Enhancing Ecosystem and Earth System Science via Remote Sensing and Machine Learning

by

Sadegh Ranjbar

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biological Systems Engineering)

at the

University of Wisconsin-Madison

2025

Date of final oral examination: April 01, 2025

The dissertation is approved by the following members of the Final Oral Committee:

Paul C. Stoy, Professor, Biological Systems Engineering

Ankur R. Desai, Professor, Atmospheric and Oceanic Sciences

Mutlu Ozdogan, Associate Professor, Forest & Wildlife Ecology

Annemarie Schneider, Professor, Nelson Inst/Environmental Studies

Table of Contents

Dedication

To my beloved wife, **Shiva**, whose unwavering love, patience, and encouragement have been my constant source of strength.

To my wonderful parents, **Mohammad and Fatemeh**, whose sacrifices, wisdom, and boundless support have shaped my journey.

To my siblings, whose companionship and belief in me have been a source of inspiration.

This work is a testament to your love, guidance, and unwavering faith in me.

"ای که دانش به مردم آموزی آنچه گویی به خلق خود بنیوش
خویشتن را علاج می‌نکنی باری از عیب دیگران خاموش"

از ابومحمد مشرف‌الدین مصلح بن عبدالله شیرازی، متخلص به سعدی شیرازی

Translation from Persian to English:

"O you who impart knowledge to mankind, Heed well the wisdom you enjoin
You fail to remedy your own soul, Yet speak of others' flaws—better be silent"

Short verse from Abū-Muhammad Muslih al-Dīn bin Abdallāh Shīrāzī, Nicknamed Saadi Shirazi

Acknowledgments

I am deeply grateful to my advisor, Prof. Paul Stoy, for his constant support, patience, and guidance throughout my academic journey. From helping me shape my research ideas to guiding me through challenges, his advice has been invaluable. He always asked thoughtful questions, gave constructive feedback, and encouraged me to think independently, which has greatly influenced my growth as a researcher. Beyond academics, his kindness and encouragement have been a constant source of motivation, reminding me to stay curious, resilient, and passionate about discovery. I feel very lucky to have had him as a mentor and role model. I also sincerely appreciate my committee members, Prof. Ankur Desai, Prof. Annemarie Schneider, and Dr. Mutlu Ozdogan, for their insightful feedback, thoughtful discussions, and continuous support, which have greatly enriched my work.

Beyond my mentors, I am incredibly thankful for my labmates and colleagues—Daniele Losos, Sophie Hoffman, Dr. Bethany Blakeley, and Ojaswee Shrestha. Their collaboration, advice, and friendship have contributed to my research and also made this journey far more enjoyable and rewarding. From brainstorming ideas and troubleshooting challenges to sharing laughs and encouragement, their presence has truly made a difference.

Moreover, I am deeply grateful to the principal investigators and research teams of the eddy covariance towers for their dedication to data collection, as well as to the AmeriFlux Management Project (AMP) and the National Ecological Observatory Network (NEON) for their meticulous efforts in organizing, maintaining, and providing quality control for these valuable datasets. Special thanks to the GOES-R data providers for making satellite observations accessible for research.

This work was supported by the University of Wisconsin–Madison Office of the Vice Chancellor for Research and Graduate Education, with funding from the Wisconsin Alumni Research Foundation, the U.S. National Science Foundation (Macrosystems Biology award 2106012 and Hydrological Sciences award 2422397), NOAA award NA22OAR4310223, and the USDA Hatch program.

Abstract

Understanding and monitoring the Earth's water and carbon cycles are essential for addressing climate change, land-use changes, and environmental sustainability. This dissertation investigates leveraging geostationary satellite imagery and machine learning to improve the frequency mapping and applicability of geostationary remote sensing for developing an integrated near real time system for water and carbon cycle monitoring. A key innovation of this work is the use of high-frequency observations from geostationary satellites, which offer an unprecedented temporal resolution compared to traditional polar-orbiting sensors, enabling near real-time monitoring of critical environmental variables.

This research utilized advanced machine learning methodologies for deriving high-frequency downwelling shortwave radiation (DSR, Chapter 2) and land surface temperature (LST, Chapter 3) products using geostationary satellite data collected by Advanced Baseline Imager (ABI) sensor mounted on the Geostationary Operational Environmental Satellite-R (GOES-R). These products serve as essential inputs for modeling energy, water, and carbon fluxes, providing a more comprehensive understanding of ecosystem dynamics. By integrating these satellite-derived datasets as well as satellite observations with machine learning techniques, we advance the estimation of gross primary productivity (GPP, Chapter 5), ecosystem respiration (RECO, Chapter 5), and evapotranspiration (ET, Chapter 6), demonstrating that high-frequency satellite observations significantly enhance remote sensing-based monitoring capabilities. Additionally, this study reveals that incorporating shortwave infrared (SWIR) reflectance into GPP estimation models improves the accuracy of vegetation productivity assessments (Chapter 4), offering new insights into terrestrial carbon fluxes.

The findings highlight the transformative potential of geostationary satellite observations for developing near real-time, automated environmental monitoring systems. This dissertation shows how high-resolution satellite data and machine learning models can improve monitoring water and carbon cycles, contributing to scalable, data-driven solutions for ecosystem and Earth system science. The final chapter discusses key insights, limitations, and future directions for expanding these methodologies, showing the way for next-generation remote sensing applications in climate resilience and sustainability.

Keywords: Remote Sensing, Carbon Cycle, Water Cycle, Machine Learning, Gross Primary Productivity (GPP), Ecosystem Respiration (RECO), and Evapotranspiration (ET)

List of tables

List of figures

Chapter 1. Brief literature review of geostationary satellites for terrestrial ecosystem

1.1. Introduction

Remote sensing is central to our understanding of the global carbon and water cycles and its responses to anthropogenic pressures including global climate change (Schimel et al., 2015; Xiao et al., 2019). Many of these advancements have come from polar orbiting satellites like Landsat (Wulder et al., 2019) and MODIS (Heinsch et al., 2006; Running et al., 2004), which observe the land surface on times scales of days to multiple weeks and provide estimates of terrestrial ecosystem variables including (the gross primary productivity, GPP) and loss (ecosystem respiration, RECO), and ecosystem-atmosphere water flux (evapotranspiration, ET) on time scales of days to years. In a world where extreme events and rapid ecosystem changes are becoming ever more prevalent (Diffenbaugh et al., 2017), and where societal decisions are increasingly tied to carbon dynamics (Dilling, 2007; Grace, 2004), rapid and real-time inference about terrestrial ecosystem function may provide an important additional tool for understanding the Earth system. The new generation of imagers on geostationary ("weather") satellites (Table 1.1) around the world can estimate GPP, RECO, and ET on the order of minutes (Jeong et al., 2023a; A. Khan et al., 2022) and address important questions about the seasonality of vegetation function and its changes (Hashimoto et al., 2021; Miura et al., 2019; MIURA and NAGAI, 2019; Wheeler and Dietze, 2021, 2019). This begs the question: what are the advantages of such "hypertemporal" (Miura et al., 2019) remote sensing of the carbon and water cycles, and why would one bother to estimate surface-atmosphere fluxes in real time?

Table 1.1 List of Current Operational Geostationary Meteorological Satellites from Seven Major Countries, Along with Their Respective Agencies Involved in Operations and Launching

Country	Current Geostationary Meteorological Satellites	Agencies Involved in Operations and Launch
USA	GOES-14 / GOES-16 / GOES-17 / GOES-18 / GOES-19	National Oceanic and Atmospheric Administration (NOAA) National Aeronautics and Space Administration (NASA)

Japan	Himawari-8 / Himawari-9	Japan Meteorological Agency (JMA) Japan Aerospace Exploration Agency (JAXA)
Europe	Meteosat-9 / Meteosat-10 / Meteosat-11 / MTG-I1	European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) European Space Agency (ESA)
India	INSAT-3D / INSAT-3DR / INSAT-3DS	India Meteorological Department (IMD) Indian Space Research Organisation (ISRO)
China	FY-2G / FY-2H / FY-4A / FY-4B	China Meteorological Administration (CMA) National Remote Sensing Center of China (NRSCC)
Russia	Electro-L N2 / Electro-L N3 / Electro-L N4	Russian Federal Service for Hydrometeorology and Environmental Monitoring (RosHydroMet) Russian Federal Space Agency (Roscosmos)
South Korea	GEO KOMPSAT-2A / GEO KOMPSAT-2B	Korea Meteorological Administration (KMA) Korea Aerospace Research Institute (KARI)

This review explains why near-global, real time geostationary observations can be uniquely useful for understanding the contemporary carbon and water cycles in our era of global change. We first provide a brief background of geostationary satellites and their sensors. Then, we describe terrestrial ecosystem modeling and a number of ongoing and emerging applications that highlight the unique capabilities of geostationary surface-atmosphere flux estimates. Finally, we discuss the challenges, needs, and opportunities ahead, focusing on the technological and data-driven hurdles to improve the accuracy and scalability of geostationary flux estimates. We also highlight the potential for advancements in satellite instrumentation, data processing, and the use of machine learning and AI to enhance these estimates, enabling more effective applications in global change research and environmental decision-making.

Solutions to some of these challenges are underway with the Geostationary-NASA Earth Exchange (GeoNEX) initiative (Nemani et al., 2020; W. Wang et al., 2020) that will help synthesize global geostationary datasets (Vandal et al., 2021). Ultimately, one of the greatest uses of hypertemporal remote sensing of terrestrial ecosystems may be its fusion with high spatial resolution and hyperspectral data products (Shen et al., 2021) to help us understand carbon fluxes "everywhere and all the time" (Chu et al., 2017). With forthcoming geostationary missions like GeoXO in their planning stages with further spectral and spatial improvements in sight (Schmit et al., 2022), geostationary satellites will be with us for a long time. Finding ways for them to help us understand the terrestrial carbon cycle will only enhance their immeasurable benefits to society and provide new insights into the ever-changing Earth system (Khan et al., 2021a; Xiao et al., 2019).

1.2. Geostationary satellites and their sensors development for Earth observation

Geostationary satellites have a rich history, rooted in early 20th-century scientific theories. The concept was first proposed by Russian scientist Konstantin Tsiolkovsky in 1928 (Zee, 2013), but a critical milestone occurred in 1963 with NASA's successful launch of the first geostationary satellite, *Syncom 2* (Schroer, 2008; Zee, 2013). This event marked the practical beginning of utilizing geostationary orbits for Earth observation and telecommunications (Schroer, 2008; Straubel, 1992). Over the ensuing decades, geostationary satellites became integral to global communication (Furqan and Goswami, 2022), meteorology (Menzel, 2020; Rao et al., 1990), and navigation (Giorgi et al., 2019). Figure 1.1 illustrates the lifespan and data availability of geostationary satellites, which have primarily been designed for Earth observation purposes.

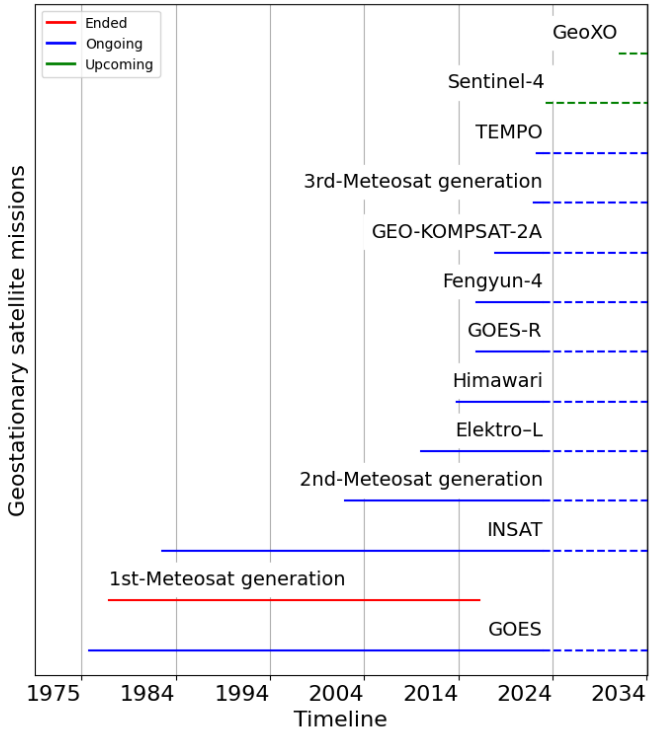

Figure 1.1: Lifespan and data availability of geostationary satellites for Earth observation.

These satellites have expanded their role beyond communications to include vital Earth observation applications. Continuous monitoring from geostationary orbits enables valuable insights into terrestrial ecosystems, demonstrated by various satellite missions such as the *Geostationary Operational Environmental Satellite* (GOES) series (Goodman et al., 2019; Ludwig and Johnson, 1981; Menzel and Purdom, 1994), Himawari-8/9, and Fengyun-4A/B (Khan et al., 2021a; Menzel, 2020). Geostationary satellites are distinct in their fixed positioning relative to a specific point on the Earth's surface, allowing them to provide near-global observational coverage via a strategic network (Figure 1.2). This unique stationary configuration underscores the pivotal role of geostationary satellites in our interconnected world, where they contribute significantly to scientific research, meteorological forecasting, and ecosystem modeling.

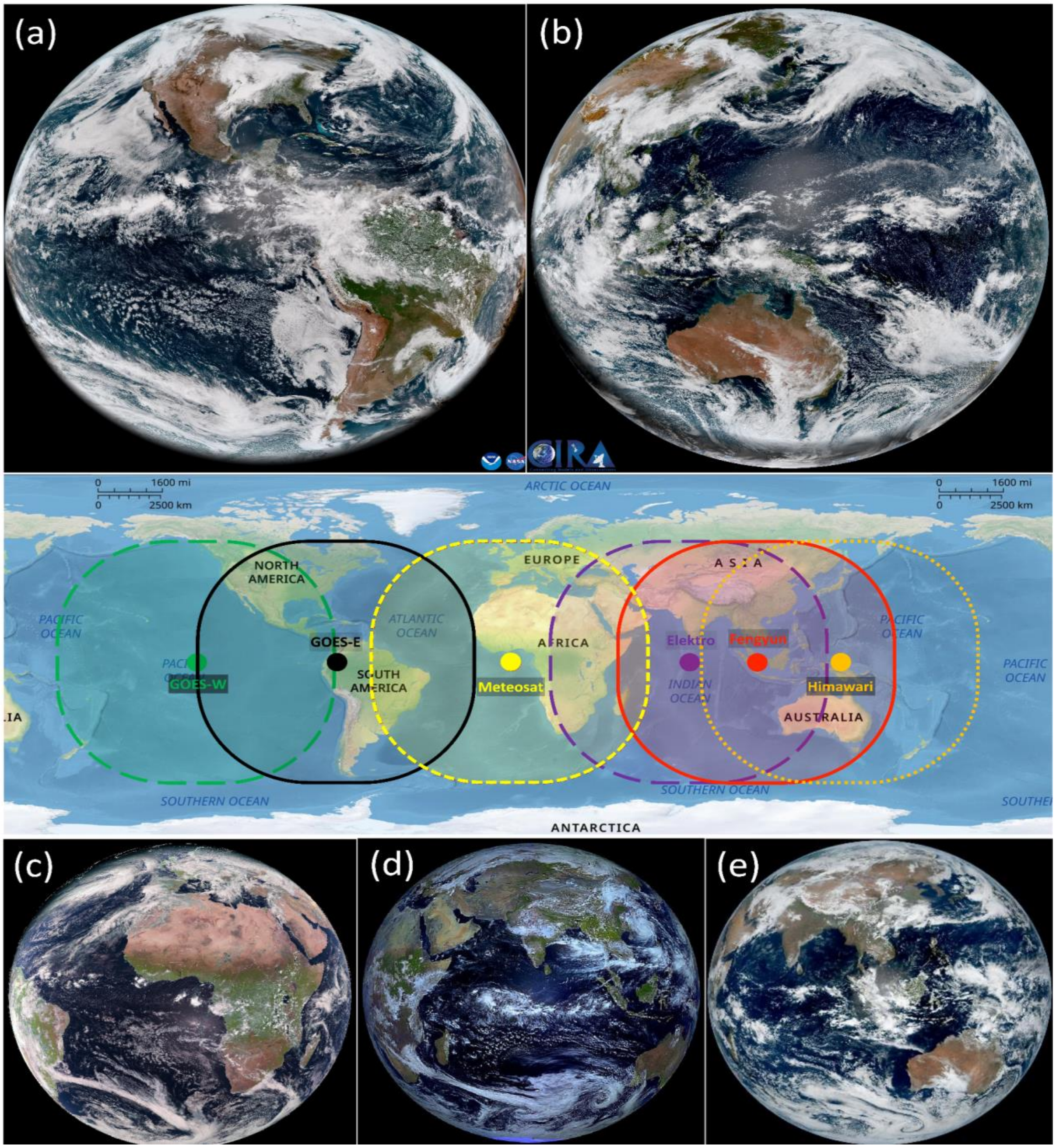

Figure 1.2: Near-global coverage of the new generation of geostationary satellites for Earth observation, showcasing full-disk images captured by *GOES-E*, *Himawari*, *Meteosat*, *Elektro*, and *Fengyun* (a–e), respectively.

The launch of *Syncom 2* in 1963, primarily designed as a communications satellite, also carried a rudimentary radiometer capable of measuring cloud cover (Newell, 1965; Schroer, 2008). This early dual-purpose design foreshadowed the potential for geostationary satellites to contribute to Earth observation (Ilčev and Ilčev, 2019). However, the true milestone in Earth observation via geostationary satellites came in 1975 with the launch of *Meteosat 1* by the European Space Agency (ESA), the first satellite dedicated to Earth observation (Ellrod and Pryor, 2019; Ilčev and Ilčev, 2019). Equipped with a visible and infrared (VIS/IR) sensor, *Meteosat 1* marked a significant leap in Earth observation capabilities, allowing for imaging of the Earth's surface and measurement of critical atmospheric parameters, such as temperature and cloud cover (Brooks et al., 1984; Menzel, 2020). This sensor allowed for imaging of the Earth's surface and the measurement of various critical atmospheric parameters, including temperature and cloud cover (Brooks et al., 1984; Davis, 2007; Menzel, 2020). This laid the groundwork for geostationary satellites' role in climate and weather monitoring (Menzel, 2020).

As the 1980s and 1990s progressed, geostationary Earth observation satellites continued to evolve, particularly in spatial and spectral resolution. The integration of advanced sensors, such as water vapor and ozone detectors, expanded their observational capabilities, enabling monitoring of a broader range of Earth phenomena (Davis, 2007; Menzel, 2020). These advancements improved spatial resolution, reaching up to 1 km, providing more detailed and accurate data for scientists and meteorologists (Khan et al., 2021a; Menzel, 2020).

Geostationary satellites like *GOES* and the *Himawari* series began to serve a dual role in weather monitoring and observing ecosystem functions (Khan et al., 2021a; Menzel, 2020). These satellites provided real-time, high-temporal-resolution imagery, which was particularly useful for monitoring vegetation health, land cover changes, and atmospheric interactions, in line with ecological concepts of ecosystem functions and services (Pettorelli et al., 2014). As ecological research gained prominence, geostationary satellites extended their applications to encompass ecosystem monitoring, focusing on the detection of rapid environmental changes and responses to disturbances such as extreme weather events and deforestation (Jansen et al., 2010).

In the 2000s and 2010s, the launch of advanced geostationary satellites such as *GOES-R*, *Himawari-8/9*, and *Fengyun-4A/B* marked another significant advancement in Earth observation (Khan et al., 2021a;

Menzel, 2020). These satellites were equipped with an array of sensors, including VIS/IR, near-infrared (NIR), shortwave infrared (SWIR), and mid-wave infrared (MWIR) sensors (Ellrod and Pryor, 2019; Khan et al., 2021a; Menzel, 2020). The enhanced sensors provided unparalleled detail and accuracy, improving climate and weather forecasting, as well as enabling the monitoring of land use and environmental changes (Goodman et al., 2019; Khan et al., 2021a; Küçük et al., 2022). The latest generation of geostationary satellites, with their ability to collect NIR and SWIR data, has become an essential tool for monitoring terrestrial ecosystems (Khan et al., 2021a; Xu et al., 2017). For instance, in tropical forests, these satellites now play a key role in detecting changes in ecosystem functions, such as carbon storage and climate regulation, resulting from disturbances like defaunation (Osuri et al., 2016; Peres et al., 2016).

1.3. Terrestrial ecosystem modeling with geostationary remote sensing

Terrestrial ecosystem functions, a term often used interchangeably with ecosystem services, ecological processes, and ecosystem processes, lack a universally agreed-upon definition, posing challenges for precise monitoring and scientific discourse (Dominati et al., 2010; Lawton et al., 1993; Mace et al., 2012; Srivastava and Vellend, 2005). Pettorelli et al., (2018) defined terrestrial ecosystem functions as attributes intrinsically linked to ecosystem performance, arising from one or multiple ecosystem processes. These functions provide direct and indirect benefits to various species, including humans, and encompass critical processes such as nutrient regulation, food production, and water supply (Pettorelli et al., 2018, 2014). The ability to monitor and model these functions has been enhanced by leveraging the sensor advancements in geostationary remote sensing, which provides continuous, high-temporal-resolution observations of the Earth's surface. The historical development of terrestrial ecosystem modeling using remote sensing data is closely tied to the evolution of satellite sensor technologies. Early sensors, such as radiometers, provided limited spectral information, but over time, advancements in visible, infrared, and NIR sensors enabled more comprehensive monitoring of ecosystem dynamics (AghaKouchak et al., 2015; Xiao et al., 2019). Over the subsequent decades, advancements in sensor technologies on geostationary satellites transformed them into indispensable tools for monitoring the terrestrial ecosystem, including weather patterns (Lovett et al., 2005; Menzel, 2020), climate change (AghaKouchak et al., 2015; Anderegg et al., 2020; Nemani et al., 2003), deforestation (AghaKouchak et al., 2015; Sokolik et al., 2019), wildfire detection and natural disasters (Mohapatra and Trinh, 2022; Ye, 2022), carbon assimilation and fluxes (Anderson et al., 2000; Xiao et al., 2019), surface-atmosphere interactions (Sokolik et al., 2019; Veroustraete et al., 1996), vegetation dynamics (Dechant et al., 2022a; Khan et al., 2021a; Massetti et al., 2019; Zhang et al., 2006), and water resources (Ellrod and Pryor, 2019; Khan et al., 2021a; Ranjbar et al., 2024c).

Unlike polar-orbiting satellites, geostationary platforms offer the unique ability to monitor diurnal and seasonal variations in key terrestrial ecosystem variables. This capability has enabled the development of advanced models that leverage geostationary data to improve estimates of carbon and water fluxes, supporting applications in climate research, agriculture, and environmental management. One of the pioneering models in this domain is the ALEXI (Atmosphere-Land Exchange Inverse) model, which estimates evapotranspiration (ET) by coupling canopy conductance to carbon dioxide uptake and using land surface temperature (LST) data (Anderson et al., 2000). ALEXI employs a two-source energy balance approach to partition heat fluxes from soil and vegetation, reducing sensitivity to LST errors by focusing on morning temperature changes (Kustas and Norman, 1999; Norman et al., 2003). This model has been widely applied for drought monitoring, water resource management, and crop productivity estimation, with its DisALEXI extension further enhancing spatial resolution by integrating Landsat thermal imagery (Anderson et al., 2016, 2013; Otkin et al., 2018, 2014, 2013). To provide gridded ET information at field scales for many applications in agriculture and water resource management, the ALEXI

ET estimates at 4-km resolution are disaggregated using the DisALEXI approach and Landsat thermal and optical imagery (Anderson et al., 2012; Norman et al., 2003). The ET data at 30-m resolution have been incorporated into the OpenET ensemble simulation of ET for water management in the western U. S. (Melton et al., 2022).

Similarly, the SEVIRI instrument on the METEOSAT Second Generation satellites provides operational gross primary productivity (GPP) products that incorporate a water stress coefficient (Cws) to account for reduced photosynthetic activity during droughts (Martínez et al., 2018). These products have demonstrated exceptional performance in detecting water stress in Mediterranean and savanna ecosystems (Martínez et al., 2020; Sanchez-Ruiz et al., 2017). Other notable models include the MODIS GPP product, which uses light-use efficiency (LUE, (Wu et al., 2010)) principles to estimate global vegetation productivity, and the BIOME-BGC model, which simulates carbon, nitrogen, and water fluxes across terrestrial ecosystems (Running et al., 1999; Sanchez-Ruiz et al., 2018). Additionally, the LPJ (Lund-Potsdam-Jena) dynamic global vegetation model and the ORCHIDEE land surface model have been instrumental in simulating vegetation distribution and carbon-water interactions under changing climate conditions (Chang et al., 2013; Jung et al., 2017).

ALEXI, LUE-based GPP products, SEVIRI/MSG, and several other approaches have been developed to enhance the accuracy of ecosystem monitoring under water-limited conditions. The GPP product derived from data has demonstrated its capability to capture short-term declines in photosynthetic activity due to water stress, incorporating a water stress coefficient (Cws) based on the ratio of actual evapotranspiration (AET) to potential evapotranspiration (PET) (Sanchez-Ruiz et al., 2018). This approach improves sensitivity to drought effects by integrating a localized water budget into the GPP estimation algorithm. However, studies have shown that vapor pressure deficit (VPD)-based downregulation of light-use efficiency (LUE), as implemented in the MOD17 algorithm, may not adequately represent water stress impacts in semi-arid ecosystems, such as Mediterranean savannas (W. Chen et al., 2023; Pettorelli et al., 2014). Instead, soil moisture-based stress factors have been found to enhance model performance in these environments. At site-level scales, the MSG GPP product has proven effective in detecting drought events across different latitudes, such as the severe El Niño-induced drought at the Kruger National Park FLUXNET site (ZA-Kru) in 2015/2016 and the extreme dry conditions in Spain's Las Majadas site (ES-LMa) during 2017 (Marshall et al., 2018; Martínez et al., 2020; Sanchez-Ruiz et al., 2017). Furthermore, operational GPP products, such as the 8-day 1 km MODIS GPP and the 10-day Gross Dry Matter Productivity (GDMP) from the Copernicus Global Land Service, have been compared against MSG-derived estimates, with SEVIRI-based retrievals demonstrating robust responses to seasonal water stress (Martínez et al., 2020). Similarly, land surface drought monitoring using satellite-based evapotranspiration (ET) models remains an ongoing challenge. While conventional ET models have primarily been applied at regional scales (Amani and Shafizadeh-Moghadam, 2023; Zhang et al., 2016), recent advancements, such as SEVIRI-based reference and real ET products, have been leveraged to assess drought occurrence in regions with strong land–atmosphere coupling, such as Southeastern Europe (Stoyanova and Georgiev, 2010). These models, combined with geostationary data, have enabled more frequent monitoring of ecosystem responses to environmental stressors, such as droughts and heatwaves, and have supported large-scale assessments of ecosystem services. Despite these advancements, challenges remain in improving the spatial resolution and accuracy of geostationary-based models, particularly in heterogeneous landscapes.

The high temporal resolution of geostationary satellite observations, capturing diurnal and seasonal variations at minute-scale intervals, provides unprecedented insights into key ecosystem processes. These data are essential for monitoring diurnal cycling of terrestrial photosynthesis, improving phonological assessments, and understanding plant water stress dynamics across diverse ecosystems, including the Amazon (Hashimoto et al., 2021). Geostationary-based models enable real-time crop monitoring, facilitating early detection of water stress and optimizing agricultural management strategies. They also play a crucial role in assessing the impacts of extreme climate events, such as flash droughts, which significantly influence global GPP. Additionally, geostationary data enhance our understanding of ecosystem responses to synoptic weather patterns, where even short periods of favorable conditions can drive annual variability in carbon fluxes (Khan et al., 2021a; Ranjbar et al., 2024c). Studies have demonstrated how synoptic weather events control carbon exchange across regions like the Great Lakes (Randazzo et al., 2020), while geostationary monitoring improves tracking of episodic disturbances such as air pollution episodes. These data also support ecological forecasting and data assimilation efforts, advancing near-real-time predictions of ecosystem function. Geostationary-based evapotranspiration models, including PT-JPL approaches that leverage NDVI and SAVI, provide key insights into land-atmosphere interactions (Amani and Shafizadeh-Moghadam, 2023; Anderson et al., 2016). Furthermore, the capacity to monitor fire events and burn scars allows for improved assessments of their carbon cycle implications, offering critical information for climate mitigation and land management. Despite existing challenges in refining spatial resolution, the integration of geostationary observations with advanced ecosystem models continues to transform our ability to monitor and predict environmental changes with greater accuracy and timeliness.

1.4. Challenges, needs, and opportunities

Geostationary satellites offer unique advantages for monitoring terrestrial ecosystems, yet several challenges must be addressed to fully harness their potential. These challenges span sensor limitations, data processing constraints, and the need for improved algorithms and international collaboration (Ellrod and Pryor, 2019; Khan et al., 2021a; Shen et al., 2021). At the same time, emerging opportunities present pathways to enhance geostationary remote sensing applications for ecosystem modeling and climate studies.

One of the major challenges in geostationary satellite remote sensing is subpixel variability (Khan et al., 2021a; Ranjbar et al., 2024c), particularly in heterogeneous landscapes. Existing products operate at varying spatial resolutions, which may not adequately capture fine-scale ecosystem processes. Geostationary sensors typically provide kilometer-scale observations (Goodman et al., 2019; Menzel, 2020), which can obscure small-scale variations in vegetation structure, land cover, and microclimate. Additionally, regions with complex terrain, such as mountainous areas, introduce further difficulties due to parallax effects (Losos et al., 2024a; Ranjbar et al., 2024c). These distortions arise from the oblique viewing angles of geostationary satellites, affecting geolocation accuracy and the interpretation of surface properties (Losos et al., 2024a). Advanced topographic corrections and fusion with other satellite datasets are necessary to mitigate these issues.

Cloud contamination remains a critical limitation in geostationary observations. Current cloud masks often misclassify thin clouds and aerosols, leading to errors in surface parameter retrievals. Enhancing cloud mask algorithms, incorporating machine learning techniques, and utilizing multi-temporal and multi-spectral data could significantly improve the reliability of cloud-cleared observations. The high temporal resolution of geostationary satellites generates vast amounts of data, posing challenges for storage, processing, and real-time analysis (Ranjbar et al., 2024c, 2024d). Computational demands increase with higher spatial and temporal resolutions, requiring efficient data handling solutions such as cloud computing, distributed processing, and scalable storage architectures (Nemani et al., 2020; Ranjbar et al., 2024d).

Existing products, such as GOES-R Downward Shortwave Radiation (DSR), are typically produced at hourly time steps due to historical computational constraints. However, higher-frequency products, such as those generated at five or ten-minute intervals, which are the native temporal resolution of GOES-R satellite observations (Ranjbar et al., 2024d), could better capture rapid changes in surface energy balance and improve applications in climate and ecosystem modeling. Expanding the suite of geostationary-derived products presents new research opportunities. For example, diffuse radiation estimates, which are critical for ecosystem productivity modeling, remain underdeveloped. Additionally, fusion products that integrate geostationary and polar-orbiting satellite observations could enhance data accuracy and coverage. Open-source algorithms and transparent processing workflows can enhance reproducibility and foster collaborative research. Platforms such as Google Earth Engine provide an opportunity to improve data usability and accessibility by hosting geostationary datasets and enabling cloud-based analyses (Kovács et al., 2023; Tamiminia et al., 2020). Coordinated efforts across international space agencies and research institutions can accelerate progress in geostationary remote sensing. By sharing data, algorithms, and best practices, global initiatives can facilitate harmonized products that benefit climate science, agriculture, and disaster response applications. Integrating geostationary satellite observations with near real-time CO_2 emissions data enables more effective tracking of human impacts on the carbon cycle, improving our ability to monitor and respond to changes in greenhouse gas dynamics (Xiao et al., 2019). Additionally, developing algorithms for estimating ecosystem respiration and non-CO_2 greenhouse gases, such as methane and nitrous oxide, could enhance our understanding of biogeochemical feedback under changing climate conditions. Addressing these challenges and leveraging new opportunities will be essential for advancing geostationary remote sensing in terrestrial ecosystem studies. By improving resolution, data processing, and algorithm development, geostationary observations can play a pivotal role in monitoring and modeling ecosystem dynamics at high temporal scales.

Chapter 2. High-Frequency Mapping of Downward Shortwave Radiation from GOES-R Using Gradient Boosting

This chapter is published in the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing: S. Ranjbar, D. Losos, S. Hoffman and P. C. Stoy, "High-Frequency Mapping of Downward Shortwave Radiation From GOES-R Using Gradient Boosting," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 11958-11968, 2024, doi: 10.1109/JSTARS.2024.3420148

Abstract

This study investigates high-frequency mapping of downward shortwave radiation (DSR) at Earth's surface using the Advanced Baseline Imager (ABI) instrument mounted on Geostationary Operational Environmental Satellite - R Series (GOES-R). The existing GOES-R DSR product (DSR_{ABI}) offers hourly temporal resolution and spatial resolution of 0.25°. To enhance these resolutions, we explore machine learning (ML) for DSR estimation at the native temporal resolution of GOES-R Level-2 Cloud and Moisture Imagery (CMI) product (five minutes) and its native spatial resolution of two kilometers at nadir. We compared four common ML regression models through the Leave-One-Out Cross-Validation (LOOCV) algorithm for robust model assessment against ground measurements from AmeriFlux and SURFRAD networks. Results show that Gradient Boosting Regression (GBR) achieves the best performance ($R^2 = 0.916$, RMSE = 88.05 W m^{-2}) with more efficient computation compared to Long Short-Term Memory (LSTM), which exhibited similar performance. DSR estimates from the GBR model through the ABI Live Imaging of Vegetated Ecosystems workflow (DSR_{ALIVE}) outperform DSR_{ABI} across various temporal resolutions and sky conditions. DSR_{ALIVE} agreement with ground measurements at SURFRAD networks exhibits high accuracy at high temporal resolutions (five-minute intervals) with R^2 exceeding 0.85 and RMSE=122 W m^{-2}. We conclude that GBR offers a promising approach for high-frequency DSR mapping from GOES-R, enabling improved applications for near-real-time monitoring of terrestrial carbon and water fluxes.

2.1. Introduction

Downward shortwave radiation (DSR) at the Earth's surface has long been estimated using geostationary ("weather") satellites by combining measured top-of-atmosphere (TOA) reflectances with radiative transfer models (Diak and Gautier, 1983; Gautier et al., 1980; H.-Y. Kim et al., 2020; Laszlo et al., 2020a; Zhang et al., 2021a). The NOAA operational algorithm using observations from the Advanced Baseline Imager (ABI) on the Geostationary Operational Environmental Satellite - R Series (GOES-R) provides hourly estimates of DSR (Laszlo et al., 2020a) (hereafter DSR_{ABI}) despite scanning on five minute intervals across the conterminous United States (CONUS) and ten-minute intervals across the Western Hemisphere "full disk" (Schmit et al., 2017a; Schmit and Gunshor, 2020). This hourly cadence complicates nascent efforts to apply geostationary remote sensing to estimate terrestrial carbon and water cycling in near-real time (A. Khan et al., 2022; Ranjbar et al., 2023a), because additional approaches must be taken to infer hourly DSR at more frequent intervals, which introduce uncertainty (Losos et al., 2024a). The DSR_{ABI} product creates maps with $0.25^{o} \times 0.25^{o}$ (0.25^{o}=27-28 km) spatial resolution over CONUS (Laszlo et al., 2020b) when most ABI products are created at two-kilometer resolution at nadir, with many missing estimates that require uncertain gapfilling to obtain continuous data (Figure 2.1). Reduced solar radiation, cloud cover, aerosol concentration, atmospheric pollution, atmospheric moisture content, variations in surface albedo (reflectivity) and various other environmental factors can contribute to data loss in the downward shortwave radiation product. For instance, during winter, the DSR_{ABI} typically experiences approximately two hours with 90% or more of valid pixels in an image, reflecting periods of relatively clearer atmospheric conditions conducive to accurate data collection.

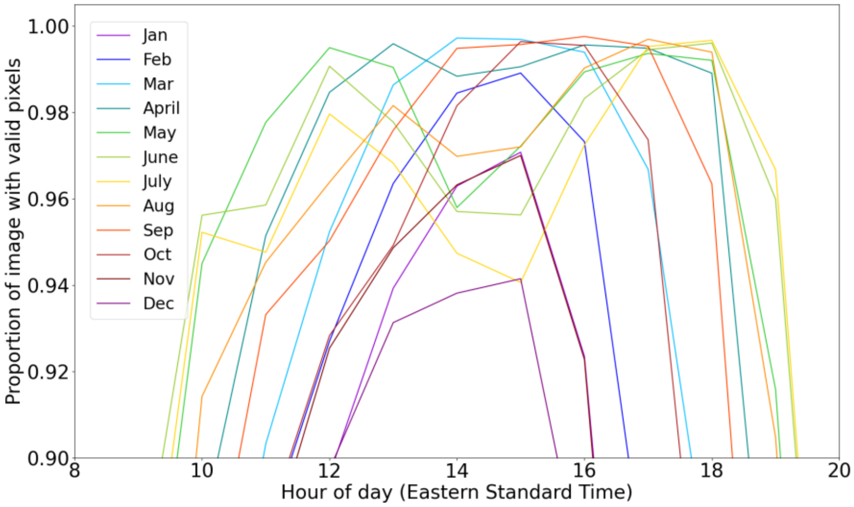

Figure 2.1. The proportion of valid pixels in the DSR$_{ABI}$ product from GOES-16 across the CONUS view in 2022, averaged over the course of the day for each month.

Recent work has applied machine learning (ML) to estimate DSR and/or photosynthetically-active photon flux density (PPFD) from the observations of multiple geostationary satellites (Hao et al., 2019; Jiang et al., 2019; Yeom et al., 2019; Zhang et al., 2021a; Zhao et al., 2022). Artificial Neural Networks (ANNs), recurrent ANNs like Long Short-Term Memory (LSTM), Random Forest (RF), and deep learning approaches are often determined to have improved skill (Chen et al., 2021; Jiang et al., 2019; R. Li et al., 2023; Ma et al., 2020; Peng et al., 2020; Ranjbar et al., 2021; Yeom et al., 2019), due in part to their ability to capture nonlinear relationships amongst multiple variables. Here, following the approach of (Ranjbar et al., 2023a) for carbon dioxide flux estimation using GOES-R, we demonstrate how gradient boosting regression (GBR) can give comparable skill in DSR estimation compared to more computationally expensive approaches like LSTM, Multi-Layer Perceptron Neural Networks (MLP), and RF to efficiently estimate DSR in near-real time from Level 2 Cloud and Moisture Imagery (CMI) reflectance factors and brightness temperatures from the ABI across CONUS.

2.2. Data

We utilized ground measurements from the Ameriflux and NEON tower networks and SURFRAD networks, along with geostationary satellite observations obtained from the ABI sensor mounted on GOES-R series satellites. To create an efficient estimate of DSR using ABI data, we sampled time series of ABI Level 2 products at 99 sites across CONUS from pixels where Ameriflux eddy covariance towers are located that measure incident shortwave radiation (SW-IN) for the period 2020-2022 (Figure 2.2). Time series data collection followed the method proposed in (Losos et al., 2024a), which applies a point-based terrain-correction to the ABI Fixed Grid projection to mitigate the effect of parallax displacement. The 2020-2022 period captures a range of atmospheric conditions across multiple climate zones for training and testing our model, hereafter $\text{DSR}_{\text{ALIVE}}$, following the ABI Live Imaging of Vegetated Ecosystems workflow (Ranjbar et al., 2023a). The SW-IN measurements underwent consistent quality control checks through the standardized Ameriflux processing scheme (Chu et al., 2023). These SW-IN data were combined with GOES-R ABI Cloud and Moisture Imagery (CMI) observations, solar zenith angle (SZA), and solar azimuth angle (SAA) (SZA and SAA were calculated using NOAA solar position calculator following https://gml.noaa.gov/grad/solcalc/calcdetails.html) to build a database of predictor variables (CMIs, SZA, & SAA) and target variables (SW-IN) to explore ML methods to estimate $\text{DSR}_{\text{ALIVE}}$. ABI CMI observations include 16 bands collected across the visible, near-infrared, and infrared ranges of the electromagnetic spectrum (see (Schmit et al., 2017a; Schmit and Gunshor, 2020) for more specific information about the bands). The model performance was compared against independent SW-IN observations from the Ameriflux and NEON tower networks and SURFRAD (Augustine et al., 2000; Chu et al., 2023; Metzger et al., 2019). The SURFRAD network measures DSR ($\text{DSR}_{\text{SURFRAD}}$) every minute, providing high-quality data at seven sites ideal for evaluating $\mathrm{DSR}_{\mathrm{ALIVE}}$ at various temporal intervals (refer to (Augustine et al., 2000) for further information). The SURFRAD sites are located in seven different states across US (Figure 2.2), including Illinois (with site code of BND), Colorado (TBL), Nevada (DRA), Montana (FPK), Mississippi (GWN), Pennsylvania (PSU), South Dakota (SXF), and Oklahoma (SGP). We further compared $\mathrm{DSR}_{\mathrm{ALIVE}}$ with $\mathrm{DSR}_{\mathrm{ABI}}$, and $\mathrm{DSR}_{\mathrm{ABI}}$ to SW-IN, to understand differences that may arise between the different ABI-based data products. We analyzed data with respect to clear and cloudy skies using the ABI Clear Sky Mask product, and also explored the ability of the ABI Aerosol Optical Depth (AOD) product to improve predictions (Knapp et al., 2005; H. Zhang et al., 2020), noting that AOD estimates from the ABI also include frequent missing observations.

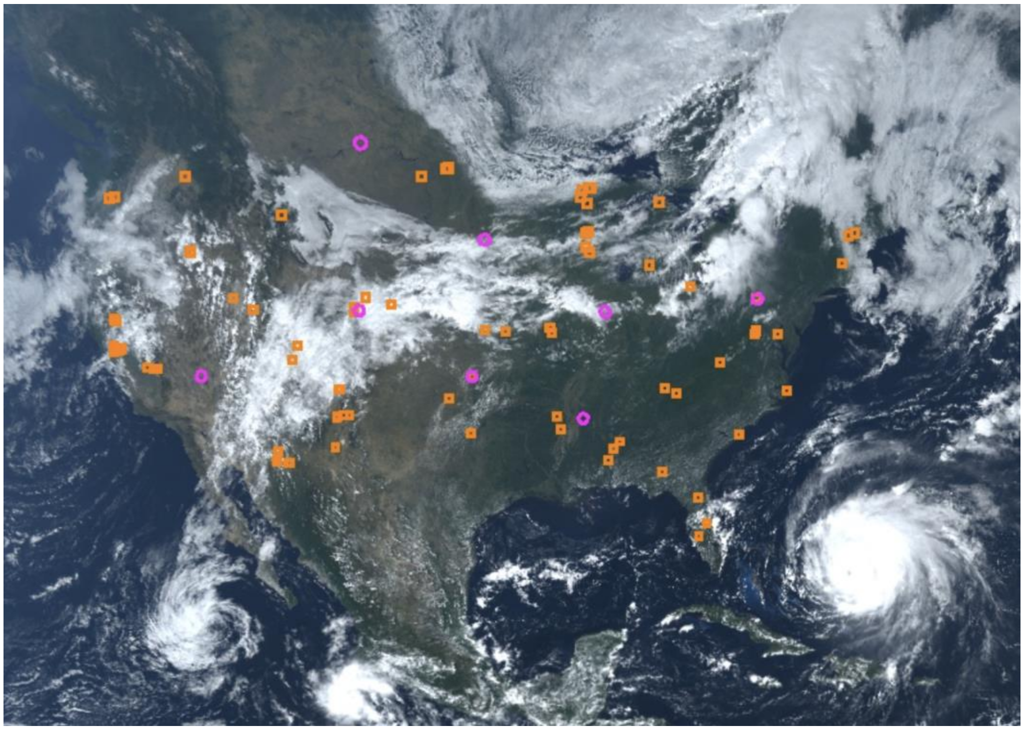

Figure 2.2. The location of Ameriflux towers (orange squares) and SURFRAD sites (magenta circles) used in this study, mapped onto GOES-16 CONUS "true color" imagery from Sept 21, 2022.

2.3. Methods

2.3.1 Regression models: RF, GBR, LSTM, and MLP

We explored four common ML regression models to create $\mathrm{DSR}_{\mathrm{ALIVE}}$: RF, GBR, LSTM, and MLP

(Diaz-Gonzalez et al., 2022; Thapa et al., 2023; Varghese et al., 2023).

RF and GBR, both ensemble models, blend decision trees via bootstrapping and feature bagging (Friedman, 2001). As applied in remote sensing (Bahrami et al., 2021a; Sahin, 2020; Saini and Ghosh, 2017; Zhang et al., 2022a), they gauge feature importance but differ in focus. RF prioritizes simplicity and minimizing training set loss, while GBR emphasizes optimizing a loss function for error reduction. GBR iteratively constructs an additive model through a forward stage-wise approach, facilitating the optimization of diverse differentiable loss functions, while RF combines trees independently (Bentéjac et al., 2021; Sahin, 2020). Both require hyperparameter tuning, influencing model size and depth. In both models, shared hyperparameters include "number of estimators," "maximum depth," and "minimum samples per leaf." These parameters collectively determine model complexity, regularization, and generalization ability (Bentéjac et al., 2021; Ranjbar et al., 2021; Zhang et al., 2022a). The "number of estimators" controls the quantity of decision trees in the ensemble, while "maximum depth" limits tree complexity to mitigate overfitting. "Minimum samples per leaf" sets the threshold for node splitting, aiding in regularization (Bahrami et al., 2021a; Zhang et al., 2022a). Finally, the learning rate parameter modulates the contribution of each tree to the ensemble, fostering a balanced trade-off between convergence speed and model stability (Bentéjac et al., 2021; Saini and Ghosh, 2017).

LSTM networks have proven instrumental in handling sequential data, showcasing proficiency in time series forecasting (Hochreiter and Schmidhuber, 1997). Their efficacy in capturing long-term dependencies, facilitated by unique memory cell architecture (Ghanbari et al., 2021; Sutskever et al., 2014), addresses a key limitation of traditional Recurrent Neural Networks (RNNs) and can handle non-linear relationships within sequences (Sutskever et al., 2014). Despite their advantages, the complexity of LSTM architecture demands higher computational resources compared to simpler RNNs (Schuster and Paliwal, 1997; Yuan et al., 2020). Additionally, achieving optimal performance often requires careful hyperparameter tuning, adding intricacy to the training process (Hochreiter and Schmidhuber, 1997). Researchers must weigh these computational costs and tuning challenges against the benefits when applying LSTMs, to tailor an approach aligned with dataset characteristics and research goals (Ghanbari et al., 2021; Schuster and Paliwal, 1997; Sutskever et al., 2014; Yuan et al., 2020). Our LSTM architecture utilized two LSTM layers with an identical number of units. The first layer (`return_sequences=True`) retained and returned the entire sequence of hidden states for each input sequence, crucial for capturing temporal dependencies. Conversely, the second layer (`return_sequences=False`) only returned the output at the last time step, effectively summarizing the learned information. Following the LSTMs, two dense layers were added. The first dense layer, with 25 units, facilitated further feature extraction and representation learning. The final dense layer, with a single unit, produced the model's output.

MLP is an ANN for learning patterns through backpropagation (Baraldi et al., 2001; Gardner and Dorling, 1998; Suliman and Zhang, 2015). With layers of interconnected neurons and non-linear activation functions, MLP is versatile in approximating non-linear relationships (Bahrami et al., 2021b; Baraldi et al., 2001; Suliman and Zhang, 2015). Hyperparameters contribute to model complexity, non-linear mapping, convergence speed, and preventing overfitting (Shirmard et al., 2022).

MLPs and LSTM networks rely on hyperparameters to optimize their performance. For MLPs, activation functions, solver algorithms, hidden layer configurations (including size), and regularization terms significantly impact the network's architecture, learning dynamics, and overfitting tendencies (Ghanbari et al., 2021; Suliman and Zhang, 2015). In LSTMs, the number of units (neurons) determines memory capacity and computational complexity, while activation functions modulate information flow within memory cells (Ghanbari et al., 2021). Optimizer choice steers training dynamics and convergence, and batch size regulates efficiency and memory usage during training iterations (Dhake et al., 2023; Ghimire et al., 2022). A mean squared error loss function was used for both MLP and LSTM.

2.3.2 Model Development

We used Google Colab Pro with random access memory (RAM) up to 32 GB and an A100 graphics processing unit (GPU) for model development. To optimize model performance, we employed a grid search algorithm, systematically testing hyperparameter combinations within each ML model to identify configurations that maximize correlation (in terms of R^2) and prediction speed (Maxwell et al., 2018; Varghese et al., 2023). Table 2.1 outlines the hyperparameter search specifications and ranges for each ML model.

Table 2.1 Grid Search Specifications and Hyperparameter Setting for Machine Learning Models

Model	**Hyperparameter**	**Grid Setting**
RF	Number of estimators	100, 300, 500, 800, 1000
	Maximum depth	5, 8, 10, 15, 20, 30
	min_samples_leaf	50, 100, 200, 500
GBR	Number of estimators	100, 300, 500, 800
	Maximum depth	3, 5, 8, 10
	min_samples_leaf	50, 100, 200, 500
	Learning rate	0.01,0.05,0.1
MLP	Activation	'logistic', 'tanh', 'relu'
	Solver	'adam','sgd'
	Hidden layer	2, 3, 5
	Hidden layer size	50, 100, 200, 500

	Alpha	0.0001, 0.001, 0.01, 0.1
LSTM	Units (neurons)	32, 64, 128, 256, 512
	Activation	'logistic', 'tanh', 'relu'
	Optimizer	'adam','sgd'
	Batch Size	32, 64, 128, 256

For validation and evaluation, we adopted a Leave-One-Out Cross-Validation (LOOCV) algorithm, reserving 20% of the data exclusively for testing, while the remaining data underwent a four-fold cross-validation process with a 75:25 training-validation split (Maxwell et al., 2018). To enhance reliability and robustness, we reran the algorithm ten times and averaged the results using the Scikit-learn Python library (Pedregosa et al., 2011) and used the coefficient of determination (R^2) and root mean squared error (RMSE) to evaluate and compare models. Following model comparison, we used feature importance results to identify key variables for modeling DSR in a SHapley Additive exPlanations (SHAP) analysis (Lundberg and Lee, 2017).

SHAP is a powerful tool rooted in game theory for unraveling the opaque nature of ML predictions (Lundberg and Lee, 2017). It uncovers individual feature contributions by quantifying the direction and magnitude of their influence to identify synergistic or antagonistic relationships that steer model behavior (Lundberg et al., 2020). SHAP is not a universal solution; computational costs may scale with data size, and its effectiveness can be influenced by model complexity (Molnar et al., 2018). Nevertheless, SHAP can clarify individual predictions, unveil feature interactions, and demystify the "black box" nature of ML models. SHAP values are a unified measure to explain the output of machine learning models by quantifying the contribution of each feature to the prediction. Higher SHAP values denote features with more significant impact and importance on the model's predictions.

2.4 Results

2.4.1 ML performance

The GBR and LSTM models displayed better performance than the other models with R^2 of 0.92, an improvement over the relationship between DSR_{ABI} against SW-IN observations for 314 Ameriflux sites (R^2 of 0.83 reported by (Losos et al., 2024a)). The prediction time for LSTM is 19.5 times that of GBR (Table 2.2), which decreases to 4.2 times that of GBR with the use of Google Colab Pro A100 GPU. Given our objective of generating hypertemporal DSR maps from ABI imagery captured every 5 minutes, prediction time is critical. In this context, GBR emerged as the optimal choice among the ML models due to its similar accuracy and reduced prediction time compared to LSTM.

Table 0.2 Models Performance vs. Prediction Time for ML model training using Google Colab Pro. Prediction time unit is seconds per sample. The number in parentheses is the relative prediction time compared to GBR. Units for RMSE are W m^{-2}.

Model	Prediction Time	Train R^2	Train RMSE	Test R^2	Test RMSE
RF	CPU: 2.7 E-05 (2.2) A100 GPU: 2.7 E-05 (2.2)	0.90	93.41	0.884	105.68
GBR	CPU: 1.2 E-05 (1) A100 GPU: 1.2 E-05 (1)	0.93	84.33	0.916	88.05
MLP	CPU: 3.41 E-05 (2.8) A100 GPU: 2.5 E-05 (2.1)	0.90	95.78	0.888	101.15
LSTM	CPU: 23.4 E-05 (19.5) A100 GPU: 5.1E-05 (4.2)	0.92	85.09	0.914	87.52

2.4.2 GBR model efficiency

To enhance efficiency of prediction time vs accuracy, we conducted training experiments on the GBR model, varying the number of samples per site (with 20,965 available samples per site), the model parameters maximum depth (m_depth), and the number of estimators (n_estimators). We explored training with 200, 500, 1000, 2000, 4000, and 8000 samples per site, combinations of m_depth between 3, 5, 8, 10, and 12, and n_estimators of 100, 300, 500, 800, and 1000.

Models trained with 2000 samples per site exhibited better performance in terms of prediction time versus accuracy compared to those trained with 200, 500, or 1000 samples per site (see Figure 2.3). Moreover, our analysis reveals that the model's performance stabilizes with 2000 samples per site, suggesting diminishing returns with larger sample sizes. Therefore, optimizing the sample size to balance accuracy and computational efficiency is paramount for our dataset generation objectives. Through iteration, we identified the most efficient GBR model, balancing prediction time (1.18E-5 second per sample) and correlation (R^2 of 0.91). This optimal model was trained with 2000 samples per site with m_depth = 8 and n_estimator = 300. The selection of this model is visually depicted by the black cross in Figure 2.3. We used this model for DSR_{ALIVE}.

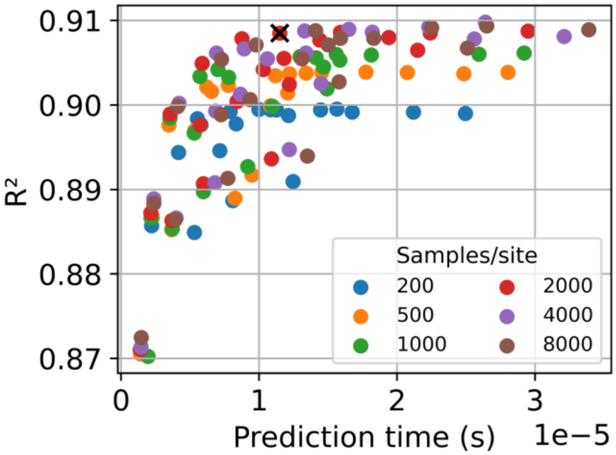

Figure 2.3. R^2 score vs. prediction time for GBR models (see Table 2.2) with different m_depth, n_estimator, and training size. The black cross shows the selected GBR model that balances prediction time and accuracy.

A density scatter plot illustrating the relationship between SW-IN and DSR_{ALIVE} is shown in Figure 2.4. The R^2 between half-hourly SW-IN and DSR_{ALIVE} is 0.92 with an RMSE of 86.12 W m^{-2}.

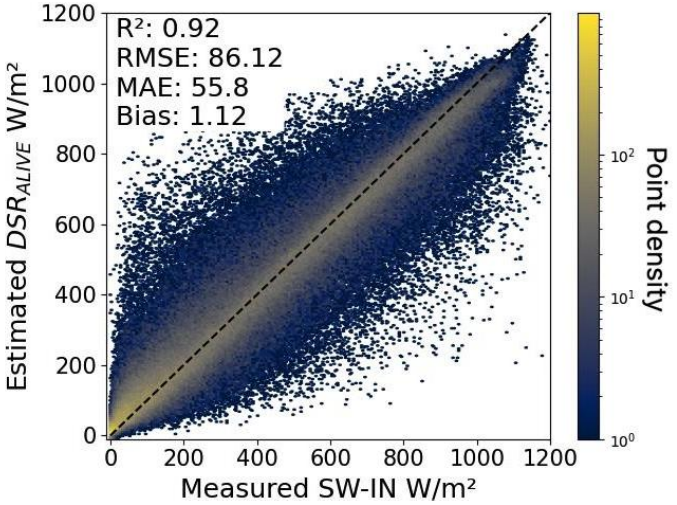

Figure 2.4. Density scatter plot between half-hourly DSRALIVE and SW-IN measurements from eddy covariance towers.

Table 2.3 presents the agreement of DSRALIVE vs SW-IN from the average of running LOOCV algorithm ten times. We computed R^2 and RMSE metrics by taking the mean DSR across different temporal resolutions (half-hourly, daily, and weekly) and sky conditions (clear and cloudy). Moreover, daily analysis revealed a more robust performance of the GBR model when comparing estimated DSRALIVE to SW-IN, with R^2 values of 0.99, RMSE of 10.30 W m^{-2}, MAE 8.19 and bias of -1.06 W m^{-2}.

Table 2.3 DSRALIVE vs Ameriflux Towers Measurements at Temporal Resolution and Sky Condition

Temporal Resolution	**Sky Condition**	**R²**	**RMSE**
Half hourly	clear	0.98	48.76
	cloudy	0.88	89.86
	all	0.92	86.12
Daily	all	0.99	10.30
Weekly	all	0.99	6.68

2.4.3 DSRALIVE performance

The relationship between half hourly DSRALIVE and SW-IN has an R^2 value of 0.98 for clear skies and 0.88 for cloudy skies, with little bias as a function of local (standard) time (Figure 2.5). Bias for cloudy skies is larger, especially in the early morning.

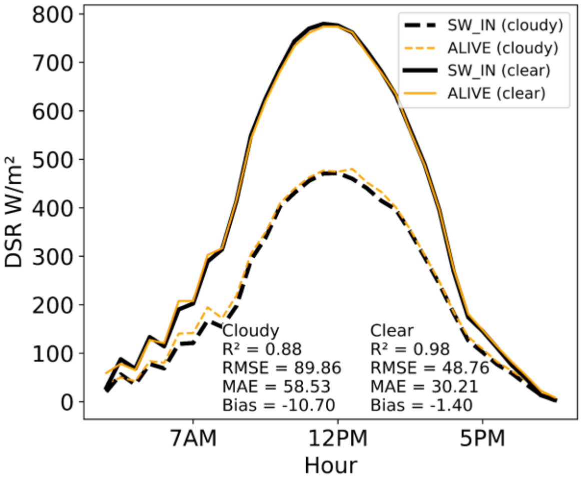

Figure 2.5. Temporal performance of the GBR model at half-hourly intervals throughout the day under clear and cloudy conditions.

Mean DSR_{ALIVE} largely matches SW-IN across the range of SZA from 11 to 88 degrees (Figure 2.6) but slightly overestimates SW-IN at low SZA, especially in cloudy conditions (Figure 2.6d), noting that less training data are available at low SZA across CONUS (Figure 2.2).

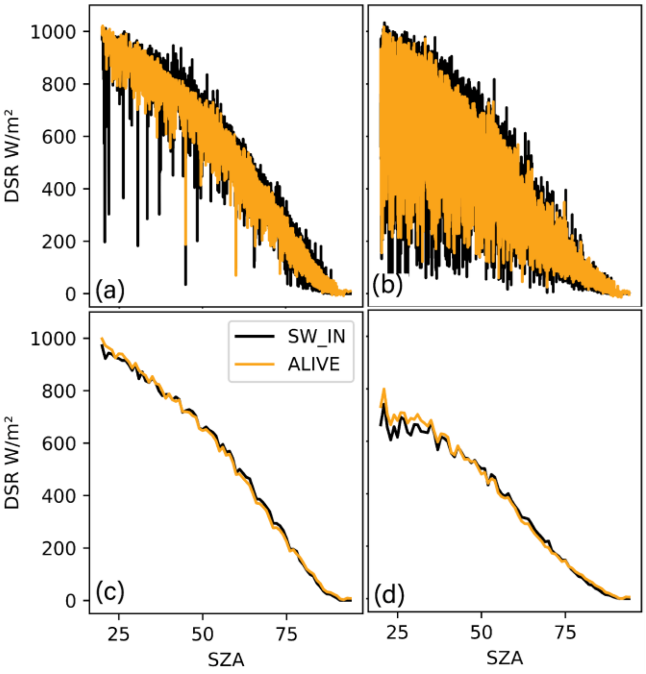

Figure 2.6. Performance of the GBR model across varying SZAs on clear (a and c) and cloudy

(b and d) days. SZA increments are 1-degree for panels c and d and 0.001 degree for panels a and b.

2.4.4 Feature analysis

We conducted SHAP analyses separately for clear and cloudy days and studied the top eight most impactful features on the GBR model output for DSR predictions. SZA with the SHAP value of 0.79 is most important, and CMI-C01 at wavelength of 0.47 µm (with SHAP value of 0.12) and SAA (with SHAP value of 0.04) consistently rank among the top three most important features for prediction (Figure 2.7). These features exhibit a positive correlation with the model prediction, where higher values result in higher DSR. Following these, CMI-C06 at wavelength of 2.2 µm, CMI-C15 at 12.3 µm, and CMI-C12 at 9.6 µm emerge as the next most important features for clear skies (Figure 2.7a), while for cloudy skies (Figure 2.6b), the order of importance shifts to CMI-C15, CMI-C12, then CMI-C06. CMI-C15 displays a negative correlation impact on the model output, and CMI-C02 and CMI-C04 exhibit similar but smaller impacts as CMI-C15.

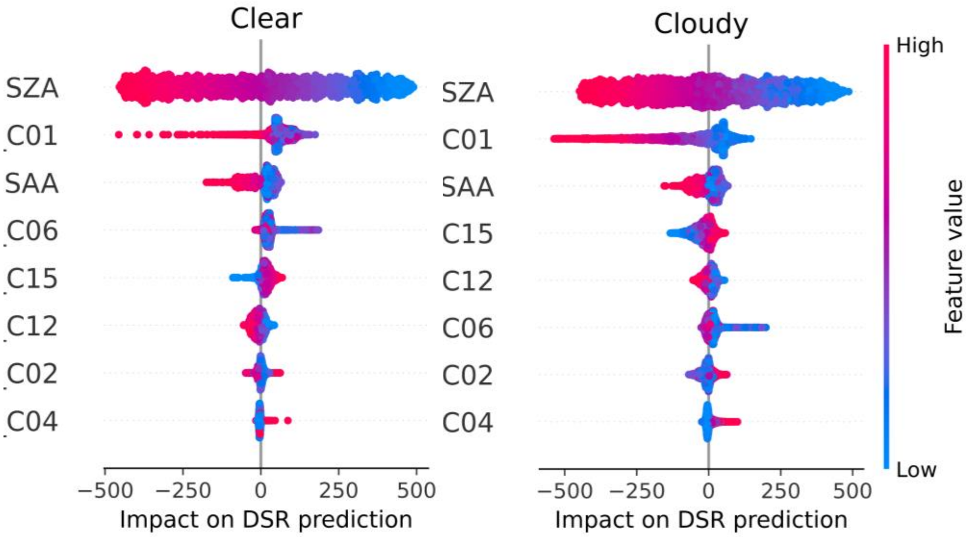

Figure 2.7. SHAP analysis for relative feature importance for ML model inputs to predict DSR under clear (left) and cloudy (right) conditions.

Our findings revealed a modest increase of 0.02 in the average R^2 when adding AOD to the GBR model. However, the standard deviation of R^2 values at a specific hour across all DOY increased by 0.05. Furthermore, 58% of our data was lost due to missing AOD pixel values. We expect there to be similar proportions of missing AOD data in the future because the ABI AOD algorithm has not changed. Using AOD as a predictor would therefore prevent DSR prediction at many timestamps, creating the same DSR data outages that we are aiming to prevent.

2.4.5 DSR_{ALIVE} vs DSR_{ABI} vs SW-IN: Assessment at discrete hours across CONUS

Figure 2.8 presents an analysis of the GBR model performance trained on data from 2021 and tested on 2022 data at a local standard time. By focusing on a specific hour on different DOY in 2022, from 99 sites spatially distributed across CONUS, we aimed to reduce the impact of time-based variations across the distributed sites. Each box plot in Figure 2.8 represents the distribution of R^2 values obtained for a particular hour across all 365 DOYs in 2022, excluding periods without at least 80 sites in each computation.

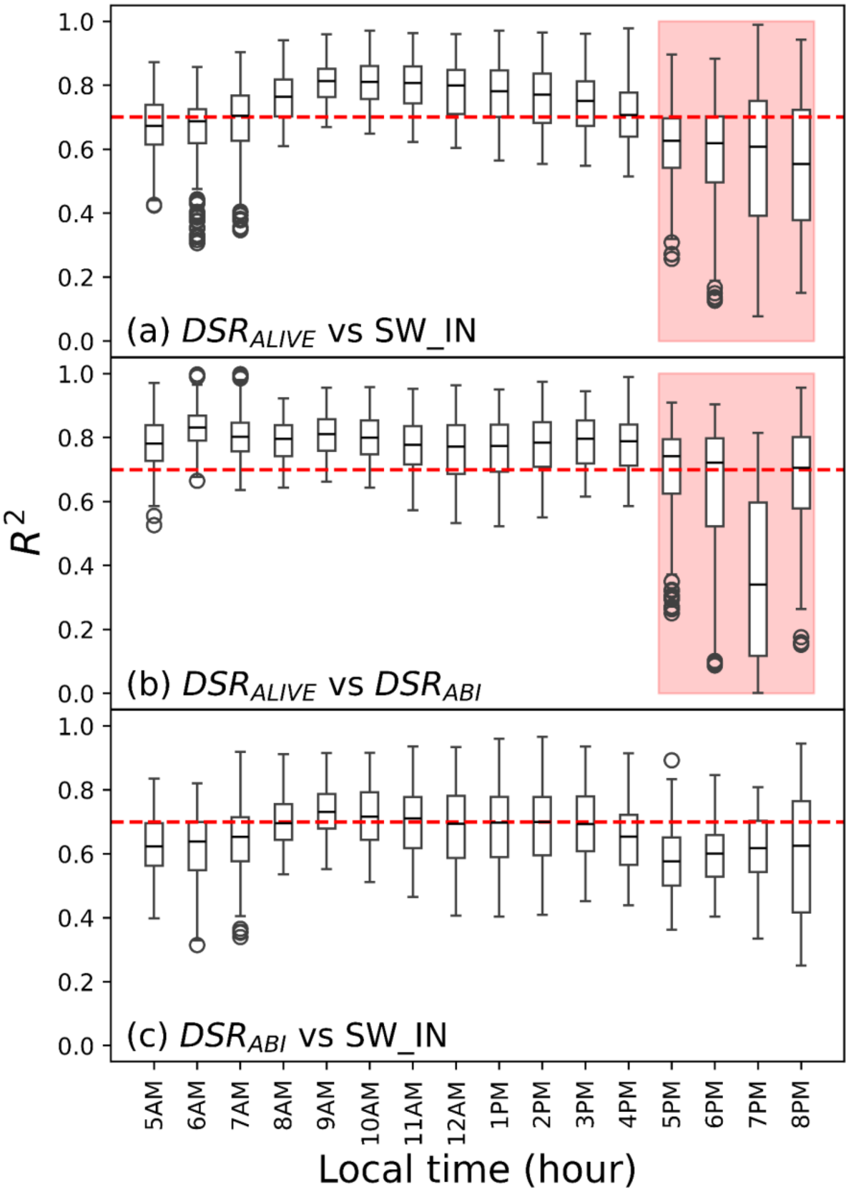

Figure 2.8. The R^2 values obtained for a specific hour (local standard time) across all DOY in 2022 using all the 99 sites distributed across CONUS. Plots present DSR_{ALIVE} predictions vs (a) tower SW-IN measurements and (b) the DSR_{ABI} product, and (c) the DSR_{ABI} product vs SW-IN measurements. The red dashed lines show an R^2 of 0.70. The red box signifies hours DSR_{ALIVE} showed lower performance than other hours.

DSR_{ALIVE} demonstrates higher R^2 values compared to DSR_{ABI} when compared with SW-IN throughout most hours, except for 6 - 8 PM local standard time (red box in Figure 2.8), during which SZAs are characteristically low across the ABI CONUS scene (Figure 2.2). DSR_{ALIVE} and DSR_{ABI} deviate from SW-IN measurements after 5 PM (Figure 2.8a and b), and the mean R^2

between SW-IN and both DSR$_{\text{ALIVE}}$ and DSR$_{\text{ABI}}$ is less than 0.7 from 5 and 6 AM. DSR averages less than 100 W m^{-2} during these times (Figure 2.5).

Figure 2.9 presents the CONUS maps for estimated DSR$_{\text{ALIVE}}$ and DSR$_{\text{ABI}}$ at 6PM UTC on Sept 21, 2022, providing a comparison extending beyond specific site locations. The analysis yielded R^2 of 0.83, RMSE of 132.15 W m^{-2}, and bias (DSR$_{\text{ABI}}$-DSR$_{\text{ALIVE}}$) of 85.86 W m^{-2}, notable differences between the two products across space and with respect to cloud characteristics revealed by storm systems. (Tropical Storm Ophelia can be seen on the right hand side of the imagery.)

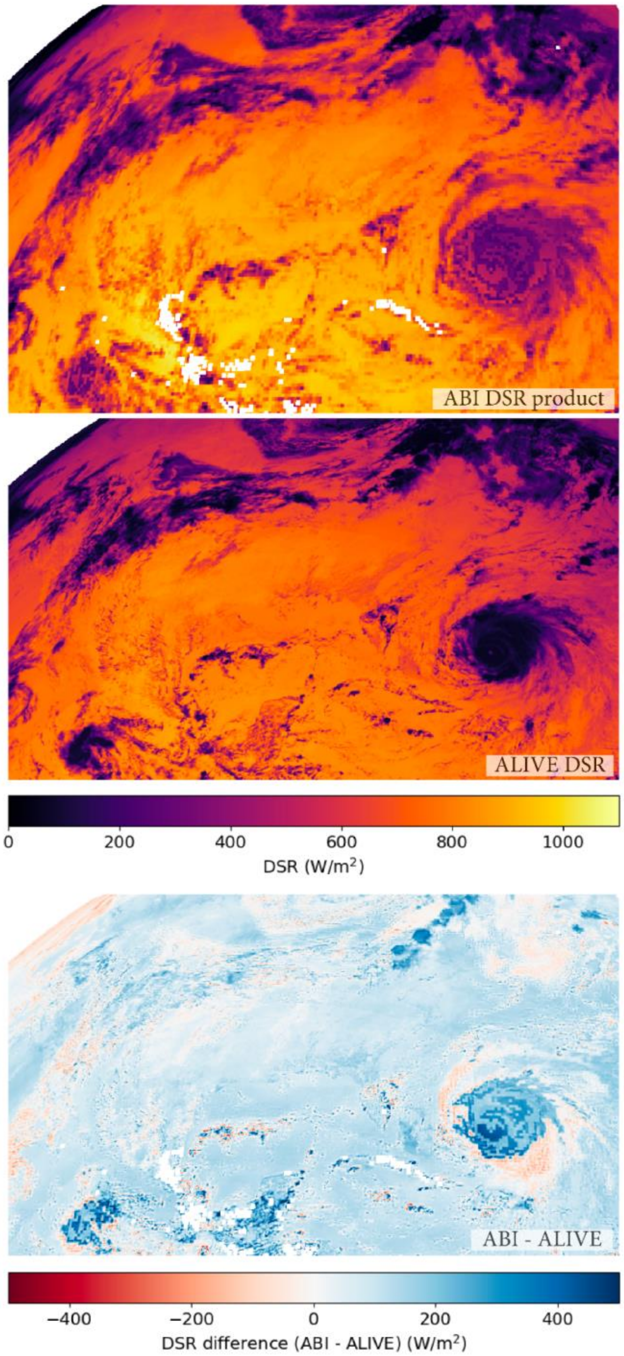

Figure 2.9. $\mathrm{DSR}_{\mathrm{ABI}}$, estimated $\mathrm{DSR}_{\mathrm{ALIVE}}$, and their differences (ABI-ALIVE) maps at hour 6PM UTC on Sep 21, 2022, with a CONUS view from the GOES satellite.

Figure 2.10 compares the model performance at Ameriflux sites (Figure 2.10a) with the full images (Figure 2.10b) from Figure 2.9 using kernel density estimates. SW-IN and $\mathrm{DSR}_{\mathrm{ALIVE}}$ exhibit closer correspondence compared to their relationship with $\mathrm{DSR}_{\mathrm{ABI}}$ (Figures 10a). $\mathrm{DSR}_{\mathrm{ABI}}$

typically shows higher estimates compared to the DSR$_{ALIVE}$ (Figures 9 and 8b). Figures 10a and b depict density plots related to Sep 21, 2022, at tower sites, where DSR$_{ALIVE}$ and DSR$_{ABI}$ maps are available (see Figure 2.9). However, there is an observed bias of 85.86 W m^{-2} between DSR$_{ALIVE}$ and DSR$_{ABI}$ maps which is evident in both the density plot (Figure 2.10b) and the difference between the two maps, DSR$_{ABI}$ - DSR$_{ALIVE}$ (Figure 2.9 (ABI-ALIVE)).

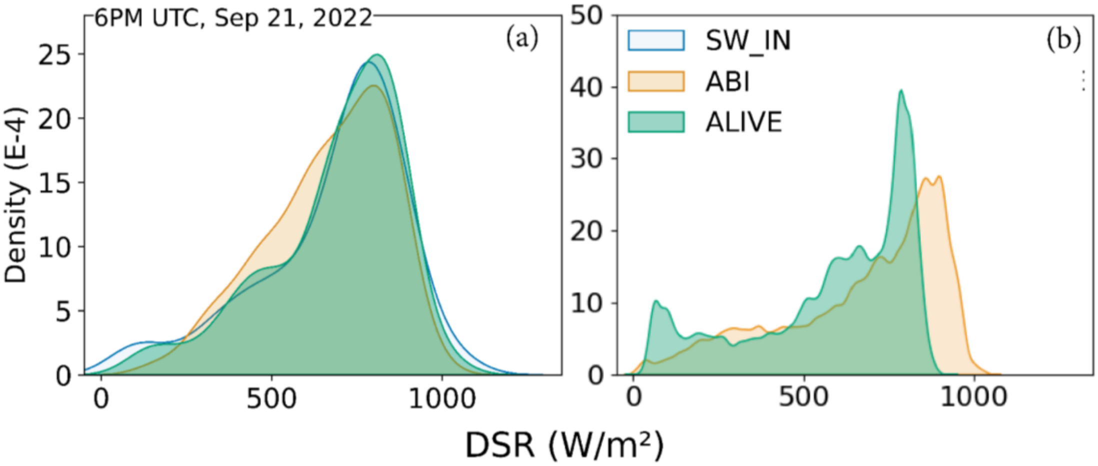

Figure 2.10. Comparison of DSR distribution using kernel density estimate: SW-IN vs DSR$_{ABI}$ vs DSR$_{ALIVE}$ at 6 PM UTC on Sept 21, 2022 at (a) Ameriflux sites and (b) across CONUS from maps presented in Figure 2.9.

2.4.6 DSR$_{ALIVE}$ performance at 5 minutes frequency vs DSR$_{SURFRAD}$

We exclusively utilized DSR$_{SURFRAD}$ for testing DSR$_{ALIVE}$; none were used in training. We selected the first week of July 2022 for time series comparison (Figure 2.11) and assessed DSR$_{ALIVE}$ performance across time intervals of 5, 15, and 30 minutes, as well as one and two-hours (Table 2.4). Figure 2.11 illustrates significant fluctuations in DSR$_{SURFRAD}$ at most sites, indicating the potential influence of aerosols or clouds (Ma et al., 2020; Zhang et al., 2021b). Analysis of AOD plots on the NOAA website, accessible at (National Oceanic and Atmospheric Administration, 2022), further demonstrates substantial fluctuations in AOD values at these times and locations.

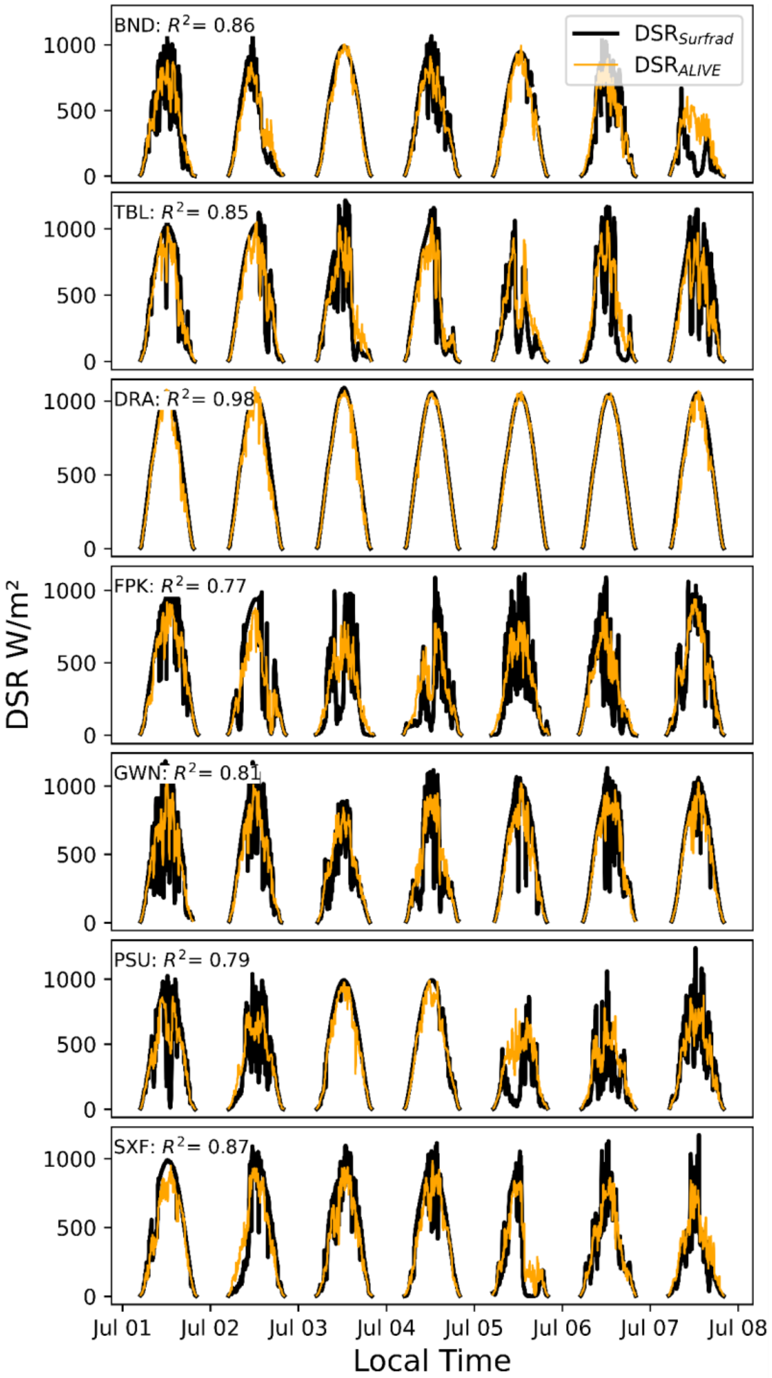

Figure 2.11. Time series of DSR_{ALIVE} vs. DSR_{SURFRAD} at 7 sites with a 5-minute frequency during the first week of July 2022. Sites abbreviations and R^2 between DSR_{ALIVE} and DSR_{SURFRAD} measurements are noted.

Figure 11 indicates that the site DRA exhibits the highest agreement with an R^2 value of 0.98, followed closely by SXL, BND, and TBL, each with R^2 values exceeding 0.85. In contrast, FPK and PSU display lower R^2 values of 0.77 and 0.79, respectively.

Table 2.4 presents the accuracy of DSR_{ALIVE} vs DSR_{SURFRAD} across temporal intervals ranging from 5 minutes to 2 hours. As temporal resolution increases, there is a corresponding improvement in model performance, demonstrated by higher R^2 values and reduced error metrics such as RMSE, MAE, and Bias. However, one may consider R^2 values of 0.85 at a 5-minute temporal resolution to be promising for hypertemporal analysis, effectively capturing short-term variations in DSR, noting that variability at shorter temporal scales, due to sub-pixel clouds and/or fast moving clouds, will not be able to be resolved by the ABI.

Table 2.4. DSR_{ALIVE} vs SURFRAD Measurements

	5 Minutes	**15 Minutes**	**30 Minutes**	**1 Hour**	**2 Hours**
R^2	0.85	0.86	0.88	0.90	0.91
RMSE	121.82	108.83	103.02	96.85	92.65
MAE	84.32	74.1	71.85	69.14	65.65
Bias	8.02	6.12	6.23	6.1	6.41

2.5 Discussion

2.5.1 ML performance and efficiency

Our study highlights the importance of selecting efficient ML models for predicting high frequency DSR maps from ABI imagery. GBR emerged as the optimal choice, offering similar accuracy to LSTM but with substantially faster prediction time, crucial for real-time applications (Liang et al., 2023; Liu et al., 2021). Through optimization experiments, we identified a GBR model configuration that balances prediction time and accuracy, making it well-suited for generating ‘hypertemporal’ DSR maps for our objective.

The findings from the GBR model efficiency experiment suggest that models trained with a larger volume of training data per site (2000 samples compared to 200, 500, or 1000) enhance the balance between prediction time and accuracy, thereby increasing overall model efficiency. Although it is well-known that a larger sample size improves ML performance (Maxwell et al., 2018), our study shows that it also increases efficiency in terms of prediction time. However, 2,000 samples per site appears to be the saturation point for efficiency (Figure 2.3). Increasing the number of samples beyond this point increases training time with little improvement in fit. Furthermore, we observed that models with higher complexity (m_depth) and lower ensemble size (n_estimators) achieve optimization in both computational efficiency and predictive accuracy (see Figure 2.3).

2.5.2 SHAP analysis

SZA emerges as the most critical predictor for modeling DSR, as widely demonstrated in previous studies (Hao et al., 2019; R. Li et al., 2023; Ma et al., 2020; Zhang et al., 2021b). Following SZA, CMI-C01 and SAA are significant predictors (Figure 2.7). These features demonstrate positive correlations with DSR, where higher values indicate higher DSR. CMI-C01, which measures blue visible reflectance at 0.47 µm; its importance is consistent with Rayleigh scattering (Hao et al., 2019; Paxton et al., 2021; Rayleigh, 1899; Tan et al., 2020). SAA improved time alignment of ABI observations with SW-IN measurements acquired at towers (see Figure 2.5). This improvement addresses the timing issue previously described in detail by (Losos et al., 2024a); in brief, the Ameriflux SW-IN observations represent an average between observations made at the top of the hour and half past the hour (or vice-versa) such that 15 and 45 minutes past the hour represents the average time over which observations are averaged. Time alignment was poorer before including SAA (results not shown).

Accurately estimating DSR presents a challenge influenced by factors beyond just cloud cover. Research by (Chen et al., 2012; Iqbal, 2012; Knapp et al., 2005) underscores the significant impact of cloud properties—like type, thickness, and composition—on solar radiation reaching the Earth's surface. Additionally, AOD, as described by (Hao et al., 2019; Iqbal, 2012; Knapp et al., 2005; H. Zhang et al., 2020; Zhang et al., 2021a) in measuring atmospheric particle concentration, also affects solar radiation attenuation. Although (Hao et al., 2019; Hou et al., 2020; Ma et al., 2020) showed the significance of using AOD in DSR modeling, incorporating AOD into the model results in only a modest increase in R^2 (0.02) for temporal patterns, but decreases R^2 between spatially distributed sites at specific hours, indicating prediction instability. Furthermore, the significant loss of data due to missing AOD values (approximately 58%) raises concerns about the reliability of the final DSR map at 5-minute frequency. In addition to AOD, many studies (Chen et al., 2012; Hao et al., 2019; Li et al., 2015; Ma et al., 2020) show that cloud optical depth (COD), and cloud fraction (CF) also play crucial roles in DSR/PAR modeling. Here we assume that the ML model incorporates observations from the ABI TOA product (CMIs) for representing the impacts of AOD, COD, and CF in DSR modeling. Previous research indicates that CMI observations, particularly the blue band, contain valuable information regarding AOD levels. In 2022, Yoojin in (Kang et al., 2022) utilized GOCI geostationary satellite data to improve retrievals of AOD using a light GBR (LGBM) model. They achieved R^2 of 0.92 by showing CH01 (the GOCI blue band at 0.412 μm) as the most informative feature. Remote sensing, combined with ML, offers an effective tool for retrieving cloud microphysical parameters such as COD and CF (Minnis et al., 2016; Yang et al., 2022).

The comparison between DSR_{ABI} and DSR_{ALIVE} maps (see Figure 2.9) reveals notable disparities, particularly in regions with high cloud cover, such as during hurricanes (Tropical Storm Ophelia is depicted on the southeast corner of Figure 2.9). It is important to note that DSR_{ABI} products may not be a perfect reference for validation (with absolute error of up to 125 W m^{-2} (Laszlo et al., 2020b)), especially in areas with frequent cloud cover or over oceans due to the lack of ground stations. Incorporating additional variables related to clouds, such as COD and/or CF, may enhance performance. However, relying on additional datasets would make DSR_{ALIVE} more vulnerable to data outages. Therefore, we have concluded that maintaining the simplicity of our algorithm is advisable, especially considering its effective performance compared to measurements of SW-IN measurements from towers.

2.5.3 Data sources and uncertainty: Ameriflux vs SURFRAD

SURFRAD DSR measurements have lower uncertainty (±2% to ±5%), compared to AmeriFlux's stated ±10% uncertainty for net radiation observations (Augustine et al., 2000). However, AmeriFlux implements the AmeriFlux quality control (QC) method to ensure data quality (Chu et al., 2023; Pastorello et al., 2020, 2014) and employs a multi-step QC process to guarantee the integrity of eddy covariance and micrometeorological data (Aubinet et al., 2012; Pastorello et al., 2014). Following initial processing by tower teams, data undergo rigorous assessment encompassing individual variable inspection, evaluation of correlations between coupled variables (e.g., solar radiation and photosynthetically active radiation), analysis of temporal patterns, and in-depth examinations of interrelated variable variations (Aubinet et al., 2012; Chu et al., 2023).

Interestingly, despite the difference in uncertainty, our research showed good agreement (R^2 of 0.85 to 0.91) between models trained on AmeriFlux data and DSR measurements from the SURFRAD network. This agreement held even at 5-minute intervals, even though the models were trained on half-hourly data (Table 2.4). This suggests that AmeriFlux SW-IN data, despite higher reported uncertainty, can be valuable for modeling purposes.

2.5.4 $\mathrm{DSR}_{\mathrm{ALIVE}}$ performance against SW-IN, $\mathrm{DSR}_{\mathrm{ABI}}$ and $\mathrm{DSR}_{\mathrm{SURFRAD}}$

We observed that $\mathrm{DSR}_{\mathrm{ALIVE}}$ generally outperformed $\mathrm{DSR}_{\mathrm{ABI}}$ when compared with SW-IN measurements, except during specific evening hours characterized by high SZAs (Figure 2.8). In fact, both models performed more poorly compared to SW-IN measurements after 5 PM, suggesting limitations in ABI to capture solar radiation dynamics during certain times of the day with high SZAs. DSR maps across CONUS in Figure 2.9 revealed notable spatial differences between $\mathrm{DSR}_{\mathrm{ALIVE}}$ and $\mathrm{DSR}_{\mathrm{ABI}}$ products, particularly regarding their response to cloud characteristics. While $\mathrm{DSR}_{\mathrm{ALIVE}}$ showed promising performance under clear sky conditions for data collected at the Ameriflux sites, differences persisted especially during periods of low solar radiation and in the presence of clouds.

Our findings demonstrate that DSR_{ALIVE} exhibits promising performance in estimating DSR at high temporal resolutions (5-minute intervals) compared to ground-based measurements from the SURFRAD network (Table 2.4). This is particularly significant because, to our knowledge, this is the first study to explore DSR mapping at such fine time scales. Traditionally, DSR retrievals have been limited to coarser temporal resolutions (half-hourly or hourly) due to constraints imposed by satellite data availability and/or processing techniques (Chen et al., 2012; Hao et al., 2019; Ma et al., 2020; Zhang et al., 2021a; Y. Zhang et al., 2020). The ability of DSR_{ALIVE} to capture observed short-term variations holds significant value for various applications. For instance, monitoring rapid phenomena like heatwaves (Jahan et al., 2023; X. Li et al., 2023), wildfires (Massetti et al., 2019; Mohapatra and Trinh, 2022; Xu and Zhong, 2017), water stresses and flash droughts (Hu et al., 2020; Otkin et al., 2018), and other fast-changing environmental events often necessitates high-frequency data. Similarly, studies investigating the carbon cycle or other processes sensitive to short-term fluctuations in solar radiation benefit from high-resolution DSR data (S. Chen et al., 2023; Jeong et al., 2023b; A. M. Khan et al., 2022; Ranjbar et al., 2023a). Our work paves the way for utilizing DSR_{ALIVE} as a tool to analyze these critical Earth system processes at unprecedented temporal detail.

It is important to note that while SURFRAD provides valuable ground-truth data, its network primarily focuses on mid-latitude regions (Figure 2.2). This poses a limitation for validating models like DSR_{ALIVE}, which might struggle to simulate DSR accurately at lower latitudes with lower solar zenith angles (e.g. Figure 2.6). These limitations in the SURFRAD network highlight the need for complementary validation strategies, potentially including future ground stations at lower latitude sites, to ensure the robustness of $\mathrm{DSR}_{\mathrm{ALIVE}}$ and similar approaches across more geographical regions.

2.5.5 Comparing against recent studies on DSR

At half-hourly resolution, our GBR models showed impressive performance metrics against Ameriflux sites (Table 2.3). Under clear sky conditions, our models achieve an R^2 of 0.98 with an RMSE of 48.76 W/m^2, surpassing the performance of DenseNET and CNNGRUnor architectures (R. Li et al., 2023). In cloudy sky conditions, although the R^2 slightly decreased to 0.88, the RMSE increased to only 89.86 W/m^2. Overall, our model's performance under all conditions at half-hourly resolution yields an R^2 of 0.92 and an RMSE of 86.12 W/m^2, indicating improved accuracy in comparison to recent studies (Chen et al., 2021; Letu et al., 2022; R. Li et al., 2023; Zhang et al., 2021b; Y. Zhang et al., 2020). For hourly DSR mapping, the GBR model demonstrates robust accuracy (see Figure 2.4) compared to previous studies. In (Zhang et al., 2021b), researchers achieved an R^2 of 0.90 using a combination of GOES-16 and MODIS data and employing a look-up table (LUT) at a 500 m spatial resolution. Additionally, in another study, researchers achieved an R^2 of 0.82 using a combination of DSCOVR-EPIC data and employing an RF model for hourly DSR mapping at a 10 km spatial resolution (Hao et al., 2019).

Shifting to performance on daily time scales, our model also outperforms compared to (Hao et al., 2019; Zhang et al., 2021a), who achieved an R^2 of 0.88 and 0.93 compared to our daily R^2 of 0.99 (see Table 2.3). Our GBR models exhibit exceptional accuracy with an R^2 of 0.99 and an RMSE of 10.30 W/m^2, outperforming the RF model from (Chen et al., 2021), which achieved an R^2 of 0.97 and an RMSE of 17.64 W/m^2. At a weekly resolution, our models maintain high accuracy, with an R^2 of 0.99 and an RMSE of 6.68 W/m^2. The robustness of our GBR models in terms of prediction time vs model performance for estimating DSR from GOES-R ABI data has significant implications for near-real-time carbon uptake mapping and solar energy applications.

When comparing $\text{DSR}_{\text{ALIVE}}$ against $\text{DSR}_{\text{SURFRAD}}$, high agreement ($R^2$ of 0.85) is achieved even at 5-minute frequency, although the GBR model was trained on half-hourly data from the AmeriFlux network. To the best of our knowledge, there is no other DSR product available at a 5-minute frequency for comparison.

2.6 Conclusion

We used public, quality-checked SW-IN ground measurements to train a ML model for predicting DSR at the native temporal and spatial scale of GOES-16 ABI CMI measurements. Our model outperformed the existing DSR_{ABI} algorithm under most solar conditions, with opportunities for improvement, especially at lower SZA and early morning/late evening periods. Additional surface SW-IN observations in subtropical and tropical regions for model training, and ongoing efforts to use ML to improve physically-based atmospheric attenuation models, would likewise advance DSR product development to better realize the extensive benefits of geostationary satellite observations.

Chapter 3. Near Real-time Mapping of All-Sky Land Surface Temperature from GOES-R Using Machine Learning

This chapter is under review in the Wiley-Journal of Geophysical Research: Machine Learning and Computation, see the preprint here: Sadegh Ranjbar, Danielle Losos, Sophie Hoffman, et al. Near Real-time Mapping of All-Sky Land Surface Temperature from GOES-R using Machine Learning. ESS Open Archive . October 04, 2024. DOI: 10.22541/essoar.172801403.30077549/v1

Abstract

Land surface temperature (LST) is crucial for understanding earth system processes. We expanded the Advanced Baseline Imager Live Imaging of Vegetated Ecosystems (ALIVE) framework to estimate LST in near-real-time for both cloudy and clear sky conditions at a five-minute resolution. We compared two machine learning (ML) models, Long Short-Term Memory (LSTM) networks and Gradient Boosting Regressor (GBR), using top-of-atmosphere observations from the Advanced Baseline Imager (ABI) on the GOES-16 satellite against observations from hundreds of observation sites for a five-year period. LSTM outperformed GBR, especially at coarser resolutions and under challenging conditions, with a clear sky R^2 of 0.96 (RMSE 2.31 K) and a cloudy sky R^2 of 0.83 (RMSE 4.10 K) across CONUS, based on 10-repeat Leave-One-Out Cross-Validation (LOOCV). GBR maintained high accuracy and ran 5.3 times faster, with only a 0.01-0.02 R^2 drop. Feature importance revealed infrared bands were key in both models, with LSTM adapting dynamically to atmospheric changes, while GBR utilized more time information in cloudy conditions. A comparative analysis against the physically-based ABI_{LST} product showed strong agreement in winter, particularly under clear sky conditions, while also highlighting the challenges of summer LST estimation due to increased thermal variability. This study underscores the strengths and limitations of data-driven models for LST estimation and suggests potential pathways for integrating ML models to enhance the accuracy and coverage of LST products.

3.1 Introduction

Land surface temperature (LST) is a critical variable recognized by the Global Climate Observing System for its essential role in various land surface processes and land–atmosphere interactions, including biophysical, hydrological, and ecological function (Firozjaei et al., 2022; Li et al., 2023; Li et al., 2013). Obtaining LST data at sub-daily resolutions is critical for understanding terrestrial ecosystem function that varies dynamically in response to its thermodynamic environment (Jia et al., 2022; Liu et al., 2023; Yan et al., 2023). High frequency LST data enable detailed monitoring of rapid ecosystem changes, such as dynamic carbon uptake processes, ecosystem respiration, and evapotranspiration, all of which are sensitive to temperature fluctuations (Chen & Liu, 2020; Jia et al., 2020; Li et al., 2023; Wang et al., 2024). Precise LST measurements help assess the carbon and water balance of ecosystems, providing insights into how temperature, water availability, and plant physiological processes interact over time (Firozjaei et al., 2022; Li et al., 2013; Mavrovic et al., 2023). This understanding is crucial for predicting ecosystem responses to climatic changes and managing water resources effectively.

Satellite remote sensing can derive LST with high temporal and spatial resolution on continental and global scales (Firozjaei et al., 2022; Li et al., 2023; Li et al., 2013; Zarei et al., 2021; Zhan et al., 2013). Geostationary satellites can provide hemispheric scan coverage on the order of minutes, offering sufficient data for subdaily LST retrievals (Jia et al., 2022, 2024; Liu et al., 2023; Yan et al., 2023). While algorithms for estimating LST under clear sky conditions using thermal infrared (TIR) data are well established (Li et al., 2013; Zarei et al., 2021; Zhan et al., 2013), the inability of TIR signals to penetrate thick clouds presents a significant challenge (Jia et al., 2022; Liu et al., 2023; Weng, 2009; Yan et al., 2023; Zarei et al., 2021; Zhan et al., 2013). This limitation results in substantial data gaps, especially during cloud cover, which typically obscures 67% of the Earth's surface (Wang et al., 2020). Several recent studies have leveraged geostationary satellites for estimating LST under cloudy conditions by integrating multi-source and/or hypertemporal observations. For example, Jia et al., (2023) produced a global hourly five-km all-sky LST dataset using a refined SEB-based cloudy sky LST recovery method, demonstrating improved accuracy over existing clear sky products. Similarly, Ding et al., (2022) developed the RTG method, which integrates reanalysis data with geostationary TIR observations to reconstruct hourly all-weather LST, achieving high reliability over the Tibetan Plateau. Jia et al., (2022) introduced a 2-km, all-sky, hourly LST product from ABI data, combining SEB modeling with cloud correction techniques to enhance retrieval accuracy across diurnal cycles. Additional studies by Zhou et al., (2022) and Liu et al., (2023) further highlight the role of geostationary satellite data fusion in minimizing cloud-related gaps. Hyper-temporal geostationary observations can be used to enhance under-cloud LST estimation by enabling continuous monitoring, data fusion, and diurnal cycle modeling. Unlike polar-orbiters, they provide high-frequency LST datasets with fewer cloud-induced gaps, benefiting climate monitoring, hydrology, and extreme weather analysis.

Moreover, many efforts have focused on estimating LSTs under all-sky conditions using different methods like surface energy balance (SEB), data fusion, and similar pixel techniques. Recently, machine learning (ML) algorithms have gained prominence, attracting growing interest from the geoscience community for their potential in improving all-sky LST estimates (Chen et al., 2012; Cho et al., 2022; Jia et al., 2021, 2024; Li et al., 2021; Liu et al., 2023; Tang et al., 2023; Zhang et al., 2024). Relationships between LST and remotely sensed variables under all-sky conditions can be empirically modeled using machine learning algorithms (Ranjbar et al., 2021; Varghese et al., 2023; M. Zheng et al., 2022). Various algorithms have been employed effectively for LST estimation, such as neural networks (NNs), including recurrent and convolutional NNs (RNNs and

CNNs) and tree-based ensemble learning models (Cho et al., 2022; Duan et al., 2023; Fu et al., 2022; Wu et al., 2022; Zhang et al., 2024; Zhang et al., 2023; Zheng et al., 2022). For instance, Shwetha & Kumar (2016) utilized NNs to predict high-resolution LST using microwave polarization data, achieving a correlation coefficient (r) up to 0.96 and RMSE as low as 1.7 K under clear skies. Under cloudy conditions, the model performed well with RMSE values lower than those observed in clear skies. Similarly, Wu et al., (2022) developed a two-step deep learning framework combining thermal infrared and passive microwave data to create gapless all-weather LST over China's landmass, yielding RMSEs between 1.71 and 3.87 K, with R^2 values ranging from 0.88 to 0.98 across different sky conditions. Cho et al., (2022) used a LightGBM model to reconstruct 1 km MODIS LST in South Korea, achieving an R^2 of 0.89–0.91 during the day and 0.96–0.97 at night for clear sky conditions; under cloudy conditions, the model maintained an R^2 of 0.55–0.63 during the day and 0.70–0.74 at night. Li et al., (2021) used Random Forest models incorporating MODIS, reanalysis, and in situ data to estimate all-sky LST across the continental United States, achieving R^2 of 0.943 to 0.963 and RMSE around 2.76 K. Additionally, Fu et al., (2022) leveraged a 3D-CNN model to reconstruct high spatiotemporal LSTs in urban areas, reducing RMSE to 0.61 K. These studies demonstrate the capability of machine learning models to generate accurate and gapless all-weather LST products using efficient models like tree based ensemble learning and neural NNs (Gong et al., 2023; Li et al., 2021; Wu et al., 2022).

Despite their widespread use, these machine learning methods have limitations, such as a lack of explanation for physical mechanisms and issues with generalizability (Zhang et al., 2024). Zhang et al. (2024) proposed a two-step framework for all-sky LST estimation that combines physical models with gradient boosting, achieving an RMSE of 2.67 K for clear skies and 2.60 K for cloudy skies, both with R^2 of 0.96. However, the method is not real-time, and does not utilize geostationary satellite observations for sub-daily inference. Most studies have been limited to small areas and short time periods, with minimal focus on real-time or continental to global-scale applications. Additionally, validation is often done with limited data, raising concerns about broader applicability (Fu et al., 2022; Gong et al., 2023; Shwetha and Kumar, 2016; Zhao and Duan, 2020). In this study, we demonstrate how machine learning can be used to estimate LST at a five-minute temporal resolution across the Contiguous United States (CONUS) and evaluate the model against ground observations from more than 100 AmeriFlux and Surface Radiation Budget Network (SURFRAD) sites. We apply SHapley Additive exPlanations (SHAP) analysis (Lundberg and Lee, 2017) to uncover individual feature contributions by quantifying the direction and magnitude of their influence, identifying synergistic or antagonistic relationships that affect model behavior (Lundberg et al., 2020; Lundberg and Lee, 2017; Temenos et al., 2023).

The Advanced Baseline Imager Live Imaging of Vegetated Ecosystems (ALIVE,) framework develops near-real-time estimates of terrestrial carbon dioxide and water flux from a geostationary platform ((Losos, Ranjbar, et al., 2024a; Ranjbar et al., 2024b; Ranjbar, et al., 2024c)), which this study aims to extend to LST estimation (here called ‘ALIVE$_{\text{LST}}$’). We use top-of-atmosphere observations from the Advanced Baseline Imager (ABI) sensor onboard the Geostationary Operational Environmental Satellite-R Series (GOES-R). This research seeks to generate high-frequency CONUS LST maps, with temporal resolution of up to 5 minutes, under various sky conditions. To achieve this, we will utilize two well-established machine learning models: Long Short-Term Memory (LSTM) networks and Gradient Boosting Regression (GBR) models. We explore these approaches because GBR consistently outperforms other tested methods for downwelling shortwave radiation estimation (Ranjbar, et al., 2024b) and surface-atmosphere carbon dioxide fluxes (Ranjbar, et al., 2024c), but LSTM may better simulate the time-varying nature of LST that includes lagged effects due to ecosystem heat capacity, and both feature in contemporary LST research (Jia et al., 2024). Our approach evaluates these models' performance across different sky conditions, seasons, and climatic regions, including comparisons against the operational GOES-R and MODIS LST products.

3.2 Materials and Methods

3.2.1 Overview

This study integrates tower-based micrometeorological observations and geostationary satellite imagery, with machine learning techniques to estimate and evaluate LST across diverse environmental conditions. The workflow begins with data collection from AmeriFlux, NEON, and SURFRAD networks (Figure 3.1, section 2.2) and ensuring temporal alignment with GOES-R ABI satellite observations at tower locations (section 2.4). LST is derived from ground-based radiative flux measurements at AmeriFlux, NEON, and SURFRAD network towers using the Stefan-Boltzmann equation (Section 2.3). Tower-based derived LST serves as both training data and an in-situ reference for estimating and validating satellite-derived LST products. Data from AmeriFlux and NEON are utilized for training and evaluating machine learning models (section 2.5), while SURFRAD network data is exclusively employed for validation purposes (section 2.6).

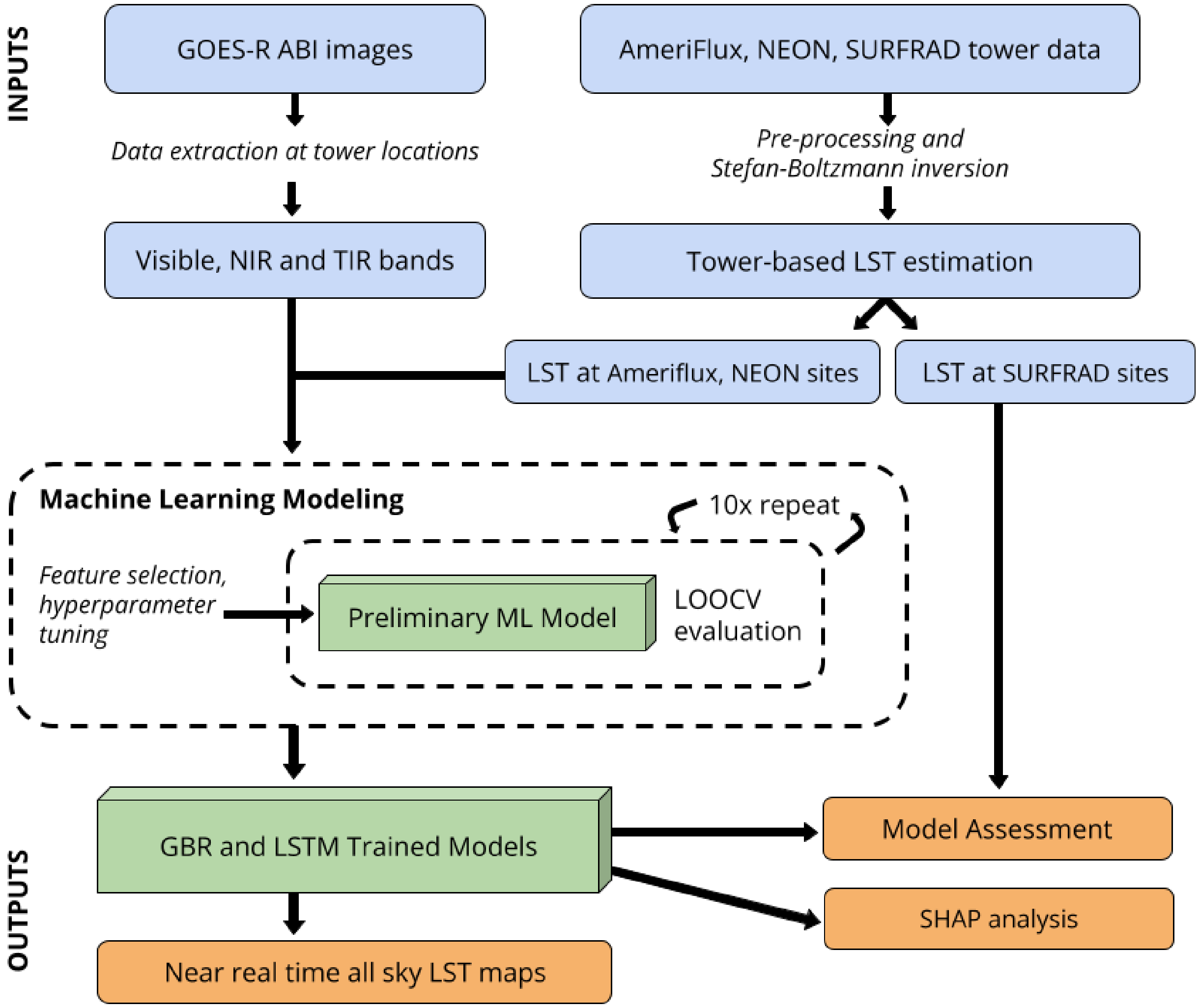

Figure 3.1. Schematic representation of the methodology used in this study, illustrating the integration of tower observations, GOES-R ABI satellite data, and machine learning-based LST estimation and validation.

3.2.2 Tower observations

We compiled micrometeorological data from AmeriFlux and NEON eddy covariance towers following the methodology described in Losos et al. (2024b). We selected sites with data available post-2019 to align with GOES-16 data availability (section 2.4), resulting in 101 sites distributed across CONUS (see Figure 3.2 for locations and Table A1 for details). Tower data availability relies on the efforts of individual tower-operating teams who provide publicly available data that undergo quality control checks by Ameriflux and NEON (Pastorello et al., 2020, 2014; Sturtevant et al., 2022). These data include half-hourly (or occasionally hourly) sums of carbon dioxide, water, sensible heat, meteorological, micrometeorological, and radiometric variables, along with other trace gas fluxes. All data undergo quality control via common algorithms. These files are updated shortly after new data are uploaded, though in practice, updates may be delayed by months to years after data collection as a result of quality control processes and time allocation.

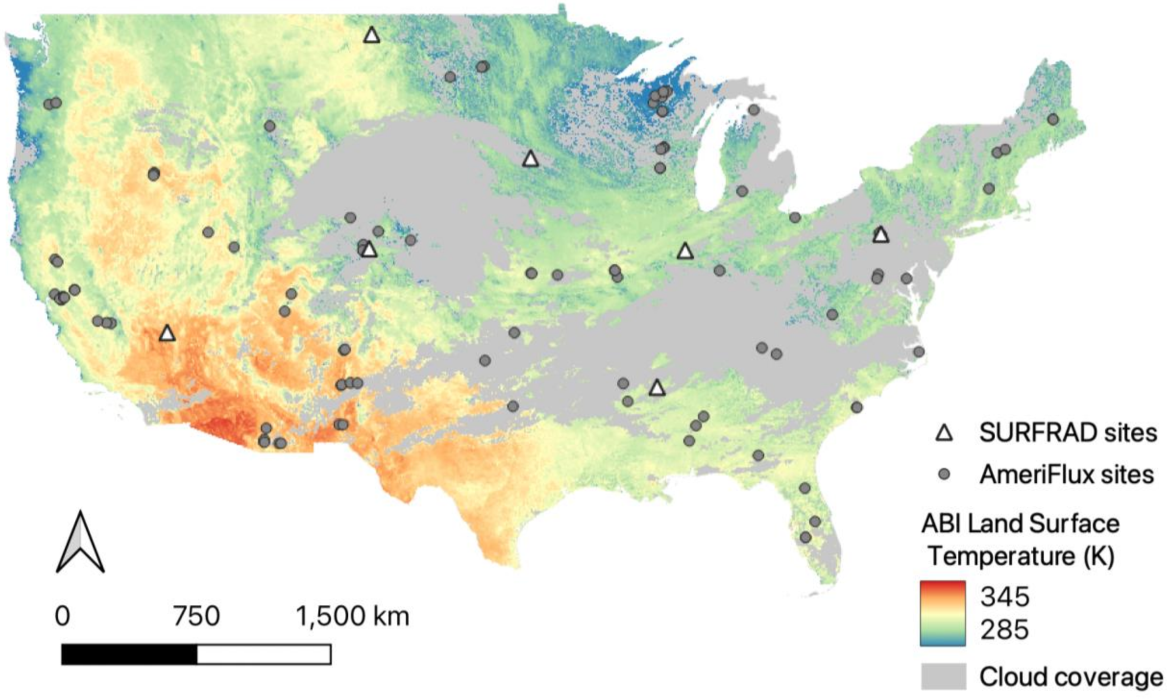

Figure 3.2. Locations of AmeriFlux, NEON, and SURFRAD sites used in this study overlaid on the GOES-R ABI LST product from July 20, 2023, at 18:02 UTC. Areas with no LST values, resulting from cloud cover, are displayed in gray.

We also utilized one-minute ground measurements from the SURFRAD network to evaluate the estimated LST at five-minute intervals, aligning with the temporal resolution of CONUS ABI observations. Established by NOAA in 1993, SURFRAD provides accurate, continuous, long-term surface radiation measurements across diverse climatic regions in the United States. For this analysis, data from seven stations, which collect independent observations of solar and infrared radiation along with meteorological parameters, were used (see Figure 3.2). To capture seasonal variations, we selected one week of data centered around January 15, April 15, July 15, and October 15, averaging one-minute SURFRAD observations to five-minute intervals for comparison.

3.2.3 LST calculations at tower locations

LST can be estimated from tower-measured longwave incoming radiation (LW_{IN}) and outgoing radiation (LW_{OUT}) while accounting for emissivity using Eq. 3.1, proposed by (Thakur et al., 2022). This approach is based on sites where incoming and outgoing broadband longwave radiation are measured, usually with four-component net radiometers (Verma et al., 2016), and through inverse application of the Stefan-Boltzmann (SB) law (Lhomme et al., 1988).

$$LST = \sqrt[4]{\frac{LW_{IN}}{\sigma} - \frac{LW_{IN}}{\varepsilon\sigma} + \frac{LW_{OUT}}{\varepsilon\sigma}} \qquad \text{Eq. 3.1}$$

Where ε is broadband emissivity and $\sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the Stefan-Boltzmann constant (Lhomme et al., 1988). For emissivity (ε), we used a classification based approach (Gillespie et al., 1998; Snyder et al., 1998). Theoretically, this method can accurately estimate the LST for properly classified surfaces with well-known emissivities (Ren et al., 2017; Snyder et al., 1998; Sobrino et al., 2016; Sobrino et al., 2008; Valor, 1996). We utilized the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) spectral library for calculating emissivity values, incorporating the International Geosphere-Biosphere Programme (IGBP) classification designated at each tower site. This approach involved using a lookup table, which assigns emissivity values based on specific IGBP land cover categories (Table 3.1).

Table 3.1 The emissivity values for the respective IGBP categories as derived from the ASTER spectral library.

IGBP Category	Wavelength: 10.25 - 10.95 µm	Wavelength: 10.95 - 11.65 µm

Evergreen Needleleaf Forest	0.971 ± 0.004	0.980 ± 0.003
Evergreen Broadleaf Forest	0.973 ± 0.003	0.982 ± 0.002
Deciduous Needleleaf Forest	0.969 ± 0.005	0.977 ± 0.004
Deciduous Broadleaf Forest	0.972 ± 0.003	0.981 ± 0.002
Mixed Forest	0.970 ± 0.004	0.979 ± 0.003
Closed Shrublands	0.967 ± 0.006	0.976 ± 0.004
Open Shrublands	0.965 ± 0.007	0.974 ± 0.005
Woody Savannas	0.968 ± 0.006	0.977 ± 0.004
Savannas	0.966 ± 0.007	0.975 ± 0.005
Grasslands	0.964 ± 0.007	0.973 ± 0.006
Permanent Wetlands	0.972 ± 0.004	0.981 ± 0.003
Croplands	0.965 ± 0.007	0.974 ± 0.005
Urban and Built-up	0.960 ± 0.008	0.969 ± 0.006
Cropland/Natural Vegetation Mosaic	0.966 ± 0.006	0.975 ± 0.004
Snow and Ice	0.983 ± 0.002	0.992 ± 0.001
Barren or Sparsely Vegetated	0.960 ± 0.008	0.969 ± 0.006

These emissivity values offer a detailed characterization of the land cover types, which is essential for accurate LST estimation under various environmental conditions. To compute the broadband emissivity, we utilized the mean value of the emissivity from both bands (Wilber, 1999).

3.2.4 ABI observations from GOES-R satellite

We leveraged top of atmosphere observations obtained by the ABI sensor onboard the GOES-16 satellite, operational since November 2017. ABI captures imagery across 16 spectral bands with a spatial resolution ranging from 0.5 to 2 km at nadir, depending on the band (Goodman et al., 2019). It provides full-disk observations every 10 minutes and CONUS observations every 5 minutes in its typical scan mode (Mode 6, Goodman et al., 2019; He et al., 2019). Our research synchronized ABI data, specifically the GOES-R ABI L1 Cloud and Moisture Imagery (CMI) products from

2019 to 2022, with ground-based eddy covariance measurement from 101 AmeriFlux and NEON sites as detailed in (Losos et al., 2024b) and added information from 7 SURFRAD sites. The CMI product measures reflectances and brightness temperatures at 16 bands spanning visible, near-infrared (NIR), shortwave infrared (SWIR) and infrared (IR) wavelengths (Table 3.2, Goodman et al., 2019; Khan et al., 2021). We used all CMI channels for model training and testing, as each band can provide unique information on surface thermal properties and/or cloud cover (Goodman et al., 2019; Ranjbar et al., 2024a; Schmit et al., 2017). We also anticipated that visible and NIR bands would impact LST prediction because they are related to vegetation greenness, which implies water availability and thereby differences in thermal capacitance.

Table 0.2 Summary of ABI bands and their primary uses.

ABI Band	Wavelength (microns)	Band Type	Primary Uses
C01	0.47	Visible	Monitoring aerosols (smoke, haze, dust) Air quality monitoring through aerosol optical depth measurements
C02	0.64	Visible	Daytime monitoring of clouds (0.5-km spatial resolution) Volcanic ash monitoring
C03	0.86	NIR	High contrast between water and land Assess land characteristics, including flooding, burn scars, and hail damage
C04	1.37	NIR	Thin cirrus detection during the day Volcanic ash monitoring
C05	1.6	NIR	Daytime snow, ice, and cloud discrimination Input to “Snow/Ice vs. Cloud” RGB
C06	2.24	NIR	Cloud particle size, snow, and cloud phase Hot spot detection at emission temperatures >600K
C07	3.9	IR	Low stratus and fog detection Fire/hot spot detection and volcanic ash
C08	6.2	IR	Upper-level feature detection (jet stream, waves)
C09	6.9	IR	Mid-level feature detection
C10	7.3	IR	Low-level feature detection (EML, fronts)
C11	8.4	IR	Cloud-top phase and type products Volcanic ash (SO_2 detection) and dust
C12	9.6	IR	Dynamics near the tropopause Input to Airmass RGB
C13	10.3	IR	Less sensitive to atmospheric moisture brightness temperatures warmer than IR
C14	11.2	IR	IR window Low stratus and fog detection
C15	12.3	IR	Greater sensitivity to moisture

			cooler brightness temperatures
C16	13.3	IR	Mean tropospheric air temperature estimation Highlights high, cold, likely icy clouds

To further evaluate and validate the performance of our machine learning models, we incorporated the physically-based GOES-R ABI LST product (hereafter 'ABI_{LST}') into our analysis. ABI_{LST} is generated using the split-window technique applied to ABI bands 14 (11.2 µm) and 15 (12.3 µm), providing LST estimates for land surfaces under clear sky conditions, with a local zenith angle of up to 85 degrees for both daytime and nighttime with a mission requirement of 2.5 K in precision and 2.3 K in accuracy (Laszlo et al., 2020b; Yu and Yu, 2020a). This product is generated hourly for the ABI Full Disk, the CONUS, and Mesoscale regions, with spatial resolutions of 2 km for CONUS view and 10 km for Full Disk (Yu and Yu, 2020a). We compared our empirical machine learning approach with the physically-based ABI_{LST} retrieval method to assess its consistency with standard satellite-derived products. The method employed by NOAA for creating ABI_{LST} is a well-established approach in the remote sensing community for retrieving land surface temperature from ABI observations (Duan et al., 2023; Li et al., 2021; Zhang et al., 2024).

We also utilized the NOAA Enterprise Cloud Mask, an advanced cloud detection scheme derived from the naive Bayesian cloud mask initially developed for the Pathfinder Atmospheres Extended (PATMOS-x) Cloud Properties Climate Data Record (Heidinger et al., 2020). The Enterprise Cloud Mask supports various sensors, including the ABI on GOES-16, -17, -18, and -19, VIIRS, and MODIS, and provides a four-level cloud mask and cloud probability (Heidinger et al., 2020; Lin et al., 2021). This allows for precise separation of cloud-affected and cloud-free pixels, enhancing the accuracy of LST and sea surface temperature estimations. For our analysis, we used the "cloudy" and "clear" classifications, treating "probably clear" as clear and "probably cloudy" as cloudy.

3.2.5 Machine learning regression models

We explored two popular machine learning regression models: Gradient Boosting Regressor (GBR) and Long Short-Term Memory (LSTM). We found that GBR outcompetes other common algorithms for estimating downwelling shortwave radiation (Ranjbar, et al., 2024b), and surface-atmosphere carbon dioxide flux (Ranjbar, et al., 2024c), striking a balance between accuracy and prediction time for real-time applications. Given these strengths, we aimed to evaluate GBR's performance against LSTM models, which may offer improved accuracy due to the effectiveness of GBR with time series data that may result in better consideration of ecosystem heat capacity (Hulley et al., 2019).

GBR is an ensemble model that combines decision trees through bootstrapping and feature bagging (Friedman, 2001). GBR aims to minimize errors by optimizing a loss function, iteratively improving weak models by addressing errors (Bentéjac et al., 2021; Sahin, 2020). GBR requires hyperparameter tuning, influencing model accuracy, size, depth, and speed. The hyperparameters include "number of estimators," "maximum depth," and "minimum samples per leaf". The "number of estimators" controls the quantity of decision trees in the ensemble, while "maximum depth" limits tree complexity to mitigate overfitting. "Minimum samples per leaf" sets the threshold for node splitting, aiding in regularization. Finally, the learning rate parameter modulates the contribution of each tree to the ensemble, fostering a balanced trade-off between convergence speed and model stability (Bentéjac et al., 2021; Friedman, 2001; Sahin, 2020).

LSTM networks excel at handling sequential data, particularly in time series forecasting (Hochreiter & Schmidhuber, 1997). They can effectively capture long-term dependencies thanks to their unique memory cell architecture, addressing a major limitation of traditional Recurrent Neural Networks (RNNs) and managing non-linear relationships within sequences (Ghanbari et al., 2021; Sutskever et al., 2014). However, their complex architecture requires more computational resources compared to simpler RNNs (Sutskever et al., 2014). Achieving optimal performance with LSTMs also involves careful hyperparameter tuning, adding complexity to the training process (Ghanbari et al., 2021; Sutskever et al., 2014; Yuan et al., 2020). Our LSTM architecture included 2 LSTM layers with the same number of units. The first layer, with `return_sequences` set to ‘true’, retained and returned the entire sequence of hidden states for each input sequence, which is crucial for capturing temporal dependencies. The second layer, with `return_sequences` set to ‘false’, returns only the output at the last time step, effectively summarizing the learned information. After the LSTM layers, we added two dense layers. The first dense layer, with 25 units, facilitated further feature extraction and representation learning. The final dense layer, with a single unit, produced the model's output.

3.2.6 Model development and assessment

We developed our models using Google Colab Pro, which provides up to 32 GB of RAM and an A100 GPU. We used a grid search algorithm to systematically test different hyperparameter combinations for each machine learning model to optimize performance. This approach helped us identify the best configurations for maximizing correlation determination (R^2) and prediction speed. Table 3.3 shows the specific hyperparameters and their ranges that we explored for each model.

Table 0.3. Grid search specifications and hyperparameter setting for machine learning models.

Model	Hyperparameter	Grid Setting
GBR	Number of estimators	100, 300, 500, 800
	Maximum depth	5, 8, 10, 12, 15
	min_samples_leaf	50, 100, 200, 500
	Learning rate	0.01,0.05,0.1
LSTM	Units (neurons)	128, 256, 512, 1024
	Activation	'logistic', 'tanh', 'relu'
	Optimizer	'adam','sgd'
	Batch Size	32, 64, 128, 256, 512

For validation and assessment, we utilized a Leave-One-Out Cross-Validation (LOOCV) technique, setting aside LST observations from 20% of the eddy covariance sites solely for testing. The remaining sites underwent a four-fold cross-validation procedure using a 75:25 training-validation ratio (Maxwell et al., 2018) to ensure reliable and robust performance across various subsets of data. To further enhance model robustness, we repeated this process ten times, each time using different random seeds to introduce variation in the training-validation splits. In each repetition, the 20% of eddy covariance sites reserved for testing were randomly reselected, and the training-validation splits in the four-fold cross-validation procedure were also adjusted accordingly. This approach ensured that different data subsets were used in each iteration, reducing potential biases and increasing the reliability of our results. The final outcomes were then averaged to provide a more stable evaluation using the Scikit-learn Python library (Pedregosa et al., 2011).

In addition to the LOOCV evaluation process, we tested our selected model against LST observations from the SURFRAD network, which provides five-minute observations. Unlike AmeriFlux data, which is available at a half-hourly resolution, SURFRAD data was solely used for validation to align with the temporal resolution of the GOES-R ABI CMI observations. For evaluating model performance, we relied on standard metrics like coefficient of determination

(R^2), Root Mean Squared Error (RMSE), and bias to measure predictive accuracy and compare different models.

After assessing performance, we turned to feature importance analysis to identify key variables influencing LST modeling. We used SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017), a tool rooted in game theory, to help us understand the impact of each feature on our predictions. SHAP analysis is crucial for uncovering the intricate details of ML predictions, revealing not just which features are important, but also how they interact and contribute to the overall model behavior (Lundberg et al., 2020; Lundberg and Lee, 2017; Temenos et al., 2023). SHAP offers a unified approach to explain and interpret machine learning outputs, bringing transparency to what can sometimes be seen as a 'black box' in predictive modeling (Descals et al., 2023; Temenos et al., 2023).

Additionally, we assessed the impact of time, land cover, and elevation-related features on model performance by selectively incorporating and removing them during estimation. Specifically, we tested the inclusion of solar zenith angle (SZA), solar azimuth angle (SAA), day of year (DOY), International Geosphere-Biosphere Programme (IGBP) land cover classification, and elevation. Based on these assessments, we selected the most influential features for modeling LST in all-sky conditions.

3.3 Results

3.3.1 ML models performance

Table 3.4 summarizes the accuracy of LST estimated from the GBR and LSTM models ($\text{ALIVE}_{\text{LST}}$), compared against calculated LST from Ameriflux and NEON towers for all-sky, clear, and cloudy conditions. These results, averaged over ten runs of LOOCV, focus on sites entirely excluded from the training process. Figure 3.3 presents the density plots comparing ALIVE$_{LST}$ from both LSTM and GBR models against calculated LST from tower measurements for both clear and cloudy sky conditions in daytime and nighttime. Table 3.5 displays model performance under all-sky conditions across various temporal resolutions, from 30 minutes to one week, with longer time periods obtained by averaging the time series. The results highlight that the LSTM model outperforms the GBR model under cloudy conditions, particularly in terms of RMSE, with differences exceeding 0.5 K.

Table 0.4. Model accuracies at daytime/nighttime under different sky conditions (all-sky, cloudy, and clear) for both GBR and LSTM models.

Condition	R^2		RMSE (K)		Bias (K)	
	LSTM	GBR	LSTM	GBR	LSTM	GBR
All-sky (day & night)	0.91	0.9	3.66	3.78	-0.26	-0.24
All-sky (day)	0.9	0.89	3.75	4.02	-0.35	-0.48
All-sky (night)	0.9	0.88	3.56	3.66	-0.17	0.01
Clear (day)	0.96	0.95	2.42	2.76	-0.36	0.05
Clear (night)	0.95	0.94	2.13	2.48	-0.25	0.67
Clear (day & night)	0.96	0.94	2.31	2.60	-0.35	0.39
Cloudy (day)	0.83	0.84	3.99	4.4	-0.35	-0.98
Cloudy (night)	0.83	0.81	4.17	4.73	-0.82	-0.78
Cloudy (day & night)	0.83	0.83	4.10	4.61	-0.52	-0.81

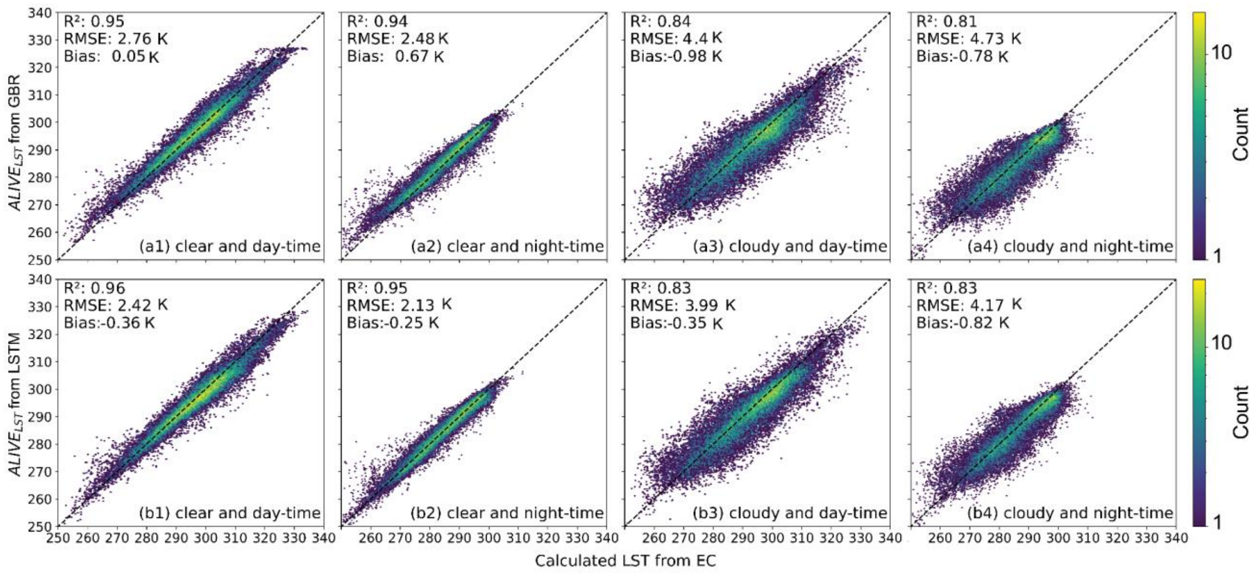

Figure 3.3. Density plots between the calculated LST from eddy covariance (EC) measurements and ALIVE$_{LST}$ from the (a1-a4) GBR model and (b1-b4) LSTM model for clear/cloudy skies and daytime/nighttime.

At the 30-minute resolution, the LSTM model consistently outperforms the GBR model. Under all-sky conditions, LSTM achieves an R^2 of 0.91 and an RMSE of 3.66 K, slightly surpassing GBR's performance (R^2 of 0.90 and RMSE of 3.78 K). During the daytime, LSTM is more accurate under clear skies, showing a slightly better fit with a higher R^2 (0.96) and lower error (2.42 K) compared to GBR (R^2 of 0.95 and RMSE of 2.76 K). In cloudy conditions, both models have similar R^2 values, but LSTM has a lower RMSE (3.99 K compare to 4.4 K). At night, the advantage of LSTM becomes even more apparent, especially under clear skies. Even in cloudy nighttime conditions, LSTM outperforms GBR with higher R^2 and a lower RMSE. Overall, the LSTM model demonstrates stronger performance across all weather conditions, both day and night.

The LSTM model demonstrates slightly better performance across different temporal resolutions (Table 3.5). As the temporal resolution becomes coarser, both models continue to show higher accuracy. For example, at the one-week resolution, LSTM achieves an R^2 of 0.97, compared to

GBR's R² of 0.94. Both models' performances at the one-day resolution have R² values of more than 0.92 and RMSEs of less than 3 K. Moreover, both models exhibit low bias values (less than 1 K), indicating reliable predictions overall. However, in terms of prediction speed, the GBR model is 5.3 times faster than the LSTM model when tested on Google Colab Pro, using a system with 32 GB of RAM and an A100 GPU.

Table 3.5. Model performance under all-sky conditions, both day and night, at different temporal resolutions.

Interval	R²		RMSE (K)		Bias (K)	
	LSTM	GBR	LSTM	GBR	LSTM	GBR
30 minutes	0.91	0.90	3.66	3.78	-0.26	-0.24
1 hour	0.92	0.91	3.53	3.89	-0.27	-0.24
1 day	0.95	0.92	2.37	2.99	-0.24	-0.22
1 week	0.97	0.94	2.27	2.51	-0.24	-0.18

3.3.2 Feature importance

Table 3.6 summarizes the impacts of additional features on model performance, including features related to time/solar geometry (SZA, SAA, and DOY), land cover (IGBP classification), and elevation. Incorporating time-related features (SZA, SAA, and DOY) into the model resulted in an improvement in performance compared to using only ABI CMI observations. Specifically, the R² for the training data increased from 0.92 to 0.95 and RMSE dropped from 3.14 K to 2.82 K, while the R² for the test sites improved from 0.91 to 0.92 and the RMSE was unchanged. When IGBP classification (or elevation) was added to the CMI features, the model showed an increase in R² for the training data, reaching 0.96 (0.95). However, the test site performance exhibited a slight decline, with R² dropping to 0.90, with an increase in RMSE from 2.44 K to 3.59 K. This suggests that while the model fit the training data more closely, its ability to generalize to new data was somewhat compromised, potentially due to overfitting. In another scenario, when all additional features (elevation, IGBP, and timing) were added to the CMI features, the model achieved the highest R² for the training data at 0.96 and the lowest RMSE at 2.43 K. However, the test site performance saw no gains, with R² of 0.92 and RMSE of 3.26 K. In our final scenario, we selected the best CMI features based on feature importance analysis, removing the visible and shortwave infrared channels, as well as channels C08 and C09, due to their weak performance in both LSTM and GBR models (see Table 3.2 for band details). Additionally, IGBP and elevation were excluded, as they did not improve the model fit for test sites. As a result, our final feature set included the selected CMI thermal infrared channels (C07, C10, C11, C12, C13, C14, C15, and C16), along with SZA, SAA, and DOY. Results from the final scenario show comparable accuracy to the full feature set, with an R² of 0.92 and an RMSE of 3.23 K for the test set (Table 3.6).

Table 0.6. Performance metrics (R² and RMSE) of the LSTM model across different feature combinations for training and test sites.

Feature sets	**R²**		**RMSE (K)**	
	Training	**Test**	**Training**	**Test**
CMIs	0.92	0.91	3.14	3.23
CMIs + SZA + SAA + DOY	0.95	0.92	2.82	3.22
CMIs + IGBP	0.96	0.90	2.44	3.59
CMIs + Elevation	0.95	0.90	2.62	3.63
CMIs + Elevation + IGBP + SZA + SAA + DOY	0.96	0.92	2.43	3.26
Selected CMIs (C07, C10, C11, C12, C13, C14, C15, C16)	0.93	0.90	2.96	3.37
Selected CMIs (C07, C10, C11, C12, C13, C14, C15, C16) + SZA + SAA + DOY	0.94	0.92	2.88	3.23

Figure 3.4 provides a detailed look at SHAP feature importance analysis, illustrating how different features from the chosen feature set contribute to LST prediction using the LSTM and GBR models under both clear and cloudy sky conditions. For clear sky conditions, both models show similar feature importance. In the GBR model, the top three features are C07 (3.9 µm), C13 (10.3 µm), and C12 (9.6 µm, see Table 3.2 for more details), with DOY and other thermal infrared

bands (C10, C15) rounding out the top six. In the LSTM model, the best-performing features are the same three in a different order, followed by C11, DOY, and C14. The overlap in the top three features between the two models highlights the critical role of these thermal infrared bands for clear sky LST predictions. However, LSTM shows a slightly higher reliance on C11 (8.4 µm), the 'cloud top phase' band (Table 3.2) compared to GBR.

Under cloudy sky conditions, feature importance shifts. The GBR model still places high importance on C07, but DOY and C12 emerge as critical predictors. In contrast, the LSTM model prioritizes a different top three: C11, C13, and C10. Interestingly, the SAA consistently appears near the bottom of the feature rankings for both models and sky conditions despite known morning/afternoon differences in LST.

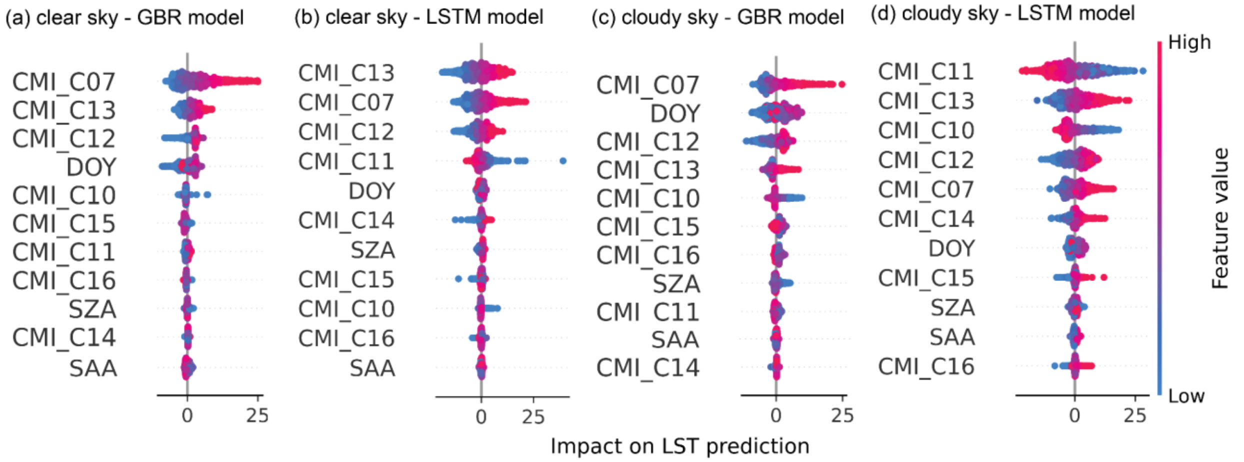

Figure 3.4. SHAP features importance analysis for the LSTM and GBR models under clear and cloudy skies.

3.3.3 ALIVE$_{LST}$ at five-minute frequency against SURFRAD network measurements

To assess model performance at a five-minute frequency, we used SURFRAD ground measurements, which provide one-minute observations. Following our approach in Section 2.3, we selected one week of data per season—centered around January 15, April 15, July 15, and

October 15, 2022—to capture diverse LST conditions. Importantly, SURFRAD data were reserved solely for testing, ensuring an independent evaluation in line with our LOOCV strategy (Figure 3.1).

Figure 3.5 presents density scatter plots comparing estimated $ALIVE_{LST}$ from the LSTM and GBR models with calculated LST from EC measurements, both for individual months and all weeks combined. The combined data density plots show that both models performed well, with R^2 values of 0.90 and RMSEs of 4.61 K and 4.42 K for the LSTM and GBR models, respectively. Model performance at a five-minute frequency ranged from R^2 of 0.67 to 0.90 across different months. Both models demonstrated weaker performance during the winter and summer months, with R^2 values of 0.67 and 0.70 in January and 0.76 and 0.78 in July for the LSTM and GBR models, respectively. The results show that the GBR model outperformed the LSTM model at a five-minute frequency in January and July, while both models performed similarly in April and October.

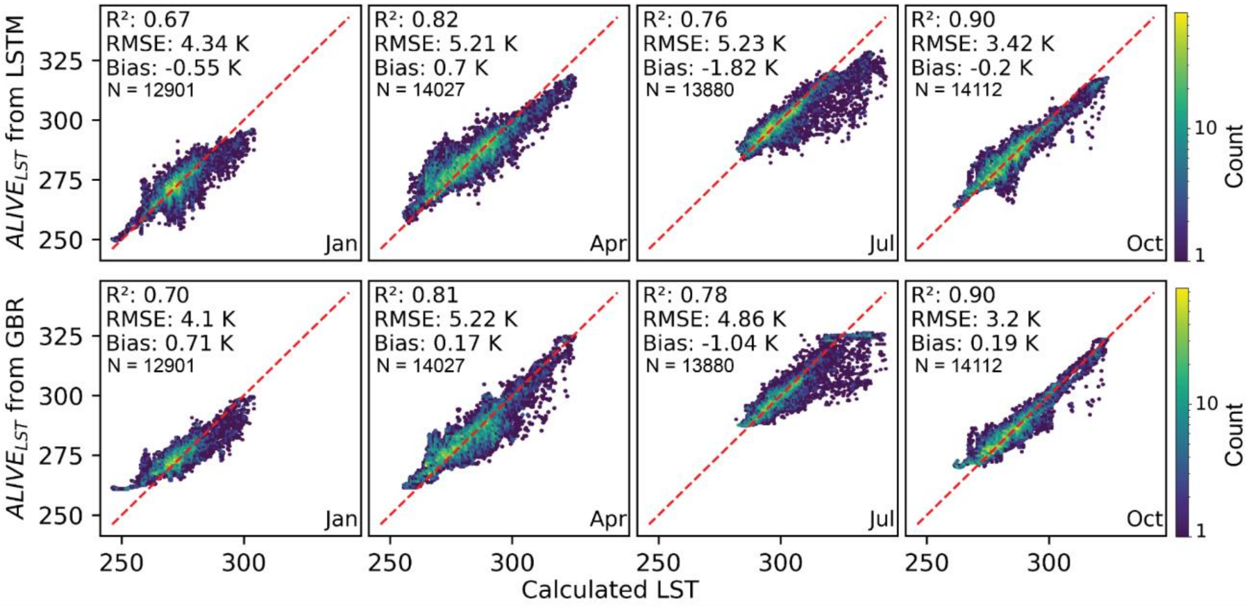

Figure 3.5. Density scatter plots between estimated ALIVE$_{\text{LST}}$ from LSTM and GBR models and calculated LST from EC measurements at five-minute frequency at seven SURFRAD sites for four weeks of data centered around January 15, April 15, July 15, and October 15, 2022.

Time series plots in Figure 3.6 show how the ALIVE$_{\text{LST}}$ estimates from both models compare with calculated LST from EC measurements for both daytime and nighttime at specific SURFRAD sites. ALIVE$_{\text{LST}}$ estimates from both models closely matched ground-based LST measurements on clear days with minimal fluctuations. The best match was observed at the TBL site in Colorado, while the poorest match occurred at the FPK site in Montana. However, the models' accuracy declined on days with significant LST fluctuations, often due to cloud cover. Notably, a large bias was observed during the week of April at the FPK site, with a similar bias present at the DRA site in Nevada during July. Large biases are observed for FPK in April, likely due to continuous multi-day cloud cover (see e.g. subplot FPK(Apr) in Figure 3.6). Additionally, the GBR model showed significant biases on January 15 and October 18 at the SXF site in South Dakota, whereas the LSTM model performed well, providing accurate estimates. A similar bias from the GBR model was also observed on January 15 at the PSU site in Pennsylvania. However, LSTM exhibited larger overall fluctuations, as indicated by the red dotted line corresponding to the LSTM estimates.

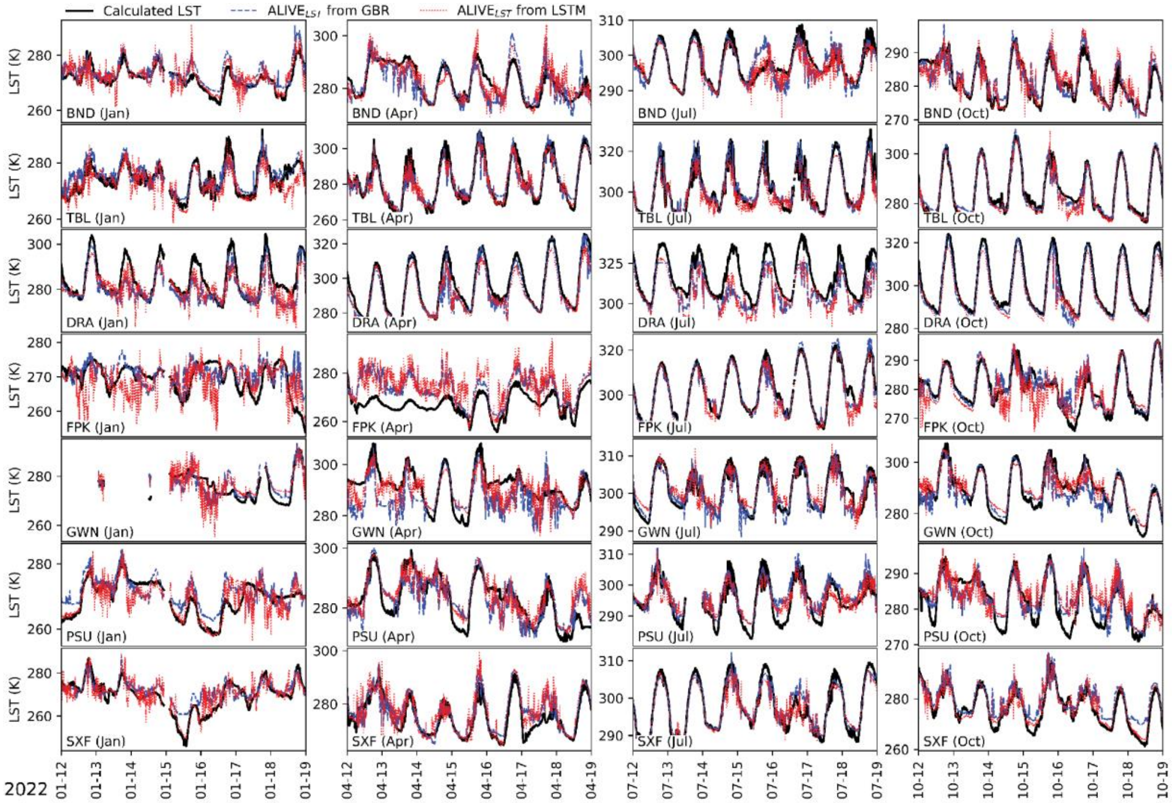

Figure 3.6. Time series of $\text{ALIVE}_{\text{LST}}$ from LSTM and GBR model vs. LST calculated from EC measurements in five-minutes frequency at seven SURFRAD sites in four weeks of data centered around January 15, April 15, July 15, and October 15, 2022.

3.3.4 ALIVE$_{LST}$ performance by month

Figure 3.7 presents the monthly performance of ALIVE$_{LST}$ from the LSTM model under clear and cloudy sky conditions. The model performs well under clear skies but shows reduced accuracy in cloudy conditions, with the largest errors in February and March (RMSE ~4 K) and stable performance in spring and autumn. Figure 3.7d illustrates a higher percentage of cloud cover during the winter months.

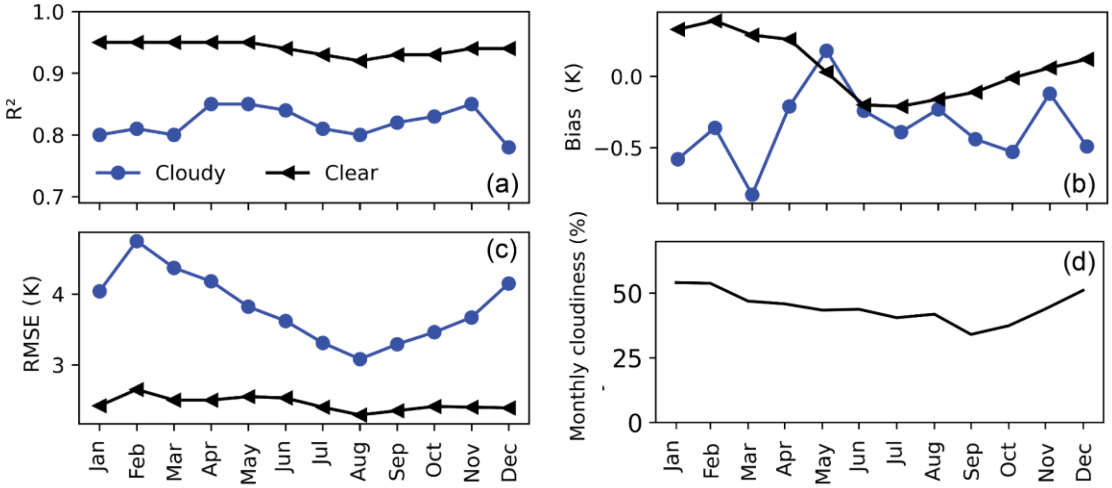

Figure 3.7. Monthly performance of the ALIVE$_{LST}$ from the LSTM model under clear and cloudy sky conditions. Panels (a), (b), and (c) present the R^2, bias, and RMSE, respectively, while panel (d) illustrates the average monthly cloud cover percentage across the 101 AmeriFlux sites.

The analysis of monthly model performance under clear sky versus cloudy sky conditions reveals significant differences. For clear sky conditions, the model consistently achieves high R^2 values, ranging from 0.92 to 0.95, indicating a strong correlation between predicted and observed values across all months. In contrast, the cloudy sky scenarios present a lower R^2 range of 0.78 to 0.85, reflecting a diminished predictive capacity. The RMSE values reinforce this disparity, with clear sky RMSE values fluctuating between 2.29 K and 2.65 K, compared to higher RMSE values of 3.08 K to 4.75 K for cloudy skies, suggesting greater prediction errors in the presence of cloud cover.

Examining seasonal variations, the model performs best in the summer months (particularly May to August) for clear skies, maintaining R^2 values around 0.95 and RMSE values ranging from 2.29 to 2.55 K. In contrast, the cloudy sky performance shows a notable decline in accuracy during the winter months (January to March), with R^2 values peaking at only 0.80 to 0.81 and RMSE values significantly higher, ranging from 4.04 to 4.75 K. As the seasons transition to spring and summer, cloudy sky performance improves slightly, with the best values recorded in May (R^2 of 0.85, RMSE of 3.82 K). However, it does not reach the reliability of clear sky conditions.

3.3.5 Intercomparison: ALIVE$_{LST}$ from LSTM and GBR vs. ABI$_{LST}$

We compared ALIVE$_{LST}$ from LSTM and GBR against the physically-based ABI$_{LST}$ product (Figure 3.8). This analysis was limited to clear sky conditions as the physically-based ABI$_{LST}$ product is only available for land surfaces under clear skies, with up to a local zenith angle of 85 degrees for both daytime and nighttime (Figure 3.8). January 21, 2023, and July 20, 2023, were specifically selected for this comparison to capture the full range of LST values, representing a winter day and a summer day in the northern hemisphere, respectively, while also minimizing cloud cover.

On January 21, 2023, during night (09:02 UTC), ALIVE$_{LST}$ from the GBR model showed strong agreement with the ABI$_{LST}$ product, with an R^2 of 0.88, an RMSE of 3.65 K, and a cold bias of -2.14 K. The LSTM model performed better than the GBR model during the same period, achieving a higher R^2 of 0.94, a lower RMSE of 3.0 K, and a reduced bias of -1.39 K. During the daytime (18:02 UTC) on the same day, the agreement between ALIVE$_{LST}$ from the GBR model and ABI$_{LST}$ improved. The GBR model achieved an R^2 of 0.96 with a lower RMSE of 2.79 K and a negligible bias of 0.04 K. The LSTM model also maintained a high R^2 of 0.95, although with a slightly higher RMSE of 3.06 K and a warm bias of 0.7 K.

For July 20, 2023, the comparison results shifted significantly, reflecting the impact of warmer summer conditions. At nighttime (09:02 UTC), the GBR model's performance remained consistent with an R^2 of 0.86, an RMSE of 2.04 K, and a bias of 0.3 K. The LSTM model performed similarly, albeit with a slightly lower R^2 of 0.83 and a slightly higher RMSE of 2.34 K, alongside a warm bias of 1.13 K. Daytime results (18:02 UTC) on the same day indicated a notable drop in performance, particularly for the LSTM model. The GBR model retained a moderate R^2 of 0.82 with an RMSE of 3.61 K and a bias of only -0.21 K. In contrast, the LSTM model showed a more pronounced decline, with an R^2 of 0.68, an RMSE of 4.62 K, and a warm bias of 0.85 K. These results suggest that while both models struggled more in capturing daytime LST during the summer, the GBR model remained more consistent than the LSTM model.

When comparing the two models directly, the scatter plots between $ALIVE_{LST}$ from GBR and LSTM showed strong agreement, particularly on the January 21, 2023 test date, with R^2 values of 0.96 (nighttime) and 0.98 (daytime), and RMSEs of 2.5 K and 1.72 K, respectively. During July 20, 2023, the models exhibited slightly reduced agreement, with R^2 values of 0.93 (nighttime) and 0.89 (daytime), and RMSEs of 1.47 K and 2.71 K, respectively.

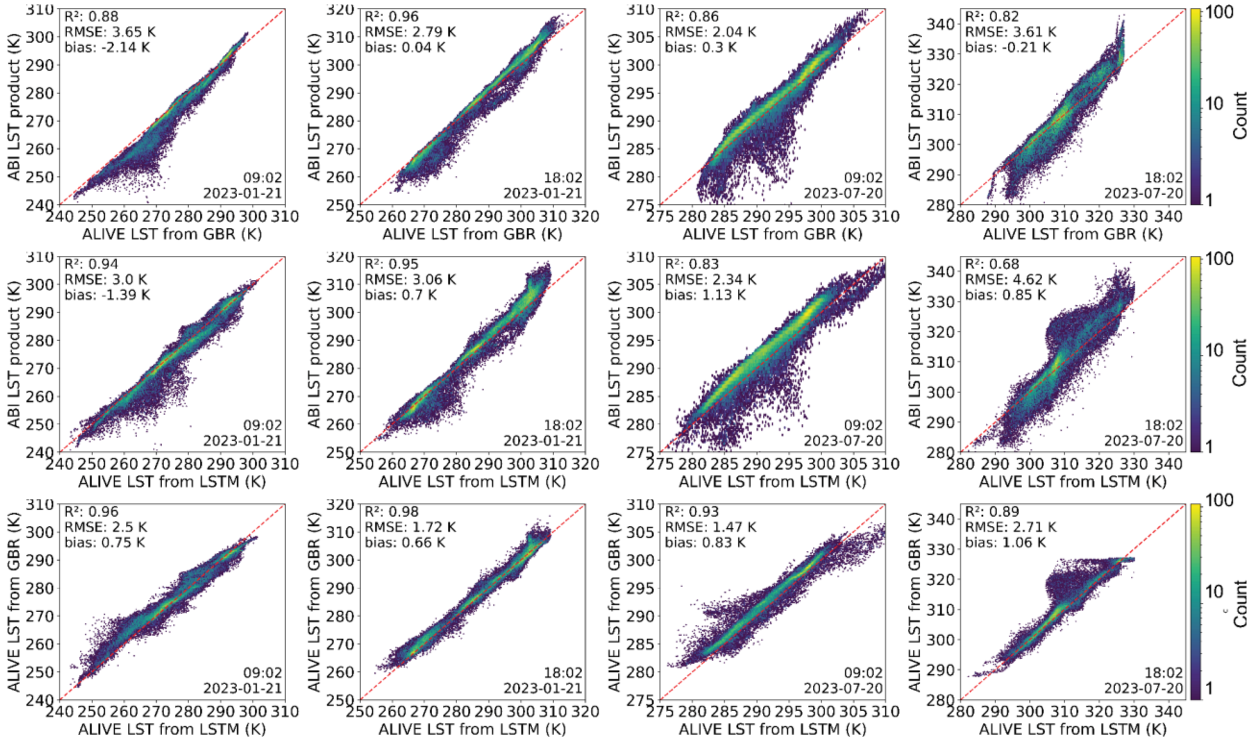

Figure 3.8. Density scatter plots comparing ABILST and ALIVELST estimates from GBR and LSTM models, as well as between LSTM and GBR estimates. The plots are shown for nighttime (09:02 UTC) and daytime (18:02 UTC) on January 21 and July 20, 2023. A random subset of 100,000 samples was selected from the total 3,750,000 samples for visualization, as repeated observations confirmed that this subset effectively represents the entire dataset.

Figures 8 and 9 present a comparison of LST maps from the ABILST product and the estimated ALIVELST generated by the GBR and LSTM models. Figure 3.9 shows LST maps for both nighttime (09:02 UTC) and daytime (18:02 UTC) on January 21 and July 20, 2023, highlighting how effectively the GBR and LSTM models capture spatial LST patterns relative to the ABILST product. While the ABILST is limited to clear sky conditions, resulting in significant data gaps due to cloud cover, the ALIVELST estimates from both models provide continuous coverage under all-sky conditions. Figure 3.9 provides difference maps that illustrate the deviations between the ABILST and ALIVELST estimates from the GBR and LSTM models, as well as the differences between the two models themselves. These maps reveal how closely the models align with the

ABI_{LST} product and each other, while also exposing biases, particularly under cloudy conditions, such as on and around the US-Mexico border on July 20, 2023 during daytime (18:02 UTC).

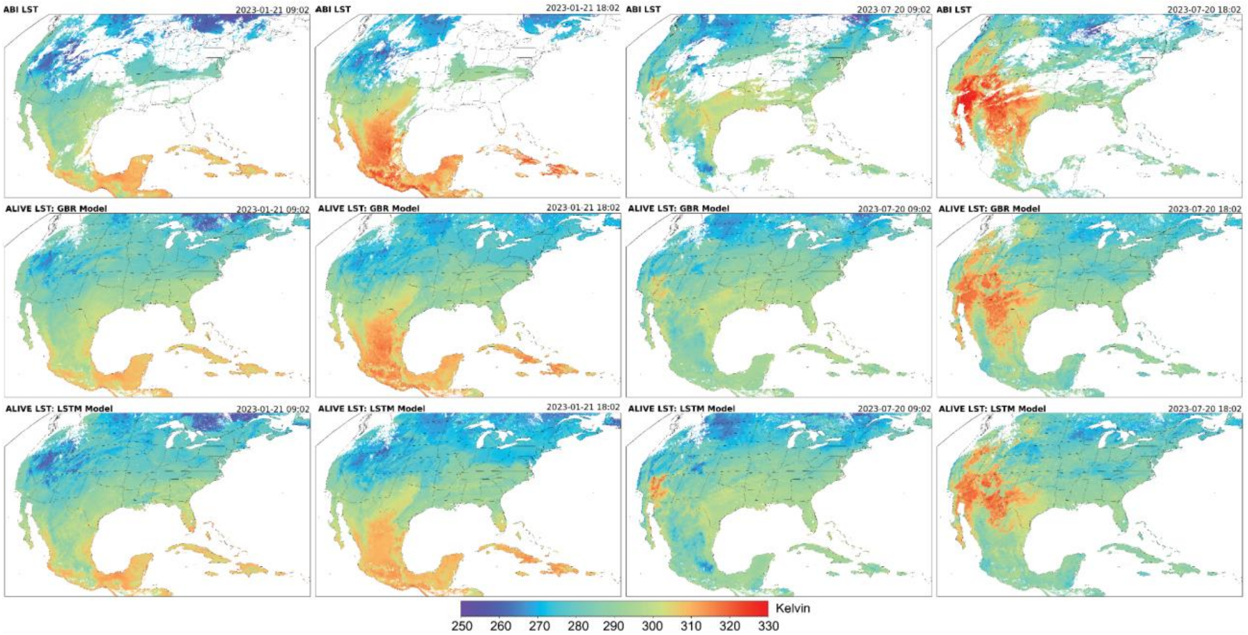

Figure 3.9. Comparison of the ABI$_{LST}$ product with estimated ALIVE$_{LST}$ maps from GBR and LSTM models, shown for nighttime (09:02 UTC) and daytime (18:02 UTC) on January 21 and July 20, 2023.

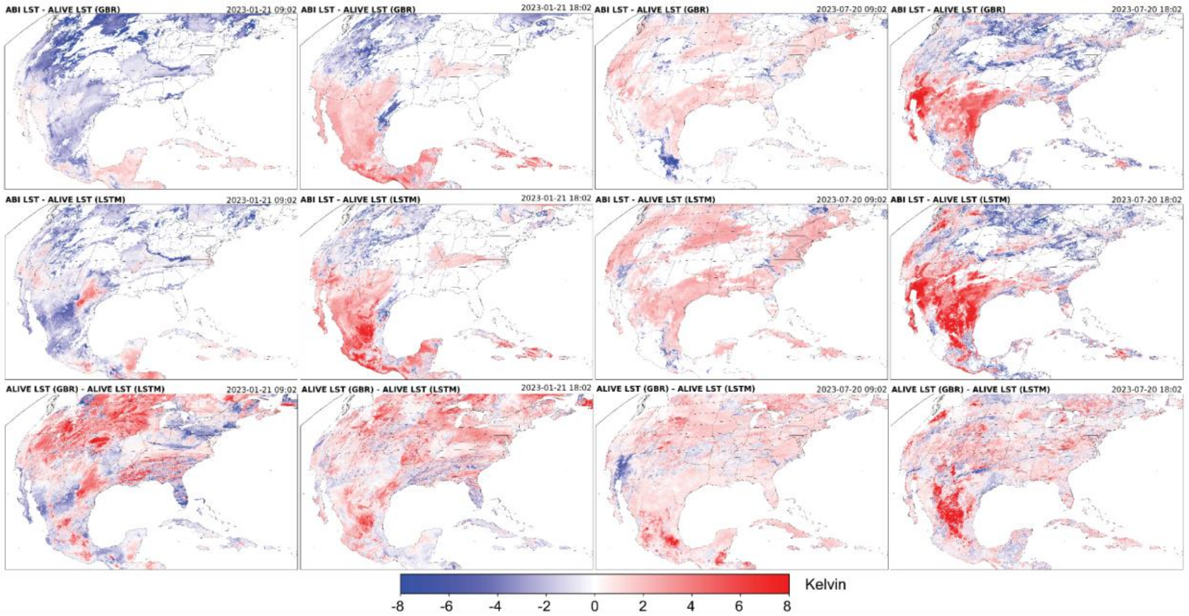

Figure 3.10. Difference maps showing the deviations between ABI$_{\text{LST}}$ and ALIVE$_{\text{LST}}$ estimates derived from the GBR and LSTM models, as well as the differences between the LSTM and GBR estimates. The maps provide a comparative analysis for both nighttime (09:02 UTC) and daytime (18:02 UTC) conditions on January 21 and July 20, 2023, highlighting seasonal and diurnal variations in model performance.

3.4 Discussion

3.4.1 ML performance

The results emphasize the LSTM model's strong performance, particularly under challenging conditions like nighttime and cloudy skies, where temporal dependencies play a crucial role (Dhake et al., 2023; Lara-Alvarez et al., 2024). The LSTM frequently outperforms the GBR model as evidenced by the slightly higher R^2 (0.01-0.02) and lower RMSE values reported across both daytime and nighttime conditions, for both clear and cloudy skies (Table 3.4). The LSTM model's architecture, designed to handle sequential data, allows it to capture nuanced temporal patterns that are critical for accurate LST estimation under variable environmental conditions, supporting the model's slightly better performance (Lara-Alvarez et al., 2024; Miller et al., 2024; Moskolaï et al., 2021; Yuan et al., 2020). While the GBR model performs slightly poorer than the LSTM model on average, it still delivers promising estimates and can outperform LSTM during certain periods. The higher RMSE for the GBR model under both cloudy daytime and nighttime conditions reinforces the observation that boosting weak learners, while effective in many cases, may not capture the same level of temporal complexity as LSTM, leading to less accurate predictions under more variable sky conditions (Bentéjac et al., 2021). LSTM's improved performance under cloudy conditions (Table 3.4) is consistent with its ability to account for temporal dependencies, which is crucial in situations with fluctuating cloud cover.

Both models show improved accuracy at longer temporal resolutions, but a key tradeoff exists between the LSTM model's superior accuracy and the GBR model's faster computation speed. While the LSTM model continues to outperform GBR at daily and weekly resolutions (Table 3.5), the GBR model remains reliable (Liu et al., 2022). With the GBR model being 5.3 times faster than LSTM in terms of prediction speed, it is an attractive option for real-time applications where speed is critical and minor trade-offs in accuracy are acceptable (Ranjbar et al., 2024c). While this speed difference likely reflects both algorithmic efficiency and available computational resources, it is uncertain how much additional GPUs or memory would impact the relative performance. For example, adding more GPUs or memory might reduce the gap, but the fundamental differences between the models would still play a role. The 5x speed advantage is particularly significant for real-time applications, where faster predictions enable more responsive systems and lower latency in decision-making processes. The choice between the LSTM's higher accuracy and the GBR's faster performance should be carefully weighed based on the specific requirements of the application.

3.4.2 Feature importance

The results offer a nuanced view of how additional features, particularly timing (SZA, SAA, and DOY) and physical features (IGBP classification and elevation), affect model performance. Incorporating timing features enhances the LSTM model's ability to capture temporal variations (Table 3.6), for both the training and test datasets. This improvement reinforces the importance of solar geometry and seasonal context in accurately predicting LST, as these factors drive surface temperature changes over time. However, the results also highlight a trade-off when adding static and physical features like land cover and elevation. While these features improve the fit to the training data, they appear to contribute to overfitting, reducing generalization to unseen test sites, suggesting that a simpler model might generalize better in this case (Dhake et al., 2023; Lara-Alvarez et al., 2024; Miller et al., 2024).

The SHAP analysis provides deeper insights into feature prioritization across varying atmospheric conditions. For both the LSTM and GBR models, the dominance of thermal infrared bands like C12 and C13 (see Table 3.2 for band details) under clear skies underscores their pivotal role in capturing surface thermal properties. C12, sensitive to mid-level water vapor, and C13, key for detecting temperature variations, consistently emerge as crucial features for accurate LST predictions (Goodman et al., 2019). Interestingly, neither model prioritized bands C14 (11.2 μm) and C15 (12.3 μm), which are the key bands for the ABI_{LST} product that uses the physically-based split-window algorithm (Yu and Yu, 2020a). This discrepancy suggests that the machine learning models leverage different feature interactions compared to the physically-based split-window method. The LSTM model's higher reliance on C11, compared to the GBR model, may reflect its ability to integrate more complex relationships between the infrared bands, further refining its predictions.

Under cloudy conditions, the LSTM model's increased reliance on C11 and C10, compared to the GBR model's continued emphasis on C07 and DOY, highlights its different ability to handle cloud interference. The shift in feature importance between clear and cloudy skies illustrates how the LSTM model dynamically adjusts to different atmospheric contexts, emphasizing features more relevant to capturing the effects of moisture and cloud cover on surface temperature. C07, which is sensitive to low-level clouds, plays a key role in the GBR model under both cloudy and nighttime conditions, where accurate cloud detection becomes critical. The 3.9 μm band (C07) appears important in our SHAP analysis, likely due to its sensitivity to rapid surface temperature changes and emissivity variations, especially in low-humidity conditions or during dawn and dusk (Goodman et al., 2019). Moreover, it is useful for identifying fog and low clouds, detecting volcanic ash, and estimating atmospheric conditions such as low-level vector winds (Goodman et al., 2019; Heidinger et al., 2020). While typically used for fire detection, cloud characterization, and sea surface temperature estimation (Goodman et al., 2019), its correlation with key thermal bands (C12, C13) may enhance LST predictions. Given its mixed emitted and reflected signals, further investigation is needed to confirm its role in ML-based LST retrieval. Additionally, the GBR model effectively leverages timing features such as DOY and SZA, using them to account for seasonal and diurnal variations in solar radiation, further enhancing its predictions under challenging conditions (Figure 3.4).

The exclusion of less impactful features like certain visible/shortwave infrared channels in the final feature selection improves the model's balance between accuracy, generalization, and prediction speed. Removing lower impact channels helps avoid overfitting without compromising prediction quality. The final simplified feature set, with thermal CMI channels and timing elements, achieves comparable accuracy to the full feature set, demonstrating that optimal feature selection can streamline model complexity while maintaining high performance across varying sky conditions. While timing features like SZA and DOY provide essential context for solar-driven temperature changes, thermal bands remain the primary drivers of accuracy, especially under variable sky conditions. The nuanced differences in feature importance between the LSTM and GBR models further reflect their distinct approaches to handling complex environmental factors in LST prediction.

3.4.3 ALIVE$_{\text{LST}}$ performance: temporal and seasonal comparisons

The performance of the ALIVE$_{\text{LST}}$ model was evaluated using both datasets, five-minute SURFRAD network data and half-hourly AmeriFlux and NEON data, highlighting both the strengths and limitations of the LSTM and GBR models across different temporal and seasonal conditions. When evaluated against the five-minute SURFRAD data, both the LSTM and GBR models showed overall strong performance. However, challenges arose during extreme conditions, particularly in January and July, where both models struggled to capture the large LST fluctuations. The GBR model performed better during these months, suggesting a greater ability to handle rapid temperature changes. However, biases at specific sites, such as SXF and PSU (Figure 3.6), point to sensitivity to local conditions like sudden temperature change, snow and/or cloud cover, which introduce variability in surface radiation, including emissivity (Chen et al., 2012; Jia et al., 2021). In contrast, the LSTM model exhibited greater stability with fewer biases under variable conditions, particularly in January and October, demonstrating robustness in handling LST fluctuations over time. Despite this, the LSTM model showed larger fluctuations in its estimates during periods of high LST variability, such as at FPK and DRA sites, likely due to its recurrent nature, which can amplify errors when faced with noisy data or rapid changes (Moskolaï et al., 2021; Yuan et al., 2020). The LSTM model's ability to retain information from previous timesteps offers dynamic memory advantages, though it can also lead to instability in rapidly changing environments, while the GBR model's efficiency is tempered by biases under extreme conditions.

When compared against the half-hourly AmeriFlux and NEON data, the seasonal variation reveals that the model struggles particularly during winter, with heightened errors in February and March, where RMSE values approached 4 K. These errors likely result from the combination of low solar radiation and the presence of snow cover, both of which introduce significant challenges (Robledano et al., 2022; Thiebault and Young, 2020; Williamson et al., 2017). The presence of snow, especially at off-nadir angles, introduces errors due to its lower (and variable) emissivity values compared to vegetative materials (Hori et al., 2006), complicating LST retrieval. Northern Hemisphere winter’s inherent atmospheric variability, particularly under cloudy skies, further compounds these challenges, leading to a performance drop, as seen in the increased error margins during the colder months. Additionally, the variability in snow emissivity likely contributes to this effect, as emissivity for fresh snow is ~0.98–0.99, but older snow at large view zenith angles exhibits lower emissivity, introducing further uncertainties in LST retrieval (Warren, 2019). Interestingly, the $\text{ALIVE}_{\text{LST}}$ model maintains relatively stable performance during spring and autumn, but a slight decline in accuracy is observed during the summer months, especially in July and August. This decline contrasts with the expected consistent performance and may be attributed to complex atmospheric conditions such as increased near-surface humidity and variable cloud cover, including fair-weather cumulus clouds, which can affect LST retrieval (Jia et al., 2020; Wang et al., 2024).

The reduced accuracy observed under cloudy skies across all seasons, particularly during winter and autumn, reinforces the challenge of retrieving LST when cloud cover interferes with surface thermal signals (Cho et al., 2022; Zhang et al., 2024; Zhang et al., 2023). As illustrated by the R^2 decline under cloudy conditions, ranging from 0.78 to 0.85, the model’s decreased performance suggests the need for improved techniques to handle cloud-related atmospheric complexities (Shwetha and Kumar, 2016; Tang et al., 2023). Addressing these challenges, particularly during winter and transitional seasons, could lead to significant improvements in the model's overall robustness and accuracy under varying sky conditions. Continuous cloud cover reduces estimation accuracy and introduces large biases (see e.g. subplot FPK(Apr) in Figure 3.6). This suggests that future work could include a sensitivity analysis to evaluate model performance under varying cloud cover durations, ranging from hourly to multi-day occlusions. Additionally, examining RMSE variations at different times of the day, particularly during morning hours, could help refine LST retrieval techniques and address spatiotemporal mismatches.

3.4.4 Comparing $\mathrm{ALIVE}_{\mathrm{LST}}$ against $\mathrm{ABI}_{\mathrm{LST}}$ and recent studies

The comparison between $\mathrm{ALIVE}_{\mathrm{LST}}$ derived from the LSTM and GBR models against $\mathrm{ABI}_{\mathrm{LST}}$ highlights both the strengths and limitations of ML-based approaches relative to a physically-based LST product. The ABI instrument provides a reliable source for generating LST products through a split-window technique using thermal infrared bands (Goodman et al., 2019; Yu and Yu, 2020a) and its stated precision, 2.5 K, is similar to our models under clear conditions Figure 3.7c), though as noted, unlike our approach, is not designed to work under clouds. $\mathrm{ABI}_{\mathrm{LST}}$ offers high temporal and spatial resolution, making it an essential benchmark for evaluating the performance of the LSTM and GBR models.

The strong agreement between $\mathrm{ALIVE}_{\mathrm{LST}}$ and $\mathrm{ABI}_{\mathrm{LST}}$ during January, particularly under nighttime and clear sky conditions, suggests that the data-driven models generally align well with the physically-based model in cold environments. However, the presence of melting snow introduces complexities, as the snow's surface temperature can hover around 273.15 K, while the rest of the snowpack may be colder. This variation, along with challenges in accurately determining the snow's emissivity (Hori et al., 2006), may impact the precision of LST estimates in a physically-based algorithm. These factors likely make it harder for the models to perform consistently across the full range of observed values (Jia et al., 2022; Jia et al., 2020). The GBR model's strong daytime performance (Figure 3.8), with negligible bias and lower RMSE, emphasizes its robustness under clear, consistent conditions.

In contrast, the agreement between $\mathrm{ALIVE}_{\mathrm{LST}}$ and $\mathrm{ABI}_{\mathrm{LST}}$ for both ML models declined during July, with the LSTM model showing a more pronounced drop (Figure 3.8), especially during the daytime. This decline underscores the challenges of predicting LST under the more variable conditions of summer, where higher solar radiation, diurnal temperature variations, and complex surface dynamics come into play. These factors introduce a greater influence on the various input variables used in both data-driven and physically based models for LST prediction. The better consistency of the GBR model during this period can likely be attributed to its ensemble learning framework, which tends to be more resilient to non-linearities and variability (Bahrami et al., 2021a; Bentéjac et al., 2021). The notable warm bias in the LSTM model during summer suggests that it struggles more with extreme daytime conditions, possibly due to over-reliance on learned temporal patterns that may not fully capture the increased heterogeneity of summer landscapes (Dhake et al., 2023). Min et al., (2021) identified a warm and dry bias in the summer and an extreme nocturnal cold bias in the winter in the High-Resolution Rapid Refresh (HRRR) model using data from the New York State Mesonet over the course of a year. The summer warm bias is attributed to inaccuracies in land-surface and cloud processes, while the winter cold bias is linked to misrepresentations in energy partitioning and soil hydrology (Min et al., 2021). This raises the question of whether universal model training is misrepresenting key periods, leading to a regression toward the mean.

The difference maps further highlight key spatial biases between the models and ABI$_{LST}$, particularly under cloudy conditions, where both models provide continuous coverage while the clear sky ABI$_{LST}$ product has significant data gaps. The biases along the southeast U.S. coast during January and around the US-Mexico border in July point to the ongoing challenges of cloud interference (Figure 3.10), which both models struggled with. This reinforces the need for better integration of cloud dynamics in machine learning models to improve their performance in cloudy conditions. The ABI$_{LST}$ product's physically-based approach provides a critical benchmark, while the data-driven LSTM and GBR models offer the advantage of continuous coverage, underscoring the potential for integrated approaches in future LST research and applications (Cho et al., 2022; Zhang et al., 2024).

Recent advances in all-sky LST estimation have focused on data fusion Sun et al., (2019), machine learning-based retrieval (Cho et al., 2022; B. Li et al., 2021; Zhang et al., 2024, 2023), and spatiotemporal modeling (Jia et al., 2024), to improve accuracy and gap-filling under cloud cover. However, many approaches rely on complex multi-source inputs or operate at lower temporal resolutions, limiting their real-time applicability. The ALIVE framework addresses these gaps by leveraging near-real-time ABI observations and a lightweight GBR model, streamlining all-sky LST retrieval for high-frequency monitoring while maintaining competitive accuracy.

3.5 Conclusion

This study successfully demonstrates the capability of two well-known machine learning algorithms, Long Short-Term Memory (LSTM) networks and Gradient Boosting Regression (GBR) models, to estimate LST under all-sky conditions for day and night time using top-of-atmosphere ABI satellite data and timing features. The LSTM model excelled in capturing complex temporal patterns, achieving high accuracy, particularly in clear skies with an R^2 of 0.96 and an RMSE of 2.31 K, and under cloudy conditions with an R^2 of 0.83 and an RMSE of 4.10 K. Despite its slightly lower accuracy, the GBR model proved to be a more practical choice for real-time applications, performing 5.3 times faster than LSTM with only a minimal reduction (0.01-0.02) in R^2 and lower bias during key periods. The feature importance analysis revealed that the infrared bands at 3.9 μm (C07) and 10.3 μm (C13) are more influential than the commonly used bands at 11.2 μm (C14) and 12.3 μm (C15) in physically-based LST estimations, highlighting the model's ability to adaptively extract valuable information from various spectral bands. It shows that ML takes a distinct approach by incorporating different feature interactions than the physically-based method. These findings emphasize the strengths and limitations of data-driven approaches in LST estimation, providing valuable insights for improving near-real-time LST mapping. Future research should focus on refining these models and exploring different machine learning architecture to enhance the prediction speed, accuracy and spatial coverage of LST products across the entire world, especially in challenging conditions such as cloudy skies and regions with high thermal variability.

Chapter 4. Using Geostationary Satellite Observations and Machine Learning Models to Estimate Ecosystem Carbon Uptake and Respiration at Half Hourly Time Steps at Eddy Covariance Sites

This chapter is published in the Wiley-Journal of Advances in Modeling Earth Systems (JAMES): Ranjbar, S., Losos, D., Hoffman, S., Cuntz, M., & Stoy, P. C. (2024). Using geostationary satellite observations and machine learning models to estimate ecosystem carbon uptake and respiration at half hourly time steps at eddy covariance sites. Journal of Advances in Modeling Earth Systems, 16(10), e2024MS004341. https://doi.org/10.1029/2024MS004341

Abstract

Polar-orbiting satellites have significantly improved our understanding of the terrestrial carbon cycle, yet they are not designed to observe sub-daily dynamics that can provide unique insight into carbon cycle processes. Geostationary satellites offer remote sensing capabilities at temporal resolutions of 5-minutes, or even less. This study explores the use of geostationary satellite data acquired by the Geostationary Operational Environmental Satellite - R Series (GOES-R) to estimate terrestrial gross primary productivity (GPP) and ecosystem respiration (RECO) using machine learning. We collected and processed data from 126 AmeriFlux eddy covariance towers in the Contiguous United States synchronized with imagery from the GOES-R Advanced Baseline Imager (ABI) from 2017 to 2022 to develop ML models and assess their performance. Tree-based ensemble regressions showed promising performance for predicting GPP (R^2 of 0.70 ± 0.11 and RMSE of 4.04 ±1.65 μmol m^{-2} s^{-1}) and RECO (R^2 of 0.77 ± 0.10 and RMSE of 0.90 ± 0.49 μmol m^{-2} s^{-1}) on a half-hourly time step using GOES-R surface products and top-of-atmosphere observations. Our findings align with global efforts to utilize geostationary satellites to improve carbon flux estimation and provide insight into how to estimate terrestrial carbon dioxide fluxes in near-real time.

4.1. Introduction

Satellite remote sensing has revolutionized our understanding of the terrestrial carbon cycle (Xiao et al., 2019) and demonstrated the critical role of terrestrial carbon fluxes in determining the trends and variability of atmospheric carbon dioxide concentration (Friedlingstein et al., 2022; Zhu et al., 2016). Many of these insights come from polar-orbiting satellites like MODIS (see Table 4.1), whose observations have been used to estimate key carbon cycle processes like gross primary productivity (GPP) and ecosystem respiration (RECO), typically on time scales of weeks to years (Running et al., 2004; Running and Zhao, 2015; Turner et al., 2006). These algorithms are being continually improved using GPP and RECO estimates from eddy covariance measurements (Heinsch et al., 2006; Huang et al., 2021; Wang et al., 2017), all of which have led to a greater appreciation of the role ecosystems play in the global carbon cycle and the critical climate services they provide (Novick et al., 2022).

Despite the key contributions of polar-orbiting satellites in carbon cycle research, a number of challenges remain, especially with respect to changes to carbon cycle processes that happen quickly. Rapid phenological transitions remain difficult to observe and model (Richardson et al., 2013). Extreme events play a disproportionate role in carbon cycling at ecosystem to global scales (Reichstein et al., 2013; Zscheischler et al., 2014), but their full impacts can be difficult to ascertain with intermittent measurements, especially when cloudiness reduces data availability (Jeong et al., 2023a). Observing diurnal patterns in variables linked to plant stress help fully understand the impacts of drought and other disturbances (Li et al., 2023; Xiao et al., 2021). Model-data integration efforts are increasingly recognizing the importance of sub-daily processes. Recent developments in data-driven terrestrial land-atmosphere flux modeling through the FLUXCOM-X framework and X-BASE products (Nelson et al., 2024) offer high-resolution estimates of terrestrial carbon dioxide and water vapor fluxes, with up to 0.05° spatial and hourly temporal resolution.

To enable near real-time carbon cycle prediction, geostationary (“weather”) satellites are increasingly being used to estimate land surface and carbon cycle processes on time scales as short as minutes (Desai et al., 2021; Jeong et al., 2023; Khan et al., 2022; Li et al., 2023; R. Nemani et al., 2020; Ranjbar et al., 2024; Wang et al., 2020; Xiao et al., 2021). “Hypertemporal” remote sensing (Miura et al., 2019) may better capture the impacts of rapid changes on land surface properties including those related to carbon cycling (Khan et al., 2021), and serve as a powerful new tool for understanding our changing carbon cycle (Xiao et al., 2021).

Measuring variables related to carbon cycle processes from geostationary satellites has become increasingly feasible, thanks to the enhanced spectral capabilities of new imagers in capturing data in near-infrared spectrum (Schmit et al., 2017a; Schmit and Gunshor, 2020). The latest generation of sensors mounted on geostationary satellites, like the Advanced Baseline Imager (ABI) mounted on the Geostationary Operational Environmental Satellite - R Series (GOES-R) and Advanced Himawari Imager (AHI) mounted on Himawari 8/9, can capture data in the visible, near-infrared and infrared (Schmit and Gunshor, 2020). These observations can be used to calculate the normalized difference vegetation index (NDVI, Wheeler and Dietze, (2021)), which is related to the leaf area index and therefore to GPP. Geostationary satellite data products can also be used to derive the near infrared reflectance of vegetation (NIRv), which is more closely related to GPP than NDVI alone, even in pixels with mixed vegetation types (Badgley et al., 2019, 2017). Such vegetation indices can be coupled with key products like land surface temperature (LST) and downwelling shortwave radiation (DSR) that have been developed from geostationary satellite observations for decades (Diak & Gautier, 1983; Gautier et al., 1980; Wang & Li, 2022). Coupling these variables can help develop models for GPP and RECO that are capable of capturing diurnal, seasonal, and interannual patterns in the terrestrial carbon cycle (Jeong et al., 2023; Khan et al., 2022).

There is no current synthesis on how to best employ geostationary satellite observations to estimate GPP and RECO across multiple ecosystems, but notable recent progress has been made in using geostationary satellites to estimate GPP at single sites or a suite of sites. Xiao et al. (2021) demonstrated how observations from the Advanced Himawari Imager can be used to infer diurnal patterns in regional GPP using a light use efficiency model similar to the MODIS GPP algorithm across large swaths of Australia. Khan et al. (2022) noted that a model based on the NIRv multiplied by photosynthetically active radiation (NIRvP, Dechant et al., (2022)) outperformed light use efficiency models when estimating GPP from eddy covariance measurements at a savanna ecosystem in California using observations from the ABI on the GOES-R series satellites. Jeong et al. (2023) demonstrated how a GPP model also based on NIRvP derived from GeoKompsat-2A observations can capture the diurnal and seasonal patterns of GPP estimated from multiple ecosystems in Korea. These findings highlight the potential for geostationary satellites to estimate GPP on time scales of minutes using visible and near infrared surface reflectance products coupled with DSR as an estimate of photosynthetically active radiation. It remains unclear how well geostationary satellites can estimate RECO, or the degree to which Machine Learning (ML) techniques can improve predictive skill over established methods given their promise for integrating eddy covariance and remote sensing observations (Jung et al., 2020; Li et al., 2022; Yu et al., 2021). It is also unclear if additional measurements from geostationary satellites – like measurements in the shortwave infrared and infrared related to atmospheric characteristics that are infrequently used for carbon cycle science – may further improve terrestrial carbon flux estimates.

It is crucial to develop a comprehensive understanding of how to effectively utilize geostationary satellite observations to enhance terrestrial ecosystem science (Khan et al., 2021).

Here, we start to answer these questions by asking: (1) How accurately can ML techniques developed using GOES-R ABI-measured reflectances, brightness temperatures, and data products model GPP and RECO? (2) How do these models perform when estimating GPP and RECO from hundreds of eddy covariance tower sites? (3) What is the impact of different input variables on the performance of these machine learning models? We test the ability of different ML approaches that use different inputs from both surface data products and top-of-atmosphere (TOA) reflectances to estimate GPP and RECO using a Leave-One-Out Cross-Validation (LOOCV) algorithm (Maxwell et al., 2018) and goodness-of-fit metrics. We focus our discussion on the approaches and datasets that are necessary to further improve geostationary satellite carbon cycle science with an eye toward developing real-time estimates of terrestrial carbon fluxes and their associated uncertainties.

Table 4.1. A list of abbreviations used in this study.

Abbreviation	**Full Form**
ABI	Advanced Baseline Imager
BRF	Bidirectional Reflectance Factor
CMI	Cloud and Moisture Imagery
DSR	Downwelling Solar Radiation
EC	Eddy Covariance
EDI	Environmental Data Initiative
GBR	Gradient Boosting Regression
GOES-R	Geostationary Operational Environmental Satellite-R Series
GPP	Gross Primary Productivity
IGBP	International Geosphere Biosphere Programme
LST	Land Surface Temperature
LOOCV	Leave-One-Out Cross-Validation
MODIS	Moderate Resolution Imaging Spectroradiometer
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
ML	Machine Learning
NN	Neural Network
NDVI	Normalized Difference Vegetation Index
NIRv	Near Infrared Reflectance of Vegetation
NIRvP	NIRv multiplied by PAR
PAR	Photosynthetically Active Radiation
R^2	Coefficient of Determination (R-squared)
RECO	Ecosystem Respiration
RF	Random Forest
RMSE	Root Mean Square Error
SVR	Support Vector Regression
SNR	Signal-to-Noise Ratio
TOA	Top-of-Atmosphere

4.2. Methodology

4.2.1. Overview

The study follows a structured approach consisting of four main steps, as illustrated in Figure 4.1. We initially collected data from two sources – Ameriflux eddy covariance towers and the GOES-R ABI – and then preprocessed the data as described in sections *2.2 Estimating GPP and RECO from eddy covariance data* and *2.3 GOES-R and the ABI*. Subsequently, we aligned the processed data in time and stored it in a repository with the Environmental Data Initiative (EDI), resulting in a unified, accessible dataset as described in Losos et al., (2024). We employed ML algorithms to establish relationships between synchronized GOES-R observations and flux data, including GPP and RECO, as detailed in section *2.4 Machine learning models*. In the final step, we trained the ML models and evaluated their performance using statistical metrics, specifically R^2 and RMSE, as described in section *2.5 Model development and performance assessment*. The performance of the best model was assessed across different vegetation cover and seasons, while also investigating the model's sensitivity to missing measurements of different lengths ('gaps') in the eddy covariance data records following Moffat et al. (2007).

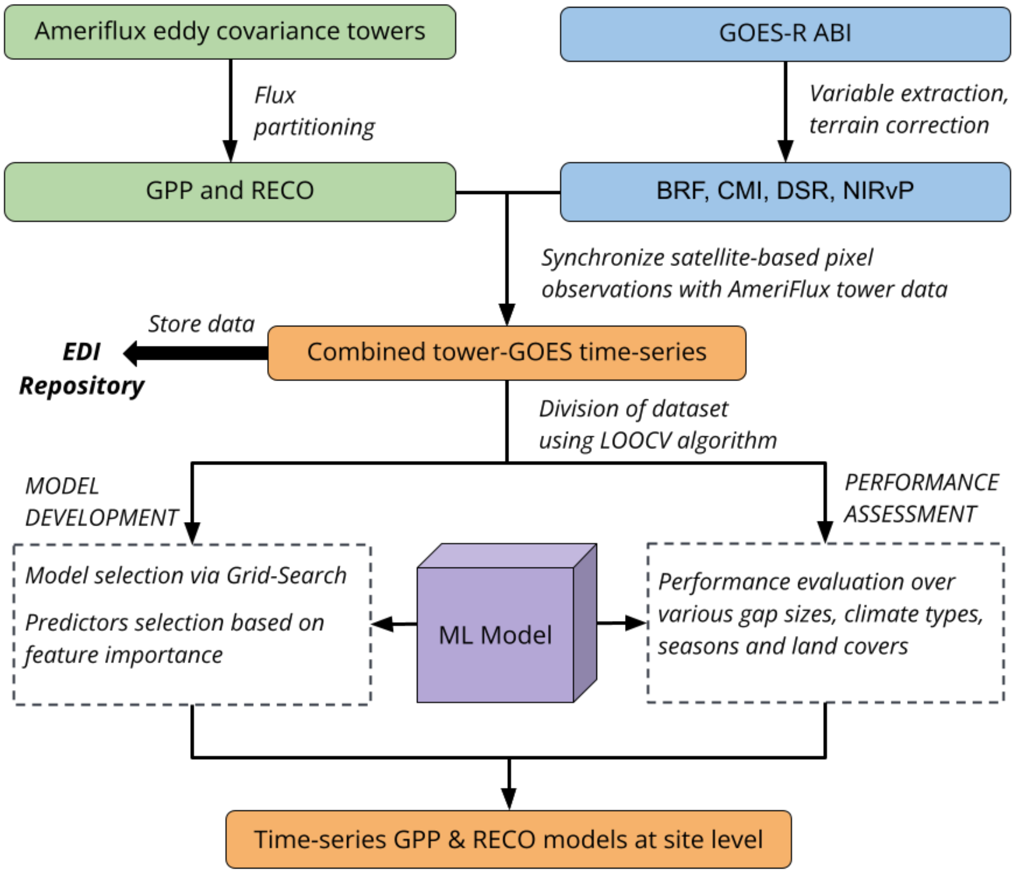

Figure 4.1. The workflow used in the present study.

4.2.2. Estimating GPP and RECO from eddy covariance data

From a total of 314 AmeriFlux and NEON eddy covariance towers discussed in Losos et al. (2024), we selected sites with a sufficient amount of eddy covariance data to partition carbon dioxide flux measurements into GPP and RECO for the period 2017 to 2022, which resulted in 126 sites (see Figure 4.2). We partitioned fluxes using the *hesseflux* library (Cuntz, 2020) in Python, which was created to process eddy covariance data in a similar way to the R-package REddyProc (Wutzler et al., 2018). *Hesseflux* is designed to perform the standardized processing steps described in Papale et al. (2006), for calculating half-hourly (or hourly) GPP and RECO and their uncertainty estimates, using both daytime (Lasslop et al., 2010) and nighttime (Reichstein et al., 2005) partitioning approaches.

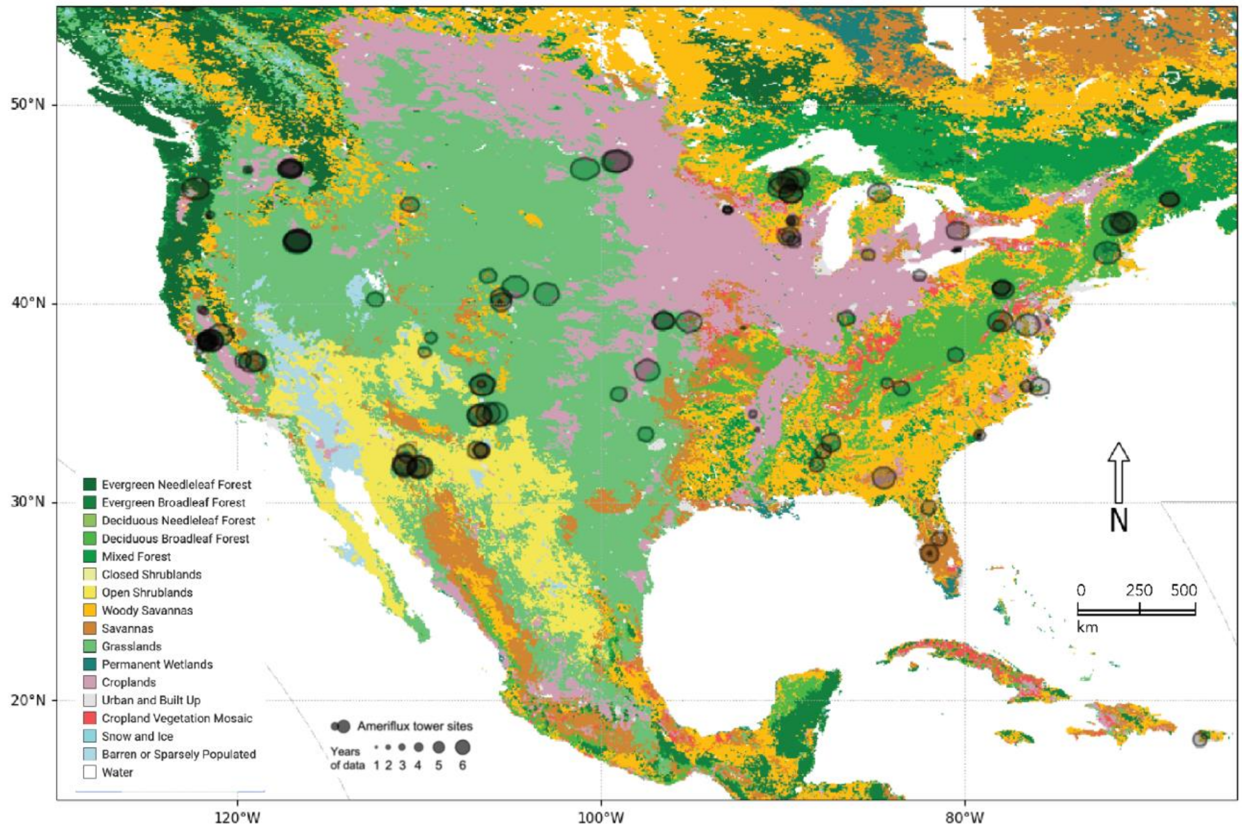

Figure 4.2. The spatial distribution of AmeriFlux and NEON eddy covariance sites used in this study, mapped over IGBP vegetation types for sites in the GOES-16 Full Disk projection.

This partitioning is completed after filtering the data for varying levels of data quality, including insufficient turbulence at night using a friction velocity (u*) filter (Reichstein et al., 2012a). We elected to exclude forest sites that were unable to provide the in-canopy CO_2 concentrations used to calculate the storage flux in addition to the eddy flux, as GPP and RECO may be underestimated when excluding carbon storage from forest site measurements (Fu et al., 2018). We describe the workflow used to partition GPP and RECO at the Ameriflux and NEON network scale in the Supplementary Material. In total, observations from 1,392,490 time points collected from the 126 sites dating back to 2017 were used in this study.

4.2.3. GOES-R and the ABI

The ABI measures radiation at 16 spectral bands from the visible to infrared (0.47 to 13.3 μm) spectra, with nominal spatial resolutions between 0.5 and 2 km at nadir. In the operational scan mode (Mode 6), the ABI images the Western Hemisphere (full disk) at 10 minute intervals, the Contiguous United States (CONUS) at 5 minute intervals, and 1000 × 1000 km mesoscale domains at 30 second intervals that can be repositioned to monitor key meteorological events (Schmit et al., 2017a; Schmit and Gunshor, 2020). We used observations from ABI aboard GOES-16 in the present analysis because it became operational first on December 18, 2017, and because GOES-17 was replaced by GOES-18 in 2022 due to overheating caused by a malfunctioning loop heat pipe (McCorkel et al., 2019). Prior to April 2019, the full disk was scanned every 15-minutes (Mode 3) rather than every 10-minutes. For the entire period, we averaged the two or three full disk measurements over the half-hour, and aligned the half-hourly ABI imagery with the half-hourly Ameriflux data. Additionally, we spatially aligned ABI pixels with tower locations by terrain-correcting to account for the parallax effect. Both data alignment processes are described in Losos et al., (2024). For further information about Level 1b and 2+ ABI products and their projections see Supplementary Material (Text S2).

In this study, besides the visible and NIR electromagnetic spectrum's surface products, including ABI bidirectional reflectance factors (BRF) bands 1, 2, and 3 (BRF1, BRF2, and BRF3 at 0.47 µm, 0.64 µm, and 0.86 µm), we also examined short-wave infrared (SWIR) bands to determine whether they might provide additional information that could be valuable for estimating GPP and RECO. Within the SWIR electromagnetic spectrum, the ABI sensor acquires two BRF bands: BRF5, Short-Wave Infrared 1 (SWIR1) at 1.6 µm, and BRF6, Short-Wave Infrared 2 (SWIR2) at 2.2 µm. Additionally, we used BRF2, BRF3, and DSR (Laszlo et al., 2020c) to calculate NIRvP following (Losos et al., 2024). We note that this approximation introduces uncertainty, primarily related to the conversion from the DSR product to PAR, which is influenced by various factors, including atmospheric conditions and spectral differences in solar irradiance (Dechant et al., 2022; Losos et al., 2024). However, the practical impact of these uncertainties is generally minimal (Hwang et al., 2022; Laliberté et al., 2016).

We further explored 16 TOA bands, including visible, NIR, and TIR, to analyze their impact on GPP and RECO estimation. Specifically, we focused on the significance of TOA infrared bands, influenced by water vapor, clouds, atmospheric, and land surface temperature, as discussed in previous studies (Ai et al., 2018, 2020; Yu et al., 2008, 2011; Yu & Yu, 2020). However, a challenge arose due to the differing time series lengths of BRF and CMI products. While BRF data became operational in August 2021, CMI records date back to July 2017. Despite BRF's robust measurements less influenced by atmospheric processes, its shorter archive lacks the historical record richness of CMI time series, impacting model training. Therefore, we explored both products individually and in combination within our ML approaches.

4.2.4. Machine learning models

We considered four commonly used ML regression models for handling nonlinear problems; Random Forest (RF), Gradient Boosting Regression (GBR), Support Vector Regression (SVR), and Multi-Layer Perceptron Neural Network (MLP). All of these models are widely employed for estimation problems in data-driven approaches (Diaz-Gonzalez et al., 2022; Thapa et al., 2023; Varghese et al., 2023).

RF and GBR are ensemble learning models that combine decision trees through bootstrapping and feature bagging (Friedman, 2001), and are frequently employed in remote sensing (Ranjbar et al., 2024; Sahin, 2020; Saini & Ghosh, 2017; Zhang et al., 2022). They offer feature importance measures, enabling us to assess the significance of different features. However, they differ in their approach and emphasis. RF focuses on simplicity and minimizing training set loss, while GBR puts more emphasis on optimizing a loss function to minimize errors by creating an ensemble of weak models, sequentially improving upon the mistakes made by previous models (Bentéjac et al., 2021; Sahin, 2020). RF and GBR both require hyperparameter tuning for optimized modeling. The number of trees in the ensemble determines the size and complexity of the model, and the maximum depth of each tree determines the depth of the decision trees in the ensemble (Sahin, 2020; Saini & Ghosh, 2017; Zhang et al., 2022).

SVR, a regression algorithm based on Support Vector Machines (SVM), identifies the best hyperplane in a high-dimensional feature space in order to maximize the margin between data points (Mountrakis et al., 2011; Smola and Schölkopf, 2004), minimizing prediction error

(Mountrakis et al., 2011). The SVR algorithm can model both linear and non-linear relationships using various kernels, including Radial Basis Function (RBF), sigmoid, polynomial, and linear. Key hyperparameters for SVR include penalty (“C”), gamma, epsilon, and, in the case of polynomial kernels, the degree parameter (Ranjbar et al., 2021). The C parameter controls regularization strength, gamma influences the shape of the decision function, epsilon sets the tolerance margin, and the degree determines the complexity of the polynomial kernel (Mountrakis et al., 2011).

Lastly, MLP is an artificial neural network that learns complex patterns and relationships through backpropagation, making it a flexible tool for approximating non-linear relationships (Baraldi et al., 2001; Gardner and Dorling, 1998; Ranjbar et al., 2021; Suliman and Zhang, 2015). It consists of multiple layers of interconnected “neurons” and utilizes non-linear activation functions to model intricate relationships between inputs and outputs (Baraldi et al., 2001; Suliman and Zhang, 2015). MLP also has hyperparameters, including the number of hidden layers and neurons, activation functions, learning rate, and regularization. The hidden layers and neurons determine the network's complexity, activation functions control nonlinear mapping, the learning rate affects convergence speed, and regularization prevents overfitting (Shirmard et al., 2022).

4.2.5. Model development and performance assessment

We used a grid search algorithm to ensure optimal performance of the models, aiding in the identification of the most suitable configuration for maximizing predictive accuracy (Maxwell et al., 2018; Varghese et al., 2023). Table 4.2 presents the search area for hyperparameter settings of each ML model. To train and test the ML models, we adopted a Leave-One-Out Cross-Validation

(LOOCV) algorithm and two key statistical metrics, namely the coefficient of determination (R^2, Eq. 4.1) and root mean squared error (RMSE, Eq. 4.2).

$$R^2 = 1 - \frac{RSS}{TSS}, \qquad \text{Eq. 4.1}$$

where RSS is the sum of squares of residuals and TSS is the total sum of squares.

$$RMSE = \sqrt{\frac{\sum_{N}^{i=1}(x_i - \hat{x})}{N}}, \qquad \text{Eq. 4.2}$$

where i variable i, N is the number of non-missing data points, x_i are actual observations in the time series, and $\hat{x}$ is the estimated time series.

In LOOCV, 20% of the available data was held out exclusively for testing purposes. Subsequently, the remaining data underwent a four-fold cross-validation process, increasing the models' generalization capabilities (Maxwell et al., 2018). To further enhance the reliability and robustness of the results, the algorithm was rerun ten times, and the results were averaged for assessment. We used the *Scikit-learn* python library to implement the grid search algorithm, ML models, and LOOCV algorithm (Pedregosa et al., 2011).

Table 0.2 Grid search specifications and searching range for hyperparameters of the machine learning models.

Model	**Hyperparameter**	**Search Range**
RF	Number of estimators	100, 300, 500, 800, 1000
	Maximum depth	5, 10, 15, 20, 50, 100
GBR	Number of estimators	50, 100, 200, 400
	Maximum depth	3, 5, 8, 10
	Learning rate	0.01,0.05,0.1
MLP	Activation	'logistic', 'tanh', 'relu'
	Solver	'adam','sgd'

	Hidden layer	2, 3, 5
	Hidden layer size	50, 100, 200, 500
	Alpha	0.0001, 0.001, 0.01, 0.1
SVR	Penalty	0.1, 10, 100, 1000
	Epsilon	0.001, 0.01, 0.2
	Gamma	0.01, 0.1, 0.5,1,10
	kernel	'RBF', 'sigmoid', 'poly', 'linear'

We further evaluated our model performance across diverse ecological and seasonal conditions, namely the International Geosphere Biosphere Programme (IGBP) land cover types (Sulla-Menashe et al., 2019) and month of the year (De Beurs and Henebry, 2005; Sulla-Menashe et al., 2019). To do so, we calculated a signal-to-noise ratio (SNR) for each month. SNR was defined as the mean GPP and RECO values divided by the mean absolute error (MAE), representing the clarity of the signal in the presence of noise. We then converted SNR to decibels (dB) units using multiplying the logarithm with base 10 by 10. Positive SNR values indicate that the signal is stronger than the noise, while negative SNR values suggest that the noise is stronger than the signal.

In addition to evaluating model performance across diverse ecological and seasonal conditions, we also assessed the models' adaptability under different gap sizes to test their suitability for gap filling purposes (Gao et al., 2023; Zhu et al., 2022). We chose 12 gap lengths ranging from one hour to one year and randomly selected data from our available dataframe for each gap length from all sites. We then kept these data out of training for testing purposes and trained the ML model using the rest of the data. Lastly, we compared the diurnal dynamics of GPP estimated from ABI observation against partitioned from tower measurements. Given RECO's minimal diurnal variation, we only presented the model performance for GPP, and RECO results are presented in

Supplementary Materials. For this analysis, we selected four AmeriFlux sites, focusing separately on GPP and RECO, ensuring that the chosen sites exhibited diurnal variability in the DOY during the year 2022 with minimum, median, and maximum values.

4.3. Results

4.3.1. Assessment of the machine learning models performance

Figure 4.3 shows the comparison of the GPP and RECO modeled using four ML algorithms by averaging the results of ten runs of the LOOCV algorithm using surface products (i.e. BRFs and DSR) as inputs for ML models. As GOES-16 surface product data are only available starting in August 2021, we utilized data from August 2021 through December 2022 from a total of 89 sites with available eddy covariance observations. RF and GBR demonstrated higher performance for both GPP and RECO estimation with statistics presented in Table 4.3. MLP exhibited the worst performance, and SVR was not as effective as the tree-based learning models (Table 4.3).

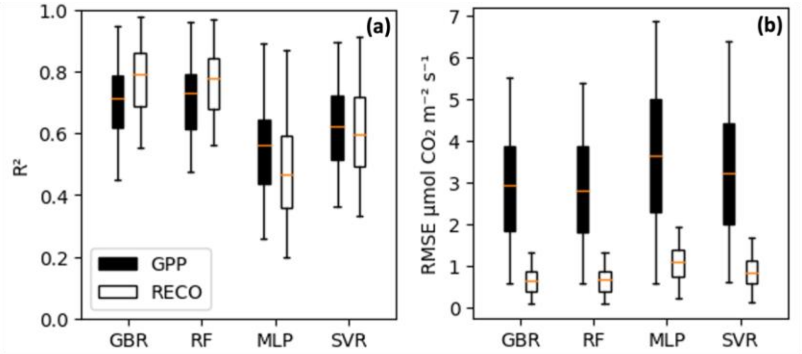

Figure 4.3. GPP and RECO estimation with four machine learning models using GOES-R surface reflectance products. (a) R^2 and (b) RMSE, based individual models trained on two years of data from 89 eddy covariance sites. The median is depicted on the box plots.

With a mean (± standard deviation) R^2 value of 0.71 ± 0.11, and corresponding RMSE of 4.04 ±1.65 µmol m^{-2} s^{-1}, RF estimated GPP with slightly better fit than GBR, which had a mean R^2 of 0.70 ± 0.12, and a slightly higher RMSE of 4.15 ± 1.73 µmol m^{-2} s^{-1}. For RECO estimation, similar trends emerged as observed in GPP estimation; RF and GBR continued to exhibit better performance with R² of 0.77 ± 0.10 and RMSE of 0.90 μmol m^{-2} s^{-1}. Despite this equivalency in accuracy, GBR exhibited notable advantages in terms of speed, with faster model training and prediction times, as well as a smaller model size compared to RF.

Table 0.3. R² and RMSE for GPP and RECO estimation using two years of data from 89 eddy covariance sites with multiple ML models (Mean ± Standard Deviation).

	GPP				**RECO**			
	Test		Cross validation		Test		Cross validation	
	R²	RMSE	R²	RMSE	R²	RMSE	R²	RMSE
RF	**0.71 ± 0.11**	4.04 ± 1.65	0.68 ± 0.12	3.28 ± 1.74	**0.77 ± 0.10**	0.90 ± 0.52	0.74 ± 0.10	0.77 ± 0.50
GBR	**0.70 ± 0.12**	4.15 ± 1.73	0.68 ± 0.13	3.34 ± 1.81	**0.77 ± 0.10**	0.90 ± 0.49	0.75 ± 0.11	0.76 ± 0.47
MLP	0.55 ± 0.15	5.23 ± 2.06	0.55 ± 0.15	4.20 ± 2.20	0.56 ± 0.16	1.42 ± 0.60	0.49 ± 0.20	1.20 ± 0.60
SVR	0.62 ± 0.14	4.70 ± 2.09	0.61 ± 0.13	3.81 ± 2.14	0.61 ± 0.15	1.18 ± 0.60	0.62 ± 0.14	0.99 ± 0.59

Results from tests and cross-validation can assess model performance and generalization across datasets. All models exhibit close test and cross-validation results, indicating comparable predictive ability for both GPP and RECO. RF and GBR have slightly higher R² values in the test set (0.71 ± 0.11 and 0.70 ± 0.12) compared to cross-validation (0.68 ± 0.12 and 0.68 ± 0.13), suggesting an increased variance in predicting GPP for test data.

4.3.2. Different ML models response to diurnal variability

Figure 4.4 presents the results of assessing ML models for estimating diurnal GPP dynamics, comparing time series EC measurements with GOES-estimated values using the ML models in three days during 2022 at four AmeriFlux sites. The three days chosen for this analysis represent days with minimum, median, and maximum GPP values at the AmeriFlux sites in 2022. This selection ensures that the analysis captures the full range of GPP variability. Since RECO exhibits little diurnal variation, we focused on evaluating the model's performance for GPP at the half-hourly scale. Findings for GPP indicate that, among the ML models, GBR and RF exhibit the highest agreement, while MLP displays the least. MLP and SVR consistently underestimate the maximum GPP values at AmeriFlux sites US-xTR and US-Rms, with more pronounced underestimation at US-Tw4. Additionally, MLP and SVR overestimate the minimum GPP values at US-Rms and US-Tw4. All models excel in estimating peak GPP values around 12 PM (local standard time) across the four sites, with GBR emerging as the most accurate model. For RECO, GBR and RF also perform better than MLP and SVR across all sites and for minimum, median, and maximum values. Similar to GPP, MLP and SVR consistently have underestimation and overestimation problems.

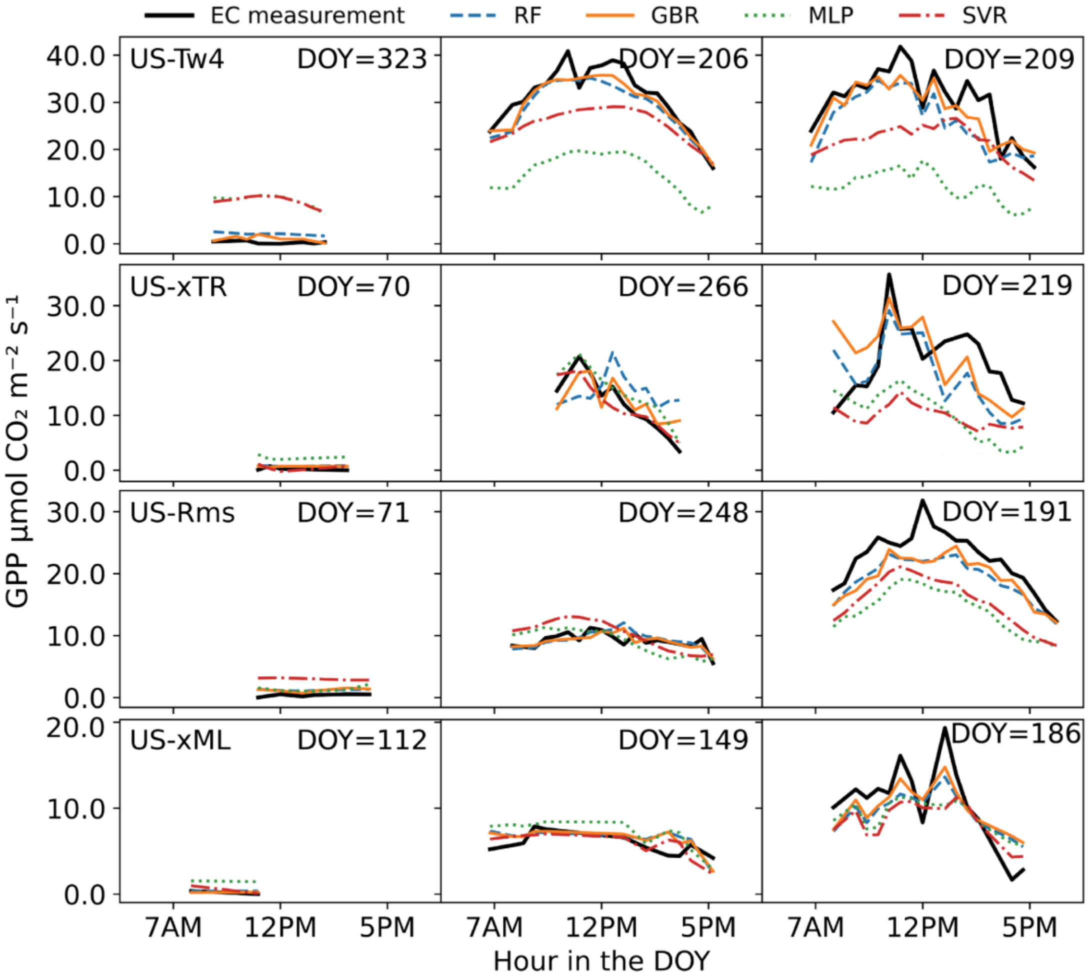

Figure 4.4. Time series GPP estimated at AmeriFlux towers using Eddy Covariance (EC) measurements versus their corresponding GOES-estimated values using machine learning models on three DOY with minimum, median, and maximum GPP values at the AmeriFlux sites in the year 2022. The AmeriFlux sites are labeled in the top-left corner of the plots.

4.3.3. Feature importance

NIRvP, NIRv, and BRF3 emerged as the most important features in the GBR model (Figure 4.5), followed by DSR and BRF6 (SWIR 2.2μm). BRF5 (SWIR 1.6μm) and BRF1 (Blue) were the least important features of those selected. DSR showed its importance after NIR-based features, even though NIRvP, the top-ranked feature, was derived using PAR, which is computed by DSR.

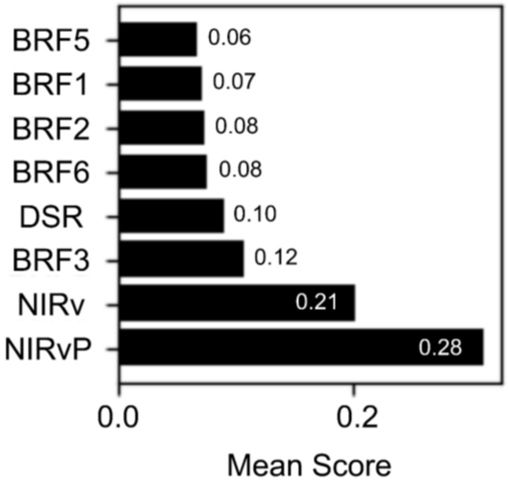

Figure 4.5. Feature importance from the selected GBR model trained on GPP estimates from 89 eddy covariance sites. Abbreviations are listed in Table 4.1.

Further investigation for optimal model input led us to train and test three different GBR models using only BRF features, only CMI features, and a combination of BRF and CMI features separately. Comparing the results shows that the best feature combination for modeling GPP is the combination of BRF and CMI, yielding an R^2 score of 0.69 ± 0.12 and an RMSE of 2.92 ± 1.27 $\mu mol\ m^{-2}\ s^{-1}$ (Figure 4.6). The results for RECO are similar to those for GPP. The combination of BRF and CMI exhibits the best R^2 score of 0.74 ± 0.11 with a RMSE of 0.67 ± 0.31 $\mu mol\ m^{-2}\ s^{-1}$.

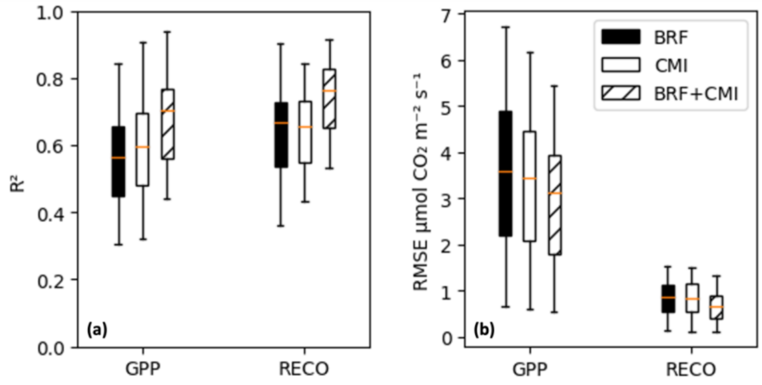

Figure 4.6. A comparison of (a) R^2 and (b) RMSE for GPP and RECO estimation using up to two years of observations from 89 eddy covariance sites and the selected GBR model using different BRF and CMI feature configurations. The median is depicted on the boxes.

When considering only surface reflectance, the most influential features are BRF3 (NIR), BRF2 (red), and BRF6 (SWIR). Conversely, when solely utilizing TOA observations, the prominent features become CMI_C3 (NIR), CMI_C12 (IR), and CMI_C6 (SWIR). Intriguingly, in the context of combined feature utilization, surface reflectance features retain their significance, outweighing TOA observations in importance for the three most important variables. However, it is noteworthy that TOA observations still maintain relatively high feature scores, with BRF3, BRF2, BRF6, CMI_C12, CMI_C3, and CMI_C6 emerging as the foremost informative features (feature importance figures are provided in the Supplementary Material, Section S2)

4.3.4. The impacts of longer historical records on model performance

Using the CMIs features, which extend back to 2017 from GOES-16, we evaluated the impact of longer data records on predicting GPP and RECO. Our analysis, depicted in Figure 4.7, shows that by extending the historical data for training the GBR model, we can consistently improve the predictive accuracy.

We considered historical data intervals of 1, 2, 3, and 4 years for training, ranging from 2020-2021 to 2017-2021. Comparing these models to those trained solely on 2021 data, we found that incorporating longer historical records yielded a steady increase in accuracy. The R² increased by up to 0.09 ± 0.07, indicating better variance explanation, while the RMSE decreased by down to -0.11 ± 0.17 µmol m^{-2} s^{-1}, highlighting improved accuracy.

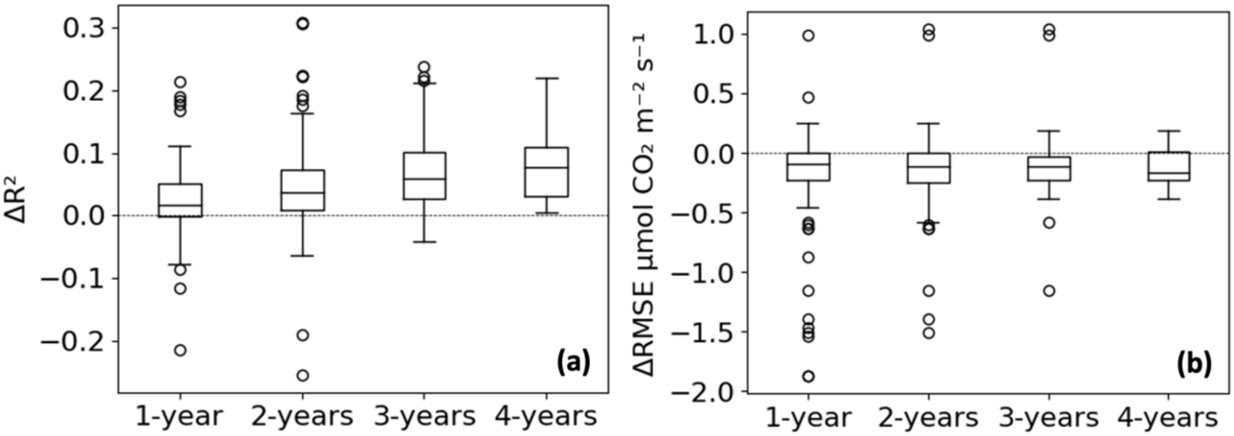

Figure 4.7. Impact of historical GOES-R CMI data record length on GPP and RECO estimation accuracy using a GBR model trained on eddy covariance data from 126 sites: (a) ΔR² and (b) ΔRMSE differentiate the accuracy of models trained using 1, 2, 3, and 4 years more data than only one year data. The median is depicted on the boxes.

4.3.5. Sensitivity of models performance to seasonal variations

For GPP estimations, July stands out as the month with the highest absolute uncertainty with a standard deviation of 3.30 µmol m^{-2} s^{-1}. In contrast, January emerges as the month with the lowest absolute uncertainty, having a lower standard deviation of 1.22 µmol µmol m^{-2} s^{-1}. The SNR analysis reaches a peak in July (SNR=8.13), signifying a relatively robust and reliable signal compared to January with the lowest SNR at 4.47, indicating a weaker signal relative to the background noise. Similarly, August is the month with the highest estimation error for RECO, with a standard deviation of 0.773 µmol m^{-2} s^{-1}. Conversely, January emerges with lowest estimation errors, characterized by standard deviation of 0.296 µmol m^{-2} s^{-1}. Similar to GPP, the SNR value for August was the highest at 9.53, indicating a notably stronger signal during this month compared to the lowest SNR of 6.94 observed in January.

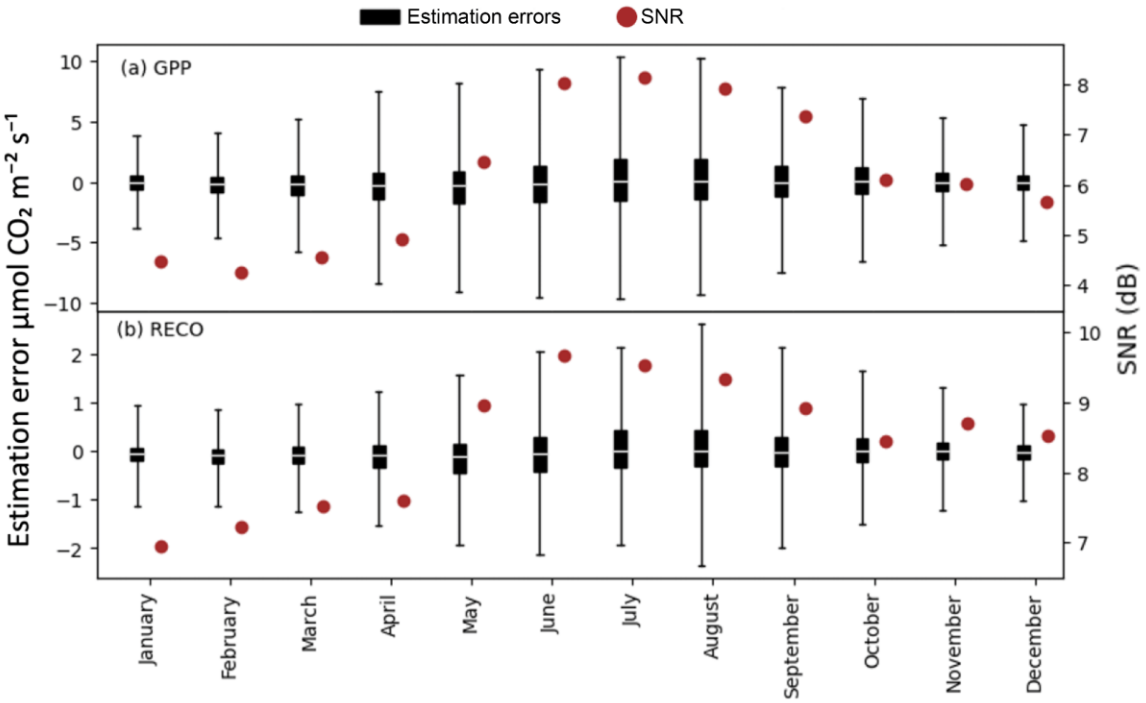

Figure 4.8. The monthly variability in estimation errors and signal-noise ratio (SNR) for (a) GPP and (b) RECO grouped by month from the GBR model.

Examining flux estimation across seasons reveals important insights into the model's performance. In winter, the standard deviation of GPP and RECO estimation errors is relatively low, at 1.31 and 0.299 μmol m^{-2} s^{-1}, respectively. This increases in spring, indicating more variability in model estimations. During summer, the standard deviation peaks at 3.20 and 0.726 μmol m^{-2} s^{-1} for GPP and RECO, respectively. In autumn, the standard deviation decreases again. The mean SNR values for each season reflect this trend, with lower values in winter and spring, and higher values in summer and autumn for both GPP and RECO.

4.3.6. Sensitivity of model performance to vegetation type and magnitude of GPP

Our analysis reveals notable variations in the model's performance for GPP estimation across different vegetation types (see Figure 4.9). Mixed Forests and Permanent Wetlands emerge as the top-performing categories, exhibiting a mean R^2 value of 0.78 for both. Conversely, Savannas display the lowest mean R² value, at 0.64, followed by Shrublands with an R² of 0.68 (Tables S3 and S4). Moving on to other vegetation categories, R² for GPP estimation is > 0.72, where the lowest is for Evergreen Needleleaf.

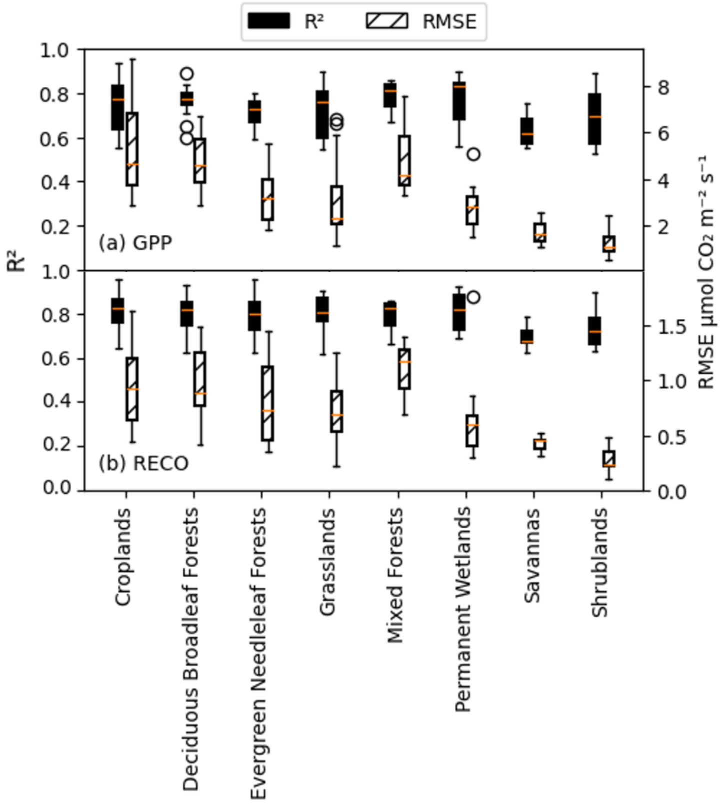

Figure 4.9. Performance of the GBR model trained on data from 89 eddy covariance sites for (a) GPP and (b) RECO estimation, employing a combination of BRF and CMI features across various IGBP vegetation types.

Similar trends are observed in the model's performance for RECO estimation. Permanent Wetlands, Deciduous Broadleaf Forests, Croplands, Evergreen Needleleaf Forests, Grasslands, and Mixed Forests have the highest mean R² value of 0.80. On the contrary, Savannas exhibit the lowest mean R² value for RECO, at 0.69, pointing to relatively reduced accuracy but still predictive skill even in this challenging vegetation type. Similar to GPP estimation, Shrublands showed the lowest accuracy after Savannas in RECO estimation, with mean R² values of 0.73.

We further investigated the sensitivity of model performance with respect to the magnitude of (Figure 4.10). The best GPP estimation was achieved for GPP values between 35 and 52 µmol µmol m^{-2} s^{-1}, followed by values between 23 and 35 µmol µmol m^{-2} s^{-1}. The poorest performance was observed at sites with the lowest and very highest GPP (between 5 to 23 and 52 to 52< µmol m^{-2} s^{-1}, respectively). For RECO estimation, a similar trend was observed; sites with intermediate values exhibited the best performance, while those with the highest and lowest biomass demonstrated lower R² values.

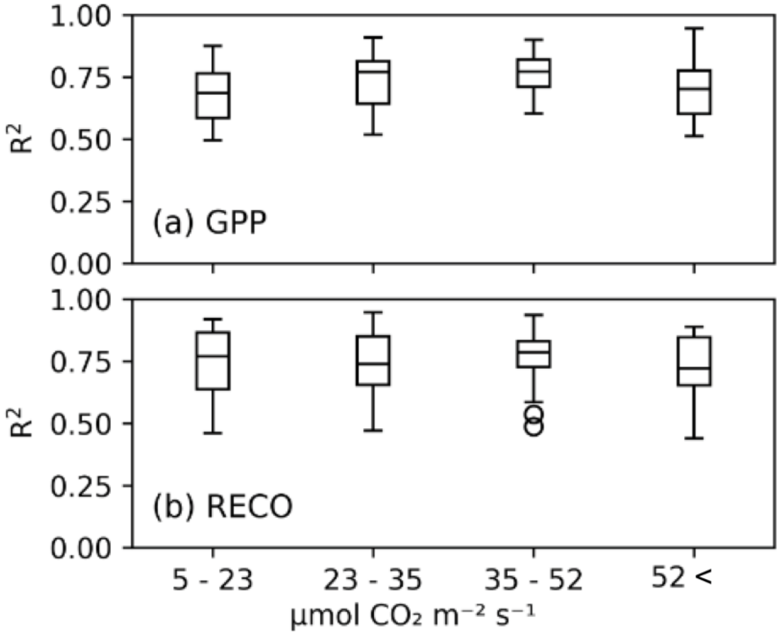

Figure 4.10. Performance of the GBR model trained on data from 89 eddy covariance sites for (a) GPP and (b) RECO estimation across different GPP bins.

4.3.7. Models performance for gap filling

The accuracy of our GBR model for gap filling over various gap lengths reveals variations in accuracy relative to gap length (Figure 4.11). As expected, the one-, two-, and four-hours intervals exhibit the highest mean R² of 0.70 for GPP and 0.75 for RECO (Tables S5 and S6) and gap sizes greater than one week exhibit the lowest mean R².

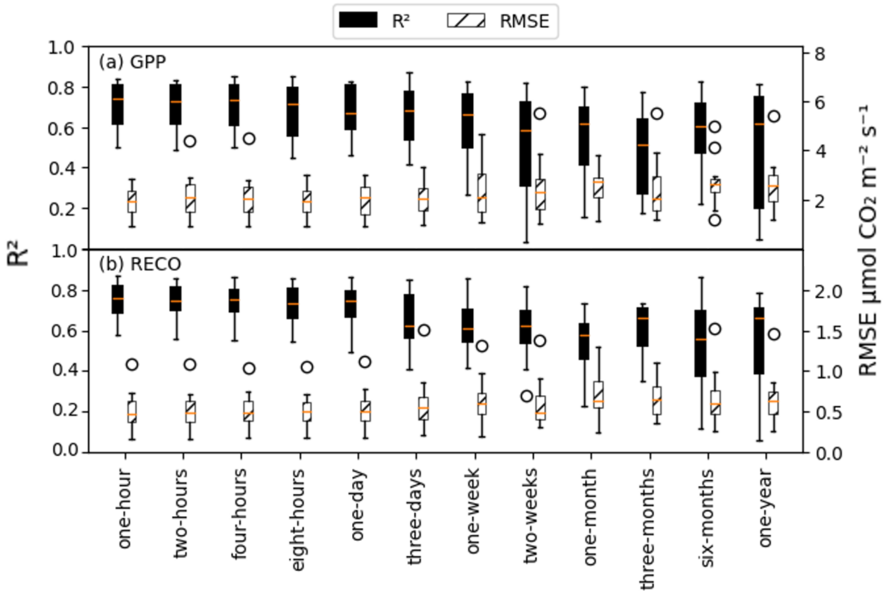

Figure 4.11. The GBR model performance for (a) GPP and (b) RECO gap filling for different gap lengths.

Figure 4.11 shows that accuracy gradually decreases as the gap length increases beyond one day, with a more pronounced decline observed after the three-day interval. However, according to GPP results, the maximum R^2 values of 0.82 and 0.81 for six-month and one-year intervals, respectively, demonstrate that the model maintained accuracy, even for longer gap sizes. The same trends are observed for RECO, with maximum values of R^2 of 0.86 and 0.79 for six-month and one-year intervals (Tables S5 and S6).

4.4. Discussion

4.4.1. Performance of the machine learning models

Our analysis of ML models with ABI data reveals that ensemble methods like RF and GBR outperform individual models like MLP and SVR. This underscores the need for ensemble techniques to capture complex relationships between surface-atmosphere fluxes and GOES-R data, aligning with past research favoring ensemble models in remote sensing applications (Yu et al., 2021; Zhang et al., 2022). For example, Yu et al. (2021) compared five ML models, including RF, SVR and ANN, to upscale GPP using MODIS observations. Their results highlighted the superiority of RF, achieving the highest accuracy (R^2 of 0.86 and RMSE of 0.99 g C m^{-2} d^{-1}) compared to the other models. Additionally, Li et al., (2023) used a Cubist regression tree model with GOES-R data to predict daily GPP variations during the 2020 U.S. heatwave. They found the regression tree effective in forecasting carbon-related variables, emphasizing the importance of multiple factors in GPP estimation. This improvement is likely due to ensemble learning's ability to merge predictions from multiple models, reducing variance and enhancing accuracy (Sagi & Rokach, 2018; Zhang et al., 2022).

In ensemble models like RF and GBR, decision trees are used for predictions. GBR stands out for employing a boosting algorithm to refine predictions. Ensemble models excel at managing missing data and noise (Kursa, 2014), common in ecological datasets (e.g. Moffatt et al., 2007). RF and GBR handle missing data by assigning weights to observed variable values' frequencies based on their proximity. This technique benefits from post-training information from the mean-imputed dataset (Ezzine and Benhlima, 2018). These learners are skilled at managing noisy data (Elbeltagi et al., 2023) by averaging estimations from multiple decision trees, reducing the impact of noise on overall performance (Jun, 2021).

Comparing ML approaches for time series analysis of diurnal GPP and RECO dynamics reveals similar trends. GBR and RF consistently exhibit the highest agreement, while MLP shows the least agreement, suggesting limitations in its ability to predict diurnal carbon flux. Notably, MLP and SVR consistently underestimate maximum values and overestimate minimum values for both GPP and RECO, which is a common challenge in ML modeling when estimating small and large values in the trained dataframe (Belitz and Stackelberg, 2021; Ye et al., 2021). Despite these nuances, all models excel in accurately estimating peak GPP values around 12 PM, with GBR emerging as the most precise. However, all models demonstrate fluctuations in RECO estimation, attributed to the use of highly dynamic satellite observations as modeling inputs and the limited variability of tower estimated values through partitioning EC measurements.

GBR demonstrates performance similar to RF but with a smaller digital model size and faster prediction time, making it favorable for data-intensive geostationary workflows. Although GBR models require more construction time than RF models due to tuning additional hyperparameters and the sequential building process (Golden et al., 2019), they often exhibit superior predictive capabilities (Golden et al., 2019; Shen & Yong, 2021; Zhang et al., 2022). By optimizing a reduced set of predictors and the model's maximum depth, GBR achieves comparable predictive accuracy to RF models but with fewer trees and maximum depths (Zhang et al., 2022), leading to expedited training and a more compact digital model size.

4.4.2. Feature importance

The feature importance analysis in this study highlights NIRvP, NIRv, and BRF3 as crucial factors influencing GPP and RECO estimation using GOES-R measurements and data products (Khan et al., 2022; Todd & Hoffer, 1998). These features, all derived from the NIR spectral band, underscore the significant role of vegetation reflectance in carbon exchange processes within ecosystems (Khan et al., 2022; Richardson et al., 2013). Baldocchi et al., (2020) further demonstrate a robust correlation between reflected near-infrared radiation and canopy photosynthesis, emphasizing the effectiveness of near-infrared radiation as a cost-effective tool for monitoring carbon assimilation across various conditions and time scales. Following these features, DSR is vital in estimating GPP and RECO as it serves as a proxy for PAR—the solar radiation essential for photosynthesis (Dechant et al., 2022; Khan et al., 2022; Wohlfahrt et al., 2010). DSR allows assessment of available light energy, positively impacting GPP when stressors are absent, enabling plants to harness more energy for photosynthesis (Bosiö et al., 2014; Oliphant et al., 2011). Although the relationship between PAR and RECO is more nuanced, it is influenced by GPP, which, in turn, is driven by PAR (Austin and Vivanco, 2006; Jia et al., 2020).

After DSR, SWIR2 at 2.2 μm (BRF6) stands out as a crucial feature, emphasizing the importance of shortwave irradiance and spectral bands for vegetation functioning. The 2.2 μm channel aids atmospheric correction, reducing noise and atmospheric interference (Hese et al., 2005). SWIR-VIs, particularly when creating VIs using SWIR and NIR bands, perform comparably to VNIR-VIs (creating VIs using visible and NIR bands) and excel in certain cases for vegetation monitoring (Benseghir and Bachari, 2021; Massetti et al., 2019). For instance, SWIR-VIs explain 40% more variability than VNIR-VIs in aboveground biomass estimation (Benseghir and Bachari, 2021). SWIR-VIs have also been used for understanding the impacts of snow, phenological change, and water availability on terrestrial ecosystem functioning. Whereas the particular mechanism by which the GBR model identifies BRF6 as an important input is difficult to ascertain from ML approaches, and a combination of the above factors is likely the cause of its importance at ecosystems distributed across CONUS, our findings suggest that incorporating SWIR bands could significantly improve the accuracy of geostationary satellite-based carbon cycle products.

Three distinct GBR models constructed by different combinations of TOA and BRF features shows that the combination of both BRF and CMI features demonstrates superior performance over models utilizing either TOA products or surface reflectances alone. However, we acknowledge the complexity of converting TOA to surface reflectance, especially when correcting for BRDF effects, which presents significant challenges in accurately capturing surface properties. Achieving good sub-daily estimations of GPP and respiration using only TOA observations demonstrates the potential for simpler and more efficient remote sensing methods. CMIs show comparable performance to BRFs, highlighting their potential as an alternative data source in carbon cycle science. While TOA information is less utilized in vegetation science (Bayat et al., 2020; Kovács et al., 2023; Parente, 2013; Yuanyong and Shenghui, 2013), we argue that it holds promise for enhancing geostationary satellite-based carbon cycle products. For training and testing, our study used filtered data to exclude low-quality data using ABI quality flags. Further research is needed to evaluate the utility of TOA observations under various data quality conditions, including cloudiness and optically thick aerosol atmospheres. The analysis highlighted the importance of surface reflectance features, even when combined with TOA observations, while the thermal infrared band CMI_C12 also showed significance, warranting further investigation. While the specific reason for CMI_C12's explanatory power in machine learning models remains unclear, it is discussed in more detail in the Supplementary Materials.

4.4.3. Longer historical record effects on model performance

By assessing the influence of extending historical data records on the estimation accuracy of GPP and RECO, we have demonstrated that extended historical data better explains their variance. Several studies have shown the positive impact of longer data records on model performance in ecological and environmental modeling (Han et al., 2023; Houghton et al., 1983; Nemani et al., 2003). Furthermore, historical remote sensing and ground-based data improve predictive power for greenhouse gasses and provide historical flux data, establishing reference emission levels adjusted for national circumstances (COP15—Copenhagen, 2009). This enhanced understanding can contribute to more informed decision-making and policy formulation regarding carbon sequestration, ecosystem health, and climate resilience (Anderegg et al., 2020; COP15—Copenhagen, 2009; Millar et al., 2007).

4.4.4. Sensitivity of the model performance to seasonal variability

Assessment of model performance to seasonality offers valuable insights into the model's accuracy and consistency across different seasons, which align with established ecological patterns. July, characterized by the highest standard deviation in error estimations and SNR for GPP, corresponds to the peak of Northern Hemisphere summer when factors such climate and weather anomalies including temperature, water stress, precipitation, and photosynthesis complicate accurate estimation of GPP from satellite remote sensing (Felton and Goldsmith, 2023; Wei et al., 2017). In contrast, the lowest estimation errors and SNR values in winter are likely a result of reduced environmental variability and near-zero photosynthesis in many ecosystems during this season (Irisarri et al., 2012; Qiu et al., 2020). The contrast between high summer SNR and low winter SNR values suggests that the discernibility of the carbon flux signal diminishes in winter due to environmental factors. This makes winter data less reliable compared to summer. Transitioning to spring, there is a noticeable surge in variability in estimation errors compare to winter, which may be attributed to the onset of new growth and changing environmental conditions (Weltzin et al., 2020), and result in spatial heterogeneity in carbon fluxes at fine spatial scales that challenge the relatively coarse pixels of geostationary satellites (Khan et al., 2022).

Similar to GPP, RECO estimations exhibit seasonal variations. The highest estimation errors and SNR values during summer months can potentially be explained by the intricate interplay of temperature, moisture, and plant activity in this season (Felton and Goldsmith, 2023; Gong et al., 2022; Younes et al., 2020). During winter, RECO estimates exhibit lower errors and SNR values

(Figure 4.8), reflecting the impact of reduced environmental variability (at least under snow) and, often, near-zero photosynthesis to which RECO is coupled. These lower errors are attributed to smaller RECO values and reduced standard deviation during winter months (Gong et al., 2022; Irisarri et al., 2012; Wei et al., 2017). The increase in variability during spring compared to winter (Figure 4.8) aligns with the transition period when vegetation begins to emerge across much of the study domain, resulting in rapid increases in both GPP and RECO. This growth and heightened photosynthetic activity also contribute to an increase in SNR values, indicating stronger signals over background noise for carbon flux data in spring (Gong et al., 2022; Younes et al., 2020). These findings underscore the importance of seasonality in model skill, requiring further research and model refinement during variable periods (Gong et al., 2022; Wei et al., 2017; White and Hastings, 2020).

4.4.5. Sensitivity of the model performance to vegetation type and magnitude of GPP

Our analysis of the GBR model's performance across various vegetation types has revealed variations in the accuracy of GPP and RECO estimation (Figure 4.9). Mixed Forests and Permanent Wetlands consistently stand out as top-performing categories, while Savannas and Shrublands exhibit comparatively lower predictability. This differential performance is critical to understanding the model's adaptability within diverse ecological settings.

The challenges faced in predicting GPP and RECO in Savannas and Shrublands arise from their inherent complexity and environmental variability. Savannas encompass a mix of grasses and scattered trees, resulting in intricate and fluctuating carbon exchange dynamics that are challenging to model accurately (Littlefield and D’Amato, 2022; Mistry, 2000). Similarly, Shrublands comprise a mix of shrubs and grasses, leading to high spatial and temporal variability in carbon exchange processes which also makes them challenging to model (Li et al., 2021). Moreover, the reduced estimation accuracy within Shrublands and Savannas can be attributed in part to the sub-pixel variability in green-up (Gessner et al., 2013; Ma et al., 2013), resulting in fluctuating NIR reflectance and challenges in remote sensing modeling.

Deciduous Broadleaf Forests, Croplands, Evergreen Needleleaf Forests, and Grasslands demonstrate a similar degree of predictability, the model's performance varies across these vegetation types. The forested vegetation types represent diverse ecosystems with distinct physiological characteristics (Zhu et al., 2023) that the GBR model appears to simulate well in Deciduous Broadleaf Forests, with poorer average fit in Evergreen Needleleaf Forests that consistently challenge remote sensing modeling of their carbon cycle (Liu et al., 2016).

The variations in model performance among these vegetation types – when using satellite data products alone – can be speculated to a combination of factors such as environmental conditions, vegetation structure, and ecological dynamics. Deciduous Broadleaf Forests often exhibit well-defined seasonal patterns of carbon exchange, making them more predictable in this regard (Muraoka, 2022) but noting that phenological transitions remain difficult to model (Richardson et al., 2013). Evergreen Needleleaf Forests have more subtle changes in greenness and canopy structure, Grasslands may have different ecological dynamics including rapid recovery from disturbances and fine-scale heterogeneity (Novick et al., 2004), which can be difficult to estimate from space (Chen et al., 2023; Qiu et al., 2020).

4.4.6. The model performance for gap filling missing eddy covariance data

The eddy covariance method is the main monitoring tool for measuring the net ecosystem exchange (NEE), defined as the net flux of CO2, balancing ecosystem respiration (release, RECO) and photosynthesis (uptake, GPP) (Gao et al., 2023; Zhu et al., 2022). These measurements are reported on a half-hourly or hourly basis. However, calibrations, equipment failures, and low turbulence conditions often result in gaps in these data series, with 20-60% or more of annual data typically missing (Irvin et al., 2021; Moffat et al., 2007; Zhu et al., 2022). Testing the models' adaptability under different gap sizes is crucial for ensuring the reliability and accuracy of gap-filling techniques. Continuous NEE datasets are essential for estimating ecosystem carbon budgets, evaluating process model predictions, and comparing with biometric measurements (Falge et al., 2001). In the domain of strategies for filling gaps in C-cycle flux time series, a diverse array of approaches including non-linear and linear regression techniques, semi-parametric models, marginal distribution sampling, look-up tables, and the application of ML models has been widely utilized (Irvin et al., 2021; Y. Kim et al., 2020; Moffat et al., 2007; Zhu et al., 2022). Nevertheless, a nuanced comparison between these approaches showed that ML models, including NNs, RF, and SVR, highlight their superiority over alternative approaches (Gao et al., 2023; Irvin et al., 2021; Y. Kim et al., 2020). This study also investigates the potential of the GBR model for gap-filling purposes. The model exhibits higher accuracy over shorter gap lengths (e.g., one hour to three days), benefiting from more complete training data for three consecutive days (Figure 4.11), which facilitates effective learning (Chen et al., 2018; Maxwell et al., 2018). Shorter intervals provide less fragmented data, enhancing the GBR model's ability to capture short-term variations and dependencies. The high level of accuracy achieved for gap sizes of up to one day is promising (Figure 4.11, and Tables S5 and S6 shows R^2 of 0.75 for RECO and 0.7 for GPP), as it demonstrates the model's capability to effectively fill approximately 99% of the gaps in time series data at the site-level observation, as depicted in Figure 4.11. Nevertheless, the model maintains acceptable performance in longer gap lengths, with an R^2 of more than 0.5 for gaps up to six months and one year. The maximum observed R^2 for longer gap lengths, including those up to one year, was more than 0.85 for GPP and RECO at some sites, underscoring the ability of geostationary-based carbon cycle estimates to fill extended data gaps alone, in the absence of micrometeorological data inputs.

Several methods offer distinct advantages and trade-offs for carbon flux observation gap filling (LeCun et al., 2015; Moffat et al., 2007). Simple interpolation like mean diurnal variation have the potential for bias and even more advanced techniques like the dual unscented Kalman filter, multiple imputation, and terrestrial biosphere models still exhibit significant biases in gap filling, lowering reliability (Moffat et al., 2007; Vekuri et al., 2023). ML models like the GBR exhibit higher accuracy, particularly in capturing short-term variations and dependencies, without overly taxing computational resources. While deep ML models like deep NNs are powerful (LeCun et al., 2015; Yuan et al., 2020), they require substantial data and computational resources (Guicquero and Sicard, 2022; Simran et al., 2020; Yuan et al., 2020). Physical methods offer accuracy but demand extensive parameterization and domain-specific knowledge (Moffat et al., 2007; Zhu, 2014). The GBR model, known for its versatility, strikes a balance between accuracy and efficiency (Sagi and Rokach, 2018; Sahin, 2020), making it arguably suitable for both short and long gaps in GPP and RECO data across various ecosystem types.

4.4.7. Comparing GPP and RECO predictability using GOES-R data

Although both GPP and RECO were estimated with an R^2 of more than 0.70 by the GBR model, RECO estimates were more accurate (Table 4.3); the RMSE of RECO (0.90 µmol m-2 s-1) was significantly lower than the RMSE of GPP (4.15 µmol m-2 s-1). The discrepancy in the RMSE between GPP and RECO estimations obtained through the GBR can be attributed to two primary factors.

Firstly, the superior estimation of RECO over GPP using GOES-R satellite data underscores the advantage of employing the spectral bands to capture essential facets of carbon cycling. The blue and red bands are sensitive to vegetation chlorophyll and leaf structural properties, providing valuable insights into photosynthesis processes (Gitelson et al., 2014; Magney et al., 2017). The NIR and SWIR bands, on the other hand, offer critical information on vegetation and ecosystem water content and canopy structure, factors closely linked to respiration (RECO) (Lees et al., 2018; M. Wang et al., 2020). Moreover, in ecosystems where heterotrophic respiration significantly contributes to the carbon budget, like forested environments with relatively high soil respiration rates (Harmon et al., 2011; Mayer et al., 2017), the SWIR band's capacity to monitor soil moisture (Tian and Philpot, 2015) and temperature (Yang et al., 2011) may become particularly advantageous. These features are consistent with the drivers of RECO, which is dependent on photosynthesis, soil and vegetation water content, and canopy structure in addition to temperature (Lei et al., 2018).

Secondly, the broader range of actual values for GPP (8.72 ± 9.14 µmol m-2 s-1) in comparison to RECO (3.11 ± 2.73 µmol m-2 s-1) leads to a larger RMSE for GPP, as RMSE gauges the average magnitude of estimation errors relative to the actual values. To provide a more equitable basis for comparison, we can explore the normalized RMSE (nRMSE) by dividing the RMSE values by the respective mean values for GPP and RECO. This normalization allowed us to consider the scale differences and, thus, revealed that the nRMSE for GPP (0.46) and RECO (0.30) reflects their relative accuracy after accounting for the variable ranges. This finding is reinforced by our comparison of SNR values between GPP and RECO. RECO consistently exhibits higher values, indicating superior signal-to-noise performance compared to GPP. Lastly, the higher R^2 values in estimating RECO indicate superior performance compared to GPP (Table 4.3), similar to the difference in nRMSE between RECO and GPP.

4.4.8. Strengths, Limitations, Suggestions and Opportunities

GOES-R data offer several strengths that bolster their utility in ecological modeling (Khan et al., 2021; Losos et al., 2024). They provide "hypertemporal" coverage of 5-10 minutes or less across key domains, and together with other geostationary satellites continuously monitor much of the Earth's surface. This extensive coverage enhances our ability to capture high temporal ecological processes over a wide geographic area. Furthermore, the multispectral nature of GOES-R observations provides diverse information from surface and atmospheric characteristics synchronized over time and space, making them valuable inputs for ML modeling (Khan et al., 2022; Khan et al., 2021; Li et al., 2023). GOES-R data integration with ML models has shown promise in estimating GPP and RECO, highlighting the potential of satellite-derived information in ecological research (Li et al., 2023). Despite these strengths, the coarse spatial resolution of GOES-R data can be a constraint, particularly when analyzing fine-scale ecological phenomena. Moreover, GOES-R satellites collect data at off-nadir angles for most areas in its scene, leading to angular distortions in observations, requiring careful preprocessing and quality control before using the data (Khan et al., 2022; Losos et al., 2024). Therefore, to further leverage GOES-R data, future research can explore spatial downscaling techniques and data fusion (Li et al., 2022; Schmitt & Zhu, 2016; Zhang, 2010) to enhance their resolution for more localized ecological modeling. While the modeling approach from this study holds great potential for regional mapping and applications such as near real-time carbon flux monitoring, exploring these applications is beyond the scope of this paper.

Eddy covariance data serve as robust validation tools for ecological models, offering the best available ground-truth information. These measurements are collected from (often) strategically placed sites representing various ecosystems and provide accurate and continuous records (Running et al., 1999). However, limited spatial coverage and temporal gaps in the data can constrain their global-scale applicability. Therefore, utilizing ML and/or satellite data to fill temporal gaps is an important research problem which has been widely investigated (Irvin et al., 2021; Y. Kim et al., 2020; Moffat et al., 2007). Furthermore, various global GPP products exist, such as MODIS with an 8-day temporal resolution and 1 km spatial resolution or FluxCom with higher temporal resolution and a coarser spatial resolution (0.5° to 1°) (Jung et al., 2019; Z. Sun et al., 2019). This study demonstrates the accurate modeling of GPP and RECO at the site level using observations from the geostationary satellite GOES-R ABI and ML models. However, the absence of global flux products from geostationary satellites poses a significant research challenge. Developing ML models for global flux maps utilizing remote sensing geostationary satellite data can effectively address the scalability challenge in carbon cycle modeling, enabling the creation of frequent (up to half-hourly) planetary-scale maps (Ai et al., 2020; Pabon-Moreno et al., 2022; Turner et al., 2006; Yu et al., 2021).

Unlike our study's approach, global map generation combines data from all sites rather than treating each site separately. Incorporating site-specific data in generating global maps improves model interpretability and relevance by addressing local variations in factors like land use, vegetation, climate, elevation, and location. Global maps also identify regions with high uncertainty in carbon flux dynamics, potentially requiring additional monitoring. However, it may pose challenges in terms of generalizability to ecosystems underrepresented in the site database and variations in data quality. For example, if the number of sites in one dominant climate or ecosystem vastly outweighs those in other areas, the map generality for less-represented regions becomes difficult to achieve, potentially leading to overfitting of the machine learning model in that particular area. Conversely, if there's a significant lack of data in specific regions, it can introduce bias into the global maps, affecting the accuracy of estimations in those underrepresented areas.

We note that other ML models not used in our study have demonstrated promising performance for monitoring complex ecological processes, including deep learning (Nathaniel et al., 2023; Yuan et al., 2020), recurrent neural networks (Nathaniel et al., 2023; Wong et al., 2023), and Gaussian process regression (Ma et al., 2022), which can improve estimation accuracy. For instance, Long Short-Term Memory (LSTM) is a well-known recurrent neural network used for modeling time series data due to its potential to capture complex patterns (Hochreiter and Schmidhuber, 1997; Talib et al., 2021), making it well-suited for tasks that involve capturing temporal relationships and nuanced dependencies within the data (Reddy and Prasad, 2018). However, because LSTM is computationally expensive (Guicquero and Sicard, 2022; Simran et al., 2020; Zeng et al., 2020), we opted for MLP among NN models given its lower computational requirements, while acknowledging the potential of LSTM and other deep learning approaches for future exploration.

4.5. Conclusion

Our study demonstrates the potential of leveraging surface reflectance products (BRF) and top-of-atmosphere products (TOA) data from the geostationary satellite GOES-R to estimate GPP and RECO across 126 AmeriFlux sites at sub-daily scales. Gradient boosting regression (GBR) model proves effective in this estimation. Notably, surface reflectance features, especially those in the NIR spectral band, play a crucial role, while visible, SWIR, and IR reflectances, as well as radiances, further enhance performance. Combining TOA products with surface reflectance improves estimation, with longer historical records proving valuable in capturing temporal patterns. Moving forward, we envision an exciting path for further advancements in carbon flux estimation and ecosystem understanding using GOES-R and other geostationary satellite observations. Integrating meteorological and site-specific data with satellite imagery will help account for local environmental variations to construct near-global scale machine learning models. This expansion beyond site-level modeling promises a more comprehensive understanding of carbon flux dynamics. We recommend diversifying modeling approaches to further enhance estimation accuracy and accommodate temporal dependencies, which could involve integrating computationally expensive deep learning techniques such as Long Short-Term Memory (LSTM). Accurate prediction of GPP and RECO informs more robust ecological models, deepening our understanding of the planet's carbon cycle and its environmental implications.

Chapter 5. Harnessing Information from Shortwave Infrared Reflectance Bands to Improve Satellite-based Estimates of Gross Primary Productivity

This chapter is published in the Wiley-Journal of Geophysical Research: Biogeosciences: Ranjbar, S., Losos, D., Dechant, B., Hoffman, S., Başakın, E. E., & Stoy, P. C. (2024). Harnessing information from shortwave infrared reflectance bands to enhance satellite-based estimates of gross primary productivity. Journal of Geophysical Research: Biogeosciences, https://doi.org/10.1029/2024JG008240

Abstract

Monitoring Gross Primary Productivity (GPP), the rate at which terrestrial ecosystems fix atmospheric carbon dioxide, is crucial for understanding global carbon cycling. Remote sensing offers a powerful tool for monitoring GPP using vegetation indices (VIs) derived from visible and near-infrared reflectance (NIRv). While promising, these VIs often suffer from sensitivity to soil background, moisture, and variations in solar and view zenith angle (SZA and VZA). This study investigates the potential of incorporating shortwave infrared (SWIR) reflectance from MODIS and GOES-R advanced baseline imager (ABI) sensors to improve GPP estimation. We evaluated various formulations for creating **S**WIR-enhanced **N**ear-**I**nfra**R**ed reflectance of **V**egetation (sNIRv) by integrating SWIR information into established VIs across 96 Ameriflux and NEON research sites. Our findings reveal that sNIRv improves correlation with GPP for ABI data by up to 0.19 on a half-hourly basis for normalized difference vegetation index (NDVI) values below 0.25, with diminishing gains as NDVI values rise. Using MODIS data, sNIRv matches r values of NIRv for NDVI above 0.25, with a slight 0.05 increase for NDVI below 0.25. Analyses using SCOPE model simulations further support the ability of sNIRv to capture fPAR (fractional photosynthetically active radiation), a proxy for GPP, especially for ecosystems with low LAI. Results highlight that sNIRv-based VIs are less sensitive to soil background, SZA, and VZA compared to NIRv. SHapley Additive exPlanations (SHAP) value analysis also identifies sNIRv

as the best feature for GPP estimation using machine learning modeling across different land covers, NDVI ranges, and soil water content (SWC) levels.

5.1 Introduction

Terrestrial vegetation plays a pivotal role in the Earth system, serving as the primary engine of photosynthetic carbon fixation (Kustas and Norman, 1999; Tucker et al., 1986; Veroustraete et al., 1996). Ecosystem carbon uptake (the gross primary productivity, GPP), underpins ecosystem function, drives food webs, and helps regulate global climate (Chen et al., 2023; Nemani et al., 2003; Wang et al., 2020; Weltzin et al., 2020). Monitoring GPP dynamics across space and time is crucial for understanding the response of terrestrial ecosystems to environmental change (Gitelson et al., 2014; Lei et al., 2018; Yang et al., 2007). Remote sensing has revolutionized our understanding of the terrestrial carbon cycle, allowing for large-scale assessment of vegetation productivity and highlighting the critical role of terrestrial carbon dioxide fluxes in determining trends and variability in atmospheric carbon dioxide concentration (Friedlingstein et al., 2022; Xiao et al., 2019). Ongoing innovations to remote sensing algorithms continue to improve our ability to monitor GPP.

Vegetation indices (VIs) calculated using passive remote sensing are powerful tools for extracting ecologically relevant information from vegetation and soil reflectances across multiple spectral bands (Massetti et al., 2019; Tucker et al., 1986; Zeng et al., 2022). These mathematical formulae condense spectral reflectance patterns into single values, acting as proxies for biophysical attributes that offer crucial insights into plant structure and function (Tucker et al., 1986; Zeng et al., 2022). In the context of GPP estimation, a diverse array of VIs has emerged, each with distinct strengths and limitations, and therefore opportunities for improvement in specific ecological contexts and environmental conditions (Zeng et al., 2022). We briefly describe VIs as proxies for GPP, focusing on the challenges posed by them, particularly in relation to soil backgrounds, snow, saturation, and variations in solar zenith angle (SZA) and view zenith angle (VZA) over various vegetation cover conditions.

The normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) are widely used proxies for the leaf area index (LAI) and the fraction of absorbed photosynthetically active radiation (fPAR), both of which have a direct relationship with GPP. Recent approaches have aimed to develop formulations more closely related to the processes that control GPP. Prominent among these are the near infrared reflectance of vegetation (NIRv, Badgley et al., 2017, 2019), the NDVI multiplied by near infrared reflectance, especially once multiplied again by photosynthetically active radiation (PAR) to calculate the NIRvP (Chen et al., 2023; Chen et al., 2023; Dechant et al., 2020, 2022; Jeong et al., 2023; Khan et al., 2022). Despite remarkably strong relationships with tower-estimated GPP that can under certain situations exceed that of solar induced fluorescence (SIF, (Baldocchi et al., 2020; Chen et al., 2023; Dechant et al., 2022; Liu et al., 2023), NIRvP's dependence on canopy parameters may hinder grasp of the full complexity of GPP.

One limitation of NIRvP, for example, is its sensitivity to soil background reflectance, particularly in ecosystems with variable meteorology and canopy leaf area (Chen et al., 2023). Although NIRv has demonstrated success in estimating LAI, even in mixed (Badgley et al., 2017), it remains challenged by soil background, especially in sparsely vegetated areas such as shrublands, savannas, and grasslands due in part to the persistent influence of soil background reflectance (Dechant et al., 2022; Jiang et al., 2021), and variability in the carbon flux of dryland ecosystems is disproportionately important for the interannual variability of the global atmospheric CO2 concentrations (Ahlström et al., 2015; Poulter et al., 2014).

To address the issue of soil background, recent research has proposed practical strategies that leverage the robustness of NIRv against soil contamination. Zeng et al., (2021) proposed NIRv from Hyperspectral data (NIRvH), an index for estimating vegetation NIR reflectance from hyperspectral data, leveraging spectral shape variations in the red-edge region to mitigate the impact of soil effects. NIRvH outperforms NIRv and the difference vegetation index (DVI) in isolating vegetation NIR reflectance, crucial for accurate photosynthesis estimation. This strategy paves the way for improved accuracy in NIRvP retrievals, particularly in challenging landscapes with significant soil background interference. Moreover, the Global Vegetation Moisture Index (GVMI) was developed to retrieve vegetation water content rapidly and reliably on a global scale using SPOT-VEGETATION data (Ceccato et al., 2002). Indices utilizing NIR and SWIR bands have shown better performance in drought monitoring, especially in grasslands, by being sensitive to soil moisture conditions (Zhu et al., 2024). Furthermore, the Vegetation-water Resistant Soil Moisture Index (VRSMI) incorporates a physical relationship between NIR and SWIR reflectance, supporting soil moisture content estimation even under low-moderate crop cover (Yue et al., 2024).

Beyond soil, snow presence can also hinder accurate GPP estimation. Traditional VIs like NDVI often struggle with snow cover, as the high reflectance in the NIR band confounds the vegetation signal (Wang et al., 2017; Xu et al., 2021). The normalized difference phenology index (NDPI) was developed in part to address this challenge by replacing the red band in the NDVI with a weighted red-shortwave infrared (SWIR) combination designed to minimize the influence of both soil and snow backgrounds (Wang et al., 2017; Xu et al., 2021). A combination of lower sensitivity at sparse vegetation and saturation in dense vegetation is a common limitation for many VIs (Huang et al., 2021; Sun et al., 2021), which also demands attention. For example, while EVI excels in dense vegetation stands, it falters in sparse grasslands due to lower sensitivity (Huang et al., 2021; Sun et al., 2021; Xu et al., 2021) while NDVI saturates at the high LAI values for which EVI is less prone to saturation (Huete et al., 1997). This highlights the importance of tailoring VI selection to specific vegetation types and coverage levels to ensure accurate GPP estimates across diverse landscapes. Angular effects also exert a significant influence on optical vegetation indices, which varies across spectral bands and can be particularly pronounced during senescence, highlighting the importance of selecting indices with minimized potential angular effects for vegetation monitoring (Gu et al., 2021).

The quest for accurate GPP estimation extends beyond VIs. Approaches that combine the strengths of multiple indices and spectral bands hold immense potential. For instance, the Vegetation Photosynthesis Model (VPM) harnesses the ability of EVI to represent the light-harvesting capacity of photosynthetic organs, quantified by the fPAR, while incorporating SWIR bands for improved soil moisture retrieval (Xiao et al., 2004). Such innovative combinations offer a more holistic perspective on GPP, but introduce modeling assumptions including temperature and water stress scalars. While NIRvP and other VIs face challenges with soil backgrounds, snow, saturation, and variations in SZA and VZA, the combination of SWIR bands with them stands out as a promising contender, particularly in sparse vegetation, savannas, shrublands, and croplands (Wang et al., 2017; Xu et al., 2021). SWIR bands collected by commonly-used satellites, primarily centered around 1630 nm and/or 2200 nm (Khan et al., 2021; McCorkel et al., 2020), can be effective for discriminating between soils with varying conditions (e.g. moisture content, see Figure 5.1).

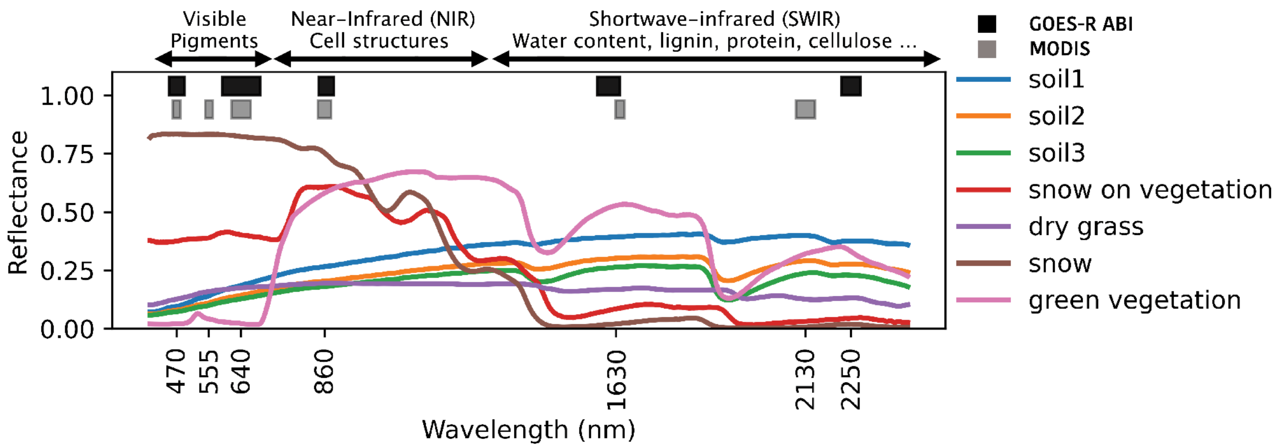

Figure 5.1. Typical spectral signatures of soil, dry and green vegetation, and snow. Gray and black bars show the spectral range measured by MODIS and GOES-R ABI bands, respectively. Three soil spectra were obtained from the SCOPE model soil simulation tool (Van der Tol et al., 2009), while the remaining spectral values depicted here were acquired from the freely available United States Geological Survey (USGS) Spectroscopy Laboratory's Spectral Library 07a (Splib07a) (Kokaly et al., 2017).

While these indices have made significant contributions to our understanding of vegetation dynamics, particularly in the context of moisture content, drought, and snow, the combination of SWIR bands with NIRv, to our knowledge, has not been extensively explored in a single, unified index focused on GPP prediction. Here, we propose the SWIR-enhanced near-infrared reflectance (NIR) of Vegetation (sNIRv) that integrates the strengths of NIRv with the potential of SWIR bands measured using common remote sensing platforms (Fig. 1). Incorporating SWIR bands in VIs may enhance their performance for more accurate monitoring of vegetation, particularly in areas with varying soil conditions, by reducing sensitivity to soil background signals (Wang et al., 2017; Xu et al., 2021).

SWIR wavelengths penetrate deeper into the leaf mesophyll, providing information on leaf internal structure, water content, and canopy structure composition (Croeser et al., 2022). These characteristics are crucial for determining photosynthetic efficiency, which directly influences GPP (Jahan et al., 2023; Ranjbar et al., 2023; Xiong & Nadal, 2020). By incorporating SWIR information into NIRv, we aim to develop a sNIRv that captures a more comprehensive picture of the factors influencing GPP across multiple ecosystems, and over time. We test different mathematical formulations for sNIRv and evaluate their performance against established GPP measurements at 96 AmeriFlux and NEON eddy covariance towers that capture a wide range of vegetation type and seasonality. Additionally, we assess sNIRv performance against fractional photosynthetically active radiation (fPAR) using hyperspectral data simulated by the SCOPE model (Van der Tol et al., 2009). Finally, we conducted a feature importance analysis for GPP modeling using different indices and machine learning (ML) to better understand what inputs are important for prediction.

5.2 Materials and Methods

5.2.1 Eddy covariance

We compiled eddy covariance flux data from AmeriFlux and NEON eddy covariance towers as described in Losos et al., (2024b) and selected sites that have data after 2021 to align with GOES-16 ABI surface bidirectional reflectance factor (BRF) and MODIS surface reflectance observations, described in 2.2, which results in 96 sites distributed across the Conterminous United States (CONUS), Canada, and Puerto Rico (see Figure 5.2). The partitioning of carbon dioxide flux measurements into GPP was conducted using the hesseflux library in Python (Cuntz, 2020) following established algorithms for calculating GPP (Papale et al., 2006), along with uncertainty estimates, utilizing both daytime (Lasslop et al., 2010) and nighttime (Reichstein et al., 2005) partitioning approaches with an established friction velocity (u*) filter to exclude eddy covariance measurements taken during periods of insufficient turbulence at night (Reichstein et al., 2012). We use nighttime flux partitioning in the present analysis. Forest sites lacking in-canopy CO2 concentration measurements to calculate storage fluxes were excluded to prevent underestimation of GPP, following the recommendations of Fu et al., (2018). In total, our analysis included 231,702 time points collected from the 96 AmeriFlux and NEON sites from August 2021 - December 2023.

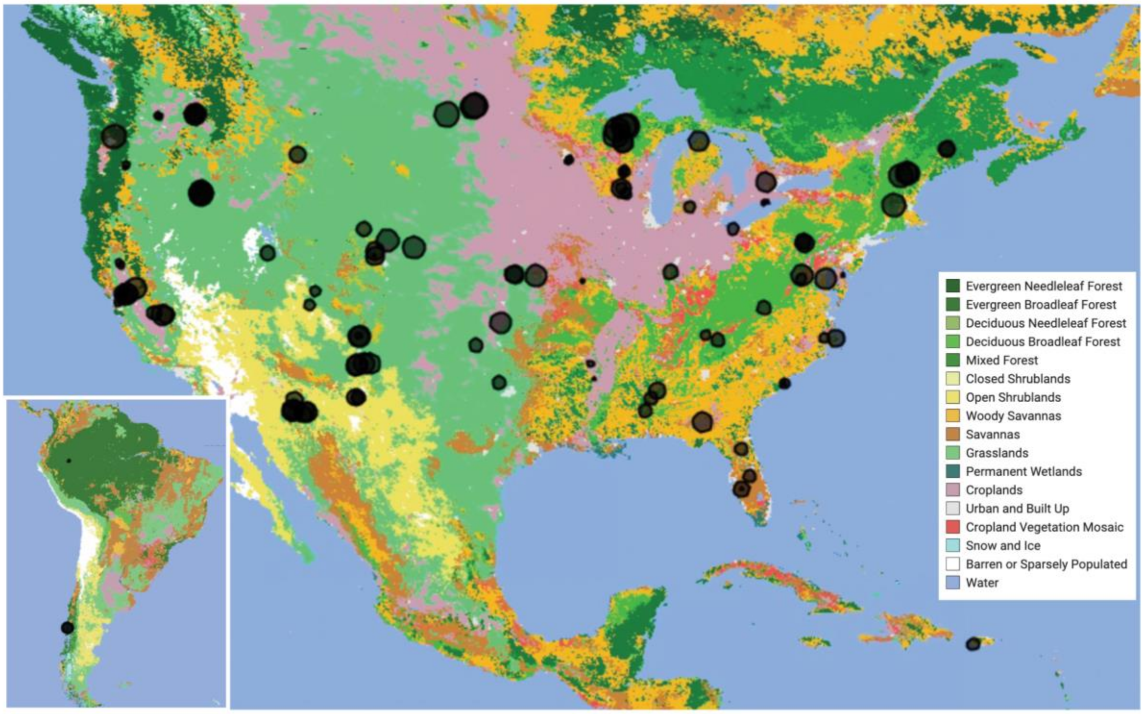

Figure 5.2. The spatial distribution of AmeriFlux and NEON eddy covariance sites used in this study, mapped over IGBP vegetation types, for sites in the GOES-16 Full Disk projection. Larger circle sizes represent sites with data size availability.

5.2.2 Remote sensing

5.2.2.1 GOES-R

This study harnessed remote sensing data from GOES-R and MODIS satellites, with an emphasis on the Advanced Baseline Imager (ABI) aboard GOES-R for monitoring terrestrial ecosystems. We use ABI observations from GOES-16, operational since November 2017, with BRF estimates provided continuously since August, 2021. The ABI captures imagery across 16 spectral bands with a nadir resolution of 0.5-2 km, providing full-disk observations every 10 minutes and CONUS observations every 5 minutes in its most common mode (Mode 6). This research synchronized

ABI data, particularly the GOES-R ABI L2 BRF products (He et al., 2019), spanning August 2021 to December 2023, with ground-based GPP data from 96 AmeriFlux and NEON sites as described in Losos et al., (2024b). These BRFs encompass bands across visible, near-infrared (NIR), and shortwave infrared (SWIR) wavelengths in the blue (BRF1, centered at 470 nm, see Figure 5.1), red (BRF2, 640 nm), NIR (BRF3, 860 nm), shortwave infrared 1 (SWIR1, 1630 nm) and shortwave infrared 2 (SWIR2, 2200 nm), provided every five minutes across CONUS and ten minutes across the Western Hemisphere Full Disk. ABI's BRF product offers a clever way to estimate ground reflectivity (surface reflectance) despite clouds. To do so, the BRFs consider both sunlight angles and the satellite's view, which is important because the satellite's position is fixed. The key is in retrieving Bidirectional Reflectance Distribution Function (BRDF) parameters, which can then be used to model surface reflectance even on cloudy days (He et al., 2019). This is achieved by analyzing clear-sky data (when available) or using BRDF information from previous clear days (Losos et al., 2024b), and inferring BRFs during cloudy periods.

We also used the Downward Shortwave Radiation (DSR) product (Laszlo et al., 2020) for our analysis. Despite potential uncertainty arising from differences between DSR and Photosynthetically Active Radiation (PAR), this impact is generally minimal compared to natural variations in solar irradiance (Dechant et al., 2022; Losos et al., 2024b; Ranjbar et al., 2024b). To ensure data quality, ABI observations were processed by averaging within each half-hour period and applying established techniques for temporal and spatial alignment with AmeriFlux data as described in (Losos et al., 2024b). We analyze the relationship between VIs based on BRFs at half-hourly intervals.

5.2.2.2 MODIS

MODIS, onboard the Terra and Aqua satellites, provides a complementary view of land surface dynamics across the continental United States (CONUS) with seven spectral bands and moderate spatial resolution (250 m). MODIS offers near-daily observations for monitoring a wide range of land surface characteristics by capturing surface reflectance at seven visible, NIR, and SWIR bands (Kim et al., 2017; Waggoner & Sokolik, 2010), and whereas it is past its expected mission duration, it provided observations during the measurement period for comparison against the ABI. The multi-year, near-daily coverage of MODIS enabled robust characterization of seasonal and inter-annual variations in land surface phenomena. We utilized surface reflectance bands from MODIS product MCD43A4.061 (MODIS Nadir BRDF-Adjusted Reflectance Daily 500m (Schaaf et al., 2002)), including blue, red, NIR, and SWIR2 band (at 2200 nm), leveraging their sensitivity to various land cover features and biophysical processes. By investigating MODIS data alongside high-temporal-resolution GOES-R observations, our aim was to assess the robustness of indices that incorporate the SWIR across multiple scales, capturing both detailed spatial features and temporal patterns.

5.2.3 Simulation of canopy hyperspectral data using SCOPE

Vegetation models serve as effective tools for comprehending diverse plant physiological phenomena. Among these models, radiative transfer models (RTMs) are extensively employed in remote sensing due to their ability to establish a direct link between observations at the top of the canopy (TOC) and vegetation characteristics such as chlorophyll content and leaf area index (Houborg et al., 2007; Jacquemoud et al., 2009). The Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) model (Van der Tol et al., 2009), a physical model, offers deeper insights, integrating radiative transfer and energy balance to simulate TOC reflected radiation, fluorescence emission, photosynthesis, and flux characteristics such as fPAR (Schull et al., 2007; van der Tol et al., 2016; Van der Tol et al., 2009).

SCOPE has been pivotal in understanding remotely sensed data including canopy photosynthesis, and for utilizing reflectance and fluorescence to quantify plant traits (Jacquemoud et al., 2009; van der Tol et al., 2016). Our study uses SCOPE to generate synthetic TOC reflected radiation, photosynthesis, and total fPAR as a GPP proxy. By simulating data across various vegetation characteristics (Table 5.1), we seek to gain deeper insights into how the sNIRv index is related to fPAR.

Table 5.1. Ranges of SCOPE input parameters.

Parameter	**Explanation**	**Unit**	**Range**
Cab	Chlorophyll a + b content	$\mu g\ cm^{-2}$	5 to 85
Cdm	Leaf mass per unit area	$g\ cm^{-2}$	0 to 0.03
Cw	Equivalent water thickness	cm	0 to 0.03
Cs	Senescence material (brown pigments)	fraction	0 to 1
Cca	Carotenoid content	$\mu g\ cm^{-2}$	0 to 25
Nl	Leaf structure parameter	-	1 to 3
LAI	Leaf area index	-	0 to 7
LIDFa	Leaf inclination function parameter a	-	−1 to 1
LIDFb	Leaf inclination function parameter b	-	−1 to 1
θs	Solar zenith angle (SZA)	°	0 to 60
θv	View zenith angle (VZA)	°	0 to 60

In addition to analyzing vegetation characteristics, our study also examines three distinct soil spectra (Soils 1, 2, and 3) and one snow spectrum, as shown in Figure 5.1. Soil 1 is dry, Soil 2 has 20% volume soil moisture, and Soil 3 has 30% volume soil moisture, contributing to the diverse soil backgrounds explored in this study (Van der Tol et al., 2009). We utilized the Brightness–Shape–Moisture (BSM) soil reflectance model (Verhoef et al., 2018; Yang et al., 2020) in SCOPE simulations to generate soil spectral simulations and obtained the snow spectrum from the USGS Spectroscopy Laboratory's Spectral Library 07a (Splib07a, Kokaly et al., 2017).

5.2.4 sNIRv formulations

We incorporated several established vegetation indices, including NIRv (Badgley et al., 2017), bNIRv, and EVIv (refer to Equations (Eq.) 5.1-5.3), into our analysis. These established indices serve as benchmarks for evaluating the performance of our newly proposed **S**WIR-enhanced near-infrared reflectance (**NIR**) of **V**egetation (**sNIRv**) indices, denoted as $sNIRv_{LSWI}$, $sNIRv_{NDPI}$, $sNIRv_{SWIR^2}$, $sNIRv_{NDVI\times LSWI}$, and $sNIRv_{NDVI+LSWI}$ (see Eq. 5.4-5.8) that leverage information from SWIR bands. We experimented with various formulations for developing sNIRv, utilizing both experimental methods and machine learning approaches. Ultimately, we developed sNIRv based on existing SWIR-based indices from the literature and experimental methods. Specifically, we employed the established vegetation indices NDPI (Xu et al., 2021), NDVI (Kriegler, 1969), and Land Surface Water Index (LSWI, Chandrasekar et al., 2010) to propose and evaluate five sNIRv formulations utilizing NDPI, SWIR^2, LSWI, NDVI×LSWI and NDVI+LSWI. LSWI has similar formulation as the Normalized Difference Water Index (NDWI, Gao, 1996)). NDPI incorporates a parameter, α, taken to be 0.74 (Xu et al., 2021), to mitigate soil and snow background effects. Further details on α's formulation and significance are provided in (Xu et al., 2021). For sNIRv formulations, following sensitivity analyses, we employed the ‘SWIR2’ band at 2100 nm. We also investigated the ‘SWIR1’ band at 1630 nm in section 3.6, where we developed a machine learning model to analyze the feature importance for GPP estimation.

$$NIRv = NDVI \times N = \frac{N-R}{N+R} \times N\,, \qquad \text{Eq. 5.1}$$

$$bNIRv = bNDVI \times N = \frac{N-B}{N+B} \times N, \qquad \text{Eq. 5.2}$$

$$EVIv = 2.5 \times \frac{N-R}{N + 6\times R - 7.5\times B + 1} \times N\,, \qquad \text{Eq. 5.3}$$

$$sNIRv_{LSWI} = LSWI \times N = \frac{N-S}{N+S} \times N\,, \qquad \text{Eq. 5.4}$$

$$sNIRv_{NDPI} = NDPI \times N = \frac{N-(\alpha R+(1-\alpha)\,S)}{N+(\alpha R+(1-\alpha)\,S)} \times N\,, \qquad \text{Eq. 5.5}$$

$$sNIRv_{SWIR^2} = \frac{N - R - S^2}{N + R + S^2} \times N \text{ ,} \qquad \text{Eq. 5.6}$$

$$sNIRv_{NDVI \times LSWI} = (NDVI \times LSWI) \times N = \left(\frac{N - R}{N + R} \times \frac{N - S}{N + S}\right) \times N = \frac{N^2 - N \times S - N \times R + R \times S}{N^2 + N \times S + N \times R + R \times S} \times N \text{ ,} \qquad \text{Eq. 5.7}$$

$$sNIRv_{NDVI + LSWI} = (NDVI + LSWI) \times N = \left(\frac{N - R}{N + R} + \frac{N - S}{N + S}\right) \times N = \frac{2N^2 - 2R \times S}{N^2 + N \times S + N \times R + R \times S} \times N, \qquad \text{Eq. 5.8}$$

Where B, R, N, and S represent blue, red, NIR, and SWIR reflectances, respectively, in the equations.

5.2.5 Feature importance analysis

To identify informative variables for refining predictions of GPP using machine learning, we employed SHapley Additive exPlanations (SHAP, Lundberg & Lee, 2017). SHAP helps pinpoint features that may contain unexplained variance, potentially guiding further exploration of the GPP prediction using sNIRv compared to other VIs. SHAP is a powerful tool rooted in game theory for unraveling the opaque nature of ML predictions (Lundberg & Lee, 2017) to understand individual features' influence on a machine learning model's predictions. By calculating the direction and strength of each feature's contribution, SHAP can reveal both collaborative (synergistic) and opposing (antagonistic) relationships between features that ultimately guide the model's behavior (Lundberg et al., 2020; Lundberg & Lee, 2017). SHAP can clarify individual predictions, unveil feature interactions, and demystify the "black box" nature of ML models.

For the trained ML model utilized in the SHAP analysis, we employed a Gradient Boosting Regression (GBR) model. GBR is proficient at capturing nonlinear relationships between variables by integrating decision trees within an ensemble learning paradigm (Bentéjac et al., 2021; Golden et al., 2019). Its applicability in environmental remote sensing, such as GPP and carbon respiration (RECO) prediction (Cai et al., 2020; Ma et al., 2022; Ranjbar et al., 2023), DSR estimation (Ranjbar et al., 2024b), LAI estimation (Liu et al., 2021), drought monitoring (Li et al., 2023; Ranjbar et al., 2021) and more, has been well-established. The critical facet of hyperparameter tuning for GBR involves adjustment of the number of trees, maximum tree depth, learning rate and minimum leaf samples executed through a systematic grid search algorithm. This fine-tuning process is aimed to govern the model's dimensions, intricacy, and mitigate the potential for overfitting (Bentéjac et al., 2021; Li et al., 2023; Ranjbar et al., 2021). The validity and robustness of the trained GBR model were evaluated through the implementation of a leave-one-out cross-validation (LOOCV) algorithm (Ranjbar et al., 2021; Yates et al., 2023).

5.3 Results and Discussion

5.3.1 Comparison of VIs

We first investigated the relationship between tower-based GPP estimates and different VIs across all 96 eddy covariance sites across CONUS. Figure 5.3 utilizes kernel density estimation (KDE) plots to visually compare the Pearson correlation coefficients (r values) between GPP and the VIs with the highest correlation (NIRv, $sNIRv_{NDPI}$, $sNIRv_{SWIR^2}$, $sNIRv_{NDVI \times LSWI}$, and $sNIRv_{NDVI+LSWI}$) at each study site. An analysis of half-hourly ABI products with KDE plots (Figure 5.3a) revealed that VIs incorporating NIR, red, and SWIR bands exhibited a tendency towards higher r values compared to other indices. Among the sNIRv indices using ABI BRFs, $sNIRv_{NDVI \times LSWI}$ displayed the strongest correlation with GPP with peak r value of 0.78 in the KDE plot, followed by $sNIRv_{SWIR^2}$ with a peak r value of 0.74. This pattern held true for daily and weekly averaged data (Figure 5.3b & d). Table 5.2 additionally presents the median values and interquartile ranges (IQRs) of the r values depicted in the KDE plot, due to their highly skewed distribution (Figure 5.3).

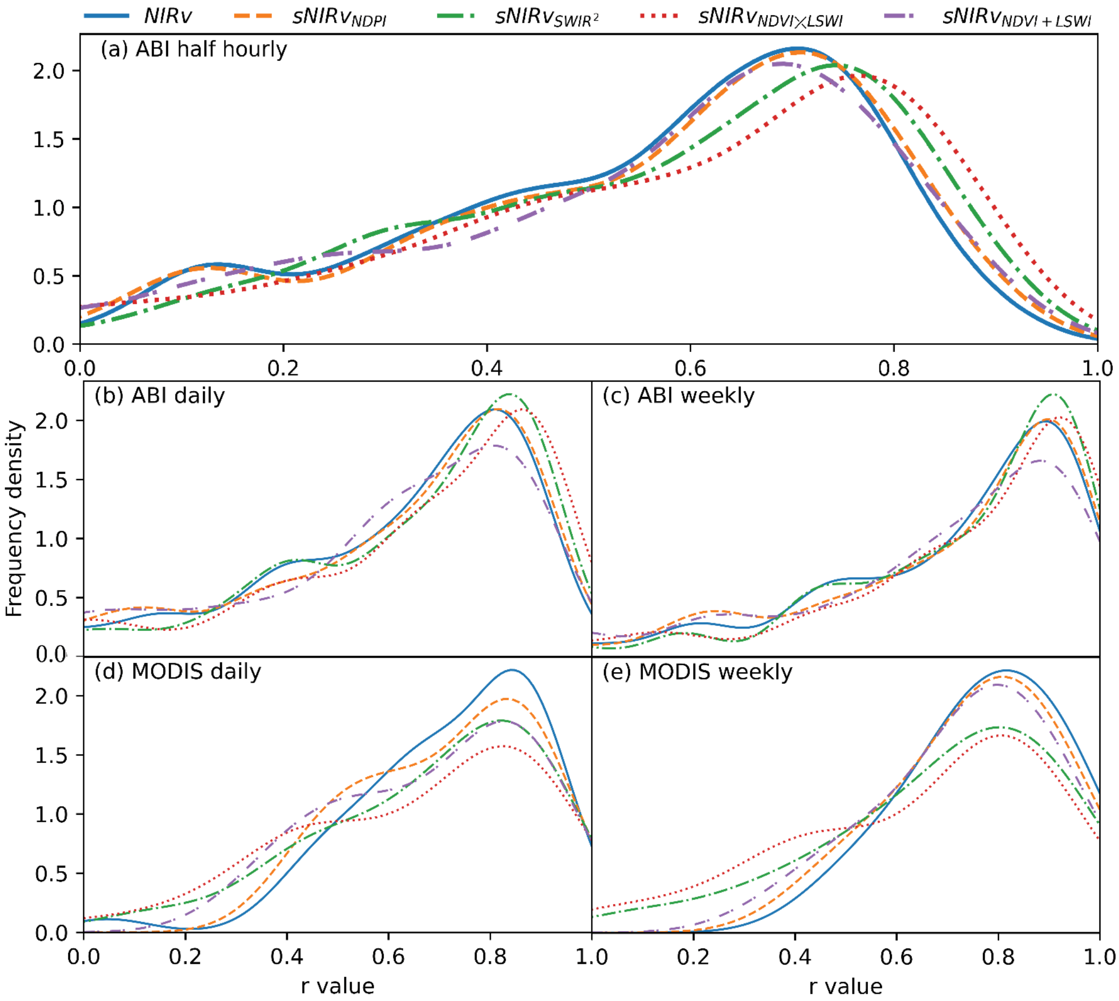

Figure 5.3: Kernel Density Estimation (KDE) plots comparing Pearson correlation coefficients (r values) of the VIs with tower-based GPP estimates across 96 CONUS sites.

However, for daily MODIS data, NIRv exhibited superior performance compared to sNIRv indices (Table 5.2), with a peak r value of 0.85 in the KDE plot. Among the sNIRv indices, $sNIRv_{NDPI}$, with a peak r value of 0.83, and closely followed by $sNIRv_{SWIR^2}$, with a peak r value of 0.81. Similar trends are observed for weekly MODIS data, where NIRv has the highest correlation closely **followed** by $sNIRv_{NDPI}$ and $sNIRv_{SWIR^2}$. Notably, bNIRv and $sNIRv_{LSWI}$ derived from both satellites exhibited the weakest correlations with GPP. Although EVIv derived from ABI data showed the worst agreement with GPP at half-hourly, daily, and weekly resolutions, it demonstrates the best performance for MODIS data. For weekly MODIS data, a median r-value of 0.76, equal to NIRv, was achieved (Table 5.2).

Table 5.2 The median correlation coefficients (r-values) of the VIs with GPP for all sites using (a) half-hourly ABI data, (b) daily ABI data, and (c) daily MODIS data. Interquartile Ranges (IQR) also are presented in parentheses.

Indices	**GOES-R ABI**			**MODIS**	
	Half hourly	**Daily**	**Weekly**	**Daily**	**Weekly**
NIRv	0.61 (0.33)	0.70 (0.37)	0.80 (0.35)	**0.75 (0.25)**	**0.76 (0.21)**
$\mathrm{sNIRv_{NDPI}}$	0.63 (0.31)	0.71 (0.37)	0.82 (0.32)	0.74 (0.29)	**0.76 (0.27)**
$\mathrm{sNIRv_{SWIR^2}}$	0.62 (0.34)	**0.74 (0.41)**	**0.85 (0.32)**	0.73 (0.38)	0.74 (0.32)
$\mathrm{sNIRv_{NDVI \times LSWI}}$	**0.64 (0.35)**	**0.74 (0.41)**	**0.86 (0.31)**	0.71 (0.44)	0.73 (0.43)
$\mathrm{sNIRv_{NDVI+LSWI}}$	0.62 (0.35)	0.65 (0.39)	0.76 (0.39)	0.74 (0.37)	**0.76 (0.30)**
$\mathrm{sNIRv_{LSWI}}$	0.57 (0.44)	0.62 (0.55)	0.73 (0.61)	0.68 (0.41)	0.74 (0.42)
bNIRv	0.52 (0.28)	0.62 (0.35)	0.71 (0.37)	0.71 (0.24)	0.75 (0.25)
EVIv	0.52 (0.44)	0.58 (0.57)	0.71 (0.58)	0.72 (0.29)	**0.76 (0.19)**

We further compute root mean square error (RMSE) values for the linear model fitted on the VIs for GPP estimation. RMSE values also support our findings, where $sNIRv_{NDVI \times LSWI}$ performs best for ABI data at all three temporal resolutions of half hourly, daily and weekly (median RMSE of 2.23 to 4.40 μmol m^{-2} s^{-1}). Moreover, EVIv had the worst performance for ABI data (median RMSE of 2.96 to 5.13 μmol m^{-2} s^{-1}). However, EVIv has the best performance for MODIS daily and weekly data (median RMSE of 2.06 to 3.79 μmol m^{-2} s^{-1}).

Our analysis of the VIs derived from ABI and MODIS data against tower-based GPP estimates reveals a critical role for VI choice and sensor characteristics in capturing GPP dynamics. A key difference between ABI and MODIS data is that MODIS provides Normalized Bidirectional Reflectance Factor (NBAR) products, while ABI data are BRFs, which are susceptible to viewing angle effects. Different bands and VIs are affected by view zenith angle, leading to increases or decreases in their correlation with GPP (Gu et al., 2021; Lobell et al., 2002). The higher correlation coefficients observed with VIs for ABI data incorporating NIR, red, and SWIR bands highlight the importance of these spectral regions for accurate BRF-based GPP estimation. Notably, the strong performance of $sNIRv_{NDVI \times LSWI}$ across ABI data suggests its potential for robust GPP estimation at finer temporal scales. However, the superiority of NIRv over sNIRv indices in daily MODIS data underscores the influence of sun-sensor-view geometry on VI-GPP relationships. This finding emphasizes the necessity of new analysis for different factors affecting satellite observations including SZA, VZA, soil water content (SWC), and soil background. Additionally, the persistent underperformance of bNIRv and $sNIRv_{LSWI}$ relative to other VIs – both which exclude the Red band – underscores their limitations in accurately capturing GPP dynamics across diverse temporal resolutions and ecosystem types.

5.3.2 Index performance over different ranges of NDVI

Figure 5.4 depicts the correlation between various VIs and GPP across different ranges of the NDVI. We used the NDVI values derived from the ABI and MODIS data themselves (Eq. 5.1). This analysis aims to investigate the influence of soil background effects on the VI-GPP relationship as low NDVI values imply less vegetation and more likely soil background effects. Figures 4a and 4b compare the VI-GPP correlations derived from ABI data at half-hourly and daily temporal resolutions, respectively. We noted higher r-values for indices such as $sNIRv_{NDVI \times LSWI}$, $sNIRv_{SWIR^2}$, and $sNIRv_{NDPI}$ when compared to NIRv for ABI data. $sNIRv_{NDVI \times LSWI}$ exhibits the strongest correlation with GPP across all NDVI ranges with an improvement in r of 0.19 for NDVI values below 0.25, albeit with diminishing improvements as NDVI values increase. In the context of MODIS data, $sNIRv_{SWIR^2}$ demonstrates comparable r values to sNIRv when NDVI exceeds 0.25, with a more modest enhancement of 0.05 in mean r values for NDVI values below 0.25. Conversely, NIRv surpasses $sNIRv_{NDVI \times LSWI}$ in correlation with GPP for all NDVI ranges

Similarly, in alignment with findings from ABI data, EVIv, bNIRv, and $sNIRv_{LSWI}$ exhibit lower correlations with GPP.

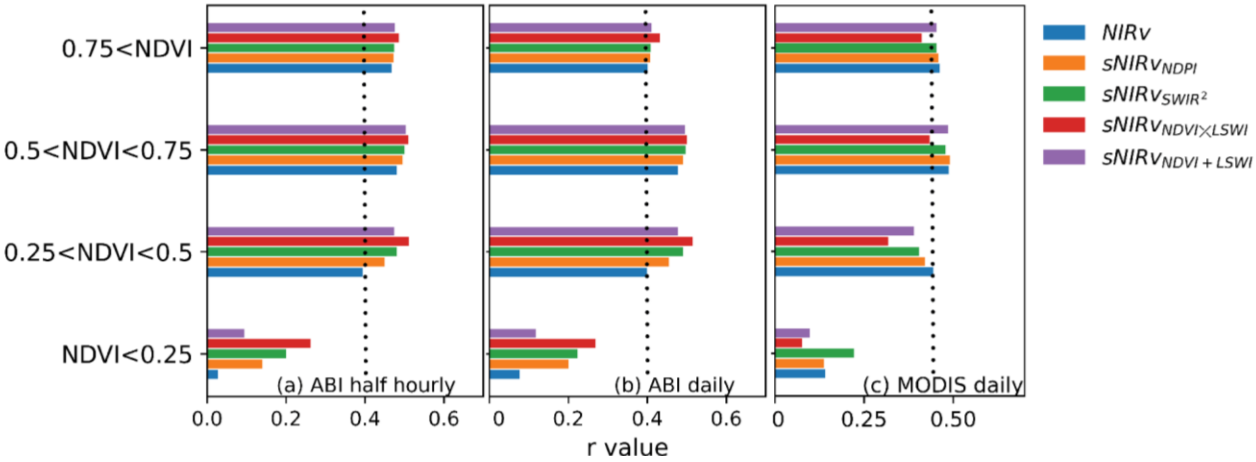

Figure 5.4. Correlation coefficients (r-values) of the VIs with GPP for all the sites in (a) half-hourly GOES-16 ABI data, (b) daily ABI data, and (c) daily MODIS data. The black dotted lines show r values of 0.4 for visual reference.

As noted, lower NDVI values typically indicate less vegetation cover, leading to a greater presence of background signals, primarily from soils. This scenario presents a challenge for some VIs, as demonstrated in previous studies that have shown drawbacks of NIRv in sparse vegetation conditions (Dechant et al., 2022; Jiang et al., 2021), which can be due to its sensitivity to background signals. In contrast, the current findings suggest that sNIRv, particularly $sNIRv_{NDVI \times LSWI}$ for ABI data and $sNIRv_{SWIR^2}$ for MODIS data, exhibits a more robust correlation with GPP across low levels of NDVI. This resilience of sNIRv in ecosystems that can be expected to have stronger soil background signals highlights its potential for accurate GPP estimation, especially in environments with sparse vegetation cover or heterogeneous land surfaces. These insights underscore the importance of considering both NDVI ranges and the spectral characteristics of VIs when assessing their suitability for GPP estimation, ultimately contributing to enhanced understanding of ecosystem productivity dynamics in response to environmental changes.

5.3.3 sNIRv sensitivity to solar zenith angle (SZA) and view zenith angle (VZA)

We conducted an analysis to investigate the sensitivity of sNIRv indices to SZA. We analyzed GOES half-hourly data and found that as SZA increases, the correlation between NIRv and GPP decreases for half hourly ABI data. However, sNIRv indices maintained a stronger correlation with GPP even at SZAs exceeding 60 degrees (Figure 5.5a), where the r value of $sNIRv_{NDVI \times LSWI}$ increased by 0.02 and that of NIRv decreased by 0.04. However, daily ABI data do not have a similar trend; the correlation between sNIRv indices and GPP decreases as SZA increases similar to the NIRv trend (Figure 5.5b). The impact of SZA was more pronounced in the analysis of daily data, although the differences between the VIs were relatively smaller than half hourly data. When examining daily data, it is apparent that there exists a confounding effect attributed to diurnal versus seasonal variations in SZA, whereas the daily dataset strongly reflects the seasonal component.

The analysis of daily MODIS data revealed that sNIRv indices follow the same trend as NIRv, offering no improvement (Figure 5.5c). This highlights the importance of investigating VZA sensitivity for GOES data, particularly considering some eddy covariance research sites are located in pixels with VZA up to 70 degrees when viewed by GOES-16 (Losos et al., 2024b).

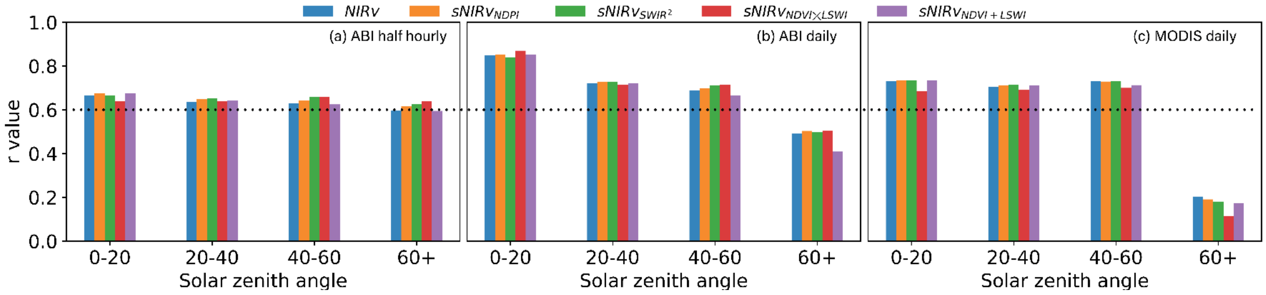

Figure 5.5. sNIRv indices performance in different solar zenith angles compared to NIRv for half-hourly and daily GOES-16 ABI data and daily MODIS data. The black dotted line shows r values of 0.6 for visual reference..

For ABI data, while the correlation between NIRv and GPP decreases notably as SZA increases, sNIRv indices exhibit greater resilience to high SZAs. Particularly noteworthy is the slower rate of the decrease in correlation observed for sNIRv indices compared to NIRv as SZA increases, indicating the potential of sNIRv indices for robust GPP estimation at lower solar elevation angle. This finding underscores the importance of considering SZA sensitivity when selecting vegetation indices for GPP estimation, especially in environments where lower solar elevation angles are prevalent, such as regions at higher latitude on the Earth or during certain seasons.

The VZA analysis revealed an interesting trend. As VZA increases, the sNIRv becomes more reliable than NIRv in its correlation with GPP (Figure 5.6). If the initial sNIRv correlation is lower than NIRv, the gap between them narrows with increasing VZA. Conversely, if the initial sNIRv correlation is higher than NIRv, the difference increases with increasing VZA. This trend is evident in both half-hourly and daily data, suggesting that sNIRv reduces sensitivity to VZA compared to NIRv. In other words, VZA has a greater impact on NIRv correlations with GPP compared to sNIRv. Lobell et al., (2002) used the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to show that SWIR reflectance, after normalizing reflectance measurements to SWIR wavelengths at 2080 nm, exhibited minimal influence from the viewing angle. However, viewing angular effects can be greater for SWIR bands (Gu et al., 2021; Lobell et al., 2002), which raises the question of whether including SWIR bands could potentially mitigate the large anisotropy from the red band.

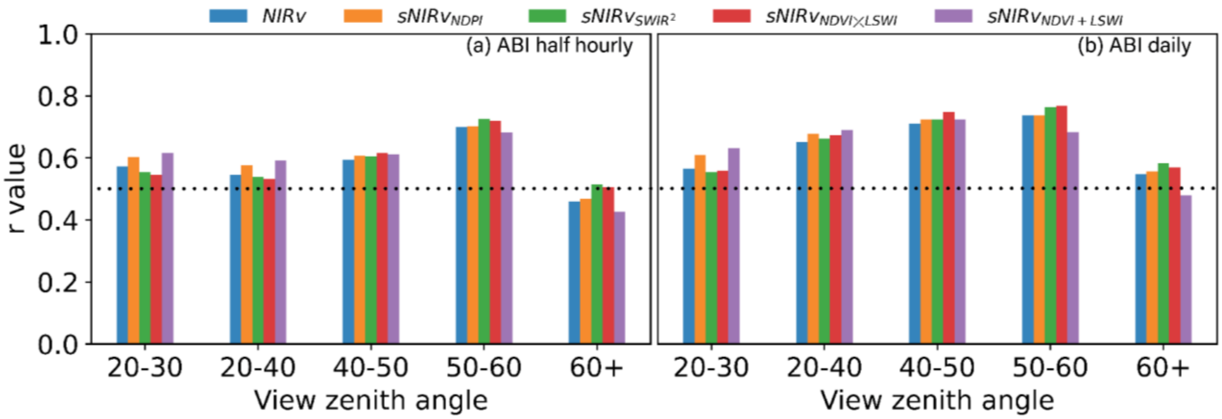

Figure 5.6. sNIRv indices performance in different view zenith angles compared to NIRv for half-hourly and daily GOES-16 ABI data. The black dotted line shows r values of 0.5 for visual reference.

5.3.4 Time series analysis across different plant functional types

We conducted an evaluation of the four most accurate VIs in the previous analysis: NIRv, $sNIRv_{NDVI\times LSWI}$, $sNIRv_{SWIR^2}$, and $sNIRv_{NDPI}$, at both half-hourly intervals (scatterplots depicted on the left side of Figure 5.7) and daily resolution (time series plots presented on the right side of Figure 5.7) across different International Geosphere-Biosphere Programme (IGBP) categories from April to September 2022, a period with significant changes in phenological stages across CONUS. We also conducted the same analysis for the full year of 2022. Our analysis revealed that $sNIRv_{NDVI\times LSWI}$ exhibited the strongest agreement with GPP across all categories, followed closely by $sNIRv_{SWIR^2}$.

Savannas displayed the most significant enhancement in r values with improvements of 0.33 for daily resolution and 0.23 for half-hourly resolution when comparing sNIRv to NIRv. Similarly, shrublands (open and closed shrublands were merged into one class) demonstrated considerable enhancement, with an increase of 0.08 in daily resolution, reaching an r value of 0.85. Additionally, sNIRv exhibited notable improvements, particularly in permanent wetlands, across both half-hourly and daily time scales. However, it is noteworthy that permanent wetlands displayed comparatively lower r values compared to all other IGBP categories.

While Deciduous Broadleaf Forests and Croplands showcased exceptionally high r values—up to 0.95 for DBF and 0.85 for croplands at the daily scale—the improvements in r values when comparing sNIRv to NIRv were relatively modest for these vegetation types.

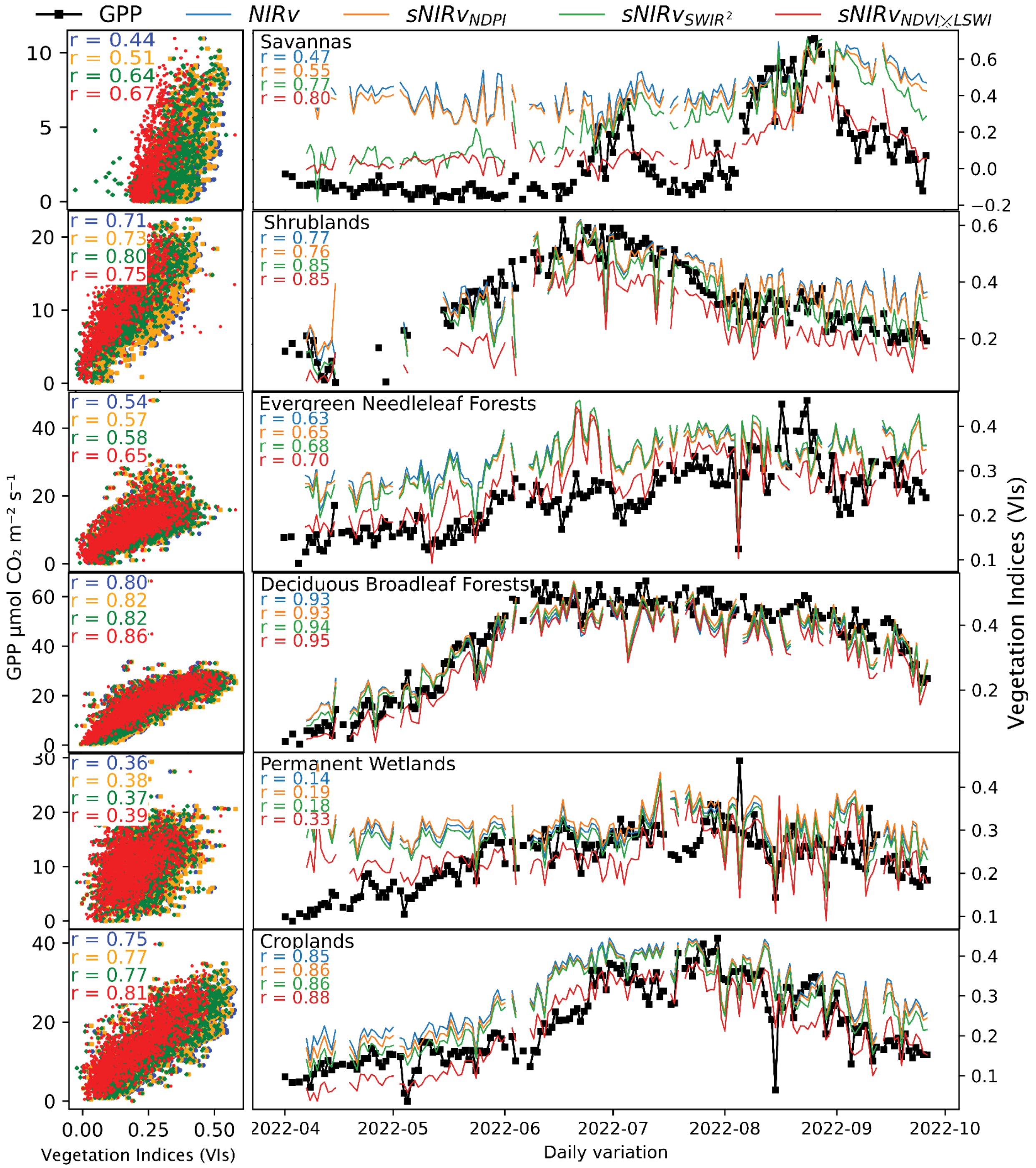

Figure 5.7: Time series plots depicting GPP and the vegetation indices (VIs) at daily resolution, alongside scatterplots between them using half-hourly ABI data across different vegetation covers categorized by the IGBP.

Across half hourly and daily temporal resolutions and land cover categories, the evaluation of VIs demonstrated $sNIRv_{NDVI\times LSWI}$ and $sNIRv_{SWIR^2}$ as having the strongest correlations with GPP. This highlights the effectiveness of these indices in capturing vegetation dynamics and productivity variations over time. Notably, the significant enhancement observed in r values for savannas and shrublands, particularly with sNIRv compared to NIRv, highlights the utility of the SWIR band in VIs for characterizing vegetation responses in sparsely vegetated ecosystems. GPP in savannas and shrublands, often characterized by sparse vegetation and exposed soil backgrounds, exhibited notable improvements in correlation with sNIRv, which was also observed for NDVI values lower than 0.25. This finding aligns with previous studies (Dechant et al., 2020, 2022; Jeong et al., 2023), which have shown a lower correlation of NIRv with GPP in savannas and shrublands. The enhanced performance of sNIRv in these ecosystems is consistent with its reduced sensitivity to background signals, leading to higher correlations with GPP. The enhanced performance of sNIRv in these ecosystems can be attributed, in part, to its reduced sensitivity to background signals, resulting in higher correlations with GPP.

The high seasonal changes in GPP in Deciduous Broadleaf Forests, attributed to leaf-out in spring and leaf fall in autumn, contribute to the pronounced variations in GPP, and relatively high r values in these ecosystems suggest that sNIRv is largely able to capture these dynamics. Similarly, croplands, subject to intensive agricultural practices and seasonal planting cycles, exhibit significant fluctuations in GPP, leading to high correlations with the selected VIs. However, the relatively modest improvements in r values when comparing sNIRv to NIRv in these categories imply that while the existing indices perform well, there is room for refinement to better capture the nuances of vegetation dynamics in such environments. These findings highlight the importance of selecting appropriate VIs tailored to specific land cover types and temporal resolutions to accurately monitor and assess ecosystem productivity and phenological changes.

5.3.5 sNIRv and fPAR relationship in SCOPE simulation

Figure 5.8 illustrates the r values between different VIs and fPAR across various categories of soil conditions, LAI, the leaf inclination function parameter 'a' (LIDFa), SZA, and VZA ranges obtained from SCOPE simulations. (LIDFa describes the orientation of leaves relative to the ground, with a more negative value indicating a more horizontal leaf angle distribution.) Among the indices evaluated, $sNIRv_{SWIR^2}$ and $sNIRv_{NDVI+LSWI}$ consistently exhibit the highest r values across all categories unless LAI is larger than 2 m^2 m^{-2}, followed closely by $sNIRv_{NDPI}$ with an average of 0.03 lower r values. Interestingly, $sNIRv_{NDVI\times LSWI}$, despite being identified as one of the best indices and consistently outperforming NIRv in previous analyses based on ABI data, displays lower r values compared to NIRv in the SCOPE simulations, similar to MODIS data.

When analyzing various soil conditions, the performance of $sNIRv_{SWIR^2}$ is notably enhanced in soil condition 1 (soil1), maintaining higher r values compared to NIRv in soils 2 and 3. SCOPE simulations using snow spectra demonstrated the highest correlation between fPAR and $sNIRv_{NDVI+LSWI}$, closely followed by NIRv. Data from snow spectra significantly reduced the correlation between fPAR, $sNIRv_{SWIR^2}$, and $sNIRv_{NDPI}$.

Analysis of LAI indicates that r values for sNIRv indices are highest for LAI values less than 2 m^2 m^{-2}, with a considerable drop observed for LAI values exceeding 2 m^2 m^{-2}. Notably, for LAI values greater than 2 m^2 m^{-2}, NIRv demonstrates higher r values compared to all other indices. Moreover, the $sNIRv_{SWIR^2}$ method shows a significant increase in r values when leaf angle distribution is more horizontal (LIDFa <-0.5). Horizontal leaves allow for better canopy light capture, explaining the observed improvement in r values.

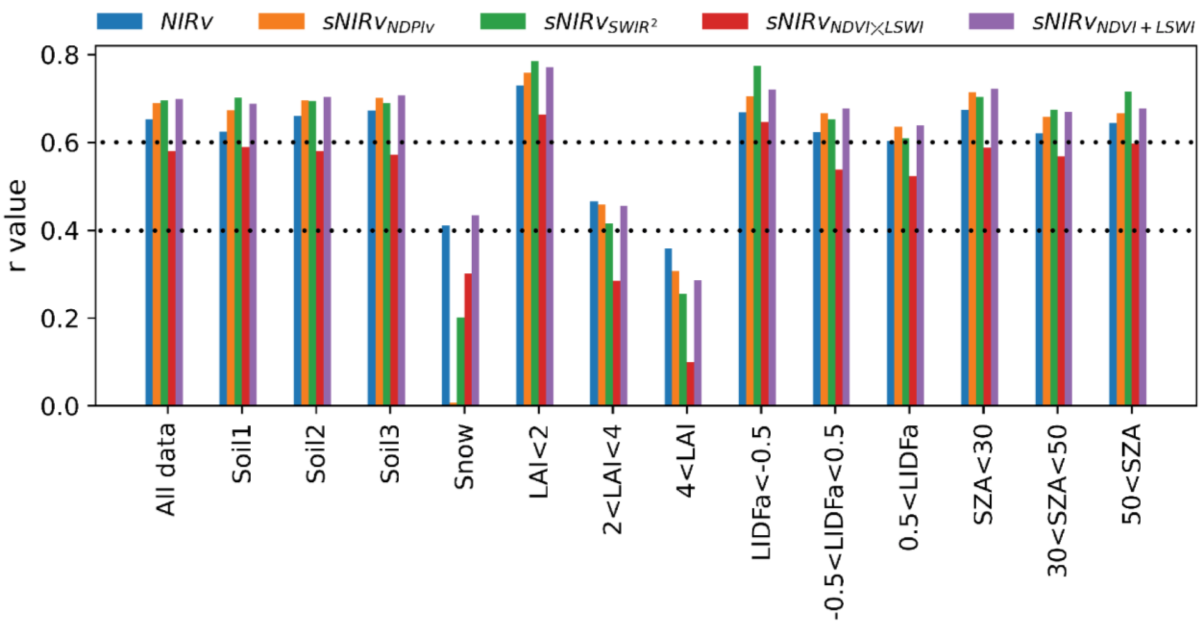

Figure 5.8. Correlation coefficients (r-values) of the VIs with fPAR as a proxy of GPP using data from SCOPE simulation. The black dotted lines show r values of 0.4 and 0.6 for visual reference.

Our findings from SCOPE also suggest a potential link between vegetation cover density and the effectiveness of VIs for fPAR estimation, particularly at LAI values less than 2 m^2 m^{-2}. Sparser vegetation cover, characteristic of low LAI, leads to stronger background signals in remote sensing data. Prior analysis also showed that sNIRv is less sensitive to background information compared to NIRv. This could explain the superior performance of sNIRv-based VIs ($sNIRv_{SWIR^2}$ and $sNIRv_{NDVI+LSWI}$) in these conditions. These observations highlight the importance of considering the interplay between vegetation density and background effects when selecting VIs for fPAR retrieval.

Furthermore, the data simulations using snow backgrounds provide valuable insights into VI performance under distinct environmental conditions. Notably, only $sNIRv_{NDVI+LSWI}$ exhibited improved performance compared to NIRv under snow cover. All other sNIRv indices displayed low correlations with fPAR in this scenario. In particular, the correlation values for $sNIRv_{SWIR^2}$ and $sNIRv_{NDPI}$ dropped significantly, indicating their high susceptibility to snow. This suggests that VI effectiveness in fPAR estimation can vary considerably depending on the presence of snow cover. Therefore, careful consideration of environmental factors is crucial for robust vegetation monitoring and modeling efforts. Snow cover of course is associated with reduced or negligible GPP, but understanding how different environmental conditions impact VIs provides additional insight into their characteristics.

Moreover, we explored the data for individual soil types. Specifically, Figure 5.4 shows the r values between different VIs and fPAR across various categories of LAI, the leaf inclination function parameter ‘a’ (LIDFa), SZA, and VZA ranges obtained from SCOPE simulations for soil1 (dry) and soil 3 (SWC=30%). Our findings revealed that variations across different LAI and angles including LIDFa, SZA, and VZA are factors beyond soil background that influence NIRv and sNIRv differences. Notably, we observed that SWC played a significant role in this context. For instance, soil1, characterized by a dried soil, exhibited a more pronounced enhancement in $sNIRv_{SWIR^2}$ compared to NIRv, with an increase of 0.18 in r-values for LIDFa values below -0.5. Conversely, soil3, with a higher SWC of 30%, demonstrated a comparatively lesser improvement, with only a 0.08 increase in r-values for LIDFa values below -0.5. These results suggest that soil background exerts a more pronounced influence on SWIR band behavior under conditions of higher SWC, leading to reduced correlation with fPAR as a proxy for GPP.

Figure 5.9 presents a density scatter plot visualizing the relationships between VIs (NIRv and four sNIRv) and fPAR derived from the SCOPE simulation data. $sNIRv_{SWIR^2}$ has the strongest correlation (with highest r-value of 0.77), closely followed by $sNIRv_{NDPI}$ and $sNIRv_{NDVI+LSWI}$. Interestingly, NIRv exhibits higher r-value of 0.71 compared to $sNIRv_{NDVI\times LSWI}$ (the top performance index in previous analysis for ABI data) with r value of 0.66, suggesting a potentially more robust linear relationship between NIRv and fPAR.

Comparing RMSE of the fitted linear regressions offers further insights. Notably, $sNIRv_{SWIR^2}$, $sNIRv_{NDPI}$, and $sNIRv_{NDVI+LSWI}$ achieved RMSE values of 0.10. Conversely, $sNIRv_{NDVI\times LSWI}$ exhibited a higher RMSE (0.12), indicating a potentially disproportionate increase in fPAR estimates with increasing $sNIRv_{NDVI\times LSWI}$ values. This observation aligns with the visual assessment in Figure 5.9, where the density scatter plots depict the relationships between various vegetation indices (VIs) and fPAR.

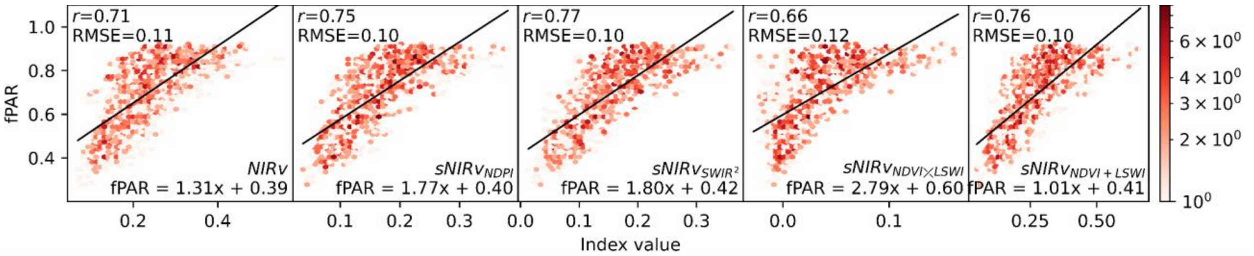

Figure 5.9. Density scatter plots between the VIs and fPAR using data from SCOPE simulations. The black line represents the linear fitted line on the heatmap.

5.3.6 Unveiling key drivers of GPP prediction using ML through SHAP analysis

Building upon the exploration of sNIRv-GPP relationships, this section evaluates the effectiveness of these VIs as predictor variables for GPP modeling using GBR at half-hourly temporal resolution. To account for the influence of photosynthetically active radiation (PAR) on vegetation productivity, we incorporated Downward Shortwave Radiation (DSR) as a feature into the GBR model. This approach aligns with previous research suggesting the effectiveness of PAR-adjusted VIs for GPP estimation (Baldocchi et al., 2020; Chang et al., 2020; Dechant et al., 2020, 2022; Mallick et al., 2024). Here, we considered two sets of feature inputs: one based on ABI reflectance products plus the DSR feature, and the other including the VIs, the DSR feature, and the SWIR1 band from ABI reflectances that was not used in the VIs. In the first scenario, we achieved an average R of 0.80 and RMSE of 5.86 μmol m^{-2} s^{-1} through ten iterations of the LOOCV algorithm. The second scenario achieved an average R of 0.82 and RMSE of 5.72 μmol m^{-2} s^{-1}. To gain insights into feature importance, we applied SHAP analysis on the trained model using categorized data by different NDVI values, soil water content (SWC) levels, and IGBP categories. The results are visualized in Figure 5.10 for the first scenario and in Figure 5.11 for the second scenario.

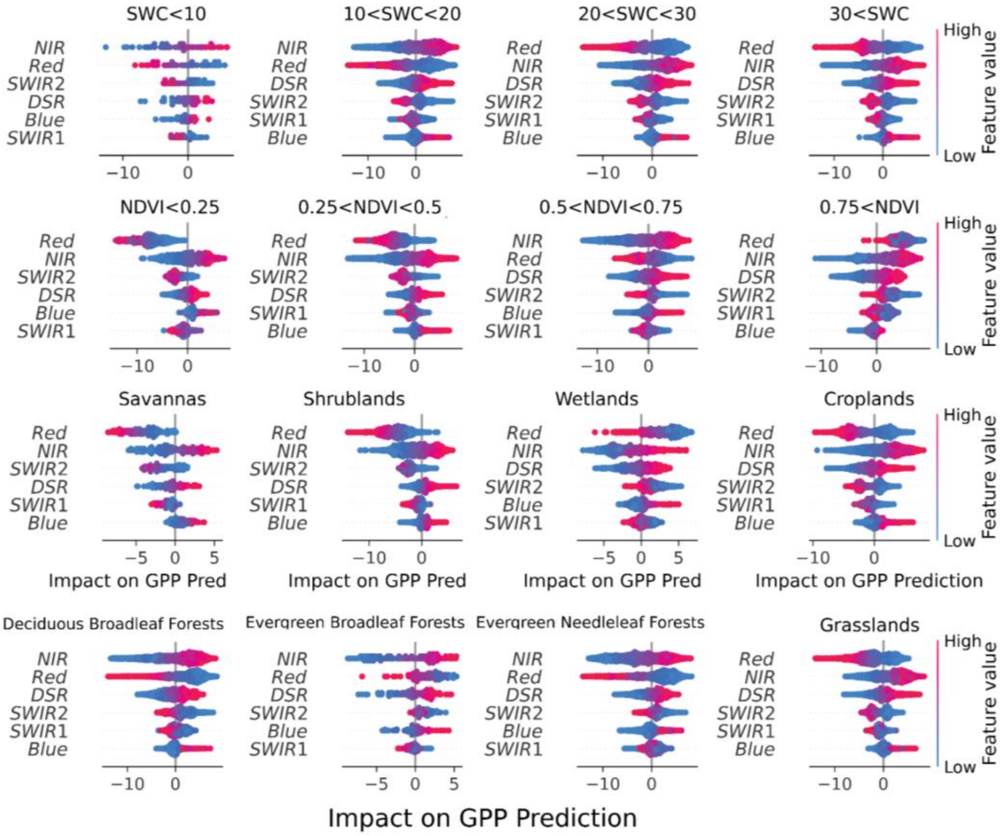

Figure 5.10: SHAP feature importance analysis for the first scenario categorized by different soil water content (SWC, %) levels (top row), NDVI values (second row), and IGBP categories (third and fourth rows). Variables are ordered from top to bottom based on their overall importance rank.

In the first modeling scenario, using ABI reflectance products with the DSR feature, Red and NIR bands exhibited consistent prominence across all categories, with SWIR2 surpassing DSR for SWC below 10% and NDVI below 0.5 (see Figure 5.10). Observations from NDVI and IGBP categories also support previous findings regarding the lower sensitivity of sNIRv to soil background signals. Notably, in scenarios with NDVI less than 0.5 and within savanna and shrubland IGBP categories, SWIR2 emerges as the third most important feature, suggesting its effectiveness in mitigating soil background effects. Interestingly, SWIR1 exhibits lower effectiveness in GPP prediction compared to SWIR2 across all categories, supporting our selection of SWIR2 for VI investigation. Further analysis of SWC reveals that SWIR2 is more effective when SWC is less than 10%, as it is less sensitive to soil background and exhibits stronger correlation with GPP. Conversely, higher SWC levels exacerbate the influence of soil background on SWIR2, resulting in reduced correlation with GPP. Different studies have demonstrated that SWIR bands are sensitive to SWC (Babaeian et al., 2019; Norouzi et al., 2021; Tian & Philpot, 2015). As SWC increases, the influence of the soil background on the SWIR2 index also intensifies (Lobell & Asner, 2002; Lingli Wang et al., 2008), which can lead to a heightened sensitivity of the sNIRv indices to the soil background. Intriguingly, Red's superior influence on GPP prediction over NIR emerged notably for SWC exceeding 20%, which suggests that NIR is also being impacted by soil background with higher soil moisture levels (Amani et al., 2016, 2017; D. B. Lobell & Asner, 2002). In the context of land cover types, SWIR2 emerged as a pivotal feature, especially in Savannas and Shrublands, surpassing NIRv performance. In these land cover types, NIRv showed the lowest performance (Dechant et al., 2022; Jiang et al., 2021). In other categories, following NIR, Red, and DSR, SWIR2 ranked as the fourth most influential feature for GPP prediction.

In the second scenario, which utilizes VIs, the DSR feature, and the SWIR1 band, $sNIRv_{NDVI\times LSWI}$ consistently ranks as the most informative feature for GPP modeling across all SWC, NDVI and IGBP categories in Figure 5.11. This suggests the GBR model heavily relies on this combined information for model development. DSR consistently ranks second in importance for 13 out of 16 categories. However, in Savannas and Shrublands, SWIR1 alone showed a greater influence on GPP prediction compared to DSR. Interestingly, features like EVIv and SWIR1 emerged as more important than NIRv, ranking among the top four most influential features for GPP prediction. Conversely, $sNIRv_{NDPI}$ and $sNIRv_{NDVI+LSWI}$ consistently rank as the least important features across all categories.

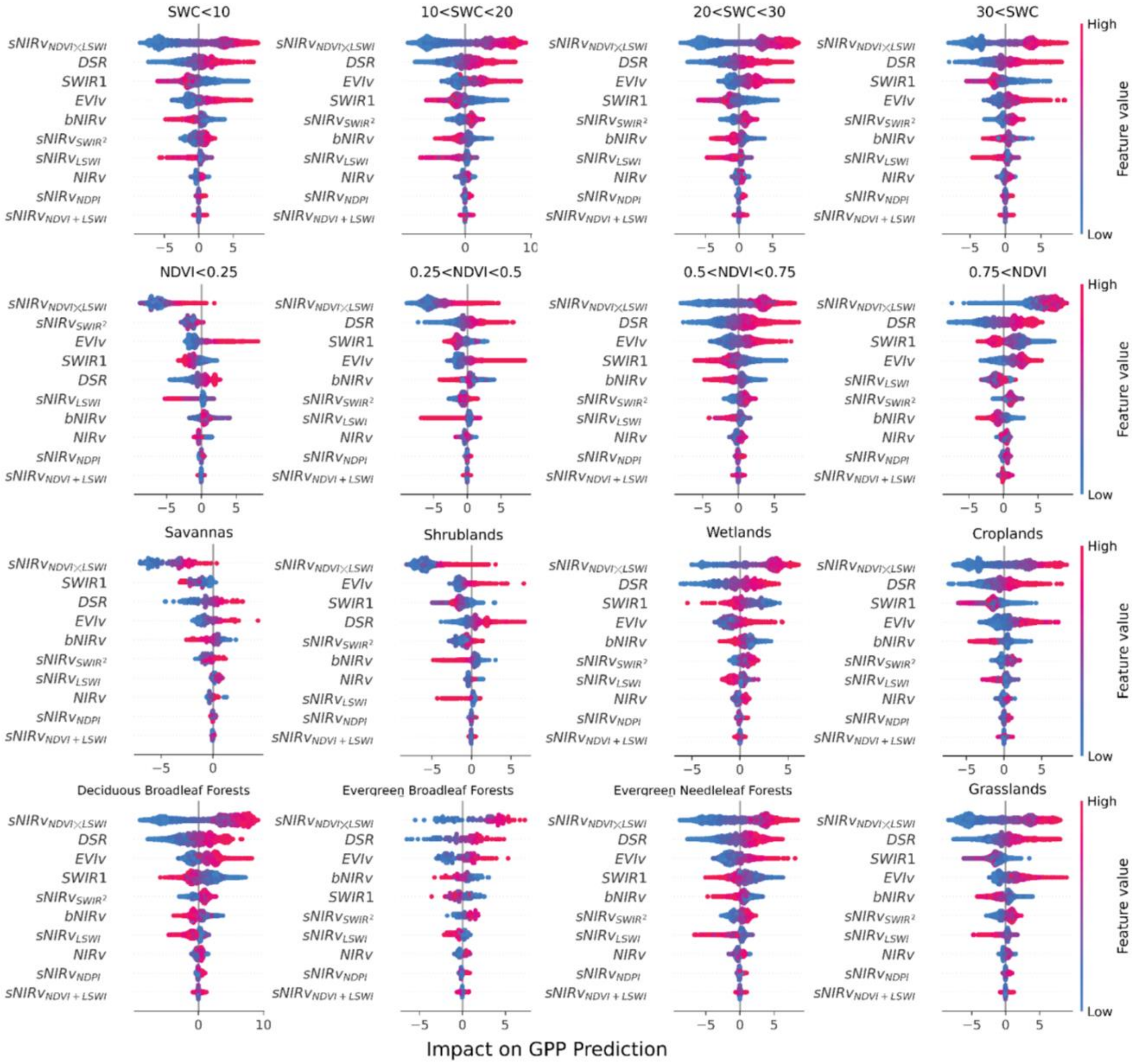

Figure 5.11: SHAP feature importance analysis for the second scenario categorized by different soil moisture (SWC) levels, NDVI values, and IGBP categories. Variables are ordered from top to bottom based on their overall importance rank.

5.3.7 sNIRv for vegetation analysis: gains, conditions, and mechanisms

Through our analysis of various VIs, which include or exclude SWIR band information under different environmental conditions, the enhanced ability of $sNIRv_{NDVI\times LSWI}$ and $sNIRv_{SWIR^2}$ to capture GPP dynamics using ABI becomes apparent, particularly for low NDVI values. From our analysis, we determined that using SWIR bands, especially SWIR2, improves GPP estimation using ABI data but not the MODIS Nadir BRDF-Adjusted product. We briefly describe key findings with an eye toward a mechanistic explanation of observations for improving satellite GPP monitoring using geostationary observations.

The observed correlation patterns highlights the influence of NDVI ranges on VI-GPP correlations, with $sNIRv_{NDVI\times LSWI}$ exhibiting strong correlations across low NDVI values, suggesting its potential for accurate GPP estimation in environments with sparse vegetation cover (Dechant et al., 2022; Jiang et al., 2021). Furthermore, SWC impacts the GPP estimation using sNIRv. While SWIR2 offers strong predictive power for GPP at lower SWC levels (SWC<10%), its effectiveness weakens as soil moisture increases due to enhanced soil background influences. This is consistent with the finding SWIR bands are highly sensitive to higher soil moisture levels (Amani et al., 2016, 2017; Babaeian et al., 2019), leading to a stronger soil background effect on sNIRv and a consequent reduction in the correlation between sNIRv and GPP. A similar trend is observed for the NIR band, where SWC exceeding 20% weakens its efficacy for GPP prediction due to increased soil background effects.

Angular effects are greater for SWIR bands (Gu et al., 2021; Lobell et al., 2002) yet they can also provide key information on the soil background signal and ecosystem water content. It is likely that the improvement introduced by adding SWIR to VIs, especially from SWIR2 from geostationary platforms, is a combination of angular, soil, and hydrological effects. Notably, $sNIRv_{SWIR^2}$ emerges as a strong VI, particularly in scenarios with low LAI and vertical and/or horizontal leaf angles, showcasing its effectiveness in mitigating background signals and capturing vegetation dynamics accurately.

Finally, comparing VI-GPP correlations for daily resolution data derived using ABI data at a single time (at noon time, see Figure 5.4) and averaged half-hourly data (Figure 5.5) demonstrates that the one derived from half hourly data capture subtle changes in GPP better than single measurements, resulting in a more accurate representation of GPP dynamics (evident in the higher r-values of Figure 5.5, with an average increase of 0.12).

5.4 Conclusions

Understanding GPP is crucial for studying global carbon cycling and its impact on Earth's climate. The VIs based solely on NIR and Red bands often struggle in challenging scenarios like low vegetation cover, high solar angles, and varying viewing geometries. This study introduces a novel class of VIs, sNIRv, which incorporates SWIR bands alongside NIR and Red. Our findings demonstrate that sNIRv-based VIs, particularly the $sNIRv_{NDVI\times LSWI}$ outperform NIRv when using ABI data by establishing stronger correlations with GPP across diverse land covers and improving GPP estimation at low NDVI values which represent higher levels of soil background reflectance. While further research is needed to fully understand the implications of SWIR bands for nadir viewing geometries in dense vegetation, this work highlights the potential of sNIRv-type VIs for robust GPP monitoring across various ecosystems and GOES-R ABI observations. By introducing pathways to integrate SWIR bands, this research paves the way for the development of improved VIs for more accurate GPP estimation, ultimately contributing to a better understanding of global carbon cycling.

Chapter 6. Near Real-Time Estimation of Daytime and Nighttime Evapotranspiration Using GOES-R Observations and Machine Learning Models

This chapter is drafted and intended to submitted to the Wiley-Water Resources Research:

Abstract

Evapotranspiration (ET) is a critical component of the water cycle, influencing climate, agriculture, and water resource management. However, most satellite-derived ET products are limited to daily or coarser temporal resolutions, despite the strong diurnal variability of ET processes. Existing satellite-based ET retrievals are largely restricted to daytime conditions, when nighttime ET is a small but often non-trivial flux. In this study, we introduce the Advanced Baseline Imager Live Imaging of Vegetated Ecosystems ET (ALIVE_{ET}), a near real-time, five-minute ET estimation framework, leveraging geostationary satellite observations from the GOES-R Advanced Baseline Imager (ABI) and machine learning models under both clear and cloudy conditions. We test Gradient Boosting Regression (GBR) and Long Short-Term Memory (LSTM) models to assess their ability to estimate ET variations across the diurnal cycle. GBR captures daytime ET with an R^2 of 0.74 (RMSE of 0.059 mm hh^{-1}) while maintaining low computational cost. For nighttime ET, where R^2 decreases by about 0.50 compared to daytime, LSTM models trained on time-series observations perform better, achieving an R^2 of 0.24 (RMSE of 0.014 mm hh^{-1}) by leveraging temporal dependencies in land surface temperature (LST) and past ABI observations. Comparisons against daily ET estimates from the physically-based ALEXI model demonstrates good agreement but opportunities for improvement. This study demonstrates the potential of integrating machine learning with geostationary remote sensing to advance high-temporal-resolution ET estimation.

Keywords - Evapotranspiration, Machine Learning, GOES-R, Nighttime ET, Near Real-Time Water Cycle Monitoring

6.1 Introduction

As the second-largest flux in the terrestrial water cycle, evapotranspiration (ET) returns approximately 60–80% of terrestrial precipitation to the atmosphere, eventually recycling nearly all of it, thereby influencing regional and global climate patterns, water availability, and ecosystem dynamics (Peterson et al., 1995; Tateishi and Ahn, 1996; Van Der Ent et al., 2010). It also plays a fundamental role in the carbon cycle through the coupling of its dominant term – transpiration through vegetation – with carbon dioxide uptake via stomatal function (Katul et al., 2012; Pan et al., 2020). Accurate ET estimation is essential for hydrological modeling, drought assessment, and sustainable agricultural water management, particularly in the face of increasing global food demand and freshwater scarcity (Sabir et al., 2024; Wanniarachchi and Sarukkalige, 2022). ET varies dynamically over the course of a typical day in response to environmental variability and plant hydrological stresses. Ecosystem models struggle to simulate these dynamics, suggesting gaps in our knowledge of key processes at the ecosystem scale (Brighenti et al., 2019; Matheny et al., 2014). Retrieving sub-daily ET estimates from satellite observations to estimate these dynamics at larger spatial scales also remains challenging due to cloud cover, sensor limitations, and the need for robust methodologies to address data gaps (Qin et al., 2022; Ranjbar et al., 2024c; Wang et al., 2023; Wanniarachchi and Sarukkalige, 2022). Moreover, most studies focused on satellite-based ET retrievals are limited to daytime, assuming that nighttime ET negligible, when it is often small but non-zero and important to understand especially in arid and semi-arid ecosystems (Krishnan et al., 2012; Tabari et al., 2012) for improving water balance assessments and land-atmosphere exchange modeling. Addressing these challenges is critical for advancing climate monitoring, optimizing irrigation strategies, and improving our understanding of land surface processes.

Remote sensing has revolutionized ET estimation, offering large-scale assessments through optical, thermal, and microwave sensors (Wanniarachchi and Sarukkalige, 2022). Methods such as the Surface Energy Balance Algorithm for Land (SEBAL, (Bastiaanssen et al., 1998)) and the Atmosphere-Land Exchange Inverse (ALEXI, (Anderson, 1997)) model leverage thermal infrared observations to estimate ET by solving the surface energy balance equation (Anderson et al., 2012; Wanniarachchi and Sarukkalige, 2022). Similarly, models like the MODIS-based Penman-Monteith (McColl, 2020; Penman, 1948) and the Priestley-Taylor Jet Propulsion Laboratory (PT-JPL, (Fisher et al., 2008; Ling et al., 2022)) frameworks integrate satellite-measured vegetation indices and land surface temperature (LST) with meteorological inputs for global ET estimation (Fisher et al., 2008).

Existing ET products face inherent trade-offs between spatial and temporal resolution, as well as data latency. MODIS-based ET products, for instance, provide global coverage but are constrained to daily or eight-day resolutions, limiting their effectiveness in monitoring diurnal ET variations (C. Zheng et al., 2022). Similarly, Landsat-derived ET estimates offer higher spatial resolution (30–100 m) but have extended revisit times of 8 to 16 days that are unable to observe sub-daily processes and may miss rapid temporal changes in ET (Bai et al., 2017; Yang et al., 2013). ECOSTRESS, on the International Space Station, provides unique opportunities to estimate ET across different times of day, but often requires extensive extrapolation to infer diurnal patterns (Fisher et al., 2020; Hu et al., 2022; Meerdink et al., 2019).

Thermal infrared-based estimates of ET have a strong physical basis but are additionally challenged by clouds, which obviously have a strong impact on ET by altering the surface energy balance. Microwave-based ET retrievals, such as those from SMAP or AMSR-E (Sun et al., 2012; Walker et al., 2019), offer all-weather capabilities but operate at coarser spatial scales (~10–50 km) (Sun et al., 2012). ET products derived from data fusion methods, such as OpenET, integrate outputs from different remote sensing-based ET models to overcome existing limitations. OpenET utilizes six well-established models to provide high-resolution (30-m spatial, daily temporal) ET data, designed to support efficient water management and agricultural decision-making (Melton et al., 2022). However, OpenET's dependence on multiple data sources makes it vulnerable to data gaps and its daily resolution does not capture sub-daily ET variations, restricting its ability to monitor ET dynamics on a diurnal scale.

Geostationary satellites, such as the GOES-R series, present a transformative opportunity for high-frequency ET mapping by providing reflectances, brightness temperatures and derived environmental variables every 5-10 minutes (Khan et al., 2021b; Ranjbar et al., 2024c, 2024d). Recent advancements in DSR and LST estimation show that brightness temperature data—even under clouds—can be used through both physically-based and machine learning approaches (Liu et al., 2023; Ranjbar et al., 2024b, 2024d; Zhao and Duan, 2020). The Advanced Baseline Imager (ABI) onboard GOES-R captures spectral information from visible to thermal wavelengths across all sky conditions, enabling the inference of environmental variables based on cloudiness, radiation, and moisture (Table 6.1). This capability allows for continuous monitoring of diurnal ET variations, providing unique insights into water and energy fluxes at sub-hourly scales (Khan et al., 2021b). Physical remote sensing approaches struggle with missing data due to cloud contamination and limited nighttime observations (Ranjbar et al., 2024c, 2024d; Zhao and Duan, 2020). To overcome these limitations, we leverage machine learning techniques to estimate ET under various sky conditions, including nighttime retrievals. Specifically, we employ Gradient Boosting Regression (GBR, (Friedman, 2001)) to capture nonlinear relationships between input features (Cai et al., 2020; Dhake et al., 2023) and Long Short-Term Memory (LSTM, (Hochreiter and Schmidhuber, 1997)) networks for temporal dependencies to model ET on a five-minute interval across the GOES-16 CONUS scene. The integration of high-frequency observations from geostationary satellites with machine learning models enhances near real-time ET estimation under all-sky conditions, spanning regional to continental scales (He et al., 2019; Jeong et al., 2023b; Khan et al., 2021b; Ranjbar et al., 2024c). We discuss the benefits and limitations of our approach, which we call ALIVE$_{ET}$ (Advanced Baseline Imager Live Imaging of Vegetated Ecosystems) and compare against a physically-based model that also uses geostationary observations, ALEXI, with an eye toward clarifying the role of machine learning-based models in water cycle science.

6.2 Materials and Methods

6.2.1 Tower observations and ET calculations

We compiled micrometeorological data from AmeriFlux and NEON eddy covariance towers following the methodology outlined in (Losos et al., 2024a). These publicly available datasets include half-hourly (or occasionally hourly) fluxes of carbon dioxide, water, heat, meteorological parameters, and radiometric variables, which undergo standard quality control (Pastorello et al., 2020, 2017; Sturtevant et al., 2022). Data were filtered using quality control criteria, including the application of a friction velocity (u*) threshold to address insufficient nighttime turbulence (Reichstein et al., 2012b). We then synchronized tower data with GOES-16 observations and data products (section 2.2), resulting in a time series for 101 locations across the Contiguous United States (CONUS, Figure 6.1 and Table 6.A1).

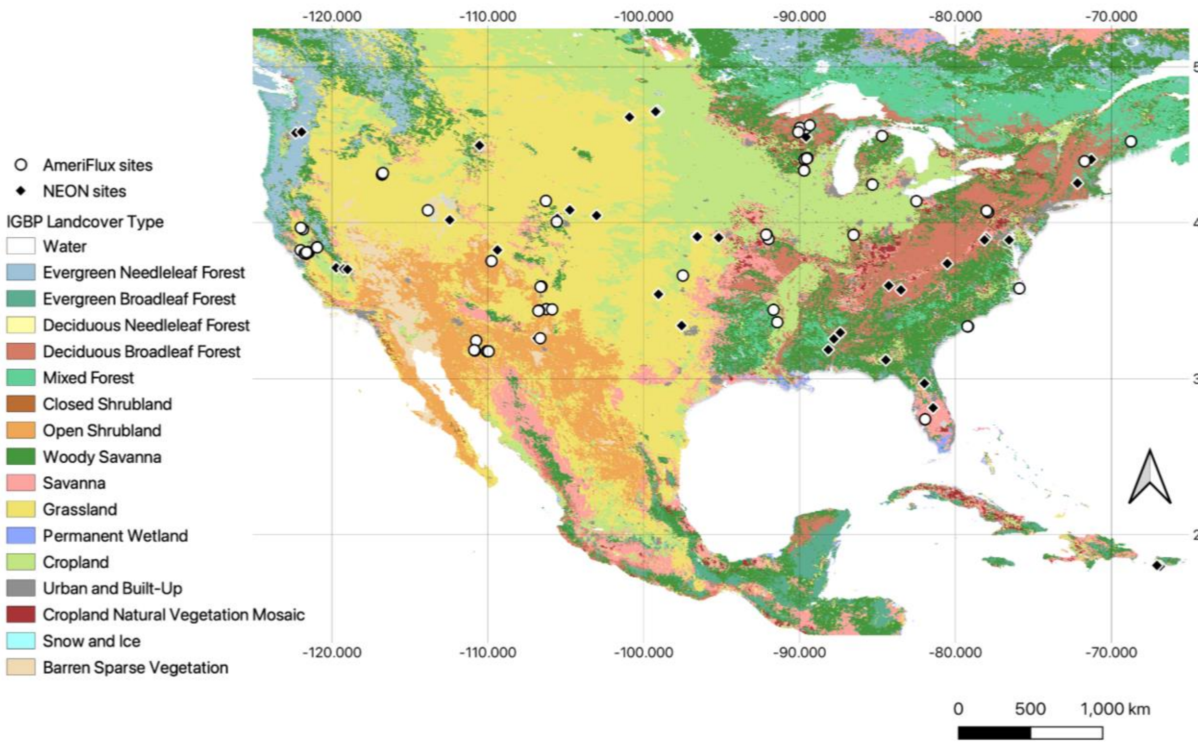

Figure 6.1. Locations of AmeriFlux and NEON sites used in this study, overlaid on the International Geosphere-Biosphere Programme (IGBP) land cover classification.

We used latent heat flux (LE) measured by the towers to estimate actual ET using Eq. 6.1 (Allen et al., 1998), which we used as the training target and in situ data for validation and testing.

$$ET = \frac{LE}{\lambda} \qquad \text{Eq. 6.1}$$

Where λ is the latent heat of vaporization. Assuming a water density of 1000 kg m^{-3}, this yields ET in mm per time. The latent heat of vaporization varies slightly with temperature and can be computed using Eq. 6.2 (Allen et al., 2006, 1998):

$$\lambda = 2.501 - (2.361 \times 10^{-3}) \times Ta \qquad \text{Eq. 6.2}$$

Where Ta is air temperature in °C, and λ is in MJ kg^{-1}.

6.2.2 ABI observations from GOES-R satellite

We used the Level 2 Cloud and Moisture Imagery (CMI) observations and land surface bidirectional reflectance factor (BRF) products from the ABI aboard the GOES-16 satellite, which has been operational since November 2017. The ABI captures data across 16 spectral bands, spanning the visible, near-infrared (NIR), shortwave infrared (SWIR), and infrared (IR) wavelengths, with spatial resolutions ranging from 0.5 to 2 km at nadir, depending on the specific band (Table 6.1, Goodman et al., 2019; Khan et al., 2021; Schmit et al., 2017). The CMI product provides top-of-atmosphere observations whereas the BRF product offers atmospherically surface reflectance data for the reflective bands at visible, NIR and SWIR channels (BRF01, BRF02, BRF03, BRF05, BRF06, detailed in Table 6.1). The ABI provides full-disk imagery every 10 minutes and CONUS imagery every 5 minutes in its typical scan mode (Mode 6, (Goodman et al., 2019; He et al., 2019)). For our analysis, we synchronized ABI data from 2019 to 2022 with ground-based eddy covariance measurements from 101 AmeriFlux and NEON sites (Losos et al., 2024a) across the GOES-R CONUS scene.

Table 6.1. Summary of ABI bands and their primary uses (Schmit et al., 2017b).

ABI Band	**Wavelength (µm)**	**Band Type**	**Primary Uses**
C01/BRF01	0.47	Visible	Monitoring aerosols (smoke, haze, dust) Air quality monitoring through aerosol optical depth measurements
C02/BRF02	0.64	Visible	Daytime monitoring of clouds (0.5-km spatial resolution) Volcanic ash monitoring
C03/BRF03	0.86	NIR	High contrast between water and land Assess land characteristics, including flooding, burn scars, and hail damage
C04	1.37	NIR	Thin cirrus detection during the day Volcanic ash monitoring
C05/BRF05	1.6	NIR	Daytime snow, ice, and cloud discrimination Input to “Snow/Ice vs. Cloud” RGB
C06/BRF06	2.24	NIR	Cloud particle size, snow, and cloud phase Hot spot detection at emission temperatures >600K
C07	3.9	IR	Low stratus and fog detection Fire/hot spot detection and volcanic ash
C08	6.2	IR	Upper-level feature detection (jet stream, waves)
C09	6.9	IR	Mid-level feature detection
C10	7.3	IR	Low-level feature detection (EML, fronts)
C11	8.4	IR	Cloud-top phase and type products

			Volcanic ash (SO$_2$ detection) and dust
C12	9.6	IR	Dynamics near the tropopause Input to Airmass RGB
C13	10.3	IR	Less sensitive to atmospheric moisture brightness temperatures warmer than IR
C14	11.2	IR	IR window Low stratus and fog detection
C15	12.3	IR	Greater sensitivity to moisture cooler brightness temperatures
C16	13.3	IR	Mean tropospheric air temperature estimation Highlights high, cold, likely icy clouds

In addition to ABI observations, we incorporated Solar Zenith Angle (SZA) data, ALIVE-derived all-sky Downwelling Shortwave Radiation (ALIVE$_{\text{DSR}}$, (Ranjbar et al., 2024d)) and ALIVE-derived all-sky Land Surface Temperature (ALIVE$_{\text{LST}}$, (Ranjbar et al., 2024b)) estimates. From the BRF bands, we calculated the NIR reflectance of vegetation (NIRv, (Dechant et al., 2022b)) and SWIR-enhanced NIRv (sNIRv, (Ranjbar et al., 2024a)) indices as proxies for vegetation productivity due to their strong correlation with canopy gross primary productivity and its strong relationship to ET. For the ET modeling, we used different configurations of variables for the daytime and nighttime. For nighttime, we focused on CMI bands along with SZA, ALIVE$_{\text{DSR}}$, ALIVE$_{\text{LST}}$, and the mean values of NIRv and sNIRv from the previous day (NIRv_daily_mean and sNIRv_daily_mean). For daytime, we included CMI, BRF, SZA, ALIVE$_{\text{DSR}}$, ALIVE$_{\text{LST}}$, NIRv, and sNIRv. We explored two feature strategies: one considering single real-time observations and another incorporating time series features from the previous 24 hours.

To assess and validate the performance of our machine learning models, we compared against the physically-based Atmosphere-Land Exchange Inverse (ALEXI, (Anderson, 1997)) ET product into our analysis. ALEXI, developed with support from NASA, NOAA, and USDA ARS since 2003, estimates daily land-surface energy fluxes at a 5–10 km resolution using thermal infrared observations and vegetation indices from satellites. By combining a two-source (soil + canopy) energy balance model (TSEB) with an atmospheric boundary layer model, ALEXI provides accurate ET estimates under both clear and cloudy conditions (Anderson et al., 2012, 2007; Talib et al., 2021; Wanniarachchi and Sarukkalige, 2022). ALEXI is expanding globally through integration with international satellites, and it has been validated using flux tower data (RMSE of 35-40 W/m^2), making it a critical tool for ET and climate monitoring (Pan et al., 2020; Wanniarachchi and Sarukkalige, 2022; C. Zheng et al., 2022). To ensure consistency, we compared our $ALIVE_{ET}$ estimates from the empirical machine learning approach with the physical-based ALEXI (Sun et al., 2017).

6.2.3 Machine learning modeling and assessment

After synchronizing satellite-based observations and derived $ALIVE_{DSR}$ and $ALIVE_{LST}$ products with tower-based measurements, we used the tower-derived ET measurements (described in subsection 2.1) as the target for estimation. For predictors, we utilized the satellite-based features outlined in subsection 2.2 in two strategies. In the first, we considered the single-time observations at the time of measurement, while in the second, we incorporated time series features from the preceding 24 hours.

We applied two machine learning regression models: GBR and LSTM. GBR has demonstrated strong performance in estimating downwelling shortwave radiation (Ranjbar et al., 2024d), LST (Ranjbar et al., 2024b), and surface-atmosphere carbon dioxide flux (Ranjbar et al., 2024c), providing a balance between accuracy and computational efficiency for real-time applications. We compared GBR against LSTM, given its effectiveness with time series data (Dhake et al., 2023), which excels in capturing long-term dependencies in sequential data (Dhake et al., 2023; Hochreiter & Schmidhuber, 1997; Reddy & Prasad, 2018) which we felt may improve estimation of nighttime ET, dominated by soil evaporation, which is often modeled as a simple function of time since precipitation (Brutsaert, 2014).

GBR is an ensemble learning model, refining predictions iteratively by optimizing a loss function through decision trees (Friedman, 2001). Key hyperparameters such as number of estimators, maximum depth, minimum samples per leaf, and learning rate influence the model's accuracy, complexity, and convergence speed (Bentéjac et al., 2021; Sahin, 2020). LSTM networks, designed for sequential data, utilize a memory cell architecture to address non-linear dependencies in time series forecasting (Ghanbari et al., 2021; Sutskever et al., 2014). Our LSTM architecture included two LSTM layers, followed by dense layers for feature extraction and output generation. The first LSTM layer retained the entire sequence (return_sequences=True), while the second layer summarized the learned information (return_sequences=False). The time series features in LSTM models are structured in a 3D input format, consisting of samples, time steps, and features. In contrast, since the GBR model accepts a 2D input, we explicitly included past values by incorporating the preceding 24 hours of time series features as individual predictors, with each hour treated as a separate feature. Models were trained on Google Colab Pro (32 GB RAM, A100 GPU) using a grid search algorithm to optimize hyperparameters for correlation determination (R^2) and prediction speed (P.T., Table 6.2).

Table 6.2 Grid search specifications and hyperparameter setting for machine learning models.

Model	**Hyperparameter**	**Grid Setting**
GBR	Number of estimators	100, 300, 500, 800
	Maximum depth	5, 8, 10, 12, 15
	min_samples_leaf	50, 100, 200, 500
	Learning rate	0.01,0.05,0.1
LSTM	Units (neurons)	128, 256, 512, 1024
	Activation	'logistic', 'tanh', 'relu'
	Optimizer	'adam','sgd'
	Batch Size	32, 64, 128, 256, 512

For validation, we applied Leave-One-Out Cross-Validation (LOOCV), reserving 20% of eddy covariance sites for testing while training on the remaining sites using a four-fold cross-validation (75:25 training-validation split) (Maxwell et al., 2018).

This procedure was repeated ten times with different random seeds to enhance robustness and minimize bias. Final results, including correlation determination (R^2, Eq. 6.3), root mean square error (RMSE, Eq. 6.4), and normalized RMSE (nRMSE, Eq. 6.5) were averaged to ensure reliability, using the Scikit-learn Python library (Pedregosa et al., 2011).

$$R^2 = 1 - \frac{RSS}{TSS}, \qquad \text{Eq. 6.3}$$

where RSS is the sum of squares of residuals and TSS is the total sum of squares.

$$RMSE = \sqrt{\frac{\sum_N^{i=1}(x_i - \hat{x})}{N}}, \qquad \text{Eq. 6.4}$$

$$\text{n}RMSE = \frac{RMSE}{x_{max}}, \qquad \text{Eq. 6.5}$$

where x_{max} is the maximum of the observations, i variable i, N is the number of non-missing data points, x_i are actual observations in the time series, and $\hat{x}$ is the estimated time series.

6.3 Results

6.3.1 Model Performance for ALIVE$_{ET}$ Estimates

Table 6.3 and Figure 6.2 illustrate the performance of LSTM and GBR models in estimating half-hourly (hh) ALIVE$_{ET}$ compared to EC-derived ET under daytime and nighttime conditions. During the daytime (Figure 6.2a), the GBR model outperformed LSTM when using single-time features, achieving the highest R^2 (0.74) and the lowest nRMSE (0.040). Incorporating time series features improved the LSTM model's performance (R^2 of 0.72 and nRMSE of 0.046), making it comparable to the GBR model (R^2 of 0.71 and nRMSE of 0.051). While the LSTM model benefits significantly from GPU acceleration—leveraging the A100 GPU (40GB/80GB VRAM, Tensor Cores, High Throughput) in Google Colab Pro (32GB RAM) to reduce training and inference time—it remains computationally expensive. The GBR model, on the other hand, primarily runs on the CPU and does not gain performance boosts from GPU acceleration. Even with GPU acceleration, the LSTM model still requires 5.3 times more computation time than GBR and fails to surpass GBR in modeling accuracy for daytime ET estimation. At night (Figure 6.2b), GBR model performance was notably lower. The LSTM model, trained on time series features, performed slightly better (R^2 of 0.24 and nRMSE of 0.008) than GBR (R^2 of 0.21 and nRMSE of 0.011).

Table 6.3. Performance metrics (R^2, nRMSE, and prediction time (P.T.)) of LSTM and GBR models for half-hourly (hh) ALIVE_{ET} vs. EC-derived ET for daytime and nighttime estimates.

		LSTM			GBR		
	Feature set	R^2	nRMSE	P.T. (sec)	R^2	nRMSE	P.T. (sec)
Single time features	Daytime	0.69	0.062	3.5 (3.8T)	**0.74**	**0.040**	0.9 (T)
	Nighttime	0.12	0.020	2.9 (3.2T)	0.11	0.019	0.8 (0.9T)
Time series features	Daytime	0.72	0.046	4.8 (5.3T)	0.71	0.051	1.2 (1.3T)
	Nighttime	**0.24**	**0.008**	4.1 (4.5T)	0.21	0.011	1.2 (1.3T)

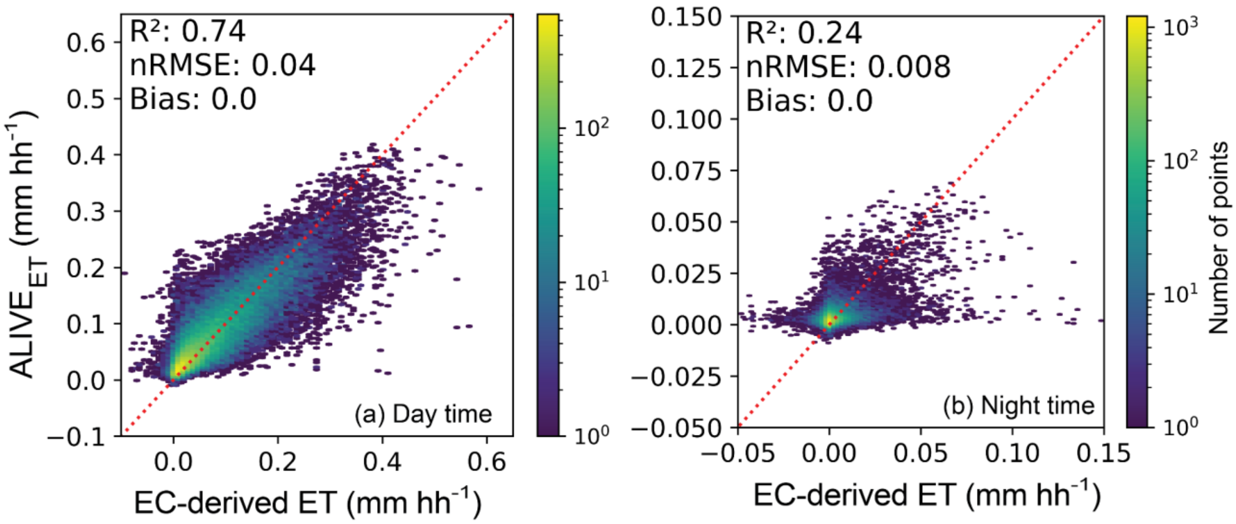

Figure 6.2. Density scatter plots of EC-derived ET (mm hh^{-1}) (a) vs. daytime ALIVE$_{ET}$ estimated from GBR trained on single time observations and (b) nighttime ALIVE$_{ET}$ estimated from LSTM trained on time series observations.

6.3.2 Model performance in representing diurnal dynamics and seasonal variability

Figure 6.3 demonstrates the ALIVE$_{ET}$ model's capability to capture diurnal variations in ET by comparing its estimates with EC-derived ET across different local hours for both daytime and nighttime conditions. Values are averaged across all sites for each local hour. During the daytime, the model exhibits strong agreement with observed ET. ET values rise in the morning, peak around midday, and gradually decline in the afternoon, closely mirroring the EC-derived ET trend. However, during nighttime, ET values drop significantly due to the absence of solar radiation, and while the model captures some fluctuations, biases from EC-derived ET are more noticeable.

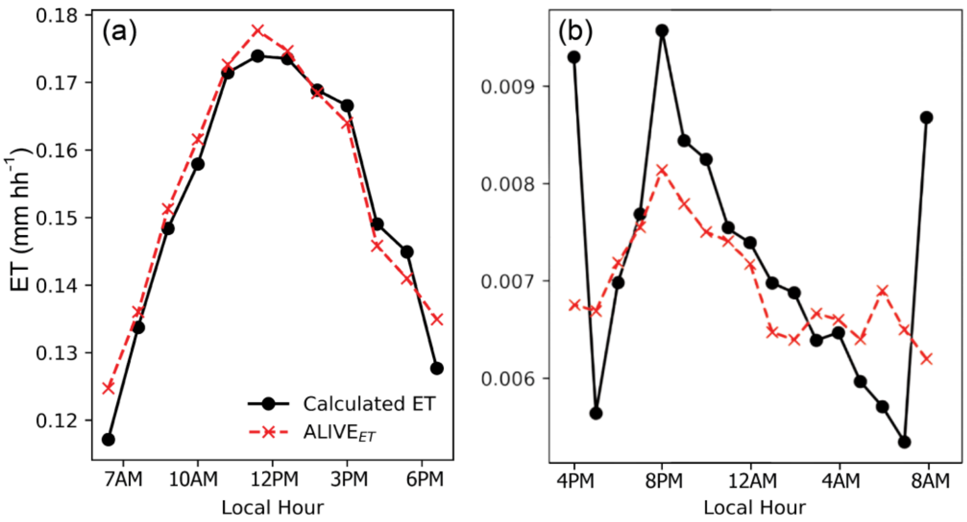

Figure 6.3. Mean values of EC-derived ET (mm hh^{-1}) (a) vs. daytime ALIVE$_{ET}$ estimated from GBR trained on single time observations and (b) nighttime ALIVE$_{ET}$ estimated from LSTM trained on time series observations. Values are averaged across all sites for each local hour.

Figure 6.4 evaluates the model's performance across different months. The R^2 follows a distinct seasonal pattern, with lower values during late winter and early spring, and improved performance from May to September. Additionally, the seasonal trend in LST, a key predictor variable, aligns with model performance, as warmer months exhibit more accurate ET estimates while colder months introduce greater uncertainty. Figure 6.4c shows the one year time series of daily ALIVE$_{ET}$ estimates against calculated ET from EC measurements, which shows the model's robustness, with an annual R^2 of 0.99 and nRMSE of 0.028.

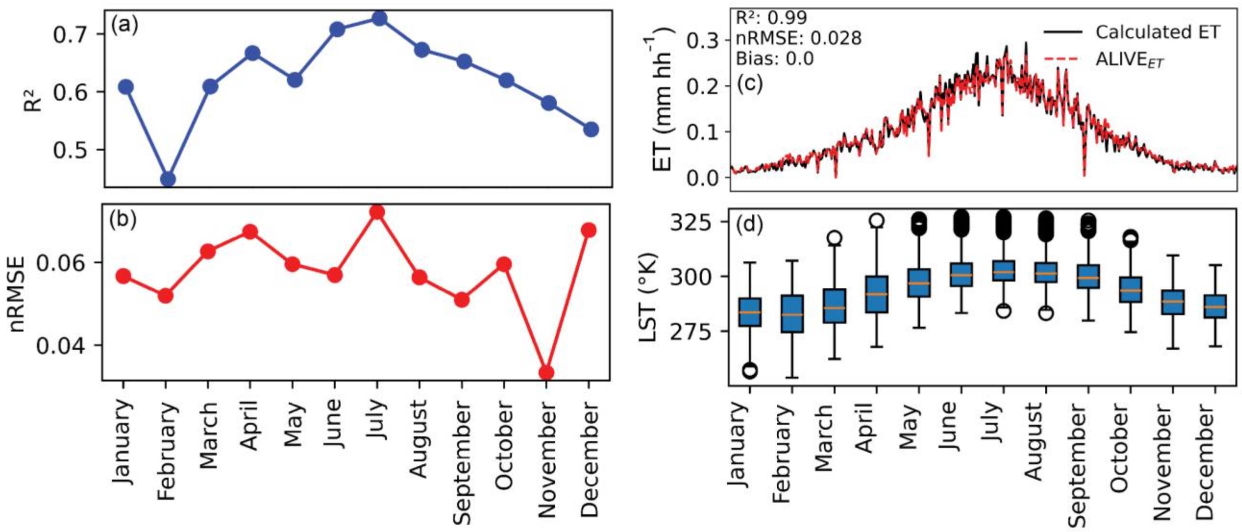

Figure 6.4. Model performance across different months: (a) R^2 values, (b) nRMSE values, (c) model fit for mean daily ET (mm hh^{-1}) over the entire year of 2023, and (d) LST distribution showing monthly variability with boxplots.

6.3.3 Feature importance

For daytime ET modeling (Figure 6.5c), the most influential feature is NIRvP, followed by ALIVE$_{LST}$ and sNIRvP. Other important predictors include ALIVE$_{DSR}$, CMI_C03 (top-of-atmosphere NIR at 0.86 μm), and BRF5 (surface SWIR reflectance at 1.6 μm). SHAP values indicate that higher NIRvP and ALIVE$_{LST}$ values positively affect ET predictions, underscoring the importance of thermal and vegetation-related features. Lower-ranked features, such as CMI_C04 and BRF3, have minimal impact on prediction, a point that is further discussed in the discussion section.

For nighttime ET modeling (Figure 6.5a), a different set of predictors emerges, with ALIVE$_{LST}$, sNIRv_daily_mean, and NIRv_daily_mean standing out as the most dominant variables. Unlike the daytime model, the absence of the surface reflectance product and DSR values near or equal to zero shifts the focus toward top-of-atmosphere SWIR and thermal observations. The lower SHAP values associated with specific CMI variables, such as CMI_C09 and CMI_C08, indicate that these spectral bands have less influence on nighttime ET estimation compared to other variables. Additionally, the time series analysis in Figure 6.5b emphasizes the significance of remote sensing observations taken 2 and 4 hours before the ET prediction. In contrast, observations from 12 to 18 hours prior to ET have the least influence, while those from 22 to 24 hours also gain some importance.

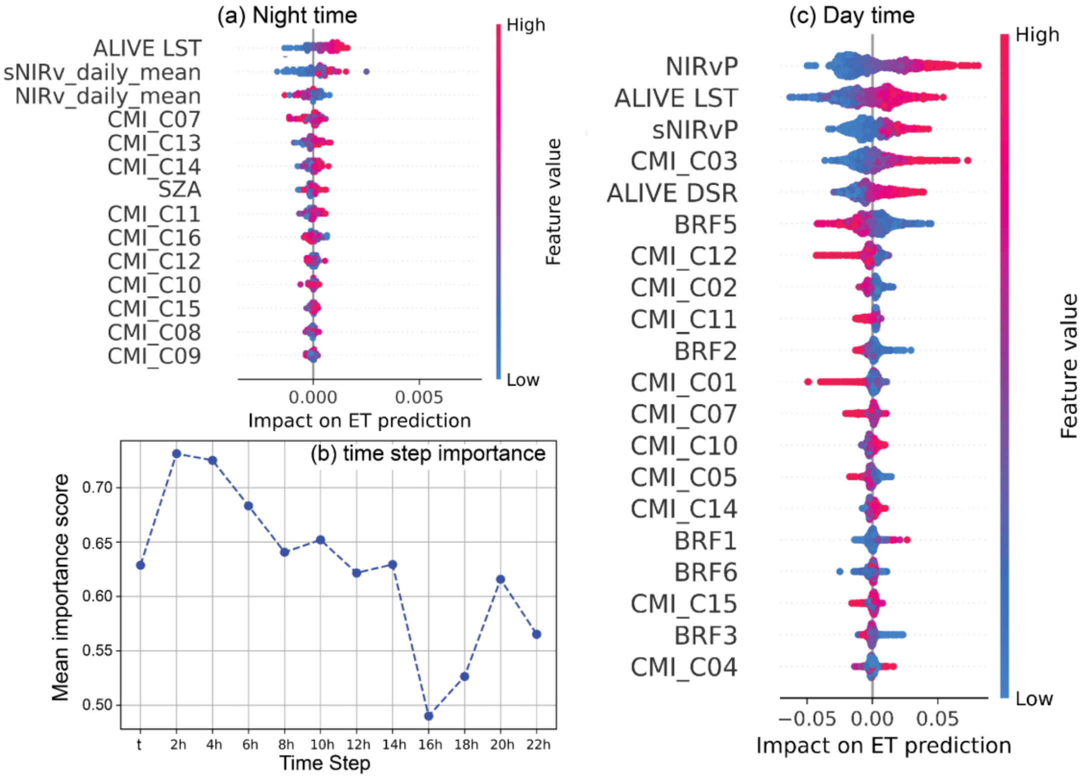

Figure 6.5. Feature importance analysis using SHAP values for (a) nighttime, (b) nighttime time-step importance, and (c) daytime. Panels (a) and (c) rank the top features by importance for nighttime and daytime, respectively, while (b) highlights time-step importance in the nighttime model, where the LSTM with time series features performed best.

6.3.4 Model Performance Across different Köppen climate classes and IGBP vegetation covers

We evaluated the model's performance across different Köppen climate classifications (Figure 6.6) and IGBP land cover types (Figure 6.7) by comparing ALIVE_{ET} estimates with EC-derived ET. We averaged the site-level estimates within each climate classification or land cover type. For this analysis, we focused on the five most prevalent climate types that cover the largest land areas over CONUS. The Figure 6.6 left panels display density scatter plots comparing half-hourly ALIVE_{ET} estimates to EC-derived ET, while the right panels illustrate daily time series for the years 2022 and 2023.

Across different climate zones, ALIVE_{ET} demonstrates varying levels of agreement with EC-derived ET. In Mediterranean climates (Csa/Csb), the model achieves the highest agreement, with R^2 values of 0.78 at the half-hourly scale and 0.91 at the daily scale. Similarly, in humid subtropical (Cfa) and humid continental (Dfb/Dfa) climates, ALIVE_{ET} performs well, with R^2 values of 0.73 and 0.89 (Cfa) and 0.73 and 0.86 (Dfb/Dfa) at half-hourly and daily scales, respectively. The normalized RMSE (nRMSE) remains low across these regions, ranging from 0.034 to 0.05, indicating strong accuracy in capturing ET variability. In drier climates, performance slightly declines. In semi-arid steppe (Bsk) regions, ALIVE_{ET} maintains reasonable accuracy, with R^2 values of 0.64 (half-hourly) and 0.81 (daily), while the nRMSE increases slightly to 0.047 and 0.03. The lowest agreement occurs in arid desert climates (Bwk), where R^2 values drop to 0.57 (half-hourly) and 0.77 (daily), and the nRMSE reaches its highest values (0.093 and 0.046).

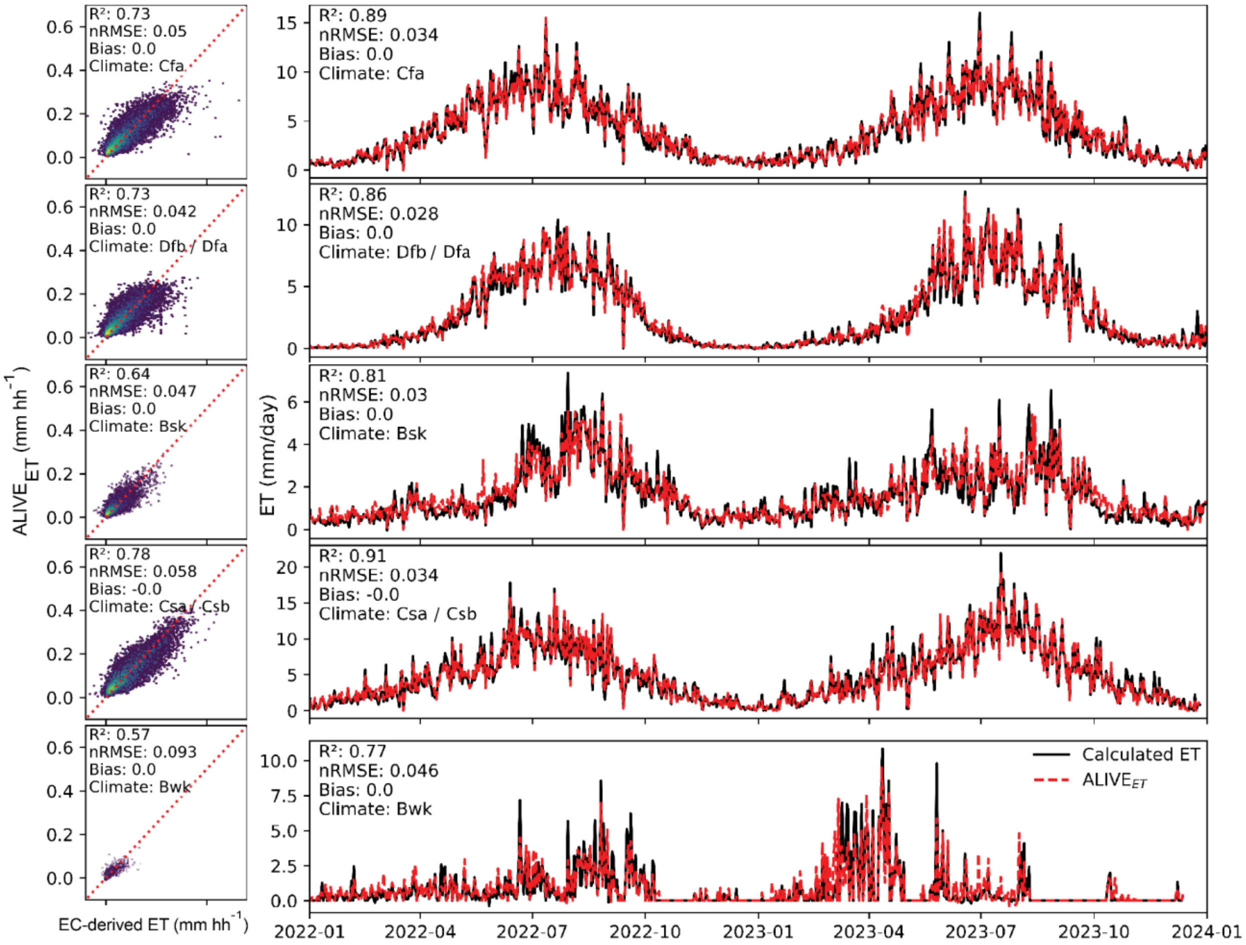

Figure 6.6. ALIVE$_{\mathrm{ET}}$ estimates compared to EC-derived ET across different Köppen climate classes (see Table 6.A2 in the appendix for climate class full name). Left panels display density scatter plots of half-hourly estimates, while right panels show daily time series of ALIVE$_{\mathrm{ET}}$ vs. EC-derived ET for 2022 and 2023.

Figure 6.7 presents a comparative evaluation of ALIVE$_{\mathrm{ET}}$ estimates against EC-derived ET across different land cover types classified according to the IGBP. Similar to Figure 6.6, the left panels illustrate density scatter plots of half-hourly ET estimates, while the right panels depict daily time series comparisons over two years (2022–2023). The performance varies across land cover types, with wetlands (WET) exhibiting the highest R^2 value of 0.81 and the lowest nRMSE (0.055), suggesting a strong correlation and minimal error. Grasslands (GRA) and croplands (CRO) also show relatively high agreement, with R^2 values of 0.77 and 0.72, respectively. However, evergreen broadleaf forests (EBF) and savannas (SAV) exhibit lower R^2 values of 0.44

and 0.40, respectively, suggesting a reduced predictive capability in these ecosystems, which are not well-represented in the tower dataset (Figure 6.1). The time series plots on the right further highlight the temporal consistency between ALIVE$_{ET}$ (dashed red) and EC-derived ET (black). For WET, GRA, and CRO, ALIVE$_{ET}$ effectively captures seasonal variations, with peak ET values aligning well with those observed from EC measurements. In contrast, EBF and SAV display greater errors, with ALIVE$_{ET}$ tending to underestimate peak ET fluxes and exhibit higher variability.

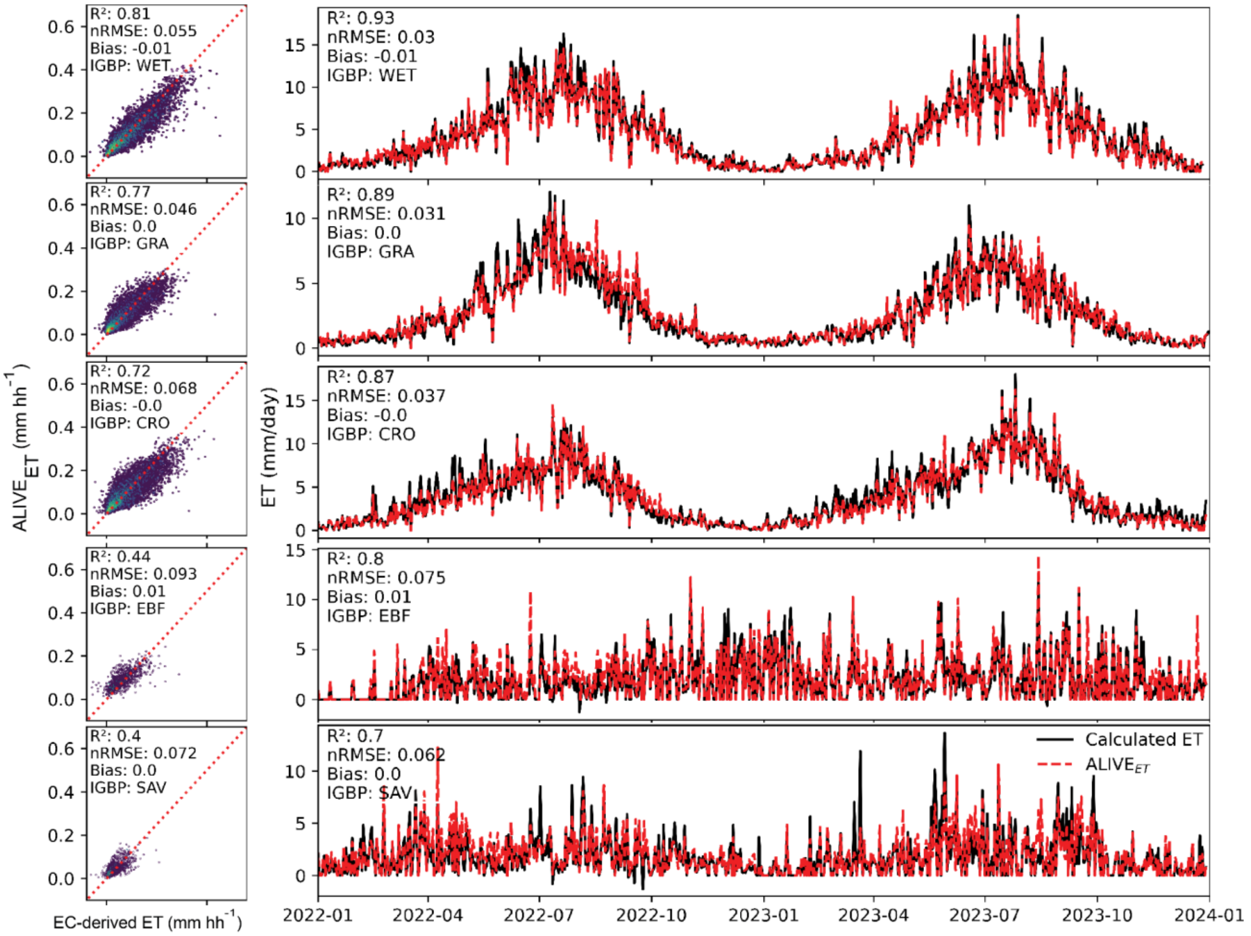

Figure 6.7. Comparison of ALIVE$_{ET}$ estimates with EC-derived ET across different IGBP land cover types (see Table 6.A2 in the appendix for full name). Left panels show density scatter plots of half-hourly estimates, while right panels display daily time series of ALIVE$_{ET}$ vs. EC-derived ET for 2022 and 2023.

6.3.5 Comparing ALIVE$_{ET}$ against ALEXI

To assess the performance of the ALIVE$_{ET}$ model spatially, we compared its daily aggregated ET estimates against those derived from the physically-based ALEXI model. We selected four days (DOY 117, 149, 179, and 218) from mid-May to mid-August 2022, ensuring a minimum 30-day interval between them for a spatial and statistical comparison. These days were chosen based on minimal cloud cover to ensure a fair comparison between the ET estimations. Since ALEXI estimates ET under cloudy conditions using post-processing techniques, selecting clear-sky days allows for a more direct evaluation of the models without the influence of these adjustments. The first and second columns in Figure 6.8 illustrate the spatial distributions of ALEXI and ALIVE$_{ET}$, respectively, while the third column depicts their differences (ALIVE$_{ET}$ – ALEXI) alongside density scatter plots and kernel density estimation (KDE) histograms.

ALIVE$_{ET}$ exhibits strong agreement with ALEXI across different regions and time periods, capturing similar spatial patterns of ET variability. However, noticeable differences emerge in certain areas, particularly in the central and eastern United States, where ALIVE$_{ET}$ estimates are generally lower than ALEXI (Figure 6.8, Column 3). These differences are more pronounced on DOY 149 and 179, where negative biases dominate, suggesting potential underestimation of ALIVE$_{ET}$ relative to ALEXI in regions with relatively high vegetation density.

The density scatter plots indicate a generally strong correlation between ALIVE$_{ET}$ and ALEXI, with R^2 ranging from 0.62 to 0.74. RMSE values vary between 0.77 mm/day and 1.17 mm/day, with biases fluctuating between -0.16 mm/day and -0.67 mm/day (about 2.8% to 11.9% relative to the median ALEXI). These values highlight a consistent but slightly underestimated ET prediction by ALIVE$_{ET}$ that we discuss further in the Discussion section. The KDE histograms further reveal that ALIVE$_{ET}$ and ALEXI share similar ET distributions, though ALIVE$_{ET}$ exhibits a higher density of lower ET values, which may be attributed to differences in model parameterizations or the input data utilized by the machine learning framework.

Despite these differences, the results indicate that ALIVE$_{ET}$ successfully captures the overall spatial and temporal trends of ALEXI. The observed discrepancies warrant further investigation, particularly in refining the model's ability to represent high-ET regions more accurately. Future work will focus on incorporating additional constraints to mitigate biases and enhance the robustness of ALIVE$_{ET}$ across diverse landscapes and climatic conditions.

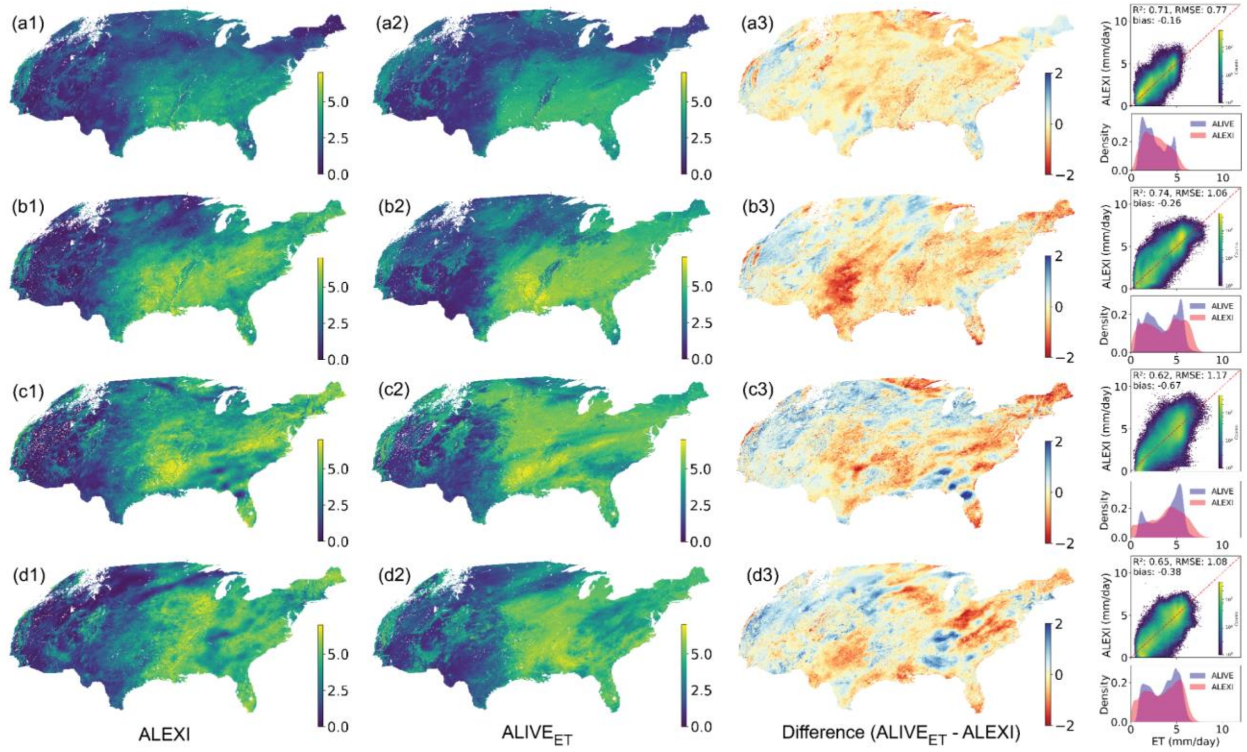

Figure 6.8. Comparison of daily ET estimates from ALEXI and ALIVE. Column 1 shows ALEXI$_{ET}$ maps, Column 2 presents ALIVE$_{ET}$ maps, and Column 3 displays their differences in units of mm/day, along with density scatter plots and KDE histogram distributions. Days of the year (DOY) with the lowest cloud cover, selected from mid-May to mid-August in 2022, include (a) 117, (b) 149, (c) 179, and (d) 218.

6.4 Discussion

6.4.1 Model Performance for ALIVE$_{ET}$ Estimates

The comparative analysis of GBR and LSTM models in estimating ALIVE$_{ET}$ reveals a clear trade-off between predictive accuracy, computational efficiency, and adaptability to near real-time half hourly estimation. GBR outperforms LSTM during daytime, likely due to its ability to model nonlinear interactions without requiring sequential dependencies (Bentéjac et al., 2021; Dhake et al., 2023; Sahin, 2020). GBR efficiently captures the relationship between solar radiation, temperature, and vegetation indices, providing accurate estimates of daytime ET with minimal computational cost. LSTM, despite its capability to model temporal dependencies, fails to improve predictions. The computational burden of LSTM further limits its practicality for large-scale, near real-time applications (Ranjbar et al., 2024d).

At night, both models struggle, reflecting the inherent challenges of predicting nocturnal ET that is most strongly related to environmental drivers that are difficult to discern from space, namely near-surface wind speed and vapor pressure deficit as well as air temperature (Fisher et al., 2007; Novick et al., 2009). The absence of direct solar radiation reduces variability, making it difficult for data-driven models to learn meaningful patterns. LSTM exhibits a slight advantage over GBR, possibly due to its ability to capture residual heat flux and time since CMI observations that are consistent with precipitation events through the sequential dependencies in time series data (Dhake et al., 2023; Ghimire et al., 2022). However, this improvement remains marginal, indicating that neither model fully accounts for the complex boundary layer processes governing nighttime ET. Incorporating other data sources for meteorological and soil parameters, such as humidity gradients and soil heat flux, could improve performance (Katul et al., 2012; Walker et al., 2019; Wanniarachchi and Sarukkalige, 2022).

These results underscore the importance of aligning model choice with application constraints. GBR is more suitable for real-time ET monitoring due to its efficiency, while LSTM, despite its potential for sequential modeling, remains limited by high computational costs (Ranjbar et al., 2024d). A promising direction is the development of hybrid models that leverage GBR's efficiency for daytime ET and LSTM's temporal learning for nocturnal predictions. Additionally, integrating physical constraints into machine learning models could enhance reliability, reducing dependence on purely empirical learning. Future work should explore attention-based architectures as a computationally efficient alternative to LSTM for capturing temporal dependencies in nocturnal ET estimation and seek to incorporate meteorological datasets, like the High Resolution Rapid Refresh (HRRR) (Dowell et al., 2022; James et al., 2022), to better simulate the processes that control ET.

6.4.2 Model performance in representing diurnal dynamics and seasonal variability

The ALIVE$_{ET}$ model effectively captures diurnal ET dynamics, closely mirroring the expected daytime trend—rising in the morning, peaking at midday, and declining in the afternoon. This alignment with observed patterns is largely due to the model's empirical, data-driven nature, which leverages statistical relationships between environmental variables and ET (Angelov and Gu, 2019; Belitz and Stackelberg, 2021; Brunton and Kutz, 2022). However, its reliance on mean values across sites limits its adaptability to local-scale variations, potentially leading to oversimplifications in heterogeneous landscapes (Brunton and Kutz, 2022).

At night, biases become more pronounced, with ET values dropping sharply in the absence of solar radiation. While the model captures some nocturnal fluctuations, it struggles to estimate early morning and late afternoon transitions accurately. This challenge likely stems from reduced variability in key input features (Angelov and Gu, 2019), such as LST and radiation, which dominate ET estimation during daylight but provide weaker predictive signals at night. Addressing this limitation requires integrating additional environmental drivers—such as humidity, soil moisture, and boundary layer meteorology—to better characterize latent heat exchange under low-energy conditions (Katul et al., 2012; Talib et al., 2021).

Seasonally, the model performs better during warmer months (May to September) and exhibits reduced accuracy in colder months. This trend reflects the direct relationship between LST and ET, where higher temperatures correspond to increased evaporative demand and more stable model predictions due to higher ET variability. In contrast, winter months introduce uncertainty due to lower LST values, as well as reductions in ET activity (Chen and Liu, 2020; Ling et al., 2022). The model's robustness at the annual scale, indicated by high R^2 for daily ET estimates, suggests it effectively generalizes long-term ET trends despite seasonal variations.

6.4.3 Feature importance

During daytime, vegetation-related and thermal features—particularly NIRvP and $\text{ALIVE}_{\text{LST}}$—emerge as the most influential predictors, reinforcing the strong dependence of ET on plant activity and temperature (Chen and Liu, 2020; Sun et al., 2012). The high SHAP values for NIRvP suggest that increased photosynthetic activity, as captured by red absorption and NIR reflectance, leads to higher ET rates (C. Zheng et al., 2022). Other relevant predictors, such as $\text{ALIVE}_{\text{DSR}}$ and CMI_C03, contribute by incorporating solar radiation and top-of-atmosphere near-infrared signals. BRF3 is utilized in the computation of NIRvP and sNIRvP, both of which are key predictors in the model. Therefore, BRF3 is assigned a lower importance ranking, which is a characteristic outcome of machine learning models when handling highly correlated input features. In such cases, the model prioritizes the most informative variable while diminishing the influence of redundant features to optimize predictive performance (Angelov and Gu, 2019; Brunton and Kutz, 2022).

In contrast, the best nighttime ET estimation was achieved from time series features using LSTM modeling. The model was primarily driven by surface temperature and the mean value of vegetation indices (NIRv and sNIRv) from prior day rather than instantaneous radiation inputs. $\text{ALIVE}_{\text{LST}}$ remains the dominant predictor, aligning with the idea that nighttime ET is primarily governed by residual surface heat and micrometorological variables like wind, temperature, and atmospheric dryness rather than direct solar-driven processes (McColl, 2020; Wanniarachchi and Sarukkalige, 2022). The significance of daily-averaged vegetation indices (sNIRv_daily_mean and NIRv_daily_mean) indicates that plant water loss continues into the night, albeit at much lower magnitudes, influenced by prior daytime conditions. Unlike daytime modeling, DSR is absent as a predictor, leading to increased reliance on CMI observations.

Furthermore, the time-step analysis for nighttime modeling highlights the temporal dependencies within the LSTM framework. Observations from 2 to 4 hours before the prediction time contribute the most, emphasizing the short-term persistence of environmental signals in driving nighttime ET. Conversely, observations from 12 to 18 hours prior show minimal influence, likely due to the decoupling of past daytime energy inputs. Interestingly, inputs from 22 to 24 hours earlier gain some relevance, potentially capturing the effects of previous nighttime cooling trends or delayed surface moisture responses (Labedzki, 2011) but also consistent with correlation amongst nighttime conditions. These findings highlight the need for tailored ET models, with nighttime models relying on time-series features, particularly LST and indices, while daytime models benefit from instantaneous reflectance, solar inputs, and LST. Future improvements could involve adding new features, particularly for nighttime ET, where additional environmental variables might enhance predictive accuracy (Fisher et al., 2007; Katul et al., 2012; Novick et al., 2009; Walker et al., 2019).

6.4.4 Model Performance Across different Köppen climate classes and IGBP vegetation covers

The varying performance of ALIVE$_{\text{ET}}$ across climate classifications and land cover types highlights the inherent challenges in modeling ET in diverse environmental conditions. The strong performance in Mediterranean (Csa/Csb) and humid subtropical (Cfa) climates suggests that the model effectively captures ET dynamics in regions with consistent seasonal cycles and relatively high vegetation cover (Feng et al., 2019; Zhou et al., 2021). This is likely because these climates exhibit strong and stable relationships between radiation, temperature, and vegetation indices, which are key drivers of ET (Feng et al., 2019; Labedzki, 2011). Although, Mediterranean climates present additional complexity due to pronounced shifts between energy-limited and water-limited periods within the same year, where ET dynamics are strongly modulated by seasonal soil moisture availability and growing season length (Ryu et al., 2008), the strong seasonal greenness signal and surface temperature variability in these regions may enhance the ability of machine learning models to capture ET dynamics.

However, the accuracy of ALIVE$_{\text{ET}}$ declines in more arid regions such as semi-arid steppe (Bsk) and arid desert (Bwk) climates. The higher errors in these environments can be attributed to the increased influence of soil evaporation, sparse vegetation cover and its subgrid variability, the pronounced spatial heterogeneity of these ecosystems, and the episodic nature of precipitation, all of which introduce greater variability in ET rates (Krishnan et al., 2012; Tabari et al., 2012). The model's reliance on vegetation indices and thermal features may lead to oversimplifications in these water-limited regions, where plant water use efficiency and soil moisture interactions play crucial roles in governing ET.

Another critical factor affecting model performance in drier climates is the role of advection and non-local energy sources. ET is not solely controlled by local surface conditions but is also influenced by large-scale atmospheric dynamics and horizontal energy transport, factors that are difficult to capture using data-driven models based on local predictors (Krishnan et al., 2012; Pan et al., 2020; Weiß and Menzel, 2008). Furthermore, the increased nRMSE in these regions suggests that the model may struggle with capturing sub-daily variability, particularly in conditions where rapid changes in atmospheric demand or soil moisture availability occur. Future enhancements may involve integrating soil moisture retrievals, atmospheric boundary layer dynamics, or explicitly accounting for advection by incorporating upstream pixels in the model training process to improve performance in these challenging regions.

The land cover analysis reveals similar challenges. While the model performs well in wetlands (WET), grasslands (GRA), and croplands (CRO), it struggles in evergreen broadleaf forests (EBF) and savannas (SAV). The strong performance in wetlands can be attributed to the dominance of surface water evaporation, which is well captured by thermal and vegetation indices (Bao et al., 2021; Drexler et al., 2004; Fleischmann et al., 2023). In contrast, the lower R^2 values in EBF and SAV suggest that canopy structure, water use strategies, and plant functional diversity introduce complexity that is not adequately represented by the current feature set (Y. Liu et al., 2022; Pan et al., 2024), and there are currently few measurements in the CONUS scene with which to train models. Dense canopies in EBF likely lead to discrepancies between surface temperature and actual transpiration rates, as the thermal signal captured by satellites may not directly reflect the evaporative demand within the canopy (Pan et al., 2024). Additionally, in savannas, the coexistence of vegetation and grass – with differing phenology – and bare soil likely leads to different ET responses, further complicating estimation accuracy. Moreover, the underestimation of peak ET fluxes in EBF and SAV suggests that ALIVE$_{\text{ET}}$ may not fully account for plant physiological responses. Trees in these ecosystems often exhibit deep rooting systems that allow them to access groundwater, enabling sustained transpiration even when surface moisture is low (Miller et al., 2010).

The findings emphasize the need for climate- and ecosystem-specific modeling approaches. While ALIVE$_{\text{ET}}$ effectively captures ET dynamics in regions with predictable climate and land cover characteristics, its limitations in arid and heterogeneous landscapes underscore the importance of incorporating additional process-based constraints. Future advancements should explore hybrid modeling frameworks that combine machine learning with physically-based constraints, ensuring that key hydrological and physiological processes are better represented across diverse environmental conditions.

6.4.5 Comparing ALIVE$_{\text{ET}}$ against physically-based ALEXI$_{\text{ET}}$

While ALIVE$_{\text{ET}}$ demonstrates a strong agreement with ALEXI$_{\text{ET}}$ in capturing broad spatial and temporal trends, systematic biases and discrepancies highlight areas requiring further refinement and investigation. A key observation is the underestimation of ALIVE$_{\text{ET}}$ relative to ALEXI$_{\text{ET}}$, particularly in regions with high vegetation density and complex moisture dynamics. This tendency is most pronounced on the DOY 149 and 179 comparisons, suggesting that ALIVE$_{\text{ET}}$ may struggle to fully capture the ET dynamics in peak growing seasons when evapotranspiration rates are at their highest. The KDE histograms show that while ALIVE$_{\text{ET}}$ successfully captures the overall shape of the ET distribution, it may not fully replicate the higher-end variability observed in ALEXI$_{\text{ET}}$. Previous studies have noted that machine learning models, despite their adaptability, often exhibit difficulties in accurately representing extreme values due to the reliance on training data distributions (Amani and Shafizadeh-Moghadam, 2023; Maxwell et al., 2018; Ranjbar et al., 2021; Thapa et al., 2023). The underestimation at higher ET values could stem from insufficient representation of extreme conditions in the training dataset or limitations in the predictor variables used by ALIVE_{ET}. ALIVE_{ET} also relies on eddy covariance measurements which on average do not close the surface-atmosphere energy balance such that a small underestimation might be expected (Mauder et al., 2024; Stoy et al., 2013; Wilson et al., 2002). Despite these limitations, ALIVE_{ET}'s ability to closely track ALEXI trends across multiple time periods and spatial regions demonstrates its potential as an empirical alternative for large-scale ET estimation for model benchmarking, or to further improve gapfilling of missing observations. The observed biases, while systematic, are relatively modest and may be addressable through targeted model improvements including additional filters for eddy covariance energy balance closure. Future work should explore additional input features, such as soil moisture retrievals and vegetation water content, to enhance model performance in high-ET regions (Katul et al., 2012; Walker et al., 2019).

Another possible explanation for the observed differences is the different parameterization strategies employed by the two models. ALEXI_{ET}, a physically based model, explicitly accounts for surface energy balance constraints and vegetation stress conditions (Anderson et al., 2012). In contrast, ALIVE_{ET}, driven by machine learning algorithms, relies on statistical relationships inferred from current and historical data, potentially leading to discrepancies in regions where these relationships deviate from physical constraints. For instance, previous research has discussed that machine learning models tend to generalize well under typical conditions but may struggle with spatial heterogeneity, especially in highly dynamic landscapes such as forested or irrigated agricultural areas (Amani and Shafizadeh-Moghadam, 2023; Talib et al., 2021). The inclusion of physics-informed constraints within the ALIVE_{ET} framework could help mitigate these discrepancies and improve the model's generalization capabilities.

The R^2 ranging from 0.62 to 0.74 indicate a moderately strong relationship between the two models; however, the variations in RMSE (0.77 mm/day to 1.17 mm/day) and biases (-0.16 mm/day to -0.67 mm/day) suggest that performance inconsistencies exist across different climatic and land cover conditions. These findings align with previous evaluations of machine learning and biophysical models for ET estimation, which have discussed that the model performance degrades in areas with complex hydrological and biophysical interactions (Bellocchi et al., 2010; García et al., 2013; Oliveira et al., 2024; Verhoef et al., 2018). The systematic negative bias observed in ALIVE_{ET} could indicate a need for model recalibration, potentially through region-specific tuning or transfer learning techniques that leverage site-specific data to adjust model parameters (Amani and Shafizadeh-Moghadam, 2023; Oliveira et al., 2024) or additional consideration of energy balance closure as noted. Future research should focus on incorporating physical constraints such as energy balance closure and mass conservation, improving representation of extreme conditions in training data, and developing hybrid models that seamlessly integrate machine learning with process-based simulations to enhance generalizability and physical consistency (Ranjbar et al., 2024d).

6.5 Conclusion

In this study, we demonstrate the potential of machine learning models, particularly GBR and LSTM networks, for estimating high-frequency ET under various sky conditions, including nighttime, using geostationary satellite data. Our findings reveal that GBR outperforms LSTM during the daytime, offering greater efficiency and accuracy for real-time ET estimation due to its lower computational cost. However, during nighttime, both models encounter challenges, as the lack of solar radiation and reduced variability in input features limit predictive performance. LSTM slightly outperforms GBR in these conditions, capturing residual heat flux and subtle atmospheric dynamics using the previous 24 hours of time series features. Despite this, the improvements are marginal, highlighting the complexity of nocturnal ET processes. The results emphasize the importance of adapting model choice to the specific application context. While GBR is more suitable for operational, real-time ET monitoring, LSTM's ability to capture temporal dependencies could enhance future efforts to improve nighttime ET predictions. Furthermore, the integration of additional environmental variables, such as humidity, soil moisture, and boundary layer parameters, could strengthen model performance, particularly during nighttime hours. Hybrid models combining the strengths of GBR and LSTM, and fusion approaches with physically-based models, could further enhance both daytime and nighttime ET estimates.

Chapter 7. Conclusion and Future Directions

7.1 Summary and Key Insights

This dissertation has explored the application of geostationary satellite data to address environmental challenges for high-temporal monitoring. By integrating high-resolution temporal data with machine learning algorithms, this work advances our ability to monitor and model key Earth system variables. We have demonstrated that geostationary remote sensing provides invaluable, near real-time insights into ecosystem dynamics (Chapter 1). The research highlights the power of geostationary satellite monitoring while emphasizing the need for continuous refinement in our models and methodologies to overcome challenges in sub-daily observations.

Our studies have particularly focused on solar radiation (Chapter 2), land surface temperature (Chapter 3), carbon fluxes (Chapter 4, Chapter 5), and evapotranspiration (Chapter 6). This dissertation showcases how machine learning and high-frequency geostationary satellite data can overcome common remote sensing limitations, such as cloud cover and data gaps. The high-frequency mapping of shortwave radiation using GOES-R (Chapter 2) represents an advancement, offering a near real time observation of atmospheric energy flows. This capability has practical implications for applications like energy forecasting, diurnal plant productivity, and water cycle monitoring. Similarly, the estimation of high-frequency land surface temperature under various sky conditions (Chapter 3) is critical for understanding diurnal and seasonal temperature variations that influence everything from crop productivity and water cycle to climate modeling.

Another major insight from this research is the ability to use geostationary satellites to estimate diurnal carbon fluxes at fine temporal scales (Chapter 4, Chapter 5). In the context of ongoing climate change, this enables near real-time tracking of the carbon cycle, providing crucial updates on terrestrial ecosystem health. High-frequency satellite observations help monitor carbon cycle and offer early warnings for changes in ecosystem productivity or environmental stressors such as wildfire, early frost, or flash drought. Likewise, our approach to near real-time evapotranspiration estimation (Chapter 6) provides new opportunities for monitoring water use, with significant implications for agricultural management and water resource planning.

Despite these findings, challenges remain for using geostationary satellites for terrestrial ecosystem monitoring. While improvements in data quality and model accuracy have been achieved, geostationary-based measurements are still affected by cloud interference, atmospheric distortions, and other sources of uncertainty such as parallax effects. This dissertation demonstrates the advances that can be achieved by leveraging the high-frequency observations of geostationary satellites, but these limitations must be acknowledged and addressed in future research. By advancing gap-filling techniques and incorporating novel machine learning approaches, we can continue refining our ability to make real-time, reliable environmental assessments.

7.2 Future Directions

Looking ahead, geostationary satellite data has the potential to transform how we monitor and model the environment. But there are still important challenges to address. One of the biggest is improving the machine learning models that process this data. This dissertation has shown that existing models can estimate environmental variables well, but handling uncertainty remains a challenge. Future research needs to focus on better ways to quantify uncertainty, making predictions more reliable—especially for critical applications like disaster response and climate adaptation. Understanding how confident we are in model predictions will be key for making fast, informed decisions, whether it’s helping farmers manage crops during a drought or responding to extreme heatwaves.

Moreover, as geostationary satellites evolve, they will offer higher spatial resolution in addition to their already impressive temporal frequency. Integrating such data with high-resolution atmospheric models will allow for an even more detailed examination of Earth system processes. However, this also introduces challenges related to processing vast amounts of data and storage. Advances in computing power, alongside innovations in deep learning, will be essential for handling large-scale datasets. Leveraging deep learning techniques such as CNNs and RNNs could unlock further insights by identifying complex patterns in satellite images and time-series data.

Beyond refining models, combining data from multiple satellite sources is an exciting direction to explore. By merging geostationary data with high-resolution observations from polar-orbiting satellites, we could create hybrid models that take advantage of the high frequency observations from geostationary platforms and the detailed spatial data from polar-orbiting systems. Adding ground-based observations and sensor networks into this mix would further boost the accuracy and reliability of Earth system models. By bringing together data from multiple sources, we can build more complete models that enhance our understanding of key global processes, like local water cycles and carbon fluxes, leading to a more holistic approach to Earth system science.

Looking further into the future, the development of self-sustaining, autonomous systems for environmental monitoring could become a reality. Self-reinforcing machine learning models that continuously update based on incoming data from geostationary satellites could enable real-time monitoring, supporting autonomous decision-making systems for applications ranging from disaster response to ecosystem management. While these advancements are promising, challenges related to data integration, model scalability, and computational efficiency remain. Future research should prioritize the development of efficient, automated data pipelines that process satellite and ground-based measurements in real time. By addressing these challenges and embracing the latest advances in machine learning and data science, we can continue to refine geostationary remote sensing capabilities, pushing the boundaries of Earth system science and environmental monitoring.

References

AghaKouchak, A., Farahmand, A., Melton, F.S., Teixeira, J., Anderson, M.C., Wardlow, B.D., Hain, C.R., 2015. Remote sensing of drought: Progress, challenges and opportunities. Rev. Geophys. 53, 452–480.

Ai, J., Jia, G., Epstein, H.E., Wang, H., Zhang, A., Hu, Y., 2018. MODIS-based estimates of global terrestrial ecosystem respiration. J. Geophys. Res. Biogeosciences 123, 326–352.

Ai, J., Xiao, S., Feng, H., Wang, H., Jia, G., Hu, Y., 2020. A global terrestrial ecosystem respiration dataset (2001-2010) estimated with MODIS land surface temperature and vegetation indices. Big Earth Data 4, 142–152.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. FAO Irrigation and drainage paper No. 56. Rome Food Agric. Organ. U. N. 56, e156.

Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J.B., 2006. A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. Agric. Water Manag. 81, 1–22.

Amani, S., Shafizadeh-Moghadam, H., 2023. A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. Agric. Water Manag. 284, 108324.

Amani, M., Parsian, S., MirMazloumi, S. M., & Aieneh, O. (2016). Two new soil moisture indices based on the NIR-red triangle space of Landsat-8 data. International Journal of Applied Earth Observation and Geoinformation, 50, 176–186.

Amani, M., Salehi, B., Mahdavi, S., Masjedi, A., & Dehnavi, S. (2017). Temperature-vegetation-soil moisture dryness index (TVMDI). Remote Sensing of Environment, 197, 1–14.

Anderegg, W.R., Trugman, A.T., Badgley, G., Anderson, C.M., Bartuska, A., Ciais, P., Cullenward, D., Field, C.B., Freeman, J., Goetz, S.J., 2020. Climate-driven risks to the climate mitigation potential of forests. Science 368, eaaz7005.

Anderson, M., 1997. A Two-Source Time-Integrated Model for Estimating Surface Fluxes Using Thermal Infrared Remote Sensing. Remote Sens. Environ. 60, 195–216. https://doi.org/10.1016/S0034-4257(96)00215-5

Anderson, M., Zolin, C., Sentelhas, P., Hain, C., Semmens, K., Yilmaz, M., Gao, F., Otkin, J., Tetrault, R., 2016. Assessing correlations of satellite-derived evapotranspiration, precipitation, and leaf area index anomalies with yields of major Brazilian crops. Remote Sens Env. 174, 82–99.

Anderson, M.C., Hain, C., Otkin, J., Zhan, X., Mo, K., Svoboda, M., Wardlow, B., Pimstein, A., 2013. An intercomparison of drought indicators based on thermal remote sensing and NLDAS-2 simulations with US Drought Monitor classifications. J. Hydrometeorol. 14, 1035–1056.

Anderson, M.C., Kustas, W.P., Alfieri, J.G., Gao, F., Hain, C., Prueger, J.H., Evett, S., Colaizzi, P., Howell, T., Chávez, J.L., 2012. Mapping daily evapotranspiration at Landsat spatial scales during the BEAREX'08 field campaign. Adv. Water Resour. 50, 162–177.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. J. Geophys. Res. Atmospheres 112, 2006JD007506. https://doi.org/10.1029/2006JD007506

Anderson, M.C., Norman, J.M., Meyers, T.P., Diak, G.R., 2000. An analytical model for estimating canopy transpiration and carbon assimilation fluxes based on canopy light-use efficiency. Agric. For. Meteorol. 101, 265–289.

Angelov, P.P., Gu, X., 2019. Empirical Approach to Machine Learning, Studies in Computational Intelligence. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-02384-3

Aubinet, M., Vesala, T., Papale, D., 2012. Eddy covariance: a practical guide to measurement and data analysis. Springer Science & Business Media.

Augustine, J.A., DeLuisi, J.J., Long, C.N., 2000. SURFRAD–A national surface radiation budget network for atmospheric research. Bull. Am. Meteorol. Soc. 81, 2341–2358.

Austin, A.T., Vivanco, L., 2006. Plant litter decomposition in a semi-arid ecosystem controlled by photodegradation. Nature 442, 555–558. https://doi.org/10.1038/nature05038

Babaeian, E., Sidike, P., Newcomb, M. S., Maimaitijiang, M., White, S. A., Demieville, J., et al. (2019). A new optical remote sensing technique for high-resolution mapping of soil moisture. Frontiers in Big Data, 2, 37.

Badgley, G., Anderegg, L.D., Berry, J.A., Field, C.B., 2019. Terrestrial gross primary production: Using NIRV to scale from site to globe. Glob. Change Biol. 25, 3731–3740.

Badgley, G., Field, C.B., Berry, J.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Sci. Adv. 3, e1602244.

Bahrami, H., Homayouni, S., Mahdianpari, M., Safari, A., 2021a. Ensemble Learning for Crop Monitoring from Multitemporal Optical and Synthetic Aperture Radar Earth Observations. Presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, pp. 5929–5932.

Bahrami, H., Homayouni, S., Safari, A., Mirzaei, S., Mahdianpari, M., Reisi-Gahrouei, O., 2021b. Deep learning-based estimation of crop biophysical parameters using multi-source and multi-temporal remote sensing observations. Agronomy 11, 1363.

Bai, L., Cai, J., Liu, Y., Chen, H., Zhang, B., Huang, L., 2017. Responses of field evapotranspiration to the changes of cropping pattern and groundwater depth in large irrigation district of Yellow River basin. Agric. Water Manag. 188, 1–11. https://doi.org/10.1016/j.agwat.2017.03.028

Baldocchi, D.D., Ryu, Y., Dechant, B., Eichelmann, E., Hemes, K., Ma, S., Sanchez, C.R., Shortt, R., Szutu, D., Valach, A., 2020. Outgoing near-infrared radiation from vegetation scales with canopy photosynthesis across a spectrum of function, structure, physiological capacity, and weather. J. Geophys. Res. Biogeosciences 125, e2019JG005534.

Bao, Y., Liu, T., Duan, L., Tong, X., Zhang, L., Singh, V.P., Lei, H., Wang, G., 2021. Comparison of an improved Penman-Monteith model and SWH model for estimating evapotranspiration in a meadow wetland in a semiarid region. Sci. Total Environ. 795, 148736.

Baraldi, A., Binaghi, E., Blonda, P., Brivio, P.A., Rampini, A., 2001. Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data. IEEE Trans. Geosci. Remote Sens. 39, 994–1005.

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. J. Hydrol. 212–213, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4

Bayat, B., van der Tol, C., Verhoef, W., 2020. Retrieval of land surface properties from an annual time series of Landsat TOA radiances during a drought episode using coupled radiative transfer models. Remote Sens. Environ. 238, 110917.

Belitz, K., Stackelberg, P., 2021. Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environ. Model. Softw. 139, 105006.

Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2010. Validation of biophysical models: issues and methodologies. A review. Agron. Sustain. Dev. 30, 109–130.

Benseghir, L., Bachari, N.E.I., 2021. Shortwave infrared vegetation index-based modelling for aboveground vegetation biomass assessment in the arid steppes of Algeria. Afr. J. Range Forage Sci. 38, 281–290.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54, 1937–1967.

Bosiö, J., Stiegler, C., Johansson, M., Mbufong, H.N., Christensen, T.R., 2014. Increased photosynthesis compensates for shorter growing season in subarctic tundra—8 years of snow accumulation manipulations. Clim. Change 127, 321–334.

Brighenti, T.M., Bonumá, N.B., Srinivasan, R., Chaffe, P.L.B., 2019. Simulating sub-daily hydrological process with SWAT: a review. Hydrol. Sci. J. 64, 1415–1423. https://doi.org/10.1080/02626667.2019.1642477

Brooks, D.R., England, C.F., Hunt, G.E., Minnis, P., 1984. An intercalibration of Meteosat-1 and GOES-2 visible and infrared measurements. J Atmos Ocean. Technol 1, 283–286.

Brunton, S.L., Kutz, J.N., 2022. Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press.

Brutsaert, W., 2014. Daily evaporation from drying soil: Universal parameterization with similarity. Water Resour. Res. 50, 3206–3215. https://doi.org/10.1002/2013WR014872

Cai, J., Xu, K., Zhu, Y., Hu, F., Li, L., 2020. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Appl. Energy 262, 114566.

Chang, J.F., Viovy, N., Vuichard, N., Ciais, P., Wang, T., Cozic, A., Lardy, R., Graux, A.-I., Klumpp, K., Martin, R., 2013. Incorporating grassland management in ORCHIDEE: model description and evaluation at 11 eddy-covariance sites in Europe. Geosci. Model Dev. 6, 2165–2181.

Chen, J., Zhu, W., Yu, Q., 2021. Estimating half-hourly solar radiation over the Continental United States using GOES-16 data with iterative random forest. Renew. Energy 178, 916–929.

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sens. Environ. 237, 111594. https://doi.org/10.1016/j.rse.2019.111594

Chen, L., Yan, G., Wang, T., Ren, H., Calbó, J., Zhao, J., McKenzie, R., 2012. Estimation of surface shortwave radiation components under all sky conditions: Modeling and sensitivity analysis. Remote Sens. Environ. 123, 457–469.

Chen, S., Mihara, K., Wen, J., 2018. Time series prediction of CO2, TVOC and HCHO based on machine learning at different sampling points. Build. Environ. 146, 238–246.

Chen, S., Sui, L., Liu, L., Liu, X., Li, J., Huang, L., Li, X., Qian, X., 2023. NIRvP as a remote sensing proxy for measuring gross primary production across different biomes and climate zones: Performance and limitations. Int. J. Appl. Earth Obs. Geoinformation 122, 103437. https://doi.org/10.1016/j.jag.2023.103437

Chen, W., Liu, S., Zhao, S., Zhu, Y., Feng, S., Wang, Z., Wu, Y., Xiao, J., Yuan, W., Yan, W., 2023. Temporal dynamics of ecosystem, inherent, and underlying water use efficiencies of forests, grasslands, and croplands and their responses to climate change. Carbon Balance Manag. 18, 1–14.

Cho, D., Bae, D., Yoo, C., Im, J., Lee, Y., Lee, S., 2022. All-Sky 1 km MODIS Land Surface Temperature Reconstruction Considering Cloud Effects Based on Machine Learning. Remote Sens. 14, 1815. https://doi.org/10.3390/rs14081815

Chu, H., Baldocchi, D.D., John, R., Wolf, S., Reichstein, M., 2017. Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. J. Geophys. Res. Biogeosciences 122, 289–307.

Chu, H., Christianson, D.S., Cheah, Y.-W., Pastorello, G., O'Brien, F., Geden, J., Ngo, S.-T., Hollowgrass, R., Leibowitz, K., Beekwilder, N.F., 2023. AmeriFlux BASE data pipeline to support network growth and data sharing. Sci. Data 10, 614.

COP15—Copenhagen, U., 2009. Methodological guidance for activities relating to reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries.

Cuntz, M., 2020. hesseflux: a Python library to process and post-process Eddy covariance data.

Davis, G.K., 2007. History of the NOAA satellite program. J. Appl. Remote Sens. 1, 012504.

De Beurs, K.M., Henebry, G.M., 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. Glob. Change Biol. 11, 779–790.

Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., Rossini, M., 2022a. NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales. Remote Sens. Environ. 268, 112763.

Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., Rossini, M., 2022b. NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales. Remote Sens. Environ. 268, 112763.

Desai, A.R., Khan, A.M., Zheng, T., Paleri, S., Butterworth, B., Lee, T.R., Fisher, J.B., Hulley, G., Kleynhans, T., Gerace, A., 2021. Multi-sensor approach for high space and time resolution land surface temperature. Earth Space Sci. 8, e2021EA001842.

Descals, A., Verger, A., Yin, G., Filella, I., Peñuelas, J., 2023. Local interpretation of machine learning models in remote sensing with SHAP: the case of global climate constraints on photosynthesis phenology. Int. J. Remote Sens. 44, 3160–3173. https://doi.org/10.1080/01431161.2023.2217982

Dhake, H., Kashyap, Y., Kosmopoulos, P., 2023. Algorithms for hyperparameter tuning of lstms for time series forecasting. Remote Sens. 15, 2076.

Diak, G.R., Gautier, C., 1983. Improvements to a simple physical model for estimating insolation from GOES data. J. Clim. Appl. Meteorol. 505–508.

Diaz-Gonzalez, F.A., Vuelvas, J., Correa, C.A., Vallejo, V.E., Patino, D., 2022. Machine learning and remote sensing techniques applied to estimate soil indicators–review. Ecol. Indic. 135, 108517.

Diffenbaugh, N.S., Singh, D., Mankin, J.S., Horton, D.E., Swain, D.L., Touma, D., Charland, A., Liu, Y., Haugen, M., Tsiang, M., 2017. Quantifying the influence of global warming on unprecedented extreme climate events. Proc. Natl. Acad. Sci. 114, 4881–4886.

Dilling, L., 2007. Towards science in support of decision making: characterizing the supply of carbon cycle science. Environ. Sci. Policy 10, 48–61.

Ding, L., Zhou, J., Li, Z.-L., Ma, J., Shi, C., Sun, S., Wang, Z., 2022. Reconstruction of Hourly

All-Weather Land Surface Temperature by Integrating Reanalysis Data and Thermal Infrared Data From Geostationary Satellites (RTG). IEEE Trans. Geosci. Remote Sens. 60, 1–17. https://doi.org/10.1109/TGRS.2022.3227074

Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol. Econ. 69, 1858–1868.

Dowell, D.C., Alexander, C.R., James, E.P., Weygandt, S.S., Benjamin, S.G., Manikin, G.S., Blake, B.T., Brown, J.M., Olson, J.B., Hu, M., 2022. The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. Weather Forecast. 37, 1371–1395.

Drexler, J.Z., Snyder, R.L., Spano, D., Paw U, K.T., 2004. A review of models and micrometeorological methods used to estimate wetland evapotranspiration. Hydrol. Process. 18, 2071–2101. https://doi.org/10.1002/hyp.1462

Duan, S.-B., Lian, Y., Zhao, E., Chen, H., Han, W., Wu, Z., 2023. A Novel Approach to All-Weather LST Estimation Using XGBoost Model and Multisource Data. IEEE Trans. Geosci. Remote Sens. 61, 1–14. https://doi.org/10.1109/TGRS.2023.3324481

Elbeltagi, A., Srivastava, A., Deng, J., Li, Z., Raza, A., Khadke, L., Yu, Z., El-Rawy, M., 2023. Forecasting vapor pressure deficit for agricultural water management using machine learning in semi-arid environments. Agric. Water Manag. 283, 108302.

Ellrod, G.P., Pryor, K., 2019. Applications of geostationary satellite data to aviation. Pure Appl. Geophys. 176, 2017–2043.

Ezzine, I., Benhlima, L., 2018. A study of handling missing data methods for big data. Presented at the 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), IEEE, pp. 498–501.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. For. Meteorol. 107, 43–69. https://doi.org/10.1016/S0168-1923(00)00225-2

Felton, A.J., Goldsmith, G.R., 2023. Timing and magnitude of drought impacts on carbon uptake across a grassland biome. Glob. Change Biol. 29, 2790–2803.

Feng, X., Thompson, S.E., Woods, R., Porporato, A., 2019. Quantifying Asynchronicity of Precipitation and Potential Evapotranspiration in Mediterranean Climates. Geophys. Res. Lett. 46, 14692–14701. https://doi.org/10.1029/2019GL085653

Firozjaei, M.K., Kiavarz, M., Alavipanah, S.K., 2022. Satellite-derived land surface temperature spatial sharpening: A comprehensive review on current status and perspectives. Eur. J. Remote Sens. 55, 644–664. https://doi.org/10.1080/22797254.2022.2144764

Fisher, J.B., Baldocchi, D.D., Misson, L., Dawson, T.E., Goldstein, A.H., 2007. What the towers don’t see at night: nocturnal sap flow in trees and shrubs at two AmeriFlux sites in California. Tree Physiol. 27, 597–610.

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resour. Res. 56, e2019WR026058. https://doi.org/10.1029/2019WR026058

Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Fleischmann, A.S., Laipelt, L., Papa, F., Paiva, R.C.D. de, De Andrade, B.C., Collischonn, W., Biudes, M.S., Kayser, R., Prigent, C., Cosio, E., 2023. Patterns and drivers of evapotranspiration in South American wetlands. Nat. Commun. 14, 6656.

Friedlingstein, P., Jones, M.W., O'Sullivan, M., Andrew, R.M., Bakker, D.C., Hauck, J., Le Quéré, C., Peters, G.P., Peters, W., Pongratz, J., 2022a. Global carbon budget 2021. Earth Syst. Sci. Data 14, 1917–2005.

Friedlingstein, P., O'sullivan, M., Jones, M.W., Andrew, R.M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I.T., Olsen, A., Peters, G.P., 2022b. Global carbon budget 2022. Earth Syst. Sci. Data Discuss. 2022, 1–159.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Fu, H., Shao, Z., Fu, P., Huang, X., Cheng, T., Fan, Y., 2022. Combining ATC and 3D-CNN for reconstructing spatially and temporally continuous land surface temperature. Int. J. Appl.

Earth Obs. Geoinformation 108, 102733. https://doi.org/10.1016/j.jag.2022.102733

Fu, Z., Gerken, T., Bromley, G., Araújo, A., Bonal, D., Burban, B., Ficklin, D., Fuentes, J.D., Goulden, M., Hirano, T., 2018. The surface-atmosphere exchange of carbon dioxide in tropical rainforests: Sensitivity to environmental drivers and flux measurement methodology. Agric. For. Meteorol. 263, 292–307.

Furqan, M., Goswami, B., 2022. Satellite communication networks. Handb. Real-Time Comput. Tian -C Levy DC Eds 1–22.

Gao, D., Yao, J., Yu, S., Ma, Y., Li, L., Gao, Z., 2023. Eddy Covariance CO2 Flux Gap Filling for Long Data Gaps: A Novel Framework Based on Machine Learning and Time Series Decomposition. Remote Sens. 15, 2695.

García, M., Sandholt, I., Ceccato, P., Ridler, M., Mougin, E., Kergoat, L., Morillas, L., Timouk, F., Fensholt, R., Domingo, F., 2013. Actual evapotranspiration in drylands derived from in-situ and satellite data: Assessing biophysical constraints. Remote Sens. Environ. 131, 103–118.

Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636.

Gautier, C., Diak, G., Masse, S., 1980. A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. J. Appl. Meteorol. Climatol. 19, 1005–1012.

Gessner, U., Machwitz, M., Conrad, C., Dech, S., 2013. Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. Remote Sens. Environ. 129, 90–102.

Ghanbari, H., Mahdianpari, M., Homayouni, S., Mohammadimanesh, F., 2021. A meta-analysis of convolutional neural networks for remote sensing applications. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3602–3613.

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Sharma, E., Ali, M., 2022. Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement 202, 111759.

Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, J.S., Hook, S., Kahle, A.B., 1998. A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. IEEE Trans. Geosci. Remote Sens. 36, 1113–1126. https://doi.org/10.1109/36.700995

Giorgi, G., Schmidt, T.D., Trainotti, C., Mata-Calvo, R., Fuchs, C., Hoque, M.M., Berdermann, J., Furthner, J., Günther, C., Schuldt, T., 2019. Advanced technologies for satellite navigation and geodesy. Adv. Space Res. 64, 1256–1273.

Gitelson, A.A., Peng, Y., Arkebauer, T.J., Schepers, J., 2014. Relationships between gross primary production, green LAI, and canopy chlorophyll content in maize: Implications for remote sensing of primary production. Remote Sens. Environ. 144, 65–72.

Golden, C.E., Rothrock Jr, M.J., Mishra, A., 2019. Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms. Food Res. Int. 122, 47–55.

Gong, Y., Li, H., Shen, H., Meng, C., Wu, P., 2023. Cloud-covered MODIS LST reconstruction by combining assimilation data and remote sensing data through a nonlocality-reinforced network. Int. J. Appl. Earth Obs. Geoinformation 117, 103195. https://doi.org/10.1016/j.jag.2023.103195

Gong, Y., Staudhammer, C.L., Wiesner, S., Zhang, Y., Cannon, J.B., Starr, G., 2022. Uncertainty in parameterizing a flux-based model of vegetation carbon phenology using ecosystem respiration. Ecosphere 13, e4101.

Goodman, S.J., Schmit, T.J., Daniels, J., Redmon, R.J., 2019. The GOES-R series: a new generation of geostationary environmental satellites. Elsevier.

Grace, J., 2004. Understanding and managing the global carbon cycle. J. Ecol. 92, 189–202.

Guicquero, W., Sicard, G., 2022. BILLNET: A Binarized Conv3D-LSTM Network with Logic-gated residual architecture for hardware-efficient video inference. Presented at the 2022 IEEE Workshop on Signal Processing Systems (SiPS), IEEE, pp. 1–6.

Han, L., Wu, P., Chu, C., 2023. A distributionally robust production planning model for maximizing customer satisfaction with budget and carbon emissions constraints. Comput. Ind. Eng. 182, 109412.

Hao, D., Asrar, G.R., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., Chen, M., 2019. Estimating hourly land surface downward shortwave and photosynthetically active radiation from DSCOVR/EPIC observations. Remote Sens. Environ. 232, 111320.

Harmon, M.E., Bond-Lamberty, B., Tang, J., Vargas, R., 2011. Heterotrophic respiration in disturbed forests: A review with examples from North America. J. Geophys. Res. Biogeosciences 116.

Hashimoto, H., Wang, W., Dungan, J.L., Li, S., Michaelis, A.R., Takenaka, H., Higuchi, A., Myneni, R.B., Nemani, R.R., 2021. New generation geostationary satellite observations support seasonality in greenness of the Amazon evergreen forests. Nat. Commun. 12, 684.

He, T., Zhang, Y., Liang, S., Yu, Y., Wang, D., 2019. Developing land surface directional reflectance and albedo products from geostationary GOES-R and Himawari data:

Theoretical basis, operational implementation, and validation. Remote Sens. 11, 2655.

Heidinger, A.K., Pavolonis, M.J., Calvert, C., Hoffman, J., Nebuda, S., Straka, W., Walther, A., Wanzong, S., 2020. ABI Cloud Products from the GOES-R Series, in: The GOES-R Series. Elsevier, pp. 43–62. https://doi.org/10.1016/B978-0-12-814327-8.00006-8

Heinsch, F.A., Zhao, M., Running, S.W., Kimball, J.S., Nemani, R.R., Davis, K.J., Bolstad, P.V., Cook, B.D., Desai, A.R., Ricciuto, D.M., 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. IEEE Trans. Geosci. Remote Sens. 44, 1908–1925.

Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, T., Schröter, K., 2005. Global biomass mapping for an improved understanding of the CO2 balance—the Earth observation mission Carbon-3D. Remote Sens. Environ. 94, 94–104.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Hori, M., Aoki, T., Tanikawa, T., Motoyoshi, H., Hachikubo, A., Sugiura, K., Yasunari, T.J., Eide, H., Storvold, R., Nakajima, Y., Takahashi, F., 2006. In-situ measured spectral directional emissivity of snow and ice in the 8–14 μm atmospheric window. Remote Sens. Environ. 100, 486–502. https://doi.org/10.1016/j.rse.2005.11.001

Hou, N., Zhang, X., Zhang, W., Wei, Y., Jia, K., Yao, Y., Jiang, B., Cheng, J., 2020. Estimation of surface downward shortwave radiation over china from himawari-8 ahi data based on random forest. Remote Sens. 12, 181.

Houghton, R., Hobbie, J., Melillo, J.M., Moore, B., Peterson, B., Shaver, G., Woodwell, G., 1983. Changes in the Carbon Content of Terrestrial Biota and Soils between 1860 and 1980: A Net Release of CO" 2 to the Atmosphere. Ecol. Monogr. 53, 235–262.

Hu, T., Mallick, K., Hulley, G.C., Planells, L.P., Göttsche, F.M., Schlerf, M., Hitzelberger, P., Didry, Y., Szantoi, Z., Alonso, I., 2022. Continental-scale evaluation of three ECOSTRESS land surface temperature products over Europe and Africa: Temperature-based validation and cross-satellite comparison. Remote Sens. Environ. 282, 113296.

Hu, T., Renzullo, L.J., van Dijk, A.I., He, J., Tian, S., Xu, Z., Zhou, J., Liu, T., Liu, Q., 2020. Monitoring agricultural drought in Australia using MTSAT-2 land surface temperature retrievals. Remote Sens. Environ. 236, 111419.

Huang, X., Xiao, J., Wang, X., Ma, M., 2021. Improving the global MODIS GPP model by optimizing parameters with FLUXNET data. Agric. For. Meteorol. 300, 108314.

Hulley, G.C., Ghent, D., Göttsche, F.M., Guillevic, P.C., Mildrexler, D.J., Coll, C., 2019. Land Surface Temperature, in: Taking the Temperature of the Earth. Elsevier, pp. 57–127. https://doi.org/10.1016/B978-0-12-814458-9.00003-4

Hwang, D.J., Frouin, R., Tan, J., Ahn, J.-H., Choi, J.-K., Moon, J.-E., Ryu, J.-H., 2022. Algorithm to estimate daily PAR at the ocean surface from GOCI data: description and evaluation. Front. Mar. Sci. 9, 924967.

Ilčev, S.D., Ilčev, S.D., 2019. Satellite Meteorological Networks. Glob. Satell. Meteorol. Obs. GSMO Appl. Vol. 2 1–63.

Iqbal, M., 2012. An introduction to solar radiation. Elsevier.

Irisarri, J.G.N., Oesterheld, M., Paruelo, J.M., Texeira, M.A., 2012. Patterns and controls of above-ground net primary production in meadows of Patagonia. A remote sensing approach. J. Veg. Sci. 23, 114–126.

Irvin, J., Zhou, S., McNicol, G., Lu, F., Liu, V., Fluet-Chouinard, E., Ouyang, Z., Knox, S.H., Lucas-Moffat, A., Trotta, C., 2021. Gap-filling eddy covariance methane fluxes:

Comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands. Agric. For. Meteorol. 308, 108528.

Jahan, E., Sharwood, R.E., Tissue, D.T., 2023. Effects of leaf age during drought and recovery on photosynthesis, mesophyll conductance and leaf anatomy in wheat leaves. Front. Plant Sci. 14, 1091418.

James, E.P., Alexander, C.R., Dowell, D.C., Weygandt, S.S., Benjamin, S.G., Manikin, G.S., Brown, J.M., Olson, J.B., Hu, M., Smirnova, T.G., 2022. The High-Resolution Rapid Refresh (HRRR): an hourly updating convection-allowing forecast model. Part II: Forecast performance. Weather Forecast. 37, 1397–1417.

Jansen, P.A., Muller-Landau, H.C., Wright, S.J., 2010. Bushmeat hunting and climate: an indirect link. Science 327, 30–30.

Jeong, S., Ryu, Y., Dechant, B., Li, X., Kong, J., Choi, W., Kang, M., Yeom, J., Lim, J., Jang, K., 2023a. Tracking diurnal to seasonal variations of gross primary productivity using a geostationary satellite, GK-2A advanced meteorological imager. Remote Sens. Environ. 284, 113365.

Jeong, S., Ryu, Y., Dechant, B., Li, X., Kong, J., Choi, W., Kang, M., Yeom, J., Lim, J., Jang, K., Chun, J., 2023b. Tracking diurnal to seasonal variations of gross primary productivity using a geostationary satellite, GK-2A advanced meteorological imager. Remote Sens. Environ. 284, 113365. https://doi.org/10.1016/j.rse.2022.113365

Jia, A., Liang, S., Wang, D., 2022. Generating a 2-km, all-sky, hourly land surface temperature product from Advanced Baseline Imager data. Remote Sens. Environ. 278, 113105. https://doi.org/10.1016/j.rse.2022.113105

Jia, A., Liang, S., Wang, D., Ma, L., Wang, Z., Xu, S., 2023. Global hourly, 5 km, all-sky land surface temperature data from 2011 to 2021 based on integrating geostationary and polar-orbiting satellite data. Earth Syst. Sci. Data 15, 869–895. https://doi.org/10.5194/essd-15-869-2023

Jia, A., Liang, S., Wang, D., Mallick, K., Zhou, S., Hu, T., Xu, S., 2024. Advances in Methodology and Generation of All-Weather Land Surface Temperature Products From Polar-Orbiting and Geostationary Satellites: A comprehensive review. IEEE Geosci. Remote Sens. Mag. 2–43. https://doi.org/10.1109/MGRS.2024.3421268

Jia, A., Ma, H., Liang, S., Wang, D., 2021. Cloudy-sky land surface temperature from VIIRS and MODIS satellite data using a surface energy balance-based method. Remote Sens. Environ. 263, 112566. https://doi.org/10.1016/j.rse.2021.112566

Jia, X., Mu, Y., Zha, T., Wang, B., Qin, S., Tian, Y., 2020. Seasonal and interannual variations in ecosystem respiration in relation to temperature, moisture, and productivity in a temperate semi-arid shrubland. Sci. Total Environ. 709, 136210.

Jiang, H., Lu, N., Qin, J., Tang, W., Yao, L., 2019. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renew. Sustain. Energy Rev. 114, 109327.

Jun, M.-J., 2021. A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area. Int. J. Geogr. Inf. Sci. 35, 2149–2167.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. Sci. Data 6, 74.

Jung, M., Reichstein, M., Schwalm, C.R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., 2017. Compensatory water effects link yearly global land CO2 sink changes to temperature. Nature 541, 516–520.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach.

Kang, Y., Kim, M., Kang, E., Cho, D., Im, J., 2022. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia. ISPRS J. Photogramm. Remote Sens. 183, 253–268.

Katul, G.G., Oren, R., Manzoni, S., Higgins, C., Parlange, M.B., 2012. Evapotranspiration: A process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. Rev. Geophys. 50, 2011RG000366. https://doi.org/10.1029/2011RG000366

Khan, A., Stoy, P., Joiner, J., Baldocchi, D., Verfaillie, J., Chen, M., Otkin, J., 2022. The Diurnal Dynamics of Gross Primary Productivity Using Observations From the Advanced Baseline Imager on the Geostationary Operational Environmental Satellite-R Series at an Oak Savanna Ecosystem. J. Geophys. Res. Biogeosciences 127, e2021JG006701.

Khan, A.M., Stoy, P.C., Douglas, J.T., Anderson, M., Diak, G., Otkin, J.A., Hain, C., Rehbein, E.M., McCorkel, J., 2021a. Reviews and syntheses: Ongoing and emerging opportunities to improve environmental science using observations from the Advanced Baseline Imager on the Geostationary Operational Environmental Satellites. Biogeosciences 18, 4117–4141.

Khan, A.M., Stoy, P.C., Douglas, J.T., Anderson, M., Diak, G., Otkin, J.A., Hain, C., Rehbein, E.M., McCorkel, J., 2021b. Reviews and syntheses: Ongoing and emerging opportunities to improve environmental science using observations from the Advanced Baseline

Imager on the Geostationary Operational Environmental Satellites. Biogeosciences 18, 4117–4141.

Khan, A.M., Stoy, P.C., Joiner, J., Baldocchi, D., Verfaillie, J., Chen, M., Otkin, J.A., 2022. The Diurnal Dynamics of Gross Primary Productivity Using Observations From the Advanced Baseline Imager on the Geostationary Operational Environmental Satellite-R Series at an Oak Savanna Ecosystem. J. Geophys. Res. Biogeosciences 127, e2021JG006701.

Kim, H.-Y., Laszlo, I., Liu, H., 2020. Shortwave Radiation Budget Products from GOES-R Series ABI. Presented at the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp. 6698–6701.

Kim, Y., Johnson, M.S., Knox, S.H., Black, T.A., Dalmagro, H.J., Kang, M., Kim, J., Baldocchi, D., 2020. Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. Glob. Change Biol. 26, 1499–1518.

Knapp, K.R., Frouin, R., Kondragunta, S., Prados, A., 2005. Toward aerosol optical depth retrievals over land from GOES visible radiances: determining surface reflectance. Int. J. Remote Sens. 26, 4097–4116.

Kovács, D.D., Reyes-Muñoz, P., Salinero-Delgado, M., Mészáros, V.I., Berger, K., Verrelst, J., 2023. Cloud-Free Global Maps of Essential Vegetation Traits Processed from the TOA Sentinel-3 Catalogue in Google Earth Engine. Remote Sens. 15, 3404.

Krishnan, P., Meyers, T.P., Scott, R.L., Kennedy, L., Heuer, M., 2012. Energy exchange and evapotranspiration over two temperate semi-arid grasslands in North America. Agric. For. Meteorol. 153, 31–44.

Küçük, Ç., Koirala, S., Carvalhais, N., Miralles, D.G., Reichstein, M., Jung, M., 2022. Characterizing the response of vegetation cover to water limitation in Africa using geostationary satellites. J. Adv. Model. Earth Syst. 14, e2021MS002730.

Kursa, M.B., 2014. Robustness of Random Forest-based gene selection methods. BMC Bioinformatics 15, 1–8.

Kustas, W.P., Norman, J.M., 1999. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agric. For. Meteorol. 94, 13–29.

Labedzki, L., 2011. Evapotranspiration. BoD–Books on Demand.

Laliberté, J., Bélanger, S., Frouin, R., 2016. Evaluation of satellite-based algorithms to estimate photosynthetically available radiation (PAR) reaching the ocean surface at high northern latitudes. Remote Sens. Environ. 184, 199–211.

Lara-Alvarez, C., Flores, J.J., Rodriguez-Rangel, H., Lopez-Farias, R., 2024. A literature review on satellite image time series forecasting: Methods and applications for remote sensing. WIREs Data Min. Knowl. Discov. 14, e1528. https://doi.org/10.1002/widm.1528

Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob. Change Biol. 16, 187–208.

Laszlo, I., Liu, H., Kim, H.-Y., Pinker, R.T., 2020a. Shortwave Radiation from ABI on the GOES-R Series, in: The GOES-R Series. Elsevier, pp. 179–191.

Laszlo, I., Liu, H., Kim, H.-Y., Pinker, R.T., 2020b. Chapter 15 - Shortwave Radiation from ABI on the GOES-R Series, in: Goodman, S.J., Schmit, T.J., Daniels, J., Redmon, R.J. (Eds.), The GOES-R Series. Elsevier, pp. 179–191. https://doi.org/10.1016/B978-0-12-814327-8.00015-9

Laszlo, I., Liu, H., Kim, H.-Y., Pinker, R.T., 2020c. Shortwave Radiation from ABI on the GOES-R Series, in: The GOES-R Series. Elsevier, pp. 179–191.

Lawton, J.H., Brown, V.K., Schulze, E.-D., Mooney, H.A., 1993. Biodiversity and ecosystem function. Biodivers. Ecosyst. Funct. 255–270.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436–444.

Lees, K.J., Quaife, T., Artz, R.R.E., Khomik, M., Clark, J.M., 2018. Potential for using remote sensing to estimate carbon fluxes across northern peatlands–A review. Sci. Total Environ. 615, 857–874.

Lei, L., Xia, J., Li, X., Huang, K., Zhang, A., Chen, S., Weng, E., Luo, Y., Wan, S., 2018. Water response of ecosystem respiration regulates future projection of net ecosystem productivity in a semiarid grassland. Agric. For. Meteorol. 252, 175–191.

Letu, H., Nakajima, T.Y., Wang, T., Shang, H., Ma, R., Yang, K., Baran, A.J., Riedi, J., Ishimoto, H., Yoshida, M., 2022. A new benchmark for surface radiation products over the East Asia–Pacific region retrieved from the Himawari-8/AHI next-generation geostationary satellite. Bull. Am. Meteorol. Soc. 103, E873–E888.

Lhomme, J.-P., Katerji, N., Perrier, A., Bertolini, J.-M., 1988. Radiative surface temperature and convective flux calculation over crop canopies. Bound.-Layer Meteorol. 43, 383–392. https://doi.org/10.1007/BF00121714

Li, B., Liang, S., Liu, X., Ma, H., Chen, Y., Liang, T., He, T., 2021. Estimation of all-sky 1 km land surface temperature over the conterminous United States. Remote Sens. Environ. 266, 112707. https://doi.org/10.1016/j.rse.2021.112707

Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., Chanussot, J., 2022. Deep learning in multimodal remote sensing data fusion: A comprehensive review. Int. J. Appl. Earth Obs. Geoinformation 112, 102926.

Li, L., Xin, X., Zhang, H., Yu, J., Liu, Q., Yu, S., Wen, J., 2015. A method for estimating hourly photosynthetically active radiation (PAR) in China by combining geostationary and polar-orbiting satellite data. Remote Sens. Environ. 165, 14–26.

Li, L., Zeng, Z., Zhang, G., Duan, K., Liu, B., Cai, X., 2022. Exploring the Individualized Effect of Climatic Drivers on MODIS Net Primary Productivity through an Explainable Machine Learning Framework. Remote Sens. 14, 4401.

Li, R., Wang, D., Liang, S., 2023. Comparison between deep learning architectures for the 1 km, 10/15-min estimation of downward shortwave radiation from AHI and ABI. Remote Sens. Environ. 295, 113697.

Li, X., Ryu, Y., Xiao, J., Dechant, B., Liu, J., Li, B., Jeong, S., Gentine, P., 2023. New-generation geostationary satellite reveals widespread midday depression in dryland photosynthesis during 2020 western US heatwave. Sci. Adv. 9, eadi0775.

Li, Y., Li, L., Dong, J., Bai, J., 2021. Assessing MODIS carbon and water fluxes in grasslands and shrublands in semiarid regions using eddy covariance tower data. Int. J. Remote Sens. 42, 595–616.

Li, Z., Wu, H., Duan, S., Zhao, W., Ren, H., Liu, X., Leng, P., Tang, R., Ye, X., Zhu, J., Sun, Y., Si, M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B., Yan, G., Zhou, C., 2023. Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications. Rev. Geophys. 61, e2022RG000777. https://doi.org/10.1029/2022RG000777

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-derived land surface temperature: Current status and perspectives. Remote Sens. Environ. 131, 14–37. https://doi.org/10.1016/j.rse.2012.12.008

Liang, H., Zhang, Z., Hu, C., Gong, Y., Cheng, D., 2023. A Survey on Spatio-temporal Big Data Analytics Ecosystem: Resource Management, Processing Platform, and Applications. IEEE Trans. Big Data.

Lin, L., Hao, X., Zhang, B., Zou, C.-Z., Cao, C., 2021. Assessment of the Reprocessed Suomi NPP VIIRS Enterprise Cloud Mask Product. Remote Sens. 13, 2502. https://doi.org/10.3390/rs13132502

Ling, M., Yang, Y., Xu, C., Yu, L., Xia, Q., Guo, X., 2022. Temporal and Spatial Variation Characteristics of Actual Evapotranspiration in the Yiluo River Basin Based on the Priestley–Taylor Jet Propulsion Laboratory Model. Appl. Sci. 12, 9784. https://doi.org/10.3390/app12199784

Littlefield, C.E., D'Amato, A.W., 2022. Identifying trade-offs and opportunities for forest carbon and wildlife using a climate change adaptation lens. Conserv. Sci. Pract. 4, e12631.

Liu, F., Wang, X., Sun, F., Wang, H., Wu, L., Zhang, X., Liu, W., Che, H., 2022. Correction of Overestimation in Observed Land Surface Temperatures Based on Machine Learning Models. J. Clim. 35, 5359–5377. https://doi.org/10.1175/JCLI-D-21-0447.1

Liu, S., Fan, H., Ferianc, M., Niu, X., Shi, H., Luk, W., 2021. Toward full-stack acceleration of deep convolutional neural networks on FPGAs. IEEE Trans. Neural Netw. Learn. Syst. 33, 3974–3987.

Liu, W., Cheng, J., Wang, Q., 2023. Estimating Hourly All-Weather Land Surface Temperature From FY-4A/AGRI Imagery Using the Surface Energy Balance Theory. IEEE Trans.

Geosci. Remote Sens. 61, 1–18. https://doi.org/10.1109/TGRS.2023.3254211

Liu, Y., Wu, C., Peng, D., Xu, S., Gonsamo, A., Jassal, R.S., Arain, M.A., Lu, L., Fang, B., Chen, J.M., 2016. Improved modeling of land surface phenology using MODIS land surface reflectance and temperature at evergreen needleleaf forests of central North America. Remote Sens. Environ. 176, 152–162.

Liu, Y., Zhang, Y., Shan, N., Zhang, Z., Wei, Z., 2022. Global assessment of partitioning transpiration from evapotranspiration based on satellite solar-induced chlorophyll fluorescence data. J. Hydrol. 612, 128044.

Losos, D., Hoffman, S., Stoy, P.C., 2024a. GOES-R land surface products at Western Hemisphere eddy covariance tower locations. Sci. Data 11, 277. https://doi.org/10.1038/s41597-024-03071-z

Losos, D., Ranjbar, S., Hoffman, S., Abernathey, R., Desai, A.R., Otkin, J.A., Zhang, H., Ryu, Y., Stoy, Paul.C., 2024b. Rapid Changes in Terrestrial Carbon Dioxide Uptake Captured in Near-Real Time from a Geostationary Satellite: The Alive Framework. https://doi.org/10.2139/ssrn.4884876

Lovett, G.M., Jones, C.G., Turner, M.G., Weathers, K.C., 2005. Ecosystem function in heterogeneous landscapes, in: Ecosystem Function in Heterogeneous Landscapes. Springer, pp. 1–4.

Ludwig, G., Johnson, D., 1981. Geostationary operational environmental satellite system performance. Adv. Space Res. 1, 23–31.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Ma, R., Letu, H., Yang, K., Wang, T., Shi, Chong, Xu, J., Shi, J., Shi, Chunxiang, Chen, L., 2020. Estimation of surface shortwave radiation from Himawari-8 satellite data based on a combination of radiative transfer and deep neural network. IEEE Trans. Geosci. Remote Sens. 58, 5304–5316.

Ma, R., Xiao, J., Liang, S., Ma, H., He, T., Guo, D., Liu, X., Lu, H., 2022. Spatial parameter optimization of a terrestrial biosphere model for improving estimation of carbon fluxes for deciduous forests in the eastern United States: an efficient model-data fusion method. Geosci. Model Dev. Discuss. 1–35.

Ma, X., Huete, A., Yu, Q., Coupe, N.R., Davies, K., Broich, M., Ratana, P., Beringer, J., Hutley, L.B., Cleverly, J., 2013. Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect. Remote Sens. Environ. 139, 97–115.

Mace, G.M., Norris, K., Fitter, A.H., 2012. Biodiversity and ecosystem services: a multilayered relationship. Trends Ecol. Evol. 27, 19–26.

Magney, T.S., Frankenberg, C., Fisher, J.B., Sun, Y., North, G.B., Davis, T.S., Kornfeld, A., Siebke, K., 2017. Connecting active to passive fluorescence with photosynthesis: a method for evaluating remote sensing measurements of Chl fluorescence. New Phytol. 215, 1594–1608.

Marshall, M., Tu, K., Brown, J., 2018. Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. Remote Sens. Environ. 217, 258–271.

Martínez, B., Gilabert, M.A., Sánchez-Ruiz, S., Campos-Taberner, M., García-Haro, F.J., Brümmer, C., Carrara, A., Feig, G., Grünwald, T., Mammarella, I., 2020. Evaluation of the LSA-SAF gross primary production product derived from SEVIRI/MSG data (MGPP). ISPRS J. Photogramm. Remote Sens. 159, 220–236.

Martínez, B., Sanchez-Ruiz, S., Gilabert, M., Moreno, A., Campos-Taberner, M., García-Haro, F.J., Trigo, I.F., Aurela, M., Brümmer, C., Carrara, A., 2018. Retrieval of daily gross primary production over Europe and Africa from an ensemble of SEVIRI/MSG products. Int. J. Appl. Earth Obs. Geoinformation 65, 124–136.

Massetti, A., Rüdiger, C., Yebra, M., Hilton, J., 2019. The Vegetation Structure Perpendicular Index (VSPI): A forest condition index for wildfire predictions. Remote Sens. Environ. 224, 167–181.

Matheny, A.M., Bohrer, G., Stoy, P.C., Baker, I.T., Black, A.T., Desai, A.R., Dietze, M.C., Gough, C.M., Ivanov, V.Y., Jassal, R.S., Novick, K.A., Schäfer, K.V.R., Verbeeck, H., 2014. Characterizing the diurnal patterns of errors in the prediction of evapotranspiration by several land-surface models: An NACP analysis. J. Geophys. Res. Biogeosciences 119, 1458–1473. https://doi.org/10.1002/2014JG002623

Mauder, M., Jung, M., Stoy, P., Nelson, J., Wanner, L., 2024. Energy balance closure at FLUXNET sites revisited. Agric. For. Meteorol. 358, 110235.

Mavrovic, A., Sonnentag, O., Lemmetyinen, J., Baltzer, J.L., Kinnard, C., Roy, A., 2023. Reviews and syntheses: Recent advances in microwave remote sensing in support of terrestrial carbon cycle science in Arctic–boreal regions. Biogeosciences 20, 2941–2970. https://doi.org/10.5194/bg-20-2941-2023

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39, 2784–2817.

Mayer, M., Sandén, H., Rewald, B., Godbold, D.L., Katzensteiner, K., 2017. Increase in heterotrophic soil respiration by temperature drives decline in soil organic carbon stocks after forest windthrow in a mountainous ecosystem. Funct. Ecol. 31, 1163–1172.

McColl, K.A., 2020. Practical and Theoretical Benefits of an Alternative to the Penman-Monteith Evapotranspiration Equation. Water Resour. Res. 56, e2020WR027106. https://doi.org/10.1029/2020WR027106

McCorkel, J., Van Naarden, J., Lindsey, D., Efremova, B., Coakley, M., Black, M., Krimchansky, A., 2019. GOES-17 advanced baseline imager performance recovery summary. Presented at the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp. 1–4.

Meerdink, S.K., Hook, S.J., Roberts, D.A., Abbott, E.A., 2019. The ECOSTRESS spectral library version 1.0. Remote Sens. Environ. 230, 111196.

Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B., 2022. OpenET: Filling a critical data gap in water management for the western United States. JAWRA J. Am. Water Resour. Assoc. 58, 971–994.

Menzel, W.P., 2020. History of geostationary weather satellites, in: The GOES-R Series. Elsevier, pp. 5–11.

Menzel, W.P., Purdom, J.F., 1994. Introducing GOES-I: The first of a new generation of geostationary operational environmental satellites. Bull. Am. Meteorol. Soc. 75, 757–782.

Metzger, S., Ayres, E., Durden, D., Florian, C., Lee, R., Lunch, C., Luo, H., Pingintha-Durden, N., Roberti, J.A., SanClements, M., 2019. From NEON field sites to data portal: a community resource for surface–atmosphere research comes online. Bull. Am. Meteorol. Soc. 100, 2305–2325.

Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17, 2145–2151.

Miller, G.R., Chen, X., Rubin, Y., Ma, S., Baldocchi, D.D., 2010. Groundwater uptake by woody vegetation in a semiarid oak savanna. Water Resour. Res. 46, 2009WR008902. https://doi.org/10.1029/2009WR008902

Miller, L., Pelletier, C., Webb, G.I., 2024. Deep Learning for Satellite Image Time-Series Analysis: A review. IEEE Geosci. Remote Sens. Mag. 2–45. https://doi.org/10.1109/MGRS.2024.3393010

Min, L., Fitzjarrald, D.R., Du, Y., Rose, B.E.J., Hong, J., Min, Q., 2021. Exploring Sources of Surface Bias in HRRR Using New York State Mesonet. J. Geophys. Res. Atmospheres 126, e2021JD034989. https://doi.org/10.1029/2021JD034989

Minnis, P., Hong, G., Sun-Mack, S., Smith Jr, W.L., Chen, Y., Miller, S.D., 2016. Estimating nocturnal opaque ice cloud optical depth from MODIS multispectral infrared radiances using a neural network method. J. Geophys. Res. Atmospheres 121, 4907–4932.

Mistry, J., 2000. Savannas. Prog. Phys. Geogr. 24, 601–608.

MIURA, T., NAGAI, S., 2019. Monitoring terrestrial vegetation and the environment with new-generation geostationary satellites. J. Remote Sens. Soc. Jpn. 39, 377–383.

Miura, T., Nagai, S., Takeuchi, M., Ichii, K., Yoshioka, H., 2019. Improved characterisation of vegetation and land surface seasonal dynamics in central Japan with Himawari-8 hypertemporal data. Sci. Rep. 9, 1–12.

Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Beckstein, C., Braswell, B.H., Churkina, G., Desai, A.R., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric. For. Meteorol. 147, 209–232.

Mohapatra, A., Trinh, T., 2022. Early wildfire detection technologies in practice—a review. Sustainability 14, 12270.

Molnar, C., Casalicchio, G., Bischl, B., 2018. iml: An R package for interpretable machine learning. J. Open Source Softw. 3, 786.

Moskolaï, W.R., Abdou, W., Dipanda, A., Kolyang, 2021. Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review. Remote Sens. 13, 4822. https://doi.org/10.3390/rs13234822

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

Muraoka, H., 2022. Phenology of Photosynthesis in a Deciduous Broadleaf Forest: Implications for the Carbon Cycle in a Changing Environment. River Basin Environ. Eval. Manag. Conserv. 3–27.

Nathaniel, J., Liu, J., Gentine, P., 2023. MetaFlux: Meta-learning global carbon fluxes from sparse spatiotemporal observations. Sci. Data 10, 440.

National Oceanic and Atmospheric Administration, 2022. Aerosol optical depth at SURFRAD networks.

Nelson, J.A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R.L., Zhang, W., Hamdi, Z.M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op De Beeck, M., Billdesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J.R., Desai, A., Dong, G., El-Madany, T.S., Euskirchen, E.S., Feigenwinter, I., Galvagno, M., Gerosa, G., Gielen, B., Goded, I., Goslee, S., Gough, C.M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M.A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korkiakoski, M., Mammarella, I., Mana, G., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C.R., Shekhar, A., Silberstein, R., Silveira, M.L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E.R., Wang, Y., Woodgate, W., Yepez, E.A., Zhang, J., Zona, D., Jung, M., 2024. X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X. https://doi.org/10.5194/egusphere-2024-165

Nemani, R., Wang, W., Hashimoto, H., Michaelis, A., Vandal, T., Lyapustin, A., Zhang, J., Lee, T., Kalluri, S., Takenaka, H., 2020. GeoNEX: A Geostationary Earth Observatory at NASA Earth Exchange: Earth Monitoring from Operational Geostationary Satellite Systems. Presented at the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp. 128–131.

Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B., Running, S.W., 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. science 300, 1560–1563.

Newell, H.E., 1965. The NASA Space Science and Applications, in: NASA University Program Review Conference. Scientific and Technical Information Division, National Aeronautics and …, p. 27.

Norman, J., Anderson, M., Kustas, W., French, A., Mecikalski, J., Torn, R., Diak, G., Schmugge, T., Tanner, B., 2003. Remote sensing of surface energy fluxes at 101-m pixel resolutions. Water Resour. Res. 39.

Novick, K.A., Metzger, S., Anderegg, W.R., Barnes, M., Cala, D.S., Guan, K., Hemes, K.S., Hollinger, D.Y., Kumar, J., Litvak, M., 2022. Informing Nature-based Climate Solutions for the United States with the best-available science. Glob. Change Biol.

Novick, K.A., Oren, R., Stoy, P.C., Siqueira, M.B.S., Katul, G.G., 2009. Nocturnal evapotranspiration in eddy-covariance records from three co-located ecosystems in the Southeastern US: implications for annual fluxes. Agric. For. Meteorol. 149, 1491–1504.

Novick, K.A., Stoy, P.C., Katul, G.G., Ellsworth, D.S., Siqueira, M.B.S., Juang, J., Oren, R., 2004. Carbon dioxide and water vapor exchange in a warm temperate grassland. Oecologia 138, 259–274. https://doi.org/10.1007/s00442-003-1388-z

Oliphant, A., Dragoni, D., Deng, B., Grimmond, C., Schmid, H.-P., Scott, S., 2011. The role of sky conditions on gross primary production in a mixed deciduous forest. Agric. For. Meteorol. 151, 781–791.

Oliveira, W.K.M., Maggi, M.F., Venancio, L.P., Bazzi, C.L., Nóbrega, L.H.P., Oliveira, A.S.D.M., Mercante, E., Schenatto, K., 2024. Remote Sensing-Based Algorithms for Estimating Evapotranspiration in Agricultural Systems: A systematic literature review. IEEE Geosci. Remote Sens. Mag.

Osuri, A., Ratnam, J., Varma, V., Alvarez-Loayza, P., Astaiza, J., Bradford, M., Fletcher, C., Ndoundou-Hockemba, M., Jansen, P., Kenfack, D., 2016. Contrasting effects of defaunation on aboveground carbon storage across the global tropics. Nat. Commun. 7, 11351.

Otkin, J.A., Anderson, M.C., Hain, C., Mladenova, I.E., Basara, J.B., Svoboda, M., 2013. Examining rapid onset drought development using the thermal infrared–based evaporative stress index. J. Hydrometeorol. 14, 1057–1074.

Otkin, J.A., Anderson, M.C., Hain, C., Svoboda, M., 2014. Examining the relationship between drought development and rapid changes in the evaporative stress index. J. Hydrometeorol. 15, 938–956.

Otkin, J.A., Svoboda, M., Hunt, E.D., Ford, T.W., Anderson, M.C., Hain, C., Basara, J.B., 2018. Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. Bull. Am. Meteorol. Soc. 99, 911–919.

Pabon-Moreno, D.E., Migliavacca, M., Reichstein, M., Mahecha, M.D., 2022. On the potential of Sentinel-2 for estimating Gross Primary Production. IEEE Trans. Geosci. Remote Sens. 60, 1–12.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrol. Earth Syst. Sci. 24, 1485–1509.

Pan, X., Yang, Z., Liu, Y., Yuan, J., Wang, Z., Liu, S., Yang, Y., 2024. A non-parametric method combined with surface flux equilibrium for estimating terrestrial evapotranspiration: Validation at eddy covariance sites. J. Hydrol. 631, 130682.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences 3, 571–583.

Parente, C., 2013. TOA reflectance and NDVI calculation for Landsat 7 ETM+ images of Sicily. Presented at the Proceedings in Electronic International Interdisciplinary Conference-The 2nd Electronic International Interdisciplinary Conference, EDIS-Publishing Institution of the University of Zilina, pp. 351–354.

Pastorello, G., Agarwal, D., Papale, D., Samak, T., Trotta, C., Ribeca, A., Poindexter, C., Faybishenko, B., Gunter, D., Hollowgrass, R., 2014. Observational data patterns for time series data quality assessment. Presented at the 2014 IEEE 10th International Conference on e-Science, IEEE, pp. 271–278.

Pastorello, G., Papale, D., Chu, H., Trotta, C., Agarwal, D., Canfora, E., Baldocchi, D., Torn, M., 2017. A new data set to keep a sharper eye on land-air exchanges. Eos Trans. Am. Geophys. Union Online 98.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 225.

Paxton, L.J., Zhang, Y., Kil, H., Schaefer, R.K., 2021. Exploring the upper atmosphere: Using optical remote sensing. Up. Atmosphere Dyn. Energ. 487–522.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, Z., Letu, H., Wang, T., Shi, C., Zhao, C., Tana, G., Zhao, N., Dai, T., Tang, R., Shang, H., 2020. Estimation of shortwave solar radiation using the artificial neural network from Himawari-8 satellite imagery over China. J. Quant. Spectrosc. Radiat. Transf. 240, 106672.

Penman, H., 1948. Natural evaporation from open water, bare soil and grass. Proc. R. Soc. Lond. Ser. Math. Phys. Sci. 193, 120–145. https://doi.org/10.1098/rspa.1948.0037

Peres, C.A., Emilio, T., Schietti, J., Desmoulière, S.J., Levi, T., 2016. Dispersal limitation induces long-term biomass collapse in overhunted Amazonian forests. Proc. Natl. Acad. Sci. 113, 892–897.

Peterson, T.C., Golubev, V.S., Groisman, P.Y., 1995. Evaporation losing its strength. Nature 377, 687–688.

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. J. Appl. Ecol. 51, 839–848.

Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A.M., Keith, D.A., Wegmann, M., Schrodt, F., Stellmes, M., 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. Remote Sens. Ecol. Conserv. 4, 71–93.

Qin, L., Yan, C., Yu, L., Chai, M., Wang, B., Hayat, M., Shi, Z., Gao, H., Jiang, X., Xiong, B., 2022. High-resolution spatio-temporal characteristics of urban evapotranspiration measured by unmanned aerial vehicle and infrared remote sensing. Build. Environ. 222, 109389.

Qiu, R., Han, G., Ma, X., Xu, H., Shi, T., Zhang, M., 2020. A comparison of OCO-2 SIF, MODIS GPP, and GOSIF data from gross primary production (GPP) estimation and seasonal cycles in North America. Remote Sens. 12, 258.

Randazzo, N.A., Michalak, A.M., Desai, A.R., 2020. Synoptic Meteorology Explains Temperate

Forest Carbon Uptake. J. Geophys. Res. Biogeosciences 125, e2019JG005476. https://doi.org/10.1029/2019JG005476

Ranjbar, S., Losos, D., Dechant, B., Hoffman, S., Başakın, E.E., Stoy, P.C., 2024a. Harnessing Information From Shortwave Infrared Reflectance Bands to Enhance Satellite-Based Estimates of Gross Primary Productivity. J. Geophys. Res. Biogeosciences 129, e2024JG008240. https://doi.org/10.1029/2024JG008240

Ranjbar, S., Losos, D., Hoffman, S., Arabi, S., Desai, A.R., Stoy, P.C., 2024b. Near Real-time Mapping of All-Sky Land Surface Temperature from GOES-R using Machine Learning. https://doi.org/10.22541/essoar.172801403.30077549/v1

Ranjbar, S., Losos, D., Hoffman, S., Cuntz, M., Stoy, P., 2023a. Geostationary Satellite Observations Can Accurately Estimate Ecosystem Carbon Uptake and Respiration at Half Hourly Time Steps at Eddy Covariance Sites.

Ranjbar, S., Losos, D., Hoffman, S., Cuntz, M., Stoy, P., 2023b. Geostationary Satellite Observations Can Accurately Estimate Ecosystem Carbon Uptake and Respiration at Half Hourly Time Steps at Eddy Covariance Sites. https://dx.doi.org/10.2139/ssrn.4671918

Ranjbar, S., Losos, D., Hoffman, S., Cuntz, M., Stoy, P.C., 2024c. Using Geostationary Satellite Observations and Machine Learning Models to Estimate Ecosystem Carbon Uptake and Respiration at Half Hourly Time Steps at Eddy Covariance Sites. J. Adv. Model. Earth Syst. 16, e2024MS004341. https://doi.org/10.1029/2024MS004341

Ranjbar, S., Losos, D., Hoffman, S., Stoy, P.C., 2024d. High-Frequency Mapping of Downward Shortwave Radiation From GOES-R Using Gradient Boosting. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 11958–11968. https://doi.org/10.1109/JSTARS.2024.3420148

Ranjbar, S., Zarei, A., Hasanlou, M., Akhoondzadeh, M., Amini, J., Amani, M., 2021. Machine learning inversion approach for soil parameters estimation over vegetated agricultural areas using a combination of water cloud model and calibrated integral equation model. J. Appl. Remote Sens. 15, 018503–018503.

Rao, P.K., Holmes, S.J., Anderson, R.K., Winston, J.S., Lehr, P.E., 1990. Weather satellites: Systems, data, and environmental applications.

Rayleigh, Lord, 1899. XXXIV. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. Lond. Edinb. Dublin Philos. Mag. J. Sci. 47, 375–384.

Reddy, D.S., Prasad, P.R.C., 2018. Prediction of vegetation dynamics using NDVI time series data and LSTM. Model. Earth Syst. Environ. 4, 409–419.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., 2013. Climate extremes and the carbon cycle. Nature 500, 287–295.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Change Biol. 11, 1424–1439.

Reichstein, M., Stoy, P.C., Desai, A.R., Lasslop, G., Richardson, A.D., 2012a. Partitioning of net fluxes. Eddy Covariance Pract. Guide Meas. Data Anal. 263–289.

Reichstein, M., Stoy, P.C., Desai, A.R., Lasslop, G., Richardson, A.D., 2012b. Partitioning of net fluxes. Eddy Covariance Pract. Guide Meas. Data Anal. 263–289.

Ren, H., Liu, R., Qin, Q., Fan, W., Yu, L., Du, C., 2017. Mapping finer-resolution land surface emissivity using Landsat images in China. J. Geophys. Res. Atmospheres 122, 6764–6781. https://doi.org/10.1002/2017JD026910

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156–173.

Robledano, A., Picard, G., Arnaud, L., Larue, F., Ollivier, I., 2022. Modelling surface temperature and radiation budget of snow-covered complex terrain. The Cryosphere 16, 559–579. https://doi.org/10.5194/tc-16-559-2022

Running, S.W., Baldocchi, D., Turner, D., Gower, S.T., Bakwin, P., Hibbard, K., 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. Remote Sens. Environ. 70, 108–127.

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience 54, 547–560.

Running, S.W., Zhao, M., 2015. Daily GPP and annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS land algorithm. MOD17 User's Guide 2015, 1–28.

Ryu, Y., Baldocchi, D.D., Ma, S., Hehn, T., 2008. Interannual variability of evapotranspiration and energy exchange over an annual grassland in California. J. Geophys. Res. Atmospheres 113, 2007JD009263. https://doi.org/10.1029/2007JD009263

Sabir, R.M., Sarwar, A., Shoaib, M., Saleem, A., Alhousain, M.H., Wajid, S.A., Rasul, F., Adnan Shahid, M., Anjum, L., Safdar, M., Muhammad, N.E., Waqas, R.M., Zafar, U., Raza, A., 2024. Managing Water Resources for Sustainable Agricultural Production, in: Kanga, S., Singh, S.K., Shevkani, K., Pathak, V., Sajan, B. (Eds.), Transforming Agricultural

Management for a Sustainable Future, World Sustainability Series. Springer Nature Switzerland, Cham, pp. 47–74. https://doi.org/10.1007/978-3-031-63430-7_3

Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8, e1249.

Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Appl. Sci. 2, 1308.

Saini, R., Ghosh, S.K., 2017. Ensemble classifiers in remote sensing: A review. Presented at the 2017 International Conference on Computing, Communication and Automation (ICCCA), IEEE, pp. 1148–1152.

Sanchez-Ruiz, S., Chiesi, M., Fibbi, L., Carrara, A., Maselli, F., Gilabert, M.A., 2018. Optimized Application of Biome-BGC for Modeling the Daily GPP of Natural Vegetation Over Peninsular Spain. J. Geophys. Res. Biogeosciences 123, 531–546. https://doi.org/10.1002/2017JG004360

Sanchez-Ruiz, S., Moreno, A., Piles, M., Maselli, F., Carrara, A., Running, S., Gilabert, M.A., 2017. Quantifying water stress effect on daily light use efficiency in Mediterranean ecosystems using satellite data. Int. J. Digit. Earth 10, 623–638. https://doi.org/10.1080/17538947.2016.1247301

Schimel, D., Pavlick, R., Fisher, J.B., Asner, G.P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., Cox, P., 2015. Observing terrestrial ecosystems and the carbon cycle from space. Glob. Change Biol. 21, 1762–1776.

Schmit, T.J., Griffith, P., Gunshor, M.M., Daniels, J.M., Goodman, S.J., Lebair, W.J., 2017a. A closer look at the ABI on the GOES-R series. Bull. Am. Meteorol. Soc. 98, 681–698.

Schmit, T.J., Griffith, P., Gunshor, M.M., Daniels, J.M., Goodman, S.J., Lebair, W.J., 2017b. A closer look at the ABI on the GOES-R series. Bull. Am. Meteorol. Soc. 98, 681–698.

Schmit, T.J., Gunshor, M.M., 2020. ABI imagery from the GOES-R series, in: The GOES-R Series. Elsevier, pp. 23–34.

Schmit, T.J., Li, Z., Gunshor, M.M., Iturbide-Iturbide, F., Yoe, J.G., McCorkel, J., Heidinger, A., 2022. US Plans for Geostationary Hyperspectral Infrared Sounders. Presented at the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp. 5411–5414.

Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: An ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4, 6–23.

Schroer, R., 2008. Satellite communications (SATCOMs)[Part Two, NASA at 50]. IEEE Aerosp. Electron. Syst. Mag. 23, 32–37.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45, 2673–2681.

Shen, Y., Zhang, X., Wang, W., Nemani, R., Ye, Y., Wang, J., 2021. Fusing geostationary satellite observations with Harmonized Landsat-8 and Sentinel-2 time series for monitoring field-scale land surface phenology. Remote Sens. 13, 4465.

Shen, Z., Yong, B., 2021. Downscaling the GPM-based satellite precipitation retrievals using gradient boosting decision tree approach over Mainland China. J. Hydrol. 602, 126803.

Shirmard, H., Farahbakhsh, E., Heidari, E., Beiranvand Pour, A., Pradhan, B., Müller, D., Chandra, R., 2022. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. Remote Sens. 14, 819.

Shwetha, H.R., Kumar, D.N., 2016. Prediction of high spatio-temporal resolution land surface temperature under cloudy conditions using microwave vegetation index and ANN. ISPRS J. Photogramm. Remote Sens. 117, 40–55. https://doi.org/10.1016/j.isprsjprs.2016.03.011

Simran, K., Sriram, S., Vinayakumar, R., Soman, K., 2020. Deep learning approach for intelligent named entity recognition of cyber security. Presented at the Advances in Signal Processing and Intelligent Recognition Systems: 5th International Symposium, SIRS 2019, Trivandrum, India, December 18–21, 2019, Revised Selected Papers 5, Springer, pp. 163–172.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Snyder, W.C., Wan, Z., Zhang, Y., Feng, Y.-Z., 1998. Classification-based emissivity for land surface temperature measurement from space. Int. J. Remote Sens. 19, 2753–2774. https://doi.org/10.1080/014311698214497

Sobrino, J.A., Jimenez-Munoz, J.C., Soria, G., Romaguera, M., Guanter, L., Moreno, J., Plaza, A., Martinez, P., 2008. Land Surface Emissivity Retrieval From Different VNIR and TIR Sensors. IEEE Trans. Geosci. Remote Sens. 46, 316–327. https://doi.org/10.1109/TGRS.2007.904834

Sobrino, J.A., Jiménez-Muñoz, J.C., Sòria, G., Ruescas, A.B., Danne, O., Brockmann, C., Ghent, D., Remedios, J., North, P., Merchant, C., Berger, M., Mathieu, P.P., Göttsche, F.-M., 2016. Synergistic use of MERIS and AATSR as a proxy for estimating Land Surface Temperature from Sentinel-3 data. Remote Sens. Environ. 179, 149–161. https://doi.org/10.1016/j.rse.2016.03.035

Sokolik, I., Soja, A., DeMott, P., Winker, D., 2019. Progress and challenges in quantifying wildfire smoke emissions, their properties, transport, and atmospheric impacts. J. Geophys. Res. Atmospheres 124, 13005–13025.

Srivastava, D.S., Vellend, M., 2005. Biodiversity-ecosystem function research: is it relevant to conservation? Annu Rev Ecol Evol Syst 36, 267–294.

Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M.A., Arneth, A., Aurela, M., Bernhofer, C., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. Agric. For. Meteorol. 171, 137–152.

Stoyanova, J., Georgiev, C., 2010. Drought and vegetation fires detection using MSG geostationary satellites. Presented at the Proceedings of the 2010 EUMETSAT Meteorological Satellite Conference, Córdoba, Spain, pp. 20–24.

Straubel, M.S., 1992. Telecommunication Satellites and Market Forces: How Should the Geostationary Orbit Be Regulated by the FCC. NCJ Intl Com Reg 17, 205.

Sturtevant, C., DeRego, E., Metzger, S., Ayres, E., Allen, D., Burlingame, T., Catolico, N., Cawley, K., Csavina, J., Durden, D., Florian, C., Frost, S., Gaddie, R., Knapp, E., Laney, C., Lee, R., Lenz, D., Litt, G., Luo, H., Roberti, J., Slemmons, C., Styers, K., Tran, C., Vance, T., SanClements, M., 2022. A process approach to quality management doubles NEON sensor data quality. Methods Ecol. Evol. 13, 1849–1865. https://doi.org/10.1111/2041-210X.13943

Suliman, A., Zhang, Y., 2015. A review on back-propagation neural networks in the application of remote sensing image classification. J. Earth Sci. Eng. 5, 52–65.

Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P., Friedl, M.A., 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product.

Remote Sens. Environ. 222, 183–194.

Sun, D., Li, Y., Zhan, X., Houser, P., Yang, C., Chiu, L., Yang, R., 2019. Land Surface Temperature Derivation under All Sky Conditions through Integrating AMSR-E/AMSR-2 and MODIS/GOES Observations. Remote Sens. 11, 1704. https://doi.org/10.3390/rs11141704

Sun, J., Salvucci, G.D., Entekhabi, D., 2012. Estimates of evapotranspiration from MODIS and AMSR-E land surface temperature and moisture over the Southern Great Plains. Remote Sens. Environ. 127, 44–59. https://doi.org/10.1016/j.rse.2012.08.020

Sun, L., Anderson, M.C., Gao, F., Hain, C., Alfieri, J.G., Sharifi, A., McCarty, G.W., Yang, Yun, Yang, Yang, Kustas, W.P., McKee, L., 2017. Investigating water use over the C hoptank R iver W atershed using a multisatellite data fusion approach. Water Resour. Res. 53, 5298–5319. https://doi.org/10.1002/2017WR020700

Sun, Z., Wang, X., Zhang, X., Tani, H., Guo, E., Yin, S., Zhang, T., 2019. Evaluating and comparing remote sensing terrestrial GPP models for their response to climate variability and CO2 trends. Sci. Total Environ. 668, 696–713.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 27.

Tabari, H., Aeini, A., Talaee, P.H., Some'e, B.S., 2012. Spatial distribution and temporal variation of reference evapotranspiration in arid and semi-arid regions of Iran. Hydrol. Process. 26, 500–512. https://doi.org/10.1002/hyp.8146

Talib, A., Desai, A.R., Huang, J., Griffis, T.J., Reed, D.E., Chen, J., 2021. Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the Midwest US. J. Hydrol. 600, 126579.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. 164, 152–170.

Tan, W., He, H., Chen, X., Qi, W., 2020. Analyzing the influence of atmosphere on optical remote sensing in 400 to 2500 nm wavelength spectrum. Presented at the AOPC 2020: Optical Spectroscopy and Imaging; and Biomedical Optics, SPIE, pp. 103–108.

Tang, Y., Wang, Q., Atkinson, P.M., 2023. Filling Then Spatio-Temporal Fusion for All-Sky MODIS Land Surface Temperature Generation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 1350–1364. https://doi.org/10.1109/JSTARS.2023.3235940

Tateishi, R., Ahn, C.H., 1996. Mapping evapotranspiration and water balance for global land surfaces. ISPRS J. Photogramm. Remote Sens. 51, 209–215.

Temenos, A., Temenos, N., Kaselimi, M., Doulamis, A., Doulamis, N., 2023. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing Using SHAP. IEEE Geosci. Remote Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3251652

Thakur, G., Schymanski, S.J., Mallick, K., Trebs, I., Sulis, M., 2022. Downwelling longwave radiation and sensible heat flux observations are critical for surface temperature and emissivity estimation from flux tower data. Sci. Rep. 12, 8592. https://doi.org/10.1038/s41598-022-12304-3

Thapa, B., Lovell, S., Wilson, J., 2023. Remote sensing and machine learning applications for aboveground biomass estimation in agroforestry systems: a review. Agrofor. Syst. 1–15.

Thiebault, K., Young, S., 2020. Snow cover change and its relationship with land surface temperature and vegetation in northeastern North America from 2000 to 2017. Int. J.

Remote Sens. 41, 8453–8474. https://doi.org/10.1080/01431161.2020.1779379

Tian, J., Philpot, W.D., 2015. Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra. Remote Sens. Environ. 169, 280–289.

Todd, S.W., Hoffer, R.M., 1998. Responses of spectral indices to variations in vegetation cover and soil background. Photogramm. Eng. Remote Sens. 64, 915–922.

Tucker, C., Fung, I.Y., Keeling, C., Gammon, R., 1986. Relationship between atmospheric CO2 variations and a satellite-derived vegetation index. Nature 319, 195–199.

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M.H., Kirschbaum, A.A., Ham, J.M., Saleska, S.R., 2006. Evaluation of MODIS NPP and GPP products across multiple biomes. Remote Sens. Environ. 102, 282–292.

Valor, E., 1996. Mapping land surface emissivity from NDVI: Application to European, African, and South American areas. Remote Sens. Environ. 57, 167–184. https://doi.org/10.1016/0034-4257(96)00039-9

Van Der Ent, R.J., Savenije, H.H.G., Schaefli, B., Steele-Dunne, S.C., 2010. Origin and fate of atmospheric moisture over continents. Water Resour. Res. 46, 2010WR009127. https://doi.org/10.1029/2010WR009127

Vandal, T.J., McDuff, D., Wang, W., Duffy, K., Michaelis, A., Nemani, R.R., 2021. Spectral synthesis for geostationary satellite-to-satellite translation. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Varghese, R., Cherukuri, A.K., Doddrell, N.H., Doss, C.G.P., Simkin, A.J., Ramamoorthy, S., 2023. Machine learning in photosynthesis: Prospects on sustainable crop development. Plant Sci. 335, 111795.

Vekuri, H., Tuovinen, J.-P., Kulmala, L., Papale, D., Kolari, P., Aurela, M., Laurila, T., Liski, J., Lohila, A., 2023. A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates. Sci. Rep. 13, 1720.

Verhoef, W., Van Der Tol, C., Middleton, E.M., 2018. Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX–Sentinel-3 tandem mission multi-sensor data. Remote Sens. Environ. 204, 942–963.

Verma, M., Fisher, J., Mallick, K., Ryu, Y., Kobayashi, H., Guillaume, A., Moore, G., Ramakrishnan, L., Hendrix, V., Wolf, S., Sikka, M., Kiely, G., Wohlfahrt, G., Gielen, B., Roupsard, O., Toscano, P., Arain, A., Cescatti, A., 2016. Global Surface Net-Radiation at 5 km from MODIS Terra. Remote Sens. 8, 739. https://doi.org/10.3390/rs8090739

Veroustraete, F., Patyn, J., Myneni, R., 1996. Estimating net ecosystem exchange of carbon using the normalized difference vegetation index and an ecosystem model. Remote Sens. Environ. 58, 115–130.

Walker, E., García, G.A., Venturini, V., Carrasco, A., 2019. Regional evapotranspiration estimates using the relative soil moisture ratio derived from SMAP products. Agric. Water Manag. 216, 254–263. https://doi.org/10.1016/j.agwat.2019.02.009

Wang, A., Zhang, M., Chen, E., Zhang, C., Han, Y., 2024. Impact of seasonal global land surface temperature (LST) change on gross primary production (GPP) in the early 21st century. Sustain. Cities Soc. 110, 105572. https://doi.org/10.1016/j.scs.2024.105572

Wang, D., Li, R., 2022. A GeoNEX-based 1km hourly land surface downward shortwave radiation (DSR) and photosynthetically active radiation (PAR) product, Zenodo [data set]. Zenodo Data Set 10.

Wang, L., Tian, F., Huang, K., Wang, Y., Wu, Z., Fensholt, R., 2020. Asymmetric patterns and temporal changes in phenology-based seasonal gross carbon uptake of global terrestrial ecosystems. Glob. Ecol. Biogeogr. 29, 1020–1033.

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., Du, Q., 2017. Evaluation of the latest MODIS GPP products across multiple biomes using global eddy covariance flux data. Remote Sens. 9, 418.

Wang, M., Sun, R., Zhu, A., Xiao, Z., 2020. Evaluation and comparison of light use efficiency and gross primary productivity using three different approaches. Remote Sens. 12, 1003.

Wang, W., Li, S., Hashimoto, H., Takenaka, H., Higuchi, A., Kalluri, S., Nemani, R., 2020. An introduction to the Geostationary-NASA Earth Exchange (GeoNEX) Products: 1. Top-of-atmosphere reflectance and brightness temperature. Remote Sens. 12, 1267.

Wang, X., Zhong, L., Ma, Y., Fu, Y., Han, C., Li, P., Wang, Z., Qi, Y., 2023. Estimation of hourly actual evapotranspiration over the Tibetan Plateau from multi-source data. Atmospheric Res. 281, 106475.

Wang, Y., Wang, X., He, H., Cao, C., 2020. Modeling and simulation techniques of cloud radiation characteristics for space-based remote sensing, in: Comerón, A., Kassianov, E.I., Schäfer, K., Picard, R.H., Weber, K., Singh, U.N. (Eds.), Remote Sensing of Clouds and the Atmosphere XXV. Presented at the Remote Sensing of Clouds and the Atmosphere XXV, SPIE, Online Only, United Kingdom, p. 8. https://doi.org/10.1117/12.2573689

Wanniarachchi, S., Sarukkalige, R., 2022. A review on evapotranspiration estimation in agricultural water management: Past, present, and future. Hydrology 9, 123.

Warren, S.G., 2019. Optical properties of ice and snow. Philos. Trans. R. Soc. Math. Phys. Eng.

Sci. 377, 20180161. https://doi.org/10.1098/rsta.2018.0161

Wei, S., Yi, C., Fang, W., Hendrey, G., 2017. A global study of GPP focusing on light-use efficiency in a random forest regression model. Ecosphere 8, e01724.

Weiß, M., Menzel, L., 2008. A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. Adv. Geosci. 18, 15–23.

Weltzin, J.F., Betancourt, J.L., Cook, B.I., Crimmins, T.M., Enquist, C.A., Gerst, M.D., Gross, J.E., Henebry, G.M., Hufft, R.A., Kenney, M.A., 2020. Seasonality of biological and physical systems as indicators of climatic variation and change. Clim. Change 163, 1755–1771.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. ISPRS J. Photogramm. Remote Sens. 64, 335–344. https://doi.org/10.1016/j.isprsjprs.2009.03.007

Wheeler, K.I., Dietze, M.C., 2021. Improving the monitoring of deciduous broadleaf phenology using the Geostationary Operational Environmental Satellite (GOES) 16 and 17. Biogeosciences 18, 1971–1985.

Wheeler, K.I., Dietze, M.C., 2019. A statistical model for estimating midday NDVI from the geostationary operational environmental satellite (GOES) 16 and 17. Remote Sens. 11, 2507.

White, E.R., Hastings, A., 2020. Seasonality in ecology: Progress and prospects in theory. Ecol. Complex. 44, 100867.

Wilber, A.C., 1999. Surface Emissivity Maps for Use in Satellite Retrievals of Longwave Radiation, NASA technical paper. NASA.

Williamson, S.N., Hik, D.S., Gamon, J.A., Jarosch, A.H., Anslow, F.S., Clarke, G.K.C., Scott Rupp, T., 2017. Spring and summer monthly MODIS LST is inherently biased compared to air temperature in snow covered sub-Arctic mountains. Remote Sens. Environ. 189, 14–24. https://doi.org/10.1016/j.rse.2016.11.009

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol. 113, 223–243.

Wohlfahrt, G., Pilloni, S., Hörtnagl, L., Hammerle, A., 2010. Estimating carbon dioxide fluxes from temperate mountain grasslands using broad-band vegetation indices. Biogeosciences 7, 683–694.

Wong, A., Yang, Z., Zhou, H., Huang, Y., Han, M., Lai, J.-H., 2023. Knowledge-Guided Recurrent Neural Networks for Monthly Forest Carbon Uptake Estimation. Presented at the AAAI Workshop.

Wu, C., Niu, Z., Gao, S., 2010. Gross primary production estimation from MODIS data with vegetation index and photosynthetically active radiation in maize. J. Geophys. Res. Atmospheres 115, 2009JD013023. https://doi.org/10.1029/2009JD013023

Wu, P., Su, Y., Duan, S., Li, X., Yang, H., Zeng, C., Ma, X., Wu, Y., Shen, H., 2022. A two-step deep learning framework for mapping gapless all-weather land surface temperature using thermal infrared and passive microwave data. Remote Sens. Environ. 277, 113070. https://doi.org/10.1016/j.rse.2022.113070

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., 2019. Current status of Landsat program, science, and applications. Remote Sens. Environ. 225, 127–147.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15, 5015–5030.

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., 2019a. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. Remote Sens. Environ. 233, 111383.

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., 2019b. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. Remote Sens. Environ. 233, 111383.

Xiao, J., Fisher, J.B., Hashimoto, H., Ichii, K., Parazoo, N.C., 2021. Emerging satellite observations for diurnal cycling of ecosystem processes. Nat. Plants 7, 877–887.

Xu, G., Zhong, X., 2017. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. Remote Sens. Lett. 8, 1052–1061.

Xu, W., Wooster, M.J., Kaneko, T., He, J., Zhang, T., Fisher, D., 2017. Major advances in geostationary fire radiative power (FRP) retrieval over Asia and Australia stemming from use of Himarawi-8 AHI. Remote Sens. Environ. 193, 138–149.

Yan, J., Ni, L., Li, X., Cheng, Y., Wu, H., 2023. A framework for reconstructing 1km all-weather hourly LST from MODIS data. Int. J. Remote Sens. 1–24. https://doi.org/10.1080/01431161.2023.2242591

Yang, F., Ichii, K., White, M.A., Hashimoto, H., Michaelis, A.R., Votava, P., Zhu, A.-X., Huete, A., Running, S.W., Nemani, R.R., 2007. Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach. Remote Sens. Environ. 110, 109–122.

Yang, G., Pu, R., Zhao, C., Huang, W., Wang, J., 2011. Estimation of subpixel land surface temperature using an endmember index based technique: A case examination on ASTER and MODIS temperature products over a heterogeneous area. Remote Sens. Environ. 115, 1202–1219.

Yang, X., Ren, L., Jiao, D., Yong, B., Jiang, S., Song, S., 2013. Estimation of Daily Actual Evapotranspiration from ETM+ and MODIS Data of the Headwaters of the West Liaohe Basin in the Semiarid Regions of China. J. Hydrol. Eng. 18, 1530–1538. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000537

Yang, Y., Sun, W., Chi, Y., Yan, X., Fan, H., Yang, X., Ma, Z., Wang, Q., Zhao, C., 2022. Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data. Remote Sens. Environ. 273, 112971.

Ye, P., 2022. Remote Sensing Approaches for Meteorological Disaster Monitoring: Recent Achievements and New Challenges. Int. J. Environ. Res. Public. Health 19, 3701.

Ye, Q., Yu, S., Liu, J., Zhao, Q., Zhao, Z., 2021. Aboveground biomass estimation of black locust planted forests with aspect variable using machine learning regression algorithms. Ecol. Indic. 129, 107948.

Yeom, J.-M., Park, S., Chae, T., Kim, J.-Y., Lee, C.S., 2019. Spatial assessment of solar radiation by machine learning and deep neural network models using data provided by the COMS MI geostationary satellite: A case study in South Korea. Sensors 19, 2082.

Younes, N., Northfield, T.D., Joyce, K.E., Maier, S.W., Duke, N.C., Lymburner, L., 2020. A novel approach to modelling mangrove phenology from satellite images: A case study from northern Australia. Remote Sens. 12, 4008.

Yu, T., Zhang, Q., Sun, R., 2021. Comparison of machine learning methods to up-scale gross primary production. Remote Sens. 13, 2448.

Yu, Y., Tarpley, D., Privette, J.L., Flynn, L.E., Xu, H., Chen, M., Vinnikov, K.Y., Sun, D., Tian, Y., 2011. Validation of GOES-R satellite land surface temperature algorithm using SURFRAD ground measurements and statistical estimates of error properties. IEEE Trans. Geosci. Remote Sens. 50, 704–713.

Yu, Y., Tarpley, D., Privette, J.L., Goldberg, M.D., Raja, M.R.V., Vinnikov, K.Y., Xu, H., 2008. Developing algorithm for operational GOES-R land surface temperature product. IEEE Trans. Geosci. Remote Sens. 47, 936–951.

Yu, Y., Yu, P., 2020a. Land Surface Temperature Product from the GOES-R Series, in: The GOES-R Series. Elsevier, pp. 133–144. https://doi.org/10.1016/B978-0-12-814327-8.00012-3

Yu, Y., Yu, P., 2020b. Land surface temperature product from the GOES-R series, in: The GOES-R Series. Elsevier, pp. 133–144.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.

Yuanyong, D., Shenghui, F., 2013. Simulation analysis of vegetation TOA reflectance based on coupled leaf-canopy-atmosphere radiative transfer model. Remote Sens. Nat. Resour. 25, 30–37.

Zarei, A., Shah-Hosseini, R., Ranjbar, S., Hasanlou, M., 2021. Validation of non-linear split window algorithm for land surface temperature estimation using Sentinel-3 satellite imagery: Case study; Tehran Province, Iran. Adv. Space Res. 67, 3979–3993. https://doi.org/10.1016/j.asr.2021.02.019

Zee, C.-H., 2013. Theory of geostationary satellites. Springer Science & Business Media.

Zeng, Z., Amin, M.G., Shan, T., 2020. Arm motion classification using time-series analysis of the spectrogram frequency envelopes. Remote Sens. 12, 454.

Zhan, W., Chen, Y., Zhou, J., Wang, J., Liu, W., Voogt, J., Zhu, X., Quan, J., Li, J., 2013. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. Remote Sens. Environ. 131, 119–139. https://doi.org/10.1016/j.rse.2012.12.014

Zhang, H., Dong, X., Xi, B., Xin, X., Liu, Q., He, H., Xie, X., Li, L., Yu, S., 2021a. Retrieving high-resolution surface photosynthetically active radiation from the MODIS and GOES-16 ABI data. Remote Sens. Environ. 260, 112436.

Zhang, H., Dong, X., Xi, B., Xin, X., Liu, Q., He, H., Xie, X., Li, L., Yu, S., 2021b. Retrieving high-resolution surface photosynthetically active radiation from the MODIS and GOES-16 ABI data. Remote Sens. Environ. 260, 112436.

Zhang, H., Kondragunta, S., Laszlo, I., Zhou, M., 2020. Improving GOES Advanced Baseline Imager (ABI) aerosol optical depth (AOD) retrievals using an empirical bias correction algorithm. Atmospheric Meas. Tech. 13, 5955–5975.

Zhang, H., Tang, B.-H., Li, Z.-L., 2024. A practical two-step framework for all-sky land surface temperature estimation. Remote Sens. Environ. 303, 113991. https://doi.org/10.1016/j.rse.2024.113991

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. Int. J. Image Data Fusion 1, 5–24.

Zhang, X., Friedl, M.A., Schaaf, C.B., 2006. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. J. Geophys. Res. Biogeosciences 111.

Zhang, Y., Liang, S., He, T., Wang, D., Yu, Y., Ma, H., 2020. Estimation of land surface incident shortwave radiation from geostationary advanced Himawari imager and advanced baseline imager observations using an optimization method. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Zhang, Y., Liu, J., Shen, W., 2022a. A review of ensemble learning algorithms used in remote sensing applications. Appl. Sci. 12, 8654.

Zhang, Y., Liu, J., Shen, W., 2022b. A review of ensemble learning algorithms used in remote sensing applications. Appl. Sci. 12, 8654.

Zhang, Y., Song, C., Sun, G., Band, L.E., McNulty, S., Noormets, A., Zhang, Q., Zhang, Z., 2016. Development of a coupled carbon and water model for estimating global gross primary productivity and evapotranspiration based on eddy flux and remote sensing data. Agric. For. Meteorol. 223, 116–131.

Zhang, Y., Yang, Y., Pan, X., Ding, Y., Hu, J., Dai, Y., 2023. Multiinformation Fusion Network for Mapping Gapless All-Sky Land Surface Temperature Using Thermal Infrared and Reanalysis Data. IEEE Trans. Geosci. Remote Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3269622

Zhao, S., Wu, L., Xiang, Y., Dong, J., Li, Z., Liu, X., Tang, Z., Wang, H., Wang, X., An, J., 2022. Coupling meteorological stations data and satellite data for prediction of global solar radiation with machine learning models. Renew. Energy 198, 1049–1064.

Zhao, W., Duan, S.-B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. Remote Sens. Environ. 247, 111931.

https://doi.org/10.1016/j.rse.2020.111931

Zheng, C., Jia, L., Hu, G., 2022. Global land surface evapotranspiration monitoring by ETMonitor model driven by multi-source satellite earth observations. J. Hydrol. 613, 128444. https://doi.org/10.1016/j.jhydrol.2022.128444

Zheng, M., Zhang, J., Wang, J., Yang, S., Han, J., Hassan, T., 2022. Reconstruction of 0.05° all-sky daily maximum air temperature across Eurasia for 2003–2018 with multi-source satellite data and machine learning models. Atmospheric Res. 279, 106398. https://doi.org/10.1016/j.atmosres.2022.106398

Zhou, R., Wang, H., Duan, K., Liu, B., 2021. Diverse responses of vegetation to hydroclimate across temporal scales in a humid subtropical region. J. Hydrol. Reg. Stud. 33, 100775.

Zhou, S., Cheng, J., Shi, J., 2022. A Physical-Based Framework for Estimating the Hourly All-Weather Land Surface Temperature by Synchronizing Geostationary Satellite Observations and Land Surface Model Simulations. IEEE Trans. Geosci. Remote Sens. 60, 1–22. https://doi.org/10.1109/TGRS.2022.3222563

Zhu, S., Clement, R., McCalmont, J., Davies, C.A., Hill, T., 2022. Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes. Agric. For. Meteorol. 314, 108777.

Zhu, S., Xu, J., Zeng, J., Feng, X., Wang, Y., Bao, S., Shi, J., 2023. Explainable machine learning confirms the global terrestrial CO 2 fertilisation effect from space. IEEE Geosci. Remote Sens. Lett.

Zhu, X., 2014. Generating High-Quality Landsat Time-Series and Its Applications in Forest Studies.

Zhu, Z., Piao, S., Myneni, R.B., Huang, M., Zeng, Z., Canadell, J.G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., 2016. Greening of the Earth and its drivers. Nat. Clim. Change 6, 791–795.

Zscheischler, J., Mahecha, M.D., Von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J.T., Schölkopf, B., Seneviratne, S.I., Tomelleri, E., 2014. A few extreme events dominate global interannual variability in gross primary production. Environ. Res. Lett. 9, 035001.