






Article

Generating High Spatial and Temporal Surface Albedo with Multispectral-Wavemix and Temporal-Shift Heatmaps

Sagthitharan Karalasingham ^{1,2} , Ravinesh C. Deo ^{1,2,*} , Nawin Raj ¹ , David Casillas-Perez ³ 
and Sancho Salcedo-Sanz ^{1,4} 

- ¹ Artificial Intelligence Applications Laboratory, School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia; d9630120@uamail.usq.edu.au (S.K.); nawin.raj@unisq.edu.au (N.R.); sancho.salcedo@uah.es (S.S.-S.)
- ² Centre for Applied Climate Science, University of Southern Queensland, Toowoomba, QLD 4350, Australia
- ³ Department of Signal Processing and Communications, Universidad Rey Juan Carlos, 28942 Fuenlabrada, Madrid, Spain; david.casillas@urjc.es
- ⁴ Department of Signal Processing and Communications, Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain
- * Correspondence: ravinesh.deo@unisq.edu.au

Abstract: Surface albedo is a key variable influencing ground-reflected solar irradiance, which is a vital factor in boosting the energy gains of bifacial solar installations. Therefore, surface albedo is crucial towards estimating photovoltaic power generation of both bifacial and tilted solar installations. Varying across daylight hours, seasons, and locations, surface albedo is assumed to be constant across time by various models. The lack of granular temporal observations is a major challenge to the modeling of intra-day albedo variability. Though satellite observations of surface reflectance, useful for estimating surface albedo, provide wide spatial coverage, they too lack temporal granularity. Therefore, this paper considers a novel approach to temporal downscaling with imaging time series of satellite-sensed surface reflectance and limited high-temporal ground observations from surface radiation (SURFRAD) monitoring stations. Aimed at increasing information density for learning temporal patterns from an image series and using visual redundancy within such imagery for temporal downscaling, we introduce temporally shifted heatmaps as an advantageous approach over Gramian Angular Field (GAF)-based image time series. Further, we propose Multispectral-WaveMix, a derivative of the mixer-based computer vision architecture, as a high-performance model to harness image time series for surface albedo forecasting applications. Multispectral-WaveMix models intra-day variations in surface albedo on a 1 min scale. The framework combines satellite-sensed multispectral surface reflectance imagery at a 30 m scale from Landsat and Sentinel-2A and 2B satellites and granular ground observations from SURFRAD surface radiation monitoring sites as image time series for image-to-image translation between remote-sensed imagery and ground observations. The proposed model, with temporally shifted heatmaps and Multispectral-WaveMix, was benchmarked against predictions from models image-to-image MLP-Mix, MLP-Mix, and Standard MLP. Model predictions were also contrasted against ground observations from the monitoring sites and predictions from the National Solar Radiation Database (NSRDB). The Multispectral-WaveMix outperformed other models with a Cauchy loss of 0.00524, a signal-to-noise ratio (SNR) of 72.569, and a structural similarity index (SSIM) of 0.999, demonstrating the high potential of such modeling approaches for generating granular time series. Additional experiments were also conducted to explore the potential of the trained model as a domain-specific pre-trained alternative for the temporal modeling of unseen locations. As bifacial solar installations gain dominance to fulfill the increasing demand for renewables, our proposed framework provides a hybrid modeling approach to build models with ground observations and satellite imagery for intra-day



Academic Editors: Ismail Gultepe and Gorden Videen

Received: 8 October 2024

Revised: 31 December 2024

Accepted: 2 January 2025

Published: 29 January 2025

Citation: Karalasingham, S.; Deo, R.C.; Raj, N.; Casillas-Perez, D.; Salcedo-Sanz, S. Generating High Spatial and Temporal Surface Albedo with Multispectral-Wavemix and Temporal-Shift Heatmaps. *Remote Sens.* **2025**, *17*, 461. <https://doi.org/10.3390/rs17030461>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

surface albedo monitoring and hence for intra-day energy gain modeling and bifacial deployment planning.

Keywords: surface albedo downscaling; temporal downscaling; image time series; temporally-shifted heatmaps; image-to-image translation; bifacial solar photovoltaic system

1. Introduction

Declared as a terrestrial essential climate variable [1], surface albedo plays a vital role in shaping our surface radiation budget and hence the surface energy balance. Sensitive to changes in land cover characteristics [2], increases in the albedo effect due to land surface transformation [3], and the rise of urban-scale climate phenomena such as the Urban Heat Islands (UHI) [4,5] underline its importance in our climate system. However, as a determinant of the portion of incoming solar irradiance that is reflected [2] (Figure 1d), surface albedo could also be harnessed by a new generation of solar technologies, the bifacial solar, boosting its energy density. Surface albedo is the most important factor influencing the energy gain of bifacial performance [6,7]. Especially in northern locations dominated by cooler climatic conditions and cloudy skies where diffused irradiance (Figure 1b) might be as high as 60 percent [8], the reflected irradiance is vital. Similarly, it has been noted that bifacial output is relatively higher during early mornings and late afternoons due to diffused irradiance [9], reinforcing the potential role of surface albedo.

As solar becomes the dominant renewable, surpassing wind [10], the drive to increase the energy density of solar installations is becoming an economic and environmental imperative. The high land usage of traditional solar installations in comparison to other renewables and non-renewables [11–13], the associated competition for land with agricultural, livestock farming, and ecological interests [14–16], the need to extend transmission infrastructure to new solar production regions [17], and societal concerns [18,19] are key drivers towards increasing solar energy density. Since bifacial panels are capable of generating electricity from both sides of the panel, harnessing both direct (Figure 1a) and reflected irradiance (Figure 1c), the technology has demonstrated the potential to increase energy production by up to 17 percent [20,21] while lowering the costs of energy production per land unit area [21–23]. The maximization of their energy generation potential is highly dependent on their ability to generate electricity from solar radiation incident on the back of the panel. In particular, harnessing the ground-reflected irradiation (GRI) (Figure 1c), which represents the surface-reflected component of the incoming solar irradiance, remains crucial to the back of the panel electricity generation. However, GRI is dependent on the level of reflectivity of the surface, i.e., surface albedo, whose modeling of intra-day variability at high spatial scales remains challenging.

However, the effective modeling of such variability is difficult due to the constant albedo assumption [24]. Further, obtaining accurate ground observations of surface albedo is challenging for the following reasons. Ground observations, considered the gold standard, are sparse, require specialized equipment such as albedometers or pyranometers, and necessary expertise, and are costly. Though field testing of bifacial installations has been carried out in various regions of the world, including the USA, Germany, France, and Chile, with bifacial test beds, studies note their limited utility due to testing scenarios, season, landscape, and temporal range [25,26]. Further studies on obtaining surface albedo observation noted that longer-duration (1 year or more) observations from albedometers are required for effective bifacial yield modeling [27]. While satellite observations from platforms such as MODIS, Sentinel-2A/B, and LandSat provide a wider coverage in the

form of surface reflectance imagery, depending on the sensing platform, it too suffers from inadequate temporal coverage (5–16 days) due to extended revisit times. Their inability to provide continuous observations throughout daylight hours, the spatial scale (10 m–2 km) of sensing, cloud coverage, aerosols, and other atmospheric factors limit their utility [28]. Specifically, the higher pixel size of coarse observation from satellite platforms such as MODIS limits comparison with spatially smaller ground sites [27]. Therefore, the modeling of daily and diurnal surface albedo with satellite-sensed data has been primarily focused on the use of parametric modeling approaches. Such approaches include using the weighted mean of instantaneous albedo retrieved from MODIS Satellite albedo product to derive daily means of surface albedo [29]; regression model trained on simulated radiative transfer outputs for the estimation of daily land surface albedo from instantaneous albedo provided by Visible Infrared Imaging Radiometer Suite (VIIRS) aboard NOAA-20/21 Satellites over snow-free land surfaces [30]; combining instantaneous albedo retrievals with simulated radiative transfer outputs in a Look-up Table (LUT) to generate daily means with lower errors than local noon surface albedo from MODIS [29]; and obtaining the product of land surface and topographical factors on diurnal surface albedo with instantaneous albedo, which does not account for topographical effects, to improve diurnal albedo predictions over complex terrains [31]. Though regression-based estimation approaches of daily mean have reduced bias from the use of single local noon albedo derived from 16-day satellite composites, geostationary satellites such as EUMETSAT and the GOES constellation have the distinct advantage of being able to provide continuous measurements due to their geostationary nature. Utilizing the Advanced Baseline Imager (ABI) on board the GOES-R [32], we developed a surface albedo estimation approach that generated observations acceptable for climate and weather requirements. However, the medium-scale spatial resolution (0.5 km, 1 km, and 2 km) did not accede to a fair comparative analysis with finer-scale ground observations [32]. Further, the positioning of the satellite platform limits the region of observation, restricting it to a circular zone of radius 70° (approx.) directly below the satellite [33]. In addition, a study on EUMETSAT and the multi-agency effort towards creating a global land surface albedo record with geostationary satellites noted the challenges in cloud detection and aerosol effects required a minimum of cloud-free, daylight observations for albedo retrieval, which was a particular challenge over high vegetation regions [34,35]. While physically based parametric models have emerged, they too rely on coarse and discontinuous satellite observations and are limited by inadequate parameterization, computational intensity, and uncertainties arising from data inputs [36]. More recently, the emergence of harmonized observations from multiple finer spatial resolution satellite sensors aboard LandSAT and Sentinel constellations paves the way for novel data-driven approaches for modeling intra-day variations in surface albedo, especially for bifacial solar applications.

Increasingly, the installation flexibility of bifacial solar has also given rise to vertical, tilted, and integrated installations, opening the way for models such as AgriSolar [37] and mixed land use. Such installation scenarios require a wider spatial and temporal coverage of surface albedo observations for the effective exploitation of solar energy potential. More importantly, the inevitable shift to bifacial technologies increases the climate dependence of solar technologies, as more climatic factors now broadly shape the power output of solar installations. Similarly, recent studies have noted the expansion of solar farms has the effect of reshaping the surface albedo profiles of barren lands [38]. And more broadly, surface albedo influences surface energy budget and is a key input in climate modeling of radiative force and its effects on climate [39,40]. Therefore, for solar power output planning, effective grid planning, assurance of grid stability, efficient energy market operations, and, more

broadly, climate modeling, the granular temporal scale modeling of intra-day variations in surface albedo is a strategic necessity.

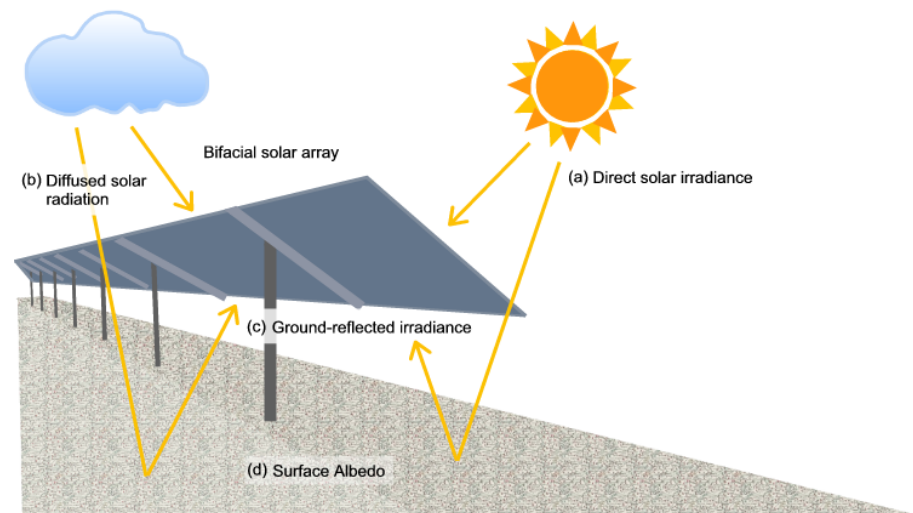


Figure 1. Solar irradiance received by a bifacial solar array.

Considering this necessity and the potential of pre-trained deep learning models to fill gaps in ground observations, this research studies image-to-image translation from remotely sensed, low spatio-temporal, multispectral imagery to high-temporal ground observations. The research envisages a future where fractional ground observation could be utilized to fine-tune pre-trained deep learning models to obtain more localized predictions. Short-term or fractional monitoring of climate and environmental factors such as pollution, greenhouse gas emissions from energy projects, and global horizontal irradiance for solar projects demonstrates the practical use of fractional monitoring. Utilizing such high temporal resolution short-term observation of surface albedo with remote-sensed surface reflectance has the potential to help deep-learning models generate observations of high spatial and temporal relevancy. Such an approach would be especially advantageous in scenarios where influencing variables such as surface type and elevation are stable over the longer term.

Motivated by the physical relationships that exist between the spectral composition of incident radiation, land cover, solar zenith angle, and surface albedo, we consider pairing satellite-sensed surface reflectance imagery with high temporal ground observations from surface radiation monitoring stations for image-to-image translation for temporal downscaling. In the context of generating high temporal intra-day surface albedo predictions, this study aims to develop a novel temporal downscaling method to generate surface albedo of 1 min temporality during daylight hours based on a satellite sensor to ground station observations deep learning. Therefore, the novelty and contribution of this paper concerning the temporal downscaling are as follows:

1. To develop a method for encoding temporal patterns into heatmaps in the form of temporal-shift heatmaps to enable the application of computer vision approaches to temporal downscaling.
2. To develop a new image-to-image translation approach, Multispectral-Wave Mix, for temporal downscaling of satellite-sensed surface reflectance across multiple spectral bands to generate surface albedo of high temporality. Motivated by the convolutions-free alternative modeling approaches inspired by vision transformers being applied in image classification, the aim is to develop an architecture that learns between remote-sensed spectral imagery transformed into frequency distributions and ground-

observed modalities transformed from time series into image series for image-to-image deep learning.

3. To learn temporal trends with image patches from frequency distributions. Since climate variables tend to exhibit trends within time blocks (hourly) and across time blocks, image patches of tokens enable learning within (intra) and across (inter) pixels for trends.
4. To apply the developed model, Multispectral-Wave Mix, for the temporal downscaling of satellite-sensed harmonized surface reflectance imagery at a 30 m scale across multiple spectral bands from Landsat and Sentinel-2A and 2B as inputs and high-temporal surface albedo observations from SURFRAD and National Renewable Energy Laboratories (NREL) surface radiation monitoring as targets.
5. Finally, the proposed Multispectral-Wave Mix is assessed against emerging modeling approaches MLP-Mix [41], image-to-image Mix [42], and Standard MLP across performance metrics including Cauchy loss function (CLF) [43], Structural Similarity Index (SSIM) [44], and signal-to-noise ratio (SNR) [45], and against observations from surface radiation monitoring stations and physical predictions from the National Solar Radiation Database (NSRDB).

The objectives of this study are considered in the context of study areas in the states of Illinois, Montana, Colorado, Mississippi, Pennsylvania, and Nevada, United States, with the presence of ground-based high temporal surface albedo monitoring stations.

2. Background and Related Work

Downscaling of low-resolution imagery to its high-resolution equivalent with deep learning approaches has been the subject of several studies. Pioneered by [46], the single-image super-resolution convolutional neural network (SRCNN) learns a mapping between low- and high-resolution image pairs in a data-driven approach with convolutional neural networks (CNNs). The simple modeling approach was further enhanced with additional convolutional layers in Very Deep Super-Resolution (VDSR) for contextual learning by mapping residuals between high- and low-resolution image pairs, thereby enhancing accuracy [47]. Further efforts addressed the computational intensity through the application of task-specific convolutions and a deconvolution layer for upscaling [48]. The Fast Super-Resolution Convolutional Neural Network (FSRCNN) by [48] also replaced the Rectified Linear Unit (ReLU) function in the antecedent models with the Parametric Rectified Linear Unit (PReLU). Similarly, the efficient sub-pixel convolutional neural network (ESPCN), focused on real-time super-resolution, utilized low-resolution images for feature learning and introduced sub-pixel convolution for combining and upscaling the low-resolution feature maps into a high-resolution output [48]. Subsequent model evolution considered the relative importance of features with attention layers. Residual Channel Attention Network (RCAN), for example, applied residual connections and channel attention [49], while Densely Residual Laplacian Attention Network (DRLN) applied Laplacian attention [50], increasing downscaling performance at the expense of computational intensity.

The emergence of Generative Adversarial Networks (GANs), which utilize a generator and a discriminator in a game-theoretic approach to iteratively generate realistic images, was applied to the multi-scale super-resolution challenge. Super-resolving low-resolution images to higher multiples (x4), the Super Resolution Generative Adversarial Networks (SRGANs) applied the GAN architecture with adversarial loss to generate perceptually similar images [51]. The approach was further enhanced in Single Image Super Resolution with Feature Discrimination (SRFeat), which employed a second feature discriminator to encourage the generation of structurally similar higher-resolution images [52]. Iterating upon SRGAN, the Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs)

introduced dense blocks, residual connections, and a relativistic discriminator [53]. The latter enabled the probabilistic discrimination of real vs. fake images [54]. The GAN-based approaches have generally attained new levels of perceptual similarity compared with the previous models.

However, Diffusion-based modeling approaches that introduce noise in steps to a base image in a forward process and learn the denoising in the backward process have set new benchmarks in image super-resolution. Accordingly, the Denoising Diffusion Probabilistic Models (DDPMs) based Super-Resolution via Repeated Refinement (SR3) model introduces a low-resolution image with Gaussian noise for progressive denoising by a U-Net network [55]. Subsequent models have attempted to reduce the computational intensity of the Diffusion process with Latent Diffusion [56] and by the learning of residuals between low- and high-resolution pairs [57].

Several works have extended the standard downscaling approaches to the spatio-temporal downscaling of climate imagery. Proposing a stacked SRCNN (Stacked SRCNN) [58] downscaled the gridded precipitation data for the United States generated by the Parameter elevation Regression on Independent Slopes Model (PRISM) from 100 km to 12.5 km. Subsequent works developed the Laplacian pyramid super-resolution network (LapSRN) [59] for the downscaling of precipitation data, adversarial super-resolution [60] for downscaling Global Climate Model (GCM) outputs on wind and solar, and the Super Resolution Deep Residual Network (SRDRN) [61] for downscaling of daily temperature and precipitation.

In contrast to the downscaling of climate model-generated data, direct downscaling of satellite imagery is fraught with challenges due to weather and cloud effects on the imagery, the spatial resolution of the imagery, and the impossibility of obtaining matching pairs of low- and high-resolution imagery of sufficient spatial and temporal coverage. To overcome some of these challenges, auxiliary high-resolution imagery from all-weather sensing Synthetic Aperture Radar (SAR) sensors, such as Sentinel-1, was paired with target variable imagery, land surface temperature imagery from Landsat-8 [62], and MODIS-sensed Normalized Difference Vegetation Index (NDVI) [63]. Yet others employed cloud-free imagery for the downscaling of MODIS-sensed spectral imagery [64] or harnessed higher-resolution aggregated sensor inputs from the panchromatic band for pan-sharpening [65,66].

Despite these advances in spatial downscaling, the lack of temporal coverage due to satellite revisit times gives rise to time gaps in satellite coverage. Existing studies have applied deep learning approaches to gridded datasets for temporal downscaling of ERA5 reanalysis data in [67] and atmospheric fields [68]. Ref. [69] utilized radar imagery of rainfall for the temporal downscaling of rainfall to finer temporal resolution. However, achieving the degree of temporal coverage with satellite observations remains an open challenge, especially for surface climate indicators, which are vital to large-scale physical climate models, climate-driven energy production, agriculture, and other climate-dependent sectors. Recent efforts have focused on the combining of multiple satellites into constellations with Sentinel-2A and 2B for more regular ground observations between 5 days apart. Similarly, the harmonization of observations of similar satellite sensors distributed across earth-sensing platforms has been undertaken. The Harmonised Surface Reflectance LandSat-Sentinel Project has combined Landsat-8 and Sentinel-2A and 2B imagery to build observations 2–3 days apart. Yet observations from single visits 2–3 days apart are inadequate for applications such as bifacial solar energy generation planning and operations, which require more granular intra-day observations for site selection, site curation, and energy gain modeling. At present, datasets such as the National Solar Radiation Database (NSRDB) provide global solar radiation projections, including daily surface albedo predictions at a spatial resolution of 2 km for the Continental United

States and 4 km spatial resolution for other regions [36]. Generated by the Physical Solar Model (PSM), the model utilizes spatially scaled-up surface albedo observations of 8-day temporality from MODIS (500 m–1 km) to accommodate the larger grid sizing of NSRDB (2–4 km) [70]. However, the daily albedo product is unable to capture the intra-day variations during sunlight hours.

Though, as noted above, deep learning-based downscaling approaches have proved skillful in downscaling certain climate variables, the lack of availability of high-resolution spatio-temporal satellite imagery to be paired with low-resolution imagery hinders current modeling approaches. While standard deep learning-based downscaling approaches employ synthetically generated low-resolution images through algorithmic approaches such as bicubic interpolation to create paired high-resolution and low-resolution imagery, the underlying distributions of such imagery may not reflect a low-resolution ground truth. Furthermore, the generalizability of such models is limited. Similarly, models designed for natural imagery, which contains inherent characteristics such as edges and contrasting features, underperform on climate imagery captured across spectral bands [71]. Furthermore, the limitations of assessment metrics, which are designed to enhance perceptual similarity to account for the human visual system [44], may not faithfully reproduce climate imagery representing physical factors. Significantly, while current modeling approaches excel in generating perceptually pleasing imagery, they continue to suffer from hallucinations [72]. However, such hallucinations could distort physical climate imagery. Lastly, transfer learning, or the use of foundational datasets, has benefited natural image downscaling. The unique characteristics of multispectral imagery from remote sensing platforms may not be conducive to such approaches. Overall, the spatio-temporal downscaling of physical climate imagery, such as surface albedo, requires specialized approaches, especially in the context of temporal downscaling.

Therefore, in this study, we consider the challenge of high spatio-temporal downscaling as an image-to-image transfer challenge between the Harmonised Surface Reflectance satellite product and high-resolution ground observations from SURFRAD towers across seven locations. Considered to be the gold standard in climate monitoring, we transform ground observations into image time series. The high-resolution image time series is paired with the low-resolution image time series based on the observations from the Harmonised Surface Reflectance satellite product. The transformation of time series into images with approaches such as Gramian Angular Fields (GAF) has been applied in pattern recognition [73,74], classification [75–78], and time-ahead prediction [79]. However, unlike classification and recognition tasks, predictions on a granular time scale using image time series require the successful reconstruction of pixels, especially from the edges of images. As such, we propose a means of increasing information density via encoding time series into temporally-shifted heatmaps and harnessing information redundancy.

In the context of image-to-image translation, GANs have proven valuable. However, transformers, including image transformers, have excelled in modeling sequential data while capturing long-term and cross-variables dependencies with attention mechanisms and parallel processing. Drawing on the transformer architecture, a new class of models, such as the Mixers models, have shown high performance when it comes to sequential data. Such models were first proposed [80] in 2023, the TSMixer, which outperformed the Temporal Fusion Transformer [81], FEDFormer [82], Autoformer [83], Informer [84], and univariate Linear and PatchTST [85] models across long-term forecasting datasets from multiple domains. Hence this research focuses on the new architecture with attentionless learning from sequence data using token mixing and less computationally intensive MLPs as a basis for learning from sequence data. We undertake a model design to explore satellite imagery across frequency components to facilitate greater feature representations.

Similarly, the addition of patch-embedding in combination with frequency components enhances learning across multi-scales, frequency, channel, and patch to capture the distant characteristics of remote sensing imagery in contrast to natural imagery.

The sections that follow present the methodological approaches that underpin the proposed model, followed by the results and discussion, conclusions, and further avenues for research.

3. Materials and Methods

3.1. Study Area

For this study, seven sites (Table 1) from the states of Illinois, Colorado, Mississippi, Montana, Nevada, and Pennsylvania, United States, were chosen. The study sites were Bondville, Illinois; Golden, Colorado; Goodwin Creek, Mississippi; Fort Peck, Montana; Desert Rock, Nevada; Penn. State Univ., Pennsylvania; and Table Mountain, Colorado. These sites represent subtropical and temperate zone land classifications (Table 2) and elevations between 1000 and 1850 m at the sites (Table 1).

Table 1. Geographical details of the study area for the proposed Multispectral-WaveMix model training and validation.

Test Site	Latitude	Longitude	Elevation (m)
Bondville, Illinois	40°3'6.9"	88°22'23.12"	230
Desert Rock, Nevada	33°37'33.6"	116°01'04.8"	1007
Fort Peck, Montana	48°18'28.18"	105°6'6.12"	634
Golden, Colorado	39°44'31.2"	105°10'48.0"	1829
Goodwin Creek, Mississippi	40°15'16.92"	89°52'22.44"	98
Penn. State Univ., Pennsylvania	40°43'12.43"	77°55'51.06"	376
Table Mountain, Colorado	40°07'30.0"	105°14'13.2"	1689

Table 2. Land cover details of the study area for the proposed Multispectral-WaveMix model training and validation.

Test Site	Land Classification
Bondville, Illinois	Grassland
Desert Rock, Nevada	Subtropics
Fort Peck, Montana	Grassland
Golden, Colorado	Subtropics
Goodwin Creek, Mississippi	Grassland
Penn. State Univ., Pennsylvania	Cropland
Table Mountain, Colorado	Temperate Zone

The study utilized 1 min averages of up-welling and down-welling global solar radiations recorded at sites by the SURFRAD solar radiation monitoring network and Solar Radiation Research Laboratory of the NREL datasets for the generation of 1 min scale surface albedo ground observations. Similarly, surface reflectance imagery from the Harmonized Landsat and Sentinel-2 (HLS) project from NASA was obtained. Harmonizing surface reflectance observations from the Operational Land Imager (OLI) onboard Landsat-8/9 and the Multispectral Instrument (MSI) from Sentinel-2A and 2B, the HLS project creates a higher temporal coverage of 2–3 days at resolutions of 30 m. It processes Landsat-8/9 observations at 30 m over 16-day revisit times with Sentinel-2A/B at similar spatial resolution over 5-day revisit times to generate a seamless product. Multispectral data (blue, green, red, and near-infrared) across bands 1, 2, 3, 4, and 5 (Table 3) were obtained, each with a spatial resolution of 30 m.

For each site, shapefiles were generated and were used to extract site-specific HLS surface reflectance and land cover imagery from the satellite image datasets for the

study period. Similarly, a 1 min surface albedo time series was generated from up-welling and down-welling global solar (W/m^2) observations for local station time between 07:00 and 17:00 hours covering daylight hours as per the following formula (MCD43A3 Albedo Product):

As described in an earlier section, the univariate 1 min surface albedo time series was transformed into an image time series, with temporal shift and Gramian Angular Field, encoding hourly temporal dependencies into an image. Following the transformation, temporally overlapping surface reflectance and surface albedo imagery were identified to be utilized in this study. A total of 20,790 matching sets of imagery across seven sites were identified, providing an average of 118 days of coverage per year for the daylight hours.

Table 3. Harmonized Landsat Sentinel-2 (HLS) spectral bands utilized for model development.

Band No	Wavelength (Micrometers)	Band	Resolution	Frequency
B01	0.43–0.45	Ultra Blue/Aerosol	30 m	2–3 Days
B02	0.45–0.51	Blue	30 m	2–3 Days
B03	0.53–0.59	Green	30 m	2–3 Days
B04	0.64–0.67	Red	30 m	2–3 Days
B08	0.78–0.88	Near-Infrared	30 m	2–3 Days

3.2. Model Design

The objective model, Multispectral-Wave Mix, and benchmark models were developed on an Nvidia RTX A6000 (48Gb) high-performance workstation under a PyTorch machine learning environment. PyTorch provides the necessary computer vision libraries for the experimental development, training, and evaluation of novel deep learning models (<https://pytorch.org/> (Last Accessed: 23 January 2025)). Further, the model design environment consisted of Kornia, a geometric computer vision and spatial AI package; PyWavelets, a wavelet transform framework; QGIS, a geospatial information system; and GDAL, a geospatial data abstraction package.

The data processing pipeline of the model development consisted of the extraction of surface reflectance imagery for each study site from the HLS surface reflectance product (S30), the extraction of solar zenith angle across time from NSDRB, the transformation of surface reflectance with the cosine of solar zenith angles, and the transformation of 1 min surface albedo time series into image time series with GAF transformations. An input, a cosine of solar zenith angle-transformed surface reflectance image series and matching target, a temporal-shift heatmap, and a GAF time series of surface albedo image series, was assembled for the daylight duration between 7:00 and 17:00 hours for the years 2019–2022, across all months for the three study locations.

The data preparation was followed by the splitting of input and target image series into training and validation sets. A 78:20:2 split was utilized in model training, evaluation, and testing.

Objective and benchmark models were subsequently developed. Model-specific hyperparameters for benchmark models were retained as originally proposed. All models were trained to epoch 50, and the best-performing epochs were used for inferencing. Further, the following common settings were used across the models for training and evaluation:

1. **Performance Metrics:** All models were assessed with a common set of metrics: Cauchy loss function (CLF), Structural Similarity Index (SSIM), and signal-to-noise ratio (SNR).
2. **Optimizer:** The Adaptive Moment Estimation (*Adam*) optimizer [86] was used with an initial learning rate of 0.001 and a learning schedule with a decayed step of 15 and a decay gamma of 0.1.

3. **Patch Sizes:** A patch size of 30 was applied across the patch embedding layer for an image size of 60 pixels across objective and benchmark models.

3.3. Performance Criteria

The performance of the Multispectral-WaveMix and comparative models was assessed with model performance evaluation metrics, namely the Cauchy loss function (CLF), Structural Similarity Index (SSIM), and peak signal-to-noise ratio (PSNR). Further model predictions were assessed against ground observations from the SURFRAD surface radiation monitoring sites within the study area, the ground observations from the bifacial PV test bed of the National Renewable Energy Laboratory, and predictions from the National Solar Radiation Database (NSRDB).

Generally, deep learning model performance is assessed with Mean Squared Error (MSE). However, the assumption of noise distribution as Gaussian under MSE may not suit if outliers are present in a distribution [43]. Since outlier events are more likely due to the nature of the climate, a Cauchy distribution, which has demonstrated better performance compared with MSE [43], was adopted. Mathematically a CLF is [43]:

$$L_{CLF} = \frac{c^2}{2} \log \left(1 + \left(\frac{y - \hat{y}}{c} \right)^2 \right) \quad (1)$$

In addition, SSIM was adopted as an additional metric to evaluate the quality of the image-to-image performance. In contrast to MSE and its derived function PSNR, SSIM was noted to better compare the structural similarities between the source and target images [44]. Mathematically, the Structural Similarity Index Measure (SSIM) is defined as [87]:

$$SSIM = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

Related to the MSE, the PSNR is a quality measure indicative of the quality between the predicted and observed imagery. Peak signal-to-noise ratio (PSNR) is mathematically formulated as follows:

$$PSNR = \frac{10 \log_{10} \left(\max(\max(x), \max(y))^2 \right)}{|x - y|^2} \quad (3)$$

This section presents a theoretical overview of the methods behind the objective model, The Multispectral Wave Mix, as featured in Figure A1. The modeling framework consists of a method to transform time series into imaging time series to increase information density and harness visual redundancy, wavelet transformation for frequency decomposition, tokenization with image patch embedding, channel and token mixing, and component aggregation. Additional details of the peer models Image to Image Mix and MLP-Mix are extensively addressed by [42] and [41], respectively.

3.4. Information Density and Visual Redundancy

The challenges, such as temporal downscaling utilizing matching satellite imagery and ground observations, face significant data constraints. These constraints arise due to satellite revisit times, atmospheric effects such as cloud cover, equipment errors, and limited spatial distribution of ground monitoring stations. Further, the cost of deployment and maintenance of high temporal resolution ground monitoring, such as albedometers and pyranometers, constrains data availability. Therefore, the transformation of limited datasets into imagery of higher information density, a measure of the information content encoded in an image, is vital.

Unlike natural languages, images suffer from lower information density [88]. The presence of noise, distortions, and artifacts contributes to lowering the informational content in sensor imagery, especially with RADAR and LIDAR. While lowering the information density serves the purpose of obfuscation of sensitive information [89], it poses challenges in image detection, classification, reconstruction, and other related tasks. Though compression algorithms have proved effective in enriching the information density of natural images [88], the preserving of temporal positioning is essential for image-based temporal downscaling.

In the context of images, the information density is measured in terms of mean information density (\bar{K}). For an image with an area, A , and information content, I , the mean information density [90]:

$$\bar{K} = \frac{I_{Total}}{A_{Total}} \quad (4)$$

and assuming uniform quantity of information per pixel, I_{Total} is defined as [90]:

$$I_{Total} = N \times H(X) \quad (5)$$

for an image with N pixels and random variable x with a stochastic distribution p has a Shannon entropy $H(x)$ representing the pixel intensity.

From a modeling perspective, transformers have emerged as a leading architecture in natural language processing with an equivalent architecture, Vision Transformers, for computer vision applications. Akin to transformers, MLP-Mix-like architecture has also emerged. However, such architectures require larger datasets for achieving higher model performance at a lower computational cost [41]. Therefore, augmenting information density is introduced using temporal shift heatmaps, to be utilized as an imaging time series.

Further in traditional deep learning, redundant information that is repetitive is considered as not useful. However, the use of repetitive patterns or self-similarity inherent within an image has been exploited successfully for image enhancement, image reconstruction and decomposition. Employing deep internal learning similarities within a single image was exploited with deep internal learning by [91] for the super-resolution of the image without the need for a large dataset of imagery. Similarly, the visual redundancy inherent within images also formed the basis of image decomposition in [92] and temporal super-resolution of videos recorded with lower frame rates [93]. To exploit the benefits of visual redundancy, the proposed image series, temporal shift maps, incorporates redundant temporal information via temporal shifts of the primary time series to augment learning.

3.5. Imaging Time Series

Unlike traditional time series analysis, computer vision approaches such as image-to-image translation require both inputs and targets in the imaging modality for the identification of visual features for representation. Therefore, first, we investigate a multi-step approach proposed by [94] for the transformation of 1D time series data into image time series (Figure 2), which has been subsequently applied in other domains [95–97]. In such a transformation process, Figure 3, the time series is encoded into an image series with Gramian Angular Field (GAF) under a polar coordinate system [94]. Unlike a Cartesian coordinate system, this approach enables the preservation of temporal relationships between time points while representing each time point uniquely in a 2D plane [94]. Further, it enables the analysis of correlations across time points using Gramian Summation Angular Field (GASF) and Gramian Difference Angular Field (GADF) approaches [94].

Lending itself to extracting visual features representing temporal relations with computer vision approaches, a time series is first scaled to be between $[-1,1]$ as follows [94]:

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (6)$$

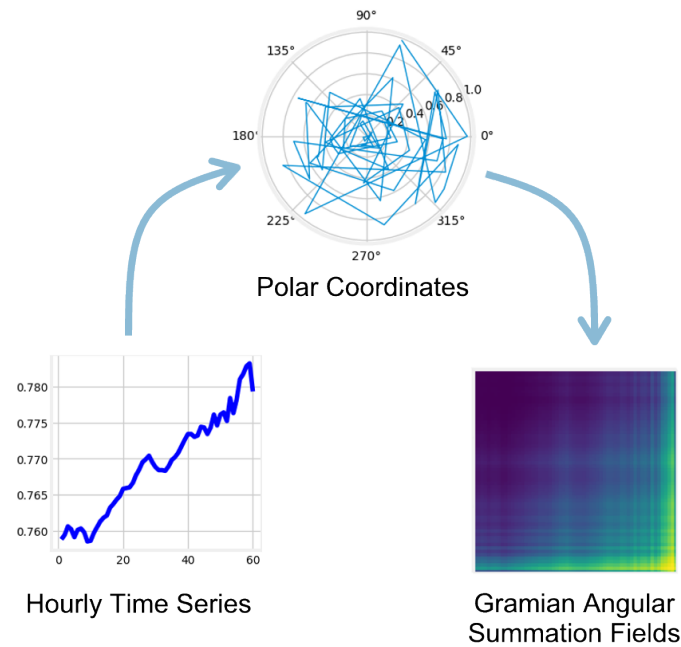


Figure 2. Transformation of 1D time series into an imaging time series with Gramian Angular Function.

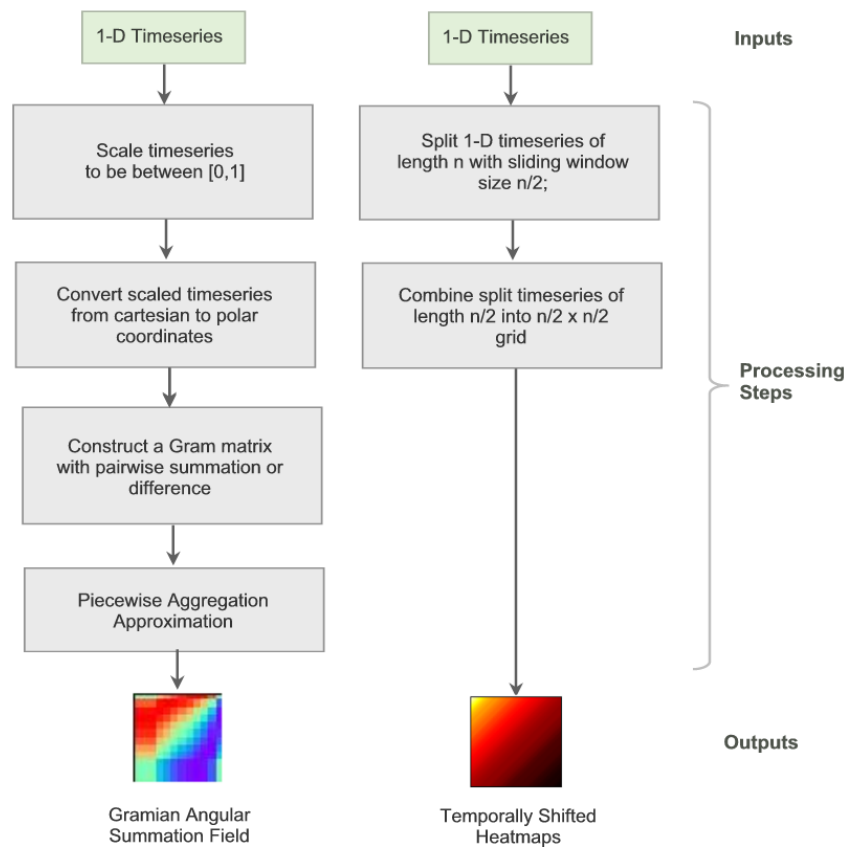


Figure 3. Flowchart of Gramian matrix and temporal shift heatmaps construction.

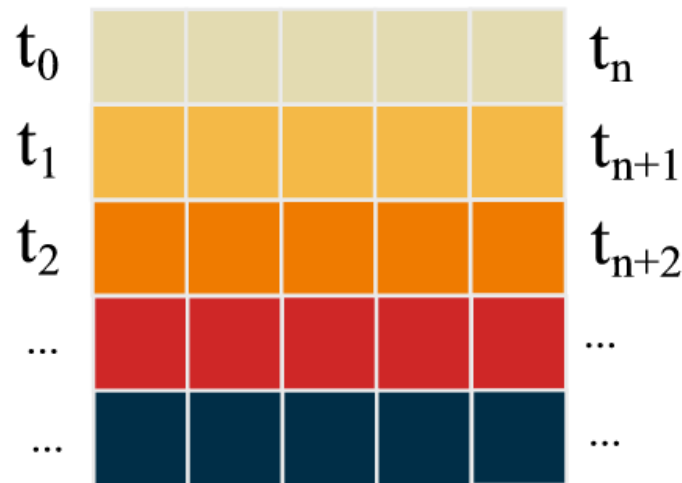


Figure 4. Transformation of 1D time series into an imaging time series with temporal shift heatmaps.

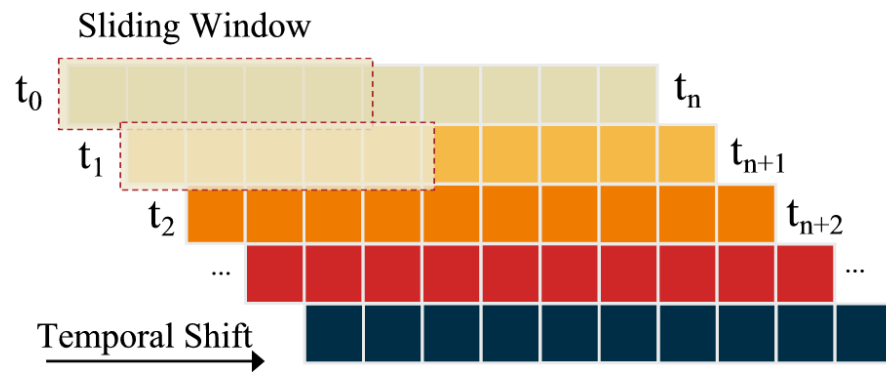


Figure 5. Transformed temporal shift heatmaps.

3.6. Wavelet Transformation

Emerging as a novel approach in signal processing, wavelet transformation is the representation of a mathematical function in terms of a set of coefficients, the wavelet coefficients. Known as wavelets, the mathematical functions are used to analyze an image across spatial and frequency components. The decomposition of an image, with wavelet transformation, results in sub-bands consisting of an approximation and detail components of an image, which can be used to reconstruct the original image via inverse wavelet transform. Practically, the wavelet transform is utilized for removing high-frequency noise and analyzing images.

In this research, we apply a 2D Discrete Wavelet Transform using Haar wavelets for the decomposition of the temporal shift heatmap-based image by each spectral band (e.g., Band 1) into four sub-bands (LL, LH, HL, and HH), one containing the approximation and three with detail coefficients (Figure 6). The Haar wavelet, mathematical functions (f), are represented as follows [98]:

$$f \rightarrow (a^L | d^L) \quad (10)$$

$$a^L = (a_1, a_2, \dots, a_{(N/2)}) \quad (11)$$

$$d^L = (d_1, d_2, \dots, a_{(N/2)}) \quad (12)$$

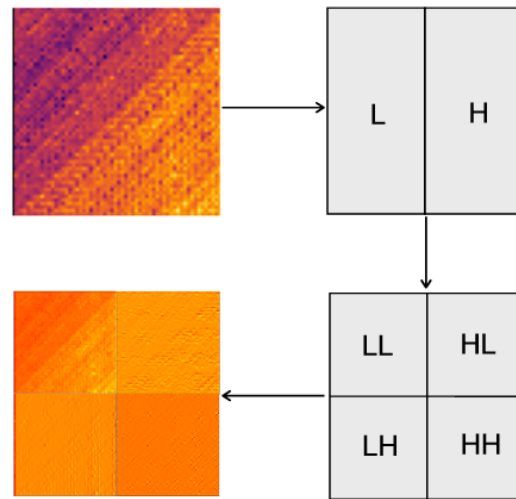


Figure 6. After Level 1 decomposition with Haar wavelet.

For decomposition level (L), the approximation (a) and details (d) are defined as follows [98]:

$$a_m = \frac{f_{(2m)} - f_{(2m-1)}}{\sqrt{2}} \quad \text{for } m = 1, 2, \dots, N/2 \quad (13)$$

$$d_m = \frac{f_{(2m)} - f_{(2m-1)}}{\sqrt{2}} \quad \text{for } m = 1, 2, \dots, N/2 \quad (14)$$

The Haar wavelet operations are applied in stages, with level one resulting in 4 decomposed sub-band images (LL, HL, LH, HH) as featured in (Figure 6). The approximation (LL) represents the original image with its low-frequency details, while the details feature high-frequency signals horizontally (HL), vertically (LH), and diagonally (HH). The sub-bands (LL, LH, HL, and HH) representing the spatial and frequency components of each spectral band would be collated across bands (1, 2, 3, 4, and 8) to form a cross-spectral band stack for feature learning across spectral bands in subsequent modeling stages (Figure 7).

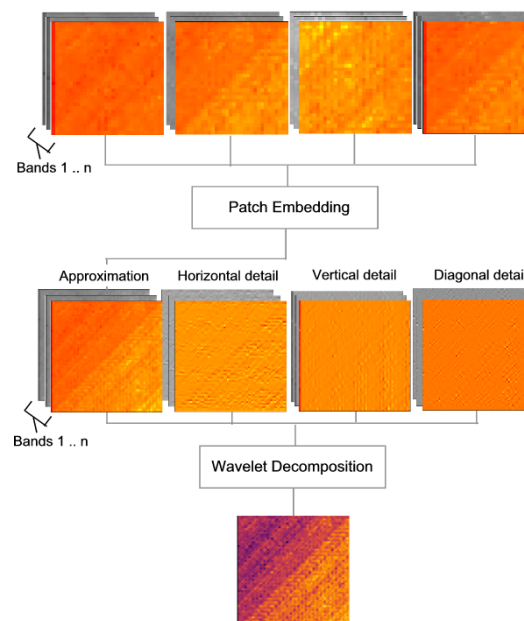


Figure 7. Patch embedding following wavelet decomposition of the input image into frequency subsets.

3.7. Patch Embedding

Equivalent to tokens used in the transformer architectures in Natural Language Processing (NLP), the splitting of images into a sequence of “patches” as input embedding was successfully applied in vision transformers (ViT) while largely retaining the antecedent architecture [97]. Enlarging the receptive field through down-sampling, enabled the learning of spatial relations across patches, thereby encoding global representations [97,99]. Subsequent convolution-free approaches such as MLP-Mix [41] and image-to-image MLP-mixer [42] adopted the approach of improving model scalability over Convolutional Neural Networks (CNNs).

Unlike the other models, in Multispectral-Wave Mix, patch embedding is applied on frequency distributions as a second layer for feature learning within and between time blocks in subsequent layers through token and channel mixing. Each frequency distribution image of size $n \times n$ is divided into $(n \times n) / (p \times p)$, where p is patch size. The generated patches were then subjected to channel and token mixing.

3.8. Channel and Token Mixing

The standard transformer architecture in ViT consists of attention heads for weighing feature importance within patches and images as a whole for global representation [100]. However, both convolutions and attention-free architectures based on multi-layer perceptrons (MLPs) have proven competitive to ViTs and CNNs [41,42]. MLP-Mix introduced a mixer layer that converts image patches into vectors. The stack of vectors from the image patches is then subjected to patch-wise (token mixing), where the vector stack is transformed by channels and fed to an MLP. Token-mixer enables cross-patch feature learning and weight sharing. Channel mixing further facilitates cross-channel feature learning on a patch basis. Both token mixing and channel mixing are carried out through MLP blocks.

The tokens generated by patch embeddings are fed to a token and channel mixing layer in a Multispectral-Wave Mix. However, unlike ViT, MLP-Mixer, and image-to-image MLP Mixers, Multispectral-Wave Mix replaces Gaussian error linear units (GELU) with Sigmoid-weighted Linear Units (SiLU) with MLP Blocks (Figure 8). Formulated as the multiplication of input by the sigmoid function [101] below, the modified MLP block with SiLU has demonstrated superior performance.

$$\text{silu}(x) = x * \sigma(x) \quad (15)$$

where logistic sigmoid = $\sigma(x)$.

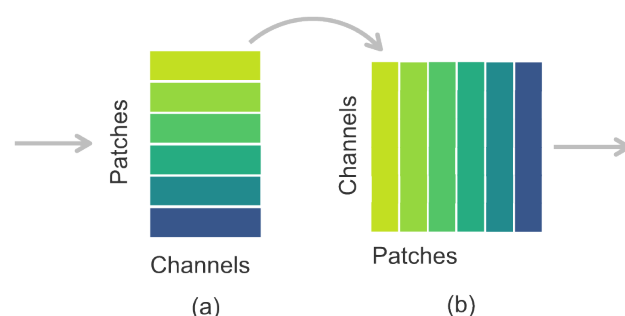


Figure 8. Mixer layers incorporating (a) channel and (b) token (patch) mixing.

3.9. Wavelet Component Aggregation, Fusion, and Exponential Smoothing

Parametric methods of surface albedo estimation from remotely sensed surface reflectance have employed bandwidth-weighted methods. Refs. [102,103] developed linear equations for spectral bands obtained from remote sensing platforms such as Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER), Landsat-7, and Mod-

erate Resolution Imaging Spectroradiometer (MODIS). Drawing on this prior work, we aggregate the spectral-wise decomposed wavelet components from the token and channel mixing stage of the model into cross-spectral (bands 1, 2, 3, 4, 8) detail and approximation components through a combination of maximum, minimum, mean, and summation approaches. As noted in the section on ablation studies, the aggregation approaches applied to cross-spectrum approximation and detail components enhanced the model accuracy while mitigating outliers during prediction.

Following wavelet reconstruction, the aggregated components are fused via trainable parameters before exponential smoothing. Unlike the traditional parametric approach, fusion through trainable parameters enables the model to learn the parameters during the data-driven training process, as parameters will be updated via backpropagation. Further inspired by [104], we integrate exponential smoothing, a traditional method for adjusting forecasts by applying exponential weight decaying to past observations, to prioritize the effect of more recent data on predictions and trends.

4. Results

Despite the importance of the primary climatic driver, surface albedo, to ground-reflected irradiance and its role in boosting the energy gain of bifacial solar PV, it remains challenging to obtain observations at finer spatial and temporal scales. As a result, at present, surface albedo is considered to be relatively constant throughout the day in models such as NSRDB. This is contrary to observations indicating the variability of surface albedo throughout the day across various land cover types [2,8]. To fill the spatial and temporal gaps in surface albedo observations emerging from satellite revisit times and cloud and weather effects, multi-satellite observations of surface reflectance have been harmonized to produce higher temporal observations. The harmonized Landsat and Sentinel surface reflectance product, for example, combines surface reflectance imagery at a 30 m spatial scale from Landsat 8 and Sentinel-2A and 2B to reduce the time gaps in satellite coverage to 2–3 days. Therefore, the combination of short-term ground observations of sufficient temporal length with remote sensing and pre-trained AI models could provide intra-day predictions of surface albedo for energy gain modeling.

This research explored the potential of deep learning approaches that pair remote-sensed surface reflectance observations with a limited set of SURFRAD observations for generating intra-day surface albedo predictions. We contrasted the performance of such modeling approaches, including the objective model, Multispectral-Wave Mix, against peer models MLP-Mix, image-to-image Mix, and Standard MLP for the temporal downscaling of satellite-sensed surface reflectance imagery across spectral bands denoted as B01, B02, B03, B04, and B08, as shown in Table 3 to generate the 1 min scale surface albedo predictions.

Overall, the objective model, Multispectral-WaveMix, outperformed comparative models, MLP-Mix, image-to-image Mix, and Standard MLP, in generating predictions that are closer to the ground truth with lower noise distortion and higher structural similarity as summarized in Table 4 and detailed in Section 4.1. Utilizing Cauchy loss that is robust to outliers, Multispectral-WaveMix achieved 0.00469 and 0.00524 for training and validation, respectively. Similarly, the objective model attained a higher signal-to-noise ratio (SNR) of 72.95060 and 72.56970 on training and validation, respectively. On faithful reconstruction of image time series assessed with the structural similarity index (SSIM), the Multispectral-WaveMix model scored 0.9990 and 0.9989 on training and validation. Contrasting the performance between models utilizing the proposed TSH against the GASF as image time series for applying computer vision approaches, models trained with TSH demonstrated higher skill than GASF, as discussed in Section 4.2. Further analysis of model performance across study sites with root mean square error found the objective model,

Multispectral-WaveMix, achieved lower predictive errors across Desert Rock, Nevada; Goodwin Creek, Mississippi; Penn. State Univ., Pennsylvania; and Golden, Colorado test sites and second-best performance after the image-to-image Mix model in Table Mountain, Colorado, and Fort Peck, Montana sites (Section 4.3). Assessing predictive skill at 1-min temporal intervals with hourly plots revealed that Multispectral-WaveMix produced lower predictive errors between 07:00 and 09:00, followed by between 14:00 and 16:00 hours. As detailed in Section 4.4, the objective model captured morning and afternoon albedo trends with greater skill than predictions from the NSRDB or other peer models.

Table 4. Harmonized Landsat Sentinel-2 (HLS) spectral bands utilized for model development.

Models	Cauchy Loss	SSIM	PSNR
Temporally Shifted Heatmaps (TSH) (Bands 1, 2, 3, 4, 8)			
Multispectral-WaveMix	0.00469/0.00524	0.99904/0.99887	72.95060/72.56970
Image-to-Image MLP Mix	0.00619/0.00587	0.99849/0.99859	71.52140/71.89770
MLP Mix	0.00595/0.00616	0.99856/0.99851	71.89280/71.62610
MLP	0.00634/0.00622		71.50470/71.52330
Gramian Angular Summation Field (Bands 1, 2, 3, 4, 8)			
Multispectral-WaveMix	0.01786/0.01799	0.9943/0.9942	69.61/69.89
Image-to-Image MLP Mix	0.01939/0.019841	0.9936/0.9934	69.34/69.43
MLP Mix	0.01942/0.01854	0.9936/0.9938	69.11/69.55
Bandwise Comparison (Multispectral-WaveMix with TSH)			
Bands 1, 2, 3	0.00514/0.00598	0.9988/0.9986	72.51/71.47
Bands 1, 2, 3, 4	0.00573/0.00550	0.9986/0.9987	71.99/72.45
Bands 1, 2, 3, 8	0.00533/0.00495	0.9988/0.9988	72.43/72.54

In contrasting the performance improvements, we note that the methods employed in the aggregation of multispectral approximation and detail components of the decomposed wavelets prior to reconstruction contributed to achieving the best training performance. Using a combination of aggregation approaches while mitigating image artifacts in the predicted TSH image time series also ensured the model was more robust to outlier forecasts in contrast to peer models, MLP-Mix and image-to-image Mix (Section 5.1). Furthermore, the transformation of predicted image time series into their equivalent 1D time series of surface albedo predictions is highly reliant on the effective reconstruction of the predicted image time series. However, since image time series are susceptible to artifacts, the incorporation of redundant temporal information via temporal shifts of the primary time series augmented the model learning in contrast to the GASF-based approach.

Despite the greater skill demonstrated by the objective model, Multispectral-WaveMix, across study sites with land classifications of grasslands, subtropics, cropland, and temperate zones, such an approach would be especially appropriate where influencing variables such as land cover remain stable over the long term. This limitation arises due to the dependence of surface albedo on surface characteristics [2]. However, in the context of large-scale bifacial solar applications, homogeneity of land cover could be assumed given their commercial nature. Further, this study is limited to selected study sites within the Continental United States with overlapping ground and satellite observations.

Technically, the primary challenge remains the development of highly generalizable models that are adaptable to the land surface and seasonal dynamics. While the building of datasets incorporating heterogeneous landscapes and greater temporal range for training deep learning models could improve generalizability, the combination of climate outputs from physically parametrized models with ground and satellite models provides promising

avenues for research. With the ability to capture climate processes that act as casual drivers, physically parametrized models analogous to ClimSim [105] highlight the possibilities for evolving existing models into physics-informed hybrid deep learning approaches to develop more generalizable surface albedo models.

4.1. Overall Comparison of Multispectral-WaveMix and Comparative Model Performance

The performance of the objective model, Multispectral-WaveMix, is assessed against benchmark models, MLP-Mix [41], image-to-image MLP-Mix [42], and the Standard MLP model to assess their image-to-image translation skill utilizing the time-series-encoded imagery, the temporal-shift heatmaps. Model training and validation performances were evaluated for image-to-image translation performance based on quantitative metrics, Cauchy loss, SSIM, and PSNR. Overall, the objective model, Multispectral-WaveMix, outperformed all other models across metrics. On Cauchy loss, a metric that is robust to outliers, the Multispectral-WaveMix achieved training and validation loss of 0.00469 and 0.00524, respectively (Figure A2). In contrast, the training vs. validation loss of image-to-image ML Mix, MLP Mix, and Standard MLP were 0.00619/0.00587, 0.00595/0.00616, and 0.00634/0.00622, respectively. Trained across all spectral bands (bands 1, 2, 3, 4, and 8), the results indicate the degree to which the objective model Multispectral-WaveMix can generate predictions that are closer to the ground truth, followed by the image-to-image MLP Mix, the next best-performing model based on the validation set.

Utilizing PSNR as a signal-to-noise metric, the quality of the pixels reconstructed by the models in contrast to noise distortion was measured. The Multispectral-WaveMix achieved a higher signal-to-noise ratio on both training and validation with 72.95060 and 72.56970, respectively (Figure A3), while peer models performed comparatively poorly, with image-to-image MLP Mix achieving PSNR of 71.52140 and 71.89770 on training and validation, respectively. Similarly, in training and validation, the MLP Mix model achieved a signal-to-noise ratio of 71.89280 and 71.62610, followed by the Standard MLP, 71.50470 and 71.52330, respectively.

Complementing the absolute error measures such as Cauchy loss and PSNR, the structural similarity between the reconstructed image and ground truth was assessed with SSIM. Capable of capturing perceptive similarity Multispectral-WaveMix achieved higher structural similarity of 0.9990 and 0.9989 on training and validation (Figure A4). Image-to-image MLP Mix obtained 0.9985 and 0.9986, and MLP Mix achieved 0.9986 and 0.9985 on training and validation, respectively. Due to the one-dimensional nature of inputs, SSIM was not applied to the Standard MLP model.

The objective model, Multispectral-WaveMix, performance was subjected to band-wise comparison as listed in (Table 3). With a varying mix of spectral bands, the effect of surface reflectance inputs across spectral bands and the predictive accuracy of surface albedo were evaluated. The Multispectral-WaveMix model trained on bands 1–3 (Ultra Blue, Blue, and Green) performed poorly across all metrics. On Cauchy loss, a model trained on bands 1–3 achieved training and validation loss of 0.00514 and 0.00598, respectively. PSNR on training and validation was 72.51 and 71.47 and SSIM was 0.9988 and 0.9986, respectively. A similar performance was noted when trained on bands 1–4 (ultra-blue, blue, green, and red) with a Cauchy loss of 0.00573 and 0.00550, PSNR of 71.99 and 72.45, and SSIM of 0.9986 and 0.9987 for training and validation, respectively. The model trained on bands 1, 2, 3, and 8 achieved the second-best performance on training and validation with a Cauchy loss of 0.00533 and 0.00495, a PSNR of 72.43 and 72.54, and an SSIM of 0.9988 and 0.9988. Notably, all band-wise variations of the Multispectral-WaveMix model outperformed the peer models, image-to-image MLP Mix, MLP Mix, and MLP, demonstrating the higher model skill.

4.2. Model Performance with GASF and TSH Image Time Series as Inputs

Analysis of temporal patterns with computer vision approaches is reliant on the encoding and accurate decoding of temporal patterns within image time series. As noted earlier, GAF is one such approach to coding temporal information as images. In this section, we contrast the performance of GASF-based image time series against TSH utilizing objective models, Multispectral-WaveMix, and peer models MLP-Mix and image-to-image MLP-Mix (Table 3). The Multispectral-WaveMix trained on GASF-based image time series achieved training and validation losses of 0.01786 and 0.01799, respectively. And SSIM of 0.9943 and 0.9942 and PSNR of 69.61 and 69.89 for training and validation. Meanwhile, the objective model, Multispectral-WaveMix, performed with TSH, surpassing the GASF-based model across all metrics. Similarly, the peer models, image-to-image MLP Mix and MLP-Mix trained on GASF, achieved training vs. validation loss of 0.01939/0.019841 and 0.01942/0.01854, respectively, underperforming compared with TSH-based models. Both models scored lower on SSIM with 0.9936 and 0.9934 for image-to-image MLP Mix and 0.9936 and 0.9938 for MLP Mix, respectively, indicating lower structural similarity. Similarly, a PSNR of 69.34 and 69.43 was noted for image-to-image MLP Mix training and validation, with MLP Mix obtaining 69.11 and 69.55. The PSNR results indicate poor image reconstruction between training and validation inputs and predicted outputs.

The underperformance of GASF imagery across metrics, and in particular the low peak signal-to-noise ratio, was consistent with the challenge of reconstructing the complex pattern inherent in such imagery. In particular, the lack of information redundancy proved to be challenging for the modeling approaches to learn the unique patterns from the images from a limited dataset. While the faithful reproduction of the complete target imagery may not undermine classification-related challenges based on image time series, regression tasks that are sensitive to loss of information during image reconstruction benefited greatly from the information density and redundancy inherent in TSH image time series.

4.3. Overall Site-Wide Model Performance on Surface Albedo

The overall performance of the objective model, Multispectral-WaveMix, and benchmark models, MLP-Mix [41], image-to-image MLP-Mix [42], and Standard MLP model, were assessed with Root Mean Square Error (RMSE) metrics. RMSE, sensitive to outliers, is utilized to assess its effect on model performance across all observations by site. Overall, the objective model outperformed all other models, achieving lower RMSE errors across the Desert Rock, Nevada; Goodwin Creek, Mississippi; Penn. State Univ., Pennsylvania; and Golden, Colorado test sites. The models achieved the second-best performance in two other sites, Table Mountain, Colorado, and Fort Peck, Montana.

The closest predictions to the ground truth were noted for Desert Rock, Colorado (Figure A5) across most study periods and seasons. More than 87.5% of the objective model, Multispectral-WaveMix, predictions generated RMSE less than 0.001, followed by 84.6%, 81.5%, and 80% for MLP-Mix, Standard MLP, and image-to-image Mix models, respectively. Similarly in Goodwin Creek, Mississippi (Figure A6), the Multispectral-WaveMix produced lower RMSE errors with 82.5% of the errors lower than 0.001 while 15% were within the range of 0.001 and 0.005. In contrast, MLP-Mix, Image-to-Image Mix, and the Standard MLP generated higher errors of 38%, 35%, and 43%, respectively, between 0.001 and 0.0049. The RMSE errors below 0.001 were also noted for the test site Penn. State Univ., Pennsylvania (Figure A7), representing 72.0%, 60.3%, 56.8%, and 53.4% of the objective model, Multispectral-WaveMix, MLP-Mix, Stand MLP, and image-to-image mix models, respectively. In Table Mountain, Colorado (Figure A8) and Fort Peck, Montana (Figure A9), Multispectral-WaveMix achieved the second lowest errors with 69.3% and 75.6% of errors below 0.001. The best-performing models, image-to-image Mix and MLP-Mix, in Table

Mountain, Colorado, and Fort Peck, Montana, respectively, achieved 73.3% and 80% of the errors below 0.001.

An additional test site, Golden, Colorado (Figure A10), which was not part of the model training corpus, was also assessed using RMSE errors using the pre-trained model. Notably, the objective model, Multispectral-WaveMix, outperformed all other models, with 69.6% of the errors falling below 0.001, while image-to-image Mix, MLP-mix, and Standard MLP achieved 60.8%, 56.9%, and 44.3% of the errors, respectively.

Across all sites except for Bondville, Illinois (Figure A11), the Standard MLP with 1D temporal series as inputs underperformed other models utilizing image time series based on TSH as input.

4.4. Comparative Model Performance on Surface Albedo Predictions at 1 min Intervals

The performance of Multispectral-WaveMix and peer models trained on all spectral bands (bands 1–8) was assessed for their predictive skill at 1 min temporal intervals with the help of hourly plots. One-minute surface albedo predictions represented by row-wise pixels from the predicted temporal-shift heatmaps were extracted across hourly time intervals for each study location for the year 2022. Where comparative data were available, the 1 min performance was considered for the period from 7:00 to 17:00 hours.

The objective model Multispectral-WaveMix generated the closest predictions to the ground truth for the Desert Rock, Colorado, across most study periods and seasons. We noted that the objective model predictions had lower errors between the hours of 07:00–09:00 (63% between 0.0030 and 0.0099) when compared with peer models with image-to-image mix, MLP-mix, and MLP with 19%, 31%, and 31%, respectively, and NSRDB at 13%. Similarly, 42% of the Multispectral-WaveMix model predictions between 14:00 and 16:00 were more agreeable to ground truth than image-to-image Mix, MLP-Mix, MLP, and NSRDB, with 21% of predictions within 0.002 and 0.009. Since surface albedo exhibits diurnal trends as the solar zenith angle varies throughout the day [28] and the influence of diurnal-specific atmospheric factors (water vapor/aerosols) or cloud cover, the model skill is in capturing the diurnal trends that would deliver greater insights on trends in reflected solar irradiance. As plots Figures A12–A20 highlight, the objective model has demonstrated greater skill comparatively to its peer models in capturing surface albedo trends during the morning and afternoon hours. This is in contrast to the predictions from NSRDB, image-to-image mixers, and the Standard MLP with 1D time series inputs. Further, MLP-Mixer with TSH inputs also demonstrated the ability to capture trends at certain time periods; however, with a lower predictability.

Also of note is that higher predictive errors across models were observed during the early morning and late afternoon. According to cloud coverage data, the daylight hours were subjected to cloud cover of types Cumulus humilis, Stratus nebulosus, Stratus fractus, and Stratocumulus. The effect of cloud cover across spectral bands varies between cloud types [106], impacting remotely sensed surface reflectance observations. Further cloud cover in return reshapes the thermal profile due to radiative feedback [107], potentially contributing to prediction errors.

The proposed model, Multispectral-WaveMix, and peer model performances were also tested against the Golden, Colorado site, which was not included in the model training dataset, to evaluate the potential of pre-trained models to generalize to unseen locations. As plot (Figure A21) for a day in January 2022 indicates, the model produced predictions closest to the ground truth while capturing the surface albedo trend. However, none of the models was able to capture the rapid variation noted in the later afternoon (16:00–17:00) period. The NSRDB model over-predicted surface albedo compared with any of the models and the ground truth.

Despite a high degree of variability in surface albedo observations in March (Figures A23 and A24), the objective model generated predictions that were comparable to ground truth and trend. Notably, predictions from MLP-Mixer were closest to the ground truth across time periods 10:00–12:00. Similarly, the objective model, Multispectral-WaveMix, predictions were closer to the ground truth across the 7:00–8:00 and 13:00–16:00 time periods. MLP-Mixer generated lower error predictions in periods between 09:00 and 12:00, followed by image-to-image mix between 8:00 and 9:00. The proposed model predictions across Autumn (Figures A25 and A26) captured the trend of ground truth observations across all except late afternoon and produced the nearest observation in periods 07:00–09:00 and 13:00–15:00. MLP-Mixer generated the close predictions between 11:00 and 13:00, followed by the Standard MLP between 10:00 and 11:00.

4.5. Model Uncertainty Assessment

The predictive uncertainty of the Multispectral-WaveMixer model was assessed by generating 1 min predictions over 50 cycles for the location of Golden, Colorado, across seasons. The 1 min predictions were used to obtain the mean prediction across 50 predictions and plotted against ground observation, NSRDB predictions, and a single prediction from the Multispectral WaveMixer. Predictions for the selected day in January show the spread of predictions was narrowly centered around the mean (Figures A27 and A28). Further, for the time period 13:00–16:00, the predictions overlapped with ground truth observations. During the spring season, (Figures A29 and A30) predictions over a day in March showed a wider spread around the mean, with a higher number of outliers, consistent with higher variability noted in the ground observations of surface albedo.

However, a significant portion of the model predictions between the hours of 8:00–12:00 overlapped with ground observations while the mean of the predictions followed the general trend of the ground observations. Similarly, the prediction for the earlier and later hours of the summer season (Figures A31 and A32) showed greater variability around the mean with several outlier predictions. The mean of the predictions was consistent with the trend of the ground observations. The predictive uncertainty plots from an autumn day in October (Figures A33 and A34) show predictions closer to the mean with a greater number of outliers in the hours between 11:00 and 15:00. Furthermore, the model predictions during the period between 08:00 and 0:00 and 14:00 and 16:00 included ground truth observations from the site. Notably, across all seasons, the model predictability was poor during the afternoon hours of 16:00 and 17:00. Overall, the uncertainty plots indicate the potential of the Multispectral WaveMixer model to generate a range of predictions within a constrained margin across the mean, providing a practical means to model various seasonal scenarios.

5. Discussion

5.1. Ablation Studies

The influence of the model components on its overall performance was assessed through ablation studies. Such studies require the removal of model components to benchmark the contribution of each component to the overall model performance. Firstly, the influence of the Wavelet component over the Standard MLP-Mixer approach was addressed earlier, with the overall model comparison demonstrating the performance advantages with lower Cauchy loss, higher peak signal-to-noise ratio, and enhanced structural similarity between the ground truth and model predictions.

Next, the contribution of various methods for aggregating the convolved spectral approximation and detail components of the wavelet was assessed using model performance

metrics (Table 5), RMSE plots (Figures A36 and A37), and predicted temporal shift heat maps (Figure A35).

Table 5. Multispectral-WaveMix with different component and aggregation approaches.

Component/Aggregation Methods	Cauchy Loss	SSIM	PSNR
With Max, min, mean and sum	0.00469/0.00524	0.99904/0.99887	72.9506/72.5697
Without any aggregation	0.00530/0.00548	0.99892/0.99878	72.2601/72.1256
Maximum as aggregation	0.00544/0.00494	0.99887/0.99891	72.1611/72.7204
Minimum as aggregation	0.00552/0.00459	0.99884/0.99899	72.2390/72.9432
Mean as aggregation	0.00540/0.00518	0.99889/0.99888	72.2374/72.355
Summation as aggregation	0.00546/0.00529	0.99887/0.99883	72.1828/72.4626
Without aggregation/Elastic Smoothing	0.00538/0.00669	0.99899/0.99878	71.6345/70.2665
Without Elastic Smoothing	0.00455/0.00508	0.99900/0.99884	73.0805/72.9114

The objective model, Multispectral-WaveMix, utilizing a combined maximum, minimum, mean, and summation for the aggregation of multispectral approximation and detail components before the reconstruction of wavelets, demonstrated the best training performance metrics compared with models constructed without any aggregation methods or singular approaches (maximum, minimum, mean, or summation). Though models with maximum and minimum aggregation approaches showed better validation performance, a visual analysis of predicted temporal shift heat maps (Figure A35) shows the presence of significant artifacts compared with predictions from the proposed Multispectral-WaveMix mode. A further analysis with RMSE plots for all observation time points by sites was conducted. RMSE as a metric is highly sensitive to the presence of outlier values due to the squaring of the efforts between ground truth and model predictions [108]. This enables the assessment of model sensitivity to outlier forecasts. In this context, the RMSE plots (Figures A36 and A37) indicate the use of combined maximum, minimum, mean, and summation for the aggregation results in a model that is comparatively least affected by outlier forecasts.

5.2. Computational Complexity and Efficiency

The computational complexity and efficiency of the objective model, Multispectral-WaveMix, were benchmarked against peer models according to trainable parameter counts, count of Multiply-Accumulate Operations (MACs), and the number of floating-point operations (FLOPs). As indicators of computational cost, FLOPs encompass all mathematical operations on floating-point numbers. Representing computational complexity, MACs provide a count of multiply and accumulate operations, which are synonymous with linear algebra operations during forward and backward propagation in neural networks. The benchmarking was carried out using the calcflops [109] library.

As listed in Table 6 among the mixer models, the Multispectral-WaveMix has comparatively lower computational complexity with lower multiply-accumulate and floating-point operations. The Standard MLP model with 1D time series as input has the lowest computational complexity, while the image-to-image model was the most computationally intensive.

Table 6. Computational efficiency of modeling approaches

Models	Trainable Parameters	MACS (Forward + Backward)	FLOPs (Forward + Backward)
Multispectral-WaveMix	6.8569 M	90.5902 M	186.526 M
Image-to-Image MLP Mix	61.5546 M	3.722 G	7.5844 G
MLP Mix	1.4558 M	98.5498 M	203.781 M
MLP	6.8612 M	1.8739 M	3.7709 M

6. Conclusions, Limitations, and Further Research Outlook

6.1. Conclusions

In this research, we focused on generating granular temporal scale surface albedo with remote sensing and ground station observations for modeling its intra-day variability for use in bifacial solar and environmental applications. Leveraging satellite-sensed multispectral surface reflectance imagery at a 30 m scale from Landsat and Sentinel-2A and 2B satellites and granular ground observations from SURFRAD surface radiation monitoring sites, we proposed a novel imaging time series and deep learning approach for sensor-to-ground temporal downscaling. Focused on seven study sites with varying degrees of land cover, the developed approach was validated with model performance metrics, visual assessment, and uncertainty plots.

As an essential climate variable, surface albedo is known for its influence on the thermal state of our environment. However, it is also a factor in the emerging new energy generation landscape with bifacial solar panels. Since surface albedo influences the amount of ground-reflected irradiance reaching the back side of bifacial panels, it is a key driver of energy gain of bifacial solar installations. Boosting the energy density of solar installations while mitigating land usage intensity, bifacial solar has demonstrated energy gains of 17% or more [20,21]. However, surface albedo varies during the sunlight hours, land cover, and other cloud and atmospheric factors. Further, the high-resolution monitoring of surface albedo requires extensive presence of ground-based monitoring or high temporal satellite coverage, both of which are either costly or technically challenging. Therefore, for solar power output planning, effective grid planning, assurance of grid stability, efficient energy market operations, and, more broadly, climate modeling, the granular temporal scale modeling of intra-day variations in surface albedo is essential. Considering this necessity, this research has (a). proposed a novel approach to generating visually intuitive image time series, temporally shifted heatmaps for increasing temporal information density and harnessing visual redundancy for deep learning. (b). utilized satellite-sensed multispectral surface reflectance imagery at a 30 m scale from Landsat and Sentinel-2A and 2B satellites and high temporal ground observations from SURFRAD surface radiation monitoring sites for intra-day surface albedo modeling, (c). developed a deep learning approach, Multispectral-WaveMix, for the sensor-to-ground temporal downscaling using wavelet decomposition and multispectral aggregation, and (d) assessed the potential of a pre-trained model for novel locations was evaluated for a SURFRAD site with granular ground observations.

The objective model, Multispectral-WaveMix, was benchmarked against peer models MLP-Mix and image-to-image mix using the 2D temporal shift heatmap image series and the Standard MLP model with 1D time series. With wavelet decomposition of multi-band satellite imagery (bands 1, 2, 3, 4, and 8), patch embedding, channel and token mixing, and the aggregation of multispectral approximation and detail components before reconstruction, the proposed model has demonstrated higher skill in capturing surface albedo trends across seasons while generating higher accuracy predictions comparable to intra-day ground truth surface albedo observations. Specifically, the main findings are as follows:

- Overall, the objective model, Multispectral-WaveMix, trained across all spectral bands (bands 1, 2, 3, 4, and 8), outperformed peer models MLP-Mix, Image-to-Image Mix, and Standard MLP with lower error margins compared with the ground truth. The objective model achieved Cauchy loss, a metric that is robust to outliers, of 0.00469 and 0.00524 for training and validation, respectively, the lowest among all other peer models tested.

- The Multispectral-WaveMix also achieved higher values on PSNR, an indicator of the quality of the pixels reconstructed by the models in contrast to noise distortion. It achieved higher values on both training and validation with 72.95060 and 72.56970, respectively, while peer models performed comparatively poorly, with image-to-image MLP Mix achieving 71.52140 and 71.89770, 75.38; MLP Mix, 71.89280 and 71.62610; and Standard MLP, 71.50470 and 71.52330, respectively.
- The model's ability to reconstruct the image consistent with the ground truth demonstrated the higher skill of the proposed Multispectral-WaveMix with structural similarity of 0.9990 and 0.9989 on training and validation. Image-to-image MLP Mix obtained 0.9985 and 0.9986, and MLP Mix achieved 0.9986 and 0.9985 on training and validation, respectively.
- Analysis of the performance of the temporal shift heatmaps as image time series noted models utilizing GASF significantly underperformed across all model performance metrics. The Multispectral-WaveMix trained on GASF-based image time series achieved training and validation losses of 0.01786 and 0.01799, respectively, while peer models, image-to-image MLP Mix and MLP-Mix trained on GASF, achieved training vs. validation losses of 0.01939/0.019841 and 0.01942/0.01854, respectively, underperforming compared with TSH-based models.
- The overall performance of the objective model, Multispectral-WaveMix, also exceeded other benchmark models on root mean square error (RMSE) metrics when assessed across all observations by site. Overall, the objective model outperformed all other models, achieving lower RMSE errors across the Desert Rock, Nevada; Goodwin Creek, Mississippi; Penn. State Univ., Pennsylvania; and Golden, Colorado test sites. The models achieved the second-best performance in two other sites, Table Mountain, Colorado, and Fort Peck, Montana.
- On assessing the generalizability of the pre-trained model on a location, Golden, Colorado, that wasn't part of the training corpus, the objective model, Multispectral-WaveMix, outperformed all other models with 69.6% of the RMSE errors falling below 0.001, while image-to-image Mix, MLP-mix, and Standard MLP achieved 60.8%, 56.9%, and 44.3% of the errors, respectively.

6.2. Limitations, Further Outlook, and Future Research

The temporally-shifted heatmaps proposed in this research and the higher predictive performance of the proposed Multispectral-WaveMix and benchmark models trained on heatmaps present a better alternative to GASF-based approaches for temporal downscaling. The potential of the temporally-shifted heatmaps and the modeling approaches has the potential for further research, which is yet to be explored in this paper:

- In the context of estimating surface albedo via remote sensing, the solar irradiance reaching a surface is subjected to distortions from atmospheric factors including cloud cover, aerosol, and water vapor [110,111]. The combination of atmospheric and surface effects, such as vegetation cover, on the incoming irradiance is captured through satellite sensors in the form of multispectral surface reflectance. The effects of cloud cover vary across the electromagnetic spectrum, with cloud absorption affecting near-infrared more than the visible portion [112]. Therefore, spatially congruent, temporally high-resolution cloud cover observations could enhance model skill, especially for bands in the near-infrared range.
- Improving model generalizability requires training inputs encompassing a wider range of spatial locations and a greater temporal range to account for domain shifts and capture causal mechanisms. Such efforts are at present limited by the higher level of computational resources required for image processing model training and dataset

development. However, physics-informed deep learning methods that harness both physical parametrization and data-driven learning approaches have greater potential in bridging the gap between the need for greater computation and larger representative datasets. The latter approach would be of greater benefit to data-constrained regions of the world.

- Self-attention mechanisms are considered integral to the high performance of vision transformers. The attention-free modeling approaches, such as Multispectral-WaveMix, MLP-Mix, and their variants, have achieved equivalent performance without an attention mechanism. However, the combination of attention mechanisms with MLP-Mix has been explored to enhance global attention and positional encoding for image segmentation [113,114], image decoding [115], and image recognition [116].

Overall, the future of generating high temporal surface albedo for both solar energy applications and for the modeling of thermal environments could utilize pre-trained deep learning models. With large-scale inclusion of training samples from a diverse set of land cover and atmospheric conditions, such models could fill the gap created by the sparse deployment of high-temporal ground stations and high-spatial and low-temporal coverage of remote sensing satellites.

Author Contributions: S.K.: Conceptualization, Visualization, Investigation, Methodology, Software, Data curation, Writing—original draft. R.C.D.: Conceptualization, Visualization, Investigation, Supervision, Writing—review and editing. D.C.-P.: Writing—review and editing. N.R.: Supervision, Writing—review and editing. S.S.-S.: Supervision, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The first author is grateful for the University of Southern Queensland (UniSQ) Domestic PhD Scholarship, the Research and Training Scheme fee scholarship from the Australian Government, and the Australian Postgraduate Research Intern (APR.intern) opportunity funded by the Senetas Corporation for professional development in AI/ML for signal processing. This research was partially supported by project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation (MICINN) to build synergies between Professor Ravinesh Deo (UniSQ, Australia) and Sancho Salcedo-Sanz (UAH, Spain).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors thank the European Commission Joint Research Centre FWC932059 (part of the Global Component of the European Union’s Copernicus Land Monitoring Service GBOV “Ground Based Observation for Validation” (Access Date: 23 January 2025) <https://land.copernicus.eu/global/gbov>) for GBOV data that are managed by ACRI-ST with support from University College London, University of Leicester, University of Southampton, University of Valencia, and Informus GmbH.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Model Diagram

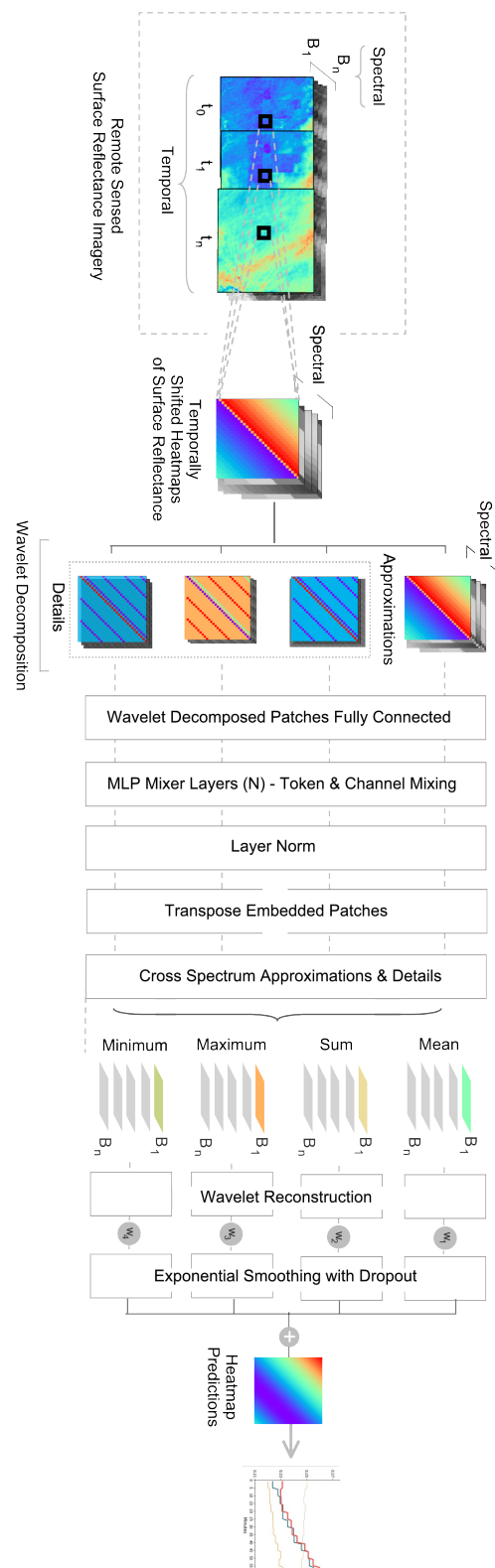


Figure A1. Overall model diagram of Multispectral-Wave Mix.

Appendix B. Model Performance Assessment—Training and Validation

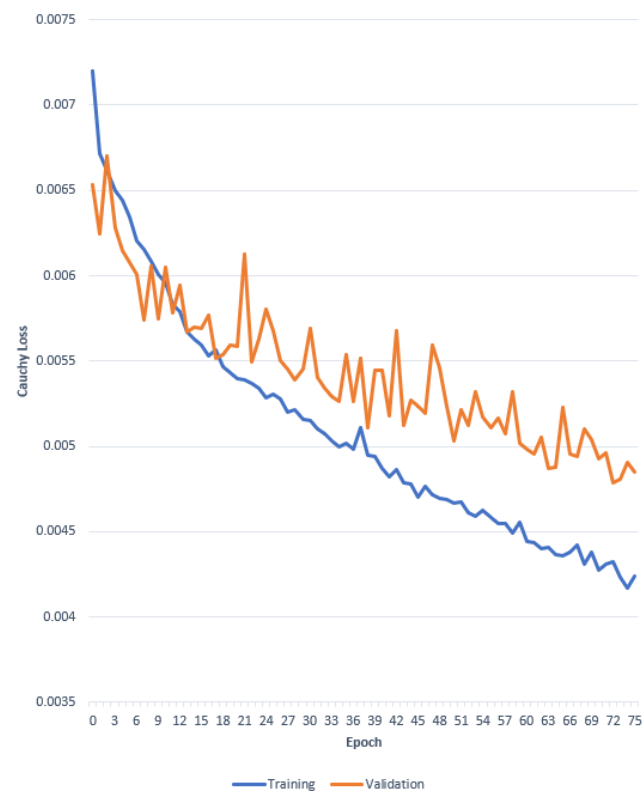


Figure A2. Training and validation performance of the proposed Multispectral-WaveMix measured by Cauchy loss vs. epochs.

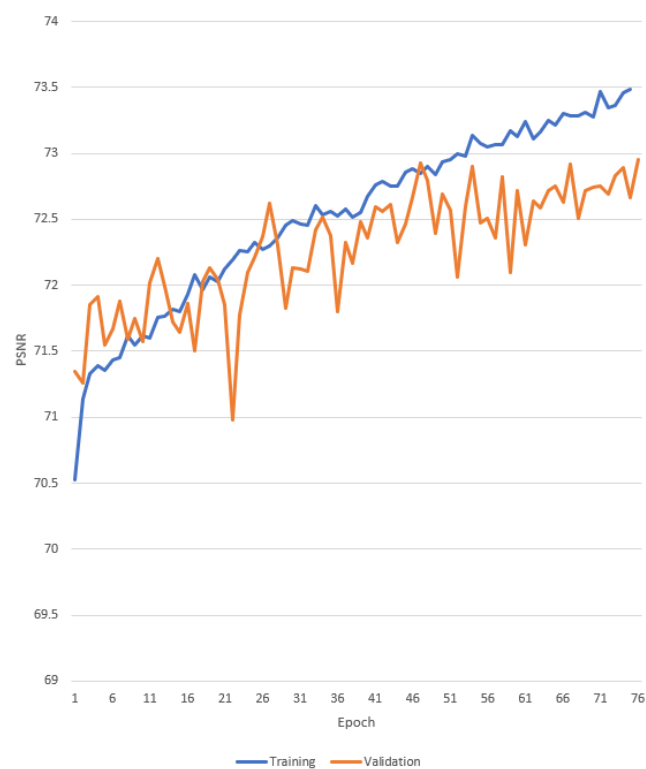


Figure A3. Training and validation performance of the proposed Multispectral-WaveMix measured by signal-to-noise ratio (SNR) vs. epochs.

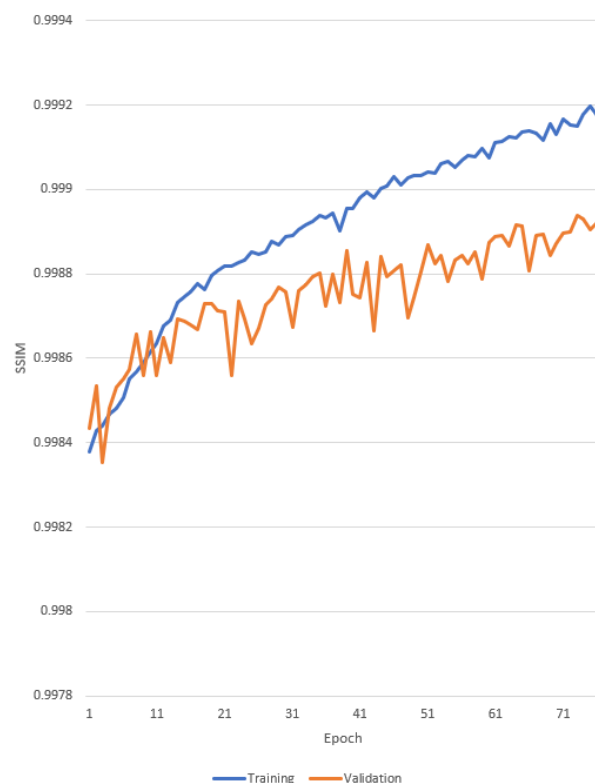


Figure A4. Training and validation performance of the proposed Multispectral-WaveMix measured by Structural Similarity Index (SSIM) vs. epochs.

Appendix C. Model Performance Assessment Plots—RMSE

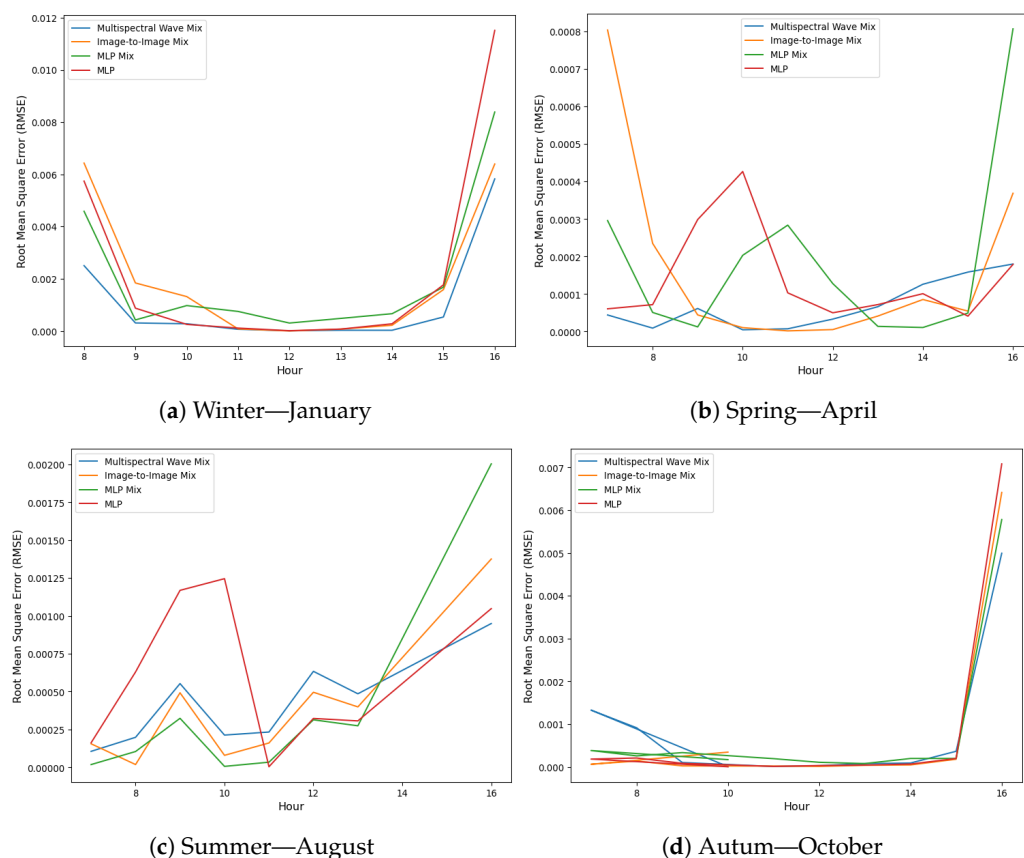


Figure A5. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Desert Rock, Nevada—2022.

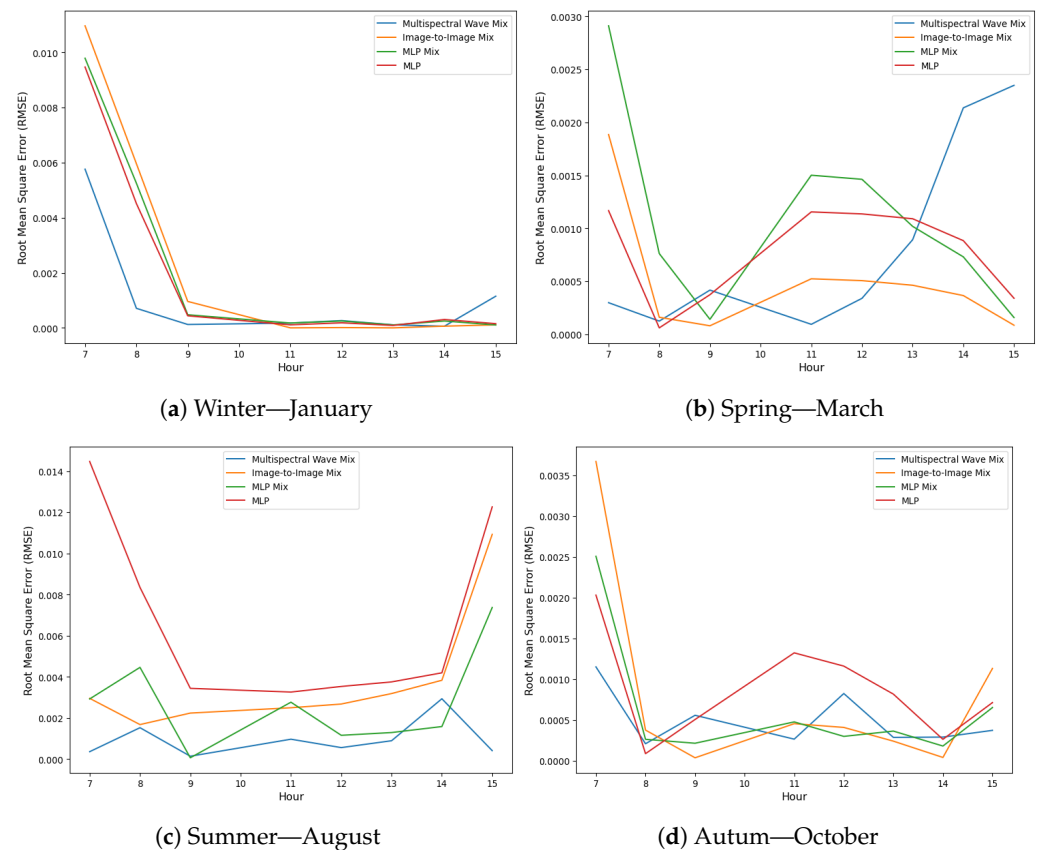


Figure A6. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Goodwin Creek, Mississippi—2022.

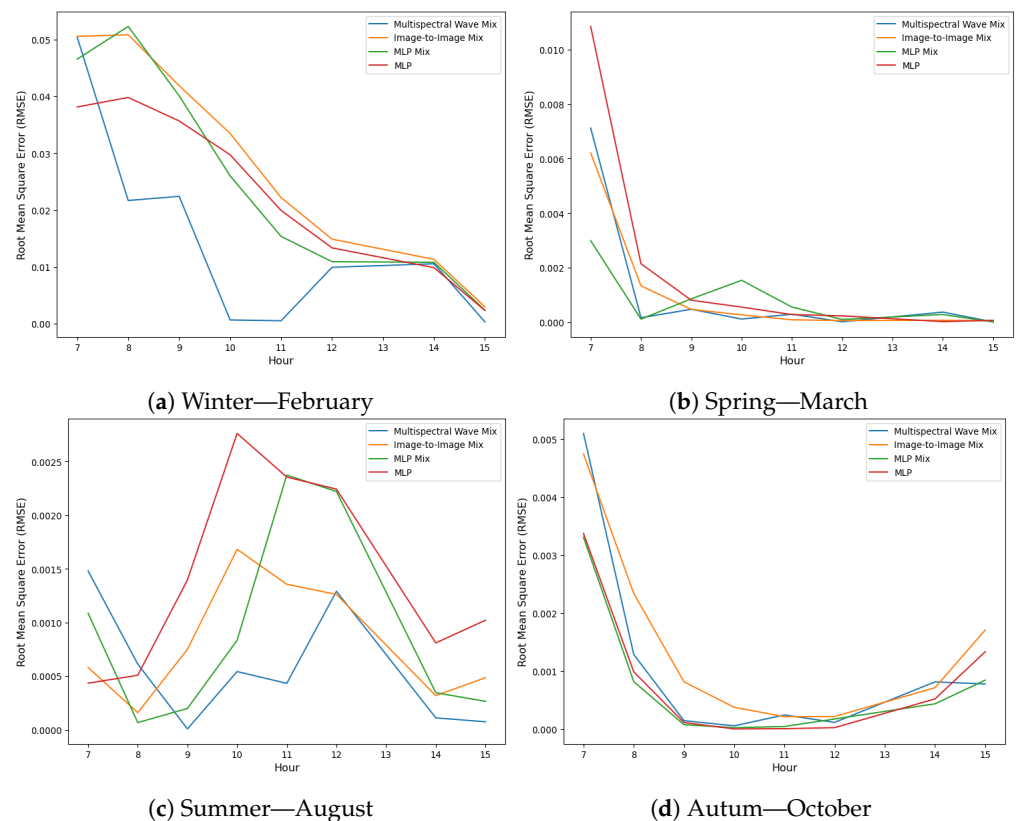


Figure A7. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Penn. State Univ., Pennsylvania—2022.

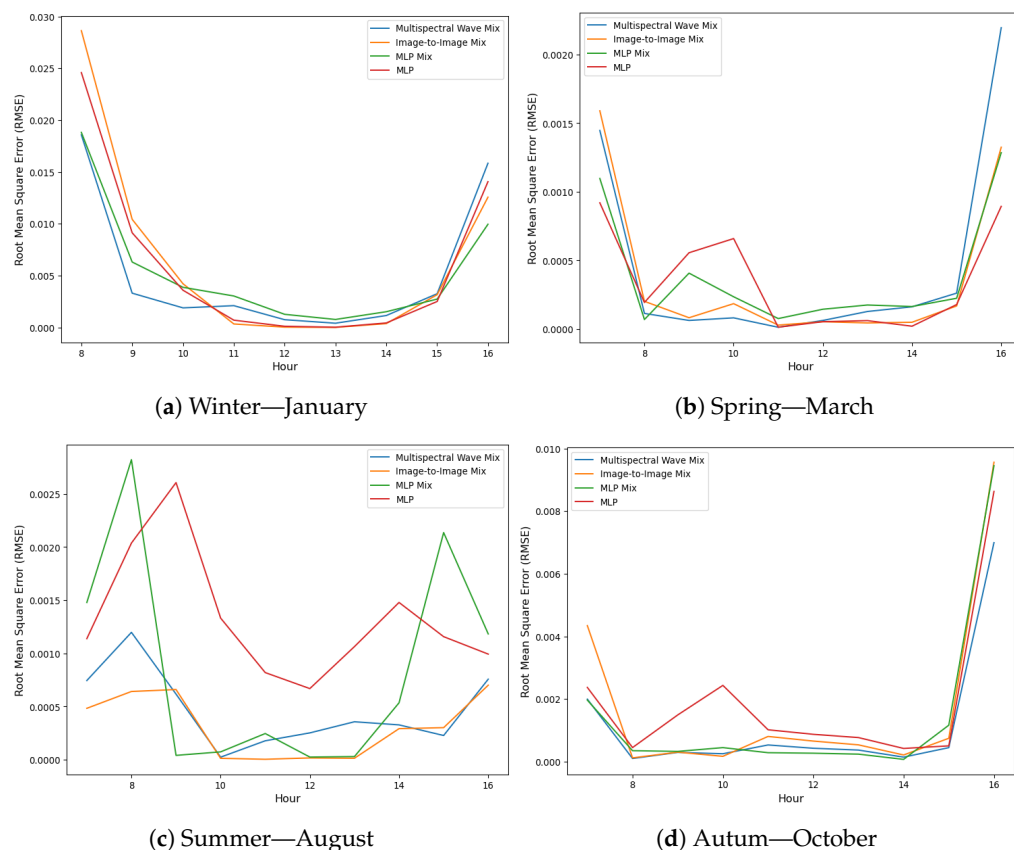


Figure A8. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Table Mountain, Colorado—2022.

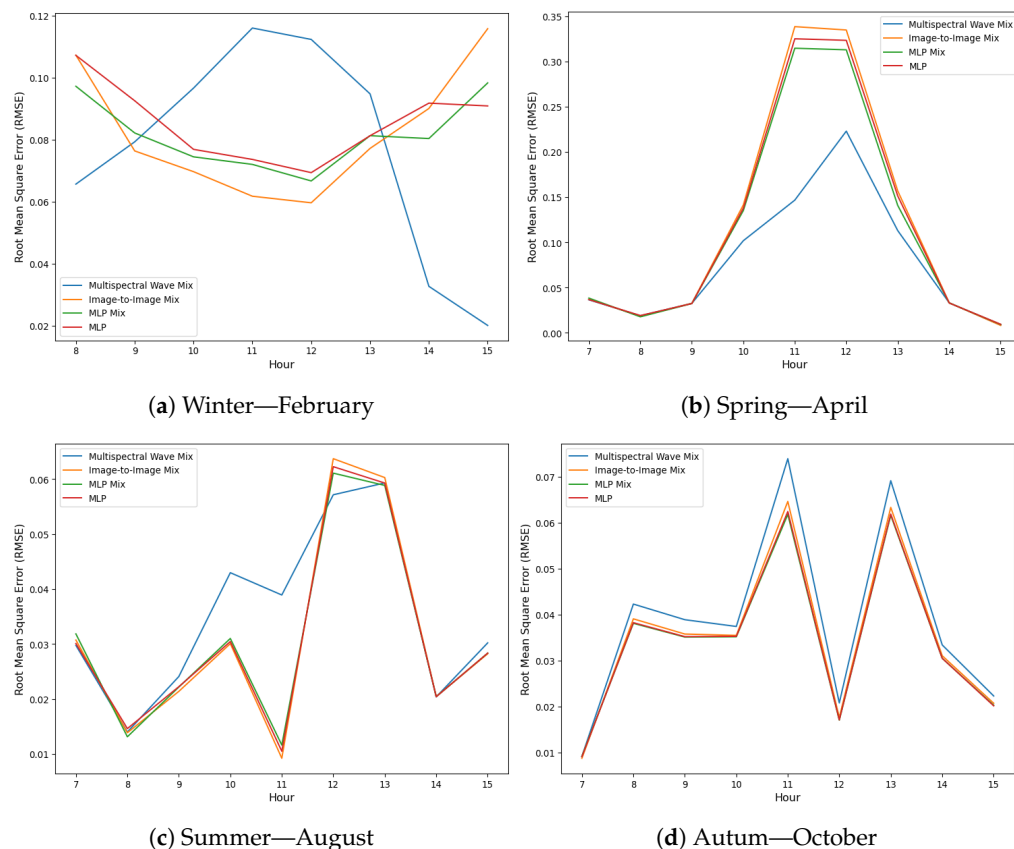


Figure A9. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Fort Peck, Montana—2022.

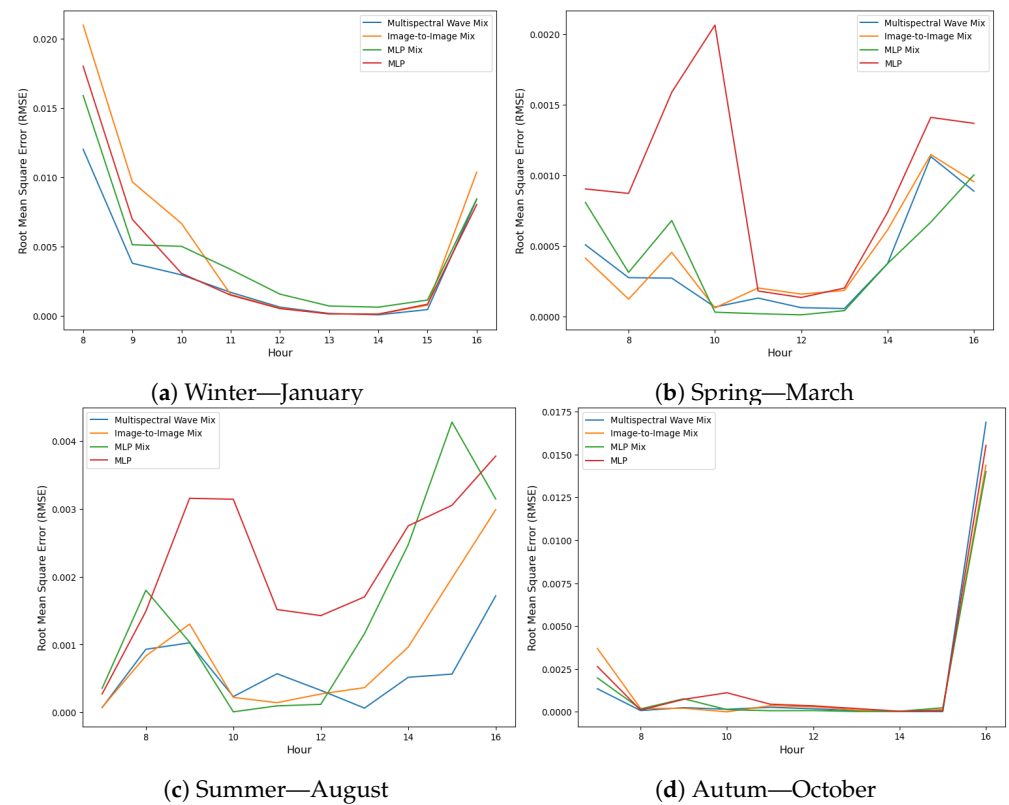


Figure A10. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Golden, Colorado—2022.

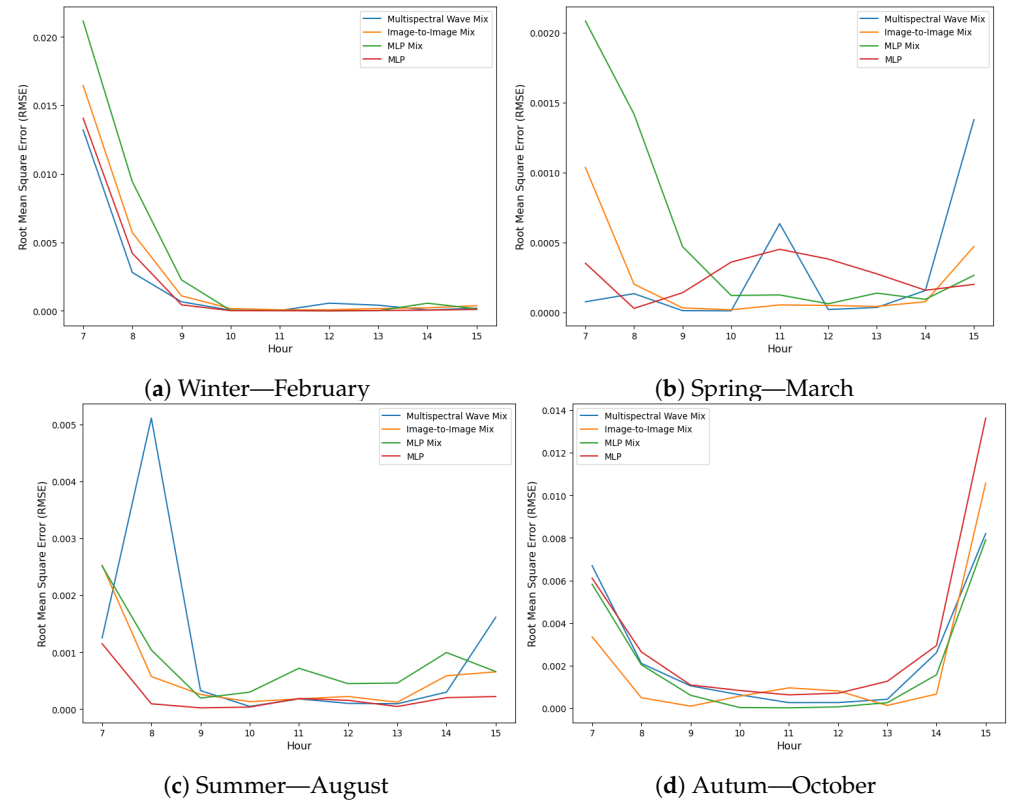


Figure A11. Plots of Root Mean Square Error (RMSE) of objective model, Multispectral-Wave Mix, peer models MLP-Mix, Image to Image MLP-Mix, and Standard MLP model for Bondville, Illinois—2022.

Appendix D. Model Performance Assessment—1 min Predictions by Time and Location

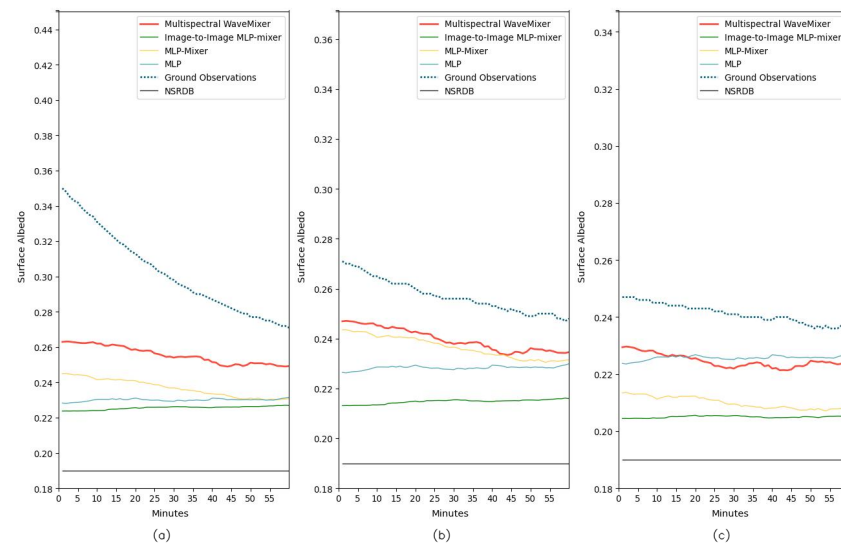


Figure A12. Plots of 1 min surface albedo predictions for periods (a) 8:00–9:00 a.m., (b) 9:00–10:00 a.m., (c) 10:00–11:00 a.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mixer, MLP-Mixer, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—January 2022.

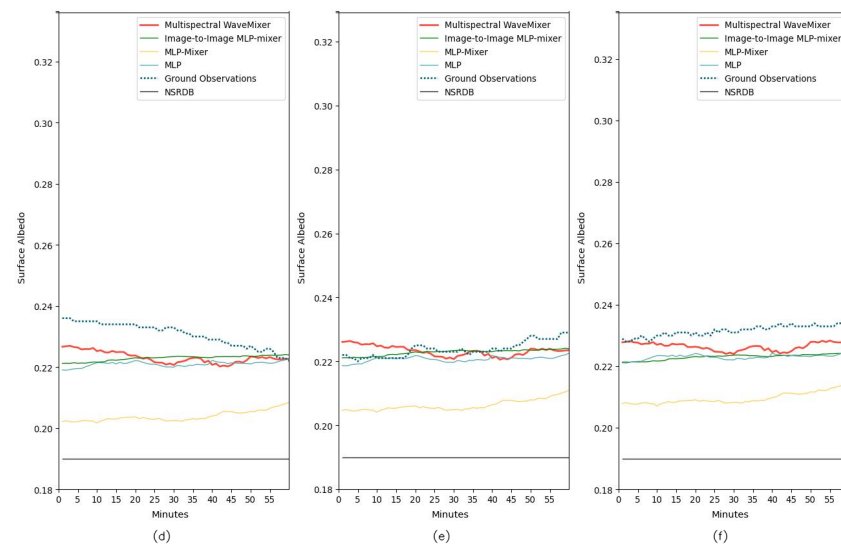


Figure A13. Cont.

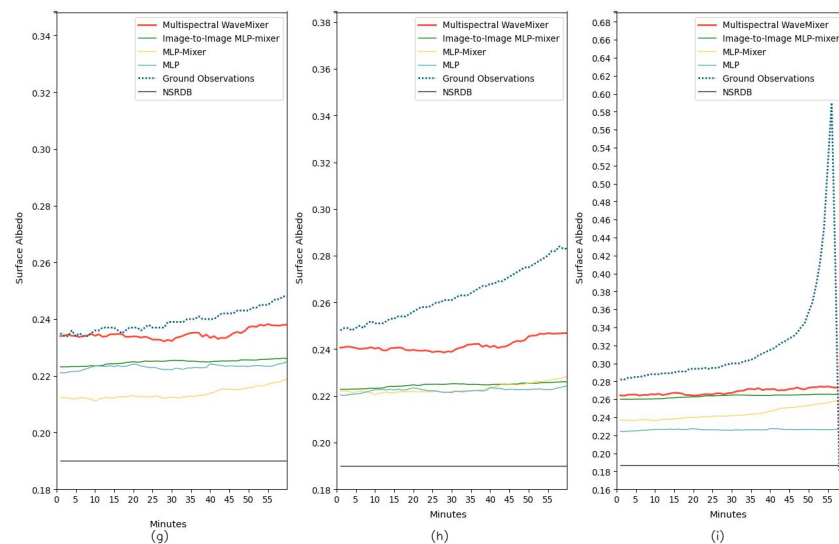


Figure A13. (Note—this figure continues from the previous figure) (d) 11:00–12:00 p.m., (e) 12:00–13:00 p.m., (f) 13:00–14:00 p.m., (g) 14:00–15:00 p.m., (h) 15:00–16:00 p.m., and (i) 16:00–17:00 p.m.

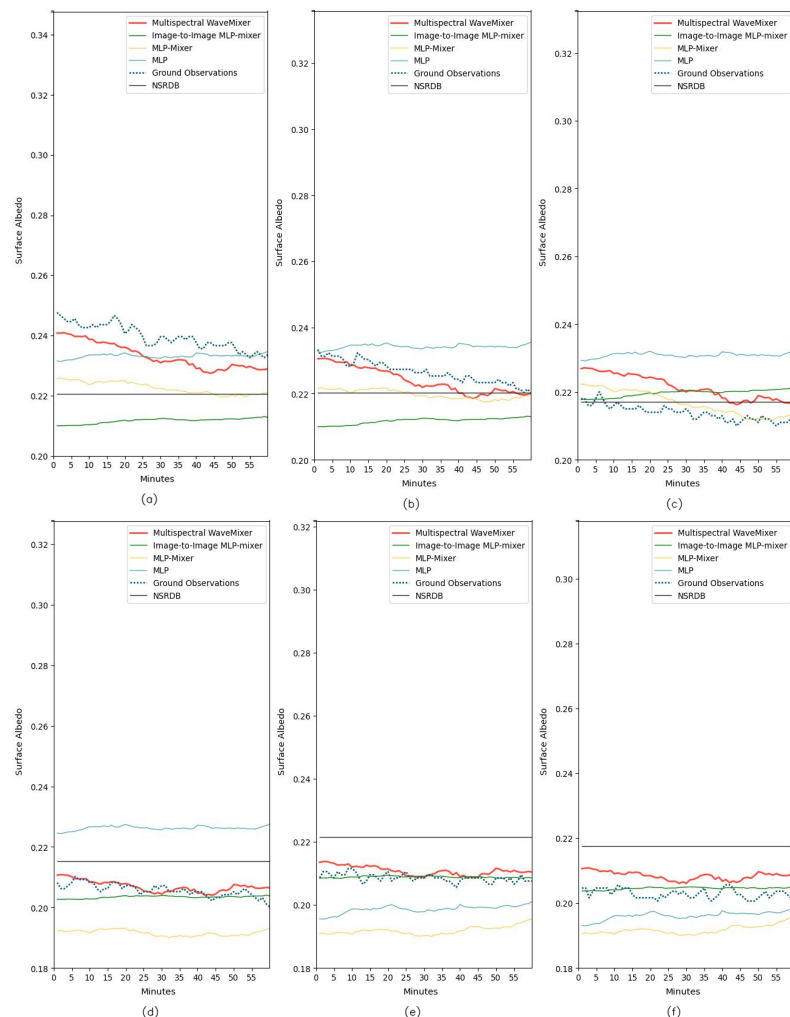


Figure A14. Plots of 1 min surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mixer, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—April 2022.

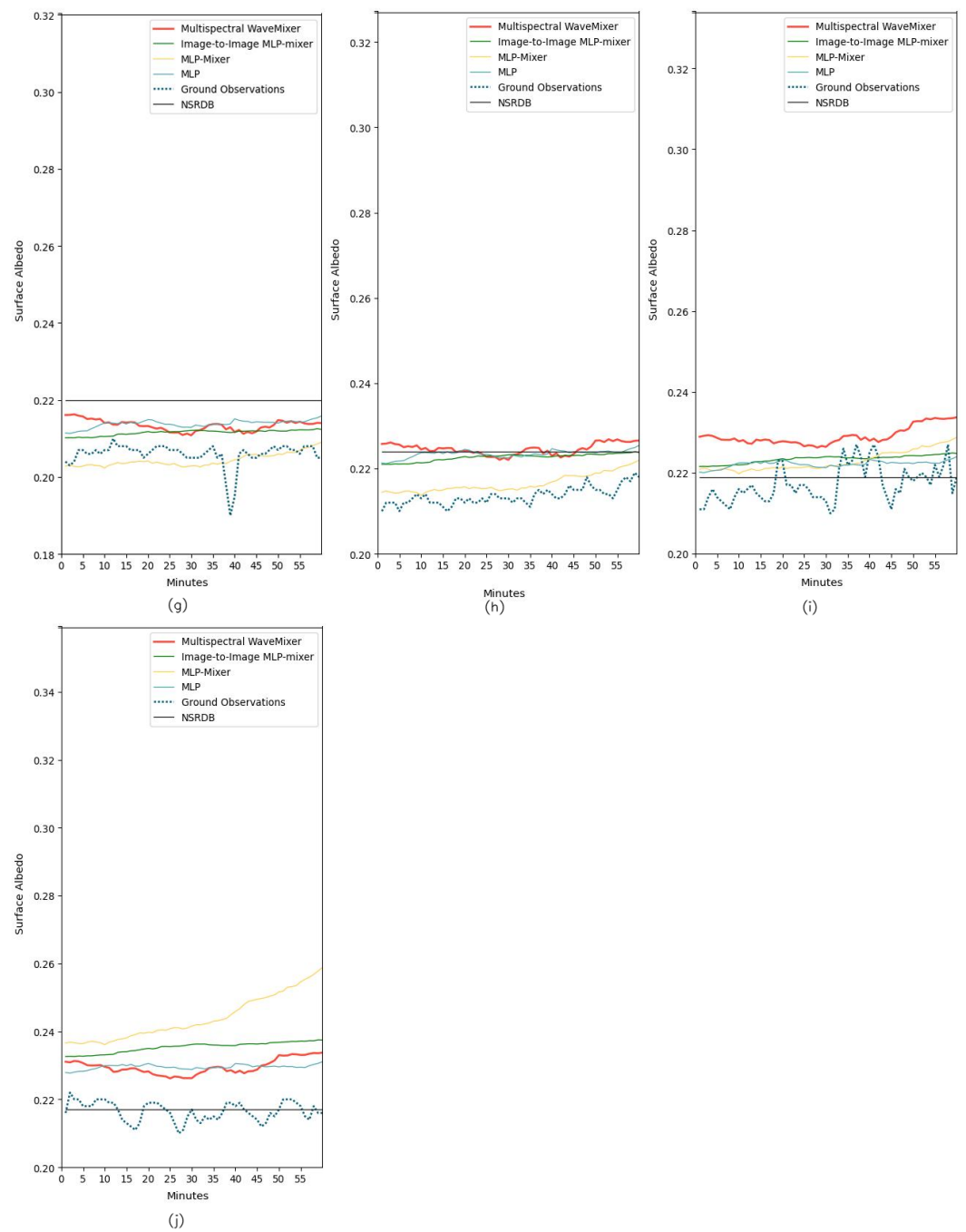


Figure A15. (Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

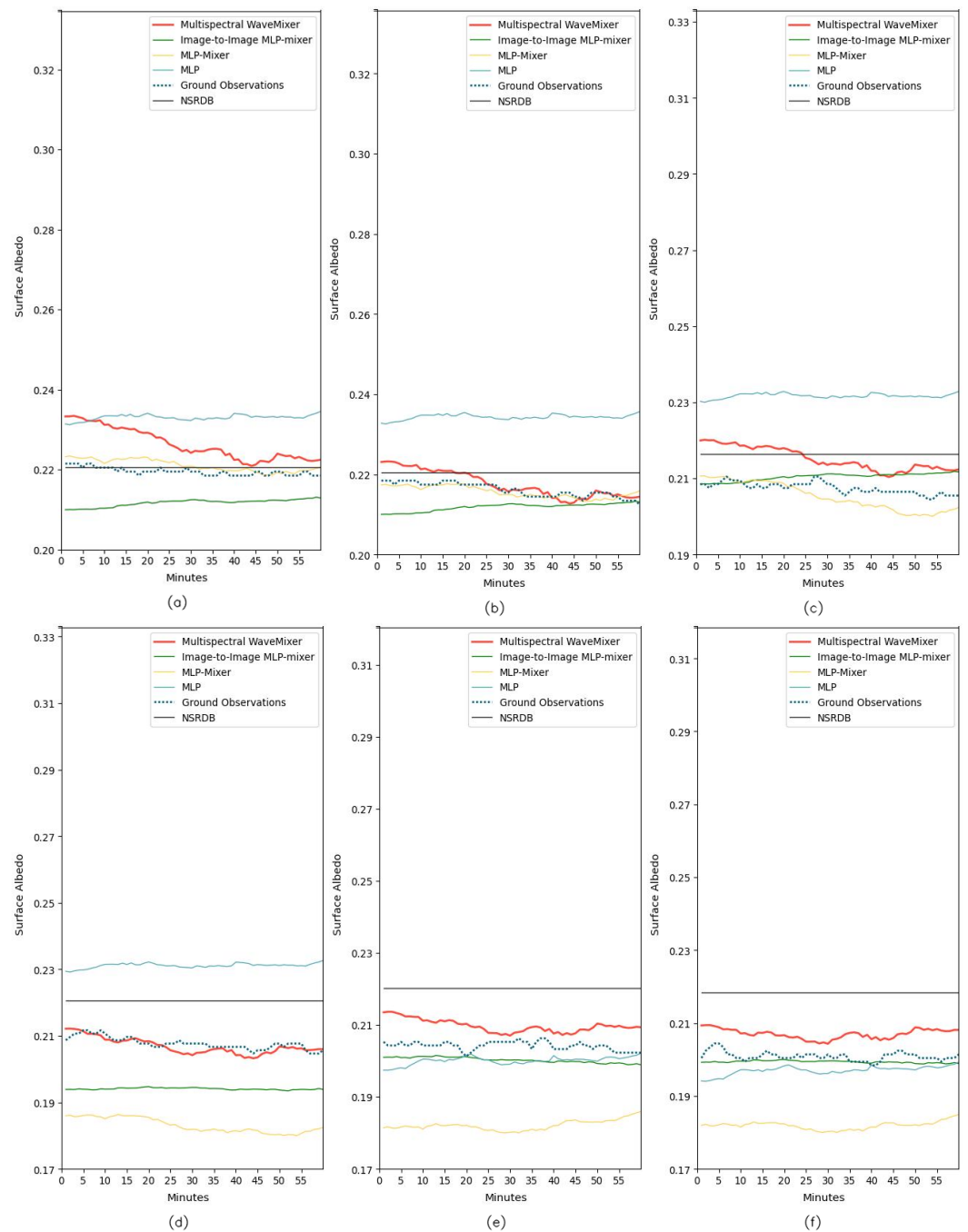


Figure A16. Plots of 1 min surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—June 2022.

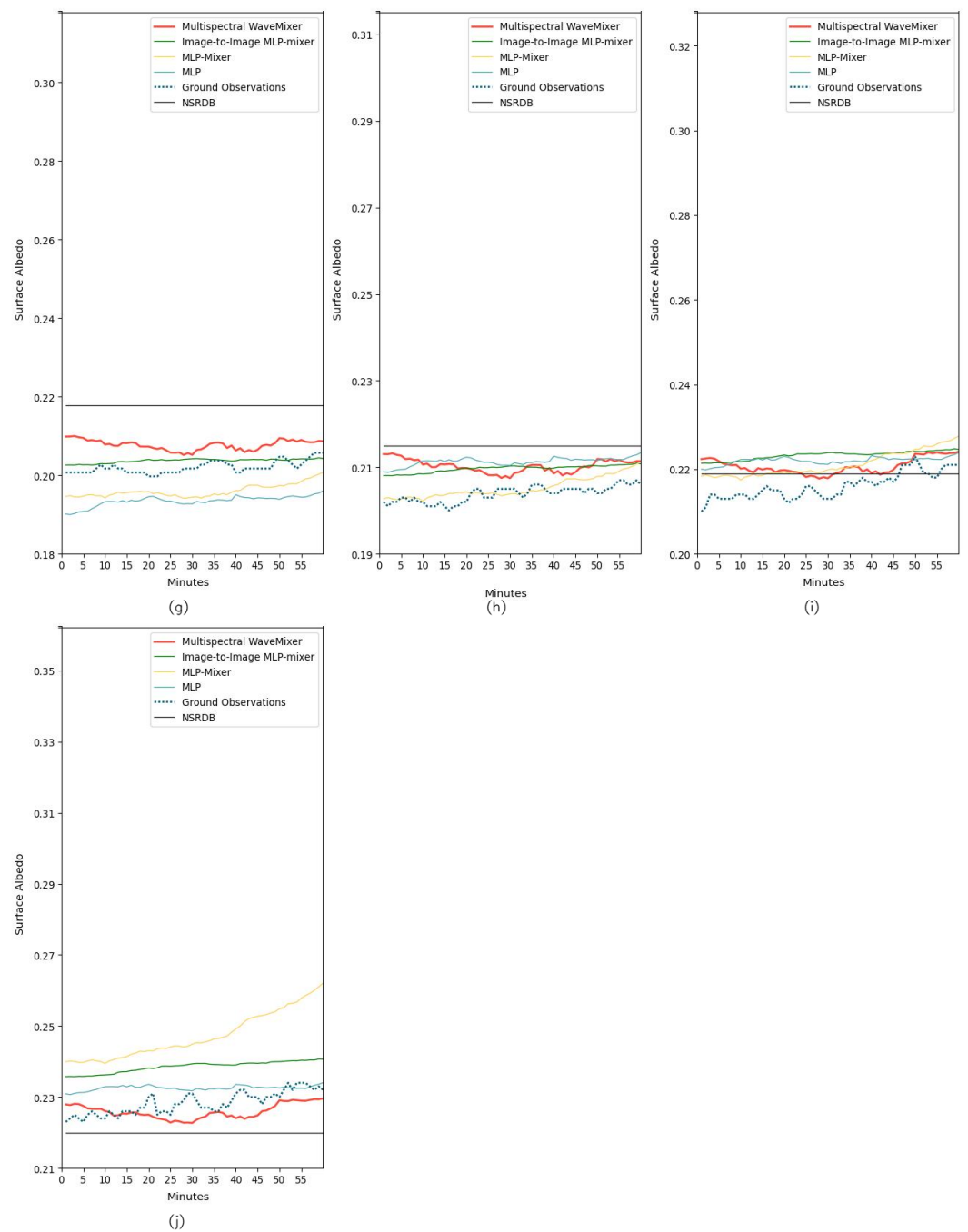


Figure A17. (Note—this figure continues from the previous figure)(g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

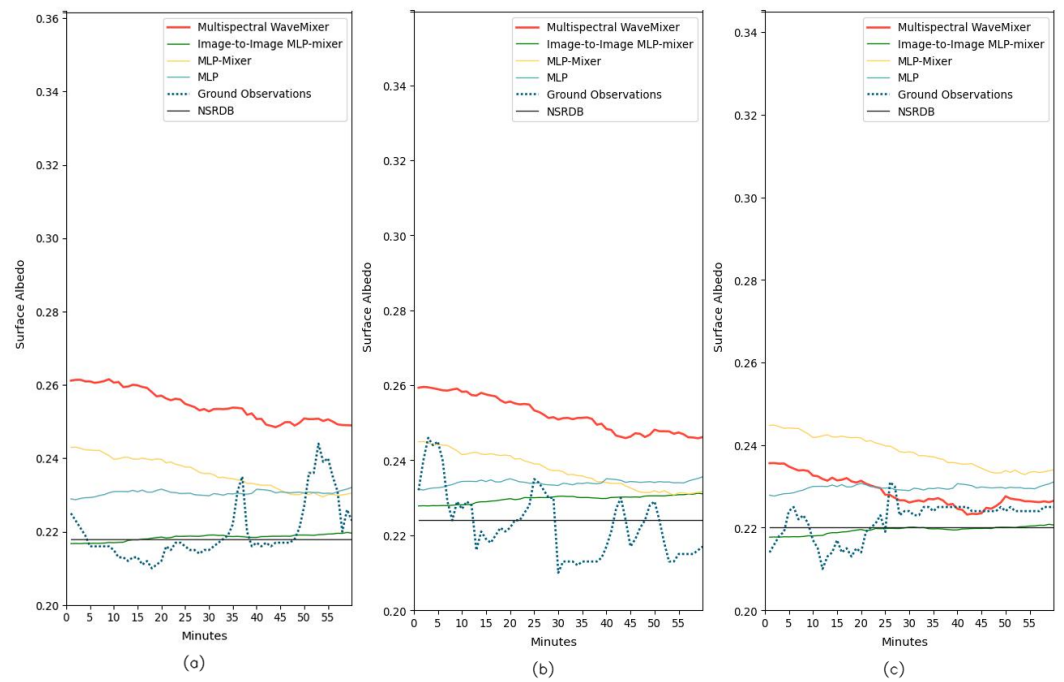


Figure A18. Plots of 1 min surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., and (c) 9:00–10:00 a.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—October 2022.

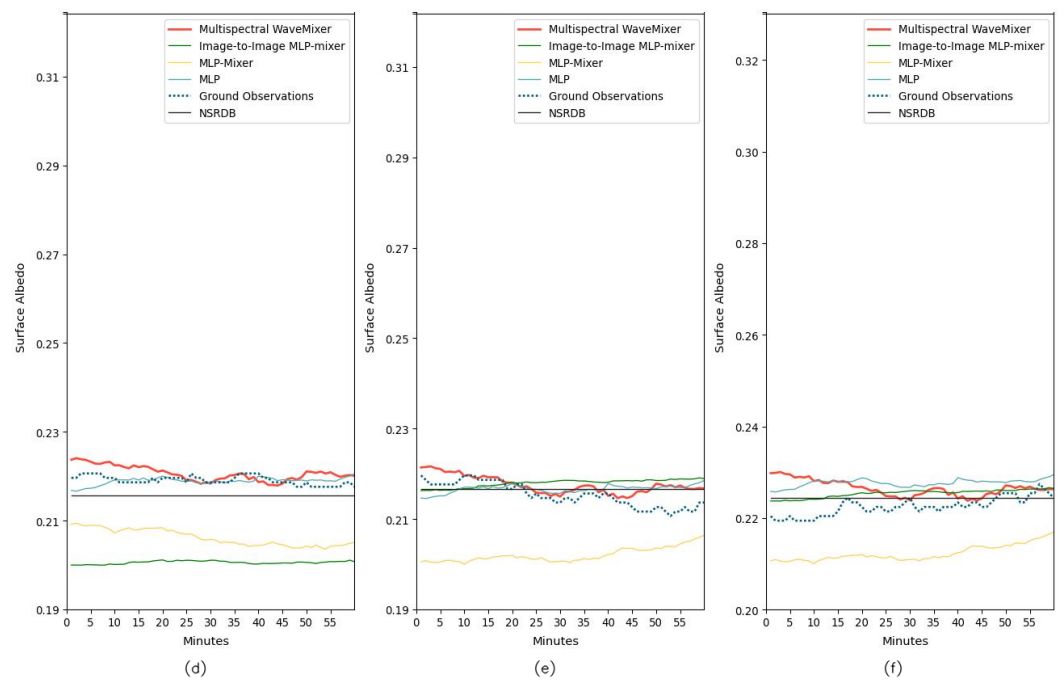


Figure A19. Cont.

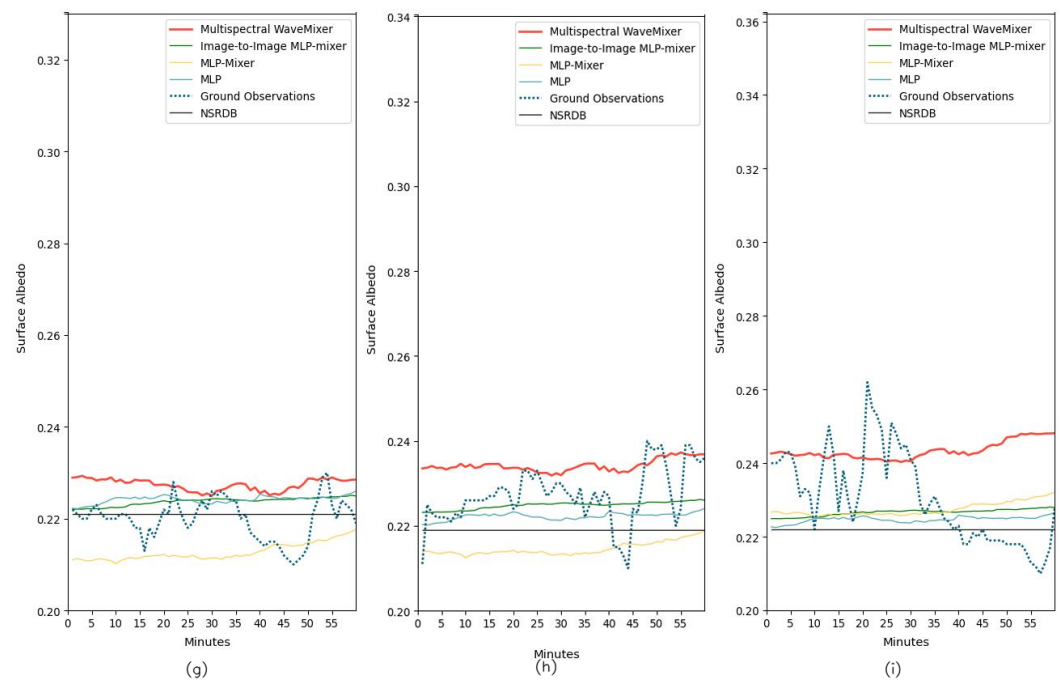


Figure A19. (Note—this figure continues from the previous figure) (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., (f) 12:00–13:00 p.m., (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., and (i) 15:00–16:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—October 2022.

Appendix E. Plots of Uncertainty Surface Albedo Predictions for Time Periods and Location

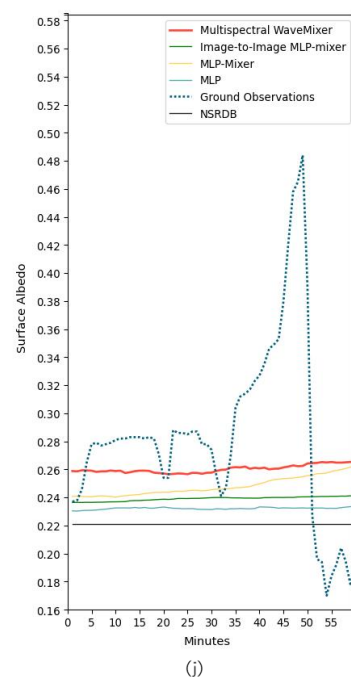


Figure A20. Plots of 1 min surface albedo predictions for time periods 16:00–17:00 p.m. PM from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Desert Rock, Colorado—October 2022.

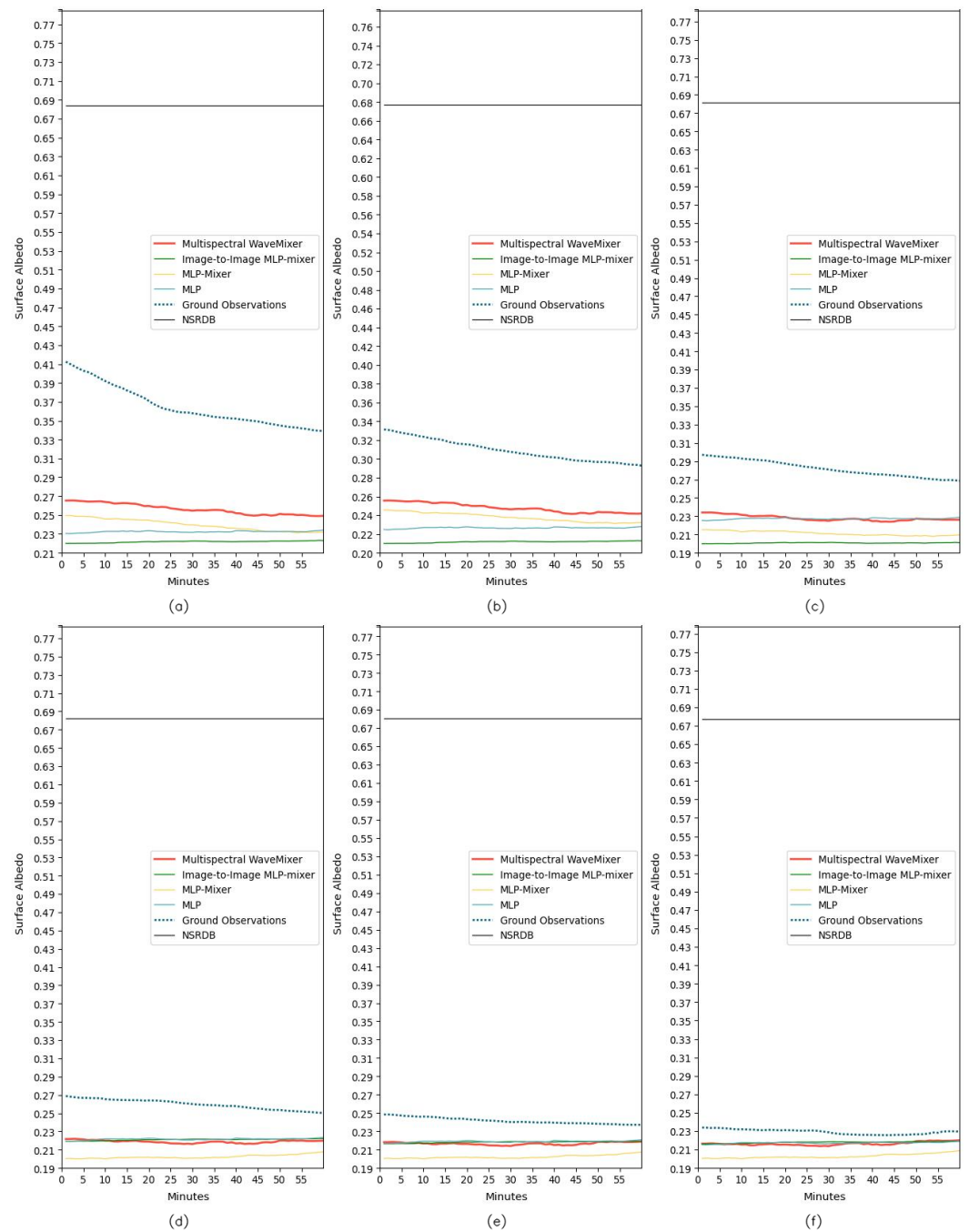


Figure A21. Plots of 1 min surface albedo predictions for time periods (a) 8:00–9:00 a.m., (b) 9:00–10:00 a.m., (c) 10:00–11:00 a.m., (d) 11:00–12:00 p.m., (e) 12:00–13:00 p.m., and (f) 13:00–14:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—January 2022.

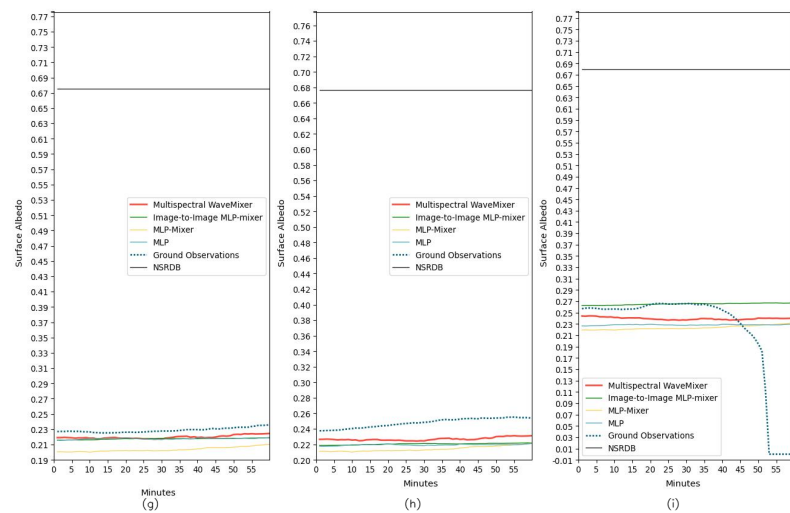


Figure A22. (Note—this figure continues from the previous figure) (g) 14:00–15:00 p.m., (h) 15:00–16:00 p.m., and (i) 16:00–17:00 p.m.

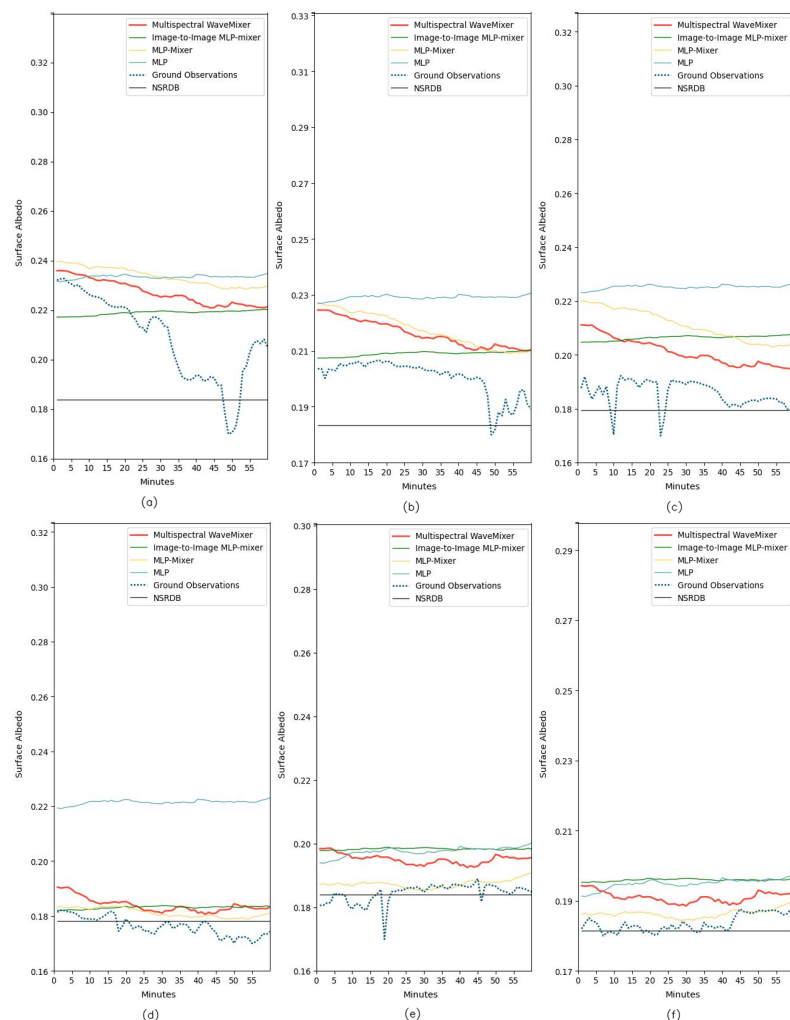


Figure A23. Plots of 1 min surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—March 2022.

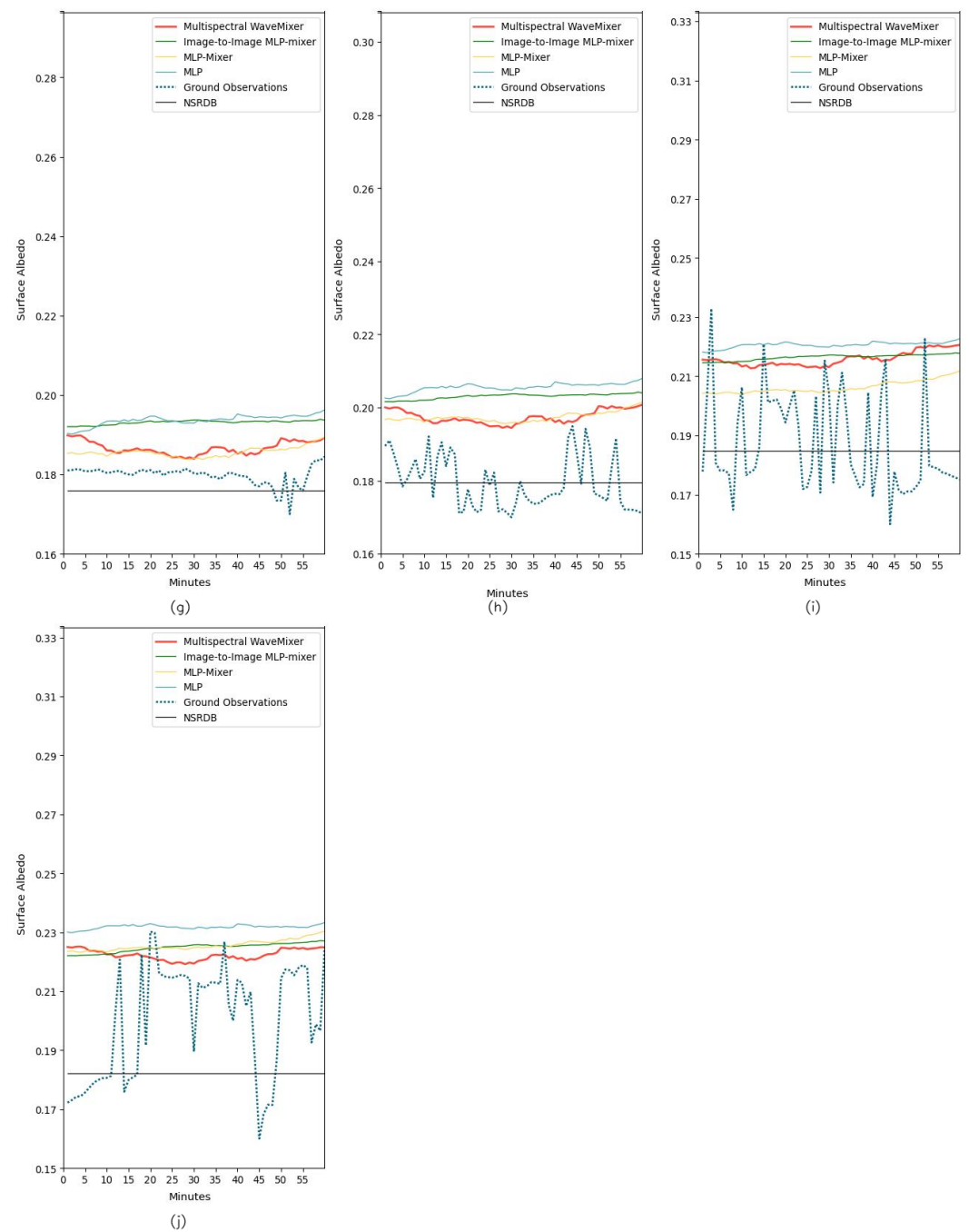


Figure A24. (Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

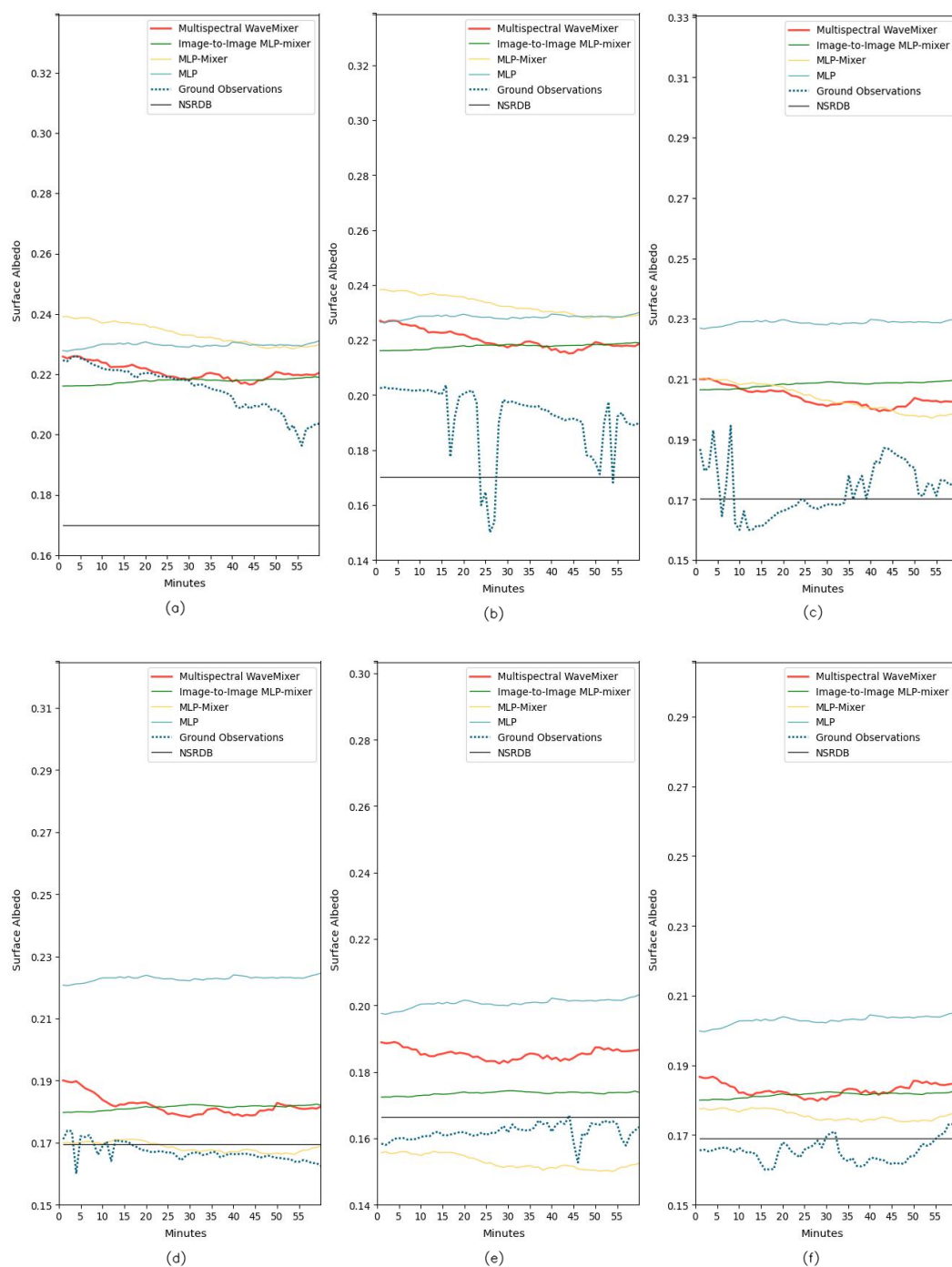


Figure A25. Plots of 1 min surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—August 2022.

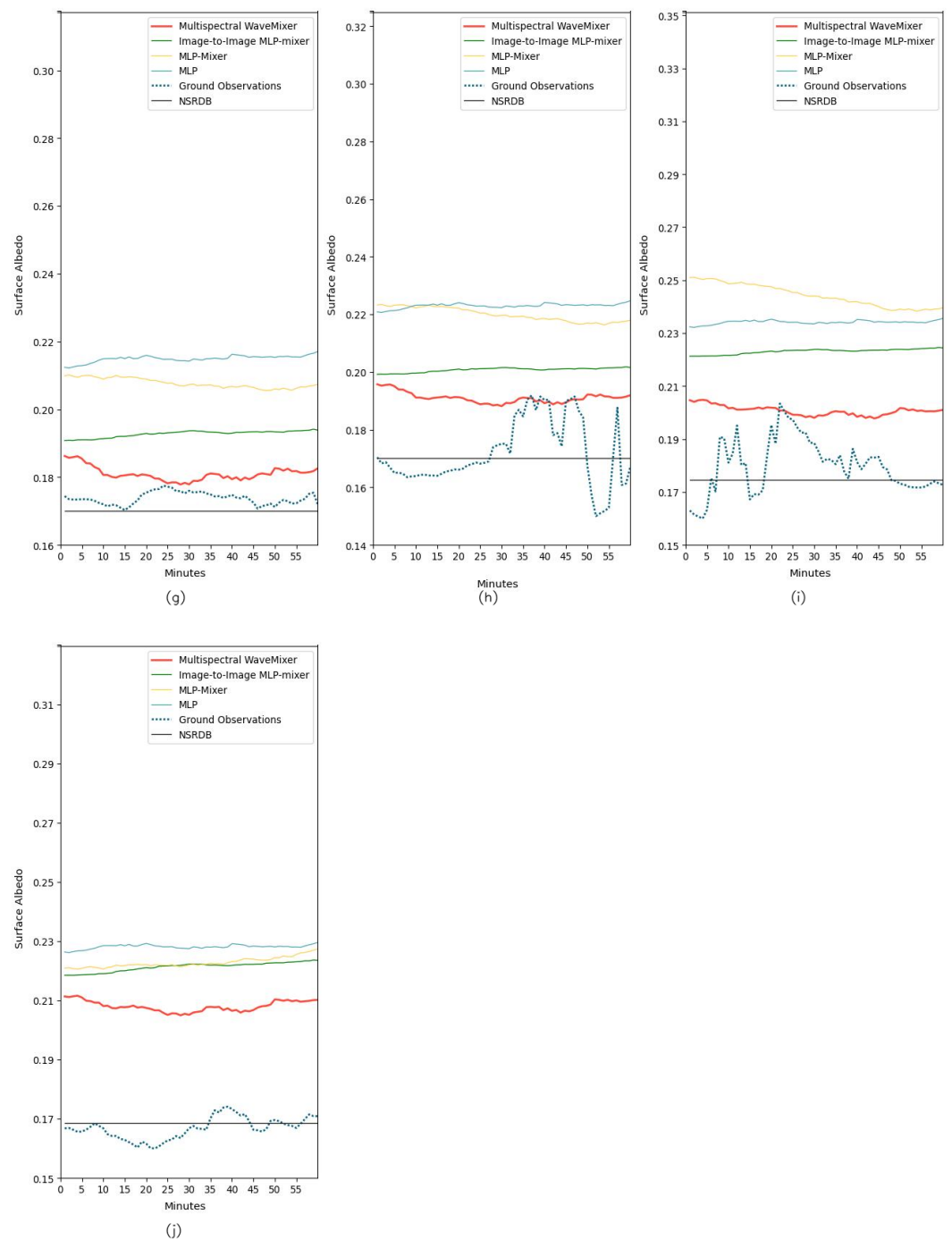


Figure A26. (Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

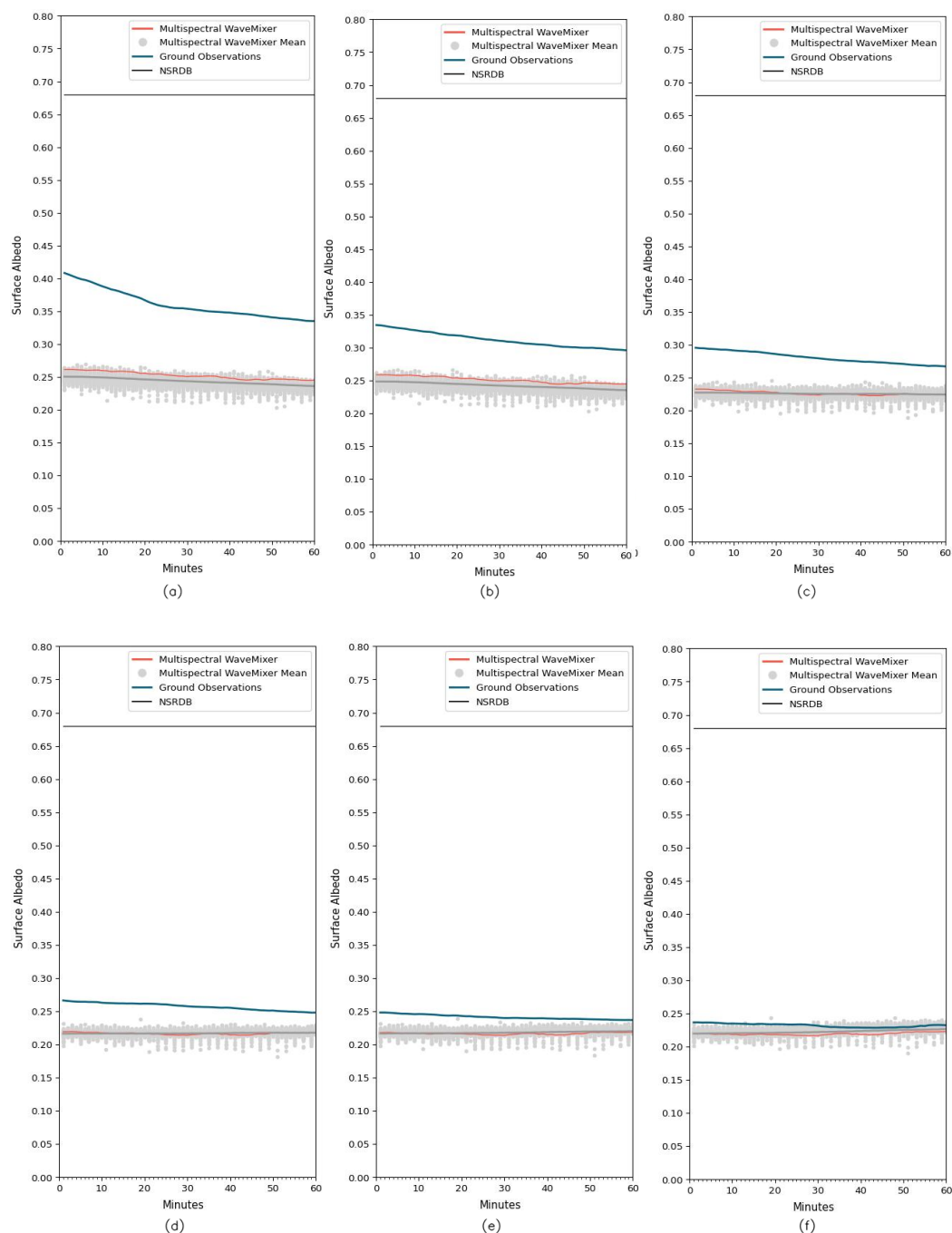


Figure A27. Plots of uncertainty surface albedo predictions for time periods (a) 8:00–9:00 a.m., (b) 9:00–10:00 a.m., (c) 10:00–11:00 a.m., (d) 11:00–12:00 p.m., (e) 12:00–13:00 p.m., (f) 13:00–14:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—January 2022.

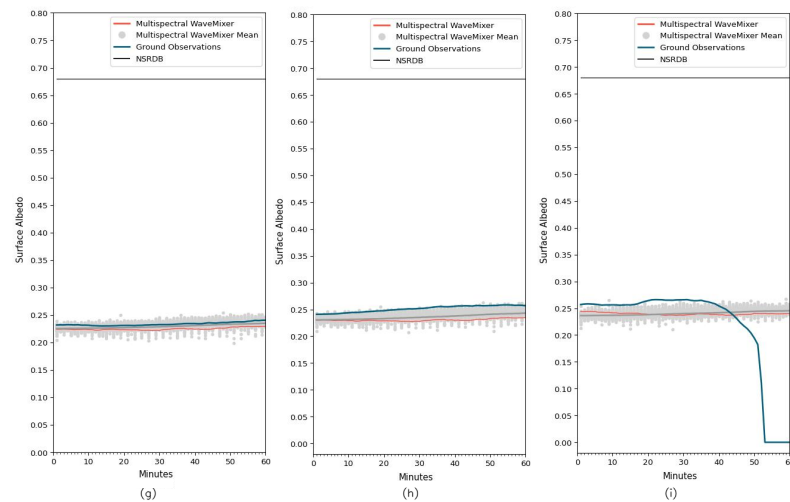


Figure A28. (Note—this figure continues from the previous figure) (g) 14:00–15:00 p.m., (h) 15:00–16:00 p.m., and (i) 16:00–17:00 p.m.

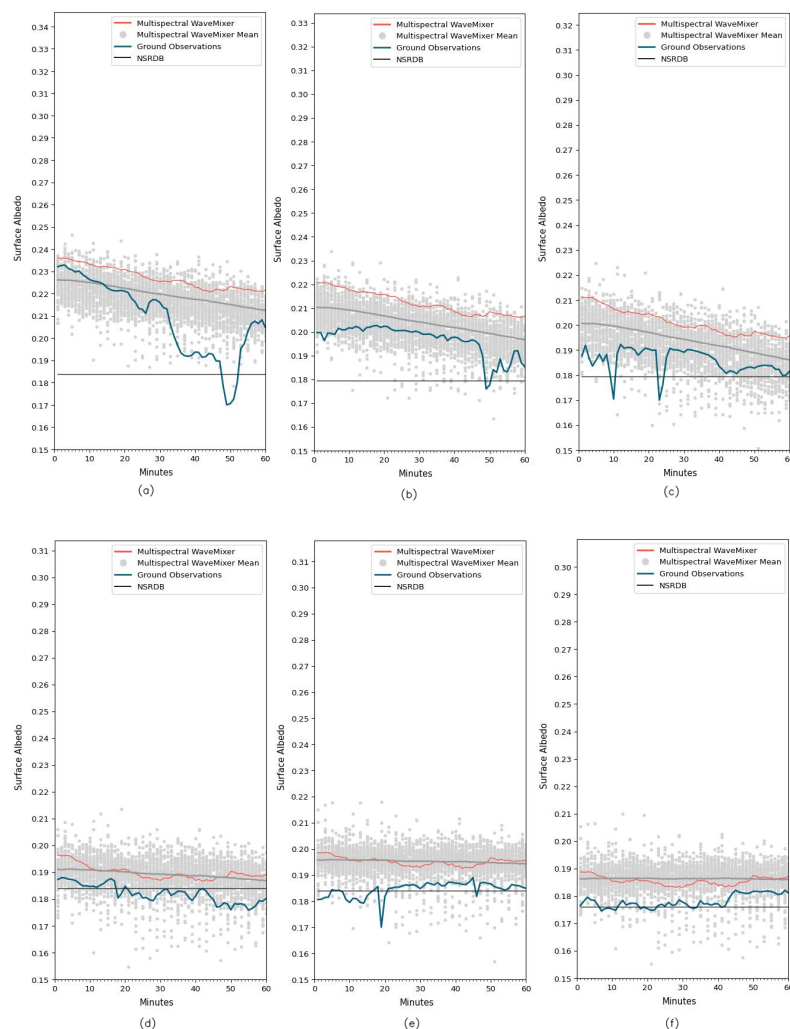


Figure A29. Plots of uncertainty surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—March 2022.

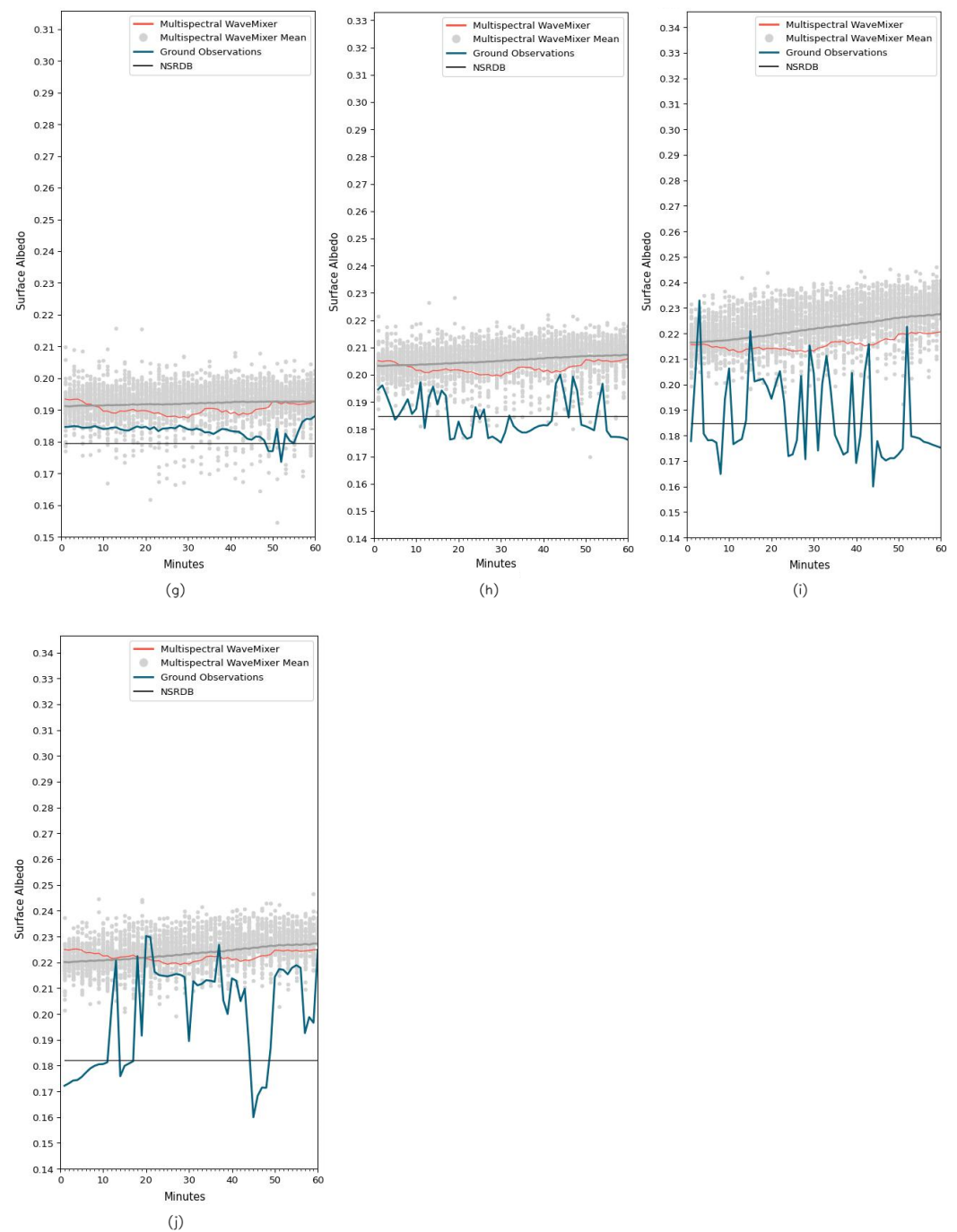


Figure A30. (Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

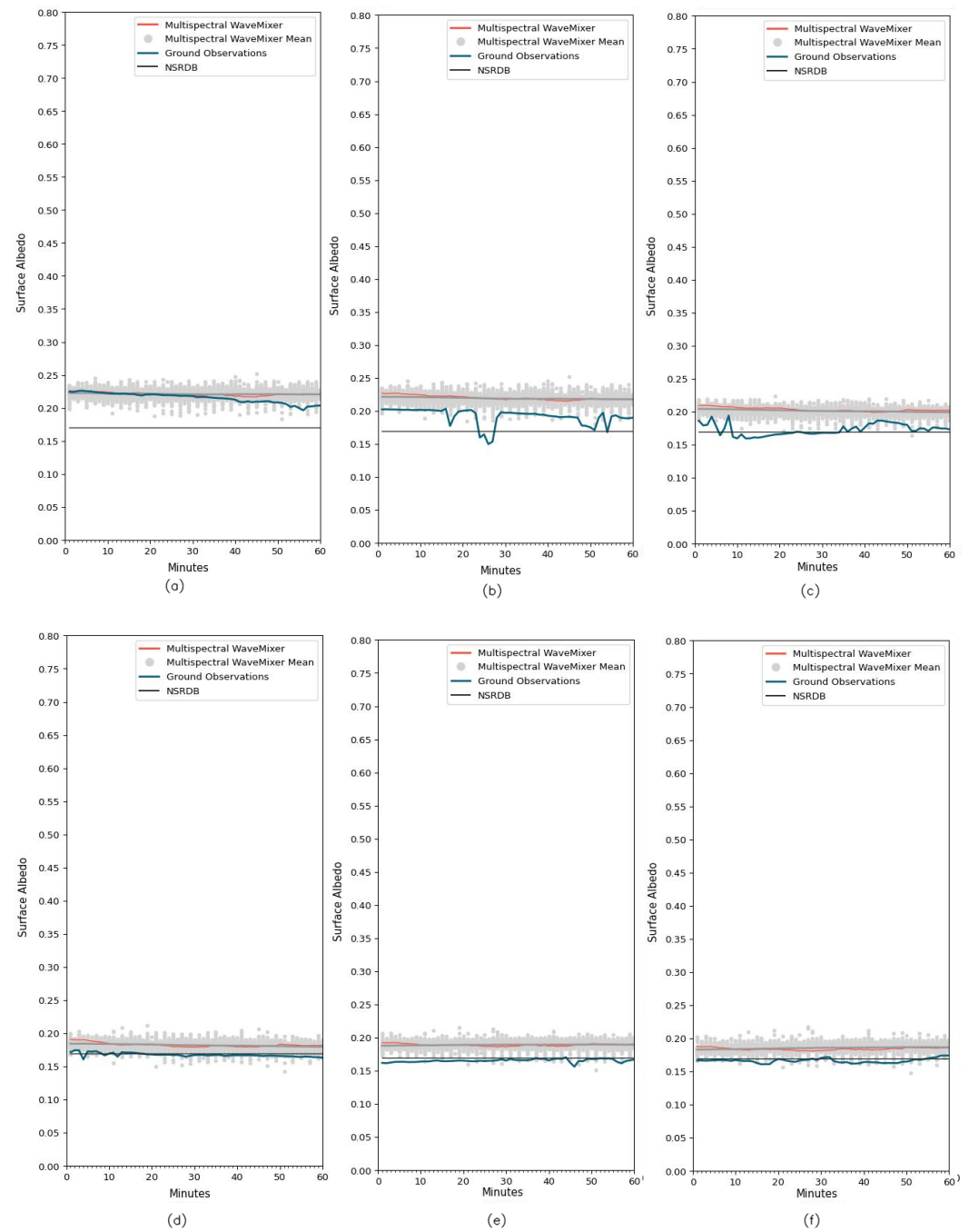


Figure A31. Plots of uncertainty surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—August 2022.

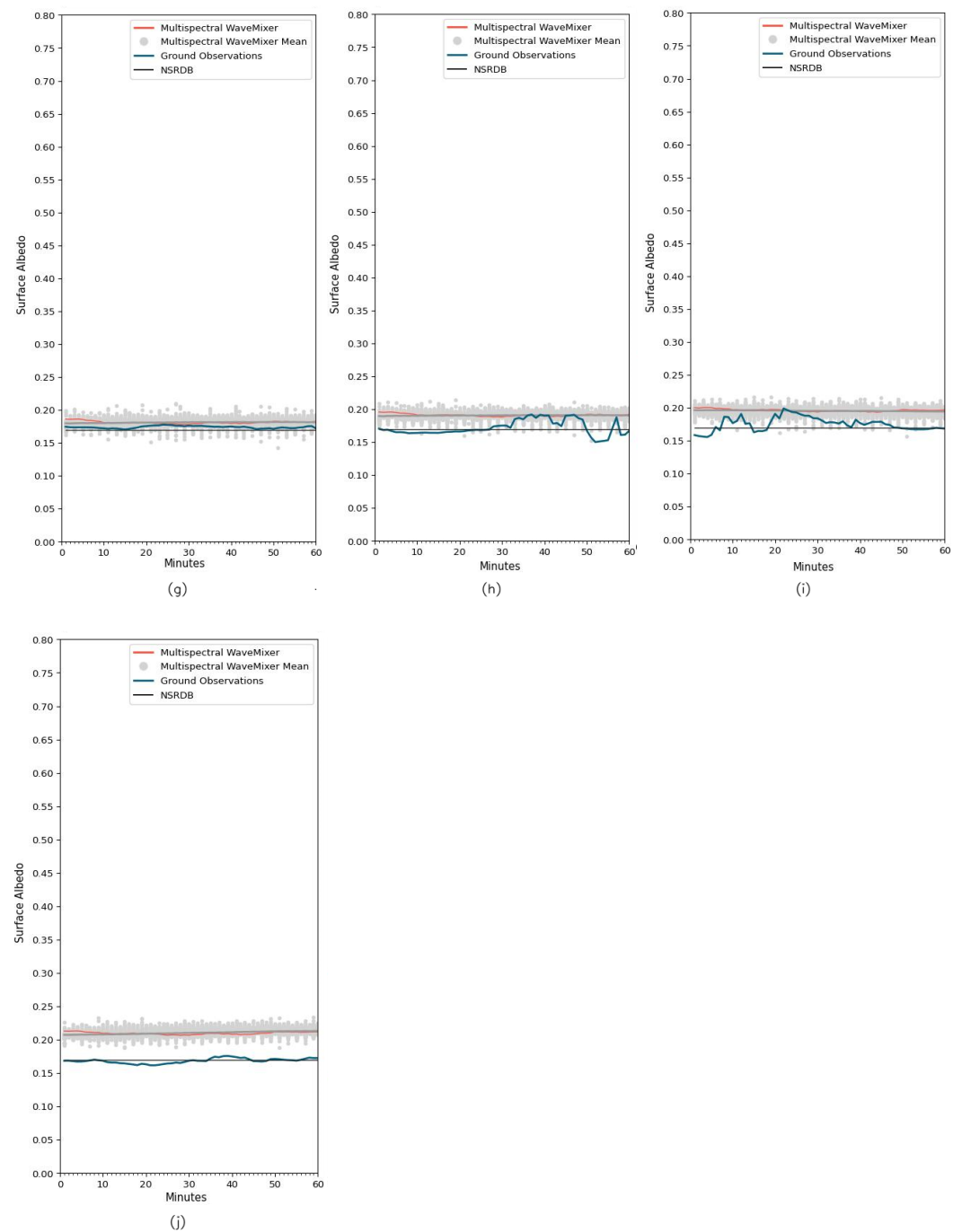


Figure A32. Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

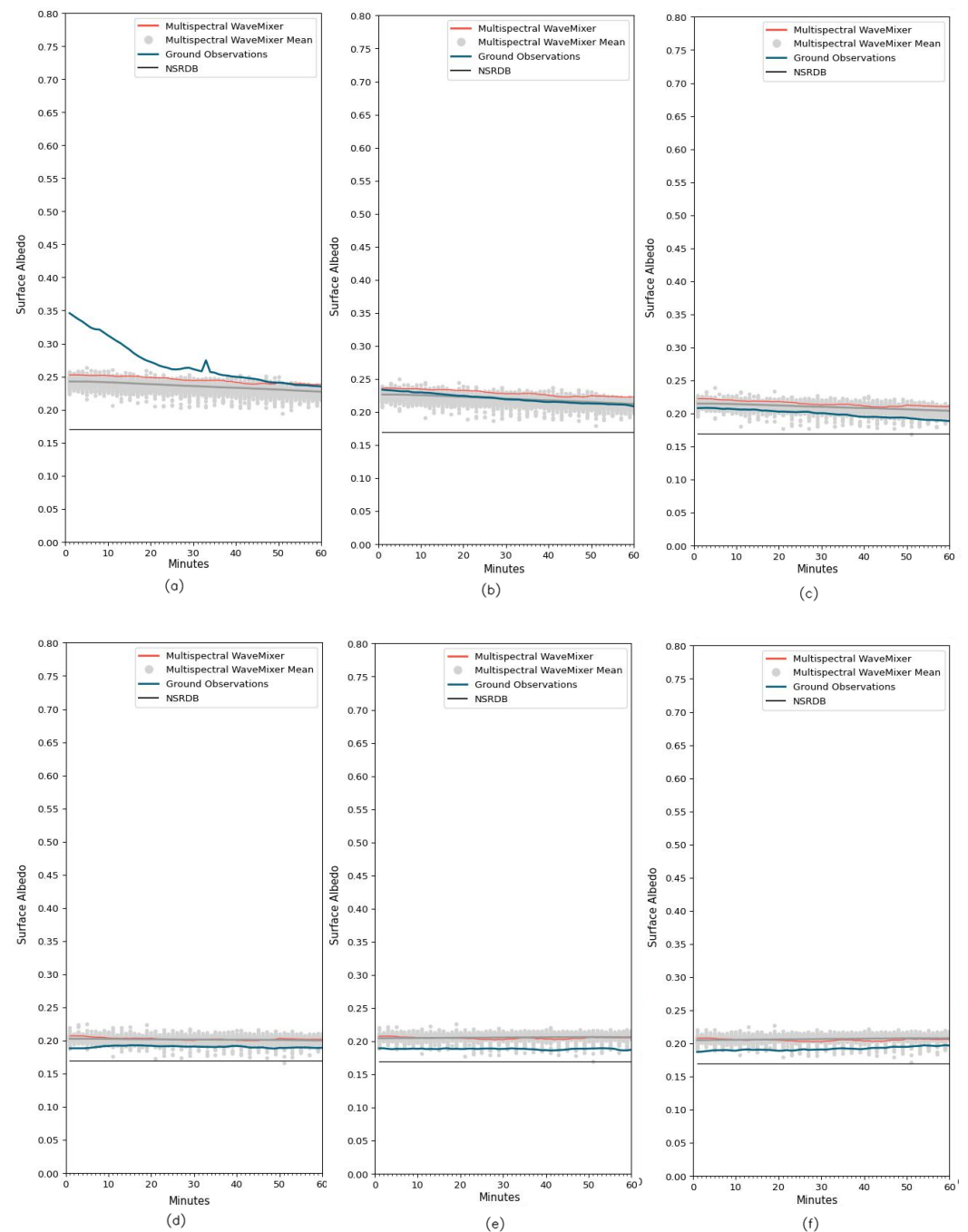


Figure A33. Plots of uncertainty surface albedo predictions for time periods (a) 7:00–8:00 a.m., (b) 8:00–9:00 a.m., (c) 9:00–10:00 a.m., (d) 10:00–11:00 a.m., (e) 11:00–12:00 p.m., and (f) 12:00–13:00 p.m. from objective model, Multispectral-Wave Mix, peer models Image to Image MLP-Mix, MLP-Mix, Standard MLP model against observations from SURFRAD monitoring stations and the National Solar Radiation Database Predictions for Golden, Colorado—October 2022.

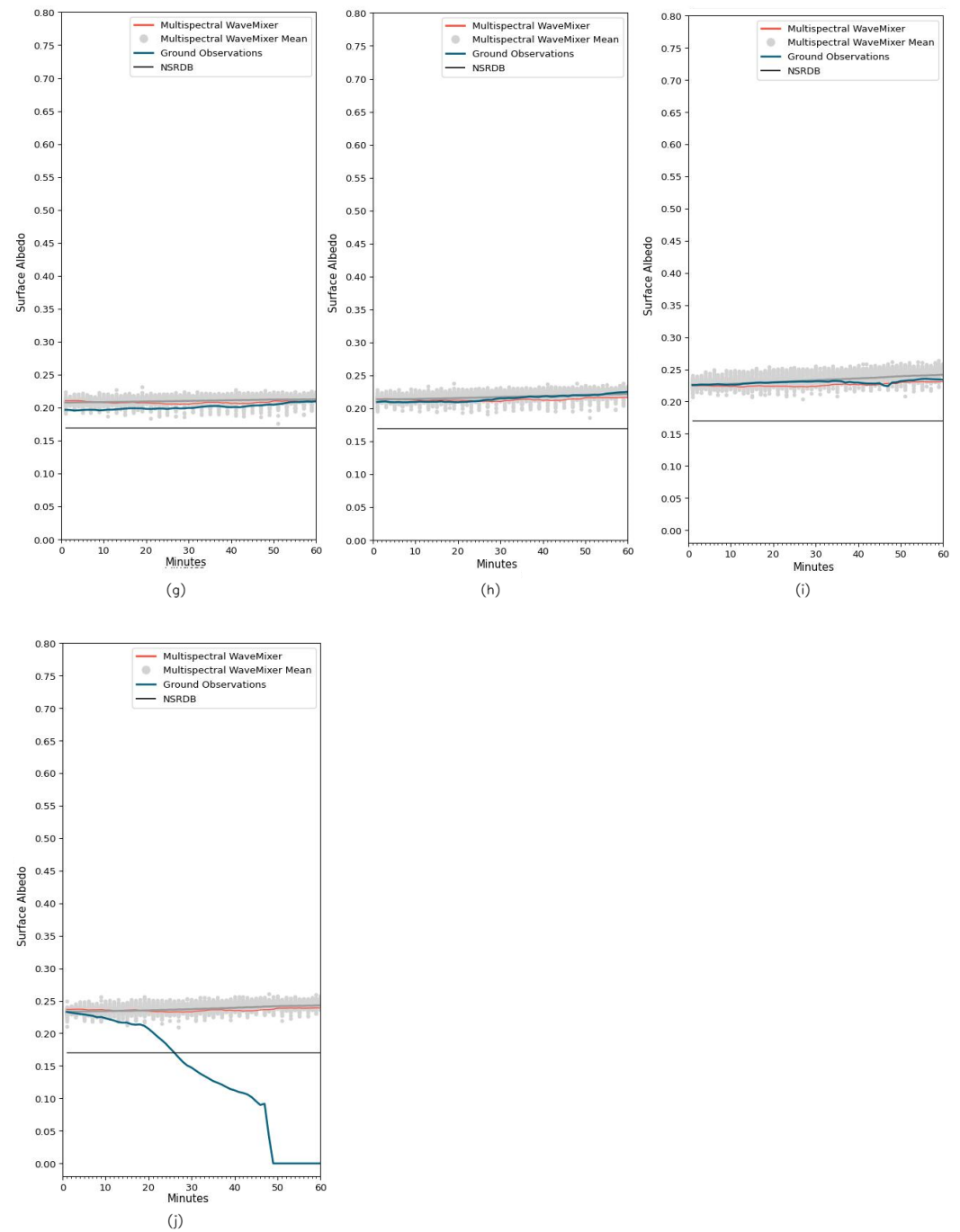


Figure A34. (Note—this figure continues from the previous figure) (g) 13:00–14:00 p.m., (h) 14:00–15:00 p.m., (i) 15:00–16:00 p.m., and (j) 16:00–17:00 p.m.

Appendix F. Predicted Temporal Shift Heatmaps

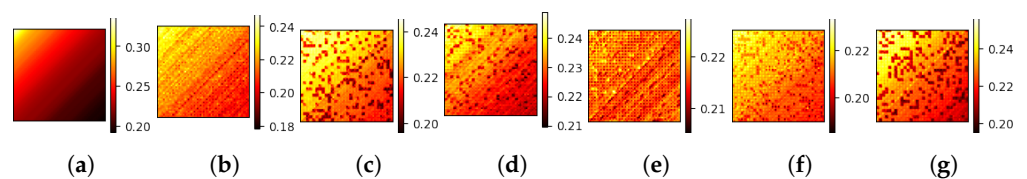


Figure A35. Predicted temporal shift heatmaps from the Multispectral-WaveMix model across spectral approximation and detail aggregation methods, minimum, maximum, sum, mean, and their combinations compared against the ground truth. (a) Ground Truth (b) Combined max, min, mean, and sum (c) without any aggregation (d) Maximum (e) Minimum (f) Mean (g) Summation.

Appendix G. Ablation Studies—Comparative Plots of Root Mean Square Error Between Wavelet Methods

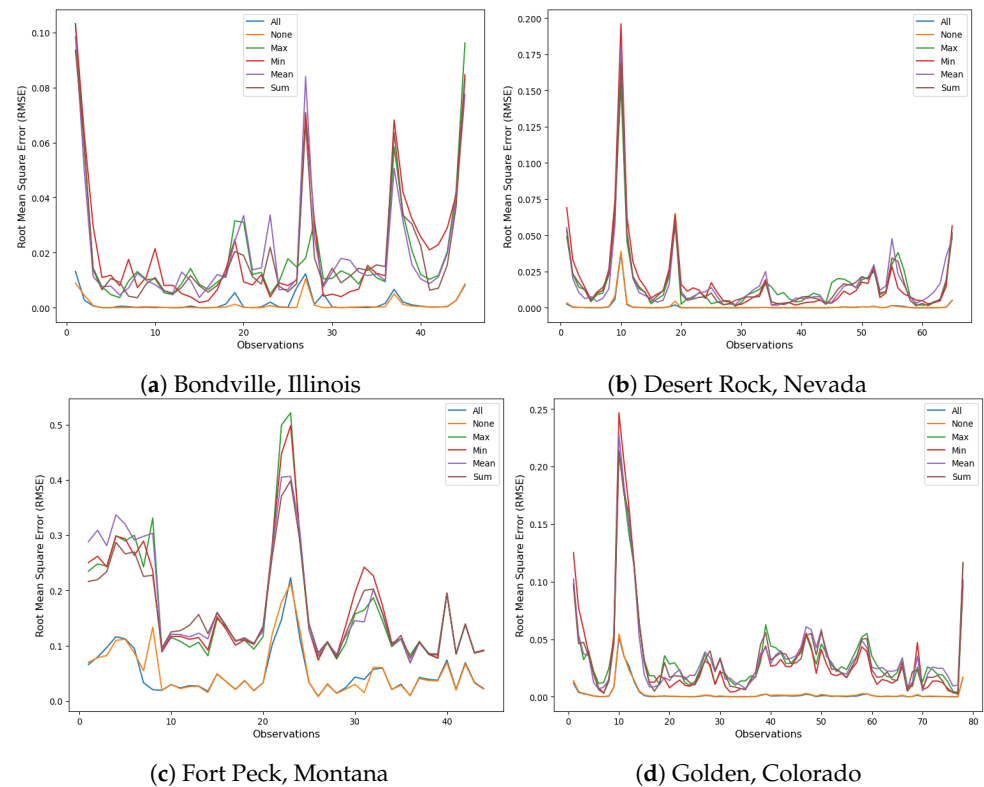


Figure A36. Comparative plots of Root Mean Square Error between wavelet methods—All (combination of minimum, maximum, sum, and mean), maximum, minimum, mean, sum, or none for: Bondville, Illinois, Desert Rock, Nevada, Fort Peck, Montana and Golden, Colorado.

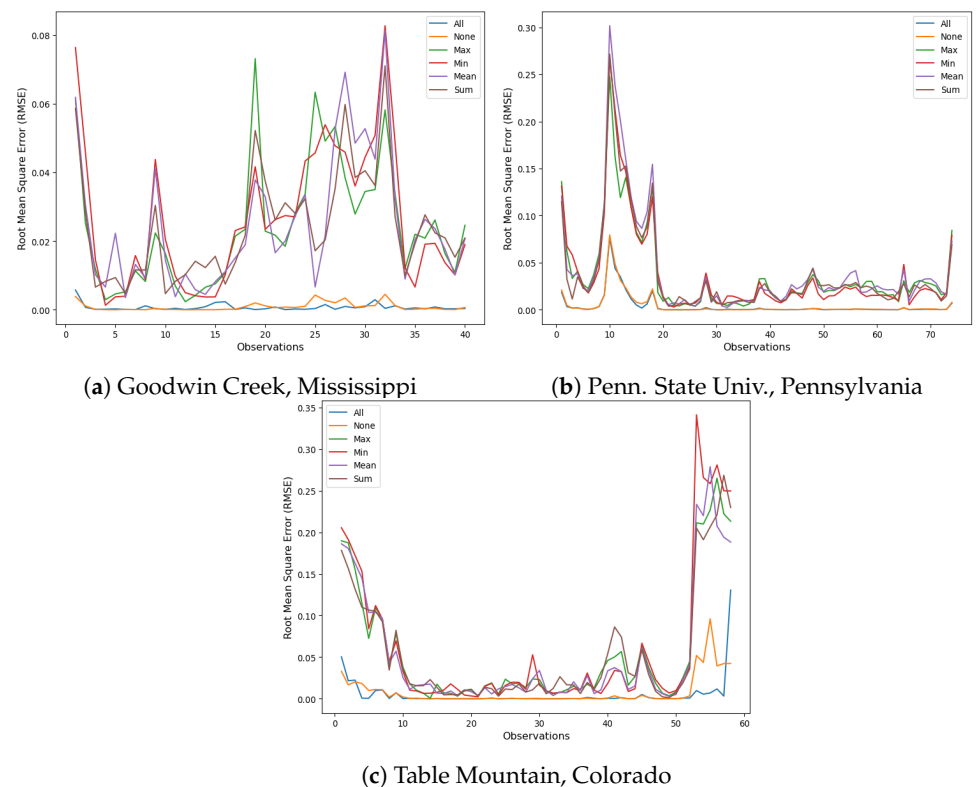


Figure A37. Caption same as previous, with results for: Goodwin Creek, Mississippi, Penn. State Univ., Pennsylvania and Table Mountain, Colorado.

References

1. Schaaf, C.B. Albedo—Albedo and Reflectance Anisotropy. 2009. Available online: <https://lpvs.gsfc.nasa.gov/PDF/TerrestrialECV/T08.pdf> (accessed on 7 December 2024)
2. Coakley, J. Reflectance and albedo, surface. *Encycl. Atmos. Sci.* **2003**, *12*, 1914–1923.
3. Myhre, G.; Shindell, D.; Bréon, F.M.; Collins, W.; Fuglestvedt, J.; Huang, J.; Koch, D.; Lamarque, J.F.; Lee, D.; Mendoza, B.; et al. Anthropogenic and natural radiative forcing. In *Climate Change 2013-The Physical Science Basis*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 659–740.
4. Trlica, A.; Hutyra, L.; Schaaf, C.; Erb, A.; Wang, J. Albedo, land cover, and daytime surface temperature variation across an urbanized landscape. *Earth's Future* **2017**, *5*, 1084–1101. [\[CrossRef\]](#)
5. Jandaghian, Z.; Akbari, H. The effect of increasing surface albedo on urban climate and air quality: A detailed study for Sacramento, Houston, and Chicago. *Climate* **2018**, *6*, 19. [\[CrossRef\]](#)
6. Riedel-Lyngskær, N.; Ribaconka, M.; Pó, M.; Thorseth, A.; Thorsteinsson, S.; Dam-Hansen, C.; Jakobsen, M.L. The effect of spectral albedo in bifacial photovoltaic performance. *Sol. Energy* **2022**, *231*, 921–935. [\[CrossRef\]](#)
7. Russell, T.C.R.; Saive, R.; Augusto, A.; Bowden, S.G.; Atwater, H.A. The Influence of Spectral Albedo on Bifacial Solar Cells: A Theoretical and Experimental Study. *IEEE J. Photovoltaics* **2017**, *7*, 1611–1618. [\[CrossRef\]](#)
8. Marsh, G. Is UK Solar Thermal about to take off? *Renew. Energy Focus* **2013**, *14*, 28–29. [\[CrossRef\]](#)
9. Wang, S.; Wilkie, O.; Lam, J.; Steeman, R.; Zhang, W.; Khoo, K.S.; Siong, S.C.; Rostan, H. Bifacial photovoltaic systems energy yield modelling. *Energy Procedia* **2015**, *77*, 428–433. [\[CrossRef\]](#)
10. IEA. *Renewable electricity—Renewables 2022—Analysis*; IEA: Paris, Italy, 2022.
11. Nøland, J.K.; Auxepaules, J.; Rousset, A.; Perney, B.; Falletti, G. Spatial energy density of large-scale electricity generation from power sources worldwide. *Sci. Rep.* **2022**, *12*, 21280. [\[CrossRef\]](#)
12. Van de Ven, D.J.; Capellan-Peréz, I.; Arto, I.; Cazcarro, I.; De Castro, C.; Patel, P.; Gonzalez-Eguino, M. The potential land requirements and related land use change emissions of solar energy. *Sci. Rep.* **2021**, *11*, 2907. [\[CrossRef\]](#)
13. Lovering, J.; Swain, M.; Blomqvist, L.; Hernandez, R.R. Land-use intensity of electricity production and tomorrow's energy landscape. *PLoS ONE* **2022**, *17*, e0270155. [\[CrossRef\]](#)
14. Krishnamurthy, A.; Serpell, O. HARVESTING THE SUN, Klienman Center for Energy Policy, University of Pennsylvania. Available online: <https://kleinmanenergy.upenn.edu/wp-content/uploads/2021/07/KCEP-Harvesting-the-Sun.pdf> (accessed on 7 October 2024).
15. Zhang, P.; Yue, C.; Li, Y.; Tang, X.; Liu, B.; Xu, M.; Wang, M.; Wang, L. Revisiting the land use conflicts between forests and solar farms through energy efficiency. *J. Clean. Prod.* **2024**, *434*, 139958. [\[CrossRef\]](#)
16. Richardson, J.; Kirk Hall, P.; Morgan, W. Land Use Conflicts Between Wind and Solar Renewable Energy and Agricultural Uses. In *WVU College of Law Research Paper No. 2022-004*; 2022. Available online: <https://ssrn.com/abstract=4042235> (accessed on 7 October 2024).
17. Graham, P.; Hayward, J.; Foster, J.; Havas, L. GenCost 2022-23: Final Report, 2023. Available online: <https://www.csiro.au/en> (accessed on 7 December 2024).
18. Cousse, J. Still in love with solar energy? Installation size, affect, and the social acceptance of renewable energy technologies. *Renew. Sustain. Energy Rev.* **2021**, *145*, 111107. [\[CrossRef\]](#)
19. Torma, G.; Aschemann-Witzel, J. Social acceptance of dual land use approaches: Stakeholders' perceptions of the drivers and barriers confronting agrivoltaics diffusion. *J. Rural. Stud.* **2023**, *97*, 610–625. [\[CrossRef\]](#)
20. Badran, G.; Dhimish, M. A Comparative Study of Bifacial versus Monofacial PV Systems at the UK Largest Solar Plant. *Clean Energy* **2024**, *8*, 248–260. [\[CrossRef\]](#)
21. Kumbaroğlu, G.S.; Çamlıbel, M.E.; Avcı, C. Techno-economic comparison of bifacial vs monofacial solar panels. *Eng. Struct. Technol.* **2021**, *13*, 7–18. [\[CrossRef\]](#)
22. Patel, M.T.; Khan, M.R.; Sun, X.; Alam, M.A. A worldwide cost-based design and optimization of tilted bifacial solar farms. *Appl. Energy* **2019**, *247*, 467–479. [\[CrossRef\]](#)
23. Rodríguez-Gallegos, C.D.; Bieri, M.; Gandhi, O.; Singh, J.P.; Reindl, T.; Panda, S. Monofacial vs bifacial Si-based PV modules: Which one is more cost-effective? *Solar Energy* **2018**, *176*, 412–438. [\[CrossRef\]](#)
24. Ziar, H.; Sönmez, F.F.; Isabella, O.; Zeman, M. A comprehensive albedo model for solar energy applications: Geometric spectral albedo. *Appl. Energy* **2019**, *255*, 113867. [\[CrossRef\]](#)
25. Stein, J.; Reise, C.; Castro, J.B.; Friesen, G.; Maugeri, G.; Urrejola, E.; Ranta, S. *Bifacial Photovoltaic Modules and Systems: Experience and Results from International Research and Pilot Applications*; Technical Report, Sandia National Lab. (SNL-NM); Fraunhofer: Albuquerque, NM, USA, 2021.
26. Alam, M.; Gul, M.S.; Muneer, T. Performance analysis and comparison between bifacial and monofacial solar photovoltaic at various ground albedo conditions. *Renew. Energy Focus* **2023**, *44*, 295–316. [\[CrossRef\]](#)

27. Blakesley, J.C.; Koutsourakis, G.; Parsons, D.E.; Mica, N.A.; Balasubramanyam, S.; Russell, M.G. Sourcing albedo data for bifacial PV systems in complex landscapes. *Solar Energy* **2023**, *266*, 112144. [\[CrossRef\]](#)
28. Grant, I.; Prata, A.; Cechet, R. The impact of the diurnal variation of albedo on the remote sensing of the daily mean albedo of grassland. *J. Appl. Meteorol.* **2000**, *39*, 231–244. [\[CrossRef\]](#)
29. Wang, D.; Liang, S.; He, T.; Yu, Y.; Schaaf, C.; Wang, Z. Estimating daily mean land surface albedo from MODIS data. *J. Geophys. Res. Atmos.* **2015**, *120*, 4825–4841. [\[CrossRef\]](#)
30. Wang, D.; Liang, S.; Zhou, Y.; He, T.; Yu, Y. A New Method for Retrieving Daily Land Surface Albedo From VIIRS Data. *IEEE Trans. Geosci. Remote. Sens.* **2017**, *55*, 1765–1775. [\[CrossRef\]](#)
31. Han, Y.; Wen, J.; You, D.; Xiao, Q.; Hao, D.; Tang, Y.; Piao, S.; Liu, G.; Liu, Q. Modeling Diurnal Variation of Land Surface Albedo Over Rugged Terrain. *IEEE Trans. Geosci. Remote. Sens.* **2024**, *62*, 1–13. [\[CrossRef\]](#)
32. He, T.; Zhang, Y.; Liang, S.; Yu, Y.; Wang, D. Developing land surface directional reflectance and albedo products from geostationary GOES-R and Himawari data: Theoretical basis, operational implementation, and validation. *Remote Sens.* **2019**, *11*, 2655. [\[CrossRef\]](#)
33. Lattanzio, A.; Schulz, J.; Matthews, J.; Okuyama, A.; Theodore, B.; Bates, J.J.; Knapp, K.R.; Kosaka, Y.; Schüller, L. Land surface albedo from geostationary satellites: A multiagency collaboration within SCOPE-CM. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 205–214. [\[CrossRef\]](#)
34. Lattanzio, A.; Schulz, J.; Roebeling, R.; Fell, F.; Bennartz, R.; Muller, J.P.; Shane, N.; Trigo, I.; Watson, G. Surface Albedo CDR From Geostationary Satellites: Validation and Further Prospects. In Proceedings of the EUMETSAT Conferences, Würzburg, Germany, 28 September–2 October 2020.
35. EUMETSAT. *Geostationary Surface Albedo (GSA) Release 2 Validation Report*; EUMETSAT: Darmstadt, Germany, v1d e-Signed ed.; 2020. Available online: https://user.eumetsat.int/s3/eup-strap-media/pdf_gsa_r2_vr_b67f4b9ca8.pdf (accessed on 7 December 2024).
36. Sengupta, M.; Xie, Y.; Lopez, A.; Habte, A.; Maclaurin, G.; Shelby, J. The national solar radiation data base (NSRDB). *Renew. Sustain. Energy Rev.* **2018**, *89*, 51–60. [\[CrossRef\]](#)
37. Wydra, K.; Vollmer, V.; Busch, C.; Prichta, S. *Agrivoltaic: Solar Radiation for Clean Energy and Sustainable Agriculture with Positive Impact on Nature*; IntechOpen: London, UK, 2023.
38. Wei, S.; Ziegler, A.D.; Qin, Y.; Wang, D.; Chen, Y.; Yan, J.; Zeng, Z. Small reduction in land surface albedo due to solar panel expansion worldwide. *Commun. Earth Environ.* **2024**, *5*, 474. [\[CrossRef\]](#)
39. Bright, R.M.; Lund, M.T. CO₂-equivalence metrics for surface albedo change based on the radiative forcing concept: a critical review. *Atmos. Chem. Phys.* **2021**, *21*, 9887–9907. [\[CrossRef\]](#)
40. Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K.; Tignor, M.; Miller, H. *IPCC Fourth Assessment Report (AR4)*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2007; Volume 374.
41. Tolstikhin, I.O.; Hounsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.
42. Mansour, Y.; Lin, K.; Heckel, R. Image-to-image MLP-mixer for image reconstruction. *arXiv* **2022**, arXiv:2202.02018.
43. Mlotshwa, T.; van Deventer, H.; Bosman, A.S. Cauchy loss function: Robustness under gaussian and Cauchy noise. In Proceedings of the Southern African Conference for Artificial Intelligence Research, Stellenbosch, South Africa, 5–9 December 2022; Springer: Cham, Switzerland, 2022; pp. 123–138.
44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 2366–2369.
46. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
48. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
49. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
50. Anwar, S.; Barnes, N. Densely residual laplacian super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1192–1204. [\[CrossRef\]](#)

51. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
52. Anwar, S.; Khan, S.; Barnes, N. A deep journey into super-resolution: A survey. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–34. [\[CrossRef\]](#)
53. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 63–79.
54. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
55. Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D.J.; Norouzi, M. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4713–4726. [\[CrossRef\]](#)
56. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
57. Shang, S.; Shan, Z.; Liu, G.; Zhang, J. Resdiff: Combining cnn and diffusion model for image super-resolution. *arXiv* **2023**, arXiv:2303.08714. [\[CrossRef\]](#)
58. Vandal, T.; Kodra, E.; Ganguly, S.; Michaelis, A.; Nemani, R.; Ganguly, A.R. Deepsd: Generating high resolution climate change projections through single image super-resolution. In Proceedings of the 23rd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1663–1672.
59. Cheng, J.; Kuang, Q.; Shen, C.; Liu, J.; Tan, X.; Liu, W. ResLap: Generating high-resolution climate prediction through image super-resolution. *IEEE Access* **2020**, *8*, 39623–39634. [\[CrossRef\]](#)
60. Stengel, K.; Glaws, A.; Hettinger, D.; King, R.N. Adversarial super-resolution of climatological wind and solar data. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16805–16815. [\[CrossRef\]](#)
61. Wang, F.; Tian, D.; Lowe, L.; Kalin, L.; Lehrter, J. Deep learning for daily precipitation and temperature downscaling. *Water Resour. Res.* **2021**, *57*, e2020WR029308. [\[CrossRef\]](#)
62. Patel, N.; Aghababaei, H.; Osei, F.B.; Stein, A.; Mahour, M. Downscaling Land Surface Temperature Using SAR Images: A Machine Learning Framework. *Preprints.org* **2023**. [\[CrossRef\]](#)
63. Mukherjee, R.; Liu, D. Downscaling MODIS spectral bands using deep learning. *Gisci. Remote. Sens.* **2021**, *58*, 1300–1315. [\[CrossRef\]](#)
64. Nomura, R.; Oki, K. Downscaling of MODIS NDVI by using a convolutional neural network-based model with higher resolution SAR data. *Remote Sens.* **2021**, *13*, 732. [\[CrossRef\]](#)
65. Wei, Y.; Yuan, Q.; Shen, H.; Zhang, L. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 1795–1799. [\[CrossRef\]](#)
66. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A deep network architecture for pan-sharpening. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5449–5457.
67. Wang, L.; Li, Q.; Peng, X.; Lv, Q. A Temporal Downscaling Model for Gridded Geophysical Data with Enhanced Residual U-Net. *Remote Sens.* **2024**, *16*, 442. [\[CrossRef\]](#)
68. Wang, L.; Li, Q.; Lv, Q.; Peng, X.; You, W. TemDeep: A Self-Supervised Framework for Temporal Downscaling of Atmospheric Fields at Arbitrary Time Resolutions. *EGUphere* **2023**, *2023*, 1–16.
69. Glawion, L.; Polz, J.; Kunstmann, H.; Fersch, B.; Chwala, C. spateGAN: spatio-temporal downscaling of rainfall fields using a cGAN approach. *Earth Space Sci.* **2023**, *10*, e2023EA002906. [\[CrossRef\]](#)
70. Maclaurin, G.; Sengupta, M.; Xie, Y.; Gilroy, N. *Development of a MODIS-Derived Surface Albedo Data Set: An Improved Model Input for Processing the NSRDB*; Technical Report; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2016.
71. Karalasingham, S.; Deo, R.C.; Casillas-Pérez, D.; Raj, N.; Salcedo-Sanz, S. Downscaling Surface Albedo to Higher Spatial Resolutions With an Image Super-Resolution Approach and PROBA-V Satellite Images. *IEEE Access* **2023**, *11*, 5558–5577. [\[CrossRef\]](#)
72. Sdraka, M.; Papoutsis, I.; Psomas, B.; Vlachos, K.; Ioannidis, K.; Karantzas, K.; Gialampoukidis, I.; Vrochidis, S. Deep learning for downscaling remote sensing images: Fusion and super-resolution. *IEEE Geosci. Remote. Sens. Mag.* **2022**, *10*, 202–255. [\[CrossRef\]](#)
73. Xu, H.; Li, J.; Yuan, H.; Liu, Q.; Fan, S.; Li, T.; Sun, X. Human Activity Recognition Based on Gramian Angular Field and Deep Convolutional Neural Network. *IEEE Access* **2020**, *8*, 199393–199405. [\[CrossRef\]](#)
74. CHEN, J.; WANG, X. Non-intrusive Load Monitoring Using Gramian Angular Field Color Encoding in Edge Computing. *Chin. J. Electron.* **2022**, *31*, 595–603. [\[CrossRef\]](#)

75. Thanaraj, K.P.; Parvathavarthini, B.; Tanik, U.J.; Rajinikanth, V.; Kadry, S.; Kamalanand, K. Implementation of Deep Neural Networks to Classify EEG Signals using Gramian Angular Summation Field for Epilepsy Diagnosis. *arXiv* **2020**, arXiv:cs.CV/2003.04534.
76. Bragin, A.D.; Spitsyn, V.G. Electroencephalogram analysis based on Gramian angular field transformation. In Proceedings of the CEUR Workshop Proceedings, Conference: 29th International Conference on Computer Graphics, Image Processing and Computer Vision, Visualization Systems and the Virtual Environment GraphiCon 2019, Tomsk Polytechnic University, Tomsk Oblast, Russia, 23–26 September 2019; Volume 24852019, pp. 273–275.
77. Zhang, G.; Si, Y.; Wang, D.; Yang, W.; Sun, Y. Automated Detection of Myocardial Infarction Using a Gramian Angular Field and Principal Component Analysis Network. *IEEE Access* **2019**, *7*, 171570–171583. [\[CrossRef\]](#)
78. Lee, H.; Yang, K.; Kim, N.; Ahn, C.R. Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. *Autom. Constr.* **2020**, *120*, 103390. [\[CrossRef\]](#)
79. Hong, Y.Y.; Martinez, J.J.F.; Fajardo, A.C. Day-Ahead Solar Irradiation Forecasting Utilizing Gramian Angular Field and Convolutional Long Short-Term Memory. *IEEE Access* **2020**, *8*, 18741–18753. [\[CrossRef\]](#)
80. Chen, S.A.; Li, C.L.; Yoder, N.; Arik, S.O.; Pfister, T. TSMixer: An All-MLP Architecture for Time Series Forecasting. *arXiv* **2023**, arXiv:cs.LG/2303.06053
81. Lim, B.; Arik, S.Ö.; Loeff, N.; Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [\[CrossRef\]](#)
82. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Proceedings of the International Conference on Machine Learning, PMLR, Bellevue, Baltimore, MD, USA, 17–23 July 2022; pp. 27268–27286.
83. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
84. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11106–11115.
85. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv* **2022**, arXiv:2211.14730.
86. Diederik, P.K. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
87. Nilsson, J.; Akenine-Möller, T. Understanding ssim. *arXiv* **2020**, arXiv:2006.13846.
88. Chen, X.; Qin, Y.; Xu, W.; Bur, A.M.; Zhong, C.; Wang, G. Explicitly Increasing Input Information Density for Vision Transformers on Small Datasets. *arXiv* **2022**, arXiv:2210.14319.
89. Hsu, H.; Asoodeh, S.; Calmon, F. Obfuscation via information density estimation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, Online, 26–28 August 2020; pp. 906–917.
90. Kamerman, G. Image Information Density and Lidar Object Interpretability. In *Bulletin of NTUU “KPI”. INSTRUMENT BUILDING Series.—Issue 44*; Kiev Polytechnic Institute: Kiev, Ukraine, 2012; pp. 44–48. Available online: <https://ela.kpi.ua/server/api/core/bitstreams/590756e8-895f-4548-abdc-0938e3401990/content> (accessed on 7 October 2024).
91. Shocher, A.; Cohen, N.; Irani, M. “zero-shot” super-resolution using deep internal learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3118–3126.
92. Zhang, Q.; Zhou, J.; Zhu, L.; Sun, W.; Xiao, C.; Zheng, W.S. Unsupervised intrinsic image decomposition using internal self-similarity cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9669–9686. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Zuckerman, L.P.; Naor, E.; Pisha, G.; Bagon, S.; Irani, M. Across scales and across dimensions: Temporal super-resolution using deep internal learning. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part VII 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 52–68.
94. Wang, Z.; Oates, T. Imaging Time-Series to Improve Classification and Imputation. *arXiv* **2015**, arXiv:cs.LG/1506.00327
95. Keogh, E.J.; Pazzani, M.J. Scaling up dynamic time warping for datamining applications. In Proceedings of the Sixth ACM sigkdd International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, 20–23 August 2000; pp. 285–289.
96. Ren, H.; Liu, M.; Li, Z.; Pedrycz, W. A piecewise aggregate pattern representation approach for anomaly detection in time series. *Knowl.-Based Syst.* **2017**, *135*, 29–39. [\[CrossRef\]](#)
97. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
98. Walker, J.S. *A Primer on Wavelets and Their Scientific Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2008.
99. Trockman, A.; Kolter, J.Z. Patches are all you need? *arXiv* **2022**, arXiv:2201.09792.
100. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–10.

101. Elfving, S.; Uchibe, E.; Doya, K. Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning. *arXiv* **2017**, arXiv:cs.LG/1702.03118. [\[CrossRef\]](#) [\[arXiv:cs.LG/1702.03118\]](#).
102. Liang, S. Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote. Sens. Environ.* **2001**, *76*, 213–238. [\[CrossRef\]](#)
103. Salleh, S.A.; Chan, A. Land Surface Albedo Determination: Remote Sensing and Statistical Validation. In Proceedings of the FIG Congress 2014, Engaging the Challenges–Enhancing the Relevance, Kuala Lumpur, Malaysia, 16–21 June 2014; pp. 1–13.
104. Woo, G.; Liu, C.; Sahoo, D.; Kumar, A.; Hoi, S. Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv* **2022**, arXiv:2202.01381.
105. Yu, S.; Hu, Z.; Subramaniam, A.; Hannah, W.; Peng, L.; Lin, J.; Bhouiri, M.A.; Gupta, R.; Lütjens, B.; Will, J.C.; et al. ClimSim-Online: A Large Multi-scale Dataset and Framework for Hybrid ML-physics Climate Emulation. *arXiv* **2024**, arXiv:cs.LG/2306.08754. [\[arXiv:cs.LG/2306.08754\]](#).
106. Shine, K. Parametrization of the shortwave flux over high albedo surfaces as a function of cloud thickness and surface albedo. *Q. J. R. Meteorol. Soc.* **1984**, *110*, 747–764. [\[CrossRef\]](#)
107. He, M.; Hu, Y.; Chen, N.; Wang, D.; Huang, J.; Stamnes, K. High cloud coverage over melted areas dominates the impact of clouds on the albedo feedback in the Arctic. *Sci. Rep.* **2019**, *9*, 9529. [\[CrossRef\]](#)
108. Chai, T.; Draxler, R. Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [\[CrossRef\]](#)
109. Ye, X. calflops: A FLOPs and Params Calculate Tool for Neural Networks in Pytorch Framework. 2023. Available online: <https://github.com/MrYxJ/calculate-flops.pytorch> (accessed on 1 December 2024).
110. Rutan, D.; Rose, F.; Roman, M.; Manalo-Smith, N.; Schaaf, C.; Charlock, T. Development and assessment of broadband surface albedo from Clouds and the Earth’s Radiant Energy System Clouds and Radiation Swath data product. *J. Geophys. Res. Atmos.* **2009**, *114*, D08125. [\[CrossRef\]](#)
111. Jian, B.; Li, J.; Wang, G.; He, Y.; Han, Y.; Zhang, M.; Huang, J. The impacts of atmospheric and surface parameters on long-term variations in the planetary albedo. *J. Clim.* **2018**, *31*, 8705–8718. [\[CrossRef\]](#)
112. Schlundt, C.; Kokhanovsky, A.; von Hoyningen-Huene, W.; Dinter, T.; Istomina, L.; Burrows, J. Synergetic cloud fraction determination for SCIAMACHY using MERIS. *Atmos. Meas. Tech.* **2011**, *4*, 319–337. [\[CrossRef\]](#)
113. Lai, H.P.; Tran, T.T.; Pham, V.T. Axial attention mlp-mixer: A new architecture for image segmentation. In Proceedings of the 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), Nha Trang, Vietnam, 27–29 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 381–386.
114. Le, H.M.Q.; Le, T.K.; Pham, V.T.; Tran, T.T. Amg-mixer: A multi-axis attention mlp-mixer architecture for biomedical image segmentation. In Proceedings of the Conference on Information Technology and Its Applications, Da Nang City, Vietnam, 28–29 July 2023; Springer: Cham, Switzerland, 2023; pp. 169–180.
115. He, Y.; Lu, Z.; Wang, J.; Ying, S.; Shi, J. A Self-Supervised Learning Based Channel Attention MLP-Mixer Network for Motor Imagery Decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2406–2417. [\[CrossRef\]](#)
116. Tang, C.; Zhao, Y.; Wang, G.; Luo, C.; Xie, W.; Zeng, W. Sparse MLP for image recognition: Is self-attention really necessary? In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 2344–2351.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.