Air Surface Temperature Estimation Using MODIS Land Surface Temperature Data in Northwest Vietnam

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr.rer.nat.)

of the Georg-August-Universität Göttingen

within the doctoral program of Geoscience/Geography

of the Georg-August University School of Science (GAUSS)

submitted by

Phan Thanh Noi

From Hatinh, Vietnam

Göttingen, 2018

To my family

Acknowledgements

This journey began in November 2013, when I officially started with the Cartography, GIS and Remote Sensing Department, Faculty of Geoscience and Geography, at the University of Göttingen. While on this journey, I have fulfilled many life goals. I started out as a young single man and developed into a husband and father, with a lovely wife and children.

My utmost appreciation and heartfelt thanks go to my supervisor, Professor Martin Kappas, for his excellent advice, guidance, patience, encouragement, and kindness. After only one year in Göttingen, he supported my decision to return to Vietnam to marry. Seven months later, in April 2015, I came back to Göttingen with my wife, and we had our first child in January 2016. After that, I was finally able to focus on my research! This was a very stressful time, however, because I had spent more than two years without any research results. Prof. Kappas encouraged and helped me to work through this difficult time. Professor, I cannot thank you enough for your encouragement and advice, and allowing me to fulfill many life goals. Without your support, this thesis would not have been completed!

I owe my sincere and earnest thanks to my dissertation committee member, Prof.Dr. Gerhard Gerold for his insightful comments, suggestions, and suggestions that helped improve the overall quality of this dissertation.

I would also like to express my gratitude to my other committee member, Prof. Dr. Christoph Dittrich, Prof. Dr. Heiko Faust, and Prof. Dr. Daniela Sauer.

My gratitude extends to all my colleagues of the Department of Cartography, GIS and Remote Sensing for supporting and assisting me during my time in Germany.

My heartfelt thanks go to Dr. Jan Degener. Your office was always open to me with friendly advice, insightful discussions, and thoughtful suggestions. I learned a lot from you. As I have told you, when I had you check even just a small assignment for me, I felt significantly more confident.

I am very grateful to Dr. Stefan Erasmi and Dr. Michael Schlund for your valuable comments and timely feedback. You have been available whenever I needed advice regarding remote sensing techniques. I am also thankful to Dr. Ammar Rafiei Emam, Dr. Daniel Wyss, and Dr. Michael Klinge for your useful comments, nice discussion, and proofreading. A special thanks also goes to MSc. Dzung, MSc. Hung, and MSc. Tung, for providing a warm working environment when working late at night during the harshest winter weather (sometimes even below minus ten degrees centigrade). I would also like to thank the Secretary of the

Preface

The present dissertation entitled "Air surface temperature estimation using MODIS land surface temperature data in northwest Vietnam" has been submitted in partial fulfilment of the requirements for the PhD degree at the University of Göttingen (Germany). The main supervisor was Prof. Dr. Martin Kappas and the second supervisor was Prof. Dr. Gerhard Gerold.

This dissertation is a compilation of seven chapters. Besides the Introduction and the Conclusions, the five remaining chapters are based on papers or manuscripts that have been published in, submitted to, or under reviewed in, ISI-indexed journals.

The study was conducted at the Department of Cartography, GIS and Remote Sensing, from November 2013 to November 2018.

Göttingen, 25.09.2018

Phan Thanh Noi

Short summary

There is increasing demand for air surface temperature (Ta) data that can capture information for a large area or for a region, since this kind of data is an important parameter for a wide range of applications. However, due to the sparse distribution of meteorological stations, especially in developing countries and remote regions (e.g. mountainous areas, the Arctic, or tropical rainforests), the spatial coverage information of Ta is often limited. Fortunately, remote sensing satellites can provide land surface temperature (LST) data, which is considered one of the most important and useful data sources for Ta retrieval over a region or large area. Among various remote sensors that can provide LST data (e.g. the Advanced Very High Resolution Radiometer (AVHRR) sensor onboard NOAA satellites, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor onboard Terra satellites, Landsat (TM, ETM, and TIRS sensors) of Landsat satellites), the most popular and most often used is the LST from the MODerate Resolution Imaging Spectroradiometer (MODIS). AVHRR, ASTER, and Landsat each have their own limitations, since AVHRR lacks metadata files, ASTER is only available upon request and payment, and Landsat has a coarse temporal resolution of 16 days. Meanwhile two MODIS instruments, the first launched on 18 December 1999 and the second on 4 May 2002 aboard the Terra and Aqua platforms, respectively, are uniquely designed to provide free LST data with a moderate spatial resolution of about one kilometer and a very high temporal resolution (i.e. up to four global observations per day including daytime and nighttime data). Over the last two decades, MODIS LST data has successfully been used for Ta estimation in many regions, such as Europe, the United States, Canada, Africa, and the Tibetan Plateau. However in Vietnam, a developing country with very sparse meteorological stations, MODIS LST has rarely been applied to retrieve Ta.

This research presents a comprehensive evaluation of the application of MODIS LST data for Ta estimation in northwestern Vietnam. From the increasing number of studies in the literature, several methods have been proposed, applied, and evaluated to retrieve Ta from MODIS LST data. However, to the best of our knowledge, there are no studies that present an overview of the application of MODIS LST data. We therefore conducted the first review of all methods that have been developed and applied over nearly the last two decades, as well as discussed the advantages and disadvantages of these methods.

It is known that LST changes rapidly in both space and time, and that different regions can exhibit specific variances, since each region has a unique terrain. A number of studies have

reported that land cover type and elevation are the most important variables that affect the relationship between LST and Ta, as well as the accuracy of Ta estimation using LST data. Therefore, we conducted a study to evaluate and investigate the variation in LST due to changes in elevation to create an overview about the LST data in northwest Vietnam. The results showed that the quantity of temperature change varied with increasing elevation from January to December in both the daytime and the nighttime. The LST increased from 3.8 °C to 6.1 °C and from 1.5 °C to 5.8 °C with a 1,000 m decrease in elevation at daytime and nighttime, respectively. In addition, land use/cover types also affected the variability of LST with changes in elevation. Therefore, in studies using MODIS LST data for Ta estimation, elevation, Julian day, and land cover types should be taken into consideration.

There are four types of MODIS LST data available each day, however, only a handful of studies have compared the performance of each individual MODIS LST between the two different sky conditions (i.e. all clear sky condition and only good LST data conditions) as well as the different combination of the four MODIS LST data for Ta estimation with the same estimation methods, in the same study areas. Therefore, we implemented the next study, which evaluated and tested each individual LST data as well as all possible combinations of the four MODIS LST data from two distinct land surface characteristics and two sky conditions in northwestern Vietnam for 10 years (from 2004 to 2013), for daily Ta estimation. The results showed that Terra LST has higher correlation with Ta than Aqua LST (in both sky conditions), meaning that having a closer overpass time with Ta occurrence time does not guarantee a higher correlation. Using only good LST data produced higher accuracy of Ta estimation, however, if the percentage of good data is low (i.e. less than 30%), using all clear sky data will produce higher Ta-max estimations. In addition, it should be noted that the trade-off between good LST data and the spatial coverage of LST data should be taken into account when selecting LST data for Ta estimation. In the next study, we used all four MODIS LST data and ten auxiliary variables to estimate Ta-max and Ta-min in northwestern Vietnam. We evaluated the performance of MODIS LST both exclusively and with auxiliary variables. The results showed that not all variables improved the accuracy of Ta estimation. Besides the four MODIS LST, elevation and longitude were considered the most important variables for Ta-max estimation. However, for Ta-min estimation, the relative performance of the simplest model (using one variable) and the most complicated model (using ten variables) was similar. At best, Ta-min/Ta-max estimation achieved results of $R^2 = 0.88/0.93$ and RMSE = 2.08/1.43 °C. It is clearly seen that with the most popular methods (linear regression model/statistical approaches) we can estimate Ta-max with very high accuracy by introducing auxiliary variables into the models, however, the accuracy of

Ta-min estimation cannot be improved using this approach. Moreover, in recent years, the application of machine learning methods to Ta estimation using MODIS LST has received great attention from scientists because they can handle the complicated relationship between LST and Ta under different conditions. Therefore, we conducted a study to compare the performance of LM, RF, and CB for Ta (Ta-max, Ta-min, and Ta-mean estimation) in northwestern Vietnam for five years (from 2009 to 2013). The results suggested that when all four MODIS LST were used with or without auxiliary variables, the performance of LM, CB, and RF were similar. This study confirmed that the very high accuracy of Ta estimation ($R^2 > 0.93/0.80/0.89$ and RMSE ~1.5/2.0/1.6 °C of Ta-max, Ta-min, and Ta-mean, respectively) could be achieved with a simple combination of the four LST, elevation, and Julian day data using a suitable algorithm. Obviously, the results of Ta-min were not as good as Ta-max estimation with any of the utilized approaches. For further research, other seldomly used variables such as nighttime light data, percentage of urban land cover, or distance to coasts should be considered and evaluated in order to improve the results of Ta estimation using MODIS LST.

List of manuscripts and publications

Chapter 2	**Noi, P.T**.; Kappas, M. Application of MODIS Land Surface Temperature Data: A Systematic Literature Review and Analysis (accepted)
	Noi, P.T.; Kappas, M.; Erasmi, S. Application of MODIS Land Surface Temperature Data for Air Surface Temperature Estimation: A Review (submitted).
Chapter 3	**Noi, P.T**.; Kappas, M.; Tran, T.P. Land Surface Temperature Variation Due to Changes in Elevation in Northwest Vietnam. Climate **2018**, *6*, 28. doi.org/10.3390/cli6020028
Chapter 4	**Noi, P.T**; Kappas, M. et al. Evaluation of MODIS LST Product Collection 6 for Daily Air Surface Temperature Estimation in Northwest Vietnam (accepted).
Chapter 5	**Noi, P.T**.; Kappas, M.; Degener, J. Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam. Remote Sens. **2016**, *8*, 1002. doi.org/10.3390/rs8121002
Chapter 6	**Noi, P.T**.; Degener, J.; Kappas, M. 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 Sens. **2017**, *9*, 398. doi.org/10.3390/rs9050398

Editorial notes:

In this dissertation, to obtain the common style throughout the text, all publications and manuscripts were edited and reformatted. The modifications and editing are as follows:

(i). Fonts, layout of the text and tables.

(ii). Numbering of chapters, figures, and tables.

(iii). A common citations and references style were used throughout the dissertation, which may differ from the journals' style.

Table of Contents

List of Figures

List of Tables

Abbreviations & symbols

Acronyms	Definition
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
DEM	Digital Elevation Model
Dlth	Day length in hours
Ele	Elevation (m)
ENVISAT	ENVIronment SATellite
ETM+	Enhanced Thematic Mapper+
GIS	Geographical Information Systems
HDF	Hierarchical Data Format
Jd	Julian day
K	Kelvin
Lat	Latitude
Long	Longitude
LST	Land surface temperature
LSTad	Aqua daytime LST
LSTan	Aqua nighttime LST
LSTtd	Terrra daytime LST
LSTtn	Terra nighttime LST
MODD	Terrra daytime LST
MODIS	Moderate Resolution Imaging Spectroradiometer
MODN	Terra nighttime LST
MYDD	Aqua daytime LST
MYDN	Aqua nighttime LST
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
SEVIRI	Spinning Enhanced Visible and Infrared Imager
Ta	Air surface temperature
Ta-max	Daily maximum Ta
Ta-mean	Daily mean Ta
Ta-min	Daily minimum Ta
TES	Temperature Emissivity Separation
TIR	Thermal InfraRed
TM	Thmatic Mapper
TOA	Top of Atmosphere
UHI	Urban Heat Island
USGS	United States Geological Survey
VIIRS	Visible and Infrared Imagery Radiometer Suite
VZA	View zenith angle

Chapter 1

1.1. General Introduction

Air surface temperature (Ta), also called air temperature, land air surface temperature, or near surface temperature, is usually measured at meteorological stations (~ 1.5 – 2 m above the land surface) and is a key measurement in describing terrestrial environmental conditions (Rosenberg et al., 1983; Prihodko and Goward, 1997; Zhang et al., 2016). It is one of the most frequently recorded measurements (Stisen et al., 2007; Zhu, W. et al., 2017) because most terrestrial life inhabits the zone between the ground surface and the near surface (Prihodko and Goward, 1997). In addition, most of the discussion about the climate change (e.g. global warming) is based on Ta (Hansen et al., 1981; Prihodko and Goward, 1997; Lofgren et al., 2011). However, since Ta is strongly affected by land surface characteristics (which vary both spatially and temporally), ground-based station observations cannot represent the Ta of an entire region (Benali et al., 2012; Zhu, W. et al., 2017; Phan et al., 2018). Nevertheless certain application fields such as agriculture, hydrology, ecology, and climate change/atmospheric science, require spatiotemporal continuous Ta information, so demand is high (Rhee and Im, 2014; Meyer at al., 2016; Zhu, S. et al., 2017; Noi et al., 2017; Oyler et al., 2015; Yang et al., 2017).

To obtain Ta information for an entire region, one commonly used method is spatial interpolation of meteorological station records (Ishida and Kawashima, 1993; Jarvis and Stuart, 2001; Stahl et al., 2006; Hofstra et al., 2008; Appelhans et al., 2015). However, the performance of this interpolation method is highly dependent on the weather station network, i.e. spatial distribution and density (Vogt et al., 1997; Chan and Paelinchx, 2008). According to Bolstad et al. (1998), if sufficient sample points are given, interpolation methods can accurately estimate annual, monthly, or daily mean Ta. Additionally, if the sample data are abundant, most interpolation methods give similar and satisfying results (Burrough and McDonnell, 1998). In less developed, remote, or mountainous areas, however, the distribution of meteorological stations is low, making interpolation unsuitable (Prihodko and Goward, 1997, Noi et al., 2016; Shi et al., 2017). Therefore alternate ways to retrieve spatiotemporal Ta information need to be researched and developed (Sun et al., 2005; Zhu, S. et al., 2017). Remote sensing techniques offer the potential to provide detailed Ta information over large areas (Goetz et al., 2000; Carlson, 2007; Zeng et al., 2015).

Land surface temperature (LST), is alike to a "skin" temperature of the ground surface (Vancutsem et al., 2010; Kloog et al., 2014), and can be retrieved directly from remotely sensed radiance data by various algorithms such as the generalized split-window algorithm (Wan and Dozier, 1996), the physics-based day/night operational method (Wan and Li, 1997), and the two-step physical retrieval method (Li et al., 2007). However, according to Czajkowski et al. (2000), Ta is difficult to directly retrieve from the visible and infrared bands of a remotely sensed image. Therefore, two steps must be implemented: (1) retrieval of LST data from the thermal infrared signals of the remote sensing data, and (2) estimation of Ta from LST data (Janatian et al., 2017). Since land and air have different heat capacities and are affected by weather conditions differently, Ta and LST are significantly different (Jin and Dickinson, 2010; Shen and Leptoukh, 2011). The correlation between Ta and LST, due to heat exchange (energy exchange) between the land surface and air surface, allows for LST to be considered as a proxy for Ta. In fact, many studies have estimated Ta with very high accuracy using LST data (i.e. RMSE ~ 1 – 2 ℃) (Noi et al., 2016, 2017).

These are the popular sensors for LST: the Advanced Very High Resolution Radiometer (AVHRR) sensor onboard NOAA satellites, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor onboard the Terra satellite, the Landsat sensors (TM, ETM, and TIRS) of the Landsat satellites, and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard the Terra and Aqua satellites. AVHRR, ASTER, and Landsat each have their own limitations, for instance: AVHRR lacks nighttime LST data, ASTER is only available upon request and purchase, and Landsat has a coarse temporal resolution of 16 days (Tomlinson et al., 2011). Therefore the most suitable way to estimate Ta is by processing MODIS LST data, which has global availability at moderate spatial resolutions of about one km, very high spatial resolution, and two overpass times (daytime and nighttime).

Many approaches have been developed and applied to estimate Ta using MODIS LST data, and the three popular groups of methods are: energy-balance modeling, temperature-vegetation index (TVX), and statistical methods (Noi et al., 2016). Among these, statistical methods are the most popular, because compared to energy-balance methods, they have simpler operability. When compared to the TVX method, statistical methods also have a wider applicability to different land surface characteristics (Chen et al., 2016). The statistical methods comprise three distinct methods: simple statistic (using one variable), advance statistic (using more than two variables, usually MODIS LST with auxiliary variables), and machine learning.

Vietnam has 170 surface meteorological stations, including 97 synoptic and 26 international exchange stations (Dinh, 2005), which is inadequate for a country with an area of 331,688 km^2 comprising 40% mountains, 40% hills, and 20% lowlands. Since interpolation techniques are not suitable for Vietnam, the best alternative method is to use satellite LST data. However, according to the Scopus database (accessed on March 01, 2018), there are only two studies using MODIS LST data for Ta estimation (Noi et al., 2016, 2017). It is therefore practical to evaluate the performance of MODIS LST data for Ta estimation.

For a preliminary application, we focused on an area in northwestern Vietnam where the topography is quite complex, with elevations ranging from sea level to about 3000 m.

1.2. Objectives

The main objectives of this research can be summarized as follows:

(1) To investigate the spatiotemporal variation of LST from MODIS data according to changes in elevation in northern Vietnam.

(2) To analyze the relationship between Ta and MODIS LST (i.e. LST_{ad}, LST_{an}, LST_{td}, LST_{tn}), and compare the performance of different MODIS LST data from the two satellites (Aqua and Terra).

(3) To estimate maximum and minimum Ta (Ta-max and Ta-min) using all four MODIS LST products (Aqua and Terra, daytime and nighttime) and auxiliary data, to solve the discontinuity problem of ground measurements.

(4) To compare three different algorithms: multiple linear regression (LM), Cubist regression (CB), and random forests models (RF), for Ta estimation using MODIS LST.

1.3. Outline of the dissertation

This dissertation is a compilation of eight chapters. Apart from the Introduction and Conclusion, the six remaining chapters are based on papers or manuscripts that have been published in, submitted to, or are under review for, ISI-indexed journals.

The eight chapters are described as follows:

Chapter 1 introduces the dissertation, background context, motivations, and expression of the research objectives and approaches.

Chapter 2 presents a comprehensive review of Ta estimation using MODIS LST data since it became available in the early 2000s.

Chapter 3 presents the spatiotemporal variability of MODIS LST in northern Vietnam.

Chapter 4 evaluates MODIS LST data for Ta estimation in northwestern Vietnam.

Chapter 5 estimates Ta-max and Ta-min in northern Vietnam using MODIS LST.

Chapter 6 compares different methods for Ta estimation using MODIS LST data.

Chapter 7 summarizes the results obtained from Chapters 2 – 7 and answers the research questions in Chapter 1. Recommendations and future research directions are also provided.

References

Appelhans, T., Mwangomo, E., Hardy, D.R., Hemp, A., Nauss, T. (2015). Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spatial Statistics, 14A, 91–113.

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124, 108-121

Bolstad, P.V., Swift, L., Collins, F, Regniere, J. (1998). Measured and predicted air temperatures at basin to regional scales in the southern Appalachian Mountains. Agricultural and Forest Meteorology, 91, 161 – 176.

Burrough, P.A., McDonnell, R.A. (1998). Principles of geographical information systems. Oxford University Press, New York.

Chan, J. C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery. Remote Sensing of Environment, 112, 2999–3011.

Chen, Y., Quan, J., Zhan, W., & Guo, Z. (2016). Enhanced Statistical Estimation of Air Temperature Incorporating Nighttime Light Data. Remote Sensing, 8, 656

Czajkowski, K.P., Goward, S.N., Stadler, S.J., & Walz, A. (2000). Thermal remote sensing of near surface environmental variables: Application over the Oklahoma Mesonet. The Professional Geographer, 52, 345–357.

Hansen, J., Johnson, D., Laeis, A., Lebedeff, S., Lee, P., Rind, D., & Russel, G. (1981). Climatic impact of increasing atmospheric carbon dioxide. Science, 213, 957-966.

Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. Journal of Geographical Research: Atmospheres, 113, D21110.

Ishida, T., Kawashima, S. (1993). Use of cokriging to estimate surface air temperature from elevation. Theoretical and Applied Climatology, 47:147–157.

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. International Journal of Climatology, 37, 1181-1194

Jarvis, C.H., Stuart, N. A (2001). Comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method. Journal of Applied Meteorology, 40, 1075–1084.

Jin, M., & Dickinson, R. E. (2010). Land surface skin temperature climatology: benefitting from the strengths of satellite observations. Environmental Research Letters, 5, 044004.

Kloog, I., Nordio, F., Coull, B.A., & Schwartz, J. (2014). Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. Remote Sensing of Environment, 150, 132-139

Li, J., Li, J., Weisz, E., & Zhou, D.K. (2007). Physical retrieval of surface emissivity spectrum from hyperspectral infrared radiances. Geophysical Research Letters, 34

Lofgren, B.M., Hunter, T.S., Wilbarger, J. (2011). Effects of using air temperature as a proxy for evapotranspiration in climate change scenarios of Great Lakes basin hydrology. Journal of Great Lakes Research, 37, 744–752.

Meyer, H., Katurji, M., Appelhans, T., Müller, M., Nauss, T., Roudier, P., & Zawar-Reza, P. (2016). Mapping Daily Air Temperature for Antarctica Based on MODIS LST. Remote Sensing, 8, 732

Noi, P., 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

Noi, P., Kappas, M., & Degener, J. (2016). Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam. Remote Sensing, 8, 1002

Phan, T.N., Kappas, M., & Tran, T.P. (2018). Land Surface Temperature Variation Due to Changes in Elevation in Northwest Vietnam. Climate, 6, 28.

Oyler, J. W., Ballantyne, A., Jencso, K., Sweet, M., & Running, S.W. (2015). Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. International Journal of Climatology, 35, 2258–2279.

Prihodko, L., & Goward, S.N. (1997). Estimation of air temperature from remotely sensed surface observations. Remote Sensing of Environment, 60, 335–346.

Rhee, J., & Im, J. (2014). Estimating High Spatial Resolution Air Temperature for Regions with Limited in situ Data Using MODIS Products. Remote Sensing, 6, 7360-7378.

Rosenberg, N. J., Blad, B. L., & Venna, S. B. (1983), Microclinulte: The Biological Environment. Wiley, New York.

Shen, S., & Leptoukh, G.G. (2011). Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environmental Research Letters, 6, 045206

Stahl, K., Moore, R., Floyer, J., Asplin, M., & McKendry, I. (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and Forest Meteorology, 139, 224–236.

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., & Eklundh, L. (2007). Estimation of diurnal air temperature

Tomlinson, C.J., Chapman, L., Thornes, J.E., & Baker, C. (2011). Remote sensing land surface temperature for meteorology and climatology: a review. Meteorological Applications, 18, 296-306

Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S.J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114, 449-465

Vogt, J., Viau, A. A., & Paquet, F. (1997). Mapping Regional Air Temperature Fields Using Satellite Derived Surface Skin Temperatures. International Journal of Climatology, 17, 1559–1579.

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(4), 892-905.

Wan, Z., & Li, Z.L. (1997). A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. IEEE Transactions on Geoscience and Remote Sensing, 35(4), 980-996.

Yang, Y., Cai, W., & Yang, J. (2017). Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China. Remote Sensing, 9, 410

Zhu, S., Zhou, C., Zhang, G., Zhang, H., Hua, J. (2017). Preliminary verification of instantaneous air temperature estimation for clear sky conditions based on SEBAL. Meteorology and Atmospheric Physics, 129, 71-81.

Zhu, W., Lű, A., Jia, S., Yan, J., & Mahmood, R. (2017). Retrievals of all-weather daytime air temperature from MODIS products. Remote Sensing of Environment, 189, 152-163

Chapter 2

This chapter is based on two manuscirpts:

Thanh Noi Phan,[a,b,] * and Martin Kappas,[a]

Thanh Noi Phan [a,b,], Martin Kappas [a], and Stefan Erasmi [a]

[a] Department of Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen, Göttingen 37077, Germany

[b] Department of Cartography and Geodesy, Faculty of Land Management, Vietnam National University of Agriculture, Hanoi, Vietnam.

Correspondence: thanh-noi.phan@geo.uni-goettingen.de; Tel.: +49 551 399805, Department of Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen, Goldschmidt Street 5, Göttingen 37077, Germany.

Abstract

There is an increasing demand of air surface temperature (Ta) data which can capture information for a large area or for a region, because this kind of data is an important parameter for a wide range of applications, such as environment and climate science, hydrology, agriculture, and weather forecasting. However, due to the sparse distribution of meteorological stations, particularly in developing countries and remote regions (e.g. mountainous areas, the Arctic, or tropical rainforests), the spatial coverage of Ta is often limited. Remotely sensed MODIS land surface temperature (LST) data, which is freely available with global coverage and has very high temporal resolutions (twice daily observations from two satellites, Aqua and Terra), is considered one of the most suitable and important data sources for Ta estimation. Since MODIS data became available in the early 2000s, there has been a rapid increase in applications of MODIS LST data for Ta estimation. To date, several methods have been proposed, applied, and evaluated to retrieve Ta from MODIS LST data. However, to the best of our knowledge, there are no studies that provide an overall review of the broad field of algorithms and applications for the estimation of Ta from MODIS LST data. In this context, this paper provides a review of methods that have been developed and applied over the last 15 years. Furthermore, we also discuss the advantages, limitations, potential, and future direction of Ta estimation using MODIS LST.

Keywords: MODIS, LST, air surface temperature, land surface temperature, estimation, statistical methods

2.1. Introduction

Air surface temperature (Ta, also called "near-surface air temperature" or "air temperature"), usually measured at ~ 2.0 m above the ground surface, is a fundamental descriptor of the earth surface conditions (Prihodko and Goward, 1997; Peon et al., 2014). In land-atmosphere systems, Ta is considered as a fundamental parameter (Zhang et al., 2015) that provides a key input to physical models in a wide range of applications, such as environment and climate science (Richard and Gratton, 2001; Lofgren, Hunter, and Wilbarger, 2011), hydrology (Vallet-Coulomb et al., 2001; Anamarija et al., 2015), agriculture (de Wit and van Diepen, 2008; Balaghi et al., 2008; Panayi, Peters, and Kyriakides, 2017), and weather forecasting (Smith et al., 1988; Christiansen, 2005). According to Rosenzweig et al. (2014) and Oyler et al. (2015), there has been a steadily rising demand for Ta data, particularly at large scales.

Traditionally, Ta has been measured at standard meteorological stations with high accuracy and temporal resolutions (Hubbard and Hollinger, 2005). However, Ta is affected by land surface characteristics (e.g. land use/cover, topography, vegetation, soil moisture), which vary in both space and time. Therefore, the observations of Ta from these stations do not represent the large areas or regions (Stisen et al., 2007) that is required in many earth system models (Stisen et al., 2007; Nieto et al., 2011). In the literature, three methods have been widely used for retrieving continuous surfaces of Ta: (1) geostatistical methods (using interpolation techniques), (2) the vertical lapse method (where the change of temperature with respect to altitude, called the lapse rate, is assumed to be constant in the vertical direction), and (3) remotely sensed land surface temperature (LST) based methods. The first two methods have been widely applied (Li and Heap, 2008). However, they contain some drawbacks, such as that geostatistical methods cannot perform well in study areas with sparse weather stations or with a lack of observation data (Prihodko and Goward, 1997; Shi et al., 2017). It should be mentioned that a number of products available today (e.g. WorldClim, CHELSA http://chelsa-climate.org) model Ta from station data and other available estimates using co-variables like topography (to achieve spatial downscaling to comparable resolutions of 1 km). The main limitation of those data is not the quality/spatial resolution but the temporal resolution. In the vertical lapse method, different weather station locations have different surface conditions; therefore, it is difficult to determine the lapse rate, which will result in a large error in Ta estimation (Stahl et al., 2006).

Currently, the remote sensing based method is considered as the most suitable way to retrieve Ta from LST data at local, regional, or global coverage (Li et al., 2013; Chen et al., 2016).

LST is defined as the "skin" temperature of the ground (Vancutsem et al., 2010; Kloog et al., 2014), and it can be retrieved directly from remotely sensed radiance data in thermal infrared spectral region by various algorithms. It is worth noting that many LST retrieval algorithms have been expanded from the split-window (SW) algorithm proposed by McMillin in 1975, which was originally used for ocean temperature retrieval, and does not require the atmospheric profiles at the time of acquisition. By adding the zenith angle and surface emissivity observed by the sensor to the radiation transfer equation, Coll et al. (1994) has made the SW algorithm become a novel retrieval algorithm. Wan and Dozier (1996) improved the SW algorithm by taking into account the observation angle (view angle) of the sensor, calling it the generalized SW algorithm. Xia et al. (2014) reviewed the historical development of LST retrieval algorithms and concluded that the SW algorithms using thermal infrared data are very mature. It has been developed, expanded, and applied to various kinds of remote sensing data, e.g. MODIS (Wan, 2014), AVHRR (Pinheiro et al., 2006), FY-2C (Tang et al., 2008), GOES-8 (Sun and Pinker, 2003) and Landsat 8 (Rosenzweig et al., 2014).

In remote sensing techniques, radiometers do not measure Ta directly, but rather the estimated LST data are obtained by translating thermal infrared (TIR) observations into skin temperature of the observable land surface using a split-window algorithm (Wan & Dozier, 1996). LST and Ta are different due to the energy exchange between the land surface and the near surface atmosphere, but there is a correlation between LST and Ta. This correlation is strongly affected by land surface characteristics. Over vegetated areas, for example, emitted radiance is relatively low. Therefore, this relationship is improved. Whereas over the surfaces with high emittance (e.g. bare land), the disparity between LST and Ta become greater (Prihodko & Goward, 1997). As a result, to retrieve Ta data, a supplementary estimation technique is necessary. Many methods have been developed and applied for Ta estimation using LST, such as the energy-balance methods (Meteotest, 2003; Pape and Löffler, 2004; Sun et al., 2005), the temperature – vegetation index (TVX; Nemani and Running, 1989; Goward et al., 1994; Prihodko and Goward, 1997; Shah et al., 2013; and Sun et al., 2014) and statistical methods - including basic statistics, advanced statistics, and machine learning methods (Chen et al., 2015; Kloog et al., 2017; Zhang et al., 2016b; Noi et al., 2017).

Four sensors are the most popular for the acquisition of LST data in the remote sensing community: The Advanced Very High Resolution Radiometer (AVHRR) sensor onboard NOAA satellites, the Advanced Spaceborne Thermal Emission and the Reflection Radiometer (ASTER) sensor onboard Terra satellites, Landsat (i.e. TM, ETM, and TIRS)

sensors of the Landsat satellites, and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard Terra and Aqua satellites. Among these sensors, AVHRR holds the longest operating record (operational since 1979), very high temporal resolution (twice daily), and high spatial resolution (approximately 1.1 km). It was one of the main LST data sources for Ta estimation (Prihodko and Goward, 1997; Lakshmi et al., 2001; Jang et al., 2004; Florio et al., 2004; Riddering and Queen, 2006). However, according to Frey, Kuenzer and Dech (2012), in the AVHRR products, there is a lack of metadata layers (time of acquisition, satellite view zenith and azimuth angle, quality flags), which lead to serious difficulties in the use and interpretation of the data. In comparison, the quality of MODIS LST is expected to be higher because it uses an updated algorithm. Landsat provides a very high spatial resolution (Landsat TM/ETM+/TIRS with thermal infrared bands are acquired at 120/60/100 m resolution, but products are resampled to 30 m); however, the temporal resolution of 16 days limits its use in time series applications. In addition, Landsat lacks nighttime data; therefore, it is not widely applied for Ta estimation. The ASTER sensor can provide very high spatial resolution (90 m) and a 16 day revisit cycle, however, the data can only be acquired upon request and must be purchased (Tomlinson et al., 2011). Consequently, the MODIS sensors, with its advantages, i.e. freely available global coverage, moderate spatial resolution (1 km), and very high temporal resolutions with both daytime and nighttime LST data (four-times daily, 8 day average, or monthly composites), have been widely applied by the remote sensing community. MODIS is regarded as the most popular data source for spatially explicit, global Ta estimations (Metz et al., 2017).

In the literature, based on Scopus and Web of Science (accessed on May 30, 2018), we have found only one review article of MODIS LST data applications (Tomlinson et al., 2011). Therefore, it is reasonable and practical to review MODIS LST data applications for Ta estimation. In this review, we addressed the major trends as well as developments of Ta estimation using MODIS LST data looking at: (1) the development of methodologies, (2) the integration of auxiliary variables with LST for accurate Ta estimation, (3) the effects of different combination of MODIS LST, and (4) discussion of the remaining issues as well as the future directions of Ta estimation using MODIS LST data.

2.2. Overview of MODIS LST Data

Two MODIS (Moderate Resolution Imaging Spectroradiometer) instruments, Terra and Aqua, are the main sensors onboard the NASA EOS (Earth Observation System) series of satellites, which started collecting data in early 2000 and mid-2002, respectively. The unique

design of these two satellites incorporates: wide spectral range (36 spectral bands, ranging from visible to thermal infrared wavelength), high spatial resolution, and near-daily global coverage, i.e. in the morning Terra passes from north to south, whereas in the afternoon, Aqua passes from south to north over the equator. This twin-MODIS design also helps optimize cloud-free imaging and the optical effects of shadows and glare that might result from sunlight in the morning and afternoon. The data collected by these two satellites will improve understanding of the changing world, including processes occurring on the land, in the oceans, and in the atmosphere.

Currently, both Terra and Aqua satellites provide nine MODIS LST products at different spatial and temporal resolutions (Table 1), including daily (MOD/MYD11A1, MOD/MYD11B1, MOD/MYD11C1, MOD/MYD11_L2), eight-day (MOD/MYD11A2, MOD/MYD11B2, MOD/MYD11C2), and monthly (MOD/MYD11B3, MOD/MYD11C3) composite data. Each of these products is available in two different versions (so-called "collection"), version 5 (V005) and version 6 (V006), with the exception of M*D11B2 and M*D11B3, which are only available with V006. However, The Land Processes Distributed Active Archive Center (LP DAAC) has recently announced (on 09 February 2018) that all V005 of these products will be decommissioned starting 09 April 2018.

Table 1: Summary of the V006 MODIS LST data products

Earth Science Data Type	Product Level	Nominal Data Array	Spatial Resolution	Temporal Resolutions	Map Projection
MOD/MYD11_L2	L2	2030 or 2040 lines by 1354 pixels per line	1 km at nadir	Swath (scene)	None. (Latitude, Long referenced)
MOD/MYD11A1	L3	1200 rows by 1200 columns	1 km (actual 0.928 km)	daily	Sinusoidal
MOD/MYD11B1 MOD/MYD11B2 MOD/MYD11B3	L3	200 rows by 200 columns	6 km (actual 5.568 km)	daily eight days monthly	Sinusoidal
MOD/MYD11A2	L3	1200 rows by 1200 columns	1 km (actual 0.928 km)	eight days	Sinusoidal
MOD/MYD11C1 MOD/MYD11C2 MOD/MYD11C3	L3	360° by 180° (global)	0.05° by 0.05°	daily eight days monthly	Equal-angle geographic

Among these nine products of MODIS LST, the daily data (MOD/MYD11A1) and the eight-day data (MOD/MYD11A2) are the most popularly used. MOD/MYD11A2 data are retrieved by averaging the MOD11A1 product over two to eight days, depending on the number of cloud free MOD/MYD11A1 products available (in the eight days).

It is worth noting that most studies in the literature use MODIS LST V005 for Ta estimation because version V006 have only been available for download since August 2015 (https://lpdaac.usgs.gov).

Regarding the different versions (V005, V006) of M*D11 LST, Wan (2014) stated that V006 is much better than previous versions (V4.1, V005) due to many fixes and improvements have been made. The most significant improvement is the removal of cloud-contaminated LST data. Furthermore, the coefficients for the split-window algorithms and the classification based surface emissivity values were updated (Metz et al., 2017).

2.3. Overview of MODIS LST Data Applications

2.3.1. Selecting publication using MODIS LST data

To produce and overview of MODIS LST data application, we used the Scopus database to search for publications. It should be noted that by using the same search keywords (Fig. 2), there are 1101 publications within the timespan 1999 – 2018, however, most of the studies (991 publications or ~90%) were published in the period from 2009 to 2018. Therefore, we analyzed publications within this timespan (2009 – 2018).
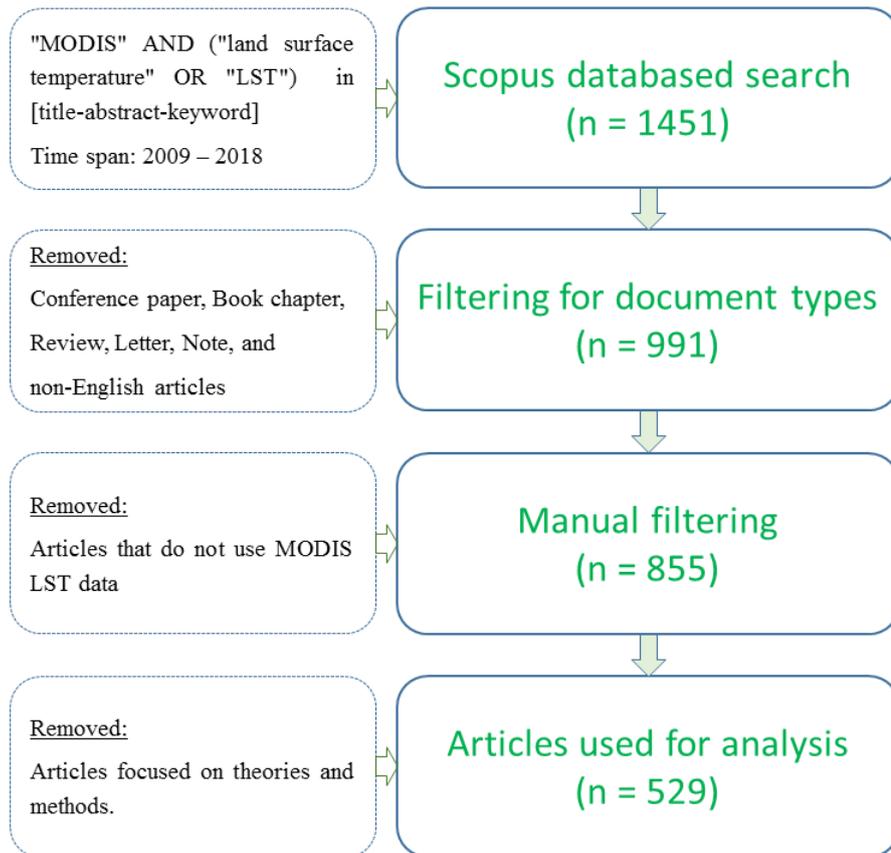
Figure 2.1. The procedure for selecting articles from Scopus database

First, we performed an advanced search in the Scopus database on 12 July 2018 using the keywords {"MODIS" AND ("land surface temperature" OR "LST")} in titles, abstracts, and keywords (including author keywords and index keywords) from 2009 to 2018. This search resulted in 1451 records. Next, we excluded conference papers, book chapters, review articles, letters, notes, and non-English articles, which resulted in 991 peer-reviewed articles (including published articles and articles in press) published in 159 journals. We further filtered this list by removing 136 articles that did not use MODIS LST data. Next, based on the titles, abstracts, and full texts (if necessary), we removed articles that focused more on theories and methods, or only analyzed the spatial and temporal variations of LST. As a result, 529 articles were selected for the final list (Figure 2.1).

2.3.2. General Research Output Based on Selected Publications

In this section, we present an overview of research application trends using MODIS LST data based on the identified 529 articles published in 159 journals from January 2009 to July 2018.
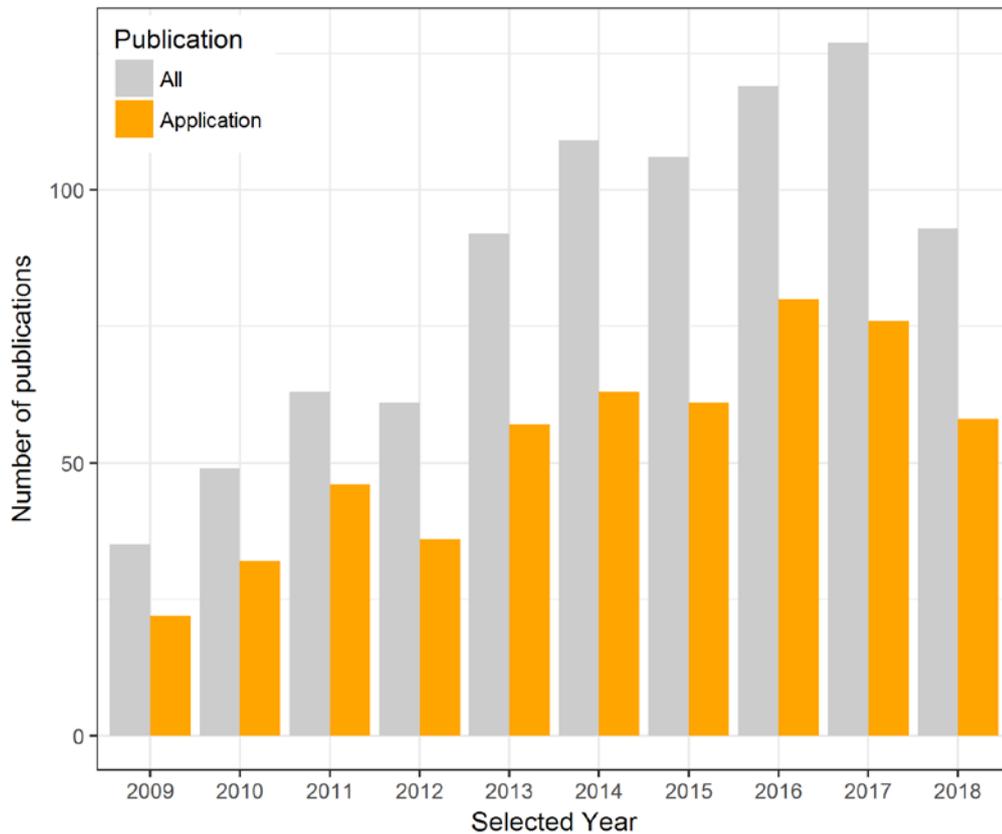
Figure 2.2. The total publications (grey) and applied for a specific field (orange) in Scopus database from 2009 to 2018.

Generally, the number of publications that used MODIS LST data increased from 2009 to 2017. It should be noted that the publications focusing on theories, methodologies, or validation purposes have been removed from this review study. In 2018, by the date of the Scopus search (12 July), there were 55 articles already published (or in press). It is reasonable to predict that this increasing trend will continue through 2018.

Regarding the subject classification of the publications, 529 articles were published within 19 subject areas in 159 journals. The top three subject areas were Earth and Planetary Sciences (397 articles – 75%), Environmental Science (172 articles – 32.5%), and Agricultural and Biological Sciences (164 articles – 31%). The total number of articles was not equal to 529 because many belong to multiple subject areas. In addition, the studies of MODIS LST data were published in both general research area journals (e.g. J-STARS, Hydrology, and Earth System Sciences) and specific journals (e.g. Remote Sensing of Environment and International Journal of Climatology), indicating a wide ranges applications of MODIS LST data.

Table 2 shows the 11 most active journals that have published MODIS LST data-based research. Among the 529 articles, 255 articles (48.2%) were published in these 11 journals,

while the remaining 274 (51.8%) were published in the other 148 journals. In addition, 65.8% of total citations (TC) were received for these 11 journals, whereas only 34.2% were received for the other 148 journals. The three major publication outlets for MODIS LST data-based research were Remote Sensing of Environment, Remote Sensing, and International Journal of Remote Sensing, with a total number of articles (% articles) of 62 (11.7%), 56 (10.6%), and 30 (5.7%), respectively. Table 2.1 shows that the journals with the highest TC values were Remote Sensing of Environment with 2808 citations (33.5%), Remote Sensing with 397 citations (4.7%), and the International Journal Of Applied Earth Observation And Geoinformation with 353 citations (4.2%).

Table 2.1. The top most active journals publishing research based on MODIS LST data.

Ranking	Journal	Total Articles (%)	Total Citations (%)	Citations per Articles
1	Remote Sensing Of Environment	62 (11.7)	2808 (33.5)	45.3
2	Remote Sensing	56 (10.6)	397 (4.7)	7.1
3	International Journal Of Remote Sensing	30 (5.7)	273 (3.3)	9.1
4	Journal Of Geophysical Research Atmospheres	20 (3.8)	293 (3.5)	14.7
5	ISPRS Journal Of Photogrammetry And Remote Sensing	17 (3.2)	344 (4.1)	20.2
6	International Journal Of Applied Earth Observation And Geoinformation	15 (2.8)	353 (4.2)	23.5
7	IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing (J-STARS)	13 (2.5)	73 (0.9)	5.6
8	Theoretical and Applied Climatology	12 (2.3)	353 (4.2)	29.4
9	Journal of Hydrology	10 (1.9)	215 (2.6)	21.5
9	Agricultural and Forest Meteorology	10 (1.9)	289 (3.4)	28.9
9	Hydrology and Earth System Sciences	10 (1.9)	126 (1.5)	12.6

Table 2.2. Top 10 most cited articles.

	Articles	Cited by	References
1	Remote sensing of the urban heat island effect across biomes in the continental USA	344	Imhoff et al.(2010)
2	Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa	199	Vancutsem et al.(2010)

3	Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data	141	Rhee et al.(2010)
4	Downscaling SMOS-derived soil moisture using MODIS visible/infrared data	139	Piles et al.(2011)
5	Estimating air surface temperature in Portugal using MODIS LST data	127	Benali et al.(2012)
6	Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI	109	Mallick et al.(2009)
7	Exploring indicators for quantifying surface urban heat islands of European cities with MODIS land surface temperatures	103	Schwarz et al.(2011)
8	Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products	86	Zhu et al. (2013)
9	Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest	85	Wu et al. (2010)
10	Estimation of net radiation from the MODIS data under all sky conditions: Southern Great Plains case study	84	Bisht and Bras (2010)

Table 2.2 presents the 10 most cited articles, which cover 5 application topics: UHI effect, Ta estimation, drought monitoring, soil moisture, estimating gross primary production, and net radiation estimation. Among the 10 most cited articles, 3 articles focus on Ta estimation, 2 articles focus on UHI effect, and 2 articles focus on soil moisture applications. Again, it is should be noted that the citation count in this review only used the Scopus database (other databases, e.g. Googlescholar, could show a higher number of citations). As shown in Table 2.2, the most cited article was written by Imhoff et al.(2010) with 344 citations, followed by Vancutsem et al. (2010) with 199 citations and Rhee et al.(2010) with 141 citations. These articles studied urban heat island, Ta estimation, and drought monitoring. An interesting finding is that these top three articles were published in the same year (2010), suggesting that these three topic applications could be the most applications of MODIS LST data.

2.3.3. Main Applications of MODIS LST Data
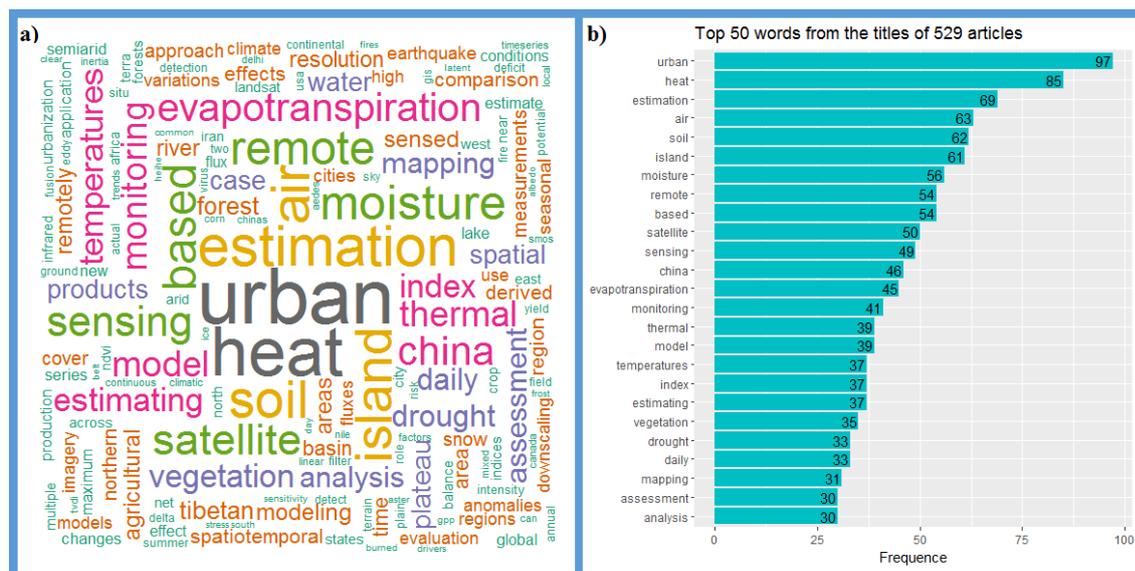
Figure 2.3. (a) the cloud of the words contained in the titles of 529 selected articles. The size of each word is indicative of its relative frequency of occurrence; (b) the top 25 most frequently used words.

By investigating the titles, abstracts, and full text (if required) of the 529 selected articles, we found that the top five applications were urban heat island (63 articles – 11.9%), air temperature estimation (52 articles – 9.8%), soil moisture estimation (50 articles – 9.5%), evapotranspiration (48 articles – 9.1%), and drought (33 articles – 6.2%). This is consistent with the hypothesis we have made based on the citations trend shown in Table 2.2. The remaining (approximately 50%) were applied for various application fields, such as risk mapping (e.g. earthquakes, forest fires, mosquitoes, floods); agriculture (rice yields, paddy rice mapping, rice quality); water (surface water temperature, lake temperature, water pollution); hydrological models; urban expansion; snow melt; snow cover; and many others (Zoran, 2012; Bisquert et al., 2012; Juri et al., 2015, Johnson, 2014; Kimura et al., 2015; Crosman & Horel, 2009; Song et al., 2016; Kuchment et al., 2010; Chakraborty et al. 2015).

Among these top 5 applications of MODIS LST, 3 applications could be considered significantly more popular because more than 50% of publications for these applications were published in the last 3 years. These applications are drought monitoring/assessment (21 articles – 63.6%), UHI effect (35 articles – 55.6%), and Ta estimation (27 articles – 52%). However, in comparison to UHI or Ta estimation studies, drought studies used additional data from MODIS sensors (and other sensors). In other words, MODIS LST is considered a key input for UHI and Ta studies, however, in drought studies, MODIS LST is just one input among many other important inputs, i.e. NDVI from MODIS, precipitation from TRMM (Rhee et al., 2010; Son et al., 2012; Hao et al., 2015; Park et al., 2016; Zhang et al., 2017).

Furthermore, Deilami et al.(2018) published a systematic review of spatio-temporal factors, data, methods, and mitigation measures of the UHI effects. Their results show that Landsat TM are the most popular data sources for UHI studies (54.6%), followed by Landsat ETM (34.6%), and MODIS just accounted for 28.0%.

Therefore, in the next section, we mainly focus on Ta estimation using MODIS LST data.

2.4. Selecting publications of Ta estimation using MODIS LST data

In this study, in order to find the suitable articles, we search in two databases, Web of Science (WoS) and Scopus, using criteria in Table 2.3.

The search resulted in 203 records (119 records in WoS and 84 records in Scopus). By reading the titles and abstracts, we removed articles that did not used MODIS LST data nor relating to air temperature, as result, there remained 81 articles for detailed check. The detailed check was implemented in order to find whether the articles meet two criteria: (1) using LST from MODIS sensor (of two satellites – Terra and Aqua), and (2) Ta was retrieved from LST data. The detailed analysis resulted in 62 eligible publications for this review study.

Table 2.3. Criteria used to select publications

Key word	MODIS LST, land surface temperature, air surface temperature, estimation
Document type	Journal article (included article in press)
Language	English
Period time	Until August 2018
Query string:	((TS = ("MODIS" OR "Moderate Resolution Imaging Spectroradiometer" AND ("LST" OR "land surface temperature"))) AND (TI = ("air" AND "temperature"))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)
	(TITLE-ABS-KEY ("MODIS" OR "Moderate Resolution Imaging Spectroradiometer" AND ("LST" OR "land surface temperature"))) AND (TITLE ("air" AND "temperature")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ip")))
Search date	August 2018

2.5. Methodology

The method applied for Ta estimation using MODIS LST data can be classified into four groups: energy-balanced method (1 articles – 1.6%), temperature-vegetation index – TVX method (6 articles – 9.7%), statistical methods (47 articles – 75.8%), and machine learning (8 articles – 12.9%).

2.5.1. Energy-Balance Method

The energy-balance method (i.e. physically based approach) was quite popular before 2005 (Pape and Löffler, 2004; Sun et al., 2005). In the energy-balance method, Ta is set as a function of many variables that describe the dynamics of the Earth system, e.g. radiation balance, soil, sensible and latent heat fluxes (Sun et al., 2005). With this method, the accuracy of Ta estimation generally ranges from 1 K to 3 K (Chen et al., 2016). Theoretically, this method is promising. However, according to Mostovoy et al. (2006), Cristobal et al. (2008), and Zhu et al. (2017), in order to accurately estimate Ta, a large of number inputs are required that are often not provided by remote sensing data, such as solar radiation, cloud-cover, wind speed, soil moisture, and land surface type.

Among the 62 articles, there is one study using energy-balance method with some modifications, called the Surface Energy Balance Algorithm for Land (SEBAL) for estimating Ta (Zhu et al., 2017). The method was applied to different study areas, the plains area, Jiangsu province (estimated Ta at 1000 m resolution), and the rugged terrain, Fujian province (estimated Ta at 30 m resolution). The results show that the RMSE at 1000 m and 30 m was 1.7 and 2.6 °C, respectively. Although the study used all input data from remote sensing images, however, one of the main criteria to choose the images is that it is acquired under clear weather conditions and there are no precipitations 3 days before the selected data. This is one of the shortcomings of the method. Furthermore, within the plain regions, the accuracy (RMSE) of the Ta estimation (at 1000 m) was only 1.7 °C, which is not as accurate as many other studies (i.e. Yoo et al., 2018; Zhang et al., 2016).

2.5.2. Temperature–Vegetation Index (TVX)

The original theory of this method based on the combined impact of vegetation cover on the average surface thermal characteristic (Friedl and Davis, 1994) and the evaporation process, which takes place to balance the energy exchange (Goward and Hope, 1989). Dense vegetation and surrounding air have a similar heat capacity. This means that with increasing vegetation cover, LST tends to approach air temperature, and the fully vegetated canopy

(NDVI-max) will have a radiometric temperature equal to Ta within the canopy. Therefore, if a negative linear relationship between vegetation index (mostly used NDVI) and LST (from remotely sensed images) can be determined, then the pure canopy temperature can be estimated from the linear relationship and the NDVI of full vegetation cover (Vancutsem et al., 2010)

The TVX method was often applied in studies implemented before MODIS LST data were available, such as the studies of Nemani and Running (1989), Goward et al. (1994), and Prihodko and Goward (1997). Since 2002, many studies have shown that TVX shows various shortcomings. One of the main drawbacks is that due to the variation of seasons, ecosystem types, and soil moisture conditions, the linear and negative slope between LST and NDVI is not always determined (Sandholt et al., 2002; Bayarjargal et al., 2006; Sun and Kafatos, 2007; Vancutsem et al., 2010). Another drawback is that the moving window calculations could result in spatial autocorrelations of the Ta estimation (Stisen et al., 2007).

Among the 62 selected articles, there are 6 articles that used TVX method for Ta estimation (i.e. Vancutsem et at., 2010; Sun et al., 2011; Zhu et al., 2013; Shah et al., 2013; Sun et al., 2014; and Kitsara et al., 2018).

Nemani et al. (1993) and Czajkowski et al. (1997) stated that, at daytime, the sharp topo-climatic gradients and variations of land cover obscure the results of Ta estimation based on the relationship between NDVI and LST. Therefore, the TVX method is not recommended for a wide or complex study area. This is demonstrated in the study of Vancutsem et al. (2010), which was implemented in different ecosystems in Africa. In this study, the authors concluded that at the continental scale, the TVX method presents limitations that affect the Ta estimation accuracy.

On the other hand, the TVX method was reported to work well in the 5 remains studies (Sun et al., 2011; Wenbin et al., 2013; Shah et al., 2013; Sun et al., 2014, and Kitsara et al., 2018). Sun et al. (2011) did not explain the advantage of the TVX methods, in comparison to other methods (i.e. statistical method) for Ta estimation. They stated that there were few studies that examined the effect of the sampling window size on the accuracy of estimated Ta. Therefore, they implemented the study in order to evaluate the effect of the sampling window size. The results showed that the accuracy of Ta estimation depend on the sampling window. For the North China Plain area, in 2003, with sampling window width of several tens of kilometers (46.0 – 83.6 km), Ta could be estimated with reliable result. Shah et al. (2013) used two methods, linear regression method for Ta-min estimation, while, using TVX method for Ta-max estimation. The results showed that the RMSE of Ta-min and Ta-max

estimation were 2.2 ad 2.16 °C, respectively. Most recently, Kitsara et al. (2018) used TVX method for Ta estimation in Greece during the growing season (May – October). They used TVX method because it is rarely applied.

2.5.3. Statistical Methods

Among the 62 selected articles, there are 47 articles (75.8%) using statistical methods for Ta estimation. This result confirms the statement of Chen et al.(2016), statistical methods are now the most popular methods for Ta estimation because compared to energy–balance methods, statistical method have simpler operability; and compared to the TVX method, they have a wider applicability for different surface covers.

In the statistical methods, it can be further classified into two categories: the simple statistical method, and the multiple (advanced) statistical method.

2.5.3.1 Simple Statistical Methods

It is known that due to the energy exchange between land surface and near surface atmosphere, Ta is strongly influenced by LST, particularly under clear sky conditions (Benali et al., 2012; Xu et al., 2018). In addition, Zakek and Schroedter-Homscheidt (2009) stated that Ta is stronger driven by LST than by direct solar radiation, therefore in Ta estimation, LST is the most important variable. In fact, a number of studies that show the strong correlation between Ta and LST, leading to Ta estimation based on a simple linear regression between Ta and LST. In the literature, a handful of studies use this method solely (Shen & Leptoukh, 2011; Yao et al., 2012; Huang et al., 2015; Cai et al., 2017; Zhang et al., 2018), rather than used together with advanced statistical methods for comparison purposes (Benali et al., 2012; Zeng et al., 2015; Noi et al., 2016; Noi et al., 2017; Janatian et al., 2017).

Shen and Leptoukh (2011) used MODIS Terra (MOD11A1, V005) for daily Ta-max and Ta-min estimation over two Eurasian regions, northern China and the fUSSR (former Union of Soviet Socialist Republics), with the simple statistic method. The effect of land cover was considered for Ta estimation. Their results showed that land cover had less impact on the relationship between LST nighttime and Ta-min than those of LST daytime and Ta-max. The accuracy (i.e. MAE – Mean Absolute Error) of Ta-min estimation was approximately 3.0 °C, and Ta-max varied from 2.4 °C to 3.2 °C depended on the land cover type. It is worth noting that in this study, LST daytime and LST nighttime were used for Ta-max and Ta-min estimation, respectively.

Zhang et al. (2016a) used both Terra and Aqua MODIS LST data (M*D11A1, V005) for daily Ta-max and Ta-min estimation over the Tibetan Plateau from 2007 to 2010. In order to analyse and interpret the effect of clouds on the accuracy of MODIS LST data and its effect on the Ta estimation accuracy, they used a simple linear regression model with LST variable.

Recently, the simple statistical methods have been mainly used for comparison with advanced statistical approaches, and even with machine learning approaches (e.g. random forest, support vector machine, and the Cubist regression). In most of these studies, the advanced statistics or machine learning approaches have shown superior performance capability. For example, in the study of Benali et al. (2012), simple statistics using only LSTtd or LSTtn achieved an R^2/RMSE for Ta-max, Ta-min, and Ta-mean of 0.875/2.27, 0.646/2.89, and 0.833/2.24, or 0.854/2.48, 0.856/1.84, and 0.923/1.50. Whereas advanced statistics including LSTtd, LSTtn, and the number of valid observations in an eight-day period (in total, there were three variables) have achieved results of 0.917/1.86, 0.871/1.74, and 0.937/1.37, respectively. Consistent with Benali et al. (2012), Zeng et al. (2015) also compared the performance of simple linear (using only LSTtd, LSTad, LSTan, or LSTtn) and advanced statistic models (combined LSTtd and LSTtn, and auxiliary Julian day, Latitude, Elevation, or Solar Zenith Angle). It should be noted that in this study, the authors have separated and analyzed different land cover types (crops, forest, and developed areas). The results showed that advanced statistical models are always superior to simple statistical models. Noi et al. (2017) compared the performance of linear regression models (including simple and multiple regression) with machine learning models (random forest and cubist regression). The results showed that simple linear regression models (using LST solely) always showed the lowest and most unstable accuracy.

However, in contrast, Meyer et al. (2016) compared the performance of simple linear regression with three machine learning algorithms (random forest, generalized boosted regression, and cubist) for daily Ta estimation using MODIS LST in Antarctica. The results showed that the machine learning approaches ($R^2 = 0.71$, RMSE =10.51 °C) were just slightly better than simple linear estimation ($R^2 = 0.64$, RMSE = 11.02 °C). Obviously, the accuracy of Ta estimation was lower than most of the Ta estimation studies (approximately between 1 K and 6 K for different methods and geographical regions). However, according to Wan (2008), MODIS LST data are stable and accurate (mostly below 1 K) in the ranges of -10 °C to 58 °C. In colder regions, such as Svalbard in Norway, Westermann et al. (2012) found that MODIS LST data has a bias of 3 K. Therefore, it can be explained that the lower accuracy

(in comparison to most of the other studies) in Ta estimation of Meyer at al. (2016) is because the average temperature of this area is approx. -20 °C.

2.5.3.2 Multiple Statistical Methods

As mentioned earlier, although there is a high correlation between LST and Ta, this correlation is complex and strongly influenced by surface and atmospheric conditions (Jin et al., 1997). Therefore, in order to improve the Ta estimation accuracy, various additional auxiliary variables have been introduced. Table 2.4 shows the studies that applied advanced statistical methods with the most popular auxiliary variables for Ta estimation using MODIS LST data.

Table 2.4. The most common auxiliary variables that have been used along with MODIS LST data to estimate Ta.

Auxiliary variables	References
Elevation	Benali et al., 2012; Lin et al., 2012; Kloog et al., 2012; Zhang et al., 2013; Niclos et al., 2014; Parmentier et al., 2014; Kloog et al., 2014; Zeng et al., 2015; Chen et al., 2015; Noi et al., 2016; Zhang et al., 2016; Chen et al., 2016; Shi et al., 2016; Didari et al., 2017; Yang et al., 2017; Noi et al., 2017; Kloog et al., 2017; Rosenfeld et al., 2017; Yoo et al., 2018; Meyer et al., 2016; Zhou et al., 2017; Huang et al., 2017
NDVI	Cristóbal et al., 2008; Kloog et al., 2012; Zheng et al., 2013; Niclos et al., 2014; Xu et al., 2014; Kloog et al., 2014; Noi et al., 2016; Zhang et al., 2016; Lin et al., 2016; Chen et al., 2016; Shi et al., 2016; Didari et al., 2017; Yang et al., 2017; Huang et al., 2017; Kloog et al., 2017; Wang et al., 2017; Rosenfeld et al., 2017; Yoo et al., 2018
Latitude/ Longitude	Cristóbal et al., 2008; Benali et al., 2012; Zheng et al., 2013; Niclos et al., 2014; Xu et al., 2014; Zeng et al., 2015; Noi et al., 2016; Zhang et al., 2016; Yang et al., 2017; Janatian et al., 2017; Wang et al., 2017; Yoo et al., 2018
Julian day	Benali et al., 2012; Peón et al., 2014; Zeng et al., 2015; Noi et al., 2016; Yang et al., 2017; Noi et al., 2017; Janatian et al., 2017;
Solar zenith angle (SZA)	Xu et al., 2012; Lin et al., 2012; Niclos et al., 2014; Zeng et al., 2015; Noi et al., 2016; Zhang et al., 2016; Yang et al., 2017; Janatian et al., 2017;
Day length	Benali et al., 2012; Weiss et al., 2014; Noi et al., 2016; Yang et al., 2017;
Distance to coast/ water body/ ocean	Benali et al., 2012; Niclos et al., 2014; Xu et al., 2014; Parmentier et al., 2014; Shi et al., 2016; Yang et al., 2017; Rosenfeld et al., 2017;
Solar radiation	Cristóbal et al., 2008; Xu et al., 2012; Xu et al., 2014; Janatian et al., 2017; Yoo et al., 2018.
Aspect	Niclos et al., 2014; Parmentier et al., 2014; Meyer et al., 2016; Rosenfeld et al., 2017; Yoo et al., 2018
Impervious surfaces/ percent urban	Kloog et al., 2012; Kloog et al., 2014; Chen et al., 2016; Shi et al., 2016; Kloog et al., 2017; Yoo et al., 2018

As shown in Table 2.4, the most popular auxiliary variables are elevation, NDVI, latitude, longitude, and Julian day, meaning that land cover (i.e. NDVI), elevation, location (latitude, longitude), and season variations (i.e. Julian day) are the factors that have the most effect on Ta estimation using MODIS LST data. Regarding the NDVI variable, different studies with different geographical locations of study areas have different conclusions. In the study by Cristobal et al. (2008), among the six auxiliary variables used (i.e. altitude, latitude, continentality, solar radiation, and NDVI), NDVI was ranked third for daily Ta estimation, and second for both monthly and annual Ta estimation as the most important variables after LST. In contrast, Benali et al. (2012) had pre-tested the effect of NDVI and stated that NDVI did not improve the accuracy of Ta estimation; therefore, they did not use NDVI as an auxiliary variable.

Chen et al. (2016) introduced a new predictor that had never been used for Ta estimation, nighttime stable light (NSL). The study was implemented in China from 2001 to 2012 with 333 weather stations. The authors concluded that by adding NSL as a predictor, the results increased by 0.18 – 2.6 K in RMSE and 0.00 – 0.15 in R^2 in comparison to other models which did not include NSL data. However, to the best of our knowledge, there are no other studies that have used NSL for Ta estimation, until now.

Most studies have been using MODIS LST data as the key variables for Ta estimation. However, Lin et al. (2012) presented a different experiment. First, they proposed to use five auxiliary variables: PW, SZA, NDVI, EVI, and elevation. They applied a stepwise regression for selecting the variables. Finally, based on the results of the stepwise processing, they used only three variables (PW, SZA, and elevation) with MODIS LST (MYD11A1, C5, from 2009 to 2010). Based on the correlation between elevation and the difference between LST and Ta (i.e. LST-Ta), they stated that it was unreliable for Ta-max estimation using only LST data, and that elevation has the strongest correlation with (LST-Ta-max) and Ta-min. Therefore, elevation played the most important role for Ta estimation. Using data from 2010 for model construction and data from 2009 for testing, the results showed that by considering only elevation, Ta-max and Ta-min could be obtained by simple regression models with an MAE (Mean Absolute Error)/AI (Agreement Index) of 1.9 °C/ 0.79 and 1.9 °C/ 0.92, respectively. The authors also concluded that due to the cloud cover (two-thirds of the LST were absent) and small variations in time (δ = 2.1 °C) over Africa, the model with elevation and LST performed slightly better than that of only elevation. Therefore, the model without LST (only elevation) is recommended in particular applications.

Another important point is about the methods for dividing, calibrating, and validating data. Most advanced statistical studies randomly divide the total dataset into two parts with proportions of 70% for model calibration and 30% for model validation (Zhang et al., 2011; Benali et al., 2012; Xu et al., 2012; Zeng et al., 2015; Chen et al., 2015; Noi et al., 2016; Janatian et al., 2017, Yang et al., 2017). A handful of studies do not randomly divide the dataset into testing and validating sub-datasets. For example, Lin et al. (2012) used data from 2010 for model construction and data from 2009 for testing. Huang et al. (2015) used data from 2003 to 2011 for calibrating the model, and data from 2012 to 2013 for model validation. Lin et al. (2016) used data from 2000 to 2005 and from 2006 to 2010 for calibration and validation, respectively.

2.5.4. Machine Learning Approaches

Among the 62 selected articles, there are only 8 articles using machine learning approaches for Ta estimation using MODIS LST data (Emamifar et al., 2013; Xu et al., 2014; Chen et al., 2015; Meyer et al., 2016; Zhang et al., 2016b; Noi et al., 2017; Xu et al., 2018; and Yoo et al., 2018).

Emamifar et al. (2013) used the M5 model tree to estimate Ta using MOD11A1 in the Khuzestan province of southwestern Iran for the year 2007. The Ta data (Ta-max and Ta-min) were collected from 29 weather stations throughout the province. Unlike most Ta estimation studies where the calibration and validation data are chosen randomly; in this study, data from 17 stations were chosen for training and data from the remaining 12 stations were chosen for testing. Before training the model, the LSTtd and LSTtn were compared to Ta-max and Ta-min, respectively. As a result, among the 17 stations, the R^2 ranged from 0.84 to 0.93 and from 0.84 to 0.94 (i.e. for LSTtd vs. Ta-max and LSTtn vs. Ta-min, respectively). Based on this correlation, the authors stated that daily mean Ta could be estimated using either daytime or nighttime data. Finally, the model was used to estimate for each one of the 12 stations. The results showed that throughout all stations, R^2 ranged from 0.92 to 0.98 (average 0.96) and RMSE ranged from 1.8 °C to 3.2 °C (average 2.3 °C). Compared to other studies, such as Zhang et al. (2016a) and Janatian et al. (2017), this is a high accuracy result for Ta estimation using only one LST term and two auxiliary variables (i.e. extraterrestrial solar radiation and Julian day). This could be explained because the model was trained from 17 stations but tested for one station, and the data were collected for one year (2007), therefore, the data could have lower variation compared to other studies. To the best of our knowledge, there is only one study using the M5 model tree to estimate Ta using MODIS LST data.

Xu et al. (2014) compared the performance of linear regression and random forest (RF) for daily Ta maximum estimation for the summer period (June, July, and August from 2003 to 2012), using Aqua MODIS LST in a mountainous region with complex terrain in British Columbia, Canada. They used nine environmental variables (LST, NDVI, MNDWI, latitude, longitude, distance to the ocean, altitude, albedo, and solar radiation) to estimate daily maximum Ta. These variables were chosen based on the stepwise multiple linear regression analysis. The Aqua MODIS (MYD11A1) was used for the study, however, the authors did not mention whether LST daytime or nighttime was used. Their result showed that RF outperformed linear regression with MAE/R^2 of 2.02 °C/0.74 and 2.41/0.64, respectively.

Meyer et al. (2016) compared the performance of a simple linear regression model with three Machine Learning algorithms: RF, generalized boosted regression models (GBM), and Cubist. Their result showed that GBM produced the highest results (R^2 = 0.71, RMSE = 10.51 °C), followed by Cubist (R^2 = 0.69, RMSE = 10.85 °C), simple linear regression (R2 = 0.64, RMSE = 11.02 °C), and RF (R^2 = 0.56, RMSE = 11.95 °C). In this study, the performance of simple linear regression compared with machine learning was explained for a study area location (Antarctica), where no land cover differences occurred, and therefore the effect of land cover types was minimized. The uncommonly very low RMSE resulted from the use of the Leave-One-Station-Out Cross-Validation (LOSOCV) for accuracy assessment. According to Gasch et al. (2015) and Meyer et al. (2016), the LOSOCV is a more strict validation strategy. It is worth noting that Zhang et al. (2016b) also used LOSOCV for Ta-mean estimation and the RMSE ranged from 2.03 °C to 2.43 °C for a separate combination of four MODIS LST data.

Zhang et al. (2016b) compared different combinations of four MODIS LST data with auxiliary variables (solar zenith and NDVI) using six statistical models for daily Ta-mean estimation (ML, the partial least squares (PLS) regression, back propagation neural network (BPNN), support vector regression (SVR), random forest (RF), and Cubist regression (CB). The authors divided the data into two cases: Case 1 – only good quality LST data for combination, and Case 2 – all LST data under clear sky conditions. The results showed that in both cases, LST nighttime plays an important role in guaranteeing high accuracy Ta-mean estimation. In Case 1 with the availability of nighttime, the simple models (i.e. ML, PLS) can estimate Ta comparable with machine algorithms (BPNN, SVR, RF, or CB). However, in Case 2, machine learning algorithms always produce better accuracy.

Noi et al. (2017) compared the performance of ML, CB, and RF for daily Ta (Ta-max, Ta-min, and Ta-mean) estimation in rural and mountainous area in northwest Vietnam. The

authors evaluated the models based on two datasets: Dataset 1 – only four MODIS LST, and Dataset 2 – four MODIS LST data with auxiliary data (elevation and Julian day). The results showed that using Dataset 2, the RF and CB algorithms produced stable and high accuracy results with all combinations of the four LSTs. However, with the LM algorithm, the more LST terms (especially LST nighttime) used in the models caused higher accuracy to be achieved. A larger, separate impact of different combinations of LST was observed more in Dataset 1 than in Dataset 2. Compared between Ta-max, Ta-mean, and Ta-min; with Dataset 1, Ta-mean achieved the highest accuracy, and with Dataset 2, Ta-max achieved the highest accuracy. Among the three algorithms, CB always showed the highest accuracy Ta estimation for all combinations.

Xu et al. (2018) used ten machine learning algorithms (Bayesian regularized neural networks, support vector machines - SVM, least absolute shrinkage and selection operator - LASSO, ridge regression - Ridge, generalized linear model - GLM, multivariate adaptive regression splines - MARS, conditional inference tree - CIT, random forest - RF, eXtreme gradient boosting - XGB, and cubist - CB) for monthly mean Ta estimation using MOD11A1 and 11 environmental variables. The results showed that the CB algorithm outperformed other algorithms with the highest accuracy and lowest sensitivity to the cloud contamination of LST data. The results also showed that among the 11 variables, LSTtn was the most important variable, followed by elevation and solar radiation. It is should be noted that in this study, only daily LST data from the Terra satellite (MOD11A1) was used with both daytime and nighttime LST, however, the accuracy of Ta estimation was very high (RMSE = 1.00 ℃, MAE = 0.73 ℃). The authors used daily LST data, which is only available under clear sky condition, whereas monthly mean Ta was calculated from daily mean Ta, which was measured from 135 weather stations under all sky conditions; however, the authors did not discuss this difference. Therefore, in future research, if weekly or monthly Ta data (from weather stations) are used for testing (validating) Ta estimation from MODIS LST, the effect of different sky conditions should be taken into account.

Yoo et al. (2018) presented a new idea about MODIS LST data combination. In order to account for the time lag between transporting heat from the ground surface (land surface) to the 2 m air temperature, they also used the LST data of the day before. In total, there were eight LST data (four times per day and four times from the day before) together with seven popular auxiliary variables (elevation, solar radiation, NDVI, latitude, longitude, aspect, and the percentage of impervious area) used for Ta-max and Ta-min estimation. Two different characteristic climate cities, Los Angeles (USA) and Seoul (South Korea), were chosen to

implement the study. The results showed that in the urban landscape, the LST data of the day before plays a crucial role in Ta estimation. Although the study was applied during hot summer days (July and August), meaning the effect of season variation is minimized, the best results were R^2 = 0.850 and 0.777/0.728 and 0.767; RMSE = 1.7 °C and 1.2 °C/1.1 °C and 1.2 °C for Ta-max and Ta-min, in Los Angeles/Seoul, respectively. It is suggested that the complexity of the urban landscape has an impact on the accuracy of Ta estimation. Therefore, more studies in urban areas should be implemented and investigated.

2.6. Different Combinations of the four MODIS LST

2.6.1. MODIS LST Data Selection

As we have already mentioned in Section 2.1, there are four MODIS LST data: Terra Day (LSTtd), Terra Night (LSTtn), Aqua Day (LSTad), and Aqua Night (LSTan). However, most studies do not use all four of these LST data. The MODIS LST products chosen for a specific study are based on some of the following criteria.

The most popular method for choosing MODIS LST variables is based on the closest time with Ta (i.e. LST nighttime for Ta-min estimation, and LST daytime for Ta-max estimation). Xu et al. (2014) used only LSTad (and nine auxiliary variables) for daily Ta-max estimation in British Columbia, Canada, simply because the overpass time of is around 13:00 local time, closest to the maximum Ta time. Alfieri et al. (2013) stated that the overpass time of LSTtd of around 11:30 local solar time is close to the Ta-max time in Southern Italy, therefore they used only LSTtd for Ta-max estimation. Similarly, Zhang et al. (2016a) used LSTad (overpass time 13:30 solar local time) and LSTtn (approximately 22:30 solar local time) for daily Ta-max and Ta-min estimation, respectively. Shi et al. (2017) reviewed previous studies (Zhang et al., 2011), and stated that the land surface type has more impact on the correlation between daytime LST and Ta-max than on nighttime LST and Ta-min. This statement is consistent with the study of Shen and Leptoukh (2011), who used daily (MOD11A1 V005) LST daytime for Ta-max estimation at the Yangtze River Delta in China by considering different land cover types.

Most recently, Yang et al. (2017) chose Aqua MODIS LST for Ta estimation in northeastern China because the daytime and nighttime passing time of the Aqua satellite is closer to the maximum and minimum Ta occurrence than the Terra satellite.

Cai et al. (2017) stated that although LSTad has closer overpass time to Ta-max, previous studies (e.g. Fu et al., 2011; Wenbin et al., 2013) concluded that Terra LST produced better

Ta estimation results, and they therefore used daytime Terra LST for Ta-max estimation at the middle and lower reaches of the Yangtze River (in China) from 1 December 2008 to 31 January 2010.

Mostovoy et al. (2006) found that in the Mississippi State, USA, area, the overpass time of Terra and Aqua has little impact on the accuracy of Ta estimation. Moreover, in many studies, although LST daytime always has a closer overpass time to Ta-max than LST nighttime, the performance of LST nighttime for Ta-max estimation was better than LST daytime. Zhang et al. (2011) built empirical models for daily Ta-max, Ta-min, and Ta-mean estimation using both Terra and Aqua LST products. They concluded that nighttime LST (LSTan, LSTtn) was the optimal variables for Ta-min, Ta-mean, and even Ta-max estimation. Consistent with Zhang et al. (2011) is the study of Zeng et al. (2015), where both LST daytime and nighttime were used for Ta-max and Ta-min estimation in the United States Corn Belt during the growing season (May-September) from 2008 to 2012. The results showed that for both Ta-max and Ta-min estimation, MODIS LST nighttime always showed better results than daytime. The authors also concluded that the combination of daytime and nighttime LST produced the highest accuracy in comparison to LST daytime/nighttime use solely. However, the combination only compared the same satellite (Aqua or Terra). Another important point is that in this study, Zeng et al. (2015) constructed the "more variables" model based on the better performance of the "fewer variables" model. For example, when (LSTtd + LSTtn) was better than (LSTad + LSTan) for Ta-max, they chose only (LSTtd + LSTtn) to combine with DOY (date of year), SZA (solar zenith angle), Lat (latitude), or Ele (elevation). In contrast, because the combination (LSTad + LSTan) was better than (LSTtd + LSTtn) for Ta-min estimation, they chose only (LSTad + LSTan) to combine with DOY, SZA, Lat, or Ele. In addition, although DOY, SZA, Lat, and Ele were used as auxiliary variables in order to integrate with MODIS LST for Ta estimation, unlike other studies, these auxiliary variables were used separately with MODIS LST (i.e. not all selected auxiliary variables were included in one model).

The combination of MODIS LST for Ta estimation was further evaluated by Zhang et al. (2016 b) and Noi et al. (2016, 2017). In these studies, the authors were not only combining LST nighttime and daytime of the same satellites but also different satellites (i.e. daytime of Terra and nighttime of Aqua, or both daytime and nighttime of both Terra and Aqua). As a result, if more LST variables are combined then a higher accuracy of Ta estimation can be achieved. Particularly, the use of LST nighttime LST was considered a guarantee for receiving high accuracy of Ta estimation (Zhang et al., 2016b; Noi et al., 2016; Noi et al., 2017).

Other studies have shown different reasons to choose LST data for Ta estimation. For example, Benali et al. (2012) used MODIS LST data from Terra because it covers the entire study period; from 2000 to 2009 (Aqua was not selected for this study because it has only been available since mid-2002). Kloog et al. (2017) stated they had checked both LSTtd and LSTtn, and decided to use LSTtn as the main predictor without any further explanation. Lin et al. (2016) tested all four MODIS LST (MOD11C3 and MYD11C3; daytime and nighttime) data for monthly Ta (Ta-max, Ta-min, and Ta-mean) estimation. Strong correlations between nighttime LST (of Terra and Aqua), and Ta-mean and Ta-min ($R^2 > 0.90$) were observed. For Ta-max, LSTan was better than LSTtn, however, due to the lack of LSTan data (due to cloud cover), the LSTtn was chosen for further analysis. Some studies, simply used the suggestion of previous studies (in different geographical locations) to choose MODIS LST data, such as Zheng et al. (2013) choosing nighttime LST of MOD11A2 product for monthly Ta-max, Ta-mean, and Ta-min estimation in northern China based on studies by Wang et al. (2006) and Zhang et al. (2009) using it.

2.6.2. Effects of Different Combination of the four MODIS LST

As mentioned in Section 4.1, there are a number of studies that have chosen one LST variable for Ta estimation. Among these, LST daytime was used for Ta-max estimation and LST nighttime for Ta-min estimation. However, with the recent development of Ta estimation methods, a combination of multiple MODIS LST data is being used more and more often.

To increase the accuracy of Ta estimation, additional auxiliary variables have been introduced and complex models have been applied, yet surprisingly, most of these studies have used only one LST variable (among the four available LST: LSTtd, LSTtn, LSTad, LSTan) with other auxiliary variables. However, more recently, there are studies comparing the different combination of MODIS LST data (using more than one LST term in a model) for Ta estimation.

Inconsistent with previous studies, Benali et al. (2012) showed that the nighttime LST was better than daytime LST for Ta estimation, and the combination of both daytime and nighttime was better than a single LST of daytime or nighttime. However, they only combined two MODIS LST from the same satellite (LSTtd and LSTtn) for Ta estimation because Aqua MODIS LST did not cover the entire study period (from 2000 to 2009). Zeng et al. (2015) estimated Ta-max and Ta-min with three different land cover types (crops, deciduous forest, and developed areas), with various combinations between MODIS LST of daytime and nighttime from both Terra and Aqua sensors. They found that in all land cover

types, LST nighttime showed a higher correlation with T-max than LST daytime. For dynamic combination, they showed a higher level of Ta estimation methods by using both Terra and Aqua MODIS LST (daytime and nighttime) for Ta-max and Ta-min estimation. However, the combination of daytime and nighttime LST data was only with the same sensor (i.e. LSTtd + LSTtn, or LSTad + LSTan).

Zhang et al. (2011) seem to be the first authors to have the idea of combining all four MODIS LST. In their study, MODIS LST in the year 2003 was used for Ta-max, Ta-min, and Ta-mean estimation in China. To construct models, they calculated the linear regression correlation between Ta and all four MODIS LST products. Based on the correlations they concluded that LST nighttime was better than LST daytime for Ta estimation, and the combination of daytime and nighttime achieved higher correlation with Ta than LST daytime or LST nighttime solely. Furthermore, they also tested the combination of all four MODIS LST, however, based on the correlation with Ta they concluded that these combinations of four LST data did not provide further improvement of Ta estimation. This is an incomplete conclusion, because in later studies (e.g. Zhang et al., (2016b) and Noi et al. (2016, 2017)) have shown that the combination of all four LST would produce the highest accuracy of Ta estimation. Even just combinations of LSTtn and LSTan would produce better results than the coupling of LSTtd and LSTtn, or LSTad and LSTan.

Most recent studies combine all four MODIS LST for Ta estimation (Noi et al., 2016; Zhou et al., 2017; Noi et al., 2017; Zhang et al., 2017; Yoo et al., 2018).

2.7. Factors effect on Ta estimation using MODIS LST

In Ta estimation, using MODIS LST data, cloud cover, land use/cover, seasons change, and complex terrain is considered as one of the main factors effect on the accuracy of estimated Ta.

2.7.1. Cloud Cover Effect

It is known that the optical band signal (of MODIS sensor) cannot penetrate clouds; therefore, land surface information will be affected. According to Jin (2000), day-by-day, cloudy sky conditions are present for more than half the globe. This means that MODIS LST data, which have excluded all the cloudy sky pixels from M*D11 products, lose more than half of their time series LST data (Sun et al., 2017).

In order to provide more information about this problem and to increase the accuracy of Ta estimation, a number of reconstruction methods have been developed and applied by researchers. For example, Yu et al. (2013) used neighboring-pixel and MODIS net surface shortwave radiation products to reconstruct MODIS LST in the semi-arid and arid regions of northern China. The reconstructed LST was compared to ground stations data, yielding an absolute error under 2.6 K. According to Fan et al. (2014), two key factors have the most influence on LST variations, topography and land cover. In fact, when reconstructing MODIS LST in mountain regions, the influence of topography was considered and confirmed in studies by Neteler (2010) and Frey and Kuenzer (2014). However, for fragmented landscape and flat terrain regions, where the relationship between LST and topography is weak, the land cover and soil conditions are the most important and influential variables (Fan et al., 2014).

According to Wan and Dozier (1996), clouded pixels can be removed from MODIS LST products by the generalized split-window algorithm. If some pixels are covered by thin clouds, however, the algorithm may not detect and remove them. Therefore, the quality information (QC file) of each product must be provided for choosing quality MODIS LST data. QC flag ranges were set from $0 - 3$, the average error of MODIS LST data ranges from $<1, 1 - 2, 2 - 3$, and > 3 K (Zhang et al., 2016a). In the literature, many studies chose only good LST data (based on QC data) for the experiment without any pretest or evaluation (Zhang et al., 2011; Benali et al., 2012; Zeng et al., 2015; Zhang et al., 2016a, Cai et al., 2017; Zhang et al., 2018). However, there are also a number of studies using all LST under all clear sky conditions for Ta estimation (Cristóbal et al., 2008; Vancutsem et al., 2010; Shen and Leptoukh, 2011; Emamifar et al., 2013; Chen et al., 2015; Lin et al., 2016; Oyler et al., 2016; Huang et al., 2017; Janatian et al., 2017). Some studies compared all clear sky LST data with good LST data and concluded that good quality LST is better for Ta estimation. For example, Shi et al. (2017) compared the results of Ta estimation using all valid LST data (MAE ≈ 2.5 ºC) and the best-quality LST (error < 1 K) for Ta estimation at the Yangtze River Delta (in China) and concluded that the error was reduced by 0.4 ºC (15%) when using the best-quality LST data. However, Zhang et al. (2016 b) compared different combinations of all four MODIS LST data (i.e. LSTtd, LSTtn, LSTad, LSTan) with different LST data quality (all clear sky data, only good data), and six different statistic models ranging from simple (e.g. multiple linear regression – MLR and partial least squares regression – PLS) to complex models (e.g. support vector regression – SVR, random forests – RFs, and Cubist regression – CB). The results showed that in most cases, good quality data produced higher accuracy for Ta estimation. However, in the models which used only daytime LST data (LSTtd, LSTad,

or LSTad + LSTtd) and advanced models (i.e. Cubist regression), the mixed-quality (all clear sky condition) was better for producing a high accuracy of Ta estimation than only good data.

Therefore, the performance of all clear sky conditions LST and only good LST data for Ta estimation is still limited and needs to be further investigated.

2.7.2. Land Use/Cover Effect

It is known that the land surface emissivity is sensitive to land cover type, particularly during the daytime, and MODIS LST is a measurement of the surface radiation and is calculated based on surface emissivity (Wan and Dozier, 1996). Therefore, the land cover type does have an effect on the relationship between LST and Ta.

The impact of land cover on the performance of Ta estimation using MODIS LST could be clearly seen from the studies that used simple statistic methods or studies using one LST data (i.e. LST daytime for Ta-max estimation and nighttime for Ta-min estimation). Regarding the accuracy of Ta estimation, although the models were developed for specific land cover types, however, the R^2 usually ranges from 0.75 to 0.95 and MAE ranges from 2 to 4 °C. For example, Shen and Leptoukh (2011) used Terra MODIS LST data (MOD11A1, collection 5) over two Eurasian regions (northern China and fUSSR, in 2009 and 2000, respectively) to estimate Ta. They evaluated the relationship between LST and Ta and found that LST nighttime and Ta-min have a slightly higher correlation than LST daytime and Ta-max. In addition, the relationship between LST daytime and Ta-max depends significantly on land cover types, whereas LST nighttime and Ta-min does not. Based on this pre-evaluation, the Ta-min and Ta-max were estimated with and without considering land cover types, respectively. The results of this study showed that the mean absolute error (MAE) for Ta-max estimation ranged from 2.4 °C (closed shrublands) to 3.2 °C (grassland), and was approximately 3.0 °C for Ta-min.

To evaluate the impact of land cover types and seasonal variation on Ta-max estimation using MODIS LST data (MOD11A1, only good data with an average LST error of no more than 1 K), Cai et al. (2017) used data at the middle and lower reaches of the Yangtze River in China. The results showed that the accuracy of Ta-max estimation was highly dependent on land cover types and season variations. Regarding the land cover types, the performance of the model including all land cover types ($R^2 = 0.87$, MAE = 2.04) was better than that of only forest ($R^2 = 0.85$, MAE = 2.35) or impervious surface ($R^2 = 0.84$, MAE = 1.98). The remaining land cover types (i.e. cropland, grassland, shrub, and water), had better results than

the model including all land cover types. Particularly, the models using grassland (R^2 = 0.92, MAE =1.89) and water (R^2 = 0.96, MAE =1.93) had the best results. It is worth noting that these two land cover types also had the lowest number of observations. Among 3132 observations of six land cover types, there were only 26 (0.83%) and 92 (2.94%) observations for water and grassland, respectively.

Meanwhile, most of the studies using advanced statistic methods (using between one and four LST terms with auxiliary variable data) produced a higher accuracy of Ta estimation with all land cover types, even in a larger area with complex land surface characteristics (Janatian et al., 2017). In other words, in these studies, the effects of land cover was considered by adding an auxiliary variable, such as NDVI, elevation, reflectance, or zenith angle. According to Cai et al. (2017), these earlier-mentioned auxiliary variables could indirectly account for the effect of land cover variations. Janatian et al. (2017) used Terra MODIS LST to estimate Ta-mean for a large spatiotemporal extent over the eastern part of Iran with 20 synoptic stations in five years, from March 2000 to December 2004. The best model (included LSTtn and five auxiliary variables, Julian day, solar zenith angle, NDVI, Lat, and Alt) produced a high accuracy of Ta estimation, with MAE of 2.3 and 1.8 at the daily and weekly scales, respectively. It should be mentioned that Janatian et al. (2017) had compared the performance of only LST daytime and nighttime for Ta estimation. Because LST nighttime produced better results of Ta estimation than LST daytime, therefore, they had chosen only LST nighttime to integrate with the auxiliary variables.

2.7.3. Seasonal Effect

In the literature, a number of studies show that the seasons not only affect the relationship between LST and Ta, but also the accuracy of Ta estimation using LST data. For example, Mutiibwa et al. (2015) analyzed the influence of seasons on the relationship between LST and Ta in complex terrain of two eco-regions of Nevada, USA, and found that the strongest correlation was achieved in fall (R^2 = 0.92), followed by summer (R^2 = 0.92), spring (R^2 = 0.86), and winter (R^2 = 0.74). The effect of seasons on the relationship between LST and Ta was also reported in the study of Mostovoy et al. (2006).

In order to reduce the effect of seasons on Ta estimation accuracy, additional auxiliary variables were used. Among these, the most popular auxiliary variables are Julian day and solar zenith angle. Julian day was used to reflect the variation of Ta due to the change in season (Jang et al., 2004, Benali et al., 2012; Zeng et al., 2015; Yang et al., 2017), while solar

zenith angle was used to reflect the diurnal difference of Ta (Jang et al., 2004; Xu et al., 2012; Zeng et al., 2015).

A clear proof of the seasonal effect on Ta estimation using MODIS LST is given by the study of Janatian et al. (2017), which was conducted in a large study area in the eastern part of Iran over a period of five years (2000 to 2004). Since the author used only LSTtn with many different models (different auxiliary variable combinations), the improvement of estimation when Julian day was included is clearly seen. The one-step-ahead stepwise regression was used in order to explore the significance of each variable in the improvement of Ta estimation. Among 11 auxiliary variables (see Janatian et al., 2017), the Julian day showed up as the most important variable (after LSTtn), so it was included in most estimated models. In addition, a direct comparison between the model using LSTtn solely (model 2) and model using LSTtn and one auxiliary variable, Julian day (model 3), showed that an improvement (R^2 increased from 0.778 to 0.859, and MAE decreased from 3.6 ℃ to 2.8 ℃) had been made.

Another example is a study using Julian day and solar zenith angle as auxiliary variables; however, the study was implemented in a context with less seasonal variation. Zeng et al. (2015) included DOY and SZA into the models, however, these variables only slightly increased the accuracy of Ta-max estimation (compared to models using only LST data) in all three land cover types, and the Ta-min estimation was not improved. This can be explained by the study of Zeng et al. (2015), which only considered a specific season (growing season, May to September, from 2008 to 2012). Obviously, during this time (May to September), the three land cover types (crops, forest, and developed) did not appear a significantly change.

Another auxiliary variable that has been used in the study of Zhang et al. (2011) for reducing the effect of seasonal variation is solar declination (Sd). However, it should be noted that Sd is calculated based on Julian day. (For more detail, please see Zhang et al., 2011). By introducing the Sd variable into the models, the authors concluded that Sd had improved the model performances by significantly reducing the seasonal estimation biases and the overall errors of Ta estimation.

2.7.4. Elevation and Terrain Effect

According to Wan (2008), the valid MODIS LST under all clear sky condition, which is defined by MODIS cloud mask, has a confidence of 66% and 95% for a land area with an elevation above and below 2000 m, respectively.

Regarding the relationship between temperature and elevation, for the atmosphere at a station location, when elevation increases, the air pressure and Ta will subsequently decrease. This decrease is referred to as the environmental lapse rate, and can be seen in high-altitude station measurements. According to Khandelwal et al. (2018), depending on the moisture conditions, the environmental lapse rate varies from 5 °C to 10 °C when the elevation increases by 1000 m. In the literature, a number of studies have shown the relationship between elevation and Ta (You et al., 2008; Liu et al., 2009) as well as elevation and LST (Khandelwal et al., 2018; Phan et al., 2018). It is worth noting that LST and Ta are not equivalent and the relationship between elevation and Ta is different from that of elevation and LST (Phan et al., 2018). Therefore, the elevation influences the relationship between LST and Ta as well as the accuracy of Ta estimation using LST data. This influence has been reported in many studies (Lin et al., 2012; Benali et al., 2012; Zeng et al., 2015). According to Mutiibwa et al. (2015), the relationship between LST and Ta varies with atmospheric changes, seasons, and geographical locations.

In various studies, elevation is reported as one of the most important variables that helps to improve the accuracy of Ta estimation. Xu et al. (2018) used MODIS LST (MOD11A1) and 11 auxiliary variables to map monthly mean-Ta in the Tibetan Plateau from 2001 to 2015. Among these variables, the authors used the forward variable selection method to selection the most important variables in order to minimize the overfitting problem. The result showed that LSTtn was the most important variable, followed by elevation and solar radiation. However, Zeng et al. (2015) showed a contrasting result of the effect of elevation on the accuracy of Ta estimation. Together with MODIS LST data; other variables, such as Julian day, solar zenith angle, latitude, and elevation were also taken into consideration. The results showed that when latitude and elevation were included in the model, the accuracy of Ta estimation was relatively unchanged. It can be explained because this study area elevation (the US Corn Belt) is not too complicated, with an elevation ranging from 87 to 666 m from east to west. Moreover, it is worth noting that the impact of elevation on Ta-max, Ta-min, and Ta-mean estimation accuracy is not similar. Yang et al. (2017) evaluated MODIS LST (MYD11A2) and seven auxiliary variables (see Table 3 of Yang et al., 2017 for more detail) for eight-day Ta estimation in a complex terrain area in northeastern China. Among various variables, they used the Standardized Regression Coefficients methods to identify the relative importance of predictors for Ta estimation. The best models for Ta-max estimation included LSTad and five variables (the days in clear sky conditions and with valid daytime LST, Latitude, Julian day, Day length, and NDVI). The best models for Ta-min estimation included LSTan and two variables (Elevation and Day length); while the best models for Ta-

mean estimation included LSTan and two variables (Day length, the days in clear sky conditions and with valid nighttime LST). Although there were more variables including in the Ta-max estimation (six variables) than in the Ta-min (3 variables) and Ta-mean estimated models (three variables), the results of Ta-min ($R^2 = 0.94$, RMSE = 3.99 °C) and Ta-mean ($R^2 = 0.94$, RMSE = 3.60 °C) estimation were better than those for Ta-max ($R^2 = 0.90$, RMSE = 4.63 °C). It is suggested that the elevation has more impact on the accuracy of Ta-min estimation than on Ta-max and Ta-mean estimation.

2.8. Conclusions and Future Research Directions

Most of the studies have shown that LST is the most important variable for Ta estimation. However, some other factors; such as elevation, land cover types, vegetation indexes, seasons, soil moisture, solar radian, geographical locations (Lat, Long, Altitude), day length, and distance to the coast also have an effect on the accuracy of Ta estimation. In order to account for the effects of these factors, advanced statistic models (e.g. multiple variable linear models) have been developed and applied. For example, Julian day and elevation (DEM) were integrated into models to reflect the annual (seasonal) variation and lapse rate of temperature. NDVI or distance to the coast were used to account for the effect of land cover. However, the relationship between Ta and environmental variables is complicated, and usually not linear (Xu et al., 2014). Therefore, a linear model (either single or multiple linear), cannot handle this complicated relationship well. That is why in all studies have compared the performance of linear regression with machine learning approaches, since machine learning approaches always show a better result than linear regression approaches (Emamifar et al., 2013; Xu et al., 2014; Zhang et al., 2016; Noi et al., 2017). Currently, however, there are a handful of studies using machine learning methods for Ta estimation. Therefore, for the future research, more machine learning approaches and more machine learning studies of different land surface characteristics, different terrain, and different geographical locations should be applied and evaluated.

The second issue of concern is that most Ta estimation using MODIS LST data has typically focused on non-urban/built-up areas (Vancutsem et al., 2010; Fu et al., 2011; Zeng et al., 2015; Huang et al., 2017; Cai et al., 2017). Meanwhile, Ta information with spatiotemporal extensions in urban landscapes is of crucial importance because it closely relates to many urban problems, such as the urban heat island effect (Li et al., 2004; Kruger and Emmanuel, 2013), air pollution (Aw and Kleeman, 2003), and human mortality (Liu et al., 2011). Since urbanization is taking place in many regions around the world, the resulting number of

weather stations in urban areas is increasing (Muller et al., 2013), including newly installed stations and many stations which were located in non-urban areas before, are now surrounded by urban surfaces (Shao et al., 2011). Moreover, urban areas usually have a large population and infrastructure is complex, therefore, a small change in temperature may significantly affect both humans and the environment (Hondula et al., 2012; Schuster et al., 2014). Therefore, understanding the spatiotemporal variation of urban Ta is very important. However, in urban landscapes, changing of Ta could take place over a short distance. MODIS (both Aqua and Terra) satellites can only provide LST data with a spatial resolution of 1 km. Therefore, it is important to downscale temperature into a higher spatial resolution (i.e. Landsat resolution).

There is no doubt that by integrating some other factors/auxiliary variables, the accuracy of Ta estimation has been increasing. However, the correlation between these auxiliary variables should not be neglected (Xu et al., 2012; Kloog et al., 2012), as correlations between variables could lead to the variables parameter bias, and therefore the results of Ta estimation accuracy could be inflated. In addition, the degree of improvement of machine learning approaches in comparison to linear methods depends on different regions, different auxiliary variables included in the models, and different combinations of MODIS LST.

Another issue that should be taken into consideration is the LST quality of MODIS LST. As mentioned in the discussion section, the results of Ta estimation using only good quality data and all clear sky data still vary between studies. Therefore, more studies to evaluate the effect of MODIS LST quality on the accuracy of Ta estimation should be considered and implemented in different regions with different methods. Another important point is that the number of missing data pixels will increase significantly if using only good (or best) LST data quality (Shi et al., 2017).

Selection of the correct MODIS LST product (i.e. Terra/Aqua and daytime/nighttime), auxiliary variables, and methods depends upon the characteristics of each individual area to be studied. For example, for a complex terrain area, elevation is one of the most important auxiliary variables to consider. Furthermore, when only one LST is used, the simple method (i.e. linear regression) should be applied, rather than complex methods like the support vector machine or the random forest algorithm. However, a complex combination of MODIS LST products and multiple variables may require a complex method. In conclusion, there are remaining issues that should be addressed in future research: (i) reconstruction of MODIS LST data; (ii) fusing MODIS with higher spatial LST data/downscaling MODIS LST data

(i.e. to Landsat resolution); (iii) evaluating the effect of LST data quality (i.e. the tradeoff between spatial coverage and accuracy) on the accuracy of Ta estimation.

References

Alfieri, S.M., De Lorenzi, F., & Menenti, M. (2013). Mapping air temperature using time series analysis of LST: the SINTESI approach. Nonlinear Processes in Geophysics, 20, 513-527

Anamarija, R., Marijana H.N., & Marija, S. (2015) Modelling river temperature from air temperature: case of the River Drava (Croatia). Hydrological Sciences Journal, 60:9, 1490-1507.

Aw, J. (2003). Evaluating the first-order effect of intraannual temperature variability on urban air pollution. Journal of Geophysical Research, 108

Babu K.V, S., Roy, A., & Prasad, P.R. (2017). Forest fire risk modeling in Uttarakhand Himalaya using TERRA satellite datasets. European Journal of Remote Sensing, 49, 381-395

Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10, 438–452.

Bayarjargal, Y., Karnieli, A., Bayasgalan, M., Khudulmur, S., Gandush, C., & Tucker, C. (2006). A comparative study of NOAA–AVHRR derived drought indices using change vector analysis. Remote Sensing of Environment, 105, 9-22

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124, 108-121

Bhardwaj, A., Singh, S., Sam, L., Bhardwaj, A., Martín-Torres, F.J., Singh, A., & Kumar, R. (2017). MODIS-based estimates of strong snow surface temperature anomaly related to high altitude earthquakes of 2015. Remote Sensing of Environment, 188, 1-8

Bisht, G., & Bras, R.L. (2010). Estimation of net radiation from the MODIS data under all sky conditions: Southern Great Plains case study. Remote Sensing of Environment, 114, 1522-1534

Bisquert, M., Caselles, E., Snchez, J.M., & Caselles, V. (2012). Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. International Journal of Wildland Fire, 21, 1025-1029

Bisquert, M., Sánchez, J.M., & Caselles, V. (2016). Evaluation of Disaggregation Methods for Downscaling MODIS Land Surface Temperature to Landsat Spatial Resolution in Barrax Test Site. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(4), 1430-1438.

Blackett, M., Wooster, M.J., & Malamud, B.D. (2011). Exploring land surface temperature earthquake precursors: A focus on the Gujarat (India) earthquake of 2001. Geophysical Research Letters, 38, n/a-n/a

Bonafoni, S. (2016). Downscaling of Landsat and MODIS Land Surface Temperature Over the Heterogeneous Urban Area of Milan. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), 2019-2027.

Cai, Y., Chen, G., Wang, Y., & Yang, L. (2017). Impacts of Land Cover and Seasonal Variation on Maximum Air Temperature Estimation Using MODIS Imagery. Remote Sensing, 9, 233

Cammalleri, C., & Vogt, J. (2015). On the Role of Land Surface Temperature as Proxy of Soil Moisture Status for Drought Monitoring in Europe. Remote Sensing, 7, 16849-16864

Chakraborty, S.D., Kant, Y., & Mitra, D. (2015). Assessment of land surface temperature and heat fluxes over Delhi using remote sensing data. Journal of Environmental Management, 148, 143-152

Chen, F., Liu, Y., Liu, Q., & Qin, F. (2015). A statistical method based on remote sensing for the estimation of air temperature in China. International Journal of Climatology, 35, 2131-2143

Chen, W., Shen, H., Huang, C., & Li, X. (2017). Improving Soil Moisture Estimation with a Dual Ensemble Kalman Smoother by Jointly Assimilating AMSR-E Brightness Temperature and MODIS LST. Remote Sensing, 9, 273

Chen, Y., Quan, J., Zhan, W., & Guo, Z. (2016). Enhanced Statistical Estimation of Air Temperature Incorporating Nighttime Light Data. Remote Sensing, 8, 656

Christiansen, B. (2005). Downward propagation and statistical forecast of the near-surface weather. Journal of Geophysical Research, 110, D14104.

Cristobal, J., Ninyerola, M., & Pons, X. (2008). Modeling air temperature through a combination of remote sensing and GIS data. Journal of Geophysical Research, 113

Crosman, E.T., & Horel, J.D. (2009). MODIS-derived surface temperature of the Great Salt Lake. Remote Sensing of Environment, 113, 73-81

Czajkowski, K.P., Goward, S.N., Stadler, S.J., & Walz, A. (2000). Thermal remote sensing of near surface environmental variables: Application over the Oklahoma Mesonet. The Professional Geographer, 52, 345–357.

Czajkowski, K.P., Mulhern, T., Goward, S.N., Cihlar, J., Dubayah, R.O., & Prince, S.D. (1997). Biospheric environmental monitoring at BOREAS with AVHRR observations. Journal of Geophysical Research: Atmospheres, 102, 29651-29662

Dantur Juri, M.J., Estallo, E., Almirón, W., Santana, M., Sartor, P., Lamfri, M., & Zaidenberg, M. (2015). Satellite-derived NDVI, LST, and climatic factors driving the distribution and abundance of Anopheles mosquitoes in a former malarious area in northwest Argentina. Journal of Vector Ecology, 40, 36-45

De Wit, A.J.W., & van Diepen, C.A. (2008). Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. International Journal of Applied Earth Observation and Geoinformation, 10, 414–425.

Dhorde, A.G., & Patel, N.R. (2016). Spatio-temporal variation in terminal drought over western India using dryness index derived from long-term MODIS data. Ecological Informatics, 32, 28-38

Duan, S. B., & Li, Z. L. (2016). Spatial Downscaling of MODIS Land Surface Temperatures Using Geographically Weighted Regression: Case Study in Northern China. IEEE Transactions on Geoscience and Remote Sensing, 54(11), 6458-6469.

Emamifar, S., Rahimikhoob, A., & Noroozi, A.A. (2013). Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. International Journal of Climatology, 33, 3174-3181

Fan, X.-M., Liu, H.-G., Liu, G.-H., & Li, S.-B. (2014). Reconstruction of MODIS land-surface temperature in a flat terrain and fragmented landscape. International Journal of Remote Sensing, 35, 7857-7877

Florio, E.N., Lele, S.R., Chi Chang, Y., Sterner, R., & Glass, G.E. (2004). Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach. International Journal of Remote Sensing, 25, 2979-2994

Frey, C.M., & Kuenzer, C. (2014). Land-surface temperature dynamics in the Upper Mekong Basin derived from MODIS time series. International Journal of Remote Sensing, 35, 2780-2798

Frey, C.M., Kuenzer, C., & Dech, S. (2012). Quantitative comparison of the operational NOAA-AVHRR LST product of DLR and the MODIS LST product V005. International Journal of Remote Sensing, 33, 7165-7183

Fu, G., Shen, Z., Zhang, X., Shi, P., Zhang, Y., & Wu, J. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. Acta Ecologica Sinica, 31, 8-13

Gasch, C.K., Hengl, T., Gräler, B., Meyer, H., Magney, T.S., Brown, D.J. (2015). Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spatial Statistics, 14A, 70–90

Goward, S. N., Waring, R. H., Dye, D. G., & Yang, J. (1994). Ecological remote sensing at OTTER: Satellite macroscale observations. Ecological Applications, 4, 322–34

Guangmeng G., & Mei Z. (2004). Using MODIS land surface temperature to evaluate forest fire risk of northeast China. IEEE Geoscience of Remote Sensing Letter, 1(2):98–100.

Hao, C., Zhang, J., & Yao, F. (2015). Combination of multi-sensor remote sensing data for drought monitoring over Southwest China. International Journal of Applied Earth Observation and Geoinformation, 35, 270-283

Hondula, D.M., Davis, R.E., Leisten M.J., Saha, M.V., Veazey, L.M., & Wegner, C.R. (2012). Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983–2008: a case-series analysis. Environmental Health, 11, p. 16

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. Advances in Atmospheric Sciences, 34, 650-662

Huang, R., Zhang, C., Huang, J., Zhu, D., Wang, L., & Liu, J. (2015). Mapping of Daily Mean Air Temperature in Agricultural Regions Using Daytime and Nighttime Land Surface Temperatures Derived from TERRA and AQUA MODIS Data. Remote Sensing, 7, 8728-8756

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. Remote Sensing of Environment, 178, 127-141

Imhoff, M.L., Zhang, P., Wolfe, R.E., & Bounoua, L. (2010). Remote sensing of the urban heat island effect across biomes in the continental USA. Remote Sensing of Environment, 114, 504-513

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. International Journal of Climatology, 37, 1181-1194

Jang, J.D., Viau, A.A., & Anctil, F. (2004). Neural network estimation of air temperatures from AVHRR data. International Journal of Remote Sensing, 25, 4541-4554

Jang, K., Kang, S., Kimball, J., & Hong, S. (2014). Retrievals of All-Weather Daily Air Temperature Using MODIS and AMSR-E Data. Remote Sensing, 6, 8387-8404

Jin, M., Dickinson, R.E., & Vogelmann, A.M. (1997). A comparison of CCM2–BATS skin temperature and surface-Air temperature with satellite and surface observations. Journal of Climate, 10, 1505–1524.

Johnson, D.M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sensing of Environment, 141, 116-128

Jung, C., Lee, Y., Cho, Y., & Kim, S. (2017). A Study of Spatial Soil Moisture Estimation Using a Multiple Linear Regression Model and MODIS Land Surface Temperature Data Corrected by Conditional Merging. Remote Sensing, 9, 870

Khandelwal, S., Goyal, R., Kaul, N., & Mathew, A. (2018). Assessment of land surface temperature variation due to change in elevation of area surrounding Jaipur, India. The Egyptian Journal of Remote Sensing and Space Science, 21(1), 87-94

Kimura, R., Moriyama, M., & Bandou, S. (2015). The relationship between land-surface temperature and rice quality in Tottori prefecture, Japan. International Journal of Remote Sensing, 36, 5690-5706

Kitsara, G., Papaioannou, G., Retalis, A., Paronis, D., & Kerkides, P. (2017). Estimation of air temperature and reference evapotranspiration using MODIS land surface temperature over Greece. International Journal of Remote Sensing, 39, 924-948

Kloog, I., Chudnovsky, A., Koutrakis, P., & Schwartz, J. (2012). Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. Science of the Total Environment, 432, 85-92

Kloog, I., Nordio, F., Coull, B.A., & Schwartz, J. (2014). Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. Remote Sensing of Environment, 150, 132-139

Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., & Schwartz, J. (2017). Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. International Journal of Climatology, 37, 296-304

Krüger, E., & Emmanuel, R. (2013). Accounting for atmospheric stability conditions in urban heat island studies: The case of Glasgow, UK. Landscape and Urban Planning, 117, 112-121

Lakshmi, V., Czajkowski, K., Dubayah, R., & Susskind, J. (2001). Land surface air temperature mapping using TOVS and AVHRR. International Journal of Remote Sensing, 22, 643-662

Li, J., & Heap, A. (2008). A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, p. 137

Li, J., Li, J., Weisz, E., & Zhou, D.K. (2007). Physical retrieval of surface emissivity spectrum from hyperspectral infrared radiances. Geophysical Research Letters, 34

Li, Q., Zhang, H., Liu, X., & Huang, J. (2004). Urban heat island effect on annual mean temperature during the last 50 years in China. Theoretical and Applied Climatology, 79, 165-174

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., & Sobrino, J.A. (2013). Satellite-derived land surface temperature: Current status and perspectives. Remote Sensing of Environment, 131, 14-37

Lin, S., Moore, N.J., Messina, J.P., DeVisser, M.H., & Wu, J. (2012). Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. International Journal of Applied Earth Observation and Geoinformation, 18, 128-140

Lin, X., Zhang, W., Huang, Y., Sun, W., Han, P., Yu, L., & Sun, F. (2016). Empirical Estimation of Near-Surface Air Temperature in China from MODIS LST Data by Considering Physiographic Features. Remote Sensing, 8, 629

Liu, L., Breitner, S., Pan X., et al. (2011). Associations between air temperature and cardio-respiratory mortality in the urban area of Beijing, China: a time-series analysis. Environmental health, 10151

Liu, X., Cheng, Z., Yan, L., & Yin, Z. (2009). Elevation dependency of recent and future minimum surface air temperature trends in the Tibetan Plateau and its surroundings. Global Planet Change, 68, 164–174

Lofgren, B.M., Hunter, T.S., Wilbarger, J. (2011). Effects of using air temperature as a proxy for evapotranspiration in climate change scenarios of Great Lakes basin hydrology. Journal of Great Lakes Research, 37, 744–752.

Mallick, K., Bhattacharya, B.K., & Patel, N.K. (2009). Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI. Agricultural and Forest Meteorology, 149, 1327-1342

Meteotest, 2003. Meteonorm handbook, Part III: Theory Part 2, http://www.meteotest.ch

Metz, M., Andreo, V., & Neteler, M. (2017). A New Fully Gap-Free Time Series of Land Surface Temperature from MODIS LST Data. Remote Sensing, 9, 1333

Meyer, H., Katurji, M., Appelhans, T., Müller, M., Nauss, T., Roudier, P., & Zawar-Reza, P. (2016). Mapping Daily Air Temperature for Antarctica Based on MODIS LST. Remote Sensing, 8, 732

Mildrexler, D., Yang, Z., Cohen, W.B., & Bell, D.M. (2016). A forest vulnerability index based on drought and high temperatures. Remote Sensing of Environment, 173, 314-325

Miles, V.E., Igor (2017). Seasonal and Spatial Characteristics of Urban Heat Islands (UHIs) in Northern West Siberian Cities. Remote Sensing, 9, 989

Mostovoy, G.V., King, R.L., Reddy, K.R., Kakani, V.G., & Filippova, M.G. (2006). Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi. GIScience & Remote Sensing, 43, 78-110

Mukherjee, S., Joshi, P.K., & Garg, R.D. (2014). A comparison of different regression models for downscaling Landsat and MODIS land surface temperature images over heterogeneous landscape. Advances in Space Research, 54, 655-669

Mukherjee, S., Joshi, P.K., & Garg, R.D. (2016). Analysis of urban built-up areas and surface urban heat island using downscaled MODIS derived land surface temperature data. Geocarto International, 32, 900-918

Muller, C.L., Chapman, L., Grimmond, C.S.B., Young, D.T., & Cai, X. (2013). Sensors and the city: a review of urban meteorological networks. International Journal of Climatology, 33, 1585-1600

Mutiibwa, D., Strachan, S., & Albright, T. (2015). Land Surface Temperature and Surface Air Temperature in Complex Terrain. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(10), 4762-4774.

Nemani, R.R., Pierce, L.; Running, S.W. (1993). Developing satellite-derived estimates of surface moisture status. Journal of Applied Meteorology and Climatology, 32, 548–557.

Nemani, R.R., Running, S.W. (1989). Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR Data. Journal of Applied Meteorology and Climatology, 28, 276–284.

Neteler, M. (2010). Estimating Daily Land Surface Temperatures in Mountainous Environments by Reconstructed MODIS LST Data. Remote Sensing, 2, 333-351

Niclos, R., Valiente, J.A., Barbera, M.J., & Caselles, V. (2014). Land surface air temperature retrieval from EOS-MODIS images. IEEE Geoscience of Remote Sensing Letter, 11, 1380–1384.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., & Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. Remote Sensing of Environment, 115, 107-116

Noi, P., 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

Noi, P., Kappas, M., & Degener, J. (2016). Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam. Remote Sensing, 8, 1002

Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agricultural and Forest Meteorology, 216, 157-169

Parmentier, B., McGill, B., Wilson, A.M., Regetz, J., Jetz, W., Guralnick, R.P., Tuanmu, M.N., Robinson, N., & Schildhauer, M. (2014). An assessment of methods and remote-sensing derived

covariates for regional predictions of 1 km daily maximum air temperature. Remote Sensing, 6, 8639-8670

Phan, T.N., Kappas, M., & Tran, T.P. (2018). Land Surface Temperature Variation Due to Changes in Elevation in Northwest Vietnam. Climate, 6, 28.

Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y.H., & Walker, J. (2011). Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. IEEE Transactions on Geoscience and Remote Sensing, 49, 3156-3166

Olivera-Guerra, L., Mattar, C., Merlin, O., Durán-Alarcón, C., Santamaría-Artigas, A., & Fuster, R. (2017). An operational method for the disaggregation of land surface temperature to estimate actual evapotranspiration in the arid region of Chile. ISPRS Journal of Photogrammetry and Remote Sensing, 128, 170-181

Oyler, J. W., Ballantyne, A., Jencso, K., Sweet, M., & Running, S.W. (2015). Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. International Journal of Climatology, 35, 2258–2279.

Oyler, J.W., Dobrowski, S.Z., Holden, Z.A., & Running, S.W. (2016). Remotely Sensed Land Skin Temperature as a Spatial Predictor of Air Temperature across the Conterminous United States. Journal of Applied Meteorology and Climatology, 55, 1441-1457

Panayi, E., Peters, G.W., & Kyriakides, G. (2017). Statistical modelling for precision agriculture: A case study in optimal environmental schedules for Agaricus Bisporus production via variable domain functional regression. PLoS ONE 12(9): e0181921.

Panda, S.K., Choudhury, S., Saraf, A.K., & Das, J.D. (2007). MODIS land surface temperature data detects thermal anomaly preceding 8 October 2005 Kashmir earthquake. International Journal of Remote Sensing, 28, 4587-4596

Pape, R., & Löffler, J. (2004). Modelling spatio-temporal near-surface temperature variation in high mountain landscapes. Ecological Modelling, 178, 483-501

Peón, J., Recondo, C., & Calleja, J.F. (2014). Improvements in the estimation of daily minimum air temperature in peninsular Spain using MODIS land surface temperature. International Journal of Remote Sensing, 35, 5148-5166

Prihodko, L., & Goward, S.N. (1997). Estimation of air temperature from remotely sensed surface observations. Remote Sensing of Environment, 60, 335–346.

Rhee, J., Im, J., & Carbone, G.J. (2010). Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. Remote Sensing of Environment, 114, 2875-2887

Riddering, J.P., Queen, L.P. (2006). Estimating near-surface air temperature with NOAA AVHRR. Canadian Journal of Remote Sensing, 32, 33–43

Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C., & Kloog, I. (2017). Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. Environmental Research, 159, 297-312.

Rosenzweig, C., Elliott, J., Deryng, D., et al. (2014). Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Inter-comparison. Proc. Natl. Acad. Sci. USA, 111, 3268–3273.

Sandholt, I., Rasmussen, K., & Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sensing of Environment, 79, 213–224.

Sánchez, N. et al. (2017). Preliminary assessment of an integrated SMOS and MODIS application for global agricultural drought monitoring. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 2000-2003.

Schuster, C., Burkart, K., & Lakes, T. (2014). Heat mortality in Berlin – Spatial variability at the neighborhood scale. Urban Climate, 10, 134-147

Schwarz, N., Lautenbach, S., & Seppelt, R. (2011). Exploring indicators for quantifying surface urban heat islands of European cities with MODIS land surface temperatures. Remote Sensing of Environment, 115, 3175-3186

Shah, D.B., Pandya, M.R., Trivedi, H.J., & Jani, A.R. (2013). Estimating minimum and maximum air temperature using MODIS data over Indo-Gangetic Plain. Journal of Earth System Science, 122, 1593–1605.

Shah, M. & Jin, S. (2016). Pre-seismic thermal anomalies of the 2015 Mw = 7:8 Gorkha (Nepal) earthquake from MODIS surface temperature. Progress in Electromagnetic Research Symposium (PIERS), Shanghai, 2016, 4755-4758.

Shao, Q., Sun, C., Liu, J., He, J., Kuang, W., & Tao, F. (2011). Impact of urban expansion on meteorological observation data and overestimation to regional air temperature in China. Journal of Geographical Sciences, 21, 994-1006

Shen, S., & Leptoukh, G.G. (2011). Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environmental Research Letters, 6, 045206

Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., & Schwartz, J. (2016). Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. Environmental Research, 146, 51-58

Shi, Y., Jiang, Z., Dong, L., & Shen, S. (2017). Statistical estimation of high-resolution surface air temperature from MODIS over the Yangtze River Delta, China. Journal of Meteorological Research, 31, 448-454

Smith, W.L., Leslie, L.M., Diak, G.R., Goodman, B.M., Velden, C.S., Callan, G.M., Raymond, W., & Wade, G.S. (1988). The integration of meteorological satellite imagery and numerical dynamical forecast models. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 324, 317–323.

Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y., & Minh, V.Q. (2012). Monitoring agricultural drought in the lower mekong basin using MODIS NDVI and land surface temperature data. International Journal of Applied Earth Observation and Geoinformation, 18, 417-427

Song, K., Wang, M., Du, J., Yuan, Y., Ma, J., Wang, M., & Mu, G. (2016). Spatiotemporal variations of lake surface temperature across the Tibetan Plateau using MODIS LST product. Remote Sensing, 8

Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G., & McKendry, I.G. (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and Forest Meteorology, 139, 224-236

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., & Eklundh, L. (2007). Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sensing of Environment, 110, 262-274

Sun, D., & Kafatos, M. (2007). Note on the NDVI-LST relationship and the use of temperature-related drought indices over North America. Geophysical Research Letters, 34

Sun, H., Chen, Y., Gong, A., Zhao, X., Zhan, W., & Wang, M. (2013). Estimating mean air temperature using MODIS day and night land surface temperatures. Theoretical and Applied Climatology, 118, 81-92

Sun, Y.J., Wang, J.F., Zhang, R.H., Gillies, R.R., Xue, Y., & Bo, Y.C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. Theoretical and Applied Climatology, 80, 37-48

Sun, Z., Wang, Q., Matsushita, B., Fukushima, T., Ouyang, Z., Watanabe, M., & Gebremichael, M. (2011). Evaluation of the VI–Tsmethod for estimating the land surface moisture index and air temperature using ASTER and MODIS data in the North China Plain. International Journal of Remote Sensing, 32, 7257-7278

Tomlinson, C.J., Chapman, L., Thornes, J.E., & Baker, C. (2011). Remote sensing land surface temperature for meteorology and climatology: a review. Meteorological Applications, 18, 296-306

Urban, M., Eberle, J., Hüttich, C., Schmullius, C., & Herold, M. (2013). Comparison of Satellite-Derived Land Surface Temperature and Air Temperature from Meteorological Stations on the Pan-Arctic Scale. Remote Sensing, 5, 2348-2367

Vallet-Coulomb, C.; Legesse, D.; Gasse, F.; Travi, Y.; Chernet, T. (2001). Lake evaporation estimates in tropical Africa (Lake Ziway, Ethiopia). Journal of Hydrology, 245, 1–18.

Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S.J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114, 449-465

Wan, Z. (2008). New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. Remote Sensing of Environment, 112, 59-74

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(4), 892-905.

Wan, Z., & Li, Z.L. (1997). A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. IEEE Transactions on Geoscience and Remote Sensing, 35(4), 980-996.

Wang, K., Li, Z., & Cribb, M. (2006). Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley–Taylor parameter. Remote Sensing of Environment, 102, 293-305

Wang, M., He, G., Zhang, Z., Wang, G., Zhang, Z., Cao, X., Wu, Z., & Liu, X. (2017). Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China. Remote Sensing, 9

Weiss, D.J., Bhatt, S., Mappin, B., Van Boeckel, T.P., Smith, D.L., Hay, S.I., & Gething, P.W. (2014). Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: A high-resolution spatiotemporal prediction. Malaria Journal, 13

Wenbin, Z., Aifeng, L., Shaofeng, J. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sensing of Environment, 130, 62-73

Westermann, S., Langer, M., & Boike, J. (2012). Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway. Remote Sensing of Environment, 118, 162-167

Wilson, K., Goldstein, A., Falge, E., et al. (2002). Energy balance closure at FLUXNET sites. Agricultural and Forest Meteorology, 113, 223–243.

Wu, C., Munger, J.W., Niu, Z., & Kuang, D. (2010). Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. Remote Sensing of Environment, 114, 2925-2939

Wu, D., Qu, J.J., & Hao, X. (2015). Agricultural drought monitoring using MODIS-based drought indices over the USA Corn Belt. International Journal of Remote Sensing, 36, 5403-5425

Xu, Y., Knudby, A., & Ho, H.C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. International Journal of Remote Sensing, 35, 8108-8121

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

Xu, Y., Qin, Z., & Shen, Y. (2012). Study on the estimation of near-surface air temperature from MODIS data by statistical methods. International Journal of Remote Sensing, 33, 7629-7643

Yang, Y., Cai, W., & Yang, J. (2017). Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China. Remote Sensing, 9, 410

Yoo, C., Im, J., Park, S., & Quackenbush, L.J. (2018). Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. ISPRS Journal of Photogrammetry and Remote Sensing, 137, 149-162

You, Q., Kang, S., Pepin, N., & Yan, Y. (2008). Relationship between trends in temperature extremes and elevation in the eastern and central Tibetan Plateau, 1961–2005. Geophysical Research Letters, 35, L14704

Zakšek, K., & Schroedter-Homscheidt, M. (2009). Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. ISPRS Journal of Photogrammetry and Remote Sensing, 64, 414-421

Zeng, L., Wardlow, B., Tadesse, T., Shan, J., Hayes, M., Li, D., & Xiang, D. (2015). Estimation of Daily Air Temperature Based on MODIS Land Surface Temperature Products over the Corn Belt in the US. Remote Sensing, 7, 951-970

Zhang, H., Zhang, F., Zhang, G., He, X., & Tian, L. (2016a). Evaluation of cloud effects on air temperature estimation using MODIS LST based on ground measurements over the Tibetan Plateau. Atmospheric Chemistry and Physics, 16, 13681-13696

Zhang, H., F. Zhang, M. Ye, T. Che, and, & Zhang, G. (2016b). Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data Journal of Geophysical Research: Atmospheres, 121, 1,425–411,441,

Zhang, H., Zhang, F.A.N., Zhang, G., Ma, Y., Yang, K.U.N., & Ye, M. (2018). Daily air temperature estimation on glacier surfaces in the Tibetan Plateau using MODIS LST data. Journal of Glaciology, 64, 132-147

Zhang, J., Hu, Y., Xiao, X., Chen, P., Han, S., Song, G., & Yu, G. (2009). Satellite-based estimation of evapotranspiration of an old-growth temperate mixed forest. Agricultural and Forest Meteorology, 149, 976-984

Zhang, L., Jiao, W., Zhang, H., Huang, C., & Tong, Q. (2017). Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices. Remote Sensing of Environment, 190, 96-106

Zhang, R.; Rong, Y.; Tian, J.; Su, H.; Li, Z.-L.; Liu, S. (2015). A Remote Sensing Method for Estimating Surface Air Temperature and Surface Vapor Pressure on a Regional Scale. Remote Sensing, 7, 6005-6025

Zhang, W., Huang, Y., Yu, Y., & Sun, W. (2011). Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures. International Journal of Remote Sensing, 32, 9415-9440

Zhang, Y., Jiang, Z., Huang H., et al. (2014). Thermal anomalies detection before 2013 Songyuan earthquake using MODIS LST data. 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, 2014, 3009-3012.

Zheng, X., Zhu, J., & Yan, Q. (2013). Monthly Air Temperatures over Northern China Estimated by Integrating MODIS Data with GIS Techniques. Journal of Applied Meteorology and Climatology, 52, 1987-2000

Zhou, W., Peng, B., Shi, J., Wang, T., Dhital, Y., Yao, R., Yu, Y., Lei, Z., & Zhao, R. (2017). Estimating High Resolution Daily Air Temperature Based on Remote Sensing Products and

Climate Reanalysis Datasets over Glacierized Basins: A Case Study in the Langtang Valley, Nepal. Remote Sensing, 9, 959

Zhu, W., Lű, A., Jia, S., Yan, J., & Mahmood, R. (2017). Retrievals of all-weather daytime air temperature from MODIS products. Remote Sensing of Environment, 189, 152-163

Zoran, M. (2012). MODIS and NOAA-AVHRR l and surface temperature data detect a thermal anomaly preceding the 11 March 2011 Tohoku earthquake. International Journal of Remote Sensing, 33, 6805-6817

Chapter 3

This chapter is based on the published paper: Noi, P.T.; Kappas, M.; Tran, T.P. **Land Surface Temperature Variation Due to Changes in Elevation in Northwest Vietnam**. *Climate* **2018**, 6, 28.

Abstract

Land surface temperature (LST) is one of the most important variables for applications relating to the physics of land surface processes. LST rapidly changes in both space and time, and knowledge of LST and its spatiotemporal variation is essential to understand the interactions between human activity and the environment. This study investigates the spatiotemporal variation of LST according to changes in elevation. The newest version (version 6) of MODIS LST data for 2015 was used. An area of 40,000 km2 (200 × 200 km2) in northwest Vietnam with elevations ranging from 8 m to 3165 m was chosen as a case study. Our results showed that the drop in LST with increased elevation varied throughout the year during both the daytime and nighttime. The monthly averages in 2015 and an altitude increase of 1000 m resulted in a decrease in LST ranging from 3.8 °C to 6.1 °C and 1.5 °C to 5.8 °C for the daytime and nighttime, respectively. This suggests that in any study relating to the spatial distribution of LST, the effect of elevation on LST should be considered. In addition, the effects of land use/cover and elevation distribution on the relationship between LST and elevation are discussed.

3. 1. Introduction

Land surface temperature (LST), defined as the "skin" temperature of the earth surface, is a key parameter for land-surface processing studies at both a regional and global scale (Anderson et al., 2008; Karnieli et al., 2010; Khandelwal et al., 2017; Duan et al., 2014a). Knowledge of LST and its spatiotemporal variation within regions is of fundamental importance to understanding the interactions between human activity (e.g., change in land use) and the environment (Stathopoulou and Cartalis, 2009). As the LST varies both spatially and temporally, ground-based station observations cannot represent the LST of a region, particularly in developing countries where weather stations are very sparse. Currently, remote sensing satellite data is the most suitable way to study LST spatial and temporal variations (Li et al., 2013; Shwetha and Kumar, 2016; Duan et al., 2014b). The most popular remote sensing images used to retrieve LST are moderate resolution imaging spectroradiometer (MODIS), advanced very high resolution radiometer (AVHRR), enhanced thematic mapper plus (ETM+), and the advanced spaceborne thermal emission and reflection radiometer (ASTER). Due to the high temporal resolutions (four times per day) of MODIS LST, its data is widely applied in remote sensing communities.

MODIS LST data has been available since early 2000 and mid-2002 from Terra and Aqua satellites, respectively. There are seven MODIS LST products with temporal resolutions of daily (MOD/MYD11A1, MOD/MYD11B1, MOD/MYD11C1, and MOD/MYD11_L2), eight-day (MOD/MYD11A2, MOD/MYD11C2), and monthly (MOD/MYD11C3) data. Each product consists of daytime and nighttime LST. Thus, by using both the Terra and Aqua daily products, the highest temporal resolution (i.e., four times per day) can be achieved. In August 2015, version 6 (V6) of MODIS LST was released and free to download (LP DAAC). Wan (2014) studied and evaluated MODIS LST version 6 (using Aqua MODIS daily swath data). The results showed that the accuracy was superior to the previous versions (i.e., V4.1 and V5).

Since available, MODIS LST data has been used for various studies, such as evaluating and monitoring urban heat islands (Rajasekar and Weng, 2009; Keramitsoglou et al., 2011; Yanev and Filchev, 2016; Gawuc and Struzewska, 2016; Miles and Esau, 2017), estimating air surface temperatures (Ta) (Noi et al., 2016; Meyer et al., 2016; Yang et al., 2017; Noi et al., 2017), retrieving soil moisture (Jung et al., 2017; Chen et al., 2017), drought assessment (Sánchez et al., 2016), and hydrology applications (Parinussa et al., 2016). Most of these studies have shown that the changes of land surface properties (e.g., normalized difference

vegetation index (NDVI), elevation, and land use/cover types) will result in the variations of LST (Khandelwal et al., 2017). Among these, the relationship between LST and NDVI or the effects of land use/cover on LST have been widely studied (Yanev and Filchev, 2016; Shah et al., 2013; Sun et al., 2014). A few studies have investigated the direct effect of elevation on LST; however, elevation was found to have an impact on most studies using MODIS LST, particularly those conducted over a large area where the terrain is variable. For example, in Ta estimation using MODIS LST data, along with LST, elevation was considered one of the most impactful variables effecting the results of Ta estimation (Gawuc and Struzewska, 2016; Chen et al., 2016; Huang et al., 2017).

It is known that for a stationary atmosphere, an increase in elevation leads to a subsequent decrease in air pressure and Ta. This decrease is referred to as the environmental lapse rate and occurs at any location along the vertical direction (Khandelwal et al., 2017). For every 1000 m increase, the environmental lapse rate varies between 5 °C to 10 °C, depending on the moisture condition (Khandelwal et al., 2017). According to El Kenawy et al. (2009), different regions show a different variability of Ta, as each region has a unique terrain (Hu et al., 2003; McElwain and Sweeney, 2007; Limsakul and Goes, 2008). Moreover, the variation of Ta was not consistent between high and low land areas. Beniston (2003) found that Ta increased more rapidly at higher than at lower elevations. Being consistent with Beniston (2003), some high mountainous areas in the Tibetan Plateau showed a more rapid change in temperature than in lower regions (You et al., 2008). Various studies have investigated the relationship between Ta and elevation. The majority of these studies used point data from weather stations located 1.5 m to 2.0 m above the land surface (You et al., 2008; Liu et al., 2009; Vancutsem et al., 2010). LST and Ta are not equivalent, and their relationship is multifaceted due to the complex surface energy budget and the multiple related variables between them (Meyer et al., 2016; Yang et al., 2017; Noi et al., 2017; Lin et al., 2016). Therefore, the relationship between Ta and elevation is different from those of LST and elevation. As mentioned above, the effect of land surface characteristics (e.g., NDVI and soil moisture) on LST has been studied (Khandelwal et al., 2017); meanwhile, the effects of elevation on LST are rarely documented. To the best of our knowledge, few studies have investigated the direct relationship between LST and elevation using remote sensing LST data. Recently, Khandelwal et al. (2017) assessed the variation of LST with changes of elevation in Jaipur, India. However, the study used only nighttime MODIS Aqua LST for three seasons with a narrow elevation range (215 m to 530 m). Furthermore, no studies from Vietnam have investigated the relationship between LST and elevation using remote sensing

image data, particularly MODIS LST data. It is therefore necessary to assess the spatiotemporal variations of LST with changes of elevation in northern Vietnam using the MODIS LST V6 data.

The main objectives of this study were to: (i) evaluate and assess the relationship between monthly LST and elevation for 2015 in northern Vietnam, and (ii) discuss the effects of land use/cover, elevation distribution, and latitude and longitude on the relationship between LST and elevation.

3.2. Materials and Methods

3.2.1. Study Area

The study was conducted in a 40,000 km2 (200 × 200 km2) area in northwest Vietnam (Figure 3.1). The elevation ranges from 8 m to 3165 m with the majority of areas below 1000 m and a mean elevation of 800 metres above sea level (m a.s.l). The northwest regions of Vietnam were selected to reduce the influence of the urban heat island (UHI) effect due to the impact of the capital, Hanoi, on the results. Uniform study areas of 200 × 200 km2 were chosen to balance the effects of longitude and latitude.
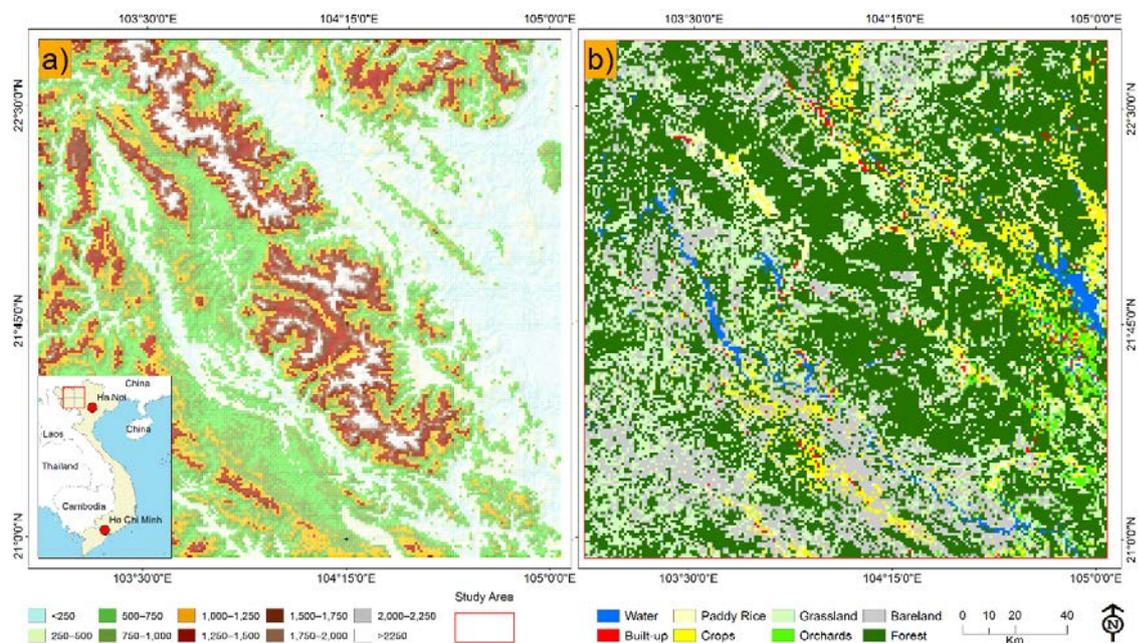
Figure 3.1. Location of the study area in northwest Vietnam. (a) Elevation range from advanced spaceborne thermal emission and reflection radiometer (ASTER) global Digital Elevation Model (DEM) and (b) land cover from Japanese Aerospace Exploration Agency (JAXA) distribution in the study area.

56

3.2.2. Data

3.2.2.1. MODIS LST

MODIS aboard Terra and Aqua satellites provide daily, eight-day, and monthly LST data at resolutions of typically 1 km and 6 km. Wan (2014) studied and evaluated LST V6 data using Aqua MODIS and stated that the accuracy was better than for previous versions (i.e., V4.1 and V5) due to fixes and improvements. The most important improvements were removing cloud-contaminated LST pixels from MODIS level 2 and 3 products and updating split-window algorithm coefficients (Wan, 2014). In this study, we assessed the variation of LST due to changes in elevation using eight-day LST data (daytime and nighttime) from the Aqua satellite MYD11A2. The eight-day MYD11A2 data were achieved by averaging all valid daily LST data under clear sky conditions. The number of daily LST data involved in calculating the eight-day LST data vary from two to eight, depending on data availability (Wan, 2014). There are 12 science data set (SDS) layers within MYD11A2 data, including LST, quality control, view time, view angle, clear sky conditions for day and night, and emissivity bands 31 and 32. For this study, 46 MODIS images (h27v06, V6, 2015) of MYD11A2 product in hierarchical data format (HDF) were downloaded from the Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov).

3.2.2.2. Elevation

The elevation across the study area ranged from 8 m to over 3165 m (Figure 3.1). Elevation data with a spatial resolution of approximately 30 m (1 arc) was retrieved from the U.S. Geological Survey (USGS) ASTER global DEM. To align the elevation data with LST data, it was resampled to a resolution of 1000 m using the nearest neighbor sampling approach in ArcMap 10.5 software.

3.2.2.3. Land Cover

Land use and land cover data were downloaded from the Japanese Aerospace Exploration Agency (JAXA) (http://www.eorc.jaxa.jp). Data was available for spatial resolutions of 15 m and 250 m for 2007 and 2015. The 2015 land use/cover map had a high overall accuracy (89.1%) with a κ coefficient of 0.872. It included nine different land use/cover classes, with the following eight classes found within the study area: water, urban and built-up, rice paddy, crops, grassland, orchards, bare land, and forest (Figure 3.1b). To be consistent with other data, the map was resampled to 1000 m using the nearest neighbor sampling approach in ArcMap 10.5 software.

3.2.3. Preprocessing and Data Analysis

The MYD11A2 data was downloaded from the LP DAAC website in HDF format using the sinusoidal projection system (https://lpdaac.usgs.gov). MODIS reprojection tools provided by LP DAAC were used to reproject MYD11A2 product to the UTM Zone 48 N projection system with WGS84 datum. Among the 12 science data set layers of MYD11A2, LST (day and night) and the quality control (QC) layer were selected and reformatted to GeoTIFF for further analysis. ArcMap 10.5 was used to analyze the data in GeoTIFF format.

MODIS LST and QC data in GeoTIFF with a resolution of 926.65 m were resampled to 1000 m using the nearest neighbor resampling approach. The images were then clipped to the study area extent using ArcMap 10.5 software.

Cloudy pixels from MODIS LST were removed by the generalized split-window algorithm (Wan and Dozier, 1996); however, there were still pixels associated with thin clouds which the algorithm was unable to detect and remove. The quality information (QC file) of each image was therefore used for removing such data, and only LST data with an average error below or equal to 2 K were used for further analysis.

All valid LST pixels (in digital number—DN value) were converted to Celsius (°C) using the following equation:

$$\text{LST (°C)} = 0.02 \times \text{DN} - 273.15 \tag{1}$$

where °C is the temperature in Celsius and 0.02 is the scale factor of the MODIS LST product (https://lpdaac.usgs.gov).

3.2.4. Relationship between LST and Elevation

The spatial and temporal relationship between monthly average LST (°C) and elevation (m) were investigated. Based on the Julian calendar of 2015, we divided the 46 eight-day LST images into 12 months. We then applied the same rule of calculating an eight-day LST from a daily LST for calculating the average monthly LST; that is, the monthly LST data was averaged from one to four eight-day LST data. Linear regression models were used to evaluate the relationship between LST and elevation.

3.3. Results

3.3.1. Spatiotemporal Variation of LST

Figure 3.2 shows the nighttime and daytime LST average from January to December in 2015 in northwest Vietnam. In general, the LST ranged from 11 °C to 25 °C during the nighttime and 21 °C to 35 °C during the daytime. It should be noted that, because LST data are not available for a location (pixel) covered by cloud, therefore, there are some blank pixels (white color) in the daytime and nighttime (e.g., in June). For all months, the LST for the highest topographic area—from the northwest to the southeast region—was lower than other areas in the study site. This was clearly seen for both the daytime and nighttime. The lowest elevation area (northeast) always showed the highest nighttime LST. However, the highest daytime LST were observed in the south-center region from March to June. This indicates that not only elevation influences LST, but also other factors, such as land use/cover. As shown in Figure 3.1, this area is mainly covered by bare land and crops; however, crop planting started between April and May, and therefore in the period from March to June, the main land cover was bare land. This is consistent with the study of Xu et al. (2013), which was implemented in the Tibetan Plateau from 2003 to 2010. In this study, Xu et al. (2013) concluded that bare land has the highest mean LST in comparison to four others land cover types: forest, grassland, water, and snow/ice. The daytime and nighttime differences indicate that the topography had a higher influence on the nighttime LST than the daytime LST.
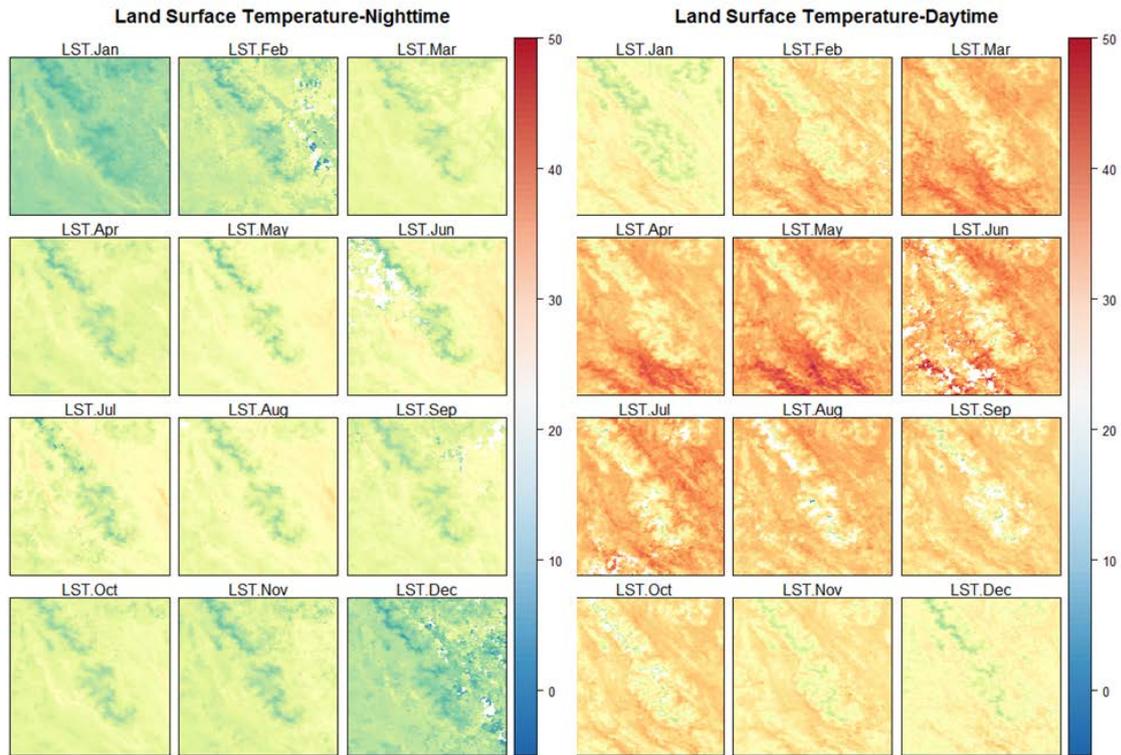
Figure 3.2. The spatial patterns of the monthly average nighttime (left) and daytime (right) land surface temperature (LST) in 2015 in northwest Vietnam.

In addition, the different LST between high and low elevations varied across different months and different times of the day. Figure 3.2 shows that while the difference in LST between high and low elevation was not clear for the daytime from June to October, there was a clear difference at nighttime.

Figure 3.3 shows the temporal variation of monthly LST from January to December in 2015, in which the daytime LST was larger than that of the nighttime. This is consistent with the results shown in Figure 3.2. It should be mentioned that the average monthly LST were calculated from 40,000-pixel values (200 × 200) containing different land covers and elevations. Therefore, the values of LST varied in all months and regions (Figure 3.2 and Figure 3.3). LST increased from January to April, reached the highest in June, and decreased thereafter. June was the hottest month during both the daytime and nighttime, whereas January and December had the lowest daytime and nighttime LST.
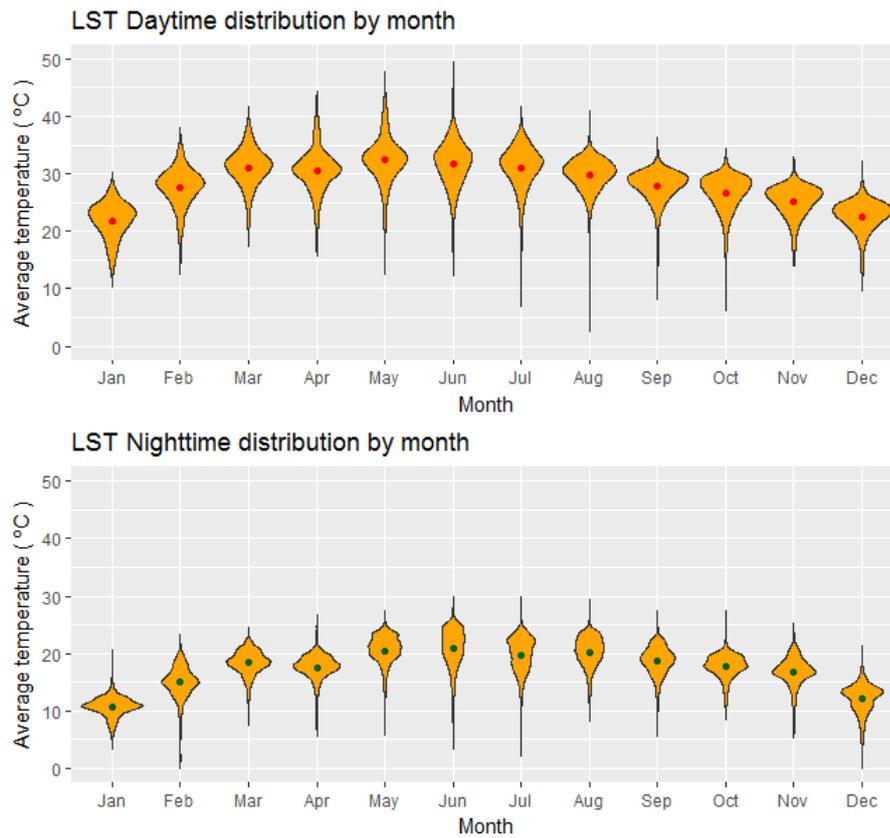
Figure 3.3. Violin plots show the monthly average LST of the daytime (top) and nighttime (bottom) in 2015 in northwest Vietnam.

3.3.2. LST and Elevation

Figure 3.4 shows the linear regression models between averaged monthly daytime (Figure 3.4a) and nighttime (Figure 3.4b) LST and elevations in northwest Vietnam. The negative correlation confirms the well-known relationship between temperature and topography (Stroppiana et al., 2014). As shown in Figure 3.4a,b, the correlation between LST and elevation was stronger at nighttime than daytime, except for four months: February, September, November, and December. This may reflect that several pixels within low regions (0 m to 1000 m) had a very low LST, ranging from 0 °C to 10 °C in December and February and 10 °C to 15 °C in September and November.
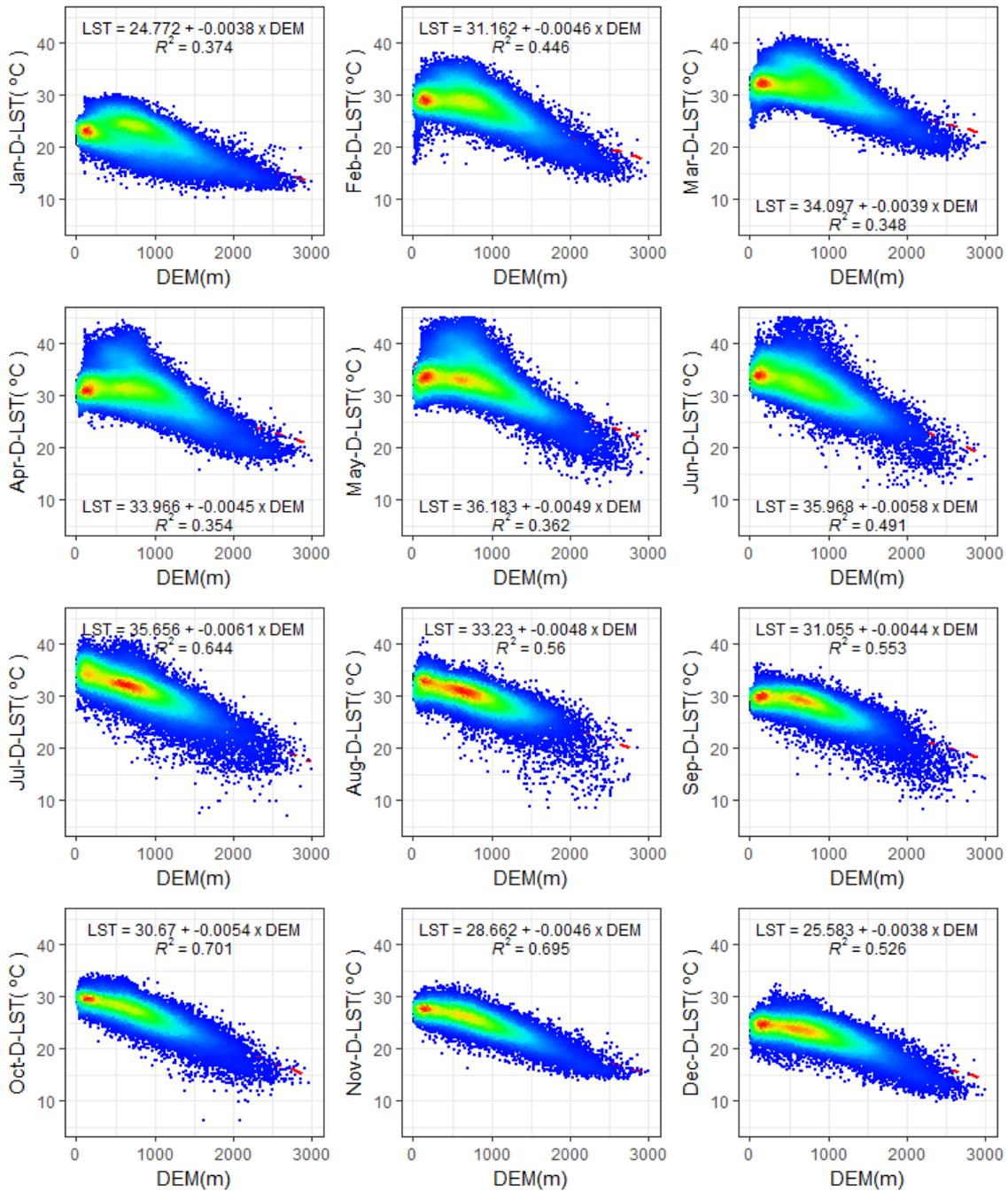
Figure 3.4. (a) Scatter plots showing the relationship between average monthly daytime LST (D-LST) and elevation in northwest Vietnam in 2015.
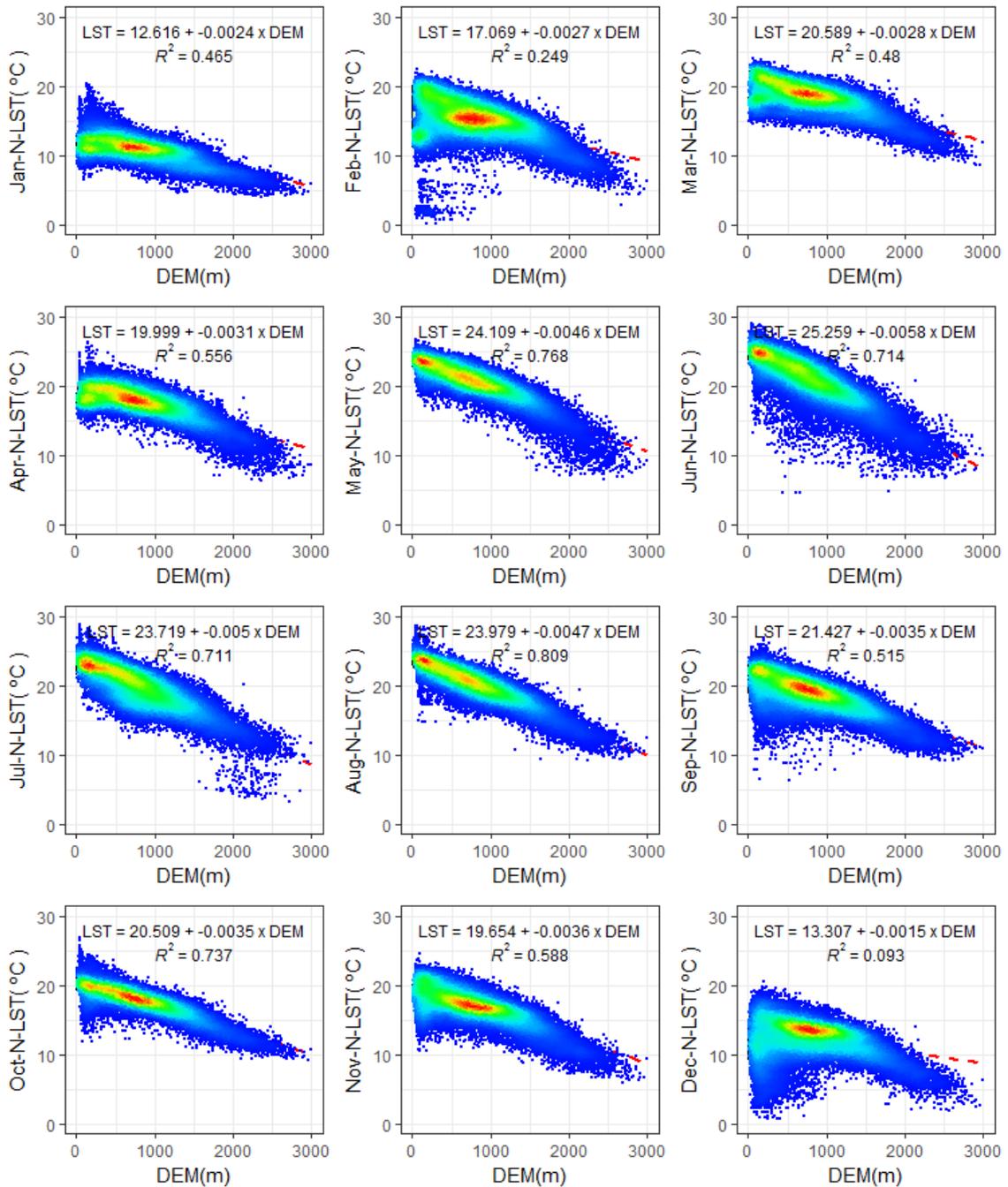
Figure 3.4. (b) Scatter plots showing the relationship between average monthly nighttime LST (N-LST) and elevation in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.

The relationship between LST and elevation not only varied between months, but also between the daytime and nighttime. Figure 3.4a,b shows that there were stronger correlations in hotter months (May to August) during the nighttime, whereas a weaker correlation for the hotter months was seen for the daytime. The highest correlations were observed in October ($R^2 = 0.701$) and in August ($R^2 = 0.809$) at daytime and nighttime, respectively.

Figure 3.4a shows that during the day, the linear regression slope ranges from −0.0038 to −0.0061. This indicates that for each 1000 m elevation increase, the LST decreases from 3.8 °C to 6.1 °C. Based on the R^2 value and the slope, the smallest and largest changes of LST corresponding to the elevation are in January and July, respectively.

The intercept varies from 24.77 to 36.18 (daytime) and 12.62 to 25.26 (nighttime) each month and is slightly higher than the average monthly LST (Figure 3.3) of the respective month. This was consistent across all months in 2015 for the daytime and nighttime, indicating a good relationship between intercept value and average LST value.

For the nighttime (Figure 3.4b), the linear regression slope varied from −0.0015 to −0.0058; that is, for each 1000 m elevation increase, the LST decreased from 1.5 °C to 5.8 °C. The lowest and highest slopes were with LST in December and June, respectively. A strong correlation (R^2 of 0.711 to 0.809) between LST and elevation was observed in May, June, July, August, and October. The correlation in January, March, September, and November ranged from 0.465 to 0.588. However, in February and December, these correlations were very low, with an R^2 of 0.093 and 0.249 for December and February, respectively. Figure 3.4b highlights the many low LST points in February and December, suggesting another local factor may have strongly reduced LST.

Except for the trends in February and December, LST at nighttime was generally more closely distributed along the regression line than that of the daytime LST. This may indicate that solar radiation affects the LST during the daytime.

3.4. Discussion

Although this study focuses on the effects of elevation on LST due to its infrequent documentation in literature, an analysis on the relationship between LST and the most popular vegetation index (normalized difference vegetation index, NDVI) was additionally discussed. The relationship between LST and NDVI is shown in Figure A3 and Figure A4 (Appendix C).

Figure A2 shows the monthly distribution of NDVI in 2015 for the study area. As shown in Figure 3.1, the major land cover of the study area was forest (51.7%); consequently, the NDVI was quite high, ranging from approximately 0.4 to 0.9 with a monthly average from 0.6 to 0.8 (Figure A2).

The correlation between LST and NDVI (Figure A3 and Figure A4) showed similar results with previous studies; on average LST and NDVI displayed a negative correlation, and the coefficient of determination (R^2) of the daytime was slightly higher than those of the nighttime.

In the literature, various studies have shown that land use/cover has a strong impact on LST (Jiang and Tian, 2010; Kayet et al., 2016; Faqe Ibrahim, 2017) and Ta (Katpatal and Kute, 2008; Perugini et al., 2017). Therefore, in this section, we presented further detail on the relationship between LST and elevation in northwest Vietnam and other potential factors that could have an effect on this relationship, such as land cover and elevation distribution.

3.4.1. The Influence of Land Use/Cover

In order to assess the effects of land use/cover on the variation of LST according to elevation, we used the high-resolution land use and land cover map products over northern Vietnam in 2015, obtained from the Japanese Aerospace Exploration Agency.

As shown in Figure 3.5a,b, there were four main land cover types (water, built-up, crops, and orchards) from 8 m to approximately 1000 m. Paddies, grassland, and bare land had a wider elevation range (2000 m to 2200 m). Forests showed the widest elevation range, from 8 m to approximately 3000 m. Figure 3.5 also shows that the number of pixels for each land cover type were different, ranging from 264 pixels to 20,671 pixels. The major distribution of land use/cover across the study area was forest, grassland, and bare land, which accounted for 51.7%, 18.4%, and 16.0%, respectively (Figure 3.5). The remaining five land use/cover types only accounted for 13.9% of the total area. The linear regression results of the major land use/cover types were more consistent with the results shown in Section 3 than for the other five land use/cover types. This indicates that the distribution of land use/cover type had an impact on the relationship between LST and elevation in the area.
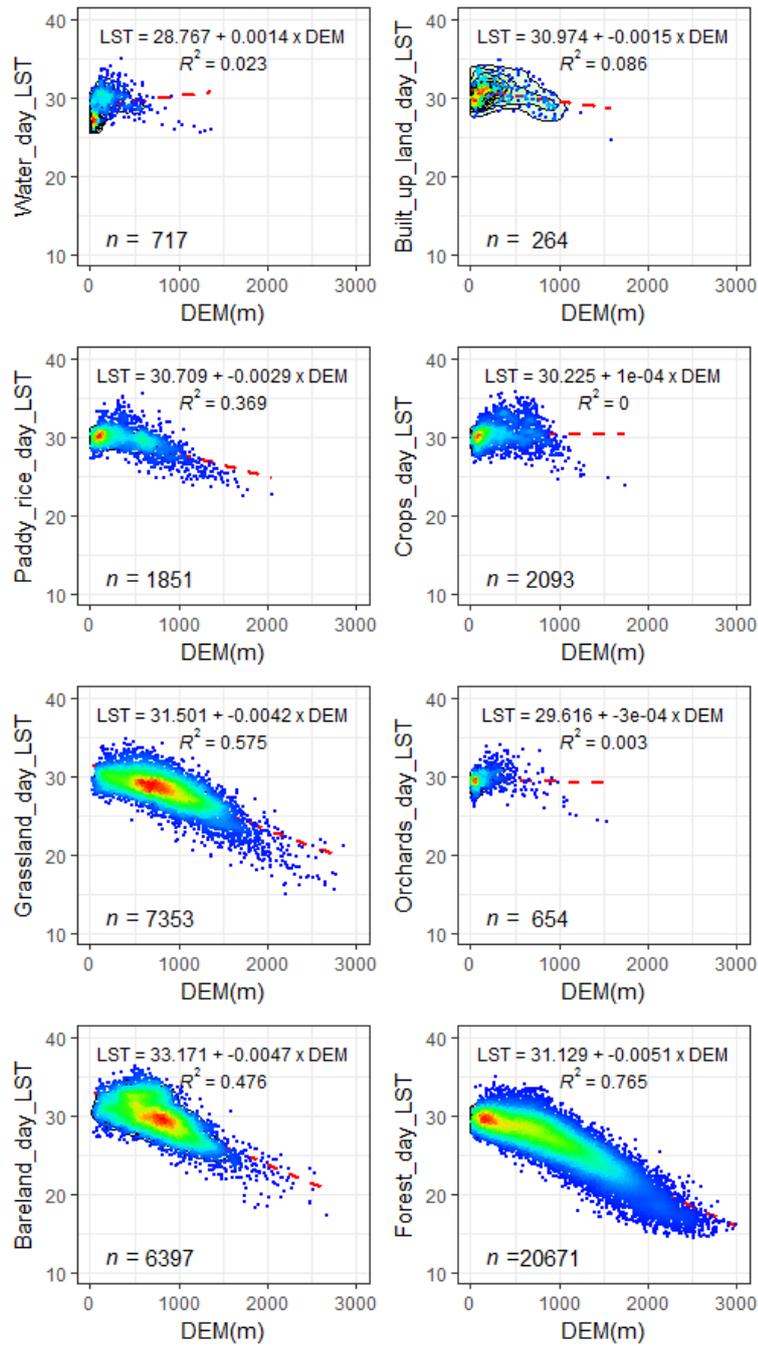
Figure 3.5. (a) Scatter plots showing the relationship between average monthly daytime LST of different land cover types and elevation in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.
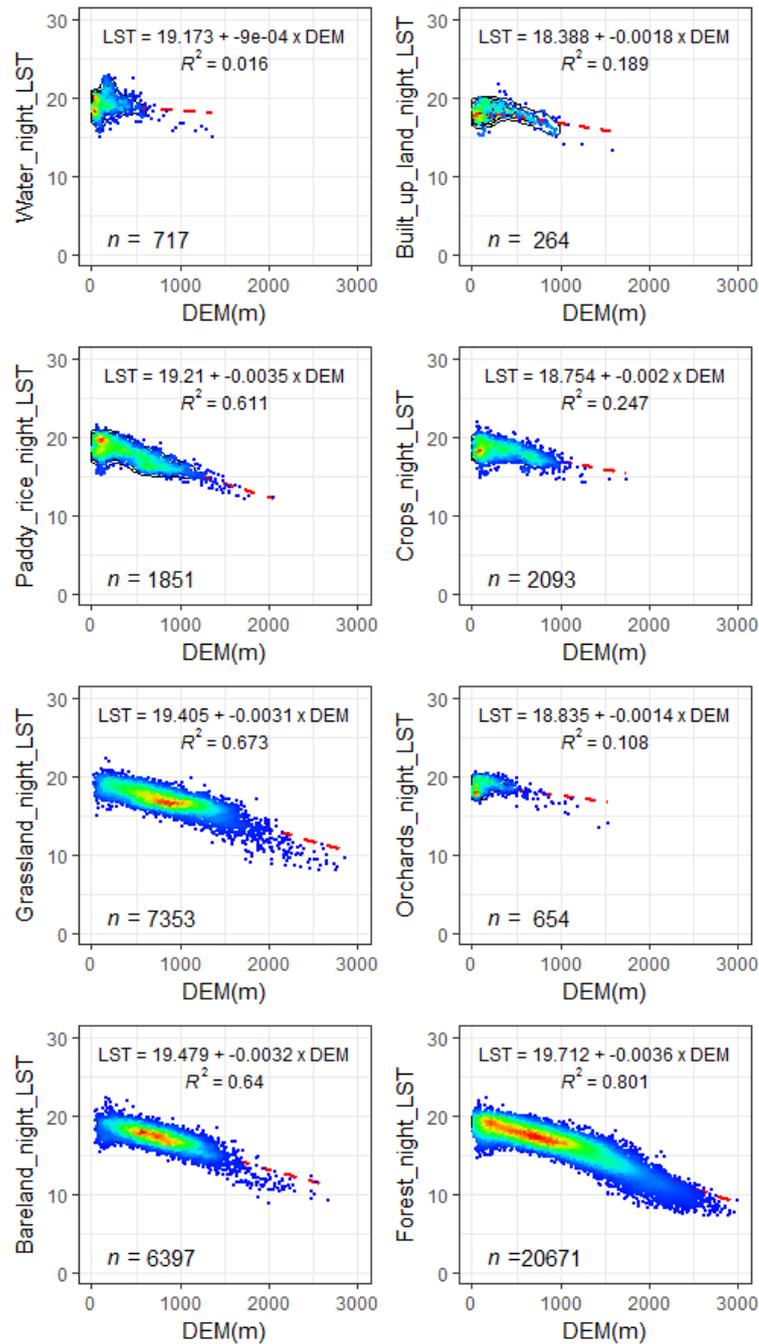
Figure 3.5. (b) Scatter plots showing the relationship between average monthly nighttime LST of different land cover types and elevation in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.

Another important point is that the change of LST and elevation of each land cover type was significantly different between the daytime and nighttime. Here, we focused on the three major land cover types (grassland, bare land, and forest); the proportion of the other land cover types was low with poorly distributed elevations. As shown in Figure 3.5a, when the elevation increased by 1000 m, the LST of grassland, bare land, and forest decreased from

4.2 °C, 4.7 °C, and 5.1 °C, respectively, at daytime. However, at nighttime (Figure 3.5b), these changes were much lower: 3.1 °C, 3.2 °C, and 3.6 °C, respectively.

3.4.2. The Influence of Topographic Distribution

Figure 3.4a,b shows that the majority of observations were located at lower elevations (8 m to 1000 m) and that the spatial distribution was imbalanced across the elevations. Figure A1 (Appendix A) shows the number of pixels for the elevation intervals. It is clearly seen that the number of observations at an elevation from 2000 m to 3000 m was much lower than the number of observations at an elevation from 8 m to 2000 m. Therefore, to assess the influence of elevation distribution, all pixels with an elevation over 2000 m were removed. As shown in Figure A1, the number of pixels distributed from 8 m to 2000 m were imbalanced. Therefore, to assess the influences of LST and elevation, a balanced number of observations at four elevation intervals were used: below 500 m (Level 1), 500 m–1000 m (Level 2), 1000 m–1500 m (Level 3), and 1500 m–2000 m (Level 4). For each elevation interval, 1000 pixels were randomly chosen for analysis.

Figure 3.6a,b shows that the relationship between LST and elevation was much stronger (R^2 ranges from 0.372 to 0.748) for all months at both daytime and nighttime than the results of all data shown in Figure 3.4a,b. The LST dropped from 3.9 °C to 5.9 °C (at daytime) and 2.4 °C to 4.6 °C (at nighttime) for every 1000-m elevation increase.
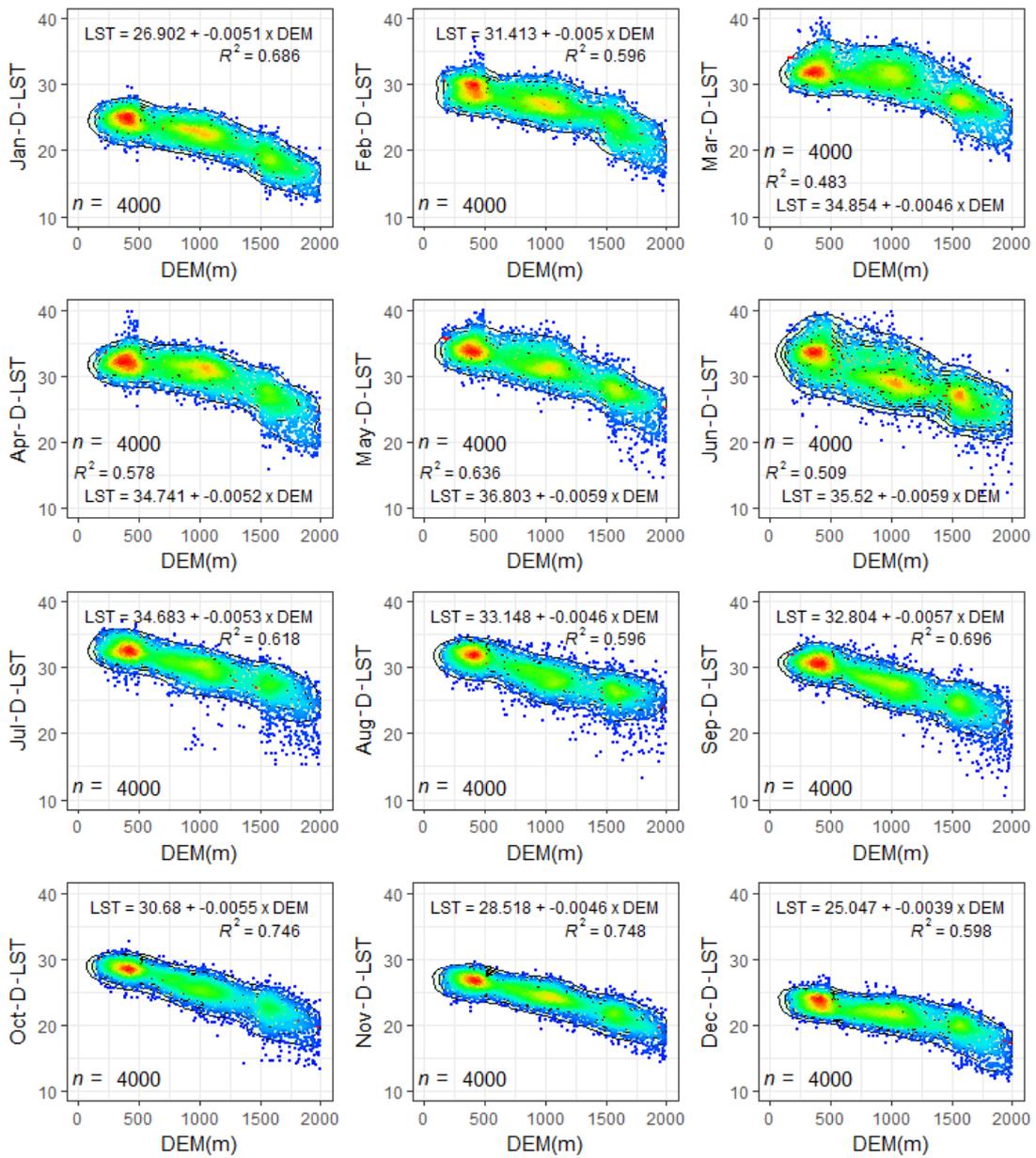
Figure 3.6. (a) Scatter plots showing the relationship between the average monthly daytime LST of a balanced number of observation-based elevations and the elevation in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.
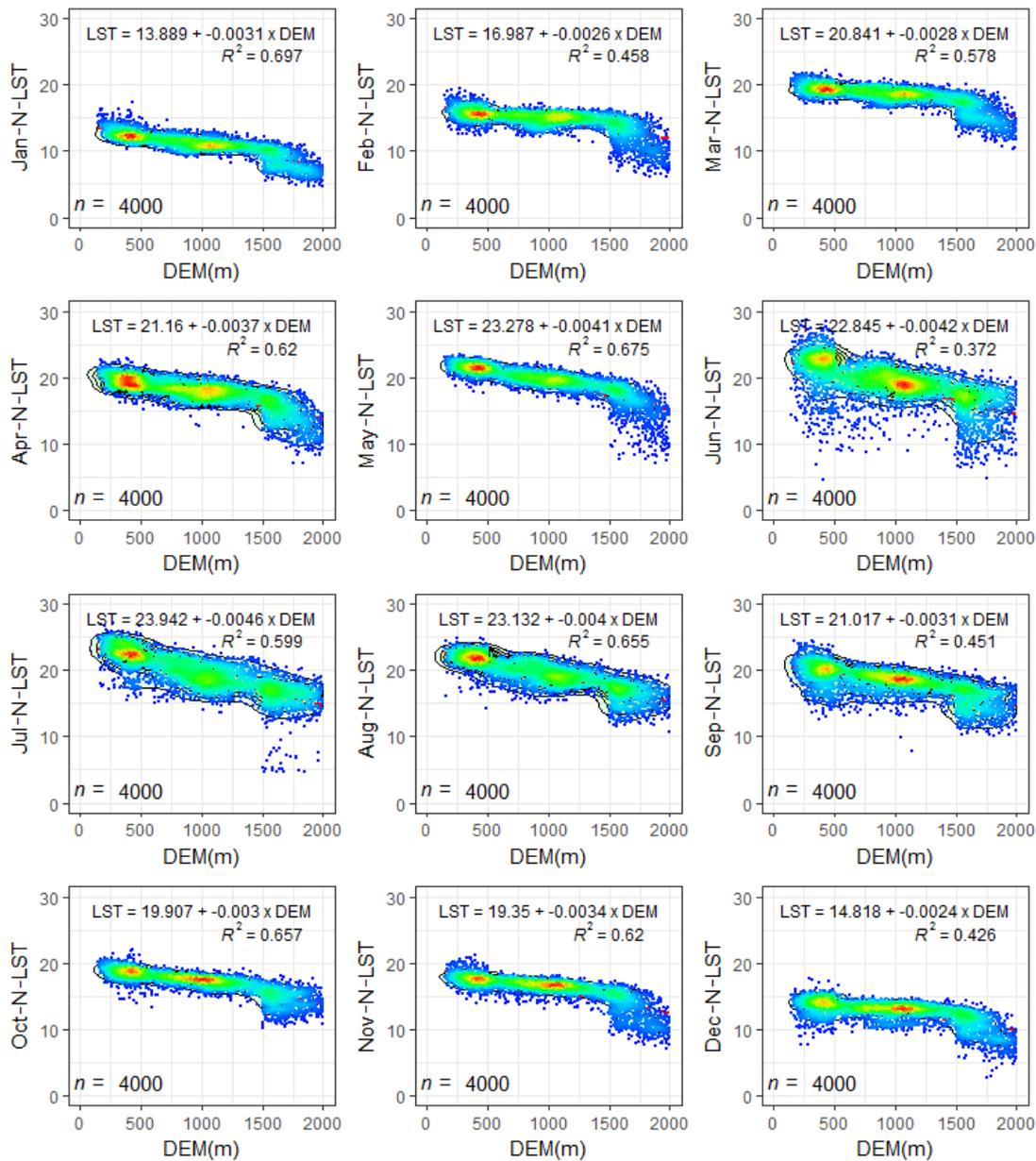
Figure 3.6. (b) Scatter plots showing the relationship between the average monthly nighttime LST of a balanced number of observation-based elevations and the elevation in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.

In this study, the relationship between LST and elevation and changes of longitude and latitude were additionally assessed. Two straight lanes were chosen for evaluation: Lane 1 (north to south) and Lane 2 (west to east). To reduce the effects of latitude and longitude, a line of 5 km width was used. Figure 3.7a,b shows that the elevations of the two lanes were consistent and ranged from 0 m to approximately 2000 m. During the daytime, LST decreased from 4.9 °C to 7.7 °C as elevation increased by 1000 m (Figure 3.7a). At nighttime, however, the LST dropped from 2.0 °C to 5.2 °C. Lane 2 (Figure 3.7b) shows the relationship between LST and elevation, where the longitude was assumed to be stable and the latitude changed by approximately 2° (200 km). During the daytime, the correlation between LST

and elevation was much lower than those of Lane 1 (Figure 3.7a). At nighttime, however, this correlation was high, except in February and December. The shift in LST due to a 1000 m change of elevation was also lower than Lane 1, ranging from 2.3 °C to 5.4 °C and 0 °C to 5.9 °C at daytime and nighttime, respectively. However, it should be noted that in Figure 3.7b, there were several observations in February and December at low elevations with very low LST in comparison to other months. This may result from unique weather events during these months. To assess the relationship between LST and elevation, such weather events should be considered.
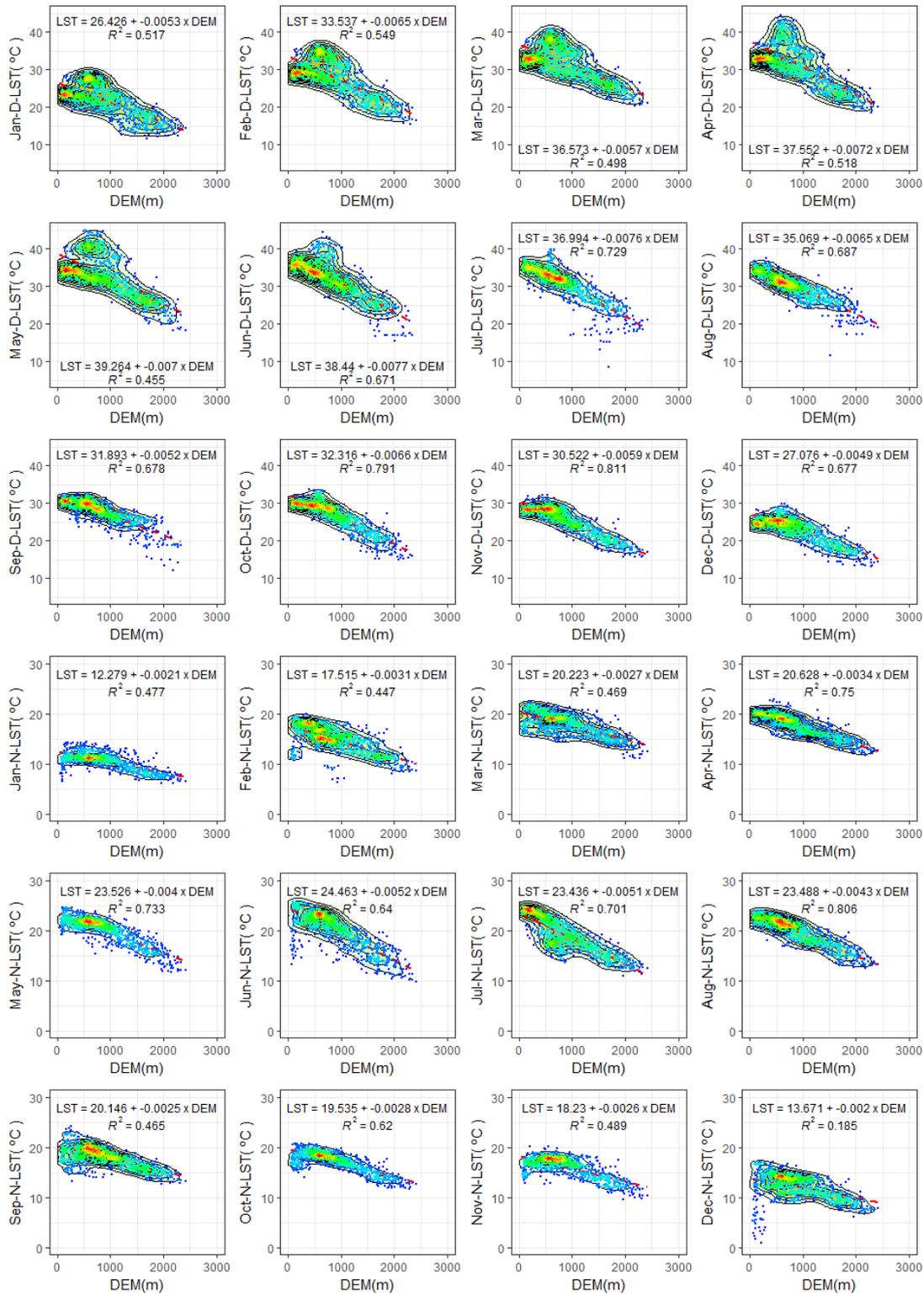
Figure 3.7. (a) Scatter plots showing the relationship between the average monthly LST (daytime and nighttime) and elevation of Lane 1 in 2015. The color ramp from blue to red expresses the point density from low to high.
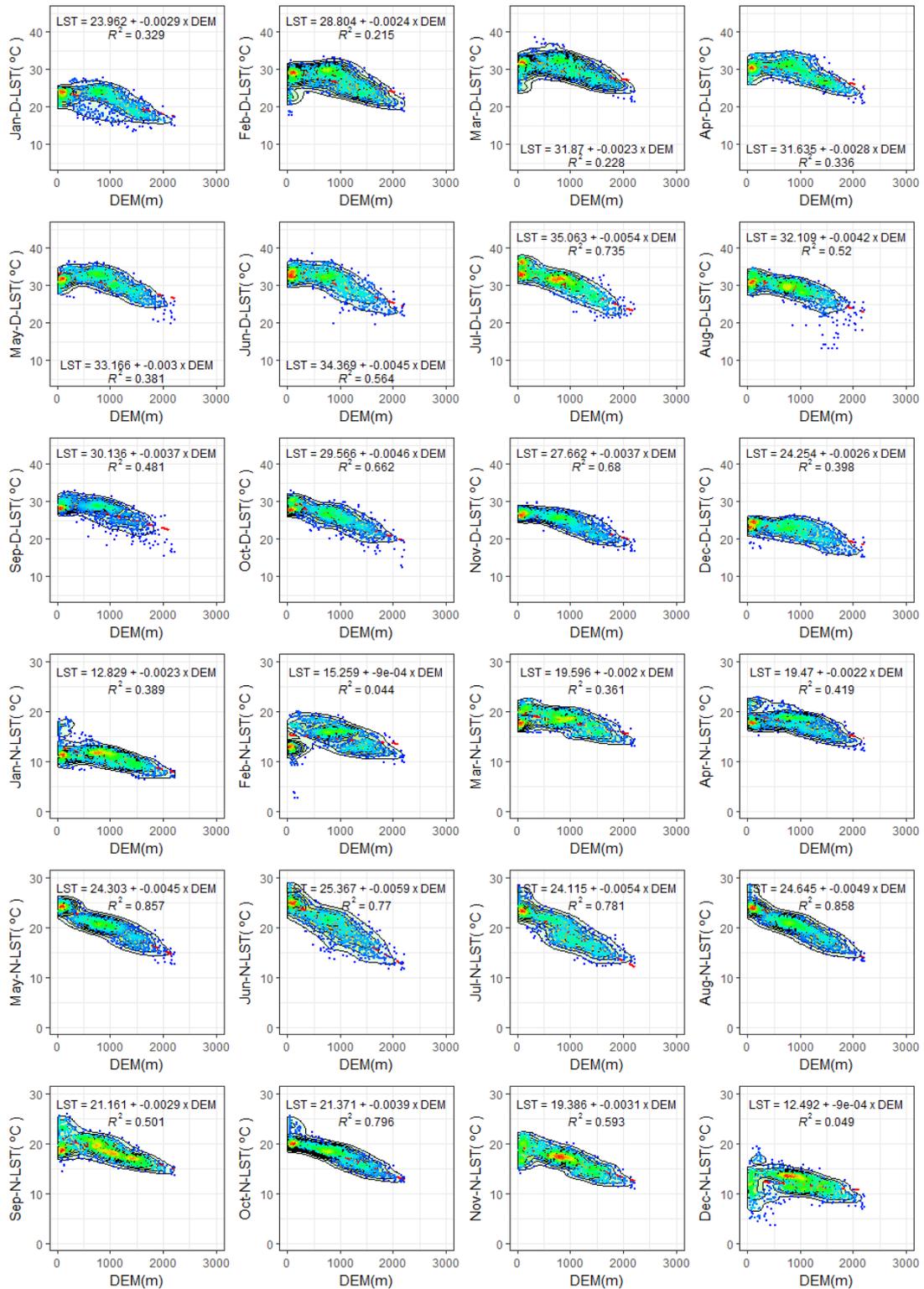
Figure 3.7. (b) Scatter plots showing the relationship between the average monthly LST (daytime and nighttime) and elevation of Lane 2 in 2015. The color ramp from blue to red expresses the point density from low to high.

73

3.5. Conclusions

In this study, we investigated the influence of elevation on LST of the daytime and nighttime from MODIS LST version 6 data over an area of 40,000 km^2 (200 × 200 km^2) in northwest Vietnam for 2015. The results of the average monthly daytime and nighttime LST showed a linear correlation between the average monthly LST and elevation. This correlation was stronger for the nighttime than for the daytime. However, the daytime variation (reducing) of LST was greater than those at nighttime when the elevation increased by 1000 m. For both the daytime and nighttime, the degree of change in LST due to an increase in elevation varied from January to December. The LST decreased from 3.8 °C to 6.1 °C and from 1.5 °C to 5.8 °C with a 1000 m increase in elevation at daytime and nighttime, respectively. Our results also showed that the type of land cover played an important role in the variation of LST due to changes in elevation. For both the daytime and nighttime, forest and bare land had the highest variations, while water and orchards showed the lowest variations. This suggests that both the elevation and land/use cover should be carefully considered and investigated when studying spatiotemporal LST distributions. In addition, it is suggested that for a more accurate representation of this relationship, all data should be used rather than an interval number of observations. Another interesting result was that the longitude and latitude influenced the relationship between LST and elevation, whereby latitude had a stronger effect. Furthermore, any special weather conditions within the study area during the study periods should be carefully considered, as they may bias the overall relationship between LST and elevation.
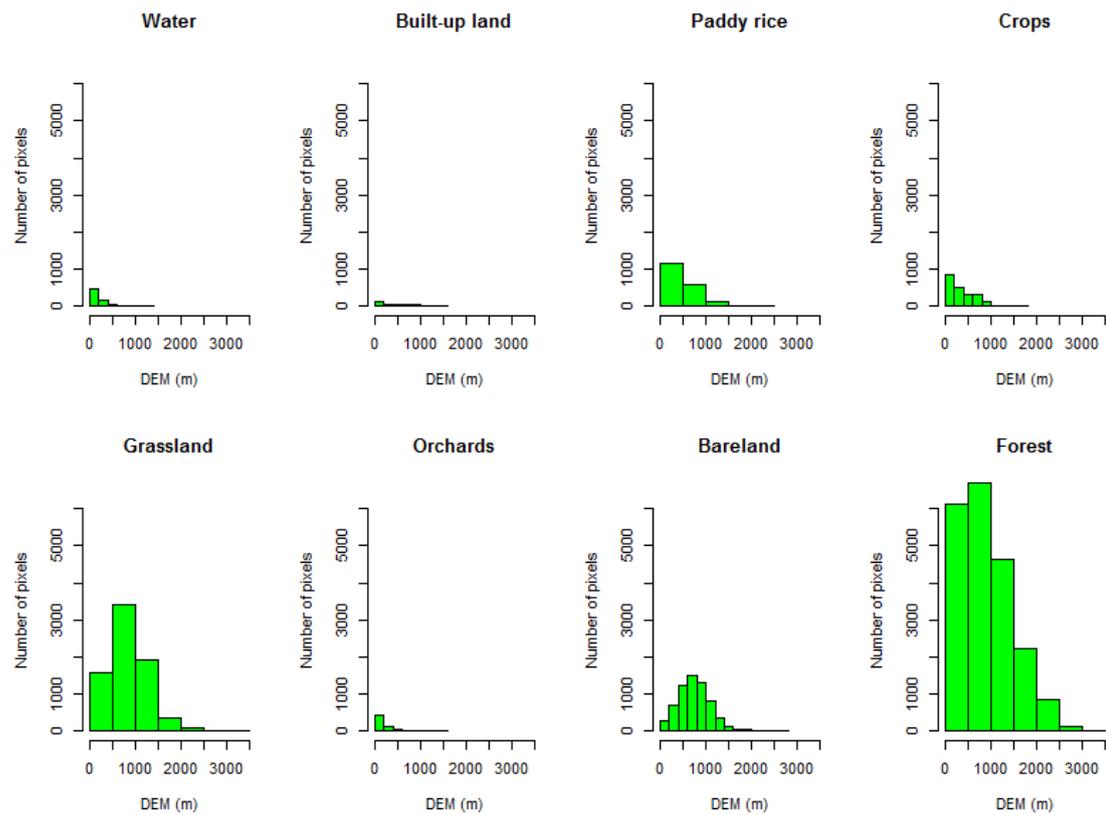
Appendix A

Figure A1. Histogram representing the number of pixels for each elevation (m a.s.l.) level for different land cover classes.

Appendix B
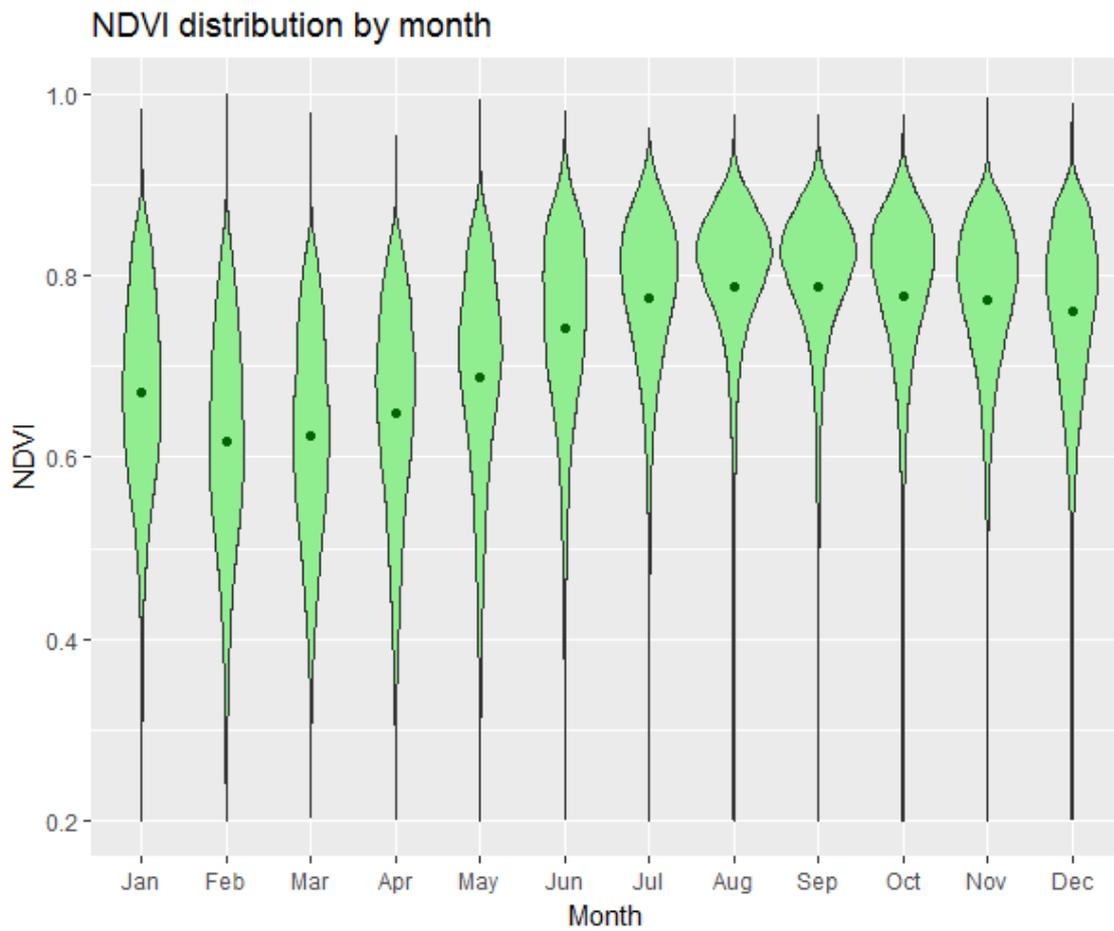
Figure A2. Violin plot of monthly average normalized difference vegetation index (NDVI) in northwest Vietnam in 2015.

Appendix C
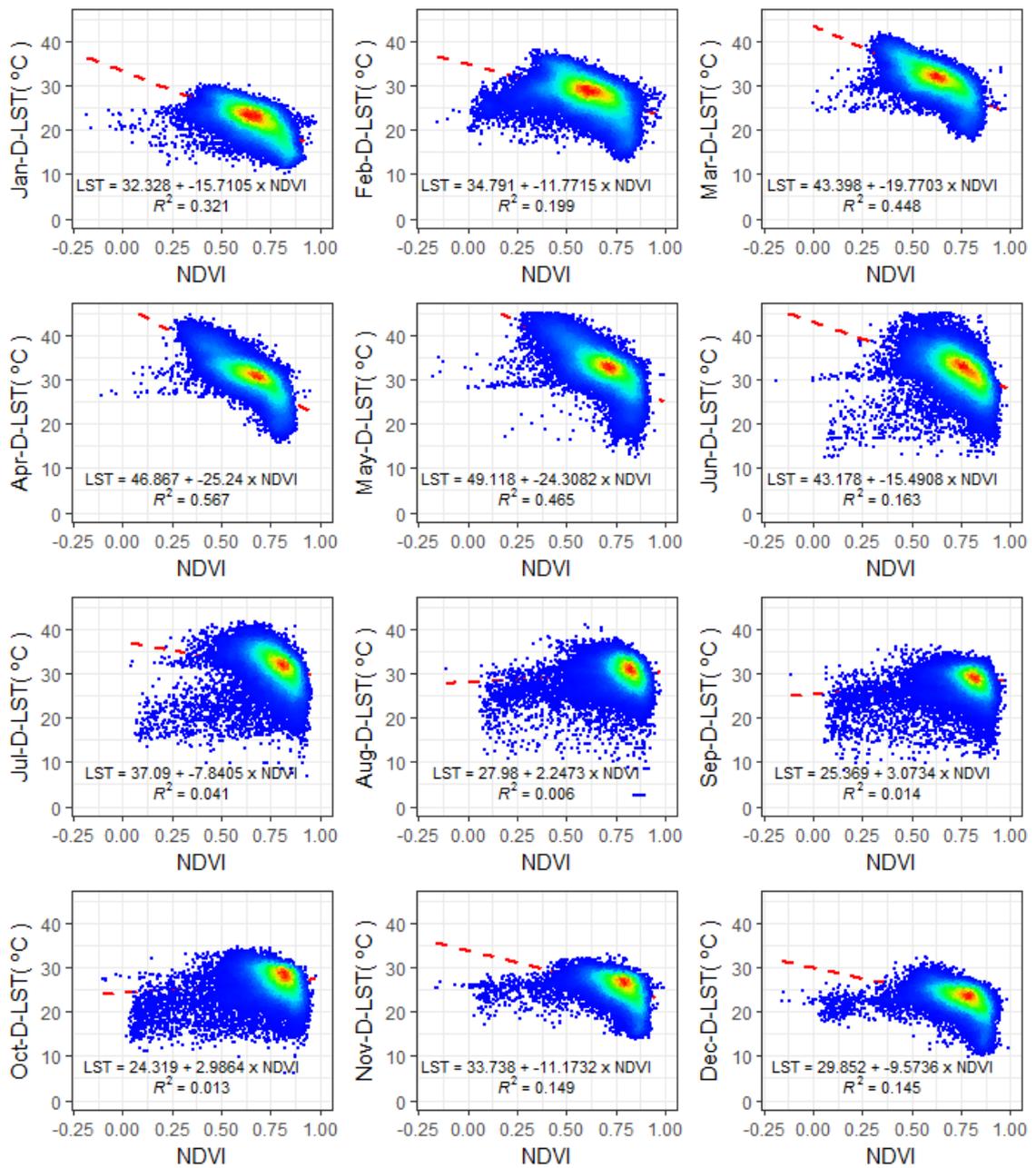
Figure A3. Scatter plots showing the relationship between the average monthly daytime LST and NDVI in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.
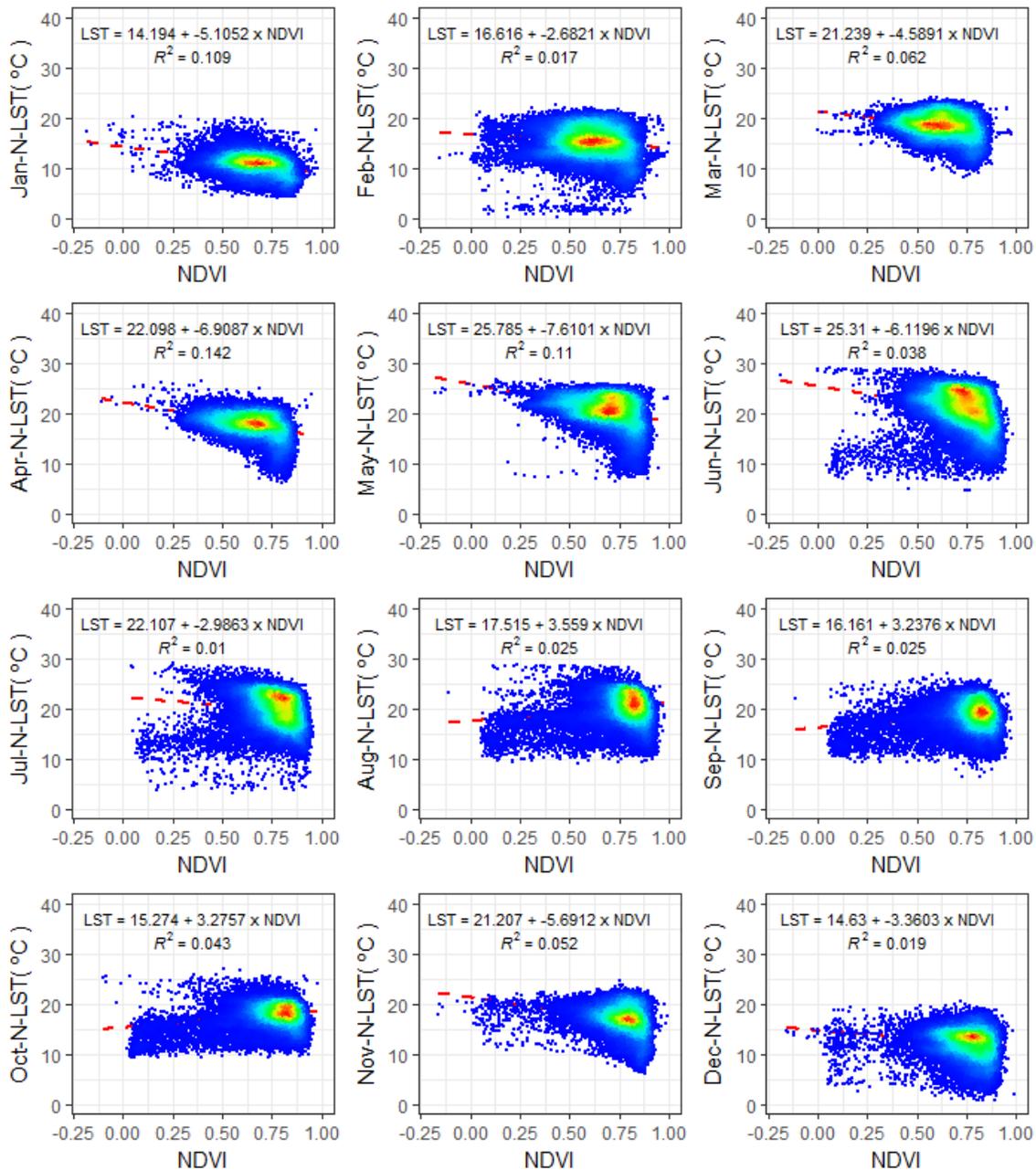
Figure A4. Scatter plots showing the relationship between the average monthly nighttime LST and NDVI in northwest Vietnam in 2015. The color ramp from blue to red expresses the point density from low to high.

References

Anderson, M.; Norman, J.; Kustas, W.; Houborg, R.; Starks, P.; Agam, N. (2008). A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. Remote Sensing of Environment, 112, 4227–4241.

Beniston, M. (2003). Climatic change in mountain regions: A review of possible impacts. Climatic Change, 59, 5–31.

Chen, W.; Shen, H.; Huang, C.; Li, X. (2017). Improving Soil Moisture Estimation with a Dual Ensemble Kalman Smoother by Jointly Assimilating AMSR-E Brightness Temperature and MODIS LST. Remote Sensing, 9, 273.

Chen, Y.; Quan, J.; Zhan, W.; Guo, Z. (2016). Enhanced Statistical Estimation of Air Temperature Incorporating Nighttime Light Data. Remote Sensing, 8, 656.

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

Duan, S.-B.; Li, Z.-L.; Tang, B.-H.; Wu, H.; Tang, R.L.; Bi, Y.; Zhou, G. (2014). Estimation of diurnal cycle of land surface temperature at high temporal and spatial resolution from clear-sky MODIS data. Remote Sensing, 6, 3247–3262.

El Kenawy, A.; Lopez-Moreno, J.; Vicente-Serrano, S.; Mekled, M. (2009). Temperature trends in Libya over the second half of the 20th century. Theoretical and Applied Climatology, 98, 1–8.

Faqe Ibrahim, G.R. (2017). Urban Land Use Land Cover Changes and Their Effect on Land Surface Temperature: Case Study Using Dohuk City in the Kurdistan Region of Iraq. Climate, 5, 13.

Gawuc, L.; Struzewska, J. (2016). Impact of MODIS Quality Control on Temporally Aggregated Urban Surface Temperature and Long-Term Surface Urban Heat Island Intensity. Remote Sensing, 8, 374.

Hu, Z.Z.; Yang, S.; Wu, R. (2003). Long-term climate variations in China and global warming signals. Journal of Geophysical Research: Atmospheres, 108, 4614–4626.

Huang, F.; Ma, W.; Wang, B.; Hu, Z.; Ma, Y. (2017). Air temperature estimation with MODIS data over the Northern Tibetan Plateau. Advances in Atmospheric Sciences, 34, 650–662.

Japanese Aerospace Exploration Agency. Available online: http://www.eorc.jaxa.jp/ALOS/en/lulc/lulc_vnm.htm (accessed on 20 November 2017).

Jiang, J.; Tian, G. (2010). Analysis of the impact of land use/land cover change on land surface temperature with remote sensing. Procedia Environmental Sciences, 2, 571–575.

Jung, C.; Lee, Y.; Cho, Y.; Kim, S. (2017). A Study of Spatial Soil Moisture Estimation Using a Multiple Linear Regression Model and MODIS Land Surface Temperature Data Corrected by Conditional Merging. Remote Sensing, 9, 870.

Karnieli, A.; Agam, N.; Pinker, R.T.; Anderson, M.; Imhoff, M.L.; Gutman, G.G.; Panov, N.; Goldberg, A. (2010). Use of NDVI and Land Surface Temperature for Drought Assessment: Merits and Limitations. Journal of Climate, 23, 618–633.

Katpatal, Y.B.; Kute, A.; Satapathy, D.R. (2008). Surface- and air-temperature studies in relation to land use/land cover of Nagpur urban area using Landsat 5 TM data. Journal of Urban Planning and Development, 134, 110–118.

Kayet, N.; Pathak, K.; Chakrabarty, A.; Sahoo, S. (2016). Spatial impact of land use/land cover change on surface temperature distribution in Saranda Forest, Jharkhand. Modeling Earth Systems and Environment, 2, 1–10.

Keramitsoglou, I.; Kiranoudis, C.T.; Ceriola, G.; Weng, Q.; Rajasekar, U. (2011). Extraction and Analysis of Urban Surface Temperature Patterns in Greater Athens, Greece, Using MODIS Imagery. Remote Sensing of Environment, 115, 3080–3090.

Khandelwal, S.; Goyal, R.; Kaul, N.; Mathew, A. (2017). Assessment of land surface temperature variation due to change in elevation of area surrounding Jaipur, India. The Egyptian Journal of Remote Sensing and Space Sciences, in press.

Land Processes Distributed Active Archive Center (LP DAAC). Available online: https://lpdaac.usgs.gov/ (accessed on 1 November 2017).

Li, Z.-L.; Tang, B.-H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. (2013). Satellite-derived land surface temperature: Current status and perspectives. Remote Sensing of Environment, 131, 14–37.

Limsakul, A.; Goes, J.I. (2008). Empirical evidence for interannual and longer period variability in Thailand surface air temperatures. Atmospheric Research, 87, 89–102.

Lin, X.; Zhang, W.; Huang, Y.; Sun, W.; Han, P.; Yu, L.; Sun, F. (2016). Empirical Estimation of near-Surface Air Temperature in China from MODIS LST Data by Considering Physiographic Features. Remote Sensing, 8, 629.

Liu, X.; Cheng, Z.; Yan, L.; Yin, Z. (2009). Elevation dependency of recent and future minimum surface air temperature trends in the Tibetan Plateau and its surroundings. Global and Planetary Change, 68, 164–174.

McElwain, L.; Sweeney, J. (2007). Key Meteorological Indicators of Climate Change in Ireland; Environmental Research Centre Report; Environmental Protection Agency: Washington, DC, USA.

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.

Miles, V.; Esau, I. (2017). Seasonal and Spatial Characteristics of Urban Heat Islands (UHIs) in Northern West Siberian Cities. Remote Sensing, 9, 989.

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.

Noi, P.T.; Kappas, M.; Degener, J. (2016). Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam. Remote Sensing, 8, 1002.

Parinussa, R.M.; Lakshmi, V.; Johnson, F.; Sharma, A. (2016). Comparing and Combining Remotely Sensed Land Surface Temperature Products for Improved Hydrological Applications. Remote Sensing, 8, 162.

Perugini, L.; Caporaso, L.; Marconi, S.; Cescatti, A.; Quesada, B.; de Noblet-Ducoudré, N.; Arneth, A. (2017). Biophysical effects on temperature and precipitation due to land cover change. Environmental Research Letters, 12, 053002.

Rajasekar, U.; Weng, Q. (2009). Urban heat island monitoring and analysis by data mining of MODIS imageries. ISPRS Journal of Photogrammetry and Remote Sensing, 64, 86–96.

Sánchez, N.; González-Zamora, Á.; Piles, M.; Martínez-Fernández, J. (2016). A New Soil Moisture Agricultural Drought Index (SMADI) Integrating MODIS and SMOS Products: A Case of Study over the Iberian Peninsula. Remote Sensing, 8, 287.

Shah, D.B.; Pandya, M.R.; Trivedi, H.J.; Jani, A.R. (2013). Estimating minimum and maximum air temperature using MODIS data over Indo-Gangetic Plain. Journal of Earth System Science, 122, 1593–1605.

Shwetha, H.; Kumar, D.N. (2016). Prediction of high spatio-temporal resolution land surface temperature under cloudy conditions using microwave vegetation index and ANN. ISPRS Journal of Photogrammetry and Remote Sensing, 117, 40–55.

Stathopoulou, M.; Cartalis, C. (2009). Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. Remote Sensing of Environment, 113, 2592–2605.

Stroppiana, D.; Antoninetti, M.; Brivio, P.A. (2014). Seasonality of modis lst over southern Italy and correlation with land cover, topography and solar radiation. European Journal of Remote Sensing, 47, 133–152.

Sun, H.; Chen, Y.; Gong, A.; Zhao, X.; Zhan, W.; Wang, M. (2014). Estimating mean air temperature using MODIS day and night land surface temperatures. Theoretical and Applied Climatology, 118, 81–92.

Vancutsem, C.; Ceccato, P.; Dinku, T.; Connor, S.J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114, 449–465.

Wan, Z. (2014). New refinements and validation of the collection -6 MODIS land-surface temperature/emissivity product. Remote Sensing of Environment, 140, 36–45.

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.

Xu, Y.; Shen, Y.; Wu, Z. (2013). Spatial and temporal variations of land surface temperature over the Tibetan Plateau based on Harmonic analysis. Mountain Research and Development, 33, 85–94.

Yanev, I.; Filchev, L. (2016). A comparative analysis between MODIS LST level-3 product and in-situ temperature data for estimation of urban heat island of Sofia. Aerosp. Res. Bulg, 28, 77–92.

Yang, Y.Z.; Cai, W.H.; Yang, J. (2017). Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China. Remote Sensing, 9, 410.

You, Q.; Kang, S.; Pepin, N.; Yan, Y. (2008). Relationship between trends in temperature extremes and elevation in the eastern and central Tibetan Plateau, 1961–2005. Geophysical Research Letters, 35, L14704.

Chapter 4

Evaluation of MODIS land surface temperature products for daily air surface temperature estimation in northwest Vietnam (under review**)**

Thanh Noi Phan [a,b,], Martin Kappas [a], Khac Thoi Nguyen [b], Trong Phuong Tran [b], Quoc Vinh Tran [c], and Ammar Rafiei Emam [a]

[a] Department of Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen, Göttingen 37077, Germany

[b] Department of Cartography and Geodesy, Faculty of Land Management, Vietnam National University of Agriculture, Hanoi, Vietnam.

[c] Department of Land Information, Faculty of Land Management, Vietnam National University of Agriculture, Hanoi, Vietnam.

Correspondence: thanh-noi.phan@geo.uni-goettingen.de; Tel.: +49 551 399805, Department of Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen, Goldschmidt Street 5, Göttingen 37077, Germany.

Thanh Noi Phan's ORCID: https://orcid.org/0000-0002-2747-5028

Abstract

Recently, the MODerate resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) is considered one of the most suitable ways to retrieve air surface temperature (Ta) – one of the most important and widely used climate variables for a wide range of applications. In fact, many successful studies have been reported from many regions of the world. Each day, four MODIS LST data are available; from two sensors (Terra and Aqua) at two local overpass times (daytime and nighttime). However, due to their different overpass times, most studies have used LST daytime and nighttime for Ta_max and Ta_min estimation, respectively. Therefore, the performance of each individual LST data, the effect of the dynamic combination of these four LST data, the effect of land surface characteristics, and the effect of LST quality on the same estimation in the same region on the accuracy of estimated Ta remains unclear. In this study, we evaluated and tested all individual LST data as well as all possible combinations of the four MODIS LST data from two separate stations with distinct land surface characteristics in northwest Vietnam for 10 years (from 2004 to 2013) under two sky conditions (all clear sky conditions and only good data—i.e. Quality Control (QC) value of 0) for daily Ta (Ta_max, Ta_min, and Ta_mean) estimations. In addition, the mixed data of the two stations were also evaluated. Our results showed that Terra LST data have a higher correlation with Ta than Aqua LST; which is consistent for both stations and both quality conditions (all clear sky and only good data). A closer overpass time with Ta_max or Ta_min occurrence time did not guarantee a higher accuracy of Ta estimation. Using only good LST data produced a higher accuracy of Ta estimation than using all clear sky data. However, if the percentage of good data is low (i.e. less than 30%), then the all clear sky data will provide better results for Ta_max estimation. Combinations including at least one nighttime LST produced stable and high accuracy Ta_min and Ta_mean estimates, while the combinations with only daytime LST produced very low accuracy results. For Ta_max estimation, the results were less impacted by LST quality; however, they were strongly impacted by different combinations and land surface characteristics.

4.1. Introduction

Air surface temperature (Ta), which plays an important role in a wide range of applications, is a key input of most models relating to land-atmosphere interaction. Among these, the most popular applications are crop yield prediction (Balaghi et al. 2008; de Wit and van

Diepen, 2008), hydrology (Mo and Lettenmaier 2014; Rabi, Nyarko and Sperac 2015), urban heat island (Rajasekar and Weng 2009; Pichierri et al. 2012; Keramitsoglou et al. 2016), and drought monitoring (Sheffield et al. 2012; Dai 2013; Cehn and Sun 2015; Bazrafshan 2017).

Traditionally, Ta has been measured 1.5 to 2.0 m above the ground surface by weather stations and can be very accurate and temporally resolved (e.g. hourly measurements). However, in less developed regions, the spatial distribution of weather stations is sparse. In these areas, information about Ta over large regions is limited, and therefore it may affect the spatial modelling of the above-cited applications (Stisen et al. 2007; Nieto et al. 2011; Lin et al. 2012). Consequently, some effects of this problem can be observed as food insecurity, improper management of natural resources, or poor understanding of ecosystems (Maeda 2014). In other words, the demand for spatiotemporal Ta information with high accuracy is still rising (Oyler et al. 2015; Yang et al. 2017).

Land surface temperature (LST), which can be directly retrieved from remotely sensed radiance data, is considered one of the most important and useful data sources for Ta retrieval over a region or large area (Li et al. 2013; Chen et al. 2016). In fact, various studies have used LST data for Ta estimation with high accuracy, i.e. root mean square error (RMSE) ranges from 1 ℃ to 6 ℃, depending on the study area, methods, and LST data sources (Goward et al. 1994; Prihodko and Goward 1997; Pape and Loffler 2004; Mostovoy et al. 2006; Zhang et al. 2011; Benali et al. 2012; Xu et al. 2014; Noi et al. 2016; Janatian et al. 2017; Yoo et al. 2018).

There are four popular satellite sensors which provide LST data: the Advanced Very High Resolution Radiometer (AVHRR) sensor onboard NOAA satellites, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor onboard Terra satellites, Landsat (TM, ETM, and TIRS sensors) of Landsat satellites, and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard Terra and Aqua satellites. AVHRR, ASTER, and Landsat each have their own limitations, such as AVHRR lacks nighttime LST data, ASTER is only available upon request and payment, and Landsat has a coarse temporal resolution of 16 days (Tomlinson et al. 2011). Therefore, MODIS LST data is considered the most suitable data source for Ta estimation because it is freely available, has a moderate spatial resolution of about 1 km, and has a very high temporal resolution (up to four observations per day with daytime and nighttime data).

Since becoming available in early 2000 (Terra) and mid-2002 (Aqua), MODIS LST data has received much attention and has been used for Ta estimation in many regions, such as Europe (Benali et al. 2012; Niclos et al. 2014; Peon et al. 2014), USA (Zeng et al. 2015; Oyler

et al. 2015), Canada (Xu et al. 2014), Africa (Vancutsem et al. 2010; Lin et al. 2012), India (Shah et al. 2013), the Tibetan Plateau (Zhang et al. 2016a, 2016b; Huang et al. 2017; Xu et al. 2018), western Asia (Emamifar et al. 2013), and southeast Asia (Noi et al. 2016, 2017).

To estimate Ta using MODIS LST data, many approaches have been developed and applied. However, according to Benali et al. (2012) and Noi et al. (2016), they can be classified into three popular groups of methods: energy-balance modelling, temperature-vegetation index (TVX), and statistical methods. Among these, the statistical methods are the most popular in recent studies, because compared to energy-balance methods, statistical methods have simpler operability. When compared to TVX methods, statistical methods have a wider applicability to different land surface characteristics (Chen et al. 2016). These statistical methods can be divided into three specific methods: simple statistic (using one variable), advanced statistic (using more than two variables, usually MODIS LST with auxiliary variables), and machine learning methods.

Four MODIS LST measurements are taken each day: Aqua daytime (MYDD), Aqua nighttime (MYDN), Terra daytime (MODD), and Terra nighttime (MODN). However, most studies simply use one LST data (among the four available) for Ta estimation—LST daytime and LST nighttime for Ta_max and Ta_min estimation, respectively (Xu et al. 2014; Zhang et al. 2016b; Shi et al. 2017; Cai et al. 2017). Since LST daytime and LST nighttime have overpass times (local solar time) close to the time of Ta_max and Ta_min occurring, respectively. In contrast, Zhang et al. (2011) have shown that LST nighttime (MYDN/MODN) produced a higher accuracy of Ta_min, Ta_mean, and even Ta_max estimation than LST daytime. In addition, Zeng et al. (2015) concluded that the combination of LST daytime and nighttime (MYDD with MYDN/MODD with MODN) produced the highest accuracy of Ta estimation in comparison to solely using LST daytime or LST nighttime. Most recently, Noi et al. (2016, 2017), Zhou et al. (2017), and Yoo et al. (2018) have shown that the combination of all four MODIS LST data could produce the highest accuracy of Ta estimation. However, these studies used MODIS LST data together with auxiliary variables to estimate Ta, or MODIS LST data were used with the same sky conditions (all data under clear sky conditions or only good data—i.e. Quality Control (QC) value of 0). Therefore, the performance of each individual MODIS LST data with each sky condition remains unknown.

To our knowledge, no study has compared all four LST products (MODD, MODN, MYDD, MYDN) in Ta_max, Ta_min, and Ta_mean estimation for different land surface characteristic conditions (i.e. heterogeneous vs. homogeneous land cover) and different

quality of MODIS LST data (i.e. all clear sky condition vs. only good condition LST data). Another motivation for us to conduct this study is that most of the study (which related to choosing LST data) was implemented with MODIS LST C5 (Collection 5) data and no comparisons have been conducted with MODIS LST C6 data. Therefore, it is necessary and practical to implement a study to evaluate MODIS LST data for Ta estimation.

The objectives of this study were to: (1) compare the performance of individual MODIS LST daytime versus nighttime, MODIS LST Terra versus Aqua, MODIS LST under all clear sky versus good quality for Ta_max and Ta_min estimation, (2) find the best combination for Ta estimation using MODIS LST data, and (3) evaluate the effects of MODIS LST quality and land surface characteristic conditions on the accuracy of Ta estimation.

4.2. Materials and Methods

4.2.1. Study Area
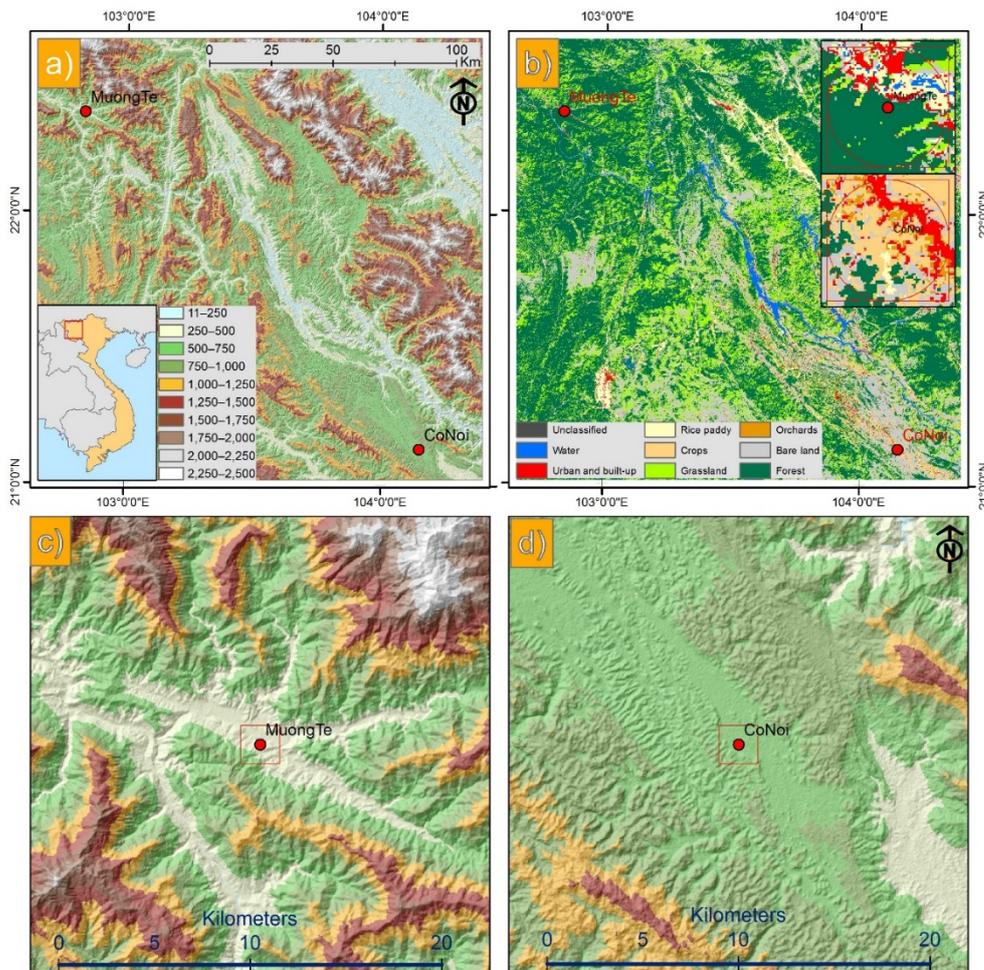
Figure 4.1. Location of the study area in northwest Vietnam. (a, c, d) Elevation range from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (DEM)

and (b) land cover in 2015 with 15 m resolution derived from the Japanese Aerospace Exploration Agency (JAXA) distribution in the study area and 1 km buffered around the two stations.

A number of studies have concluded that along with MODIS LST, land cover type and elevation were considered the most important variables for Ta estimation using MODIS LST data (Huang et al. 2017; Kloog et al. 2017). Therefore, to fulfill the objectives of this study, we selected two weather stations with different land surface characteristics, the Muongte station and the Conoi station (Figure 4.1). The location of each weather station was selected based on the available station network, a Japanese Aerospace Exploration Agency (JAXA) land cover map (15 m spatial resolution for 2015), and elevation from ASTER Global DEM (30 m spatial resolution, downloaded from the U.S. Geological Survey (USGS) website).

Muongte station is located in the western part of northern Vietnam with an altitude of 329 m. As shown in Figure 4.1b (a buffer zone of 1 km from the station location), the main land cover of this station is forest with some built-up area at the north. Conoi station is about 100 km southeast of Muongte station, with an altitude of 671 m. Within a 1 km radius of Conoi station, the land cover is more heterogeneous than around Muongte, consisting of mainly of crops, built-up, and bare land. It is worth noting that a buffer area of 1 km around the stations was chosen because it is consistent with the spatial resolution of MODIS LST data, and helps to understand the land cover surrounding the weather stations.

Figure 4.1c and 1d show that within 1 km the terrain of both stations was similar and flat. However, the terrain is more complex within a 10 km radius of the Muongte station compared to the Conoi station, with high mountains (up to 2000 m and higher) surrounding the station.

4.2.2. Data

4.2.2.1. MODIS Data

In this study, we used the newest collection of daily MODIS LST (collection 6 – C6) aboard Terra (MOD11A1) and Aqua (MYD11A1) from 01 January 2004 to 31 December 2013. The overpass times of Terra and Aqua through the region containing the two stations (Muongte and Conoi) were close together; approximately 11:20 a.m./10:15 p.m. and 12:50 p.m./2:30 a.m., respectively. Therefore, each day four LST datasets were created: Terra daytime (MODD), Terra nighttime (MODN), Aqua daytime (MYDD), and Aqua nighttime (MYDN).

All MODIS data were collected from the USGS Land Processes Distributed Active Archive Center (LP DAAC) in HDF format.

Table 4.1. All possible combinations of the four LST data and the number of observations under all clear sky conditions (CS_data) and only good data (G_data) at Muongte and Conoi stations.

No.	Combination			Muongte station			Conoi station		
				CS_data	G_data	%	CS_data	G_data	%
C01	MODD			883	470	53.23	933	593	63.56
C02	MODN			1120	630	56.25	1056	802	75.95
C03	MYDD			1188	576	48.48	1014	507	50.00
C04	MYDN			1021	262	25.66	909	613	67.44
C05	MODD	+ MODN		476	207	43.49	536	308	57.46
C06	MYDD	+ MYDN		630	129	20.48	512	225	43.95
C07	MODD	+ MYDD		546	210	38.46	579	262	45.25
C08	MODN	+ MYDN		692	151	21.82	608	396	65.13
C09	MODD	+ MYDN		455	128	28.13	518	303	58.49
C10	MODN	+ MYDD		682	258	37.83	584	279	47.77
C11	MODN	+ MYDD	+MYDN	455	87	19.12	377	169	44.83
C12	MODD	+ MYDD	+MYDN	309	80	25.89	355	148	41.69
C13	MODD	+ MODN	+MYDN	315	82	26.03	379	208	54.88
C14	MODD	+ MODN	+MYDD	329	118	35.87	375	162	43.20
C15	MODD + MODN	+MYDD	+MYDN	216	52	24.07	262	109	41.60

4.2.2.2. Weather Station Data

To fulfill the objectives of this study, we collected daily Ta (Ta_max, Ta_min, and Ta_mean) of two weather stations, Muongte and Conoi stations (Figure 4.1) from the Vietnam Institute of Meteorology, Hydrology, and the Environment (IMHEN) for the 10 years from 01 January 2004 to 31 December 2013. Ta was recorded hourly, in which Ta_max and Ta_min are the maximum and minimum temperatures of the day, respectively. Ta_mean is the average temperature of all 24 hourly measurements in a single day.

4.2.3. Methods

4.2.3.1. Pre-processing MODIS Data.

Each HDF of MOD11/MYD11 A1 data image includes twelve data layers, however for this study purpose, we extracted only four layers of datasets: LST daytime, QC_Day, LST nighttime, and QC_Night. The MODIS reprojection tools, which are freely provided by LP DAAC, were used to convert from HDF (sinusoidal projection system) to TIFF (the UTM Zone 48 N projection system with WGS84 datum).

ArcMap 10.5 was used to extract LST pixels of weather stations, and convert from a DN (Digital Number) value to Celsius (ºC) using the following equation:

$$LST\ (°C) = 0.02 \times DN - 273.15 \tag{1}$$

where °C is the temperature in Celsius and 0.02 is the scale factor of the MODIS LST product (LP DACC).

According to Wan et al. (2014), significant changes and improvements have been made for MODIS LST C6. Among these improvements, removing cloud-contaminated LSTs pixels is considered one of the most important changes of C6 in comparison to previous collections (C4.1, C5). There were still some pixels produced under thin cloud cover, however, that the algorithm was not able to detect and remove. Therefore, the quality information (QC file) of each image dataset was provided for user-purpose selection. In this study, to evaluate the MODIS LST data, we used two types of data: all valid LST data (which was produced under all clear sky conditions), and only good LST data (which was defined by the QC file data).

To evaluate the performance of different LST data from different satellites (MODD, MODN, MYDD, MYDN) and to minimize the effects of other factors such as land cover, elevation, or overpass time; we evaluated data from each station (which rarely was evaluated in previous studies), and combined the two stations (which commonly was evaluated in literature). Consequently, we have three pairs of datasets:

- Set1: Muongte station data: CS_data (all clear sky), and G_data (only good).
- Set2: Conoi station data: CS_data, and G_data.
- Set3: Combined two stations: CS_data, and G_data.

The detail of these datasets is shown in Table 4.1. In general, Conoi station has a higher percentage of good data than Muongte station, particularly at nighttime (Table 4.1).

It should be noted that based on QC flag data, all MODIS LST at two stations (under all clear sky data) have an error $\leq$ 2K.

4.2.3.2. Ta Estimation using MODIS LST Data

For each dataset (Set1, Set2, and Set3), we applied the same procedure for Ta estimation. All possible combinations (15 combinations from 4 LST data – Table 4.1) were applied for all Ta_max, Ta_min and Ta_mean estimation.

As mentioned in the Introduction section (Section 1), due to some drawbacks of TVX and energy-balance methods, we used the statistical approach (simple and multiple linear regression) for Ta estimation. Furthermore, to have a clear view of each LST data (MYDD, MYDN, MODD, MODN) we used only LST data (without auxiliary) for Ta estimation. In total, there are 15 different models (Table 4.1) for Ta_max, Ta_min, and Ta_mean estimation.

To evaluate the performance of each model, we used two popular criteria: the coefficient of determination (R^2) and the root mean square error (RMSE).

Furthermore, we randomly divided all observations of each dataset into two parts: calibration and validation with 75% and 25% proportions, respectively. The R^2 and RMSE were calculated from the estimated and measured Ta of the validation datasets. These procedures were repeated 100 times with the averaged R^2 and RMSE values recorded.

4.3. Results

4.3.1. The relationship between MODIS LST and Ta, and The Influence of Land Surface Characteristic.
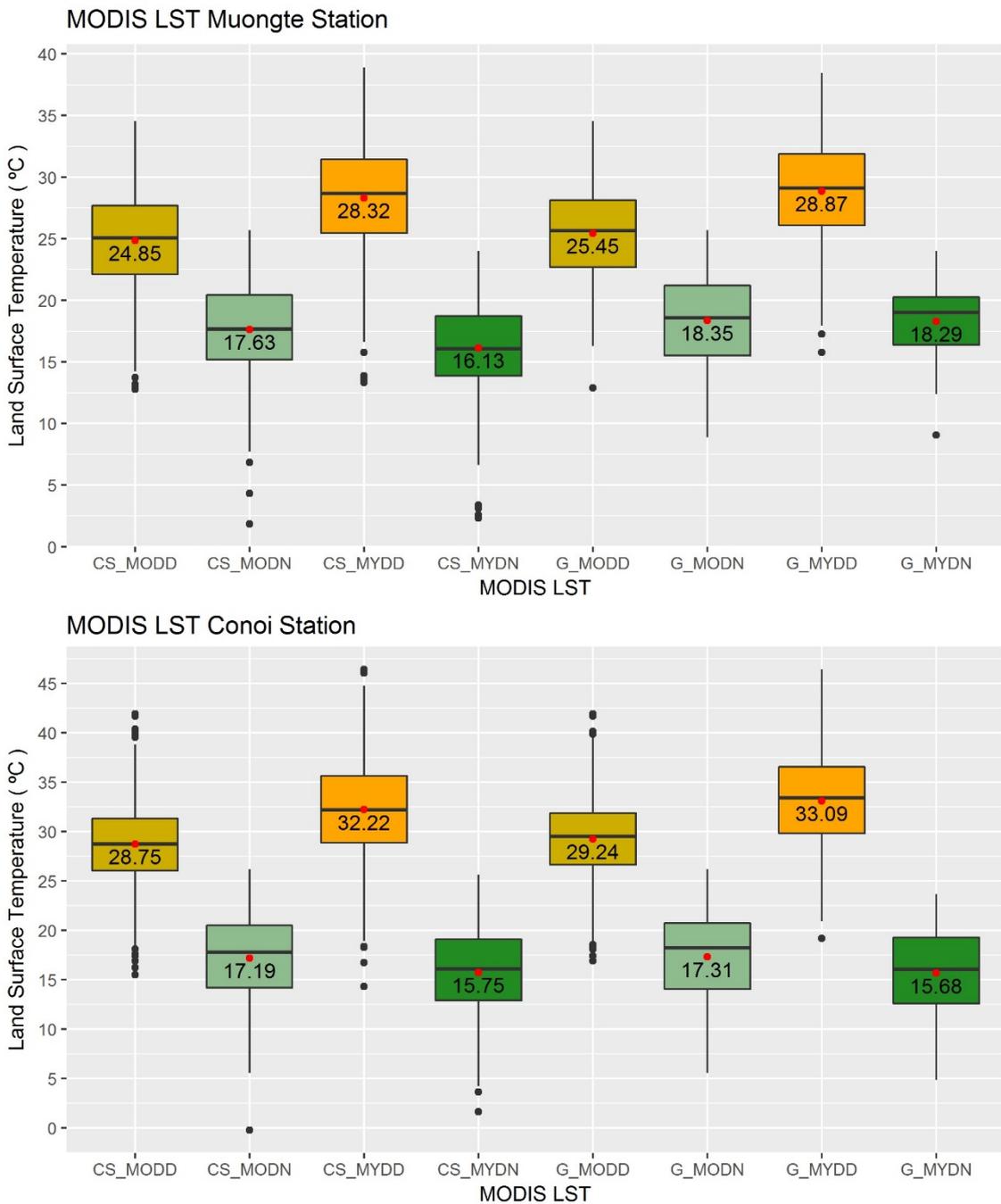
Figure 4.2. The average temperature of the four MODIS LST data at two stations, under two sky conditions: all clear sky condition (CS_*) and only good condition (G_*).

As shown in Figure 4.2, the nighttime LST were consistent between the two stations, however, the daytime LST values have a large difference, approximately 4 °C for both Terra and Aqua daytime. This indicates that land surface characteristics have a strong impact on

LST at daytime. Regarding the time-dependency, both stations have a close overpass time (the overpass time of Terra is approximately 11:20 a.m. and Aqua is approximately 12:50 p.m.), yet the difference between MYDD and MODD is almost 4 °C at both stations (Figure 4.2). It is suggested that the direction of solar radiation (at daytime) has a strong effect on LST, and this effect is quite similar at different land surface characteristics. However at nighttime, the difference of LST between MODN (overpass time approximately 10:15 p.m.) and MYDN (overpass time approximately 2:30 a.m.), lowered to about 1.5 °C (of all clear sky data) for both stations. This indicates that overpass time has some effects on the LST. It should be noted that the lowering of LST between MODN and MYDN were different between all clear sky data and good data. At Conoi station, the difference of all clear sky data (which lowered 1.5 °C) was consistent with good data (which lowered 1.6 °C); whereas at Muongte station, the lowering was only 1.5 °C and 0.06 °C with all clear sky data and good data, respectively. This can be explained by only 25.7% of all clear sky data at Muongte station having good quality, compared to 67.5% at Conoi station (Table 4.1).

Figure 4.2 also shows that the differences between daytime and nighttime of Aqua LST were larger than those of Terra LST. The larger differences between both Terra and Aqua LST were observed at Conoi station. This confirms that land surface characteristics have an impact on LST.
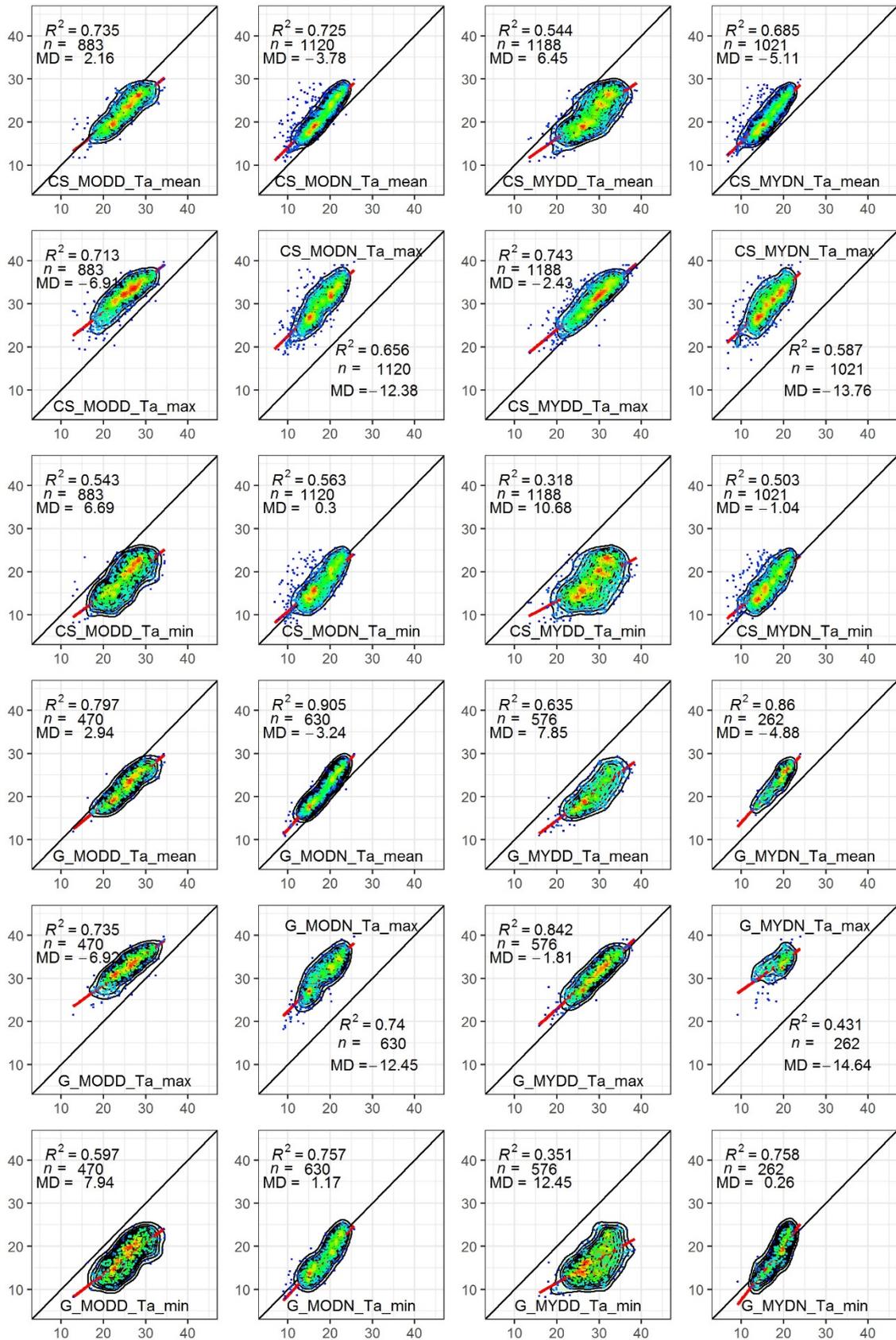
Figure 4.3. The point density plots of the LST (all clear sky data/ good data) products and Ta (Ta_mean, Ta_max, and Ta_min) at MuongTe station. The point density from low to high is expressed by the colour ramp from blue to red.

In general, Terra LST has a higher correlation with Ta_mean, Ta_max, and Ta_min than that of Aqua LST. This is true for both datasets; all clear sky LST (CS-data) and good LST (G-data).

Figure 4.3 also shows that both LST daytime and nighttime of both satellites were distributed above and below of 1:1 line of Ta_max and Ta_min, respectively. This is indicates that all most LST was lower than Ta_max and higher than Ta_min. However, with both CS-data and G-data, Terra LST was distributed closer to the 1:1 line than Aqua LST. This could be explained by the overpass time (local solar time) of Terra (11:20 a.m., 10:15 p.m.) being closer to Ta_max and Ta_min than Aqua.

Furthermore, looking at the Mean Difference (MD) between MODIS LST and Ta, it is clearly seen that MODD (11:20 a.m.) is closest to Ta_mean (MD = +2.16 °C) and MYDD (12:50 p.m.) is closest to Ta_max (MD = -2.43 °C). This is consistent with both data conditions; all clear sky and only good LST data. For Ta_min, there is a difference: the all clear sky data MODN (10:15 p.m.) is closest to Ta_min (MD = +0.3 °C), and the only good data MYDN (2:30 a.m.) is closest to Ta_min (MD = +0.26 °C). It is worth noting that the variation of the smallest MD from MODD (11:20 a.m.) to MYDD (12:50 p.m.) was larger (+2.16 to +6.45 with CS-data, and +2.94 to +7.85 °C with G-data) than that of the MD from MODN (10:15 p.m.) to MYDN (2:30 a.m.) (for all clear sky data, MD ranges from -0.3 °C to -1.04 °C; for only good data, from +1.17 °C to +0.26 °C). This indicates that the variation of LST at daytime is much higher than at nighttime (approximately 4 °C compared to 1 °C).

The G-dataset (of both Terra and Aqua) show a higher correlation with Ta than the CS-data (except MYDN with Ta_max). The exception of MYDN and Ta_max could be explained by the lowest percent (25.66%) of good data in the MYDN data (Table 4.1).
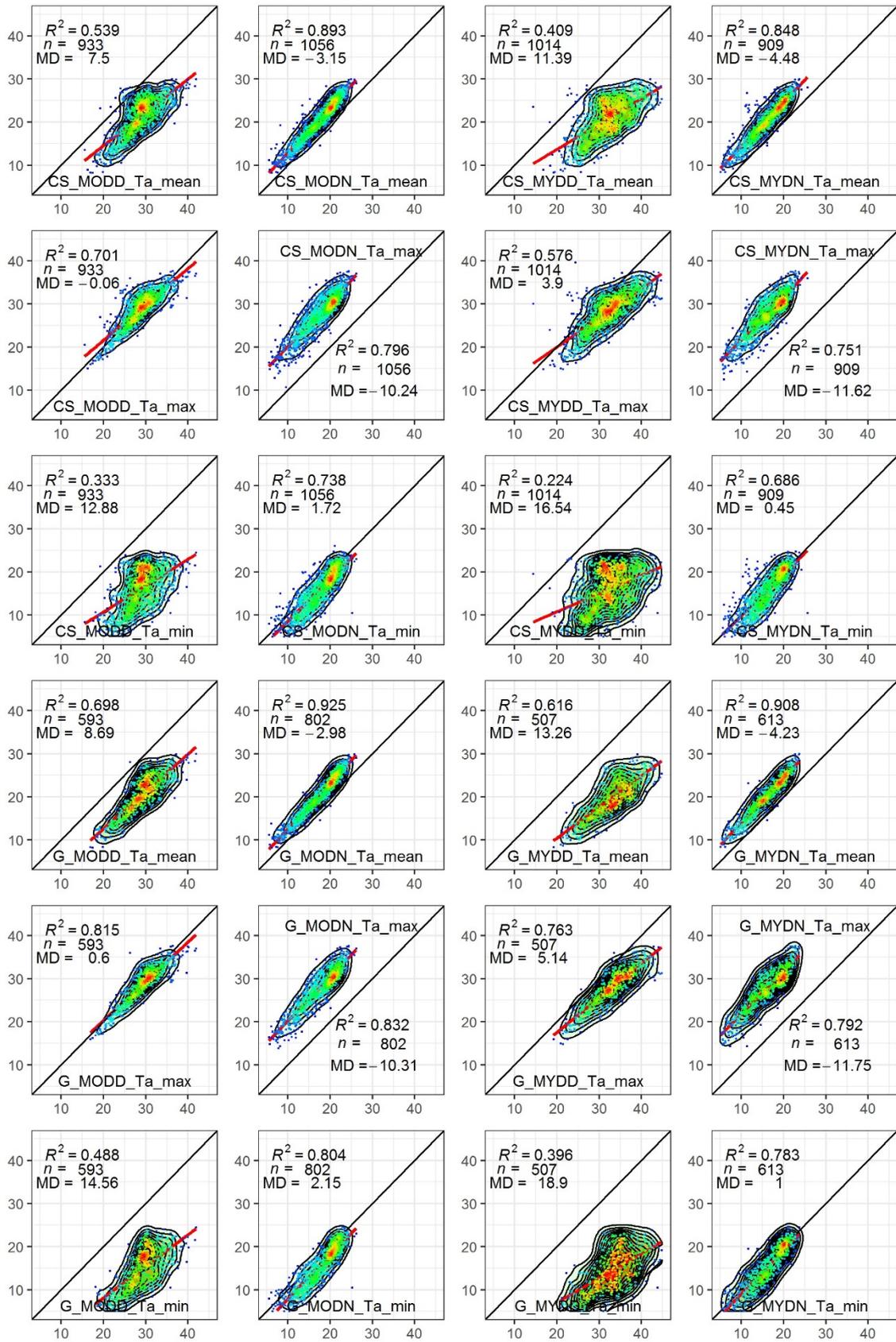
Figure 4.4. The point density plots of the LST (clear sky data/good data) products and Ta (Ta_mean, Ta_max, and Ta_min) at Conoi station. The point density from low to high is expressed by the colour ramp from blue to red.

95

At Conoi station, the rising absolute values of the MD for the two datasets (CS-data, G-data) were consistent with each other, and with Ta_mean, Ta_max, and Ta_min (Figure 4.4). This consistency could be explained by looking at Table 4.1, the percentage of good LST (over the total clear sky data) data ranged from 50 to 75.95%. The ascending order of Ta_mean, Ta_max, and Ta_min were MODN – MYDN – MODD – MYDD, MODD – MYDD – MODN – MYDN, and MYDN – MODN – MODD – MYDD.

Figure 4.3 and Figure 4.4 show that the difference between nighttime LST and Ta were consistent with respect to both stations. In general, all Ta (Ta_max, Ta_min, and Ta_mean) were higher than LST nighttime (i.e. MD > 0), while only MD of MYDN and Ta_min of the Muongte station was below zero. However, when comparing MYDN and Ta_min of CS-data and G-data (Figure 4.3), it is clearly seen that the inconsistent results were because of thin cloud pixels. In contrast, the difference between LST daytime and Ta were larger. At Muongte station, the largest difference was observed between MYDD and Ta_min (MD = +10.68 and +12.45 for CS-data and G-data, respectively), while the smallest differences were observed between MODD and Ta_mean (MD = +2.16 for CS-data), and between MYDD and Ta_max (MD = -1.81 for G-data). At Conoi station, the largest difference was observed between MYDD and Ta_min (MD = +16.54, and +18.90 for CS-data and G-data, respectively). The smallest difference was observed between MODD and Ta_max (MD = -0.06 for CS-data, and MD = +0.6 for G-data). MYDD was closest to Ta_max with MD = +3.9 for CS-data, and +5.14 for G-data.

4.3.2. MODIS LST data for Ta estimation

4.3.2.1. Muongte Station
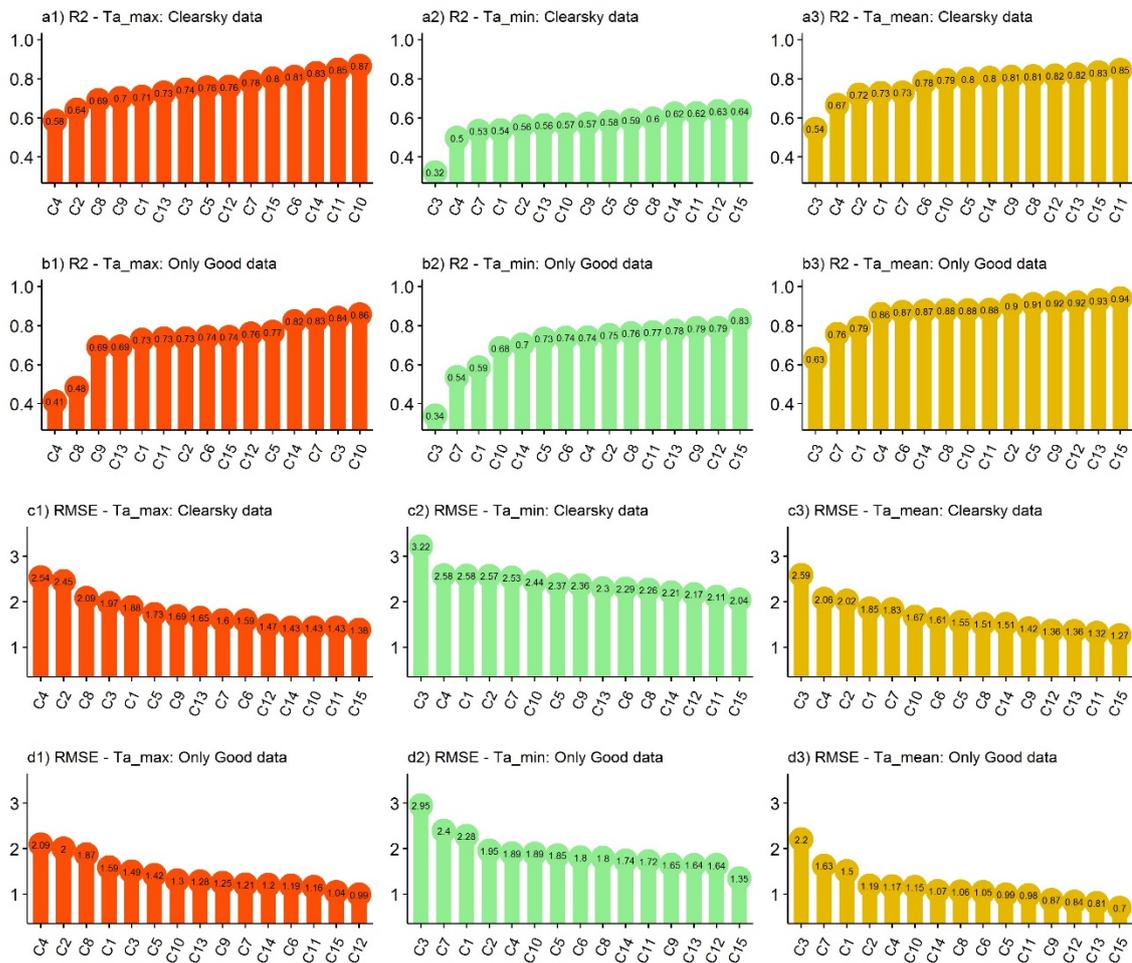
Figure 4.5. Averaged results for Ta_max, Ta_min, and Ta_mean estimation using two different quality conditions at Muongte station. The y-axis shows the value of R² and RMSE (°C), the x-axis shows the different combinations.

In general, the results of Ta_min and Ta_mean estimation using only good data were better (higher R^2, and lower RMSE) than when using all clear sky data. However, the results of Ta_max was similar when using all clear sky data or only good data. Although in this case, the results of using all clear sky data were slightly higher than of using good data only. It is suggested that the quality of valid MODIS LST data has the highest impact on Ta_min estimation, followed by Ta_mean estimation, and has less impact on Ta_max estimation.

As shown in Figure 4.5, the lowest results were achieved when using the C3 model (for Ta_min and Ta_mean estimation) and the C4 model (for Ta_max estimation). This is consistent with both all clear sky data and only good data. It is indicated that the overpass time of MODIS LST data has some impacts on the accuracy of Ta estimation.

4.3.2.2. Conoi station
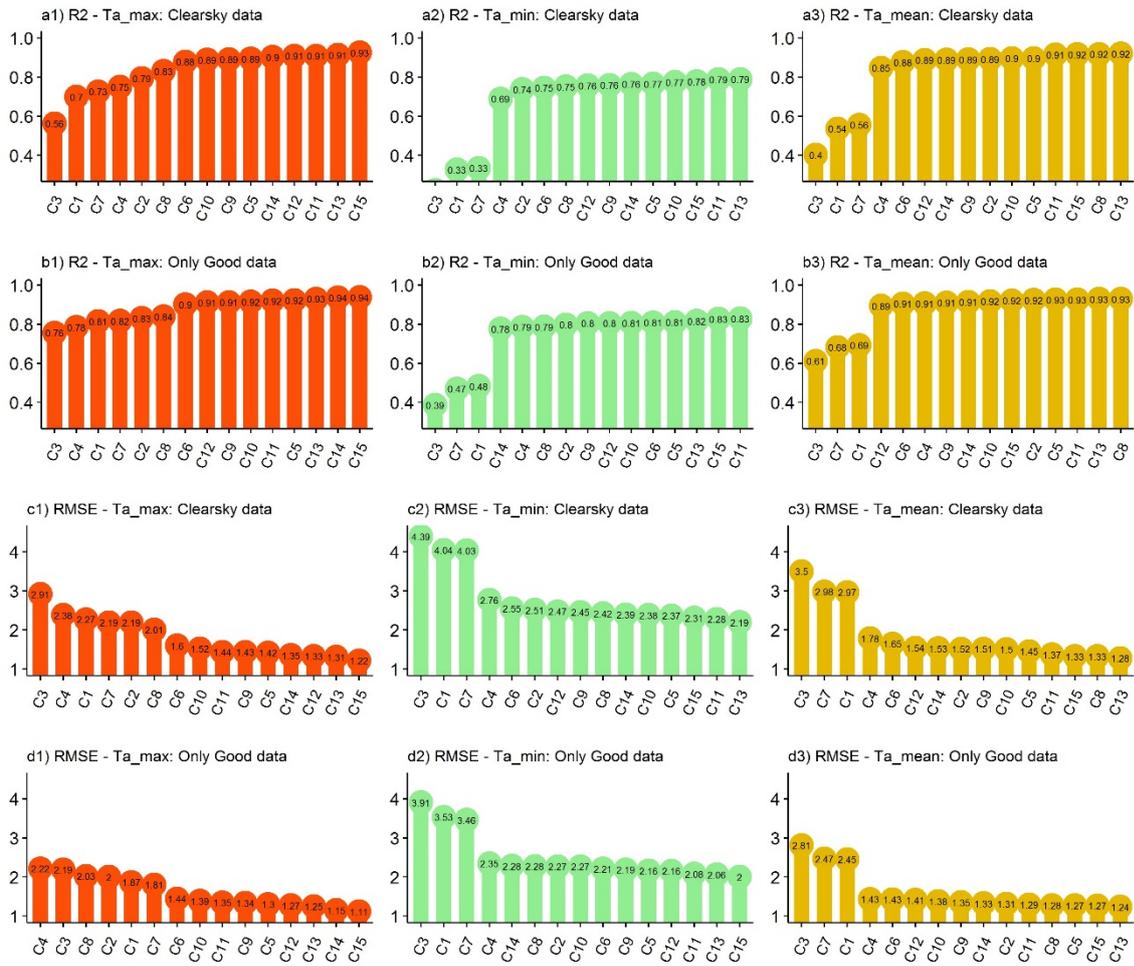
Figure 4.6. Averaged results for Ta_max, Ta_min, and Ta_mean estimation using two different quality conditions at Conoi station. The y-axis shows the value of R² and RMSE (°C), the x-axis shows the different combinations.

As shown in Figure 4.6, the results of Ta_max and Ta_min estimation for the Conoi station were different than for the Muongte station. The improvement of Ta_min estimation using only good data was not as high as at the Muongte station, and contrasting results were observed with Ta_max estimation. Unlike at the Muongte station, the results of Ta_max estimation at the Conoi station using only good data was higher than when using all clear sky data with all combinations.

In general, for all Ta estimation (Ta_max, Ta_min, and Ta_mean), the difference between the results of using all clear sky data and using only good data was largest at the Muongte station, which can be explained by the percentage of good data at Conoi station being higher than at Muongte station (Table 4.1). At this station, the results of Ta estimation can be divided into two groups of accuracy. For Ta_min and Ta_mean estimation, the lower accuracy group include C1, C3, and C7; while the higher accuracy group was consistent and

similar, containing the rest of the combinations. However, for Ta_max estimation, the lower accuracy group contains C1-C4, C7, and C8; while the higher accuracy group includes the rest of the combinations where the accuracy was consistent and similar (R^2 ranges from 0.88 to 0.93/0.90 to 0.94, RMSE reduces from 1.60 to 1.22/1.44 to 1.11 with all clear sky data/only good data, respectively).

4.3.2.3. Mixed-stations
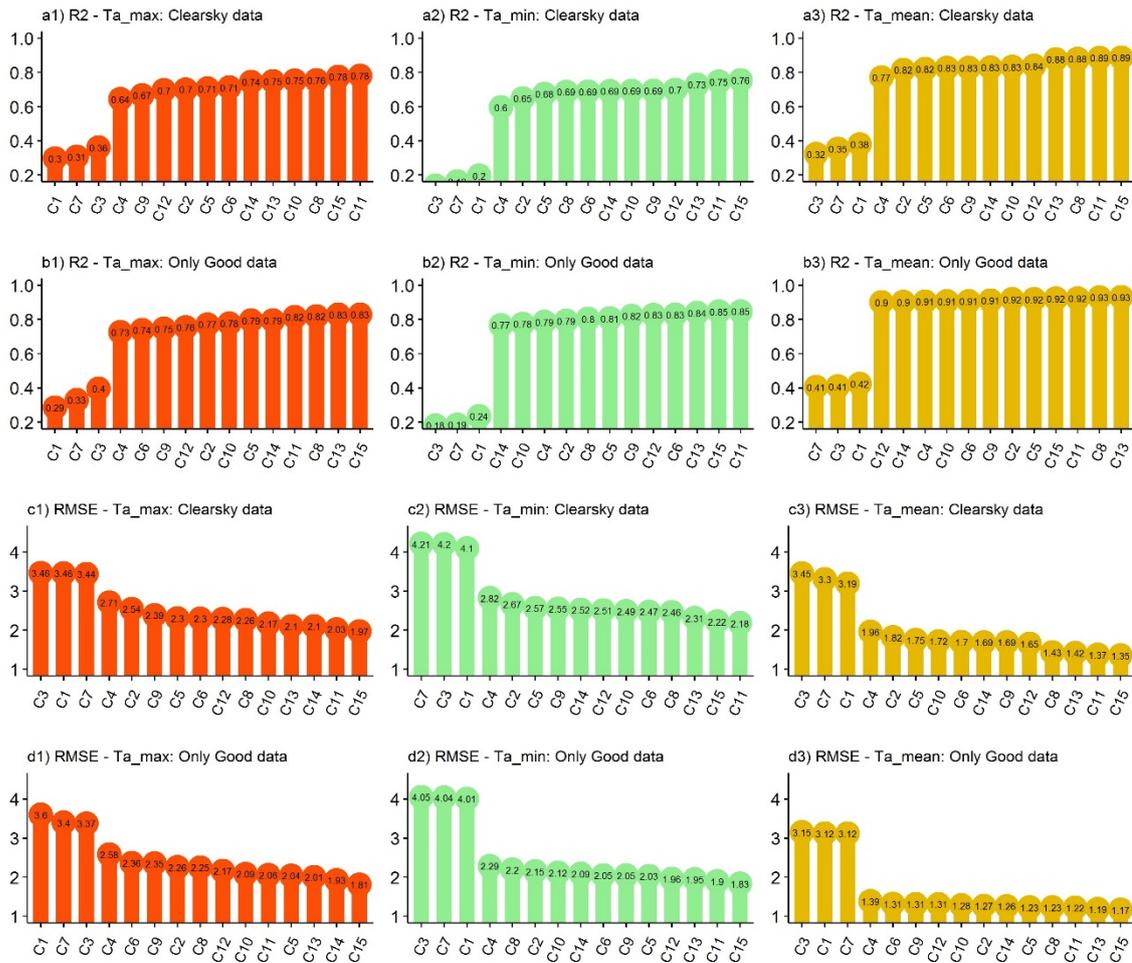
Figure 4.7. Averaged results for Ta_max, Ta_min, and Ta_mean estimation using two different quality conditions with mixed stations. The y-axis shows the value of R^2 and RMSE (°C), the x-axis shows the different combinations.

It can be clearly seen from Figure 4.7 that all the results of Ta estimation (Ta_max, Ta_min, and Ta_mean) with both data qualities (all clear sky data or only good data) can be divided into two groups, a lower accuracy group (containing C1, C3, and C7 combinations) and a higher accuracy group (containing the remaining combinations). Regarding the lower accuracy group, Ta_mean estimation achieved the highest accuracy followed by Ta_max and Ta_min. For the higher accuracy group, Ta_mean estimation always achieved the highest accuracy, however, there is a difference between the results of all clear sky data and only

good data for Ta_max and Ta_min estimations. With the all clear sky data, the results of Ta_max estimation were slightly higher those of Ta_min, and the contrasting results were observed with only good data (Figure 4.5). It also clearly seen that in the higher accuracy group, the difference between combinations of only good data was similar (R^2 ranges from 0.73 to 0.83, 0.77 to 0.85, and 0.90 to 0.93, and the RMSE reduces from 2.58 °C to 1.81 °C, 2.29 °C to 1.83 °C, and 1.39 °C to 1.17 °C; for Ta_max, Ta_min, and Ta_mean estimation, respectively). Meanwhile, the difference between combinations of all clear sky data is slightly larger in comparison to only good data (R^2 ranges from 0.64 to 0.78, 0.60 to 0.76, and 0.77 to 0.89, the RMSE reduces from 2.71 °C to 1.97 °C, 2.62 °C to 2.18 °C, and 1.96 °C to 1.35 °C; for Ta_max, Ta_min, and Ta_mean estimation, respectively). This indicates that the different combinations have more effect on the accuracy of Ta estimation when using all clear sky data than when using only good data.

4.4. Discussions

4.4.1. Overpass time, Correlations and Estimations

In the literature, numerous studies have chosen MODIS LST for Ta estimation based on the shortest time between overpass time of MODIS satellites and the occurrence of Ta_max or Ta_min time. For example, Xu et al. (2014) chose MYDD, which has an overpass time closest to the Ta_max, for Ta_max estimation in British Columbia, Canada. For the same reason, Alfieri et al. (2013) chose MODD for Ta_max estimation in southern Italy. However, LST time closest to the Ta_max or Ta_min does not always mean that it has the highest correlation with Ta_max or Ta_min, because it also depends on the land surface characteristic and other factors. In the present study, although MYDD (12:50 p.m.) has a much closer time with Ta_max occurrence than MODN (10:15 p.m.), the correlation of MYDD and MODN and Ta_max were different at the two stations. At Muongte station (Figure 4.3), MYDD had a higher correlation (R^2 = 0.743) with Ta_max than MODN (R^2 = 0.656); but at Conoi station (Figure 4.4), a contrasting result was observed, where the R^2 of MYDD and MODN with Ta_max were 0.576 and 0.796, respectively. This was consistent with all clear sky data and only good data. In addition, the result at Conoi station (nighttime LST has a higher correlation with Ta_max than daytime LST) was consistent with studies by Zhang et al. (2011) and Zeng et al. (2015). Consequently, correlation between overpass time of MODIS LST data and Ta is different from the station to station. In other words, having a closer overpass time to Ta does not necessarily yield a higher correlation. It is stated that

land surface characteristics or solar radiation have some impact on the relationship between LST and Ta.

Results also differ between stations with respect to the overpass time and accuracy of Ta estimation. Looking at the accuracy of Ta estimation using only MYDD (C3) and only MODN (C2) for Ta_max estimation, Muongte station (Figure 4.5) C3 produced higher accuracy of Ta estimation than C2, while at Conoi station, C2 produced higher accuracy than C3. This is consistent with both all clear sky data and only good data. It is worth noting that with mixed station data, the results of C2 were much higher than C3. This is similar to the study of Benali et al. (2012), which used data from 106 weather stations. Since Benali et al. (2012) used MODIS LST all clear sky data, the authors stated that the thin cloud cover pixels could decrease LST and incorrect modelling resulted in a low accuracy of Ta estimation. However, our results showed that with both datasets (all clear sky data and only good LST data), the model C2 always showed much higher accuracy than C3. It is suggested that the complexity in land cover, elevation, and solar radiation at daytime could have resulted in a low accuracy of Ta estimation rather than the LST quality.

4.4.2. Effect of LST Data Quality and Dynamic Combinations

From Figure 4.5, Figure 4.6, and Figure 4.7 it is clearly seen that for Ta_min and Ta_mean estimations, the results of only good data are always better than those of all clear sky data. However, for Ta_max estimation at Muongte station, most combinations (models) showed that the results of all clear sky were better than only good data. This result can be explained by looking at Table 4.1. All of the combinations that have a percentage of good LST (over all clear sky LST data) below 30% showed a worse result (i.e. lower R^2 and larger RMSE). In addition, at the Conoi station and the mixed station, the results of Ta_max estimation from only good data were slightly higher than the results from all clear sky data. This indicates the thin cloud LST data has a stronger impact on Ta_min and Ta_mean estimated accuracy than Ta_max. This is consistent with the study by Zhang et al. (2016a).

Comparing the results between the single stations and the mixed station for Ta_max, the results of the single stations (Muongte and Conoi) were always better than the mixed station with both all clear sky and only good data, indicating that land surface characteristics have a stronger impact on the Ta_max estimation accuracy than on Ta_min and Ta_mean estimation. This finding was also supported by looking at Figure 4.5 and Figure 4.6, where the results of Ta_max estimation were always better than Ta_min when using all clear sky

data. However, with the mixed station (Figure 4.7), the results of Ta_max and Ta_min were similar with both types of data (all clear sky and only good data).

4.5. Conclusions

Overpass time of each individual MODIS LST, thin cloud effects (LST quality), and dynamic combinations of MODIS LST collection 6 for daily Ta_max, Ta_min, and Ta_mean estimation at two separate stations as well as mixed stations in northern Vietnam for 10 years were analyzed in this study. Using observation data from two stations to validate the accuracy of Ta estimation using MODIS LST data showed that the average of RMSE (repeated 100 times) were 1.38/0.99 °C, 2.04/1.35 °C, and 1.27/0.70 °C; R^2 were 0.87/0.86, 0.64/0.83, and 0.85/0.94 for Ta_max, Ta_min, and Ta_mean estimation with all clear sky data/only good data at Muongte station, respectively. At Conoi station and mixed station, the results (in the same order) were 1.22/1.11 °C, 2.19/2.00 °C, and 1.28/1.24 °C; 0.93/0.94, 0.79/0.83, and 0.92/0.93; 1.97/1.81 °C, 2.16/1.83 °C, and 1.35/1.17 °C; 0.78/0.83, 0.76/0.85, and 0.89/0.93, respectively. Comparing each individual LST data showed that the overpass time has less impact on the performance of Ta estimation than land surface characteristics does. Terra LST has higher correlation with Ta than Aqua LST. With the same estimation models, the only good data showed a better performance than the all clear sky data for Ta_min and Ta_mean estimation. However, with Ta_max estimation, if the percentage of only good data (over all valid LST) was too low (i.e. less than 30%), all clear sky data would produce higher estimated Ta_max than only good data. Different combinations had less influence on the results of only good data than on the results of all clear sky data. Between Ta_max, Ta_min and Ta_mean; the different combinations had the strongest impact on Ta_max, followed by Ta_min and Ta_mean, with exception of the models with only daytime LST (C1, C3, and C7).

References

Alfieri, S. M., F. De Lorenzi, and M. Menenti. 2013. "Mapping air temperature using time series analysis of LST: the SINTESI approach." Nonlinear Processes in Geophysics 20 (4):513-27. doi: 10.5194/npg-20-513-2013.

Balaghi, Riad, Bernard Tychon, Herman Eerens, and Mohammed Jlibene. 2008. "Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco." International Journal of Applied Earth Observation and Geoinformation 10 (4):438-52. doi: 10.1016/j.jag.2006.12.001.

Bazrafshan, Javad. 2017. "Effect of Air Temperature on Historical Trend of Long-Term Droughts in Different Climates of Iran." Water Resources Management 31 (14):4683-98. doi: 10.1007/s11269-017-1773-8.

Benali, A., A. C. Carvalho, J. P. Nunes, N. Carvalhais, and A. Santos. 2012. "Estimating air surface temperature in Portugal using MODIS LST data." Remote Sensing of Environment 124:108-21. doi: 10.1016/j.rse.2012.04.024.

Cai, Yulin, Gang Chen, Yali Wang, and Li Yang. 2017. "Impacts of Land Cover and Seasonal Variation on Maximum Air Temperature Estimation Using MODIS Imagery." Remote Sensing 9 (3):233. doi: 10.3390/rs9030233.

Chen, H., and J. Sun. 2015. "Drought Response to Air Temperature Change over China on the Centennial Scale." Atmospheric and Oceanic Science Letters 8: 113–119.

Chen, Yunhao, Jinling Quan, Wenfeng Zhan, and Zheng Guo. 2016. "Enhanced Statistical Estimation of Air Temperature Incorporating Nighttime Light Data." Remote Sensing 8 (8):656. doi: 10.3390/rs8080656.

Dai, A. 2013. "Increasing drought under global warming in observations and models." Nature Climate Change, 3, 52–58.

de Wit, A. J. W., and C. A. van Diepen. 2008. "Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs." International Journal of Applied Earth Observation and Geoinformation 10 (4):414-25. doi: 10.1016/j.jag.2007.10.004.

Emamifar, Saeed, Ali Rahimikhoob, and Ali Akbar Noroozi. 2013. "Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree." International Journal of Climatology 33 (15):3174-81. doi: 10.1002/joc.3655.

Filippova, Marina G., V. Gopal Kakani, K. Raja Reddy, Roger L. King, and Georgy V. Mostovoy. 2006. "Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi." GIScience & Remote Sensing 43 (1):78-110. doi: 10.2747/1548-1603.43.1.78.

Goward, S. N., Waring, R. H., Dye, D. G., and Yang, J. 1994. "Ecological remote sensing at OTTER: Satellite macroscale observations." Ecological Applications, 4, 322–34

Huang, Fangfang, Weiqiang Ma, Binbin Wang, Zeyong Hu, Yaoming Ma, Genhou Sun, Zhipeng Xie, and Yun Lin. 2017. "Air temperature estimation with MODIS data over the Northern Tibetan Plateau." Advances in Atmospheric Sciences 34 (5):650-62. doi: 10.1007/s00376-016-6152-5.

Janatian, Nasime, Morteza Sadeghi, Seyed Hossein Sanaeinejad, Elham Bakhshian, Ali Farid, Seyed Majid Hasheminia, and Sadegh Ghazanfari. 2017. "A statistical framework for estimating air temperature using MODIS land surface temperature data." International Journal of Climatology 37 (3):1181-94. doi: 10.1002/joc.4766.

Keramitsoglou, Iphigenia, Chris Kiranoudis, Panagiotis Sismanidis, and Klemen Zakšek. 2016. "An Online System for Nowcasting Satellite Derived Temperatures for Urban Areas." Remote Sensing 8 (4):306. doi: 10.3390/rs8040306.

Kloog, Itai, Francesco Nordio, Johanna Lepeule, Andrea Padoan, Mihye Lee, Annick Auffray, and Joel Schwartz. 2017. "Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data." International Journal of Climatology 37 (1):296-304. doi: 10.1002/joc.4705.

Li, Zhao-Liang, Bo-Hui Tang, Hua Wu, Huazhong Ren, Guangjian Yan, Zhengming Wan, Isabel F. Trigo, and José A. Sobrino. 2013. "Satellite-derived land surface temperature: Current status and perspectives." Remote Sensing of Environment 131:14-37. doi: 10.1016/j.rse.2012.12.008.

Lin, Shengpan, Nathan J. Moore, Joseph P. Messina, Mark H. DeVisser, and Jiaping Wu. 2012. "Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa." International Journal of Applied Earth Observation and Geoinformation 18:128-40. doi: 10.1016/j.jag.2012.01.004.

Maeda, Eduardo Eiji. 2014. "Downscaling MODIS LST in the East African mountains using elevation gradient and land-cover information." International Journal of Remote Sensing 35 (9):3094-108. doi: 10.1080/01431161.2014.903442.

Mo, Kingtse C., and Dennis P. Lettenmaier. 2014. "Hydrologic Prediction over the Conterminous United States Using the National Multi-Model Ensemble." Journal of Hydrometeorology 15 (4):1457-72. doi: 10.1175/jhm-d-13-0197.1.

Niclos, R., Valiente, J.A., Barbera, M.J., and Caselles, V. 2014. "Land surface air temperature retrieval from EOS-MODIS images." IEEE Geoscience of Remote Sensing Letter, 11, 1380–1384.

Nieto, Héctor, Inge Sandholt, Inmaculada Aguado, Emilio Chuvieco, and Simon Stisen. 2011. "Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula." Remote Sensing of Environment 115 (1):107-16. doi: 10.1016/j.rse.2010.08.010.

Noi, Phan, Jan Degener, and Martin Kappas. 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 (5):398. doi: 10.3390/rs9050398.

Noi, Phan, Martin Kappas, and Jan Degener. 2016. "Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam." Remote Sensing 8 (12):1002. doi: 10.3390/rs8121002.

Oyler, Jared W., Ashley Ballantyne, Kelsey Jencso, Michael Sweet, and Steven W. Running. 2015. "Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature." International Journal of Climatology 35 (9):2258-79. doi: 10.1002/joc.4127.

Pape, Roland, and Jörg Löffler. 2004. "Modelling spatio-temporal near-surface temperature variation in high mountain landscapes." Ecological Modelling 178 (3-4):483-501. doi: 10.1016/j.ecolmodel.2004.02.019.

Peón, Juanjo, Carmen Recondo, and Javier F. Calleja. 2014. "Improvements in the estimation of daily minimum air temperature in peninsular Spain using MODIS land surface temperature." International Journal of Remote Sensing 35 (13):5148-66. doi: 10.1080/01431161.2014.935831.

Pichierri, Manuele, Stefania Bonafoni, and Riccardo Biondi. 2012. "Satellite air temperature estimation for monitoring the canopy layer heat island of Milan." Remote Sensing of Environment 127:130-8. doi: 10.1016/j.rse.2012.08.025.

Prihodko, L., and Goward, S.N. 1997. "Estimation of air temperature from remotely sensed surface observations." Remote Sensing of Environment, 60, 335–346.

Rabi, Anamarija, Marijana Hadzima-Nyarko, and Marija Šperac. 2015. "Modelling river temperature from air temperature: case of the River Drava (Croatia)." Hydrological Sciences Journal 60 (9):1490-507. doi: 10.1080/02626667.2014.914215.

Rajasekar, Umamaheshwaran, and Qihao Weng. 2009. "Urban heat island monitoring and analysis using a non-parametric model: A case study of Indianapolis." ISPRS Journal of Photogrammetry and Remote Sensing 64 (1):86-96. doi: 10.1016/j.isprsjprs.2008.05.002.

Shah, D. B., Pandya, M. R., Trivedi, H. J., and Jani, A. R. 2013. "Estimating Minimum and Maximum Air Temperature Using MODIS Data over Indo-Gangetic Plain." Journal of Earth System Science 122 (6): 1593–1605.

Sheffield, J., E. F. Wood, and M. L. Roderick. 2012. "Little change in global drought over the past 60 years." Nature 491 (7424):435-8. doi: 10.1038/nature11575.

Shi, Yi, Zhihong Jiang, Liangpeng Dong, and Suhung Shen. 2017. "Statistical estimation of high-resolution surface air temperature from MODIS over the Yangtze River Delta, China." Journal of Meteorological Research 31 (2):448-54. doi: 10.1007/s13351-017-6073-y.

Stisen, Simon, Inge Sandholt, Anette Nørgaard, Rasmus Fensholt, and Lars Eklundh. 2007. "Estimation of diurnal air temperature using MSG SEVIRI data in West Africa." Remote Sensing of Environment 110 (2):262-74. doi: 10.1016/j.rse.2007.02.025.

Tomlinson, Charlie J., Lee Chapman, John E. Thornes, and Christopher Baker. 2011. "Remote sensing land surface temperature for meteorology and climatology: a review." Meteorological Applications 18 (3):296-306. doi: 10.1002/met.287.

Vancutsem, Christelle, Pietro Ceccato, Tufa Dinku, and Stephen J. Connor. 2010. "Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa." Remote Sensing of Environment 114 (2):449-65. doi: 10.1016/j.rse.2009.10.002.

Wan, Zhengming. 2014. "New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product." Remote Sensing of Environment 140:36-45. doi: 10.1016/j.rse.2013.08.027.

Xu, Yongming, Anders Knudby, and Hung Chak Ho. 2014. "Estimating daily maximum air temperature from MODIS in British Columbia, Canada." International Journal of Remote Sensing 35 (24):8108-21. doi: 10.1080/01431161.2014.978957.

Xu, Y., Knudby, A., Shen, Y., and 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. doi: 10.1109/JSTARS.2017.2787191

Yang, Yuan, Wen Cai, and Jian Yang. 2017. "Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China." Remote Sensing 9 (5):410. doi: 10.3390/rs9050410.

Yoo, Cheolhee, Jungho Im, Seonyoung Park, and Lindi J. Quackenbush. 2018. "Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data." ISPRS Journal of Photogrammetry and Remote Sensing 137:149-62. doi: 10.1016/j.isprsjprs.2018.01.018.

Zeng, Linglin, Brian Wardlow, Tsegaye Tadesse, Jie Shan, Michael Hayes, Deren Li, and Daxiang Xiang. 2015. "Estimation of Daily Air Temperature Based on MODIS Land Surface Temperature Products over the Corn Belt in the US." Remote Sensing 7 (1):951-70. doi: 10.3390/rs70100951.

Zhang, Hongbo, Fan Zhang, Guoqing Zhang, Xiaobo He, and Lide Tian. 2016a. "Evaluation of cloud effects on air temperature estimation using MODIS LST based on ground measurements over the Tibetan Plateau." Atmospheric Chemistry and Physics 16 (21):13681-96. doi: 10.5194/acp-16-13681-2016.

Zhang, H., F. Zhang, M. Ye, T. Che, and, and Zhang, G. 2016b. "Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data." Journal of Geophysical Research: Atmospheres, 121, 1,425–411,441

Zhang, Wen, Yao Huang, Yongqiang Yu, and Wenjuan Sun. 2011. "Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures." International Journal of Remote Sensing 32 (24):9415-40. doi: 10.1080/01431161.2011.560622.

Zhou, Wang, Bin Peng, Jiancheng Shi, Tianxing Wang, Yam Dhital, Ruzhen Yao, Yuechi Yu, Zhongteng Lei, and Rui Zhao. 2017. "Estimating High Resolution Daily Air Temperature Based on Remote Sensing Products and Climate Reanalysis Datasets over Glacierized Basins: A Case Study in the Langtang Valley, Nepal." Remote Sensing 9 (9):959. doi: 10.3390/rs9090959.

Chapter 5

Noi, P.T.; Kappas, M.; Degener, J. **Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam**. Remote Sens. **2016**, 8, 1002.

Abstract

This study aims to evaluate quantitatively the land surface temperature (LST) derived from MODIS (Moderate Resolution Imaging Spectroradiometer) MOD11A1 and MYD11A1 Collection 5 products for daily land air surface temperature (Ta) estimation over a mountainous region in northern Vietnam. The main objective is to estimate maximum and minimum Ta (Ta-max and Ta-min) using both TERRA and AQUA MODIS LST products (daytime and nighttime) and auxiliary data, solving the discontinuity problem of ground measurements. There exist no studies about Vietnam that have integrated both TERRA and AQUA LST of daytime and nighttime for Ta estimation (using four MODIS LST datasets). In addition, to find out which variables are the most effective to describe the differences between LST and Ta, we have tested several popular methods, such as: the Pearson correlation coefficient, stepwise, Bayesian information criterion (BIC), adjusted R-squared and the principal component analysis (PCA) of 14 variables (including: LST products (four variables), NDVI, elevation, latitude, longitude, day length in hours, Julian day and four variables of the view zenith angle), and then, we applied nine models for Ta-max estimation and nine models for Ta-min estimation. The results showed that the differences between MODIS LST and ground truth temperature derived from 15 climate stations are time and regional topography dependent. The best results for Ta-max and Ta-min estimation were achieved when we combined both LST daytime and nighttime of TERRA and AQUA and data from the topography analysis.

Keywords: land surface temperature (LST); MODIS LST products; northern Vietnam

5.1. Introduction

Land air surface temperature (Ta, also called "air temperature" or "near surface temperature") data are usually collected as point data from weather station locations, typically at 2 m above the land surface. It is an important parameter in a wide range of fields, such as agriculture, e.g., crop evapotranspiration (de Bruin et al., 2010), crop yield prediction (Balaghi et al., 2008; de Wit et al., 2008), hydrology (Lehman, 2002; Vallet-Coulomb et al., 2001), ecology, environment and climate change (IPCC, 2001; IPCC, 2007). Generally, Ta values obtained from weather stations have a very high accuracy and temporal resolution (Fu et al., 2011), but do not capture information for a whole region and may therefore be unsuitable for spatial modelling applications (Stisen et al., 2007; Nieto et al., 2011; Lin et al., 2012).

In order to obtain Ta information for a region, researchers have proposed various methods of interpolation based on known weather station sites (Basist et al., 1998; Peterson et al., 2000; Florio et al., 2004). These interpolation methods' accuracy is highly dependent on the weather station network density, as well as the scale of spatial and temporal variability (Vogt et al., 1997; Vancutsem et al., 2010). Furthermore, station geometry and topography (elevation) change also affects the accuracy of interpolation, especially in regions with a wide range of elevation (Lai et al., 2012; Sun et al., 2014). However, the spatial distribution of weather stations is often limited in developing countries. Our study area of Vietnam has 170 surface meteorological observing stations, including 97 synoptic and 26 international exchange stations (Dinh, 2005), which is obviously inadequate for a country with an area of 331,688 km2 in which about 40% is mountainous, 40% hill and the remaining 20% lowland. Therefore, interpolation techniques may not be suitable for Vietnam.

Fortunately, remote sensing data provide a promising solution to overcome the limitation of interpolation techniques in mountainous areas and a sparsity of weather station areas. The successful launch of the Advanced Very High-Resolution Radiometer (AVHRR) in 1981 and the Moderate Resolution Imaging Spectroradiometer (MODIS) on board TERRA (December 1999) and AQUA (May 2002) has driven researchers to study new satellite-based methods, as a hot topic in recent years (Vogt et al., 1997; Coll et al., 1994; Wan and Dozier, 1996; Cresswell et al., 1999; Williamson et al., 2014; Prince et al., 1998; Mostovoy et al., 2006; Xu et al., 2012; Huang et al., 2015).

In recent years, there have been more and more studies employing land surface temperature (LST) obtained from remotely-sensed images for Ta estimation because of high spatial and temporal resolution, free availability and easy access. Particularly, MODIS on board TERRA and AQUA can provide daily LST data with high temporal (four times per day, TERRA LST daytime, TERRA LST nighttime, AQUA LST daytime, AQUA LST nighttime, which overpass local time at around 10:30 a.m., 10:30 p.m., 1:30 a.m. and 1:30 p.m., respectively) and very high spatial resolution (1 km) are widely applied. The difference between LST and Ta is strongly influenced by the surface characteristics and atmospheric conditions (Jin and Dickinson, 2010; Lin et al., 2016). In some regions, the difference between LST and Ta is high (Gallo et al., 2011; Lai et al., 2012). However, researchers from all over the world state that there is a strong linear correlation between MODIS LST and Ta over many regions, e.g., in Africa (Vancutsem et al., 2010), in Portugal (Benali et al., 2012), over the U.S. (Crosson et al., 2012; Zeng et al., 2015) and in Southeast Asia (Xu et al., 2012; Huang et al., 2015). The

detailed information of this difference, as well as the possible causes of this difference are still limited and need to be studied.

Some researchers (Benali et al., 2012; Zeng et al., 2015; Zaksek and Schroedter-Homscheidt, 2009) reviewed the types of commonly-used methods for Ta estimation based on LST. There are three main distinct types of methods:

The first type is the temperature-vegetation index method (TVX), which is based on the assumption that in an absolutely thick canopy, the temperature at the top of the canopy is the same as within the canopy (Prihodko and Goward, 1997); and there is a strong negative correlation between LST and the vegetation index, such as NDVI (Goetz, 1997; Goetz et al., 2000; Stisen et al., 2007; Nieto et al., 2011; Zhu et al., 2013; Sun et al., 2013). However, in some cases, this method is not satisfying due to the assumption that it often does not fit to the reality or the effect of seasonality, land cover type or soil moisture (Vancutsem et al., 2010; Sandholt et al., 2002).

The second type includes the surface energy-balance-based methods. The sum of in-coming net radiation and anthropogenic heat fluxes is considered equal to the sum of the surface's sensible and latent heat fluxes (Sun et al., 2005). The major drawback of these methods is that they require large amounts of information often not provided by remote sensing (Prince et al., 1998; Mostovoy et al., 2006).

The last type is using statistical methods that are based on regression techniques. These methods include various levels of complexity, from basic approaches that only use LST for Ta estimation (Vogt et al., 1997; Mostovoy et al., 2006) to advanced approaches that use more than one independent variable, such as elevation, NDVI, land cover, distance to water body, solar zenith angle, day length in hours, latitude and altitude (Vancutsem et al., 2010; Benali et al., 2012; Xu et al., 2012; Zhang et al., 2011; Good, 2015; Zhang et al., 2015). One of the biggest advantages of this method is that the systematic regional errors in satellite data can be reduced (Janatian et al., 2016).

The most recently popular studies of Ta estimation using statistic approaches are shown in Table 5.1.

Table 5.1. List of daily T_a temperature estimation studies using MODIS LST products in recent years.

Authors	Methods	Accuracy of T_{a-max}, T_{a-min} Estimation (°C)	Study Region
Vancutsem et al. (2010)	Statistical approach	RMSE = 2.1–2.76	Africa

Table 5.1. List of daily T_a temperature estimation studies using MODIS LST products in recent years.

Authors	Methods	Accuracy of $T_{a\text{-}max}$, $T_{a\text{-}min}$ Estimation (°C)	Study Region
Shen and Leptoukh (2011)	Statistical approach	Daily $T_{a\text{-}max}$: MAE: 2.4–3.2 Daily $T_{a\text{-}min}$: MAE: 3.0	Central and eastern Eurasia
Zhu et al. (2013)	TVX	$T_{a\text{-}max}$:RMSE = 3.79, MAE = 3.03, r = 0.83 $T_{a\text{-}min}$: RMSE = 2.97, MAE = 2.37, r = 0.94	Xiangride River Basin of China
Emamifar et al. (2013)	M5 model tree	Daily $T_{a\text{-}mean}$: RMSE = 2.3, r^2 = 0.96	Southwest of Iran
Xu et al. (2014)	Statistical approach	$T_{a\text{-}max}$: MAE: 2.02 ; r = 0.74	western Canada
Zeng et al. (2015)	Statistical approach	$T_{a\text{-}max}$: RMSE = 2.15–4.27, MAE = 1.71–3.35 $T_{a\text{-}min}$: RMSE = 1.75–5.13, MAE = 1.30–4.06	The Corn Belt over U.S.
Huang et al. (2015)	Statistical approach	Daily $T_{a\text{-}mean}$: RMSE = 2.41, MAE = 1.84	Central China

RMSE: root mean square error; MAE: mean absolute error; r: correlation coefficient; r^2: determination coefficients. TVX, temperature-vegetation index.

However, most of these studies have only used LST daytime and LST nighttime solely for Ta maximum (Ta-max) and Ta minimum (Ta-min) estimation, respectively. In a recent study (Zeng et al., 2015), both LST nighttime and daytime were used for Ta-max and for Ta-min estimation. However, this study was only applied for the growing season (May–September) from 2008–2012, and the elevation of the study site (the Corn Belt region of the Midwestern U.S.) ranging from 87–666 m and mostly covered by crops has a small vegetation index range.

There are several researchers who have studied the effect of the time of observation on the relationship between LST and Ta, but the conclusions are quite different in various time and geographical regions of the study areas. For instance, Mostovoy et al.(2006) found that the overpass time of TERRA and AQUA has little impact on the accuracy of Ta estimation in Mississippi State. Zhang et al.(2011) concluded that for daily Ta estimation, TERRA LST and AQUA LST give the same results. Benali et al.(2012) showed that the use of both AQUA LST daytime and LST nighttime gives better accuracy of Ta-max and Ta-min estimation, respectively, than TERRAs in Portugal. In contrast, Zhu et al.(2013) stated that TERRA LST daytime and TERRA nighttime were better than AQUA LST daytime and nighttime for Ta estimation. These differences could be understood as not only the time of observation, but also geographical location affecting the relationship between LST products and Ta and, therefore, affecting the accuracy estimation of Ta based on LST products. Because of all of these different conclusions about the relationship between Ta and LST of TERRA and

AQUA, in this study, we analyzed the relationship between Ta with both TERRA LST and AQUA LST products.

In addition, as far as we know, there are no studies over Vietnam that have integrated both TERRA and AQUA LST of daytime and nighttime for Ta estimation (using all four MODIS LST datasets).

The main objective of this research was to estimate daily Ta (maximum and minimum) using both TERRA and AQUA MODIS LST products (daytime and nighttime) and auxiliary data, solving the discontinuity problem of ground measurements. In addition, to find out which variables, among 14 predefined variables, are the most effective to describe the relationship between LST and Ta, we have tested several methods, such as: the Pearson correlation coefficient, forward selection, backward elimination, stepwise, Bayesian information criterion (BIC), adjusted R-squared and the principal component analysis (PCA) of 14 variables (including: LST products (four variables), NDVI, elevation, latitude, longitude, day length in hours, Julian day and four variables of the view zenith angle). Finally, we applied nine models for Ta-max and nine models for Ta-min estimation.

5.2. Materials and Methods

5.2.1. Study Area

The study area is located in northern Vietnam and covers more than 37,000 km2, comprising the provinces of Hoa Binh, Ha Noi, Vinh Phuc, Thai Nguyen, Yen Bai, Phu Tho and Son La (see Figure 5.1). The study area extends from 20°18'N–22°40'N and from 103°12'E–106°18'E. The elevation ranges from sea level to over 3000 m.
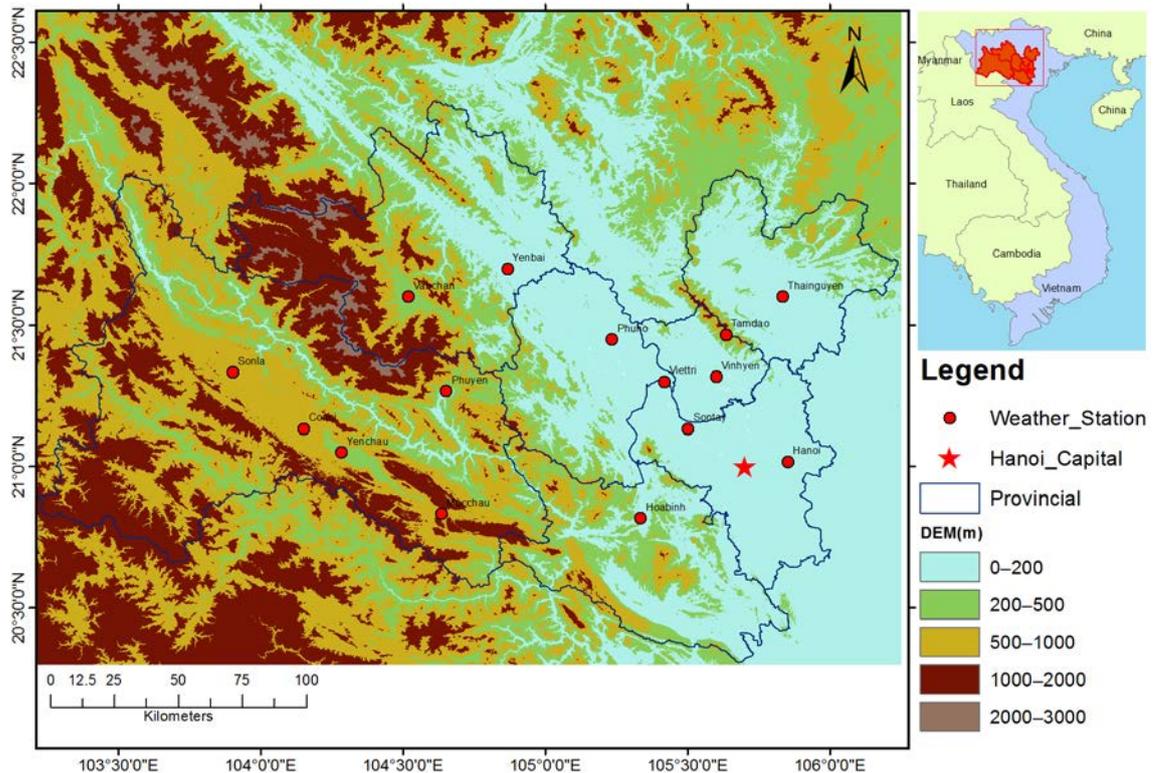
Figure 5.1. Location of the 15 meteorological stations in northern Vietnam and the range of elevations in the study area.

The northern Vietnam study site was chosen for the following reasons:

(1)   Good distribution of meteorological station network in comparison to southern Vietnam.

(2)   Wide range of elevation (from approximately sea level to more than 3000 m).

(3)   The spatial heterogeneity of land cover.

(4)   There is no study about Ta estimation using all 4 MODIS LST datasets in northern Vietnam.

In this area, the topography is quite complex, increasing from southeast to northwest, and mostly divided into two regions: lowland and highland. The low part includes: Hoabinh, Hanoi, Phutho, Vinhphuc and Thainguyen provinces. The high part includes 2 large provinces: a part of Yenbai and Sonla.

The climate and weather in this area also vary depending on the elevation and type of landscape. The humidity is high, with the average ranging around 84% a year.

In late October and early November 2008, there was torrential rain in northern and central Vietnam that triggered severe floods in this region (NASA Visible Earth).

The location and elevation of meteorological stations are shown in Table 5.2.

Table 5.2. Geographical description of weather stations used in this study.

Weather Station	Latitude (°)	Longitude (°)	Elevation (m)
Conoi	21.13	104.15	671
Hanoi	21.02	105.80	6
Hoabinh	20.82	105.33	23
Mocchau	20.83	104.68	972
Phuho	21.45	105.23	54
Phuyen	21.27	104.63	169
Sonla	21.33	103.90	675
Sontay	21.13	105.50	16
Tamdao	21.47	105.65	934
Thainguyen	21.60	105.83	35
Vanchan	21.58	104.52	275
Viettri	21.30	105.42	30
Vinhyen	21.32	105.60	10
Yenbai	21.70	104.87	56
Yenchau	21.05	104.30	314

5.2.2. Data

5.2.2.1. Remote Sensing Data

• MODIS LST Data

Two MODIS LST products (h27v06, Collection 5, from 2003–2013, over northern Vietnam) were used in this study: MOD11A1 daily land surface temperature and emissivity from the TERRA satellites and MYD11A1 daily land surface temperature and emissivity from the AQUA satellites. There exist 4 LST data records per day, two from the TERRA satellites and two others from the AQUA satellites, which pass over the study site (local solar time) mostly around 10:30 a.m., p.m. and 1:45 a.m. and p.m., respectively. These times are relatively close to the times of maximum and minimum Ta daily data. In total, there were more than 8000 images (in HDF format, from 1 January 2003–31 December 2013) downloaded from the U.S. Geological Survey (http://earthexplorer.usgs.gov). The MODIS LST is generated using a split-window algorithm (Wan et al., 2002) with two thermal infrared bands, 31 (10.78–11.28 µm) and 32 (11.77–12.27 µm). These two products (MOD11A1 and MYD11A1) have been validated, and the accuracy was reported better than 1 K under clear sky conditions (http://landval.gsfc.nasa.gov).

• Elevation

In this study area, the elevation is quite complex; it ranges from sea level to above 3000 m (see Figure 5.1). Generally, higher elevations are associated with lower temperatures. Elevation data were obtained from ASTER Global DEM. This data are available from the U.S. Geological Survey (USGS) with a spatial resolution of 30 m. These elevation data were resized to 1-km resolution using the nearest neighbor resampling type in order to be associated with other data resolutions, such as MODIS LST or NDVI data.

• Vegetation Based on NDVI

Several studies have shown that the vegetation cover on the surface can affect LST values (Shi et al., 2016). Vegetation cover (NDVI) data are provided every 16 days at 1-kilometer spatial resolution from the MODIS satellite, including MOD13A2 (TERRA 16-days period starting Day 001) and MYD13A2 (AQUA 16-day period starting Day 009). There are some studies that use both MOD13A2 and MYD13A2 in order to composite 8-day period data by averaging these two products (Sun et al., 2014). Miura and Yoshioka (2014) stated that MOD13A2 and MYD13A2 could be interchangeable. We assumed that NDVI did not change within 16 days. Therefore, in this study, to correct the influence of vegetation, we used NDVI 16 days data at 1-km resolution obtained from the MOD13A2 product.

5.2.2.2. *Meteorological Data*

Daily maximum and minimum air temperature data have been collected from 15 meteorological stations in northern Vietnam (see Figure 5.1), from 2003–2013. The data were obtained from the Vietnam Institute of Meteorology, Hydrology and ENvironment (IMHEN).

5.2.2.3. Auxiliary Data

Day length (Dlth) is the total time that any portion of the Sun is above the horizon. Typically (for low and mid-latitude locations), this will be the elapsed time beginning at sunrise and ending at sunset (http://aa.usno.navy.mil/data/). In this study, the day length (in hours) was downloaded from the Astronomical Applications Department of the U.S. Naval Observatory website (http://aa.usno.navy.mil/data/).

The Julian day (Jd) was extracted from NASA server (http://landweb.nascom.nasa.gov/). Both Julian day and day length are proxies for the fraction of solar energy absorption during the day and emitted energy during the night, influencing the diurnal amplitude of Ta throughout the year. Jang et al. (2004) showed that the Julian day was a more significant parameter than altitude or the solar zenith angle in the inter-seasonal estimation of Ta. In

order to understand the effect of viewing angle in the temporal variations in LST, especially in rugged regions, the view zenith angle (VZA) of daytime and nighttime should be taken into consideration (Wan, 2006). In this study, there are four types of VZA of TERRA daytime, TERRA nighttime, AQUA daytime, AQUA nighttime (VZAtd, VZAtn, VZAad, VZAan, respectively). These data were collected from MOD11A1 and MYD11A1 products. We also took the effects of the latitude, longitude and elevation of each station, collected from IMHEN, into consideration to estimate Ta accuracy. All key terms and their explanations are summarized in Table 5.3.

Table 5.3. Description of the key terminology used in this study.

Used Terms	Description
T_a (°C)	Land air surface temperature
$T_{a\text{-max}}$ (°C)	Daily Maximum T_a
$T_{a\text{-min}}$ (°C)	Daily Minimum T_a
LST (°C)	Land surface temperature
LST_{td} (°C)	TERRA LST daytime
LST_{tn} (°C)	TERRA LST nighttime
LST_{ad} (°C)	AQUA LST daytime
LST_{an} (°C)	AQUA LST nighttime
$T_{a\text{-min-es}}$ (°C)	Daily minimum T_a estimation
$T_{a\text{-max-es}}$ (°C)	Daily maximum T_a estimation
Ele (m)	Elevation of weather stations
NDVI	Normalized Difference Vegetation Index
Long (°)	Longitude
Lat (°)	Latitude
Dlth (hours)	Day length in hours
Jd	Julian day
VZA_{td}	View zenith angle of TERRA daytime
VZA_{tn}	View zenith angle of TERRA nighttime
VZA_{ad}	View zenith angle of AQUA daytime
VZA_{an}	View zenith angle of AQUA nighttime

5.2.3. Preprocessing Data

There were 8036 files of MODIS HDF format (MOD11A1 and MYD11A1, Version 5) from 1 January 2003–31 December 2013 that were downloaded for this study. Because all MODIS data were downloaded from USGS in HDF format (Hierarchical Data Format), with each file containing 12 data layers (Wan, 2006), we used the MODIS Reprojection Tool to extract the corresponding bands (LST_Day_1km, LST_Night_1km and view zenith angle) from MOD11A1 and MYD11A1. Next, we used the layer-stacking tool in Envi5.0 to stack images

for time series analysis purposes, before clipping the images to fit to our study area using ArcGIS 10.1.

5.2.4. LST Data Retrieval at Weather Stations

LST data under clear sky conditions at weather stations were retrieved by the following steps:

(1) MODIS LST day and night was extracted from MOD11A1 and MYD11A1 data for the pixel in which the meteorological stations are located. It should be noted that in the MODIS LST (v005) product, LST values derived from a single clear-sky MODIS observation are selected from MOD11_L2 files if the viewing zenith angles are small or larger zenith angles have LST errors smaller than 2 K (Wan, 2006). This means that MOD11A1 and MYD11A1 LST were not produced for pixels that are classed as cloudy or cloud contamination remaining.

(2) However, Wan (2006) also noted that pixels that are lightly or modestly cloud contaminated are not removed by using cloud-removing masks if the contamination is smaller than the normal temporal variations in clear-sky LST. Typically, these are thin and subpixel clouds that are difficult to detect by the algorithm. In this case, we applied two steps to remove those types of errors. First, we simply filter and remove all unrealistic LST data that had values greater than 100 °C and below −50 °C. Second, we calculated the difference between Ta-max versus LST daytime and Ta-min versus LST nighttime. Then, we applied a statistic outlier removal based on these differences' histograms to detect and remove data with unusually large differences.

(3) All valid LST day and nighttime data of TERRA and AQUA were statistical compared with Ta-max and Ta-min, respectively.

5.2.5. Estimation of Land Air Surface Temperature Using MODIS LST and Auxiliary Data

The method for this study was based on a multivariate linear regression analysis. This method was chosen because it could be applied for sparse meteorological station networks like northern Vietnam. Although some interpolation methods may give a higher accuracy result, they are not possible for regions with poorly-distributed station networks (Kilibarda et al., 2014); or physical models (Sun et al., 2005) that require an unreliable amount of input data.

5.2.5.1. Variable Selection

Zaksek and Schroedter-Homscheidt (2009) stated that Ta is driven more by LST than by direct solar radiation, meaning that LST is the most important variable for Ta estimation; in reference to previous studies (Vancutsem, 2010; Benali et al., 2012; Zeng et al., 2015; Zhang et al., 2011; Jang et al., 2004), also in consideration of all of the available data that potentially have an effect on the accuracy of Ta estimation of the study area. We have collected 14 variables, including: LST products (4 variables), NDVI, elevation (DEM), latitude, longitude, day length in hours and Julian day, and 4 variables of the view zenith angle as the potential variables (candidate variables) for Ta estimation.

Besides the 4 LST variables, we chose NDVI because it influences the land surface vegetation properties. Elevation, latitude and longitude were chosen for capturing the variability of climatic conditions between different regions. Day length in hours was chosen because we considered that it would affect the received solar radiation. Following (Jang et al., 2004; Janatian et al., 2016), we chose Julian day because it reflects seasonal variation in air temperature. The view zenith angle was considered as a factor affecting the accuracy of LST data.

Since we have 14 potential variables (predictors), there are 2^{14} (= 16,384) possible models that could be applied for Ta-max and Ta-min estimation. Clearly, we cannot apply all of these possible models and test them one by one. Consequently, to save the time of data processing and to find the smallest model that best fits the data, we selected the best subset of variables among the 14 available variables.

In this study, we used several popular methods for variable selection, such as: (1) the Pearson correlation coefficient; (2) stepwise; (3) Bayesian information criterion (BIC); (4) adjusted R-squared; and (5) principal component analysis (PCA).

• Pearson Correlation Coefficient

To select the variables for the multivariate linear regression models, we calculated the Pearson's correlation coefficient (r) of all related variables. According to (Lin et al., 2012), r close to ±1 indicates a strong correlation; r = 0 means that there is no correlation; $0.25 \leq r \leq 0.75$, means there is a moderate degree of correlation and $0 \leq r < 0.25$, a weak correlation. Based on this, we took only variables with r greater than 0.25 for Ta estimation.

The results in Table 5.4 show that we should use 7 variables for Ta-min estimated models (4 LST, Ele, Long and Dlth); and 7 other variables for Ta-max estimated models (4 LST, Ele, Dlth and Jd).

Table 5.4. Pearson correlation coefficients of all variables considered in models for daily Ta-max and Ta-min estimation.

	LST_{td}	LST_{tn}	LST_{ad}	LST_{an}	NDVI	Ele	Lon	Lat	Dlth	Jd	VZA_{td}	VZA_{tn}	VZA_{ad}	VZA_{an}
$T_{a\text{-}min}$	0.68	0.92	0.61	0.92	0.03	−0.29	0.27	−0.01	0.75	−0.17	0.00	0.01	−0.03	−0.01
$T_{a\text{-}max}$	0.85	0.86	0.86	0.82	−0.18	−0.28	−0.10	−0.08	0.68	−0.36	−0.01	0.03	−0.03	−0.03

* All variables are described in Table 5.3

● Variable selection based on adjusted R-squared ($\overline{R}2$) and BIC:

$$\overline{R}^2 = 1 - (1 - r^2)\frac{N-1}{N-k-1},$$

(1)

The best model will be the one with the largest value of $\overline{R}^2$:

$$BIC = N\log\left(\frac{SSE}{N}\right) + (k+2)\log(N),$$

(2)

Minimizing the BIC is intended to give the best model.

N is the number of observations; k is the number of predictors (variables); r^2 is the coefficient of determination; and SSE is the sum of squared error.

In Figure 5.2, colored parts (green for $\overline{R}2$ and orange for BIC) indicate that a variable was included in the model and white color that a variable was not included in the model. All of these models include the intercept.
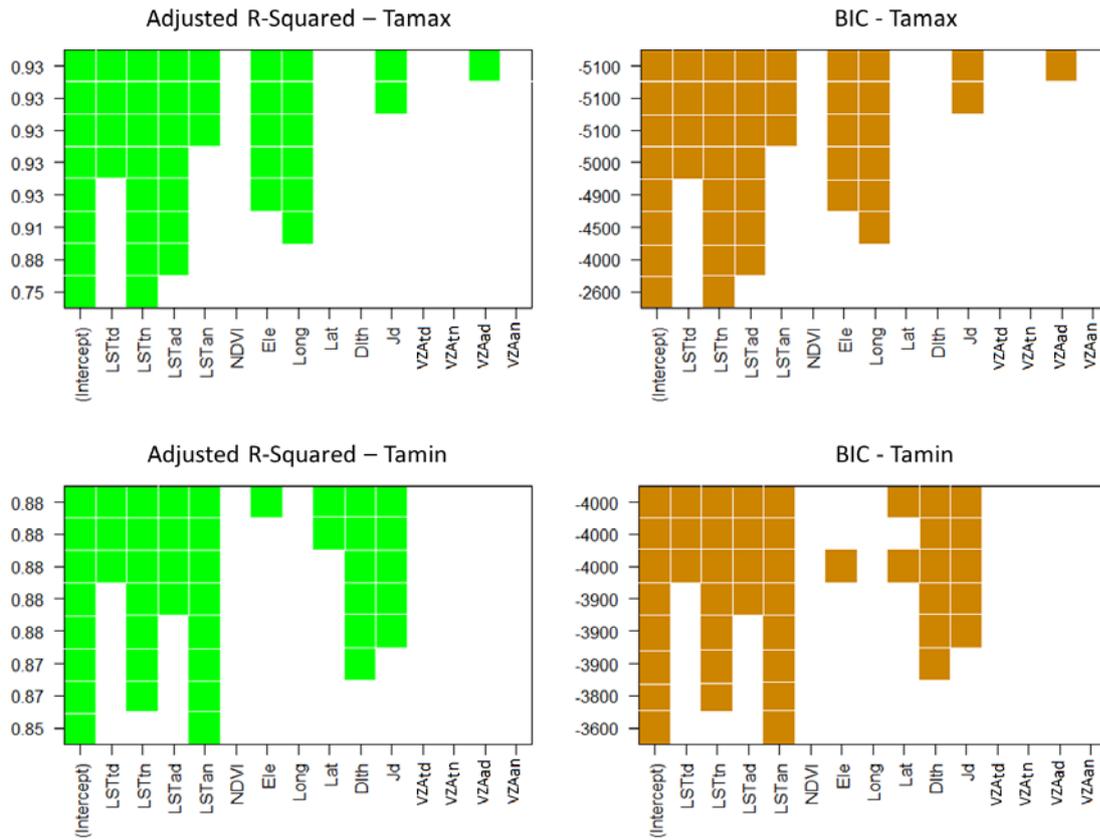
Figure 5.2. Ranking of models based on $\overline{R}2$ and BIC criteria for Ta-max (upper row panel) and Ta-min (lower row panel) estimation.

Looking at the y-axis of Figure 5.2, it can be clearly seen that, in both cases of $T_{a\text{-max}}$ and $T_{a\text{-min}}$, the top three models have the same results of $\overline{R}^2$ and BIC.

In this case, we chose all top 3 models for T_a estimation. These model forms are as follows:

$$T_{a\text{-max}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g \times Jd + h \times VZA_{ad} + i, \qquad (3)$$

$$T_{a\text{-max}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g \times Jd + h, \qquad (4)$$

$$T_{a\text{-max}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g, \qquad (5)$$

$$T_{a\text{-min}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Lat + g \times Jd + h \times Ele + i, \qquad (6)$$

$$T_{a\text{-min}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Lat + g \times Jd + h, \qquad (7)$$

$$T_{a\text{-min}} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Jd + g. \qquad (8)$$

• Variable Selection Using Forward, Backward and Stepwise

Another popular method that is used for variable selection is the stepwise method. In this method, there are three types of operative steps: forward, backward and stepwise.

Forward selection starts with no variable in the model (intercept only model). In the next steps, the most significant variables are added to the model one by one. The process stops when all of the variables not in the model have a p-value greater than 0.15.

Backward elimination starts with the model including all variables. In the next step, the least significant variable will be removed. The procedure continues until all of the variables in the model have p-values less than or equal to 0.15.

The stepwise method adds or removes a variable in each step, depending on its p-value. This process continues until all variables within the model have a p-value ≤0.15, and all of the variables that were not in the model have a p-value >0.15.

Based on the results of forward, backward and stepwise selection, the Ta-max estimated model should include 12 variables (using all variables except VZA$_{tn}$ and VZA$_{an}$); the Ta-min estimated model should include 10 variables (except long, VZA$_{an}$, VZA$_{td}$, VZA$_{tn}$).

The Ta-max and Ta-min estimated models were as follows:

$$T_{a-max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times NDVI + f \times Ele + g \times Long + h \times Lat + i \times Dlth + j \times Jd + k \times VZA_{td} + l \times VZA_{ad} + m, \tag{9}$$

$$T_{a-min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times NDVI + f \times Ele + g \times Lat + h \times Dlth + i \times Jd + j \times VZA_{ad} + k, \tag{10}$$

• Variable Selection-Based Principal Component Analysis (PCA)

The result of the PCA showed that both Ta-max and Ta-min estimation models should include 4 LST, elevation, day length in hours and Julian day.

$$T_{a-max} = T_{a-min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Dlth + g \times Jd + h, \tag{11}$$

From all of the analyses above (Pearson correlation, $\overline{R}^2$, BIC, stepwise and PCA), it is indicated that 4 LSTs (LST$_{td}$, LST$_{tn}$, LST$_{ad}$, LST$_{an}$) play an important role in Ta-max and Ta-min estimation, because they are always included in the top models of all above-mentioned methods.

To identify which LST data (LST$_{td}$, LST$_{tn}$, LST$_{ad}$, LST$_{an}$) are the best variables for Ta-max and Ta-min estimation, we have tested the $\overline{R}^2$, BIC and stepwise analysis within 4 LST product. These results gave us 4 models (Models 1–4 for Ta-max, Models 10–13 for Ta-min estimation) in Table 5.5.

Table 5.5. Model equations for T$_{a-max}$ (1–9) and T$_{a-min}$ (10–18) estimations.

	Models for T$_{a-max}$ estimations (*)		Models for T$_{a-min}$ estimations
1	$T_{a-max} = a \times LST_{tn} + b$	10	$T_{a-min} = a \times LST_{an} + b$
2	$T_{a-max} = a \times LST_{tn} + b \times LST_{ad} + c$	11	$T_{a-min} = a \times LST_{tn} + b \times LST_{an} + c$
3	$T_{a-max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d$	12	$T_{a-min} = a \times LST_{tn} + b \times LST_{ad} + c \times LST_{an} + d$
4	$T_{a-max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e$	13	$T_{a-min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e$

5	$T_{a\text{-}max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g$	14	$T_{a\text{-}min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Jd + g$
6	$T_{a\text{-}max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g \times Jd + h$	15	$T_{a\text{-}min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Jd + g \times Lat + h$
7	$T_{a\text{-}max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Long + g \times Jd + h \times VZA_{ad} + i$	16	$T_{a\text{-}min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Dlth + f \times Jd + g \times Lat + h \times Ele + i$
8	$T_{a\text{-}max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times NDVI + f \times Ele + g \times Long + h \times Lat + i \times Dlth + j \times Jd + k \times VZA_{td} + l \times VZA_{ad} + m$	17	$T_{a\text{-}min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times NDVI + f \times Ele + g \times Lat + h \times Dlth + i \times Jd + j \times VZA_{ad} + k$
9	$T_{a\text{-}max} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Dlth + g \times Jd + h$	18	$T_{a\text{-}min} = a \times LST_{td} + b \times LST_{tn} + c \times LST_{ad} + d \times LST_{an} + e \times Ele + f \times Dlth + g \times Jd + h$

From Equations (3), (4), (5), (9) and (11), we build Models 5–9 for Ta-max estimation; Models 14–18 of Ta-min estimation were built based on Equations (6–8), (10) and (11).

We evaluated the models using the coefficient of determination (r^2), the root mean square error (RMSE) and the mean absolute error (MAE). Because RMSE is very sensitive to outliers, hence MAE was chosen as an additional measure of the model quality.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Ta,i - Tes,i)^2}, \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Ta,i - Tes,i|, \tag{4}$$

where Ta,i is the observed land air surface temperature from weather stations and Tes,i is the corresponding land air surface temperature estimated using linear regression analysis methods.

5.2.5.2. Model Calibration and Validation

For calibration and validation purposes, all data were randomly divided into two parts: calibration and validation datasets using a 70%/30% proportion, respectively.

To calculate the model coefficients a–m (see Table 5.5) for the models, we applied a least squares fitting analysis to the calibration dataset. The constant coefficients are listed in Appendix A and Appendix B. To validate the models, we applied the constant coefficients, which were calculated based on the calibration dataset, to the validate dataset.

5.3. Results

5.3.1. The Relationship between MODIS LST and Ta

In order to analyze the relations between Ta and all MODIS LST data, we calculated the coefficient of determination (r^2), root mean square error (RMSE) and mean absolute error (MAE) of each type of MODIS LST (TERRA daytime, TERRA nighttime, AQUA daytime and AQUA nighttime) solely with Ta measured from weather stations.

Figure 5.3a shows that LST nighttime of both TERRA and AQUA showed a stronger correlation with Ta (both Ta-max and Ta-min) than for LST daytime. This can be explained by the fact that during the night, solar radiation does not affect the thermal infrared signal. Looking at the overpasses time of the satellite, the overpass time of AQUA daytime (around 1:45 p.m.) is later than the overpass time of TERRA (10:30 a.m.), as more solar radiation had been received by the land surface. As a result, LST AQUA daytime is hotter than LST TERRA daytime. Similarly, the overpass time of TERRA LST nighttime is around 10.30 p.m., three hours earlier than that of AQUA (1.45 a.m.), and the LST nighttime of TERRA is slightly higher than that of AQUA. This could be reflected by the cooling process of the land's surface at night.
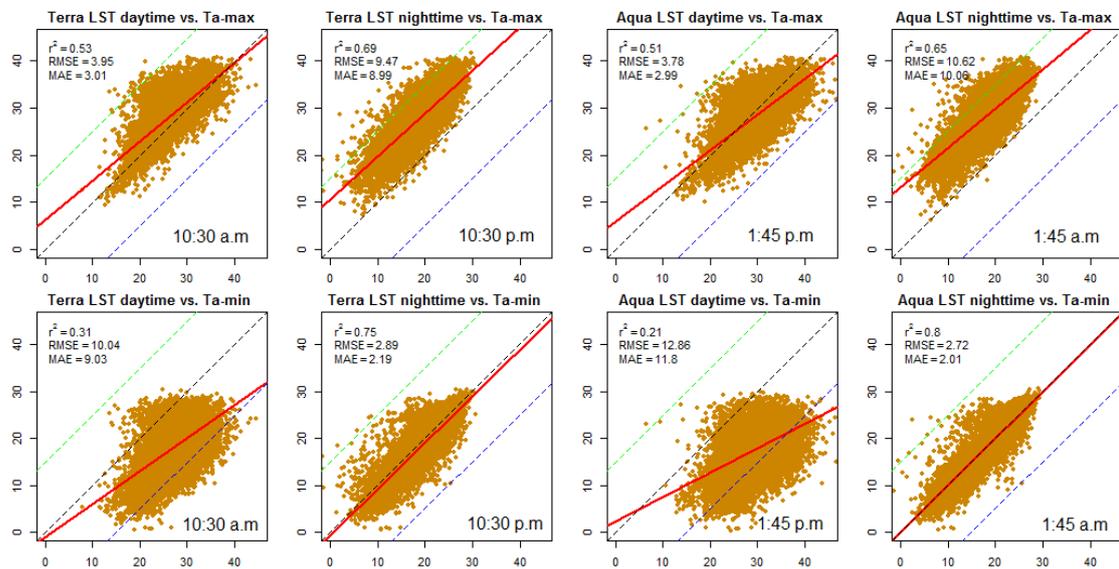
Figure 5.3. (a) The relationship between LST (x-axis) and Ta-max (upper panels), Ta-min (lower panels) of all meteorological stations from 2003–2013. The dashed green line indicates that the difference between Ta and LST is over +15 °C; the dashed blue line indicates that the difference between Ta and LST is lower than −15 °C. The dashed black line indicates the 1:1 line. The solid red line shows the regression line.
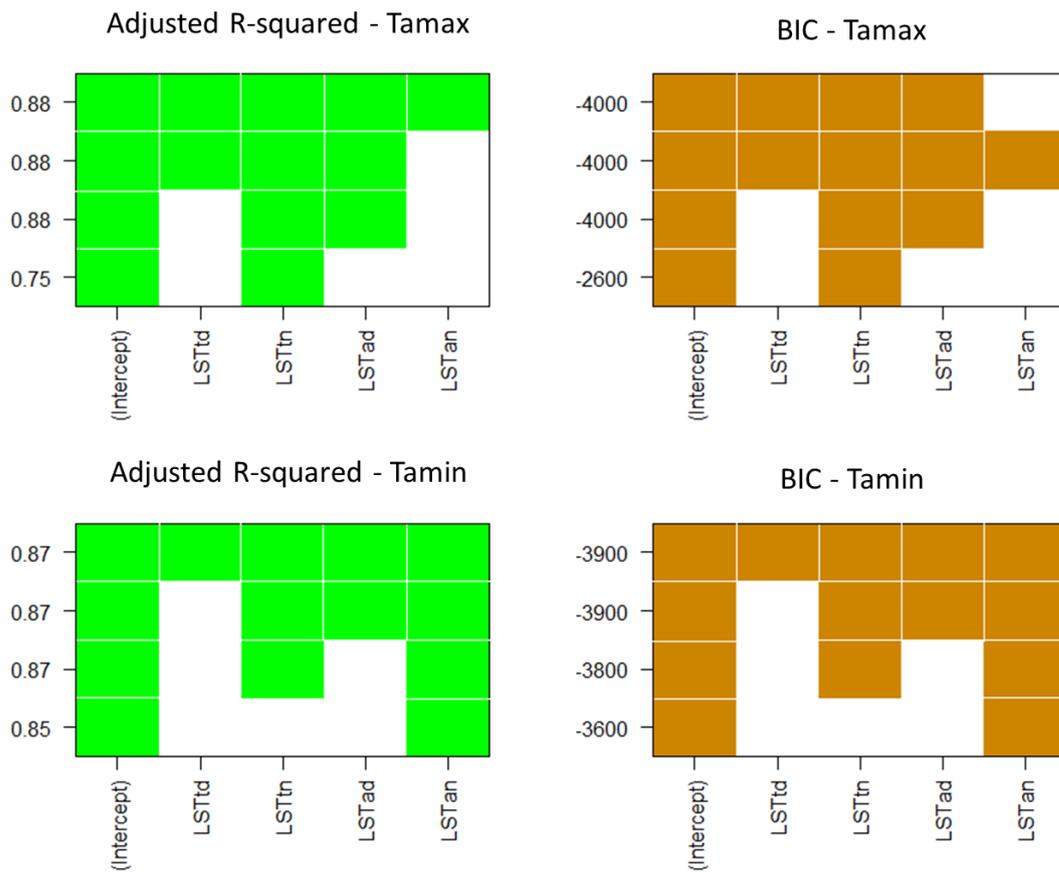
Figure 5.3. (b) Adjusted R-squared and BIC criteria show the top four models of Ta-max (upper panel) and Ta-min (lower panel) estimation using only the four LST dataset.

It also showed that LST nighttime of TERRA seems to perform better than LST nighttime of AQUA for Ta-max estimation. In contrast, LST nighttime of AQUA seemed better than LST nighttime of TERRA for Ta-min estimation.

In order to analyze which of the four LST variables is the best suited Ta estimation, we used the adjusted R-squared and BIC criteria (Figure 5.3b). Among the resulting models, we chose the one that only contained one variable. This variable is as such the best to estimate Ta.

It can be clearly seen that, based on the adjusted R-squared and BIC criteria (Figure 5.3b), the LST nighttime of TERRA and of AQUA were the best variables for Ta-max and Ta-min estimation, respectively. This result was consistent with the result shown in Figure 5.3a.

This result also indicates that the overpass time of satellites and the time of Ta-max, Ta-min recorded had no key influence on the relationship between Ta and LST. This is consistent with other research (Fu et al., 2011; Mostovoy et al., 2006).

5.3.2. Ta-Max Estimation

The parameters of models for Ta-max estimation were determined when we applied Models 1–9 to the calibration dataset. This result is shown in Appendix A.

Figure 5.4 show that the results of the nine models were consistent with the processing in Section 2.5. Model 1 showed the lowest result, followed by Models 2–4.
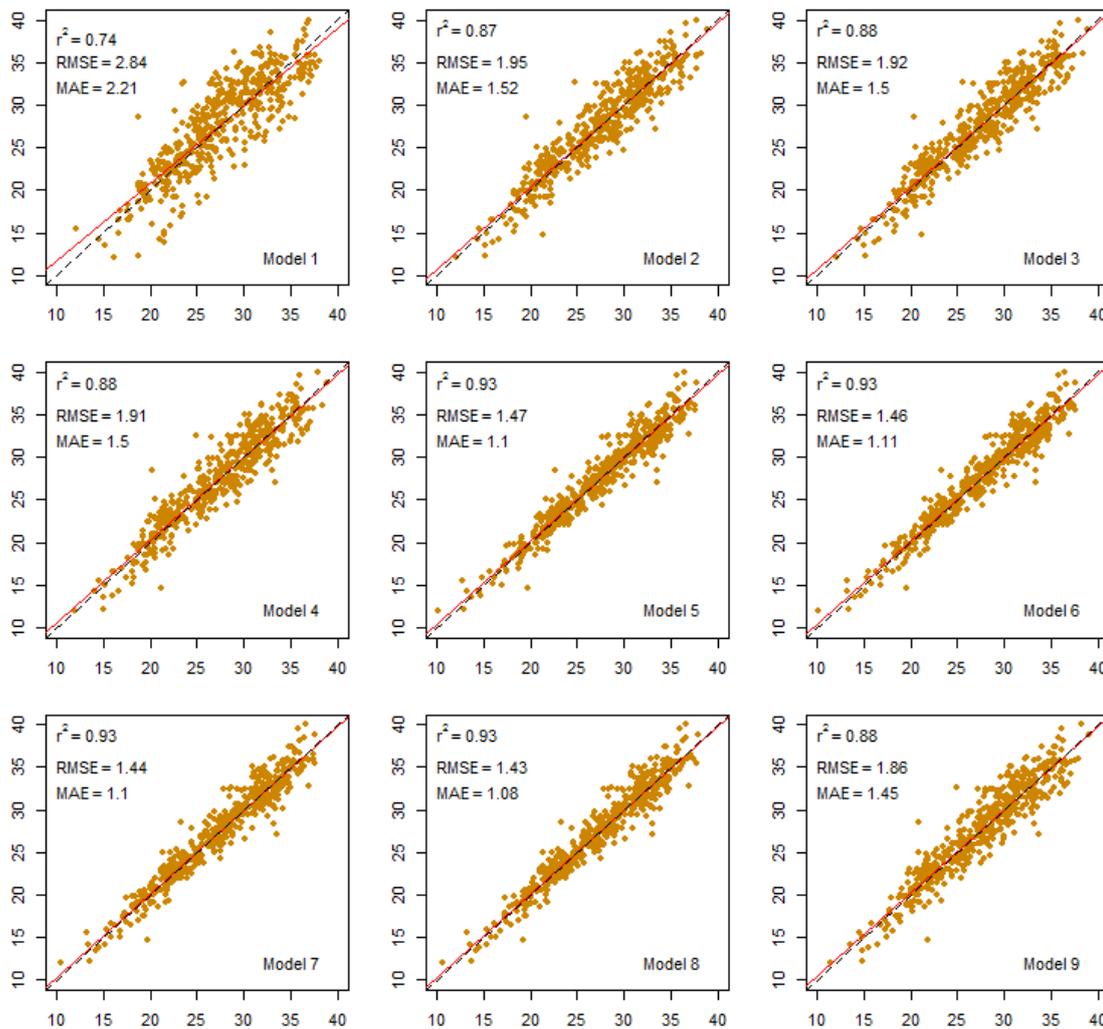
Figure 5.4. The relationship between Ta observed (y-axis) and Ta estimated (x-axis) using Models 1–9. The dashed black line indicates the 1:1 line. The solid red line shows the regression line.

Models 5–8 presented similar high results. However, Model 5 used only six variables; Models 6, 7 and 8 used 7, 8 and 12 variables, respectively. Based on this point, Model 5 would be chosen as the best model of Ta-max estimation.

Combining LST TERRA nighttime and LST AQUA daytime (Model 2), the result of Ta-max estimation was significantly better than Model 1 using LST TERRA nighttime solely. However, when we added LST TERRA daytime (Model 3, three variables), LST TERRA

daytime and LST AQUA nighttime (Model 4, four variables), the performance of these models was not significantly improved in comparison to Model 2 (using two variables).

Comparing Model 4 and Model 5, it can be seen that, by combining four LSTs and elevation (Ele), longitude (Long) into the model, the result was much higher than using four LSTs only. This indicates that elevation (Ele), as well as the location (Long) of the weather station play an important role in Ta-max estimation.

5.3.3. Ta-Min Estimation

The parameters of models for Ta-min estimation were determined when we applied Models 10–18 to the calibration dataset. This result is shown in Appendix B.

In general, Figure 5.5 shows that all models gave similar results of Ta-min estimation. From the simple model with two variables to the complex model (Model 17) using 10 variables, the results are similar. Figure 5.5 also shows that Ta-min estimation can reach a very high accuracy (r^2 = 0.85, RMSE = 2.31, MAE = 1.75) when using only one variable: LSTan. However, it was difficult to increase the accuracy of Ta-min estimated even with 10 variables or all 14 variables (see Figure 5.5).
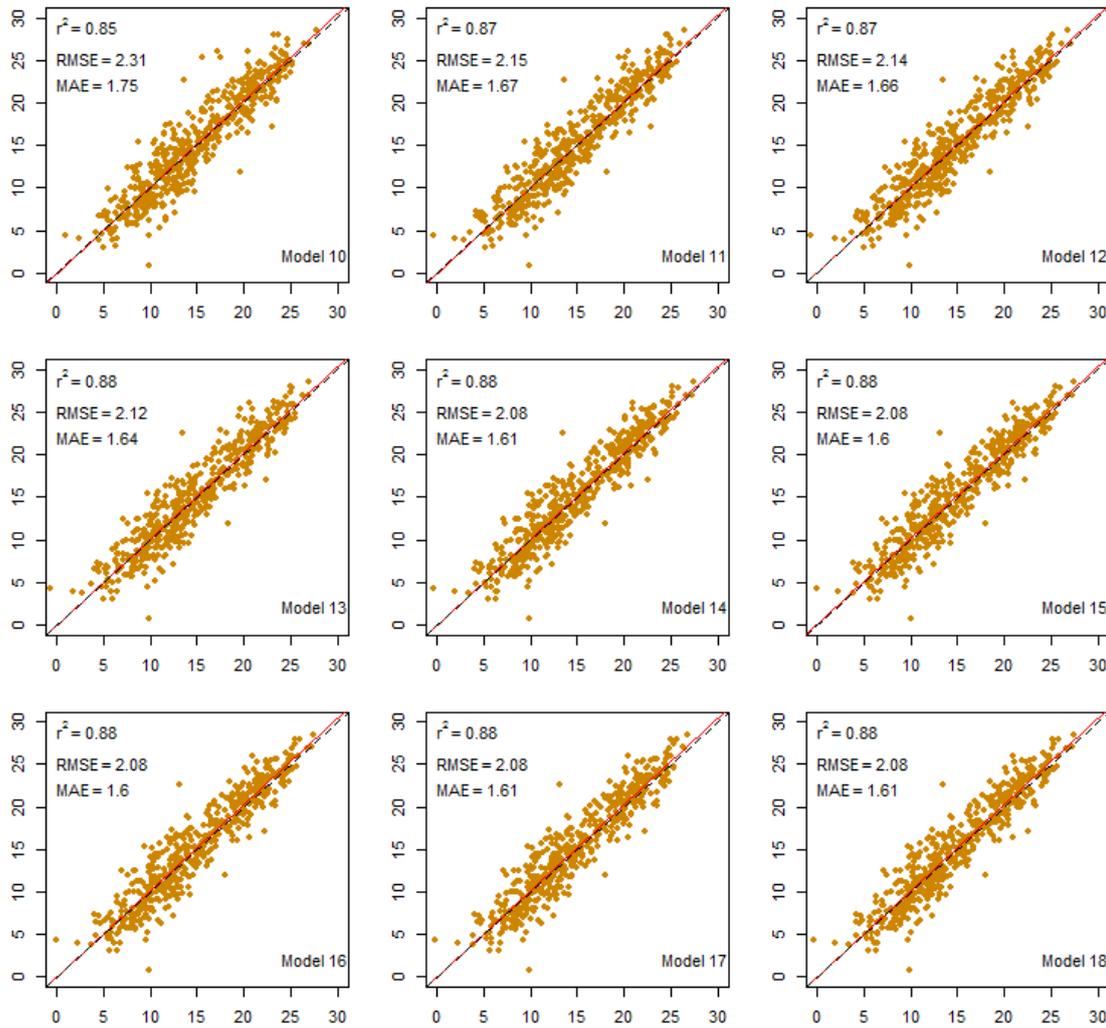
Figure 5.5. The relationship between Ta observed (y-axis) and Ta estimated (x-axis) using Models 10–18. The dashed black line indicates the 1:1 line. The solid red line shows the regression line.

5.3.4. Performance of the Best Model

In order to test the effect of weather station location, as well as the seasonal change or any other factor related to station characteristics, we applied the best model to all datasets. The results are shown in Figure 5.6.
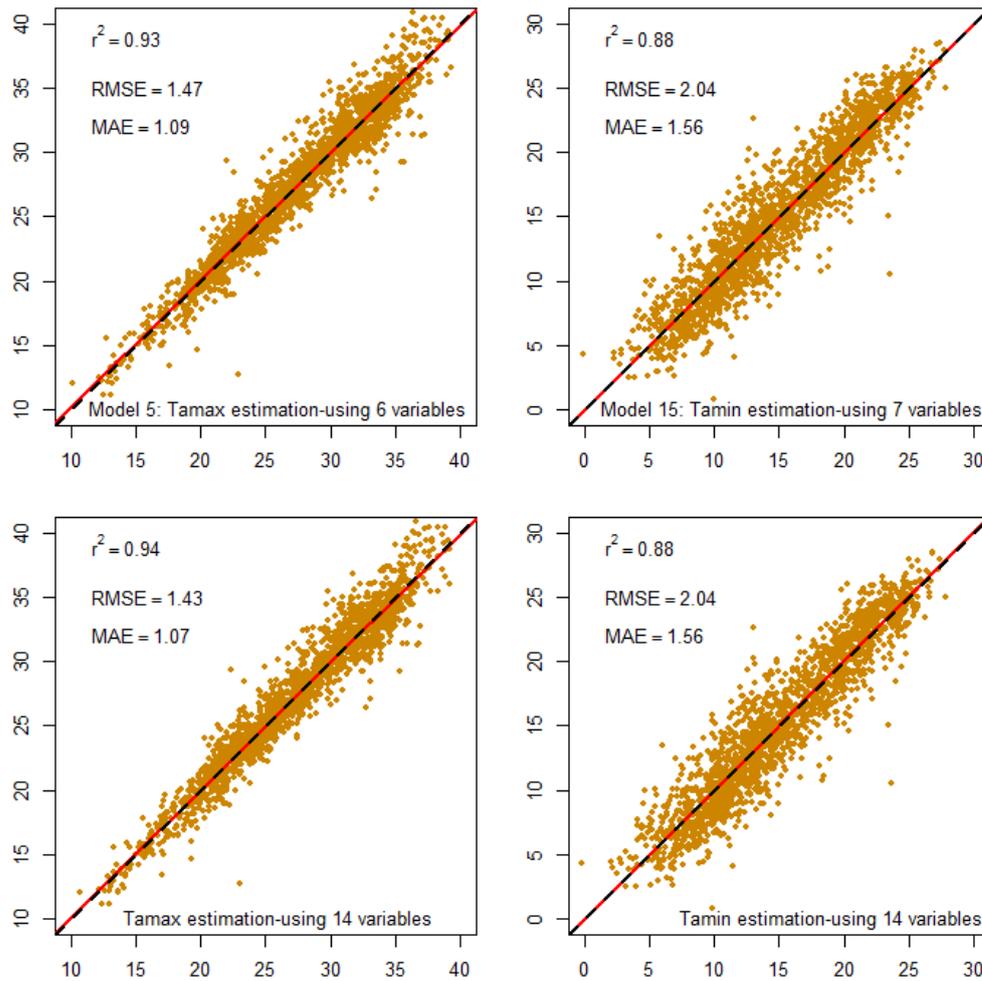
Figure 5.6. The comparison of estimated Ta using Model 5, Model 15 and the model using 14 variables. The dashed black line indicates the 1:1 line. The solid red line shows the regression line.

Looking at Figure 5.4, Figure 5.5 and Figure 5.6, it can be clearly seen that there are similar results of Model 5 and Model 15 when we applied them to the validation dataset or to all datasets. This consistent result indicates that there is no significant factor that could affect the accuracy of Ta estimation when Model 5 and Model 15 were applied for Ta-max and Ta-min estimation, respectively.

5.4. Discussion

5.4.1. MODIS LST Products for Ta Estimation

There are two MODIS sensors, TERRA and AQUA, which provides LST data four times daily (i.e., LSTtd, LSTtn, LSTad, LSTan). Most previous studies have used LST daytime for Ta-max and LST nighttime for Ta-min estimation. Some studies used only TERRA MODIS LST for Ta estimation (Janatian et al., 2016; Emamifar et al., 2013). There is a handful studies

using LST daytime for Ta-min and LST nighttime for Ta-max estimation (Zhang et al., 2011; Benali et al., 2012).

As far as we know, there have also been several studies using both LST daytime and nighttime of both MODIS sensors on TERRA and AQUA for their Ta estimated models (Vancutsem et al., 2010; Huang et al., 2015; Zhang et al., 2011; Shamir and Georgakakos, 2014). However, the purpose of using four times daily LST was for filling the missing LST value. In our study, it is required that all four LST data have to be available.

Zhang et al. (2011) combine TERRA LST daytime and nighttime (two variables), AQUA LST daytime and nighttime (also two variables) for Ta estimation and concluded that: (i) nighttime LST was better than daytime for deriving daily Ta; and (ii) incorporating daytime and nighttime LST significantly improved the estimation of Ta, as compared to using LST nighttime or daytime solely. Our results were consistent with this study (Zhang et al., 2011); however, our analysis could provide an even deeper insight into the matter. Model 2 and Figure 5.3b show that the combination of TERRA LST nighttime and AQUA LST daytime is even better than that of TERRA LST daytime and TERRA LST nighttime.

For Ta-max estimation, Figure 5.4 shows that if only LST products are used (without auxiliary data) the best result was achieved when we combined all four LST data (model 4). However, the Model 4 result (using four LSTs) was just slightly better than Model 2 (using only LSTtn and LSTad).

Similarly, the best result of Ta-min estimation was achieved when we combined all four LSTs (Model 13). The combination of LSTtn and LSTan (Model 11) made the result of Ta-min estimation better than using LSTan solely (Model 10).

As we can see from Figure 5.2, Ta-min has a stronger correlation with LST nighttime of both TERRA and AQUA than Ta-max. However, as the final result, Ta estimation and Ta-max estimation show a much higher accuracy than Ta-min.

5.4.2. Effect of Seasonal on the Accuracy of Ta Estimation

To examine the effect of seasons on the relationship between LST and Ta (Ta-max and Ta-min), we separated the data into four seasons: spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January, February).

Looking at the nine upper panels (Ta-max estimation) of Figure 5.7, it can be seen that the accuracy of Ta-max estimation was significantly improved through different models. Models

1, 2, 3 and 4 show a low accuracy of Ta-max estimated in summer. Models 5, 6, 7 and 8 show a similar accuracy of Ta-max estimation. Model 9 again shows a low accuracy in summer. This means that, if we estimate Ta-max using LST data without other auxiliary data, the results would not be accurate. This can be explained by, as in the summer, the land surface receiving more solar radiation, and as such, the relation between Ta and LST becomes more complex.
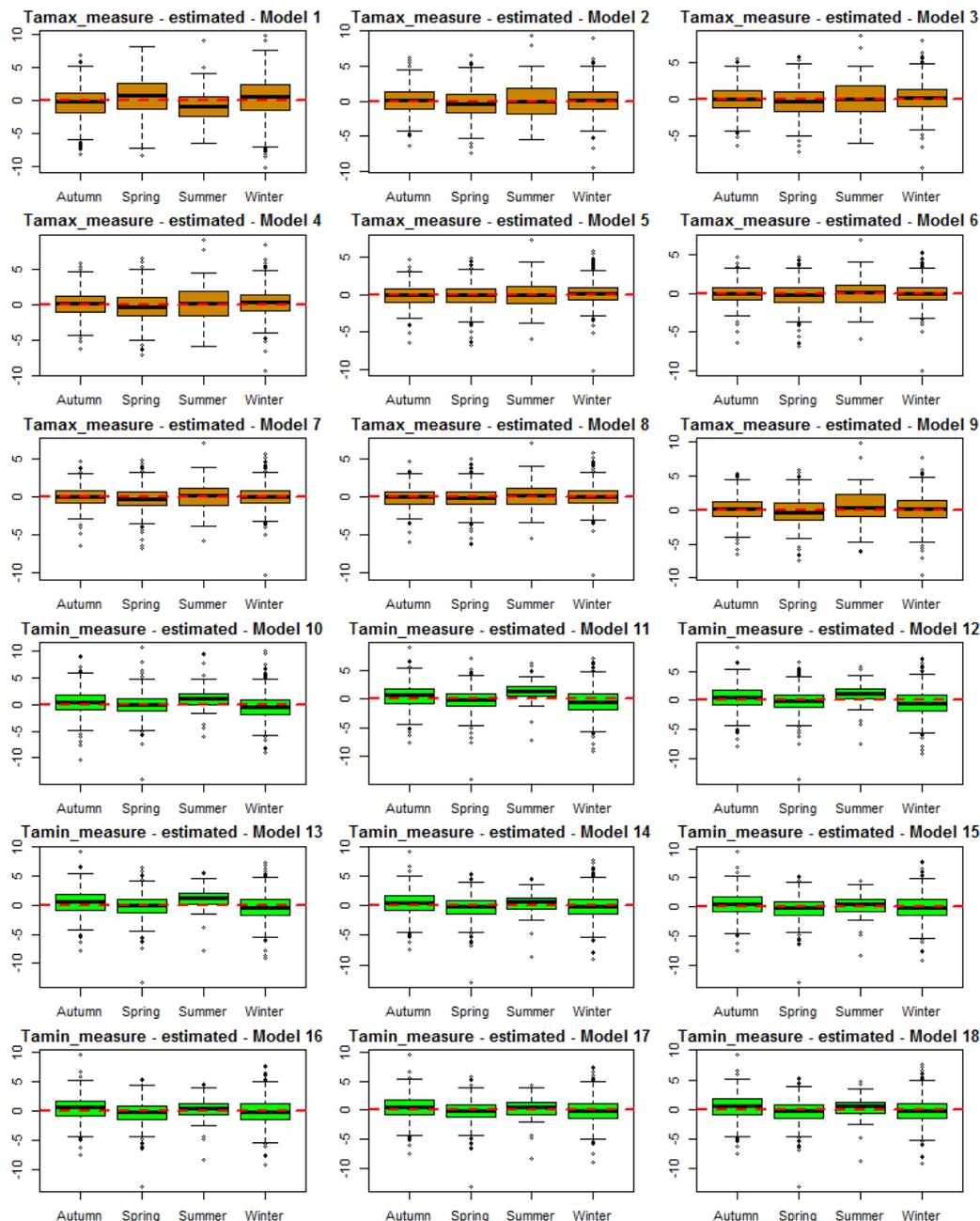
Figure 5.7. Box-plots with the difference between Ta-estimated and Ta-measured by season in years. The line within the box indicates the median. The bottom of the box is the first quartile, and the top is the third quartile. Whiskers represent the lowest value still within 1.5 IQR (IQR = third quartile − first quartile) and the highest value still within 1.5 IQR. The black plus mark indicates the outliers.

130

In general, the changing season had no significant effect on the accuracy of Ta estimation if we use all four LST data and other variables for linear regression.

5.4.3. Effect of View Zenith Angle on the Accuracy of Ta Estimation

As mentioned in Section 2.4, LST data were collected at smaller viewing zenith angles or LST retrievals at larger zenith angles, but with LST errors smaller than at least 2 K.

In order to check the effect of VZA on the relationship between MODIS LST and Ta, we divided the data into three groups based on the range of VZA: bl40 ($0° \leq$ VZA $\leq 40°$), f41t90 ($41° \leq$ VZA $\leq 90°$) and f91t130 ($91° \leq$ VZA $\leq 130°$). Using all linear regression models in Table 5.5 for Ta-max and Ta-min estimation, we then calculated the difference between Ta estimated and Ta measured (difference = Ta measured − Ta estimated). These differences are shown in Figure 5.8. Although the differences might vary from Models 1–9 of Ta-max estimation and Models 10–18 of Ta-min estimation, it was similar between the three groups through all models. This indicates that the effect of VZA on the result of Ta estimation was not significant.
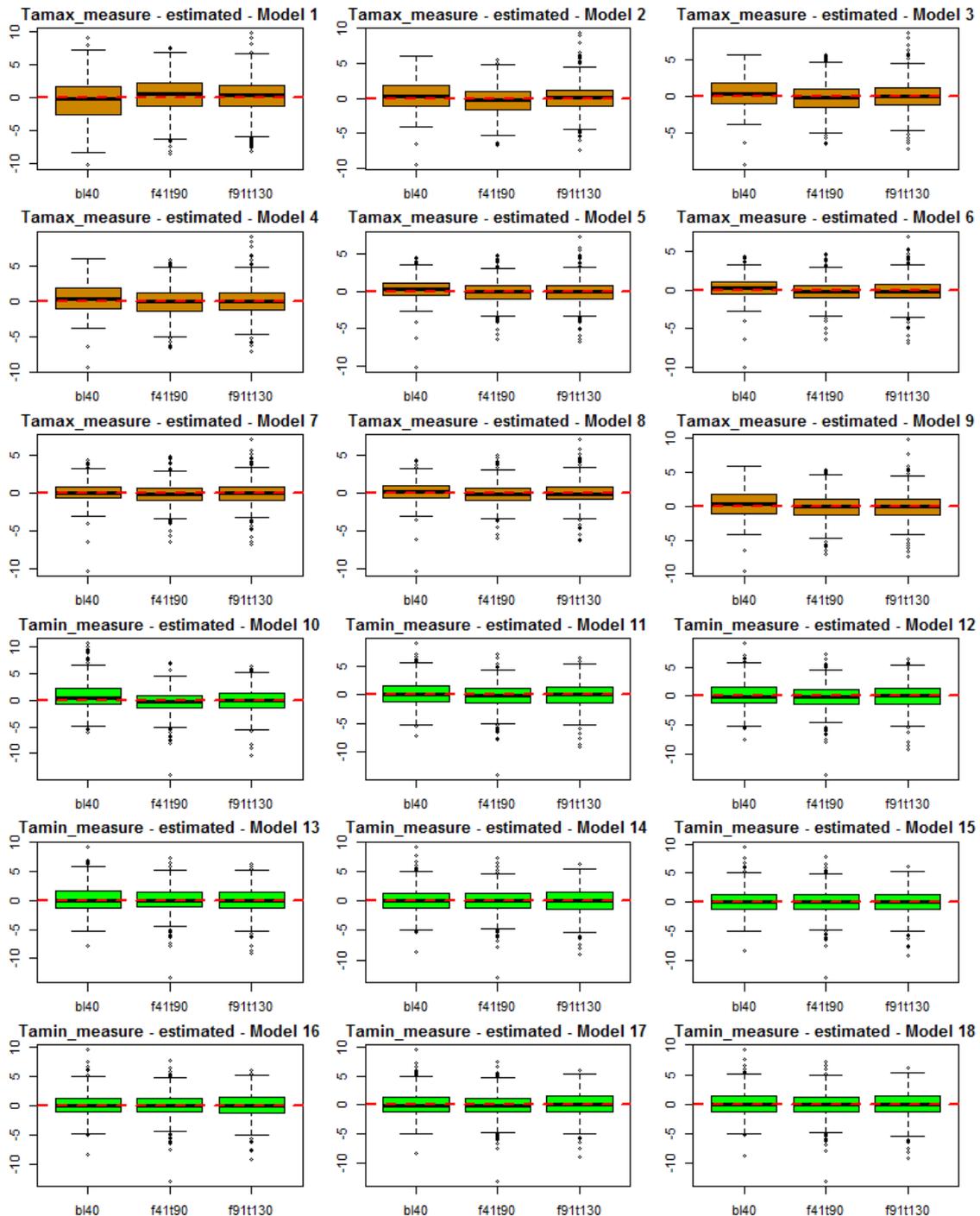
Figure 5.8. Box-plots with the difference between Ta-estimated and Ta-measured by the view zenith angle (VZA). The line within the box indicates the median. The bottom of the box is the first quartile, and the top is the third quartile. Whiskers represent the lowest value still within 1.5 IQR (IQR = third quartile − first quartile) and the highest value still within 1.5 IQR. The black plus mark indicates outliers. bl40 (0° ≤ VZA ≤ 40°), f41t90 (41° ≤ VZA ≤ 90°), f91t130 (91° ≤ VZA ≤ 130°).

5.4.4. Effect of Station Elevation on Accuracy

To test the effect of weather station elevation on the estimation results, we divided the data into three regions: bl200m, f200t500m and ov500m. The bl200m region includes stations that have an elevation below 200 m; the f200t500m region includes stations having an

elevation from 200–500 m; and ov500m region includes stations that have an elevation higher than 500 m. It should be noted that this division is just one option to test the effect of station altitude that might affect the results. It could be divided into more parts for more detail (the more parts are divided, the more detail on the differences could be achieved). We calculated the difference between Ta estimated and Ta measured of these three groups of elevation (difference = Ta measured − Ta estimated). These differences are shown in Figure 5.9.
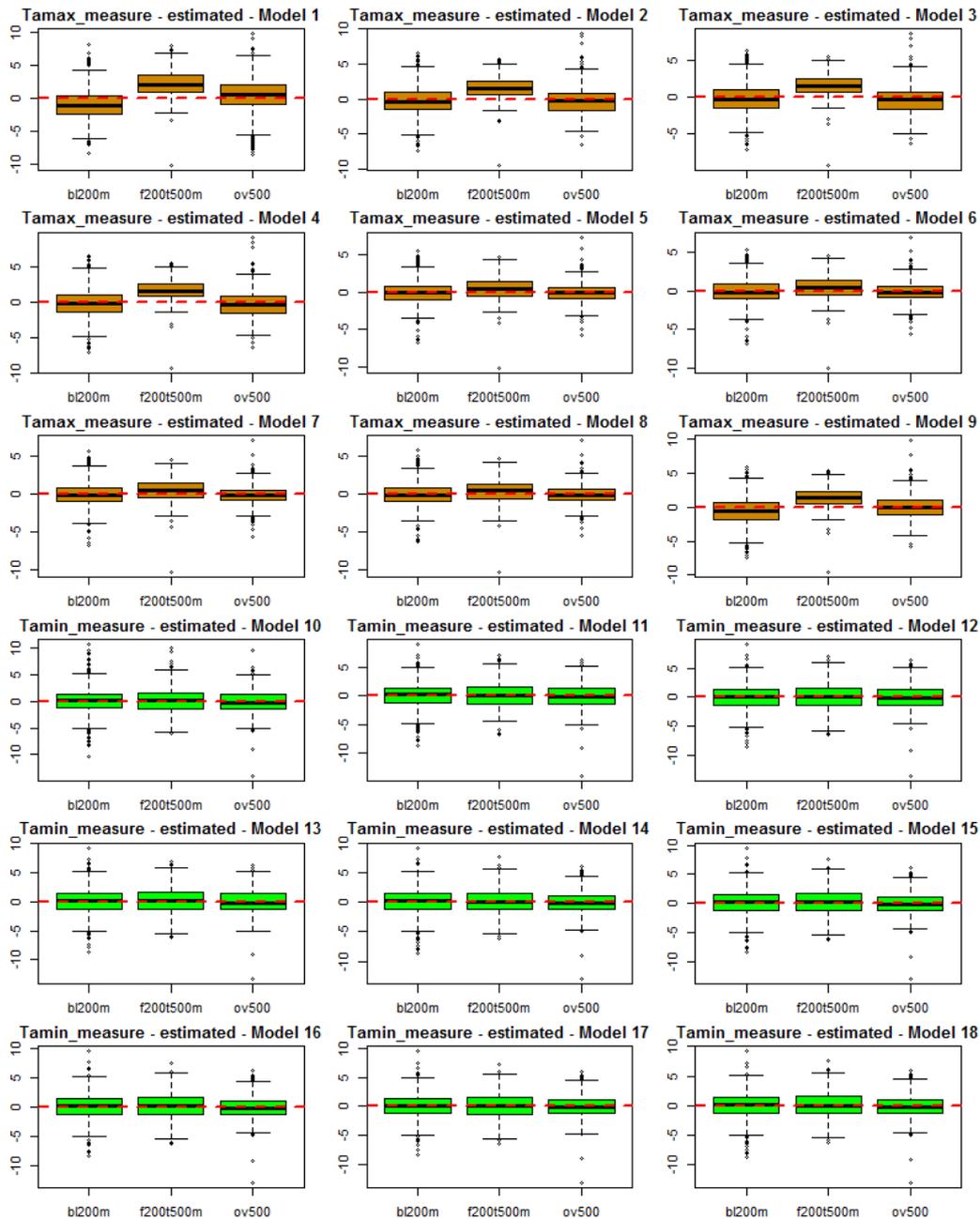
Figure 5.9. Box-plots with the difference between Ta-estimated and Ta-measured by the elevation of the station (Ele). The line within the box indicates the median. The bottom of the box is the first quartile, and the top is the third quartile. Whiskers represent the lowest value still within 1.5 IQR (IQR = third quartile − first quartile) and the highest value still within 1.5 IQR. The black plus mark indicates outliers. bl200m (Ele < 200 m), f200t500 (200 m ≤ Ele ≤ 500 m), ov500 (Ele > 500 m).

As we can see in Figure 5.9, for Ta-max, the differences of three groups were variations from Models 1–9. The most significant was found in Model 1 and Model 9. Models 5, 6, 7 and 8 showed a similar difference between three groups. This indicates that we can reduce the effect of the station's elevation by using Model 5, 6, 7 or 8.

Looking at Table 5.5, we found that all of Models 5, 6, 7 and 8 include longitude and elevation variable in the models. Model 9 also has the elevation variable in it, but Figure 5.9 shows that there is a difference between the three elevation groups. It could be concluded that the longitude variable affects the Ta-max estimation.

Regarding Ta-min estimation, Figure 5.9 shows a similar result of all models (from Models 10–18) and a similar performance between the three elevation groups. This indicates that the station's elevation has no effect on Ta-min estimation.

5.4.5. Accuracy Improvement by Integrating Four LST Products and Auxiliary Variables

5.4.5.1. For Ta-max Estimation

When comparing the results of Model 4 (using four LST data) versus Models 5–9 (using four LST data and auxiliary data; see Figure 5.4), it can be clearly seen that the results were significantly improved. The coefficient of determination (r^2) was increased from 0.88–0.93; RMSE and MAE were decreased from 1.91 down to 1.43 and 1.50 down to 1.08 °C respectively.

The performances of Models 5, 6 and 7 were similar and had very high accuracy. As shown in the variable selection section (Section 2.5.1), these models came from the top three models of adjusted R-squared and BIC criteria.

Model 8, which was chosen from stepwise regression, showed a similar result of accuracy. However, this model used up to 12 variables.

Model 9, which was chosen based on PCA analysis, also showed a high accuracy result with $r^2 = 0.88$, RMSE = 1.86, and MAE = 1.45; however, this result was not as high as the results of Models 5, 6, 7, and 8.

Looking at Table 5.5, it can be clearly seen that Models 5–8 always include elevation and longitude variables. It would be expected that because Model 9 did not use the longitude variable, therefore the accuracy was not as good as Models 5, 6, 7 and 8.

This indicates that elevation and longitude are the most important variables for Ta-max estimation. This result is consistent with adjusted R-squared, BIC criteria and stepwise analysis.

5.4.5.2. For Ta-min Estimation

Looking at Figure 5.5, it can be clearly seen that the results of all nine models (Models 10–18) were not significantly different. In other words, the accuracy of the model was not increased from the simplest (Model 10, using one variable) to the most complex model (Model 17 with 10 variables).

In comparison to previous studies (see Table 5.1), we achieved better results (similar r^2, but smaller RMSE and MAE) of Ta-max estimation due to the combination of four LST data and auxiliary variables. However, this combination just made a slight improvement for the Ta-min estimation (comparing the result of Model 10, versus Models 11–18). In a further study, a better method for increasing the accuracy of Ta-min estimation should be examined.

Considering the coefficient of determination (r^2), the accuracy (RMSE, MAE) and the number of variables used per model, we would regard Model 5 (for Ta-max estimation) and Model 15 (for Ta-min estimation) as the best models.

5.5. Conclusions

In this study, we have analyzed and discussed the relationship between Ta-max, Ta-min and four LST products (LSTtd, LSTtn, LSTad, LSTan). The simple method of multiple linear regression analysis was used, and a high accuracy was achieved with $r^2 = 0.93$, RMSE = 1.43, MAE = 1.08 and $r^2 = 0.88$, RMSE = 2.08, MAE = 1.60, for Ta-max and Ta-min, respectively.

When estimating Ta using one LST datum solely, Ta-min showed a better result than Ta-max (Model 1 versus Model 10 in Figure 5.4 and Figure 5.5). Multiple linear regressions always give better results than simple linear.

An interesting result is that when we directly compared LST data versus Ta, LST nighttime showed a stronger correlation with both Ta-max and Ta-min than LST daytime; Ta-min had a better correlation with LST data than Ta-max (see Figure 5.3a). However, the results of modeling showed that Ta-max can be estimated with better results (higher r^2 and lower RMSE, MAE) than Ta-min when adding auxiliary variables into the models. It could be concluded that in Ta estimation, it is not possible to see the relationship between Ta and

LST from a directed comparison, because there are other factors that also affect that relationship.

Several model analyses indicate that MODIS LST represents the most important variables for Ta estimation. However, to achieve the best results, other variables, such as day length (in hours), Julian day, longitude, latitude and elevation, should be taken into consideration and put into the models.

Appendix A

Table A1. Parameters of Models for T$_{a-max}$ Estimation.

		Estimate	Std. Error	*t*-value	*p*-value
Model 1	(Intercept)	10.7057	0.2954	36.2400	0.0000
	LST$_{tn}$	0.9826	0.0165	59.7200	0.0000
Model 2	(Intercept)	2.1651	0.3159	6.8530	0.0000
	LST$_{tn}$	0.5778	0.0161	35.8360	0.0000
	LST$_{ad}$	0.5133	0.0145	35.4500	0.0000
Model 3	(Intercept)	1.8313	0.3206	5.7110	0.0000
	LST$_{td}$	0.1234	0.0258	4.7810	0.0000
	LST$_{tn}$	0.5482	0.0171	32.0060	0.0000
	LST$_{ad}$	0.4329	0.0221	19.5830	0.0000
Model 4	(Intercept)	1.7664	0.3226	5.4750	0.0000
	LST$_{td}$	0.1283	0.0259	4.9470	0.0000
	LST$_{tn}$	0.6028	0.0363	16.6280	0.0000
	LST$_{ad}$	0.4305	0.0221	19.4510	0.0000
	LST$_{an}$	−0.0570	0.0334	−1.7080	0.0880
Model 5	(Intercept)	265.4000	9.1650	28.9520	0.0000
	LST$_{td}$	0.1646	0.0197	8.3610	0.0000
	LST$_{tn}$	0.5297	0.0274	19.3710	0.0000
	LST$_{ad}$	0.2505	0.0176	14.2410	0.0000
	LST$_{an}$	0.1419	0.0260	5.4600	0.0000
	Ele	−0.0028	0.0001	−19.8260	0.0000
	Long	−2.4810	0.0865	−28.6960	0.0000
Model 6	(Intercept)	269.5000	9.1030	29.6050	0.0000
	LST$_{td}$	0.1712	0.0195	8.7690	0.0000
	LST$_{tn}$	0.5073	0.0274	18.5120	0.0000
	LST$_{ad}$	0.2238	0.0182	12.3170	0.0000
	LST$_{an}$	0.1649	0.0261	6.3190	0.0000
	Ele	−0.0029	0.0001	−20.6540	0.0000
	Long	−2.5100	0.0857	−29.2840	0.0000
	Jd	−0.0019	0.0004	−5.1030	0.0000
Model 7	(Intercept)	270.1000	9.0700	29.7730	0.0000
	LST$_{td}$	0.1868	0.0201	9.3050	0.0000
	LST$_{tn}$	0.4746	0.0292	16.2390	0.0000
	LST$_{ad}$	0.2178	0.0182	11.9700	0.0000
	LST$_{an}$	0.1886	0.0271	6.9660	0.0000
	Ele	−0.0029	0.0001	−20.8160	0.0000
	Long	−2.5130	0.0854	−29.4290	0.0000
	Jd	−0.0020	0.0004	−5.3020	0.0000
	VZA$_{ad}$	−0.0037	0.0012	−3.1350	0.0018
Model 8	(Intercept)	285.4000	9.9020	28.8230	0.0000
	LST$_{td}$	0.1812	0.0218	8.3240	0.0000
	LST$_{tn}$	0.4523	0.0314	14.3870	0.0000
	LST$_{ad}$	0.2445	0.0212	11.5410	0.0000

	LST$_{an}$	0.2013	0.0294	6.8500	0.0000
	NDVI	1.6500	0.3829	4.3090	0.0000
	Ele	−0.0033	0.0002	−19.8170	0.0000
	Long	−2.5080	0.0847	−29.6040	0.0000
	Lat	−0.7172	0.1754	−4.0890	0.0000
	Dlth	−0.1648	0.0950	−1.7350	0.0829
	Jd	−0.0024	0.0004	−6.0140	0.0000
	VZA$_{td}$	0.0021	0.0013	1.6460	0.1001
	VZA$_{ad}$	−0.0038	0.0012	−3.2710	0.0011
Model 9	(Intercept)	4.6214	1.2432	3.7170	0.0002
	LST$_{td}$	0.1701	0.0259	6.5730	0.0000
	LST$_{tn}$	0.5597	0.0367	15.2720	0.0000
	LST$_{ad}$	0.4074	0.0225	18.0830	0.0000
	LST$_{an}$	−0.0463	0.0337	−1.3720	0.1704
	Ele	−0.0013	0.0002	−7.2070	0.0000
	Dlth	−0.1617	0.1235	−1.3090	0.1907
	Jd	−0.0013	0.0005	−2.5570	0.0107

Appendix B

Table B1. Parameters of Models for T$_{a\text{-min}}$ Estimation.

		Estimate	**Std. Error**	*t*-value	*p*-value
Model 10	(Intercept)	−1.5176	0.2085	−7.2800	0.0000
	LST$_{an}$	1.0157	0.0121	83.9300	0.0000
Model 11	(Intercept)	−2.4950	0.2178	−11.4600	0.0000
	LST$_{tn}$	0.4055	0.0371	10.9200	0.0000
	LST$_{an}$	0.6492	0.0355	18.2900	0.0000
Model 12	(Intercept)	−1.4608	0.3340	−4.3730	0.0000
	LST$_{tn}$	0.4496	0.0385	11.6920	0.0000
	LST$_{ad}$	−0.0620	0.0153	−4.0640	0.0001
	LST$_{an}$	0.6541	0.0353	18.5420	0.0000
Model 13	(Intercept)	−1.6875	0.3420	−4.9350	0.0000
	LST$_{td}$	0.0800	0.0275	2.9080	0.0037
	LST$_{tn}$	0.4417	0.0384	11.4950	0.0000
	LST$_{ad}$	−0.1139	0.0235	−4.8570	0.0000
	LST$_{an}$	0.6425	0.0354	18.1590	0.0000
Model 14	(Intercept)	−10.9000	1.3100	−8.3200	0.0000
	LST$_{td}$	0.0544	0.0271	2.0060	0.0450
	LST$_{tn}$	0.4210	0.0382	11.0130	0.0000
	LST$_{ad}$	−0.0931	0.0239	−3.9020	0.0001
	LST$_{an}$	0.5800	0.0358	16.2000	0.0000
	Dlth	0.8850	0.1280	6.9320	0.0000
	Jd	0.0020	0.0005	3.6870	0.0002
Model 15	(Intercept)	6.4498	5.2374	1.2310	0.2184
	LST$_{td}$	0.0479	0.0271	1.7700	0.0769
	LST$_{tn}$	0.4187	0.0380	11.0110	0.0000
	LST$_{ad}$	−0.0998	0.0238	−4.1890	0.0000

	LST$_{an}$	0.5885	0.0357	16.4770	0.0000
	Dlth	0.9112	0.1273	7.1560	0.0000
	Jd	0.0020	0.0005	3.7160	0.0002
	Lat	−0.8214	0.2397	−3.4280	0.0006
Model 16	(Intercept)	7.0965	5.2666	1.3470	0.1781
	LST$_{td}$	0.0529	0.0274	1.9320	0.0537
	LST$_{tn}$	0.4097	0.0388	10.5540	0.0000
	LST$_{ad}$	−0.1027	0.0239	−4.2870	0.0000
	LST$_{an}$	0.5864	0.0358	16.3970	0.0000
	Dlth	0.9479	0.1312	7.2230	0.0000
	Jd	0.0019	0.0005	3.5110	0.0005
	Lat	−0.8599	0.2419	−3.5540	0.0004
	Ele	−0.0002	0.0002	−1.1540	0.2487
Model 17	(Intercept)	10.9860	5.3900	2.0380	0.0418
	LST$_{td}$	0.0740	0.0284	2.6070	0.0093
	LST$_{tn}$	0.3827	0.0412	9.2900	0.0000
	LST$_{ad}$	−0.0913	0.0245	−3.7250	0.0002
	LST$_{an}$	0.5887	0.0376	15.6640	0.0000
	NDVI	1.6296	0.5405	3.0150	0.0026
	Ele	−0.0006	0.0002	−2.5960	0.0096
	Lat	−1.0290	0.2479	−4.1510	0.0000
	Dlth	0.8447	0.1343	6.2910	0.0000
	Jd	0.0015	0.0005	2.6990	0.0071
	VZA$_{ad}$	−0.0027	0.0016	−1.6590	0.0973
Model 18	(Intercept)	−11.0000	1.3200	−8.3410	0.0000
	LST$_{td}$	0.0575	0.0275	2.0890	0.0369
	LST$_{tn}$	0.4160	0.0390	10.6610	0.0000
	LST$_{ad}$	−0.0946	0.0240	−3.9460	0.0001
	LST$_{an}$	0.5790	0.0359	16.1250	0.0000
	Ele	−0.0001	0.0002	−0.6680	0.5043
	Dlth	0.9060	0.1310	6.8960	0.0000
	Jd	0.0019	0.0005	3.5520	0.0004

References

Balaghi, R.; Tychon, B.; Eerens, H.; Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10, 438–452.

Basist, A.N.; Peterson, N.C.; Peterson, T.C.; Williams, C.N. (1998). Using the special sensor microwave/imager to monitor land surface temperature, wetness, and snow cover. Journal of Applied Meteorology, 37, 888–911.

Benali, A.; Carvalho, A.C.; Nunes, J.P.; Carvalhais, N.; Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124, 108–121.

Coll, C.; Caselles, V.; Sobrino, J.A.; Valor, E. (1994). On the atmospheric dependence of the split-window equation for land surface temperature. Int. J. Remote Sensing of Environment, 27, 105–122.

Cresswell, M.P.; Morse, A.P.; Thomson, M.C.; Connor, S.J. (1999). Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. International Journal of Remote Sensing, 20, 1125–1132.

Crosson, W.L.; Al-Hamdan, M.Z.; Hemmings, S.N.; Wade, G.M. (2012). A daily merged MODIS Aqua–Terra land surface temperature data set for the conterminous United States. Remote Sensing of Environment, 119, 315–324.

De Bruin, H.A.R.; Trigo, I.F.; Jitan, M.A.; TemesgenEnku, N.; van der Tol, C.; Gieske, A.S.M. (2010). Reference crop evapotranspiration derived from geo-stationary satellite imagery: A case study for the Fogera flood plain, NW-Ethiopia and the Jordan Valley, Jordan. Hydrology and Earth System Sciences, 14, 2219–2228.

De Wit, A.J.W.; van Diepen, C.A. (2008). Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. International Journal of Applied Earth Observation and Geoinformation, 10, 414–425.

Dinh, V.V. (2005). Country Report the hydro meteorological service of Vietnam and its modernization plan. In Proceedings of the 5th Global Precipitation Measurement (GPM) International Planning Workshop, Tokyo, Japan, 7–9 November 2005.

Emamifar, S.; Rahimikhoob, A.; Noroozi, A. (2013). Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. International Journal of Climatology, 33, 3174–3181.

Florio, E.N.; Lele, S.R.; Chi Chang, Y.; Sterner, R.; Glass, G.E. (2004). Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: A statistical approach. International Journal of Remote Sensing, 25, 2979–2994.

Fu, G.; Shen, Z.X.; Zhang, X.Z.; Shi, P.L.; Zhang, Y.J.; Wu, J.S. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. Acta Ecologica Sinica, 21, 8–13.

Gallo, K.; Hale, R.; Tarpley, D.; YU, Y. (2011). Evaluation of the relationship between air and land surface temperature under clear- and cloudy-sky conditions. Journal of Applied Meteorology and Climatology, 50, 767–775.

Goetz, S.J. (1997). Multi-sensor analysis of NDVI, surface temperature and biophysical variables at amixed grassland site. International Journal of Remote Sensing, 18, 71–94.

Goetz, S.J.; Prince, S.D.; Small, J. (2000). Advances in satellite remote sensing of environmental variables for epidemiological applications. Advances in Parasitology, 47, 289–307.

Good, E. (2015). Daily minimum and maximum surface air temperatures from geostationary satellite data. Journal of Geophysical Research: Atmospheres, 120, 2306–2324.

Huang, R.; Zhang, C.; Huang, J.; Zhu, D.; Wang, L.; Liu, J. (2015). Mapping of daily mean air temperature in agricultural regions using daytime and nighttime land surface temperatures derived from Terra and Aqua MODIS data. Remote Sensing, 7, 8728–8756.

Intergovernmental Panel on Climate Change (IPCC). Climate Change 2001: The Scientific Basis; Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK, p. 881.

Intergovernmental Panel on Climate Change (IPCC). Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Cli-mate Change; Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L., Eds.; Cambridge University Press: Cambridge, UK, 2007.

Janatian, N.; Sadeghi, M.; Sanaeinejad, S.H.; Bakhshian, E.; Farid, A.; Hasheminia, S.M.; Ghazanfari, S. (2016). A statistical framework for estimating air temperature using MODIS land surface temperature data. International Journal of Climatology (article in press).

Jang, J.D.; Viau, A.A.; Anctil, F. (2004). Neural network estimation of air temperatures from AVHRR data. International Journal of Remote Sensing, 25, 4541–4554.

Jin, M.; Dickinson, R.E. (2010). Land surface skin temperature climatology: Benefitting from the strengths of satellite observations. Environmental Research Letters, 5, 044004.

Kilibarda, M.; Hengl, T.; Heuvelink, G.; Gräler, B.; Pebesma, E.; Perčec Tadić, M.; Bajat, B. (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. Journal of Geophysical Research: Atmospheres, 119, 2294–2313.

Lai, Y.J.; Li, C.F.; Lin, P.H.; Wey, T.H.; Chang, C.S. (2012). Comparison of MODIS land surface temperature and ground-based observed air temperature in complex topography. International Journal of Remote Sensing, 33, 7685–7702.

Land Quality Assessment Site of NASA. Julian Day. Available online: http://landweb.nascom.nasa.gov/browse/calendar.html (accessed on 2 November 2015).

Lehman, J.T. (2002). Application of satellite AVHRR to water balance mixing dynamics and the chemistry of Lake Edward, East Africa. In The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity; Odada, E.O., Olago, D.O., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 235–260.

Lin, S.; Moore, N.J.; Messina, J.P.; de Visser, M.H.; Wu, J. (2012). Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. International Journal of Applied Earth Observation and Geoinformation, 18, 128–140.

Lin, X.; Zhang, W.; Huang, Y.; Sun, W.; Han, P.; Yu, L.; Sun, F. (2016). Empirical estimation of near-surface air temperature in China from MODIS LST data by considering physiographic features. Remote Sensing, 8, 629.

Miura, T.; Yoshioka, H.; Fujiwara, K.; Yamamoto, H. (2008). Inter-comparison of ASTER and MODIS surface reflectance and vegetation index products for synergistic applications to natural resource monitoring. Sensors, 8, 2480–2499.

Mostovoy, G.V.; King, R.L.; Reddy, K.R.; Kakani, V.G.; Filippova, M.G. (2006). Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi. GIScience & Remote Sensing, 43, 78–110.

NASA Visible Earth. Available online: http://visibleearth.nasa.gov/view.php?id=35791 (accessed on 12 October 2015).

Nieto, H.; Sandholt, I.; Aguado, I.; Chuvieco, E.; Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. Remote Sensing of Environment, 115, 107–116.

Peterson, T.C.; Basist, A.N.; Williams, C.; Grody, N. (2000). A blended satellite/in situ near-global surface temperature data set. Bulletin of the American Meteorological Society, 81, 2157–2164.

Prihodko, L.; Goward, S.N. (1997). Estimation of air temperature from remotely sensed surface observations. Remote Sensing of Environment, 60, 335–346.

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, 213, 230–249.

Sandholt, I.; Rasmussen, K.; Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sensing of Environment, 79, 213–224.

Shamir, E.; Georgakakos, K.P. (2014). MODIS land surface temperature as an index of surface air temperature for operational snowpack estimation. Remote Sensing of Environment, 152, 83–98.

Shen, S.; Leptoukh, G.G. (2011). Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environmental Research Letters, 6, 045206.

Shi, L.; Liu, P.; Kloog, I.; Lee, M.; Kosheleva, A.; Schwartz, J. (2016). Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. Environmental Research, 146, 51–58.

Stisen, S.; Sandholt, I.; Norgaard, A.; Fensholt, R.; Eklundh, L. (2007). Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sensing of Environment, 110, 262–274.

Sun, H.; Chen, Y.; Gong, A.; Zhao, X.; Zhan, W.; Wang, M. (2014). Estimating mean air temperature using MODIS day and night land surface temperatures. Theoretical and Applied Climatology, 118, 81–92.

Sun, H.; Zhao, X.; Chen, Y.; Gong, A.; Yang, J. (2013). A new agricultural drought monitoring index combining MODIS NDWI and day–night land surface temperatures: A case study in China. International Journal of Remote Sensing, 34, 8986–9001.

Sun, Y.J.; Wang, J.F.; Zhang, R.H.; Gillies, R.R.; Xue, Y.; Bo, Y.C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. Theoretical and Applied Climatology, 80, 37–48.

The Report of MOD11A1 and MYD11A1 Validation Accuracy. Available online: http://landval.gsfc.nasa.gov/ProductStatus.php?ProductID=MOD11 (accessed on 2 November 2015).

The United States Naval Observatory (USNO). Day Length in Hours Data. Available online: http://aa.usno.navy.mil/data/docs/Dur_OneYear.php#formb (accessed on 1 December 2015).

The U.S. Geological Survey. MODIS LST Data. Available online: http://earthexplorer.usgs.gov (accessed on 1 August 2015).

Vallet-Coulomb, C.; Legesse, D.; Gasse, F.; Travi, Y.; Chernet, T. (2001). Lake evaporation estimates in tropical Africa (Lake Ziway, Ethiopia). Journal of Hydrology, 245, 1–18.

Vancutsem, C.; Ceccato, P.; Dinku, T.; Connor, S.J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114, 449–465.

Vogt, J.V.; Viau, A.A.; Paquet, F. (1997). Mapping regional air temperature fields using satellite-derived surface skin temperatures. International Journal of Climatology, 17, 1559–1579.

Wan, Z. (2006). MODIS Land Surface Temperature Products Users' Guide. Available online: http://www.icess.ucsb.edu/modis/LstUsrGuide/MODIS_LST_products_Users_guide_C5.pdf (accessed on 19 November 2015).

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.

Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.L. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sensing of Environment, 83, 163–180.

Williamson, S.; Hik, D.; Gamon, J.; Kavanaugh, J.; Flowers, G. (2014). Estimating temperature fields from MODIS land surface temperature and air temperature observations in a sub-arctic alpine environment. Remote Sensing, 6, 946–963.

Xu, Y.; Knudby, A.; Ho, H.C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. International Journal of Remote Sensing, 35, 8108–8121.

Xu, Y.; Qin, Z.; Shen, Y. (2012). Study on the estimation of near-surface air temperature from MODIS data by statistical methods. International Journal of Remote Sensing, 33, 7629–7643.

Zakšek, K.; Schroedter-Homscheidt, M. (2009). Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. ISPRS J. Photogramm. Remote Sensing, 64, 414–421.

Zeng, L.; Wardlow, B.D.; Tadesse, T.; Shan, J.; Hayes, M.J.; Li, D.; Xiang, D. (2015). Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US. Remote Sensing, 7, 951–970.

Zhang, R.; Rong, Y.; Tian, J.; Su, H.; Li, Z.L.; Liu, S. (2015). A remote sensing method for estimating surface air temperature and surface vapor pressure on a regional scale. Remote Sensing, 7, 6005–6025.

Zhang, W.; Huang, Y.; Yu, Y.; Sun, W. (2011). Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures. International Journal of Remote Sensing, 32, 9415–9440.

Zhu, W.; Lǚ, A.; Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sensing of Environment, 130, 62–73.

Chapter 6

Noi, P.T.; Degener, J.; Kappas, M. **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 Sens. 2017, 9, 398.

Abstract

Recently, several methods have been introduced and applied to estimate daily air surface temperature (T_a) using MODIS land surface temperature data (MODIS LST). Among these methods, the most common used method is statistical modeling, and the most applied algorithms are linear/multiple linear regression models (LM). There are only a handful of studies using machine learning algorithm models such as random forest (RF) or cubist regression (CB). In particular, there is no study comparing different combinations of four MODIS LST datasets with or without auxiliary data using different algorithms such as multiple linear regression, random forest, and cubist regression for daily T_{a-max}, T_{a-min}, and T_{a-mean} estimation. Our study examines the mentioned combinations of four MODIS-LST datasets and shows that different combinations and differently applied algorithms produce various T_a estimation accuracies. Additional analysis of daily data from three climate stations in the mountain area of North West of Vietnam for the period of five years (2009 to 2013) with four MODIS LST datasets (AQUA daytime, AQUA nighttime, TERRA daytime, and TERRA nighttime) and two additional auxiliary datasets (elevation and Julian day) shows that CB and LM should be applied if MODIS LST data is used solely. If MODIS LST is used together with auxiliary data, especially in mountainous areas, CB or RF is highly recommended. This study proved that the very high accuracy of T_a estimation ($R^2 >$ 0.93/0.80/0.89 and RMSE ~1.5/2.0/1.6 °C of T_{a-max}, T_{a-min}, and T_{a-mean}, respectively) could be achieved with a simple combination of four LST data, elevation, and Julian day data using a suitable algorithm.

Keywords:

MODIS LST; daily air surface temperature; northwest Vietnam; linear regression (LM); random forest (RF); cubist regression (CB)

6.1. Introduction

Air surface temperature (T_a) with high spatial and temporal resolution plays an important role in various applications, such as crop growth monitoring and simulations (de Wit and van Diepen, 2008), hydrological, ecological, and environmental studies (Daly, 2006; Stahl et al., 2006; Izady et al., 2015), weather forecasting (Smith et al., 1988; Christiansen, 2005), and climate change (IPCC, 2007; Lofgren et al., 2011). It is used as a key input variable and directly affects the accuracy of these applications. Traditionally, T_a is usually measured by

weather stations (often at 2 m above the ground) and usually limited in spatial coverage. Especially in mountainous areas of Vietnam, weather station coverage is extremely sparse.

Meanwhile, satellite data available at various spatial and temporal resolutions, such as Landsat, the Advanced Very High Resolution Radiometer (AVHRR), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and especially Moderate-resolution Imaging Spectroradiometer (MODIS), which was launched in the early 2000s, have marked a significant increase in the quality and quantity of thermal data. The advantage of MODIS is that it can provide Land Surface Temperature (LST) data directly. However, there is a difference between T_a and LST because of the complex surface energy budget and multiple related variables between them.

Recently, several methods have been introduced and applied to estimate T_a using satellite data such as the temperature–vegetation index method—TVX (Stisen et al., 2007; Nieto et al., 2011; Zhu et al., 2013), surface energy-balance-based methods (Sun et al., 2005), and statistical methods (Mostovoy et al., 2006; Vancutsem et al., 2010; Benali et al., 2012; Good, 2015; Noi et al., 2016; Zhang et al., 2016) using different satellite datasets such as Landsat— ETM+ (Wloczyk et al., 2011; Ho et al., 2014), AVHRR (Prince et al., 1998), or MODIS LST (Vancutsem et al., 2010; Zhu et al., 2013; Shen and Leptoukh, 2011; Zeng et al., 2015). Among these satellite data, the most used is MODIS LST because it is freely available and can be obtained easily (Zhang et al., 2016). In addition, MODIS satellite provides four LST datasets daily, including: TERRA daytime (LST_{td}), TERRA nighttime (LST_{tn}), AQUA daytime (LST_{ad}), and AQUA nighttime (LST_{an}), which overpass local time at around 10 a.m., 10 p.m., 1 p.m., and 1 a.m. (our study area), respectively.

Looking at the current literature, there are plentiful T_a estimation studies; however, studies using machine learning techniques such as cubist regression (CB) or random forests (RF) are very rare as far as we know, only (Zhang et al., 2016; Emamifar et al., 2013; Xu et al., 2014; Meyer et al., 2016). However, all of these studies used MODIS LST integrating auxiliary data and estimated only T_{a-max} or T_{a-mean}. Furthermore, their conclusions are also different. Meyer et al.(2016) stated that RF algorithms show the weakest results among linear regression, generalized boosted regression models (GBM), and Cubist regression. In contrast, Xu et al. (2014) concluded that RF outperforms the linear regression. Zhang et al. (2016) divided their data record into two groups (group S1 contains all four MODIS LST under good quality and group S2 had at least one LST with poor quality). The results based on the two datasets are different: in group S1, RF shows the best results in almost all combinations, but in group S2

the best algorithm is the Cubist regression. As a final result, the best algorithm for daily $T_{a\text{-}max}$, $T_{a\text{-}min}$, and $T_{a\text{-}mean}$ estimation remains unknown.

Regarding MODIS LST data (v005), LST data are not available for a location (pixel) if cloudiness is present inside the pixel (Wan, 2008). Due to the differences in satellite overpass times, the valid observation data at a specific location (pixel) varies between LST_{ad}, LST_{an}, LST_{td}, and LST_{tn}. Therefore, it is important to compare the dynamic combination of one to four LST data that are available at different times and locations as well as the most suitable algorithm to apply for T_a estimation. Furthermore, a rising question using LST MODIS solely is the kind of relationship (linear or nonlinear) between LST and T_a, especially in mountainous areas.

Therefore, in this research, we investigate all 15 (i.e., $2^4 - 1$) possible dynamic combinations of four LST with or without auxiliary data for daily T_a estimation using three different algorithms: multiple linear regression (LM), cubist regression (CB), and random forests models (RF). Finally, the accuracies of these T_a-estimated are evaluated by comparison with T_a-measured data, which are collected from weather stations. Root mean square error (RMSE) and coefficient of determination (R^2) were used as the model evaluation scores.

6.2. Materials and Methods

6.2.1. Study Area and Weather Station Data

The study area is located in northwest Vietnam inside two large provinces: Lai Chau and Dien Bien. It covers an area of 18,600 km^2 (Figure 6.1). The study area presents a rural and mountainous region in northwest Vietnam with a sparse distribution of weather stations. There are only four weather stations (Figure 6.1) within these two provinces. However, due to the lack of data measurement, we chose only three stations, Sin Ho, Dien Bien, and Lai Chau, for this study (Table 6.1). In each station, the T_a data were recorded hourly. $T_{a\text{-}max}$ and $T_{a\text{-}min}$ are the highest (maximum) and lowest (minimum) air surface temperatures that occur on a diurnal cycle (24 h cycle), respectively; $T_{a\text{-}mean}$ was calculated by averaging all 24 hourly measurements in a day. Generally, $T_{a\text{-}max}$ occurs after solar noon from one to two hours, and $T_{a\text{-}min}$ usually occurs shortly before dawn. In this study, we collected daily $T_{a\text{-}max}$, $T_{a\text{-}min}$, and $T_{a\text{-}mean}$ from 2009 to 2013 from the Vietnam Institute of Meteorology, Hydrology, and the Environment (IMHEN).
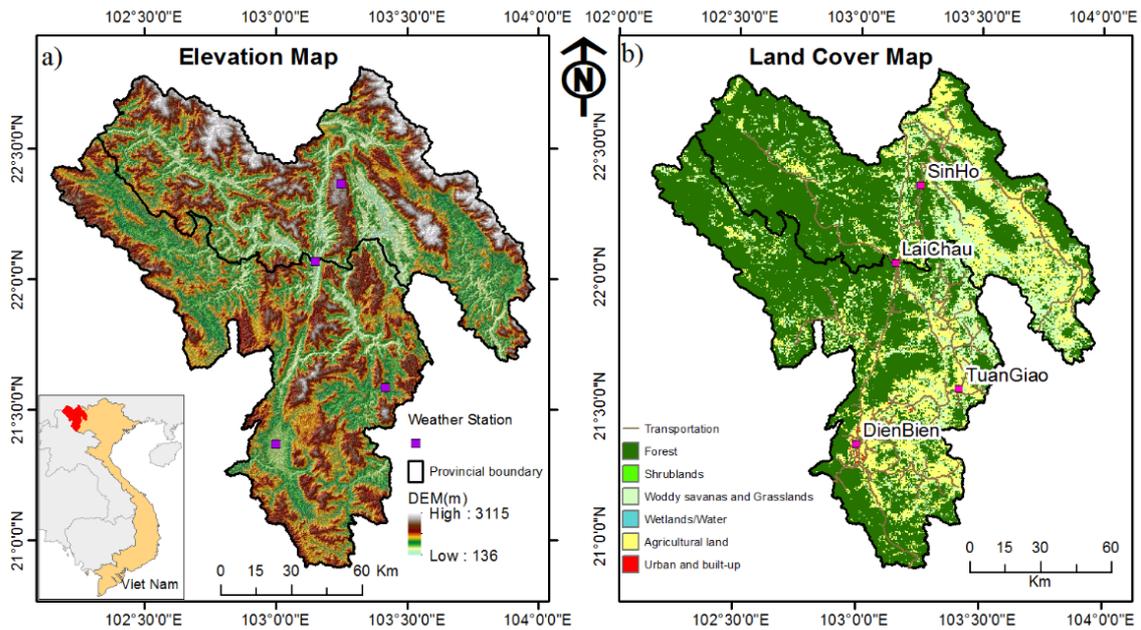
Figure 6.1. Location of the weather stations and range of elevation (**a**) and land cover (**b**) from MODIS MCD12Q1 data in 2010 of the study area.

Table 6.1. Geographical description and land cover type of weather stations used in this study.

No.	Station	Lat (°)	Long (°)	Elevation (m)	Land Cover
1	Sin Ho	22.37	103.25	1534	Forest
2	Dien Bien	21.37	103.00	475	Crop land
3	Lai Chau	22.07	103.15	243	Forest

Based on the MODIS land cover type product (MCD12Q1 data of 2010), the major land cover type in this area is forest, covering approximately 64% (Figure 6.1).

6.2.2. Data

6.2.2.1. MODIS LST

All MODIS LST data used in this study were acquired from the U.S. Geological Survey (USGS) website (http://earthexplorer.usgs.gov).

We used two MODIS LST products (v005, h27v06), MOD11A1 and MYD11A1 from TERRA and AQUA satellites, respectively. The MODIS LST consists of daytime and nighttime data at a spatial resolution of 1 km. Thus, in total there are four LST datasets: AQUA daytime (LST$_{ad}$), AQUA nighttime (LST$_{an}$), TERRA daytime (LST$_{td}$), and TERRA nighttime (LST$_{tn}$).

In the literature, there are some studies that use eight-day LST averages to estimate T$_a$ (Colombi et al., 2007; Mostovoy et al., 2006; Vancutsem et al., 2010). It should be considered

that eight-day-average LST is calculated by averaging all valid data under clear sky conditions, the number of participant data points varying from one to eight days depending on availability. Meanwhile, eight-day-average T_a is calculated by averaging the data under changing sky conditions. Therefore, if we compare eight-day-average LST and eight-day-average T_a, the sampling may introduce uncertainty (Shen and Leptoukh, 2011). Taking this difference into consideration, in this study we decided to use daily LST under clear sky conditions instead of eight-day-average LST data.

6.2.2.2. MODIS Land Cover

The MODIS Land Cover Type Product (MCD12Q1) is downloaded from the Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov). In order to use this product easily in the community, four main classification schemes were provided, including IGBP (International Geosphere–Biosphere Programme), UMD (University of Maryland), LAI/fPAR (Leaf Area Index/fraction of Photosynthetically Active Radiation), and NPP (Net Primary Productivity) (Liang et al., 2015). For our study, we use the primary land cover scheme, which is provided by the IGPB land cover classification. Based on this scheme, our study has 13 types of land cover classes. However, in order to make it easy to use and distinguish between each class, consistent with the land cover of the study area we combined and reduced the classes to six types (Figure 6.1). As is shown in Figure 6.1, the majority of land cover in the study area is forest and cropland.

In addition, based on the results of our previous study (Noi et al., 2016), we take two more variables into account for T_a estimation in northern Vietnam: station elevation (el) and Julian day data. Elevations of stations were obtained from the Vietnam Institute of Meteorology, Hydrology and Environment (IMHEN). The Julian day (jd) was extracted from the NASA server (http://landweb.nascom.nasa.gov/).

6.2.3. Methods

6.2.3.1. Calculating LST of Weather–Station–Location

LST data under clear sky conditions at weather stations are retrieved by the following steps:

- A total of 3652 MODIS images (MOD11A1 and MYD11A1, h27v06, Collection 5, from 1 January 2009 to 31 December 2013, over northern Vietnam) in HDF (Hierarchical Data Format) format were reprojected to WGS_1984_UTM_zone_48N using the nearest neighbor resampling method with the MODIS Re-Projection Tool. The corresponding layers (LST_Day_1km, LST_Night_1km, Daytime LST observation time, and Nighttime

LST observation time) were extracted in TIF format. However, Daytime and Nighttime LST observation time were used in order to identify the approximate overpass time of MODIS at local time.

- MODIS LST data for the pixels in which the weather stations are located are extracted from 7304 TIF format MODIS images (3652 daytime and 3652 nighttime images) using batch processing of extract multi value to points in ArcGIS 10.3.

- All these LST data (DN value) were converted to Celsius temperature using the following equation:

°C = 0.02 * DN − 273.15,

where °C is the Celsius temperature and 0.02 is the scale factor of the MODIS LST product.

- Removing outlier data: MODIS LST products are not available for a location (pixel) if clouds are present (Wan, 2008). However, there are some pixels that are lightly covered or contaminated by clouds. These pixels are not removed because the contamination is very small and cannot be detected by the cloud-removing mask algorithm (Ackerman et al., 2008; Williamson et al., 2013). To avoid this kind of data, we studied and developed a similar method that was used in (Williamson et al., 2017). This approach includes two steps: First, we simply filter and remove all unrealistic LST data that had values greater than 100 °C and/or below −50 °C. Second, we calculated the difference between T$_{a\text{-}max}$ versus LST daytime and T$_{a\text{-}min}$ versus LST nighttime. Then, we applied statistical outlier removal based on these differences' histograms to detect and remove data with unusually large differences (the histogram does not follow a normal distribution).

6.2.3.2. Estimation Air Temperature Using MODIS LST Data

- Dynamic Combination of MODIS LST data

To estimate daily T$_a$, we used all possible combinations of four LST data (LST$_{ad}$, LST$_{an}$, LST$_{td}$, and LST$_{tn}$). These 15-combinations are shown in Table 6.2.

Table 6.2. All possible combinations of four LST data and the valid number of observations.

No.	Combination		SinHo	DienBien	LaiChau	Total
C01	LST$_{ad}$		488	572	571	1631
C02	LST$_{an}$		420	321	261	1002
C03	LST$_{td}$		427	500	507	1434
C04	LST$_{tn}$		562	593	528	1683
C05	LST$_{ad}$	+LST$_{an}$	254	219	190	663
C06	LST$_{td}$	+LST$_{tn}$	255	286	298	839
C07	LST$_{ad}$	+LST$_{td}$	297	318	348	963

C08	LST_{an}	$+LST_{td}$		231	193	176	600
C09	LST_{ad}	$+LST_{tn}$		283	348	340	971
C10	LST_{an}	$+LST_{tn}$		294	224	193	711
C11	LST_{td}	$+LST_{tn}$	$+LST_{ad}$	195	200	229	624
C12	LST_{td}	$+LST_{tn}$	$+LST_{an}$	176	132	131	439
C13	LST_{ad}	$+LST_{an}$	$+LST_{td}$	184	138	137	459
C14	LST_{ad}	$+LST_{an}$	$+LST_{tn}$	198	159	143	500
C15	LST_{ad}	$+LST_{an}$	$+LST_{td}$ $+LST_{tn}$ 141	92	100	333

Due to the cloud cover effect, the number of valid observations from each station and each combination (C01–C15) are various (Table 6.2).

In order to investigate the difference between dynamic combinations, as well as the performance of different algorithms, we used two datasets: Dataset A, MODIS LST data only; and Dataset B, MODIS LST together with elevation (ele) and Julian day (jd) data.

- Algorithms used

Linear/Multiple Linear Regression Model (LM) is a model that represents the relationship between one response variable and one predictor variable (Simple Linear Regression) or more than one predictor variable (Multiple Linear Regression) by using parameters entered linearly and estimated by the least squares method. So far, LM is one of the most popular statistical models for T_a estimation using MODIS LST (Vancutsem et al., 2010; Noi et al., 2016; Shen and Leptoukh, 2011; Xu et al., 2014; Huang et al., 2015; Shi et al., 2016). Although it was found that the correlation between LST and T_a is high, this relationship may not actually be linear (Zhang et al., 2016). Therefore, our current knowledge might be incomplete if we do not try machine learning algorithms. Machine learning algorithms promise a better estimation of T_a using MODIS LST because they can handle non-linearity and highly correlated predictor variables (Meyer et al., 2016; Kuhn and Johnson, 2013; James et al., 2013). Furthermore, based on the conceptual designs of machine learning algorithms, they are able to deal with data that have a different relationship between predictor and response variables under different conditions such as season, elevation, and land cover characteristic (Meyer et al., 2016).

Random Forests (RF), which was proposed by Breiman (2001), is a nonparametric and ensemble technique. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It is different from traditional statistical methods that contain a parametric model for prediction. In RF, it contains many decision trees, where each tree is built from a random subset of training data with a random subset of predictor

variables. The final predicted values are produced by the aggregation of the results of all the individual trees that make up the forest (Xu et al., 2014).

Cubist regression (CB) is a rule-based regression technique that was developed based on a combination of the ideas of Quinlan (1992, 1993a, 1993b). CB does not retrieve one final model like RF, but a set of rules associated with sets of multi-variate models. Then, a specific set of predictor variables will choose an actual prediction model based on the rule that best fits the predictors (Appelhans et al., 2015). Cubist is a commercial, proprietary product and has the least algorithmic documentation in comparison to linear regression and random forest (Walton, 2008). However, it is currently a popular and widely used regression and classification method because it was ported into R by Kuhn et al. (2013). Most recently, it was used in T$_a$ estimation research and showed very good results in the research of Meyer et al. (2016) and Zhang et al. (2016).

Therefore, in this study, to estimate T$_a$ and assess the accuracy of estimation, three different methods were employed: linear regression (LM), cubist regression (CB) and random forests (RF). All methods are performed in the R statistical software.

6.2.3.3. Comparison of Different Combination and Algorithms

- Assessment Criteria

To assess the performance of models, we used and compared the values of the two most popular criteria: the coefficient of determination (R^2) and the root mean square error (RMSE) that were calculated from the measured and estimated T$_a$ values from three algorithms: LM, CB, and RF.

- Comparison

Being one of the most popular validation methods, cross-validation was used in order to compare different combinations and different algorithms.

In order to implement the cross-validation, the dataset is divided into k groups (k-fold) of approximately the same size. Then, $k-1$ groups of the dataset are used as the training set, and the left-out group is used for validation. When the number of groups (k) equals the number of observations (n), it is called "leave-one-out cross-validation".

Due to the high number of observations, we used 10-fold cross-validation (k = 10) and repeated it twice for cross-validation.

6.3. Results

6.3.1. The Relationship between Ta and LST MODIS

In order to evaluate the MODIS LST data for T_a estimation, we first test the relationship between T_a and LST MODIS of all three weather stations.

Figure 6.2 shows the scatter plot of T_a and LST of daytime and nighttime. It was found that: The coefficients of determination were high, ranging from 0.43 to 0.72, and the correlation between LST nighttime and T_{a-min} were higher than LST daytime and T_{a-max} at Dien Bien and Sin Ho stations. However, at Lai Chau station, the correlation between LST daytime versus T_{a-max} was slightly higher than nighttime versus T_{a-min}. This indicates that the relationship between MODIS LST and T_a of this study area is complex.
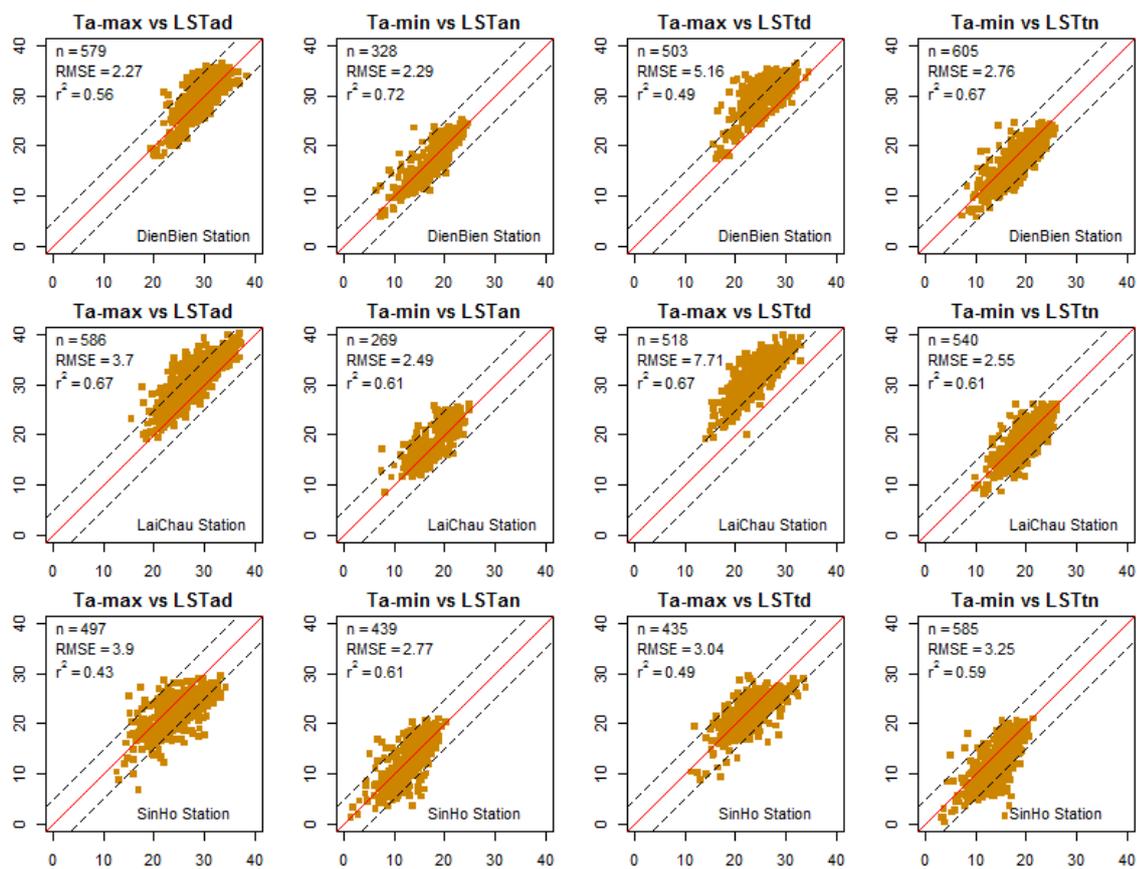
Figure 6.2. The relationship between LST (x-axis) and T_{a-max} (first and third columns), T_{a-min} (second and last columns) of all meteorological stations from 2009 to 2013. The dashed line indicates that the difference between T_a and LST is over ±5 °C (±5 line). The red line indicates the 1:1 line.

The relationships between T_{a-min} versus LST_{an} and LST_{tn} were quite similar at all three stations. However, the relationships between T_{a-max} versus LST daytime were different; at Lai Chau station most T_{a-max} values are higher than the LST_{ad} and LST_{td} values (most of the points lie above the red line). Meanwhile, at Dien Bien station, T_{a-max} is quite similar to LST_{ad} but T_{a-}

154

$_{max}$ was higher than LST$_{td}$. At Sin Ho station, there is not much difference between T$_a$ versus LST but there are a lot of data points lying outside the "±5 lines".

Due to the cloud effect, the number of valid observations changes from daytime to nighttime; during the daytime the AQUA sensor gives more observations than TERRA. However, at night, TERRA gives more observations than AQUA.

6.3.2. Different Combinations of MODIS LST for Ta Estimation

As shown in Figure 6.2, the linear correlation between T$_a$ and LST are strong for both Terra LST and Aqua LST of daytime and nighttime. Furthermore, in Section 1 we also showed that there are plenty of studies using MODIS LST data for T$_a$ estimation using the LM method.

Therefore, in order to estimate the effect of different MODIS LST data combinations, we applied all three methods (linear regression, cubist regression, and random forest models) to the 15 combinations as shown in Table 6.2.

In the LM method, the equations of 15 combinations (C01–C15) are shown in Appendixe A and Appendixe B. However, regarding CB and RF, which are nonparametric methods, equations cannot be provided as for the LM method.

6.3.2.1. Combinations Using One LST Variable

Figure 6.3a,b show the coefficient of determination (R^2) and root mean square error (RMSE) of combinations C01–C04 using three algorithms (LM, CB, and RF) with Dataset A and Dataset B, respectively. It can be clearly seen that there is a large difference between Figure 6.3a (using LST solely) and Figure 6.3b (using LST with elevation and Julian day data). At Figure 6.3a, LM and CB show similar results and higher accuracy than the RF algorithm in all four combinations (C01–C04). In contrast, Figure 6.3b shows similar results for CB and RF in all four combinations and slightly higher values than with the LM algorithm. It is suggested that when one LST is used with an auxiliary data for T$_a$ estimation, RF and CB performance are better than LM.
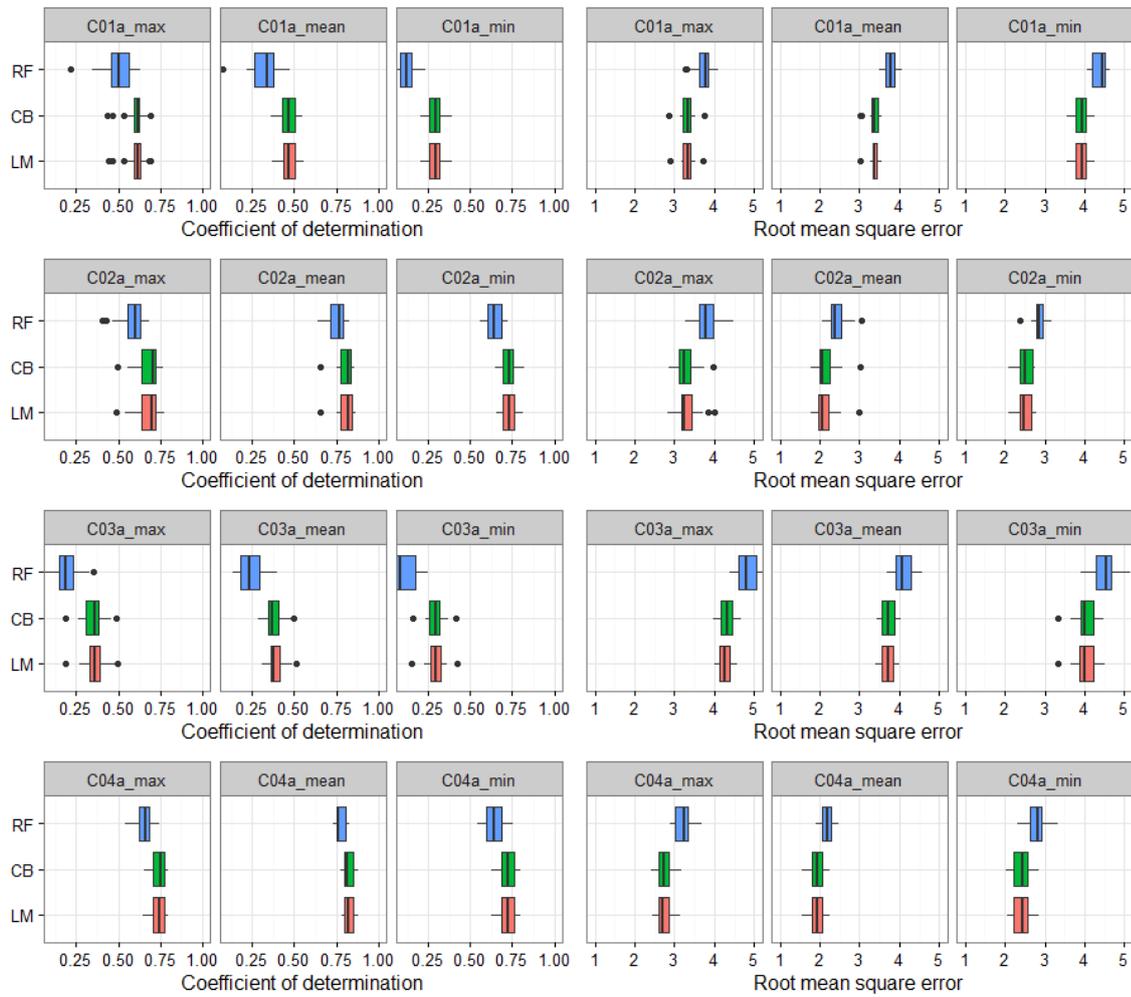
Figure 6.3. (a) Cross-validation results for one-LST-combination (C01–C04) using Dataset A, and multiple comparisons of the three algorithms. The x-axis shows the value of R^2 and RMSE (°C), the y-axis shows the model types. The box and whiskers plots show the distributions of R^2 and RMSE.
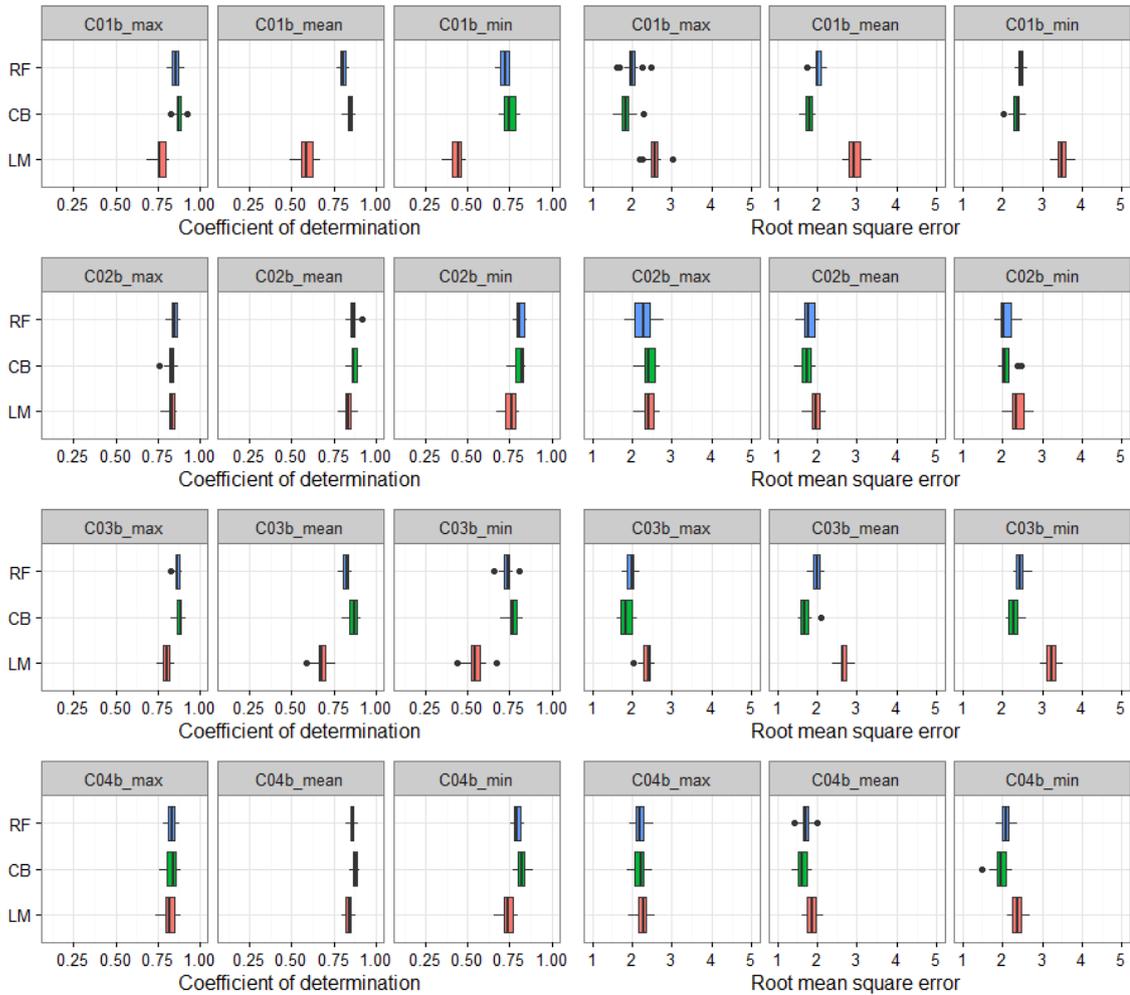
Figure 6.3. (**b**) Cross-validation results for one-LST-combination (C01–C04) using Dataset B, and multiple comparisons of the three algorithms. The x-axis shows the values of R² and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R² and RMSE.

Both Figure 6.3a,b show that the accuracy of C02 and C04 is much higher than for C01 and C03 (higher value of R^2 and lower value of RMSE). It can be stated that nighttime LST was better than daytime for deriving daily T$_a$. This result is consistent with (Noi et al., 2016; Zhang et al., 2011). Regarding the two datasets used, in all combinations (C01–C04) the accuracies of T$_a$ estimation using Dataset B are much higher than when using Dataset A.

For T$_{a-min}$ and T$_{a-mean}$ estimation, Figure 6.3a shows that the combinations using LST nighttime (C02 and C04) have significantly higher accuracy than the combinations using LST daytime (C01 and C03). However, these differences are not clearly shown in Figure 6.3b (except for in the LM results).

Regarding Dataset A, AQUA daytime (C01) shows better results for T$_{a-max}$ estimation than TERRA daytime (C03). However, at night AQUA and TERRA show similar results for T$_a$ estimation. The results of both daytime and nighttime of TERRA and AQUA are consistent and similar in T$_a$ estimation (Figure 6.3b).

157

6.3.2.2. Combinations Using Two-LST Variables

In this case, we used all possible combinations with LST to estimate T_a. As shown in Table 6.2, we applied six possible combinations of LST for T_a estimation.

In general, Figure 6.4a,b show that both results of T_a estimation using Dataset A and B are higher than the one-LST-combination (Figure 6.3a,b). Figure 6.4a shows that the difference between the three algorithms is not as large as in the results shown in Figure 6.3a (except for C07).
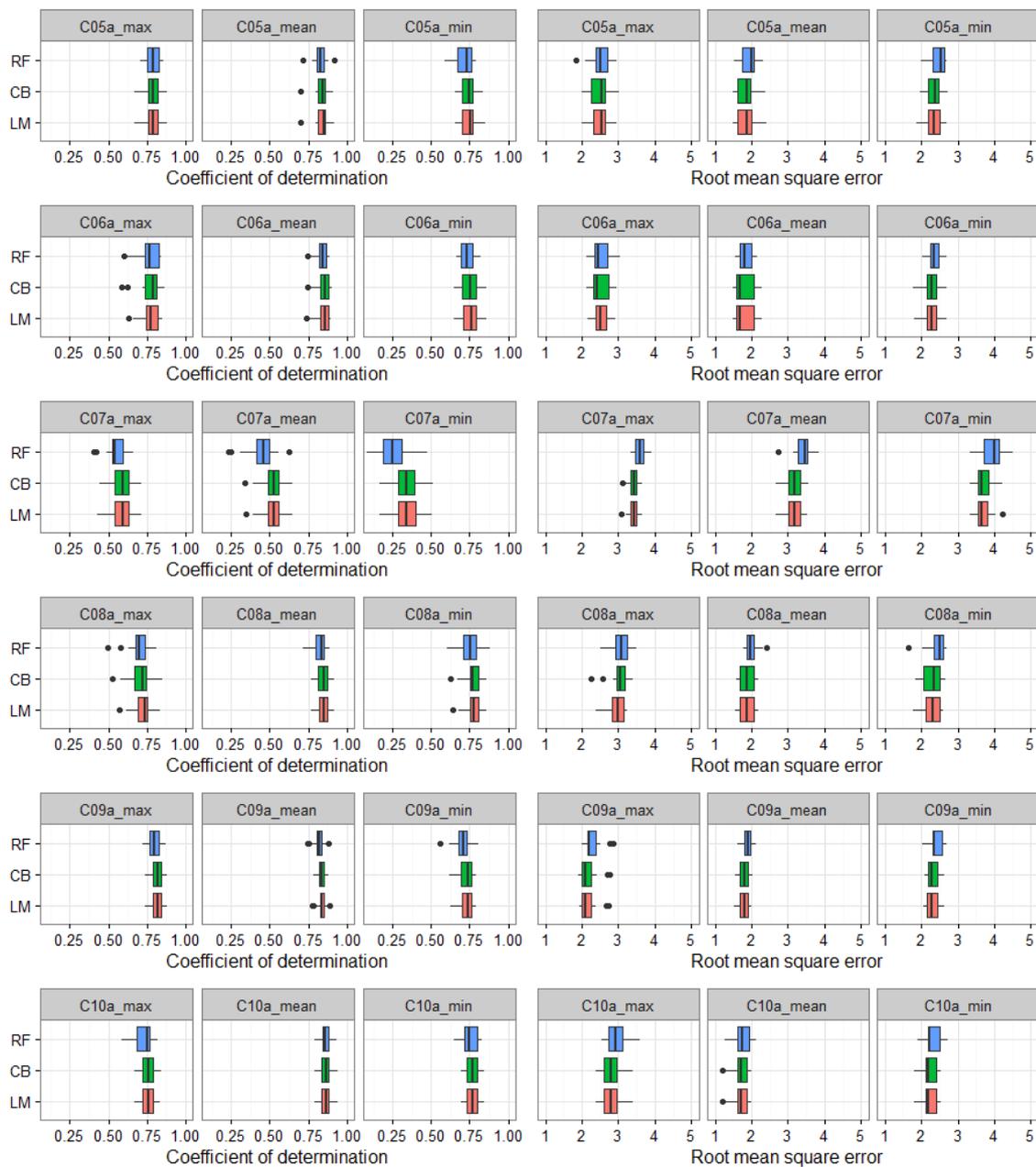
Figure 6.4. (a) Cross-validation results for two-LST-combinations (C05–C10) using Dataset A and multiple comparisons of the three algorithms. The x-axis shows the value of R2 and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R2 and RMSE

Figure 6.4. (b) Cross-validation results for two-LST-combinations (C05–C10) using Dataset B and multiple comparisons of the three algorithms. The x-axis shows the values of R2 and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R2 and RMSE.

In these combinations (C05–C10), CB and LM show similar and slightly higher accuracies than RF for Dataset A. The contrast is also evident in Dataset B: the CB and RF results are similar and slightly higher than LM. Especially in C07, the results of LM are much lower than those of CB and RF (Figure 6.4b). The results of all T_a estimations using Dataset B are still higher than using Dataset A.

Looking at the first two rows of Figure 6.4a,b (C05, combined LST_{ad} + LST_{an}; and C06, combined LST_{td} + LST_{tn}), there are similar results for T_a estimations between them. It is indicated that the overpass times of AQUA and TERRA do not significantly affect the result

159

of T_a estimation when we combine daytime and nighttime LST. This is true for all three methods (LM, CB, and RF). These results are also consistent with previous studies (Zhang et al., 2011; Benali et al., 2012; Noi et al., 2016), which used LM as the statistical model for T_a estimation.

The most interesting finding of two-LST-combined is the combination of C07. The results of Dataset A (panel row 3, Figure 6.4a) show the lowest accuracy in comparison to five other two-LST-combined (R^2 approximately 0.6, 0.5 and 0.35; RMSE approximately 3.5, 3.2, and 3.7 °C for T_{a-max}, T_{a-mean}, and T_{a-min}, respectively). In addition, among the three algorithms, RF shows the lowest results with lower R^2 and higher RMSE. In contrast, the results of Dataset B are absolutely different (Figure 6.4b, panel row 3). The results of C07 (using Dataset B) are similar to the five other two-LST-combined (R^2 approximately 0.88, 0.80, and 0.73; RMSE approximately 1.8, 1.9, and 2.5 °C for T_{a-max}, T_{a-mean}, and T_{a-min}, respectively, except for the results of LM) and much higher than using Dataset A. Among the three algorithms, the lowest result for T_a estimation is LM (Figure 6.4b). Meanwhile, CB and RF show higher results, especially for T_{a-min} and T_{a-mean} estimation. It should be noted that C07 is the combination of TERRA and AQUA daytime LST, which is the most complicated in the relationship between T_a and LST in comparison to the rest of the combinations. The difference between the results of Datasets A and B indicates that elevation and Julian day (i.e., season) also affect the relationship between LST and T_a. This is consistent with the results from (Benali et al., 2012; Zeng et al., 2015; Peón et al., 2014; Janatian et al., 2016). The high accuracy of T_a estimation using the RF and CB algorithms in Figure 6.4b also indicates that RF and CB can account for the complicated relationship between predictor and response variables under different conditions, especially in mountainous area. This finding is consistent with the studies by Zhang et al. (2016) and Xu et al. (2014).

6.3.2.3. Combinations Using Three-LST Variables

In general, Figure 6.5a,b show that all three-combined LST result in a very high accuracy of T_a estimation and the differences in accuracy between the three different algorithms are not significant (p-value > 0.05). However, the results of T_a estimation using Dataset B are much higher than using Dataset A. In both datasets, the results of T_{a-max} and T_{a-mean} are always better than T_{a-min} (except C12 and C14 of Dataset A). This can be explained by the fact that, because of two LST nighttime variables (LST_{tn} and LST_{an}) in C12 and C14, the accuracy of T_{a-min} estimation could be increased. However, in Dataset B, by introducing the two variables elevation and Julian day, the accuracy of all T_{a-max}, T_{a-min}, and T_{a-mean} estimations has increased

(T$_{a\text{-max}}$ and T$_{a\text{-mean}}$ is increased more significantly than T$_{a\text{-min}}$ when elevation and Julian day data were introduced).
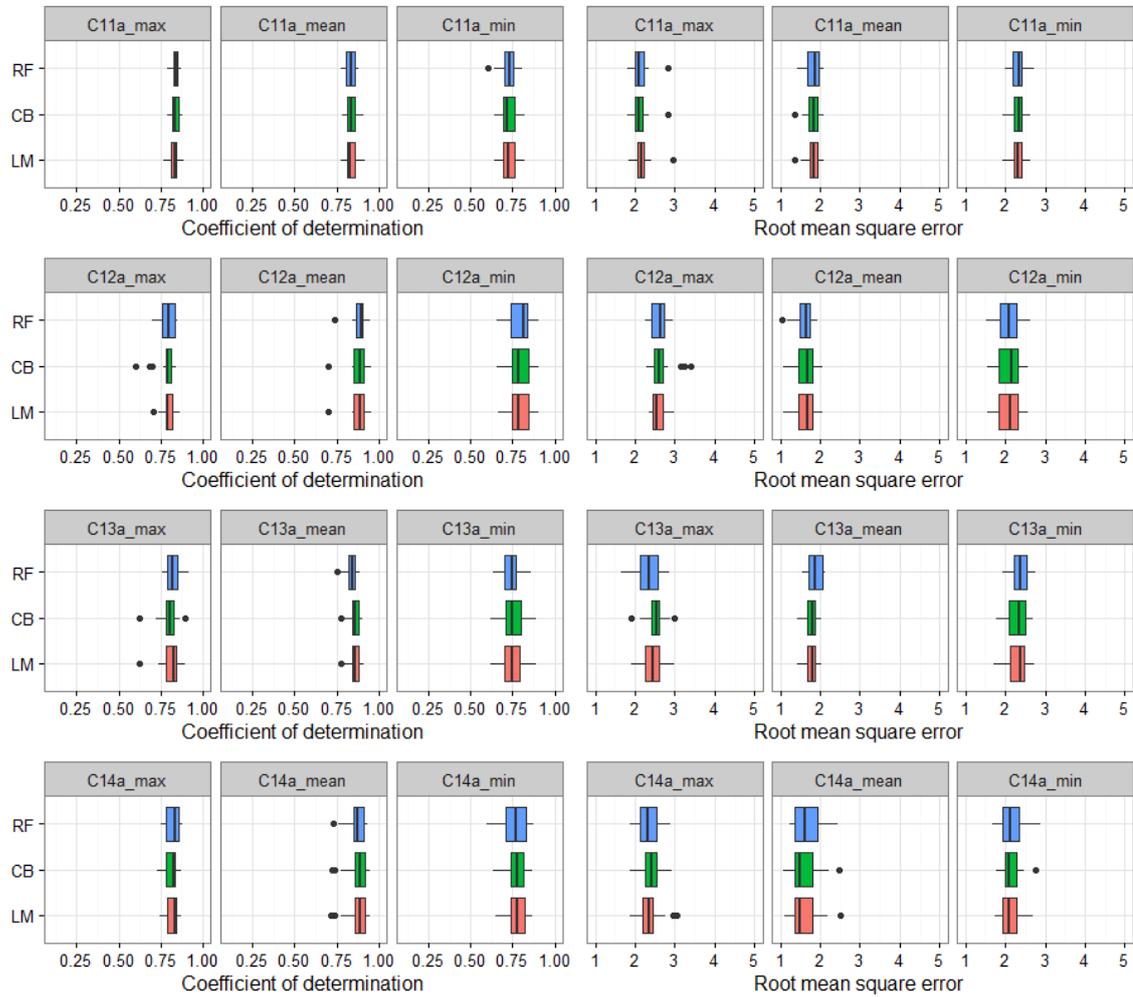
Figure 6.5. (a) Cross-validation results for three-LST-combinations (C11–C14) using Dataset A and multiple comparisons of the three algorithms. The x-axis shows the values of R2 and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R2 and RMSE.
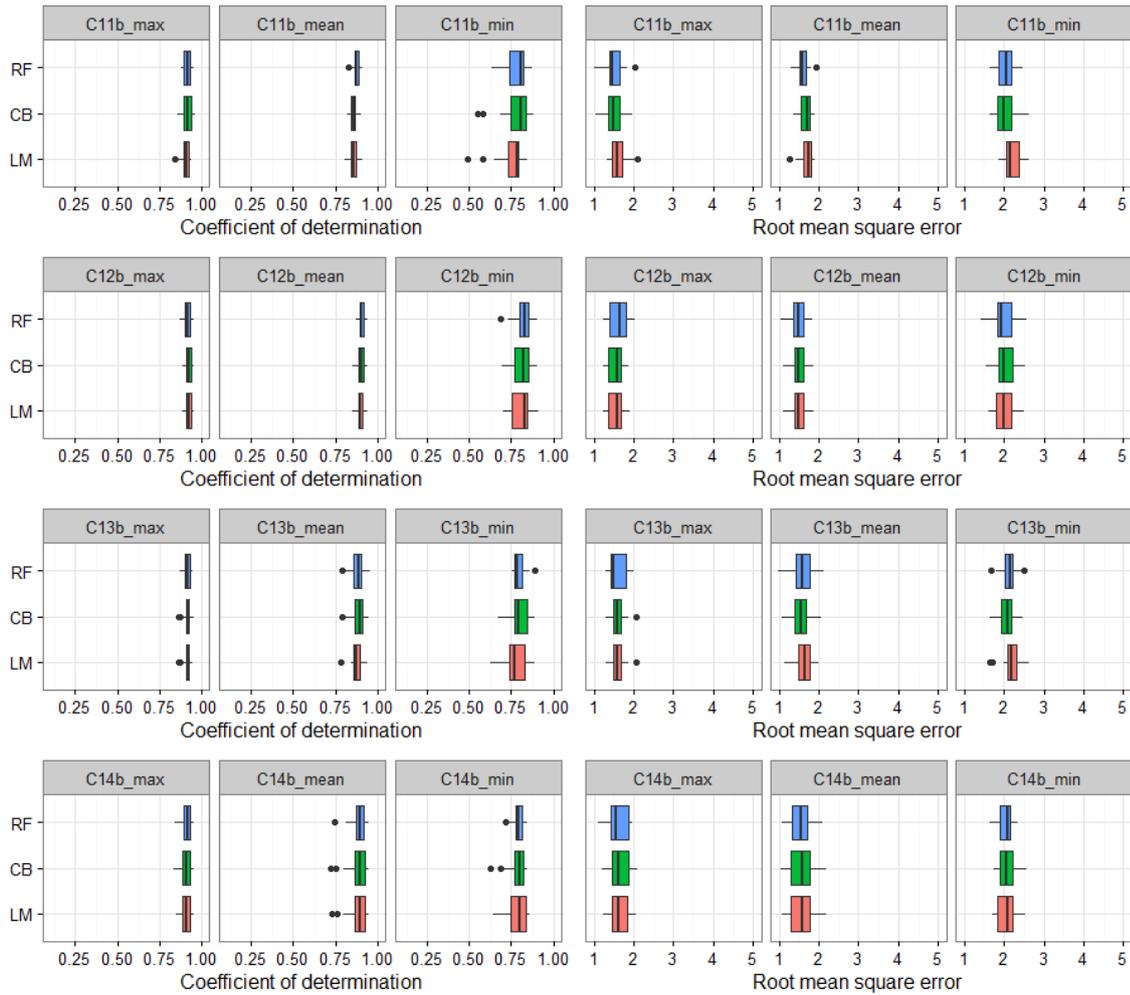
Figure 6.5. (b) Cross-validation results for three-LST-combinations (C11–C14) using Dataset B and multiple comparisons of the three algorithms. The x-axis shows the value of R² and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R² and RMSE.

6.3.2.4. Combinations Using Four-LST Variables

The first result clearly seen from Figure 6.6 is that all three algorithms show a similar accuracy of T_a estimation in both Dataset A and B. However, the results of Dataset B (R^2 approximately 0.93, 0.89 and 0.8, RMSE approximately 1.5, 1.6, and around 2.0 °C for $T_{a\text{-max}}$, $T_{a\text{-mean}}$, and $T_{a\text{-min}}$, respectively) are much higher than the results of Dataset A (R^2 approximately 0.84, 0.88, and 0.75; RMSE roughly 2.2, 1.7, and 2.2 °C for $T_{a\text{-max}}$, $T_{a\text{-mean}}$, and $T_{a\text{-min}}$, respectively).
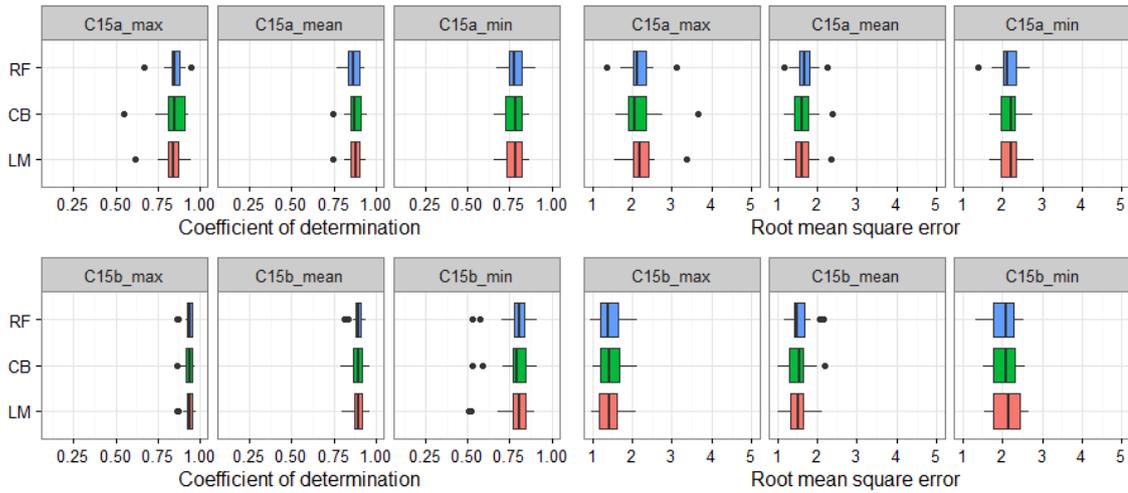
Figure 6.6. Cross-validation results for four-LST-combinations (C15) using Dataset A (upper rows) and B (lower rows) and multiple comparisons of the three algorithms. The x-axis shows the values of R^2 and RMSE (°C); the y-axis shows the model types. The box and whiskers plots show the distributions of R^2 and RMSE.

In addition, the statistical results also indicate that the difference between the three algorithms is not significant (p-value > 0.05) in either Dataset A or B.

6.4. Discussion

6.4.1. Model Calibration and Validation

In several previous studies (Zeng et al., 2015; Huang et al., 2015), one of the most common validation methods is that the sample data is randomly divided into a calibration and a validation dataset (e.g., 70% and 30% respectively). Calibration data were used for training data and validation data were used to assess the model performance. However, there is a drawback with this random choice: If we use a local dataset to train the model (i.e., a dataset that does not represent all dataset characteristics), then we apply a fitted model to the validation data. This could be misleading in the accuracy assessment. Especially in machine learning algorithms like CB or RF, this could lead to overfitting problems (e.g., the accuracy of the training part is very high; however, the model cannot be applied successfully to the validation dataset).

In this paper, we studied this problem in T_a estimation using MODIS LST. First, we randomly divide the data of all 15 combinations into two datasets: calibration and validation (70% and 30%, respectively). Next, we fitted the model using a calibration dataset, and then we applied the fitted model to the validation dataset and the entire dataset. Finally, we assessed the accuracies of validation data, full data, and cross-validation.

These processes are applied to both Dataset A and Dataset B.

In Figure 6.7, the LM algorithm shows consistent results between the validation data, the total data, and the cross-validations of both Dataset A and B. The results of T_a estimation using Dataset B (right-hand panel) are slightly higher than with Dataset A (left-hand panel). It could be suggested that when LST data alone were used (without auxiliary data), the accuracy of T_a estimation could be affected by a change in season or the elevation of the weather station. This is consistent with previous studies (Huang et al., 2015; Noi et al., 2016). In the CB method (Figure 6.8), the results of validation, full data, and cross-validation are also consistent with each other. However, in both algorithms LM and CB, the results of Dataset A and Dataset B showed a significant difference, especially the combinations 1, 3, and 7 (C01, C03, and C07), where there is only LST daytime data. It is suggested that if LST nighttime is not available then the accuracy of T_a estimation could be improved by adding auxiliary data. Comparing Figure 6.7 and Figure 6.8, it can be clearly seen that CB produces better results for T_a estimation than LM.
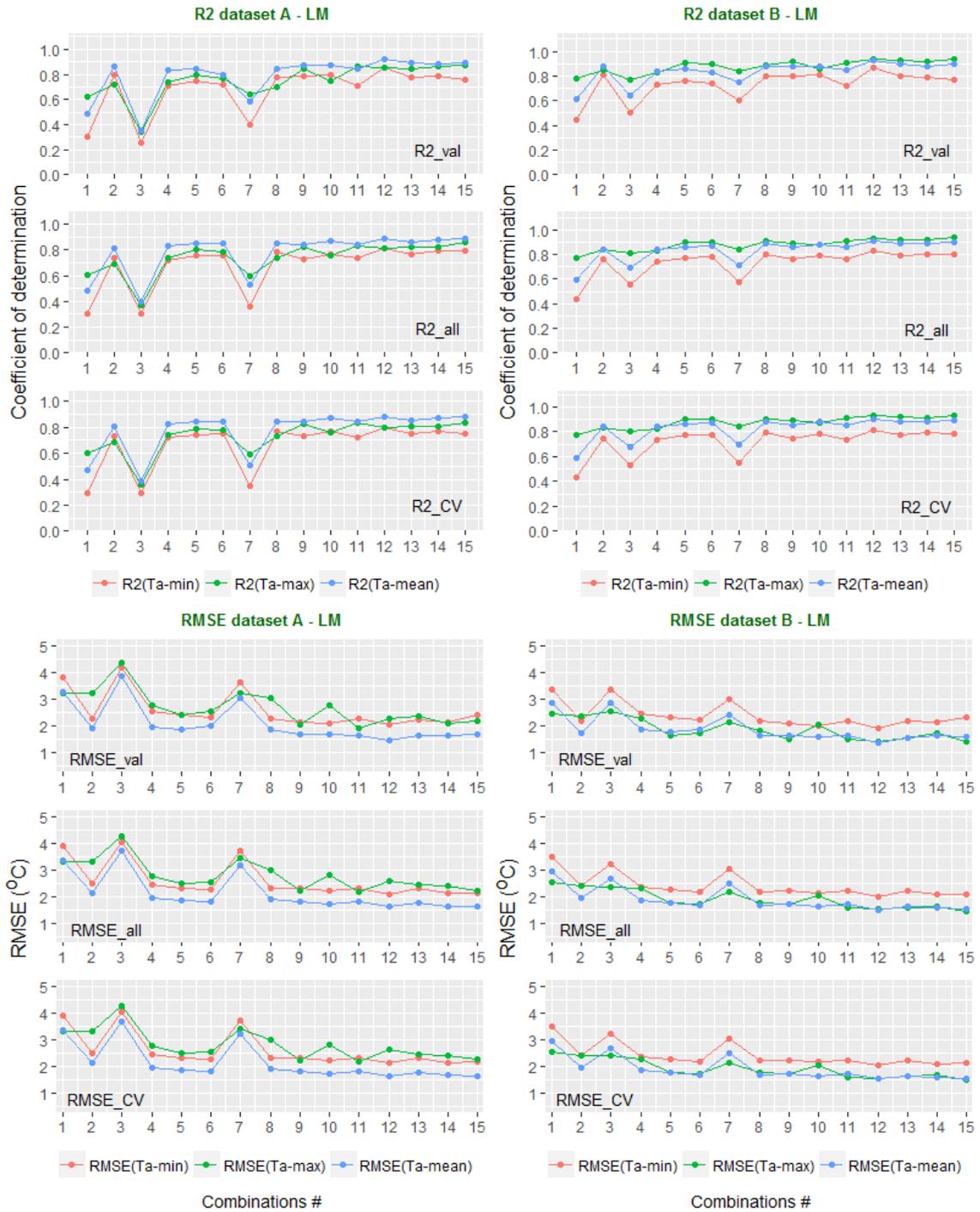
Figure 6.7. Comparison of accuracy (R² and RMSE) when applying the LM algorithm to the validation dataset (_val), the full dataset (_all), and a cross-validation (_cv) of all combinations. The x-axis shows the combination number. The y-axis shows the values of RMSE (°C) and R².
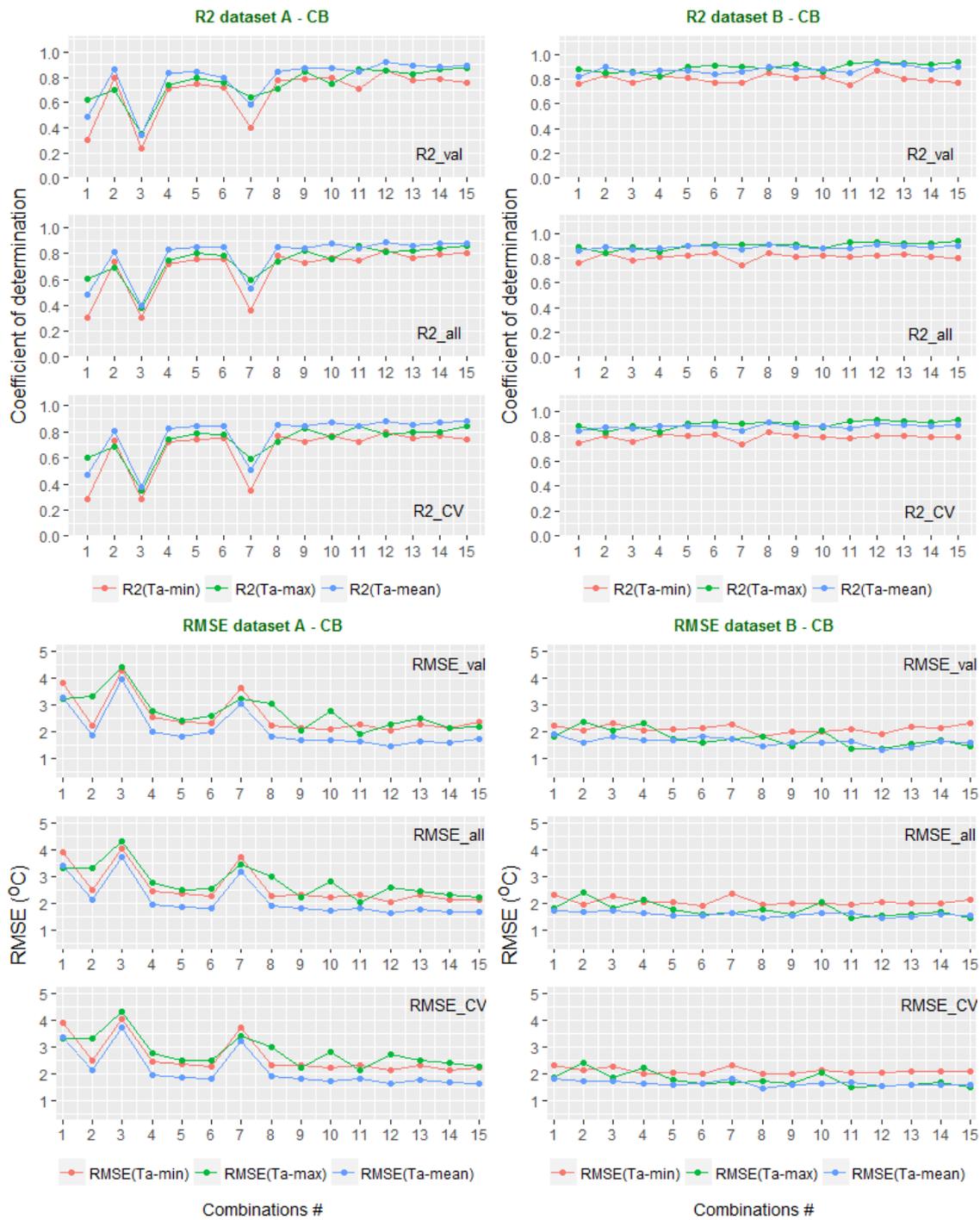
Figure 6.8. Comparison of accuracy (R² and RMSE) when applying the CB algorithm to the validation dataset (_val), the full dataset (_all), and a cross-validation (_cv) of all combinations. The x-axis shows the combination number. The y-axis shows the values of RMSE (°C) and R².

Unlike LM and CB, the results of RF algorithm (Figure 6.9) are not consistent when applied to the validation data, full data, and cross-validation using Dataset A or Dataset B. As is shown in Figure 6.9, the results of cross-validation and the results using the validation data are similar and lower than when using the full data. It is suggested that the RF algorithm could be overfitting the T_a estimation using MODIS LST. It is also clearly seen that the

results of T$_a$ estimation using Dataset B are much higher than Dataset A, especially the combinations C01, C03, and C07. Again, the results of RF confirm that auxiliary data (i.e., elevation and Julian day) together with the RF algorithm can increase the accuracy of T$_a$ estimation, especially in the case of missing LST nighttime data (i.e., combinations C01, C03, and C07).
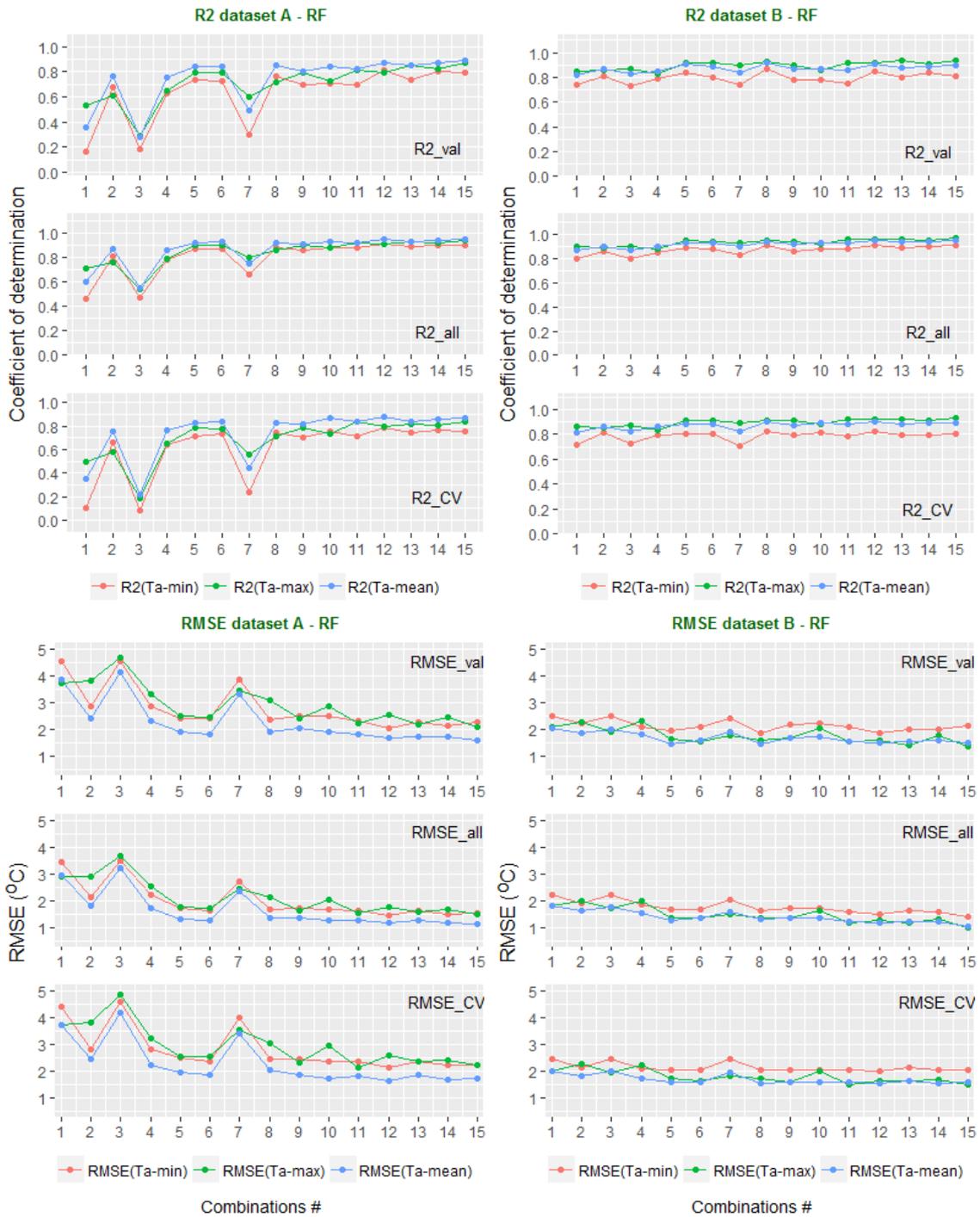
Figure 6.9. Comparison of accuracy (R^2 and RMSE) when applying the RF algorithm to the validation dataset (_val), the full dataset (_all), and a cross-validation (_cv) of all combinations. The x-axis shows the combination number. The y-axis shows the values of RMSE (°C) and R^2.

6.4.2. Effects of Different Combinations and Statistical Model Applications

Figure 6.10 shows a comparison between the 15 combined LST datasets when applied to three different algorithms (LM, CB, and RF), based on the criteria of R^2 and RMSE.
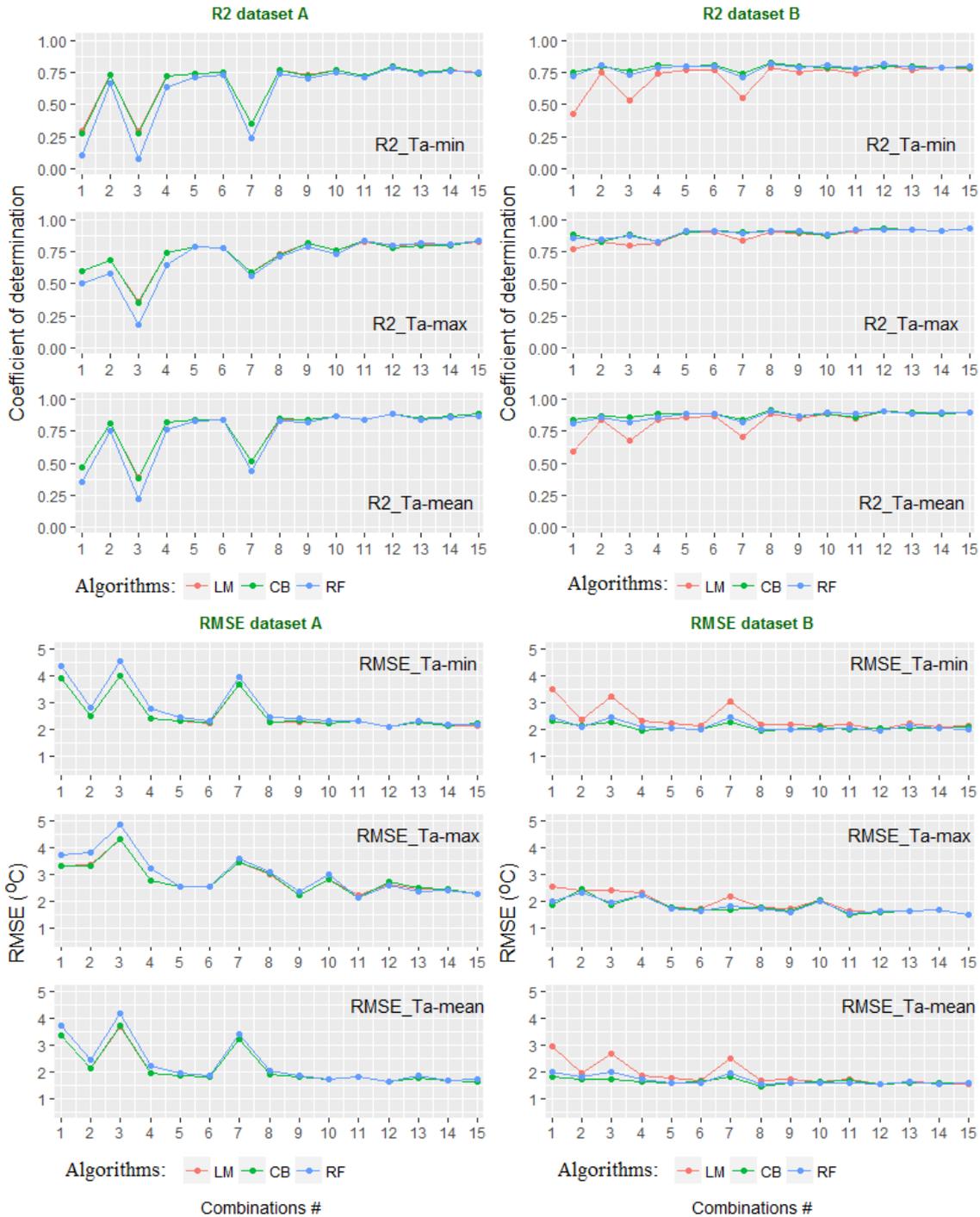
Figure 6.10. Different performance of the algorithms LM (red), CB (green), and RF (blue) through 15 combinations of Dataset A and Dataset B. The x-axis shows the combination number. The y-axis shows the values of RMSE (°C) and R².

Regarding Dataset A, in all combinations (C01–C15) for all T_{a-max}, T_{a-min}, and T_{a-mean} estimations, the results of the LM and CB algorithms are similar and higher than RF.

However, from C10 to C15, the differences between the three algorithms are not clear. The results of combinations C01, C03, and C07 are much lower than the rest of the combinations in all three algorithms.

Considering Dataset B, the results are very different to those of Dataset A. Especially, in combinations C01, C03, and C07, the results of CB and RF are similar and much higher than LM. This can be explained by the fact that during the daytime, solar radiation affects the thermal infrared signal, and the relationship between T$_a$ and LST becomes more complicated. That is why simple models like C01, C03, and C07 (of Dataset A) cannot handle this relationship well. The results of all combinations (C01 to C15) were quite similar when the CB and RF algorithms were applied.

It can be clearly seen that, in all combinations (C01–C15) of Dataset B, the cubist regression always shows the highest accuracy of T$_a$ estimation (slightly higher than RF and much higher than LM). This is consistent with the studies of (Zhang et al., 2016; Xu et al., 2014). It should be remembered that Xu et al. (2014) used MODIS LST and many other auxiliary variables like NDVI, longitude, latitude, etc. In this case, it could be explained by the complex terrain of the study area. It is suggested that the differences in topography, land surface properties, solar radiation, and many other factors could affect the relationships between T$_a$ and LST (Vancutsem et al., 2010; Jin and Dickinson, 2010; Shreve, 2010; Fu et al., 2011). Therefore, a linear regression model, considered as a single global model, could not handle the complicated relationship between T$_a$ and the abovementioned variables under different conditions (Xu et al., 2014). In contrast, CB and RF can account for the nonlinear and complicated relationship between the predictor and response variables under different conditions. That is why, in this mountainous study area, the cubist regression and random forest algorithms always show better results than LM in all 15 combinations (Figure 6.10, right panel).

However, from combination number C02 and C04 to 15 (except number 7—C07), which have at least one nighttime LST term in the combination, the performances of all three methods are good (high correlation and low errors).

Another point is that in Dataset A, the different combinations of LST had a similar effect on all three algorithms. However, in Dataset B, the different combinations of LST had a similar effect on RF and CB but a significantly different effect on the LM algorithm. The largest difference was found in T$_{a-min}$ estimation, follow by T$_{a-mean}$ and T$_{a-max}$ estimation.

6.5. Conclusions

This study proved that the very high accuracy of T_a estimation ($R^2 > 0.93/0.80/0.89$ and RMSE ~1.5/2.0/1.6 °C of $T_{a\text{-max}}$, $T_{a\text{-min}}$, and $T_{a\text{-mean}}$, respectively) could be achieved with a simple combination of four LST data, elevation, and Julian day data using a suitable algorithm.

Using Dataset B (MODIS LST, elevation, and Julian day) with RF or CB algorithms would give a stable and high accuracy in all combinations (C01–C15). With the LM algorithm, the more LST terms (especially LST nighttime) are presented the higher the accuracy that can be achieved.

The impact of the different combinations is larger in Dataset A than in Dataset B. However, in Dataset B, this impact was also large when using the LM algorithm.

LST nighttime data of both AQUA and TERRA play an important role in daily T_a estimation, guaranteeing higher accuracy. Depending on LST data availability, it could be used in any combination from C02, C04, and C05 to C15 (except C07 and C09) to achieve the highest results solely with MODIS LST using any of the three mentioned algorithms. However, when MODIS LST and auxiliary (elevation and Julian day) are available, any combination (C01–C15) can be applied with the CB or RF algorithm.

Among $T_{a\text{-max}}$, $T_{a\text{-min}}$, and $T_{a\text{-mean}}$, using Dataset A, $T_{a\text{-mean}}$ was estimated with the highest accuracy, followed by $T_{a\text{-min}}$ and $T_{a\text{-max}}$. However, the difference between $T_{a\text{-max}}$ and $T_{a\text{-min}}$ was not significant. Considering Dataset B, $T_{a\text{-max}}$ was estimated with the highest accuracy, followed by $T_{a\text{-mean}}$ and $T_{a\text{-min}}$. This means that the highest improvement for $T_{a\text{-max}}$ is made by introducing elevation and Julian day data, followed by $T_{a\text{-mean}}$ and $T_{a\text{-min}}$. However, the difference between $T_{a\text{-max}}$ and $T_{a\text{-mean}}$ was not significant.

Appendix A

Table A1. Parameters of LM models for T_a estimation using Dataset A.

	Combination	a_0	a_1	a_2	a_3	a_4
$T_{a\text{-min}}$ Estimation	C01	−0.4567	0.6037			
	C02	−0.2678	1.0020			
	C03	−1.1905	0.7170			
	C04	−1.7601	1.0184			
	C05	0.1561	−0.0656	1.0647		
	C06	−3.6700	0.0382	1.0329		
	C07	−4.1084	0.4906	0.2442		
	C08	−1.4168	1.0031	0.0277		
	C09	−2.1783	−0.0425	1.0769		
	C10	−2.2857	0.4799	0.5784		
	C11	−2.8733	−0.0336	0.0347	1.0349	
	C12	−2.4495	0.5464	−0.0378	0.5552	
	C13	−1.0977	−0.0344	0.9997	0.0496	
	C14	−0.5283	−0.1538	0.6645	0.5408	
	C15	−1.6045	−0.0714	0.6659	−0.0020	0.4556
$T_{a\text{-max}}$ Estimation	C01	0.7418	0.9849			
	C02	8.4402	1.1748			
	C03	6.1865	0.9026			
	C04	5.8675	1.2125			
	C05	−0.0367	0.5587	0.7505		
	C06	4.3759	0.1263	1.1694		
	C07	−0.0708	1.0098	−0.0068		
	C08	8.5918	1.1432	0.0458		
	C09	−0.7751	0.4757	0.8778		
	C10	5.5651	0.3821	0.9083		
	C11	1.0850	0.5573	−0.2434	0.9824	
	C12	7.5080	0.4518	−0.1274	0.9481	
	C13	3.7089	0.6542	0.8513	−0.3246	
	C14	−1.1526	0.4704	0.3212	0.6027	
	C15	3.2723	0.5978	0.4074	−0.4465	0.7015
$T_{a\text{-mean}}$ Estimation	C01	−0.3329	0.7579			
	C02	3.0973	1.0630			
	C03	0.9103	0.8154			
	C04	1.1378	1.0888			
	C05	−0.4702	0.2122	0.9074		
	C06	−1.6236	0.1523	1.0316		
	C07	−3.2005	0.6693	0.1964		
	C08	1.3821	1.0028	0.1121		
	C09	−2.2374	0.1935	0.9691		
	C10	0.7231	0.4639	0.6828		
	C11	−2.3500	0.2016	0.0036	0.9516	
	C12	−0.0214	0.5094	0.0377	0.6325	
	C13	−0.2995	0.2344	0.8818	−0.0172	
	C14	−1.5659	0.1396	0.5060	0.5450	
	C15	−1.1587	0.1968	0.5392	−0.0702	0.4952

a_0 is the intercept of each model (combination), a_1–a_4 are parameters of LST variables with the same order as shown in Table 6.2.

Appendix B

Table A2. Parameters of LM models for T_a estimation using Dataset B.

	Combination	a_0	a_1	a_2	a_3	a_4	Elevation	Julian day
$T_{a\text{-}min}$ Estimation	C01	4.1126	0.4728				−0.0029	0.0066
	C02	0.3258	0.9505				−0.0008	0.0057
	C03	3.1331	0.6298				−0.0042	0.0046
	C04	−1.6854	0.9873				−0.0005	0.0050
	C05	0.4293	−0.0318	0.9822			−0.0007	0.0050
	C06	−4.5116	0.1075	0.9595			−0.0006	0.0053
	C07	1.7992	0.0921	0.5553			−0.0041	0.0042
	C08	−1.4452	0.9067	0.0887			−0.0009	0.0054
	C09	−2.3238	0.0098	0.9868			−0.0007	0.0049
	C10	−2.7464	0.4843	0.5678			−0.0003	0.0056
	C11	−3.0229	−0.0266	0.1405	0.8891		−0.0011	0.0045
	C12	−3.2945	0.5450	−0.0031	0.5286		−0.0003	0.0053
	C13	−0.7924	−0.0512	0.8894	0.1366		−0.0010	0.0044
	C14	−0.7538	−0.1054	0.6304	0.4971		−0.0005	0.0044
	C15	−1.8881	−0.0530	0.6303	0.0650	0.3815	−0.0007	0.0041
$T_{a\text{-}max}$ Estimation	C01	10.4393	0.7387				−0.0043	0.0007
	C02	18.4850	0.8308				−0.0045	−0.0048
	C03	15.9842	0.7267				−0.0066	−0.0048
	C04	13.5620	0.9526				−0.0032	−0.0040
	C05	10.5450	0.4115	0.5496			−0.0038	−0.0010
	C06	12.3927	0.3408	0.6526			−0.0044	−0.0055
	C07	11.3235	0.3628	0.4616			−0.0058	−0.0027
	C08	16.0125	0.5685	0.3214			−0.0052	−0.0056
	C09	6.3793	0.4536	0.6361			−0.0031	−0.0007
	C10	15.0605	0.3058	0.6539			−0.0038	−0.0045
	C11	8.8810	0.2941	0.1853	0.5654		−0.0041	−0.0032
	C12	15.3856	0.3071	0.2084	0.4250		−0.0048	−0.0060
	C13	13.7642	0.1994	0.4982	0.2098		−0.0049	−0.0043
	C14	8.8056	0.3859	0.2534	0.4067		−0.0037	−0.0008
	C15	12.8016	0.2253	0.3015	0.1265	0.3065	−0.0047	−0.0044
$T_{a\text{-}mean}$ Estimation	C01	5.4211	0.6044				−0.0030	0.0035
	C02	6.6191	0.9322				−0.0017	0.0003
	C03	7.1288	0.6996				−0.0048	−0.0001
	C04	3.4497	1.0007				−0.0011	0.0006
	C05	2.4255	0.1864	0.8229			−0.0013	0.0016
	C06	0.6160	0.2471	0.8388			−0.0016	0.0003
	C07	4.1489	0.2239	0.5289			−0.0043	0.0007
	C08	3.8781	0.7786	0.2243			−0.0020	−0.0002
	C09	−0.5674	0.2103	0.8726			−0.0011	0.0019
	C10	3.1115	0.4446	0.6126			−0.0011	0.0004
	C11	−0.1692	0.1288	0.1705	0.7716		−0.0016	0.0009
	C12	2.0954	0.4650	0.1490	0.4676		−0.0015	−0.0002
	C13	2.9232	0.0875	0.7353	0.1781		−0.0019	0.0001
	C14	0.6955	0.1371	0.4779	0.4833		−0.0011	0.0014
	C15	1.5190	0.0943	0.4956	0.1184	0.3575	−0.0016	0.0001

a_0 is the intercept of each model (combination), a_1–a_4 are parameters of LST variables with the same order as shown in Table 6.2.

References

Ackerman, S.A.; Holz, R.E.; Frey, R.; Eloranta, E.W.; Maddux, B.C.; McGill, M. (2008). Cloud detection with MODIS. Part II: Validation. Journal of Atmospheric and Oceanic Technology, 25, 1073–1086.

Appelhans, T.; Mwangomo, E.; Hardy, D.R.; Hemp, A.; Nauss, T. (2015). Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spatial Statistics, 14, 91–113.

Benali, A.; Carvalho, A.C.; Nunes, J.P.; Carvalhais, N.; Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124, 108–121.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Christiansen, B. (2005). Downward propagation and statistical forecast of the near-surface weather. Journal of Geophysical Research, 110, D14104.

Colombi, A.; De Michele, C.; Pepe, M.; Rampini, A. (2007). Estimation of daily mean air temperature from MODIS LST in Alpine areas. EARSeL eProceding, 6, 38–46.

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology, 26, 707–721.

De Wit, A.J.W.; van Diepen, C.A. (2008). Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. International Journal of Applied Earth Observation and Geoinformation, 10, 414–425.

Emamifar, S.; Rahimikhoob, A.; Noroozi, A. (2013). Daily mean temperature estimation from MODIS land surface temperature. International Journal of Climatology, 33, 3174–3181.

Fu, G.; Shen, Z.; Zhang, X.; Shi, P.; Zhang, Y.; Wu, J. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. Acta Ecologica Sinica, 31, 8–13.

Good, E. (2015). Daily minimum and maximum surface air temperatures from geostationary satellite data. Journal of Geophysical Research: Atmospheric, 120, 2306–2324.

Ho, H.C.; Knudby, A.; Sirovyak, P.; Xu, Y.; Hodul, M.; Henderson, S.B. (2014). Mapping maximum urban air temperature on hot summer days. Remote Sensing of Environment, 154, 38–45.

Huang, R.; Zhang, C.; Huang, J.; Zhu, D.; Wang, L.; Liu, J. (2015). Mapping of daily mean air temperature in agricultural regions using daytime and nighttime land surface temperatures derived from TERRA and AQUA MODIS data. Remote Sensing, 7, 8728–8756.

Intergovernmental Panel on Climate Change (IPCC). Climate Change 2007: The Physical Science Basis: Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change; Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L., Eds.; Cambridge University Press: Cambridge, UK.

Izady, A.; Davary, K.; Alizadeh, A.; Ziaei, A.N.; Akhavan, S.; Alipoor, A.; Joodavi, A.; Brusseau, M.L. (2015). Groundwater conceptualization and modeling using distributed SWAT-based recharge for the semi-arid agricultural Neishaboor plain. Iran. Hydrogeology Journal, 23, 47–68.

James, G.; Witten, D.; Hastie, T.; Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R, 1st ed.; Springer: New York, NY, USA.

Janatian, N.; Sadeghi, M.; Sanaeinejad, S.H.; Bakhshian, E.; Farid, A.; Hasheminia, S.; Ghazanfari, S. (2016). A statistical framework for estimating air temperature using MODIS land surface temperature data. International Journal of Climatology, 37, 1181–1194.

Jin, M.; Dickinson, R.E. (2010). Land surface skin temperature climatology: Benefitting from the strengths of satellite observations. Environmental Research Letters, 5, 041002.

Kuhn, M.; Johnson, K. (2013). Applied Predictive Modeling, 1st ed.; Springer: New York, NY, USA.

Kuhn, M.; Weston, S.; Keefer, C.; Coulter, N. (2013). Cubist: Rule- and Instance-Based Regression Modeling; R Package Version 0.0.13; CRAN: Wien, Austria.

Land Processes Distributed Active Archive Center. Available online: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1 (accessed on 18 October 2016).

Land Quality Assessment Site of NASA. Julian Day. Available online: http://landweb.nascom.nasa.gov/browse/calendar.html (accessed on 2 December 2016).

Liang, D.; Zuo, Y.; Huang, L.; Zhao, J.; Teng, L.; Yang, F. (2015). Evaluation of the consistency of MODIS Land Cover Product (MCD12Q1) based on Chinese 30 m GlobeLand30 datasets: A case study in Anhui Province, China. ISPRS International Journal of Geo-Information, 4, 2519–2541.

Lofgren, B.M.; Hunter, T.S.; Wilbarger, J. (2011). Effects of using air temperature as a proxy for evapotranspiration in climate change scenarios of Great Lakes basin hydrology. Journal of Great Lakes Research, 37, 744–752.

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.

Mostovoy, G.V.; King, R.L.; Reddy, K.R.; Kakani, V.G.; Filippova, M.G. (2006). Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi. GIScience & Remote Sensing, 43, 78–110.

Nieto, H.; Sandholt, I.; Aguado, I.; Chuvieco, E.; Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. Remote Sensing of Environment, 115, 107–116.

Noi, P.T.; Kappas, M.; Degener, J. (2016). Estimating daily maximum and minimum land air surface temperature using MODIS land surface temperature data and ground truth data in Northern Vietnam. Remote Sensing, 8, 1002.

Peón, J.; Recondo, C.; Calleja, J.F. (2014). Improvements in the estimation of daily minimum air temperature in peninsular Spain using MODIS land surface temperature. International Journal of Remote Sensing, 35, 5148–5166.

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, 213, 230–249.

Quinlan, J.R. (1992). Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Tasmania, 16–18 November 1992; 343–348.

Quinlan, J.R. (1993). Combining instance-based and model-based learning. In Proceedings of the Tenth International Conference on Machine Learning, Amherst, MA, USA, 27–29 June 1993; 236–243.

Quinlan, J.R. (1993). C4.5: Programs For Machine Learning; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA.

Shen, S.-H.; Leptoukh, G.-G. (2011). Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environmental Research Letters, 6, 045206.

Shi, L.; Liu, P.; Kloog, I.; Lee, M.; Kosheleva, A.; Schwartz, J. (2016). Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. Environmental Research, 146, 51–58.

Shreve, C. (2010). Working towards a community-wide understanding of satellite skin temperature observations. Environmental Research Letters, 5, 044004.

Smith, W.L.; Leslie, L.M.; Diak, G.R.; Goodman, B.M.; Velden, C.S.; Callan, G.M.; Raymond, W.; Wade, G.S. (1988). The integration of meteorological satellite imagery and numerical dynamical forecast models. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 324, 317–323.

Stahl, K.; Moore, R.D.; Floyer, J.A.; Asplin, M.G.; McKendry, I.G. (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and Forest Meteorology, 139, 224–236.

Stisen, S.; Sandholt, I.; Nørgaard, A.; Fensholt, R.; Eklundh, L. (2007). Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sensing of Environment, 110, 262–274.

Sun, Y.J.; Wang, J.F.; Zhang, R.H.; Gillies, R.R.; Xue, Y.; Bo, Y.C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. Theoretical and Applied Climatology, 80, 37–48.

The U.S. Geological Survey. MODIS LST Data. Available online: http://earthexplorer.usgs.gov (accessed on 1 October 2016).

Vancutsem, C.; Pietro, C.; Tufa, D.; Stephen, J.C. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114, 449–465.

Walton, J.T. (2008). Subpixel urban land cover estimation: Comparing cubist, random forests, and support vector regression. Photogrammetric Engineering and Remote Sensing, 74, 1213–1222.

Wan, Z. (2008). New refinements and validation of the MODIS land-surface temperature/emissivity products. Remote Sensing of Environment, 112, 59–74.

Wloczyk, C.; Borg, E.; Richter, R.; Miegel, K. (2011). Estimation of instantaneous air temperature above vegetation and soil surfaces from Landsat 7 ETM+ data in northern Germany. International Journal of Remote Sensing, 32, 9119–9136.

Williamson, S.N.; Hik, D.S.; Gamon, J.A.; Kavanaugh, J.L.; Koh, S. (2013). Evaluating cloud contamination in clear-sky MODIS Terra daytime land surface temperatures using ground-based meteorology station observations. Journal of Climate, 26, 1551–1560.

Williamson, S.N.; Hik, D.S.; Gamon, J.A.; Jarosch, A.H.; Anslow, F.S.; Clarke, G.K.C.; Rupp, T.S. (2017). Spring and summer monthly MODIS LST is inherently biased compared to air temperature in snow covered sub-Arctic mountains. Remote Sensing of Environment, 189, 14–24.

Xu, Y.; Knudby, A.; Ho, H.C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. International Journal of Remote Sensing, 35, 8108–8121.

Zeng, L.; Wardlow, B.D.; Tadesse, T.; Shan, J.; Hayes, M.J.; Li, D.; Xiang, D. (2015). Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US. Remote Sensing, 7, 951–970.

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. Journal of Geophysical Research: Atmospheric, 121, 11425–11441.

Zhang, W.; Huang, Y.; Yu, Y.Q.; Sun, W.J. (2011). Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures. International Journal of Remote Sensing, 32, 9415–9440.

Zhu, W.; Lű, A.; Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sensing of Environment, 130, 62–73.

Chapter 7

7.1. A summary of key findings

As mentioned at the beginning of this dissertation, the references are based on papers or manuscripts that have been published in, submitted to, or are under review for, ISI-indexed journals.

The following is a summary of key findings from the six papers.

▪ Over the last ten years, from 2009 to 2018, LST data from MODIS sensor is the most commonly used in comparison to LST data from ASTER, AVHRR, or Landsat. The number of publications using MODIS LST data has been on the rise between 2009 and 2018. Among various applications of MODIS LST data, the top five applications were urban heat island, air temperature estimation, soil moisture estimation, evapotranspiration, and drought monitoring/ assessment.

▪ In nearly the last two decades, several methods have been developed and applied; however, they can be grouped into three main approaches, namely energy-balance modeling, temperature-vegetation index (TVX), and statistical approaches. Among these, the statistical approaches are the most popular, comprising simple statistic (using one variable), advanced statistic (using two or more variables), and machine learning methods. Of these, the advanced statistic methods are the most popular.

▪ Reviewing publications in the literature (chapter 2) we found that for different study areas with different land surface characteristics, the performance of each individual MODIS LST is different, and in some studies LST daytime and nighttime were the best variables for Ta-max and Ta-min estimation, respectively. In others, LST nighttime was better for both Ta-max and Ta-min estimation. The best choice, which is mostly reported, is to combine both daytime and nighttime data. In addition, we discussed the solutions for problems that have been reported about how cloud cover, land cover, season, and elevation affect the performance of Ta estimation.

Therefore, in our experiments, we considered the aforementioned concerns when relating to our northwestern Vietnam study area.

▪ In northwest Vietnam, both daytime and nighttime monthly LST have a linear correlation with elevation. Between them, nighttime has a stronger correlation with elevation, and daytime has a larger variation when elevation increases.

▪ For both daytime and nighttime, the change in LST with increased elevation varied throughout the year. From the monthly averages in 2015, an altitude increase of 1000 m resulted in a decrease in LST from 6.1 °C to 3.8 °C and from 5.8 °C to 1.5 °C for daytime and nighttime, respectively.

▪ The results also showed that land use/cover types played an important role in the variation of LST due to changes in elevation. Forest and bare land had the highest variations, while water and orchards showed the lowest variations. This is consistent for both daytime and nighttime.

▪ Another interesting find in this study area is that the variation of LST was larger with changes in latitude than with changes in longitude.

▪ For both LST data quality conditions (all clear sky data and only good LST data), Terra LST has a higher correlation with Ta than Aqua;

▪ A closer overpass time with Ta-max or Ta-min occurrence does not indicate a higher correlation with Ta;

▪ On average, using only good data will produce higher accuracy Ta estimation, however, if the percentage of good data is low (i.e. less than 30%), using all clear sky data will produce higher Ta-max accuracy;

▪ To guarantee an accurate Ta estimation, at least one LST nighttime should be included in each model, as the models with only LST daytime produced very low accuracy. For Ta-max estimation specifically, the impact of different combinations and land surface characteristics was stronger than the effects of LST quality.

▪ To increase the accuracy of Ta estimation, various variables (i.e. NDVI, elevation, latitude, longitude, day length, Julian day, and view zenith angle) were included in the linear regression models. The results showed that the effects of these variables differed between Ta-max and Ta-min estimations, and the improvement of each variable is also different. For Ta-max estimation, elevation and longitude were the most important variables (excluding the four LST variables), whereas the accuracy of Ta-min estimation was not significantly improved with the inclusion of any using auxiliary variables. The accuracy of Ta-min estimation was

similar whether the simplest model (using one variable) or the most complex model (using 10 variables) was used.

When integrating all four MODIS LST and auxiliary variables (e.g. elevation and longitude), Ta-max and Ta-mean estimation was significantly improved. However, auxiliary variables and other approaches to help improve the accuracy of Ta-min estimation remains unknown. The best results for Ta-min ($R^2 = 0.88$, RMSE = 2.08) were not as high as for Ta-max ($R^2 = 0.93$, RMSE = 1.43).

The impact of different combinations is larger when using only MODIS LST data than when using MODIS LST with auxiliary data (e.g. elevation and Julian day).

LST nighttime data of both Aqua and Terra played an important role in daily Ta estimation, guaranteeing higher accuracy. Whether combining the two LST nighttime data (without auxiliary variables) or at least three LST variables, the performance of LM, CB, and RF are similar.

When auxiliary data were used alongside MODIS LST, RF and CB algorithms gave a more stable and higher accuracy Ta estimate than LM (for all combinations).

Very high accuracy of T_a estimation ($R^2 > 0.93/0.80/0.89$ and RMSE ~1.5/2.0/1.6 °C of $T_{a\text{-max}}$, $T_{a\text{-min}}$, and $T_{a\text{-mean}}$, respectively) could be achieved with a simple combination of four LST data, elevation, and Julian day data using a suitable algorithm.

7.2. Discussions about the results

Throughout this research, data quality was an issue for both the satellite and meteorological station data. Due to cloud cover, the satellite data available in northwestern Vietnam was better than in northeastern Vietnam, however, measurements at stations on the northeast side were better than measurements on the northwest side. In addition, at some meteorological stations, lack of distinct Ta-min, Ta-max, and T-mean values was one of the main issues. Therefore, the study areas (and weather station selection) in Chapter 5 and Chapter 6 were different. Along with this problem, we had a chance to test and confirm results between different areas and time spans for consistency. The results from Chapter 5 and Chapter 6 (the best results of Ta-max/Ta-min estimation) were similar with R^2 ~ 0.93/0.80 and RMSE ~ 1.5/2.0 °C , respectively.

Another important point is that the effects of thin cloudy pixels, which were not removed by algorithms, remain unknown. In this thesis, we only evaluated the thin cloudy pixel effects

for MODIS LST C6 data at individual stations and found that it had a stronger effect on Ta-min estimation than on Ta-max or Ta-mean. This could be the reason for the lower results of Ta-min estimation reported in Chapter 5 and Chapter 6. However, the decision whether to use all clear sky data or only good MODIS LST data depends on the specific study area. In practice, the trade-off between spatiotemporal continuous coverage and accuracy at individual stations should be considered.

Due to the availability of both satellite and weather station data, all experiments were implemented in non-urban areas. To gain more understanding of other Ta applications using MODIS LST data, further research that extends to the urban landscape should be conducted.

7.3. The outlook for future research

In the review paper, we found there is a growing demand in various sectors for environmental monitoring, trend and pattern analysis, and forecasting. To fulfill these demands, Ta, which affects many biological and physical processes between the land surface and near surface atmosphere, is an essential climatic and ecological driver. Therefore, the availability of spatiotemporal continuous Ta is crucial for a wide range of fields and applications, such as hydrology, climate change, vegetation monitoring, urban climate, and evapotranspiration.

It is worth to mention that, together with MOD/MYD11 data products, which were produced over nearly the last two decades, a new product, MOD/MYD21 LST and emissivity product will be available soon. It will bring more chances to MODIS LST application, because according to MODIS Land Team (https://modis-land.gsfc.nasa.gov/), the new product is generated with an adaptation of the temperature–emissivity separation (TES) method which will produce more accurate LST data over semiarid and arid areas - where MOD/MYD11 has less accuracy.

This research is one of the first remote sensing applications for Ta retrieval implemented in Vietnam. Furthermore, this research used state of the art methods to estimate Ta using MODIS LST data. We are also one of the first to use machine learning methods and different combinations of the four MODIS LST data. Therefore, some limitations were discovered and should be addressed in future research. Along with the findings of this research, several other topics of research have developed that could contribute to better results for Ta retrieval.

Our recommendations for future work:

Results of this research showed that a very high accuracy of Ta estimation could be achieved (R^2 > 0.93/0.80/0.89 and RMSE ~ 1.5/2.0/1.6 °C of Ta-max/Ta-min/Ta-mean, respectively) over the complex terrain and mountainous area of northwestern Vietnam using the MODIS LST C5 data. It is reasonable and practical to apply this study to other parts of Vietnam with different land cover types, weather conditions, and topography. However, the accuracy of Ta-min estimation was not as good as Ta-max and Ta-mean. Other approaches or additional auxiliary variables should therefore be considered and investigated to increase the accuracy of Ta-min estimation.

While cloud cover is one of the largest problems when using the optical band signal, the effects of this problem are not clearly reported in the literature. A handful of studies have compared the performance of MODIS LST data (good quality and data under all clear sky conditions) for Ta estimation, however definite conclusions still remain unknown. Due to cloud cover, there is a lack of gap-free time series for MODIS LST, affecting all of the mentioned applications. Therefore, investigating and developing a suitable and practical method for data reconstruction is necessary.

Another consideration is that most studies for Ta estimation using MODIS LST were implemented in non-urban areas to complement the sparse weather station distribution. However, in Vietnam, the distribution of weather stations in urban areas is also sparse. Spatiotemporal continuous information of Ta is important for many urban problems, such as the urban heat island effect, air pollution, and public health. Ta can change over short distances in the urban landscape, yet the spatial resolution of MODIS is only one km. Therefore, downscaling or fusing LST data into higher spatial resolutions (i.e. Landsat resolution) is an important application for future research.

CURRICULUM VITAE

PHAN THANH NOI

(Updated 01.9.2018)

Address:	Albrecht-Thaer-Weg, 8b 37075 Göttingen, Germany
Telephone:	+49-17681569399
Email:	tphan1@gwdg.de
Date of Birth:	01.07.1984
Place of Birth:	Hatinh, Vietnam

EDUCATION

11/2013 – ongoing	PhD student at Department of Cartography, GIS and Remote Sensing, University of Göttingen, Germany Research topic: **Air surface temperature estimation using MODIS land surface temperature data in northwest Vietnam** Supervisor: Prof. Dr. Martin Kappas
08/2010 – 06/2012	Master of Science, School of the Environment, Resources & Development, Asian Institute of Technology, Thailand Major: Natural Resources Management Thesis title: Detecting Land Use and Land Cover Change using Remote Sensing and GIS: Case Study of Tam Nong District, Phu Tho Province, Vietnam
09/2002 – 07/2007	Bachelor of Science, Hanoi University of Mining and Geology, Vietnam. Major: Geodesy

SCHOLARSHIPS

11/2017 – 04/2018	Family Oriented Finishing Grant from GAUSS - University of Göttingen, Germany.
11/2013 – 10/2017	Scholarship from the Vietnam government (Project 911) for PhD program at the University of Göttingen, Germany.
2010 – 2012	Scholarship from Ford Foundation for Master Degree at Asian Institute of Technology (AIT), Bangkok, Thailand.
2005	Scholarship for outstanding student from Aerial Photo - Topography Company and Military Map Department of Vietnam.
2002 - 2007	Scholarship for excellent students, Hanoi University of Mining and Geology in the academic years: 2002/2003, 2003/2004, 2004/2005, 2005/2006 and 2006/2007.

EMPLOYMENT

07/2007 – 03/2008	**Geodetic and Topographic Division, Hanoi, Vietnam**
	Geodesy Engineer, Land/GPS Surveyor.
03/2008 – Now	**Vietnam National University of Agriculture, Hanoi, Vietnam**
	Lecturer and researcher on Land Surveying, GPS, Remote Sensing and Geographical Information Systems.

PUBLICATIONS

[1]. **Noi, P.T.;** Kappas, M.; Degener, J. Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam. Remote Sens. **2016**, 8, 1002. https://doi.org/10.3390/rs8121002

[2]. **Noi, P.T.;** Degener, J.; Kappas, M. 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 Sens. **2017**, 9, 398. https://doi.org/10.3390/rs9050398

[3]. **Noi, P.T.;** Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors **2018**, 18, 18. https://doi.org/10.3390/s18010018

[4]. **Noi, P.T.;** Kappas, M.; Tran, T.P. Land Surface Temperature Variation Due to Changes in Elevation in Northwest Vietnam. Climate **2018**, 6, 28. https://doi.org/10.3390/cli6020028

[5]. **Noi, P.T.;** Kappas, M. Application of MODIS Land Surface Temperature Data: A Systematic Literature Review and Meta-analysis (accepted).

[6]. **Noi, P.T;** Kappas, M. et al. Evaluation of MODIS LST Product Collection 6 for Daily Air Surface Temperature Estimation in Northwest Vietnam (accepted).

[7]. **Noi, P.T.;** Kappas, M.; Erasmi, S. Application of MODIS Land Surface Temperature Data for Air Surface Temperature Estimation: A Review (submitted).

POSTERS

Thanh Noi Phan, Martin Kappas, and Jan Degener. Different combination of MODIS land surface temperature data for daily air surface temperature estimation in Northwest Vietnam. EGU General Assembly 2017, 23–28 April 2017, Vienna, Austria.

MEMBERSHIPS IN SCIENTIFIC SOCIETIES

Remote Sensing & Photogrammetry Society (RSPSoc)

European Geosciences Union (EGU)

International Society for Photogrammetry and Remote Sensing (ISPRS)

REVIEWER ACTIVITIES

[*Journal - # manuscripts*]

ISPRS Journal of Photogrammetry and Remote Sensing – 2

Geoscience and Remote Sensing Letters – 1

Remote Sensing (MDPI) – 3

Sustainability - 2