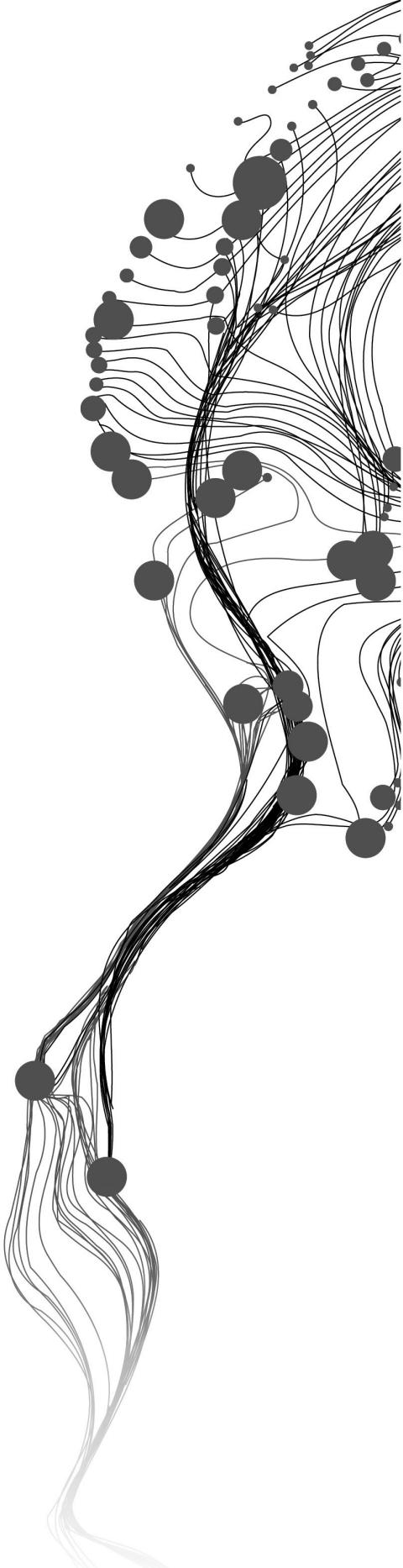


# **DOWNSCALING LAND SURFACE TEMPERATURE USING SAR IMAGES: A MACHINE LEARNING FRAMEWORK**

NISHIT PATEL  
August, 2023

SUPERVISORS:  
dr. ing. H. Aghababaei  
dr. F.B. Osei





# **DOWNSCALING LAND SURFACE TEMPERATURE USING SAR IMAGES: A MACHINE LEARNING FRAMEWORK**

**NISHIT PATEL**

Enschede, The Netherlands, August, 2023

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.  
Specialization: M-GEO

**SUPERVISORS:**

dr. ing. H. Aghababaei  
dr. F.B. Osei

**THESIS ASSESSMENT BOARD:**

Prof. dr. ir. A. Stein (chair)  
Prof. Giampaolo Ferraioli (External Examiner - Università degli Studi di Napoli Parthenope)

#### Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

## ABSTRACT

Land Surface Temperature (LST) is significant for climatological and environmental studies. LST products acquired from satellites, however, suffer from the tradeoff between spatial and temporal resolution. Spatial downscaling has emerged as a well explored field aiming to overcome limitations arising from this tradeoff. Previous research on regression based LST downscaling models focused on utilizing predictors derived from optical imagery for constructing such spatial downscaling models. Weather-dependent nature of optical imagery data, however, can influence downscaling models and render them ineffective during bad weather conditions like high cloud cover. To cope with this issue, in this research, we involve predictors derived from the weather-independent Sentinel-1 Synthetic Aperture Radar (SAR) imagery to downscale Landsat-8 LST and MODIS LST products. In this context, we propose to use machine learning techniques, namely Random Forest (RF) and Convolutional Neural Networks (CNN) as base regression algorithms to develop radar-based LST downscaling models. To demonstrate the applicability and performance of the proposed method, extensive experimental analyses were conducted over Zuid-Holland in the Netherlands. From the experiments, we found that the results obtained with radar predictors were comparable both quantitatively and qualitatively to those achieved using optical predictors. This confirms that the proposed method indeed paves a new way for mapping LST using SAR images.

In addition to the aforementioned contributions, we present innovative methodologies for engineering features from SAR and land cover datasets, aiming to enhance the performance of base models. Through conducting various experiments with distinct feature sets, we found out that these engineered features effectively address limitations stemming from the inherent ‘noisy’ nature of SAR images. Consequently, they contribute to the improvement of downscaling models’ performance.

Moreover, we propose a novel CNN-based architecture for downscaling LST. Through quantitative comparison with the widely adopted RF model, we have determined that the CNN-based downscaling architecture outperforms RF. However, upon conducting a qualitative assessment of the downscaled images generated by the CNN-based approach, we identified certain performance shortcomings associated with the architecture’s design. In light of these findings, we recommend specific modifications to the proposed architecture with the aim of addressing the aforementioned shortcomings. By implementing these modifications, we anticipate further increase in performance and capabilities of the proposed CNN-based downscaling methodology.

### Keywords

*Land Surface Temperature (LST), Synthetic Aperture Radar (SAR), Downscaling, Regression, Random Forest, Convolutional Neural Networks*

## ACKNOWLEDGEMENTS

I would like to express my profound gratitude and appreciation to all individuals who have contributed to the completion of this thesis. Their support and involvement have been instrumental in shaping this work, and I am sincerely grateful for their contributions.

First and foremost, I extend my deepest gratitude to my first supervisor, Dr. Ing. Hossein Aghababaei. Without his invaluable guidance, this thesis would not have been possible. His unwavering support and positive demeanor served as a catalyst for my motivation. Our regular meetings and discussions were a constant source of inspiration, propelling me to present part of this thesis at a conference and submit a research paper for publication. These achievements exceeded my initial expectations, and I fully credit his mentorship for this success. I am truly grateful for the comfort and openness of our discussions, enabling productive exchange of ideas. I could not have asked for a better first supervisor.

I would also like to express my gratitude to my second supervisor, Dr. Frank Osei. His invaluable feedback emphasized the significance of meticulous attention to detail in writing, presenting, and overall implementation. His guidance has helped me strengthen the foundation of this work, ensuring its quality and coherence.

I extend my heartfelt appreciation to my dear friends, Jorges, Yajan, Onur, and Shreya, for their unwavering kindness, support, and availability. Their companionship and the meaningful conversations we shared have been a source of joy and fulfillment. I treasure these connections and look forward to fostering further dialogue with them in the future.

I cannot express enough gratitude to my parents for their unyielding support and provision throughout this academic journey. Their unwavering belief in me has been a constant source of motivation and strength.

Lastly, I would like to thank Dr. Ir. Rolf de By for developing this L<sup>A</sup>T<sub>E</sub>X template for ITC thesis. This template enabled me to overcome initial worries on working with L<sup>A</sup>T<sub>E</sub>X and provided me with a solid foundation to confidently utilize L<sup>A</sup>T<sub>E</sub>X not only for this thesis but also in future endeavors.

To all individuals mentioned above, as well as anyone else who has contributed directly or indirectly to this thesis, I extend my sincerest thanks for your support. I hope that the content presented herein does justice to your efforts.

Thank you, and I wish you a pleasant reading experience.

## TABLE OF CONTENTS

---

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Introduction to Downscaling . . . . .	3
1.3 Literature Review . . . . .	4
1.3.1 Spatio-temporal Fusion Models . . . . .	4
1.3.2 Physical Models . . . . .	5
1.3.3 Regression Models . . . . .	6
1.4 Research Gap . . . . .	8
1.5 Research Objectives and Questions . . . . .	8
1.5.1 Main Objective . . . . .	11
1.5.2 Sub-Objectives . . . . .	11
1.6 Contributions . . . . .	11
1.6.1 Highlights . . . . .	12
1.7 Thesis Structure . . . . .	12
<b>2 Study Area, Datasets, and Tools</b>	<b>13</b>
2.1 Study Area . . . . .	13
2.2 Datasets . . . . .	15
2.2.1 Dataset Description and Use Case . . . . .	15
2.2.2 Google Earth Engine . . . . .	17
2.2.3 Pre-processing . . . . .	17

2.2.4	Selection of Dates of Interest . . . . .	17
2.3	Tools . . . . .	18
<b>3</b>	<b>Methodology</b>	<b>19</b>
3.1	Basic Spatial Downscaling Algorithm . . . . .	19
3.2	Random Forest . . . . .	22
3.2.1	Radar-based RF downscaling model . . . . .	22
3.2.2	Optical-based RF downscaling model . . . . .	24
3.2.3	Integration of Land Cover information . . . . .	25
3.3	Convolutional Neural Networks . . . . .	25
3.4	MODIS LST as a target . . . . .	27
3.5	Evaluation Framework . . . . .	28
<b>4</b>	<b>Results</b>	<b>29</b>
4.1	Random Forest . . . . .	29
4.1.1	Radar-based downscaling . . . . .	29
4.1.2	Optical-based downscaling . . . . .	30
4.1.3	Integration of Land Cover information . . . . .	35
4.2	Convolutional Neural Networks . . . . .	38
4.2.1	Radar-based Downscaling . . . . .	38
4.2.2	Optical-based Downscaling . . . . .	41
4.3	Application to MODIS LST . . . . .	41
<b>5</b>	<b>Discussion</b>	<b>45</b>
5.1	Efficacy of radar data in LST downscaling . . . . .	45
5.2	Efficacy of engineered radar features in LST downscaling . . . . .	46
5.3	Integration of Land Cover information . . . . .	49
5.4	CNN: shortcomings and potential architecture modifications . . . . .	51
5.5	Application to MODIS LST: shortcomings and a potential correction procedure	52

5.6	Challenge in modelling time . . . . .	53
<b>6</b>	<b>Conclusion</b>	<b>55</b>
6.1	Research Questions - Answered . . . . .	55
6.2	Future Recommendations . . . . .	56
6.3	Conclusion . . . . .	57
<b>A</b>	<b>Other figures</b>	<b>59</b>
	<b>List of References</b>	<b>63</b>

## LIST OF FIGURES

---

1.1	A simple representation of the ‘Downscaling’ process. Here, $I_c$ refers to image at coarse resolution, whereas, $I_f$ refers to image at fine resolution. ‘ $\times 4$ ’ denotes the downscaling factor (in this case: ‘4 times’) . . . . .	4
1.2	Chronological progression of the regression-based LST downscaling studies discussed in the Literature Review section . . . . .	9
2.1	Map showing the study area: Zuid-Holland province (left) and various municipalities in Zuid-Holland province (right) . . . . .	14
3.1	Methodology Flowchart . . . . .	20
3.2	Representation of inclusion of neighboring values as features for the downscaling models. Here, ‘Connection A’ in the image illustrates conventional pixel-to-pixel mapping between coarse resolution predictor image $P_c$ and coarse resolution target image $T_c$ , whereas, ‘Connection B’ illustrates pixel-to-( $5 \times 5$ )patch mapping between coarse resolution predictor image $P_c$ and coarse resolution target image $T_c$ . . .	23
3.3	Representation of integration of land cover proportion as features to the downscaling models. Here, ‘Connection A’ highlighted in (a) illustrates a pixel-to-patch mapping where $P_o$ represents the original land cover image collected at 10 m spatial resolution and $T_c$ refers to the coarse resolution target image. ‘Connection B’ in (b) demonstrates the same concept but at a fine resolution to facilitate inference. Here, $T_f$ refers to the target image to be estimated by the model during inference . . .	26
3.4	proposed CNN-based downscaling architecture . . . . .	27
4.1	Results of the ‘Case 3’ radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downsampled LST images (100 m) to the Landsat-8 validation LST images (100 m) . . . . .	32
4.2	Results of the optical-based RF downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downsampled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b) . . . . .	34

4.3 Results displaying the efficacy of different land cover integration approaches. Here, (a) refers to the aggregated Landsat-8 LST (1000 m), (b) refers to the RF model estimation (1000 m) with the conventional approach, and (c) refers to the RF model estimation (1000 m) with the proposed approach. (d) and (e) displays the histograms of (b) and (c) in comparison to (a), respectively. Here, 'Conventional Approach' refers to utilizing aggregated or upscaled version of land cover image as a predictor, whereas, 'Proposed Approach' refers to utilizing proportion of land cover classes as predictors . . . . .	36
4.4 Results achieved by integrating land cover information to 'Case 3' of the radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m) . . . . .	37
4.5 Results achieved by integrating land cover information to the optical-based RF downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b) . . . . .	39
4.6 Results of the radar-based CNN downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m) . . . . .	40
4.7 Results of the optical-based CNN downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b) . . . . .	42
4.8 Results of the downscaling experiments conducted using MODIS LST as the target variable. Here, (a) refers to the Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m) using (5 × 5) VV, (5 × 5) VH, homogeneity index, and land cover as predictors, and (c) refers to the achieved downscaled LST (100 m) using six optical bands as predictors. (d) and (e) displays the histograms of (b) and (c) in comparison to (a), respectively, whereas, (f) and (g) displays the scatterplot comparison between the downscaled LST (100 m) and the Landsat-8 validation LST (100 m) . . . . .	43

5.1 Comparison between the spatial patterns exhibited by different datasets. Here, the red box in (a) shows the location of the zoomed-in images. (b) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020, (c) refers to the VV band image of Sentinel-1 GRD dataset (100 m) acquired on 25/03/2020, and (d) refers to the optical true color composite of Sentinel-2 MSI dataset (100 m) acquired on 26/03/2020. As can be inferred through the images, in terms of spatial patterns, (d) shows a better agreement with (b) as compared to (c). Moreover, (c) demonstrates a higher presence of pixel value variations over local regions, such as homogeneous green areas, in contrast to (d) . . . . .	47
5.2 Comparison between the downscaled LST images (100 m) obtained by different feature inclusion cases. Here, the red box in (a) shows the location of the zoomed-in images. (b) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020, (c) refers to the downscaled LST (100 m) obtained by utilizing VV and VH bands as predictors, (d) refers to the downscaled LST (100 m) obtained by utilizing $(5 \times 5)$ VV and $(5 \times 5)$ VH patches as predictors, and (e) refers to the downscaled LST (100 m) obtained by utilizing $(5 \times 5)$ VV patch, $(5 \times 5)$ VH patch, and homogeneity index as predictors . . . . .	48
5.3 Impact of incorporating land cover information into the downscaling framework. Here, the red box in (a) shows the location of the zoomed-in images. (b) shows the tree cover class from the 'ESA WorldCover' dataset (green color) and (c) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020. (d) and (f) refer to the achieved downscaled LST (100 m) by utilizing $(5 \times 5)$ VV patch, $(5 \times 5)$ VH patch, and homogeneity index as predictors, whereas (e) and (g) refer to the achieved downscaled LST (100 m) by utilizing $(5 \times 5)$ VV patch, $(5 \times 5)$ VH patch, homogeneity index, and land cover as features. Here, (d) and (e) are achieved downscaled LST images without residual correction, whereas, (f) and (g) with residual correction . . . . .	50
5.4 Comparison between (a) MODIS LST (1000 m) and (b) aggregated Landsat-8 LST (1000 m). The correlation coefficient $r$ between (a) and (b) is 0.56 . . . . .	52
A.1 Results of the 'Case 1' radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m) . . . . .	60
A.2 Results of the 'Case 2' radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m) . . . . .	61

## LIST OF TABLES

---

1.1	Spatial and Temporal characteristics of some widely-used TIR-based satellites . . . . .	2
2.1	Summary statistics for Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020 . . . . .	14
2.2	Overview of the datasets used in this research (Part 1) . . . . .	16
2.3	Overview of the datasets used in this research (Part 2) . . . . .	16
4.1	Evaluation metrics for ‘Case 1’, ‘Case 2’, and ‘Case 3’ of the radar-based RF downscaling experiments performed on all selected dates of interest . . . . .	30
4.2	<i>RMSE</i> of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 25/03/2020) for all the land cover classes in the study area . . . . .	30
4.3	<i>RMSE</i> of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 10/04/2020) for all the land cover classes in the study area . . . . .	31
4.4	<i>RMSE</i> of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 28/05/2020) for all the land cover classes in the study area . . . . .	31
4.5	Evaluation metrics for the optical-based RF downscaling experiment. Here, six bands from the optical dataset were used as predictors . . . . .	33
4.6	<i>RMSE</i> of the optical-based RF downscaling experiment for each land cover class in the study area. Here, six bands from the optical dataset were used as predictors . . . . .	33
4.7	Evaluation metrics achieved by comparing RF model estimate with the original target LST image at 1000 m. Here, ‘Conventional Approach’ refers to utilizing aggregated or upscaled version of land cover image as a predictor, whereas, ‘Proposed Approach’ refers to utilizing proportion of land cover classes as predictors . . . . .	35
4.8	Evaluation metrics achieved by integrating land cover information to ‘Case 3’ of the radar-based RF downscaling experiment . . . . .	35
4.9	<i>RMSE</i> achieved by integrating land cover information to ‘Case 3’ of the radar-based RF downscaling experiment for all the land cover classes in the study area . . . . .	36
4.10	Evaluation metrics achieved by integrating land cover information to the optical-based RF downscaling experiment . . . . .	38
4.11	<i>RMSE</i> achieved by integrating land cover information to the optical-based RF downscaling experiment for all the land cover classes in the study area . . . . .	38

4.12	Evaluation metrics for the radar-based CNN downscaling experiment. Here, VV and VH images from the radar dataset were utilized as predictors for the CNN downscaling model . . . . .	41
4.13	Evaluation metrics for the optical-based CNN downscaling experiment. Here, six bands from the optical dataset were utilized as predictors for the CNN downscaling model . . . . .	41
4.14	Evaluation metrics for the downscaling experiments conducted with MODIS LST as the target variable . . . . .	44
5.1	Quantitative metrics without and with residual correction process for radar-based RF downscaling experiments (date of interest: 25/03/2020) . . . . .	49
5.2	<i>RMSE</i> values of radar (VV and VH) and optical (six bands) based RF and CNN downscaling experiments before and after residual correction . . . . .	51
5.3	Evaluation metrics for the RF based downscaling experiment conducted with MODIS LST as the target variable before and after residual correction . . . . .	53

## List of Abbreviations

- ANN** Artificial Neural Network
- API** Application Programming Interface
- BSI** Bare Soil Index
- CNN** Convolutional Neural Networks
- DB** Dry Biomass
- DEM** Digital Elevation Model
- dpRVI** dual-pol Radar Vegetation Index
- EPSG** European Petroleum Survey Group
- ESA** European Space Agency
- GEE** Google Earth Engine
- GNNWR** Geographically Neural Network Weighted Regression
- GRD** Ground Range Detected
- GWAR** Geographically Weighted AutoRegressive
- GWR** Geographically Weighted Regression
- HUTS** High-Resolution Thermal Sharpener
- LST** Land Surface Temperature
- LSTM** Long Short Term Memory
- MFGWML** Multi-Factor Geographically Weighted Machine Learning
- MLR** Multiple Linear Regression
- MODIS** Moderate Resolution Imaging Spectroradiometer
- MSE** Mean Squared Error
- MSI** Multi Spectral Instrument
- NDBI** Normalized Difference Building Index
- NDVI** Normalized Difference Vegetation Index

**NIR** Near Infrared

**PAI** Plant Area Index

**PMW** Passive Microwave

**ReLU** Rectified Linear Unit

**RF** Random Forest

**RMSE** Root Mean Square Error

**RNNs** Recurrent Neural Networks

**RFI** Radar Vegetation Index

**SAR** Synthetic Aperture Radar

**SD** Spatial Downscaling

**SLC** Single Look Complex

**SLSTR** Sea and Land Surface Temperature Radiometer

**SVM** Support Vector Machines

**SWIR** Short-Wave Infrared

**TIR** Thermal Infrared

**UAVs** Unmanned Aerial Vehicles

**VIS** Visible

**VWC** Vegetation Water Content

## Chapter 1

# Introduction

This chapter is divided into seven main sections: ‘Background’ (Section 1.1), ‘Introduction to Downscaling’ (Section 1.2), ‘Literature Review’ (Section 1.3), ‘Research Gap’ (Section 1.4), ‘Research Objectives and Questions’ (Section 1.5), ‘Contributions’ (Section 1.6), and ‘Thesis Structure’ (Section 1.7). The chapter commences with Section 1.1, emphasizing the significance of high spatio-temporal Land Surface Temperature (LST) products and the constraints encountered in their attainment. This sets the foundation for Section 1.2, where the process of ‘Spatial Downscaling’ is introduced as a solution to address the constraints in achieving high spatio-temporal LST products. Section 1.3 delves into noteworthy studies conducted in the field of LST downscaling, which subsequently aids in identifying the research gap discussed in Section 1.4. Drawing inspiration from Section 1.4, Section 1.5 establishes the groundwork for formulating the research objectives and the corresponding research questions. Lastly, the chapter concludes by highlighting the contributions made by this research in Section 1.6 and providing a description of the thesis structure to the reader in Section 1.7.

### 1.1 BACKGROUND

LST holds significant importance in comprehending and simulating the energy distribution on the Earth’s surface. It serves as a valuable variable for predicting the effects of climate change and land use-land cover alterations on energy budgets at both global and local scales. The analysis of LST provides crucial insights into the distribution and transfer of thermal energy within the Earth’s system. One noteworthy application of LST data lies in the monitoring and analysis of urban heat islands (Zakšek & Oštir, 2012). Urban areas exhibit higher temperatures compared to their surrounding rural regions, primarily due to human activities and the built environment. Through the examination of LST patterns, researchers can evaluate the extent and intensity of urban heat islands. This assessment plays a pivotal role in urban planning, the design of heat mitigation strategies, and the enhancement of overall comfort and well-being for urban residents. Furthermore, LST data also proves to be beneficial in studying soil moisture (Bai et al., 2019; Zhang et al., 2014). The correlation between LST and soil moisture content aids researchers in understanding water availability in different areas. By estimating soil moisture through LST data, researchers can monitor agricultural productivity, assess drought conditions, and make informed decisions regarding irrigation practices, crop management, and water resource allocation. Drought monitoring (Wan et al., 2004) represents another critical application of LST information. LST data provides insights into the intensity and spatial extent of drought events, facilitating timely and targeted mitigation measures. Through the analysis of LST patterns, researchers and policymakers can assess the severity of drought conditions, anticipate potential impacts on agriculture and

**Table 1.1** Spatial and Temporal characteristics of some widely-used TIR-based satellites

Satellite	Spatial Resolution (Bands)	Temporal Resolution
MODIS	500m (VIS, NIR, SWIR) 1000m (TIR)	1 day
Sentinel-3 SLSTR	500m (solar reflectance bands) 1000m (TIR)	1 day
Landsat-8	30m (VIS, NIR, SWIR) 100m (TIR)	16 days

ecosystems, and implement appropriate measures to manage water resources and minimize socio-economic consequences. LST data also enables the study of vegetation dynamics as highlighted by Julien and Sobrino (2009). Alterations in LST patterns indicate variations in plant growth, phenology, and overall vegetation health. By correlating LST with vegetation indices derived from satellite imagery, researchers can monitor vegetation productivity, assess the impact of climate change on ecosystems, and enhance our understanding of the Earth's carbon cycle. In addition to these applications, LST data is instrumental in estimating heat loss from individual buildings. By analyzing temperature differences between building surfaces and the surrounding environment, researchers can evaluate energy efficiency, identify areas of heat loss, and optimize building design and insulation strategies to reduce energy consumption.

While conventional ground weather stations have traditionally been used to collect LST data, their spatial distribution often exhibits sparsity, limiting their effectiveness in large-scale LST mapping (Z. L. Li et al., 2013). To overcome this limitation, satellite-derived LST data has gained prominence. Satellites equipped with thermal sensors have the capability to capture LST measurements over vast areas, providing detailed information at both spatial and temporal scales. This satellite-based remote sensing approach facilitates comprehensive monitoring, modeling, and analysis of LST, thereby enabling a wide range of applications across various scientific disciplines and environmental management practices. This space-based LST information can be retrieved using both Thermal Infrared (TIR) and Passive Microwave (PMW) satellites (Hu et al., 2023); however, the majority of space-based LST information is derived via TIR-based satellites.

Within the domain of space-based data, the Moderate Resolution Imaging Spectroradiometer (MODIS) has emerged as the most extensively investigated source for deriving LST products. MODIS provides daily satellite images with a spatial resolution of 500 m for Visible (VIS), Short-Wave Infrared (SWIR), and Near Infrared (NIR) bands, and 1000 m for the TIR band. Another viable option for computing LST products, offering similar spatial and temporal characteristics to MODIS, is the Sentinel-3 Sea and Land Surface Temperature Radiometer (SLSTR), developed by the European Space Agency (ESA). In the context of finer spatial resolution, Landsat-8 offers an appealing alternative, presenting a TIR channel with a spatial resolution of 100 m for computing LST information. However, it is worth noting that Landsat-8 exhibits a temporal resolution of 16 days, resulting in less frequent data acquisitions compared to MODIS and Sentinel-3 SLSTR. Table 1.1 provides information on the spatial and temporal characteristics of these satellite products. As seen from Table 1.1, currently available, and operational remote sensing sensors produce TIR products (e.g., LST) at a much coarser resolution than VIS, NIR, and SWIR products (e.g., spectral indices) from the same satellite. The main reason of this disparity in the band resolutions is that the TIR bands capture radiation at longer wavelengths (Guo et al., 2022). Also, as the emitted wave from the ground surface (TIR) carries less energy compared to the reflected wave (VIS) from the

VISible region, the spatial resolution of TIR channels is often coarse in order to get enough level of emitted energy at each pixel.

Furthermore, retrieving LST information through remote sensing techniques often encounters a fundamental trade-off between spatial and temporal resolution. The acquisition of high spatial resolution LST products from space-based platforms generally comes at the expense of reduced temporal resolution, and vice versa. Consequently, irrespective of the chosen LST product, users are confronted with uncertainties arising from either the spatial or temporal domain. To alleviate this conflict, aerial methods, such as airborne or Unmanned Aerial Vehicles (UAVs), can be employed to collect LST data as needed. Nevertheless, these approaches tend to be costly and often lack practicality for large-scale applications. Alternatively, image processing techniques offer optimal solutions to tackle the inherent trade-off between spatial and temporal resolutions associated with space-based LST products. A particular class of image processing techniques, known as Spatial Downscaling (SD) methods, have demonstrated effectiveness in overcoming these constraints. SD methods employ statistical and mathematical algorithms to downscale LST data from coarser spatial resolutions to finer resolutions, thereby enhancing the level of detail and precision. The application of SD methods enables a more comprehensive understanding of LST patterns and processes at finer scales, thereby facilitating enhanced analysis and decision-making across various domains, including urban planning, agriculture, hydrology, and climate studies. By effectively addressing the trade-off between spatial and temporal resolutions inherent in space-based LST products, SD methods serve as valuable tools to bridge the gap between detailed spatial information and frequent temporal coverage. Ultimately, the integration of these techniques serves as a catalyst for the advancement of our understanding and utilization of LST data.

## 1.2 INTRODUCTION TO DOWNSCALING

Spatial downscaling can be defined as the process of translating spatial information from coarse to fine spatial resolutions. Figure 1.1 provides a simple intuition of the spatial downscaling process. By utilizing spatial downscaling techniques, users can overcome the inherent trade-offs between spatial and temporal resolutions, thereby attaining high spatio-temporal LST information. This capability holds particular significance in urban studies, where the demand for precise and detailed LST products is of utmost importance. Urban areas are characterized by their intricate and diverse land surface properties, with built-up regions exhibiting high heterogeneity in terms of land cover types, surface materials, and anthropogenic activities. Consequently, these variations often correspond to substantial variability in LST (Z. L. Li et al., 2013). To effectively comprehend and analyze the thermal behavior of urban environments, acquiring LST data with high spatio-temporal resolution becomes imperative.

High spatio-temporal resolution LST data facilitates the observation and monitoring of LST changes across distinct land covers within urban areas at both fine spatial and temporal resolution. Whether it pertains to fields, roads, or buildings, each land cover type exhibits its unique thermal characteristics, contributing to the overall thermal dynamics of the urban landscape. By capturing these nuanced variations through fine-scale LST data, researchers and policymakers can derive valuable insights into the temporal patterns and changes in LST across different land covers. Conversely, LST products derived from coarse spatial resolutions, such as those with the resolution of 1000 m, encounter limitations in accurately capturing LST variations across various land covers. The lack of spatial detail in these coarse-resolution products tends to obscure the distinct thermal signatures

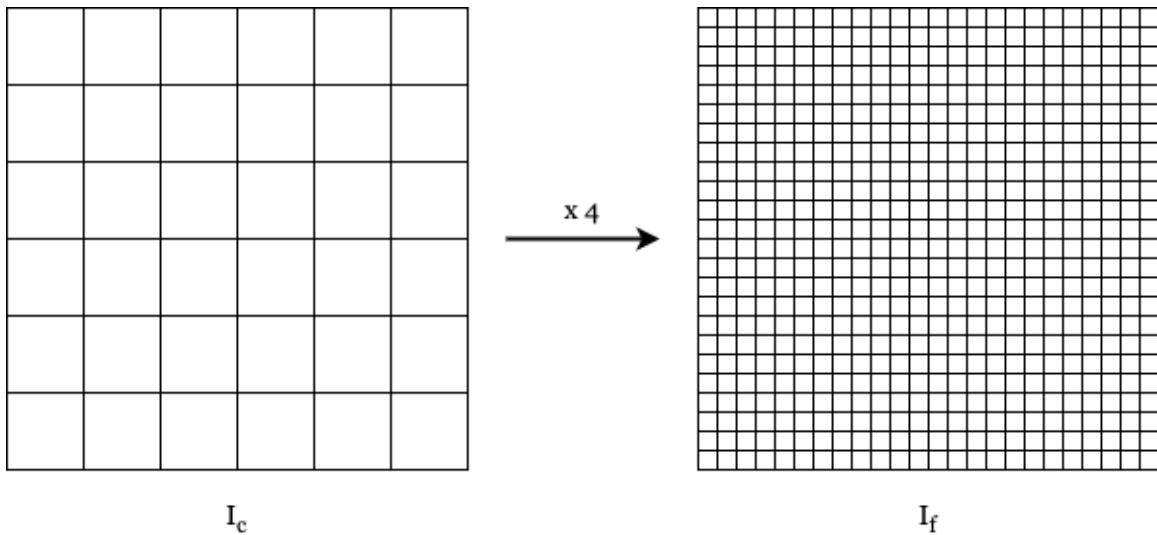


Figure 1.1: A simple representation of the ‘Downscaling’ process. Here,  $I_c$  refers to image at coarse resolution, whereas,  $I_f$  refers to image at fine resolution. ‘ $\times 4$ ’ denotes the downscaling factor (in this case: ‘4 times’)

associated with different land cover types. Consequently, conducting detailed assessments of LST patterns and their relationship to specific land covers becomes challenging.

In light of these challenges, spatial downscaling techniques assume a critical role in improving the spatial resolution of LST products. By augmenting the spatial detail, these techniques enable a comprehensive analysis of thermal behavior within urban areas. The application of spatial downscaling methods serves to overcome the ambiguities and limitations of coarse-resolution LST products, ensuring that fine-scale LST variations across diverse land covers can be accurately captured and analyzed.

Spatial downscaling methods can be divided into various categories such as: physical models (Hu et al., 2023), spatio-temporal fusion models (Weng et al., 2014; Zhu et al., 2010), and regression-based models (Agam et al., 2007; Huetengs & Vohland, 2016; W. Li et al., 2019).

### 1.3 LITERATURE REVIEW

This section briefly introduces various developed techniques in the context downscaling LST. Here, the section on regression-based models (1.3.3) has been provided in greater detail as compared to sections on physical (1.3.2) and spatio-temporal fusion models (1.3.1), as the scope of this thesis is limited to regression-based modelling.

#### 1.3.1 Spatio-temporal Fusion Models

Spatio-temporal fusion models are a class of downscaling techniques that attempts to leverage the fine spatial qualities of fine spatial resolution LST products and the fine temporal qualities of the

coarse spatial resolution LST products collected from different sensors, to construct fine spatio-temporal LST products. The base of these models lie in pixel mixture theory where the reflectance of a mixed coarse spatial resolution pixel can be decomposed into weighted linear combination of the reflectance of overlapping fine spatial resolution endmembers. After a series of assumption, the reflectance value of a fine spatial resolution pixel at time  $t_p$  is found to be dependent upon:

1. the reflectance value of the same fine resolution pixel at time  $t_o$
2. the reflectance value of the resampled coarse resolution pixel at time  $t_o$
3. the reflectance value of the resampled coarse resolution pixel at time  $t_p$
4. conversion coefficient obtained by fitting a linear regression model between the same end-member reflectance change from time  $t_m$  to time  $t_n$  and the coarse resolution reflectance change from time  $t_m$  to time  $t_n$

The above short summary forms the base of Zhu et al. (2010) and Weng et al. (2014). It is also worth noting that unlike the physical and regression models (discussed in Sections 1.3.2 and 1.3.3), spatio-temporal fusion models directly address the enhancement of temporal resolution. This is evident from the above explanation where the model directly incorporates changes in reflectance values over time as its parameters. Furthermore, since the process of obtaining the conversion coefficient parameter listed above is distinct from that of other parameters, the time  $t$  is labelled in a different manner.

### 1.3.2 Physical Models

Physical downscaling models are a class of spatial downscaling techniques that incorporate physical principles and processes to estimate high-resolution variables based on coarser-resolution data. These models have proven particularly valuable in the context of downscaling LST, as they leverage knowledge of energy transfer, heat conduction, radiation, and other physical phenomena to accurately estimate LST values at finer spatial resolutions. To achieve this, physical downscaling models take into account various factors that influence thermal patterns, including land cover characteristics, topography, atmospheric conditions, and surface energy balance components. By incorporating these factors into their algorithms, these models simulate the intricate fine-scale variations of LST, providing detailed estimations that reflect the underlying physical mechanisms at play (X. Li et al., 2022).

It is important to acknowledge that physical downscaling models also have their limitations. One notable challenge is the inherent complexity and demanding computational requirements associated with their design. Implementing these models often requires advanced expertise and substantial computational resources, which can be a barrier to their widespread adoption. Furthermore, since the incorporated physical processes are governed by a lot of meteorological and physical variables, these models require multiple data inputs which extends the complexity of these models.

### 1.3.3 Regression Models

Regression-based models are a type of spatial downscaling models that aim to establish a statistical relationship between input predictors and a target variable at a coarse resolution. The primary objective of these models is to leverage this established relationship to predict the target variable at a more detailed spatial resolution, utilizing fine-resolution predictors as inputs. The underlying assumption of regression-based models is scale-invariance, which suggests that the relationship between predictors and the target variable remains consistent despite spatial scale variations. In practical terms, researchers typically train regression-based models using data collected at a specific spatial resolution denoted as X. After model training and validation at spatial resolution X, it can be utilized to estimate the target variable at a different spatial resolution, denoted as Y (where  $X > Y$ ). This transferability of the model across resolutions constitutes a key advantage of regression-based downscaling, enabling the estimation of the target variable in regions or time periods where only coarse-resolution data is available.

In context of downscaling LST, various predictors are commonly employed by researchers to capture the relationship between these predictors and LST. These predictors typically include spectral indices derived from remote sensing data. Additionally, terrain factors such as elevation, slope, and aspect, as well as land use and land cover information, are also often integrated as predictors in these models. The inclusion of such predictors aims to account for and model as much variability as possible in the target variable, LST. However, one of the challenges in regression-based downscaling is the selection of appropriate predictors. Since the choice of predictors is usually subjective, uncertain, and specific to the application (Hu et al., 2023), researchers must carefully consider which predictors to include in the model. The selection process may involve expert knowledge, statistical analysis, and exploration of previous studies.

Nevertheless, regression-based modeling remains the most commonly adopted and experimented method in downscaling LST studies (X. Li et al., 2022). The success of regression models can be attributed to their practical ease of implementation compared to other methods (Hutengs & Vohland, 2016).

Extensive research has been undertaken in the field of regression-based LST downscaling models during the last two decades. Kustas et al. (2003) proposed a technique called DisTrad, which utilized least square regression between Normalized Difference Vegetation Index (NDVI) and LST. This approach aimed to downscale MODIS LST products from a resolution of 1000 m to MODIS NDVI resolution of 250 m. Subsequently, Agam et al. (2007) refined the DisTrad algorithm and introduced a new method called TsHARP. In TsHARP, fractional vegetation cover was used as a predictor instead of NDVI, which reduced the DisTrad algorithm's Root Mean Square Error (RMSE) from 1.5 °C to a range of 0.67–1.35 °C. These methods demonstrated good performance in areas with high vegetation cover. However, their effectiveness in urban areas was limited due to the absence of predictors associated with urban LST variability. To address this limitation, both NDVI and surface albedo were utilized as predictors by Dominguez et al. (2011) in the development of a new method called High-Resolution Thermal Sharpener (HUTS) with a focus on urban areas. Comparative analysis between HUTS and TsHARP revealed superior performance of HUTS. This observation highlights that the inclusion of additional predictors correlated with LST in downscaling models tends to enhance their performance. Nonetheless, it is important to note that assuming a linear relationship between all incorporated predictors and LST would be erroneous. This underscores the necessity of employing regression models capable of capturing non-linear and complex relationships between predictors and the target variable.

Given the advantages of machine learning methods in image processing and image regression, in the last decade several studies explored the benefits of applying such advanced techniques to downscale LST images. For example, Yang et al. (2010) employed Artificial Neural Network (ANN) to downscale LST from a resolution of 360 m to 90 m. This ANN-based downscaling approach outperformed the DisTrad method in all conducted experiments. Bindhu et al. (2013) proposed a Non-Linear DisTrad (NL-DisTrad) algorithm, which combined a hot edge model with an ANN. The NL-DisTrad algorithm achieved an RMSE of less than 0.96 °C for the task of downscaling MODIS LST. Similar to previous statistical methods, NL-DisTrad utilized NDVI as a predictor for LST downscaling. Hutengs and Vohland (2016) employed Random Forest (RF) regression, a well-known machine learning algorithm, to downscale LST. Their approach incorporated surface reflectances (red, blue, green, short-wave infrared, and near-infrared channels), topography variables derived from Digital Elevation Model (DEM), and a land use map as predictors for the downscaling model. Furthermore, W. Li et al. (2019) applied popular machine learning algorithms such as RF and Support Vector Machines (SVM) to downscale MODIS LST. The achieved RMSE values ranged from 2 to 3 °C. Their study leveraged multi-source predictors derived from land use land cover components, terrain factors, and different spectral indices. The selected machine learning techniques outperformed the TsHARP method. R. Wang et al. (2020) also observed a similar trend where the RF model outperformed the Multiple Linear Regression (MLR) model and TsHARP. These findings support the conclusion that non-linear models capable of capturing complex relationships between predictors and LST, particularly in urban areas, are necessary to achieve optimal downscaling performance.

The aforementioned studies have predominantly utilized global models, which do not explicitly consider spatial non-stationarity. However, considering the spatially varying nature of LST, it becomes evident that models accounting for spatial non-stationarity are necessary. Consequently, regression models that accommodate this aspect in LST downscaling have been seldom proposed. For instance, Duan and Li (2016) employed Geographically Weighted Regression (GWR) to downscale MODIS LST from 990 m to 90 m. In this approach, NDVI and Digital Elevation Model (DEM) were employed as predictors for the GWR model. GWR, proposed by Fotheringham et al. (2003), extends ordinary least square regression by fitting localized regression models using weighted neighboring data points, enabling the modeling of spatial non-stationarity by considering varying relationships between variables across different locations. To account for spatial autocorrelation in addition to spatial non-stationarity, S. Wang et al. (2020) proposed a Geographically Weighted AutoRegressive (GWAR) model to downscale MODIS LST from 1000 m to 100 m. GWAR extends GWR by incorporating a spatial autoregressive component to account for spatial autocorrelation or dependence. The GWAR model, utilizing Normalized Difference Building Index (NDBI) and DEM as predictors, exhibited superior performance compared to the GWR model. Furthermore, Xu et al. (2021) proposed the Multi-Factor Geographically Weighted Machine Learning (MFGWML) downscaling approach, which extends the concept of GWR by incorporating multiple predictors and leveraging the non-linear nature of machine learning algorithms. A common limitation of models based on the concept of GWR is the a priori assumption of the weight distribution function (Du et al., 2020). To address this issue, Du et al. (2020) proposed the Geographically Neural Network Weighted Regression (GNNWR), which employs a neural network to estimate the spatial weights based upon the input spatial distances within a neighborhood. Subsequently, Liang et al. (2023) utilized GNNWR to downscale Landsat-8 LST, incorporating predictors such as NDVI, NDBI, DEM, and slope data.

In summary, the application of machine learning methods like ANN, SVM, and RF in LST downscaling has yielded promising outcomes, outperforming traditional statistical approaches.

The flexibility of these techniques to model non-linear relationships between predictors and LST proves advantageous, particularly in urban areas. Furthermore, the discussed studies highlight the necessity of regression models that account for spatial non-stationarity. GWR, GWAR, MFGWML, and GNNWR represent notable advancements in addressing this concern and have showcased improved performance in capturing the complex spatial relationships between predictors and LST. Figure 1.2 displays the chronological progression of all the mentioned studies in this section.

## 1.4 RESEARCH GAP

As outlined in Section 1.3, regardless of the employed method—be it statistical, machine learning, or deep learning techniques—the primary predictors utilized for downscaling LST include surface reflectances and spectral indices (such as: NDVI, NDBI, Bare Soil Index (BSI), and more) derived from optical images. However, these predictors derived from optical data possess a significant drawback: their dependence on clear weather conditions and their inability to provide observations during night-time. Cloudy conditions often result in a substantial portion of the optical images being obscured by clouds, which can pose limitations when constructing a downscaling model and when using the model to estimate fine-scale LST. In a nutshell, the aforementioned models fail to produce accurate results under cloudy conditions and during night-time due to the inherent characteristics of optical imagery.

## 1.5 RESEARCH OBJECTIVES AND QUESTIONS

In this research, a solution is proposed to address the aforementioned research gap by employing Synthetic Aperture Radar (SAR) data for deriving predictors in order to downscale LST images. SAR systems utilize active imaging mode and microwave electromagnetic waves, enabling them to penetrate through clouds and acquire information from the Earth's surface during both daytime and night-time, irrespective of weather conditions (J. Li et al., 2018). The influence of radar parameters such as incidence angle, wavelength, and polarization, as well as surface parameters including surface roughness, soil moisture, and surface dielectric constant, on radar backscatter has been extensively studied by Benallegue et al. (1995), Hoeben et al. (1997), and Ulaby (1974). Notably, models such as the 'Oh' Model (Oh et al., 1992) and the 'IEM' model (Fung et al., 1992) have been thoroughly researched for soil moisture retrieval studies. Additionally, a Radar Vegetation Index (RVI) analogous to NDVI has been developed by Kim and van Zyl (2001). As vegetation cover influences SAR backscatter, NDVI can also be estimated from SAR backscatter (Filgueiras et al., 2019; Veloso et al., 2017). These studies rely on backscattering coefficients derived from different polarization modes (HH, VV, HV, VH), where polarization modes denote the direction of the electric field of the radar signal (i.e., V for Vertical and H for Horizontal). The aforementioned parameters of vegetation cover and soil moisture are critical factors explaining the variation in surface LST (Song et al., 2018; Sun & Kafatos, 2007; Zhang et al., 2014). Consequently, it is reasonable to assume the existence of a relationship between the backscattering coefficients obtained from different polarization channels of SAR and LST. Therefore, the backscattering coefficients derived from SAR can be modelled as predictors to elucidate the variability in LST. This anticipated relationship between radar-derived surface parameters and temperature motivates exploration of SAR images' potential for downscaling LST products. Leveraging the advantages of SAR data, the proposed utilization of predictors derived from SAR has the potential to significantly

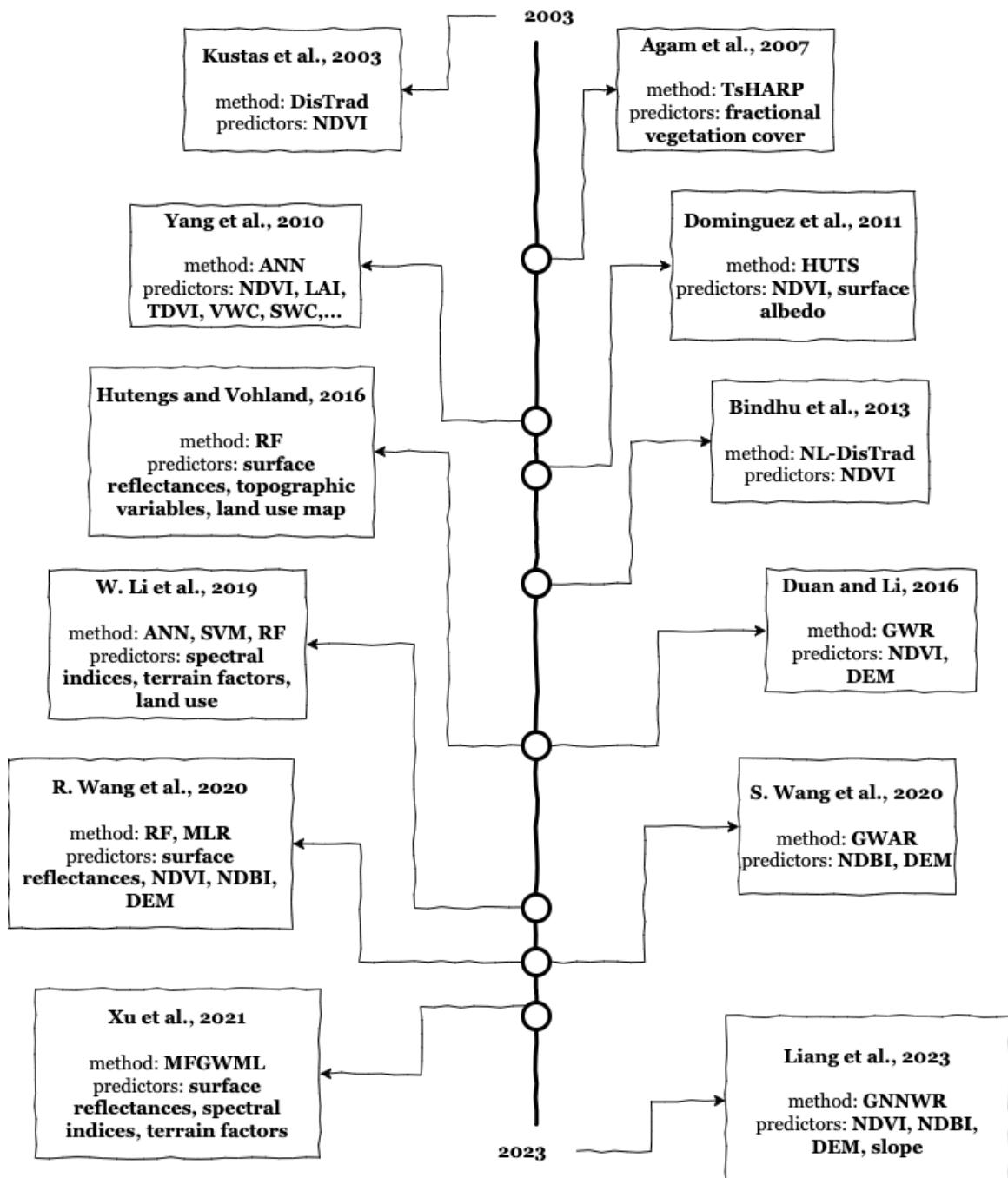


Figure 1.2: Chronological progression of the regression-based LST downscaling studies discussed in the Literature Review section

impact LST mapping and monitoring, potentially replacing or supplementing optical images in delivering high spatio-temporal LST data.

This brings us to the primary research objective, i.e., to downscale LST from 1000 m spatial resolution to 100 m spatial resolution using SAR data.

Furthermore, it is essential to acknowledge that LST is a complex phenomenon that cannot be adequately modeled using a single data source. Consequently, many downscaling studies discussed in the literature review section aim to integrate multiple sources of information correlated with LST to achieve more accurate estimations. Consistent with this approach, this research incorporates land cover information as additional predictors for the downscaling models. However, instead of following the conventional approach of integrating land cover information directly into the downscaling model, this research utilizes a novel approach to integrate land cover information as predictors into these models. This novel approach is discussed in Section 3.2.3.

In addition to utilizing ancillary data such as land cover information to account for maximum possible variation in LST, it is also intriguing to explore whether additional features can be engineered from the primary data source, i.e., SAR, to address this issue and subsequently improve the performance of the downscaling models.

As emphasized in the literature review section, traditional machine learning algorithms have been extensively studied in the context of downscaling LST. In line with these studies, RF, proposed by Breiman (2001), is selected as the base regression algorithm for the developed downscaling models in this research. The RF algorithm functions by constructing an ensemble of decision trees that are trained on random subsets of data. The predictions generated by these individual trees are subsequently combined through voting or averaging methods, resulting in the final prediction. The choice of RF regression is motivated by its success in LST downscaling studies (Hutengs & Vohland, 2016; W. Li et al., 2019; X. Li et al., 2022; Njuki et al., 2020; R. Wang et al., 2020), as well as the advantages highlighted in these studies, such as its non-linear and non-parametric nature, reduced risk of overfitting, ability to handle high-dimensional datasets, incorporation of both continuous and categorical variables, and efficient computation time.

In addition to the RF model, this research also employs Convolutional Neural Networks (CNN) to downscale LST. CNN belong to a class of deep learning models that have been significantly experimented with in the fields of computer VISion and image processing (Krizhevsky et al., 2012; Lecun et al., 2015). These networks adopt a hierarchical approach, progressively learning low-level features, such as edges and textures, in initial layers and gradually advancing to higher-level concepts like shapes and objects in deeper layers. CNN have emerged as the preferred choice for various tasks such as, object detection, image classification, and image segmentation. It is noteworthy that CNN are not solely restricted to classification tasks but can also be effectively utilized for regression problems, such as the regression-based downscaling considered in this study. By modifying the architecture and output layer of a CNN, it becomes adept at predicting continuous values instead of discrete class labels. In contrast to RF, CNN excels in extracting spatial information from input data. CNN possess an inherent ability to learn and exploit the spatial relationships present in the data, rendering them particularly valuable for downscaling tasks. The convolutional layers of a CNN capture local patterns and features from the input, while subsequent layers, such as fully connected layers, facilitate the regression process by making predictions based on these learned features. The integration of deep learning, in general, into the domain of LST downscaling has been relatively limited. Studies like Yang et al. (2010) and Liang et al. (2023) have integrated neural

networks for downscaling LST, however, these studies are based on fully connected networks and do not incorporate CNN for regression-based downscaling. It is important to note that a super-resolution convolutional neural network architecture (Dong et al., 2016), which employs CNN to learn the end-to-end mapping between input images, has been utilized for downscaling NDVI in the study conducted by Nomura and Oki (2021). However, to the best of our knowledge, the incorporation of a regression-based CNN architecture for LST downscaling has not yet been investigated. In light of this, this research proposes a novel regression model for downscaling LST based on the CNN architecture with the aim of incorporating intrinsic spatial patterns that are pertinent to the downscaling process. The developed architecture is compared with the well-studied RF model to assess its performance. This proposed novel architecture is discussed in Section 3.3.

Finally, to evaluate the effectiveness of the predictors derived from SAR data, the performance of the developed downscaling models with predictors derived from optical data is used as a benchmark. In summary, the preceding paragraphs describing the research can be outlined in terms of its objectives and the corresponding research questions as follows:

#### **1.5.1 Main Objective**

To downscale LST using predictors derived from SAR data

#### **1.5.2 Sub-Objectives**

1. To utilize RF regression algorithm as the base model for downscaling LST
  - What new features can be engineered from the SAR data to achieve better performance for the task of downscaling?
2. To integrate land cover information into the developed downscaling models
  - How does the proposed novel approach of incorporating land cover information into the downscaling models perform compared to the traditional approach?
3. To downscale LST with the proposed Convolutional Neural Network (CNN)-based regression architecture
  - How effective is the proposed novel CNN-based architecture for the task of downscaling?
4. To downscale LST using predictors derived from optical data
  - How does the performance of SAR-based LST downscaling models compare to the performance of optical-based LST downscaling models?

#### **1.6 CONTRIBUTIONS**

Based on the above described research objectives (Section 1.5), the contributions made by this research can be summarized as follows:

1. We propose a novel way of estimating fine-scale LST from coarse-scale LST using SAR data
2. We engineer innovative features from SAR dataset to increase the performance of base models
3. We propose a novel way of incorporating land cover information into LST downscaling models
4. We propose a novel CNN-based regression architecture for LST downscaling

### 1.6.1 Highlights

1. Preliminary results from this research were presented at the Nederlands Aardwetenschappelijk Congres (NAC) GEO conference held in Utrecht on March 2023 under the title of ‘Spatial Downscaling of Land Surface Temperature (LST) using radar data’
2. A part of this research was submitted for publication in Remote Sensing (MDPI) on June 2023. The preprint (Patel et al., 2023) is available at: <https://www.preprints.org/manuscript/202306.1391/v1>
3. To contribute to the advancement of open source academia, the code corresponding to all the developed and implemented models is hosted on GitHub: [https://github.com/ohheynish/DWNSCL\\_LST\\_MSc](https://github.com/ohheynish/DWNSCL_LST_MSc).

## 1.7 THESIS STRUCTURE

This thesis, excluding ‘Introduction’ (Chapter 1), consists of five other chapters: ‘Study Area, Datasets, and Tools’ (Chapter 2), ‘Methodology’ (Chapter 3), ‘Results’ (Chapter 4), ‘Discussion’ (Chapter 5), and ‘Conclusion’ (Chapter 6). The study area and the datasets utilized in this research are introduced in Chapter 2. The proposed methodology for experimentation is described in Chapter 3, whereas the experimental results are provided in Chapter 4. The discussion of the achieved results is presented in Chapter 5. Finally, Chapter 6 is dedicated to the conclusion of the thesis.

## Chapter 2

# Study Area, Datasets, and Tools

This chapter serves as an introduction to the study area, datasets, and tools that have been taken into account for this research. The sections are correspondingly titled Study Area (Section 2.1), Datasets (Section 2.2), and Tools (Section 2.3). The Datasets section is subsequently divided into four sub-sections, providing a comprehensive description of the use case (Section 2.2.1), the data collection procedure (Section 2.2.2), the undertaken pre-processing steps (Section 2.2.3), and the selection of dates of interest for the research (Section 2.2.4).

### 2.1 STUDY AREA

As shown in Figure 2.1, the considered study area for this research is Zuid-Holland province in the Netherlands. Zuid-Holland covers an area of approximately 3400 km<sup>2</sup> and it is one of the most densely populated and urbanized regions in the country, with a population of over 3.5 million people. Based upon the global surface temperature data collected by the world bank (<https://climateknowledgeportal.worldbank.org/>), in 2020, Zuid-Holland province experienced an annual minimum temperature of approximately 281 K, an annual maximum temperature of 289 K, and an annual mean temperature of 285 K. These statistics provide a broad overview of the temperature regime in the region, but it is important to note that the LST is measured in a different way than the Global Surface Temperature. The statistics for LST over Zuid-Holland collected from Landsat-8 LST at 100 m for the selected dates of interest are shown in Table 2.1. Furthermore, it is also important to note that the actual temperature dynamics at the local scale can vary significantly depending on factors such as topography, land cover, and urbanization.

According to the ESA world cover dataset i.e., ‘ESA WorldCover v100’, the landscape of Zuid-Holland province is primarily characterized by grassland and built-up areas, accounting for approximately 52 % and 17 % of the total land area, respectively. Cropland and tree cover also contribute significantly to the land cover, each covering approximately 14 % of the total land area. Bare/sparse vegetation and herbaceous wetland are relatively less abundant, each covering approximately 1 % of the total land area. The province is also characterized by a coastal zone, polders, and river deltas. The coastline of Zuid-Holland is about 70 km long and consists of sandy beaches, dunes, and tidal flats. The polders in the province are low-lying areas that have been reclaimed from the sea and are used for agriculture and horticulture. The river deltas of the Rhine and Maas rivers are also important features of the landscape of Zuid-Holland. Such a variety of land covers in the study case can expectedly change LST from region to region. High spatio-temporal resolution remote sensing images offer the possibility to track these LST changes, enabling a better understanding of the temperature dynamics in Zuid-Holland.

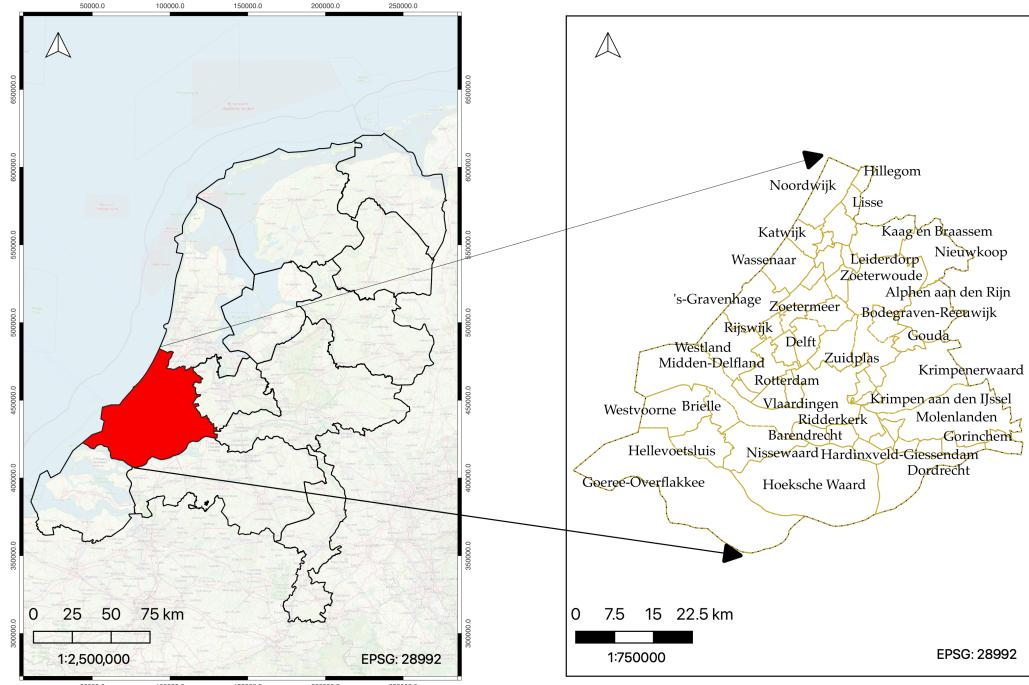


Figure 2.1: Map showing the study area: Zuid-Holland province (left) and various municipalities in Zuid-Holland province (right)

Table 2.1 Summary statistics for Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020

Landsat-8 Level 2, Collection 2, Tier 1 LST (100 m)			
	25/03/2020	10/04/2020	28/05/2020
<b>Minimum</b>	275.60	267.68	276.61
<b>Maximum</b>	302.02	313.16	319.19
<b>Mean</b>	286.59	291.94	299.60
<b>Standard deviation</b>	3.30	6.19	6.08
<b>1<sup>st</sup> quartile</b>	284.48	290.97	296.36
<b>3<sup>rd</sup> quartile</b>	287.37	313.16	303.61

## 2.2 DATASETS

### 2.2.1 Dataset Description and Use Case

To accomplish the primary objective of this research, which is to downscale LST using SAR images, various sets of satellite remote sensing data were employed. The utilized datasets consist of Landsat-8 LST (100 m), MODIS LST (1000 m), Sentinel-1 Ground Range Detected (GRD) SAR (10 m), Sentinel-2 Multi Spectral Instrument (MSI) (10 m), and ESA WorldCover v100 land type information (10 m).

As indicated by the research objectives, the downscaling models developed in this study can be categorized into two types: radar-based and optical-based. Radar-based models utilize predictors derived from radar data as inputs, while optical-based models utilize predictors derived from optical data.

For the construction of radar-based downscaling models, Landsat-8 LST data was employed as the target and reference variable, and Sentinel-1 GRD SAR data served as the basis for deriving the predictor variables. The models were trained using the coarse resolution version of both datasets where the datasets were aggregated to a spatial resolution of 1000 m. Subsequently, the constructed models were used to produce a high-resolution radar-derived LST image using the available high-resolution SAR images. Additionally, for subsequent experiments, ESA WorldCover data was also incorporated as a predictor variable in the radar-based downscaling models. To assess the efficacy of radar-derived predictors for the task of downscaling, the performance of the radar-based models was evaluated and compared with downscaling models constructed based on the optical datasets. In this regard, Sentinel-2 MSI optical data was used as the basis for deriving the predictor variables. Furthermore, the use of aggregated Landsat-8 LST data as a target variable does not correspond to a practical real-world scenario, as Landsat-8 LST is not originally collected at a coarse spatial resolution. Consequently, the experiments were repeated using MODIS LST (1000 m) as the target variable, which aligns better with real-world conditions. Further details regarding this are elaborated in Section 3.4. The use case for all the considered datasets in this research is provided in Table 2.2.

It is worth noting that the acquired Landsat-8 LST product was sourced from Landsat-8 Level-2 Collection-2 Tier-1 dataset. The ‘Level-2’ refers to products that have undergone atmospheric correction and have been converted to surface reflectance values from digital numbers (Level-1 product). ‘Collection-2’ denotes the reprocessed collection of Landsat products, characterized by improved data quality compared to Collection-1 (which has been discontinued by USGS since December 30, 2022). Finally, ‘Tier-1’ data signifies products that meet predetermined geometric and radiometric quality requirements. As previously mentioned, the radar data was obtained from the Sentinel-1 GRD SAR product. GRD data represents a multi-looked and ground range projected version of the Single Look Complex (SLC) data collected by the Sentinel-1 SAR satellite. The GRD product does not retain the phase information present in the raw SLC data. Essentially, the Sentinel-1 GRD data comprises two bands: VV and VH, with pixel values corresponding to the measured backscattering coefficient. Lastly, the Sentinel-2 MSI data used for the optical-based downscaling model also originated from a Level-2A dataset, where ‘Level-2A’ refers to atmospherically corrected data derived from Level-1C products. Further information about the utilized products can be found on their respective Google Earth Engine (GEE) pages. The GEE tags of all the datasets used in this research are provided in Table 2.3.

**Table 2.2** Overview of the datasets used in this research (Part 1)

Dataset	Original Spatial Resolution	Use Case
Landsat-8 LST	100 m	Target variable (Aggregated 1000 m LST)
		Validation data (Original 100 m LST)
MODIS LST	1000 m	Target Variable
Sentinel-1 SAR	10 m	Predictor variable
Sentinel-2 MSI	10 m	Predictor Variable
ESA WorldCover v100	10 m	Predictor Variable

**Table 2.3** Overview of the datasets used in this research (Part 2)

Dataset	GEE Tag	Acquisition Date and Time
Landsat-8 LST	'LANDSAT/LC08/C02/T1_L2'	25-03-2020 (T-10:33)
		10-04-2020 (T-10:33)
		28-05-2020 (T-10:33)
		25-03-2020 (T-11:30)
MODIS LST	'MODIS/061/MOD11A1'	-
Sentinel-1 SAR	'COPERNICUS/S1_GRD'	25-03-2020 (T-17:25) 11-04-2020 (T-17:33) 29-05-2020 (T-17:33) 26-03-2020 (T-10:46)
Sentinel-2 MSI	'COPERNICUS/S2_SR_HARMONIZED'	-
ESA WorldCover	'ESA/WorldCover/v100'	One image for the entire year of 2020

### 2.2.2 Google Earth Engine

All the aforementioned datasets were collected and processed using the GEE Python Application Programming Interface (API). GEE is a powerful platform extensively utilized by researchers to analyze satellite data for a wide range of applications. It offers an extensive collection of geospatial data and a robust processing infrastructure, enabling users to explore, visualize, and extract valuable insights from satellite imagery. Furthermore, GEE facilitates the direct transfer of coarse resolution data to Python data structures, eliminating the need for data download and storage, thus enhancing the reproducibility of the data collection process. However, there are certain limits imposed by GEE on transferring large volumes of data. Consequently, the fine resolution images utilized in this research had to be downloaded separately. Nonetheless, GEE also provides a convenient method for direct data download via Google Drive. Another notable advantage of utilizing GEE as a data repository is the ability to work directly with pre-processed datasets. For instance, in the case of Sentinel-1 GRD SAR data, several crucial pre-processing steps, such as applying orbit file, border noise removal, thermal noise removal, and terrain correction, are necessary whenever analyzing this dataset. Fortunately, these pre-processing steps are inherently incorporated within the GEE processing pipeline. Therefore, Sentinel-1 GRD SAR images downloaded or queried through GEE do not require additional implementation of these pre-processing steps.

### 2.2.3 Pre-processing

Since this research aims to develop models using data or images acquired from multiple sources, certain pre-processing steps are crucial to ensure accurate function approximation across the images. These pre-processing steps include essential procedures such as clipping, resampling, and reprojection of the datasets, which are performed to achieve proper coregistration between the datasets. For the purpose of clipping all the datasets to the same extent, the ‘Zuid-holland’ feature from the Global Administrative Unit Layers (GAUL) feature collection was utilized as the boundary/extent mask. Additionally, all the datasets were reprojected to the European Petroleum Survey Group (EPSG):4326 coordinate system to ensure consistency. This coregistration procedure is fundamental for establishing an accurate correspondence between the predictor variables and the target variables in the downscaling models.

Furthermore, noise removal is another fundamental process that was implemented in the pre-processing stage. All the acquired datasets except ‘ESA WorldCover v100’ were filtered using a median filter with  $3 \times 3$  kernel to remove noise.

### 2.2.4 Selection of Dates of Interest

Another important aspect to consider for ensuring accurate function approximation between images is their date of acquisition. A closer proximity in acquisition dates enhances the accuracy of the approximated function. However, one significant challenge in achieving this is the presence of cloud contamination in LST and optical images. To address this issue, dates with low cloud cover were initially identified based on the MODIS quality indicator band (‘QC’). Subsequently, from this subset of dates, only those dates where all the considered satellites acquired data were selected as the dates of interest for the analysis.

From Table 1.1, it is evident that Landsat-8 acquires images every 16 days, while MODIS acquires images on a daily basis. Furthermore, the temporal resolution of Sentinel-1 is 5 days, whereas that of Sentinel-2 is 10 days. Due to these temporal configurations and the weather-dependent nature of LST and optical images, finding intersecting acquisition dates among these satellites is rare. In cases where intersecting dates were absent, a one-day buffer was considered as the criteria to select the dates of interest. Following this approach, the selected dates of interest for Landsat-8 LST products were determined as 25<sup>th</sup> March, 2020, 10<sup>th</sup> April, 2020, and 28<sup>th</sup> May, 2020. Correspondingly, the selected dates of interest for Sentinel-1 SAR products were identified as 25<sup>th</sup> March, 2020, 11<sup>th</sup> April, 2020, and 29<sup>th</sup> May, 2020. Here, the SAR images collected on 25<sup>th</sup> March, 11<sup>th</sup> April and 29<sup>th</sup> May map to the LST images collected on 25<sup>th</sup> March, 10<sup>th</sup> April, and 28<sup>th</sup> May, respectively. For the optical-based downscaling experiment, only one image acquired on 26<sup>th</sup> March, 2020, was considered, and this image was mapped to the LST image acquired on 25<sup>th</sup> March, 2020. Similarly, for the downscaling LST experiment using MODIS LST, only one MODIS LST image acquired on 25<sup>th</sup> March, 2020, was considered. The selected acquisition dates for the datasets utilized in this research are summarized in Table 2.3.

### 2.3 TOOLS

The data collection process and the associated pre-processing steps, including clipping, resampling, and reprojection, were performed utilizing the GEE Python API. The downscaling models were developed and quantitatively evaluated using the Python programming language. As for qualitative evaluation purposes, QGIS software was employed. It is noteworthy that all Python scripts were developed and tested on the CRIB geospatial computing platform.

## Chapter 3

# Methodology

This chapter introduces the methodology devised for this research. Section 3.1 introduces the fundamental downscaling algorithm, while Sections 3.2 and 3.3 describe the adopted methodologies for utilizing RF and CNN as downscaling models, respectively. In Section 3.4, a brief overview is provided regarding the utilization of MODIS LST as the target variable instead of the aggregated Landsat-8 LST. Finally, Section 3.5 introduces the adapted quantitative evaluation framework to assess the performance of the developed models. The entirety of this chapter's methodology can be succinctly summarized by referring to the flowchart displayed in Figure 3.1.

As depicted in Figure 3.1, the initial step involves aggregating the datasets considered for this research to either a spatial resolution of 100 m or 1000 m, depending on their specific use-case. To achieve this, median-based aggregation was employed. Subsequently, the datasets utilized as features or predictors are separated from the target variable (LST). The predictor datasets then undergo certain data transformations and feature engineering processes, resulting in the creation of two distinct sets: training features and prediction features. The training features are employed to train the model using the coarse spatial resolution (1000 m) target variable (LST). Conversely, the prediction features are used during the inference stage to estimate the target variable (LST) at a finer spatial resolution (100 m). Further elaboration on this methodology can be found in subsequent sections of this chapter.

### 3.1 BASIC SPATIAL DOWNSCALING ALGORITHM

As previously illustrated, Figure 1.1 depicts a fundamental spatial downscaling process. This process can be decomposed into several mathematical equations, beginning with:

$$M_c = f(P_c, T_c) \quad (3.1)$$

where:

$M_c$  = regression model between the sets of predictors  $P_c$  and the target values  $T_c$

$f$  = mapping function approximated by training the model  $M_c$  with predictors  $P_c$  and the target values  $T_c$

$P_c$  = predictors collected at coarse spatial resolution

$T_c$  = target values collected at coarse spatial resolution

Equation 3.1 describes the process of training the regression model, denoted as  $M_c$ , using the predictors and target variables obtained at a coarse spatial resolution. Here, the subscript 'c' denotes

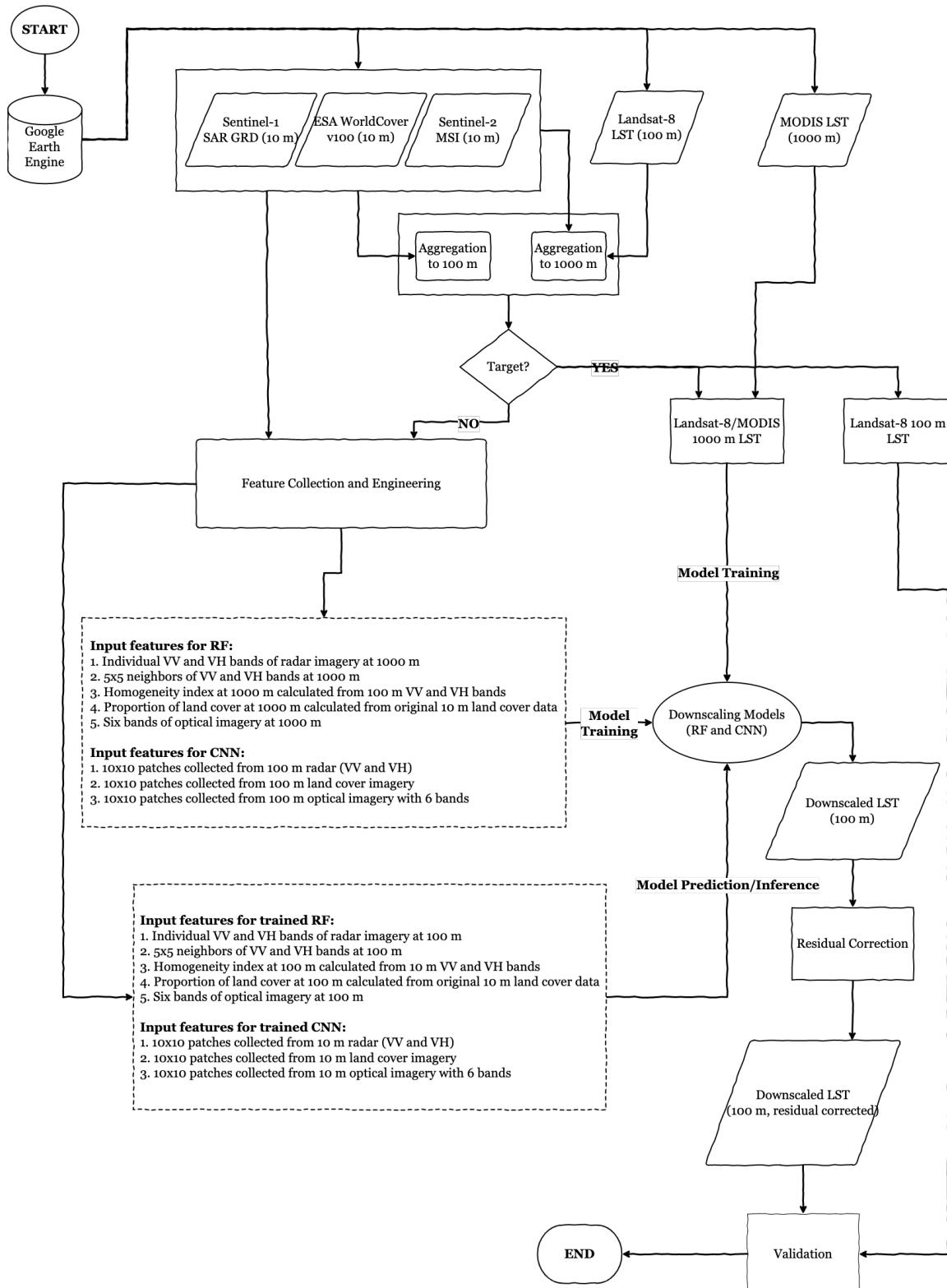


Figure 3.1: Methodology Flowchart

the coarse resolution. In the context of downscaling LST, the variable  $T_c$  represents the target LST to be downscaled, while  $P_c$  refers to the predictor variables collected from radar, optical, or other ancillary datasets.

Once the model  $M_c$  is trained at the coarse resolution, it can be utilized to generate or estimate the target values at a fine spatial resolution using the predictors collected at that resolution. This can be expressed as:

$$T_f = M_c(P_f) \quad (3.2)$$

where:

$M_c$  = trained regression model from equation 3.1

$P_f$  = predictors collected at fine spatial resolution

$T_f$  = target values estimated by the trained regression model  $M_c$

Equations 3.1 and 3.2 embody the ‘scale-invariance’ assumption of regression-based downscaling models. In practical terms, it is assumed that the model trained at a coarse resolution can estimate the target variable at a fine resolution, presuming that the relationship between the predictors and target variables remains invariant across different scales.

Moreover, typical regression models employed for downscaling often fail to capture all the variation present in the target variable. To address this issue, a residual correction procedure was introduced by Kustas et al. (2003), which can be described as:

$$\Delta T = T_{fc} - T_c \quad (3.3)$$

where:

$T_{fc}$  = reaggregated  $T_f$  from fine spatial resolution to coarse spatial resolution

$T_c$  = target values collected at coarse spatial resolution

$\Delta T$  = residuals collected at coarse spatial resolution

Once these residuals are collected at the coarse spatial resolution, they are added back to the overlapping pixels at fine spatial resolution. In other words, the estimates of all the fine spatial resolution pixels covering one coarse spatial resolution pixel are adjusted by the corresponding coarse spatial resolution residual value.

Finally, the downscaled target value, denoted as  $T'_f$ , can be estimated as:

$$T'_f = T_f + \Delta T \quad (3.4)$$

where:

$T'_f$  = residual-corrected downscaled target variable

The set of equations: 3.1, 3.2, 3.3, and 3.4, collectively represents the entire spatial downscaling process. Although the specific predictors ‘ $P$ ’, target variable ‘ $T$ ’, coarse resolution ‘ $c$ ’, fine resolution ‘ $f$ ’, and regression model ‘ $M$ ’ may vary across different studies, the underlying algorithm remains the same.

### 3.2 RANDOM FOREST

In accordance with sub-objective 1 (1.5.2-1), this research utilizes RF as the fundamental model for achieving the main objective (1.5.1), i.e., to downscale LST using radar data. This section is organized into three subsections, each focusing on a specific aspect of the developed downscaling models. These subsections delve into the radar-based downscaling model, the optical-based downscaling model, and the integration of land cover information into both models, respectively.

#### 3.2.1 Radar-based RF downscaling model

As outlined in Section 1.5, the downscaling models developed in this research can be classified into two main categories: radar-based models and optical-based models. Furthermore, this research attempts to downscale LST from 1000 m spatial resolution to 100 m spatial resolution. Consequently, in the context of this research, the Equations 3.1, 3.2, 3.3, and 3.4 can be reformulated for the radar-based downscaling models as follows:

$$RF_{1000} = f(SAR_{1000}, LST_{1000}) \quad (3.5)$$

$$LST_{100} = RF_{1000}(SAR_{100}) \quad (3.6)$$

$$\Delta LST_{1000} = LST_{100 \rightarrow 1000} - LST_{1000} \quad (3.7)$$

$$DLST_{100} = LST_{100} + \Delta LST_{1000} \quad (3.8)$$

where:

$RF_{1000}$  = random forest regression model trained at 1000 m coarse resolution

$SAR_{1000}$  = radar predictors collected from the aggregated 1000 m coarse resolution Sentinel-1 SAR dataset

$LST_{1000}$  = 1000 m aggregated coarse resolution LST from Landsat-8 LST dataset

$SAR_{100}$  = radar predictors collected from the aggregated 100 m fine resolution Sentinel-1 SAR dataset

$LST_{100}$  = LST estimated at 100 m fine resolution from the trained  $RF_{1000}$  model

$LST_{100 \rightarrow 1000}$  = reaggregated  $LST_{100}$  from 100 m fine resolution to 1000 m coarse resolution

$\Delta LST_{1000}$  = LST residuals collected at 1000 m coarse resolution

$DLST_{100}$  = downscaled LST at 100 m fine resolution

As stated in the above description regarding the variables used in the radar-based RF downscaling model, the predictor variables  $SAR_{1000}$  and  $SAR_{100}$  are derived from the Sentinel-1 GRD SAR dataset. Specifically, the VV and VH bands available in the dataset were utilized as predictors ( $SAR_{1000}$  and  $SAR_{100}$ ) for the initial downscaling experiments. Due to the distinct scattering mechanisms of radar waves, even minor alterations in ground features' structure or intrinsic characteristics can lead to significant variations in backscattering values for both VV and VH bands. This inherent limitation of radar data results in spatial patterns observed in radar images being less smooth compared to those observed in optical or LST images (Figure 5.1).

One approach to address this challenge is to apply excessive spatial filtering. However, this can potentially lead to information loss, hindering the model's ability to capture all the variations

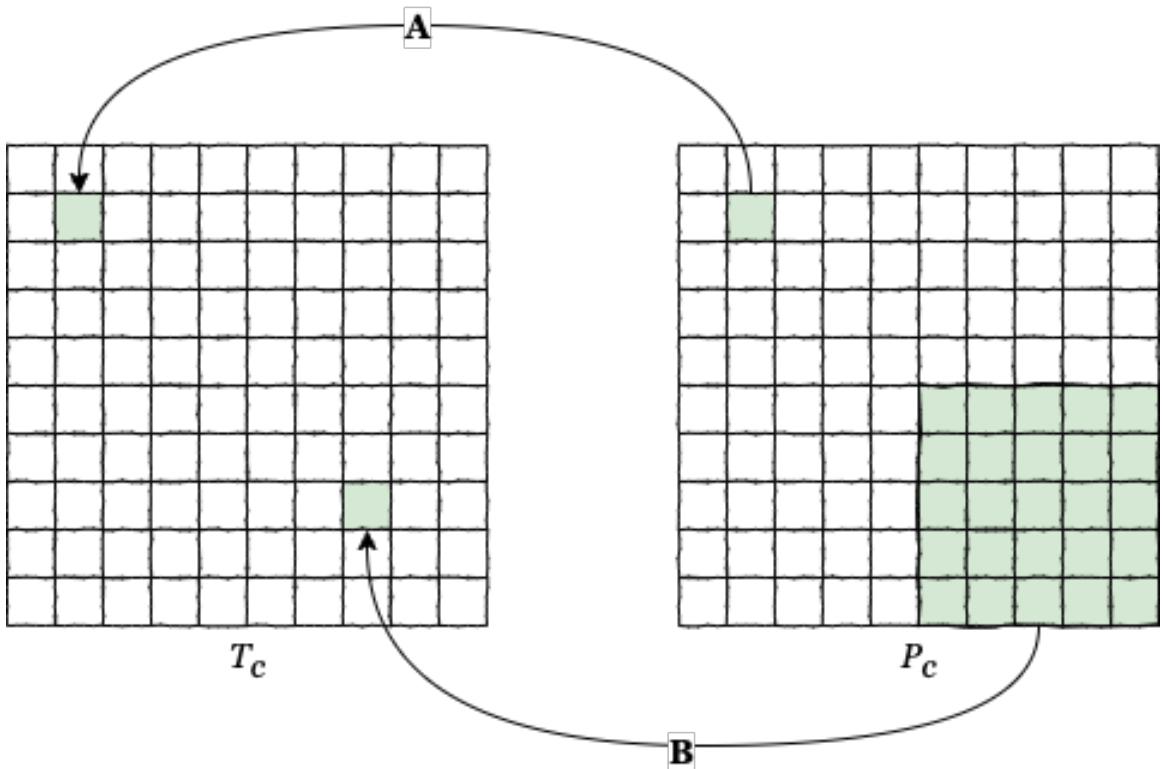


Figure 3.2: Representation of inclusion of neighboring values as features for the downscaling models. Here, ‘Connection A’ in the image illustrates conventional pixel-to-pixel mapping between coarse resolution predictor image  $P_c$  and coarse resolution target image  $T_c$ , whereas, ‘Connection B’ illustrates pixel-to-( $5 \times 5$ )patch mapping between coarse resolution predictor image  $P_c$  and coarse resolution target image  $T_c$

in the target LST. To mitigate this limitation, an additional feature engineering technique was employed, incorporating the neighboring values of all pixels within a  $5 \times 5$  patch for both VV and VH channels as supplementary predictors for the downscaling model. Figure 3.2 visually illustrates this concept of feature inclusion and engineering.

To further address the limitation of noisy spatial patterns in radar data, a novel radar homogeneity index was proposed and employed as an additional predictor for the radar-based downscaling experiments. This radar homogeneity index can be described mathematically as:

$$H.I.(X) = \frac{|\{x_i : x_i \in X, i \in [1, n]\}|}{|\{x_j : x_j \in A, j \in [1, N]\}|} \quad (3.9)$$

where:

$H.I.(X)$  = homogeneity index value of a patch  $X$

$x$  = pixel value

$X$  = a  $10 \times 10$  patch in an image  $A$

$n$  = Total number of pixel values in a patch  $X$

$N$  = Total number of pixel values in an image  $A$

As represented by Equation 3.9, the value of the homogeneity index  $H.I.(X)$  is calculated for a  $10 \times 10$  patch. When overlapping a 1000 m image with a 100 m image, a  $10 \times 10$  patch from the 100 m image covers a single pixel in the 1000 m image. Leveraging this concept, the homogeneity index for the patch or the 1000 m coarse resolution pixel can be computed by dividing the total number of distinct pixel values within the patch by the total number of distinct pixel values in the entire image. The underlying assumption being that homogeneous patches would yield lower values for this index, while heterogeneous patches would result in higher values. This process is repeated using patches extracted from the original 10 m Sentinel-1 SAR image to facilitate inference at the finer 100 m spatial resolution.

In summary, the radar-based RF downscaling model encompassed three key experiments:

1. with VV and VH bands as predictors
2. with  $(5 \times 5)$  VV and  $(5 \times 5)$  VH bands as predictors
3. with  $(5 \times 5)$  VV,  $(5 \times 5)$  VH, and homogeneity index as predictors

Correspondingly, the feature dimension of the first experiment is 2 (VV and VH band), the second experiment is 50 (25 values from both VV and VH bands), and the third experiment is 52 (25 values from both VV and VH bands + 2 homogeneity index features corresponding to VV and VH band).

All the conducted radar-based downscaling experiments were validated both quantitatively and qualitatively using the Landsat-8 LST data collected at 100 m original fine resolution. The corresponding results are displayed in Section 4.1.1.

### 3.2.2 Optical-based RF downscaling model

In order to evaluate the effectiveness of the radar-based RF downscaling model, a comparison was made with an optical-based RF downscaling model. Previous studies, as seen in Section 1.3.3, have demonstrated that many research efforts in downscaling LST rely on predictors derived from optical data. Consequently, the RF downscaling model constructed using optical predictors was employed as a benchmark for comparing the results obtained by the radar-based RF downscaling model. Specifically, six optical bands, namely Red, Blue, Green, SWIR1, SWIR2, and NIR, extracted from the Sentinel-2 MSI data, were utilized as predictors to develop the optical-based RF downscaling model. In this context, Equations 3.1 and 3.2 can be rewritten as:

$$RF_{1000} = f(S2_{1000}, LST_{1000}) \quad (3.10)$$

$$LST_{100} = RF_{1000}(S2_{100}) \quad (3.11)$$

where:

$RF_{1000}$  = random forest regression model trained at 1000 m coarse resolution

$S2_{1000}$  = optical predictors collected from the aggregated 1000 m coarse resolution Sentinel-2 MSI dataset

$LST_{1000}$  = 1000 m aggregated coarse resolution LST from Landsat-8 LST dataset

$S2_{100}$  = optical predictors collected from the aggregated 100 m fine resolution Sentinel-2 MSI dataset

$LST_{100}$  = LST estimated at 100 m fine resolution from the trained  $RF_{1000}$  model

It is important to highlight that the modification introduced in this case pertains to the utilization of a distinct set of predictors. As a result, the remaining equations, specifically 3.7 and 3.8, can also be employed in this scenario, enabling the generation of an optical-based downscaled LST at a resolution of 100 m. The results derived from this particular experiment are provided in Section 4.1.2.

### 3.2.3 Integration of Land Cover information

It is crucial to acknowledge that LST is a complex phenomenon that cannot be adequately modeled using a single source of data like optical or radar. Consequently, in many downscaling studies, researchers strive to integrate multiple sources of information that are correlated with LST to achieve more accurate estimations. In line with this approach, this research also incorporates land cover information as predictors for the radar-based downscaling models. However, instead of adopting the conventional approach, which involves aggregating the land cover image to a coarser resolution and training a model as indicated by Equation 3.1, this research introduces a novel method for incorporating land cover information as a feature. The proposed approach calculates the proportion of land cover classes within a coarse spatial resolution pixel, aiming to provide a more precise representation of the inclusion of land cover information. The primary distinction between the proposed approach and the conventional approach in incorporating land cover information lies in the handling of discretization. In the conventional approach, the estimated LST is discretized due to the discrete nature of the features extracted from the land cover information (i.e., different land cover classes). In contrast, the proposed method transforms these discrete classes into a continuous representation, which is assumed to be comparatively more accurate. Before integrating the proportion of land cover classes as features into both the radar-based and optical-based downscaling models, the performance of this approach is evaluated. A comparison is made against the conventional approach of utilizing land cover information as predictors. The corresponding results of the land cover experiments are presented comprehensively in Section 4.1.3.

The process of calculating the proportion of land cover information is visually depicted in Figure 3.3. As illustrated in Figure 3.3, a  $4 \times 4$  patch extracted from a  $12 \times 12$  land cover image covers or overlaps a single pixel in a  $3 \times 3$  image (Figure 3.3a). In the context of this research, considering the land cover image's spatial resolution of 10 m and the coarse-scale LST's spatial resolution of 1000 m, a  $100 \times 100$  patch extracted from the land cover image would correspond to a single 1000 m coarse resolution LST pixel. Within this patch, the proportion of each land cover class is computed. For example, if 2000 pixels out of the  $100 \times 100$  patch correspond to land cover class A, the proportion of this class is calculated as  $2000/10000 = 0.2$ . Subsequently, this value is utilized as a feature value for the respective coarse resolution pixel. This process is applied to all available land cover classes in the ESA WorldCover v100 image. Furthermore, this process is repeated with a patch size of  $10 \times 10$  to facilitate inference at the finer spatial resolution of 100 m (represented by Figure 3.3b in context of a  $12 \times 12$  image).

## 3.3 CONVOLUTIONAL NEURAL NETWORKS

As seen in Section 1.5, there is a gap in terms of integration of deep learning to the domain of downscaling LST. This gap lead us to establish one of the sub-objectives of this research (1.5.2-3), i.e., to downscale LST using CNN. Our proposed regression-based CNN architecture aims to

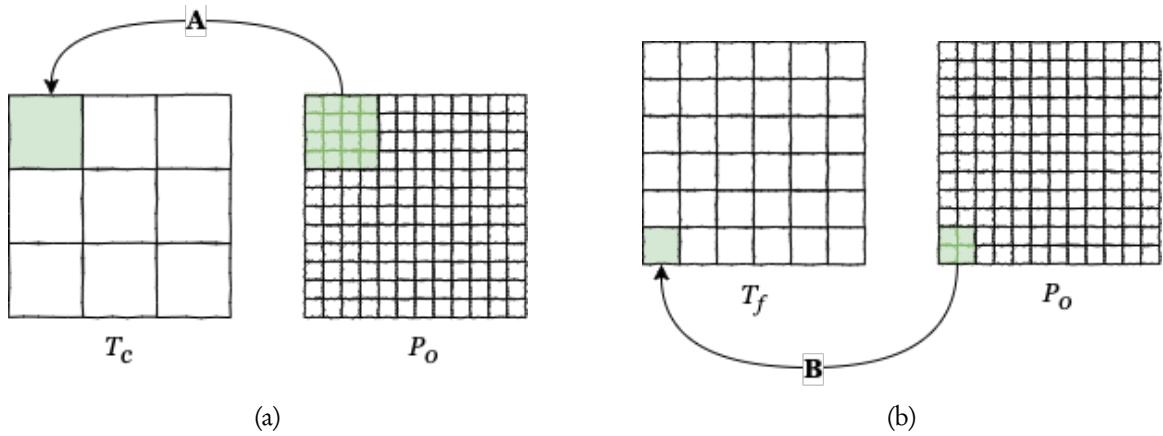


Figure 3.3: Representation of integration of land cover proportion as features to the downscaling models. Here, ‘Connection A’ highlighted in (a) illustrates a pixel-to-patch mapping where  $P_o$  represents the original land cover image collected at 10 m spatial resolution and  $T_c$  refers to the coarse resolution target image. ‘Connection B’ in (b) demonstrates the same concept but at a fine resolution to facilitate inference. Here,  $T_f$  refers to the target image to be estimated by the model during inference

establish a mapping between a coarse resolution target image and fine resolution predictor images. To achieve this, the predictor images derived from radar, land cover, and optical imageries are initially aggregated from their original 10 m resolution to 100 m resolution. Subsequently,  $10 \times 10$  patches, covering the corresponding 1000 m target pixel are generated, and considered as inputs for the model. Once the model is trained, the  $10 \times 10$  patches are generated again from the original 10 m predictor images and fed into the trained model to generate the corresponding 100 m target LST image. Thus, the model receives inputs with a dimension of  $10 \times 10 \times k$ , where  $k$  represents the number of predictors. For each input patch, the output of the model is a singular value representing the downscaled temperature.

The model architecture commences with a stack of convolutional layers. The initial convolutional layer consists of 32 filters of size  $10 \times 10$ , matching the input patch size. The subsequent two convolutional layers employ 32 filters of size  $3 \times 3$  with consistent activation and padding settings. Rectified Linear Unit (ReLU) activation function is employed in each convolutional layer, introducing non-linearity into the architecture. Max pooling is then applied with a pool size of  $2 \times 2$  to downsample the feature maps. After max pooling, a similar pattern is followed by the subsequent block of convolutional layers. The block begins with a convolutional layer containing 64 filters of size  $5 \times 5$ , followed by two convolutional layers with 64 filters of size  $3 \times 3$ . The final block of convolutional layers comprises three convolutional layers with 16 filters of size  $3 \times 3$ . Another round of max pooling is performed to further downsample the feature maps. The output of the convolutional layers is then flattened into a 1-dimensional vector. This vector passes through a series of fully connected (dense) layers. The first dense layer consists of 512 units with the ReLU activation function. Dropout regularization with a rate of 0.2 is applied after this layer. The subsequent two dense layers have 128 and 64 units, respectively, with ReLU activation and dropout layers utilizing the same dropout rate. Finally, the output layer is a dense layer with a single unit, representing the regression prediction. The model is compiled using the Adam optimizer and the Mean Squared Error (MSE) loss function. The learning rate is set to 0.001 and the number of epochs for training is set to 30. Furthermore, to ensure numerical stability in the model, all the

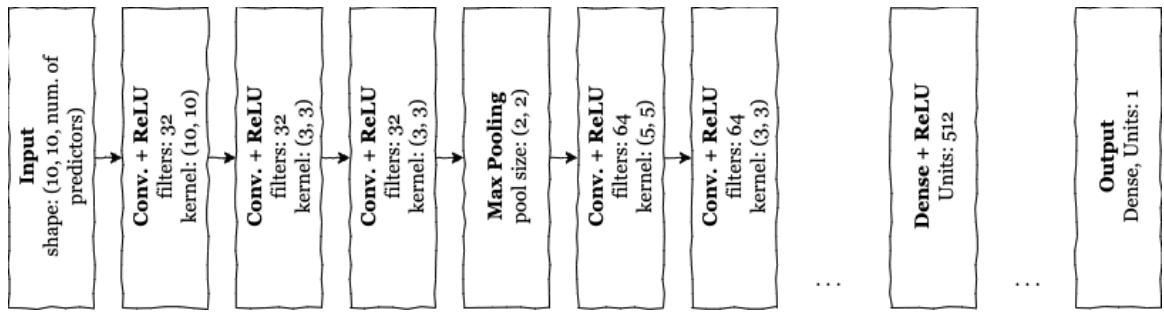


Figure 3.4: proposed CNN-based downscaling architecture

feature and target values were normalized between 0 and 1 using min-max normalization. The model architecture is presented in Figure 3.4.

In summary, the proposed CNN architecture attempts to capture spatial patterns in images through convolutional layers and utilizes dense layers for higher-level abstraction and regression prediction. Figure illustrates the representation of this regression-based CNN architecture for downscaling LST.

Consistent with the RF experiments, the experiments from CNN-based downscaling model utilized SAR (VV and VH) and optical (six bands) images collected from Sentinel-1 GRD and Sentinel-2 MSI datasets, respectively as predictor images for training. Thus, for the CNN-based downscaling model, two main experiments were conducted:

1. with VV and VH images as predictors
2. with six bands of optical images as predictors

The results obtained from the CNN-based downscaling experiments are presented in Section 4.2.

### 3.4 MODIS LST AS A TARGET

As outlined in Sections 3.2 and 3.3, the downscaling models developed in this study employed the aggregated 1000 m Landsat-8 LST as the target variable to generate downscaled LST products at 100 m resolution. However, it is worth noting that this aggregated version of LST, derived from data originally collected at 100 m resolution, does not reflect a practical scenario. To address this concern, all the models and experiments were retrained and conducted using MODIS LST as the target variable. Unlike Landsat-8 LST, MODIS acquires daily LST images at 1000 m spatial resolution, which can be utilized within the developed downscaling framework to generate fine-resolution LST products at 100 m. Nevertheless, a challenge arises in selecting suitable 100 m validation LST data to assess the performance of the models trained using MODIS LST as the target variable. In the case of models and experiments employing the aggregated Landsat-8 LST as the target variable, Landsat-8 LST at 100 m resolution was used for validation. However, using Landsat-8 LST to validate the performance of models trained with MODIS LST can introduce errors, as differences can arise due to distinct sensor characteristics. This challenge or limitation is further discussed in

Section 5.5. The results obtained by utilizing the 1000 m MODIS LST as the target variable are presented in Section 4.3.

### 3.5 EVALUATION FRAMEWORK

The evaluation framework for this research encompasses the utilization of three key metrics: 1) Root Mean Square Error (*RMSE*), 2) Correlation Coefficient (*r*), and 3) Coefficient of Determination (*R*<sup>2</sup>). These metrics serve as comprehensive measures to assess the performance and accuracy of the proposed research methodology. *RMSE* quantifies the average deviation between the predicted values and the ground truth values, providing an insight into the model's overall predictive accuracy. The correlation coefficient (*r*) evaluates the strength and direction of the linear relationship between the predicted and actual values, enabling an assessment of how well the model captures the underlying patterns in the data. Additionally, the coefficient of determination (*R*<sup>2</sup>) offers a measure of the proportion of the variance in the target variable that can be explained by the model, indicating the goodness of fit. Together, these evaluation metrics provide a robust framework to gauge the effectiveness and reliability of the generated downscaled LST products, enabling valuable insights into the performance and predictive capabilities of the proposed downscaling models. In the context of downscaling LST, these metrics can be expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (T'_f - T_v)^2}{n}} \quad (3.12)$$

$$r = \frac{\sum_{i=1}^n (T_v - \bar{T}_v)(T'_f - \bar{T}'_f)}{\sqrt{\sum_{i=1}^n (T_v - \bar{T}_v)^2 \sum_{i=1}^n (T'_f - \bar{T}'_f)^2}} \quad (3.13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (T_v - T'_f)^2}{\sum_{i=1}^n (T_v - \bar{T}_v)^2} \quad (3.14)$$

where:

$T'_f$  = final residual corrected downscaled LST (100 m)

$\bar{T}'_f$  = mean of  $T'_f$

$T_v$  = validation data obtained from Landsat-8 LST (100 m)

$\bar{T}_v$  = mean of  $T_v$

$n$  = number of pixels

It is crucial to acknowledge that the aforementioned metrics provide quantitative insights into the performance of the developed models. As for qualitative assessment, the generated downscaled images were visually evaluated in comparison to the validation images.

## Chapter 4

# Results

This chapter is dedicated to presenting the results obtained from a series of LST downscaling experiments. The chapter consists of three main sections: ‘Random Forest’ (Section 4.1), ‘Convolutional Neural Networks’ (Section 4.2), and ‘Application to MODIS LST’ (Section 4.3). As indicated by their respective titles, the ‘Random Forest’ and ‘Convolutional Neural Networks’ sections showcase the outcomes achieved using RF and CNN as the downscaling models, respectively. The final section, titled ‘Application to MODIS LST’ focuses on the results obtained by employing MODIS LST (1000 m) as the target variable instead of aggregated Landsat-8 LST (1000 m).

### 4.1 RANDOM FOREST

This section presents the results obtained from the experiments conducted using the RF regression model as the primary model for the downscaling process. The section is divided into three subsections, corresponding to the experiments performed with different datasets, namely radar, optical, and land cover.

#### 4.1.1 Radar-based downscaling

As explained in Section 3.2.1, the initial radar-based downscaling experiment considered three main cases. These cases were designed to address the research question associated with sub-objective 1 (1.5.2-1). These cases can be labelled as follows:

- **Case 1:** Utilizing VV and VH bands as predictors.
- **Case 2:** Utilizing  $(5 \times 5)$  VV and  $(5 \times 5)$  VH bands as predictors.
- **Case 3:** Utilizing  $(5 \times 5)$  VV,  $(5 \times 5)$  VH, and homogeneity index as predictors.

Table 4.1 presents the quantitative metrics achieved for downscaling aggregated Landsat-8 LST from a coarse spatial resolution of 1000 m to a finer spatial resolution of 100 m for all these cases. As mentioned earlier, the experiments were repeated for three selected dates of interest (25/03/2020, 10/04/2020, and 28/05/2020), and the metrics were obtained by validating the downscaled LST (100 m) against the original Landsat-8 LST (100 m).

**Table 4.1** Evaluation metrics for ‘Case 1’, ‘Case 2’, and ‘Case 3’ of the radar-based RF downscaling experiments performed on all selected dates of interest

	25/03/2020			10/04/2020			28/05/2020		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
	1	2	3	1	2	3	1	2	3
<i>RMSE</i>	1.44	1.25	1.22	2.10	1.93	1.83	2.76	2.61	2.64
<i>r</i>	0.89	0.92	0.92	0.94	0.95	0.95	0.89	0.91	0.90
<i>R</i> <sup>2</sup>	0.80	0.84	0.85	0.88	0.90	0.91	0.80	0.82	0.82

**Table 4.2** *RMSE* of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 25/03/2020) for all the land cover classes in the study area

	25/03/2020		
	Case 1	Case 2	Case 3
Tree Cover	1.73	1.51	1.43
Shrubland	3.11	3.03	2.89
Grassland	1.14	1.00	0.95
Cropland	1.61	1.42	1.36
Built-Up	1.83	1.62	1.60
Bare/sparse vegetation	2.81	2.58	2.53
Permanent water bodies	1.37	1.09	1.04
Herbaceous wetland	1.93	1.65	1.57

The downscaled LSTs (100 m) obtained for all these cases were quantitatively validated for each land cover class derived from the ‘ESA WorldCover v100’ dataset. The corresponding *RMSE* values for each land cover class in the study area, for the selected dates of interest, are presented in Tables 4.2, 4.3, and 4.4.

After examining Tables 4.1, 4.2, 4.3, and 4.4, it is evident that Case 3, which incorporates the highest number of radar features, outperforms the other cases. Consequently, Figure 4.1 illustrates the downscaled LST images (100 m), the Landsat-8 LST used for validation (100 m), as well as the corresponding histograms and scatterplots pertaining to Case 3. The results obtained for the other cases (i.e., Case 1 and Case 2) are presented in a similar manner in Figures A.1 and A.2, respectively. Another observation from examining these tables is that the downscaling models built for the dates of 10/04/2020 and 28/05/2020 exhibit poorer performance compared to the model for 25/03/2020. This discrepancy could be attributed to the greater temporal difference between the acquisition of LST and radar images for the dates of 10/04/2020 and 28/05/2020 (1 day) compared to that of 25/03/2020 (same day).

#### 4.1.2 Optical-based downscaling

As discussed in Section 3.2.2, the optical-based RF downscaling experiment considered six bands (red, blue, green, SWIR1, SWIR2, and NIR) of the optical image as predictors for the downscaling model. Table 4.5 presents the quantitative metrics achieved for downscaling aggregated Landsat-8 LST from a resolution of 1000 m to 100 m using these six bands as predictors. Additionally, Table 4.6 displays the obtained *RMSE* values for each land cover class in the study area. Unlike the

**Table 4.3** *RMSE* of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 10/04/2020) for all the land cover classes in the study area

	10/04/2020		
	Case 1	Case 2	Case 3
<b>Tree Cover</b>	2.30	2.19	2.01
<b>Shrubland</b>	4.41	4.48	4.10
<b>Grassland</b>	1.67	1.56	1.47
<b>Cropland</b>	2.55	2.18	2.02
<b>Built-Up</b>	2.25	2.22	2.14
<b>Bare/sparse vegetation</b>	3.22	3.14	2.97
<b>Permanent water bodies</b>	2.31	2.01	1.95
<b>Herbaceous wetland</b>	2.88	2.68	2.58

**Table 4.4** *RMSE* of ‘Case 1’, ‘Case 2’, and ‘Case 3’ downscaling experiments (date of interest: 28/05/2020) for all the land cover classes in the study area

	28/05/2020		
	Case 1	Case 2	Case 3
<b>Tree Cover</b>	3.19	3.14	3.04
<b>Shrubland</b>	6.52	6.60	5.37
<b>Grassland</b>	2.35	2.19	2.14
<b>Cropland</b>	3.34	3.32	3.30
<b>Built-Up</b>	2.81	2.76	3.40
<b>Bare/sparse vegetation</b>	4.85	4.47	4.28
<b>Permanent water bodies</b>	2.77	2.45	2.27
<b>Herbaceous wetland</b>	3.39	3.07	2.88

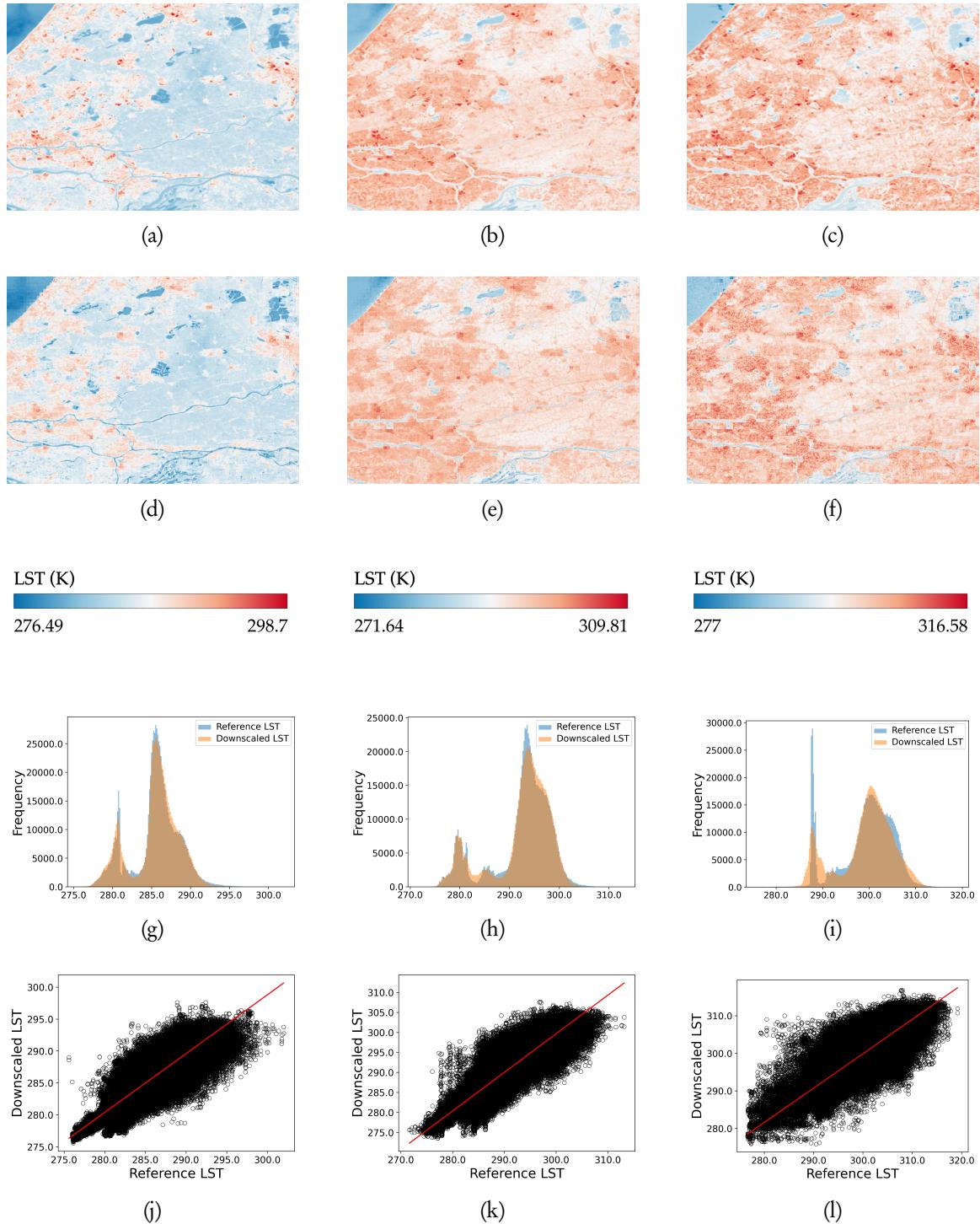


Figure 4.1: Results of the 'Case 3' radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m)

**Table 4.5** Evaluation metrics for the optical-based RF downscaling experiment. Here, six bands from the optical dataset were used as predictors

25/03/2020	
<b>RMSE</b>	1.12
<b>Correlation Coefficient (<math>r</math>)</b>	0.93
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.87

**Table 4.6**  $RMSE$  of the optical-based RF downscaling experiment for each land cover class in the study area. Here, six bands from the optical dataset were used as predictors

25/03/2020	
<b>Tree Cover</b>	1.22
<b>Shrubland</b>	2.10
<b>Grassland</b>	0.95
<b>Cropland</b>	0.96
<b>Built-Up</b>	1.68
<b>Bare/sparse vegetation</b>	1.81
<b>Permanent water bodies</b>	0.92
<b>Herbaceous wetland</b>	1.68

radar-based downscaling experiment, the optical-based downscaling experiment was conducted only for a single date of interest due to the unavailability of cloud-free optical images for the other selected dates.

Furthermore, Figure 4.2 presents the downscaled LST image (100 m), the Landsat-8 LST used for validation (100 m), and the corresponding histograms and scatterplots related to the optical-based RF downscaling experiment.

To address sub-objective 4 (1.5.2-4), a comparison is required between the quantitative metrics of the radar-based downscaling experiment and the optical-based downscaling experiment. Upon reviewing Tables 4.1 and 4.5, it becomes apparent that the radar-based downscaling model exhibits slightly inferior performance compared to the optical-based downscaling model. Furthermore, from Tables 4.2, 4.3, and 4.4, it becomes apparent that the  $RMSE$  values for various land cover classes exhibit similar trends across all radar-based RF downscaling models. The shrubland class consistently yields the highest error, while the grassland class demonstrates the lowest error. This pattern is also observed in the optical-based RF model (Table 4.6); however, the lowest error for this model occurs in the permanent water bodies class. This can be attributed to the fact that LST over water bodies varies smoothly, aligning with the smooth spatial patterns observed in the optical images. Conversely, the absence of this smoothness over water surfaces in radar images could contribute to higher error values. Notably, the radar-based RF downscaling model outperforms the optical-based RF downscaling model for built-up areas, a finding consistent with the CNN-based downscaling model. However, the observed increase in performance is not significant, making it challenging to attribute this characteristic to a specific reason. Section 5.1 discusses various reasons associated with the suboptimal performance of the radar-based downscaling model when compared with the performance of the optical-based downscaling model.

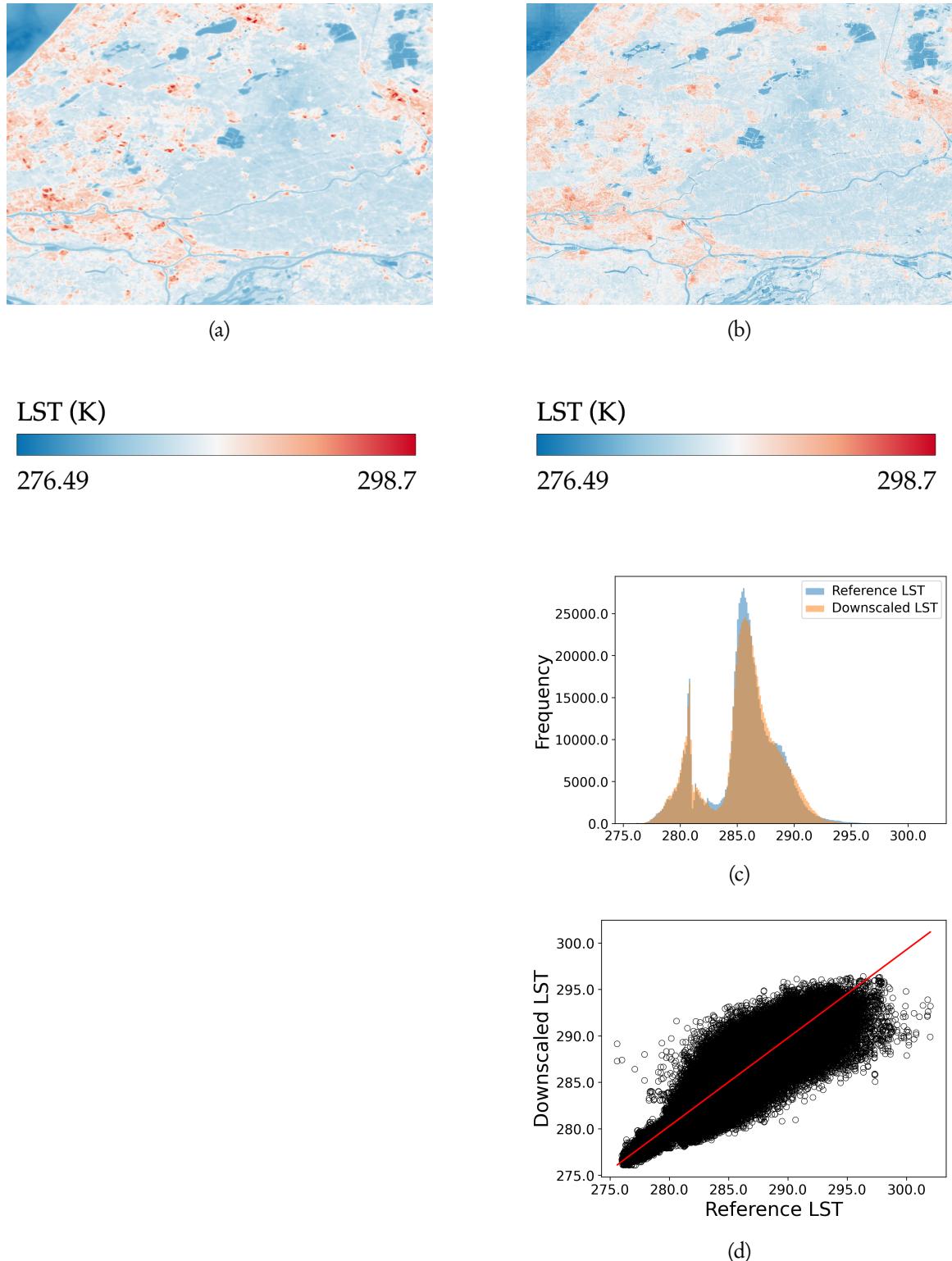


Figure 4.2: Results of the optical-based RF downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b)

**Table 4.7** Evaluation metrics achieved by comparing RF model estimate with the original target LST image at 1000 m. Here, ‘Conventional Approach’ refers to utilizing aggregated or upscaled version of land cover image as a predictor, whereas, ‘Proposed Approach’ refers to utilizing proportion of land cover classes as predictors

	Conventional Approach	Proposed Approach
<b><i>RMSE</i></b>	1.6	0.69
<b>Correlation Coefficient (<i>r</i>)</b>	0.83	0.97

**Table 4.8** Evaluation metrics achieved by integrating land cover information to ‘Case 3’ of the radar-based RF downscaling experiment

	25/03/2020	10/04/2020	28/05/2020
<b><i>RMSE</i></b>	1.20	1.66	2.33
<b><i>r</i></b>	0.93	0.96	0.92
<b><i>R</i><sup>2</sup></b>	0.86	0.93	0.86

#### 4.1.3 Integration of Land Cover information

In order to evaluate the efficacy of the proposed approach of utilizing the proportion of land cover as a feature, compared to the conventional method (described in Section 3.2.3), it is necessary to conduct a preliminary test before integrating the land cover dataset with the developed radar and optical downscaling models (as discussed in subsection 3.2.3). Visualizing and comparing the performance of these predictors at coarse resolution (Equation 3.1) can aid in assessing their effectiveness. The traditional approach employs discrete class labels as predictors, resulting in a discretized representation of the target LST map, which does not accurately reflect the continuous nature of LST variations. Upon reviewing Figure 4.3, it can be concluded that this assumption is indeed valid. Furthermore, the quantitative metrics obtained by comparing the model estimates/predictions at a coarse resolution of 1000 m to the aggregated Landsat-8 LST (1000 m) (Table 4.7) support the finding that the proposed method of incorporating land cover information into the downscaling models outperforms the conventional approach.

Having established the superiority of the proposed approach, the land cover information was integrated with the developed radar-based and optical-based downscaling models. The subsequent sub-subsections present the results obtained for these cases.

##### 4.1.3.1 Incorporating land cover proportion features to radar-based RF downscaling model

As suggested by the title of the sub-subsection, this experiment involves integrating the land cover information into the developed radar-based downscaling model that utilizes  $(5 \times 5)$  VV,  $(5 \times 5)$  VH, and homogeneity index as predictors. The corresponding quantitative results are presented in Tables 4.8 and 4.9. Additionally, Figure 4.4 illustrates the achieved results in a similar manner to Figure 4.1.

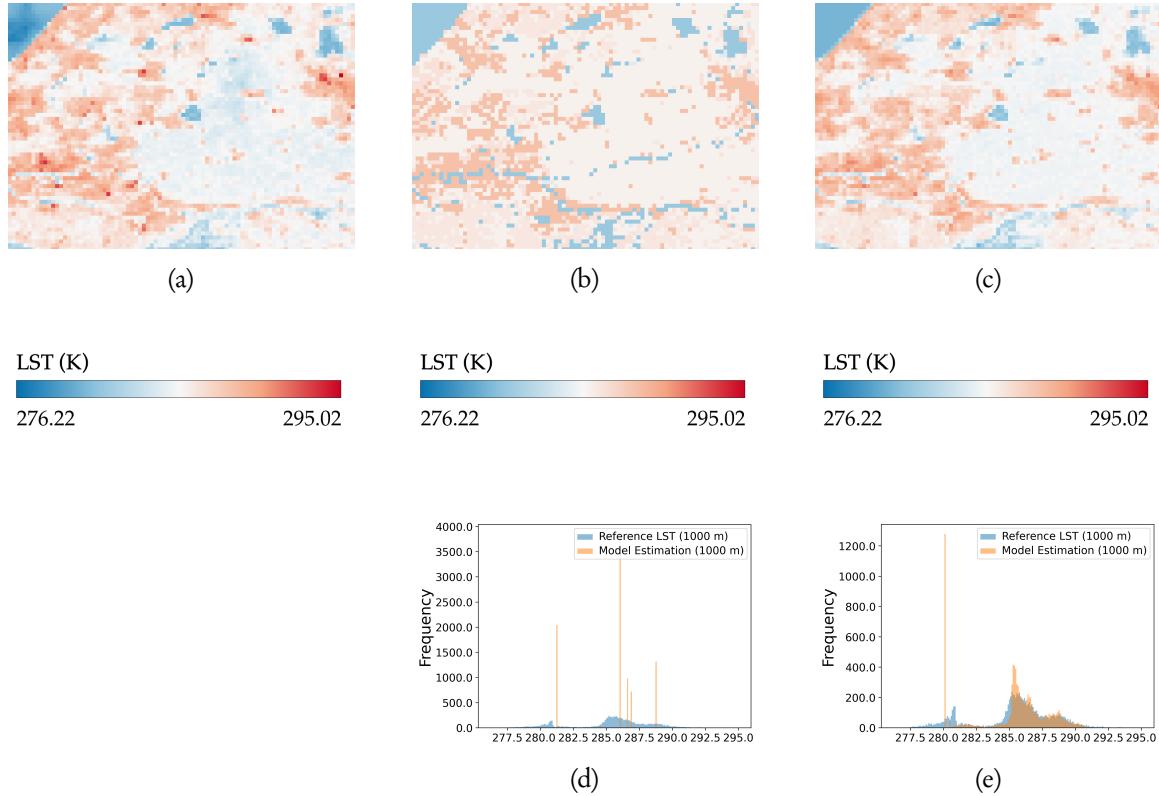


Figure 4.3: Results displaying the efficacy of different land cover integration approaches. Here, (a) refers to the aggregated Landsat-8 LST (1000 m), (b) refers to the RF model estimation (1000 m) with the conventional approach, and (c) refers to the RF model estimation (1000 m) with the proposed approach. (d) and (e) displays the histograms of (b) and (c) in comparison to (a), respectively. Here, ‘Conventional Approach’ refers to utilizing aggregated or upscaled version of land cover image as a predictor, whereas, ‘Proposed Approach’ refers to utilizing proportion of land cover classes as predictors

**Table 4.9** *RMSE* achieved by integrating land cover information to ‘Case 3’ of the radar-based RF downscaling experiment for all the land cover classes in the study area

	25/03/2020	10/04/2020	28/05/2020
<b>Tree Cover</b>	1.20	1.65	3.10
<b>Shrubland</b>	2.30	2.49	3.07
<b>Grassland</b>	0.96	1.39	1.98
<b>Cropland</b>	1.05	1.66	2.92
<b>Built-Up</b>	2.00	2.22	2.86
<b>Bare/sparse vegetation</b>	2.08	2.76	3.39
<b>Permanent water bodies</b>	1.04	1.70	1.68
<b>Herbaceous wetland</b>	1.35	1.88	2.16

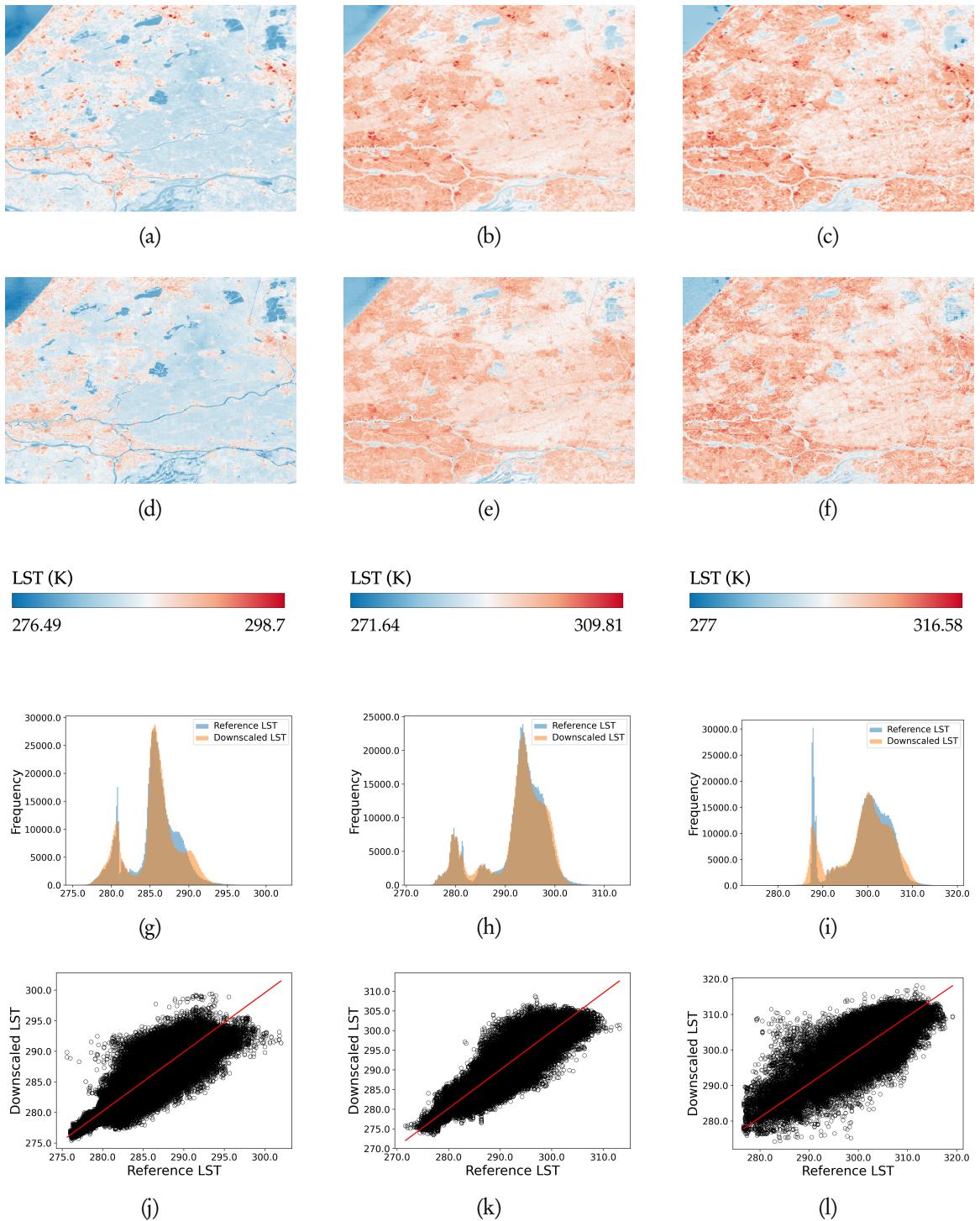


Figure 4.4: Results achieved by integrating land cover information to 'Case 3' of the radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m)

**Table 4.10** Evaluation metrics achieved by integrating land cover information to the optical-based RF downscaling experiment

	25/03/2020
<b>RMSE</b>	1.10
<b>Correlation Coefficient (<math>r</math>)</b>	0.94
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.88

**Table 4.11**  $RMSE$  achieved by integrating land cover information to the optical-based RF downscaling experiment for all the land cover classes in the study area

	25/03/2020
<b>Tree Cover</b>	1.20
<b>Shrubland</b>	1.89
<b>Grassland</b>	0.93
<b>Cropland</b>	0.80
<b>Built-Up</b>	1.78
<b>Bare/sparse vegetation</b>	1.53
<b>Permanent water bodies</b>	0.97
<b>Herbaceous wetland</b>	1.28

#### 4.1.3.2 Incorporating land cover proportion features to optical-based RF downscaling model

In this case, the land cover information was integrated into the developed optical-based downscaling model that utilizes six optical bands as predictors. The corresponding quantitative metrics are presented in Table 4.10. Furthermore, Figure 4.5 demonstrates the achieved results in a similar manner to Figure 4.2.

After comparing Table 4.8 with 4.1 and Table 4.10 with 4.5, it becomes apparent that integrating land cover information into both the radar-based and optical-based downscaling models leads to improved performance. The specific advantages offered by the incorporation of land cover information are further discussed in Section 5.3.

## 4.2 CONVOLUTIONAL NEURAL NETWORKS

This section presents the results obtained from the downscaling experiments conducted using the proposed CNN-based downscaling architecture. The section is further divided into two subsections corresponding to the radar-based and optical-based downscaling.

### 4.2.1 Radar-based Downscaling

In the primary radar-based CNN downscaling experiment, the VV and VH bands from the radar imagery were utilized as predictors. The achieved quantitative metrics are presented in Table 4.12. Additionally, Figure 4.6 illustrates the achieved results in a similar manner to Figure 4.1.

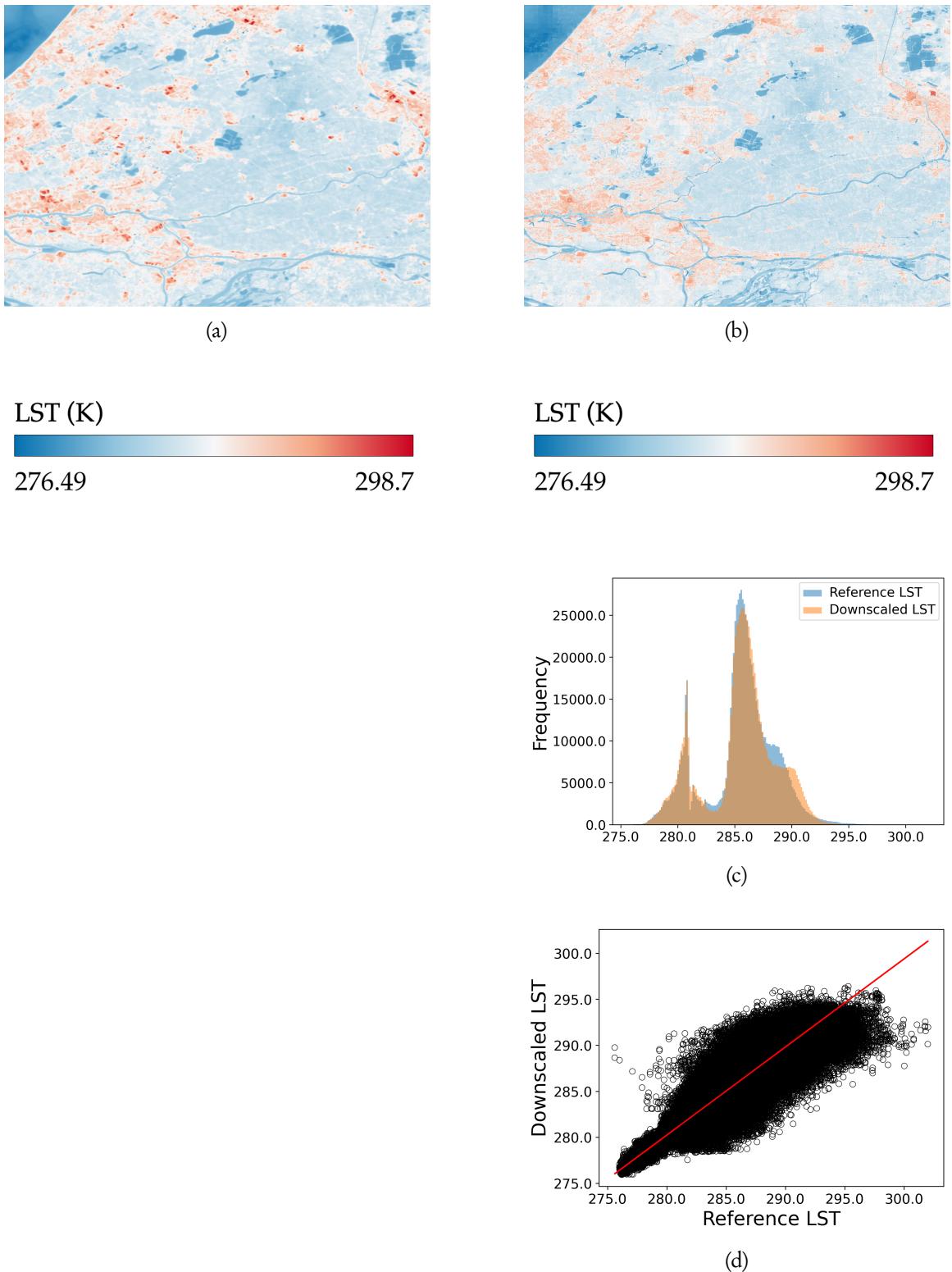


Figure 4.5: Results achieved by integrating land cover information to the optical-based RF downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b)

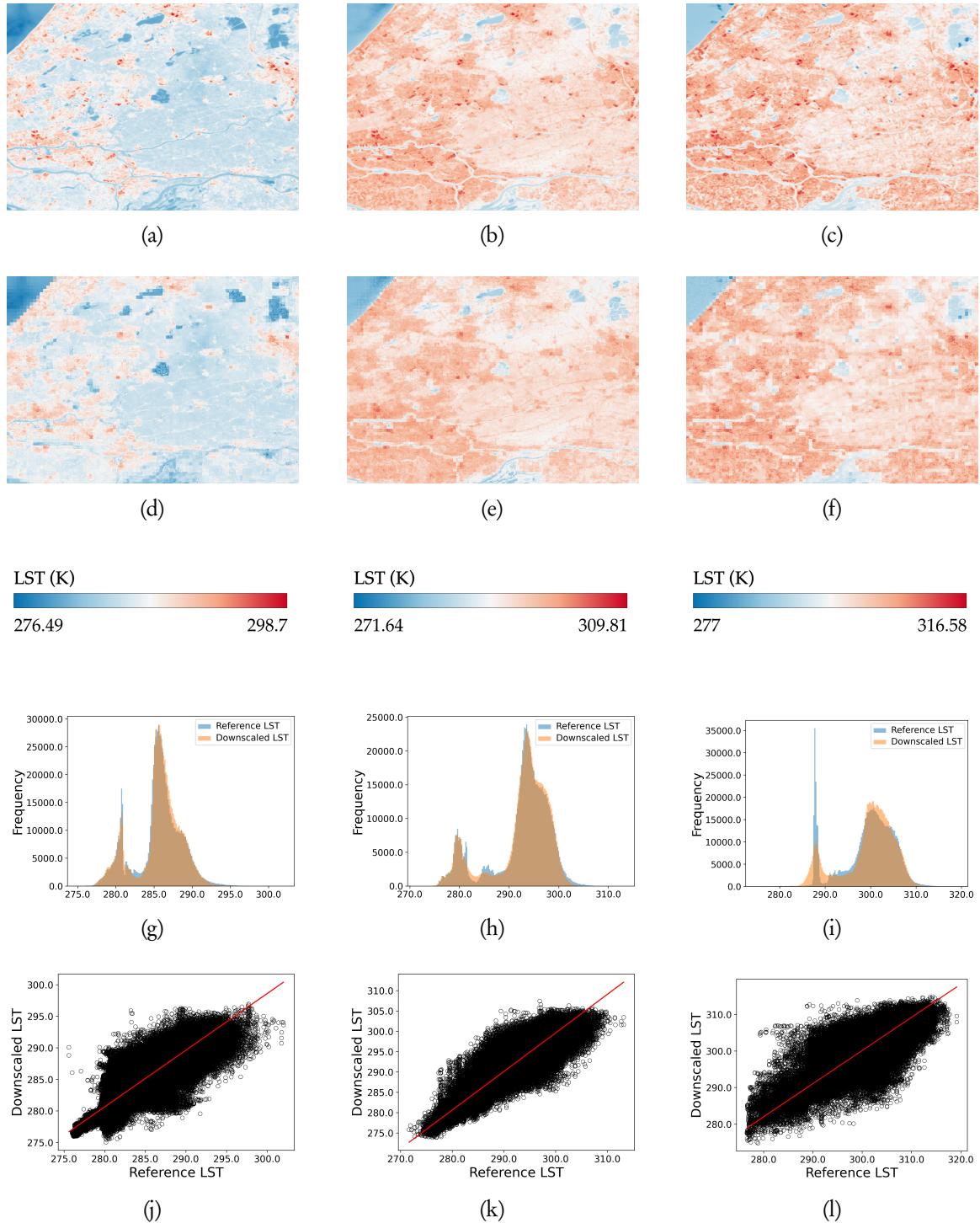


Figure 4.6: Results of the radar-based CNN downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m)

**Table 4.12** Evaluation metrics for the radar-based CNN downscaling experiment. Here, VV and VH images from the radar dataset were utilized as predictors for the CNN downscaling model

	25/03/2020	10/04/2020	28/05/2020
<b><math>RMSE</math></b>	1.21	1.68	2.33
<b><math>r</math></b>	0.92	0.96	0.92
<b><math>R^2</math></b>	0.85	0.93	0.86

**Table 4.13** Evaluation metrics for the optical-based CNN downscaling experiment. Here, six bands from the optical dataset were utilized as predictors for the CNN downscaling model

	25/03/2020
<b><math>RMSE</math></b>	1.09
<b>Correlation Coefficient (<math>r</math>)</b>	0.94
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.88

Upon comparing the quantitative metrics from Tables 4.1 and 4.12, it is evident that the CNN models outperform the RF model. However, upon analysing the downscaled LST obtained through the proposed CNN architecture without residual correction, certain shortcomings related to the proposed architecture are revealed. These shortcomings are discussed in Section 5.4.

#### 4.2.2 Optical-based Downscaling

In the primary optical-based CNN downscaling experiment, six optical bands were utilized as predictors. The achieved quantitative metrics are presented in Table 4.13. Additionally, Figure 4.7 illustrates the achieved results in a similar manner to Figure 4.2.

Similar to the observation made for the radar-based CNN downscaling experiment, the CNN model outperforms the RF model in the optical-based downscaling experiment (Tables 4.13 and 4.5).

### 4.3 APPLICATION TO MODIS LST

As mentioned in Section 3.4, all the experiments were conducted again utilizing MODIS LST (1000 m) as the target variable to demonstrate the applicability of the developed methods in practical scenarios. The achieved results exhibited similar trends to those obtained using aggregated Landsat-8 LST (1000 m) as the target variable and have not been included here to avoid redundancy. However, to illustrate for the reader, the example results achieved for two RF-based MODIS LST downscaling experiments, namely utilizing  $(5 \times 5)$  VV,  $(5 \times 5)$  VH, homogeneity index, and land cover as predictors, and utilizing six bands of optical imagery and land cover as predictors, are displayed in Figure 4.8 and presented in Table 4.14.

As can be seen in Table 4.14, utilizing MODIS LST (1000 m) as the target variable yields identical quantitative performance for radar-based and optical-based experiments. This contrasts with the findings reported in Tables 4.8 and 4.10, where a noticeable difference in performance is evident

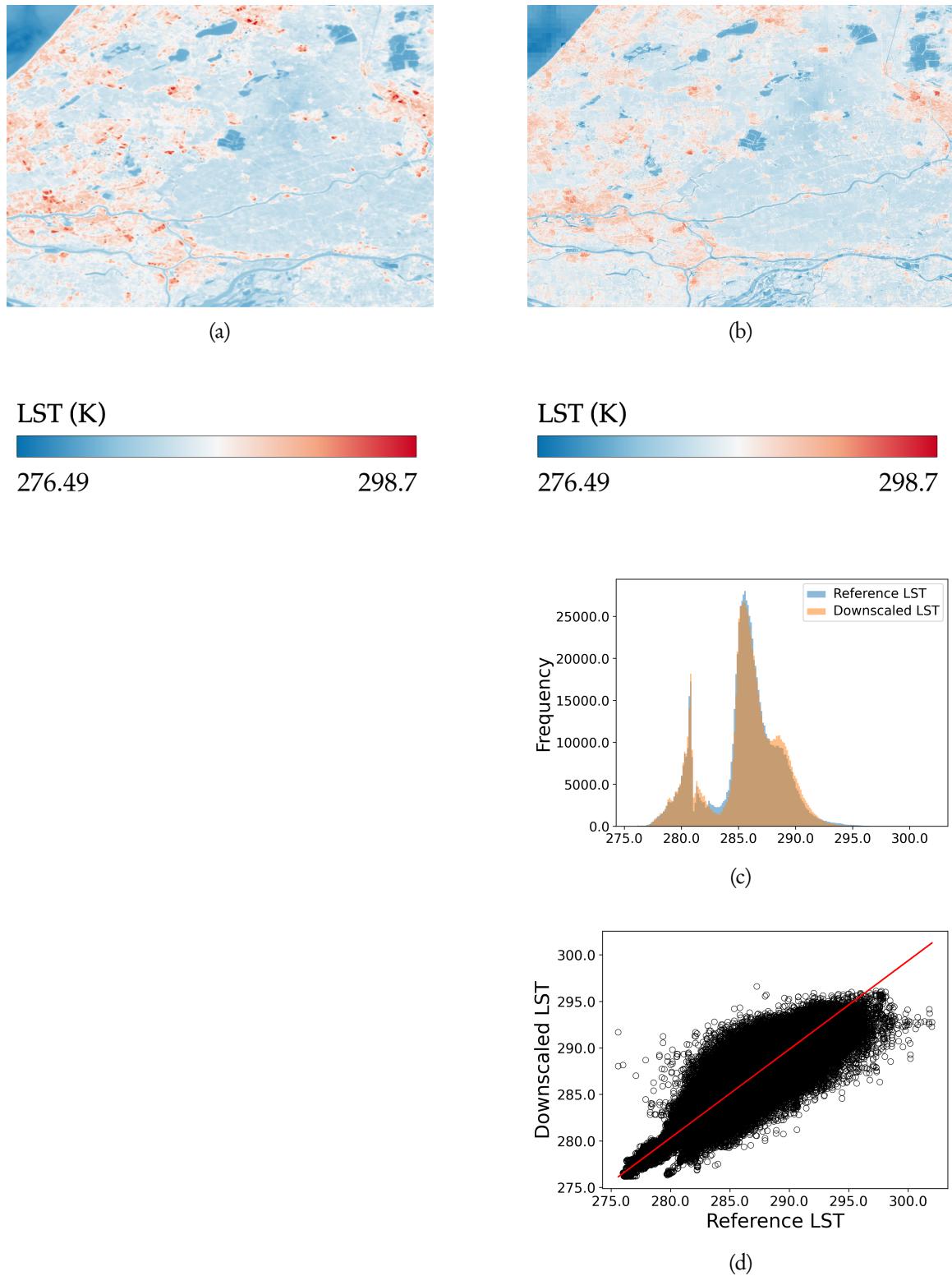


Figure 4.7: Results of the optical-based CNN downscaling experiment. Here, (a) refers to Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m), (c) refers to the histograms of (a) and (b), and (d) refers to the scatterplots between (a) and (b)

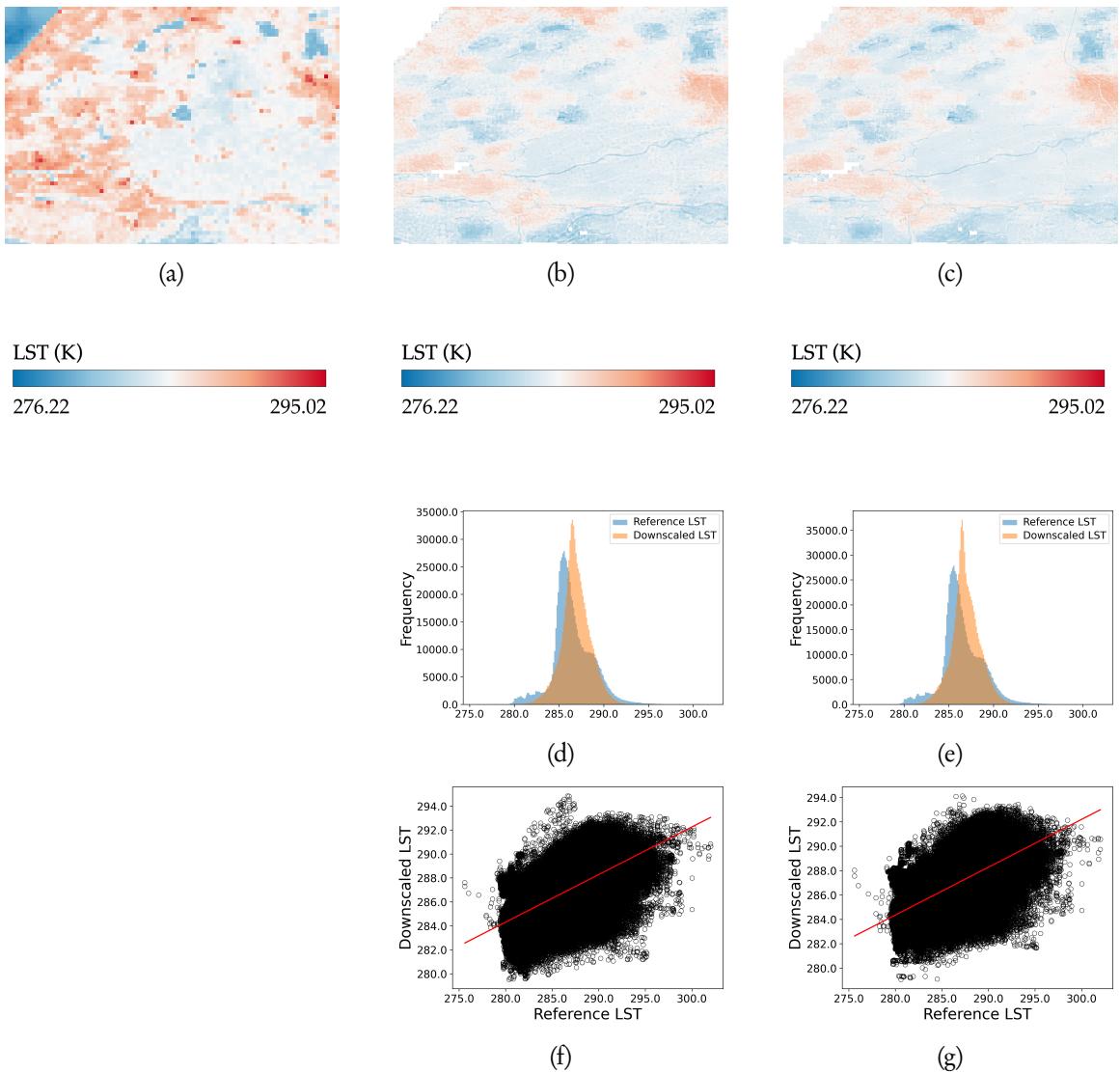


Figure 4.8: Results of the downscaling experiments conducted using MODIS LST as the target variable. Here, (a) refers to the Landsat-8 validation LST (100 m) (acquisition date: 25/03/2020), (b) refers to the achieved downscaled LST (100 m) using  $(5 \times 5)$  VV,  $(5 \times 5)$  VH, homogeneity index, and land cover as predictors, and (c) refers to the achieved downscaled LST (100 m) using six optical bands as predictors. (d) and (e) displays the histograms of (b) and (c) in comparison to (a), respectively, whereas, (f) and (g) displays the scatterplot comparision between the downscaled LST (100 m) and the Landsat-8 validation LST (100 m)

**Table 4.14** Evaluation metrics for the downscaling experiments conducted with MODIS LST as the target variable

	with (5 × 5) VV, (5 × 5) VH, homogene- ity index, and land cover as predictors	with six optical bands and land cover as predictors
<b>RMSE</b>	1.92	1.92
<b>Correlation Coefficient <math>r</math></b>	0.56	0.56
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.31	0.31

between the radar-based and optical-based downscaling experiments. Moreover, the quantitative performance achieved by performing the same experiments utilizing aggregated Landsat-8 LST as target (Tables 4.8 and 4.10) demonstrates superiority compared to that shown in Table 4.14. It is important to emphasize that the predictors and the downscaling model remain the same for the experiments presented in Tables 4.8, 4.10, and 4.14. The decline in performance when using MODIS LST (1000 m) as the target variable can be attributed to the differences in sensor characteristics between MODIS and Landsat-8. Specifically, since the sensor characteristics are distinct, it may be erroneous to directly compare the performance of downscaled MODIS LST (100 m) to Landsat-8 LST (100 m). Section 5.5 delves further into the details of this matter.

## Chapter 5

# Discussion

The objective of this chapter is to offer insights into both the advantages and shortcomings of the overall research and propose potential avenues to pursue in order to achieve enhanced performance. The insights presented in Section 5.1 focus specifically on radar data in general, while Section 5.2 provides insights into the effectiveness of the innovative features derived from the radar dataset. Section 5.3 highlights the benefits attained by incorporating land cover information into the downscaling model. Furthermore, Section 5.4 is dedicated to discussing the limitations of the proposed CNN-based downscaling model and suggests potential modifications to its architecture to improve performance. The discrepancy between MODIS and Landsat-8 LST products is examined in Section 5.5, and finally, Section 5.6 addresses the challenge of modeling the temporal aspect in regression-based downscaling frameworks.

### 5.1 EFFICACY OF RADAR DATA IN LST DOWNSCALING

The primary objective of this research was to evaluate the efficacy of radar-derived predictors in downscaling LST (1.5.1). The quantitative evaluation metrics presented in Table 4.1 demonstrate the effectiveness of using the VV and VH bands of Sentinel-1 GRD data as predictors for LST downscaling. The downscaled LST maps generated using the radar-based RF downscaling model exhibit favorable agreement with the validation data, as depicted in Figure 4.1. This observation remains consistent in the case of a CNN-based downscaling model, as evidenced by the results presented in Table 4.12 and Figure 4.6. When comparing the performance of the predictors derived from the Sentinel-1 radar dataset (Tables 4.1 and 4.12) with those derived from the Sentinel-2 optical dataset (Tables 4.5 and 4.13), it becomes evident that the downscaling models based on optical data display a slightly superior performance when compared to their radar-based counterparts.

Several factors may contribute to the optical dataset outperforming the radar dataset. Firstly, radar images typically display more pixel value variations over local regions, whereas optical bands exhibit smoother variations in pixel values (Figure 5.1). Although efforts were made to address this issue by incorporating neighboring values of radar bands as features, datasets with inherent smooth patterns generally yield better results compared to those without such patterns. Additionally, as can be inferred from Figure 5.1, the boundary delineation observed in the pixel values of optical images between different features in the study area, such as urban and green areas, aligns with the spatial patterns of observed differences in LST values; however, the same cannot be observed for radar images. Another characteristic that may contribute to the slightly better performance of the optical dataset is the difference in sensor characteristics between radar and optical data, particularly the disparity in viewing geometry. Sentinel-1 SAR satellites collect images using a

side-looking geometry, whereas both Landsat-8 LST and Sentinel-2 optical data are acquired using a nadir-looking geometry, which may introduce some spatial pattern mismatches and limit the performance of the radar dataset for downscaling.

Furthermore, as highlighted in Section 1.3.3, LST downscaling studies commonly employ spectral indices, such as NDVI derived from optical datasets as predictors, rather than individual spectral bands. It should be noted that similar spectral indices can also be calculated using radar datasets, such as Sentinel-1. For instance, Mandal et al. (2020) derived a dual-pol Radar Vegetation Index (dpRVI) for the Sentinel-1 SAR dataset, and observed a strong correlation with biophysical variables such as Dry Biomass (DB), Plant Area Index (PAI), and Vegetation Water Content (VWC). However, the calculation of such indices relies on the covariance matrix, which can only be computed from the Sentinel-1 SLC dataset. In this study, the Sentinel-1 GRD dataset was employed, which lacks phase information, thus preventing the calculation of these indices for use as features. Additionally, notable attempts have been made, such as those by Bhogapurapu et al. (2022), to compute the Radar Vegetation Index (RVI) from Sentinel-1 GRD data for soil moisture estimation. Incorporating such information in the developed downscaling models could potentially lead to improved results. Moreover, since the Sentinel-1 satellite is dual-pol, it only offers VV and VH polarization channels. The inclusion of quad-pol datasets, providing four polarization channels (VV, VH, HH, and HV), may further enhance the outcomes.

## 5.2 EFFICACY OF ENGINEERED RADAR FEATURES IN LST DOWNSCALING

As seen through Tables 4.1 and 4.8, the incorporation of feature engineering techniques, such as including neighboring values and radar-based homogeneity index enhances the quantitative performance of the radar-based RF downscaling models and yields improved qualitative results. Figure 5.2 depicts the impact of feature inclusion within the radar-based downscaling framework. From Figure 5.2, it can be inferred that the downscaled LST image (100 m), generated by incorporating neighboring values (Figures 5.2d and 5.2e), exhibits smoother and more gradual variations compared to the version without such inclusion (Figure 5.2c). However, the difference observed in the downscaled LST, generated by including the radar-based homogeneity index (Figure 5.2e), compared to the version without this inclusion (Figure 5.2d), is not significantly substantial. As discussed before, the feature dimension increases from 2 (VV and VH bands) to 50 ( $5 \times 5$  VV,  $5 \times 5$  VH) between Figures 5.2c and 5.2d, and subsequently from 50 ( $5 \times 5$  VV,  $5 \times 5$  VH) to 52 ( $5 \times 5$  VV,  $5 \times 5$  VH, VV homogeneity index, and VH homogeneity index) between Figures 5.2d and 5.2e. Hence, the performance increase between Figures 5.2c and 5.2d is expected to be more significant compared to that between Figures 5.2d and 5.2e. This qualitative observation is consistent with the findings presented in Table 4.1, where the performance increase between ‘Case 1’ and ‘Case 2’ is more substantial than that between ‘Case 2’ and ‘Case 3’. Furthermore, the inclusion of neighboring values and radar-based homogeneity index also mitigates numerous false high LST estimates that may arise from elevated values of the backscattering coefficient. It is important to note that high backscatter values do not necessarily correspond to high LST values, as the raw radar backscatter is influenced by various variables beyond LST alone.

Although the quantitative performance gains achieved through the inclusion of these features in the models are not substantial, qualitatively, these engineered features effectively address several important challenges.

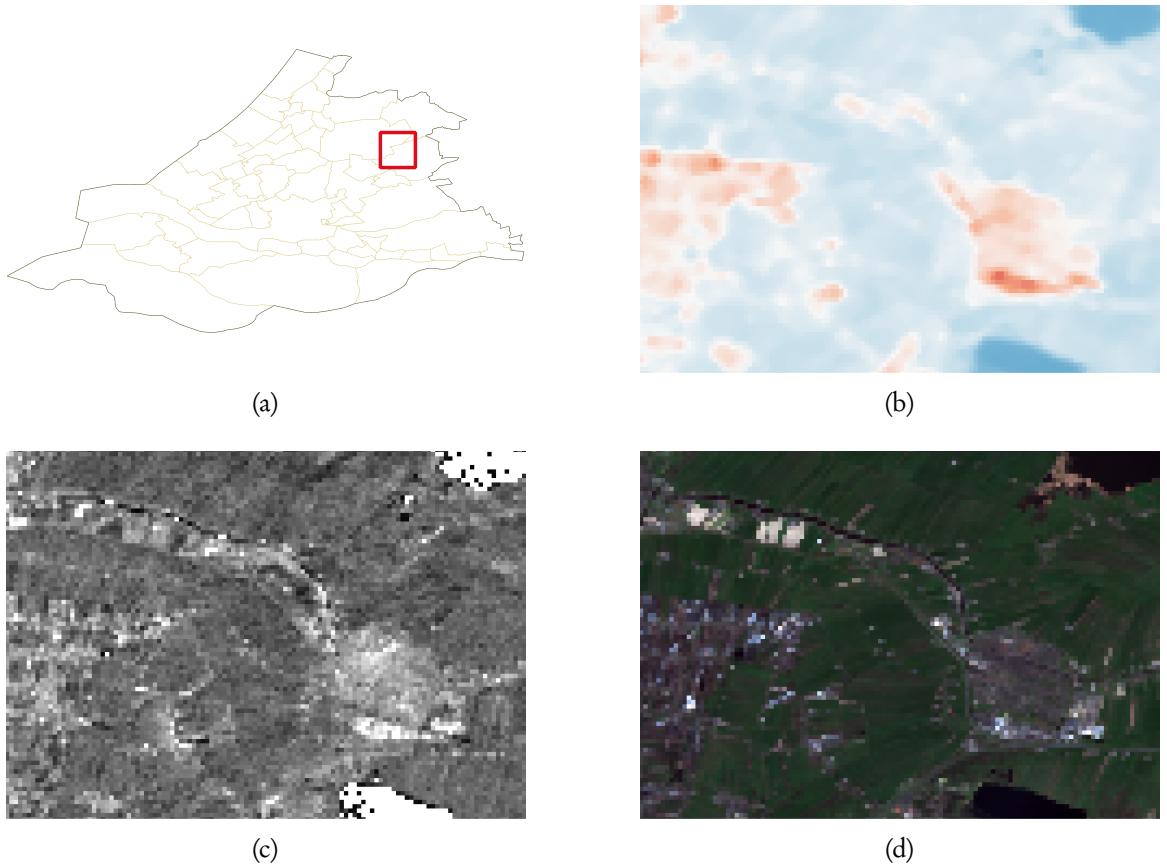


Figure 5.1: Comparison between the spatial patterns exhibited by different datasets. Here, the red box in (a) shows the location of the zoomed-in images. (b) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020, (c) refers to the VV band image of Sentinel-1 GRD dataset (100 m) acquired on 25/03/2020, and (d) refers to the optical true color composite of Sentinel-2 MSI dataset (100 m) acquired on 26/03/2020. As can be inferred through the images, in terms of spatial patterns, (d) shows a better agreement with (b) as compared to (c). Moreover, (c) demonstrates a higher presence of pixel value variations over local regions, such as homogeneous green areas, in contrast to (d).

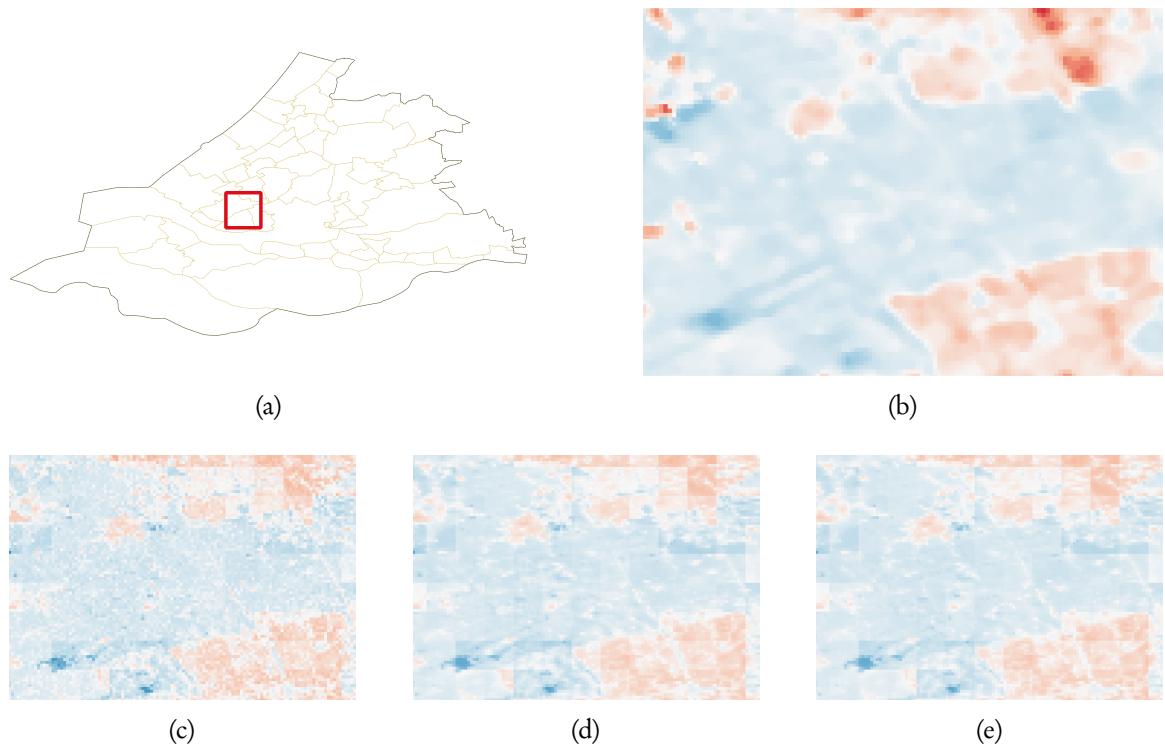


Figure 5.2: Comparison between the downscaled LST images (100 m) obtained by different feature inclusion cases. Here, the red box in (a) shows the location of the zoomed-in images. (b) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020, (c) refers to the downscaled LST (100 m) obtained by utilizing VV and VH bands as predictors, (d) refers to the downscaled LST (100 m) obtained by utilizing (5 × 5) VV and (5 × 5) VH patches as predictors, and (e) refers to the downscaled LST (100 m) obtained by utilizing (5 × 5) VV patch, (5 × 5) VH patch, and homogeneity index as predictors

**Table 5.1** Quantitative metrics without and with residual correction process for radar-based RF downscaling experiments (date of interest: 25/03/2020)

	Case 1		Case 2		Case 3	
	without	with	without	with	without	with
	$\Delta T$					
<b>RMSE</b>	1.80	1.44	1.63	1.25	1.60	1.22
<b>Correlation Coefficient (<math>r</math>)</b>	0.82	0.89	0.85	0.92	0.86	0.92
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.67	0.80	0.72	0.84	0.73	0.85

It is worth noting that the residual correction process, as described by Equations 3.3 and 3.4, effectively addresses certain challenges mentioned above. However, it is important to acknowledge that the application of this process can introduce boxy patterns in regions where the model's predictive capability is limited. These boxy patterns are also visible in Figure 5.2. This occurs because the residuals are obtained at a coarser scale, and the corresponding fine-scale pixels are adjusted using a constant residual value. The significance of this issue varies depending on the practical use of these downscaled products. Consequently, it is generally preferable to enhance predictive power through the incorporation of additional features rather than relying heavily on the residual correction method. Table 5.1 offers insights into the performance improvement achieved through the residual correction process.

### 5.3 INTEGRATION OF LAND COVER INFORMATION

In radar imagery, urban regions typically exhibit high backscatter values due to complex scattering interactions, while green areas tend to show lower backscatter values. However, it should be noted that high backscatter values do not necessarily correspond to high LST values, as the raw radar backscatter is influenced by various variables beyond LST alone. Consequently, during the training process, as most of the high backscattering values originate from urban regions, they tend to be associated with high LST values. This situation introduces errors when dealing with ground features that exhibit high backscatter values but do not align with high LST values, as is the case with forested areas.

Forested areas typically demonstrate high backscatter values due to various scattering mechanisms, while their LST values remain low. To address this issue, integrating land cover information into the downscaling models becomes necessary. Figure 5.3 illustrates the impact of incorporating land cover information within the proposed framework. As depicted in Figure 5.3, the inclusion of land cover features successfully resolves the issue related to forested areas (compare Figure 5.3d with Figure 5.3e and Figure 5.3f with Figure 5.3g). Additionally, incorporating land cover features aids in better boundary delineation between different land cover categories, addressing a limitation observed when using predictors derived solely from Sentinel-1 GRD data (VV and VH). Furthermore, land cover information is relatively static and independent of weather conditions. Therefore, integrating land cover information into downscaling models does not pose significant processing challenges.

Comparing Figure 5.3f with Figure 5.3d, it can be observed that residual correction attempts to solve this issue of false mapping without the incorporation of land cover information. However, since the predictive capacity of the model is low the area looks more boxy as compared to Figure 5.3g.

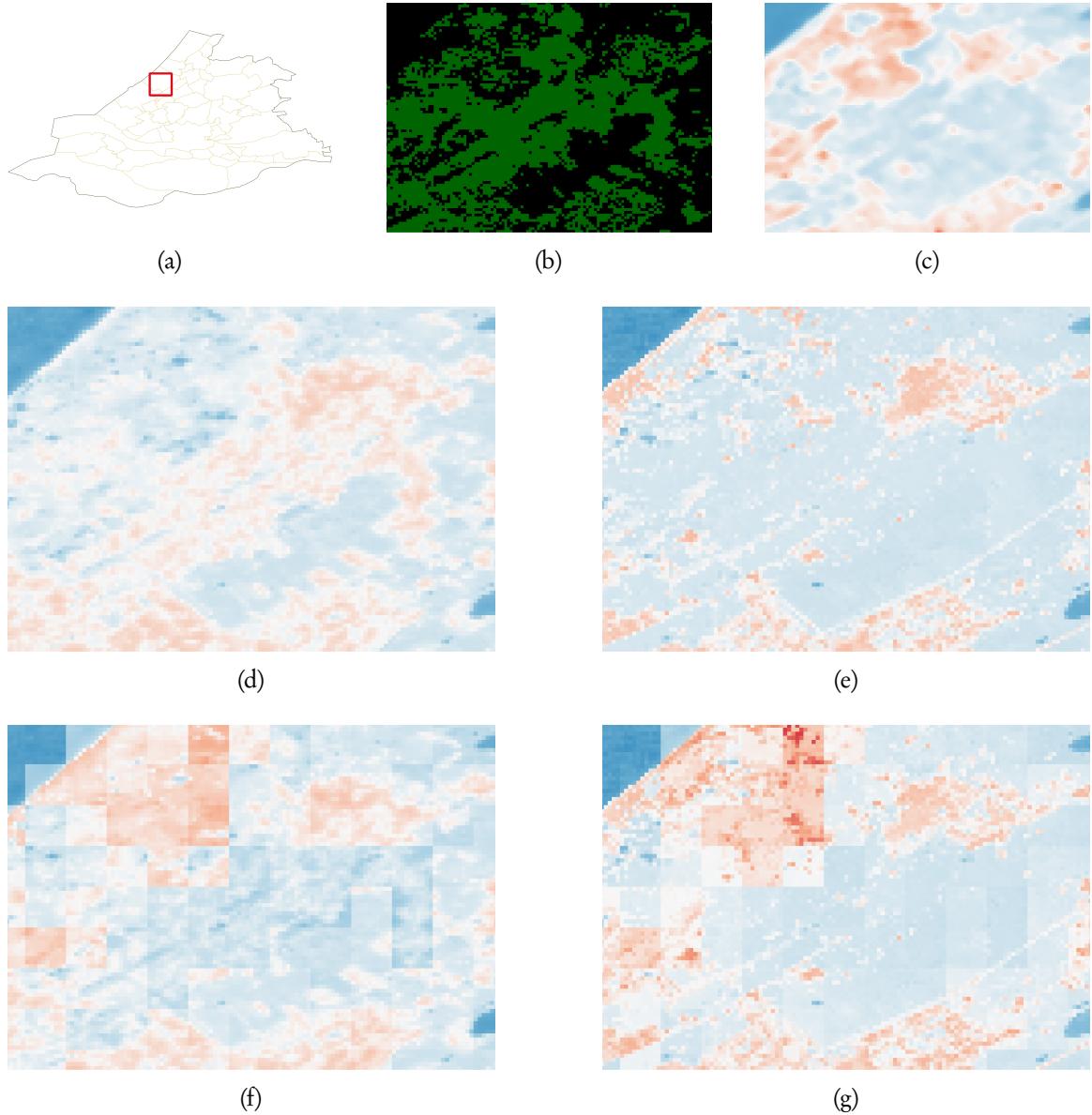


Figure 5.3: Impact of incorporating land cover information into the downscaling framework. Here, the red box in (a) shows the location of the zoomed-in images. (b) shows the tree cover class from the ‘ESA WorldCover’ dataset (green color) and (c) refers to the Landsat-8 LST (100 m) acquired on 25/03/2020. (d) and (f) refer to the achieved downscaled LST (100 m) by utilizing  $(5 \times 5)$  VV patch,  $(5 \times 5)$  VH patch, and homogeneity index as predictors, whereas (e) and (g) refer to the achieved downscaled LST (100 m) by utilizing  $(5 \times 5)$  VV patch,  $(5 \times 5)$  VH patch, homogeneity index, and land cover as features. Here, (d) and (e) are achieved downscaled LST images without residual correction, whereas, (f) and (g) with residual correction

**Table 5.2** *RMSE* values of radar (VV and VH) and optical (six bands) based RF and CNN downscaling experiments before and after residual correction

	RF		CNN	
	without	with	without	with
	$\Delta T$	$\Delta T$	$\Delta T$	$\Delta T$
with VV and VH as predictors	1.80	1.44	2.93	1.21
with six bands of optical dataset as predictors	1.39	1.12	1.35	1.09

#### 5.4 CNN: SHORTCOMINGS AND POTENTIAL ARCHITECTURE MODIFICATIONS

In addition to the conventional RF downscaling model, this research also introduced a CNN-based regression architecture for LST downscaling. The primary objective was to establish an end-to-end mapping between coarse-scale target values and fine-scale predictor values. As indicated by Tables 4.5 and 4.13, the CNN-based architecture exhibits slightly better quantitative performance compared to the random forest model when utilizing the same optical input. However, the difference in performance is not substantial. From a qualitative standpoint, the CNN-based downscaling model preserves the structural characteristics of features within the study area, in contrast to the RF-based model. Consequently, there are minimal drastic variations observed in the CNN-based downscaled LST values across local regions (Figure 4.7b). A noteworthy observation is that when using VV and VH bands as predictors without residual correction, the evaluation metrics for the CNN-based downscaling model are inferior to those of the RF-based downscaling model (Table 5.2). For optical data, before applying the residual correction, the evaluation metrics for both RF and CNN downscaling models exhibit similarities, with CNN displaying slightly superior performance compared to RF. This discrepancy might be attributed, in part, to the fact that the feature inputs for the radar-based RF downscaling model, i.e., the neighbors and the homogeneity index do not correlate with the features learned by the convolution layers. Consequently, it may be erroneous to directly compare these models. However, this discrepancy could also be attributed to the fact that incorporation of any images that exhibit noisy spatial patterns like the radar dataset would always result in bad estimates due to the nature of the developed architecture. Specifically, the proposed CNN architecture attempts to map the intrinsic fine structure obtained from the predictor image to the value of corresponding coarse resolution target pixel. Consequently, inputs lacking a smoothly varying pattern can pose challenges for the proposed CNN downscaling architecture in identifying appropriate features. If this hypothesis holds true, it highlights a limitation in the developed architecture and suggests the need for modifications.

One potential modification could involve increasing the number of convolutional layers in the architecture. This adjustment could enhance the architecture's ability to identify intricate features from a complex data source like radar. However, it is worth noting that the proposed architecture takes a  $10 \times 10$  patch of predictor images as input. Consequently, utilizing a deeper network might prove redundant since there may not be sufficient pixels remaining to convolve after pooling.

Another potential modification could involve integrating a fully connected network that functions as a traditional downscaling regression algorithm, where the inputs consist of the coarse-resolution predictors and target images, akin to the random forest regression algorithm. Combining this approach with the proposed CNN-based regression architecture, which excels at preserving structures, in some form of a weighted output may yield improved results. However, definitive conclusions cannot be drawn without conducting further experimentation.

The primary objective behind employing a CNN architecture was to enable the model to learn and understand the spatial relationships present in the data. However, it is worth noting that the proposed architecture, as it stands, does not explicitly consider the spatial structure of the target variable, namely LST. In other words, there is no direct mapping between the spatial structure of the predictor image and the spatial structure of the target image. As a consequence, the performance of the proposed CNN architecture may be limited in capturing the intricate spatial dependencies crucial for accurate LST downscaling. In this context, a fully convolutional neural network architecture can be adapted alongside the proposed CNN architecture to learn the spatial relationship between predictor and target images.

By considering the above mentioned modifications, future iterations of the proposed CNN architecture can potentially achieve superior results for the task of LST downscaling.

## 5.5 APPLICATION TO MODIS LST: SHORTCOMINGS AND A POTENTIAL CORRECTION PROCEDURE

According to the data presented in Table 4.14, the performance of the developed RF downscaling framework on MODIS LST (1000 m) was found to be inferior compared to the results obtained from aggregated Landsat-8 LST (1000 m) (Tables 4.8 and 4.10). As discussed in Section 4.3, this difference can be partly attributed to the distinct sensor characteristics of MODIS and Landsat-8. Therefore, comparing the downscaled MODIS LST with Landsat-8 validation data is incorrect. The issue of spatial pattern mismatch between MODIS LST (1000 m) and aggregated Landsat-8 LST (1000 m) on the same day is highlighted in Figure 5.4. Additionally, it is worth mentioning that MODIS and Landsat-8 LST images were acquired at different time, as indicated in Table 2.3. This temporal disparity could also contribute to the observed differences.

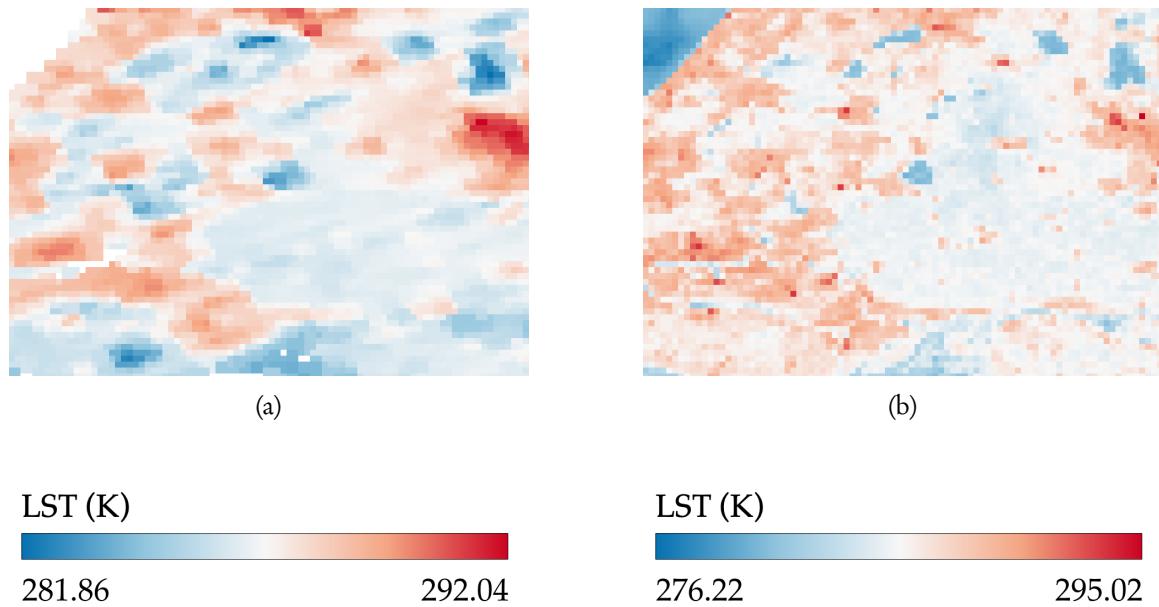


Figure 5.4: Comparison between (a) MODIS LST (1000 m) and (b) aggregated Landsat-8 LST (1000 m). The correlation coefficient  $r$  between (a) and (b) is 0.56

**Table 5.3** Evaluation metrics for the RF based downscaling experiment conducted with MODIS LST as the target variable before and after residual correction

	with (5 × 5) VV, (5 × 5) VH, homogeneity index, and land cover as predictors	
	without $\Delta T$	with $\Delta T$
<b>RMSE</b>	1.67	1.92
<b>Correlation Coefficient (<math>r</math>)</b>	0.70	0.56
<b>Coefficient of Determination (<math>R^2</math>)</b>	0.48	0.31

Moreover, since the spatial patterns of the two datasets differ (Figure 5.4), the process of residual correction (Equation 3.3) would yield poorer performance. This assertion is supported by quantitative evidence presented in Table 5.3.

To address this disparity and obtain a more representative validation set for MODIS LST at 100 m, a possible solution is to employ linear regression correction between the MODIS and Landsat-8 datasets. The basic spatial downscaling algorithm described in Section 3.1 can facilitate the implementation of this correction procedure. Similar to this study's approach of utilizing aggregated radar data at 1000 m as a predictor to model the variability in the target LST variable, the aggregated Landsat-8 LST (1000 m) can be used as a predictor, and MODIS LST (1000 m) can serve as the target to train a linear regression model. Considering that both the predictor and target variables are LST, a linear relationship between them is reasonable, justifying the selection of linear regression as the downscaling model. Once the model is trained, it can be utilized to generate a MODIS equivalent dataset at 100 m using the original Landsat-8 LST (100 m) as the input for the model.

However, it is essential to note that the suggested correction procedure is not without uncertainties, as it does not directly address the issue of spatial pattern mismatch. Therefore, applying the suggested correction process does not guarantee improved results.

## 5.6 CHALLENGE IN MODELLING TIME

A crucial challenge in generating fine-scale LST data from coarse-scale predictors is the temporal aspect. The main assumption underlying this research is that the high temporal resolution of satellites like MODIS, which acquire LST products at coarse spatial resolution, allows spatial downscaling models to be applied over time for accurate temporal estimation. However, this assumption faces practical challenges in real-life applications. The primary challenge arises from the weather dependency of TIR-based satellites. Consequently, building accurate models on a daily basis becomes infeasible as the daily data quality of the LST products acquired from TIR-based satellites could be affected due to bad weather conditions. Even if daily data is transformed using interpolation or other processing techniques to improve its quality, training models daily with new data can render the models unscalable. While this approach of constructing daily models using changing data may be practical, it would be more advantageous to develop an architecture that explicitly incorporates the temporal aspect.

Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM) networks, and similar

architectures could serve as potential candidates for addressing the temporal aspect. However, the challenge remains that TIR-based LST products are weather dependent, which may limit the availability of sufficient data to train these architectures effectively.

On the other hand, radar data is independent of weather conditions, suggesting that a spatial downscaling model trained on multi-temporal radar data could potentially provide accurate estimations over time. However, the temporal variations in LST values may not necessarily correlate with the temporal changes in predictors derived from radar datasets such as Sentinel-1 GRD images. To overcome this temporal challenge, one approach could involve considering the incorporation of phase information available in the Sentinel-1 SLC product. By mapping the phase disparity between two images acquired at different time, it may be possible to capture differences in LST values over time. Unfortunately, the phase information is not retained in the Sentinel-1 GRD product. Thus, addressing this limitation presents an avenue for future research exploration.

## Chapter 6

# Conclusion

In order to conclude this thesis, this chapter comprises of three sections that address the research questions (Section 6.1), provide future recommendations (Section 6.2), and offer a conclusion (Section 6.3).

### 6.1 RESEARCH QUESTIONS - ANSWERED

**Q1: *What new features can be engineered from the SAR data to achieve better performance for the task of downscaling?***

Considering the Sentinel-1 GRD SAR dataset employed in this research, two innovative features, namely neighboring values and the homogeneity index, were engineered to enhance the performance of the RF-based downscaling models. The results, presented in Section 4.1.1 and further discussed in Section 5.2 demonstrate that the incorporation of these features indeed leads to improved performance. Alternatively, the utilization of the Sentinel-1 SLC SAR dataset could offer additional features, such as spectral indices, dielectric constant, and others, due to its retention of phase information. These features could be considered in the development of radar-based LST downscaling models.

**Q2: *How does the proposed novel approach of incorporating land cover information into the downscaling models perform compared to the traditional approach?***

As detailed in Section 3.2.3, this research introduced a novel approach to incorporate land cover information into the downscaling models. The results presented in Section 4.1.3 clearly demonstrate that the performance of the proposed approach surpasses that of the traditional approach. However, it is important to acknowledge that this novel approach has a limitation in its robustness, as it is only applicable when downscaling LST to a resolution coarser than 10 m ( $>10$ ), which is the original spatial resolution of the ‘ESA WorldCover v100’ dataset. This limitation arises from the fact that the proposed approach aims to calculate the proportion of land cover within a pixel. At a resolution of 10 m, there are no further overlapping higher resolution pixels available to accurately calculate the proportions of different land cover classes. While practically, the approach can still be adapted for downscaling to a fine resolution of 10 m, the underlying significance of the approach may get compromised.

**Q3: *How effective is the proposed novel CNN-based architecture for the task of downscaling?***

As outlined in Section 4.2, the quantitative results obtained from the proposed novel CNN-based

downscaling architecture exhibited a slight superiority compared to the RF-based downscaling approach. However, upon conducting analysis of the downscaled images without residual correction, certain limitations stemming from the design of the architecture were identified. Potential modifications aimed at enhancing the proposed architecture were discussed in Section 5.4.

**Q4: How does the performance of SAR-based LST downscaling models compare to the performance of optical-based LST downscaling models?**

This research endeavors to forge a new path in the development of LST downscaling models utilizing SAR or radar data. Consequently, answering the above question is of utmost importance to this study. Despite the acknowledged limitations associated with radar data, as deliberated in Section 5.1, the performance of LST downscaling models employing predictors derived from radar data only falls a bit short of the performance achieved by predictors derived from optical data. As a result, by sacrificing only a negligible degree of performance, the utilization of radar data renders LST downscaling models independent of weather conditions. This performance is expected to witness improvement in future investigations through the incorporation of more comprehensive radar datasets, such as Sentinel-1 SLC or even quad-pol radar datasets.

Moreover, radar data can serve as a supplementary dataset to the optical dataset, potentially leading to enhanced performance. This notion was informally tested in one of the experiments and the attained performance of the fusion of optical and radar datasets fell somewhere between the performance derived from individual radar and optical datasets. It is worth noting, however, that rigorous testing is required to conclusively support this assertion, which is why it was not included as part of the thesis.

## 6.2 FUTURE RECOMMENDATIONS

Based on the answers of Section 6.1 and the overall content of Chapter 5, the future recommendations can be summarized as:

1. Incorporate additional features obtained from the Sentinel-1 SLC SAR dataset, such as RVI, dielectric constant, soil moisture, and others, to enhance the construction of LST downscaling models.
2. Enhance the proposed CNN architecture by integrating a fully connected neural network that performs a basic downscaling algorithm, thereby improving the model's capabilities.
3. Refine the proposed CNN architecture by incorporating a CNN architecture that aims to capture the spatial structure of LST in relation to the predictor variables, leading to improved downscaling performance.
4. Explore the use of advanced architectures, such as RNNs and LSTM, to model the temporal aspect of LST in combination with the spatial downscaling models.
5. Utilize the phase information derived from the Sentinel-1 SLC dataset to effectively model the temporal behavior of LST, enabling a more accurate representation of temporal changes over time.

These recommendations serve as valuable insights for future research endeavors, aiming to enhance the performance and applicability of LST downscaling models.

### 6.3 CONCLUSION

This research introduced a novel approach for estimating LST using SAR data. Two machine learning techniques, namely Random Forest and Convolutional Neural Networks, were employed to downscale coarse-resolution LST images from 1000 m to 100 m. The Sentinel-1 SAR images were utilized as predictors, enabling the generation of high-resolution LST images. The performance of the developed models was evaluated against Landsat-8 LST data at 100 m spatial resolution. Furthermore, innovative feature engineering techniques were proposed, including the incorporation of neighboring values and a radar-based homogeneity index, to address limitations observed in preliminary radar-based downscaling experiments.

Moreover, a comparison was made between the performance of the proposed radar-based downscaling method and an optical-based approach. Remarkably, despite the inherent limitations of radar data, the downscaling models built using radar predictors achieved comparable performance to those constructed using optical predictors. Notably, radar data exhibits the advantage of weather independence, making the downscaling models resilient to the effects of weather conditions.

To further enhance the downscaling models, a novel approach for incorporating land cover information was introduced. This approach successfully addressed one of the main limitations of radar data, which is the potential for inaccurate estimation of LST for specific land cover classes due to the wrong approximation between LST and radar backscatter.

The proposed CNN-based downscaling architecture demonstrated promising results, although certain modifications are required to address observed limitations.

Further, to improve the performance of radar-based downscaling models, it is recommended to incorporate additional information from more comprehensive radar datasets, such as Sentinel-1 SLC.

One of the primary challenges in generating high-resolution LST data from coarse-scale predictors is the temporal aspect. While radar images have shown potential for spatial downscaling, they are not suitable for accurate temporal estimation. Temporal variations in LST values may not align with changes in predictors derived from Sentinel-1 GRD intensity images. To overcome this issue, incorporating phase information from the Sentinel-1 SLC product can help capture temporal disparities in LST values over time.

In conclusion, this research contributes a new perspective to the field of LST estimation by leveraging SAR data. The findings highlight the suboptimal performance of radar-based downscaling models, emphasize the advantages of radar data in terms of weather independence, and propose strategies for further improvement. By considering the recommendations and addressing the identified limitations, future studies can advance the accuracy and applicability of radar-based LST downscaling models.



## Appendix A

### Other figures

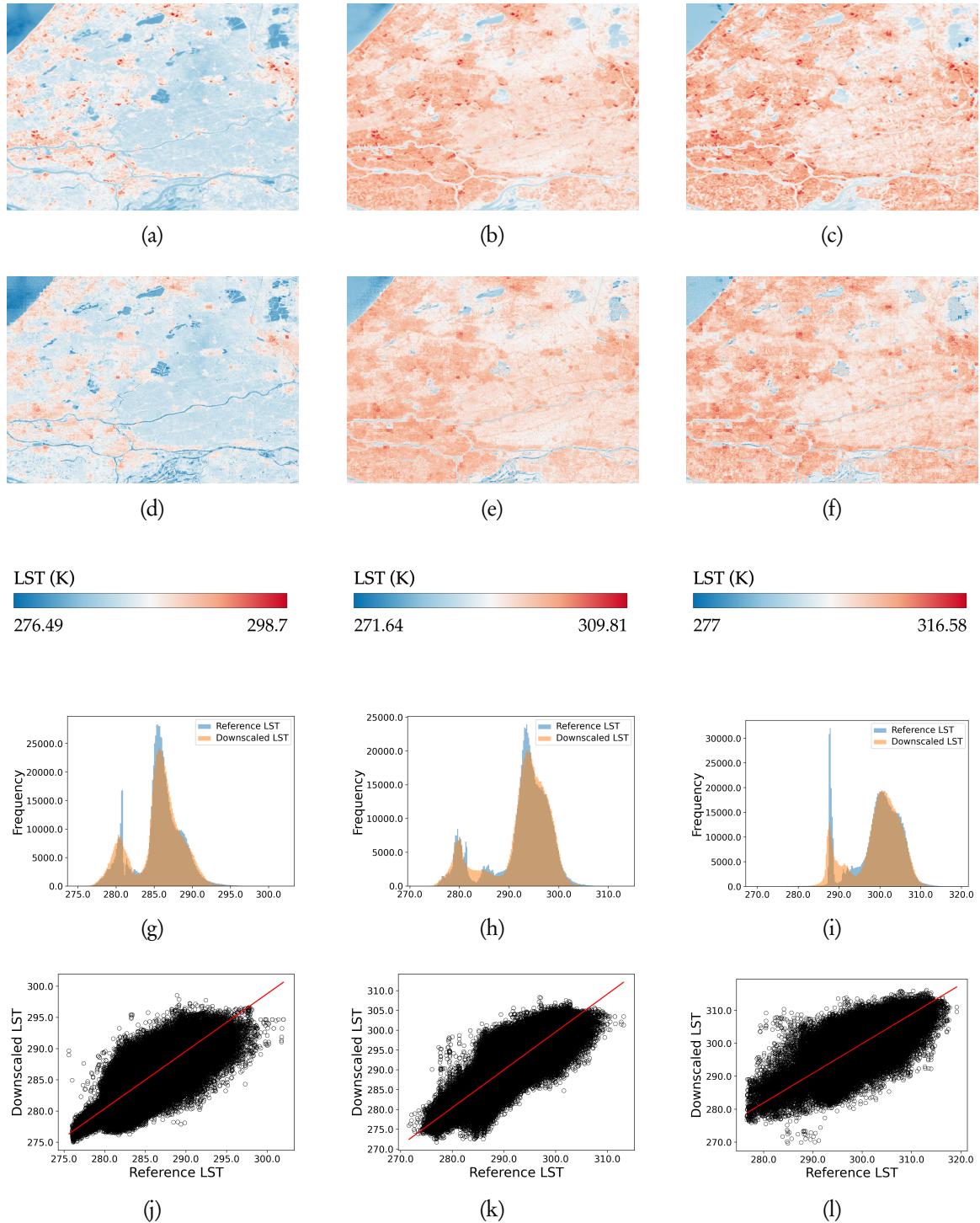


Figure A.1: Results of the ‘Case 1’ radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m)

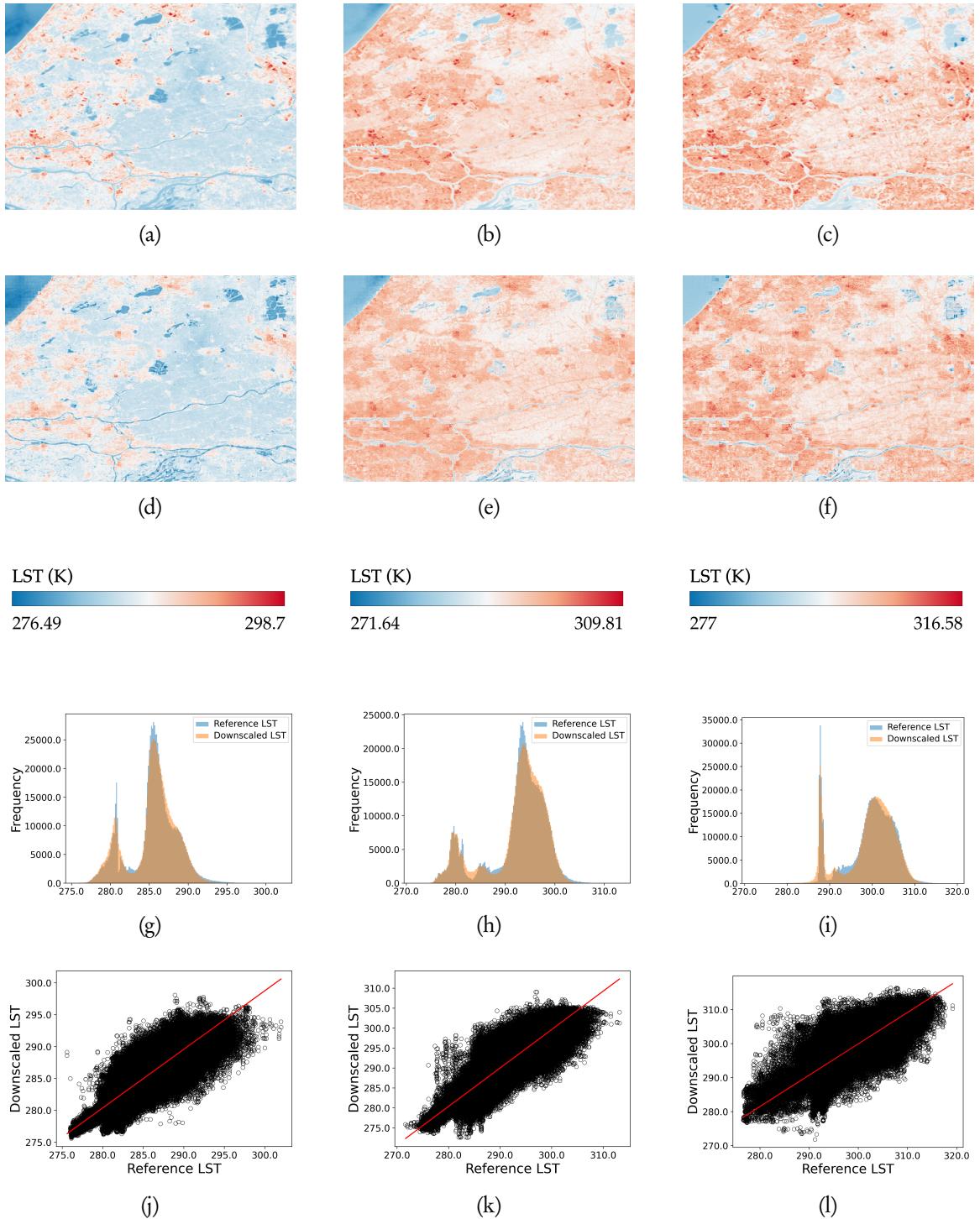


Figure A.2: Results of the 'Case 2' radar-based RF downscaling experiment. Here, (a), (b), and (c) refer to validation Landsat-8 LST (100 m) acquired on 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (d), (e), and (f) refer to the achieved downscaled LST images (100 m) for 25/03/2020, 10/04/2020, and 28/05/2020, respectively. (g), (h) and (i) refer to the histograms and (j), (k), and (l) refer to the scatterplots obtained on comparing downscaled LST images (100 m) to the Landsat-8 validation LST images (100 m)



## List of References

- Agam, N., Kustas, W. P., Anderson, M. C., Li, F., & Neale, C. M. (2007). A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sensing of Environment*, 107(4), 545–558. <https://doi.org/10.1016/j.rse.2006.10.006>
- Bai, L., Long, D., & Yan, L. (2019). Estimation of surface soil moisture with downscaled land surface temperatures using a data fusion approach for heterogeneous agricultural land. *Water Resources Research*, 55(2), 1105–1128. <https://doi.org/https://doi.org/10.1029/2018WR024162>
- Benallegue, M., Taconet, O., Vidal-Madjar, D., & Normand, M. (1995). The use of radar backscattering signals for measuring soil moisture and surface roughness. *Remote Sensing of Environment*, 53(1), 61–68. [https://doi.org/10.1016/0034-4257\(94\)00113-2](https://doi.org/10.1016/0034-4257(94)00113-2)
- Bhogapurapu, N., Dey, S., Homayouni, S., Bhattacharya, A., & Rao, Y. S. (2022). Field-scale soil moisture estimation using sentinel-1 GRD SAR data. *Advances in Space Research*, 70(12), 3845–3858. <https://doi.org/10.1016/j.asr.2022.03.019>
- Bindhu, V., Narasimhan, B., & Sudheer, K. (2013). Development and verification of a non-linear disaggregation method (nl-distrad) to downscale modis land surface temperature to the spatial scale of landsat thermal data to estimate evapotranspiration. *Remote Sensing of Environment*, 135, 118–129. <https://doi.org/10.1016/j.rse.2013.03.023>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Dominguez, A., Kleissl, J., Luvall, J. C., & Rickman, D. L. (2011). High-resolution urban thermal sharpener (HUTS). *Remote Sensing of Environment*, 115(7), 1772–1780. <https://doi.org/10.1016/j.rse.2011.03.008>
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307. <https://doi.org/10.1109/TPAMI.2015.2439281>
- Du, Z., Wang, Z., Wu, S., Zhang, F., & Liu, R. (2020). Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. *International Journal of Geographical Information Science*, 34(7), 1353–1377. <https://doi.org/10.1080/13658816.2019.1707834>
- 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. <https://doi.org/10.1109/TGRS.2016.2585198>
- Filgueiras, R., Mantovani, E. C., Althoff, D., Fernandes Filho, E. I., & da Cunha, F. F. (2019). Crop NDVI monitoring based on sentinel 1. *Remote Sensing*, 11(12). <https://doi.org/10.3390/rs11121441>
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.

- Fung, A., Li, Z., & Chen, K. (1992). Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 356–369. <https://doi.org/10.1109/36.134085>
- Guo, F., Hu, D., & Schlink, U. (2022). A new nonlinear method for downscaling land surface temperature by integrating guided and gaussian filtering. *Remote Sensing of Environment*, 271(August 2021), 112915. <https://doi.org/10.1016/j.rse.2022.112915>
- Hoeben, R., Troch, P. A., Su, Z., Mancini, M., & Chen, K. S. (1997). Sensitivity of radar backscattering to soil surface parameters: A comparison between theoretical analysis and experimental evidence. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3, 1368–1370. <https://doi.org/10.1109/igarss.1997.606449>
- Hu, Y., Tang, R., Jiang, X., Li, Z. L., Jiang, Y., Liu, M., Gao, C., & Zhou, X. (2023). A physical method for downscaling land surface temperatures using surface energy balance theory. *Remote Sensing of Environment*, 286(December 2022), 113421. <https://doi.org/10.1016/j.rse.2022.113421>
- Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127–141. <https://doi.org/10.1016/j.rse.2016.03.006>
- Julien, Y., & Sobrino, J. A. (2009). The yearly land cover dynamics (ylcd) method: An analysis of global vegetation from ndvi and lst parameters. *Remote Sensing of Environment*, 113(2), 329–334. <https://doi.org/10.1016/j.rse.2008.09.016>
- Kim, Y., & van Zyl, J. (2001). Comparison of forest parameter estimation techniques using sar data. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, 3(100), 1395–1397. <https://doi.org/10.1109/IGARSS.2001.976856>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.). Curran Associates, Inc.
- Kustas, W. P., Norman, J. M., Anderson, M. C., & French, A. N. (2003). Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sensing of Environment*, 85(4), 429–440. [https://doi.org/10.1016/S0034-4257\(03\)00036-1](https://doi.org/10.1016/S0034-4257(03)00036-1)
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, J., Wang, S., Gunn, G., Joosse, P., & Russell, H. A. (2018). A model for downscaling smos soil moisture using sentinel-1 sar data. *International Journal of Applied Earth Observation and Geoinformation*, 72(July), 109–121. <https://doi.org/10.1016/j.jag.2018.07.012>
- Li, W., Ni, L., Li, Z. L., Duan, S. B., & Wu, H. (2019). Evaluation of machine learning algorithms in spatial downscaling of modis land surface temperature. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2299–2307. <https://doi.org/10.1109/JSTARS.2019.2896923>
- Li, X., Zhang, G., Zhu, S., & Xu, Y. (2022). Step-By-Step Downscaling of Land Surface Temperature Considering Urban Spatial Morphological Parameters. *Remote Sensing*, 14(13). <https://doi.org/10.3390/rs14133038>
- 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. <https://doi.org/10.1016/j.rse.2012.12.008>
- Liang, M., Zhang, L., Wu, S., Zhu, Y., Dai, Z., Wang, Y., Qi, J., Chen, Y., & Du, Z. (2023). A High-Resolution Land Surface Temperature Downscaling Method Based on Geographically

- Weighted Neural Network Regression. *Remote Sensing*, 15(7). <https://doi.org/10.3390/rs15071740>
- Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J. M., McNairn, H., & Rao, Y. S. (2020). Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, 247(January), 111954. <https://doi.org/10.1016/j.rse.2020.111954>
- Njuki, S. M., Mannaerts, C. M., & Su, Z. (2020). An improved approach for downscaling coarse-resolution thermal data by minimizing the spatial averaging biases in random forest. *Remote Sensing*, 12(21), 1–23. <https://doi.org/10.3390/rs12213507>
- Nomura, R., & Oki, K. (2021). Downscaling of modis ndvi by using a convolutional neural network-based model with higher resolution sar data. *Remote Sensing*, 13(4), 1–20. <https://doi.org/10.3390/rs13040732>
- Oh, Y., Sarabandi, K., & Ulaby, F. (1992). An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 370–381. <https://doi.org/10.1109/36.134086>
- Patel, N., Aghababaei, H., Osei, F. B., Stein, A., & Mahour, M. (2023). Downscaling Land Surface Temperature Using SAR Images : A Machine Learning Framework. <https://doi.org/10.20944/preprints202306.1391.v1>
- Song, Z., Li, R., Qiu, R., Liu, S., Tan, C., Li, Q., Ge, W., Han, X., Tang, X., Shi, W., Song, L., Yu, W., Yang, H., & Ma, M. (2018). Global land surface temperature influenced by vegetation cover and PM 2.5 from 2001 to 2016. *Remote Sensing*, 10(12), 1–18. <https://doi.org/10.3390/rs10122034>
- 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(24), 1–4. <https://doi.org/10.1029/2007GL031485>
- Ulaby, F. T. (1974). Radar measurement of soil moisture content. *IEEE Transactions on Antennas and Propagation*, 22(2), 257–265. <https://doi.org/10.1109/TAP.1974.1140761>
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J. F., & Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199, 415–426. <https://doi.org/10.1016/j.rse.2017.07.015>
- Wan, Z., Wang, P., & Li, X. (2004). Using modis land surface temperature and normalized difference vegetation index products for monitoring drought in the southern great plains, usa. *International Journal of Remote Sensing*, 25(1), 61–72. <https://doi.org/10.1080/0143116031000115328>
- Wang, R., Gao, W., & Peng, W. (2020). Downscale MODIS land surface temperature based on three different models to analyze surface urban heat island: A case study of Hangzhou. *Remote Sensing*, 12(13). <https://doi.org/10.3390/rs12132134>
- Wang, S., Luo, X., & Peng, Y. (2020). Spatial Downscaling of MODIS Land Surface Temperature Based on Geographically Weighted Autoregressive Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 2532–2546. <https://doi.org/10.1109/JSTARS.2020.2968809>
- Weng, Q., Fu, P., & Gao, F. (2014). Generating daily land surface temperature at landsat resolution by fusing landsat and modis data. *Remote Sensing of Environment*, 145, 55–67. <https://doi.org/10.1016/j.rse.2014.02.003>
- Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., Wen, M., & Zhang, N. (2021). Spatial Downscaling of Land Surface Temperature Based on a Multi-Factor Geographically Weighted Machine Learning Model. *Remote Sensing*, 13(6), 1186. <https://doi.org/10.3390/rs13061186>

- Yang, G., Pu, R., Huang, W., Wang, J., & Zhao, C. (2010). A novel method to estimate subpixel temperature by fusing solar-reflective and thermal-infrared remote-sensing data with an artificial neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4 PART 2), 2170–2178. <https://doi.org/10.1109/TGRS.2009.2033180>
- Zakšek, K., & Oštir, K. (2012). Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sensing of Environment*, 117, 114–124. <https://doi.org/10.1016/j.rse.2011.05.027>
- Zhang, F., Zhang, L. W., Shi, J. J., & Huang, J. F. (2014). Soil moisture monitoring based on land surface temperature-vegetation index space derived from modis data. *Pedosphere*, 24(4), 450–460. [https://doi.org/10.1016/S1002-0160\(14\)60031-X](https://doi.org/10.1016/S1002-0160(14)60031-X)
- Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11), 2610–2623. <https://doi.org/10.1016/j.rse.2010.05.032>