permite calcular la PPB, proporcionando las estimaciones más fiables. Sin embargo, este método resulta costoso y su alcance es limitado. La percepción remota (PR) ofrece alternativas accesibles para estimar la PPB a diferentes escalas. Una técnica reciente involucra la medición de fluorescencia inducida por el sol (SIF) mediante sensores remotos. Dado que la SIF está directamente relacionada con la actividad fotosintética de la vegetación, puede utilizarse como indicador para estimar la PPB. Este estudio evaluó la capacidad de productos satelitales para estimar la PPB en un ambiente semiárido en Bernal, Querétaro, Mx. Se consideraron los productos SIF\_MODIS, PPB\_MODIS y PPB\_FluxSat. Cada producto se evaluó mediante regresión lineal simple, comparándolos con estimaciones de PPB de una torre de EC en el sitio de estudio, durante 2017 y 2018. Además, se generaron tres modelos de aprendizaje automático (ML) para estimar la PPB basados en cada producto satelital, índices y variables meteorológicas de PR. La variable de respuesta en cada ML fue la PPB de la torre de EC. Cada modelo ML se entrenó con el 85% de los datos y validado con el 15% restante. Los resultados de la regresión lineal mostraron un ajuste deficiente al comparar los productos con mediciones directas de EC ( $r^2=0.65$  SIF\_MODIS,  $r^2=0.59$  PPB\_MOD17 y  $r^2=0.54$  PPB\_FluxSat). Las estimaciones de PPB a través de ML, se compararon favorablemente con las mediciones de la torre de EC. Destacando el modelo basado en SIF\_MODIS con un rendimiento robusto ( $r^2=0.93$ ), a pesar de ser una técnica novedosa. En conclusión, estas técnicas de PR tienen el potencial de contribuir de manera práctica y rentable al análisis de la captura de carbono tanto en el sitio de estudio.

*Palabras clave: productividad primaria bruta (PPB), fluorescencia inducida por el sol (SIF), ecosistemas semiáridos, estimación de carbono, aprendizaje automático (ML)*

## Abstract

Gross primary production (GPP) represents the process by which vegetation captures and stores carbon through photosynthesis within ecosystems. In arid and semiarid regions, where water availability becomes limited, precise GPP estimations are essential to comprehend its response to climate change. Unfortunately, these ecosystems are inadequately represented within models focused on carbon dynamics. The Eddy Covariance (EC) method directly quantifies gas exchange between the biosphere and the atmosphere, allowing the calculation of GPP providing the most reliable estimates. However, this method is characterized by high costs and limited coverage and distribution. Remote sensing (RS) techniques offer accessible alternatives to assess GPP across different scales. A recent approach involves retrieving solar-induced fluorescence (SIF) using RS. Given the direct association between SIF and the photosynthetic activity of vegetation, it serves as a direct indicator for GPP estimations. This study evaluates the capacity of different satellite products to estimate GPP in a semi-arid setting Bernal, Querétaro, Mx. The evaluated products include SIF\_MODIS, PPB\_MODIS, and PPB\_FluxSat. For each product, a simple linear regression was employed for assessment, comparing with GPP measurements derived from an EC tower in the study site between 2017 and 2018. Furthermore, three machine learning (ML) models were generated to predict GPP, based on each satellite product alongside vegetation indices and meteorological variables acquired through RS. The response variable in each ML model was the GPP measured by the EC tower. Each ML model was trained with 85% of the dataset and validated with the remaining 15%. Results demonstrated deficient adjustment when comparing with direct EC measurements ( $r^2=0.65$  for SIF\_MODIS,  $r^2=0.59$  for PPB\_MODIS, and  $r^2=0.54$  for PPB\_FluxSat). Moreover, GPP ML predictions exhibited a favorable alignment with *in-situ* EC measurements in Bernal. Notably, the ML model based on SIF\_MODIS exhibited remarkable performance with high degree of robustness ( $r^2=0.93$ ), despite being a relatively novel technique. In conclusion, these RS techniques hold the potential to practically and cost-effectively contribute to carbon capture analysis in the study area.

*Key words: gross primary production (GPP), solar-induced fluorescence (SIF), semiarid ecosystems, carbon estimates, machine learning (ML)*

## I. Introducción

Las actividades antropogénicas contribuyen al cambio climático global lo cual genera nuevas condiciones climáticas para los ecosistemas, dichas condiciones se encuentran fuera del rango histórico de variabilidad climática (Seastedt *et al.*, 2008). Estos cambios generalmente tienen consecuencias negativas para la biodiversidad y los servicios ecosistémicos, fundamentales para el bienestar humano (Locatelli y Kanninen, 2010). Entre estos servicios ecosistémicos, los de regulación y soporte se destacan por su capacidad para controlar el clima, controlar la contaminación atmosférica, mantener la calidad del agua, regular enfermedades, así como la captura y almacenamiento de carbono y otros gases de efecto invernadero (Uribe-Botero, 2015).

La dinámica del ciclo del carbono desempeña un papel fundamental en los servicios ecosistémicos de regulación, ya que influye en el balance de carbono a nivel global (Reichstein *et al.*, 2013). Uno de los componentes clave de esta dinámica es la productividad primaria bruta (PPB), la cual representa el proceso mediante el cual la vegetación asimila el carbono a través de la fotosíntesis en los ecosistemas (Beer *et al.*, 2010). Por lo tanto, la PPB representa la capacidad de los ecosistemas para contrarrestar parcialmente las emisiones de CO<sub>2</sub> causadas por actividades humanas (Beer *et al.*, 2010).

La dinámica de la PPB en diversos ecosistemas está determinada por una interacción compleja de factores ambientales y biológicos (Baldochi, 2014). En bosques tropicales, la PPB tiende a ser alta debido a las condiciones climáticas favorables durante todo el año y a la diversidad biológica (Beer *et al.*, 2010; Ahlström *et al.*, 2015). En regiones templadas, la PPB tiende a ser más alta durante los meses cálidos de primavera y verano, mientras que en invierno disminuye debido a la falta de luz solar y las bajas temperaturas (Beer *et al.*, 2010; Ahlström *et al.*, 2015).

Los ecosistemas áridos y semiáridos, cubren hasta el 40% de la superficie terrestre, desempeñan un papel significativo en la regulación del carbono a nivel global, debido a su extensa presencia territorial y variabilidad hidrometeorológica (Rotenberg y Yakir, 2010; Ahlström *et al.*, 2015). A pesar de que la captura de carbono a nivel global está principalmente influenciada por áreas altamente productivas, como los bosques tropicales, se ha observado que las tendencias anuales y las variaciones en la captura global de carbono están especialmente influenciadas por los ecosistemas semiáridos (Ahlström *et al.*, 2015). Sin embargo, la comprensión de la dinámica del carbono en estos ambientes sigue siendo limitada, ya que la mayoría de los estudios y modelos actuales a escala global y regional se centran en ecosistemas más húmedos (Biederman *et al.*, 2017; Villarreal *et al.*, 2018).

Hasta la fecha no se ha desarrollado una técnica completamente precisa para medir de manera directa los flujos de carbono a diferentes escalas espaciales y temporales (Baldochi, 2014). Actualmente, las torres de flujo de covarianza de vórtices, también conocidas como Eddy Covariance (EC en inglés), se consideran el método más preciso para medir el intercambio de carbono a nivel de ecosistema (Baldochi, 2014). Sin embargo, estos sistemas son relativamente costosos de adquirir y presentan limitaciones en la cobertura espacial de observación (Wu *et al.*, 2018). Además, la disponibilidad de sitios de distribución de la FLUXNET 2015, que constituye una red global de sitios de observación de EC, es limitada a nivel global, regional y nacional. (Wu *et al.*, 2018; Villarreal *et al.*, 2018; Villareal y Vargas, 2021).

Como alternativa para complementar la información obtenida a partir de las torres de EC, se han desarrollado modelos de aprendizaje automático (ML) basados en datos de percepción remota (PR) (Yan *et al.*, 2019). Estos modelos han surgido como herramientas efectivas para estimar la PPB a diversas escalas con resultados consistentes (Yan *et al.*, 2019). Estas alternativas resultan ser valiosas, debido a que la mayoría de los datos

están disponibles de forma gratuita, lo que promueve el acceso y la investigación de manera más accesible, especialmente en áreas donde no se han instalado torres de EC para el registro de datos (Holloway y Mengersen, 2018; Smith *et al.*, 2018).

Por otro lado, la fluorescencia inducida por el sol (SIF) es una señal emitida por la clorofila de las plantas y refleja la actividad fotosintética de la vegetación (Köhler *et al.*, 2018). Esta señal, también detectada mediante PR y puede ser utilizado como indicador directo para estimar PPB (Mohammed *et al.*, 2019). A pesar de ser una técnica relativamente reciente, la SIF ha surgido como una herramienta precisa para evaluar procesos fotosintéticos en distintos ecosistemas y para obtener estimaciones de PPB (Porcar-Castell *et al.*, 2014; Mohammed *et al.*, 2019).

Por lo tanto, este estudio tiene como objetivo evaluar dos productos satelitales de PPB (PPB\_MODIS y PPB\_FluxSat) y un producto de fluorescencia inducida por el sol (SIF\_MODIS), utilizando modelos de aprendizaje automático (ML). Esta evaluación se llevará a cabo en comparación con mediciones obtenidas por una torre de EC en un ecosistema semiárido. Al abordar estas metodologías alternativas, se espera avanzar en la comprensión de la dinámica del carbono en estos ecosistemas y mejorar nuestra capacidad para monitorear los flujos de carbono.

## **II. Marco teórico y conceptual**

### **2.1 Dinámica del carbono en ecosistemas terrestres**

La dinámica del carbono en los ecosistemas es esencial para evaluar su desempeño y salud, ya que proporciona información valiosa sobre cómo la fisiología del ecosistema responde a diversas presiones ambientales (Baldochi, 2014). Para entender esto, es crucial analizar varios componentes. La productividad primaria bruta (PPB) representa la tasa de absorción o la asimilación de carbono a través del proceso de la fotosíntesis que realiza la vegetación en un tiempo dado dentro de un ecosistema



(Chapin, 2011). Por su parte, la respiración de un ecosistema (Reco) es la suma de la respiración de la vegetación (Rveg) y de los organismos heterótrofos (Rhet), es decir, microfauna, mesofauna y microorganismos (p.ej. hongos micorrízicos y bacterias fijadoras de nitrógeno (Figura 1) (Chapin, 2011).

El intercambio neto del ecosistema (NEE) es una medida del intercambio neto de carbono entre un ecosistema y la atmósfera por unidad de área (Chapin, 2011). Es decir, el flujo de PPB menos la Reco entre biosfera y atmósfera, y es un indicador principal de la capacidad de sumidero de carbono de un ecosistema (Figura 1) (Kramer *et al.*, 2002). El NEE se puede ver representado por la siguiente relación:

$$NEE = PPB - Reco \quad (1)$$

Donde,

$$Reco = Rveg + Rhet \quad (2)$$

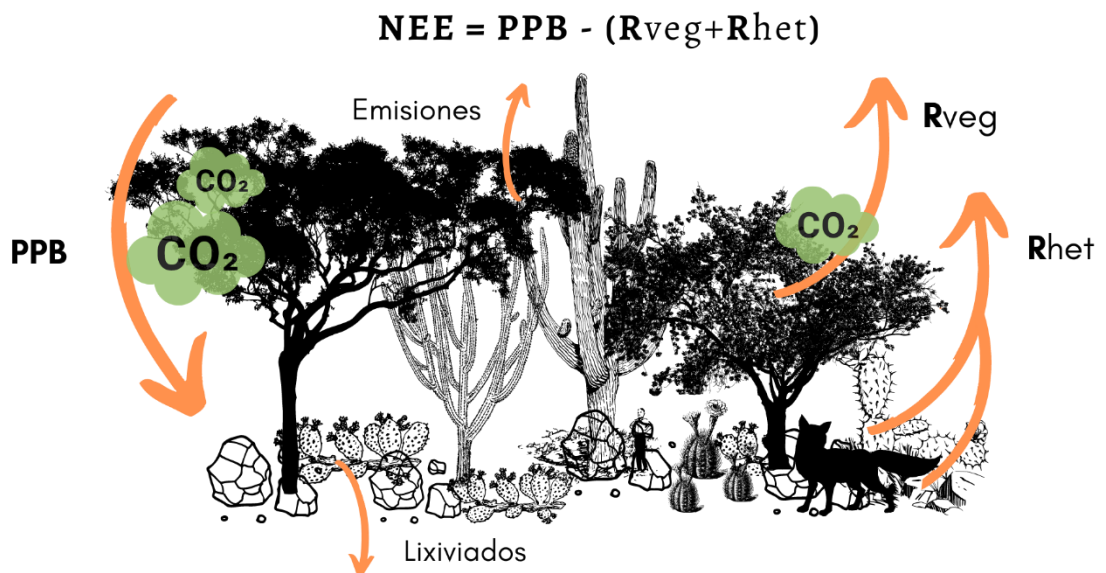


Figura 1. Representación esquemática de los principales componentes del intercambio neto de carbono del ecosistema (NEE). Existen entradas relativamente menores, como carbono orgánico disuelto por la precipitación (lixiviados) y emisiones por disturbios. Créditos: Elaboración propia.

Es importante mencionar que, en este intercambio de carbono, el ecosistema experimenta pérdidas de carbono a través de diferentes vías, además de la respiración (Chapin, 2011). Estas vías incluyen las emisiones de carbono resultantes de incendios forestales y otros disturbios (Piao *et al.*, 2019) (Figura 1). Otra fuente de pérdida es la muerte de plantas o partes de plantas; el consumo de plantas por herbívoros; la transferencia de carbono por microorganismos asociados a la vegetación; por la secreción de compuestos orgánicos volátiles y por la formación de carbonatos originados por carbono inorgánico, los cuales pueden ser lixiviados hacia las aguas subterráneas (Lal *et al.*, 2004; Chapin, 2011) (Figura 1).

La PPB está controlada por factores que regulan los patrones de la fotosíntesis a corto y largo plazo (Chapin, 2011). Los factores abióticos que intervienen son la radiación solar, las fluctuaciones de temperatura, la precipitación, el régimen de humedad en el suelo, el déficit de vapor, la disponibilidad de nutrientes y disturbios ambientales (Chapin, 2011).

Los factores bióticos que desempeñan un papel crucial en la dinámica del carbono en los ecosistemas, incluyen al tipo de vegetación, la estructura del dosel y los cambios en la fenología de la vegetación (Chapin, 2011). La fenología de la vegetación, en particular, determina el momento y la duración de un dosel fotosintéticamente activo e influye en la magnitud de los flujos de carbono y agua a lo largo de una temporada (Wolf *et al.*, 2016).

## 2.2 Torres de flujo de Eddy Covariance

Actualmente, el método de covarianza de vórtices o método de Eddy Covariance (EC) se destaca como una de las técnicas más precisas, directas y ampliamente utilizadas para medir los flujos de carbono entre biósfera y atmósfera a nivel de ecosistema (Burba, 2013; Baldocchi, 2014). Esta técnica implica el uso de torres equipadas con sensores micrometeorológicos que registran las fluctuaciones turbulentas verticales del flujo atmosférico, como la velocidad del viento y las concentraciones de gases (por ejemplo, CO<sub>2</sub>, H<sub>2</sub>O, CH<sub>4</sub> y N<sub>2</sub>O) (Burba, 2013; Baldocchi, 2014).

El método de EC emplea un anemómetro sónico para medir la velocidad del viento y un analizador de gases específico para la medir la concentración de los de interés (Burba, 2013). Estas torres de EC monitorean de manera continua estas fluctuaciones a nivel del dosel sin perturbar la vegetación, con una cobertura espacial de muestreo de cientos de metros cuadrados (Capioli *et al.*, 2016).

En la actualidad, existe una red global de monitoreo llamada FLUXNET, que engloba sistemas de EC distribuidos en América del Norte, Central y del Sur, Europa, Asia, África y Australia (Pastorello *et al.*, 2020). En México, la red de torres de EC conocida como MexFlux cuenta con varios sitios de medición en diferentes regiones, enfocados en el monitoreo del intercambio de CO<sub>2</sub> y energía entre la biósfera y la atmósfera (Villarreal *et al.*, 2018).

Los sensores en las torres de EC registran la velocidad vertical del viento y su contenido de CO<sub>2</sub> a alta frecuencia, por ejemplo, diez o veinte veces por segundo, en el aire ascendente y descendente (Burba, 2013). Estos datos se promedian en ventanas de 30 minutos. Al acumular estos flujos durante períodos como una hora o un día, se obtiene un registro del flujo neto de CO<sub>2</sub> entre la biosfera y la atmósfera (NEE) (Burba, 2013).

El registro del NEE se puede descomponer en sus componentes principales, la PPB y la Reco de varias maneras utilizando diferentes métodos basados en modelos conceptuales ecosistémicos (Reichstein *et al.*, 2005; Wulter *et al.*, 2018). El método de Reichstein *et al.*, (2005), es generalmente usado para esta partición, se basa en la obtención de Reco nocturno sensible a la temperatura, para extrapolar valores de Reco diurno. Una vez obtenido el Reco diurno, la PPB se calcula restando el Reco del NEE (Reichstein *et al.*, 2005) (Ver ecuación 1).

Los sistemas de EC ofrecen ventajas significativas, ya que proporcionan una comprensión detallada de los flujos de carbono, sin embargo, la adquisición de estos sistemas requiere de inversiones considerables (Burba, 2013). Adicionalmente, se presenta un desafío en cuanto a la

representación geográfica de los sitios de la FLUXNET 2015, ya que existe una escasez de sitios de observación en regiones áridas y semiáridas (Pastorello *et al.*, 2020; Villarreal *et al.*, 2018). Más aun, estas regiones carecen de registros de datos continuos a largo plazo, lo que dificulta obtener estimaciones confiables de los flujos de carbono en estos entornos (Pastorello *et al.*, 2020; Villarreal *et al.*, 2018). Por tal motivo, surge la necesidad de profundizar y desarrollar métodos precisos para la estimación de PPB.

## 2.3 Productos satelitales para estimar la PPB

### 2.3.1 Producto MOD17A2H

El Espectrorradiómetro de Imágenes de Resolución Moderada (MODIS por su sigla en inglés) es un instrumento a bordo de los satélites Terra y Aqua de la NASA. Es un sensor diseñado para capturar imágenes de alta resolución de la superficie y la atmósfera terrestre en una variedad de bandas espectrales, proporcionando datos y productos valiosos para una amplia gama de investigaciones y aplicaciones de ciencias de la Tierra (Running y Zhao, 2019).

El producto MOD17A2H v006 de MODIS (PPB\_MODIS), es uno de los productos satelitales de PPB más usados en estudios sobre la dinámica carbono a diferentes escalas (Running y Zhao, 2019; Yan *et al.*, 2019). Este producto es una composición acumulativa de valores de PPB globales, cuenta con datos de 2000 a 2022, a una resolución espacial de 500 m<sup>2</sup> y una resolución temporal de diez días, con unidades de kg C m<sup>-2</sup> (Running y Zhao, 2019).

El producto PPB\_MODIS deriva de un algoritmo que se basa en la lógica que sugiere que la productividad de cultivos en condiciones de riego y fertilización adecuadas está linealmente relacionada con la cantidad de radiación fotosintéticamente activa absorbida (APAR) (Monteith, 1972; Runnin y Zhao, 2019). La estimación de la productividad a partir de APAR se realiza a través de un parámetro de eficiencia de conversión ( $\epsilon$ ), que

varía según el tipo de vegetación y las condiciones climáticas (Running y Zhao, 2019). En consecuencia, para estimar la PPB, las principales entradas de datos al algoritmo incluyen: fracción de radiación fotosintéticamente activa (fPAR) e índice de área foliar (LAI) de MODIS; un conjunto de datos meteorológicos; una clasificación de cobertura global terrestre; y los valores de  $\epsilon$  máxima para los diferentes tipos de vegetación (Running y Zhao, 2019).

### *2.3.2 Producto FluxSat V2.0*

Recientemente, ha surgido un producto de PPB conocido como FluxSat V2.0 (PPB\_FLUXSAT), que se ha desarrollado mediante un algoritmo de redes neuronales basado en reflectancias del sensor MODIS, con la ventaja de estar calibrado con información de las torres EC de FLUXNET 2015 (Joiner *et al.*, 2018; Joiner y Yoshida, 2020). Para generar este producto, se utilizaron las reflectancias del producto MCD43C4v006 de MODIS, que proporciona datos diarios de reflectancias ajustadas a la distribución de la reflectancia bidireccional de nadir (BRDF Bidirectional Reflectance Distribution Function en inglés) a una resolución espacial de 0.05 grados. Además, FluxSat V2.0 se basa en los datos de torres de EC de FLUXNET 2015, que incluyen mediciones de flujos de CO<sub>2</sub>, vapor de agua y calor sensible de más de 200 torres distribuidas en todo el mundo (Joiner *et al.*, 2018; Joiner y Yoshida, 2020). FluxSatV2.0, ha sido utilizado con menos frecuencia en comparación con PPB\_MODIS, sin embargo, por su origen y validación es un producto prometedor, y han surgido algunos estudios que demuestran resultados positivos de este producto (Byrne *et al.*, 2020; Bennet *et al.*, 2021).

### *2.3.3 Fluorescencia Inducida por el Sol*

La Fluorescencia Inducida por el Sol (SIF), es la radiación electromagnética emitida directamente por el sistema fotosintético de la vegetación (Gu, 2019) (Figura 2). La emisión de esta señal se produce después de que la clorofila de las plantas absorbe radiación fotosintéticamente activa (PAR),

cierta cantidad de esta energía es utilizada en las reacciones fotoquímicas que convierten el CO<sub>2</sub> y agua en glúcidos y oxígeno, mientras que parte de la energía que no se utiliza de inmediato se reemite en forma de fluorescencia, a longitudes de onda entre los 650 y 800nm (Meroni *et al.*, 2009; Porcar-Castell *et al.*, 2014) (Figura 2).

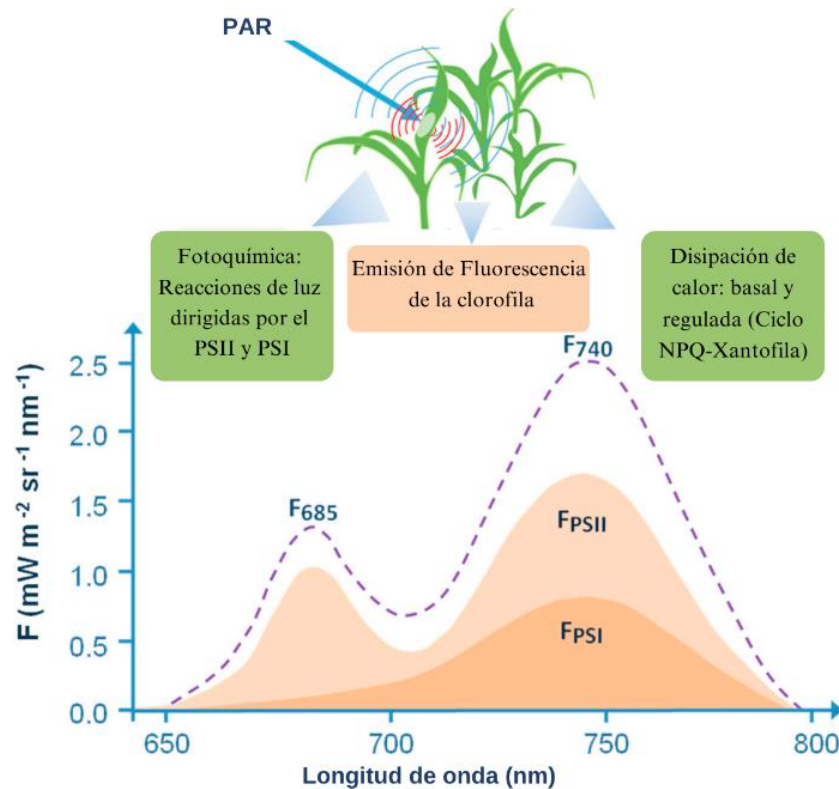


Figura 2. Conceptualización de la emisión de fluorescencia. La luz absorbida puede ser utilizada para la fotoquímica, disiparse térmicamente o reemitirse como fluorescencia de la clorofila. En la figura se presenta una figura conceptual de la emisión de fluorescencia en las hojas, con picos máximos en las regiones espectrales rojas y rojas lejanas, y provenientes de los fotosistemas PSII y PSI. PSII contribuye tanto a las emisiones rojas (~685 nm) como a las rojas lejanas (~740nm), mientras que PSI contribuye principalmente a la región roja lejana. Modificado de: Mohammed *et al.*, (2019).

Debido a que existe una estrecha relación entre la SIF y la actividad biológica de la vegetación, ésta es una herramienta emergente en el campo de la PR para estudiar el proceso de fotosíntesis y la PPB (Guanter *et al.*, 2012; Porcar-Castell *et al.*, 2014). Dicha relación se ha observado a lo largo de varias escalas, abarcando el nivel de la hoja y la planta (Damm *et al.*,

2015), hasta el dosel (Zarco-Tejada *et al.*, 2013) y el ecosistema en conjunto (Guanter *et al.*, 2012; Porcar-Castell *et al.*, 2014).

La SIF se encuentra en el espectro de longitudes de onda de 650 a 800 nm, con dos picos máximos de emisión característicos en las regiones del rojo y rojo lejano (también conocido como infrarrojo cercano o NIR) (Yang *et al.*, 2017; Joiner *et al.*, 2014) (Figura 2). Los avances en la PR han permitido que varios satélites espaciales sean empleados para la observación de la SIF (Parazoo *et al.*, 2019). A pesar de que estos satélites fueron inicialmente diseñados para el monitoreo de la química atmosférica, su amplia cobertura espectral los convierte en herramientas valiosas para la medición de la SIF (Parazoo *et al.*, 2019; Mohammed *et al.*, 2019).

Uno de los primeros satélites en contribuir a las mediciones de la SIF fue el satélite GOME-2 (Global Ozone Monitoring Experiment-2). Su principal objetivo es el monitoreo de la composición atmosférica, especialmente la concentración de ozono (Joiner *et al.*, 2013). Este satélite ofrece mediciones de SIF a una resolución espacial de 40 km x 40 km y temporal cada 29 días, desde 2007 a la fecha. GOME-2 ofrece una visión global de la SIF, pero poco detallada para estudios locales (Joiner *et al.*, 2013).

Por otro lado, GOSAT (Greenhouse Gases Observing Satellite), fue diseñado principalmente para monitorear gases de efecto invernadero, pero también contribuyen con mediciones de SIF (Frankberg *et al.*, 2011). Este satélite ofrece mediciones de SIF globales pero discontinuas espacialmente, a una resolución espacial de 10x10 km y resolución temporal cada tres días (Frankberg *et al.*, 2011). GOSAT ofrece mediciones desde 2009 hasta la fecha presente (Frankberg *et al.*, 2011).

El satélite OCO-2 (Orbiting Carbon Observatory-2) es una misión dedicada a monitorear las concentraciones de CO<sub>2</sub> en la atmósfera (Frankberg *et al.*, 2014). Destaca por su alta resolución espacial de aproximadamente 1.3 por 2.25 km, y resolución temporal de 16 días (Frankberg *et al.*, 2014). Sin embargo, a pesar de su ventaja en la resolución espacial, OCO-2 no proporciona una cobertura espacialmente continua de la superficie terrestre

debido a su franja de observación, que tiene un ancho de 10 kilómetros. Como resultado, OCO-2 necesita realizar múltiples visitas para recopilar suficientes datos de un sitio específico (Frankberg *et al.*, 2014). OCO-2 cuenta con mediciones de SIF desde 2014 a la fecha.

El instrumento TROPOMI (TROPOspheric Monitoring Instrument) a bordo del satélite Sentinel-5P tiene como objetivo la observación de la composición atmosférica (Guanter *et al.*, 2015). Cuenta con la capacidad de medir la SIF a alta resolución de 7 x 3.5 km, y a una resolución temporal de 16 días. Esta resolución espacial permite a TROPOMI proporcionar una visión detallada de los patrones de SIF en la Tierra (Guanter *et al.*, 2015). TROPOMI fue lanzado a orbita en octubre de 2017, y ofrece mediciones desde abril de 2018 a diciembre de 2021 (Guanter *et al.*, 2015).

La segunda versión del producto SIF<sub>OCO2-005</sub> de MODIS (MODIS SIF), proporciona estimaciones de SIF globales y espacialmente continuas a una resolución espacial de 0.05° (aproximadamente 5 km) y temporal de 16 días, desde septiembre de 2014 hasta julio de 2020, con unidades en  $\text{mW m}^{-2}\text{sr}^{-1} \text{nm}^{-1}$  (Li y Xiao, 2019; Yu *et al.*, 2021). Este producto es resultado de una red neuronal artificial (ANN) entrenada con datos de SIF de OCO-2 y reflectancias de superficie de siete bandas de MODIS (Li y Xiao, 2019; Yu *et al.*, 2021). Esta ANN permitió predecir la SIF en áreas donde no se dispone de observaciones de OCO-2 para brindar datos espacialmente continuos (Li y Xiao, 2019; Yu *et al.*, 2021).

Es importante destacar que, entre las opciones disponibles, MODIS SIF resultó ser la selección más apropiada. Esta decisión se tomó después de considerar principalmente a OCO-2 y TROPOMI. Sin embargo, OCO-2 se descartó debido a la falta de continuidad espacial en sus datos, lo que resultó en una insuficiencia de datos para este estudio. TROPOMI también se excluyó debido a la falta de concordancia entre las fechas de los datos disponibles y las requeridas para este estudio. Por otro lado, GOME-2 y GOSAT no fueron tomados en cuenta debido a que no cumplen con la resolución requerida para este trabajo.



## 2.4 Aprendizaje automático para la estimación de PPB

En las últimas décadas las técnicas de aprendizaje automático o machine learning (ML), han sido empleadas ampliamente en el campo de la ecología y ciencias ambientales (Olden *et al.*, 2008; Thessen, 2016). Estas técnicas son lo suficientemente flexibles para manejar la interacción de múltiples variables, haciéndolas ideales para modelar procesos en los sistemas ecológicos (Olden *et al.*, 2008; Thessen, 2016).

Paralelamente, la aplicación de técnicas de ML a la información obtenida por PR, ha permitido distintas perspectivas de estudio en la ecología (Lary *et al.*, 2016). Estos enfoques han generado una comprensión más profunda de los patrones y procesos ecológicos, como la distribución de especies, cambios de cobertura vegetal, y la evaluación de flujos de carbono en los ecosistemas (Thessen, 2016, Xiao *et al.*, 2019).

Las técnicas de ML han hecho posible la modelación global y regional de la PPB gracias a integración de variables meteorológicas, y de fisiología vegetal, obtenidas *in situ* y captadas por PR (Yuan *et al.*, 2007; Xiao *et al.*, 2019; Running y Zhao, 2019). Por ejemplo, la disponibilidad de conjuntos de datos de la FLUXNET y el ML han permitido la extrapolación de observaciones en los sitios de distribución para generar estimaciones de PPB y Reco a escala regional y global (Xiao *et al.*, 2019; 2018; Bai *et al.*, 2021).

### III. Antecedentes

La capacidad de la vegetación para asimilar carbono a través de la fotosíntesis, conocida como producción primaria bruta a nivel de ecosistema, representa el intercambio más significativo de carbono entre la biosfera y la atmósfera (Beer *et al.*, 2010). Para obtener mediciones y estimaciones precisas de la PPB, es fundamental utilizar diversas técnicas que nos permitan monitorear de cerca el ciclo del carbono y evaluar de manera cuantitativa los impactos del cambio climático y las actividades humanas que repercuten en el medio ambiente (Cheng *et al.*, 2014).

Entre estas técnicas, el método de Eddy Covariance proporciona una medición directa y confiable de la dinámica de carbono a nivel de ecosistema, permitiendo la generación de información de alta precisión, la cual ha sido estandarizada e integrada por medio de diversas redes nacionales, regionales y globales de monitoreo ambiental (p. ej., MexFlux; AmeriFlux; FLUXNET) (Baldocchi, 2003; Baldocchi, 2014; Tarin *et al.*, 2020). Las cuales, a través de distintos productos de valor agregado como la generación de bases de datos, modelos conceptuales, estudios de síntesis espaciales, entre otras, han contribuido de manera sustancial a la comprensión de los flujos de carbono en distintos ecosistemas y a distintas escalas de tiempo (Goulden *et al.*, 1996; Liu *et al.*, 2019; Jung *et al.*, 2020).

Dada la limitada cobertura espacial de la técnica de EC, surge la necesidad de desarrollar métodos alternativos para obtener una comprensión más completa de la dinámica de carbono (Cheng *et al.*, 2014; Running y Zhao, 2019). Una estrategia de obtener una mayor cobertura espacial es mediante el surgimiento de modelos de PR basados en la captura óptima de la PAR por la vegetación y el uso eficiente de la luz (LUE) de la vegetación (Cheng *et al.*, 2014; Running y Zhao, 2019). EL LUE se refiere a la cantidad de energía solar que la vegetación utiliza para producir biomasa (Cheng *et al.*, 2014). Estas metodologías han dado lugar a productos como MOD17 de MODIS para estimar la PPB proporcionando una cobertura espacial más amplia y contribuyendo significativamente a nuestra comprensión de los flujos de carbono en diversos ecosistemas (Cheng *et al.*, 2014; Running y Zhao, 2019).

Además, la integración de datos de EC y datos de PR con técnicas de aprendizaje automático (ML), surgieron trabajos sobre la estimación de PPB en distintos ecosistemas y a escalas regionales y globales, mejorado nuestra comprensión de los procesos que regulan la captura de carbono (Yuan *et al.*, 2007; Wang *et al.*, 2010; Jin *et al.*, 2013; Yan *et al.*, 2019; Zhang *et al.*, 2019). Estudios más recientes basados en técnicas de aprendizaje automático (ML) conjuntan variables meteorológicas,

mediciones de reflectancia superficial, e índices de vegetación por medio de PR, siendo una herramienta valiosa para estimar la PPB global, encontrando poca incertidumbre en comparación con torres de EC de la red FLUXNET (Tramontana *et al.*, 2016; Joiner *et al.*, 2018; Joiner y Yoshida, 2020).

Actualmente, una de las líneas más activas de investigación en la PPB se enfoca en el estudio de la SIF, la cual ha demostrado un importante estimador de la PPB por su estrecha relación con la actividad fotosintética (Porcar-Castell *et al.*, 2014; Mohammed *et al.*, 2019; Köhler *et al.*, 2020). Joiner *et al.*, (2014) probó la capacidad que tiene la SIF para observar el ciclo estacional de la PPB global en comparación con mediciones provenientes de las torres de EC de la FLUXNET, obteniendo resultados prometedores. Damm *et al.*, (2015), concluyeron que el uso de la SIF puede disminuir la ambigüedad en las estimaciones de PPB entre ecosistemas, debido a que la SIF es sensible a cambios de vegetación. Con esto se espera que los modelos de estimación de PPB basado en SIF mejoren a medida que se disponga de más datos de los satélites especializados en captación de la señal de la SIF como OCO-2 o TROPOMI (Li y Xiao, 2019; Mohammed *et al.*, 2019).

Algunos de los trabajos basados en SIF para la estimación de la PPB, se han enfocado en los ecosistemas áridos y semiáridos. El trabajo de Smith *et al.*, (2018), evaluó la SIF en comparación con otros índices de vegetación para estimar la PPB, y encontraron que la SIF estima con mayor precisión la PPB estacional en ambientes semiáridos de Norte América. Yan *et al.*, (2019), evaluaron la relación entre IVs de MODIS y la PPB a distintas escalas espaciales y de resolución, donde destacan que la heterogeneidad de la vegetación en las zonas áridas puede complicar la modelación PPB utilizando IVs por PR en dicho tipo de ecosistemas. Sin embargo, aún son limitados los estudios de la dinámica de carbono en estos ambientes debido a su heterogeneidad estructural y funcional, así como a la falta de mediciones continuas (Biederman *et al.*, 2017; Smith *et al.*, 2019).

#### **IV. Justificación**

Es fundamental comprender y monitorear la dinámica del carbono para hacer frente a los desafíos del cambio climático y garantizar la provisión de servicios ecosistémicos que fomenten el bienestar socioambiental. Para lograrlo, es necesario llevar a cabo estudios que permitan el desarrollo de métodos de PR más precisos, y accesibles de medición de los flujos de carbono en los ecosistemas.

Además, se ha observado una falta de estudios con respecto a la dinámica del carbono que se centran en los ecosistemas áridos y semiáridos. Lo anterior limita nuestra capacidad para comprender plenamente como estos ecosistemas contribuyen a la regulación del carbono global. Por lo tanto, llevar a cabo estudios, como el que se describe en este documento, es crucial para aplicar métodos convenientes que nos permitirán obtener una visión más amplia de la dinámica del carbono en estos entornos.

Al contar con métodos confiables (p.ej. Eddy Covariance y modelos basados en datos de percepción remota) para estimar la captura de carbono en los ecosistemas áridos y semiáridos, se podrá brindar a los tomadores de decisiones una herramienta para promover prácticas de conservación, mitigación y uso sostenible del suelo, con el objetivo de reducir emisiones por pérdida de los ecosistemas. Asimismo, los productos pueden servir para su posible incorporación a los mecanismos de mercados de carbono, lo que podría impulsar la protección y el incremento los niveles actuales de captura carbono.

#### **V. Hipótesis**

Los modelos de aprendizaje automático basados en la señal de la fluorescencia inducida por el sol (SIF) son una mejor alternativa para predecir la productividad primaria bruta (PPB) en un ecosistema semiárido, en comparación con los modelos basados en productos satelitales de PPB originados por índices de vegetación y reflectancias, debido a que la SIF es

una medida directa de la actividad fotosintética, lo que la hace una mejor opción para predecir la PPB.

## **VI. Objetivo general**

Evaluar distintos modelos de ML para la estimación de PPB a partir de productos satelitales y a partir de un producto de SIF en un ecosistema semiárido.

### **6.1 Objetivos particulares**

- Validar dos productos satelitales de PPB (MOD17 Y FluxSat) y uno de SIF contra las mediciones de EC para determinar su concordancia como punto de referencia.
- Montar y entrenar algoritmos de aprendizaje automático para modelar la PPB en el sitio a partir de dos productos satelitales de PPB (MOD17 Y FluxSat) y diferentes variables ambientales.
- Montar y entrenar algoritmos de aprendizaje automático para modelar la PPB en la zona de estudio a partir de datos SIF y diferentes variables ambientales.
- Calcular parámetros para evaluar la bondad de ajuste de los modelos generados considerando observaciones de una torre de flujos de Eddy Covariance.

## **VII. Materiales y Métodos**

### **7.1 Sitio de estudio**

El sitio de estudio corresponde a la localidad de Bernal, en el municipio de Ezequiel Montes, el cual se encuentra en la franja semiárida del Estado de Querétaro, México (Figura 3). La cual a su vez forma parte de la eco región elevaciones mediterráneas meridionales (CONABIO, 2018). El observatorio se ubica a  $20^{\circ}7'17''$  N,  $99^{\circ}9'41''$  O, y a una altitud de 2050 m.s.n.m. Se caracteriza por un clima árido con lluvias en verano (Bsk) con temperatura promedio anual de  $17.1^{\circ}\text{C}$ , y precipitaciones de 476mm anuales (CICESE,

2015). La vegetación presente corresponde a un matorral secundario dominado por géneros como *Acacia*, *Prosopis*, y *Opuntia*, entre otros.

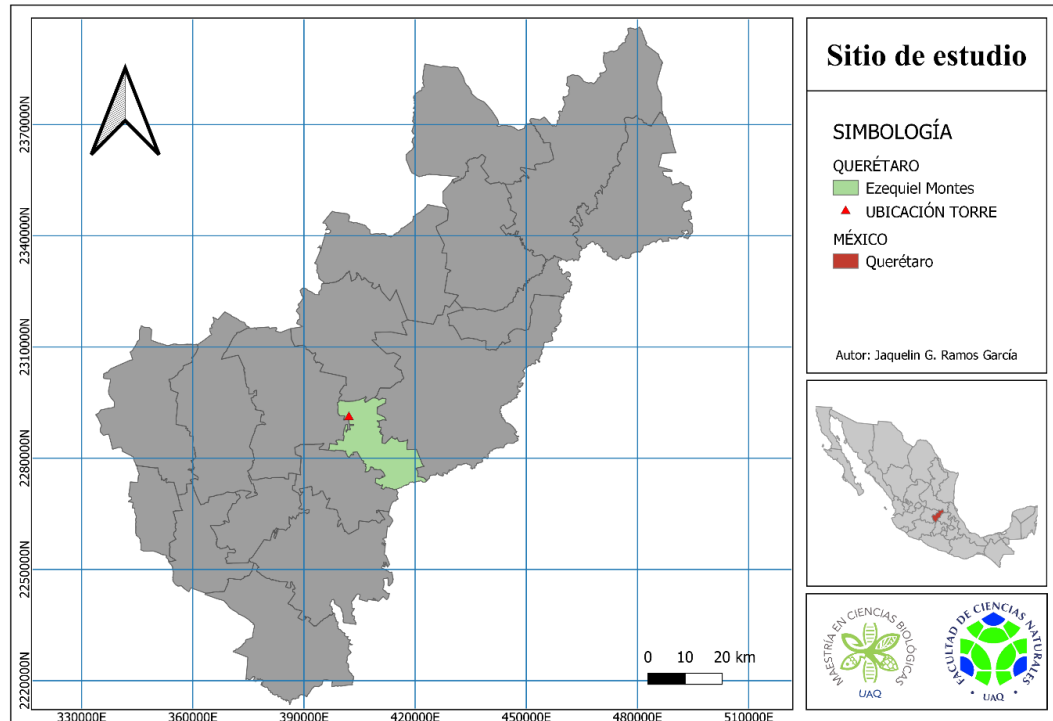


Figura 3. Localización geográfica de la zona de estudio.

## 7.2 Obtención de datos en campo

Durante el periodo de enero de 2017 a diciembre de 2018, una torre flujo de EC estuvo instalada en el sitio de estudio. El sistema de EC, con una cobertura espacial de 600 m<sup>2</sup>, midió los flujos de NEE a una altura de 6m con los siguientes instrumentos: un sistema Biomet (LI-COR Biosciences, USA) para medir flujos de H<sub>2</sub>O y CO<sub>2</sub> mediante un analizador de circuito abierto IRGASON-EC-150, un anemómetro sónico CSAT3 y un higrómetro de criptón KH20; estos se conectaron a un registrador de datos CR3000 (Campbell Scientific Inc., Logan, UT, EE. UU.) (Guevara *et al.*, 2021).

Variables como, humedad relativa y la temperatura del aire se midieron con una sonda HMP155A (Vaisala Corporation, Helsinki, Finlandia), la radiación neta se midió con un radiómetro NR-Lite2 (Kipp and Zonen BV Delft, Países Bajos) y la radiación fotosintéticamente activa (PAR) se midió con un

sensor cuántico (SKP215; Skye Instruments, Llandrindod Wells, Reino Unido). Las mediciones del flujo de calor del suelo se implementaron con cuatro placas HFP01SC auto calibradas a 80 mm de profundidad y en cuatro posiciones representativas del paisaje (Hukseflux Thermal Sensors BV, Delft, Países Bajos). El contenido volumétrico de agua en el suelo se midió con tres sondas de reflectometría en el dominio del tiempo (TDR) (CS616), y dos conjuntos de sondas TCAV (termopar de suelo promedio) midieron la temperatura a 60 y 40 mm de profundidad y por encima de las placas HFP01SC (Campbell Scientific Inc. Logan, UT, EE.UU.). La precipitación se midió con el pluviómetro TE525 (Texas Electronics, Dallas, TX, EE. UU.). Todas estas variables se midieron cada a 10MHz (10 mediciones por segundo) cada 5s y se procesaron con el paquete REddyProc en Rstudio, (R Development Core Team, 2009) para registrar valores promedio en intervalos de 30 minutos; la precipitación se acumuló durante el mismo intervalo de tiempo (Guevara *et al.*, 2021).

El NEE ( $\text{mmol m}^{-2} \text{ s}^{-1}$ ) se post procesó utilizando el paquete REddyProc de R) para dividir el flujo NEE en sus componentes: PPB y Reco. La Reco se calculó en base a la temperatura del suelo y del aire siguiendo el procedimiento de Reichstein *et al.*, (2005) (Guevara *et al.*, 2021). A partir del Reco se calcularon los valores de PPB, y se convirtieron a  $\text{g C m}^{-2}$  al día utilizando la relación molar del carbono (C). Únicamente se reportaron las mediciones continuas para el período de abril de 2017 a agosto de 2018, las cuales se tomaron en cuenta para el presente estudio.

### 7.3 Obtención de datos por percepción remota

#### 7.3.1 Productividad Primaria Bruta MODIS

Se descargó el producto MODIS de PPB MOD17A2H v006 (PPB\_MODIS) para el sitio de estudio y para el periodo de 2017 al 2018 a partir del Land Processes Distributed Active Archive Center (LPDAAC) de la National Aeronautics and Space Administration (NASA; <https://lpdaac.usgs.gov/>). Los productos obtenidos presentaron una resolución espacial de  $500 \text{ m}^2$ , y una resolución temporal de 8 días. Los valores del producto PPB\_MODIS

están expresados en  $\text{kg C m}^{-2}$ , para obtener valores de PPB en  $\text{g C m}^{-2}$ , se multiplicaron los valores del producto MOD17 por el factor de escalamiento de 0.0001, que se encuentra en la guía de usuarios (Running y Zhao, 2019).

### *7.3.2 Productividad Primaria Bruta FluxSat*

Se obtuvo el producto satelital de PPB FluxSat V2.0 (PPB\_FluxSat), para el sitio de estudio y para el periodo de 2017 al 2018 a partir de la plataforma LPDAAC de la NASA. Este producto ofrece valores diarios de PPB en archivos mensuales, cuenta con una resolución de  $0.05^\circ$  (aproximadamente 5 km) y unidades de  $\text{g C m}^{-2}$ , y con disponibilidad de datos desde 2014 a 2020. Se descargaron un total de 14 archivos mensuales del producto PPB FluxSat, y se generó un código para extraer los valores diarios de PPB en el sitio de estudio, manteniendo las unidades de  $\text{g C m}^{-2} \text{ d}^{-1}$ .

### *7.3.3 Fluorescencia Inducida por el Sol*

Se obtuvo el producto  $\text{SIF}_{\text{OCO2-005}}$  de MODIS (SIF\_MODIS), para el sitio de estudio y para el periodo de 2017 al 2018. Este producto cuenta con una resolución espacial de  $0.05^\circ$  (aproximadamente 5 km) y temporal de 16 días, y con disponibilidad de datos desde 2014 a 2020. La descarga del producto MODIS SIF se realizó a través de la plataforma LPDAAC de la NASA.

Para estandarizar las bases de datos a valores diarios, a los productos satelitales obtenidos cada 8 o 16 días (PPB\_MODIS y SIF\_MODIS), se les aplicó una interpolación a los datos basada en la técnica de Thin Plate Spline usando la función spline en Rstudio. Esta técnica se basa en la interpolación y suavizado de datos en un espacio bidimensional (en este caso la PBB o SIF diaria a partir los datos obtenidos cada 8 o 16 días, en la ubicación geográfica de la zona de estudio), que permite estimar datos continuos a partir de datos conocidos en un espacio (Green y Silverman, 1994).

### *7.3.4 Índices de vegetación y variables meteorológicas*



Con la finalidad de mejorar la representación de la dinámica del ecosistema, se descargaron diferentes subproductos de PR de la NASA. Se obtuvieron las bandas de reflectancias diarias de la B1 a B14 que cubren un rango espectral desde 405 hasta 2155 nm (MOD09GA y MOD09GA de MODIS), así como índices de vegetación de MODIS (se describe de manera detalla la resolución espectral y temporal de cada uno de estos en el Anexo I). A partir de las bandas de reflectancias se generaron índices de vegetación de acuerdo a Hill *et al.*, (2006), así como los índices de vegetación de acuerdo a la metodología de Henrich *et al.*, (2012), descritos en el Cuadro I.

**Cuadro I. Índices de vegetación diarios a partir de las bandas de reflectancias de MODIS.**

Índice	Resolución espacial	Referencia
Simple ratio (SimpleR)	500 m	(Hill et al., 2006)
Moisture stress (MoistS)	500 m	(Hill et al., 2006)
Disease stress index (DSI)	500 m	(Hill et al., 2006)
Red-green ratio index (RGR)	500 m	(Hill et al., 2006)
Normalized difference vegetation index (NDVI)	500 m	(Hill et al., 2006)
Normalized difference water index (NDWI <sub>w</sub> )	500 m	(Hill et al., 2006)
Green leaf index (GLI)	500 m	(Hill et al., 2006)
Green atmospherically resistant vegetation (GAR)	1 km	(Henrich et al., 2012)
Enhanced vegetation index (EVI)	1 km	(Henrich et al., 2012)

Finalmente se obtuvieron parámetros meteorológicos como, datos de precipitación de la misión Global Precipitation Measurement (GPM) obtenidos a través del portal Giovanni (<https://giovanni.gsfc.nasa.gov/giovanni>). Los parámetros climatológicos se obtuvieron a partir de la plataforma Daymet de la NASA (Cuadro II).

**Cuadro II. Detalle de las variables meteorológicas.**

Variable	Resolución espacial	Resolución temporal	Plataforma de origen
Precipitación (Dayprc)	1 km	Diario	Daymet
Radiación de onda corta (Daysrad)	1 km	Diario	Daymet
Temperatura máxima del aire (DayTmax)	1 km	Diario	Daymet
Temperatura mínima del aire (DayTmin)	1 km	Diario	Daymet
Presión de vapor de agua (Dayvp)	1 km	Diario	Daymet
Precipitación (3IMERGDF v006)	11 km	Diario	GPM

Una vez obtenidos los datos, tanto *in situ* obtenidos de la torre de EC, como la información obtenida por PR la cual incluye, PPB\_MODIS y PPB\_

FluxSat, SIF\_MODIS, reflectancias, índices de vegetación y variables meteorológicas; se procedió a integrarlas en una base de datos con formato CSV.

## 7. 4 Modelación

### *7.4.1 Regresión simple*

Es posible validar la precisión de estimaciones de productos satelitales en comparación con mediciones registradas por torres de EC mediante métodos de regresión. El propósito principal de un análisis de regresión es interpretar y predecir la relación entre una variable de interés, y una o más variables explicativas (Green y Silverman, 1994). El objetivo del modelo de regresión simple, es encontrar el ajuste a la tendencia de los datos, de modo que pueda utilizarse para estimar valores de la variable de respuesta en función de los valores de la variable explicativa (Gotelli y Ellison, 2004). Estos modelos evalúan con el cálculo de un coeficiente de determinación ( $r^2$ ) que permite analizar la bondad de ajuste del modelo y proporciona la información para realizar una comparación entre las variables (Gotelli y Ellison, 2004; Crawley, 2012).

En este estudio, se construyeron tres modelos de regresión simple en Rstudio con la función `lm()`. El propósito de estos modelos es el establecer una relación entre los datos proporcionados por los productos satelitales (PPB\_MODIS, PPB\_FluxSat y SIF\_MODIS) y los datos obtenidos en el campo a partir de la torre de EC en la localidad de Bernal, durante el intervalo de tiempo intermitente que abarca desde abril a junio y octubre a diciembre de 2017 y de enero a agosto de 2018. Es importante destacar que se utilizó el 100% del conjunto de datos en todos los casos. Finalmente, se calculó el coeficiente de determinación ( $r^2$ ) como medida de bondad de ajuste para evaluar cada uno de los modelos.

### *7.4.2 Modelos de aprendizaje automático para predecir la PPB*

En el presente estudio se exploraron modelos de aprendizaje automático (ML) debido a su capacidad para ofrecer estimaciones más robustas, al

identificar patrones cuando se utilizan grandes conjuntos de datos (Holloway y Mengersen, 2018). En particular, el trabajo se centró en modelos ML para predictores de PPB, basados en los productos satelitales y las variables captadas por PR que conforman el conjunto de datos.

Para desarrollar los modelos ML se empleó con la función “AutoML” del paquete H2O en Rstudio (Hall *et al.*, 2019). Esta metodología se dividió en dos etapas clave: entrenamiento y evaluación de los modelos, donde “AutoML” automatiza el proceso completo (LeDell y Poirie, 2020). La versión actual de AutoML entrena y valida los siguientes algoritmos: tres modelos predefinidos XGBoost GBM (Gradient Boosting Machine), Modelos Lineales Generalizados (GLM), un Distributed Random Forest (DRF) predeterminado, una Red Neuronal Profunda predeterminada, un bosque extremadamente aleatorio (XRT), una cuadrícula aleatoria de GBM XGBoost, una cuadrícula aleatoria de GBM de H2O y una cuadrícula aleatoria de redes neuronales profundas (LeDell y Poirie, 2020).

Durante la fase de entrenamiento, la función “AutoML” entrena y ajusta la amplia gama de algoritmos de aprendizaje automático permitiendo así una búsqueda exhaustiva de soluciones que se ajuste mejor a los datos proporcionados (LeDell y Poirie, 2020). En esta fase del proceso, AutoML mantiene una tabla de líderes que registra los modelos y sus métricas de rendimiento correspondiente (LeDell y Poirie, 2020).

Al concluir el proceso de entrenamiento, se utilizó la función “`automl_leaderboard`” para acceder a una lista de los modelos generados junto con sus métricas de evaluación correspondientes. Esta función seleccionó automáticamente el modelo que demostró el mejor rendimiento en la métrica especificada, que en este caso se consideró el valor más bajo de devianza.

Adicionalmente, para comprender las variables más influyentes en las predicciones, se empleó la función “`h2o.varimp`” del paquete. Esta función proporciona información sobre las variables más relevantes para cada

modelo, que permite realizar una evaluación más profunda de su impacto en las predicciones. (Hall *et al.*, 2019).

Se desarrollaron tres ML individuales: uno basado en PPB\_MODIS, otro en PPB\_FluxSat y el tercero en SIF\_MODIS. Para cada uno de estos modelos, se utilizaron valores de reflectancia provenientes de los productos MODIS MOD09GA, que abarcan las bandas B1 a B7, y MODOCGA, que incluye las bandas B8 a B14, ambos procesados a nivel L2 (Anexo I). Además de estas reflectancias, se incorporaron diversos índices de vegetación como, Fracción de radiación fotosintéticamente activa (fPAR), Índice de área foliar (LAI) Temperatura de la superficie diurna y nocturna (LST day y LST night) Índice de vegetación mejorado (EVI), y Fotosíntesis neta (PsnNet), por mencionar algunos (Anexo I y Cuadro I). Además, se incluyeron parámetros meteorológicos obtenidos a través de PR, como Precipitación, Radiación de Onda Corta (Daysrad), Temperatura Máxima y Mínima del Aire (DayTmax y DayTmin), y Presión de Vapor de Agua (Dayvp) (Cuadro II). En cada uno de los modelos, la variable de respuesta fue la PPB registrada por la torre de EC en la localidad de Bernal.

Los modelos se entrenaron con el 85% del conjunto de datos y el 15% restante se utilizó para su evaluación. Es importante destacar que los modelos anteriormente descritos se generaron para el periodo de tiempo continuo de 2017 a 2018.

Adicionalmente, la serie de tiempo se segmentó con el propósito de llevar a cabo el mismo análisis tanto para la temporada de lluvias como para la de sequía. Dado que los datos de PPB de la torre de EC para el período de 2017 a 2018 presentaban discontinuidades, se llevó a cabo la división por temporadas como se presenta en el Cuadro III.

**Cuadro III. Temporada de lluvias y sequía en la zona de estudio. Amarillo: sequía, verde: lluvias. NA indica la ausencia de datos de PPB de EC. Se incluye la precipitación (P) mensual promedio de acuerdo a la base de datos climatológica nacional (sistema CLICOM)**

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
<b>2017</b>	NA	NA	NA				NA	NA	NA			
<b>P (mm)</b>	0	0.19	1.01	0.1	0.5	4.05	3.83	4.97	3.29	0.36	0.08	0
<b>2018</b>									NA	NA	NA	NA
<b>P (mm)</b>	60.1	78.8	37.2	75.4	63.8	96.5	66.2	93.8	60.4	31.6	1.01	0

Finalmente, para evaluar cada uno de los modelos, se calcularon medidas de bondad de ajuste entre valores observados y predichos. Para ello, se calculó el coeficiente de determinación ( $r^2$ ), la raíz cuadrada del cuadrado medio del error (RMSE), devianza y sesgo (Bland y Altman, 2010). Se emplearon las funciones “automl\_leaderboard” y “h2o.varimp”, para conocer el mejor modelo y las variables de mayor importancia en cada caso.

### **VIII. Resultados**

Los modelos de regresión simple permitieron obtener una primera aproximación sobre la precisión de los productos satelitales contemplados en este trabajo, en comparación con las mediciones registradas por la torre de EC en el sitio de estudio. Se observó un coeficiente de determinación ( $r^2$ ) de 0.59 ( $p < 2.2 \times 10^{-16}$ ) para los datos generales entre las estimaciones de PPB\_MODIS y las mediciones de PPB de EC (Figura 4a). Para el caso de las estimaciones de PPB\_FluxSat y PPB de EC se obtuvo un coeficiente de determinación de  $r^2=0.54$  ( $p < 2.2 \times 10^{-16}$ ) (Figura 4b). Por otra parte, para el conjunto de datos generales de la señal de la SIF del producto SIF\_MODIS y la PPB registrada por la misma torre de EC, se obtuvo un coeficiente de  $r^2=0.65$  ( $p < 2.2 \times 10^{-16}$ ) (Figura 4c). La significancia de la prueba en todos los casos indica que las relaciones observadas son significativas e improbable que sean resultado al azar.

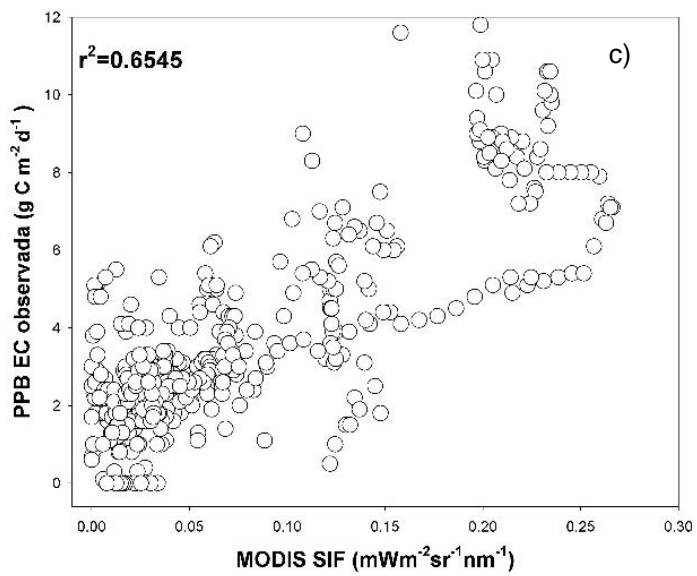
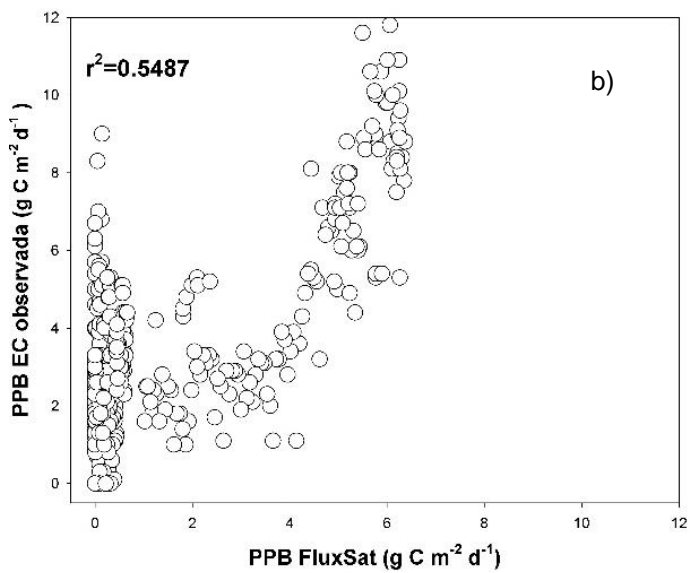
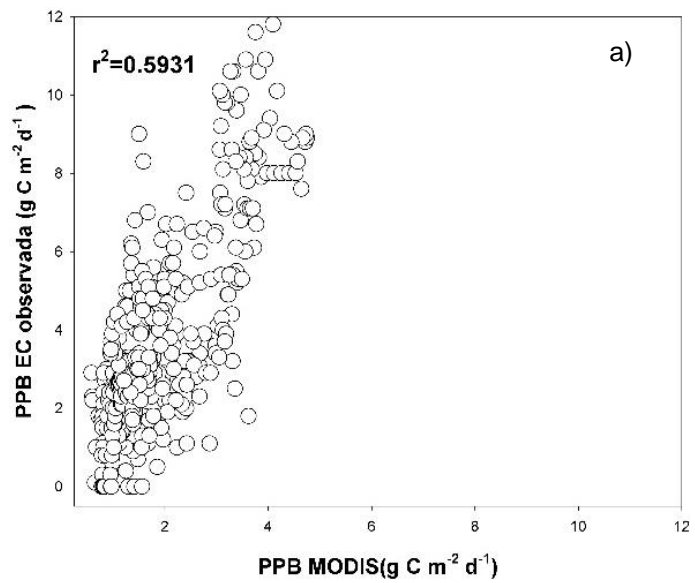
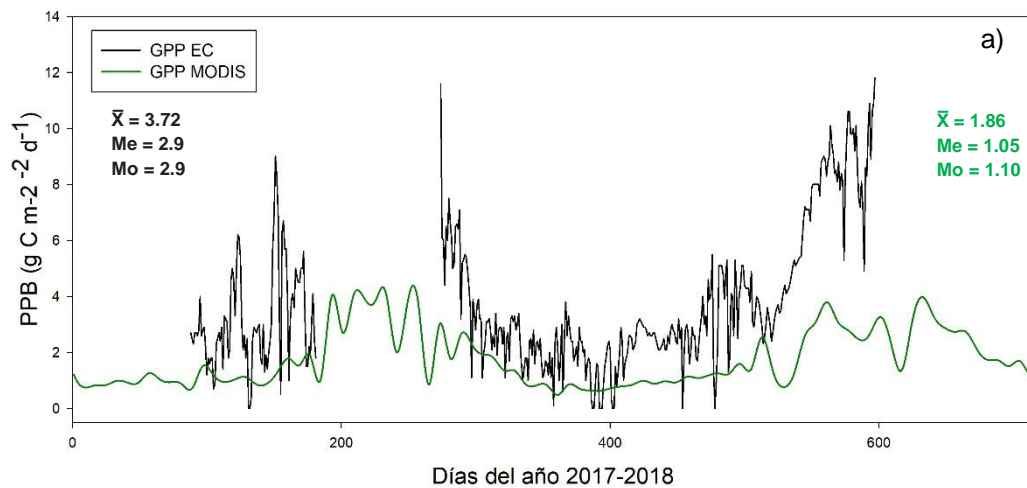


Figura 4. Modelos de regresión simple entre la PPB registrada por la Torre de EC en el sitio de estudio y los productos satelitales: a) PPB\_MODIS diaria versus las mediciones diarias de PPB de EC; b) PPB\_FluxSat diaria en comparación con las mediciones diarias de PPB de EC; c) SIF diaria del producto MODIS en relación con las mediciones diarias de PPB de EC. En todos los casos, se empleó el conjunto de datos completo para llevar a cabo las comparaciones.

Los valores obtenidos de  $r^2$  son relativamente bajos en todos los casos, señalando una notable variabilidad entre las mediciones de la PPB de la torre de EC que no puede ser explicada por los productos satelitales analizados en esta sección del estudio. Resulta interesante observar que, a pesar de este bajo ajuste, se identifican patrones similares en las series temporales de los datos de PPB de EC y los datos satelitales, especialmente para valores altos del producto SIF\_MODIS (Figura 5a, b y c). Estos resultados, aunque muestran una tendencia, resaltan la necesidad de una investigación más profunda y la consideración de otros factores y métodos para mejorar la precisión de las estimaciones de PPB utilizando datos de los productos satelitales considerados.



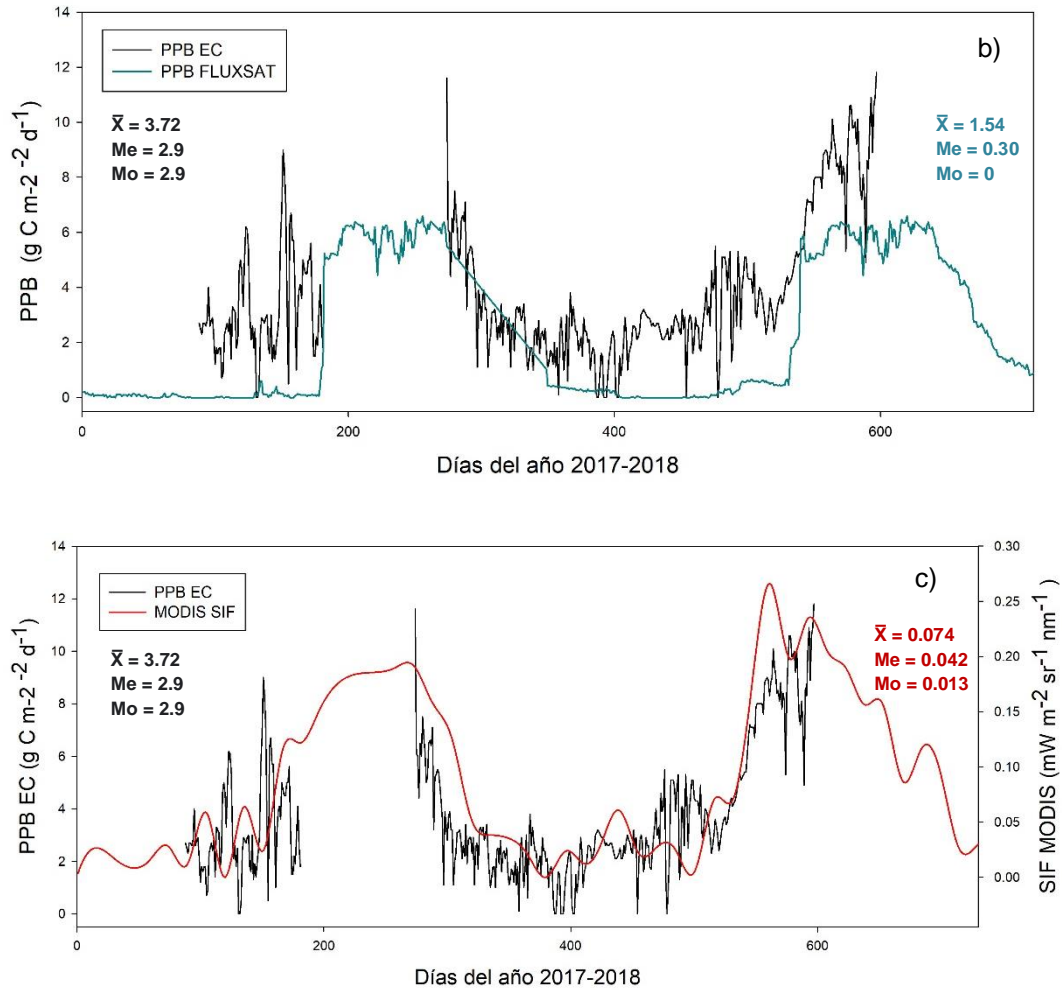


Figura 5. Comparación y medidas de tendencia central de las series temporales entre los valores de PPB de la torre de EC y los datos satelitales: a) PPB\_MODIS y mediciones de PPB de EC; b) PPB\_FluxSat en comparación con las mediciones de PPB de EC; c) SIF del producto MODIS en relación con las mediciones de PPB de EC. En cada figura se incluyen medidas de tendencia central: promedio  $\bar{X}$ , mediana Me, y moda Mo, con el color correspondiente de cada producto.

### 8.1 Modelos ML para predecir la PPB

De acuerdo con el modelo ML para predecir la PPB basado en las estimaciones de PPB\_MODIS y las variables derivadas de PR, se obtuvo una sólida relación estadística con un  $r^2=0.87$  y un RMSE de  $0.85 \text{ g C m}^{-2} \text{ día}^{-1}$  (Figura 6). La función AutoML generó un total de 509 modelos, y el mejor modelo fue seleccionado basándose en el valor mínimo de devianza. Este modelo pertenece a la familia "XGBoost" y presentó una devianza de 0.72. Las variables más relevantes que influyeron en las estimaciones de PPB, ordenadas de mayor a menor importancia, fueron: el Índice de Área



Foliar (LAI) del producto MCD15A2H de MODIS (LAI MCD) (54.9%), la Banda 14 (R\_14) del producto MODOCGA que cubre una resolución espectral de 673-683 nm (7.2%), y PPB\_MODIS (6.2%); mientras que el resto de las variables tuvieron una importancia inferior al 5% (Anexo II).

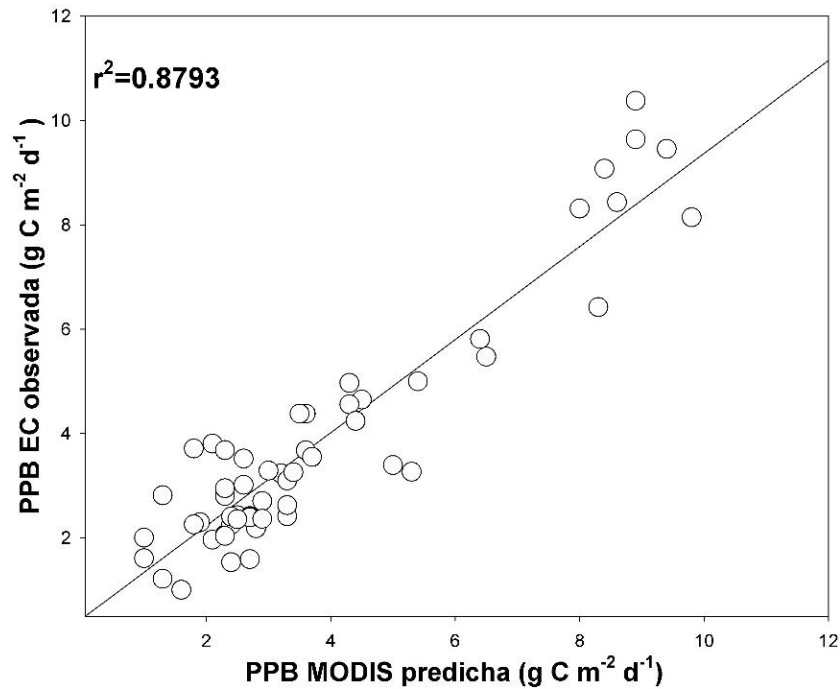


Figura 6. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en PPB\_MODIS y las observaciones de PPB derivadas de EC en el sitio Bernal, validadas con el 15% del conjunto de datos.

En el caso del ML utilizado para predecir la PPB basado en el producto FluxSat, se obtuvo un coeficiente de determinación ( $r^2$ ) de 0.88 y un RMSE = 0.86 g C m<sup>-2</sup> día<sup>-1</sup> (Figura 7). Durante el proceso de entrenamiento, se generaron un total de 643 modelos, y se seleccionó un ensamble basándose en el valor mínimo de devianza (0.73), el cual corresponde a un modelo de la familia "XGBoost". Las variables más influyentes para las predicciones de PPB en este modelo fueron: PPB\_FluxSat (38.4%), el Índice de Área Foliar (LAI MCD) (26.2%), y el Índice de Vegetación Mejorado (EVI MYD) del producto MYD13A1 de MODIS (4.7%). Las demás variables tuvieron una importancia menor al 4% (Anexo II).

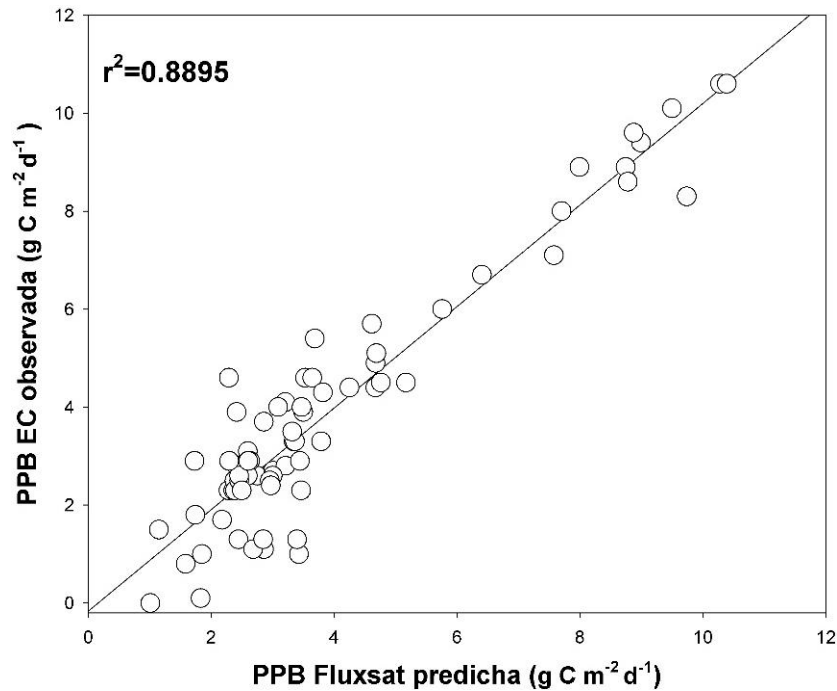


Figura 7. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en PPB\_Fluxsat y las observaciones de PPB derivadas de EC en el sitio Bernal, validadas con el 15% del conjunto de datos.

Con respecto al ML basado en la SIF de MODIS y las variables derivadas de PR, se obtuvieron las siguientes medidas de bondad de ajuste:  $r^2=0.93$  y  $RMSE = 0.60 \text{ g C m}^{-2} \text{ día}^{-1}$  (ver Figura 8). En la fase de entrenamiento, se generaron un total de 517 modelos y se seleccionó el mejor modelo en función del valor mínimo de devianza, correspondiente a un modelo de la familia Gradient Boosting Machine (GBM) con una devianza de 0.37. Las variables que más influyeron en este modelo para la estimación de la PPB, en orden descendente de importancia, fueron SIF\_MODIS (48.9%), la fracción de radiación fotosintéticamente activa (fPAR) del producto MCD15A2H de MODIS (15.4%), y el Índice de Área de Foliar (LAI MCD) (5.8%). El resto de las variables contribuyeron con menos del 5% de importancia (Anexo II).

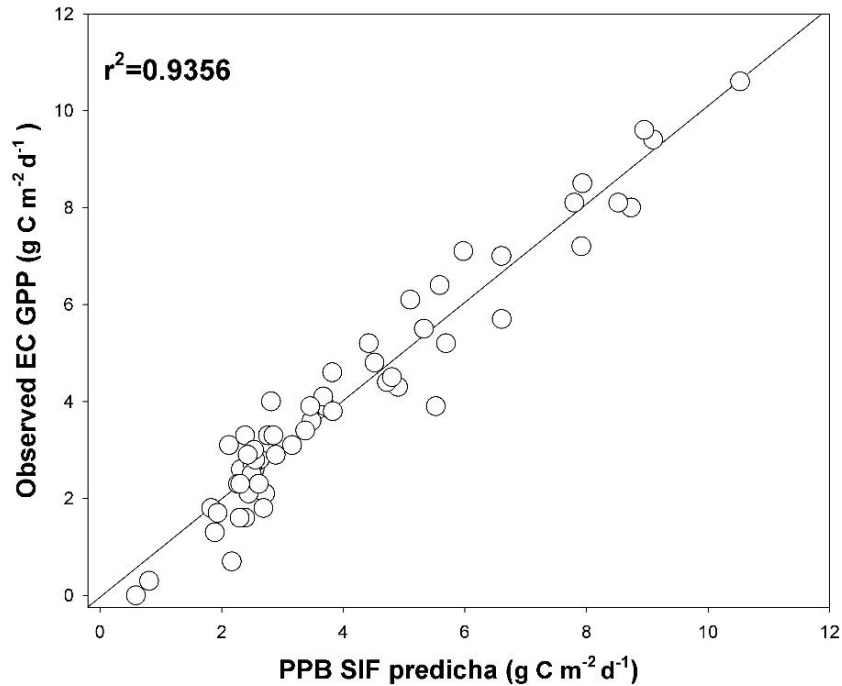


Figura 8. Concordancia entre las predicciones de PPB del modelo de aprendizaje automático basado en SIF de MODIS y las observaciones de PPB derivadas de EC en el sitio Bernal, validado con el 15% del conjunto de datos.

La evaluación de los modelos de estimación, PPB MODIS, PPB FluxSat y PPB SIF en comparación con PPB EC, reveló un desempeño sólido en sus estimaciones. Notablemente, el modelo PPB SIF destacó al obtener el coeficiente de determinación más alto y las mejores medidas de bondad de ajuste. En el Cuadro IV, se presentan a detalle las relaciones estadísticas entre las estimaciones de PPB y las observaciones de PPB de la torre de EC. Estos resultados subrayan la precisión del modelo PPB SIF en comparación con los otros modelos evaluados.

**Cuadro IV. Medidas de bondad de ajuste calculadas de los modelos ML basados en los productos satelitales y mejor modelo durante el proceso para estimar la PPB en la localidad de Bernal de 2017 a 2018.**

Modelo	$r^2$	RMSE	Sesgo	Devianza	Mejor modelo
PPB MODIS	0.87	0.84	0.06	0.71	XGBoost
PPB FluxSat	0.88	0.85	-0.08	0.73	Best of Family Stacked Ensemble
SIF MODIS	0.93	0.60	-0.02	0.36	GBM

## 8.2 Análisis estacional

### 8.2.1 Temporada de lluvias

A partir del del análisis en la temporada de lluvias, los resultados del modelo ML basado en el producto PPB\_MODIS mostró un buen desempeño con un  $r^2=0.90$  y  $RMSE = 0.86 \text{ g C m}^{-2} \text{ día}^{-1}$  (Figura 9). Se generaron un total de 913 modelos en la fase de entrenamiento, de los cuales se seleccionó el mejor modelo del tipo "XGBoost" con una devianza igual a 0.73. Las variables más influyentes que estimaron la PPB en este modelo fueron: Fotosíntesis neta (PsnNet) del producto MOD17A2H de MODIS (30.61%), el Índice de Área Foliar (LAI MCD) (17.85%), y el Índice de Vegetación Mejorado (EVI MYD) (11.89%). Las demás variables tuvieron una importancia menor al 10% (consulte Anexo III).

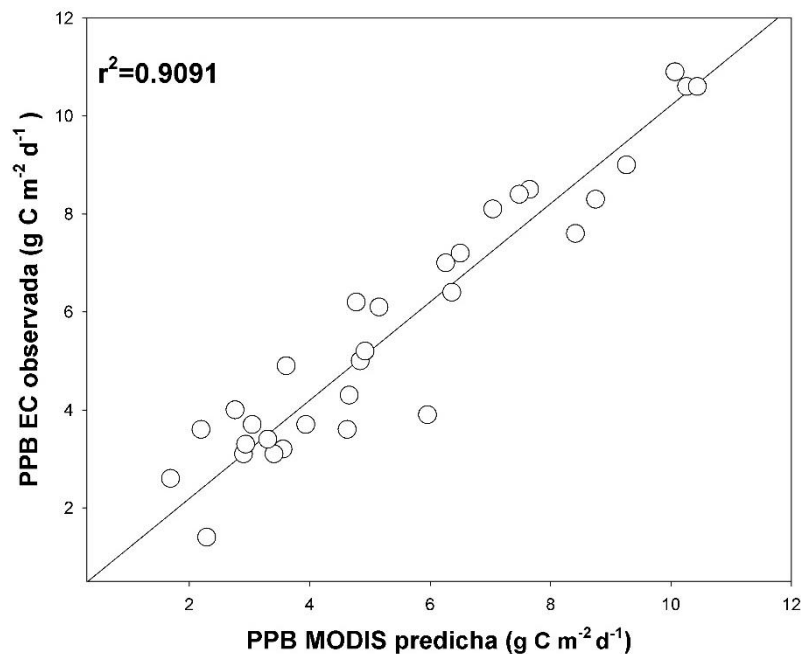


Figura 9. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB de MODIS frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.

En relación a la PPB predicha por el ML basado en FluxSat y su comparación con la PPB derivada de la torre de EC presenta un  $r^2=0.88$  y  $RMSE=1.02 \text{ g C m}^{-2} \text{ día}^{-1}$  (Figura 10). Durante el proceso de entrenamiento, se generaron un total de 529 modelos, y el mejor modelo, del tipo

"XGBoost", fue seleccionado con una devianza de 1.04. Las variables más influyentes en la estimación de la PPB en este modelo, ordenadas de mayor a menor importancia, fueron: PPB\_FluxSat (52.23%), el Índice de Reflectancia Global de Vegetación (GARI) (18.60%) y la Banda 14 (R\_14) (4.33%). Las demás variables tuvieron una relevancia menor al 4% (Anexo III).

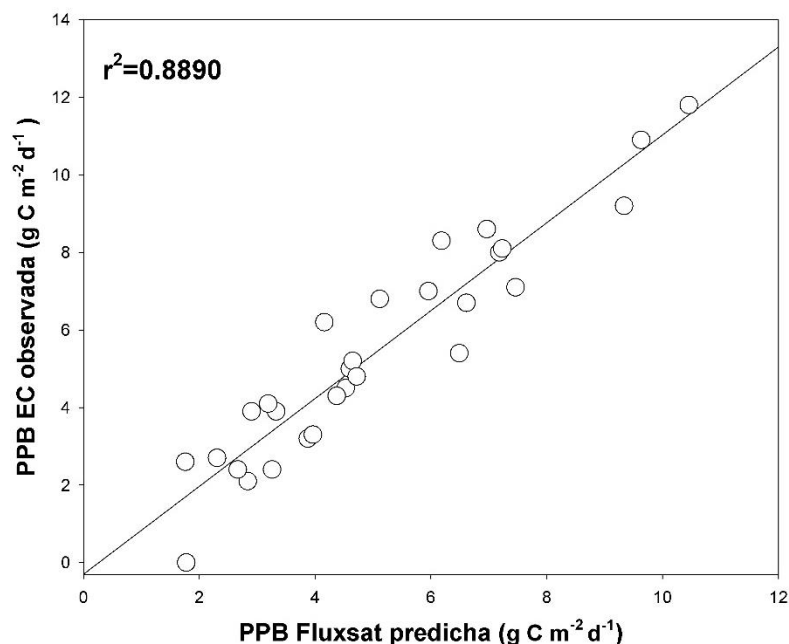


Figura 10. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB FluxSat frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.

Finalmente, el modelo ML para predecir la PPB durante la temporada de lluvias, basado en la señal de la SIF, exhibió un  $r^2=0.86$  y  $RMSE=0.80$  g C m<sup>-2</sup> día<sup>-1</sup> (Figura 11). Durante el proceso de entrenamiento, se generaron un total de 576 modelos, y se seleccionó el mejor modelo del tipo "XGBoost" con una devianza de 0.65. Las variables de mayor importancia en la estimación de la PPB en este modelo, clasificadas de mayor a menor relevancia, fueron: Índice de Área Foliar (LAI MCD) (53.46%), Señal de Fluorescencia del Producto MODIS (SIF MODIS) (19.83%), y Banda 14 (R\_14) (5.28%). Por otro lado, las demás variables tuvieron una relevancia inferior al 5% (Anexo III).

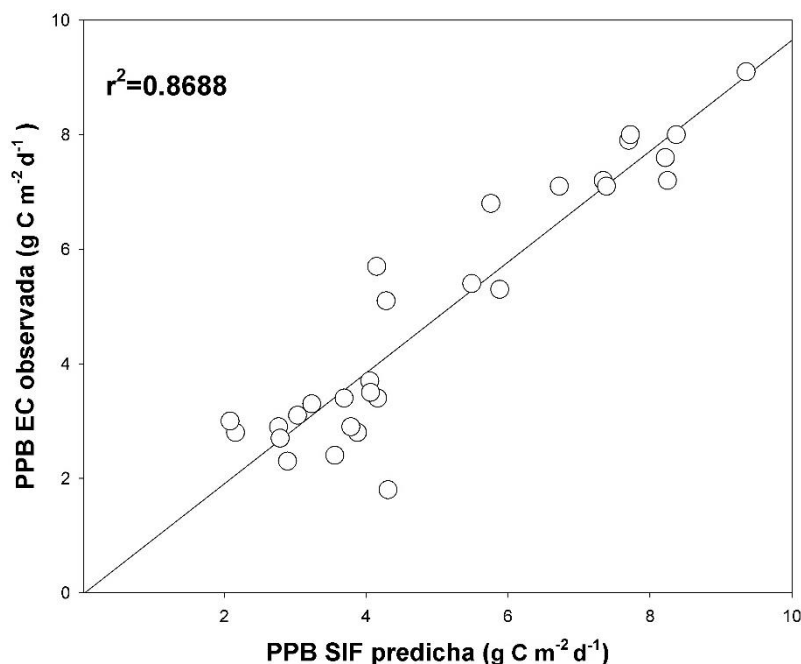


Figura 11. Concordancia entre las predicciones de PPB del modelo ML basado en la SIF de MODIS frente a las observaciones de PPB derivadas de EC durante la temporada de lluvias.

En el Cuadro V se presentan las relaciones estadísticas entre las estimaciones de PPB de los modelos ML y las observaciones de PPB de EC para la época de lluvias. El coeficiente de determinación más alto se presentó en modelo PPB MODIS, destacando la precisión en las predicciones. Sin embargo, al evaluar el modelo basado en SIF MODIS, es crucial considerar el RMSE, sesgo y devianza, ya que los valores bajos que presenta indican una alta precisión con respecto a los datos observados (Cuadro V).

**Cuadro V. Comparación de las medidas de bondad de ajuste de los modelos ML basados en los productos satelitales y mejor modelo para estimar la PPB en la localidad de Bernal en la temporada de lluvias de 2017 a 2018.**

Modelo	$r^2$	RMSE	Sesgo	Devianza	Mejor modelo
PPB MODIS	0.90	0.85	-0.20	0.73	XG Boost
PPB FluxSat	0.88	1.02	0.19	1.04	XG Boost
SIF MODIS	0.86	0.80	0.19	0.65	XG Boost

### 8.2.2 Temporada de sequía

El análisis estacional del modelo predictivo basado en PPB\_MODIS, dio como resultado las siguientes medidas de bondad de ajuste:  $r^2= 0.48$  y un  $RMSE=0.54 \text{ g C m}^{-2} \text{ día}^{-1}$ . La gráfica de dispersión se puede observar en la Figura 12. Se generaron un total de 544 modelos durante el proceso de entrenamiento y el mejor modelo de acuerdo a su valor de devianza (0.29) fue de tipo “GBM”. En este caso las variables de mayor importancia fueron, la temperatura de la superficie diurna del producto MODIS MOD11A2 (LST day) (18.84%), GARI con (12.62%), PPB\_MODIS (12.13%) y la Banda 13 (R\_13) del producto MODOCGA MODIS (11.98%) y el resto de las variables con menos del 10% (Anexo IV).

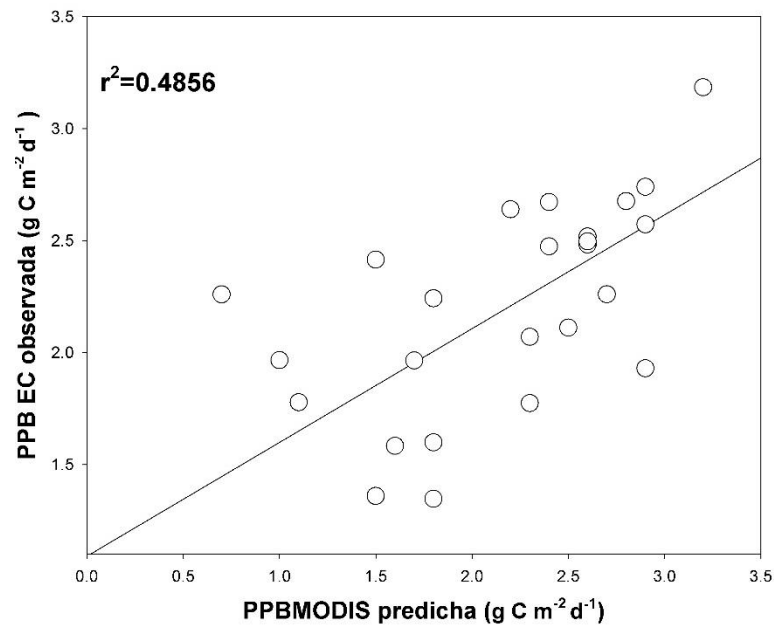


Figura 12. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB MODIS y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía.

El modelo ML basado en PPB FluxSat obtuvo un  $r^2= 0.58$  y un  $RMSE=0.56 \text{ g C m}^{-2} \text{ día}^{-1}$ . La relación estadística se puede observar en la Figura 13. Se generaron un total de 632 modelos y se seleccionó un “DRF” como mejor modelo por su valor de devianza =0.31. Las variables de mayor importancia para construir el modelo fueron la R\_13 (12.86%), LAI MCD (9.79%),

PsnNet (8.88%), LST day (8.72) y el resto de las variables obtuvieron menos de 7% de importancia (Anexo IV).

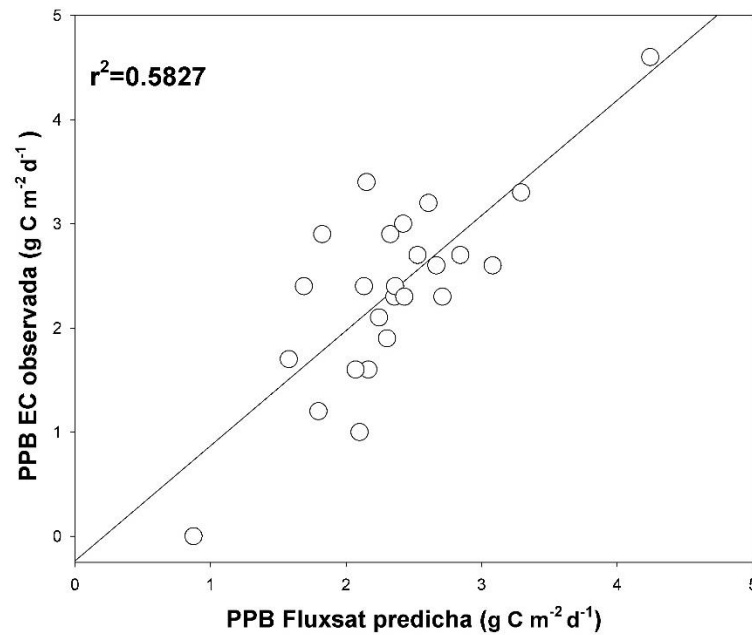


Figura 13. Concordancia entre las predicciones de PPB del modelo ML basado en la PPB FluxSat y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía

El modelo de PPB SIF vs PPB EC obtuvo un  $r^2= 0.40$  y  $RMSE = 0.69$  g C m<sup>-2</sup> día<sup>-1</sup> (Figura 11). Se produjeron un total de 607 modelos durante el proceso de entrenamiento y se seleccionó un “XGboost” como mejor modelo por su valor de devianza =0.48. Las variables de mayor importancia para construir dicho modelo fueron, la fPAR (15.89%), PsnNet (11.71%), LAI MCD (9.39%), SIF\_MODIS (9.03%) y el resto de las variables obtuvieron menos del 5% (Anexo IV).



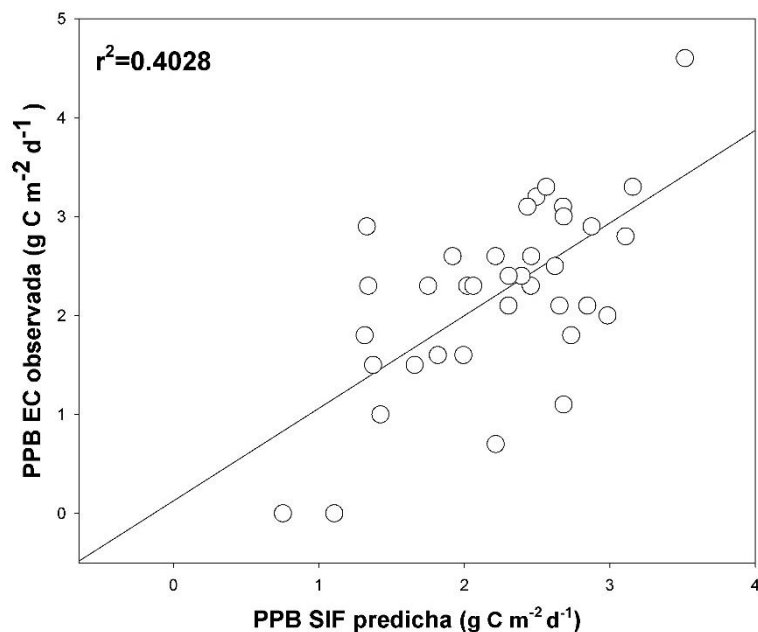


Figura 14. Concordancia entre las predicciones de PPB del modelo ML basado en la SIF de MODIS y las observaciones de PPB obtenidas mediante EC durante la temporada de sequía.

El Cuadro VI detalla las relaciones estadísticas obtenidas entre las estimaciones de PPB resultante de los modelos ML y las observaciones de PPB de EC para la temporada de sequía. En general, de acuerdo a las métricas obtenidas, los modelos de aprendizaje ML para la temporada de sequía exhibieron baja precisión en el desempeño de sus predicciones.

**Cuadro VI. Comparación de las medidas de bondad de ajuste de los modelos ML basados en los productos satelitales y mejor modelo durante el proceso de entrenamiento para estimar la PPB en la localidad de Bernal en la temporada de secas de 2017 a 2018.**

Modelo	$r^2$	RMSE	Sesgo	Devianza	Mejor modelo
PPB MOD17	0.48	0.54	0.07	0.29	GBM
PPB FluxSat	0.58	0.56	0.17	0.31	DRF
SIF MODIS	0.40	0.69	0.02	0.48	XG Boost

Cabe destacar que el coeficiente de determinación más alto se obtuvo en las estimaciones del modelo basado en PPB\_FluxSat (consultar Cuadro VI). Además, el modelo basado en PPB\_MODIS exhibió los valores más bajos de RMSE, sesgo y devianza (Cuadro VI).

## IX. Discusión

A lo largo de la presente investigación, se llevaron a cabo diferentes modelos de estimación de PPB. Para ello, se utilizaron dos productos satelitales de PPB, específicamente PPB\_MODIS y PPB\_FLuxsat y datos de fluorescencia proporcionados por el producto MODIS\_SIF. Con el fin de validar los hallazgos encontrados, comparamos con los registros de PPB recopilados por la torre de EC ubicada en el mismo sitio de estudio, caracterizado por un clima semiárido. En términos generales, los resultados sugieren que el modelo basado en SIF MODIS es competitivo y constituye una alternativa precisa para predecir la PPB en un ecosistema semiárido, confirmando así la hipótesis planteada.

### 9.1 Regresión simple

En el análisis general, realizado como punto de referencia, contempla una serie de tiempo intermitente de abril de 2017 a agosto de 2018, comprendiendo la variabilidad estacional de la PPB, encontramos una relación deficiente entre los productos satelitales seleccionados para este estudio y las mediciones observadas de PPB EC (PPB\_MODIS  $r^2 = 0.59$ , PPB\_Fluxsat  $r^2 = 0.54$ , y SIF\_MODIS  $r^2 = 0.65$ ). Cabe señalar que el entorno semiárido, agrega un nivel adicional de complejidad a las estimaciones por parte de los productos satelitales empleados en el presente trabajo.

Se ha documentado que los productos satelitales de PPB pueden presentar discrepancias en sus estimaciones en comparación con observaciones de PPB hechas por las torres de EC en ecosistemas semiáridos (Smith *et al.*, 2018). Estas discrepancias pueden surgir debido a variaciones en la resolución espacial y a las limitaciones presentes en los algoritmos utilizados en el desarrollo de dichos productos satelitales (Smith *et al.*, 2018). Además, la baja densidad vegetal y la alta variabilidad espacial y temporal en el área de estudio también pueden contribuir a estas diferencias (Huenneke *et al.*, 2001).

Una de las principales razones detrás del sesgo en las estimaciones de PPB\_MODIS en zonas semiáridas se relaciona con el enfoque de su algoritmo, el cual se basa en la suposición típica de un solo tipo de bioma por píxel, en lugar de una mezcla de tipos funcionales y estructurales, y en la entrada de datos de baja resolución para generar estimaciones globales (Wang *et al.*, 2019, Biederman *et al.*, 2017). Dado que Bernal se distingue por una vegetación espacialmente heterogénea, el producto de PPB\_MODIS, al asumir una vegetación homogénea, resulta en estimaciones poco precisas.

Además, Biederman *et al.*, (2017) encontró que el producto MODIS\_PPB puede subestimar la variación interanual de la PPB en comparación con EC en zonas semiáridas de América del Norte. Lo anterior puede deberse a la heterogeneidad de la vegetación presente en estos ecosistemas, así como a la existencia de especies con tolerancia a sequía, que los índices de vegetación utilizados en el algoritmo de PPB\_MODIS no siempre capturan, y no coinciden con la PPB real durante momentos de estrés hídrico (Biederman *et al.*, 2017; Smith *et al.*, 2018; Yan *et al.*, 2019). La diversidad botánica y la variabilidad espacial presente en Bernal (Guevara *et al.*, 2021), subrayan la necesidad de valores más específicos o locales que consideren las condiciones climáticas y el tipo de vegetación para obtener estimaciones más precisas de la PPB.

El producto PPB\_Fluxsat ha demostrado coherencia con la mayoría de los sitios de la FLUXNET y con la mayoría de las regiones del mundo que se consideraron en la generación de este producto (Bennett *et al.*, 2021; Lv *et al.*, 2023). Sin embargo, una de las limitaciones que se pueden atribuir las estimaciones de PPB\_Fluxsat con respecto a las observaciones en el sitio de estudio, es por el tamaño de píxel que ofrece a este producto (0.05° o 5.55 km por píxel). Además, la baja representatividad de FLUXNET correspondiente al monitoreo de ecosistemas semiáridos en Norte América y en especial para México, dificulta la estimación y la validación de PPB para la zona de estudio (Joiner *et al.*, 2018; Villarreal *et al.*, 2018)

La relación general de SIF\_MODIS con respecto a las observaciones de EC *in situ* fue ligeramente mejor ( $r^2=0.65$ ). Se ha comprobado una relación casi lineal entre la SIF captada por PR y la PPB en escalas de tiempo de un año continuo o más, a escala global (Joiner *et al.*, 2011; Guanter *et al.*, 2012; Köhler *et al.*, 2020). Sin embargo, para el presente estudio se contó con un año de datos intermitentes de PBB que pudieron haber influido en la relación SIF-GPP.

Es relevante destacar que la técnica de EC tiene una cobertura espacial típica de 1 km<sup>2</sup> para medir el CO<sub>2</sub> (Burba, 2013). La PPB de EC, con la cual se compara el producto SIF, cuenta con una cobertura de observación de 600 m<sup>2</sup> en el sitio de estudio. El producto SIF\_MODIS utilizado, presenta una resolución de 0.05°, lo cual es otro aspecto a considerar. La resolución espacial de las observaciones satelitales de SIF actualmente disponibles, sigue siendo un desafío para evaluar de manera completa la relación entre SIF y PPB. (Smith *et al.*, 2017). En estudios recientes también se ha observado una baja relación de SIF\_MODIS con la PPB en sitios de vegetación heterogénea, como matorral, caracterizados por ambiente árido, especialmente en temporada de sequía (Shekhar *et al.*, 2021). La baja relación observada entre SIF\_MODIS y la PPB en Bernal puede atribuirse a una combinación de factores. Dado que se trata de un matorral secundario, se ve afectado por el sobrepastoreo, por sus condiciones climáticas específicas y la estructura de la vegetación, dominada por herbáceas en este entorno semiárido (Guevara *et al.*, 2021)

## 9.2 Modelos de aprendizaje automático para predecir la PPB

En el análisis subsecuente, al emplear modelos de aprendizaje automático para estimar la PPB a partir de cada uno de los productos satelitales, las variables meteorológicas, reflectancias e índices de vegetación, se observó una marcada mejora en la precisión de las estimaciones de PPB (ML PPB\_MODIS  $r^2 =0.87$ , ML PPB\_Fluxsat  $r^2 =0.88$ , y ML SIF\_MODIS  $r^2 =0.93$ ). Asimismo, estos resultados respaldan la hipótesis de la actual investigación, que plantea que los modelos de aprendizaje automático

basados en la señal de la SIF mejoran las estimaciones de PPB en un ecosistema de clima semiárido en comparación con los otros productos satelitales evaluados en este estudio.

Los resultados de los modelos ML, sugieren que estos poseen la capacidad de capturar de manera efectiva las relaciones entre las variables predictoras y la PPB observada en la zona de estudio. Lo anterior, respalda la perspectiva de que los flujos de carbono pueden ser representados mediante algoritmos de aprendizaje automático utilizando variables captadas por percepción remota, tanto a escala local y ecosistémica (Tramontana *et al.*, 2016), como a escala global (Jung *et al.*, 2011; Joiner y Yoshida, 2020).

Comprender las variables que influyen significativamente en las estimaciones de la PPB es crucial para mejorar la precisión de los modelos, pero también nos proporciona un mejor entendimiento de los procesos implicados en los ecosistemas áridos y semiáridos. El análisis de las variables de importancia, nos permitió identificar dentro de cada modelo las variables clave que influyeron en la predicción de la PPB.

El mejor modelo basado en SIF evidenció como variables de mayor importancia la SIF\_MODIS, fPAR y LAI. La SIF\_MODIS fue una variable significativa de actividad fotosintética de la vegetación en el sitio de estudio para modelar la PPB. Lo anterior contribuye a las investigaciones sobre uso de la SIF como potencial indicador para estimar la PPB en ecosistemas, sobre todo en ecosistemas de zonas áridas (Porcar-Castel *et al.*, 2014; Smith *et al.*, 2017; Smith *et al.*; 2018).

La detección de SIF se ve poco afectada por la presencia de las nubes y reflectancia del suelo (Damm *et al.*, 2015; Smith *et al.*, 2018). Esta última propiedad cobra relevancia en zonas áridas, donde la baja densidad de cobertura vegetal amplifica la influencia de los suelos expuestos, y la luminosidad puede interferir en la captura de algunos IVs mediante técnicas de PR (Smith *et al.*, 2018). También, se ha demostrado que la SIF presenta

mayor sensibilidad a la actividad de la vegetación caducifolia (Zuromski et al., 2018). Las limitaciones de la SIF\_MODIS se asocian con su resolución del pixel del producto (0.05°o 5.5km) que pueden no capturar la heterogeneidad de la vegetación en el sitio (Shekhar et al., 2021). Por lo que es recomendable integrar este sensor a otros con mayor resolución espaciotemporal para mejorar la representatividad (Zhang et al., 2023).

Además, la influencia de variables como la fPAR y el índice LAI sugiere la importancia del área foliar y la radiación absorbida por la vegetación del sitio, en conjunto con la SIF, para la estimación de PPB a partir del modelo ML. El LAI es un índice que brinda información sobre la densidad y estructura de dosel, y es valioso para captar variaciones estacionales (Fensholt et al., 2004). En trabajos anteriores, se ha observado una asociación entre SIF y LAI (Yang et al., 2017), resaltando su relevancia conjunta de obtener observaciones terrestres para mejorar las estimaciones de la PPB.

En los modelos ML basados en PPB\_MODIS y PPB\_Fluxsat, las variables de mayor importancia fueron: LAI, la Banda R\_14 y PPB MODIS; así como PPB\_Fluxsat, LAI y EVI, respectivamente. Estos resultados destacan al índice LAI una vez más como unas de las variables de mayor impacto en la predicción de PPB en ambos casos. El índice LAI de MODIS utilizado en este trabajo, asume una estructura del dosel y del suelo son constantes en un área determinada (Running y Zhao, 2019) lo que puede no ser siempre válido en zonas áridas debido a la heterogeneidad de la vegetación. Sin embargo, la literatura señala que LAI es capaz de capturar los patrones de LAI característicos de estos ecosistemas si es a un intervalo de tiempo prolongado (Fensholt et al., 2004). Esto sugiere que, a pesar de las limitaciones, el LAI empleado puede proporcionar información valiosa sobre la vegetación en el sitio de estudio para estimar la PPB.

Es importante destacar, la importancia que presentaron las bandas R\_13 (662-672 nm) y R\_14 (673-683 nm) como variables para estimar la PPB en este estudio. Estas bandas, sensibles a la reflectancia de la clorofila y otros

pigmentos en la vegetación, han demostrado ser esenciales para monitorear la actividad fotosintética en otros trabajos, especialmente al calcular índices como el Photochemical Reflectance Index (PRI) (Liu *et al.*, 2022; Balde *et al.*, 2023). Por otro lado, los picos de fluorescencia de la SIF alrededor de 680 nm y 740 nm, asociados a la emisión de clorofila después de la absorción de luz, señalan la actividad fotosintética en sí (Porcar-Castell *et al.*, 2014; Mohammed *et al.*, 2019). Aunque la SIF destaca como un predictor efectivo, la combinación con las bandas espectrales 13 y 14, por su posición espectral, pueden servir como estrategia para una estimación completa de la PPB (Liu *et al.*, 2022; Balde *et al.*, 2023).

## 9.2 Análisis estacional

El análisis estacional mostró buen ajuste para las estimaciones de los modelos ML para la temporada de lluvias (PPB\_MODIS  $r^2 = 0.90$ , PPB\_Fluxsat  $r^2 = 0.88$ , y SIF\_MODIS  $r^2 = 0.86$ ), demostrando la capacidad de explicar gran parte de la variabilidad de las observaciones de PPB en esta temporada. Sin embargo, para la temporada de sequía se observa menor precisión en las estimaciones de PPB a partir de los modelos ML (MODIS\_PPB  $r^2 = 0.48$ , ML PPB\_Fluxsat  $r^2 = 0.58$ , y ML SIF\_MODIS  $r^2 = 0.40$ ).

La comparación de los resultados en ambas temporadas y durante todo el período de tiempo, proporciona una amplia visión de la capacidad de los modelos de ML para estimar la PPB en ecosistemas áridos. Esto nos permite comprender cómo estos modelos se ajustan a las variaciones estacionales.

En la temporada de lluvias, destacan variables clave y frecuentes en los modelos predictores de la PPB, como índices de vegetación y reflectancias, tales como el LAI, GARI y la reflectancia en la Banda 14. A pesar de que la precipitación no fue una variable destacada en esta sección del análisis estacional, las lluvias tienen un impacto significativo en la estructura y función de la vegetación, y, por ende, en la PPB (Zhang *et al.*, 2020). Lo

anterior puede atribuirse a la evidencia de la presencia de desfases temporales entre la precipitación y la respuesta de la actividad de la vegetación (Jeong *et al.*, 2017; Barnes *et al.*, 2021). Destacando la necesidad de explorar condiciones previas de precipitación, para lograr una representación más precisa de la dinámica de agua y carbono, especialmente en zonas áridas (Barnes *et al.*, 2021).

En términos generales, podemos señalar que, durante la temporada de sequía, los modelos exhibieron un rendimiento inferior debido a diversos factores. La calidad de los datos podría haber influido en el análisis considerando la intermitencia de la serie de tiempo en el periodo de 2017 a 2018. Por otro lado, los modelos basados en PR podrían presentar una baja precisión, ya que se ha observado que variables como fPAR, LAI, EVI entre otras, no capturan de manera efectiva las señales durante situaciones de estrés hídrico (Ha *et al.*, 2015; Sims *et al.*, 2014; Biederman *et al.*, 2017).

Las variables de mayor importancia en los distintos modelos para estimar la PPB durante la temporada de sequía se relacionan con índices de vegetación y reflectancias como: LstDay, GARI, Banda 13, LAI, y fPAR. A pesar de que estos índices no se asocian directamente con la actividad de la vegetación, pueden mostrar de manera indirecta las condiciones del entorno. Sin embargo, en este trabajo no fueron suficientes para representar la PPB en temporada de sequía.

## **X. Conclusiones**

A lo largo de esta investigación se evaluaron productos satelitales de estimación de PPB. Estos productos incluyeron PPB\_MODIS, PPB\_Fluxsat y SIF\_MODIS, y se validaron a escala local. Los resultados se compararon con registros de PPB obtenidos de una torre de EC en un entorno semiárido queretano. Los resultados del análisis general mostraron una baja relación entre las estimaciones de los productos satelitales y las mediciones de PPB registradas *in situ*



Posteriormente se emplearon modelos de aprendizaje automático (ML), que utilizaron como variables predictoras los datos de los productos satelitales, IVs y variables meteorológicas también captadas por PR. Este procedimiento mejoró significativamente la precisión de estimación de PPB. Los modelos basados en SIF mostraron mejoras notables en la precisión de las estimaciones en comparación con otros productos satelitales. Esto respalda la idea de que los algoritmos de ML basados en la SIF y variables pueden modelar de manera efectiva PPB del sitio de estudio.

El trabajo se amplió mediante un análisis estacional, cuyo propósito era evaluar si los modelos ML mantenían su rendimiento en la estimación de PPB durante temporada de lluvias y de sequía. Los resultados sugieren que los modelos de ML obtuvieron mejores resultados al utilizar el conjunto de datos del período completo; no obstante, también demostraron un desempeño satisfactorio durante la temporada de lluvias en comparación con la temporada de sequía. La precisión durante la temporada de sequía puede relacionarse con la baja sensibilidad para capturar de manera efectiva las variables por PR en condiciones de estrés hídrico.

El estudio también identificó variables de importancia que influyen en las estimaciones de PPB, como el índice de área foliar LAI y fracción de radiación fotosintéticamente activa fPAR. Estas variables y otras, surgieron de manera recurrente en los modelos de estimación, subrayando su relevancia en la predicción precisa del PPB. Estos resultados pueden proporcionar una guía para simplificar la base de datos y mejorar los procesos de estimación.

Se destacó que los productos satelitales, como PPB\_MODIS, pueden presentar sesgos en zonas semiáridas debido a la entrada de datos en su algoritmo, específicamente con respecto a los valores máximos de eficiencia de uso de la luz determinado por tipo de bioma. Por lo tanto, se recomienda considerar valores específicos para zonas áridas y locales a fin de mejorar la precisión. En el caso de PPB\_FLUXSAT, a pesar de la precisión de sus estimaciones globales, la baja resolución espacial y la falta

de sitios de referencia de la FLUXNET para zonas semiáridas pueden limitar su validez en el sitio de estudio y sitios con cualidades similares.

Se espera que el producto SIF\_MODIS evolucione para ofrecer valores locales y una mayor resolución espacial. Además, se sugiere llevar a cabo una evaluación a lo largo de un intervalo temporal más amplio con SIF\_MODIS, con el fin de analizar su capacidad para representar de las variaciones interanuales de PPB.

De manera general, la metodología empleada ofrece una forma efectiva y accesible de analizar la captura de carbono en el área de estudio. El monitoreo continuo de la captura de carbono ofrece una herramienta valiosa para comprender la dinámica de carbono el ecosistema estudiado. Finalmente, los datos adquiridos mediante estas técnicas pueden tener un impacto significativo en los tomadores de decisiones sobre la administración de recursos naturales y las estrategias que permitan una planificación sustentable.

## Referencias

- Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneeth, A., Jung, M., Canadell, J. G. P., Jain, AK., Kato, E., Poulter, B., Sitch, S., Stocker, BD, Viovy, N., Wang, YP, Wiltshire, A., Zaehle, S. y Zeng, N. 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO<sub>2</sub> sink. *Science* (80-.), 348(6237), 895-899.
- Bai, Y., Liang, S., y Yuan, W. 2021. Estimating global gross primary production from sun-induced chlorophyll fluorescence data and auxiliary information using machine learning methods. *Remote Sensing*, 13(5), 963.
- Balde, H., Hmimina, G., Goulas, Y., Latouche, G., & Soudani, K. (2023). Synergy between TROPOMI sun-induced chlorophyll fluorescence and MODIS spectral reflectance for understanding the dynamics of gross primary productivity at Integrated Carbon Observatory System (ICOS) ecosystem flux sites. *Biogeosciences*, 20(7), 1473-1490.
- Baldocchi, D. 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology* 9:479–492.
- Baldocchi, D. 2014. Measuring fluxes of trace gases and energy between ecosystems and the atmosphere—the state and future of the eddy covariance method. *Global change biology*, 20(12), 3600-3609.
- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., y Papale, D. 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, 329(5993), 834-838.
- Bennett, B., Joiner, J., y Yoshida, Y. 2021. Validating satellite based FluxSat v2. 0 Gross Primary Production (GPP) trends with FluxNet 2015 eddy covariance observations. In *AGU Fall Meeting Abstracts* (Vol. 2021, pp. B55F-1272).
- Biederman, J. A., Scott, R. L., Bell, T. W., Bowling, D. R., Dore, S., Garatuza-Payan, J., y Goulden, M. L. 2017. CO<sub>2</sub> exchange and evapotranspiration across dryland ecosystems of southwestern North America. *Global Change Biology*, 23(10), 4204-4221.
- Balde, H., Hmimina, G., Goulas, Y., Latouche, G., y Soudani, K. 2023. Synergy between TROPOMI sun-induced chlorophyll fluorescence and MODIS spectral reflectance for understanding the dynamics of gross primary

- productivity at Integrated Carbon Observatory System (ICOS) ecosystem flux sites. *Biogeosciences*, 20(7), 1473-1490.
- Bland, J. M., y Altman, D. G. 2010. Statistical methods for assessing agreement between two methods of clinical measurement. *International journal of nursing studies*, 47(8), 931-936.
- Byrne, B., Liu, J., Bloom, A. A., Bowman, K. W., Butterfield, Z., Joiner, J., y Yin, Y. 2020. Contrasting regional carbon cycle responses to seasonal climate anomalies across the east-west divide of temperate North America. *Global biogeochemical cycles*, 34(11), e2020GB006598.
- Cheng, Y. B., Zhang, Q., Lyapustin, A. I., Wang, Y., y Middleton, E. M. 2014. Impacts of light use efficiency and fPAR parameterization on gross primary production modeling. *Agricultural and Forest Meteorology*, 189, 187-197.
- CICESE, (Centro de Investigación Científica y de Educación Superior de Ensenada) 2015. Base de datos climatológica nacional (CLICOM)., Villa Bernal, Querétaro. <http://clicom-mex.cicese.mx> (Consulta: 28 de agosto de 2019)
- Campoli, M., Malhi, Y., Vicca, S., Luysaert, S., Papale, D., Peñuelas, J., y Janssens, I. A. 2016. Evaluating the convergence between eddy-covariance and biometric methods for assessing carbon budgets of forests. *Nature communications*, 7(1), 1-12.
- Crawley, M. J. 2012. *The R book*. John Wiley & Sons.
- Damm, A., Guanter, L., Paul-Limoges, E., Van der Tol, C., Hueni, A., Buchmann, N., y Schaepman, M. E. 2015. Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sensing of Environment*, 166, 91-105.
- Fensholt, R., Sandholt, I., y Rasmussen, M. S. 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote sensing of Environment*, 91(3-4), 490-507.
- Frankenberg, C., Fisher, J.B., Worden, J., Badgley, G., Saatchi, S.S., Lee, J.-E., Toon, G.C., Butz, A., Jung, M., Kuze, A., Yokota, T., 2011. New global observations of the terrestrial carbon cycle from GOSAT: patterns of plant fluorescence with gross primary productivity. *Geophys. Res. Lett.* 38, L17706.

- Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollack, R., Taylor, T.E., 2014. Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sens. Environ.* 147, 1–12.
- Green, P. J., y Silverman, B. W. 1994. In *Nonparametric Regression and Generalized Linear Models* (pp. 1-10). Springer US.
- Green, P. J., y Silverman, B. W. 1994. Thin plate splines. In *Nonparametric Regression and Generalized Linear Models* (pp. 137-161). Springer US.
- Gotelli, N. J., y Ellison, A. M. (2004). *Regression. A primer of ecological statistics. Vol 1.* (pp 237-287) Sunderland: Sinauer Associates.
- Gu, L., Wood, J. D., Chang, C. Y.-Y., Sun, Y., y Riggs, J. S. 2019. Advancing terrestrial ecosystem science with a novel automated measurement system for sun-induced chlorophyll fluorescence for integration with Eddy covariance flux networks. *Journal of Geophysical Research: Biogeosciences*, 124, 127–146.
- Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., 2012. Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sensing of Environment*, 121, 236–251
- Guanter, L., Aben, I., Tol, P., Krijger, J.M., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., Landgraf, J., 2015. Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. *Atmos. Meas. Tech.* 8, 1337–1352.
- Guevara-Escobar, A., González-Sosa, E., Cervantes-Jiménez, M., Suzán-Azpiri, H., Queijeiro-Bolaños, M. E., Carrillo-Ángeles, I., y Cambrón-Sandoval, V. H. 2021. Machine learning estimates of eddy covariance carbon flux in a scrub in the Mexican highland. *Biogeosciences*, 18(2), 367-392.
- Hall, P., Gill, N., Kurka, M., Phan, W., y Bartz, A. 2019. *Machine Learning Interpretability with H2O Driverless AI*, edited by: Bartz, A., H2O.ai Inc., California, US.
- Henrich, V., Krauss, G., Götze, C., y Sandow, C. 2012. *Entwicklung einer Datenbank für Fernerkundungsindizes.* Bochum: AK Fernerkundung, 15.
- Hill, M. J., Held, A. A., Leuning, R., Coops, N. C., Hughes, D., y Cleugh, H. A. 2006. MODIS spectral signals at a flux tower site: Relationships with high-

- resolution data, and CO<sub>2</sub> flux and light use efficiency measurements. *Remote Sensing of Environment*, 103(3), 351-368.
- Holloway, J., y Mengersen, K. 2018. Statistical machine learning methods and remote sensing for sustainable development goals: a review. *Remote Sensing*, 10(9), 1365.
- Hoover, D. L., Pfennigwerth, A. A., & Duniway, M. C. 2021. Drought resistance and resilience: The role of soil moisture–plant interactions and legacies in a dryland ecosystem. *Journal of Ecology*, 109(9), 3280-3294.
- Huenneke, L. F., Clason, D., y Muldavin, E. 2001. Spatial heterogeneity in Chihuahuan Desert vegetation: implications for sampling methods in semi-arid ecosystems. *Journal of Arid Environments*, 47(3), 257-270.
- Jeong, S. J., Schimel, D., Frankenberg, C., Drewry, D. T., Fisher, J. B., Verma, M., y Joiner, J. 2017. Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests. *Remote sensing of Environment*, 190, 178-187.
- Jin, C., Xiao, X., Merbold, L., Arneeth, A., Veenendaal, E., y Kutsch, W. L. 2013. Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. *Remote Sensing of Environment*, 135, 189-201.
- Joiner, J., Yoshida, Y., Vasilkov, A.P., Yoshida, Y., Corp, L.A., Middleton, E.M., 2011. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences* 8, 637–651.
- Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A.P., Middleton, E.M., Huemmrich, K.F., Yoshida, Y., Frankenberg, C., 2013. Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2. *Atmos. Meas. Tech.* 6, 2803–2823.
- Joiner, J., Yoshida, Y., Vasilkov, A., Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S., Middleton, E.M., Huemmrich, K.F., Gu, L., Marchesini, L.B., 2014. The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange. *Remote Sens. Environ.* 152, 375–391.
- Joiner, J., Yoshida, Y., Zhang, Y., Duveiller, G., Jung, M., Lyapustin, A., y Tucker, C. J. 2018. Estimation of terrestrial global gross primary production (GPP)

- with satellite data-driven models and eddy covariance flux data. *Remote Sensing*, 10(9), 1346.
- Joiner, J., y Yoshida, Y. 2020. Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales. *Agricultural and Forest Meteorology*, 291, 108092.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., y Williams, C. 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research: Biogeosciences*, 116(G3).
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., y Walker, A. 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach.
- Köhler, P., Behrenfeld, M. J., Landgraf, J., Joiner, J., Magney, T. S., Frankenberg, C. 2020. Global retrievals of solar-induced chlorophyll fluorescence at red wavelengths with TROPOMI. *Geophysical Research Letters*, 47(15), e2020GL087541.
- Lal, R., Griffin, M., Apt, J., Lave, L., y Morgan, M. G. 2004. Managing soil carbon. *Science*, 304(5669), 393-393.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., y Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3-10.
- LeDell, E., y Poirier, S. 2020. H2o automl: Scalable automatic machine learning. In *Proceedings of the AutoML Workshop at ICML (Vol. 2020)*. ICML.
- LI-COR, B. 2019. Eddypro Software instruction manual, Nebraska, US, version 7.0.
- Li, X., y Xiao, J. 2019. A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data. *Remote Sensing*, 11(5), 517.
- Liu, P., Zha, T., Jia, X., Black, T. A., Jassal, R. S., Ma, J., ... y Wu, Y. 2019. Different effects of spring and summer droughts on ecosystem carbon and water exchanges in a semiarid shrubland ecosystem in Northwest China. *Ecosystems*, 22, 1869-1885.
- Liu, Y., Wu, C., Tian, F., Wang, X., Gamon, J. A., Wong, C. Y., y Jassal, R. S. 2022. Modeling plant phenology by MODIS derived photochemical reflectance index (PRI). *Agricultural and Forest Meteorology*, 324, 109095.

- Locatelli, B., y Kanninen, M. 2010. Servicios ecosistémicos y adaptación al cambio climático. En *Adaptación al cambio climático y servicios ecosistémicos en América Latina*, 11-20.
- Lv, Y., Liu, J., He, W., Zhou, Y., Nguyen, N. T., Bi, W., y Chen, H. 2023. How well do light-use efficiency models capture large-scale drought impacts on vegetation productivity compared with data-driven estimates?. *Ecological Indicators*, 146, 109739.
- Mauder, M. y Foken, T. 2011. Documentation and instruction manual of the eddy-covariance software package TK3, Universität Bayreuth Abteilung Mikrometeorologie, Bayreuth, Germany.
- Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., & Moreno, J. 2009. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote sensing of environment*, 113(10), 2037-2051.
- Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., ...y Zarco-Tejada, P. J. 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote sensing of environment*, 231, 111177.
- Monteith, J. 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, 9(3), 747-766.
- Olden, J. D., Lawler, J. J., y Poff, N. L. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, 83(2), 171-193.
- Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., y Law, B. 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific data*, 7(1), 1-27.
- Parazoo, N. C., Frankenberg, C., Köhler, P., Joiner, J., Yoshida, Y., Magney, T. 2019. Towards a harmonized long-term spaceborne record of far-red solar-induced fluorescence. *Journal of Geophysical Research: Biogeosciences*, 124, 2518–2539.
- Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., y Zhu, X. 2019. Plant phenology and global climate change: Current progresses and challenges. *Global change biology*, 25(6), 1922-1940.
- Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfündel, E. E., y Berry, J. A. 2014. Linking chlorophyll a fluorescence to



- photosynthesis for remote sensing applications: mechanisms and challenges. *Journal of experimental botany*, 65(15), 4065-4095.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., ... y Valentini, R. 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global change biology*, 11(9), 1424-1439.
- Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., y Wattenbach, M. 2013. Climate extremes and the carbon cycle. *Nature*, 500(7462), 287-295.
- Rotenberg, E., y Yakir, D. 2010. Contribution of semi-arid forests to the climate system. *Science*, 327(5964), 451-454
- Running, S. W., y Zhao, M. 2019. User's guide daily GPP and annual NPP (MOD17A2H/A3H) and year-end gap-filled (MOD17A2HGF/A3HGF) products NASA Earth Observing System MODIS land algorithm. (For Collection 6). University of Montana, Missoula, MT.
- Seastedt, T. R., Hobbs, R. J., y Suding, K. N. 2008. Management of novel ecosystems: are novel approaches required?. *Frontiers in Ecology and the Environment*, 6(10), 547-553.
- Shekhar, A., Buchmann, N., & Gharun, M. (2022). How well do recently reconstructed solar-induced fluorescence datasets model gross primary productivity?. *Remote Sensing of Environment*, 283, 113282.
- Smith, W. K., Biederman, J. A., Scott, R. L., Moore, D. J. P., He, M., Kimball, J. S., y Litvak, M. E. 2018. Chlorophyll fluorescence better captures seasonal and interannual gross primary productivity dynamics across dryland ecosystems of southwestern North America. *Geophysical Research Letters*, 45, 748–75
- Smith, W. K., Dannenberg, M. P., Yan, D., Herrmann, S., Barnes, M. L., Barron-Gafford, G. A., y Yang, J. 2019. Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. *Remote Sensing of Environment*, 233, 111401.
- Smith, W. K., Biederman, J. A., Scott, R. L., Moore, D. J. P., He, M., Kimball, J. S., y Litvak, M. E. 2018. Chlorophyll fluorescence better captures seasonal and interannual gross primary productivity dynamics across dryland ecosystems of southwestern North America. *Geophysical Research Letters*, 45(2), 748-757.

- Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., Drewry, D. T., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science*, 358(6360), eaam5747.
- Tarin-Terrazas, T., Alvarado-Barrientos, S., Cueva-Rodríguez, A., Hinojo-Hinojo, C., del Castillo, E. G., Sánchez-Mejía, Z., y Yépez-González, E. A. 2020. MexFlux: sinergias para diseñar, evaluar e informar soluciones climáticas naturales. *Elementos para Políticas Públicas*, 4(2), 99-118.
- Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, 1, e8621.
- Turner, D.P., S. Urbanski, D. Bremer, S.C. Wofsy, T. Meyers . 2003. A cross-biome comparison of daily light use efficiency for gross primary production. *Global Change Biology* 9:383–395.
- Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., y Papale, D. 2015. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sensing of Environment*, 168, 360-373.
- Uribe -Botero, E. 2015. El cambio climático y sus efectos en la biodiversidad en América Latina.
- Villarreal, S., y Vargas, R. 2021. Representativeness of FLUXNET sites across Latin America. *Journal of Geophysical Research: Biogeosciences*, 126(3), e2020JG006090.
- Villarreal, S., Guevara, M., Alcaraz-Segura, D., Brunsell, N. A., Hayes, D., Loescher, H. W., y Vargas, R. 2018. Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. *Agricultural and Forest Meteorology*, 262, 423-433.
- Wang, Z., Xiao, X., y Yan, X. 2010. Modeling gross primary production of maize cropland and degraded grassland in northeastern China. *Agricultural and Forest Meteorology*, 150(9), 1160-1167.
- Wang, H., Li, X., Ma, M., & Geng, L. 2019. Improving estimation of gross primary production in dryland ecosystems by a model-data fusion approach. *Remote Sensing*, 11(3), 225.
- Wolf, S., Keenan, T. F., Fisher, J. B., Baldocchi, D. D., Desai, A. R., Richardson, A. D., y van der Laan-Luijkx, I. T. 2016. Warm spring reduced carbon cycle

- impact of the 2012 US summer drought. *Proceedings of the National Academy of Sciences*, 113, 5880–5885.
- Wu, X., Xiao, X., Zhang, Y., He, W., Wolf, S., Chen, J., y Blanken, P. D. 2018. Spatiotemporal consistency of four gross primary production products and solar-induced chlorophyll fluorescence in response to climate extremes across CONUS in 2012. *Journal of Geophysical Research: Biogeosciences*, 123(10), 3140-3161.
- Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., y Zhang, X. 2019. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, 233, 111383.
- Yan, D., Scott, R. L., Moore, D. J. P., Biederman, J. A., y Smith, W. K. 2019. Understanding the relationship between vegetation greenness and productivity across dryland ecosystems through the integration of PhenoCam, satellite, and eddy covariance data. *Remote sensing of environment*, 223, 50-62.
- Yang, H., Yang, X., Zhang, Y., Heskell, M. A., Lu, X., Munger, J. W., y Tang, J. 2017. Chlorophyll fluorescence tracks seasonal variations of photosynthesis from leaf to canopy in a temperate forest. *Global Change Biology*, 23(7), 2874-2886.
- Yu, L., J. Wen, C.Y. Chang, C. Frankenberg, y Y. Sun. 2021. High Resolution Global Contiguous SIF Estimates from OCO-2 SIF and MODIS, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA
- Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., y Wofsy, S. C. 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agricultural and Forest Meteorology*, 143(3-4), 189-207.
- Zarco-Tejada, P. J., Catalina, A., González, M. R., & Martín, P. 2013. Relationships between net photosynthesis and steady-state chlorophyll fluorescence retrieved from airborne hyperspectral imagery. *Remote Sensing of Environment*, 136, 247-258.
- Zhang, L., Zhou, D., Fan, J., Guo, Q., Chen, S., Wang, R., y Li, Y. 2019. Contrasting the performance of eight satellite-based GPP models in water-limited and temperature-limited grassland ecosystems. *Remote Sensing*, 11(11), 1333.

- Zhang, L., Xiao, J., Zheng, Y., Li, S., y Zhou, Y. 2020. Increased carbon uptake and water use efficiency in global semi-arid ecosystems. *Environmental Research Letters*, 15(3), 034022.
- Zhang, Y., Gentine, P., Luo, X., Lian, X., Liu, Y., Zhou, S., y Keenan, T. F. 2022. Increasing sensitivity of dryland vegetation greenness to precipitation due to rising atmospheric CO<sub>2</sub>. *Nature communications*, 13(1), 4875.
- Zhang, R., Zhou, Y., Hu, T., Sun, W., Zhang, S., Wu, J., y Wang, H. 2023. Detecting the Spatiotemporal Variation of Vegetation Phenology in Northeastern China Based on MODIS NDVI and Solar-Induced Chlorophyll Fluorescence Dataset. *Sustainability*, 15(7), 6012.
- Zuromski, L. M., Bowling, D. R., Köhler, P., Frankenberg, C., Goulden, M. L., Blanken, P. D., y Lin, J. C. 2018. Solar-induced fluorescence detects interannual variation in gross primary production of coniferous forests in the Western United States. *Geophysical Research Letters*, 45(14), 7184-7193.

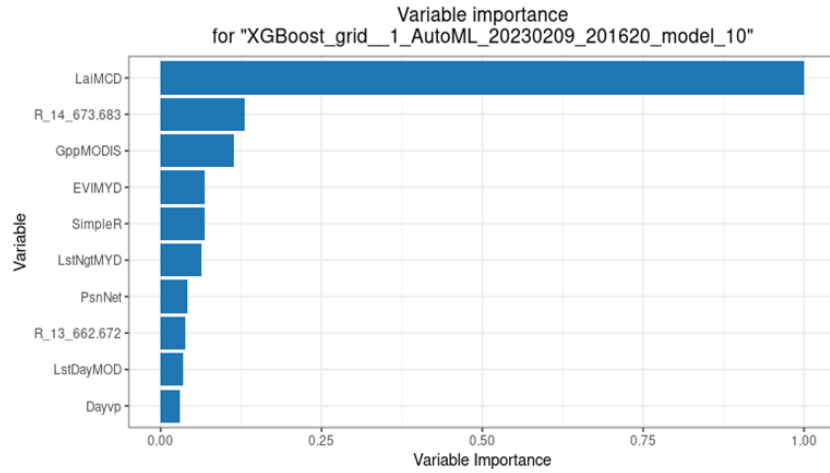
## Anexos

### Anexo I. Características de las reflectancias e índices de MODIS

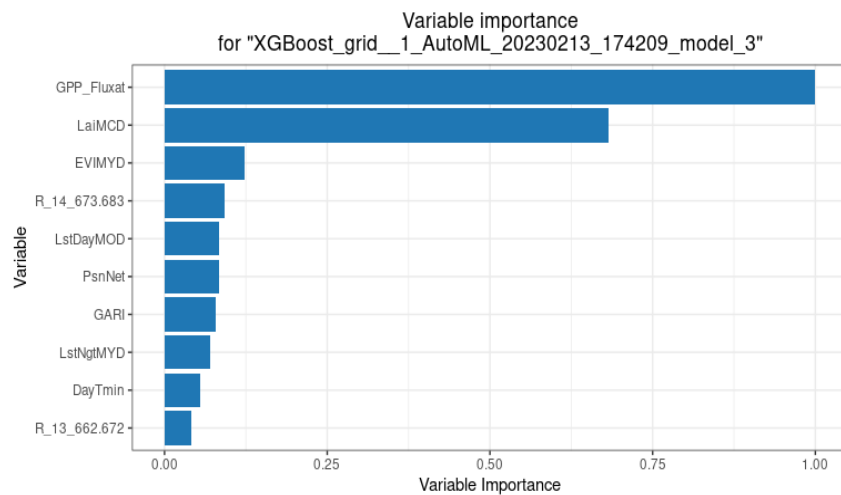
Nombre del producto	Capa satelital	Resolución espacial	Resolución temporal (d)	Resolución espectral (nm)
MOD09GA	Banda 1	500 m	1	620–670
	Banda 2			841–876
	Banda 3			459–479
	Banda 4			545–565
	Banda 5			1230–1250
	Banda 6			1628–1652
	Banda 7			2105–2155
MOD0CGA	Banda 8	1 km		405–420
	Banda 9			438–448
	Banda 10			483–493
	Banda 11			526–536
	Banda 12			546–556
	Banda 13			662–672
	Banda 14			673–683
MCD15A2H	Fracción de radiación fotosintéticamente activa (fPAR) Índice de área foliar (LAI)	1 km	8	NA
MOD11A2	Temperatura de la superficie y emisividad del día (LST day)	1 km	8	
MYD11A2	Temperatura de la superficie y emisividad nocturna (LST night)			
MYD13A1	Índice de vegetación mejorado (EVI)	500m	16	
MOD17A2H	Productividad Primaria Bruta (GPP) Fotosíntesis neta (PsnNet)	500m	8	

**Anexo II. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB de 2017 a 2018: a) Modelo XG Boost basado en PPB\_MODIS; b) Best of Family Stacked Ensemble basado en PPB\_FluxSat; c) GBM basado en la SIF\_MODIS.**

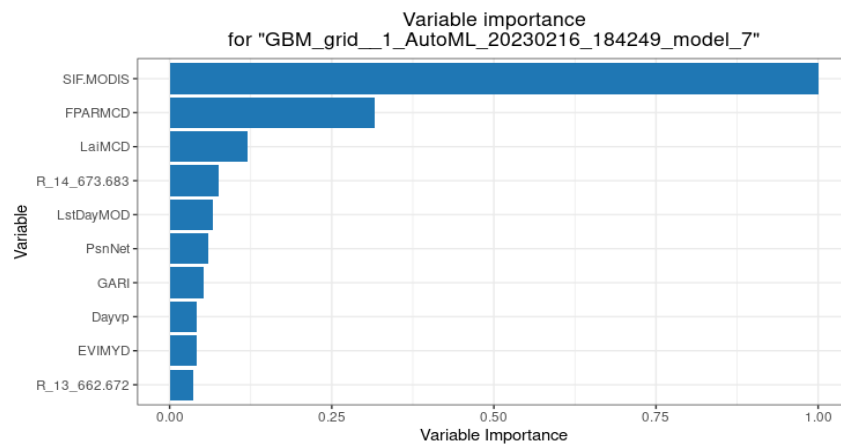
**a)**



**b)**

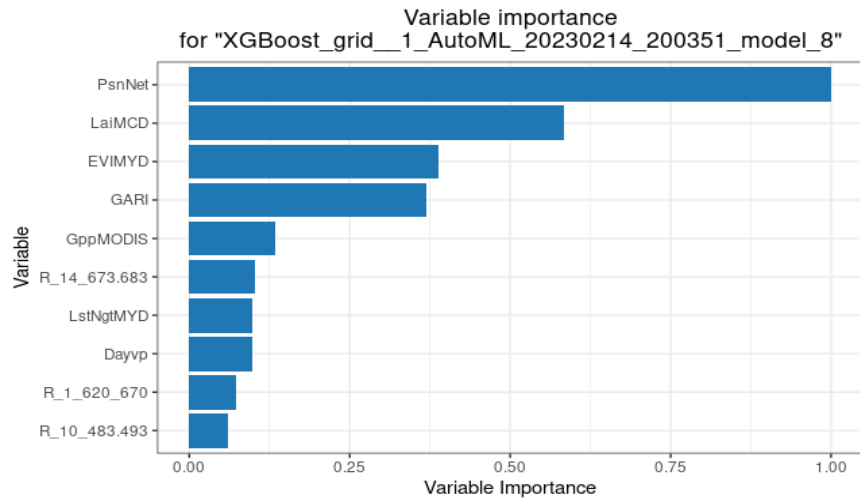


**c)**

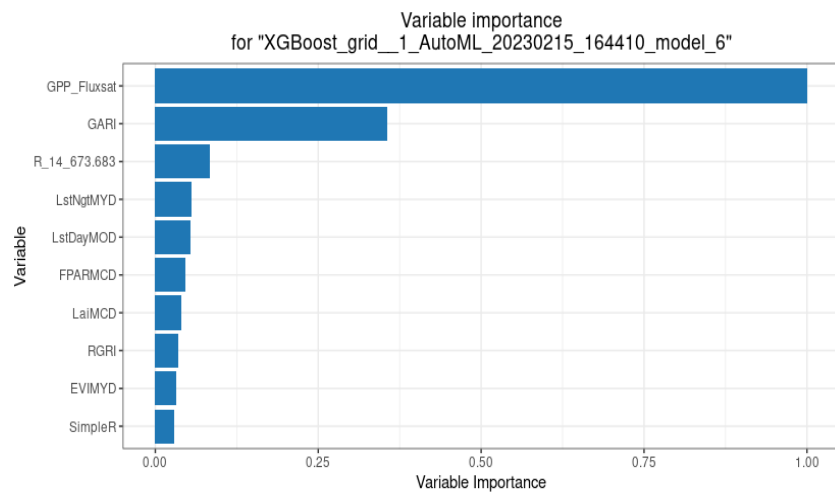


Anexo III. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB en temporada de lluvias: a) XG Boost basado en PPB\_MODIS; b) XG Boost basado en PPB\_FluxSat; c) XG Boost basado en la SIF\_MODIS.

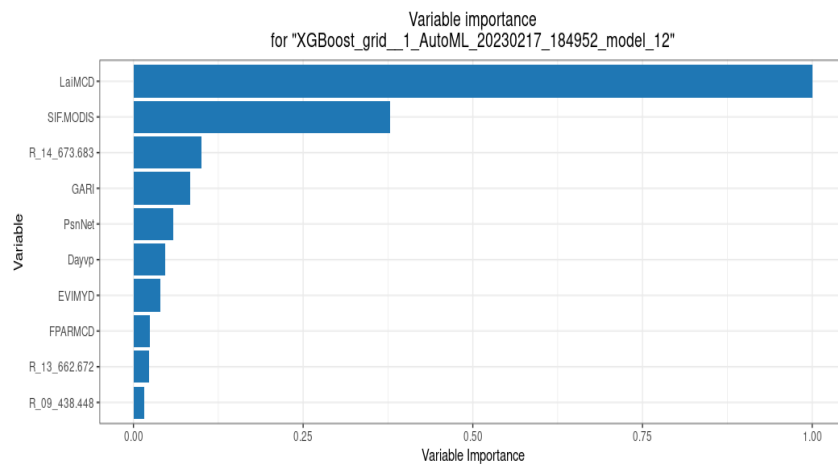
a)



b)

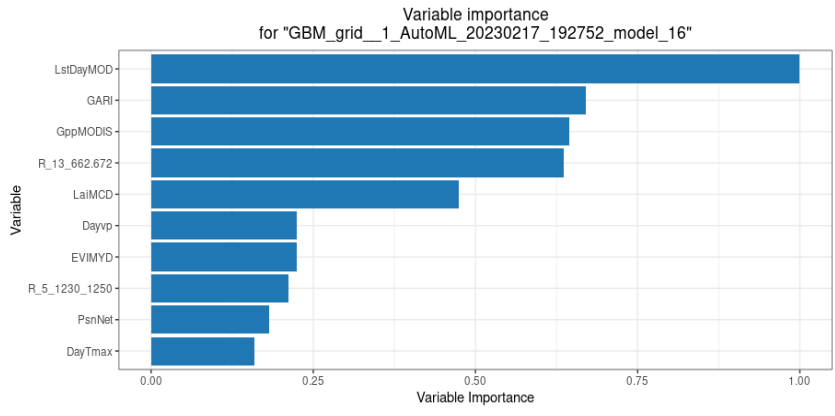


c)

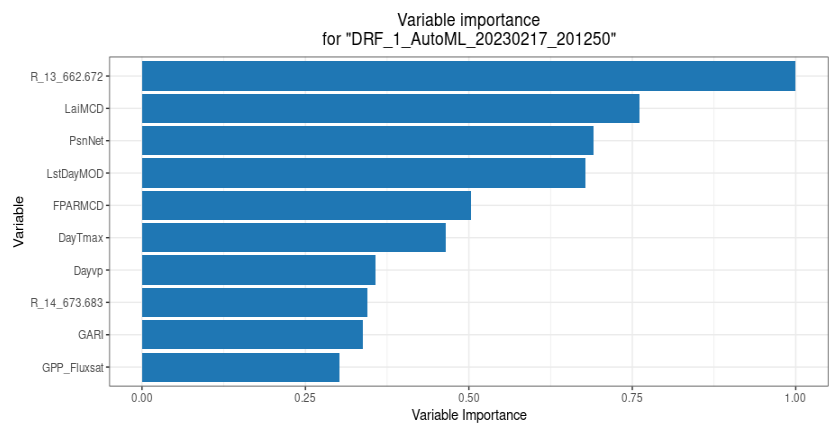


**Anexo IV. Gráficos de la importancia relativa de las variables que tuvieron mayor influencia en los modelos individuales para la predicción de la PPB en temporada de sequía: a) GBM basado en PPB\_MODIS; b) DRF basado en PPB\_FluxSat; c) XGBoost basado en la SIF\_MODIS.**

**a)**



**b)**



**c)**

