ABSTRACT

Title of Dissertation: DEEP LEARNING APPROACHES FOR ESTIMATING AND FORECASTING SURFACE DOWNWARD SHORTWAVE RADIATION FROM SATELLITE DATA

Ruohan Li, Doctor of Philosophy, 2023

Dissertation directed by: Dongdong Wang, Associate Professor
Department of Geographical Sciences

Surface downward shortwave radiation (DSR) designates solar radiation with a wavelength from 300 to 4000 nm received at the Earth's surface. DSR plays a pivotal role in the surface energy and radiation budget, serving as the primary driver for hydrological, ecological, and biogeochemical cycles (Liang et al., 2010, 2019), and the important input for various earth models (Huang et al., 2019; Liang et al., 2010; Stephens et al., 2012). Given the rising demand for renewable energy, as well as accelerated advancements in solar energy technologies on both utility-scale and residential scale, the precision and resolution in estimating and forecasting DSR have become indispensable for planning and administering solar power plants (Gueymard, 2014; Jiang et al., 2019). This dissertation delves into the potential of integrating deep learning with satellite observations to address the deficiencies in current DSR estimation and forecasting methods, aiming to cater to the evolving needs of solar radiation estimation. The research begins by examining current DSR satellite products, emphasizing their limitations, particularly concerning spatial resolution and performance in snowy, cloudy, and high-latitude areas. In such

regions, challenges arise from the degradation of radiative transfer models, band saturation, the pronounced effects of 3D cloud dynamics, and temporal resolution constraints (Li et al., 2021). Identifying these gaps, the study introduces the concept of transfer learning to tackle cases where physical methods degrade and limited training data is available. By combining data from physical simulations and ground observations, the proposed models enhance both the accuracy and adaptability of DSR predictions on a global scale. The investigation further reveals the influence of training data volume on model performance, illustrating how transfer learning can ameliorate these effects (Li et al., 2022). Moreover, the dissertation compares the application of DenseNet, Gated Recurrent Unit (GRU), and a hybrid of Convolutional Neural Network (CNN) and GRU (CNNGRU) to geostationary satellite data, achieving precise and timely DSR estimates. These models underscore their prowess in tackling 3D cloud effects and reducing dependency on additional data sources by the spatial and temporal structure of DL (Li et al., 2023b). Finally, the dissertation introduces the SolarFormer, a space-time transformer neural network adept at forecasting solar radiation up to three hours in advance at 15-minute intervals. By harnessing solely geostationary satellite imagery without the need for ground measurements, this model facilitates expansive DSR predictions, which are crucial for optimizing solar energy distribution at both utility and micro scales. This chapter also highlights the Transformer model's potential for extended forecasting due to its computational and memory efficiency.

DEEP LEARNING APPROACHES FOR ESTIMATING AND FORECASTING SURFACE DOWNWARD SHORTWAVE RADIATION FROM SATELLITE DATA

by

Ruohan Li

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
2023

Advisory Committee:
Dr. Dongdong Wang, Chair
Dr. Giovanni Baiocchi
Dr. Matthew C. Hansen
Dr. Zhanqing Li, Dean's Representative
Dr. Yiqun Xie

Dedication

To my parents, *Jun Li* and *Lingling Ren*, whose unconditional love and support have been the cornerstone of my PhD journey. Life is a circle; once, you acknowledged me in your dissertation, and now, with deep honor, I acknowledge you in mine.

To my grandparents, *Ronghua Li, Minyan Cao, Zhenfang Li*, and *Bingle Ren* — your strength, wisdom, integrity, and love have shaped the family I hold with immense pride.

Acknowledgements

This dissertation represents not just my effort but is the culmination of support, guidance, and companionship from numerous individuals.

Firstly, I extend my deepest gratitude to my advisor, Prof. Dongdong Wang, for his unwavering support, mentorship, and invaluable guidance throughout my Ph.D. journey. He has always respected my choices—be it in research direction, career path, or personal life—offering advice and resources with my best interests at heart. His approach to thinking, questioning, and conducting research has been enlightening. His generosity in sharing knowledge has been immensely appreciated, making my Ph.D. journey a truly enjoyable experience under his supervision.

My sincere appreciation is also directed to my committee members. I thank Prof. Zhanqing Li for his thorough examination of this dissertation and the collaboration opportunities he has provided. I am grateful to Prof. Yiquan Xie for his expertise in the cutting-edge field of AI and for continually sparking new avenues for my research. I also wish to express my gratitude to Prof. Giovanni Baiocchi and Prof. Matthew Hansen for their expertise and commitment to ensuring the quality of this dissertation. It was a great honor to have them all in my committee.

I am thankful to Prof. Shunlin Liang for initially providing me with this opportunity and for his ongoing guidance throughout my Ph.D. His insights into quantitative remote sensing have been invaluable. Prof. Sinead Farrell deserves special mention for offering me my first GRA position and for her consistent encouragement and support. She is a great female academic inspiration to me.

Special thanks go to Jiena and Zhihao, who have been my pillars during the most challenging times of my Ph.D. journey. Jiena's relaxed demeanor and warmth have been healing. Her unwavering support, both emotionally and physically, has filled me with happiness and a sense of fortune to have her by my side. For Zhihao, the in-depth discussions about our past, present, and future have been profoundly motivating. I look forward to continuing our journey supporting each other's efforts into the future.

I also wish to extend my gratitude to Yanjun, Yuehui, Zhili, Weiye for our daily dining hall chats about both research and life, and for the memorable potluck dinners. You all nailed the cooking. I am thankful for my colleagues Amber, Xinyuan, Guangxiao, Yiming, and Baobao for being there and supporting each other from the beginning of this journey. My gratitude also goes to my graduated colleagues Yanjia, Diyang, Jiaming, Aolin, Jing, Zheng, and Lei for their guidance and care. Lastly, I thank Di and Hanwen for their companionship and encouragement during my two-year stay in the Bay Area.

I would also like to extend my gratitude to the Surface Radiation Budget Network (SURFRAD), Baseline Surface Radiation Network (BSRN), FLUXNET, and the Greenland Climate Network (GC-NET) for providing the essential data that made this research possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Aerosol optical depth (AOD)

Advanced Baseline Imager (ABI)

Advanced Himawari Imager (AHI)

Breathing Earth System Simulator (BESS)

Blue sky albedo (BSA)

Clouds, Albedo and Radiation Edition 2 data (CLARA),

Clouds and the Earth's Radiant Energy System (CERES)

Convolutional neural network (CNN)

Convolutional neural network with gated recurrent unit (CNNGRU)

Deep Learning (DL)

Downward Shortwave Radiation (DSR)

Global Land Surface Satellite Downward Shortwave Radiation (GLASS)

Gated Recurrent Unit (GRU)

Machine learning (ML)

Moderate Resolution Imaging Spectroradiometer (MODIS)

Look-Up Tables (LUT)

Photovoltaic (PV)

Radiative transfer models (RTMs)

Surface Radiation Budget Network (SURFRAD)

Transfer learning (TL)

Top-Of-Atmosphere (TOA)

Visible Infrared Imaging Radiometer Suite (VIIRS)

Moderate Resolution Imaging Spectroradiometer (MODIS)

MODIS land surface Downward Shortwave Radiation (MCD18)

MODIS daily surface albedo product (MCD43)

Chapter 1: Introduction

1.1 Background and motivations

Surface Downward Shortwave Radiation (DSR) holds a pivotal role in shaping the surface energy and radiation budget, serving as a crucial driver for numerous hydrological, ecological, and biogeochemical surface processes (Liang et al., 2010, 2019). This component fosters the exchange of water, energy, carbon, and additional elements across the land surface, atmosphere, and oceans, thus establishing its significance in modeling assorted land surface processes (Huang et al., 2019; Liang et al., 2010; Stephens et al., 2012). Given the accelerated advancements in solar energy technologies paired with a rising demand for clean, renewable energy, the precision in estimating and forecasting DSR has become indispensable for adeptly planning and administering solar power plants (Gueymard, 2014; Jiang et al., 2019).

As a critical component of many land and atmospheric modeling, the estimation of DSR can be dated back to years (Pinker, 1995; Liang et al., 2006; Li et al., 1993). DSR, influenced by atmospheric and surface attenuation, and interactions with extraterrestrial solar irradiance, fluctuates with the Earth-Sun distance and solar elevation angle (Wang et al., 2020). Numerical models yield spatiotemporally continuous DSR, with advantages being their completeness and consistency, crucial for long-term climate monitoring and analysis. Amongst DSR products derived from numerical models, commonly used are reanalysis products, such as the Modern Era Retrospective analysis for Research and Applications (MERRA) and the ERA-Interim reanalysis from the European Centre for Medium-Range Weather Forecasts (Decker et al., 2012). However, these models notably struggle with accurate simulation or prediction of cloud quantities, casting doubt on the quality of DSR products, particularly those with high temporal resolution (Zhao et al., 2013).

Progress in estimating DSR from satellite observations has been significant (Mueller et al., 2009; Rutan et al., 2015; Wang et al., 2020; Zhang X. et al., 2014). Existing methods fall into two categories: forward computation from simplified atmospheric and surface parameters input (Huang et al., 2018; Xie et al., 2016) and inverse modeling from Top-Of-Atmosphere (TOA) satellite data (Pinker, 1995). Inverse methods vary: some, grounded in atmospheric radiative transfer theory, derive DSR through inversion of physical models like Look-Up Tables (LUT) (Liang et al., 2006) and optimization (Zhang et al., 2018). Others apply a more statistical approach, leveraging empirical relationships between satellite data and DSR, e.g., direct estimation (Zhang et al., 2019) and machine learning methods (Brown et al., 2020). Wang et al. (2020) offer a detailed method comparison.

Numerous satellite DSR products exist, with several global products derived from polar-orbiting satellite data or a combination with geostationary data, including the Clouds and the Earth's Radiant Energy System (CERES; Kato et al., 2013), Global Energy and Water Cycle Experiment - Surface Radiation Budget (GEWEX/SRB; Pinker & Laszlo, 1992), and International Satellite Cloud Climatology Project -Flux Data (ISCCP-FD; Zhang et al., 2004). Additionally, regional products have emanated from geostationary satellites like Himawari-8 (Yu et al., 2018), METEOSAT Second Generation (Pfeifroth et al., 2018), and Goddard Earth Observing System (GEOS).

Distinct applications and research studies impose varied demands concerning DSR estimation in terms of spatial resolution, temporal scale, coverage, and quality levels. Table 1-1 summarizes user-defined requirements for DSR for different application, collected by WMO Observing System Capability Analysis and Review Tool (available at: https://www.wmo sat.info/oscar/variables/view/50). The apex spatial resolution objective is 1 km, primarily for applications in agricultural meteorology. On the other hand, the temporal resolution goals is 60 seconds, with a minimum of one hour resolution for solar nowcasting. The target for uncertainty is established at 1 W/m^2, while the inferior threshold for quality levels is 20 W/m^2. Moreover, the increasing deployment of micro-scale photovoltaic (PV) systems, attributed to their minimized environmental impact (Rishee et al., 2017), propels the demand for DSR estimation at higher spatial and temporal resolutions. The appetite for economic and stable integration of PV into pre-existing grids (Marcos et al., 2014), requires precise DSR estimation and forecasting. Whether current DSR estimation methods and products can fulfill the requirements especially in solar energy production still need fully examined. Additionally, the current state of DSR estimation confronts several issues, impeding its practical applications. For instance, degradation issues in high-latitude areas and snow-covered regions adversely affect climate monitoring and energy balance studies, particularly over the Arctic and Antarctic regions (Hollmann et al., 2006). The assumption encapsulated in the Independent Pixel Approximation (IPA), which neglects the three-dimensional (3D) cloud effects—such as non-local cloud shadows and photon diffusion from adjacent columns—is deemed insufficient for high-resolution products (Chen et al., 2019). A study by Wang et al. (2017)

highlighted that the shadowing effect, induced by the solar-cloud-satellite geometry, can potentially lead to errors of up to 80% in DSR estimation. Furthermore, the reliance on additional inputs restricts the application of real-time estimation and likewise impacts real-time nowcasting adversely (Li et al., 2023).

Table 1-1. Quantitative requirements for DSR collected by WMO OSCAR in weather, water, and climate related application areas (Huang et al. 2019)

Application areas	Spatial resolution(km)			Temporal scale			Uncertainty (RMSE Wm^{-2})		
	Goal	Break	Thres.	Goal	Break	Thres.	Goal	Break	Thres.
Numerical Weather Prediction	10	30	100	60m	3 h	12 h	1	10	20
Agricultural meteorology	1	5	20	24 h	2 d	7 d	–	–	–
Nowcasting	5	15	50	60 s	10m	60m	1	10	20
Climate monitoring	25	50	100	24 h	2 d	5 d	5	6.5	10

Machine learning (ML) especially the deep learning (DL) has seen dramatic advances in recent years across numerous domains including audio, visual data processing, and natural language processing, presents a potentially transformative approach to improving DSR estimation and forecasting, especially given its adeptness at approximating complex nonlinear relationships through multi-layer learning with ample training data (Bengio et al., 2013; Alzubaidi et al., 2021). In light of the growing of "big data" from satellite observations, DL models offer a significant opportunity to explore spectral, spatial, and temporal relationship within satellite obaservations to deliver enriched outcomes in remote sensing retrieval, fusion, downscaling, and other applications, thereby providing fresh ways for novel methods in earth and environmental monitoring (Yuan et al., 2020). While numerous studies have successfully employed DL in various aspects of environmental monitoring—including in the realms of atmosphere, vegetation, hydrology, and land surface types—there remains a gap concerning the application and optimization of DL in enhancing DSR estimation and forecasting. Notwithstanding, the exploration of DL models, such as Convolutional Neural Networks (CNNs) for spatial pattern recognition, Recurrent Neural Networks (RNNs) for deciphering temporal dependencies in DSR time-series data, and Generative Adversarial Networks (GANs) for generating synthetic satellite imagery under varied atmospheric conditions, could fortify DSR estimation and forecasting, opening up promising pathways for future research and development in this domain. Moreover, challenges related to model generalization across diverse geographical and temporal contexts, and robustness against scarce conditions to affirm the practicality and dependability of DL in DSR applications has not been examined yet.

1.2 Research objectives and dissertation structure

The overarching goal is to explore the potential for deep learning models to improve DSR estimation and forecasting, utilizing satellite observations and addressing limitations inherent in existing physical approaches. The overarching question of the dissertation is to what degree the neural network can enhance DSR estimation and prediction capabilities compared with physical methods. The dissertation is structured across six chapter. Chapter 1 outlines the background, its motivation, the questions driving this research, and an overview of the dissertation's layout.

Chapter 2 addresses the research question: "Do current DSR products meet the demands of solar radiation applications?" To answer this, the chapter offers a thorough evaluation of five contemporary global daily DSR satellite products. The aim is to gauge the performance of these satellite products considering varying spatial resolutions, cloud and snow conditions, latitudes, and both intra-annual and inter-annual accuracy. Notably, data from polar regions, often overlooked in similar studies, are incorporated. By doing so, the chapter exposes the limitations of current DSR satellite products, setting the trajectory for subsequent chapters.

Chapter 3 addresses the research question: "How can transfer learning tackle the degradation of DSR estimation in challenging scenarios?" Given that traditional physical methods degrade in accurately estimating DSR in cloudy, snowy, and high-latitude areas – coupled with the scarcity of training data for these conditions using traditional machine learning – this chapter brings concept of transfer learning to handle the issue. By training data from both physical simulations and ground measurements, the objective is to boost the accuracy and generalizability of DSR predictions. Additionally, it examines the impact of training data volume on model efficacy and how transfer learning can help offset these effects.

Chapter 4 seeks to answer: "How can deep neural networks navigate the challenges posed by 3D cloud effects and dependency on supplemental data for granular DSR estimations?" The chapter aims to explore the potential of DL structure on geostationary satellite observations and enable the DSR estimation at high accuracy and low latency at 15/10 minute interval with 1km spatial resolution—ideal for solar nowcasting. DenseNet, Gated Recurrent Unit (GRU) and Convolutional neural network (CNN) with gated recurrent unit (CNNGRU) models, emphasizing their capabilities to address the 3D cloud effects and minimizing reliance on auxiliary data inputs by their spatial and temporal structure both individually and in conjunction.

Chapter 5 seeks to answer: "How can transformer facilitate intra-day DSR nowcasting using only satellite imagery?" Implementing a space-time transformer neural network, this chapter demonstrates the feasibility of forecasting solar radiation three hours ahead with 15-minute intervals, building upon near real-time DSR estimates from Chapter 4. By exclusively leveraging geostationary satellite images, the proposed SolarFormer model does not use ground measurements as inputs, facilitating large-scale DSR predictions vital for optimizing solar energy dispatching for both utility scale and micro scale. The chapter also explores the scalability and lead-time potential of the Transformer model.

Chapter 6 summarizes the main findings across all chapters and lists potential directions for future research.

Chapter 2: Comprehensive Assessment of Five Global Daily Downward Shortwave Radiation Satellite Products

Five updated global satellite daily DSR products, namely the Earth's Radiant Energy System Synoptic TOA and surface fluxes and clouds (CERES), Clouds, Albedo and Radiation Edition 2 data (CLARA) (Karlsson et al., 2017), Global Land Surface Satellite Downward Shortwave Radiation (GLASS) (Zhang et al., 2014), Breathing Earth System Simulator (BESS) (Ryu et al., 2018) shortwave radiation product, and Moderate Resolution Imaging Spectroradiometer land surface Downward Shortwave Radiation (MCD18) (Wang, 2021), were assessed using in situ measurements from 142 global sites, with an emphasis on high latitudes in 2004. The products were compared at various spatial resolutions across distinct subareas. Their overall accuracy improved upon aggregation to 100 km. At this resolution, the global root-mean-square errors for CERES, CLARA, GLASS, BESS, and MCD18 were 27.6, 29.1, 30.3, 29.6, and 31.6 W/m^2, respectively. Most products displayed accuracies about 7 W/m^2 lower in high-latitude areas. All products revealed a noticeable seasonal variation in accuracy, especially pronounced in high-latitude regions. Both GLASS and MCD18 performed commendably over coastal regions but struggled in snow-covered areas. All products degrades over cloudy conditions. Over the long term, both in situ data and BESS and CERES products indicated negligible trends, whereas CLARA and GLASS demonstrated a dimming trend. This study points to potential enhancements in current high-resolution DSR retrieval algorithms to bolster accuracy (Li et al., 2021). This study has been published (Li et al., 2021).

2.1 Introduction

DSR is a critical component of the surface energy and radiation budget. It is also a fundamental driving force of many hydrological, ecological, and biogeochemical surface processes (Liang et al., 2010, 2019). The DSR promotes the exchange of water, energy, carbon, and other agents among the land surface, atmosphere, and oceans and thus is of particular importance for modeling various land surface processes (Huang et al., 2019; Liang et al., 2010; Stephens et al., 2012). With the rapid development of solar energy technologies and the increasing demand for renewable and clean energy, it is necessary to precisely estimate the incident solar energy to plan and manage solar power plants (Gueymard, 2014; Jiang et al., 2019). Many DSR validations and assessments have been conducted (Boilley & Wald, 2015; Gui et al., 2010; Jia et al., 2013; Pfeifroth et al., 2018; Posselt et al., 2012; Sanchez-Lorenzo et al., 2015; Sun et al., 2018; Urraca et al., 2017; Wu & Fu, 2011; Xia et al., 2006; Yu et al., 2018, 2019; X. Zhang et al., 2015; X. Zhang et al., 2016). However, most of them were focused on regional areas such as Europe or East Asia. With respect to global analysis, the polar region has rarely been fully investigated because of the limited availability of field measurement stations over high-latitude areas. However, the shortwave radiation over high-latitude areas is extremely important for studies of global climate change due to the unique location and weather characteristics of the polar region (Sun et al., 2018). It influences the variability of the polar climate and ecological environment by directly or indirectly affecting snow melt and refreezing, the sea ice extent, surface albedo, and oceanic circulation (Kapsch et al., 2016; Zhang et al., 2011). In addition, it is difficult to accurately estimate the DSR by satellites because of the low solar incident angle and high snow coverage over high-latitude areas and thus comprehensive analyses are needed. Previous global assessments were mostly based on CERES, GEWEX/SRB, ISCCP-FD, or reanalysis data such as ERA-Interim (Gui et al., 2010; T. Zhang et al., 2013; X. Zhang et al., 2015). However, several of these products have not been updated since the 2000s and have a coarse resolution (Urraca et al., 2017; X. Zhang et al., 2015). In recent years, several new DSR products with higher resolution have been developed for land surface applications. Clouds, Albedo and Radiation Edition 2 data (CLARA-A2) with a 25 km resolution were published in 2017 (Karlsson et al., 2017). The Global Land Surface Satellite Downward Shortwave Radiation (GLASS-DSR) product was first generated in 2013 and has a 5 km resolution (X. Zhang et al., 2014). The Moderate Resolution Imaging Spectroradiometer (MODIS) land surface DSR product (MCD18A1) produced by the National Aeronautics and Space Administration (NASA) provides global 1-km resolution DSR estimates over land surfaces and has been made available to the public in 2021 (D. Wang, 2021). The BESS shortwave radiation product was released in 2017 with 5 km resolution and covers a temporal period from 2000 till now (Ryu et al., 2018). The Earth Polychromatic Imaging Camera (EPIC) data onboard the Deep Space Climate Ovservatory (DSCOVER/EPIC) also generate the shortwave radiation dataset with 0.1 degree resolution from 2015 to 2019 (Hao et al., 2020). However, the analyses of these high-resolutions products remain limited.

Therefore, the aim of this study was to comprehensively validate global daily DSR satellite products including the CERES-SYN, CLARA-A2, GLASS, BESS, and

MCD18A1. The primary goal of this study was to update the DSR satellite validation and fill data gaps over high-latitude areas. In this study, 142 global ground measurement stations were employed among which 41 are located at latitudes ranging from 60° to 90°. The spatial distributions of the data availability and accuracy, monthly variations among latitudes, and long-term trend were examined. In addition, the MCD18A1 product was used as an example to demonstrate how potential algorithm refinements can improve the accuracy of the DSR estimation.

This study is organized as follows. The satellite products and ground measurement data are described in Section 2.2. The data quality control, data aggregation, and validation processes are described in Section 2.3. The overall accuracy of the daily DSR estimates of the five products and the analysis of the spatial distribution, intraannual analysis, interannual analysis, and quantitative analysis of potential error sources are presented and discussed in Section 2.4. Several strategies regarding the improvement of future algorithms are discussed in Section 2.5. The main conclusions are summarized in Section 2.6.

2.2 Data

2.2.1 Satellite DSR products

Five global satellite DSR products with different spatial resolutions were evaluated in this study. All five products are based on different methods and input data. The products are compared in detail in Table 2-1.

Table 2-1. Summary of the five satellite DSR products.

Product	**Spatial resolution**	**Coverage**	**Main input**	**Method**

CERES-SYN	1° (~100 km)	2000–present	CERES, MODIS, GEOS-4/5, etc.	Radiative Transfer Code
CLARA-A2	0.25° (~25 km)	1982–2015	AVHRR	LUT
GLASS-DSR	0.05° (~5 km)	2000–2019	MODIS	Direct estimation
MCD18A1	1 km	2000–present	MODIS	LUT
BESS	0.05° (~5 km)	2000-2019	MODIS, MERRA	ANN

The CERES Synoptic (SYN1deg) product provides a global dataset of radiant fluxes at the surface, TOA, and in various atmospheric layers. The Langley Fu-Liou radiative transfer code was employed to calculate the surface shortwave radiant fluxes based on satellite observations from CERES, MODIS, geostationary data, and reanalysis data from the GEOS and other sources (Rutan et al., 2015). The CERES-SYN data span from 2000 to the present and have a 1° spatial resolution. The data have been widely used and the data accuracy has been compared with various ground measurements and other satellite products (Riihelä et al., 2017; Rutan et al., 2015; Sun et al., 2018; Yu et al., 2018).

The CLARA-A2 data were published by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application on Climate Monitoring (CM SAF) in 2017. They are based on polar-orbiting National Oceanic and Atmospheric Administration and Metop satellite measurements from the Advanced Very High Resolution Radiometer (AVHRR). They have a global coverage and a spatial resolution of 25 km. The surface downwelling shortwave radiation was retrieved by employing the mesoscale atmospheric global irradiance code, which includes an updated version of the LUT (Mueller et al., 2009) considering the effects of aerosols, water vapor, and the surface albedo (Karlsson et al., 2013). In the second edition of CLARA data, the preprocessing of the input AVHRR level-1 TOA radiance data was improved (Karlsson et al., 2017). The CLARA-A2 now covers all cloud information, surface albedo, and surface radiation from 1983–2019.

The GLASS-DSR product (Liang et al., 2021) covers the period from 2000 to present with a spatial resolution of 5 km on a daily basis. The data are generated from the MODIS TOA spectral reflectance and surface albedo using a direct estimation approach (X. Zhang et al., 2019). The retrieval algorithm is composed of two major steps. First, the surface shortwave net radiation (SSNR) is calculated from the MODIS TOA reflectance based on a linear regression (D. Wang et al., 2015). Subsequently, the GLASS broadband albedo product is used to estimate the DSR (X. Zhang et al., 2019). The accuracy of the product has been validated by using Baseline Surface Radiation Network (BSRN), the Global Energy Balance Archive, and the Climate Data Center of the China Meteorological Administration (X. Zhang et al., 2019).

The MCD18A1 is a recently developed global high-resolution product of the DSR based on the combination of MODIS Terra and Aqua. The product employs the LUT approach in which the TOA reflectance is used as the main input (D. Wang et al., 2020). Without further input, such as aerosol or cloud optical parameters, the native resolution of MODIS can be estimated (D. Wang et al., 2020). Version 6, which provides data from 2000 to the present at a 5 km spatial resolution and 3 h temporal resolution, is available online (D. Wang, 2017). Version 6.1 with a 1 km resolution was used in this study (D. Wang, 2021). This version is available from 2000 till now. BESS is a simplified process-based model which combines RTM with an artificial neural network to generate daily DSR at 0.05 degree from 2000 to the present (Ryu et al., 2018). The training data are simulated from a series of MODIS atmosphere and land products as well as MERRA reanalysis data using RTM Forest Light Environmental Simulator. ANN is trained and replace the RTM to generate shortwave transmittance and then calculate DSR. The generated data has been tested over BSRN and FLUXNET stations (Ryu et al., 2018).

2.2.2 Ground measurements

Data were collected from 142 stations and 6 networks, with full global coverage, as shown in Figure 2-1. The ancillary information of the networks is summarized in Table 2-2. The Surface Radiation Budget Network (SURFRAD) data from 2000 to 2019 were downloaded for long-term analysis. For the rest of the networks, the data from 2004 were used for the daily satellite product analysis.

The FLUXNET (Pastorello et al., 2020) provides global data that were collected at sites of multiple regional flux networks over a variety of surface types. All the stations are operated and maintained by their science teams. The temporally aggregated data after quality control are submitted to the Regional Networks and the FLUXNET Data Portal. All data were processed based on a uniform and detailed procedure developed by the European and AmeriFlux teams including data format standardization, data quality checks, gap-filling, and flux partitioning (Pastorello et al., 2020). The daily aggregated data are directly used in this study.

The SURFRAD provides data from 1993 collected at 7 sites in the United States. It has adopted measurement standards set by the BSRN, which is sponsored by the World Climate Research Program of the World Meteorological Organization. Accuracies of 2%–5% have been achieved (Augustine et al., 2000).

The Greenland Climate Network (GCNET) consists of 24 automatic weather stations in Greenland. It provides hourly radiation observations from 2000 till now. The possible error sources have been acknowledged and the data have been calibrated by the GC-NET team. The instrument accuracy is about 5%-15% (Steffen et al., 1996). To ensure the accuracy of these high latitude measurements, further quality control procedure is conducted which will be discussed in Section 3. The study also adds seven sites in the Antarctic area from the McMurdo Dry Valleys Long-term Ecological Research (LTER) project (McMurdo Dry Valleys LTER et al., 2019a, 2019b, 2019c, 2019d, 2019e, 2019f, 2019g). Both 15 min (level 1) and daily summaries (level 2) data are available, and those data have been post-processed by the LTER. Daily data from LTER are directly used in this study.

To incorporate more stations over high latitude areas, the data from the Finnish Meteorological Institute (FMI) (FMI, 2016), and the Norwegian Institute of Bioeconomy Research for emergency services and agricultural research (NIBIO) (NIBIO-LMT, 2016) are also included. As no additional documents have been found for these five datasets about their data quality, critical quality control procedures have been applied to these data which will also be discussed in Section 2.3.

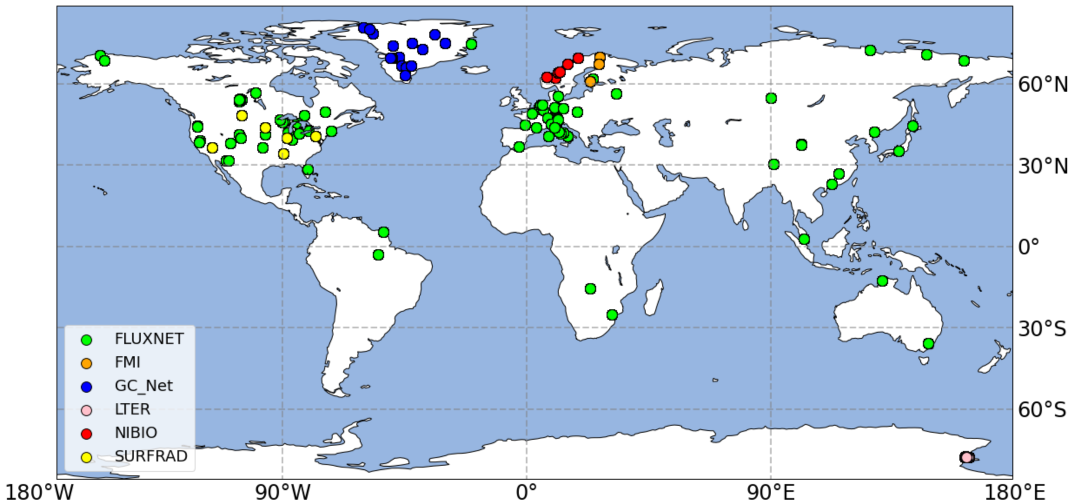

Figure 2-1. Locations of the 142 ground measurement stations.

2.2.3 Other Data

To examine the influence of snow cover, the Interactive Multisensor Snow and Ice Mapping System (IMS) Daily Northern Hemisphere Snow and Ice Analysis at 4 km resolution data (U.S. National Ice Center, 2008) are employed in this study. The land surface is classified into five different surface types, which are the sea, land without snow, sea ice, and snow-covered land. The data is generated by the U.S. National Ice Center and covers the northern hemisphere on a daily basis from 2004 (U.S. National Ice Center, 2008).

Table 2-2. Description of the 6 monitoring networks.

Network	Number	Region	Temporal resolution	Temporal coverage
NIBIO	6	Norway	1 h	2008–2018
FMI	3	Finland	1 min	1998–2019
GC-Net	17	Greenland	1 h	1995–2019
FLUXNET	103	Global	30 min, daily	1991–2014
LTER	7	Antarctic	15 min, daily	1993–2018
SURFRAD	6	United States	1 min	1995–2020

2.3 Methods

2.3.1 Quality control of the field measurements

Quality control procedures were necessary for ground measurements especially over high-latitude areas because severe conditions are common in these areas, which can lead to systematic or operational errors. Stations located over high-latitude areas generally have low solar zenith angles. The area affected by the shadow increases with the solar zenith angle and thus the measurement error increases (Muneer et al., 2007). In addition, snow cover and melt as well as freezing caused by the low temperature will lead to spurious readings during the DSR measurements.

The quality control is firstly conducted on hourly or sub-hour measurements, which are from SURFRAD, GC-Net, NIBIO, and FMI networks. The procedures include the setting of irradiance values which initially have a negative solar elevation angle as 0. Subsequently, all data pass the two limit tests of BSRN quality check (Roesch et al., 2011), which are described below:

$$Physically\ possible\ limits: -4W/m^2 < R_s < S_0 * 1.5 * \mu^{1.2} + 100\ W/m^2 \quad (1)$$

$$Extremely\ rare\ limits: -2W/m^2 < R_s < S_0 * 1.2 * \mu^{1.2} + 50W/m^2 \quad (2)$$

where $R_s$is the DSR, S_0 is the solar constant adjusted for the Earth–Sun distance, and μ is the cosine of the solar zenith angle. Because of the unavailability of other parameters, the final procedure of the BSRN test, that is, the "across quantities" procedure, is not implemented (Roesch et al., 2011). After temporal data aggregation, the samples are also compared with satellite measurements. The stations which has more than 50 samples with standardized residuals above three for all satellite products are eliminated.

2.3.2 Data aggregation

Temporal data aggregation is a necessary preprocessing step, which is used to handle inconsistent temporal resolutions between in situ data and satellite products. The hourly data is firstly aggregated to daily. In previous research, generally a simple method was used, such as averaging hourly data, if a specific number of slots were available (Urraca et al., 2017). However, such a procedure will lead to the underestimation of daily values in cases with missing daytime hourly measurements (Urraca et al., 2017) and may cause large data loss during, which is a critical problem over high-latitude areas. Gap-filling is a possible solution, but arbitrary filling may also cause bias. A new method was developed in this study considering the tradeoff between the data availability and data accuracy. The hourly data in one day were divided into eight slots by averaging three-hourly data, but only slots with more than or equal to two hours of data were used. Subsequently, the daily average including all eight available slots was calculated. By applying this method, a day with random data loss can be included without compromising the data accuracy.

For the long-term trend analysis in Section 4, all daily data were aggregated into months by calculating the mean following the methodology reported in X. Zhang et al. (2015). The monthly value was calculated if daily data were missing for less than nine days. The equation is as follows:

$$R_m = \frac{\left(\sum_{i=1}^{n} R_d^i\right)}{n} * N \tag{3}$$

where R_m and R_d are the monthly and daily average DSR, respectively; n is the number of R_d available in one month; and N is the total number of days in the month.

2.3.3 Assessment metrics

The daily and monthly satellite data were validated against the ground measurement data. Several statistical indicators are available for the evaluation of the performance of solar radiation products:

R squared (R^2):

$$R^2 = \frac{\sigma_{R_p R_g}}{\sigma R_p \sigma R_g} \qquad (4)$$

$R^2 = \frac{\sigma_{R_p R_g}}{\sigma R_p \sigma R_g}$Mean Bias Difference (MBD), relative Mean Bias Difference (rMBD)

$$MBD = \frac{1}{N}\sum_{i=1}^{N}(R_s^i - R_g^i) \qquad (5)$$

$$rMBD = MBD/\overline{R_g} \qquad (6)$$

Root-Mean-Square Error (RMSE), Relative Root-Mean-Square Error (rRMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_s^i - R_g^i)^2} \qquad (7)$$

$$rRMSE = RMSE/\overline{R_g} \qquad (8)$$

where R_s represents the values predicted from satellites, R_g represents the ground measurement values, σR_p represents the standard deviation of the predicted values for the testing samples, σR_g represents the standard deviation of the ground measurements for the testing samples, $\sigma_{R_p R_g}$ represents covariance of the predicted values and ground measurements for the testing samples, and N is the total number of observations. The rMBD and rRMSE are the main statistics; they facilitate better comparisons among different regions and seasons among these five products.

2.4 Results

2.4.1 Data availability

Different products present different levels of data loss, which strongly depends on the latitude and seasonality. Figure 2-2 depicts the spatial distribution of the data availability, which is defined as the number of available data divided by the total number of ground measurement data at each station. The CERES and BESS present a continuous and full coverage. The MCD18 fully covers the continents as well. The only data loss is at the edge of the continents where were classified as ocean . The availabilities of CLARA and GLASS generally reach 100% at mid- and low latitudes, but data loss occurs over high-latitude areas. To examine the reasons for this data loss, the data availability at different latitudes among different months for the five products is plotted in Figure 2-3. The GLASS data availability strongly correlates with both the latitude and seasonality. The higher the latitude is, the longer the data loss exists in winter in the Northern Hemisphere. Polar night period are colored in white, and it shows that the data loss periods are longer than the polar night period, leading to a low data availability or the lack of data in winter between 50 and 70° in the Northern Hemisphere. CLARA data has low availability at higher latitudes as well, but the data availability does not strongly correlate with the latitude.

Furthermore, CLARA estimations over snow-covered areas have been excluded due to the reduced data accuracy (Karlsson et al., 2017).

Considering the various data availabilities of the five products, especially over high-latitude areas, only the estimated DSR, which is available for all five products, was used in the remainder of this study.

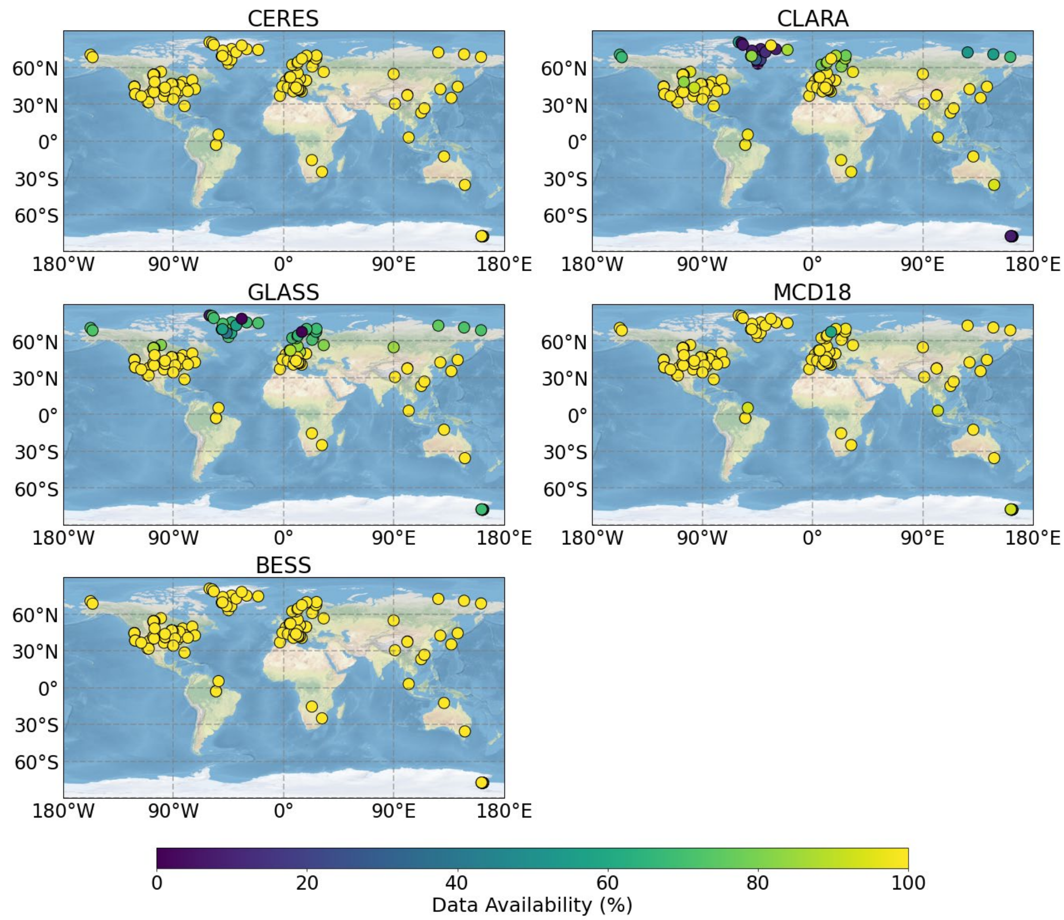

Figure 2-2. Spatial distribution of the data availability on a daily time scale for all five products.

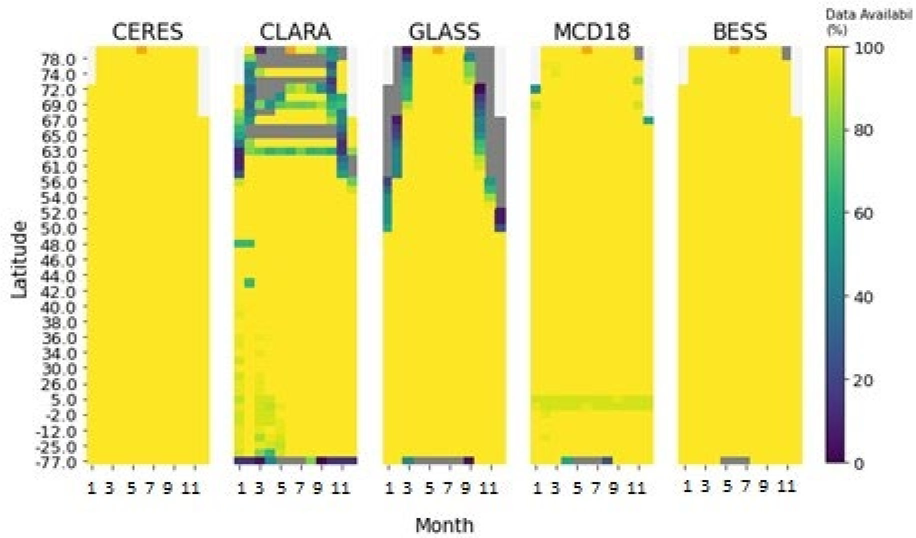

Figure 2-3. Monthly distribution of the availability of five products. Grey colors represent areas for which no satellite data are available. Orange colors represent areas for which no insitu data are available. The white colors represent polar night period.

2.4.2 Overall accuracies

The validations were conducted at the finest resolutions of each product, that is, at a resolution of 1, 5, 25, and 100 km, and the products with higher resolutions were aggregated for the intercomparison. It is noticeable that we use the single point ground measurement for the validation in this study. The results are shown in Table 2-3. Overall, the RMSE of all products decreases when they are aggregated to a coarser resolution, possibly due to the ignorance of the 3D cloud effects which is pronounced at fine scale (Chen et al., 2019). The MCD18 is the only product that has a spatial resolution ranging from 1 to 100 km. The difference between the products with a 1 and 100 km resolution is ~5% in terms of the rRMSE. The RMSE decreases more rapidly when the products are aggregated from 1 to 5 km, but the rate of the decrease subsequently declines. Spatial aggregation process exert considerable influence on BESS data accuracy. At its original resolution, 5km, the rRMSE of BESS is higher than both MCD18 and GLASS, but it is lower than MCD18 and

GLASS after aggregating to 100km. Note that the accuracies of CLARA, GLASS, and MCD18 are higher at 25 km than at 100 km over high-latitude areas.

A more consistent comparison of the five products was conducted based on a spatial resolution of 1°. The detailed statistics are presented in Table 2-4. The performance of the five products after aggregation is similar. The overall RMSE falls in the normal range of the daily validation, that is <35 W/m^2 (Huang et al. 2019; D. Wang et al. 2020), but the rRMSE is slightly higher than that reported in previous studies (Urraca et al., 2017; D.Wang et al., 2020). This is mostly due to the fact that more stations were located over high-latitude areas. The CERES presents the lowest RMSE of 27.6 W/m^2 and highest R^2 of 0.916. The high-resolution products have a slightly higher RMSE, which is 29.1, 30.3, 29.6, and 31.6 W/m^2 for CLARA, GLASS, BESS, and MCD18, respectively. One possible reason for this difference might be that these four DSR products with high spatial resolution are mainly derived from polar-orbiting satellite data with limited daily observations, which is a major error source in cases with dynamically atmospheric conditions. In addition, three-dimensional cloud radiative effects will lead to more uncertainties in high-resolution products combined with the limited daily observations (Chen et al., 2019). The CERES tends to overestimate the DSR, whereas the other four products underestimate the DSR. The CLARA is the least biased among all five products.

The reliability of the results also strongly depends on the accuracies of the ground measurements. In Figure 2-4, the retrievals of different products are plotted against the measurements of different networks. FLUXNET stations have a full global coverage and provides most measurements, and consequently validation results over

FLUXNET stations dominate the overall results in Table 2-4. The SURFRAD only have limited stations in the United States but the accuracy of this network is the highest because it provides 1 minute measurements and have processed the highest level of calibrations set by the BSRN. Based on the comparison of these two networks, CERES achieves the most consistent results with a rRMSE of 16.8%. The GLASS performs the worst over SURFRAD stations, whereas the MCD18 presents the lowest accuracy over FLUXNET stations. The NIBIO and FMI are two networks located at high latitudes, which are monitored by European countries. The accuracies of these two networks have been verified (Urraca et al., 2017). With respect to these two networks, CLARA achieves the highest accuracies with a rRMSE of 17.8% and 18.4%, respectively. BESS has the highest rRMSE which are 22.3% and 22.4% respectively over these two networks. The GC-Net and LTER stations are in Arctic and Antarctic regions, respectively. The rRMSE over these two networks significantly increases for all five products and MCD18 is significantly biased compared with the other four products. This is partly due to the limitations of existing algorithms over high-latitude areas. However, the accuracies of the ground measurements must be also considered. The performances of these five products slightly vary over different networks, but the overall results are comparable.

Table 2-3. Accuracies of each product at different spatial resolutions.

		100 km					**25 km**					**5 km**					**1 km**				
Latitude	**Product**	R^2	**MBD W/m^2**	**rMBD (%)**	**RMS (W/m^2**	**rRMSE (%)**	R^2	**MBD (W/m^2**	**rMBD (%)**	**RMS (W/m^2**	**rRMSE (%)**	R^2	**MBD (W/m^2**	**rMBD (%)**	**RMS (W/m^2**	**rRMSE (%)**	R^2	**MBD (W/m^2**	**rMBD (%)**	**RMS (W/m^2**	**rRMSE (%)**
0–30	CERES	0.86	11.8	6.6	30.1	16.8															
	CLARA	0.85	21.5	12	36.1	20.2	0.83	23.7	13.3	41	23										
	GLASS	0.81	10.2	5.7	33.6	18.8	0.81	13.6	7.6	35.5	19.9	0.79	13.3	7.5	37.3	20.9					
	BESS	0.73	-1.1	-0.6	38.9	21.8	0.8	3	1.7	33.7	18.9	0.76	3.9	2.2	38.7	21.7					
	MCD18	0.82	0	0	30.8	17.2	0.83	3.7	2.1	31.6	17.7	0.81	2.5	1.4	34.6	19.4	0.78	2.1	1.2	37.6	21.1
30–60	CERES	0.93	0.2	0.1	25.3	15.8															
	CLARA	0.93	-3.6	-2.2	26.1	16.2	0.93	-4	-2.5	27.5	17.1										
	GLASS	0.91	-4.8	-3	28.7	17.9	0.9	-5	-3.1	30.6	18.9	0.9	-4.9	-3	31.8	19.7					
	BESS	0.92	-3.9	-2.5	27.3	17	0.91	-3.7	-2.3	30.7	19	0.89	-3.6	-2.2	34.5	21.4					
	MCD18	0.92	-8.5	-5.3	29.6	18.4	0.91	-7.7	-4.7	31.6	19.6	0.9	-7.8	-4.8	35.2	21.8	0.88	-7.9	-4.9	38.3	23.7
60–90	CERES	0.86	6.7	4.6	34.3	23.8															
	CLARA	0.85	-5.4	-3.7	36.2	25.1	0.85	-3.2	-2.2	35.6	24.8										
	GLASS	0.88	2.6	1.8	31.5	21.8	0.88	3.7	2.6	30.3	21.2	0.88	3.6	2.5	30.5	21.2					
	BESS	0.87	-11	-7.6	33.7	23.4	0.84	-9.9	-6.9	36.5	25.5	0.81	-9.7	-6.8	40.6	28.3					
	MCD18	0.84	-14.2	-9.8	38.7	26.8	0.86	-10.2	-7.1	34.4	24	0.85	-10.1	-7	35.8	24.9	0.82	-10.9	-7.6	40.3	28
All	CERES	0.92	2.2	1.3	27.6	17															
	CLARA	0.91	-1.5	-0.9	29.1	17.9	0.9	-1.6	-1	30.4	18.7										
	GLASS	0.9	-2.8	-1.7	30.3	18.6	0.89	-2.9	-1.8	31.9	19.5	0.88	-2.9	-1.8	33.1	20.3					
	BESS	0.91	-3.4	-2.1	29.6	18.2	0.89	-3.1	-1.9	32.3	19.8	0.87	-3.1	-1.9	36.4	22.3					
	MCD18	0.9	-8	-4.9	31.6	19.4	0.9	-6.8	-4.2	32.6	20	0.88	-7.1	-4.4	36	22.1	0.86	-7.2	-4.4	39.2	24

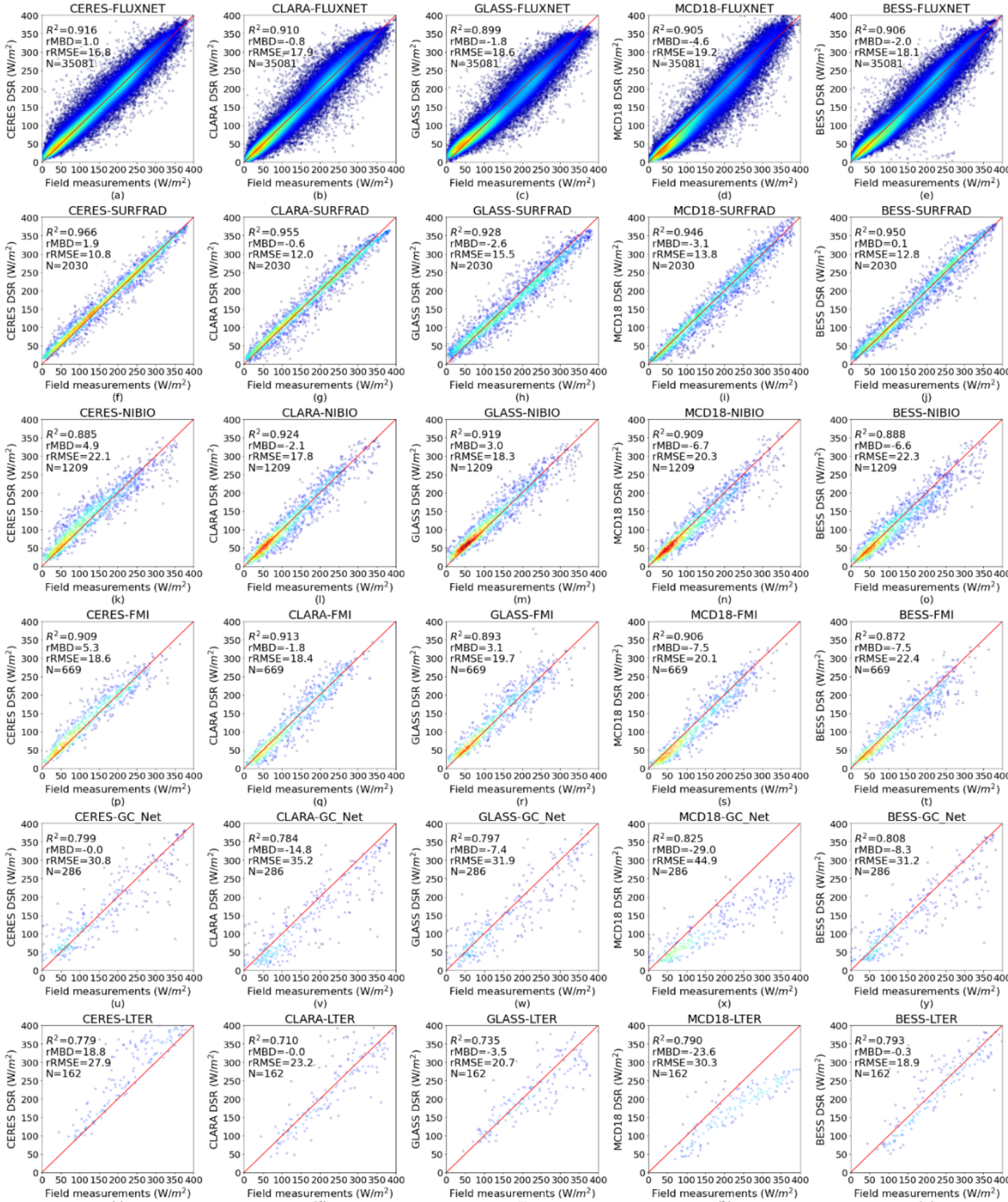

Figure 2-4. Retrievals of the five products over different networks.

Table 2-4. Summary of the statistics of each product for different subareas.

Latitude (°)	Product	R^2	MBD (W/m^2)	rMBD (%)	RMSE (W/m^2)	rRMSE (%)	MAD (W/m^2)	rMAD (%)
All	CERES	0.916	2.2	1.3	27.6	17.0	19.6	12.0
	CLARA	0.911	-1.5	-0.9	29.1	17.9	20.3	12.5
	GLASS	0.899	-2.8	-1.7	30.3	18.6	21.7	13.3
	MCD18	0.904	-8.0	-4.9	31.6	19.4	22.6	13.9
	BESS	0.907	-3.4	-2.1	29.6	18.2	21.3	13.1
60–90	CERES	0.860	6.7	4.6	34.3	23.8	24.1	16.7
	CLARA	0.846	-5.4	-3.7	36.2	25.1	23.5	16.2
	GLASS	0.878	2.6	1.8	31.5	21.8	22.0	15.3
	MCD18	0.840	-14.2	-9.8	38.7	26.8	26.1	18.1
	BESS	0.875	-11.0	-7.6	33.7	23.3	24.1	16.7
30–60	CERES	0.932	0.2	0.1	25.3	15.8	18.3	11.4
	CLARA	0.931	-3.6	-2.2	26.1	16.2	18.6	11.6
	GLASS	0.915	-4.8	-3.0	28.7	17.8	20.7	12.9
	MCD18	0.922	-8.5	-5.3	29.6	18.4	21.4	13.3
	BESS	0.924	-3.9	-2.4	27.3	17.0	20.1	12.5
0–30	CERES	0.856	11.8	6.6	30.1	16.8	22.8	12.7
	CLARA	0.852	21.5	12.0	36.1	20.2	28.9	16.1
	GLASS	0.807	10.2	5.7	33.6	18.8	26.1	14.6
	MCD18	0.824	0.0	0.0	30.8	17.2	23.8	13.3
	BESS	0.726	-1.14	-0.6	38.9	21.6	24.7	13.8

2.4.3 Spatial distribution

Figure 2-5 depicts the spatial distributions of the rMBD and rRMSE of the five products. The figure shows that all five products present latitudinal variations, with the lowest deviations and RMSEs in mid-latitude regions. The statistics for different latitude zones are listed in Table 2-2. Over low-latitude areas, CERES and MCD18 have the highest accuracies with a RMSE of 30.1 W/m^2 (16.8%) and 30.8 W/m^2 (17.2%), respectively. Almost no bias was observed for the MCD18 product. In contrast, BESS yields the worst accuracy over low-latitude areas, with the largest RMSE of 38.9 W/m^2 (20.2%) and MBD of -1.14 (-0.6%), followed by CLARA with the largest MBD of 21.5 W/m^2 (12%) and RMSE of 36.1 W/m^2 (18.8%). Both the

GLASS and CERES tend to overestimate the DSR over low-latitude areas and have a bias of 11.8 W/m^2 (6.6%) and 10.2 W/m^2 (5.7%), respectively. The accuracies of the CERES data are outstanding, with the lowest RMSE for each subregion. The estimation accuracy is superior over mid-latitude areas. As shown in Table 2-2, the MBD over mid-latitude areas is close to 0.2 W/m^2 (0.1%) and the RMSE is the lowest among all products. In addition, CLARA and BESS also yield low RMSE over mid-latitude areas, which are, 26.1 W/m^2 (16.2%) and 27.3 (17%) respectively. Over high-latitude areas, all product degrade with significant increase of rRMSE. GLASS shows the best performance, with a bias of 2.6 W/m^2 (1.8%) and the lowest RMSE of 31.5 W/m^2 (21.8%). The MCD18 significantly underestimates the DSR, with a MBD of -14.2 W/m^2 (-9.8%). However, because only data that are available for all products are kept over high-latitude areas, several degraded results of one product may be excluded the data loss of other products. In summary, the latitude influences the products significantly, especially for CLARA, BESS, and MCD18. The magnitude of underestimation of BESS and MCD18 amplified from tropical areas to high latitude areas, while CLARA overestimates in low latitude area but tend to underestimate as latitude increases.

The products also have several other similar spatial distribution characteristics. First, significant deviations are mostly observed near the coastal region. Most of the overestimations of CERES and CLARA were observed along the coastline. These results agree well with those of previous studies (Urraca et al., 2017; X. Zhang et al., 2015). Ocean fraction data from the CERES-SYN product were employed to further examine the effect of the presence of ocean in the pixel. An ocean fraction above and below 70% was defined as coastal and inland pixel, respectively. Figure 2-6 depicts the rRMSEs and rMBDs of the five satellite products for inland and coastal areas. Significant overestimation can be observed for CLARA and CERES over coastal regions compared with inland regions. The GLASS, BESS, and MCD18 tend to underestimate the DSR in the coastal region, but the magnitude of the underestimation insignificantly differs depending on the location. BESS is most influenced by the ocean, with rMBD increase from -2.0% to -4.5% and rRMSE increase from 18.3% to 21.3%. Although the rRMSEs of the other four products are less influenced by the presence of the ocean, a difference can be noted over the inland area, where CERES and CLARA perform better than GLASS and MCD18. Opposite results were obtained over the coastal area. The rRMSEs of CERES, CLARA, GLASS, and MCD18 over coastal areas are 17.6%, 16.7%, 15.8%, and 15.7%, respectively, which demonstrates that the estimation accuracy of the DSR in these regions is higher at higher product resolution. These results may indicate that the identification of land or ocean along the coast improves based on the use of high-resolution satellite products, as it can mitigate the surface albedo effect from the ocean.

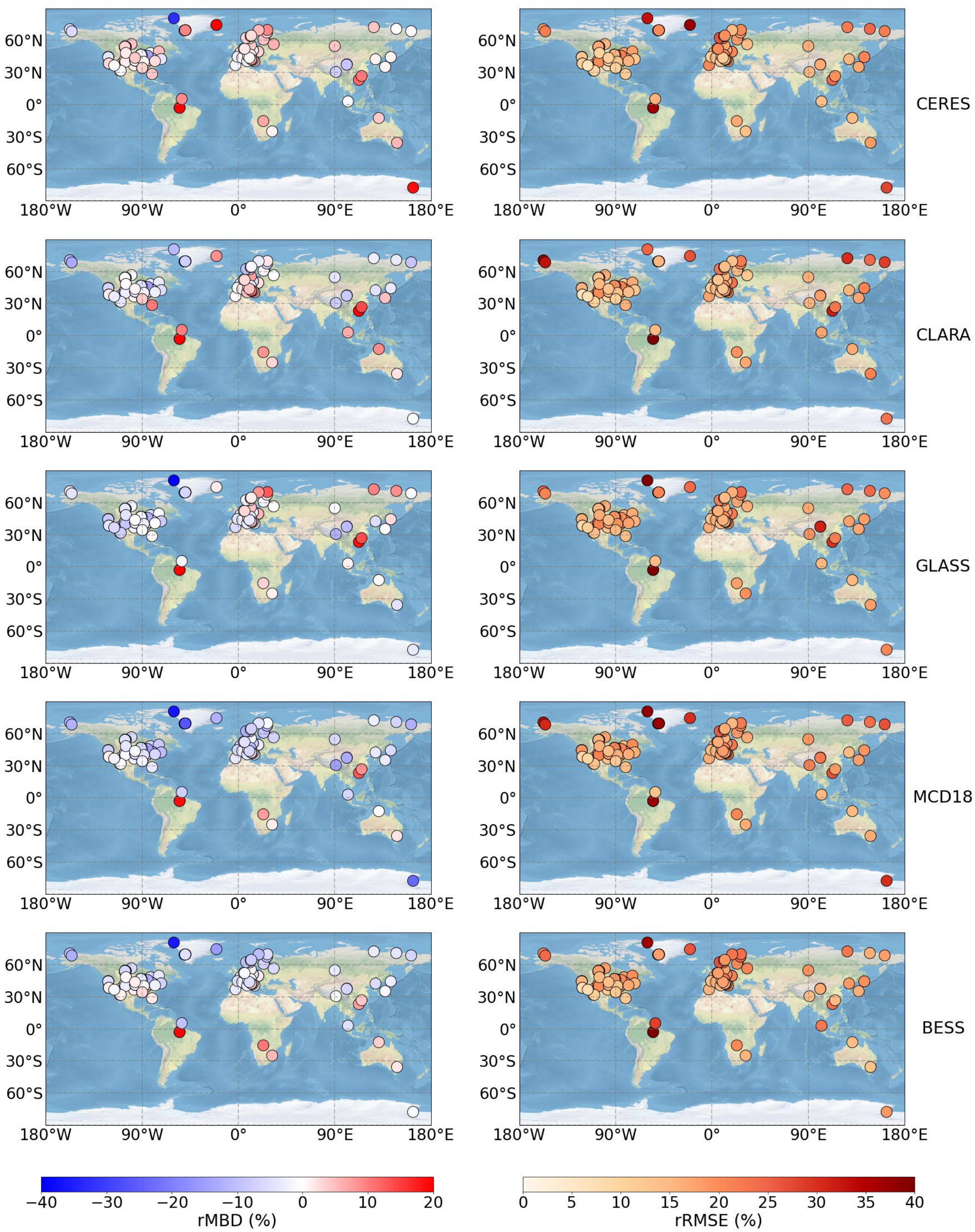

Figure 2-5. Spatial distributions of the rMBDs and rRMSEs of the satellite-estimated DSR and ground measurements for CERES, CLARA, GLASS, MCD18, and BESS on a daily time scale.

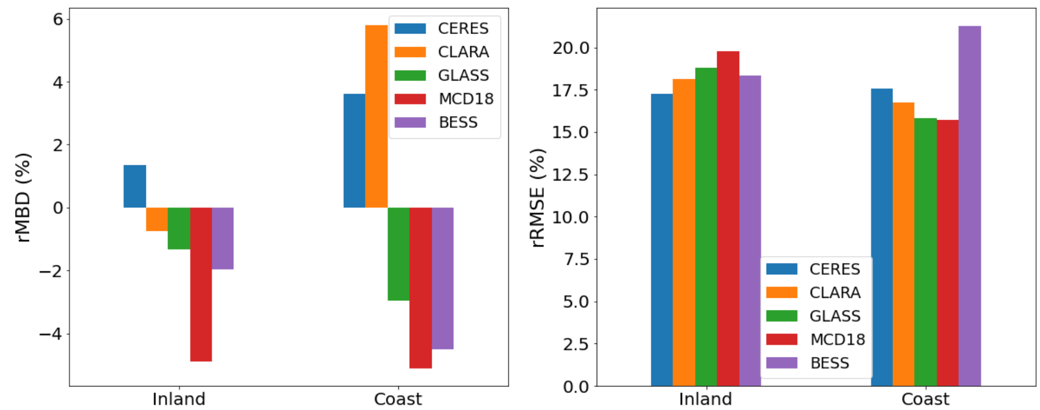

Figure 2-6. Comparison of the rMBD and rRMSE over inland and coastal areas.

2.4.4 Intraannual analysis

Figure 2-7 shows the heatmap of the intraannual variation of the monthly rMBD and rRMSE for different latitude intervals in the Northern Hemisphere. Grids without data were plotted in grey for all five products. The rMBDs of all products over areas with latitudes ranging from 0 to 30° do not indicate notable intraannual trends. For the area between 30°–55°, the products tend to underestimate the DSR during the winter, but the results are comparatively accurate from April to September, which agrees with previous results obtained for CLARA (Urraca et al., 2017). This trend is especially pronounced for MCD18. The CERES is less dependent on the seasonal change at mid-latitudes. When the latitude reaches 55°–80°, a seasonal change of the rMBD can be observed for all five products. The CERES, CLARA, and GLASS slightly overestimate the DSR from June to August, whereas the BESS yields comparatively accurate results. In winter, CLARA, GLASS, BESS, and MCD18 tend to underestimate the DSR over high-latitude areas and the MCD18 is the most biased among the products.

With respect to the rRMSE, all products present a similar pattern. The rRMSE is low from June to August and subsequently increases in fall and winter. The higher the latitude is, the longer the high rRMSE lasts. The lowest rRMSEs were observed over mid-latitude areas, which agrees with our previous conclusions. The rRMSE slightly increases in low-latitude areas, but a seasonal trend, similar to that in the high-latitude region, is not notable.

Based on the combination of the results from all plots, we can conclude that the season affects the DSR estimation. The accuracies of all five products decrease in winter and the seasonal effect slightly varies with the latitude. The higher the latitude is, the more notable is the annual trend in terms of the rMCD and rRMSE. All five products yield comparable results during the summer over mid-latitude areas. However, the absolute rMBD increases in winter, which leads to a high rRMSE. The seasonal dependence of each product varies. The MCD18 has responds the earliest to seasonal changes, followed by GLASS, CLARA, and BESS. The seasonal and spatial variations may indicate that snow is one of the major factors contributing to large biases and inaccuracies in satellite DSR estimations, especially for MCD18, because the similar optical and reflection characteristics between snow and clouds lead to difficulties in the identification of snow-covered periods and thus to underestimations over snow-covered surfaces (Huang et al., 2019).

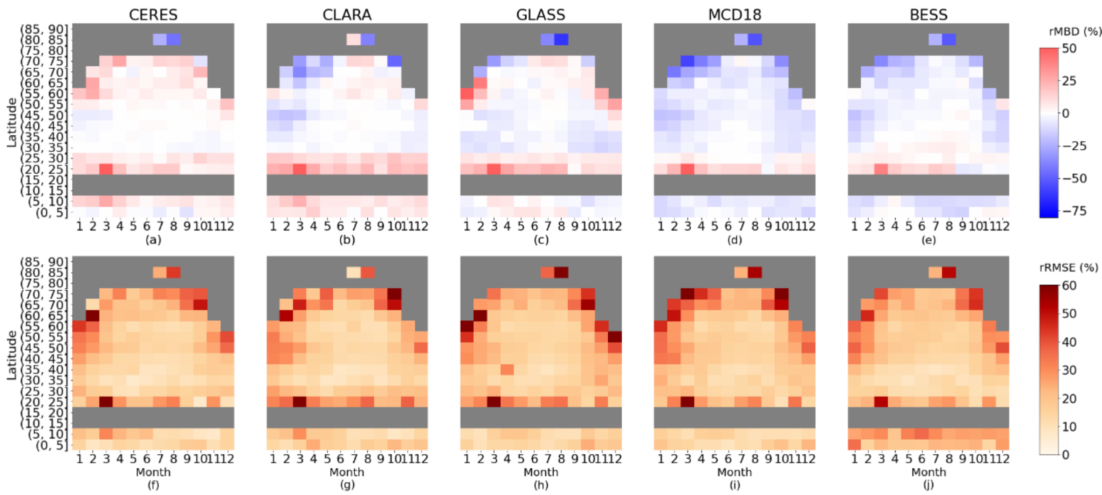

Figure 2-7. The rMBDs and rRMSEs of daily values over the Northern Hemisphere were calculated for each month and each latitude interval. The rMBD is shown in (a), (b), (c), (d), and (e). The rRMSE is shown in (f), (g), (h), (i), and (j).

To better support this argument, more analyses of snow were conducted. The IMS Daily Northern Hemisphere Snow and Ice Analysis at 4 km resolution data (U.S. National Ice Center, 2008) were employed. Figure 2-8 presents the rMBD and rRMSE over snow-covered and snow-free areas. Because snowy regions are generally located over high-latitude areas in which other sources may also lead to uncertainties, we only selected SURFRAD stations for the comparison. The results show that snow indeed leads to uncertainties in the DSR estimation. All five products yield higher RMSEs over snow-covered surfaces, but the magnitudes vary depending on the product. The MCD18 is most influenced by the snow, with a rRMSE of 13.1% over land and a rRMSE of 40.3% over snow-covered regions. The MCD18 also significantly underestimates the DSR over snow-covered areas. The GLASS is also substantially affected by snow, with a rRMSE of 14.9% over land and a rRMSE of 38.6% over snow-covered regions. However, overestimations were observed over snow-covered regions. Although CLARA already excludes strongly degraded estimates over snow-covered surfaces, the current rRMSE still reaches 36.0%. BESS is less influenced by the snow in terms of rRMSE, but BESS indeed significantly underestimate DSR when snow present with rMBD changing from positive 0.4% to negative 8.1%. The CERES performs the best over snow-covered regions, with a rRMSE of 24.3%.

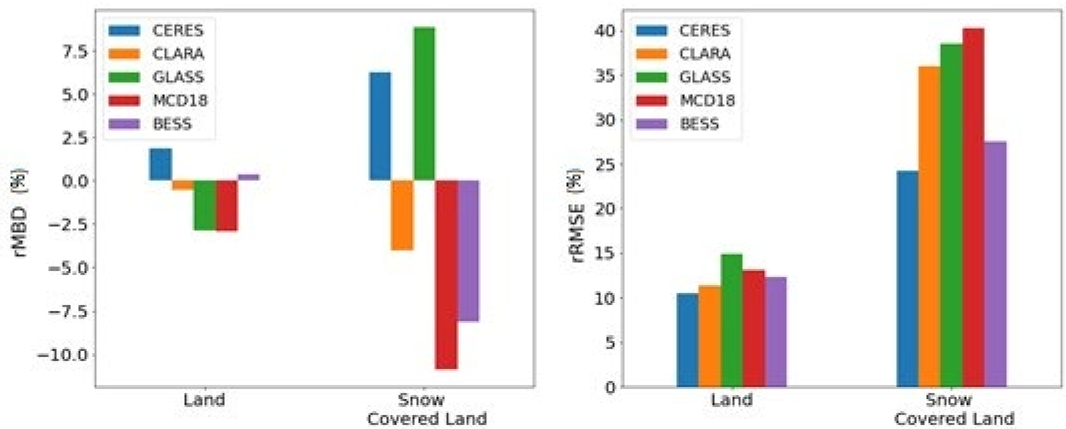

Figure 2-8. Comparison of the rMBD and rRMSE values obtained over snow-covered and snow-free areas.

2.4.5 Interannual analysis

The results of previous studies indicated a pronounced brightening trend during the 1990s, but the trend diminishes or even reverses starting in 2000 (Hatzianastassiou et al., 2005; Hinkelman et al., 2009; Wild et al., 2009; X. Zhang et al., 2015). However, the latest global trend analysis based on satellite products was conducted by using data obtained before 2012 (X. Zhang et al., 2015). The ~20-year long data records of four satellite products and associated long-term measurements archived by three SURFRAD stations allow us to revisit this matter. Because limited MCD18 data have been released, only the long-term trend of CERES, GLASS, CLARA, and BESS were analyzed in this study. Figure 2-9 depicts the annual average DSR estimate and values measured from 2001 to 2018. All four satellite products are consistent with the ground measurements. The CERES, BESS, and CLARA can explain most of the interannual variabilities, with correlation coefficients of 0.95, 0.91, and 0.90, respectively. This further demonstrates that satellites can be used to monitor the interannual variability. In our study, in situ data show insignificant brightening or dimming trends over the past two decades, with an annual slope of 0.004 W/m^2. The CERES and BESS data agree well with the in situ results, with an annual slope of 0.049 $W/m^2$and 0.034 W/m^2. CLARA demonstrates a slight dimming trend (-0.193 W/m^2per year) and GLASS data present a significant dimming trend (-0.383 W/m^2per year). Figure 2-9 also proves that CERES tends to overestimate the DSR, whereas CLARA, BESS, and GLASS have underestimated the DSR during the past two decades.

To examine the stabilities of the current satellite DSR products and determine if they can detect long-term trends, which is especially useful for climate change analysis, the bias and RMSE were monitored over the long term in this study using SURFRAD data. The interannual evolution of the annual aggregated bias and the RMSE from 2001 to 2009 are plotted for four products in Figure 2-10. BESS and CERES show stable consistencies with the ground measurement values because both the bias and RMSE have not changed during the past two decades. However, CLARA and GLASS data exhibit abnormal changes. Both the deviation and RMSE of these two products increase over the years. The slope of the increase in the annual bias reaches -0.250 and -0.387 $W/m^2$for CLARA and GLASS, respectively, which contributes to a change of ~7.8 $W/m^2$for GLASS after two decades and a change of 3.75 $W/m^2$for

CLARA after 15 years. The annual RMSEs of CLARA and GLASS also significantly change, with slopes of 0.357 and 0.310 W/m^2, respectively. The changes may be related to sensor degradation. In previous publications, it has been pointed out that Terra MODIS presents a significant degradation, especially at shorter wavelengths (D. Wang, et al., 2012). The stability of BESS and CERES also indicates that using higher level of MODIS products and the incorporation of various inputs from different data sources may help avoid the influence of sensor degradation. The long term drifting trend disagrees with the results reported by Urraca et al. (2017) who reported the temporal stability for CLARA data from 2005 to 2015. This might be due to several reasons. First, our study includes a much longer time scale, which may lead to a more notable trend. Furthermore, Urraca et al.'s (2017) research focused on the European area, with latitudes ranging from 35° to 65°, whereas our study includes data for the United States, with latitudes ranging from 30° to 50°. As mentioned before, the latitude affects the accuracy of the satellite products, which may lead to different long-term trends. This also explains why both CLARA and GLASS have exhibited a dimming trend over the past two decades in contrast to the ground measurement results shown in Figure 2-9.

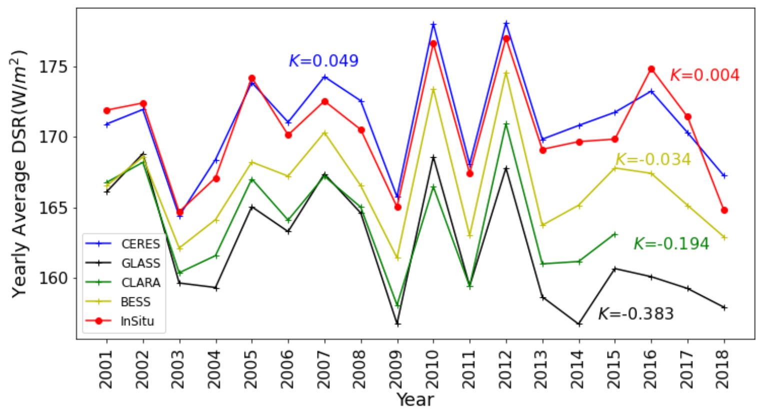

Figure 2-9. Annual average DSR from 2001 to 2019 over the SURFRAD sites. The slope (K) is labeled for each product.

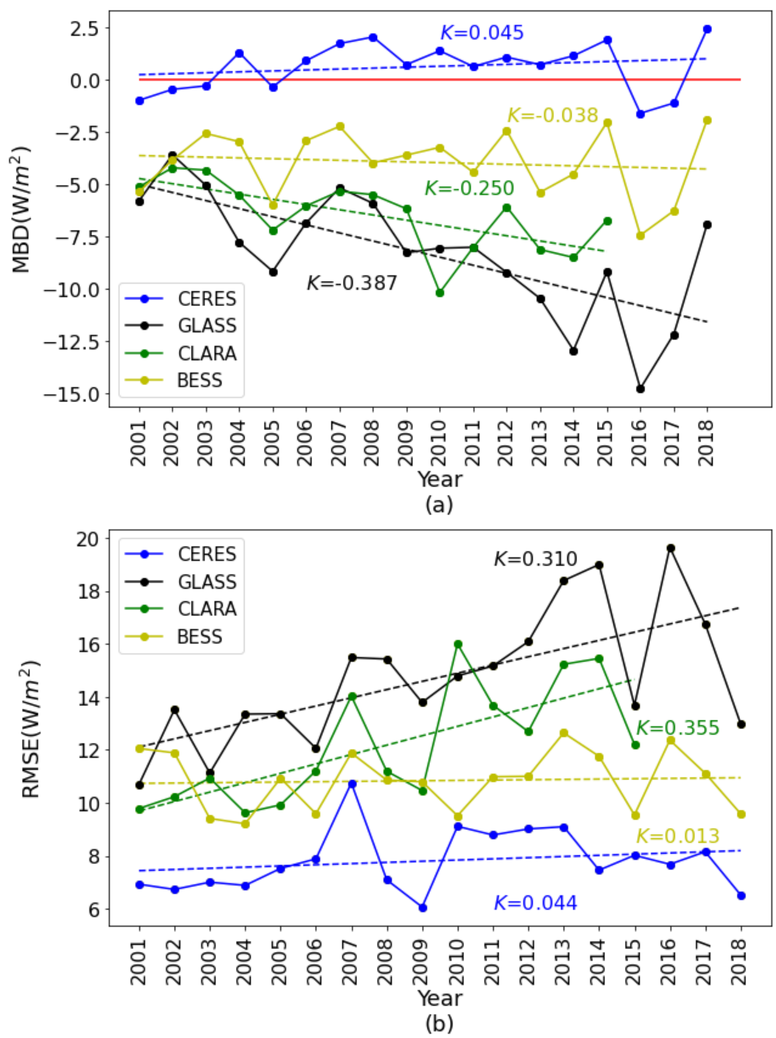

Figure 2-10. (a) Annual bias (MBD) from 2001–2019, (b) Annual RMSE from 2001–2019. The slope (K) is labeled for each product.

2.4.6 Cloud analysis

Cloud coverage and cloud visible optical depth also directly affect the estimation of DSR. In this study, cloud coverage and cloud visible optical depth data from the CERES-SYN dataset were used to examine the cloud effect. As shown in Figure 2-11, the rRMSE linearly increases with increasing cloud coverage for all five satellite products. The coefficients of the correlations between the rRMSE and cloud coverage are 0.94, 0.95, 0.94, 0.97, and 0.93 for CERES, CLARA, GLASS, MCD18, and BESS respectively. Similar strong correlations were observed between the rRMSE and cloud optical depth. In addition, an extreme high cloud fraction and cloud optical depth also lead to significant biases in all five products. The MCD18 tends to underestimate the DSR in areas that are fully covered by clouds or in which the cloud optical depth is high. The CERES, CLARA, BESS, and GLASS tend to overestimate the DSR when the cloud optical depth is high, but correlations with the cloud area fraction were not observed. Comparing the five products, the CLARA is less biased when the cloud area fraction is high and MCD18 is less biased when the cloud visible optical depth is high.

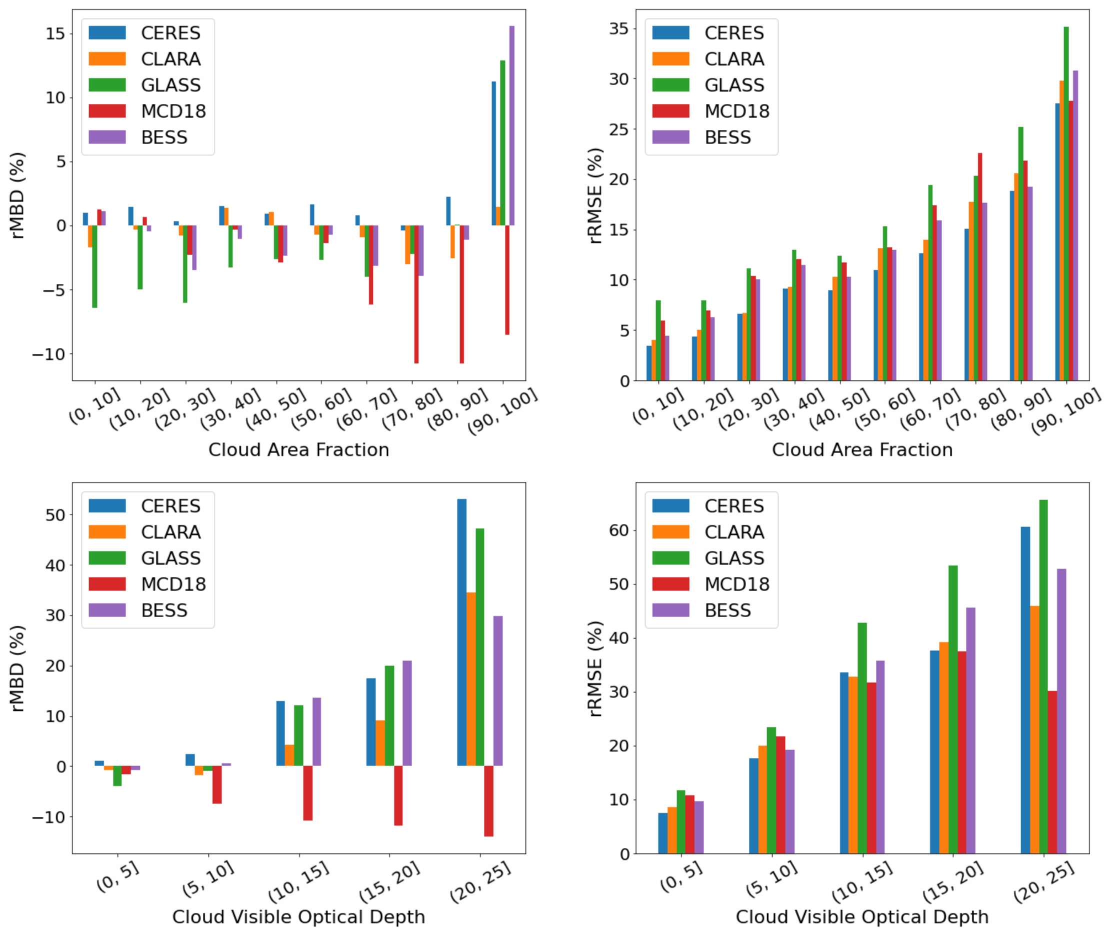

Figure 2-11. Correlations between the rMBD and rRMSE and the cloud area fraction and cloud visible optical depth.

2.5 Potential algorithm refinements

The DSR products with high spatial resolution, such as MCD18A1, CLARA-A2, BESS, and GLASS, can resolve the fine spatial variability of solar radiation. However, the current accuracies of these products remain slightly lower than that of the mature CERES product. In response to the issues of current high resolution DSR products, two possible refinements are suggested in this study. The MCD18 is used as an example. As indicated by the above-mentioned analysis, the compromised performance of the retrieval algorithm over snow-covered surfaces will cause the overall degradation of the accuracy. The TOA reflectance of the blue band is used in the original LUT algorithm implemented in the MCD18 data generation to derive atmospheric transmittance and scattering parameters. Although blue band data provide an informative contrast between surface and atmosphere contributions over dark surfaces, the blue band-based approach loses its effectiveness over bright surfaces such as snow-covered pixels. To address this issue related to the reduced sensitivity of the TOA surface of the blue band to the DSR, the shortwave infrared channel should be used for snow-covered surfaces (Figure 2-12). Snow-covered pixels observed by Aqua/MODIS data over seven SURFRAD sites in 2013 were used in the comparison. The validation confirms that the DSR over snow-covered pixels is significantly underestimated based on the use of the blue band method, which aligns with the assessment results of the MCD18A1 product. Based on the comparison, the use of shortwave infrared band will substantially improve the DSR estimation. The RMSE decreases by 82.8 W/m^2 and the bias declines by 54.6 W/m^2.

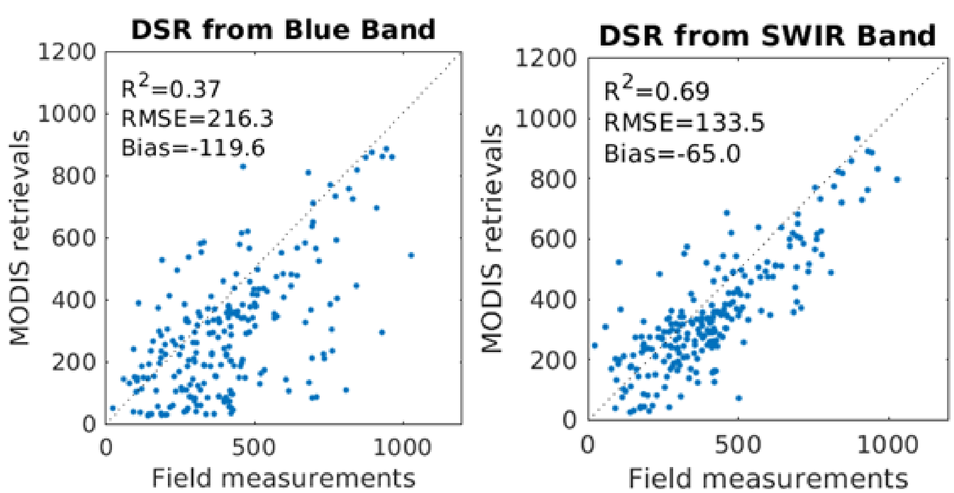

Figure 2-12. Comparison of the DSR retrieval results based on different band selection strategies.

Another possible reason for the performance difference between CERES and other four high spatial resolution DSR products might be that all four DSR products with high spatial resolution evaluated in this study are mainly based on polar-orbiting satellite data, whereas the CERES product employs geostationary data for the diurnal interpolation (Doelling et al., 2013). Although polar-orbiting sensors enable global coverage and an increased sampling frequency over high latitudes, they can only provide limited observations in low- and mid-latitude areas. The limited number of daily observations could be a major source of uncertainties in cases with dynamically changing atmospheric conditions. Previous studies have applied different methods on geostationary sensors like advanced Himawari imager (AHI) and Advanced Baseline Imager (ABI) to estimate DSR and achieved promosing results (Ma et al., 2020; Yi Zhang et al., 2020). In this study, we applied the LUT algorithm to Advanced Baseline Imager (ABI) data acquired by the Geostationary Operational Environmental Satellite (GOES) – 16 and combined with MODIS data to improve the estimation accuracy. .

Figure 2-13 shows an example of a time series of diurnal DSR changes measured at Fort Peck, MT, using the SURFRAD network. In this example, the ABI 15-min retrievals agree very well with the field measurements. The decrease in the DSR in the local morning caused by morning clouds is accurately captured by the ABI data with high-temporal resolution. In contrast, although MODIS data yield accurate instantaneous values corresponding to the satellite overpass time, the daily value suffers from large uncertainties. The feasibility of geostationary satellite data for the estimation of the daily DSR can be demonstrated based on the quantitative validation with the SURFRAD measurements (Figure 2-14). The RMSE of the ABI daily results is low (24.7 W/m^2), that is, 8.6 W/m^2 less than that of the MODIS results, and is comparatable with previous studies on geostationary sensors (Ma et al., 2020; Yi Zhang et al., 2020). Subsequently, the optimal interpolation approach was applied to combine the ABI and MODIS data (D. Wang & Liang, 2014), which further reduced the RMSE to 19.3 W/m^2.

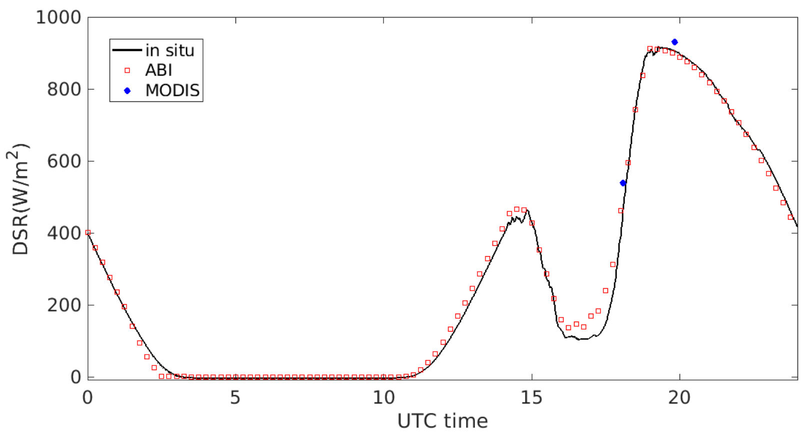

Figure 2-13. Diurnal variation of the downward shortwave radiation based on in situ measurements and satellite estimates obtained on June 10, 2018, at Fort Peck, MT.

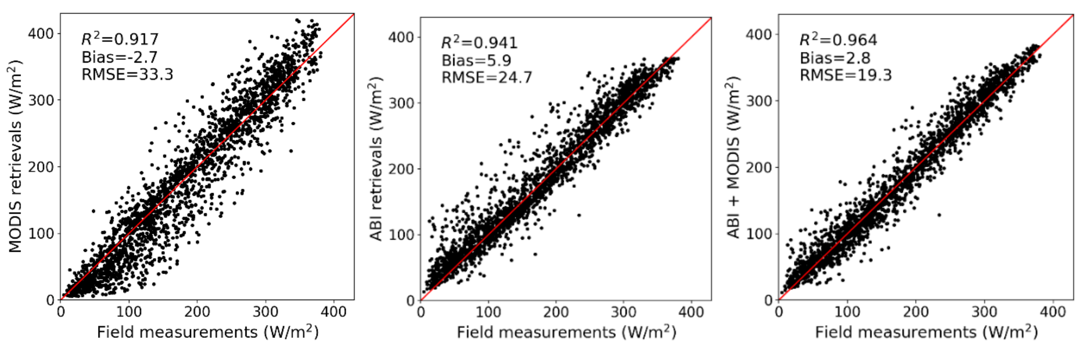

Figure 2-14. Validation results of the daily downward shortwave radiation retrieved from ABI and MODIS data.

2.6 Conclusions

In this study, daily DSR estimated derived from five global satellite products, that is, CERES-SYN, CLARA-A2, GLASS-DSR, MCD18A1, and BESS were evaluated. The products were compared at various resolutions ranging from 1 to 100 km. A full examination was conducted at a resolution of 1° in terms of the spatial, intraannual, and inter-annual analyses of the global product accuracies with focus on high-latitude areas.

CERES, BESS, and MCD18 have a full global coverage, while GLASS and CLARA present data loss over high latitude area in winter and summer over the northern hemisphere respectively. The accuracies of all products increased when the data were aggregated from 1 to 100 km, and the rate of the increase declined with decreasing resolution. The most accurate results were retrieved at 25 km over high latitude areas for most products. After aggregation to a resolution of 1°, the RMSEs of CERES, CLARA, GLASS, BESS, and MCD18 daily data were 27.6, 29.1, 30.3, 29.6, and 31.6 W/m^2, respectively. Although the high-resolution products present an overall lower RMSE, they have a potential capacity for accurately estimating the DSR over complex terrain such as the coastal region. All products indicate that the latitude has a significant effect on the DSR estimation. The most accurate DSR retrievals were obtained over mid-latitude areas. The latitude also affects the intraannual change of the estimation accuracy. A seasonal trend was not observed over low-latitude areas, whereas high rRMSE and absolute rMBD are presented in winter over mid- to high-latitude areas. The high rRMSEs over high-latitude regions in winter in the Northern Hemisphere strongly correlate with the snow-covered area, which is indistinguishable from clouds. The results show that the rRMSEs of all high-resolution products increase more than 20% in snow-covered areas than those in snow-free areas. Clouds also affect the DSR estimations based on all five products. A higher cloud visible optical depth is associated with a higher rRMSE for all products. In terms of the rMBD, CLARA is the least influenced by the high cloud area fraction and MCD18 is the least affected by the high cloud visible optical depth.

An analysis examining the long-term trend from 2001 to 2018 was conducted based on the monthly data from four satellite products. Both in situ, BESS, and CERES data have exhibited a flat trend in the past two decades, whereas CLARA and GLASS data show a dimming trend. The absolute bias and RMSE of CLARA and GLASS have continuously increased over the past two decades, while both CERES and BESS remain stable. Note that this study is only based on a limited number of SURFRAD stations in mid-latitude areas. The observed changes of different satellite products may be influenced by the latitude as well. Additional ground measurement data should be used in future research to examine global changes.

The four global DSR products with high spatial resolution that were evaluated in this study have slightly worse accuracies than the mature CERES product, mainly because of their degraded performance over snow-covered cases and the limited temporal sampling provided by polar-orbiting satellite data. Based on the proposed algorithm refinements, which address these two issues, the DSR retrieval accuracy can be substantially improved.

Chapter 3: Estimating Global Downward Shortwave Radiation from VIIRS data using a Transfer-learning Neural Network

This chapter introduces transfer learning (TL) to achieve global DSR estimation using only top-of-atmosphere and surface albedo data at local solar noon. The developed approach utilizes Visible Infrared Imaging Radiometer Suite (VIIRS) data to estimate both instantaneous and daily DSR at a 750-m resolution. These estimates were then validated against measurements from 40 independent stations worldwide. The root mean-square error and relative root mean-square error for instantaneous DSR, validated across various networks in 2013, ranged from 75.0 (24.2%) to 106.3 (18.3%) W/m^2. Daily validations achieved error metrics ranging between 30.8 (15.5%) to 33.5 (17.6%) W/m^2. Impressively, the method showcased high accuracy in polar regions, with performance comparable to traditional machine learning models, physics models, and other DSR products in other regions. Furthermore, when the algorithm was tested on VIIRS swath data for global coverage, the instantaneous mapping mirrored the spatial pattern of the cloud-mask product. Daily mapping presented patterns resembling the CERES TOA and surface fluxes and cloud products, but with finer details. Notably, after implementing transfer learning, the model's performance exhibited reduced sensitivity to the volume of training data. Overall, this chapter underscores the merits of using TL to enhance the accuracy and generalizability of DSR estimation, suggesting its potential applicability for other variable retrievals (Li et al., 2022). This study has been published (Li et al., 2022).

3.1 Introduction

Previous methods that estimate surface DSR from satellite data are largely based on the physical process of RTM. These methods can be classified as either inverse or forward algorithms (Wang et al., 2021) which are described chapter 2. ML methods have been actively used to estimate DSR over the past few years, owing to their superior capability in capturing non-linear relationships with higher efficiency. Random-forest, gradient-boosting regression tree (GBRT), and neural-network (NN) models are widely used to estimate DSR (Babar et al., 2020; Chen et al., 2019; Hou et al., 2020; Yang et al., 2018). In recent years, some deep-learning models, such as CNN (Jiang et al., 2019) and generalized regression NNs (Şenkal, 2010) have been employed (Yuan et al., 2020). Several studies have compared these ML models using the same inputs and demonstrated that NN and GBRT presented the best performances (Brown et al., 2020; Wei et al., 2019).

However, the accuracy of ML methods greatly depends on the quality and quantity of training data (Zhang et al., 2018). Some ML models were trained on ground-measurement data (Babar et al., 2020; Brown et al., 2020; Hao et al., 2019; Peng et al., 2020; Wei et al., 2019; Yang et al., 2018). Such models generally achieve high accuracy when trained at regional scale or trained with data that are representative enough. The capabilities of these models when implemented at larger spatial scales or with more scarce conditions have not been fully demonstrated as sufficient representative training data is hard to obtain, especially for polar-orbiting sensors. Thus, the applications of MLs trained on ground-measurement data to estimate DSR globally is still limited. Other ML models are trained using radiative transfer simulations which cover sufficient instances under different atmospheric conditions and over different surface types (Bue et al., 2019; Chen et al., 2019; Ma et al., 2020; Ryu et al., 2018; Shi et al., 2019; Takenaka et al., 2011; Wang et al., 2012). The data size is usually large, which is preferred by NNs. However, owing to tradeoffs between the input numbers of driving parameters and RTM model efficiencies, gaps still exist between the RTM simulated data and the real-world data. For example, most simulations are still based on less computationally expensive 1-dimensional RTM, which consider cloud as horizontally homogeneous and infinite (Chen et al., 2019), meaning that he multilayer and 3-dimensional cloud effects are ignored. At regional scale, especially over areas with sufficient measurement data, simulation-based ML models will present less accuracies.

It would be ideal to combine both simulated and ground-measurement data in the training model which can be achieved by adapting the concept of transfer learning (TL). TL uses the model trained on source domain data as a starting point, and is fine-tuned in the process of developing a model for target domain (Pan & Yang, 2010). It is a powerful tool to transfer pre-known knowledge to a target model with improved generalization ability, even when the target training datasets are limited (Pan & Yang, 2010). Learning from simulations is a good practice of TL (Malmgren-Hansen et al., 2017), as simulations provide sufficient data covering comprehensive conditions and carries traceable physics knowledge which is valuable for initial model training. The gaps between the simulated data and real-world data can be filled when the model is further fine-tuned. TL has been applied in some studies of remote sensing, mostly in the area of land cover classification (Huang et al., 2018; Tong et al., 2020), where transferred CNN models were adapted to classify optical remote sensing images. Other areas include extending regional models (Wei, Zheng, & Yang, 2016) or extending one variable to similar variables estimation (Khaki et al., 2021). Transferring from simulated data has also conducted to classify synthetic aperture radar target (Zhou et al., 2020). However, to the best of our knowledge, TL has not been employed to improve DSR estimation.

This study proposes a robust NN method that employs TL to estimate instantaneous and daily DSR based on Visible Infrared Imaging Radiometer Suite (VIIRS) TOA reflectance and Moderate Resolution Imaging Spectroradiometer (MODIS) black-sky albedo (BSA) at local solar noon only. The proposed method leverages the advantages of both using radiative transfer simulation data and ground measurement data for training, which enables accurate estimation of DSR under various climate regions, even over polar regions. The remainder of this paper is organized as follows: Section 3.2 introduces the input data and data separation method; Section 3.3 describes the methodology; Section4 presents and discusses validation results; the main conclusions are drawn in Section 3.5.

3.2 Data

3.2.1 Real-world data

Let $\boldsymbol{D_{real}}$ be the real-world data serving as the target data domain in this study. $\boldsymbol{D_{real}}$ is composed of satellite data and in-situ measurements which are matched based on their locations.

The TOA reflectance, geometry information, and surface BSA are extracted from satellite data: VIIRS/Suomi NPP Moderate Resolution Bands (VNP02MOD), VIIRS/Suomi NPP Moderate Resolution Terrain-Corrected Geolocation (VNP03MOD), and MODIS/Terra+Aqua Albedo (MCD43A3), respectively. VNP02MOD provides calibrated TOA spectral reflectance and pixel-level quality flags at 750-m resolution. Bands M1–M11 (except for M6) are used as inputs. VNP03MOD includes corresponding geographical coordinate locations and viewing geometry information. Three geometry angles are extracted. The surface BSA at local solar noon for MODIS bands 1 through 7 are retrieved from MCD43A3, which are produced daily at 500-m resolution. 16 days of Terra and Aqua data are weighted in MCD43, consequently containing more available data.

The in-situ observations of DSR are collected from four different networks serving as different purposes in this study. 96 sites from FLUXNET global network of micrometeorological tower sites in 2013 are used as training/fine-tuning datasets. The results are validated over 25 Baseline Surface Radiation Network (BSRN) stations, eight Greenland Climate Network (GC-NET) stations with data in 2013, and seven Surface Radiation Network (SURFRAD) stations with data in both 2013 and 2014. All the validation data are independent from the training data. Both SURFRAD and BSRN adopted high-level measurement standards set by the World Climate Research Program of the World Meteorological Organization (Augustine et al., 2000). The distribution of the in-situ sites is shown in Figure 3-1. The instantaneous DSR is calculated by averaging the nearest half-hour data from the VIIRS overpass time, and the daily DSR is the mean of all instantaneous measurements in one day and is calculated as in Li et al. (2021). Measurements with a solar zenith angle (SZA) higher than 87 degrees are excluded.

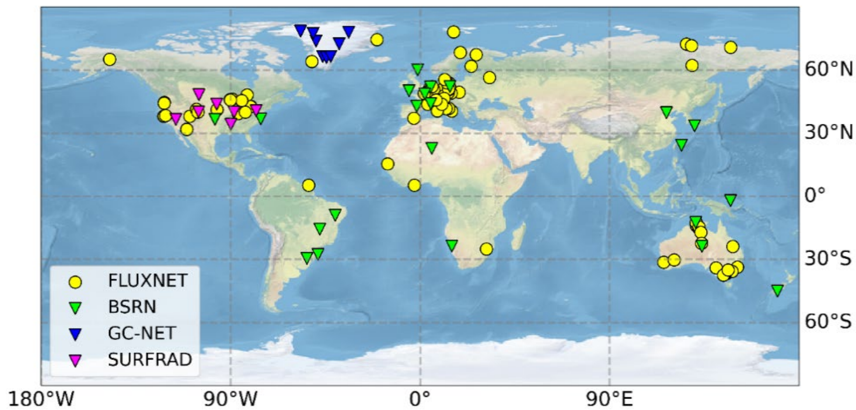

Figure 3-1. Distribution of ground measurement stations. Training data are labeled as circles. Validating and testing data are labeled as triangles.

3.2.2 Simulated data

Let $\boldsymbol{D_{sim}}$ be the source domain of this study. MODTRAN5 (Berk et al., 2004) is employed to generate the simulated database. Following Wang et al. (2015), our radiative transfer model is operated for 294 cases of viewing geometry, including seven values of SZA, six values of viewing zenith angle (VZA), and seven values of relative azimuth angle (RAA). Different combinations of surface and atmospheric parameters are used as inputs for each view. The corresponding TOA reflectance of VIIRS bands and simulated instantaneous and daily DSR measures are also recorded. To cover comprehensive surface types, a library of 245 surface albedo is used, which is obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (Baldridge et al., 2009) and the U.S. Geological Survey (Clark et al., 2007) surface spectra databases. In terms of atmospheric inputs, different driving parameters are considered under clear and cloudy sky conditions. The aerosol types, aerosol optical depths (AODs), and water-vapor loads are modeled under clear skies.

For cloudy skies, cloud optical depths (CODs), types, heights, and water-vapor loads are set as input parameters. The scattering and absorption of atmospheric particulate matter are ignored when clouds are present.

Daily DSRs are averaged from instantaneous DSRs throughout the day, assuming that the atmospheric conditions remain the same. Under the same VZA, RAA, surface, and atmospheric conditions, instantaneous DSRs at any daytime, t, in a day are only related to SZA. Consequently, daily DSR simulations includes two more parameters (i.e., latitude and solar declination angle) to calculate SZA(t). Detailed settings are listed in Table 3-1.

In summary, MODTRAN is operated with the combination of 294 cases of view geometry, 245 surface albedo, 576 cloudy-sky atmospheric conditions, and 210 clear-sky atmospheric conditions to simulate a comprehensive database. However, not all MODTRAN variables are used as NN inputs. Sensitivity analysis of previous studies have shown that SZA, albedo, and cloud optical depth are the top-three factors influencing the DSR estimation under all sky conditions (Brown et al., 2020; Chen et al., 2019). Hence, only spectral TOA reflectance, surface albedo, and geometry information data are required for the proposed model. The other driving parameters for the simulation model are flattened. We then have our simulation database, $\boldsymbol{D_{sim}}$.

Table 3-1. Driving parameters used in MODTRAN

Parameters	**Values**
View geometry	
Solar zenith angle (SZA)	0,10,20,30,45,60,75
View zenith angle (VZA)	0,10,20,30,45,60,75
Relative azimuth angle (RAA)	0,30,60,90,120,150,180
Clear sky	
Aerosol types	5 values
Aerosol optical depth (AOD)	0.0,0.05,0.10,0.2,0.4,0.6,1.0

Water vapor	0.5,1.0,1.5,3.0,5.0,7.0
Cloud sky	
Cloud types	stratus, cumulus, altostratus, nimbostratus
Cloud height	3 values
Cloud optical depth (COD)	0.25,1,2,3,5,10,20,30,60,120,180,240
Water vapor	1.5,3.0,7.0,12.0
Daily	
Latitude	0,10,20,30,40,50,60,70,80,90
Declination angle	-23.5:23.5/4:23.5

3.2.3 Additional data

The results of proposed method is compared with existing high resolution DSR products including the DSR product based on the Earth Polychromatic Imaging Camera (EPIC), the Breathing Earth System Simulator (BESS) shortwave radiation product,the Global Land Surface Satellite (GLASS) DSR product, and MODIS/Terra+Aqua Surface Radiation (MCD18). The EPIC uses the random forest method to generate DSR at 0.1-degree spatial resolution and 1-2h temporal resolution. The BESS product is based on NN trained from RTM simulations. A series of MODIS atmosphere and land product data are included as the inputs of the RTM (Ryu et al., 2018). The product generates 5 km DSR globally. The GLASS also provides daily DSR at 5 km; it applies the direct estimation approach based on MODIS TOA spectral reflectance and GLASS broadband albedo (Zhang et al., 2019). The MCD18A1 Version 6.1 generates 3h DSR at 1km; it employs the LUT approach in which the MODIS TOA reflectance and albedo are used as the main inputs (Wang et al., 2020). It has been used as inputs of some surface variable retrieval methods (Jia et al., 2021; Wu et al., 2021). Clouds and the Earth's Radiant Energy System (CERES) for Synoptic TOA and surface fluxes and clouds (SYN) DSR data are employed in this study as a reference for daily estimation. It offers the accurate satellite DSR estimation at 1 degree worldwide (Li et al., 2021; Zhang et al., 2015). CLDMSK_L2_VIIRS_SNPP data are also used. This product inherits the algorithms designed for MOD35 at a spatial resolution of 750 m. Both snow and cloud information used in this study are derived from the CLDMSK_L2_VIIRS_SNPP product.

3.3 Methods

3.3.1 NN structure

Simple back-propagation NNs have been widely used to estimate DSR in various studies (Peng et al., 2020; Ryu et al., 2018; Wang et al., 2012). The theoretical base of the algorithm was described by Takenaka et al. (2011). The input layer of the proposed model has 20 neurons for the instantaneous estimation model, which includes 10 bands of TOA reflectance, seven bands of surface BSA, and three geometry angles. For daily models, two more neurons are added: latitude and declination angles. The sensitivity analysis of the input parameters is discussed in Section 4.1. To determine the hyper-parameters of the NN model, the models have been trained on different combinations of different hyper-parameters. SURFRAD data are used to validate and select the model parameters. To balance the model accuracies and the computational efficiencies, a four-layers (i.e., one input, two hidden, and one output) NN model is selected. The first hidden layer has 25 neurons followed by another hidden layer of five neurons with a rectified linear unit activation function (Nair & Hinton, 2010). The output layer has one neuron corresponding to the

DSR value of linear activations. The weights of the model, w, are initialized following the Glorot uniform:

$$w \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_l+n_{l+1}}}, \frac{\sqrt{6}}{\sqrt{n_l+n_{l+1}}}\right], \text{ for } l = 0, \ldots, \text{L-1}, \quad (1)$$

where L is the number of layers in the NN, and n_l is the number of neurons in the layers. Because the final outputs of the model are the numerical values of the DSR, the mean-square error (MSE) function is used as the error function:

$$E(w) = \frac{\sum_{i=1}^{n}[y(w,x_i)-h_i]^2}{n}, \quad (2)$$

where h_i is the simulated/measured DSR for the ith sample. The error function is further minimized using RMSprop (Hinton, Srivastava, & Swersky, 2012), which is a type of stochastic gradient descent that enables a stable learning curve and boosts the convergence speed by calculating the differential square weighted average of the gradient.

3.3.2 Transfer learning model construction

The construction of the transfer learning model requires two rounds of training (Figure 3-2). It is firstly trained from the source domain data, which is D_sim in this study. The model here is referred to as the base model (BM). Then the model is further fine-tuned by the target domain data, D_real. This is the final model used as DSR estimation, and is referred to as a transfer learning model (BM_TL).

BM was constructed following the NN configuration, with training datasets consisting completely of simulation data. BM can be summarized as $y(x, w), x \in D_{sim}$. We randomly separated the simulation data into 80% training and

20% as training and validation data. Since the data size is significantly large, no cross-validation was applied. We also included SURFRAD measurements as additional validation data to ensure that the selected BM model is not overfitted to the ground measurements. BM learned the knowledge from the simulation data. To adapt the model to real-world data, the model is fine-tuned by the matching data between satellite products and FLUXNET measurements. Previous studies revealed that the first layers of the NN models are not usually specific to a particular task but instead general to many related tasks, with the last few layers being task-specific (Yosinski et al., 2014). In this study, the weights of the first layer, w_1, were directly copied from BM, as shown in Figure 3-2, meaning that they are initialized from $argmin\ E(w, x), for\ x \epsilon D_{sim}$. These weights were frozen and were not changed. The remaining weights from other layers were initialized by (1), and updated using RMSprop during the minimization process of (2), but with $x \epsilon D_{real}$. Consequently, the BM_TL can be summarized as $y_{tl}(x, w_{tl}), x \epsilon D_{real}$.

3.3.3 Evaluation

In this study, the proposed method was trained and evaluated on independent sites (i.e., no overlap sites exist between the training and testing data) as DSR estimation is related to surface properties. FLUXNET data served as training/tuning data, as they have comparatively sufficient measurements globally. SURFRAD sites were used for model selection and tuning hyper-parameters. The models were validated on BSRN and GCNET sites separately. FLUXNET, SURFRAD, and BSRN presented similar variable distributions while GCNET showed a different distribution with the peak of the visible bands of TOA and surface albedo located at larger values. GCNET provides unique measurements over high latitude areas where some existed DSR products degrade (Li et al., 2021). Hence, it helps to investigate the model's performance on under-representative conditions. The root mean square error (RMSE), relative root mean square error (rRMSE), mean bias difference (MBD) and R-square (R^2) measures are the main assessment metrics of this study.

3.3.4 Other methods

To compare the traditional NN method with the proposed method, a baseline model, BM_SA, is trained purely from in-situ measurements. BM_SA can be summarized as $y_{sa}(x, w_{sa}), x \epsilon D_{real}$. The same NN configuration as BM_TL is applied. FLUXNET data are used as the only training data and SURFRAD data are still used as validation datasets.

The narrow-band LUT approach was first proposed by (Liang et al., 2006), which includes a reflectance LUT to determine the atmospheric condition and a flux LUT to finalize surface DSR estimation. The LUT stores the pre-calculated results from the same RTM simulations. The method is further developed by Wang et al. (2020) to generate MCD18 products. It is noticeable that the original LUT method was applied to the MODIS TOA band. In this study, the VIIRS band is used instead. Divergence between VIIRS and MODIS bands may lead to the results being slightly different from the published ones (Wang et al., 2020).

Direct estimation utilizes the linear relationship between TOA reflectance and surface shortwave net radiation of narrow bands under cloudy and clear-sky conditions (Wang et al., 2021). The linear coefficients are also retrieved from the same RTM simulation data. In this case, we generate coefficients for each VIIRS band. The surface shortwave net radiation is then estimated, and the DSR is calculated by combining broadband albedos. This method employs cloud-mask information from the CLDMSK_L2_VIIRS_SNPP products.

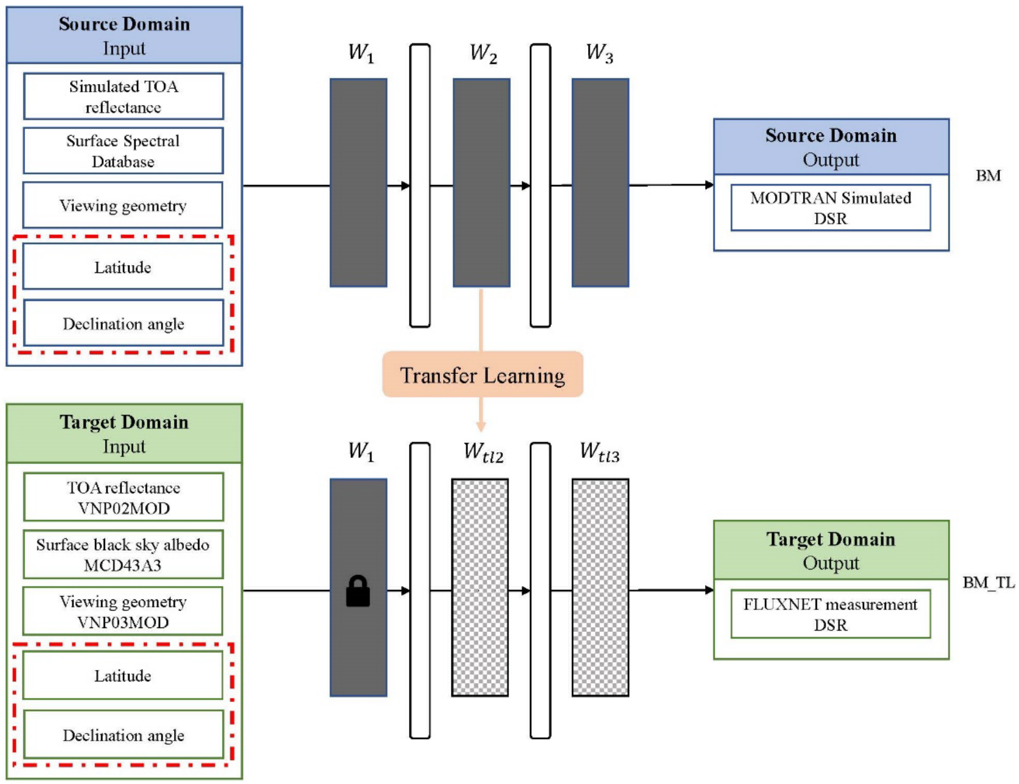

Figure 3-2. Flowchart for constructing BM_TL.

3.4 Results and discussion

3.4.1 Sensitivity analysis

Since using VIIRS TOA reflectance as inputs for DSR estimation is new, a sensitivity analysis was conducted to investigate the impact of the input variables for both instantaneous and daily DSR estimation. The Sobol analysis method (Sobol, 2001) from the SALib library in Python was employed. Both first-order indices (S1) and total-order indices (ST) were calculated. S1 measures the contribution to the output variances caused by a single input alone. ST is S1 plus all higher-order interactions of this input. The results are shown in Figure 3-3. For instantaneous DSR, the visible band 1-4 with the reflected range from 0.4-0.6 μm carry the mostof the information needed for the DSR estimation. Among them, band3 and band4 account for more than 65% of the total variance. The near infrared to medium-wave infrared bands 5-11 also account for 12.5% in total, but most of the variance are from the interaction with other inputs. Unlike previous sensitivity analysis based on forward methods (Li et al., 2015; Chen et al., 2012), the SZA is not the first dominant factors but still plays an important role with 6.8%. Surface BSA band 3 and 4, combined, also influence the DSR estimation with 6.5%, which provides required surface conditions and multi-scattering information. The influence of other bands of surface BSA are negligible. The daily model has similar sensitivity analysis results as instantaneous one, but it is noticeable that more variances are from the interactions with other variables, as ST has a smaller portion of S1 compared with instantaneous model. Besides, SZA occupies smaller part. More percentages are given to latitude and declination angles.

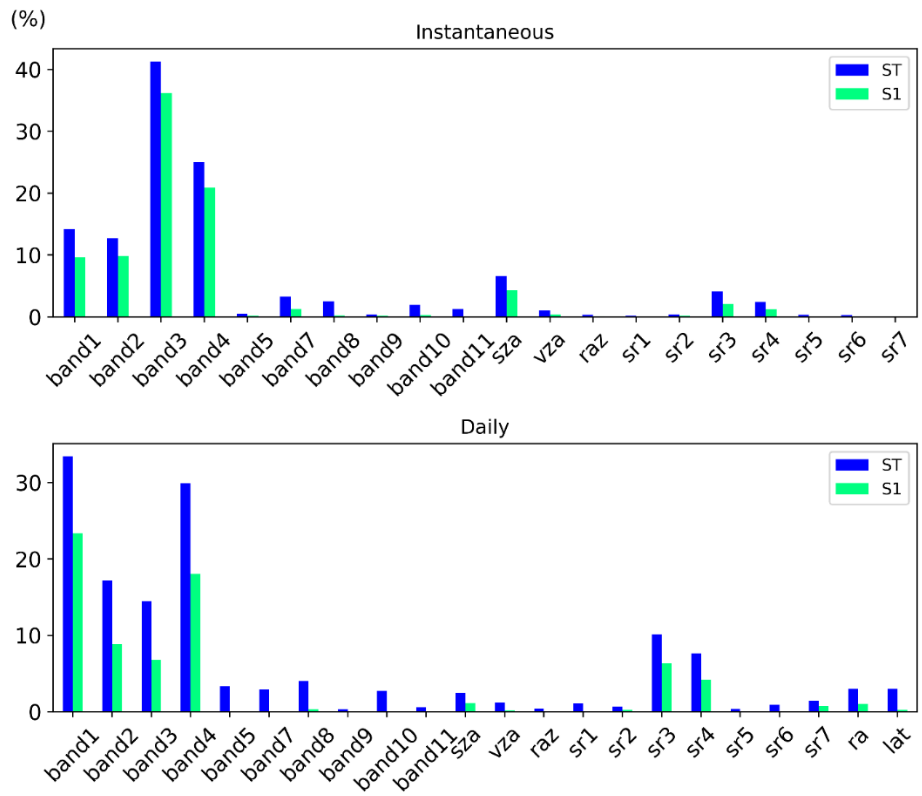

Figure 3-3. The first-order indices (S1) and total-order indices (ST) for input variables of daily and instantaneous BM_TL.

3.4.2 Comparison with traditional NN methods

The proposed method is tested on the various in-situ networks in 2013 and the results are shown separately due to the different purposes of each network in this study. FLUXNET data serve as the training data of BM_SA and BM_TL model and hence represent the training accuracies of these two models. To demonstrate how TL improves the traditional ML, the proposed model (BM_TL) is also compared with BM and BM_SA. Figure 3-4 and Figure 3-5 show the validation results for instantaneous and daily estimations, respectively.

For instantaneous estimation, BM had a higher RMSE than BM_SA over FLUXNET, SURFRAD and BSRN stations. The RMSE of the BM model over FLUXNET, SURFRAD and BSRN data scored 127.0, 126.2 and 104.5 W/m^2respectively, whereas BM_SA achieved an RMSE of 104.3, 101.1 and 92.4 W/m^2. Over the GC-NET stations, BM had a lower RMSE of 81.9 W/m^2compared with BM_SA which had an RMSE of 100.2 W/m^2. The RMSE of BM_TL were 109.7 (22.7%), 106.3 (18.3%), 91.2 (16.1%), 75.0 (24.2%) W/m^2over FLUXNET, SURFRAD, BSRN, and GCNET respectively. Compared with BM_SA, BM_TL had a slightly higher RMSE over FLUXET and SURFRAD, but lower over BSRN and GCNET stations. BM_TL also had better accuracies than BM over all the sites.

The comparisons between the estimated daily DSR are shown in Figure 3-5. Similar to the instantaneous estimation, BM_SA achieved significantly better results than BM over the FLUXNET, SURFRAD and BSRN station data. The RMSE of the BM at FLUXNET, SURFRAD and BSRN stations were 54.2, 54.5, and 49.3 W/m^2, respectively. The RMSE of BM_SA at FLUXNET, SURFRAD and BSRN stations was 34.3, 35.7, and 32.5 W/m^2. It is noticeable that the difference between BM and BM_SA over these two networks was larger than the instantaneous estimations. This may indicate that the simulated daily samples have more systematic deviations from the truth than the instantaneous samples did due to the assumption of stable atmospheric conditions throughout a day. Owing to the insufficient training data over polar regions, the overfitting issues of the BM_SA model persisted, which led to the highest RMSE of 52.8 W/m^2over the GC-NET stations among the three models. The employment of BM_TL presented significant improvements over GC-NET stations and slightly better results for the FLUXNET, SURFRAD and BSRN stations, The RMSEs of BM_TL were 33.2 (18.1%), 33.5 (17.6%), 30.8 (15.5%), and 31.3 (14.4)

W/m^2 over FLUXNET, SURFRAD, BSRN, and GC-NET, respectively, at raw resolutions for daily DSR estimations.

In summary, BM_SA achieved a higher accuracy for the FLUXNET, SURFRAD and BSRN stations for both instantaneous and daily estimations, but it degraded dramatically for GC-NET stations. This corresponds to the overfitting issues of the ML model trained only on in-situ data only. The SURFRAD and BSRN stations have a similar sample distribution as with the training data from FLUXNET, which led to high accuracy. Over polar regions, no sufficient training samples were available; hence, BM_SA fell short. On the other hand, BM displayed superior performance over high latitude areas. No obvious bias as observed, and the RMSE was comparable to previous methods (Zhang et al., 2018), demonstrating the strength of training using simulation data that cover sufficient atmospheric and surface cases. However, the accuracy of general cases was low, especially for daily estimations. BM_TL successfully combined the advantages of both BM and BM_SA, boosting its generality as it performed stably over all networks and resolved the issues of inaccurate daily simulated samples; it improves the accuracies as well. These results support the previous findings that TL can leverage both source and target data to better tune the model, and it demonstrates the feasibility of applying TL to DSR estimation.

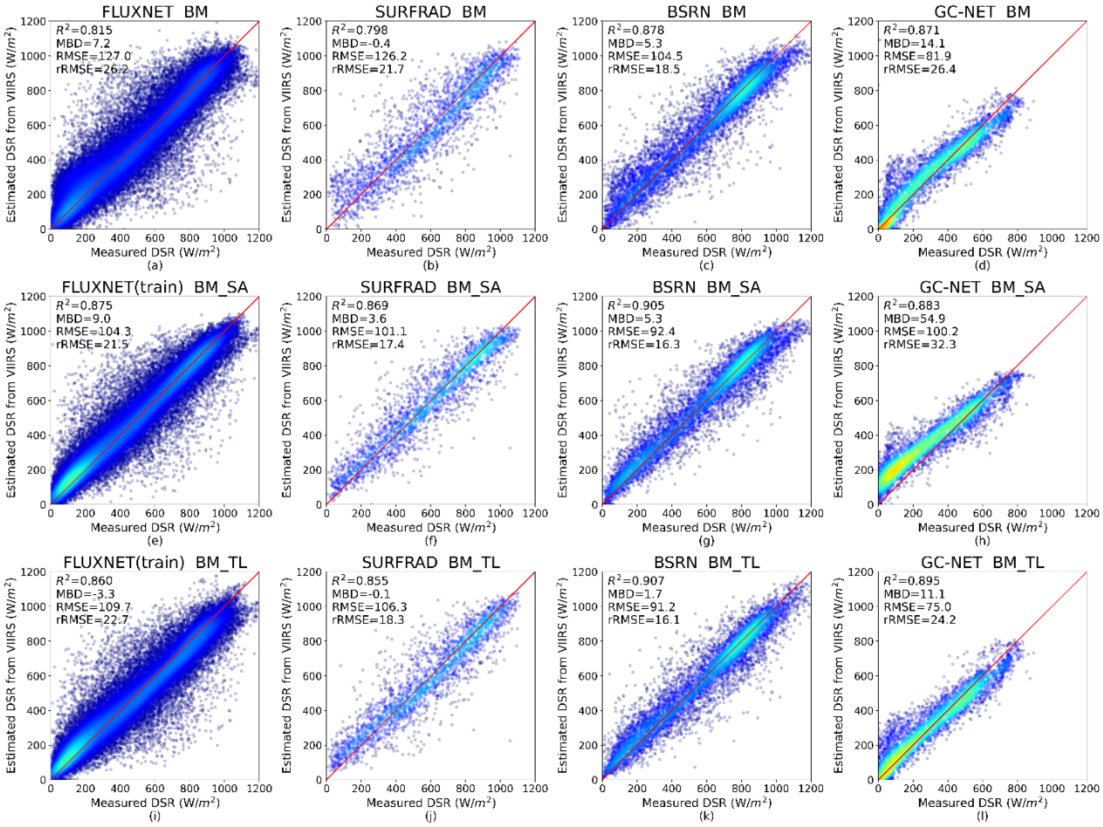

Figure 3-4. Comparison among the estimated instantaneous DSR by BM, BM_SA, and BM_TL models over ground measurements at SURFRAD, BSRN, and GC-NET stations in 2013.

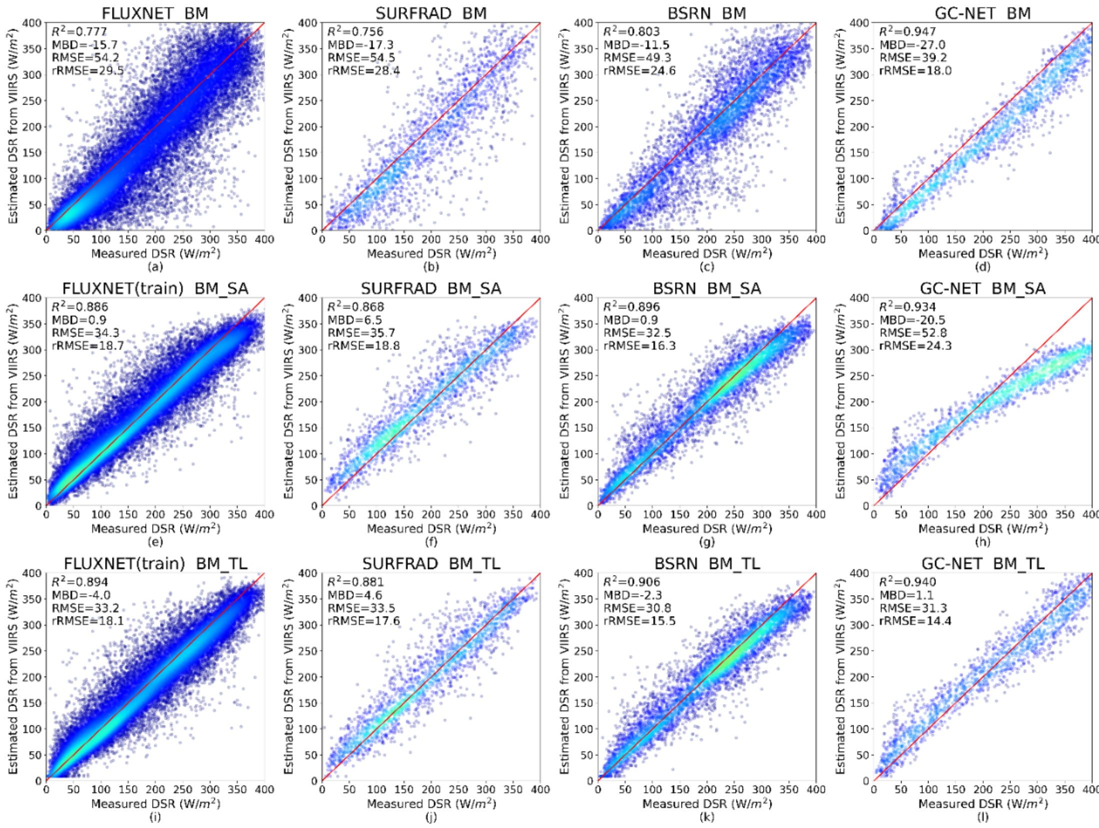

Figure 3-5. Comparisons among the estimated daily DSR by BM, BM_SA, and BM_TL models over ground measurements at SURFRAD, BSRN, and GC-NET stations in 2013.

To further investigate how BM_TL improved the traditional ML, the BM, BM_SA, and BM_TL models were compared at different latitudes and SZAs, as shown in Figure 3-6. The relative RMSE (rRMSE) of BM and BM_SA were similar over low- to mid-latitudes. BM_TL achieved comparable or better results for those areas, but the improvements were limited. Over areas greater than 70°, the rRMSE of all three models dramatically increased, especially for BM_SA. The difference between BM_SA and BM was large in this case. BM_TL still achieved better results than BM over high latitude areas, and the same conclusions were drawn from the comparisons of different SZA. BM_SA, BM, and BM_TL all showed similar performance when

SZA was low. When SZA was greater than 60°, BM_SA tended to fail as rRMSE increased. The rRMSE of BM increased as well, but the magnitude was small compared to BM_SA. The rRMSE of BM_TL drops in the middle of BM and BM_SA for high SZA samples.

Figure 3-7 shows the comparison of the three instantaneous models under different snow and cloud conditions. Both snow and cloud information are derived from CLDMSK_L2_VIIRS_SNPP product. First, the performance of all three models varied with the testing data. For GCNET stations, regardless of conditions, BM_SA had the highest rRMSE followed by BM. BM_TL significantly improved this by more than 5 % rRMSE under all conditions. For BSRN and SURFRAD stations, the three models had different performances under different conditions. For non-snow samples, the rRMSE of all three models were similar. For snow-covered pixels, the rRMSE of all three models increased, but BM_TL achieved the lowest rRMSE. For atmospheric conditions, the rRMSE of all models increase when it pertained to cloud over the BSRN and SURFRAD stations. When conditions were clear or probably clear, the three models present similar estimation accuracies. As it was probably cloudy or cloudy, BM showed the worst performance. This may originate from the simulation data which did not consider the effect of AOD, multilayer cloud, or 3-dimensional cloud. This issue is addressed by BM_TL through incorporating the ground measurements especially for completely cloudy conditions. However, BM_TL still not function as well as BM_SA for probably cloudy conditions. Overall, the results demonstrate that BM_TL normally aligns well with the lower RMSE achieved by the two traditional models, leading to significant improvement under scarce atmospheric and surface conditions mostly over high latitudes and high SZAs areas compared with BM_SA, and over cloudy cases compared with BM.

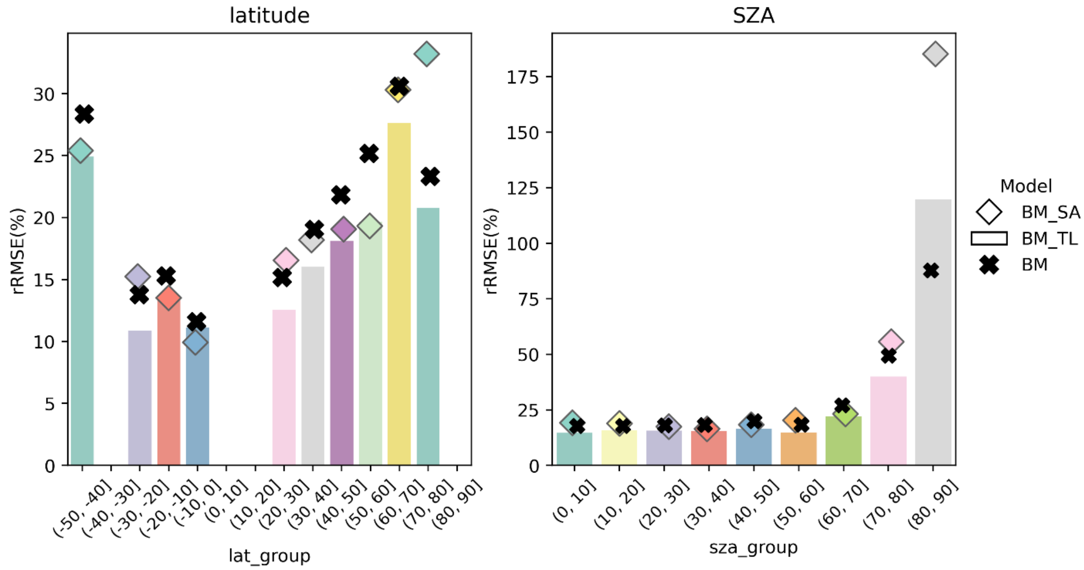

Figure 3-6. Comparisons of the estimated instantaneous DSR by BM, BM_SA, and BM_TL models at different latitudes and SZA.

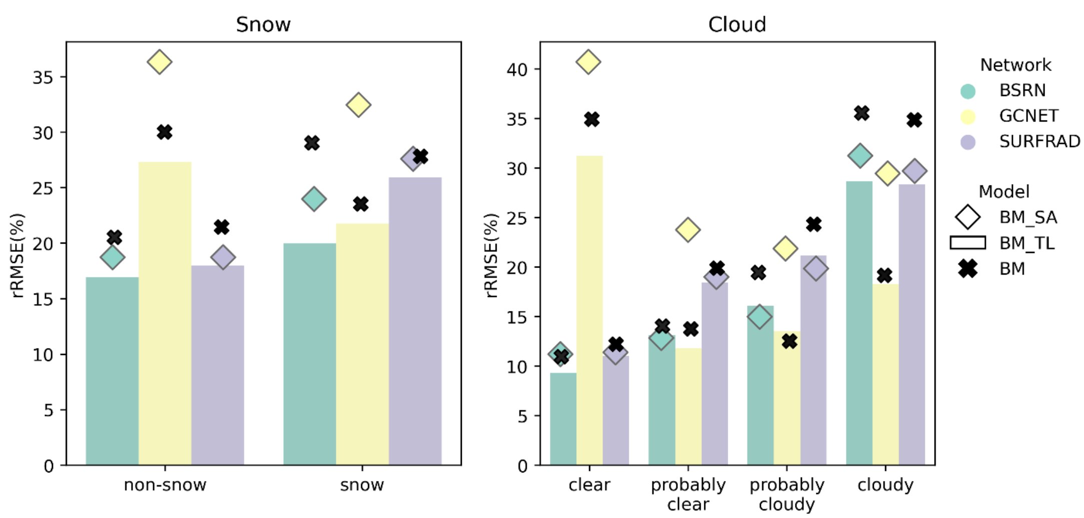

Figure 3-7. Comparison between the estimated instantaneous DSR by BM, BM_SA, and BM_TL model at different snow and cloud conditions for each testing network.

3.4.3 Comparison with physical methods

In this study, the BM_TL method was compared with two traditional physical methods, narrow-band LUT and direct estimation methods, using the same input data to ensure consistent comparisons. The detailed description of these two methods can be found in Wang. et al (2021). The coefficients of LUT and the direct estimation method were also retrieved from the same RTM simulated data. We kept all the available estimations for each method in the comparison. Figure 3-8 and Figure 3-9 show the validation results over SURFRAD, BSRN, and GC-NET stations for instantaneous and daily estimation, respectively. For instantaneous DSR estimation, the RMSE of LUT and the direct estimation method for SURFRAD stations were 140.0 and 143.2 W/m^2, respectively, and for BSRN stations, they were 108.4 and 116.8 W/m^2respectively. The BM_TL model achieved a much lower RMSE with 106.3 and 91.2 W/m^2over SURFRAD and BSRNs. For daily DSR estimation, the RMSE of the LUT and the direct estimation method for SURFRAD stations were 48.6 and 44.9 W/m^2respectively, and for BSRN stations, they were 43.5 and 42.9 W/m^2, respectively. Again, the BM_TL model achieved a lower RMSE with 30.8 and 31.3 W/m^2over SURFRAD and BSRN, respectively. There are some possible reasons for the superior performance of BM_TL. Though the three methods are based on the same simulation data, both LUT and the direct estimation method only use linear interpolation to calculate DSR, while BM_TL can capture the non-linear relationship. Furthermore, in addition to the simulation data, BM_TL leverages information from the ground measurements, which can minimize the uncertainties that originate from the ignored conditions in the simulation data.

The superiority of BM_TL was also particular to the case of estimating over GC-NET stations. The RMSE of LUT and the direct estimation method instantaneous estimation were 124.4 and 149.6 W/m^2, respectively. Invalid estimations of the direct estimation method exist at samples with SZA higher than 75 degrees which explains the higher RMSE while lower rRMSE compared with LUT. BM_TL has a much lower RMSE with 75 W/m^2 (24.2%) and, notably, R^2 was dramatically improved. For daily estimation, the RMSEs were 52.1 and 40.8 W/m^2, respectively, for LUT and direct estimation methods, whereas BM_TL achieved a RMSE of 38.0 W/m^2. In addition to the above-mentioned reasons, the improvements of BM_TL over LUT at GCNET stations are also due to the combination of multiple bands ranging from visible to medium-wave infrared to better discriminate the snow from cloud. The validation results further reveal the dilemmas faced by traditional inversion methods can be improved by BM_TL especially over high-latitude areas.

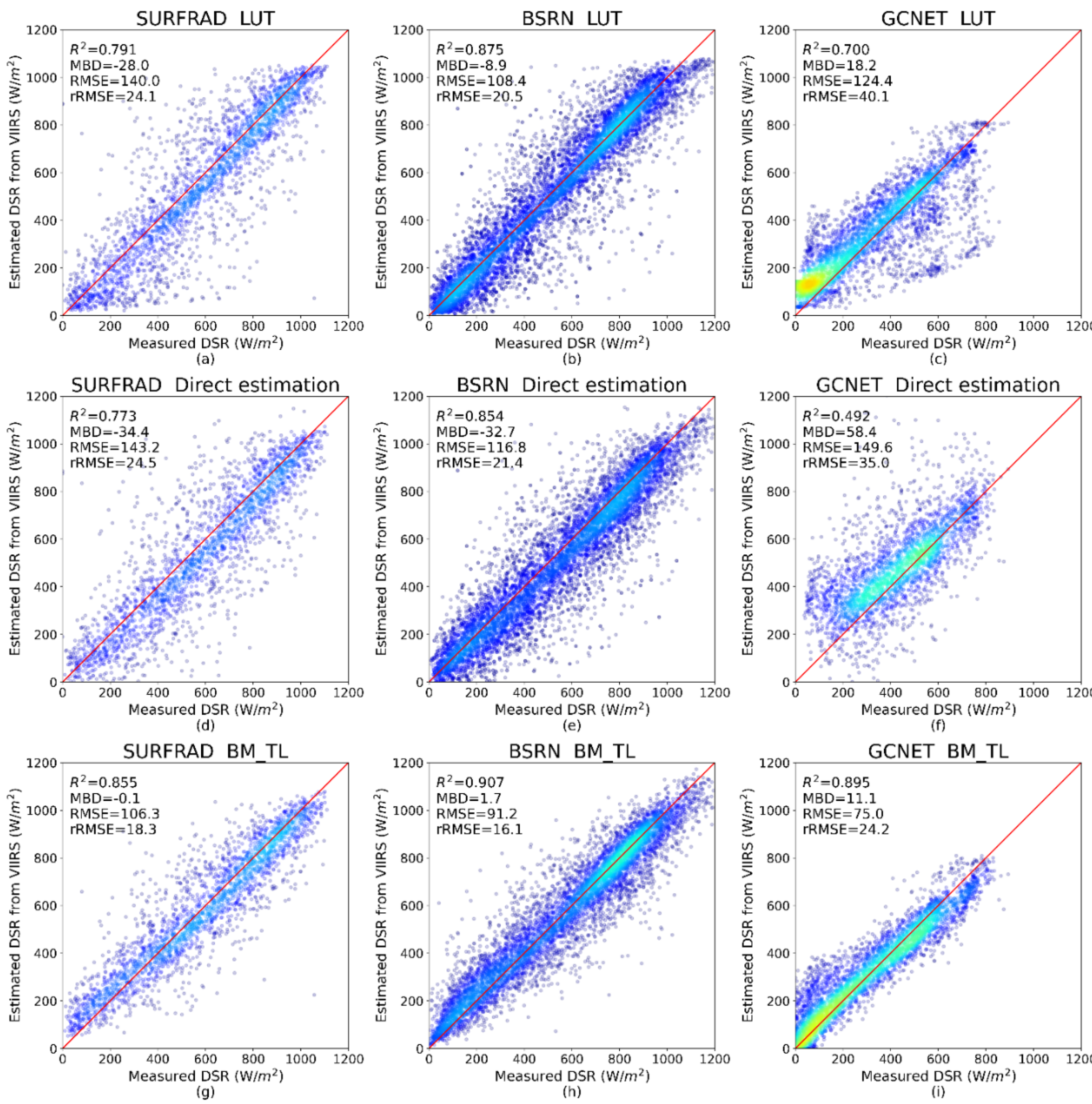

Figure 3-8. Comparison between the estimated instantaneous DSR by LUT, Direct estimation, and BM_TL model with ground measurements at SURFRAD, BSRN, GC-NET stations.

Table 3-2. Summary of comparisons between BM_TL estimated instantaneous DSR with EPIC, GLASS, and BESS products.

Products	**Time scale**	R^2	**RMSE** (W/m^2)	**Bias** (W/m^2)	**rRMSE (%)**	**Reference**
EPIC	Hourly	0.83	113.21	1.99	23.44	(Hao et al., 2019)
GLASS	Instantaneous	0.83	115	-6.5	-	(Ryu et al., 2018)
BESS	Instantaneous	0.84	123.3	-0.3	-	(Ryu et al., 2018)
BM_TL	Instantaneous	0.91	89.6	4.3	18.30	

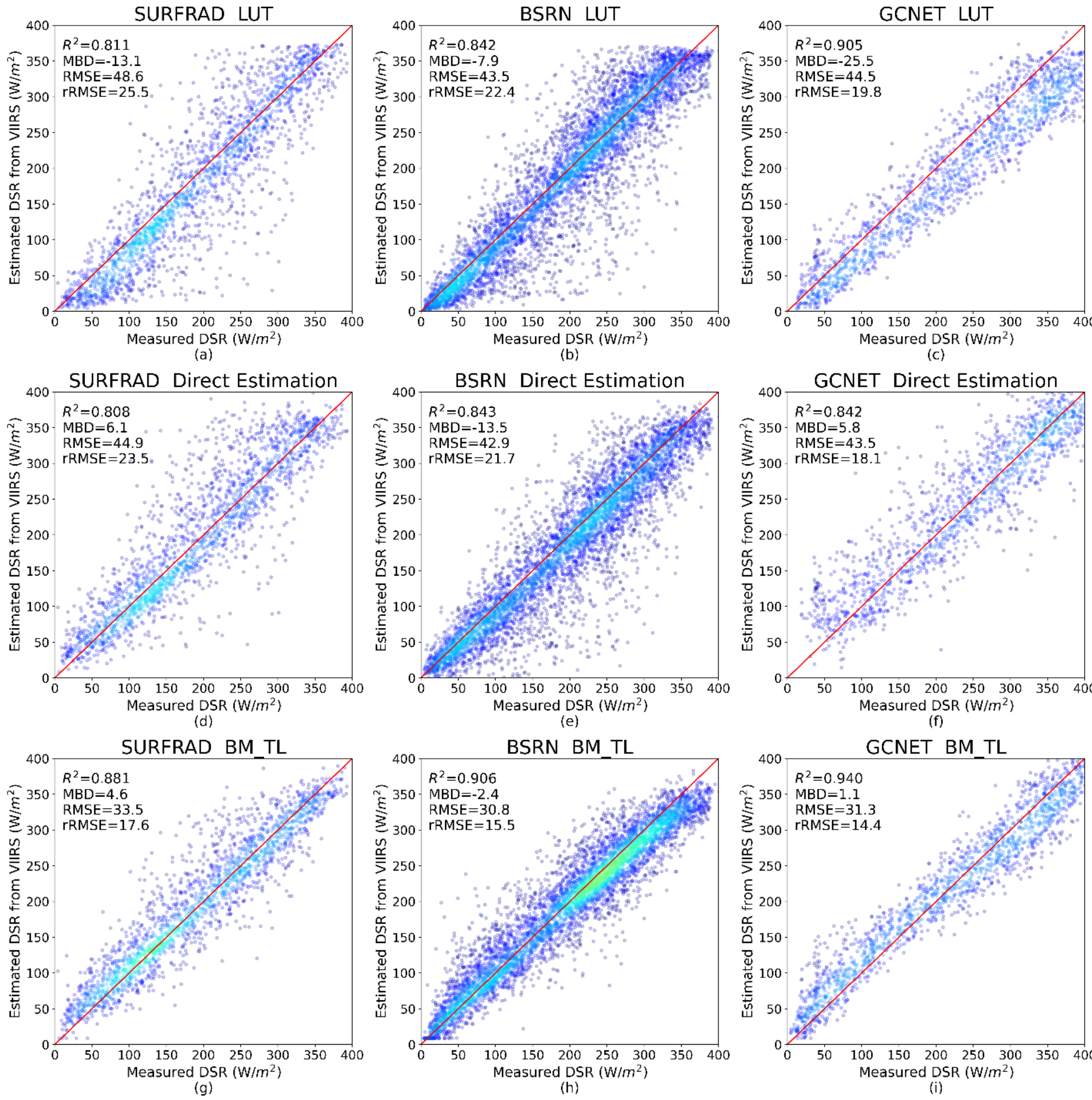

Figure 3-9. Comparison between the estimated daily DSR by LUT, Direct estimation, and BM_TL model with ground measurements at SURFRAD, BSRN, GC-NET stations.

3.4.4 Comparison with existed DSR products

The validation results of the proposed model were also compared with existing high resolution DSR products. Table 3-2 presents the comparison results with GLASS, BESS, and EPIC at an instantaneous time-scale. The statistics were collected from several recently published literatures (Hao et al., 2019; Ryu et al., 2018). The proposed method had a significantly higher R^2 and lower RSME and rRMSE than existed products. However, the results are not based on the same validation sites during the same time period, hence it only suggests the comparability between the products.

For the daily DSR estimation, we validated and compared BESS, GLASS, and MCD18 with our proposed method over the same sites from SURFRAD, BSRN and GCNET (Table 3-3). Invalid values were removed from the comparison. BM_TL has a comparable RMSE and rRMSE over all stations, and is least biased compared to the other three products. Furthermore, it presents significantly better accuracies over the GCNET stations with the lowest RMSE and bias. Over the BSRN and SURFRAD sites, the advantages of the propose method are not pronounced. Some possible reasons can be considered. Firstly, spatial resolution influences the DSR validation results (Li et al. 2021). The lower the spatial resolution is, the higher the validation accuracies are. BESS and GLASS are 5km products and MCD18 are 1km products, while the proposed method estimated daily DSR at 750m. Secondly, the data source may influence the retrieval accuracies. BESS, GLASS, and MCD18 are all based on MODIS data while the proposed method is using new VIIRS TOA data. BESS, GLASS, and MCD18 are mature products while the proposed method is still at the development stage. No adjustments procedures have been added for product generation.

Table 3-3. Summary of comparisons between BM_TL estimated daily DSR with BESS, GLASS, and MCD18.

Network	Products	R^2	RMSE (W/m^2)	Bias (W/m^2)	rRMSE (%)
All	BESS	0.91	-3.2	32.0	15.4
	GLASS	0.83	-16.6	45.2	21.7
	MCD18	0.77	-19.4	53.9	25.9
	BM_TL	0.90	0.2	31.7	15.2

BSRN and SURFRAD	BESS	0.91	-1.9	30.6	15.2
	GLASS	0.93	-7.7	27.8	13.8
	MCD18	0.89	-7.8	35.1	17.4
	BM_TL	0.90	0.2	31.6	15.7
GCNET	BESS	0.89	-9.2	37.9	16.0
	GLASS	0.64	-57.1	88.4	37.2
	MCD18	0.55	-72.1	102.6	43.2
	BM_TL	0.92	-0.2	32.2	13.6

3.4.5 Temporal analysis

Since the training data of the proposed models are completely from 2013, we also validated the models with measurements in 2014 over SURFRAD sites to test its capability for future inter-annual analysis. The results are shown in Figure 3-10. The rRMSE are 17.5% and 17.1% respectively for instantaneous and daily estimation, and no pronounced bias are observed. It presents a similar accuracy for both instantaneous and daily estimation compared with the validation results over the same sites in 2013 which are 18.3% and 17.6% in terms of rRMSE (Figure 3-4 and Figure 3-5).

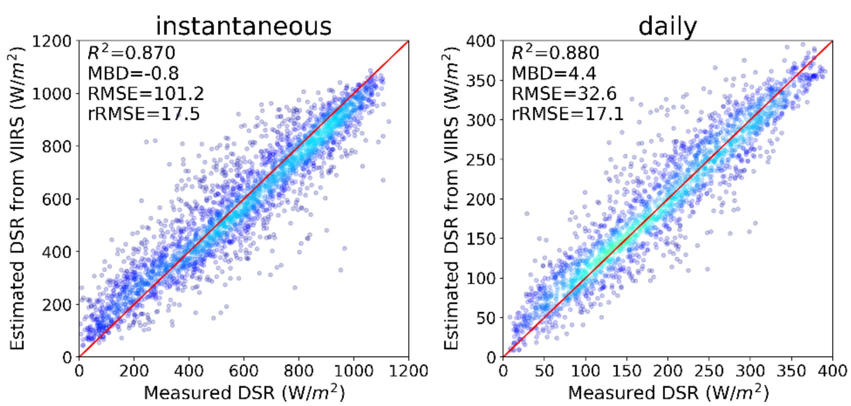

Figure 3-10. Validation results of instantaneous and daily BM_TL over SURFRAD stations in 2014

To evaluate the intra-annual stability of the proposed method, the rRMSE and rBias were calculated for each month over all validation sites. These results are shown in Figure 3-11. Similar to previous analysis (Li et al., 2021), comparatively higher rRMSE were observed from November to March, and the lowest rRMSE was achieved from May to August. The difference between the highest and lowest rRMSE of each month was about 7%. In terms of rBias, the model tended to underestimate DSR from November to February, and overestimate DSR from August to October.

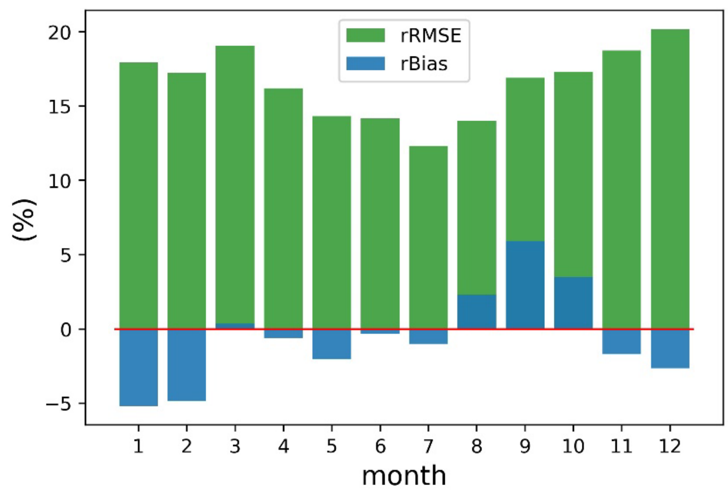

Figure 3-11. The rRMSE and rBias of estimated DSR for each month.

Previous analysis also suggests that the higher the latitude, the more obvious seasonal trend was in current DSR products (Li et al., 2021). To evaluate the temporal stability of the proposed method over high latitude areas, the daily variation of DSR over eight GCNET stations in 2013 are plotted in Figure 3-12. Consistencies are observed between measured and estimated DSR. The variability and magnitude of DSR are captured in most cases. By adopting satellite observations, the proposed method can also fill the gaps of ground measurements. For example, no measurements are available at Humboldt stations before May but continual estimations can be provided by the proposed method. The accurate retrieval of DSR over polar regions was challenging and limited even for ground measurements due to the severe weather;

hence, the stable and accurate estimation of TL over high latitude areas will be of great importance in future study to understand polar climate.

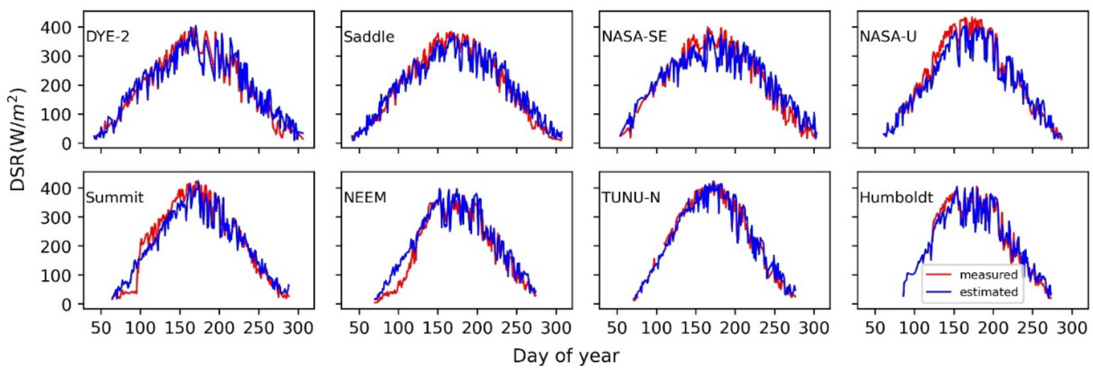

Figure 3-12. Daily variation of estimated and measured DSR over eight GCNET stations in 2013.

3.4.6 DSR mapping

BM_TL was also applied to satellite imagery to demonstrate global effectiveness. Three VIIRS swaths over Greenland, Ural, and Middle East regions were chosen, where few stations are available to provide training data and where traditional DSR estimation methods degrade. A random day, June 6th, 2013, was chosen for both instantaneous and daily mapping.

Figure 3-13 presents the estimated instantaneous DSR mapping. VIIRS red–green–blue (RGB) images are plotted as references for instantaneous estimation because the existence of clouds always corresponds to low DSR. To mask the cloud cover over snowy areas on RGB images, CLDMSK_L2_VIIRS_SNPP data were also used. It shows that the BM_TL based DSR estimations presented similar spatial patterns with cloud over all three regions. BM_TL was also able to distinguish different cloud phase to estimate DSR values properly. For example, higher DSR were estimated over periphery areas than the middle area of the clouds in Middle East. The BM_TL model also performed well over snow-covered regions. The cloud mask product reveaed the presence of cloud in the middle of Greenland, which is hard to observe visually in the VIIRS RGB image. The proposed model successfully discriminated clouds from snow and estimated DSR properly. This agrees well with previous validation results over Greenland regions. Figure 3-14 compares the estimated daily DSR mapping and corresponding CERES-SYN results over the same three areas on June 6, 2013. Similar spatial patterns and color scale were presented between the daily DSR estimated by BM_TL and that retrieved from CERES-SYN. Moreover, owing to the high spatial resolution of VIIRS, more detailed information could be observed from the BM_TL maps. However, some differences persist, such as those over the west part of Greenland and those southwest of the Middle East.

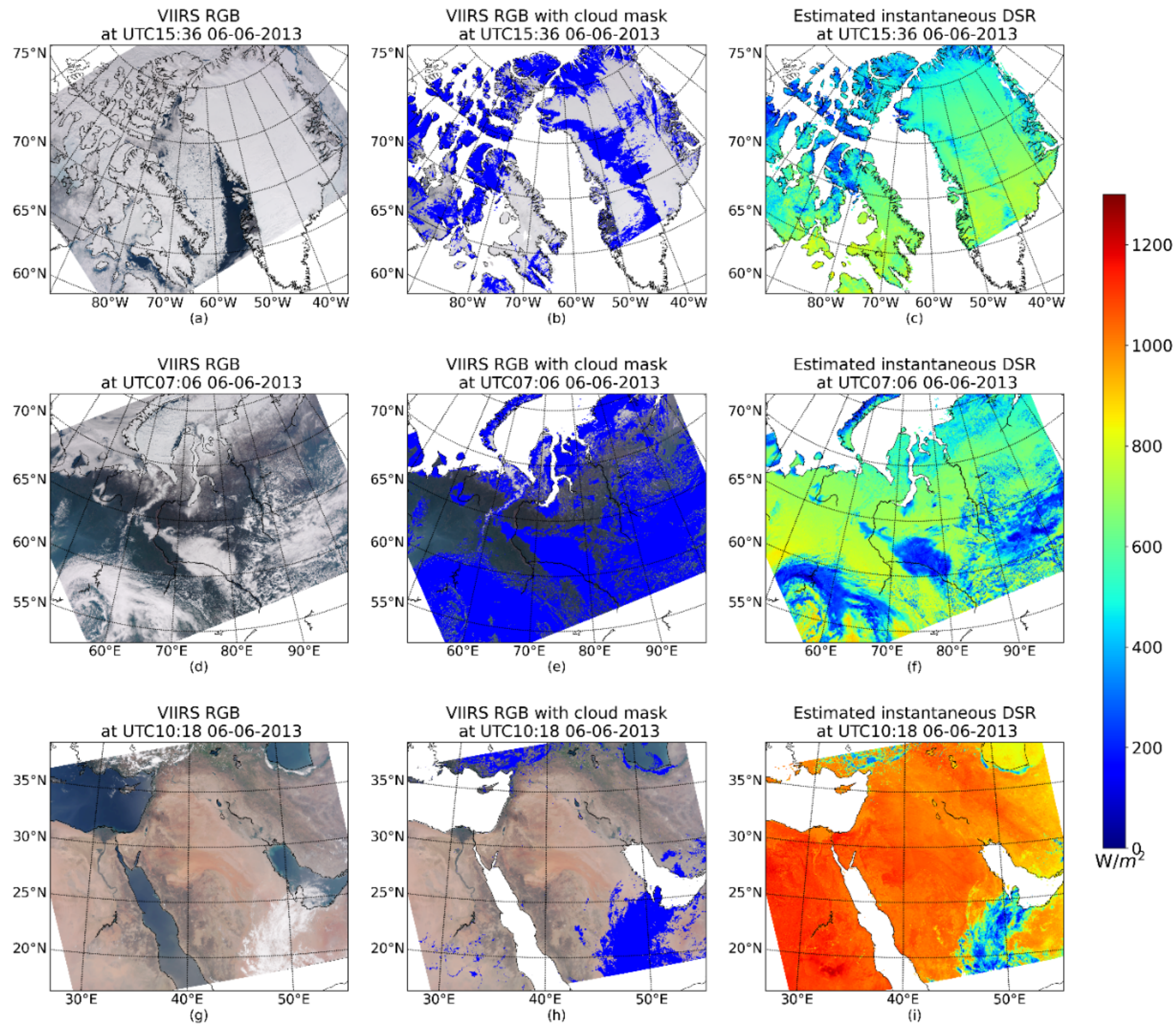

Figure 3-13. Estimated instantaneous DSR and VIIRS RGB images and those with cloud-cover masks over Greenland, Ural, and Middle East areas on June 6th, 2013.

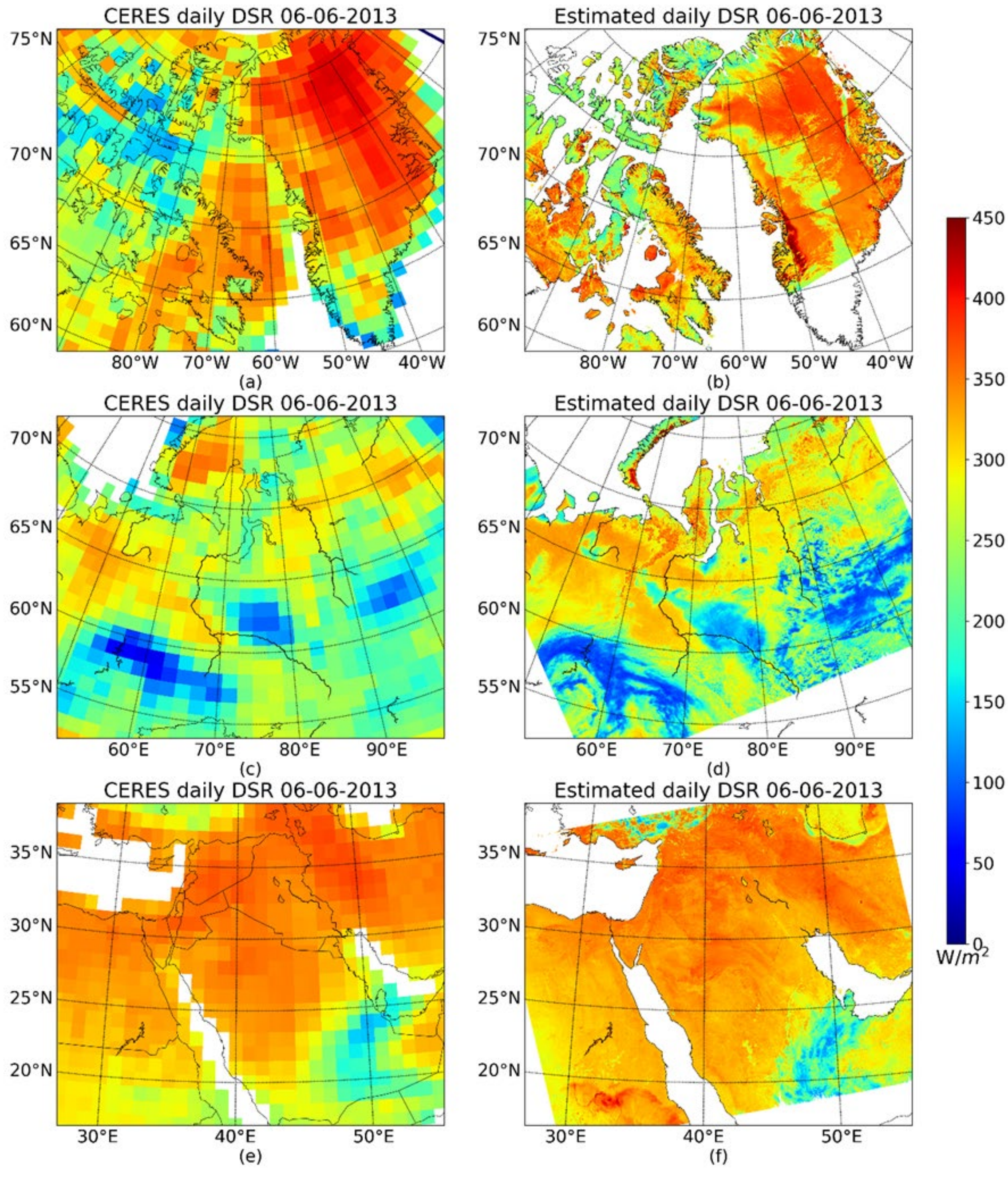

Figure 3-14. Estimated daily DSR and comparison with CERES-SYN daily DSR over Greenland, Ural, and Middle East areas on June 6th, 2013.

3.4.7 Impact of the training data size

As TL was first introduced to combine RTM simulated data and in-situ data, more analysis was performed to show how the size of training data influence the TL performance compared with traditional NN. We used instantaneous estimations as illustration here.

Six random subsets of the full data are employed by using a different number of stations (5, 10, 30, 50, 70, and 96) to train the BM_SA and BM_TL models. For each number, the stations were randomly chosen from all FLUXNET stations ten times to minimize the influence of training data representativeness. The models were tested over SURFRAD and GCNET stations. The same model structure and configuration were retained, except for the frozen layers and epoch number, which were justified separately for each training dataset. The results are presented in Figure 3-15 and Table 3-4. Several conclusions can be drawn. First, a lower mean RMSE was observed for BM_TL, especially when the training data size was smaller than 50 stations. For SURFRAD stations, the rRMSE of BM_TL remains 17%-19% regardless of the training size while rRMSE of BM_SA ranged from 19%-28%. The difference was more obvious when testing over GCNET stations; the rRMSE of BM_SA reached 71.7% while BM_TL achieved nearly half of BM_SA's, at 36.6%, for GCNET sites when there was 5 training stations. Moreover, lower standard deviations (STDs) by 10 times and less outliers were observed for BM_TL. Over both SURFRAD and GCNET stations, almost 0 W/m^2 STD was observed for BM_TL when trained from all data. When training from small sample size, BM_TL still has a low STD comparing with BM_SA. The STD was lower than 6 W/m^2 for BM_TL, while it reached 100 W/m^2 for BM_SA over SURFRAD stations, and was lower than 34 W/m^2 for BM_TL while it reachef 62.6 W/m^2 for BM_SA over GCNET stations. One possible explanation is that, unlike BM_SA, BM_TL will not be influenced by the random initial weights. Moreover, when randomly selecting data from the training data pool, it is possible that some stations are more representative and have better measurement quality. The results suggest that BM_TL is less sensitive to the representativeness of training data. It is also noticeable that the difference between BM_TL and BM_SA diminishes as training data size increases, especially over SURFRAD stations. This is possible because, when including all of the training data from FLUXNET, the data has already covered sufficient conditions for the testing sites. The comparatively simple structure of the proposed method may also be another reason.

In summary, TL has superior results when facing limited training data compared with traditional NN. These superior performances include the improvement of the accuracy as well as the stability of the model. The advantages of TL are less pronounced when the training data size increases, but it still performs better over under-representative conditions, such as high latitude areas. These results indicate that with the knowledge learned from simulation data, TL can achieve increased performances even when the size of the training data is small. This characteristic will potentially help estimate more peculiar geographical variables when ground-measurement data are limited.

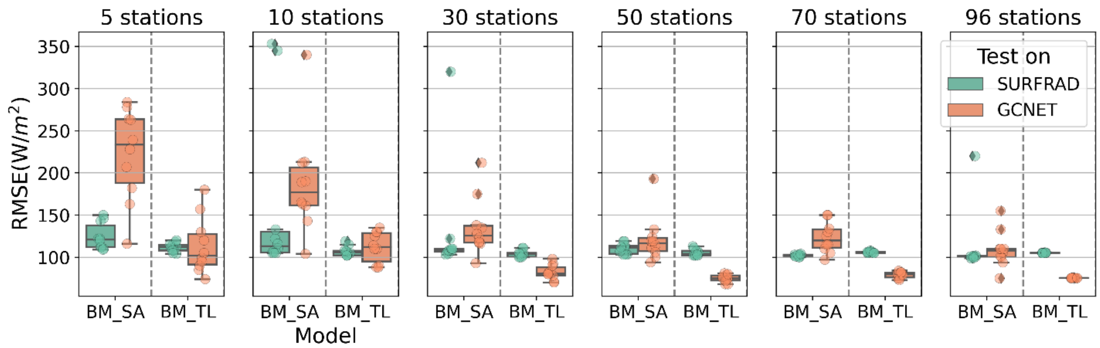

Figure 3-15. Comparison between BM_TL, BM_SA models trained on 5, 10, 30, 50, 70, and all (96) FLUXNET stations.

Table 3-4. Evaluation of BM_SA and BM_TL trained on 10, 30, and 50 FLUXNET stations ten times over SURFRAD and GC-NET stations.

N	Model	SURFRAD RMSE (W/m^2)	SURFRAD rRMSE (%)	SURFRAD STD (W/m^2)	GCNET RMSE (W/m^2)	GCNET rRMSE (%)	GCNET STD (W/m^2)
5	BM_SA	124.6	21.5	15.8	222.4	71.7	55.1
	BM_TL	111.7	19.2	5.4	113.5	36.6	33.6
10	BM_SA	160.2	27.6	100.0	188.0	60.6	62.6
	BM_TL	107.0	18.4	5.8	112.0	36.1	18.2
30	BM_SA	129.9	22.4	67.0	135.5	43.7	34.1
	BM_TL	104.2	17.9	4.1	82.2	26.5	8.9
50	BM_SA	110.3	19.0	6.2	121.3	39.1	27.7
	BM_TL	105.5	18.2	4.1	74.9	24.1	4.6
70	BM_SA	101.9	17.5	1.4	123.1	39.7	18.0
	BM_TL	105.8	18.2	0.9	79.0	25.5	3.6
96	BM_SA	112.6	19.4	37.8	109.4	35.3	21.7
	BM_TL	105.1	18.1	0.3	75.4	24.3	0.2

3.5 Conclusion

This study proposed an NN combined with TL to estimate global instantaneous and daily DSR from VIIRS TOA reflectance sources. A four-layer NN model was constructed and first trained on RTM simulated data. It transfers the knowledge learned from the simulated data to the real data by freezing one layer of the model and re-training it on FLUXNET measurements. The results were independently validated over seven SURFRAD stations, 25 BSRN stations, and eight GC-NET stations. For instantaneous estimation, the RMSEs were 106.3 (18.3%), 91.2 (16.1%), 75.0 (24.2%) W/m^2, respectively. For daily estimation, the RMSEs were 30.8 (15.5%), 33.5 (17.6%), and 31.3 (14.4) W/m^2, respectively. The results were also compared with traditional ML algorithms and physical algorithms, and presented significant improvements, especially for GC-NET data. The proposed method was also compared with existing DSR products and showed comparable accuracies over BSRN and SURFRAD stations and better results over GC-NET stations.

There are several advantages of the proposed algorithm. It only requires TOA and surface albedo as inputs, which avoids uncertainties inherently brought by other high-level products and enables retrievals at VIIRS raw resolutions (~ 750 m). TL is a bridge that combines the strengths of models trained by physical simulations and real-word data. Generality and accuracy were both improved, which enables a global application. The method was applied on three VIIRS swaths where no training data were collected. The instantaneous mapping captured the spatial pattern of the clouds, and the daily estimation corresponded well to the daily results of CERES-SYN. BM_TL also presents high accuracies over polar regions where some existing methods degrade. Lastly, BM_TL revealed the potential of training a model that mitigates the restrictions of training data size. BM_TL significantly improved the accuracy when limited stations were used as training data. For geographical science studies where both physical models and ground measurements exist, TL revealed unique values when combining model simulations with real-world data.

The current algorithm still has room for further improvement. For example, the model did not consider topography as an explicit input which has been noted by previous studies that influences DSR estimation dramatically through shadowing and multi-scattering (Hao et al., 2018; Proy et al., 1989). Due to the data limitation, BSA at noon rather than the blue sky albedo is used as input in this study. Nevertheless, invalid estimations still exist in the proposed method due to the lack of different surface albedo bands as inputs. Future studies can incorporate blue sky albedo and climatology albedo data or train the model with less albedo bands to address the issues. In this study, a simple four-layers NN was employed with relatively sufficient training data. The potential of TL may not be fully leveraged with all available training data as discussed in Section 3.4.7. The combination of TL with a deeper and more sophisticated NN model can potentially deliver further improved performance.

Chapter 4: Comparison between deep learning architectures for the 1km, 10/15-min estimation of downward shortwave radiation from AHI and ABI

This chapter aims to utilize the spatial and temporal structure of DL models to estimate DSR at high spatial and temporal resolutions. Unlike the problem encountered in the chapter 3, geostationary satellite provide comparatively enough and homogeneous observations for training large models. The challenges are how to estimate near-real-time DSR for future solar radiation forecasting and deal with the 3D cloud impacts at fine scales. This chapter compares three DL methods: DenseNet to extract spatial information, BiGRU to extract temporal information, and CNNGRU to extract both spatial and temporal information. These methods were employed to estimate DSR at 1 km and 10/15 min resolutions using top-of-atmosphere reflectance data from the Advanced Himawari Imager (AHI) on Himawari-8 and the Advanced Baseline Imager (ABI) on GOES-16. The results showcased high precision, with instantaneous root mean square errors (RMSEs) for the models being 68.4 (16.1%), 69.4 (16.3%), and 67.1 (15.7%) W/m^2. These values were superior to the baseline machine learning model, the multilayer perceptron (MLP), which had an RMSE of 76.8 W/m^2 (18.0%). Hourly accuracies for the DL models ranged between 57.3 (13.8%) to 58.6 (14.1%) W/m^2, aligning well with RMSE values from established datasets such as the Earth's Radiant Energy System (CERES) and GeoNEX. This chapter highlights the potential of DL models with temporal components to obviate the need for surface albedo input. This is paramount for real-time DSR monitoring and nowcasting. Furthermore, the inclusion of spatial data can enhance accuracy during cloudy conditions, while adding infrared bands can bolster DSR estimation accuracy (Li et al., 2023b). This study has been published (Li et al., 2023b).

4.1 Introduction

Recent years have witnessed increasing demand for high spatiotemporal surface DSR with high accuracies (Liang et al., 2010; Stephens et al., 2012; Liang et al., 2019). As DSR is the main input for surface ecological, hydrological, and biogeochemical models, high spatiotemporal resolution is a prerequisite for estimating many surface variables at a fine scale, such as hourly ground-level ozone (Wei et al., 2022), land surface temperature (Jia et al., 2022b; Jia et al., 2023), and evapotranspiration (Huang et al., 2019). DSR with high spatiotemporal resolution is also critical for monitoring PV output and the economic and stable operation of the solar grid (Buster et al., 2021). Short-term fluctuations in the DSR can adversely affect power grid quality (Gandhi et al., 2020). Power system models, such as the Global Solar Energy Estimator (Pfenninger and Staffell, 2016) and the Integrated Grid Modeling System (Palmintier et al., 2017), require hourly DSR data as input. Analysis of storage requirements for large grid-connected PV plants requires sub-hour or even minute-by-minute DSR data (Marcos et al., 2014; Nguyen et al., 2016). The World Meteorological Organization (WMO) suggests that a 10-min temporal resolution would be a break-through for DSR nowcasting and very short-range forecasting (Huang et al., 2019). In addition, coarse spatial resolution DSR cannot accommodate the current trend of extensive small-distributed PV deployment, particularly at the residential scale (Jain et al., 2017). Reportedly, issues of spatial heterogeneity for

DSR exist on an hourly scale and are more pronounced in mountainous regions as the spatial scale increases (Li et al., 2023). High accuracy is also demanded by DSR nowcasting (Huang et al., 2019). However, achieving accurate retrievals at high temporal and spatial resolutions remains a challenging topic.

Accurately estimating DSR with high spatiotemporal resolution currently presents various challenges and limitations. Existing DSR retrieval methods rely on high-level satellite-derived products that require complex processing algorithms for data extraction from raw satellite observations (Zhang et al., 2020). Although parameterization of atmospheric radiative transfer is an accurate and stable method, its performance still heavily relies on the choice of satellite-derived products as inputs, limiting its adaptability to high spatial and temporal resolution sensors (Van Laake and Sanchez-Azofeifa, 2004; Huang et al., 2018; Tang et al., 2019; Wang et al., 2021). The LUTs approach pose challenges owing to their reliance on various satellite-derived products. Zhang et al. (2021) employed MODIS aerosol optical depth (AOD), albedo, and GOES-16 Clear Sky Mask products to the LUT to estimate hourly photosynthetically activated radiation (PAR). Li et al. (2015) used MODIS AOD, water vapor, ozone, surface albedo, and snow cover products to estimate PAR over China. Letu et al. (2022) also required a serious of Himawari8 atmosphere products and ERA5 reanalysis data when estimating surface radiation from the Advanced Himawari Imager (AHI). The production of GeoNEX is comparatively less dependent on the high-level products, but still requires MODIS daily surface albedo product (MCD43) as inputs (Li et al., 2023). Some ML methods also estimate DSR forwardly similar to parameterization method and have been applied to generate DSR products. Hao et al. (2019) included MCD43 and products from a series of the Earth Polychromatic Imaging Camera (EPIC) onboard the Deep Space Climate Observatory (DSCOVR) to generate global hourly DSR and PAR products using random forest. Ryu et al. (2018) produced Breathing Earth System Simulator DSR using a neural network based on a series of MODIS products. The resolutions of high-level products restrict the retrieval of DSR with high spatiotemporal resolutions and prolong the latency for in time DSR estimation and nowcasting. With the increasing analyses on current solar energy supply and demand (Tong et al., 2021), storage (Gandhi et al., 2020), and solar power prediction (Bansal et al., 2022), developing methods that rely less on input data and provide more timely information on DSR is critical.

Cloud complexity has been identified as another factor constraining the enhancement of DSR estimation accuracy (Huang et al., 2019). Current products and retrieval algorithms have primarily been developed from one-dimensional RT theory, which is based on the assumption of independent pixel approximation (IPA) owing to its relatively high efficiency. It ignores the three-dimensional (3D) effect of clouds including the nonlocal cloud shadows and photon diffusion from neighboring columns, which works well for large pixel retrieval but is inadequate for high-resolution products (Chen et al., 2019). Wang et al. (2017) demonstrated that shadowing effect brought by the solar-cloud-satellite geometry may lead to up to 80% errors in DSR estimation. The errors are more prominent at high temporal resolutions, such as instantaneous estimation, and indirectly influence current global daily high-spatial resolution products extrapolated from instantaneous estimations. With the increasing demand for DSR with high spatial and temporal resolution, the complex effect of clouds on DSR retrieval can no longer be ignored.

Deep learning (DL) can extract spatial and temporal features on top of spectral information from satellite data by incorporating specific structures, such as convolutional layers or forget gates, presenting a potential solution to address the challenges posed by cloud heterogeneity and the reliance on high-level products when estimating DSR. The use of DL in DSR is still at its infancy. Few studies have utilized DL for estimating DSR particularly with both spatial and temporal structures, while most have focused on a single DL model for accuracy improvement. Jiang et al. (2019) employed residual convolutional neural network (CNN) models to estimate DSR over mainland China. Zhang et al. (2022) used a spectral-wise CNN to extract the relationship between different bands for DSR estimation from the Visible Infrared Imager-Radiometer Suite. Although certain studies have applied recurrent neural network (RNN) to solar energy forecasting (Sønderby et al., 2020; Bansal et al., 2022), their estimation capability has not been investigated.

The current study developed and compared three DL methods that accurately estimate 1km DSR from AHI onboard Himawari-8 with 10-min frequency and the Advanced Baseline Imager (ABI) onboard GOES-16 with 15-min frequency. The models in this study were designed with minimal inputs. The DenseNET model requires both TOA reflectance and surface albedo as inputs, whereas the bidirectional gated recurrent unit (BiGRU) and CNN with gated recurrent unit (CNNGRU) models only require TOA reflectance, henceforth referred as BiGRU_norand CNNGRU_nor models, respectively. These three DL models with spatial, temporal, or combined structures were compared and analyzed for their accuracy under different temporal resolutions, model efficiency, and ability to address issues related to dependency on high-level products and cloud effects. The remainder of this manuscript is organized as follows: Section 4.2 describes the input and ground measurement data; Section 4.3 elaborates on the data preprocessing methods, model structures, and validation methods; Section 4.4 discusses the overall validation results and analysis; Section 4.5 summarizes the study conclusions.

4.2 Data

4.2.1 Input data

The input variables and the corresponding data source are listed in Table 4-1. The gridded TOA reflectance was retrieved from GOES16/HIMAWARI8 Level 1G archived in the NASA GeoNEX data portal (Wang W. et al., 2020). The 15 bands shared by ABI onboard GOES-16 and AHI onboard Himawari-8, ranging from visible to infrared, were included in this study. ABI and AHI sensors cover areas ranging from 60°N, 138°W to 60°S, 18°W and 60°N, 78°E to 60°S, 198°E, respectively (Figure 4-1). The study also includes the geometric angles of each scan, namely solar zenith angles (SZA), viewing zenith angles (VZA), and relative azimuth angles (RAA), as inputs. The Global 30 Arc-Second Elevation (GTOPO30) dataset provided the static surface elevation data (EROS, 2017). The surface albedo information used in this study is carried by the black sky albedo (BSA) of MODIS bands 1-7 (MCD43). The blanks left by MCD43 are filled with the surface albedo climatology data which are generated by Jia et al. (2022a) through averaging the available surface albedo from 2001 to 2020 for the existing MODIS products. Surface albedo and BSA are used synonymously in this study.

Table 4-1. The Summary of input variables and corresponding datasets.

Variables	Dataset	Dimension	Spatial resolution	Temporal resolution	References
TOA reflectance	GOES16/ HIMAWARI8 Level 1G	15 (bands)	1km	10min or 15min	(Wang W. et al., 2020)
Geometry angles	GOES16/ HIMAWARI8 Level 1G	3 (angles)	1km	10min or 15min	(Wang W. et al., 2020)
Elevation	GTOPO30	1	30 arc seconds	Static	(EROS, 2017)
Surface albedo	MCD43A3	7 (bands)	500m	Daily	(Schaaf and Wang, 2015)
Surface albedo (climatology)	Blue-sky Actual and Snow-free Albedo Climatology	7 (bands)	1km	Daily	(Jia et al., 2022a)

4.2.2 Ground measurement data

Ground measurement data cover 110 sites measured by four networks: AMERIFLUX, BSRN, FLUXNET, and SURFRAD, in 2018 (Figure 4-1). All data used as training and testing samples first passed the BSRN quality check (Roesch et al., 2011). The data were subjected to a quality check following the procedure described by Li et al. (2021). BSRN and SURFRAD provide measurements at 1 min intervals, whereas FLUXNET and AMERIFLUX measure DSR at 30 min intervals. To derive FLUXNET and AMERIFLUX measurements at every 10/15 min, we first generated the per minute measurements by downscaling through forward filling. For example, the 8:01 to 8:30 am data were all filled with 8:30am measurements. The final in-situ measurements of all four networks used as the true value of training or validating were averaged from the 15 min measurements centered at the satellite passing time to eliminate the representative errors of the point-scale DSR measurement (Huang et al., 2016). The hourly measurements were directly averaged, and the daily aggregation DSR were calculated as per procedure stated by Li et al. (2021).

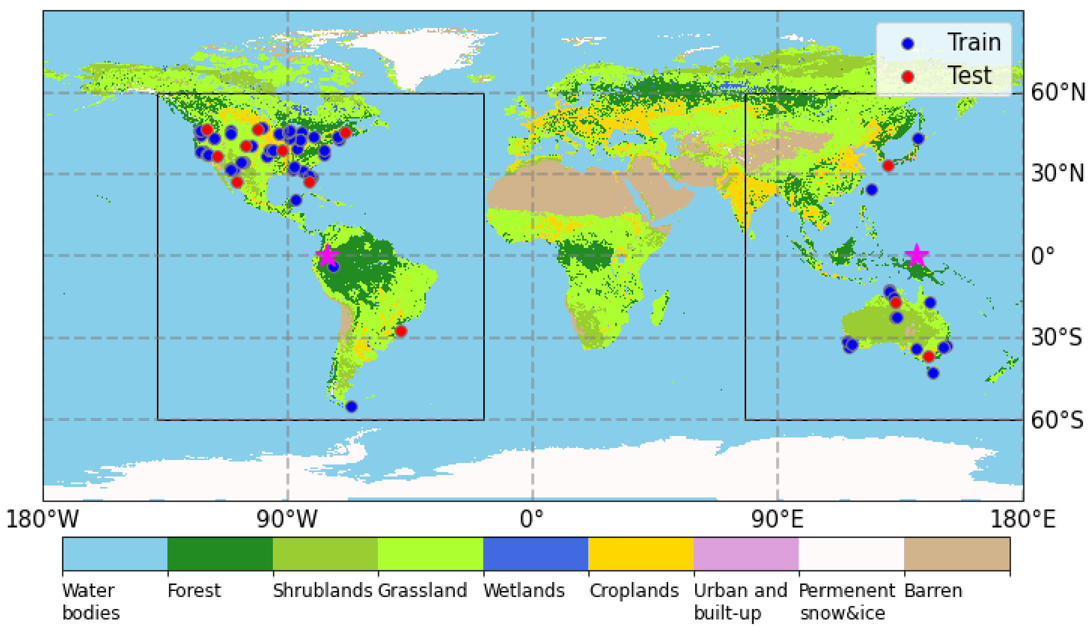

Figure 4-1. Distribution of ground measurement stations. Training sites are labeled in blue and testing sites are labeled in red. The MCD12 International Geosphere-Biosphere Programme (IGBP) maps are used as the background.

4.2.3 Ancillary data sets

In this study, we compare our DSR estimations with hourly CERES SYN1deg product (CERES). Owing to its high accuracy and global coverage, the validation accuracy of CERES has been used as the baseline result in many studies (Riihelä et al., 2017; Zhang et al., 2020; Li et al., 2022). GeoNEX DSR is a newly developed product with the highest spatiotemporal resolution among the existing DSR products, with spatial and temporal resolutions of 1 hour and 1 km (Li et al., 2023). It employs the LUT approach on the ABI and AHI TOA bands and aggregates the instantaneous results into an hourly scale. It demonstrates greater accuracy than CERES at both the hourly and daily scales and is regarded to have the highest accuracy in retrieving DSR from satellites (Li et al., 2023).

In addition, the NOAA GOES-R Series Level 2 clear sky mask and cloud top height product were used to analyze the performance of DSR retrievals under different cloud conditions. Full disk versions of the two products were generated at the same temporal resolution as the ABI observations. The clear sky mask was retrieved at a horizontal spatial resolution of 2km, and cloud top height of 10 km. Both were matched with the ground measurement after conversion to the geographic coordinate system. The products were only available for ABI coverage.

4.3 Methods

4.3.1 Data preprocessing and evaluation methods

The inputs and outputs for each model are listed in Table 4-2. All of the input variables were matched with the geographic locations of the in-situ sites to generate all the samples. Measurements and satellite observations in 2018 for each 10/15 min over 110 sites were used in this study. The total sample size was approximately 1.2 million. The extensive ground measurement data used in this study also ensured that samples covered as many different land cover types as possible (Figure 4-1). The broadband BSA ranged from 0 to 0.8, and the elevation of the sites ranged from 0 to 3000 m.

Carefully splitting training and testing data are critical for testing and ensuring the model generality. Samples from geospatially independent sites were used as training and testing data since surface-reflecting properties play a critical role in estimating DSR. We manually selected 12 sites as testing sites that were geospatially independent from the training sites and covered different surface types (Figure 4-1). Prior to feeding the data into any of the models, the input variables were normalized by the mean and standard deviation calculated over the entire training sites. The models were trained on 80% of the training sites data and validated on the remaining 20% to determine their configurations.

Two evaluation methods based on sites were employed in this study. The models were firstly tested on the 12 sites that were completely left out during the model tuning process. The second evaluation metric comprised 10-fold cross-validation across sites. All the sites were randomly divided into ten groups. The selected group of sites served as validation data, and the model was trained on the remaining data. This process was repeated for each of the ten groups. The mean of the ten iteration scores was used as the final accuracy.

The evaluation metrics for the validation and comparison of the models includes R^2, mean bias difference (MBD), root mean square error (RMSE), relative root-mean-square error (rRMSE), and ratio of variance (ROV) (Tramontana et al., 2016). Among them, RMSE and rRMSE are the most widely used metrics of DSR estimation studies; hence, they were primarily used to compare the models among each other and with existing products. The standard deviation of RMSE among the 10-fold validation (STD) was also included to quantify the uncertainties of each model. The evaluation metrics includes R^2, MBD, RMSE, rRMSE. Additionally, we calculate ratio of variance (ROV)

$$ROV = \frac{(\sigma R_g)^2}{(\sigma R_p)^2} \tag{5}$$

where σR_p represents the standard deviation of the predicted values for the testing samples, σR_g represents the standard deviation of the ground measurements for the testing samples.

4.3.2 Model construction

The study used an NVIDIA Tesla V100 GPU and the TensorFlow DL framework in Python for implementation. Detail configuration of each model can be found at https://github.com/Ruohan-Li/DeepLearningDSREstimation.git (https://doi.org/10.5281/zenodo.7926773).

4.3.2.1 MLP

MLP denotes a fully connected feedforward neural network (NN), one of the most extensively used NN models. Over the last few years, it has been used to estimate DSR (Takenaka et al., 2011; Ryu et al., 2018; Brown et al., 2020; Peng et al., 2020). A four-layer MLP model was developed as the baseline model based on previous studies where it achieved high accuracy (Li et al., 2022). The input layer included 26 neurons. The first hidden layer had 25 neurons, followed by another hidden layer of five neurons. One neuron in the output layer corresponded to the DSR value of linear activation. All layers were activated by the rectified linear unit (ReLU) activation function (Nair and Hinton, 2010). The model used Root Mean Square Propagation (RMSprop) optimizer and mean squared error (MSE) loss function.

4.3.2.2 DenseNET

The structure of the convolutional and pooling layers in CNN enables learning the spatial relationships from the input data with fewer trainable parameters. CNNs have been used for cloud classification from TOA, demonstrating their ability to detect cloud spatial structure, which is critical for improving DSR estimation (Chai et al., 2019; Jappesen et al., 2019). Many deep CNN models such as LeNet-5 (LeCun et al., 1998), AlexNet (Krizhevsky et al., 2017), and VGG-16 (Simonyan and Zisserman, 2014) have been developed and tested for high efficiency in image classification. As the input features include the dominant information of DSR estimation (Wang et al., 2021) and the spatial information are the complementary knowledge, a dense convolutional network (DenseNET) (Huang et al., 2017) was employed for our specific task to address the issue of vanishing or losing information (He et al., 2016). This allowed constantly concatenation of new features from all preceding features as new input layers, and ensured that the information from the inputs or the first few layers is passed to the end of the structure. These concatenation operations were encapsulated in dense blocks (Figure 4-2a). The developed model has the potential to achieve deep feature extraction with fewer variables and has a regularizing effect, which mitigates overfitting issues.

The structure of the DenseNET is shown in Figure 4-2b. The growth rate k was set to a small number, k = 4, owing to the large input features. The DenseNET comprised three dense blocks, each with three layers. The transition layers consisting of a 1 × 1 convolutional layer were followed by an average pooling layer employed to connected the dense blocks to reduce the dimensions. The outputs of the last pooling layer were flattened to form a fully connected layer; the output layer with one node estimated the final estimated DSR value. The RMSprop with learning rate of 0.001 and MSE loss function were used to train the model. Previous studies have suggested that the 3D cloud effect should be considered when the spatial scale is smaller than approximately 24 km (Chen et al., 2019). The highest accuracy of DSR estimation for current 1 km sensors is normally achieved when averaging over window size of 25 km (Li et al., 2021); hence, the input sample size of this study was set at 26 × 26 pixels, corresponding to 26 km × 26 km in reality.

4.3.2.3 $\mathrm{BiGRU_{nor}}$

The RNNs, which study the relationships between time-series data in DL, are adapted to eliminate the need for BSA as additional inputs by leveraging multiple observations made within a single day of geostationary sensors. Various RNNs have been developed that use temporal information from time-series data (Sherstinsky, 2020), including the simple RNN (Hopfield, 1982), long short-term memory network (Hochreiter and Schmidhuber, 1997), gated recurrent unit (GRU) (Cho et al., 2014), and RNNs equipped with bidirectional flows (Schuster and Paliwal, 1997). After comparing the above mentioned RNNs, bidirectional GRU (BiGRU) produced the best results with the highest efficiencies; hence, BiGRU was used in this study. Given that the overall variation in DSR in a day is asymmetrical, the BiGRU has the ability to take advantage of the former and the latter input values to adjust current DSR estimation. The BiGRU includes two GRU layers that are chained in opposite directions, enabling learning information flow from both sides (Figure 4-2c).

In this study, we applied a many-to-many BiGRU structure that takes observations from an entire day as input and simultaneously generates a DSR estimate at each time step in a day. The inputs of the $\mathrm{BiGRU_{nor}}$ model do not include BSA (Table 4-2); the structure of this model is shown in Figure 4-2c. All observations of one day from 6 am to 8pm were reconstructed as one sample and were aligned by their local time with a 10/15 min interval that corresponds to a time step. Invalid measurements or observations were filled with -1. The outputs of one sample correspond to the DSR estimations at each time step within a day. The model is not designed to be that deep because the diurnal variation of DSR are still heavily influenced by diurnal cloud conditions and not by the temporal correlation within one day. The model includes one BiGRU layer with 32 units each, two dense layers with 25 and 5 nodes each, and one output layer. The adaptive moment estimation (Adam) with learning rate of 0.001 optimization function and MSE loss function were employed for this model.

4.3.2.4 $\mathrm{CNNGRU_{nor}}$

To investigate whether the combination of temporal and spatial information can further improve the estimation accuracy, a combination of CNN and RNN models, namely the CNNGRU structure, was developed as part of this study owing to the unique temporal and spatial characteristics brought by satellite observations on DSR estimations (Figure 4-2d). As the model did not include BSA as inputs, it is referred as $\mathrm{CNNGRU_{nor}}$. The inputs of this model are listed in Table 4-2. It used a CNN for spatial feature extraction and a GRU for time sequence analyses. This type of model is currently in the early stages of predicting solar energy or precipitation (Sønderby et al., 2020; Bansal et al., 2022); however, to the best of our knowledge, it has not been used to estimate DSR directly. The input data were first fed into the CNN structure via three convolutional layers, followed by batch normalization (BN) and pooling layers. The flattened nodes were then fed into BiGRU structure, similar to the ones described above but comprising one dense layer including eight neurons. Finally, the model was made to produce the diurnal DSR simultaneously after digesting the spatial and temporal information of all spatial satellite images for one day. The Adam with learning rate of 0.001 optimization function and MSE loss function were employed for this model.

Table 4-2. The Inputs and outputs of the models compared in this study; the models-of-choice are highlighted in gray.

Model	**Input data dimension**	**Input data description**	**Feature description**	**Output data dimension**	**Output data description**
MLP	(26)	(Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
$\mathrm{MLP_{nor}}$	(19)	(Feature)	15 TOA bands, 3 angles, elevation	(1)	(DSR)
$\mathrm{MLP_{sw}}$	(16)	(Feature)	5 TOA shortwave bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
BiGRU	(60,26)	(Timestep, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)
$\mathrm{BiGRU_{nor}}$	(60,19)	(Timestep, Feature)	15 TOA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)

DenseNET	(26,26,26)	(Spatial width, Spatial height, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
$\mathrm{DenseNET}_{nor}$	(26,26,19)	(Spatial width, Spatial height, Feature)	15 TOA bands, 3 angles, elevation	(1)	(DSR)
CNNGRU	(60,26,26,26)	(Timestep, Spatial width, Spatial height, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)
CNNGRU_{nor}	(60,26,26,19)	(Timestep, Spatial width, Spatial height, Feature)	15 TOA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)

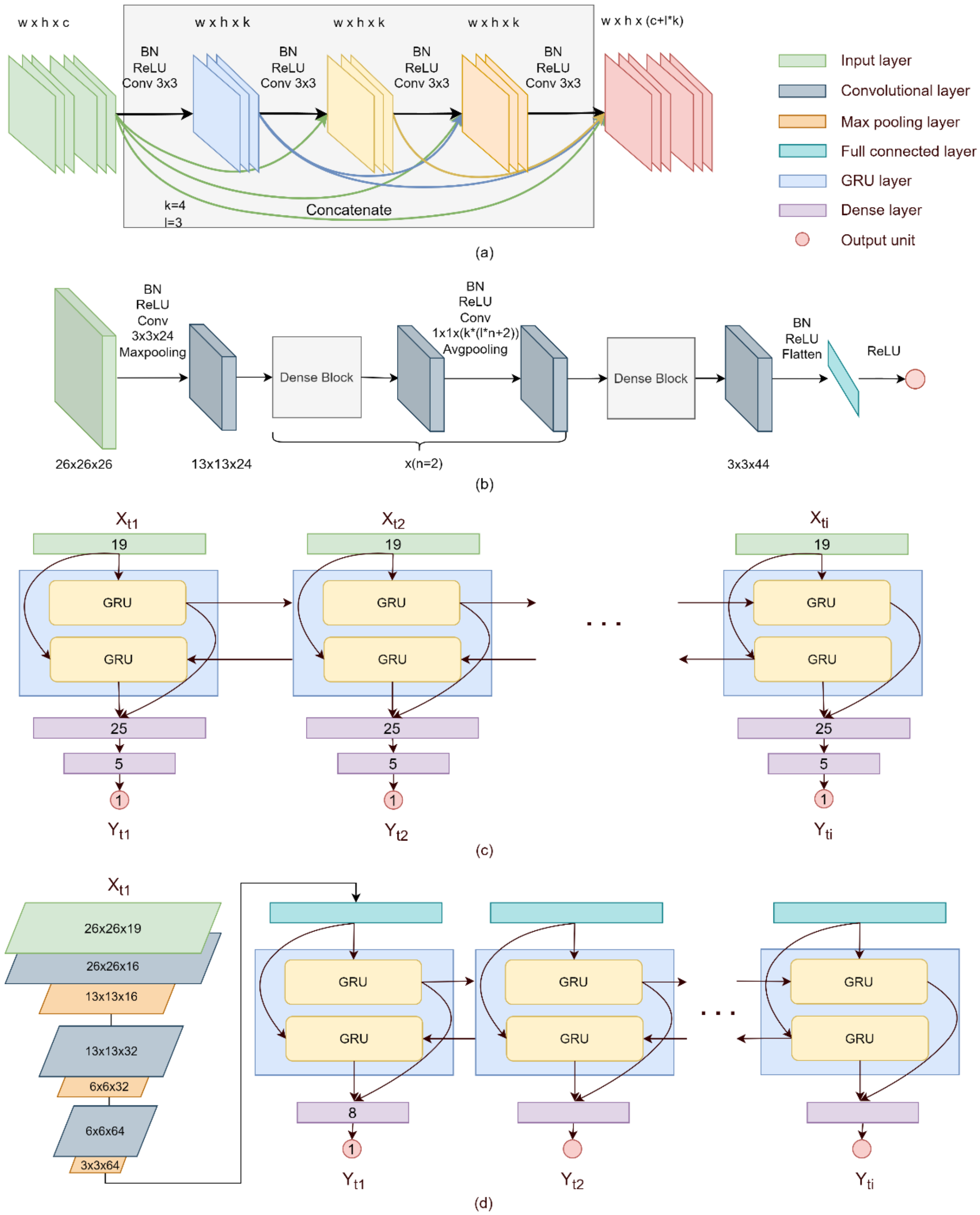

Figure 4-2.(a) Structure of one DenseNET (b) Structure of the DenseNET. (c)Structure of CNNGRU

4.4 Results

4.4.1 Comparison among the models and with the existing products

The 10-fold cross-validation results have been shown in Figure 4-3. All the models were compared over the same samples. CNNGRU_nor achieved the highest accuracies with an RMSE of 67.1 $W/m^2$and an rRMSE of 15.7%, followed by BiGRU_nor with an RMSE of 68.4 (16.1%) $W/m^2$and the DenseNET with an RMSE of 69.4 (16.3%) W/m^2. MLP shows the worst RMSE at 76.8 (18%) W/m^2 but the lowest MBD at 0.1 W/m^2. The ROV for all four models were higher than 0.9, indicating that the estimations have a similar level of variance as the ground truth. The STD of 10-fold validation results of the four models were similar, ranging from 6.9 to 7.8 W/m^2. Figure 4-4 shows the validation results on 12 independent sites. The overall accuracy was found to be similar to that of the 10-fold validation, which further demonstrated the generality of the proposed models. CNNGRU_nor still achieved the highest accuracy with an RMSE of 69.8 $W/m^2$and an rRMSE of 16.3%. However, it presented the largest MBD at -7.4 W/m^2. BiGRU_nor presented an RMSE of 70.1 (16.3%) W/m^2 and the DenseNET model showed an RMSE of 70.7 (16.5%) W/m^2. The MLP method showed the lowest accuracy, with an RMSE of 78.8 (18.4%) W/m^2. Compared to the baseline MLP model, the results demonstrated the ability of DL methods to achieve higher accuracy in retrieving DSR at 1km and 10/15 minutes with no overfitting issues. It is noticeable that both the BiGRU_nor and CNNGRU_nor models did not include surface BSA as inputs but achieved comparable accuracies.

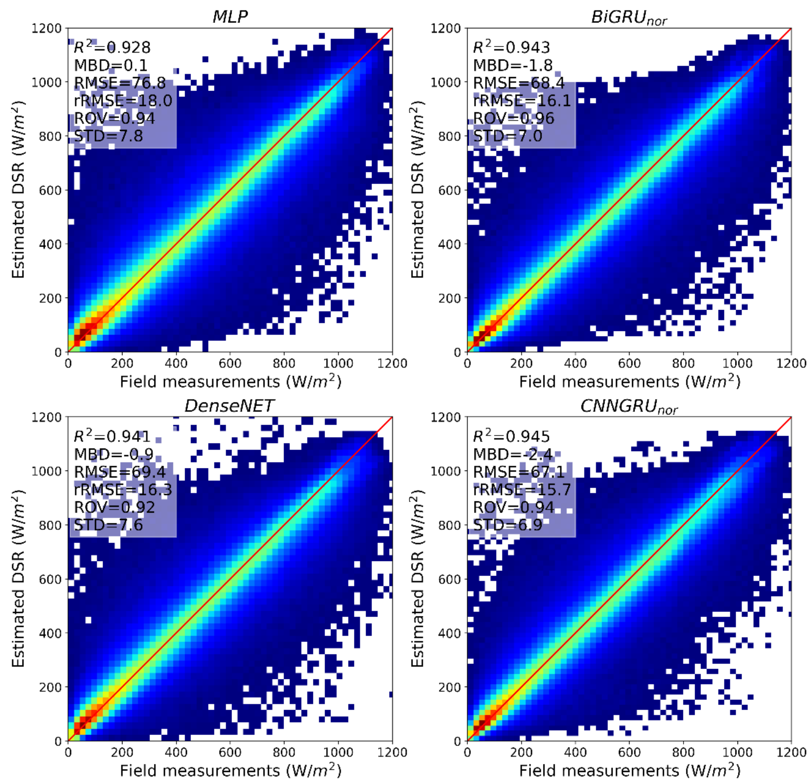

Figure 4-3. The 10-fold cross validation results of MLP , BiGRU_nor, DenseNET , and CNNGRU_nor models.

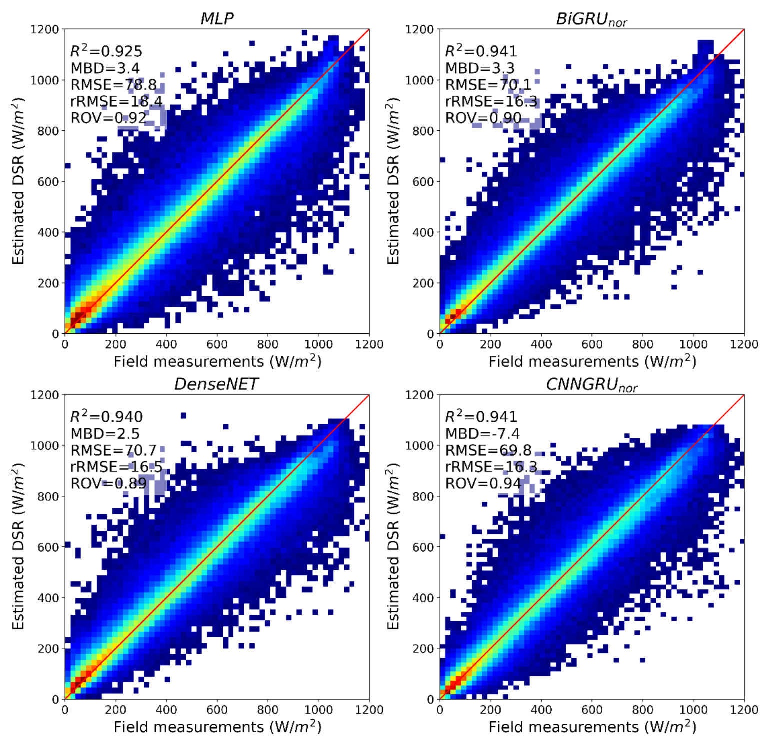

Figure 4-4. Validation results of MLP , BiGRU_nor, DenseNET , and CNNGRU_nor models over 12 independent sites.

Hourly validation results were compared with CERES and GeoNEX. CERES provides hourly DSR estimations with global coverage. Notably, the spatial resolution of CERES is 1 degree which is different from that of other products and methods included in this comparison. GeoNEX is a newly developed product based on AHI and ABI with 1 km and 1 hour spatiotemporal resolution. These two products present greater accuracy than existing DSR hourly products and hence were compared in this study (Li et al., 2023).

As shown in Figure 4-5, the 10/15 min retrievals of the proposed DL models were aggregated into hourly scales for comparison. CERES showed the highest RMSE at 88.8 (21.4%) W/m^2but with no bias. GeoNEX presents better accuracies than CERES, with an RMSE of 77.8 (18.8%) W/m^2 which aligned well with existing validations (Li et al., 2022). MLP shows higher accuracies than existing two products, with an RMSE of 63.1 (15.2%) W/m^2. The reasons for the high accuracy of the simple MLP method could be the sufficient training data and comparably small coverage areas in this specific geostationary sensor-based DSR retrieval. The other DL methods showed high accuracies with RMSE values of 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.9%) $W/m^2$for BiGRU_nor, DenseNET , and CNNGRU_nor, respectively. To the best of our knowledge, these accuracies are higher than those of most of the existing hourly products or operational results (Letu et al., 2022; Chen et al., 2021; Zhang et al., 2020). Notably, when compared with the instantaneous estimation shown in Figure 4-3, the hourly accuracies improved by approximately 10 $W/m^2$for all ML models. Among them, BiGRU_nor and CNNGRU_nor showed slightly smaller improvements, suggesting that the instantaneous estimation of the

GRU-related model incorporated some information from the temporal averaging results. As DSR normally exhibits pronounced diurnal cycles, we analyzed the intra-hour anomalies by employing the methodology described in Tramontana et al. (2016). These anomalies are determined as the deviation of an instantaneous value from the climatological averaged hourly value. To calculate the climatological DSR values for this study, we utilized the CERES 2017-2019 hourly data, extracting the mean hourly cycle for each site, each day, and each hour. Figure 4-6 demonstrates the successful reproduction of the intra-hour variation in DSR by all four models. Notably, the DL models exhibit higher R^2 values, approximately 0.8, surpassing that of the MLP model.

We further aggregated hourly DSR into daily scale for comparison (Figure 4-7). For DSR retrieval at daily scale, the superiority of DL was reduced, particularly the model with temporal structure. DenseNET showed the lowest RMSE at 15.1 W/m^2 with no bias. MLP, BiGRU_nor, and CNNGRU_nor show similar performance with RMSE values of 17.0 (8.2%), 16.1 (7.0%), and 16.0 (7.7%) W/m^2respectively. Nevertheless, they presented lower RMSE than CERES at 24.0 (11.5%) $W/m^2$with no bias and GeoNEX at 19.6 (9.4%) W/m^2

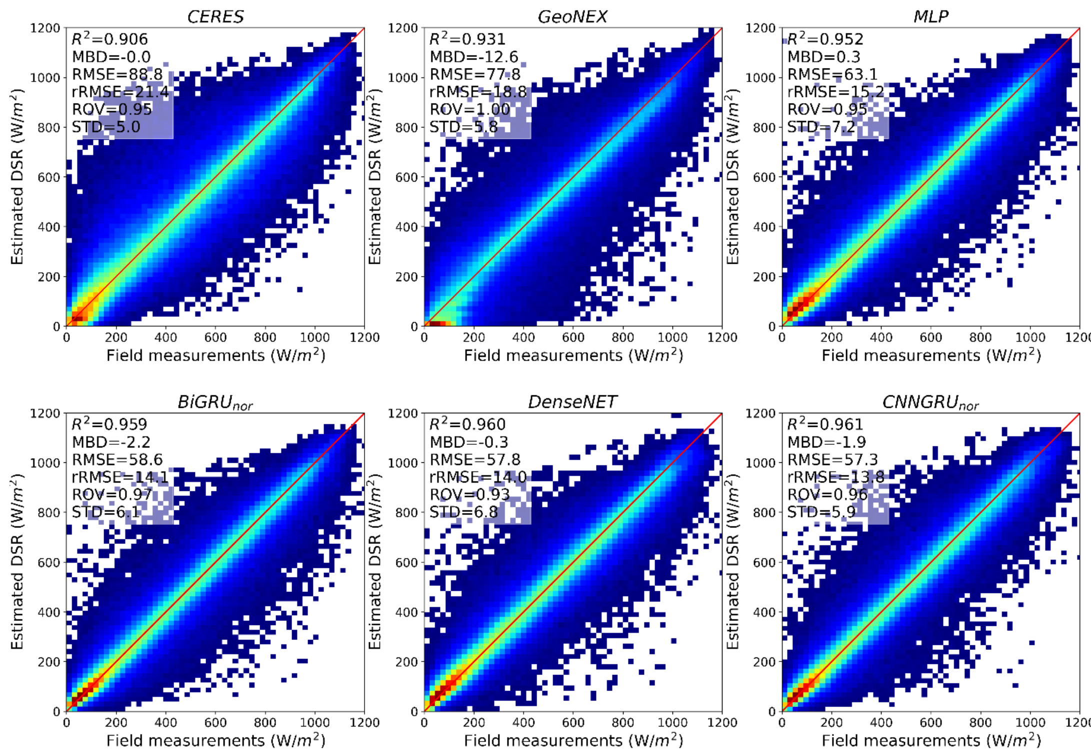

Figure 4-5. Comparison of the 10-fold validation results for MLP, BiGRU_nor, DenseNET, and CNNGRU_nor with CERES and GeoNEX at an hourly scale.

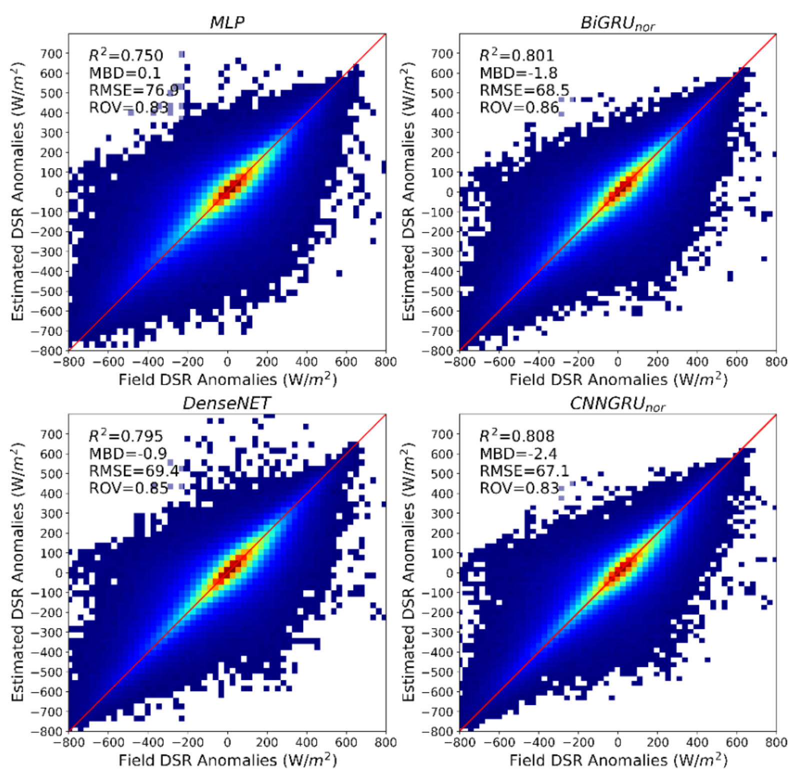

Figure 4-6. The anomalies of the 10-fold validation results for MLP, BiGRU_nor, DenseNET, and CNNGRU_nor at an hourly scale.

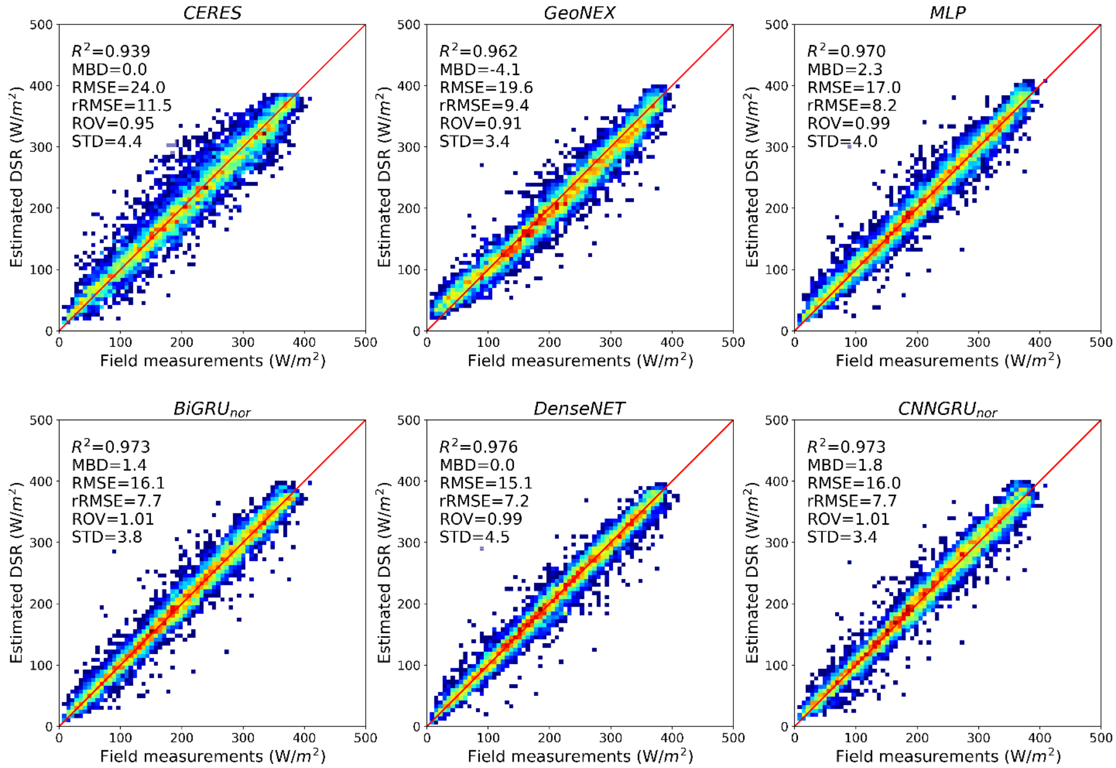

Figure 4-7. Comparison of the 10-fold validation results of MLP, BiGRU_nor, DenseNET, and CNNGRU_nor with CERES and GeoNEX at the daily scale.

Figure 4-8 shows the diurnal variations in the proposed methods at an hourly scale for comparison with the GeoNEX and CERES products. The results were plotted from day of year (DOY) 190 to 210 across six sites covering different land surface types. The results show that both GeoNEX and the proposed methods can capture the diurnal patterns in DSR variation, while CERES is more prone to overestimation under cloudy conditions. GeoNEX displayed more variations than in-situ measurements such as DOY 196 over US-Pnp sites and DOY 209 over US-SRG sites, which is also partly observed in MLP methods, likely owing to single pixel retrieval. DL-based methods performed better in these cases of overcast conditions and barren surfaces. Nevertheless, even with spatial information, DL methods lost some sudden changes in DSR like DOY 209 over Collie sites. Even when DL methods could capture the change, they tend to smooth the change.

Figure 4-9 shows the spatial distribution of the 10-fold validation rRMSE of the proposed methods, CERES and GeoNEX, over all sites. In general, the rRMSE of all sites improved with the proposed ML models. The results further demonstrated that similar patterns observed for both the proposed methods and existing products eliminated the overfitting issues. DL methods showed a more evenly distributed rRMSE than MLP. High rRMSE values in South America and over coastal regions were observed for all the models and products. One possible reason is that more diverse cloud conditions over coastal regions and less accurate or representative ground measurements may contribute to the low accuracy of all models and products. The VZA-related radial degradation that resulted in a high rRMSE at sites distant from AHI and ABI owing to high VZA in GeoNEX was also less pronounced in the DL-based methods.

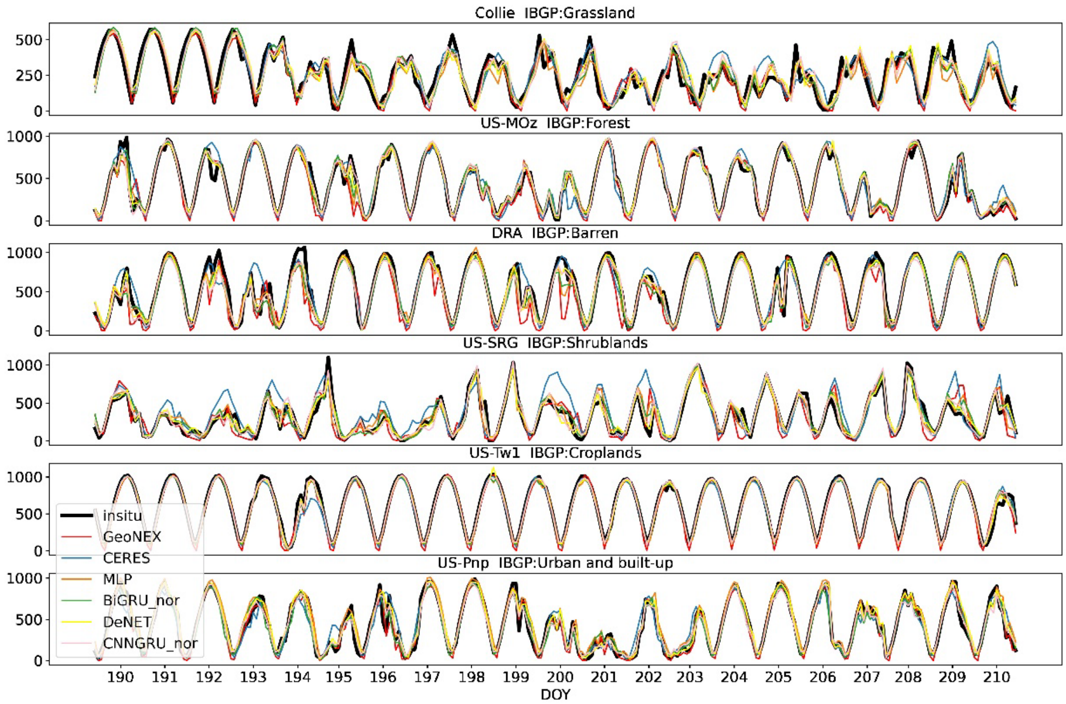

Figure 4-8. Diurnal variation in DSR of MLP, BiGRU_nor, DenseNET, and CNNGRU_nor with CERES and GeoNEX from day of year (DOY) 190 to 210 across six sites.

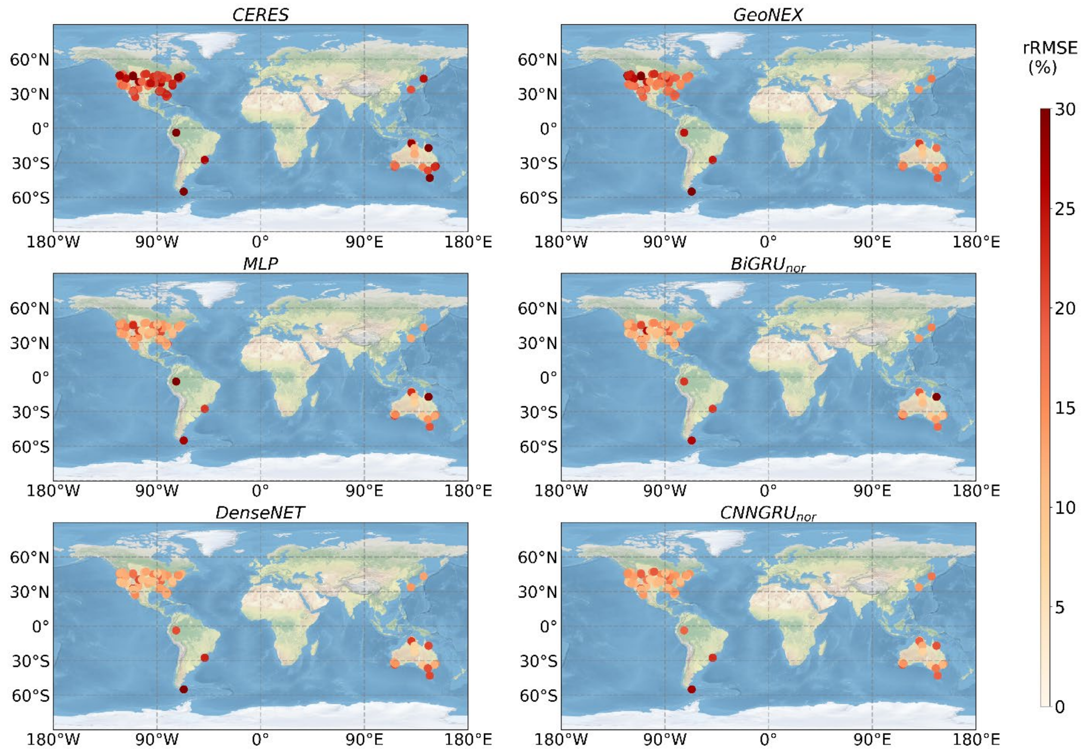

Figure 4-9. Spatial distribution of the rRMSE for MLP, BiGRU_nor, DenseNET, and CNNGRU_nor with CERES and GeoNEX.

Figure 4-9. Spatial distribution of the rRMSE for MLP, BiGRU_nor, DenseNET, and CNNGRU_nor with CERES and GeoNEX.

To conduct a comprehensive comparison of the models, this study included a list of parameter numbers and floating point operations (FLOPs) for each model in Table 4-3. BiGRU_nor and CNNGRU_nor both simultaneously estimated 60 values in a day; hence the average of each estimation was calculated and compared. MLP shows the lowest parameter numbers and FLOPs, followed by BiGRU_nor for each single estimation. DenseNET presented the most complex structure.

Table 4-3 compares the model accuracies at different temporal scales, providing a useful reference for deciding which model to use under different conditions. CNNGRU_nor with both temporal and spatial structures showed the highest accuracy at temporal resolutions higher than an hour. The superiority of GRU-related temporal structures disappeared at daily scales, as DenseNET surpassed both BiGRU_nor and CNNGRU_nor. However, both CNNGRU_nor and DenseNET have high numbers of parameters and FLOPs. Considering the speed and computing resources, BiGRU_nor would be the top choice for DSR estimation under all three temporal scales.

Table 4-3. Summary of models' parameter numbers, FLOPs, and accuracy at 10/15 minutes, hourly, and daily scale; the number in brackets represent the average for each time step in GRU-related model.

Model	**Parameters number**	**FLOPs**	**10/15 minutes RMSE (W/m^2)**	**Hourly RMSE (W/m^2)**	**Daily RMSE (W/m^2)**
MLP	811	795	76.8	63.1	17.0
$BiGRU_{nor}$	18,013 (300)	115,610 (1927)	68.4	58.6	16.1
$DenseNET$	18,913	4,614,984	69.4	57.8	15.1
$CNNGRU_{nor}$	143,761 (2,396)	198,746,606 (3,312,443)	67.1	57.3	16.0

4.4.2 Impact of surface albedo

Surface albedo is a crucial variable in DSR estimation and currently accounts for a minimal input in existing DSR products (Wang et al., 2019; Wang et al., 2021). However, the retrieval of surface albedo degrades in the presence of clouds, snow, and other bright surfaces (Schaaf and Wang, 2015; He et al., 2018). Data gaps exist in current surface albedo data, even in the widely used MCD43, and using surface albedo as an explicit variable may introduce further uncertainties (Schaaf et al., 2002).

The proposed $\text{BiGRU}_{\text{nor}}$ and $\text{CNNGRU}_{\text{nor}}$ models successfully achieved stable and accurate DSR estimations without using surface BSA as inputs. To further testify whether the time-sequence structure helps eliminate the dependencies on surface albedo, we trained and compared the proposed DL models with and without BSA as inputs (Figure 4-10). The inputs of the models compared in this section are summarized in Table 4-2. We also compared the DL models with the LUT method (Wang et al., 2021). When BSA was high, LUT produced extremely poor results and increased the interquartile range (IQR) among 10-fold sites, primarily because both the unreliable sources of surface albedo and limited bands which brought difficulties in distinguishing snow and clouds contribute to the degradation of physical methods over bright surfaces (Karlsson et al., 2017; Wang et al., 2020). Compared to LUT, ML improved the accuracy for all BSA intervals. ML models present a similar accuracy when the surface BSA was lower than 0.2, comparable to most cases in reality. As the surface BSA increased, the difference between inputs with and without surface BSA became visible. For both the MLP and $\mathrm{DenseNET}$ models, the absence of surface BSA as input, led to a higher rRMSE than that of the same model structure with surface BSA, particularly when the surface albedo was higher than 0.4. The difference in such cases can exceed 10% when the surface BSA ranges from between 0.6 to 0.8 for both models. The models without surface BSA as inputs either tend to have larger IQR. However, the same conclusion was not observed in the GRU-related models. $\mathrm{BiGRU}_{\mathrm{nor}}$ achieved the highest accuracy than all the models without surface BSA as inputs at all intervals. Additionally, the differences between BiGRU and $\mathrm{BiGRU}_{\mathrm{nor}}$ in terms of median rRMSE and IQR were found to be small in all cases and substantially smaller than those of other methods when the surface BSA was large. Compared with all models without surface BSA as inputs, the BiGRU structure has better generalities as it is less affected by surface albedo. This is likely owing to the incorporation of time series information in one day, which provides more reliable information on the stable surface. The $\mathrm{CNNGRU_{nor}}$ model displayed satisfactory results when the surface BSA is between 0.4 and 0.6. However, it degrades when the surface BSA was higher than 0.6. One possible reason for this was the depth of the models. Some critical temporal information was lost in the CNN layers, which could not be passed on to the GRU layers.

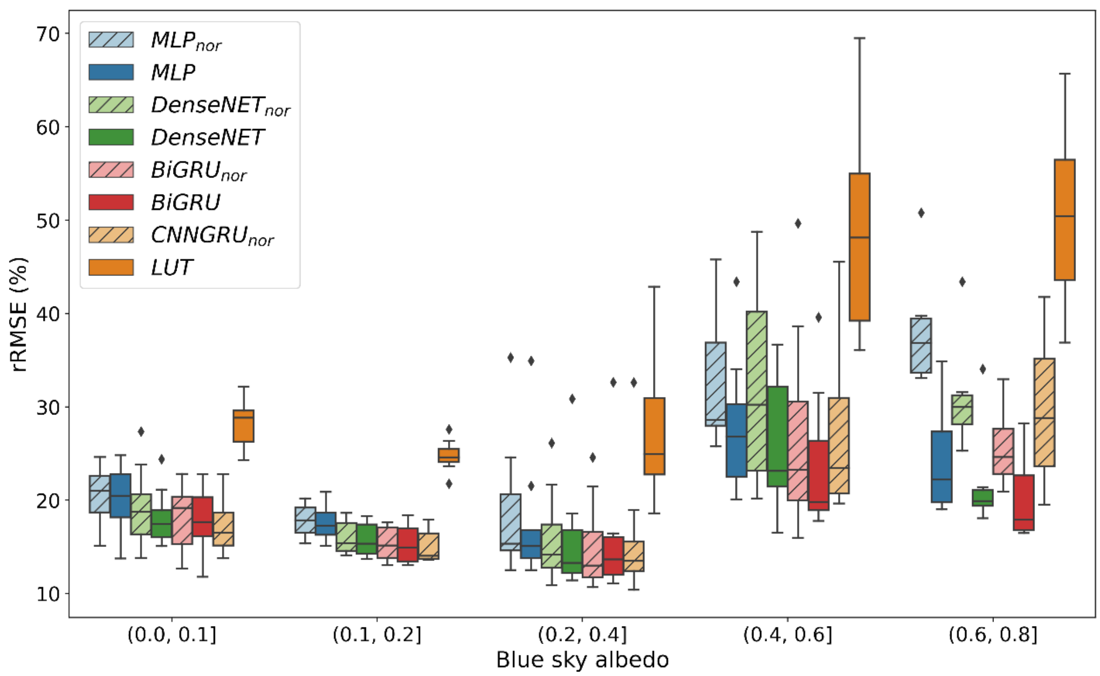

Figure 4-10. The boxplot of model rRMSE with and without surface BSA as inputs during the 10-fold validation; the model without surface BSA as inputs include nor as subscript and highlighted with the shadow.

4.4.3 Impacts of cloud conditions

The study also compared all models with LUT under clear and cloudy conditions (Figure 4-11a). For all models, accuracies under clear-sky conditions were significantly higher than those under cloud-sky conditions. Under clear sky conditions, the difference between different neural networks was trivial, but varied across different models under cloudy sky conditions. The LUT methods presented the highest rRMSE (37.2%) under cloudy conditions, which corresponds well with previous findings (Li et al., 2023). It is notable that LUT do not need any ground measurement data as inputs hence it is with the highest generalibility. MLP presented the second-highest rRMSE (25.4%) and the highest IQR. DL models were shown to improve the DSR estimation accuracies under cloudy conditions. Single-pixel-based BiGRU_nor showed high accuracy, with a rRMSE of 22.3%. One possible reason is that the temporal information before and after the current moment contributes useful information in terms of cloud type and cloud positions. DenseNET and CNNGRU_nor had lower rRMSE under cloudy skies of 22.1% and 21.5%, respectively. These two models incorporate spatial information into their input, which aided in recognizing different cloud conditions. All DL models showed less IQR than MLP which demonstrate their high generality under cloudy condition.

Furthermore, we examined the performance of different models for different cloud top heights (Figure 11b). ML models presented similar accuracy when the cloud height was lower than 2000m. With the increase in cloud height, attributed to the solar-cloud-satellite geometry effects are more prone to occur (Wang et al., 2017), the DL-based models particularly the models with the spatial structure showed high accuracies, indicating that the convolutional filter may help models distinguish the cloud structure and hence improve the DSR estimations.

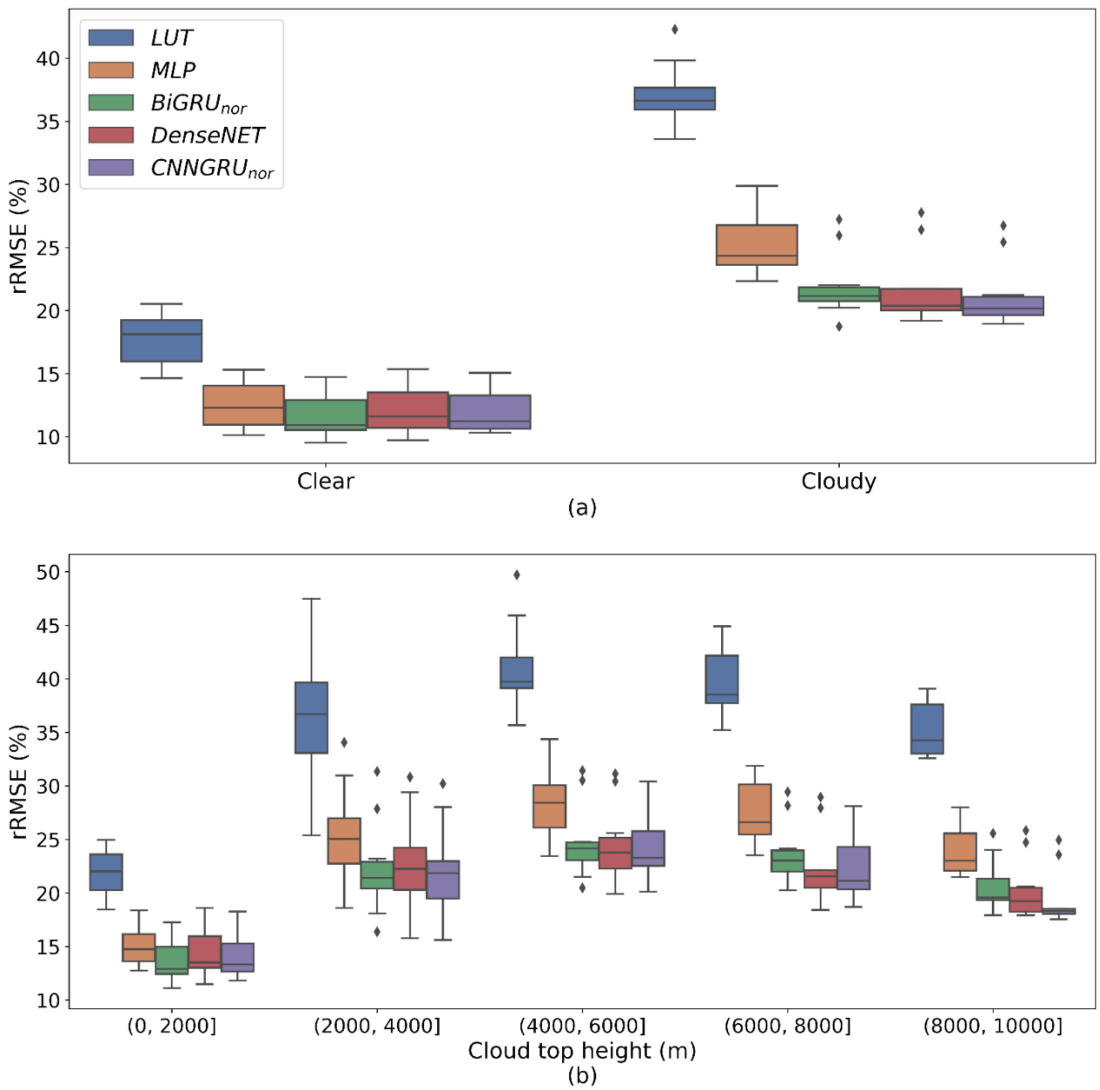

Figure 4-11. The boxplot of models (a) under clear and cloudy sky conditions and (b) under different cloud top heights during the 10-fold validation.

4.4.4 Influence of infrared bands

Existing DSR retrieval methods, particularly physics-based methods, primarily use shortwave bands of TOA reflectance. With flexibility in the inputs of DL models, this study tested the importance of infrared (IR) bands in retrieving DSR. Figure 4-12a shows the 10-fold validation results of MLP trained on all bands ranging from 0.47 to 13.3 µm shared by AHI and ABI sensors, and Figure 4-12b shows the validation results of the same model but trained on shortwave bands only (0.47-3.9μm) (MLP_sw). The inputs difference between MLP and MLP_sw are also summarized in Table 4-2. The total RMSE of MLP_sw is 4 W/m^2lower than MLP . We further analyzed the impacts of infrared bands under different surface conditions as illustrated in Figure 4-12c. Our findings indicate that when the surface albedo is less than 0.4, the differences between MLP and MLP_sw are negligible. However, for surface albedo greater than 0.4, MLP_sw exhibited significantly higher accuracy. This can be attributed to the complementary function of IR bands to the visible bands, enabling the distinction of clouds from high albedo surface types. To investigate the relative contributions of each channel, a sensitivity analysis was applied to all normalized inputs via the Sobol analysis method (Sobol, 2001) based on the MLP method (Figure 4-12d). First-order indices (S1) measure the contribution of a single input to the output variances, while total-order indices (ST) are the total of all single and higher-order interactions of this input. This shows that the blue band (0.47 μm) contributed the most to the surface shortwave radiation estimation, which is consistent with previous findings (Wang et al., 2020). IR bands especially between 12.3-13.3 μm, also showed great contributions in estimating DSR. Both bands were used to generate products such as the cloud mask and the cloud-top height (Menzel and Stravala, 1997; Chang and Xiong, 2019) owing to their ability to delineate the upper level of the cloud (Dupont et al., 2009). Previous studies have reported similar results for IR bands in estimating the diffuse portion of the DSR (Chen et al., 2022).

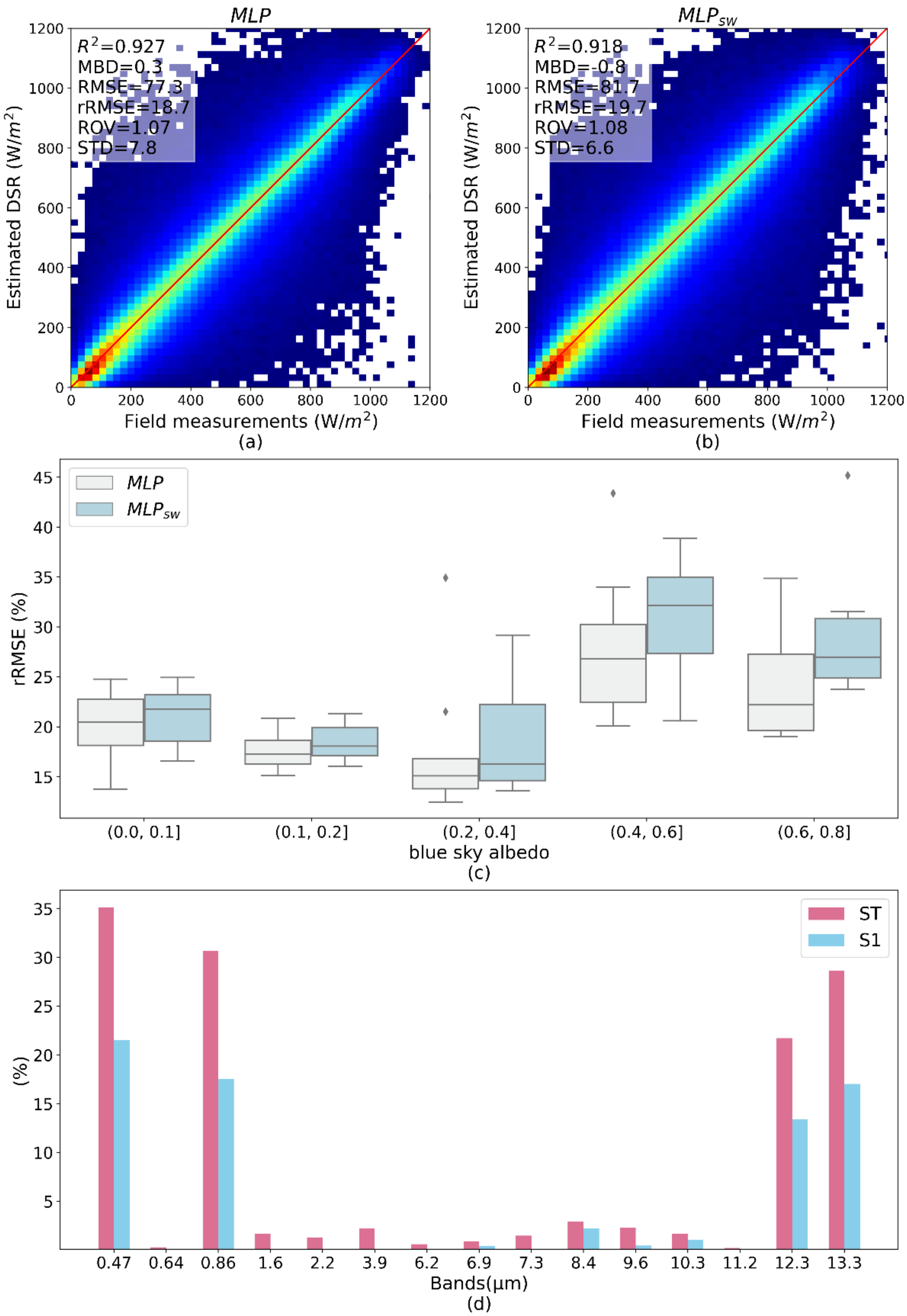

Figure 4-12. (a) The scatterplots of MLP model. (b) The scatterplots of MLP_sw model. (c) The boxplots label the median, quartiles, and outliers of model rRMSE comparison over different surface BSA (d) Sensitivity analysis of all bands in MLP models

4.5 Conclusions

Recently, there has been an increasing need for the retrieval of DSR with high spatiotemporal resolution and less latency. However, the issues of cloud heterogeneity and the dependence of existing methods on high-level products hamper the development of fine-resolution DSR retrieval. DL, which has been widely used in different areas, provides a promising way by incorporating spatial and temporal information in addition to the spectral information used by conventional approaches. This study developed and compared three DL models for retrieving DSR with high spatiotemporal resolution from both ABI and AHI TOA reflectance. DenseNET used convolutional layers to extract spatial information and employed densely connected layers to pass the critical band and angle information to the end of the model. BiGRU_nor employed hidden states to transmit information from previous or future time series, as well as two gates to determine the amount of information to be retained. CNNGRU_nor combined spatial and temporal structures. Both BiGRU_nor and CNNGRU_nor did not include surface albedo as input. The ground measurement data covered 110 stations over AHI and ABI coverage areas, with approximately 1.2 million samples in total. To avoid data overfitting, the models were evaluated over 12 independent sites that covered sufficient land types as well as a 10-fold site-based validation. Both independent site validation and 10-fold site-based validation presented similar accuracies, which demonstrated that the proposed methods do not have overfitting issues, hence we do not apply the transfer learning technique introduced in chapter 3.

DL methods show high accuracy for estimating high spatial- and temporal-resolution DSR. BiGRU_nor, DenseNET , and CNNGRU_nor achieved an instantaneous RMSE of 68.4 (16.1%), 69.4 (16.3%), 67.1 (15.7%) W/m^2, respectively. The hourly results were compared with those of CERES-SYN and GeoNEX. BiGRU_nor, DenseNET , and CNNGRU_nor achieve 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.9%) respectively, which is higher than CERES with 88.8 (21.4%) and GeoNEX (18.8%) with 77.8 (18.8%). Different DL models exhibit different strengths under different conditions. CNNGRU_nor presented the highest accuracy at hourly or sub-hourly resolution. DenseNET showed the highest accuracy at daily scale. BiGRU_nor was considered to be the best choice for DSR estimation at all temporal resolution considering the model efficiency. Models, such as DenseNET and CNNGRU_nor, incorporating spatial structures, presented higher accuracies under cloudy and particularly high cloud conditions. The use of temporal structure in DL models, such as BiGRU_nor, not only allows for similar accuracies without surface albedo but also presents higher accuracies when surface albedo is high. The incorporation of IR bands can increase estimation accuracy, particularly over high albedo surface.

Chapter 5: A transformer-based approach to nowcasting solar radiation from satellite images

Solar energy has gained prominence with the evolution of PV technologies and the surging demand for renewable energy. An accurate forecast of solar radiation with low latency is critical for stable and efficient PV and grid operation. However, solar radiation is highly variated due to both the Earth's distance to the sun as well as the cloud movement. This study aims to employ a space-time Transformer model to bridge the gap of 0-3 hour lead time with 15-minute intervals of solar nowcasting from geostationary satellite observations only. For hourly forecasting, the proposed SolarFormer model achieves the RMSE (rRMSE) values of 93.8 W/m^2 (15.0%), 118.9 W/m^2 (19.8%), and 129.1 W/m^2 (24.2%) validation over seven SURFRAD stations. The accuracy is better than existing machine learning models and comparable with numerical weather prediction results as well as commercial products. The study also reveals the superiority of the Transformer for its high computing and memory efficiency and shows the potential of the Transformer on longer lead time forecasting.

5.1 Introduction

Solar energy has gained prominence with the evolution of PV technologies and the surging demand for renewable energy. Over the past decade, solar power's share has risen consistently, accounting for 27% of renewable generation (Statistical Review of World Energy, 2021). Forecasts indicate that by 2050, solar and wind will contribute almost 42% to the grid, with solar driving nearly 80% of this growth (EIA, 2021). As installations surge, optimizing solar energy use becomes pivotal for economic and ecological progress.

Accurate forecast of solar radiation with low latency is critical for stable and efficient PV and grid operation. Unlike more consistent power generation methods such as hydropower or nuclear energy, which can reliably produce and adjust their output based on demand, solar PV systems often face power fluctuations due to the inherent variability of solar radiation (Qin et al., 2022). This variability arises from the Earth's distance from the sun and fluctuating cloud cover (Dong et al., 2013). It is observed a notable surge in solar irradiance, approximately by 700 W/m^2, was associated with a disruption in the dense, fragmented cloud cover, all within a 30-minute timeframe (Miller et al., 2018). Solar PV devices respond immediately to irradiance fluctuations, resulting in the grid's power output being highly dependent on these conditions (Sharma et al., 2016). Accurately short term forecasting (0-6 hours) surface solar irradiance aids in the strategic planning and deployment of electricity across various units (Fu & Cheng, 2013). It also aids the reduction of battery size incorporated in the PV plants facilitates (Pérez et al., 2021). Perez et al. (2014) highlighted that precise forecasts could diminish the expenses linked to short-term variability mitigation by almost 50%

Over the past decades, site-based forecasting has become a staple for short-term predictions. For ultra-short-term solar radiation forecasts (within an hour), the primary data sources are surface sky camera networks (Lin, Zhang, & Wang, 2023). For broader short-term forecasts, the data inputs encompass historical site-specific solar radiation records and weather information. Various forecasting methodologies have been employed, including numerical weather prediction (NWP) (Chow et al., 2011) statistical approaches (Yang, Jirutitijaroen, & Walsh, 2012) machine learning techniques (Sharma et al., 2016), 18-19], and deep learning strategies integrated with data from neighboring sites (Reikard, 2009). Nonetheless, given their constrained data sources, site-based forecasts typically suit short timescales or predictions requiring lower accuracy. In environments with significant variability or for longer prediction lead times, the accuracy of these forecasts remains suboptimal (Sharma et al., 2016).

Satellite platforms offer insights into cloud patterns beyond the scope of local, surface-based networks (Miller et al., 2018). Recognizing this, recent research has investigated integrating satellite data with site-based observations for enhanced short-term forecasting. For instance, Choi, Rachunok, & Nateghi (2021) employed a CNN model on NOAA's Geostationary Operational Environmental Satellite 8 (GOES8) infrared channels for site-specific solar radiation predictions. Gallo et al. (2022) utilized 3D CNN and ConvLSTM with the SEVIRI instrument on the Meteosat Second Generation (MSG) satellite, merging it with ground measurements during clear conditions. Likewise, Qin et al. (2022) combined CLSTM methodologies with satellite and ground data. Yet, these methods still lean heavily on ground measurements, potentially limiting their applicability in expansive solar energy operations, especially as both macro and micro solar PV penetrations rise.

In large-scale forecasting, Numerical Weather Prediction (NWP) models, which utilize geographical and meteorological data, are commonly employed with notable accuracy. However, despite advances in hourly-updated NWPs, their spatial prediction resolution remains less refined than what is ideal for photovoltaic grid operators (Fu & Cheng, 2013). For short-term (0-3h) solar insolation predictions, satellite-based techniques are particularly suitable (Miller et al., 2018). Cloud motion vector (CMV) has been developed for satellite-image-based forecasting Lorenz et al. (2007). The SUNY solar forecasting model, incorporated into the SolarAnywhere software, refines CMV using NWP models, boasting a 25.5% accuracy for 3-hour forecasts (Perez et al., 2010). However, this product isn't freely accessible. Parallelly, with cloud movement calculations at their core, data-driven machine learning methods offer alternative solutions. Pérez et al. (2021) applied a convolutional neural network to solar radiation datasets from Meteosat's SEVIRI, although its applicability remains limited to a single test site. Nielsen et al. (2021) employed a ConvLSTM model on effective cloud albedo data from SEVIRI over Europe. To our knowledge, this is the sole method employing spatiotemporal deep learning for satellite-based image-to-image forecasting. Yet, ConvLSTM faces scalability issues, particularly when larger images and longer temporal range challenge computational and memory capacities (Gao et al., 2022)

In recent times, Transformers leveraging the attention mechanism have garnered significant interest, mainly due to their uncomplicated design coupled with their adeptness at recognizing extensive correlation structures. Many researchers have adapted this primarily language-focused architecture to suit visual and video-related tasks, achieving commendable outcomes. Notable works in video related tasks include divided attention (Bertasius et al., 2021), axial attention (Ho et al., 2019), video Swin transformer (Liu et al., 2022), and others. Gao et al. (2022) introduced

EarthFormer, a general space-time transformer based on video transformers. Specifically designed for earth system predictions, EarthFormer has demonstrated its prowess in forecasting precipitation. This model harnesses attention across varying spatiotemporal patterns, resulting in reduced computational needs and facilitating parallel processing. Nevertheless, the potential of these data-centric methods in short-term solar irradiance forecasting remains underexplored. Furthermore, detailed examinations into the ideal spatial and temporal dimensions, as well as the most effective patterns for spatial-temporal separation, remain pending in the context of solar radiation forecasting

This paper primarily emphasizes enhancing satellite-based intraday solar forecasting, aimed at bolstering PV generations for utility's real-time dispatch procedures. By integrating the transformer with cutting-edge geostationary satellite observations, our proposed model, SolarFormer, bypasses the need for ground measurements, facilitating comprehensive estimations. Demonstrating both high accuracy and efficiency, SolarFormer overcomes data scarcity issues commonly encountered in most machine learning models. This study also underscores the transformer's promising capabilities in long-term forecasting. This chapter structure unfolds as follows: Section 5.2 delves into insitu and satellite data; Section 5.3 elaborates on the methodology; Section 5.4 showcases and interprets validation results, while Section 5.5 concludes the discussion.

5.2 Data

Insitu data from seven SURFRAD stations within the CONUS from 2018 were employed to validate the SolarFormer model (Figure 5-1). SURFRAD records DSR measurements at intervals of 1 minute. This data underwent a quality assessment based on the methodology proposed by Li et al. (2021). Any data with an SZA exceeding 85 degrees or values that were zero or negative were discarded. To mitigate potential errors in point-scale DSR measurement, the data were averaged over 15-minute intervals, aligned with the satellite's pass time, as recommended by Huang et al. (2016).

The BiGRU part uses 15 spectral bands from the GeoNEX L1G TOA reflectance data sourced from the ABI on board the GOES-16, relevant geometry and static elevation information derived from GTOPO30 (Li et al., 2023). The GeoNEX L1G TOA data can retrieved from NASA's Advanced Supercomputing (NAS) Division GeoNEX platform, as detailed by Wang et al. (2020). The data is archived in a consistent global tile gridding system with other geostationary sensors. It produces full disk scanning every 10 or 15 min depending on the sensor and years. We included seven tiles covering the seven stations of SURFRAD in 2018, 2019, and 2021 (Figure 5-1). All the samples from 2019 and 2021 served as training data, while those from 2018 were utilized for testing.

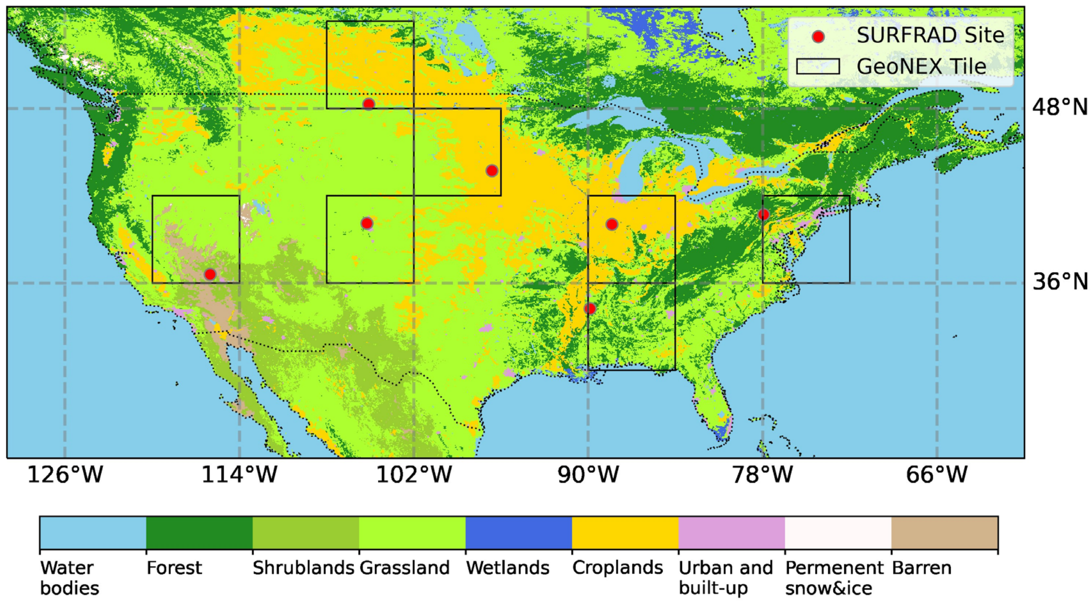

Figure 5-1. The distribution of seven SURFRAD stations. The black rectangles outline the GeoNEX tile coverage areas.

5.3 Methods

5.3.1 SolarFormer

The SolarFormer model consists of two major parts, the estimation and prediction parts (Figure 5-2). The estimation are realized by applying previously developed and publicly available Bi-directionaly GRU models (Figure 5-2a) (Li et al., 2023). For the BiGRU, the model takes all available GeoNEX observations in one day with their 15 bands information and geometry as well as the elevation for each pixel. The outputs are the estimated DSR for the corresponding pixel. After running through each pixel of one image, we can retrieve the DSR for the whole image at each time step in one day. The detailed description and configuration of the BiGRU model can be found in Li et al. (2023)

The retrieved DSR images are with an original spatial resolution as GeoNEX of 600km spanning 120 pixels. The image is divided into 4 patches, which leaves each of these samples covering an area of 300km x 300km and contains 60 x 60 pixels. To eliminate the impact of SZA (θ) on the change of diurnal DSR, before feeding into the transformer part, all DSR values R_i are normalized by the theoretically clearest condition in a day $E_{\max,i}$ at time i. The retrieved samples can be interpreted as clearness index x_i, which is calculated as:

$$E_{\max,i} = E * COS(\theta_i)$$
$$x_i = \frac{R_i}{E_{max,i}}$$

Where E represents the solar constant which is 1360 W/m^2. The input tensor with size [T, W, H, C]. T is the input time series. W and H are width and height. C is the input channel number, which is 1. The input images first go through a CNN downsample blocks, and encoded by position embedding.

The forecasting part are achieved by the hierarchical encoder-decoder transformer (Figure 5-2b). The main mechanism is about the cuboid self attention block. The cuboid cross attention block is similar with self-attention block but has additional cross attention layer based on the input both from the tensor and the memory in the encoder. For each cuboid attention block, there are several attention layers depends on the choice of the cuboid pattern to decompose the input tensor. The cuboid decomposition is the essence of transformer applying on earth science spatial-temporal studies as it reduces the complexity significantly but still is able to capture different types of data correlation (Gao et al., 2022). In this study, we choose the Divided Space-Time pattern to split the input tensor based on their spatial and temporal axial separately, leading to two attention layers in one block. For each attention layer, the decomposed sequence are flattened and apply multi-head self-attention in parallel. An attention function can be described as mapping a query (Q) and a set of key (K) - value(V) pairs to an output where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key, which are also referred to the attentions core. For multi-head attention, a head specific projection w_Q, w_K, w_V is applied to Q,K,V respectively. Hence the multi-head attention can be summarized as

$$Attention(Q, K, V) = softmax\left(\frac{Qw_K K}{\sqrt{C}}\right) w_V V$$

For self-attention, Q,K,V are all same as input tensor. For cross-attention, Q is the input tensor. K and V are the memory from the output of the same hierarchies of encoder. The output of the attention is with the same shape as each x tensor. The output of each cuboid attention block is stacked with the residual connection. After D blocks of attention, introduced by Earthformer, downscaling layers are added to further decrease the shape to increase the efficiency in the encoder. Corresponding upscaling layers are added in the decoder. The combination of attention blocks and upscaling/downscaling are repeated from M times. The output of the final upsample layer are multiplied by the E_{max} to generate the final DSR values. The detailed configuration of the model can be found in table.

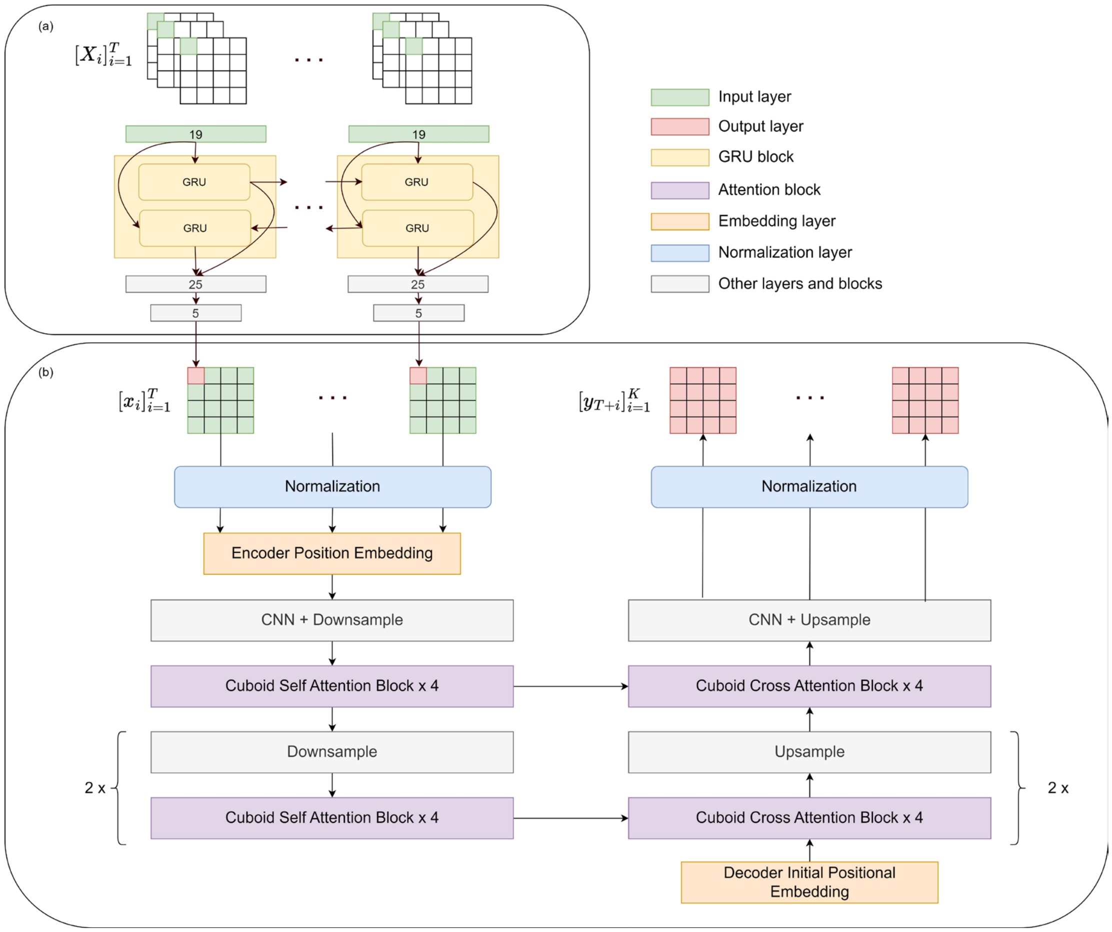

Figure 5-2. The structure of SolarFormer (a) the Bi-GRU part. (b) the space-time Transformer part.

5.3.2 Evaluation metrics

The models were evaluated using several metrics, including the coefficient of determination (R^2), mean bias difference (Bias), root mean square error (RMSE), relative root-mean-square error (rRMSE). Notably, RMSE and rRMSE are served as the primary metrics for comparing the models against existing benchmarks. The formulas used to compute these evaluation metrics are delineated below:

$$R^2 = \frac{\sigma_{R_p R_g}}{\sigma R_p \sigma R_g}$$

$$Bias = \frac{1}{N}\sum\nolimits_{i=1}^{N}(R_p^i - R_g^i)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_p^i - R_g^i)^2}$$
$$rRMSE = RMSE/\overline{R_g}$$

Where σR_p denotes the standard deviation of the predicted values for the test samples, while σR_g stands for the standard deviation of the ground-truth measurements for the same samples. Furthermore, $\sigma_{R_p R_g}$ represents covariance of the predicted values and ground measurements for the testing samples. For a given observation, R_p^i and R_g^i denote the predicted value and the corresponding ground-truth measurement, respectively. The mean value of the ground-truth measurements for the test samples is represented $\overline{R_g}$, and N is the total number of the testing samples.

Additionaly, we construct the smart persistence forecasting model (PFM) as the baseline model, and calculate the forecasting skill (FS) based on the baseline model. The PFM assumes the static cloud conditional at the last input time x_T throughout the forecast lead time and the predicted DSR (R^{PFM}) is only influenced by the $E_{\max}$. The FS presents the predict ability of the proposed model compared with the PFM.

$$R_{T+i}^{PFM} = x_T * E_{\max,T+i}$$
$$FS = 1 - \frac{RMSE}{RMSE_{PFM}}$$

5.4 Results

5.4.1 Images validation

In this section, we use the value generated by Bi-GRU as the ground truth to validate the predictive capability of the transformer component. Figure 5-3 presents a test sample of the input, target, and SolarFormer-predicted values for the GWN site on January 13, 2018. The input time frame spans from 9:00 AM to 10:45 AM, while the output time frame extends from 11:00 AM to 1:45 PM. Overall, the images predicted by SolarFormer exhibit a DSR pattern similar to the targets over the three-hour forecast, capturing the evolution of DSR due to cloud dispersion. For instance, we observe the vanishing of low DSR on the right side of the image. However, certain detailed patterns are inaccurately predicted, especially for longer forecasting durations.

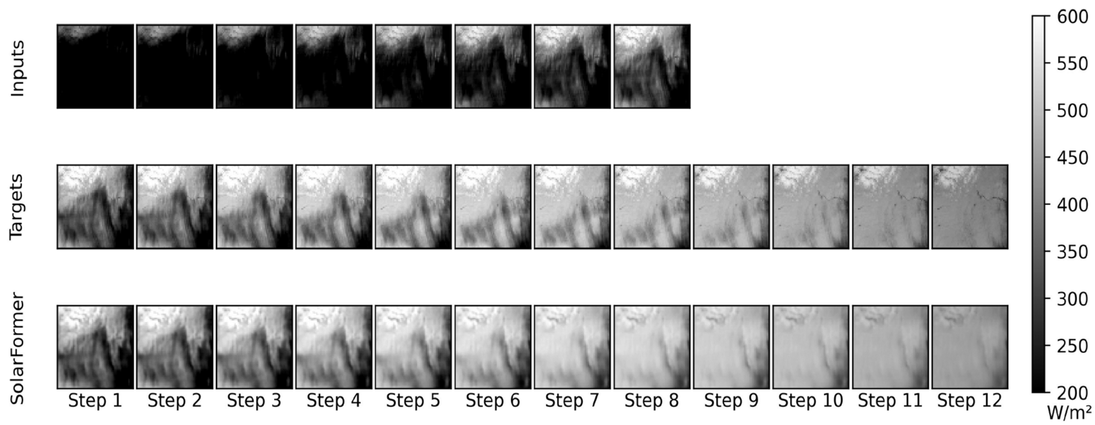

Figure 5-3. One testing sample of input, target, and SolarFormer predicted values over GWN site on January 13th in 2018. The input time step ranges from 9:00 AM to 10:45 AM and the output time steps ranges from 11:00 AM to 1:45PM.

Each pixel in the predicted images is validated with the targeted images for different lead time and compared with the baseline model as well (Figure 5-4). Both predicted and target images are normalized by $E_{\max}$. As expected, the rRMSE of both methods increase as the forecasting time step increase, but SolarFormer shows significant slower increasing speed and higher accuracy for longer time step forecasting. They both start from rRMSE around 3%-5% from the first 15 min forecasting, with SolarFormer maintains rRMSE lower than 20% when forecasting 3 hours ahead while the baseline model show 10% larger. The FS of SolarFomer shows a slightly decrease from 15 min to 3hour prediction, from 40 to 31.

Figure 5-4. The validation results of clear sky ratio (CSR) at each pixel within one image for both the SolarFormer and baseline models.

5.4.2 Insitu validation

The study also validated the SolarFormer's predictions with ground measurements from seven SURFRAD sites to assess the model's overall efficacy and compared it with the baseline model (Figure 5-5). We first validated the BiGRU-retrieved GHI estimation at the last input timeframe, shown as 0 min. The estimated RMSE (rRMSE) is 92.1 W/m2 (15.9%), which aligns with the accuracy reported by Li et al. (2023), representing the systematic uncertainties introduced by the estimation methods. Unlike the image-scale validation, the baseline model exhibits high accuracy in the first 30 min, surpassing the estimation accuracy within the initial 15 min. This is likely due to the relatively stable atmospheric conditions in such a short period combined with the spatiotemporal representativeness issues of point-based ground measurements compared to satellite pixels as well as the systematic uncertainties introduced by the estimation part. After 30 min, the SolarFormer demonstrates superior performance. The overall RMSE (rRMSE) of the SolarFormer for the 3h prediction is within 150 W/m2 (30%), compared to the baseline model, which achieved 198 W/m2 (38%). Regarding FS, the most significant increase occurs within the first hour, after which it largely stabilizes at approximately 25% regardless of lead time, further demonstrating the transformer's advantages in handling long-time series data. However, SolarFormer exhibits increasing bias as the forecasting lead time extends, likely due to the inevitable error accumulation in long time-series projections inherent to data-driven methods. To mitigate this, we plan to adopt the multi-step forecasting method developed by Wang et al. (2023) in future studies.

To assess the overall efficacy of the SolarFormer model, we compared its predictions with the ground truth obtained from the seven SURFRAD stations Figure 5-6 showcases scatterplots that juxtapose the results of the SolarFormer with the baseline model, all analyzed at an hourly granularity across the various sites. For the baseline model, the RMSE (rRMSE) values for the first through third hour forecasting are 124.5(19.9%), 159.1(26.5%), and 170.5(31.9%) W/m^2 respectively. In contrast, the SolarFormer model demonstrates improved RMSE (rRMSE) values of 93.8(15.0%), 118.9(19.8%), and 129.1(24.2%) W/m^2 for the same forecasting durations. The FS of the SolarFormer consistently hovers around 25 for all three forecasting hours. Table 5-1 details the corresponding metrics for each station. Notably, the RMSE for most sites, both at the one-hour and three-hour forecasts, fall within the upper and lower bounds as defined by Yang (2022). This underscores the utility of the proposed models. Furthermore, these hourly results are not only comparable but, in some instances, surpass the best outcomes previously recorded in NWP modeling or those reported for SolarAnywhere v2 in earlier research. It's important to note, however, that the results might not have been compared across identical years.

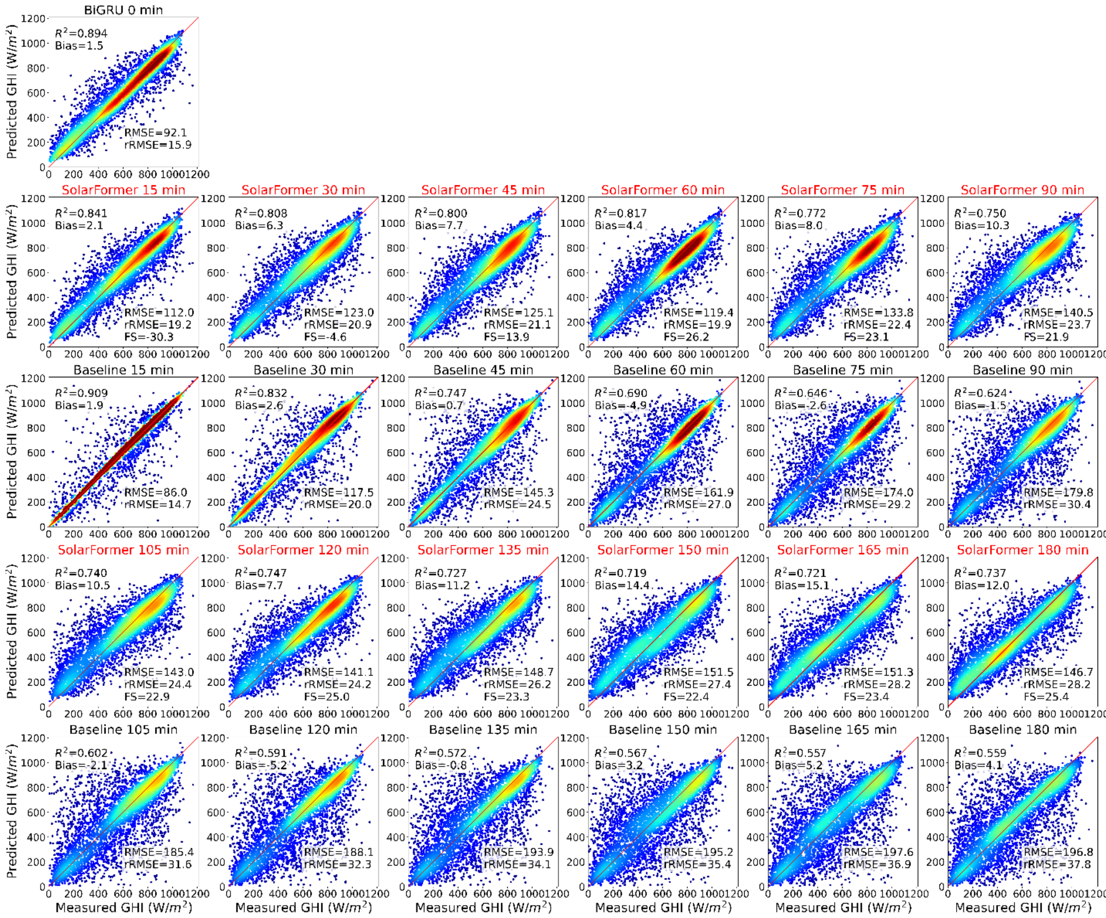

Figure 5-5. The validation results of solar radiation GHI over seven surface radiation budget observing network (SURFRAD) sites for both the SolarFormer and baseline models. Step 00 min represents the last input time frame, which we validate over BiGRU generated GHI

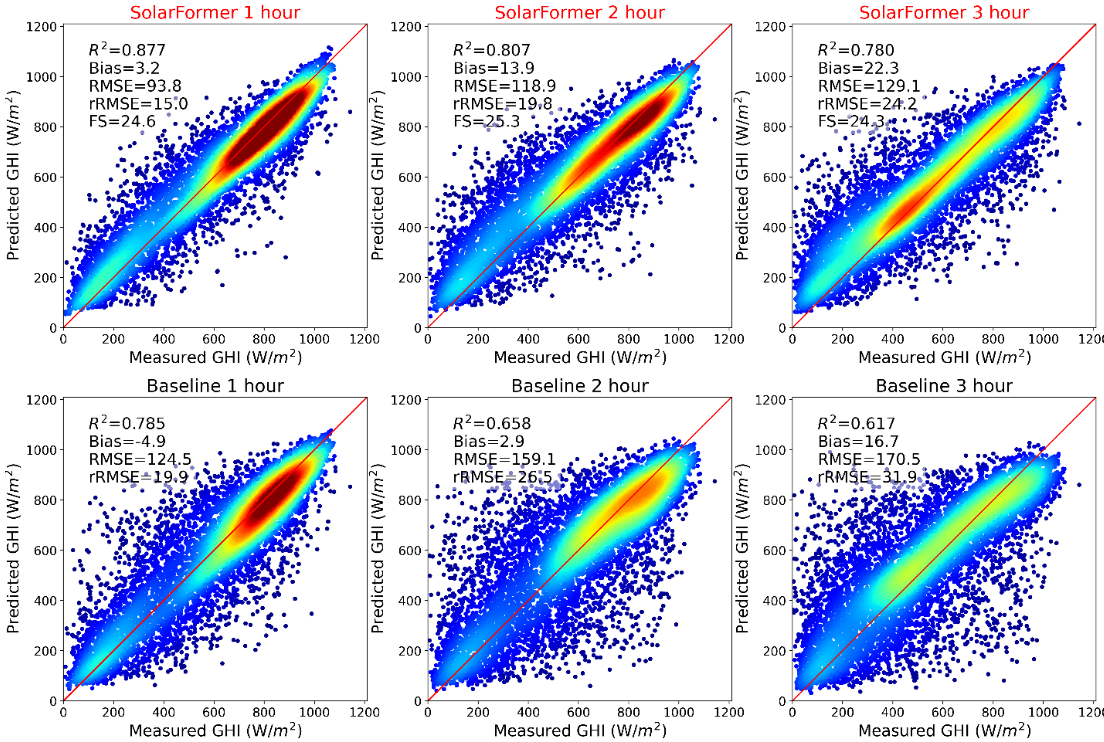

Figure 5-6. The validation results on all ground truth for both baseline model and SolarFormer at hourly granularity.

Table 5-1. The validation metrics of SolarFormer over seven SURFRAD sites at hourly granularity

	1 hour				2 hour				3 hour			
	FS (%)	R^2	RMSE (w/m^2)	rRMSE (%)	FS (%)	R^2	RMSE (w/m^2)	rRMSE (%)	FS (%)	R^2	RMSE (w/m^2)	rRMSE (%)
ALL	24.63	0.88	93.82	14.99	25.26	0.81	118.87	19.80	24.27	0.78	129.10	24.18
BON	32.50	0.91	77.73	12.98	32.46	0.85	101.39	17.56	28.65	0.82	114.60	22.21
DRA	22.99	0.83	86.68	10.72	25.74	0.78	101.42	13.15	23.83	0.81	107.54	15.77
FPK	22.26	0.88	88.81	14.67	22.06	0.83	108.74	18.87	19.84	0.81	116.86	22.68
GWN	26.67	0.89	87.78	14.12	24.47	0.80	119.30	19.51	23.73	0.76	130.43	23.46
PSU	16.78	0.84	110.60	21.39	20.44	0.75	137.74	26.91	23.34	0.72	141.94	30.43
SXF	30.31	0.90	86.56	15.22	32.34	0.85	106.67	19.89	33.48	0.83	113.92	24.36
TBL	23.97	0.83	112.75	17.38	22.54	0.73	147.74	24.42	19.49	0.68	167.40	32.04

We then assessed the results and depicted the rRMSE and FS for each SURFRAD station, along with their mean values, spanning time up to 3h at 15-min intervals (Figure 5-7). Among all the sites, DRA records the lowest rRMSE but also the lowest FS (Figure 5-7). This is attributable to the DRA sites being located in Nevada's arid desert, where atmospheric conditions are comparatively stable and clear. Consequently, both the SolarFormer and the baseline models can achieve high prediction accuracy, but the advantages of the SolarFormer in dealing with dynamic cloud movement are also diminished in these regions. For the other sites, although the prediction rRMSE may vary slightly, the FS is similar, demonstrating the stability of the SolarFormer across different climate types. We conducted further analysis to understand whether the forecasting accuracy of the SolarFormer is influenced by its initial conditions, namely the estimation accuracy of the BiGRU at the last input time frame. We calculated the rRMSE across all lead times and the absolute error (AE) at the initial condition for each prediction, then aggregated the prediction rRMSE with the initial AE (Figure 5-8). Our analysis revealed a significant correlation between the prediction rRMSE and the initial AE. A larger initial AE tends to result in a higher rRMSE and a greater standard deviation in the rRMSE of the SolarFormer predictions. Notably, the upper error bar of the prediction rRMSE corresponds to the highest value within each initial AE bin. This suggests that achieving superior accuracy at the initial condition can lead to more accurate forecasting outcomes.

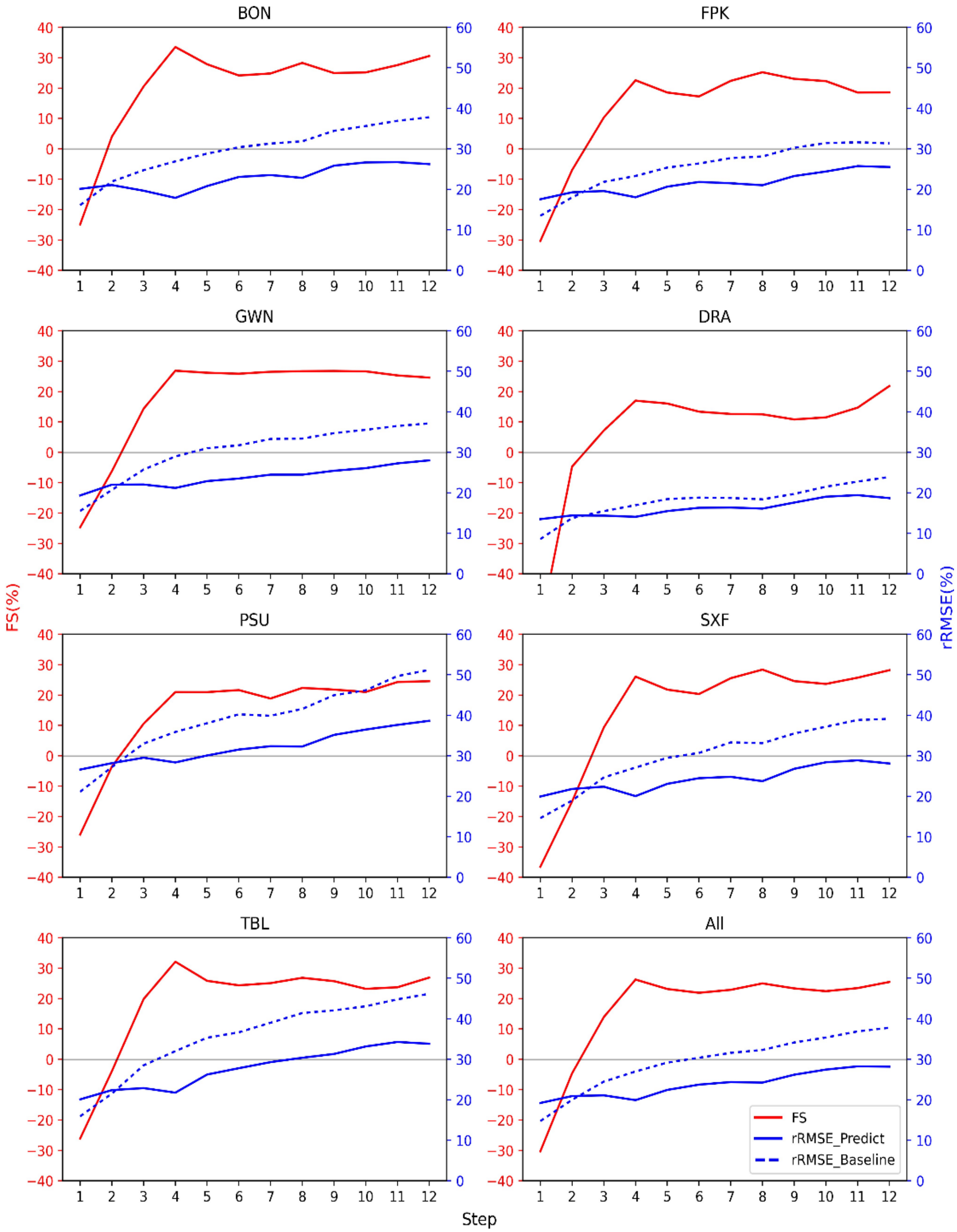

Figure 5-7. The rRMSE and FS over seven SURFRAD stations and all sites from time step 1 to 12. Step 0 represent the validation results of BiGRU estimation at the last input time step.

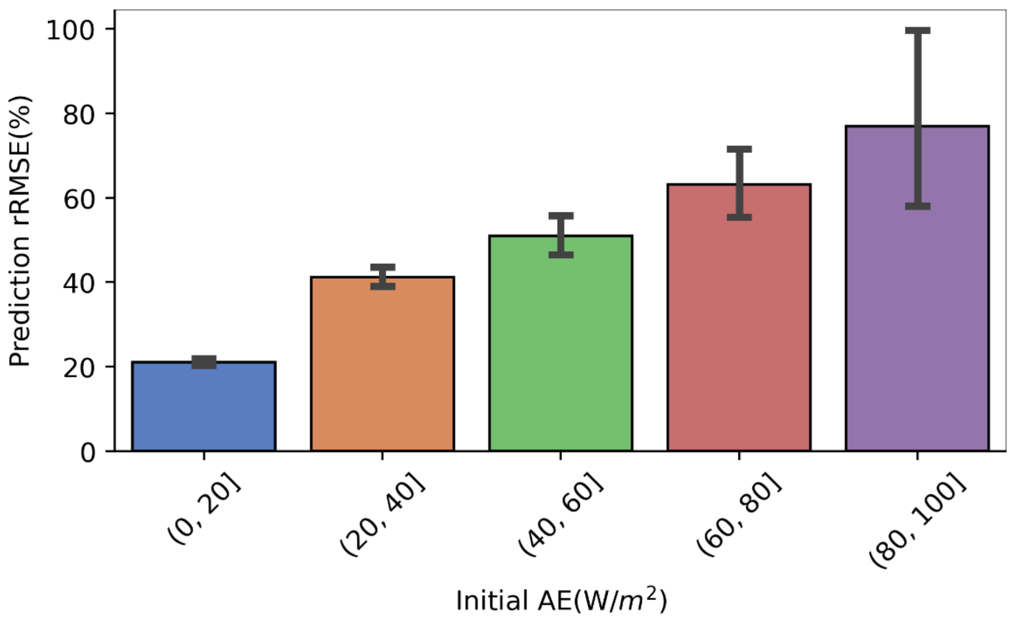

Figure 5-8. The change of prediction rRMSE(%) with different absolute error (AE) range at the initial condition. The bar represents the standard deviation of the rRMSE.

The study explores the performance of SolarFormer under rapidly changing cloud conditions. We segregated the results based on whether the cloud mask changed or remained consistent, as illustrated in Figure 5-9. For situations where the cloud mask remained unchanged, the forecast skill during the initial two steps was either below or equal to 0. This indicates that, within a forecasting window of 30 minutes and in stable atmospheric conditions, SolarFormer's performance is on par with or not superior to the baseline model. After the 1-hour mark, the forecast skill becomes consistent as the forecasting interval extends at around 15. In contrast, for scenarios where the cloud mask underwent changes, the forecast skill consistently exceeds 0 and grows with increasing forecasting intervals. This trend highlights the enhanced capability of SolarFormer over the baseline model, particularly as the forecasting duration expands. Moreover, the forecast skill for varied cloud masks consistently outperforms that of unchanged masks across all intervals. This not only underscores SolarFormer's proficiency during significant atmospheric shifts but also suggests that the model is more adept at handling variable conditions compared to consistent atmospheric scenarios.

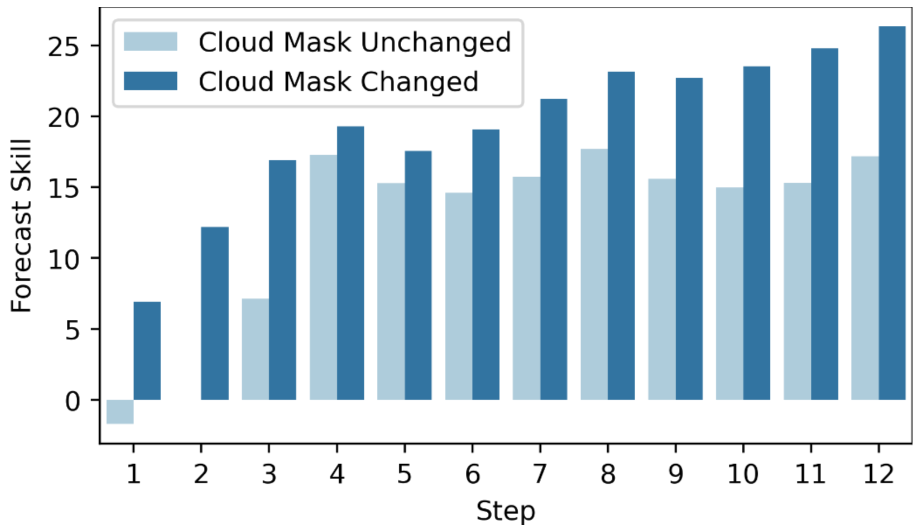

Figure 5-9. The forecast skill of SolarFormer at different predict steps with cloud mask changed or unchanged.

5.4.3 Analysis and evaluation of the SolarFormer components and forecasting potentials

The study evaluates the capability of SolarFormer in forecasting over extended time periods and identifies factors affecting its accuracy, as depicted in Figure 5-10. The combination of input, output frames, and spatial dimensions has varying effects on performance. Generally, longer output frames correlate with reduced accuracy, especially for shorter lead time combined with smaller spatial size. When dealing with larger spatial size, the accuracy differences between 12 and 16 output frames start to converge by the 12th time step, which is not observed for smaller spatial sizes. Increasing the input frames typically leads to enhanced accuracy. This trend becomes particularly pronounced when paired with a larger spatial size for forecasts that span beyond three hours. The findings indicate that the SolarFormer consistently maintains its forecasting capability over longer durations without a marked decline in accuracy. For short-term forecasts, up to 3 hours, the configuration adopted in this study is optimal. However, for more extended forecasting horizons, expanding both the input frames and spatial size can bolster the reliability of long-term predictions.

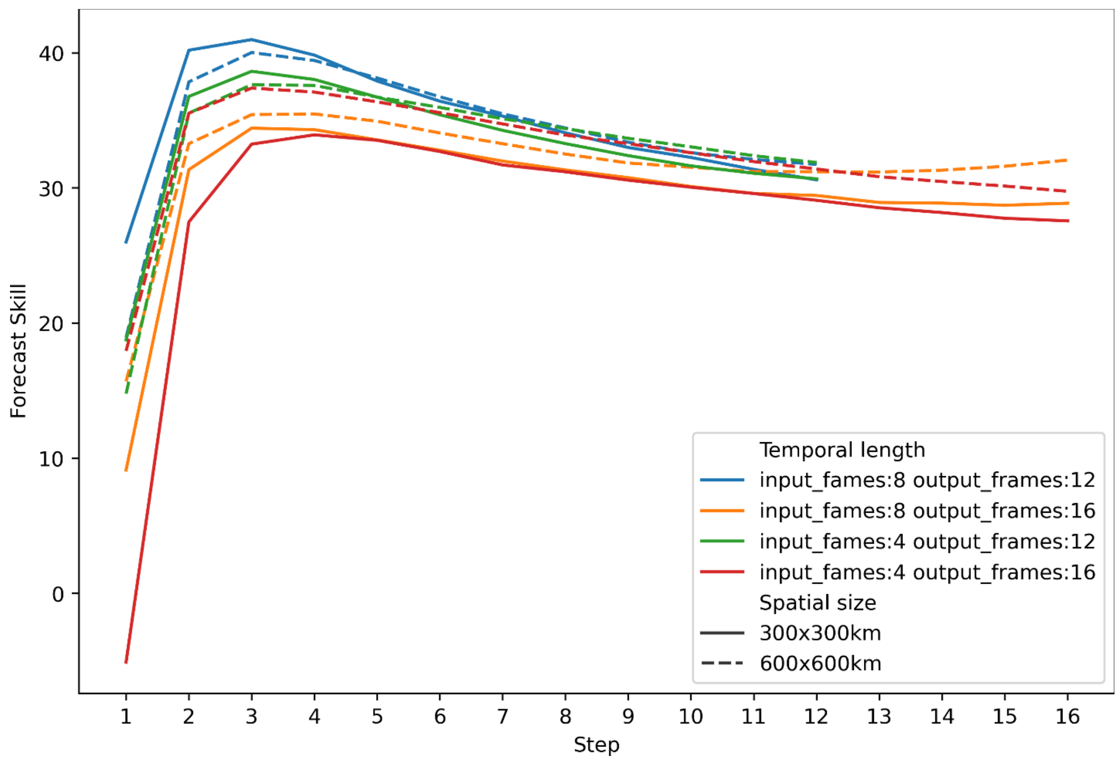

Figure 5-10. The forecast skill of different input and output frames and input spatial size at different time step.

Table 5-2. Comparison of different model structures and their performance metrics for DSR estimation

Model Structure		rRMSE	FLOPs	#Para
Cuboid Types	divided_st	12.7	22.95G	4.81M
	spatial_lg_v1	13.5	30.15G	6.59M
	video_swin	13.4	23.14G	4.81M
	axial	13.4	29.55	6.59M
Global Vector	0	12.7	22.95G	4.81M
	8	12.8	22.96G	5.48M
Downsample scales	2	12.7	22.95G	4.81M
	3	13	13.7G	4.87M

In our study, we tested various model structures to identify the optimal configuration for the space-time transformer when applied to solar forecasting (Table 5-2). We assessed the model's performance across different cuboid types, global vectors, and downsample scales. As delineated in Gao et al. (2022), cuboid types include diverse data correlation methodologies, instrumental for the stacking of multiple elementary attention layers. Our comparative analysis encompassed several types: divided_st (implementing separate temporal and spatial attentions), spatial_lg (segmenting tensors into 8x4x4 patches), video_swin (utilizing 2x4x4-sized patches), and axial (segmenting along the tensor's temporal, height, and width dimensions). The global vector, as introduced in Gao et al. (2022) serves as a linkage between different cuboids. Meanwhile, downsample scales represent preliminary steps that reduce the spatial resolution of the input. In contrast to the Earthformer, originally designed for precipitation forecasting and best paired with an axial cuboid type and 8 global vectors, our results indicated that the divided_st type without a global vector yielded the highest accuracy for solar forecasting. Moreover, our findings highlighted that increased downsample layers tend to decrease accuracy but significantly reduce FLOPs. Future studies should thoughtfully weigh these trade-offs when determining the choice of convolutional layers.

SolarFormer's distinct advantage lies in its reduced computational demands and memory consumption. To illustrate this, we use two metrics: floating point operations per second (FLOPs) and the number of parameters. These are employed to compare various input frames, output frames, and input image spatial dimensions of SolarFormer against those of ConvLSTM (Table 5-3). For a spatial size set at 300km, SolarFormer consistently exhibits fewer FLOPs. However, it does have a larger parameter count. Notably, despite this larger parameter number, SolarFormer achieves a lower rRMSE and a higher FS in almost all scenarios. When the input spatial dimension is increased to 600km, both the FLOPs and the parameter count for SolarFormer rise, but they remain at levels feasible for a single GPU. In contrast, ConvLSTM's FLOPs and parameters surge dramatically, surpassing the memory capabilities of a single GPU. As highlighted in section 5.4.3, future forecasting endeavors, especially over more extended lead times, will necessitate enlarging both input timeframes and image dimensions. Given this context, SolarFormer emerges as a promising candidate, showcasing significant potential for extended period forecasting.

Table 5-3. Performance comparison of SolarFormer and ConvLSTM across various input-output frames and spatial sizes.

			SolarFormer				ConvLSTM			
Input frames	Output frames	Spatial size	FLOPs (G)	#Para (M)	rRmse (%)	FS	FLOPs (G)	#Para (M)	rRmse (%)	FS
8	12	300	22.95	4.81	12.7	34.2	106	2.6	13.0	32.8
8	16	300	27.57	4.81	15.6	30.1	131	2.6	15.4	31.0
4	12	300	18.82	4.81	12.9	34.2	90	2.6	13.3	32.4
4	16	300	23.85	4.81	15.1	31.8	116	2.6	15.4	30.6
8	12	600	113.96	6.56	12.7	34.2	1697	10.6	-	-
8	16	600	141.68	6.59	15.1	32.4	2101	10.6	-	-
4	12	600	95.97	6.59	13.0	33.9	1455	10.6	-	-
4	16	600	121.09	6.59	15.0	32.6	1860	10.6	-	-

5.5 Conclusion

In this study, we created the SolarFormer, a model that combines a BiGRU with a space–time transformer deep learning framework. This integration enables near-real-time nowcasting of GHI, ranging from 15 min to 3 h at 15-min intervals, utilizing geostationary satellite data. Notably, the SolarFormer eliminates the requirement for ground measurement data as inputs and is not limited by the volume of data available for training. The model is able to forecast at image scale in areas covered by both

AHI and ABI. These features make the SolarFormer particularly advantageous for the dispatch modeling of both macro- and micro-energy systems.

The model has been rigorously validated at the image scale and across seven SURFRAD stations. At the image scale, with 15-min intervals and treating BiGRU-generated results as true values, the rRMSE ranges from 3.2% to 18.3%, and the FS ranges from 31.1 to 44.5 (%). When validated against ground measurements at 15-min intervals, the rRMSE ranges from 19.2% to 28.2%. Hourly validation of RMSE (rRMSE) at these stations yields results of 93.82 W/m2 (14.99%), 118.87 W/m2 (19.80%), and 129.10 W/m2 (24.18%) for 1-h, 2-h, and 3-h lead times, respectively. Forecast skills were recorded at 24.63, 25.26, and 24.27 (%) for the corresponding intervals, highlighting the SolarFormer's superiority over the smart persistent model. These metrics rival or surpass those of existing methods and NWP commercial data offerings. Our analysis underscores the critical role of BiGRU's estimation accuracy in enhancing forecasting precision. Higher accuracy in initial conditions invariably leads to improved forecasting outcomes, a dimension that merits further investigation. Furthermore, the SolarFormer's performance is particularly notable under highly variable cloudy conditions and also demonstrates increased accuracy under stable atmospheric conditions beyond the initial 30 min.

The study reveals the potential of SolarFormer to forecast over extended periods. The findings indicate that the SolarFormer reliably sustains its forecasting accuracy over longer durations without significant deterioration. For longer time series forecasting, incorporating broader spatial dimensions and longer input time frames can enhance predictive precision. Comparative analysis also demonstrates SolarFormer's superior performance relative to ConvLSTM regarding accuracy, computational requirements, and memory efficiency. The SolarFormer's efficient architecture enhances its prospects, positioning it as a viable option for longer forecasting horizons in the future.

The proposed algorithm shows promise but could benefit from further enhancements. A notable issue is the accumulation of bias during long lead-time forecasting. Adopting a multi-branch structure could mitigate error accumulation. Additionally, refining the model to extend its forecasting lead time holds significant potential. Separating the loss function to enable high accuracy at each time frame, rather than achieving a global averaged minimum, could enhance long-term forecasting. Regarding data, incorporating large spatial images and longer input time frames is necessary. Furthermore, future models should enable probabilistic forecasting rather than determinate results.

Chapter 6: Conclusion

6.1 Major findings

The dissertation identified the issues and limitation of existing DSR estimation products and traditional estimation methods, and aims to address these issues by exploring various state of art neural network to improve both daily, 15 minute, and instantaneous DSR estimation. The efforts include employing transfer learning concepts by learning from both RTM simulations and ground measurement data, and applying DL methods to utilize spatial and temporal information from high spatial temporal resolution satellite observations. Moreover, the dissertation enables solar radiation nowcasting with lead time from 15min to 3 hours with 15min interval based on the improved high spatial-temporal DSR estimation methods and the space-time transformer method.

Chapter 2 compared five updated global satellite daily DSR products including CERES, CLARA, GLASS, BESS, and MCD18 were evaluated using in situ measurements from 142 global sites with a special focus on high latitudes in 2004. The products were aggregated and compared at various spatial resolutions over different subareas. The overall accuracy increased after the products were aggregated to 100 km. When all products were evaluated at a resolution of 100 km, the global root-mean-square error of CERES, CLARA, GLASS, BESS, and MCD18 was 27.6, 29.1, 30.3, 29.6, and 31.6 W/m^2, respectively, and the mean bias difference was 2.2, −1.5, −1.8, −3.4, and −8.0 W/m^2, respectively. The RMSE still not satisfied the requirements on different DSR applications listed in Chapter 1. The accuracies of most products are ~7 W/m^2lower over high-latitude areas. A seasonal variation of the accuracies was observed for all products. It is particularly pronounced over high-latitude areas. With respect to the long term, both in situ data, BESS, and CERES show insignificant trends, while CLARA and GLASS present dimming trend. Besides, CLARA and GLASS exhibit slight annual changes of -0.250 and -0.387 W/m^2 in the bias and 0.357 and 0.310 W/m^2in the RMSE in the past two decades. GLASS and MCD18 exhibit a superior performance over coastal regions but degrade over snow-covered areas. Potential refinements of current high-resolution DSR retrieval algorithms are suggested, which will improve the retrieval accuracy. Chapter 3 applied the transfer-learning (TL) concept to utilize both radiative transfer simulations and ground measurement data, achieving global DSR estimation with only top-of-atmosphere and surface albedo at local solar noon as inputs. The proposed method estimates both instantaneous and daily DSR from Visible Infrared Imaging Radiometer Suite (VIIRS) data at 750-m resolution, and both the estimates are validated by 40 independent stations globally. The root mean-square error and relative root mean square error of instantaneous DSR validation over 25 Baseline Surface Radiation Network, seven Surface Radiation Network, and eight Greenland Climate Network stations in 2013 were 91.2 (16.1%), 106.3 (18.3%), 75.0 (24.2%) W/m^2, respectively, and the daily validation achieved 30.8 (15.5%), 33.5 (17.6%), and 31.3 (14.4) W/m^2, respectively. The proposed method presents significant high accuracy over polar regions and similar performances over other areas compared with traditional ML models, physics models (e.g., look-up tables and direct estimations), and existing DSR products. The algorithm is also applied to VIIRS swath data to test its global efficacy. Instantaneous mapping captures the spatial pattern of the cloud-mask product, and daily mapping shows spatial patterns similar to the CERES TOA and surface fluxes and clouds product, but with more detail. Further analysis indicates that model performance is less sensitive to the quantity of training data after TL has been incorporated. This study demonstrates the advantages of TL on boosting both the generality and accuracy of DSR estimation, which can potentially be applied to other variable retrievals.

Chapter 4 developed and compared three DL methods, namely the DenseNet, the BiGRU_nor without surface albedo as inputs, and the CNNGRU_nor without surface albedo as inputs. These methods were used to estimate DSR at 1 km and 10/15 min resolutions directly from top-of-atmosphere reflectance over the Advanced Himawari Imager (AHI) onboard Himawari-8 and the Advanced Baseline Imager (ABI) onboard GOES-16 coverage, achieving high accuracies. The instantaneous root mean square error (RMSE) and relative RMSE for the three models were 68.4 (16.1%), 69.4 (16.3%), and 67.1 (15.7%) W/m^2, respectively, which are lower than the baseline machine learning method, the multilayer perceptron model (MLP), with RMSE at 76.8 W/m^2 (18.0%). Hourly accuracies for the three DL methods were 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.8%) W/m^2, which are within the DSR RMSEs that we estimated for existing datasets of the Earth's Radiant Energy System (CERES) (88.8 W/m^2, 21.4%) and GeoNEX (77.8 W/m^2, 18.8%). The study illustrates that DL models that incorporate temporal information can eliminate the need for surface albedo as an input, which is crucial for timely monitoring and nowcasting of DSR. Incorporating spatial information can enhance the retrieval accuracy in overcast conditions, and incorporating infrared bands can further improve the accuracy of DSR estimation.

Chapter 5 employs a space-time Transformer model to address the challenge of forecasting solar radiation in a 0-3 hour lead time at 15-minute intervals using only geostationary satellite observations. For hourly predictions, the proposed SolarFormer model achieves RMSE (rRMSE) values of 93.8 W/m^2 (15.0%), 118.9 W/m^2 (19.8%), and 129.1 W/m^2 (24.2%) in validation across seven SURFRAD stations. This accuracy surpasses existing machine learning models and is on par with numerical weather predictions and commercial products. Furthermore, the study highlights the efficiency and memory capabilities of the Transformer, indicating its potential for longer lead-time forecasting.

6.2 Future research

The dissertation has explored various ML and DL methods on improving DSR estimation and forecasting. Building on the developed forecasting products and the insights into integrating solar radiation with DL, future research are aims on three perspectives: enhancing the estimation of the diffuse component of DSR, investigating the broader implications of DSR on various sectors in the context of climate change, and furthering the domain of scientific machine learning.

The diffuse component of DSR plays an integral role in processes such as photosynthesis and solar energy conversion. However, current methods often falter in accurately estimating this component. Traditional physical methods show signs of degradation, and machine learning models face issues of insufficient insitu measurement. To address these challenges, upcoming research will harness physical-guided DL methods, merging the strengths of both physical models and machine learning. This approach seeks to offer a refined estimation of the diffuse portion, tailored for high spatial-temporal resolutions.

Turning to applications of DSR, its most immediate utility is evident in solar power conversion, particularly in photovoltaic systems. With the aid of high-resolution DSR estimations, studies into the geophysical constraints of renewable energy are feasible. Future research, combining the derived DSR product with macro-energy models and other energy sources, will emphasize large-scale renewable energy dispatch and installation design, aiming to circumvent the current limitations posed by reliance on reanalysis data (Davidson et al., 2022). Moreover, given its important role in various earth cycles, DSR's influence sectors like crop yield, carbon sequestration, and early drought detection. However, the altering landscape of DSR, spurred by climate change and escalating anthropogenic pollution (He et al., 2022), necessitates an in-depth understanding of its evolving interplay with the broader earth system (Sweerts et al., 2019).

Scientific machine learning is another pathway of my future research. The dissertation shows the ability of space time transformer on solar radiation forecasting, highlighting its latent potential for long-term predictions. The same mechanism can be used to forecast various other earth phenomena, ranging from aerosols and wildfires to hurricanes (Bi et al., 2023). Furthermore, the transfer learning applications showcased in this dissertation seamlessly combine the benefits of training on both in-situ data and simulations, merging physical knowledge into the model and bolstering its generalizability. Subsequent research will continue to unearth the potential of physics guided machine learning. Physics-based models surpass DL in terms of explainability and lower data requirements, whereas DL excels over physics-based models in computing efficiency especially with higher resolution and its capacity to capture complex relationships. Physics-informed machine learning offers a seamless integration of data with governing physical laws, accommodating even models with partially missing physics, in a cohesive manner. Future research direction will tackle earth science challenges by encompassing parameterization, inverse modeling, and the resolution of partial differential equations (Willard et al., 2020; Shen et al., 2023).

Additionally, insitu data is essential for the whole dissertation, no matter for training new models or model validations. Ground measurements of DSR is sufficient compared with other environmental studies, but the collected data is not evenly distributed spatially as well as over different surface and climate types. This issue can be better address both from data and algorithm perspectives. In terms of algorithms, spatial fairness mechanisms will be considered in the future model development. More strict validation approaches like leaving out some surface types for independent testing will be included. Collecting more data as well as detailed data preprocessing will also be essential for model developments. More data over high latitude regions, western Asia, as well as Eastern African will greatly boost the model generalization and the understanding of its ability at global scale.

Bibliography

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

Augustine, J. A., DeLuisi, 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.

Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. Solar Energy, 198(1), 81–92. https://doi.org/10.1016/j.solener.2020.01.034

Baldridge, A. M., Hook, S. J., Grove, C. I., & Rivera, G. (2009). The aster spectral library version 2.0. Remote Sensing of Environment, 113(4), 711–715. https://doi.org/10.1016/j.rse.2008.11.007

Bansal, A. S., Bansal, T., & Irwin, D. (2022). A moment in the sun: solar nowcasting from multispectral satellite data using self-supervised learning. In Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (pp. 251-262).

Berk, A., Cooley, T. W., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., ... & Lewis, P. E. (2004). MODTRAN5: A reformulated atmospheric band model with auxiliary species and practical multiple scattering options. In Remote Sensing of Clouds and the Atmosphere IX (Vol. 5571, pp. 78-85). International Society for Optics and Photonics.

Bertasius, G., Wang, H., & Torresani, L. (2021). Is space-time attention all you need for video understanding?. In ICML (Vol. 2, No. 3, p. 4).

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970), 533-538.

Boilley, A., & Wald, L. (2015). Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface. Renewable Energy, 75, 135–143. https://doi.org/10.1016/j.renene.2014.09.042

Brown, M. G. L., Skakun, S., He, T., & Liang, S. (2020). Intercomparison of machine-learning methods for estimating surface shortwave and photosynthetically active radiation. Remote Sensing, 12(3), 372. https://doi.org/10.3390/rs12030372

Bue, B. D., Thompson, D. R., Deshpande, S., Eastwood, M., Green, R. O., Natraj, V., Mullen, T., & Parente, M. (2019). Neural network radiative transfer for imaging spectroscopy. Atmospheric Measurement Techniques, 12(4), 2567–2578. https://doi.org/10.5194/amt-12-2567-2019

Buster, G., Rossol, M., Maclaurin, G., Xie, Y., & Sengupta, M. (2021). A physical downscaling algorithm for the generation of high-resolution spatiotemporal solar irradiance data. Solar Energy, 216, 508-517

Chang, T., & Xiong, X. (2019). GOES-16/ABI thermal emissive band assessments using GEO-LEO-GEO double difference. Earth and Space Science, 6(12), 2303-2316.

Chai, D., Newsam, S., Zhang, H. K., Qiu, Y., & Huang, J. (2019). Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. Remote sensing of environment, 225, 307-316.

Chen, J., He, T., Jiang, B., & Liang, S. (2020). Estimation of all-sky all-wave daily net radiation at high latitudes from modis data. Remote Sensing of Environment, 245, 111842. https://doi.org/10.1016/j.rse.2020.111842

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

Chen, J., Zhu, W., & Yu, Q. (2022). High-spatiotemporal-resolution estimation of solar energy component in the United States using a new satellite-based model. Journal of Environmental Management, 302, 114077.

Chen, L., Yan, G., Wang, T., Ren, H., Calbó, J., Zhao, J., & McKenzie, R. (2012). Estimation of surface shortwave radiation components under all sky conditions: Modeling and sensitivity analysis. Remote Sensing of Environment, 123, 457-469.

Chen, L., Yan, G., Wang, T., Ren, H., Hu, R., Chen, S., & Zhou, H. (2019). Spatial scale consideration for estimating all-sky surface shortwave radiation with a modified 1-D radiative transfer model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(3), 821-835.

Cho, K., Van Merri¨enboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation arXiv preprint arXiv:1406.1078.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258.

Choi, M., Rachunok, B., & Nateghi, R. (2021). Short-term solar irradiance forecasting using convolutional neural networks and cloud imagery. Environmental Research Letters, 16(4), 044045.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., & Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy, 85(11), 2881-2893.

Clark, R. N., Swayze, G. A., Wise, R. A., Livo, K. E., Hoefen, T. M., Kokaly, R. F., & Sutley, S. J. (2007). USGS Digital Spectral Library splib06a (Data Series). US Geological Survey. http://dx.doi.org/10.3133/ds231

Davidson, M. R., & Millstein, D. (2022). Limitations of reanalysis data for wind power applications. Wind Energy, 25(9), 1646-1653.

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., Wielicki, B. A., Young, D. F., & Sun, M. (2013). Geostationary Enhanced Temporal Interpolation for CERES Flux Products. Journal of Atmospheric and Oceanic Technology, 30(6), 1072–1090. https://doi.org/10.1175/JTECH-D-12-00136.1

Dong, Z., Yang, D., Reindl, T., & Walsh, W. M. (2013). Short-term solar irradiance forecasting using exponential smoothing state space model. Energy, 55, 1104-1113.

Dupont, J. C., Haeffelin, M., & Long, C. N. (2009). Cirrus cloud radiative effect on surface-level shortwave and longwave irradiances at regional and global scale. Atmospheric Chemistry and Physics Discussions, 9(6), 26777-26832.

Earth Resources Observation And Science (EROS) Center. (2017). Global 30 Arc-Second Elevation (GTOPO30), U.S. 465 Geological Survey [Data set], https://doi.org/10.5066/F7DF6PQS.

FMI. (2016). Finish Meteorological Institute. http://en.ilmatieteenlaitos.fi.

Fu, C. L., & Cheng, H. Y. (2013). Predicting solar irradiance with all-sky image features via regression. Solar Energy, 97, 537-550.

Gao, Z., Shi, X., Wang, H., Zhu, Y., Wang, Y. B., Li, M., & Yeung, D. Y. (2022). Earthformer: Exploring space-time transformers for earth system forecasting. Advances in Neural Information Processing Systems, 35, 25390-25403.

Gallo, R., Castangia, M., Macii, A., Macii, E., Patti, E., & Aliberti, A. (2022). Solar radiation forecasting with deep learning techniques integrating geostationary satellite images. Engineering Applications of Artificial Intelligence, 116, 105493.

Gandhi, O., Kumar, D. S., Rodríguez-Gallegos, C. D., & Srinivasan, D. (2020). Review of power system impacts at high PV penetration Part I: Factors limiting PV penetration. Solar Energy, 210, 181-201.

Gueymard, C. A. (2014). A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. Renewable and Sustainable Energy Reviews, 39, 1024–1034. https://doi.org/10.1016/j.rser.2014.07.117

Gui, S., Liang, S., Wang, K., Li, L., & Zhang, a. X. (2010). Assessment of Three Satellite-Estimated Land Surface Downwelling Shortwave Irradiance Data Sets. IEEE Geoscience and Remote Sensing Letters, 7(4), 776–780. https://doi.org/10.1109/LGRS.2010.2048196

Hatzianastassiou, N., Matsoukas, C., Fotiadi, A., Pavlakis, K. G., Drakakis, E., Hatzidimitriou, D., & Vardavas, I. (2005). Global distribution of Earth’s surface shortwave radiation budget. Atmospheric Chemistry and Physics, 5, 2847–2867.

Ham, S. H., Kato, S., Barker, H. W., Rose, F. G., & Sun-Mack, S. (2014). Effects of 3-D clouds on atmospheric transmission of solar radiation: Cloud type dependencies inferred from A-train satellite data. Journal of Geophysical Research: Atmospheres, 119(2), 943-963.

Han, J., & Moraga, C. (1995, June). The influence of the sigmoid function parameters on the speed of backpropagation learning. In International workshop on artificial neural networks (pp. 195-201). Springer, Berlin, Heidelberg.

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

He, T., Liang, S., Wang, D., Cao, Y., Gao, F., Yu, Y., & Feng, M. (2018). Evaluating land surface albedo estimation from Landsat MSS, TM, ETM+, and OLI data based on the unified direct estimation approach. Remote Sensing of Environment, 204, 181-196.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

He, L., Wei, J., Wang, Y., Shang, Q., Liu, J., Yin, Y., ... & Yung, Y. L. (2022). Marked impacts of pollution mitigation on crop yields in China. Earth's Future, 10(11), e2022EF002936.

Hinkelman, L. M., Stackhouse Jr, P. W., Wielicki, B. A., Zhang, T., & Wilson, S. R. (2009). Surface insolation trends from satellite and ground measurements: Comparisons and challenges. Journal of Geophysical Research, 114(D10).

Hinton, G., Srivastava, N., & Swersky, K. (2012). Lecture 6a Overview of Mini–Batch Gradient Descent, Coursera Lecture Slides. [Online]. Available: https://class.coursera.org/neuralnets-2012-001/lecture

Ho, J., Kalchbrenner, N., Weissenborn, D., & Salimans, T. (2019). Axial attention in multidimensional transformers. arXiv preprint arXiv:1912.12180.

Hollmann, R., Mueller, R.W., Gratzki, A., 2006. CM-SAF surface radiation budget: first results with AVHRR data. Atmospheric Remote Sensing: Earth's Surface, Troposphere, Stratosphere and Mesosphere - Ii 37, 2166–2171.

Hou, N., Zhang, X., Zhang, W., Wei, Y., Jia, K., Yao, Y., Jiang, B., & Cheng, J. (2020). Estimation of surface downward shortwave radiation over china from himawari-8 ahi data based on random forest. Remote Sensing, 12(1), 181. https://doi.org/10.3390/rs12010181

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences, 79(8), 2554-2558.

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

Huang, G., Li, X., Huang, C., Liu, S., Ma, Y., & Chen, H. (2016). Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products. Remote sensing of environment, 181, 198-206.

Huang, G., Liang, S., Lu, N., Ma, M., Wang, D., (2018). Toward a Broadband Parameterization Scheme for Estimating Surface Solar Irradiance: Development and Preliminary Results on MODIS Products. J. Geophys. Res.-Atmos. 123, 12180–12193.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., & Zhang, Y. (2019). Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sensing of Environment, 233, 111371.

Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. Remote sensing of environment, 229, 247-259.

Jia, B., Xie, Z., Dai, A., Shi, C., & Chen, F. (2013). Evaluation of satellite and reanalysis products of downward surface solar radiation over East Asia: Spatial and seasonal variations. Journal of Geophysical Research: Atmospheres, 118(9), 3431–3446. https://doi.org/10.1002/jgrd.50353

Jiang, H., Lu, N., Qin, J., Tang, W., & Yao, L. (2019). A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renewable and Sustainable Energy Reviews, 114, 109327. https://doi.org/10.1016/j.rser.2019.109327

Jia, A., Wang, D., Liang, S., Peng, J., & Yu, Y. (2022a). Global Daily Actual and Snow-Free Blue-Sky Land Surface Albedo Climatology From 20-Year MODIS Products. Journal of Geophysical Research: Atmospheres, 127(8), e2021JD035987.

Jia, A., Liang, S., & Wang, D. (2022b). Generating a 2-km, all-sky, hourly land surface temperature product from Advanced Baseline Imager data. Remote Sensing of Environment, 278, 113105.

Jia, A., Liang, S., Wang, D., Ma, L., Wang, Z., & Xu, S. (2023). Global hourly, 5 km, all-sky land surface temperature data from 2011 to 2021 based on integrating geostationary and polar-orbiting satellite data. Earth System Science Data, 15(2), 869-895.

Jiang, H., Lu, N., Qin, J., Tang, W., & Yao, L. (2019). A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renewable and Sustainable Energy Reviews, 114, 109327

Jiang, H., Lu, N., Huang, G., Yao, L., Qin, J., & Liu, H. (2020). Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. Applied Energy, 270, 115178.

Kapsch, M.-L., Graversen, R. G., Tjernström, M., & Bintanja, R. (2016). The Effect of Downwelling Longwave and Shortwave Radiation on Arctic Summer Sea Ice. Journal of Climate, 29(3), 1143–1159. https://doi.org/10.1175/JCLI-D-15-0238.1

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., & Hollmann, R. (2017). CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. Atmospheric Chemistry and Physics, 17(9), 5809–5828. https://doi.org/10.5194/acp-17-5809-2017

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J. F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., & Wolters, E. (2013). CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data. Atmospheric Chemistry and Physics, 13(10), 5351–5367. https://doi.org/10.5194/acp-13-5351-2013

Kalman, B. L., & Kwasny, S. C. (1992, June). Why tanh: choosing a sigmoidal function. In [Proceedings 1992] IJCNN International Joint Conference on Neural Networks (Vol. 4, pp. 578-581). IEEE.

Kato, S., Loeb, N. G., Doelling, D. R., Rose, F. G., Rutan, D. A., Caldwell, T. E., Yu, L., & Weller, R. A. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. Journal of Climate, 26(9), 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

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. https://doi.org/10.1038/s41598-021-89779-z

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

Letu, H., Nakajima, T. Y., Wang, T., Shang, H., Ma, R., Yang, K., ... & Shi, J. (2022). A new benchmark for surface radiation products over the East Asia–Pacific region retrieved from the Himawari-8/AHI next-generation geostationary satellite. Bulletin of the American Meteorological Society, 103(3), E873-E888.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

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

Li, R., Wang, D., & Liang, S. (2021). Comprehensive assessment of five global daily downward shortwave radiation satellite products. Science of Remote Sensing, 4, 100028.

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.

Li, R., Wang, D., Wang, W., & Nemani, R. (2023a). A GeoNEX-based high-spatiotemporal-resolution product of land surface downward shortwave radiation and photosynthetically active radiation. Earth System Science Data, 15(3), 1419-1436.

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

Li, Z., Leighton, H. G., Masuda, K., & Takashima, T. (1993). Estimation of SW flux absorbed at the surface from TOA reflected flux. Journal of Climate, 6(2), 317-330.

Liang, S., Wang, D., He, T., Yu, Y., 2019. Remote sensing of earth's energy budget: synthesis and review. Int. J. Digital Earth 12, 737–780.

Liang, S., Wang, K., Zhang, X., Wild, M., 2010. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 3 (3), 225–240. https://doi.org/10.1109/JSTARS.2010.2048556.

Liang, S., Zheng, T., Liu, R., Fang, H., Tsay, S.C., Running, S., 2006. Estimation of incident photosynthetically active radiation from Moderate Resolution Imaging Spectrometer data. J. Geophys. Res.-Atmos. 111.

Lin, F., Zhang, Y., & Wang, J. (2023). Recent advances in intra-hour solar forecasting: A review of ground-based sky image methods. International Journal of Forecasting, 39(1), 244-265.

Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., & Hu, H. (2022). Video swin transformer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3202-3211).

Lorenz, E., Heinemann, D., Wickramarathne, H., Beyer, H.G., Bofinger, S., (2007). Forecast of ensemble power production by grid-connected PV systems. In: Proc. 20th European PV Conference, Milano, Italy.

Ma, R., Letu, H., Yang, K., Wang, T., Shi, C., Xu, J., Shi, J., Shi, C., & Chen, L. (2020). Estimation of surface shortwave radiation from himawari-8 satellite data based on a combination of radiative transfer and deep neural network. IEEE Transactions on Geoscience and Remote Sensing, 58(8), 5304–5316. https://doi.org/10.1109/TGRS.2019.2963262

Malmgren-Hansen, D., Kusk, A., Dall, J., Nielsen, A. A., Engholm, R., & Skriver, H. (2017). Improving sar automatic target recognition models with transfer learning from simulated data. IEEE Geoscience and Remote Sensing Letters, 14(9), 1484–1488. https://doi.org/10.1109/LGRS.2017.2717486

Marcos, J., Storkël, O., Marroyo, L., Garcia, M., & Lorenzo, E. (2014). Storage requirements for PV power ramp-rate control. Solar Energy, 99, 28-35.

Menzel, W. P., & Strabala, K. (1997). Cloud top properties and cloud phase algorithm theoretical basis document. University of Wisconsin--Madison.

Miller, S. D., Rogers, M. A., Haynes, J. M., Sengupta, M., & Heidinger, A. K. (2018). Short-term solar irradiance forecasting via satellite/model coupling. Solar Energy, 168, 102-117.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019a. Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Canada Glacier Meteorological Station (CAAM) in Taylor Valley, Antarctica from 1994 to Present. https://doi.org/10.6073/PASTA/6B40F960061394C157C89B111D963492.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019b. Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Lake Bonney Meteorological Station (BOYM) in Taylor Valley, Antarctica from 1993 to Present.

https://doi.org/10.6073/PASTA/3220523D0AAB8489375A9997856EC1A9. McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019c.

Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Lake Fryxell Meteorological Station (FRLM) in Taylor Valley, Antarctica from 1993 to Present. https://doi.org/10.6073/PASTA/1AC8DA354BFF946F8CB14EEC393737F0.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019d. Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Lake Hoare Meteorological Station (HOEM) in Taylor Valley, Antarctica from 1987 to Present. https://doi.org/10.6073/PASTA/BE2F27D49D9E0289FFD64871A6E1F3D1.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019e. Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Lake Vida Meteorological Station (VIAM) in Victoria Valley, Antarctica from 1995 to Present. https://doi.org/10.6073/PASTA/8864645B519274A28F6F94D9619A2958.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019f. Mcmurdo Dry Valleys LTER: Daily Measurement Summaries from Miers Valley Meteorological Station (MISM) in Miers Valley, Antarctica from 2012 to Present. https://doi.org/10.6073/PASTA/65234DF2904F948F1EBCEBF12CB27D09.

McMurdo Dry Valleys LTER, Doran, P., Fountain, A., 2019g. Mcmurdo Dry Valleys LTER: daily measurement summaries from taylor glacier meteorological station (TARM) in taylor valley, Antarctica from 1994 to present. https://doi.org/10.6073/PASTA/F0F5860D88BEEF27032CEF66C0BC363B.

Mueller, R. W., Matsoukas, C., Gratzki, A., Behr, H. D., & Hollmann, R. (2009). The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance — A LUT based eigenvector hybrid approach. Remote Sensing of Environment, 113(5), 1012–1024. https://doi.org/10.1016/j.rse.2009.01.012

Muneer, T., Younes, S., & Munawwar, S. (2007). Discourses on solar radiation modeling: AS. Renewable and Sustainable Energy Reviews, 11(4), 551–602. https://doi.org/10.1016/j.rser.2005.05.006

NIBIO-LMT. (2016). Norwegian Institute of Bioceonomy Research) - Landbruksmeteorologisk Tjeneste. http://lmt.nibio.no/

Nguyen, A., Velay, M., Schoene, J., Zheglov, V., Kurtz, B., Murray, K., ... & Kleissl, J. (2016). High PV penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations. Solar Energy, 132, 221-235.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Palmintier, B. (2015). Integrated Grid Modeling System (IGMS) for Combined Transmission and Distribution Simulation (No. NREL/PR-5D00-65552). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., . . . Papale, D. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific Data, 7(1), 225. https://doi.org/10.1038/s41597-020-0534-3

Peng, Z., Letu, H., Wang, T., Shi, C., Zhao, C., Tana, G., Zhao, N., Dai, T., Tang, R., Shang, H., Shi, J., & Chen, L. (2020). Estimation of shortwave solar radiation using the artificial neural network from himawari-8 satellite imagery over china. Journal of Quantitative Spectroscopy and Radiative Transfer, 240, 106672. https://doi.org/10.1016/j.jqsrt.2019.106672

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K., Renné, D., & Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy, 84(12), 2161-2172.

Perez, R., Kankiewicz, A., Schlemmer, J., Hemker, K., & Kivalov, S. (2014). A new operational solar resource forecast model service for PV fleet simulation. In 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC) (pp. 0069-0074). IEEE.

Pérez, E., Pérez, J., Segarra-Tamarit, J., & Beltran, H. (2021). A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. Solar Energy, 218, 652-660.

Pfenninger, S., & Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy, 114, 1251-1265.

Pinker, Frouin, Li, 1995. A review of satellite methods to derive surface shortwave irradiance. Remote Sens. Environ. 51 (1), 108–124.

Qin, J., Jiang, H., Lu, N., Yao, L., & Zhou, C. (2022). Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning. Renewable and Sustainable Energy Reviews, 167, 112680.

Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar energy, 83(3), 342-349.

Riihel¨a, A., Key, J.R., Meirink, J.F., Kuipers Munneke, P., Palo, T., Karlsson, K.-G., 2017. An intercomparison and validation of satellite-based surface radiative energy flux estimates over the Arctic. J. Geophys. Res.: Atmosphere 122 (9), 4829–4848. https://doi.org/10.1002/2016JD026443.

Roesch, A., Wild, M., Ohmura, A., Dutton, E.G., Long, C.N., Zhang, T., 2011. Assessment of BSRN radiation records for the computation of monthly means: BSRN. Atmospheric Measurement Techniques 4 (2), 339–354. https://doi.org/10.5194/amt-4-339-2011.

Rosenstein, M. T., Marx, Z., Kaelbling, L. P., & Dietterich, T. G. (2005). To transfer or not to transfer. In NIPS 2005 workshop on transfer learning (Vol. 898, pp. 1-4).

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Le Nguyen, T., Caldwell, T. E., & Loeb, N. G. (2015). CERES Synoptic Product: Methodology and Validation of Surface

Radiant Flux: CERES1. *Journal of Atmospheric and Oceanic Technology*, *32*(6), 1121–1143. https://doi.org/10.1175/JTECH-D-14-00165.1

Ryu, Y., Jiang, C., Kobayashi, H., & Detto, M. (2018). Modis-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. Remote Sensing of Environment, 204(D09104), 812–825. https://doi.org/10.1016/j.rse.2017.09.021

Sanchez-Lorenzo, A., Wild, M., Brunetti, M., Guijarro, J. A., Hakuba, M. Z., Calbó, J., Mystakidis, S., & Bartok, B. (2015). Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). Journal of Geophysical Research: Atmospheres, 120(18), 9555–9569. https://doi.org/10.1002/2015JD023321

Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., ... & Kalchbrenner, N. (2020). Metnet: A neural weather model for precipitation forecasting. arXiv preprint arXiv:2003.12140.

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X.W., Tsang, T., Strugnell, N.C., Zhang, X.Y., Jin, Y.F., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R.P., Hu, B.X., Liang, S.L., Privette, J.L., Roy, D., (2002). First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sens. Environ. 83, 135–148.

Schaaf, C. and Wang, Z. (2015). MCD43A3 MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global - 500m V006, NASA 550 EOSDIS Land Processes DAAC [Data set], https://doi.org/10.5067/MODIS/MCD43A3.006

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681.

Sharma, V., Yang, D., Walsh, W., & Reindl, T. (2016). Short term solar irradiance forecasting using a mixed wavelet neural network. Renewable Energy, 90, 481-492.

Sengupta, A., Ye, Y., Wang, R., Liu, C., Roy, K., 2019. Going deeper in spiking neural networks: VGG and residual architectures. Front. Neurosci. 13, 95.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., ... & Lawson, K. (2023). Differentiable modeling to unify machine learning and physical models and advance Geosciences. arXiv preprint arXiv:2301.04027.

Shi, H., Xiao, Z., & Tian, X. (2019). Exploration of machine learning techniques in emulating a coupled soil–canopy–atmosphere radiative transfer model for multi-parameter estimation from satellite observations. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 8522–8533. https://doi.org/10.1109/TGRS.2019.2921392

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and computers in simulation, 55(1-3), 271-280.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sweerts, B., Pfenninger, S., Yang, S., Folini, D., Van der Zwaan, B., & Wild, M. (2019). Estimation of losses in solar energy production from air pollution in China since 1960 using surface radiation data. Nature Energy, 4(8), 657-663.

Steffen, K., Box, J. E., & Abdalati, W. (1996). Greenland climate network: GC-Net. US Army Cold Regions Reattach and Engineering (CRREL), CRREL Special Report, 98–103.

Stephens, G.L., Li, J., Wild, M., Clayson, C.A., Loeb, N., Kato, S., L'Ecuyer, T., Stackhouse, P.W., Lebsock, M., Andrews, T., 2012. An update on Earth's energy balance in light of the latest global observations. Nat. Geosci. 5 (10), 691–696. https://doi.org/10.1038/ngeo1580.

Sun, D., Ji, C., Sun, W., Yang, Y., & Wang, H. (2018). Accuracy assessment of three remote sensing shortwave radiation products in the Arctic: Arctic. Atmospheric Research, 212, 296–308. https://doi.org/10.1016/j.atmosres.2018.01.003

U.S. National Ice Center. (2008). IMS Daily Northern Hemisphere Snow and Ice Analysis at 1 km, 4 km, and 24 km Resolutions, Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N52R3PMC

Urraca, R., Gracia-Amillo, A. M., Koubli, E., Huld, T., Trentmann, J., Riihelä, A., Lindfors, A. V., Palmer, D., Gottschalg, R., & Antonanzas-Torres, F. (2017). Extensive validation of CM SAF surface radiation products over Europe. Remote Sensing of Environment, 199, 171–186. https://doi.org/10.1016/j.rse.2017.07.013

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, 116(D8), 2003. https://doi.org/10.1029/2009JD013337

Tang, W.J., Yang, K., Qin, J., Li, X., Niu, X.L., (2019). A 16-year dataset (2000-2015) of high-resolution (3 h, 10 km) global surface solar radiation. Earth System Science Data 11, 1905–1915.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., ... & Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences, 13(14), 4291-4313.

Tong, D., Farnham, D. J., Duan, L., Zhang, Q., Lewis, N. S., Caldeira, K., & Davis, S. J. (2021). Geophysical constraints on the reliability of solar and wind power worldwide. Nature communications, 12(1), 1-12.

Van Laake, P.E., Sanchez-Azofeifa, G.A., 2004. Simplified atmospheric radiative transfer modelling for estimating incident PAR using MODIS atmosphere products. Remote Sens. Environ. 91, 98–113.

Wang, D. (2021). MODIS/Terra+Aqua Surface Radiation Daily/3-Hour L3 Global 1km SIN Grid V061. https://doi.org/10.5067/MODIS/MCD18A1.061

Wang, D., & Liang, S. (2014). Improving LAI Mapping by Integrating MODIS and CYCLOPES LAI Products Using Optimal Interpolation. IEEE Journal of Selected

Topics in Applied Earth Observations and Remote Sensing, 7(2), 445–457. https://doi.org/10.1109/jstars.2013.2264870

Wang, D., Liang, S., He, T., & Shi, Q. (2015). Estimation of daily surface shortwave net radiation from the combined MODIS data. IEEE Transactions on Geoscience and Remote Sensing, 53(10), 5519–5529.

Wang, D., Liang, S., Zhang, Y., Gao, X., Brown, M. G., & Jia, A. (2020). A new set of MODIS land products (MCD18): Downward shortwave radiation and photosynthetically active radiation. Remote Sensing, 12(1), 168.

Wang, D., Liang, S., Li, R., & Jia, A. (2021). A synergic study on estimating surface downward shortwave radiation from satellite data. Remote Sensing of Environment, 264, 112639.

Wang, T., Shi, J., Husi, L., Zhao, T., Ji, D., Xiong, C., & Gao, B. (2017). Effect of solar-cloud-satellite geometry on land surface shortwave radiation derived from remotely sensed data. Remote Sensing, 9(7), 690.

Wang, W., Li, S., Hashimoto, H., Takenaka, H., Higuchi, A., Kalluri, S., & Nemani, R. (2020). An introduction to the Geostationary-NASA Earth Exchange (GeoNEX) Products: 1. Top-of-atmosphere reflectance and brightness temperature. Remote Sensing, 12(8), 1267.

Wei, Y., Zhang, X., Hou, N., Zhang, W., Jia, K., & Yao, Y. (2019). Estimation of surface downward shortwave radiation over china from avhrr data based on four machine learning methods. Solar Energy, 177(10), 32–46. https://doi.org/10.1016/j.solener.2018.11.008

Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer knowledge between cities. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1905-1914).

Wild, M., Trüssel, B., Ohmura, A., Long, C. N., König-Langlo, G., Dutton, E. G., & Tsvetkov, A. (2009). Global dimming and brightening: An update beyond 2000. Journal of Geophysical Research: Atmospheres, 114(D10).

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. arXiv preprint arXiv:2003.04919, 1(1), 1-34.

Wyser, K., O'Hirok, W., & Gautier, C. (2005). A simple method for removing 3-D radiative effects in satellite retrievals of surface irradiance. Remote sensing of environment, 94(3), 335-342.

Wu, F., & Fu, C. (2011). Assessment of GEWEX/SRB version 3.0 monthly global radiation dataset over China. Meteorology and Atmospheric Physics, 112(3-4), 155–166. https://doi.org/10.1007/s00703-011-0136-x

Xia, X. A., Wang, P. C., Chen, H. B., & Liang, F. (2006). Analysis of downwelling surface solar radiation in China from National Centers for Environmental Prediction reanalysis, satellite estimates, and surface observations. Journal of Geophysical Research, 111(D9), L06803. https://doi.org/10.1029/2005JD006405

Yang, D., Jirutitijaroen, P., & Walsh, W. M. (2012). Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy, 86(12), 3531-3543.

Yang, D. (2022). Correlogram, predictability error growth, and bounds of mean square error of solar irradiance forecasts. Renewable and Sustainable Energy Reviews, 167, 112736. Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K., Renné, D., & Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy, 84(12), 2161-2172.

Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., & Jia, A. (2018). Estimating surface downward shortwave radiation over china based on the gradient boosting decision tree method. Remote Sensing, 10(2), 185. https://doi.org/10.3390/rs10020185

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?

Yu, Y., Shi, J., Wang, T., Letu, H., Yuan, P., Zhou, W., & Hu, L. (2019). Evaluation of the Himawari-8 Shortwave Downward Radiation (SWDR) Product and its Comparison With the CERES-SYN, MERRA-2, and ERA-Interim Datasets. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(2), 519–532. https://doi.org/10.1109/JSTARS.2018.2851965

Yu, Y., Wang, T., & Shi, J. (Eds.) (2018). Assessment of two Satellite-Based Land Surface Shortwave Downward Radiation Datasets Over the Tibetan Plateau. IEEE.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment, 241(11), 111716. https://doi.org/10.1016/j.rse.2020.111716

Zhang, H., Dong, X., Xi, B., Xin, X., Liu, Q., He, H., ... & Yu, S. (2021). Retrieving high-resolution surface photosynthetically active radiation from the MODIS and GOES-16 ABI data. Remote Sensing of Environment, 260, 112436.

Zhang, T., Stackhouse, P. W., Gupta, S. K., Cox, S. J., Colleen Mikovitz, J., & Hinkelman, L. M. (2013). The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. Journal of Quantitative Spectroscopy and Radiative Transfer, 122, 127–140. https://doi.org/10.1016/j.jqsrt.2012.10.004

Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B., & Cheng, J. (2016). Evaluation of the Reanalysis Surface Incident Shortwave Radiation Products from NCEP, ECMWF, GSFC, and JMA Using Satellite and Surface Observations. Remote Sensing, 8(3), 225. https://doi.org/10.3390/rs8030225

Zhang, X., Liang, S., Wild, M., & Jiang, B. (2015). Analysis of surface incident shortwave radiation from four satellite products. Remote Sensing of Environment, 165, 186–202. https://doi.org/10.1016/j.rse.2015.05.015

Zhang, X., Liang, S., Zhou, G., Wu, H., & Zhao, X. (2014). Generating Global LAnd Surface Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. Remote Sensing of Environment, 152, 318–332. https://doi.org/10.1016/j.rse.2014.07.003

Zhang, X., Zhao, X., Li, W., Liang, S., Wang, D., Liu, Q., Yao, Y., Jia, K., He, T., Jiang, B., Wei, Y., & Ma, H. (2019). An Operational Approach for Generating the Global Land Surface Downward Shortwave Radiation Product From MODIS Data. IEEE Transactions on Geoscience and Remote Sensing, 57(7), 4636–4650. https://doi.org/10.1109/TGRS.2019.2891945

Zhang, Y., Liang, S., He, T., Wang, D., Yu, Y., & Ma, H. (2020). Estimation of land surface incident shortwave radiation from geostationary advanced Himawari imager and advanced baseline imager observations using an optimization method. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-11.

Zhang, Y., Liang, S., & He, T. (2022). Estimation of Land Surface Downward Shortwave Radiation Using Spectral-Based Convolutional Neural Network Methods: A Case Study From the Visible Infrared Imaging Radiometer Suite Images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-15.

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. https://doi.org/10.1109/TGRS.2020.2983201