ABSTRACT

Title of Dissertation:	QUANTIFYING THE SPATIAL AND TEMPORAL VARIATION OF LAND SURFACE WARMING USING SATELLITE DATA
	Yuhan Rao, Doctor of Philosophy, 2019
Dissertation directed by:	Professor Shunlin Liang Department of Geographical Sciences

The global mean surface air temperature (SAT) has demonstrated the "unequivocal warming". To understand the impact of the global warming, it is very important to quantify the spatial and temporal patterns of the surface air temperature change. Currently, most observational studies rely on in situ temperature measurements over the land and ocean. But the uneven and sparse nature of these temperature measurements may cause large uncertainty for the climate analysis especially at local and regional scales. With the rapid development of satellite data, it is possible to estimate spatial complete surface air temperature from satellite data using advanced statistical models. The satellite data-based estimation can serve as a better data source for local and regional climate analysis to reduce analysis uncertainty.

In this dissertation, I firstly examined the uncertainty of four mainstream gridded SAT datasets over the global land area (i.e., BEST-LAND, CRU-TEM4v, NASA-GISS, NOAA-NCEI). The comprehensive assessment of these datasets concludes that different data coverage may cause remarkable differences (i.e., -0.4 ~ 0.6°C) of calculated large scale (i.e., global, hemispheric) average SAT anomaly using different datasets. Moreover, these datasets show even larger differences at regional and local scale (5°×5°). The local and regional data differences can lead to statistically significant differences on linear trends of SAT estimated using different datasets. The correlation analysis shows strong relationship between the uncertainty of estimated SAT trends and the density of in situ measurements across different regions.

To reduce the uncertainty of surface air temperature data, I developed a statistical modelling framework which can estimate daily surface air temperature using remote sensing land surface temperature and radiation products. The framework uses machine learning models (i.e., rule-based Cubist regression model and multivariate adaptive regression spline) to characterize the physical difference between land surface temperature and surface air temperature by including radiation products at both surface and the top of the atmosphere. The model was firstly developed for the Tibetan Plateau using Cubist model trained with Chinese Meteorological Administration station measurements. Comprehensive evaluation show that the Cubist model can estimate the surface air temperature with nearly zero degree Celsius bias and small RMSEs between 1.6 °C ~ 2.1 °C. The estimated SAT over the entire Plateau for 2000-2015 show that the warming of the western part of the Plateau has been more prominent than the rest of the region. This result show the potential

underestimation of conventional station measurements based studies because there are no station measurements to represent the rapid warming region.

The machine learning model is then extended to the northern high latitudes with necessary modification to account for the regional difference of the diurnal temperature cycle as well as the large data volume of the northern high latitudes. The MARS model trained using data over the northern high latitudes from the Global Historical Climatology Network daily data archive show a reasonable model performance with the bias of around -0.2 °C and the RMSE ranging between 2.1 – 2.6 °C. Further evaluation shows that the model performs worse over permanent snow and ice surface due to the insufficient training data to represent this specific surface conditions.

Overall, this research demonstrated that leveraging advanced statistical methods and satellite products can help generating high quality surface air temperature data which can provide much needed spatial details to reduce the uncertainty of local and regional climate analysis. The model developed in this research is generic and can be further extended to other regions with proper modification and training using high quality local data.

QUANTIFYING THE SPATIAL AND TEMPORAL VARIATION OF LAND SURFACE WARMING USING IN SITU, SATELLITE DATA

by

Yuhan Rao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:

Professor Shunlin Liang, Chair
Professor George Hurtt
Professor Giovanni Baiocchi
Dr. Dongdong Wang
Dr. Yunyue Yu
Professor Wayne McIntosh

Preface

The dissertation contains six different chapters. The first chapter provides the overview of the scope and background of the research, while Chapter 2 to Chapter 4 describe the results of three inter-connected research. The Chapter 5 summarizes the lessons learned regarding applying machine learning models to reduce data uncertainty for local and regional climate studies, and the last chapter provides the conclusion of the dissertation and lay out future work to enhance the research.

Chapter 2 has been published on the Journal of Geographical Research: Atmosphere. Chapter 2 and Chapter 5 have been submitted to the journal Remote Sensing of Environment for publication. The submitted manuscript have contributions from Drs. Zhen Song, Yuan Zhou, Miaogen Shen, and Baiqing Xu who aided on accessing and processing station measurements and satellite observations. Chapter 4 will be submitted to peer reviewed journal for publication in the future. The co-author of this manuscript also includes Drs. Zhen Song and Yuan Zhou who provided the access to satellite derived radiation products.

Dedication

To my parents

Qiaoping Hu (胡巧萍) and *Huaming Rao* (饶华明)

To everyone who is committed to promote

open data, open source projects, and open scholarly research

Acknowledgements

First and foremost, I would like to acknowledge my advisor, Dr. Shunlin Liang, my supervisor at NOAA, Dr. Yunyue Yu for their continuous support and guidance on both academic and personal matters throughout my PhD. I still remember the first day when I arrived at the Dulles airport on the same flight with Dr. Liang. He kindly took me back to College Park in a taxi and drove me to my house late in the night.

I also want to share my appreciation to my dissertation committee members for their continuous guidance and help throughout the journey of my dissertation research. They are always very patient and responsive when I wanted to discuss any research related questions. They have always challenged me to improve the implementation and presentation of my research as an independent scholar, which I really appreciate and value.

I would love to thank all my group members and visiting scholars for the wonderful companion and support throughout this journey. I would share special thanks to Dongdong, Tao, Danxia, Zhen, Xiaona, Yunfeng, Fengming, Hongya, Xin, Yi, Meredith, and Yuan, who has supported me for the challenging transition time during the first year at UMD.

Pursuing a PhD degree could sometimes be boring and lonely. However, this is not my case thanks to the great cohort of graduate students and colleagues in the department. I am very grateful for all the peer support from my cohort throughout the journey. Wish all of you a great journey ahead no matter where you go next. A special thank you to my fellow LeFrak cohort members, Meredith, Cole, Kelly, Diana, and Cortney for wonderful lunch conversations and fun weekend gatherings.

Also, thanks to all my friends and colleagues in the department who made my time at UMD special and memorable. A special note to my dear friends from the Pineway International House, Ciccio, Anna, Ricky, Vale, Hulya, Gamze, Romina, Maya, Nur, Aakash.

Thanks to all my previous and current housemates who need to occasionally listen to my complaints about the PhD life. Also, thanks to all the wonderful trivia teammates who made Tuesday nights exciting and fun!

I would also like to acknowledge the Chinese Scholarship Council, National Center for Atmospheric Research, the Cooperative Institute for Climate & Satellites – Maryland, and Dr. Jingli Yang for financially supporting my dissertation research.

Thanks to all the faculty and staff in the department for their committed daily support, so I can do my research worry-free. I would also like to thank all my friends and colleagues from BSOS Dean's Graduate Student Advisory Council and Graduate School for giving me the opportunity to explore new ideas to improve our own graduate experiences here.

Lastly, I want to thank my parents for their unconditional love and support throughout my time here at Maryland. Without their love and support, it will be impossible for me to complete the wonderful PhD journey.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background and motivation

The global mean surface temperature (GMST) has demonstrated the "unequivocal warming" in the climate system (IPCC, 2013), but the warming rate of GMST in the last two decades has slowed down because of internal variability (Yan et al., 2016). Research argued that this "warming hiatus" is the result of energy redistribution caused by large atmospheric and oceanic circulations (Trenberth, 2015; Xie, 2016). However, it should be noted that although GMST is a useful indicator for monitoring climate change, this integrated indicator has very little direct impact on the human society and ecosystems. It is the local, regional and seasonal variations of temperature that are vital for assessing impacts of climate change (Nature Geoscience, 2014). Different evidence has shown varying warming rate at different regions, such as the "warming hole" in central continental United States which has experienced cooling trends that is significantly different than rapid warming patterns in neighboring regions (Pan et al., 2004). However, due to incomplete coverage and uncertainty in data, variations of surface temperature at different spatial and temporal scales remain unclear (Ji et al., 2014).

Currently, studies mainly rely on model simulations and in-situ observations from meteorological stations to quantify surface temperature changes (Dai, 2013; Hegerl et al., 2014; Huntington, 2006). Stations over land, the majority of which located in Europe and North America, are sparsely and unevenly distributed across different continents (Hegerl et al., 2014). To minimize the uncertainty caused by incomplete

station coverage, efforts have been made to fill gaps at coarse scales (e.g., 1°-5°) by interpolating observations or simply averaging observations from neighboring stations (Hansen et al., 2010; Jones et al., 2012; Rohde et al., 2013; Vose et al., 2012). However, these spatially filled data may still introduce large uncertainty in the tropics, high latitudes, and mountainous regions where long-term high-quality observations could be rare. This data uncertainty could affect the estimation of spatial and temporal variations of the surface warming rate and its impacts (Thorne et al., 2016).

For model based research, different assumptions and parameterizations could lead to uncertainty in model simulations (Trenberth, 2015; Xie, 2016; Yan et al., 2016). Therefore, quantifying the surface warming rate at regional and local scales based on different global or regional models could produce different or sometimes contradictory results (Trenberth, 2015; Trenberth et al., 2014; Trenberth and Fasullo, 2013). The multi-model ensemble mean is usually used to reduce impacts of model uncertainty at large scale studies. Although multi-model ensemble mean provides complete spatial coverage, models could differ notably at regional and local scales, thus hampering the confidence of warming analysis at different spatial and temporal scales.

Recently, remotely sensed data have been valued for climate studies and model improvements (Yang et al., 2013, 2016). Efforts have been made to derive essential climate variables (ECVs) for climate studies, such as air temperature, land surface temperature, precipitation, and Earth energy budget using remotely sensed data acquired since late 1970s (Ashouri et al., 2014; Shi et al., 2016; Shi and Bates, 2011; Yu et al., 2008). These satellite products provide unique opportunities to estimate the spatial and temporal variations of the surface warming rate because of the global

coverage and various spatial and temporal resolutions. However, the data length and inconsistency of single-sensor products, due to the short lifetime of satellites/sensors, has limited their applications in climate studies so far (National Research Council, 2004). To address this issue, National Research Council has proposed the development of climate data records (CDR) from environmental satellites by combining data from multiple platforms and sensors for long-term climate studies (National Research Council, 2004). Although there would still be uncertainty in CDRs, the complete data coverage can provide valuable insights to quantify the surface warming rate and its impacts across different spatial and temporal scales.

With the increasing amount of satellite observations available to the research community, researchers have turned the attention to use advanced statistical modeling approach to provide an alternative of the surface temperature datasets, especially for the regions that are experiencing dramatic change in the recent decades (Alfieri et al., 2013; Good, 2015; Hall et al., 2013; Meyer et al., 2016; Rayner et al., 2018; Shen and Leptoukh, 2011; Squintu et al., 2019; Zhang et al., 2016). Particularly, machine learning models have become widely popular to estimate the surface air temperature with the remote sensing LST as the main input (Meyer et al., 2016; Noi et al., 2017; Xu et al., 2018; Zhang et al., 2016). The application of machine learning on mapping SAT is mainly because of the strong correlation between the satellite estimated land surface temperature and the station measured surface air temperature across different landscapes (Good et al., 2017; Lu et al., 2018; Nielsen-Englyst et al., 2019). However, these studies all share the same limitation that the estimated surface air temperature is only available for clear sky conditions due to the input land surface temperature data

(usually derived from thermal infrared satellite data) are prone to cloud contamination. This so called “clear sky bias” makes it difficult to apply the estimated temperature data for climate studies in an unbiased manner. As land surface temperature has been recognized by the Global Climate Observing System (GCOS) as one of the ECVs (WMO, 2016), it becomes more important to address this limitation caused by cloud contamination.

The research of my dissertation is mainly motivated by two factors. The first factor is the low confidence of our current understanding of surface air temperature change at regional and local scales, which can be attributed to the lack of high quality temperature datasets. The second factor is the great potential of leveraging machine learning and remote sensing products to provide high quality climate datasets to enhance our confidence in local and regional climate analysis.

In the remaining of this chapter, I will first briefly review the status of current surface temperature datasets, including station measurements, global gridded datasets based on station measurements, model-based temperature datasets, and remote sensing temperature datasets. Afterwards, the overarching research question and three different research objectives of this dissertation are outlined before the overview of the dissertation structure.

1.2 Review of current surface temperature data

1.2.1 Weather station measurements

It is the invention of the Six’s thermometer (or the maximum-minimum thermometer) in 1780 make it possible to accurately measure and record the temperature of a local

area over a period of time, for example 24 hours. In 1781, the Meteorological Observatory Hohenpeißenberg, the world's oldest mountain weather station, was established (Gantner et al., 2000). It has been continuously measuring temperature since then. However, it is not until 1850s that the instrumental records of thermometer measured temperature at weather stations become available for different parts of the world (Hansen et al., 2010). Since then, the number of available global weather stations has been increasing steadily. The archived temperature measurements of these stations have become the basis for climate scientists to reconstruct the historical record of global surface temperature over the land. Since these temperature measurements were collected by different authorities (e.g., states, countries, research groups), there are multiple efforts led by different international institutions to collect, curate, archive, and distribute the historical temperature measurements for climate studies. Table 1-1 lists the major collections of historical temperature measurements that are freely available for public use. With the increasing volume of temperature measurements, the uncertainty of the reconstructed time series of GMST has decreased notably (Hansen et al., 2010; IPCC, 2014; Jones, 2016).

Table 1-1. The summary of the available collections of weather station measured historical near surface air temperature data.

Name	Source	Data frequency	Number of stations
GHCN-M	NOAA NCEI	Monthly	~26,000
GHCN-D	NOAA NCEI	Daily	~100,000
ISD	NOAA NCEI	Sub-daily/ daily/monthly	~35,000
ISTI	ISTI	Daily	~36,000
BEST	Berkeley Earth	Daily	~39,000

Although the number of the weather stations worldwide has been increasing steadily, both the distribution of the station and the distribution of the increase has been rather uneven. Most of the stations are clustered at the regions with higher level of economic development before the 1950s, such as, North America, Europe, and the coastal regions of Australia. On the contrary, the less developed and less populated regions of the world have been neglected for very long time period, such as, Africa, South America, high mountainous regions, and high latitudes. Unfortunately, these regions are also very vulnerable to anthropogenic climate change. Although the distribution of the historical and current weather stations is uneven, many studies have concluded that the current archive of weather station data can capture the change of GMST over land with limited uncertainty with proper data processing (Hansen et al., 2010; IPCC, 2014; Jones, 2016). The high confidence at the global scale can be attributed to the spatial autocorrelation of the near surface air temperature. Theoretically, I need at least 400 weather stations that are located at the selected locations of the world over the land to accurately capture the trend of GMST (Thorne et al., 2017). However, the lack of weather station measurements has led to low level of confidence for the regional and climate analysis.

1.2.2 Station based gridded temperature data

Although the number of weather stations is growing, the station measured data can only represent the temperature of a limited region. To infer the temperature of any given corner of the world, researchers have developed spatial interpolation methods to generate the temperature of any given locations based on available station temperature measurements. Here, I want to emphasize the difference between “measurement” and “estimation”. As defined by the Oxford English Dictionary, “measurement” is “the

size, length, or amount of something, as established by using an instrument or device marked in standard units"; "estimation" is "a rough calculation of the value, number, quantity, or extent of something." In this dissertation research, "measurement" only refers to the temperature data measured by weather stations. Thus, all SAT datasets other than the station measurements are the estimation of the temperature of any given locations using predefined mathematical methods. These datasets are all subject to errors and uncertainty associated with the data and methods that they are using.

With different spatio-temporal interpolation methods, several global gridded temperature datasets are developed by different institutions independently (Table 1-2). Each of these datasets is developed by different institutions independently. The Berkeley Earth uses more than 36,000 globally distributed weather station data and a modified Gaussian process regression (or Kriging interpolation) to generate the BEST-LAND (Rohde et al., 2013). The Climate Research Unit at the University of East Anglia (UEA-CRU) uses the temperature measurements of about 5,600 weather stations worldwide to produce the CRU-TEM4v, which is used as the land component of the global temperature dataset maintained by United Kingdom's Met Office Hadley Center (HadCRU) (Jones et al., 2012). Difference from other datasets, the CRU-TEM4v does not fill the data gap for any months and any 5°×5° grids with no station measurements (Jones et al., 2012). The Goddard Institute for Space Studies at the National Aeronautics and Space Administration (NASA-GISS) generates its long-term global gridded temperature estimation by weighted averaging the temperature measurements of the weather stations (within 1200 km radius for each 2°×2° grid) from the GHCN-M archive (Hansen et al., 2010). The NOAA National Center for Environmental

Information (NOAA-NCEI) also maintains its long-term global gridded temperature data using the temperature measurements of the GHCN-M via the empirical orthogonal teleconnection (EOT) method which minimize the data noise and preserve large scale dynamics and trends (Smith et al., 2008; Vose et al., 2012). These four datasets are created to routinely monitor the GMST change and are updated monthly to reflect the most recent dynamic of the GMST.

Table 1-2. The summary of the global gridded near surface air temperature datasets.

Name	Grid size	Interpolation method	Temporal range	Reference
BEST-LAND	1°×1°	Gaussian process regression*	1753-present	Rohde et al., (2013)
CRU-TEM4v	5°×5°	Not applicable	1850-present	Jones et al., (2012a)
NASA-GISS	2°×2°	Inverse distance weighting	1880-present	Hansen et al., (2010)
NOAA-NCEI	5°×5°	Empirical orthogonal teleconnection	1880-present	Vose et al., (2012)
CRU-Ts.4.02	0.5°×0.5°	Angular distance weighting	1901-present	Harris et al., (2014)

*: *The BEST-LAND dataset uses a modified version of Gaussian process regression by accounting for the station measurement error in their iterative weighting process* (Rohde et al., 2013).

Besides the CRU-TEM4v, the UEA-CRU also produces a high resolution monthly climate dataset (i.e., CRU-Ts.4.02) using the similar source of weather station temperature measurements (Harris et al., 2014). The purpose of the CRU-Ts.4.02 is not to monitor the global temperature change, but to provide a high resolution climatology dataset that contain useful spatial details for climate impact studies. Therefore, the CRU-Ts.4.02 includes not only temperature but also precipitation, solar radiation and other important meteorological variables. With this purpose, the CRU-Ts.4.02 has been widely popular for both global, regional and local studies in ecology, biogeography,

and other related fields. However, the accuracy and quality of the CRU-Ts.4.02 has not been comprehensively evaluated for regions with limited station measurements. Particularly, the analysis of CRU-Ts.4.02 surface temperature data has found that the estimated SATs of many 0.5°×0.5° grids are derived only using stations outside of the target grids within certain radius (Harris et al., 2014).

1.2.3 Model-based data

In addition to the station-based temperature datasets, there are also many climate model or land surface model reanalysis data which also contains the estimation of temperature at various spatial and temporal resolutions. These datasets are produced via the reanalysis process, which is a consistent reprocessing of historical observations from both weather stations and satellite sensors using various numerical models (Dee et al., 2013, 2011; Trenberth et al., 2008). In other words, the reanalysis process uses a statistical framework combine the observational information and model simulation via physical constraints (Dee et al., 2013). It firstly started in early 1980s and has evolved remarkably since then to produce high quality gridded datasets which are fundamental to research in the Earth sciences (Dee et al., 2013). There are several generations of reanalysis datasets that are produced by different international institutions, including NOAA National Centers for Environmental Prediction (NCEP), NASA Global Modeling and Assimilation Office (GMAO), the European Center for Medium-Range Weather Forecasts (ECMWF), and the Japanese Meteorological Administration (JMA). The major differences between different generations are the statistical frameworks and the observation data used during the reanalysis process.

Table 1-3 lists the major global reanalysis datasets that has been produced by different agencies. Considering the scope of this dissertation, I only summarized the most recent version of reanalysis products from various agencies instead of listing all historical reanalysis datasets. ECMWF released its newest version of reanalysis data, ECMWF Reanalysis version 5 (ERA-5) with its Integrated Forecast System (IFS) (Hersbach and Dee, 2016). The ERA-5 has replaced its predecessor ERA-Interim with notable improvements, such as, using climate-appropriate inputs, higher spatial and temporal resolution, providing uncertainty with 10-member ensemble data assimilation, and improved model system and bias correction scheme (Hersbach and Dee, 2016). The ERA-5 provides the estimation of SAT at the grid size of 0.25°×0.25° globally for every 3 hours since 1979 and it will be further extended to 1950s for climate studies. The NASA GMAO generates the Modern-Era Retrospective analysis for Research and Application version 2 (MERRA-2) reanalysis dataset at the grid size of 0.5°×0.625° globally spanning the satellite observing ear since 1980 till present (Gelaro et al., 2017). The NOAA NCEP generates its Climate Forecast System Reanalysis (NCEP-CSFR) at the coarse grid size of 2.5°×2.5° globally using its most recent operational climate forecast system (CFSv2). It has remarkable improvements comparing to its predecessor NCEP Reanalysis 2 (NCEP-R2) developed in cooperation with the Department of Energy (DOE) with a coupled global atmosphere, ocean, land surface and cryosphere reanalysis scheme and improved input data (Saha et al., 2013, 2010). The Japanese Meteorological Agency (JMA) also produce a global reanalysis data at the Gaussian grid T319l60 (with the approximate grid size of 0.56°×0.56°) spanning from 1958. Since JMA's original project is to develop a 55-year reanalysis dataset, it is named as

JRA-55 since then even though the dataset has exceeded its original time span (Kobayashi et al., 2015). At last, the NASA Global Land Data Assimilation System (GLDAS) generates a suite of datasets at different resolutions for global land only using various land surface models (LSM), such as, Noah LSM, Community Land Model (CLM), Variable Infiltration Capacity (VIC) LSM, Catchment LSM, and Mosaic LSM. The purpose of GLDAS is to produce high quality datasets of fundamental variables for the terrestrial hydrological cycle to improve the understanding of land-surface-atmosphere interactions (Beaudoing and Rodell, 2016; Rodell et al., 2004).

Table 1-3. The list of the global model-based datasets produced by different international institutions.

Name	Temporal range/ frequency	Spatial resolution	Source	Reference
ERA-5	1979-present/ hourly	0.25°×0.25°	ECMWF	(Hersbach and Dee, 2016)
MERRA2	1980-present/ hourly	0.5°×0.625°	NASA GMAO	(Gelaro et al., 2017)
NCEP-CSFR	1979-present/ 6-hourly	2.5°×2.5°	NOAA NCEP	(Kanamitsu et al., 2002)
JRA-55	1958-present/ 3-hourly	~0.56°×0.56°	JMA	(Kobayashi et al., 2015)
GLDAS	1979-present/ 3-hourly*	0.25°×0.25°/ 1°×1°	NASA LDAS (Land only)	(Rodell et al., 2004)

With substantial efforts, these model based datasets provide valuable spatial information of fundamental variables of the climate system, including the SAT. However, the uncertainty of these datasets are still not fully understood. The uncertainty of these model based datasets is the combination of the uncertainty inherited from the model and the uncertainty introduced by the input data. Many of these model based datasets use observations from different instruments and satellite platforms as a part of the inputs. The change of instruments may introduce biases and

errors into the model which could further propagate into the final model outputs. Various studies have demonstrated the large biases and regional discontinuities of the current model based datasets at different geographical regions and time periods (Dee et al., 2013; Trenberth et al., 2008).

1.2.4 Remotely sensed data

Table 1-4 lists existing global temperature datasets that are derived or collected by various satellite missions. In the list, there are three different types of remote sensing temperature datasets, i.e., brightness temperature (BT), land surface temperature (LST), and temperature profile. The remote sensing BT data are the observations of the radiance of the radiation traveling upward from the top of the atmosphere to satellite sensors. BT data are usually the raw data containing the information of both Earth surface and the atmosphere column. The Advanced Very High Resolution Radiometer (AVHRR) onboard the series of NOAA Polar Orbiting Environmental Satellites (POES) has been collecting thermal infrared BT data since late 1970s.

Table 1-4. The summary of remote sensing temperature datasets including temperature profiles, land surface temperature (LST), and brightness temperature (BT) of different satellite platforms.

Dataset Name	**Time coverage**	**Spatial resolution**	**Variables***	**Satellite Platform(s)**
AIRS/AMSU	2002-present/ daily	1°×1°	Temperature profile	EOS Aqua
MOD07/MYD07	2000-present/ monthly	5 km/ 1°×1°	Temperature profile	EOS Terra/Aqua
HIRS CDR	1979-2017/ daily	~30 km	Temperature profile	NOAA POES
MOD11/MYD11	2000-present/ daily	1 km/ 0.05°×0.05°	LST	EOS Terra/Aqua
(A)ATSR CDR	1995-2012/ daily	1 km/ 0.05°×0.05°	LST	Envisat

AVHRR CDR	1981-present/ daily	4 km/ 0.05°×0.05°	BT	NOAA POES
VIIRS LST	2012-present/ daily	750 m	LST	Suomi NPP/JPSS

Although the thermal infrared BT contains a mixed signal of surface and atmosphere, BT data has been used to estimate the LST data to monitor the temperature of the "skin" of Earth since 1980s. Most thermal infrared LST products are estimated using split-window algorithms developed for different satellite sensors onboard satellite platforms, such as, MODIS, AVHRR, VIIRS, and ATSR/AATSR (Scarino et al., 2017; Wan and Dozier, 1996; Yu et al., 2008). These LST datasets have been available since late 1990s and early 2000s, which provide useful thermal information of Earth's surface. However, the thermal infrared LST data are only available for clear sky conditions and they become unavailable when there is cloud presenting during satellite overpassing time. Despite this limitation, it has become more popular in recent years for regional and local climate analysis because its strong correlation with the near surface air temperature. However, it should be noted that the LST is physically different from the SAT which are the common climate variables that are used for the majority of climate studies at different scales.

Additionally, there are also temperature profile products derived from thermal infrared sounders onboard different satellite platforms, such as, High-resolution InfraRed Sounder (HIRS) and Atmospheric InfRared Sounder (AIRS). These temperature profile products contain temperature estimation of multiple vertical layers of the atmosphere at different pressure levels (including 2m height). NOAA NCEI generated a long-term climate data records (CDR) temperature profile products using HIRS data ranging from

1979 to 2017 with the horizontal resolution of ~30 km (Peng et al., 2016; Shi et al., 2016). Other temperature profile data include AIRS/AMSU-A monthly gridded temperature data at 1°×1° (Chahine et al., 2006) and MODIS products (MOD07/MYD07) (Seemann et al., 2003; Sobrino et al., 2015), both starting from early 2000s. Similarly with the LST data, these thermal infrared temperature profile products are only available during clear sky conditions, thus leading to possible clear sky bias in the climate analysis. Additionally, the uncertainty of these products are still not fully understood because of the insensitivity of the designed thermal sounder channels to surface conditions.

1.3 Research questions and design

In this dissertation research, I would like to answer the research question that how I can use advanced statistical models and remote sensing data to reduce the uncertainty of surface air temperature data at regional and local scales. To answer this research question, the dissertation research has been further separated into three research objectives.

The first research objective is to quantify the uncertainty of major existing surface air temperature datasets that are estimated based on station temperature measurements at different spatial and temporal scales. Because of the importance of the GMST, many research has focused on reconciling the difference of among different station based temperature datasets. However, little attention has been paid to understand how different surface air temperature datasets may differ at different spatial and temporal scales, and how the difference at different scales may affect the confidence of the climate analysis using these data. To answer these questions, I plan to comprehensively

quantify the differences of four major global gridded SAT datasets at different spatial and temporal scales, and evaluate the impact of the data difference on trend analysis at regional and local scales.

After quantifying the uncertainty of the existing global gridded SAT data, the second research objective is to develop a statistical framework to estimate SAT at high spatial and temporal resolution using remote sensing product for the Tibetan Plateau. As identified by my research of the first research objective, the existing global gridded SAT datasets show large uncertainty over the high mountainous regions. The large uncertainty is likely caused by the lack of station measurements of these regions. As the world's "Third Pole", the Tibetan Plateau has experienced dramatic climate change in the last decades. However, the uncertainty of existing SAT datasets lead to low confidence of the regional and local climate analysis over the Tibetan Plateau. A high quality SAT dataset at high spatial and temporal resolution of the Tibetan Plateau is of urgent need to improve our understanding of how the climate, ecosystem, hydrological cycle have changed over this vulnerable region. Therefore, I developed a machine learning based modeling framework to generate a daily SAT data of the Tibetan Plateau using remote sensing LST data and radiation products.

Besides the high mountainous regions, the high latitudes also suffer from large data uncertainty related to the lack of station observations as identified by the research of my first research objective. The third research objective is to adapt the developed machine learning based framework to larger geographical extents. The northern high latitudes has seen amplified warming as a whole comparing to the rest of the world. However, the lack of in situ measurements has limited the comprehensive

understanding of the warming amplification and its impacts on the environment at local, regional and global scales. With the successful development of the machine learning based framework in the second research objective, I adapted the generic framework to the entire northern high latitudes to demonstrate the applicability of the machine learning framework to other regions.

1.4 Structure of the dissertation

With these predefined three research objectives, the dissertation is organized as follow. Chapter 2 describes the research of quantifying the uncertainty of existing global gridding SAT datasets over the land area and how the uncertainty affects the confidence of local and regional warming trend analysis. Additionally, Chapter 2 also analyzes the potential factor that contributes to the uncertainty among these datasets. Chapter 3 summarizes the research of developing a machine learning based framework to estimate the daily SAT over the Tibetan Plateau. Within Chapter 3, different modeling strategies are firstly described and then compared. After comprehensive evaluation of different modeling strategies, the results of the best modeling strategy are then presented. Chapter 3 also presents the warming analysis of the Tibetan Plateau using the newly estimated SAT dataset generated from the machine learning model. Chapter 4 presents the research results of the third research objective by adapting the machine learning framework developed in Chapter 3 to the northern high latitudes. This chapter firstly describes the importance of high quality SAT data on improving the understanding of the climate change over the northern high latitudes. It then presents the details on how the machine learning model framework developed over the Tibetan Plateau is adapted for the northern high latitudes considering some unique features of

the northern high latitudes. Chapter 5 then summarizes the lessons learned from the machine learning modeling research on how to avoid potential pitfalls of the machine learning model through cautious design and training of a machine learning model with the support of physical mechanisms. Last, Chapter 6 summarizes the research of the dissertation and emphasize the key findings and future research directions.

Chapter 2 Inequality of Weather Station Data Cause Large Uncertainty of Regional and Local Climate Analysis

2.1 Summary

Several groups routinely produce gridded land surface air temperature (LSAT) data sets using station measurements to assess the status and impact of climate change. The Intergovernmental Panel on Climate Change Fifth Assessment Report suggests that estimated global and hemispheric mean LSAT trends of different data sets are consistent. However, less attention has been paid to the inter-comparison at local/regional scales, which is important for local/regional studies. In this study I comprehensively compared four data sets at different spatial and temporal scales, including Berkley Earth Surface Temperature land surface air temperature data set (BEST-LAND), Climate Research Unit Temperature Data Set version 4 (CRUTEM4v), National Aeronautics and Space Administration Goddard Institute for Space Studies data (NASA-GISS), and National Oceanic and Atmospheric Administration National Center for Environmental Information data (NOAA-NCEI). The mean LSAT anomalies are remarkably different because of the data coverage differences, with the magnitude nearly 0.4 °C for the global and Northern Hemisphere and 0.6 °C for the Southern Hemisphere. This study additionally finds that on the regional scale, northern high latitudes, southern middle-to-high latitudes, and the equator show the largest differences nearly 0.8 °C. These differences cause notable differences for the trend calculation at regional scales. At the local scale, four data sets show significant variations over South America, Africa, Maritime Continent, central Australia, and Antarctica, which leads to remarkable differences in the local trend

analysis. For some areas, different data sets produce conflicting results of whether warming exists. Our analysis shows that the differences across scales are associated with the availability of stations and the use of infilling techniques. Our results suggest that conventional LSAT data sets using only station observations have large uncertainties across scales, especially over station-sparse areas. In developing future LSAT data sets, the data uncertainty caused by limited and unevenly distributed station observations must be reduced.

2.2 Background

The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) concludes that it is confident that the global land surface air temperature (LSAT) has warmed since the 1900s, and the increase after the 1970s has been much faster than previous years (IPCC, 2014). This high confidence is based on consistent results using four LSAT datasets produced independently by Berkley Earth, NASA Goddard Institute of Space Studies (GISS), NOAA National Center for Environmental Information (NCEI), and Climate Research Unite (CRU) at University of East Anglia (Hartmann et al., 2013; Jones, 2016). This consistency is only achieved after improvements of these datasets in the recent decade. Jones (2016) asserts that the consistency of the large-scale temperature estimates mainly resulted from 1) similar raw input station data, 2) similar methods for correcting biases and adjusting inhomogeneity of the raw data, and 3) spatial autocorrelation of the temperature data.

Despite the global LSAT being one of the most direct indicators of climate change, it has very little direct impact on ecosystems and human societies, which are mainly influenced by local and regional temperature variations. An analysis based on the

global LSAT reflects the general status of the surface temperature over the global land, but misses the crucial spatial pattern of the surface temperature changes that directly influence ecosystems and millions of people (Editorial, 2017). This spatial pattern of the LSAT change will directly affect the essential functions of human and natural systems, such as vegetation productivity, hydrological events (e.g., snow melting and surface run-off), and human health. To produce global temperature records, these institutions usually generate gridded datasets first with various methods using preprocessed station-based observations at coarse grid-boxes (e.g., 1°–5°) (Hansen et al., 2010; Jones et al., 2012; Muller et al., 2013; Vose et al., 2012). These gridded datasets have been used in various studies to quantify LSAT changes and assess LSAT's impact on human and natural systems at different spatial scales.

Unfortunately, the confidence regarding spatial details of the LSAT change is still low, especially for regions with sparse stations (IPCC, 2014). Most regional- and local-scale studies mainly focus on regions with abundant ground-based observations, such as Continental United States (CONUS), China, Australia, and Europe. Regional studies like these could draw relatively confident conclusions of the regional mean LSAT change. However, the confidence is usually low when it comes to the spatial details of the LSAT change, which could partially be attributed to the station data quality and different preprocessing and gridding methods. Fall et al. (2011) question the potential large biases in observations collected by the United States Historical Climatology Network, of which many stations cannot meet the official World Meteorology Organisation (WMO) siting guidance. However, the overall biases of the network in recent decades can be better explained by instrumental changes rather than siting biases

(Hartmann et al., 2013; Menne et al., 2010).The confidence of regional analysis is worse for observation-sparse regions, such as the Antarctic, high mountains, and other sparsely populated areas. Research can usually agree on the sign of the LSAT change (i.e., warming or cooling) for the regional mean LSAT by interpolating available sparse ground-based observations. However, a significant inconsistency or even disagreement in the magnitude and spatial pattern of the LSAT change has been observed (Thorne et al., 2016).

Surprisingly, little attention has been paid to quantify differences among different gridded temperature datasets and to assess their impact on the LSAT trend calculation at regional and local scales with the importance and wide application of these datasets (Thorne et al., 2016). Since IPCC's Fourth Assessment Report (AR4), many efforts have been made to improve the data quality of individual gridded LSAT dataset by improving spatial coverage, preprocessing, and gridding methods (Hansen et al., 2010; Jones et al., 2012; Muller et al., 2013; Vose et al., 2012). The ultimate goal of these improvements is to reconcile the differences of the global mean LSAT calculated from different data sets, which has been proven successful (Thorne et al., 2016). However, no comprehensive quantification of the dataset differences and the impact on the LSAT trend analysis has been made at the regional and local scales (Thorn et al., 2016). Vose et al. Vose et al. (2005) conducted an inter-comparison between the gridded LSAT datasets produced by CRU, GISS, and NCEI at the global, hemispheric, and grid-box levels. Their study focused on the impact of different gridding techniques and averaging methods on the global and hemispheric mean LSAT (Vose et al., 2005). They only provide a comparison of the estimated linear LSAT trends at the grid-box level

for the CRU and NCEI data, and conclude that a general agreement is met at the grid-box level with large regional variations (Vose et al., 2005). Furthermore, with the recent improvements of these datasets and the new development of the Berkley Earth LSAT dataset, this topic must be comprehensively revisited (Muller et al., 2013).

In addition, gridded LSAT datasets are all constructed using available observations collected by national and regional station networks, which are constantly changing through time (Menne and Williams, 2009). In general, the availability of observations has significantly increased, especially during 1950–1980 (Hansen et al., 2010). However, the change of the data availability is not steady through time, and not even across continents (Hegerl et al., 2014). Different datasets have extended their spatial coverage over different regions by including networks from various agencies and research groups over high latitudes, such as the Antarctic and Greenland (Hansen et al., 2010; Jones et al., 2012). Despite these "data-hunting" efforts, the number of stations used in most, if not all, datasets even decreased in the recent decades (Hansen et al., 2010; Hegerl et al., 2014). This reduction is mainly caused by the elimination of stations in South America and Africa. The reduction could be alarming for both developers and users of these datasets mainly because this will increase the dependence of the LSAT datasets on specific stations over data-sparse regions. With recent effort, the GHCN monthly data (version 4) reverse this decreasing pattern including more station data since 1980s (Rennie et al., 2014). This new station data will be used as main input for both NOAA and NASA for future products. It is still worthwhile to understand how the variation across datasets evolves through time in response to the changing data availability.

The present study focuses on quantifying the differences among four gridded LSAT datasets and their impact on the LSAT trend estimation. Even though most of these datasets are using similar raw input data (except BEST which uses much more stations than others), different quality control procedures, homogenization methods, and gridding techniques could lead to different spatial coverages and values at the grid-box levels (Jones, 2016). This analysis intends to perform a comprehensive assessment to (i) evaluate the data coverage biases of LSAT at the global and regional scales and (ii) quantify the dataset differences and their impact on the LSAT trend estimation at the regional and local scales. In this study, I use local scale and grid-box scale interchangeably since the grid-box level is the finest scale of gridded datasets. Section 2 summarizes the information of individual datasets necessary for the inter-comparison and preprocessing procedures to make the inter-comparison meaningful. Section 3 presents the difference of global and hemispheric mean LSAT caused by data coverage differences. Sections 4 and 5 present the inter-comparison results. The discussion and the conclusion are provided to summarize the inter-comparison results and the implications for the future LSAT change analysis and the new LSAT data development.

2.3 Data and methods

Table 2-1 summarizes the basic information of four major gridded LSAT datasets used in this comparison. These four datasets have been widely used for assessing the status of climate change and its impact on ecosystem and society (IPCC, 2014; Jones, 2016) because of their rigorous quality control, routine (monthly) updates, good documentation and completeness of their archive. Each dataset is briefly described in the following section to ensure appropriate interpretation of current inter-comparison.

More detailed information of individual data sets should be directed to the references listed in Table 2-1.

Table 2-1. The Summary of BEST-LAND, CRU-TEM4v, NASA-GISS, and NOAA-NCEI.

	Grid size	Climatology period	No. of stations	Homogenization method	Interpolation method	reference
BEST-LAND	1°×1°	1951-1980	36,866	"scalpel": split time series using detected break points and automatically adjust weight for each time series	Gaussian process regression/Kriging	Muller et al. 2013; Rohde et al. 2013
CRU-TEM4v	5°×5°	1961-1990	5,583	Comparing with neighbor stations	No interpolation implemented	Jones et al., 2012
NASA-GISS	2°×2°	1951-1980	7,290	Comparing with neighbor stations; urbanization adjustment	Distance-dependent weighted average of station observations within 1200 km radius	Hansen et al., 2010
NOAA-NCEI	5°×5°	1961-1990	7,280	Comparing with neighbor stations	Two-step (low and high frequency) reconstruction using Empirical Orthogonal Teleconnection	Smith et al., 2008; Vose et al., 2012

2.3.1 Berkley Earth Surface Temperature dataset (BEST-LAND)

The Berkley Earth Surface Temperature land surface air temperature dataset (BEST-LAND) combines station temperature measurements from 14 different sources with a total archive of 44,455 sites (Muller et al., 2013; Rohde et al., 2013). A total of 36,866 sites have been kept for the final BEST-LAND process after removing duplicate stations in different datasets and stations with measurements less than 1 year or missing location meta data (Rohde et al., 2013). The largest data source used by BEST-LAND is the GHCN-Daily (more than 25,000 stations) and GHCN-Monthly (GHCNM-v3

with 7280 stations, which will be replaced by GHCNM-v4 later this year) data archive managed by NOAA NCEI. These datasets contain temperature measurements from 180 countries (Menne et al., 2012). After preprocessing and monthly averaging, the station measurements are interpolated into 15,984 equal-area grid cells (with nearly 1.25° resolution at the equator) for the earth surface using Gaussian process regression (i.e., Kriging interpolation), then regridded into 1°×1° grid-boxes (Rohde et al., 2013; Thorne et al., 2016).

The station data used in BEST-LAND are raw data from each data source with no homogenization and limited data quality control. Before the interpolation process, it uses a pair-wise method to identify statistical breakpoints within the original data for each station compared with its neighboring stations (Rohde et al., 2013). Unlike other groups, BEST-LAND does not correct the detected discontinuities potentially caused by site relocation, instrument changes, and urbanization effects. Instead, it separates original data into different fragments at detected breakpoints and treats these fragments as data from different stations. This process, called "scalpel", is intentionally designed to reduce the human bias caused by adjustment (Muller et al., 2013; Rohde et al., 2013). An original archive of 36,866 stations produces 179,928 data fragments with the "scalpel" process (Rohde et al., 2013). These fragments are then used to produce the temperature data field for each cell using Kriging interpolation with an integrated iterative bias adjustment and outlier de-weighting process. This process is designed to be tolerant of data records with a limited length, thereby allowing majority of the reliable station observations to be used in the analysis. BEST-LAND provides

temperature anomalies from 1850 to present time comparing to the climatology period of 1951–1980.

2.3.2 Climate Research Unit temperature dataset (CRU-TEM4v)

CRU-TEM4v is a gridded LSAT dataset provided by the Climate Research Unit at the University of East Anglia with the variance adjusted for changing station numbers within each grid (i.e., 5° × 5°). The station temperature measurements used in CRU-TEM4v are combined from multiple data sources with a total of 5583 stations (Jones et al., 2012). The main data source is the National Meteorology Services (NMSs) of countries across the world comprising 2444 stations in the final archive. Another important data source for CRU-TEM4v is the decadal World Weather Report (WWR) starting from 1950s onwards, which provides data records for underrepresented nations in NMS data (mainly South America, Africa, Asia and many island groups) (Jones et al., 2012). With its most recent data archive update, CRU-TEM4v has improved its data coverage over the Arctic compared to its predecessor (i.e., CRU-TEM3v) by including more data for the Arctic region provided by Bekryaev et al. (2010) and the Danish Meteorological Institute reports.

Instead of performing homogenization for all stations, CRU mostly relies on homogenized temperature records provided by NMSs (Jones et al., 2012). Additionally, CRU identifies 219 data records for additional adjustment of homogeneity caused by various factors, such as change of instruments, site locations, and local environments. The adjustment is performed by comparing with multiple neighboring stations (Jones et al., 2012). The time series of absolute temperature are then converted into anomalies by simply subtracting the long-term average for each month derived from a base period

(i.e., 1961–1990 for CRU-TEM4v) for each station, which is referred to as the climate anomaly method (CAM). Finally, a temperature anomaly of each grid-box is generated by simple averaging of all the available station anomalies within each grid-box (Jones et al., 2012). Station availability changes through time for some grid-boxes; hence, CRU-TEM4v has also adjusted the variance of each grid-box to account for this factor using the method outlined in Brohan et al. (2006).

2.3.3 NASA GISS temperature dataset (NASA-GISS)

The NASA Goddard Institute for Space Studies produces a global land surface air temperature dataset using the reference station method (Hansen et al., 2010). The majority of station temperature time series used by the NASA-GISS is obtained from GHCN-Monthly version 3 with a total of 7280 stations. By selecting stations with at least 20 years of records, nearly 6300 stations are kept for further analysis. Hansen et al. (2010) also use monthly data collected by the Scientific Committee on Antarctic Research (SCAR) since 1957 to fulfill data gaps in the GHCN station archives. Similar with CRU-TEM4v, the original time series are inspected for homogeneity and adjusted for inhomogeneity if necessary. Moreover, Hansen et al. (2010) utilize satellite night-light radiance data to identify the stations affected by urban effects. They correct these station measurements by comparing them with their neighboring rural stations.

After the adjustments, the station measurements are converted to anomalies compared to their long-term average for each month derived from the base period of 1951–1980. The temperature anomalies for 2°×2° grid cells are generated by weighted averaging of the anomalies for all stations within 1200 km of that grid. The weight for each station

is a linear function decreasing with its distance from the grid center (Hansen et al., 2010).

2.3.4 NOAA NCEI temperature dataset (NOAA-NCEI)

The NOAA National Center for Environmental Information Surface Temperature Dataset at a 5° × 5° grid-box is created by separately constructing low-frequency and high-frequency variations. The station observations used by NOAA-NCEI are from the GHCN-Monthly dataset (GHCNM-v3). The inhomogeneity adjustments for each time series are implemented for each station using a pair-wise method compared with its neighboring stations (Menne & Williams, 2009). All homogenized data records are then converted to anomalies with the base period for 1961–1990. For each grid-box, the anomalies for the stations within the grid-box are averaged into a "super-observation," which is used for the final reconstruction process (Vose et al., 2012).

The reconstruction for the NOAA-NCEI assumes that temperature anomaly time series can be divided into two different components: low frequency variations that reflect long-term changes and high frequency variations that represent temperature variability over short time periods (Smith et al., 2008; Smith and Reynolds, 2005; Vose et al., 2012). Therefore, the low frequency component is first derived by a simple spatio-temporal smoothing method (Smith et al., 2008). The residual anomaly time series for each grid-box is then fitted to a group of large-scale spatial-covariance modes derived from the modern era data (1982–1991, with a maximum spatial coverage) using the empirical orthogonal teleconnections (Smith and Reynolds, 2005; van den Dool et al., 2000). The final temperature anomaly time series for each grid-box is obtained by simply adding the smoothed low-frequency time series and the fitted high-frequency

time series. The residuals of the reconstructed data compared with the original data are treated as background noise potentially arising from uneven sampling, observation errors, etc. (Vose et al., 2012). The reconstruction is designed to capture the key trends and patterns while neglecting the local, short-term irregularities, and provide anomalies in unsampled areas by identifying spatial-covariance modes (Smith et al., 2008). For consistency with its ocean counterpart, the anomalies over the land grid-box (with base period of 1961–1990) are converted to anomalies comparing to base period of 1971–2000 (Vose et al., 2012).

2.3.5 Data processing

Each dataset has its own spatial resolution and climatology period; hence, all datasets must be adjusted to the same spatial resolution and climatology period to ensure a meaningful inter-comparison. First, NOAA-NCEI and CRU-TEM4v are adjusted to the LSAT anomaly values against the monthly climatology of 1951–1980 by subtracting the 30-year mean value (1951-1980) for each month from the original anomaly. After the climatology adjustment, fine resolution datasets (i.e., BEST-LAND and NASA-GISS) are aggregated to the spatial resolution of 5°×5° by weighted average considering the grid area change caused by the latitude.

The comparison is performed at different spatial scales, including global, hemispheric, latitudinal, regional, and local scales. The spatial average is calculated using weighted average considering the grid area change caused by the latitude. The latitudinal average is calculated for each 20° latitudinal band, while the regional average is calculated based on the regions described in Table 2-2. In addition, the spatial coverage is different across datasets because of the different gridding methods and source data used by each

group. Hence, I define two valid spatial coverages for calculating the spatial average: native and common data coverage. The native coverage for a dataset includes all grid-boxes, of which the dataset has a non-missing anomaly value, whereas the common coverage only includes grid-boxes, of which all four datasets have non-missing anomaly values. All spatial averages are calculated based on both native and common coverages.

The datasets are also compared at different temporal scales, including annual and seasonal mean. Therefore, the monthly anomaly is averaged through a given year or season. In this study, I define four seasons as December–January–February (DJF), March–April–May (MAM), June–July–August (JJA), and September–October–November (SON), where the December value is from the previous year.

For trend calculation, I use ordinary the least square (OLS) method to estimate the linear trend for the given time periods (i.e., 1901-2017, 1951-2017, 1981-2017 and 1998-2017). However, temperature data are usually strongly autocorrelated, which will lead to underestimation of standard error for OLS-estimated trends (Hausfather et al., 2017; Lee and Lund, 2004; Santer et al., 2000). To address this issue, I consider an autoregressive-moving-average model with the order of 1 for each component (i.e., ARMA(1,1)) to adjust standard errors of estimated trends for global, hemispheric, regional and local scales. More details of the adjustment method can be found in Hausfather et al. (2017) and Lee & Lund (2004).

To test whether LSAT differences across datasets cause significant impact on linear trend calculation, I adopt the method from Hausfather et al. (2017) which calculates the statistical significance of linear trends of dataset difference time series. In theory, when

data coverage is the same, difference between trends estimated from two time series is the same as trends estimated from the difference time series. This method can avoid the dependency of different datasets caused by similar source data (Hausfather et al., 2017). Because BEST-LAND uses much more stations to generate the gridded product, I use it as our reference for the difference time series calculation. The difference time series of CRU-TEM4v, NASA-GISS, and NOAA-NCEI comparing to BEST-LAND are then used to estimate the difference trends using OLS. I also use ARMA(1,1) model to address the autocorrelation in difference time series.

Table 2-2. Geographical Boundaries of the Regions Used for Calculating the Spatial Average.

Continent	Min. latitude (degrees North)	Max. latitude (degrees North)	Min. longitude (degrees East)	Max. longitude (degrees East)
North America 1 (NA1)	15	50	−165	−50
North America 2 (NA2)	50	85	−165	−50
South America 1 (SA1)	−23.5	15	−90	−30
South America 2 (SA2)	−60	−23.5	−80	−40
Europe (EUR)	35	80	−15	60
Africa (AFR)	−35	30	−20	50
Asia 1 (AS1)	5	50	45	150
Asia 2 (AS2)	50	80	60	180
Maritime Continent (MCT)	−10	5	90	165
Australia (AUS)	−50	−10	110	155
Antarctica (ANT)	−90	−60	−180	180
Greenland (GRL)	60	90	−70	−10

2.4 Differences of the global and hemispheric mean LSATs

I first examine how different spatial coverages affect the calculation of large-scale mean LSAT (i.e., global and hemispheric averages) by comparing the spatial average calculated using different coverages (i.e., native and common coverages) for each dataset. Figure 2-1 shows the differences between the global and hemispheric averages of the native coverage and the ones of the common coverage. The CRU-TEM4v generally has the lowest differences in the mean LSATs calculated using different data coverages because CRU-TEM4v does not use interpolation techniques to fill in the grid-boxes with no observations (Cowtan & Way, 2014; Jones et al., 2012). Thus, it has the smallest native coverage similar with the common data coverage for all datasets.

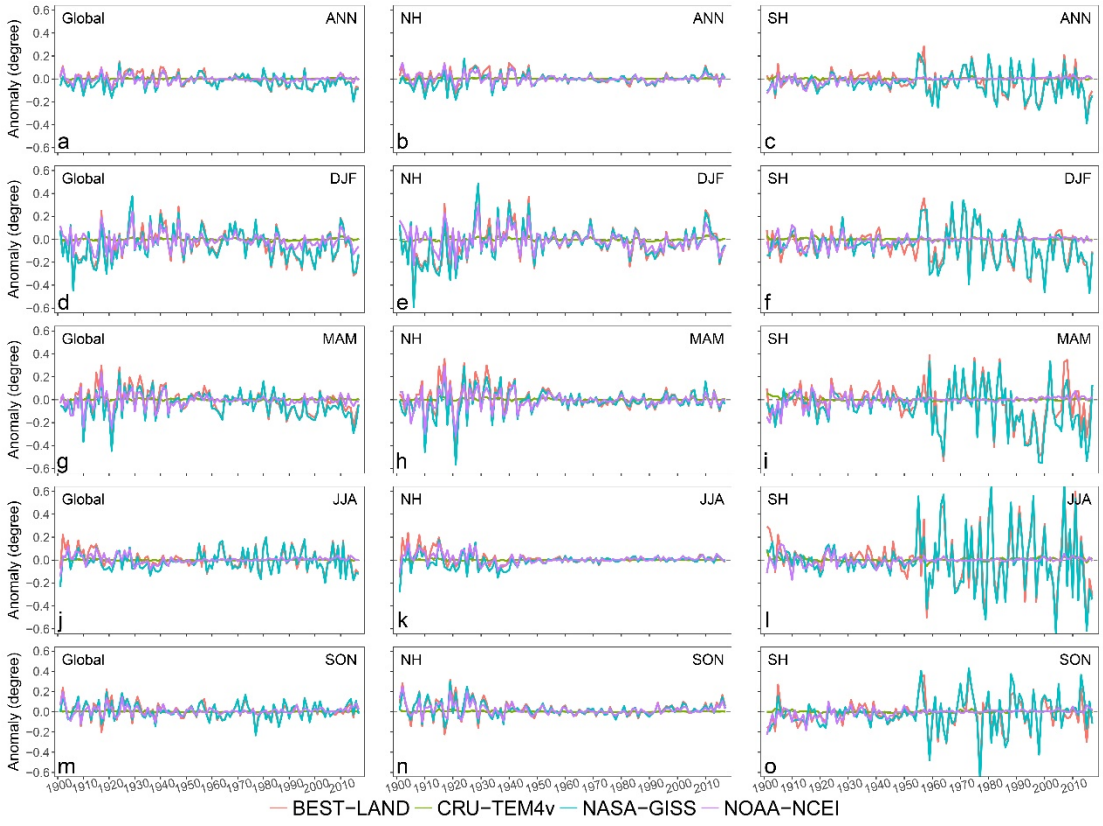

Figure 2-1. Differences between the temperature anomalies calculated using the native coverage and the common coverage for four datasets (ANN: annual; DJF: December–

January–February; MAM: March–April–May; JJA: June–July–August; and SON: September–October–November).

For the global scale, the data coverage caused differences are relatively stable for the annual mean, ranging between ±0.2°C, but vary remarkably for the seasonal mean, especially for DJF and MAM before 1960, which vary between ±0.4°C. The difference caused by the data coverage is substantial considering that the magnitude of global warming is around 1.8°C since 1900 (Jones, 2016). NASA-GISS and BEST-LAND have the largest differences because they both use infilling techniques to estimate temperatures of the grid-boxes with no stations. Meanwhile, NOAA-NCEI has smaller differences, followed by CRU-TEM4v. Figure 2-2 depicts that BEST-LAND and NASA-GISS have the highest percentage of valid land area for the whole study period because of spatial infilling, especially over high latitudes and the tropics.

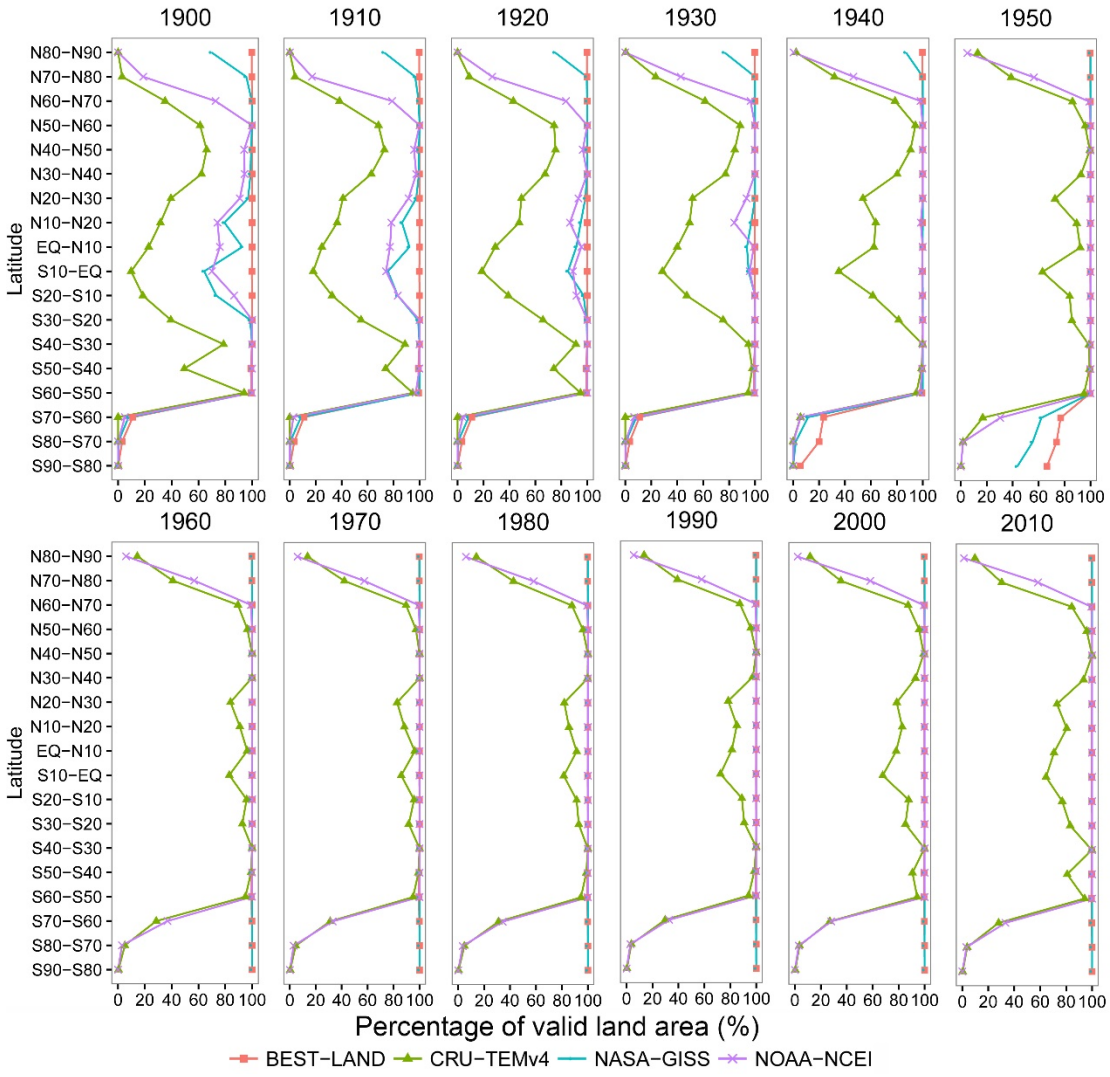

Figure 2-2. Decadal change of the percentage of land areas with valid LSAT anomaly values for different 10-degree latitudinal bands for each dataset.

The difference of the northern hemisphere (NH) mean LSAT has the largest variation before 1950s for both annual and seasonal mean LSATs. The variation is relatively smaller for the annual mean and warm seasons (JJA and SON) (i.e., ±0.2°C), but much larger for the cold seasons (DJF and MAM) ranging from −0.6°C to 0.4°C. The large variation before 1950s are mostly caused by the data coverage differences for 0–70°N (Figure 2-2). The difference caused by the data coverage of this region is mostly caused by the low data coverage of CRU-TEM4v and NOAA-NCEI, which do not provide

LSAT for grid-boxes with no station observations (Jones et al., 2012; Smith et al., 2008; Vose et al., 2012).

In contrast with the northern hemisphere, the difference for the southern hemisphere mean LSAT shows the largest variations after the 1950s. The difference ranges from −0.4°C to 0.3°C for the annual mean LSAT and DJF mean LSAT. It varies within (−0.6, 0.4)°C for MAM/SON and (−0.6, 0.6)°C for JJA. The large variation after 1950s is mostly caused by the rapid increase of the data coverage in BEST-LAND and NASA-GISS (Figure 2-2). This data coverage improvement is the result of the extensive efforts to add new station observations over the southern high latitudes (Hansen et al., 2010; Rohde et al., 2013). Although the number of stations over the southern high latitudes is still very limited, BEST-LAND and NASA-GISS use spatial interpolation methods to fill in the data gaps, which provides complete data coverage for this region. Notably, the difference in data coverage for the SH results in an obvious trend for certain time periods, such as in MAM 1970-2000 (Figure 2-1).

Furthermore, I examine the impact of dataset differences on the estimated trend for different time periods and seasons (Table A1). For global mean LSATs, CRU-TEM4v, NASA-GISS, and NOAA-NCEI all show significant difference trends comparing to BEST-LAND for both annual and seasonal during 1951-2017. The difference trends range from 0.01 to 0.02 degree/decade. This positive difference trend pattern also presents in NH for almost all seasons and datasets, while it only appears in NASA-GISS and NOAA-NCEI for SH. During 1981-2017, CRU-TEM4v shows a significant trend difference with BEST-LAND for both global and NH mean, while NASA-GISS and NOAA-NCEI mainly differ with BEST-LAND for SH. Due to relative short time

period for 1998-2017, I only find that CRU-TEM4v differs significantly from BEST-LAND for NH trends of mean LSATs.

Although a previous analysis claims that large-scale mean LSAT should be robust against the dataset choice (Jones, 2016), our analysis shows that it is sensitive to the data coverage of different datasets. The difference causes significant trend differences estimated from different data. This issue should be examined more carefully in the future IPCC assessment. The users of these datasets should be cautious in terms of the conclusions inferred from the global/hemispheric mean LSAT using only single dataset, particularly for seasonal mean temperatures.

2.5 Latitudinal and regional LSAT differences

2.5.1 Decadal and annual mean LSAT comparison

Figure 2-3 and Figure 2-4 demonstrate the differences of the regional mean LSAT anomalies for different latitudinal zones and predefined regions in Table 2-2. I only show the decadal mean values instead of the individual years or months to focus on the systematic differences rather than on the inter-annual variability. Figure 2-3 shows that overall LSAT changed remarkably for all regions in all four datasets, especially in recent decades, but with strong latitudinal variations. However, the differences among the four data sets are evident despite the agreement on the general warming patterns.

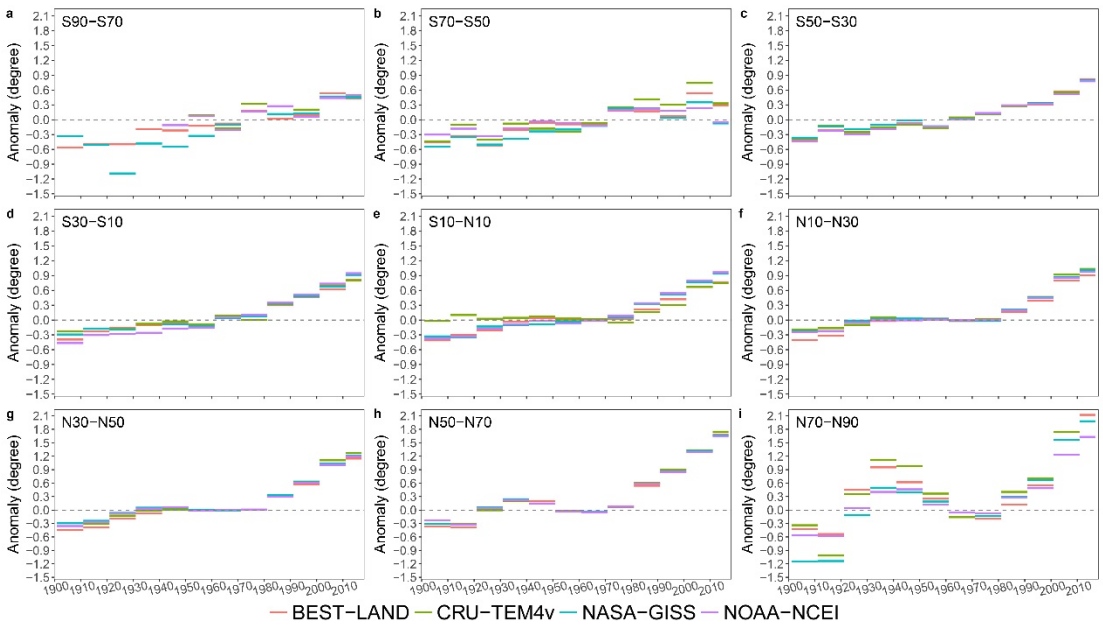

Figure 2-3. Differences of the decadal annual mean LSAT anomaly between the four data sets for different latitudinal bands.

As shown in Figure 2-3, the four datasets have the smallest differences at the southern mid-latitude and northern mid-to-high latitude (i.e., 30–50°S and 30–70°N), which are regions with rich ground stations (Figure 2-2). The differences are larger near the equator and the largest in the southern high latitude and polar regions (i.e., 50–70°S, 10°S–10°N, and 70–90°N/S), where only a very limited number of, if any, station observations are available. The differences are as large as 0.8°C for some time periods. Although previous research suggests that the differences among these data sets are reduced at a global scale because of the introduction of more stations (Jones, 2016), I find that the latitudinal differences do not necessarily decrease through time, especially for high latitudes. For example, the difference between CRU-TEM4v and NOAA-NCEI at 70–90°N is approximately 0.6°C for the 21st century, which is much larger than those in the 1980s and 1990s. One possible reason for this pattern is the decline of available stations in certain regions (Figure 2-2). Additionally, different interpolation

methods also contribute to this difference. BEST-LAND and NASA-GISS assign large weights to very high latitude stations to represent unsampled regions in high latitude regions, while CRU-TEM4v and NOAA-NCEI tend to give larger weights to closer regions with more stations available.

For the regional mean LSAT, Figure 2-4 shows that the four datasets have the highest degree of agreement for North America, Europe, Asia, and Australia. For less-populated regions, such as South America, Greenland, and Antarctica, the differences are much larger than the others.

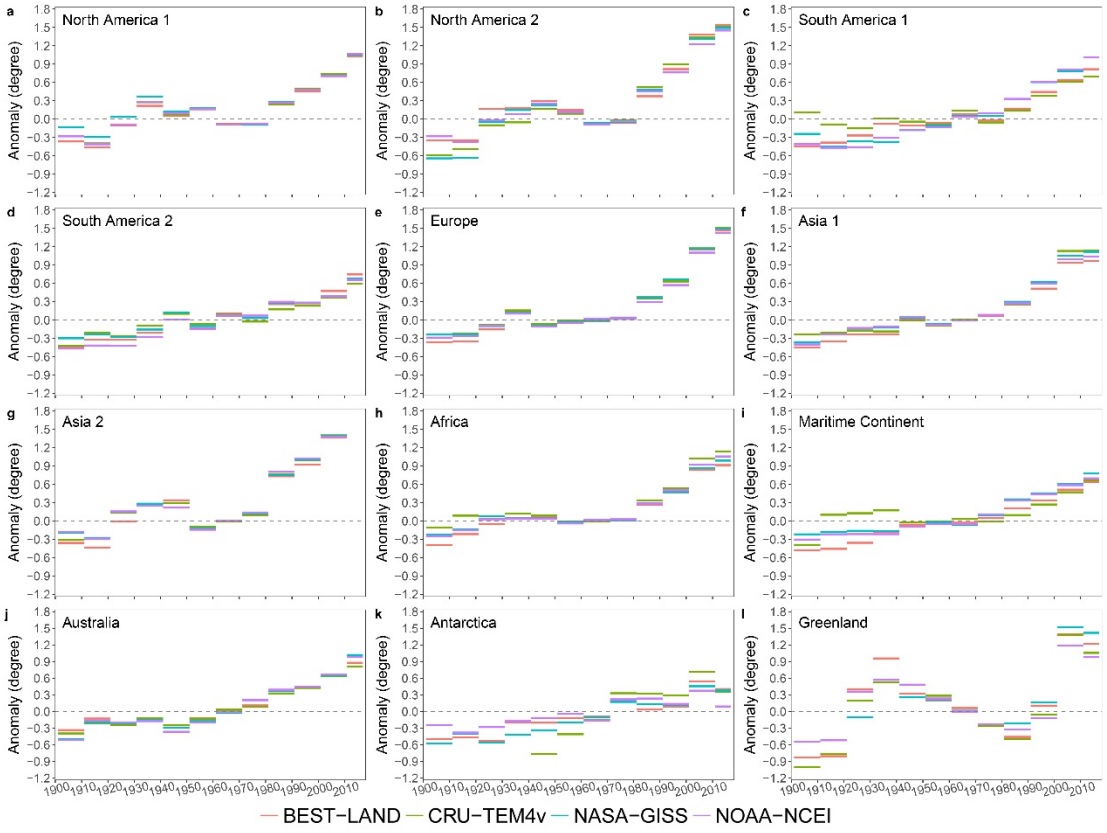

Figure 2-4. Same as Figure 2-3, but for different regions described in Table 2-2.

This pattern is highly correlated with the spatial distribution of available ground stations (Figure A1 in Hansen et al. (2010)). Similar with the latitudinal comparison, the differences among these datasets show notable increases in recent decades for most

continents, including high-latitude North America (NA2), South America, Asia, Africa, Maritime Continent, Australia, Antarctica, and Greenland.

Figure 2-5 shows the latitudinal and regional comparison using the Taylor diagram (Taylor, 2001). The annual mean LSAT anomaly time series are used in these comparisons. BEST-LAND has the best spatial coverage, and the values calculated from BEST-LAND are used as the reference for the diagram because no "true" values exist for the latitudinal/regional mean LSAT. For the latitudinal comparison (Figure 2-5a) the Taylor diagram confirms that high-latitude and polar regions have the lowest degree of agreement among these datasets, followed by the tropics. In addition, CRU-TEM4v and NOAA-NCEI tend to have larger deviations from the reference (i.e., BEST-LAND), whereas NASA-GISS has better agreement with BEST-LAND. For the regional comparison (Figure 2-5b), Antarctica, Greenland, Maritime Continent, and South America show the largest variations, while Europe, Asia, and low-to-mid latitude North America exhibit the best agreement among the four datasets.

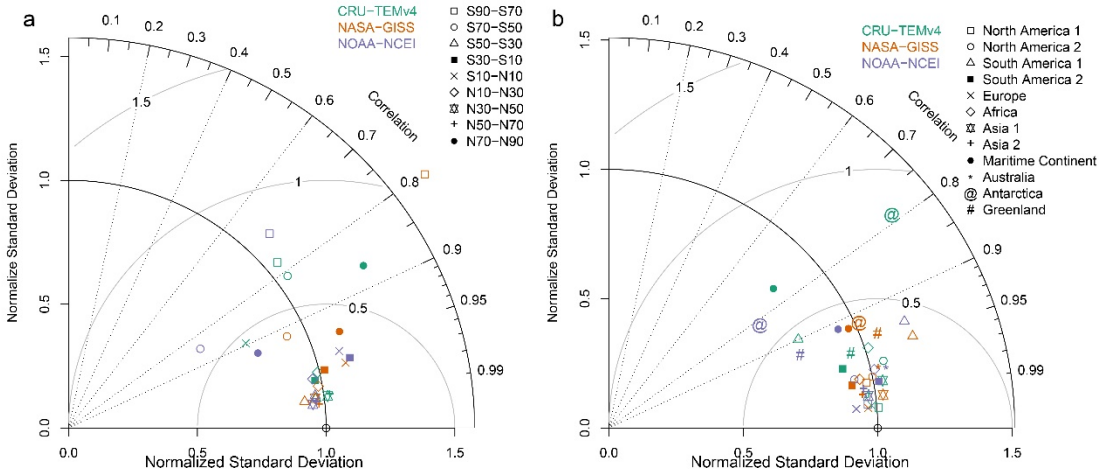

Figure 2-5. The normalized Taylor diagram for (a) latitudinal zones and (b) selected regions as described in Table 2-2 (both using BEST-LAND as the reference).

2.5.2 LSAT warming trend comparison for latitudinal bands

I compare the latitudinal/regional linear trends estimated using these datasets for different time periods (i.e., 1901–2017, 1951–2017, 1981–2017, and 1998–2017) to examine how the temperature differences among these datasets influence the surface warming trend analysis at the latitudinal/regional scales. Figure 2-6 illustrates a comparison of the annual and seasonal trends for different latitudes. All datasets generally show consistent latitudinal patterns of the LSAT trends. For the annual trends, the northern mid-high latitudes (i.e., 50–90°N) experience the highest warming rates for all time periods. In addition, the warming trends for this region accelerated in the recent decades, especially since 1981. However, the increasing warming trend does not exist in other latitudes. Some latitudes even experience smaller LSAT trends in recent decades. For instance, the LSAT trend of 30–50°N is not significant (around 0.1°C/decade) for 1998–2017, but it is above 0.2°C/decade for 1951–2017 and nearly 0.35°C/decade for 1981–2017. This slowdown of warming also happens for 90–50°S. The trend for other latitudes do not notably differ in the recent decades.

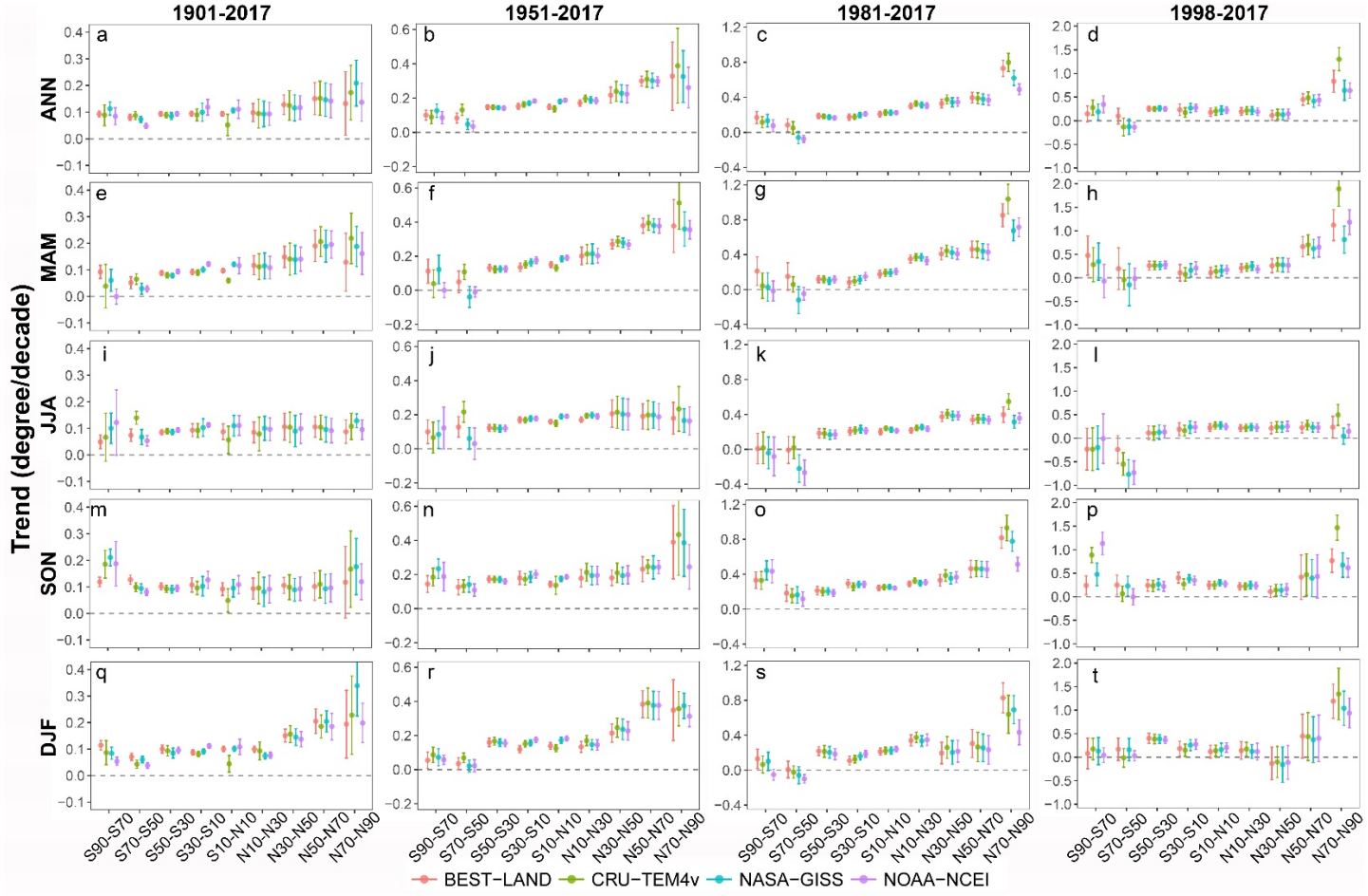

Figure 2-6. The comparison of the linear trends (unit: degree per decade) of the annual and seasonal mean LSAT for different latitudinal bands. The error bar around each point indicates the adjusted standard error of the linear trend estimation. The linear trends are calculated for different time periods.

The high latitudes (70–90°S/N) have the largest seasonal variations with a larger magnitude of warming during the cold seasons (i.e., DJF and MAM for the NH and JJA and SON for the SH) and a smaller magnitude of warming during the warm seasons. The "warming acceleration" for the northern-high latitudes exists in almost all seasons, except in the summer time (i.e., JJA). The southern-high latitudes (i.e., 90–70° S) only shows a significant warming trend during its spring season (i.e., SON). In addition, this region exhibits the "warming acceleration" phenomenon with the warming rate reaching almost 1°C/decade in 1998–2017, which is much larger than 0.3°C/decade for the other time periods. The LSAT trends of 30°S–50°N during DJF and MAM for 1998-2017 are insignificant or even negative (i.e., cooling) for these latitudinal bands. However, most of them are significantly positive (i.e., warming) for the other time periods.

Despite the general agreements among the four datasets, notable differences also exist for the LSAT trends across different latitudes and time periods. For the equator (i.e., 10°S–10°N), the LSAT trends estimated from CRU-TEM4v are consistently smaller by more than 50% than those of the other datasets for the whole study period (1901–2017). The LSAT trends for the equator of CRU-TEM4v and BEST-LAND for 1951–2017 are similar, but always smaller than the ones of NASA-GISS and NOAA-NCEI for all seasons. For the southern hemisphere mid-to-high latitudes (i.e., 70–50°S), BEST-LAND has the highest LSAT trends followed by CRU-TEM4v and NOAA-NCEI. Meanwhile, NASA-GISS has the lowest LSAT trend for 1901–2017. The relative difference of the LSAT trend for this region between BEST-LAND and NASA-GISS ranges from 62% to 93% for different seasons. For DJF and MAM, the sign of

the estimated LSAT trends for this region differs across datasets. This inter-dataset differences also occur during the time period from 1951 to 2017. Moreover, the estimated LSAT trends for the high latitudes in both hemispheres differ across datasets. CRU-TEM4v and BEST-LAND usually have larger estimated trends than NASA-GISS and NOAA-NCEI for the northern high latitude (i.e., 70–90°N) during most seasons and time periods. In the recent decades (i.e., 1981–2017 and 1998–2017), the trend differences across different datasets become more remarkable, especially for the annual trend and the seasonal trends of summer and fall (i.e., JJA and SON). The largest relative differences reach 48%, 61%, and 51% for the annual, summer, and fall LSAT trends during 1981–2017.

Using BEST-LAND as the reference, I test the significance of trend differences for different time periods and different seasons (Table A2). CRU-TEM4v has consistently significant lower trends for both annual and seasonal mean LSATs during 1901-2017 over Equator (i.e., S10-N10). During 1951-2017, CRU-TEM4v shows significantly higher trends over northern hemisphere (i.e., N10-N70) for annual mean LSATS, while this pattern expands to high-latitudes for fall season (i.e., SON). In the recent decades (1981-2017 and 1998-2017), CRU-TEM4v shows significant positive trend differences for northern latitude bands for most seasons and negative trend differences for southern high latitudes (i.e., S90-S50). Meanwhile, NASA-GISS shows significant negative trend differences during 1901-2017 for most southern latitudinal bands. It also presents significant trend differences for northern latitudes during summer time (i.e., JJA). However, the significant trend differences between NASA-GISS and BEST-LAND expand to much broader regions from 1951 over the tropics and subtropics (i.e., S30-

N30). NOAA-NCEI shows the similar pattern to NASA-GISS but with larger magnitude of trend differences. In addition, during 1901-2017, NOAA-NCEI also appears to have significant trend differences during MAM season for S30-N50 latitudinal belts. This trend difference analysis indicates that choice of different datasets leads to significant differences for warming trend calculation.

2.5.3 LSAT warming trend analysis for the selected regions

Similar with the latitudinal band analysis, I also compare the LSAT trends estimated using different datasets of different time periods for the predefined regions (Table 2-2). The estimated LSAT trends generally have the highest degree of agreements over North America, Europe, and Asia, while they notably differ over South America, Africa, Maritime Continent, Australia, Antarctica, and Greenland (Figure 2-7). This pattern generally agrees with the spatial distribution of the ground stations commonly used for generating these datasets.

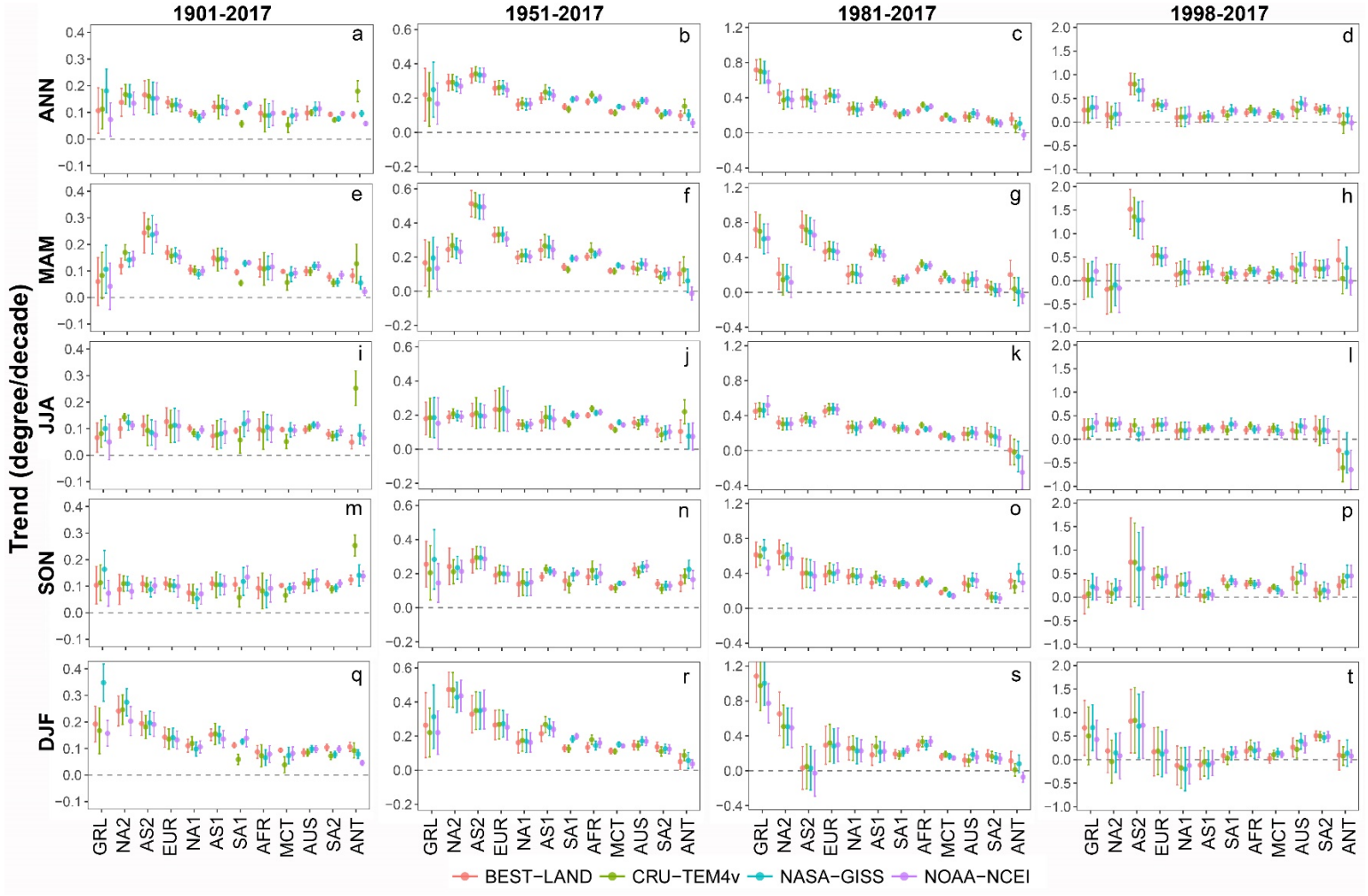

Figure 2-7. Same as Figure 2-6, but for different regions described in Table 2-2.

For South America, the LSAT trends estimated using NASA-GISS and NOAA-NCEI are consistently higher than those of BEST-LAND. Meanwhile, CRU-TEM4v has the lowest LSAT trends for the long-term analysis (i.e., 1901–2017 and 1951–2017). The differences are the largest over tropical South America (i.e., SA1), Greenland (GRL) and Antarctica (ANT). A large part of the tropical South America is covered by dense forests, in which setting up ground stations is difficult and leads to a large uncertainty of the observation-based datasets because of the substantial data gaps (Frenne and Verheyen, 2016).

Trends estimated from four datasets differ notably over Antarctica for both annual and seasonal mean LSATs. The most notable difference for the annual trends is during 1981–2017 when BEST-LAND suggests that Antarctica is warming at a rate of 0.2°C/decade, while other datasets all indicate that it experiences a pause of warming or even slight cooling (insignificant) during the last four decades. For the fall season over Antarctica (i.e., MAM), the LSAT trends estimated from different datasets often disagree with others in terms of whether Antarctica is warming or cooling for different time periods.

For Greenland, these four datasets have a large discrepancy for both annual and seasonal trend calculation. NASA-GISS has the highest LSAT trends for the long-term (i.e., 1901–2017 and 1951–2017) for all seasons, while this pattern does not persist in recent decades (i.e., 1981–2017 and 1998–2017). For the spring time (i.e. MAM) during the whole study period (i.e., 1901–2017), NASA-GISS presents an evidence of a warming Greenland with a significant warming rate of 0.15°C/decade, while NOAA-NCEI shows no significant warming during the same periods.

Using BEST-LAND as reference, CRU-TEM4v shows significantly different trends over high latitude North America (NA2), Europe (EUR), and South America (SA1, SA2) during 1901-2017. More regions, including the lower part of Asia (AS1), Africa (AFR), maritime continent (MCT) and Antarctica (ANT), also show significant trend differences after 1951, especially since 1981 (Table A3). Although NASA-GISS has better agreement with BEST-LAND in terms of absolute anomalies, it also appears to have significant trend differences for lower part of Asia, South America, Africa, maritime continent, and Antarctica mostly after 1951. Surprisingly, NASA-GISS has significant negative trend differences comparing with BEST-LAND over lower part of North America (NA1) during 1901-2017, which is a region with abundant station observations. The substantial trend difference is likely the combined impact of different station data (BEST uses a much larger station data archive than CHCNMv3) and different homogenization methods. Further research is necessary to further detangle contributions of individual factors. For NOAA-NCEI, it also exhibits significant trend differences with BEST-LAND over South America, Africa, high latitude regions in Asia and North America, Europe, and Antarctica.

2.6 The comparison of grid-box LSAT and local trend analysis

Existing studies rarely examine the LSAT difference across datasets at the local scale and its impact on local trend calculation. I present herein the comparison results for both the LSAT and estimated trend at the grid-box scale.

2.6.1 Grid-box LSAT comparison

Figure 2-8 demonstrates the coefficients of variations (COV) for the LSATs of the four datasets. I show the temporal evolution of the LSAT variations across datasets by separating the whole study periods into six 20-year periods (i.e., 1901–1920, 1921–1940, 1941–1960, 1961–1980, 1981–2000, and 2001–2017). Figure 2-8 (a–f) shows the spatial pattern of the COV for different time periods, while Figure 2-8 (g–l) demonstrate the histogram of the COV for each time period. The COV is only calculated for the land grid-boxes, where the four datasets have valid values during the time period.

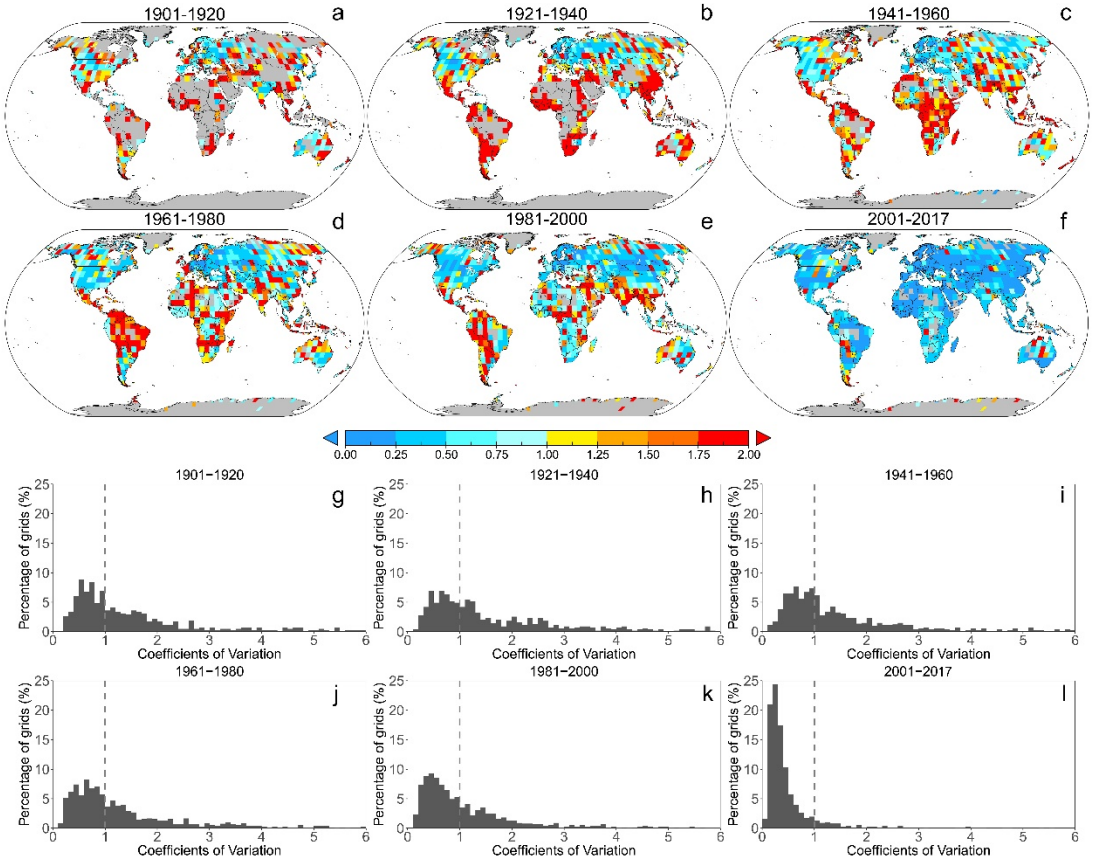

Figure 2-8. The spatial pattern of the coefficients of variation (COV) of the annual mean LSAT between four data sets for different time periods (i.e., 1901–1920, 1921–1940, 1941–1960, 1961–1980, 1981–2000, and 2001–2017) (a–f). Only common data

coverage areas are shown in the map; and the corresponding histograms of the COV of the annual mean LSAT between the four data sets for different time periods (g-l).

The continental United States and Europe have the lowest LSAT variations across datasets for all time periods, while other parts of the world experience gradual decreases of the LSAT variations across datasets through time. The decrease of the LSAT variations for most parts of the land, including South America, Africa, and majority of Asia, happens during 1981–2000, which might have been a result of the introduction of new ground stations worldwide. However, some regions still have relatively large variations since 1981 despite more stations being used for these datasets, including the west part of South America, Sahel, Indian subcontinents, Southeast Asia as well as the west and central Australia. The variation drops substantially in the last two decades (i.e., 2001–2017) for most regions. Only a small portion of the grid-boxes over central Australia, central Africa, and high latitudes still have large variations. This temporal evolution is clearly captured by the leftward shifting of the histograms across different time periods (Figure 2-8 g–l).

2.6.2 Grid-box LSAT trend comparison

Figure 2-9 shows the spatial pattern of the annual LSAT trend for the common data coverage areas during different time periods (i.e., 1901–2017, 1951–2017, 1981–2017, and 1998-2017). BEST-LAND has the highest spatial resolution and full data coverage; hence, I use the trends estimated from BEST-LAND as the references because of the lack of "true values" of LSAT trends.

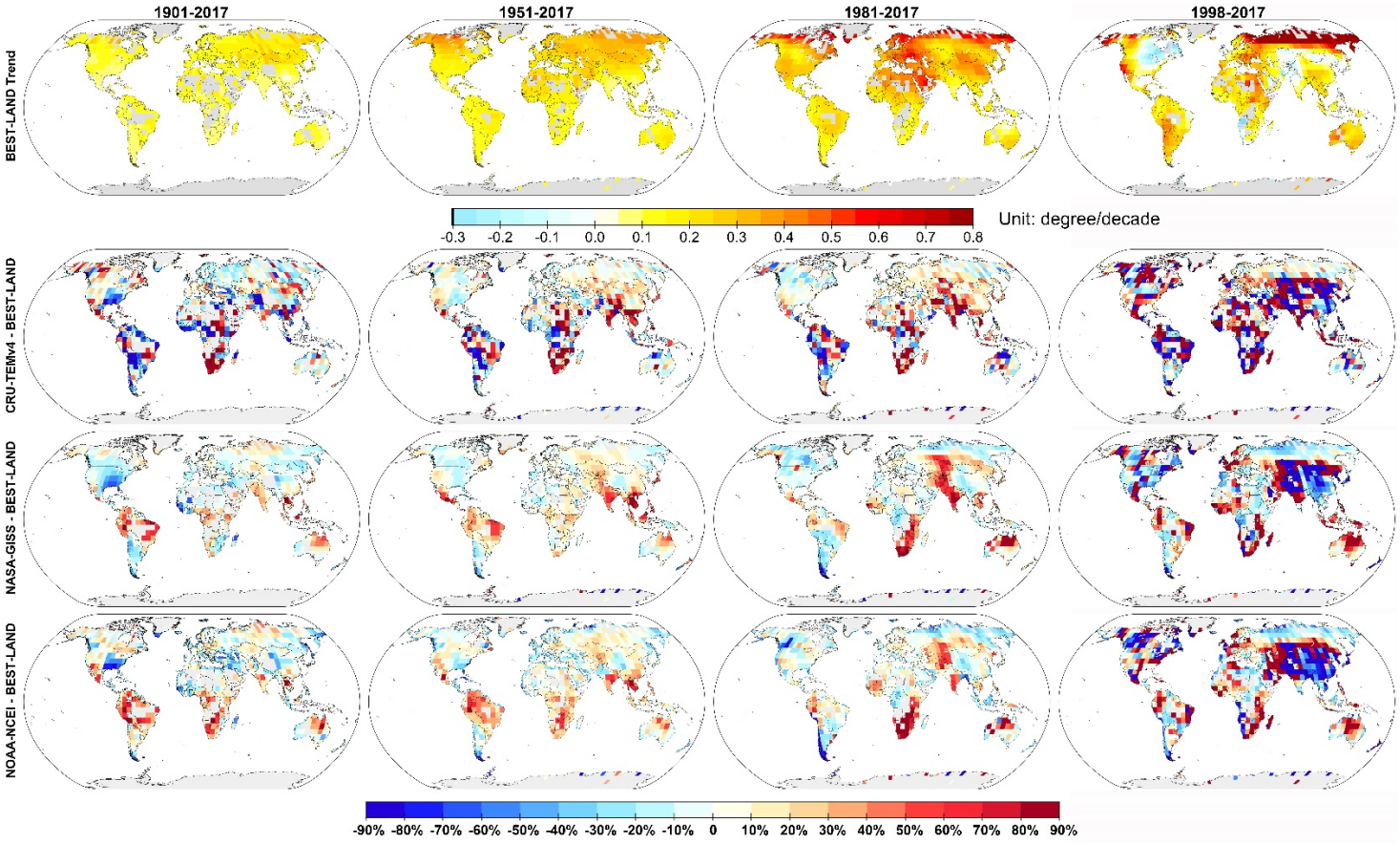

Figure 2-9 Upper panel: the spatial pattern of the linear trends (unit: degrees per decade) of the annual mean LSAT of BEST-LAND for 1901–2017, 1951–2017,1981–2017, and 1998-2017. Bottom panel: relative trend differences of the annual mean LSATs for CRU-TEM4v, NASA-GISS, and NOAA-NCEI compared to BEST-LAND over 1901–2017, 1951–2017, 1981–2017, and 1998-2017. Only common data coverage areas are shown in the maps.

The upper panel in Figure 2-9 shows a clear spatial pattern of the surface warming for different time periods based on BEST-LAND. Northern mid-to-high latitudes experienced the highest rate of surface warming. The surface warming for majority of the land grid-boxes has accelerated since 1951, with the most profound acceleration occurring at Europe, North Africa, North China, Central Asia, and Siberia. However, certain regions experience smaller warming trends or even cooling trends in the recent two decades (1998-2017). These regions include northeast part of North America, southwest tip of Africa continent, central Asia, and northern China. The spatial patterns of the surface warming based on other datasets (i.e., CRU-TEM4v, NASA-GISS, NOAA-NCEI) are similar with the one of BEST-LAND (not presented here).

The lower panel of Figure 2-9 shows the relative difference maps of the annual trends of CRU-TEM4v, NASA-GISS, and NOAA-NCEI compared to BEST-LAND. CRU-TEM4v shows the largest differences compared to BEST-LAND, especially over station-sparse regions, such as Africa. The relative trend difference can even reach 95% (in the central part of Africa). Other regions, such as South America, high latitudes in North America, and Asia also show large relative differences over different time periods. The large discrepancy of the LSAT trends occurred at Africa and South America.

For NASA-GISS and NOAA-NCEI, they both show relatively small trend differences over a long period (1901-2017, 1951-2017). But station sparse regions like tropical South America, Africa and north Australia demonstrate large trend differences reaching 50%. In recent decades (1981-2017), trend differences increase remarkably over central Asia, India, southern Africa continent and north Australia (more than 80%). For the

past two decades, due to relative short data length, the estimated trends show substantial differences across all datasets.

Figure 2-10 demonstrates the scatter plots of the annual trend comparison among different datasets for different time periods using BEST-LAND as the reference. CRU-TEM4v appeared to be the most inconsistent with BEST-LAND with the lowest correlation coefficients (R), largest mean differences (MD), and largest root mean square differences (RMSD) for all time periods. In contrast, NASA-GISS has the highest degree of agreement with BEST-LAND for the annual trend supported by the highest R and smallest RMSD for all time periods. For some grid-boxes, different datasets disagree on the sign of estimated LSAT trends. The degree of this disagreement increases in recent decades because of the large standard errors of the estimated trends caused by short data records and large inter-annual variability.

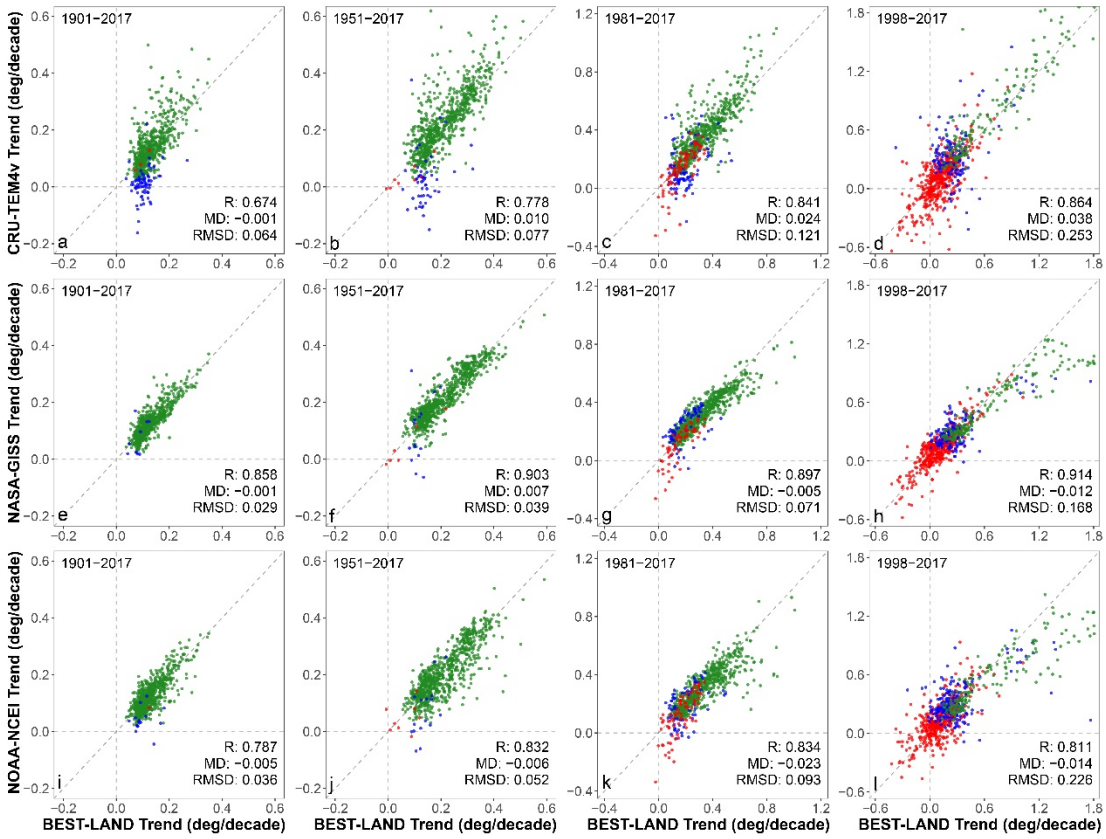

Figure 2-10 The scatterplots of the linear trend of the annual mean LSAT for different time periods (i.e., 1901–2017, 1951–2017, 1981–2017, and 1998-2017) for CRU-TEM4v, NASA-GISS, and NOAA-NCEI compared to BEST-LAND. Red points indicate the estimated trends are not significant for either dataset; blue points represent that only one estimated trend is significant; green points indicate both estimated trends are significant.

Table 2-3 presents the detailed statistics for the annual and seasonal trend comparison across datasets using BEST-LAND's estimations as the reference. CRU-TEM4v generally has the largest differences compared to BEST-LAND, while NASA-GISS has the smallest differences. For the seasonal trend comparison, all datasets seem to have a higher degree of agreement for MAM and DJF for all datasets. Despite the high correlation coefficients in the recent decades, the variations of the LSAT trends increase with time, which are shown by the increasing RMSD values for both annual and seasonal LSAT trends for all datasets. The RMSD for CRU-TEM4v increases from

Table 2-3. Statistics of the estimated trend differences (unit: degrees per decade) among CRU-TEM4v, NASA-GISS, and NOAA-NCEI using BEST-LAND as the reference for 1901–2017, 1951–2017, 1981-2017, and 1998–2017. (R: correlation coefficient; MD: mean difference; RMSD: root mean square difference; ANN: annual; MAM: March–April–May; JJA: June–July–August; SON: September–October–November; and DJF: December–January–February).

		CRU-TEM4v			NASA-GISS			NOAA-NCEI		
		R	MD	RMSD	R	MD	RMSD	R	MD	RMSD
1901–2017	ANN	0.674	-0.001	0.064	0.858	-0.001	0.029	0.787	-0.005	0.036
	MAM	0.781	0.000	0.069	0.923	0.000	0.033	0.872	-0.006	0.044
	JJA	0.602	-0.001	0.065	0.752	0.000	0.035	0.633	-0.001	0.042
	SON	0.689	-0.001	0.066	0.842	-0.002	0.032	0.746	-0.004	0.042
	DJF	0.761	-0.001	0.078	0.906	0.001	0.038	0.857	-0.006	0.051
1951–2017	ANN	0.778	0.010	0.077	0.903	0.007	0.039	0.832	-0.006	0.052
	MAM	0.861	0.004	0.084	0.943	0.005	0.047	0.895	-0.013	0.065
	JJA	0.610	0.008	0.082	0.768	0.004	0.047	0.660	-0.004	0.058
	SON	0.783	0.016	0.086	0.899	0.009	0.042	0.754	-0.001	0.064
	DJF	0.855	0.012	0.096	0.939	0.010	0.056	0.871	-0.005	0.082
1981–2017	ANN	0.841	0.024	0.121	0.897	-0.005	0.071	0.834	-0.023	0.093
	MAM	0.911	0.014	0.127	0.942	-0.008	0.091	0.899	-0.029	0.123
	JJA	0.772	0.027	0.127	0.839	0.003	0.089	0.796	-0.006	0.105
	SON	0.875	0.030	0.143	0.930	-0.001	0.085	0.823	-0.024	0.130
	DJF	0.838	0.020	0.162	0.928	-0.012	0.092	0.833	-0.027	0.137
1998–2017	ANN	0.864	0.038	0.253	0.914	-0.012	0.168	0.811	-0.014	0.226
	MAM	0.922	0.021	0.266	0.957	-0.015	0.216	0.895	-0.033	0.292
	JJA	0.620	0.052	0.253	0.648	-0.002	0.196	0.557	0.014	0.221
	SON	0.846	0.049	0.307	0.897	0.000	0.192	0.730	-0.011	0.285
	DJF	0.893	0.010	0.345	0.936	-0.030	0.231	0.858	-0.022	0.323

(0.064, 0.078) °C/decade for 1901–2017 to (0.253, 0.345) °C/decade for 1998–2017. The RMSD for NASA-GISS also increases from (0.029, 0.038) °C/decade (1901–2017) to (0.068, 0.216) °C/decade (1981–2017).

2.7 Discussion and conclusion

In this study, I thoroughly examine the differences of four LSAT datasets (i.e., CRU-TEM4v, BEST-LAND, NASA-GISS, and NOAA-NCEI) at different spatial and temporal scales and their potential impacts on the trend calculation. The data coverage used for calculating the large-scale mean LSAT at global and hemispheric scales has a strong impact on the final time series. For the global annual mean LSAT, different data coverages introduce an LSAT anomaly difference at the magnitude of 0.15°C. This difference is even larger for different seasons (i.e., 0.4°C for DJF and MAM and 0.2°C for JJA and SON). For the hemispheric mean LSAT, the anomaly differences caused by differences in data coverages are nearly 0.6°C and 0.3°C for the cold and warm seasons, respectively.

The mean LSAT differences at the latitudinal scale are most prominent at high latitudes (e.g., 70–90°N and 50–90°S) and at the equator (i.e., 10°S–10°N). The decadal mean LSAT differences are as large as 0.6°C and 0.3° for high latitudes and the equator, respectively. These large differences lead to notable differences for the LSAT trend estimation. The relative difference of the LSAT trend for the high latitudes and the equator ranges from 35% to 60% for different seasons and time periods. Meanwhile, the LSAT differences across datasets for the southern high latitudes cause different signs for the estimated LSAT trends for the recent decades. This notable disagreement would lead to different conclusions of whether the warming of the region is accelerating

or slowing down, which is essential for understanding the status of the current climate change and its impacts on the ecosystem and the society.

At the regional scale, these datasets have the highest degree of agreement over North America, Europe, and Asia, but notably differ over South America, Antarctica, Africa, and Maritime Continent. This difference may be attributed to the skewed distribution of the ground station data used to generate these datasets and different interpolation methods. Most stations are clustered in regions that are more developed and populated. The regional mean LSAT differences are nearly 0.4°C even after 2000. As a result, the trends estimated from different datasets differ substantially for those regions. The relative trend difference across datasets ranges from 28% to 93% over different regions and time periods. The LSAT differences across datasets for Antarctica and Greenland even cause different signs of the estimated trends, thereby leading to contrasting conclusions on these most vulnerable regions.

At local scale, our analysis show that the regions with the largest LSAT variation across datasets are grid-boxes over South America, Africa, Indian subcontinent, central and northern Australia, southeast Asia, and Siberia, which was consistent with the regional-scale analysis. The LSAT variation across datasets decreases with time, with the lowest variation among datasets seen after 2000. For the trend analysis, the largest variation of the trends often occurs at the grid-boxes with the largest LSAT variation across datasets. The relative difference of trends estimated from different datasets can reach nearly 90% over different regions and time periods. CRU-TEM4v generally appears to have the largest grid-box scale differences, while NASA-GISS has the smallest differences compared to BEST-LAND. The uncertainty of the LSAT trend estimation

caused by the dataset differences (i.e., RMSD) ranges from 0.035 to 0.086°C/decade for the long-term trend (i.e., 1901–2017) to 0.097–0.305°C/decade for recent decades (i.e., 1981–2017).

Based on the previous comparison across different scales, the dataset differences strongly depend on the station data availability at any scales. Indeed, the main challenge of generating observation-based LSAT data sets is to obtain homogenized station observations and ingest them into the final product. The stations used for each dataset vary significantly because of different quality control procedures and raw data sources (Table 2-2). Examining the variation of the number of stations used by each dataset over different regions and grid-boxes would further advance the understanding of the dataset variation. CRU provides the number of stations used in each grid-box to generate CRU-TEM4v, which is used herein as a proxy since CRU uses least amount of stations in the final dataset.

Figure 2-11 demonstrates the coefficients of variation (COV) of the grid-box LSAT trend for different time periods and its relationship with the number of stations available in each grid-box. The box plots showed that the spread and the mean value of the COV when there are less than 5 stations in the entire 5°×5° grid-box were significantly larger than the grid-boxes with more stations. This large cross-dataset variation could be expected because using a very limited number of stations to capture the full LSAT dynamic of LSAT across such a large area (e.g., approximately 500 km × 500 km at low latitudes), especially over regions with complex topography or heterogeneous landscapes, is very challenging (EDW, 2015). Additionally, the within-grid-box distribution can also contribute to this variation since interpolation methods used by

BEST-LAND and NASA-GISS tend to give less weights to the clustered stations while they give more weights to the isolated stations in the grid-box.

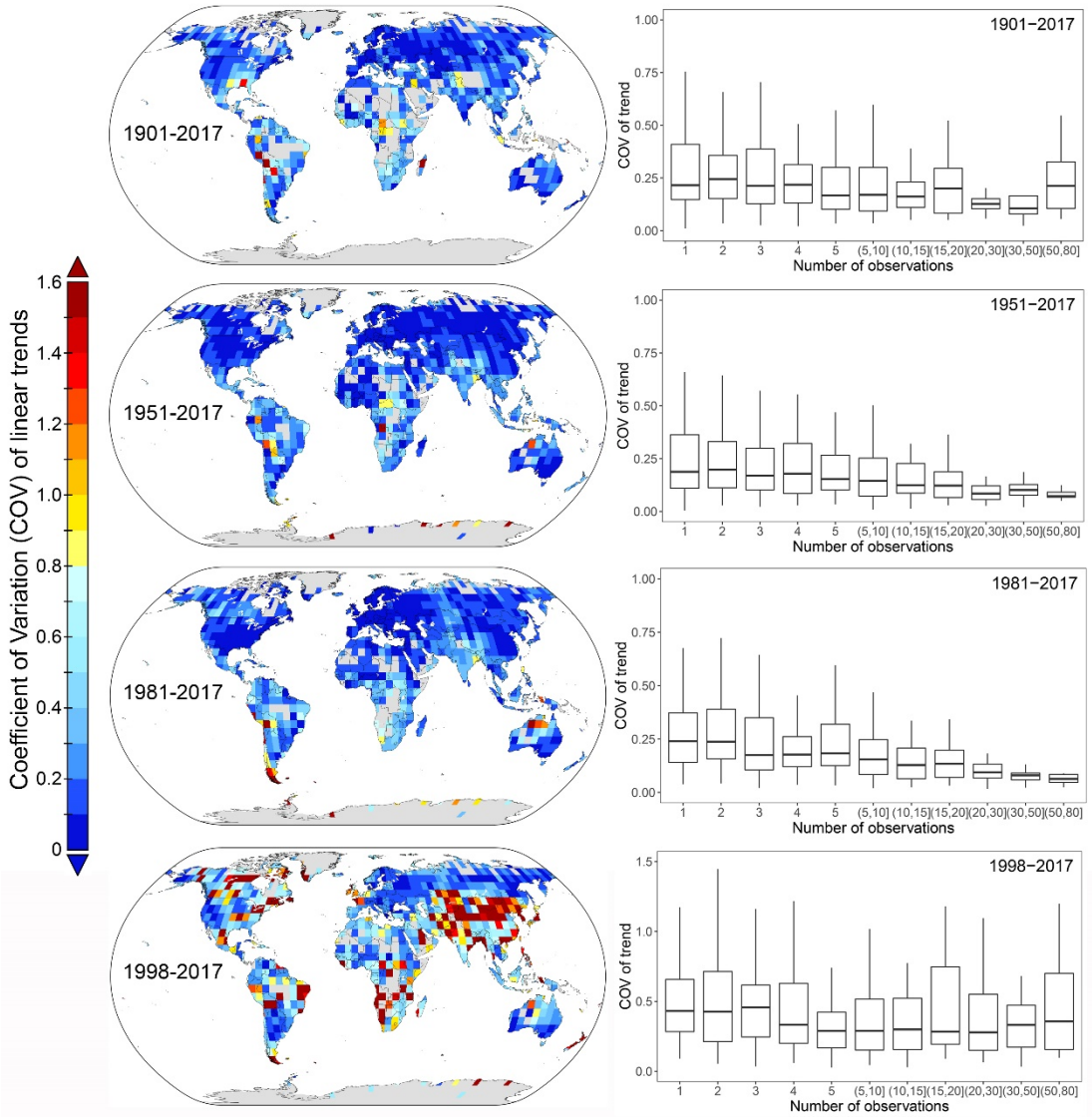

Figure 2-11. Left panel: the spatial patterns of the linear trend of the annual mean LSAT for different time periods (i.e., 1901–2017, 1951–2017, 1981–2017, and 1998-2017) among the four datasets. Gray areas are grids, where at least two datasets do not have significant trends for that time periods. Right panel: corresponding box-plots of the coefficients of variation of the linear trend for the grid-boxes grouped by the number of stations available in the grid boxes used by CRU-TEM4v.

This limitation of the observation-based data sets needs to be addressed to increase the confidence of climate change studies based on these data sets. One way to address this

issue is to improve the design and implementation of the global station network. Different international initiatives have already started the process of improving the density and the quality of station measurements, such as the International Surface Temperature Initiative (ISTI, Rennie et al. (2014)) and the new network implementation plan described by the Global Climate Observing System (Thorne et al., 2017b; WMO, 2016).

Although continuing this improvement is critical and necessary, doing so is time and resource consuming. Moreover, the improvement would mostly benefit the data set for the future, which cannot directly reduce the variations across datasets for the past. In contrast, remote sensing data are in a unique niche to provide nearly spatial-complete information over land surface. Several off-the-shelf global surface temperature products at various spatial resolutions for the recent decades (since the 1980s) are currently available, including both land surface temperature and air temperature profiles. Remotely sensed products suffer from their own limitations, such as the observation time change across satellites and warm biases due to cloud contamination. On the other hand, atmospheric reanalysis data have also been very popular in various applications. Despite its known uncertainties, reanalysis data provide valuable complete global temperature data at various resolutions. With appropriate statistical methods, combining global station network observations, reanalysis temperature data and remotely sensed temperature products to generate a spatial-complete LSAT data is possible. Efforts in this aspect are already ongoing, which could significantly benefit climate change studies requiring LSAT data sets (Merchant et al., 2013; Thorne et al., 2017b).

Chapter 3 Estimating the Near Surface Air Temperature of the Tibetan Plateau

3.1 Summary

The Tibetan Plateau (TP) has experienced rapid warming in recent decades. However, the meteorological stations of the TP are scarce and mostly located at the eastern and southern parts of the TP where the elevation is relatively low, which increases the uncertainty of regional and local climate studies. Recently, the remotely sensed land surface temperature (LST) has been used to estimate the surface air temperature (SAT). However, the thermal infrared based LST is prone to cloud contamination, which limits the availability of the estimated SAT. This study presents a novel all sky model based on the rule-based Cubist regression to estimate all sky daily SAT using LST, incident solar radiation (ISR), top-of-atmosphere (TOA) albedo and outgoing longwave radiation (OLR). The model is trained using station data of the Chinese Meteorological Administration (CMA) and corresponding satellite products. The output is evaluated using independent station data with the bias of -0.07 °C and RMSE of 1.87°C. Additionally, the 25-fold cross validation shows a stable model performance (RMSE: 1.6-2.8 °C). Moreover, the all sky Cubist model increases the availability of the estimated SAT by nearly three times. I used the all sky Cubist model to estimate the daily SAT of the TP for 2002-2016 at 0.05°×0.05°. I compared our all sky Cubist model estimated SAT with three reanalysis datasets (i.e., GLDAS, CLDAS, CMFD). Our model estimation shows similar spatial and temporal dynamics with these existing data but outperforms them with lower bias and RMSE when benchmarked against CMA station data. The estimated SAT data could be very useful for regional and local climate

studies over the TP. Although the model is developed for the TP, the framework is generic and may be extended to other regions with proper model training using local data.

3.2 Background

The Tibetan Plateau (TP) is the world's highest plateau in central Asia with an average elevation higher than 4000 meters above sea level (ASL) (Figure 3-1(a)) (Yang et al., 2014). As the world's "Third Pole", it is the origins of major rivers in Asia and regulates regional and global weather patterns (Yao et al., 2018). Many previous studies reported that the Tibetan Plateau, similar with other high mountainous areas, has experienced more rapid surface temperature change comparing to many other parts of the world, especially after early 1950s. The reported warming exists for both mean, minimum, and maximum surface air temperatures (SAT), leading to the decreasing diurnal temperature range of the TP (Duan and Xiao, 2015; Li et al., 2005; Liu et al., 2009; Liu and Chen, 2000; Yang and Ren, 2017). As a consequence of the SAT change, the TP has shown remarkable changes of its cryosphere, hydrological cycles, and ecosystems. For example, Shen et al. (2015) reported that the snow cover of the TP has reduced by 5.7% during 1997-2012; Yang et al. (2014) demonstrated that the central TP experiences more convective precipitation and more surface runoff while the southern and eastern regions experience reduction in both precipitation and surface run off in recent decades; multiple studies observed that the vegetation activity shows strong response to surface temperature change over the TP (Cao et al., 2018; Cong et al., 2017; Shen et al., 2015).

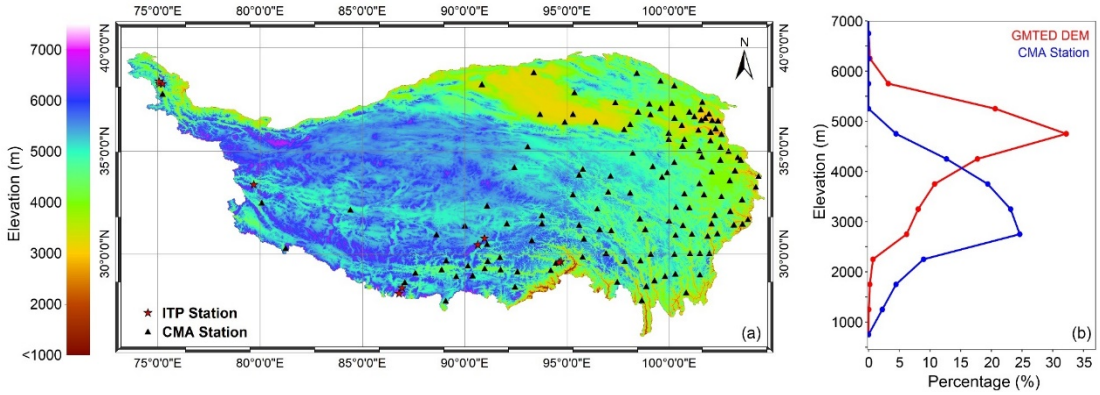

Figure 3-1. (a) The elevation map of the Tibetan Plateau and the location of the China Meteorological Administration (CMA) stations (black triangles) and the Institute of Tibetan Plateau Research (ITP) stations (red pentagram) within the Tibetan Plateau; (b) the elevation distributions of the CMA stations (blue line) and the GMTED DEM for the entire TP (red line).

However, previous studies heavily rely on SAT data measured by unevenly distributed meteorological stations (Figure 3-1 (a)). Figure 3-1 (b) shows the elevation distribution of the CMA stations and the elevation distribution of the radar-based digital elevation model (DEM) of the entire TP. Over 70% of the CMA stations are located at relatively low elevation (< 4,000 meters ASL) and the eastern part of the TP, while almost no operational CMA stations are placed beyond 5000 meters ASL. The sparse and biased station samples may increase the uncertainty of local and regional climate analysis and corresponding impact studies (Pepin et al., 2015; Rao et al., 2018; Yao et al., 2018). Alternatively, previous studies also use several spatially complete SAT datasets which are produced through either data interpolation or data assimilation, such as, Global Land Data Assimilation System (GLDAS) data, Chinese Meteorological Forcing Data (CMFD), Chinese Land Data Assimilation System (CLDAS), and Climate Research Unit (CRU) high resolution climate dataset. These datasets provide important information for the regions with no station measurements. However, the spatial resolutions of these datasets are relatively coarse, which may cause large uncertainty

in applications, especially for the region of the TP with complex terrain (An et al., 2018). Additionally, these datasets, developed at global or national scales, have not been comprehensively validated for the TP and usually have large uncertainty associated with the methods or land surface models used during their production.

Meanwhile, remotely sensed land surface temperature (LST) has been widely used to study regional and global climate change due to its strong correlation with SAT and its global coverage from multiple satellite missions (Good et al., 2017; Pepin et al., 2016). Using monthly LST data of Moderate resolution Imaging Spectroradiometer (MODIS), Qin et al. (2009) reported that the warming rate of the TP has shown notable dependency on the elevation during 2000-2006. Despite the strong correlation between LST and SAT, they are two distinct variables with different physical definitions. To overcome this limitation, researchers have developed various methods to estimate SAT using LST of various sensors, such as, MODIS (F. Huang et al., 2017; Lu et al., 2018; Zhang et al., 2016), Spinning Enhanced Visible and Infrared Imager (SEVIRI) (Good, 2015), and Advanced Very High Resolution Radiometer (AVHRR) (Prince et al., 1998). Most of these methods are based on linear regression using LST and other auxiliary input, such as, land cover, surface roughness, day length, and evapotranspiration (Good, 2015; F. Huang et al., 2017; Meyer et al., 2016; Noi et al., 2017; Zhang et al., 2016). Recently, studies have also explored different machine learning (ML) models (e.g., support vector machine, artificial neural network, random forest, Cubist regression, etc.) to estimate SAT using LST (Meyer et al., 2016; Noi et al., 2017; Zhang et al., 2016). Generally, the ML models perform better than linear regression models because ML models can better capture the complex relationship

between LST and SAT. Besides ML models, spatiotemporal interpolation methods, such as, geographically weighted regression, hierarchical Bayesian model, and kriging regression, have also been used to generate high resolution SAT using LST and other auxiliary inputs (Chen et al., 2014; Li et al., 2018; Lu et al., 2018).

Despite the recent progress, the LST-based SAT estimation still suffers a major limitation caused by the cloud contamination. Since most of current LST data are derived from thermal infrared data, the LST data are unavailable when cloud exists during satellite overpassing time. The cloud contamination has strong impacts on the availability and the quality of the SAT estimated using existing methods. Noi et al. (2017) reported that using four instantaneous MODIS LSTs (i.e., daytime and nighttime LSTs of both Terra MODIS and Aqua MODIS products) can accurately estimate daily SAT with the root-mean-squared-error (RMSE) less than 2 K. However, the RMSE of the estimated SAT increases (larger than 3 K) when cloud contamination occurs (Noi et al., 2017). To account for cloud contamination, Zhang et al. (2016)'s framework estimate the daily SAT by dynamically integrating available MODIS LSTs based on their quality. Although Zhang el al. (2016)'s method can increase the availability of the estimated SAT, it still requires at least one high quality clear sky LST and the estimated SAT has different levels of uncertainty due to the changing availability and quality of MODIS LSTs (ranging from 1.5 to 3.5 K) (Zhang et al., 2016). Zhang et al. (2008) reported that the annual average cloud coverage of the TP ranges from 40% - 60% during 1971-2004. The frequent cloud contamination can have serious implications on the quality and availability of the estimated SAT using existing methods.

The main objective of this study is to develop a method that can produce daily SAT of the TP with relatively high resolution (i.e., 0.05°×0.05°) that are not or less prone to frequent cloud contamination. Different from existing studies, I propose a ML model to estimate daily SAT using all available LSTs and remotely sensed radiation variables at both the surface and top-of-atmosphere (TOA) levels. These radiation variables are available for both clear sky and cloudy sky conditions and contain important information about surface energy exchange. Theoretically, the surface energy exchange regulates SAT and its difference with LST. Thus, including these radiation variables may help capturing the physical process of surface heat exchange thus improving the model performance. In this study, I choose the rule-based Cubist regressing (hereafter referred as Cubist) as our base model since previous studies all reported that the Cubist has the best performance on estimating SAT using LSTs over different regions including the TP (Noi et al., 2017; Zhang et al., 2016). To robustly estimate the all sky SAT, I also compare two different strategies using 1) one generic model for both clear sky and cloudy sky conditions or 2) two separate models for clear sky and cloudy sky conditions separately. To the best of our knowledge, this study is the first study using machine learning models to estimate daily all sky condition SAT with remotely sensed products. The estimated all sky SAT dataset can be very important for climate analysis and relevant impact studies for the TP. The structure of this manuscript is organized as follow: section 2 describes the data and necessary data processing used in this study; section 3 summarizes the overall research method, Cubist regression model, and the evaluation strategies of this study; model training and validation results are reported in section 4, while section 5 describe the results of cross comparison with existing

datasets; section 6 discusses the advantages and limitations of this study while the conclusion is presented in section 7.

3.3 Data

The data used in this study include 1) the station measured SAT for model training and evaluation, 2) the SAT of various reanalysis/forcing datasets for cross comparison, and 3) the remotely sensed variables as the model inputs (i.e., elevation, LST, surface variables and radiation variables). Each category of the data (i.e., station data, remotely sensed data, and reanalysis/forcing data) is further described in the corresponding subsections with the details that are meaningful to this study.

3.3.1 Station data

In this study, the station measured SAT data are used for both model training and evaluation (Table 3-1). The main source of the daily average SAT used in this study is 135 meteorological stations of the TP managed by the Chinese Meteorological Administration (CMA). The data between 2002 and 2015 were downloaded from the CMA's National Meteorological Information Center (NMIC) (http://data.cma.cn). Additionally, I also collected daily average SAT of 10 individual experiment stations managed by different research groups of the Institute of Tibetan Plateau Research (ITP) to independently evaluate the Cubist model performance. Different from CMA stations, ITP stations have various length of data records since most of these stations are not operational meteorological stations. Moreover, three out of 10 ITP stations are located in regions with elevation above 5,000 meters ASL, which are used to evaluate Cubist

model performance over high elevation regions. The location of the CMA stations and ITP stations is presented in Figure 3-1 (a).

Table 3-1. The summary of observational and model-based surface air temperature data used in this chapter.

Dataset	Data Type	Resolution (Spatial/temporal)	Data Source	Reference
CMA	Station data	- / Daily	NMIC	-
ITP	Station data	- / Daily	TPE	Yao et al. (2012)
CMFD	Reanalysis data	0.10° / 3-hourly	TPE	Chen et al. (2011)
CLDAS	Reanalysis data	0.0625°/ hourly	NMIC	Shi et al. (2011)
GLDAS	Reanalysis data	0.25° / 3-hourly	NASA GES DISC	Rodell et al. (2004)

3.3.2 Remote sensing data

The remote sensing data used in this study are listed in Table 3-2. The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) was downloaded from the United States Geological Survey (USGS). It is produced by combining multiple high-quality DEM datasets from various international institutions. The GMTED2010 data, with an original resolution of 7.5 arc-seconds, were resampled to 0.05°×0.05° by simple averaging to match with the resolution of other remotely sensed data.

Table 3-2. The summary of the remote sensing data used in this chapter.

Variable	Dataset(s)	Variable Category	Resolution (Spatial/temporal)	Data Source	References
Elevation	GMTED2010	Geo-location	7.5″ / Static	USGS	Danielson & Gesch (2011)

Land surface temperature (LST)	MOD11C1, MYD11C1	Clear sky only	0.05°/ Daily	NASA LP DAAC	Wan et al. (2015a, b)
Incident solar radiation (ISR)	GLASS05B01	All sky	0.05°/ Daily	UMD	Zhang et al. (2014)
Outgoing longwave radiation (OLR)	AVHOLR	All sky	0.05°/ Daily	UMD	Zhou et al. (2019)
Top-of-atmosphere albedo (TOAALB)	AVHALB	All sky	0.05°/ Daily	BNU	Song et al. (2018)
Land surface albedo (SFCALB)	MCD43C1	Clear sky only	0.05°/ Daily	NASA LP DAAC	Schaaf & Wang (2015)
Normalized Difference Vegetation Index (NDVI)	MOD13C1, MYD13C1	Clear sky only	0.05°/ 16-day	NASA LP DAAC	Didan (2015a, b)
Normalized Difference Snow Index (NDSI)	MOD10C1, MYD10C1	Clear sky only	0.05°/ Daily	NSIDC	Hall & Riggs (2015a, b)

In this study, I use MODIS daily composite LST data in a 0.05°×0.05° grid (i.e., MOD11C1 and MYD11C1), which were downloaded from NASA Land Process Distributed Active Archive Center (i.e., LP DAAC, https://lpdaac.usgs.gov/) (Wan et al., 2015a, 2015b). These products are generated by aggregating MODIS Level 2 LST products (i.e., MOD11_L2 and MYD11_L2) with strict quality control. Each product (MOD11C1 and MYD11C1) contains both daytime and nighttime LSTs from different satellite viewing time. Previous studies have proven that combining all four LST values can improve the accuracy of the estimated SAT (Noi et al., 2017; Zhang et al., 2016).

In this study, I also use three remotely sensed radiation products, including Global LAnd Surface Satellite (GLASS) incident solar radiation (ISR) at the surface, University of Maryland's (UMD) TOA outgoing longwave radiation (OLR), and Beijing Normal University's (BNU) TOA albedo (TOAALB). The GLASS ISR data are derived from multiple satellites' data, including AVHRR, MODIS and available geostationary satellites' data (Zhang et al., 2014). The OLR data are produced using AVHRR and MODIS thermal infrared data based on linear regression models derived from radiative transfer model (RTM) simulations (Zhou et al., 2019). The TOAALB data are also produced using AVHRR and MODIS data with linear models derived from RTM simulations (Song et al., 2018). All radiation products are daily data with the same spatial resolution of 0.05°×0.05° for all sky conditions.

The surface variables used in this study include MODIS surface albedo (SFCALB), Normalized Difference Vegetation Index (NDVI) and Normalized Difference Snow Index (NDSI). The MODIS surface albedo product (MCD43C1) provides daily surface albedo in a 0.05°×0.05° grid (Schaaf and Wang, 2015). Although MCD43C1 is a daily product, it estimates daily albedo using MODIS data within a 16-day moving window. Therefore, it might not reflect the real surface information of a specific day especially over regions with rapid surface dynamics. The MODIS NDVI data include MOD13C1 and MYD13C1, which are aggregated 16-day products in the same 0.05°×0.05° grid derived using Terra and Aqua MODIS data respectively (Didan, 2015a, 2015b). Both MCD43C1 and MOD13C1/MYD13C1 data were downloaded from NASA LP DAAC. Furthermore, the daily MODIS NDSI data (i.e., MOD10C1 and MYD10C1) were

acquired from National Snow and Ice Data Center (NSIDC, https://nsidc.org/) (Hall and Riggs, 2015a, 2015b) in the same 0.05°×0.05° grid.

Since LST has strong correlation with SAT, I use the all available LSTs (i.e., four instantaneous MODIS) to better capture the diurnal cycle of the surface temperature. Because the difference between LST and SAT is related with surface heat exchange, I propose to include radiation variables (i.e., ISR, OLR, and TOAALB) to reflect the crucial process that may improve the accuracy of estimated SAT. Since surface conditions can also affect the difference between LST and SAT, the surface variables are also used as candidate inputs for the model. However, the radiation variables are available for all sky conditions, while the surface variables are only available for clear sky conditions. Thus, including radiation variables would likely increase the data availability of the estimated SAT.

3.3.3 Model-based data

In this study, the SAT data of three reanalysis/meteorological forcing datasets (see Table 3-1) are also used to assess the performance of the Cubist model estimated SAT of the TP. NASA GLDAS produces reanalysis datasets regularly using multiple land surface models at different spatial resolutions (Rodell et al., 2004). I downloaded the SAT data from the GLDAS NOAH reanalysis dataset via NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC). The spatial and temporal resolution of this dataset is 0.25°×0.25° and 3-hourly respectively with the complete coverage over the global land area except the Antarctic since 2000. In addition, I also downloaded the SAT data of the CMA's CLDAS reanalysis data. The CLDAS data are produced hourly in a 0.0625°×0.0625° grid since 2008 (Shi et al., 2011; Xie et al.,

2011). Lastly, the CMFD SAT data were downloaded from the ITP's Third Pole Environment Database (TPE). The CMFD dataset contains 3-hourly SAT in a 0.1°×0.1° grid from 1979 to 2016, which is generated by dynamically adjusting the bias of GLDAS reanalysis SAT data to match CMA station observations via spline interpolation (Chen et al., 2011; Yang et al., 2010).

3.3.4 Data processing

As mentioned earlier, the GMTED2010 elevation data were aggregated to the 0.05°×0.05° grid which is the same with all other remotely sensed datasets. To train and evaluate the model, I extracted all remotely sensed data for all CMA and ITP stations aforementioned (Table 3-1) via nearest neighborhood method. These extracted remotely sensed data were then paired with the corresponding station SAT data. The station-satellite data pairs were labeled as either clear sky or cloudy sky observations based on the quality flags of the satellite data. Considering different satellite data may have different cloud masks in their quality flags, I only marked the data as clear sky observations when all satellite data' quality flags were cloud free; otherwise, the data were labeled as cloudy observations.

Since MODIS LSTs are only available under the cloud free condition, the missing values were replaced using a temporal moving window (i.e., ± five days) method. When there is at least one clear sky LST within this 11-day time period, the clear sky LST value which is temporally closest to the target date is used to replace the missing value. The purpose of this step is not to accurately predict LST under the cloudy conditions, but rather to provide a first guess of LST that can be used by the Cubist model to estimate SAT. This step could be replaced by more complex spatio-temporal gap filling

of LST data, but it is out of the scope of this study. The sensitivity of the data availability and model performance on the moving window size will be discussed later. To generate daily NDVI for each grid, 16-day MODIS NDVI data (i.e., MOD13C1 and MYD13C1) were firstly merged into one NDVI time series with corresponding date information for each grid. The merged NDVI time series were then filtered using Savitzkey-Golay method to further remove possible cloud contamination (Chen et al., 2004). Finally, the filtered time series were interpolated into daily time series based on the double sigmoid model. For all reanalysis and meteorological forcing datasets, I aggregated their sub-daily SAT values to daily mean SAT by averaging all estimations within the same day for each grid but leave them as their native spatial resolutions.

3.4 Methods

The overall design of this study is presented in Figure 3-2. First, all station-satellite data pairs were extracted and processed as described in section 2 for model training and evaluation. Only part of the CMA station-satellite data (2004-2013) were used for model training, while the rest of the CMA station-satellite data and the ITP station-satellite data were kept for independent model evaluation. The Cubist model was trained using the leave-one-station-out (LOSO) strategy to determine the parameters of the final model. I use LOSO to reduce the risk of overfitting by mimicking the process of estimating SAT for unknown regions with no station data (Meyer et al., 2016). After the model parameters were determined using LOSO, I compared two different modeling strategies to estimate daily SAT of all sky conditions (i.e., a universal all sky model vs. two separate models for clear and cloudy sky separately). The model strategy with the best accuracy evaluated using station data was chosen as the final model.

Lastly, the final model was evaluated by comparing with independent station data, the 25-fold cross validation, and cross comparing with the reanalysis/forcing data. The basis of the Cubist model and the model training/evaluation methods are further descripted in the following subsections.

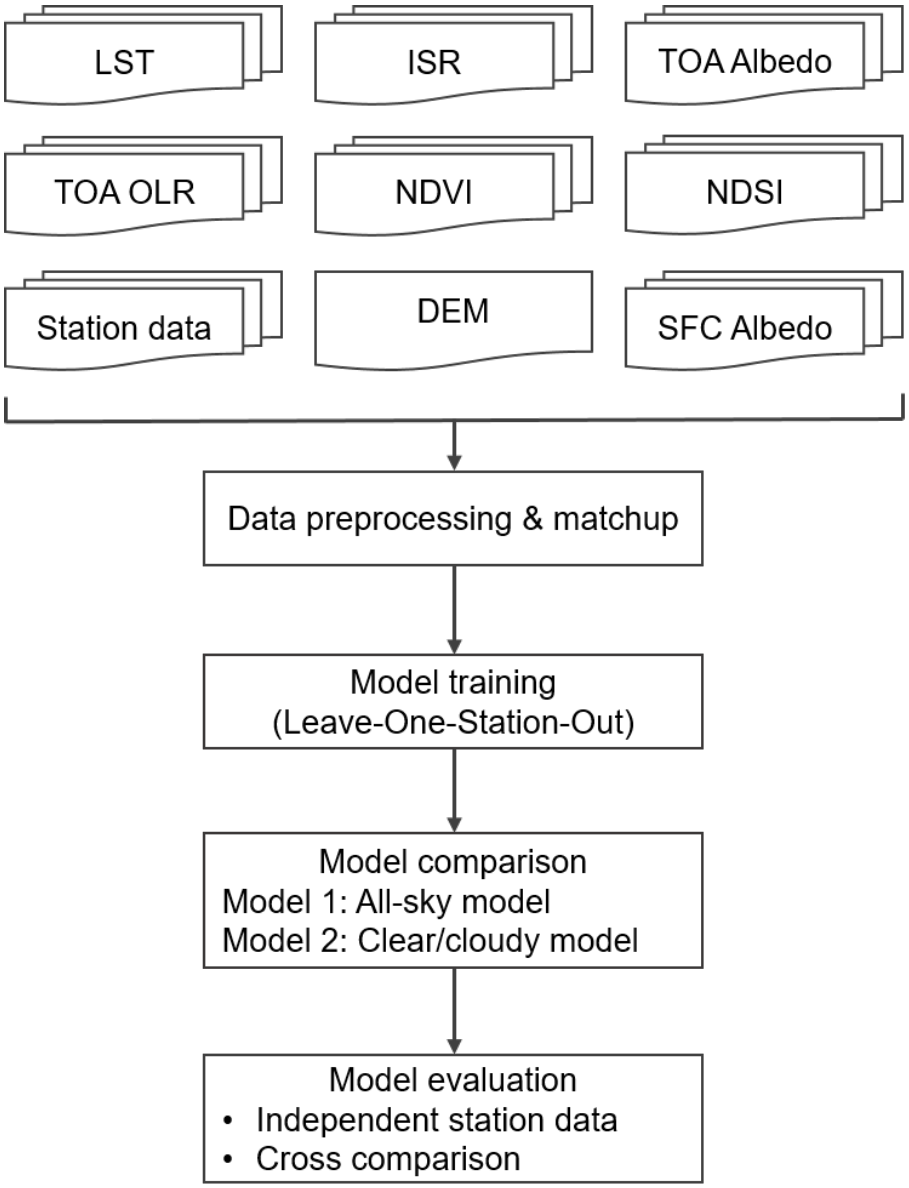

Figure 3-2. The overall flowchart of the model training and evaluation strategies of this chapter.

3.4.1 The basis of the rule-based Cubist regression

The Cubist model is a rule-based regression method developed by (Quinlan, 1993a, 1993b, 1992). The Cubist model does not give one final model like other machine learning methods, but it generates a set of rules and multi-variate predictive models associated with the rules based on the independent variables used. Once the rules and rule-associated models are determined, a specific set of independent variables will correspond to predictive models based on rules that best suits this set of independent variables. It is originally developed as a commercial software with limited documentation comparing to other popular machine learning methods. It has been adapted by researchers using open source statistical language R and become a popular model in different disciplines (Kuhn et al., 2018; Kuhn and Johnson, 2013). Despite the lack of documentation, the Cubist model is summarized herein based on existing research and documentations.

The Cubist model originates from the M5 model tree (Quinlan, 1992), which is an improved version of the simple decision tree model. M5 firstly develops a full tree by recursively partitioning all samples into different nodes, which is called tree growing process. After the tree is produced, a multiple linear regression model is fitted for each node. However, some nodes might have large model errors due to insufficient samples. Then, the M5 creates a smaller set of generalized regression models considering the training error, the numbers of sample and the goodness-of-fit for each node, which is called tree pruning process. After the pruning process, the M5 creates a set of rules and corresponding multi-variate linear regression models (Figure 3-3). Specifically, a rule is constructed using one or multiple input variables (i.e., variables listed in Table 3-2).

For instance, a rule may be set as “elevation > 3500 meters and OLR < 87 W/m2”, which creates a subset of data with the elevation higher than 3500 meters ASL and OLR lower than 87 W/m^2. Although the original tree partitioning is recursive, final rule sets after pruning may be overlapping, which means a sample may be assigned into multiple subsets based on different rules. In these cases, predictions from different subsets will be averaged to generate the merged prediction.

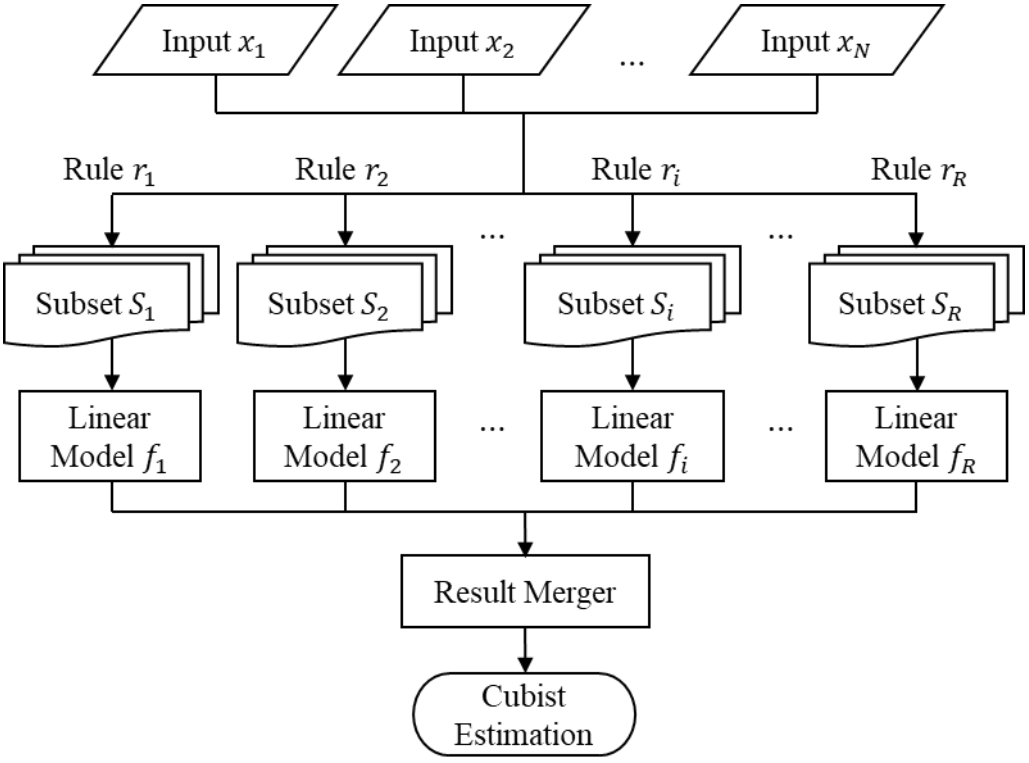

Figure 3-3. The schematic of the non-committee Cubist regression model.

Next, the Cubist model utilize a boosting-like technique to enhance its prediction performance with a process named committee prediction (Kuhn and Johnson, 2013). Since the M5 model tree does not produce perfect prediction, the Cubist model uses the model error of the initial model to adjust the original dependent variable (i.e., SAT in this study) and creates a new M5 model tree using same inputs and processes. This process repeats multiple times which is predefined to create the final model and each

individual M5 model is considered a committee (Figure 3-4). The final prediction of the Cubist model is calculated by averaging corresponding predictions of each committee.

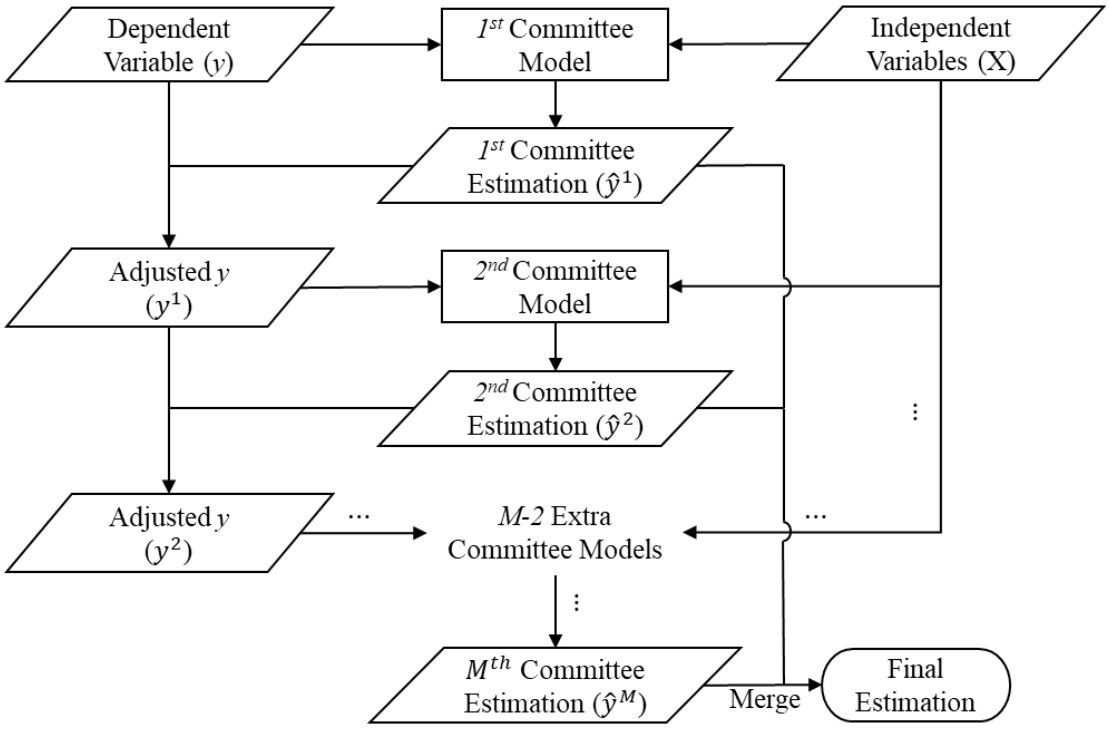

Figure 3-4. The structure of the rule-based Cubist regression model with M committees.

3.4.2 Model training and evaluation strategies

To build and evaluate the proposed model, the station-satellite data pairs were separated into two sets: 1) the training set and 2) the validation set. The training set contains CMA station-satellite data pairs from 2004 to 2013, while the validation set includes all ITP station-satellite data pairs and the ones of CMA of the year 2002, 2003, 2014, and 2015. As mentioned earlier, I compare two different strategies to estimate SAT of all sky conditions (strategy I: a universal all sky model; strategy II: two models for clear/cloudy sky conditions separately). Table 3-3 lists all candidate models for both strategies using different combinations of variables listed in Table 3-2, including,

elevation, LSTs, radiation variables, and surface variables. Since surface variables are only available for clear sky conditions, only Strategy II's clear sky models include them (i.e., NDVI, NDSI, and SFCALB) as model inputs.

Table 3-3. The summary of variables used in different Cubist models in this chapter.

Model Type	#	Geolocation & Elevation	Day of Year	LST	Radiation Variables	Surface Variables
Clear Sky (CLR)	0	Yes	Yes	Yes	-	-
	1	Yes	Yes	-	Yes	-
	2	Yes	Yes	Yes	-	Yes
	3	Yes	Yes	-	Yes	Yes
	4	Yes	Yes	Yes	Yes	-
	5	Yes	Yes	Yes	Yes	Yes
Cloudy Sky (CLD)	0	Yes	Yes	Gap-filled	-	-
	1	Yes	Yes	-	Yes	-
	4	Yes	Yes	Gap-filled	Yes	-
All Sky (ALL)	0	Yes	Yes	Gap-filled	-	-
	1	Yes	Yes	-	Yes	-
	4	Yes	Yes	Gap-filled	Yes	-

In the Cubist model, two parameters need to be determined through training, i.e., the number of committees and the number of neighbors. During the training process, I used the LOSO strategy to select model parameters to avoid the overfitting issue as mentioned earlier (Meyer et al., 2016). Firstly, the training data were grouped by stations. For each iteration, a series of Cubist models were fitted using different combinations of model parameters using data of all stations except one which was randomly chosen. Then, the models were evaluated using the data of the left-out station. After each station has been used as the left-out station to evaluate different model

parameters, the final model parameters were selected based on the model performance across all iterations.

To evaluate the final Cubist model, I first used the validation dataset of CMA stations of the year 2002, 2003, 2014 and 2015 to assess the model performance when it is applied to data of different years. Additionally, the model was also evaluated using independent data of 10 ITP stations. Furthermore, I carried out a 25-fold cross validation experiment to examine the robustness of our model.

Lastly, I applied the final Cubist model to the entire TP for the year of 2014. The estimated SAT of the TP was cross compared with three reanalysis/forcing datasets listed in Table 3-1. The main purpose of the cross comparison is to evaluate the spatial and temporal (i.e., seasonal) pattern of the estimated SAT. Additionally, I also compared the accuracy of our Cubist estimation and the existing datasets using CMA station data as the reference since all datasets have their own uncertainty.

3.5 Model training and evaluation results

3.5.1 The comparison of different modeling strategies

The statistics of all candidate models listed in Table 3-3 are presented in Table 3-4. For clear sky models, the full model (i.e., CLR-5) has the best performance with the lowest RMSE and the highest R^2. The clear sky model without surface variables (i.e., CLR-4) also achieves comparable performance with the full model (CLR-5). However, when LSTs are replaced by TOA radiations (i.e., CLR-0 vs. CLR-1, CLR-2 vs. CLR-3), the performance of the models without LSTs are worse than the models with LSTs. The clear sky models indicate that LSTs have strong impacts on the Cubist model performance while radiation variables can be good supplemental variables to improve

the model performance. Figure 3-5 demonstrates the density scatter plots of all Cubist models for the estimated daily mean SATs against the CMA station measurements.

Table 3-4. The comparison of the training statistics for all Cubist models listed in Table 3-3.

Model Type	Model Number	Bias (℃)	RMSE (℃)	R^2	Data Count
Clear Sky Model (CLR)	0	-0.134	1.373	0.978	102,457
	1	-0.027	1.937	0.955	102,457
	2	-0.133	1.344	0.979	102,457
	3	-0.039	1.864	0.958	102,457
	4	-0.111	1.291	0.980	102,457
	5	-0.108	1.265	0.981	102,457
Cloudy Sky Model (CLD)	0	-0.091	1.618	0.969	268,938
	1	-0.018	2.058	0.949	268,938
	4	-0.096	1.484	0.974	268,938
All Sky Model (ALL)	0	-0.104	1.549	0.972	371,395
	1	-0.021	2.048	0.951	371,395
	4	-0.106	1.434	0.976	371,395

For cloudy sky models, the model with the temporally gap-filled LSTs and radiation variables (i.e., CLD-4) has the best performance with lowest RMSE. However, the model performance deteriorates when either gap-filled LSTs or radiation variables are dropped out from the model (i.e., CLD-0, CLD-1). Nonetheless, the model with only gap-filled LSTs (CLD-0) still outperforms the model with only radiation variables (CLD-1), which is similar with clear sky models. Moreover, the best cloudy sky model (CLD-4) performs slightly worse than the corresponding clear sky model (CLR-4) which may be caused by the uncertainty of gap-filled LSTs.

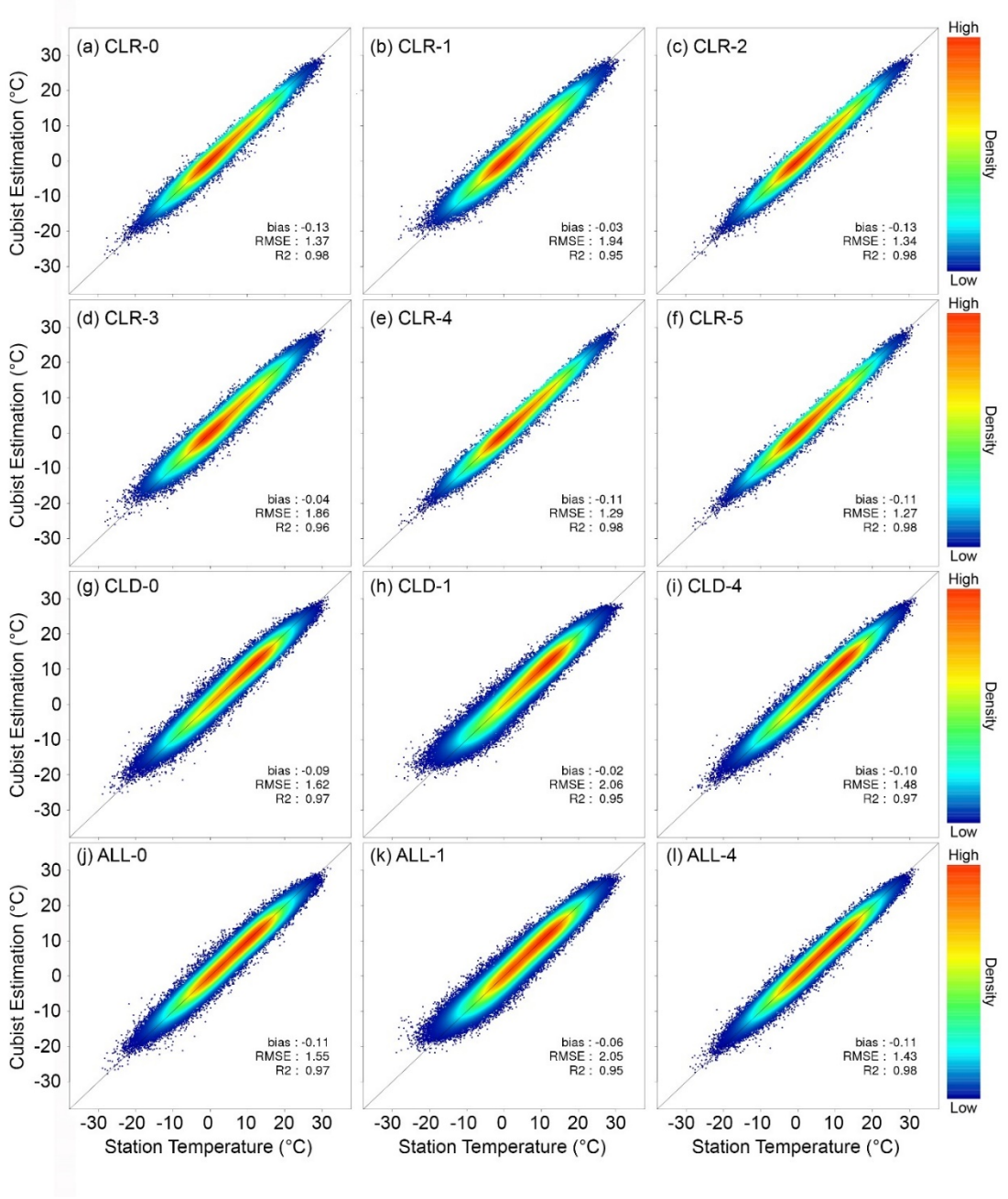

Figure 3-3-5. The density scatter plots of all 12 Cubist models listed in Table 3-3 for training results: (a) CLR-0, (b) CLR-1, (c) CLR-2, (d) CLR-3, (e) CLR-4, (f) CLR-5, (g) CLD-0, (h) CLD-1, (i) CLD4, (j) ALL-0, (k) ALL-1, (l)ALL-4.

The best all sky model is the one with both gap-filled LSTs and radiation variables (ALL-4), followed by the model with only gap-filled LSTs (ALL-0) and the model with only radiation variables (ALL-1). Moreover, the best all sky model (ALL-4) has

better overall performance than the best cloudy-sky model (CLD-4) but underperforms the best clear sky model (CLR-5). However, the al sky model, ALL-4, can estimate daily SAT for much more observations instead of the clear sky only model (CLR-5) (i.e., number of data points: 371,395 vs. 102,457) with comparable overall accuracy. In practice, this advantage can largely increase the data availability without notably sacrificing the data quality.

Table 3-5 summarizes the validation results for all 12 Cubist models using temporally independent CMA station data (of the year 2002, 2003, 2014, and 2015). For all sky models, I further calculated the statistics for clear sky and cloudy sky observations separately to directly compare with the results of clear/cloudy sky models. For clear sky condition, all models with LSTs as inputs show comparable performance with each other but outperforms the models without LSTs. This further confirms that LSTs have major contribution to accurately estimate daily mean SAT. The best model is still the full model with all variables as inputs (CLR-5). For cloudy and all sky models, the models with both gap-filled LSTs and radiation variables are the best model of its category (i.e., CLD-4 and ALL-4). Figure 3-6 shows the density scatter plots of the best models within each category (i.e., CLR-5, CLD-4, and ALL-4). For the model ALL-4, the density scatter plots of clear sky and cloudy sky observations are also presented separately to directly compare with the results of CLD-4 and CLR-5 (Figure 3-6 (c-d)).

Table 3-5. The comparison of the statistics for the validation results for different cubist models listed in Table 3-3. In this table, the validation for all sky models is further separated for clear sky and cloudy sky data.

Model Type	Model Number	Data Type	Bias (℃)	RMSE (℃)	R^2	Data Count

Clear Sky Model (CLR)	0	-	-0.141	1.638	0.967	40,528
	1	-	-0.085	2.373	0.931	40,528
	2	-	-0.144	1.637	0.967	40,528
	3	-	-0.096	2.372	0.932	40,528
	4	-	-0.116	1.643	0.967	40,528
	5	-	-0.113	1.631	0.967	40,528
Cloudy Sky Model (CLD)	0	-	-0.027	1.983	0.951	109,713
	1	-	0.021	2.489	0.924	109,713
	4	-	-0.025	1.917	0.955	109,713
All Sky Model (ALL)	0	All	-0.067	1.884	0.957	150,241
		Clear	0.028	1.647	0.967	40,528
		Cloudy	-0.106	1.986	0.952	109,713
	1	All	-0.024	2.460	0.927	150,241
		Clear	0.097	2.362	0.932	40,528
		Cloudy	-0.080	2.496	0.924	109,713
	4	All	-0.059	1.837	0.959	150,241
		Clear	-0.069	1.634	0.967	40,528
		Cloudy	-0.058	1.922	0.954	109,713

For all sky models, when the samples are separated into clear/cloudy conditions, the estimation of all sky models can achieve similar or even better accuracy with the estimation of the corresponding clear/cloudy sky models. For example, the statistics of ALL-4 for clear sky and cloudy sky observations are similar with the statistics of the corresponding clear sky model (CLR-4; RMSE: 1.634 °C vs. 1.643 °C; Bias: -0.069 °C vs. -0.116 °C; R^2: 0.967 vs. 0.967) and cloudy sky model (CLD-4; RMSE: 1.922 °C vs. 1.917 °C; Bias: -0.058 °C vs. -0.025 °C; R^2: 0.954 vs. 0.955). In general, the all sky model with gap-filled LSTs and radiation variables (i.e., ALL-4) shows the best overall performance with satisfactory accuracy and the capability of overcoming cloud

contamination issue. Therefore, I only evaluate the ALL-4 model in the remaining part of this study.

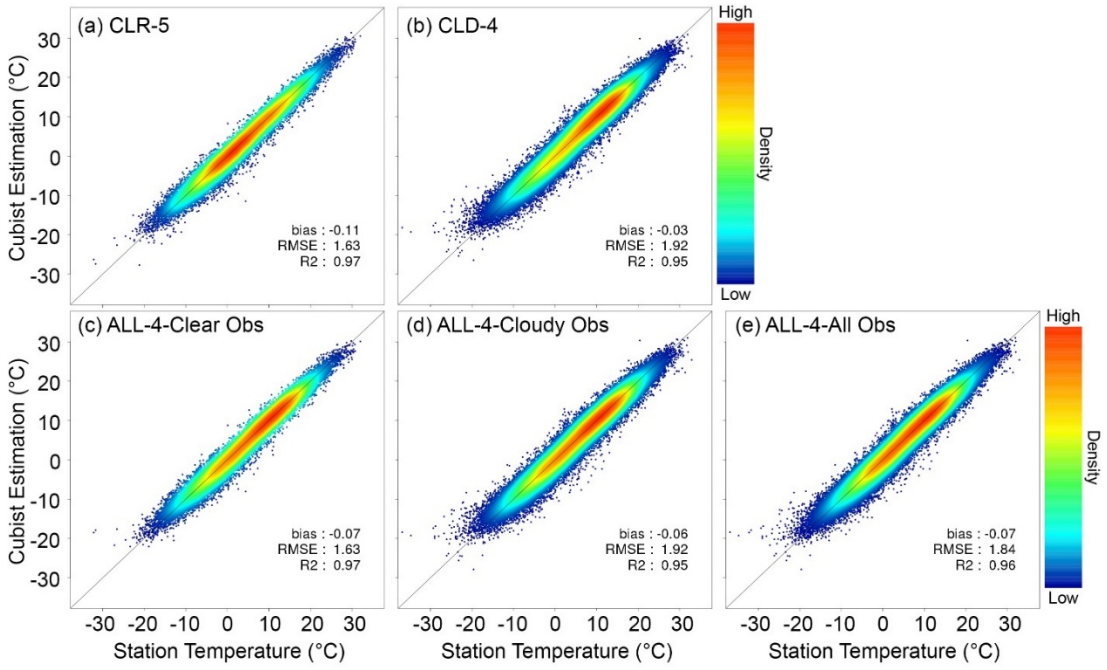

Figure 3-6. The density scatter plots of the validation results for the best model in each category (a) CLR-5, (b) CLD-4, (c) clear sky observations of ALL-4, (d) cloudy sky observations of ALL-4, and (e) all sky observations of ALL-4.

3.5.2 The Independent evaluation with ITP station measurements

To further validate the all sky model independently, the data of 10 ITP stations of varying time periods were used in this study. Out of these 10 stations, three of them are located at elevation higher than 5,000 ASL. Figure 3-7 presents the validation results using these ITP station data. The estimated daily mean SAT show good agreement with the station measurements with nearly zero bias. However, the RMSE is slightly larger than the ones of model training and validation results using the CMA station data (RMSE: 2.18°C vs. 1.84°C). Furthermore, the accuracy of estimated SAT for stations with elevation higher than 5,000 meters ASL is slightly worse than other ITP stations (RMSE: 2.29°C vs. 2.05°C). This result is possible considering that the training data

do not contain any stations above 5,000 meters ASL. This lack of representation could increase the uncertainty of the estimated SATs over high elevation regions (Zhang et al., 2016).

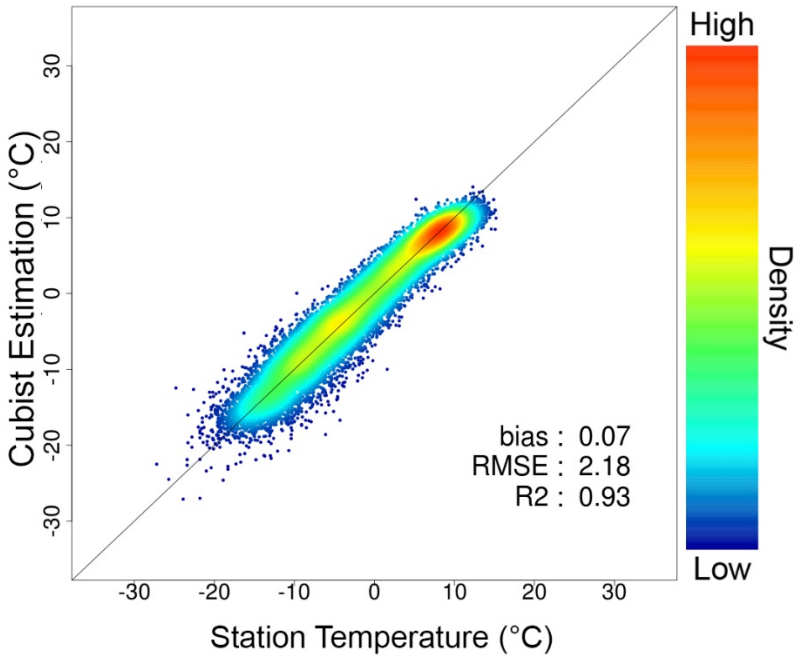

Figure 3-7. The density scatter plot of the independent validation for the final Cubist model using data of 10 Institute of Tibetan Plateau Research (ITP) stations.

3.5.3 The cross comparison with model-based data

In addition to evaluate the proposed model with station measurements, I also compared the Cubist estimation of the TP with three reanalysis/meteorological forcing datasets listed in Table 3-1. Figure 3-8 compares the spatial pattern of the monthly mean SAT of our Cubist model estimation (Figure 3-8 (a-d)) with GLDAS (Figure 3-8 (e-h)), CLDAS (Figure 3-8 (i-l)), and CMFD (Figure 3-8 (m-p)) for January, April, July, and October 2014. Overall, all four datasets show very similar spatial and temporal SAT gradients across the entire TP. Generally, the SAT is higher at the regions with low altitudes (i.e., the northern and southeast parts of the TP) while the high elevation

regions (e.g., the western and central areas of the TP) have lower temperature. Additionally, all datasets show the same seasonal SAT dynamics.

Despite the consistency, there are still notable differences among these datasets. Even though GLDAS may capture the overall spatial pattern of the SAT, it does not have the same level of spatial details as the Cubist estimation because GLDAS's resolution is very coarse (0.25° vs. 0.05°). The lack of the spatial details can be troublesome because parts of the TP have very complex terrain. It is not suitable to use such coarse resolution to represent the climate and ecosystem processes of those regions. Additionally, the CLDAS SAT appears to have larger spatial gradients, especially for April and July 2014.

Considering all these datasets have their own uncertainties, I use validation years' CMA station measurements as a reference to compare the accuracy of these four datasets. Both CMFD and GLDAS show substantial underestimation (Bias: -2.47°C vs. -3.11°C) when compared to the reference CMA station data. The CLDAS data notably overestimates the surface temperature with a bias of 1.07°C while the Cubist model estimation shows nearly zero bias (-0.07°C). Furthermore, the Cubist model estimation shows smaller uncertainty than other three datasets (RMSE: 1.84°C (Cubist) vs. 4.82°C (GLDAS) vs. 4.20 (CMFD) vs. 3.31°C (CLDAS)). In summary, the Cubist model estimated SAT of the TP has better accuracy than existing reanalysis and forcing datasets and can capture the SAT's spatial and temporal dynamics of the TP.

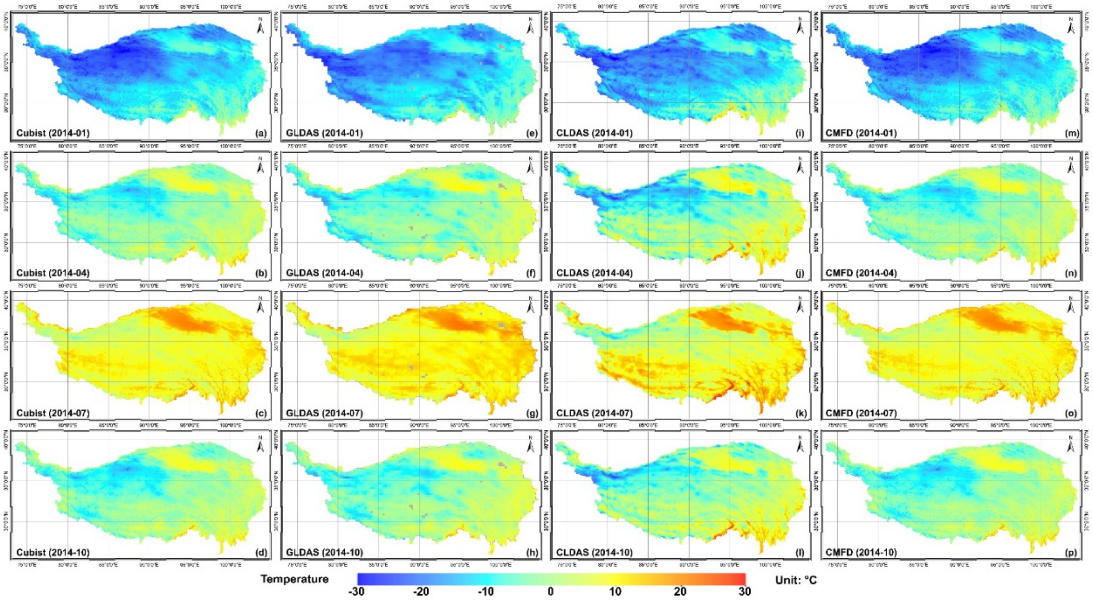

Figure 3-8. The spatial and temporal patterns of (a-d) the Cubist model estimated surface air temperature (SAT), (e-h) the GLDAS SAT, (i-l) the CLDAS SAT, and (m-p) the CMFD SAT data of January, April, July and October 2014.

3.6 Warming analysis of the Tibetan Plateau

I applied the final all sky Cubist regression model to the entire Tibetan Plateau for 2000-2015. The estimated SAT is then used to estimate the surface warming rate of the TP. Figure 3-9 demonstrates the spatial pattern of the estimated surface warming rate over the TP for annual mean SAT, cold season (from December till May) mean SATA and warm season (from June till November) mean SAT. The estimation of both annual and seasonal mean SAT warming rate shows rapid warming of the TP during the past 16 years with fastest annual warming occurs over the western part of the Plateau. If compared with the elevation map of the Plateau, this fast warming region is located at high elevation mountainous areas, which is also missed by the CMA station measurements network. In other words, the warming studies depending only on the station measurements would likely underestimate the warming rate of the Plateau. This pattern persists through all seasonal mean SAT warming analysis. Moreover, the

underestimation of station measurements of the cold season mean can be more severe due to the warming rate of the western part of the Plateau is much higher than the annual mean or warm season mean SAT.

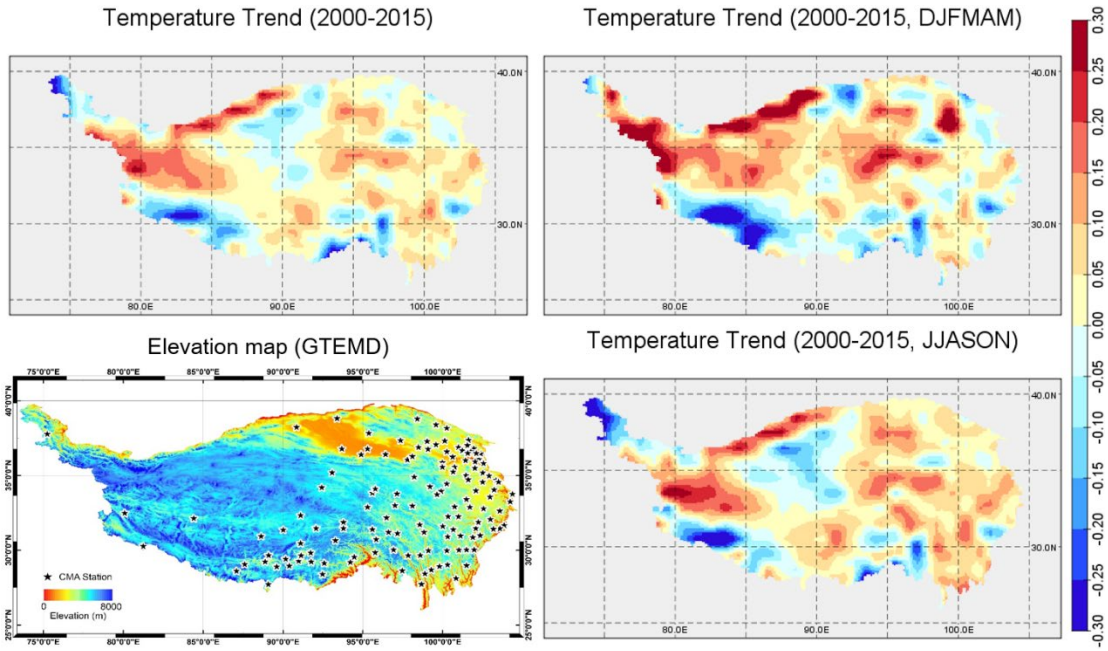

Figure 3-9. The surface warming rate of the annual mean temperature, cold season mean temperature (DJFMAM), and warm season mean temperature (JJASON) and its reference to the elevation and station distribution of the Tibetan Plateau.

To further demonstrate the advantage of the Cubist-based SAT estimation, I calculated the regional annual mean SAT anomalies across the entire Tibetan Plateau between 2000-2015. Figure 3-10 shows the comparison of the time series of the annual mean SAT anomalies of TP (against climatology period 2001-2010) using the CMA station measurements and the Cubist-based regional dataset generated by this research. The Cubist-based time series show a faster warming pattern than the station-based time series. If the warming rate is quantified using the estimated linear trend, the Cubist-based data suggest that the TP has been warming at a faster rate for 2000-2015 than previously estimated (i.e., 0.512 °C/decade v.s. 0.465 °C/decade). This difference may

be caused by the undersampling of the western part of the plateau by CMA station network, especially over the high elevation regions.

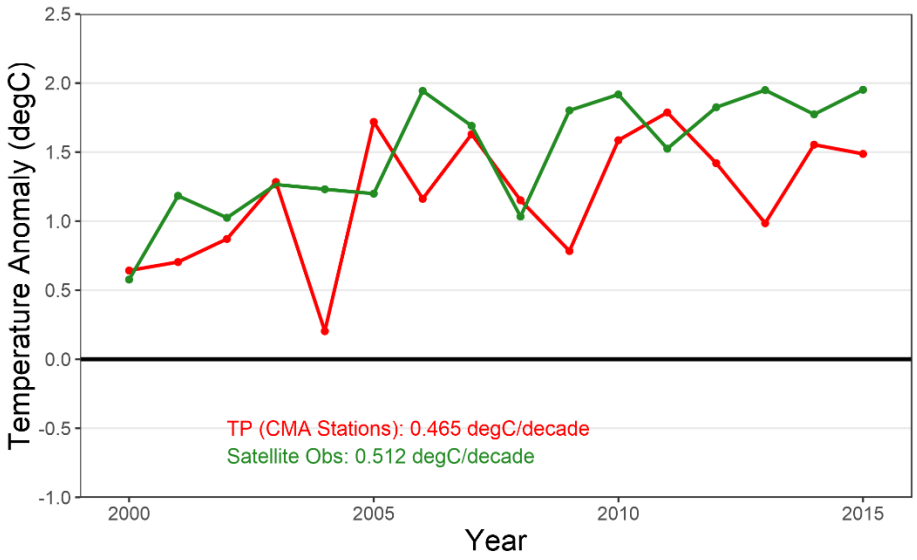

Figure 3-10. The time series of estimated regional annual mean surface air temperature anomalies over the entire Tibetan Plateau using Chinese Meteorological Stations measurements (red line) and the Cubist-based dataset (green line).

3.7 Conclusion

This study demonstrates that combining LSTs and radiation variables at both the surface and TOA levels can produce daily all sky SAT data with high accuracy. With a reasonably defined temporal moving window to fill the gap of missing LST caused by cloud contamination, the all sky Cubist model can largely increase the data availability of estimated daily mean SAT over the Tibetan Plateau. The model has been validated using spatially and temporally independent station data and cross validation with nearly zero bias and reasonable RMSEs (1.8-2.2 °C). When cross compared with the existing reanalysis/forcing datasets, the Cubist model estimated SAT can represent the spatial and temporal dynamics of the surface temperature of the TP and retain important spatial details. When all datasets were benchmarked against the CMA station

data, the Cubist model show better performance with no notable bias and much smaller RMSEs. However, the 25-fold cross-validation practice suggests that the representativeness of the training dataset is of great importance to produce a high-quality machine learning model with no built-in bias.

With the all sky Cubist model, I generated a 0.05°×0.05° daily average surface air temperature dataset for the entire Tibetan Plateau for 2002-2016. The resulting dataset is of great value to study recent climate warming and corresponding impacts over the entire Tibetan Plateau. However, as mentioned earlier, the training data are only from the finite CMA stations. Therefore, users should be aware of the potential larger uncertainty for the regions of which the weather/climate patterns might not be represented by the CMA station data, such as, the regions with very high elevations (e.g., above 6,000 meters ASL) or with complex topography. Although the model is developed for the Tibetan Plateau, the framework of this model could be extended to other regions since the underlining mechanism should be similar. However, when the framework is applied to other areas, the model should be properly retrained using the data of the target area to ensure that the model are built correctly with representative training data.

Despite the improved accuracy and data availability of the daily SAT dataset, there are still uncertainties require further investigations to improve the resulted data. First, all input satellite data have different level of uncertainties which can be propagated into this empirical-based estimation. Therefore, it will be beneficial to understand the sensitivity of the estimated SAT regarding to the uncertainty of each individual input variables. Secondly, due to the complex terrain of the TP, there are grids with large

elevation gradients. It can be very challenging to assess the accuracy of these regions. Hence, it is necessary to use more independent data of the regions with very high elevations and complex topography, if available, to comprehensively evaluate the quality of the estimated SAT data. Moreover, there are growing demands of grid-level uncertainty assessment to improve the confidence of local and regional applications using various climate datasets. Thus, it is of the best interest to provide the uncertainty value associated with each grid for the estimated SAT data using advanced statistical methods, such as, Markov Chain Monte Carlo (MCMC), bootstrapping etc. Lastly, I are planning to extend this model using AVHRR data to generate long-term SAT climate data records of the TP (since 1982) to enable climate applications for the last four decades.

Chapter 4 Estimating the Surface Air Temperature of the Northern High Latitudes Using Machine Learning Model

4.1 Summary

The northern high latitude consists of a complex physical system with atmosphere, cryosphere, ocean, and biosphere. It has experienced dramatic climate and environmental changes in the past decades. Overall, the temperature of the northern high latitudes has warmed faster than the rest of the world based on historical observations and model simulations. However, the lack of high quality surface air temperature dataset has led to low confidence of the climate analysis at local and regional scales over the northern high latitudes. Recently, research has used remote sensing land surface temperature and ice surface temperature products to assess the temperature change over the northern high latitudes and its associated changes. However, the different physical meaning of remote sensing land/ice surface temperature has limited its application for climate analysis. To address this issue, I adapted the machine learning framework developed in the previous chapter for the Tibetan Plateau by considering the unique features of the northern high latitudes. Because of the large data volume of this study area, the rule-based Cubist regression model in the original framework has been replaced by the multivariate adaptive regression splines model. The adapted machine learning model is trained and evaluated using the Global Historical Climatology Network daily station data archive. Using data of 642 stations, the model estimated surface air temperature has good performance with the bias of -0.1 ~ -0.2 °C and the RMSE of 2.1 ~ 2.6 °C. However, due to lack of stations over the permanent ice surface, the model performs worse over the permanent

ice surface than other surface conditions (RMSE: 2.8 °C vs. 2.4 °C). When the model is applied to the entire northern high latitudes, the estimated SAT show reasonable spatial and seasonal dynamics when compared with existing datasets but it can provide much detailed spatial information for climate analysis.

4.2 Background

The northern high latitudes (NHL), defined as the geographical regions poleward of 60°N, consists of a complex physical system with atmosphere, cryosphere, ocean, and biosphere (Groisman and Soja, 2007; Hwang et al., 2018). The NHL is an integral part of Earth system and plays a fundamental role on regulating climate system, hydrological cycle, ecosystem dynamics, and societal activities at local, regional and even global scales (Hwang et al., 2018). The terrestrial snow and ice host within the NHL account for a majority of the fresh water resource of the world (J. Huang et al., 2017). The NHL is also home to a diverse set of the Earth's biome, such as, boreal forest, tundra, steppe, taiga, and desert. The permafrost under the ground of the NHL stores a substantial amount of organic carbon, which is an important major carbon sink. Through large scale atmospheric and oceanic circulations, the NHL strongly affects the weather and climate pattern of the Northern Hemisphere. Studies have demonstrated the influence of the Arctic on the extreme weather events at mid-latitude (Cohen et al., 2014; Overland et al., 2016).

Due to anthropogenic climate change, the NHL has experienced remarkable environmental change across all components (Bhatt et al., 2017; Hwang et al., 2018; Loranty et al., 2018). Both observational information and model simulations have reported rapid temperature warming over the NHL during the last few decades (J.

Huang et al., 2017). Moreover, the warming over the NHL has been reported to be much faster than other regions. More specifically, the analysis based on station temperature measurements has reported that the SAT over the NHL warmed 2-3 times faster than the global average since 1970s, while similar results are also found by other climate model simulation based studies (J. Huang et al., 2017; Pithan and Mauritsen, 2014). Moreover, climate model simulations further suggest that this warming amplification over the NHL will likely continue through the future at a rate of 1.5-4.5 times faster than other parts of the world (Pithan and Mauritsen, 2014). This warming amplification has already caused remarkable environmental change within the NHL and other regions. For example, the warming SAT and prolonged growing season for vegetation has led to the change of vegetation phenology within the NHL, thus affecting the carbon cycle locally (Bhatt et al., 2017; Pearson et al., 2013; Swann et al., 2010). Additionally, the increasing temperature also change the freeze/thaw cycle of the active soil layers of the NHL (Loranty et al., 2018; Watts et al., 2014; Zhang et al., 2013). Observations also reported the advance of snow and ice melting during the summer and the delay of snow and ice accumulation during the winter, which results to reduced amount of snow and ice and shortened snow/ice accumulation season (Chen et al., 2018). These changes notably impact the hydrological regime and associated processes, such as, methane emission, seasonal flooding, and animal migration (Watts et al., 2014; Woods and Caballero, 2016; Zhang et al., 2013). Recently, studies have also attributed the increasing occurrence of the extreme weather events at the mid-latitude to the warming NHL, such as, extreme cold winters and extreme precipitation events (Cohen et al., 2014; Overland et al., 2016). However, the analysis of these

impact studies are not noncontroversial due to the lack of high quality climate observations.

Despite the high confidence of the overall warming of the NHL, the confidence of local and regional climate analysis remains low for the NHL due to the lack of observational data and the lack of comprehensive understanding of the complex physical processes of the NHL (Döscher et al., 2014; Jung et al., 2016; Laudon et al., 2017). Observation-based warming studies predominantly use station measured temperature data to assess the warming rate of the NHL. However, the weather stations with thermometers are very sparse over the NHL (Figure 4-1). In the Global Historical Climatology Network daily data archive (GHCN-D), there are total 642 weather stations within the NHL with the measurements of daily mean SAT. Of these stations, nearly 30% of them are clustered within Alaska and 15%-20% of stations are within the northern Europe. The rest of the stations are sparsely distributed across the vast majority of the land mass over the NHL, leaving Greenland and the eastern part of the Eurasia undersampled. The inequality of the station measurements may lead to large uncertainty for the local and regional climate analysis over the undersampled regions as pointed out in the second chapter. Studies have also attempted to conduce climate analysis using global and regional reanalysis datasets, such as, ERA-5, MERRA-2, GLDAS, and JRA-55 etc. These datasets suffer large uncertainty and notable biases for the polar regions.

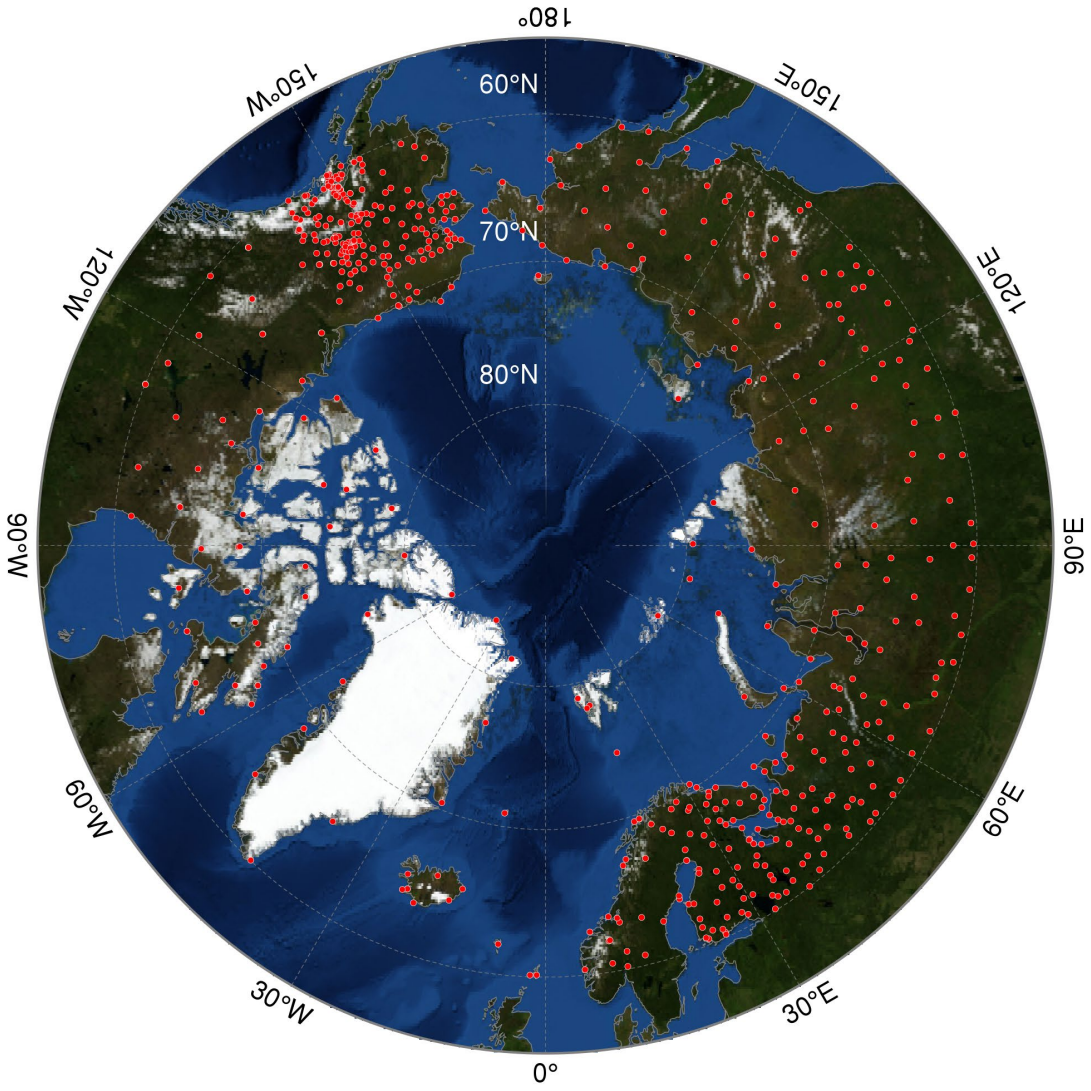

Figure 4-1. The distribution of Global Historical Climatology Network - Daily (GHCN-D) for the northern high latitude. The red dots are the weather stations of the GHCN-D with daily mean temperature measurements over the northern high latitude. The background image is the blue marble image of 2014/08 created using NASA EOS MODIS data.

As an alternative, researchers also explored satellite remote sensing products to study the warming of the northern high latitude because of the global coverage of the remote sensing data. Among remote sensing data, the LST and IST data have been mostly popular because of their strong correlation with the SAT. For example, Hall et al. (2013) examined the warming of Greenland using MODIS IST products for 2000-2012

and found dramatic warming of the Greenland and faster melting. Westergaard-Nielsen et al. (2018) used MODIS LST products to analyze the temperature change of ice-free part of Greenland and found varying seasonal and latitudinal pattern of the warming for 2001-2015. Despite the strong correlation between remotely sensed LST/IST and SAT, they are physically different variables as pointed out in earlier chapters. Good (2016) found that the satellite LST data are usually several degrees Celsius higher than station measured SAT over the NHL during summer months based on data of Atmospheric Radiation Measurement stations (ARM). But Nielsen-Englyst et al. (2019) used multi-year temperature measurements of 29 in situ sites over the northern Alaska, Greenland, and Arctic Ocean and documented that the satellite IST data are nearly 2°C cooler than the SAT measurements, while the largest differences occur when the temperature is below 0°C or when the ice is meltin. The difference between satellite LST/IST and SAT would likely cause biases while the LST or IST produces were used as the proxy for the SAT.

Given the importance of the NHL and the observed dramatic change over the NHL, one would expect there were substantial efforts on providing high quality climate data to improve our understanding of the critical change over the NHL. However, past and current research efforts mostly focus on improving the capability of modeling and forecasting the change of the NHL, while only limited progress has been made to provide high quality and spatial complete climate data based on remotely sensed data. With the increasing attention of using satellite LST to approximate SAT, different methods have been developed to estimate SAT from remotely sensed LST including linear regression models, machine learning models and spatio-temporal interpolation

models as described in Chapter 3. Using machine learning models to map SAT using remotely sensed LST has become more popular in recent years because of the ability to account for the nonlinearity nature of the difference between SAT and LST. However, as outline in Chapter 3, these methods fail to address the issue of cloud contamination on remotely sensed LST products which may lead to reduced data quality and availability. Therefore, I developed a machine learning based modeling framework to estimate SAT using remotely sensed LST and satellite radiation products for all sky conditions in Chapter 3. Although the model is developed using the data of the Tibetan Plateau, I would argue that the generic modeling framework can be adapted to the NHL to generate high quality all sky daily SAT data of the NHL using satellite products.

The purpose of this chapter is to adapt the machine learning based all sky model described in Chapter 3 to generate a daily all sky SAT dataset over the NHL at the grid size of 0.05°×0.05°. However, the model needs to be retrained using appropriate station-satellite data pairs over the NHL in order to ensure the accuracy of the final estimation. Additionally, there is another situation need to be taken into consideration while adapting the model developed in Chapter 3 for the NHL. Parts of the NHL periodically experience polar day and polar night when the diurnal cycle of the surface temperature is different from a typical temperature diurnal cycle. Meanwhile, the satellite products of the TOA albedo and incident solar radiation are unavailable for polar night situation since there is no incoming solar radiation. Therefore, the all sky model for the NHL ideally need to handle this situation. In this chapter, the section 4.3 summarizes the station measurements, remote sensing data, and model-based data. The

section 4.4 describes the multivariate adaptive regression spline model and the strategy of model training and evaluation used in this chapter. The section 4.5 presents the results of model training and evaluation using both station measurements and existing SAT datasets. The section 4.6 shows the analysis of the surface warming rate over the NHL using the SAT data estimated from the MARS model with satellite products as inputs.

4.3 Data

Similar with the Chapter 3, this chapter also includes three different categories of data for model training, evaluation and analysis. The three categories include station temperature measurements, remote sensing products, and existing SAT datasets. Since most of the data used in this chapter has already been previously introduced in other chapters, I will focus on the data that have not been described before

In this chapter, the station temperature measurements are used for both model training and evaluation. The main source of the daily average SAT used in this study is 642 meteorological stations within the archive of GHCN-D over the NHL. The station measurements from 2002 to 2017 were downloaded from the NOAA NCEI. The spatial distribution of these GHCN-D stations over the NHL is presented in Figure 4-1. Although there are more than 642 stations over the NHL within the GHCN-D data archive, these are the only stations providing the daily mean SAT measurements after 2000.

The remote sensing data used in this study are listed in Table 4-1. The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) was downloaded from the United States Geological Survey (USGS). It is produced by combining multiple high-

quality DEM datasets from various international institutions. The GMTED2010 data, with an original resolution of 7.5 arc-seconds, were resampled to 0.05°×0.05° by simple averaging to match with the resolution of other remotely sensed data.

Table 4-1. The summary of the remote sensing data used in this chapter.

Variable	Dataset(s)	Variable Category	Resolution (Spatial/temporal)	Data Source	References
Elevation	GMTED2010	Geo-location	7.5″ / Static	USGS	Danielson & Gesch (2011)
Land surface temperature (LST)	MOD11C1, MYD11C1	Clear sky only	0.05°/ Daily	NASA LP DAAC	Wan et al. (2015a, b)
Incident solar radiation (ISR)	GLASS05B01	All sky	0.05°/ Daily	UMD	Zhang et al. (2014)
Outgoing longwave radiation (OLR)	AVHOLR	All sky	0.05°/ Daily	UMD	Zhou et al. (Submitted)
Top-of-atmosphere albedo (TOAALB)	AVHALB	All sky	0.05°/ Daily	BNU	Song et al. (2018)

In this study, I use MODIS daily composite LST data in a 0.05°×0.05° grid (i.e., MOD11C1 and MYD11C1), which were downloaded from NASA Land Process Distributed Active Archive Center (i.e., LP DAAC, https://lpdaac.usgs.gov/) (Wan et al., 2015a, b). These products are generated by aggregating MODIS Level 2 LST products (i.e., MOD11_L2 and MYD11_L2) with strict quality control. Each product (MOD11C1 and MYD11C1) contains both daytime and nighttime LSTs from different satellite viewing time. It should be noted that there are no nighttime/daytime LST values for the regions experiencing polar day/night situation.

In this study, I also use three remotely sensed radiation products, including Global LAnd Surface Satellite (GLASS) incident solar radiation (ISR) at the surface, University of Maryland's (UMD) TOA outgoing longwave radiation (OLR), and Beijing Normal University's (BNU) TOA albedo (TOAALB). The GLASS ISR data are derived from multiple satellites' data, including AVHRR, MODIS and available geostationary satellites' data (Zhang et al., 2014). The OLR data are produced using AVHRR and MODIS thermal infrared data based on linear regression models derived from radiative transfer model (RTM) simulations (Zhou et al., 2019). The TOAALB data are also produced using AVHRR and MODIS data with linear models derived from RTM simulations (Song et al., 2018). All radiation products provide daily data with the same spatial resolution of 0.05°×0.05° for all sky conditions except for regions experiencing polar day/night condition.

Since LST has strong correlation with SAT, I use the all available LSTs (i.e., four instantaneous MODIS) to better capture the diurnal cycle of the surface temperature. Because the difference between LST and SAT is related with surface heat exchange, I propose to include radiation variables (i.e., ISR, OLR, and TOAALB) to reflect the crucial process that may improve the accuracy of estimated SAT. The experiments in Chapter 3 have demonstrated that the all sky model without surface variables can achieve satisfactory accuracy, thus I did not include the surface variables in this chapter. Additionally, I calculated the day length for each 0.05°×0.05° grid based on the location and the day of year as an indicator for the polar day/night.

To evaluate the spatial and temporal pattern of the estimated SAT, two global gridded SAT datasets with relative high resolution are used in this chapter. The SAT of the

CRU-Ts.4.02 is derived by spatially interpolating the measurements of around 6,000 weather stations worldwide via the angular distance weighting (ADW) method (Harris et al., 2014). It provides the estimation of monthly SAT with the grid size of 0.5°×0.5° for the global land area. The GLDAS provide the 3-hourly SAT estimation at the grid size of 0.25°×0.25° for the global land area using the Noah land surface model (Rodell et al., 2004).

Table 4-2. The summary of observational and model-based surface air temperature data used in this chapter.

Dataset	Data Type	Resolution (Spatial/temporal)	Data Source	Reference
CRU-Ts.4.02	Station interpolation data	0.5° / monthly	UAE CRU	(Harris et al., 2014)
GLDAS	Reanalysis data	0.25° / 3-hourly	NASA GES DISC	Rodell et al. (2004)

4.4 Methods

In Chapter 3, the model is developed using the rule-based Cubist regression model which is a tree-based regression model. Although Cubist model has shown great strengths in estimating the SAT at different regions, it requires substantial computational resources to train the model, especially for the model structure with committee enhancements. The high computational cost can limit the model's application to larger geographical extents. Thus, in this chapter I choose to replace the Cubist regression model with a non-parametric method with better computational efficiency, i.e., the multivariate adaptive regression spline model (MARS). Studies have shown that the MARS is suitable of handling large data volume without

sacrificing the performance of model output and the MARS is also very flexible and easy to understand and interpret \?references?\.

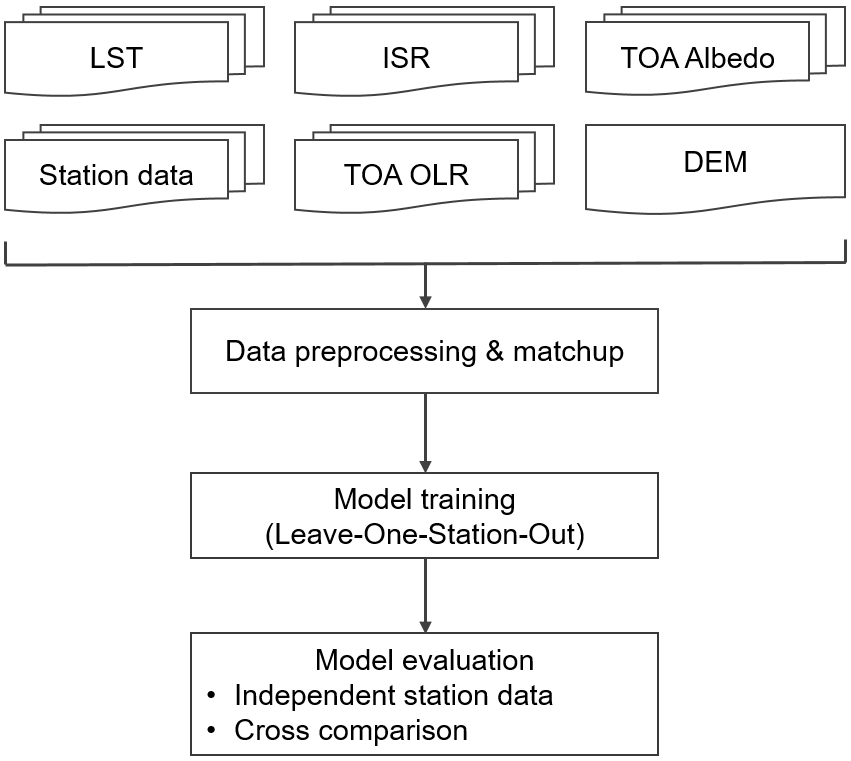

Figure 4-2. The overall research road map of Chapter 4.

4.4.1 The multivariate adaptive regression spline model (MARS)

MARS is a non-parametric regression model which can be viewed as the extension of a general linear regression model automatically accounting for the nonlinearities of a data set (Friedman, 1991; Hastie et al., 2009). It does not require that the relationship between the predictor and dependent variables is linear. In general, MARS modeling split the feature hyperspace of predictors into separate hyper-regions and then use a linear regression to characterize the relationship between dependent and independent variables. The joint point where the slope changes among different hyper-regions is defined as a knot and the set of knots determined by the MARS model is used to create a set of basis functions (or splines), which represents either transformations of a single

independent variable or interactions of multiple variables. The determination of the knots and basis functions is completely based on the data sets used for training and is specific to the problem in MARS. This data-based feature makes MARS a powerful and flexible adaptive regression method (Hastie et al., 2009; Zhou et al., 2018). Additionally, the MARS model also allows the inclusion of the product of multiple simple basis functions to represent the interactions between different variables. To what degree of the interactions between different variables may occur may be constrained by a user-defined parameter of the MARS model. This feature is unique to MARS and its later modifications (Kuter et al., 2015).

During the model building process, MARS uses a two-stage strategy, namely, "forward pass" and "backward pass". The forward pass is used to select a suitable collection of joins and corresponding basis functions, and it is repeated till a predefined maximum number has been reached. The foreword pass keeps the knot that gives the best fit and then fits the response using linear functions that are both nonzero on one size of the know. Once one variable is selected, the splits on other variables can depend on the existing splits. The forward pass will continue till a user predefined maximum number of basis functions. Once the forward pass has completed, there is a full model tree with a suite of knot and basis functions. The backward pass then evaluates the contribution of these splits and associated basis functions and decide whether they should be removed from the model or not (i.e., pruning) (Friedman, 1991; Hastie et al., 2009; Kuter et al., 2015). The pruning is designed to remove the basis functions which give the smallest increase in the residual sum of squares step-by-step. The purpose of this backward elimination process is to reduce the model complexity thus decrease the risk

of over-fitting. The final model is chosen based on a generalized cross-validation (GCV) measure of the mean square error (MSE). The procedure of the GCV is designed to evaluate which variables should be kept in the final model by introducing a penalty on including more variables to the model (Friedman, 1991; Hastie et al., 2009). Moreover, the CSV is also used to determine the importance of individual variables and rank the variables by computing the GCV with and without each variable in the model.

It should also be noted that "MARS" has been trademarked and licensed exclusively to Salford Systems. Although it is okay to use MARS as an abbreviation for the statistical model, it cannot be used as the name for other software that carrying out this model unless authorized by Salford Systems. In this chapter, I used the R software package named "*earth*" to carry out the MARS model and to apply the model to the entire NHL.

4.4.2 The strategy of model training and evaluation

To train the all sky MARS model, I first extracted the station-satellite data pairs for all 642 GHCN-D stations following the same data processing procedure documented in Chapter 3. Considering the stations have different length of measurements record (1-18 years since 2000), only the stations with the data records longer than 10 years are used for model training, which accounts for about 82% of the stations over the NHL. The rest of the station data of which the length of records is shorter than 10 years are kept as an independent set of data to evaluate the quality of MARS estimated SAT. The distribution of the data length of all 642 GHCN-D stations is presented in Figure 4-3.

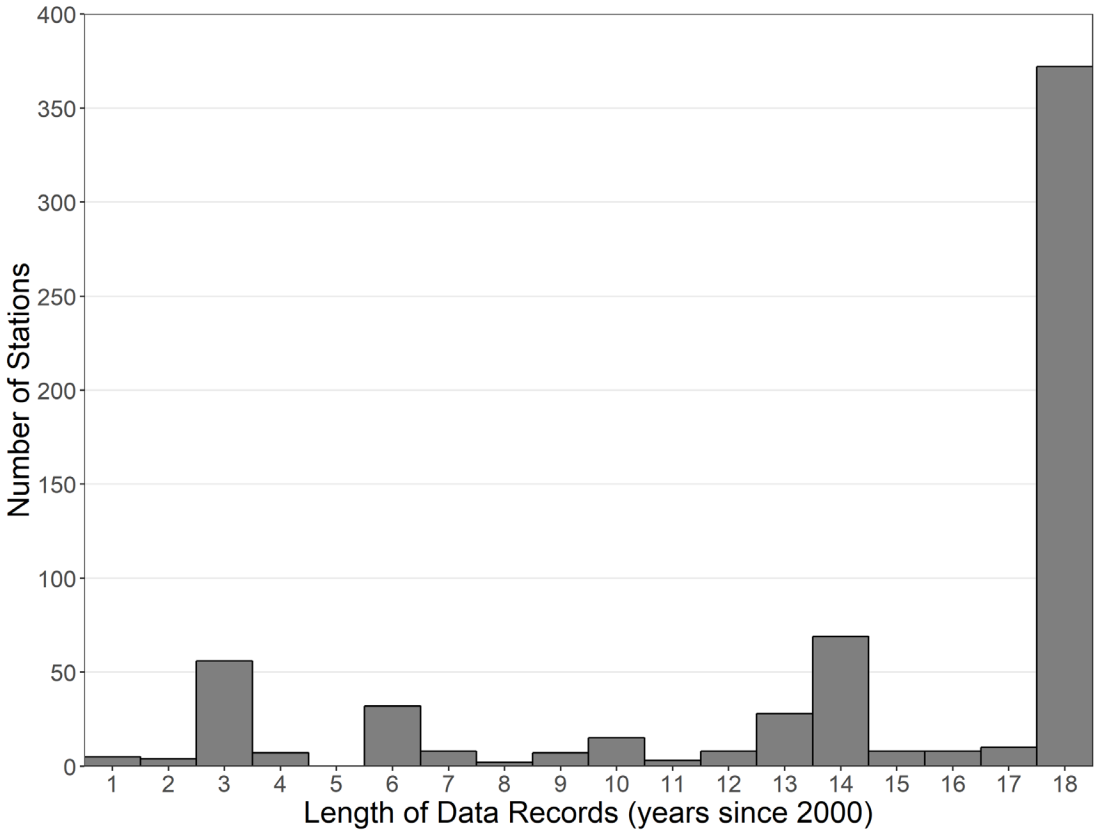

Figure 4-3. The histogram of the length of temperature records for all 642 stations within the GHCN-Daily archive.

In the MARS model, two parameters need to be determined through training, i.e., the degree of interactions allowed, and the maximum number of pruned terms allowed in the final model. During the training process, I used the LOSO strategy to select model parameters to avoid the overfitting issue as mentioned earlier (Meyer et al., 2016). Firstly, the training data were grouped by stations. For each iteration, a series of MARS models were fitted using different combinations of model parameters using data of all stations except one which was randomly chosen. Then, the models were evaluated using the data of the left-out station. After each station has been used as the left-out station to evaluate different model parameters, the final model parameters were selected based on the model performance across all iterations.

To evaluate the final MARS model, I first used the validation dataset of the GHCN-Daily stations with less than 10-year's data to assess the performance of the final MARS model when it is applied to the data of different stations. Then, I applied the final MARS model to the satellite data of the entire NHL for the year of 2002-2017. The estimated SAT of the NHL was cross compared with two existing datasets listed in Table 4-2. The main purpose of the cross comparison is to evaluate the spatial and temporal (i.e., seasonal) pattern of the estimated SAT. Additionally, I also compared the accuracy of our Cubist estimation and the existing datasets using GHCN-D station data as the reference since all datasets have their own uncertainty.

4.5 Estimating the SAT of the northern high latitudes

4.5.1 The results of model training and evaluation using station measurements

Figure 4-4 presents the density scatter plot of the training results of the final MARS model for the NHL. Generally, the MARS model estimated daily SAT show strong agreement with the GHCN-D station SAT measurements with the majority of the points distributing around the 1:1 line. The overall model accuracy and precision is good with the bias of -0.24 °C and the RMSE of 2.21 °C. There are still scattering around the 1:1 line, especially over the low temperature range (i.e., < -20 °C). Additionally, when the statistics of the MARS model are compared to the ones of the Cubist model listed in Chapter 3, the MARS model show slightly worse performance with the bias 0.1 °C higher and the RMSE nearly 0.4 °C higher than the Cubist model. This is likely caused by the increased uncertainty of the satellite products over the high latitude.

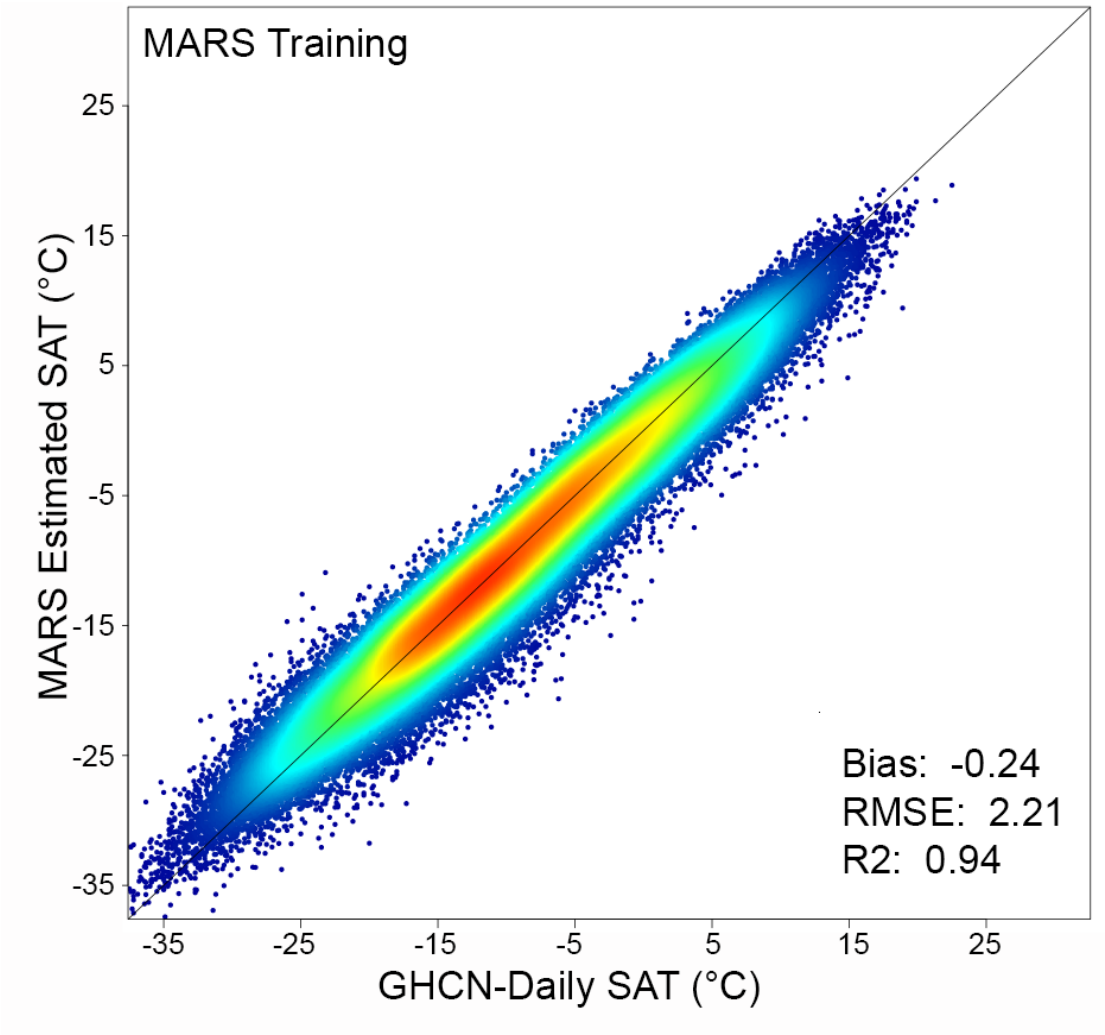

Figure 4-4. The density scatter plot of the training results for the MARS model. The red color indicates high point density while the blue color represents low point density.

Figure 4-5 exhibits the performance of the final MARS model for the NHL when it is evaluated using the set of independent GHCN-D station measured SAT. The density scatter plot of the independent evaluation shows similar pattern of the one of the model training result. The points are mostly clustered along the 1:1 line with larger scattering at the low temperature range. However, the RMSE of the independent evaluation is notably higher than the model training result by 0.4 °C. To further diagnosing this larger uncertainty, I suspect there might be systematic differences of the model

performance related to surface conditions. Therefore, I further separated the evaluation into two surface conditions: the permanent ice surface and the rest of the surface conditions. The data of the stations located at the permanent ice surface show similar bias with the rest of the data (i.e., -0.19 °C vs. -0.14 °C). But there are notable RMSE differences between these two surface conditions. The MARS estimated SAT show larger RMSE (2.87 °C) over the permanent ice surface when comparing to the GHCN-D station measurements while the RMSE of the rest data is only 2.35 °C.

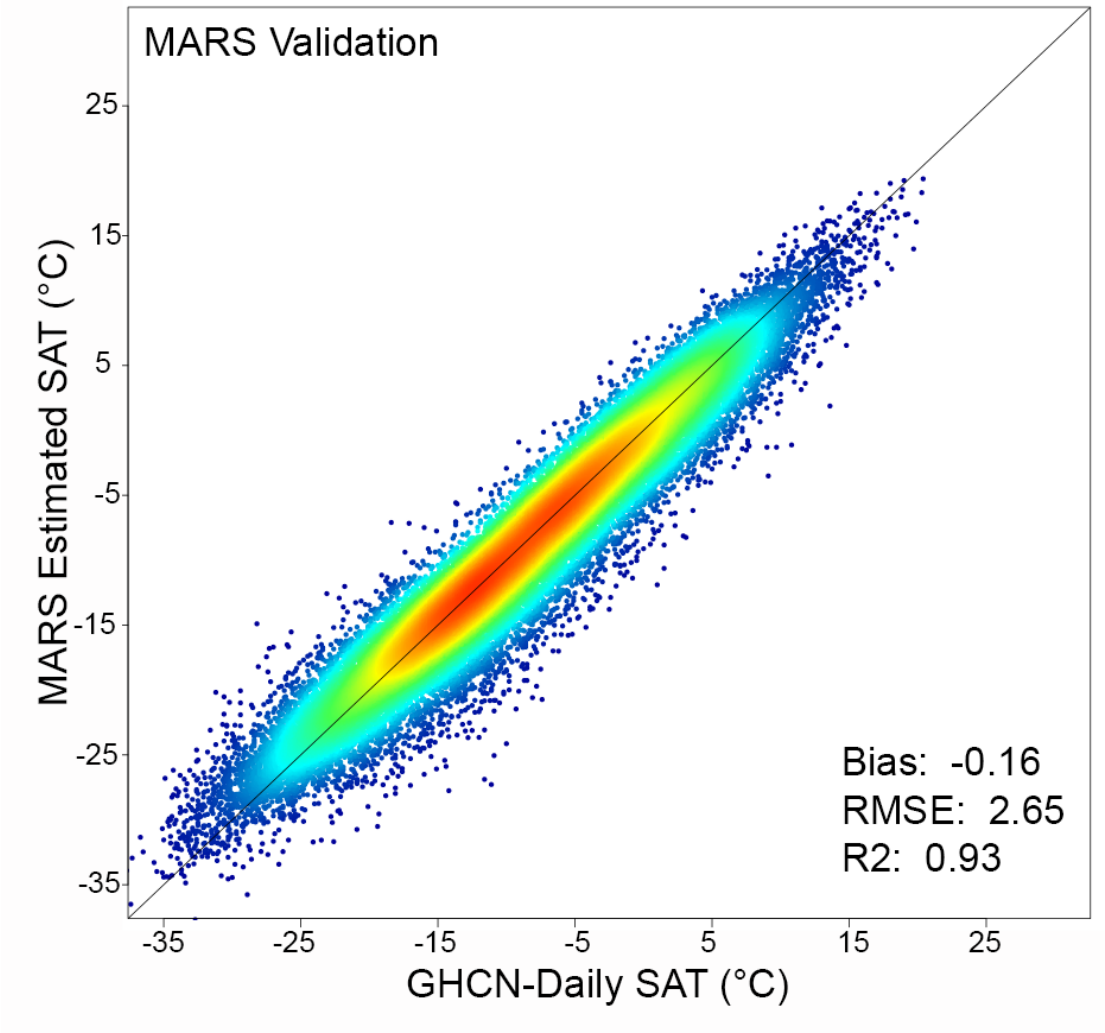

Figure 4-5. Similar with Figure 4-4, but for the validation results using the data of the independent GHCN-D stations.

However, the number of stations that are located over the permanent ice surface is very limited. There is only about 6% of the GHCN-D stations used in this chapter belongs to this surface condition. This could be one of the reason why the MARS model performs worse over permanent ice surface. Nonetheless, the overall model performance has shown satisfactory results for the NHL. The MARS model can provide a much needed high resolution daily all sky SAT dataset of the NHL for local and regional climate studies.

4.5.2 The cross comparison with model-based data

The all sky MARS model has been applied to the MODIS satellite data of the entire NHL. Before the MARS estimated SAT can be applied to relevant studies, it is cross compared with two existing gridded SAT datasets which covers the NHL at various spatial resolution, i.e., CRU-Ts.4.02 and GLDAS. Figure 4-6 presents the cross comparison of these three SAT estimations at monthly mean level for four different months of the year 2014 (i.e., January, April, July and October, 2014). These four months are chosen to represent four different seasons of the NHL. Generally, the MARS estimated SAT shows better agreement with GLDAS SAT estimation than with CRU-Ts.4.02, especially over the winter and summer months (i.e., 2014-01 and 2014-07). Overall, the MARS SAT estimation shows reasonable latitudinal and altitudinal gradients and seasonal dynamics of the estimated SAT.

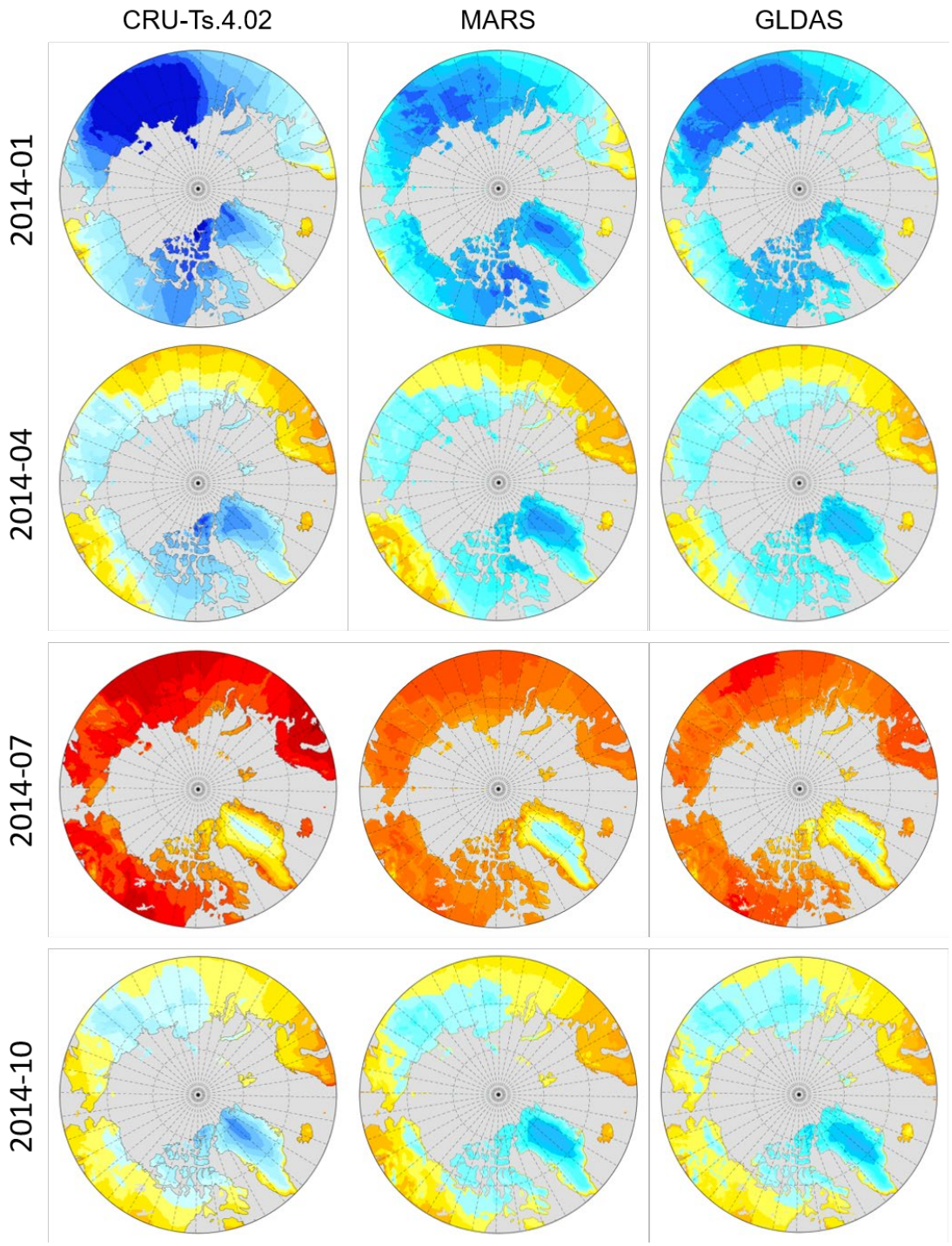

Figure 4-6. The cross comparison of the monthly mean SAT between the MARS model estimation with the SATs of CRU-Ts.4.02 and GLDAS for January, April, July, and October, 2014.

When compared with CRU-Ts.4.02, both MARS and GLDAS SAT estimations show smaller seasonal change of the SAT over the eastern part of Eurasia and the northern

Europe. Additionally, the spatial and seasonal patterns of the SAT estimation of Greenland demonstrated by CRU-Ts.4.02 show larger differences with the other two SAT estimations. Although there is no evidence supporting which estimation of the SAT is better, previous studies have questioned the quality of the CRU-Ts.4.02 over station sparse regions including the NHL. Over these station sparse regions, the 0.5°×0.5° grid of the CRU-Ts.4.02 may use station measurements that are far from the target grid to estimate the local SAT, which could lead to large data uncertainty.

4.6 Conclusion

In this chapter, the previously developed machine learning model for the Tibetan Plateau is adapted to the NHL with proper modifications to account for the unique features of the NHL. In order to handle with the large data volume, the original rule-based Cubist regression model is replaced by the MARS model for the NHL. Additionally, the model is designed to be able to handle the unique feature of polar day/night over the NHL by including the day length variable into the model as an indicator. The model is trained using the station temperature measurements over the NHL accessed through the GHCN-Daily data archive. The model training and evaluation results show that the modified machine learning model performs relatively well overall but has moderate dependency on the surface conditions. Specifically, the model estimated SAT shows larger RMSE and bias over the permanent ice surface. This surface condition dependency is likely caused by the insufficient amount of permanent ice surface data used during mode training. When the model is applied to the entire NHL, the estimated SAT data show reasonable spatial and seasonal dynamics of surface temperature comparing with two existing datasets, i.e., CRU-Ts.4.02 and

GLDAS. The successful adaption of the machine learning framework to the NHL demonstrated the genericity and applicability of this framework on estimating the surface air temperature using remote sensing products. This framework can be very beneficial when applied to regions with limited station temperature measurements. However, the model need to be carefully modified and trained using high quality of station measurements.

Chapter 5 Strengths and Pitfalls of Machine Learning Applications: The Lessons Learned

5.1 Leveraging machine learning and remote sensing

In recent decades, artificial intelligence (AI) technology has unequivocally accelerated scientific discovery in most, if not all, fields, including earth system science. As a subset of AI, machine learning has been gaining ground in earth system science. In the Web of Science database, there are 4052 publications related to the term "machine learning" within earth science related fields from 1991 to 2018 (Figure 5-1). Both the number of publications and citations has been growing exponentially. Additionally, the average number of citations for all previous publications is also increasing steadily, which indicates, in my opinion, the increasing popularity of machine learning in earth sciences.

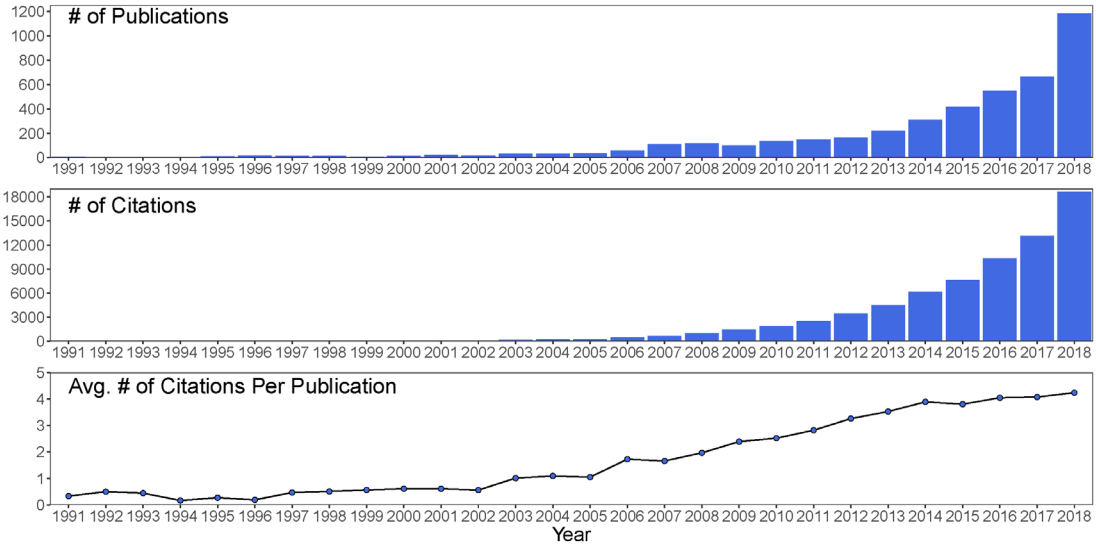

Figure 5-1. The growth of scientific publications on the topic of machine learning and earth sciences on the Web of Science between 1991 and 2018.

The name of “machine learning” has been coined in 1950s by computer scientists. Despite its early debut, machine learning has only gained its popularity since early 2000s in earth sciences. As statistical models, machine learning models are usually trained using existing data to answer questions of interest with good performance. The statistical model used for machine learning has always been around. In fact, most machine learning techniques are based on statistical models developed in the 1970s and 1980s. The major power engine of the machine learning explosion in earth science is the data and computation capability. Free access to the incredible amount of remote sensing data is the fuel, while the rapid development of high performance computation is the engine powering this rapid growth of machine learning applications.

In my dissertation research, I tested the framework of using well developed machine learning model (i.e., rule-based Cubist regression model and MARS model) to achieve the objective of generating high spatial resolution all sky daily SAT datasets for the Tibetan Plateau and the northern high latitudes, which are two regions experienced dramatic climate change. The research is driven by the fact that these two regions are lack of high quality climate data to reduce the uncertainty of regional and local climate analysis. Remote sensing data contains valuable observational information of the Earth’s surface and atmosphere, which can be used to estimate fundamental physical variables of the complex processes of the Earth system. I combined machine learning models with remote sensing products using the station measurements as the reference data to constrain the model. With careful design and proper training and evaluation, the machine learning model provides the estimation of the daily SAT for the two study regions. Different from existing studies, my model SAT estimations for both clear sky

and cloudy sky conditions with reasonable quality. This feature is very important for improving the quality of remote sensing based SAT datasets to address the issue of cloud contamination on thermal infrared observations. This improvement is a direct results of a physically meaningful machine learning model by using remote sensing radiation variables to account for the physical differences between remote sensing LST and SAT.

Although machine learning models can create great benefits as demonstrated in previous chapters, it is very important to understand the sensitivity of the model and what may cause large model uncertainty. In this chapter, I use the all sky Cubist model developed in Chapter 3 as an example to show how the quality and representativeness of the training datasets may lead to an unstable model thus causing large model uncertainty. I will also discuss the trade-off between model complexity, the risk of overestimation, and the computational cost for large scale applications from my own experiences while developing the model for the Tibetan Plateau. These are all lessons learned during the process of this dissertation research.

5.2 An unstable model and how to avoid it

While developing the all sky Cubist regression model, I used the LOSO strategy to avoid the risk of model overfitting caused by model training. Figure 5-2 shows the RMSE and R^2 of the final LOSO result of each individual CMA stations used during the model training process. In general, the final Cubist model performs well for most stations with RMSE lower than 2 °C and R^2 higher than 0.95 (Figure 5-2 (a, d)). However, there are some stations with relatively large uncertainty. These stations appear to be located at the regions with relatively complex terrain. Figure 5-2(b) shows

that stations above 4,000 meters ASL may show larger RMSE during this LOSO analysis. Figure 5-2(c) demonstrates that the Cubist model estimation of nearly 80% of the CMA stations has RMSE less than 2.1°C. This result shows comparable or slightly better performance than previous studies' clear sky only models. Overall, the LOSO result suggests that the Cubist model trained with limited amount of station data may be applied to other regions of the Plateau with acceptable accuracy for all sky conditions.

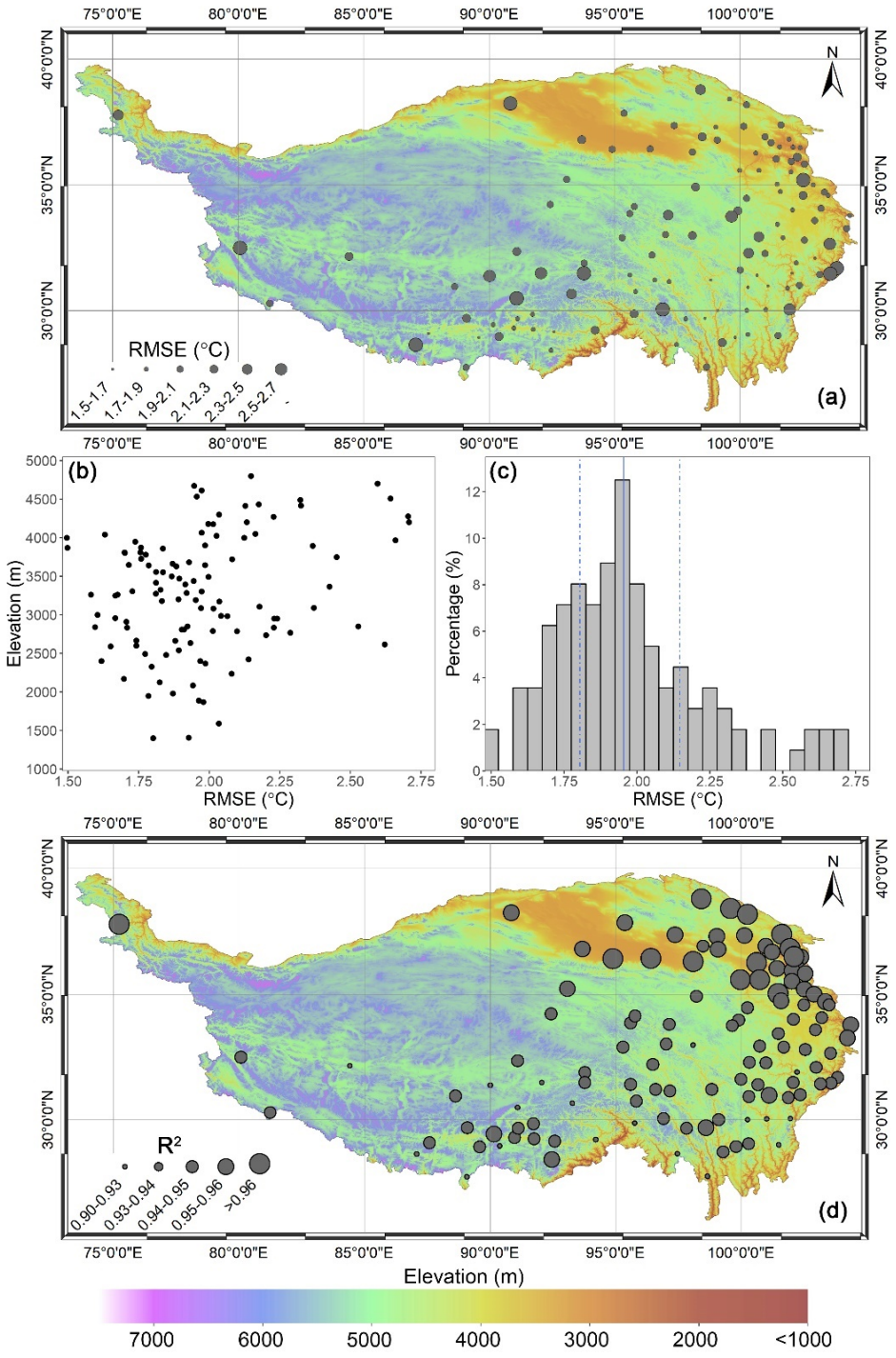

Figure 5-2. The results of the leave-one-station-out (LOSO) experiment: (a) the spatial distribution of the RMSE of all CMA stations; (b) the scatter plot between the RMSE and the elevation of all CMA stations; (c) the histogram of the RMSE of all CMA stations, where the solid blue vertical line indicates the median value and the two dashed vertical lines refer to the 25% and 75% quantiles respectively; (d) the

spatial distribution of the R^2 of all CMA stations. The background colors of (a) and (d) are the elevation of the GMTED2010 data.

To further investigate the robustness of the machine learning model, I carried out a 25-fold cross validation experiments. In this experiment, all CMA station-satellite data pairs were randomly divided into 25 folds by station ID. During each iteration, a Cubist model was fitted using 24 folds of data with the same parameters previously determined by the LOSO. This model was then evaluated using the left-out fold of data. This process was repeated 25 times until all 25 folds of data have been used to independently evaluate a Cubist model. This cross validation process is used to examine the sensitivity of the Cubist model on the training datasets. Figure 5-3 exhibits the RMSE of all 25 models trained with slightly different set of training data. In general, most models in the 25-fold cross validation experiment have relatively low RMSE (1.6°C – 2.1°C). However, some models (i.e., model No. 2, 9, 19) show relatively large RMSEs (>2.5°C) compared with other models.

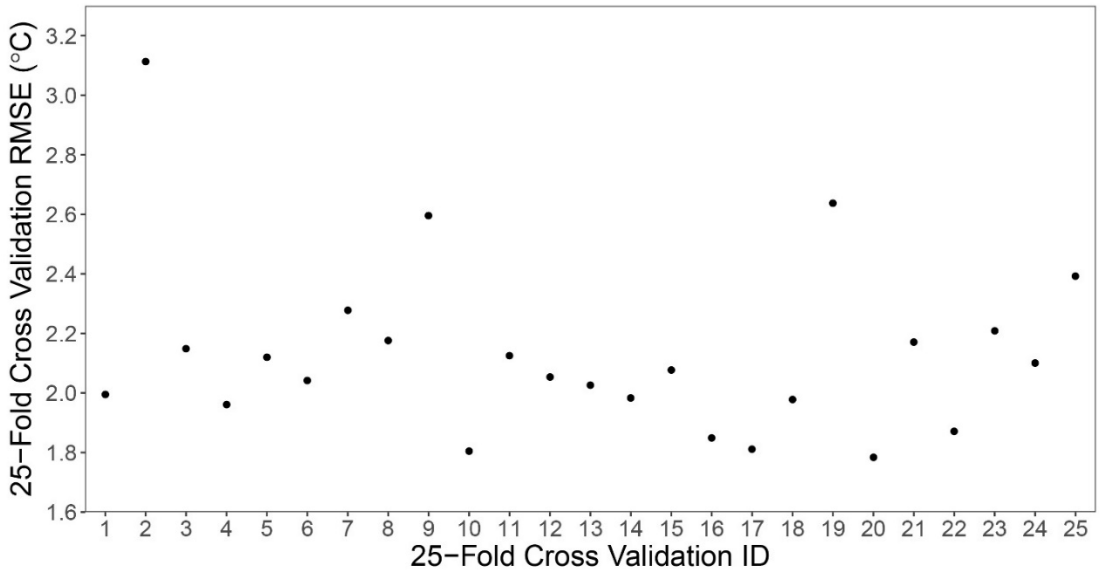

Figure 5-3. The RMSE of the 25-fold cross validation experiment using all CMA station data.

To better understand potential error sources, I examined the four Cubist models with largest RMSEs in the 25-fold cross validation experiment (i.e., Fold-2, 9, 19, 25). Figure 5-4 presents the density scatter plots of validation results, the comparison of data distributions of the training and validation data, and the quantile-quantile (Q-Q) plots between the training and validation datasets. For most of these cases (except model 2), the data distributions of training and validation data are notably different. For example, Fold-19 and Fold-25 all show that the distributions of their validation data are shifted rightwards from the distributions of their training data; Fold-9 exhibits a double-peak distribution of its validation data which is different from the near normal distribution of it training data. The Q-Q plots also confirm these differences among training/validation data's distribution. This common characteristic of these three cases underpins the assumption of machine learning models. Machine learning models are designed to predict unknown situations by learning from existing data/observations. The underlying assumption of most machine learning models is that the training data should represent the overall data distribution reasonably well. If this assumption is invalid, like in these cases, the performance of the prediction/estimation can be notably affected. Therefore, it is very important to ensure the representativeness of the data during model training process. Nevertheless, when I examined the distributions of the training data for all four cases, it is quite assuring to see that they share almost the same distribution despite the difference of their training data. This implies that when the amount of training data is large enough the sample distribution may be very close to the real data distribution. However, it is always the best practice to examine and

increase, if possible, the representativeness of the training data to ensure the trained machine learning model is not biased from the beginning.

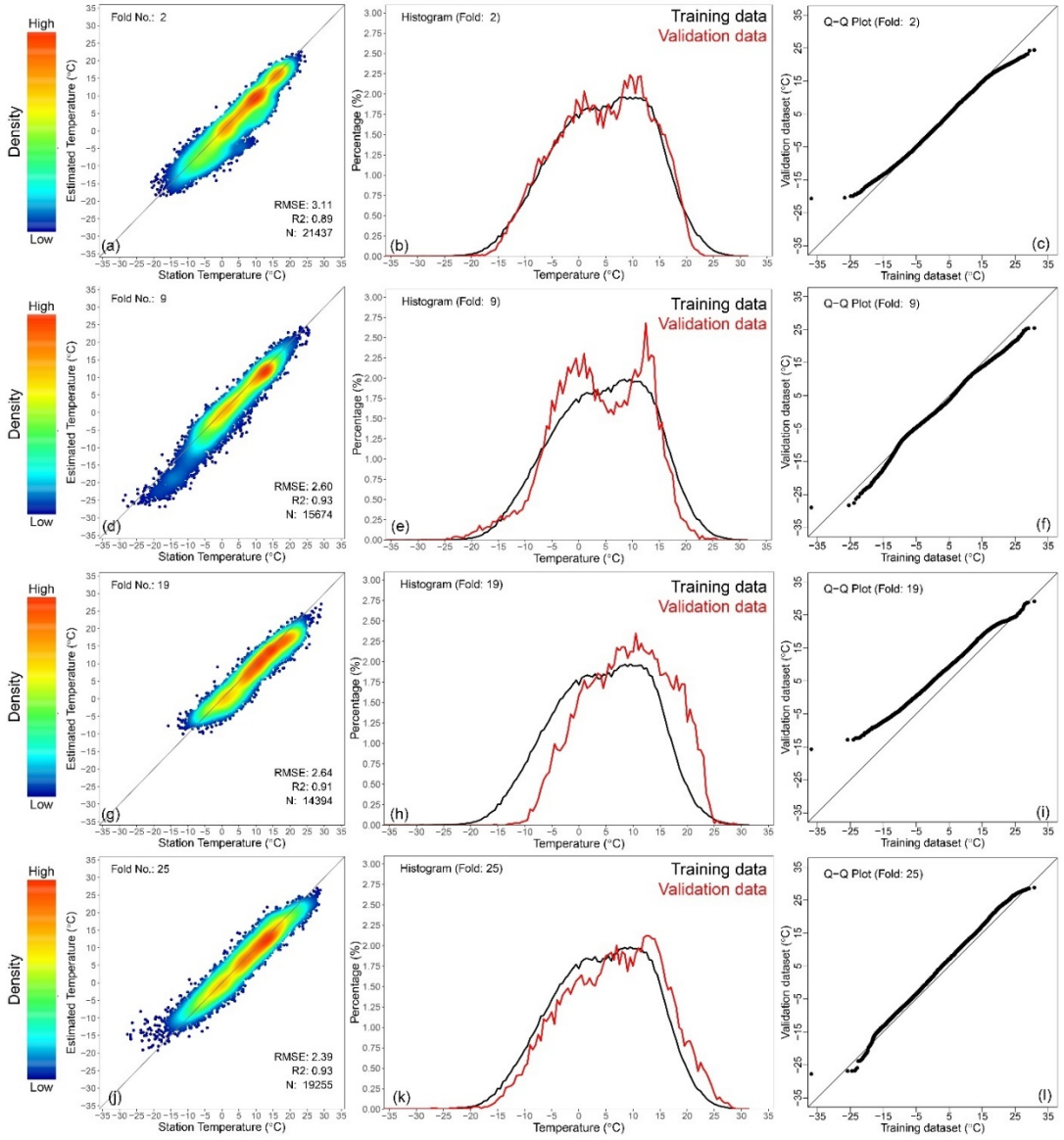

Figure 5-4. First column (a, d, g, j): the density scatter plots of four Cubist models with largest RMSEs in Figure 5-3 (i.e., No. 2, 9, 19, 25); second column (b, e, h, k): the comparisons of the data distributions of SATs from the training and validation datasets of each model; third column (c, f, i, l): the quantile-quantile (Q-Q) plots of SATs between the training and validation datasets of each model.

5.3 The model complexity trade-off

While developing a machine learning model, the users almost always have the freedom to define certain parameters of the model which may affect the performance of the model. In the Cubist model case, there are two key parameters need to be determined, i.e., the number of committees used and the number of neighbors used in the model. Among these two, the number of committees strongly affects the complexity of the final model since it defines how many iterations of a tree-based regression model are used in the model. The higher the number is the more complex the model is. As there is no theoretical way to determine the value of these parameters for machine learning model, users are always suggested to tune the parameters using either random or systematic searching approach. The random searching generates a set of parameters randomly and then chooses the final parameter value by selecting the one with the best model performance. The systematic searching choose the final parameter value from a comprehensive set of parameters defined through a systematic way, which is usually through a uniform sampling of the potential parameter spaces. In this dissertation, I choose systematic searching to determine the parameters of the Cubist model and MARS from a large set of parameters (typically 100~200 different combinations of the two used parameters). For the Cubist model described in Chapter 3, the model uncertainty defined as the RMSE shows a decreasing pattern with the increasing model complexity no matter how many neighbors are used to modify the final model estimation (Figure 5-5).

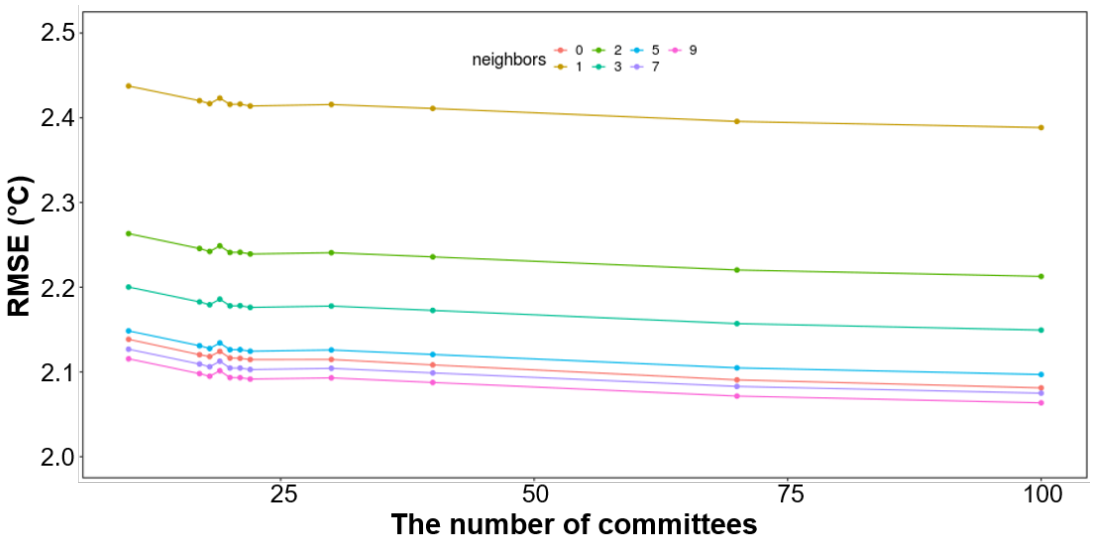

Figure 5-5. The relationship between the all-sky Cubist model uncertainty quantified using the root-mean-square-error (RMSE) and the model complexity defined by the number of committees used in the Cubist model.

However, there are also associated costs of the increasing model complexity. The first cost is the computational cost. The computational time of a Cubist model is linearly related to the number of committees used in the model. With more committees, the number of single tree-based model increases linearly. For example, using the same computational resources it cost roughly 17 hours to train a Cubist regression model with 100 committees while it only took 3.2 hours to train a Cubist model with 10 committees. Although the increasing model complexity reduces the model uncertainty, the trade-off between computational costs and accuracy gain need to be taken into consideration. In the Cubist case, the model performance only improves at a level of 0.05 °C while the computational cost more than doubled between a 50-committee model and a 100-committee model (Figure 5-5). This issue can be much more server when using more computational intense models or apply models to large volume of data.

Finally, the increasing model complexity may also cause higher risk of model overfitting. Theoretically, machine learning are statistical models trained with existing data. By increasing the model complexity, it can perfectly fit the data used during model training process. However, once the model is applied to a new set of data, it might perform poorly because the overfitted model cannot handle the information of the new data. To avoid this issue, different studies have suggested slightly different solutions. For example, while choosing the parameter for final models, users may choose the parameter yielding acceptable model accuracy (within 2% range or one standard deviation of all model accuracy values) and with lower complexity instead of choosing the best model with higher model complexity. Moreover, users can strategically design a model training process that theoretically can prevent the model overfitting issue from occurring, like the leave one station out cross validation strategy that was used in this dissertation research (Meyer et al., 2016; Noi et al., 2017).

Chapter 6 Conclusion and Future Work

6.1 Summary of the key findings

In this dissertation, I conducted a series of research to answer the overarching question of how I can use advanced statistical methods and remote sensing data to reduce the uncertainty of surface air temperature data at regional and local scales.

I firstly comprehensively compared four major global gridded land surface air temperature datasets, i.e., BEST-LAND, CRU-TEM4v, NASA-GISS, and NOAA-NCEI, at different spatial and temporal scales. The large scale mean LSAT anomalies are remarkably different because of the data coverage differences, with the magnitude nearly 0.4 °C for the global and Northern Hemisphere and 0.6 °C for the Southern Hemisphere. This study additionally finds that on the regional scale, northern high latitudes, southern middle-to-high latitudes, and the equator show the largest differences nearly 0.8 °C. These differences may cause notable differences for the trend calculation at regional scales. At the local scale, four datasets show significant variations over South America, Africa, Maritime Continent, central Australia, and Antarctica, which leads to remarkable differences in the local trend analysis. For some areas, different data sets produce conflicting results of whether warming exists. Our analysis shows that the differences across scales are associated with the availability of stations and the use of infilling techniques. My results suggest that conventional LSAT data sets using only station observations have large uncertainties across scales, especially over station-sparse areas. In developing future LSAT data sets, the data uncertainty caused by limited and unevenly distributed station observations must be reduced. To the best of my knowledge, this is the first comprehensive assessment of

the four major SAT datasets at local and regional scales since the IPCC Fifth Assessment Report, providing strong evidence for the future research to reduce the uncertainty of global gridded SAT datasets.

To reduce the uncertainty of regional and local surface air temperature datasets, I then developed a machine learning framework to estimate the daily SAT over the Tibetan Plateau with a high spatial resolution. This machine learning model is a novel all sky model based on the rule-based Cubist regression to estimate all sky daily SAT using LST, ISR, TOA albedo and OLR. The model is trained using station data of the Chinese Meteorological Administration and corresponding satellite products. The output is evaluated using independent station data with the bias of -0.07 °C and RMSE of 1.87°C. Additionally, the 25-fold cross validation shows a stable model performance (RMSE: 1.6-2.8 °C). Moreover, the all sky Cubist model increases the availability of the estimated SAT by nearly three times. I used the all sky Cubist model to estimate the daily SAT of the TP for 2002-2016 at 0.05°×0.05°. I also compared the all sky Cubist model estimated SAT with three reanalysis datasets (i.e., GLDAS, CLDAS, CMFD). My model estimation shows similar spatial and temporal dynamics with these existing data but outperforms them with lower bias and RMSE when benchmarked against CMA station data. Using the estimated SAT over the Tibetan Plateau, I found that the western part of the Plateau has experienced dramatic warming for the time period of 2002-2015, which did not reflect by any CMA station temperature measurements. The estimated SAT data could be very useful for regional and local climate studies over the TP.

Although the machine learning model is developed for the TP, the framework is generic and may be extended to other regions with proper model training using local data. Thus, I extended the machine learning model to a larger geographical extent, the northern high latitudes which has experienced dramatic temperature change in the last decades. To apply the machine learning model to the northern high latitudes, I modified the model to account for the differences between the TP and the NHL. Because of the large data volume of this study area, the rule-based Cubist regression model in the original framework has been replaced by the multivariate adaptive regression splines model. The adapted machine learning model is trained and evaluated using the Global Historical Climatology Network daily station data archive. Using data of 642 stations, the model estimated surface air temperature has good performance with the bias of -0.1 ~ -0.2 °C and the RMSE of 2.1 ~ 2.6 °C. However, due to lack of stations over the permanent ice surface, the model performs worse over the permanent ice surface than other surface conditions (RMSE: 2.8 °C vs. 2.4 °C). When the model is applied to the entire northern high latitudes, the estimated SAT show reasonable spatial and seasonal dynamics when compared with existing datasets but it can provide much detailed spatial information for climate analysis.

Overall, the research in this dissertation has demonstrated the great potential of leveraging machine learning and remote sensing products to estimate the SAT at high spatial resolution. The applications of this modeling framework to the Tibetan Plateau and the northern high latitudes show that the estimated SAT data provide unique value for local and regional climate analysis, demonstrating the important spatial details which are otherwise missing from the station-based datasets. However, the machine

learning model need to be trained and evaluated with extreme cautions to ensure that the model is not biased or unstable because of model overfitting or poor representativeness of training datasets.

6.2 Future research plan

Although this model framework has been applied to the time period starting from 2000 till 2017, the MODIS era, 18 years of data are not necessarily sufficient for climate studies considering the natural variability and other factors. Therefore, I plan to extend the model for both the Tibetan Plateau and the northern high latitudes to data from multiple satellite platforms, including (A)ATSR and AVHRR data. With the AVHRR data spanning from the 1980s, the model can generate a long-term time series of high resolution SAT dataset for nearly four decades for both the Tibetan Plateau and the northern high latitudes. However, attention need to be paid to address the inconsistency among different satellite platforms caused by different satellite overpassing time as well as the inconsistent sensor calibration. Additionally, I would like to extend the machine learning framework to estimate daily maximum/minimum temperature as well as the diurnal temperature range over these study regions since the extreme values of the SAT also play fundamental role on affecting ecosystem activities and regional environmental change.

Moreover, all input satellite data have different level of uncertainties which can be propagated into this statistical-based estimation. Therefore, it will be beneficial to understand the sensitivity of the estimated SAT regarding to the uncertainty of each individual input variables. In addition, there are growing demands of grid-level uncertainty assessment to improve the confidence of local and regional applications

using various climate datasets. Thus, it is of the best interest to provide the uncertainty value associated with each grid for the estimated SAT data using advanced statistical methods, such as, Markov Chain Monte Carlo (MCMC), bootstrapping etc.

Lastly, I would like to explore the possibility of extending the developed machine learning model to even larger geographical extent, e.g., continents where stations are sparse (like South America, Antarctica, and Africa). The successful application of the machine learning model for those continents will be very important to increase the confidence of climate analysis over these station sparse regions.

Appendix

Table A1. Estimated trends of difference series (using BEST-LAND as the reference) using common data coverage for global mean LSAT and hemispheric mean LSAT data. Number with bold red font indicate the difference trend is significant ($p<0.05$). The significance test use ARMA(1,1) to address autocorrelation issue in temperature time series (SH: southern hemisphere; NH: northern hemisphere; ANN: annual mean LSAT; MAM: March-April-May mean LSAT; JJA: June-July-August mean LSAT; SON: September-October-November mean LSAT; DJF: December-January-February mean LSAT).

		CRU-TEM4v					NASA-GISS					NOAA-NCEI				
		ANN	**MAM**	**JJA**	**SON**	**DJF**	**ANN**	**MAM**	**JJA**	**SON**	**DJF**	**ANN**	**MAM**	**JJA**	**SON**	**DJF**
1901-2017	**Global**	-0.005	-0.003	-0.005	-0.006	-0.004	-0.005	-0.004	-0.004	-0.008	-0.004	0.000	0.000	0.002	-0.001	-0.002
	SH	**-0.007**	**-0.006**	-0.002	**-0.008**	-0.010	0.005	0.009	**0.010**	-0.001	0.003	**0.018**	**0.022**	**0.016**	**0.015**	**0.018**
	NH	-0.004	-0.002	-0.006	-0.005	-0.002	-0.008	-0.008	-0.008	-0.010	-0.006	-0.006	**-0.007**	-0.003	-0.006	-0.008
1951-2017	**Global**	**0.014**	**0.010**	**0.011**	**0.019**	**0.018**	**0.014**	**0.013**	**0.011**	**0.015**	**0.016**	**0.012**	**0.006**	**0.008**	**0.017**	**0.015**
	SH	0.001	0.000	-0.003	-0.004	**0.011**	**0.020**	**0.024**	**0.011**	**0.014**	**0.032**	**0.026**	**0.030**	**0.012**	0.021	**0.043**
	NH	**0.019**	**0.013**	**0.016**	**0.027**	**0.020**	**0.012**	**0.009**	**0.011**	**0.016**	0.011	**0.007**	-0.002	**0.006**	**0.016**	**0.006**
1981-2017	**Global**	**0.027**	**0.022**	**0.030**	**0.029**	**0.025**	**0.008**	0.008	**0.016**	0.010	-0.002	0.002	-0.003	0.006	0.005	0.002
	SH	0.002	0.003	**0.021**	-0.017	0.003	**0.021**	**0.021**	**0.022**	0.011	**0.031**	**0.020**	**0.040**	0.008	-0.012	**0.044**
	NH	**0.035**	**0.028**	**0.033**	**0.044**	**0.032**	0.003	0.003	0.014	0.010	-0.014	-0.004	**-0.018**	0.005	0.011	-0.013
1998-2017	**Global**	0.019	0.012	**0.031**	**0.023**	0.005	0.009	0.011	0.018	0.019	-0.016	0.007	-0.004	0.021	0.011	0.002
	SH	-0.038	-0.040	-0.003	**-0.062**	**-0.042**	0.041	0.043	0.041	0.025	0.053	0.032	0.060	0.040	-0.026	0.059
	NH	**0.038**	**0.028**	**0.041**	**0.051**	0.021	-0.002	0.001	0.010	0.017	-0.039	-0.001	**-0.025**	0.014	0.023	-0.016

Table A2. Estimated trends of difference series (using BEST-LAND as the reference) using common data coverage for latitudinal band mean LSATs. Number with bold red font indicate the difference trend is significant ($p<0.05$). The significance test use ARMA(1,1) to address autocorrelation issue in temperature time series (ANN: annual mean LSAT; MAM: March-April-May mean LSAT; JJA: June-July-August mean LSAT; SON: September-October-November mean LSAT; DJF: December-January-February mean LSAT).

		1901-2017					1951-2017					1981-2017					1998-2017				
		ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF
CRU-TEM4v	S90-S70	-0.049	**-0.138**	-0.062	0.004	0.005	-0.049	**-0.138**	-0.062	0.004	0.005	-0.102	**-0.222**	-0.011	0.027	**-0.153**	**0.195**	-0.090	**0.240**	**0.523**	0.121
	S70-S50	-0.013	-0.005	-0.018	**-0.026**	**-0.015**	-0.002	0.015	-0.051	**-0.025**	0.016	**-0.082**	**-0.160**	**-0.154**	-0.038	-0.031	**-0.168**	**-0.249**	**-0.390**	**-0.099**	-0.047
	S50-S30	-0.004	**-0.008**	**0.004**	-0.005	-0.005	-0.002	**-0.009**	-0.001	-0.001	**0.006**	-0.005	-0.003	-0.002	**-0.013**	-0.004	-0.011	-0.001	-0.006	-0.011	-0.020
	S30-S10	-0.003	0.000	0.001	-0.008	-0.003	0.011	**0.017**	0.004	-0.004	**0.029**	0.004	0.016	0.017	-0.026	0.011	-0.059	-0.049	-0.026	**-0.100**	-0.054
	S10-N10	**-0.040**	**-0.035**	**-0.031**	**-0.039**	**-0.055**	-0.010	-0.016	-0.006	-0.005	**-0.015**	0.023	0.023	**0.048**	0.012	0.009	0.028	0.030	**0.047**	0.016	0.016
	N10-N30	-0.001	-0.005	0.003	0.002	-0.004	**0.036**	**0.026**	**0.040**	**0.044**	**0.037**	**0.047**	**0.038**	**0.042**	**0.058**	**0.042**	**0.045**	0.055	**0.043**	0.025	0.024
	N30-N50	-0.003	-0.003	-0.010	-0.004	0.004	**0.021**	**0.016**	0.008	**0.029**	**0.029**	**0.044**	**0.034**	**0.033**	**0.055**	**0.059**	**0.024**	**0.019**	**0.027**	**0.036**	0.025
	N50-N70	0.001	**0.007**	-0.004	-0.003	**0.004**	**0.009**	0.005	0.008	**0.020**	0.006	0.010	0.003	**0.017**	0.023	-0.005	0.036	-0.002	**0.052**	0.078	0.003
	N70-N90	0.034	0.003	0.015	0.056	0.032	0.022	-0.006	0.066	**0.033**	-0.025	**0.089**	**0.055**	**0.177**	**0.081**	**-0.084**	**0.126**	-0.017	**0.340**	0.157	**-0.263**
NASA-GISS	S90-S70	**-0.052**	**-0.132**	**-0.083**	0.004	0.001	**-0.052**	**-0.132**	**-0.083**	0.004	0.001	-0.088	**-0.241**	**-0.100**	0.082	-0.067	0.072	**-0.198**	-0.150	**0.545**	0.069
	S70-S50	**-0.032**	-0.034	**-0.039**	**-0.032**	**-0.032**	-0.026	-0.043	**-0.078**	**-0.020**	0.002	**-0.112**	**-0.210**	**-0.188**	-0.024	**-0.054**	**-0.175**	**-0.325**	**-0.413**	-0.031	0.008
	S50-S30	**-0.008**	**-0.010**	0.001	**-0.011**	**-0.014**	-0.004	-0.007	-0.003	-0.002	-0.002	-0.014	-0.018	-0.015	-0.008	-0.014	0.009	0.003	0.017	0.024	-0.013
	S30-S10	0.003	0.008	0.008	-0.005	0.004	**0.021**	**0.031**	0.009	0.009	**0.036**	**0.031**	**0.042**	**0.028**	0.006	**0.052**	0.055	0.080	0.058	0.013	0.074
	S10-N10	0.012	0.023	**0.020**	0.005	0.000	**0.034**	**0.035**	**0.034**	**0.034**	**0.032**	**0.018**	**0.019**	**0.026**	0.016	0.010	0.041	0.028	0.043	0.053	0.034
	N10-N30	-0.002	-0.006	**0.013**	-0.003	-0.012	**0.021**	**0.015**	**0.029**	**0.020**	**0.018**	**0.019**	**0.019**	**0.038**	**0.017**	-0.003	**0.022**	**0.041**	0.023	0.031	-0.016
	N30-N50	-0.014	-0.013	**-0.019**	-0.015	-0.009	**0.009**	**0.008**	-0.003	**0.013**	0.020	0.012	0.014	0.012	**0.020**	0.004	0.012	0.018	0.020	0.032	-0.030
	N50-N70	-0.007	-0.007	**-0.015**	-0.011	0.003	0.003	0.001	0.008	**0.012**	-0.008	-0.011	-0.014	0.014	0.003	-0.040	-0.029	-0.040	0.006	-0.010	-0.070
	N70-N90	0.012	0.016	-0.010	0.008	**0.056**	-0.024	-0.059	-0.028	-0.018	0.031	-0.176	**-0.273**	**-0.150**	-0.119	-0.185	**-0.445**	**-0.670**	-0.302	**-0.375**	-0.405
NOAA-NCEI	S90-S70	0.018	**-0.091**	0.049	**0.089**	0.007	0.018	**-0.091**	0.049	**0.089**	0.007	-0.005	-0.115	-0.036	0.156	-0.087	**0.249**	**-0.450**	**0.360**	**0.623**	-0.030
	S70-S50	**-0.026**	**-0.028**	-0.029	**-0.025**	-0.015	-0.042	**-0.080**	-0.080	0.004	-0.019	**-0.160**	**-0.228**	**-0.313**	-0.025	**-0.106**	**-0.224**	-0.160	**-0.618**	**-0.233**	-0.015

S50-S30	0.001	0.004	**0.008**	-0.006	-0.004	-0.006	-0.005	-0.001	-0.012	-0.005	**-0.019**	-0.002	-0.011	**-0.030**	**-0.033**	-0.004	0.017	0.024	-0.020	-0.035
S30-S10	**0.020**	**0.025**	**0.016**	**0.017**	**0.022**	**0.031**	**0.039**	0.012	0.022	**0.051**	**0.035**	**0.059**	0.017	-0.002	**0.068**	0.048	0.094	0.066	-0.038	0.078
S10-N10	**0.017**	**0.019**	**0.023**	**0.018**	0.007	**0.038**	**0.038**	**0.036**	**0.041**	**0.038**	0.014	**0.028**	0.017	-0.004	0.019	0.033	0.042	0.012	0.024	0.062
N10-N30	-0.002	**-0.011**	**0.017**	0.001	-0.015	**0.012**	0.000	**0.021**	**0.015**	0.008	0.003	**-0.021**	**0.017**	0.015	-0.003	-0.009	-0.033	0.009	0.003	-0.027
N30-N50	-0.011	**-0.011**	**-0.011**	-0.010	**-0.013**	**0.005**	-0.002	-0.005	**0.017**	**0.011**	**0.016**	0.000	0.012	**0.034**	0.017	**0.031**	0.005	**0.044**	**0.056**	0.016
N50-N70	-0.005	-0.002	**-0.014**	-0.009	0.003	-0.002	-0.009	-0.005	0.011	-0.004	**-0.023**	**-0.030**	-0.004	-0.006	**-0.044**	-0.015	**-0.047**	0.003	0.015	-0.040
N70-N90	0.006	-0.019	0.017	-0.020	0.062	-0.078	-0.126	-0.007	**-0.125**	-0.040	**-0.257**	**-0.375**	-0.069	**-0.292**	**-0.305**	**-0.530**	**-0.872**	-0.041	**-0.543**	-0.506

Table A3. Estimated trends of difference series (using BEST-LAND as the reference) using common data coverage for regional mean LSATs. Number with bold red font indicate the difference trend is significant ($p<0.05$). The significance test use ARMA(1,1) to address autocorrelation issue in temperature time series. (ANN: annual mean LSAT; MAM: March-April-May mean LSAT; JJA: June-July-August mean LSAT; SON: September-October-November mean LSAT; DJF: December-January-February mean LSAT).

		1901-2017					1951-2017					1981-2017					1998-2017				
		ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF	ANN	MAM	JJA	SON	DJF
CRU-TEM4v	Greenland	0.000	0.001	0.010	0.008	-0.005	-0.007	-0.013	0.014	0.005	-0.004	0.037	0.026	0.025	0.061	-0.015	**0.121**	**0.130**	0.115	**0.135**	0.101
	North America 2	**0.032**	**0.050**	**0.035**	**0.023**	0.023	0.003	**0.011**	**0.019**	0.003	**-0.021**	**-0.023**	-0.004	-0.004	**-0.019**	**-0.073**	-0.034	-0.006	-0.032	-0.004	-0.128
	Asia 2	-0.008	-0.002	-0.014	-0.018	0.001	**0.016**	0.002	0.009	**0.032**	**0.023**	0.026	0.001	0.032	0.044	**0.028**	0.064	0.032	**0.099**	0.119	**0.066**
	Europe	-0.010	**-0.011**	**-0.018**	-0.008	-0.004	0.004	0.002	-0.003	**0.013**	0.003	**0.028**	**0.017**	**0.025**	**0.038**	**0.028**	**0.032**	0.010	**0.029**	0.057	0.021
	North America 1	-0.004	-0.002	**-0.017**	-0.003	0.006	**0.008**	**0.012**	0.001	**0.011**	**0.011**	0.010	**0.021**	0.004	0.015	0.003	0.015	**0.039**	0.023	**0.049**	**-0.045**
	Asia 1	-0.004	-0.009	0.003	0.003	-0.007	**0.036**	**0.024**	**0.029**	**0.047**	**0.045**	**0.067**	**0.044**	**0.058**	**0.078**	**0.092**	0.021	0.001	0.033	0.008	0.059
	South America 1	**-0.044**	-0.036	**-0.035**	**-0.047**	**-0.053**	-0.014	-0.016	-0.016	**-0.020**	-0.006	-0.019	-0.021	-0.010	-0.026	-0.016	**-0.082**	-0.077	**-0.076**	**-0.120**	**-0.062**
	Africa	-0.006	-0.001	-0.002	-0.012	-0.006	**0.041**	**0.039**	**0.044**	**0.042**	**0.043**	**0.067**	**0.076**	**0.088**	**0.051**	**0.046**	**0.089**	**0.113**	**0.114**	0.070	0.033
	Maritime Continent	-0.042	-0.041	**-0.045**	-0.029	**-0.054**	-0.008	-0.005	-0.022	-0.001	-0.002	**0.040**	**0.064**	0.025	**0.035**	**0.025**	**0.088**	**0.123**	**0.074**	**0.080**	**0.090**
	Australia	0.003	0.000	0.006	0.004	0.001	**-0.011**	**-0.011**	**-0.011**	**-0.014**	-0.007	-0.010	-0.011	-0.004	**-0.018**	-0.007	**-0.046**	**-0.058**	-0.016	**-0.056**	**-0.045**
	South America 2	**-0.020**	**-0.021**	-0.007	**-0.019**	**-0.030**	**-0.030**	**-0.037**	**-0.027**	**-0.030**	**-0.027**	**-0.026**	-0.022	-0.035	**-0.034**	-0.014	**-0.044**	-0.015	**-0.080**	**-0.075**	-0.009
	Antarctica	0.014	0.017	-0.040	0.014	0.026	-0.013	0.006	**-0.074**	-0.010	0.025	**-0.114**	**-0.222**	**-0.157**	-0.037	**-0.079**	**-0.096**	**-0.275**	**-0.272**	-0.082	-0.015
NASA-GISS	Greenland	0.003	-0.002	**-0.021**	0.006	**0.039**	0.024	**0.037**	0.009	0.026	**0.050**	0.001	-0.026	0.023	0.032	0.025	0.062	0.059	**0.142**	0.091	0.059
	North America 2	0.001	0.007	0.003	-0.005	-0.004	0.012	0.011	0.005	-0.005	**-0.057**	**-0.055**	-0.032	-0.023	-0.031	**-0.134**	0.013	0.103	-0.022	-0.057	-0.035
	Asia 2	-0.016	**-0.019**	**-0.033**	-0.022	0.008	0.006	-0.015	-0.001	**-0.023**	0.018	-0.008	**-0.042**	-0.004	0.013	-0.004	**-0.104**	**-0.177**	-0.056	-0.097	-0.091
	Europe	-0.007	**-0.008**	-0.013	-0.008	0.000	**-0.006**	-0.004	-0.005	-0.009	-0.007	-0.012	-0.012	**-0.029**	-0.020	-0.008	-0.006	-0.020	-0.028	-0.011	-0.045
	North America 1	**-0.020**	-0.017	**-0.029**	**-0.022**	**-0.013**	-0.002	**-0.012**	**-0.013**	-0.001	-0.008	-0.009	-0.013	**-0.019**	-0.003	-0.028	-0.009	-0.067	-0.012	-0.035	-0.077
	Asia 1	0.000	-0.004	**-0.010**	-0.001	-0.006	**0.029**	**0.018**	**0.023**	**0.035**	**0.036**	**0.035**	**0.027**	**0.044**	**0.039**	**0.033**	**0.028**	-0.014	-0.050	**-0.043**	-0.004
	South America 1	0.020	0.032	**0.024**	0.010	0.012	**0.040**	**0.045**	**0.030**	**0.034**	**0.048**	**0.012**	0.011	0.015	0.000	**0.019**	0.025	0.023	-0.050	-0.015	-0.042
	Africa	-0.018	-0.014	-0.005	-0.028	-0.025	**-0.011**	**-0.015**	**-0.015**	-0.004	**-0.012**	**-0.019**	**-0.037**	**-0.038**	-0.006	-0.001	-0.023	**-0.062**	-0.009	-0.021	-0.005

	Maritime Continent	-0.005	-0.005	0.002	-0.003	-0.017	**-0.033**	**-0.034**	**-0.025**	**-0.030**	**-0.043**	-0.001	-0.009	0.000	-0.021	-0.006	-0.072	**-0.088**	-0.030	-0.037	**-0.139**
	Australia	**-0.014**	**-0.019**	**-0.016**	-0.007	**-0.013**	-0.020	**-0.024**	-0.015	-0.016	-0.026	-0.038	-0.026	-0.022	-0.043	**-0.063**	-0.119	-0.086	**-0.114**	-0.146	-0.137
	South America 2	**-0.015**	**-0.019**	-0.002	**-0.013**	**-0.026**	**-0.013**	**-0.018**	-0.015	**-0.010**	-0.010	**-0.037**	**-0.041**	**-0.044**	**-0.036**	**-0.029**	-0.019	-0.011	-0.020	-0.006	-0.043
	Antarctica	-0.003	-0.058	-0.054	-0.004	-0.052	-0.018	-0.052	**-0.073**	-0.003	-0.014	**-0.113**	**-0.258**	**-0.189**	-0.026	-0.047	**-0.131**	**-0.374**	**-0.425**	**-0.188**	-0.036
NOAA-NCEI	**Greenland**	**-0.031**	-0.032	**-0.028**	**-0.034**	-0.027	-0.029	-0.006	-0.019	**-0.055**	-0.063	-0.107	-0.116	**-0.087**	-0.070	-0.216	-0.130	-0.179	**-0.245**	-0.254	-0.064
	North America 2	-0.005	-0.016	**-0.011**	-0.004	-0.010	**-0.017**	-0.015	-0.006	-0.008	**-0.049**	**-0.046**	-0.022	**-0.027**	**-0.043**	**-0.081**	-0.016	-0.014	-0.004	-0.057	-0.032
	Asia 2	-0.015	-0.012	**-0.034**	-0.021	-0.010	-0.004	-0.015	-0.008	-0.016	-0.028	**-0.038**	**-0.072**	-0.017	-0.013	**-0.038**	**-0.079**	**-0.153**	-0.044	-0.066	-0.035
	Europe	**-0.012**	**-0.014**	**-0.016**	-0.010	-0.008	-0.010	**-0.021**	**-0.010**	-0.008	-0.013	-0.017	-0.001	**-0.021**	**-0.045**	-0.010	-0.024	-0.004	-0.042	-0.058	-0.023
	North America 1	**-0.005**	**-0.005**	**-0.006**	-0.005	-0.006	-0.006	-0.006	-0.003	**-0.012**	-0.001	-0.001	-0.002	-0.004	-0.012	**-0.027**	**-0.044**	**-0.049**	-0.031	**-0.084**	-0.001
	Asia 1	-0.007	**-0.013**	**-0.010**	-0.004	-0.021	**-0.015**	-0.003	**-0.011**	**-0.024**	**-0.021**	-0.013	-0.010	**-0.018**	**-0.020**	-0.029	-0.001	-0.055	-0.026	-0.011	-0.017
	South America 1	**-0.033**	**-0.034**	**-0.034**	**-0.030**	**-0.028**	**-0.046**	**-0.049**	**-0.028**	-0.043	**-0.064**	-0.010	-0.027	-0.006	-0.022	**-0.042**	-0.009	-0.010	-0.038	-0.062	-0.055
	Africa	-0.000	-0.002	-0.003	-0.003	-0.002	**-0.021**	**-0.023**	**-0.020**	**-0.019**	**-0.021**	**-0.036**	**-0.047**	**-0.043**	**-0.025**	**-0.029**	**-0.039**	**-0.081**	-0.038	-0.025	-0.007
	Maritime Continent	-0.013	-0.014	-0.010	-0.009	-0.021	**-0.023**	**-0.020**	-0.008	**-0.031**	**-0.031**	-0.021	-0.008	-0.021	-0.041	-0.015	-0.009	-0.054	-0.054	-0.039	-0.098
	Australia	-0.012	**-0.015**	**-0.014**	**-0.010**	**-0.009**	-0.016	-0.018	-0.013	-0.013	**-0.021**	-0.019	-0.020	-0.004	-0.020	-0.030	-0.076	-0.064	-0.094	-0.076	-0.074
	South America 2	-0.004	-0.007	-0.013	-0.002	**-0.007**	-0.012	-0.013	-0.009	-0.011	-0.013	**-0.047**	**-0.042**	**-0.063**	**-0.047**	**-0.039**	-0.020	-0.024	-0.048	-0.035	-0.016
	Antarctica	**-0.052**	**-0.114**	**-0.057**	**-0.036**	-0.022	-0.030	**-0.100**	-0.067	**-0.045**	-0.003	**-0.103**	**-0.210**	**-0.205**	**-0.068**	**-0.102**	-0.068	**-0.267**	**-0.323**	-0.103	-0.008

References

Alfieri, S.M., Lorenzi, F.D., Menenti, M., 2013. Mapping air temperature using time series analysis of LST: the SINTESI approach. Nonlinear Processes in Geophysics 20, 513–527. https://doi.org/10.5194/npg-20-513-2013

An, S., Zhu, X., Shen, M., Wang, Y., Cao, R., Chen, X., Yang, W., Chen, J., Tang, Y., 2018. Mismatch in elevational shifts between satellite observed vegetation greenness and temperature isolines during 2000–2016 on the Tibetan Plateau. Global Change Biology 24, 5411–5425. https://doi.org/10.1111/gcb.14432

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2014. PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. Bull. Amer. Meteor. Soc. 96, 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Beaudoing, H., Rodell, M., 2016. GLDAS Noah Land Surface Model L4 monthly 0.25 x 0.25 degree V2.1, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).

Bekryaev, R.V., Polyakov, I.V., Alexeev, V.A., 2010. Role of Polar Amplification in Long-Term Surface Air Temperature Variations and Modern Arctic Warming. J. Climate 23, 3888–3906. https://doi.org/10.1175/2010JCLI3297.1

Bhatt, U.S., Walker, D.A., Raynolds, M.K., Bieniek, P.A., Epstein, H.E., Comiso, J.C., Pinzon, J.E., Tucker, C.J., Steele, M., Ermold, W., Zhang, J., 2017. Changing seasonality of panarctic tundra vegetation in relationship to climatic variables. Environ. Res. Lett. 12, 055003. https://doi.org/10.1088/1748-9326/aa6b0b

Brohan, P., Kennedy, J.J., Harris, I., Tett, S.F.B., Jones, P.D., 2006. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. J. Geophys. Res. 111, D12106. https://doi.org/10.1029/2005JD006548

Cao, R., Shen, M., Zhou, J., Chen, J., 2018. Modeling vegetation green-up dates across the Tibetan Plateau by including both seasonal and daily temperature and precipitation. Agricultural and Forest Meteorology 249, 176–186. https://doi.org/10.1016/j.agrformet.2017.11.032

Chahine, M.T., Pagano, T.S., Aumann, H.H., Atlas, R., Barnet, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E.J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F.W., Kakar, R., Kalnay, E., Lambrigtsen, B.H., Lee, S.-Y., Le Marshall, J., McMillan, W.W., McMillin, L., Olsen, E.T., Revercomb, H., Rosenkranz, P., Smith, W.L., Staelin, D., Strow, L.L., Susskind, J., Tobin, D., Wolf, W., Zhou, L., 2006. AIRS: Improving Weather Forecasting and Providing New Data on Greenhouse Gases. Bull. Amer. Meteor. Soc. 87, 911–926. https://doi.org/10.1175/BAMS-87-7-911

Chen, F., Liu, Y., Liu, Q., Qin, F., 2014. A statistical method based on remote sensing for the estimation of air temperature in China. International Journal of Climatology 35, 2131–2143. https://doi.org/10.1002/joc.4113

Chen, J., Jönsson, Per., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set

based on the Savitzky–Golay filter. Remote Sensing of Environment 91, 332–344. https://doi.org/10.1016/j.rse.2004.03.014

Chen, X., Long, D., Hong, Y., Hao, X., Hou, A., 2018. Climatology of snow phenology over the Tibetan plateau for the period 2001–2014 using multisource data. International Journal of Climatology 38, 2718–2729. https://doi.org/10.1002/joc.5455

Chen, Y., Yang, K., He, J., Qin, J., Shi, J., Du, J., He, Q., 2011. Improving land surface temperature modeling for dry land of China. Journal of Geophysical Research: Atmospheres 116. https://doi.org/10.1029/2011JD015921

Cohen, J., Screen, J.A., Furtado, J.C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., Jones, J., 2014. Recent Arctic amplification and extreme mid-latitude weather. Nature Geoscience 7, 627–637. https://doi.org/10.1038/ngeo2234

Cong, N., Shen, M., Piao, S., 2017. Spatial variations in responses of vegetation autumn phenology to climate change on the Tibetan Plateau. J Plant Ecol 10, 744–752. https://doi.org/10.1093/jpe/rtw084

Dai, A., 2013. Increasing drought under global warming in observations and models. Nature Clim. Change 3, 52–58. https://doi.org/10.1038/nclimate1633

Dee, D.P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A.J., Thépaut, J.-N., 2013. Toward a Consistent Reanalysis of the Climate System. Bull. Amer. Meteor. Soc. 95, 1235–1248. https://doi.org/10.1175/BAMS-D-13-00043.1

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828

Didan, K., 2015a. MOD13C1 MODIS/Terra Vegetation Indices 16-Day L3 Global 0.05Deg CMG V006. https://doi.org/10.5067/MODIS/MOD13C1.006

Didan, K., 2015b. MYD13C1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 0.05Deg CMG V006. https://doi.org/10.5067/MODIS/MYD13C1.006

Döscher, R., Vihma, T., Maksimovich, E., 2014. Recent advances in understanding the Arctic climate system state and change from a sea ice perspective: a review. Atmospheric Chemistry and Physics 14, 13571–13600. https://doi.org/10.5194/acp-14-13571-2014

Editorial, 2017. Expanding research views. Nature Climate Change 7, 229. https://doi.org/10.1038/nclimate3270

Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J.R., Pielke, R.A., 2011. Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. J. Geophys. Res. 116, D14120. https://doi.org/10.1029/2010JD015146

Frenne, P.D., Verheyen, K., 2016. Weather stations lack forest data. Science 351, 234–234. https://doi.org/10.1126/science.351.6270.234-a

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. Ann. Statist. 19, 1–67. https://doi.org/10.1214/aos/1176347963

Gantner, L., Winkler, P., Köhler, U., 2000. A method to derive long-term time series and trends of UV-B radiation (1968–1997) from observations at Hohenpeissenberg (Bavaria). Journal of Geophysical Research: Atmospheres 105, 4879–4888. https://doi.org/10.1029/1999JD900907

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Good, E., 2015. Daily minimum and maximum surface air temperatures from geostationary satellite data. J. Geophys. Res. Atmos. 120, 2014JD022438. https://doi.org/10.1002/2014JD022438

Good, E.J., 2016. An in situ-based analysis of the relationship between land surface "skin" and screen-level air temperatures. J. Geophys. Res. Atmos. 121, 2016JD025318. https://doi.org/10.1002/2016JD025318

Good, E.J., Ghent, D.J., Bulgin, C.E., Remedios, J.J., 2017. A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series. J. Geophys. Res. Atmos. 122, 2017JD026880. https://doi.org/10.1002/2017JD026880

Groisman, P., Soja, A., 2007. Northern Hemisphere high latitude climate and environmental change. Environ. Res. Lett. 2, 045008. https://doi.org/10.1088/1748-9326/2/4/045008

Hall, D.K., Comiso, J.C., DiGirolamo, N.E., Shuman, C.A., Box, J.E., Koenig, L.S., 2013. Variability in the surface temperature and melt extent of the Greenland ice sheet from MODIS. Geophysical Research Letters 40, 2114–2120. https://doi.org/10.1002/grl.50240

Hall, D.K., Riggs, G.A., 2015a. MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG, Version 6. https://doi.org/10.5067/MODIS/MOD10C1.006

Hall, D.K., Riggs, G.A., 2015b. MODIS/Aqua Snow Cover Daily L3 Global 0.05Deg CMG, Version 6. https://doi.org/10.5067/MODIS/MYD10C1.006

Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global Surface Temperature Change. Rev. Geophys. 48, RG4004. https://doi.org/10.1029/2010RG000345

Harris, I., Jones, P. d., Osborn, T. j., Lister, D. h., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int. J. Climatol. 34, 623–642. https://doi.org/10.1002/joc.3711

Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A., Soden, B.J., Thorne, P.W., Wild, M., Zhai, P.M., 2013. Observations: Atmosphere and Surface, in: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working

Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 159–254. https://doi.org/10.1017/CBO9781107415324.008

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics. Springer New York.

Hausfather, Z., Cowtan, K., Clarke, D.C., Jacobs, P., Richardson, M., Rohde, R., 2017. Assessing recent warming using instrumentally homogeneous sea surface temperature records. Science Advances 3, e1601207. https://doi.org/10.1126/sciadv.1601207

Hegerl, G.C., Black, E., Allan, R.P., Ingram, W.J., Polson, D., Trenberth, K.E., Chadwick, R.S., Arkin, P.A., Sarojini, B.B., Becker, A., Dai, A., Durack, P.J., Easterling, D., Fowler, H.J., Kendon, E.J., Huffman, G.J., Liu, C., Marsh, R., New, M., Osborn, T.J., Skliris, N., Stott, P.A., Vidale, P.-L., Wijffels, S.E., Wilcox, L.J., Willett, K.M., Zhang, X., 2014. Challenges in Quantifying Changes in the Global Water Cycle. Bull. Amer. Meteor. Soc. 96, 1097–1115. https://doi.org/10.1175/BAMS-D-13-00212.1

Hersbach, H., Dee, D., 2016. ERA5 reanalysis is in production. ECMWF newsletter 147.

Huang, F., Ma, W., Wang, B., Hu, Z., Ma, Y., Sun, G., Xie, Z., Lin, Y., 2017. Air temperature estimation with MODIS data over the Northern Tibetan Plateau. Adv. Atmos. Sci. 34, 650–662. https://doi.org/10.1007/s00376-016-6152-5

Huang, J., Zhang, X., Zhang, Q., Lin, Y., Hao, M., Luo, Y., Zhao, Z., Yao, Y., Chen, X., Wang, L., Nie, S., Yin, Y., Xu, Y., Zhang, J., 2017. Recently amplified arctic warming has contributed to a continual global warming trend. Nature Climate Change 7, 875. https://doi.org/10.1038/s41558-017-0009-5

Huntington, T.G., 2006. Evidence for intensification of the global water cycle: Review and synthesis. Journal of Hydrology 319, 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003

Hwang, J., Choi, Y.-S., Kim, W., Su, H., Jiang, J.H., 2018. Observational estimation of radiative feedback to surface air temperature over Northern High Latitudes. Clim Dyn 50, 615–628. https://doi.org/10.1007/s00382-017-3629-6

IPCC, 2014. Climate Change 2014: Synthesis Report.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1017/CBO9781107415324

Ji, F., Wu, Z., Huang, J., Chassignet, E.P., 2014. Evolution of land surface air temperature trend. Nature Clim. Change 4, 462–466. https://doi.org/10.1038/nclimate2223

Jones, P., 2016. The reliability of global and hemispheric surface temperature records. Adv. Atmos. Sci. 33, 269–282. https://doi.org/10.1007/s00376-015-5194-4

Jones, P., Lister D. H., Osborn T. J., Harpham C., Salmon M., Morice C. P., 2012. Hemispheric and large-scale land-surface air temperature variations: An

extensive revision and an update to 2010. Journal of Geophysical Research: Atmospheres 117. https://doi.org/10.1029/2011JD017139

Jones, P.D., Lister, D.H., Osborn, T.J., Harpham, C., Salmon, M., Morice, C.P., 2012. Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. J. Geophys. Res. 117, D05127. https://doi.org/10.1029/2011JD017139

Jung, T., Gordon, N.D., Bauer, P., Bromwich, D.H., Chevallier, M., Day, J.J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H.F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I.A., Smith, G., Svensson, G., Tolstykh, M., Yang, Q., 2016. Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. Bull. Amer. Meteor. Soc. 97, 1631–1647. https://doi.org/10.1175/BAMS-D-14-00246.1

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J.J., Fiorino, M., Potter, G.L., 2002. NCEP–DOE AMIP-II Reanalysis (R-2). Bull. Amer. Meteor. Soc. 83, 1631–1643. https://doi.org/10.1175/BAMS-83-11-1631

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. Journal of the Meteorological Society of Japan. Ser. II 93, 5–48. https://doi.org/10.2151/jmsj.2015-001

Kuhn, M., Johnson, K., 2013. Regression trees and rule-based models, in: Applied Predictive Modeling. Springer, pp. 173–220.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, J.R., 2018. Cubist: Rule- And Instance-Based Regression Modeling.

Kuter, S., Weber, G.-W., Akyürek, Z., Özmen, A., 2015. Inversion of top of atmospheric reflectance values by conic multivariate adaptive regression splines. Inverse Problems in Science and Engineering 23, 651–669. https://doi.org/10.1080/17415977.2014.933828

Laudon, H., Spence, C., Buttle, J., Carey, S.K., McDonnell, J.J., McNamara, J.P., Soulsby, C., Tetzlaff, D., 2017. Save northern high-latitude catchments. Nature Geoscience 10, 324–325. https://doi.org/10.1038/ngeo2947

Lee, J., Lund, R., 2004. Revisiting simple linear regression with autocorrelated errors. Biometrika 91, 240–245. https://doi.org/10.1093/biomet/91.1.240

Li, X., Zhou, Y., Asrar, G.R., Zhu, Z., 2018. Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States. Remote Sensing of Environment 215, 74–84. https://doi.org/10.1016/j.rse.2018.05.034

Loranty, M.M., Abbott, B.W., Blok, D., Douglas, T.A., Epstein, H.E., Forbes, B.C., Jones, B.M., Kholodov, A.L., Kropp, H., Malhotra, A., Mamet, S.D., Myers-Smith, I.H., Natali, S.M., O'Donnell, J.A., Phoenix, G.K., Rocha, A.V., Sonnentag, O., Tape, K.D., Walker, D.A., 2018. Reviews and syntheses: Changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions. Biogeosciences 15, 5287–5313. https://doi.org/10.5194/bg-15-5287-2018

Lu, N., Liang, S., Huang, G., Qin, J., Yao, L., Wang, D., Yang, K., 2018. Hierarchical Bayesian space-time estimation of monthly maximum and minimum surface air temperature. Remote Sensing of Environment 211, 48–58. https://doi.org/10.1016/j.rse.2018.04.006

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An Overview of the Global Historical Climatology Network-Daily Database. J. Atmos. Oceanic Technol. 29, 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Menne, M.J., Williams, C.N., 2009. Homogenization of Temperature Series via Pairwise Comparisons. J. Climate 22, 1700–1717. https://doi.org/10.1175/2008JCLI2263.1

Menne, M.J., Williams, C.N., Palecki, M.A., 2010. On the reliability of the U.S. surface temperature record. Journal of Geophysical Research: Atmospheres 115. https://doi.org/10.1029/2009JD013094

Meyer, H., Katurji, M., Appelhans, T., Müller, M.U., Nauss, T., Roudier, P., Zawar-Reza, P., 2016. Mapping Daily Air Temperature for Antarctica Based on MODIS LST. Remote Sensing 8, 732. https://doi.org/10.3390/rs8090732

Muller, R.A., Rohde, R., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., Wickham, C., 2013. A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview 2013. https://doi.org/10.4172/2327-4581.1000101

National Research Council, 2004. Climate Data Records from Environmental Satellites: Interim Report. https://doi.org/10.17226/10944

Nature Geoscience, 2014. Hiatus in context. Nature Geoscience 7, 157–157. https://doi.org/10.1038/ngeo2116

Nielsen-Englyst, P., Høyer, J.L., Madsen, K.S., Tonboe, R., Dybkjær, G., Alerskans, E., 2019. In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic. The Cryosphere 13, 1005–1024. https://doi.org/10.5194/tc-13-1005-2019

Noi, P.T., Degener, J., Kappas, M., 2017. Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data. Remote Sensing 9, 398. https://doi.org/10.3390/rs9050398

Overland, J.E., Dethloff, K., Francis, J.A., Hall, R.J., Hanna, E., Kim, S.-J., Screen, J.A., Shepherd, T.G., Vihma, T., 2016. Nonlinear response of mid-latitude weather to the changing Arctic. Nature Climate Change 6, 992–999. https://doi.org/10.1038/nclimate3121

Pan, Z., Arritt, R.W., Takle, E.S., Gutowski, W.J., Anderson, C.J., Segal, M., 2004. Altered hydrologic feedback in a warming climate introduces a “warming hole.” Geophys. Res. Lett. 31, L17109. https://doi.org/10.1029/2004GL020528

Pearson, R.G., Phillips, S.J., Loranty, M.M., Beck, P.S.A., Damoulas, T., Knight, S.J., Goetz, S.J., 2013. Shifts in Arctic vegetation and associated feedbacks under climate change. Nature Climate Change 3, 673–677. https://doi.org/10.1038/nclimate1858

Peng, G., Shi, L., Stegall, S.T., Matthews, J.L., Fairall, C.W., 2016. An Evaluation of HIRS Near-Surface Air Temperature Product in the Arctic with SHEBA Data.

J. Atmos. Oceanic Technol. 33, 453–460. https://doi.org/10.1175/JTECH-D-15-0217.1

Pepin, N., Bradley, R.S., Diaz, H.F., Baraer, M., Caceres, E.B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M.Z., Liu, X.D., Miller, J.R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M.B., Williamson, S.N., Yang, D.Q., 2015. Elevation-dependent warming in mountain regions of the world. Nature Climate Change 5, 424.

Pepin, N.C., Maeda, E.E., Williams, R., 2016. Use of remotely sensed land surface temperature as a proxy for air temperatures at high elevations: Findings from a 5000 m elevational transect across Kilimanjaro. J. Geophys. Res. Atmos. 121, 2016JD025497. https://doi.org/10.1002/2016JD025497

Pithan, F., Mauritsen, T., 2014. Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nature Geoscience 7, 181–184. https://doi.org/10.1038/ngeo2071

Prince, S.D., Goetz, S.J., Dubayah, R.O., Czajkowski, K.P., Thawley, M., 1998. Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. Journal of Hydrology 212–213, 230–249. https://doi.org/10.1016/S0022-1694(98)00210-8

Qin, J., Yang, K., Liang, S., Guo, X., 2009. The altitudinal dependence of recent rapid warming over the Tibetan Plateau. Climatic Change 97, 321. https://doi.org/10.1007/s10584-009-9733-9

Quinlan, J.R., 1993a. Constructing Decision Trees, in: C4.5: Programs for Machine Learning. Elsevier, pp. 17–26.

Quinlan, J.R., 1993b. Combining instance-based and model-based learning, in: Proceedings of the Tenth International Conference on Machine Learning. pp. 236–243.

Quinlan, J.R., 1992. Learning with continuous classes, in: 5th Australian Joint Conference on Artificial Intelligence. World Scientific, pp. 343–348.

Rao, Y., Liang, S., Yu, Y., 2018. Land Surface Air Temperature Data Are Considerably Different Among BEST-LAND, CRU-TEM4v, NASA-GISS, and NOAA-NCEI. Journal of Geophysical Research: Atmospheres 123, 5881–5900. https://doi.org/10.1029/2018JD028355

Rayner, N.A., Auchmann, R., Bessembinder, J., Bronnimann, S., Brugnara, Y., Capponi, F., Conway, E.A., Dodd, E., Englyst, P.N., Ghent, D., Good, E., Hoyer, J., Kennedy, J., van der Linden, P., Lindgren, F., Madsen, K.S., Merchant, C.J., Mitchelson, J., Morice, C.P., Remedios, J.J., Squintu, A., van der Schrier, G., Stephens, A., Tonboe, R.T., Veal, K.L., Waterfall, A.M., Woolway, I., 2018. The EUSTACE project: delivering global, daily information on surface air temperature. AGU Fall Meeting Abstracts 24.

Rennie, J.J., Lawrimore, J.H., Gleason, B.E., Thorne, P.W., Morice, C.P., Menne, M.J., Williams, C.N., de Almeida, W.G., Christy, J. r., Flannery, M., Ishihara, M., Kamiguchi, K., Klein-Tank, A.M.G., Mhanda, A., Lister, D.H., Razuvaev, V., Renom, M., Rusticucci, M., Tandy, J., Worley, S.J., Venema, V., Angel, W., Brunet, M., Dattore, B., Diamond, H., Lazzara, M.A., Le Blancq, F.,

Luterbacher, J., Mächel, H., Revadekar, J., Vose, R.S., Yin, X., 2014. The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. Geosci. Data J. 1, 75–102. https://doi.org/10.1002/gdj3.8

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The Global Land Data Assimilation System. Bull. Amer. Meteor. Soc. 85, 381–394. https://doi.org/10.1175/BAMS-85-3-381

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., Mosher, S., 2013. Berkeley Earth Temperature Averaging Process. Geoinformatics & Geostatistics: An Overview 2013. https://doi.org/10.4172/2327-4581.1000103

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, Jiande, Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, Jun, Hou, Y.-T., Chuang, H., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010. The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc. 91, 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E., 2013. The NCEP Climate Forecast System Version 2. J. Climate 27, 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Santer, B.D., Wigley T. M. L., Boyle J. S., Gaffen D. J., Hnilo J. J., Nychka D., Parker D. E., Taylor K. E., 2000. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. Journal of Geophysical Research: Atmospheres 105, 7337–7356. https://doi.org/10.1029/1999JD901105

Scarino, B.R., Minnis, P., Chee, T., Bedka, K.M., Yost, C.R., Palikonda, R., 2017. Global clear-sky surface skin temperature from multiple satellites using a single-channel algorithm with angular anisotropy corrections. Atmos. Meas. Tech. 10, 351–371. https://doi.org/10.5194/amt-10-351-2017

Schaaf, C., Wang, Z., 2015. MCD43C1 MODIS/Terra+Aqua BRDF/AlbedoModel Parameters Daily L3 Global 0.05Deg CMG V006. https://doi.org/10.5067/MODIS/MCD43C1.006

Seemann, S.W., Li, J., Menzel, W.P., Gumley, L.E., 2003. Operational Retrieval of Atmospheric Temperature, Moisture, and Ozone from MODIS Infrared Radiances. J. Appl. Meteor. 42, 1072–1091. https://doi.org/10.1175/1520-0450(2003)042<1072:OROATM>2.0.CO;2

Shen, S., Leptoukh, G.G., 2011. Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environ. Res. Lett. 6, 045206. https://doi.org/10.1088/1748-9326/6/4/045206

Shen, S.S.P., Yao, R., Ngo, J., Basist, A.M., Thomas, N., Yao, T., 2015. Characteristics of the Tibetan Plateau snow cover variations based on daily data during 1997–2011. Theor Appl Climatol 120, 445–453. https://doi.org/10.1007/s00704-014-1185-0

Shi, C., Xie, Z., Qian, H., Liang, M., Yang, X., 2011. China land soil moisture EnKF data assimilation based on satellite remote sensing data. Sci. China Earth Sci. 54, 1430–1440. https://doi.org/10.1007/s11430-010-4160-3

Shi, L., Bates, J.J., 2011. Three decades of intersatellite-calibrated High-Resolution Infrared Radiation Sounder upper tropospheric water vapor. J. Geophys. Res. 116, D04108. https://doi.org/10.1029/2010JD014847

Shi, L., Matthews, J.L., Ho, S., Yang, Q., Bates, J.J., 2016. Algorithm Development of Temperature and Humidity Profile Retrievals for Long-Term HIRS Observations. Remote Sensing 8, 280. https://doi.org/10.3390/rs8040280

Smith, T.M., Reynolds, R.W., 2005. A Global Merged Land–Air–Sea Surface Temperature Reconstruction Based on Historical Observations (1880–1997). J. Climate 18, 2021–2036. https://doi.org/10.1175/JCLI3362.1

Smith, T.M., Reynolds, R.W., Peterson, T.C., Lawrimore, J., 2008. Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006). J. Climate 21, 2283–2296. https://doi.org/10.1175/2007JCLI2100.1

Sobrino, J.A., Jiménez-Muñoz, J.C., Mattar, C., Sòria, G., 2015. Evaluation of Terra/MODIS atmospheric profiles product (MOD07) over the Iberian Peninsula: a comparison with radiosonde stations. International Journal of Digital Earth 8, 771–783. https://doi.org/10.1080/17538947.2014.936973

Song, Z., Liang, S., Wang, D., Zhou, Y., Jia, A., 2018. Long-term record of top-of-atmosphere albedo over land generated from AVHRR data. Remote Sensing of Environment 211, 71–88. https://doi.org/10.1016/j.rse.2018.03.044

Squintu, A.A., van der Schrier, G., Brugnara, Y., Klein Tank, A., 2019. Homogenization of daily temperature series in the European Climate Assessment & Dataset. International Journal of Climatology 39, 1243–1261. https://doi.org/10.1002/joc.5874

Swann, A.L., Fung, I.Y., Levis, S., Bonan, G.B., Doney, S.C., 2010. Changes in Arctic vegetation amplify high-latitude warming through the greenhouse effect. PNAS 107, 1295–1300. https://doi.org/10.1073/pnas.0913846107

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106, 7183–7192.

Thorne, P.W., Allan, R.J., Ashcroft, L., Brohan, P., Dunn, R.J.H., Menne, M.J., Pearce, P.R., Picas, J., Willett, K.M., Benoy, M., Bronnimann, S., Canziani, P.O., Coll, J., Crouthamel, R., Compo, G.P., Cuppett, D., Curley, M., Duffy, C., Gillespie, I., Guijarro, J., Jourdain, S., Kent, E.C., Kubota, H., Legg, T.P., Li, Q., Matsumoto, J., Murphy, C., Rayner, N.A., Rennie, J.J., Rustemeier, E., Slivinski, L.C., Slonosky, V., Squintu, A., Tinz, B., Valente, M.A., Walsh, S., Wang, X.L., Westcott, N., Wood, K., Woodruff, S.D., Worley, S.J., 2017.

Towards an integrated set of surface meteorological observations for climate science and applications. Bull. Amer. Meteor. Soc. https://doi.org/10.1175/BAMS-D-16-0165.1

Thorne, P.W., Donat, M.G., Dunn, R.J.H., Williams, C.N., Alexander, L.V., Caesar, J., Durre, I., Harris, I., Hausfather, Z., Jones, P.D., Menne, M.J., Rohde, R., Vose, R.S., Davy, R., Klein-Tank, A.M.G., Lawrimore, J.H., Peterson, T.C., Rennie, J.J., 2016. Reassessing changes in diurnal temperature range: Intercomparison and evaluation of existing global data set estimates. J. Geophys. Res. Atmos. 121, 2015JD024584. https://doi.org/10.1002/2015JD024584

Trenberth, K.E., 2015. Has there been a hiatus? Science 349, 691–692. https://doi.org/10.1126/science.aac9225

Trenberth, K.E., Dai, A., van der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. Nature Clim. Change 4, 17–22. https://doi.org/10.1038/nclimate2067

Trenberth, K.E., Fasullo, J.T., 2013. An apparent hiatus in global warming? Earth's Future 1, 19–32. https://doi.org/10.1002/2013EF000165

Trenberth, K.E., Koike, T., Onogi, K., 2008. Progress and Prospects for Reanalysis for Weather and Climate. Eos Trans. AGU 89, 234–235. https://doi.org/10.1029/2008EO260002

van den Dool, H.M., Saha, S., Johansson, Å., 2000. Empirical Orthogonal Teleconnections. J. Climate 13, 1421–1435. https://doi.org/10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2

Vose, R.S., Arndt, D., Banzon, V.F., Easterling, D.R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J.H., Menne, M.J., Peterson, T.C., Reynolds, R.W., Smith, T.M., Williams, C.N., Wuertz, D.B., 2012. NOAA's Merged Land–Ocean Surface Temperature Analysis. Bull. Amer. Meteor. Soc. 93, 1677–1685. https://doi.org/10.1175/BAMS-D-11-00241.1

Vose, R.S., Wuertz, D., Peterson, T.C., Jones, P.D., 2005. An intercomparison of trends in surface air temperature analyses at the global, hemispheric, and grid-box scale. Geophys. Res. Lett. 32, L18718. https://doi.org/10.1029/2005GL023502

Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Transactions on Geoscience and Remote Sensing 34, 892–905. https://doi.org/10.1109/36.508406

Wan, Z., Hook, S., Hulley, G., 2015a. MOD11C1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006. https://doi.org/10.5067/MODIS/MOD11C1.006

Wan, Z., Hook, S., Hulley, G., 2015b. MYD11C1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006. https://doi.org/10.5067/MODIS/MYD11C1.006

Watts, J.D., Kimball, J.S., Bartsch, A., McDonald, K.C., 2014. Surface water inundation in the boreal-Arctic: potential impacts on regional methane emissions. Environ. Res. Lett. 9, 075001. https://doi.org/10.1088/1748-9326/9/7/075001

Westergaard-Nielsen, A., Karami, M., Hansen, B.U., Westermann, S., Elberling, B., 2018. Contrasting temperature trends across the ice-free part of Greenland. Scientific Reports 8, 1586. https://doi.org/10.1038/s41598-018-19992-w

WMO, 2016. The Global Observing System for Climate: Implementation Needs (Technical Report No. 200), GCOS. World Meteorological Organization.

Woods, C., Caballero, R., 2016. The Role of Moist Intrusions in Winter Arctic Warming and Sea Ice Decline. J. Climate 29, 4473–4485. https://doi.org/10.1175/JCLI-D-15-0773.1

Xie, S.-P., 2016. Oceanography: Leading the hiatus research surge. Nature Clim. Change 6, 345–346. https://doi.org/10.1038/nclimate2973

Xie, Y., Koch, S., McGinley, J., Albers, S., Bieringer, P.E., Wolfson, M., Chan, M., 2011. A Space–Time Multiscale Analysis System: A Sequential Variational Analysis Approach. Mon. Wea. Rev. 139, 1224–1240. https://doi.org/10.1175/2010MWR3338.1

Xu, Y., Knudby, A., Shen, Y., Liu, Y., 2018. Mapping Monthly Air Temperature in the Tibetan Plateau From MODIS Data Based on Machine Learning Methods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 345–354. https://doi.org/10.1109/JSTARS.2017.2787191

Yan, X.-H., Boyer, T., Trenberth, K., Karl, T.R., Xie, S.-P., Nieves, V., Tung, K.-K., Roemmich, D., 2016. The global warming hiatus: Slowdown or redistribution? Earth's Future 4, 2016EF000417. https://doi.org/10.1002/2016EF000417

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nature Clim. Change 3, 875–883. https://doi.org/10.1038/nclimate1908

Yang, K., He, J., Tang, W., Qin, J., Cheng, C.C.K., 2010. On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau. Agricultural and Forest Meteorology 150, 38–46. https://doi.org/10.1016/j.agrformet.2009.08.004

Yang, K., Wu, H., Qin, J., Lin, C., Tang, W., Chen, Y., 2014. Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: A review. Global and Planetary Change 112, 79–91. https://doi.org/10.1016/j.gloplacha.2013.12.001

Yang, W., John, V.O., Zhao, X., Lu, H., Knapp, K.R., 2016. Satellite Climate Data Records: Development, Applications, and Societal Benefits. Remote Sensing 8, 331. https://doi.org/10.3390/rs8040331

Yao, T., Thompson, L.G., Mosbrugger, V., Zhang, F., Ma, Y., Luo, T., Xu, B., Yang, X., Joswiak, D.R., Wang, W., Joswiak, M.E., Devkota, L.P., Tayal, S., Jilani, R., Fayziev, R., 2012. Third Pole Environment (TPE). Environmental Development 3, 52–64. https://doi.org/10.1016/j.envdev.2012.04.002

Yao, T., Xue, Y., Chen, D., Chen, Fahu, Thompson, L., Cui, P., Koike, T., Lau, W.K.-M., Lettenmaier, D., Mosbrugger, V., Zhang, R., Xu, B., Dozier, J., Gillespie, T., Gu, Y., Kang, S., Piao, S., Sugimoto, S., Ueno, K., Wang, L., Wang, W., Zhang, F., Sheng, Y., Guo, W., Ailikun, Yang, X., Ma, Y., Shen, S.S.P., Su, Z., Chen, Fei, Liang, S., Liu, Y., Singh, V.P., Yang, K., Yang, D., Zhao, X., Qian, Y., Zhang, Y., Li, Q., 2018. Recent Third Pole's rapid warming accompanies cryospheric melt and water cycle intensification and interactions between monsoon and environment: multi-disciplinary approach with observation, modeling and analysis. Bull. Amer. Meteor. Soc. https://doi.org/10.1175/BAMS-D-17-0057.1

Yu, Y., Privette, J.L., Pinheiro, A.C., 2008. Evaluation of Split-Window Land Surface Temperature Algorithms for Generating Climate Data Records. IEEE Transactions on Geoscience and Remote Sensing 46, 179–192. https://doi.org/10.1109/TGRS.2007.909097

Zhang, H., Zhang, F., Ye, M., Che, T., Zhang, G., 2016. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. J. Geophys. Res. Atmos. 121, 2016JD025154. https://doi.org/10.1002/2016JD025154

Zhang, X., He, J., Zhang, J., Polyakov, I., Gerdes, R., Inoue, J., Wu, P., 2013. Enhanced poleward moisture transport and amplified northern high-latitude wetting trend. Nature Climate Change 3, 47–51. https://doi.org/10.1038/nclimate1631

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

Zhang, X., Peng, L., Zheng, D., Tao, J., 2008. Cloudiness variations over the Qinghai-Tibet Plateau during 1971–2004. J. Geogr. Sci. 18, 142–154. https://doi.org/10.1007/s11442-008-0142-1

Zhou, W., Wang, T., Shi, J., Peng, B., Zhao, R., Yu, Y., 2018. Remotely Sensed Clear-Sky Surface Longwave Downward Radiation by Using Multivariate Adaptive Regression Splines Method, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 5571–5574. https://doi.org/10.1109/IGARSS.2018.8519297