

Using Satellite Land Surface Temperature to Parameterize Sub-grid Tiling Schemes and
Enable Tile-level Calibration of Land Surface Model

by

Jiaxuan Cai

Department of Civil and Environmental Engineering
Duke University

Date: _____

Approved:

Nathaniel Chaney, Advisor

Andrew D. Bragg

Wenhong Li

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School
of Duke University

2021

ABSTRACT

Using Satellite Land Surface Temperature to Parameterize Sub-grid Tiling Schemes and
Enable Tile-level Calibration of Land Surface Model

by

Jiaxuan Cai

Department of Civil and Environmental Engineering
Duke University

Date: _____

Approved:

Nathaniel Chaney, Advisor

Andrew D. Bragg

Wenhong Li

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School
of Duke University

2021

Copyright by
Jiaxuan Cai
2021

Abstract

To better simulate terrestrial energy and water processes, great efforts have been made to improve the representation of spatial heterogeneity in land surface models. HydroBlocks, a field-scale resolving land surface model, was developed to address this challenge while minimizing increases in computational demands by employs a new tiling framework – the hierarchical multivariate clustering (HMC) approach. The HMC is used to cluster grid cells with similar surface characteristics into hydrologic response units (HRUs, i.e., sub-grid tiles) using high-resolution data. HRUs model land surface processes separately and HRUs are connected through subsurface flow. Hence, the HRU configuration plays a critical role in the model performance. Furthermore, since the outputs for each HRU can be mapped in space, it is possible to explicitly calibrate the model at tile level based on satellite-derived data. To investigate the advantages that satellite observations can have to tune HydroBlocks, the model was set up to run from 2013 to 2019 at the 30-meter spatial resolution and hourly temporal resolution over the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP), U.S. This study demonstrated that HydroBlocks benefits from considering the temporal mean and standard deviation map of Landsat LST when generating HRUs. The effects of inputting the albedo, emissivity, and leaf area index derived directly from MODIS into the model were also explored. Finally, six soil and vegetation parameters in HydroBlocks were

calibrated by maximizing the linear correlation coefficient of the time-series of simulated LST and observed MODIS LST at each HRU. The regional-level and site-level evaluation indicate the effectiveness of the calibration approach.

Contents

Abstract	iv
List of Tables.....	viii
List of Figures.....	ix
1. Introduction.....	1
1.1 Background and Motivation.....	1
1.2 Objectives	8
2. Methodology	10
2.1 Overview of the HydroBlocks land surface model.....	10
2.1.1 Generation scheme of hydrologic response units	11
2.1.2 Parameters considered in the calibration	12
2.2 Study Area.....	17
2.3 Data	19
2.3.1 Satellite Retrievals.....	19
2.3.2 Primary Model Inputs	22
2.3.3 In Situ Measurements.....	23
2.4 Model Experiments.....	25
2.4.1 Convergence Analysis Experiment	25
2.4.2 Intra-band Clustering with Land Surface Temperature.....	26
2.4.3 Replacement of Albedo, Emissivity, and LAI.....	28
2.4.4 Parameter Calibration with Latin Hypercube Sampling	28

3. Results	30
3.1 Convergence Analysis	30
3.2 Intra-band clustering with LST	35
3.3 Performance with MODIS Parameters.....	41
3.4 Tile-level Calibration	41
4. Discussion	45
4.1 Heterogeneity Representation.....	45
4.2 Problems of Using MODIS Parameters.....	47
4.3 Uncertainties of Calibration.....	49
5. Conclusions.....	53
Appendix A	54
Appendix B.....	55
References	57

List of Tables

Table 1: Summary of some LST instruments commonly used to generate high-resolution LST data.....	8
Table 2: Parameters selected for calibration	16
Table 3: HMC parameter sets used for convergence analysis.....	26

List of Figures

Figure 1: The study area consists of a 1.0-degree box in central northern Oklahoma and southern Kansas. The Google satellite map, DEM, and land cover of this domain are presented. The location of Atmospheric Radiation and Measurement (ARM) Southern Great Plains (SGP) Central Facility (site E14) is shown in the DEM map.	18
Figure 2: Time series of measured latent heat flux of site E14 and E37 after quality control.....	25
Figure 3: Simulations of annual mean land surface temperature over the study area. Each panel shows the simulation run with different parameter sets of the hierarchical multivariate clustering scheme in HydroBlocks, and the total number of HRUs is shown as the title. The specific parameters of each simulation are shown in Table 3.	33
Figure 4: Annual mean of the spatial mean and spatial standard deviation of land surface temperature (LST), latent heat flux (LE), and sensible heat flux (H) for nine different HRU configurations.	34
Figure 5: Temporal mean and standard deviation of Landsat land surface temperature.	37
Figure 6: The domain is separated into five characteristic basins (a). Height bands are determined within each characteristic basin. For characteristic basin 1, six height bands are generated based on the setting that the ratio between the area of a height band and its adjacent height band below it is 2 (b). Within each height band, the grid cells are clustered by latitude, longitude, and DEM (HRUs-Baseline); and latitude, longitude, DEM, temporal mean and standard deviation of Landsat LST (HRUs-LST), respectively. HRU configuration is done throughout the domain, but only the characteristic basin 1 is shown to present the details better.	38
Figure 7: Comparison of the distribution of correlation coefficients (R) between 189 MODIS LST maps and corresponding mapped LST simulations of model run with (a) HRUs-Baseline, (b) HRUs-LST, and (c) HRUs-LST_MODIS.....	39
Figure 8: Probability distribution of correlation coefficients (R) between MODIS LST maps and LST simulations.	39

Figure 9: Model simulations versus ARM observations, (a) land surface temperature (LST), (b) sensible heat flux (H), and (c) latent heat flux (LE). R and KGE are shown in each panel for each variable.40

Figure 10: The range of the correlation coefficients between time series of simulated and observed LST for each HRU using 50 parameter sets. The performance of the model run with default and calibrated parameters is also shown.43

Figure 11: Comparison of the spatial performance between default and calibrated model. The linear correlation of pixel values for mapped LST simulation and MODIS image is calculated and the total 189 correlation coefficients (R) are summarized into probability distribution.43

Figure 12: Site-level comparison between model simulated (default and calibrated) and observed (a) land surface temperature (LST), (b) sensible heat flux (H), and (c) latent heat flux (LE). R and KGE are shown in each panel for each variable.44

Figure 13: Time series of default and MODIS emissivity and LAI used in the model at..48

Figure 14: Examples of MODIS LST images and the remapped LST based on that of each HRU estimated from the original data.52

1. Introduction

1.1 *Background and Motivation*

Terrestrial energy and water fluxes play a pivotal role in the Earth's climate system. Monitoring, simulation, and prediction of land surface components and processes are important for water resources management, food production, as well as weather forecasts and warnings (Seneviratne et al., 2006). One key source of uncertainty in hydrologic prediction remains the parameterization of spatial heterogeneity. Spatial heterogeneity impacts and is impacted by surface features at different spatial scales (e.g., the overlying atmosphere, the surface roughness, the type and density of vegetation cover, and the soil physical properties). It has profound influences on the water and energy cycles over the land surface (Bierkens et al., 2015; Brunsell et al., 2011; Clark et al., 2015; Friedl, 2002; Mahrt, 2000; Richardson et al., 2013). As a result, there have been concerted efforts to improve its representation in Earth system models (ESMs) since the 1980s.

Over the past decades, great efforts have been made to account for sub-grid land surface heterogeneity in large-scale models via parameterizations (Avissar & Pielke, 1989; Koster & Suarez, 1992; Liang et al., 1994; Franks & Beven, 1997; Peters-Lidard et al., 1998). There have been many attempts, including "mosaic" approaches and "effective parameter" approaches (Rouholahnejad Freund et al., 2020). These approaches,

however, are typically not able to capture fine-scale land surface processes (e.g., runoff generation, and hillslope effect on infiltration and drainage) or fully represent the water and energy transfers between the atmosphere and surface due to the too coarse spatial resolution ranging from 10 to 100 km (Singh et al., 2015). For example, Aminzadeh & Or (2017) found that to better estimate the evapotranspiration over the heterogeneous area covered by bare soil and vegetation patches, the heterogeneity length scale should not exceed the thickness of convective planetary boundary layer. Singh et al. (2015) reported that with the same 1 km soil and topographic information, running the National Center for Atmospheric Research Community Land Model (CLM v4.0) at finer-resolution (from 100 to 25 to 1 km) over the Southwestern U.S. leads to significant improvements in the simulation of soil moisture, sensible heat, and snow water equivalent as well.

To address this weakness, land surface models (LSMs) with finer resolution are required. The fact that higher-resolution modeling contributes to a more realistic simulation of the terrestrial surface has already been demonstrated (Kollet & Maxwell, 2008; Maxwell, 2010; Singh et al., 2015; Wood et al., 2011). Consequently, with the development of computational power and the availability of high-resolution forcing data, more and more hyper-resolution (i.e., 1 km or finer) LSMs (Wood et al., 2011) have been performed at regional scales. When it comes to catchment scales, even higher resolutions have been applied (Kumar et al., 2006). Of course, it would be perfect if we

can model the land surface with every meter-scale grid cell over the globe. Unfortunately, it remains impossible in the near future due to the excessive computational demands (i.e., CPU time and storage requirements) (Bierkens et al., 2015; Kollet et al., 2010), the epistemic uncertainty (Beven & Cloke, 2012), and the lack of global data with such high resolution. However, improving the surface heterogeneity in land surface models can effectively reduce the computational demands. As a result, there is renewed effort focusing on better representing spatial heterogeneity in high-resolution Earth system models (Chaney et al., 2020; Clark et al., 2015; Newman et al., 2014).

Calibration is another problem in land surface modeling. It is worthwhile to note that calibration is necessary for LSMs to accurately replicate critical land surface processes (Beven, 1993; Troy et al., 2008). Currently, land surface models rely heavily on a lengthy list of physically conceptualized parameters that are globally applicable or associated with the environment, such as land cover and soil classifications. The predefined parameter values, however, are often inappropriate for the climatic conditions of the study area or not measured at the applied modeling scales, and thus the physical meaning is lost (Hou et al., 2015; Rosero et al., 2010; Santanello et al., 2013). Hence, attempts have been made to calibrate these parameters that are full of uncertainties.

In the past, calibration of land surface models was conducted solely based on in situ measurements such as flux towers, soil moisture probes, or stream gauges. In this case, the model could work accurately after calibration at point scale while the model performance over the region remains unknown. For instance, when modeling the watersheds, flow is usually calibrated based on one monitoring site at the watershed outlet. However, many arid and semi-arid watersheds are not continuous in most cases. The outlet's flow characteristics, therefore, do not represent the entire watershed though the simulation might be appealing (Niraula et al., 2012). In addition, measurements from eddy-covariance flux towers normally represent an area of 100 m² to 1 km² varying with the weather condition and surface features; soil moisture has even lower spatial correlation and in situ probes may only represent a support volume of about 4 dm³ according to some researches (Blöschl & Sivapalan, 1995; Iwema et al., 2017; Kurc & Small, 2007). Thus, neither can reflect the overall performance, especially over the heterogeneous surface. Aimed at ameliorating the issue, calibration methods using flux towers or stream gauges at multiple locations were proposed (Niraula et al., 2012; Williams et al., 2009). Meanwhile, multi-objective calibration is employed to fully evaluate and calibrate the land surface models since they generate a variety of outputs (McCabe et al., 2005). Nevertheless, the critical terrestrial states and fluxes such as sensible heat flux, evapotranspiration, and soil moisture still cannot be fully represented

because they have significant spatial heterogeneities at fine scales which by insufficient coverage of in situ networks can barely capture. Hence, the limitation of the calibration approaches makes it attractive to calibrate the model with spatially distributed observations.

The advancement in satellite remote sensing has offered the community a unique opportunity to evaluate and calibrate land surface models in a spatially explicit way using large-scale monitoring of surface temperature, soil moisture, and evapotranspiration, even though some uncertainties should be addressed (Liu et al., 2011; Wagner et al., 2012). Actually, the idea of using satellite-derived data to evaluate and calibrate the models is not new. Research into calibration against remote sensing data has been ongoing for years (Prata et al., 1995). For example, Kunnath Poovakka et al. (2013) used microwave soil moisture retrievals and satellite-derived evapotranspiration to calibrate the parameters of the Australian Water Resources Assessment Landscape model ($\sim 5 \text{ km} \times 5 \text{ km}$), leading to better predictions of evapotranspiration and runoff. Corbari et al. (2015) reported that the calibration based on MODIS LST retrieved from satellite data outperformed the calibration based on discharge. Among the available data, satellite images of land surface temperature have received particular attention as LST is measured more directly than evapotranspiration, and LST is a critical variable of the Earth's surface. LST drives the outgoing longwave

radiation and exerts control over the partitioning of sensible heat flux and latent heat flux (Hulley et al., 2019; Wang et al., 2014). Sensible heat is the heat that warms or cools the air in contact with the surface, while latent heat accounts for the heat that water vapor evaporates from or condenses onto the land (Shuttleworth, 2012). The sensible and latent heat fluxes, part of the energy cycle, also have a profound effect in determining the hydrological cycle, boundary layer, weather, and climate (Wilson et al., 2002). LST is therefore of fundamental importance in the simulation of energy and water budgets. At the same time, LST itself is widely studied due to its versatility. For example, LST is a strong indicator for climate change (Yan et al., 2020), an important variable for the retrieval of air temperature and soil moisture (Oyler et al., 2016; Yao et al., 2011), and a factor in predicting agriculture (Marques da Silva et al., 2015; Son et al., 2012).

In recent years, satellite LST data have been widely used to calibrate and validate model parameters. Some spaceborne instruments used to measure high-resolution surface temperatures on a global scale are summarized in Table 1. Dozens of LST products with different levels have been generated using diverse approaches. Thanks to the daily temporal resolution, MODIS LST data ($\sim 1 \text{ km} \times 1 \text{ km}$) is widely used. For example, Gutmann and Small (2010) calibrated the NOAH LSM with MODIS LST; Corbari and Mancini (2014) used it to calibrate the soil and vegetation parameters of the

pixel lack of in situ measurements of discharge. Samaniego et al. (2010) proposed a multiscale parameter regionalization technique that enables using MODIS LST to calibrate parameters at coarser scales (i.e., 2 km, 4 km, and 8 km).

Although many studies have calibrated the model parameters with satellite LST, there is no effort to do it at sub-grid scale level. This is partly because tiles in the past are most abstract concept – they are not collocated in space. HydroBlocks land surface model, however, developed a new tiling framework and the tiles can be mapped and evaluated. HydroBlocks is a field-scale (30 m) physically-based land surface model that learn spatial heterogeneity from the high-resolution environment datasets (30 – 100 m) (Chaney et al., 2016). To this end, HydroBlocks clusters areas of similar surface characteristics and hydrologic behavior into hydrologic response units (HRUs), or tiles in other words, allowing the model to effectively simulate important surface processes while maintaining computation efficiency. The clustering mechanism allows for the results of each HRU to be mapped over the region, making it possible to assess the model performance explicitly. Additionally, satellite LST data with high spatial resolution might be a good choice for the generation of HRUs.

Table 1: Summary of some LST instruments commonly used to generate high-resolution LST data.

Satellite	Sensor	Launch Time	Decommission Time	Pixel size at nadir (m)	Revisit time (days)
ISS	ECOSTRESS	2018		69 × 38	3–5
Landsat 5	TM	1984	2013	120	16
Landsat 7	ETM+	1993		60	16
Landsat 8	TIRS	2013		100	16
Terra	MODIS	1999		1000	1–2
Aqua	MODIS	2002		1000	1–2
Terra	ASTER	1999		90	16
ERS-1	ATSR-1	1991	2000	1000	35
ERS-2	ATSR-2	1995	2008	1000	35
Envisat	AATSR	2002	2012	1000	35
Sentinel-3A	SLSTR	2016		1000	2
Sentinel-3B	SLSTR	2018		1000	2
MetOp-A	AVHRR/3	2006		1100	1
MetOp-B	AVHRR/3	2012		1100	1
MetOp-C	AVHRR/3	2018		1100	1

1.2 Objectives

This study aims to investigate how the HydroBlocks land surface model can benefit from the available satellite observations. This is achieved through analyzing the following specific questions: (1) Can satellite-derived high-resolution LST data improve the heterogeneity representation of HydroBlocks LSM? (2) Is it possible to improve the simulation by replacing important parameters computed by the model with satellite retrievals? (3) How reliable is a calibration based on satellite-derived LST data? To

answer these questions, we conducted this study over the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site. The HydroBlocks LSM was set up to run from 2013 to 2019 at the 30-meter spatial resolution and hourly temporal resolution. To improve the tiles representation of the landscape, the clustering algorithm used to assemble HydroBlocks' sub-grid tiles was modified to allow the model to learn the heterogeneity from the temporal mean and standard deviation of Landsat LST. The effects of inputting the albedo, emissivity, and leaf area index (LAI) from MODIS into the model were also investigated. Finally, the MODIS LST was employed to perform tile-level calibration, and the effectiveness was discussed.

2. Methodology

2.1 Overview of the HydroBlocks land surface model

HydroBlocks is a field-scale land surface model (Chaney et al., 2016) which represents the spatial heterogeneity of the water, energy, and carbon cycles explicitly via HRUs. A hierarchical multivariate clustering (HMC) approach, which will be discussed in detail later, is applied to divide and cluster points with similar characteristics into HRUs (Chaney et al., 2018, 2020). The HMC is a framework employed to capture the heterogeneity and allow HydroBlocks to work as a semi-distributed model while performing comparably to a fully distributed model (i.e., one HRU per grid cell). On this account, the application of these HRUs enables HydroBlocks to handle mass and energy exchanges thoroughly and realistically while ensures computational efficiency, thus making it possible to model hydrological, geophysical, and biophysical processes at high spatial and temporal resolutions over regional to continental extents (Vergopolan et al., 2020). HydroBlocks applies Noah-MP (Niu et al., 2011) at each HRU to simulate the surface process explicitly. At each time step, the land surface scheme updates the hydrological states at each HRU; and the HRUs dynamically interact laterally via subsurface flow (Chaney et al., 2020).

2.1.1 Generation scheme of hydrologic response units

The generation scheme of HRU in the HydroBlocks is keeping being updated. The original scheme was replaced by the hierarchical multivariate clustering (HMC) approach proposed in 2018 to represent the riparian zones properly (Chaney et al., 2018). Recently, HMC was modified to enable better capture of riparian zone dynamics and adequate coupling with the river network (Chaney et al., 2020). The steps of HMC include (a) defining characteristic basins, (b) discretizing height bands, and (c) determining intra-band clusters. The key information of HMC is outlined below while detailed description can be found in Chaney et al. (2020).

(a) Characteristic basins: The K-means clustering algorithm is used to partition basins, defined based on a 30-meter digital elevation model (DEM), into k (user-defined) clusters. The clusters (i.e., characteristic basins) are determined by topographic and hydrologic features such as latitude, longitude, and flow accumulation area. Figure 5a shows an example of dividing the domain into five characteristic basins. An empirical cumulative distribution function (CDF) of the 30-meter height above nearest drainage (HAND) values is fitted for each characteristic basin respectively. The empirical CDFs of basins in the given characteristic basin are then matched to the characteristic basin's CDF. In this way, the similarity of those basins is increased.

(b) Height bands: For each characteristic basin, all channel grid cells are grouped into the first height band. All the other grid cells are then sorted according to the HAND values. The area of the second height band should be n (user-defined) times larger than the first height band; the third height band is n times larger than the second height band, and so on. Therefore, the larger n defined, the fewer height bands generated. Figure 5b shows an example of the partitioning of the height bands for one characteristic basin.

(c) Intra-band clusters: The grid cells within each height band are then clustered into HRUs, the smallest modeling units, based on environmental characteristics (e.g., position, elevation, soil properties, and land use). The user can define the average number of HRUs – p , and the exact number of HRUs for each height band is then assigned according to the ratio of its area to the characteristic basin's total area. The HRU configuration is determined by both the number of intra-band clusters and the covariate features used (Figure 5c).

2.1.2 Parameters considered in the calibration

The HydroBlocks LSM simulates the surface processes through Noah-MP (Niu et al., 2011), which builds on the Noah land surface model (Ek et al., 2003). Noah-MP provides options for parameterization for some land surface states and processes. The parameterization schemes chosen for this study are listed in Appendix A. Since the primary focus here is to understand the important processes and parameters related to

land surface temperature, this section provides a brief overview of them. The core of the energy balance is that in the energy fluxes between the near-surface atmosphere, vegetation canopy, vegetated ground surface, and bare ground surface, LST is the representative equilibrium temperature that is able to solve these equations and close the energy balance.

The LST is calculated as the weighted-sum of bare ground surface temperature ($T_{g,b}$), vegetated ground surface temperature ($T_{g,v}$), and vegetation canopy surface temperature (T_v). The calculations of surface temperature and energy fluxes over different surfaces are slightly different, but the main principles are similar. So here we only show the formulations for vegetated ground. For a more comprehensive description, see Niu et al. (2011). For vegetated ground, the net longwave radiation is calculated as:

$$L_{ag,v} = -\alpha_{gL}L_v \downarrow + \epsilon_g \sigma T_{g,v}^4$$

where ground surface absorption (α_{gL}) and the ground surface emissivity (ϵ_g) are considered to be equal, downward longwave radiation ($L_v \downarrow$) includes that from the atmosphere transmitted through the canopy and emitted by the canopy; and σ is the Stefan-Boltzmann constant.

The sensible heat flux ($H_{g,v}$) is calculated as:

$$H_{g,v} = \rho C_p \frac{T_{g,v} - T_{ac}}{r_{ah,g}}$$

where the heat difference between the ground surface and the canopy air is computed based on the temperature difference between them ($T_{g,v} - T_{ac}$), air density (ρ), and the dry-air specific heat capacity (C_p); $r_{ah,g}$ denotes the aerodynamic resistance for heat.

For latent heat flux ($LE_{g,v}$), the calculation is similar with sensible heat flux except it accounts for the heat caused by water vapor instead of the air:

$$LE_{g,v} = \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_{g,v})h_g - e_{ac}}{r_{aw,g} + r_{soil}}$$

where γ is the psychrometric constant, $e_{sat}(T_{g,v})$ is the saturated water vapor pressure, h_g is the relative humidity, e_{ac} is the water vapor pressure of the canopy air, $r_{aw,g}$ and r_{soil} are the aerodynamic and soil surface resistance for water vapor.

The ground heat flux (G_v) is calculated as:

$$G_v = \frac{2\lambda_{isno+1}}{\Delta z_{isno+1}} (T_{g,v} - T_{isno+1})$$

where λ_{isno+1} is the thermal conductivity of the surface layer (snow or soil), Δz_{isno+1} denotes the layer thickness of the surface layer, and T_{isno+1} represents the temperature of the surface layer.

Given the LST computed by solving the surface energy balance is then used to calculate the heat fluxes, it is encouraging that model calibration can be performed with LST by tune the parameters to which heat fluxes are sensitive. Both soil and plant parameters have been demonstrated as sensitive parameters in land surface models no matter for water or energy fluxes (Cuntz et al., 2016; Yang et al., 2020). So, it is necessary to consider both soil and plant parameters when we conduct the model calibration. There are a large number of soil and plant parameters, and it demands considerable computational time to tune all of them. For example, only the stomatal resistance and the photosynthesis equations already require 20 parameter values (Jefferson et al., 2017). To reduce the computational demands, only the parameters that have been commonly examined and might greatly impact the simulation of land surface temperature are selected for calibration. Table 2 provides a brief summary of those parameters and their descriptions as well as ranges are adapted from literature (Jefferson et al., 2017; Li et al., 2018; Rosero et al., 2010; Santanello et al., 2013; Zhang et al., 2017).

Table 2: Parameters selected for calibration

Parameter	Description	Units	Range	Range in calibration
bb	b parameter in hydraulic function	-	3.5, 10	$\max(0.5bb, 3.5)$, $\min(2bb, 10)$
satdk	Saturated soil hydraulic conductivity	m s^{-1}	3.5E-7, 4E-5	0.01satdk, 100satdk
maxsmc	Saturated soil moisture content (i.e., porosity)	$\text{m}^3 \text{m}^{-3}$	0.43, 0.7	$\max(0.5\text{maxsmc}, \text{refsmc}^a)$, $\min(2\text{maxsmc}, 1)$
mp	Slope for conductance-to-photosynthesis relationship	-	4, 12	0.25mp, 4mp
bp	Minimum leaf conductance	$\mu\text{mol m}^{-2} \text{s}^{-1}$	1000, 10000	0.01bp, 100bp
vcmx25	Maximum rate of carboxylation at 25°C	$\mu\text{molCO}_2 \text{m}^{-2} \text{s}^{-1}$	20, 65	0.1vcmx25, 10vcmx

^a refsmc refers to the soil moisture threshold for maximum transpiration

2.2 Study Area

The study area is located in the Southern Great Plains (SGP), and covers Northern Oklahoma and Southern Kansas, extending from 36.2° to 37.2° N, and 97.1° to 98.1° W (see Figure 1). The domain is characterized by relatively flat terrain with an average elevation of about 340 m while the elevation decreases slightly from west to east. SGP has a continental to subhumid climate presenting a large seasonal variation in temperature and precipitation with the greatest rainfall in the summer (Batra et al., 2006). The soil types throughout this domain are primarily the silty loam, along with loamy sand, sandy loam, and loam (Wan et al., 2004). The area has a heterogeneous land cover. There are rivers traversing the region with permanent wetlands along the rivers. The majority is a mixture of croplands (e.g., winter wheat, barley, and corn) and native perennial grasslands (Bagley et al., 2017). There are also forests spread over the southeast corner. In addition, Enid located in the southwest corner is the largest town in the domain while the other small urban areas are scattered.

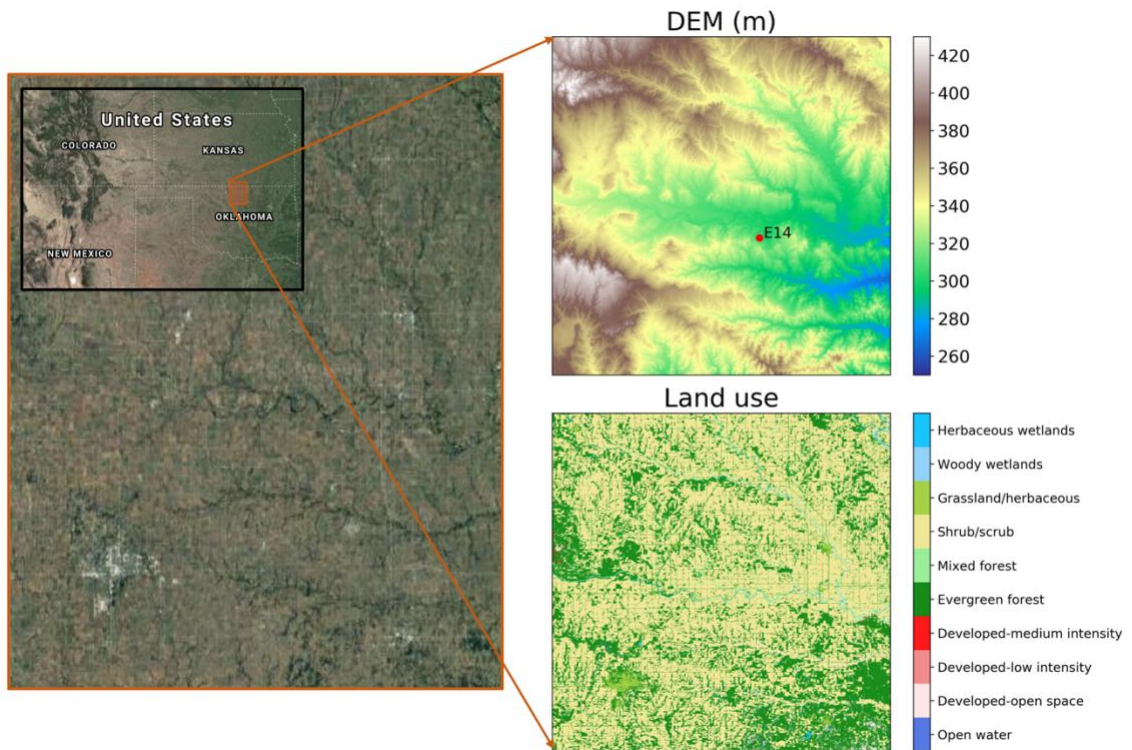


Figure 1: The study area consists of a 1.0-degree box in central northern Oklahoma and southern Kansas. The Google satellite map, DEM, and land cover of this domain are presented. The location of Atmospheric Radiation and Measurement (ARM) Southern Great Plains (SGP) Central Facility (site E14) is shown in the DEM map.

2.3 Data

2.3.1 Satellite Retrievals

a. Landsat land surface temperature

Landsat land surface temperature is employed in the HRU configuration. The U.S. Geological Survey (USGS) provides Landsat LST product generated from Landsat 4-5 Thematic Mapper, Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Thermal Infrared Sensor (TIRS) (Department of the Interior U.S. Geological Survey, 2016; NASA, 2019). The Landsat series of Earth Observation satellites have continuously acquired moderate-resolution multispectral data of the Earth's land surface since 1972. The data required in the study is from 2013 to 2019. So, it is only from Landsat 7 launched in 1999 and Landsat 8 launched in 2013, which are still in commission. Both of them are in a near-polar orbit and repeat the cycle every 16 days while there is an 8-day offset between the two satellites. So, the combination of the two results in a revisit time of 8 days. In addition, the data acquisition time (i.e., crossing time) of the study area is 17:00 ± 25 minutes (GMT). Landsat 7 carries the ETM+ sensor which has a 60-m spatial resolution in the thermal channel (10.40 - 12.50 μm) (Hook et al., 2020). Landsat 8's TIRS acquires two thermal channels, TIRS1 (10.6 - 11.19 μm) and TIRS2 (11.5 - 12.51 μm), at a spatial resolution of 100 m. Landsat LST dataset contains Surface Temperature bands that represent the temperature of the Earth's land surface in Kelvin (K), three Quality

Assessment (QA) bands, and several intermediate bands. Each pixel of Landsat LST is 30-meters across. Further details are available in the Landsat Provisional Surface Temperature Product Guide (USGS, 2018).

Landsat LST product provides a wealth of information, but the uncertainties should not be ignored. Thus, quality control must be conducted. The following criteria are applied: (a) for each scene, pixels where LST = -1000, the filling value, were discarded; (b) pixels where uncertainty is above 10% were discarded; (c) scenes with LST difference greater than 30 were excluded; (d) scenes with less than 5% pixels remain were excluded. These requirements led to a reduction in the number of scenes from 278 to 42 over the study period.

b. MODIS land surface temperature

MODIS level-2 land surface temperature product (MOD11_L2 and MYD11_L2 Version 6) is used for evaluation and calibration (Wan et al., 2015). The LST is retrieved from MODIS sensor onboard the satellites Terra and Aqua with a pixel size of ~1000 m in the thermal infrared bands. The product is produced daily in 5-minute temporal increments of satellite acquisition using the generalized split-window algorithm from 2 bands, band 31 (10.78-11.28) and band 32 (11.77-12.27) (Duan et al., 2019a). MODIS/Terra and MODIS/Aqua both view each spot on the Earth almost every day. The data acquisition time of the two spacecraft for the study domain is about 1:00 pm and 11:00

pm respectively (local time), so that day-time and night-time LST are both provided. Similar quality control was also conducted for MODIS LST product. More specifically, (a) pixels where the QA bands (i.e., QA is 0 and 1) indicate good quality were kept and the others were discarded; (b) scenes with LST difference greater than 10 were excluded; (c) scenes with less than 10% pixels remained were excluded. In total, 3642 images of MODIS LST products were collected for the whole calibration period from January 2016 to December 2019, while 189 images meet the criterion.

c. MODIS albedo

In land surface models, downwelling radiative flux may be partitioned into direct radiation and diffuse radiation. Accordingly, black-sky albedo and white-sky albedo are defined as the albedo when there is only direct radiation or diffuse radiation. The downwelling radiation is also divided into visible and near-infrared bands so that corresponding albedo values are required. The 1000 m daily white-sky albedos and the black-sky albedos for the visible (0.3-0.7 μm) and near-infrared (0.7-5.0 μm) broadband are provided by the MOD43 Version-6 MODIS/Terra+Aqua BRDF/Albedo product (Schaaf & Wang, 2015). The gaps of the daily albedo value were filled with the average value of the available data. Then the hourly albedo was estimated from daily value by linear interpolation.

d. MODIS emissivity

The MOD11A1 Version 6 product provides daily spectral emissivity in MODIS band 31 (ϵ_{31}) and band 32 (ϵ_{32}) with 1000 m spatial resolution (Wan et al., 2015). The surface broadband emissivity (ϵ) was estimated based on the approach proposed by Jin and Liang (2006):

$$\epsilon = 0.4587 \times \epsilon_{31} + 0.5414 \times \epsilon_{32}$$

Similarly, hourly surface emissivity was estimated in the same way as albedo.

e. MODIS leaf area index

Leaf area index (LAI) retrieved from MCD15A2H Version-6 product (Myneni et al., 2015), an 8-day composite dataset with ~500-meter pixel size. This means that the value for each pixel is the best value chosen from all the available data from both Terra and Aqua within the 8 days. Each 8-day composite LAI was assigned to the corresponding hourly periods.

2.3.2 Primary Model Inputs

The meteorological inputs to the model came from the Princeton CONUS Forcing (PCF) dataset which provides a 3-km ($1/32^\circ$) hourly meteorological product over CONUS from 2002 to the present (Pan et al., 2016).

The 1 arc-second (~ 30 meters) USGS National Elevation Dataset (NED) was used in this study (Gesch et al., 2009). A series of processes were conducted to remove pits, derive slope, topographic index, flow direction, and flow accumulation area.

The soil properties used in this study was from the 30-m Probabilistic Remapping of SSURGO (POLARIS) dataset (Chaney et al., 2016), which provides saturated soil moisture content (maxsmc), wilting point, field capacity, saturated hydraulic conductivity (satdk), bubbling pressure, as well as the b parameter (bb) used in the water retention curve.

The National Land Cover Database (NLCD) is a 30-m land cover database with a 16-class land cover classification (Homer et al., 2015). The land cover Noah-MP vegetation categories were converted from NLCD.

2.3.3 In Situ Measurements

Model outputs are also evaluated against ARM data from the eddy correlation (ECOR) flux measurement system and the surface energy balance system (SEBS) of ARM. The ECOR provides half-hour measurements of sensible heat and latent heat fluxes. The SEBS dataset consists of the upwelling and downwelling shortwave and longwave radiation, and the ground surface heat flow. Hence the surface energy closure can be analyzed.

There are two sites in the study domain: E14 (36.605216° N, 97.487684° W) and E37 (36.310422° N, 97.927405° W). The E14 site is located in the middle of pastures and wheat fields while the surrounding landscape of E37 site is made up of grassy field with

a cultivated field to the South. Although ARM data provides a wealth of information, uncertainties in the data can lead to misleading conclusions. To ensure the quality of the data used as ground truth, the following flags were applied at each site: 1) timesteps with no data of sensible heat flux, latent heat flux or ground heat flux were discarded; 2) timesteps in which the error in energy balance closure was above 10% were discarded. Gap-filling was not applied and only available observation data were used. Then, all the data were aggregated to hourly intervals.

As seen in Figure 2, there are substantial unqualified data at E37 in 2017, which accounts for nearly one fourth of the evaluation period. The E37 was hence not used in this study. After preprocessing, there were 3627 timesteps remained from 2016 to 2019 at E14. In addition, the upwelling and downwelling shortwave measured by SEBS together with the emissivity retrieved from MODIS are used to estimate the surface temperature of this site (Duan et al., 2019b):

$$LST_{obs} = \left[\frac{LW_u - (1 - \varepsilon)LW_d}{\sigma\varepsilon} \right]^{1/4}$$

where LST_{obs} is the in situ LST, LW_u and LW_d are the upwelling and downwelling longwave radiation, ε is the surface broadband emissivity calculated in Section 2.2.1, and σ is the Stefan-Boltzmann constant.

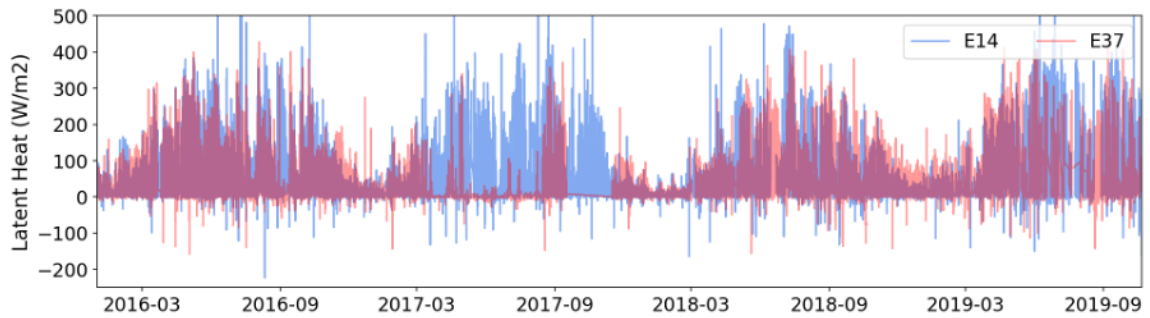


Figure 2: Time series of measured latent heat flux of site E14 and E37 after quality control.

2.4 Model Experiments

For all experiments, the HydroBlocks LSM was run over the SGP domain at the 30-meter spatial resolution and hourly temporal resolution from 2013 to 2019. The first three years were used for spin-up, and hence the simulation for the rest period 2016-2019 was used for analysis.

2.4.1 Convergence Analysis Experiment

In order to figure out the number of HRUs that are sufficient to simulate the whole study area almost explicitly while achieving high computational efficiency, a convergence experiment should be conducted. Similar to the approach used in Chaney et al. (2020), a series of different model experiments were run to evaluate how the HRU configuration parameters impact the representation of heterogeneity and the model simulations. For all experiments, the clustering covariates for clustering the

characteristic basins were latitude, longitude, flow accumulation, and the natural logarithm of the flow accumulation area. The clustering covariates for the intra-band clusters (i.e., HRUs) were the 30-m resolution latitude, longitude, and DEM map. The experiment ID and the corresponding set of HMC parameters are listed in Table 3 where k is the number of characteristic basins, n is the ratio of the area of adjacent height bands, and p is the number of intra-band clusters. It is noted that the set of HMC parameters used for all the other experiments was determined according to the convergence analysis.

Table 3: HMC parameter sets used for convergence analysis

ID	k	n	p
a	1	1000	3
b	5	1000	3
c	10	1000	3
d	10	10	6
e	10	5	6
f	10	2	6
g	5	2	15
h	5	2	30
i	5	2	60

2.4.2 Intra-band Clustering with Land Surface Temperature

Both the number and the configuration of HRUs play a pivotal role in balancing the goal of simulating as realistically as possible and minimizing computational demands. The convergence analysis in the last section focused on the number of HRUs

while HRU configuration will be investigated in this section. As mentioned above, the HRUs are determined by the selected proxies of the drivers of spatial heterogeneity. To check if Landsat LST can improve the assembling of grids with similar characteristics, the model was forced to learn the heterogeneity from the temporal mean and standard deviation of Landsat LST (2013-2019). Specifically, comparisons of two experiments were conducted. For the baseline, the HRUs (HRUs-Baseline) were clustered by latitude, longitude, and DEM. The experiment (HRUs-LST) included not only those used for HRUs-Baseline but also the temporal mean and temporal standard deviation of Landsat LST. HydroBlocks was then run with the two sets of HRUs severally. To examine the influence of using different HRU assembling, the simulated LST of the two experiments were then compared against the MODIS LST data. This means that the hourly simulated LST data which was closest to the satellite crossing time was extracted for evaluation. What's more, the spatial resolution of mapped simulated LST is 30 m while that of MODIS LST is 1000 m. So, simulated LST was upscaled to keep the same pixel size with observations. Aside from evaluation at the regional level, the land surface temperature, sensible heat flux, and latent heat flux were compared against the site observations to provide more information on the model performance.

The agreement between the simulated values and the observed values was measured using the following performance metric: Pearson correlation coefficient (R)

and the Kling-Gupta efficiency (KGE) metric which combines linear correlation, temporal variability, and mean bias (Gupta et al., 2009). KGE ranges from negative infinity to 1 achieved when the model has perfect simulation.

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_{model}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{model}}{\mu_{obs}} - 1\right)^2}$$

2.4.3 Replacement of Albedo, Emissivity, and LAI

HydroBlocks simulates the surface processes through Noah-MP which consists of models to compute albedo, emissivity, and LAI according to vegetation type, time, land cover, whether the soil is saturated, and so on. They are important parameters that play a critical role in controlling the land surface temperature and the consequent surface energy balance. MODIS provides satellite-derived albedo, emissivity, and LAI products. In this study, we also investigated the effects of using MODIS parameters instead of the values generated from the model. The model performance was assessed in the same way as the section 2.4.2.

2.4.4 Parameter Calibration with Latin Hypercube Sampling

The Latin hypercube sampling (LHS) technique (McKay et al., 1979) was employed to achieve optimal model performance. LHS is based on stratification – it divides the range of each parameter (assuming uniform distributions in this study) into several disjoint intervals of equal probability and then takes a value randomly from each interval. The values for each parameter are then combined to generate several parameter

sets. Therefore, LHS ensures that the parameters are sampled in a relatively even way. To simplify the calibration process, the LHS was used to generate 50 sets of scalars for six parameters, b parameter in hydraulic function (bb), saturated soil hydraulic conductivity (satdk), saturated soil moisture content (maxsmc), slope for conductance-to-photosynthesis relationship (mp), minimum leaf conductance (bp), and maximum rate of carboxylation at 25°C (vcmx25), per HRU. Multiply the scalars and the default parameter values together, and we got a 50-member parameter ensemble which was used to run the HydroBlocks land surface model. The range of the scalars and the feasible bounds are shown in Table 3.

To assess the performance for those parameter sets, the observed LST for each HRU should be estimated from the MODIS LST data first. As a result, MODIS LST image was first downscaled to 30 meters, and then the observed LST per HRU was computed based on the HRU map (i.e., HRUs_LST). For each HRU, the Pearson correlation coefficient between the time series of observed LST and simulated LST with every parameter set was recorded. The strategy here was to calibrate the model by maximizing the correlation coefficient. Hence the parameter set which led to the best simulation was found. Subsequently, HydroBlocks was then run with the calibrated parameter values. The regional performance and site-level performance were also evaluated.

3. Results

3.1 Convergence Analysis

To determine how many HRUs are sufficient to represent the heterogeneity of the domain, a convergence analysis was performed. Hence, the model was run with nine sets of HRU generation parameters leading to the increasing total number of HRUs. In specific, simulations a-c increase the number of characteristic basins (increase k) from 1 to 5 to 10 while setting $p = 3$ and $n = 1000$; simulations d-f increase the number of height bands (decrease n from 10 to 5 to 2) while setting $p = 6$ and $k = 10$; simulations g-i increase the average number of intra-band clusters (increase p) from 15 to 30 to 60 while setting $n = 2$ and $p = 5$. To illustrate how the increase in heterogeneity complexity impacts the fine-scale simulated land surface temperature, the annual mean land surface temperature over the domain is shown in Figure 3.

Figure 3a-c show the influence of increasing the number of characteristic basins. It is clear that when the number of characteristic basins is too low, all the channels are grouped into one HRU which cannot represent the hydrologic processes such as flood and inundation realistically. With five characteristic basins, the main channels and the tributaries are able to be separated; the riparian areas are recognized meanwhile. Increasing the number of characteristic basins from 5 to 10 also leads to an increase in heterogeneity as the number of total HRUs increases. Figure 3c begins to clearly show

the south-north gradient of annual mean land surface temperature taking advantage of 60 HRUs.

With the decreasing of n parameter, the number of height bands increases. From Figure 3d to Figure 3f, there is a significant tendency for the more detailed representation of the channel and riparian zone with the more height bands. When the number of height bands is insufficient, it makes no or little distinction between the area immediately adjacent to the channel and the area away from the channel, thus resulting in unrealistic simulation of hydrologic processes such as flood and recharge.

Increasing the average number of intra-band clusters from 15 to 30 to 60 leads to a substantial increase in the heterogeneity even though the covariate features for clustering are just the latitude, longitude, and DEM. It is noted that the boundaries between the characteristic basins and height bands disappear effectively.

In addition, the temporal means of the spatial mean and standard deviation of the simulated land surface temperature, latent heat flux, and sensible heat flux maps are shown as a function of the number of HRUs in Figure 4. It is intended to quantify the impact of a more complex HRU configuration at the macroscale. According to the diminishing impact, the convergence happens quickly with the increase of HRUs. It is encouraging to believe that the use of ~900 HRUs is nearly sufficient to model the fine-scale features while maintaining computational efficiency. The corresponding LST map

is Figure 3h which is able to represent the region well. Therefore, the HMC parameter set, $k = 5$, $n = 2$, and $p = 30$, was then used for all the experiments in this study.

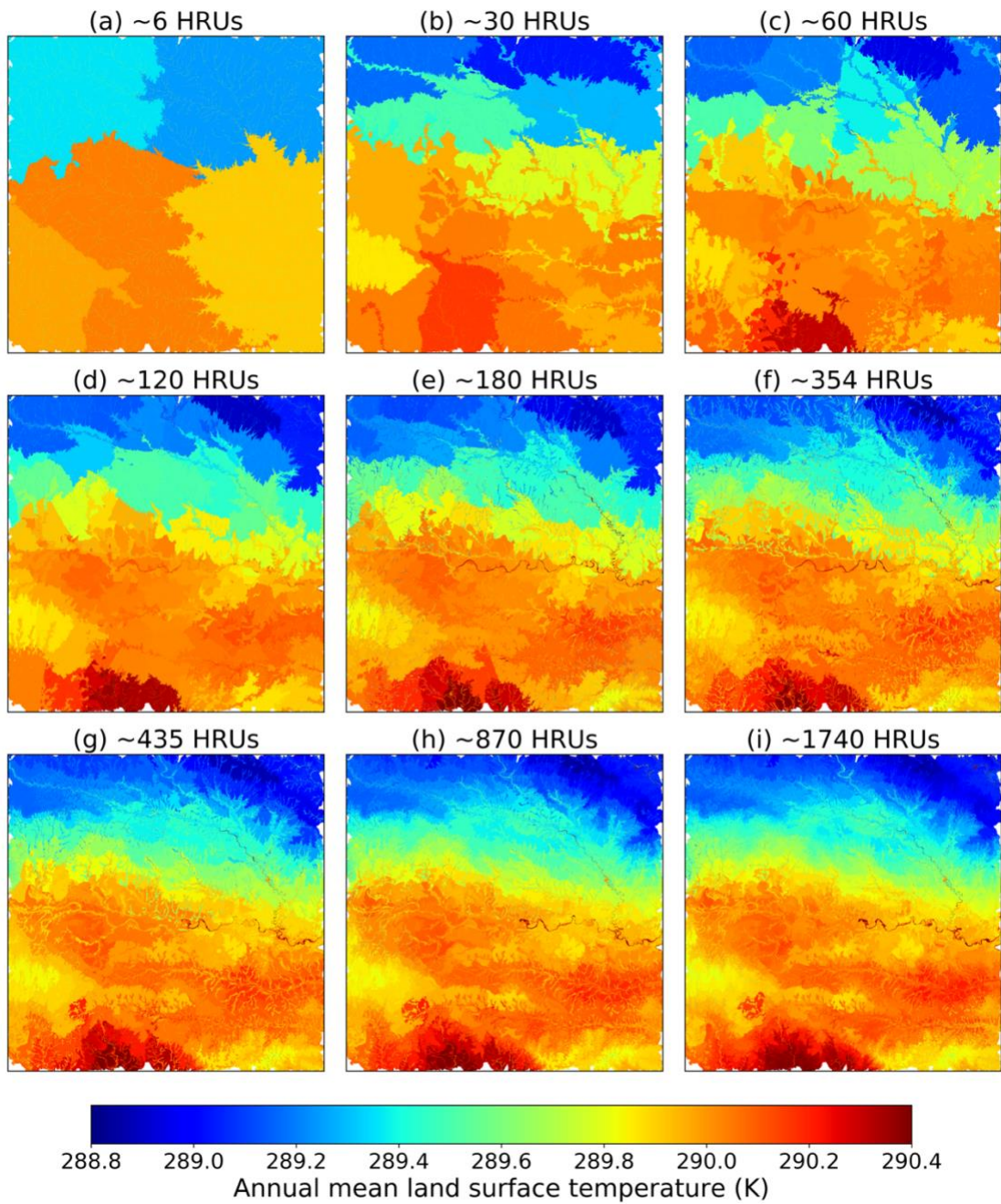


Figure 3: Simulations of annual mean land surface temperature over the study area. Each panel shows the simulation run with different parameter sets of the hierarchical multivariate clustering scheme in HydroBlocks, and the total number of HRUs is shown as the title. The specific parameters of each simulation are shown in Table 3.

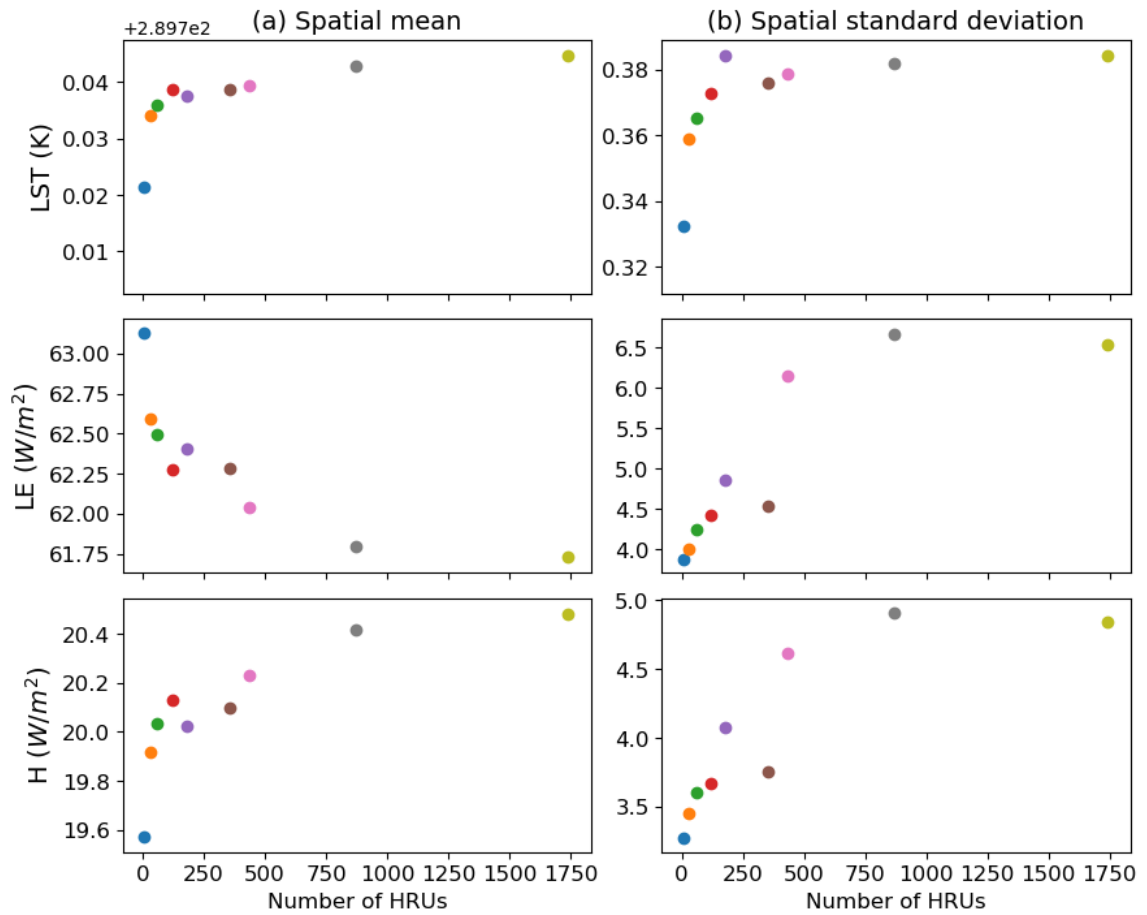


Figure 4: Annual mean of the spatial mean and spatial standard deviation of land surface temperature (LST), latent heat flux (LE), and sensible heat flux (H) for nine different HRU configurations.

3.2 Intra-band clustering with LST

As seen in Figure 5, the temporal mean and standard deviation of Landsat land surface temperature data reflected the surface features well. The rivers and the biggest town located in the southwest corner are distinct. The standard deviation of land surface temperature can reflect the seasonal variation of some vegetation types. For example, the grassland located in the west and southeast corner was characterized by relatively low variation in LST throughout the year compared with cropland. In addition, the riparian zones and the heterogeneous land covers separated by the roads were also captured by the satellite observations. The influences of the type and density of the land cover, the surface roughness, the soil moisture, and other surface features were all embedded in the land surface temperature.

To check if Landsat LST can improve the assembling of grids with similar characteristics, the model was forced to learn the heterogeneity from temporal mean and standard deviation of LST from Landsat (2013-2019). For the baseline, the HRUs (HRUs-Baseline) were clustered by latitude, longitude, and DEM. The HRUs-LST was clustered by latitude, longitude, DEM, temporal mean of Landsat LST, and temporal standard deviation of Landsat LST. The impact of incorporating LST data in the covariate features is illustrated in Figure 6c. With the same number of intra-band clusters, the HRUs-LST was much more fragmentary and has clear boundaries, which was consistent with the

characteristic of the temporal mean and standard deviation of LST maps. According to the satellite map of this study domain (Figure 1), HRUs_LST is more likely to have a realistic representation of the heterogeneous land covers separated by roads and scattered small urban areas.

To compare the difference of using HRUs-Baseline and HRUs-LST, their simulated LST were evaluated against MODIS LST, and the correlation of the pixel values for each available time step was computed (shown in Appendix B). With the purpose of having an overall understanding of the model performance, the distribution of the total 189 Pearson correlation coefficients is presented in Figure 7. In terms of the mean value of correlation coefficients, HRUs-LST performed better. To have a clearer comparison, the probability distribution is summarized into five ranges displayed in Figure 8. We can observe that HRUs-LST was not able to further improve the simulation with excellent performance ($R \geq 0.75$) or turn the bad simulation ($R < 1$) around. However, HRUs-LST could effectively increase the proportion of good simulations ($0.5 < R \leq 0.75$).

In order to further investigate the performance of other simulated variables, the outputs of the two experiments are also compared against the point scale observations (E14). Figure 9 summarizes the distribution of the simulated land surface temperature, sensible heat flux, and latent heat flux of each experiment in comparison to the in-situ

observations. The two experiments both represented the land surface temperature well. As for the simulation of heat fluxes, the latent heat was reasonable but the sensible heat was not appreciable. HRUs-Baseline outperformed HRUs-LST in the simulation of LST, while HRUs-LST showed better performance for latent heat and sensible heat in terms of KGE. Overall, the performance of HRUs-LST was comparable to or better than HRUs-Baseline.

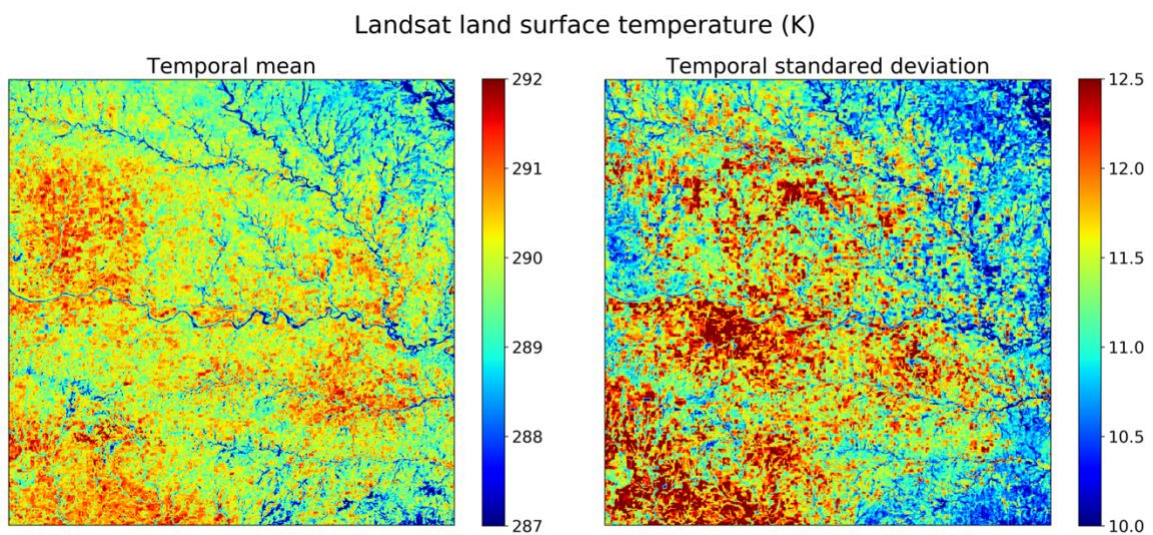


Figure 5: Temporal mean and standard deviation of Landsat land surface temperature.

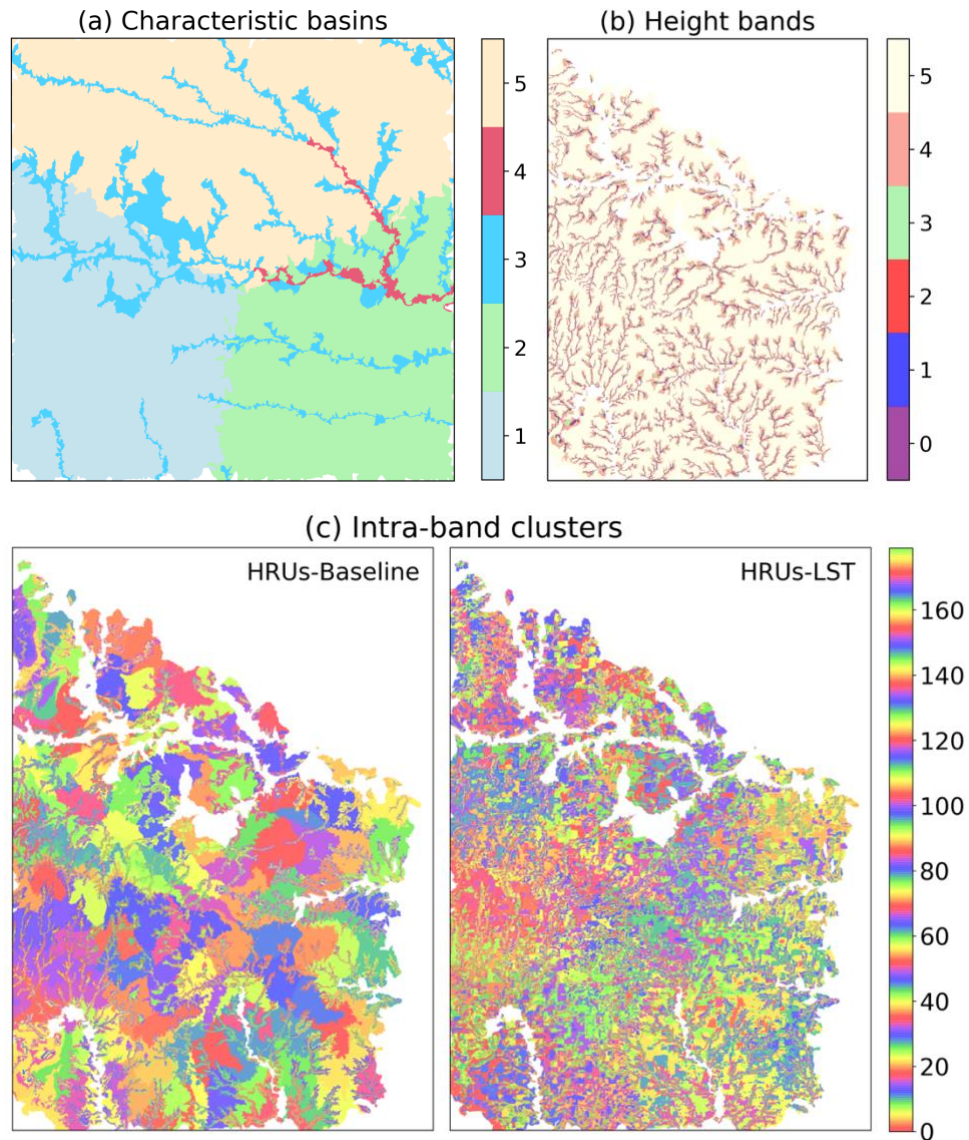


Figure 6: The domain is separated into five characteristic basins (a). Height bands are determined within each characteristic basin. For characteristic basin 1, six height bands are generated based on the setting that the ratio between the area of a height band and its adjacent height band below it is 2 (b). Within each height band, the grid cells are clustered by latitude, longitude, and DEM (HRUs-Baseline); and latitude, longitude, DEM, temporal mean and standard deviation of Landsat LST (HRUs-LST), respectively. HRU configuration is done throughout the domain, but only the characteristic basin 1 is shown to present the details better.

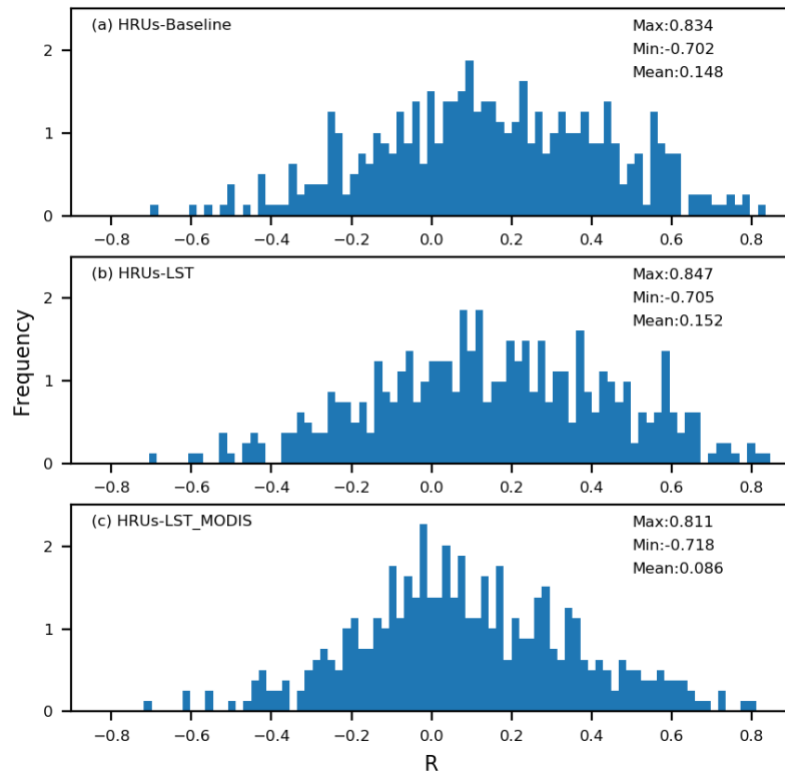


Figure 7: Comparison of the distribution of correlation coefficients (R) between 189 MODIS LST maps and corresponding mapped LST simulations of model run with (a) HRUs-Baseline, (b) HRUs-LST, and (c) HRUs-LST_MODIS.

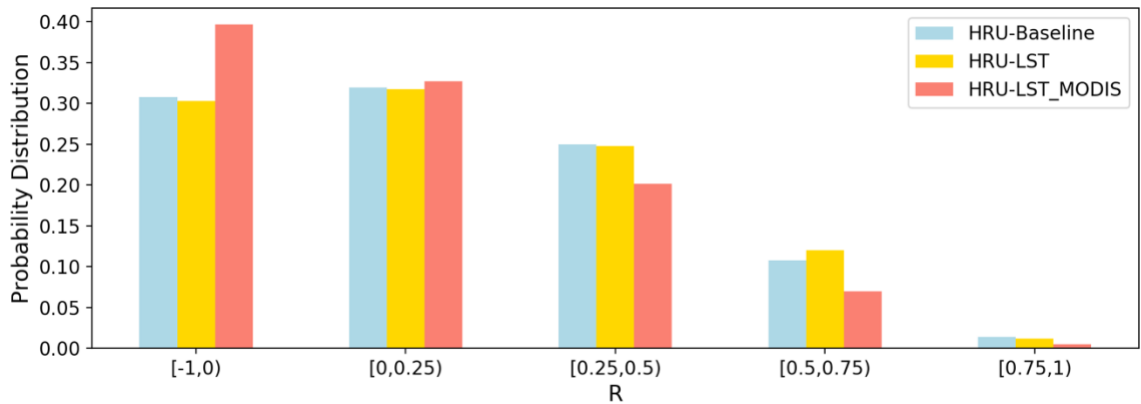


Figure 8: Probability distribution of correlation coefficients (R) between MODIS LST maps and LST simulations.

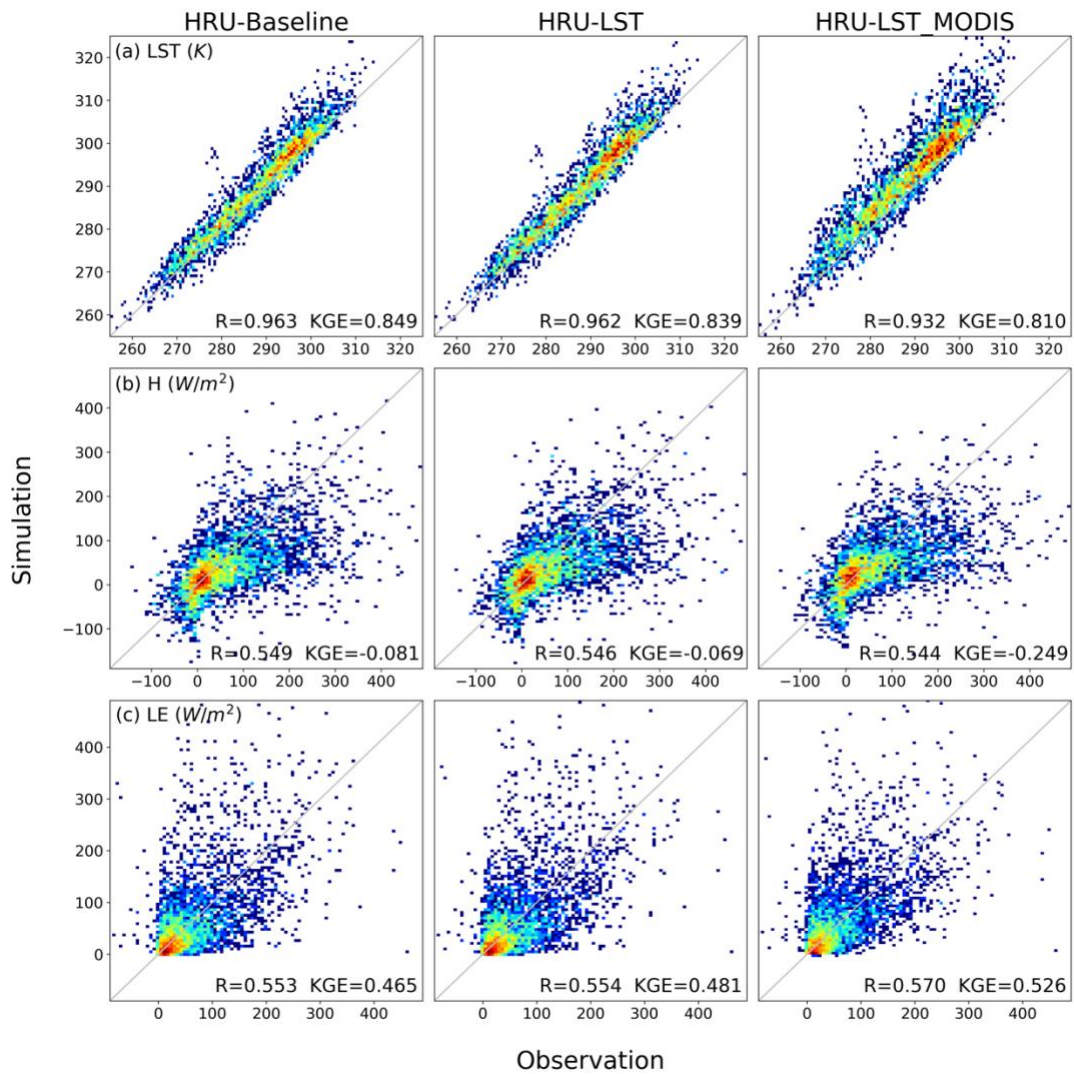


Figure 9: Model simulations versus ARM observations, (a) land surface temperature (LST), (b) sensible heat flux (H), and (c) latent heat flux (LE). R and KGE are shown in each panel for each variable.

3.3 Performance with MODIS Parameters

Albedo, emissivity and, LAI are all crucial factors in the land surface processes. The computation of those three parameters relies on the lookup table and schemes such as dynamic vegetation and snow albedo. Using satellite data to move beyond lookup tables seems to be a good way. This approach, however, did not lead to good results in the study. The performance of the HydroBlocks with MODIS parameters is also displayed in Figure 7 and Figure 8. Unexpectedly, using the MODIS parameters resulted in bad LST simulations with regards to the spatial pattern. The mean correlation coefficient of simulated LST and observed LST decreased from 0.148 to 0.086 (Figure 7). The bad simulation ($R < 1$) increased significantly after switching to MODIS parameters. As shown in Figure 9, the only improvement was the latent heat flux at site level at the cost of deteriorated land surface temperature and sensible heat flux.

3.4 Tile-level Calibration

The performance of the model run with a 50-member parameter ensemble was investigated, and the correlation coefficients between the time series of observed and simulated LST per HRU were recorded (Figure 10). The parameter sets selected for every HRU were those that maximized the correlation coefficient. The performance of the model run with the default parameters and the calibrated parameters is also included in

Figure 10. For most HRUs, the calibrated model performed almost as well as the best simulation. However, for a few HRUs, especially those located in the riparian zone, there were noticeable differences as HRUs are connected via subsurface flow. Overall, even though the HRUs are connected and their performance will be influenced by connected HRUs, calibrating each HRU separately works well.

After calibration, the number of mapped simulated LST highly correlated with observed LST increased slightly (Figure 11). Although the portion of good ($0.5 < R \leq 0.75$) and moderate simulations ($0.25 < R \leq 0.5$) decreased, calibration also reduced bad simulations ($R < 0$) notably. Therefore, it is encouraging to improve the spatial pattern of simulation over the domain by calibrating the time series of LST per HRU. Figure 12 investigated the influence of calibration further by plotting the comparison between model simulation and in situ observation for more variables. The calibrated model showed better performance in simulating latent heat flux while not in land surface temperature. It is worth noting that the simulation of sensible heat flux improved appreciably with the KGE increasing from -0.069 to 0.490. The decreasing of the ratio of simulated and observed temporal variability from 1.742 to 0.975 primarily contributed to the improvement. In general, the site-level evaluation indicates the efficacy of calibration.

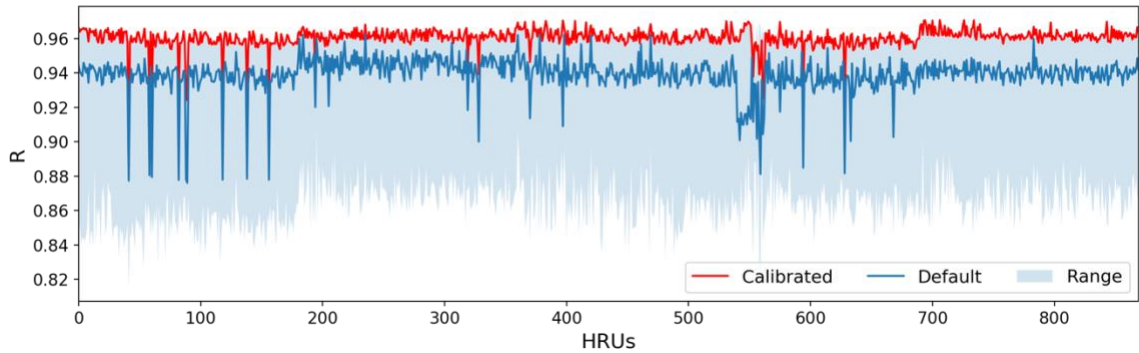


Figure 10: The range of the correlation coefficients between time series of simulated and observed LST for each HRU using 50 parameter sets. The performance of the model run with default and calibrated parameters is also shown.

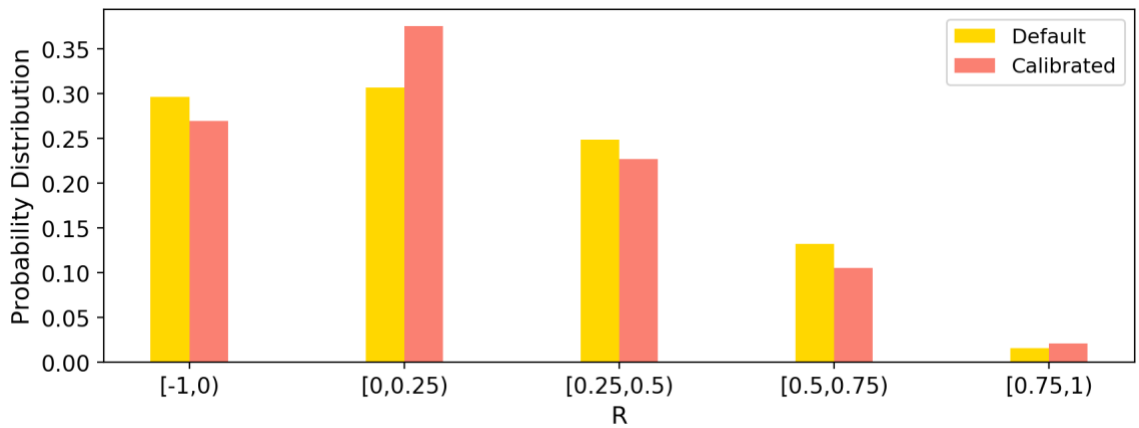


Figure 11: Comparison of the spatial performance between default and calibrated model. The linear correlation of pixel values for mapped LST simulation and MODIS image is calculated and the total 189 correlation coefficients (R) are summarized into probability distribution.

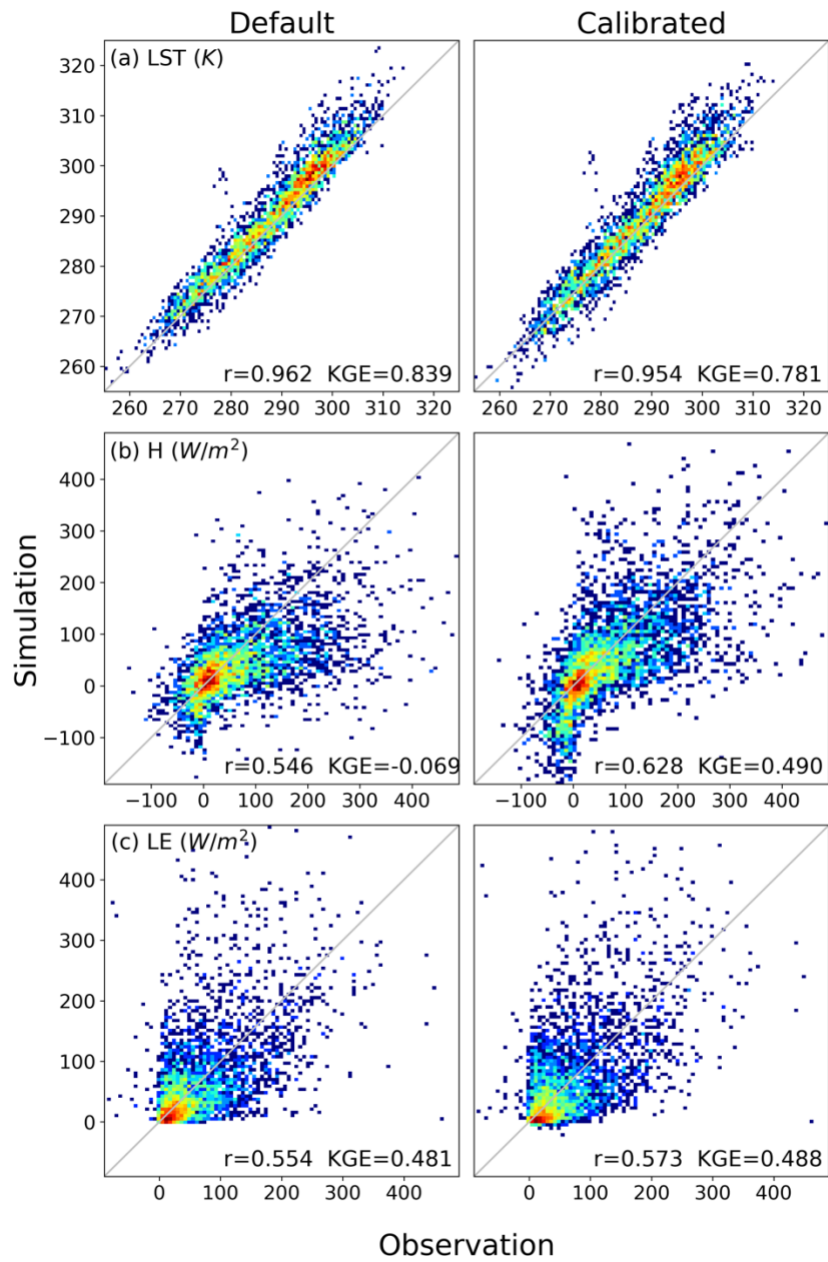


Figure 12: Site-level comparison between model simulated (default and calibrated) and observed (a) land surface temperature (LST), (b) sensible heat flux (H), and (c) latent heat flux (LE). R and KGE are shown in each panel for each variable.

4. Discussion

4.1 *Heterogeneity Representation*

As mentioned above, the intra-band clustering aims to generate HRUs by combining grid cells with similar environmental characteristics within each height band. Therefore, besides the number of clusters, the covariate features used for clustering play a critical role in determining how well the model can perform. Land surface characteristics that have been used in the HRU generation include elevation, land use, precipitation, and soil properties (Chaney et al., 2018; Chaney, et al., 2016). They are good indicators for spatial heterogeneity not only because they are important surface features that have a great impact on the water and energy cycles over the land surface but also because of the availability of high-resolution data. However, as Chaney et al. (2018) mentioned, the real goal of the HMC approach is to characterize the multi-scale heterogeneity of land surface processes. Using the proxies of the heterogeneity (e.g., elevation and soil properties) is an indirect way to realize the objective. Satellite-derived states and fluxes, such as land surface temperature and soil moisture, offer a considerable chance for us to learn the observed heterogeneity. Moreover, there are other constraints of using classification data (e.g., land cover). Even if there is no misclassification, learning heterogeneity from land use data can be problematic in some cases. For example, barren areas account for a large part of the land surface while they

are all grouped into one class. Although this limit can be compensated partly by taking the soil properties into consideration, it is subject to the fact that land use can only reflect the type of vegetation cover while the density of it is not considered. Another problem would be that the changing of the land use sometimes is not shown in the dataset in time. Taken all those factors into consideration, the high-resolution satellite-derived land surface temperature is definitely a great choice for the intra-band clustering although quality control should be conducted carefully. Moreover, with the development of remote sensing, it can be expected that more high-resolution LST products with high quality will be available in the near future.

In spite of the promising future, some limits with the tiling approach used in the study should be noted. To simulate as well as fully distributed models with a minimal computational expense, both the features used for clustering and the number of total HRUs are critical. The land surface temperature was only combined with location and elevation to inform the heterogeneity. Features, such as land cover, precipitation, and soil property, were not included considering the feature weighting for clustering. It is worthwhile to explore the effective combination of different features as well as the influence on convergence. The number of HRUs required might also vary with different covariate features employed for clustering.

4.2 Problems of Using MODIS Parameters

There are three possible reasons for the deteriorated performance. Firstly, the data used, especially the emissivity, is not as good as expected. As shown in Figure 13, the default emissivity appears much more reasonable than the MODIS retrieval as emissivity is supposed to change with the weather and seasonal variation of vegetation. Also, the spatial resolutions of the three parameters are 500 or 1000 meters which are far coarser than HydroBlocks. Secondly, the simple and rough interpolation approach used to generate hourly data could be problematic. For example, the change of snow albedo is complicated while it was just calculated from daily data via linear interpolation in this study. It was not able to competently represent the influence of snow cover. Another possible reason is the mismatch of model required variable and observed data. The emissivity inputted is the averaged broadband emissivity. However, there are several specified emissivities in the model, including the vegetation canopy emissivity and vegetated ground emissivity. Using averaged emissivity to replace vegetation canopy emissivity and vegetated ground emissivity can lead to bias.

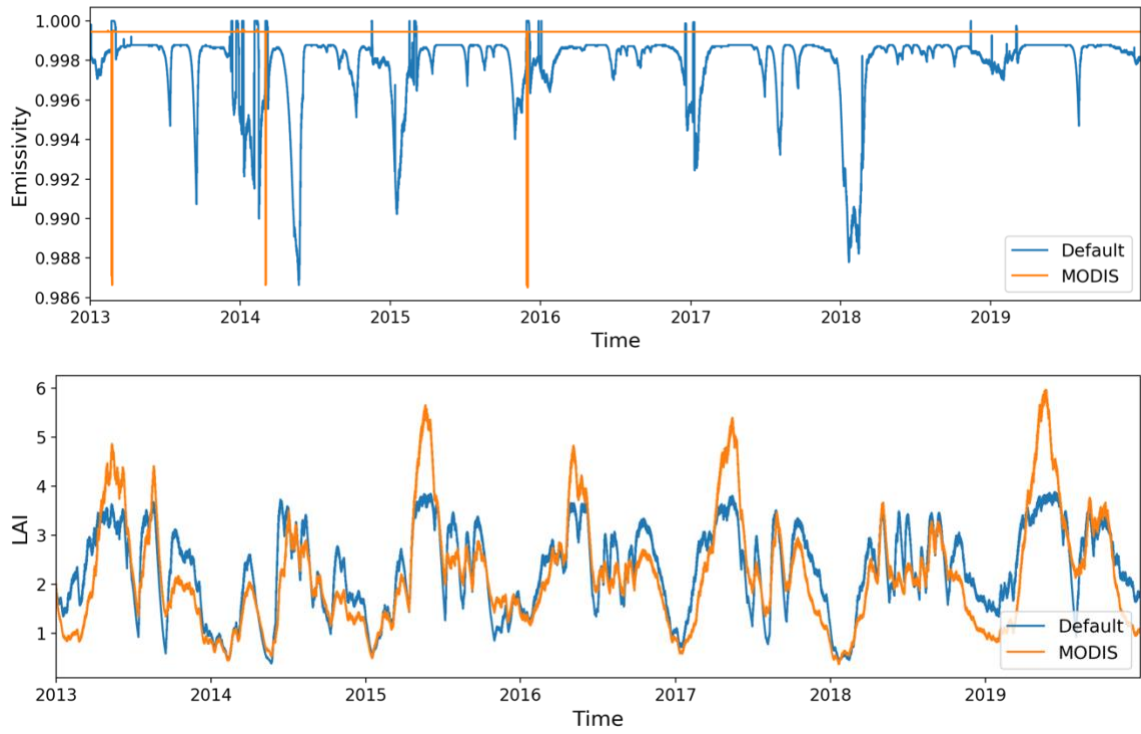


Figure 13: Time series of default and MODIS emissivity and LAI used in the model at site E14.

4.3 Uncertainties of Calibration

Encouragingly, time series of simulated LST can be calibrated effectively, and the spatial pattern will be improved to some extent consequently. Despite the promising results, the calibration approach has some limitations.

The uncertainties lie to a large extent in the LST data considered as ground-truth. To calibrate the model, the time series of observed and simulated LST for each HRU should be compared. So, it is necessary to estimate the LST for every HRU based on the MODIS LST images. Figure 14 shows the examples of the original MODIS LST images and their corresponding remapped LST from estimated HRU LST. The accuracy of HRU LST is of fundamental importance for the effectiveness of calibration. The accuracy and uncertainties of MODIS LST data itself have been fully investigated. What we want to address here is the problem caused by the coarse resolution. MODIS LST product is chosen due to the short revisit time, or in other words, more available data. However, it would be better if more high-quality satellite LST data is available. Estimating observed LST for each HRU from LST images would introduce uncertainties a certain extent. As shown in Figure 14, the remap of LST of HRUs is what the model will try to reproduce through calibration. If the remapped LST is not highly consistent with the original data, it is impossible to improve the overall performance of the model through calibration.

The quality of the remapping is related to the spatial resolution and the HRU configuration. In respect of the HRU configuration, the similarity between original LST images and remapped LST might be able to work as an indicator for heterogeneity representation. As for the spatial resolution, thanks to the development of remote sensing and retrieval algorithm, there are more and more high-quality high-resolution LST data. Recently, the new MODIS LST product (MOD21 v6) is just released (Hulley & Hook, 2017). The ASTER Temperature Emissivity Separation algorithm for MOD21 is considered more advanced than the split-window algorithm for MOD11 that was employed in the study. The only reason that ECOSTRESS and Landsat LST data was not used for calibration is the inadequate number of qualified images due to the limited operation time or long revisit time. Therefore, they will become ideal resources in the future. Additionally, the upcoming satellites with state-of-art instruments are promising, such as the Landsat 9 scheduled to launch in 2021.

Another uncertainty is the criterion for calibration. Aside from the linear correlation, the effectiveness of using KGE and root mean square error remains to be studied. On the other hand, assessing the spatial pattern via the linear correlation of the pixel values is rarely comprehensive. There are other ways to quantify spatial similarity.

It should be noted that the calibration here was not performed in an exclusive way as the primary goal is to explore the feasibility and general efficacy of calibrating

HydroBlocks using the satellite LST product. Otherwise, sensitivity analysis should be conducted first as there are so many parameters in the model. What's more, 50-member LHS is inadequate.

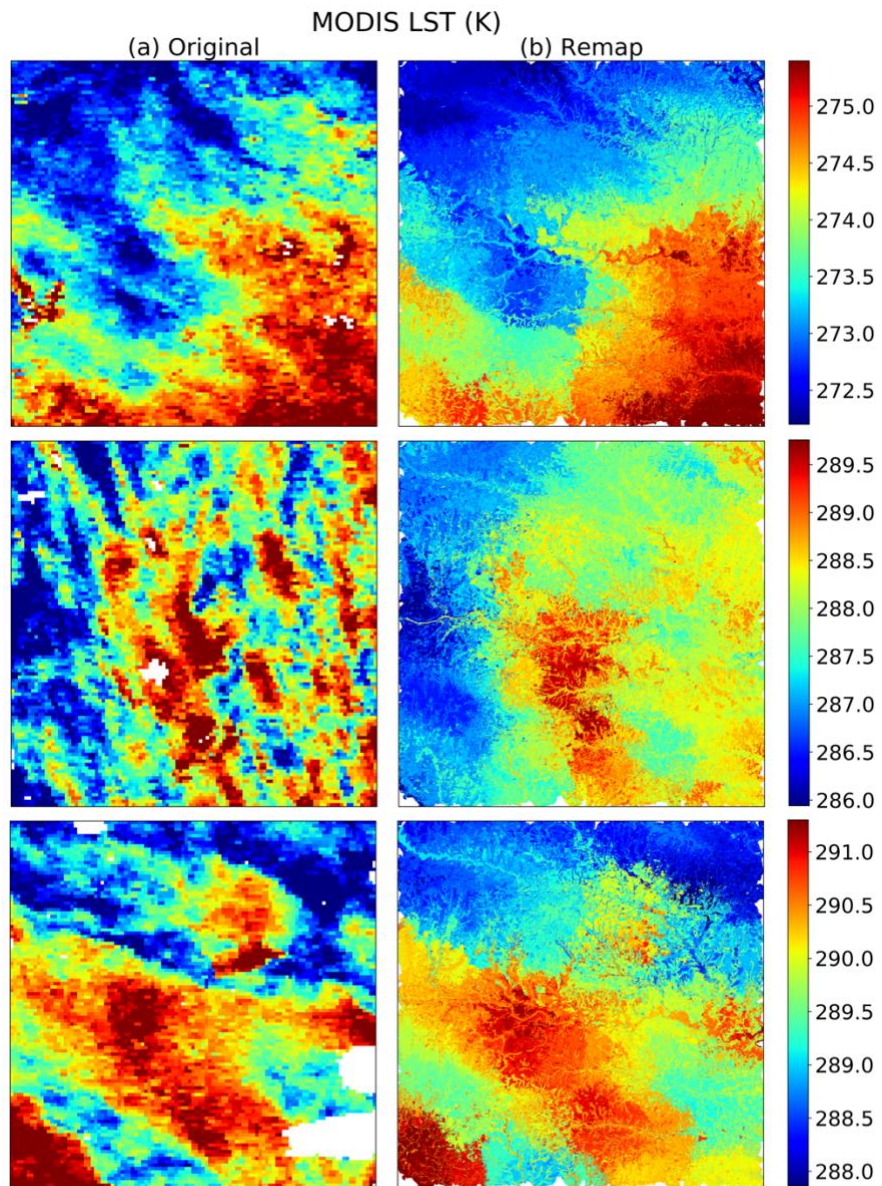


Figure 14: Examples of MODIS LST images and the remapped LST based on that of each HRU estimated from the original data.

5. Conclusions

This study demonstrated the benefits of letting the HydroBlocks land surface model learn the spatial heterogeneity from Landsat LST, in particular, the temporal mean and standard deviation map of LST. The feasibility of improving the model simulation through replacing the important variables (i.e., albedo, emissivity, and LAI) with the values retrieved from MODIS was also studied. However, it indicated that the replacement would result in deteriorated performance. Based on the analysis conducted at the site level, the primary reason might be that the quality of the data employed here is not as high as expected, especially the emissivity. The poor interpolation of albedo and the mismatch of the observed and model required emissivity value could be problematic as well. Furthermore, HydroBlocks was calibrated at tile-level by maximizing the linear correlation coefficient of the time-series of simulated LST and observed MODIS LST per HRU. The calibration was assessed at regional level as well as the point level. The spatial pattern of the simulated LST over the domain was partly improved while the sensible heat simulation enhanced considerably at the site. Therefore, the approach introduced in the study shed the light on the promising prospect of tile-level calibration in spite of the uncertainties that require further study.

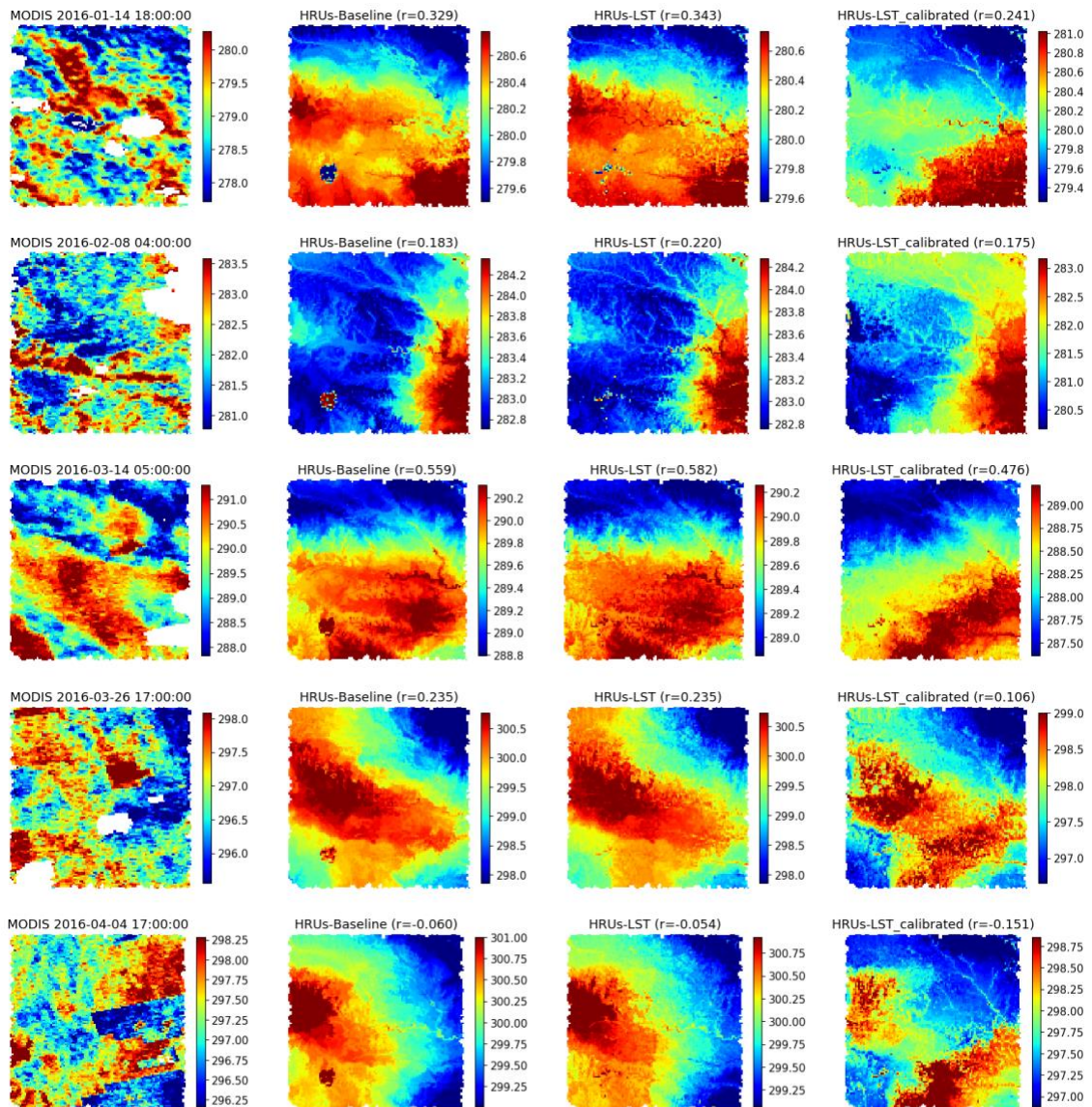
Appendix A

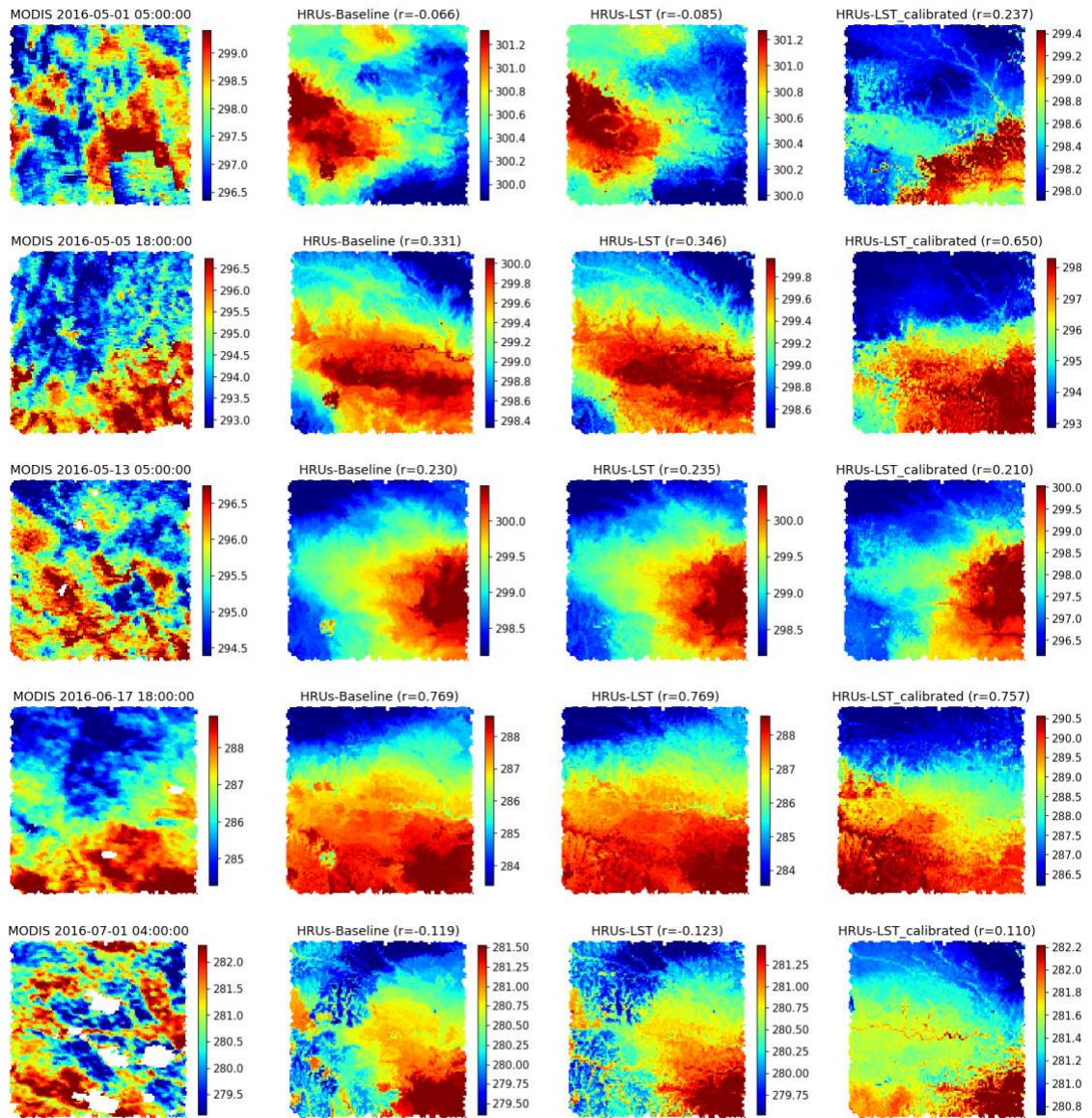
Noah-MP parameterization options used throughout the study

Process	Option selected	Description
Dynamic vegetation	5	On (use maximum vegetation fraction)
Stomatal resistance	1	Ball-Berry
Soil moisture factor for stomatal resistance	1	Noah type
Runoff and ground water	2	TOPMODEL with an equilibrium water table
Surface layer drag coefficient	1	Monin–Obukhov
Supercooled liquid water	2	Koren's iteration
Frozen soil permeability	2	Nonlinear effects
Radiation transfer	1	Modified two-stream radiation
Snow albedo	2	CLASS
Precipitation partitioning (rainfall & snowfall)	3	Based on the freezing point
Lower boundary condition of soil temperature	1	No heat flux from bottom
Snow/soil temperature time scheme	2	Full implicit
Surface resistant to evaporation/sublimation	1	Sakaguchi and Zeng (2009)
Crop model	0	No crop model

Appendix B

20 examples of MODIS LST and corresponding model simulated LST are shown below. The timestep is shown in the title of MODIS LST. The correlation coefficient of observation and simulation is shown in the title of simulated LST map.





References

- Aminzadeh, M., & Or, D. (2017). The complementary relationship between actual and potential evaporation for spatially heterogeneous surfaces. *Journal of the American Water Resources Association*, 53, 580–601. <https://doi.org/10.1111/j.1752-1688.1969.tb04897.x>
- Avissar, R., & Pielke, R. A. (1989). A Parameterization of Heterogeneous Land Surfaces for Atmospheric Numerical Models and Its Impact on Regional Meteorology. *Monthly Weather Review*, 117, 2113–2136.
- Bagley, J. E., Kueppers, L. M., Billesbach, D. P., Williams, I. N., Biraud, S. C., & Torn, M. S. (2017). The influence of land cover on surface energy partitioning and evaporative fraction regimes in the U.S. Southern Great Plains. *Journal of Geophysical Research*, 122(11), 5793–5807. <https://doi.org/10.1002/2017JD026740>
- Batra, N., Islam, S., Venturini, V., Bisht, G., & Jiang, L. (2006). Estimation and comparison of evapotranspiration from MODIS and AVHRR sensors for clear sky days over the Southern Great Plains. *Remote Sensing of Environment*, 103(1), 1–15. <https://doi.org/10.1016/j.rse.2006.02.019>
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1), 41–51. [https://doi.org/10.1016/0309-1708\(93\)90028-E](https://doi.org/10.1016/0309-1708(93)90028-E)
- Beven, K. J., & Cloke, H. L. (2012). Comment on “hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al. *Water Resources Research*, 48(1), 1–10. <https://doi.org/10.1029/2010WR010090>
- Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., et al. (2015). Hyper-resolution global hydrological modelling: What is next?: “Everywhere and locally relevant” M. F. P. Bierkens et al. Invited Commentary. *Hydrological Processes*, 29(2), 310–320. <https://doi.org/10.1002/hyp.10391>
- Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9(3–4), 251–290. <https://doi.org/10.1002/hyp.3360090305>

- Brunsell, N. A., Mechem, D. B., & Anderson, M. C. (2011). Surface heterogeneity impacts on boundary layer dynamics via energy balance partitioning. *Atmospheric Chemistry and Physics*, 11(7), 3403–3416. <https://doi.org/10.5194/acp-11-3403-2011>
- Chaney, N., Torres-Rojas, L., Vergopolan, N., & Fisher, C. (2020). Two-way coupling between the sub-grid land surface and river networks in Earth system models. *Geoscientific Model Development Discussions*, (October), 1–31. <https://doi.org/10.5194/gmd-2020-291>
- Chaney, N. W., Metcalfe, P., & Wood, E. F. (2016). HydroBlocks: a field-scale resolving land surface model for application over continental extents. *Hydrological Processes*, 30(20), 3543–3559. <https://doi.org/10.1002/hyp.10891>
- Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, 274, 54–67. <https://doi.org/10.1016/j.geoderma.2016.03.025>
- Chaney, N. W., Van Huijgevoort, M. H. J., Shevliakova, E., Malyshev, S., Milly, P. C. D., Gauthier, P. P. G., & Sulman, B. N. (2018). Harnessing big data to rethink land heterogeneity in Earth system models. *Hydrology and Earth System Sciences*, 22(6), 3311–3330. <https://doi.org/10.5194/hess-22-3311-2018>
- Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, 5929–5956. <https://doi.org/10.1002/2015WR017096>
- Corbari, C., Mancini, M., Li, J., & Su, Z. (2015). Can satellite land surface temperature data be used similarly to river discharge measurements for distributed hydrological model calibration? *Hydrological Sciences Journal*, 60(2), 202–217. <https://doi.org/10.1080/02626667.2013.866709>
- Corbari, Chiara, & Mancini, M. (2014). Calibration and validation of a distributed energy-water balance model using satellite data of land surface temperature and ground discharge measurements. *Journal of Hydrometeorology*, 15(1), 376–392. <https://doi.org/10.1175/JHM-D-12-0173.1>
- Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the

- Noah-MP land surface model. *Journal of Geophysical Research*, 121(18), 10,676–10,700. <https://doi.org/10.1002/2016JD025097>
- Department of the Interior U.S. Geological Survey. (2016). Landsat 8 Data Users Handbook. *Nasa*, 8(June), 97. Retrieved from <https://landsat.usgs.gov/documents/Landsat8DataUsersHandbook.pdf>
- Duan, S. B., Li, Z. L., Li, H., Göttsche, F. M., Wu, H., Zhao, W., et al. (2019a). Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote Sensing of Environment*, 225(September 2018), 16–29. <https://doi.org/10.1016/j.rse.2019.02.020>
- Duan, S. B., Li, Z. L., Li, H., Göttsche, F. M., Wu, H., Zhao, W., et al. (2019b). Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2019.02.020>
- Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *Journal of Geophysical Research*, 108, 1–16. <https://doi.org/10.1029/2002JD003296>
- Franks, S. W., & Beven, K. J. (1997). Estimation of evapotranspiration at the landscape scale: A fuzzy disaggregation approach. *Water Resources Research*, 33(12), 2929–2938. <https://doi.org/10.1029/97WR01963>
- Friedl, M. A. (2002). Forward and inverse modeling of land surface energy balance using surface temperature measurements. *Remote Sensing of Environment*, 79(2–3), 344–354. [https://doi.org/10.1016/S0034-4257\(01\)00284-X](https://doi.org/10.1016/S0034-4257(01)00284-X)
- Gesch, D., Evans, G., Mauck, J., Hutchinson, J., & William J. Carswell Jr. (2009). *The National Map—Elevation: U.S. Geological Survey Fact Sheet 2009-3053*. <https://doi.org/10.5860/choice.50-0008>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>

- Gutmann, E. D., & Small, E. E. (2010). A method for the determination of the hydraulic properties of soil from MODIS surface temperature for use in land-surface models. *Water Resources Research*, 46(6), 1–16.
<https://doi.org/10.1029/2009WR008203>
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States-Representing a Decade of Land Cover Change Information Landsat-based mapping project View project Create new project “Forest Inventory and Analysis” View project. *Photogrammetric Engineering & Remote Sensing*, 81(5), 345–354. <https://doi.org/10.14358/PERS.81.5.345>
- Hook, S. J., Cawse-Nicholson, K., Barsi, J., Radocinski, R., Hulley, G. C., Johnson, W. R., et al. (2020). In-Flight Validation of the ECOSTRESS, Landsats 7 and 8 Thermal Infrared Spectral Channels Using the Lake Tahoe CA/NV and Salton Sea CA Automated Validation Sites. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2), 1294–1302. <https://doi.org/10.1109/TGRS.2019.2945701>
- Hou, T., Zhu, Y., Lü, H., Sudicky, E., Yu, Z., & Ouyang, F. (2015). Parameter sensitivity analysis and optimization of Noah land surface model with field measurements from Huaihe River Basin, China. *Stochastic Environmental Research and Risk Assessment*, 29(5), 1383–1401. <https://doi.org/10.1007/s00477-015-1033-5>
- Hulley, G. C., Ghent, D., Göttsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). Land Surface Temperature. In *Taking the Temperature of the Earth* (pp. 57–127). <https://doi.org/10.1016/b978-0-12-814458-9.00003-4>
- Hulley, G., Hook, S. (2017). *MOD21 MODIS/Terra Land Surface Temperature/3-Band Emissivity 5-Min L2 1km V006* [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-03-10 from <https://doi.org/10.5067/MODIS/MOD21.006>
- Iwema, J., Rosolem, R., Rahman, M., Blyth, E., & Wagener, T. (2017). Land surface model performance using cosmic-ray and point-scale soil moisture measurements for calibration. *Hydrology and Earth System Sciences*, 21(6), 2843–2861.
<https://doi.org/10.5194/hess-21-2843-2017>
- Jefferson, J. L., Maxwell, R. M., & Constantine, P. G. (2017). Exploring the Sensitivity of Photosynthesis and Stomatal Resistance Parameters in a Land Surface Model. *Journal of Hydrometeorology*, 18(3), 897–915. <https://doi.org/10.1175/jhm-d-16-0053.1>

- Jin, M., & Liang, S. (2006). An Improved Land Surface Emissivity Parameter for Land Surface Models Using Global Remote Sensing Observations. *Journal of Climate*, 19(12), 2867–2881. <https://doi.org/10.1175/JCLI3720.1>
- Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, 44(2), 1–18. <https://doi.org/10.1029/2007WR006004>
- Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., & Simmer, C. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, 46(4), 1–7. <https://doi.org/10.1029/2009WR008730>
- Koster, R. D., & Suarez, M. J. (1992). Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *Journal of Geophysical Research*, 97(D3), 2697–2715. <https://doi.org/10.1029/91JD01696>
- Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling and Software*, 21(10), 1402–1415. <https://doi.org/10.1016/j.envsoft.2005.07.004>
- Kunnath Poovakka, A., Ryu, D., Renzullo, L. J., Pipunic, R., & George, B. (2013). Calibration of land surface model using remotely sensed evapotranspiration and soil moisture predictions. *Proceedings - 20th International Congress on Modelling and Simulation, MODSIM 2013*, (January 2016), 3113–3119. <https://doi.org/10.13140/RG.2.1.2670.0886>
- Kurc, S. A., & Small, E. E. (2007). Soil moisture variations and ecosystem-scale fluxes of water and carbon in semiarid grassland and shrubland. *Water Resources Research*, 43(6), 1–13. <https://doi.org/10.1029/2006WR005011>
- Li, J., Chen, F., Zhang, G., Barlage, M., Gan, Y., Xin, Y., & Wang, C. (2018). Impacts of Land Cover and Soil Texture Uncertainty on Land Model Simulations Over the Central Tibetan Plateau. *Journal of Advances in Modeling Earth Systems*, 10(9), 2121–2146. <https://doi.org/10.1029/2018MS001377>
- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7). <https://doi.org/10.1029/94jd00483>

- Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., M. Van Dijk, A. I. J., et al. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15(2), 425–436. <https://doi.org/10.5194/hess-15-425-2011>
- Mahrt, L. (2000). Surface heterogeneity and vertical structure of the boundary layer. *Boundary-Layer Meteorology*, 96(1–2), 33–62. <https://doi.org/10.1023/A:1002482332477>
- Marques da Silva, J. R., Damásio, C. V., Sousa, A. M. O., Bugalho, L., Pessanha, L., & Quaresma, P. (2015). Agriculture pest and disease risk maps considering MSG satellite data and land surface temperature. *International Journal of Applied Earth Observation and Geoinformation*, 38, 40–50. <https://doi.org/10.1016/j.jag.2014.12.016>
- Maxwell, R. M. (2010). Infiltration in Arid Environments: Spatial Patterns between Subsurface Heterogeneity and Water-Energy Balances. *Vadose Zone Journal*, 9(4), 970–983. <https://doi.org/10.2136/vzj2010.0014>
- McCabe, M. F., Franks, S. W., & Kalma, J. D. (2005). Calibration of a land surface model using multiple data sets. *Journal of Hydrology*, 302(1–4), 209–222. <https://doi.org/10.1016/j.jhydrol.2004.07.002>
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- Myneni, R., Knyazikhin, Y., Park, T. (2015). MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-10-05 from <https://doi.org/10.5067/MODIS/MCD15A2H.006>
- NASA. (2019). Landsat 7 (L7) Data Users Handbook, 7(November), 106. Retrieved from https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1927_L7_Data_Users_Handbook-v2.pdf
- Newman, A. J., Clark, M. P., Winstral, A., Marks, D., & Seyfried, M. (2014). The use of similarity concepts to represent subgrid variability in land surface models: Case study in a snowmelt-dominated watershed. *Journal of Hydrometeorology*, 15(5), 1717–1738. <https://doi.org/10.1175/JHM-D-13-038.1>

- Niraula, R., Norman, L. M., Meixner, T., & Callegary, J. B. (2012). Multi-gauge Calibration for modeling the Semi-Arid Santa Cruz watershed in Arizona-Mexico border area using SWAT. *Air, Soil and Water Research*, 5, 41–57. <https://doi.org/10.4137/ASWR.S9410>
- Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, 116(12), 1–19. <https://doi.org/10.1029/2010JD015139>
- Oyler, J. W., Dobrowski, S. Z., Holden, Z. A., & Running, S. W. (2016). Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. *Journal of Applied Meteorology and Climatology*, 55(7), 1441–1457. <https://doi.org/10.1175/JAMC-D-15-0276.1>
- Pan, M., Cai, X., Chaney, N. W., Entekhabi, D., & Wood, E. F. (2016). An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophysical Research Letters*, 43(18), 9662–9668. <https://doi.org/10.1002/2016GL069964>
- Peters-Lidard, C. D., Blackburn, E., Liang, X., & Wood, E. F. (1998). The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *Journal of the Atmospheric Sciences*, 55(7), 1209–1224. [https://doi.org/10.1175/1520-0469\(1998\)055<1209:TEOSTC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<1209:TEOSTC>2.0.CO;2)
- Prata, A. J., V. Casellescoll, C., Sobrino, J. A., & Otle, C. (1995). Thermal remote sensing of land surface temperature from satellites: current status and future prospects. *Remote Sensing Reviews*, 12(3–4), 175–224. <https://doi.org/10.1080/02757259509532285>
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169, 156–173. <https://doi.org/10.1016/j.agrformet.2012.09.012>
- Rosero, E., Yang, Z. L., Wagener, T., Gulden, L. E., Yatheendradas, S., & Niu, G. Y. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition

- zones during the warm season. *Journal of Geophysical Research Atmospheres*, 115(3), 1–21. <https://doi.org/10.1029/2009JD012035>
- Rouholahnejad Freund, E., Fan, Y., & W. Kirchner, J. (2020). Global assessment of how averaging over spatial heterogeneity in precipitation and potential evapotranspiration affects modeled evapotranspiration rates. *Hydrology and Earth System Sciences*, 24(4), 1927–1938. <https://doi.org/10.5194/hess-24-1927-2020>
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), 1–25. <https://doi.org/10.1029/2008WR007327>
- Santanello, J. A., Kumar, S. V., Peters-Lidard, C. D., Harrison, K., & Zhou, S. (2013). Impact of land model calibration on coupled land-atmosphere prediction. *Journal of Hydrometeorology*, 14(5), 1373–1400. <https://doi.org/10.1175/JHM-D-12-0127.1>
- Schaaf, C., Wang, Z. (2015). *MCD43 MODIS/Terra+Aqua BRDF/Albedo White Sky/Black Sky Albedo VIS/NIR Daily L3 Global 30ArcSec CMG V006* [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-10-05 from <https://doi.org/10.5067/MODIS/MCD43D59.006>.
- Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land-atmosphere coupling and climate change in Europe. *Nature*, 443(7108), 205–209. <https://doi.org/10.1038/nature05095>
- Shuttleworth, W. J. (2012). *Terrestrial hydrometeorology*. Hoboken (N. J.): Wiley-Blackwell.
- Singh, R. S., Reager, J. T., Miller, N. L., & Famiglietti, J. S. (2015). Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations over the Southwestern U.S. *Journal of the American Water Resources Association*, 51, 2648–2667. <https://doi.org/10.1002/2014WR015686>
- Son, N. T., Chen, C. F., Chen, C. R., Chang, L. Y., & Minh, V. Q. (2012). Monitoring agricultural drought in the lower mekong basin using MODIS NDVI and land surface temperature data. *International Journal of Applied Earth Observation and Geoinformation*, 18(1), 417–427. <https://doi.org/10.1016/j.jag.2012.03.014>
- Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, 44(9), 1–13. <https://doi.org/10.1029/2007WR006513>

- Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sensing of Environment*, 242(February), 111740. <https://doi.org/10.1016/j.rse.2020.111740>
- Wagner, W., Dorigo, W., De Jeu, R., Fernandez, Di., Benveniste, J., Haas, E., & Ertl, M. (2012). FUSION of ACTIVE and PASSIVE MICROWAVE OBSERVATIONS to CREATE AN ESSENTIAL CLIMATE VARIABLE DATA RECORD ON SOIL MOISTURE. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(September), 315–321. <https://doi.org/10.5194/isprsannals-I-7-315-2012>
- Wan, Z., Hook, S., Hulley, G. (2015). *MOD11_L2 MODIS/Terra Land Surface Temperature/Emissivity 5-Min L2 Swath 1km V006* [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-10-05 from https://doi.org/10.5067/MODIS/MOD11_L2.006
- 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, A., Barlage, M., Zeng, X., & Draper, C. S. (2014). Comparison of land skin temperature from a land model, remote sensing, and in situ measurement. *Journal of Geophysical Research: Atmospheres*, 119, 3093–3106. <https://doi.org/10.1002/2014JD021606>
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., et al. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, 6(7), 1341–1359. <https://doi.org/10.5194/bg-6-1341-2009>
- Wilson, K. B., Baldocchi, D. D., Aubinet, M., Berbigier, P., Bernhofer, C., Dolman, H., et al. (2002). Energy partitioning between latent and sensible heat flux during the warm season at FLUXNET sites. *Water Resources Research*, 38(12), 301–3011. <https://doi.org/10.1029/2001wr000989>
- Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47(W05301). <https://doi.org/10.1029/2010WR010090>

- Yan, Y., Mao, K., Shi, J., Piao, S., Shen, X., Dozier, J., et al. (2020). Driving forces of land surface temperature anomalous changes in North America in 2002–2018. *Scientific Reports*, 10(1), 1–13. <https://doi.org/10.1038/s41598-020-63701-5>
- Yang, Y., Guan, K., Peng, B., Pan, M., Jiang, C., & Franz, T. E. (2020). High-resolution spatially explicit land surface model calibration using field-scale satellite-based daily evapotranspiration product. *Journal of Hydrology*, (October), 125730. <https://doi.org/10.1016/j.jhydrol.2020.125730>
- Yao, Z., Li, J., Li, J., & Zhang, H. (2011). Surface emissivity impact on temperature and moisture soundings from hyperspectral infrared radiance measurements. *Journal of Applied Meteorology and Climatology*, 50(6), 1225–1235. <https://doi.org/10.1175/2010JAMC2587.1>
- Zhang, G., Zhou, G., & Chen, F. (2017). Analysis of parameter sensitivity on surface heat exchange in the Noah land surface model at a temperate desert steppe site in China. *Journal of Meteorological Research*, 31(6), 1167–1182. <https://doi.org/10.1007/s13351-017-7050-1>