From Hyperspectral Indices to Global Fluorescence: PACE Vegetation Indices as Predictors of Terrestrial Photosynthesis

Caleb Sloan Harmon

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Forestry

Randolph H. Wynne, Co-Chair
Valerie A. Thomas, Co-Chair
E. Fred Huemmrich

December 8, 2025
Blacksburg, VA

Keywords: Solar-induced chlorophyll fluorescence, PACE mission, Ocean Color Instrument, Hyperspectral vegetation indices, Enhanced Vegetation Index, Chlorophyll Index Red Edge, Gross Primary Productivity, Global terrestrial ecosystems, Carbon cycle monitoring

From Hyperspectral Indices to Global Fluorescence: PACE Vegetation Indices as Predictors of Terrestrial Photosynthesis

Caleb Sloan Harmon

ABSTRACT

Solar-induced chlorophyll fluorescence (SIF) serves as a direct remotely sensed indicator of photosynthetic activity, making it a valuable tool for assessing terrestrial productivity. However, the practical application of satellite-derived SIF is hindered by spatial resolution limitations and data gaps. The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission presents an opportunity to overcome these challenges through its hyperspectral Ocean Color Instrument and globally distributed Land Vegetation Index (LANDVI) product suite. This study's primary contribution lies in demonstrating that PACE vegetation indices alone can reliably predict SIF from global retrievals obtained using data from the Copernicus Sentinel-5P Tropospheric Monitoring Instrument (TROPOMI), even in the absence of PACE BRDF and albedo products (and their associated corrections).

Using 8-day global composites from 2024, we establish a temporal and spatial correlation between PACE indices and TROPOMI SIF (TROPOSIF) across fifteen biome-stratified study regions, including forests, grasslands, agricultural systems, and xeric landscapes. Simple univariate linear models reveal that the Enhanced Vegetation Index (EVI) and the Chlorophyll Index Red Edge (CIRE) are the most reliable global predictors of SIF, accounting for 80% and 77% of the variance, respectively. Notably, the stability of the EVI-SIF and CIRE-SIF relationships across seasons further emphasizes the significant role of canopy structure and chlorophyll information captured by these two indices in explaining global SIF variability. Seasonal analyses indicate that while EVI and CIRE are most effective in most forests and agricultural systems, moisture-sensitive indices and pigment indices perform better during dry seasons and transitional periods in water-limited ecosystems. Spatial residual analyses suggest minimal global bias but systematic underestimation in boreal forests and selected tropical regions, which aligns with established effects of canopy architecture and fluorescence escape probability. Considering that EVI, in particular, is available even from moderate-resolution Earth resource satellite missions such as Sentinel-2 and Landsat, there is substantial potential for SIF downscaling to management and policy-relevant scales.

From Hyperspectral Indices to Global Fluorescence: PACE Vegetation Indices as Predictors of Terrestrial Photosynthesis

Caleb Sloan Harmon

GENERAL AUDIENCE ABSTRACT

New satellite technology from NASA's PACE mission helps us study Earth's plants from space. This project uses PACE data to measure how much plants glow faintly when they photosynthesize, a process called SIF, which shows how active and healthy plants are worldwide. By combining PACE data with information about different landscapes, like forests and grasslands, we created a way to predict this plant glow across the globe. Our findings show that certain measurements, like the greenness of plants, strongly match the glow, especially when looking at seasonal changes, plants behave differently in spring versus winter, for example. These results could help us monitor how plants respond to climate change, improve farming, and better understand how plants store carbon, which is vital for a healthy planet.

Table of Contents

List of Figures

List of Tables

1. Introduction

Terrestrial ecosystems are critical components of the global carbon cycle, with photosynthetic activity serving as a fundamental measure of ecosystem productivity and resilience. SIF is a faint optical signal emitted by chlorophyll molecules during photosynthesis, typically within the 650-850 nm spectral range, as a byproduct of absorbed light energy not used for photochemistry or dissipated as heat (Porcar-Castell et al., 2014). Because SIF is directly linked to photosynthetic electron transport and light use efficiency (LUE), it provides a mechanistic proxy for gross primary productivity (GPP), the total carbon fixed by vegetation (Li et al., 2022). In contrast, traditional vegetation indices (VIs) such as the normalized difference vegetation index (NDVI) and EVI infer photosynthetic activity indirectly through canopy greenness. EVI, which incorporates near-infrared, red, and blue reflectance bands, reduces soil and atmospheric effects to better capture canopy structure and greenness (Huete et al., 2002). However, because canopy greenness often remains relatively stable, especially in evergreen systems, VIs may not fully represent the physiological seasonality of photosynthesis. Empirical analyses have shown that EVI and related indices exhibit moderate correlations with GPP ($R^2 \approx 0.5\text{-}0.7$), though the strength and form of these relationships vary across vegetation types and environmental conditions (Sims et al., 2008; Liu et al., 2021).

SIF, by contrast, captures the dynamic efficiency of the photosynthetic apparatus and has been shown to track GPP seasonality across a range of temporal scales. In winter-dormant conifer forests, SIF correlates strongly with GPP ($R^2 = 0.62\text{-}0.92$) due to changes in photoprotective pigments and photosystem II efficiency rather than variations in absorbed photosynthetically active radiation (APAR) (Magney et al., 2019). Similarly, Wu et al. (2022) found that in soybean fields, SIF peaks 14-17 days earlier than GPP, a difference explained by canopy structural development and chlorophyll content. These findings emphasize that while SIF directly reflects the efficiency of light use in photosynthesis, its seasonal dynamics may precede carbon fixation due to biophysical canopy processes.

At larger scales, Tao et al. (2024) provided a high-resolution, multi-decadal SIF dataset demonstrating strong coherence between SIF and GPP across Chinese ecosystems, further validating the potential of satellite-based SIF as a reliable photosynthetic indicator. Satellite-based SIF retrievals have also demonstrated high correspondence with tower-measured GPP across biomes, with global correlations typically ranging from $R^2 \approx 0.6\text{-}0.9$ (Frankenberg et al., 2018; Li et al., 2022).

The estimation of SIF from space relies on detecting the weak fluorescence signal embedded within the reflected solar spectrum. Early satellite missions such as GOSAT and GOME-2 exploited Fraunhofer lines, narrow spectral regions of reduced solar irradiance, to isolate SIF signals (Joiner et al., 2011). Later missions, including the TROPOspheric Monitoring Instrument (TROPOMI) aboard Sentinel-5 Precursor, substantially improved retrieval accuracy and spatial resolution, providing near-daily global SIF coverage at ~3.5 × 7 km resolution at nadir (Köhler et al., 2018). These advances have enabled high-frequency mapping of photosynthetic activity and support applications in carbon-cycle modeling, drought monitoring, and crop productivity assessment.

However, while TROPOMI provides high-quality SIF observations, it does not distribute derived vegetation indices, and generating such indices from reflectance data requires additional processing. The PACE mission, launched in 2024, fills this gap with its publicly available LANDVI product, which includes ten precomputed vegetation indices: NDVI, EVI, CIRE, Normalized Difference Water Index (NDWI), Normalized Difference Infrared Index (NDII), Normalized Difference Snow Index (NDSI), Chlorophyll-Carotenoid Index (CCI), Carotenoid Content Index (CAR), Photochemical Reflectance Index (PRI), and Modified Anthocyanin Reflectance Index (MARI) (Werdell et al., 2024; Caplan & Huemmrich., 2024).

The ten vegetation indices provided from the PACE satellite target distinct structural, biochemical, and physiological properties of vegetation. CIRE uses red-edge reflectance to estimate total canopy chlorophyll content and is particularly sensitive to variations in photosynthetic capacity. NDWI and NDII both exploit near- and shortwave-infrared reflectance to monitor vegetation water content, with NDWI emphasizing liquid water absorption near 860 nm and NDII targeting deeper canopy moisture. NDSI distinguishes snow from vegetation and soil using green and shortwave-infrared bands, aiding phenological studies in high-latitude and mountainous regions. CCI and CAR track the balance between chlorophyll and carotenoid pigments, providing insight into photosynthetic downregulation and stress responses. PRI, calculated from reflectance at 531 nm and a reference band (usually 570 nm), serves as a well-established proxy for xanthophyll-cycle activity and light-use efficiency at both leaf and canopy scales. Lastly, MARI detects anthocyanin accumulation in leaves, which often signals senescence, nutrient stress, or protective responses to high light or cold. Together, these hyperspectral indices extend traditional metrics by capturing pigment pool dynamics, water status, and rapid physiological adjustments that directly influence photosynthetic efficiency and carbon uptake (Caplan & Huemmrich., 2024).

Vegetation indices derived from hyperspectral reflectance, such as those in PACE’s LANDVI dataset, offer complementary insight into GPP by capturing both structural and physiological properties of vegetation. The red edge-based CIRE, as stated, is sensitive to canopy chlorophyll concentration and has been shown to align closely with photosynthetic capacity and SIF in croplands and forests (Wu et al., 2024). Similarly, machine-learning approaches integrating EVI, CIRE, and PRI have demonstrated the ability to reconstruct SIF or GPP signals with relatively high accuracy (R^2 up to ~0.8) in certain ecosystems (Zhang et al., 2023; Wu et al., 2024). Several of these indices, particularly CCI and PRI, provide physiologically relevant measures of photosynthetic efficiency by tracking pigment dynamics associated with the xanthophyll cycle and chlorophyll-carotenoid balance (Gamon et al., 2016). The inclusion of such hyperspectral indices in PACE’s LANDVI dataset enables monitoring of both canopy structure and photoprotective responses in an all-in-one product, offering more convenient insights into vegetation function.

PACE’s Ocean Color Instrument collects hyperspectral measurements from 340-890 nm, providing global 8-day averages at a consistent 0.1° (~11 km) resolution. This temporal aggregation minimizes cloud-related data gaps, enabling consistent global coverage suitable for monitoring vegetation phenology (Werdell et al., 2024). The enhanced spectral fidelity of PACE improves red-edge and xanthophyll-based indices such as CIRE, and PRI, while EVI benefits from more precise atmospheric correction. Although PACE does not directly measure SIF, its hyperspectral LANDVI dataset provides the necessary inputs to model SIF and bridge spatial or temporal gaps in existing SIF products, offering a new pathway for GPP estimation and vegetation monitoring.

Despite the availability of high-quality satellite SIF products such as TROPOMI’s ~3.5 × 7 km retrievals, modeling SIF from hyperspectral vegetation indices remains essential for several practical and scientific reasons. First, harmonizing SIF and reflectance-based datasets across missions requires spatial resampling, which can introduce artifacts and increase processing complexity; deriving SIF directly from the vegetation indices yields a native 0.1° dataset that aligns exactly with PACE's global 8-day averages without additional spatial transformations. Second, PACE also distributes higher-resolution Ocean Color Instrument reflectance data and LANDVI product at 1.2 km, offering the potential to apply the same SIF-VI model to generate finer-resolution SIF products than currently available from TROPOMI. Finally, reconstructing SIF from reflectance is valuable because SIF remains one of the most direct remote sensing proxies for GPP; producing consistent SIF estimates across sensors, spatial scales, and time periods improves comparability, supports data fusion, and extends the utility of SIF for predicting ecosystem productivity and stress responses.

2. Objective

This thesis aims to develop a predictive framework for estimating global SIF from the PACE LANDVI product using statistical modeling, building on approaches like those of Gentine and Alemohammad (2018), who reconstructed SIF from MODIS reflectance. This analysis also aims to improve our understanding of seasonal vegetation dynamics and their sensitivity to environmental variability, with broader applications in ecosystem health assessment, agricultural productivity, and global carbon-cycle research.

Our approach is deliberately simple: we test whether new PACE indices alone, without any sun-angle or APAR correction, can predict SIF well across biomes and seasons. The novelty would show that structural and pigment information embedded in these new indices largely captures the variation in SIF, making complex corrections unnecessary for many applications.

If successful, this dramatically simplifies global monitoring of photosynthesis and carbon uptake of CO_2. Carbon-cycle models, drought early-warning systems, and agricultural monitoring programs could use near-weekly PACE data with minimal processing, giving us much better seasonal and interannual tracking of ecosystem health than we have today.

3. Materials and Methods

3.1. Data Acquisition and Temporal Resolution

Data for this study were derived from two primary sources: SIF measurements from the Tropospheric Monitoring Instrument (TROPOMI) sensor, processed by Rui Cheng, and the PACE mission's Level 3 mapped LANDVI product, which computes various vegetation indices using the equations in table 1. To ensure global coverage in each temporal step, 8-day average averages were selected for both datasets. Satellite SIF data from the TROPOMI instrument on board European Space Agency's Sentinel-5p satellite was gridded using L2B TROPOMI near-infrared SIF into a global product of 0.01-degree resolution and 8-day interval from January 1, 2024, to December 31, 2024. The gridding algorithm was established in a Julia package (https://github.com/cfranken/gridding), which has been used effectively in Cheng et al. (2022a, b) and Kunik et al. (2023). The analysis period spanned from the earliest available PACE 8-day average March 5-12, 2024, to December 18-25, 2024. The partial period from December 26-31, 2024, was excluded to maintain temporal alignment, as recalculating new 8-day averages would introduce possible inconsistencies between a new averaging method and the method imbedded in the LANDVI product.

Each vegetation index was calculated from PACE Ocean Color Instrument (OCI) surface reflectance values following the formulations in Table 1 from Caplan and Huemmrich (2024). Each ρ_i represents the surface reflectance measured by OCI at wavelength or band i. When a specific wavelength is required, single-band reflectance values are used. For indices requiring broad spectral regions (blue, green, red, and near-infrared), reflectance values are averaged across all OCI bands within the corresponding wavelength interval according to Equation 1:

$$\rho_{\text{aggregated}} = \frac{\sum_i^n \rho_{i,\text{OCI}}}{N} \qquad \text{(Eq. 1)}$$

where N is the number of OCI bands within the wavelength interval used for aggregation. As described by Caplan and Huemmrich (2024), "*N is the number of OCI bands in the wavelength interval from i to n. The wavelengths defining each interval are: NIR, 847–877 nm; red, 620–671 nm; green (1), 526–537 nm; green (2), 545–566 nm; and blue, 459–479 nm.*"

Table 1. Equations for each VI included in the LANDVI suite (Caplan & Huemmrich., 2024)

Index	Equation	Reference
NDVI	$(\rho\text{NIR} - \rho\text{red})/(\rho\text{NIR} + \rho\text{red})$	Huete et al. (2002)
EVI	$2.5\times(\rho\text{NIR}-\rho\text{red})/(\rho\text{NIR}+6.0\times\rho\text{red}-7.5\times\rho\text{blue}+1)$	Huete et al. (2002); Jiang et al. (2008)
NDWI	$(\rho\text{NIR} - \rho1250)/(\rho\text{NIR} + \rho1250)$	Gao (1996)
NDII	$(\rho\text{NIR} - \rho1618)/(\rho\text{NIR} + \rho1618)$	Hardisky, Klemas, and Smart (1983); Ji et al. (2011)
CCI	$(\rho\text{green}(1) - \rho\text{red})/(\rho\text{green}(1) + \rho\text{red})$	Gamon et al. (2016); Middleton et al. (2016)
NDSI	$(\rho\text{green}(2) - \rho1618)/(\rho\text{green}(2) + \rho1618)$	Hall, Riggs, and Salomonson (2001)
PRI	$(\rho530 - \rho570)/(\rho530 + \rho570)$	Grace et al. (2007); Coops et al. (2010)
CIRE	$(\rho800/\rho705) - 1$	A. A. Gitelson et al. (2005) & A. Gitelson and Solovchenko (2017)
Car	$[(1/\rho495) - (1/\rho705)] \times \rho800$	Gitleson et al. (2007)
mARI	$[(1/\rho550) - (1/\rho705)] \times \rho800$	Gitleson et al. (2007)

3.2. Reclassification of Biomes and Land Cover

The coarse spatial resolution of the datasets often encompassed multiple biome and land cover types within a single pixel, potentially confounding analyses of vegetation dynamics. To enhance homogeneity in study areas, biomes and land cover classes were reclassified into broader, generalized categories. Biome classifications were based on the World Wildlife Fund (WWF) Terrestrial Ecoregions of the World dataset (Olson et al., 2001). Land cover data originated from the Copernicus Global Land Service (2022 release), aligned with the United Nations Food and Agriculture Organization's (FAO) Land Cover Classification System (LCCS).

Biomes were aggregated into six categories: forested tropical (1), forested temperate (2), taiga (3), grasslands (4), Mediterranean (5), and xeric shrublands (6). The Mediterranean category was retained separately due to its heterogeneous land cover mosaic, distinct from other biomes. Tundra was excluded because minimal vegetation cover would not inform photosynthetic proxies like SIF. Mangroves were also omitted, as their fragmented and sparse distribution prevented homogeneous study areas at the spatial scale of the data (Table 2).

Table 2. Reclassification of WWF terrestrial ecoregions into the six biome categories used for site selection and analysis. Tundra and mangroves were excluded from the study.

New Biome Class	WWF Biome Class
Tropical Forests	Tropical and Subtropical Moist Broadleaf Forests Tropical and Subtropical Dry Broadleaf Forests Tropical and Subtropical Coniferous Forests
Temperate Forests	Temperate Broadleaf and Mixed Forests Temperate Coniferous Forests
Boreal Forests/Taiga	Boreal Forests/Taiga
Grasslands	Tropical and Subtropical Grasslands, Savannas, and Shrublands Temperate Grasslands, Savannas, and Shrublands Flooded Grasslands and Savannas Montane Grasslands and Shrublands
Mediterranean	Mediterranean Forests, Woodlands, and Scrub
Xeric	Deserts and Xeric Shrublands
Removed	Mangroves Tundra

Land cover classes were similarly aggregated into three categories: agriculture (1), forest (2), grassland (3). Non-vegetated classes (e.g., bare soil, water) were masked out to focus on vegetated areas. Wetlands were ultimately excluded from final analyses due to their limited spatial extent and inability to support homogeneous study regions (Table 3).

Table 3. Reclassification of Copernicus Global Land Cover (UN-LCCS) classes into the three functional categories used in this study. Non-vegetated classes were masked; wetlands/flooded classes were considered but not used in the final site selection.

New Cover Class	Copernicus LCCS
Agriculture	Cropland, rainfed Cropland, irrigated or post-flooding Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%) Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
Forested	Tree cover, broadleaved, evergreen, closed to open (>15%) Tree cover, broadleaved, deciduous, closed to open (>15%) Tree cover, needle leaved, evergreen, closed to open (>15%) Tree cover, needle leaved, deciduous, closed to open (>15%) Tree cover, mixed leaf type (broadleaved and needle leaved) Mosaic tree and shrub (>50%) / herbaceous cover (<50%) Tree cover, flooded, fresh or brackish water Tree cover, flooded, saline water
Herbaceous/Grasslands	Mosaic herbaceous cover (>50%) / tree and shrub (<50%) Shrubland Grassland Shrub or herbaceous cover, flooded, fresh/saline/brackish water

Non-Veg	Lichens and mosses Sparse vegetation (tree, shrub, herbaceous cover) (<15%) Urban areas Bare areas Water bodies Permanent snow and ice No Data

3.3. Study Area Selection

Figure 1. Fifteen ~1,100-km² study sites were strategically selected to span the global range of biome types, productivity levels, and phenological regimes while maintaining high within-site homogeneity (≥80% single land-cover class for most sites). Locations of the 15 study areas (red dots) used for all analyses. Numbered labels correspond to the legend (1–15). Three sites (Eunápolis Mixed Region, Madrid Mixed Region, Central France Forests) were deliberately chosen as heterogeneous mosaics.

To balance data volume with representativeness, study areas were defined as 100 x 100-pixel boxes (approximately 1,100 km² at 0.1° resolution). Given the challenges of random sampling in achieving homogeneity, particularly across mixed land covers areas were selected visually using GIS software, prioritizing alignment between reclassified biome and land cover maps. Three areas were intentionally designed to capture mixed cover types (e.g., agriculture, forest, grasslands mosaics), while the remainder targeted near-homogeneous compositions. An R script was employed using the sf 1.0-19 package and the raster 3.6-32 package to quantify the proportional composition of each land cover class within candidate areas. Given coarse spatial resolution, homogeneous sites were prioritized; xeric sites were included to counterbalance productivity bias and widen the range of data available for modelling to avoid a bias towards highly vegetated areas. Homogeneous sites required ≥80% dominance by a single class, while mixed sites targeted 30% (±10%) for primary classes. Multiple rounds were run on each study area, and the study area was relocated or removed until all sites met the standard. This approach yielded diverse, biome-stratified regions suitable for global-scale SIF-EVI modeling (Figure 1).

Table 4. Study areas with biomes and cover classes.

Study Area	Biome	Cover
Northeast Appalachia	Temperate Forests	Forested
Northeast US Ag	Temperate Forests	Agriculture
Canadian Taiga	Boreal forests/Taiga	Forested
Southern Mexico Ag	Tropical Forests	Agriculture
Yucatan Forests	Tropical Forests	Forested
Peru Grasslands	Grasslands	Grasslands
Chapada Diamantina	Xeric	Agriculture
Eunápolis Mixed Region	Tropical Forests	Mixed
Brazilian Highlands	Grasslands	Agriculture
Central France Forests	Temperate Forests	Mixed
Iberian Grasslands	Mediterranean	Grasslands
Madrid Mixed Region	Mediterranean	Mixed
Congo Forests	Xeric	Forests
Namibia Scrublands	Xeric	Grasslands
Kazakh Steppe	Grasslands	Mixed

Certain land cover/biome combinations could not be created due to lack of sufficient homogeneity of land cover within the biome. For example, the boreal forest biome was comprised almost entirely of either forests or non-vegetated areas due to its lack of suitability for agriculture and other forms of low-lying vegetation. Agricultural and forested areas were too interspersed with non-vegetated areas in the Mediterranean area due to large portions of this area being urbanized.

Northeast Appalachia (Temperate Forests, Forested)

Located in the Allegheny Mountains, spanning parts of Monongahela and George Washington National Forests, this study area features dense temperate forests with a rich mix of deciduous and evergreen species (Lovett et al., 2006). Dominant deciduous trees include sugar maple (*Acer saccharum*), red oak (*Quercus rubra*), and American beech (*Fagus grandifolia*), which display vibrant fall colors as leaves shed in autumn (Braun, 1950). Evergreen species such as eastern hemlock (*Tsuga canadensis*) and white pine (*Pinus strobus*) provide year-round canopy cover. The forest understory is lush with ferns, wildflowers, and shrubs like mountain laurel (*Kalmia latifolia*), which thrive in the moist, well-drained soils typical of the region (USFS, n.d.).

Northeast US Agriculture (Temperate Forests, Agriculture)

Centered on Lake Loraine State Park in Ohio, this study area lies within a temperate forest biome but is dominated by agricultural fields. Corn (*Zea mays*) and soybeans (*Glycine max*) are the primary crops (USDA NASS, 2023). Remnant patches of temperate deciduous forests, featuring species like white oak (*Quercus alba*), hickory (*Carya* spp.), and black walnut (*Juglans nigra*), persist along field edges and in protected areas (ODNR, 2018). The region experiences four distinct seasons, with agricultural cycles aligned to spring planting and fall harvest.

Canadian Taiga (Boreal Forests/Taiga, Forested)

Situated just south of Fort McMurray in Alberta's Dillon River Wildland, this study area is a quintessential boreal forest dominated by evergreen conifers. Black spruce (*Picea mariana*), white spruce (*Picea glauca*), and tamarack (*Larix laricina*) form dense stands typical of Alberta's boreal region (Natural Regions Committee, 2014). Understory vegetation includes mosses, lichens, and shrubs such as Labrador tea (*Rhododendron groenlandicum*) adapted to acidic, nutrient-poor soils (Bonan & Shugart, 1989). Seasonality is extreme, with long, harsh winters and short, cool summers permitting a brief growing season.

Southern Mexico Agriculture (Tropical Forests, Agriculture)

Located in southern Veracruz, this study area lies within a tropical forest biome but is dominated by agricultural production. Corn (*Zea mays*) and beans (*Phaseolus* spp.) are widely cultivated in fields cleared from the original vegetation (Toledo, 1990). Remnant tropical forest patches include species such as ceiba (*Ceiba pentandra*), mahogany (*Swietenia macrophylla*), and strangler figs (*Ficus* spp.), supported by the region's warm, humid climate (Pennington & Sarukhán, 2005).

Yucatán Forests (Tropical Forests, Forested)

Centered on the Yucatán Peninsula, this study area is characterized by semi-deciduous and evergreen tropical forests (Flores & Espejel, 1994). Dominant species include sapodilla (*Manilkara zapota*), mahogany (*Swietenia macrophylla*), and gumbo-limbo (*Bursera simaruba*). The region experiences a tropical climate with a pronounced wet season and a dry season during which many trees shed leaves (Duch, 1988).

Peru Grasslands (Grasslands, Grasslands)

Located in the rain shadow of the Andes northwest of Lake Titicaca, this region contains high-altitude tropical grasslands known as Puna. Vegetation consists of perennial grasses such as *Festuca* and *Stipa*, as well as cushion plants adapted to cold, arid conditions (Young, 1997). The area experiences strong seasonality, with a wet season supporting grass growth and a dry season limiting productivity (Rundel, 2011).

Chapada Diamantina (Xeric, Agriculture)

Situated in Bahia, Brazil, within Chapada Diamantina National Park, this region contains xeric vegetation dominated by *campo rupestre* communities, grass-shrub mosaics adapted to rocky, nutrient-poor soils (Silveira et al., 2016). Vegetation includes drought-tolerant grasses, sedges, and shrubs such as *Vellozia* spp., with local agriculture occurring in scattered patches (Rapini et al., 2008).

Eunápolis Mixed Region (Tropical Forests, Mixed)

Straddling the Rio Jequitinhonha in Bahia, this study area blends tropical forests, grasslands, and agriculture. Forests contain species such as Brazil nut (*Bertholletia excelsa*), *Tabebuia* spp., and palms (*Euterpe* spp.) characteristic of the Atlantic Forest biome (Ribeiro et al., 2009). Agricultural components include cacao and coffee, which are commonly cultivated in shaded agroforestry systems in the region (Sambuichi, 2006).

Brazilian Highlands (Grasslands, Agriculture)

Located between the Rio Grande and Rio Araguari near Uberlândia, this landscape is part of the cerrado biome, a fire-adapted savanna with grasses, shrubs, and scattered trees (Oliveira & Marquis, 2002). Dominant species include grasses such as *Andropogon* spp. and shrubs such as *Curatella americana*. Agricultural fields of soybeans and corn replace portions of the natural vegetation (Klink & Machado, 2005).

Central France Forests (Temperate Forests, Mixed)

In the northern Central Massif, this region features temperate forests dominated by European beech (*Fagus sylvatica*), sessile oak (*Quercus petraea*), and hornbeam (*Carpinus betulus*) (Ellenberg, 1988). Mosaic landscapes of forest, agriculture, and grassland are typical of the Continental biogeographic region (EEA, 2016).

Iberian Grasslands (Mediterranean, Grasslands)

Located in northern Castile and León, this area consists of Mediterranean grasslands dominated by perennial grasses such as *Stipa tenacissima* and *Brachypodium* spp., along with low shrubs adapted to dry, rocky soils (Canteiro et al., 2018). The Mediterranean climate produces hot, dry summers and mild, wetter winters that support seasonal vegetation dynamics (Blondel et al., 2010).

Madrid Mixed Region (Mediterranean, Mixed)

Centered around Madrid, this region includes forests, agriculture, and grasslands characteristic of the Mediterranean biome. Holm oak (*Quercus ilex*) and cork oak (*Quercus suber*) dominate wooded areas (Pulido & Díaz, 2005). Agricultural elements such as wheat and olive groves reflect land-use practices adapted to the region's hot, dry summers and cool, wet winters (Díaz et al., 1997).

Congo Forests (Xeric Forests)

Located between Musumba and Kahinda in the Kapanga locality, this area contains xeric woodlands dominated by drought-tolerant species such as *Brachystegia* and *Julbernardia*, typical of miombo ecosystems (Frost, 1996). Semi-deciduous species occur along rivers, where higher moisture allows denser vegetation (White, 1983).

Namibia Scrublands (Xeric, Grasslands)

Southeast of Windhoek, this landscape consists of xeric scrublands featuring grasses such as *Aristida* spp., succulents, and acacia shrubs (*Vachellia* spp.) adapted to arid conditions (Mendelsohn et al., 2002). Seasonal rainfall drives short wet season greening, followed by prolonged drought during the dry season (Burke, 2006).

Kazakh Steppe (Grasslands, Mixed)

Centered on Ust-Kamenogorsk in eastern Kazakhstan, this region includes grasslands dominated by *Stipa* and *Festuca* species, with scattered forests of Siberian larch (*Larix sibirica*) and birch (*Betula pendula*) (Wesche et al., 2016). The climate is strongly continental, with cold, snowy winters and warm summers that drive pronounced vegetation seasonality (Lavrenko & Karamysheva, 1993).

3.4. Multivariate Correlation Analysis

Pixel-level values for SIF and LANDVI rasters were extracted from the selected study areas for each 8-day timestep and compiled into time-series datasets. Initial exploratory analyses involved simple linear regression with 1:1 models between SIF and each vegetation index across the full temporal range. R^2 values were computed for each timestep and visualized as temporal line plots to assess seasonal variations in correlation strength. Average R^2 values were also calculated per index to identify the most robust predictors.

To assess multicollinearity among candidate vegetation indices and guide the selection of predictors for multivariate models, variance inflation factors (VIFs) were calculated for all possible combinations of indices. VIF for each predictor was computed using equation 2 where R^2 is obtained by regressing the predictor of interest against all other predictors.

$$VIF = \frac{1}{1 - R^2} \qquad \text{(Eq. 2)}$$

This determined which indices had high variance inflation and should not be used in a model together as well as models that could be used in conjunction to produce multivariate models.

3.5. Other Indices and SIF

To establish baseline global relationships between SIF and the vegetation indices, two simple linear regression models were fitted using the entire pooled dataset from all study areas and all available dates. Specifically, Equation 3 was obtained by fitting a linear model of the form SIF ~ EVI using the lm() function in R version 4.4.2, i.e., lm(SIF ~ EVI, data = global_dataset). Similarly, Equation 4 was derived by fitting lm(SIF ~ CIRE, data = global_dataset). Model coefficients, standard errors, 95% confidence intervals, R^2 values, and p-values reported in Equations 3 and 4 come directly from the summary() output of these lm objects. Spatiotemporal autocorrelation could slightly inflate model fit statistics, but the variance across biomes was the major source of variability. Residuals from both global models were subsequently mapped for the period 1-11 August 2024 to visualize spatial patterns in model performance.

To examine how the relationships between SIF and all the vegetation indices vary throughout the growing season, the data for each study area were stratified into three temporal periods. The phenological differences varied between each study site as addressed in section 3.3 and were used to inform the results. When creating the seasonal analysis, to ensure that differences among sites and seasons reflected true variation in the relationships between SIF and vegetation indices rather than differences in scale, we normalized all SIF and index values using the pooled distribution of values across all study areas. Specifically, each variable was transformed using min-max normalization calculated from the complete dataset (i.e., each value was rescaled between 0 and 1 based on the global minimum and maximum across all regions and time periods). Normalizing in this way allowed the analysis to place tropical forests, temperate systems, croplands, and arid shrublands on a common scale despite their vastly different absolute reflectance and SIF magnitudes. This step was critical for comparing values across sites, ensuring that differences in predictive strength reflected ecological or phenological variation rather than artifacts of disparate measurement ranges. It also prevented high-SIF or high-reflectance biomes (such as humid forests) from dominating the structure of the models simply because they operate on naturally larger numeric scales. Separate scatterplots with associated linear regression lines and R^2 values were then generated for each seasonal period, biome, and land-cover type (Figure 9). This seasonal partitioning allowed us to identify which vegetation index exhibited the strongest and most consistent correlation with SIF in a given biome or cover type at different stages of the year, revealing important temporal shifts in predictive performance that were masked when using annual or non-stratified data.

4. Results

4.1. Correlation Analysis

EVI and CIRE were the only indices that maintained strong linear relationships with SIF ($R^2 > 0.75$) through most of the year, confirming their potential superiority as global SIF proxies. Several other indices (NDVI, NDII, NDWI) periodically achieved moderate explanatory power ($R^2 > 0.55$), particularly during peak growing-season months, suggesting that biome- or season-specific models could occasionally benefit from alternative predictors.

All indices exhibited synchronized declines in performance in late March and again in late November - early December, likely reflecting rapid phenological transitions (leaf-out and senescence) that temporarily exclude reflectance-based indices from instantaneous photosynthetic uptake. NDSI performed poorly year-round (R^2 consistently < 0.25) because permanent snow or ice was absent from nearly all study sites during the growing season (Figure 2).

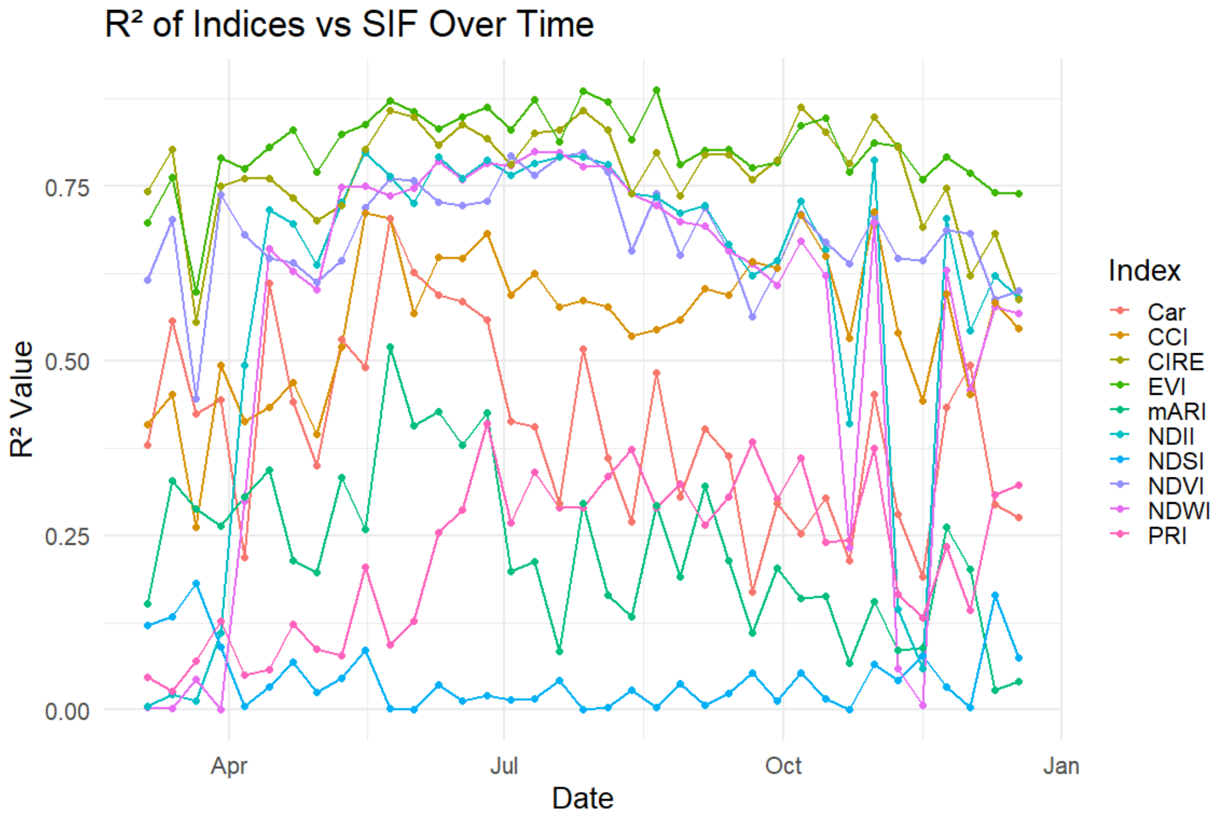


Figure 2. EVI and CIRE outperform all other vegetation indices in tracking SIF throughout the year on a global scale, with R² values exceeding 0.75 for most 8-day periods. Temporal dynamics of linear model performance (R²) for each LANDVI index when regressed against SIF across all study sites (2024-2025).

Averages were compiled from each index to definitively determine which indices were best. This helped solidify EVI and CIRE were the best indices to use in predictive modeling of SIF, more than just a visual assessment of the line graph (Table 5).

Table 5. EVI and CIRE are the clearest winners: only they achieve average $R^2 > 0.75$ with SIF. Mean R^2 values from linear regressions of each vegetation index against SIF across all 8-day periods and study sites in 2024-2025.

Index	**NDVI**	**EVI**	**NDWI**	**NDII**	**CCI**	**NDSI**	**PRI**	**CIRE**	**CAR**	**MARI**
Average R^2	0.68	0.80	0.56	0.60	0.56	0.04	0.23	0.77	0.41	0.23

Multicollinearity among the candidate vegetation indices was pronounced (Table 4). Pairwise variance inflation factors exceeded 5 and in several cases 8 for many structurally similar indices, most notably between EVI and NDVI (VIF = 9.41), NDII and NDWI (VIF = 9.59), EVI and CIRE (VIF = 8.20), NDVI and CIRE (VIF = 6.58), and CAR and MARI (VIF = 5.74). These high values indicate substantial redundancy in the physiological and structural information captured by these predictors. EVI and NDVI are both capturing structural information from the canopy while NDII and NDWI are both capturing moisture sensitive properties (Table 6).

Table 6. Strong multicollinearity prevents meaningful multivariate models: EVI and CIRE, despite being the two best individual predictors, cannot be combined because of high mutual collinearity (VIF > 6.5). Pairwise variance inflation factors (VIF) calculated among all candidate vegetation indices and SIF. Values > 5 (red shading) indicate severe collinearity; values < 2 (green shading) indicate negligible collinearity.

Index	*SIF*	*NDVI*	*EVI*	*NDWI*	*NDII*	*CCI*	*NDSI*	*PRI*	*CIRE*	*CAR*	*MARI*
SIF		3	5.11	1.63	2.05	2.24	1	1.26	4.25	1.61	1.27
NDVI	3		9.41	1.36	2.02	2.98	1.04	1.28	6.58	2.02	1.65
EVI	5.11	9.41		1.53	2.24	3.05	1.01	1.34	8.2	1.78	1.43
NDWI	1.63	1.36	1.53		9.59	2.04	1.78	1.51	1.6	1.13	1.02
NDII	2.05	2.02	2.24	9.59		3.37	1.36	1.7	2.41	1.28	1.08
CCI	2.24	2.98	3.05	2.04	3.37		1.08	3.59	2.96	1.22	1.04
NDSI	1	1.04	1.01	1.78	1.36	1.08		1.32	1.01	1.12	1.38
PRI	1.26	1.28	1.34	1.51	1.7	3.59	1.32		1.3	1	1.09
CIRE	4.25	6.58	8.2	1.6	2.41	2.96	1.01	1.3		2.07	1.56
CAR	1.61	2.02	1.78	1.13	1.28	1.22	1.12	1	2.07		5.74
MARI	1.27	1.65	1.43	1.02	1.08	1.04	1.38	1.09	1.56	5.74	

Although EVI and CIRE individually emerged as the strongest univariate predictors of SIF across the global dataset, their strong mutual collinearity rendered them unsuitable for inclusion in the same multivariate model. Attempts to construct multivariate models using two non-collinear index combinations yielded adverse changes in explanatory power relative to the best univariate models. All models with two or more dependent variables had lower R^2s than the best one-variable models. It is possible that having more than one dependent variable could improve season- or biome-specific models but checking each combination was beyond our study scope. For this reason, multivariate modeling was not pursued further, and subsequent analyses and predictions relied on the two best-performing univariate models: SIF ~ EVI and SIF ~ CIRE.

4.2. SIF Modeling

These scatter plots illustrates the relationship between SIF and EVI. The density scatter plot of SIF versus EVI reveals a strong positive linear relationship (Figure 3), with the fitted global univariate model yielding equation 3. The second shows the relation of the predicted and actual SIF values across various study areas for all time steps, with data points color coded by region (Figure 4). We attempted to add confidence intervals but confidence intervals of 95% around the fitted relationship were extremely narrow due to the large sample size (n ≈ 56,000) and are were not visually distinguishable from the regression line.

$$SIF = -0.535 + 3.722EVI \qquad \text{(Eq. 3)}$$

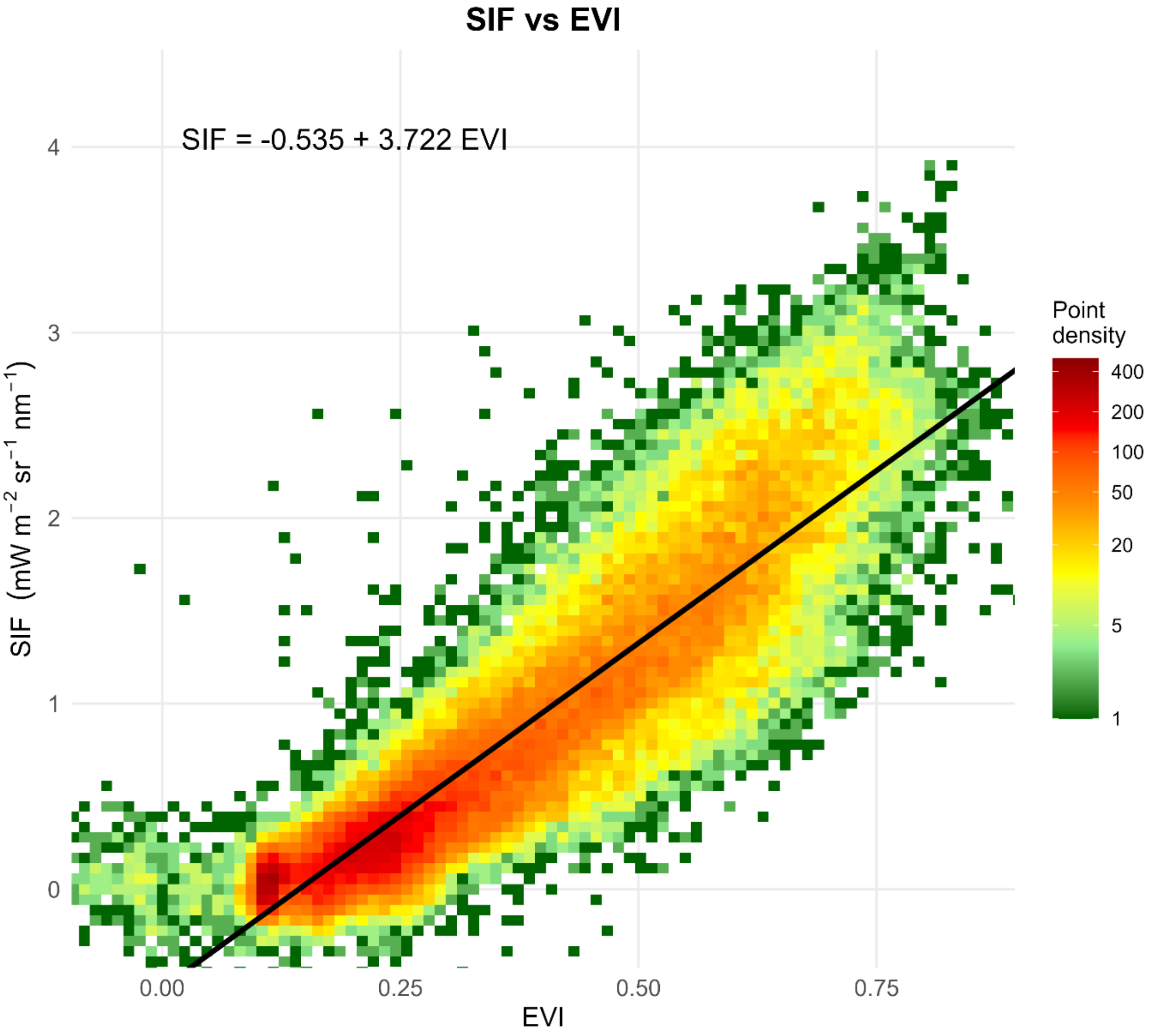


Figure 3. A density scatter plot shows the relationship between observed SIF and EVI across all study areas and time steps. Points are binned and colored by density (green: low; red: high) on a log scale. Black lines represent global univariate linear regressions with the equation annotated. Both indices explain substantial variance in SIF, supporting their utility as simple proxies for solar-induced fluorescence.

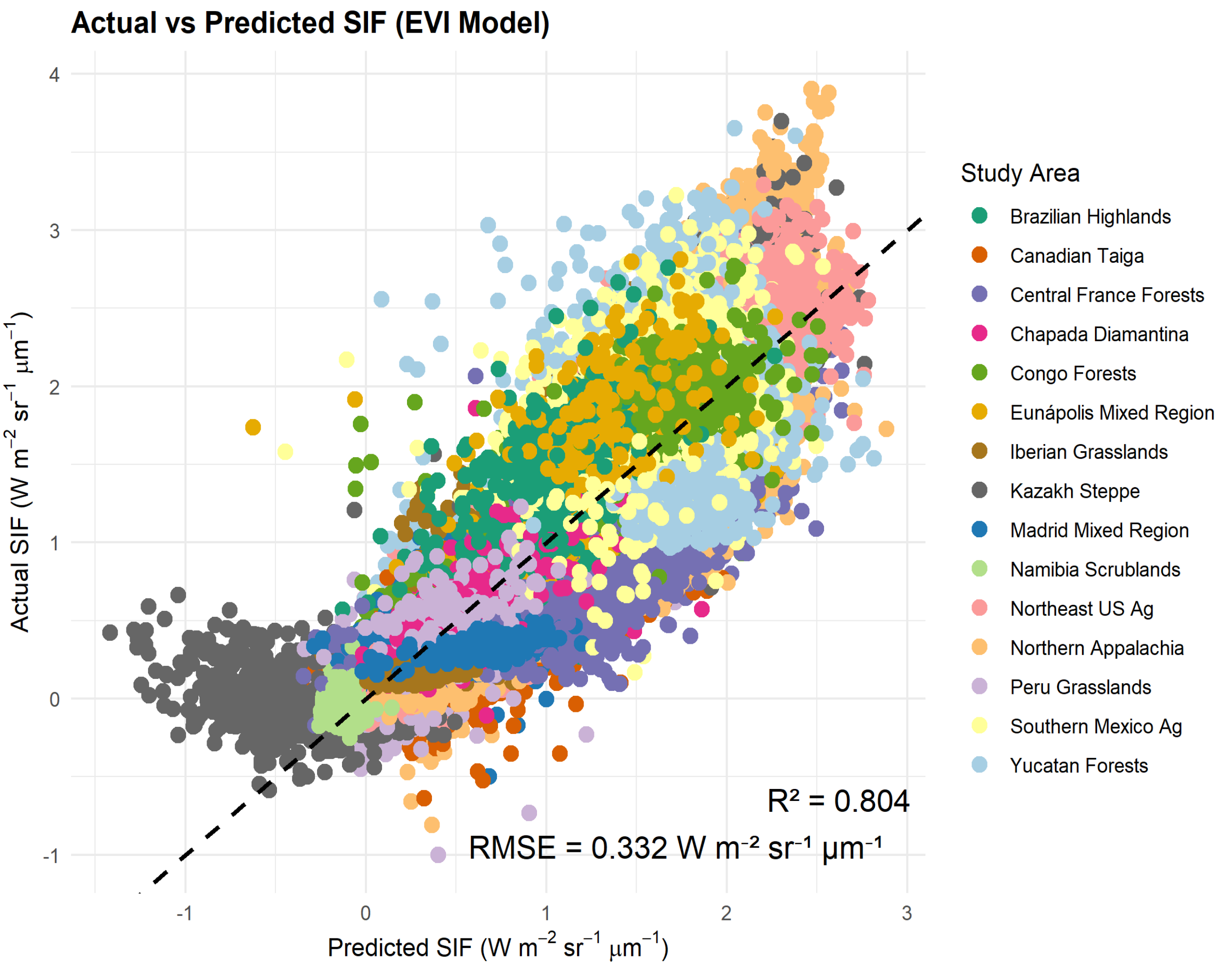


Figure 4. A simple global linear model using only EVI explains 80.4% of the variance in SIF across all biomes and seasons. Observed versus predicted SIF from the global univariate linear regression SIF ~ EVI (Equation 3). Each point represents an 8-day pixel-level observation from one of the 15 study areas, colored by site. The red dashed line is the 1:1 line; the black line is the fitted regression. $R^2 = 0.80$, *RMSE* $= 0.33\ \mathrm{W\ m^{-2}\ sr^{-1}\ \mu m^{-1}}$.

This indicated a positive correlation between the two variables. The model explains 80.4% of the variance ($R^2 = 0.804$) with a root mean square error (RMSE) of 0.33W $\mathrm{m^{-2}\ sr^{-1}\ \mu m^{-1}}$), suggesting a reasonably good fit despite some scatter in the data. The shape of this graph does suggest the possibility of a non-linear fit, but quadratic models were tried and only resulted in a < 0.02 improvement in the R^2 value (0.82) and a non-statistically significant difference in RMSE (0.329 W $\mathrm{m^{-2}\ sr^{-1}\ \mu m^{-1}}$).

Figure 5 shows the global spatial distribution of residuals (observed SIF - predicted SIF) from the EVI-based model for 04-11 August 2024. The model performs well over most tropical and temperate vegetated surfaces, with residuals near zero (pale yellow to light green) prevailing across croplands, grasslands, and seasonal forests. For this current time-step the model is underestimating in boreal areas with an exception in Greenland and the tropical forests of South America. Overestimation seems to be much less of an issue being mostly relegated to tropical islands, Greenland, and some parts of the Northeastern United States. This pattern was fairly consistent across all time steps. The distribution of residuals had a standard deviation being ~ 0.4 W m^{-2} sr^{-1} μm^{-1}. The other 36 time-steps had roughly similar means and standard deviations with some variation of 0 ± 0.05 for the mean and ±0.1 W m^{-2} sr^{-1} μm^{-1} for the standard deviation.

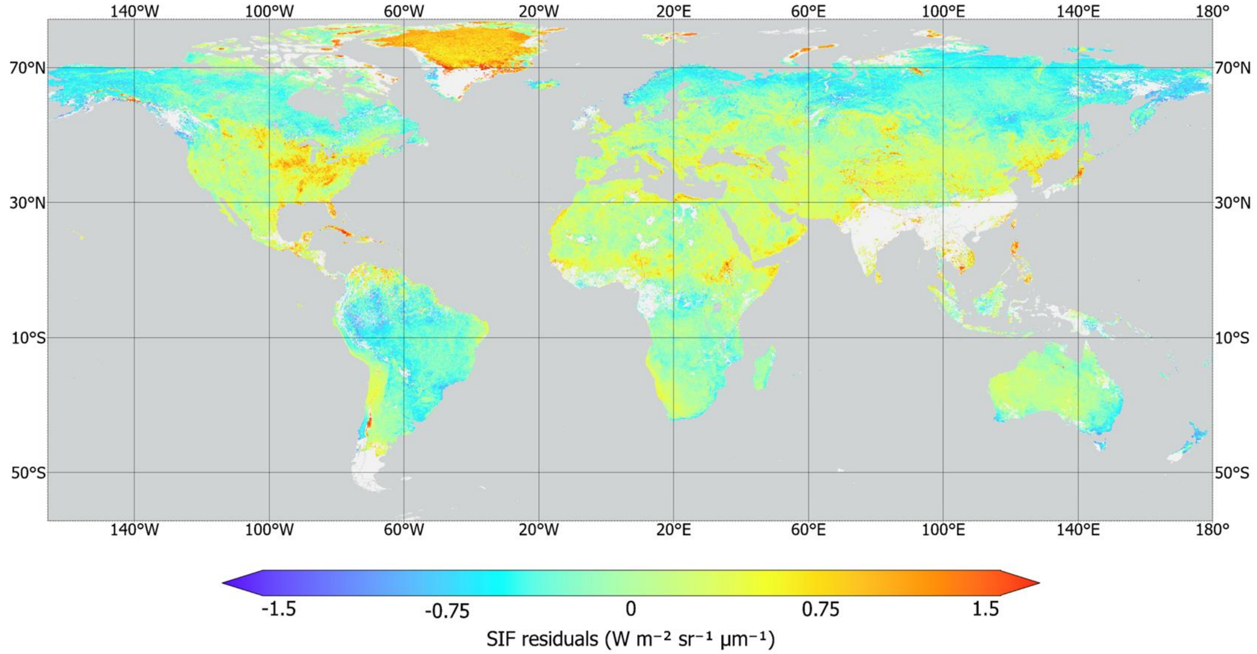


Figure 5. The global SIF ~ EVI model shows almost no bias in most biomes but consistently underestimates SIF in boreal forests and parts of the Amazon while slightly overestimating in a few high-latitude and island regions. Global map of residuals (observed - predicted SIF) from the univariate SIF ~ EVI model (Equation 3) for 4 - 11 August 2024. Color scale ranges from -1.5 (blue, strong underestimation) to +1.5 (red, strong overestimation).

These scatter plots illustrate the relationship between SIF and CIRE across various study areas for all time steps. The density scatter plot for SIF versus CIRE shows a clear positive correlation, described by the linear equation Eq 4. Dense point clusters align well with the regression line, particularly at low to mid CIRE values (Figure 6). For the second graph with data points color coded by region to show how the different study areas cluster depending on the physiology of the site. (Figure 7). A linear regression model in R version 4.4.2 yielded the equation represented in Eq. 4

$$SIF = -0.047 + 0.801CIRE \qquad \text{(Eq. 4)}$$

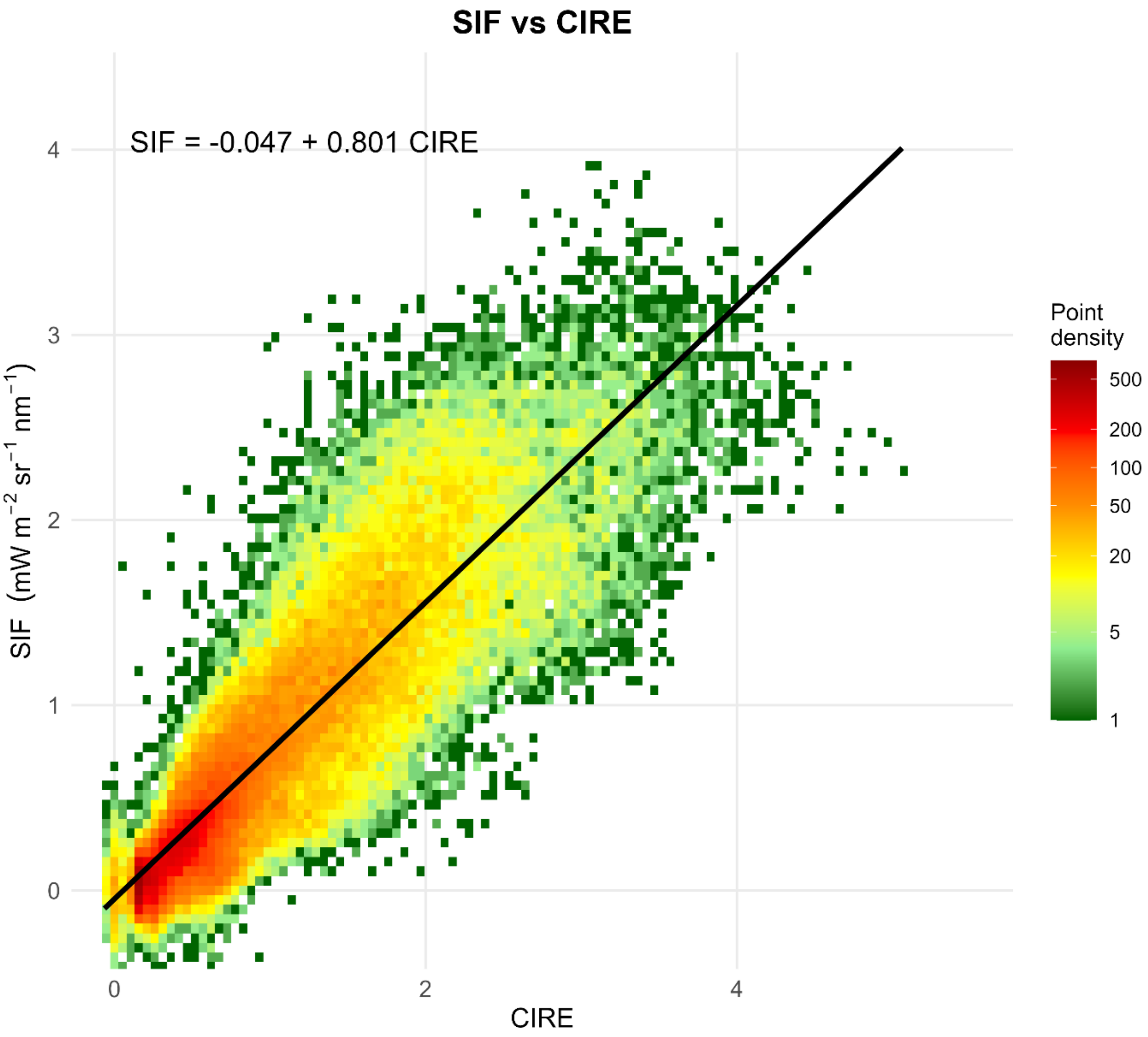


Figure 6. A density scatter plot shows the relationship between observed SIF and EVI across all study areas and time steps. Points are binned and colored by density (green: low; red: high) on a log scale. Black lines represent global univariate linear regressions with the equation annotated. Both indices explain substantial variance in SIF, supporting their utility as simple proxies for solar-induced fluorescence.

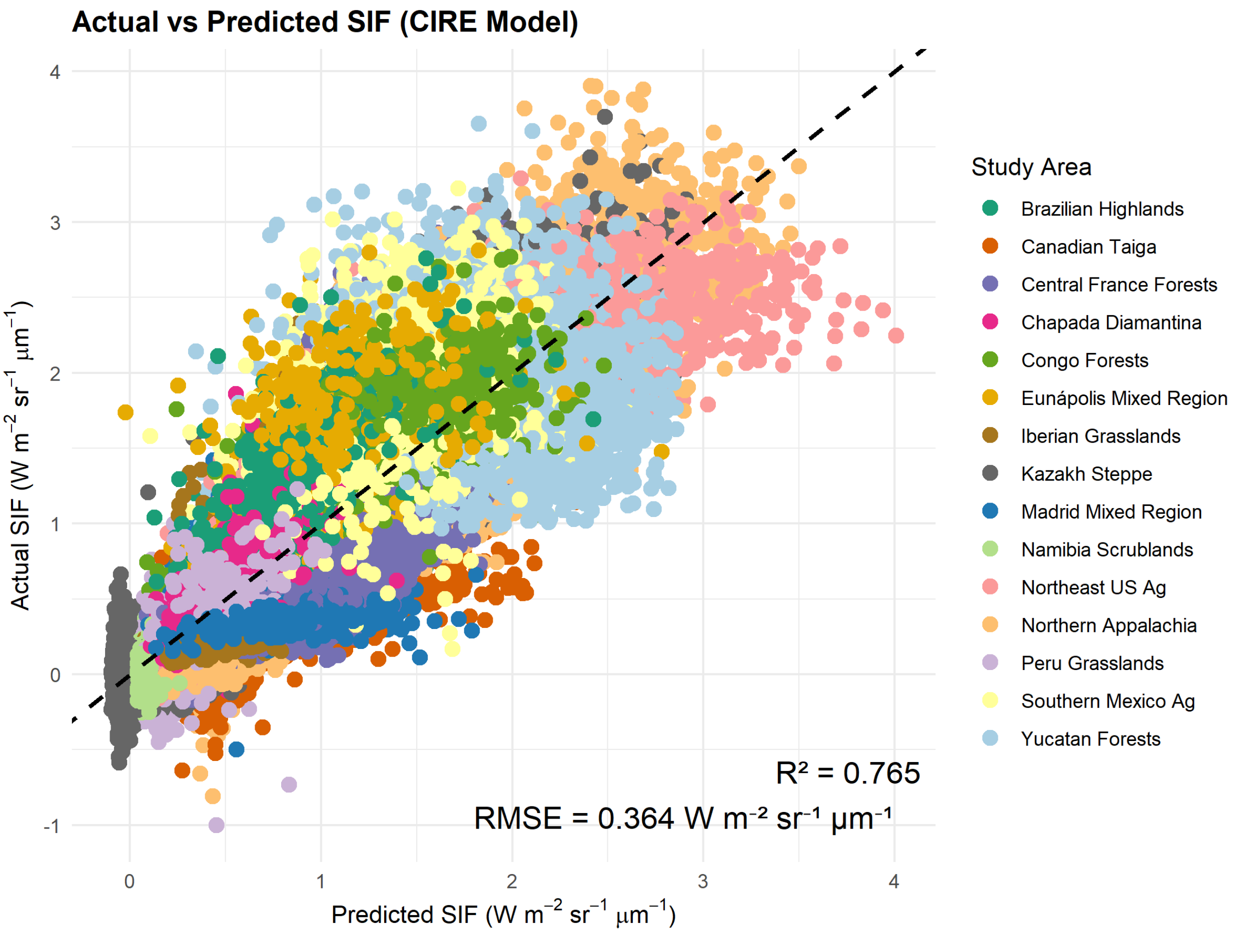


Figure 7. A simple global linear model using only CIRE explains 76.5% of the variance in SIF across all biomes and seasons. Observed versus predicted SIF from the global univariate linear regression SIF ~ CIRE (Equation 4). Each point represents an 8-day pixel-level observation from one of the 15 study areas, colored by site. The red dashed line is the 1:1 line; the black line is the fitted regression. $R^2 = 0.77$, RMSE = 0.36 W m^{-2} sr^{-1} μm^{-1}.

This indicated a positive correlation between the two variables. The model explains 76.5% of the variance ($R^2 = 0.765$) with a RMSE of 0.36 W m^{-2} sr^{-1} μm^{-1}, suggesting a slightly poorer fit than the model utilizing EVI. Once again there is some clustering of points toward the origin within the Kazakh Steppe. This could be due to it being a more northern location and having sparse vegetation during the winter months or an issue with satellite retrievals.

Figure 8 shows the global spatial distribution of residuals (observed SIF - predicted SIF) from the CIRE-based model for 04-11 August 2024. The model performs similarly to the EVI model over most tropical and temperate vegetated surfaces, with residuals near zero (pale yellow to light green) prevailing across croplands, grasslands, and seasonal forests. For this current time-step the model is once again underestimating in boreal areas, more heavily in Russia, and the tropical forests of South America. Overestimation seems to be more of an issue with the CIRE model around equatorial latitudes, Asiatic mountains, and some select spots in Ontario and Quebec. This pattern was fairly consistent across all time steps. The distribution of this data had a mean that showed a slightly positive bias, with a standard deviation of ~0.37 W m^{-2} sr^{-1} μm^{-1}.

The other 36 time-steps had roughly similar means and standard deviations with some variation of ± 0.1 for the mean and ±0.07 for the standard deviation.

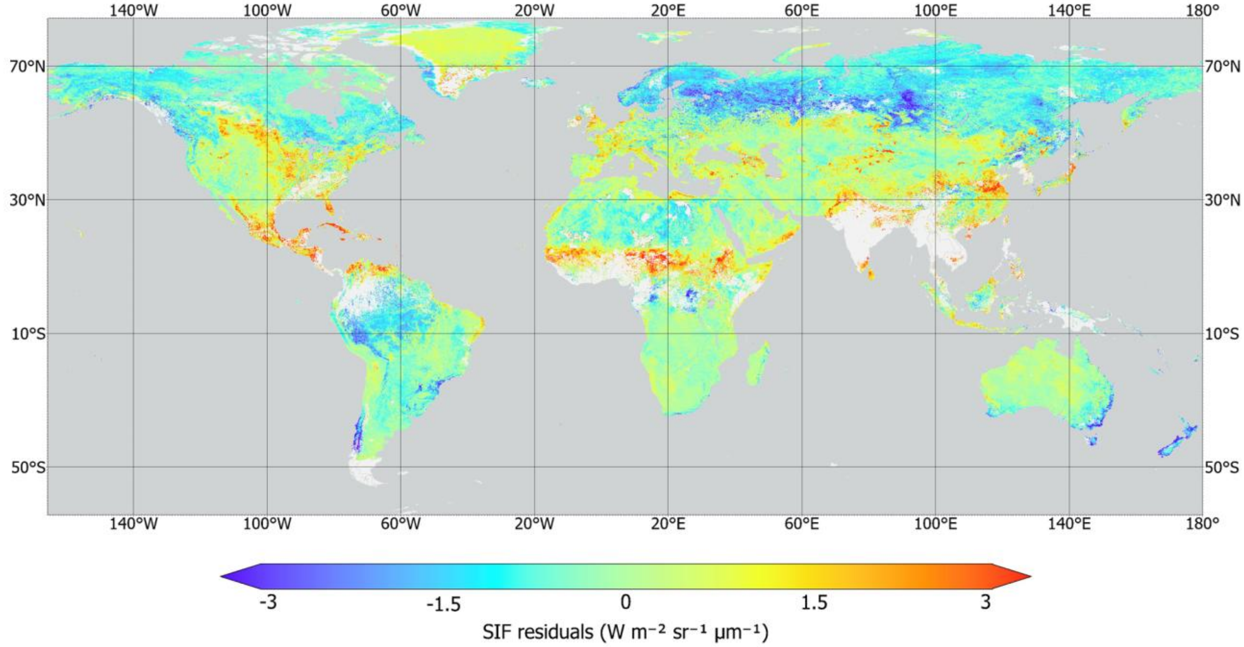


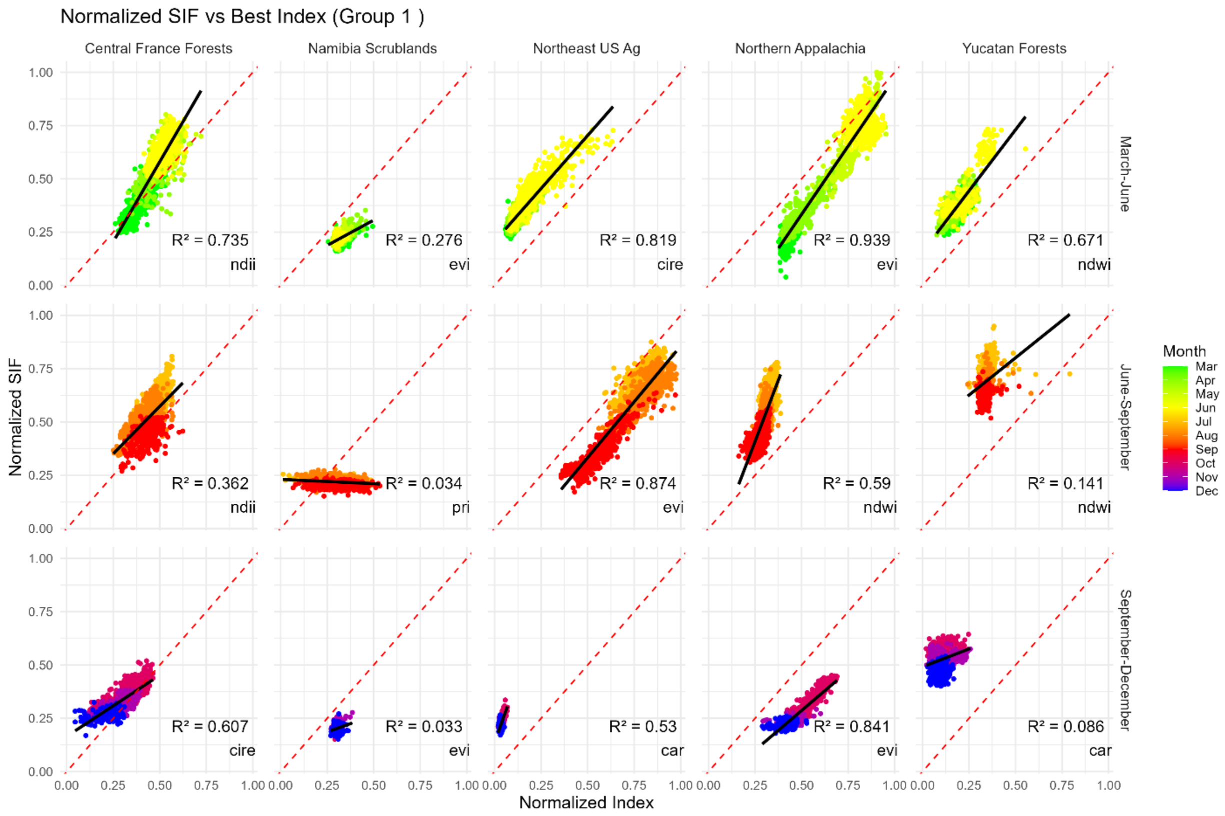
Figure 8. The global SIF ~ CIRE model is essentially unbiased over most of Earth's vegetated land surface but exhibits stronger and more widespread overestimation near the equator and in mountainous regions compared to the EVI model. Global map of residuals (observed - predicted SIF) from the univariate SIF ~ CIRE model (Equation 4) for 4 - 11 August 2024. Color scale ranges from -3.0 (blue, underestimation) to +3.1 (red, overestimation).

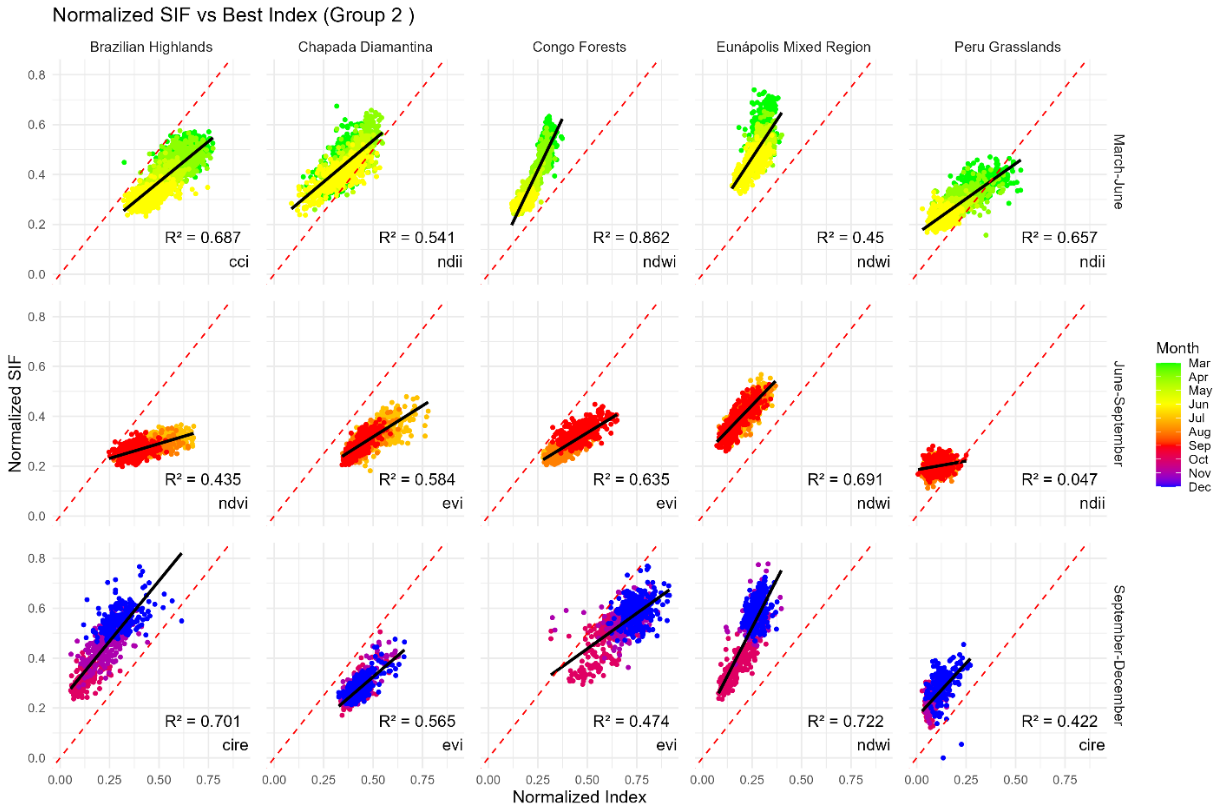
4.3. Seasonal Analysis

While EVI and CIRE remained strong predictors in many scenarios for the first grouping NDWI outperformed CIRE significantly in Yucatan Forest for the dry season (CIRE $R^2 = 0.40$) and marginally in the summer months of the Appalachians (CIRE $R^2 = 0.55$). All indices were poor predictors in the xeric Namibia Scrublands likely due to a lack of vegetation in the area. NDII also marginally outperformed EVI and CIRE in the spring and summer for the mixed cover in Central France (spring EVI $R^2 = 0.61$, summer CIRE $R^2 = 0.32$). The strongest correlation of any seasonal grouping can also be observed on this graph during the spring months in the Appalachians showing an R^2 of 0.94 for a normalized SIF~EVI.

EVI and CIRE remained strong predictors in many of the study areas for the second grouping, yet we still see NDWI outperforming in certain tropical areas. For all seasons in the Eunápolis Mixed Region NDWI was marginally higher than EVI, however much of the wet season (November-April) was excluded due to availability of data (EVI $R^2 = 0.26$, 0.65, and 0.63 respectively). NDWI also marginally outperformed in the xeric forests of the Congo after the end of the wet season in April into the early portions of the dry season of May through September (CIRE $R^2 = 0.80$). In the Brazilian Highlands CCI marginally outperformed other indices (EVI $R^2 = 0.64$) which is the only substantial instance of CCI being a better predictor than the rest. NDII also marginally outperformed other indices once again in the rain shadow of the Andes for the early portions of the dry season in the Peru Grasslands (EVI $R^2 = 0.58$) and the early portions of the dry season in the Chapada Diamantina (EVI $R^2 = 0.54$) where herbaceous plants and crops are beginning to enter dormancy.

It is in group 3 where there is little deviation from what we expect where EVI and CIRE remain the two best indicators of SIF. For many of the groupings we see R^2 values > 0.7 during the growing seasons while in the drier environments like Spain and Mexico we see a drop in the predictive power of all indices. Even though there is a drop in the predictive power for Southern Mexico and Madrid the variation in the data seems limited as many of the points remain in tight clusters. Across all 15 study sites and three phenological seasons, EVI or CIRE was the best univariate predictor of normalized SIF in 81% of cases (37 of 46 site–season combinations), routinely achieving $R^2 > 0.75$ and often > 0.90 during peak growing season (Figure 9).


Normalized SIF vs Best Index (Group 1)
Central France Forests
Namibia Scrublands
Northeast US Ag
Northern Appalachia
Yucatan Forests
March-June
June-September
September-December
Normalized SIF
Normalized Index
Month
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
R² = 0.735
ndii
R² = 0.276
evi
R² = 0.819
cire
R² = 0.939
evi
R² = 0.671
ndwi
R² = 0.362
ndii
R² = 0.034
pri
R² = 0.874
evi
R² = 0.59
ndwi
R² = 0.141
ndwi
R² = 0.607
cire
R² = 0.033
evi
R² = 0.53
car
R² = 0.841
evi
R² = 0.086
car
0.00
0.25
0.50
0.75
1.00


Normalized SIF vs Best Index (Group 2)
Brazilian Highlands
Chapada Diamantina
Congo Forests
Eunápolis Mixed Region
Peru Grasslands
March-June
June-September
September-December
Normalized SIF
Normalized Index
Month
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
R² = 0.687
cci
R² = 0.541
ndii
R² = 0.862
ndwi
R² = 0.45
ndwi
R² = 0.657
ndii
R² = 0.435
ndvi
R² = 0.584
evi
R² = 0.635
evi
R² = 0.691
ndwi
R² = 0.047
ndii
R² = 0.701
cire
R² = 0.565
evi
R² = 0.474
evi
R² = 0.722
ndwi
R² = 0.422
cire
0.0
0.2
0.4
0.6
0.8
0.00
0.25
0.50
0.75

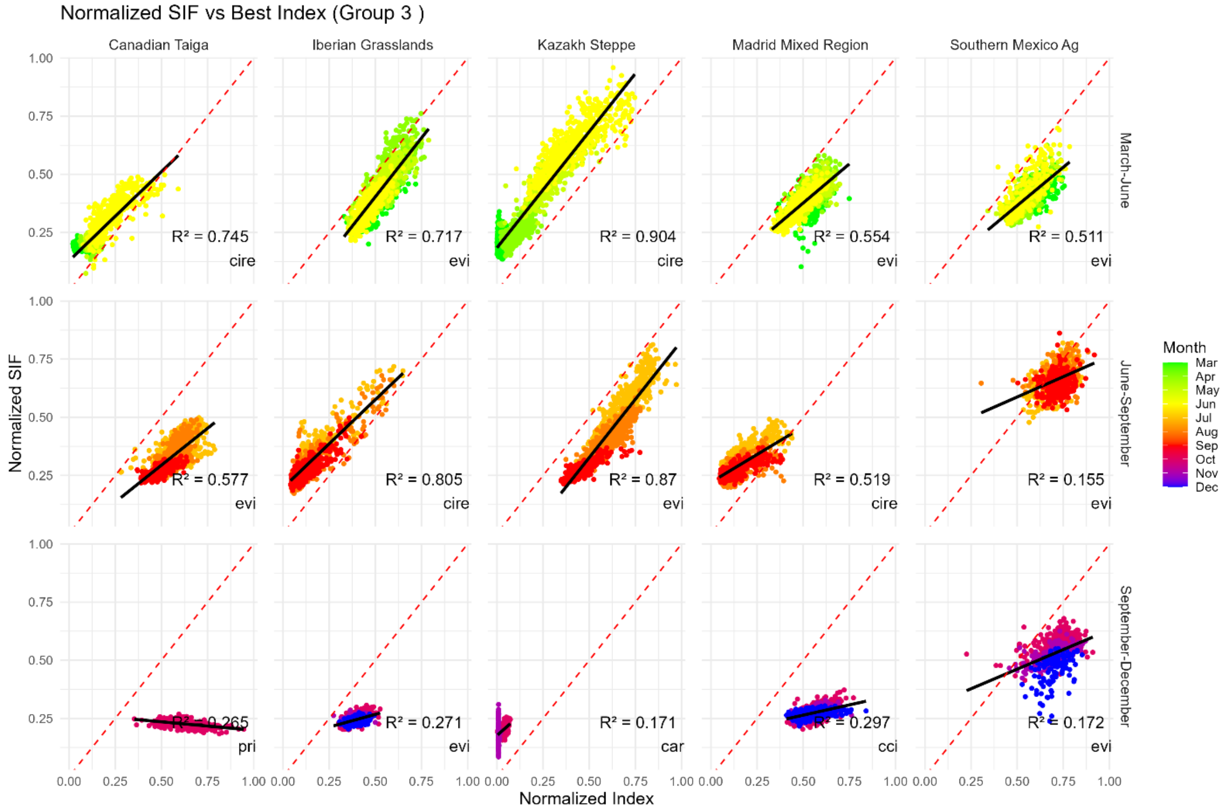
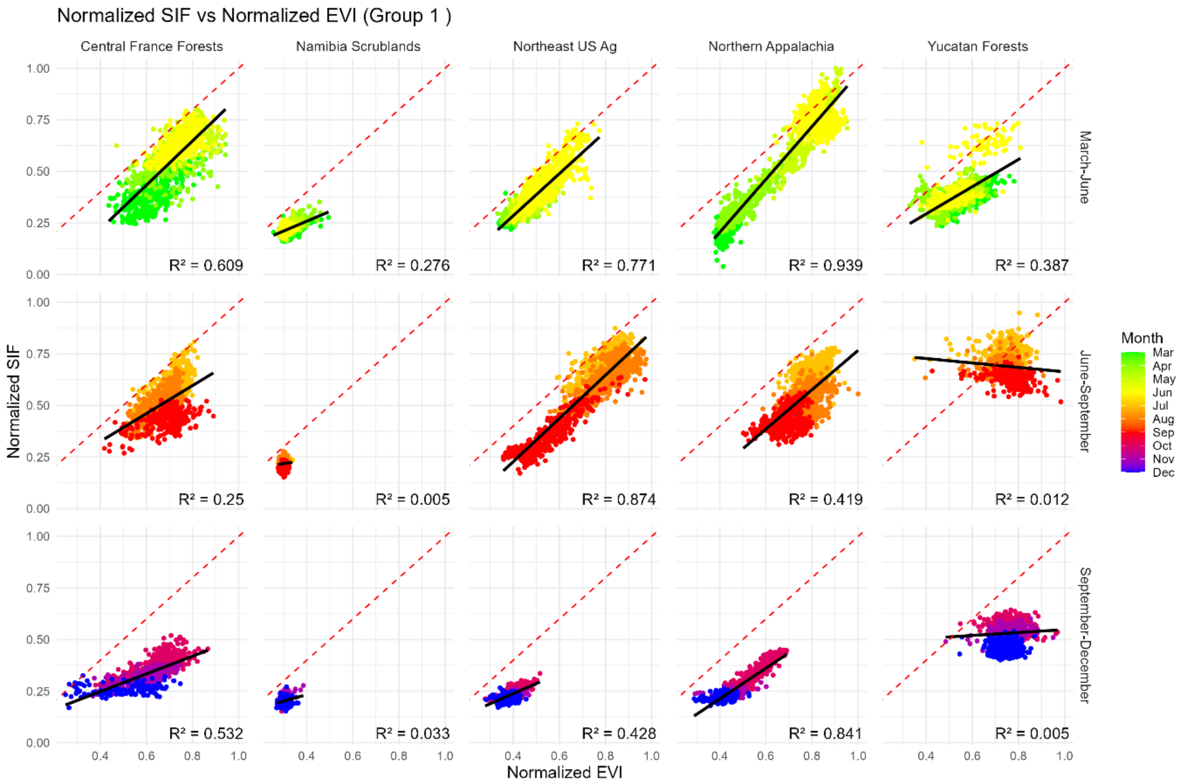
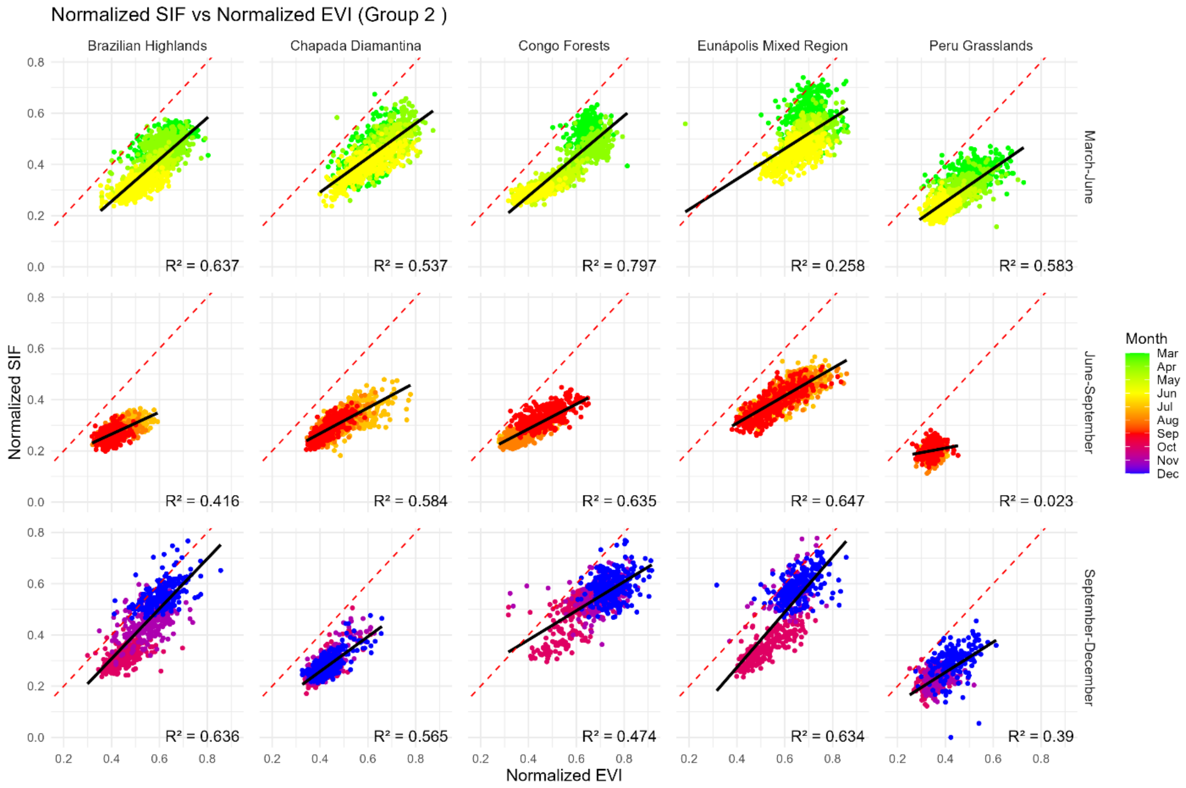


Figure 9. EVI and CIRE dominate as the best SIF predictors in almost all biomes and seasons, but water- and pigment-based indices (NDWI, NDII, CCI) occasionally outperform them during dry-season or early-senescence periods in moisture-limited ecosystems. Normalized SIF plotted against the single best-performing vegetation index for each site and phenological season (early-, peak-, and late-year). Points are colored by month; black line = linear fit; red dashed line = 1:1. R² and identity of the best index are shown in each panel.

Because EVI consistently emerged as the strongest global predictor of SIF, we also present a separate figure that focuses exclusively on the normalized SIF~EVI relationships across all study areas and seasons. While the full set of indices provides important insight into local deviations and biome-specific behavior, isolating EVI allows a clearer visualization of its dominant global performance. (Figure 10).


Normalized SIF vs Normalized EVI (Group 1)
Central France Forests
Namibia Scrublands
Northeast US Ag
Northern Appalachia
Yucatan Forests
March-June
June-September
September-December
Normalized SIF
Normalized EVI
Month
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
R² = 0.609
R² = 0.276
R² = 0.771
R² = 0.939
R² = 0.387
R² = 0.25
R² = 0.005
R² = 0.874
R² = 0.419
R² = 0.012
R² = 0.532
R² = 0.033
R² = 0.428
R² = 0.841
R² = 0.005
0.00
0.25
0.50
0.75
1.00
0.4
0.6
0.8
1.0


Normalized SIF vs Normalized EVI (Group 2)
Brazilian Highlands
Chapada Diamantina
Congo Forests
Eunápolis Mixed Region
Peru Grasslands
March-June
June-September
September-December
Normalized SIF
Normalized EVI
Month
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
R² = 0.637
R² = 0.537
R² = 0.797
R² = 0.258
R² = 0.583
R² = 0.416
R² = 0.584
R² = 0.635
R² = 0.647
R² = 0.023
R² = 0.636
R² = 0.565
R² = 0.474
R² = 0.634
R² = 0.39
0.0
0.2
0.4
0.6
0.8

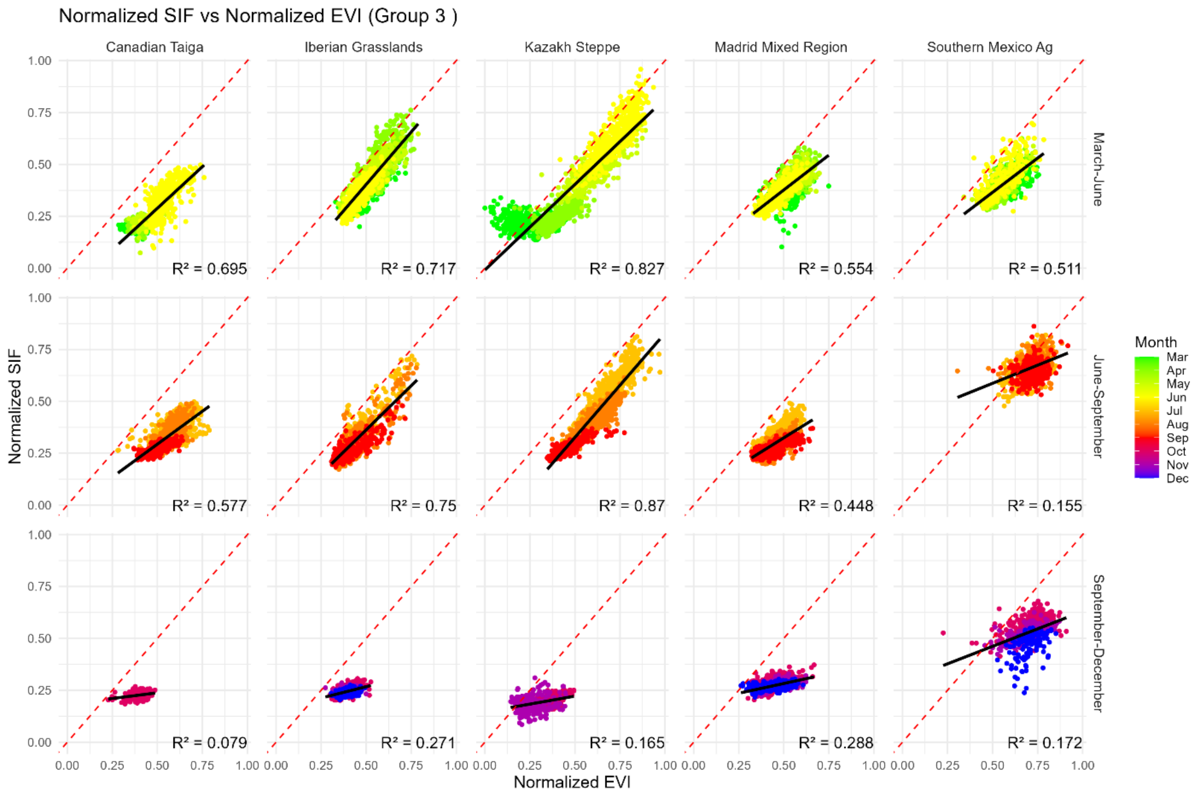


Figure 10. While some Vis improve predictions slightly from EVI, it still remains a strong predictor throughout most sites and times. Normalized SIF plotted against the Normalized EVI for each site and phenological season (early-, peak-, and late-year). Points are colored by month; black line = linear fit; red dashed line = 1:1.

5. Discussion

This study demonstrates the utility of vegetation indices for modeling SIF across diverse global biomes. The generally strong correlations observed between SIF and both EVI and CIRE align with previous work showing that greenness and red-edge based vegetation indices capture meaningful variation in photosynthetic activity (Huete et al., 2002; Gitelson et al., 2005) however the temporal analysis indicated there are some areas and times of year where that is not the case. The strength of the SIF~EVI relationship ($R^2 = 0.80$) is consistent with studies linking EVI to canopy structure and APAR, which together support stable predictive performance across ecosystems (Sims et al., 2008). Similarly, the performance of CIRE reflects its sensitivity to chlorophyll content and canopy photosynthetic capacity, echoing findings from crop and forest studies where red-edge indices have proven to correlate highly with SIF and GPP (Gitelson & Solovchenko, 2017; Wu et al., 2024).

The performance of EVI and CIRE can be directly linked to the radiative transfer processes that govern SIF emission and escape. SIF originates from the re-emission of absorbed photons by chlorophyll molecules, meaning that the absorption features surrounding the red edge (~700 nm) exert strong links with fluorescence level (Meroni et al., 2009). Because CIRE responds to subtle shifts in chlorophyll absorption and leaf biochemical status, it is highly sensitive to the same pigment-driven processes that modulate fluorescence yield. By contrast, EVI is primarily sensitive to canopy structure, leaf area index, and APAR, which together determine both the total pool of absorbed energy and the efficiency with which fluorescence can escape through the canopy. These layered radiative effects explain why both indices track SIF robustly at global scales. Each represents a distinct but complementary component of the fluorescence signal biochemical absorption at the leaf level and radiative transport at the canopy level. However because SIF is highly related to escape it may be the case that the areas in which the model is overpredicting or underpredicting, the inclusion of a variable that tracks escape probability may further improve predictions.

The use of 8-day average data from PACE and TROPOMI provided a robust temporal framework for capturing vegetation seasonality. This is particularly advantageous compared to the 16-day temporal resolution of MODIS vegetation indices, which can obscure short-term physiological changes (Huete et al., 2002). Moreover, the shared 8-day interval between PACE and the SIF dataset eliminated the need for custom temporal aggregation procedures, reducing potential artifacts associated with resampling (Tao et al., 2024). Two of the 8-day average averages could be averaged together to create a product with the same temporal resolution as MODIS. This temporal aggregation would also likely help reduce cloud-related data gaps since two different averages are being used. It is also important to note because fluorescence emission responds rapidly to environmental stressors, the improved temporal pairing between PACE and TROPOMI allows these short-term radiative and physiological changes to be captured more reliably without the averaging effects present in 16-day averages.

The reclassification of biomes and land cover into generalized categories allowed for controlled sampling within the constraints of the 0.1° resolution. Although this approach ensured homogeneity in most locations, it may also have introduced bias toward productive, vegetated regions, which is why the xeric biomes were included. The exclusion of tundra and mangroves removes biomes known to exhibit unique photosynthetic phenology. The tundra, for example, often displays short but intense seasonality detectable by SIF even when greenness indices underperform (Magney et al., 2019). Thus, future work may benefit from incorporating alternative biome classifications or higher-resolution sensors to include these regions more effectively. This may be difficult however as the orbit path of PACE can cause seasonal data gaps for extreme latitudes.

Study area selection likely influenced model performance. Xeric regions exhibited weaker correlations, consistent with low canopy cover and high soil background influence, which is known to degrade VI~photosynthesis relationships (Gao, 1996; Hardisky et al., 1983). In these open-canopy systems, the fluorescence escape fraction is high, but the total amount of absorbed energy is low, making structural indices less reliable predictors of SIF variation. This may help explain the comparatively weak EVI correlations in xeric regions. The reliance on visually selected regions introduces potential observer bias, although this was mitigated through the ≥80% homogeneity / ~30% heterogeneity thresholds and R-based compositional checks. Even so, automated sampling approaches, such as stratified random selection or object-based segmentation, could improve reproducibility and reduce bias in future studies.

A key insight from the seasonal stratification analysis is that NDWI often outperformed EVI and CIRE in seasonally dry regions. NDII also showed improved correlations in transitions to the dry season in several grassland and xeric sites, consistent with its established sensitivity to leaf and canopy water content (Hardisky et al., 1983). This aligns with prior work showing strong coupling between water status and photosynthetic efficiency under drought conditions (Li et al., 2022; Wu et al., 2024). NDWI's superior performance in tropical dry season forests and semi-arid landscapes suggests that moisture sensitive indices may provide critical complementary information where guard cell metabolism and stomatal conductance limit fluorescence emission (Flexas et al., 2002). It has been shown that water stress reduces stomatal conductance and CO_2 diffusion into the leaf, which downregulates photosynthetic electron transport and increases non-photochemical quenching, thereby suppressing chlorophyll fluorescence (Porcar-Castell et al., 2014). Biochemical limitations, including reduced Rubisco activity in C_3 plants or PEP carboxylase capacity in C_4 plants, further decrease electron transport demand, diminishing fluorescence yield under drought or thermal stress (Cornic, 1994; von Caemmerer, 2000; Sun et al., 2018). SIF represents the fraction of absorbed energy not used for photochemistry or dissipated through heat, any environmental factor that alters the balance between these pathways such as pigment content, water limitation, or thermal stress, directly modulates fluorescence yield. Vegetation indices that track these same processes therefore provide physically meaningful proxies for SIF variation. These findings demonstrate that a single global predictor may not fully capture the biome-specific drivers of SIF, reinforcing the need for tailored or multivariate models. The focused breakdown of EVI based on separate spatial and temporal distributions highlights how reliably EVI tracks SIF during periods of active vegetation growth and how its predictive strength remains comparatively robust even in regions where other indices may outperform. Presenting the EVI-only results separately also helps emphasize the broader conclusion of this analysis, despite regional nuances, EVI remains the most stable and widely applicable univariate proxy for SIF at the global scale

The spatial residual analysis further highlighted systematic underestimation in boreal forests and some tropical regions. Boreal underestimation has been documented in prior SIF~VI studies due to evergreen phenology, photoprotective pigment dynamics, and snow related reflectance effects that weaken VI sensitivity to physiological changes (Magney et al., 2019; Middleton et al., 2016). Conversely, overestimation in some tropical regions may relate to subcanopy shading or cloud contamination, both of which have been shown to influence SIF retrievals and reflectance indices differently (Joiner et al., 2011; Köhler et al., 2018).

Another important factor influencing these relationships is the fluorescence escape fraction, which varies strongly by canopy architecture, leaf angle distribution, and background reflectance. Dense forests, particularly evergreen canopies, exhibit low escape fractions due to strong multiple scattering and reabsorption of emitted fluorescence, which can dampen the apparent SIF signal even when photosynthetic activity remains high (Malenovský et al., 2019). Greenness indices such as EVI may saturate under these conditions, while red-edge indices partially compensate due to their sensitivity to optical depth. These radiative constraints likely contributed to the residual underestimation observed in boreal biomes, where fluorescence escape is known to be seasonally modulated by pigment composition and snow-affected reflectance.

Importantly, even with strong overall performance, global linear models are limited in capturing nonlinear or saturation effects, especially in dense tropical forests where red-edge indices may be more informative than greenness metrics alone (Gamon et al., 2016). Although quadratic models only marginally improved fit (<0.02 increase in R^2), nonparametric or machine-learning approaches such as those used by Gentine and Alemohammad (2018) have shown substantial gains in reconstructing SIF from multispectral data and may prove beneficial when PACE datasets extend beyond a single year.

This study underscores the potential of vegetation indices to contribute meaningfully to global SIF modeling. Taken together, these findings highlight that the predictive success of EVI and CIRE is not merely empirical but grounded in radiative and physiological mechanisms. Red-edge indices succeed because they track chlorophyll absorption, the biochemical gateway for both photochemistry and fluorescence, while greenness indices succeed because they capture canopy structure and light interception, the physical determinants of photon availability and radiative transfer. By linking these indices to the underlying processes that regulate fluorescence emission, this study identifies measurable proxies that stand in for the causal drivers of SIF at global scales. With its hyperspectral measurements, consistent temporal resolution, and global coverage, PACE provides a valuable platform for advancing our understanding of terrestrial photosynthesis. However, limitations such as land cover misclassification or change from the 2022 Copernicus dataset, cloud impacts, and partial year coverage in some regions highlight the need for continued refinement. Incorporating multiple years of PACE data, integrating higher resolution reflectance products, and exploring multivariate or machine-learning models will likely improve global SIF predictions even further and expand the usage cases for this dataset.

6. Conclusions

In conclusion, this study demonstrates the potential of using other vegetation indices to enhance global SIF predictions, leveraging EVI's correlation with photosynthetic activity. The methodology provides a replicable framework for integrating remote sensing data across biomes, with applications in environmental monitoring and climate adaptation strategies. While this study has a global focus, future studies that narrow down the exact land cover/biome combinations could be even more effective at accurately predicting SIF. Future research should focus on expanding temporal coverage to multiple years, incorporating advanced machine learning models, including escape probabilities and addressing data gaps to further refine global SIF estimation.

7. References

Blondel, J., Aronson, J., Bodiou, J. Y., & Boeuf, G. (2010). *The Mediterranean Region: Biological Diversity through Time and Space*. Oxford University Press.

Bonan, G. B., & Shugart, H. H. (1989). Environmental factors and ecological processes in boreal forests. *Ecological Monographs, 59*(2), 113–150.

Braun, E. L. (1950). *Deciduous Forests of Eastern North America*. Blakiston.

Burke, A. (2006). Savanna ecosystems in Namibia. *Journal of Arid Environments, 64*, 459–479.

Canteiro, C., et al. (2018). Dry grasslands of the Iberian Peninsula. *Journal of Vegetation Science, 29*, 270–284.

Caplan, S., & Huemmrich, K. F. (2025). Unveiling PACE OCI's hyperspectral terrestrial data products. *Remote Sensing Letters, 16*(4), 422–433. https://doi.org/10.1080/2150704X.2025.2470905

Coops, N. C., Hilker, T., Hall, F. G., Nichol, C. J., & Drolet, G. G. (2010). Estimation of light-use efficiency of terrestrial ecosystems from space: A synthesis. *Remote Sensing of Environment, 113*(12), 2537–2549. https://doi.org/10.1016/j.rse.2009.07.023

Cornic, G. (1994). Drought stress inhibits photosynthesis by decreasing the stomatal and mesophyll conductances to CO_2. *Photosynthetica, 30*(2), 1–11.

Cheng, R., et al. (2022a). Evaluating photosynthetic activity across Arctic-Boreal land cover types using solar-induced fluorescence. *Environmental Research Letters, 17*(11), 115009. https://doi.org/10.1088/1748-9326/ac9dae

Cheng, R., et al. (2022b). Impact of radiation variations on temporal upscaling of instantaneous solar-induced chlorophyll fluorescence. *Agricultural and Forest Meteorology, 327*, 109197. https://doi.org/10.1016/j.agrformet.2022.109197

Díaz, M., Campos, P., & Pulido, F. (1997). The Spanish dehesas: A diversity in land-use and wildlife. *Forest Ecology and Management, 94*, 155–166.

Duch, J. (1988). *Geografía de la Península de Yucatán*. UNAM.

Ellenberg, H. (1988). *Vegetation Ecology of Central Europe* (4th ed.). Cambridge University Press.

European Environment Agency. (2016). *Biogeographical Regions – Europe*. EEA Report.

Flexas, J., & Medrano, H. (2002). Drought-inhibition of photosynthesis in C_3 plants: Stomatal and non-stomatal limitations revisited. *Annals of Botany, 89*(2), 183–189. https://doi.org/10.1093/aob/mcf027

Flores, J. S., & Espejel, I. (1994). Vegetation of the Yucatán Peninsula. *Journal of Vegetation Science, 5*, 355–366.

Frankenberg, C., et al. (2018). The Chlorophyll Fluorescence Imaging Spectrometer (CFIS), mapping far-red fluorescence from aircraft. *Remote Sensing of Environment, 217*, 523–536. https://doi.org/10.1016/j.rse.2018.08.032

Frost, P. (1996). The ecology of miombo woodlands. In Campbell (Ed.), *The Miombo in Transition*. CFI.

Gamon, J. A., Huemmrich, K. F., Wong, C. Y. S., Ensminger, I., Garrity, S., Hollinger, D. Y., Noormets, A., & Peñuelas, J. (2016). A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. *Proceedings of the National Academy of Sciences, 113*(48), 13087–13092. https://doi.org/10.1073/pnas.1606162113

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gentine, P., & Alemohammad, S. (2018). Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophysical Research Letters, 45*(7), 3136–3146. https://doi.org/10.1002/2017gl076294

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2005). Relationships between leaf chlorophyll content and spectral reflectance with different sampling methods. *Journal of Plant Physiology, 161*(2), 165–173. https://doi.org/10.1078/0176-1617-01186

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2007). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology, 74*(1), 38–45. https://doi.org/10.1562/0031-8655(2001)074

Gitelson, A. A., & Solovchenko, A. (2017). Remote estimation of chlorophyll content in higher plant leaves. *Advances in Botanical Research, 81*, 229–262. https://doi.org/10.1016/bs.abr.2016.11.006

Grace, J., Nichol, C., Disney, M., Lewis, P., Quaife, T., & Bowyer, P. (2007). Can we measure terrestrial photosynthesis from space directly, using spectral reflectance and fluorescence? *Global Change Biology, 13*(7), 1484–1497. https://doi.org/10.1111/j.1365-2486.2007.01352.x

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (2001). Algorithm theoretical basis document (ATBD) for the MODIS snow and sea ice-mapping algorithms. *NASA Goddard Space Flight Center*.

Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora*. *Photogrammetric Engineering and Remote Sensing, 49*(1), 77–83.

Huete, A., et al. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Ji, L., Zhang, L., Wylie, B., & Rover, J. (2011). On the terminology of the spectral vegetation index (VI) family. *International Journal of Remote Sensing, 32*(21), 6901–6907. https://doi.org/10.1080/01431161.2010.510485

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment, 112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Joiner, J., et al. (2011). First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences, 8*(3), 637–651. https://doi.org/10.5194/bg-8-637-2011

Klink, C. A., & Machado, R. B. (2005). Conservation threats in the Brazilian Cerrado. *Conservation Biology, 19*(3), 707–713.

Köhler, P., et al. (2018). Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: First results and intersensor comparison to OCO-2. *Geophysical Research Letters, 45*(19), 10,456–10,463. https://doi.org/10.1029/2018gl079031

Köhler, P., et al. (2020). Global retrievals of solar-induced chlorophyll fluorescence at red wavelengths with TROPOMI. *Geophysical Research Letters, 47*(15). https://doi.org/10.1029/2020gl087541

Kunik, L., et al. (2023). Satellite-based solar-induced fluorescence tracks seasonal and elevational patterns of photosynthesis in California's Sierra Nevada Mountains. *Environmental Research Letters, 19*(1), 014008. https://doi.org/10.1088/1748-9326/ad07b4

Lavrenko, E. M., & Karamysheva, Z. V. (1993). *Steppes of the Former USSR*. Cambridge University Press.

Li, C., et al. (2022). SIF-based GPP as a useful index for assessing impacts of drought on vegetation: An example of a mega-drought in Yunnan Province, China. *Remote Sensing, 14*(6), 1509. https://doi.org/10.3390/rs14061509

Liu, F., et al. (2021). Can vegetation index track the interannual variation in gross primary production of temperate deciduous forests? *Ecological Processes, 10*(1). https://doi.org/10.1186/s13717-021-00324-2

Lovett, G. M., et al. (2006). Forest ecosystems of the northeastern United States. *BioScience, 56*(10), 873–886.

Magney, T. S., et al. (2019). Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *PNAS, 116*(24), 11640–11645. https://doi.org/10.1073/pnas.1900278116

Malenovský, Z., van Wittenberghe, S., Verrelst, J., Romero, J. M., Alonso, L., Miguel, E., ... & Moreno, J. (2019). A new hyperspectral SIF model: Model structure and validation using airborne data. *Remote Sensing of Environment, 231*, 111-114.

Mendelsohn, J., Jarvis, A., Roberts, C., & Robertson, T. (2002). *Atlas of Namibia*. David Philip Publishers.

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

Middleton, E. M., Huemmrich, K. F., Landis, D. R., Black, T. A., Barr, A. G., McCaughey, H., & Kljun, N. (2016). Photosynthetic efficiency of northern forest ecosystems from canopy-level chlorophyll fluorescence. *Remote Sensing of Environment, 187*, 334–351. https://doi.org/10.1016/j.rse.2016.10.014

Natural Regions Committee. (2014). *Natural Regions and Subregions of Alberta*. Government of Alberta.

ODNR (Ohio Department of Natural Resources). (2018). *Ohio's Forests 2016*.

Oliveira, P. S., & Marquis, R. J. (2002). *The Cerrado of Brazil: Ecology and Natural History of a Neotropical Savanna*. Columbia University Press.

Olson, D. M., et al. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience, 51*(11), 933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Pennington, T. D., & Sarukhán, J. (2005). *Árboles Tropicales de México*. UNAM/FCE.

Porcar-Castell, A., et al. (2014). Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *Journal of Experimental Botany, 65*(15), 4065–4095. https://doi.org/10.1093/jxb/eru191

Pulido, F. J., & Díaz, M. (2005). Regeneration of holm oak woodlands in Spain. *Agroforestry Systems, 67*, 181–194.

Rapini, A., et al. (2008). Biodiversity of the Chapada Diamantina. *Journal of Plant Ecology, 1*, 37–52.

Ribeiro, M. C., et al. (2009). The Brazilian Atlantic Forest: A shrinking biodiversity hotspot. *Biological Conservation, 142*, 1141–1153.

Rundel, P. W. (2011). High-elevation tropical ecosystems of the Andes. *Annals of the Missouri Botanical Garden, 98*, 159–170.

Sambuichi, R. H. R. (2006). Structure of cacao-agroforestry in Bahia. *Biodiversity & Conservation, 15*, 177–197.

Silveira, F. A. O., et al. (2016). Ecology and evolution of campo rupestre. *Plant and Soil, 403*, 15–36.

Sims, D., et al. (2008). A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment, 112*(4), 1633–1646. https://doi.org/10.1016/j.rse.2007.08.004

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., & Magney, T. (2018). Overview of solar-induced chlorophyll fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sensing of Environment, 209*, 808–823. https://doi.org/10.1016/j.rse.2018.02.016

Tao, S., et al. (2024). A high-resolution satellite-based solar-induced chlorophyll fluorescence dataset for China from 2000 to 2022. *Scientific Data, 11*(1). https://doi.org/10.1038/s41597-024-04101-6

Toledo, V. M. (1990). The ecological anthropology of traditional agriculture in Veracruz. *Human Ecology, 18*(3), 333–360.

USDA NASS. (2023). *Crop Production Summary*. U.S. Department of Agriculture.

USFS (U.S. Forest Service). (n.d.). *Ecological Units of the Eastern United States*.

von Caemmerer, S. (2000). *Biochemical models of C_4 photosynthesis.* CSIRO Publishing.

Werdell, P. J., et al. (2019). The Plankton, Aerosol, Cloud, Ocean Ecosystem mission: Status, science, advances. *Bulletin of the American Meteorological Society, 100*(9), 1775–1794. https://doi.org/10.1175/BAMS-D-18-0056.1

Wesche, K., Ambarlı, D., Kamp, J., Török, P., & Treiber, J. (2016). The Central Asian steppes. *Biodiversity and Conservation, 25*, 2187–2212.

White, F. (1983). *The Vegetation of Africa*. UNESCO.

Wu, G., et al. (2022). Difference in seasonal peak timing of soybean far-red SIF and GPP explained by canopy structure and chlorophyll content. *Remote Sensing of Environment, 279*, 113104. https://doi.org/10.1016/j.rse.2022.113104

Wu, Y., et al. (2024). Solar-induced chlorophyll fluorescence tracks canopy photosynthesis under dry conditions in a semi-arid grassland. *Agricultural and Forest Meteorology, 356*, 110174. https://doi.org/10.1016/j.agrformet.2024.110174

Young, K. R. (1997). Puna landscapes of the central Andes. In *The Andes: A Geographical Portrait*.

Zhang, Y., & Peñuelas, J. (2023). Combining solar-induced chlorophyll fluorescence and optical vegetation indices to better understand plant phenological responses to global change. *Journal of Remote Sensing, 3*. https://doi.org/10.34133/remotesensing.0085

Zhao, R., et al. (2023). Solar-induced chlorophyll fluorescence extraction based on heterogeneous light distribution for improving in-situ chlorophyll content estimation. *Computers and Electronics in Agriculture, 215*, 108405. https://doi.org/10.1016/j.compag.2023.108405