

ABSTRACT

Title of Dissertation: ARTIFICIAL INTELLIGENCE-DRIVEN
INTEGRATION OF SATELLITE, IN-SITU, AND
SIMULATION DATA TO ENHANCE UNI-
TEMPORAL AND MULTI-TEMPORAL
VARIABLE ESTIMATION

Shuo Xu, Doctor of Philosophy, 2026

Dissertation directed by: Professor, Yiqun Xie, Department of Geographical
Sciences

Reliable estimation of environmental variables is critical for understanding land–atmosphere interactions, ecosystem functioning, and the global carbon cycle, and for supporting applications such as natural hazard prediction, resource management, and sustainable development. However, current environmental datasets derived from satellite observations, in-situ measurements, and process-based ecosystem models have inherent limitations. Satellite observations provide broad spatial coverage but may be affected by observational constraints and the indirect nature of satellite measurements. In-situ observations provide highly accurate measurements but are spatially sparse and often temporally incomplete. Process-based models simulate ecosystem dynamics with physical interpretability but often depend on simplified assumptions. Therefore, effectively integrating these complementary data sources remains a key challenge for Earth system monitoring.

This dissertation develops artificial intelligence (AI)-driven integration frameworks that combine satellite, in-situ, and simulation data, particularly through knowledge-guided machine learning (KGML), to improve environmental variable estimation. The proposed frameworks address both uni-temporal and multi-temporal variable estimation tasks, with two representative case studies: land surface temperature (LST) and gross primary productivity (GPP), where LST serves as a concrete example of uni-temporal (snapshot) variable estimation, while GPP serves as a concrete example of multi-temporal (time-series) variable estimation.

The first study conducts a systematic evaluation of existing satellite-based LST products and ecosystem model GPP simulations using satellite observations, model outputs, and in-situ measurements. The analysis identifies key limitations in current environmental datasets, including the strong dependence of LST retrieval on surface properties and the sensitivity of GPP estimation to vegetation composition.

The second study applies a machine learning (ML)-based framework for improving snapshot retrieval of LST. A transfer learning (TL) approach integrates radiative transfer simulations (RTMs) with real-world observations. By combining shortwave and longwave top-of-atmosphere (TOA) observations, the proposed TL method reduces reliance on external emissivity inputs and improves retrieval accuracy compared with conventional single-channel (SC) and split-window (SW) algorithms.

The third study proposes a KGML framework for time-series estimation of GPP. The framework introduces a decomposition-and-resembling (DERE) strategy to link process-based ecosystem simulations with satellite-derived vegetation composition observations. In addition, a probabilistic label expansion module based on diffusion modeling is introduced to address the limited temporal coverage of flux tower observations. Experiments across multiple observational networks

demonstrate that the proposed DERE framework improves estimation performance for GPP compared with conventional baseline and KGML algorithms.

Overall, results show that the proposed methods consistently improve estimation accuracy compared with existing approaches across both tasks, highlighting the effectiveness of AI-driven and knowledge-guided data integration. By integrating satellite, in-situ, and simulation data, the proposed approaches improve environmental variable estimation and contribute to more reliable Earth system monitoring.

ARTIFICIAL INTELLIGENCE-DRIVEN INTEGRATION OF SATELLITE, IN-SITU, AND
SIMULATION DATA TO ENHANCE UNI-TEMPORAL AND MULTI-TEMPORAL
VARIABLE ESTIMATION

by

Shuo Xu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:

Professor Yiqun Xie, Chair
Professor George Hurtt
Professor Leila De Floriani
Professor Xianfeng Yang
Dr. Yuling Liu

© Copyright by
Shuo Xu
2026

Dedication

To my family, with love.

Acknowledgements

This dissertation marks the end of my PhD journey, a path shaped by challenges, growth, and many unexpected turns. Looking back, I realize how much I have learned—not only in research, but also in life.

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Yiqun Xie, for his guidance and support throughout my PhD journey. After joining his group, he introduced me to new research directions, particularly in artificial intelligence. He was always patient and supportive, offering constant encouragement and demonstrating strong confidence in my abilities, which greatly strengthened my self-confidence and helped me grow into a more independent researcher. Under his mentorship, I learned many important lessons about both research and academic life. I would also like to thank my former advisor, Dr. Dongdong Wang, who supervised my research during the early years of my PhD until his retirement. Under his guidance, I developed a solid foundation in remote sensing research.

I would also like to sincerely thank my research assistantship supervisor, Dr. Yuling Liu, with whom I worked closely during my first four years at the University of Maryland on the NOAA project focused on improving VIIRS Land Surface Temperature (LST) products. I learned a great deal from this collaboration, and her encouragement always made me feel supported.

I am deeply grateful to my other dissertation committee members: Professor Leila De Floriani, Professor George Hurtt, and Professor Xianfeng Yang. Professor De Floriani has been my committee member since I first arrived at UMD and has witnessed my academic growth throughout this journey. I truly appreciate her encouragement and guidance. I also thank Professor Hurtt for introducing me to the field of carbon monitoring and for his valuable guidance that

broadened my perspective. I am also grateful to Professor Yang for his insightful advice and support in the areas of machine learning.

I would also like to thank several mentors and collaborators who supported my academic development along the way, including Dr. Shunlin Liang, Dr. Peng Yu, and Dr. Yunyue Yu, for their valuable advice and encouragement. My appreciation also goes to the staff members of the Department of Geographical Sciences. I would especially like to thank Dr. Rachel Haber for her support and guidance throughout my entire journey at UMD, from my application to the present. I also appreciate her care as my GEOG buddy.

I am also grateful to my colleagues and collaborators. I would like to thank Aolin for many insightful research discussions and for his generous support when I first arrived in the United States. I would also like to thank Lei for his guidance in carbon research. His support has helped me adapt to this new area and made the transition much easier. I also thank Zhihao, Ruohan, Zhili, and Ruichen for their collaboration.

My heartfelt thanks go to my friends who accompanied me throughout this journey. I would especially like to thank Shuyu Jin and Shuyu Ji, whose friendship brought warmth and support to my life. I also thank Quan, Mengyu, and Xin for sharing conversations and experiences about PhD life. I would like to thank the senior students who graduated before me—Jiaming, Diyang, Zheng, Yao, and Peiqi—for their support and for helping me navigate graduate life. I also thank my housemates for their companionship and support.

I would also like to thank my friends from my master's studies, Xiuxia and Linqing, for their support during my UMD application process, and Yamin, Jing, Junxiong, Yijia, Xue, and Yue for the many conversations and shared experiences. I am also grateful to my other friends, including Xun, Rui, Yiheng, and Meng, for their support and companionship.

Most importantly, I would like to thank my family—my parents, grandparents, aunts, siblings, and cousins—for their unconditional love and support throughout my life. Their encouragement has always been my strongest foundation. I would also like to thank my boyfriend, who constantly encourages me, believes in me, and supports me with patience and care. I would also like to thank my cat July, who has been a constant companion throughout this journey. From long hours of study to quiet nights of rest, she has always been by my side. She often tries to jump to higher places, sometimes failing but never giving up. Watching her persistence has been a steady source of encouragement, reminding me that progress often comes through patience and repeated effort.

This dissertation was supported by the Ann G. Wylie Dissertation Fellowship from the University of Maryland Graduate School. My graduate research assistantship was supported by the NOAA VIIRS LST project. Additional support for this research was provided by NSF, NASA, and Google.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	vi
List of Tables	x
List of Figures	xi
List of Abbreviations	xiv
Chapter 1: Introduction	1
1.1 Motivation	1
1.2 Data and Modeling Frameworks	2
1.2.1 Multiple Data Sources	3
1.2.2 Physics-Based and Data-Driven Modeling	5
1.2.3 Snapshot and Time-Series Models	7
1.3 Scientific Background	7
1.3.1 Earth System Monitoring	8
1.3.2 Energetic Processes: Land Surface Temperature (LST)	9
1.3.3 Ecological Processes: Gross Primary Productivity (GPP)	10
1.4 Research Objectives and Questions	11
1.5 Dissertation Structure	13
Chapter 2: Gap Analysis of Existing Datasets	14
2.1 Background	14
2.2 LST Data Analysis	15

2.2.1 LST Datasets	15
2.2.2 LST Results.....	21
2.3 GPP Data Analysis.....	23
2.3.1 GPP Datasets.....	23
2.3.2 GPP Results	25
2.4 Conclusion	27
Chapter 3: Artificial Intelligence (AI)-Driven Snapshot Data Integration to Enhance Variable	
Retrieval: A Case Study with LST.....	29
3.1 Background.....	29
3.2 Data	32
3.2.1 In-Situ Measurements	32
3.2.2 Satellite Top-Of-Atmosphere (TOA) Data	35
3.2.3 Satellite LST Data from Single-Channel (SC) and Split-Window (SW) Algorithms .	36
3.3 Methodology	37
3.3.1 Simulation Data Generation.....	38
3.3.2 Machine Learning (ML) Modeling.....	42
3.3.3 Transfer Learning (TL) Modeling	43
3.4 Results.....	45
3.4.1 Performance of ML and TL Models	45
3.4.2 Comparison with SC and SW Algorithms	47
3.4.3 Spatial Distribution of Different LSTs	48
3.5 Discussion	51
3.5.1 The Effectiveness of Incorporating Shortwave Data	51

3.5.2 Accuracy Improvement for Different Land Cover Types.....	53
3.5.3 Application to Other Landsat Missions	55
3.6 Conclusion	57
Chapter 4: AI-Driven Time-Series Data Integration to Enhance Variable Prediction: A Case	
Study with GPP.....	60
4.1 Background.....	60
4.2 Data	64
4.3 Methodology.....	66
4.3.1 Knowledge-aligned decomposition: explicit representation of intermediate ecological processes	69
4.3.2 Knowledge-aligned resembling: integrating satellite observations with ecosystem processes	70
4.3.3 Probabilistic label expansion and uncertainty-aware fine-tuning.....	71
4.4 Results.....	73
4.4.1 Candidate Methods	73
4.4.2 Overall Performance Evaluation.....	74
4.4.3 Time-Series and Scatter Plot Analysis.....	80
4.4.4 Probabilistic Label Expansion Strategy	81
4.5 Discussion.....	82
4.5.1 Ablation Study	82
4.6 Conclusion	84
Chapter 5: Conclusion.....	86
5.1 Summary of the Dissertation	86

5.2 Main Contributions	89
5.3 Limitations	90
5.4 Future Research Directions.....	91
Appendices.....	92
References	100

List of Tables

Table 2.1 Key information for the temperature datasets used in this study.	17
Table 2.2 Key information for the station observation used in this study.	20
Table 2.3 The percentages of valid pixels and overall root mean square errors (RMSEs) for the LST datasets assessed in this study.....	21
Table 2.4 Valid pixel percentages and overall root mean square errors (RMSEs) for land surface, air, and near-surface soil temperatures derived from the ERA5 dataset.....	23
Table 3.1 The parameters used in the radiative transfer models (RTMs).	39
Table 3.2 The performance of different algorithms on different land cover (LC) types.	53
Table 4.1 Gross primary productivity (GPP) result comparison (top results in bold; runners-up with underlines).	75
Table 4.2 Ecosystem respiration (RECO) result comparison (top results in bold; runners-up with underlines).....	75
Table 4.3 Net ecosystem exchange (NEE) result comparison (top results in bold; runners-up with underlines).....	76
Table 4.4 Absolute mean-based normalized RMSE (NRMSE) and its average across networks.	77
Table 4.5 Ablation study (top results in bold; runners-up with underlines).	83

List of Figures

Figure 1.1 The overall diagram of the dissertation.	12
Figure 2.1 Spatial distribution of the flux tower stations used for gross primary productivity (GPP) measurements.....	24
Figure 2.2 Comparison of gross primary productivity (GPP) between the Ecosystem Demography (ED) model and in-situ measurements.....	26
Figure 2.3 Comparison of plant functional types (PFTs) between the Ecosystem Demography (ED) model and the European Space Agency (ESA) Climate Change Initiative (CCI) product for four classes: broadleaf forest (BL), needleleaf forest (NL), grass and shrub (GS), and others (OT).	27
Figure 3.1 Distributions of the land surface temperature (LST) measurement stations used in this study.....	35
Figure 3.2 Flowchart of the transfer-learning (TL)-based method to retrieve land surface temperature (LST).....	38
Figure 3.3 Generating and matching longwave and shortwave simulation data.	40
Figure 3.4 Comparison between the Landsat-8 land surface temperature (LST) product generated by the single-channel (SC) algorithm, the Landsat-8 LST data generated by the split-window (SW) algorithm, the Landsat-8 LST predicted by the machine learning (ML) model, and the Landsat-8 LST predicted by the transfer-learning (TL) model over in-situ measurements.	46
Figure 3.5 The spatial comparison of various winter (a–f) and summer (g–l) land surface temperature (LST) data. (a) and (g): transfer-learning (TL) LST; (b) and (h): single-channel (SC) LST; (c) and (i): split-window (SW) LST; and (d) and (j): MODIS LST. Additionally, (e) and (k): their histograms; and (f) and (l): MODIS land cover (LC) types (1: Evergreen Needleleaf Forests;	

2: Evergreen Broadleaf Forests; 3: Deciduous Needleleaf Forests; 4: Deciduous Broadleaf Forests; 5: Mixed Forests; 8: Woody Savannas; 9: Savannas; 10: Grasslands; 11: Permanent Wetlands; 12: Croplands; 13: Urban and Built-up Lands; 14: Cropland/Natural Vegetation Mosaics; 16: Barren; 17: Water Bodies).	50
Figure 3.6 Comparison of thermal land surface emissivity (LSE) and predicted LSE from shortwave top-of-atmosphere (TOA) observations.	52
Figure 4.1 Example of in-situ data availability from the ABoVE dataset.	61
Figure 4.2 Multi-source data.	64
Figure 4.3 Spatial distribution of in-situ observation sites.	66
Figure 4.4 Overview of the proposed decomposition-and-resembling (DERE) framework for integrating process-based ecosystem simulations with satellite-derived plant functional types (PFTs) and flux tower observations, illustrated using gross primary productivity (GPP) as an example.	68
Figure 4.5 Spatial comparison of gross primary productivity (GPP) normalized RMSE (NRMSE) for the baseline, knowledge-guided machine learning (KGML), and decomposition-and- resembling (DERE). White circles indicate locations where DERE achieves an NRMSE reduction greater than 0.5 compared to the baseline or KGML.	79
Figure 4.6 Example of time-series predictions of gross primary productivity (GPP). Both the baseline and the knowledge-guided machine learning (KGML) models are based on Transformer architecture.	81
Figure 4.7 Scatter plot examples of observed versus predicted gross primary productivity (GPP) for the ABoVE dataset. Both the baseline and knowledge-guided machine learning (KGML) models are implemented using iTransformer architecture.	81

Figure 4.8 Probabilistic label expansion examples. The green shaded regions represent the 5%–95% confidence interval based on sampled time series. 82

List of Abbreviations

ABoVE	Arctic Boreal Vulnerability Experiment
AI	Artificial Intelligence
AIRS	Atmospheric Infrared Sounder
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AMSR2	Advanced Microwave Scanning Radiometer 2
AOD	Aerosol Optical Depth
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AT	Air Temperature
ATC	Annual Temperature Cycle Component
ATMS	Advanced Technology Microwave Sounder
BL	Broadleaf Forest
BSRN	Baseline Surface Radiation Network
BT	Brightness Temperature
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CLDAS	China Meteorological Administration Land Data Assimilation System
CrIS	Cross-track Infrared Sounder
CSDI	Conditional Score-based Diffusion Imputation
DOE	Department of Energy
DSR	Downward Shortwave Radiation
DTC	Diurnal Temperature Cycle Component (Δ DTC)

DERE Decomposition-and-Resembling Structure

ED Ecosystem Demography

ERA5 Fifth-generation ECMWF Atmospheric Reanalysis

ESA European Space Agency

ETM+ Enhanced Thematic Mapper Plus

GAPRI Global Atmospheric Profiles from Reanalysis Information

GED Global Emissivity Database

GEE Google Earth Engine

GEO Geostationary Earth Orbit

GLDAS Global Land Data Assimilation System

GOES Geostationary Operational Environmental Satellite

GPP Gross Primary Productivity

GS Grass and Shrub

HiWATER Heihe Watershed Allied Telemetry Experimental Research

JPSS Joint Polar Satellite System

KGML Knowledge-Guided Machine Learning

LC Land Cover

LSE Land Surface Emissivity

LST Land Surface Temperature

LUT Look-Up Table

LWDN Downwelling Longwave Radiation

LWUP Upwelling Longwave Radiation

MERRA-2 Modern-Era Retrospective Analysis for Research and Applications, Version 2

MIR Mid-Infrared

MIRS Microwave Integrated Retrieval System

MKF Multi-scale Kalman Filter

ML Machine Learning

MODIS Moderate Resolution Imaging Spectroradiometer

MSE Mean Squared Error

MSG Meteosat Second Generation

MTSAT Multifunction Transport Satellite

NASA National Aeronautics and Space Administration

NDVI Normalized Difference Vegetation Index

NEE Net Ecosystem Exchange

NIR Near-Infrared

NL Needleleaf Forest

NOAA National Oceanic and Atmospheric Administration

OLI Operational Land Imager

OT Other Vegetation Types

PCA Principal Component Analysis

PFT Plant Functional Type

PMW Passive Microwave

RAA Relative Azimuth Angle

RECO Ecosystem Respiration

RMSE Root Mean Square Error

RTE Radiative Transfer Equation

RTM Radiative Transfer Model

SC Single-Channel

SRF Spectral Response Function

STD Standard Deviation

SURFRAD Surface Radiation Budget Network

SW Split-Window

SWIR Shortwave Infrared

TES Temperature/Emissivity Separation

TIR Thermal Infrared

TIRS Thermal Infrared Sensor

TM Thematic Mapper

TL Transfer Learning

TOA Top-of-Atmosphere

VZA Viewing Zenith Angle

WCRP World Climate Research Programme

WTC Weather Temperature Component

WV Water Vapor

WVC Water Vapor Content

Chapter 1: Introduction

1.1 Motivation

Earth system monitoring plays a critical role in understanding the interactions among the land surface, atmosphere, and biosphere, with important implications for climate change, ecosystem dynamics, and environmental management (Bonan & Doney, 2018). Reliable estimation of environmental variables is essential for characterizing surface energy balance and ecosystem functioning. These variables support a wide range of applications, including drought monitoring, agricultural assessment, and the evaluation of ecosystem responses to environmental variability.

Over the past decades, substantial progress has been made in observing and modeling land surface processes through satellite remote sensing, in-situ measurement networks, and process-based ecosystem models (Pasetto et al., 2018). These data sources provide complementary information across spatial and temporal scales. Satellite observations offer extensive spatial coverage but are often indirect and affected by observational constraints. In-situ measurements provide high-accuracy observations but are spatially sparse. Process-based models encode mechanistic understanding but rely on simplified representations and parameterizations. Existing approaches often struggle to effectively utilize the complementary information provided by these data sources for accurate environmental variable estimation.

Traditional modeling approaches are often constrained by trade-offs between physical consistency and data adaptability. Physics-based methods provide interpretable and physically grounded estimates but often rely on simplified assumptions. Data-driven approaches, particularly machine learning methods, offer strong flexibility in modeling complex and nonlinear relationships but may lack physical interpretability and robustness under limited data conditions. These limitations

highlight the need for approaches that can bridge the gap between physical knowledge and data-driven learning. Knowledge-guided machine learning (KGML) methods have emerged to address this need.

Furthermore, environmental variable estimation involves fundamentally different modeling paradigms depending on their temporal characteristics. Environmental variable estimation can be broadly categorized into two paradigms: uni-temporal (snapshot) and multi-temporal (time-series) variable estimation. Variables associated with energetic processes, such as land surface temperature (LST), are typically estimated from instantaneous observations using snapshot retrieval approaches. In contrast, variables associated with ecological processes, such as gross primary productivity (GPP), evolve continuously over time and require time-series modeling approaches to capture their temporal dynamics. This distinction introduces additional challenges for developing modeling approaches that can effectively address both types of variables.

By developing methods that incorporate physical knowledge into data-driven frameworks, this research aims to improve both uni-temporal and multi-temporal variable estimation, while enhancing the consistency and reliability of multi-source Earth system monitoring.

1.2 Data and Modeling Frameworks

To address the challenges associated with multi-source data integration and variable estimation, this section introduces the primary data sources and modeling frameworks. Accurate estimation of key land surface variables requires both reliable data sources and appropriate modeling approaches (Kalma et al., 2008; Liang, 2007). This section first outlines the complementary roles of satellite observations, in-situ measurements, and process-based simulations, and then describes physics-based and machine learning approaches, followed by two modeling paradigms: snapshot retrieval

and time-series prediction.

1.2.1 Multiple Data Sources

Over the past decades, significant advances in satellite remote sensing, in-situ measurement networks, and process-based modeling have greatly improved our ability to observe and simulate land surface processes (Pan et al., 2020; Pasetto et al., 2018). However, energetic and ecological processes operate across a wide range of spatial and temporal scales, and no single observational system can fully capture their complexity (Balsamo et al., 2018). Accurate monitoring of surface dynamics relies on information derived from multiple complementary data sources (S. Xu, D. Wang, et al., 2024).

Satellite remote sensing has become one of the primary tools for observing the Earth system at regional to global scales (Dash & Ogutu, 2016; Zhao et al., 2022). Satellite instruments provide repeated measurements of land surface properties over large spatial extents, enabling consistent monitoring of surface dynamics (Z. L. Li et al., 2023; Ustin & Middleton, 2024). Optical, thermal infrared, and microwave sensors provide information related to vegetation structure, surface temperature, soil moisture, and land cover dynamics (Jones et al., 2011; Qin et al., 2021). Because of their systematic and long-term coverage, satellite observations play a critical role in characterizing environmental variability and supporting large-scale Earth system analysis (Yang et al., 2013).

However, many land surface variables cannot be measured directly by satellites and must be inferred from radiative signals through retrieval algorithms (Kalma et al., 2008; Z. L. Li et al., 2023). In addition, satellite observations may be affected by cloud contamination, viewing geometry, and limited temporal sampling (Favrichon et al., 2019; A. Jia et al., 2021; Li et al., 2009).

Consequently, although satellite data provide extensive spatial coverage, they often require additional information or modeling frameworks to derive accurate estimates of key environmental variables.

In contrast to satellite observations, in-situ measurements provide direct and highly accurate observations of environmental variables at local scales (Albergel et al., 2012; Centurioni et al., 2019). Ground-based measurement networks, including eddy-covariance flux towers, meteorological stations, and ecological monitoring sites, provide detailed observations of land surface conditions and ecosystem processes (Balzarolo et al., 2011). These measurements offer valuable insights into local-scale dynamics and serve as essential references for validating remote sensing products and evaluating model simulations.

Despite their high accuracy, in-situ measurements are inherently limited in spatial coverage. This limitation is partly due to the high cost and logistical challenges associated with establishing and maintaining measurement networks, resulting in sparse measurement sites across many regions (Wang, 2011). Consequently, while in-situ observations provide high-quality information at specific locations, they cannot independently support large-scale monitoring of land surface processes.

In addition to observational data, theory-driven simulations provide an important complementary source of information (Maass et al., 2018; Taghikhah et al., 2021). Process-based land surface and ecosystem models represent the underlying physical, ecological, and biogeochemical mechanisms governing land–atmosphere interactions (Duveiller et al., 2018; Zhao & Li, 2015). These models include radiative transfer models and ecosystem process models, which are formulated based on physical laws and mechanistic representations of radiative, ecological, and biogeochemical processes. By integrating environmental drivers, vegetation processes, and surface energy balance,

they enable the simulation of variables such as ecosystem productivity, carbon exchange, and surface energy fluxes.

Such simulation frameworks offer several advantages for Earth system analysis (Flato, 2011). They enable consistent representation of environmental processes across spatial scales and allow exploration of system responses under diverse environmental conditions (H. Wang et al., 2024). However, model simulations inevitably rely on parameterizations and simplified representations of complex processes (Luo & Schuur, 2020; McFarlane, 2011), which may introduce structural uncertainties. In addition, representing multiple interacting processes often makes these models computationally expensive (J. Zhang et al., 2018).

Taken together, satellite observations, in-situ measurements, and process-based simulations provide complementary perspectives on land surface processes. Therefore, integrating information from these different sources plays a crucial role in improving the estimation of key land surface variables.

1.2.2 Physics-Based and Data-Driven Modeling

Traditional approaches for estimating environmental variables can generally be categorized as physics-based methods or data-driven approaches (Erge & Van Oort, 2022; Willard et al., 2022). These approaches represent different strategies for translating observational information into quantitative representations of land surface processes (X. Li et al., 2024).

Physics-based methods rely on radiative transfer theory or process-based ecosystem modeling to estimate environmental variables (Verrelst et al., 2025). By explicitly representing underlying physical and ecological processes, these approaches ensure physical consistency and interpretability (J. Wang et al., 2022). However, they often rely on simplified assumptions and

predefined parameterizations of complex environmental processes (Camporese & Girotto, 2022). In addition, physics-based models typically require pre-derived parameters and accurate ancillary inputs, such as atmospheric conditions, surface properties, or vegetation characteristics (Hong & Lee, 2019). Furthermore, the rigid structure of physics-based models can limit their flexibility in fully exploiting the rich information contained in satellite observations.

Data-driven approaches provide an alternative strategy by learning relationships directly from data (Bergen et al., 2019; Liu et al., 2022). In recent years, machine learning techniques have become one of the most widely used data-driven approaches in environmental research (Sun & Scanlon, 2019; Zhu et al., 2023). These methods have demonstrated strong capability in modeling nonlinear and high-dimensional relationships and can integrate heterogeneous inputs from multiple observational sources (Strielkowski et al., 2023). Because machine learning models do not require explicit specification of mechanistic equations, they offer greater flexibility in capturing complex dependencies among environmental variables (Pichler & Hartig, 2023). However, purely data-driven models typically require large labeled datasets for training and may lack physical interpretability (Karniadakis et al., 2021). Their performance may also degrade when observational data is limited (Ahmadi et al., 2025).

To address the limitations of both paradigms, recent methodological advances seek to integrate physical knowledge with the flexibility of data-driven models (Kasilingam et al., 2024). Physics-guided machine learning incorporates physical constraints, process-based insights, or intermediate representations within data-driven frameworks to improve robustness, interpretability, and physical consistency (Karpatne et al., 2024; Wang & Yu, 2025). By combining the strengths of physics-based and data-driven approaches, these hybrid frameworks provide a promising direction for improving environmental variable estimation in multi-source Earth system monitoring (J.

Wang et al., 2022).

1.2.3 Snapshot and Time-Series Models

In environmental research, two broad modeling paradigms are commonly adopted based on their temporal characteristics: snapshot retrieval models and time-series prediction models (Kuenzer et al., 2015; Xu et al., 2015).

Snapshot retrieval models estimate variables from uni-temporal observations acquired at a single time step (Fernández-Manso et al., 2020). In this framework, the model uses observations collected at a specific time point to estimate the corresponding environmental variable at the same time (Carmona et al., 2015; Xiao et al., 2011). Because snapshot models rely on observations at a single time point, they primarily focus on estimating spatial states of environmental variables and typically do not explicitly exploit temporal dependencies.

Time-series prediction models, in contrast, incorporate multi-temporal observations to capture temporal dynamics in environmental processes (Koehler & Kuenzer, 2020). Instead of relying on a single observation, these models use sequences of historical inputs to learn temporal dependencies and describe how environmental variables evolve over time (Zhou et al., 2025). In practice, time-series models are often initialized using historical observations and generate predictions for subsequent time steps (Kooistra et al., 2024; Zrira et al., 2024).

1.3 Scientific Background

As mentioned earlier, snapshot retrieval models emphasize instantaneous estimation of environmental states across space, whereas time-series prediction models focus on modeling the temporal evolution of environmental processes. To explore these two paradigms, case studies are needed to enable concrete analysis.

Two widely used variables representing fundamentally different land surface processes are selected as case studies in this dissertation. Snapshot retrieval uses LST as a case study, whereas time-series prediction uses GPP as a case study.

1.3.1 Earth System Monitoring

Earth system monitoring provides the observational and analytical foundation for understanding the dynamic interactions among the land surface, atmosphere, biosphere, and hydrosphere (Balsamo et al., 2018; Steffen et al., 2020; Ustin & Middleton, 2021). The land surface functions as a critical interface where energy, water, and carbon are continuously exchanged with the atmosphere (Betts et al., 1996; Blyth et al., 2021; Crisp et al., 2022). These exchanges regulate surface temperature, ecosystem productivity, and biogeochemical cycling, thereby shaping regional environmental variability and global system dynamics (Denman et al., 2007; B. Wang et al., 2024).

Because the land surface plays a central role in the Earth system, accurate monitoring of its states and processes is essential for both scientific understanding and practical applications (Balsamo et al., 2018; Z. L. Li et al., 2023). Land surface dynamics are governed by both energetic exchanges that regulate the surface energy balance and ecological processes that control ecosystem productivity. Reliable estimates of land surface variables facilitate drought detection, agricultural productivity assessment, ecosystem disturbance monitoring, and resource management (Crow et al., 2012; Fensholt et al., 2013; Mashala et al., 2023). Diagnosing surface energy balance dynamics helps characterize energetic processes at the land surface (Brunsell & Gillies, 2003; Friedl, 2002), while quantifying terrestrial carbon uptake provides critical insights into ecosystem productivity, the carbon cycle, and ecosystem resilience (Frank et al., 2015; Runge et al., 2025; Trumbore, 2006).

Together, these applications highlight the central role of Earth system monitoring in environmental science and ecosystem management.

Beyond these applications, Earth system monitoring also enables the diagnosis of environmental variability by providing systematic observations of land surface conditions across spatial and temporal scales. In particular, monitoring energetic processes and ecological processes provides complementary perspectives on land–atmosphere interactions and ecosystem functioning.

1.3.2 Energetic Processes: Land Surface Temperature (LST)

Energetic processes describe how incoming solar radiation is absorbed, reflected, stored, and redistributed at the land surface (Gray et al., 2010; S. Liang et al., 2019). These processes regulate the partitioning of available energy into sensible, latent, and ground heat fluxes and play a central role in land–atmosphere interactions (Gentine et al., 2019; Mauder et al., 2020). Surface heating influences atmospheric boundary layer development, evapotranspiration dynamics, and hydrological processes, making energetic processes a fundamental component of Earth system dynamics (Niyogi, 2019).

At the land surface, energetic processes are influenced by multiple environmental factors, including surface albedo, vegetation structure, soil moisture, and atmospheric forcing (Zhao & Li, 2015). Variations in these factors modify the balance between incoming and outgoing radiation and consequently affect surface thermal conditions (Halawa et al., 2014). These interactions illustrate the complexity of surface energy balance and the need for accurate characterization of energetic processes.

LST serves as a representative variable for energetic processes (Z. L. Li et al., 2023; Taheri et al., 2022). LST reflects the thermal state of the land surface and integrates the combined effects of

radiative forcing, surface properties, and environmental conditions (Reiners et al., 2023; Weng, 2009). Because of its integrative nature, LST has been widely used to investigate surface energy balance, evapotranspiration, drought monitoring, and land–atmosphere interactions (Karnieli et al., 2010; Liou & Kar, 2014). In addition, LST provides important information for diagnosing surface heat extremes and understanding feedback between land surface conditions and atmospheric processes (Jia et al., 2024; Reiners et al., 2023).

Accurate estimation of LST is essential for improving our understanding of energetic processes and their impact on the Earth system. Therefore, LST is selected as a representative case study for investigating snapshot retrieval models in this dissertation.

1.3.3 Ecological Processes: Gross Primary Productivity (GPP)

Ecological processes govern ecosystem functioning, vegetation growth, and carbon assimilation within terrestrial ecosystems (De Deyn et al., 2008; Luo et al., 2015; Schimel et al., 2015). Through photosynthesis, plants convert atmospheric carbon dioxide into organic carbon, forming the primary entry point of carbon into the terrestrial biosphere (Baldocchi, 2008; Hari & Tyagi, 2022). These processes regulate terrestrial carbon cycling and influence ecosystem productivity, vegetation dynamics, and carbon–atmosphere interactions (Heimann & Reichstein, 2008).

Ecological activity is controlled by a combination of environmental drivers, including solar radiation, temperature, soil moisture, nutrient availability, and vegetation phenology (Bonato Asato et al., 2023; Comas et al., 2010). These controls vary across climate zones, vegetation types, and disturbance regimes, resulting in strong spatial and temporal variability in ecosystem productivity (Liu et al., 2020; Potter et al., 2008). Understanding these dynamics is essential for evaluating ecosystem responses to environmental variability and assessing the role of terrestrial

ecosystems in the global carbon cycle.

GPP represents a key variable for characterizing ecological processes (Anav et al., 2015). GPP quantifies the total carbon uptake by vegetation through photosynthesis and serves as a fundamental indicator of ecosystem productivity (Liao et al., 2023). Because GPP integrates the effects of environmental drivers and vegetation physiology, it provides critical information for understanding ecosystem functioning and carbon exchange between the land surface and the atmosphere (Anav et al., 2015; Yuan et al., 2025).

Unlike energetic variables such as LST, which often reflect near-instantaneous surface states, GPP exhibits strong temporal dependencies driven by phenology, environmental conditions, and ecosystem memory effects (Zhang et al., 2025). Seasonal vegetation cycles, environmental variability, and disturbance events can all influence GPP dynamics over time (De Pue et al., 2023). Capturing these temporal dynamics is essential for accurately characterizing ecosystem productivity and terrestrial carbon cycling. Therefore, GPP is selected as a representative case study for investigating time-series prediction models in this dissertation.

1.4 Research Objectives and Questions

The previous sections highlight a fundamental challenge in Earth system monitoring. Although multiple data sources and modeling paradigms are available, effective integration across satellite observations, in-situ measurements, and process-based simulations remains limited.

To address this limitation, approaches that can integrate physical knowledge with data-driven learning are needed. This dissertation investigates whether physics-guided machine learning can enable the integration of multi-source information to improve environmental variable estimation.

In particular, the research focuses on two representative variables: LST and GPP.

Figure 1.1 illustrates the overall research framework of this dissertation, including the three interconnected studies. The central question of this dissertation is whether physics-guided machine learning can provide a framework for integrating multi-source information to improve both snapshot retrieval of energetic variables and time-series prediction of ecological variables.

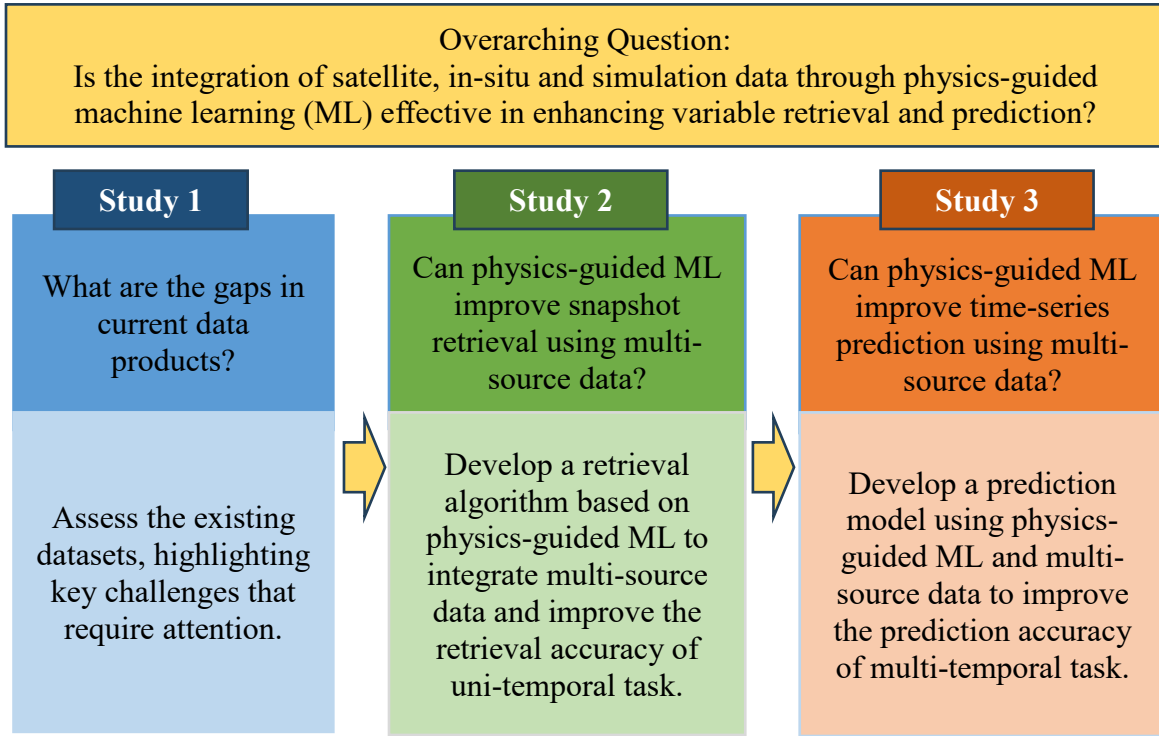


Figure 1.1 The overall diagram of the dissertation.

To address this question, this dissertation focuses on three specific objectives. First, the dissertation systematically analyzes gaps in existing LST and GPP datasets. These identified gaps motivate the methodological developments in the subsequent studies.

Second, the dissertation develops an artificial intelligence (AI)-driven framework for improving snapshot LST retrieval. This study focuses on integrating multi-source observations to enhance LST estimation accuracy while reducing dependence on ancillary inputs.

Third, the dissertation proposes a physics-guided machine learning framework to improve time-

series prediction of GPP. By incorporating complementary information from satellite observations, in-situ measurements, and process-based ecosystem simulations, the proposed approach aims to better capture the temporal dynamics of ecosystem productivity.

Together, these objectives establish a methodological framework centered on physics-guided multi-source integration for environmental variable estimation.

1.5 Dissertation Structure

This dissertation is organized into five chapters and presents three interconnected studies that collectively address the central research question.

Chapter 1 introduces the scientific background, data and modeling frameworks, research questions, and overall objectives of the dissertation.

Chapter 2 evaluates existing LST and GPP datasets. By comparing datasets derived from satellite observations, in-situ measurements, and model simulations, this chapter identifies key gaps and limitations in current datasets. These findings motivate the methodological developments presented in the following chapters.

Chapter 3 focuses on snapshot retrieval and presents an AI-based framework for improving LST estimation using multi-source satellite observations.

Chapter 4 addresses time-series prediction and develops a physics-guided machine learning framework to improve GPP prediction through the integration of satellite observations, in-situ measurements, and ecosystem model simulations.

Finally, Chapter 5 summarizes the major findings of the dissertation, highlights the main contributions, discusses limitations, and outlines future research directions for physics-guided multi-source integration in Earth system monitoring.

Chapter 2: Gap Analysis of Existing Datasets

2.1 Background

Over the past decades, numerous datasets have been developed to estimate key land surface variables such as LST and GPP (Z. L. Li et al., 2023; Liao et al., 2023). These datasets are derived from a variety of sources, including satellite observations, in-situ measurements, and process-based ecosystem models (Pandey et al., 2024; Worden et al., 2025). Advances in remote sensing technologies and modeling approaches have significantly improved the availability of environmental datasets and enabled large-scale monitoring of land surface processes (Balsamo et al., 2018; Chen et al., 2022; S. Xu, Y. Liu, et al., 2024).

Despite these advances, existing datasets often differ substantially in spatial resolution, temporal coverage, retrieval algorithms, and underlying assumptions (Zhao & Li, 2015). Satellite-based products may be affected by observational constraints such as cloud contamination or limited temporal sampling, while model-based and machine learning-derived datasets may exhibit uncertainties related to model structure, training data, and input variables (Bishop, 2013). As a result, different datasets may produce different estimates of the same environmental variable, leading to uncertainties in environmental analysis and limiting their effective use in scientific applications.

Despite substantial progress, several challenges remain in the evaluation and application of these datasets. For LST, multiple satellite-derived products are available, but their validation is often conducted using different in-situ datasets and evaluation approaches, making it difficult to directly compare their performance (Y. Liu et al., 2024; Xu & Wang, 2022). A systematic evaluation of existing datasets is therefore necessary to better understand their characteristics and limitations. For GPP, process-based ecosystem models have been widely developed and successfully applied

in long-term carbon monitoring. These models provide physically interpretable simulations and have been extensively used in studies of ecosystem dynamics and the global carbon cycle. However, a focused evaluation of model simulations against in-situ observations remains necessary to examine their performance and consistency under real-world conditions. These considerations further emphasize the importance of a comprehensive evaluation of both LST and GPP datasets.

This chapter conducts a comprehensive analysis of existing LST and GPP datasets. By comparing datasets derived from satellite observations, model simulations, and in-situ measurements, this study aims to identify key gaps and limitations in current data products. These findings provide the foundation for the methodological developments presented in the following chapters, where improvements in LST retrieval and GPP prediction are further explored.

2.2 LST Data Analysis

2.2.1 LST Datasets

Near-surface temperatures, such as LST, air temperature, and soil temperature, play significant roles in surface radiation and energy balance. Among these variables, LST is an important parameter for characterizing surface energy balance and serves as a representative variable of energetic processes (Xu & Cheng, 2021; Xu et al., 2021).

This section focuses on the evaluation of LST datasets, including products derived from thermal infrared (TIR), passive microwave (PMW), reanalysis, and all-weather LST datasets. For comparison purposes, air temperature and soil temperature are also considered.

Eight gridded datasets were collected for this study, and their key characteristics are presented in Table 2.1. All available data for the year 2014 were collected. Moderate Resolution Imaging

Spectroradiometer (MODIS) official LST products consist of two separate data suites. The MXD11 and MXD21 LST data are generated using the split-window technique (Wan, 2014; Wan & Dozier, 1996), and ASTER temperature/emissivity separation (TES) techniques (Hulley et al., 2014), respectively. The quality flag of MXD11 data is determined by cloud effects, average emissivity, and LST error; while that of MXD21 data is determined by missing data and calibration conditions, cloud conditions, atmospheric opacity, LST accuracy, emissivity, min-max difference for thermal bands and algorithm iteration rate (Hulley et al., 2016; Tan et al., 2021).

The Copernicus LST version 2 datasets include LST data from multiple sensors onboard different geostationary (GEO) satellites, including the Meteosat Second Generation (MSG), Geostationary Operational Environmental Satellite (GOES), and Multifunction Transport Satellite (MTSAT). These LST data are estimated from the top-of-atmosphere (TOA) brightness temperature (BT) from the infrared spectral channels, together with the albedo, vegetation cover, and soil moisture (Freitas et al., 2013). Himawari-7 data, which covers the study area, was used in this study.

The Atmospheric Infrared Sounder (AIRS) is a grating spectrometer ($R = 1200$) onboard Aqua that consists of 2378 infrared channels and four visible/near-infrared channels. The AIRS retrieval system is designed to obtain highly accurate atmospheric temperature and moisture profiles, as well as a variety of additional earth/atmosphere products. The fundamental principle of this system is to generate clear column radiances for each channel (Susskind et al., 2014). Its products have been widely used to improve weather and drought forecasting. The AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS-only) $1^\circ \times 1^\circ$ product (v006) was used in this study (Kahn et al., 2014).

Table 2.1 Key information for the temperature datasets used in this study.

Name in this study	Short name	Product name	Variables	Product type	Spatial resolution	Temporal resolution
MXD11	MOD11A1	MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061	LST	TIR	1 km	Daily
	MYD11A1	MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061	LST	TIR	1 km	Daily
MXD21	MOD21A1 D	MODIS/Terra Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Day V061	LST	TIR	1 km	Daily
	MYD21A1 D	MODIS/Aqua Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Day V061	LST	TIR	1 km	Daily
	MOD21A1 N	MODIS/Terra Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Night V061	LST	TIR	1 km	Daily
	MYD21A1 N	MODIS/Aqua Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Night V061	LST	TIR	1 km	Daily
Himawari -7	MTSAT	Copernicus LST version 2	LST	TIR	5 km	Hourly
AIRS	AIRS3STD	AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS-only) 1 degree x 1 degree V006	LST	TIR	1 degree	Daily
MIRS	MIRSAIMG	MIRS Precipitation and Surface Products	LST	PMW	~27 km	Daily
All-weather LST (Xu)	-	1 km seamless land surface temperature dataset of China	LST	All-weather	1 km	Daily
All-weather LST (Zhang)	TRIMS LST-TP	Western China Daily 1 km spatial resolution all-weather land surface temperature data set V1	LST	All-weather	1 km	Daily
ERA5	ERA5	ERA5 hourly data on single levels from 1979 to present	Air/ Skin/ Soil temperature	Re-analysis	0.1 degree	Hourly

The Microwave Integrated Retrieval System (MIRS) is a 1-D variational and iterative physical inversion system (Boukabara et al., 2011). Some sounder products (JPSS_SND) are generated by using Cross-track Infrared Sounder (CrIS) and Advanced Technology Microwave Sounder (ATMS) sensor data records and intermediate products from the Joint Polar Satellite System (JPSS). The MIRS precipitation and surface products are JPSS_SND products that provide skin temperature. The data have a similar observation time as the MODIS/Aqua data, and its spatial resolution is approximately 25-30 km at the nadir, increasing to approximately 100 km at scan edge.

Xu's 1 km seamless LST dataset of China is generated by fusing MODIS/Aqua LST and Advanced Microwave Scanning Radiometer for EOS (AMSR-E)/Advanced Microwave Scanning Radiometer 2 (AMSR2) data (Xu & Cheng, 2021). First, an empirical retrieval algorithm based on a look-up table (LUT) and a downscaled algorithm based on geographically weighted regression (GWR) is used to generate the AMSR-E/AMSR2 LST and 1 km downscaled AMSR-E/AMSR2 LST. The cumulative distribution function (CDF) matching method is then used to adjust the downscaled AMSR-E/AMSR2 LST to improve its accuracy and fill the MODIS LST to generate the completely filled MODIS LST. Finally, the multi-scale Kalman filter (MKF) approach is used to fuse the AMSR-E/AMSR2 LST, and the completely filled MODIS LST to generate a 1 km seamless LST dataset.

Zhang's Western China daily 1 km spatial resolution all-weather LST dataset V1 is generated using a method based on temporal component decomposition (Zhang et al., 2019). In this method, LST is decomposed into three temporal components: the annual temperature cycle component (ATC), diurnal temperature cycle component (Δ DTC), and weather temperature component (WTC). Its main input data include MODIS/Aqua LST, AMSR-E/AMSR2 data, and LST from the

China Meteorological Administration Land Data Assimilation System (CLDAS) and Global Land Data Assimilation System (GLDAS). In this method, the 1-km ATC is derived from discontinuous MODIS/Aqua and AMSR-E/AMSR2 LST, and the 1-km Δ DTC is derived from CLDAS and GLDAS LST. Then, the 1-km WTC is derived from the continuous daily AMSR-E/AMSR2 LST, and the daily 1 km all-weather LST is obtained by summing the ATC, Δ DTC, and WTC.

Re-analysis data are a combination of model forecasts and newly acquired observations to form the best new estimate using data assimilation methods (Hennermann & Berrisford, 2017). The ECMWF develops several atmospheric and ocean re-analysis products (Dee et al., 2011). The Fifth-generation ECMWF atmospheric reanalysis (ERA5) is its latest atmospheric re-analysis data, offering a global improvement based on 10 years of model and data assimilation development. The ERA5 hourly data on single levels from 1979 to the present were used in this research and provide hourly estimates for a large number of atmospheric, ocean waves, and land surface quantities (Hersbach et al., 2018). In this study, 2 m air temperature, skin temperature, and soil temperature levels 1/2/3 (soil depths: 0-7, 7-28, and 28-100 cm) were used.

In addition to the gridded data, the in-situ measurements, and auxiliary data on land cover (LC) type and elevation were collected for validation and analysis. Ground measurement data were collected from a hydrometeorological observation network dataset generated by the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) project (X. Li et al., 2013; Liu et al., 2018). The data from 7 stations were utilized. These measurements were collected from multiple sites located in close proximity but representing different LC types, providing an opportunity to generate new in-situ data that matches the spatial resolution of the satellite data and help explore the spatial heterogeneity issue. The key information for these sites is provided in Table 2.2. Surface upwelling longwave radiation (LWUP), downward longwave radiation

(LWDN), 2 m air temperature, and soil temperatures at levels 1–3 from multiple sites were used in this study.

Table 2.2 Key information for the station observation used in this study.

Station	Arou	Shenshawo Sandy Desert	Dashalong	Zhangye Wetland	Bajitan Gobi	Huazhaizi Desert Steppe	Daman
Short Name	AR	SSW	DSL	ZY	BJT	HZZ	DM
Landscape (Liu et al., 2018)	subalpine meadow	sandy desert	marsh alpine meadow	reed	Reaumuria desert	Kalidium foliatum desert	maize
Longitude (°)	100.46	100.49	98.94	100.45	100.3	100.32	100.37
Latitude (°)	38.05	38.79	38.84	38.98	38.92	38.77	38.86
Elevation (m)	3033	1594	3739	1460	1562	1731	1556
MODIS LC Type	Grasslands	Barren	Grasslands	Crops	Barren	Grasslands	Crops

The selected in-situ measurements are those that are close to the pixel observation time, with a maximum time difference of 5 minutes. The air and soil temperatures are directly obtained from field measurements. To obtain LST, the LWUP and LWDN are input to the Stefan-Boltzmann (Equation 1):

$$T_s = \left[\frac{F^\uparrow - (1 - \varepsilon_b) F^\downarrow}{\sigma \varepsilon_b} \right]^{1/4} \quad (1)$$

where T_s denotes the LST in Kelvin (K); $F^\uparrow$ and $F^\downarrow$ are the measured LWUP and LWDN in watts per square meter (Wm^{-2}), respectively; ε_b represents the surface broadband emissivity (BBE); and σ is the Stefan–Boltzmann constant ($5.67 \times 10^{-8} \text{ Wm}^{-2} \text{ K}^{-4}$). ε_b is estimated from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Emissivity Database (GED) product (Hulley & Hook, 2009), coupled with a linear equation that integrates the narrowband surface emissivities from ASTER bands 10 to 14, as detailed by Cheng et al. (2013):

$$\varepsilon_b = 0.197 + 0.025\varepsilon_{10} + 0.057\varepsilon_{11} + 0.237\varepsilon_{12} + 0.333\varepsilon_{13} + 0.146\varepsilon_{14} \quad (2)$$

where $\varepsilon_{10} - \varepsilon_{14}$ are the surface narrowband emissivities of ASTER bands 10-14, respectively.

For coarse-spatial-resolution products (spatial resolution $\geq 1^\circ$), including AMSR2, MIRS, and AIRS with spatial resolutions of 25 km, 27 km, and 1° , respectively, land-cover-weighted aggregation of in-situ measurements is used for comparison. Specifically, in-situ observations are aggregated according to land-cover type and their relative fractional area within each pixel.

2.2.2 LST Results

Table 2.3 presents the percentages of valid pixels and the overall root mean square errors (RMSEs) for the LST datasets used in this study. The comparison reveals clear trade-offs among different LST datasets in terms of data availability and retrieval accuracy.

Table 2.3 The percentages of valid pixels and overall root mean square errors (RMSEs) for the LST datasets assessed in this study.

Type	Data	Available Pixel Percentage-Day (%)	Available Pixel Percentage-Night (%)	RMSE- Day	RMSE- Night
TIR	MOD11A1	57.4	54.8	4.15 K	4.03 K
TIR	MOD21A1	54.3	43.1	4.61 K	2.51 K
TIR	MYD11A1	51.4	58.3	4.68 K	3.24 K
TIR	MYD21A1	50.8	45.5	4.36 K	1.90 K
TIR	Himawari-7	44.4	56.3	6.64 K	3.54 K
TIR	AIRS	74.2	81.0	6.35 K	4.58 K
PMW	MIRS	94.9	94.7	5.51 K	7.44 K
All-weather	All-weather LST (Xu)	100	100	4.72 K	3.68 K
All-weather	All-weather LST (Zhang)	100	100	4.81 K	3.76 K
Re-analysis	ERA5	100	100	8.79 K	6.04 K

TIR products generally achieve relatively high accuracy, with RMSE values mostly below 5 K. However, their available pixel percentages are significantly limited, typically ranging from approximately 44% to 74%, mainly due to cloud contamination. The AIRS dataset shows a higher

valid pixel percentage than other TIR products because it includes both cloud-free observations and data collected under thin cloud conditions (Evan Manning et al., 2020).

In contrast, PMW products provide much higher data availability, with valid pixel percentages exceeding 94% for both daytime and nighttime observations. Nevertheless, their retrieval accuracy is lower than that of most TIR products, with RMSE values exceeding 5 K and reaching 7.44 K during nighttime.

All-weather LST datasets achieve complete spatial coverage, with valid pixel percentages reaching 100% for both daytime and nighttime conditions. Their RMSE values remain comparable to or slightly higher than those of TIR-based products, suggesting that all-weather approaches can substantially improve data availability while maintaining reasonable retrieval accuracy.

Reanalysis datasets such as ERA5 also provide full spatial coverage but exhibit substantially larger errors, with RMSE values exceeding 8 K during daytime. This result suggests that reanalysis products have limited capability to accurately represent LST.

Overall, this comparison indicates that the quality of current LST products remains limited. To examine whether this difficulty is unique to LST, land surface, air, and near-surface soil temperatures were compared within the same ERA5 framework. Table 2.4 presents the valid pixel percentages and RMSEs for land surface, air, and near-surface soil temperatures from the ERA5 dataset. The results show that LST exhibits larger RMSE values than the other temperature variables.

Table 2.4 Valid pixel percentages and overall root mean square errors (RMSEs) for land surface, air, and near-surface soil temperatures derived from the ERA5 dataset.

Data	Variables	Available Pixel Percentage-Day (%)	Available Pixel Percentage- Night (%)	RMSE- Day	RMSE- Night
ERA5	LST	100	100	8.79 K	6.04 K
	Air temperature	100	100	4.61 K	2.51 K
	Soil temperature (0 - 7 cm)	100	100	5.73 K	5.02 K
	Soil temperature (7 - 28 cm)	100	100	3.37 K	3.92 K
	Soil temperature (28 - 100 cm)	100	100	3.62 K	3.56 K

One possible explanation is that air temperature and soil temperature are thermally smoother, whereas LST exhibits a larger dynamic range. In addition, LST shows stronger dependence on surface properties. These characteristics highlight the greater complexity of LST estimation and motivate the development of improved approaches for LST retrieval by incorporating satellite-derived surface property information and in-situ observations, which will be further explored in Chapter 3.

2.3 GPP Data Analysis

2.3.1 GPP Datasets

GPP represents the total carbon uptake of ecosystems through photosynthesis and is a key indicator of ecosystem functioning and carbon cycling. Accurate estimation of GPP is therefore essential for understanding ecosystem productivity and evaluating land–atmosphere carbon exchange processes.

In this study, GPP estimates from the Ecosystem Demography (ED) model (Hurtt et al., 1998; Ma et al., 2021; Moorcroft et al., 2001) were compared with in-situ observations. The ED model is a process-based terrestrial biosphere model that simulates vegetation dynamics and carbon cycling across spatial and temporal scales. Developed over the past two decades, it has been widely used

in studies of global carbon cycling, land-use change, and ecosystem responses to climate variability. The model has also supported major scientific and policy efforts, including the Global Carbon Budget (Friedlingstein et al., 2023; Friedlingstein et al., 2024) and the National Aeronautics and Space Administration (NASA) Carbon Monitoring System (Hurt et al., 2019). To evaluate model performance, in-situ GPP measurements were collected from flux tower observations, including those from the NASA Arctic Boreal Vulnerability Experiment (ABOVE) dataset (Bill et al., 2023). The spatial distribution of the sites is shown in Figure 2.1. These measurements provide direct estimates of ecosystem carbon fluxes and are widely used for validating ecosystem models and remote sensing products.

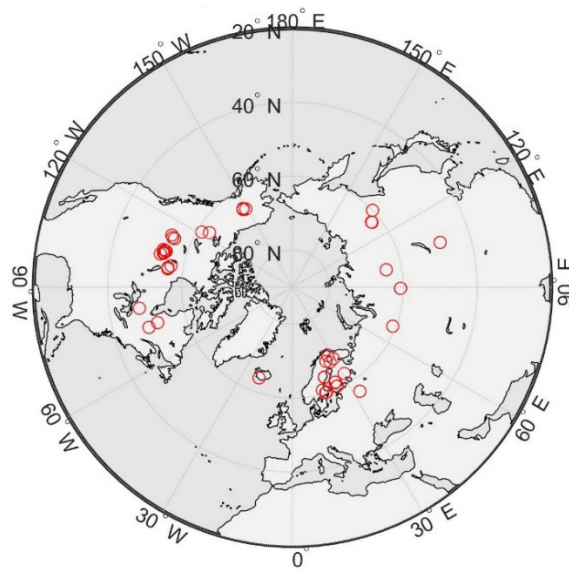


Figure 2.1 Spatial distribution of the flux tower stations used for gross primary productivity (GPP) measurements.

In addition, plant functional type (PFT) information was obtained from the European Space Agency (ESA) Climate Change Initiative (CCI) product. The dataset provides satellite-derived estimates of vegetation types and their spatial components. Because different PFTs exhibit distinct

productivity characteristics, the representation of PFT composition can strongly influence GPP simulations in ecosystem models.

By combining ED model simulations, in-situ observations, and satellite-derived PFT information, this study evaluates the consistency between modeled and observed GPP and explores potential factors influencing model performance.

2.3.2 GPP Results

Figure 2.2 shows example comparisons of GPP between the ED model simulations and in-situ measurements. In many cases (e.g., the first column), the ED model shows good agreement with the in-situ observations, capturing both the seasonal variations and the magnitude of GPP.

However, in some cases, the ED simulations reproduce similar temporal patterns but differ in magnitude from the in-situ observations. These discrepancies suggest that certain factors influencing ecosystem productivity may not be fully represented in the model simulations.

To further investigate this issue, the PFT composition used in the ED simulations was examined. Different PFTs exhibit distinct productivity patterns. For example, broadleaf forests and needleleaf forests often show different seasonal dynamics and productivity levels. As a result, variations in PFT composition may influence simulated GPP.

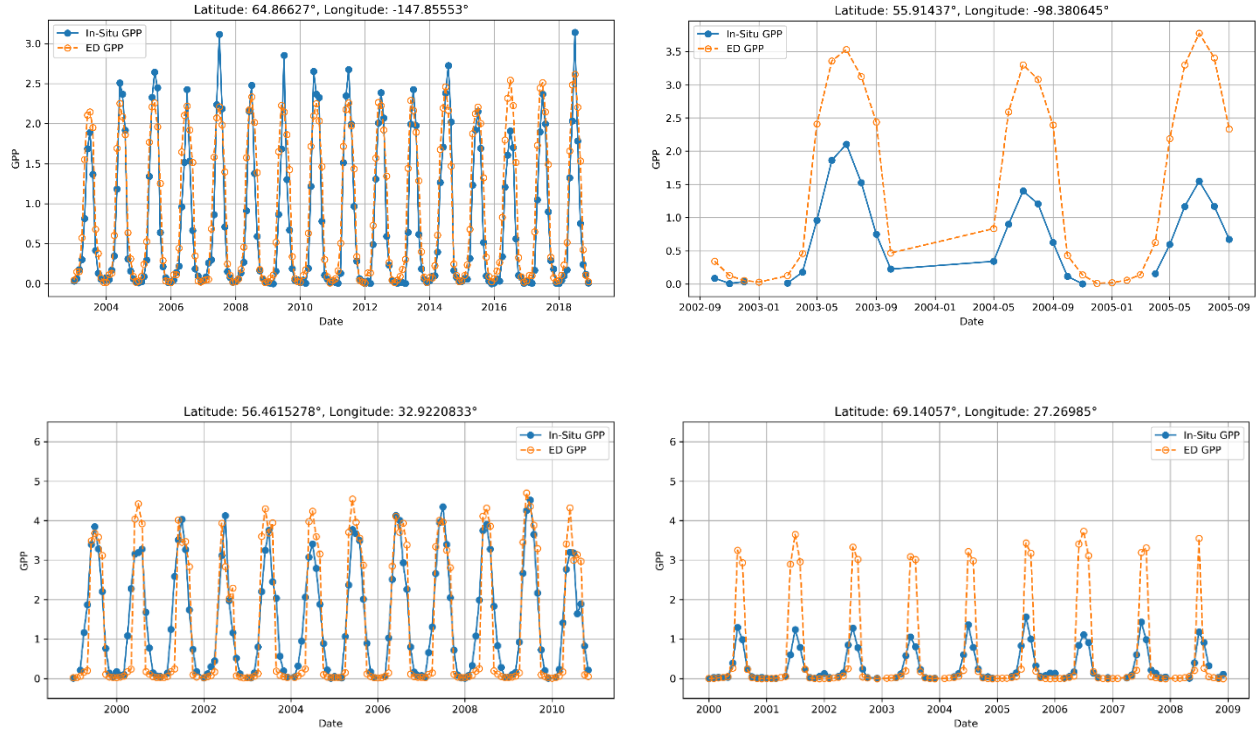


Figure 2.2 Comparison of gross primary productivity (GPP) between the Ecosystem Demography (ED) model and in-situ measurements.

Figure 2.3 compares PFT composition between the ED model and the ESA CCI product, corresponding to the two panels in the second row of Figure 2.2. Four PFT classes are considered in this comparison: broadleaf forest (BL), needleleaf forest (NL), grass and shrub (GS), and others (OT). In this figure, PFT fractions from the ED simulations and satellite observations are shown using stacked bar charts. Different colors represent different PFT classes, and the largest fraction indicates the dominant PFT.

In the first column of Figure 2.3, the two datasets identify the same dominant PFT. Under these conditions, the ED simulations generally produce GPP estimates that align well with the in-situ observations (Figure 2.2, second row, first column). However, in the second column of Figure 2.3,

the dominant PFT differs between the datasets. Further examination shows that changes in PFT composition are associated with differences in the simulated GPP (Figure 2.2, second row, second column).

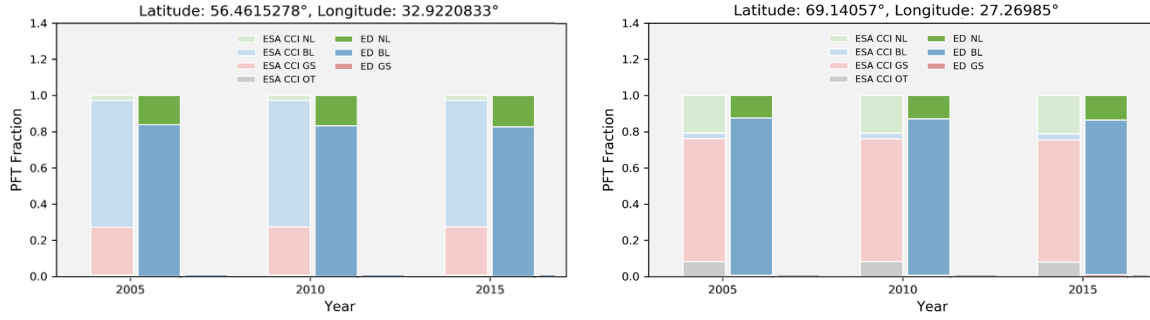


Figure 2.3 Comparison of plant functional types (PFTs) between the Ecosystem Demography (ED) model and the European Space Agency (ESA) Climate Change Initiative (CCI) product for four classes: broadleaf forest (BL), needleleaf forest (NL), grass and shrub (GS), and others (OT).

These results suggest that the representation of PFTs plays an important role in determining GPP prediction performance. This finding highlights the importance of accurately representing vegetation composition and motivates the development of improved approaches for time-series GPP prediction by integrating satellite-derived PFT information with in-situ observations, which will be further explored in Chapter 4.

2.4 Conclusion

This chapter conducted a systematic evaluation of existing datasets for two key land surface variables: LST and GPP. By comparing satellite products, model simulations, and in-situ observations, this analysis identified key challenges and limitations in current environmental datasets.

For LST, the comparison across multiple datasets revealed clear trade-offs between data

availability and retrieval accuracy. Thermal infrared products generally provide relatively high accuracy but suffer from limited spatial coverage due to cloud contamination, while passive microwave and reanalysis products offer higher data availability but often exhibit lower accuracy. Further comparison with air and soil temperatures within the ERA5 framework shows that LST exhibits larger RMSEs than other temperature variables. One possible explanation is that LST has a larger dynamic range and is strongly influenced by surface properties. These characteristics highlight the complexity of LST estimation and suggest that incorporating surface-related information may help improve LST retrieval.

For GPP, comparisons between ED model simulations and in-situ observations show that although the model can capture the seasonal dynamics of ecosystem productivity, discrepancies in magnitude occur in some cases. Further analysis indicates that these differences are associated with variations in PFT representation. Because different vegetation types exhibit distinct productivity patterns, inconsistencies in dominant PFT composition can lead to differences in predicted GPP, suggesting that incorporating PFT-related information may help improve GPP estimation.

Overall, the results of this chapter highlight two key challenges in current environmental datasets. First, LST retrieval is strongly linked to surface properties, suggesting the need to incorporate satellite-derived surface information together with in-situ observations. Second, GPP prediction varies with dominant PFT representation, highlighting the importance of accurately representing vegetation composition. These findings motivate the methodological developments presented in the following chapters. Chapter 3 focuses on improving LST retrieval by integrating satellite surface property information with in-situ observations, while Chapter 4 investigates improved approaches for GPP prediction by integrating satellite-derived PFT information with in-situ data.

Chapter 3: Artificial Intelligence (AI)-Driven Snapshot Data Integration to Enhance Variable Retrieval: A Case Study with LST

3.1 Background

Variables that are estimated from instantaneous observations are typically treated as snapshot variables. LST is a representative example, defined as the radiative skin temperature of the land surface and calculated from measured radiance (Szagri et al., 2023). It represents the thermodynamic temperature of a thin layer at the interface between soil, vegetation, or other surface components and the atmosphere (Z.-L. Li et al., 2023). LST serves as a key parameter in various studies, including research on urban heat islands (Feng et al., 2023; A. Jia et al., 2020; F. Liu et al., 2024), climate change (Eleftheriou et al., 2018; S. Xu et al., 2023), evapotranspiration (Jia et al., 2026; Moura et al., 2019; Zhang et al., 2022), hydrological cycles (Gampe et al., 2016; Jia et al., 2025; Olivera-Guerra et al., 2022), and vegetation monitoring (Ding et al., 2024; Zilin Li et al., 2024). In particular, LST's significance is pronounced in analyzing the urban heat island effect and in understanding the broader impacts of climate change. The urban heat island effect is a phenomenon characterized by higher temperatures in urban areas compared to rural areas (Oke, 1982; Yang et al., 2022). The monitoring and mitigation of urban heat islands rely heavily on accurate LST measurements, enabling policymakers to implement effective heat management strategies and improve urban livability (Voogt & Oke, 2003). Additionally, LST is recognized by the Global Climate Observing System as a vital parameter in the land biosphere (Hollmann et al., 2013). It plays a crucial role in the development of climate change indices, offering essential data for researchers and policymakers to monitor and evaluate the effects of climate variability and change on both local and global scales (Halder et al., 2021).

Therefore, high-quality LST datasets are required to provide effective scientific support for

environmental and climate change research (Jia et al., 2022; Jia et al., 2023; Xu et al., 2019). Field stations can serve as one of the sources for LST data. They offer precise measurements but are limited in spatial extent (Vancutsem et al., 2010; Shuo Xu et al., 2023). This constraint has spurred the use of TIR remote sensing, which provides spatial information crucial for environmental and climatic studies (Xu & Cheng, 2021; Zhang et al., 2020).

Various algorithms have been developed for deriving LST from TIR data, each with its strengths and limitations. The single-channel (SC) method utilizes a single radiative transfer equation (RTE) but requires accurate land surface emissivity (LSE) and atmospheric conditions (Cristóbal et al., 2009; Jimenez-Munoz et al., 2008; Qin et al., 2001). The split-window (SW) method mitigates atmospheric effects by utilizing two TIR bands, improving the retrieval process, but still dependent on accurate LSE (Mao et al., 2005; Wan & Dozier, 1996; Ye et al., 2022). Multichannel methods use multiple TIR channels for atmospheric correction, but most require precise LSE measurements, and the inclusion of mid-infrared (MIR) channels can often introduce errors (Gillespie et al., 1998; Liang, 2001; Wan, 2004). The TES algorithm retrieves the LST and LSE simultaneously without external LSE data as input but requires precise atmospheric correction (Gillespie et al., 1998; Malakar & Hulley, 2016). Machine learning (ML) techniques offer significant improvements over traditional physical models, and ML models benefit from extensive training data (Yuan et al., 2020; Y. Zhang et al., 2018). However, they struggle with data from diverse domains, such as simulations and real-world observations, and depend heavily on auxiliary data such as LSE, water vapor content (WVC), downward shortwave radiation (DSR), and vegetation indexes (H. Jia et al., 2021; Mao et al., 2007; Xu et al., 2021).

In summary, current approaches for LST retrieval face notable shortcomings: 1) conventional ML methods rely heavily on training data, but cannot effectively utilize data from distinct domains,

such as simulation and the real world; and 2) traditional LST-retrieval methods are constrained by the coupling of the LST, the LSE and the atmosphere. As a result, both traditional physical LST-retrieval methods and ML LST-retrieval methods frequently rely on auxiliary data, particularly LSE.

Transfer learning (TL) represents a cutting-edge methodology that capitalizes on the knowledge from a distinct source domain to enhance the learning tasks within a target domain (Pan & Yang, 2010). Specifically, TL can transfer the knowledge accumulated in a pre-trained model from the source domain to a target model by utilizing model parameters, and fine-tune the model for the target domain by using data from the target domain (Pan & Yang, 2010). Despite its successful application across a variety of remote sensing endeavors (Khaki et al., 2021; Li et al., 2022; Tong et al., 2020; Zhou et al., 2020; Zhou et al., 2023), the application of TL to improve LST estimation remains underexplored.

This study introduces an innovative TL-based approach for LST retrieval that effectively bridges the simulation domain and the real-world domain, serving as the source and target domains, respectively. This approach utilizes comprehensive knowledge from radiative transfer model (RTM) simulations alongside empirical observations, facilitating the development of models with enhanced accuracy. The proposed TL framework solely relies on the integration of shortwave and longwave TOA data. Strong correlations have been found between shortwave spectral reflectance or derived indices and thermal emissivity (Meng et al., 2019; Tang et al., 2015). With the addition of shortwave data, the proposed method can retrieve LST without the explicit requirement for external LSE information. The LST data generated by this study can provide valuable support for research in environmental and climate change.

The remainder of this paper is structured as follows: Section 3.2 introduces the data used in this

study, including in-situ measurements and satellite data. Section 3.3 outlines the methodology, encompassing the generation of simulation data as well as the development of both the ML and TL models. Section 3.4 provides a comprehensive examination of the LST values derived using the ML and TL techniques, comparing these with the official Landsat LST product based on SC (Cook et al., 2014; Malakar et al., 2018) and the SW-derived LST data (Meng et al., 2019), while also illustrating the spatial distribution patterns of these different LSTs. Section 3.5 discusses the advantages of incorporating shortwave data into the proposed models, examines the performance of various algorithms across different LC types, and extends the application of the proposed methodologies to other Landsat missions. Finally, a summary of findings and conclusions is provided in Section 3.6.

3.2 Data

3.2.1 In-Situ Measurements

Our research extensively utilized in-situ measurements from four prominent networks: Surface Radiation Budget Network (SURFRAD), FLUXNET, Baseline Surface Radiation Network (BSRN), and AmeriFlux Network, offering comprehensive observations across various climate regions, LC categories, and ecosystems. This extensive geographical coverage enables robust conclusions and facilitates representative analysis and discussion in this study.

Funded by the National Oceanic and Atmospheric Administration's (NOAA) Office of Global Programs, SURFRAD's primary goal is to supply ground truth references essential for climate research (<https://www.esrl.noaa.gov/gmd/grad/surfrad/overview.html>) (Augustine et al., 2000; Augustine et al., 2005). Overall, the SURFRAD network consists of seven stations located in different climate zones with different surface coverage types. FLUXNET, initiated by the

Lawrence Berkeley National Laboratory (USA), employs the eddy covariance technique to measure the cycling of carbon, water, and energy between the biosphere and atmosphere (<https://fluxnet.org/>) (Pastorello et al., 2020). The network encompasses over a thousand flux measurement sites, distributed throughout numerous climatic zones and representative biomes. Launched by the World Climate Research Program (WCRP), the radiometric network BSRN aims to detect important changes in the Earth's radiation field (<https://bsrn.awi.de>) (Driemel et al., 2018; Ohmura et al., 1998). The network covers different climatic zones from 80° N to 90° S and currently comprises 74 stations (58 active). Established by the Department of Energy (DOE) of the United States, AmeriFlux aims to quantify the variability in water and carbon exchanges between terrestrial ecosystems and the atmosphere (<https://ameriflux.lbl.gov/about/about-ameriflux/>) (Hargrove et al., 2003). This network includes more than 100 stations covering various ecosystem types.

In-situ LST data derivation involved the Stefan–Boltzmann equation (Equation 1), utilizing measurements of LWUP and LWDN. Data collection spanned more than a decade (March 2013 to November 2023). Through specific filtering criteria, focusing on the availability of Landsat-8 LST data for geographical location and minimal spatial heterogeneity, 241 sites (including 6 from SURFRAD, 40 from FLUXNET, 13 from BSRN, and 182 from AmeriFlux Network) were selected for this study, enhancing the data's representativeness and reliability for the analysis.

Given the Landsat-8 dataset's approximately 30-meter or 100-meter (TIR) spatial resolution, it was imperative to acknowledge and address the potential for significant spatial heterogeneity at certain sites, which might compromise the representativeness of their corresponding satellite pixels (Shuo Xu et al., 2023). Therefore, the spatial heterogeneity was assessed and served as a key criterion for site selection. This study utilized Landsat-8 LST data from the same period and performed a

standard deviation (STD) analysis of the pixels surrounding each site using a 3×3 window to quantify spatial heterogeneity. Sites with an average STD exceeding 4 K were deemed unsuitable and subsequently excluded from the dataset. This careful site selection and filtering process was aimed at mitigating spatial heterogeneity issues, thereby ensuring the dataset's enhanced accuracy and representativeness for subsequent analysis.

The distributions of the selected 241 sites are shown in Figure 3.1 and Table S1 in Appendices. Additional details regarding LC types are accessible via the respective networks' websites.

For the LST derivation from the chosen 241 sites, their LWUP and LWDN readings were input into the Stefan–Boltzmann equation (Equations 1 and 2 in Section 2.2.1). SURFRAD, FLUXNET, BSRN, and AmeriFlux Networks provide high-quality and continuous ground-based LWUP and LWDN measurements at temporal resolutions ranging from 1 to 3 minutes, 30 minutes, 1 to 60 minutes, and 30 minutes, respectively. This study utilized in-situ measurements that closely aligned with the Landsat-8 product pixels' observation times, ensuring temporal relevance and consistency.

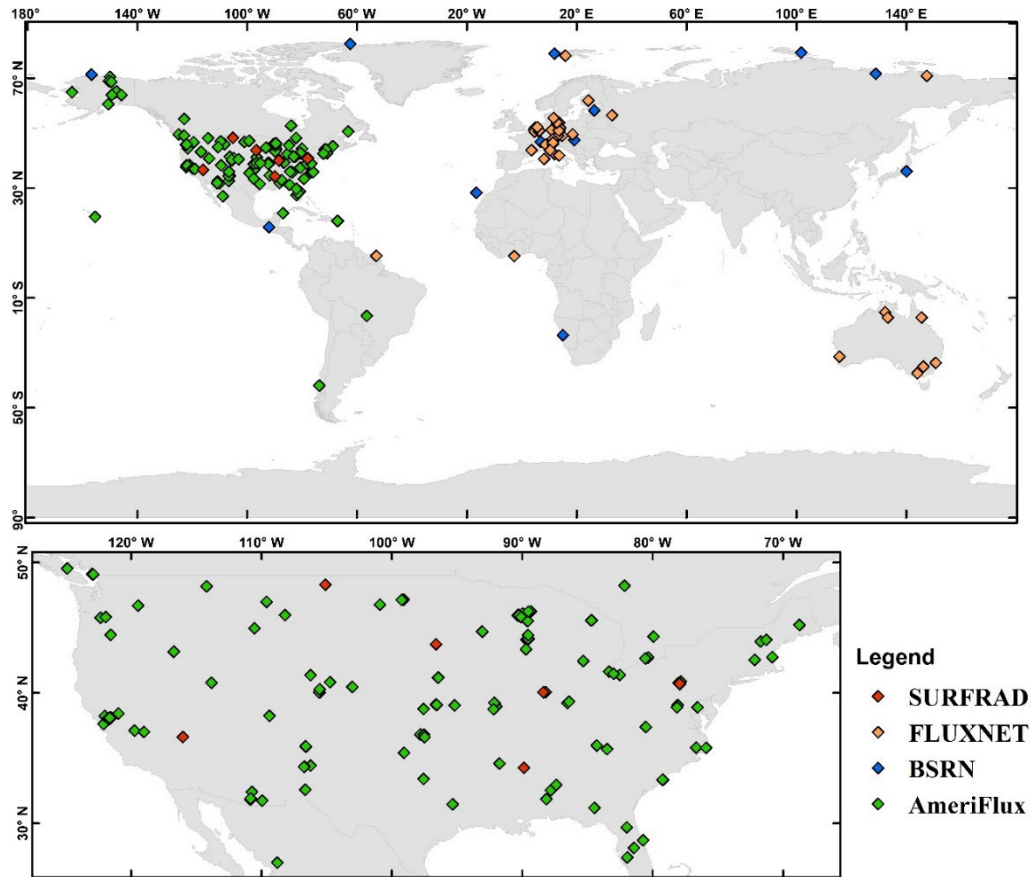


Figure 3.1 Distributions of the land surface temperature (LST) measurement stations used in this study.

This meticulous approach to data collection, site selection, and LST derivation ensures the robustness and integrity of the analysis, ensuring that the findings are based on accurate and representative datasets.

3.2.2 Satellite Top-Of-Atmosphere (TOA) Data

The Landsat program provides relatively high-spatial-resolution satellite data (30 or 100 meters), making it invaluable for various research endeavors. Landsat-8, launched on February 11, 2013, stands out with its advanced instrumentation, carrying the Operational Land Imager (OLI) and the

Thermal InfraRed Sensor (TIRS) instruments. Notably, Landsat-8 TIRS features an additional thermal band compared to Landsat 4–5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+). This additional thermal band not only enhances the extraction of thermal information but also supports the effective application of SW algorithms, expanding the potential for improved thermal analysis.

For this study, Landsat-8's TOA observations across various spectral bands were utilized, including bands 1–4 (visible), band 5 (near-infrared (NIR)), bands 6–7 (shortwave infrared (SWIR)), and bands 10–11 (TIR), alongside surface temperature (LST) data. The inclusion of these diverse spectral bands allows for an in-depth monitoring and characterization of various surface attributes. Data spanning more than a decade (March 2013 to November 2023) were used to provide a broad temporal coverage for comprehensive analysis. The TOA observations were obtained from the USGS Landsat 8 Collection 2 Tier 1 TOA Reflectance dataset available on Google Earth Engine (GEE) (https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_TOA).

3.2.3 Satellite LST Data from Single-Channel (SC) and Split-Window (SW) Algorithms

The SC and SW methods are two widely used algorithms for deriving LST from TIR data. The SC method, as exemplified by the official Landsat-8 LST data, predominantly relies on the thermal data from TIR band 10. The SC process begins with an atmospheric correction for the TIR data, followed by an emissivity correction, and concludes with an SC inversion, utilizing the LUT method (Cook et al., 2014; Malakar et al., 2018). To compare the performance of the proposed algorithm with the SC method, Landsat-8 LST data were obtained from the USGS Landsat 8 Level 2, Collection 2, Tier 1 dataset provided by GEE (<https://developers.google.com/earth->

[engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2](#)). Quality attributes were also obtained and applied for effective cloud masking, which enhanced the reliability of the satellite data used in this study.

The SW method derives LST using observations from two TIR bands (typically channels 10 and 11 in Landsat-8). There are three frequently employed SW algorithms that exclusively rely on LSE and TOA observations (Meng et al., 2019; Ye et al., 2022): the generalized linear algorithm (Wan & Dozier, 1996), the nonlinear SW algorithm (Du et al., 2015; Wan, 2014), and the NOAA JPSS enterprise algorithm (Liu et al., 2019). The research on Landsat-8 (Meng et al., 2019) indicates that the nonlinear SW algorithm (Du et al., 2015; Wan, 2014) provides the highest accuracy in most cases. To compare the performance of the proposed algorithm with the SW method, the nonlinear SW algorithm was implemented in this study. The algorithm is expressed in the following equations.

$$LST = B_0 + \left(B_1 + B_2 \frac{1-\varepsilon}{\varepsilon} + B_3 \frac{\Delta\varepsilon}{\varepsilon^2} \right) \frac{T_{10} + T_{11}}{2} + \left(B_4 + B_5 \frac{1-\varepsilon}{\varepsilon} + B_6 \frac{\Delta\varepsilon}{\varepsilon^2} \right) \frac{T_{10} - T_{11}}{2} + B_7 (T_{10} - T_{11})^2 \quad (3)$$

where T_{10} and T_{11} are the Landsat-8 brightness temperatures of channels 10 and 11, respectively; ε and $\Delta\varepsilon$ are the LSE mean and difference, respectively. $\varepsilon = (\varepsilon_{10} + \varepsilon_{11})/2$, $\Delta\varepsilon = \varepsilon_{10} - \varepsilon_{11}$; B_i ($i = 0 \sim 7$) are the algorithm coefficients determined using simulated data. The improved NDVI-based threshold method (Cheng et al., 2018) and the vegetation cover method (Valor & Caselles, 2005) were selected to estimate the LSE for vegetated and non-vegetated areas, respectively.

3.3 Methodology

The extensive simulation data generated by RTMs (source domain) encompasses diverse atmospheric conditions and surfaces. These simulations were employed to train a comprehensive

ML model, which served as a starting point for the TL model. The pre-trained ML model was then further refined using a subset of real-world data (target domain), which included satellite and in-situ data, to develop the TL model. Another subset of real-world data was employed to predict and validate LST results from both the ML and TL models. Figure 3.2 provides the overall flowchart of the TL algorithm.

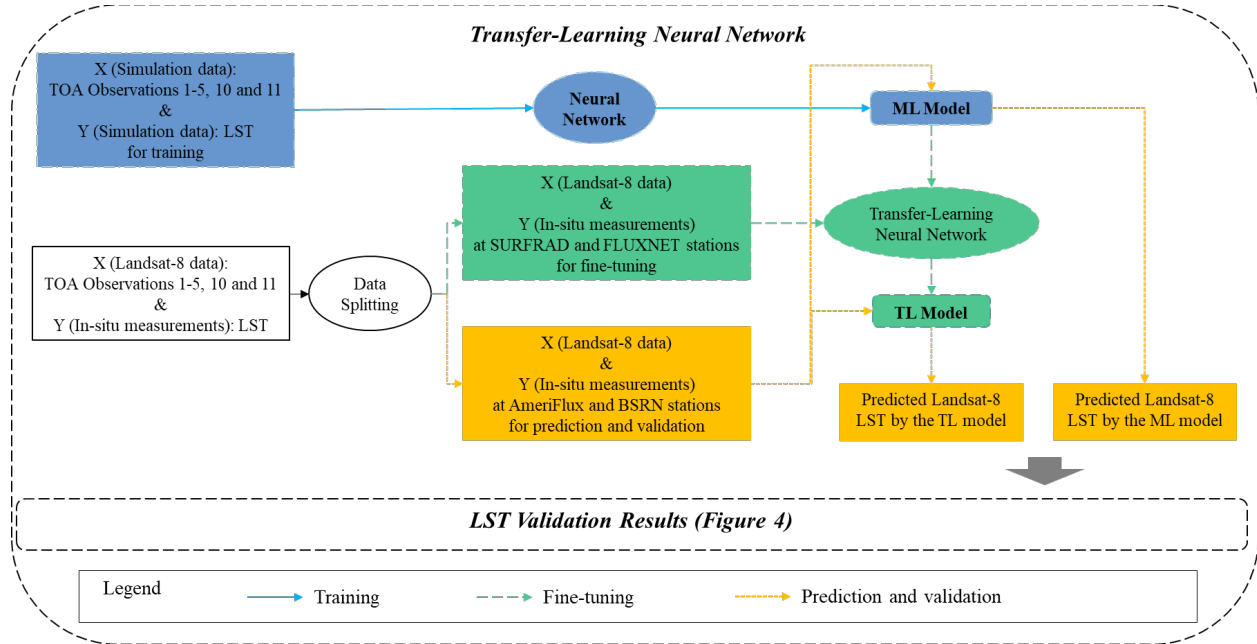


Figure 3.2 Flowchart of the transfer-learning (TL)-based method to retrieve land surface temperature (LST).

Section 3.3.1 details the simulation data generation process, while Sections 3.3.2 and 3.3.3 elaborate on the construction processes of the ML and TL models respectively.

3.3.1 Simulation Data Generation

To provide comprehensive inputs for the models, TOA observations in both longwave and shortwave spectra were utilized. Longwave measurements, particularly TIR data, are pivotal for capturing the radiation emitted from the Earth's surface, directly informing LST interpretations.

Conversely, shortwave data, encompassing the visible, NIR, and SWIR bands, reflect solar illumination and are instrumental in characterizing land surface inherent attributes.

In theory, it is possible to produce the all-wave simulation by running RTMs across the shortwave and longwave spectrum. However, it is computationally inefficient to achieve reasonably fine spectral intervals in both shortwave channels (in wavelength) and longwave channels (in wavenumber). As an alternative, simulations in shortwave and longwave channels are first conducted separately and then integrated to balance the requirement in spectral resolution and the computing demand. Table 3.1 presents the parameters used in the RTMs, and Figure 3.3 provides the schematic representation of the data generation and matching process.

Table 3.1 The parameters used in the radiative transfer models (RTMs).

Model Name	Parameters	Values
MODTRAN	LST (K)	AT0 - 5, AT0, AT0 + 5, AT0 + 10, AT0 + 15, AT0 + 20
	VZA (°)	0, 10, 20
libRadtran	AOD	0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5
	SZA (°)	0, 10, 20, 30, 40, 50, 60, 70, 80
	RAA (°)	0, 30, 60, 90, 120, 150, 180
	WV (cm)	0, 1.5, 3.0, 4.5, 6.0, 7.5, 9.0, 10.5
	Altitude (km)	0, 1, 2, 4, 6, 8
	VZA (°)	0, 10, 20

For the generation of longwave simulation data, MODTRAN 6.0 (Berk et al., 2014) was employed to generate longwave simulation data, as it is widely recognized for its accuracy in atmospheric transmission and radiative transfer simulations. Specifically, to simulate the TOA radiance of the Landsat-8 TIR channels (Band 10 and Band 11), several key parameters were considered, including emissivity, LST, water vapor (WV), altitude, and viewing zenith angle (VZA). These channels are centrally located at wavelengths of 10.89 μm for band 10 and 11.98 μm for band 11.

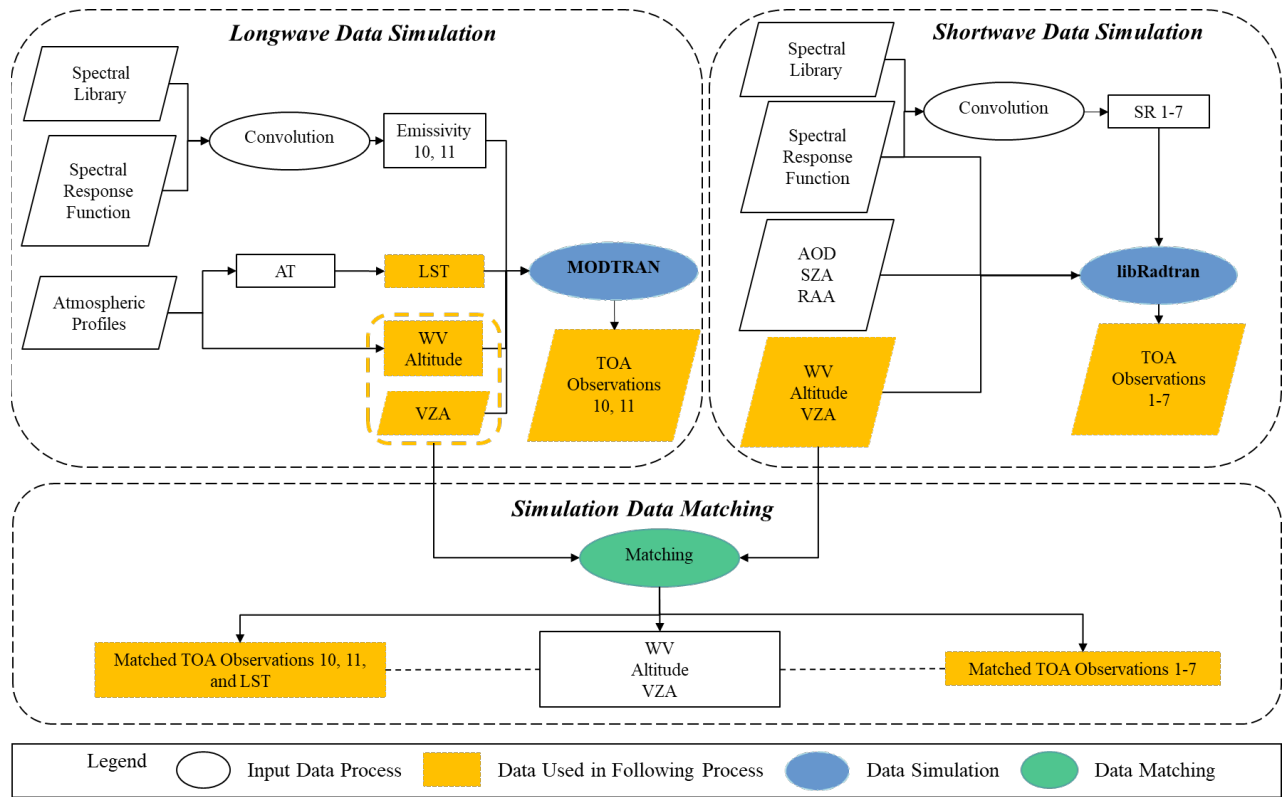


Figure 3.3 Generating and matching longwave and shortwave simulation data.

Emissivity values were derived from the ASTER spectral library (Baldrige et al., 2009) together with the spectral response functions (SRF) of the Landsat-8 TIR bands (https://nwp-saf.eumetsat.int/downloads/rtcoef_rtov13/ir_srf/rtcoef_landsat_8_tirs_srf.html). The selected spectra include a variety of types: soils (alfisol, aridisol, entisol, inceptisol, lunar, mollisol, and utisol), rocks (igneous, metamorphic, and sedimentary), vegetation (grass, and trees), and water (sea, distilled, ice, snow, and tap), etc. The atmospheric profiles, specifically regarding air temperature (AT), WV, and altitude, were informed by the clear-sky Global Atmospheric Profiles from Reanalysis Information (GAPRI) (Mattar et al., 2015). The base air temperature from these profiles (AT0) was used to set the LST values (Table 3.1), following the methodology outlined by Jiménez-Muñoz et al. (2014). Given the statistical range of VZA for Landsat-8 (0° to 20°), three

distinct VZA values were selected for simulations (Table 3.1). The detailed equation can be found at Berk et al. (1998), Berk et al. (2005), and Berk et al. (2014).

The simulation of shortwave data was conducted using libRadtran (Mayer & Kylling, 2005), recognized for its efficiency in modeling solar radiation transmission. This involved generating TOA reflectance values for Landsat-8's shortwave bands (1–7), requiring inputs such as SR, aerosol optical depth (AOD) at 550 nm, solar zenith angle (SZA), relative azimuth angle (RAA), WV, altitude, and VZA. The central wavelengths for these bands range from 0.44 μm to 2.20 μm . SR for bands 1–7 was calculated using spectral data from the ASTER spectral library (Baldrige et al., 2009) and the SRF of the Landsat-8 bands 1–7 (https://nwp-saf.eumetsat.int/downloads/rtcoef_rtov13/ir_srf/rtcoef_landsat_8_oli_srf.html). The AOD at 550 nm, SZA, RAA, WV, altitude, and VZA were set as Table 3.1, and input into libRadtran to simulate the necessary data (Mayer & Kylling, 2005). The detailed equation can be found at Mayer and Kylling (2005), and Emde et al. (2016). Optical atmospheric parameters were obtained from libRadtran, and the TOA reflectance for Landsat-8 Bands 1–7 was then calculated according to Equation 4.

$$R(\mu_0) = R_0(\mu_0) + \frac{\gamma_s \bar{\rho}}{1 - \gamma_s \bar{\rho}} \mu_0 \gamma(\mu_0) \quad (4)$$

Where, $R(\mu_0)$ is TOA reflectance, $R_0(\mu_0)$ is path reflectance, γ_s is SR, $\bar{\rho}$ is the spherical albedo, μ_0 is the cosine of the solar zenith angle, and $\gamma(\mu_0)$ is the total transmittance (Y. Zhang et al., 2018).

For both longwave and shortwave simulations, it was imperative to ensure consistency across key parameters such as WV, altitude, and VZA. This facilitated the matching of longwave and shortwave data sets, as depicted in Figure 3.3. The aligned data sets, encompassing matched LST

values and corresponding TOA observations for bands 1–7, 10, and 11, were then utilized for further analysis in the development of ML and TL models.

3.3.2 Machine Learning (ML) Modeling

The ML model was based on a neural network. It is a numerical model of the neural network of the brain, which uses interconnected nodes or neurons in a layered structure that mimics the structure of the human brain. The most important feature of neural networks is their ability to be "trained" and approximate complex relationships using learning algorithms (Takenaka et al., 2011).

To construct the models, simulations from RTMs were utilized as the source domain. TOA data from 1–7, 10, and 11 were available as independent variables (X). However, the inclusion of nine bands, although comprehensive, may introduce potential redundancies. To streamline the band selection and identify the optimal set of bands, a principal component analysis (PCA) was conducted. The first three principal components can explain 99.99% of the total variance. Specifically, bands 10 and 11 are predominant in the first and second principal components, while bands 1-5 significantly contribute to the third principal component. Consequently, the models focus on bands 1-5, 10, and 11. In the ML model, the simulated TOA observations 1–5, 10, and 11 are used as the independent variables (X), and the simulated LST is used as the dependent variables (Y), as shown in Figure 3.2.

To determine the optimal hyperparameters for the ML model, experiments were conducted with different combinations, balancing accuracy and computational efficiency. The result was the development of a four-layer neural network model, which included one input layer, two hidden layers, and one output layer. The input layer was designed with seven neurons, corresponding to

the seven selected TOA bands (1–5, 10, and 11), ensuring that each band's unique contribution to LST could be captured. The output layer consisted of a single neuron, representing the LST value. The hidden layers, comprising twenty-five neurons in the first layer and five neurons in the second, were meticulously calibrated to provide the neural network with sufficient complexity to model the relationships in the data while maintaining computational manageability.

This structured approach to ML modeling ensures that the neural network can effectively learn from the RTM-generated simulations, providing a robust foundation for accurate LST prediction and serving as a critical step toward the development of the TL model.

3.3.3 Transfer Learning (TL) Modeling

The TL algorithm utilizes the pre-trained model as a starting point, extracting and refining its features and weights, and fine-tuning them to produce an optimal model (H. Liang et al., 2019). This approach not only streamlines the model development process but also ensures adaptability to diverse datasets across different domains, thereby broadening the applicability and effectiveness of ML.

In this study, the TL paradigm was employed to adapt the previously developed ML model to better align with real-world observations, thereby transitioning from the source domain of simulated data to the target domain of satellite and in-situ measurements. This adaptation process involved integrating the learned features and weights from the ML model and meticulously fine-tuning them with a subset of real-world data, culminating in the development of an advanced TL model.

This refined TL model utilizes both theoretical simulations and empirical data. Traditional ML methods heavily rely on extensive training datasets. Simulations from RTMs are comprehensive

and cover a wide range of scenarios. Station data, representing real-world conditions, helped mitigate discrepancies between simulation data and actual observations. Fine-tuning the model with station data allowed the TL model to better approximate real-world conditions, thereby improving its predictive performance. The use of these datasets ensures both the quantity and quality needed for accurate and reliable predictions.

The real-world dataset, comprising Landsat-8 satellite observations alongside in-situ LST measurements, played a pivotal role in fine-tuning the TL model. To prevent overfitting and ensure robust model validation, the dataset was strategically divided into training and test subsets. The training data, obtained from SURFRAD and FLUXNET, were utilized to fine-tune the ML model, thereby establishing the TL model. Conversely, the test data, sourced from AmeriFlux and BSRN, were instrumental in evaluating the performance of both the ML and TL models.

Initially, the TL model adheres to the same structural framework as the pre-trained ML model, maintaining a consistent architecture that includes one input layer, two hidden layers, and one output layer. This architectural consistency ensures that the TL model retains the foundational strengths of the ML model while incorporating the necessary refinements to enhance its applicability to real-world data. Next, the learned features and weights were transferred from the ML model, with the first hidden layer frozen to preserve the general features learned during pretraining. Real-world data were then used for training, where Landsat-8 TOA observations (bands 1–5, 10, and 11) served as the independent variables (X), corresponding to seven neurons in the input layer, and the in-situ LST measurements were specified as the dependent variable (Y), corresponding to one neuron in the output layer. Finally, with the first hidden layer's weight fixed, the model undergoes further training with real-world data, allowing the weights to be refined and the model's accuracy to be enhanced.

The evaluation results of the predicted Landsat-8 LST by the ML and TL models are presented in Section 3.4.1. This evaluation not only highlights the model's enhanced predictive capabilities but also emphasizes the effectiveness of incorporating real-world data in improving the model's predictive accuracy.

3.4 Results

In this section, the performance of the proposed ML and TL models is assessed, and the proposed algorithm is compared with the SC and SW algorithms. Additionally, the spatial distribution of the retrieved LST is examined.

3.4.1 Performance of ML and TL Models

The efficacy of the ML and TL models in estimating LST was rigorously evaluated, utilizing a subset of in-situ measurement data. The validation sites are sourced from AmeriFlux and BSRN. There are 182 sites from AmeriFlux and 13 sites from BSRN. Detailed information, including surface type, elevation, latitude, and longitude, can be found in Table S1 in Appendices.

This evaluation aimed to assess the predictive accuracy and reliability of both models in rendering precise LST estimations from Landsat-8 satellite observations. Figure 3.4 depicts the validation results for Landsat-8 LST derived from these models.

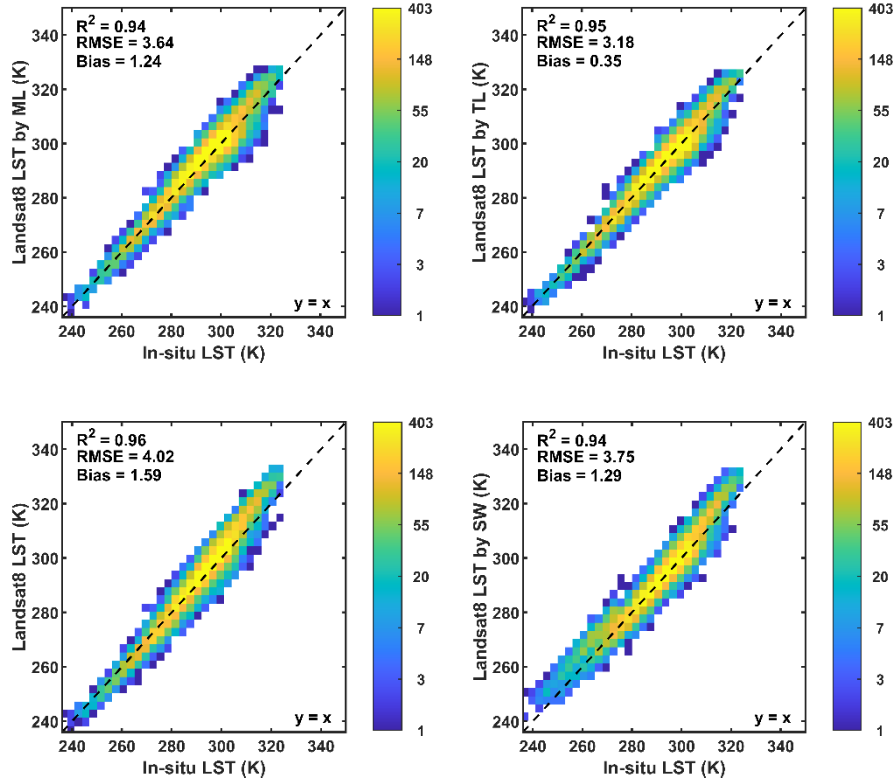


Figure 3.4 Comparison between the Landsat-8 land surface temperature (LST) product generated by the single-channel (SC) algorithm, the Landsat-8 LST data generated by the split-window (SW) algorithm, the Landsat-8 LST predicted by the machine learning (ML) model, and the Landsat-8 LST predicted by the transfer-learning (TL) model over in-situ measurements.

The performance metrics for the ML model showed an R^2 of 0.94, indicating a strong correlation between the predicted and observed LST values. The model also recorded a RMSE of 3.64 K, reflecting the average magnitude of the prediction errors, and a bias of 1.24 K, indicating the average deviation from the actual measurements.

Building upon this foundation, the TL model, which was fine-tuned using a subset of in-situ measurements, demonstrated an enhanced performance. The TL model achieved an R^2 of 0.95, an RMSE of 3.18 K, and a bias of 0.35 K. These improvements in the TL model, with a lower RMSE

and bias by 0.46 K and 0.89 K, respectively, underscore its superior predictive capability compared to the ML model.

These findings demonstrate the effectiveness of the TL approach in leveraging the foundational theoretical knowledge encapsulated within the pre-trained ML model, further enhancing its accuracy with the integration of real-world data. This highlights the promising potential of combining simulations from RTMs with real-world observational data to refine and enhance LST prediction models.

In recognition of its enhanced performance and potential for broader application, the TL model has been made publicly available on GEE. Interested readers can directly apply the code to estimate LST from Landsat 8 and 9 data.

3.4.2 Comparison with SC and SW Algorithms

In addition to evaluating the ML and TL models, comparisons were conducted with traditional LST retrieval methodologies, namely the SC algorithm (Cook et al., 2014; Malakar et al., 2018) and the SW algorithm (Du et al., 2015; Meng et al., 2019; Wan, 2014). Figure 3.4 also displays the validation results for the official Landsat-8 LST product and Landsat-8 LST data generated by the SW algorithm (Meng et al., 2019).

The Landsat-8 LST product generated by the official SC algorithm had an RMSE of 4.02 K, while the Landsat-8 LST data generated by the SW algorithm had RMSE values of 3.75 K. When compared, the LST estimations from ML and TL models demonstrated even more enhanced accuracy, with RMSEs of 3.64 K and 3.18 K, respectively.

This comparative analysis highlights that both ML LST and TL LST surpass the accuracy of the official Landsat-8 LST product. This success can be largely attributed to the models' exclusive

reliance on observed TOA data, combined with the strategic incorporation of shortwave data, which effectively enhances land surface information capture. In contrast, the official Landsat-8 LST products rely on auxiliary data such as LSE, and these intermediate data may introduce errors in the LST estimation process.

Furthermore, the accuracy of LST estimations by the proposed ML and TL models surpassed those derived from the SW algorithm. Although the ML model uses only shortwave and longwave TOA data, its accuracy is higher than the SW algorithm utilizing LSE, meaning that combining shortwave and longwave can eliminate the dependency on external LSE data. Further elaboration on this concept and its implications is provided in Section 3.5.1. In addition, the TL model demonstrated further enhanced accuracy over the SW algorithm. The TL model utilized both simulations and real-world data, where the simulation data provides a comprehensive theoretical basis and the real-world data enhances the model's alignment with actual conditions. Consequently, this method surpasses traditional techniques in accuracy, despite relying solely on TOA data, highlighting the potential of the proposed TL model in improving LST retrieval.

3.4.3 Spatial Distribution of Different LSTs

The spatial distribution of LSTs is also examined. For this purpose, two cloud-free Landsat images were selected: one from a winter day on January 23, 2022, and another from a summer day on July 5, 2014.

The LST maps generated by the TL method for these days are illustrated in Figure 3.5 (a) and (g). To evaluate the spatial distribution of LST derived using the TL method, LST products from the SC method, SW method, and MODIS dataset are also presented in Figure 3.5 (b) and (h), Figure 3.5 (c) and (i), and Figure 3.5 (d) and (j). To facilitate a visual assessment of their similarities,

histograms depicting the distribution of LST values are provided in Figure 3.5 (e) and (k). Furthermore, MODIS LC types are provided in Figure 3.5 (f) and (l) to show the diversity of LC types. The MODIS LC data is obtained from MODIS Land Cover Type Yearly Global 500m products (MCD12Q1.061). This product is downloaded from <https://lpdaac.usgs.gov/products/mcd12q1v061/>.

The spatial patterns revealed by the TL LST closely mirror those observed in SC LST, SW LST, and MODIS LST datasets, affirming the TL method's alignment with established LST estimations. Notably, the TL LST adeptly highlights temperature extremes, such as higher temperatures in cropland regions and cooler temperatures over small water bodies. This precision in capturing temperature variances attests to the TL model's ability to accurately represent geographical and thermal extremities.

Furthermore, the finer spatial resolution of the TL LST, alongside the SC and SW LST datasets, offers a more detailed observation of surface characteristics compared to the broader MODIS LST maps. This advantage underscores the importance of high-resolution data in detecting subtle geographical features and temperature nuances across diverse landscapes, enhancing our understanding and analysis of surface temperature distributions.

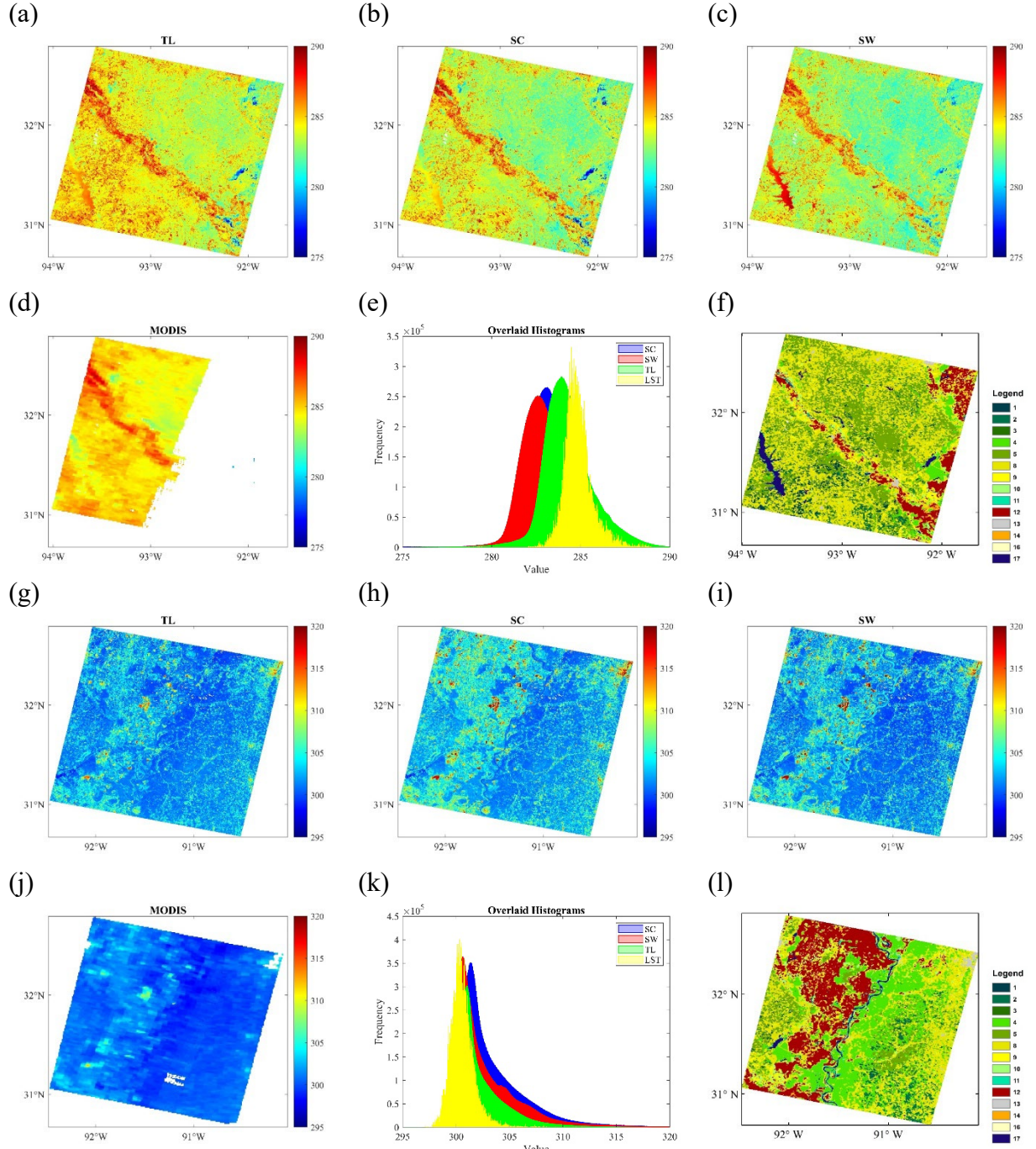


Figure 3.5 The spatial comparison of various winter (a–f) and summer (g–l) land surface temperature (LST) data. (a) and (g): transfer-learning (TL) LST; (b) and (h): single-channel (SC) LST; (c) and (i): split-window (SW) LST; and (d) and (j): MODIS LST. Additionally, (e) and (k): their histograms; and (f) and (l): MODIS land cover (LC) types (1: Evergreen Needleleaf Forests;

2: Evergreen Broadleaf Forests; 3: Deciduous Needleleaf Forests; 4: Deciduous Broadleaf Forests; 5: Mixed Forests; 8: Woody Savannas; 9: Savannas; 10: Grasslands; 11: Permanent Wetlands; 12: Croplands; 13: Urban and Built-up Lands; 14: Cropland/Natural Vegetation Mosaics; 16: Barren; 17: Water Bodies).

3.5 Discussion

This section further investigates the advantages of incorporating shortwave data into the model inputs, examines accuracy improvements across different LC types, and explores the application of the proposed algorithms to other Landsat missions.

3.5.1 The Effectiveness of Incorporating Shortwave Data

While shortwave signals are not directly linked to LST, they offer valuable information about surface emissivity within thermal ranges due to inherent correlations in surface properties across spectral regions.

Various researchers have explored the connection between shortwave parameters or indices and thermal emissivity. Notably, studies by Van de Griend and Owe (1993), Valor and Caselles (2005), and Tang et al. (2015) have demonstrated a significant correlation between LSE and the NDVI. Moreover, a physical relationship exists between LSE and spectral albedos or reflective variables, particularly over non-vegetated surfaces, as highlighted by Cheng et al. (2010), Cheng et al. (2018), and Meng et al. (2019). Therefore, incorporating shortwave data can enrich emissivity information.

To further illustrate the idea, the relationship between the shortwave TOA observations (bands 1–5) and thermal emissivity (band 10 or 11) was examined using the simulated data. As shown in Section 3.3.1, a substantial amount of simulation data was generated. Each record, as depicted in

Figure 3.3, incorporates shortwave TOA observations for different bands along with their corresponding thermal emissivities (Emissivity 10 and 11). These shortwave TOA observations and thermal LSE values were then used to train the ML method. The shortwave TOA observations were used as independent variables and the thermal LSE as dependent variables. An ML model based on a neural network was trained, allowing us to derive LSE by inputting TOA reflectances into this ML model. A scatter plot, depicted in Figure 3.6, showcases the relationship between actual LSE values and those predicted, revealing a notable correlation. This confirms that shortwave TOA data are intrinsically linked to thermal emissivity, suggesting that the integration of shortwave data into models can effectively reduce the reliance on direct LSE measurements.

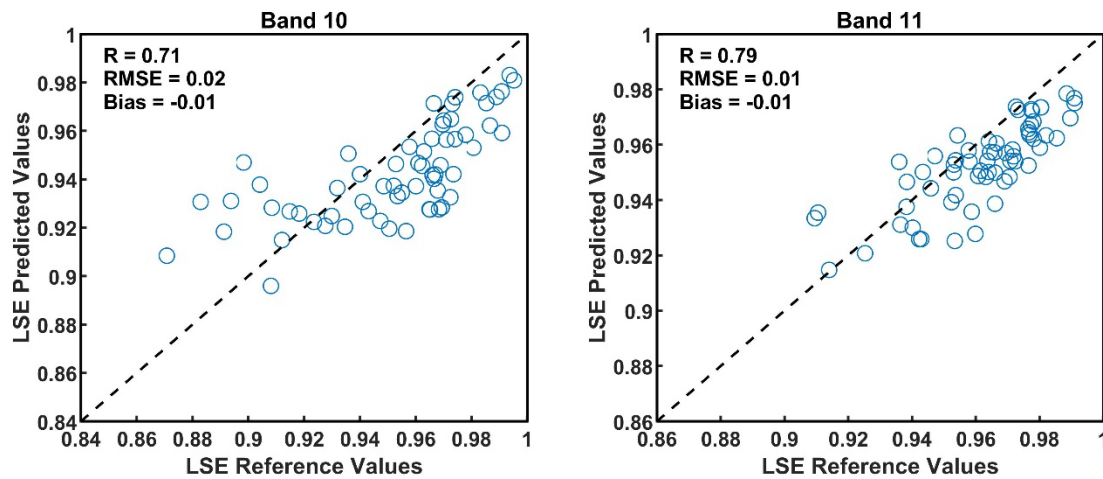


Figure 3.6 Comparison of thermal land surface emissivity (LSE) and predicted LSE from shortwave top-of-atmosphere (TOA) observations.

Additionally, longwave data play a crucial role in interpreting LST, whereas shortwave data provide valuable insights into land surface attributes, such as vegetation, moisture, and emissivity. These spectral bands cover different ranges, enabling the extraction of comprehensive information. According to Fahlen et al. (2021), the inclusion of additional channels can be beneficial for

improving the measurement of phenomena that encompass multiple spectral ranges, such as WV.

3.5.2 Accuracy Improvement for Different Land Cover Types

Our TL model's performance was evaluated across various LC types, comparing it with the official SC algorithm, SW algorithm, and ML algorithm. Their LST outputs were aligned with in-situ data to analyze RMSE, R^2 , and bias for different LC types, as detailed in Table 3.2.

Table 3.2 The performance of different algorithms on different land cover (LC) types.

LC Types	Number	Single-Channel Algorithm			Split Window Algorithm			Machine Learning Algorithm					
		SC - Landsat-8 Product			SW			ML			TL		
		RMSE (K)	R^2	Bias (K)	RMSE (K)	R^2	Bias (K)	RMSE (K)	R^2	Bias (K)	RMSE (K)	R^2	Bias (K)
Shrublands	1482	5.27	0.97	3.65	4.77	0.97	3.48	4.03	0.96	2.45	3.32	0.97	1.57
Croplands	2270	4.13	0.94	1.95	3.92	0.92	1.62	3.70	0.93	1.14	3.20	0.94	0.48
Forests	2140	3.29	0.95	0.61	2.85	0.94	0.15	3.14	0.94	0.87	2.85	0.94	-0.20
Grasslands	2784	4.32	0.94	1.84	3.91	0.93	1.12	3.77	0.92	0.70	3.41	0.94	-0.12
Permanent Wetlands	1657	3.30	0.96	1.19	3.66	0.93	1.17	3.31	0.95	1.25	2.84	0.95	0.16
Savannas	940	5.61	0.94	4.08	4.60	0.94	3.03	4.88	0.92	3.49	4.06	0.93	2.50
Barren Sparse Vegetation	142	3.07	0.97	1.10	5.21	0.94	3.81	2.99	0.97	1.63	2.40	0.97	0.59
Snow and Ice	36	5.31	0.87	3.50	5.57	0.74	3.70	5.34	0.86	3.75	4.17	0.87	2.60
Water Bodies	32	2.56	0.72	-0.81	3.29	0.51	0.19	2.55	0.71	0.56	2.52	0.72	-0.62
Rock	62	2.16	0.96	0.55	2.11	0.97	0.73	2.17	0.96	0.51	2.11	0.96	-0.16
Tundra	434	2.93	0.96	-0.71	5.51	0.88	2.68	2.67	0.96	-0.93	2.64	0.96	-0.20
Desert Gravel	188	5.25	0.89	4.37	3.88	0.91	2.96	2.80	0.85	1.32	2.09	0.90	0.45
All LC Types - Overall	12167	4.02	0.96	1.59	3.75	0.94	1.29	3.64	0.94	1.24	3.18	0.95	0.35

The SC algorithm's performance varied widely, with RMSE values ranging from 2.16 K to 5.61 K, R^2 values between 0.72 and 0.97, and biases from -0.81 K to 4.37 K. Notably, shrublands and savannas exhibited higher RMSEs at 5.27 K and 5.61 K, respectively, potentially due to the inherent uncertainties in atmospheric correction or errors in LSE estimation. The snow and ice LC types presented an RMSE of 5.31 K, attributed to the challenges in distinguishing clouds from snow and the issue of incomplete snow coverage at the pixel scale, as suggested by Duan et al. (2021).

The SW algorithm showed a slight improvement with RMSEs spanning 2.11 K to 5.57 K, R^2 from 0.51 to 0.97, and biases between 0.15 K and 3.81 K. In tundra regions, the RMSE of Landsat-8 LST generated by the SW algorithm is 5.51 K, likely due to inaccuracies in LSE. The accuracy of the SW algorithm mainly depends on the confidence of LSE. Similar to the SC algorithm, snow and ice regions had a considerable RMSE of 5.57 K, with similar reasons to the SC algorithm at snow and ice regions.

The ML algorithm's performance ranged with RMSEs from 2.17 K to 5.34 K, R^2 from 0.71 to 0.97, and bias from -0.93 K to 3.75 K. The TL model, building upon the ML algorithm, demonstrated notable improvements, achieving RMSEs from 2.09 K to 4.17 K, R^2 from 0.72 to 0.97, and biases from -0.62 K to 2.60 K, indicating the highest accuracy among the LC types when compared to the LST products from SC, SW, and ML results.

The RMSE difference between the SC algorithm and the TL algorithm has a range from 0.04 K to 3.16 K, with desert gravel showing the largest RMSE difference at 3.16 K. The low accuracy of the SC algorithm in desert gravel may be caused by the large LSE uncertainty. Desert regions exhibit significant LSE variations, influenced by aspects such as soil composition, texture, and moisture content (Yu et al., 2006). The challenges in precisely determining LSE result in high

uncertainties in LST retrieval for desert surfaces (Z.-L. Li et al., 2013). In contrast, the TL algorithm, which relies exclusively on TOA data, is not influenced by these LSE-related uncertainties.

Comparatively, the RMSE difference between the SW algorithm and the TL algorithm did not exceed 2.87 K. Except for barren sparse vegetation and tundra, the RMSE difference between the SW algorithm and TL algorithm did not exceed 1.79 K. For barren sparse vegetation and tundra, their RMSE differences are 2.81 K and 2.87 K, respectively, likely a result of accumulated LSE errors. However, the TL model, by relying solely on TOA data, remains unaffected by such error accumulation.

Lastly, the RMSE difference between the ML and TL models ranges from 0.03 K to 1.17 K, with the TL model providing more accurate LST estimates than the ML model. This enhanced accuracy in the TL model stems from utilizing the simulation data and real-world data. The simulations provide a comprehensive basis, and incorporating real-world data enhances the model's alignment with actual conditions. Additionally, despite the exclusion of specific LC types such as rock, tundra, desert gravel, barren sparse vegetation, snow and ice, and water bodies from the fine-tuning dataset, the TL approach exhibited good performance across these categories. This indicates that the TL model is effectively generalizing.

3.5.3 Application to Other Landsat Missions

Our methodology has been extended to encompass additional Landsat missions, particularly the latest Landsat-9 satellite, which was launched on September 27, 2021. This satellite is equipped with the advanced Operational Land Imager-2 (OLI-2) and Thermal Infrared Sensor-2 (TIRS-2) instruments, continuing the legacy of its predecessors in providing high-quality earth observation

data.

This analysis specifically focused on a case study using Landsat-9, covering the period from October 2021 to November 2023. The TOA observations for Landsat-9 were acquired from the USGS Landsat-9 Collection 2 Tier 1 TOA Reflectance dataset available through GEE (https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC09_C02_T1_TOA#bands), while the Landsat-9 LST data was sourced from the corresponding USGS Landsat-9 Level 2, Collection 2, Tier 1 dataset, also available through GEE (https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC09_C02_T1_L2). Quality attributes essential for accurate cloud masking were also downloaded and applied to the datasets.

In evaluating the effectiveness of various LST estimation algorithms on Landsat-9 data, the official SC algorithm, the SW algorithm, the ML algorithm, and the TL algorithm were included. The validation outcomes, illustrated in Figure S1 in Appendices, reveal a striking consistency with the results obtained from Landsat-8 data.

The LST estimations derived from the official SC algorithm for Landsat-9 data recorded an RMSE of 4.20 K, whereas the SW algorithm demonstrated a slightly improved RMSE of 3.74 K. In contrast, the developed ML and TL models yielded even more favorable RMSEs of 3.70 K and 3.32 K, respectively. This superior accuracy of the TL model highlights the efficacy of the proposed TL methodology. Strategically incorporating shortwave data significantly reduces reliance on explicit LSE data, while the combined use of simulations and real-world observations greatly enhances the precision of LST estimations.

The code for estimating LST from Landsat data is available on Google Earth Engine at <https://code.earthengine.google.com/6bd47d83730403dd4ae9f55661622162> (Landsat 8) and

<https://code.earthengine.google.com/1725716fcfdf724c362db3d6641f1226> (Landsat 9).

3.6 Conclusion

In environmental studies, the Landsat program has played a critical role by offering long-term, high-resolution LST data, which is crucial for assessing urban heat islands, climate change, evapotranspiration, hydrological cycles, and vegetation monitoring. However, conventional approaches for LST retrieval from satellite data face challenges in decoupling LST and LSE and often rely on auxiliary LSE data. Meanwhile, traditional ML methods require extensive training data and struggle to utilize information from diverse domains such as simulation and the real world. In this study, these limitations were addressed by exclusively utilizing shortwave and longwave TOA data without auxiliary data and proposing a novel TL approach that allows for the combined use of simulated and real-world data.

The essence of longwave data in interpreting LST and the invaluable perspective provided by shortwave data on land surface characteristics underscore the benefit of integrating the two types of satellite data. Despite shortwave signals not being a direct measure of LSE, empirical analyses, supported by existing literature, highlight a notable connection between shortwave inputs (like shortwave TOA observations, NDVI, spectral albedos, or reflective variables) and thermal LSE. This integration of shortwave data not only enriches LSE information but also diminishes the dependency on direct LSE measurements, addressing a significant challenge in traditional LST retrieval frameworks.

The fusion of comprehensive knowledge from RTM simulations with the nuanced realities captured in actual observations amplifies both the generalization capabilities and the accuracy of the models. In the TL framework, an ML model was first developed using simulated data generated

by RTMs, which was then used to initialize and refine the TL model with real-world data. This approach overcomes the traditional ML methods' limitations in bridging data from different sources.

Our evaluation encompassed a decade of Landsat observations and in-situ measurements across diverse geographical regions and LC types. The comparative assessment of ML model (RMSE: 3.64 K) and TL model (RMSE: 3.18 K) against the conventional SC algorithm (RMSE: 4.02 K) and SW algorithm (RMSE: 3.75 K) highlighted the superiority of TL models. The TL model's consistent outperformance across varied LC types further emphasizes its robustness, notably overcoming the SC algorithm's limitations in desert gravel due to LSE uncertainties and the SW algorithm's inaccuracies in barren sparse vegetation and tundra due to LSE error accumulation. Moreover, in the analysis of LST errors using simulated data, the TL algorithm demonstrated smaller LST errors compared to the SW algorithm for the same level of emissivity error, indicating that the TL method is more robust under these conditions.

This study presents a novel framework to estimate LST purely from satellite TOA data without external input of LSE information. The TL algorithm, by utilizing both shortwave and longwave data, minimizes the reliance on LSE and is immune to LSE-related errors or their accumulation. By utilizing RTMs' theoretical simulations and actual observational data, the TL algorithm boosts the model's accuracy and generalization capabilities.

Our observations indicate that the quality of LST retrieval varies significantly under different WV conditions, with the TL algorithm showing relatively larger errors at higher WV levels. Additionally, while the ASTER GED dataset offers comprehensive spatial coverage and relatively high spatial resolution, its coarse temporal resolution may miss seasonal emissivity changes, introducing LST errors in sparsely to moderately vegetated areas. Future research will incorporate

WV data or satellite datasets with WV bands, as well as emissivity products with higher temporal resolution, to further refine the proposed methodology.

The TL-based algorithm has been implemented on GEE and made freely available for estimating LST from Landsat 8 and 9 data. The retrieved LST data generated by the proposed method can provide valuable support for research in environmental studies.

Chapter 4: AI-Driven Time-Series Data Integration to Enhance Variable Prediction: A Case Study with GPP

4.1 Background

Time-series modeling is essential for understanding and predicting ecosystem processes that evolve continuously over time. Forest ecosystems play a central role in these long-term processes, particularly in regulating the global carbon cycle through carbon uptake, storage, and exchange with the atmosphere (Grace, 2004; Lorenz, 2010; Malhi et al., 2002). Quantifying forest carbon fluxes at regional to global scales is essential for improving our understanding of carbon–climate feedbacks and for supporting climate mitigation and forest management (Haya et al., 2023; Hurteau, 2021). Among key carbon flux variables, GPP is widely used to characterize ecosystem carbon uptake and to benchmark terrestrial biosphere models (Seiler et al., 2022; Wang et al., 2020).

Earth observation has substantially expanded our ability to monitor forest ecosystems (Hirschmugl et al., 2017; Holzwarth et al., 2020). Satellite remote sensing provides repeated observations over time at large spatial scales related to forest structure and composition, such as land cover, PFTs, and canopy properties (Zadbagher et al., 2023). Meanwhile, eddy-covariance flux towers offer direct and high-frequency measurements of ecosystem carbon exchange at local scales (Baldocchi, 2003). In parallel, process-based ecosystem models provide physically and ecologically interpretable frameworks to simulate carbon dynamics under comprehensive environmental conditions. In particular, the ED model has been developed over the past two decades and has been widely adopted in large-scale carbon monitoring efforts (Hurt et al., 1998; Ma et al., 2021; Moorcroft et al., 2001).

Despite these advances, effectively integrating multi-source time-series data, including satellite

observations and flux tower measurements, into process-based ecosystem modeling remains challenging (Xu et al., 2025). First, most process-based models, including the ED model, rely on process-based formulations. These models rely on simplified or fixed assumptions, which limit their ability to refine their predefined structures using real-world observations. Second, many remote sensing products provide aggregated, high-level information, whereas ecosystem models represent ecosystem dynamics through detailed internal structures, including PFT-specific or age-specific processes. This granularity mismatch limits the direct use of satellite observations to correct internal model states. For example, ecosystem models often simulate vegetation dynamics through a set of subprocesses corresponding to different age groups, while satellite observations typically provide aggregated ecosystem-level information. Third, flux tower observations, although highly accurate, are spatially sparse and often temporally incomplete relative to the multi-decadal duration of ecosystem simulations, as illustrated in Figure 4.1.

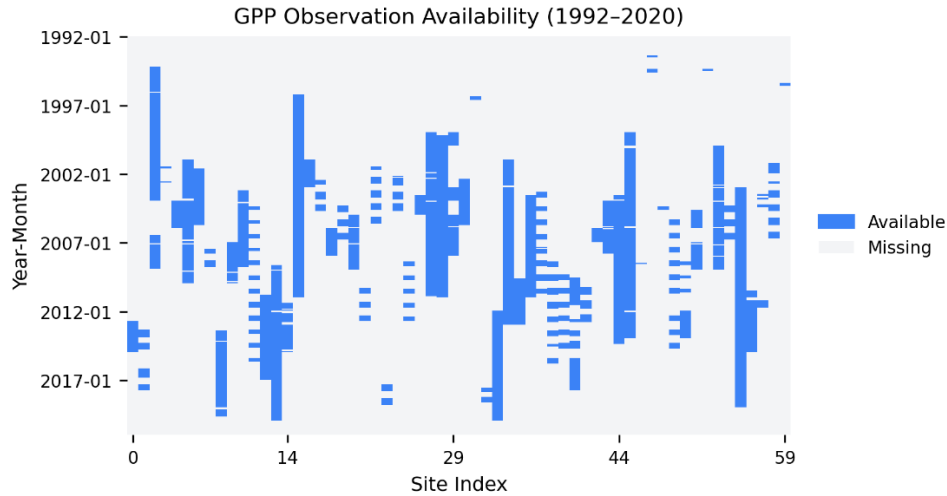


Figure 4.1 Example of in-situ data availability from the ABoVE dataset.

Recent studies have explored deep learning models for learning temporal patterns and performing time-series prediction. Transformer-based architectures have become widely adopted due to their

ability to capture long-range dependencies (Devlin et al., 2018; Dosovitskiy et al., 2021; Vaswani et al., 2017), outperforming earlier models based on recurrent structures (Chen, Kalanat, et al., 2023; Lai et al., 2018; Wang et al., 2023; L. Xu et al., 2024). Several variants of transformer-based models have been proposed to further improve performance. Examples include ProbSparse self-attention in Informer (Zhou et al., 2021), frequency-domain attention in FEDFormer (Zhou et al., 2022), cross-feature and cross-time modeling in Crossformer (Yunhao & Junchi, 2023), exogenous feature integration in TimeXer (Y. Wang et al., 2024), time-variable inversion in iTransformer (Liu et al., 2023), and multivariate modeling in SimpleTM (Chen et al., 2025). Although these models have shown strong performance in general forecasting tasks, they are primarily data-driven and often do not explicitly incorporate physical knowledge from process-based models. This limitation can affect their performance when only limited in-situ observations are available for large-scale environmental applications.

ML models have also been explored as emulators of process-based simulations. These approaches aim to approximate the behavior of complex physical models using deep learning techniques. For example, learning-based emulators have been developed to approximate climate models (Mooers et al., 2021; Rasp et al., 2018; X. Wang et al., 2022; S. Yu et al., 2024) and weather forecasting systems (Bonev et al., 2023; Kurth et al., 2023; Lam et al., 2023). Previous work has also explored learning-based emulators for ecosystem and carbon cycle simulations at the global scale (Li et al., 2026; Wang et al., 2025; Wang et al., 2023). However, these approaches mainly focus on improving computational efficiency rather than integrating physical knowledge with observational data to enhance predictive accuracy.

More broadly, KGML has emerged as an important direction for scientific modeling (Willard et al., 2022). Early approaches mainly focused on residual learning between process-based

simulations and observations (Forssell & Lindskog, 1997; Wan et al., 2018; Xu & Valocchi, 2015). More recent strategies incorporate physical knowledge through simulation-based pretraining (Ham et al., 2019; Hurtado, 2018; Read et al., 2019; Sultan et al., 2018; Yu et al., 2025), physics-constrained loss functions (X. Jia et al., 2020; Zongyi Li et al., 2024; Raissi et al., 2019; Read et al., 2019; R. Yu et al., 2024), architectural modifications inspired by physical processes (Anderson et al., 2019; Feng et al., 2022; Muralidhar et al., 2018), or multi-model integration (Chen, Kalanat, et al., 2023; Chen, Xie, et al., 2023; He et al., 2023; X. Jia et al., 2021; Z. Wang et al., 2024). While these methods improve interpretability and generalization, they focus on directly matched variables between simulations and observations. They typically do not consider satellite-based indirect labels, where high-level observations are not directly compatible with the subprocess-level representations used in ecosystem models.

Despite these advances, an important gap remains. Satellite-derived PFT products are available at large spatial scales and long time spans, but they are difficult to use for constraining internal ecological processes in age-structured ecosystem models. To address this gap, this work proposes a KGML framework that bridges indirect remote sensing observations and process-based ecosystem representations. The framework introduces a decomposition-and-resembling design to represent intermediate ecological processes and connect them with aggregated satellite PFT observations. In addition, to better utilize sparse and temporally incomplete flux tower observations, this work introduces a probabilistic label expansion strategy to estimate missing observations with associated uncertainty. In summary, this study makes the following contributions:

- A knowledge-guided decomposition-and-resembling framework that links aggregated satellite observations to intermediate ecological processes underlying carbon flux prediction.

- A multi-source integration strategy that integrates process-based simulations, satellite-derived PFT observations, and in-situ flux tower measurements.
- A diffusion-based probabilistic label expansion module with uncertainty awareness that increases the effective temporal coverage of flux tower supervision.

Together, these contributions provide a unified approach for integrating physical modeling and real-world observations to enhance large-scale carbon flux prediction and ecosystem monitoring.

4.2 Data

Multi-source datasets were used in this study, including initial states and auxiliary drivers, process-based model simulations, remote sensing PFT observations, and in-situ measurements. Figure 4.2 provides an overview of datasets used in this research. The goal of this study is to predict carbon flux ecosystem variables across large spatial domains and over long temporal periods.

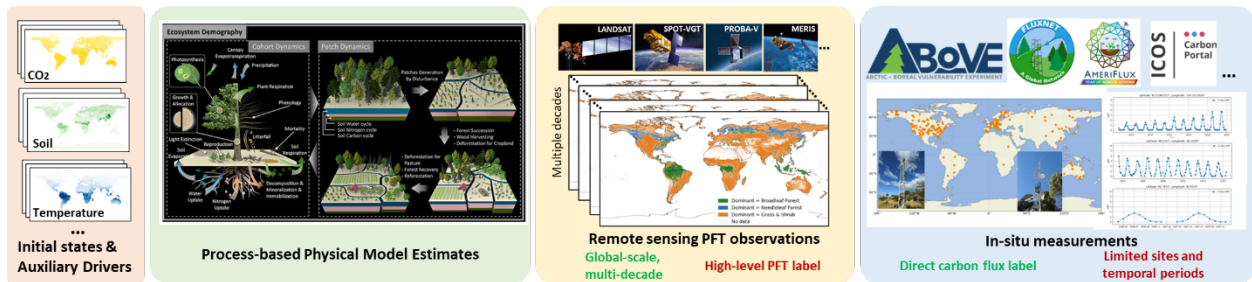


Figure 4.2 Multi-source data.

These datasets provide three complementary sources of information for ecosystem carbon flux prediction. Process-based model simulations provide physically consistent and spatially complete estimates across large spatial domains and long time periods, driven by environmental inputs such as meteorological conditions, soil properties, and atmospheric CO₂ concentration, but they may contain biases due to structural assumptions and uncertain internal ecological states. Flux tower observations provide direct measurements of ecosystem carbon fluxes, but they are available only

at a limited number of sites and often contain temporal gaps. Satellite-derived PFT observations provide large-scale information on vegetation composition with broad spatial coverage and long temporal records, but these observations are typically aggregated and not directly compatible with the subprocess-level representations used in ecosystem models.

Process-based simulations from the ED model (Hurtt et al., 1998; Ma et al., 2021; Moorcroft et al., 2001) were utilized in this study. As mentioned in Section 2.3.1, the ED model has been widely used to estimate the global carbon budget and support major carbon monitoring initiatives such as Global Carbon Budget (Friedlingstein et al., 2023; Friedlingstein et al., 2024) and NASA's Carbon Monitoring System (Hurtt et al., 2019). The model takes initial states and auxiliary environmental drivers as inputs, including meteorological variables, soil properties, and atmospheric CO₂ concentration. Meteorological data were obtained from NASA's Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al., 2017) (<https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/>). Soil property data were obtained from the ROSETTA database (Montzka et al., 2017) (<https://doi.pangaea.de/10.1594/PANGAEA.870605>). Surface CO₂ concentration data were obtained from the NOAA CarbonTracker database (Peters et al., 2007) (<https://gml.noaa.gov/ccgg/carbontracker/>). Using these inputs, the ED model can simulate multiple ecosystem variables. In this study, time-series carbon flux variables such as GPP were used as the primary prediction targets. To enforce carbon balance as a physical constraint, ecosystem respiration (RECO) and net ecosystem exchange (NEE) were also predicted.

In addition, satellite-derived PFT information from the ESA CCI product (Harper et al., 2023) was used to represent vegetation composition. As discussed in Section 2.3.2, these satellite observations provide global coverage and can help improve ecosystem variable prediction by providing additional constraints on vegetation structure.

In-situ carbon flux datasets were also collected from the ABoVE dataset (Bill et al., 2023) and the CarbonSense benchmark (Fortier et al., 2024). The CarbonSense benchmark integrates data from several major flux tower networks under diverse environmental conditions. In this study, representative datasets from the ABoVE, AmeriFlux, FLUXNET, and ICOS-WW networks were selected. Many sites contain overlapping records across different networks, and these sites were merged to form a combined dataset referred to as “Multiple”. Figure 4.3 shows the geographic distribution of sites across the different datasets.

For model training and evaluation, each dataset was spatially divided into training and testing sets, with 80% of sites used for training and the remaining 20% used for testing. This split ensures site-level separation between training and testing.

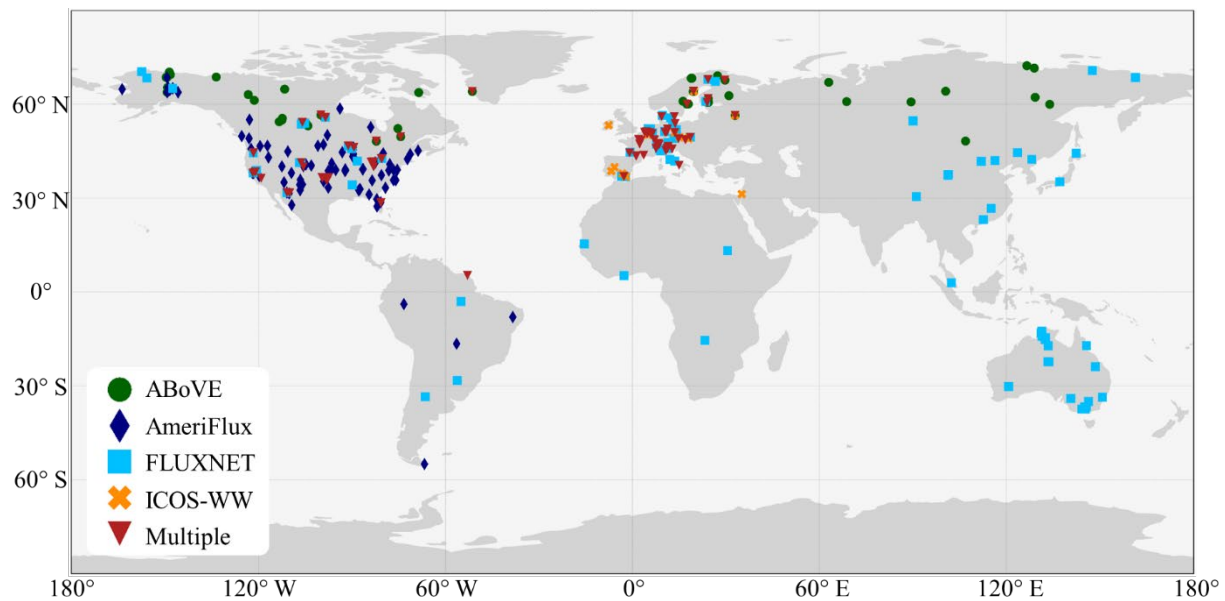


Figure 4.3 Spatial distribution of in-situ observation sites.

4.3 Methodology

To enable effective integration of indirect remote sensing observations with process-based

ecosystem modeling, a knowledge-guided learning framework with a decomposition-and-resembling structure (DERE) was established. The framework links ecosystem simulations with multiple observational constraints, including satellite-derived PFT observations and flux tower measurements, to improve large-scale carbon flux estimation.

The core challenge addressed by this framework lies in the mismatch between observation types and model representations. Process-based ecosystem models simulate carbon dynamics through age-structured cohorts and PFT-specific processes. In contrast, satellite observations typically provide aggregated PFT composition. Meanwhile, flux tower measurements offer direct observations of carbon exchange but are spatially sparse. As a result, these observational sources cannot be directly incorporated into process-based ecosystem models without mechanisms that align observational information with internal ecological processes.

To address this challenge, the proposed methodology consists of three main components, as illustrated in Figure 4.4 using GPP as an example.

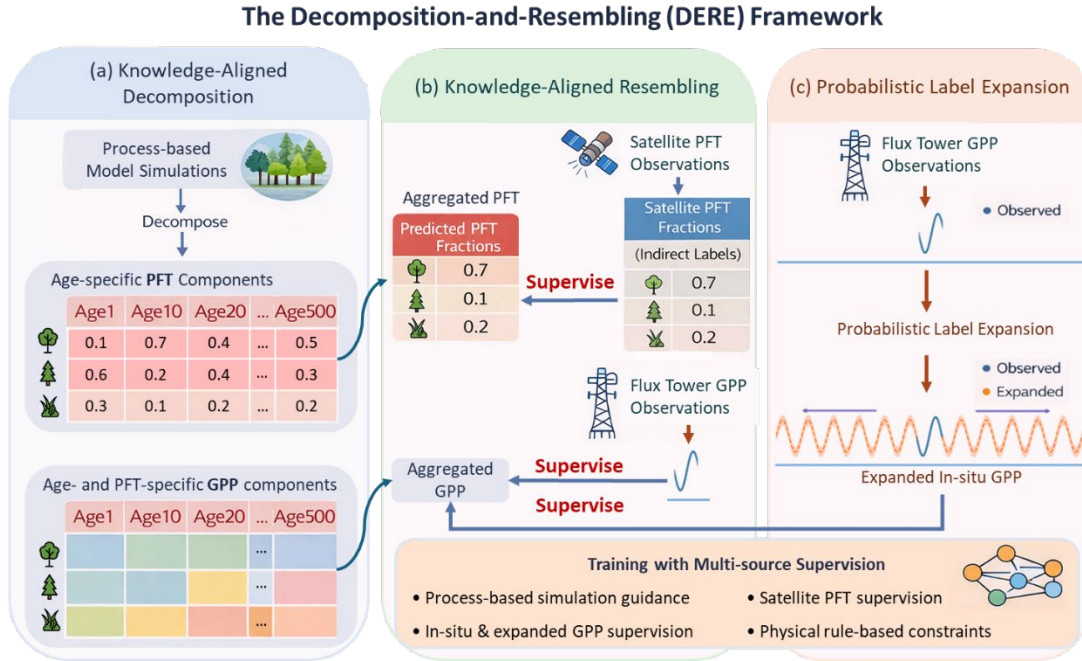


Figure 4.4 Overview of the proposed decomposition-and-resembling (DERE) framework for integrating process-based ecosystem simulations with satellite-derived plant functional types (PFTs) and flux tower observations, illustrated using gross primary productivity (GPP) as an example.

First, a knowledge-aligned decomposition explicitly represents intermediate ecological processes within age-structured ecosystem simulations. In this step, carbon flux variables are decomposed into PFT-specific contributions across different age groups, producing intermediate representations that include age-specific PFT fractions and age–PFT carbon flux contributions. This representation makes the internal ecological structure, including PFT composition and age-related processes, available for subsequent integration with observational data. Second, a knowledge-aligned resembling step integrates real-world observations with ecosystem processes. The age-specific PFT fractions derived from the decomposition step are aggregated across age groups to obtain grid-level PFT composition, which can then be aligned with satellite-observed

PFT fractions. Through this alignment, satellite observations provide indirect constraints on the internal ecological representation while maintaining consistency with ecosystem simulations. In addition, flux tower measurements are used as supervision for carbon flux prediction during model training. Finally, a probabilistic label expansion step with uncertainty-aware fine-tuning is introduced to better utilize temporally sparse flux tower records. This step estimates missing observations and incorporates the associated uncertainty to guide model training.

4.3.1 Knowledge-aligned decomposition: explicit representation of intermediate ecological processes

Process-based ecosystem models simulate carbon dynamics through multiple interacting ecological processes, including vegetation growth, competition among PFTs, and age-dependent ecosystem development. However, large-scale ecosystem simulations, such as those produced by the ED model, typically archive only carbon flux outputs. Intermediate ecological states, including PFT-specific contributions and age-cohort structure, are typically not archived in large-scale ecosystem simulations because storing these variables would significantly increase storage requirements and regenerating them requires substantial computational resources.

Many observational datasets, particularly satellite-derived forest composition products, are more directly related to these intermediate ecological processes than to aggregated carbon flux variables. Satellite observations often describe vegetation structure, such as the proportions of different PFTs, which reflect internal ecosystem composition rather than final carbon exchange.

To enable meaningful integration with satellite-based observations, the proposed framework introduces a knowledge-aligned decomposition step that explicitly represents intermediate ecological processes within ecosystem simulations. In this step, grid-level carbon fluxes are

separated into contributions associated with different PFTs and forest age groups. This representation captures how PFT-specific physiological characteristics and age-dependent ecosystem dynamics contribute to carbon flux variables such as productivity and respiration. The decomposed PFT-specific contributions remain consistent with the original grid-level carbon fluxes, since they can be recombined through proportion-weighted aggregation and competitive interactions in the subsequent resembling step.

Through this decomposition step, internal ecological structures such as PFT composition and age-related ecosystem processes are explicitly represented while remaining consistent with the behavior of process-based ecosystem simulations.

The resulting intermediate ecological representation serves as a structural bridge between ecosystem modeling outputs and satellite-derived vegetation composition observations. It provides a physically interpretable internal state that can later be constrained by large-scale remote sensing data.

4.3.2 Knowledge-aligned resembling: integrating satellite observations with ecosystem processes

After intermediate ecological processes are explicitly represented through the decomposition step, the next stage integrates observational data from multiple sources to constrain both ecosystem structure and carbon flux predictions.

Flux tower measurements provide direct observations of ecosystem carbon exchange, including variables such as GPP, NEE, and RECO. These measurements are highly accurate and process-relevant, but they are available at only a limited number of sites and often have short temporal coverage. In the proposed DERE framework, flux tower observations are used to constrain the

predicted carbon flux variables during model training, ensuring that the model remains anchored to physically observed ecosystem exchange.

In contrast, satellite observations provide broad spatial coverage and long-term information about vegetation composition, particularly PFT composition. These observations are available on a global scale and therefore contain valuable large-scale ecological information. However, satellite PFT products represent aggregated vegetation composition at the grid scale and do not resolve age-specific, or cohort-level processes represented in ecosystem models.

To address this difference in representation, the intermediate ecological outputs derived from the decomposition step are aggregated across age states to obtain grid-level estimates of vegetation composition. These model-derived estimates can then be directly compared with satellite observations of PFT fractions.

By enforcing consistency between satellite-observed PFT composition and the composition derived from ecosystem simulations, satellite observations provide indirect but physically meaningful constraints on internal ecological processes. During model training, two types of observational supervision are applied simultaneously: flux tower measurements constrain the predicted carbon flux outputs, while satellite observations constrain vegetation composition derived from the intermediate ecological representation. Through this joint supervision, the framework improves both predictive accuracy and ecological consistency.

4.3.3 Probabilistic label expansion and uncertainty-aware fine-tuning

A major limitation of flux tower observations is incomplete temporal coverage. Many towers operate only during part of the simulation period due to installation timing, maintenance interruptions, or data quality issues. As a result, observational records often cover only a fraction

of the multi-decadal duration of ecosystem simulations.

To enhance the effective use of available in-situ observations, the proposed framework introduces a probabilistic label expansion strategy to estimate missing carbon flux measurements together with associated uncertainty. Instead of performing deterministic gap filling, the method generates multiple plausible values for missing flux observations based on the available records within each site's time series. This probabilistic reconstruction is implemented based on the Conditional Score-based Diffusion Imputation (CSDI) framework (Tashiro et al., 2021), which enables the generation of realistic samples conditioned on partially observed time series.

Probabilistic reconstruction produces expanded supervisory signals along with predictive uncertainty. The estimated uncertainty is then incorporated into the training process through uncertainty-aware fine-tuning. Expanded labels associated with lower predictive uncertainty exert stronger influence during training, while those with higher uncertainty are assigned to lower weights. This strategy increases the usable temporal coverage of flux tower supervision while maintaining robustness and avoiding over-reliance on uncertain estimates.

Overall, the framework is trained using three complementary sources of information: process-based ecosystem simulations, satellite-derived vegetation composition observations, and flux tower measurements. Process-based simulations provide physically consistent large-scale ecosystem dynamics, satellite observations constrain vegetation structure across broad spatial domains, and flux tower data anchor carbon flux predictions at observed sites. By jointly utilizing these complementary data sources, the framework improves predictive accuracy for large-scale carbon flux estimation.

4.4 Results

4.4.1 Candidate Methods

In this study, four categories of candidate methods are included for comparison: (1) the process-based ED model; (2) several advanced time-series ML models; (3) knowledge-guided extensions of these time-series ML models; and (4) the proposed method, DERE. For training, the hyperparameter settings recommended in the original studies were adopted when available. All models were trained until convergence using early stopping with a patience criterion based on validation performance. The validation set comprised 10% of the training data and remained independent from the test dataset.

- Transformer (Vaswani et al., 2017): The vanilla Transformer serves as the fundamental sequence modeling framework with self-attention to help capture long-range temporal dependencies.
- Informer (Zhou et al., 2021): A Transformer variant that employs sparse attention and generative-style decoding to enhance long-sequence prediction efficiency.
- FEDFormer (Zhou et al., 2022): A Transformer variant using frequency-domain attention to strengthen series decomposition and boost forecasting accuracy.
- iTransformer (Liu et al., 2023): A Transformer variant that adopts an inverted design, emphasizing feature dimensions over time steps to exploit correlations among inputs.
- Timexer (Y. Wang et al., 2024): A Transformer variant that refines attention by integrating inter-target and input-target relations to better capture temporal dependencies.
- SimpleTM (Chen et al., 2025): A lightweight time series forecasting model that streamlines architecture and improves computational efficiency while maintaining strong predictive

performance.

For each forecasting method, two variants were considered: a default data-driven version (baseline) and an extension with KGML. The baseline models are trained directly using in-situ observations as target values. The KGML variants incorporate two commonly adopted strategies. First, models are pre-trained using ED simulation data and then fine-tuned with in-situ observations. Second, the training objective incorporates physics-constrained loss functions derived from the carbon mass balance relationship (e.g., $NEE = RECO - GPP$), in addition to the standard mean squared error (MSE) loss computed from in-situ measurements.

These strategies are also incorporated into the proposed DERE model. DERE model is implemented with the Transformer baseline serving as the backbone.

4.4.2 Overall Performance Evaluation

The models compared in this study include the process-based ED model, several baseline models, several KGML models, and the proposed DERE framework. Table 4.1–Table 4.3 present the overall testing performance for GPP, RECO, and NEE. For convenience, the number of times each method ranked within the top two across all columns (Top-2 counts) was also calculated.

Table 4.1 Gross primary productivity (GPP) result comparison (top results in bold; runners-up with underlines).

Methods		ABoVE		AmeriFlux		FLUXNET		ICOS-WW		Multiple		Top-2 Count
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
Physical	ED	0.427	0.703	0.883	1.419	0.822	1.110	0.713	0.968	0.678	1.010	0/10
Baseline	Transformer	0.502	0.684	1.030	1.405	0.850	1.147	0.955	1.235	0.736	0.997	0/10
	Informer	0.523	0.729	1.093	1.465	0.882	1.180	0.859	1.096	0.817	1.065	0/10
	FEDformer	0.451	0.626	0.917	1.347	0.768	1.025	1.154	1.524	0.790	1.080	0/10
	iTransformer	0.561	0.662	1.007	1.480	0.887	1.162	1.009	1.340	0.851	1.149	0/10
	TimeXer	0.685	1.001	1.250	1.939	1.156	1.602	1.451	1.860	1.100	1.547	0/10
	SimpleTM	0.631	0.782	1.057	1.549	0.972	1.251	1.132	1.434	1.007	1.346	0/10
KGML	Transformer	<u>0.352</u>	0.541	<u>0.651</u>	1.033	0.699	0.951	0.711	1.014	0.636	0.919	2/10
	Informer	0.373	0.555	0.604	0.947	0.756	1.011	0.839	1.141	<u>0.618</u>	0.888	4/10
	FEDformer	0.369	0.531	0.769	1.115	0.726	0.926	0.830	1.121	0.701	1.003	0/10
	iTransformer	0.363	<u>0.508</u>	0.678	<u>1.027</u>	0.654	0.885	0.641	0.919	0.621	0.909	3/10
	TimeXer	0.375	0.566	0.657	1.036	<u>0.649</u>	0.879	0.791	1.050	0.678	0.990	1/10
	SimpleTM	0.406	0.587	0.737	1.102	0.669	0.855	<u>0.651</u>	<u>0.916</u>	0.680	0.976	3/10
Proposed	DERE	0.302	0.485	0.663	1.068	0.598	<u>0.863</u>	0.682	0.888	0.585	<u>0.890</u>	7/10

Table 4.2 Ecosystem respiration (RECO) result comparison (top results in bold; runners-up with underlines).

Methods		ABoVE		AmeriFlux		FLUXNET		ICOS-WW		Multiple		Top-2 Count
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
Physical	ED	0.431	0.657	0.603	0.855	0.603	0.833	0.464	0.620	<u>0.457</u>	0.649	3/10
Baseline	Transformer	0.602	0.857	0.804	1.040	0.559	0.743	0.698	0.886	0.488	0.709	0/10
	Informer	0.498	0.719	0.720	0.941	0.546	0.749	0.697	0.927	0.530	0.740	0/10
	FEDformer	0.431	0.611	0.616	0.867	0.482	0.638	0.832	1.054	0.574	0.804	0/10
	iTransformer	0.412	0.526	0.543	0.774	0.510	0.662	0.679	0.969	0.535	0.752	0/10
	TimeXer	0.607	0.854	0.946	1.279	0.713	0.960	1.386	1.770	0.784	1.031	0/10
	SimpleTM	0.503	0.660	0.620	0.865	0.561	0.709	0.862	1.126	0.626	0.859	0/10
KGML	Transformer	0.328	<u>0.467</u>	0.526	0.705	<u>0.438</u>	<u>0.583</u>	0.586	0.756	0.466	0.655	3/10
	Informer	0.354	0.495	<u>0.448</u>	0.614	0.475	0.626	0.559	0.716	0.463	<u>0.638</u>	3/10
	FEDformer	0.319	0.443	0.533	0.736	0.506	0.645	0.582	0.781	0.541	0.727	2/10
	iTransformer	0.348	0.492	0.439	<u>0.650</u>	0.445	0.585	0.595	0.796	0.537	0.733	2/10
	TimeXer	0.388	0.543	0.591	0.769	0.512	0.640	0.739	0.914	0.555	0.744	0/10
	SimpleTM	0.385	0.555	0.492	0.694	0.487	0.643	0.519	0.707	0.545	0.733	0/10
Proposed	DERE	<u>0.320</u>	0.472	0.465	0.671	0.393	0.557	<u>0.502</u>	<u>0.639</u>	0.442	0.627	7/10

Table 4.3 Net ecosystem exchange (NEE) result comparison (top results in bold; runners-up with underlines).

Methods		ABoVE		AmeriFlux		FLUXNET		ICOS-WW		Multiple		Top-2 Count
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
Physical	ED	0.276	0.386	0.599	0.920	0.613	0.889	0.530	0.691	0.537	0.770	0/10
Baseline	Transformer	0.413	0.500	0.615	0.875	0.575	0.790	0.489	0.665	0.532	0.726	0/10
	Informer	0.217	0.287	0.626	0.860	0.561	0.787	0.466	0.589	0.554	0.741	2/10
	FEDformer	0.201	0.266	0.546	0.805	0.548	0.769	0.535	0.680	0.529	0.735	0/10
	iTransformer	0.365	0.427	0.711	1.026	0.697	0.940	0.585	0.762	0.625	0.855	0/10
	TimeXer	1.288	2.209	0.714	1.129	0.770	1.126	1.340	1.644	0.843	1.474	0/10
	SimpleTM	0.334	0.415	0.692	0.997	0.664	0.912	0.640	0.827	0.655	0.894	0/10
KGML	Transformer	<u>0.177</u>	0.244	0.507	0.733	<u>0.504</u>	0.722	0.511	0.721	<u>0.489</u>	0.691	4/10
	Informer	0.186	0.252	0.456	0.700	0.509	<u>0.721</u>	0.519	0.730	0.501	0.699	3/10
	FEDformer	0.197	0.268	0.506	0.741	0.522	0.735	<u>0.475</u>	<u>0.627</u>	0.492	0.698	2/10
	iTransformer	0.179	0.239	0.540	0.770	0.600	0.834	0.485	0.641	0.534	0.737	1/10
	TimeXer	0.192	0.264	0.529	0.800	0.532	0.760	0.563	0.735	0.505	0.714	0/10
	SimpleTM	0.187	0.255	0.582	0.842	0.588	0.790	0.527	0.709	0.492	0.698	0/10
Proposed	DERE	0.175	<u>0.244</u>	<u>0.496</u>	<u>0.713</u>	0.478	0.683	0.505	0.696	0.476	<u>0.692</u>	8/10

The purely data-driven baselines do not outperform the process-based ED model, likely due to generalization challenges under limited observational constraints. As a reference, the ED model is a well-established ecosystem model that has been widely used in major systems such as the NASA Carbon Monitoring System. In contrast, KGML approaches outperform the purely data-driven baselines, highlighting the importance of incorporating physical knowledge into machine learning frameworks.

Compared with the process-based ED model and several advanced time-series machine learning models (including both baseline and KGML approaches), the proposed DERE method achieves the best overall performance. The improved performance of DERE model can be attributed to its integration of multiple sources of supervision, including direct labels, indirect labels, probabilistic labels, and simulation data. These results demonstrate the strong potential of integrating indirect satellite-derived information to improve ecosystem carbon flux prediction.

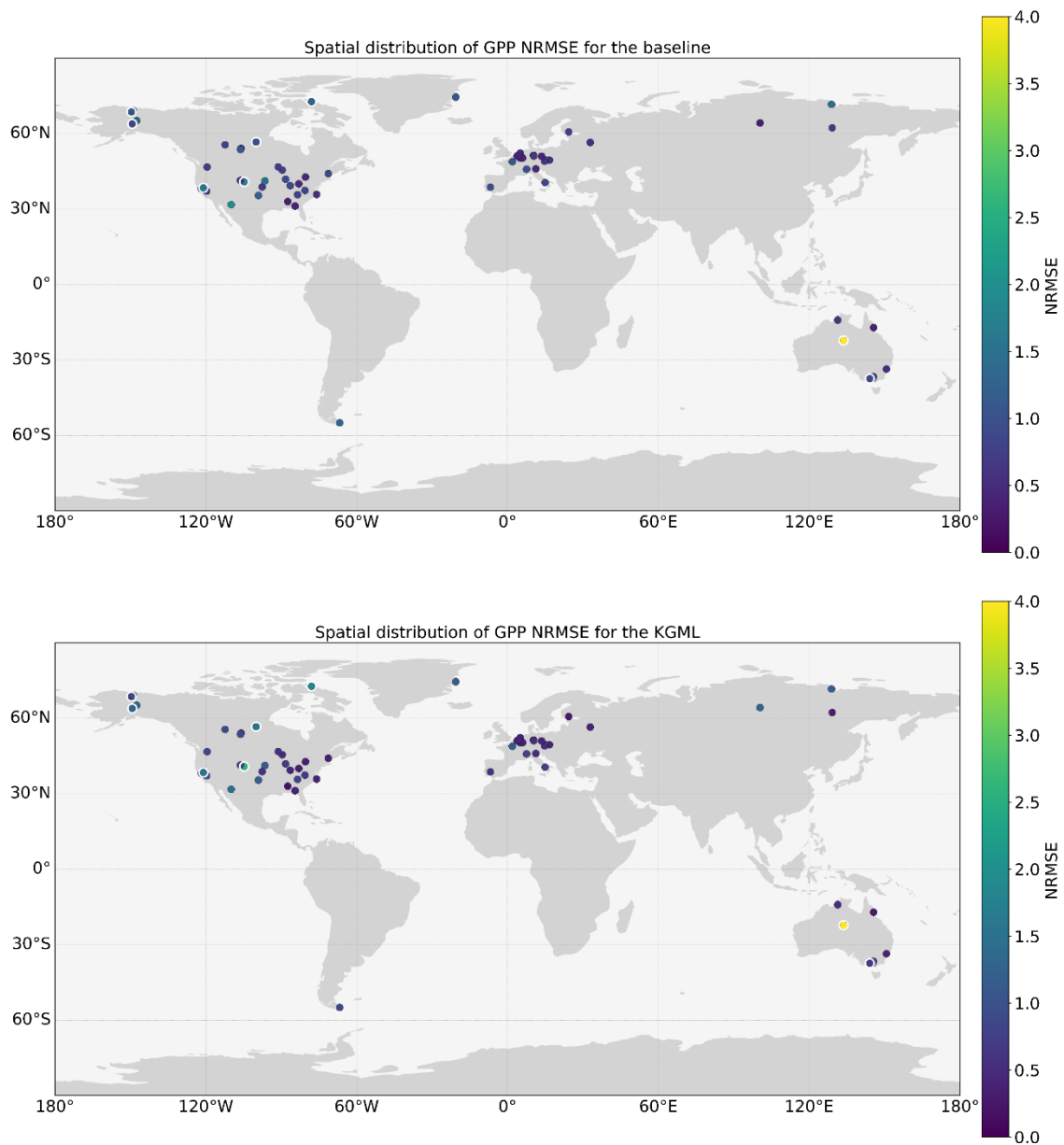
To further examine model performance beyond MAE and RMSE, performance is evaluated using absolute mean-based normalized RMSE (NRMSE). In this study, NRMSE is defined as the RMSE normalized by the mean absolute value of the true flux. This absolute-value-based normalization provides a stable and scale-independent metric, particularly for flux variables that may have small mean values or sign variability. Additionally, the average NRMSE across all networks is reported. As shown in Table 4.4, compared to both the process-based ED model and the baseline model (Transformer), the proposed DERE method consistently reduces relative error across all variables. For GPP, the average NRMSE across networks decreases from 0.82–0.85 to 0.64, and for RECO, it decreases from 0.75–0.88 to 0.60. For NEE, although the values remain higher due to its smaller magnitude and higher variability, DERE still reduces the average NRMSE from 1.43–1.46 to 1.15, indicating more accurate estimation of ecosystem carbon fluxes.

Table 4.4 Absolute mean-based normalized RMSE (NRMSE) and its average across networks.

Variable	Methods	NRMSE					Average NRMSE
		ABoVE	AmeriFlux	FLUXNET	ICOS-WW	Multiple	
GPP	ED	1.020	0.817	0.786	0.589	0.711	0.817
	Transformer	0.993	0.845	0.812	0.751	0.702	0.845
	DERE	0.704	0.643	0.611	0.540	0.626	0.643
RECO	ED	1.070	0.752	0.910	0.450	0.588	0.752
	Transformer	1.396	0.879	0.812	0.643	0.642	0.879
	DERE	0.769	0.598	0.609	0.464	0.568	0.598
NEE	ED	1.708	1.430	1.347	1.342	1.273	1.430
	Transformer	2.212	1.462	1.197	1.291	1.200	1.462
	DERE	1.080	1.152	1.035	1.351	1.144	1.152

To further examine where the proposed DERE performs better, the spatial distribution of NRMSE is compared across the baseline Transformer, KGML Transformer, and DERE for all sites, using

GPP as an example (Figure 4.5). Here, NRMSE is defined as the RMSE normalized by the mean absolute value of the true GPP flux.



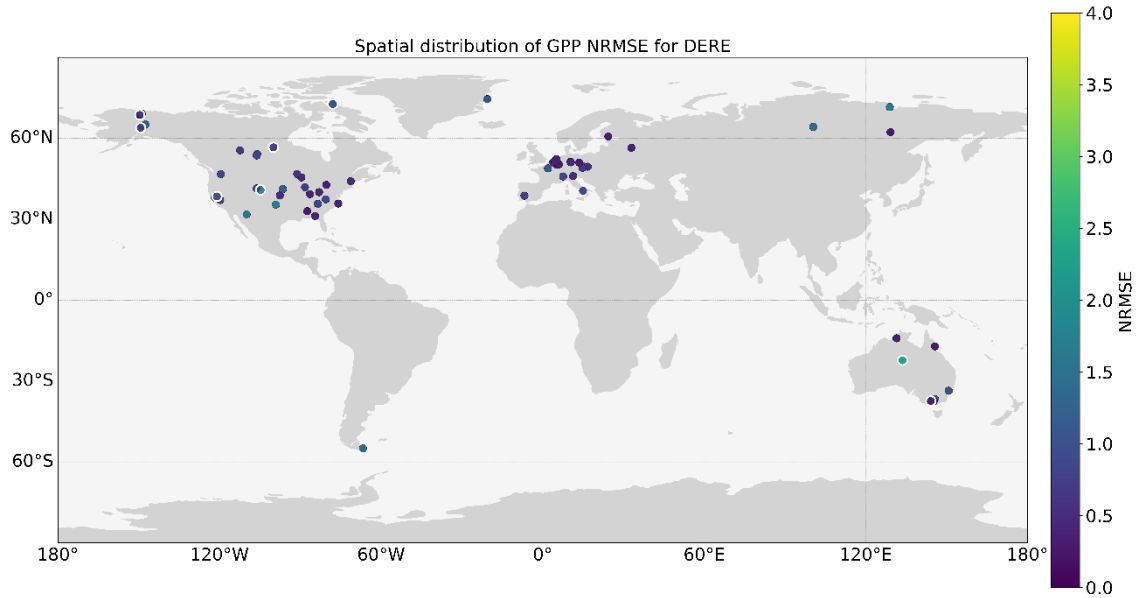


Figure 4.5 Spatial comparison of gross primary productivity (GPP) normalized RMSE (NRMSE) for the baseline, knowledge-guided machine learning (KGML), and decomposition-and-resembling (DERE). White circles indicate locations where DERE achieves an NRMSE reduction greater than 0.5 compared to the baseline or KGML.

From these maps, DERE reduces errors at several locations and produces a more consistent spatial pattern compared to the baseline and KGML. The improvements are not uniform, but are more evident in regions where the baseline and KGML exhibit relatively higher errors. To highlight stronger improvements, sites where DERE achieves an NRMSE reduction greater than 0.5 relative to the baseline or KGML are marked with white circles.

In particular, improvements are primarily observed in high-latitude North America (especially Alaska and northwestern Canada), where both baseline and KGML show relatively large errors for GPP. Additional scattered improvements are observed in other regions, including parts of Australia, particularly at locations with relatively high errors. These regions are associated with more challenging environmental conditions, such as higher climate variability, extreme

temperature regimes, and relatively sparse observations, which increase modeling difficulty. The reduction of errors in these areas indicates improved robustness under such conditions.

The improvement can be attributed to several aspects of the framework:

- (i) the decomposition design models sub-processes learned from process-based simulations, capturing more physically consistent relationships than the baseline;
- (ii) the resembling mechanism aligns model components with real-world observations (satellite and in-situ), providing stronger observational constraints; and
- (iii) the probabilistic label expansion provides additional supervision and supports learning from temporally sparse data.

In addition, physical constraints (e.g., carbon mass balance) further regularize the model.

Overall, these components enable DERE to better integrate simulation, satellite, and in-situ data while respecting physical structure, leading to improved robustness across challenging environments.

4.4.3 Time-Series and Scatter Plot Analysis

Figure 4.6 illustrates example temporal dynamics of GPP derived from the baseline Transformer model, the KGML Transformer model, the proposed DERE method, and in-situ observations. These examples provide a qualitative comparison of model performance. Overall, the proposed method shows closer agreement with the in-situ observations than the other models. Both the KGML model and the proposed DERE method capture more stable and consistent temporal patterns compared with the baseline model.

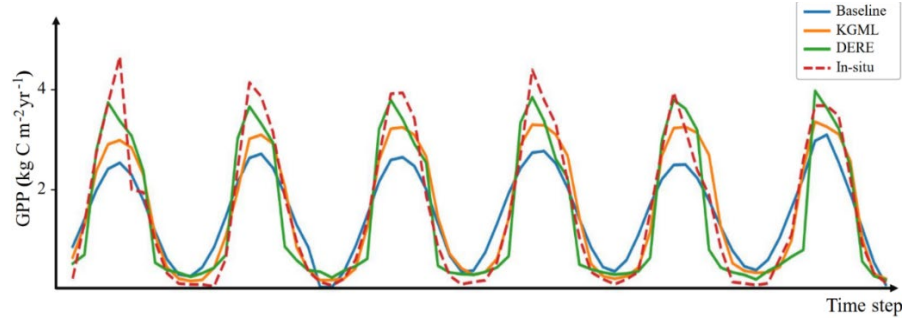


Figure 4.6 Example of time-series predictions of gross primary productivity (GPP). Both the baseline and the knowledge-guided machine learning (KGML) models are based on Transformer architecture.

Figure 4.7 further presents scatter plots comparing predicted GPP with in-situ observations across the ABoVE network. The results indicate that the proposed DERE method effectively reduces many prediction deviations highlighted by the red boxes, leading to improved agreement with the observations and overall prediction accuracy.

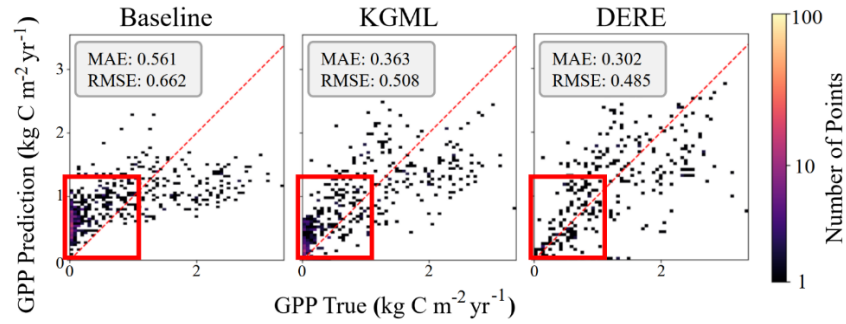


Figure 4.7 Scatter plot examples of observed versus predicted gross primary productivity (GPP) for the ABoVE dataset. Both the baseline and knowledge-guided machine learning (KGML) models are implemented using iTransformer architecture.

4.4.4 Probabilistic Label Expansion Strategy

Figure 4.8 illustrates examples of probabilistic label expansion on in-situ data using the conditional

diffusion framework. The green line represents the mean value of the sample predictions. Within the masked regions, the predicted time series closely matches the hidden in-situ observations, demonstrating the capability of the model to accurately reconstruct missing GPP values.

The green shaded region represents the 5%–95% confidence interval across the imputed time series, reflecting the uncertainty associated with the probabilistic predictions. This prediction variance can be further utilized by the uncertainty-aware fine-tuning module to improve the effectiveness of the probabilistic labels during model training.

In addition, the reconstructed temporal patterns remain relatively stable under varying proportions of missing data, suggesting that the model benefits from the incorporation of auxiliary environmental inputs.

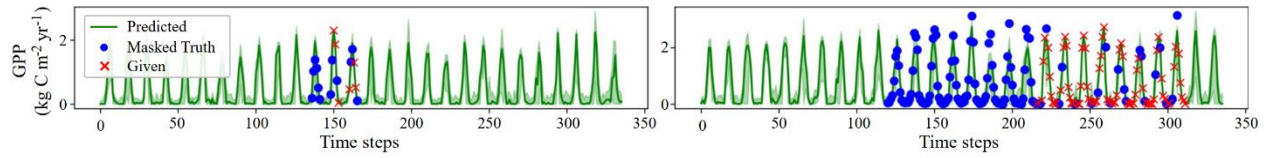


Figure 4.8 Probabilistic label expansion examples. The green shaded regions represent the 5%–95% confidence interval based on sampled time series.

4.5 Discussion

4.5.1 Ablation Study

Table 4.5 presents the ablation study results for GPP, RECO, and NEE. The evaluated models include the baseline model, the KGML model, the KGML model with indirect labels, and the proposed DERE framework. For convenience, the number of times each method ranked within the top two across all columns (Top-2 counts) was also calculated.

Table 4.5 Ablation study (top results in bold; runners-up with underlines).

	Methods	ABoVE		AmeriFlux		FLUXNET		ICOS-WW		Multiple		Top-2 Count
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
GPP	Baseline	0.502	0.684	1.030	1.405	0.850	1.147	0.955	1.235	0.736	0.997	0/10
	KGML	0.352	0.541	0.651	1.033	0.699	0.951	0.711	1.014	0.636	0.919	2/10
	KGML + indirect labels	<u>0.321</u>	<u>0.536</u>	0.677	1.084	<u>0.619</u>	0.850	<u>0.685</u>	0.888	0.573	0.888	8/10
	DERE	0.302	0.485	<u>0.663</u>	<u>1.068</u>	0.598	<u>0.863</u>	0.682	<u>0.888</u>	<u>0.585</u>	<u>0.890</u>	10/10
RECO	Baseline	0.602	0.857	0.804	1.040	0.559	0.743	0.698	0.886	0.488	0.709	0/10
	KGML	<u>0.328</u>	0.467	0.526	<u>0.705</u>	0.438	0.583	0.586	0.756	0.466	0.655	3/10
	KGML + indirect labels	0.334	0.526	<u>0.526</u>	0.721	<u>0.409</u>	<u>0.567</u>	<u>0.508</u>	<u>0.639</u>	0.402	0.596	7/10
	DERE	0.320	<u>0.472</u>	0.465	0.671	0.393	0.557	0.502	0.639	<u>0.442</u>	<u>0.627</u>	10/10
NEE	Baseline	0.413	0.500	0.615	0.875	0.575	0.790	0.489	0.665	0.532	0.726	2/10
	KGML	<u>0.177</u>	0.244	0.507	0.733	0.504	0.722	0.511	0.721	0.489	0.691	2/10
	KGML + indirect labels	0.179	0.242	0.475	<u>0.723</u>	0.471	0.677	<u>0.505</u>	0.696	<u>0.482</u>	0.698	7/10
	DERE	0.175	<u>0.244</u>	<u>0.496</u>	0.713	<u>0.478</u>	<u>0.683</u>	0.505	<u>0.696</u>	0.476	<u>0.692</u>	9/10

The results show that incorporating physical knowledge and additional supervisory signals consistently improves prediction performance. Compared with the baseline model, the KGML model significantly reduces both MAE and RMSE across most datasets, indicating that integrating physics knowledge helps improve model generalization. Further extending KGML with indirect remote sensing labels leads to additional performance gains, demonstrating the benefit of incorporating supplementary supervisory information.

The proposed DERE model with probabilistic labels and uncertainty-aware tuning achieves the best overall performance across the three variables. In particular, DERE achieves the highest Top-2 counts for GPP (10/10), RECO (10/10), and NEE (9/10). These results confirm that each component of the proposed framework contributes to improving ecosystem carbon flux prediction.

4.6 Conclusion

This study presented a knowledge-guided learning framework based on the proposed DERE approach for improving ecosystem carbon flux prediction. The framework bridges process-based ecosystem models and machine learning models to leverage multiple sources of supervision, including direct observational labels, indirect labels derived from related variables, probabilistic labels generated through CSDI, and simulation data. By integrating these complementary information sources, the proposed approach aims to improve prediction accuracy and model generalization.

Extensive experiments were conducted using datasets from multiple observational networks with global coverage, including the ABoVE dataset and the CarbonSense benchmark. The results demonstrate that the proposed DERE-based KGML framework consistently improves prediction performance for key ecosystem carbon variables, including GPP, RECO, and NEE, compared with both purely data-driven baselines and existing KGML approaches. In addition to improved quantitative metrics, qualitative analyses show that the proposed method produces more stable temporal dynamics and better alignment with in-situ observations. The ablation study further confirms the contribution of each component in the framework, highlighting the importance of integrating physical knowledge and multiple sources of supervision.

Despite these improvements, several limitations remain. The in-situ datasets used in this study have relatively dense coverage in regions such as North America and Europe due to the availability of flux tower observations, while other regions, including parts of Africa, remain sparsely monitored. As a result, the model performance in these underrepresented regions may be less well characterized. In addition, scenario-based evaluations under different climate zones, vegetation types, and environmental conditions were not explored in this study.

Future work will extend the proposed framework in several directions. First, the approach can be expanded to support forecasting tasks that predict ecosystem carbon dynamics under future environmental conditions. Second, scenario-based evaluations across different climate zones and forest types will be investigated to better understand model robustness under diverse ecological conditions. Finally, additional methodological developments will explore approaches for handling anomalous or rare environmental events. The implementation of the proposed framework will be made publicly available to support further research in data-driven ecosystem monitoring.

Chapter 5: Conclusion

5.1 Summary of the Dissertation

This dissertation investigates methods for improving the estimation of key land surface variables through the integration of multi-source data and AI approaches. Accurate estimation of environmental variables plays a central role in Earth system monitoring because it enables the characterization of interactions among the land surface, atmosphere, and biosphere. Such understanding is essential for studying land–atmosphere interactions, ecosystem dynamics, and environmental change.

In recent decades, substantial progress has been made in developing environmental datasets derived from satellite observations, in-situ measurements, and process-based ecosystem models. Satellite observations provide extensive spatial coverage and long-term monitoring capability. In-situ measurements offer highly accurate observations at local scales. Process-based models simulate ecosystem processes using physically interpretable formulations. However, each data source also has inherent limitations. Satellite observations may be affected by observational constraints such as cloud contamination and indirect measurement mechanisms. In-situ measurements are spatially sparse and often temporally incomplete. Process-based models rely on simplified assumptions and parameterizations that may introduce structural uncertainties. Consequently, effectively integrating information from these complementary data sources remains a key challenge for advancing Earth system monitoring.

This dissertation uses LST and GPP as case studies, as these variables represent two major categories of land surface processes and correspond to two distinct modeling paradigms. LST reflects energetic processes governing radiation and heat exchange at the land–atmosphere interface and is typically retrieved from satellite observations using snapshot retrieval methods. In

contrast, GPP reflects ecological processes associated with vegetation activity and ecosystem carbon uptake, which evolve continuously over time and therefore require time-series prediction approaches. Studying these two variables provides a unified framework for investigating both energetic and ecological dynamics in Earth system monitoring while examining methodological challenges associated with snapshot retrieval and time-series prediction.

To address the challenge of multi-source information integration, this dissertation investigates whether physics-guided machine learning can improve environmental variable estimation. It is organized into three interconnected studies that progressively examine environmental datasets, snapshot retrieval methods, and time-series prediction approaches.

The first study conducted a systematic evaluation of existing LST products and GPP simulations using satellite observations, model simulations, and in-situ measurements. The analysis revealed several important limitations in current environmental datasets. For LST, comparisons across multiple datasets showed a clear trade-off between data availability and retrieval accuracy. TIR products generally provide relatively accurate estimates but suffer from limited spatial coverage due to cloud contamination, while PMW and reanalysis products offer higher data availability but often exhibit larger retrieval errors. For GPP, comparisons between ecosystem model simulations and flux tower observations showed that process-based models can capture seasonal dynamics of ecosystem productivity but may exhibit discrepancies in magnitude at some sites. Further analysis indicated that these discrepancies are closely related to differences in PFT representation. These findings highlight two key challenges in current environmental datasets: the strong dependence of LST retrieval on surface properties and the sensitivity of GPP prediction to vegetation composition. The second study addressed the challenge of improving snapshot retrieval of LST. Traditional LST retrieval methods often rely on auxiliary inputs such as LSE, which is not always available or

accurately known. In addition, conventional ML approaches typically rely on large training datasets and often struggle to integrate information from different data domains. To address these limitations, this study proposed a TL-based framework that integrates radiative transfer simulations with real-world observations. The approach first uses simulation data generated by RTMs to train ML models and then fine-tunes the model using satellite observations and in-situ measurements. By integrating shortwave and longwave TOA observations, the proposed framework reduces dependence on external LSE inputs and improves LST retrieval accuracy. Extensive validation using Landsat observations and in-situ measurements demonstrated that the proposed TL model outperforms conventional SC and SW algorithms, highlighting the potential of ML-based multi-source integration for improving snapshot retrieval of energetic variables.

The third study focused on improving time-series prediction of ecosystem carbon flux variables such as GPP. Existing approaches face challenges in integrating information from satellite observations, process-based ecosystem simulations, and in-situ measurements. In particular, satellite observations often provide aggregated ecological information, whereas ecosystem models represent vegetation dynamics through detailed internal structures such as age-dependent and PFT-specific processes. To address this challenge, this study proposed a KGML framework based on DERE design. The framework first decomposes ecosystem model outputs into intermediate ecological components, including age-specific PFT fractions and carbon flux contributions across PFTs and age groups, making the internal ecosystem structure available for learning. Second, it performs a knowledge-aligned resembling step that aggregates age-specific PFT fractions to obtain overall PFT composition and aligns it with satellite-derived PFT observations, enabling satellite data to provide indirect structural constraints. Flux tower measurements are also used as supervisory signals for carbon flux prediction. Third, a probabilistic label expansion module is

introduced to better utilize temporally sparse observations by estimating missing flux values and incorporating uncertainty during model training. Experimental results across multiple observational networks demonstrate that the proposed DERE framework consistently improves prediction performance for GPP compared with both purely data-driven models and existing knowledge-guided learning approaches.

Overall, these studies demonstrate the potential of KGML to integrate multiple complementary environmental data sources. By combining process-based simulation knowledge, satellite observations, and in-situ measurements, the proposed approaches improve environmental variable estimation and enhance the effective use of multi-source data, highlighting the potential of KGML for integrating observations and process knowledge in Earth system monitoring.

5.2 Main Contributions

This dissertation makes several important contributions to environmental variable estimation and Earth system monitoring.

First, the research provides a systematic evaluation of existing LST products and GPP simulations using satellite observations, model outputs, and in-situ measurements. By identifying key limitations related to surface property dependence and vegetation composition representation, the study offers insights into the sources of uncertainty in current environmental datasets and provides guidance for improving environmental variable estimation.

Second, this dissertation develops an ML-based framework for improving snapshot retrieval of LST through the integration of radiative transfer simulations and real-world observations. The proposed TL approach demonstrates an effective strategy for combining simulation knowledge with observational data to enhance satellite-based retrieval accuracy while reducing reliance on

auxiliary inputs such as LSE.

Third, the research introduces a KGML framework for time-series prediction of ecosystem carbon flux variables. The proposed DERE strategy enables the integration of satellite-derived vegetation information with internal ecosystem model representations, bridging indirect satellite observations with ecological process-based modeling.

Finally, this dissertation demonstrates the broader potential of KGML for integrating multiple complementary environmental data sources. By combining process-based simulations, satellite observations, and in-situ measurements, the proposed approaches improve both predictive performance and effective use of multi-source environmental information.

5.3 Limitations

Despite the improvements achieved in this dissertation, several limitations remain.

First, the spatial distribution of in-situ observations used in this research is uneven. Flux tower networks are relatively dense in regions such as North America and Europe, while other regions, including parts of Africa and South America, remain sparsely monitored. As a result, model performance in these underrepresented regions may be less well characterized.

Second, although satellite observations provide valuable large-scale information, integrating satellite-derived variables with the internal representations of process-based ecosystem models remains challenging due to differences in spatial resolution and ecological representation.

Third, the proposed frameworks primarily focus on improving estimation and reconstruction of environmental variables rather than long-term forecasting under changing environmental conditions.

5.4 Future Research Directions

Future research can extend the work presented in this dissertation in several directions.

First, the proposed frameworks can be extended to support forecasting tasks that predict variables under future environmental conditions. Integrating climate projections and ecosystem simulations may provide new opportunities for long-term ecosystem prediction.

Second, future studies can conduct evaluations across different climate zones, vegetation types, and disturbance regimes to better understand model robustness under diverse environmental conditions.

Third, additional methodological developments can explore approaches for handling anomalous or rare environmental events. Improving model performance under such conditions will enhance the reliability of environmental predictions.

Finally, expanding observational datasets and incorporating new satellite missions will further improve the effectiveness of multi-source data integration in Earth system monitoring.

Appendices

Table S1. Basic information of the sites used in this study.

Network Name	Site ID	Surface Type	Elevation (m)	Latitude (°)	Longitude (°)
AmeriFlux	BR-Npw	Woody Savannas	120.00	-16.50	-56.41
AmeriFlux	CA-ARB	Permanent Wetlands	90.00	52.70	-83.95
AmeriFlux	CA-ARF	Permanent Wetlands	88.00	52.70	-83.96
AmeriFlux	CA-BOU	Permanent Wetlands	108.00	50.52	-63.21
AmeriFlux	CA-Ca3	Evergreen Needleleaf Forests	null	49.53	-124.90
AmeriFlux	CA-Cbo	Deciduous Broadleaf Forests	120.00	44.32	-79.93
AmeriFlux	CA-DB2	Permanent Wetlands	4.00	49.12	-123.00
AmeriFlux	CA-DBB	Permanent Wetlands	4.00	49.13	-122.99
AmeriFlux	CA-DSM	Permanent Wetlands	null	49.09	-122.90
AmeriFlux	CA-Gro	Mixed Forests	340.00	48.22	-82.16
AmeriFlux	CA-LP1	Evergreen Needleleaf Forests	751.00	55.11	-122.84
AmeriFlux	CA-TP4	Evergreen Needleleaf Forests	184.00	42.71	-80.36
AmeriFlux	CA-TPD	Deciduous Broadleaf Forests	260.00	42.64	-80.56
AmeriFlux	CL-SDF	Evergreen Broadleaf Forests	50.00	-41.88	-73.68
AmeriFlux	CL-SDP	Permanent Wetlands	50.00	-41.88	-73.67
AmeriFlux	MX-Aog	Deciduous Broadleaf Forests	367.28	27.00	-108.79
AmeriFlux	MX-PMm	Permanent Wetlands	10.00	20.85	-86.90
AmeriFlux	PR-xGU	Evergreen Broadleaf Forests	143.00	17.97	-66.87
AmeriFlux	PR-xLA	Grasslands	24.00	18.02	-67.08
AmeriFlux	US-A03	Barren Sparse Vegetation	5.00	70.50	-149.88
AmeriFlux	US-A10	Barren Sparse Vegetation	4.00	71.32	-156.62
AmeriFlux	US-A32	Grasslands	335.00	36.82	-97.82
AmeriFlux	US-A74	Croplands	337.00	36.81	-97.55
AmeriFlux	US-ALQ	Permanent Wetlands	null	46.03	-89.61
AmeriFlux	US-An1	Open Shrublands	600.00	68.99	-150.28
AmeriFlux	US-An2	Open Shrublands	600.00	68.95	-150.21
AmeriFlux	US-An3	Open Shrublands	600.00	68.93	-150.27
AmeriFlux	US-ARM	Croplands	314.00	36.61	-97.49
AmeriFlux	US-Bi2	Croplands	-5.00	38.11	-121.54
AmeriFlux	US-BRG	Grasslands	180.00	39.22	-86.54
AmeriFlux	US-BZB	Permanent Wetlands	100.00	64.70	-148.32
AmeriFlux	US-BZF	Permanent Wetlands	95.00	64.70	-148.31

AmeriFlux	US-CC1	Croplands	314.00	44.07	-89.68
AmeriFlux	US-CC2	Croplands	320.00	44.10	-89.62
AmeriFlux	US-CRK	Evergreen Needleleaf Forests	151.00	31.46	-95.34
AmeriFlux	US-CRT	Croplands	180.00	41.63	-83.35
AmeriFlux	US-CS1	Croplands	328.00	44.10	-89.54
AmeriFlux	US-CS2	Evergreen Needleleaf Forests	328.00	44.15	-89.50
AmeriFlux	US-CS4	Croplands	328.00	44.16	-89.55
AmeriFlux	US-CS6	Croplands	328.00	44.43	-89.54
AmeriFlux	US-DFC	Croplands	264.90	43.34	-89.71
AmeriFlux	US-DFK	Croplands	210.00	43.35	-89.72
AmeriFlux	US-Dmg	Permanent Wetlands	1.00	38.00	-121.67
AmeriFlux	US-DS3	Croplands	-7.00	38.12	-121.55
AmeriFlux	US-EDN	Permanent Wetlands	null	37.62	-122.11
AmeriFlux	US-EML	Open Shrublands	662.00	63.88	-149.25
AmeriFlux	US-GLE	Evergreen Needleleaf Forests	3197.00	41.37	-106.24
AmeriFlux	US-HB1	Permanent Wetlands	0.10	33.35	-79.20
AmeriFlux	US-HB2	Evergreen Needleleaf Forests	4.70	33.32	-79.24
AmeriFlux	US-HB3	Evergreen Needleleaf Forests	7.30	33.35	-79.23
AmeriFlux	US-HBK	Deciduous Broadleaf Forests	367.00	43.94	-71.72
AmeriFlux	US-Hn3	Open Shrublands	120.90	46.69	-119.46
AmeriFlux	US-Ho1	Evergreen Needleleaf Forests	60.00	45.20	-68.74
AmeriFlux	US-Ho2	Evergreen Needleleaf Forests	61.00	45.21	-68.75
AmeriFlux	US-HRA	Croplands	null	34.59	-91.75
AmeriFlux	US-Hsm	Permanent Wetlands	1.11	38.24	-122.02
AmeriFlux	US-HWB	Cropland/Natural Vegetation Mosaics	378.00	40.86	-77.85
AmeriFlux	US-Jo1	Open Shrublands	1188.00	32.58	-106.64
AmeriFlux	US-KFS	Grasslands	310.00	39.06	-95.19
AmeriFlux	US-KLS	Grasslands	373.00	38.77	-97.57
AmeriFlux	US-KM4	Grasslands	246.30	42.44	-85.33
AmeriFlux	US-Kon	Grasslands	417.00	39.08	-96.56
AmeriFlux	US-KPL	Permanent Wetlands	100.60	60.54	-150.51
AmeriFlux	US-KS3	Permanent Wetlands	0.00	28.71	-80.74
AmeriFlux	US-Los	Permanent Wetlands	480.00	46.08	-89.98
AmeriFlux	US-MC2	Croplands	900.00	48.18	-114.21
AmeriFlux	US-Me2	Evergreen Needleleaf Forests	1253.00	44.45	-121.56
AmeriFlux	US-MH2	Croplands	891.00	45.97	-108.20
AmeriFlux	US-Mj1	Croplands	1285.00	46.99	-109.61
AmeriFlux	US-MMS	Deciduous Broadleaf Forests	275.00	39.32	-86.41

AmeriFlux	US-Mo1	Croplands	260.00	39.23	-92.12
AmeriFlux	US-Mo2	Grasslands	275.00	38.95	-91.99
AmeriFlux	US-Mo3	Croplands	261.00	39.23	-92.15
AmeriFlux	US-MOz	Deciduous Broadleaf Forests	219.40	38.74	-92.20
AmeriFlux	US-Mpj	Woody Savannas	2196.00	34.44	-106.24
AmeriFlux	US-MtB	Evergreen Needleleaf Forests	2573.00	32.42	-110.73
AmeriFlux	US-NC2	Evergreen Needleleaf Forests	5.00	35.80	-76.67
AmeriFlux	US-NC3	Evergreen Needleleaf Forests	5.00	35.80	-76.66
AmeriFlux	US-NC4	Permanent Wetlands	1.00	35.79	-75.90
AmeriFlux	US-Ne1	Croplands	361.00	41.17	-96.48
AmeriFlux	US-Ne2	Croplands	362.00	41.16	-96.47
AmeriFlux	US-Ne3	Croplands	363.00	41.18	-96.44
AmeriFlux	US-NGB	Snow and Ice	5.27	71.28	-156.61
AmeriFlux	US-NGC	Grasslands	35.00	64.86	-163.70
AmeriFlux	US-NR1	Evergreen Needleleaf Forests	3050.00	40.03	-105.55
AmeriFlux	US-ONA	Grasslands	25.00	27.38	-81.95
AmeriFlux	US-OWC	Permanent Wetlands	174.00	41.38	-82.51
AmeriFlux	US-PAS	Grasslands	27.10	27.39	-81.95
AmeriFlux	US-PFb	Evergreen Needleleaf Forests	474.00	45.97	-90.32
AmeriFlux	US-PFc	Deciduous Broadleaf Forests	473.00	45.97	-90.31
AmeriFlux	US-PFe	Water Bodies	480.00	45.98	-90.30
AmeriFlux	US-PFf	Grasslands	464.00	45.95	-90.29
AmeriFlux	US-PFg	Evergreen Needleleaf Forests	475.00	45.97	-90.27
AmeriFlux	US-PFk	Deciduous Broadleaf Forests	476.00	45.91	-90.34
AmeriFlux	US-PFm	Deciduous Broadleaf Forests	484.00	45.92	-90.31
AmeriFlux	US-PFn	Deciduous Broadleaf Forests	478.00	45.94	-90.28
AmeriFlux	US-PFo	Water Bodies	476.00	45.92	-90.27
AmeriFlux	US-PFp	Deciduous Broadleaf Forests	497.00	45.94	-90.26
AmeriFlux	US-PHM	Permanent Wetlands	1.40	42.74	-70.83
AmeriFlux	US-Prr	Evergreen Needleleaf Forests	210.00	65.12	-147.49
AmeriFlux	US-Rls	Closed Shrublands	1608.00	43.14	-116.74
AmeriFlux	US-Ro1	Croplands	290.00	44.71	-93.09
AmeriFlux	US-Ro2	Croplands	292.00	44.73	-93.09
AmeriFlux	US-Ro4	Grasslands	274.00	44.68	-93.07
AmeriFlux	US-Ro5	Croplands	283.00	44.69	-93.06
AmeriFlux	US-Ro6	Croplands	282.00	44.69	-93.06
AmeriFlux	US-Rpf	Deciduous Broadleaf Forests	497.00	65.12	-147.43
AmeriFlux	US-Rws	Open Shrublands	1425.00	43.17	-116.71
AmeriFlux	US-Seg	Grasslands	1596.00	34.36	-106.70

AmeriFlux	US-Ses	Open Shrublands	1604.00	34.33	-106.74
AmeriFlux	US-Sne	Grasslands	-5.00	38.04	-121.76
AmeriFlux	US-Snf	Grasslands	-4.00	38.04	-121.73
AmeriFlux	US-SRM	Woody Savannas	1120.00	31.82	-110.87
AmeriFlux	US-SSH	Deciduous Broadleaf Forests	310.00	40.67	-77.90
AmeriFlux	US-Syv	Mixed Forests	540.00	46.24	-89.35
AmeriFlux	US-Ton	Woody Savannas	177.00	38.43	-120.97
AmeriFlux	US-TrB	Water Bodies	499.00	46.04	-89.69
AmeriFlux	US-TrS	Water Bodies	495.00	46.00	-89.71
AmeriFlux	US-Tw1	Permanent Wetlands	-5.00	38.11	-121.65
AmeriFlux	US-Tw2	Croplands	-5.00	38.10	-121.64
AmeriFlux	US-Tw3	Croplands	-4.00	38.12	-121.65
AmeriFlux	US-Tw4	Permanent Wetlands	-5.00	38.10	-121.64
AmeriFlux	US-Tw5	Permanent Wetlands	-5.00	38.11	-121.64
AmeriFlux	US-Uaf	Evergreen Needleleaf Forests	155.00	64.87	-147.86
AmeriFlux	US-UC1	Cropland/Natural Vegetation Mosaics	390.00	40.75	-78.01
AmeriFlux	US-UC2	Cropland/Natural Vegetation Mosaics	396.00	40.76	-78.00
AmeriFlux	US-UiA	Croplands	224.00	40.06	-88.20
AmeriFlux	US-UiB	Croplands	224.00	40.06	-88.20
AmeriFlux	US-UiC	Croplands	224.00	40.06	-88.20
AmeriFlux	US-UiD	Grasslands	224.00	40.06	-88.20
AmeriFlux	US-UM3	Water Bodies	234.00	45.57	-84.67
AmeriFlux	US-UMB	Deciduous Broadleaf Forests	234.00	45.56	-84.71
AmeriFlux	US-UMd	Deciduous Broadleaf Forests	239.00	45.56	-84.70
AmeriFlux	US-UTB	Barren Sparse Vegetation	1287.00	40.78	-113.83
AmeriFlux	US-Vcp	Evergreen Needleleaf Forests	2500.00	35.86	-106.60
AmeriFlux	US-Vcs	Evergreen Needleleaf Forests	2752.00	35.92	-106.61
AmeriFlux	US-WCr	Deciduous Broadleaf Forests	520.00	45.81	-90.08
AmeriFlux	US-Wkg	Grasslands	1531.00	31.74	-109.94
AmeriFlux	US-WPT	Permanent Wetlands	175.00	41.46	-83.00
AmeriFlux	US-Wrc	Evergreen Needleleaf Forests	371.00	45.82	-121.95
AmeriFlux	US-xAB	Evergreen Needleleaf Forests	363.00	45.76	-122.33
AmeriFlux	US-xAE	Grasslands	516.00	35.41	-99.06
AmeriFlux	US-xBA	Permanent Wetlands	6.00	71.28	-156.62
AmeriFlux	US-xBL	Deciduous Broadleaf Forests	183.00	39.06	-78.07
AmeriFlux	US-xBN	Evergreen Needleleaf Forests	263.00	65.15	-147.50
AmeriFlux	US-xBR	Deciduous Broadleaf Forests	232.00	44.06	-71.29

AmeriFlux	US-xCL	Grasslands	259.00	33.40	-97.57
AmeriFlux	US-xCP	Grasslands	1654.00	40.82	-104.75
AmeriFlux	US-xDC	Grasslands	559.00	47.16	-99.11
AmeriFlux	US-xDJ	Evergreen Needleleaf Forests	529.00	63.88	-145.75
AmeriFlux	US-xDL	Mixed Forests	22.00	32.54	-87.80
AmeriFlux	US-xDS	Cropland/Natural Vegetation Mosaics	15.00	28.13	-81.44
AmeriFlux	US-xGR	Deciduous Broadleaf Forests	579.00	35.69	-83.50
AmeriFlux	US-xHA	Deciduous Broadleaf Forests	351.00	42.54	-72.17
AmeriFlux	US-xHE	Open Shrublands	705.00	63.88	-149.21
AmeriFlux	US-xJE	Evergreen Needleleaf Forests	44.00	31.19	-84.47
AmeriFlux	US-xKA	Grasslands	1329.00	39.11	-96.61
AmeriFlux	US-xKZ	Grasslands	381.00	39.10	-96.56
AmeriFlux	US-xLE	Deciduous Broadleaf Forests	20.00	31.85	-88.16
AmeriFlux	US-xMB	Open Shrublands	1767.00	38.25	-109.39
AmeriFlux	US-xML	Deciduous Broadleaf Forests	1126.00	37.38	-80.52
AmeriFlux	US-xNG	Grasslands	578.00	46.77	-100.92
AmeriFlux	US-xNW	Evergreen Needleleaf Forests	3513.00	40.05	-105.58
AmeriFlux	US-xPU	Evergreen Broadleaf Forests	1685.00	19.55	-155.32
AmeriFlux	US-xRM	Evergreen Needleleaf Forests	2743.00	40.28	-105.55
AmeriFlux	US-xRN	Deciduous Broadleaf Forests	334.00	35.96	-84.28
AmeriFlux	US-xSB	Evergreen Needleleaf Forests	45.00	29.69	-81.99
AmeriFlux	US-xSC	Deciduous Broadleaf Forests	361.00	38.89	-78.14
AmeriFlux	US-xSE	Deciduous Broadleaf Forests	15.00	38.89	-76.56
AmeriFlux	US-xSJ	Savannas	368.00	37.11	-119.73
AmeriFlux	US-xSL	Croplands	1364.00	40.46	-103.03
AmeriFlux	US-xSR	Open Shrublands	983.00	31.91	-110.84
AmeriFlux	US-xST	Deciduous Broadleaf Forests	481.00	45.51	-89.59
AmeriFlux	US-xTA	Evergreen Needleleaf Forests	135.00	32.95	-87.39
AmeriFlux	US-xTE	Evergreen Needleleaf Forests	2147.00	37.01	-119.01
AmeriFlux	US-xTL	Permanent Wetlands	843.00	68.66	-149.37
AmeriFlux	US-xTR	Deciduous Broadleaf Forests	472.00	45.49	-89.59
AmeriFlux	US-xUK	Deciduous Broadleaf Forests	335.00	39.04	-95.19
AmeriFlux	US-xUN	Mixed Forests	518.00	46.23	-89.54
AmeriFlux	US-xWD	Grasslands	579.00	47.13	-99.24
AmeriFlux	US-xWR	Evergreen Needleleaf Forests	407.00	45.82	-121.95
AmeriFlux	US-xYE	Evergreen Needleleaf Forests	2116.00	44.95	-110.54
SURFRAD	BND	Croplands	230.00	40.05	-88.37
SURFRAD	DRA	Shrub Land	1007.00	36.62	-116.02

SURFRAD	FPK	Grasslands	634.00	48.31	-105.10
SURFRAD	GWN	Pasture	98.00	34.25	-89.87
SURFRAD	PSU	Croplands	376.00	40.72	-77.93
SURFRAD	SXF	Croplands	473.00	43.73	-96.62
BSRN	ALE	Tundra	127.00	82.49	-62.42
BSRN	BAR	Tundra	8.00	71.32	-156.61
BSRN	BUD	Grasslands	139.10	47.43	19.18
BSRN	CAB	Grasslands	0.00	51.97	4.93
BSRN	CAP	Tundra	null	79.27	101.75
BSRN	GOB	Desert Gravel	407.00	-23.56	15.04
BSRN	IZA	Rock	2372.90	28.31	-16.50
BSRN	NYA	Tundra	11.00	78.92	11.93
BSRN	PAY	Cultivated	491.00	46.81	6.94
BSRN	SEL	Grasslands	602.00	15.78	-91.99
BSRN	TAT	Grasslands	25.00	36.06	140.13
BSRN	TIK	Tundra	48.00	71.59	128.92
BSRN	TOR	Grasslands	70.00	58.26	26.46
FLUXNET	AU-Cum	Evergreen Broadleaf Forests	null	-33.62	150.72
FLUXNET	AU-Dry	Savannas	null	-15.26	132.37
FLUXNET	AU-Gin	Woody Savannas	null	-31.38	115.71
FLUXNET	AU-Rob	Evergreen Broadleaf Forests	null	-17.12	145.63
FLUXNET	AU-Stp	Grasslands	null	-17.15	133.35
FLUXNET	AU-Wom	Evergreen Broadleaf Forests	705.00	-37.42	144.09
FLUXNET	AU-Ync	Grasslands	null	-34.99	146.29
FLUXNET	BE-Bra	Mixed Forests	16.00	51.31	4.52
FLUXNET	BE-Lon	Croplands	167.00	50.55	4.75
FLUXNET	CZ-BK1	Evergreen Needleleaf Forests	875.00	49.50	18.54
FLUXNET	CZ-wet	Permanent Wetlands	426.00	49.02	14.77
FLUXNET	DE-Akm	Permanent Wetlands	-1.00	53.87	13.68
FLUXNET	DE-Geb	Croplands	161.50	51.10	10.91
FLUXNET	DE-Gri	Grasslands	385.00	50.95	13.51
FLUXNET	DE-Kli	Croplands	478.00	50.89	13.52
FLUXNET	DE-Lkb	Evergreen Needleleaf Forests	1308.00	49.10	13.30
FLUXNET	DE-Obe	Evergreen Needleleaf Forests	734.00	50.79	13.72
FLUXNET	DE-RuR	Grasslands	514.70	50.62	6.30
FLUXNET	DE-RuS	Croplands	102.76	50.87	6.45
FLUXNET	DE-SfN	Permanent Wetlands	590.00	47.81	11.33
FLUXNET	DE-Spw	Permanent Wetlands	61.00	51.89	14.03
FLUXNET	DE-Tha	Evergreen Needleleaf Forests	385.00	50.96	13.57

FLUXNET	DE-Zrk	Permanent Wetlands	0.00	53.88	12.89
FLUXNET	DK-Sor	Deciduous Broadleaf Forests	40.00	55.49	11.64
FLUXNET	FI-Hyy	Evergreen Needleleaf Forests	181.00	61.85	24.29
FLUXNET	FR-Pue	Evergreen Broadleaf Forests	270.00	43.74	3.60
FLUXNET	GF-Guy	Evergreen Broadleaf Forests	48.00	5.28	-52.92
FLUXNET	GH-Ank	Evergreen Broadleaf Forests	124.00	5.27	-2.69
FLUXNET	IT-CA1	Deciduous Broadleaf Forests	200.00	42.38	12.03
FLUXNET	IT-CA2	Croplands	200.00	42.38	12.03
FLUXNET	IT-Col	Deciduous Broadleaf Forests	1560.00	41.85	13.59
FLUXNET	IT-Isp	Deciduous Broadleaf Forests	210.00	45.81	8.63
FLUXNET	IT-Lav	Evergreen Needleleaf Forests	1353.00	45.96	11.28
FLUXNET	IT-Noe	Closed Shrublands	25.00	40.61	8.15
FLUXNET	IT-Ren	Evergreen Needleleaf Forests	1730.00	46.59	11.43
FLUXNET	IT-SR2	Evergreen Needleleaf Forests	4.00	43.73	10.29
FLUXNET	NL-Loo	Evergreen Needleleaf Forests	25.00	52.17	5.74
FLUXNET	RU-Cok	Open Shrublands	48.00	70.83	147.49
FLUXNET	RU-Fyo	Evergreen Needleleaf Forests	265.00	56.46	32.92
FLUXNET	SJ-Adv	Permanent Wetlands	17.00	78.19	15.92

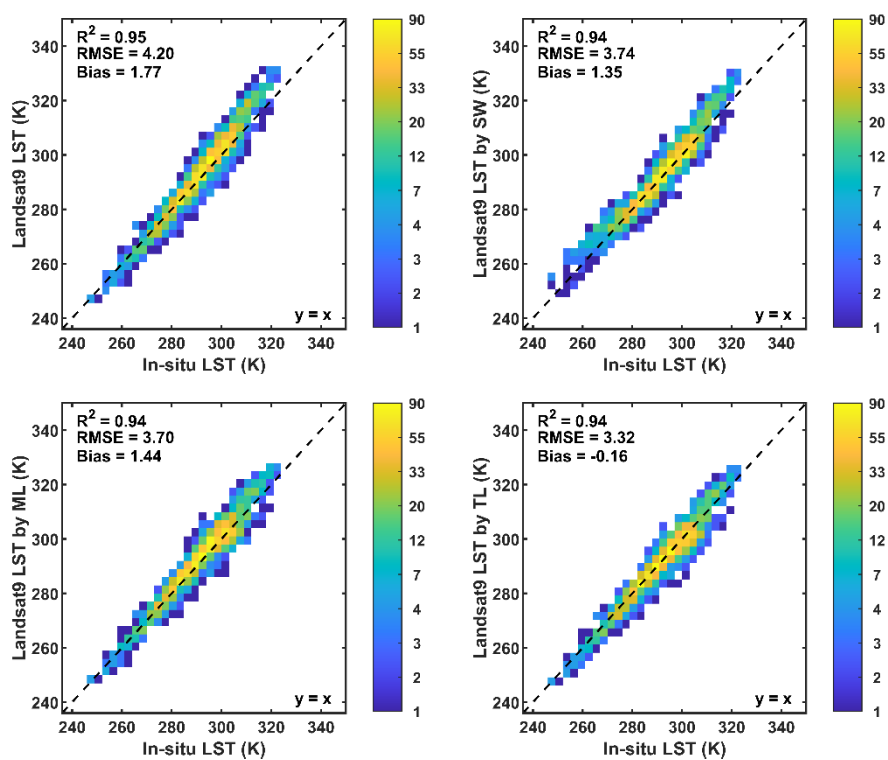


Figure S1. Comparison between the Landsat-9 land surface temperature (LST) product generated

by the single-channel (SC) algorithm, the Landsat-9 LST data generated by the split-window (SW) algorithm, the Landsat-9 LST predicted by the machine learning (ML) model, and the Landsat-9 LST predicted by the transfer-learning (TL) model over in-situ measurements.

References

- Ahmadi, M., Biswas, D., Lin, M., Vrionis, F. D., Hashemi, J., & Tang, Y. (2025). Physics-informed machine learning for advancing computational medical imaging: integrating data-driven approaches with fundamental physical principles. *Artificial intelligence review*, 58(10), 297.
- Albergel, C., De Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., . . . Wagner, W. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote sensing of environment*, 118, 215-226.
- Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., . . . Peylin, P. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, 53(3), 785-818.
- Anderson, B., Hy, T. S., & Kondor, R. (2019). Cormorant: Covariant molecular neural networks Advances in Neural Information Processing Systems. *Advances in neural information processing systems*.
- Augustine, J. A., DeLuise, J. J., & Long, C. N. (2000). SURFRAD—A national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society*, 81(10), 2341-2358.
- Augustine, J. A., Hodges, G. B., Cornwall, C. R., Michalsky, J. J., & Medina, C. I. (2005). An update on SURFRAD—The GCOS surface radiation budget network for the continental United States. *Journal of Atmospheric and Oceanic Technology*, 22(10), 1460-1472.
- Baldocchi, D. (2008). ‘Breathing’ of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, 56(1), 1-26.
- Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, 9(4), 479-492.
- Baldrige, A. M., Hook, S. J., Grove, C., & Rivera, G. (2009). The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113(4), 711-715.
- Balsamo, G., Agustí-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., . . . Brown, A. (2018). Satellite and in situ observations for advancing global Earth surface modelling: A review. *Remote Sensing*, 10(12), 2038.
- Balzarolo, M., Anderson, K., Nichol, C., Rossini, M., Vescovo, L., Arriga, N., . . . Cerasoli, S. (2011). Ground-based optical measurements at European flux sites: a review of methods, instruments and current controversies. *Sensors*, 11(8), 7954-7981.
- Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323.
- Berk, A., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., Lee, J., . . . Hoke, M. L. (2005). *MODTRAN 5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering options: update*. Paper presented at the Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery XI.
- Berk, A., Bernstein, L. S., Anderson, G. P., Acharya, P. K., Robertson, D. C., Chetwynd, J. H., & Adler-Golden, S. M. (1998). MODTRAN Cloud and Multiple Scattering Upgrades with

- Application to AVIRIS. *Remote Sensing of Environment*, 65(3), 367-375.
doi:[https://doi.org/10.1016/S0034-4257\(98\)00045-5](https://doi.org/10.1016/S0034-4257(98)00045-5)
- Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., & Van Den Bosch, J. (2014). *MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code*. Paper presented at the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS).
- Betts, A. K., Ball, J. H., Beljaars, A. C., Miller, M. J., & Viterbo, P. A. (1996). The land surface-atmosphere interaction: A review based on observational and global modeling perspectives. *Journal of geophysical research: atmospheres*, 101(D3), 7209-7225.
- Bill, K., Dieleman, C., Baltzer, J., Degre-timmons, G., Mack, M., Cumming, S., . . . Turetsky, M. (2023). Arctic-Boreal Vulnerability Experiment (ABoVE). *ORNL Distributed Active Archive Center (DAAC) dataset 10.3334/ORNLDAAC/2235 (2023, 2235*.
- Bishop, C. M. (2013). Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984).
- Blyth, E. M., Arora, V. K., Clark, D. B., Dadson, S. J., De Kauwe, M. G., Lawrence, D. M., . . . Yoshimura, K. (2021). Advances in land surface modelling. *Current Climate Change Reports*, 7(2), 45-71.
- Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, 359(6375), eaam8328.
- Bonato Asato, A. E., Wirth, C., Eisenhauer, N., & Hines, J. (2023). On the phenology of soil organisms: Current knowledge and future steps. *Ecology and Evolution*, 13(4), e10022.
- Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., & Anandkumar, A. (2023). Spherical fourier neural operators: Learning stable dynamics on the sphere International conference on machine learning. *International Conference on Machine Learning*.
- Boukabara, S.-A., Garrett, K., Chen, W., Iturbide-Sanchez, F., Grassotti, C., Kongoli, C., . . . Weng, F. (2011). MiRS: An all-weather 1DVAR satellite data assimilation and retrieval system. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), 3249-3272.
- Brunsell, N. A., & Gillies, R. R. (2003). Scale issues in land-atmosphere interactions: implications for remote sensing of the surface energy balance. *Agricultural and Forest Meteorology*, 117(3-4), 203-221.
- Camporese, M., & Girotto, M. (2022). Recent advances and opportunities in data assimilation for physics-based hydrological modeling. *Frontiers in Water*, 4, 948832.
- Carmona, F., Rivas, R., & Caselles, V. (2015). Development of a general model to estimate the instantaneous, daily, and daytime net radiation with satellite data on clear-sky days. *Remote sensing of environment*, 171, 1-13.
- Centurioni, L. R., Turton, J., Lumpkin, R., Braasch, L., Brassington, G., Chao, Y., . . . Dohan, K. (2019). Global in situ observations of essential climate and ocean variables at the air-sea interface. *Frontiers in Marine Science*, 6, 419.
- Chen, H., Luong, V., Mukherjee, L., & Singh, V. (2025). *SimpleTM: A simple baseline for multivariate time series forecasting*. Paper presented at the The Thirteenth International Conference on Learning Representations.
- Chen, J., Chen, S., Fu, R., Li, D., Jiang, H., Wang, C., . . . Hicks, B. J. (2022). Remote sensing big data for water environment monitoring: Current status, challenges, and future prospects. *Earth's Future*, 10(2), e2021EF002289.

- Chen, S., Kalanat, N., Topp, S., Sadler, J., Xie, Y., Jiang, Z., & Jia, X. (2023). Meta-transfer-learning for time series data with extreme events: An application to water temperature prediction. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*.
- Chen, S., Xie, Y., Li, X., Liang, X., & Jia, X. (2023). Physics-guided meta-learning method in baseflow prediction over large regions. *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*.
- Cheng, J., Liang, S., Nie, A., & Liu, Q. (2018). Is there a physical linkage between surface emissive and reflective variables over non-vegetated surfaces? *Journal of the Indian Society of Remote Sensing*, 46, 591-596.
- Cheng, J., Liang, S., Yao, Y., & Zhang, X. (2013). Estimating the Optimal Broadband Emissivity Spectral Range for Calculating Surface Longwave Net Radiation. *Ieee Geoscience and Remote Sensing Letters*, 10(2), 401-405. doi:10.1109/LGRS.2012.2206367
- Cheng, J., Ren, H., Liang, S., & Yan, G. (2010). Glass-Global Land Surface Broadband Emissivity Product: Algorithm Theoretical Basis Document Version 1.0. *Beijing Normal University: Beijing, China*.
- Comas, L. H., Bauerle, T., & Eissenstat, D. (2010). Biological and environmental factors controlling root dynamics and function: effects of root ageing and soil moisture. *Australian Journal of Grape and Wine Research*, 16, 131-137.
- Cook, M., Schott, J. R., Mandel, J., & Raqueno, N. (2014). Development of an Operational Calibration Methodology for the Landsat Thermal Data Archive and Initial Testing of the Atmospheric Compensation Component of a Land Surface Temperature (LST) Product from the Archive. *Remote Sensing*, 6(11), 11244-11266.
- Crisp, D., Dolman, H., Tanhua, T., McKinley, G. A., Hauck, J., Bastos, A., . . . Aich, V. (2022). How well do we understand the land-ocean-atmosphere carbon cycle? *Reviews of Geophysics*, 60(2), e2021RG000736.
- Cristóbal, J., Jiménez-Muñoz, J., Sobrino, J., Ninyerola, M., & Pons, X. (2009). Improvements in land surface temperature retrieval from the Landsat series thermal band using water vapor and air temperature. *Journal of Geophysical Research: Atmospheres*, 114(D8).
- Crow, W., Kumar, S., & Bolten, J. (2012). On the utility of land surface models for agricultural drought monitoring. *Hydrology and earth system sciences*, 16(9), 3451-3460.
- Dash, J., & Ogutu, B. O. (2016). Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems. *Progress in Physical Geography*, 40(2), 322-351.
- De Deyn, G. B., Cornelissen, J. H., & Bardgett, R. D. (2008). Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecology letters*, 11(5), 516-531.
- De Pue, J., Wieneke, S., Bastos, A., Barrios, J. M., Liu, L., Ciais, P., . . . Maignan, F. (2023). Temporal variability of observed and simulated gross primary productivity, modulated by vegetation state and hydrometeorological drivers. *Biogeosciences*, 20(23), 4795-4818.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Bauer, d. P. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553-597.
- Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., . . . Jacob, D. (2007). Couplings between changes in the climate system and biogeochemistry.

- Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change The Physical Science Basis*, 499-587.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Ding, Y., Zhang, L., He, Y., Cao, S., Wei, X., Guo, Y., . . . Filonchyk, M. (2024). Spatiotemporal evolution of agricultural drought and its attribution under different climate zones and vegetation types in the Yellow River Basin of China. *Science of The Total Environment*, 914, 169687. doi:<https://doi.org/10.1016/j.scitotenv.2023.169687>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . Hounsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*.
- Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., . . . Grobe, H. (2018). Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017). *Earth System Science Data*, 10(3), 1491-1501.
- Du, C., Ren, H., Qin, Q., Meng, J., & Zhao, S. (2015). A practical split-window algorithm for estimating land surface temperature from Landsat 8 data. *Remote Sensing*, 7(1), 647-665.
- Duan, S.-B., Li, Z.-L., Zhao, W., Wu, P., Huang, C., Han, X.-J., . . . Shang, G. (2021). Validation of Landsat land surface temperature product in the conterminous United States using in situ measurements from SURFRAD, ARM, and NDBC sites. *International Journal of Digital Earth*, 14(5), 640-660. doi:10.1080/17538947.2020.1862319
- Duveiller, G., Forzieri, G., Robertson, E., Li, W., Georgievski, G., Lawrence, P., . . . Sitch, S. (2018). Biophysics and vegetation cover change: a process-based evaluation framework for confronting land surface models with satellite observations. *Earth System Science Data*, 10(3), 1265-1279.
- Eleftheriou, D., Kiachidis, K., Kalmintzis, G., Kalea, A., Bantasis, C., Koumadoraki, P., . . . Gemitzi, A. (2018). Determination of annual and seasonal daytime and nighttime trends of MODIS LST over Greece - climate change implications. *Science of the Total Environment*, 616-617, 937-947. doi:<https://doi.org/10.1016/j.scitotenv.2017.10.226>
- Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., . . . Bugliaro, L. (2016). The libRadtran software package for radiative transfer calculations (version 2.0.1). *Geosci. Model Dev.*, 9(5), 1647-1672. doi:10.5194/gmd-9-1647-2016
- Erge, O., & Van Oort, E. (2022). Combining physics-based and data-driven modeling in well construction: Hybrid fluid dynamics modeling. *Journal of Natural Gas Science and Engineering*, 97, 104348.
- Evan Manning, J., Larrabee Strow, U., Teixeira, J., Leader, J. A. T., Fetzer, E. J., Yue, Q., & Thrastarson, H. T. (2020). AIRS-Team Retrieval For Core Products and Geophysical Parameters: Versions 6 and 7 Level 2. *Jet Propulsion*.
- Fahlen, J. E., Brodrick, P. G., Thompson, D. R., Herman, R. L., Hulley, G., Cawse-Nicholson, K., . . . Miller, C. E. (2021). Joint VSWIR-TIR retrievals of earth's surface and atmosphere. *Remote Sensing of Environment*, 267, 112727. doi:<https://doi.org/10.1016/j.rse.2021.112727>
- Favrichon, S., Prigent, C., Jimenez, C., & Aires, F. (2019). Detecting cloud contamination in passive microwave satellite measurements over land. *Atmospheric Measurement Techniques*, 12(3), 1531-1543.

- Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*.
- Feng, Z., Wang, X., Yuan, J., Zhang, Y., & Yu, M. (2023). Changes in air pollution, land surface temperature, and urban heat islands during the COVID-19 lockdown in three Chinese urban agglomerations. *Science of The Total Environment*, 892, 164496. doi:<https://doi.org/10.1016/j.scitotenv.2023.164496>
- Fensholt, R., Rasmussen, K., Kaspersen, P., Huber, S., Horion, S., & Swinnen, E. (2013). Assessing land degradation/recovery in the African Sahel from long-term earth observation based primary productivity and precipitation relationships. *Remote Sensing*, 5(2), 664-686.
- Fernández-Manso, A., Quintano, C., & Roberts, D. A. (2020). Can Landsat-derived variables related to energy balance improve understanding of burn severity from current operational techniques? *Remote Sensing*, 12(5), 890.
- Flato, G. M. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 783-800.
- Forssell, U., & Lindskog, P. (1997). Combining Semi-Physical and Neural Network Modeling: An Example of Its Usefulness. *IFAC Proceedings Volumes*.
- Fortier, M., Richter, M. L., Sonnentag, O., & Pal, C. (2024). Carbonsense: A multimodal dataset and baseline for carbon flux modelling. *arXiv preprint arXiv:2406.04940*.
- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., . . . Babst, F. (2015). Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology*, 21(8), 2861-2880.
- Freitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., & Perdigão, R. (2013). Land surface temperature from multiple geostationary satellites. *International journal of remote sensing*, 34(9-10), 3051-3068.
- Friedl, M. A. (2002). Forward and inverse modeling of land surface energy balance using surface temperature measurements. *Remote sensing of environment*, 79(2-3), 344-354.
- Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C., Hauck, J., . . . Peters, G. P. (2023). Global carbon budget 2023. *Earth System Science Data*, 15(12), 5301-5369.
- Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., . . . Olsen, A. (2024). Global carbon budget 2024. *Earth System Science Data Discussions*, 2024, 1-133.
- Gampe, D., Ludwig, R., Qahman, K., & Afifi, S. (2016). Applying the Triangle Method for the parameterization of irrigated areas as input for spatially distributed hydrological modeling — Assessing future drought risk in the Gaza Strip (Palestine). *Science of The Total Environment*, 543, 877-888. doi:<https://doi.org/10.1016/j.scitotenv.2015.07.098>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., . . . Reichle, R. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419-5454.
- Gentine, P., Massmann, A., Lintner, B. R., Hamed Alemohammad, S., Fu, R., Green, J. K., . . . Vilà-Guerau de Arellano, J. (2019). Land-atmosphere interactions in the tropics—a review. *Hydrology and earth system sciences*, 23(10), 4171-4197.

- Gillespie, A., Rokugawa, S., Matsunaga, T., Cothorn, 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 Transactions on Geoscience and Remote Sensing*, 36(4), 1113-1126.
- Grace, J. (2004). Understanding and managing the global carbon cycle. *Journal of Ecology*, 92(2), 189-202.
- Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., . . . Hood, L. (2010). Solar influences on climate. *Reviews of Geophysics*, 48(4).
- Halawa, E., van Hoof, J., & Soebarto, V. (2014). The impacts of the thermal radiation field on thermal comfort, energy consumption and control—A critical overview. *Renewable and Sustainable Energy Reviews*, 37, 907-918.
- Halder, B., Bandyopadhyay, J., & Banik, P. (2021). Evaluation of the climate change impact on urban heat island based on land surface temperature and geospatial indicators. *International Journal of Environmental Research*, 15, 819-835.
- Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*.
- Hargrove, W. W., Hoffman, F. M., & Law, B. E. (2003). New analysis reveals representativeness of the AmeriFlux network. *Eos, Transactions American Geophysical Union*, 84(48), 529-535.
- Hari, M., & Tyagi, B. (2022). Terrestrial carbon cycle: tipping edge of climate change between the atmosphere and biosphere ecosystems. *Environmental Science: Atmospheres*, 2(5), 867-890.
- Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., . . . Boettcher, M. (2023). A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models. *Earth System Science Data*, 15(3), 1465-1499.
- Haya, B. K., Evans, S., Brown, L., Bukoski, J., Butsic, V., Cabiyo, B., . . . Sanchez, D. L. (2023). Comprehensive review of carbon quantification by improved forest management offset protocols. *Frontiers in Forests and Global Change*, 6, 958879.
- He, E., Xie, Y., Liu, L., Chen, W., Jin, Z., & Jia, X. (2023). Physics guided neural networks for time-aware fairness: an application in crop yield prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Heimann, M., & Reichstein, M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, 451(7176), 289-292.
- Hennermann, K., & Berrisford, P. (2017). ERA5 data documentation. *Copernicus knowledge base*.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., . . . Rozum, I. (2018). ERA5 hourly data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 10.
- Hirschmugl, M., Gallaun, H., Dees, M., Datta, P., Deutscher, J., Koutsias, N., & Schardt, M. (2017). Methods for mapping forest disturbance and degradation from optical earth observation data: A review. *Current Forestry Reports*, 3(1), 32-45.
- Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., . . . Forsberg, R. (2013). The ESA climate change initiative: Satellite data records for essential climate variables. *Bulletin of the American Meteorological Society*, 94(10), 1541-1552.

- Holzwarth, S., Thonfeld, F., Abdullahi, S., Asam, S., Da Ponte Canova, E., Gessner, U., . . . Kuenzer, C. (2020). Earth observation based monitoring of forests in Germany: a review. *Remote Sensing*, 12(21), 3570.
- Hong, T., & Lee, S. H. (2019). Integrating physics-based models with sensor data: An inverse modeling approach. *Building and Environment*, 154, 23-31.
- Hulley, G., Freepartner, R., Malakar, N., & Sarkar, S. (2016). Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature and Emissivity Product (MxD21) User Guide. *NASA: Washington, DC, USA*.
- Hulley, G., Veraverbeke, S., & Hook, S. (2014). Thermal-based techniques for land cover change detection using a new dynamic MODIS multispectral emissivity product (MOD21). *Remote sensing of environment*, 140, 755-765.
- Hulley, G. C., & Hook, S. J. (2009). The North American ASTER Land Surface Emissivity Database (NAALSED) Version 2.0. *Remote Sensing of Environment*, 113(9), 1967-1975.
- Hurtado, D. M. (2018). Deep transfer learning in the assessment of the quality of protein models. *arXiv preprint arXiv:1804.06281*.
- Hurteau, M. D. (2021). The role of forests in the carbon cycle and in climate change *Climate change* (pp. 561-579): Elsevier.
- Hurt, G., Zhao, M., Sahajpal, R., Armstrong, A., Birdsey, R., Campbell, E., . . . Flanagan, S. (2019). Beyond MRV: high-resolution forest carbon modeling for climate mitigation planning over Maryland, USA. *Environmental Research Letters*, 14(4), 045013.
- Hurt, G. C., Moorcroft, P. R., And, S. W. P., & Levin, S. A. (1998). Terrestrial models and global change: challenges for the future. *Global Change Biology*, 4(5), 581-590.
- Jia, A., Liang, S., & Wang, D. (2022). Generating a 2-km, all-sky, hourly land surface temperature product from Advanced Baseline Imager data. *Remote Sensing of Environment*, 278, 113105. doi:<https://doi.org/10.1016/j.rse.2022.113105>
- Jia, A., Liang, S., Wang, D., Jiang, B., & Zhang, X. (2020). Air pollution slows down surface warming over the Tibetan Plateau. *Atmospheric Chemistry and Physics*, 20(2), 881-899.
- 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 System Science Data*, 15(2), 869-895.
- 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 Geoscience and Remote Sensing Magazine*, 12(4), 218-260.
- 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 sensing of environment*, 263, 112566.
- Jia, A., Mallick, K., Lin, Z., Sulis, M., Szantoi, Z., Zhang, L., . . . Roujean, J.-L. (2026). Sensitivity of thermal evapotranspiration models to surface and atmospheric drivers across ecosystems and aridity. *Agricultural and Forest Meteorology*, 376, 110930.
- Jia, A., Mallick, K., Upadhyaya, D., Hu, T., Szantoi, Z., Bhattacharya, B., . . . Ruiz, L. (2025). Deriving a clear-sky soil moisture index from ECOSTRESS land surface temperature. *Remote sensing of environment*, 329, 114945.

- Jia, H., Yang, D., Deng, W., Wei, Q., & Jiang, W. (2021). Predicting land surface temperature with geographically weighed regression and deep learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(1), e1396.
- Jia, X., Xie, Y., Li, S., Chen, S., Zwart, J., Sadler, J., . . . Read, J. (2021). Physics-guided machine learning from simulation data: An application in modeling lake and river systems. *2021 IEEE International Conference on Data Mining (ICDM)*.
- Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., . . . others. (2020). Physics-Guided Recurrent Graph Networks for Predicting Flow and Temperature in River Networks. *arXiv preprint arXiv:2009.12575*.
- Jimenez-Munoz, J. C., Cristóbal, J., Sobrino, J. A., Soria, G., Ninyerola, M., & Pons, X. (2008). Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1), 339-349.
- Jiménez-Muñoz, J. C., Sobrino, J. A., Skoković, D., Mattar, C., & Cristobal, J. (2014). Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1840-1843.
- Jones, M. O., Jones, L. A., Kimball, J. S., & McDonald, K. C. (2011). Satellite passive microwave remote sensing for monitoring global land surface phenology. *Remote sensing of environment*, 115(4), 1102-1114.
- Kahn, B., Irion, F., Dang, V., Manning, E., Nasiri, S., Naud, C., . . . Bowman, K. (2014). The atmospheric infrared sounder version 6 cloud products. *Atmospheric Chemistry and Physics*, 14(1), 399-426.
- Kalma, J. D., McVicar, T. R., & McCabe, M. F. (2008). Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, 29(4), 421-469.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422-440.
- Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., . . . Goldberg, A. (2010). Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *Journal of Climate*, 23(3), 618-633.
- Karpatne, A., Jia, X., & Kumar, V. (2024). Knowledge-guided machine learning: Current trends and future prospects. *arXiv preprint arXiv:2403.15989*.
- Kasilingam, S., Yang, R., Singh, S. K., Farahani, M. A., Rai, R., & Wuest, T. (2024). Physics-based and data-driven hybrid modeling in manufacturing: a review. *Production & Manufacturing Research*, 12(1), 2305358.
- Khaki, S., Pham, H., & Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, 11(1), 11132. doi:10.1038/s41598-021-89779-z
- Koehler, J., & Kuenzer, C. (2020). Forecasting spatio-temporal dynamics on the land surface using earth observation data—A review. *Remote Sensing*, 12(21), 3513.
- Kooistra, L., Berger, K., Brede, B., Graf, L. V., Aasen, H., Roujean, J.-L., . . . Prikaziuk, E. (2024). Reviews and syntheses: Remotely sensed optical time series for monitoring vegetation productivity. *Biogeosciences*, 21(2), 473-511.

- Kuenzer, C., Dech, S., & Wagner, W. (2015). Remote sensing time series revealing land surface dynamics: Status quo and the pathway ahead *Remote sensing time series: Revealing land surface dynamics* (pp. 1-24): Springer.
- Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., . . . Anandkumar, A. (2023). Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. *Proceedings of the platform for advanced scientific computing conference*.
- Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018). *Modeling long-and short-term temporal patterns with deep neural networks*. Paper presented at the The 41st international ACM SIGIR conference on research & development in information retrieval.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., . . . others. (2023). Learning skillful medium-range global weather forecasting. *Science*.
- Li, R., Wang, D., Liang, S., Jia, A., & Wang, Z. (2022). Estimating global downward shortwave radiation from VIIRS data using a transfer-learning neural network. *Remote Sensing of Environment*, 274, 112999. doi:<https://doi.org/10.1016/j.rse.2022.112999>
- Li, R., Xie, Y., Wang, Z., Jia, X., Mai, G., Ma, L., . . . Li, Z. (2026). EcoDiffusion: Uncertainty-Aware Emulation of Ecosystem Processes with Conditional Diffusion for Long Sequences with Single-Step Initialization. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., . . . Qi, Y. (2013). Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, 94(8), 1145-1160.
- Li, X., Liu, F., Ma, C., Hou, J., Zheng, D., Ma, H., . . . Yang, K. (2024). Land data assimilation: Harmonizing theory and data in land surface process studies. *Reviews of Geophysics*, 62(1), e2022RG000801.
- Li, Z.-L., Wu, H., Duan, S.-B., Zhao, W., Ren, H., Liu, X., . . . Zhou, C. (2023). Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications. *Reviews of Geophysics*, 61(1), e2022RG000777. doi:<https://doi.org/10.1029/2022RG000777>
- Li, Z.-L., Wu, H., Wang, N., Qiu, S., Sobrino, J. A., Wan, Z., . . . Yan, G. (2013). Land surface emissivity retrieval from satellite data. *International Journal of Remote Sensing*, 34(9-10), 3084-3127.
- Li, Z., Bai, X., Tan, Q., Zhao, C., Li, Y., Luo, G., . . . Long, M. (2024). Dryness stress weakens the sustainability of global vegetation cooling. *Science of The Total Environment*, 909, 168474. doi:<https://doi.org/10.1016/j.scitotenv.2023.168474>
- Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., . . . Maring, H. (2009). *Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective*. Paper presented at the Annales Geophysicae.
- Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., . . . Anandkumar, A. (2024). Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*.
- Li, Z. L., Wu, H., Duan, S. B., Zhao, W., Ren, H., Liu, X., . . . Zhu, J. (2023). Satellite remote sensing of global land surface temperature: Definition, methods, products, and applications. *Reviews of Geophysics*, 61(1).

- Liang, H., Fu, W., & Yi, F. (2019). *A survey of recent advances in transfer learning*. Paper presented at the 2019 IEEE 19th international conference on communication technology (ICCT).
- Liang, S. (2001). An optimization algorithm for separating land surface temperature and emissivity from multispectral thermal infrared imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2), 264-274. doi:10.1109/36.905234
- Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*, 31(5), 501-516.
- Liang, S., Wang, D., He, T., & Yu, Y. (2019). Remote sensing of earth's energy budget: Synthesis and review. *International Journal of Digital Earth*, 12(7), 737-780.
- Liao, Z., Zhou, B., Zhu, J., Jia, H., & Fei, X. (2023). A critical review of methods, principles and progress for estimating the gross primary productivity of terrestrial ecosystems. *Frontiers in Environmental Science*, 11, 1093095.
- Liou, Y.-A., & Kar, S. K. (2014). Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—A review. *Energies*, 7(5), 2821-2849.
- Liu, F., Liu, J., Zhang, Y., Hong, S., Fu, W., Wang, M., & Dong, J. (2024). Construction of a cold island network for the urban heat island effect mitigation. *Science of The Total Environment*, 915, 169950. doi:<https://doi.org/10.1016/j.scitotenv.2024.169950>
- Liu, J., Ma, X., Duan, Z., Jiang, J., Reichstein, M., & Jung, M. (2020). Impact of temporal precipitation variability on ecosystem productivity. *Wiley Interdisciplinary Reviews: Water*, 7(6), e1481.
- Liu, P., Wang, L., Ranjan, R., He, G., & Zhao, L. (2022). A survey on active deep learning: From model driven to data driven. *ACM Computing Surveys (CSUR)*, 54(10s), 1-34.
- Liu, S., Li, X., Xu, Z., Che, T., Xiao, Q., Ma, M., . . . Wang, L. (2018). The Heihe Integrated Observatory Network: A basin-scale land surface processes observatory in China. *Vadose Zone Journal*, 17(1).
- Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.
- Liu, Y., Yu, Y., Yu, P., Wang, H., & Rao, Y. (2019). Enterprise LST algorithm development and its evaluation with NOAA 20 data. *Remote Sensing*, 11(17), 2003.
- Liu, Y., Yu, Y., Yu, P., Xu, S., & Wang, H. (2024). *Validation and Performance Evaluation of NOAA-21 VIIRS LST Product*. Paper presented at the AGU Fall Meeting Abstracts.
- Lorenz, K. (2010). *Carbon sequestration in forest ecosystems*: Springer.
- Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, 21(5), 1737-1751.
- Luo, Y., & Schuur, E. A. (2020). Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Global Change Biology*, 26(3), 1109-1117.
- Ma, L., Hurtt, G., Ott, L., Sahajpal, R., Fisk, J., Lamb, R., . . . Chatterjee, A. (2021). Global evaluation of the Ecosystem Demography model (ED v3. 0). *Geoscientific Model Development Discussions*, 2021, 1-41.
- Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, 19(12), 1.

- Malakar, N. K., & Hulley, G. C. (2016). A water vapor scaling model for improved land surface temperature and emissivity separation of MODIS thermal infrared data. *Remote Sensing of Environment*, 182, 252-264.
- Malakar, N. K., Hulley, G. C., Hook, S. J., Laraby, K., Cook, M., & Schott, J. R. (2018). An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5717-5735. doi:10.1109/TGRS.2018.2824828
- Malhi, Y., Meir, P., & Brown, S. (2002). Forests, carbon and global climate. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360(1797), 1567-1591.
- Mao, K., Qin, Z., Shi, J., & Gong, P. (2005). A practical split-window algorithm for retrieving land-surface temperature from MODIS data. *International journal of remote sensing*, 26(15), 3181-3204.
- Mao, K., Shi, J., Tang, H., Li, Z.-L., Wang, X., & Chen, K.-S. (2007). A neural network technique for separating land surface emissivity and temperature from ASTER imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1), 200-208.
- Mashala, M. J., Dube, T., Mudereri, B. T., Ayisi, K. K., & Ramudzuli, M. R. (2023). A systematic review on advancements in remote sensing for assessing and monitoring land use and land cover changes impacts on surface water resources in semi-arid tropical environments. *Remote Sensing*, 15(16), 3926.
- Mattar, C., Durán-Alarcón, C., Jiménez-Muñoz, J. C., Santamaría-Artigas, A., Olivera-Guerra, L., & Sobrino, J. A. (2015). Global Atmospheric Profiles from Reanalysis Information (GAPRI): a new database for earth surface temperature retrieval. *International Journal of Remote Sensing*, 36(19-20), 5045-5060.
- Mauder, M., Foken, T., & Cuxart, J. (2020). Surface-energy-balance closure over land: a review. *Boundary-Layer Meteorology*, 177(2), 395-426.
- Mayer, B., & Kylling, A. (2005). The libRadtran software package for radiative transfer calculations-description and examples of use. *Atmospheric Chemistry and Physics*, 5(7), 1855-1877.
- McFarlane, N. (2011). Parameterizations: representing key processes in climate models without resolving them. *Wiley Interdisciplinary Reviews: Climate Change*, 2(4), 482-497.
- Meng, X., Cheng, J., Zhao, S., Liu, S., & Yao, Y. (2019). Estimating Land Surface Temperature from Landsat-8 Data using the NOAA JPSS Enterprise Algorithm. *Remote Sensing*, 11(2), 155.
- Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., & Vereecken, H. (2017). A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. *Earth System Science Data*, 9(2), 529-543.
- Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., & Gentine, P. (2021). Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real-geography boundary conditions. *Journal of Advances in Modeling Earth Systems*.
- Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological monographs*, 71(4), 557-586.

- Moura, M. M., dos Santos, A. R., Pezzopane, J. E. M., Alexandre, R. S., da Silva, S. F., Pimentel, S. M., . . . de Carvalho, J. R. (2019). Relation of El Niño and La Niña phenomena to precipitation, evapotranspiration and temperature in the Amazon basin. *Science of The Total Environment*, 651, 1639-1651. doi:<https://doi.org/10.1016/j.scitotenv.2018.09.242>
- Muralidhar, N., Islam, M. R., Marwah, M., Karpatne, A., & Ramakrishnan, N. (2018). Incorporating prior domain knowledge into deep neural networks. *IEEE Big Data*.
- Niyogi, D. (2019). Land surface processes *Current trends in the representation of physical processes in weather and climate models* (pp. 349-370): Springer.
- Ohmura, A., Dutton, E. G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., . . . Müller, G. (1998). Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, 79(10), 2115-2136.
- Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly journal of the royal meteorological society*, 108(455), 1-24.
- Olivera-Guerra, L., Quintanilla, M., Moletto-Lobos, I., Pichuante, E., Zamorano-Elgueta, C., & Mattar, C. (2022). Water dynamics over a Western Patagonian watershed: Land surface changes and human factors. *Science of The Total Environment*, 804, 150221. doi:<https://doi.org/10.1016/j.scitotenv.2021.150221>
- Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., . . . Kato, E. (2020). Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrology and earth system sciences*, 24(3), 1485-1509.
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. doi:10.1109/TKDE.2009.191
- Pandey, V., Harde, S., Rajasekaran, E., & Deb Burman, P. K. (2024). Gross primary productivity of terrestrial ecosystems: a review of observations, remote sensing, and modelling studies over South Asia. *Theoretical and Applied Climatology*, 155(9), 8461-8491.
- Pasetto, D., Arenas-Castro, S., Bustamante, J., Casagrandi, R., Chrysoulakis, N., Cord, A. F., . . . Karnieli, A. (2018). Integration of satellite remote sensing data in ecosystem modelling at local scales: Practices and trends. *Methods in Ecology and Evolution*, 9(8), 1810-1821.
- Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., . . . Papale, D. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7(1), 225. doi:10.1038/s41597-020-0534-3
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., . . . Hirsch, A. I. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proceedings of the national Academy of sciences*, 104(48), 18925-18930.
- Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14(4), 994-1016.
- Potter, C., Boriah, S., Steinbach, M., Kumar, V., & Klooster, S. (2008). Terrestrial vegetation dynamics and global climate controls. *Climate Dynamics*, 31(1), 67-78.
- Qin, Q., Wu, Z., Zhang, T., Sagan, V., Zhang, Z., Zhang, Y., . . . Xu, W. (2021). Optical and thermal remote sensing for monitoring agricultural drought. *Remote Sensing*, 13(24), 5092.

- Qin, Z., Karnieli, A., & Berliner, P. (2001). A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. *International journal of remote sensing*, 22(18), 3719-3746.
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*.
- Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the national Academy of sciences*.
- Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., . . . others. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*.
- Reiners, P., Sobrino, J., & Kuenzer, C. (2023). Satellite-derived land surface temperature dynamics in the context of global change—A review. *Remote Sensing*, 15(7), 1857.
- Runge, K., Tucker, M., Crowther, T. W., Fournier de Laurière, C., Guirado, E., Bialic-Murphy, L., & Berdugo, M. (2025). Monitoring terrestrial ecosystem resilience using earth observation data: Identifying consensus and limitations across metrics. *Global Change Biology*, 31(3), e70115.
- Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., . . . Cox, P. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, 21(5), 1762-1776.
- Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., . . . Lienert, S. (2022). Are terrestrial biosphere models fit for simulating the global land carbon sink? *Journal of Advances in Modeling Earth Systems*, 14(5), e2021MS002946.
- Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H. J., Dube, O. P., Dutreuil, S., . . . Lubchenco, J. (2020). The emergence and evolution of Earth System Science. *Nature Reviews Earth & Environment*, 1(1), 54-63.
- Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. *Energies*, 16(10), 4025.
- Sultan, M. M., Wayment-Steele, H. K., & Pande, V. S. (2018). Transferable neural networks for enhanced sampling of protein dynamics. *Journal of Chemical Theory and Computation*.
- Sun, A. Y., & Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters*, 14(7), 073001.
- Susskind, J., Blaisdell, J. M., & Iredell, L. (2014). Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. *Journal of Applied Remote Sensing*, 8(1), 084994.
- Szagri, D., Nagy, B., & Szalay, Z. (2023). How can we predict where heatwaves will have an impact? – A literature review on heat vulnerability indexes. *Urban Climate*, 52, 101711. doi:<https://doi.org/10.1016/j.uclim.2023.101711>
- Taghikhah, F., Filatova, T., & Voinov, A. (2021). Where does theory have it right? A comparison of theory-driven and empirical agent based models. *Journal of Artificial Societies and Social Simulation*.

- Taheri, M., Mohammadian, A., Ganji, F., Bigdeli, M., & Nasser, M. (2022). Energy-based approaches in estimating actual evapotranspiration focusing on land surface temperature: A review of methods, concepts, and challenges. *Energies*, 15(4), 1264.
- Takenaka, H., Nakajima, T. Y., Higurashi, A., Higuchi, A., Takamura, T., Pinker, R. T., & Nakajima, T. (2011). Estimation of solar radiation using a neural network based on radiative transfer. *Journal of Geophysical Research: Atmospheres*, 116(D8). doi:<https://doi.org/10.1029/2009JD013337>
- Tan, J., Che, T., Wang, J., Liang, J., Zhang, Y., & Ren, Z. (2021). Reconstruction of the Daily MODIS Land Surface Temperature Product Using the Two-Step Improved Similar Pixels Method. *Remote Sensing*, 13(9), 1671.
- Tang, B.-H., Shao, K., Li, Z.-L., Wu, H., & Tang, R. (2015). An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. *International Journal of Remote Sensing*, 36(19-20), 4864-4878.
- Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). CSDI: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34, 24804-24816.
- Tong, X.-Y., Xia, G.-S., Lu, Q., Shen, H., Li, S., You, S., & Zhang, L. (2020). Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237, 111322.
- Trumbore, S. (2006). Carbon respired by terrestrial ecosystems—recent progress and challenges. *Global Change Biology*, 12(2), 141-153.
- Ustin, S. L., & Middleton, E. M. (2021). Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, 10(1), 1.
- Ustin, S. L., & Middleton, E. M. (2024). Current and near-term earth-observing environmental satellites, their missions, characteristics, instruments, and applications. *Sensors*, 24(11), 3488.
- Valor, E., & Caselles, V. (2005). Validation of the vegetation cover method for land surface emissivity estimation. *Recent Research Developments in Thermal Remote Sensing*, 1-20.
- Van de Griend, A., & Owe, M. (1993). On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. *International Journal of Remote Sensing*, 14(6), 1119-1131.
- Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114(2), 449-465. doi:<https://doi.org/10.1016/j.rse.2009.10.002>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Verrelst, J., Morata, M., García-Soria, J. L., Sun, Y., Qi, J., & Rivera-Caicedo, J. P. (2025). RTM Surrogate Modeling in Optical Remote Sensing: A Review of Emulation for Vegetation and Atmosphere Applications. *Remote Sensing*, 17(21), 3618.
- Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86(3), 370-384.
- Wan, Z.-M. (2004). Land surface temperature measurements from EOS MODIS data.
- Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, 140, 36-45.

- Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), 892-905.
- Wan, Z. Y., Vlachas, P., Koumoutsakos, P., & Sapsis, T. (2018). Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLOS ONE*.
- Wang, B. (2011). Coverage problems in sensor networks: A survey. *ACM Computing Surveys (CSUR)*, 43(4), 1-53.
- Wang, B., Hua, L., Mei, H., Wu, X., Kang, Y., & Zhao, N. (2024). Impact of climate change on the dynamic processes of marine environment and feedback mechanisms: an overview. *Archives of computational methods in engineering*, 31(6), 3377-3408.
- Wang, H., Yan, H., Rong, C., Yuan, Y., Jiang, F., Han, Z., . . . Li, Y. (2024). Multi-scale simulation of complex systems: a perspective of integrating knowledge and data. *ACM Computing Surveys*, 56(12), 1-38.
- Wang, J., Li, Y., Gao, R. X., & Zhang, F. (2022). Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63, 381-391.
- 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. *Global Ecology and Biogeography*, 29(6), 1020-1033.
- Wang, R., & Yu, R. (2025). Physics-guided deep learning for dynamical systems: A survey. *ACM Computing Surveys*, 58(5), 1-31.
- Wang, X., Han, Y., Xue, W., Yang, G., & Zhang, G. J. (2022). Stable climate simulations using a realistic general circulation model with neural network parameterizations for atmospheric moist physics and radiation processes. *Geoscientific Model Development*.
- Wang, Y., Wu, H., Dong, J., Qin, G., Zhang, H., Liu, Y., . . . Long, M. (2024). Timexer: Empowering transformers for time series forecasting with exogenous variables. *Advances in neural information processing systems*, 37, 469-498.
- Wang, Z., Ma, L., Hurtt, G., Jia, X., Li, Y., Li, R., . . . Xie, Y. (2025). CarbonGlobe: A Global-Scale, Multi-Decade Dataset and Benchmark for Carbon Forecasting in Forest Ecosystems. *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*.
- Wang, Z., Xie, Y., Jia, X., Ma, L., & Hurtt, G. (2023). High-Fidelity Deep Approximation of Ecosystem Simulation over Long-Term at Large Scale. *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*.
- Wang, Z., Xie, Y., Li, Z., Jia, X., Jiang, Z., Jia, A., & Xu, S. (2024). SimFair: Physics-guided fairness-aware learning with simulation models. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), 335-344.
- Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*.
- Worden, M. A., Bilir, T. E., Bloom, A. A., Fang, J., Klinek, L. P., Konings, A. G., . . . Smallman, T. L. (2025). Combining Observations and Models: A Review of the CARDAMOM

- Framework for Data-Constrained Terrestrial Ecosystem Modeling. *Global Change Biology*, 31(8), e70462.
- Xiao, Z., Liang, S., Wang, J., Jiang, B., & Li, X. (2011). Real-time retrieval of Leaf Area Index from MODIS time series data. *Remote sensing of environment*, 115(1), 97-106.
- Xu, L., Zhang, X., Yu, H., Chen, Z., Du, W., & Chen, N. (2024). Incorporating spatial autocorrelation into deformable ConvLSTM for hourly precipitation forecasting. *Computers & Geosciences*.
- Xu, S., & Cheng, J. (2021). A new land surface temperature fusion strategy based on cumulative distribution function matching and multiresolution Kalman filtering. *Remote sensing of environment*, 254, 112256.
- Xu, S., Cheng, J., & Zhang, Q. (2019). Reconstructing All-Weather Land Surface Temperature Using the Bayesian Maximum Entropy Method Over the Tibetan Plateau and Heihe River Basin. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9), 3307-3316. doi:10.1109/JSTARS.2019.2921924
- Xu, S., Cheng, J., & Zhang, Q. (2021). A Random Forest-Based Data Fusion Method for Obtaining All-Weather Land Surface Temperature with High Spatial Resolution. *Remote Sensing*, 13(11), 2211.
- Xu, S., Liu, Y., Yu, Y., & Yu, P. (2024). *All-weather LST: Methodology and Experiment on JPSS/VIIRS LST*. Paper presented at the AGU Fall Meeting Abstracts.
- Xu, S., & Wang, D. (2022). *Validation of gridded datasets of near-surface temperature variables: uncertainties, spatial heterogeneity and clear-sky bias*. Paper presented at the AGU Fall Meeting Abstracts.
- Xu, S., Wang, D., Liang, S., Jia, A., Li, R., Wang, Z., & Liu, Y. (2024). A novel approach to estimate land surface temperature from Landsat top-of-atmosphere reflective and emissive data using transfer-learning neural network. *Science of the Total Environment*, 955, 176783.
- Xu, S., Wang, D., Liang, S., Liu, Y., & Jia, A. (2023). Assessing the reliability of the MODIS LST product to detect temporal variability. *IEEE Geoscience and Remote Sensing Letters*, 1-1. doi:10.1109/LGRS.2023.3312384
- Xu, S., Wang, D., Liang, S., Liu, Y., & Jia, A. (2023). Assessment of gridded datasets of various near surface temperature variables over Heihe River Basin: Uncertainties, spatial heterogeneity and clear-sky bias. *International Journal of Applied Earth Observation and Geoinformation*, 120, 103347. doi:<https://doi.org/10.1016/j.jag.2023.103347>
- Xu, S., Wang, Z., Wang, R., Jia, X., Ma, L., Hurtt, G. C., & Xie, Y. (2025). Knowledge-guided assimilation to bridge the gap between sensing and modeling with indirect labels for global-scale carbon monitoring.
- Xu, T., Liu, S., Xu, L., Chen, Y., Jia, Z., Xu, Z., & Nielson, J. (2015). Temporal upscaling and reconstruction of thermal remotely sensed instantaneous evapotranspiration. *Remote Sensing*, 7(3), 3400-3425.
- Xu, T., & Valocchi, A. J. (2015). Data-driven methods to improve baseflow prediction of a regional groundwater model. *Computers & Geosciences*.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., . . . Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, 3(10), 875-883.

- Yang, Y., Guangrong, S., Chen, Z., Hao, S., Zhouyiling, Z., & Shan, Y. (2022). Quantitative analysis and prediction of urban heat island intensity on urban-rural gradient: A case study of Shanghai. *Science of The Total Environment*, 829, 154264. doi:<https://doi.org/10.1016/j.scitotenv.2022.154264>
- Ye, X., Ren, H., Zhu, J., Fan, W., & Qin, Q. (2022). Split-Window Algorithm for Land Surface Temperature Retrieval From Landsat-9 Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5. doi:10.1109/LGRS.2022.3184980
- Yu, R., Qiu, C., Ladwig, R., Hanson, P., Xie, Y., & Jia, X. (2025). Physics-guided foundation model for scientific discovery: An application to aquatic science. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Yu, R., Qiu, C., Ladwig, R., Hanson, P. C., Xie, Y., Li, Y., & Jia, X. (2024). Adaptive process-guided learning: An application in predicting lake do concentrations. *2024 IEEE International Conference on Data Mining (ICDM)*.
- Yu, S., Hannah, W., Peng, L., Lin, J., Bhouri, M. A., & Gupta, R. (2024). ClimSim: A large multi-scale dataset for hybrid physics-ML climate emulation. *Advances in neural information processing systems*.
- Yu, Y., Pinheiro, A. C., & Privette, J. L. (2006). *Correcting land surface temperature measurements for directional emissivity over 3D structured vegetation*. Paper presented at the Remote Sensing and Modeling of Ecosystems for Sustainability III.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., . . . Wang, J. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote sensing of environment*, 241, 111716.
- Yuan, W., Tian, J., Wang, M., Wang, S., Xu, W., Wang, Y., . . . Zhou, S. (2025). Impacts of rising atmospheric dryness on terrestrial ecosystem carbon cycle. *Nature Reviews Earth & Environment*, 6(11), 712-727.
- Yunhao, Z., & Junchi, Y. (2023). Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. *International Conference on Learning Representations*.
- Zadbagher, E., Marangoz, A. M., & Becek, K. (2023). Characterizing and estimating forest structure using active remote sensing: An overview. *Advanced Remote Sensing*, 3(1), 38-46.
- Zhang, C., Brodylo, D., Rahman, M., Rahman, M. A., Douglas, T. A., & Comas, X. (2022). Using an object-based machine learning ensemble approach to upscale evapotranspiration measured from eddy covariance towers in a subtropical wetland. *Science of The Total Environment*, 831, 154969. doi:<https://doi.org/10.1016/j.scitotenv.2022.154969>
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, 561, 918-929.
- Zhang, Q., Wang, N., Cheng, J., & Xu, S. (2020). A Stepwise Downscaling Method for Generating High-Resolution Land Surface Temperature From AMSR-E Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5669-5681. doi:10.1109/JSTARS.2020.3022997
- Zhang, Q., Zhou, T., Zeng, J., Zhang, Y., Zhang, J., Wu, X., . . . Qu, Y. (2025). Deep learning-based identification of environmental memory effects on the gross primary productivity of alpine grasslands on the Qinghai-Tibet Plateau. *Ecological Processes*, 14(1), 82.

- Zhang, X., Zhou, J., Göttsche, F.-M., Zhan, W., Liu, S., & Cao, R. (2019). A method based on temporal component decomposition for estimating 1-km all-weather land surface temperature by merging satellite thermal infrared and passive microwave observations. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7), 4670-4691.
- Zhang, Y., He, T., Liang, S., Wang, D., & Yu, Y. (2018). Estimation of all-sky instantaneous surface incident shortwave radiation from Moderate Resolution Imaging Spectroradiometer data using optimization method. *Remote Sensing of Environment*, 209, 468-479. doi:<https://doi.org/10.1016/j.rse.2018.02.052>
- Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., & Gong, P. (2022). An overview of the applications of earth observation satellite data: impacts and future trends. *Remote Sensing*, 14(8), 1863.
- Zhao, W., & Li, A. (2015). A review on land surface processes modelling over complex terrain. *Advances in Meteorology*, 2015(1), 607181.
- Zhou, C., Ye, M., Xia, Z., Wang, W., Luo, C., & Muller, J.-P. (2025). An interpretable attention-based deep learning method for landslide prediction based on multi-temporal InSAR time series: A case study of Xinpu landslide in the TGRA. *Remote sensing of environment*, 318, 114580.
- Zhou, C., Zhang, J., Liu, J., Zhang, C., Shi, G., & Hu, J. (2020). Bayesian transfer learning for object detection in optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11), 7705-7719.
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Zhou, J., Yang, Q., Liu, L., Kang, Y., Jia, X., Chen, M., . . . Jin, Z. (2023). A deep transfer learning framework for mapping high spatiotemporal resolution LAI. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 30-48. doi:<https://doi.org/10.1016/j.isprsjprs.2023.10.017>
- Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). *Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting*. Paper presented at the International Conference on Machine Learning.
- Zhu, J.-J., Yang, M., & Ren, Z. J. (2023). Machine learning in environmental research: common pitfalls and best practices. *Environmental Science & Technology*, 57(46), 17671-17689.
- Zrira, N., Kamal-Idrissi, A., Farssi, R., & Khan, H. A. (2024). Time series prediction of sea surface temperature based on BiLSTM model with attention mechanism. *Journal of Sea Research*, 198, 102472.