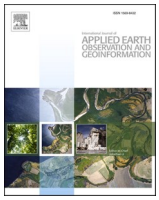
International Journal of Applied Earth Observation and Geoinformation 138 (2025) 104468

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Improved hourly all-sky land surface temperature estimation: Incorporating the temporal variability of cloud-radiation interactions

Dukwon Bae [a,1], Dongjin Cho [a,b,1], Jungho Im [a,c,d,*], Cheolhee Yoo [e], Yeonsu Lee [a], Siwoo Lee [a]

[a] Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea
[b] Environmental Planning Institute, Seoul National University, Seoul, South Korea
[c] Graduate School of Artificial Intelligence, UNIST, Ulsan, South Korea
[d] Graduate School of Carbon Neutrality, UNIST, Ulsan, South Korea
[e] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Land surface temperature
GK2A
All-sky
Temporal variation of cloud-radiation effects
Machine learning

ABSTRACT

Land surface temperature (LST) is an indispensable factor for comprehending of surface equilibrium state on the Earth. In particular, satellites can continuously provide LST data and support the large-scale monitoring of LST with a high temporal resolution; however, satellite data may be easily contaminated by clouds. Previous satellite-based all-sky LST reconstruction approaches have inherent limitations, such as low temporal resolution and insufficient consideration of cloud effects. Therefore, this study aims to propose a novel methodology for all-sky 2-km hourly LST reconstruction from GEO-KOMPSAT-2A (GK2A) using machine learning and timely weighted accumulated radiation to reflect the temporal variation of cloud effects. The light gradient boosting machine approach used the European Center for Medium-Range Weather Forecasts Reanalysis-Land variables (i.e., LST, 2 m air temperature, evaporation, and wind), GK2A products (i.e., short and longwave radiation, and binary cloud cover), and auxiliary variables including geographic variables as independent variables. The GK2A LST and in situ measurements were used as dependent variables. The proposed model showed robust spatial agreement with GK2A LST under clear-sky conditions when conducting five-fold spatial cross-validation, with coefficient of determination (R^2) values of 0.97–0.99. In the leave one station-out cross-validation using 36 in situ data under all-sky conditions, the proposed model showed high performance with R^2 values of 0.86–0.97, root mean square error values of 1.42–2.60 °C, and bias values of −0.49–0.23 °C. In a comparison of the proposed model with two scenarios and previous research investigating the effect of accumulated radiation, we demonstrated that the use of accumulated radiation was effective in reconstructing cloudy-sky LST, particularly during the daytime, as evident from the variable analysis conducted through Shapley additive explanations. Using the proposed model, we successfully reconstructed a spatiotemporally seamless LST, which can serve as a fundamental dataset for hourly heat-related research, such as hourly heat flow estimation and urban heat island analysis.

1. Introduction

Land surface temperature (LST) plays a vital role in controlling the exchange of heat between the Earth's surface and atmosphere by governing the vertical flux of terrestrial radiation (Chen et al., 2023; Peng et al., 2020). LST is a fundamental variable used in various climate and environmental studies in fields such as agriculture, meteorology, climatology, and ecology (Shen et al., 2016; Zeng et al., 2018). Moreover, spatial analysis with fine temporal resolution utilizing the diurnal temperature cycle is highly sought after within the domains of thermodynamics and environmental research (Kumar et al., 2006; Zhou et al., 2022). The findings of these analyses have direct implications for human environmental assessments, including energy policies, studies on human mortality, and disease analyses (Cao et al., 2022; Guo et al., 2022; Wang et al., 2022). Therefore, it is crucial to monitor LST across high spatiotemporal domains.

LST can be measured at ground stations as point data with high temporal intervals; however, it is difficult to monitor large areas due to

* Corresponding author.
E-mail address: ersgis@unist.ac.kr (J. Im).
[1] The first two authors equally contributed to the paper.

https://doi.org/10.1016/j.jag.2025.104468
Received 5 June 2024; Received in revised form 24 February 2025; Accepted 5 March 2025
Available online 9 March 2025
1569-8432/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

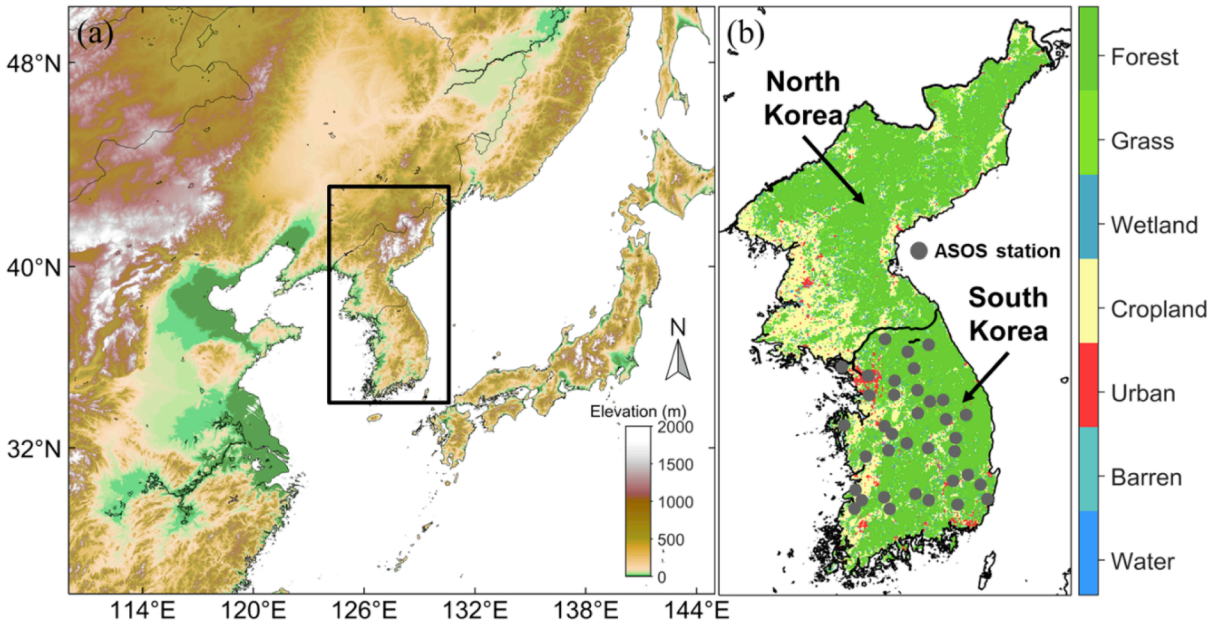


Fig. 1. (a) Geographic overview of the Korean Peninsula in East Asia by elevation derived from the Shuttle Radar Topography Mission global 30 m digital elevation model (DEM). The black box highlights the area that is shown in detail in panel (b). (b) Locations of automated synoptic observing system (ASOS) station (indicated in gray dot). The background map represents various land cover types such as forest, grass, cropland, urban, barren, and water, derived from Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data (MCD12Q1).

an insufficient number of stations (Cho et al., 2022). With advances in satellite technology, thermal infrared (TIR) sensors have made it possible to monitor areas where ground stations are scarce or sparsely distributed. TIR-based LST observations are conducted using both polar-orbiting and geostationary satellites; notably, the use of geostationary satellites with higher temporal resolution (e.g., hourly) has recently increased, enabling more consistent, continuous monitoring over extensive regions (Jia et al., 2024). TIR-based LST, on the other hand, cannot be used when there are clouds because clouds block thermal information from the land surface to passive optical sensors (Li et al., 2023; Zeng et al., 2018).

A wide range of LST reconstruction approaches have been proposed to produce spatially seamless LST, broadly categorized into spatiotemporal interpolation, surface energy balance (SEB), and data fusion—combining TIR with passive microwave (PMW) or integrating TIR with numerical models (Mo et al., 2021). The first approach, spatiotemporal interpolation, estimates cloudy-sky LST using neighboring clear-sky pixels. However, this method reconstructs a hypothetical clear-sky LST without considering cloud effects, which can result in significant deviations during cloudy conditions due to the cooling or warming influence of clouds on LST (Zeng et al., 2018). The second approach, based on the SEB equation, estimates cloudy-sky LST by calculating differences in SEB between clear-sky and cloudy-sky pixels to account for cloud impacts. However, it requires difficult-to-obtain physical parameters, such as air temperature and wind speed, and struggles with nighttime data due to low shortwave radiation inputs, making its application challenging under certain conditions (Li et al., 2023). The third approach combines TIR and PMW products to estimate LST in cloudy-sky conditions, as PMW sensors can partially penetrate clouds. This integration can make LST more accurate under cloudy conditions, but PMW satellites' lower temporal resolution (e.g., more than daily) makes them less compatible with high-frequency geostationary satellite data. In addition, PMW sensors typically measure subsurface rather than skin temperatures, resulting in additional uncertainty (Huang et al., 2019; Song et al., 2024). Finally, the fourth approach incorporates TIR-LST with numerical or reanalysis models. Reanalysis datasets, such as the European Centre for Medium-Range Weather Forecasts Reanalysis v5 Land (ERA5), provide high temporal resolution but have a coarser spatial resolution and use smoothing techniques to minimize grid noise (Ding et al., 2022). Nevertheless, these datasets hold significant potential for reconstructing the all-sky LST of geostationary satellites due to their frequent data availability, and recent studies have increasingly focused on this integration to achieve spatially seamless LST with high temporal resolution, addressing the limitations of standalone satellite data (Ding et al., 2022; Ding et al., 2023; Xu et al., 2023).

Machine learning techniques, particularly tree-based models (e.g., random forest (RF) and light gradient boosting machine (LGBM)) actively employed to perform LST reconstruction (Cho et al., 2022; Yoo et al., 2020). Machine learning outperforms traditional regression models in terms of capturing complex and nonlinear relationships (Wang et al., 2024). Moreover, recent machine learning-based LST reconstruction studies have emphasized the importance of considering cloud effects (Ermida et al., 2019). Yoo et al. (2020) and Cho et al. (2022) addressed instantaneous cloud effects using cloudy-sky in situ observations or binary cloud flags. However, when using these methods, it is difficult to adequately account for the temporal variability of cloud effects on LST. He et al. (2019) revealed that LST is not solely a result of instantaneous solar radiation but rather a cumulative outcome of solar radiation over time. Zhao and Duan (2020) studied the effects of clouds from sunrise to the target time by using accumulative radiation to reconstruct the all-sky Moderate Resolution Imaging Spectroradiometer (MODIS) LST. However, they only conducted research during the daytime. To our knowledge, no studies have explored the temporal variations in cloud radiation effects on daytime and nighttime hourly LST using machine learning approaches.

Therefore, this study aimed to reconstruct a reliable all-sky hourly Geo-Kompsat-2A (GK2A) LST based on machine learning by considering the temporal variation of cloud effects. We adopted accumulated downward shortwave and longwave radiation, along with the GK2A and ERA5 meteorological variables and auxiliary variables as inputs for the LGBM model. The proposed model used clear-sky GK2A LST and all-sky in situ measurements as targets. The main objectives of this study were to (1) develop a machine learning-based all-sky 2-km hourly GK2A LST

Table 1
Geoinformation of automated synoptic observing system stations used in this study, including station number, ID, longitude, latitude, Moderate Resolution Imaging Spectroradiometer (MODIS) 500 m land cover (MCD12Q1), resampled 2 km land cover, the calculated homogeneity (%), and elevation from Shuttle Radar Topography Mission global 30 m digital elevation model. Homogeneity is defined as the proportion of the most frequent land cover type among the 16 MODIS 500 m pixels within each 2 km grid.

Station number	Station ID	Longitude (°)	Latitude (°)	MODIS land cover (500 m)	Resampled land cover (2 km)	Homogeneity (%)	Elevation (m)
1	93	127.30	38.15	Urban	Urban	68.75	157.68
2	95	128.17	38.06	Cropland	Cropland	56.25	331.35
3	108	127.75	37.95	Urban	Urban	100	114.58
4	114	126.45	37.71	Urban	Urban	93.75	93.72
5	119	127.88	37.68	Urban	Urban	100	196.13
6	121	126.97	37.57	Grass	Grass	43.75	86.32
7	127	127.49	37.49	Cropland	Cropland	43.75	30.95
8	129	127.95	37.34	Cropland	Cropland	100	149.55
9	131	127.48	37.26	Urban	Urban	100	79.27
10	133	126.98	37.26	Urban	Urban	81.25	36.50
11	136	128.46	37.18	Urban	Urban	87.5	240.29
12	137	128.19	37.16	Urban	Urban	81.25	325.00
13	143	127.95	36.97	Urban	Urban	93.75	216.37
14	152	128.91	36.87	Urban	Forest	37.5	368.73
15	201	128.52	36.78	Cropland	Cropland	75	237.78
16	202	126.49	36.76	Urban	Urban	68.75	29.27
17	203	127.29	36.64	Urban	Cropland	81.25	73.69
18	211	127.44	36.57	Forest	Forest	56.25	66.32
19	212	128.71	36.49	Urban	Urban	81.25	147.13
20	221	127.73	36.41	Cropland	Cropland	68.75	190.50
21	226	128.16	36.37	Cropland	Forest	43.75	71.78
22	232	127.37	36.36	Cropland	Cropland	68.75	51.80
23	236	128.69	36.27	Cropland	Cropland	56.25	107.88
24	243	126.92	35.98	Cropland	Cropland	100	28.45
25	244	128.95	35.88	Cropland	Cropland	56.25	99.92
26	245	128.65	35.82	Forest	Cropland	50	68.17
27	247	129.20	35.73	Urban	Urban	68.75	40.69
28	251	126.72	35.73	Cropland	Cropland	43.75	15.20
29	271	127.91	35.67	Grass	Forest	62.5	205.55
30	272	127.29	35.61	Cropland	Grass	37.5	266.90
31	278	129.33	35.58	Grass	Urban	50	40.61
32	281	128.17	35.57	Urban	Urban	81.25	73.30
33	283	126.84	35.56	Cropland	Cropland	100	63.80
34	284	128.74	35.49	Cropland	Cropland	93.75	8.13
35	285	126.70	35.43	Forest	Urban	50	47.48
36	288	127.40	35.42	Urban	Urban	100	119.86

reconstruction model using temporally accumulated radiation information, (2) systematically evaluate the model performance under clear and cloudy skies, and (3) investigate the effects of accumulated radiation information in reconstructing all-sky LST.

2. Study area and data

2.1. Study area

The study area was the Korean Peninsula (33°–38°N, 125°–131°E), which has four distinct seasons and dynamic atmospheric conditions (Fig. 1). In summer, the region is heavily influenced by the East Asian monsoon, resulting in high humidity, frequent rainfall, and high cloud cover. In contrast, during the winter, it is strongly affected by the Siberian high-pressure system, which causes northwesterly winds and heavy snowfall. The western part of the Korean Peninsula primarily consists of gentle flatlands, whereas the eastern part is mountainous. In addition, the land cover classes vary, including forest (62.09 %), croplands (20.44 %), grass (7.94 %), wetlands (4.93 %), urban (2.99 %), water (0.28 %), barren (0.15 %), and other regions (1.18 %) (Fig. 1b).

The research period was from March 1 to November 30 in both 2020 and 2021, except for winter. We excluded the winter season from the study because the LST measurement sensors at ground stations are frequently hindered by heavy snow.

2.2. GK2A satellite data

The GK2A, launched on December 4, 2018, is a geostationary meteorological satellite operated by the National Meteorological Satellite Center to observe atmospheric and surface conditions for weather monitoring and forecasting. The GK2A is equipped with an Advanced Meteorological Imager and features 16 channels for measuring various atmospheric properties. The satellite provides products with spatial resolutions ranging from 0.5 to 2 km at 2 min intervals over the Korean peninsula and full-disk coverage at 10 min intervals. GK2A retrieves 10-min interval LST with a 2 km resolution above the study area through the split-window algorithm using channels 13 (10.3539 μm) and 15 (12.3651 μm) (Choi and Suh, 2020). For the seamless temporal incorporation of all datasets, we utilized only hourly data within the 10-min GK2A LST dataset. Compared to the in situ measurements from the Baseline Surface Radiation Network (BSRN), the root mean square error (RMSE) of the GK2A LST was less than 2 K (Choi and Suh, 2020). We utilized LST, binary cloud cover, downward shortwave (SRD), and longwave radiation (LRD) products from the hourly GK2A dataset in this study.

2.3. ERA5 reanalysis data

The ERA5 reanalysis data provided global meteorological variables, featuring a 10 km spatial resolution with 1-hour interval, combining the results of a data assimilation model using various satellite data, models, and in situ data. The ERA5 skin temperature (LST) product was utilized as a key independent variable for reconstruction because it delivers a gap-free LST regardless of cloud cover. In addition, the 2 m atmospheric temperature (T2m), evaporation (EVP), 10 m zonal (U10) and meridional (V10) wind were used because they are highly associated with LST

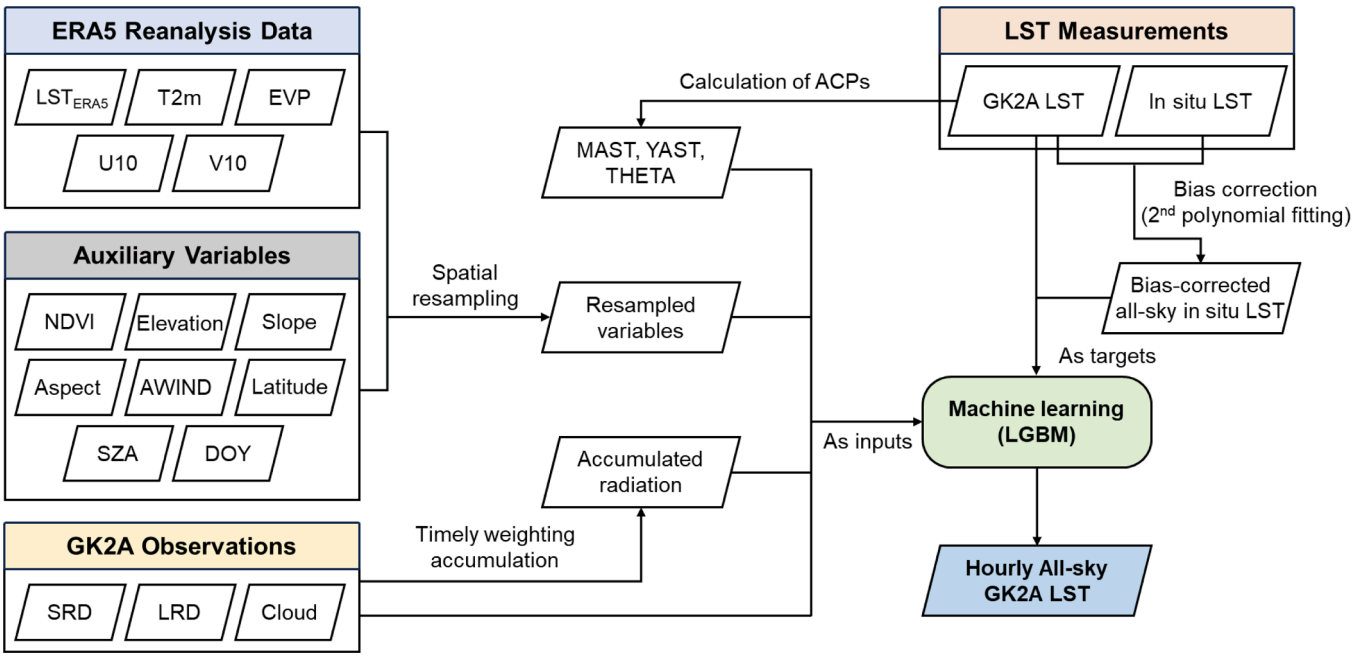


Fig. 2. Flowchart of the proposed reconstruction model for all-sky hourly 2-km hourly Geo-Kompsat-2A (GK2A) land surface temperature (LST).

(Guo et al., 2020). To align with the intrinsic 2 km spatial grid of the GK2A LST, we applied cubic interpolation to resample the ERA5 data accordingly.

2.4. Auxiliary variables

The normalized difference vegetation index (NDVI), elevation, slope, aspect, annual mean wind speed (AWIND), latitude, solar zenith angle (SZA), and day of year (DOY) were used to reflect a geographical information and temporal periodicity of LST (Bird et al., 2022; Liu et al., 2021). Derived from 1 km MODIS Terra 16-day vegetation index (MOD13A2; https://modis.gsfc.nasa.gov/), NDVI represents the greenness of the surface and effectively helps estimate the spatial difference of LST (Shafizadeh-Moghadam et al., 2020). As missing NDVI values in the 16-d composite images could exist in the summer season due to the high cloud cover rate, the NDVI data were converted to monthly NDVI values and used in this study. The 500 m MODIS land cover (MCD12Q1) dataset, derived from Terra and Aqua satellite observations, was utilized as the reference for land cover classification. The original land cover types were aggregated into four simplified categories: forest, grass, cropland, and urban. Then, the majority land cover type within a 4 × 4 pixel window was assigned as the representative land cover for the 2 km grid.

In addition, Elevation was obtained from the Shuttle Radar Topography Mission global 30 m digital elevation model (DEM) (https://earthexplorer.usgs.gov/). Slope and aspect were computed from "slope" and "aspect" tools of the Spatial Analyst Toolbox in ArcGIS, respectively. The AWIND provided by the Global Wind Atlas was used to reflect the fluctuating climate value of Earth's surface heat (https://globalwindatlas.en). The latitude, associated with solar radiation, was calculated according to the GK2A grid. However, longitude was not considered as an input to avoid being an all-sky LST reconstruction model that relies heavily on coordinate information (Sadeghi et al., 2020). Finally, ZA and DOY were used to reflect the temporal fluctuations. The DOY was scaled within −1 to 1 by applying a cosine transformation.

2.5. In situ LST data

In situ LST observed by BSRN is suitable for matching with satellite-based LST, but they were unavailable in the Korean Peninsula. Instead, we used in situ data observed at automated synoptic observing system (ASOS) stations, which directly measure the LST using a platinum-resistant temperature sensor (PT-100) shielded by a thin layer of soil to avoid the influence of direct sunlight, with a measurement range of −40 °C to 100 °C, a resolution of 0.1 °C, and a measurement error within ±0.3 °C. Following the Standards for Surface Meteorological Observations set by the Korea Meteorological Administration (KMA), ASOS stations are located on flat, open terrain to minimize interference from surrounding structures and maintain stable airflow, providing reliable temperature measurements unaffected by surrounding terrain and canopy (Fig. S1). The KMA operates 103 ASOS stations in South Korea. To reduce the inconsistency between satellite-based LST and in situ measurements, we filtered ASOS stations that had no location changes during the research period and were highly consistent with GK2A after implementing the bias correction mentioned in Section 3.1. Ultimately, 36 ASOS stations were selected for this study (Table 1).

3. Methodology

Fig. 2 summarizes the proposed all-sky GK2A LST reconstruction process. GK2A data (LST, clouds, SRD, and LRD), ERA5 data (LST, T2m, EVP, U10, and V10), and eight auxiliary variables (NDVI, elevation, slope, aspect, AWIND, latitude, ZA, and DOY) were used as independent data. We also used annual cycle parameters (ACPs), such as mean annual surface temperature (MAST), the yearly amplitude surface temperature (YAST), and the phase shift of LST in relation to the spring equinox (theta)) and temporally accumulated SRD and LRD (CUMSRD and CUMLRD) to account for the cyclic nature of LST and temporal variability of cloud effects, respectively. To consider the different relationships between input variables and LST under clear and cloudy-sky conditions, we used not only clear-sky GK2A LST but also all-sky in situ LST as dependent variables (Cho et al., 2022). The proposed all-sky GK2A LST reconstruction was performed on an hourly basis.

Notably, throughout the night, the CUMSRD takes information about the total amount of SRD over the preceding day, whereas the SRD becomes zero. This study reconstructed hourly all-sky LST data for three seasons (spring, summer, and autumn). Because SRD can be effective in reflecting changes in sunrise and sunset times depending on the season,

this study used SRD as an input variable, even during the nighttime.

3.1. Data processing

High-quality GK2A LST data were filtered using a quality flag (normal flag). GK2A data were unobserved from August 22 to December 20 at 06 UTC due to satellite wheel offloading. To ensure data continuity, LST at 06 UTC was interpolated by averaging the data from 05:50 and 06:10 UTC. To mitigate the discrepancy in spatial scale between GK2A and in situ LSTs, a second-order polynomial equation was applied (Christensen et al., 2008). After bias correction, we selected 36 ASOS stations satisfying the thresholds $R^2 \geq 0.9$ and RMSE ≤ 2.5 °C with GK2A LST for every hour.

3.2. Calculation of annual cycle parameters

To consider the annual periodicity of LST, we adopted annual cycle parameters (ACP) as model inputs: MAST, YAST, and THETA. MAST is derived as the mean of all LST values along the fitted curve, representing the annual average temperature for the region. YAST is defined as the difference between the maximum value of the curve and the MAST, indicating the amplitude of annual temperature variability. THETA represents the phase shift of the curve relative to the spring equinox, indicating the timing of peak temperature within the annual cycle. ACP can be derived through curve fitting, enabling its calculation even in missing data. However, due to its sensitivity to interannual variability, ACP should be recalculated annually to represent temporal dynamics accurately. Previous studies have established ACP as an influential independent variable capable of accurately estimating LST and representing the climatological characteristics and revealed that it provides insights into the local-scale annual variability of LST (Bechtel, 2012; Liu et al., 2019). ACP variables were retrieved using Eq. (1), based on the hourly GK2A clear-sky LST from January to December each year on each grid.

$$T(d) = MAST + YAST \times \sin\left(\frac{d2\pi}{365} + theta\right) \quad (1)$$

where d indicates the daily period relative to the spring equinox.

3.3. Accumulation of downward short- and longwave radiation

To reflect the temporal variability of cloud effects, we used instantaneous SRD and LRD as well as CUMSRD and CUMLRD in Eq. (2).

$$Rad_{cum} = \sum_{i=1}^{targettime}{}^{(Distance)} [Rad_i \bullet W_i], W_i = \sum_{i=1}^{targettime}{}^{(Distance)} \frac{1}{e^{i-1}} \quad (2)$$

where i represents the data obtained from i discrete time intervals prior to the current time and Rad_i indicates the instantaneous radiation value. W_i is composed of the inverse multiplier of e, which indicates the weight of heat attenuation over time. Through accumulation, CUMSRD and CUMLRD were calculated to effectively reflect the total shortwave and longwave radiation radiating to the surface during the daytime, and the amount re-radiated by clouds at night, respectively.

3.4. LGBM

LGBM is an ensemble machine learning method based on Gradient Boosting Decision Trees (GBDT), which offers the advantage of improved training speeds for massive datasets without compromising accuracy (Ke et al., 2017). Unlike other GBDT models, such as XGBoost, the LGBM utilizes a leaf-wise expansion strategy to guarantee well-balanced trees and enhance computational efficiency. The LGBM employs a gradient-focused sampling algorithm that keeps high-gradient samples and randomly selects those with low gradients to prioritize undertrained samples for enhanced training efficiency. In addition, the LGBM not only handles complex nonlinear relationships among multiple variables but maintains robustness against spatial autocorrelation, heteroscedasticity, and multicollinearity issues among variables (Amin et al., 2022; Brandt et al., 2024). Therefore, the LGBM improves training efficiency with low memory by reducing model. The "lightgbm" package in Python was used for implementation in this study.

Table 2
Range of each hyperparameter of light gradient boosting machine for Bayesian optimization.

Hyperparameter	max depth	min data in leaf	learning rate
Range	[4, 20]	[30, 500]	[0.01, 0.1]
Hyperparameter	**Number of estimators**	**colsample by tree**	**num leaves**
Range	[500, 2000]	[0.5 1]	[16, 500]

There are many hyperparameters (e.g., *max depth*, *min data in leaf*, *learning rate*, *n estimators*, *colsample by tree*, and *num leaves*) in the LGBM. Considerable computation time is required to determine the optimal combination of these hyperparameters using the grid search method. To reduce the computation time, we optimized them using the Bayesian optimization algorithm, which determines the optimal solution to the objective function by building a probabilistic (surrogate) model and then performing a Gaussian process (Shahriari et al., 2016; Shangguan et al., 2023). The main principle of Bayesian optimization method contains two parts: a surrogate model that performs probabilistic estimation and an acquisition function that recommends the best candidate set of hyperparameters for the optimal solution based on prior knowledge (Brochu et al., 2010). Table 2 lists the range of each hyperparameter considered for the Bayesian optimization in this study.

3.5. Model assessment and analysis

This study employed approximately 7–12 million samples for each hourly model and employed three distinct cross-validation methods to evaluate the robustness and generalization capability of the proposed model. First, we conducted five-fold spatial cross-validation (SPCV) using clear-sky GK2A LSTs to assess the spatial generalization of the model. In the SPCV process, the training and test sets were separated by a 7:3 ratio based on a random selection criterion that involved the extraction of 30 % of the total pixels from the study area. Second, southern-to-northern cross-validation (SNCV) was conducted using models trained exclusively using data from either South Korea or North Korea. The SNCV results allow the assessment of the performance of the proposed model in each area, enabling the evaluation of the model's applicability in regions where it has not been trained. Third, the leave-one-station-out cross-validation (LOSOCV) with 36 ASOSs was performed to evaluate the performance of estimated all-sky LST conditions. The LGBM was trained by dividing the training set into calibration and validation sets at an 8:2 ratio, ensuring that no test data were included to prevent overfitting.

Because satellite-based LST accuracy differs for land use/land cover (Yu et al., 2022), each land cover type in situ (e.g., forest, grass, cropland, and urban) was classified and compared for detailed validation. Quantitative validation was performed using the following statistical metrics: R^2, RMSE, and bias. The Gap-free ERA5 LST and products from previous research were used as comparative models (Jia et al., 2022b). Jia et al. (2023) conducted an hourly 5-km LST reconstruction using a statistical model incorporating instantaneous radiation and reanalysis data (https://doi.org/10.5281/zenodo.7487284). In addition, two different scenarios were compared to quantitatively examine the impact of radiation variables on the reconstruction of cloudy-sky LST. In Scenario 1, both the instantaneous and accumulated radiation components were excluded from the proposed model architecture. In Scenario 2, only

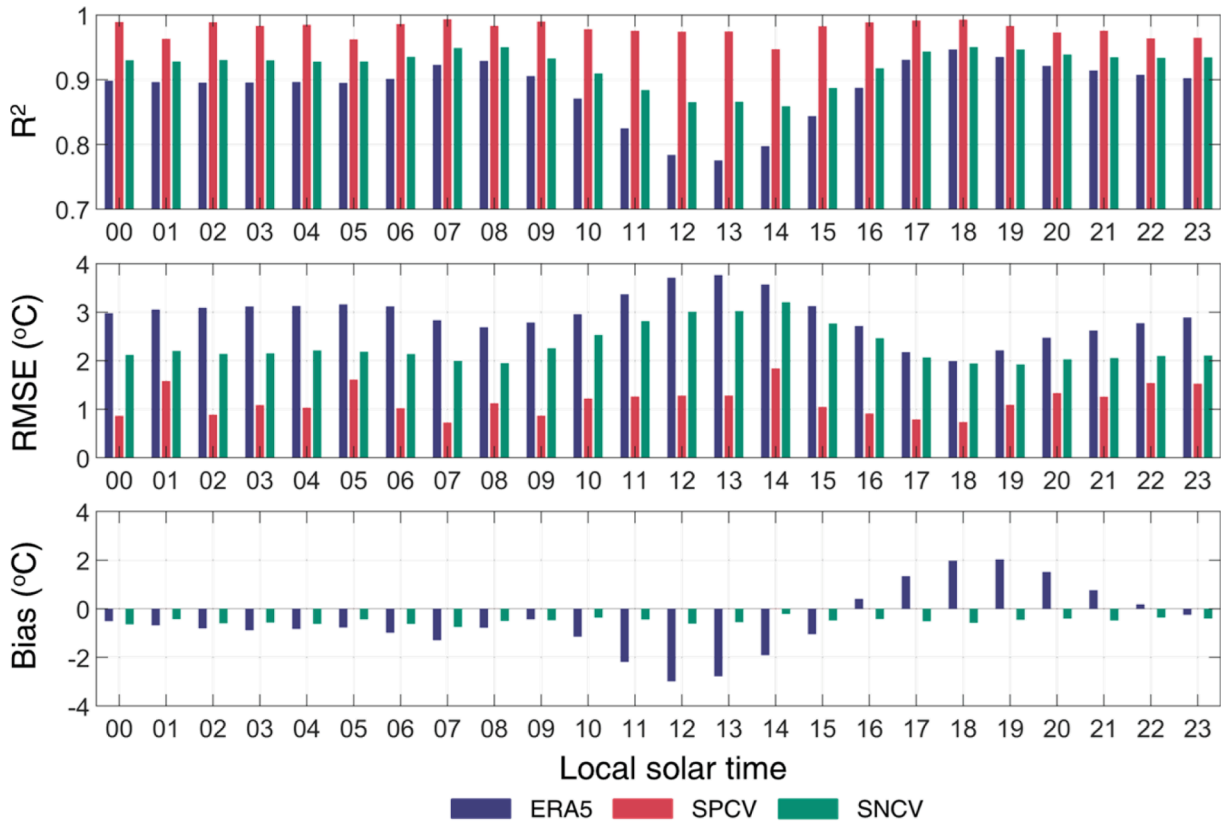


Fig. 3. Hourly performance of European Center for Medium-Range Weather Forecasts Reanalysis-Land (ERA5) and five-fold spatial cross-validation (SPCV) and Southern-to-Northern cross-validation from the proposed model compared to clear-sky GK2A LST.

accumulated radiation components within the proposed model architecture were excluded. In these scenarios, all variables except radiation information were employed.

Shapley additive explanation (SHAP) was adopted to analyze the contributions of the input variables and their interactions with the model's predictions. Based on game theory, the SHAP method quantifies the marginal effect of each feature by assessing its presence in all possible combinations (Lundberg and Lee, 2017). It produces detailed information on the magnitude of each feature's contribution and its effect on the model's final predictions. For the SHAP analysis in this study, we employed the "fasttreeshap" package in Python.

4. Results and discussion

4.1. Spatiotemporal evaluation using GK2A LST

Fig. 3 shows the performance of ERA5 and the SPCV and SNCV from the proposed model against the clear-sky GK2A LST. The ERA5 had a R^2 of 0.78–0.95, a RMSE of 1.99–3.77 °C, and a bias of −2.99–2.03 °C. On the other hand, The SPCV result shows a R^2 of 0.95–0.99, a RMSE of 0.73–1.84 °C, and an insignificant bias. The SNCV results had a R^2 of 0.85–0.95, a RMSE of 1.93–3.21 °C, and −0.75-−0.21. The two cross-validation results derived from the proposed model demonstrated high consistency with GK2A for all hours. Despite its slightly lower accuracy compared with the SPCV with more extensive training data, the SNCV maintained a high level of reliability.

All comparison results showed a diurnal cycle, with a tendency to consistently decrease during the daytime and increase at night. This was because the heat distribution and atmospheric turbulence above the ground during the daytime were more significant than those at night (Yang et al., 2017). ERA5 consistently displayed slight discrepancies compared to GK2A LST across all time, with these differences being more pronounced during daytime (e.g., R^2 of 0.93 (0.78) and RMSE of 2.18 °C (3.77°C) at 18 KST (14 KST). These tendencies can result from coarse spatial resolution of ERA5 (10 km) and uncertainty in the initial parameters including elevation, soil moisture, and skin conductivity (Trigo et al., 2021). Conversely, the proposed model exhibited insignificant differences with GK2A (e.g., R^2 values of 0.97–0.99 and RMSEs of 0.9 °C-1.55 °C with negligible biases).

ERA5 exhibited higher LST differences with GK2A spatially than the proposed model at four times in 6-h intervals (i.e., 00, 06, 12, and 18 KST) (Fig. 4). In particular, ERA5 showed a very different LST from GK2A in the coastal and high-altitude regions. This is possibly because the parameterization of land–ocean interactions is not perfect and the spatial information of the ERA5 LST has been smoothed, causing large systematic errors (Liu et al., 2015). By contrast, the proposed model exhibited significantly fewer spatial differences with GK2A than with ERA5 across all hours. These results indicate that the proposed model was spatiotemporally analogous to GK2A under clear-sky conditions.

4.2. Model evaluation using all-sky in situ LST

The proposed model showed higher consistency with in situ LST than with ERA5 and the results of Jia et al. (2022b), especially during daytime under clear-sky conditions (Fig. 5). Interestingly, GK2A LST exhibited slightly higher RMSE values at night than the other products, but the proposed model still showed high consistency, even though GK2A LST was used as a model target. This could be because the proposed model not only considered various physical variables (e.g., radiation-related, meteorological, and geographical variables) but also utilized bias-corrected in situ LST reflecting realistic cloudy-sky LST as one of the target variables. When comparing the proposed model that used in situ LST as the target variable with the model that did not, the proposed model showed a significantly higher accuracy (Fig. S4).

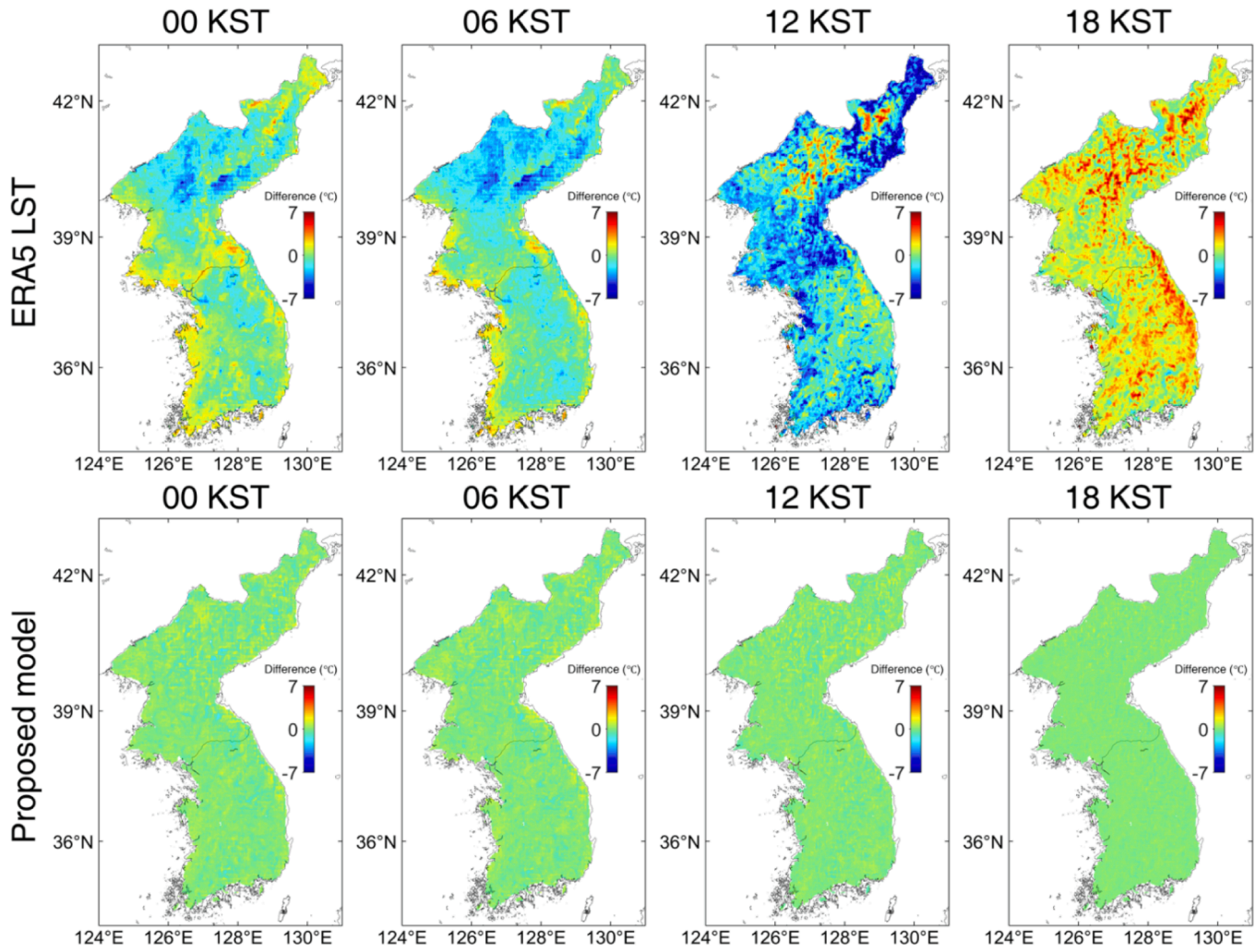


Fig. 4. LST difference maps between ERA5 and the proposed model in relation to the clear-sky GK2A LST at 00, 06, 12, and 18 KST. The top row represents the ERA5 LST differences, while the bottom row shows the differences for the proposed model. The difference map of the proposed model at each time was retrieved from the SPCV results, where smaller absolute differences indicate closer agreement with the GK2A LST. Positive values (red) and negative values (blue) represent the magnitude of the differences.

All LST products generally exhibited minimal differences compared to in situ LST during nighttime; however, ERA5 LST occasionally displayed minor discrepancies. In the daytime, the proposed model significantly outperformed the other three LST products, with an RMSE difference of less than 1 °C at 14 KST. Spatial scale mismatches between point-based in situ measurements and grid-based satellite products, combined with heterogeneous surface characteristics, are likely to contribute to larger errors during the daytime due to increased land surface temperature variability (Göttsche et al., 2013; Reiners et al., 2021; Reiners et al., 2023). Despite the slight performance decrease in the proposed model, it still outperformed other global-scale datasets such as ERA5 and Jia et al. (2022b) results. Furthermore, our proposed model consistently produced stable results across various periods and conditions within regional-scale contexts. Notably, during the daytime under cloudy conditions, the model demonstrated relatively better performance compared to clear-sky conditions. The reduction in solar radiation, which mitigates LST variability and aligns the dynamic range of LST between untrained ASOS stations and others, is responsible for this improvement. (Cho et al., 2022).

The type of land cover significantly influences LST in relation to incoming and outgoing surface energy balances (Zhao and Duan, 2020). Therefore, we analyzed the performance of clear- and cloudy-sky LST against the 36 ASOSs using MODIS land cover classes: forest (n = 4), grass (n = 2), cropland (n = 14), and urban (n = 16) (Fig. 6). The proposed model consistently demonstrated lower RMSE values than the other LST products, regardless of time and land cover. In comparison to GK2A, the proposed model demonstrated comparable performance during daytime under clear-sky conditions but exhibited superior results during nighttime, as detailed in Section 4.1. Additionally, under cloudy-sky conditions during the daytime, the proposed model achieved notably higher alignment with the in situ LST. The significant diurnal variability in RMSE across vegetation types (forest, grass, and cropland) can likely be attributed to the uneven spatial patterns of vegetation and their seasonal dynamics, which are associated with the surface cooling effect (e.g., shading and evaporative cooling) (Mal et al., 2022).

A time-series pattern of the reconstructed all-sky LST was analyzed to determine how well it matched the in situ measurements (forest [station number 21], grass [station number 30], cropland [station number 15 and 22], and urban [station number 12 and 16]) under all-sky conditions (Fig. 7). Despite the stations being situated across various land cover types, the proposed model showed a high level of concordance with the in situ LST during the clear and cloudy conditions, in contrast to the underestimation tendency of ERA5. The discrepancy between ERA5 and in situ observations may result from the inherently smoothed LST pattern within the reanalysis model and insufficient consideration of land cover characteristics, particularly vegetation (Johannsen et al.,

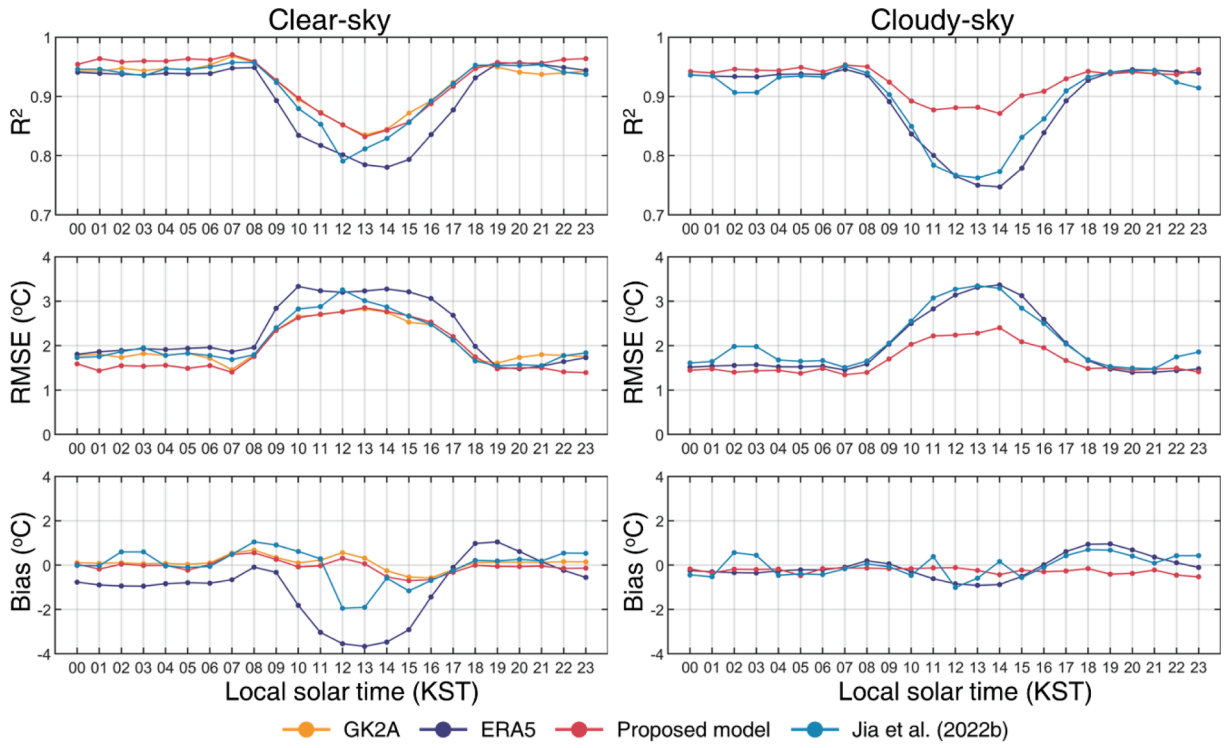


Fig. 5. Hourly leave-one-station-out cross-validation (LOSOCV) results of clear and cloudy-sky LSTs from GK2A, ERA5, proposed model, and the result of Jia et al. (2022b) compared to the bias-corrected in situ LST.

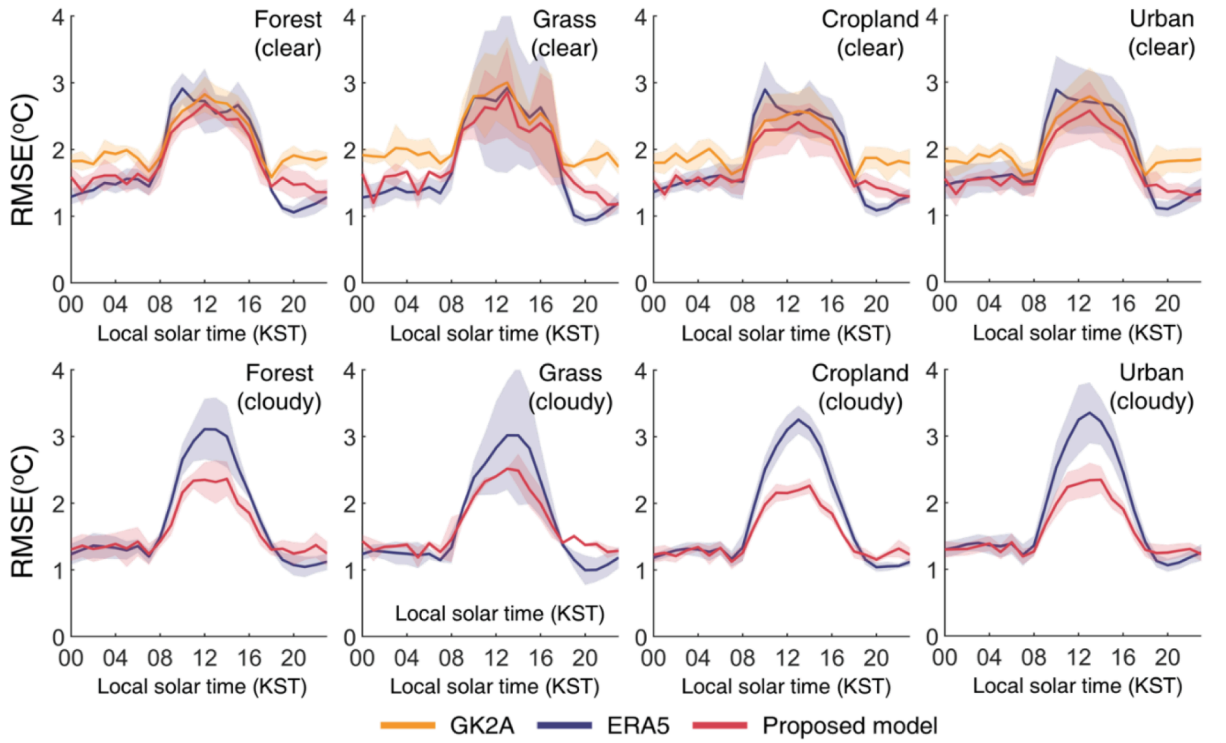


Fig. 6. Hourly root mean squared error (RMSE) distributions of GK2A, ERA5, and the proposed model compared to bias-corrected in situ LST by MODIS land covers under clear- and cloudy-sky conditions. The shaded area indicates the hourly standard deviation of RMSEs from 36 stations.

2019).

4.3. *Effects of accumulated radiation on all-sky LST reconstruction*

To explore the impact of accumulated radiation, the LOSOCV results for two scenarios (Scenarios 1 and 2) and the proposed model were

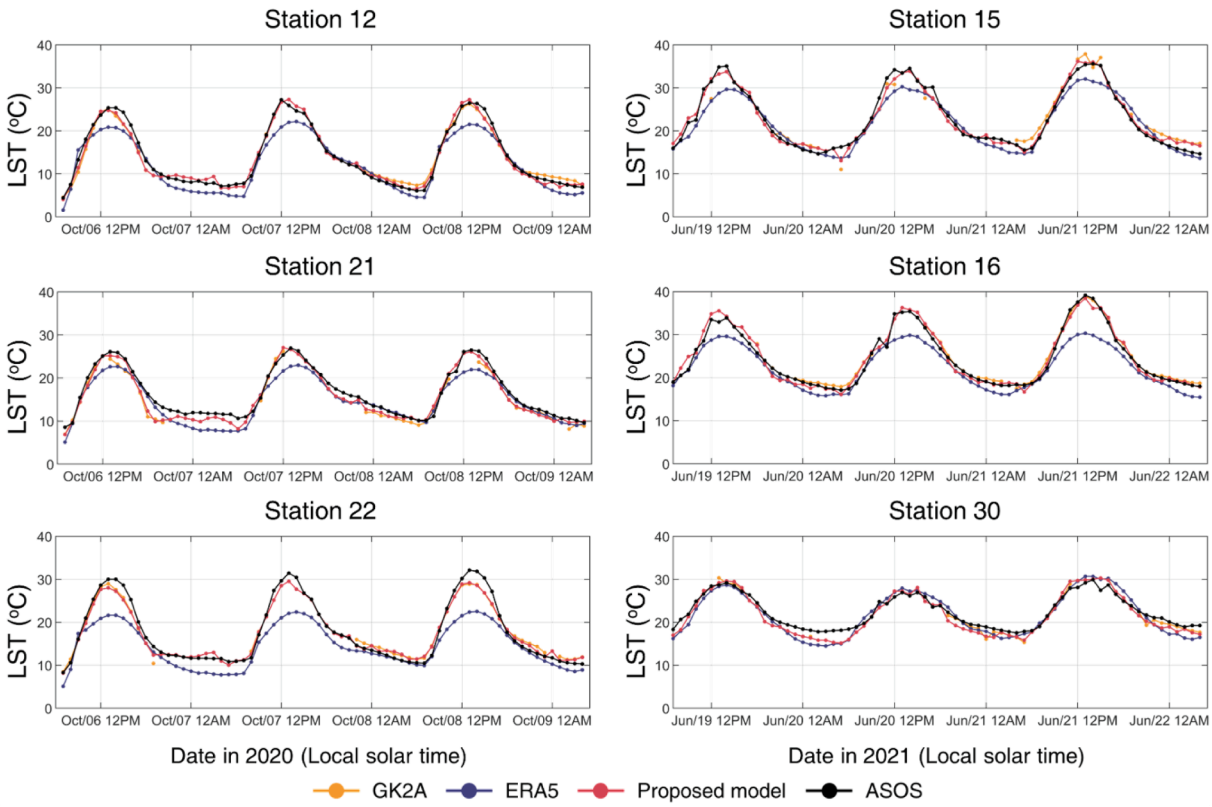


Fig. 7. Time-series of GK2A, ERA5, proposed model, and ASOS at six stations during autumn (from October 6, 07 local solar time (KST) to October 9, 04 KST in 2020) and summer (from June 19, 16 KST to June 22, 19 KST in 2021).

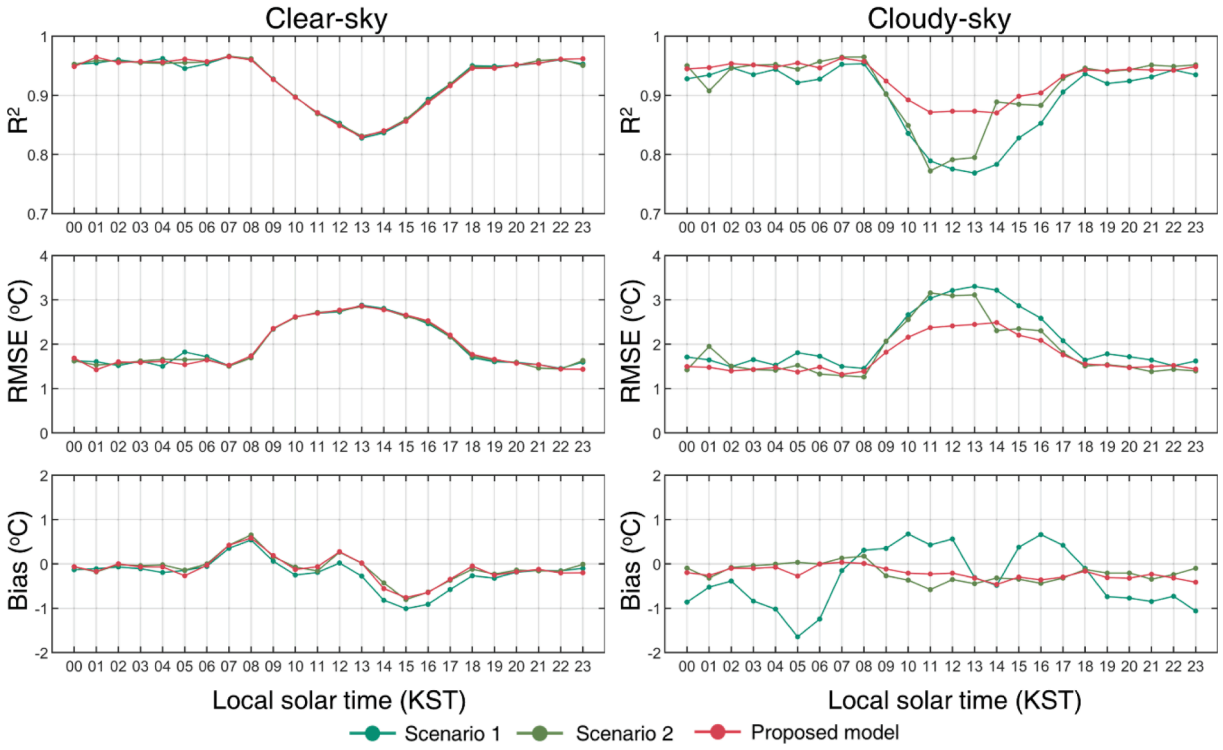


Fig. 8. LOSOCV results from scenario 1, scenario 2, and the proposed model compared to the bias-corrected all-sky in situ LST under the clear-sky and cloudy-sky conditions.

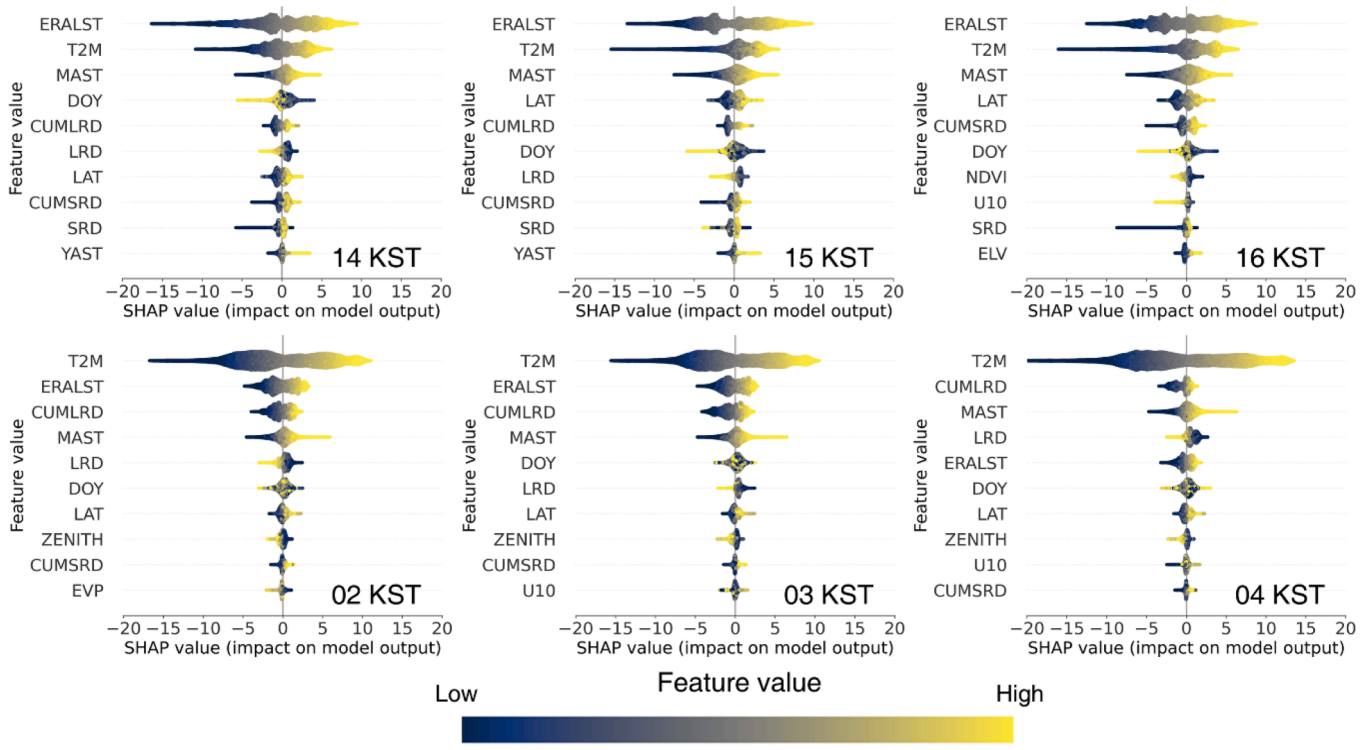


Fig. 9. Shapley additive explanations (SHAP) summary plots showing the top 10 input variables based on feature importance during the daytime (14 KST-16 KST) and the nighttime (02 KST-04 KST).

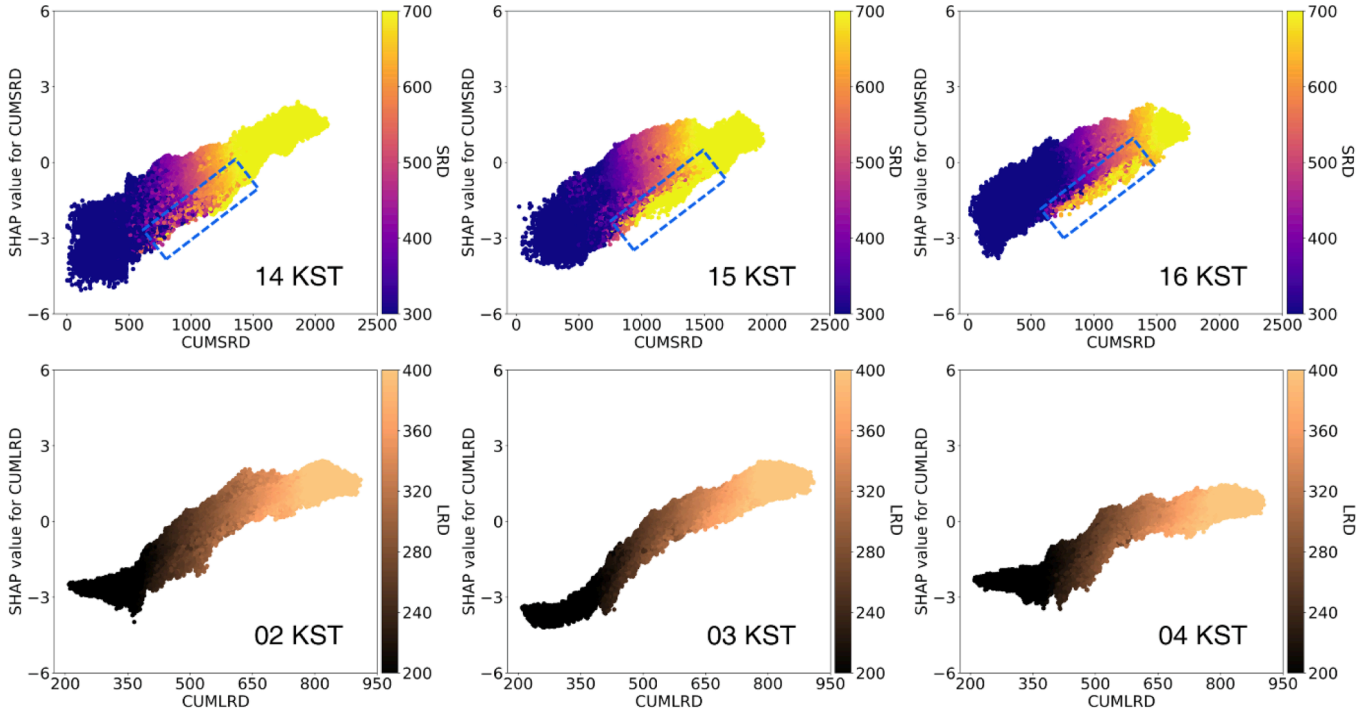


Fig. 10. SHAP dependence plots of the interaction between instantaneous radiation and accumulated radiation features during the daytime (14 KST-16 KST) and the nighttime (02 KST-04 KST). The dashed blue box indicates the section where temporally accumulated downward shortwave radiation values are close to the average of 700–1500 W/m^2, whereas downward shortwave radiation values are high with more than 700 W/m^2.

compared (Fig. 8). These three models showed similar performances under clear-sky conditions but exhibited significantly different performances under cloudy-sky conditions. Scenario 1 had the lowest R^2 and highest RMSE among the three models for all hours, with a slight warm bias during the daytime and a cold bias at night. Scenario 2 showed a better cloudy-sky LST reconstruction performance than Scenario 1 during the day. Notably, the proposed model demonstrated remarkable improvements in R^2 and RMSE compared with the other models,

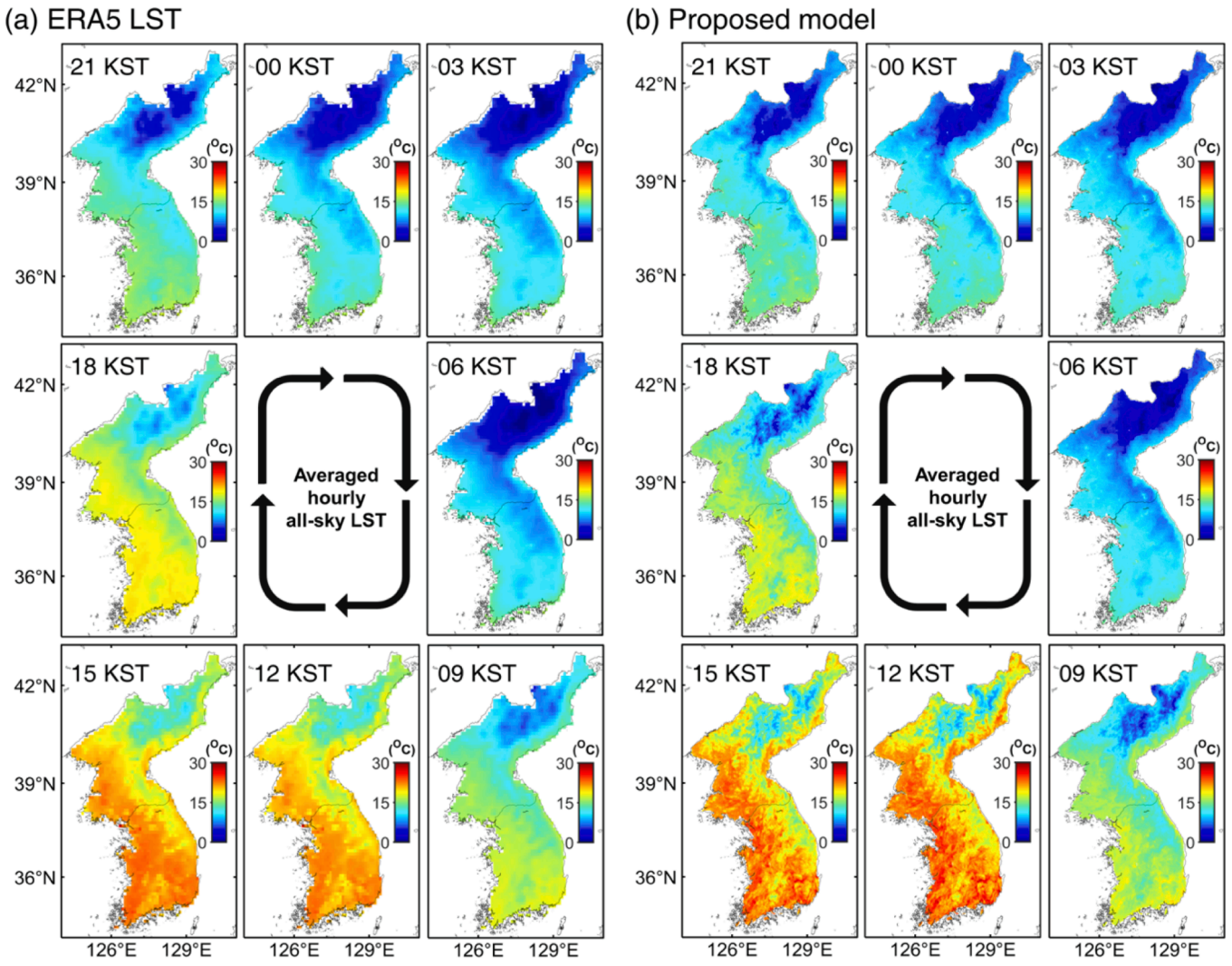


Fig. 11. Spatial distribution of hourly averaged LST from ERA5 (a) and the proposed model (b) during the study period. LST distribution maps are shown at three hour intervals from 00 KST to 21 KST. The color bar represents LST, ranging from 0 to 30°C.

especially during the daytime. Consequently, these results demonstrate that the use of accumulated radiation helps effectively consider cloud effects when reconstructing cloudy-sky LST.

SHAP analysis was conducted on both the daytime and nighttime models to examine the contributions and interactions of the input variables. Fig. 9 presents summary plots illustrating the variable contributions in the daytime (14 KST–16 KST) and nighttime (02 KST–04 KST), with the 10 most contributing input variables in the proposed model. During the day, ERA5 LST, T2m, and MAST were the primary contributors to the proposed model. In addition, geographic (e.g., elevation and latitude) and periodic (e.g., ACP and DOY) variables demonstrated relatively high variable importance. These variables have a strong correlation with the spatial and temporal distributions of LST (Liu et al., 2021).

Notably, the accumulated radiation variables (CUMSRD and CUMLRD) consistently exhibited a higher contribution to the model than the instantaneous radiation variables (SRD and LRD) during both daytime and nighttime. Note that LRD represents the total amount of longwave radiation from clouds and atmospheric gases reaching the Earth's surface, indicating cloud cover (Ahn et al., 2018). Therefore, during the daytime, lower LRD values indicate a lower presence of clouds and are associated with an increasing trend in LST. In contrast, higher CUMLRD values, suggesting the influence of long-duration cloud cover, were associated with increasing LST trends. This contrasting relationship between the two variables was more prominent at night than during the day, which is potentially related to the warm bias reported in previous studies (Jia et al., 2022a).

Using dependency plots, we further analyzed the interaction between instantaneous and cumulative radiation variables during both day and nighttime (Fig. 10). During the daytime, despite the significantly high SRD values ranging from 700 to 1500 W/m^2, a section exists where the CUMSRD values are not high (blue dashed box in Fig. 10). This section is likely to occur immediately after the dissipation of clouds, highlighting the benefit of instantaneous and accumulated radiation products in estimating LST rather than relying solely on instantaneous radiation, as in Scenario 2. However, at night, LRD and CUMLRD showed a clear linear relationship, without patterns such as the dashed blue section during the day. This may be because LRD already reflects the influence of CUMLRD at night, when longwave radiation is a dominant factor in LST. Additionally, this resulted in a performance similar to that of the proposed model and Scenario 2 at night.

4.4. Spatial distribution of reconstructed all-sky LST

ERA5 and the proposed model revealed a clear diurnal cycle pattern with an increasing trend in LST during the daytime owing to solar radiation and a decreasing trend after sunset (Fig. 11). In addition, they exhibited comparable spatial patterns over the Korean Peninsula,

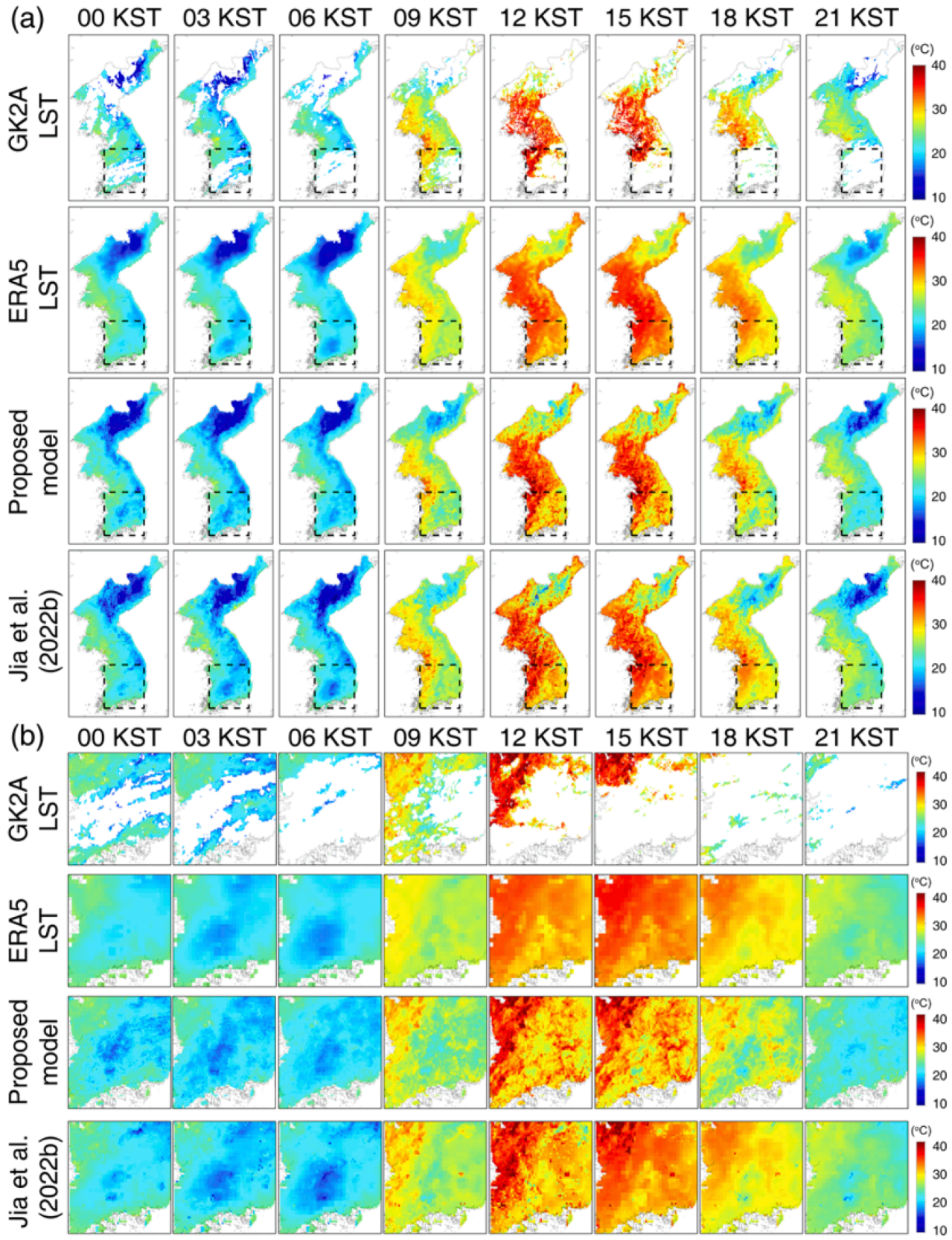


Fig. 12. (a) Spatial distribution of LST from GK2A, ERA5, the proposed model, and Jia et al. (2022b) on July 23rd, 2021, at a three-hour interval from 00 KST to 21 KST. The color bar represents LST, ranging from 10-40 °C. The white areas in the GK2A LST indicate missing values due to cloud cover within the study area. (b) Enlarged views of the LST distributions within the region highlighted by the black box for each product in (a).

Table 3
Distribution ratios of MODIS land cover classes for North Korea and South Korea.

	Land cover type (%)						
	Forest	Grass	Wetland	Cropland	Urban	Barren	Water
North Korea	60.51	9.03	4.92	23.75	1.51	0.10	0.17
South Korea	66.66	6.76	4.95	16.36	4.84	0.19	0.25

showing relatively lower LST in areas at higher altitudes than those at lower altitudes. However, ERA5 exhibited a relatively smooth spatial pattern (Fig. 11a), whereas the proposed model reasonably addressed the missing LST area without compromising spatial detail (Fig. 11b), which was evident in the hourly averaged maps. In addition, the eastern part of the study area (mainly forest type) at 21 KST and 00 KST had a low LST distribution following the mountain range, which may have resulted from the cooling effect of LST due to high vegetation and elevation (Ferreira and Duarte, 2019).

Fig. 12 presents the spatial distribution of LST for GK2A and ERA5, the proposed model, and the results of Jia et al. (2022b) on July 23, 2021. Although GK2A LST exhibits a more detailed spatial distribution than ERA5 LST, it is difficult to assert that GK2A adequately represents LST within a research area where cloud cover is substantial owing to numerous missing data points. Our proposed model effectively addressed these missing values while maintaining a continuous spatial distribution of LST. Under sustained cloudy conditions (in the southern part of the research area), a slight smoothing effect was observed in the spatial distribution of LST within ERA5 (Fig. 12 (b)) (Jia et al., 2022b). This effect may be due to the reduced spatial heterogeneity of the LSTs, which is attributed to their relatively coarse resolution and relatively small dynamic range under cloudy-sky conditions (Jia et al., 2022a). The proposed model showed improvements in the spatial pattern found in the original GK2A data and mitigated the smoothing effect observed in other LST products. These results denote the effectiveness of the proposed model in reconstructing the hourly missing GK2A LST.

4.5. Limitations

Although the proposed model showed strong spatiotemporal consistency with GK2A and hourly in situ data, it had some limitations. First, previous studies conducted LST reconstruction throughout all seasons; however, this study excluded winter from the analysis. When satellite data were unavailable owing to cloud contamination, in situ LST was used as the dependent variable in the proposed model. However, significant data contamination due to snow can result in prolonged periods of unchanged in situ LST values. The stuck LST poses difficulties during the validation process when utilizing data from affected stations to model and to determine whether snow influences the data. Although certain aspects of data contamination can be mitigated using quality control measures, establishing a consistent reference point for all ASOS observatories has inherent limitations. According to Zhou et al. (2017) the performance of LST reconstruction under cloudy-sky conditions is the lowest in summer compared to other seasons.

When valid quality assurance procedures are applied to in situ winter data in the future, our proposed model is expected to be effective for LST reconstruction during winter. Second, although this study focused on the GK2A dataset covering the Korean Peninsula, it should be noted that validation in the North Korea region is not feasible because of the unavailability of in situ station data. However, the land-cover distribution in North and South Korea predominantly comprises forests, agricultural lands, and grasslands, with significant similarity in the topographical characteristics of both regions (Table 3). The proposed model demonstrated consistent results, even in forested and grassland areas, which are predominant in North and South Korea. However, it is essential to validate these results using additional in situ measurements.

5. Conclusion

This study developed an LGBM-based hourly 2-km GK2A LST reconstruction model over the Korean Peninsula, aiming to fill in the missing values caused by cloud contamination by considering the effects of temporal variation in radiation. The SPCV for the clear-sky conditions had $R^2 > 0.97$, RMSE ≤ 1.55 °C, and negligible bias, indicating that the proposed model outperformed ERA5. Through LOSOCV results, the proposed model exhibited R^2 values of 0.87–0.92 and 0.94–0.95, and RMSE values of 1.82–2.49 °C and 1.44–1.52 °C during daytime (10 KST-16 KST) and nighttime (22 KST-04 KST) under cloudy-sky conditions, respectively, demonstrating higher consistency with in situ LST than with ERA5. In addition, when empirically compared with the latest LST reconstruction research and predefined Scenarios 1 and 2, the proposed model exhibited superior performance. Furthermore, SHAP analysis revealed that accumulated radiation, owing to its association with the presence and duration of clouds during daytime and night-time periods, is a highly effective variable for comprehending the temporal variability of cloud effects. These results emphasize the successful reconstruction of hourly LST using the proposed model, which considers cloud effects through the utilization of accumulated radiation information, particularly during the daytime. In future research, it could be possible to enhance LST estimation accuracy by exploring modifications to the accumulation method or by using an accumulative approach with more precise physical variables.

CRediT authorship contribution statement

Dukwon Bae: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Dongjin Cho:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Jungho Im:** Writing – review & editing, Supervision, Software, Resources, Funding acquisition, Conceptualization. **Cheolhee Yoo:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Yeonsu Lee:** Writing – review & editing, Formal analysis. **Siwoo Lee:** Writing – review & editing, Formal analysis.

Declaration of competing 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.

Acknowledgement

This research was supported by the Korea Meteorological Administration Research and Development Program under Grant KMIPA 2017-7010, by the National Research Foundation of Korea under Grant NRF-2021R1A2C2008561, and Korea Environment Industry & Technology Institute (KEITI) through Project for developing an observation-based GHG emissions geospatial information map, funded by Korea Ministry of Environment (MOE) (RS-2023-00232066).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.jag.2025.104468.

Data availability

Data will be made available on request.

References

Ahn, S.H., Lee, K.T., Rim, S.H., Zo, I.S., Kim, B.Y., 2018. Surface downward longwave radiation retrieval algorithm for GEO-KOMPSAT-2A/AMI. Asia Pac. J. Atmos. Sci. 54, 237–251. https://doi.org/10.1007/s13143-018-0007-1.

Amin, M.N., Salami, B.A., Zahid, M., Iqbal, M., Khan, K., Abu-Arab, A.M., Alabdullah, A. A., Jalal, F.E., 2022. Investigating the bond strength of FRP laminates with concrete using LIGHT GBM and SHAPASH analysis. Polymers (Basel) 14. https://doi.org/10.3390/polym14214717.

Bechtel, B., 2012. Robustness of annual cycle parameters to characterize the urban thermal landscapes. IEEE Geosci. Remote Sens. Lett. 9, 876–880. https://doi.org/10.1109/LGRS.2012.2185034.

Bird, D.N., Banzhaf, E., Knopp, J., Wu, W., Jones, L., 2022. Combining spatial and temporal data to create a fine-resolution daily urban air temperature product from remote sensing land surface temperature (LST) data. Atmosphere (Basel) 13. https://doi.org/10.3390/atmos13071152.

Brandt, P., Beyer, F., Borrmann, P., Möller, M., Gerighausen, H., 2024. Ensemble learning-based crop yield estimation: a scalable approach for supporting agricultural statistics. Gisci. Remote Sens. 61. https://doi.org/10.1080/15481603.2024.2367808.

Brochu, E., Brochu, T., Freitas, N., 2010. A Bayesian interactive optimization approach to procedural animation design, in: Computer Animation 2010 - ACM SIGGRAPH / Eurographics Symposium Proceedings, SCA 2010.

Cao, S., Cai, Y., Du, M., Weng, Q., Lu, L., 2022. Seasonal and diurnal surface urban heat islands in China: an investigation of driving factors with three-dimensional urban morphological parameters. Gisci. Remote Sens. 59, 1121–1142. https://doi.org/10.1080/15481603.2022.2100100.

Chen, Y., Yang, J., Yu, W., Ren, J., Xiao, X., Xia, J.C., 2023. Relationship between urban spatial form and seasonal land surface temperature under different grid scales. Sustain. Cities Soc. 89. https://doi.org/10.1016/j.scs.2022.104374.

Cho, D., Bae, D., Yoo, C., Im, J., Lee, Y., Lee, S., 2022. All-sky 1 km MODIS land surface temperature reconstruction considering cloud effects based on machine learning. Remote Sens. (Basel) 14. https://doi.org/10.3390/rs14081815.

Choi, Y.Y., Suh, M.S., 2020. Development of a land surface temperature retrieval algorithm from GK2A/AMI. Remote Sens. (Basel) 12. https://doi.org/10.3390/rs12183050.

Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35. https://doi.org/10.1029/2008GL035694.

Ding, L., Zhou, J., Li, Z.L., Ma, J., Shi, C., Sun, S., Wang, Z., 2022. Reconstruction of hourly all-weather land surface temperature by integrating reanalysis data and thermal infrared data from geostationary satellites (RTG). IEEE Trans. Geosci. Remote Sens. 60. https://doi.org/10.1109/TGRS.2022.3227074.

Ding, L., Zhou, J., Li, Z.L., Zhu, X., Ma, J., Wang, Z., Wang, W., Tang, W., 2023. Near-real-time estimation of hourly all-weather land surface temperature by fusing reanalysis data and geostationary satellite thermal infrared data. IEEE Trans. Geosci. Remote Sens. 61. https://doi.org/10.1109/TGRS.2023.3313730.

Ermida, S.L., Trigo, I.F., DaCamara, C.C., Jiménez, C., Prigent, C., 2019. Quantifying the clear-sky bias of satellite land surface temperature using microwave-based estimates. J. Geophys. Res. Atmos. 124, 844–857. https://doi.org/10.1029/2018JD029354.

Ferreira, L.S., Duarte, D.H.S., 2019. Exploring the relationship between urban form, land surface temperature and vegetation indices in a subtropical megacity. Urban Clim. 27, 105–123. https://doi.org/10.1016/j.uclim.2018.11.002.

Göttsche, F.M., Olesen, F.S., Bork-Unkelbach, A., 2013. Validation of land surface temperature derived from MSG/SEVIRI with in situ measurements at Gobabeb. Namibia. Int J Remote Sens 34. https://doi.org/10.1080/01431161.2012.716539.

Guo, A., Yang, J., Sun, W., Xiao, X., Xia Cecilia, J., Jin, C., Li, X., 2020. Impact of urban morphology and landscape characteristics on spatiotemporal heterogeneity of land surface temperature. Sustain. Cities Soc. 63. https://doi.org/10.1016/j.scs.2020.102443.

Guo, Y., Ren, Z., Dong, Y., Hu, N., Wang, C., Zhang, P., Jia, G., He, X., 2022. Strengthening of surface urban heat island effect driven primarily by urban size under rapid urbanization: national evidence from China. Gisci. Remote Sens. 59, 2127–2143. https://doi.org/10.1080/15481603.2022.2147301.

He, J., Zhao, W., Li, A., Wen, F., Yu, D., 2019. The impact of the terrain effect on land surface temperature variation based on Landsat-8 observations in mountainous areas. Int. J. Remote Sens. 40, 1808–1827. https://doi.org/10.1080/01431161.2018.1466082.

Huang, C., Duan, S.B., Jiang, X.G., Han, X.J., Leng, P., Gao, M.F., Li, Z.L., 2019. A physically based algorithm for retrieving land surface temperature under cloudy conditions from AMSR2 passive microwave measurements. Int. J. Remote Sens. 40, 1828–1843. https://doi.org/10.1080/01431161.2018.1508920.

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

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

Jia, A., Liang, S., Wang, D., Ma, L., Wang, Z., Xu, S., 2022b. Global Hourly, 5-km, All-sky Land Surface Temperature (GHA-LST) [Data set]. In Earth System Science Data (ESSD): Vol x (Version 02, Number x, p x). Zenodo. https://doi.org/10.5281/zenodo.7487284.

Jia, A., Liang, S., Wang, D., Mallick, K., Zhou, S., Hu, T., Xu, S., 2024. Advances in methodology and generation of all-weather land surface temperature products from polar-orbiting and geostationary satellites: a comprehensive review. IEEE Geosci. Remote Sens. Mag. 2–43. https://doi.org/10.1109/MGRS.2024.3421268.

Johannsen, F., Ermida, S., Martins, J.P.A., Trigo, I.F., Nogueira, M., Dutra, E., 2019. Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian Peninsula. Remote Sens. (Basel) 11. https://doi.org/10.3390/rs11212570.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Proces. Syst.

Kumar, S.V., Peters-Lidard, C.D., Tian, Y., Houser, P.R., Geiger, J., Olden, S., Lighty, L., Eastman, J.L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E.F., Sheffield, J., 2006. Land information system: An interoperable framework for high resolution land surface modeling. Environ. Model. Softw. 21, 1402–1415. https://doi.org/10.1016/j.envsoft.2005.07.004.

Li, Z.L., Wu, H., Duan, S.B., Zhao, W., Ren, H., Liu, X., Leng, P., Tang, R., Ye, X., Zhu, J., Sun, Y., Si, M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B.H., Yan, G., Zhou, C., 2023. Satellite remote sensing of global land surface temperature: definition, methods, products, and applications. Rev. Geophys. https://doi.org/10.1029/2022RG000777.

Liu, Y., Ortega-Farías, S., Tian, F., Wang, S., Li, S., 2021. Estimation of surface and near-surface air temperatures in arid northwest china using landsat satellite images. Front. Environ. Sci. 9. https://doi.org/10.3389/fenvs.2021.791336.

Liu, Y., Yu, Y., Yu, P., Göttsche, F.M., Trigo, I.F., 2015. Quality assessment of S-NPP VIIRS land surface temperature product. Remote Sens. (Basel) 7, 12215–12241. https://doi.org/10.3390/rs70912215.

Liu, Z., Zhan, W., Lai, J., Hong, F., Quan, J., Bechtel, B., Huang, F., Zou, Z., 2019. Balancing prediction accuracy and generalization ability: A hybrid framework for modelling the annual dynamics of satellite-derived land surface temperatures. ISPRS J. Photogramm. Remote Sens. 151, 189–206. https://doi.org/10.1016/j.isprsjprs.2019.03.013.

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

Mal, S., Rani, S., Maharana, P., 2022. Estimation of spatio-temporal variability in land surface temperature over the Ganga River Basin using MODIS data. Geocarto Int. 37, 3817–3839. https://doi.org/10.1080/10106049.2020.1869331.

Mo, Y., Xu, Y., Chen, H., Zhu, S., 2021. A review of reconstructing remotely sensed land surface temperature under cloudy conditions. Remote Sens. (Basel). https://doi.org/10.3390/rs13142838.

Peng, X., Wu, W., Zheng, Y., Sun, J., Hu, T., Wang, P., 2020. Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. Sci. Rep. 10. https://doi.org/10.1038/s41598-020-67423-6.

Reiners, P., Asam, S., Frey, C., Holzwarth, S., Bachmann, M., Sobrino, J., Göttsche, F.M., Bendix, J., Kuenzer, C., 2021. Validation of avhrr land surface temperature with modis and in situ lst—a timeline thematic processor. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13173473.

Reiners, P., Sobrino, J., Kuenzer, C., 2023. Satellite-derived land surface temperature dynamics in the context of global change—a review. Remote Sens. (Basel). https://doi.org/10.3390/rs15071857.

Sadeghi, G., Pisello, A.L., Safarzadeh, H., Poorhossein, M., Jowzi, M., 2020. On the effect of storage tank type on the performance of evacuated tube solar collectors: Solar radiation prediction analysis and case study. Energy 198. https://doi.org/10.1016/j.energy.2020.117331.

Shafizadeh-Moghadam, H., Weng, Q., Liu, H., Valavi, R., 2020. Modeling the spatial variation of urban land surface temperature in relation to environmental and anthropogenic factors: a case study of Tehran, Iran. Gisci. Remote Sens. 57, 483–496. https://doi.org/10.1080/15481603.2020.1736857.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2016. Taking the human out of the loop: A review of Bayesian optimization. In: Proceedings of the IEEE. https://doi.org/10.1109/JPROC.2015.2494218.

Shangguan, Y., Min, X., Wang, N., Tong, C., Shi, Z., 2023. A long-term, high-accuracy and seamless 1km soil moisture dataset over the Qinghai-Tibet Plateau during 2001–2020 based on a two-step downscaling method. Gisci. Remote Sens. 61. https://doi.org/10.1080/15481603.2023.2290337.

Shen, H., Huang, L., Zhang, L., Wu, P., Zeng, C., 2016. Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China. Remote Sens. Environ. 172, 109–125. https://doi.org/10.1016/j.rse.2015.11.005.

Song, P., Li, X., Ma, Z., Wu, S., 2024. Estimation of all-weather land surface temperature through correcting cloud-shadowing bias simulated by hourly cloud information. Int. J. Appl. Earth Obser. Geoinform. 128. https://doi.org/10.1016/j.jag.2024.103703.

Trigo, I.F., Ermida, S.L., Martins, J.P.A., Gouveia, C.M., Göttsche, F.M., Freitas, S.C., 2021. Validation and consistency assessment of land surface temperature from geostationary and polar orbit platforms: SEVIRI/MSG and AVHRR/Metop. ISPRS J. Photogramm. Remote Sens. 175, 282–297. https://doi.org/10.1016/j.isprsjprs.2021.03.013.

Wang, H., Mao, K., Shi, J., Bateni, S.M., Altantuya, D., Sainbuyan, B., Bao, Y., 2024. A normal form for synchronous land surface temperature and emissivity retrieval using deep learning coupled physical and statistical methods. Int. J. Appl. Earth Obs. Geoinf. 127. https://doi.org/10.1016/j.jag.2024.103704.

Wang, S., Shen, M., Liu, W., Ma, Y., Shi, H., Zhang, J., Liu, D., 2022. Developing remote sensing methods for monitoring water quality of alpine rivers on the Tibetan Plateau. Gisci Remote Sens 59, 1384–1405. https://doi.org/10.1080/15481603.2022.2116078.

Xu, S., Wang, D., Liang, S., Liu, Y., Jia, A., 2023. Assessment of gridded datasets of various near surface temperature variables over Heihe River Basin: Uncertainties, spatial heterogeneity and clear-sky bias. Int. J. Appl. Earth Obs. Geoinf. 120. https://doi.org/10.1016/j.jag.2023.103347.

Yang, Y.Z., Cai, W.H., Yang, J., 2017. Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. Remote Sens. (Basel) 9. https://doi.org/10.3390/rs9050410.

Yoo, C., Im, J., Cho, D., Yokoya, N., Xia, J., Bechtel, B., 2020. Estimation of all-weather 1 km MODIS land surface temperature for humid summer days. Remote Sens. (Basel) 12. https://doi.org/10.3390/RS12091398.

Yu, P., Zhao, T., Shi, J., Ran, Y., Jia, L., Ji, D., Xue, H., 2022. Global spatiotemporally continuous MODIS land surface temperature dataset. Sci. Data 9. https://doi.org/10.1038/s41597-022-01214-8.

Zeng, C., Long, D., Shen, H., Wu, P., Cui, Y., Hong, Y., 2018. A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. ISPRS J. Photogramm. Remote Sens. 141, 30–45. https://doi.org/10.1016/j.isprsjprs.2018.04.005.

Zhao, W., Duan, S.B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. Remote Sens. Environ. 247. https://doi.org/10.1016/j.rse.2020.111931.

Zhou, S., Cheng, J., Shi, J., 2022. A physical-based framework for estimating the hourly all-weather land surface temperature by synchronizing geostationary satellite observations and land surface model simulations. IEEE Trans. Geosci. Remote Sens. 60. https://doi.org/10.1109/TGRS.2022.3222563.

Zhou, W., Peng, B., Shi, J., 2017. Reconstructing spatial–temporal continuous MODIS land surface temperature using the DINEOF method. J. Appl. Remote Sens. 11, 1. https://doi.org/10.1117/1.jrs.11.046016.