

Article

Analysis of MODIS LST Compared with WRF Model and *in situ* Data over the Waimakariri River Basin, Canterbury, New Zealand

Mohammad Sohrabinia ^{1,*}, Wolfgang Rack ² and Peyman Zawar-Reza ¹

¹ Atmospheric Research Centre, Department of Geography, University of Canterbury, Christchurch, New Zealand; E-Mail: peyman.zawar-reza@canterbury.ac.nz

² Gateway Antarctica, Centre for Antarctic Studies & Research, University of Canterbury, Christchurch, New Zealand; E-Mail: wolfgang.rack@canterbury.ac.nz

* Author to whom correspondence should be addressed; E-Mail: sohrabinia.m@gmail.com; Tel.: +64-22-601-4495.

Received: 12 September 2012; in revised form: 9 November 2012 / Accepted: 12 November 2012 / Published: 19 November 2012

Abstract: In this study we examine the relationship between remotely sensed, *in situ* and modelled land surface temperature (LST) over a heterogeneous land-cover (LC) enclosed in alpine terrain. This relationship can help to understand to what extent the remotely sensed data can be used to improve model simulations of land surface parameters such as LST in mountainous areas. LST from the MODerate resolution Imaging Spectro-radiometer (MODIS), the modelled surface skin temperature by the Weather Research and Forecasting (WRF) mesoscale numerical model and the *in situ* measurements of surface temperature are used in the analysis. The test-site is located in a mountain valley in the Southern Alps of New Zealand. Geospatial analysis in GIS is used to relate pixels, grid-cells and points from the MODIS LST, model simulations and the *in situ* data, respectively. Differences between LST from MODIS, the WRF model and the *in situ* data are presented with respect to surface LC at different times of day. Initial results from regression analysis of the three datasets showed a goodness of fit R^2 coefficient of 0.77 for the model simulations and 0.35 for the MODIS LST. These values improved significantly when time-lags were considered and the few outliers were removed, giving R^2 values of 0.80 for the model and 0.73 for the MODIS LST. These results show that the WRF model correlates better with the *in situ* measurements over various LC types in this region compared with the MODIS LST. Longer time-series, however, are required to draw more robust conclusions about the applicability of the MODIS LST product for improving WRF simulations over alpine complex terrain.

Keywords: MODIS; land surface temperature; LST; WRF; mesoscale model; *in situ* data; iButtons; Waimakariri

1. Introduction

The importance of the land surface temperature (LST) in surface-based bio-geophysical processes and land-atmosphere interactions is well documented in the literature [1–8]. In complex terrain, meteorological stations and ground surveys are usually sparsely and/or irregularly distributed [6], hence, remotely sensed LST can be a critical source of data for the study of land surface processes. Skin temperature (in particular, its diurnal cycle) is needed in calculating sensible and latent heat fluxes; specifically, sensible heat flux is determined by the instantaneous difference between the near-surface air temperature (T_a) and LST [2]. LST also determines the amount of thermal heat that is vertically transported into the ground. LST is routinely retrieved from remote sensing satellite sensors such as the MODerate resolution Imaging Spectro-radiometer (MODIS), providing reasonably good resolutions in space (1 km) and time (approximately 4 observations per day) at a regional scale.

As an attribute of the land surface, LST is influenced by the local land-cover (LC). Therefore, the quality of LST retrievals from satellite observations over various LC types needs to be assured in order to use this data source in the above applications.

Comparative analysis of LST from remote sensing data and modelling approaches in the existing research can be categorized into four groups. The first group attempt to improve LST retrievals via modelling complex LC and terrain features. These studies have examined LC for the purpose of a better approximation of surface emissivity, which rely on vegetation fraction of the surface cover [8,9]. The second group have attempted to use remotely sensed thermal and LC data to improve atmospheric models for simulation of land surface parameters [7,10–12]. The third group have used remotely sensed LST to study the near-surface T_a [13–15] or surface soil moisture [16,17]. Some of these studies have related LC indirectly to T_a looking at LST as the intermediate link [18]. Finally, the last group of the published research have used modelling and *in situ* measurements for validation of the MODIS LST [3,4,19–24].

The focus of the first group is physical modelling of the parameters involved in measurement of LST, including the sensor, the atmosphere in between and the properties (such as emissivity) of the target area. The second group attempt to improve accuracy of numerical models through assimilation of satellite observational data into the models. Both of these groups focus on producing or using satellite derived data often without comparison with a similar database. The third and the last group, on the other hand, are usually concerned with validation of satellite observations based on ground-truth data. Other works in the literature, which are concerned with both LC and LST, have exploited the inverse relationship between LST and surface vegetation density (e.g., [25]), some of which have tried to account for the sub-pixel variations in LST based on higher resolution LC data (e.g., [26]).

Despite numerous research conducted on the validation of remotely sensed LST from various sensors, such as MODIS (e.g., [19,21,27]), Landsat (e.g., [23]) or a combination of sensors [3,4] in different parts

of the world, a study of LST variations over different LC types is missing in the literature. LST itself is used to derive other surface parameters, such as the near-surface soil moisture (SM). The near-surface SM is a function of soil temperature [10], therefore, remotely sensed LST is widely used to study it under different soil conditions (e.g., [16,17]). Each LC type has its own unique thermal properties with daily cycles of heating and cooling, hence, the near-surface SM derivations can also be affected by LC types of the surface. As a consequence, characterization of LST variations over different LC types is equally critical in SM derivations. Apart from the *in situ* data, weather models also can be used to analyse the effects of LC types on LST, as well as the near-surface SM. Being able to provide high-resolution data temporally and spatially, computational models of climate and weather provide an opportunity for understanding of LST over a heterogeneous land. Since LST is a key computed variable in the model schemes based on heat fluxes, it is readily available. The spatial distribution of energy and heat fluxes can only be taken into account with remote sensing methods or numerical models [28]. However, research on quality of the computed values using remotely sensed data as a basis is limited and does not include a comprehensive analysis over various LC types.

The objective of this paper, therefore, is to examine the relationship between LST observations from MODIS with the modelled dataset and the *in situ* measurements over various LC types in the study area. Taking the *in situ* measurements as reference, the aim of the study is to find out if the MODIS LST product is applicable to improve surface skin temperature simulations from the Weather Research and Forecasting (WRF) model over a valley in alpine terrain. There are very few climate stations in mountain valleys such as the one that we have studied; consequently, satellite data and numerical modelling are the only feasible solutions currently available for long-term environmental and climate studies in these areas. Nevertheless, such a terrain poses complexities on both numerical models and the retrievals from satellite observations.

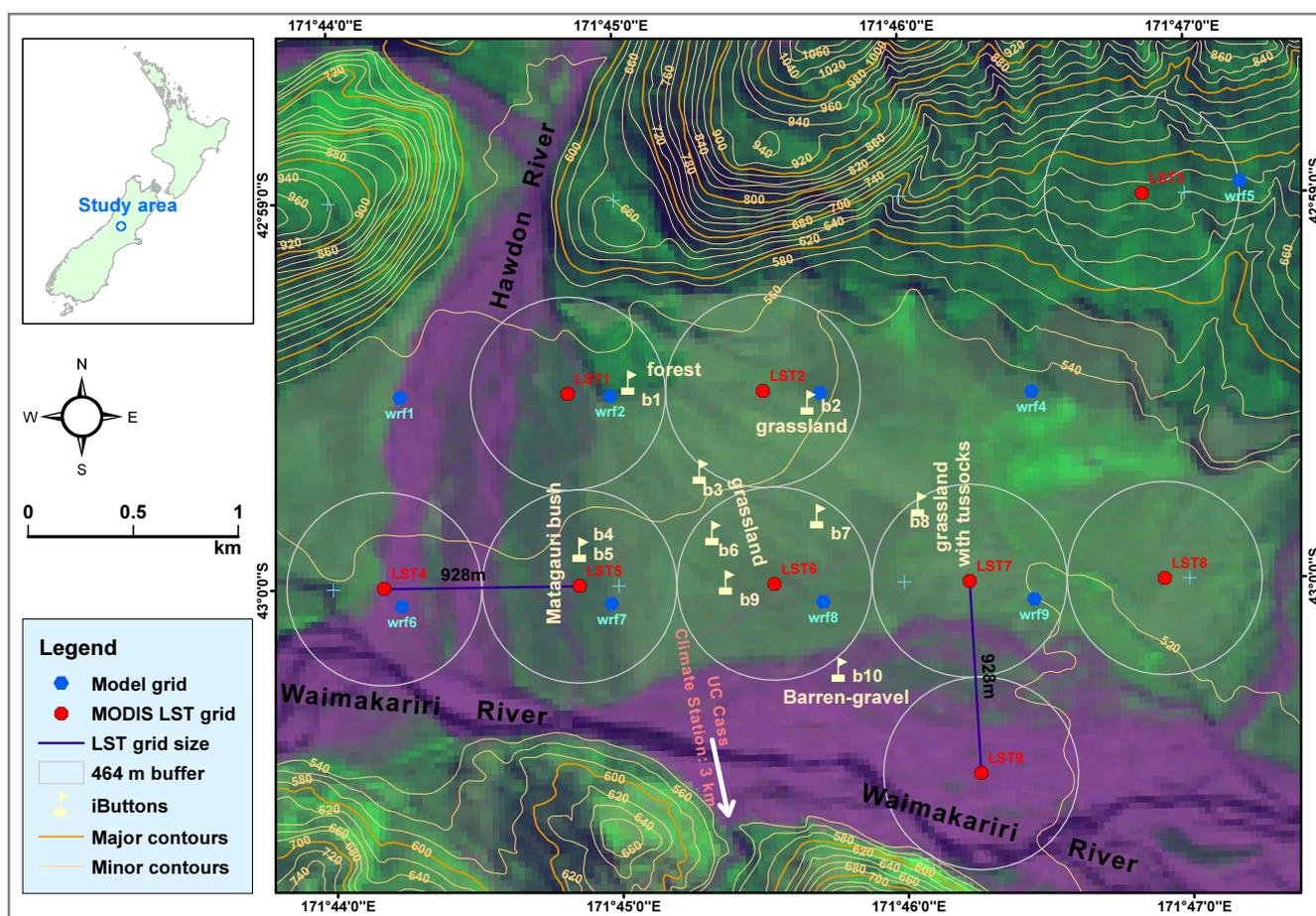
The paper is organized as follows. The study area and the used datasets (*in situ*, MODIS, WRF) are described in Sections 2 and 3, respectively. This is followed by a description of pre and post-processing methods with a focus on the effects of MODIS viewing angles and surface emissivities in Section 4. Results are provided in Section 5, where in Section 5.1, time-series of all three datasets are used as input to a correlation analysis, without looking at the spatial heterogeneity of LC types. In Section 5.2 a LC type specific correlation analysis between the three datasets is provided, and in Section 5.3 correlations characterizing day and night time-series over different LC types is presented. The interpretation of the results is discussed in Section 6, before the summary and conclusion are drawn in Section 7.

2. Study Area

The study area is located in a valley of the Waimakariri River basin in the Southern Alps of New Zealand (171°45'29"E, 42°59'39"S). The area is relatively flat with an average elevation of 550 m above sea level (a.s.l.), however, high-rise mountains border the area just to the North and to the South (Figure 1). The river's braided channel is more than a km wide, whereas the active riverbed is dynamically shifting the braids across the channel, leaving behind barren river banks on the sides. The LC types of this region are predominantly grasslands, mixed grass and tussock, a mixture of Matagauri bush with other scrubs and native forest with Beech trees as the dominant species. The spatial extent of

these LC types were interpreted from a Landsat TM 5 image (acquisition date 28 March 2011, 43 days before the field experiment) and were visually checked on the ground during the field experiment. Since the area is free of farming and agricultural activities, the natural landscape is intact with a relatively stable canopy. In mid-Autumn, the season when the field experiment was conducted, vegetation on the site maintained their leaves, and the seasonal variation of the LC types were negligible. As a result, this site provides diverse LC types and adequate spatial extent so that the LST variations over LC types can be differentiated. The rationale for choosing this test-site was to collect measurements over a relatively flat land of the mountain valleys in this region, while at the same time to avoid rugged terrain and significant elevation variations to be able to compare the measurements with satellite observations.

Figure 1. *In situ* iButtons (b1 to b10), MODIS grid points (LST1 to LST9) and the WRF model grid points (wrf1 to wrf9) overlaid on a Landsat image (TM5, 28 March 2011) of the test-site and plotted with 20 m topography contours lines.



The climate of the study area is alpine with large diurnal temperature differences at day and night and freezing temperatures in winter. Based on climate data from the University of Canterbury Cass (UC Cass) Automatic Weather Station (171°45'34"E, 43°02'05"S, 583 m a.s.l.), which is located 3 km to the South of the study area, long-term average T_a of the area is 9.4 °C (1998–2012). However, monthly minimums and maximums fluctuate between -10 and 30 °C during cold and warm months of the year, respectively.

3. Data

Data analysed in this paper fall into one of three categories: (i) *in situ* ground surface temperature (GST) measurements (LST_{inSitu}); (ii) LST retrievals from MODIS thermal infrared (TIR) observations (LST_{Modis}); and (iii) the WRF numerical model outputs (LST_{wrf}). The first two categories are observational data with different levels of accuracy and precision, while the third category is simulated data generated by the state-of-the-art WRF numerical meteorological model. These data are explained in more detail below.

(i) A field experiment for the *in situ* measurements of GST in the study area was conducted over a period of 14 days from May 10 to 23, 2011. Temperature data-loggers known as **iButtons** (Thermochron® DS1922L iButton, $-40\text{ }^{\circ}\text{C}$ to $85\text{ }^{\circ}\text{C}$ sensitivity range, $\pm 0.5\text{ }^{\circ}\text{C}$ accuracy, $0.0625\text{ }^{\circ}\text{C}$ resolution) were used in this experiment to measure GST over five different LC types of the area, allocating at least one iButton for each LC type (shown as b1 to b10 in Figure 1 and described in Table 1). These low-cost temperature sensors have already become widely used in geoscience research (e.g., [29–32]). The iButtons were cross-calibrated at indoor conditions before use, and the accuracy of all 10 matched the manufacturer’s specification. The sensors were placed over five LC types: grassland, grassland with tussock, forest, barren/gravel and Matagauri bush/scrub. Considering surface conditions and chances of exposure to direct sunlight, more than one sensor was employed over some of the LC types to ensure at least one measurement site with good quality data. The minimum distance between samples was 250 m. Sampling rate for the loggers was set to once every 30 minutes with New Zealand Standard Time (NZST) as reference. NZST is equivalent to UTC+12 hours (centred at 180° longitude), therefore, an offset of -33 minutes needed to be subtracted from NZST corresponding to the longitudinal position of the test site (171.75° E) in order to synchronise the ground measurements with the local solar time provided in the MODIS LST product for every pixel ($171.75^{\circ} - 180^{\circ} = -8.25^{\circ}$ longitude ≈ -33 min in time). Five iButtons were installed over Grassland (b2, b3, b6, b7 and b9), two iButtons on Bush-Scrub (b4, b5) and one iButton over the other LC types, which include Forest (b1), Grassland with Tussock (b8) and Barren-Gravel (b10). All the iButtons but 1 and 5 were placed at a depth of 1–2 cm in the soil for shade and protection against direct sunlight. The Forest LC type is a relatively dense canopy of native Beech trees. For a better approximation of the temperature on top of the forest measured by satellites, iButton 1 was fixed on a tree inside the forest at about 4 m height. The bush-scrub LC type is a sparse canopy of Matagauri native bush with about 50% density. To account for the temperature on top of the bush, as well as on the open ground within the bush canopy, iButton 4 was buried in the soil while iButton 5 was fixed on top of a bush, shaded from direct sunlight. We used the data collected over grassland to examine the spatial heterogeneity of GST, and to determine whether one iButton was able to representatively measure the surface temperature over a single LC type. Difference among 5 iButtons at similar times was less than $2\text{ }^{\circ}\text{C}$ for most of the measurement period. Few exceptions were observed during early morning and around noon (such as 22 May 2011 11:00 AM) when the differences exceeded $4\text{ }^{\circ}\text{C}$ (Figure 2). This indicates that one iButton could be able to capture GST over Grassland with an accuracy of $\pm 1\text{ }^{\circ}\text{C}$ for most of the period. Nevertheless, due to larger canopy height and complexity over Forest and Bush-Scrub, we estimate the magnitude of the uncertainty over these LC types to be

higher than ± 1 °C. A single averaged time-series of GST from the 5 iButtons over Grassland was also compared with the measurements from other LC types (Figure 3). Apart from the iButton on top of the bush (BushT), the range of GST over the LC types has been less than 5 °C for most of the measurement period. At warm hours the range has increased up to about 10 °C. Due to sheltering by the canopy, GST time-series from the iButton buried in the ground among the bush (BushG) showed less sensitivity for day and night extremes.

Given the shallow depth of installations, these iButton measurements are used as a proxy to compare with the MODIS LST. Although this should be presumably a good approximation, we conducted a time-lag analysis to account for the differences depending on the soil heat capacity and LC type. Besides the iButton point measurements, we also used the model simulations (explained further below) as an independent variable, which is also gridded data with a cell-size similar to the MODIS pixel spacing.

Table 1. Specifications of iButton measurement sites. All iButtons were buried at a depth of 1–2 cm in the soil except iButtons 1 & 5.

iButton	LC type	Description
b1	Forest	Dense native Beech trees with ≈ 10 m canopy height
b2	Grassland	Native grass-turf with ≈ 5 cm height
b3	Grassland	”
b4	Matagauri bush/scrub	Native bush with $\approx 50\%$ density and ≈ 1.5 m height
b5	Matagauri bush/scrub	” (this iButton was fixed on top of the bush)
b6	Grassland	Native grass-turf with ≈ 5 cm height
b7	Grassland	”
b8	Grassland with tussocks	” mixed with native tussock with ≈ 20 cm height
b9	Grassland	Native grass-turf with ≈ 5 cm height
b10	Barren-gravel	Bare soil with gravel on the inactive river banks

Figure 2. Minimum and maximum ground temperature measurements from 5 iButtons over Grassland LC type (30-min rate, May 2011). Night observations (7 pm–7 am) are distinguished by the grey-line overlaid on the plot.

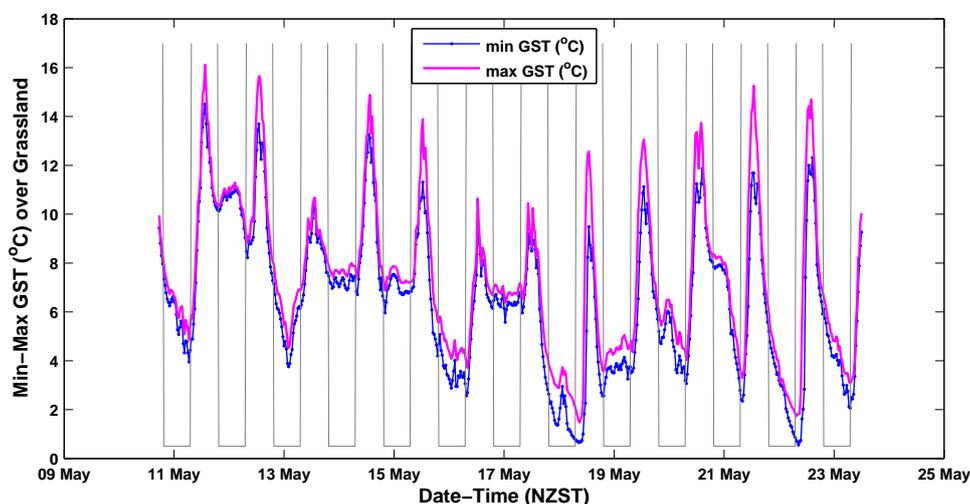
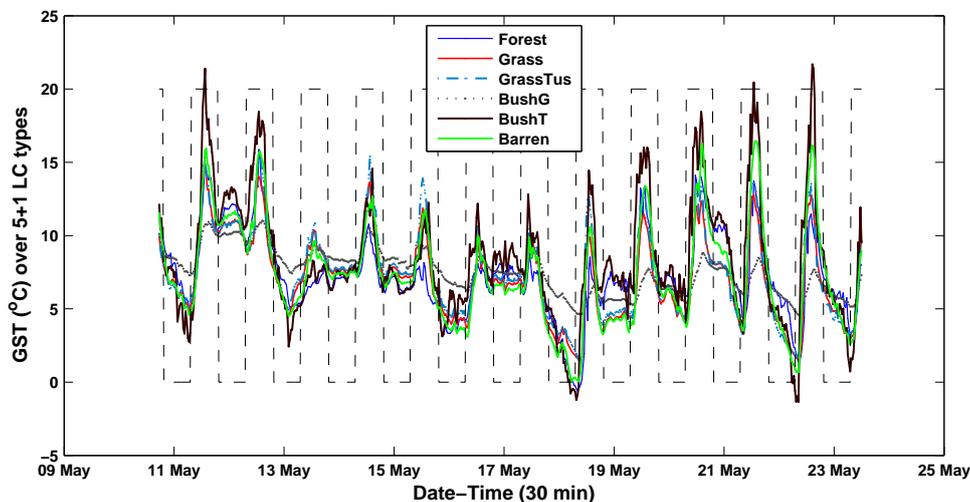


Figure 3. Comparison of GST time-series measured over 5 LC types plus the iButton on top of the bush (30-min data, May 2011). The broken grey-line is overlaid on the plot to distinguish day and night measurements; it is assigned 0 for night (7 pm–7 am) and 20 °C for day (7 am–7 pm).



(ii) MODIS LST product level 3, collection 5 (V5) archived in hierarchical data format (HDF) with approximately 1 km spatial resolution is used in this paper. This product is retrieved from the MODIS TIR observations in 10.78 to 12.27 μm range (bands 31 & 32) using the generalized split window (GSW) algorithm [1,19–21,27,33,34]. This dataset contain TIR observations for both day and night overpasses of MODIS-Terra at $\approx 10:30$ (descending) and 22:30 (ascending) and MODIS-Aqua at $\approx 13:30$ (ascending) and 1:30 (descending) local solar time. Therefore, four observations are available for each day from a combination of the two sensors and day and night overpasses of each sensor in ideal conditions (e.g., no cloud-cover). Because TIR signals cannot penetrate clouds, pixels contaminated with cloud have already been skipped in the LST processing using a cloud mask, so that LST is not mixed with cloud-top temperature [19,20], and are assigned a fill value in the LST product [35]. The level 3 V5 product is derived in clear-sky conditions at $\geq 95\%$ confidence defined in MODIS cloud-mask (MOD35) product over land with $\leq 2,000$ m elevation a.s.l. [35]. This leaves the possibility of TIR observations suspected to be cloud contaminated with $< 5\%$ confidence to be skipped from cloud masking. Inside each HDF dataset, the first and fifth data-fields are TIR observations for day and night, respectively. Digital data storage precision of the product is 16-bit unsigned, with a scale factor of 0.02 to produce temperature values in Kelvin (K) after scaling [35,36]. This allows for a quantization precision (or radiometric resolution) of 0.02 K for the LST product at data management and conversion stage. The original values of LST product were first scaled based on the product's documentation [35] using the corresponding attribute fields. Dealing with a large number of LST data from MODIS-Terra and MODIS-Aqua for day and night over the study area, a code was written to perform radiometric scaling automatically. This code used the corresponding attribute of each field for day and night in each HDF file to produce time-series of LST for each LC type scaled to K. Spatial locations of pixels from this dataset are marked by LST1 to LST9 on the map of the study area (Figure 1).

(iii) WRF simulations for *land surface skin temperature* in Kelvin (abbreviated as TSK in the WRF model outputs) were produced for the corresponding period of the field measurements. We used version 3 of the WRF modelling system for this work. WRF is a state-of-the-art community atmospheric model, which is suitable for use in a broad range of applications across scales ranging from metres to thousands of kilometres [37]. A detailed description of the WRF modelling system can be found in [37,38]. The static geographic data, which include MODIS-based 20-category LC classification at 1-km resolution (available from NCAR) are used by the model to interpolate terrestrial static fields to the prescribed domains (see [38]). National Centers for Environmental Prediction (NCEP) Final Operational Global Analysis (known as FNL) data with 1.0x1.0 degree grid resolution and six hourly frequency (see <http://rda.ucar.edu/datasets/ds083.2>) were used as input for initialization and boundary conditions of the 3-dimensional atmospheric fields (see also [39]). Terrestrial data (including the local LC) are integrated by the Geogrid component in the 2-dimensional static fields and global meteorological data are ingested by the Metgrid component into the 3-dimensional atmospheric fields of the model. Each of the static and atmospheric fields contain various terrestrial and atmospheric information about the target area for the simulation, including (but not limited to) skin temperature (K), layers of soil temperature (K), vegetation/land-use type, vegetation greenness fraction (VGF), relative humidity, soil moisture and annual mean temperature (see [38,40] for a complete list and details of these layers). The spatial resolution of the default USGS 24-category land-use data available in the static layers is relatively coarse, therefore, we used NOAH Land Surface Model (LSM) scheme with higher resolution MODIS LC categories for the best compatibility with the spatial resolution of the MODIS LST. NOAH LSM uses four soil layers (of temperature, water+ice, water) [38], one vegetation type in each grid cell without dynamic vegetation and carbon budget [40], and predicts soil moisture and temperature in four layers. Ground heat budget in NOAH LSM is calculated using a diffusion equation for soil temperature and the surface skin temperature is determined using a single linearized surface energy balance equation [41]. The input physics parameters in this scheme include Rapid Radiative Transfer Model (RRTM) for longwave [38] and Dudhia scheme [42] for shortwave atmospheric radiation (see also [40]). The first domain of the simulations covered the entire South Island of New Zealand centred on the study area. The nested domain covered the study area and the surrounding region. Considering the spatial resolution of the LST product and observation timing of MODIS, grid-size of the nested domain and the interval of the model output were set to 1 km and 30 minutes, respectively. This produced 48 values with 1 km spatial resolution for each day. Duration of the simulations covered one day before till the end of the field experiment. The first 24-hours (or 12-hours as used in the literature) is the spin-up time required by the model to reach a balanced state with the boundary conditions [39]. These settings facilitated comparisons of the model TSK with the MODIS LST. The outputs of the WRF model were in netCDF (Network Common Data Form) format. Since TSK is the simulated parameter equivalent to LST from MODIS, it will be referred to as LST_{wrf} . Spatial locations of the grid-points from this dataset are marked by wrf1 to wrf9 on the map of the study area (Figure 1).

4. Methods

4.1. LST Pre-Processing

We used raster image analysis in order to overlay the MODIS LST with the LC data from the study area. A certain number of pre-processing steps were required to convert the original LST product in HDF format to raster layers with a versatile projected coordinate system. First of all, raster subsets of the LST product were extracted based on the boundary extent of the study area. LST L3 product is gridded in the global Sinusoidal projection, and the grid containing data for the study area is located at column 30 (h30) east-west and line 13 (v13) north-south. However, in this projection New Zealand falls in the lower-right corner with distortion along east-west direction. Using ESRI ArcGIS projection conversion utilities, “New Zealand Transverse Mercator 2000” coordinate system (Spheroid GRS 1980, Datum: NZGD 2000) was applied on the LST raster subsets. With spatial overlay in GIS, coordinates of LST pixels for each LC type in the study area were determined. These coordinates were used in a code to read LST values from the entire HDF files covering the period of field experiment and constructing LST time-series for each LC type. Data for observation times affected with cloud-cover were automatically filtered out by the code using fill-value attribute of the LST product. Quality control field for each observation helped to determine the level of accuracy of that observation.

4.2. Spectral Unmixing

The linear mixing model is used for unmixing the LST pixels with mixed LC types in the study area (Equation (1)). The linear mixing model assumes that the observed reflectance spectrum of a given pixel is generated by a linear combination of a small number of unique constituent deterministic spectral signatures [43]. This model calculates the final LST values of each pixel based on the fractional abundance of each LC type (or endmember) in that pixel. This model was applied on the correlation analysis between Forest, Bush/Scrub and Barren-gravel, which are measured by b1, b4/5 and b10 *in situ* iButtons, respectively (Figure 1).

$$x = \sum_{i=1}^M a_i s_i + w \quad (1)$$

where M is the number of endmembers, a_i (i from 1 to M) is the fractional abundance vector, w is the additive observation noise, x is the value of the pixel after unmixing [43–45]. x is calculated through summation of the weighted values of the endmembers ($s_{i=1}^M$) using fractional abundance values ($a_{i=1}^M$) and adding the noise term. This method was applied on the LST pixels with a mixing of 5% or more, using fractional abundance vectors of each pixel (Table 2). Values of pixels with a homogeneous LC and less than 5% mixing have been used as the endmember (s_i) to calculate the unmixed value of each target pixel. LST5 representing LC type Bush-scrub was an exception, for which there was no ideal homogeneous pixel, hence a weighted value based on LST5 (representing 50% Bush-scrub) and LST6 (representing 40% grassland) was used. For those endmembers that $a_i = 0$, s_i will be 0 too, however, sum of fractional abundance values in each a_i vector must be equal to 1. Since we had only five pixels representing five LC types, which were validated on the ground, $w = 0$ has been assumed in

the unmixing process. Equation (1) is useful for a single-band or thematic image, but can be expanded to multi-band imagery (see [46,47]).

Table 2. Fractional abundance vectors ($a_i = \text{LST1, LST4, etc.}$) and endmembers (s_i) used in spectral unmixing.

LC-Types	LST1	LST4	LST5	LST6	LST7	s_i
Barren-gravel	0.35	0.70	0.10	0.15	0.10	LST9
Bush-scrub	0.25	0.05	0.50	0.00	0.05	LST5-6
Forest	0.25	0.00	0.00	0.00	0.00	LST3
Grassland	0.15	0.20	0.40	0.85	0.10	LST2
Grass-Tussock	0.00	0.05	0.00	0.00	0.75	LST8

4.3. Spatial Overlay of MODIS LST, WRF Simulations and in situ Points

We defined the WRF model domain's central point in a way that model grid-cell centres were located as close as possible to the LST pixel centres. Subtle tuning of the model domain's central point moved coordinates of the model grid points close to the corresponding coordinates of LST pixel in the study area. However, due to the difference in the pixel-size of the MODIS LST (928 m) with the grid cell-size of the model (1 km), a difference of about 100 to 300 m in the positions of the grid cells and LST pixels was inevitable (blue and red points in Figure 1). We used a nearest neighbour approach confined with a 462 m proximity rule (white circles in Figure 1) for the comparison. Any points (or WRF grid-centres) out of this proximity distance were not spatially related to the MODIS LST centroids (e.g., b3 in Figure 1). The distance used in the proximity rule is equal to one-half of the pixel-size of the MODIS LST and almost one-half of the model grid's cell-size for the nested domain. This technique was necessary in order to relate the pixels of $\text{LST}_{\text{Modis}}$, point layers derived from LST_{wrf} , and *in situ* measurement points. There was no LST pixel with similar LC type within the 464 m buffer of the iButton over "Barren-Gravel" LC type (b10), therefore, time-series of this iButton were related to a LST pixel (LST4) which is outside the defined buffer but has a similar LC type.

4.4. View Angle and Emissivity Analysis

Since LST is affected by viewing angle (θ) and the emissivity (ϵ) of surface LC types, we needed to analyse these parameters prior to LST correlation. Every scan of MODIS-Terra and Aqua covers a field of view extending $\pm 55^\circ$ from the perpendicular to the surface NADIR zenith line. The actual local view angle extends further to $\pm 65^\circ$ due to curvature of the Earth's surface [33]. Viewing angle for all pixels of the study area at a single overpass were identical (it could vary if the study area was larger than one degree in the across-track direction of the MODIS ground field of view, or more than $2330/130 \approx 18$ km wide). Retrieved emissivities for each pixel in bands 31 and 32 (emis31 & emis32 fields in the LST product) showed that the emissivities of each LC type varies over time (Figure 4), where the amount of standard deviations (σ) from the mean is larger for some LC types (especially in band 31, Figure 4(a)). The mean emissivities of "Forest" were higher than all other LC types from both bands

31 and 32 over the period of field measurements (Figure 4). Despite the expected lower values for bare soil, the emissivity fields showed no significant difference between “Barren-Gravel” and other LC types, such as “Bush/scrub” or “Grassland”. However, “Barren-Gravel” and “Bush-Scrub” showed the least σ from the mean over the analysed period (this is shown by the upper and lower caps of the error-bars in Figure 4). This analysis also showed that the majority of the observations have been viewed from the East (shown with negative values), some exceeding 55° from NADIR (Figure 5). On the other hand, θ from the West is under 55° for all cases. We also found that larger variations in ϵ have occurred when θ exceeded $\pm 55^\circ$ (Figure 6).

Figure 4. The mean (dots) and $\pm 1\sigma$ (bars with caps) from the mean emissivities over the period of 14 days (May 10–23, 2011) in bands 31 & 32 of MODIS over 5 LC types (BG: Barren-Gravel, BS: Bush-Scrub, NF: Native Forest, G: Grassland, GT: Grassland-Tussock).

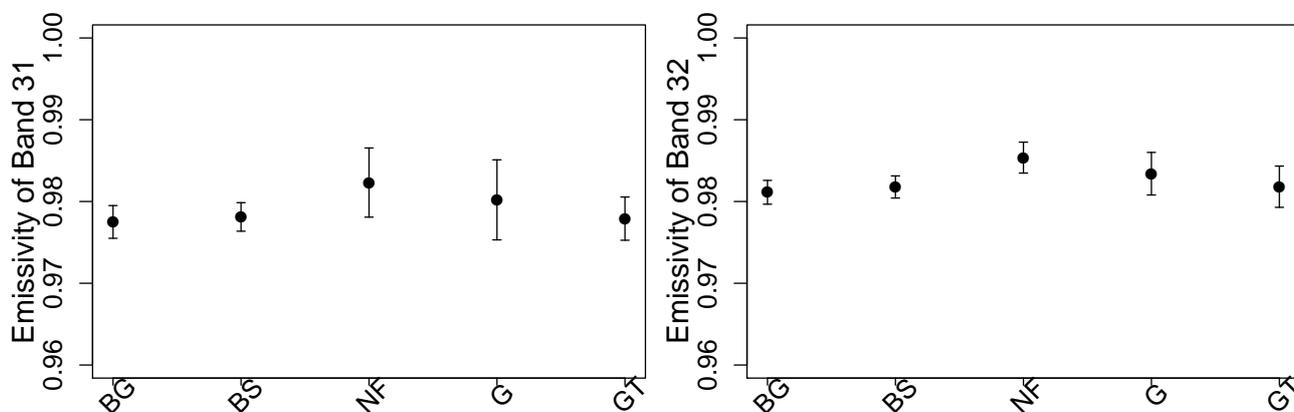


Figure 5. View zenith angles (θ) of the MODIS observations: negative values imply view from the east and positive angles are view from the west of Nadir (in total 25 observations in 14 days, dark-grey bars indicate outliers).

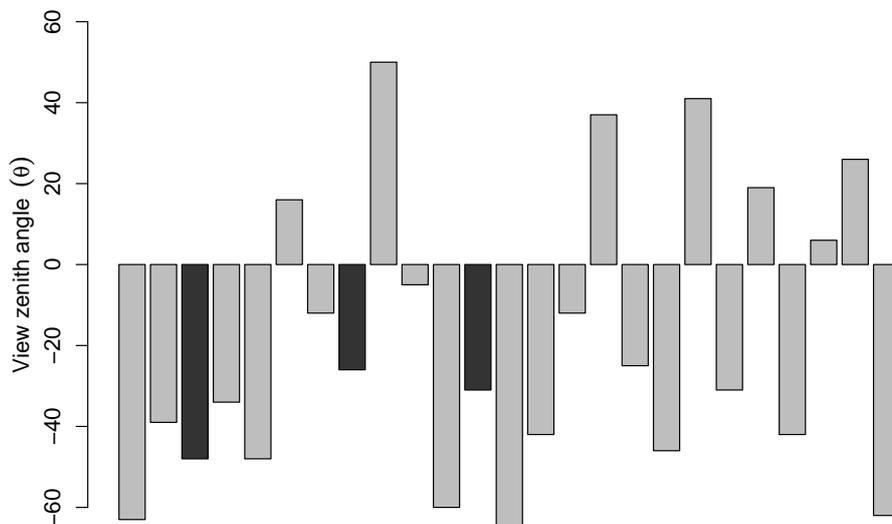
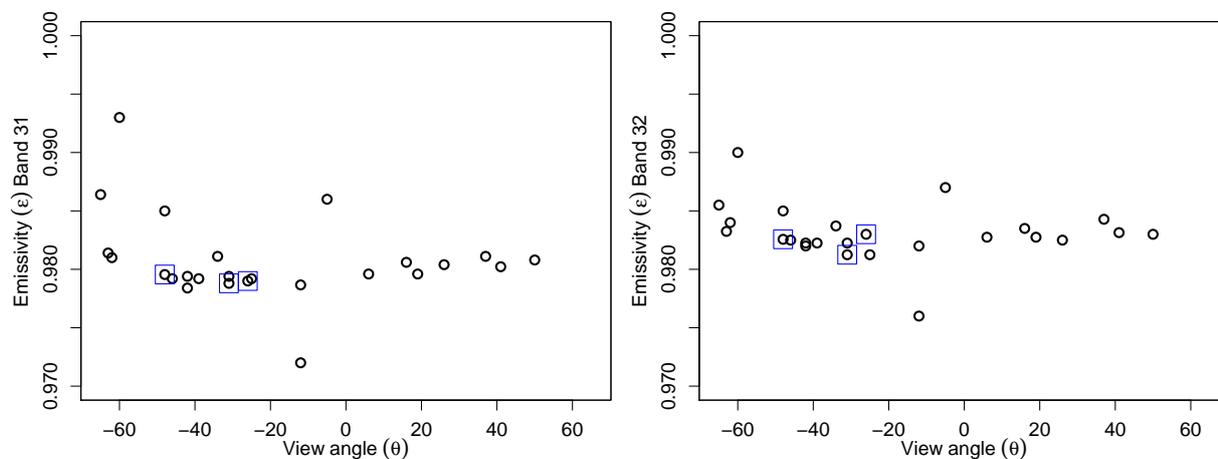


Figure 6. View zenith angles (θ) versus emissivities (ϵ) in bands 31 & 32 (outliers are marked by a square).



4.5. Geo-Statistical Analysis

Time-series of the input data sources (LST_{Modis} , LST_{wrf} and LST_{inSitu}) were constructed in order to apply correlation and regression analysis on the alternate pairs of the three datasets (listed below). Since the frequency of MODIS observations is limited to four times a day, which is further limited to cloudless days, subsets of the other two datasets were selected accordingly (depending on the latitude, more than four MODIS observations is also possible, but the L3 V5 product contains LST values from only four observations [35]). At least one series from each dataset for each LC type was constructed. Additionally, simultaneous records of all iButtons, MODIS pixels and the WRF grid points over all LC types were spatially averaged to generate three respective time-series (Equation (2)).

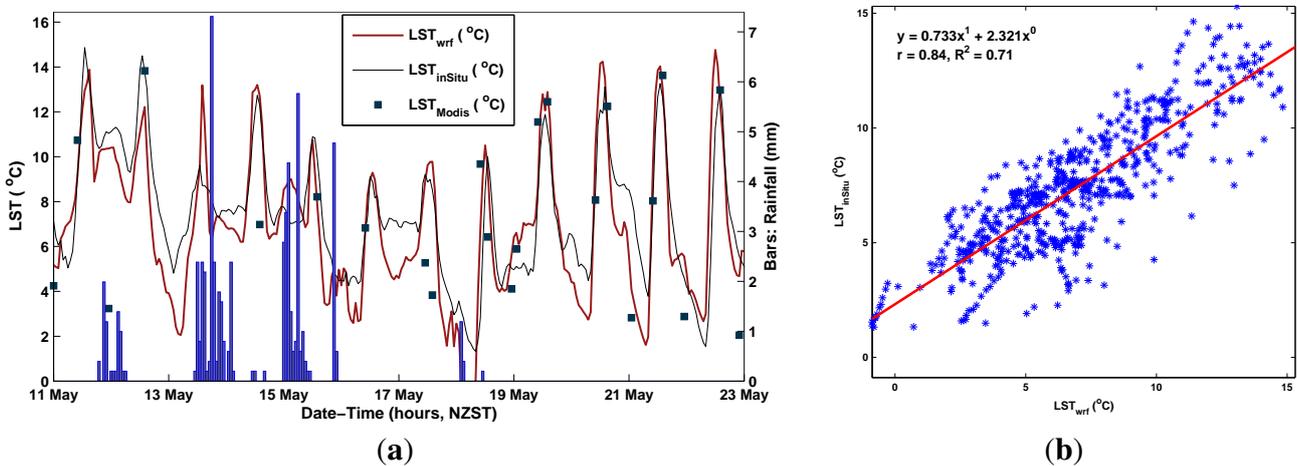
$$Ts(t) = \frac{\sum_{i=1}^n LST(i)}{n} \quad (2)$$

where Ts is the spatially averaged time-series of each LST dataset (iButtons, MODIS or WRF), t is time and n is the number of total samples (iButtons = 10, MODIS pixels = 9 and WRF grid points = 9).

Three alternate correlation calculations have been followed in the regression analysis at all times: LST_{Modis} correlated to LST_{inSitu} (denoted as MOD \sim in Situ), LST_{wrf} to LST_{inSitu} (shown as WRF \sim in Situ) and LST_{Modis} to LST_{wrf} (shown as MOD \sim WRF). The regression coefficient of determination (R^2) was used to show the strength of correlation between each of the two datasets.

We also tested how the removal of possible outliers in the measurements improves the correlation results from the regression analysis. A scatter plot by itself is a non-parametric test for the existence of the outliers [48,49]. Regression residuals' vector also can be used directly to define a cut-off margin. Any observation outside this cut-off margin is considered as an outlier. The cut-off margin was defined as $\pm 1.5\sigma \approx \pm 5$, where $\sigma = 3.57$, and is the regression's residual standard error when MODIS LST is correlated against *in situ* data (Section 5.1). This method, however, cannot determine which of the bivariate dataset was the source of the outliers, unless one or both of those datasets are correlated with a third independent variable. If one of these bivariate datasets shows the same outliers when it is correlated with a third independent variable, outliers can be further investigated in that dataset.

Figure 7. (a) Spatially averaged time-series produced from the WRF simulations and the *in situ* data (both dataset downsampled from 30-minute to hourly rate) overlaid on hourly rainfall (bars) from the UC Cass climate station; MODIS LST is also plotted as dots for those times when data were available; (b) The WRF simulations are correlated vs. *in situ* data (30-minute rate, May 2011).



5. Results

5.1. Time-Series of Spatially Averaged Datasets

In this section, the results from the correlation analysis between the spatially averaged time-series of the three datasets are presented. First, we compared spatially averaged high resolution time-series of the WRF simulations with the *in situ* measurements. Available MODIS LST observations were also overlaid on the *in situ* and WRF series (Figure 7(a)). This showed that the WRF and *in situ* datasets have captured diurnal temperature cycle very closely over the measurement period. On the other hand, MODIS LST series have shown larger bias from both WRF and *in situ* series. Hourly rainfall data from the UC Cass climate station was also analysed (Figure 7(a)). LST_{wrf} and LST_{inSitu} 30-minute data showed a relatively strong ($R^2 = 0.71$) correlation (Figure 7(b)). Afterwards, time-series of the WRF and the *in situ* measurements were filtered according to the available MODIS LST observations. The filtered time-series of the *in situ* data showed highest mean ($\mu = 8.49$) and lowest standard deviation ($\sigma = 3.26$), while the MODIS LST time-series showed the lowest mean and the highest standard deviation (Table 3). Comparison of the averaged time-series from the three datasets (which were filtered based on the available MODIS LST observations) revealed the differences in the LST at different dates (Figure 8). While model simulations generally follow the *in situ* dataset, the MODIS LST showed some significantly different observations from the *in situ* measurements (such as points 3 and 8 in Figure 8). There were a total of 25 coincident measurements over the study period, with 10 measurements over night and 15 measurements during day. During nighttime, 80% of the MODIS observations are lower compared with the *in situ* measurements. During daytime, 73% of the MODIS temperatures are similar or higher and only in 23% of the cases lower than the *in situ* data. This indicates that the lower mean LST from the MODIS time-series (Table 3) is mainly due to night-time observations. Correlation with

the *in situ* time-series with no time-lag yields $R^2 = 0.35$ for the MODIS LST and $R^2 = 0.77$ for the WRF simulations (Table 4(a)).

Figure 8. Time-series and trend-lines of LST_{inSitu} , LST_{Modis} and LST_{wrf} . Series are produced by spatially averaging points over various LC types in the study area (except for LST3, which is excluded due to its higher elevation, LST1 to 10 in Figure 1 are used to calculate the averaged LST_{Modis} series). Series only contain the values from times that were coincident with the MODIS LST observations.

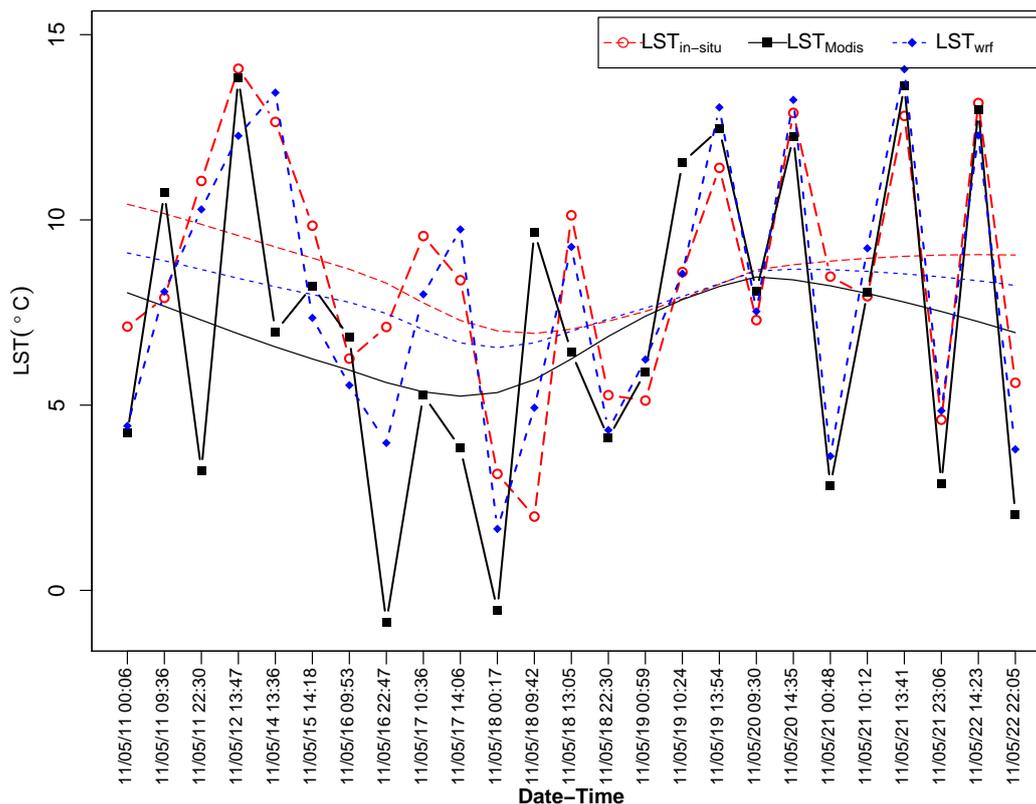


Table 3. Basic statistics from spatially averaged time-series of the three dataset.

Stat.	<i>in situ</i>	MODIS	WRF
μ	8.49	6.98	7.99
σ	3.26	4.33	3.63

Time-lags were considered to account for the delay in warming and cooling of the surface as measured by the iButtons 1–2 cm below the surface versus instantaneously observed by the satellite. Taking the MODIS acquisition time as reference, lags of ± 100 minute were applied on the *in situ* data iteratively with 1 minute intervals and the regression R^2 values were calculated for each lag. For the correlation of model simulations with the *in situ* data, lags were applied only to the *in situ* measurements. Statistics from 0, 30, 60 and 90 minutes time-lags are given in the results (Table 4).

Table 4. Regression statistics from spatially averaged data, with and without outliers. P-values from the correlations between MODIS LST and iButton series are also provided, where the smaller the **p-level**, the more significant the relationship. **(a)** Regression statistics from original data; **(b)** Regression statistics after removal of outliers.

Time-Lag	MOD~inSitu				WRF~inSitu		MOD~WRF	
	R ²	σ	p-val.	F-Stat	R ²	σ	R ²	σ
0min	0.35	3.57	0.0020	12.36	0.77	1.79	0.56	2.92
30min	0.40	3.44	0.0008	15.10	0.80	1.68	0.62	2.73
60min	0.47	3.23	0.0002	20.20	0.79	1.71	0.60	2.80
90min	0.51	3.10	0.0000	24.10	0.67	2.13	0.52	3.06

(a)

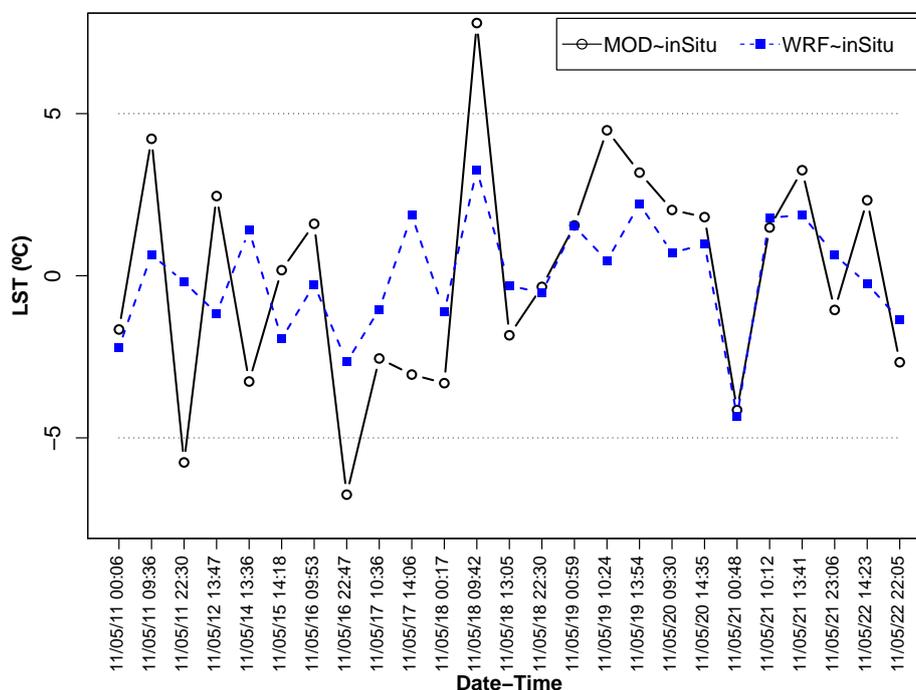
Time-Lag	MOD~inSitu				WRF~inSitu		MOD~WRF	
	R ²	σ	p-val.	F-Stat	R ²	σ	R ²	σ
0min	0.62	2.63	0.0000	32.84	0.77	1.79	0.69	2.38
30min	0.69	2.37	0.0000	45.13	0.80	1.68	0.72	2.26
60min	0.73	2.22	0.0000	54.55	0.79	1.71	0.67	2.47
90min	0.72	2.27	0.0000	51.43	0.67	2.13	0.56	2.83

(b)

Considering time-lags, correlations were generally deteriorated with lags of more than 90 minutes. The best agreement between the WRF model simulations with the *in situ* data were obtained at 30 minute time-lag while the MODIS LST did so at about 90 min time-lag (Table 4(a)). The model simulations also showed the highest agreement with the MODIS LST after adding a 30 min time-lag (Table 4(a)). It is noteworthy that the highest correlation between the MODIS LST and the *in situ* data is lower than the smallest value achieved from correlating the model simulations with the *in situ* data. Surprisingly, the MODIS LST at all times has correlated better with the model simulations than with the *in situ* data.

Several tests were made to improve the correlation values between the MODIS LST and the *in situ* data by detecting possible outliers based on the regression residuals and the scatter plots of the two datasets. Taking ± 5 ($\approx \pm 1.5\sigma$) as the upper and lower control limits for the regression residuals, three distinct outliers in the MODIS LST time-series appeared likely (Figure 9). Since the residuals from the regression between LST_{wrf} and LST_{inSitu} were well inside the cut-off margins, it turned out that the source of the outliers to be in the MODIS LST dataset. The same outliers are also visible in the scatter plots (Figure 10(a)). These points are lying farthest from the line-fit in the scatter plots of the two datasets (points bordered by a larger square in Figure 10). The first two outliers from MODIS nighttime observations were considerably colder than the measured and modelled data. The third outlier was a MODIS morning observation, which was warmer than the temperatures in the other two datasets.

Figure 9. Regression residuals' plot revealing the outliers in the MODIS LST dataset (controls are roughly equal to $\pm 1.5\sigma$, where $\sigma = 3.57$ is the residual standard error of regression when the MODIS LST time-series is correlated with the *in situ* measurements).



According to the overpass times (Figure 9), all outliers originate from the MODIS-Terra observations, and have appeared in the first half of the field experiment period. Because the view zenith angles of the outliers (48° , 26° and 31° from the East) are close or within the acceptable range of less than 45° [2], it seems unlikely that the differences are due to unfavourable imaging geometry. Emissivity values of the outliers in both bands 31 & 32 are also within the standard range (Figure 6).

We looked at relative humidity (RH) and other parameters from the model, but we did not find any meaningful pattern to explain the reason for the outliers. Since high amounts of atmospheric water vapour limits the accuracy of LST retrievals [33] and, therefore, can be a reason for the outliers, we checked MODIS Total Precipitable Water or Water Vapor (MOD05, level 2, V5.1) product. Both near-infrared (1 km pixel-size, day only) and infrared (5 km pixel-size, day and night) fields were checked, however, there was no data over the study area for those dates when the outliers have occurred. We also checked the outliers against T_a data from the UC Cass climate station. This station is close to our test-site (3 km to the South) and is approximately at a similar elevation (583 m a.s.l.). In all three cases where the MODIS LST shows an outlier, T_a data were closer to the WRF simulations and the iButton measurements. For example, at point 8 (2011/05/16 22:47 in Figure 8) the MODIS LST showed -1°C , while T_a from the UC Cass weather station was 9.3°C (average daily T_a was 7.9°C). Also, soil temperature at 10 cm depth recorded in that station was 7.4°C . These records are much closer to those from the WRF simulations and the *in situ* measurements than to the MODIS LST (Figure 8), which affirms the outliers in the latter dataset.

To prevent swamping and masking effects (see [49]), outliers were removed sequentially, followed by the calculation of the new R^2 coefficient. Removing the largest outlier (obs. 12 in Figure 9) improves R^2 from 0.35 to 0.51, removing the second largest outlier (point 8) improves R^2 to 0.53, and removing

the third outlier (point 3) improves R^2 to 0.62 (Figure 10(e)). With all the outliers removed, regression statistics were re-calculated for all time-lags (Table 5(b) and Figure 10).

Figure 10. Scatterplots with line-fits of LST_{inSitu} , LST_{Modis} and LST_{wrf} time-series, first row: all observations (outliers are bordered by a larger square), second row: outliers removed. (a) no time-lag; (b) 30 min time-lag; (c) 60 min time-lag; (d) 90 min time-lag; (e) no lag, outliers removed; (f) 30 min lag, outliers removed; (g) 60 min lag, outliers removed; (h) 90 min lag, outliers removed.

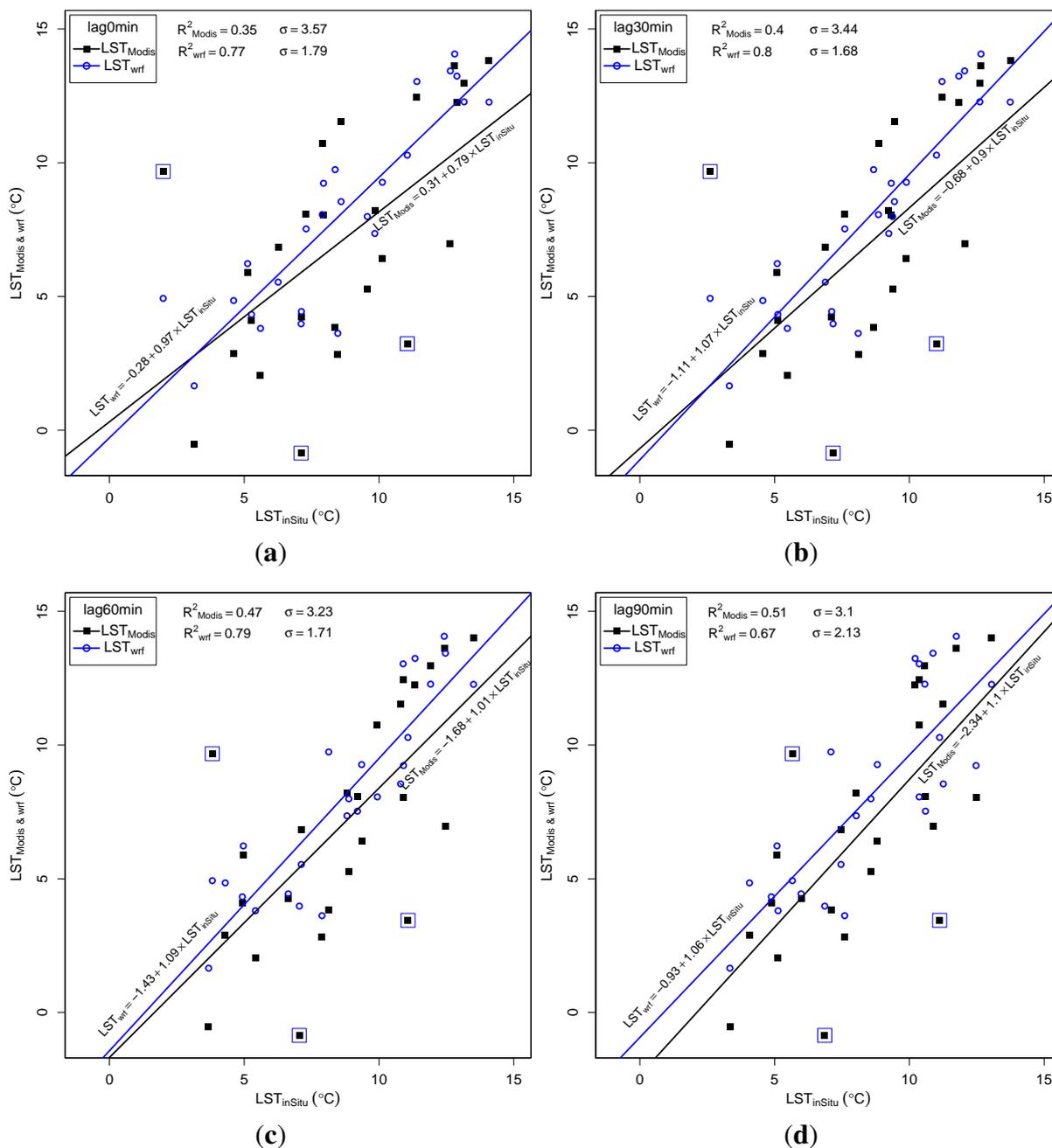
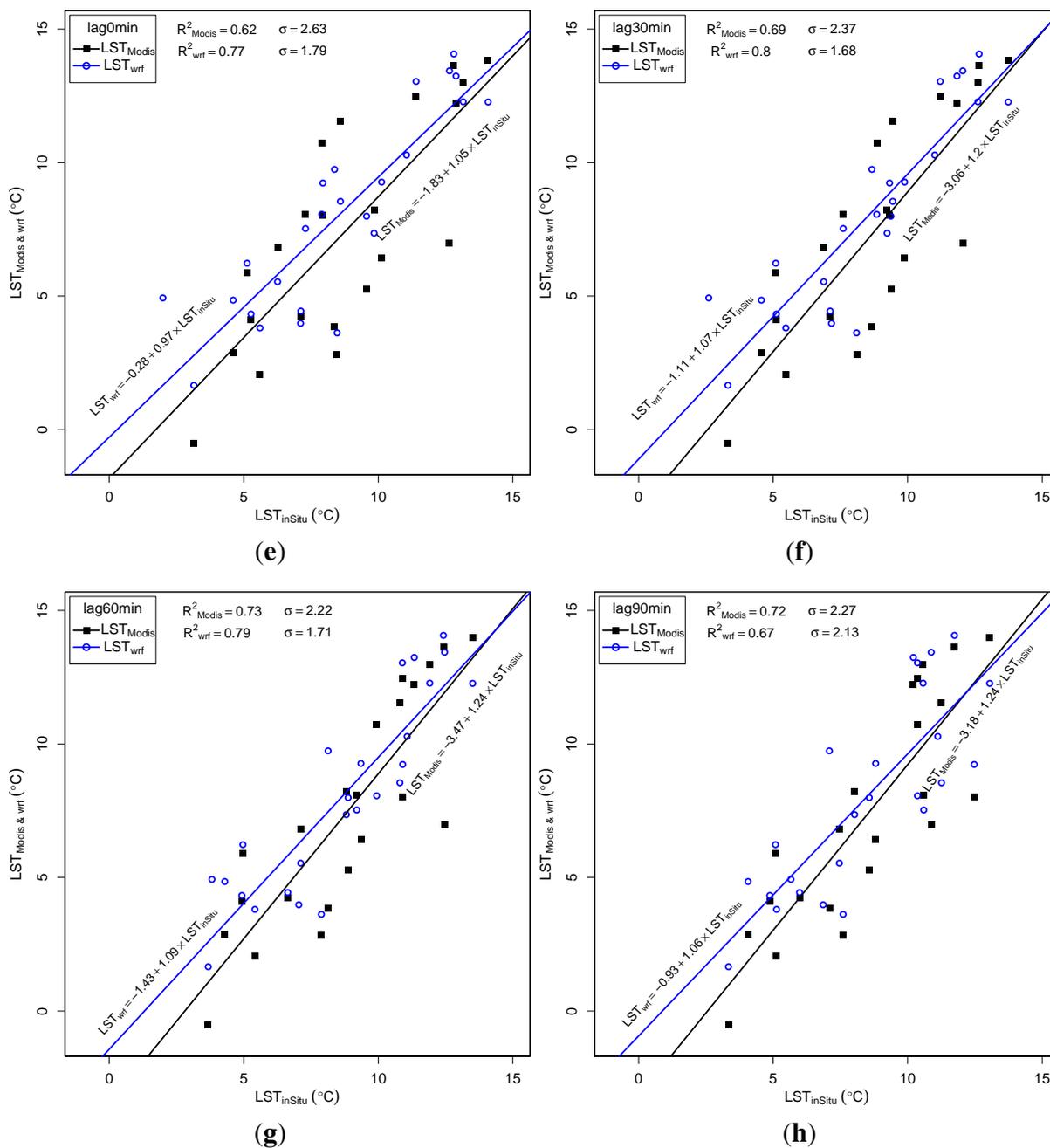


Figure 10. Cont.



5.2. Land-Cover Based Time-series

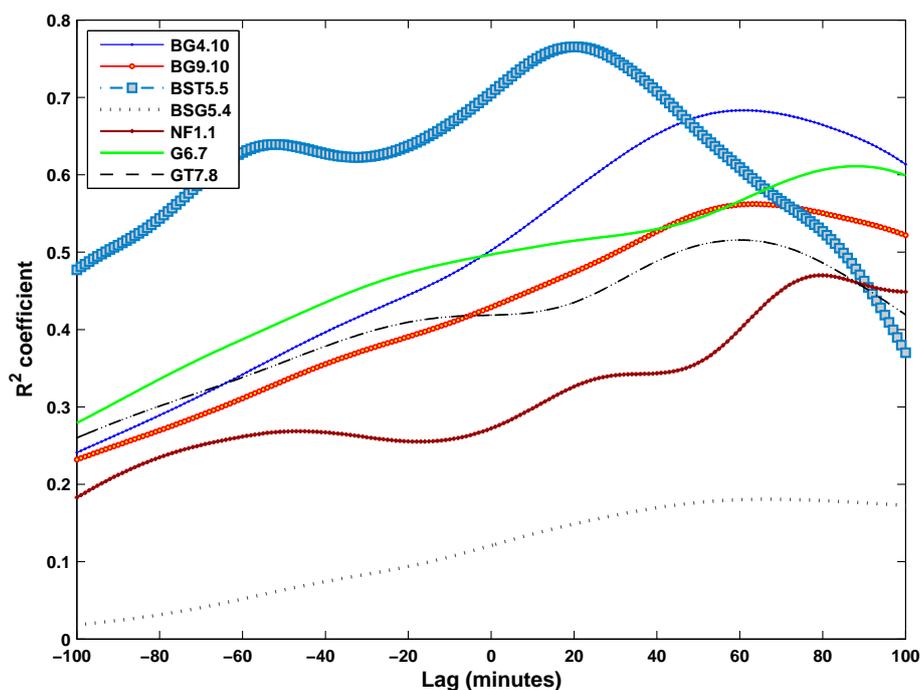
Time-series of the MODIS LST and the model simulations over 5 LC classes were correlated individually versus the corresponding series of the *in situ* data with and without time-lags (Table 5). Time-lags with a range of ± 100 minute, with 1 minute increments, were applied on the correlations between the MODIS LST and the *in situ* series over all the LC types (Figure 11). The five LC classes are Barren-Gravel (BG), Bush-Scrub (BS), Native-Forest (NF), Grassland (G) and Grass-Tussock (GT). For BG LC type there was no MODIS LST pixel overlapping the *in situ* measurement points, therefore, a non-overlapping pixel with identical LC type has been used in the analysis. The MODIS LST pixel

for the LC type NF suffered from high spectral mixing with only 25% fractional abundance (Table 2). Another pixel with almost 100% forest cover (LST3), but higher elevation, was used in the spectral mixing of this point (Figure 1).

Table 5. Regression R^2 statistics over 5 LC types, including the maximum R^2 values achieved with various time-lags (minute). Time-lags were applied in a range of ± 100 minutes, with one-minute increments. R^2 results from correlations between MODIS and *in situ* with three lags, plus the lag giving the maximum R^2 , are listed in this table. For the other two correlations, only two time-lags are given.

LC-Type/ R^2 @lag	MODIS~inSitu				WRF~inSitu		MODIS~WRF	
	0	30	60	best lag	0	30	0	30
BarrenGravel(LST4)	0.42	0.51	0.62	0.68(68)	0.71	0.75	0.73	0.75
BarrenGravel(LST9)	0.38	0.43	0.50	0.56(69)	0.71	0.75	0.43	0.46
BushScrubT(LST5)	0.51	0.62	0.66	0.77(19)	0.67	0.74	0.65	0.67
BushScrubG(LST5)	0.04	0.07	0.09	0.18(74)	0.31	0.33	0.65	0.67
Forest(LST1)	0.21	0.23	0.31	0.47(80)	0.48	0.50	0.64	0.70
Grassland(LST6)	0.42	0.46	0.48	0.61(91)	0.77	0.78	0.69	0.73
GrassTussock(LST7)	0.37	0.39	0.43	0.52(57)	0.81	0.79	0.53	0.62

Figure 11. Variations of regression R^2 statistics from correlations between iButton GST measurements over 5 LC types and time-series of the corresponding MODIS LST pixels using ± 100 minute time-lags (1 min increments). Full names of LC types in this figure are listed in Table 5, the first number is for LST pixel and the second number is for iButtons, e.g., BG4.10 stands for Barren-Gravel, LST4 correlated to b10.



Another point for consideration was the LC type BS. The height of the bush in the area is about 1 to 1.5 m from the base, with a moderate density of the scrub over the surface. Considering this point,

the authors were concerned that the values of LST recorded by MODIS observation on this particular LC type are not exactly a representative of the skin temperature of the soil, but rather affected by the temperature near the top of the bush. Although the field experiment was conducted in mid-autumn, the bush maintained considerable foliage, which could have affected temperature measurements as well as solar radiation reaching the ground. Two iButtons were placed at this site, one on top of the bush (b5) and another one on the ground (b4). The corresponding data recorded at this site are referred to as “Bush-Scrub-Ground” (BSG) and “Bush-Scrub-Top” (BST). Correlation of these measurements with the corresponding MODIS LST pixel (LST5) showed higher agreement for BST compared with BSG (Table 5). The strongest correlation from BST was observed at 19 min time-lag ($R^2 = 0.77$), with another peak at 60 min, then it dropped significantly when time-lag increased to 90 min (Figure 11 and Table 5). Also, for other LC classes correlations increased with time-lags between MODIS observations and the *in situ* measurements. For LC type BG two pixels from LST_{Modis} were analysed: LST9 with homogeneous LC and LST4 with mixed LC. Both of them showed highest agreement with the *in situ* data with ≈ 60 min time-lag; however, except for BST which is affected by the ambient T_a , LST4 showed the strongest correlation ($R^2 = 0.68$) than all LC types at all time-lags (Table 5). The maximum improvements in the correlations between LST_{wrf} and LST_{inSitu} , as well as LST_{Modis} with LST_{wrf} were observed when ≈ 30 min time-lag was applied.

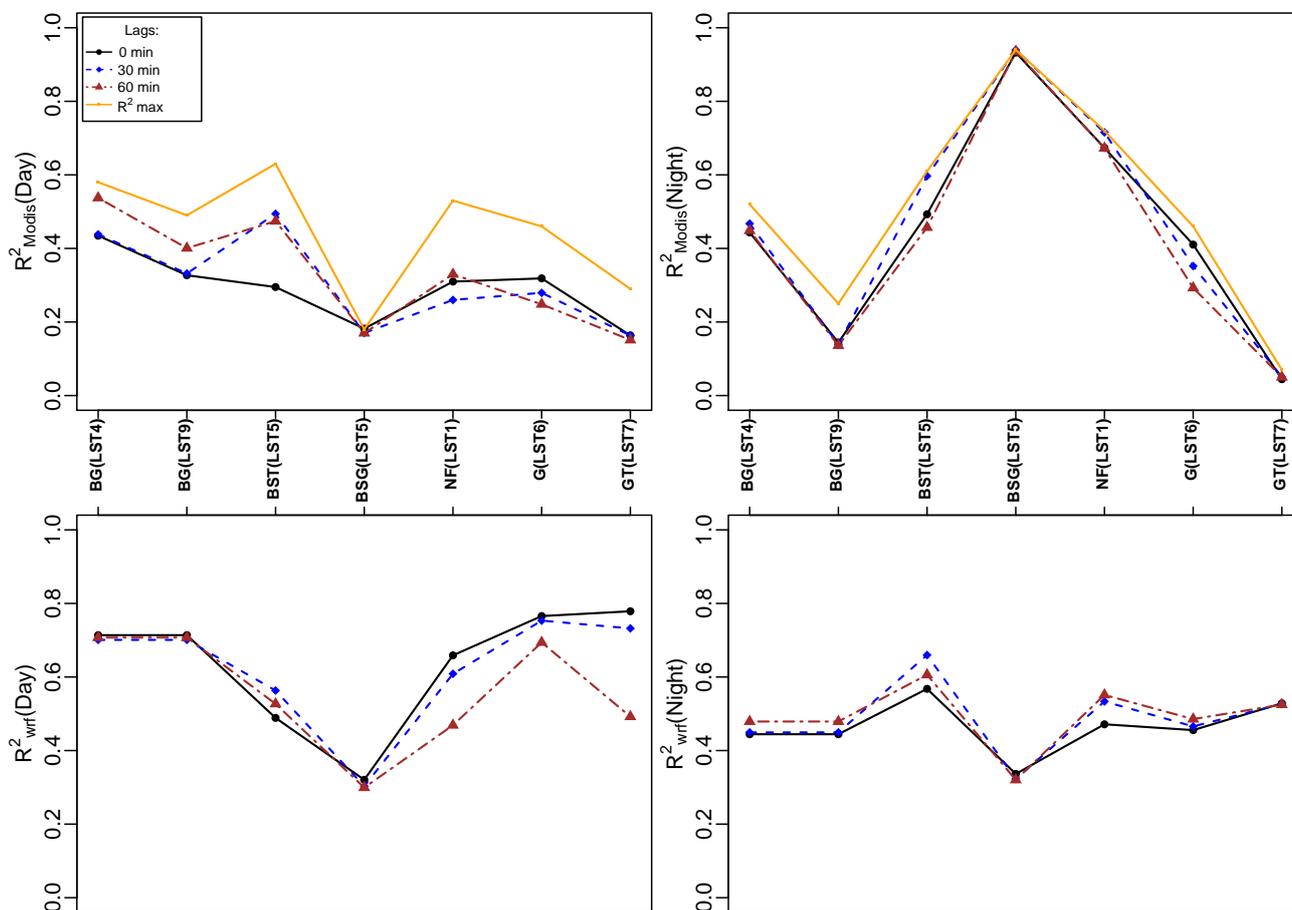
5.3. Comparison between Day and Night Measurements

In this section, the results from the regression analysis between the MODIS LST, the WRF model and the *in situ* data are presented by separating day and night observations. This analysis was necessary to discover the difference in time-lags at day/night and any possible anti-correlation due to differing surface temperature on each LC type at day or night.

Correlations between the MODIS LST and the *in situ* data are generally stronger from the night series compared with those from the day series except for LC types GT and BG. LC type BS (“Bush-scrub”) showed more complex correlation pattern from day and night observations. During daytime, higher agreement is observed between the MODIS LST and the iButton measurements on top of the bush (BST in Figure 12 with $R^2 = 0.63$) compared with the iButton on the ground (BSG in Figure 12 with $R^2 = 0.18$), whereas during night a stronger correlation ($R^2 = 0.94$) is observed at the ground (BSG in Figure 12). Barren-gravel LC type showed higher correlations over the pixel with higher mixing (LST4) both day and night. Although the other pixel in this LC class (LST9) had less mixing, correlations over that pixel are considerably lower (possible reasons can be shadowing effects of the mountain and contribution of the river water).

On the contrary, the WRF model simulations showed generally higher correlations with the *in situ* data during day compared to night. Time-lags did not improve correlations between the WRF and the *in situ* day-series (except for few cases with 30 min time-lag), however, some improvements were visible for time-lags of up to 60 min with night-series (Figure 12).

Figure 12. Correlations of day and night time-series of the three datasets over various LC types considering different time-lags (BG: Barren-Gravel, BST: Bush-Scrub-Top, BSG: Bush-Scrub-Ground, NF: Native-Forest, G: Grassland and GT: Grass-Tussock).



6. Discussion

Although the MODIS LST and the WRF model simulations are both gridded datasets, our results showed that the latter correlates better with the *in situ* measurements. We discuss first the characteristics of the *in situ* data, and then MODIS LST product over the study area as a possible explanation for this result.

6.1. Site-Specific Characteristics

The iButtons were buried very close to the surface well shielded from direct radiation. Even though they measure different physical properties, each dataset have already been adjusted to approximate the surface skin temperature. These adjustments include surface emissivity and directional effects in the MODIS LST product (see [33]), surface energy balance characterization by the WRF model [41] and the time-lag analysis on the iButton measurements. Therefore, we were expecting a close correspondence in temperatures between the MODIS LST and iButtons as well as with the model simulations. Warming or cooling at about 1–2 cm below the surface where the iButtons were buried lags behind the surface according to the soil specific heat conduction. The general pattern in the results is confirmed by

the time-lag analysis, which showed overall a significant improvement in the correlations. Time-lags improved R^2 values over different LC types at least 14% (such as GT) and up to 26% (such as BG in Table 5). Due to higher complexity over NF and BS LC types, more *in situ* measurements at different heights of the forest and bush canopy is needed to characterize the temperature measured by the MODIS from top of the canopy. Such an effort also needs to consider density of the bush and/or the forest, differences in height and even different species of the canopy. As a result, we took a simplistic and practical approach in the field experiment and, therefore, we give higher significance for the results over G, GT and BG LC types.

Consideration of geographic characteristics of the test-site also helps to understand the difference between the MODIS LST and iButton measurements, which is discussed below as how it affects the LST retrieval process.

6.2. Issues in LST Retrieval

Uncertainties involved in the retrieval of LST can be related to (i) geographic characteristics of the test-site; (ii) the emissivity of land surface types and (iii) the atmospheric corrections (see [3]). Other sources of uncertainty include view-angle effect [2].

(i) The study area is constrained inside a mountainous region, which can be a source of a bias in the MODIS LST at two stages: the initial local effects on the observations upon the sensor's overpass and the second, the effects of local orography on the LST retrieval algorithm. The effects of the rugged terrain on the view angle and multiple scattering on the observations cannot be avoided, so as the effects of local orography and scattered or sub-pixel clouds on the visibility if it is not as strong as to be eliminated in cloud masking. The second outlier in the MODIS LST (point 8 in Figure 8), which showed unusually colder value than the other datasets, can be explained as cloud-top temperature skipped during cloud masking. Cloud-top temperatures in the LST dataset typically appear in the negative °C range (see [6]). The issue of variability in view angle, as well as overpass time, had been demonstrated in LST- T_a analysis by [13]. Orography and view angle had been assessed in the cross-sensor analysis of LST in [4], where they have identified uncertainties in LST retrievals due to (1) satellite viewing angle (2) surface orography and (3) surface LC types. Synoptic weather effects on LST retrievals can be another source of uncertainties. Based on daily rainfall data from the UC Cass climate station, the first few days of the measurement period has been rainy (Figure 7(a)). The agreement between the MODIS LST and the other two datasets during the rainy episode (11–15 May) has been poor (Figure 8), while it improves during the subsequent days. The first outlier in the MODIS LST dataset coincides with a moderate rainfall event recorded by the UC Cass climate station, which starts at late afternoon on 11 May and continues till the next morning (Figure 7(a)). It is suspected that the rainfall events with higher atmospheric water content and patchy clouds during and shortly after rainy days have affected the accuracy of LST retrievals.

(ii) Emissivity and view angle effects: the issues with the accuracy of the model employed for retrieval of LST over regions where the emissivity is highly variable is indicated in the MODIS LST theoretical document [33]. Our analysis on the emissivity (emis31 & emis32) and view angle fields, which are accompanied with the LST product, showed that variations in the emissivities of all LC types were less than 0.01 for view angles under $\pm 55^\circ$ (for view angles exceeding $\pm 55^\circ$ the emissivity variations were

higher than 0.01). Although these emissivities are output from (not input to) the retrieval algorithm, it can be interpreted that the emissivity has been well defined prior to the retrieval procedure (see [33] for a description of the input emissivities in the GSW algorithm).

(iii) Atmospheric effects: with respect to the LST product's algorithm description document, the ultimate quality of the atmospheric adjustments will depend on the quality of temperature and water vapour profile retrievals in the lower troposphere up to 9 km [33]. In this respect, we showed that the agreement between the MODIS LST and the other two datasets was poor during and shortly after rainy days. Patchy clouds after intermittent rainfall events, if too small to be detected by the cloud mask, can cause problems in the modelled atmospheric profiles. Input atmospheric profiles used in the V5 LST product include the MODIS Atmospheric Profiles (MOD07) product [20]. Information of the atmospheric lower boundary temperature, or T_a , provided in the MODIS atmospheric product is also used to improve LST retrieval accuracy [27]. Results from comparison of the MODIS LST with the two iButtons, one on the top of the bush and another on the ground, can be related to the effects of the lower boundary T_a on LST retrievals. It must be noted that the iButton on the bush had been strongly affected by the near-surface T_a . Bush canopy provides protection against solar heat during day and excess emission during night, hence the iButton on the ground might have been negatively or positively biased at day and night, respectively (this can be interpreted from GST variations measured by BushG iButton in Figure 3). When the MODIS LST correlates better with the iButton on the ground protected by the bush, it shows lowest correlation with the exposed BG LC type. It is suspected that strong terrestrial emission occurs from the bare-soil skin at night in mid-Autumn (when field experiment was conducted), while T_a is relatively warmer than the surface skin. Since the MODIS LST correlates better with the warmer temperatures inside the bush at night, it is suspected that the LST product actually has a bias towards T_a at night. Similarly, higher day-time correlations of LST with the iButton on top of the bush, as well as bare-soil (which has faster heating rate), is an indication of the fact that the MODIS LST correlates better with T_a during day.

Consequently, there is the possibility that the algorithm used in the extraction of MODIS LST does not perform well in the mountainous regions, where the solar radiation regimes are different (see [50]), and the atmospheric profiles of temperature and water vapour can vary in short distances.

7. Conclusions

The MODIS LST product over a mountainous region in the Southern Alps of New Zealand was analysed in comparison with the *in situ* data and the modelled LST for the same region. Results showed a relatively significant ($R^2 = 0.35$, F-statistics = 12.36, p-value = 0.0020, 99% confidence) correlation between the MODIS LST and the *in situ* data, while at the same time a relatively strong correlation ($R^2 = 0.77$, 99.9% confidence) between the model simulations and the *in situ* measurements. It also became evident that the MODIS LST over some of the LC types has higher agreement with the *in situ* data, while for the other LC types the agreement is relatively poor. A few outliers in the spatially averaged time-series of the MODIS LST were detected using parametric and non-parametric statistical techniques. When these outliers were removed from the regression analysis, results were significantly improved. Despite these improvements as a result of adding time-lags and outliers removal, model

correlations ($R^2 = 0.80$ with 99.9% confidence) were higher than the MODIS LST ($R^2 = 0.73$ with 99.9% confidence). It is, therefore, concluded that the MODIS LST, if assimilated in the WRF model without any prior assessment for the outliers and the local effects, will not provide any improvement for LST simulations over an alpine region. Longer time-series, however, are required to draw more robust conclusions about the applicability of the MODIS LST product for improving WRF simulations over alpine complex terrain. We suggest outliers in time-series of the MODIS LST to be investigated based on *in situ* measurements (if available) and climate data, especially for areas where the terrain elevation is variable. Rainfall events, and the inherent patchy clouds during and shortly after rainy days, turned out to be the main cause of the outliers. Time-lag between the instantaneous satellite observations and the *in situ* measurements of contact temperature should be also taken into account when the latter is used as reference. In this way, we found that the temperature measured slightly below the surface by iButtons (especially over the grassland) compares well with the model simulations. This suggests that the *in situ* measurements made by these low-cost instruments may serve as a proxy for the remotely sensed and simulated LST after applying a time-lag accounting for the heat capacity of the soil.

Acknowledgments

This research is funded by the University of Canterbury (UC) in New Zealand. We acknowledge free access to the MODIS LST product provided by the US NASA. Warehouse Inventory Search Tool (WIST) and Reverb were used to download the LST product. LC satellite imagery data were obtained from US Geological Survey (USGS) publicly available web resources via Earth Explorer. The authors wish to acknowledge Rob Agnew from the New Zealand Institute for Plant & Food Research for his iButton data-loggers. We also express our thanks to Justin Harrison for his help and expert advice on the field work. We appreciate the valuable comments from four anonymous reviewers, which significantly improved this manuscript.

References

1. Wan, Z.; Dozier, J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 892–905.
2. Jin, M. Analysis of land skin temperature using AVHRR observations. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 587–600.
3. Coll, C.; Caselles, V.; Galve, J.M.; Valor, E.; Nicolòs, R.; Sánchez, J.M.; Rivas, R. Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data. *Remote Sens. Environ.* **2005**, *97*, 288–300.
4. Trigo, I.; Monteiro, I.; Olesen, F.; Kabsch, E. An assessment of remotely sensed land surface temperature. *J. Geophys. Res.* **2008**, *113*, D17108.
5. Tang, B.; Bi, Y.; Li, Z.L.; Xia, J. Generalized split-window algorithm for estimate of land surface temperature from chinese geostationary fengyun meteorological satellite (FY-2C) data. *Sensors* **2008**, *8*, 933–951.
6. Neteler, M. Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. *Remote Sens.* **2010**, *2*, 333–351.

7. Ghent, D.; Kaduk, J.; Remedios, J.; Ardö, J.; Balzter, H. Assimilation of land surface temperature into the land surface model JULES with an ensemble Kalman filter. *J. Geophys. Res.* **2010**, *115*, D19112.
8. Trigo, I.; Dacamara, C.; Viterbo, P.; Roujean, J.; Olesen, F.; Barroso, C.; Camacho-de Coca, F.; Carrer, D.; Freitas, S.; Garcia-Haro, J.; *et al.* The satellite application facility for land surface analysis. *Int. J. Remote Sens.* **2011**, *32*, 2725–2744.
9. Dash, P.; Göttsche, F.M.; Olesen, F.S.; Fischer, H. Separating surface emissivity and temperature using two-channel spectral indices and emissivity composites and comparison with a vegetation fraction method. *Remote Sens. Environ.* **2005**, *96*, 1–17.
10. Huang, C.; Li, X.; Lu, L. Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sens. Environ.* **2008**, *112*, 1320–1336.
11. Sheng, J.; Wilson, J.P.; Lee, S. Comparison of land surface temperature (LST) modeled with a spatially-distributed solar radiation model (SRAD) and remote sensing data. *Environ. Model. Softw.* **2009**, *24*, 436–443.
12. Meng, C.L.; Li, Z.L.; Zhan, X.; Shi, J.C.; Liu, C.Y. Land surface temperature data assimilation and its impact on evapotranspiration estimates from the Common Land Model. *Water Resour. Res.* **2009**, *45*, W02421.
13. Mostovoy, G.V.; King, R.L.; Reddy, K.R.; Kakani, V.G.; Filippova, M.G. Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of mississippi. *Gisci. Remote Sens.* **2006**, *43*, 78–110.
14. Colombi, A.; Michele, C.D.; Pepe, M.; Rampini, A. Estimation of daily mean air temperature from MODIS LST in alpine areas. *EARSeL eProc.* **2007**, *6*, 38–46.
15. Zhang, W.; Huang, Y.; Yu, Y.; Sun, W. Estimation of daily mean air temperature from MODISLST in alpine areas. *Int. J. Remote Sens.* **2011**, *32*, 1–26.
16. Wang, C.; Qi, J.; Moran, S.; Marsett, R. Soil moisture estimation in a semiarid rangeland using ERS-2 and TM imagery. *Remote Sens. Environ.* **2004**, *90*, 178–189.
17. Hain, C.R.; Crow, W.T.; Mecikalski, J.R.; Anderson, M.C.; Holmes, T. An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. *J. Geophys. Res.* **2011**, *116*, D15107.
18. Cheng, K.; Su, Y.; Kuo, F.; Hung, W.; Chiang, J. Assessing the effect of landcover changes on air temperature using remote sensing images-A pilot study in northern Taiwan. *Landsc. Urban Plan.* **2008**, *86*, 297–297.
19. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z. Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote Sens.* **2004**, *25*, 261–274.
20. Wan, Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sens. Environ.* **2008**, *112*, 59–74.
21. Wan, Z.; Li, Z. Radiance-based validation of the V5 MODIS land-surface temperature product. *Int. J. Remote Sens.* **2008**, *29*, 5373–5395.
22. Kabsch, E.; Olesen, F.; Prata, F. Initial results of the land surface temperature (LST) validation with the Evora, Portugal ground-truth station measurements. *Int. J. Remote Sens.* **2008**, *29*, 5329–5345.

23. Coll, C.; Galve, J.; Sanchez, J.; Caselles, V. Validation of landsat-7/etm+ thermal-band calibration and atmospheric correction with ground-based measurements. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 547–555.
24. Niclòs, R.; Galve, J.M.; Valiente, J.A.; Estrela, M.J.; Coll, C. Accuracy assessment of land surface temperature retrievals from MSG2-SEVIRI data. *Remote Sens. Environ.* **2011**, *115*, 2126–2140.
25. Karnieli, A.; Agam, N.; Pinker, R.T.; Anderson, M.; Imhoff, M.L.; Gutman, G.G.; Panov, N.; Goldberg, A. Use of NDVI and land surface temperature for drought assessment: merits and limitations. *J. Clim.* **2010**, *23*, 618–633.
26. Mostovoy, G.V.; Anantharaj, V.; King, R.L.; Filippova, M.G. Interpretation of the relationship between skin temperature and vegetation fraction: Effect of subpixel soil temperature variability. *Int. J. Remote Sens.* **2008**, *29*, 2819–2831.
27. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sens. Environ.* **2002**, *83*, 163–180.
28. Rigo, G.; Parlow, E. Modelling the ground heat flux of an urban area using remote sensing data. *Theor. Appl. Climatol.* **2007**, *90*, 185–199.
29. Holden, Z.A.; Abatzoglou, J.T.; Luce, C.H.; Baggett, L.S. Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain. *Agric. Forest Meteorol.* **2011**, *151*, 1066–1073.
30. Fitzgerald, W.B.; Fahmy, M.; Smith, I.J.; Carruthers, M.A.; Carson, B.R.; Sun, Z.; Bassett, M.R. An assessment of roof space solar gains in a temperate maritime climate. *Energy Build.* **2011**, *43*, 1580–1588.
31. Brooks, R.T.; Kyker-Snowman, T.D. Forest floor temperature and relative humidity following timber harvesting in southern New England, USA. *Forest Ecol. Manag.* **2008**, *254*, 65 – 73.
32. Gubler, S.; Fiddes, J.; Gruber, S.; Keller, M. Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain. *Cryosphere Discuss.* **2011**, *5*, 307–338.
33. Wan, Z. *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document: LST ATBD*; Institute for Computational Earth System Science, University of California: Santa Barbara, CA, USA, 1999.
34. Wan, Z.; Zhang, Y.; Li, Z.; Wang, R.; Salomonson, V.; Yves, A.; Bosseno, R.; Hanocq, J. Preliminary estimate of calibration of the moderate resolution imaging spectroradiometer thermal infrared data using Lake Titicaca. *Remote Sens. Environ.* **2002**, *80*, 497–515.
35. Wan, Z. *Collection-5 MODIS Land Surface Temperature Products Users' Guide*; ICES, University of California: Santa Barbara, CA, USA, 2009.
36. Wolfe, R.E.; Saleous, N. 7 MODIS Land Products and Data Processing. In *Earth Science Satellite Remote Sensing Volume 1: Science and Instruments*; Qu, J.J., Gao, W., Kafatos, M., Murphy, R.E., Salomonson, V.V., Eds.; Tsinghua University Press-Springer: Beijing, China, 2006; pp. 110–122.
37. NCAR. *ARW Version 3 Modeling System User's Guide*; Mesoscale & Microscale Meteorology Division, National Center for Atmospheric Research: Boulder, CO, USA, 2011.

38. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.Y.; Wang, W.; Powers, J.G. *NCAR Technical Note: A Description of the Advanced Research WRF Version 3*; Mesoscale & Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, CO, USA, 2008.
39. Maussion, F.; Scherer, D.; Finkelnburg, R.; Richters, J.; Yang, W.; Yao, T. WRF simulation of a precipitation event over the Tibetan Plateau, China - an assessment using remote sensing and ground observations. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1795–1817.
40. Jin, J.; Miller, N.L.; Schlegel, N. Sensitivity study of four land surface schemes in the wrf model. *Adv. Meteorol.* **2010**, *2010*, 167436:1–167436:11.
41. Chen, F.; Dudhia, J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.* **2001**, *129*, 569–585.
42. Dudhia, J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.* **1989**, *46*, 3077–3107.
43. Averbuch, A.; Zheludev, M. Two linear unmixing algorithms to recognize targets using supervised classification and orthogonal rotation in airborne hyperspectral images. *Remote Sens.* **2012**, *4*, 532–560.
44. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57.
45. Pu, R.; Gong, P.; Michishita, R.; Sasagawa, T. Spectral mixture analysis for mapping abundance of urban surface components from the Terra/ASTER data. *Remote Sens. Environ.* **2008**, *112*, 939–954.
46. Souza, C., Jr; Firestone, L.; Silva, L.M.; Roberts, D. Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models. *Remote Sens. Environ.* **2003**, *87*, 494–506.
47. Wang, L.; Uchida, S. Use of linear spectral mixture model to estimate rice planted area based on MODIS data. *Rice Sci.* **2008**, *15*, 131–136.
48. Gnanadesikan, R.; Kettenring, J.R. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **1972**, *28*, 81–124.
49. Ben-Gal, I. Outlier Detection. In *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*; Maimon, O., Rockach, L., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2005; pp. 1–16.
50. Proy, C.; Tanré, D.; Deschamps, P. Evaluation of topographic effects in remotely sensed data. *Remote Sens. Environ.* **1989**, *30*, 21–32.