



4-27-2017

# Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China


Yuan Z. Yang  
*Chinese Academy of Sciences, China*

Wen H. Cai  
*Chinese Academy of Sciences, China*

Jian Yang  
*University of Kentucky, jian.yang@uky.edu*

**[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)**

Follow this and additional works at: [https://uknowledge.uky.edu/forestry\\_facpub](https://uknowledge.uky.edu/forestry_facpub)

 Part of the [Forest Sciences Commons](#), [Natural Resource Economics Commons](#), [Natural Resources and Conservation Commons](#), and the [Natural Resources Management and Policy Commons](#)

## Repository Citation

Yang, Yuan Z.; Cai, Wen H.; and Yang, Jian, "Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China" (2017). *Forestry and Natural Resources Faculty Publications*. 15.  
[https://uknowledge.uky.edu/forestry\\_facpub/15](https://uknowledge.uky.edu/forestry_facpub/15)

This Article is brought to you for free and open access by the Forestry and Natural Resources at UKnowledge. It has been accepted for inclusion in Forestry and Natural Resources Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

**Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China**

**Notes/Citation Information**

Published in *Remote Sensing*, v. 9, issue 5, 410, p. 1-19.

© 2017 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Digital Object Identifier (DOI)**

<https://doi.org/10.3390/rs9050410>

Article

# Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China

Yuan Z. Yang<sup>1,2</sup>, Wen H. Cai<sup>1</sup> and Jian Yang<sup>1,3,\*</sup>

<sup>1</sup> Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China; yangyuanzheng8672@163.com (Y.Z.Y.); caiwenhua@iae.ac.cn (W.H.C.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Department of Forestry, TP Cooper Building, University of Kentucky, Lexington, KY 40546, USA

\* Correspondence: jian.yang@uky.edu; Tel.: +1-859-257-5820

Academic Editors: Parth Sarathi Roy and Prasad S. Thenkabail

Received: 4 February 2017; Accepted: 19 April 2017; Published: 27 April 2017

**Abstract:** Air temperature ( $T_{\text{air}}$ ) near the ground surface is a fundamental descriptor of terrestrial environment conditions and one of the most widely used climatic variables in global change studies. The main objective of this study was to explore the possibility of retrieving high-resolution  $T_{\text{air}}$  from the Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products, covering complex terrain in Northeast China. The All Subsets Regression (ASR) method was adopted to select the predictors and build optimal multiple linear regression models for estimating maximum ( $T_{\text{max}}$ ), minimum ( $T_{\text{min}}$ ), and mean ( $T_{\text{mean}}$ ) air temperatures. The relative importance of predictors in these models was evaluated via the Standardized Regression Coefficients (SRCs) method. The results indicated that the optimal models could estimate the  $T_{\text{max}}$ ,  $T_{\text{min}}$ , and  $T_{\text{mean}}$  with relatively high accuracies (Model Efficiency  $\geq 0.90$ ). Both LST and day length (DL) predictors were important in estimating  $T_{\text{max}}$  (SRCs: daytime LST = 0.53, DL = 0.35),  $T_{\text{min}}$  (SRCs: nighttime LST = 0.74, DL = 0.23), and  $T_{\text{mean}}$  (SRCs: nighttime LST = 0.72, DL = 0.28). Models predicting  $T_{\text{min}}$  and  $T_{\text{mean}}$  had better performance than the one predicting  $T_{\text{max}}$ . Nighttime LST was better at predicting  $T_{\text{min}}$  and  $T_{\text{mean}}$  than daytime LST data at predicting  $T_{\text{max}}$ . Land covers had noticeable influences on estimating  $T_{\text{air}}$ , and even seasonal vegetation greening could result in temporal variations of model performance. Air temperature could be accurately estimated using remote sensing, but the model performance was varied across different spatial and temporal scales. More predictors should be incorporated for the purpose of improving the estimation of near surface  $T_{\text{air}}$  from the MODIS LST production.

**Keywords:** air temperature; land surface temperature; remote sensing; MODIS; statistical models; Northeast China

## 1. Introduction

Air temperature ( $T_{\text{air}}$ ) near the ground surface is a fundamental descriptor of terrestrial environment conditions [1,2] and one of the most widely used climatic variables in global change studies. It plays an important role in multiple biological and physical processes among the hydrosphere, atmosphere, and biosphere [3–5]. Accurate estimation of  $T_{\text{air}}$  and mapping its spatial distribution are useful for predicting ecological consequences of climate change. For example, climate warming will lead to higher temperatures and an increase of extreme weather events, which are associated with changes in wildfire regime [6–8], forest biomass distribution [9] and crop yield [10,11]. The demand for accurate spatial  $T_{\text{air}}$  data over large scale has continued to rise [12,13].

Air temperature is often measured in thermometer shelters 1.5 m–2 m above the ground at meteorological stations, and is a commonly recorded form of meteorological observation data with

high accuracy and temporal resolution [5,14,15]. However, these recorded data are limited by the sparse distribution of meteorological stations. Hence, the spatial resolution of recorded air temperature data is coarse [1,3,16]. Various interpolation methods such as global interpolators, kriging, and inverse density weighting (IDW) are often carried out in order to compensate for this shortcoming [15,17]. These methods have been proven to be successful in estimating temperature near meteorological stations, but it could lead to considerable uncertainties caused by inhomogeneous distribution of meteorological stations, especially in the regions with complicated topography [4,18,19].

Satellite remote sensing can provide high spatio-temporal resolution data of land surface temperature (LST) at a global coverage [16,20]. It might provide a feasible way to improve the spatiotemporal accuracy of estimating  $T_{\text{air}}$ , because it is spatially contiguous and available on a regular basis, especially in regions where ground observations are too sparse to support reliable spatial interpolation [5,21]. Land surface temperature is defined as the “skin” temperature of the ground, which was detected by satellites via looking through the atmosphere to the ground [15,17]. The LST is not equivalent to  $T_{\text{air}}$ , and their relationship is complex from a theoretical and empirical perspective [22–24]. Most studies about estimating  $T_{\text{air}}$  using remote sensing indexes are based on simple linear regression models. For example, Tao et al. [25] showed that the correlation between nighttime LST and daily minimum  $T_{\text{air}}$  ( $R^2 = 0.93$ ) is significantly better than that between daytime LST and maximum  $T_{\text{air}}$  ( $R^2 = 0.79$ ). During daytime when there is direct solar illumination, the difference between LST and maximum  $T_{\text{air}}$  is mainly determined by the ground surface energy balance, which is a complex system depending on many additional factors such as solar radiation, cloud-cover, soil moisture and surface roughness [15,17]. Some effective measures were introduced to improve the precision of estimating daily maximum  $T_{\text{air}}$ , for example, the temperature-vegetation index (TVX) method was used in Zhu et al. [16]. This method is based on the assumption that the radiometric temperature of a fully developed vegetation canopy is in equilibrium with ambient  $T_{\text{air}}$  [3,15]. However, the performance of the TVX method mainly relies on the negative correlation between LST and the normalized difference vegetation index (NDVI), and NDVI varies greatly with different satellite sensor and land surface conditions [15,26]. In other studies, Benali et al. [5] used an optimization procedure with mixed bootstrap and jackknife resample to estimate maximum air temperature ( $T_{\text{max}}$ ), minimum air temperature ( $T_{\text{min}}$ ) and mean air temperature ( $T_{\text{mean}}$ ) based on LST and auxiliary data. However, previous studies often either lack spatial and/or temporal predictions, present relatively low predictive power or do not consider the model’s uncertainty. They also do not take into account the relative importance of the controls, such as latitude, longitude, elevation, Julian day, NDVI and so on, which might influence the relationship between LST and  $T_{\text{air}}$ .

Northeast China is a special geographical and important economic region. From southeast to northwest in this region, humid, semi-humid, and semi-arid zones distribute in turn with decreasing annual precipitation. The land covers in this region are diverse with natural forests, grassland, farmland, and urban area. Forests in Northeast China occupy about 30% of the total carbon storage and provide approximately 30% of the total timber production in China [27,28]. Boreal larch forests in this region are also fire prone, characterized by frequent surface fires, mixed with infrequent stand-replacing crown fires, with an estimated historical mean fire return interval of 30–120 years [29,30]. Regular spatial resolution of  $T_{\text{air}}$  can be used to predict possible changes in forest wildfire regimes and plant phenology [1], and is of great use for quantifying ecosystem processes and functions such as forest NPP and crop yield [31,32].

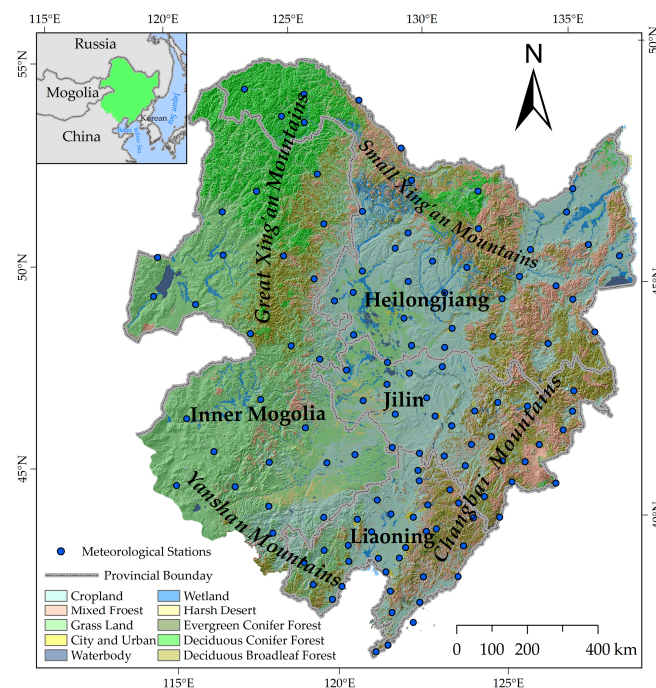
The main objective of our study was to systematically explore the relationship between  $T_{\text{air}}$  and Moderate Resolution Imaging Spectroradiometer (MODIS) LST over Northeast China. In order to estimate the eight-day  $T_{\text{max}}$ ,  $T_{\text{min}}$ , and  $T_{\text{mean}}$ , a statistical approach was implemented based on MODIS LST and auxiliary data. The specific objectives of this study were to (1) evaluate relative importance of the driving factors influencing the relationship between LST and eight-day  $T_{\text{air}}$ , (2) develop optimal models to predict  $T_{\text{air}}$  covering a large area that encompasses a broad variation of physiognomy and

land cover, and (3) estimate the optimal models' performance in temporal (i.e., month and season) and spatial (i.e., land cover) scales.

## 2. Data and Methods

### 2.1. Study Area

The study area includes Heilongjiang, Jilin, Liaoning provinces and the eastern Inner Mongolia Autonomous Region, which are located in Northeastern China ( $38^{\circ}45'N$ – $53^{\circ}34'N$ ,  $115^{\circ}20'E$ – $135^{\circ}05'E$ ), and encompasses about  $1.47 \times 10^6$  km<sup>2</sup>. The altitude ranges from 0 m to 2691 m. Geographically, this region is characterized by plains separated by four major mountain ranges (Changbai Mountains, Small Xing'an Mountains, Great Xing'an Mountains and Yanshan Mountains), and the majority of croplands and forests are distributed in these plain and mountain regions respectively (Figure 1). The climate is controlled by the high latitude East Asia monsoon climate and includes warm temperate, intermediate, and cool temperate zones distributed from south to north along latitude, and humid, semi-humid and semiarid zones from east to west. Meteorological station data sets (1984–2016) show that annual average temperature varies between  $-4.0$  °C and  $11.3$  °C, with a strong spatial gradient pattern. The monthly mean temperature ranges from the minimum  $-28.9$  °C in January to the maximum  $25.3$  °C in July. Annual average rainfall ranges from 246.1 mm to 1095.0 mm, and the precipitation from May to September accounts for 80% of the total. The vegetation in this area is dominated by grass, conifer forest, mixed, and deciduous broadleaf forest. The growing season starts in May and the vegetation reaches its maximum primary productivity in July or August, and then begins to grow slowly thereafter.



**Figure 1.** The geographic location of the study area, different vegetation types, and the distribution of the meteorological stations.

### 2.2. Data

#### 2.2.1. MODIS Data

The LST product from MODIS has been used in previous studies to derive  $T_{air}$  [5,15,16]. MODIS sensors were launched on board the National Aeronautics and Space Administration (NASA) Earth

Observing System (EOS) Terra and Aqua satellites in December 1999 and May 2002, respectively [16]. Both sensors are aboard sun-synchronous polar orbiting satellites. MODIS Terra data is available during 10:30–12:00 a.m. and p.m. (daytime/nighttime) local time, while MODIS Aqua sensor collects the imagery during 1:00–3:00 a.m. and p.m. (daytime/nighttime). The Aqua LST images were chosen instead of the scenes from the Terra platform for this study, because their acquisition time is closer to the actual occurrence of observed maximum and minimum temperatures [5,33,34]. Two MODIS land products (version 5) were chosen in this study: Aqua eight-day LST and emissivity (MYD11A2) and reflectance of bands 1–7 (MYD09A1) covered the entire study period from July 2002 to July 2016. Here, the eight-day LST, view zenith angle, and view time from MYD11A2 and reflectance of band 1 and band 2 from MYD09A1 were used in the following analysis. These data are available from the Level-1 and Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC) at the Goddard Space Flight Center (GSFC) (<https://ladsweb.nascom.nasa.gov/search/>). The eight-day LST product (MYD11A2) at 1 km resolution is the averaged LSTs of the MYD11A1 over eight days, and LST products (MYD11A1) are retrieved from the MODIS thermal and mid-infrared spectral regions using the generalized split window algorithm [35]. MYD09A1 provides MODIS bands 1–7 surface reflectance at 500 m resolution. Each pixel contains the best observation during an eight-day period selected on the basis of high observation coverage, low view zenith angle, absence of clouds or cloud shadow, and aerosol loading [36]. In this paper, MYD09A1 was used to calculate eight-day NDVI. The land cover extracted from the MODIS Land Cover Type Yearly L3 Global 500 m SIN Grid (MCD12Q1, 2000, download from <https://ladsweb.nascom.nasa.gov/search/>) were converted to 1 km resolution using a nearest neighbor resample type to match the resolution of MODIS LST products. The land cover of each meteorological station was extracted in terms of its position, and reclassified into urban, farmland, forest, and grassland.

### 2.2.2. Vegetation Index

Normalized difference vegetation index is the most common remote sensing index used to parameterize vegetation status [16,37,38]. The absorption and reflectivity of the vegetation cover are correlated with their structural properties, such as leaf area index (LAI), fractional vegetation cover (FVC), and their physiological condition [39,40]. The values of NDVI vary between  $-1$  and  $1$ , where the range between  $0.2$  and  $0.9$  is mostly common in continuous vegetation cover [39]. The NDVI is defined as:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad (1)$$

where *NIR* is near-infrared band, *RED* is red band. In MODIS product MYD09A1, band 1 is the red band and band 2 is the near-infrared band.

Applying Equation (1) to MYD09A1, eight days NDVI at 500 m resolution was produced, and then resampled to 1 km resolution.

### 2.2.3. Meteorological Data

Air temperature observations were obtained from 123 meteorological ground stations in the study area (Figure 1). The measurements included daily  $T_{\min}$ ,  $T_{\max}$ , and  $T_{\text{mean}}$ . The time period selected for our study ranged from July 2002 to July 2016 according to the availability of meteorological station record and MODIS Aqua data. The meteorological station records were obtained from the China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn/home.do>). This data was aggregated to eight-day periods using the same compositing method applied to the daily remote sensing LST data (i.e., by averaging the daily valid observations in the eight days period).

### 2.2.4. Auxiliary Data

In addition to MODIS products (LST, view time, view zenith angle, and the number of clear days) and NDVI, some auxiliary variables were used, including latitude, longitude, elevation, distance to the

coast, Julian day, and day length (in hours) (Table 1). These auxiliary variables either have a known impact on  $T_{\text{air}}$  and LST or influence the relationship between  $T_{\text{air}}$  and LST. Latitude and longitude were derived from the location file of meteorological stations. Elevation was obtained from a 90-m resolution digital elevation model (DEM) (downloaded from <http://www.glcf.umd.edu/data/srtm/>), and then converted to be the same resolution and projection as the LST dataset on the basis of the nearest-neighbor method. This method is considered to be efficient and is widely used as a resampling method in related research [41]. Distances to the coast (based on the Global Self-Consistent Hierarchical High-Resolution Shoreline (GSHHS), download from <http://www.soest.hawaii.edu/pwessel/gshhs/index.html>) were calculated in the ESRI ArcGIS 9.3. Both Julian day and day length are proxies for the fraction of solar energy absorption during the day and emission during the night, influencing the diurnal amplitude of  $T_{\text{air}}$  throughout the year. Julian day is the continuous count of days from 1 January every year. The day length was calculated [42]:

$$DayLen = \frac{2 \cos^{-1}[-\tan \delta \tan \varnothing]}{\omega} \quad (2)$$

where  $DayLen$  is the day length in hours,  $\delta$  is the solar declination in radians,  $\varnothing$  is the geographic latitude in radians, and  $\omega$  is the angular velocity of the earth's rotation ( $0.2618 \text{ rad h}^{-1}$ ). The value  $\delta$  is defined as follows:

$$\delta = \sin^{-1} \left\{ 0.4 \sin \left[ \frac{2\pi}{365} (d_n - 82) \right] \right\} \quad (3)$$

the value  $d_n$  is the day number of the year (Julian day).

**Table 1.** Data sources and variables.

Abbreviation	Variable	Source	Explanation
LD, LN	Land Surface Temperature	MODIS	8-day average Land Surface temperature layers derived over the 2002–2016 time period using MYD11A2 product
AD, AN	View Zenith Angle		
TD, TN	View Time		
CD, CN	Clear Days		
NDVI	Normalized difference vegetation index	MODIS	Calculated Normalized Difference Vegetation Index using 8-day average the surface spectral reflectance layers derived over the 2002–2016 time period using MYD09A1 product
D	Julian day	Meteorological data	The continuous count of days was from 1 January to the last day every year
DL	Day length	Meteorological data	Calculated the 8-day average day length using Julian day and latitude in the meteorological data
Lon	Longitude	Meteorological data	The geographical location of meteorological stations was extracted from meteorological metadata
Lat	Latitude		
Ele	Elevation	DEM	The altitude extracted from a 90 m resolution digital elevation model (DEM) according to the location of meteorological stations
Dis	Distance to the coast	GSHHS coastline	Distance to the coast was calculated based on the GSHHS coastline in the ESRI ArcGIS 9.3
$T_{\text{max}}$ , $T_{\text{min}}$ , $T_{\text{mean}}$	Maximum, Minimum and Average Temperature	Meteorological data	Calculated the 8-day average day maximum, minimum and average temperature according to meteorological data

The low quality pixels in LST image were removed according to quality assessment (QA) layer, then outlier pixels were eliminated using the application of several histogram and temperature gradient-based outlier filters [43,44]. All the MODIS data, including LST, view zenith angle, view time, clear days, and NDVI, were extracted from MYD11A2 and MYD09A1 products for the location of the meteorological stations in ESRI ArcGIS. The auxiliary data sets were extracted using the same method. All these data were combined to create a single dataset for each eight-day period and each meteorological station. The combining criteria between meteorological and satellite-retrieved data for the same date are based on a one-to-one relationship, relating eight-day  $T_{\text{mean}}$ ,  $T_{\text{max}}$ , and  $T_{\text{min}}$  from a station with eight-day satellite-retrieved data (daytime LST, nighttime LST, view zenith angle,

view time, clear days, and NDVI) and the auxiliary data (elevation, distance to the coast, Julian day, and day length). Finally, all the records with null values were removed, and the numbers of valid records for daytime LST, nighttime LST, and LST with both day and night are 56,507, 65,164 and 54,691 respectively. The collinearity of independent variables was detected using variance inflation factor (VIF > 10) and pair wise correlation ( $r > 0.7$ ) [45,46].

### 2.3. Methodology

#### 2.3.1. Predictor Selection and Model Building

All Subsets Regression (ASR) and Standardized Regression Coefficients (SRCs) methods were adopted to select the predictors, and then used to build different multiple linear regression models. The ASR method could overcome the limitation of the stepwise regression by inspecting every possible model when all independent variables are taken into account. The best model was selected in terms of the adjusted  $R^2$  from the possible models. The SRCs method could assess the relative importance of predictors, which describe expected change in the response variable holding the other predictors constant. We used the combination of ASR and SRCs to select the final predictors with strong interpretation and prediction power.

#### 2.3.2. Model Calibration and Validation

Cross-validation was used to evaluate the generalizability of a model for predicting air temperature with the LST data. The observations of near surface air temperature were randomly divided into two parts. 70% of observations were use for model calibration, and the rest were used as test dataset for model validation. A set of statistic measures were calculated to assess the accuracy of the predicted air temperature, including root mean square error (RMSE), model efficiency (MEF), coefficient of residual mass (CRM), mean bias error (MBE), and mean absolute error (MAE). The equations of the statistic measures followed as:

$$\text{RMSE} = \left[ \sum_{i=1}^n (y_i - \hat{y}_i)^2 / n \right]^{1/2} \quad (4)$$

$$\text{MEF} = 1 - \left( \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \right) \quad (5)$$

$$\text{CRM} = \left( \frac{\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{y}_i}{\sum_{i=1}^n y_i} \right) \quad (6)$$

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (7)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

where  $n$  is the number of records in validation datasets used in this study,  $\hat{y}_i$  is the estimated variable, and  $y_i$  is the observed variable.

The RMSE is an indicator showing the bias in the mean and in the spatial variance. The value of MEF ranges from  $-1.0$  to  $1.0$ , and would be lower if the estimation method had a poor performance. The CRM indicates overall under- or overestimation of the estimation method, with the value approaching zero for perfect estimation. A positive value of CRM indicates the tendency to underestimate the observed values, and vice versa. The MBE shows the direction of the error bias with zero indicating unbiased estimation of the model. The MAE shows the error magnitude, and lower values mean better performance [41,47].



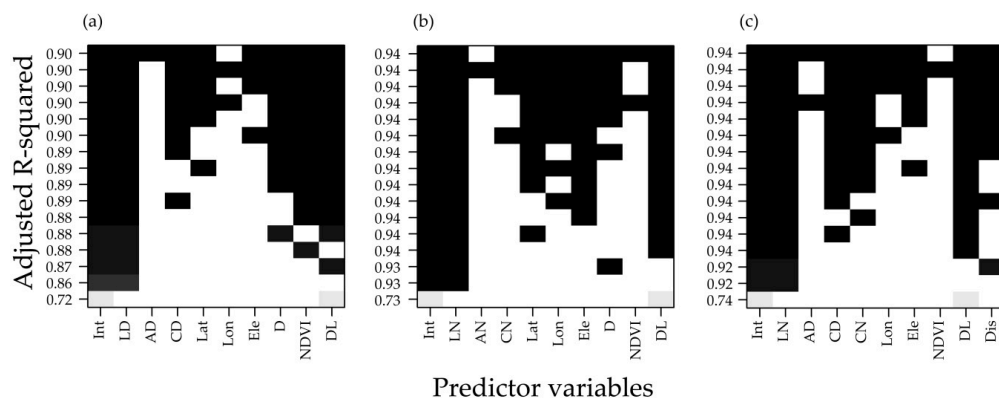
The uncertainty of predicted air temperature in spatial and temporal scales was also calculated. The pairs of observations and predictions were compared in each meteorological station. Model performance was compared across different land cover (static variable) and Julian day (dynamics variable). Moreover, the impact of seasonality on model accuracy was also evaluated.

MODIS data processing was implemented in MODIS Reprojection Tool (MRT, [https://lpdaac.usgs.gov/tools/MODIS\\_reprojection\\_tool](https://lpdaac.usgs.gov/tools/MODIS_reprojection_tool)) and software R using the “sp”, “raster” and “rgdal” packages. All the statistical analyses were performed in R using the basic function and “leap” package (R Development Core Team 2015).

### 3. Results

#### 3.1. Variable and Model Selection

The adjusted  $R^2$  of  $T_{\max}$  models ranged from 0.72 to 0.90 (Figure 2a, starting at the bottom). The models with the following six independent variables were considered optimal with the adjusted  $R^2$  fixed at 0.90. Further increasing independent variables didn't enhance the adjusted  $R^2$ . The day LST, day length, Julian day, NDVI, the number of clear days, and latitude were all determined as important independent variables with SRCs at 0.53, 0.35, 0.11, 0.11, 0.08, and  $-0.05$ , respectively (Figure 3a). The candidate models involving these six independent variables (model 6 to model 7, Table 2) were constructed to predict  $T_{\max}$ . In model selection of  $T_{\min}$ , three independent variables could elevate the adjusted  $R^2$  to 0.94 (Figure 2b). The night LST whose SRC was 0.74 could predict  $T_{\min}$  well. The SRCs of the other two predictors (day length and elevation) were 0.23 and 0.05, respectively (Figure 3b). Models 9, 10, and 11 were the candidate models to predict  $T_{\min}$  (Table 2). The combination of nighttime LST and day length could explain 94% of variation in  $T_{\text{mean}}$ , and the addition of other predictors did not increase the adjusted  $R^2$  (Figure 2c). The nighttime LST, day length, the number of clear days, and the number of clear nights were important independent variables, whose SRCs were 0.72, 0.28, 0.03, and 0.03, respectively (Figure 3c). The candidate models included models 13–15.

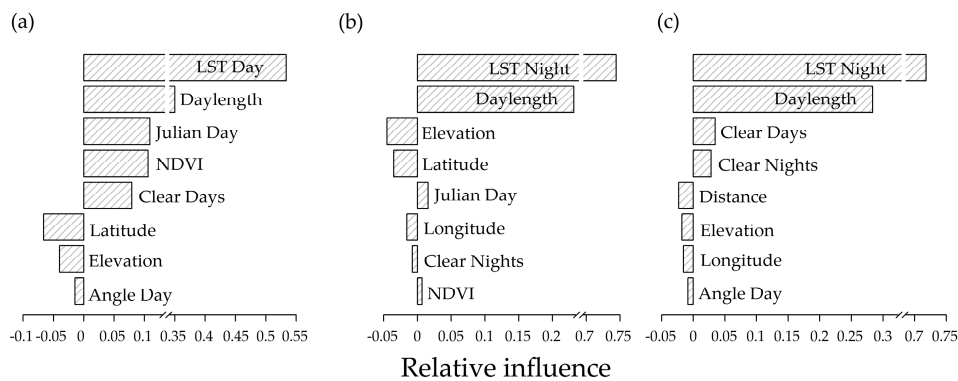


**Figure 2.** Best models for each subset size based on Adjusted R-squared, (a)  $T_{\max}$ , (b)  $T_{\min}$ , and (c)  $T_{\text{mean}}$  variable selection. LD: 8-Day daytime 1 km grid land surface temperature, AD: Average view zenith angle of daytime land surface temperature, CD: the days in clear sky conditions and with valid daytime land surface temperature observation, LN: 8-Day nighttime 1 km grid land surface temperature, AN: Average view zenith angle of nighttime land surface temperature, CN: the days in clear sky conditions and with valid nighttime land surface temperature observation, Lat: Latitude, Lon: Longitude, Ele: Elevation, Dis: The distance to the coastal line, D: Julian day, DL: Mean day length of 8-Day day length, NDVI: Normal difference vegetation index.

**Table 2.** Model equations and validation accuracy for  $T_{\max}$ , model accuracy is given by Adj  $R^2$ , RMSE, MAE, MFE and  $R^2$  change.

	Model Equations	Adj $R^2$	RMSE	MFE	CRM	MBE	MAE	2.5%	97.5%
(1)	Max = $0.8158 \times LD - 0.8104$	0.86	5.21	0.86	0.02	0.30	4.15	-10.94	9.49
(2)	Max = $0.5899 LD + 0.5421 CD + 1.1267 DL + 9.880 NDVI - 15.2347$	0.89	4.75	0.88	0.02	0.24	3.79	-9.52	9.14
(3)	Max = $0.5560 LD + 0.0130 D + 1.5538 DL + 8.2394 NDVI - 20.0725$	0.89	4.77	0.88	0.02	0.22	3.80	-10.01	8.76
(4)	Max = $0.5082 LD - 0.2578 Lat + 0.0152 D + 1.9 DL + 7.631 NDVI - 12.23$	0.89	4.75	0.88	0.01	0.20	3.77	-9.77	8.86
(5)	Max = $0.5350 LD + 0.5312 CD + 0.0128 D + 1.6670 DL + 7.7470 NDVI - 22.6600$	0.89	4.68	0.89	0.02	0.22	3.73	-9.57	8.83
(6)	Max = $0.519 LD + 0.5729 CD - 0.0033 Ele + 0.0134 D + 1.778 DL + 7.711 NDVI - 23.03$	0.90	4.64	0.89	0.02	0.21	3.69	-9.36	8.83
<b>(7)</b>	<b>Max = <math>0.4726 LD + 0.6268 CD - 0.3161 Lat + 0.0155 D + 2.112 DL + 6.913 NDVI - 13.51</math></b>	<b>0.90</b>	<b>4.63</b>	<b>0.89</b>	<b>0.01</b>	<b>0.19</b>	<b>3.69</b>	<b>-9.24</b>	<b>8.89</b>
(8)	Min = $0.9714 LN + 1.6738$	0.93	4.20	0.92	0.43	-0.48	3.22	-8.63	8.06
(9)	Min = $0.8262 LN + 1.0985 DL - 11.8179$	0.94	4.04	0.93	0.39	-0.44	3.12	-8.21	7.91
(10)	Min = $0.7933 LN - 0.1857 Lat + 1.2719 DL - 5.6884$	0.94	4.00	0.93	0.38	-0.43	3.10	-8.05	7.88
<b>(11)</b>	<b>Min = <math>0.8053 LN - 0.0026 Ele + 1.209 DL - 12.42</math></b>	<b>0.94</b>	<b>3.99</b>	<b>0.93</b>	<b>0.38</b>	<b>-0.43</b>	<b>3.09</b>	<b>-8.04</b>	<b>7.84</b>
(12)	Mean = $0.9466 LN + 7.3698$	0.92	4.09	0.91	-0.05	-0.36	3.12	-8.28	8.04
(13)	Mean = $0.7479 LN + 1.5114 DL - 11.2209$	0.94	3.67	0.93	-0.04	-0.31	2.86	-7.25	7.21
(14)	Mean = $0.7343 LN + 1.5545 DL + 0.3246 CD - 12.8276$	0.94	3.63	0.93	-0.04	-0.30	2.82	-7.23	7.16
<b>(15)</b>	<b>Mean = <math>0.7513 LN + 1.5653 DL + 0.3240 CN - 13.2565</math></b>	<b>0.94</b>	<b>3.60</b>	<b>0.93</b>	<b>-0.04</b>	<b>-0.29</b>	<b>2.80</b>	<b>-7.21</b>	<b>7.02</b>

LD: 8-Day daytime 1 km grid land surface temperature, LN: 8-Day nighttime 1 km grid land surface temperature, NDVI: Normal difference vegetation index, DL: Mean day length of 8-Day day length, D: Julian day, Ele: Elevation, Lat: Latitude, CD: the days in clear sky conditions and with valid daytime LST, CN: the days in clear sky conditions and with valid nighttime LST.



**Figure 3.** Independent variable's standardized regression coefficient for estimating air temperature, (a)  $T_{\max}$ , (b)  $T_{\min}$ , and (c)  $T_{\text{mean}}$ .

### 3.2. Model Validation

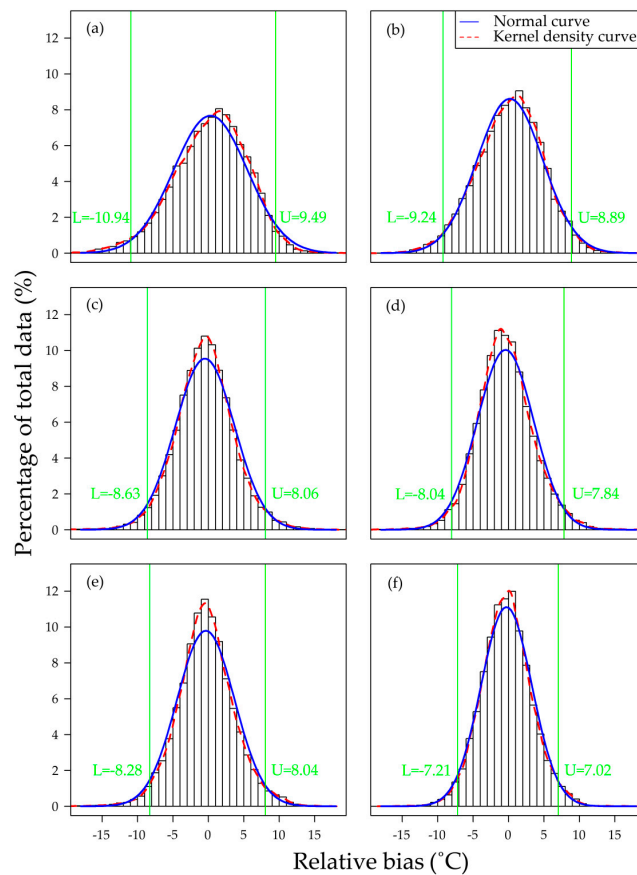
Among the candidate models of  $T_{\max}$ , model 1 (Table 2) that included only one independent variable (daytime LST) explained 86.0% of the variability in  $T_{\max}$ , while the addition of other predictors in models 2 to 7 explained 3–4% more. Among these seven models, models 6 and 7 had lowest RMSE. All  $T_{\max}$  models' CRM and MBE were positive values, indicating models overestimated the  $T_{\max}$ . Models 6 and 7 included the same number of predictors (Table 2), Model 7 was better at fitting and predicting according to the set of statistic measures, and the latitude predictor in Model 7 was more effective than elevation in Model 6 (Figure 3a).

Among candidate models of  $T_{\min}$ , model 11 was optimum. The RMSE (3.99 °C), MAE (3.09 °C), and the confidence interval (−8.04 °C–7.84 °C) of Model 11 were smaller than others. Although Model 10 and Model 11 included the same number of variables, and the MFE, CRM and MBE were the same (Table 2), the SRC of elevation is larger than latitude (Figure 3b). Comparing the statistic measures of  $T_{\text{mean}}$  models, model 15 was optimum for predicting  $T_{\text{mean}}$ . The RMSE of model 15 was 3.60 °C, which was smaller than other  $T_{\text{mean}}$  models. The other statistic measures (i.e., MBE = −0.29 °C, MAE = 2.80 °C) also illustrated that model 15 was the best (Table 2).

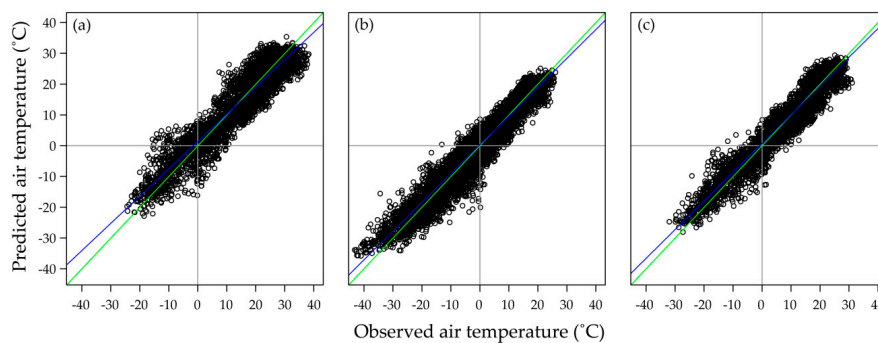
Model 7, model 11, and model 15 which used auxiliary data (e.g., day length, Julian day, and NDVI) were better than model 1, model 8, and model 12, which merely involved daytime or nighttime LST data in predicting  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  respectively. The residuals of these optimal models approximated normality. Autocorrelation of the residuals were found not significant as the auto-correlation function (ACF) was smaller than 0.2 for all lags except at zero. The distribution of relative bias (observed  $T_{\text{air}}$ —predicted  $T_{\text{air}}$ ) was closely centralized to zero and analogous to normal distribution (Figure 4). For all stations, the scatter plots between observed and predicted  $T_{\text{air}}$  distributed agreeably along the diagonal line, indicating the accuracy of model fitting (Figure 5). However, the effectiveness of optimum models was relatively poor under extreme low and high temperature conditions. The performance of models for  $T_{\min}$  and  $T_{\text{mean}}$  was better than that of  $T_{\max}$ .

### 3.3. Spatial and Temporal Performance

The optimal models were analyzed in each meteorological station to assess their effectiveness. The results showed that for the  $T_{\max}$  model about 80% of the stations had an RMSE lower than 5 °C (Figure 6a), MAE lower than 4 °C, and MEF higher than 0.85 (Figure 6b). The distribution of MEF was clearly skewed toward higher values. The stations' RMSE and MEF ranged from 3.41 °C to 5.94 °C, and 0.69 to 0.93 respectively, which were very similar to the overall results (RMSE = 4.63 °C and MFE = 0.89). The  $T_{\max}$  model MBE in most stations ranged from −1.5 °C to 1.5 °C, gathering around zero. These statistic measures illustrated that model 7 was able to predict  $T_{\max}$  accurately. The spatial performance of optimal models predicting  $T_{\min}$  (model 11) and  $T_{\text{mean}}$  (model 15) was better than model 7, because the distribution of their RMSE (Figure 6e,i) and MAE (Figure 6h,l) tended to be even, and MEF more skewed to 1 (Figure 6f,j).

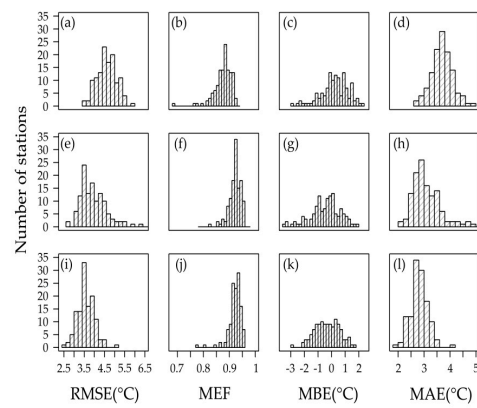


**Figure 4.** (a–f) Relative bias distribution considering all eight days  $T_{max}$  ((a) model 1, (b) model 7),  $T_{min}$  ((c) Model 8, (d) model 11) and  $T_{mean}$  ((e) model 12, (f) model 15). The blue curves are normal density distribution curves, red curves are kernel density distribution curves, and green lines show the lower endpoint and upper endpoint of the relative bias 95% confidence interval.

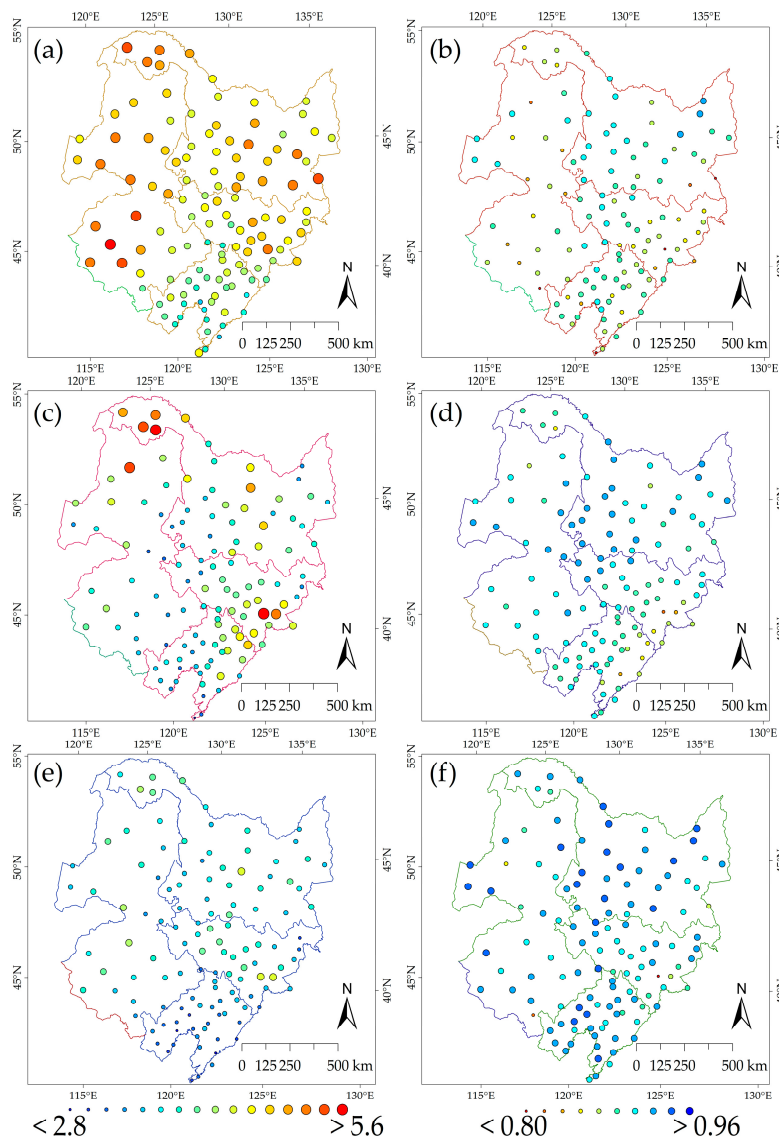


**Figure 5.** Scatter plots between predicted air temperature (°C) and observed air temperature (°C) of (a)  $T_{max}$ , (b)  $T_{min}$ , and (c)  $T_{mean}$ . The blue line was the relationship between predicted air temperature and observed air temperature, the green line is 1:1 line, which represents the relationship of observed air temperature = predicted air temperature.

The spatial distribution of RMSE (model 7, model 11, and model 15) showed that the models generally performed worse for the stations located in the mountain and plateau areas than those in the plains (Figure 7a,c,e). The spatial distribution of MFE was similar, which performed well even in plain areas (Figure 7b,d,f).



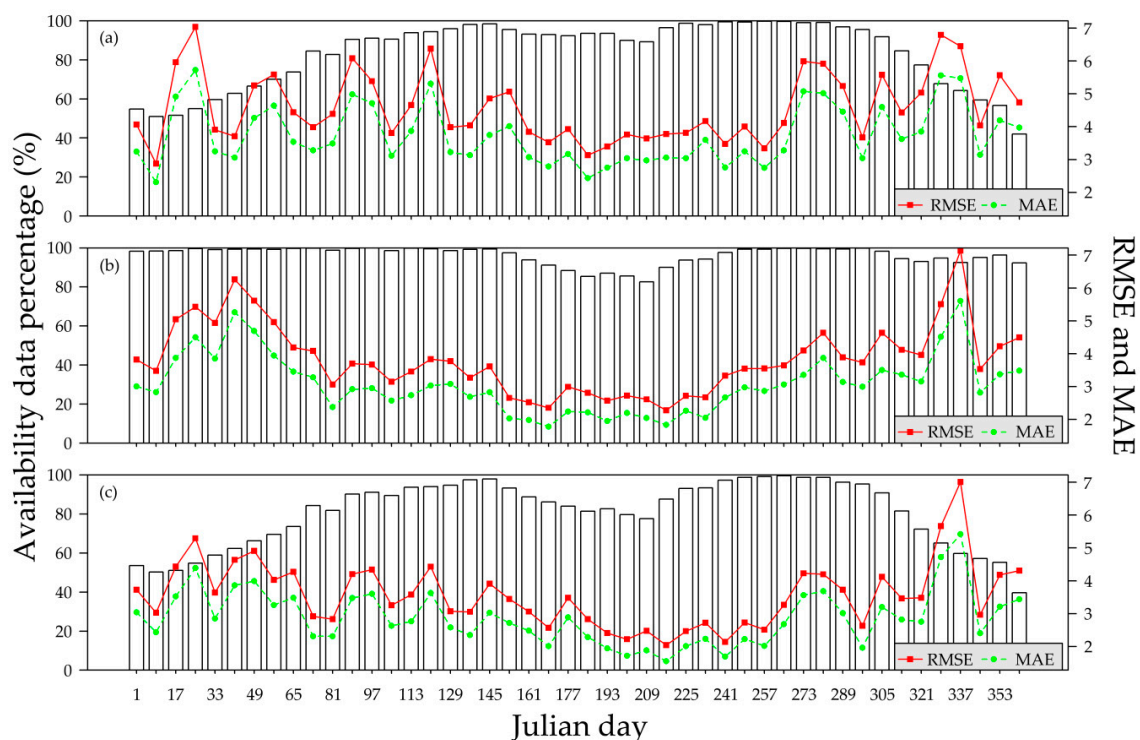
**Figure 6.** (a–l)  $T_{max}$ ,  $T_{min}$ , and  $T_{mean}$  model performance distribution for all the meteorological stations used: RMSE (a,e,i); MEF (b,f,j); MBE (c,g,k) and MAE (d,h,l).



**Figure 7.** (a–f) Spatial distribution of the optimal models RMSE: (a)  $T_{max}$ , (c)  $T_{min}$ , (e)  $T_{mean}$  and MEF: (b)  $T_{max}$ , (d)  $T_{min}$ , (f)  $T_{mean}$  for all the meteorological stations.

The models performed well for stations located in urban areas and farmland, and performed poorly in forest and grassland areas (Table 3). The RMSEs of  $T_{\max}$  model 7 in urban, farmland, forest and grassland areas were 4.62 °C, 4.55 °C, 4.77 °C, and 5.09 °C, and MFEs were 0.89, 0.89, 0.84, and 0.88, respectively. For  $T_{\min}$  model 11, the RMSEs in urban, farmland, forest, and grassland areas were 3.97 °C, 3.82 °C, 4.81 °C, and 4.33 °C, respectively, and MFEs were 0.93, 0.93, 0.90, and 0.92. The RMSEs for  $T_{\text{mean}}$  model 15 in urban, farmland, forest, and grassland areas were 3.63 °C, 3.53 °C, 3.76 °C, and 3.77 °C, respectively, and MFEs were 0.93, 0.93, 0.90, and 0.93. Model 11 and model 15, predicting  $T_{\min}$  and  $T_{\text{mean}}$ , were better than model 7 at predicting  $T_{\max}$  across all land cover types.

The temporal distributions of model performance were evaluated at season, month, and 8-day scales. A set of statistic measures were calculated for all predicted/observed pairs within a specific season (Table 4), month (Table 5), and 8-day (Figure 8) periods, comprising all the years. Model 7 for  $T_{\max}$ , model 11 for  $T_{\min}$ , and model 15 for  $T_{\text{mean}}$  performed well at predicting air temperature in summer according to RMSE (3.82 °C, 2.71 °C, and 2.70 °C, respectively), while they underperformed in terms of MFE (−0.04, 0.62, and 0.45, respectively). These showed that uncertainty existed in summer. Based on the synthesis of all statistic measures, all the models performed better in spring and worse in winter. The model 11 for  $T_{\min}$  and model 15 for  $T_{\text{mean}}$  performed better than model 7 for  $T_{\max}$ , which was similar to the model performance at the spatial scale. Model performance at the monthly scale was similar to that at the seasonal scale. At the eight-day period scale, data availability was poor in winter and late autumn, during which the RMSE and MAE of model 7 for  $T_{\max}$  were high (Figure 8a). An similar pattern was also found in model 15 for  $T_{\text{mean}}$ , and poor data availability was accompanied by higher RMSE and MAE (Figure 8c). Although the data availability used to predict  $T_{\min}$  was high in winter and late autumn, the performance of model 11 was poor with high RMSE and MAE (Figure 8b).



**Figure 8.** Relations between the temporal distribution of data availability and model performance, (a)  $T_{\max}$ , (b)  $T_{\min}$ , and (c)  $T_{\text{mean}}$ . Data availability is the percentage (%) of the valid data used in model calibration/validation.

**Table 3.** The performance of optimal models for  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  for different land cover types. Model accuracy is given by RMSE, MEF, CRM, MAE, and MBE.

Land Cover	$T_{\max}$					$T_{\min}$					$T_{\text{mean}}$				
	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE
Urban	4.62	0.89	0.03	0.44	3.71	3.97	0.93	0.75	−0.65	3.09	3.63	0.93	−0.09	−0.68	2.84
Farmland	4.55	0.89	0.02	0.27	3.61	3.82	0.93	−0.23	0.08	2.96	3.53	0.93	0.03	0.22	2.72
Forest	4.77	0.84	−0.04	−0.57	3.73	4.81	0.90	0.34	−0.92	3.61	3.76	0.90	−0.12	−1.01	2.88
Grass	5.09	0.88	−0.06	−0.67	4.06	4.33	0.92	0.11	−0.54	3.41	3.77	0.93	−0.09	−0.43	2.99

**Table 4.** The performance of optimal models for  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  in different seasons. Model accuracy is given by RMSE, MEF, CRM, MAE, and MBE.

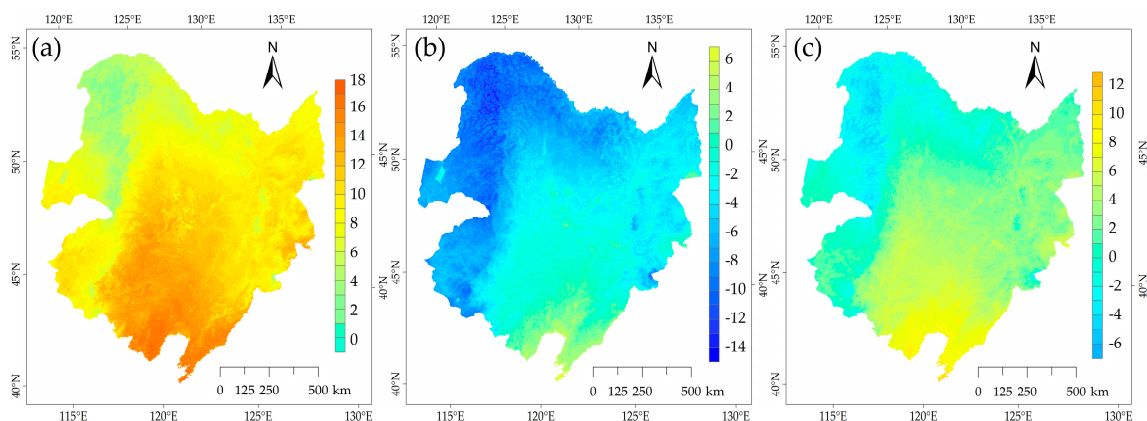
Seasons	$T_{\max}$					$T_{\min}$					$T_{\text{mean}}$				
	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE
Spring	4.80	0.75	0.01	0.09	3.85	3.64	0.85	−0.30	−0.55	2.91	3.67	0.83	0.02	0.21	2.92
Summer	3.82	−0.04	0.01	0.20	3.06	2.71	0.62	−0.04	−0.69	2.11	2.70	0.45	−0.05	−1.06	2.13
Autumn	4.96	0.78	0.05	0.58	4.00	4.15	0.82	0.11	−0.21	3.34	3.66	0.85	0.04	0.22	2.91
Winter	5.10	0.57	0.04	−0.25	4.04	5.02	0.66	0.01	−0.27	3.93	4.51	0.68	0.05	−0.61	3.51

**Table 5.** The performance of optimal models for  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  in different months. Model accuracy is given by RMSE, MEF, CRM, MAE, and MBE.

Month	$T_{\max}$					$T_{\min}$					$T_{\text{mean}}$				
	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE	RMSE	MEF	CRM	MBE	MAE
January	5.23	0.53	−0.15	1.40	4.04	4.52	0.72	0.01	−0.21	3.55	4.21	0.69	0.07	−1.05	3.36
February	4.76	0.56	−0.06	0.21	3.86	5.47	0.62	0.05	−0.77	4.43	4.34	0.68	0.05	−0.55	3.50
March	4.81	0.57	−0.05	−0.35	3.84	3.78	0.71	0.18	−1.16	3.01	3.59	0.68	−3.22	−1.01	2.87
April	4.60	0.50	0.07	1.36	3.84	3.42	0.62	−0.10	−0.27	2.73	3.69	0.54	0.15	1.60	2.94
May	4.95	0.34	−0.03	−0.61	3.88	3.68	0.61	−0.01	−0.14	2.95	3.72	0.54	0.01	0.16	2.94
June	4.13	−0.28	−0.08	−2.00	3.25	2.64	0.55	−0.07	−1.01	2.01	3.16	0.10	−0.10	−1.97	2.52
July	3.48	−0.15	0.03	0.87	2.79	2.69	0.44	−0.02	−0.35	2.11	2.51	0.26	−0.03	−0.80	1.97
August	3.81	0.07	0.07	1.70	3.11	2.80	0.62	−0.05	−0.67	2.22	2.36	0.57	−0.02	−0.38	1.87
September	4.46	0.30	0.09	1.80	3.58	3.72	0.56	−0.03	−0.21	3.06	3.25	0.59	0.06	0.81	2.62
October	5.05	0.36	0.12	1.31	4.18	4.11	0.53	0.12	−0.24	3.34	3.59	0.58	0.12	0.50	2.90
November	5.48	0.36	−23.03	−1.68	4.38	4.60	0.64	0.02	−0.19	3.64	4.23	0.62	0.16	−0.86	3.33
December	5.32	0.47	0.29	−2.21	4.23	5.02	0.60	−0.01	0.20	3.81	4.92	0.58	0.02	−0.28	3.66

### 3.4. Patterns of Estimated Annual Air Temperature

The annual spatial  $T_{\text{air}}$  pattern was consistent with prior knowledge of this region solely derived from the meteorological stations data. Predicted annual  $T_{\text{max}}$ ,  $T_{\text{min}}$  and  $T_{\text{mean}}$  exhibited marked spatial gradients decreasing from south to north (Figure 9). The spatial gradients of predicted  $T_{\text{air}}$  varied not only along the latitude, but also with the longitude. Estimated annual  $T_{\text{air}}$  presented obvious patterns with higher temperature in the center of this region, and lower in both east and west sides. Meanwhile, the patterns of estimated  $T_{\text{air}}$  were influenced by the distance to the coastline. Air temperature was relatively lower in the west of this region further away from the coastline than the east, and the lowest air temperature was found in the northwest.



**Figure 9.** Estimated annual temperature of eight-day (a)  $T_{\text{max}}$ , (b)  $T_{\text{min}}$ , and (c)  $T_{\text{mean}}$  integrated over the entire study period (2002–2016).

## 4. Discussion

Our study showed that it was feasible to estimate  $T_{\text{air}}$  by simply building a linear model between  $T_{\text{air}}$  and LST. However, this model performed poorly in extreme temperatures. The variation of model performance mainly depended on the spatial heterogeneity (i.e., topography, land cover) and time cycle (i.e., seasonal alteration).

### 4.1. Spatial Uncertainties of Models

Complex topography could affect the relationship between  $T_{\text{air}}$  and LST data, and give rise to the model uncertainty in predicting  $T_{\text{air}}$ . Cooler conditions prevail in regions with high altitude due to a steep lapse rate under the influence of cold-air drainage [48]. In stands with similar altitude, south-facing slopes tend to receive greater solar radiation than shaded areas, and have higher temperature. In the meantime, cold-air pooling or temperature inversions occur frequently in mountain regions, especially under weather conditions with calm wind and clear sky, which would lead to model uncertainties in predicting  $T_{\text{air}}$  from LST data [49].

Physiographic factors also influenced snow accumulation and melt. Yan et al. [50] found that  $T_{\text{air}}$  was lower if the land was covered with snow in winter. Additionally, the linear correlation between  $T_{\text{air}}$  and LST was weak during snow-covered periods in high mountains [22]. This is because snow cover could enhance the visible reflectance, and reduce the solar radiation absorption of ground surface. Snow cover also strongly affected vegetation greening through influences on the biotic and abiotic environment. Air temperature would be reduced if the snow cover period lengthened [49]. Shifts in vegetation greening could also change soil moisture content due to snow melt, and alter the land surface temperature. It would result in uncertainty in predicting  $T_{\text{air}}$  through LST data irrespective of physiographic factors. Thus, the poor model performance in our study in extreme low temperature during winter could be attributed to snow cover (Figure 5).



Land cover plays an important role in influencing the relationship between  $T_{\text{air}}$  and LST. Our results showed that the effectiveness of optimal models in predicting  $T_{\text{air}}$  varied in different land cover types (Table 3). This variation could be introduced by the specific heat capacities of different land covers. Air temperature mainly depended on the heat transfer process which was strongly influenced by the local radiation budget [22]. In general, barren land has lower heat capacity than forest. Hence, air was heated much quicker over barren land than forest. Vegetation could also alter latent heat flux, such as enhancing or reducing transpiration [51,52], and cool the  $T_{\text{air}}$  in forests [53,54]. The cooling effect was not explicitly incorporated in our models due to uneven distribution of meteorological stations across different vegetation types. The meteorological station was too scarce in some vegetation types. Thus, it was hard to take consider of vegetation type in our models. However, land cover also affected land surface albedo, and a slight warming was observed in winter due to dense canopy reducing the albedo of snow cover [49]. Thus, the influence of land cover on estimating  $T_{\text{air}}$  was conditional and time dependent.

#### 4.2. Temporal Uncertainties of Models

The temporal variation also had strong influence on the estimation of  $T_{\text{air}}$  from LST data. Seasonal vegetation greening changed the physical properties of land surface, such as vegetation coverage, snow cover, and soil moisture. Our study area is mainly covered with lush crops and pasture, as well as frondent forests during growing season; and it would turn into bare land, leafless forests, or snow cover afterwards (Figure 1). The variation of land cover due to seasonal vegetation greening enhanced the spatial heterogeneity of LST data, and increased the temporal uncertainties in predicting air temperature [55]. It is important to consider temporal variation caused by vegetation greening on the estimation of  $T_{\text{air}}$  from LST over complex terrain [22].

In addition to vegetation greening, air masses and aerosol affected the temporal variation in the performance of  $T_{\text{air}}$  models as well. This region is characterized by a typical temperate continental monsoon climate, and controlled by the Siberian high pressure system in winter and the Hawaiian high pressure system in summer [56,57]. The basic mechanism for estimating  $T_{\text{air}}$  with LST relies on the intense heat exchange between land surface and the atmosphere, yet the air masses strike the local energy balance due to coming from different sources [21,51]. Dry and cold air masses come from Siberia in winter; conversely wet and cool air masses come from the sea surface of the Hawaiian system in summer. Cool air moving over a warmer land surface is heated discriminately from below, due to the fact that moisture content is different in such air masses. Moisture content of air is an important control on thermal conduction between air and surface [22,58]. A time lag existed for establishing the new balance between air and land surface [51]. Both the moisture content and time lag could result in uncertainties in estimating  $T_{\text{air}}$  from LST data.

Aerosol varies with season, and is an important component of the air that essentially determines the energy relationship between  $T_{\text{air}}$  and land surface [21]. In this region, there are more aerosols in winter than summer due to burning fossil fuel for heating. Aerosols can absorb short-wave solar radiation to warm the atmosphere, while they usually scatter and reflect short-wave radiations to cool the ground surface. The residence time of aerosols in the atmosphere is regional and seasonal, which affects the radiative forcing of aerosols [21]. Huang et al. [59] also indicated that high concentration of aerosol retained the heat by acting, to some extent, as insulators that trapped outgoing radiation.

Our results showed that the performance of models estimating  $T_{\text{min}}$  and  $T_{\text{mean}}$  using nighttime LST were better than that estimating  $T_{\text{max}}$  using daytime LST (Figure 5). Nighttime LST was stable and affected by fewer factors compared with daytime LST due to the absence of solar radiation [51,59]. Generally, the land surface warms faster than ambient air in the day, and cools more rapidly during the night. The warmer land surface heated the ambient air to stimulate the turbulence of ground atmosphere in the day, while the turbulence turned weak on account of lacking solar radiation to heat land surface in the night. An isothermal and homogeneous surface was established over the ground surface in the night, and improved the estimation accuracy of  $T_{\text{air}}$ . Shen and Leptoukh [55]

indicated that the relationship between  $T_{\max}$  and daytime LST depended significantly on land cover, but  $T_{\min}$  and nighttime LST had little dependence on land cover, especially during the summer time. The influences of direct solar illumination and land cover resulted in more complex interactions between ambient air and land surface. That is why the daytime LST was not as good as the nighttime LST in predicting air temperature.

#### 4.3. Spatial Variations of Annual Air Temperature

The spatial patterns of all the predicted annual  $T_{\max}$ ,  $T_{\min}$  and  $T_{\text{mean}}$  were relatively higher in the south, and lower in the north (Figure 9). In the south, the Earth's surface could receive more solar radiation to heat the ambient air, while the energy was relatively lower in the north. In this region, the mountains covered by forests are located in the north and both sides of this region, while the cities and farm fields are located in the center (Figure 1). Air temperature becomes low where the altitude increases, and the forests are also able to cool the air temperature. Meanwhile, the artificial buildings in the plains could absorb more solar radiation to heat the ambient air. The south and east parts are closer to the coastline, so the warmer vapor coming from the sea is able to heat the ambient air through latent heat fluxes.

### 5. Conclusions

The results of this study confirm that the eight-day average of  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  can be accurately estimated using LST and auxiliary data. The simple multiple linear regression models were established for  $T_{\max}$ ,  $T_{\min}$ , and  $T_{\text{mean}}$  using different predictors for each value, but LST and day length were both relatively important predictors in all models. The performance of optimal models was efficient to estimate  $T_{\max}$  (RMSE = 4.63 °C, MFE = 0.89),  $T_{\min}$  (RMSE = 3.99 °C, MFE = 0.93) and  $T_{\text{mean}}$  (RMSE = 3.60 °C, MFE = 0.93) for all meteorological stations. The performance was better for models predicting  $T_{\min}$  and  $T_{\text{mean}}$  than that of  $T_{\max}$ . Nighttime LST data was better in predicting  $T_{\min}$  and  $T_{\text{mean}}$  than daytime LST data in predicting  $T_{\max}$ . Nighttime or combining daytime and nighttime LST should be taken into account to improve the performance of the model for  $T_{\max}$ .

The model performance varied across different spatial and temporal scales. The models performed well in the plains, and relatively poorly in the complex terrain of the mountains. Land covers had noticeable influences on estimating air temperature, and seasonal changes of land covers could result in temporal variations of model performance. The RMSE values for the models during the summer were slightly better than the other seasons. The changes in physical and chemical composition of near-surface atmosphere may also contribute to the models' temporal variations. More predictors should be considered for the purpose of improving the models estimating air temperature. It is possible that non-linear models or advanced ensemble models such as boosted regression trees may produce better estimation for  $T_{\text{air}}$ .

**Acknowledgments:** This work was funded by the National Key Research and Development Program of China (2016YFA0600804), the National Natural Science Foundation of China (41222004, 31270511, 31470517 & 41501200), CAS Interdisciplinary Innovation Team, and the CAS Pioneer Hundred Talents Program. We also acknowledge the measured air temperature data providers, namely, China Meteorological Data Sharing Service System. We acknowledge the MODIS mission scientists and associated NASA personnel for the production of the data used in this research effort.

**Author Contributions:** Y.Z.Y. led the acquisition and analysis of the LST and meteorological data and the writing of the manuscript. W.H.C. led the acquisition of the LST and meteorological data and contributed to the writing of the manuscript. J.Y. designed the project, oversaw the analysis, and wrote the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Peón, J.; Carmen, R.; Javier, F.C. Improvements in the Estimation of Daily Minimum Air Temperature in Peninsular Spain Using MODIS Land Surface Temperature. *Int. J. Remote Sens.* **2014**, *35*, 5148–5166. [[CrossRef](#)]
2. Prihodko, L.; Goward, S.N. Estimation of Air Temperature from Remotely Sensed Surface Observations. *Remote Sens. Environ.* **1997**, *60*, 335–346. [[CrossRef](#)]
3. Stisen, S.; Sandholt, I.; Nørgaard, A.; Fensholt, R.; Eklundh, L. Estimation of Diurnal Air Temperature Using Msg Seviri Data in West Africa. *Remote Sens. Environ.* **2007**, *110*, 262–274. [[CrossRef](#)]
4. Shamir, E.; Georgakakos, K.P. MODIS Land Surface Temperature as an Index of Surface Air Temperature for Operational Snowpack Estimation. *Remote Sens. Environ.* **2014**, *152*, 83–98. [[CrossRef](#)]
5. Benali, A.; Carvalho, A.C.; Nunes, J.P.; Carvalhais, N.; Santos, A. Estimating Air Surface Temperature in Portugal Using MODIS Lst Data. *Remote Sens. Environ.* **2012**, *124*, 108–121. [[CrossRef](#)]
6. Westerling, A.L.; Hidalgo, H.G.; Cayan, D.R.; Swetnam, T.W. Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. *Science* **2006**, *313*, 940–943. [[CrossRef](#)] [[PubMed](#)]
7. Chen, Y.; Randerson, J.T.; Morton, D.C.; DeFries, R.S.; Collatz, G.J.; Kasibhatla, P.S.; Giglio, L.; Jin, Y.; Marlier, M.E. Forecasting Fire Season Severity in South America Using Sea Surface Temperature Anomalies. *Science* **2011**, *334*, 787–791. [[CrossRef](#)] [[PubMed](#)]
8. Manzo-Delgado, L.; Sánchez-Colón, S.; Álvarez, R. Assessment of Seasonal Forest Fire Risk Using Noaa-Avrr: A Case Study in Central Mexico. *Int. J. Remote Sens.* **2009**, *30*, 4991–5013. [[CrossRef](#)]
9. Reich, P.B.; Luo, Y.; Bradford, J.B.; Poorter, H.; Perry, C.H.; Oleksyn, J. Temperature Drives Global Patterns in Forest Biomass Distribution in Leaves, Stems, and Roots. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13721–13726. [[CrossRef](#)] [[PubMed](#)]
10. Ruane, A.C.; McDermid, S.; Rosenzweig, C.; Baigorria, G.A.; Jones, J.W.; Romero, C.C.; Cecil, L.D. Carbon-Temperature-Water Change Analysis for Peanut Production under Climate Change: A Prototype for the Agmip Coordinated Climate-Crop Modeling Project (C3mp). *Glob. Chang. Biol.* **2014**, *20*, 394–407. [[CrossRef](#)] [[PubMed](#)]
11. Rosenzweig, C.; Elliott, J.; Deryng, D.; Ruane, A.C.; Muller, C.; Arneth, A.; Boote, K.J.; Folberth, C.; Glotter, M.; Khabarov, N.; et al. Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3268–3273. [[CrossRef](#)] [[PubMed](#)]
12. Oyler, J.W.; Ballantyne, A.; Jencso, K.; Sweet, M.; Running, S.W. Creating a Topoclimatic Daily Air Temperature Dataset for the Conterminous United States Using Homogenized Station Data and Remotely Sensed Land Skin Temperature. *Int. J. Climatol.* **2015**, *35*, 2258–2279. [[CrossRef](#)]
13. Beier, C.M.; Signell, S.A.; Luttman, A.; DeGaetano, A.T. High-Resolution Climate Change Mapping with Gridded Historical Climate Products. *Landsc. Ecol.* **2012**, *27*, 327–342. [[CrossRef](#)]
14. Urban, M.; Eberle, J.; Hüttich, C.; Schmullius, C.; Herold, M. Comparison of Satellite-Derived Land Surface Temperature and Air Temperature from Meteorological Stations on the Pan-Arctic Scale. *Remote Sens.* **2013**, *5*, 2348–2367. [[CrossRef](#)]
15. Vancutsem, C.; Ceccato, P.; Dinku, T.; Connor, S.J. Evaluation of MODIS Land Surface Temperature Data to Estimate Air Temperature in Different Ecosystems over Africa. *Remote Sens. Environ.* **2010**, *114*, 449–465. [[CrossRef](#)]
16. Zhu, W.; Lú, A.; Jia, S. Estimation of Daily Maximum and Minimum Air Temperature Using MODIS Land Surface Temperature Products. *Remote Sens. Environ.* **2013**, *130*, 62–73. [[CrossRef](#)]
17. Kloog, I.; Nordio, F.; Coull, B.A.; Schwartz, J. Predicting Spatiotemporal Mean Air Temperature Using MODIS Satellite Surface Temperature Measurements across the Northeastern USA. *Remote Sens. Environ.* **2014**, *150*, 132–139. [[CrossRef](#)]
18. Williamson, S.N.; Hik, D.S.; Gamon, J.A.; Kavanaugh, J.L.; Flowers, G.E. Estimating Temperature Fields from MODIS Land Surface Temperature and Air Temperature Observations in a Sub-Arctic Environment. *Remote Sens.* **2014**, *6*, 946–963. [[CrossRef](#)]
19. Vincent, L.A.; Mekis, E. Changes in Daily and Extreme Temperature and Precipitation Indices for Canada over the Twentieth Century. *Atmos. Ocean* **2006**, *44*, 177–193. [[CrossRef](#)]
20. Carlson, T. An Overview of the “Triangle Method” for Estimating Surface Evapotranspiration and Soil Moisture from Satellite Imagery. *Sensors* **2007**, *7*, 1612–1629. [[CrossRef](#)]

21. Zhang, W.; Huang, Y.; Yu, Y.; Sun, W. Empirical Models for Estimating Daily Maximum, Minimum and Mean Air Temperatures with MODIS Land Surface Temperatures. *Int. J. Remote Sens.* **2011**, *32*, 9415–9440. [[CrossRef](#)]
22. Lin, X.; Zhang, W.; Huang, Y.; Sun, W.; Han, P.; Yu, L.; Sun, F. Empirical Estimation of near-Surface Air Temperature in China from MODIS Lst Data by Considering Physiographic Features. *Remote Sens.* **2016**, *8*. [[CrossRef](#)]
23. Hanna, M.; Katurji, M.; Appelhans, T.; Müller, M.U.; Nauss, T.; Roudier, P.; Zawar-Reza, P. Mapping Daily Air Temperature for Antarctica Based on MODIS LST. *Remote Sens.* **2016**, *8*. [[CrossRef](#)]
24. Zhang, H.B.; Zhang, F.; Ye, M.; Che, T.; Zhang, G.Q. Estimating Daily Air Temperatures over the Tibetan Plateau by Dynamically Integrating MODIS Lst Data. *J. Geophys. Res. Atmos.* **2016**, *121*, 11425–11441. [[CrossRef](#)]
25. Tao, J.; Zhang, Y.; Zhu, J.; Jiang, Y.; Zhang, X.; Zhang, T.; Xi, Y. Elevation-Dependent Temperature Change in the Qinghai–Xizang Plateau Grassland During the Past Decade. *Theor. Appl. Climatol.* **2013**, *117*, 61–71. [[CrossRef](#)]
26. Nieto, H.; Sandholt, I.; Aguado, I.; Chuvieco, E.; Stisen, S. Air Temperature Estimation with Msg-Seviri Data: Calibration and Validation of the TVX Algorithm for the Iberian Peninsula. *Remote Sens. Environ.* **2011**, *115*, 107–116. [[CrossRef](#)]
27. Fang, J.; Chen, A.; Peng, C.; Zhao, S.; Ci, L. Changes in Forest Biomass Carbon Storage in China between 1949 and 1998. *Science* **2001**, *292*, 2320–2322. [[CrossRef](#)] [[PubMed](#)]
28. Piao, S.; Fang, J.; Ciais, P.; Peylin, P.; Huang, Y.; Sitch, S.; Wang, T. The Carbon Balance of Terrestrial Ecosystems in China. *Nature* **2009**, *458*, 1009–1013. [[CrossRef](#)] [[PubMed](#)]
29. Liu, Z.; Yang, J.; Chang, Y.; Weisberg, P.J.; He, H.S. Spatial Patterns and Drivers of Fire Occurrence and Its Future Trend under Climate Change in a Boreal Forest of Northeast China. *Glob. Chang. Biol.* **2012**, *18*, 2041–2056. [[CrossRef](#)]
30. Cai, W.; Yang, J.; Liu, Z.; Hu, Y.; Weisberg, P.J. Post-Fire Tree Recruitment of a Boreal Larch Forest in Northeast China. *For. Ecol. Manag.* **2013**, *307*, 20–29. [[CrossRef](#)]
31. Welch, J.R.; Vincent, J.R.; Auffhammer, M.; Moya, P.F.; Dobermann, A.; Dawe, D. Rice Yields in Tropical/Subtropical Asia Exhibit Large but Opposing Sensitivities to Minimum and Maximum Temperatures. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14562–14567. [[CrossRef](#)] [[PubMed](#)]
32. van Wart, J.; Grassini, P.; Cassman, K.G. Impact of Derived Global Weather Data on Simulated Crop Yields. *Glob. Chang. Biol.* **2013**, *19*, 3822–3834. [[CrossRef](#)] [[PubMed](#)]
33. Mildrexler, D.J.; Zhao, M.; Running, S.W. A Global Comparison between Station Air Temperatures and MODIS Land Surface Temperatures Reveals the Cooling Role of Forests. *J. Geophys. Res.* **2011**, *116*. [[CrossRef](#)]
34. Zhang, P.; Bounoua, L.; Imhoff, M.L.; Wolfe, R.E.; Thome, K. Comparison of MODIS Land Surface Temperature and Air Temperature over the Continental USA Meteorological Stations. *Can. J. Remote Sens.* **2014**, *40*, 110–122.
35. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.L. Validation of the Land-Surface Temperature Products Retrieved from Terra Moderate Resolution Imaging Spectroradiometer Data. *Remote Sens. Environ.* **2002**, *83*, 163–180. [[CrossRef](#)]
36. MODIS Land Surface Temperature Products User’s Guide. Available online: <http://www.ices.ucs.edu/MODIS/LstUsrGuide/usrguide.html> (accessed on December 2013).
37. Stow, D.A.; Hope, A.; McGuire, D.; Verbyla, D.; Gamon, J.; Huemmrich, F.; Houston, S.; Racine, C.; Sturm, M.; Tape, K.; et al. Remote Sensing of Vegetation and Land-Cover Change in Arctic Tundra Ecosystems. *Remote Sens. Environ.* **2004**, *89*, 281–308. [[CrossRef](#)]
38. Reynolds, M.; Comiso, J.; Walker, D.; Verbyla, D. Relationship between Satellite-Derived Land Surface Temperatures, Arctic Vegetation Types, and NDVI. *Remote Sens. Environ.* **2008**, *112*, 1884–1894. [[CrossRef](#)]
39. Bustos, E.; Meza, F.J. A Method to Estimate Maximum and Minimum Air Temperature Using MODIS Surface Temperature and Vegetation Data: Application to the Maipo Basin, Chile. *Theor. Appl. Climatol.* **2014**. [[CrossRef](#)]
40. Reynolds, M.K.; Walker, D.A.; Maier, H.A. NDVI Patterns and Phytomass Distribution in the Circumpolar Arctic. *Remote Sens. Environ.* **2006**, *102*, 271–281. [[CrossRef](#)]
41. Zheng, X.; Zhu, J.; Yan, Q. Monthly Air Temperatures over Northern China Estimated by Integrating MODIS Data with Gis Techniques. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 1987–2000. [[CrossRef](#)]

42. Neitsch, S.L.; Arnold, J.G.; Kiniry, J.R.; Williams, J.R.; King, K.W. *Soil and Water Assessment Tool: Theoretical Documentation*; Version 2009; Texas Water Resources Institute: College Station, TX, USA, 2011; pp. 30–32.
43. Neteler, M. Estimating Daily Land Surface Temperature in Mountainous Environments by Reconstructed MODIS LST Data. *Remote Sens.* **2010**, *2*, 333–351. [[CrossRef](#)]
44. Ke, L.H.; Wang, Z.X.; Song, C.Q.; Lu, Z.Q. Reconstruction of MODIS Land Surface Temperature in Northeast Qinghai-Xizang Plateau and LST Comparison with Air Temperature. *Plateau Meteorol.* **2011**, *30*, 277–287.
45. Zurr, A.F.; Ieno, E.N.; Elphick, C.S. A Protocol for Data Exploration to Avoid Common Statistical Problems. *Methods Ecol. Evol.* **2010**, *1*, 3–14. [[CrossRef](#)]
46. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 027–046. [[CrossRef](#)]
47. Willmott, C.J.; Matsuura, K. Advantages of the Mean Absolute Error (Mae) over the Root Mean Square Error (Rmse) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
48. Dobrowski, S.Z.; Abatzoglou, J.T.; Greenberg, J.A.; Schladow, S.G. How Much Influence Does Landscape-Scale Physiography Have on Air Temperature in a Mountain Environment? *Agric. For. Meteorol.* **2009**, *149*, 1751–1758. [[CrossRef](#)]
49. Van De Kerchove, R.; Lhermitte, S.; Veraverbeke, S.; Goossens, R. Spatio-Temporal Variability in Remotely Sensed Land Surface Temperature, and Its Relationship with Physiographic Variables in the Russian Altay Mountains. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *20*, 4–19. [[CrossRef](#)]
50. Yan, H.; Zhang, J.; Hou, Y.; He, Y. Estimation of Air Temperature from MODIS Data in East China. *Int. J. Remote Sens.* **2009**, *30*, 6261–6275. [[CrossRef](#)]
51. Zeng, L.; Wardlow, B.D.; Tadesse, T.; Shan, J.; Hayes, M.; Li, D.; Xiang, D. Estimation of Daily Air Temperature Based on MODIS Land Surface Temperature Products over the Corn Belt in the US. *Remote Sens.* **2015**, *7*, 951–970. [[CrossRef](#)]
52. Kaufmann, R.K.; Zhou, L.; Myneni, R.B.; Tucker, C.J.; Slayback, D.; Shabanov, N.V.; Pinzon, J. The Effect of Vegetation on Surface Temperature: A Statistical Analysis of NDVI and Climate Data. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
53. Jeong, S.J.; Ho, C.H.; Jeong, J.H. Increase in Vegetation Greenness and Decrease in Springtime Warming over East Asia. *Geophys. Res. Lett.* **2009**, *36*. [[CrossRef](#)]
54. Pouteau, R.; Rambal, S.; Ratte, J.P.; Gogé, F.; Joffre, R.; Winkel, T. Downscaling MODIS-Derived Maps Using Gis and Boosted Regression Trees: The Case of Frost Occurrence over the Arid Andean Highlands of Bolivia. *Remote Sens. Environ.* **2011**, *115*, 117–129. [[CrossRef](#)]
55. Shen, S.; Leptoukh, G.G. Estimation of Surface Air Temperature over Central and Eastern Eurasia from MODIS Land Surface Temperature. *Environ. Res. Lett.* **2011**, *6*. [[CrossRef](#)]
56. Tian, Z.; Jiang, D. Revisiting Last Glacial Maximum Climate over China and East Asian Monsoon Using Pmip3 Simulations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2016**, *453*, 115–126. [[CrossRef](#)]
57. Sun, L.; Shen, B.; Sui, B.; Huang, B. The Influences of East Asian Monsoon on Summer Precipitation in Northeast China. *Clim. Dyn.* **2016**. [[CrossRef](#)]
58. Mostovoy, G.V.; King, R.L.; Reddy, K.R.; Kakani, V.G.; Filippova, M.G. Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS Lst Data over the State of Mississippi. *Gisci. Remote Sens.* **2006**, *43*, 78–110. [[CrossRef](#)]
59. Huang, R.; Zhang, C.; Huang, J.; Zhu, D.; Wang, L.; Liu, J. Mapping of Daily Mean Air Temperature in Agricultural Regions Using Daytime and Nighttime Land Surface Temperatures Derived from Terra and Aqua MODIS Data. *Remote Sens.* **2015**, *7*, 8728–8756. [[CrossRef](#)]

