

**Reconstruction of multiple climate variables at high
spatiotemporal resolution based on Big Earth data platform**

Mingxi Zhang

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Life Sciences, Faculty of Science

University of Technology Sydney

Australia

November 2021

Certificate of Original Authorship

I, Mingxi Zhang declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences/Faculty of Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Signature of student: Production Note:
Signature removed prior to publication.

Date: 30 November 2021

Acknowledgements

This four-year Ph.D. journey has been an amazing experience for me that has changed my destiny. I would like to thank many people and institutions involved in the completion of this PhD study.

First and foremost, I would like to express my deepest gratitude to my research supervisors at UTS, Professor Qiang Yu, Dr. Xihua Yang and Professor Alfredo Huete, who have provided me generous guidance and support to accomplish this challenging project. Professor Qiang Yu, my principal supervisor, who is an inspiration in my life. He not only gave me an intellectually rich environment, but also gave me insightful guidance when having philosophical conversations with him. Without the help and support of Prof Qiang Yu, it would not be possible to get started my PhD study and bring this thesis to its completion. I am deeply grateful for the professional guidance of the principal research scientist Dr. Xihua Yang, who provided me immeasurable support and scientific motivation for my industry internship in NSW DPIE. Dr. Xihua Yang helped me with careful read and provided wealthy inputs to improve the theoretical contents of this thesis. I also greatly appreciate the help of Professor Alfredo Huete, who welcomed me to the group and expand my horizons on the state-of-the-art of technologies.

My special thanks go to the principal research scientist at NSW DPI, Dr. De Li Liu, who instilled in me the passion for scientific research and provided me with a great learning environment and invaluable support. I also owe my great gratitude to Dr. James Cleverly at UTS. I am so much in debt to his unconditional help. He helped me conceptualize the paper, instructed me on how to conduct scientific research and enrich it. Without his responsible step-by-step guidance, it would have been very difficult for me to make quick progress in my thesis.

I thank my colleagues from UTS, Dr. Bin Wang, Dr. Jie He, Dr. Puyu Feng, Dr. Qiaoyun Xie, Dr. Wenjie Zhang, and Dr. Cicong Gao, Dr. Song Leng, Dr. Lijie Shi, Dr. Rong Gan, Dr. Qinggaozi Zhu, Dr. Jianxiu Shen, Dr. Yuxia Liu, for your generosity and support, and for all

the wonderful memories we shared. I would like to thank Professor Jiandong Liu, Dr. Jonathan Grey, Dr. John Leys, Dr. Hongtao Xing, and Dr. Dongdong Kong for their insightful comments and support on my PhD study. I am also grateful to staff at UTS and NSW DPIE for their ongoing assistance in ensuring an enjoyable study environment and easy workflow during this wonderful PhD journey.

I would like to express my deepest gratitude to my beloved wife Dr. Hong Zhang and my parents in China. I thank for their perennial supported during my academic journey, and their encouragement has been a source of motivation for me to keep studying.

Finally, I would like to thank the China Scholarship Council and UTS for providing scholarships, and UTS and NSW DPIE for providing nice office environment to conduct this project.

Publications arising from this thesis

Journal papers directly related in this thesis:

Zhang, M., Wang, B., Liu, D. L., Liu, J., Zhang, H., Feng, P., ... & Yu, Q. (2020). Incorporating dynamic factors for improving a GIS-based solar radiation model. *Transactions in GIS*, 24(2), 423-441. (Chapter 2)

Zhang, M., Wang, B., Cleverly, J., Liu, D. L., Feng, P., Zhang, H., ... & Yu, Q. (2020). Creating new near-surface air temperature datasets to understand elevation-dependent warming in the Tibetan Plateau. *Remote Sensing*, 12(11), 1722. (Chapter 3)

Zhang, M., Yang, X., Cleverly, J., Huete, A., Zhang, H., & Yu, Q. (2021). Heat wave tracker: A multi-method, multi-source heat wave measurement toolkit based on Google Earth Engine. *Environmental Modelling & Software*, 105255. (Chapter 4)

Zhang, M., Yang, X., Leys, J., Gray, J., Zhu, G., & Yu, Q. (2020). The first combined water and wind erosion assessment for Australia 2000-2020. (Ready for submission) (Chapter 5)

Yang, X., Zhang, M., Oliveira, L., Ollivier, Q. R., Faulkner, S., & Roff, A. (2020). Rapid Assessment of Hillslope Erosion Risk after the 2019–2020 Wildfires and Storm Events in Sydney Drinking Water Catchment. *Remote Sensing*, 12(22), 3805. (Chapter 5)

Yang, X., Zhang, X., Lv, D., Yin, S., Zhang, M., Zhu, Q., ... & Liu, B. (2020). Remote sensing estimation of the soil erosion cover-management factor for China's Loess Plateau. *Land Degradation & Development*, 31(15), 1942-1955. (Chapter 5)

Research Grant directly related in this thesis:

Community vulnerability to extreme heat wave in Lake Macquarie area, Lake Macquarie Environmental Research Grants, 2019

Contents

Certificate of Original Authorship	II
Acknowledgements	III
Publications arising from this thesis.....	V
Contents	VI
List of Figures	IX
List of Tables	XII
Glossary.....	XIII
Abstract.....	XIV
Chapter 1. Introduction.....	1
1.1 Background to the question	1
1.1.1 Big EO data	1
1.1.2 Big EO data meets Climate Change	3
1.1.3 Big EO data and Cloud Computing.....	4
1.1.4 Big EO data with Machine Learning.....	5
1.2 Statement of significance and knowledge gaps	7
1.3 Research issues and objectives	9
1.4 Thesis outline.....	11
Chapter 2. Incorporating dynamic factors for improving a GIS-based solar radiation model.....	12
Abstract	12
2.1 Introduction	13
2.2 Materials and methods	16
2.2.1 Study area and observed solar radiation data.....	16
2.2.2 Schematic of the modelling	18
2.2.3 Distributed Global Solar Radiation (GSR) model for rugged terrain.....	19
2.2.3.1 Radiation on the horizontal surface.....	19
2.2.3.2 Radiation on the inclined surface	21
2.2.4 Spatial and temporal MODIS albedo gap-filling	22
2.2.5 Model evaluation.....	25
2.3 Results.....	27
2.3.1 Model validation in the Loess Plateau.....	27
2.3.2 Comparison with other SSR and GSR products	29
2.4 Discussion	33
2.5 Conclusion.....	38
Chapter 3. Creating new near-surface air temperature datasets to understand elevation-dependent warming in the Tibetan Plateau	39
Abstract	39
3.1 Introduction	40
3.2 Materials and Methods.....	42
3.2.1 Study area and all climate data	42

3.2.2 Methodology.....	44
3.2.2.1 Step 1: Hybrid model to estimate daily seamless MODIS LST and validation	45
3.2.2.2 Step 2: Remotely sensed indices, DEM derivatives and mountainous solar radiation	46
3.2.2.3 Step 3: Regression models and target-oriented validation.....	47
3.2.2.4 Step 4: Creating near-surface air temperature products and elevation-dependent warming analysis.....	49
3.3 Results.....	50
3.3.1 Evaluation of spatio-temporal composite LST	50
3.3.2 Model performance and variable importance	52
3.3.3 Spatial distribution of surface air temperature.....	55
3.3.4 Comparison with other Tibetan Plateau temperature products	59
3.3.5 Elevation-dependent warming	61
3.4 Discussion	62
3.5 Conclusions	67
Chapter 4. Heat wave tracker: a multi-method, multi-source heat wave measurement toolkit based on Google Earth Engine.....	68
Abstract.....	68
4.1 Introduction	69
4.2 Data and methods.....	72
4.2.1 Earth observation datasets	72
4.2.2 Heat wave indices	74
4.2.3 Non-stationary generalized extreme value analysis	77
4.2.4 Online heat wave measurement under a framework	78
4.3 Results.....	79
4.3.1 Heat Wave Tracker.....	79
4.3.2 How do the datasets differ in representing heat waves?.....	80
4.3.3 How do the methods differ in identifying and characterising heat waves?	83
4.3.4 How does the heat wave risk change in recent climates?	85
4.3.5 How does the heat wave risk change under future climate conditions?	87
4.4 Discussion	90
4.4.1 Model Comparison.....	90
4.4.2 Heat wave threshold.....	91
4.4.3 Future needs.....	92
4.5 Conclusion.....	94
Chapter 5. New assessment of water and wind erosion for Australia 2000-2020	96
Abstract.....	96
5.1 Introduction	97
5.2 Data and methods.....	99
5.2.1 Earth Observation and Soil Datasets.....	99
5.2.2 Estimates of Water Erosion by RUSLE.....	103

5.2.2.1 Rainfall erosivity (R) factor.....	103
5.2.2.2 Cover-management (C) factor.....	104
5.2.2.3 Slope-steepness (LS) factor	105
5.2.2.4 Soil erodibility (K) factor	105
5.2.3 Albedo-based wind erosion model.....	108
5.2.4 DustWatch PM10 measurements	110
5.3 Results.....	110
5.3.1 Estimation of sub-factors in RUSLE.....	110
5.3.2 Assessment and comparison of two wind erosion model outputs with DustWatch ..	114
5.3.3 Monthly and annually wind-water erosion maps.....	117
5.4 Discussion	124
5.4.1 Water and wind erosion explorer	124
5.4.2 Underlying drivers for water and wind erosion changes	125
5.4.3 Limitations and Model Uncertainties.....	127
5.5 Conclusion.....	128
Chapter 6. Final conclusions and future research	129
6.1 Final conclusions	129
6.2 Future research.....	131
Reference.....	132

List of Figures

Figure 1-1 The framework of the thesis.....	11
Figure 2-1 The study area showing the Loess Plateau located in north-central China including 10 radiation stations and 301 weather stations.....	17
Figure 2-2 Flowchart of steps for calculation of solar radiation in mountainous terrain.....	19
Figure 2-3 Albedo map of Loess Plateau at 1 January 2011 shown as an example of gap filling. Left panel shows missing values (white) in the northern and western regions of the plateau. Right panel shows Whittaker smoother gap-filled albedo map.	24
Figure 2-4 Variation of daily albedo for different land types. Missing values in raw albedo images were filled by spatio-temporal gap-filling method. Those gap values in the curves were fitted by the Whittaker smoother method, with $\lambda = 20$, iterative=3.	24
Figure 2-5 The percentage of albedo data during 2011 for the whole Loess Plateau (a), a representative validation area with 10 points (b), the temporal variation of daily albedo at point 7 with 10 randomly observed albedo (c) and cross validation for 100 samples during 2011 (d).	27
Figure 2-6 Comparison of annual observed and estimated (by mountain solar radiation model) monthly Global Solar Radiation (GSR) for 10 radiation sites on the Loess Plateau, China, during 2005 to 2009. Comparisons for direct radiation (DIR) and diffuse radiation (DFR) are shown only for YuZhong.	29
Figure 2-7 Spatial distributions of yearly solar radiation on the Loess Plateau in 2011 by mountain solar radiation produced by STMSR model, Surface Solar Radiation, and GLDAS.	31
Figure 2-8 Summary statistics for estimated daily solar radiation produced by the Spatio-temporal Mountain Solar Radiation (STMSR) model (a), the Surface Solar Radiation (SSR) model (b), and the Global Land Data Assimilation System (GLDAS) model (c) compared with observed data across 10 solar radiation stations in 2007-2013.	31
Figure 2-9 Spatial distributions of RMSE calculated between the daily solar radiation of the Spatio-temporal Mountain Solar Radiation (STMSR) model and Surface Solar Radiation (SSR) product at 1000 randomly selected points in 2011 over the Losses Plateau. Circle diameters correspond to the size of RMSE. RMSE units in the legend are $\text{MJ}\cdot\text{m}^{-2}$	32
Figure 2-10 The comparison of different Global Solar Radiation (GSR) products with in situ observations at YuZhong in 2009. STMSR: Spatio-temporal Mountain Solar Radiation, SA: Solar Analyst and r.sun: radiation integrated in GRASS.	33
Figure 2-11 The application interface for the mountain solar radiation model on the Google Earth Engine APP Platform.	34
Figure 2-12 Spatial distribution of annual astronomical solar radiation, direct solar radiation, diffuse solar radiation, and reflected solar radiation in 2011 over the Loess Plateau.	35
Figure 3-1 Location of Tibetan Plateau, distribution of 130 weather stations and A'rou station	43
Figure 3-2 Flowchart of steps for calculation of near-surface temperature over TP.....	45
Figure 3-3 shows the prevalence of available data in the two pairs of maps. Figure 3-3(a) shows	

the percentage of days for the given year for which LST day (i.e. 1:30 pm on Aqua (T2)) values are available at each pixel of the TP domain. Figure 3-3(b) shows the percentage of daily merged T2 for the given year for which daily merged T2 values are available. Figure 3-3(c) shows the percentage of days for the given year for which LST night (i.e. 1:30 am on Aqua (T4)) values are available at each pixel of the TP domain. Figure 3-3(d) shows the percentage of daily merged T4 for the given year for which daily merged T4 values are available. 51

Figure 3-4 LST maximum and minimum temperature validation with in-situ LST measurements in A'rou station..... 52

Figure 3-5 (a) and (b) show the R^2 and RMSE for maximum (Tmax), minimum (Tmin) and mean (Tmean) air temperatures using rf, cubist and xgbDART methods based on LLTO-CV, LTO-CV, LLO-CV. The boundaries of box mark the 25th and 75th percentiles; the horizontal black lines within the box indicate the median; the upper and lower whiskers mark the 90th and 10th percentiles. 54

Figure 3-6 Tmean, Tmin, and Tmax temperature residuals showed varying temporal sensitivity to physiographic drivers. Each variable was scaled to a total of 100%..... 55

Figure 3-7 Monthly Tmean based on RF model in 2003-2013 56

Figure 3-8 Spatial distribution of the seasonally averaged daily mean air temperatures for 2003-2013 in spring (a), summer (b), autumn (c), winter (d) and the full year (e)..... 57

Figure 3-9 Monthly Maximum temperature derived from RF between 2003 and 2013..... 58

Figure 3-10 Monthly Minimum temperature derived from RF between 2003 and 2013 58

Figure 3-11 The comparison of monthly Tmean in May (a) and December (b) derived from Random Forest, TerraClimate and CMFD with the observed mean temperature of 1980-2010 from in-situ measurements. 60

Figure 3-12 (a) Spatial average maps and (b) histograms of Tmean in 2003-2013 at Central TP for May and December..... 61

Figure 3-13 Tmean variation at 3 elevation zones from 01/2003 to 12/2013. The number of pixels within 1000 m elevation interval were extracted and each temperature change was computed from the mean of the pixels. 62

Figure 4-1 An example schematic of indices used to define heat wave-EHF. Short duration heat spikes less than three days in a row are not heat waves. In this figure the green line is the threshold and black line is the EHF. There are four discrete events including red and pink heat spikes (HWN); the highest red heat spikes is the heat wave amplitude (HWA); the length of the longest event is also the red heat spikes (HWD); the average heat wave magnitude is the average magnitude across four events (HWM); and the sum of four heat wave events that above the threshold is HWF. The five indices in the figure are calculated for each season and annually. 76

Figure 4-2 The online implementation of heat wave tracker toolkit based on Google Earth Engine, using a framework enables climate data integration for heat wave measurement at a continental scale..... 79

Figure 4-3 Examples of heat wave aspects derived from three different climate datasets in 2018 82

Figure 4-4 Examples of heat wave aspects of ERA5 from three different methods in 2018 ..	84
Figure 4-5 Distinct heat wave events derived from time series with EHF, TN90 and TX90 at the same point of southeastern Australia.	85
Figure 4-6 (a) Effective return level under the non-stationary assumption with mean HWA value from the continental Australia. (b) The probability density functions (PDF) of HWA under 1920-2019 and 1980-2019. (c) Return period of HWA over Australia. The distributions are fit with non-stationary GEV for the climates of 1920-2019 (red), 1980-2019 (blue).....	87
Figure 4-7 Near-future (2020–2039) and Far-future (2069-2099) projected climatology for heat wave amplitude obtained from the CMIP5 multi-GCM ensemble	89
Figure 4-8 Near-future (2020–2039) and Far-future (2069-2099) projected climatology for heat wave duration obtained from the CMIP5 multi-GCM ensemble.....	89
Figure 4-9 Heat wave metrics comparison between HWT and GHWR software tools.....	91
Figure 5-1 Maps of the RUSLE factors: Rainfall erosivity factor, Soil erodibility factor, Cover management factor, Slope length and steepness factor	112
Figure 5-2 Monthly C-factor based on MODIS fractional vegetation cover in 2001-2020..	113
Figure 5-3 Comparison between the R factor values derived from SILO and GPM and TRMM for 12 months along the Great Dividing Range in south-eastern Australia	114
Figure 5-4 Monthly wind erosion values from the albedo-based model, the RWEQ model, and the observations at Tibooburra site are compared for the period 2009 to 2019 using the GLDAS dataset as model input	116
Figure 5-5 Monthly wind erosion values from the albedo-based model, the RWEQ model, and the observations at Tibooburra site are compared for the period 2009 to 2019 using the ERA5 dataset as model input	117
Figure 5-6 Monthly water erosion based on SILO in 2001-2020	118
Figure 5-7 Monthly water erosion by State in 2001-2020.....	119
Figure 5-8 Monthly wind erosion based on Albedo-based model in 2001-2020	120
Figure 5-9 Monthly wind erosion by State based on Albedo-based model in 2001-2020....	121
Figure 5-10 Monthly wind erosion based on RWEQ model in 2001-2020	122
Figure 5-11 Monthly wind erosion by State based on RWEQ in 2001-2020.....	123
Figure 5-12 Annual water and wind erosion and uncertainty based on RUSLE and RWEQ in 2001-2020.....	124

List of Tables

Table 2-1 A comparison between the proposed tool and related tools	15
Table 2-2 Data sources for calculating and comparing solar radiation	18
Table 3-1 Overview of datasets across the TP	44
Table 4-1 Datasets used in this study	74
Table 4-2 Structural similarity index between different heat wave characteristics from three climate datasets.	82
Table 5-1 Datasets used in this study	101

Glossary

EO	Earth Observation
PB	Petabytes
GEE	Google Earth Engine
ODC	Open Data Cube
SH	Sentinel Hub
API	application programming interface
DEM	Digital Elevation Model
HPC	High-Performance Computing
CNN	convolutional neural network
DL	Deep Learning
STMSR	Spatial and Temporal Mountain Solar Radiation Modelling
SSR	Surface Solar Radiation
GSR	Global Solar Radiation
TP	Tibetan Plateau
LST	Land Surface Temperature
GWTR	geographically and temporally weighted regression
RF	Random Forest
XGBoost	eXtreme Gradient Boosting
LLTO	Leave-Location-Time-Out
LLO	Leave-Location-Out
LTO	Leave-Time-Out
FFS	Forward Feature Selection
EHF	Excess heat factor
HWN	Heat Wave Number
HWD	Heat Wave Duration
HWF	Heat Wave Amplitude
HWM	Heat Wave Magnitude
HWT	Heat Wave Tracker
GEV	Generalized Extreme Value
NEVA	Non-stationary Extreme Value Analysis
CMIP5	Coupled Model Intercomparison Project 5
WEPP	Water Erosion Prediction Project
SWEEP	Single-Event Wind Erosion Evaluation Program
RUSLE	Revised Universal Soil Loss Equation
RWEQ	Revised Wind Erosion Equation
GPM	Global Precipitation Measurement
GLDAS	Global Land Data Assimilation System
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis 5
SLGA	Soil and Landscape Grid of Australia
FVC	Fractional Vegetation Cover

Abstract

Reconstruction of climate variables with high spatio-temporal resolution is important when the meteorological observations required for environmental monitoring and modelling do not cover the study area. In addition, climate model reanalysis datasets suffer from coarse spatio-temporal resolutions, which fails to capture the complex variability of climate at fine scales. This thesis mainly reconstructed four climate datasets including: mountainous solar radiation, near-surface air temperature datasets over rugged terrain, five distinct metrics of long-term heat wave datasets, an updated database of water and wind erosion. For further use in practice, these datasets are freely accessible and online web application has been developed for academic research on climate change under accelerated global warming. The main findings of this thesis are: (1) A GIS-based solar radiation model that incorporates albedo, shading by surrounding terrain, and variations in cloudiness was developed to address the spatial variability of these factors in mountainous terrain. (2) The Tibetan Plateau has been undergoing accelerated warming over recent decades, and is considered an indicator for broader global warming phenomena. However, our understanding of warming rates with elevation in complex mountain regions is incomplete. The most serious concern is the lack of high-quality near-surface air temperature (T_{air}) datasets in these areas. To address this knowledge gap, we create new near-surface air temperature datasets to understand elevation-dependent warming in the Tibetan Plateau. (3) Under ongoing global warming due to climate change, heat waves in Australia are expected to become more frequent and severe. A Google Earth Engine-based toolkit named heat wave tracker (HWT) is developed, which can be used for dynamic visualization, extraction, and processing of complex heat wave events. The datasets, toolkit, and findings we developed contribute to global studies on heat waves under accelerated global warming. (4) Soil erosion caused by water and wind is a complicated natural process that has been accelerated by human activity. This erosion has resulted in increasing areas of land degradation which threaten the productive potential of landscapes. Consistent and continuous erosion monitoring will help identify the trends, magnitude, and location of soil erosion. We apply the water-wind erosion model to produce monthly and annual water, and wind erosion estimation at high spatial resolution (up to 90 m, 500 m) for Australia from 2000 to 2020.

Keywords: Big data; solar radiation; near-surface air temperature; heat wave; water and wind erosion; climate change; Cloud computing; China, Australia