Improved fusion model for generating hourly fine scale land surface temperature data under all-weather condition

Ibrahim Ademola Adeniran [a], Majid Nazeer [a,*], Man Sing Wong [a,b], Rui Zhu [c], Jinxin Yang [d], Pak-Wai Chan [e]

[a] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China*
[b] *Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong, China*
[c] *Systems Science Department, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore*
[d] *School of Geographical Science, Guangzhou University, China*
[e] *Hong Kong Observatory, Hong Kong, China*

A B S T R A C T

Existing Land Surface Temperature (LST) fusion models encounter some challenges due to missing data, complex weather areas, and rapid land cover changes. To overcome these limitations, we proposed the Integrated SpatioTemporal Fusion Algorithm (ISFAT). ISFAT is developed based on contemporary fusion models but in addition incorporates data from partially contaminated LSTs using the masked weight function. This helps to predict fine-scale LST on prediction date while considering error resulting from landcover changes between the base and prediction date. This algorithm also factors in the calculation of model residuals, which are distributed back to the predicted fine-scale LST using the thin-plate spline function. The fine-scale LST on prediction can thereafter employed for predicting hourly fine-scale LST images by integrating a coarse resolution LST with hourly temporal resolution. Compared to contemporary LST fusion models, ISFAT demonstrates superior performance, with mean average differences of 0.1 K and 0.27 K over SADFAT and STITFM, respectively. Additionally, diurnal LST predictions from ISFAT compare well with air temperatures from automatic weather stations. Notably, on February 18, 2020, ISFAT effectively optimized fine-scale LST for Hong Kong, achieving an RMSE of 3.33 K, despite the limitation of cloud cover in the base date. The newly developed ISFAT could facilitate better LST retrieval over a large spatial coverage under different degrees of cloud contamination.

1. Introduction

Land surface temperature (LST) is an important index in studying surface energy exchange and interaction between the atmosphere and the land surface (Gong et al., 2023). Retrieving LST data from satellite sensors using remote sensing methods help overcome the limitations of in-situ LST measurement under varying temporal and spatial variations (Adeniran et al., 2022). However, the trade-off between temporal and spatial resolution of remotely sensed LSTs has also been a constraint to the widely-used remotely sensed LSTs (Sobrino et al., 2012). For example, geostationary satellites like Himawari-8/9, Geostationary Operational Environmental Satellites (GOES), and Meteosat 2–7 can obtain LST at a high temporal resolution (~1 h or higher) suitable for diurnal LST analysis but have a relatively coarse spatial resolution (~2000 m or lower), which makes detailed spatial analysis of LST in

heterogeneous surfaces difficult (Zhan et al., 2013). In contrast, polar-orbiting satellites like Landsat 4–9, ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), which can measure LST at high spatial resolution (~100 m or higher) and are capable of detailed spatial LST analysis in heterogenous surfaces, have a relatively coarse temporal resolution (~12 h or lower) (Adeniran et al., 2022).

To address this constraint, several LST fusion methods have been proposed to optimize the spatial and temporal resolution of LSTs from satellite sensors. These methods can be classified into two broad categories (Quan et al., 2018). The approaches in the first category require data from a single satellite sensor. This method entails spatial down-scaling of geostationary satellite LSTs or temporal interpolation of polar satellite LSTs (Zhan et al., 2012). The second category uses pairs of

---

* Corresponding author.
E-mail address: majid.nazeer@connect.polyu.hk (M. Nazeer).

multi-resolution LSTs from different satellites at distinct time instances. Sparse representation, endmember unmixing, and weight function-based fusion approaches are employed.

Owing to the longer wavelength of the thermal infrared (TIR) channel, compared to the visible and near-infrared (NIR) channel, LST are retrieved at coarse spatial resolution compared to data in the NIR channel of same satellite (Herrero-Huerta et al., 2019; Reddy & Manikiam, 2017). The spatial downscaling method thus optimizes the spatial resolution of the LST data by capitalizing on the finer resolution of data in the visible and near-infrared channels. The limitation of this method, however, is that the finest spatial resolution achievable using the method is derived from other bands on the same satellite. This resolution may not be sufficient for LST analysis in heterogeneous regions (Yunhao et al., 2014; Quan et al., 2018).

Temporal interpolation methods are developed to increase the temporal resolution of LST estimations primarily derived from polar orbiting satellites, which have long revisiting periods. The interpolation involves associating temporally discrete observations with the surface energy balance model or annual/diurnal temperature cycle (ATC/DTC) (Duan et al., 2014). In contrary to the spatial downscaling, temporal interpolation method integrates data from external source to improve the temporal resolution of LST estimates. The limitation of this technique is that, for effective interpolation at the diurnal scale, at least four LST observations on the prediction day are required. However, achieving this requirement is challenging, as it relies on a limited collection of cloud-free MODIS observations with a spatial resolution of 1 km (Quan et al., 2016), which is not suitable for LST analysis in highly heterogeneous areas.

Recently, the adoption of methods which can combine the TIR data from geostationary and polar orbiting satellite sensors has gained significant attention. These methods optimize the coarse spatial/low temporal resolutions of geostationary/polar orbiting satellite sensors for optimized prediction of LST. These methods capitalize on the spatial strength of polar orbiting satellites and the temporal strength of geostationary satellites. They achieve this by combining LST or TIR data from both satellites using various techniques, such as end-member unmixing methods (EMUM), sparse representation methods, or weight function-based methods. The spatial and temporal adaptive reflectance fusion model (STARFM) developed by Gao et al. (2006) is the most representative of the fusion models that employs data from multiple satellite sensors. STARFM, originally developed for the fusion of data from the visible and NIR bands primarily relies on weight function-based method and has been revised by several studies to meet different research objectives (Zhu et al., 2010; Quan et al., 2018; Weng et al., 2014). Notably the spatiotemporal integrated temperature fusion model (STITFM) developed by Wu et al. (2015) and Spatiotemporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) by Weng et al. (2014) emerged as one of the contemporary multi sensor based LST fusion model.

As an improvement, SADFAT integrates the interpolation of Annual/Diurnal Temperature Cycle (ATC) with the STARFM methodology to predict LST data at high spatiotemporal resolution. In SADFAT, fine-scale LST is achieved by firstly optimizing the radiance data to fine scale, after which the optimized fine-scale radiance data will be inverted to achieve the synthesized LST data. On the other end, contrary to STARFM, SADFAT, and other multi-sensor fusion models that use data from two different satellites (fine and coarse scale), STITFM introduced the integration of a moderate-scale LST data into the STARFM model. This integration aims to break large-scale differences between fine and coarse resolution data. Also, as against the SADFAT methodology, STITFM inverts the TIR data to get the LST at both fine, medium, and coarse scale before employing the fusion model to achieve fine-scale LST.

Although both the SADFAT and STITFM has been applied for fine scale LST optimization and has achieved relatively high accuracy, some limitation still influences their wide range application. This includes

difficulty in the adoption of SADFAT in region with complex environmental condition that results in large scale cloud cover. This is because it will be difficult to acquire multiple pairs of fine and coarse scale (FnCs) LST that are cloud free on the same base date which is an important requirement for this model. Also, given that prediction at fine scale LST in STITFM is based on a single fine-scale observation, it becomes difficult to capture large-scale LST changes between the base and prediction date. These changes may not be captured in the moderate and coarse resolution LST data used in the final prediction (Hilker et al., 2009), resulting in poor prediction result especially in regions with rapid land cover change over short period.

Considering the limitation facing the existing model, this study aims to develop an integrated spatiotemporal fusion algorithm (ISFAT) for generating fine scale LST data suitable for diurnal LST analysis under all-weather condition. To achieve this, a two-staged model was developed, estimating fine-scale LST on the prediction date and diurnal LST on the same date. Initially, both clear and partly contaminated LST data from multiple base dates were combined for fine-scale LST prediction on the prediction date. Subsequently, recognizing the relatively insignificant diurnal land use changes on the prediction date, a single fine-scale LST data point was integrated with medium and coarse-scale LST data for diurnal prediction. Additionally, considering that the two identified benchmark LST models follow two different approaches, i.e., SADFAT optimizes radiance data before inversion to LST, while STITFM inverts radiance data before fusion, this study first experiments with the two approaches. The aim is to determine the optimum method to be adapted in the improved model. Finally, the improved fusion model developed was employed for the estimation of diurnal LST at high spatial resolution over Hong Kong which is the study area for this study.

The next section focuses on the description of the study area and materials employed for this study. Section 3 focuses on the research methodology, while results from the analysis are presented in Section 4. Discussion and conclusions are provided in Section 5.

2. Study area and materials

Hong Kong was selected as the study area, which is located between longitude $111°50'7''$ E $\sim$ $114°26'30''$ E and latitude $22°9'14''$ N $\sim$ $22°33'44''$ N. Hong Kong has undergone a tremendous urbanization trend in recent years (Wang et al., 2020). While it is characterized by the abundance of highlands and hills, resulting in remarkably high density in the shallow landscape across its 18 districts (Liu & Zhang, 2011) (Fig. 1). With a population of 7.6 million over a land area of 1,114 km^2, Hong Kong is one of the most densely populated cities in the world (Lang et al., 2019). The region is in the subtropic humid climate zone and witnesses long rainy and cloudy weather throughout the year, making the use of remotely sensed data for urban analysis difficult (Fan et al., 2008). For example, from 2006 to 2022, out of the 610 Landsat satellite acquisitions covering the study area, only 9.5 % (58 satellite acquisitions) are cloud free, most of these cloud-free data acquisitions are between November and January. Therefore, exploring the functionality of LST fusion model in this area is of high significance (Ling et al., 2021). For the assessment of both the contemporary and the improved fusion models, clear sky and partly cloudy images covering the study area were collected. These images were obtained from Landsat-8, Sentinel-3, and Himawari-8 satellite sensors at spatial resolutions of 100 m, 1000 m, and 2000 m, respectively. The acquisition dates of these satellite data are limited to dates with clear and partially clouded (cloud cover < 50 %) fine scale satellite data details presented in Table1. Landsat-8 data were sourced from the United States Geological Survey website (https://glovis.usgs.gov/app?fullscreen=1). Sentinel-3 data from the Copernicus Sci-Hub and Himawari-8 data from the JAXA where it was archived in the NetCDF format using the P-Tress system. Hourly air temperature data (AT) from automatic weather station (AWS) sparsely distributed across the study area were sourced from the Hong Kong Observatory (HKO) to validate the optimized hourly LST data. While Hong Kong has
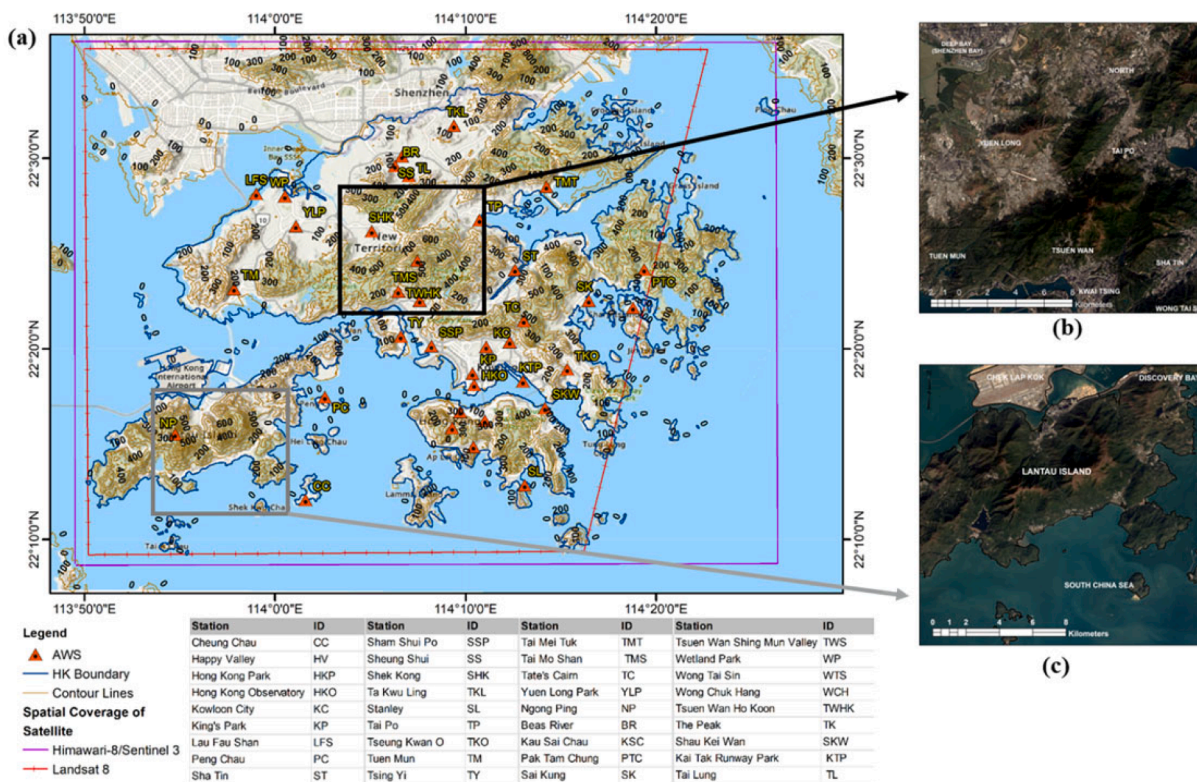
Fig. 1. Map of Hong Kong (a) locations of Automatic Weather Stations (AWS). (b) Satellite image of the relatively heterogenous region. (c) Satellite image of relatively homogenous region.

Table 1

Satellite data acquisition dates and overpass time (local time) for respective satellite sensors used in this study.

Date ID	Acquisition date	Landsat 8 (Acquisition time)	Sentinel-3 (Acquisition time)	Himawari-8 (Acquisition time)
d_1	27/09/2019	10:52	10:4822:48	00:00–23:50 (1 h interval)
d_2	29/10/2019	10:52	10:4822:48	00:00–23:50 (1 h interval)
d_3	14/11/2019	10:52	10:4822:48	00:00–23:50 (1 h interval)
d_4	30/11/2019	10:52	10:4822:48	00:00–23:50 (1 h interval)
d_p	18/02/2020	10:52	10:4822:48	00:00–23:50 (1 h interval)

a total of 50 AWS recording at an hourly temporal resolution, this study focused on 36 stations. These particular stations were selected because they fall within the tile of the fine, medium and coarse scale LST employed for this study.

2.1. Pre-processing

The digital numbers (DNs) from Band 10 and 11 of Landsat-8 satellite data were radiometrically corrected and converted to surface radiance. Simultaneously, radiance data from Sentinel-3 were retrieved from the SLSTR level 1B data. For Himawari 8, whose Thermal Infrared (TIR) data are distributed in brightness temperature, the data were converted to surface radiance. This conversion was done to facilitate the fusion of radiance. (Polehampton et al., 2022). The split window algorithm (SWA) was later adapted for the retrieval of LST from TIR data of the satellite data (Choi & Suh, 2018; Pérez-Planells et al., 2021; Rozenstein et al., 2014). This was achieved using ENVI 5.3 software. The surface

reflectance and Land Surface Temperature (LST) map retrieved from the three satellites were clipped to the geographic extent of the study area. After that, the data were geometrically corrected by projecting all the LST images to the same coordinate system (World Geodetic System 1984, UTM Zone 49). The surface radiance and LSTs from Himawari-8 (2 km resolution) and Sentinel-3 (1 km resolution) were subsequently resampled to the same spatial resolution as the LST data with the finest spatial resolution, i.e., Landsat 8 (100 m). Finally, the LST maps from the three satellites were evaluated to determine cloud-contaminated pixels using the Fmask python module for Landsat 8 and the cloud data layer provided for Sentinel-3 and Himawari-8 (Zhu & Woodcock, 2012).

3. Methodology

In this study, the improved spatiotemporal fusion algorithm (ISFAT) is aimed at optimizing the spatial and temporal resolution of LST data by learning from the merit of contemporary temperature fusion algorithms (i.e., SADFAT and STITFM). After developing the model, its accuracy was assessed over a region dominated by mixed land cover types over a relatively small area (heterogeneous) (Fig. 1b). Additionally, the assessment was conducted in a region dominated by a single land cover type over a relatively large area (homogeneous) (Fig. 1c), as extracted from Fig. 1a. The performance was then compared with the performance of the two benchmark models. In addition, the three models were employed to predict fine-scale LST of the study area, and hourly LSTs were predicted using ISFAT for diurnal analysis. In the absence of in-situ LST data in the study area, the accuracy of the models was evaluated based on the studies by Weng et al. (2014) and Bai et al. (2015). This evaluation involved comparing the optimized LST data with the LSTs obtained from the original Landsat-8 imagery. Metrics such as Root Mean Square Error (RMSE), correlation coefficient (r), Mean Difference (MD), and Mean Absolute Difference (MAD) values were employed for this assessment.

Due to the unavailability of fine-scale LST data at an hourly

resolution, the accuracy assessment of the optimized hourly LST was based on its comparison with the relative air temperature from an AWS (ΔAT). An AWS with a lower temperature served as the reference station, and the temperature value from this station was subtracted from measurements at other stations to compute ΔAT. Similarly, the optimized LST value corresponding to this reference AWS location was used as a baseline to compute ΔLST for various satellite sensors. The r and coefficient of determination R^2 of ΔLST and ΔAT was subsequently used to measure the accuracy of the optimized LST data. The study of Siu and Hart (2013) suggested the use of TKL as the preferred reference rural site for UHI and LST analysis in HK. This is based on the local climate zone the station is classified (Dispersed Low-Rise (BCZ10)) and its relatively low temperature. Further evaluations of hourly LST predictions involved comparisons across urban, suburban, and rural stations. Following the insights from Siu and Hart (2013), HKO was chosen as the urban reference, while CC and LFS stations were designated as suburban and rural reference points, respectively.

3.1. Benchmark models

3.1.1. STITFM

The STITFM algorithm by Wu et al. (2015) is one of the contemporary models designed particularly for the fusion of data in the thermal infrared region. STITFM was designed to predict fine-scale LSTs using data from multiscale geostationary and polar-orbiting satellites. The model optimizes fine-scale Land Surface Temperature (LST) on the prediction date using at least a pair of fine and medium-scale (FnMs) LST on a base date acquired at the same time (t_1). Additionally, it utilizes pairs of medium and coarse-scale (MnCs) LSTs acquired at a different time (t_2) and a coarse-scale LST acquired at the prediction time (t_p).

3.1.2. SADFAT

SADFAT, developed by Weng et al. (2014), is an improvement over the STARFM algorithm, originally developed to optimize surface reflectance to fuse radiance data in the TIR region of remotely sensed satellite data. In contrast to STARFM, which requires a pair of FnCs data on a base date (d_1) and coarse resolution data on the prediction date (d_p),

SADFAT requires an additional pair of FnCs data on another base date d_2. In addition, ATC was integrated into the SADFAT algorithm to estimate the conversion coefficient, which is a constant in the prediction of fine-scale radiance on d_p (Quan et al., 2016). The optimized fine-scale radiance data will be inverted to achieve fine-scale LST in the study area. The algorithm is based on the linear relationship between observations of radiances of a homogenous pixel (i) from two different sensors.

3.1.3. ISFAT

As presented in Fig. 2, The improved fusion model is built by two consecutive steps. The first step is the prediction of fine-scale LST on the prediction date, and the second step involves the prediction of hourly LST on the same day. The optimization of fine scale LST on the prediction date in ISFAT is based on SADFAT, a robust method that leverages the averaging of predictions from multiple base dates (Weng et al., 2014). Unlike STITFM which requires fine scale LST on a single base date, adapting SADFAT approach enables ISFAT to accurately account for LST changes resulting from land use changes between the base dates and prediction date. In contrast diurnal prediction in ISFAT learn from the STITFM model by integrating data from fine, medium, and coarse scale satellite to improve the diurnal prediction as land use throughout the data is relatively constant (Wu et al., 2015).

Step 1: Fine-scale prediction on the prediction date

The first step in the diurnal prediction of LST is to have a fine-scale LST on the prediction date. In ISFAT, improvements were made to the structure of the SADFAT algorithm to achieve a more accurate estimate on the prediction date. The improvements include (i) the estimation of residual and its distribution using fine scale land use map., and (ii) the introduction of masked weight variable to integrate information from partly clouded LSTs for prediction in complex regions.

(i) *Estimation and distribution of residuals*

Following the multi-sensor LST fusion framework (MLFF), the fine scale LST on of pixel i on d_p ($T_{F,i}(d_p)$) can be estimated using Eq. (1).
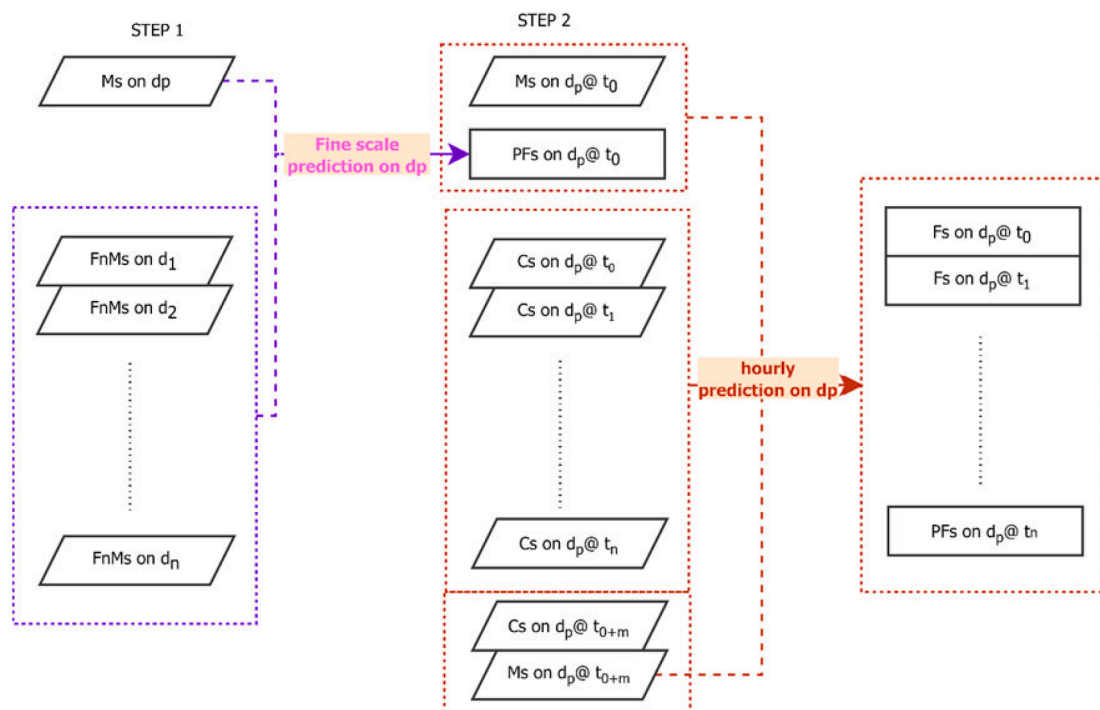

Fig. 2. Framework for fine scale diurnal LST prediction in ISFAT.

$$T_{F,i}(d_p) = T_{F,i}(d_1) + \sum_{n=1}^{N} w_n \times v_n \times \left(T_{M,i}(d_p) - T_{M,i}(d_1) \right) \qquad (1)$$

where $T_{F,i}(d_1)$ represents the fine-scale LST of the pixel on the base date, $T_{M,i}(d_p)$ and $T_{M,i}(d_1)$, represents the fine-scale LST of the pixel on the base date. w_i and v_i denote the weight and ATC of each similar pixel, as determined based on the studies by Gao et al. (2006) and Zhu et al. (2010).

Although contemporary LST fusion models have attempted to minimize the bias in fine-scale prediction by considering information from spectrally similar surrounding pixels and introducing the conversion coefficient v_n, the bias could not be completely captured. This ineptitude is attributed to factors such as incorrect selection of similar neighboring pixels and the lack of consideration for LST-related environmental factors. These factors include elevation, wind speed, and humidity differences across the study area, impacting the model's accuracy across the region (Wu et al., 2015). The resulting residual in the predicted LST at both fine and coarse scales can be estimated as presented in Eqs. (2) and (3).

$$R_{F,i,d_0}(d_p) = T_{F,i}(d_p) - T_{F,i,d_0}(d_p) \qquad (2)$$

$$R_{C,d_0}(d_p) = T_C(d_p) - T_{C,d_0}(d_p) \qquad (3)$$

where $R_{F,i,d_0}(d_p)$ and $R_{C,d_0}(d_p)$ are the model residual estimated at fine and coarse scale, respectively, $T_{F,i}(d_p)$ and $T_C(d_p)$ are the actual LST value for pixel i on d_p as estimated from the true fine and coarse scale satellite data, respectively. Whereas $T_{F,i,d_0}(d_p)$ and $T_{C,d_0}(d_p)$ represents the predicted LST for pixel i on d_p at fine and coarse scale respectively using data from base data d_0.

According to the LTMM, if the differences in the sensor are accounted for, the LST in each coarse-scale pixel should equal the aggregated sum of the fine-scale LST in the corresponding location and a systematic bias

the predicted fine scale LST on d_p, the fine scale land use data was regressed against the estimated residual to estimate the weight of the residual on each of the land use. After which the Thin Plat Spline (TPS) model was employed to distribute this residual on each of the land use within each coarse scale pixel. This approach helps in mitigating the block effect that might arise from directly adding the residual.

$$T_{F,i}(d_p) = T_{F,i,d_1}(d_p) + R_{F,i,d_1}(d_p) \qquad (8)$$

(ii) *Introduction of mask weight*

To effectively capture large scale land use change over the prediction period, the framework for contemporary LST fusion models like SADFAT recommended fine scale prediction from two pairs of FnCs cloud-free LSTs on two base dates (d_1 and d_2). In addition to a coarse-scale LST on the prediction date dp (Fig. 3(a)). However, due to complex atmospheric conditions in some regions, obtaining two pairs of cloud-free LSTs within an ATC becomes challenging. This limitation hinders the adoption of contemporary LST fusion models in these regions, as clouded areas result in missing information in the optimized LST (Fig. 3b). To overcome this limitation, ISFAT is designed to require a pair of cloud-free fine and coarse-scale Land Surface Temperature (LST) on at least one base date. Additionally, the model utilizes multiple pairs of partly clouded LSTs (with cloud cover below 70 %) at fine and coarse resolutions on other base dates, along with a coarse-scale LST on the prediction date dp (Fig. 3c).

As presented in ISFAT prediction date (d_p) framework (See Fig. 4), fine-scale LST on d_p is estimated from multiple base dates (d_0, d_1, …d_n) employing the improved multi-resolution framework in Eq. (8).

The multiple date predictions are then combined using the temporal and mask weight of each pixel prediction. The temporal weight (w_{t_x}) for each base date d_x (where x = 0, 1,2 …, n) in ISFAT is determined using Eq. (9).

$$w_{t_x} = \frac{1 / \left| \sum_{i=1}^{I} T_{M,i(d_x)} - \sum_{i=1}^{I} T_{M,i(d_p)} \right|}{\left(1 / \left| \sum_{i=1}^{I} T_{M,i(d_0)} - \sum_{i=1}^{I} T_{M,i(d_p)} \right| + 1 / \left| \sum_{i=1}^{I} T_{M,i(d_1)} - \sum_{i=1}^{I} T_{M,i(d_p)} \right| + \dots + 1 / \left| \sum_{i=1}^{I} T_{M,i(d_n)} - \sum_{s=1}^{I} T_{M,i(d_n)} \right| \right)} \quad x = 0, 1, \dots n \qquad (9)$$

(Deng & Wu, 2013). Thus, the relationship between the fine and coarse scale LST at d_p and d_0 can be presented using Eqs. (4) and (5), respectively.

$$T_C(d_p) = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{1}{a} T_{F,k}(d_p) - \frac{b}{a} \right) + \in \qquad (4)$$

$$T_C(d_0) = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{1}{a} T_{F,k}(d_0) - \frac{b}{a} \right) + \in \qquad (5)$$

The LST difference between d_p and d_0 can be estimated at coarse scale as presented in Eq. (6).

$$T_C(d_p) - T_C(d_0) = \frac{1}{Ka} \left(\sum_{k=1}^{K} T_{F,k}(d_p) - \sum_{k=1}^{K} T_{F,k}(d_0) \right) \qquad (6)$$

Eq. (6) is only true in an ideal situation when the predicted fine scale LST from base date d_0 equals the actual LST estimated from the satellite data on prediction date. The resulting residual in the predicted fine scale LST can be estimated according to Eq. (7).

$$R_{C,b}(d_p) = a(T_C(d_p) - T_C(d_0)) - \frac{1}{K} \left(\sum_{k=1}^{K} T_{F,k,d_0}(d_p) - \sum_{k=1}^{K} T_{F,k}(d_0) \right) \qquad (7)$$

Given that the estimated residual is at coarse scale, to distribute them to

Following the estimation of the temporal weight for each base date prediction, a mask is generated for each prediction date. In this mask, pixels with missing data receive a w_c value of zero, while pixels with accurately predicted LST data are assigned a w_c value of one. Combining this mask weight with the temporal weight facilitates the computation of the fine-scale LST on the prediction date ($T_{F,i}(d_p)$), as outlined in Eq. (10):

$$T_{F,i}(d_p) = \sum_{x=1}^{n} \left(w_{c_x} \times w_{t_x} + \frac{1 - \sum_{n=1}^{n} w_{c_n} \times w_{t_n}}{\sum_{n=1}^{n} w_{c_n} \times w_{t_n}} \times w_{t_x} \right) \times T_{F,i,d_x}(d_p) \qquad (10)$$

Step 2: Prediction of Hourly LST

After estimating the fine-scale LST on the prediction date ($T_{F,i}(d_p)$), the next step is the estimation of hourly LST for diurnal analysis. The ISFAT diurnal model integrates data from fine, medium, and coarse scales, addressing the significant scale gap between fine and coarse LST through medium-scale predictions. Utilizing improved MLFF in conjunction with residual estimation and distribution equation (Eq. (8)), ISFAT predicts the fine scale at each prediction time t_p (ranging from $t_{00:00}$ to $t_{23:00}$) at an hourly resolution. This prediction follows a two-stage method illustrated in Fig. 5.

Initially, a pair of medium and coarse scale LST at a base time t_m (($T_M(d_p, t_m)$ and $T_C(d_p, t_m)$) alongside a coarse scale LST at t_p is employed to predict medium scale LST at t_p ($T_M(d_p, t_p)$) using the MLFF. Subsequently $T_M(d_p, t_p)$, together with a pair of fine and medium scale LST
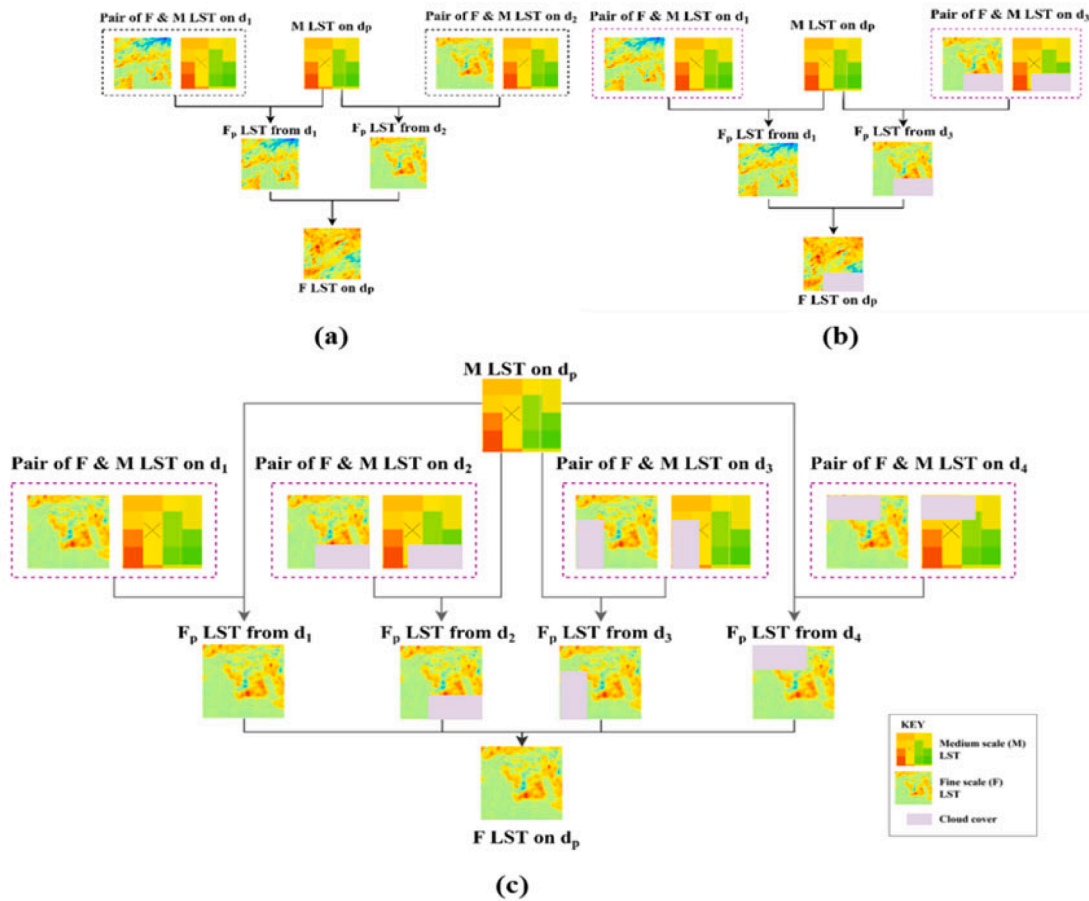
Fig. 3. Structure of LST image fusion algorithm (a) Structure of the contemporary fusion models (SADFAT) using pairs of cloud free data. (b) Structure of contemporary fusion models with partly clouded LSTs (clouded region represented with grey path on the fine and coarse LST). (c) Structure of Improved fusion model incorporating partly clouded LSTs from multiple base dates.

pairs at another base time t_o $(T_M(d_p, t_o)$ and $(T_F(d_p, t_o))$ are processed using the MLFF to predict fine scale LST at t_p denoted as $T_F(d_p, t_p)$. Here, t_0 and t_m are element of set t_p. Prediction for the other time is achieved by updating the coarse resolution LST $(T_C(d_p, t_p))$ hourly.

3.2. Assessment of ISFAT based on different LST fusion approaches

LST data optimization in literature has been estimated through two major approaches, i.e., fusion of radiance data retrieved from TIR bands of different satellites (fusion before inversion) or the fusion of LST data estimated from TIRS of different satellites (inversion before fusion). Thus, after developing the improved fusion model in this study, the accuracy of the model will be examined using the two diverse data combinations to determine the combination that will give the optimum result.

To achieve this, for the first combination, after downloading the satellite data from the fine, moderate, and coarse resolution satellites, satellite data processing will not include the inversion of the radiance data into LST. The pre-processed satellite data in the form of radiance data from fine, moderate, and coarse resolution satellites will thus be used as input data in the improved fusion model. After that, the resulting fine-scale radiance data will be inverted to achieve the LST for the study area using a suitable LST retrieval algorithm. For the other combination, the data pre-processing will be extended to include the inversion of the radiance data to get the LST data for the fine moderate and coarse resolution satellite using the appropriate LST retrieval algorithm. Furthermore, the retrieved LSTs will be used as input data for the improved image fusion model to yield fine-scale LST data on the

prediction date. The fusion result using the radiance data will thereafter be compared with the fusion result from the use of LST data as input data to determine the combination that have the better accuracy. For this study fine and medium scale LST from two base dates (d_2 and d_3) was used for the prediction of fine scale LST on the prediction date (d_p). While for STITFM a pair of fine and medium scale LST on d_3, a pair of medium and coarse scale LST d_4 and coarse scale LST on d_p was used to predict LST on the prediction date. While for ISFAT, pairs of fine and medium scale LST on d_1 to d_4 were combined with medium scale LST on d_p to predict fine scale LST on the d_p.

4. Results

4.1. Assessment of ISFAT based on the two approaches of LST fusion

To investigate the accuracy of the two different approaches adopted for image fusion, ISFAT was adopted for the prediction of fine scale LST on February 18, 2020, using both the inversion before fusion and fusion before inversion approaches. The accuracy assessment as presented in Table 2 revealed that fusion result from optimized LST is more comparable to the actual LST data with a r of 0.85 compared to 0.65 when ISFAT is used for the optimization of radiance data before inverting it to LST. This indicates that fusion of inverted LST data improves the prediction compared to the fusion of radiance data before inversion to LST. This is also evident in the improved RMSE, MD and MAD from 3.62 K to 3.07 K, 2.80 K to 1.76 K, and 2.92 K to 2.71 K respectively. This is primarily due to the intricacies of the fusion processes. Specifically, when fusion precedes inversion, errors present in the optimized radiance
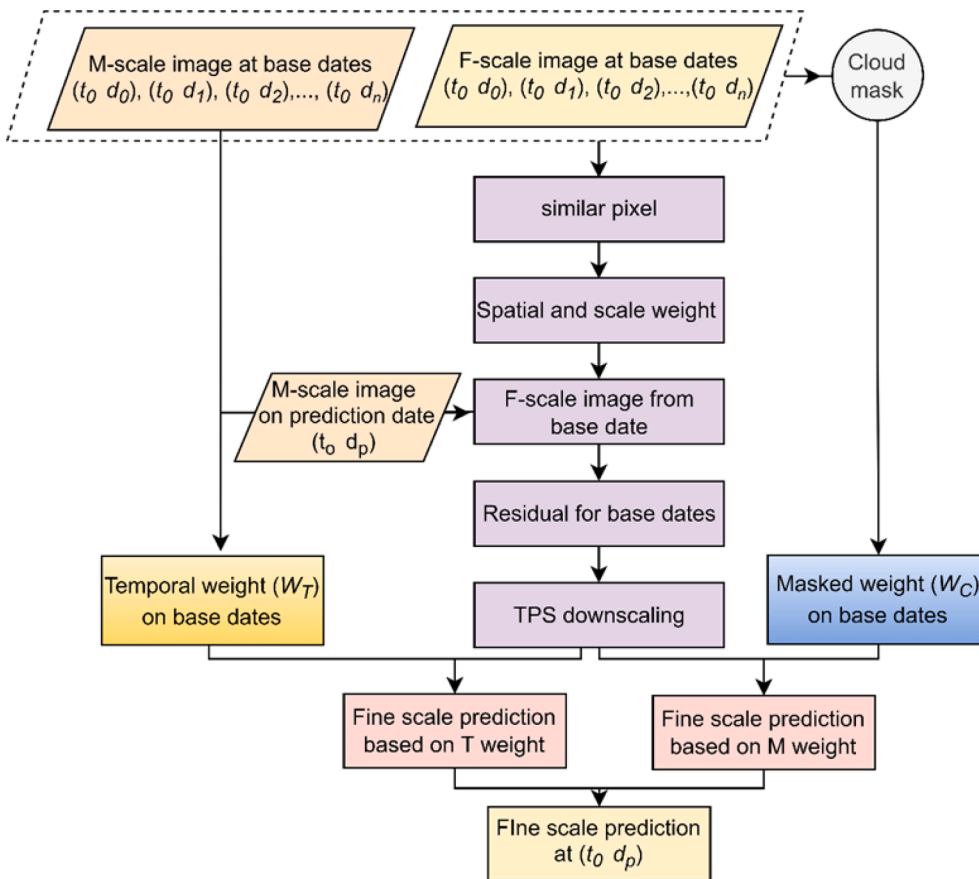
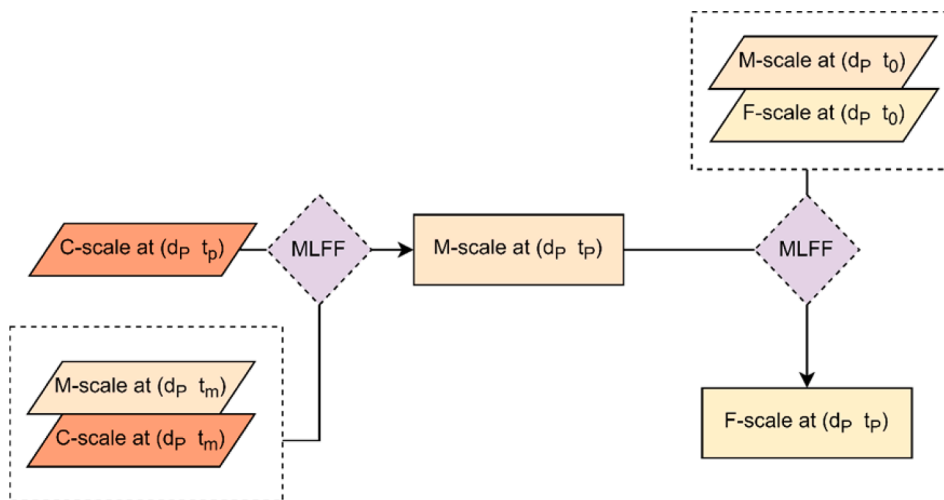**Fig. 4.** Framework for fine-scale lst prediction on d_p.

Fig. 5. Framework for hourly LST prediction on d_p.

Table 2
Assessment of different approach of LST fusion on February 18, 2020.

Approach	RMSE (K)	MD (K)	r	MAD (K)
Inversion before fusion	3.07	1.76	0.85	2.71
Fusion before inversion	3.62	2.80	0.68	2.92

already contains inherent errors, this additional layer of inversion amplifies the overall prediction bias. This is especially true when compared with the inversion of actual fine, medium, and coarse-scale LST data before fusion, which already eliminates errors resulting from differences in the spectral responses between the sensors, as data from all sensors are already inverted into LST.

data caused either by differences in spectral response functions between different sensors, errors in the adopted fusion model, among others are compounded during the inversion stage. Given that the radiance data

4.2. Assessment of LST fusion models

4.2.1. Assessment of fusion model over homogenous and heterogenous region

LST maps predicted using ISFAT, SADFAT and STITFM for Hong Kong on February 18, 2020, were assessed over both homogenous and heterogeneous regions. A quantitative assessment of fine scale LST map on d_4 highlighted a coefficient of variation (CV) of 0.82 for the heterogeneous region, contrasting with a CV of 0.47 for the homogeneous region. The higher CV in the heterogeneous region can be attributed to the diverse land uses present, resulting in high dominance of mixed pixel.

Fig. 6 offers a visual comparison: Columns 2–4 depict predictions from the respective models, while Column 1 showcases the LST data obtained from Landsat 8. Observations from both heterogeneous (Row 1) and homogeneous (Row 2) regions indicate that the predictions closely align with the actual LST values derived from Landsat 8 on the specified prediction date.

However detailed comparison as presented in Table 3, revealed that ISFAT outperformed both SADFAT and STITFM across both regions, registering the lowest RMSE values: 2.71 K for the heterogeneous region and 2.00 K for the homogeneous region. The improved performance of ISFAT is attributed to its refined methodology, incorporating a weighted combination of predictions derived from multiple fine-scale base dates. Such an approach equips the model to account for nuanced land use changes between the base and prediction dates. Additionally, the estimation of model residual, which was returned after the estimation contributed to the high performance. This residual however could not totally be eliminated because it was estimated at coarse scale and dispersed accordingly based on spatial properties. Between the two contemporary models, SADFAT demonstrated superior predictive capabilities over the heterogeneous region compared to STITFM. This advantage arises from SADFAT's reliance on a weighted combination of predictions from two pairs of fine and medium-scale (FnM) LSTs. This contrasts with STITFM's dependence on singular fine-scale data, which prevents STITFM from capturing land use changes that are not captured in the multiple pairs of medium and coarse-scale (MnCs) LSTs. Conversely, in the homogeneous region, STITFM shows better accuracy than SADFAT. This outcome can be attributed to STITFM's all-inclusive

approach, leveraging data from fine, medium, and coarse-scale LSTs. Additionally, the relatively static land use patterns in the region reduce the necessity for multiple base date predictions, as utilized by SADFAT.

4.2.2. Assessment of fusion models over study area

The three fusion model were thereafter employed to predict fine scale LST over Hong Kong. The prediction accuracy, when fusion results were compared with actual LST (Table 4) revealed varying degrees of accuracy across the models. Prediction using ISFAT is most significant with an RMSE, MD and r value of 3.33 K, 3.47 K, and 0.77, respectively. This is closely followed by prediction using SADFAT, whose prediction has a better r value + 0.01 but the RMSE and MD is higher with + 0.19 K and −0.01 K, respectively. Also, due to inaccessibility of two pairs of cloud free satellite scenes on two base dates over the prediction period optimized data using SADFAT include no data region resulting from cloud mask (Fig. 7d). This issue has however been corrected in ISFAT with the introduction of masked weight to ensure that the algorithm make use of cloud free pixel data on other dates to make up for predictions in cloud contaminated pixels. While STITFM was able to predict LST for the entire study area at fine scale because it only requires fine scale LST pairs at just a single date, but the accuracy is less than what is achievable when ISFAT or SADFAT is employed.

4.3. Evaluating diurnal fusion model result

The ISFAT diurnal algorithm was further employed for hourly fine scale LST prediction in the study area over the course of the day. As illustrated in Fig. 7, the spatial pattern of Land Surface Temperature (LST) over the course of the day revealed a noticeable increase in the LST over the study area from 01:00 UTC (9:00 am HKT) up till it reaches its peak around 05:00 UTC (1:00 pm HKT). Subsequently, the temperature began to drop until it hits the lowest LST around 21:00 UTC (5:00 am HKT). It was also observed that LST of regions along the coast in the study area are relatively high between 18:00 UTC (10:00 pm HKT) to 00:00 UTC (8:00 am HKT). This could be linked to lower heating and cooling rate of water and coastlands when compared with more urbanized regions in the study area (Li, 2020). The accuracy assessment of the predicted hourly fine-scale LST, as presented in Fig. 9, revealed that ISFAT is suitable for LST prediction in the study area. The r R^2 values,
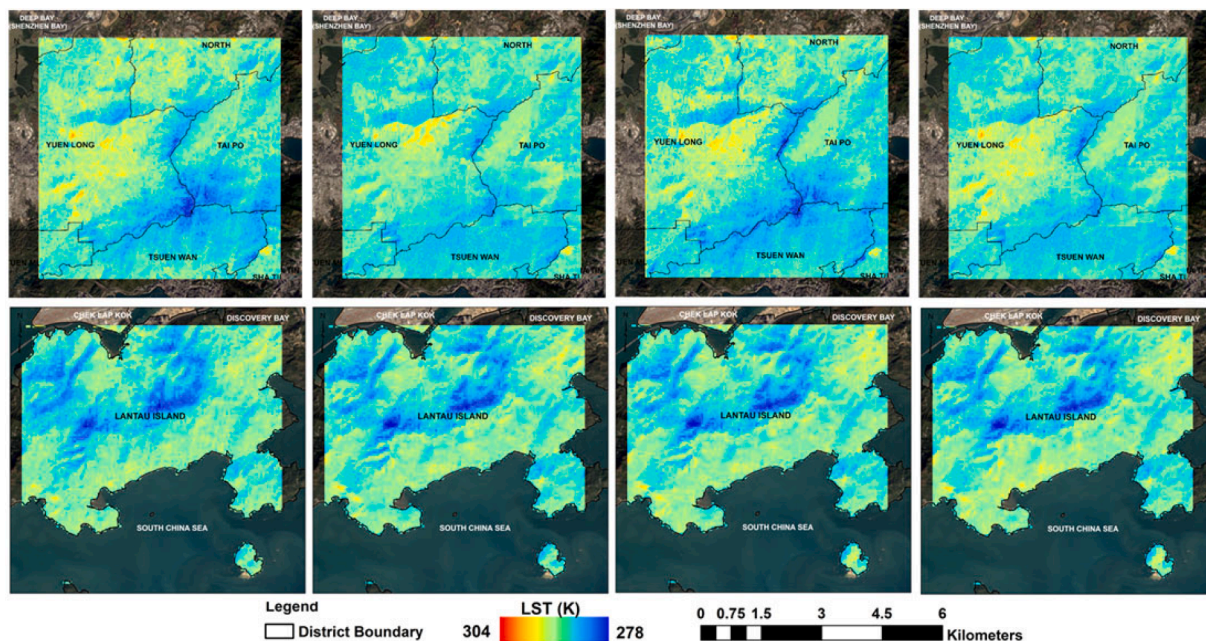

Fig. 6. Actual LST from Landsat 8 (Column-1) and predictions from STITFM (Column-2), SADFAT (Column-3), and ISFAT (Column-4) on February 18, 2020, at ~ 11:00am over heterogenous (Row-1) and homogenous (Row 2) region.

Table 3

Accuracy of the predicted fine scale LST compared to the actual LST on February 18, 2020, over heterogenous and homogenous region.

Model	Heterogenous				Homogenous			
	RMSE (K)	MD (K)	r	MAD (K)	RMSE (K)	MD (K)	r	MAD (K)
STITFM	3.91	2.07	0.70	2.98	1.27	1.97	0.91	2.02
SADFAT	3.12	2.45	0.84	2.81	2.06	3.85	0.86	4.15
ISFAT	3.07	1.76	0.85	2.71	1.20	1.95	0.93	2.00

Table 4

Accuracy of the predicted fine scale LST compared to the actual LST on February 18, 2020 over Hong Kong.

Model	RMSE (K)	MD (K)	r	MAD (K)
STITFM	4.01	3.97	0.68	4.43
SADFAT	3.52	3.48	0.78	4.19
ISFAT	3.33	3.47	0.77	3.91

when comparing ΔLST to ΔAT, ranged from 0.74 to 0.83 and 0.52 to 0.68, respectively (Fig. 9) (See Fig. 8).

As presented in Fig. 10, the comparison of relative air temperature (ΔAT) and relative LST (ΔLST) across selected urban, suburban, and rural stations reveals interesting patterns. Both ΔLST and ΔAT exhibit a sinusoidal pattern over 24 h, with the urban station exhibiting the highest relative temperature, followed by the suburban station, and the rural station displaying the least temperature difference. This observation further supports the findings depicted in Fig. 9, suggesting that the diurnal LST estimate from ISFAT is suitable for accurate fine-scale LST prediction throughout the day. However, a pattern of underestimation can be identified in the urban station, with ΔAT greater than ΔLST for most parts of the day. Conversely, the inverse pattern (overestimation) is observed in the rural station, with ΔAT lower than ΔLST for a greater part of the day. For the suburban station, it is a mixture of over and underestimation. These patterns of underestimation and overestimation in different stations highlight the influence of local factors and
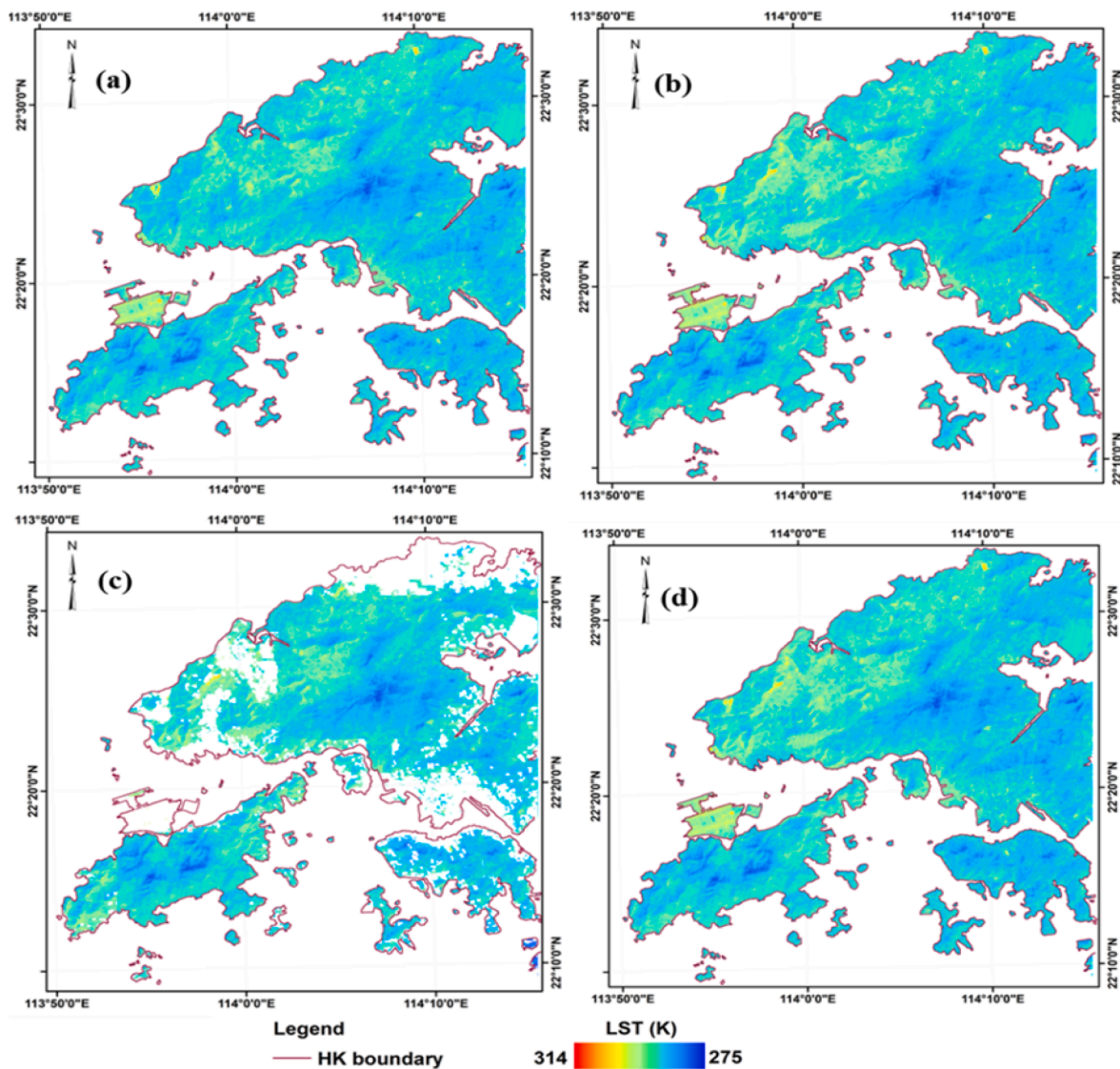

Fig. 7. Actual LST from Landsat 8 and predictions from fusion models on February 18, 2020 at ~ 11:00 am over Hong Kong (a)Actual (b) STITFM (c) SADFAT and (d) ISFAT.
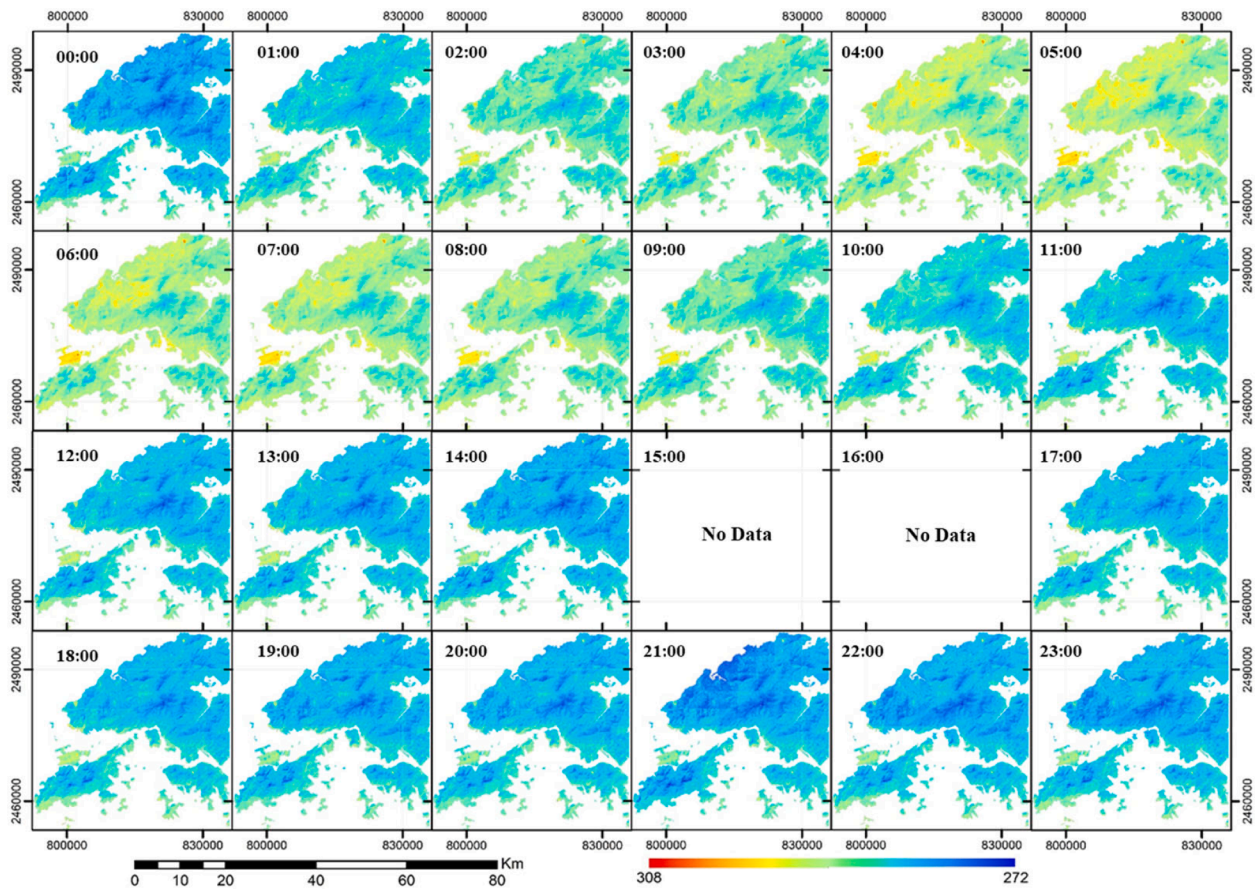
Fig. 8. Predicted LSTs at an hour temporal resolution at 100 m spatial scale on 18 February 2020, the time stamp in UTC is denoted in the top corner (local time = UTC time − 8 h). LSTs at 15:00 and 16:00 UTC could not be predicted due to missing Himawari-8 observations at these times.
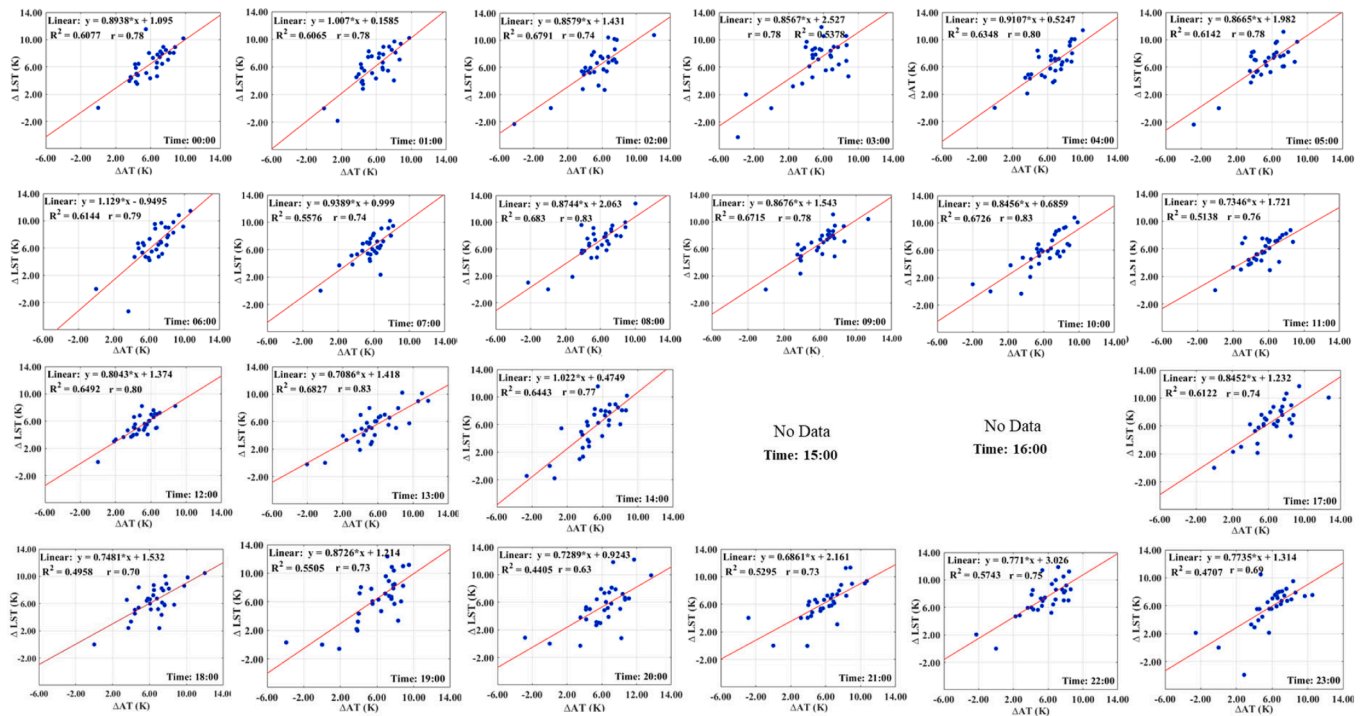

Fig. 9. Accuracy assessment of optimized hourly LST using relative optimized LST (ΔLST) and relative top of atmosphere temperature (ΔAT) from AWS.
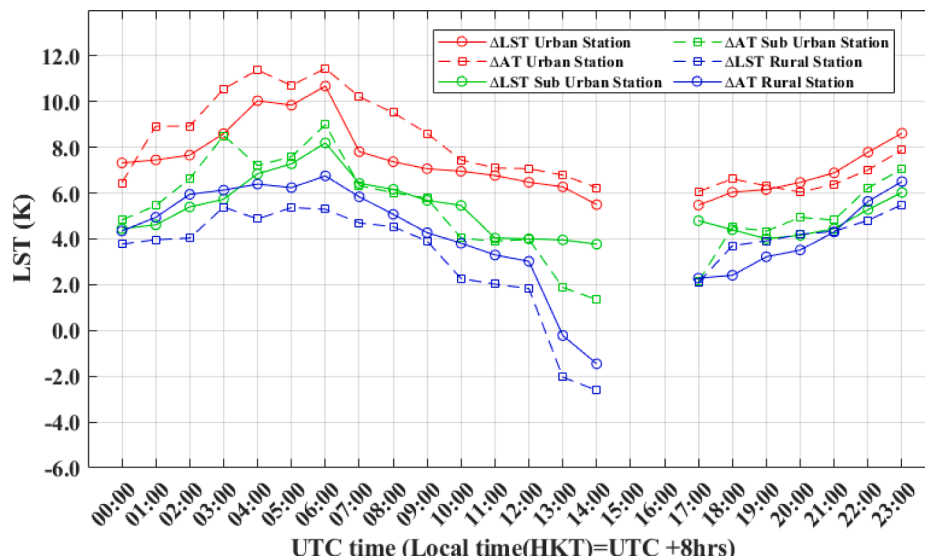
Fig. 10. Line graph of Δ LST against ΔAT at an Urban (HKO), suburban (CCH) and Rural station (LFS) over 24 h.

microclimates on the relationship between AT and LST. Factors such as urban heat island effects, variations in land cover, and local land use patterns can contribute to the observed differences. Understanding these patterns and their implications is crucial for accurate temperature modelling, urban planning, and assessing the impact of temperature variations on local environments and communities.

5. Discussion

This research introduces ISFAT, an improved LST fusion model designed to optimize LST data with a fine spatial resolution of 100 m and an hourly temporal resolution. By combining data from Landsat 8 (fine resolution), Sentinel-3 (medium resolution), and Himawari-8 (coarse resolution), ISFAT presents a multifaceted approach to diurnal LST optimization. The model's framework is subdivided into two pivotal stages. Initially, ISFAT concentrates on predicting fine-scale LST for a designated prediction date d_p. Subsequently, it focuses on diurnal prediction of fine-scale LST across the day. This subdivision allows the model to first, focus on capturing all spatial and temporal changes between prediction and base dates, followed by temporal changes across the day.

Thus, to predict fine scale LST on the prediction date, ISFAT requires pairs of fine and medium scale LSTs (FnMs) at multiple base dates, with only on cloud free LST pair mandatory as against contemporary model that require pairs of clear sky FnMs LST on two base dates. To account for missing data resulting from cloud cover in data, ISFAT introduced the mask weight variable (w_c). This variable combined with the temporal weight (w_t), bolsters the algorithm's accuracy in predicting fine-scale LST for the prediction date. Additionally, to mitigate model biases, ISFAT incorporates the estimation of prediction residuals. These residuals are subsequently integrated back into the predicted fine-scale LST, using the TPS model. Notably, when applied to LST prediction over Hong Kong in this study, ISFAT showcased enhanced accuracy, demonstrating an improved RMSE of 3.33 K. Transitioning to the algorithm's second stage, ISFAT learns from STITFM's methodology. In doing so ISFAT integrates fine, medium and coarse scale data in its prediction. In order to break the large-scale gap between fine and coarse scale LST, in diurnal prediction a pair of medium and coarse scale (MnCs) LST together with the coarse scale (Cs) LST on the prediction time t_p is combined to predict medium scale LST at the prediction time. After which pairs of FnMs LST another time stamp t_0 is combined with predicted medium scale LST at t_p. This method was able to achieve highly accurate diurnal LST predictions as revealed when compared to

relative air temperature data.

The limitation of this study however lies in the absence of in-situ LST data source to validate Fusion results. Consequently, accuracy assessments on the prediction date hinged on comparing optimized fine scale LST with LST inverted from fine scale satellite data on the prediction date. Then for diurnal LST accuracy assessment when fine scale LST is not available, diurnal LST prediction was assessed by comparing relative fusion result with relative diurnal air temperature from automatic weather stations.

6. Conclusion

The inherent trade-off between temporal and spatial resolutions in satellite-derived Land Surface Temperature (LST) has long hindered its effectiveness for urban thermal environment analyses. Despite advancements achieved by existing fusion models, challenges persist, notably the impact of missing data due to cloud cover and biases stemming from land use variations between prediction and base dates. Addressing these limitations, our study proposed a novel two staged LST fusion model, Integrated Spatiotemporal Fusion Algorithm (ISFAT). The first stage of the model focuses on fine-scale prediction for a specific date, while the second stage emphasizes hourly fine-scale prediction. ISFAT incorporate both clear sky and partly clouded fine and medium scale LST pairs from multiple base date for prediction of fine scale LST on the prediction date, with the condition of only one pair needs to be clear sky. By introducing the masked weight variable, the model effectively circumvents errors arising from missing data due to cloud cover in the base date pairs. Additionally, the model integrates the estimation of residual errors and redistributes them to the fine-scale predictions to minimize prediction inaccuracies. For the empirical application, we employed ISFAT to predict fine-scale LST over Hong Kong on February 18, 2020, utilizing data from Landsat 8 (fine scale), Sentinel 3 (medium scale), and Himawari 8 (coarse scale). An accuracy assessment of the fine-scale LST predictions on this date reveals that ISFAT outperformed SADFAT and STITFM, achieving an improved RMSE of 0.19 K and 0.68 K, respectively. Additionally, when comparing the relative diurnal fine-scale LST predictions with relative air temperature data from weather stations, ISFAT's diurnal predictions exhibited satisfactory performance, yielding an r score ranging between 0.74 and 0.83.

CRediT authorship contribution statement

Ibrahim Ademola Adeniran: Writing – review & editing,

Visualization, Methodology, Investigation, Data curation, Conceptualization. **Majid Nazeer:** Writing – review & editing, Visualization, Methodology, Investigation, Funding acquisition. **Man Sing Wong:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Investigation. **Rui Zhu:** Writing – review & editing, Visualization, Investigation. **Jinxin Yang:** Writing – review & editing, Visualization, Investigation. **Pak-Wai Chan:**.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Man Sing Wong reports financial support was provided by Research Grants Council of Hong Kong. Majid Nazeer reports financial support was provided by The Hong Kong Polytechnic University.].

Data availability

Data will be made available on request.

Acknowledgments

References

Adeniran, I.A., Zhu, R., Yang, J., Zhu, X., Wong, M.S., 2022. Cross-Comparison between Sun-Synchronized and Geostationary Satellite-Derived Land Surface Temperature: a case study in Hong Kong. Remote Sens. 14 (18).

Choi, Y.-Y., Suh, M.-S., 2018. Development of Himawari-8/Advanced Himawari Imager (AHI) land surface temperature retrieval algorithm. Remote Sens. 10 (12).

Deng, C., Wu, C., 2013. Examining the impacts of urban biophysical compositions on surface urban heat island: a spectral unmixing and thermal mixing approach. Remote Sens. Environ. 131, 262–274. https://doi.org/10.1016/j.rse.2012.12.020.

Duan, S.-B., Li, Z.-L., Tang, B.-H., Wu, H., Tang, R., 2014. Generation of a time-consistent land surface temperature product from MODIS data. Remote Sens. Environ. 140, 339–349.

Fan, F., Wang, Y., Wang, Z., 2008. Temporal and spatial change detecting (1998–2003) and predicting of land use and land cover in Core corridor of Pearl River Delta (China) by using TM and ETM+ images. Environ. Monit. Assess. 137 (1), 127–147.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44 (8), 2207–2218.

Gong, Y., Li, H., Shen, H., Meng, C., Wu, P., 2023. Cloud-covered MODIS LST reconstruction by combining assimilation data and remote sensing data through a nonlocality-reinforced network. Int. J. Appl. Earth Obs. Geoinf. 117, 103195.

Herrero-Huerta, M., Lagüela, S., Alfieri, S.M., Menenti, M., 2019. Generating high-temporal and spatial resolution TIR image data. Int. J. Appl. Earth Obs. Geoinf. 78, 149–162.

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., White, J.C., 2009. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sens. Environ. 113 (8), 1613–1627.

Lang, W., Chen, T., Chan, E.H., Yung, E.H., Lee, T.C., 2019. Understanding livable dense urban form for shaping the landscape of community facilities in Hong Kong using fine-scale measurements. Cities 84, 34–45.

Ling, J., Zhang, H., Lin, Y., 2021. Improving urban land cover classification in cloud-prone areas with polarimetric SAR images. Remote Sens. 13 (22).

Liu, L., Zhang, Y., 2011. Urban heat island analysis using the Landsat TM data and ASTER data: a case study in Hong Kong. Remote Sens. 3 (7), 1535–1552.

Pérez-Planells, L., Niclòs, R., Puchades, J., Coll, C., Göttsche, F.-M., Valiente, J.A., Valor, E., Galve, J.M., 2021. Validation of Sentinel-3 SLSTR land surface temperature retrieved by the operational product and comparison with explicitly emissivity-dependent algorithms. Remote Sens. 13 (11).

Polehampton, E., Cox, C., Smith, D., Ghent, D., Wooster, M., Xu, W., Bruniquel, J., Dransfeld, S. 2022. Sentinel-3-SLSTR-Land-Handbook Preparation and Operations of the Mission Performance Centre (MPC) for the Copernicus Sentinel-3 Mission.

Quan, J., Zhan, W., Chen, Y., Wang, M., Wang, J., 2016. Time series decomposition of remotely sensed land surface temperature and investigation of trends and seasonal variations in surface urban heat islands. J. Geophys. Res. Atmos. 121 (6), 2638–2657.

Quan, J., Zhan, W., Ma, T., Du, Y., Guo, Z., Qin, B., 2018. An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes. Remote Sens. Environ. 206, 403–423.

Reddy, S.N., Manikiam, B., 2017. Land surface temperature retrieval from LANDSAT data using emissivity estimation. Int. J. Appl. Eng. Res. 12 (20), 9679–9687.

Rozenstein, O., Qin, Z., Derimian, Y., Karnieli, A., 2014. Derivation of land surface temperature for Landsat-8 TIRS using a split window algorithm. Sensors 14 (4), 5768–5780.

Siu, L.W., Hart, M.A., 2013. Quantifying urban heat island intensity in Hong Kong SAR, China. Environ.. Monit. Assess 185 (5), 4383–4398.

Sobrino, J., Oltra-Carrió, R., Sòria, G., Bianchi, R., Paganini, M., 2012. Impact of spatial resolution and satellite overpass time on evaluation of the surface urban heat island effects. Remote Sens. Environ. 117, 50–56.

Wang, X., Yan, F., Su, F., 2020. Impacts of urbanization on the ecosystem services in the Guangdong-Hong Kong-Macao greater bay area, China. Remote Sens. 12 (19), 3269.

Weng, Q., Fu, P., Gao, F., 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sens. Environ. 145, 55–67.

Wu, P., Shen, H., Zhang, L., Göttsche, F.-M., 2015. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. Remote Sens. Environ. 156, 169–181.

Yunhao, C., Wenfeng, Z., Jinling, Q., Ji, Z., Xiaolin, Z., Hao, S., 2014. Disaggregation of remotely sensed land surface temperature: a generalized paradigm. IEEE Trans. Geosci. Remote Sens. 52 (9), 5952–5965.

Zhan, W., Chen, Y., Wang, J., Zhou, J., Quan, J., Liu, W., Li, J., 2012. Downscaling land surface temperatures with multi-spectral and multi-resolution images. Int. J. Appl. Earth Obs. Geoinf. 18, 23–36.

Zhan, W., Chen, Y., Zhou, J., Wang, J., Liu, W., Voogt, J., Zhu, X., Quan, J., Li, J., 2013. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. Remote Sens. Environ. 131, 119–139.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114 (11), 2610–2623.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.