5-14-2017

Landsat Surface Temperature Product: Global Validation and Uncertainty Estimation

Kelly Laraby
kga1099@rit.edu

Landsat Surface Temperature Product: Global Validation and Uncertainty Estimation

by

Kelly Laraby

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

May 14, 2017

Signature of the Author ______________________________

Accepted by ______________________________
Coordinator, Ph.D. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Kelly Laraby
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

____
Dr. John Schott, Dissertation Advisor

____
Dr. Sophia Maggelakis, External Chair

____
Dr. Carl Salvaggio

____
Dr. Emmett Ientilucci

____
Date

Landsat Surface Temperature Product: Global Validation and Uncertainty Estimation

by

Kelly Laraby

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Surface temperature is an important Earth system data record that is useful to fields such as change detection, climate research, environmental monitoring, and many smaller scale applications like agriculture. Earth-observing satellites can be used to derive this metric, with the goal that a global product can be established. There are a series of Landsat satellites designed for this purpose, whose data archives provides the longest running source of continuously acquired multispectral imagery. The moderate spatial and temporal resolution, in addition to its well calibrated sensors and data archive make Landsat an unparalleled and attractive choice for many research applications. Through the support of the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS), a Landsat Surface Temperature product (LST) has been developed. Currently, it has been validated for Landsat 5 scenes in North America, and Landsat 7 on a global scale. Transmission and cloud proximity were used to characterize LST error for various conditions, which showed that 30% of the validation data had root mean squared errors (RMSEs) less than 1 K, and 62% had RMSEs less than 2 K. Transmission and cloud proximity were also used to develop a LST uncertainty estimation method, which will allow

the user to choose data points that meet their accuracy requirements. For the same dataset, about 20% reported LST uncertainties less than 1 K, and 63% had uncertainties less than 2 K. Enabling global validation and establishing an uncertainty estimation method were crucially important achievements for the LST product, which is now ready to be implemented and scaled so that it is available to the public. This document will describe the LST algorithm in full, and it will also discuss the validation results and uncertainty estimation process.

Acknowledgements

There are so many people that I owe the completion of my Ph.D. to, and I hope I can express just how grateful I am to them in these few words. I have been especially honored to have Dr. Schott as an advisor, not only for his expertise and wealth of knowledge in the field of remote sensing, but also for his consistent and invaluable guidance (and also patience) throughout my project. I must also thank my committee members, Dr. Carl Salvaggio, Dr. Emmett Ientillucci, and Dr. Sophia Maggelakis, who were all instrumental in my success by offering their time and considerable wisdom. My thanks to the Center for Imaging Science and the Digital Imaging and Remote Sensing Lab for providing a fun learning environment with an abundance of resources, and I also want to thank Nina Raqueno especially for making time to answer my countless questions.

I also want to thank my friends and family, who have been more gracious and supportive than I could have ever hope for. My husband Cody kept me smiling and motivated, my parents cheered me on as they have since my childhood, and my dearest friend Brittany was always there to brighten my day and make me laugh.

I would like to thank NASA and USGS EROS as well as the Jet propulsion Laboratory for their support of this work that has been long in the making.

To my husband Cody, whose love, support, and provision of tea kept me going.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A fundamental part of understanding and monitoring the Earth's surface is knowledge of surface temperature. In the simplest sense, surface temperature is how hot the ground or water surface feels to the touch. This lends itself to many applications such as weather prediction, climate change research, and agriculture. Land surface temperature and sea surface temperature are typically derived from satellites using either a split-window approach (requires multiple adjacent thermal bands), or a single thermal band method (requires knowledge of atmosphere and surface emissivity). There have been many attempts in the past to develop an algorithm to derive surface temperature values from satellite thermal images, but very few make it to a fully-validated product.

Landsat is a series of satellites managed jointly by the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS). This suite of sensors provides continuously acquired multispectral imagery at a 30 x 30 meter resolution, with thermal archives dating back to 1982 [USGS, 2013]. Landsat's moderate spatial and temporal resolutions, in addition to its well calibrated sensors and data archive, makes it an unparalleled and attractive choice for many research applications. When considering which satellites to use for developing a surface temperature product, the Landsat sensors stand out as good candidates because it would allow the product database to span the past 35 years, which is extremely useful for studying the Earth's surface over time. Additionally, having knowledge of surface temperature at a 60-120 m resolution would open up opportunities for smaller-scale research such as monitoring lakes or small farms. Historically, all of the Landsat sensors have had a single thermal band (with the exception of the most recent sensor, Landsat 8). In order to create an algorithm that can be applied to the entire Landsat thermal archive, a single band method must be used. This requires accurate knowledge of the atmosphere as well as surface emissivity. For water surfaces emissivity is well known, but a reliable source of emissivities for land surfaces is needed in order to produce a full surface temperature product.

For the past several years, the USGS has funded the Rochester Institute of Technology to work on the development of the "Landsat Land Surface Temperature Product." The algorithm was originally developed by Cook, but since it is able to estitimate both land and sea surface temperature, we have chosen to change the product name to Landsat Surface Temperature (LST). The original algorithm is able to perform atmospheric compensation and estimate surface temperature at a per-pixel level for any Landsat thermal image [Cook, 2014]. The atmospheric compensation process utilizes certain atmospheric variables at several heights, which are obtained from the North American Regional Reanalysis (NARR) database [Dee et al., 2015]. A radiative transfer program called MODTRAN is used to determine the atmospheric parameters needed to calculate surface temperature [AFRL, 2015], and various interpolation steps are utilized to obtain values of these parameters for every pixel in a Landsat scene. Initial validation studies for this algorithm were performed over water, where the emissivity of water is a constant and well-known value. When the product is fully implemented, we will use the Advanced Spaceborne Thermal

Emission and Reflection Global Emissivity Database (ASTER GED), which was developed by the Jet Propulsion Lab and provides mean emissivity values at 100 m resolution [Hulley et al., 2015].

Cook was able to validate the LST algorithm for Landsat 5, but these studies were limited to North America because of the choice in reanalysis product and source of ground truth (the latter of which was buoy measurements). Since then, we have made it possible for the algorithm to operate at a global scale by identifying a comparable global reanalysis product, as well as an acceptable source of ground truth that is available globally. The new reanalysis product that was chosen is called Modern-Era Retrospective Analysis for Research and Applications (MERRA), and the new source of ground truth is the Sea Surface Temperature (SST) product that is derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. This expansion of the LST algorithm has allowed us to perform a thorough global validation study for Landsat 7.

Besides observing how accurate our LST algorithm is for various atmospheric conditions and different Landsat sensors, an extremely important goal of our work is to be able to provide an estimated uncertainty or confidence metric with the final product. The LST product will be much more meaningful to its users if we are able to inform them how trustworthy each prediction of surface temperature is. We have developed a method of estimating the uncertainty in the surface temperature retrievals which combines standard error propagation with observations of how the LST algorithm tends to behave under various conditions (i.e. cloud proximity and transmission levels). In order to analyze how well this method works, we will compare our "predictions" of LST error to actual observed errors from the validation process.

Chapter 2

Objectives

The process of creating a Landsat Surface Temperature (LST) product is considerably complex. There is the development of the algorithm, which involves a thorough investigation of the most appropriate methodology to use. After an algorithm has been established, there are many studies and assessments of the algorithm's accuracy to be performed. It would be benefitial to outline the individual tasks that need to be accomplished, as well as identify goals that we aim to acheive throughout our work. Section 2.1 contains a brief, high-level view of what we aim to accomplish. Section 2.2 lists specific objectives we olan to complete, while Section 2.3 provides a more in-depth description of these tasks. The final section in this chapter reviews the contribution that the finished LST product will have towards the field of remote sensing. Note that when we list our objectives, the first three have been addressed by the previous investigator, Cook [Cook, 2014]. We include them in our list of objectives to provide a complete end-to-end list of tasks that must be performed in order to produce a final Landsat Surface Temperature product, and because some of them will need to be readdressed in order to achieve all the tasks listed.

2.1 Problem Statement

The goal of our work is to develop an automated process that will generate the atmospheric parameters necessary to perform atmospheric compensation on a per-pixel level, in order to achieve a surface temperature product for all past and current Landsat scenes. In addition, we aim to assess the accuracy of the surface temperature product with the ultimate goal of including a quality or uncertainty band. This latter item will be particularly beneficial for users with different accuracy requirements.

2.2 Objectives

1. Select an appropriate source of atmospheric input variables with adequate spatial and temporal resolution for use with MODTRAN and for any current/archived Landsat scene in North America.

2. Implement an automated method of determining atmospheric parameters on a per-pixel level.

3. Validate the process using available truth data for Landsat scenes over North America.

4. Validate the LST process for Landsat scenes on a global scale, and for each Landsat sensor that provides thermal imagery

5. Develop a method of estimating error/uncertainty for LST errors on a per pixel level, in order to include a quality map in the final product.

6. Form a set of recommendations for how the LST product should be implemented by USGS, and present a final assessment of the product's expected performance using the recommended approach.

2.3 Tasks

1. **Select an appropriate source of atmospheric input variables with adequate spatial and temporal resolution for use with MODTRAN and for any current/archived Landsat scene in North America.**

   Radiosondes provide accurate atmospheric profiles, but they are only available at specific locations. Because of Landsat's spatial and temporal resolution, it would be beneficial to consider certain reanalysis products that provide atmospheric variables on regularly intervaled grids. Knowledge of pressure, temperature, and humidity variables at various heights will be required. It is necessary to identify a dataset that provides these variables over the desired time and space (i.e. from 1982 to the present on a global scale).

2. **Implement an automated method of determining atmospheric parameters on a per-pixel level.**

   Given the atmospheric profile for a given Landsat scene, MODTRAN can be used to generate the radiative transfer parameters transmission, upwelled radiance, and downwelled radiance. The particular method for generating such parameters must be chosen and justified in terms of accuracy and computational efficiency. The automated process will need to format the atmospheric profile data to a MODTRAN compatible form, perform the MODTRAN runs, and generate the parameters. Since the atmospheric data will not match the resolution of the Landsat scenes, and we also wish to avoid the computational strain of running MODTRAN on a per pixel level, a study on the optimal number of runs and their physical locations will be conducted. Certain interpolation steps will also be necessary, such as a temporal interpolation of the atmospheric profile data to a given Landsat acquisition time. Additionally, the radiative transfer parameters retrieved through MODTRAN need to be interpolated to the location and elevation of each pixel in the scene. The methods of interpolation need to be chosen carefully in order to minimize the computation time as well as errors seen in the final product. The methods that we explore are located in Section 4.5, along with an analysis of the errors introduced by each method of interpolation.

3. **Validate the process using available truth data for Landsat scenes over North America.**

   The error in the surface estimations can be determined by using measured ground truth as a comparison, but truth data over land surfaces are difficult to measure without influencing the observed temperature. The direct measurement of the surface temperature of water, however, is a more simple task since the instruments can be acclimated through submersion. Buoys often carry such instruments, and therefore can provide us with truth measurements at various geographical locations so we can calculate the error associated with the predicted surface temperature at the buoy locations; these errors will be used to justify our methodology. Note that this particular task has been completed for Landsat 5 by the previous investigator, but we include this item because it is a stepping stone to the next task.

4. **Validate the LST process for Landsat scenes on a global scale, and for each Landsat sensor that provides thermal imagery**

   Initial validation efforts were limited to Landsat scenes in North America; partly because of the coverage of the reanalysis product used, and partly because of the availability of buoy truth data. Since the LST product would be much more useful as a global product, changes to the methodology are necessary in order to realize this goal. Cook has addressed the first three tasks in regard to the LST validation of

Landsat 5 for North America, but they must now be reconsidered in order to proceed to global validation and for other Landsat sensors. This will require a search for a comparable source of reanalysis data as well as an accurate source of globally available truth data.

5. **Develop a method of estimating error/uncertainty for LST errors on a per pixel level, in order to include a quality map in the final product.**
   The final LST product should include a quality map or some estimation of uncertainty on a per-pixel level. We will begin by using standard uncertainty propagation, which considers all the sources of uncertainty involved in the calculation of LST. Based on past and current validation studies, we have observed errors in the LST retrievals that are not explained by standard error propagation. This will motivate an effort to quantify the unexplained errors using cloud proximity and transmission information. The analysis for this investigation can be found in Section 5.7.

6. **Form a set of recommendations for how the LST product should be implemented by USGS, and present a final assessment of the product's expected performance using the recommended approach.**
   After completing the previous tasks, we will have analyzed the accuracy of the LST algorithm for various conditions (e.g. global location, different Landsat sensors, various atmospheric conditions, etc.). We will also have assessed our ability to estimate the uncertainty or error in the LST retrievals. At that point, we will consider how to include this information in the final product so that it is feasible to implement and beneficial for the users. We will ultimately present our recommendations such as adding biases or which metrics to provide the user to USGS, who will scale the LST algorithm into a publicly available product.

2.4 Contribution to Field

Various land and sea surface temperature algorithms and products have been developed, but there are numerous aspects of our product that uniquely contribute to the remote sensing community. Firstly, Landsat provides a spatial resolution of 30 m and about a 8 to 16 day repeat cycle, which is unparalleled by other global temperature products. This unique combination of resolutions is ideal for many LST related applications that require data at higher resolutions than are currently available. Additionally, the Landsat archives for thermal imagery date back to 1982, which has great potential to be used for observing changes and trends in the Earth's surface. Although our original endeavor was to develop an algorithm for land surface temperature, it is also applicable to sea surface temperature. Having both land and sea temperatures with Landsat's spatial and temporal resolution will be an enormous contribution to the scientific field.

The land surface temperature products currently in existence use a split-window technique, which requires multiple thermal bands and is a relatively well demonstrated process. All of the Landsat sensors (excluding Landsat 8) only have one thermal band, so our approach uses a method that only needs a single thermal band to retrieve surface temperatures. Various single band LST algorithms have been explored over the years, but few have been implemented over large data sets, over large geographic extent, or been validated using physical truth data. Our method, however, is able to generate results on a large scale because it automatically integrates atmospheric data. Even more importantly, our product will provide an unprecedented global surface temperature solution for the Landsat archive. Currently, our product has been validated for Landsat 5 using scenes in North America, and global validation of Landsat 7 is part of this study.

It is important to note that this project makes a significant stride towards using the

Landsat satellites to their full potential. The excellent spatial and temporal resolution of Landsat has already been mentioned, but even more impressive is the consistent availability of thermal data reaching back to 1982. This archive of high quality Earth images would prove extremely useful to many scientific communities, but currently the dataset remains largely untapped. The Landsat Surface Temperature product will use this exceptional archive to provide a much needed dataset to the remote sensing community and other scientists, which in turn will lead to an increase in the number of people considering Landsat for their projects and applications.

Since this work is an expansion of work that was done by the previous investigator, Cook, we wish to highlight the new contributions that will be made. Cook's contribution was developing a baseline LST tool for Landsat 5, and for North American regions. Our current work will extend the LST tool for use with the entire Landsat thermal archive, and the process will be validated on a global scale for Landsat 7. Additionally, a quality map will be established, which will report expected errors/uncertainties to users. This will be a crucially important addition to the LST tool, because it will allow users to select data that meets their specific accuracy requirements.

Chapter 3

Background

Remote sensing allows us to gather information about some object or scene of interest without coming into contact with it. For our purposes, we want to determine the temperature of the Earth's surface using imagery from the Landsat series of satellites. This requires knowledge of radiometric theory for thermal applications, a good understanding of the atmosphere, a suitable validation method, and other information pertinent to our goals. This section is designed to provide background theory as well as introduce concepts and terminology that are directly involved in the Landsat Surface Temperature (LST) methodology (Chapter 4).

This chapter begins by reviewing radiative transfer theory and the governing equation that allows us to calculate the surface temperature from the sensor reaching radiance (Section 3.1 and 3.2). In Section 3.3 we present the history of the Landsat series of satellites since the LST product is being designed for use with Landsat imagery. This will be followed by a thorough discussion about land surface temperature, its applications, and previously used techniques and validation methods (Section 3.4). Sections 3.5 reviews the main two reanalysis products that are utilized in our process, and Section 3.6 presents a basic description of MODTRAN. Then, Section 3.7 introduces the standard error propagation of the governing equation, which will aid in our efforts to establish a quality band. Finally, Section 3.8 discusses the conversions that are made throughout the LST process.

3.1 Sensor Reaching Radiance

Thermal infrared radiation ranges from about 8-13 microns on the electromagnetic spectrum. Since the thermal bands on the Landsat satellites respond to wavelengths between 10.40μm and 12.50μm, we will only consider thermal sources of radiation that reach the sensor. Figure 3.1 illustrates the four different energy paths that contribute to the observed radiance. Path A represents upwelled radiance, which is the energy emitted from the particles within the atmosphere. Path B shows downwelled radiance, which is when energy is emitted from the atmosphere and interacts with the target before reaching the sensor. The target itself can emit energy towards the sensor, which is illustrated by path C in Figure 3.1. Path D indicates that some thermal energy can originate from background objects before reflecting off the target and reaching the sensor. For our purposes, we assume the contribution of path D is negligible because in most cases the background objects are few or they do not block much of the sky. We are then left with paths A, B, and C as significant contributors to the sensor observed radiance, which leads us to the governing equation discussed in the next section [Schott, 2007].

3.2 Governing Equation

It has been established that upwelled radiance, downwelled radiance, and radiance emitting from the target contribute significantly to the overall radiance observed by the sensor.

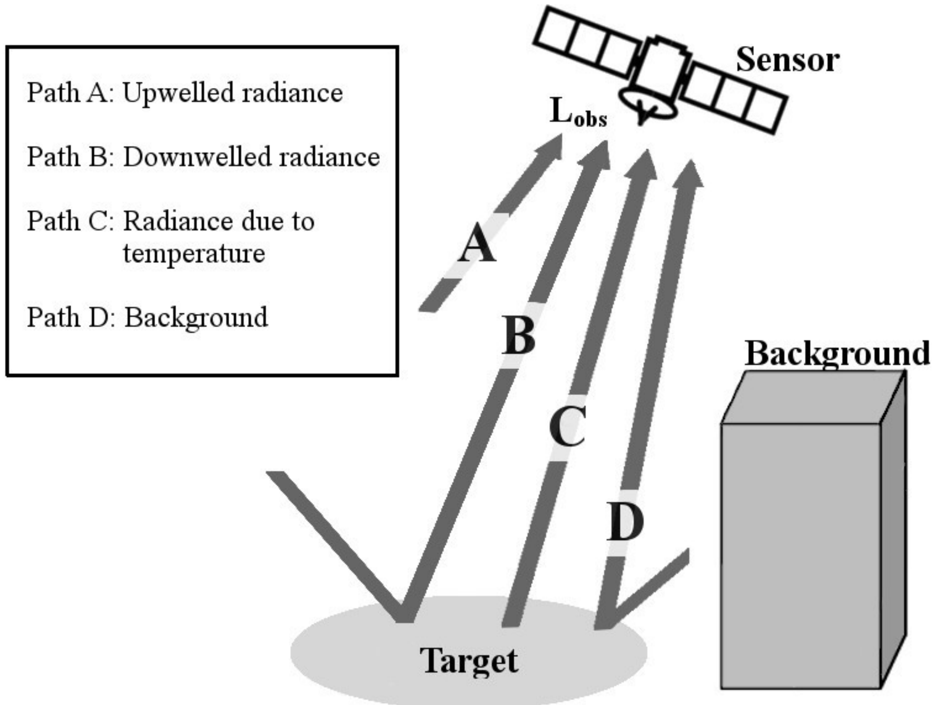

Figure 3.1: Thermal energy paths that contribute to the sensor reaching radiance.

Besides these energy paths we must consider properties such as the emissivity of the target and the transmission of the atmosphere that the light rays are traveling through. Using all these factors we can now introduce Equation 4.1 which we will refer to as the governing equation.

$$L_{obs\lambda eff} = (L_{T\lambda eff}\,\epsilon + (1-\epsilon)\ L_{d\lambda eff})\ \tau + L_{u\lambda eff} \tag{3.1}$$

In this equation, $L_{obs\lambda eff}$ is the effective spectral radiance reaching the sensor, $L_{T\lambda eff}$ is the effective spectral radiance due to temperature (Path C in Figure 3.1), ϵ is the emissivity of the surface of interest, $L_{d\lambda eff}$ is the effective spectral downwelled radiance (Path B in Figure 3.1), τ is the transmission, and $L_{u\lambda eff}$ is the effective spectral upwelled radiance (Path A in Figure 3.1). The terms that are in effective spectral radiance are marked by λeff, and incorporate the spectral response function of the sensor to obtain a single value of radiance per unit wavelength (with units $Wm^{-2}sr^{-1}\mu m^{-1}$). Equation 3.2 shows how effective spectral radiance is obtained by starting with spectral radiance L_λ and a spectral response function $R(\lambda)$.

$$L_{\lambda eff} = \frac{\int L_\lambda R(\lambda) d\lambda}{\int R(\lambda) d\lambda} \tag{3.2}$$

Since we plan to use the Landsat series of satellites to derive a surface temperature product, we can find the value of $L_{obs\lambda eff}$ on a per pixel level by converting calibrated digital numbers to radiance using Equation 3.3. Q_{cal} is the quantized pixel value in digital counts (ranging from 0 to 255 for the 8-bit sensors in Landsat 4,5, and 7, and ranging from 0 to 4097 for Landsat 8), and Q_{calmax} and Q_{calmin} are the maximum and minimum Q_{cal} values. LMAX$_\lambda$ is the effective spectral radiance that corresponds to Q_{calmax}, and LMIN$_\lambda$ is the effective spectral radiance that corresponds to Q_{calmin}. Because of the continued efforts to maintain the calibration of the Landsat sensors, we will assume that the sensors are correctly calibrated and that the calibration coefficients used to obtain radiance values can be trusted [Schott et al., 2012].

$$L_{obs\lambda eff} = \left(\frac{LMAX_\lambda - LMIN_\lambda}{Q_{calmax}} \right) Q_{cal} + LMIN_\lambda \tag{3.3}$$

The governing equation shows how various radiance terms contribute to the sensor reaching radiance, but the term of interest for our application is $L_{T\lambda eff}$, the effective spectral radiance emitting from a target or pixel of interest. Momentarily disregarding the spectral response function, we can use Planck's equation to determine the electromagnetic radiation emitting from a black body (Equation 3.4). This incorporates constants h, c, and k, which represent Planck's constant, the speed of light, and the Boltzman constant, respectively. The T in the equation is the temperature of the black body object. To reach an equation for $L_{T\lambda eff}$, we must account for the sensor spectral response function. This modification is shown Equation 3.5, where $L_T(\lambda)$ refers to Planck's equation defined in 3.4.

$$L_T(\lambda) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \tag{3.4}$$

$$L_{T\lambda eff} = \frac{M_{\lambda eff}}{\pi} = \frac{\int L_T(\lambda)R(\lambda)d\lambda}{\int R(\lambda)d\lambda} \tag{3.5}$$

We now have the equation for the term of interest, $L_{T\lambda eff}$, but what we are truly interested in finding is the temperature. It cannot be directly solved for with the Planck equation, so instead we generated a Look Up Table (LUT) by calculating $L_{T\lambda eff}$ for a range of temperatures in increments of 1 Kelvin. Then, for any value of $L_{T\lambda eff}$ we use linear interpretation to obtain the corresponding temperature.

3.3 Landsat History

Landsat was first conceived in a time where weather satellites existed, but there were no instruments specifically designed to observe various properties of the Earth's surface. The first Landsat sensor was launched by NASA in 1972, and now it provides the longest set of continuously acquired moderate resolution imagery of the Earth [Irons and Rocchio, 2015a]. This type of data, which became publicly and freely available in 2009, is extremely useful for fields such as agriculture, environmental monitoring, land mapping, and change detection [Irons and Rocchio, 2015b]. This section will present an overview of the history of the Landsat satellites as well as some details of their mechanical and optical designs.

3.3.1 Landsat 1, 2, and 3

When Landsat 1 was first being built, it was actually referred to as the Earth Resources Technology Satellite (ERTS). This instrument was launched July 1972, carrying the Return Beam Vidicon (RBV) and the Multispectral Scanner (MSS). Landsat 1 functioned until 1978, but Landsat 2 was launched in January of 1975. Landsat 2 carried the same payload, and was eventually decommissioned in 1983. Landsat 3 was launched in March, 1978, which had improved ground resolution for the RBV instrument and the MSS had a fifth spectral band (a thermal band) that failed shortly after launch. All three Landsat sensors were at an altitude of about 900 km with an inclination angle of 99.2°, and they all had polar, sun synchronous orbits and an equatorial crossing time of 9:42 AM mean local time (descending node) [Irons and Rocchio, 2015a].

3.3.2 Landsat 4

Launched in July 1982, Landsat 4 underwent significant changes such as the addition of a thermal band. On board were the MSS instrument as well as the Thematic Mapper (TM), which is a whisk-broom system that had 7 spectral bands and better resolution than the first three Landsat satellites. Whisk-broom sensors consist of an oscillating mirror that moves in the across track direction, while the satellite motion provides the along track motion. Figure 3.2 illustrates a general layout of a whisk-broom system, and Figure 3.3 shows the correction that must be made to images captured by whisk-broom systems like the TM. Considering the Thematic Mapper, the specific ranges for the spectral bands as well as their resolution can be found in Table 3.1 [Irons and Rocchio, 2015a].

Table 3.1: Spectral bands for Landsat 4 Thematic Mapper [USGS, 2014].

Bands	Spectrum Area	Wavelength Range (μ m)	Resolution (m)
Band 1	Visible	0.45 - 0.52	30
Band 2	Visible	0.52 - 0.60	30
Band 3	Visible	0.63 - 0.69	30
Band 4	Near IR	0.76 - 0.90	30
Band 5	Near IR	1.55 - 1.75	30
Band 6	Thermal	10.40 - 12.50	120
Band 7	Mid-wave IR	2.08 - 2.35	30

3.3.3 Landsat 5

Landsat 5, launched in March of 1984, carried identical versions of Landsat 4's MSS and TM instruments. The satellite was expected to be operational for at least three years, but both the MSS and TM exceeded this goal by far. The MSS instrument experienced downlinking problems and was deactivated in 1993, while the TM continued to function

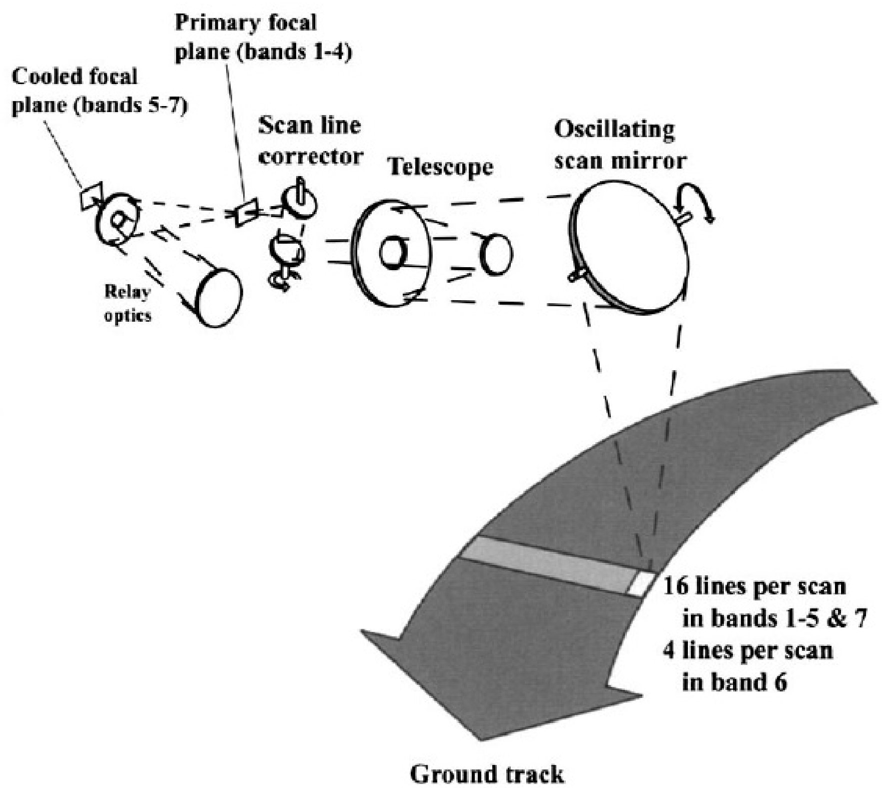

Figure 3.2: Optical layout of a line scanner [Schott, 2007].

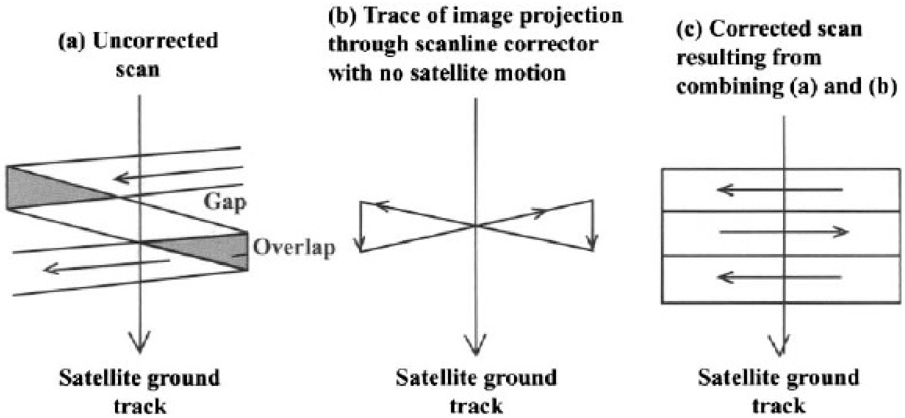

Figure 3.3: Illustration of a scan line corrector being appied to a line scanner [Schott, 2007].

until Landsat 5 was officially decommissioned in 2013. There was a point in late 2011 where image acquisition was halted because some electronic components were degrading, but the overall longevity of Landsat 5 has provided us with a vast supply of archived data [Irons and Rocchio, 2015a].

3.3.4 Landsat 6

The first three Landsat sensors were government owned and maintained, but in 1984 Landsat 4 and 5 were moved to commercial hands. Developing Landsat 6 became a privatized venture, but upon its launch in 1993 it failed to reach orbit. Problems such as gaps in the data acquisition, infrequent calibration studies, as well as the high price of imagery were becoming straining on the user community. Fortunately, the longevity of Landsat 5 prevented a gap in image acquisition, and in 2001 the operational control of Landsat sensors was returned to government hands [Irons and Rocchio, 2015a].

3.3.5 Landsat 7

Landsat 7 was launched in April, 1999, and the instrument on board was the Enhanced Thematic Mapper Plus (ETM+). The ETM+ doubles the resolution in the thermal band, and also adds a panchromatic band with an impressive 15 meter resolution. Specific wavelength ranges and resolutions can be found in Table 3.2 [USGS, 2014].

Table 3.2: Spectral bands for Landsat 7 [USGS, 2014].

Bands	**Spectrum Area**	**Wavelength Range (μm)**	**Resolution (m)**
Band 1	Visible	0.45 - 0.52	30
Band 2	Visible	0.52 - 0.60	30
Band 3	Visible	0.63 - 0.69	30
Band 4	NIR	0.77 - 0.90	30
Band 5	NIR	1.55 - 1.75	30
Band 6	Thermal	10.40 - 12.50	60
Band 7	MWIR	2.09 - 2.35	30
Band 8	Panchromatic	.52 - .90	15

Landsat 7 provides extremely accurate measurements and has remained stable and well characterized/calibrated. The excellent data quality, as well as the archive becoming free in 2009 have caused a significant increase in the number of people using Landsat data for research purposes. Unfortunately, in 2003 the scan line corrector ceased to function which introduced the bowtie effect to the imagery (see Figure 3.3). This effect is most drastic in the across track direction as pixels get farther away from the image center. These gaps in the imagery make up about 25% of each scene, which makes calibration efforts more difficult [Irons and Rocchio, 2015a]. This has steered scientists towards using the non-gap pixels from multiple acquisitions of the same location, which is often more convenient when the images are in some physical units such as reflectance or temperature. It is clear that the completed Landsat Surface Temperature product would prove useful for such activities, although it is not our main motivation.

3.3.6 Landsat 8

NASA developed the Landsat Data Continuity Mission (LDCM), which was launched in February of 2013, but three months later the control of operations switched over to USGS and LDCM was renamed to Landsat 8. Landsat 8 was much anticipated since Landsat 5 was nearing the end of its life and Landsat 7 has the scan line corrector issue. The payloads on this newest satellite include the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS), where OLI consists of nine bands and TIRS has two. Table 3.3 shows the bands that correspond with each sensor, but the most notable changes are the addition of a cirrus band as well as the second thermal band [Irons and Rocchio, 2015a].

Table 3.3: Spectral bands for Landsat 8 [USGS, 2014].

Bands	**Sensor**	**Spectrum Area**	**Wavelength Range (μm)**	**Resolution (m)**
Band 1	OLI	Coastal Aerosol	0.43 - 0.45	30
Band 2	OLI	Blue	0.45 - 0.51	30
Band 3	OLI	Green	0.53 - 0.59	30
Band 4	OLI	Red	0.64 - 0.67	30
Band 5	OLI	NIR	0.85 - 0.88	30
Band 6	OLI	SWIR 1	1.57 - 1.65	30
Band 7	OLI	SWIR 2	2.11 - 2.29	30
Band 8	OLI	Panchromatic	.50 - .68	15
Band 9	OLI	Cirrus	1.36 - 1.38	30
Band 10	TIRS	TIRS 1	10.60 - 11.19	100
Band 11	TIRS	TIRS 2	11.50 - 12.51	100

Unlike its predecessors, Landsat 8 carries push-broom sensors rather than using a whisk-broom design. Push-broom sensors use linear arrays to capture lines of data in the across track direction simultaneously, where the different linear arrays can be designed to capture different ranges of spectral wavelengths. The motion of the in-flight sensor provides the along track direction. This design increases dwell time and eliminates the need for a scan line corrector. See Figure 3.4 for an illustration of how a pushbroom sensor operates [Schott, 2007].

Even though Landsat 8 has two thermal bands, our method of obtaining LST values can still be utilized on a single band basis. Validation studies have not been applied to Landsat 8 scenes yet because of issues with stray light from outside the field of view causing banding in the images [Montanaro et al., 2014]. Currently, the Landsat team at RIT is making great efforts to develop a correction for any stray light present in an image, and once a final decision on calibration of Landsat 8 has been made the LST process can safely be validated for this sensor.

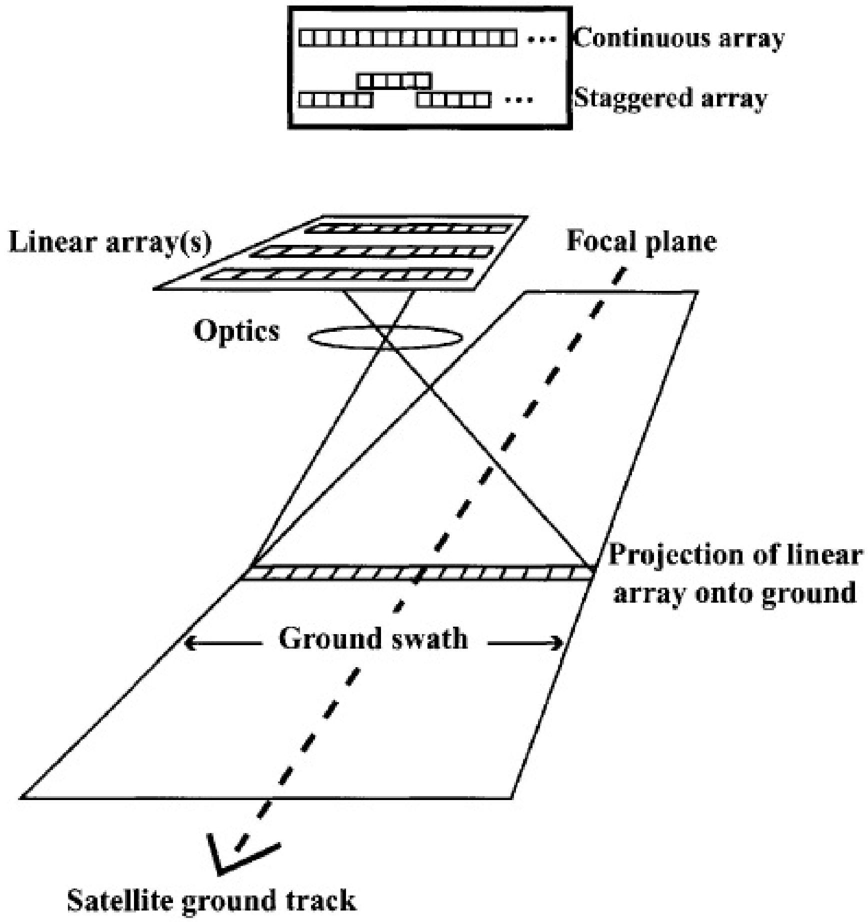

Figure 3.4: Illustration of how pushbroom sensors operate [Schott, 2007].

3.4 Land Surface Temperature

Land Surface Temperature (LST), in the most basic sense, is how hot the ground feels to the touch. From the point of view of a satellite, the land surface is the first solid surface encountered, whether it is the actual ground or something like the top of a tree canopy or building. This surface is considered to be a few millimeters in thickness, and can be any type of terrain such as grass, forest, desert, snow, water, among others [Wan and Dozier, 1996].

Problems with observing the Earth's surface often occur with the presence of clouds, which prevents the satellite from gathering accurate measurements of the land surface.

The surface temperature of the Earth, especially on a global scale, would be useful by itself as well as for use in obtaining other variables and properties of the Earth's surface and terrain. There have been various LST algorithms and products developed over the years, as well as validation and error analysis techniques that need to be understood before we can describe our own method. This section will discuss the many applications for a LST product, the differences between a single and multiple channel method, and the validation/error analysis methods that have been used in previous products.

3.4.1 Applications

Several environmental factors as well as interactions between the ground and the atmosphere lead to a measurable land surface temperature. It is intertwined with many different properties and characteristics of the Earth at a surface and atmospheric level; therefore, a LST product lends itself to a host of applications and research areas including the hydrologic process, terrestrial biosphere dynamics, climate studies, change detection, agricultural studies, and more.

One such field that would benefit from knowledge of surface temperatures is hydrology, which among other things includes monitoring soil moisture, water cycles, assessing water resources, and evapotranspiration. A particular study by Schott in 1986 investigates the uses of remotely sensed data for observing the effects of the thermal bar (a phenomenon seen in dimictic lakes) on water quality [Schott, 1986]. In 2001, a journal article detailed the process and results of the postlaunch calibration of Landsat 7, with an emphasis on the imagery's potential for water resource studies [Schott et al., 2001]. Extensive drought and evapotranspiration studies have been performed by Anderson and Kustas (2008), who have utilized several satellite-derived LST products that vary in temporal and spatial resolution in order to map various forms of biospheric stress [Anderson and Kustas, 2008].

The Landsat LST product would also be a significant boon for researchers interested in climate studies, changes, and other meteorological applications. One example is the study of weather patterns specifically relating to monsoons, which heavily relies on the contrast between land and sea temperatures [Meehl, 1994]. Another application would be observing the Earth's surface temperatures over many years in order to identify trends, such as rising temperatures in lakes [Schneider et al., 2009].

Although the large scale applications we have mentioned are numerous and important, there are also multiple smaller scale operations that would greatly benefit from a publicly available, moderate resolution LST product. In agriculture, for instance, the LST product could be used to determine water requirements for a certain plot of land throughout the year [Jackson et al., 1977], or identify frosts in orange groves [Caselles and Sobrino, 1989]. These regional or local studies require a higher resolution source of LST information than is currently available; a need which will be met by the Landsat-derived product.

3.4.2 Multiple Thermal Band Approach

The most common way to obtain LST values from satellite data is to use a split-window approach, which uses the difference in absorptions of adjacent thermal bands to compensate for the atmosphere. A split-window approach is often chosen because unlike single band approaches, there is no need for a radiative transfer model and atmospheric profile information. Even though a single band method must be used in order to obtain LST values using the Landsat satellites, it is still important to review split-window approaches and how accurate they are. This section will summarize several algorithms that use multiple thermal bands to obtain surface temperatures, as well as the satellites that the algorithm is designed for.

Wan and Dozier (1996) present a generalized split-window algorithm for obtaining land surface temperatures from AVHRR (Advanced Very High Resolution Radiometer) and MODIS (MODerate resolution Imaging Spectroradiometer) data. The AVHRR sensor used in this particular text was on board the NOAA-11 satellite, and the MODIS sensor is found on the Terra and Aqua satellites. Their main focus was to develop an algorithm that would be accurate for large viewing angles (which can be as large as $55°$ for MODIS and $69°$ for AVHRR), and they also wanted the algorithm to be insensitive to uncertainties in atmospheric properties and surface emissivities. They began with a linear version of the Wan-Dozier split-window LST algorithm [Wan and Dozier, 1989] and used radiative transfer simulations and regression analysis to obtain the coefficients necessary to calculate surface temperature. They found that the view-angle dependent method produced consistently lower errors, which was expected since the optical path is much longer for high view angles. In other words, there is more atmosphere to travel through between the ground and the sensor, which needs to be accounted for in order to reduce LST errors. They also found that including column water vapor (at 0.5 cm intervals) improved the accuracy of the algorithm so that the errors are less than 1.7 K at angles up to $69°$. A sensitivity analysis was performed in regard to emissivity, which showed that the view-angle dependent algorithm was influenced much less than the independent algorithm. Considering an uncertainty of 0.005 in emissivity, the independent and dependent algorithm was expected to reach errors of 0.8 K and 0.37 K, respectively. [Wan and Dozier, 1996].

Vasquez, et al. (1997) compared the accuracy and overall performance of four split-window algorithms for a certain set of AVHRR (aboard NOAA-12) images at a specific location. The location was chosen because of its low atmospheric water vapor content, its low probability of having clouds, and its flat terrain consisting of grass and soil. Only scenes up to a 30° view angle were used, and any cloudy pixels were removed. Images from 1994 to 1995, and for months March through December were used to span different seasons and a variety of surface temperatures. Every algorithm tested requires a priori knowledge of emissivity, so emissivities were chosen based on season and vegetation cover. Subterranean temperatures (5 cm) provided by the Meteorological Office were used as truth data to compare with LST estimations. In general, all the tested algorithms tended to overestimate the land surface temperature, where the root mean square deviations were 3.8 K, 3.0 K, 2.3 K, and 1.9 K and the mean bias deviations were 3.3 K, 1.8 K, 0.1 K, and 0.7 K, respectively. They each had different strengths and weaknesses such as sensitivity to emissivity error. It was pointed out in the text that because the algorithms were tested for specific atmospheric conditions, their performance depended on how similar the conditions of the test site were to the ones used to derive the algorithm. In addition, it was noted that using the subterranean temperatures introduced some amount of error and that further studies would be required [Vasquez et al., 1997].

Qin, et al. (2001) proposed another algorithm for retrieving LST from AVHRR images, which requires fewer parameters and claims to be more accurate than previous algorithms. As we have seen in earlier algorithms, there are various coefficients that are applied to the brightness temperatures in bands 4 and 5 of the AVHRR sensor. This particular method has three coefficients that are determined by transmittance, ground emissivity, and viewing angle, the latter of which is known for every scene. The ground emissivity was estimated using a Normalized Difference Vegetation Index (NDVI) technique as described by [Sobrino et al., 2001]. The transmittance is derived through column water vapor information and LOWTRAN7 simulations. In terms of sensitivity to emissivity uncertainties, the average LST error was 0.35° K and 0.71° K for emissivity errors of 0.005 and 0.10, respectively. The algorithm was validated using atmospheric simulations as well as a limited set of ground truth data. For the simulation case where transmittance and emissivity is known perfectly, the average LST error was 0.25 K. Using ground truth measurements without in situ knowledge of column water vapor, the average error was 1.75 K. For a different set of ground truth that included in situ atmospheric water vapors, the average error was 0.24 K [Qin et al., 001a].

Sun and Pinker (2003) apply different LST algorithms to GOES-8 (Geostationary Operational Environmental Satellite) data; the first is the generalized split window approach (with and without the inclusion of water vapor content), and the second is a three channel method. The three channel method can be applied to any three thermal bands, but in this particular effort two thermal bands and one mid-wave IR (MIR) band from GOES were used. While the inclusion of a MIR band improves atmospheric compensation, it is not ideal for daytime images because of the added influence of solar radiation. Therefore, the three channel algorithm using the MIR band is more suited for nighttime applications, where there is negligible solar radiation. The surface emissivities were obtained from the Moderate Spectral Atmosphere Radiance and Transmittance (MOSART) spectral library, and radiative transfer was performed using MODTRAN. A variety of sources were used as "truth" data to asses the accuracy of these algorithms (further discussion of this can be found in Section 3.4.4). In general, for the new split window and three channel approach, the RMS errors are typically less than 0.5 K, but can go beyond 1 K for view angles larger than 6° and temperatures greater than 300 K. Also, the proposed split window algorithm performs best for daytime scenes, while the three channel algorithm provides most accurate LST values at nighttime [Sun and Pinker, 2003].

Yu and Privette (2005) dealt with Visible Infrared Imaging Radiometer Suite (VIIRS) data, where VIIRS was an instrument on board the National Polar-orbiting Operational Environmental Satellite System (NPOESS). This particular sensor was designed to be a similar but improved version of AVHRR and MODIS, so naturally there was much interest in extending LST efforts to this newer sensor, which has a spatial resolution of 750 x 750 m. The goal of this particular work was to evaluate the VIIRS LST algorithms using real satellite data. The algorithms included two dual split window (DSW) approaches for daytime and nighttime images, and a backup split window approach. The regular split window approach uses two thermal bands, while the DSW methods use two short-mid IR and two thermal IR bands. The VIIRS LST methods were applied to MODIS data, in order to compare with the MODIS LST product. Differences between the products were typically less than 2 K, but it was discovered that the split window algorithm performed much better than the DSW approaches for both daytime and nighttime conditions. This is easily explained by the fact that the DSW methods include MIR bands, which sees a high variation in emissivity (especially over land), and during the daytime there is solar contamination which contributes to the LST error. This study reveals the difficulty in developing an accurate LST product using a DSW algorithm [Yu and Privette, 2005].

As briefly mentioned, a daily LST product exists for the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument using multiple thermal bands. MODIS provides global coverage with a high temporal frequency, a spatial resolution of 1 km for thermal bands, and accurately calibrated data. The LST algorithm utilizes a generalized split-window algorithm and a physics-based day/night algorithm. The generalized split-window algorithm uses the brightness temperature from bands 31 and 32, while the day/night algorithm utilizes seven thermal infrared bands for a pair of day/night images. Using several TIR bands allows atmospheric compensation without the need for surface data or additional atmospheric profiles. Emissivity is estimated from land-cover types using thermal infrared BRDFs and emissivity modeling. For certain land cover types with temperatures from 263 K to 300 K, the error in the MODIS LST product is less than 1 K, but in semi-arid regions where the estimated surface emissivity can be inaccurate the product often underestimate temperatures [Wan et al., 2004]. Although the LST product using MODIS has met with significant success, it has a lower spatial resolution than Landsat, and is therefore difficult to apply in certain small-scale applications such as local agriculture or irrigation studies.

Gillespie et al. (1998) developed a method of retrieving land surface temperatures as well as emissivity spectra for ASTER and MODIS imagery. The spatial resolution of ASTER is 90 x 90 m, and for MODIS it is 5 km x 5 km. The temporal frequency of ASTER is every 16 days, and for MODIS is it daily. This particular algorithm uses temperature and emissivity separation (TES), which is known for being difficult to implement because it is an underdetermined problem. In this work, the first step of the TES algorithm is to estimate emissivities and temperature using the Normalized Emissivity Method (NEM). Then the Ratio Module calculates the emissivity band ratios (this preserves the shape but not amplitude of the spectral emissivity). The final step is to recalculate surface temperature with the atmospherically corrected radiance and new emissivity values. Validation of this technique involved numerical simulations and simulated ASTER imagery, which indicate that the algorithm will perform within the 1.5 K specifications [Gillespie et al., 1998]. Directly related to this is work done by Hulley and Hook (2011), who propose a land surface temperature and emissivity product for both ASTER and MODIS. This is different than the individual LST products that were developed for these two sensors, which varied in spatial and temporal resolutions and were hard to compare. Hulley and Hook proposed the use of the TES algorithm on both ASTER and MODIS data, which combines the higher resolution but infrequent revisit cycle of ASTER with the lower resolution but high temporal frequency of MODIS. In terms of validation, this new LST method was compared to ground measurements at two sites; the Algodones Dunes and the Salton Sea. They found that the ASTER and MODIS products match closely, with differences less that 1 K

[Hulley and Hook, 2011].

The last multiple band LST method that we will mention was developed by Freitas et al. (2011). The goal of this work was to generate near real time LST values using multiple geostationary satellites. These satellites are Meteosat Second Generation (MSG), Geostationary Operational Environmental Satellite (GOES), and Multifunction Transport Satellite (MTSAT). Although a full validation was not performed, it is still interesting because of the use of multiple satellites. A generalized SW algorithm was applied to the Spinning Enhanced Visible and Infrared Imager (SEVIRI) data, which is the instrument on board the MSG. It was found that accuracy largely depended on view angle and atmospheric water vapor content. For GOES and MTSAT imagery, a dual algorithm (mono-channel for daytime, two channel method for nighttime) was applied. The mono-channel method used one TIR band, and the two channel method used one MIR and one TIR band. Since the MIR band is only used in nighttime conditions, contamination due to solar radiation is avoided. Coefficients for the dual algorithm require knowledge of TOA brightness temperature, forecasted total column water vapor, land-cover classification, and viewing angle. Results showed that the two-channel method was comparable to the generalized split window algorithm, but the single channel method had a significant increase in uncertainty. This is further evidence that either two channels must be used to perform more accurate atmospheric compensation, or the single channel must be supplemented with atmospheric profile information. Compared to an independent set of simulations, uncertainties range from 2 K for the multi-channel algorithms up to 4 K for the mono-channel method[Feitas et al., 2011].

3.4.3 Single Band Approach

As mentioned previously, single band approaches to LST retrieval is a more difficult task because it requires accurate atmospheric profile information, and a temperature emissivity separation cannot be used. We also saw studies in the previous section where an increase in uncertainty was observed when a mono-channel method was used compared to a multi-channel method. These single channel methods still need to be explored and improved, however, because it is the only hope for creating a global LST product for the entire Landsat thermal archive. This section is dedicated to reviewing several single channel algorithms that have been developed and tested to perform LST retrievals from a variety of satellites.

Sun et al (2004) investigate two LST methods for GOES satellites M-Q. One algorithm is a single channel method that requires total precipitable water data, and the second is a two channel method that uses one TIR and one MIR band. Similar to the results found by Freitas et al. (2011), this study showed that the two channel method was comparable to Wan and Dozier's generalized split window method (RMS errors around 2 K), and that the single channel algorithm was less accurate [Sun et al., 2004].

Sobrino et al. (2004) evaluate the performance of three different LST algorithms applied to Landsat 5 imagery. The first method of retrieving LST values was the simple use of radiosonde data and radiative transfer simulations , which was used as a source of "truth" to compare with the other two methods. These other methods were single channel algorithms developed by Qin et al. (2001) and Jimenez-Munoz and Sobrino (2003). The surface emissivities were estimated using the Normalized Difference Vegetation Index (NDVI) method, which uses the visible and near infrared bands to describe the amount of vegetation coverage in a particular scene. Both the algorithm proposed by Qin et al. and the one by Jimenez-Munoz and Sobrinoand require emissivity as well as an estimate of water vapor content. Qin et al.'s algorithm also needs near-surface temperatures to calculate atmospheric transmissivity. Compared to the "truth" data derived from radiosondes, MODTRAN runs, and in situ emissivity measurement, Qin et al.s algorithm had a root mean square deviation (RMSD) of 2.2 K and Jimenez-Munoz and Sobrinos single channel method had a RMSD of 0.9 K [Sobrino et al., 2004].

The following study does not involve a new LST algorithm, but rather an investigation into the various sources that contribute to LST error. Jimenez-Munoz and Sobrino (2004) used MODTRAN simulations to observe how atmospheric compensation, sensor noise, land surface emissivity, aerosols, angular effects, wavelength uncertainty and band-pass effects influence the error seen in LST retrievals. From the results it was concluded that atmospheric effects are the most impactful, and can introduce around 0.2 K if in situ data is used and 0.7 K if remote sensing data is used. Uncertainty in emissivity also has an effect on error, about 0.4 K. In general cases where in situ data is used, the minimum expected LST error was estimated to be 0.3 K, and 0.8 K for the case where remote sensing data is used [Jimenez-Munoz and Sobrino, 2004].

3.4.4 Validation Methods

Here we will review the validation methodologies used for past LST algorithms (most of which were mentioned throughout the past two sections), which have helped guide us towards the methods we have chosen to use in our work.

Vasquez et al. (1997) compared four different split window LST algorithms and how they performed when applied to NOAA-12 HRPT data (this was first discussed in Section 3.4.2). In order to evaluate the accuracy of these algorithms, subterranean temperatures measured at 5 cm below the surface were used as truth data. These measurements were managed and provided by the Meteorological Office of the Air Force, and were made available in 30 minute increments for the surrounding area. These locations were chosen for their flat terrain consisting of grassland and some patches of bare soil. The subterranean measurements that were taken closest to that of satellite morning overpass time were used. The morning overpasses were preferable because they corresponded to the times of minimum difference between the skin temperature and the 5 cm subsurface temperature. RMS deviations of approximately 2 K to 4 K were found for the tested algorithms [Vasquez et al., 1997].

Sun and Pinker (2003) analyze the performance of different LST algorithms applied to GOES-8 data, as discussed in Section 3.4.2. The accuracy of these methods were determined using three different sources of "truth" or comparison data. One of these validation studies used the Atmospheric Radiation Measurement (ARM) observations of surface skin temperature, provided by the Central Facility in Southwest Oklahoma. Another validation set involved soil temperature observations and air temperature from automated weather stations from the North Carolina Agricultural Research Service (NCARS) Weather and Climate Network. The third and final validation used the Surface Radiation Network (SURFRAD) upwelling thermal infrared radiances measured by Precision Infrared Radiometers (PIR) at four different stations, where the upwelling thermal irradiance is first converted to skin temperature, which requires an estimation of the surface emissivity. Three separate validation studies are performed comparing the algorithms to each validation data set. Errors ranged from 0.5 K to slightly greater than 2 K. The use of a variety of truth data sources allowed them to study the inherent difficulties of each type of validation data, as well as the performance of their algorithm. The ARM method yeilded an RMS error of about 1-2 K, while the soil truth method yeilded high LST errors [Sun and Pinker, 2003].

Wan et al. (2004) dealt with the validation of the MODIS LST product, as described in Section 3.4.2. This effort involves a series of specific field campaigns described in [Wan et al., 2002a] and [Wan et al., 2002b], which were conducted at Lake Titicaca in Bolivia, grasslands in Mono Lake, Bridgeport, California, rice fields in Chico, California, Walker Lake, Nevada, a silt playa in Railroad Valley, Nevada, and soybean and rice fields in Greenville, Mississsippi. In these campaigns, TIR radiometers measured lake surface kinetic temperatures, radiosonde balloons were launched from the lake shore, and winds speeds and air temperatures were recorded. Using this data and MODTRAN radiative transfer simulations, the lake surface temperature could be determined [Wan et al., 2002a]. From these field measurements, the accuracy of the "ground truth data" was found to be better than 1 K [Wan et al., 2004].

Gillespie et al. (1998) used numerical simulations to validate the ASTER TES algorithm (description of the TES algorithm can be found in Section 3.4.2). These simulations were generated for a variety of atmospheric and surface conditions, and when error-free input radiances were considered they found errors within the 1.5 K specifications [Gillespie et al., 1998]. Hulley and Hook (2011) utilized the TES algorithm to be able to compare and validate the ASTER and MODIS LST products. They utilized measured data from two sites and laboratory spectra for two sand dune sites as truth data, and found that their temperature retrievals for both ASTER and MODIS were within 1 K [Hulley and Hook, 2011].

Yu and Privette (2005) performed a validation study for their proposed LST algorithm for NPOESS VIIRS data (see Section 3.4.2). Instead of using ground truth data, the MODIS level 2 swath product was used for comparisons. The VIIRS LST algorithms were applied to MODIS radiance data and then the LST retrievals were compared to the MODIS product. Even though this a relative rather than absolute comparison, it served as a way to evaluate the algorithm performance over a large area for various atmospheric conditions [Yu and Privette, 2005].

For our validation studies in North American regions, we will use water temperatures measured by buoys and convert them to skin temperature to be used as ground truth data. This validation method has been used extensively to calibrate Landsat and other sensors [Barsi et al., 2010]. When we extend to a global LST solution, we will use the MODIS Sea Surface Temperature (SST) product as a source of ground truth.

3.4.5 Error Analysis

It is important for any LST product to be able to determine the expected errors for a variety of conditions, so that users can know the amount of uncertainty associated with the product. Traditional error analysis is difficult to do in many LST products (including ours) because several data sets and transfer codes are used. In fact, we will find that general trends and conclusions about LST error contributions are easily made, but it is rare to encounter a quantitative error analysis. This section summarizes several error analysis methods used by current LST products, which have helped shape our own approaches to error analysis. Some of these products have been mentioned in the previous sections, but the goal here is to emphasize the techniques that were used to evaluate or estimate their overall errors.

The first error analysis method we will review was used with the ASTER TES algorithm developed by Gillespie et al. (1998). The goal was to estimate the quality of their LST results in three quality assurance (QA) data planes that use eight bits. As discussed in Section 3.4.2, both numerical simulations and lab/field simulated data were used to evaluate the algorithm. Numerical simulations revealed the algorithm's performance for a variety of conditions, including changes to ground temperature, emissivity, the sensitivity to NEΔT, sky irradiance, and atmospheric compensation [Gillespie et al., 1998]. In the first data quality plane, the first four bits label data quality as good, suspect, or bad based on input data or algorithm completion. A good rating indicates that the pixel has no known defects, a suspect pixel may mean that output bands were out of range, and a bad TES value may be caused by a lack of good bands or a possible divergence of the algorithm. Bits five and six represent the cloud mask by indicating thick clouds, thin clouds, or clear conditions estimated using ASTER VNIR and SWIR data. The final two bits make up the adjacency code, which predicts the percentage of radiance that is uncorrected cloud irradiance. The percentages are placed into range categories, which are less than 10%, 10 - 20%, 20 - 30%, and greater than 30% [Gillespie et al., 1998].

The second quality data plane consists of temperature and emissivity information. The first two of eight bits specify scene conditions based on the maximum emissivity ϵ_{max}. For instance, when ϵ_{max} is ≥ 0.98, the scene is likely made up of either water, snow, vegetation, or certain moist soils. Alternatively, a default value is used when ϵ_{max} is between 0.96 and 0.98. If ϵ_{max} is between 0.94 and 0.96, this corresponds to silicate rocks, and if ϵ_{max} is less than 0.94 then there may be possible error conditions. Bits three and four indicate the algorithm convergence speed based on the number of iterations (fast to converge, nominal performance, or slow to converge. Bits five and six draw conclusions about the atmospheric compensation using the ratio of the downwelling atmospheric irradiance normalized by π to the measured land leaving radiance. For example, a ratio ≤ 0.1 indicates a high altitude scene where the compensation is probably accurate. A ratio value between 0.1 - 0.3 are known as "nominal values." If the ratio is > 0.3, the scene is likely warm with humid air or cold land, and the correction may be inaccurate. The last bits suggest whether ϵ_{max} needs to be corrected in proportion to measurements [Leff, 1999].

The third and final QA data plane is specific to temperature or emissivity. Both have two bits for accuracy and two bits for precision, categorized as poor, marginal, nominal or excellent performance. However, these bits are initially zero-filled because they are not set automatically by the processing software [Leff, 1999]. This thorough examination of the different QA planes reveals how intermediate values from the algorithm and simulation results can be indicators of quality.

Now we will discuss error analysis performed for the MODIS LST product, which as discussed in Section 3.4.2 is generated using the generalized split window algorithm [Wan and Dozier, 1996]. Additionally, this product utilizes MODIS geolocation, cloud masks, atmospheric profiles, land cover and snow cover products [Wan,]. There are many contributing sources to the overall uncertainty of the LST product, the main groups of which are the instrument, the algorithm, and the emissivity. More specifically, errors can be introduced by things such as calibration accuracy, spectral response function, optical and system noise equivalent temperature, uncertainty in the generalized split window algorithm and in the registration for the day/night algorithm, and uncertainty in the estimated emissivity for each land cover type. In order to estimate an overall LST error, a root sum square (RSS) of these uncertainties (varying with view angle and column water vapor) is calculated [Wan, 1999].

Yu et al. (2010) investigate error sources for the LST algorithm developed for the GOES-R Advanced Baseline Imager (ABI). Among the influential error sources are surface emissivity and atmospheric water vapor absorption. Also, the large local zenith viewing angles and moist atmospheric conditions are expected to cause high errors. The LST product includes product quality information derived from intermediate values in the algorithm, but does not include estimated error values nor infers the usability of the output. The quality product is defined for each pixel in 16 bits, where the first byte defines availability, surface type, and the cloud index. The categories for availability are normal, bad data, and missing data, the surface types include land, snow/ice, in-land water, and sea, and the options for cloud index are clear, probably clear, probably cloudy, or cloudy. The cloud index is determinde using the ABI cloud mask, an independent GOES-R product. The second byte specifies the atmospheric condition (dry, moist, or very moist) based on water vapor, whether it is day or night based on the solar zenith angle, the view angle (normal or large), LST quality (normal, cold or out of range), and emissivity quality (normal or historical) [Yu et al., 2010].

Now we will move on to the error analysis of the LST algorithm for NPOESS VIIRS data. Recall from Section 3.4.2 that both a regular split window algorithm using two TIR bands is used, as well as a dual split window algorithm that used two TIR and two MIR bands. For both approaches, VIIRS brightness temperature, optical thickness, cloud mask, and surface type are all used to generate the LST product. The three byte QA component is generated from various inputs and intermediate values of the LST algorithms. The first byte includes the LST quality for each pixel (high, medium, low, or no retrieval), and it specifies which algorithm was used (4-band or 2-band). It also includes bits specifying whether it is day or night, whether there is active fire, and whether thin cirrus clouds are present. These are all determined from the VIIRS cloud mask. The second byte includes the precision degradation from the brightness temperature product, an indication of whether the retrieved LST is within the acceptable range from 213 to 343 K, the confidence in the cloud mask values, whether or not the aerosol optical thickness is ≤ 1.0, the horizontal reporting interval based on the sensor zenith angle, sun glint, and terminator based on the solar zenith angle from the cloud mask. Finally, the third byte classifies the type of background (land or water) from the cloud mask, as well as the surface type from the surface type product [Ip and Siebels, 2009]. Some general conclusions are made about the effects on the LST product quality, but this does not include an estimate of quantitative error for the product.

As discussed in Section 3.4.2, Freitas et al. (2011) propose the development of a near real time LST product using multiple geostationary satellites. In their analysis, they provide an error bar associated with each LST value, which considers input errors of sensor noise, uncertainty in emissivity, statistics of total column water vapor forecast errors, and uncertainties in the retrieval algorithm. This last item is largely influenced by the optical path between sensor and surface, which changes for different viewing angle and column water vapor values. The sensitivity to each of these variables is determined using radiative transfer simulations. Comparing the results from both exact and perturbed inputs to a model allows us to observe the sensitivity of the LST algorithm to each variable. This type of modeling is used to determine the errors from each individual input, and by assuming that all sources of error are independent the LST error bar is calculated [Freitas et al., 2010] [Feitas et al., 2011].

Hulley et al. (2012) propose a temperature and emissivity uncertainty simulator, where the goal is to accurately quantify uncertainties for any sensor and algorithm combination and for a variety of atmospheric and surface conditions. Global radiosonde data from Wyoming CLAR databaseis used in conjunction with MODTRAN radiative transfer code, and emissivity is obtained from the ASTER spectral library. Similar to the previous technique utilized by Freitas et al. (2011), simulations were run with actual and perturbed atmospheres, perfect and imperfect simulated TOA radiances, and perfect inputs for simulated LST to compare to retrieved LST. These alterations to the "perfect scenario" represent adding atmospheric noise, measurement noise, and LST model error in order to observe the effect on overall LST retrieval error by calculating the RSS of these errors. This is shown in Equation 3.6, where the uncertainty conributed by the atmospheric noise, measurement noise, and model noise is represented by LST_A, LST_N, and LST_M, respectively.

$$\delta LST_{TES} = \sqrt{\delta LST_A^2 + \delta LST_N^2 + \delta LST_M^2} \tag{3.6}$$

When considering a specific sensor and algorithm, a least squares regression is performed between the simulated total LST error and a quadratic function of error contributors (e.g. total column water vapor and sensor view angle). Coefficients from this step are then applied to all pixels within a scene. Considering the LST product for ASTER and MODIS, atmospheric errors were found to be the largest source of error in both cases [Hulley et al., 2012]. This matches what was seen in the Jimenez-Munoz and Sobrino study from Section 3.4.3, who also concluded that atmospheric conditions had the greatest effect on LST uncertainty.

The final error analysis method included in this section was implemented by Hook et al. (2007). They suggest considering the errors in the atmospheric variables in order to estimate LST uncertainty. This is accomplished by perturbing the atmosphere by the atmospheric uncertainty of individual variables and conducting simulations. These alterations were made to the variables water vapor, air temperature, ozone, and visibility. Also, the path length and assumption of incorrect emissivity were varied. The final LST errors were calculated using a nominal radiance, then adjusting the profile and recalculating radiance, and finally comparing the at-sensor temperature to the nominal and adjusted radiances. The largest uncertainty was introduced with changes in visibility and column water vapor [Hook et al., 2007]. These investigations cause us to expect that, in the thermal regime, errors from atmospheric compensation will dominate the LST uncertainty.

In our efforts to estimate the error the LST retrievals produced by our algorithm, we chose to use standard error propagation. This involves evaluating various error terms and summing them in quadrature. Part of this process uses a similar method to Hook et al. (2007), where atmospheric profiles were perturbed in order to estimate uncertainty related to the atmospheric compensation process. In Chapter 5, we will also attempt to use observations from our validation results to help quantify errors that are not accounted for by the standard error propagation terms.

3.5 Reanalysis

Reanalysis is a retrospective type of analysis that aims to use observations and numerical models to generate a reliable data set for climate monitoring and research. Typically, reanalysis products utilize inputs such as radiosondes, dropsondes, aircraft, and surface data to quantify many hard-to-measure variables in the form of a spatial grid. A few examples of variables that are commonly provided include air temperature, vorticity, geopotential height, wind speed, and humidity. These are usually reported for a range of various atmospheric pressures, which correspond to heights (high pressures are closer to sea level and low pressures are higher in the atmosphere). There are several reanalysis products available that provide different variables, coverage, and spatial/temporal resolutions, which helps researchers choose one to use based on their application. For our land surface temperature project we have used the North American Regional Reanalysis (NARR) as well as the Modern-Era Retrospective analysis for Research and Applications (MERRA). NARR was used in our initial algorithm, but its coverage is limited to North America so we moved on to MERRA in order to perform global studies. The background for both NARR and MERRA will be discussed in this section.

3.5.1 NARR Data Set

The North American Regional Reanalysis (NARR) is maintained by the National Center for Environmental Prediction (NCEP), and was created to improve upon the already existing NCEP global analysis. Although the coverage was limited to North America, more variables were made available and the spatial resolution was increased to a 349 x 277 point grid with a spacing of about 32 km [Shafran, 2007]. Data at these grid points are available for any date past January 1st of 1979, and three different temporal resolutions can be chosen from. The temporal resolution used for the LST method is eight times daily, but the other options are once daily and once monthly. The NARR data is provided in GRIB 1 (General Regularly-distributed Information in Binary form) format, and is accessible through the NOMADS website ***nomads.ncdc.noaa.gov/data.php?name=access#narr_datasets*** [NOMADS, 2015].

In terms of atmospheric variables, the NARR data set provides values at 29 different pressure levels at each grid point, which essentially gives us a three-dimensional grid. The three atmospheric variables that we utilize are air temperature, geopotential height, and specific humidity at each pressure level, which after some conversions is used to create an atmospheric profile as an input into MODTRAN. The 29 pressure levels are listed below in hectopascals (hPa), where high pressures are closer to sea level and low pressures are higher in the atmosphere.

1000 hPa	850 hPa	700 hPa	400 hPa	200 hPa
975 hPa	825 hPa	650 hPa	350 hPa	175 hPa
950 hPa	800 hPa	600 hPa	300 hPa	150 hPa
925 hPa	775 hPa	550 hPa	275 hPa	125 hPa
900 hPa	750 hPa	500 hPa	250 hPa	100 hPa
875 hPa	725 hPa	450 hPa	225 hPa	

3.5.2 MERRA Data Set

MERRA is a global reanalysis product that is operated by NASA's Global Modeling and Assimilation Office (GMAO), and uses version 5 of the Goddard Earth Observing System Data Assimilation System (GEOS-5) to produce many of the same atmospheric variables that NARR provides. Like NARR, MERRA data sets are available from 1979 to the present, but the grid spacing and number of pressure levels differ. MERRA data is available in three different grid resolutions; the first is its native grid of $0.5° \times 0.667°$, the second is a reduced resolution of $1.25° \times 1.25°$, and the third as a $1.0° \times 1.25°$ option. Temporal resolution options include hourly, eight times daily, four times daily, and once monthly, but these options change by product type. For instance, the original analysis data is available in 6 hour increments (four times daily) in the native grid spacing, but assimilated fields are in 3 hour increments (eight times daily) and are in the reduced resolution of 1.25° x 1.25°. The data can be downloaded through the Modeling and AssimilationData and Information Services Center (MDISC), where the individual files are in NetCDF format. [Kempler, 2009].

Considering the three-dimensional atmospheric variables needed for our proposed LST process, we are limited to either the 6-hourly analysis at the 0.5° x 0.667° resolution or the 3-hourly product at the reduced resolution. Since it is expected that time will have a greater effect on changes in the atmosphere than spatial location, we chose to use the 3-hourly product at 1.25° x 1.25° resolution. Once again, we extract the air temperature, geopotential height, and specific humidity at the different pressure levels to perform atmospheric compensation. MERRA has 42 atmospheric pressure levels ranging from 1000 to 0.1 hPa, which extends to much higher levels than the NARR data sets. The pressure levels for MERRA are listed below.

1000 hPa	825 hPa	600 hPa	250 hPa	30 hPa	2 hPa
975 hPa	800 hPa	550 hPa	200 hPa	20 hPa	1 hPa
950 hPa	775 hPa	500 hPa	150 hPa	10 hPa	0.7 hPa
925 hPa	750 hPa	450 hPa	100 hPa	7 hPa	0.5 hPa
900 hPa	725 hPa	400 hPa	70 hPa	5 hPa	0.4 hPa
875 hPa	700 hPa	350 hPa	50 hPa	4 hPa	0.3 hPa
850 hPa	650 hPa	300 hPa	40 hPa	3 hPa	0.1 hPa

3.6 Atmospheric Compensation via MODTRAN

Radiative transfer theory, as introduced in sections 3.1 and 3.2, is an integral part of deriving land surface temperatures from satellite imagery. Several codes and programs have been developed to provide a means of propagating electromagnetic radiation through a specified atmosphere, which is often useful for simulating realistic situations without the need for expensive fly overs. One such radiative transfer code called MODTRAN (MODerate resolution atmospheric TRANsmission) has been created in a joint effort between Spectral Sciences Inc. and the US Air Force. It is capable of modeling the propagation of electromagnetic radiation between 0.2 and 100 μm, and characterizes outputs such as molecular absorption, emission, reflection, and scattering. The user has the option to provide atmospheric layer information or choose one of MODTRAN's pre-defined atmospheres, but it should be noted that the models used by MODTRAN assume a homogeneous layered atmosphere [Schott, 2007].

In our process of obtaining surface temperatures, MODTRAN is used to determine the atmospheric parameters in Equation 4.1 for individual Landsat scenes given the weather conditions and sensor geometry at the image acquisition time. We will note here that the needed parameters are not a direct output of MODTRAN, but the approach used to find them is disclosed in the methodology section (Chapter 4). The atmospheric profile that is supplied to MODTRAN is built by gathering specific atmospheric variables from a reanalysis product such as NARR or MERRA, which must first undergo some alterations in order to comply with MODTRAN's input requirements. The necessary conversions are outlined in the following section, and a more detailed discussion of how MODTRAN fits into our LST methodology can be found in Chapter 4.

3.7 Error Analysis of the Governing Equation

The governing equation illustrates how a combination of radiance paths, surface properties, and atmospheric parameters contribute to the sensor-reaching radiance. The goal is to obtain values for each of these terms in order to evaluate $L_{T\lambda eff}$, which leads us directly to a surface temperature value. Chapter 4 explains the methods used to acquire values for each term, but we should first consider how we might estimate the uncertainty in our final calculation of $L_{T\lambda eff}$. There are ways of comparing our prediction of surface temperature to certain forms of truth data, but this will only give us errors in a few specific locations. When we consider our overall goal of creating a LST product that provides the Earth's surface temperature for every pixel in every image of the Landsat archive, it reveals the need for some sort of metric that will indicate how much error or uncertainty is associated with each LST prediction. Developing a way to arrive at such a metric has become a major focus of our work.

The standard way of performing error analysis on a general function is to take the partial derivative of the function with respect to each variable, multiply the partials by the uncertainty in their respective variables, then sum all these terms in quadrature. In our case, the function we have is the governing equation, but we must rearrange it because we are interested in looking at the error in the surface leaving radiance, not the observed radiance. Equation 3.7 shows this modification of the governing equation, where the effective spectral radiance subscripts, λeff, have been omitted for simplicity.

$$L_T = \frac{L_{obs} - L_u + L_d \, \tau \, (\epsilon - 1)}{\tau \, \epsilon} \tag{3.7}$$

Now we can express the uncertainty in L_T using standard error analysis, as shown in Equation 3.8. Notice that we have also included cross correlation terms because not all of the sources of error are independent. All of the partial derivative terms in this equation are easily evaluated, and since they have no special meaning their derivations are located in Appendix C. The definitions for the uncertainty terms, S_τ, S_{L_u}, and S_{L_d} can be found in equation 3.9. This equation includes the variables S_T, S_{RH}, S_P, and S_H, which refer to the uncertainty in the air temperature, relative humidity, pressure, and geometric height profiles that are obtained from a reanalysis product. $S_{L_{obs}}$ is the uncertainty in the Landsat radiance measurement, which is different for each Landsat sensor but these values are easily found in the literature. S_ϵ is the uncertainty in emissivity values provided by the ASTER Global Emissivity Database.

$$S_{L_T} = \Bigg[\left(\frac{\delta L_T}{\delta \tau} S_\tau\right)^2 + \left(\frac{\delta L_T}{\delta L_u} S_{L_u}\right)^2 + \left(\frac{\delta L_T}{\delta L_d} S_{L_d}\right)^2 + \left(\frac{\delta L_T}{\delta L_{obs}} S_{L_{obs}}\right)^2 + \left(\frac{\delta L_T}{\delta \epsilon} S_\epsilon\right)^2 + 2\rho_{\tau L_u} \frac{\delta L_T}{\delta \tau}\frac{\delta L_T}{\delta L_u} S_\tau S_{L_u} + 2\rho_{\tau L_d} \frac{\delta L_T}{\delta \tau}\frac{\delta L_T}{\delta L_d} S_\tau S_{L_d} + 2\rho_{L_u L_d} \frac{\delta L_T}{\delta L_u}\frac{\delta L_T}{\delta L_d} S_{L_u} S_{L_d} \Bigg]^{\frac{1}{2}} \tag{3.8}$$

$$\begin{aligned} S_\tau &= \left[\left(\frac{\delta \tau}{\delta T} S_T\right)^2 + \left(\frac{\delta \tau}{\delta RH} S_{RH}\right)^2 + \left(\frac{\delta \tau}{\delta P} S_P\right)^2 + \left(\frac{\delta \tau}{\delta H} S_H\right)^2 \right]^{\frac{1}{2}} \\ S_{L_u} &= \left[\left(\frac{\delta L_u}{\delta T} S_T\right)^2 + \left(\frac{\delta L_u}{\delta RH} S_{RH}\right)^2 + \left(\frac{\delta L_u}{\delta P} S_P\right)^2 + \left(\frac{\delta L_u}{\delta H} S_H\right)^2 \right]^{\frac{1}{2}} \\ S_{L_d} &= \left[\left(\frac{\delta L_d}{\delta T} S_T\right)^2 + \left(\frac{\delta L_d}{\delta RH} S_{RH}\right)^2 + \left(\frac{\delta L_d}{\delta P} S_P\right)^2 + \left(\frac{\delta L_d}{\delta H} S_H\right)^2 \right]^{\frac{1}{2}} \end{aligned} \tag{3.9}$$

The equations presented here show, in a theoretical sense, how the error associated with L_T (S_{L_T}) can be calculated. This, however, does not prove useful unless we can create methods of evaluating the various sources of error that contribute to S_{L_T}. The details of how each error source was determined or estimated can be found in Section 4.6, where it will also be revealed that there is an additional source of error that accounts for the influence of cloud proximity.

3.8 Conversions

The variables obtained from the NARR/MERRA datasets are air temperature, geopotential height, and specific humidity at 29 different pressure levels. These are used to build the atmospheric profile layers that will be used by MODTRAN to perform atmospheric compensation, but first some alterations must be made to conform with MODTRAN's input requirements. One is that the geopotential height must be converted to geometric height, and the other is that MODTRAN does not accept specific humidity as an input. Instead, it allows one of the following humidity variables: volume mixing ratio [ppmv], mass mixing ratio [g/kg], mass density [g/m3], number density [molecules/cm3], partial pressure [mb], dew point temperature [K or °C], or relative humidity [%]. The air temperature and pressure in Kelvin and hectopascals (hPa) are already in an acceptable format for MODTRAN and do not need to be altered [Berk et al., 2003].

While geometric height is simply the elevation above mean sea level, geopotential height is a "gravity adjusted height" that depends on latitude and elevation. All the variables and constants that are used in the conversion from geopotential to geometric height are located in Table 3.4. Equation 3.10 shows how to calculate the geometric height Z from geopotential height H.

Table 3.4: Variables and constants for the conversion from geopotential height to geometric height [Wright, 1997].

Variable	**Description**	**Value [Units]**
H	given geopotential height	[m]
ϕ	latitude at location of heights	[radians]
g_0	standard acceleration due to gravity	9.80665 [m/s^2]
R_{max}	Earth's equatorial radius	6378.137 [km]
R_{min}	Earth's polar radius	6356.752 [km]

$$Z = \frac{HR_e}{GR_e - H} \tag{3.10}$$

In this conversion, G is the gravity ratio between the gravity at a specific latitude and the standard value of gravity, and can be calculated using Equations 3.11 and 3.12. R_e is the radius of the Earth at a given latitude, which is obtained using Equation 3.13.

$$G = \frac{g}{g_0} \tag{3.11}$$

$$g = 9.80616\ \left[10.002637\ cos(2\phi) + 0.0000059\ cos^2(2\phi)\right] \tag{3.12}$$

$$R_e^2 = \left(\frac{cos^2(\phi)}{R_{max}^2} + \frac{sin^2(\phi)}{R_{min}^2}\right)^{-1} \tag{3.13}$$

Specific humidity is the ratio between the the mass of water vapor and the mass of the dry air that is present, while relative humidity is ratio of the vapor partial pressure of the air to the saturation vapor partial pressure at a certain temperature. This conversion utilizes the variables and constants in Table 3.5.

Table 3.5: Variables and constants for the conversion from specific humidity to relative humidity [Kruger, 2010].

Variable	**Description**	**Value [Units]**
T_C	air temperature	[°C]
T_K	air temperature	[K]
q	specific humidity	[kg/kg]
p	pressure	[hPa]
N_L	Avogadro's constant	6.0221415 x 10^{23} [mol^{-1}]
R	universal gas constant	8.301447215 [J/(mol K)]
M_{H2O}	molar mass of water	18.01534 [g/mol]
M_{dry}	molar mass of dry air	28.9644 [g/mol]

The main equation for calculating relative humidity from specific humidity is shown in Equation 3.14, where P_{H2O} is the vapor partial pressure and e is the saturation water vapor pressure [Kruger, 2010].

$$RH = \frac{P_{H2O}}{\epsilon} \cdot 100 \tag{3.14}$$

Many methods of estimating the saturation water vapor pressure have been developed, but we have chosen to use the Grodd-Gratch equation (Equation 3.15 [Goff and Gratch, 1946]).

$$\begin{aligned}
\log(e) = &- 7.90298 \left(\frac{373.15}{T_K} - 1 \right) \\
&+ 5.02808 \log \left(\frac{373.15}{T_K} \right) \\
&- 1.3816x10^{-7} \left(10^{11.344\ (\ 1\ -\ T_K/373.15\ )} - 1 \right) \\
&+ 8.1328x10^{-3} \left(10^{-3.49149\ (\ 373.15/T_K\ -\ 1\ )} - 1 \right) \\
&+ \log(1013.25)
\end{aligned} \tag{3.15}$$

In order to calculate the partial vapor pressure, we must multiply the pressure by the volume mixing ratio, X_{H2O} (Equation 3.16). Equation 3.17 shows how to solve for X_{H2O}

using the specific humidity and the molar mass of water and dry air.

$$P_{H2O} = p\, X_{H2O} \tag{3.16}$$

$$X_{H2O} = \frac{q\, M_{dry}}{M_{H2O} - q\, M_{H2O} + q\, M_{dry}} \tag{3.17}$$

3.9 Concluding Remarks

This chapter was designed to provide sufficient background information for all aspects of this project. First, we reviewed the energy paths that contribute to the sensor reaching radiance, which led to the discussion of the governing equation. Then, a brief history of the Landsat satellites was presented, followed by an overview of LST algorithms and validation methods found in the literature. Descriptions of reanalysis products and MODTRAN were provided. Then we introduced the theory of performing standard error propagation, because a major goal is to be able to develop a way to estimate errors associated with the LST algorithm. The chapter ends with a compilation of data conversions that take place throughout the LST process.

In our work, only Landsat sensors 4 through 8 are relevant because the first three sensors did not acquire thermal imagery. The thermal instruments on board Landsat 4, 5, and 7 were designed with a single thermal band, while Landsat 8 was designed with two. In order to take advantage of the full thermal archive, we have chosen to develop the LST product using a single band method (which can also be used individually on Landsat 8's two thermal bands). The radiative transfer program known as MODTRAN will be used to perform atmospheric compensation, and various interpolation steps will be utilized to obtain atmospheric parameters on a per-pixel level. In terms of analyzing the accuracy of our algorithm, we have chosen to use a combination of buoy-measured water temperatures and the MODIS Sea Surface Temperature (SST) product as truth data. The specifics of our methodology are presented in the following chapter.

Before proceeding to the next chapter, we will reemphasize that this product will be the first global solution for obtaining land surface temperatures with the Landsat thermal archive. While other land surface temperature products exist using other sensors, they have a much coarser spatial resolution than that of Landsat, and they do not have Landsat's continuously-acquired thermal imagery dating back to 1982. A previous investigator, Cook, developed the methodology for the Landsat Surface Temperature product and validated it for North American regions using Landsat 5 [Cook, 2014]. Our current efforts will extend this baseline tool to a global product that can be applied to all Landsat sensors, which will also be accompanied by a quality map that will let the user know how trustworthy the LST values are at each pixel. When our work is supplemented with JPL's emissivity product, we will have an extremely useful tool for remote sensing and other scientific communities.

Chapter 4

Methodology

Chapter 3 was dedicated to the review of important background knowledge; namely, radiative transfer theory, historical information about the Landsat satellites, various land surface temperature retrieval and analysis methods, and an overview of reanalysis products, MODTRAN, and conversions. This chapter aims to specify the methods selected for each component of our overall process. First we will describe the chosen method of finding the atmospheric variables through MODTRAN. Then we provide the reader with an overview of the end-to-end process of estimating the land surface temperature of Landsat scenes on a per-pixel level. This is followed by sections detailing each step of the process, which includes making reanalysis data compatible with Landsat scenes, putting the reanalysis data in a MODTRAN friendly format, justifying interpolation steps, discussing error analysis methods, and reporting deliverables of the finished product.

4.1 Generating Radiative Transfer Parameters

In Section 3.2, we presented the governing equation where L_T, the radiance due to temperature, is the term we desire to solve for. L_u, L_d and τ (upwelled radiance, downwelled radiance, and transmission, respectively), are the three atmospheric parameters that must be determined through atmospheric compensation. MODTRAN is the radiative transfer program that we have chosen to use for this purpose, but unfortunately the parameters we need are not direct outputs. There is more than one way to manipulate the spectral outputs of MODTRAN to retrieve values for L_u, L_d and τ, but the method we present here maintains an optimal balance between accuracy and computation time. The details of the other approaches that were tested can be found in Appendix E of Cook's dissertation [Cook, 2014].

The chosen method for retrieving the atmospheric variables involves three separate MODTRAN runs. As a reminder, we will show again the governing equation (4.1), which requires knowledge of emissivity. In the long wave infrared portion of the spectrum, surface properties such as emissivity and temperature can be altered without impacting atmospheric characteristics. By modifying surface properties with a total of three MODTRAN runs, we are able to estimate the desired atmospheric parameters. For instance, if we let emissivity equal unity, then the governing equation reduces to Equation 4.2.

$$L_{obs} = (L_T\, \epsilon + (1 - \epsilon)\, L_d)\, \tau + L_u \tag{4.1}$$

$$L_{obs} = L_T\, \tau + L_u \tag{4.2}$$

MODTRAN requires a height in kilometers for the observing sensor, as well as a final height of what is being observed and its corresponding boundary/surface temperature. If we proceed through two MODTRAN runs where emissivity is unity and the surface temperature is altered between runs, we can use the Landsat response function to calculate the effective spectral radiance value for the outputs. This will give us a single L_{obs} and L_T value per run, which we can use to create a simple two point solution to Equation 4.2 (illustrated by Figure 4.1). As shown by the figure and indicated by Equation 4.2, the slope is equal to the transmission and the intercept is the upwelled radiance. The points in the figure are labeled by T_1 and T_2, which are the surface temperatures that directly correspond to the L_T values via Planck's equation (Equation 3.4 in Section 3.2).

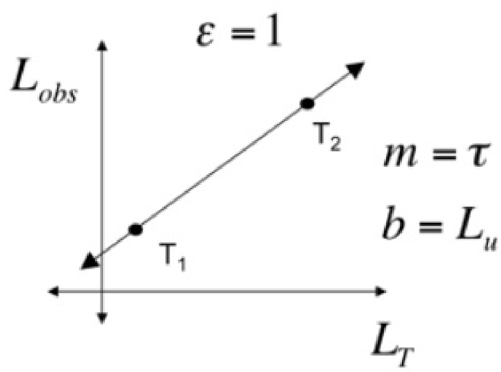

Figure 4.1: Illustration of the regression that can be made to obtain transmission and upwelled radiance [Cook, 2014].

The boundary temperatures used for the first two runs were chosen to be 273 K and 310 K, in order to span the range of surface temperatures that we expect the LST product to encounter. For the third MODTRAN run, we let emissivity equal 0.9 and we set the boundary temperature to "000" K, which uses the air temperature of the initial atmospheric layer. Since we have already obtained values for transmission and upwelled radiance, the governing equation can be rearranged to solve for downwelled radiance (Equation 4.3).

$$L_d = \frac{\frac{L_{obs} - L_u}{\tau} - L_T\ \epsilon}{1 - \epsilon} \tag{4.3}$$

After this third run, we now have values for the upwelled radiance, downwelled radiance, and transmission of the atmosphere. It may seem confusing that surface temperatures are defined by the user in MODTRAN when that is the very term being sought, so we will reiterate that we are using "hypothetical" values for emissivity and temperature because they allow us to estimate the atmospheric parameters without affecting their values.

4.2 Process Overview

The LST process was originally developed for use with the NARR reanalysis data, so much of the details discussed throughout the chapter will pertain to this dataset. We also use MERRA for our global validation efforts which requires modifying parts of the methodology. These differences will be explained where relevant.

After an appropriate reanalysis product has been selected to provide atmospheric profile information, it is important to ensure that this data is compatible with Landsat imagery as well as MODTRAN. Figure 4.2 shows an example of how NARR has a native coordinate system that differs from Landsat, leading to a nonlinear relationship between the two. MERRA, on the other hand, relates to Landsat imagery in a much more convenient manner. In either case, the number of reanalysis points used for MODTRAN runs needs to be limited to points within and just outside the Landsat scene. The atmospheric variables that are needed at each reanalysis point are air temperature, geopotential height, and specific humidity, but temporal interpolation must be applied so that the 3-hourly reanalysis data can be adjusted to the Landsat scene acquisition time. This step is illustrated by Figure 4.3.

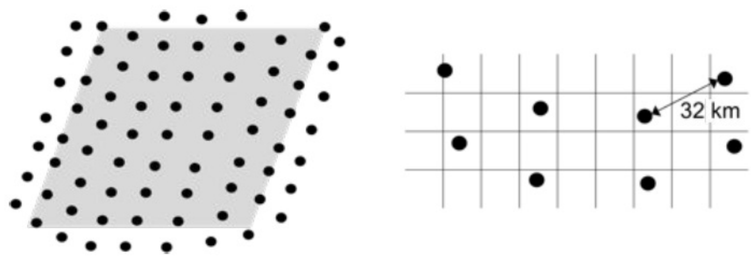

Figure 4.2: Left image shows a gray outline of a Landsat image (not to scale) with black dots representing NARR points within and just outside the scene. The right image shows a grid representing the Landsat UTM coordinate system compared to NARR point locations [Cook, 2014].

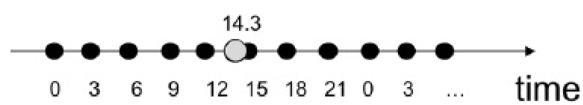

Figure 4.3: Example of temporal interpolation of reanalysis points to Landsat acquisition time. If the Landsat scene was acquired at 14.3Z, then the NARR points at the 12Z and 15Z collection times are interpolated to 14.3 [Cook, 2014].

After reanalysis profiles are obtained for grid points within and just outside a given Landsat scene and temporal interpolation has been applied, MODTRAN can be executed for several ground altitudes at each grid point in order to generate a three dimensional array of atmospheric parameters. This is visualized by Figure 4.4, where each small cube represents a reanalysis location at a specific height. With the aid of a Digital Elevation

(DEM) model for each Landsat scene, the atmospheric parameters can be interpolated to the correct height for each reanalysis point (see Figure 4.5). Then, a spatial interpolation method can be applied in order to generate parameters for every pixel in the Landsat scene (see Figure 4.6). For these three instances of interpolation, time is spent choosing the appropriate method to minimize errors and computation requirements, and we attempt to quantify the error that each step contributes to the LST process as a whole.

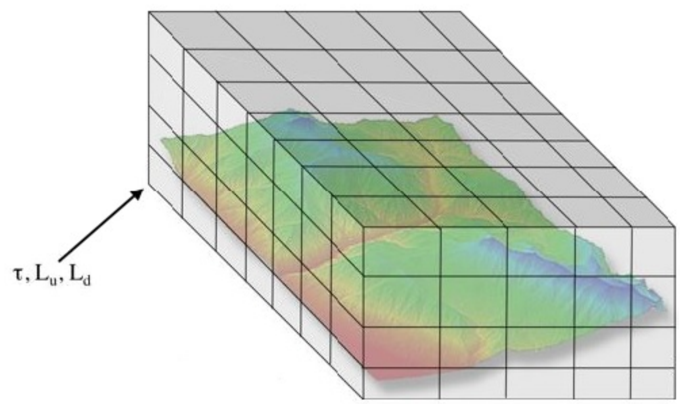

Figure 4.4: Illustration of the data cube that is generated by calculating atmospheric parameters at each reanalysis point for multiple altitudes. The data extends outside the scene itself so that spatial interpolation can be performed (Image from http://www.scisoft-gms.com).

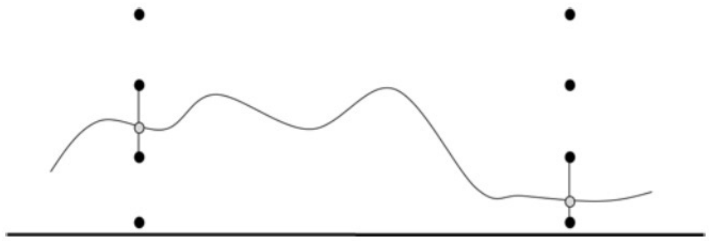

Figure 4.5: Depiction of height interpolation of reanalysis points to terrain elevation [Cook, 2014].

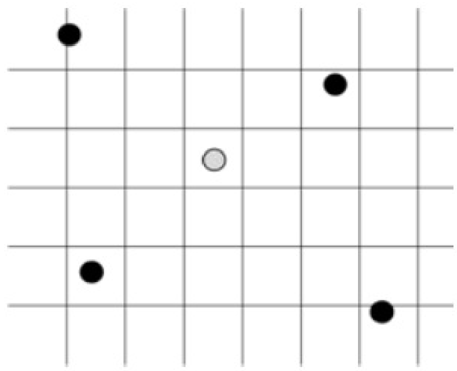

Figure 4.6: Spatial interpolation of reanalysis points to Landsat pixels [Cook, 2014].

4.3 Reanalysis Registration with Landsat

It is important to consider how the reanalysis product being used relates to Landsat imagery; particularly, we need to understand how their coordinate systems relate. Landsat images are provided in UTM (Universal Transverse Mercator) coordinates, which divides the world into 60 strips or "zones" that are each 6° wide. Zone 1 spans -180° to -174° longitude, and Zone 60 covers the area between 174° and 180°. A portion of the world map is shown in Figure 4.7, showing several zone borders. Locations in UTM coordinates are reported as an easting value in meters, a northing value in meters, and the corresponding

UTM zone. In order to avoid negative values, the center meridian of each zone is assigned an easting of 500,000 meters. In this respect, any location west of a particular zone's central meridian will have an easting less than 500,000 meters, and an easting greater than 500,000 meters if the location is east of the meridian. For locations in the northern hemisphere, the Equator is considered zero and the northing value increases in a northbound direction. If the area of interest is in the southern hemisphere, the Equator is assigned a northing value of 10,000,000 meters, which decreases in a southerly direction [USGS, 2001].

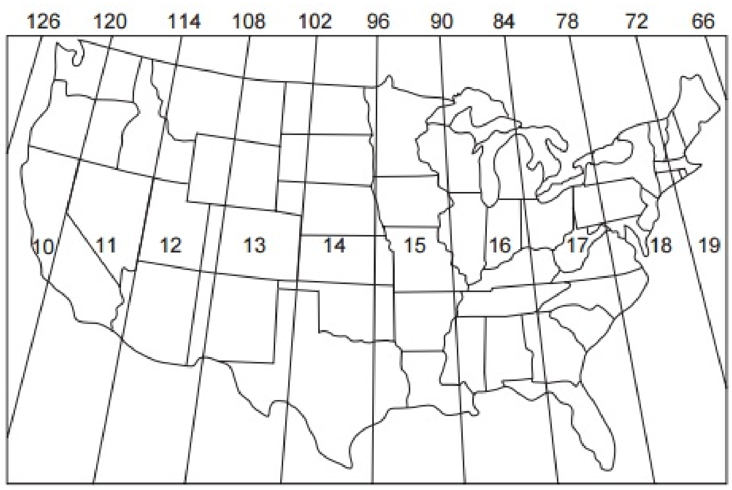

Figure 4.7: Depiction of the UTM grid over a portion of the world; zones 10 through 19 are shown, with corresponding longitude values located at the top of the image [USGS, 2001].

Reanalysis products are typically available for some large region, but it would be unreasonable to perform MODTRAN runs on every point in the grid. Instead, only the points in and immediately around each Landsat scene are needed. The point selection process will require knowledge of the reanalysis point locations in order to determine which ones should be included. We will describe this selection process for both NARR and MERRA reanalysis, which are the two products that are used in our LST method.

4.3.1 NARR

NARR reanalysis provides a 349 x 277 grid in the Lambert Conformal Conic coordinate system, and indexes each grid point with “i” and “j” integer values. Latitude and longitude coordinates for these points are made available through the U.S. Climate Prediction Center [NOMADS, 2015]. The goal is to subset the NARR data to the points that fall within and just outside a given Landsat scene, but first we must ensure that the coordinate systems match. Since we have the geographical coordinates for each NARR point, we can make use of the Landsat scene corner coordinates that can be found in the metadata file that is included in every scene download. With a simple greater or less than test, we can determine which NARR points are located within a certain scene. Because of the spatial interpolation step that will be described in Section 4.5, we also need to include NARR points that closely surround the Landsat scene. To accomplish this we simply adjust the scene corner coordinates enough to envelope at least one additional NARR point in each direction. Then, based on the maximum and minimum index values i and j of the included NARR points, we add any other points within the range of i and j values so that we form a complete rectangle of NARR points. This rectangle is only realized if we visualize the points in their native format, which is depicted in Figure 4.8.

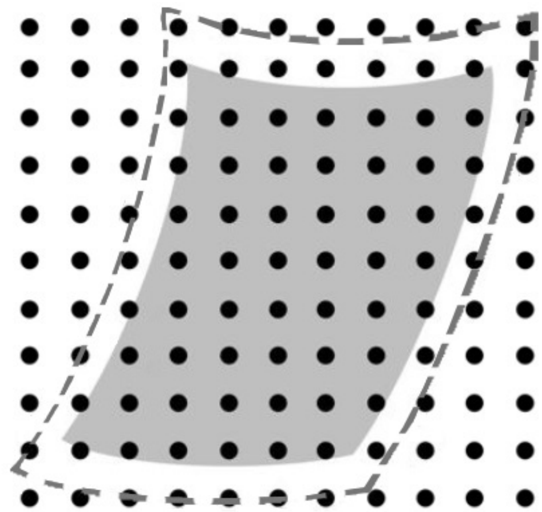

Figure 4.8: NARR points (black dots) overlaid on a Landsat scene (gray area) in NARR-native coordinates. The dotted line indicates the extended boundary made to include points outside the Landsat scene for a later interpolation step [Cook, 2014].

4.3.2 MERRA

MERRA reanalysis points are provided in a 288 x 144 grid at a 1.25° x 1.25° resolution. Latitude and longitude locations can be generated manually by marking latitudes every 1.25° from -90° to 90°, and marking longitude points every 1.25° from -180° to 180°. Geographic coordinates and the UTM coordinate system are closely related and are visually similar, and as long as the zone number is known the conversion between the two is a simple task. To select which MERRA points to use for a given Landsat scene, we can use the same method as before, where we once again extend the Landsat scene boundaries to include the rectangle of points just outside the scene. This is depicted by Figure 4.9 in order to highlight the contrast between using NARR and MERRA as reanalysis products.

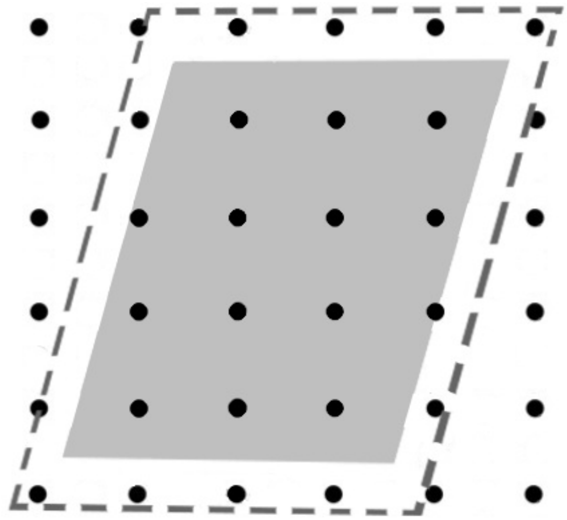

Figure 4.9: MERRA points (black dots) overlaid on a Landsat scene (gray area) in UTM coordinates. The dotted line indicates the extended boundary made to include points outside the Landsat scene for a later interpolation step.

4.4 Reanalysis Data to MODTRAN

Once the appropriate reanalysis points are chosen for a given Landsat scene, the MODTRAN runs to obtain the atmospheric parameters (as described in Section 4.1) can be executed for each point. MODTRAN has very strict formatting requirements for input data, so it is important that we ensure the reanalysis data is in a compatible form. The changes that must be made to the data obtained from NARR and MERRA are very similar, so we will refer to the reanalysis products in a general sense except for where differences occur.

For both NARR and MERRA, the three variables that are retrieved at each reanalysis point are air temperature, geopotential height, and specific humidity at a multitude of pressure levels. MODTRAN accepts the air temperature in its original form, but it requires geometric instead of geopotential height, and it only accepts certain humidity variables such as volume mixing ratio [ppmv], number density [molecules/cm3], and relative humidity [%].

Section 3.8 provides specific details on the conversion between geopotential an geometric height, and the conversion of specific humidity to relative humidity is also explained. This process is identical regardless of which reanalysis product is used, except that the number of spatial points to consider will differ as well as the number of pressure levels.

Both reanalysis products provide atmospheric variables at fixed pressure levels, which each correspond to some geometric height. Occasionally, at high pressure levels (low heights), the geopotential height conversion leads to a negative geometric height value. These instances are removed from the atmospheric profile, and only layers that have positive geometric height values are included.

MODTRAN requires atmospheric profile information to be present up to 100 km, which is much higher than any of the data obtained from NARR of MERRA. At these high altitudes where there is very little atmosphere, the impact on the values of upwelled radiance, downwelled radiance, and transmission are negligible. It is still necessary to provide information for these layers, so one of MODTRAN's standard atmospheres can be appended to the top of the reanalysis profile. We have chosen to use the mid-latitude summer profile, and linear interpolation was performed between the highest reanalysis layer and the second closest MODTRAN layer in order to create smooth transitions. We have provided Figures 4.10 through 4.15 to illustrate how the profiles look before and after interpolation, using an arbitrary NARR grid point location. Figure 4.10 shows the NARR pressure levels compared to MODTRAN, and 4.11 shows the resulting profile when the MODTRAN atmosphere profile was truncated, interpolated and appended to the NARR profile. Figures 4.12 and 4.13 show the identical process for the temperature profile, and Figure 4.14 and 4.15 contain the profiles for relative humidity. The shapes of the profiles below about 15 km are subject to change based on location and climate conditions, but little change will occur above that altitude because MODTRAN's mid-latitude summer profile takes over. Since an identical interpolation process is used for MERRA profiles, graphics are not included here. The only difference is that the MERRA profile will be dominant at higher altitudes because MERRA provides data at lower pressure levels up to 0.1 hPa.

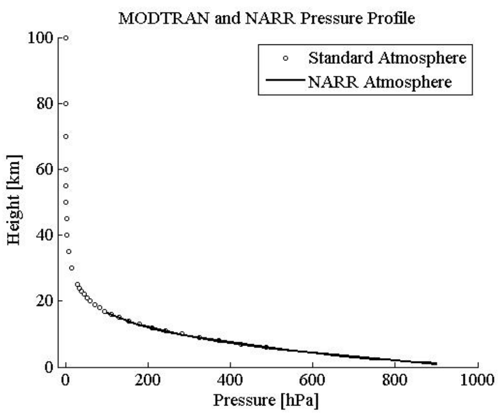

Figure 4.10: Plot of standard atmosphere and NARR pressure profiles.

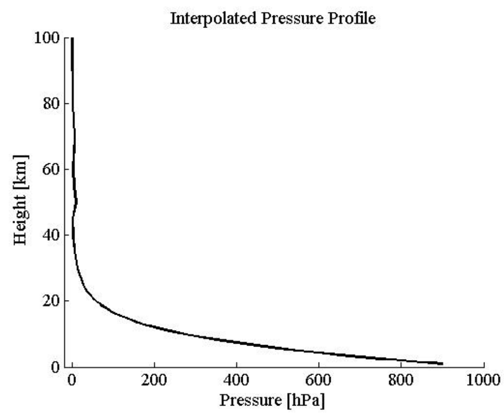

Figure 4.11: Plot of interpolated pressure profile for MODTRAN input.

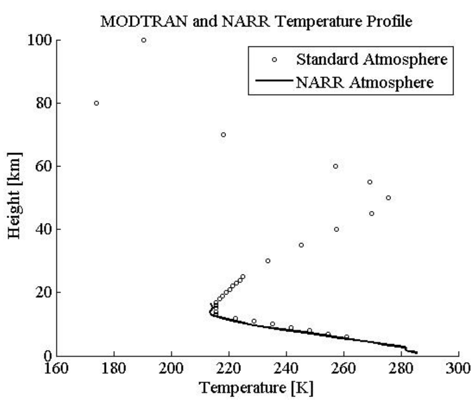

Figure 4.12: Plot of standard atmosphere and NARR temperature profiles.

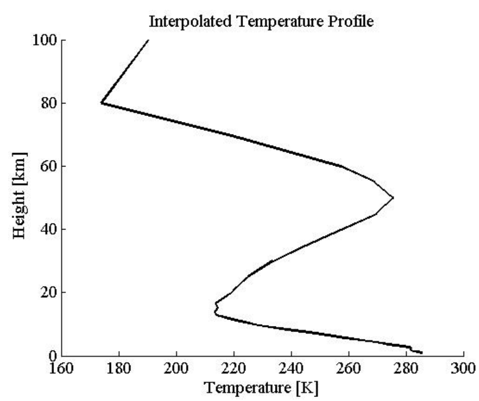

Figure 4.13: Plot of interpolated temperature profile for MODTRAN input.

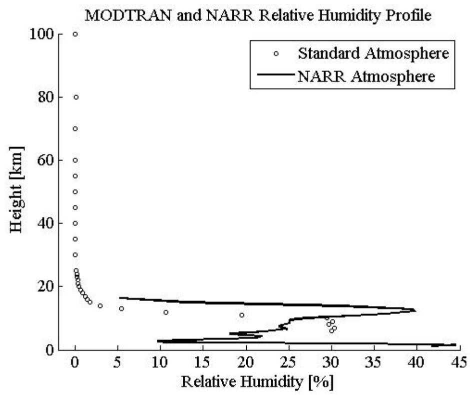

Figure 4.14: Plot of standard atmosphere and NARR humidity profiles.

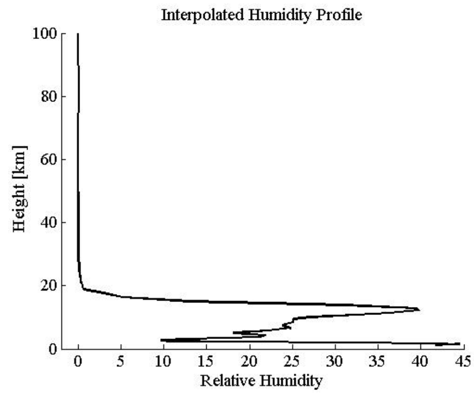

Figure 4.15: Plot of interpolated humidity profile for MODTRAN input.

4.5 Interpolation

In the overview of the LST process (Section 4.2), three occasions of interpolation were mentioned. The first was temporal interpolation of the reanalysis data to the Landsat acquisition time. The second instance of interpolation occurs after the atmospheric parameters are generated at several altitudes, but we want to estimate these values at the actual elevations within a given Landsat scene. The third and final form of interpolation takes place spatially in order to obtain atmospheric parameter values on a per-pixel level. In all these steps, we must determine which interpolation method is the most appropriate choice to minimize errors; options include linear, cubic spline, sinusoidal, or a nearest neighbor interpolator.

For each method that is chosen, we present an analysis of the errors that get introduced. All the investigations were done with NARR data and were carried out by Cook; the same studies were not performed with MERRA because once the process was chosen and validated using NARR we were able to justify it for MERRA by observing how the LST results changed when MERRA is used instead [Cook, 2014].

4.5.1 Temporal Interpolation

The analysis conducted by Dr. Cook in regards to temporal interpolation is summarized here but the full details can be found in her dissertation [Cook, 2014]. The main assumption made is that all pixels in a Landsat scene are captured at the same time; namely, the scene center scan time that is provided in every metadata file.

The structure of the NARR data and how certain variables change over time can be examined in order to find the best interpolation method. Two specific NARR point locations were chosen for the arbitrary date of August 2nd, 2007. The points used were at (42.809°N, 78.473°W) and (32.303°N, 115.453°W), northwest and southwest locations in the United States. The temperature and humidity variables at pressure levels 1000 hPa, 875 hPa, 750 hPa, and 550 hPa were plotted for each of the eight time samples throughout the day. These plots are not included here; instead we will note significant patterns. In general, the temperature values at higher pressure levels had higher values, and there was no obvious pattern as a function of time. When the same pressures were plotted for relative humidity, much more variation and nonuniformity was observed. Since the highest pressure levels have the most impact on the MODTRAN outputs and temporal pattern may become more evident, temperature and relative humidity were replotted for pressure levels 1000 hPa, 975 hPa, 950 hPa, 925 hPa, and 900 hPa. The range of temperatures seen across the different pressure levels decreased, but the variations throughout the day increased. Relative humidity once again varied with much less uniformity. Sinusoidal, cubic spline, and nearest neighbor interpolations were tested but none of them proved to be better fits across all pressure levels, even when the number of time samples was expanded. Therefore, as an initial method, piecewise linear interpolation was chosen for all three NARR variables.

The desire now is to get an initial look at the errors introduced by this form of temporal interpolation, which was done by performing several small studies. The general approach involves providing MODTRAN with a truth atmosphere, and using our method of generating atmospheric parameters we end up with $L_{obs(truth)}$, $\tau_{(truth)}$, $Ł_{d(truth)}$, $L_{u(truth)}$, and using Planck's equation we obtain T_{truth}. Then, if we execute a second run using an interpolated profile we can generate $\tau_{(NARR)}$, $Ł_{d(NARR)}$, and $L_{u(NARR)}$ using the observed radiance from the first run ($L_{obs(truth)}$). Planck's equation is used to determine T_{NARR}, which can then be compared to T_{truth} to quantify the error introduced by using the interpolated profile. This process is summarized in a visual form in Figure 4.16. At this point, we will review the various studies that are detailed in Cook's dissertation [Cook, 2014].

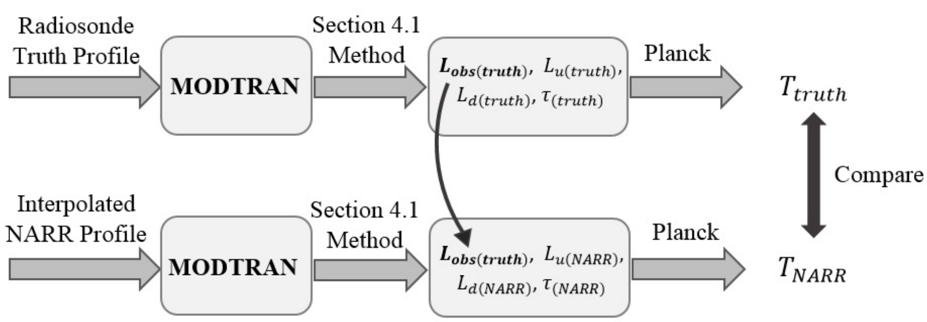

Figure 4.16: Depiction of how the error due to linear interpolation was determined.

First, the worst case scenario was considered in order to get an idea of the maximum errors that may be seen as a result of using the piecewise linear interpolation. Two NARR profiles at 9Z and 15Z on a particular day in August of 2007 were linearly interpolated to 12Z with intention of comparing the interpolated profile to the "true" profile that NARR provides at 12Z. This is considered the worst case scenario because the NARR profiles are each three hours away from the desired time; in normal operation with Landsat scenes, the longest time period between the scene acquisition time and the NARR profiles taken before and after is 1.5 hours. The interpolated profile was used with MODTRAN to reach predicted surface temperatures at several different altitudes. The same was done with the true NARR profile, and the difference between the "true" temperatures and the temperatures obtained through the interpolated profile were calculated. The errors varied as a function of height, but the error rarely exceeded magnitudes of 0.2 K which is encouraging.

The second study examined how changes in the atmosphere as time progresses affect the apparent temperature. This time, NARR profiles at 15Z and 18Z were interpolated to 16.5Z, and the apparent temperature for many altitudes was obtained through MODTRAN. This was compared to the MODTRAN results using the 15Z profile, and then to the results for the 18Z profile. Note that we are calculating the "error" between temperatures from two different times on purpose, in order to observe how much temperature changes over time and across a range of altitudes. When the temperatures obtained for the 16.5Z and 15Z profiles were plotted, the temperature increased by up to 3 K for altittudes less than 1 km. At altitudes roughly between 1 and 2 km, the temperature decreased to around -0.5 K, and at higher altitudes the temperature flattens out with slight fluctuates around 0 K. When the temperatures from the interpolated profile were compared to the 18Z profile, the patterns that were just described appeared in a reverse manner; that is, the shape of the curve was flipped so that the temperature differences such as the ones for altitudes less than 1 K was 3 K. This means that the temperature changes from 15Z and 16.Z were of similar magnitude to the changes seen between 16.5Z and 18Z, which indicates that the linear interpolation step was a reasonable choice.

Finally, a study was performed in order to explore any seasonal effects on the accuracy of temperature retrieval. For each month of the year (but not necessarily of the same year), the 15Z and 18Z NARR profiles were interpolated to 16.5Z, and compared to a radiosonde profile corrected to 16.5Z which can be obtained using a process described in F. Padula's thesis [Padula et al., 2003]. Rather than citing the results and patterns for each month, we will simply say that errors tend to be larger in the summer months, especially for instances where transmission is low or relative humidity is high. Since the uncertainty between NARR and the radiosonde profile is unknown and it is difficult to say whether all the errors seen can be attributed solely to interpolation, we are content with the fact that most of the errors were in the single digits (in Kelvin).

From these studies, we feel that linear interpolation of the NARR data in a temporal sense is a sufficient choice for our LST retrieval process. As a reminder, these tests are mostly to insure that a reasonable interpolator is being used. Overall errors in the LST algorithm will be assessed by comparing retrieved values to ground truth values.

4.5.2 Height Interpolation

The following studies can be found in full form in M. Cook's dissertion [Cook, 2014]. Terrain elevation affects the values of the NARR parameters used for computing the LST; for instance, upwelled and downwelled radiance generally decreases and transmission increases as the elevation increases. This is due to the fact that a larger volume of atmosphere is being compensated for at low elevations. It is not feasible to consider executing MODTRAN for every elevation in the image, so instead we have chosen to use a specific set of heights at each NARR location. Once the radiative transfer parameters are obtained at each height, they can be interpolated to the appropriate elevation of each pixel, which requires the use of a Digital Elevation Model (DEM). We must now determine the optimal heights at which MODTRAN should be executed.

Most pixels that will be encountered by the LST product will have an elevation between 0 and 2 km, but to ensure that we include most elevations found around the globe MODTRAN was executed at nine heights spaced uniformly from 0 to 4 km. Figure 4.17 shows an arbitrary atmospheric profile on the left with a zoomed in portion on the right.

The dashed lines in the right image indicates the elevations at which the MODTRAN runs were executed. The number of runs and the height values were chosen as an initial test, but we are now considering a change from uniform spacing to logarithmic spacing, where the number of MODTRAN runs are more dense at lower altitudes and more sparse as the altitude approaches 4 km. The appropriateness of this change will be investigated at a later date, but throughout this text we have utilized the uniform elevation spacing.

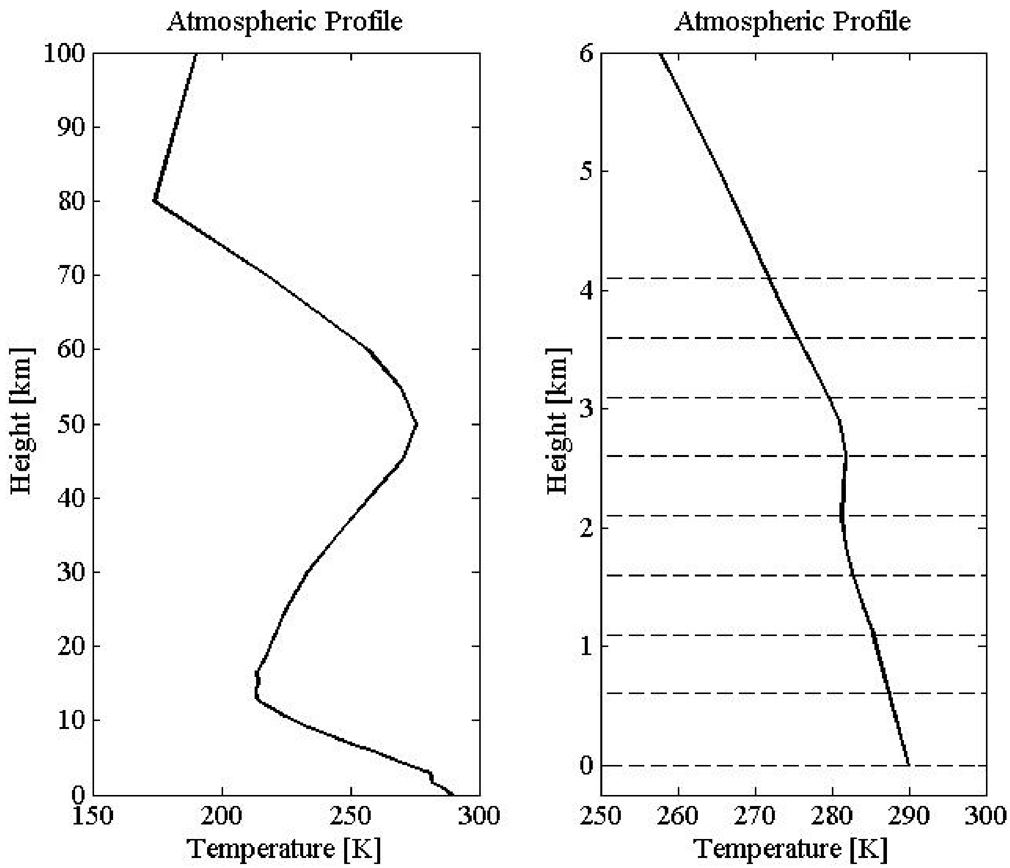

Figure 4.17: Left image: Example of an atmospheric profile. Right image: zoomed in look at the lowest elevations. The dashed lines mark the elevations at which radiative transfer parameters are currently generated [Cook, 2014].

In order for MODTRAN to generate outputs at specific heights, the atmospheric layers must be interpolated in a particular manner. The ground altitude that is input to MODTRAN needs to be identical to the geometric height of the lowest layer in the atmospheric profile, which signals to MODTRAN that the atmosphere being compensated for begins at that altitude. Interpolation is performed between the closest atmospheric layer above and below the ground altitude, which forms the layer at the desired ground altitude. Layers with geometric heights below the ground altitude selected for the current MODTRAN are eliminated.

This interpolation method was used to execute the nine MODTRAN runs consisting of heights ranging from 0 km and 4 km. When used for multiple NARR points, the pre-defined heights remain the same except for the lowest height, which is always set to the first layer of the NARR profile. At this point, the radiative transfer parameters that are generated at these nine heights must be interpolated to the elevation of each pixel based on a DEM of the Landsat scene. For now, we have implemented a simple piecewise linear interpolation that uses one point above and below the actual pixel elevation. To estimate the error introduced by this step, atmospheric parameters were generated at these nine heights, and then again for an increased resolution of 80 elevations over the same range. Using 80 elevations is very computationally taxing, and is only used for this particular study. The atmospheric parameters generated for the 80 elevations were used to create a set of truth temperatures using the same technique as usual from Section 4.1. Parameters for the nine elevations were linearly interpolated to obtain values at each of the eighty elevations, which are then converted into temperature values and compared to the "truth" data that was generated with MODTRAN.

This study was conducted for two dates in February and August, and the NARR point used was (42.809°N, 78.473°W). For both dates, the errors were higher at lower elevations, and the February results never exceeded a magnitude of 0.2 K. The August results had errors up to a magnitude of one, but this is very acceptable since summer scenes can be expected to have higher errors.

4.5.3 Spatial Interpolation

After the temporal interpolation of the NARR profiles and height interpolation of the atmospheric parameters to the correct scene elevations, we are left with the task of generating parameters on a per-pixel level. This is not an intuitive process, largely because NARR and Landsat have different native coordinate systems. Our chosen method calculates parameters for each pixel by using the values at the four surrounding NARR points. Since we have finally arrived at a point where operations are being performed for every pixel in a Landsat scene (approximately 56 million pixels total), it is important to design the method in a way that minimizes computation time. First, we will describe the way we select which NARR points to use in the interpolations to every pixel in the scene, and then we will discuss the analysis of the errors contributed by our spatial interpolation techniques. Once again, all these studies were conducted by M. Cook and can be found in her thesis, but here we present only summaries [Cook, 2014].

Pixel Iteration and Reanalysis Point Selection

For every pixel in a given Landsat scene, we will choose the four surrounding NARR points that make a square in Lambert Conformal coordinate system. As mentioned in the background section, NARR points are assigned Lambert Conformal integer values i and j that correspond to physical locations. The corresponding latitude and longitude coordinates for these points are easily accessible and it is a simple matter to convert them to UTM coordinates. In our automated process, all UTM coordinates are calculated relative to the zone for the particular Landsat scene; for example, the Easting is defined as zero meters at the left edge of a given zone.

The first attempted method of selecting NARR points involved calculating the distance from every pixel to every NARR point for the scene, and then choosing the four points that were found to be the closest. The non-linearity of the coordinate systems caused the performance of this method to be less than stellar, and the computation time was significant. A new, systematic method was developed to iterate through pixels from left to right and top to bottom, and the surrounding NARR points are selected so that they form a rectangular shape in their native coordinate system. Figure 4.18 shows a schematic of some NARR points forming "quad" areas as they appear in NARR's native coordinates, where NARR points are labeled with letters and numbers refer to different quad sections that Landsat pixels may potentially lie within.

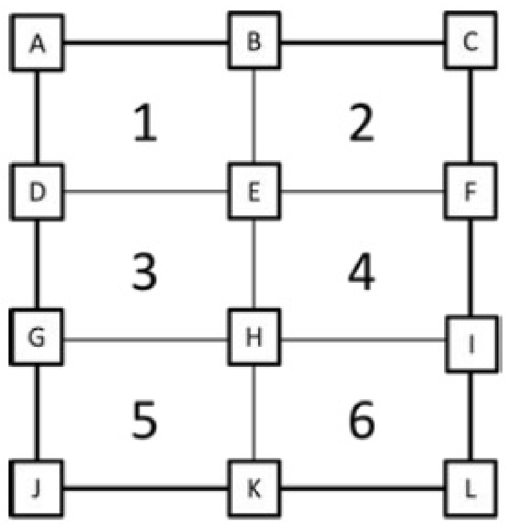

Figure 4.18: Schematic figure of NARR points and quad areas for pixel interpolation and NARR point selection. Quads are defined by the upper left NARR point [Cook, 2014].

Starting with the first pixel in the first row, the distance to every NARR point for the Landsat scene is calculated. The closest point above and left is identified, and the rest of the quad is defined using the Lambert Conformal grid notation. As an example, if point G in Figure 4.18 is the closest, pixels in quad three are interpolated from NARR points G, H, J, and K. Once we identify the quad for the first pixel in each row, six distance calculations are used to determine if the subsequent pixel falls within a new quad. Pixels can only shift to quads above or below the current one, or alternatively they can move to quads to their right. Compared to the initial method that was tested, this technique reduces the number of calculations by a factor of ten per pixel.

In order to determine if a pixel falls within a new quad, we must calculate which NARR point it is closest to. If we begin in quad 3, but the subsequent pixel is closer to point F than D, the pixel belongs to quad 4. Alternatively, if the pixel is closest to point A or J compared to D, the quad will move up or down, respectively. After all pixels have been iterated through and assigned to a quad of NARR points, we can move on to the interpolation of the parameters.

Interpolation of Radiative Transfer Parameters

Several spatial interpolation methods were examined, but we eventually settled on Shepards method. Simply put, Shephard's method applies weights to the NARR points based on their proximity to the point of interest, so that the final interpolation is influenced more by NARR points that are closer. Figure 4.19 illustrates the layout of how points are used in the method, where f_i are the values at coordinates (x_i,y_i), and F is the final interpolated value at (x,y). Equations 4.4, 4.5, and 4.6 show that mathematical expressions used to obtain F(x,y). In this set of equations, d_i is the distance that gets calculated from each NARR point to the pixel of interest, and w_i is the weight that gets applied to f_i where p is a weighting exponent with a default value of 2. Equation 4.6 provides the final interpolation value for the point of interest [Shepard, 1968].

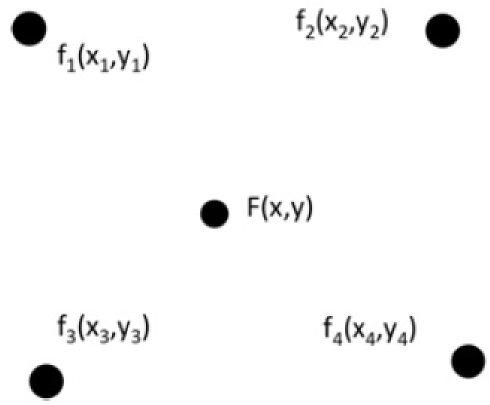

Figure 4.19: caption

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{4.4}$$

$$w_i = \frac{d_i^{-p}}{\Sigma_{j=1}^{n} d_j^{-p}} \tag{4.5}$$

$$F(x, y) = \sum_{i=1}^{n} w_i f_i \tag{4.6}$$

Yet again, we must attempt to isolate the error contributed by this particular interpolation step. As in the height interpolation studies, we can use a radiosonde profile (corrected to a surface weather station at 15Z) as "truth" data. MODTRAN and our method from Section 4.1 are used at the same nine heights given in Section 4.5.2. These are the "truth apparent ground temperatures. The location of the radiosonde weather station becomes the "pixel of interest," and the quad is identified using the technique described in the previous subsection. Using Shepard's method, the radiative transfer parameters are interpolated from NARR points (from the 15Z profile) to the radiosonde location at every height and the apparent ground temperatures are computed. The difference between these temperatures and the radiosonde-derived “truth temperatures reveals the error contributed by using Shepard’s method. The results for the same dates/locations from the height interpolation study were used; the error seen in the February case was within ± 0.1 K, and the errors for the August case reached as high as 1.5 K, but these results are reasonable enough to continue using this method of spatial interpolation.

4.6 Estimating LST Error

In Chapter 3, we introduced the general equations that would result from using standard error propagation on the governing equation. From Equation 3.8, we have created our own nomenclature to make this complex expression more intuitive. This simplified version is shown in Equation 4.7, and the definition of its parts can be found in Equation 4.8. With this new arrangement, S_A refers to the uncertainty in our estimation of the atmosphere (includes reanalysis product and atmospheric compensation uncertainties), S_I signifies the uncertainty in the Landsat radiance measurements, S_E is the error associated with the ASTER GED product, and S_P represents the cross correlation terms that are associated with S_A^2. These uncertainty terms will be calculated in radiance units, and then the value of S_{L_T} will be converted to units of Kelvin using a lookup table. At this final step, we will use the term S_{LST} to refer to the total estimated uncertainty in units of Kelvin. The next few sections will elaborate on how we aim to obtain values for each source of error/uncertainty. The actual implementation of this error analysis method is presented in Chapter 5, so we can determine how well we can predict what the uncertainty in the LST algorithm will be for any given pixel.

In the validation studies that were performed by Cook for Landsat 5, a trend between cloud proximity and LST error was observed [Cook, 2014]. Essentially, when there are clouds over or near a pixel, the LST retrieval tends to underestimate the actual surface temperature. We suspect that the standard error propagation method may not be sufficient for estimating LST uncertainties in such cases. In Section 5.7 of the Results chapter, we will investigate the LST error that is not accounted for by standard error propagation.

$$S_{L_T} = \sqrt{S_A^2 + S_I^2 + S_E^2 + S_P} \tag{4.7}$$

$$\begin{aligned}
S_A^2 &= \left(\frac{\delta L_T}{\delta \tau} S_\tau\right)^2 + \left(\frac{\delta L_T}{\delta L_u} S_{L_u}\right)^2 + \left(\frac{\delta L_T}{\delta L_d} S_{L_d}\right)^2 \\
S_I^2 &= \left(\frac{\delta L_T}{\delta L_{obs}} S_{L_{obs}}\right)^2 \\
S_E^2 &= \left(\frac{\delta L_T}{\delta \epsilon} S_\epsilon\right)^2 \\
S_P &= 2\rho_{\tau L_u} \frac{\delta L_T}{\delta \tau} \frac{\delta L_T}{\delta L_u} S_\tau S_{L_u} \\
&\quad + 2\rho_{\tau L_d} \frac{\delta L_T}{\delta \tau} \frac{\delta L_T}{\delta L_d} S_\tau S_{L_d} \\
&\quad + 2\rho_{L_u L_d} \frac{\delta L_T}{\delta L_u} \frac{\delta L_T}{\delta L_d} S_{L_u} S_{L_d}
\end{aligned} \tag{4.8}$$

4.6.1 Error due to the Atmosphere and cross correlation terms

We use the term S_A to refer to uncertainty that is associated with the use of reanalysis atmospheric profiles, as well as our method of performing atmospheric compensation. In Equation 4.8, we can see that our definition of S_A^2 is simply a gathering of the three terms in the original standard propagation equation (3.8) that directly relate to atmospheric parameters. Within this definition of S_A^2, there are partial derivative terms as well as S_τ, S_{L_u}, and S_{L_d}. The partial derivative terms are easily defined but have no special meaning, so their mathematical definitions are located in Appendix C. S_τ, S_{L_u}, and S_{L_d} are the uncertainties in our approximation of transmission, upwelled radiance, and downwelled radiance, respectively. These terms consist of more partial derivatives and uncertainty terms relating to the atmospheric profile information, and were originally defined in Equation 3.9.

Although we have a good definition of S_A^2, there is not an obvious way to calculate it. If we recall Section 3.4.5, where we discussed various error analysis methods used for LST algorithms, we can gain inspiration from the work of Hook and Hulley [Hook et al., 2007] [Hulley et al., 2012]. Their approach was to perturb the atmospheric profiles and observe the change from the "perfect scenario." Our implementation of this method involved choosing 11 sites around the world that represent four different climate types, then obtaining atmospheric profiles for several days at each location, and then perturbing the temperature and relative humidity profiles to measure the change in our calculation of transmission, upwelled radiance, and downwelled radiance. Notice that we have chosen not to consider perturbing the pressure and geometric height profiles, because the effect on τ, L_u, and L_d would be minimal. Therefore, we have simplified the definitions for S_τ, S_{L_u}, and S_{L_d} as shown in Equation 4.9. The terms S_T and S_{RH} are the uncertainties in the temperature and relative humidity profiles themselves, which are not easily defined but we have chosen preliminary values of 0.75 K and 2% based on MODIS atmospheric profile retrievals [Seemann et al., 2006]. The four climate types that were considered were "cold," "moderate," "hot and arid," and "hot and humid." The goal was to be able to determine if climatology would aid our efforts to relate atmospheric paramaters to the uncertainty in those parameters.

$$\begin{aligned} S_{\tau} &= \left[\left(\frac{\delta\tau}{\delta T} S_T \right)^2 + \left(\frac{\delta\tau}{\delta RH} S_{RH} \right)^2 \right]^{\frac{1}{2}} \\ S_{L_u} &= \left[\left(\frac{\delta L_u}{\delta T} S_T \right)^2 + \left(\frac{\delta L_u}{\delta RH} S_{RH} \right)^2 \right]^{\frac{1}{2}} \\ S_{L_d} &= \left[\left(\frac{\delta L_d}{\delta T} S_T \right)^2 + \left(\frac{\delta L_d}{\delta RH} S_{RH} \right)^2 \right]^{\frac{1}{2}} \end{aligned} \tag{4.9}$$

For a particular day and time at one of the 11 sites, we would get the temperature, relative humidity, pressure, and geometric height profiles from the MERRA database (we cannot use NARR for this particular study because we are using sites outside North America). Using the original profiles, we first generate the radiative transfer parameters using the method described in Section 4.1. Then, while keeping the relative humidity profile the same, we modify the temperature profile ± 5 Kelvin in 1 Kelvin increments and recalculate the parameters. This allows us to observe the changes in τ, L_u, and L_d as different amounts of "uncertainties" are introduced into the temperature profile. Similarly, we modify the relative humidity profile ± 30% in steps of 2%, while the temperature profile is kept constant. After plotting the change in the profiles versus the observed difference in the atmospheric parameters, we saw that the relationships were all very linear (with R^2 values around 0.95). Since they were acceptably linear, we set the slopes to be equal to the partial derivative terms in Equation 4.9. For example, we would define $\frac{\delta\tau}{\delta T}$ as the slope obtained from the plot of temperature profile change versus change in transmission. By repeating this process for many dates/times and different locations, we can generate many values for the partial derivative terms, and since S_T and S_{RH} are constants, we are now able to evaluate S_{τ}, S_{L_u}, and S_{L_d}. Taking a step back, we have the original parameters τ, L_u, and L_d for each case, and now we have a calculation of S_τ, S_{L_u}, and S_{L_d}, the uncertainty of these parameters. The original variables were plotted against their corresponding uncertainty terms, which are displayed in Figures 4.20, 4.21, and 4.22. We have applied a quadratic trendline to each of these figures, which fits very well to the data. This means that we can use these quadratic equations to predict S_τ, S_{L_u}, and S_{L_d} solely from the original transmission, upwelled radiance, and downwelled radiance values. Even more importantly, we can now calculate S_A^2 from Equation 4.8. For further information about the data used in this study as well as a more in depth description of how these terms are evaluated, see Appendix C.

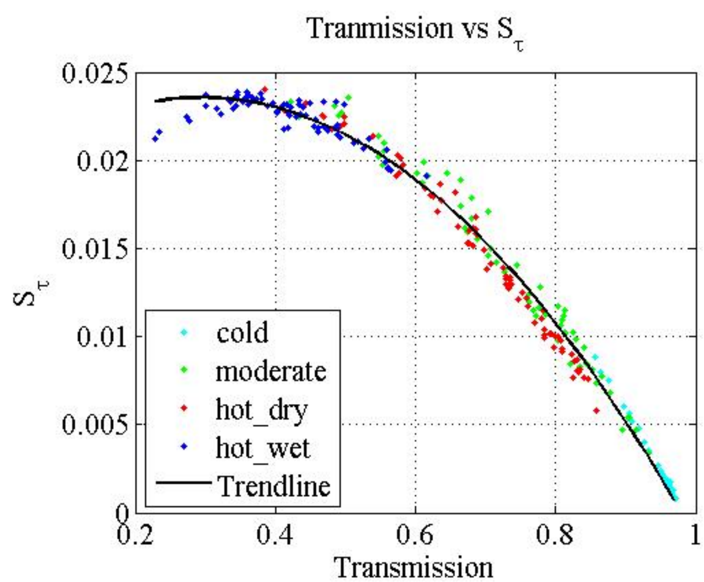

Figure 4.20: Plot of transmission versus the uncertainty in transmission, with a quadratic fit overlaid.

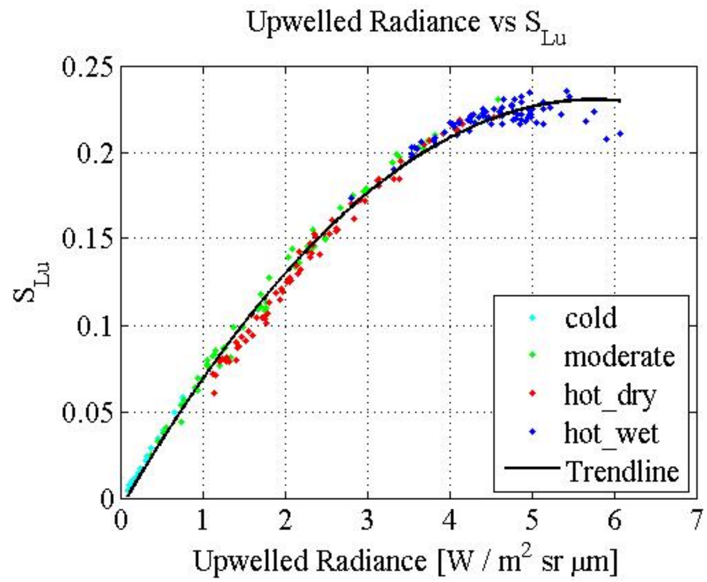

Figure 4.21: Plot of upwelled radiance versus the uncertainty in upwelled radiance, with a quadratic fit overlaid.

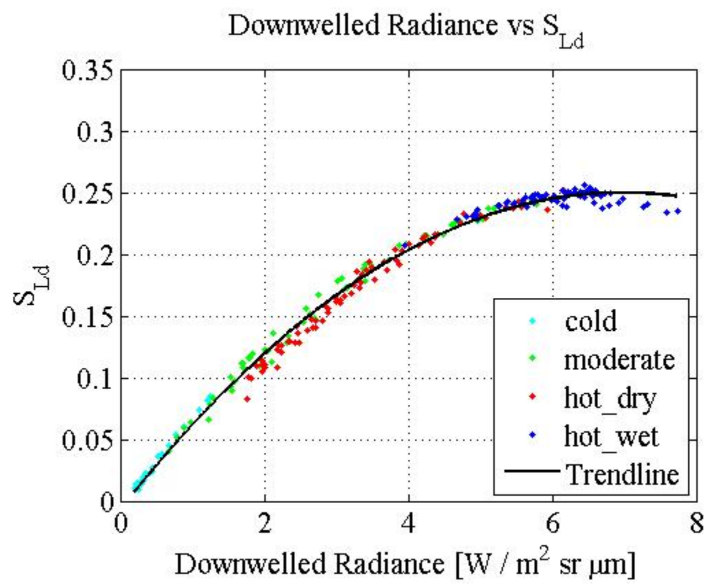

Figure 4.22: Plot of downwelled radiance versus the uncertainty in downwelled radiance, with a quadratic fit overlaid.

In Equation 4.7, we defined S_A^2 and S_P as the uncertainty due to the atmosphere and the associated cross correlation terms, respectively. They are defined separately to avoid confusion, because S_A^2 and other sources of error are expressed in terms of squared error, whereas S_P does not contain squared terms. Some of the variables contained in the definition S_P (see Equation 4.8) have already been discussed because they are part of S_A^2, namely; the terms S_τ, S_{L_u}, and S_{L_d}. The partial derivative terms are defined in Appendix C, and the variables $\rho_{\tau L_u}$, $\rho_{\tau L_d}$, and $\rho_{L_u L_d}$ signify the correlation coefficients between the variables identified in the subscripts. Since we have already revealed how the uncertainty in transmission, upwelled radiance, and downwelled radiance can be estimated, there is only the task to calculate the correlation coefficients. Since we already have τ, L_u, and L_d values from the atmosphere perturbation study, this last task becomes very simple.

4.6.2 Error due to the Instrument

A yardstick has markings to allow someone to measure a distance in inches, and even in smaller increments such as sixteenths of inches. If the distance between the markings was infinitesimally small, the user would be 100% certain that the measurement was correct. In reality, there is always some amount of uncertainty associated with measurements made by any instrument.

The Landsat satellites are far more advanced than a simple yardstick, but they still fall victim to the curse of uncertainty. These satellites have detectors aboard that translate electromagnetic radiation into a digital format, which then gets converted into radiance units. There are various sources of noise that are inherent in the detector as well as the electronics associated with it, which can all contribute to the overall uncertainty in the measurements made by Landsat.

In Equation 4.8, we have seen the mathematical form for the error introduced by the

Landsat instrument. The partial derivative is defined in Appendix C because it is a collection of variables that are known and equate to a constant, but the term $S_{L_{obs}}$ is the uncertainty in the observed radiance. This value can be found in the literature on the Landsat instruments, where it is often important to report specifications associated with various components of the imaging system. For example, the uncertainties for Landsat 5 and Landsat 7 and can be found in terms of noise-equivalent temperatures (NEΔT) in a publication by Barsi, Schott, and others [Barsi et al., 2005]. For our calculation of S_I^2, we choose the NEΔT value that corresponds to the Landsat sensor being used, and we convert this value to units of radiance.

4.6.3 Error due to Emissivity

In Section 4.1, assumptions were made about surface emissivity in order to approximate transmission, upwelled radiance, and downwelled radiance. In order to solve for the surface leaving radiance, L_T, we still need find an accurate source of surface emissivity that is available globally. For our validation efforts, we have historically only considered water pixels due to the availability of ground truth, and have therefore used the well known emissivity for water. In the official implementation of the algorithm, however, we intend to use a database developed by the Jet Propulsion Lab (JPL). This database is entitled Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Emissivity Database (ASTER GED), where ASTER is the satellite that is used to help derive surface emissivity. This database is available at *https://reverb.echo.nasa.gov/* and provides 1° x 1° maps at either 1 km or 100 m resolution [Hulley et al., 2015]. Since Landsat thermal bands have a resolution of 30 m, we would use the 100 m version of the ASTER GED, and we would also need to mosaic several esmissivity maps/images in order to cover the whole Landsat scene (see Figure 4.23).

Each image granule of the GED consists of several bands pertaining to surface emissivity;

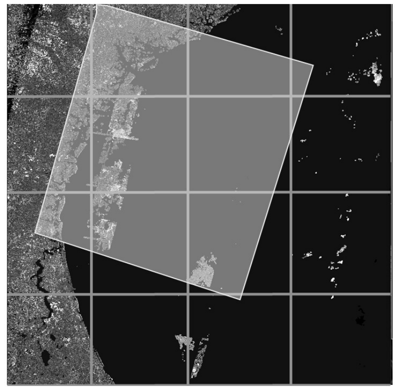

Figure 4.23: Illustration of several ASTER GED granules mosaicked to cover an entire Landsat scene, which is represented by the translucent gray rectangle.

we are particularly interested in using the mean emissivity band and the standard deviation band. The mean emissivity values will be used in the implementation of the LST algorithm, while the standard deviation will be used as the value for S_ϵ, the uncertainty in the ASTER GED product. This is half of our definition of S_E^2 in Equation 4.8, the other half being the partial derivative ($\frac{\delta L_T}{\delta \epsilon}$) that is defined in Appendix C. The developers of the ASTER GED have developed a method of quantifying the error associated with their emissivity retrievals, but it is currently not in a form that is publicly available. Until this tool is released, we will use the standard deviation as an estimate of the uncertainty in emissivity.

4.6.4 Error due to Cloud Proximity

There is another error-inducing factor that is not part of the standard error propagation; this factor is cloud proximity. When clouds are in a satellite image they obscure the Earth's surface underneath, which means the satellite observes the radiance being emitted from the clouds. Clouds are often a hindrance to the remote sensing community, and in this case they often cause the algorithm to underestimate surface temperature because clouds tend to be much cooler than the Earth's surface. There are also cases where the cloud may be warmer (possibly even very close to the Earth's surface temperature), and as a result the algorithm happens to be accurate. Either way, the LST algorithm is attempting to predict the Earth's surface temperature based off of radiance being emitted from clouds, and that is certainly not ideal. An easy solution to this would be to either not calculate LST for pixels that contain clouds, or to warn the user not to use these pixels for data analysis. Unfortunately, we have also observed that the presence of clouds not only affects LST retrievals for pixels that are directly under clouds, but they also affect pixels in the vicinity. We suspect that this indicates that the reanalysis products used tend to underestimate or fail to capture the presence of clouds and dense pockets of moisture. Since the quality of these reanalysis products are not well documented, it is more feasible to try to address this issue by relating the error seen in the LST retrieval to cloud proximity. The ultimate goal would be to create a function so that the LST error can be predicted based on how far away clouds are. We developed a metric to aid this pursuit, which is referred to as the "distance to nearest cloud" method.

The distance to nearest cloud method is a very simple process that calculates the distance from a point of interest to the nearest cloud pixel in a scene, which gets compared to the known LST error (requires a source of ground truth). If we make this comparison for a large number of points and for a variety of cloud conditions, we can determine whether a relationship between LST error and cloud proximity can be established. This method requires knowledge of where there are clouds in any given Landsat scene, so we intend to use the Landsat cloud mask product called "CFmask" that was developed at Boston University [Zhu and Woodcock, 2012]. These cloud masks are freely available for most archived Landsat scenes at *https://earthexplorer.usgs.gov/* as part of the Climate Data

Records (CDR) product. The cloud mask images contain single values that signify the presence of things in the scene such as water, land, cloud, cloud shadow, and snow. For our purposes, we only need to know which pixels contain a cloud and which do not, so we modify the cloud masks so that they contain only this information (See Table 4.1). Figure 4.24 shows an example of a modified cloud mask image that we would use for our distance to cloud method. Within that image, there would be a point where there is truth data available so we can determine the error in the LST algorithm, and we would also calculate the distance from that point to the nearest cloud pixel. Since our goals for validating the LST algorithm require comparing many LST predictions to some form of ground truth, we will plan to also calculate the distance to nearest cloud for each of these points and analyze the results. These results are located in Chapter 5.

In the beginning of this section, we presented the equation for the uncertainty in the LST retrievals, S_{L_T}, using standard error propagation (Equation 4.7). We also mentioned that there may be another contributor to this uncertainty that is not captured by the standard propagation terms, which has to do with cloud proximity (and also transmission, as we will see in the results chapter). If we use the distance to cloud approach to relate cloud proximity to LST error, we could aid our endeavor to estimate LST uncertainty. We will determine in Section 5.7 whether the standard error propagation method is sufficient or not for calculating uncertainty.

4.7 Deliverables

Besides surface temperature, atmospheric compensation is an important and useful piece of the overall product. Instead of just providing a single temperature value at each pixel, we plan to include all components necessary to calculate the LST given a known emissivity.

Table 4.1: Cloud mask pixel assignments from the CFmask, and the modified values used for the distance to nearest cloud calculations.

Pixel Type	**CFmask value**	**New Value**
Fill	255	156
Clear	0	0
Water	1	–
Shadow	2	–
Snow	3	–
Cloud	4	255

Currently, the LST process generates a five band geotiff image for each Landsat scene, at Landsat size and resolution. In Chapter 5, we will present our ability to estimate the uncertainty in the LST retrievals, which will introduce the need to also include a quality band. Table 4.2 lists all of the bands that can potentially be included in the final product. It may not be reasonable, however, to include all nine bands because it would make each file very large in size. USGS may decide to provide a subset of the bands that we have presented here.

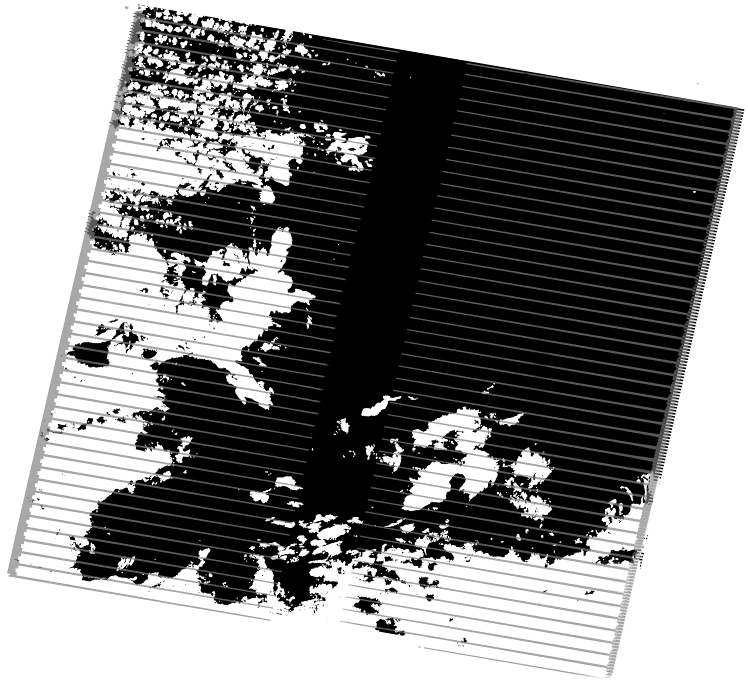

Figure 4.24: Example of a cloud mask for a Landsat scene after it has been modified so that it shows cloud pixels as white an no-cloud pixels as black. Landsat 7 images will include the scan line corrector gaps, which is shown in this example.

Table 4.2: Details on each layer of deliverables that can potentially be provided in the full LST product.

Band	Content	Units
1	Landsat thermal band radiance (L_{obs})	$Wm^{-2}sr^{-1}\mu m^{-1}$
2	Elevation	m
3	Transmission	Unitless
4	Upwelled radiance	$Wm^{-2}sr^{-1}\mu m^{-1}$
5	Downwelled radiance	$Wm^{-2}sr^{-1}\mu m^{-1}$
6	Emissivity	Unitless
7	Surface temperature	K
8	Quality/Uncertainty band	K
9	Distance to nearest cloud	km

4.8 Concluding Remarks

This chapter was designed to reveal the different components that make up the automated process of LST estimation. This started with the technique we chose to utilize to obtain atmospheric parameters through specifically designed MODTRAN runs. We then discussed the compatibility issues and solutions between the reanalysis data and both Landsat and MODTRAN. A description of the temporal, height, and spatial interpolations used throughout the process was provided, along with several studies to quantify the potential errors introduced by each interpolation method. Finally, we presented a more thorough look at how we plan to estimate the error associated with the LST algorithm. Although more work could be done to fully optimize the various parts of the process, the methods used have introduced adequately low errors in the temperatures predicted through the LST code. With this thoroughly described methodology, we can now move on to Chapter 5, where we present the bulk of validation results for various Landsat sensors and for different regions of the world, as well as an analysis of our ability to be able to estimate LST error.

Chapter 5

Results

This chapter contains our validation results, as well as a thorough discussion of their physical meanings. The first section of this chapter is designed to prepare the reader for various terminologies and notations that are used throughout the validation studies. Section 5.2 will lead with a summary of the Landsat 5 validation that was conducted by Cook, which also includes an initial attempt at observing the relationship between cloud proximity and LST error [Cook, 2014]. This effort only includes scenes in North America, using buoy temperature measurements as a source of truth data (details on the buoys used is located in Section 5.1.2). In Section 5.3, the distance to nearest cloud method is implemented for the first time based on a set of Cook's validation data. This will reveal the potential for this method to help estimate the uncertainty in the LST algorithm. Section 5.4 highlights the studies that were conducted in order to prove that global validation was possible, particularly for Landsat 7. This is followed by the actual global validation results in Section 5.5, which explores the effect of climate types on LST error and how best to quantify that effect. Section 5.6 contains the analysis of the LST uncertainty estimation method that was described in the Methodology chapter, and Section 5.8 steps through a visual example of what the final LST product will look like.

5.1 Overview of Validation Sets

A few clarifications should be made to ensure the results shown in this chapter are as straightforward as possible. First, we will explain how WRS-2 notation is used to describe where each Landsat image is captured, because we will often refer to specific sites using this notation. Secondly, we will discuss the process of calculating water surface temperatures from submerged buoy instruments, which is how we obtain truth data. Finally, we will review how the error in our LST process is calculated so that the results throughout this chapter are easily understood.

5.1.1 WRS-2 Notation for Landsat Scenes

The World Reference Stystem 2 (WRS-2) notation, which is used throughout our results, allows us to identify specific sites that Landsat passes over. A "path" integer value is assigned to different parts of the Landsat orbital track, where path 1 crosses the equator at a specific latitude and increases as the track moves from east to west. The total number of paths used in WRS-2 is 233, and some of these paths are depicted in Figure 5.1. An additional "row" value is used to describe where a particular Landsat scene was acquired along one of the paths. Row values range from 1 to 248, and they each correspond to a specific latitude that runs through the center of the Landsat frame. The combinations of path and row values represent every unique scene center, which is a convenient way to reference specific locations [Irons and Rocchio, 2015c].

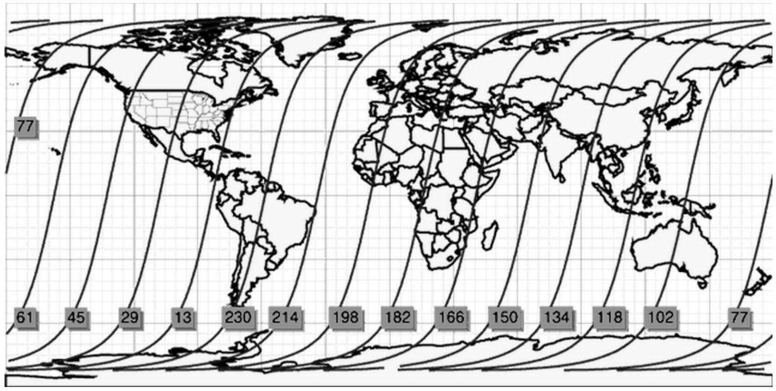

Figure 5.1: Illustration of a set of WRS-2 paths on a map [Aber, 2013].

5.1.2 Buoy Truth Measurements

The National Data Buoy Center (NDBC) operates and maintains a vast collection of buoys that record various atmospheric and oceanic measurements for a variety of purposes. Excluding international partners, these buoys are deployed in North American coastal regions. Sea temperature is a common measurement that these buoys take, but the instruments are submerged at some depth under the water's surface [NDBC, 2009]. Since the Landsat satellites are collecting information from the very top few milimeters of the water's surface, we need to correct the buoy-measured bulk temperature to a "skin" temperature. Certain meteorological data and information about the buoy's watch radius and sensor depth are required to perform this correction, which will determine which buoys can be used. The expected error for this correction was found to be 0.35 K, which is largely due to uncertainties in the thermistor that measures the water bulk temperature [Schott et al., 2012] and [Padula and Schott, 2010].

Aside from the NDBC buoys, there are two sites that are used in our validation studies that contain buoys and platforms maintained by JPL. These buoys are equipped with measurement tools similar to NDBC buoys, and an alternative method is used to correct measured bulk temperatures to skin temperature [Hook et al., 2004] and [Hook et al., 2007].

Truth data at these sites (Lake Tahoe and the Salton Sea) were provided to us by JPL. These sites are important to include in our validation results because we wish to cover a variety of locations, elevations, and climate types, and the buoy instruments are well calibrated and maintained.

As a final comment, when the term "buoy truth" or "truth data" is used throughout this text, we are referring to the skin-corrected temperature. Even though this correction introduces some error, it is acceptably small enough that we consider the calculated skin temperature to be the "truth," which is what we use to compare with our LST predictions.

5.1.3 Calculating LST Error

Throughout this chapter we refer to "LST error," which is simply the LST algorithm retrieval minus the truth temperature. The truth temperatures that we use are either derived from buoy data, or the MODIS Sea Surface Temperature product (the latter is discussed in Section 5.4.2). Equation 5.1 shows the simple calculation of LST error, which is used consistently in all our validation studies. As an example, if the error is negative, this indicates that our process underestimated the surface temperature. Another declaration we wish to make is that for all the validation results, the known emissivity of water was used rather than the ASTER Global Emissivity Database because only water pixels were used. Also, since we will be discussing mean errors throughout this chapter as a way of assessing the LST performance, we would like to inform the user that our goal for the finished product is to be accurate within $\pm$ 2 Kelvin.

$$\text{error} = \text{Predicted LST - Ground Truth Temperature} \tag{5.1}$$

5.2 Validation of Landsat 5

The reanalysis product that was initially selected for the LST process was the North American Regional Reanalysis (NARR), which limited validation studies to North American locations. Validation for Landsat 5 was performed by Dr. Monica Cook, who selected a set of Landsat scenes in North America that each contained one or more buoys. Then she simply applied the LST algorithm to those scenes and obtain a predicted surface temperature at the location of the buoy. A simple comparison between the predicted and truth temperature reveals the error in the LST process, which is computed for every scene. This validation study begins by only considering scenes where there are no clouds near the buoy, then it goes on to include a large set of scenes where there are clouds near or over the buoy.

5.2.1 Cloud Free Scenes

This validation set consists of 259 Landsat 5 scenes that were visually declared to be "cloud free," meaning that it appeared that any clouds present were far enough away from the buoy that they would not be likely to affect the LST prediction. There are nine different sites that are represented by the 259 scenes; Salton Sea, Tahoe Lake, Lake Ontario, Delaware Bay, the coast off of Georgia, the coast off Santa Maria, the coast off Santa Monica, Lake Huron, and Lake Superior. These locations have unique "path" and "row" values assigned to using the Worldwide Reference System 2 (WRS-2) notation [Irons and Rocchio, 2015c]. The specific path/rows for the mentioned locations can be found in Table 5.1, and Figure 5.2 shows these locations on a map. Error histograms for each site were created in order to show the number of scenes that fell within different ranges of errors (e.g. -1.5 to -0.5, -0.5 to 0.5, 0.5 to 1.5, etc). To be clear, this is the error in the LST process as determined by Equation 5.1. The shape of these histograms were similar between the different sites, so we will simply show the histogram for the whole set of cloud free scenes (see Figure 5.3). Ideally all the scenes would fall within the bin centered around zero Kelvin, but due to the variability present in several components of the LST process we should expect to see some spread. This type of graphic will remain consistent throughout our validation studies. It is also advantageous to review the statistics for these histograms so we can easily quantify the accuracy of the LST predictions. The statistics that correspond to Figure 5.3 are located in Table 5.1.

Figure 5.2: A map highlighting the WRS-2 locations that were used for the validation of Landsat 5. Multiple scenes were used from each site [Cook, 2014].

Table 5.1 shows that some sites consist of as few as 11 scenes, while others have as many as 89. This is largely based on the availability of cloud free scenes, which varies based on location. The absence of clouds should lead to optimal performance of the LST process, so it is doubtful that the statistics would change much even if more scenes were processed. The mean errors for each site are all within ± 1 K which is within the range of acceptability,

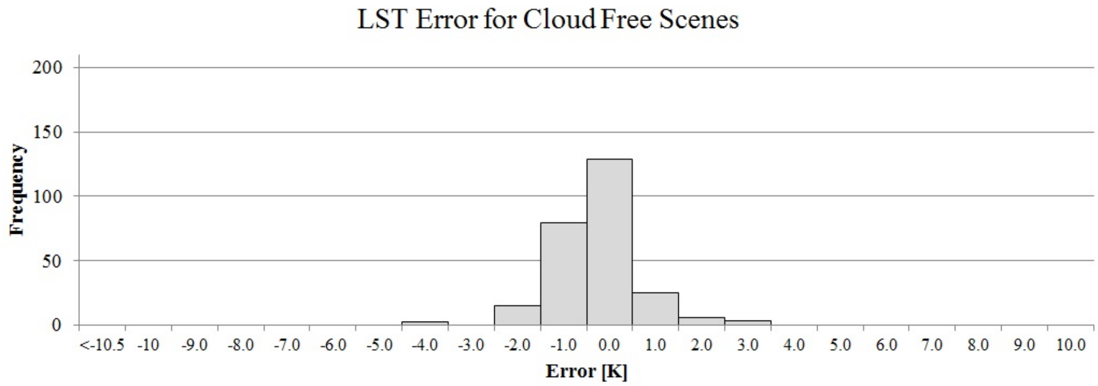

Figure 5.3: Error histogram of the LST process when only clear scenes are considered. The number of scenes for this case is 259 [Cook, 2014].

but we will also point out that the average error of -0.267 is statistically significant. A thorough investigation into the cause of this consistent underestimation can be found in Cook's dissertation [Cook, 2014], but the concluding thought is that NARR tends to very slightly underestimate the column water vapor and thus causes the negative bias. The root mean square error (RMSE) is also reported, because it is insensitive to whether the errors are positive or negative. For instance, if there were large errors that were large in magnitude but opposite in sign, the mean error might be reported as close to zero but the RMSE would be large, indicating that large errors or outliers are present.

5.2.2 Including Cloud Scenes

One of the goals of the Land Surface Temperature project is to be able to produce a confidence metric or quality map to inform users how trustworthy the reported LST is for each pixel in a scene. Also, we want the product to produce results for as many cloud-free pixels as possible. Cook attempted to determine the overall magnitude of error in the LST

Table 5.1: Initial validation results for individual sites, using cloud free scenes [Cook, 2014].

Location	**Path_Row**	**Mean Error [K]**	**Standard Deviation [K]**	**RMSE [K]**	**No. of Scenes**
Salton Sea	39_37	-0.12	0.558	0.545	11
Lake Tahoe	43_33	-0.213	0.713	0.740	89
Lake Ontario	16_30 17_30	-0.068	0.639	0.626	89
Delaware Bay	13_33 14_33	-0.447	1.179	1.245	34
Georgia (coast)	16_38	0.041	1.267	1.239	23
Santa Maria (coast)	43_36	-0.219	0.789	0.799	19
Santa Monica (coast)	41_37	-0.574	1.089	1.208	21
Lake Huron	20_29	-0.695	0.820	1.059	19
Lake Superior	24_27	-0.167	0.676	0.682	23
Total	—	**-0.267**	**0.900**	**0.905**	**259**

process by propagating errors through the input atmospheric profiles for cloud free scenes. The input temperature and relative humidity were expected to influence error the most, so she simulated cases where these input profiles were altered in order to come up with predictions of the error in the LST process. In the end, she was able create a scatter plot of predicted errors to actual errors in Kelvin, which did not show a significant correlation. This indicated that using this error propagation method would not be helpful in estimating the error that the LST values would have, and therefore not very useful for developing a quality map.

At this point there was a desire to validate the LST process for more than just cloud free scenes, and there was hope that the type and proximity of clouds would lead to a confidence metric. This spurred a large study using Landsat scenes at the same sites mentioned in the previous section, but now including both cloud free and cloudy scenes. The same validation process was utilized where buoys in the scene were used to provide truth temperatures, and by sorting the scenes by cloud type and proximity we can observe the errors seen for each group. Table 5.2 lists the six categories that were created by Dr. Cook, and it provides a short description for each. There are three main types of clouds; cumulus, stratus, and cirrus [NOAA, 2015]. Cumulus clouds are typically made up of individual and dome-like elements, stratus clouds appear as grey, flat layers that vary in thickness. Cirrus clouds are generally found at high altitudes and take the form of white, delicate wisps. Prefixes such as "alto" and "cirrus" are used to indicate the height of the clouds, but for this study texture was more important. Therefore, all cloud types that are variations of cumulus were simply labeled "cumulus," while all types of stratus and cirrus clouds were labeled as cirrus.

The cloud categories are defined based on whether there are clouds over the buoy, in the vicinity of the buoy, or not near the buoy at all. This categorization was done visually, but on average scenes were assigned a category of 3, 4, or 5 if there was a cloud within 0.5 km (17 pixels) of the buoy. For categories 1 and 2, clouds had to be present between roughly 0.5 and 5 km away from the buoy, and for category 0 all the clouds in the scene were at least 5 km (167 pixels) away from the buoy.

Now that the category types have been defined, we can begin to present the results for this study. A total of 826 scenes were used, which includes the 259 cloud free scenes from the previous section. Figure 5.4 shows the error histogram for all 826 scenes, which

Table 5.2: A description of each category used to sort scenes based on cloud type and proximity. Also includes the number of scenes belonging to each category [Cook, 2014].

Category	Description	No. of Scenes	% of Scenes
0	Cloud free	259	31.3%
1	Cumulus in vicinity of buoy	98	11.9%
2	Stratus or cirrus in vicinity of buoy	158	19.1%
3	Cumulus over buoy	60	7.3%
4	Stratus or cirrus over buoy	202	24.4%
5	Completely cloud covered image	50	6.0%

includes every cloud type. The center bins have a similar shape to the cloud free set, which is encouraging, but there are many scenes in the negative error bins which indicates that the LST underestimated the temperature. The bin on the very left side of the graph shows that there were about 150 scenes that had errors of -10.5 Kelvin or less, which is highly significant. It is suspected that the scenes in these negative bins represent cases where there are clouds directly over the buoy. Most clouds are colder than the surface of the Earth (especially when they are at high altitudes), so it is very feasible that the LST process would predict a low temperature for cloud pixels. An easy way to verify this theory is to remove the scenes that had clouds over the buoys (categories 3, 4, and 5), and observe the changes in the error histogram. This is depicted in Figure 5.5, which only has a few scenes in the left bin and therefore confirms the theory. There are still several scenes in the other negative error bins, which is most likely due to the cases where the clouds are close enough to the buoy to affect the LST prediction. If we were to also remove categories 1 and 2, the histogram would simply be that of Figure 5.3. The statistics for the different category combinations are located in Table 5.3.

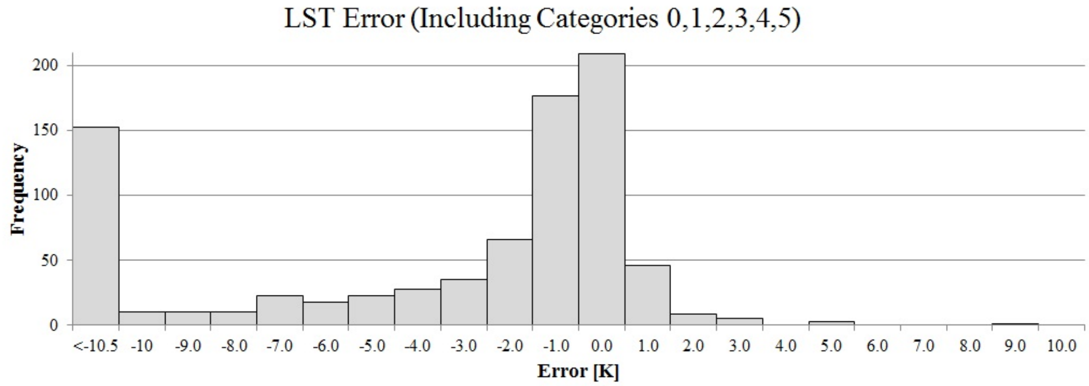

Figure 5.4: Error histogram of the LST process when all scenes are considered, regardless of cloud category. The number of scenes for this case is 826 [Cook, 2014].

The first row in Table 5.3 shows that the mean error is -8.471 K and the standard deviation is 19.313 K when all the cloud categories are used. When categories 3, 4, and 5 are removed the mean error drops to -1.538 K and the standard deviation drops to 4.174 K. This shows that significant error is introduced when there are clouds directly over the buoy, which was also indicated by Figure 5.5. When only category 0 scenes were used, the statistics matched the initial study from Section 5.2.1 because they use the same selection of scenes.

It was mentioned earlier that one of the goals of this cloud analysis was to assess the effect of cloud type on the LST error. Based on the statistics shown in Table 5.3 as well as Figure 5.6, there is no obvious correlation between cloud type and LST error other than the fact that removing scenes that have clouds near/over the buoys improves the average error. Figure 5.6 does show that by excluding category 2 scenes many of the negative error bin heights go to zero, but the heights of the center bins also decrease. Considering the vast improvement in error when the scenes with clouds over the buoys were removed, it is safe

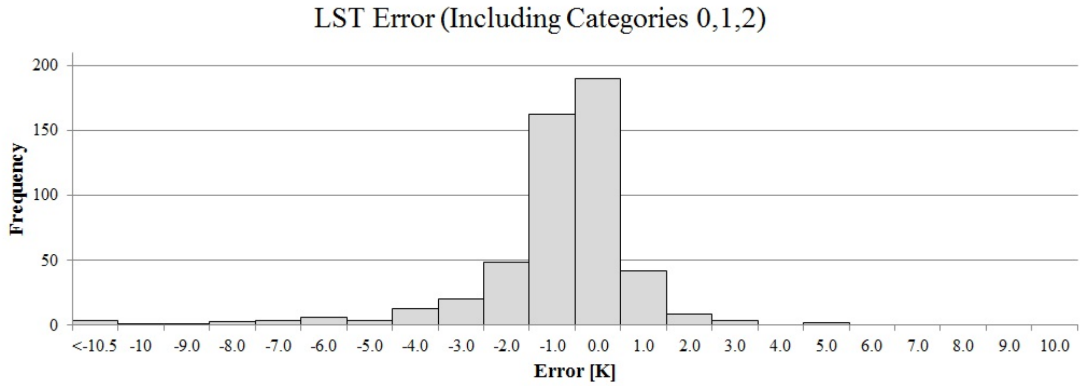

Figure 5.5: Error histogram of the LST process for clear scenes or scenes where clouds are near but not over the buoy (categories 0, 1, and 2). The number of scenes for this case is 515.

to say that cloud proximity influences error much more than cloud type. This conclusion plus the desire to automate this process led us to explore a new method of cloud analysis.

5.3 First Implementation of the Distance to Nearest Cloud Method

It has been demonstrated that there is a direct relationship between the proximity of clouds and the error seen in the LST process; however, we wish to observe this trend in an exact manner that eliminates the subjective sorting of a human user. This is where the "distance to nearest cloud" method comes in, a concept which was first introduced in Section 4.6.4. Using cloud masks that are available as part of the Landsat Surface Reflectance product, we simply calculate the distance from a point of interest (e.g. where there is truth data) to the nearest cloud pixel. Further details on this process are located

Table 5.3: Summary of statistics for different cloud categories within the dataset of 826 scenes.

Cloud Category	Mean Error [K]	Standard Deviation [K]	RMSE [K]	No. of Scenes	% of Scenes
0,1,2,3,4,5	-8.471	19.313	21.078	826	100%
0,1,2,3	-1.538	4.174	4.445	575	70%
0,1,2	-0.933	2.460	2.629	515	62%
0,1	-0.499	2.228	2.281	357	43%
0	-0.267	0.900	0.927	259	31%

in Appendix C.

5.3.1 Initial Implementation Results

This cloud analysis study is made up of 949 scenes, where 826 are Landsat 5 scenes from Cook's initial cloud investigation, and 122 are from a Landsat 7 data set that was used in a calibration study [Cook, 2014]. These scenes were chosen because the buoy truth data and the LST predictions were already available. The distance to nearest cloud metric was calculated for each scene, and by removing scenes where this number is below a certain threshold we can observe the changes in the LST error histogram. Figure 5.7 compares the LST error when all scenes are used to the case where "0 distances" are removed. A distance of zero simply means that a cloud pixel was found to be directly over the buoy. It is clear that by removing these scenes the error histogram improves greatly, but it is not ideal because we are still including scenes that have clouds just a few pixels away.

By setting a distance-to-cloud threshold, one can observe the gradual improvement of the LST error. Table 5.4 contains the statistics for Figure 5.7 as well as for different distance

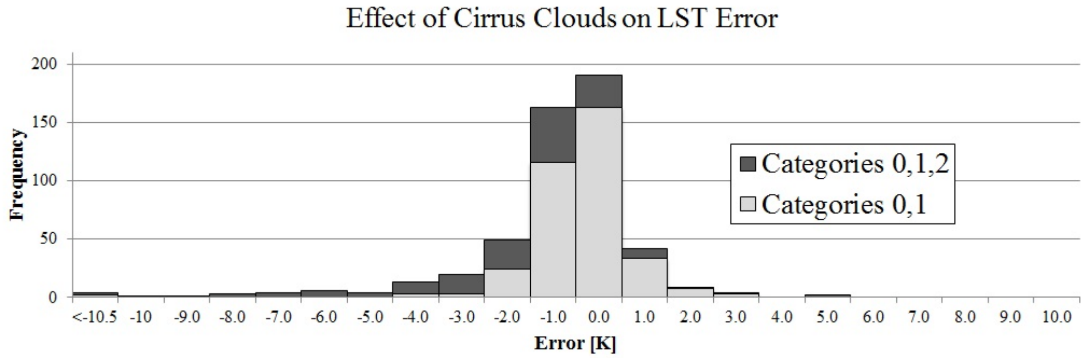

Figure 5.6: The error histogram for categories 0,1, and 2 is in dark grey and includes cirrus clouds. In light grey is the error histogram for categories 0 and 1, which only includes clear and cumulus scenes.

thresholds which are not depicted in histogram form, but to provide visual clarity we have plotted these statistics in Figure 5.8. The vertical axis on the left is used to plot the mean error and standard deviation for each distance threshold listed in Table 5.4, and the right vertical axis is used to plot the percentage of scenes being used for each threshold level. In general, we see that the initial removal of the "0 distance" cases shows the most improvement, although the percentage of scenes used drops significantly as well. In general, it appears that the percentage of scenes drops at a faster rate than the mean error and standard deviation improve. The threshold directly affects the number of scenes that can be used, so there is a trade off between the two that could help users decide which scenes to accept based on the average error they are willing to tolerate.

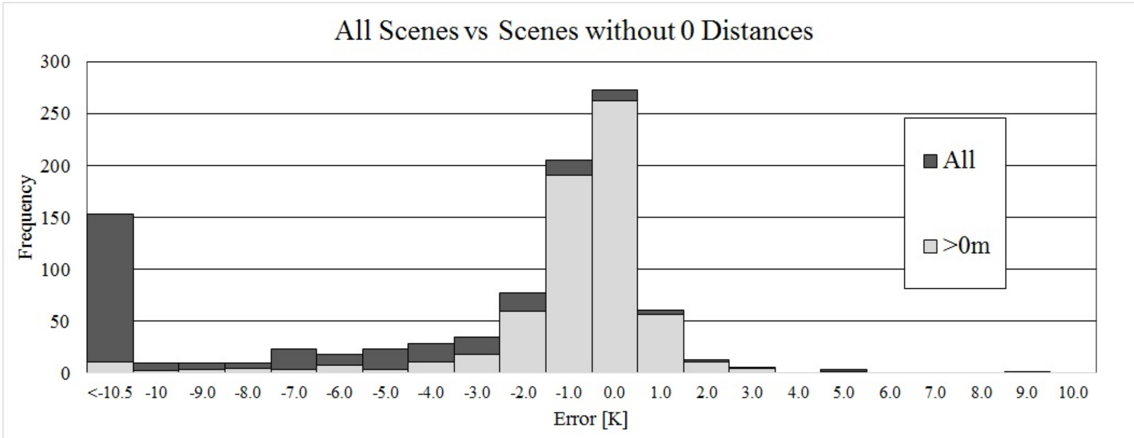

Figure 5.7: LST error from the Landsat 5 validation set, showing all scenes (dark grey bars) compared to LST error when "0 distance" scenes are removed (light grey bars). This eliminated any scene that had a cloud pixel in the same location as the buoy.

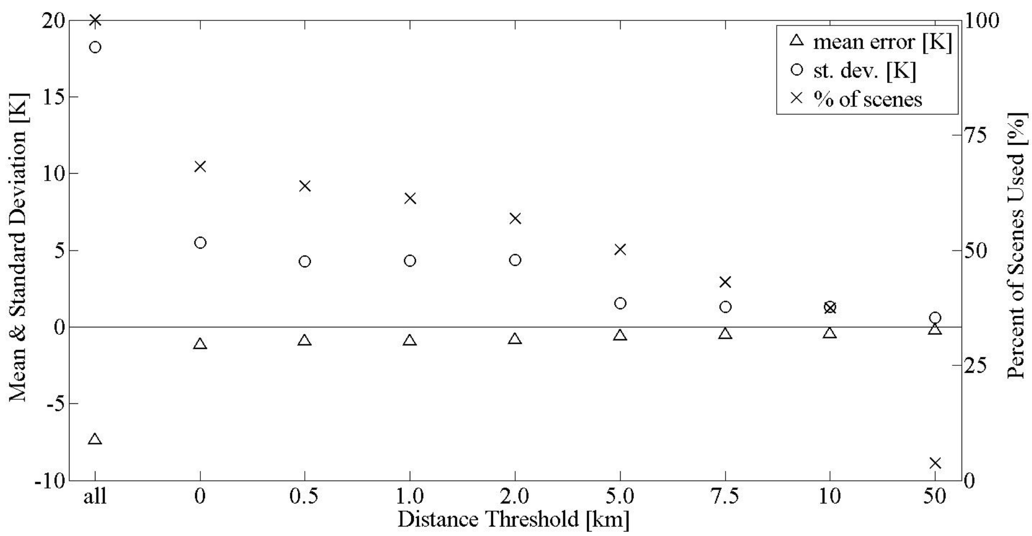

Figure 5.8: Plot of mean errors and standard deviations for different cloud distance thresholds. The thresholds are on the x axis, the mean errors and standard deviation are on the left y axis because they have the same units, and the percent of scenes used is on the right y axis.

Table 5.4: Mean error, standard deviation, RMSE, and number/percent of scenes when various distance to nearest cloud thresholds are set (Landsat 5 validation set).

Distance Threshold	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
d $\geq$ 0	-7.412	18.240	19.679	949	100.00
d > 0	-1.196	5.472	5.597	647	68.2
d > 0.5	-0.969	4.267	4.372	606	63.9
d > 1.0	-0.943	4.325	4.423	581	61.2
d > 2.0	-0.835	1.354	4.429	540	56.9
d > 5.0	-0.609	1.540	1.655	475	50.1
d > 7.5	-0.526	1.311	1.411	409	43.1
d > 10	-0.489	1.286	1.374	355	37.4
d > 50	-0.262	0.624	0.669	35	3.7

It may be more useful to examine ranges of distances to the nearest cloud and observe the average LST error for the different groups. Like before, we expect to see the error decrease as the distances increase, but the bins will give us a better idea of the errors that can be seen when the nearest cloud is located within some range away from the buoy. Figure 5.9 shows the average errors for various distance bins; for example, the first bin includes all scenes where the nearest cloud pixel was greater than 0 km away and up to 1 km away. These ranges for the bins were selected manually based on the data, because bins that are too small or too large often obscured the overall trend between cloud distance and LST error. If more data is processed, these bin edges may be redefined. The statistics for this graph are located in Table 5.5.

Table 5.5: Average errors and other statistics for various distance bins for Landsat 5 validation set. Zero distances are not included.

Cloud Distance	Avg. Error [K]	St. Dev. [K]	RMSE [K]	Samples	Percent
0-1 km	-3.424	11.183	11.614	66	10.19
1-5 km	-2.630	10.098	2.598	92	14.20
5-10 km	-1.139	2.598	2.822	80	12.35
10-40 km	-0.554	1.355	1.462	375	57.87
40-inf km	-0.262	0.624	0.669	35	5.40

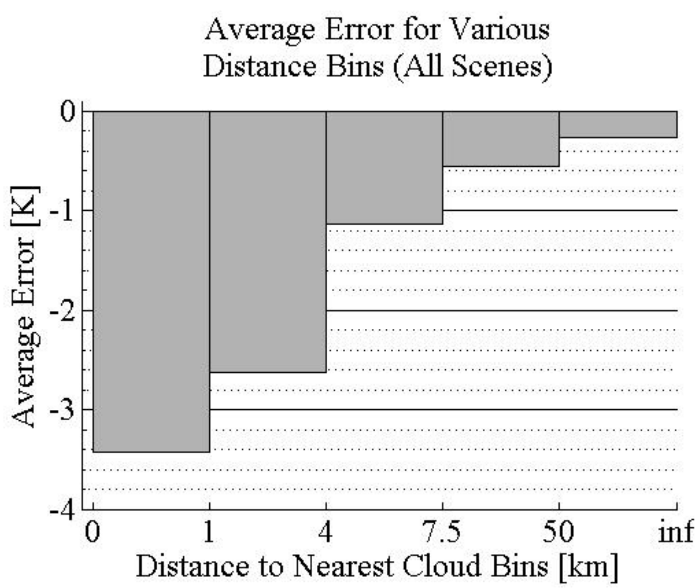

Figure 5.9: Average errors for defined distance to nearest cloud bins.

While the mean errors decrease as expected, the standard deviation and root mean squared (RMSE) errors are disturbingly high for the first two bins, indicating that there may be some outliers. Upon inspection, there were three scenes found that reported a distance to nearest cloud that was greater than zero, but the errors were extremely large, the most severe case being -93.2 Kelvin. It was discovered that for these cases the cloud mask failed to detect clouds that were actually present, which was determined by comparing the cloud mask to the true color image from the Earth Explorer website (http://earthexplorer.usgs.gov/). Figure 5.10 shows a case of the Californian coast where the distance to nearest cloud was reported to be 2.8 km, but the LST error at the buoy location was -93.2 K. It is clear that the cloud mask missed most of the clouds, which led to an erroneous distance to nearest cloud. These omissions sometimes occur in cases where the sun angle is low, as well for scenes at high latitudes. The CFmask algorithm also has difficulty capturing cirrus clouds, which is evidenced by Figure 5.11. In this instance, the distance to nearest cloud was 1.0 km, and the LST error was -16.3 K. There are several more scenes in the dataset where this type of cloud omission occurs, some of which are more drastic than others. For example, if all the cirrus clouds are far away from the buoy, it does not matter much if the cloud mask fails to identify them. When the cirrus clouds are close enough to have a negative impact on the LST prediction, the consequences are more drastic if the clouds are missed.

Since we have identified three scenes where the cloud mask omitted a severe amount of clouds, we wish to reexamine the average errors for the distance bins that were defined earlier. Figure 5.12 reflects this alteration, which largely affected the first two distance bins. Since the three scenes that were removed are considered flukes, this new histogram presents a more accurate outlook on the errors that can be expected based on how far away the nearest cloud pixel is. Table 5.6 shows how the statistics have changed since the removal of the three outlier scenes.

Throughout the rest of this chapter, we will show how we can utilize the distance to nearest cloud method to help characterize the accuracy of the LST algorithm. It will also be useful for estimating the uncertainty in the LST retrievals.

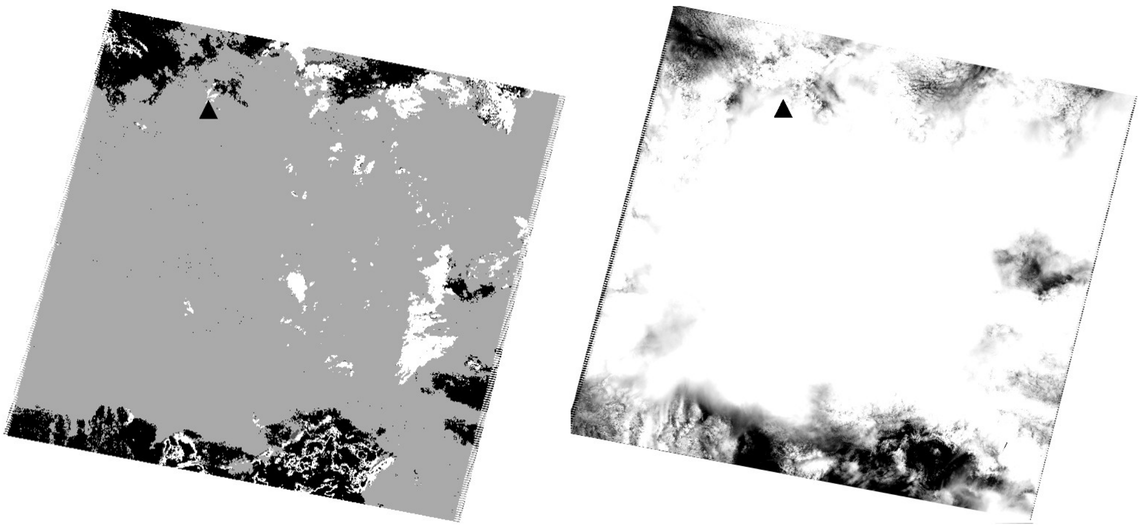

Figure 5.10: An example of the cloud mask omitting clouds. The left image is the pre-altered cloud mask and the image on the right is a grayscale version of the true color image. The black triangles indicate the location of the buoy in the scene, and are not to scale. The white areas in both images represent clouds.

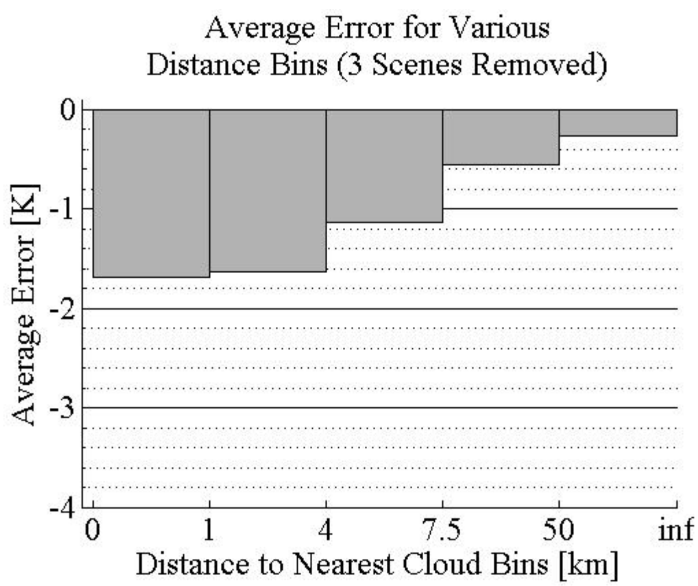

Figure 5.12: Average errors for defined distance to nearest cloud bins, with three erroneous scenes removed.

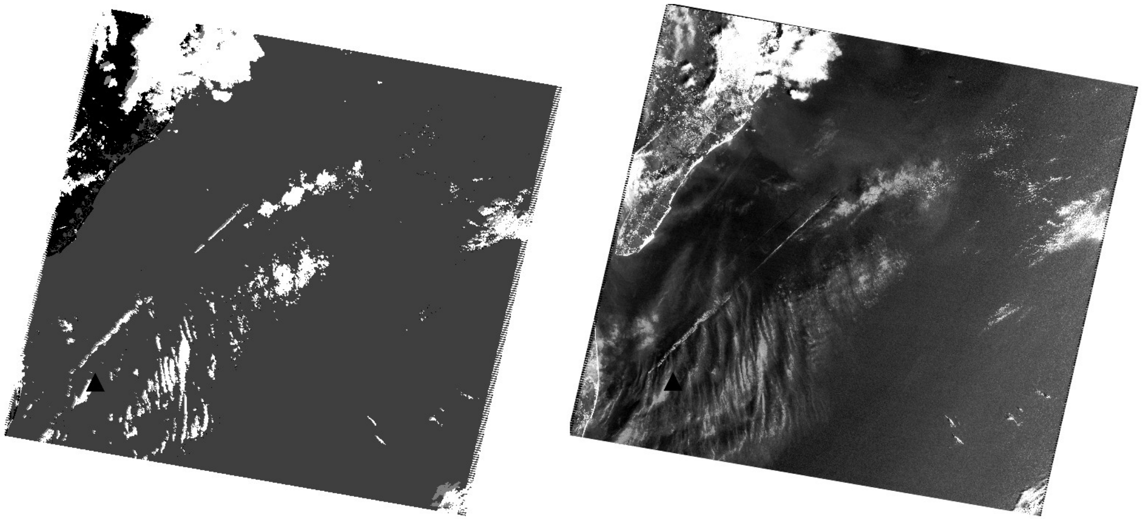

Figure 5.11: An example of the cloud mask failing to capture cirrus clouds. The left image is the pre-altered cloud mask and the image on the right is a grayscale version of the true color image. The black triangles indicate the location of the buoy in the scene.

5.4 Developing a Global Validation method

The validation results for Landsat 5 were very encouraging, and the distance to nearest cloud method proved useful for characterizing LST accuracy at different proximity levels. Although the validation of Landsat 5 was extensive, it was only for North American regions and it cannot be assumed that the LST product will behave as well on a global scale where climates and available radiosondes differ. Unfortunately, there are a few immediate obstacles that prevent us from performing global validation with our current process. The first problem is that the North American Regional Reanalysis (NARR) data is, as its name states, only available for North America. The second complication is the lack of usable buoys to provide truth data on a global scale. In order to have any hope of conducting a global validation study, we must find a global reanalysis product to replace NARR, and we

Table 5.6: Statistics for each distance bin for the Landsat 5 validation set, after scenes with extreme errors were removed. Zero distances are not included.

Cloud Distance	Avg. Error [K]	St. Dev. [K]	RMSE [K]	Samples	Percent
0-1 km	-1.691	3.449	3.817	64	9.91
1-5 km	-1.635	3.315	3.367	92	14.24
5-10 km	-1.139	2.598	2.822	80	12.38
10-40 km	-0.554	1.355	1.462	375	58.05
40-inf km	-0.262	0.624	0.669	35	5.42

must identify an acceptable source of truth data.

5.4.1 Choosing a Global Reanalysis Product

There are several global reanalysis products to choose from, but the two that show the most promise are the Modern-Era Retrospective analysis for Research and Applications (MERRA), and the Climate Forecast System Reanalysis (CFSR). These were chosen based on the date availability (for which years data exists) as well as shared characteristics with NARR. It will be noted here that NCEP reanalysis was considered briefly, but was found to be a poor candidate [Cook, 2014]. The characteristics for NARR and the two potential replacement products are listed in Table 5.7 so that differences can easily be seen. For instance, CFSR has a spatial grid resolution that is almost as fine as NARR while MERRA has a much sparser grid of reanalysis points. Another point of contrast is the fact that NARR and MERRA data are available every 3 hours, while CFSR is obtained in 6 hour increments. In order to evaluate which global reanalysis product is the best replacement for NARR, we can perform the validation process for the LST process using the different products and observe the changes in the error histogram.

Table 5.7: Comparison of the NARR reanalysis product and the two global products MERRA and CFSR.

	NARR (NOAA)	**MERRA (NASA)**	**CFSR (NOAA)**
Coverage	North America	Global	Global
Spatial	32 x 32 km (0.3° at the equator) 349 x 277 points	1.25° x 1.25° (140 km at the equator) 288 x 144 points	38 km spacing (0.5° at the equator) 720 x 361 points
Temporal	8x daily 3-hr interval	8x daily 3-hr interval	4x daily 6-hr interval
Pressure Levels	29 levels 1000-100 hPa	42 levels 1000-0.1 hPa	37 levels 1000-1 hPa

To compare NARR and MERRA, the LST process was applied to 397 Landsat 5 scenes; once using NARR for the atmospheric inputs, and once using MERRA. There were 182 scenes used to perform a similar study comparing NARR and CFSR. Both studies consist of North American scenes with buoy truth available, and are discussed in the following two subsections. We will present histograms to visually observe similarities/differences, and we will also perform t-tests to determine if the reanalysis datasets are statistically different. The t-test used assumes there are two independent samples with normal distributions and unequal variances. In the statistics tables, a simple "yes" or "no" indicates whether the t-test was passed. Passing the t-test means the datasets being compared are statistically equivalent.

We will find that MERRA is a comparable substitute for NARR, and that CFSR is not. Therefore, a more extended study was performed to determine whether MERRA was accurate enough to be used for the entire LST product, or if NARR should still be used where available. The short answer is that they produce very similar LST error distributions, but that NARR is slightly better so it should be used over MERRA where possible. The full analysis for this study can be found in Appendix D.

MERRA

The initial study comparing NARR and MERRA was performed by Dr. Cook [Cook, 2014]. A set of 397 Landsat 5 scenes containing truth buoys were used to observe changes in the LST results when MERRA is used instead of NARR for atmospheric inputs. Error histograms were produced for different combinations of cloud categories as defined in Section 5.2, but for now we will simply show the statistics in Table 5.8.

Table 5.8: Comparison of LST results when NARR and MERRA are used. Mean error, standard deviation, and number/percent of scenes are presented based on which cloud categories are included. The T-test assumes independent samples and unequal variances, and passing the test indicates that datasets are statistically equivalent.

Cloud Category	**Mean Error [K]**		**Standard Deviation [K]**		**No. of Scenes**	**Percent**	**Passes T-Test**
	NARR	**MERRA**	**NARR**	**MERRA**			
0,1,2,3,4,5	-8.470	-8.697	18.190	18.520	397	100%	Yes
0,1,2,3	-1.446	-1.474	3.724	3.610	262	66%	Yes
0,1,2	-1.002	-0.954	2.139	1.846	239	60%	Yes
0,1	-0.441	-0.513	1.471	1.086	153	39%	Yes
0	-0.235	-0.354	0.921	0.911	101	25%	Yes

The most important observation to note from Table 5.8 is that the mean errors and standard deviations are very similar between the two reanalysis products. This is a good sign, but it would also be beneficial to examine these results using the new distance-to-nearest-cloud metric. Note that with the new method, removing all the scenes that have a distance to nearest cloud of zero is most similar to the case where cloud categories 0, 1, and 2 were considered. Figure 5.13 shows the LST error histogram when NARR reanalysis was used, and when "0 distances" have been removed. Figure 5.14 shows the LST errors when MERRA was used for the same scenes. A simple visual inspection indicates that the general shapes of the histograms are similar, and Table 5.9 lists the statistics for these figures. When we compare these mean errors to the table that uses the old cloud categories, we notice that the new errors with the zero distances removed are roughly one Kelvin higher than the mean errors for the case where cloud categories 0, 1, and 2 are included. This can be easily explained since we have already discovered that the CFmask often fails to capture cirrus clouds which have a significant impact on the LST error if they are in the vicinity.

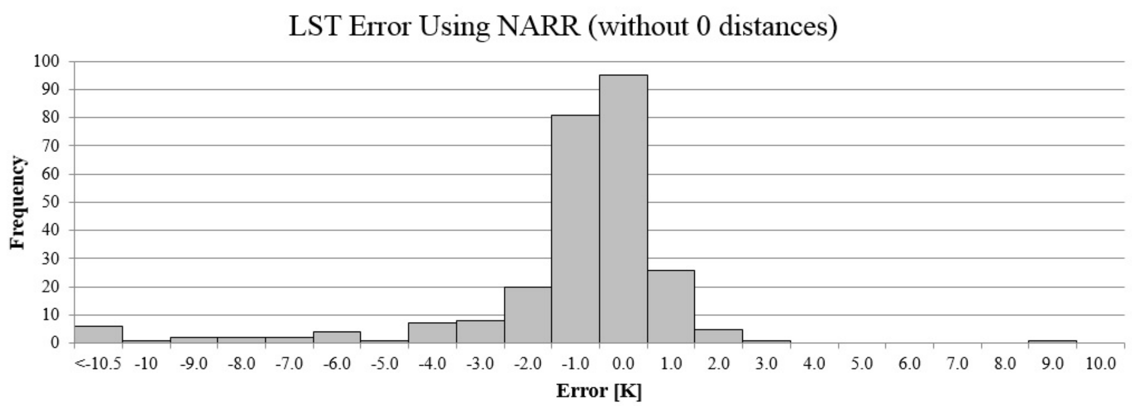

Figure 5.13: LST error histogram when NARR was used. This uses the new distance to cloud metric, where "0 distance" scenes are excluded.

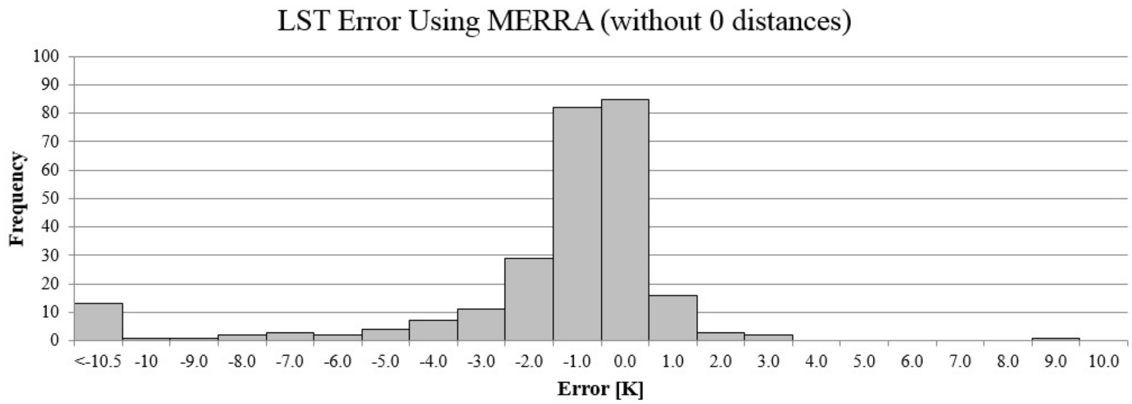

Figure 5.14: LST error histogram when MERRA was used. This uses the new distance to cloud metric, where "0 distance" scenes are excluded.

Table 5.9: Comparison of LST statistics when NARR and MERRA were used. This uses the new distance to cloud metric, and statistics are listed for different distance thresholds. The T-test assumes independent samples and unequal variances, and passing the test indicates that datasets are statistically equivalent.

Cloud Category	**Mean Error [K]**		**Standard Deviation [K]**		**No. of Scenes**	**Percent**	**Passes T-Test**
	NARR	**MERRA**	**NARR**	**MERRA**			
d $\geq$ 0 km	-8.470	-8.697	18.190	18.520	397	100%	Yes
d $>$ 0 km	-2.362	-2.344	8.509	8.290	262	66%	Yes
d $>$ 10 km	-0.640	-0.641	1.561	1.047	74	17%	Yes

CFSR

The Climate Forecast System Reanalysis (CFSR) came to our attention later on as a potential candidate for performing global validation. This particular study includes 182 Landsat 5 scenes in North America and uses the distance-to-nearest cloud metric to analyze the results. Figures 5.15 and 5.16 show the error histograms for the cases where NARR and CFSR are used, respectively. Any scene that had a distance to nearest cloud of zero was removed because they introduce the most error. Figure 5.15 looks similar to what has been seen in previous validation sets, and the bins on the left hand side represent most of the scenes that had clouds near the buoy. Figure 5.16, on the other hand, shows a much different shape and is visibly worse than the errors that were achieved using NARR. Additionally, they are visibly worse than the results obtained by using MERRA. The statistics for these two graphs can be found in Table 5.10. It is clear that MERRA is a considerably better option for our LST process, which may indicate that the temporal availability has a greater impact than the spatial resolution of the product.

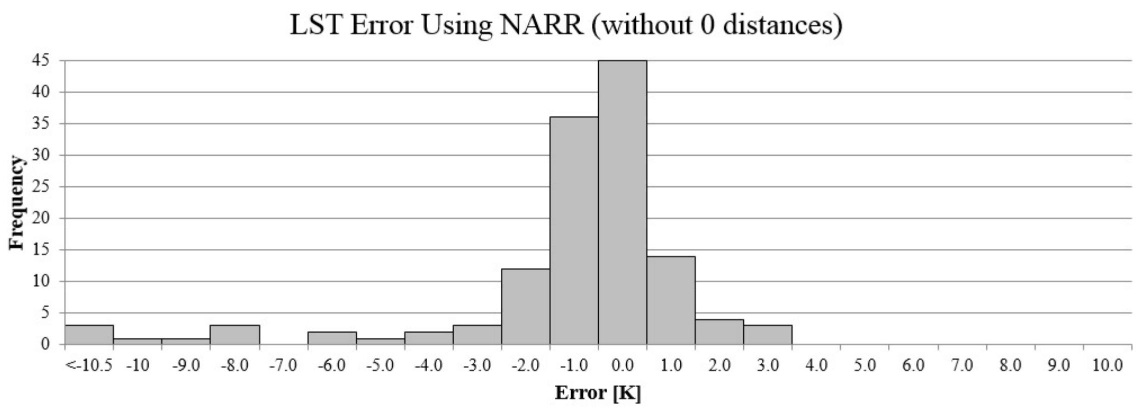

Figure 5.15: LST error histogram when NARR was used. "0 distance" scenes have been removed.

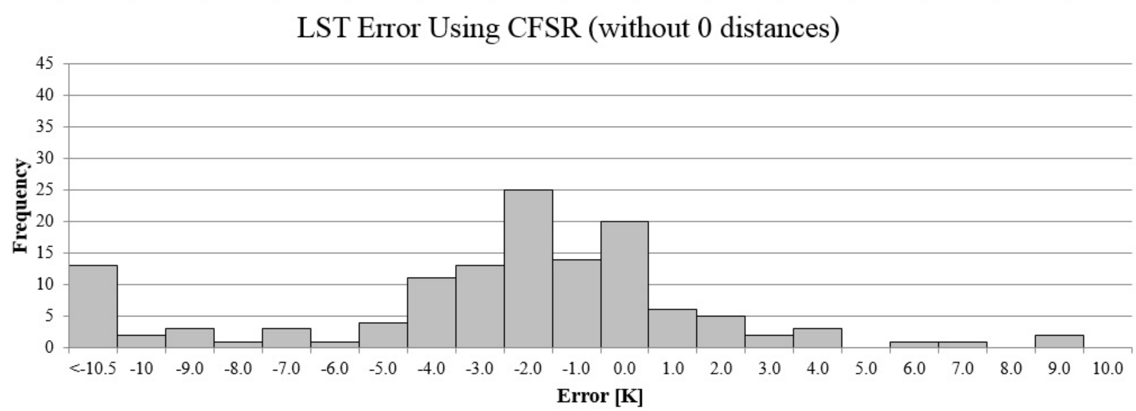

Figure 5.16: LST error histogram when CFSR was used. "0 distance" scenes have been removed.

Table 5.10: Comparison of LST statistics when NARR and CFSR were used. This uses the new distance to cloud metric, and statistics are listed for different distance thresholds. The T-test assumes independent samples and unequal variances, and passing the test indicates that datasets are statistically equivalent.

Cloud Category	Mean Error [K]		Standard Deviation [K]		No. of Scenes	Percent	Passes T-Test
	NARR	**MERRA**	**NARR**	**MERRA**			
d $\geq$ 0 km	-3.927	-7.454	9.190	18.411	182	100%	No
d $>$ 0 km	-1.692	-4.805	8.496	17.252	130	71%	No
d $>$ 10 km	-0.356	-2.413	1.779	5.316	73	40%	Yes

5.4.2 Choosing a Source of Ground Truth

Since there is a lack of buoys that are reliable and that meet our requirements on a global scale, we are driven to consider satellite-derived options for surface truth. One promising avenue is the MODIS Sea Surface Temperature (SST) product. MODIS is a sensor that is onboard the two NASA spacecrafts known as Aqua and Terra. In order to be a feasible source of ground truth, MODIS SST images must be captured within a short time from the

Landsat acquisition time, which means it may only be usable for certain Landsat sensors based on their orbit tracks. Unfortunately, the only Landsat sensor that has a similar orbit to either of the MODIS sensors is Landsat 7, which lines up quite nicely with the MODIS sensor aboard the Terra satellite. In general, they follow similar orbital tracks and TERRA is typically 15-30 minutes behind Landsat. This means that for a particular Landsat scene, the corresponding MODIS scene is typically captured 15-30 minutes after the Landsat scene. By "corresponding MODIS scene," we mean that the MODIS image includes the same area that was captured within the Landsat frame, and the pixels are viewed at similar angles. Although the SST product cannot be used for all Landsat sensors, it has the potential to be used for the global validation of Landsat 7. This section will determine whether the MODIS SST product can safely be used for this purpose.

MODIS SST Study

Once the correct MODIS scene is found, we want to compare both the SST and the LST values to the truth temperature provided by a buoy in the scene. This will allow us to assess how well the SST product emulates the true surface temperature. The basic process for this study involved selecting a set of Landsat 7 scenes that contain usable truth buoys, downloading the corresponding MODIS SST image, georeferencing that image, and subsetting it to roughly the same size and area as the Landsat scene. Then we simply average a 5 x 5 window of LST predicted temperatures around the buoy, and a 3 x 3 window around the buoy in the SST image, and compare each to the truth value. Once again, we used the emissivity of water to calculate surface temperatures. To help visualize the process, Figure 5.17 shows an example MODIS SST image that has been georeferenced, the subset SST image, and the corresponding Landsat scene. The buoys are marked by black triangles, and the white squares that surround them represent the pixels around the buoys that are averaged to obtain the SST and LST values that are compared to the buoy truth temperature (averaging gets a good estimate for the desired pixel and reduces the influence of any potential outlier pixels). The windows are different sizes because of the different resolutions between MODIS and Landsat.

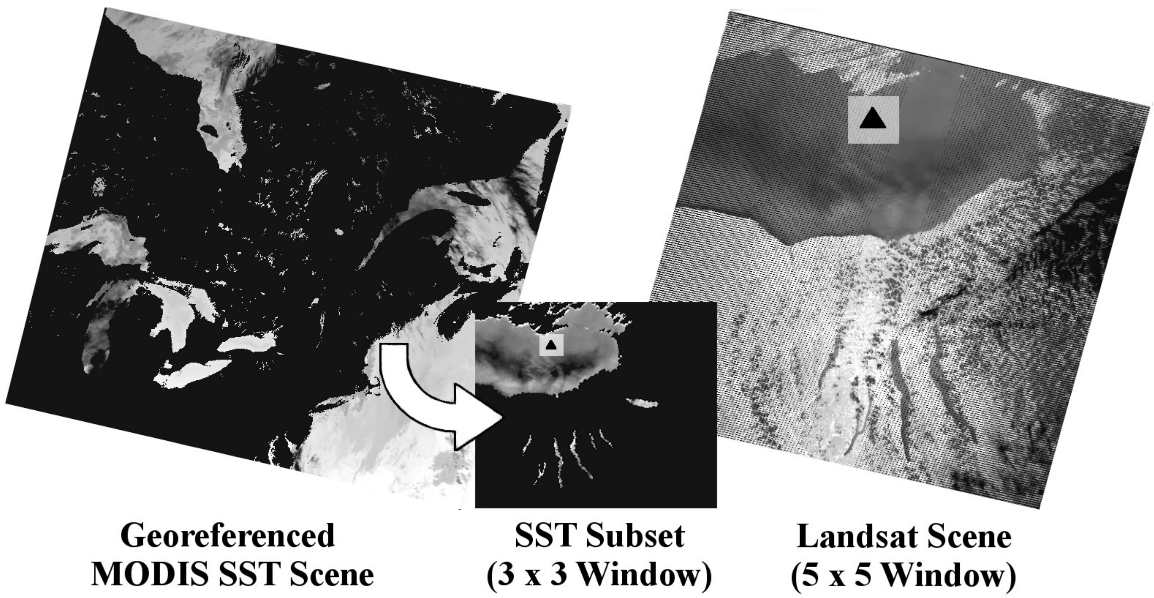

Figure 5.17: Process of comparing MODIS SST and LST values to buoy truth. The SST image is subset to reduce computation time, and a 3 x 3 window around the buoy is averaged to get the SST value at that location. A 5 x 5 window around the buoy of LST values is averaged for the Landsat scene. Buoy and window sizes are not to scale.

A total of 144 scenes were chosen from the same North American sites that were used in previous studies. Some scenes did not have buoy data available so only 118 images were able to be used. The MODIS SST product also includes a quality band, which rates pixels from 0 to 4, where 0 is the best quality and 4 is an invalid pixel. In our study, we will only consider cases where the 3 x 3 window around the buoy has an average quality metric of zero (in other words, all pixels in the window have a rating of zero). After filtering out cases where the average quality was greater than zero, we were left with 75 scenes. These best quality results can be found in Figures 5.19 and 5.21, which contain the errors seen in the LST predictions and the SST product, respectively. The statistics for this study can be found in Table 5.11.

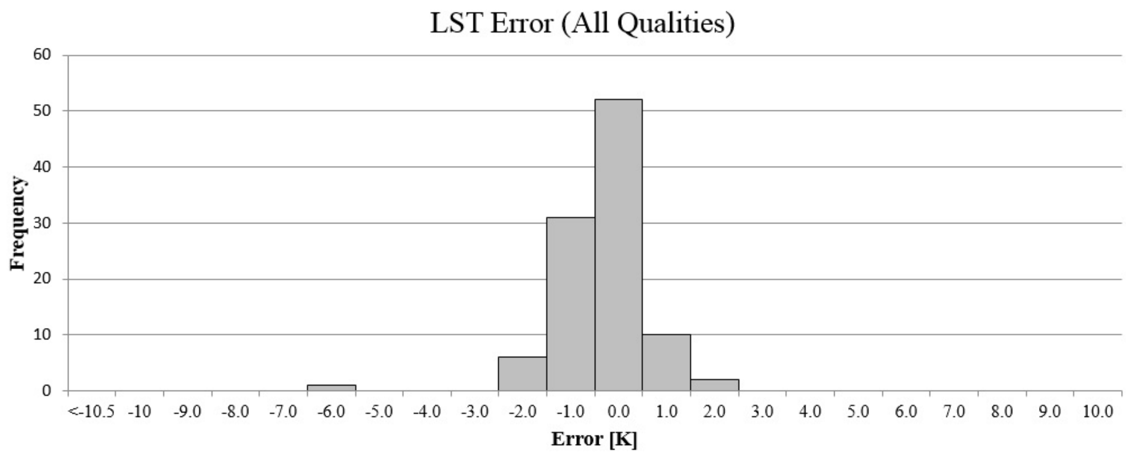

Figure 5.18: Landsat 7 LST errors using NARR reanalysis and buoy-derived truth, including all scenes.

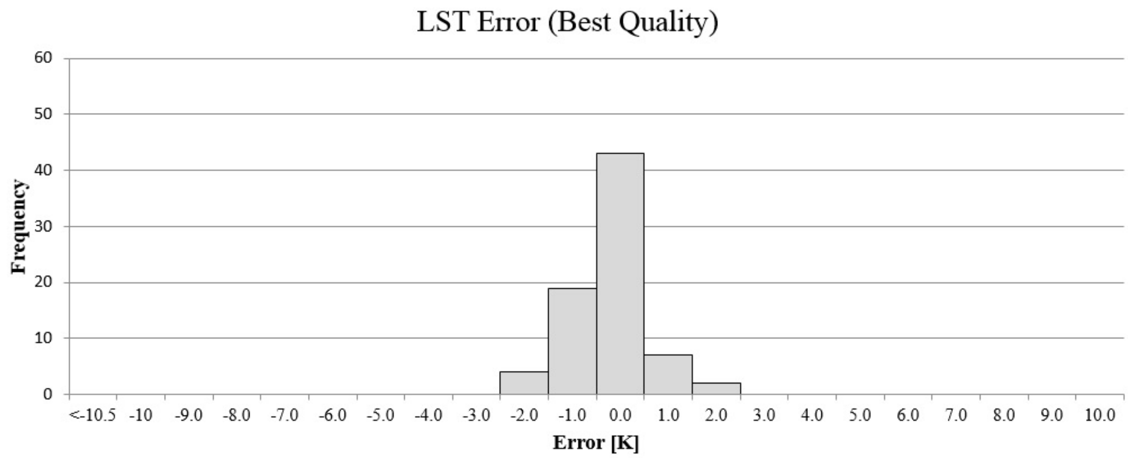

Figure 5.19: Landsat 7 LST errors using NARR reanalysis and buoy-derived truth, including only the best quality scenes.

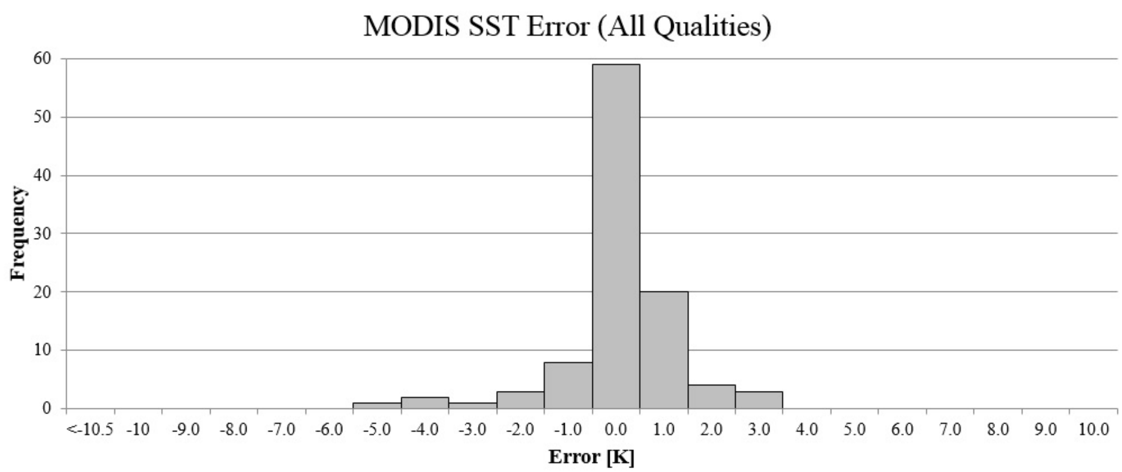

Figure 5.20: MODIS SST errors against buoy-derived truth, including all scenes.

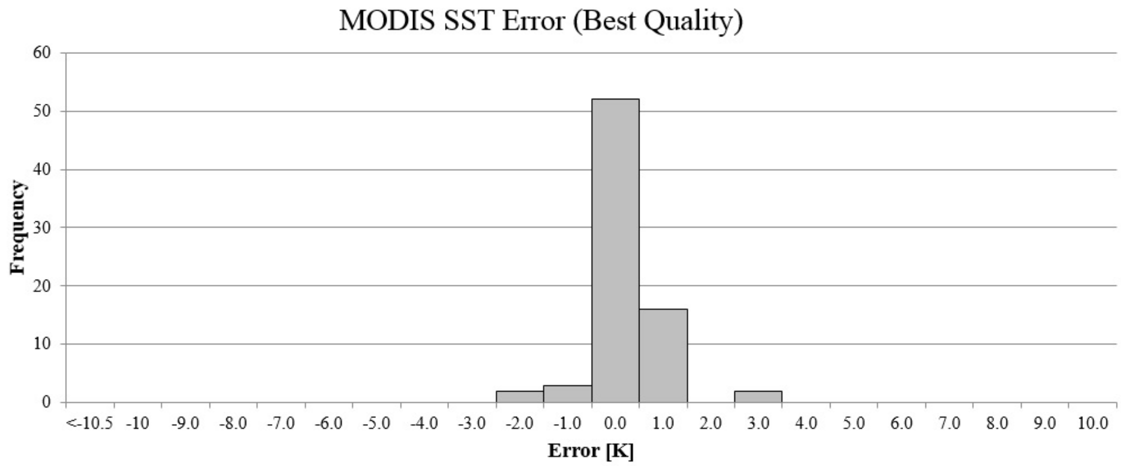

Figure 5.21: MODIS SST errors against buoy-derived truth, including only the best quality scenes.

Table 5.11: A summary LST and SST errors compared to buoy truth data.

Dataset	**No. of Scenes**	**Mean Error [K]**	**Standard Deviation [K]**	**RMSE [K]**
LST using NARR	118	-0.351	0.891	0.953
LST using NARR (Best Quality)	75	-0.241	0.701	0.737
MODIS SST	118	0.214	1.599	1.605
MODIS SST (Best Quality)	75	0.244	0.699	0.736

As one would expect, the error histograms improve when only the best quality scenes are used. The histograms for the LST errors resemble what we have seen with the Landsat 5 validation results, which shows that Landsat 7 works well with our process. For the MODIS SST results, we can see that the average error is positive, which means it tends to slightly overestimate the temperature. Also, the MODIS average error for the best quality scenes is slightly higher than when all scenes were used, but the standard deviation values and the shapes of the histograms show that while the error slightly increases, the spread decreases significantly. In other words, the accuracy decreased marginally, but the precision improved considerably. A larger set of scenes would most likely show the trend that was expected, which was an increase in both accuracy and precision. A t-test shows that although the errors are quite low, the population means between the Landsat versus buoy errors and the SST versus buoy errors are significantly different (best quality cases only). This suggests that there may be some correctable bias in one or both of the products; however, the errors are so low that any further attempts to reduce them is not a major focus of this study.

After examining the performance of the LST process and the MODIS SST product separately, we can create a new histogram that shows the LST error when MODIS is used as truth. Figure 5.22 shows this comparison for the best quality scenes, where the mean error is -0.485 K, the standard deviation is 0.795 K, and the RMSE is 0.927 K. These numbers give us enough confidence to move forward with the global validation of Landsat 7 using the MODIS SST product as a source of truth data.

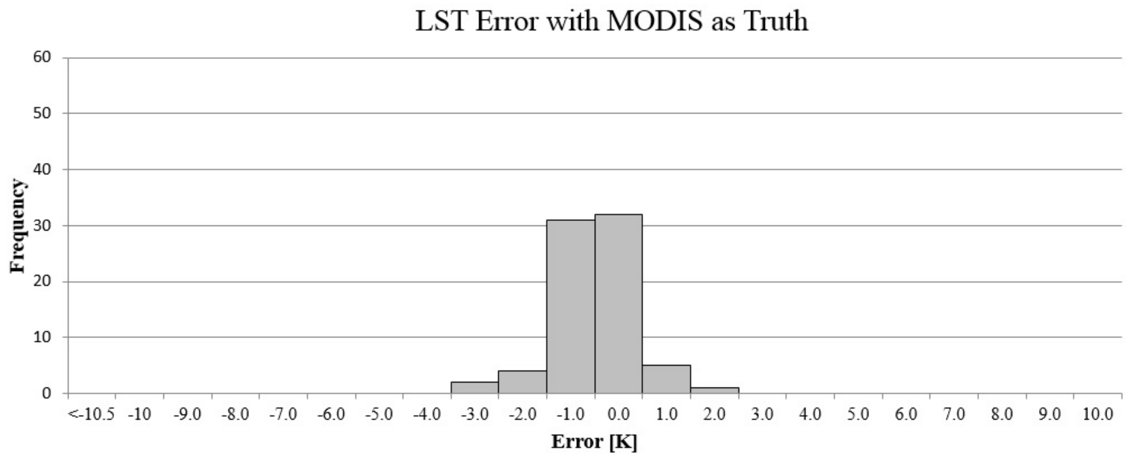

Figure 5.22: LST error histogram when MODIS SST is used as truth values.

5.5 Global Validation of Landsat 7

Now that we have reached a solution to our reanalysis and truth data problems, we can finally begin to validate Landsat 7 on a global scale. This section will walk through validation results for 14 different sites around the world, with several goals in mind. First, we will see how the LST algorithm performs for each of the sites. Second, we will group the data into different climate types to see if the algorithm performs noticeably better or worse in each case. Third, we will present the data for different cloud bins, to ensure that the distance to nearest cloud method still shows promise now that we are working on a global scale. Starting in Section 5.5.5, the concept of utilizing transmission to sort the validation data is introduced and implemented. Finally, Section 5.5.6 divides the validation data into different transmission bins as well as the cloud distance bins that have been previously established. Section 5.5.1 contains an overview of the dataset that was used for this extensive study, which also highlights alterations that were made to the dataset for various reasons.

5.5.1 Dataset Overview

There was a small, initial global validation study conducted by Cook that used MERRA reanalysis and MODIS SST as truth before they were thoroughly proven to be appropriate replacements for NARR and buoy-measured temperatures [Cook, 2014]. Cook chose 11 global sites to represent different climate types and radiosonde abundance, in order to see how the LST algorithm performed under a variety of conditions. This study consisted of a total of 63 samples that were manually identified as being cloud free, and although the average errors at each site were around ± 1 K, there were too few samples to draw any conclusions.

In Section 5.4, we described the process of how we compare points within the Landsat 7 scenes and their corresponding MODIS SST images. In that study, we chose to only permit best quality pixels in our validation (as defined by the SST product), which severely limited how much data got through. Since that study, the SST product underwent an update that included a change to the quality band. While the quality ratings remained the same, there were much fewer "best quality" pixels in general. In order to obtain a large enough dataset for global validation, we required that only 5 out of the surrounding 3 x 3 must be best quality. We also enforced other restrictions for choosing appropriate comparison points, such as an upper limit on the standard deviation for the surrounding area. Details concerning this selection process is detailed in Appendix B. In total, we ran the LST algorithm for several hundred scenes per site, and allowed up to seven comparison points per scene if they met our criteria. Most of the processed Landsat scenes were from 2009 to 2015, but some go back as far as 2004. Out of all of these scenes that were processed, a total of 3634 samples were able to be used to validate Landsat 7.

The initial global validation study implemented by Cook consisted of 11 global sites, but we have added three North American sites to make 14 total sites. The locations of these sites can be seen in Figure 5.23, and the descriptions of each site is listed in Table 5.12. In our analysis, we will first observe how the LST algorithm performs for each of these sites, and from there try to characterize LST error under various atmospheric conditions. For all of the results we will present from now on, there will be no "0 distance" cases included, and also no errors less than -10 K included. We have shown in previous sections that when there is a cloud directly above a pixel of interest, the LST retrieval for that pixel should not be trusted. In addition, we also showed that the cloud mask sometimes omits clouds, which caused us to see many LST errors that were very negative, less than -10 K. When we remove just the "0 distance" cases, we still see several cases where error was less than -10 K because of omitted clouds. Since these points are not an accurate reflection of the LST product's performance (we cannot expected LST to give appropriate surface temperatures when there are clouds in the way), we have chosen to remove them from our global validation results.

We have also added some biases to the data based on the Landsat 5 validation and the study between LST and SST. Recalling the Landsat 5 validation results, we determined that under ideal conditions (i.e. no clouds) the LST algorithm underestimates buoy truth by an average of 0.241 K. In the LST versus SST study (Section 5.4.2), we saw the the SST product tends to overestimate buoy truth by 0.244 K. We included these biases in our global validation dataset by adding 0.241 K to the LST retrievals (which accounts for the slight bias in the LST process) and by subtracting 0.244 K from the MODIS SST values to better approximate buoy values.

As we present the global validation results for Landsat 7, we will be interested in several statistics. The average error will indicate if there is a bias in the LST retrievals (for example if it consistently underestimated truth by 0.5 K). The standard deviation will relate to the spread of the data, or how precise the LST algorithm is. The root mean squared error (RMSE) encompasses both the average errors and the variation in errors, which makes it a useful metric that is sensitive to outliers. We are also interested in the number of samples there are in each plot, because large samples increase our trust in the shape of the error histograms. We want the standard deviation to be as low as possible because it relates to the precision of the algorithm. The accuracy of the algorithm can be related to the average error, but if the precision is good then a bias can be applied to get the average error close to zero.

Table 5.12: Description of each site that was selected, including climate type, WRS-2 location, and radiosonde rating.

Location	**WRS-2 path_row**	**Lat, Lon**	**Radiosonde Rating**	**Description**
North Brazil	216_63	-4.26, -37.7	1	Tropical
Mediterranean	196_30	43.3,4.8	2	Mid lat - Northern
Black Sea	174_30	43.47,38.91	2	Mid lat - Northern
India	144_54	9.0, 76.34	0	Tropical
Hong Kong	121_44	22.46, 114.9	2	Low lat - Northern
Russia	107_19	58.85,149.43	1	High lat - Northern
Australia	113_82	-31.9, 114.95	2	Mid lat - Southern
Africa	180_75	-22.0, 14.0	0	Low lat - Southern
Greenland	232_17	61.5, -41.75	1	High lat - Northern
South Brazil	218_77	-24.045, -45.18	2	Low lat - Southern
South Chili	233_93	-47.88, -75.45	1	Mid lat - Southern
Georgia Coast	16_38	31.75, -80.50	2	Mid lat - Northern
Lake Huron	20_29	44.61, -82.74	2	Mid lat - Northern
California	43_36	34.62, -121.42	2	Mid lat - Northern

Figure 5.23: Map of Landsat 7 global validation sites.

5.5.2 Validation results for each site

Within this section we will look at the LST error for each global validation site. The histograms for these sites will give a general idea of how well the algorithm estimates surface temperature, but their corresponding statistics can be found in Table 5.12.

Figures 5.24 and 5.25 are the error histograms for the Russia site and the Greenland site, which both have cold climates. They both show encouraging histogram shapes, although the Greenland site only contains 12 samples.

Figures 5.26, 5.27, 5.28, and 5.29 show the error histograms for sites that tend to be very warm and arid. They are generally similar between each other, although the Africa site has a very low average error even though its standard deviation is within about 0.1 K of the other figures just mentioned. The Africa site only has 48 samples, so it may be that there were not many cases where clouds caused the algorithm to vastly underestimate surface temperature.

Figures 5.30, 5.31, 5.32, and 5.33 show the error histograms for sites with more moderate climates. Again, we see relatively similar shapes and varying standard deviations.

Figures 5.34, 5.35, 5.36, and 5.37 are histograms for sites that generally have a very warm but very humid climate. This type of climate will tend to have low transmission because of all the moisture particles in the atmosphere, which makes it difficult to perform atmospheric compensation well. The histograms have noticeably more spread than the other sites, although they all still have their highest bin centered on zero.

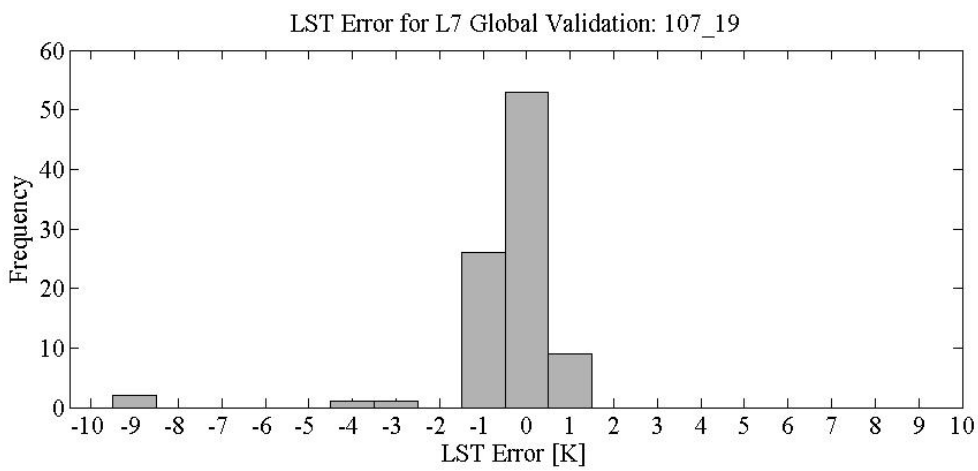

Figure 5.24: L7 global validation LST error histogram for path_row 107_19 (Russia).

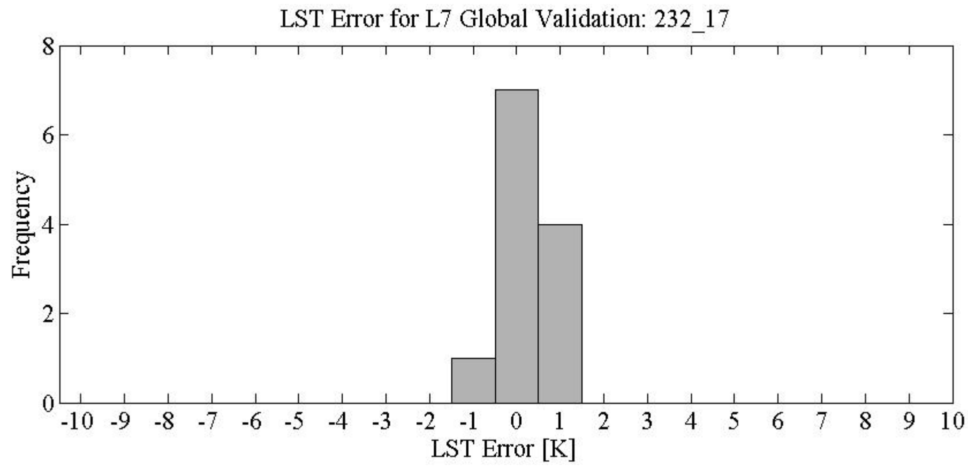

Figure 5.25: L7 global validation LST error histogram for path_row 232_17 (Greenland).

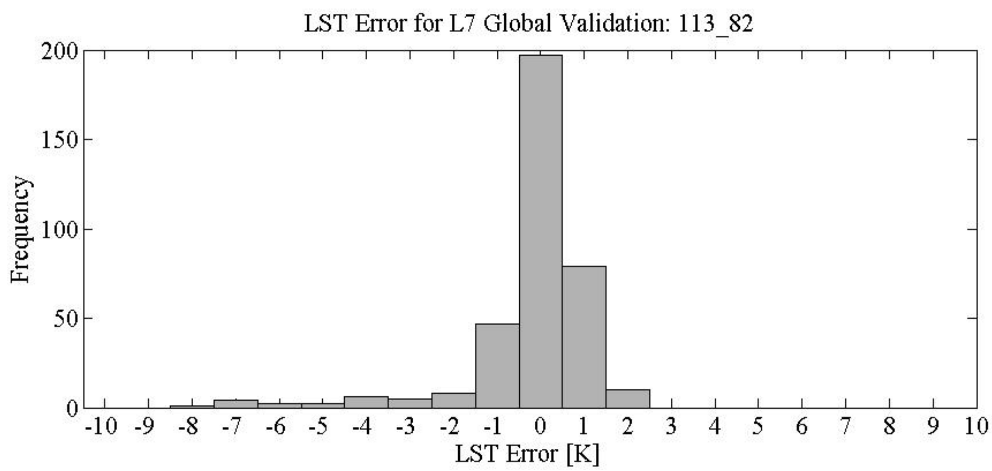

Figure 5.26: L7 global validation LST error histogram for path_row 113_82 (Australia).

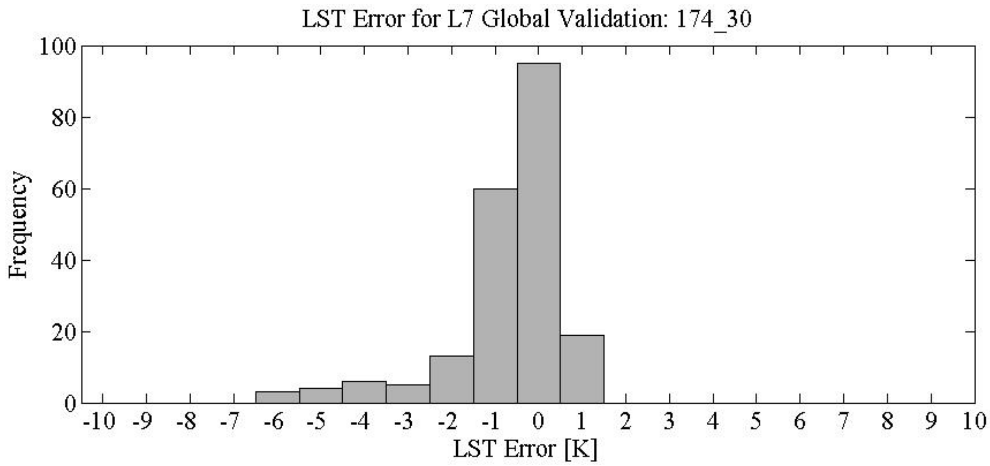

Figure 5.27: L7 global validation LST error histogram for path_row 174_30 (Black Sea).

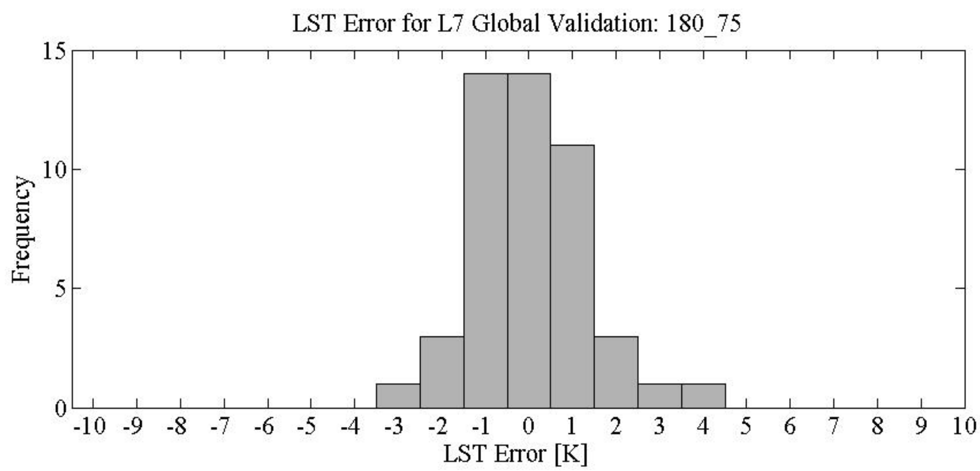

Figure 5.28: L7 global validation LST error histogram for path_row 180_75 (Africa).

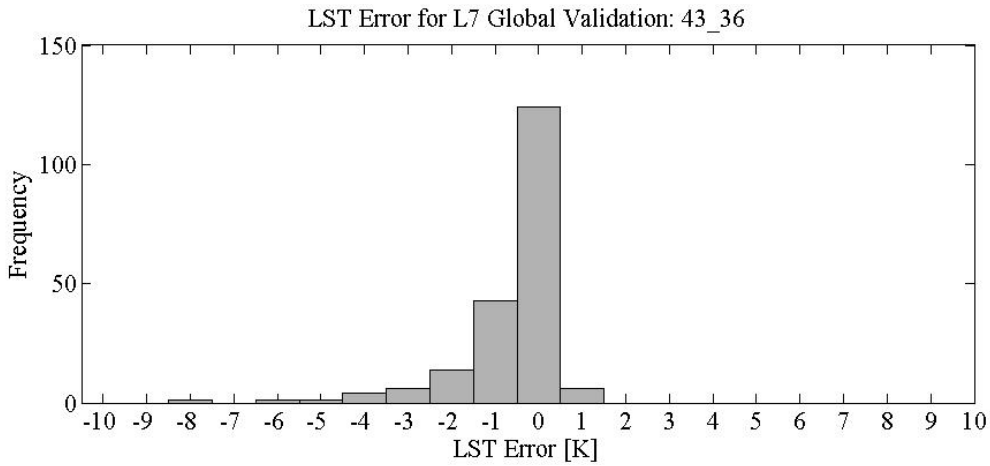

Figure 5.29: L7 global validation LST error histogram for path_row 43_36 (California).

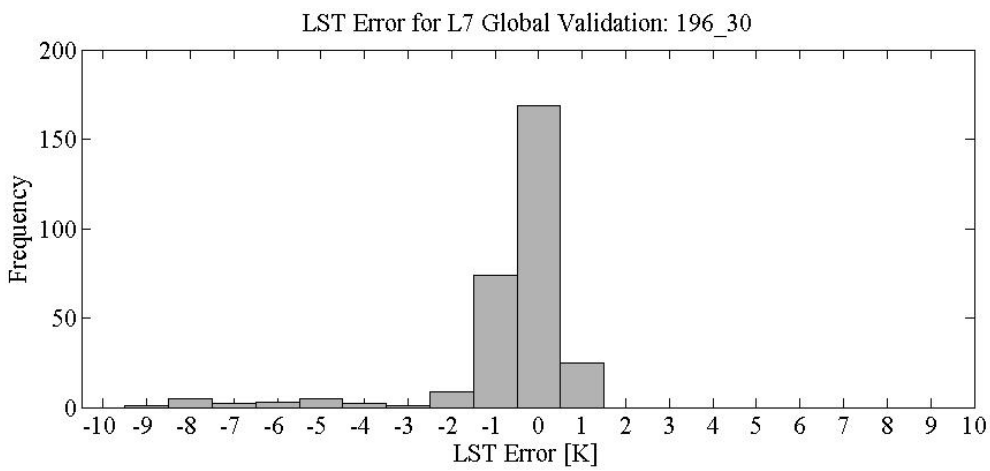

Figure 5.30: L7 global validation LST error histogram for path_row 196_30 (Mediterranean).

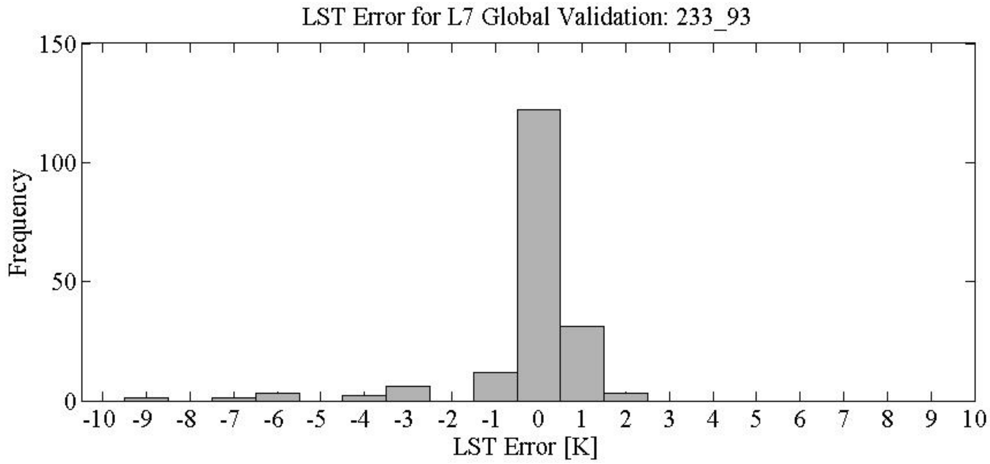

Figure 5.31: L7 global validation LST error histogram for path_row 233_93 (South Chili).

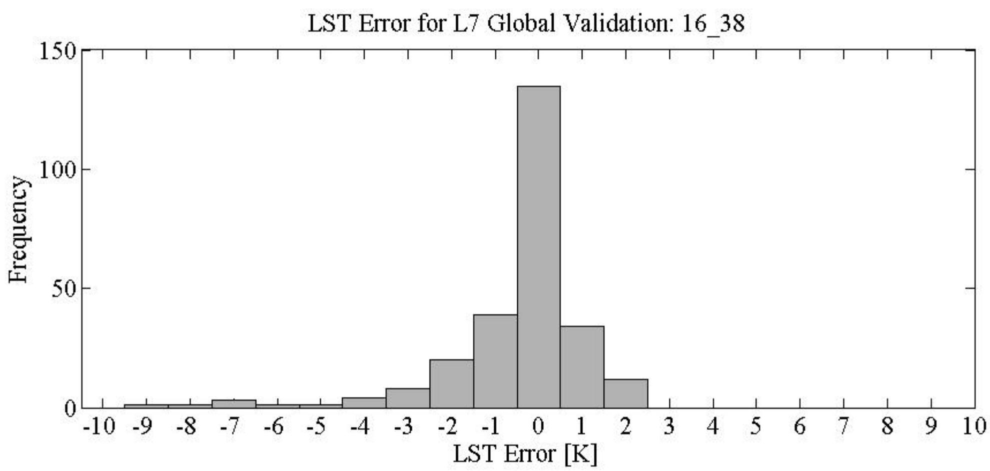

Figure 5.32: L7 global validation LST error histogram for path_row 16_38 (Georgia Coast).

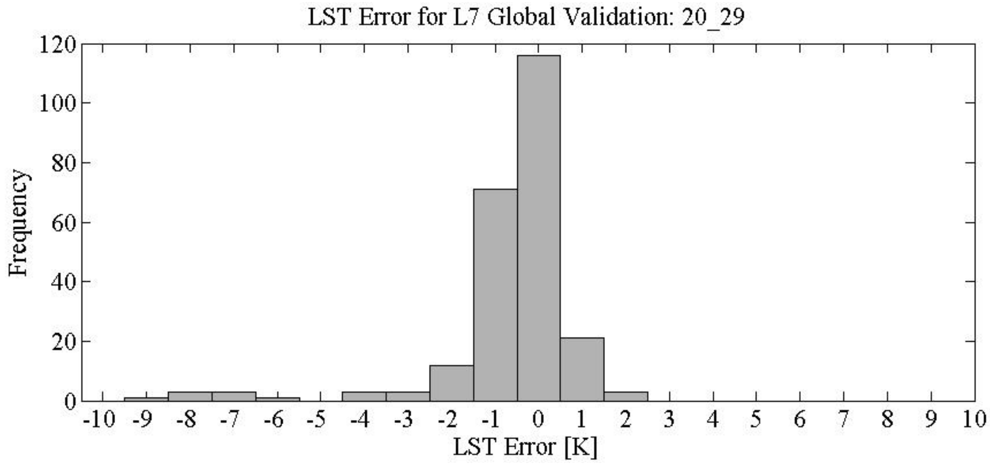

Figure 5.33: L7 global validation LST error histogram for path_row 20_29 (Lake Huron).

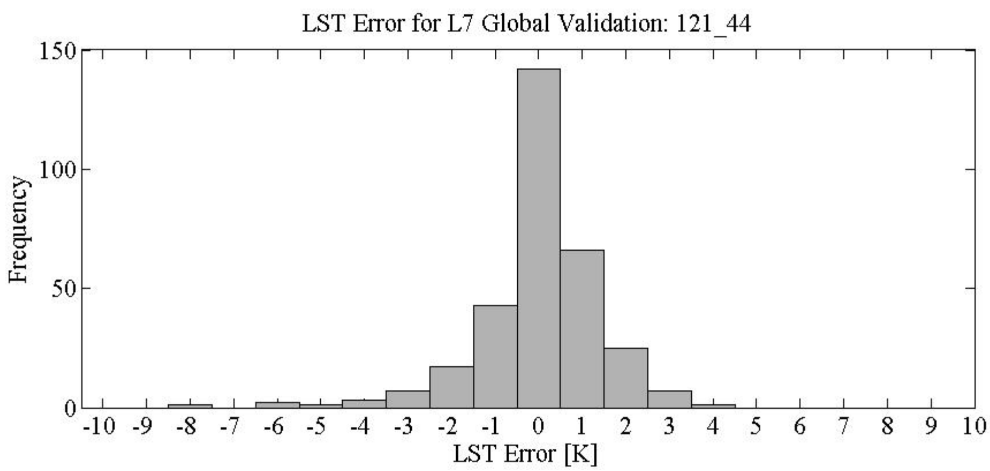

Figure 5.34: L7 global validation LST error histogram for path_row 121_44 (Hong Kong).

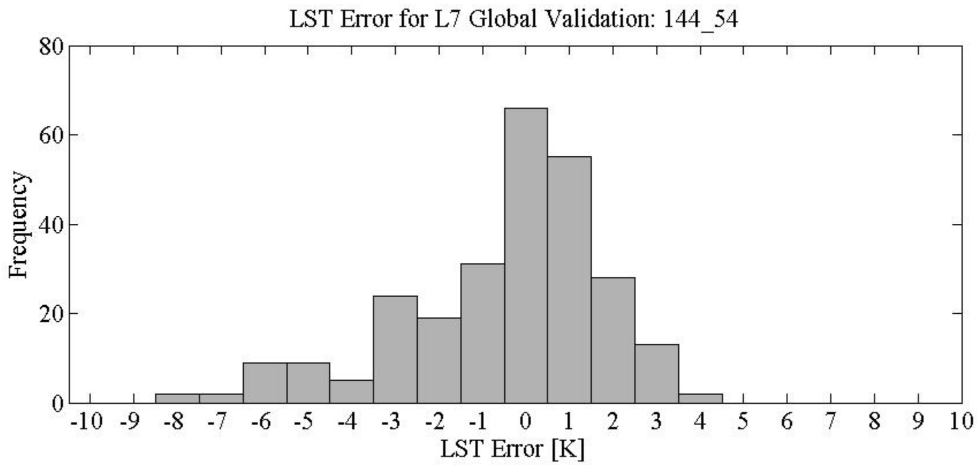

Figure 5.35: LST error histogram for path_row 144_54 (India).

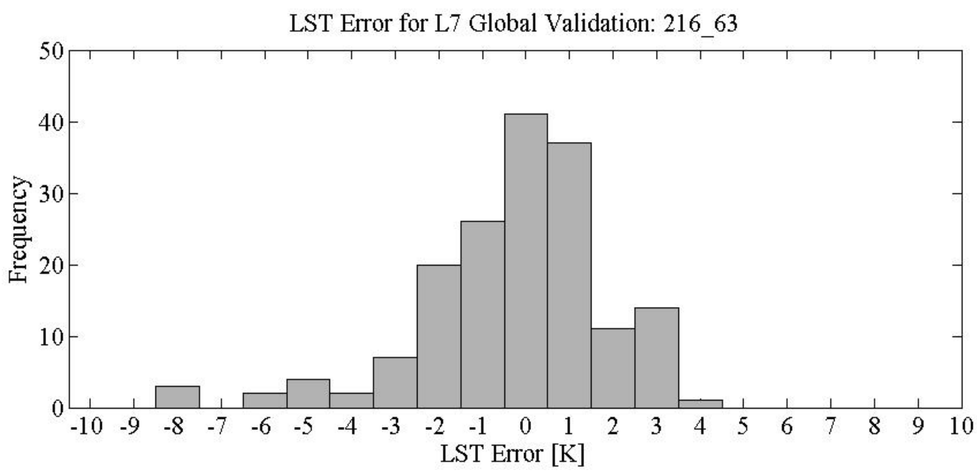

Figure 5.36: L7 global validation LST error histogram for path_row 216_63 (North Brazil).

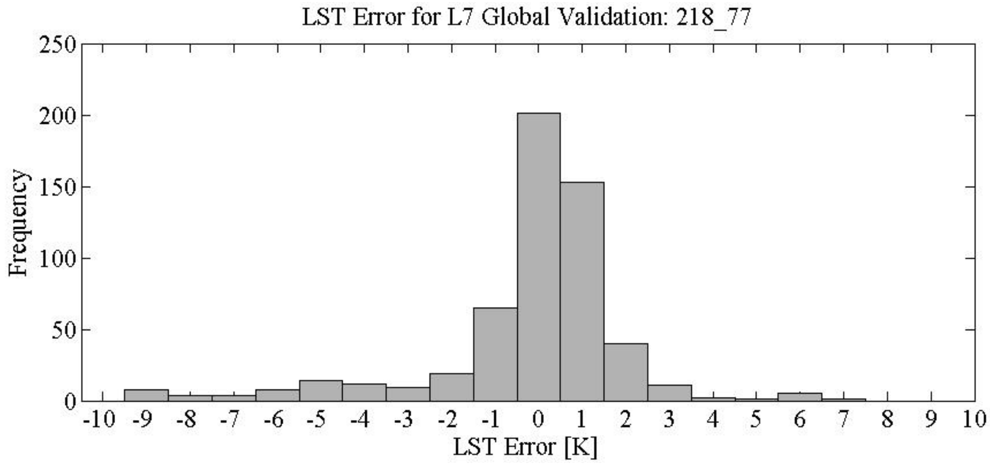

Figure 5.37: L7 global validation LST error histogram for path_row 218_77 (South Brazil).

Table 5.13: LST error statistics for each of the 14 global validation sites, excluding cases where clouds were directly over the pixel of interest, and excluding cases where LST error was less than -10 K.

Path_Row	**Location**	**Avg. Error [K]**	**St. Dev [K]**	**RMSE [K]**	**Samples**	**Percent**
107_19	Russia	-0.5043	1.4407	1.519	92	2.88
113_82	Australia	-0.1424	1.3907	1.3961	361	11.30
121_44	Hong Kong	0.0415	1.399	1.3973	315	9.86
144_54	India	-0.5139	2.326	2.3778	265	8.29
174_30	Black Sea	-0.7148	1.3191	1.4975	205	6.41
180_75	Africa	0.0765	1.3528	1.3408	48	1.50
196_30	Mediterranean	-0.6561	1.5909	1.7184	296	9.26
216_63	South America	-0.2558	2.1678	2.1764	168	5.26
218_77	South America	-0.1966	2.211	2.2178	557	17.43
232_17	Greenland	0.0993	0.5428	0.5291	12	0.38
233_93	South America	-0.1866	1.4195	1.4278	181	5.66
16_38	Georgia Coast	-0.4553	1.5546	1.617	259	8.10
20_29	Lake Huron	-0.6604	1.524	1.658	237	7.42
43_36	California Coast	0.604	1.1511	1.2974	200	6.26
TOTAL	—	-0.3562	1.7349	1.7709	3196	100.00

5.5.3 Validation results for climate types

Based on the histograms for each of the individual 14 global validation sites, we decided to sort the data into a few general climate types. The motivation behind this was to observe any noticeable trends or common histogram shapes between sites that have similar climates, which would indicate how accurate the LST product is for various climate types. We chose to divide the 14 global validation sites into four climate types: cold, moderate, hot/arid, and hot/humid. These were assigned using the guidance of the Köppen-Geiger climate classificaiton map [Rubel and Kottek, 2010]. This classification map uses a combination of letters to describe different areas of the world. There are five letters that indicate some type of main climate (e.g. equatorial, arid, polar), and there are sets of letters to describe the precipitation and the temperature of a given area. In our case, we do not need to be so specific about the climate at each site we are using; rather, we want to observe how the LST algorithm performs for "generally" cold scenes or "generally" hot and humid scenes. Table 5.14 shows the 14 global validation sites as well as the climate they were assigned.

Table 5.14: Table categorizing each of the 14 global validation sites into 4 climate types

Path_Row	**Location**	**Climate Type**
107_19	Russia	Cold
232_17	Greenland	Cold
113_82	Australia	Hot & Arid
174_30	Black Sea	Hot & Arid
180_75	Africa	Hot & Arid
43_36	California	Hot & Arid
196_30	Mediterranean	Moderate
233_93	South Chili	Moderate
16_38	Georgia Coast	Moderate
20_29	Lake Huron	Moderate
121_44	Hong Kong	Hot & Humid
144_54	India	Hot & Humid
216_63	North Brazil	Hot & Humid
218_77	South Brazil	Hot & Humid

Figures 5.38, 5.39. 5.40, and 5.41 show the LST error histograms for each climate type. They show what we began to suspect when we looked at the histograms for the individual sites; namely, that the algorithm performs the best for cold climates, and the worst for hot and humid locales. The LST errors are slightly worse for the moderate cases than the hot/arid cases, most likely because moderate climates tend to have more humidity. Table 5.15 shows the statistics for these results. Since we are seeing that the algorithm performs differently in these different climates, we suspect that the reanalysis products do not always provide an accurate depiction of atmospheric variables along the vertical column. It may be that the reanalysis products' estimates of humidity are not as reliable in climates that are especially hot and humid, which could explain why the error histograms have a larger spread.

It is apparent that climate type has an impact on the LST algorithm performance, but it is not a feasible relationship to implement in the final LST product. If we try to use climate type as a way of estimating the error in the LST retrievals, it requires that we know the climate type for every Landsat scene. Additionally, our climate types are very general and were chosen for this small-scale study; it would be unwise to try to assign every location on Earth to just four climate types. The ideal solution would be to somehow capture the effect of climate type using a metric that can be easily attained through the LST process. Our first attempt at this was to use transmission to try to categorize the data, because we already know that the algorithm performs better for cold and hot/arid scenes that tend to have high transmission values. The discussion of this approach can be found in Section 5.5.5, but first we will dedicate the next section to presenting the validation data using the distance to nearest cloud method.

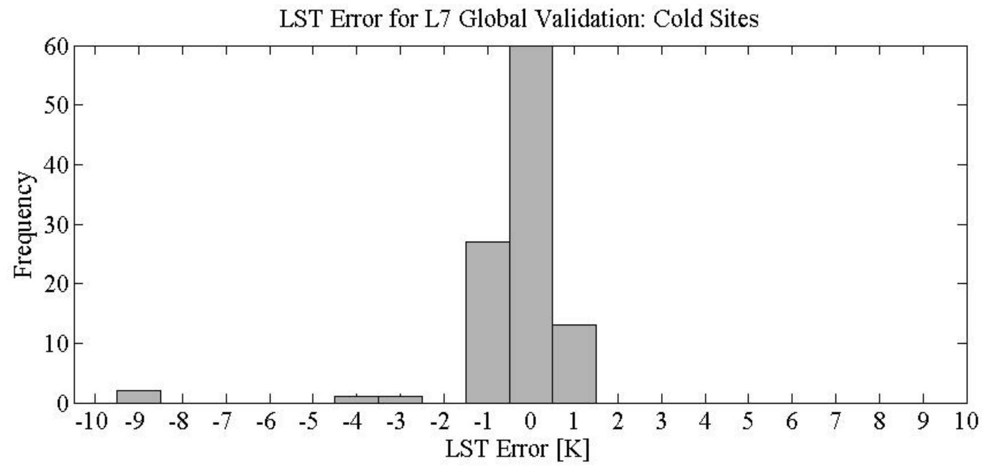

Figure 5.38: LST error histogram for global validation scenes with a "cold" climate.

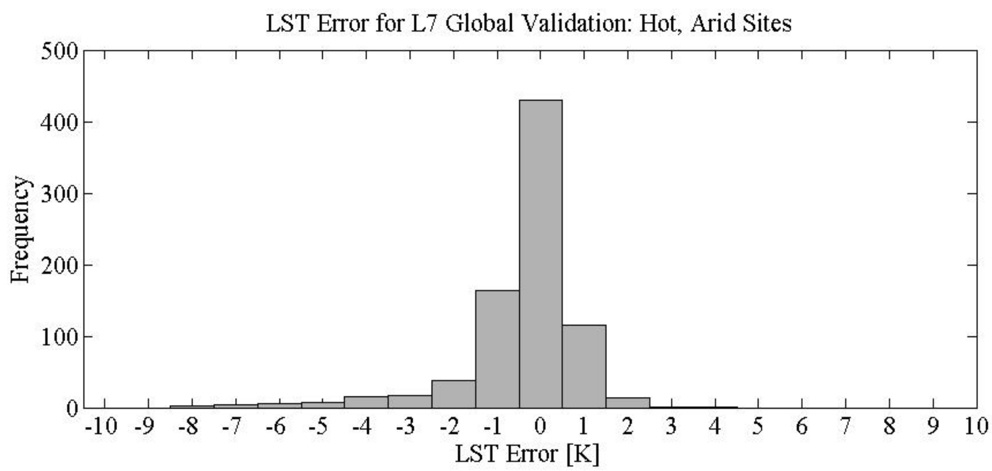

Figure 5.39: LST error histogram for global validation scenes with a "hot/dry" climate.

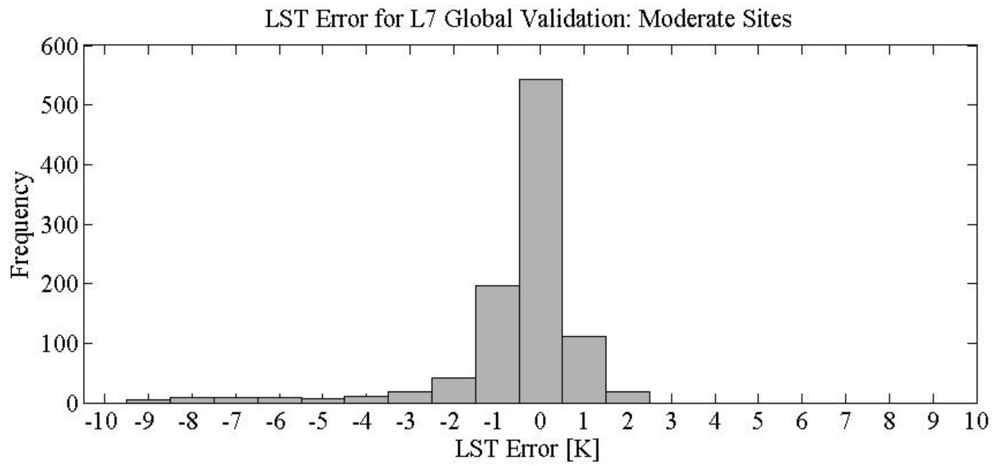

Figure 5.40: LST error histogram for global validation scenes with a "moderate" climate.

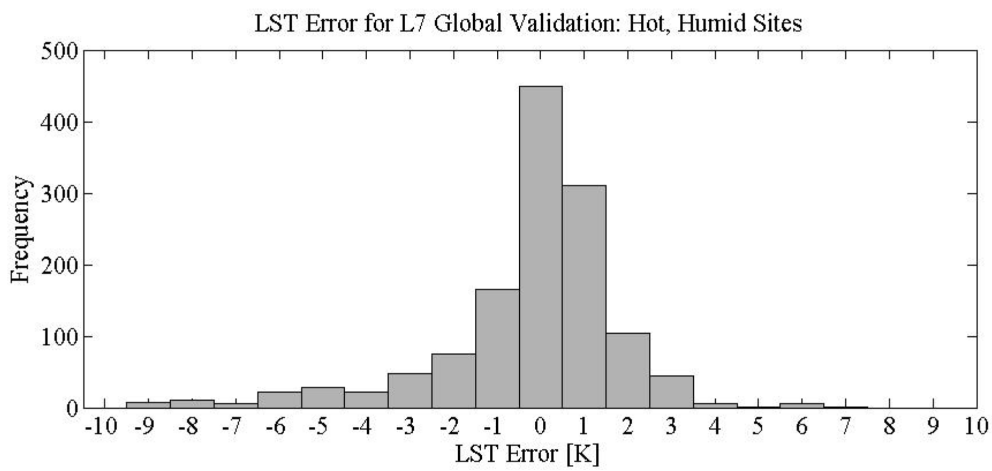

Figure 5.41: LST error histogram for global validation scenes with a "hot/humid" climate.

Table 5.15: LST error statistics for the global validation data divided into four different climate types.

Climate Type	Avg. Error [K]	St. Dev. [K]	RMSE [K]	Samples	Percent
Cold	-0.4346	1.3795	1.44	104	3.25
Hot & Dry	-0.3871	1.3421	1.396	814	25.47
Moderate	-0.5164	1.5424	1.6258	973	30.44
Hot & Wet	-0.2112	2.0706	2.0805	1305	40.83

5.5.4 Validation results using cloud proximity

In the validation of Landsat 5, we have shown the practicality of using the distance to nearest cloud metric to sort LST errors (with the goal of being able to predict LST error from cloud proximity). Since we have expanded our dataset to include global sites that are outside North America, we felt that we should briefly present the LST errors for each cloud distance bin. Note that we have assigned new cloud bin edges, because they make more sense for our much larger dataset. Previously, the bins were 0 - 1 km, 1 - 4 km, 4 - 7.5 km, 7.5 - 50 km, and 50 - infinity km. Now the cloud bins are defined as 0 - 1 km, 1 - 5 km, 5 - 10 km, 10 - 40 km, and 40 - infinity km.

Figure 5.42 shows the LST errors that fall into each of the five newly defined cloud distance bins. Once again, we see that the histograms improve in shape and become more narrow as we look at cloud distance ranges that are farther away from the pixel of interest. Table 5.16 shows the statistics for these plots, which confirms that average error, standard deviation, and root mean square error (RMSE) all consistently improve as clouds get farther away. In this manner, we have confirmed the trend that we observed when we first implemented the distance to nearest cloud method in Section 5.3. As we have explored our global validation results, however, we have begun to show that the LST algorithm is also affected by something that cannot be accounted for by looking at cloud proximity. This was initially illustrated in Section 5.5.3, where we grouped the validation data into four generalized climate types. We were able to show that the LST algorithm performed better for more arid climates, and worse for climates with significant moisture in the air. As previously stated, we do not think that using climate types to help estimate LST errors is wise. Instead, we will investigate if transmission can be used to sort the LST errors, because it is a much more feasible metric to work with. This analysis is presented in the next section.

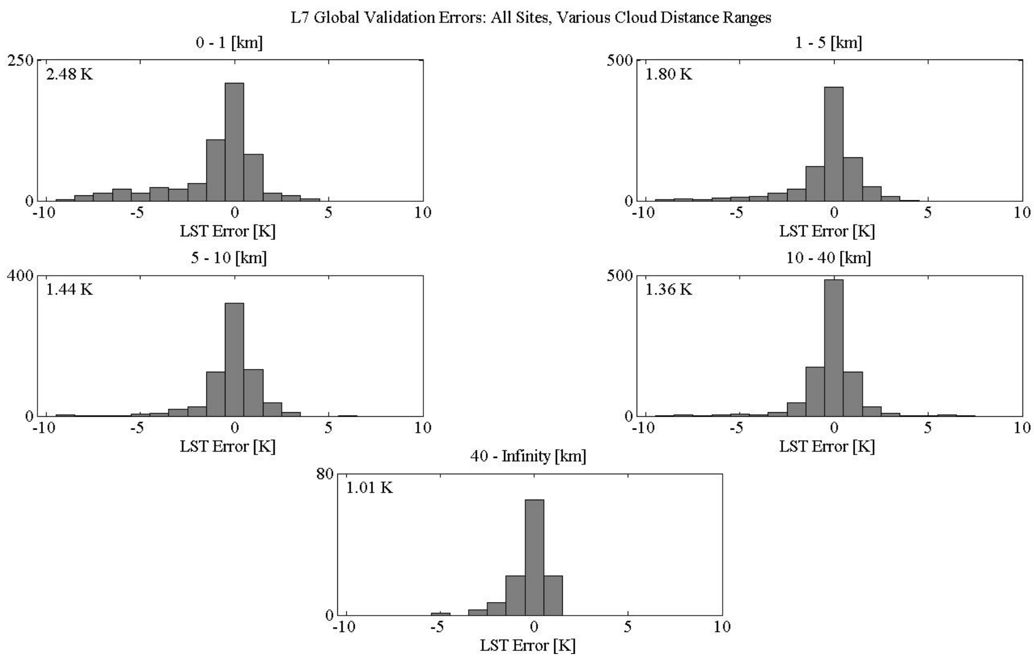

Figure 5.42: Global validation errors for various cloud distance ranges. The RMSE's for each group is located in the top left of each plot. This figure illustrates how the observed LST error histogram consistently improves as clouds are farther away.

Table 5.16: Global validation statistics for each distance to cloud bin, confirming that LST errors improve as clouds get farther away.

Cloud Distance	Avg. Error [K]	St. Dev. [K]	RMSE [K]	Samples	Percent
0 - 1 km	-1.0109	2.3847	2.5881	565	17.68
1 - 5 km	-0.3268	1.8019	1.8303	871	27.25
5 - 10 km	-0.1532	1.4313	1.4384	693	21.68
10 - 40 km	-0.1555	1.3564	1.3646	947	29.63
40 - inf km	-0.2429	0.9850	1.0106	120	3.75

5.5.5 Validation results using transmission

With the distance to nearest cloud method, we have shown that we can better define LST errors by sorting them into different ranges of cloud proximity. This knowledge will aid in our effort to be able to quantify LST uncertainty. Our initial global validation efforts have revealed, however, that climate type may also be influencing the algorithm's performance. In other words, our ability to accurately compensate for the atmosphere changes with different climate conditions. Since trying to relate LST error to climate types is not ideal, we have decided to see if atmospheric transmission can be used instead. We chose to bin the LST errors into different ranges of transmission values, similar to what was done with the distance to nearest cloud bins. Four transmission bins were defined as follows: 0.3 - 0.55, 0.55 - 0.7, 0.7 - 0.85, and 0.85 - 1.0. These ranges were manually chosen so that trends in the LST errors could be easily observed, while also having an adequate number of samples in each bin. The smallest transmission value was set to 0.3 because there are no instances in the global validation dataset that go below this number.

Figures 5.43 shows the LST error histograms for the four transmission bins. The statistics for these charts can be found in Table 5.17. The histograms show clearly that the LST errors gradually tighten around 0 K as transmission increases, which is very encouraging.

Grouping the data by transmission values instead of climate types appears to be an appropriate change. The climate types provided a general sense of how the LST algorithm performs across the globe, but weather at any given site does not always adhere to its climate type. For example, places with hot and arid climates can occasionally experience days of rain or significant moisture in the air. Using transmission can circumvent some of these issues, because we have observed that the LST algorithm performs the most poorly for low transmission values regardless of climate type.

We have observed an encouraging relationship between transmission and LST error, but these results did not include any cloud proximity information other than excluding "0 distances" (where a cloud is directly over a pixel of interest) and cases where LST error was less than -10 K. Therefore, the next logical step is to sort the global validation data to both cloud distance bins as well as transmission bins. The goal of this is to see if we can improve our description of how the LST algorithm performs under these various conditions (i.e. the goal is to produce narrower error histograms). Section 5.5.6 presents the results of this new effort.

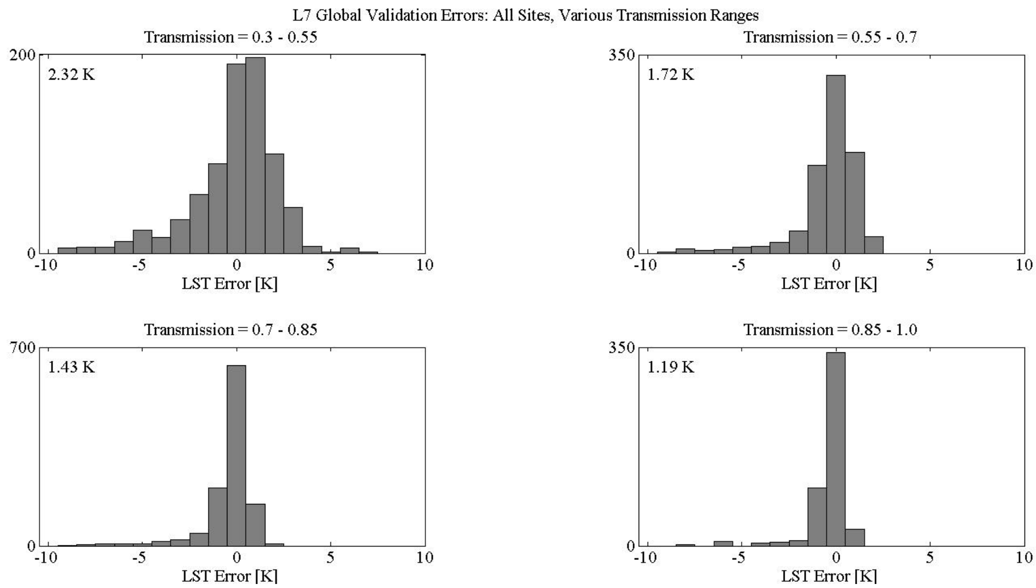

Figure 5.43: Global validation errors for various transmission ranges. The RMSE's for each group is located in the top left of each plot. This figure illustrates how the observed LST error histogram consistently improves as transmission increases.

Table 5.17: Global validation errors for various transmission ranges. The RMSE's for each group is located in the top left of each plot. This figure illustrates how the observed LST error histogram consistently improves as transmission increases.

Transmission	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
0.3 - 0.55	-0.109	2.322	2.323	799	25.00
0.55 - 0.7	-0.438	1.727	1.780	783	24.50
0.7 - 0.85	-0.447	1.437	1.505	1110	34.73
0.85 - 1.0	-0.420	1.117	1.192	504	15.77

5.5.6 Validation results with both clouds and transmission

Section 5.5.4 showed how the LST algorithm performs for different distance to nearest cloud bins, and Section 5.5.5 showed how transmission bins can also be used to help define LST performance. Now we want to utilize both of these metrics to improve our understanding of LST errors for various conditions. We divided the global validation data into 20 bins: 4 transmission bins and 5 cloud bins (the definitions of these bins have already been explained in the previous two sections). The validation data will still be depicted using histograms, but we will arrange them in a matrix layout where each row is one of the transmission bins, and each column is one of the cloud distance bins.

Figure 5.44 shows the LST error histograms for each of the 20 different combinations of transmission and cloud distance bins, and the RMSE values are displayed in the corner of each histogram. The full statistics for this plot are located in Tables 5.18 through 5.19, and a summary of the RMSEs are shown in Table 5.22. This grid of histograms is a very convenient way to view the LST errors in one graph, since it allows one to easily see the changes as transmission increases and as cloud distance increases. Figure 5.45 shows how many samples and the percent of samples present in each of the histograms within the matrix, which is an important factor to consider when interpreting these results. For example, when transmission is relatively low (0.3 - 0.7) and clouds are more than 40 km away, there are very few samples being plotted. The upper right plot in the matrix only has 4 samples that all have large negative errors, which is unusual for cases where clouds are more than 40 km away. Most likely, they represent cases where the cloud mask failed to report clouds that were near the pixel of interest (this phenomenon was fully described in Section 5.3).

One important observation to make is that for the last three columns and last two rows, the RMSE is around 1 K or less for each histogram. This represents about 30% of the entire global validation. The last four columns and last three rows have RMSE values less than 2 K, and represent around 62% of the data. This is very encouraging because it means that users would be able to utilize a significant portion of all the pixels with a reasonable amount of confidence in the accuracy of the LST retrievals for those pixels. Our goal is to be able to estimate uncertainty for these various transmission and cloud groups, so that we can provide users with uncertainty with every LST retrieval, rather then providing them with the general accuracy information that we have observed in this section. The uncertainty estimation results will be discussed in the next section.

Another thing we wish to point out is that there is still a slight negative bias in the data, which we can see from the statistics tables. If we ignore the cases where clouds are between 0 and 1 km, and the few cases where transmission is greater than 0.85 and clouds are more than 40 km, the weighted average error is -0.211 K. Remember that this is after we have already shifted the validation data by a total of 0.485 K (0.241 K and 0.244 K). This average error value is statistically different, so one option would be to shift the LST retrievals by this value as well, but for now we will leave it in since there is not a study behind it that explains where this slight bias comes from.

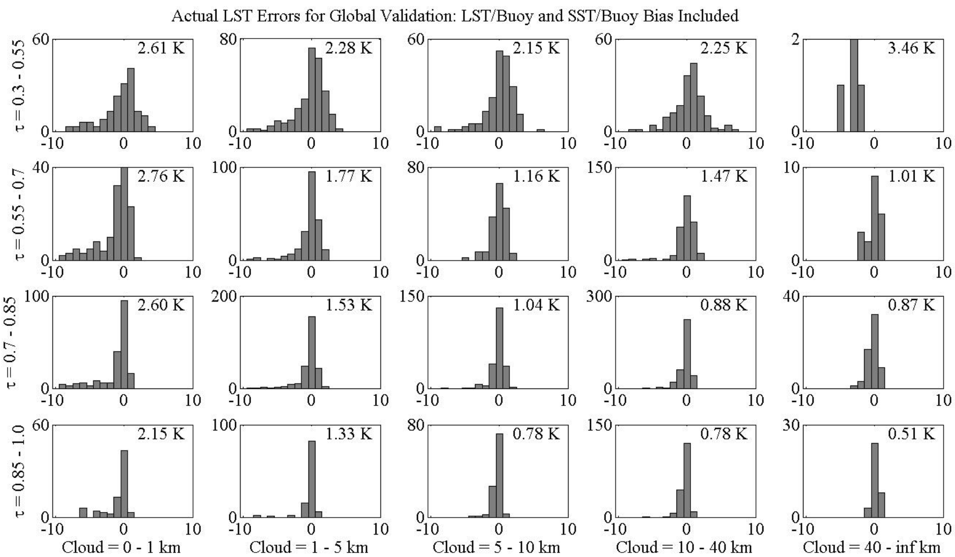

Figure 5.44: Global validation error histograms assembled in a 5 x 4 grid, where each plot in the grid corresponds to one of the 20 combinations of cloud bins and transmission bins. RMSE values are located in the top right of each plot.

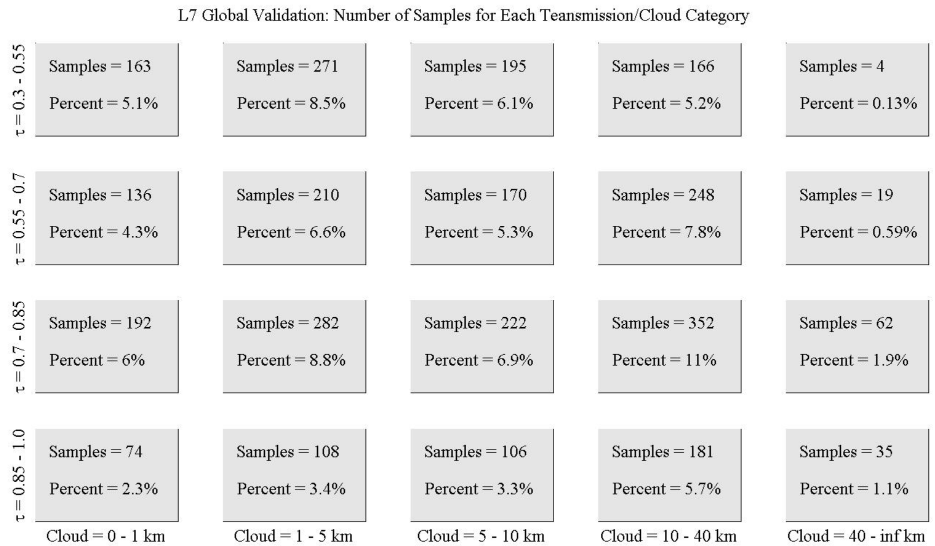

Figure 5.45: Number of samples for global validation error histograms assembled in a 5 x 4 grid, where each plot in the grid corresponds to one of the 20 combinations of cloud bins and transmission bins.

Table 5.18: Global validation error statistics for data where transmission is between 0.3 and 0.55.

Transmission = 0.3 - 0.55					
Cloud Distance	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
0-1 km	-0.520	2.566	2.611	163	5.10
1-5 km	-0.179	2.277	2.280	271	8.48
5-10 km	0.013	2.155	2.150	195	6.10
10-40 km	0.342	2.234	2.254	166	5.19
40-inf km	-3.322	1.105	3.457	4	0.13

Table 5.19: Global validation error statistics for data where transmission is between 0.55 and 0.7.

Transmission = 0.55 - 0.7					
Cloud Distance	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
0-1 km	-1.380	2.398	2.759	136	4.26
1-5 km	-0.366	1.738	1.772	210	6.57
5-10 km	-0.158	1.157	1.165	170	5.32
10-40 km	-0.194	1.462	1.471	248	7.76
40-inf km	-0.188	1.018	1.008	19	0.59

Table 5.20: Global validation error statistics for data where transmission is between 0.7 and 0.85.

Transmission = 0.7 - 0.85					
Cloud Distance	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
0-1 km	-1.166	2.329	2.599	192	6.01
1-5 km	-0.429	1.468	1.527	282	8.82
5-10 km	-0.180	1.023	1.036	222	6.95
10-40 km	-0.269	0.841	0.881	352	11.01
40-inf km	-0.273	0.828	0.866	62	1.94

Table 5.21: Global validation error statistics for data where transmission is between 0.85 and 1.0.

Transmission = 0.85 - 1.0					
Cloud Distance	**Avg. Error [K]**	**St. Dev. [K]**	**RMSE [K]**	**Samples**	**Percent**
0-1 km	-1.010	1.909	2.148	74	2.32
1-5 km	-0.356	1.292	1.334	108	3.38
5-10 km	-0.395	0.671	0.776	106	3.32
10-40 km	-0.339	0.704	0.780	181	5.66
40-inf km	0.133	0.496	0.506	35	1.10

Table 5.22: RMSEs for the L7 global validation errors, sorted by transmission and cloud proximity.

RMSES [K] for actual LST errors		Transmission Ranges			
		0.3 - 0.55	0.55 - 0.7	0.7 - 0.85	0.85 - 1.0
Cloud distance	0 - 1 km	2.611	2.759	2.599	2.148
	1 - 5 km	2.280	1.772	1.527	1.334
	5 - 10 km	2.150	1.165	1.036	0.776
	10 - 40 km	2.254	1.471	0.881	0.780
	40 - inf km	3.457	1.008	0.866	0.506

5.6 Analysis of LST Uncertainty

In the Methodology chapter, we presented a simplified expression for LST uncertainty (Equation 4.8). Through our analysis of the global validation, however, we have observed how we can use transmission in conjunction with cloud proximity to better quantify LST uncertainties. Therefore, we will investigate whether the standard propagation method is sufficient or if using transmission/cloud distance information can add any additional gains. Section 5.6.1 will go through the terms that contribute to standard error propagation and discuss how it does not sufficiently captures LST uncertainty. Section 5.7 will present our efforts to quantify an "unknwon" error/uncertainty that is not accounted for by the standard method.

5.6.1 Standard Error Propagation Method

In Chapter 4, we illustrated how each of the uncertainty terms in Equation 4.7 could be estimated or evaluated. Now that we have a large set of global validation data, we can calculate S_{LST} for each sample and see how well we are able to "predict" the error in the LST algorithm. We also want to observe the contribution that each uncertainty term has on the overall predicted error, although the "error due to emissivity" is currently excluded because the validation data consists solely of water pixels (where emissivity uncertainty is essentially zero).

Effective Error due to the Atmosphere

Using the atmospheric perturbation method described in Section 4.6.1, we were able to calculate the uncertainty/error due to the atmosphere for every global validation point. This process also allowed us to evaluate S_P, the cross correlation terms. The cross correlation terms are directly related to the atmospheric error terms, so we will first observe their separate contributions and then their combined effect. Figure 5.46 and 5.47 show the separate contributions for S_A and S_P, which are plotted against cloud distance and color-coded for the different transmission ranges. Also note that we are displaying the error magnitudes in units of Kelvin, so the values are more intuitive to interpret. These plots show that transmission is a major influence on the calculation of these terms, for as transmission decreases the uncertainty in the atmospheric compensation process increases. Additionally, the cross correlation contribution becomes more negative as transmission lowers.

Figure 5.48 shows the "effective error due to the atmosphere," which shows the magnitude of errors for S_A when the cross correlation values were included (The correct way to evaluate this is $\sqrt{S_A^2 + S_P}$, where S_P tends to be negative). From this graphic we are able to see that for high transmission values (0.85 - 1.0), the error introduced ranges from 0 - 1 Kelvin. The most significant errors are for the low transmission bin (0.3 - 0.55), at around 3.5 Kelvin. In the calculation of S_{L_T}, the error terms are expressed in units of radiance, which in our case are often less than 1 $[W/m^2sr^{-1}\mu m]$. We can expect, therefore, that when the term S_A is squared it will become very small. As we explore the contributions of the other error terms, we will determine if there is a dominant source of error.

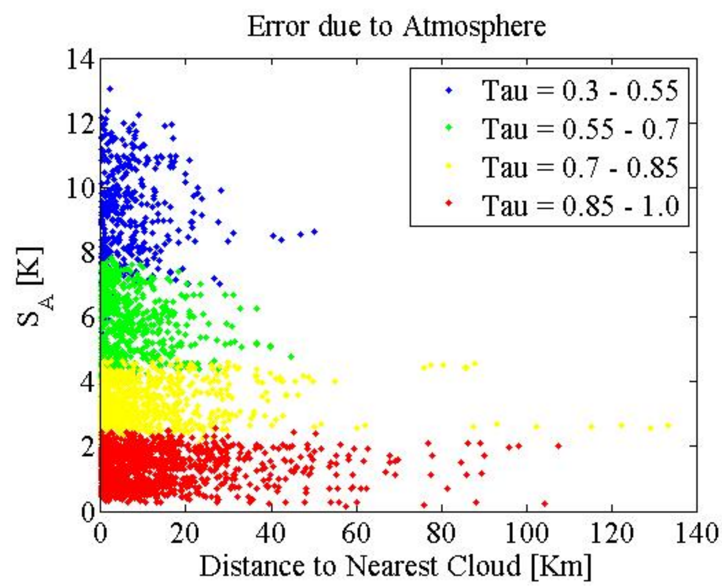

Figure 5.46: Error magnitudes for S_A, the error due to the atmospheric compensation process. These are plotted against the distance to nearest cloud values, and color-coded by transmission bin.

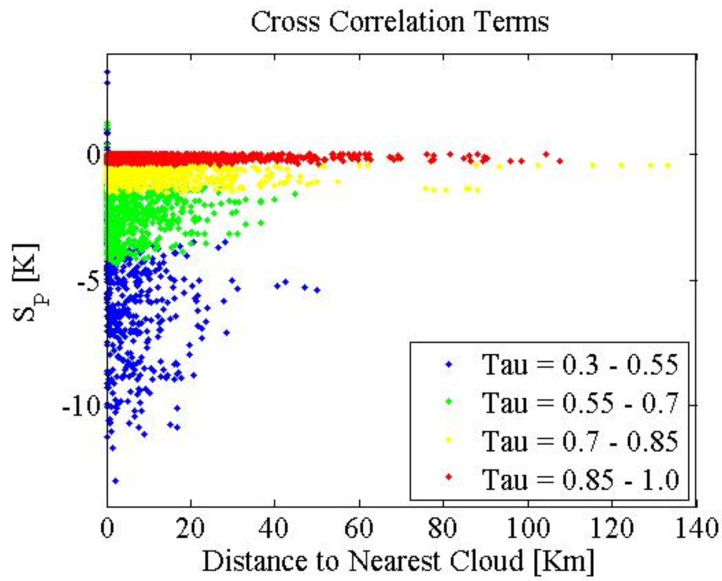

Figure 5.47: Magnitudes for S_P, the cross correlation terms associated with S_A. These are plotted against the distance to nearest cloud values, and color-coded by transmission bin.

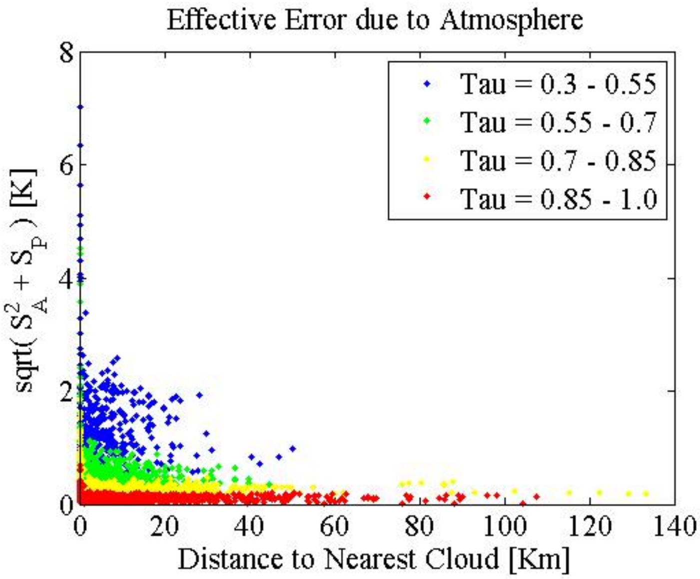

Figure 5.48: Error magnitudes for $S_{A_{eff}}$, the effective error due to the atmospheric compensation process. These are plotted against the distance to nearest cloud values, and color-coded by transmission bin.

Error due to the Instrument

The definition of S_I includes the the term $S_{L_{obs}}$, which is the uncertainty in the Landsat sensor measurements (see Equation 4.8). Since we are currently dealing with Landsat 7, we will use the Noise-Equivalent Differential Temperature (NEΔT) of 0.28 K at 280 K [Barsi et al., 2005]. Figure 5.49 shows the evaluation of S_I for every validation point, which again is plotted in units of Kelvin and color-coded by transmission bins. This graphic shows that the Error introduced by the instrument is very low (0.2 - 0.6 K) across all transmission values. The reason that S_I is not a constant value is because it includes a partial derivative term that has the transmission variable in it. Compared to the effective error due to the atmosphere, $S_{A_{eff}}$, this source of error is much less dominant.

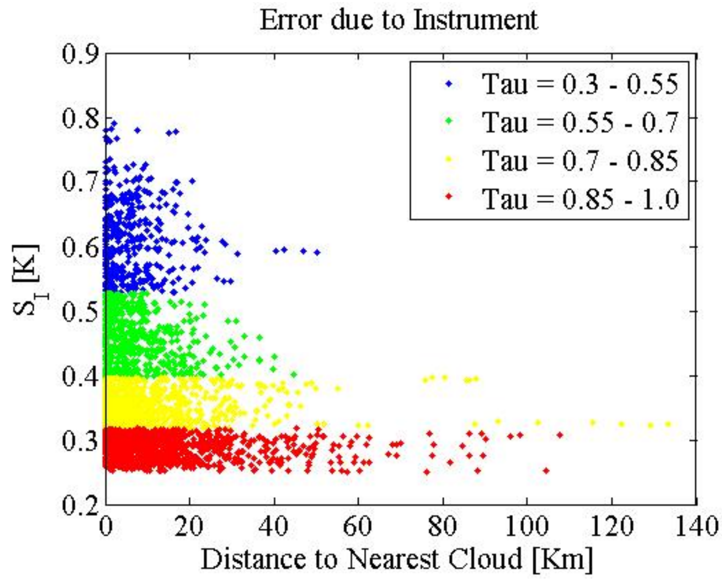

Figure 5.49: Error magnitudes for S_I, the error due to the Landsat instrument. These are plotted against the distance to nearest cloud values, and color-coded by transmission bin.

Predicted LST Errors Compared with Actual LST Errors

The individual contributions of the uncertainty terms in the standard propagation equation have been shown, so now we will look at the total estimated uncertainty. Figure 5.50 shows S_{LST} against distance to nearest cloud, and color-coded by transmission bin. We can see that the majority of estimations of LST error are less than 2 K, although there are a few low transmission cases that go above 2 K. Note that this figure include samples that have a cloud distance of zero, to show that the standard error propagation method is somewhat responsive to areas with large amounts of column water vapor (which is often accompanied by clouds and low transmission). Since we have seen firsthand that observed LST errors can often be extremely negative when clouds are nearby, we expect that these predictions will perform poorly in such situations. Figure 5.51 shows the LST error predictions as a grid of histograms for each cloud and transmission bin. Although the standard error propagation method does not directly use the cloud proximity and transmission bins to predict LST error, it is still useful to analyze the predictions this way because it will be easily comparable to the results that will be generated when these bins are used. We can see that for all 20 plots in Figure 5.51 the predicted LST error is mostly between 0 and 1 K, regardless of transmission and cloud conditions. Although many of the actual LST errors that we observed in Figure 5.44 had errors within this range, there were also many LST errors that were more extreme than this. Plainly put, the standard error propagation method could be adequate for ideal atmospheric conditions (high transmission, no clouds), but it is not sufficient for inferior conditions. Therefore, we will try to quantify the remaining error using our defined transmission and cloud groups, which we expect will improve our ability to estimate LST uncertainty.

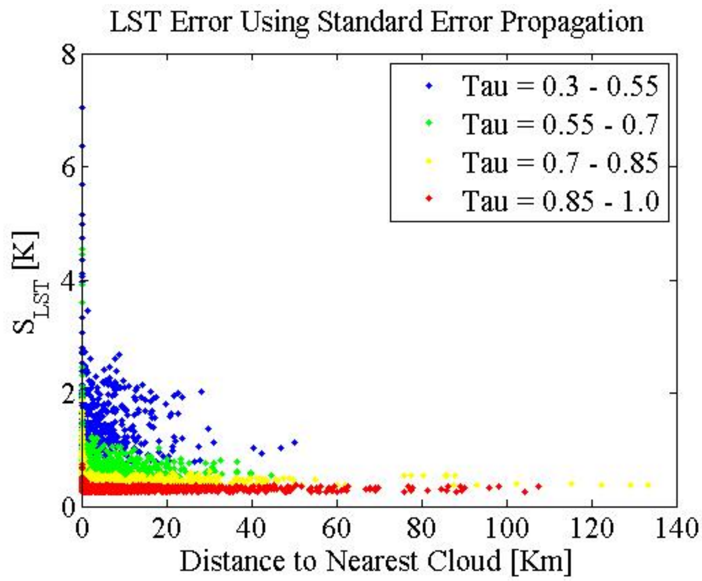

Figure 5.50: Predicted LST errors using standard error propagation, plotted against cloud distance and color-coded by transmission bins.

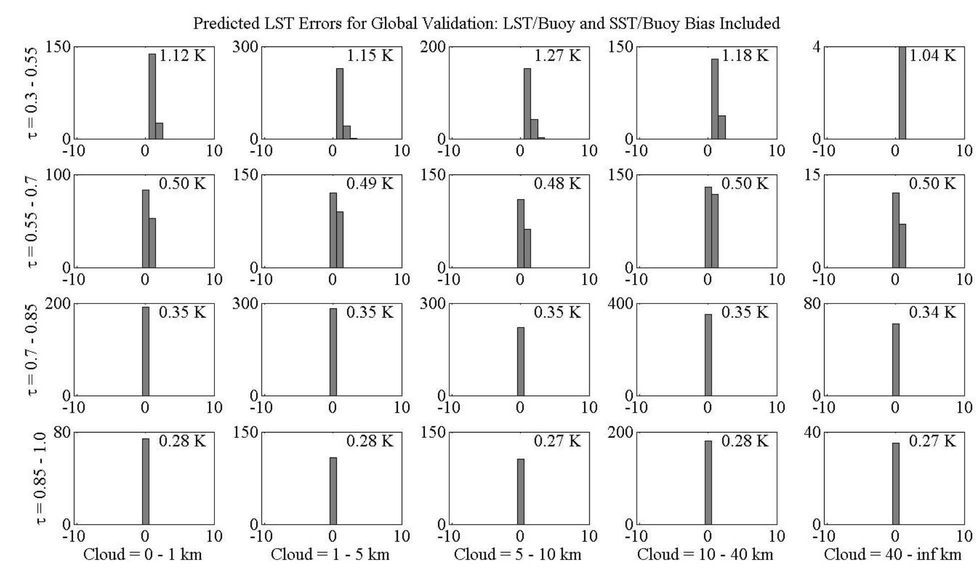

Figure 5.51: Predicted LST error histograms using standard error propagation.

5.7 Computing Unknown Error

In the previous section, we showed that the standard error propagation did not fully capture the errors that we observed from the global validation results. Although LST error histograms were consistently centered near zero, there was a varying amount of spread based on transmission levels and cloud proximity. This most likely means that MERRA'a atmospheric profiles are not as reliable for cases where either transmission is low or clouds are in the vicinity. Since we have statistics related to LST error and our predicted error/uncertainty, we can attempt to solve for the remaining error. Equation 5.2 shows how we calculated this remaining error, which we are calling "unknown" error because although it is being defined using transmission and cloud proximity levels, it is not fully understood what source of error is at play. As we have mentioned, it is plausible that MERRA becomes untrustworthy under certain atmospheric conditions, but it is virtually impossible to prove because reanalysis products such as this do not report uncertainties in the profiles. Calculating this "unknown error" gave us a value for every transmission/cloud category (20 values total). Using these values, we performed a bilinear interpolation so that we obtained the unknown error for every sample in our global validation set. Table 5.23 shows the 20 calculated values that were used for the bilinear interpolation.

We have mentioned before that the top right plot from the validation results (where transmission is 0.3 - 0.55 and clouds are more than 40 km away) is not trusted because it only has 4 samples and extreme LST retrievals. Therefore, we chose to infer the unknown error for this group by interpolating from surrounding values. This will prevent LST uncertainties from being exaggerated in the estimation process.

$$\text{Unknown Error} = \sqrt{(\text{observed RMSE})^2 - (\text{standard predicted error})^2} \tag{5.2}$$

Table 5.23: Unknown errors that were calculated for the L7 global validation dataset. Unknown errors are computed using Equation 5.2. For the bin where transmission is between 0.3 and 0.55 and clouds are more than 40 km away, the unknown error was interpolated from the surrounding values because the validation data is not trusted from that category.

Unknown Error Quantities [K]		Transmission Ranges			
		0.4 - 0.55	0.55 - 0.7	0.7 - 0.85	0.85 - 1.0
Cloud distance	0 - 1 km	2.391	2.715	2.576	1.4138
	1 - 5 km	2.016	1.703	1.487	1.305
	5 - 10 km	1.816	1.062	0.976	0.726
	10 - 40 km	1.972	1.385	0.811	0.730
	40 - inf km	3.299	0.875	0.795	0.427

Once unknown errors were calculated for the entire global validation dataset, we used Equation 5.3 to obtain the total estimated LST uncertainties. Figure 5.52 shows histograms for this new LST uncertainty, where the horizontal scale ranges from 0 - 5 K since this value cannot be negative. This shows what we would expect, that the uncertainties tend to be higher in situations where transmission is low and/or clouds are nearby.

In the final version of the Landsat Surface Temperature product, there will include a band for the LST image and a band for the LST uncertainty image. A user of the product would utilize these two bands to get LST $\pm$ uncertainty, which would help them decide which data points to used based on how certain they need the LST values to me. As an extra visual aid, Figure 5.53 shows the LST retrievals (in blue), SST values (in black), and uncertainty in LST that is plotted as a shaded region (in red) around the LST retrievals. The y-axes is in Kelvin, and the x-axes is simply the sample indexes (the data was sorted from minimun LST to maximum LST to make visualization easier). If our uncertainty estimates are adequate, then the SST lines should always fall within the shaded bounds. The vast majority of the SST samples do indeed fall within the bounds, keeping in mind that top right plot only has 4 samples and is not trusted. As transmission increases and clouds get farther away, the uncertainty bounds get tighter around the LST line, but the

SST line still continues to stay within the bounds.

In the evaluation of the global validation results, we saw that about 30% of the data had RMSEs less than 1 K, and 62% had RMSE's less than 2 K. When we consider the uncertainties in LST that we plan to include in the final product, we should note that 20% of the dataset has reported uncertainties less than 1 K, and 63% has uncertainties less than 2 K.

$$\text{LST Uncertainty} = \sqrt{(\text{interpolated unknown error})^2 + (\text{standard predicted error})^2} \quad (5.3)$$

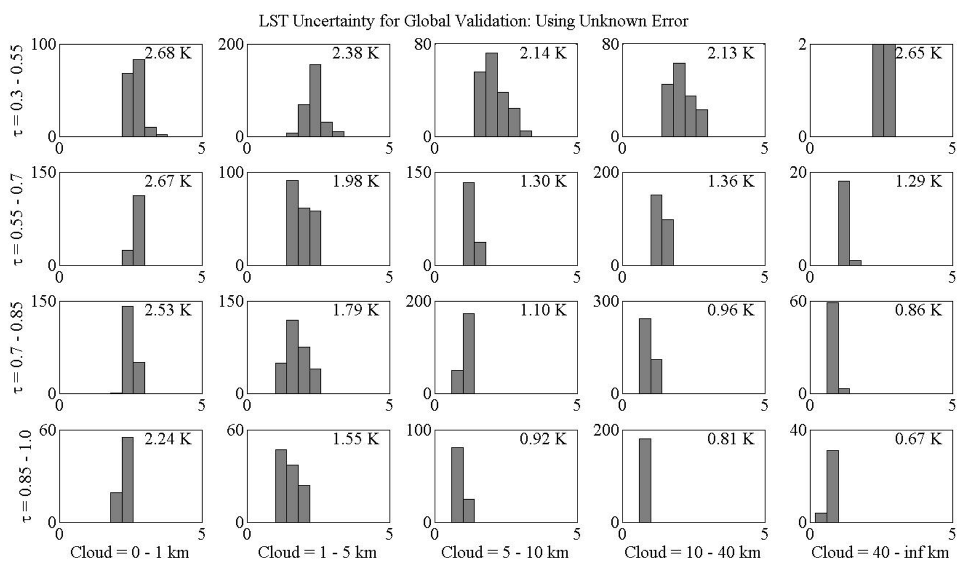

Figure 5.52: Total LST uncertainty for the L7 global validation set. The total uncertainty includes predicted error via standard error propagation and "unknown" or unaccounted for error, calculating using Equation 5.2. The RMSE values are located in the top right of each plot. The top right plot is not trusted because it only has 4 samples.

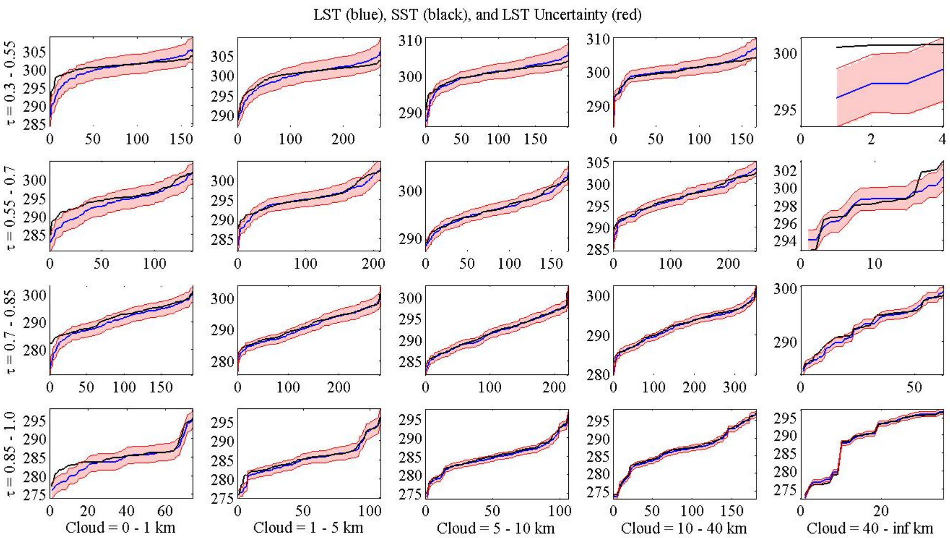

Figure 5.53: LST retrievals are plotted in blue, SST values are plotted in black, and the uncertainty range is shaded red. Each plot is sorted from min LST to max LST, therefore the x axis is the sample index. The truth (SST) is almost always within the uncertainty bounds, and the bounds get tighter for better atmospheric conditions.

5.8 Example of LST Product for a Full Scene

From our validation studies we have seen that the Landsat Surface Temperature algorithm performs very well in conditions where clouds are far away and transmission is high. In section 5.6, we showed that standard error propagation was not sufficient for estimating LST uncertainty. We introduced "unknown error," which was the portion of the observed LST error that was not explained by standard error propagation. Using this unknown error in addition to the standard error propagation proved to be a good method for estimating LST uncertainty. Now, we wish to go through an example Landsat 7 scene and show what an actual LST image looks like, as well as what the LST uncertainty image looks like.

The example Landsat 7 scene we will be using as an example is LE70160382011313EDC00. Its geographical location is off the coast of the state of Georgia. Figure 5.54 shows the true color image for our example, and Figure 5.55 shows the thermal band image. We should expect to see that in the LST image, the land tends to be hotter, and the clouds are much cooler. Indeed, we see exactly that in Figure 5.58.

In order to calculate LST uncertainty we will need to utilize the cloud mask product, so the cloud mask for our example is shown in Figure 5.57. Our uncertainty estimation process requires distance to nearest cloud, but for the validation studies it was only necessary to make the calculation for one point in the image at a time. In the full implementation, an image of distances to nearest cloud for each pixel will need to be generated. This is a very computationally heavy endeavor, so for the sake of this example this image was generated at a tenth of the resolution of the original Landsat scene. This resulting image can be seen in Figure 5.56. Notice that the original cloud mask identified a few small spots along the coast as clouds, but there does not appear to clouds in the true color or thermal image. This will affect what the LST uncertainty band reports, but a user that is manually choosing areas to use will be able to notice this and choose data points accordingly.

Finally, Figure 5.59 shows the LST uncertainty image. As we would expect, the LST uncertainty is highest where there are clouds, and then the uncertainty decreases for areas farther away. There are some block-like artifacts, which is partly due to the downsampled distance to cloud image and partly due to the block shaped nature of the transmission, upwelled and downwelled images. These are not included because they look like gradient-filled boxes, where the corners mark where the reanalysis points are, and the gradient is because of the spatial interpolation.

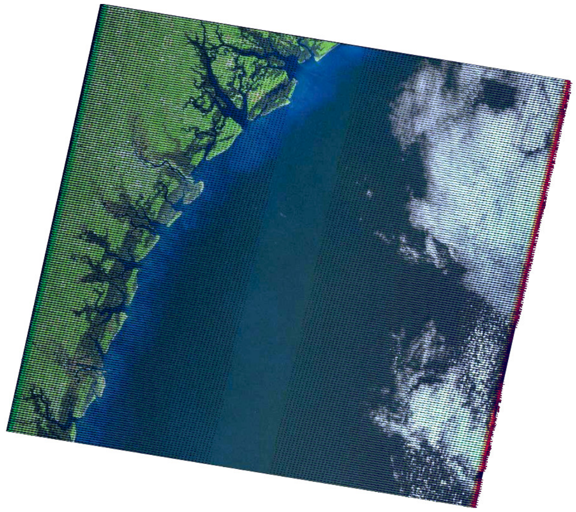

Figure 5.54: The true color image for Landsat scene LE70160382011313EDC00.

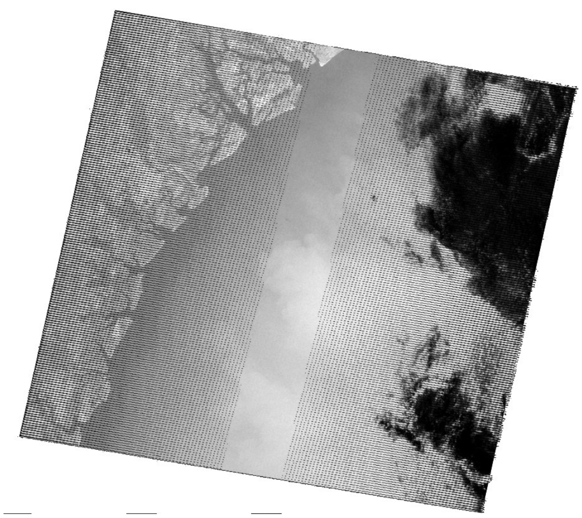

Figure 5.55: The thermal image (band 6) for Landsat scene LE70160382011313EDC00.

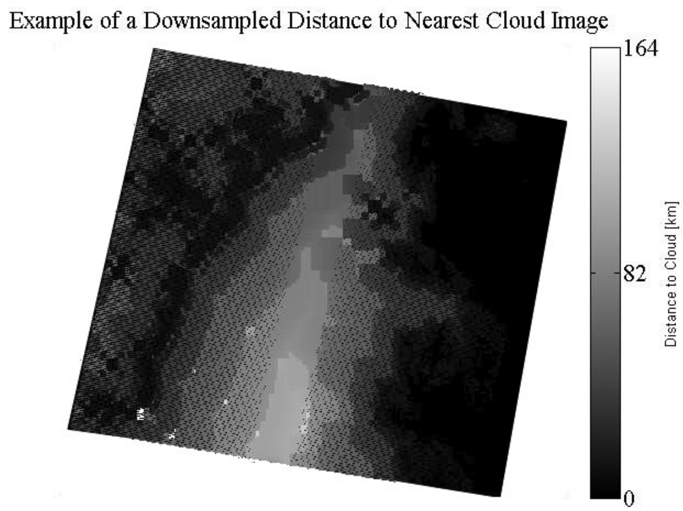

Figure 5.56: The distance to nearest cloud image for Landsat scene LE70160382011313EDC00. This was downsampled by a factor of 10 in order to make computation time more reasonable, which is the reason for the block-like texture.

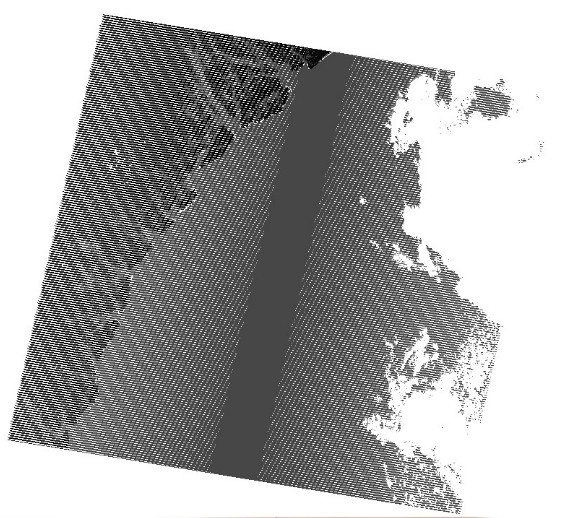

Figure 5.57: The cloud mask product for Landsat scene LE70160382011313EDC00.

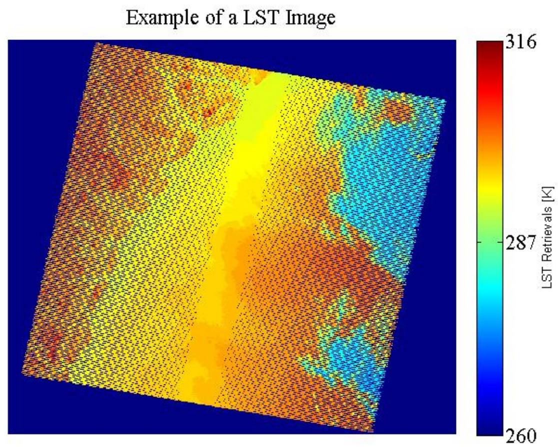

Figure 5.58: The LST image for Landsat scene LE70160382011313EDC00. The land tends to have higher surface temperatures, while clouds tend to have low temperatures.

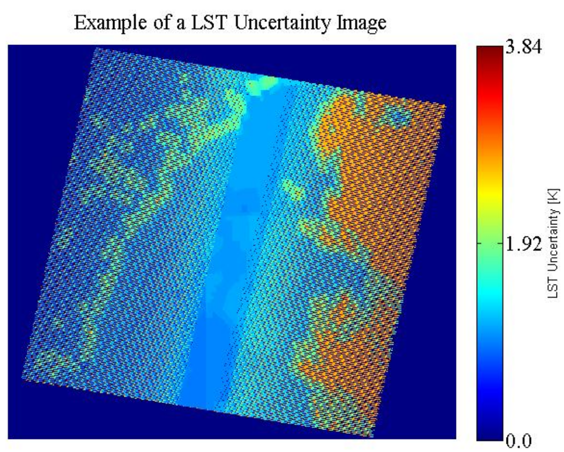

Figure 5.59: The LST uncertainty image for Landsat scene LE70160382011313EDC00. It is clear that the cloudy areas have the highest uncertainty, but the uncertainties decrease significantly in areas far away from the clouds.

5.9 Concluding Remarks

This chapter, stated briefly, presented Landsat 5 validation studies, Landsat 7 global validation results, LST uncertainty analysis, and bias removal studies. For Landsat 5, we were able to validate the LST algorithm for North American locations, and we quantified how well the algorithm performed for different ranges of cloud proximities. From this validation set, the average LST error ranged from -3.424 K (when clouds were between 0 and 1 km away) and -0.262 K (when clouds were greater than 50 km away). For the validation of Landsat 7, we were able to alter our source of reanalysis data and truth data in order to perform studies of LST error on a global scale. This revealed that the LST algorithm performance is generally worse if transmission levels are low, and generally better if they are high. By using cloud proximity information as well as transmission values, we were able to sort the global validation data into 20 different bins and observe the LST errors. This showed that 30% of the dataset had RMSEs less than 1 K, and 62% of the dataset had RMSEs less than 2 K.

We were also able to develop a method of estimating LST uncertainty, which combines standard error propagation calculations with the remaining error that was observed from the validation results. This proved to be an adequate method, where 20% of the dataset reported uncertainties less than 1 K, and 63% of the dataset had uncertainties less than 2 K. All of our results have shown very encouraging signs that the LST algorithm is very accurate under desirable atmospheric conditions, and we have seen that we have a reliable way to estimate uncertainty in the LST retrievals.

To sum up everything that has been accomplished to date, Table 5.24 lists the tasks completed by Cook, the initial investigator, and the new studies done by the current investigator. The progress in terms of validation is significant, but we are especially proud of our LST uncertainty estimation method. There were also many parts of the process that were automated, so that the code can be easily and swiftly implemented by USGS.

Table 5.24: List of tasks completed by Cook and Laraby, where Cook was the initial investigator and Laraby is the current one.

Cook	Laraby
Landsat 5 Validation	
1. LST validation 826 N.A. scenes including clouds, using NARR	1. Compared LST errors when NARR and CFSR were used for 130 N.A. scenes
2. Compared LST errors when NARR and MERRA were used for 397 N.A scenes	
Landsat 7 Validation	
1. MODIS SST vs buoys and LST with NARR vs buoys (60 N.A. scenes)	1. MODIS SST vs buoys and LST with NARR vs buoys (118 N.A. scenes)
2. LST validation using MERRA, and using MODIS SST as truth (63 global scenes)	2. LST validation using MERRA, and using MODIS SST as truth (3081 global samples)
Confidence Metric	
1. Categorized cloud types manually for 826 Landsat 5 scenes	1. Performed cloud distance analysis for 949 scenes (827 Landsat 5, 122 Landsat 7)
2. Performed preliminary investigation into the use of standard error propagation for the LST algorithm	2. Established uncertainty estimation method, which gives acceptable levels for most conditions of interest (i.e. high transmission and removed from clouds)
Process Automation	
1. Developed LST code that downloads reanalysis data and determines the atmospheric parameters at Landsat resolution	1. Created scripts to calculate cloud distances for single points and entire images.
2. Developed code to calculate surface temperature at a specific point in Landsat scene	2. Automated process of downloading MODIS SST images, georeferencing them, and subsetting them.

Chapter 6

Conclusions and Future Work

Great strides have been made towards the completion of the Landsat Land Surface Temperature Product. Firstly, were able to extend our algorithm to be able to process scenes on a global scale. We also identified the MODIS Sea Surface Temperature product as an acceptable source of ground truth, which allowed us to perform a thorough global validation study for Landsat 7. This dataset was sorted into four transmission bins and five cloud distance bins, which improved our ability to quantify the algorithm's accuracy under various atmospheric conditions. We also demonstrated our ability to estimate the error in the LST retrivals, and we implemented a bias removal technique that minimized LST errors and improved the accuracy of our error estimation method.

In Chapter 2, we listed our objectives for the development of the Landsat Land Surface Temperature product. It would be beneficial to reexamine the completion status of each objective, and also discuss any tasks that have yet to be addressed. Section 6.1 discusses the status of our original objectives, and 6.2 describes the remaining tasks to be completed.

6.1 Current Status of Objectives

1. **Select an appropriate source of atmospheric input variables with adequate spatial and temporal resolution for use with MODTRAN and for any current/archived Landsat scene in North America.**

   When the LST algorithm was first developed by Cook, the North American Regional Reanalysis dataset was used to obtain atmospheric input variables [Cook, 2014]. In our efforts to expand the algorithm's operability to the entire globe, we identified a global reanalysis product that was comparable to NARR. This new reanalysis source is known as the Modern-Era Retrospective Analysis for Research and Applications (MERRA). We have shown in Section 5.4.1 that the errors in the LST retrievals are almost identical regardless of whether NARR or MERRA was used to provide the atmospheric input variables, but our recommendation is to use NARR where available because it has a finer spatial resolution.

2. **Implement an automated method of determining atmospheric parameters on a per-pixel level.**

   The methodology for the Landsat LST algorithm was described in Section 4.1. Essentially, our approach involves inputting reanalysis profiles into MODTRAN and performing simulations at several ground altitudes and three different surface temperatures, which allows us to estimate the parameters transmission, upwelled, and downwelled radiance at each elevation, and also at each profile's spatial location within a Landsat scene. Linear interpolation is used to obtain these three parameters at the true elevations within the Landsat scene, where the true elevations are obtained from Digital Elevation Models (DEMs). Shepard's interpolation method was then used to calculate the atmospheric parameters for every pixel in the Landsat scene. After all these steps, the algorithm provides all the information necessary to calculate surface temperature at a per-pixel level, excluding emissivity. In our validation efforts we were able to use the emissivity of water, but when the final product is implemented it will be combined with the ASTER Global Emissivity Database that was developed at the Jet Propulsion Laboratory.

3. **Validate the process using available truth data for Landsat scenes over North America.**

   This task has already been completed by Cook, but it was included in our list of objectives to provide a complete outline of how we aimed to reach a complete LST product [Cook, 2014]. Around 800 Landsat 5 scenes within North America were processed, and single LST retrievals were compared to buoy-measured surface temperatures to assess the accuracy of the algorithm. We were able to see that the average error in the LST retrievals was -0.267 K for ideal conditions (i.e. clear, cloud-free scenes). There was also a preliminary study that showed how the algorithm performed when different types of clouds were over, near, or far from the pixel being validated. The cloud types and proximities were manually identified for each of the validation points, which was useful for observing general trends but not ideal because of the study's subjective and tedious nature. The conclusions that were drawn from this study were that the algorithm performance is affected more by cloud proximity than cloud type, and that the LST errors can be very extreme (even -10 K or less) when clouds are directly over or nearby the pixel of interest.

4. **Validate the LST process for Landsat scenes on a global scale, and for each Landsat sensor that provides thermal imagery.**

   Studies were performed in order to show that MERRA was an appropriate replacement for NARR, and that the MODIS Sea Surface Temperature product was an adequate source of ground truth (see Sections 5.4.1 and 5.4.2). MODIS is an instrument aboard the satellite Terra, which has a similar orbit to Landsat 7 (it captures images of the same area around 20 minutes after Landsat 7). This allowed us to perform a global validation study for Ladnsat 7, which involved processing several years worth of scenes at 14 different sites across the globe. We divided the validation data into 20 categories in order to observe how the algorithm performed under various conditions. There were four ranges of transmission levels and five ranges of "distance to nearest cloud," which was a metric we developed to help quantify LST errors based on cloud proximity. When we examined groups where clouds were more than 5 km away and transmission was at least 0.55, we saw that the Root Mean Square Error (RMSE) was around 1 K. These groups made up almost 50% of the entire validation set, which showed that the LST retrievals are encouragingly accurate about half the time. Note that we have only performed validations for Landsat 5 and Landsat 7, and have not yet addressed Landsat 4 and 8. Although this will be a useful future endeavor, we expect that the results would be very similar our other validation studies.

5. **Develop a method of predicting overall LST errors on a per pixel level, in order to include a quality map in the final product.**
   The Landsat LST Product would be much more attractive for scientific pursuits if it included a quality band that informed users how accurate each LST retrieval is. We determined that standard error propagation was an insufficient estimate of the error in the LST retrievals, but we were able to improve these estimates by using statistics from the global validation dataset. We quantified an "unknwon error" that accounted for the error was not explained by standard error propagation, and we used it to help calculate the total estimated uncertainty associated with each LST retrieval. We were

able to show that our uncertainty estimation method was very good at characterizing how the LST retrievals can be less trustworthy when transmission is low and clouds are nearby.

6. **Form a set of recommendations for how the LST product should be implemented by USGS, and present a final assessment of the product's expected performance using the recommended approach.**

   Through our global validation of Landsat 7, we showed that the LST retrievals had RMSEs around 1 K or less for about a third of the dataset where clouds were far enough away and transmission was fair. The RMSE's were less than 2 K for about 62% of the data. Our ultimate suggestion would be for the LST retrievals to be biased by 0.241 K, and for SST values to be biased by -0.244 K for any further validation studies. These biases came from the Landsat 5 validation study and the LST SST study. We would also suggest that the quality band should contain the values that our current uncertainty estimation process yields, so that users can use it as a $\pm$ to the LST values. This lends itself useful to people with different accuracy requirements.

6.2 Next Steps for the LST Product

1. **Integrate New Methods into the Automated LST Process**

   The automated scripts for the LST process currently output a stack of five bands that consist of the original thermal band, an elevation band, a transmission band, an upwelled radiance and, and a downwelled radiance band. Now that we have a method of estimating LST error, we can add this to the process so that the output also includes a quality band. The process will also need to be altered so that NARR is automatically used in North American regions and MERRA is used elsewhere. Finally, we will also aim to include the option to produce a band that provides the "distance to nearest cloud" for every pixel.

2. **Perform Basic Validation for Landsat 4**

   Up to this point, most of our validation efforts have been focused on Landsat 5 and 7. It is important to generate results for Landsat 4 as well, so that we can ensure that our process works well for as much of the thermal archive as possible. This should only require a small validation study (50 or so scenes), using North American scenes with available buoy truth. If the LST error histograms are similar to results obtained for the other Landsat sensors, we would be confident that our algorithm works well for Landsat 4 as well.

3. **Extend Validation Efforts to Landsat 8**

   The other Landsat sensor that has not been validated is Landsat 8, which is the most current sensor and it is very different compared to its predecessors. This task was put on hold because banding artifacts were observed in Landsat 8's Thermal InfraRed Sensor (TIRS) imagery. This phenomenon was not present in every scene, but when it existed it could introduce extra signals as high as 8% in band 11, and about half that amount in band 10. A team at the Rochester Institute of Technology has worked closely with NASA to determine the source of the false signals, and they made great efforts to develop a method to correct scenes that were affected [Montanaro et al., 2014]. It is expected that the correction will be implemented and available to the public in the near future. Since Landsat 8 has two thermal bands, validation would need to be performed for each of them.

6.3 Concluding Remarks

This chapter discussed the status of the original objectives that were outlined in Chapter 2, and it listed the tasks that are next in line to be addressed. These future efforts involve adding the quality band calculations to the current process, and performing validation studies for Landsat 4 and Landsat 8. Once the updated process has been handed off to USGS, they will be able to easily implement it and make the ultimate decision on which bands to provide to the user, and if a bias will be applied.

Appendix A

MODTRAN Inputs and Outputs

The LST algorithm currently uses MODTRAN 4 version 3 revision 1 (4v3r1) for the atmospheric compensation process. Now there are newer versions of MODTRAN, but to ensure that our validation results are comparable we have chosen to stay with the version we began with. Additionally, there is no indication that version 4v3r1 has any major issues that would cause our results to be invalid. Input and output files of MODTRAN are called the "card deck," and individual files are called tape files. This terminology originated from the era when punch cards were used, although now the files are now digital. The input file to MODTRAN is known as the tape5 file, and the outputs produced are the tape6, tape7, tape7scn, and tape8 files.

The input tape5 file can either be edited as a text file or using a graphical user interface to define various parameters and settings for a particular MODTRAN run. This file requires a very specific format that must be followed in order for the program to operate correctly, so it is very important to understand exactly how to create and edit a tape5 file. Figure A.1 shows an example of a tape5 file that was used, where each line refers to a "card" that has its own variables and options that need to specified in a specific way. The first four and last five lines make up the different cards, and the lines in the middle are the atmospheric layers for a particular location. We have provided a table for each card that was used in our MODTRAN runs which list each variable, describes each varible, displays the option/input that was used, and a brief explanation of why that input was chosen. These are located in Tables A.2 through A.10. For more information on the card inputs or formatting requirements, see the MODTRAN 4v3r1 manual [Berk et al., 2003].

```
TM  7    3    2    1    0    0    0    0    0    0    1    1    0 000,000   0,10
T   8F    0 360,00000        0         0 F F F         0,000
    1    0    0    0    0    0     0,000      0,000      0,000      0,000      0,213
   47    0    0
     0,213 1,000E+03 2,770E+02 4,952E+01 0,000E+00 0,000E+00AAH
     0,418 9,750E+02 2,752E+02 5,341E+01 0,000E+00 0,000E+00AAH
     0,627 9,500E+02 2,756E+02 4,753E+01 0,000E+00 0,000E+00AAH
     0,844 9,250E+02 2,778E+02 2,466E+01 0,000E+00 0,000E+00AAH
     1,068 9,000E+02 2,792E+02 1,700E+01 0,000E+00 0,000E+00AAH
     1,299 8,750E+02 2,796E+02 1,443E+01 0,000E+00 0,000E+00AAH
     1,536 8,500E+02 2,796E+02 1,249E+01 0,000E+00 0,000E+00AAH
     1,781 8,250E+02 2,792E+02 1,015E+01 0,000E+00 0,000E+00AAH
     2,033 8,000E+02 2,786E+02 6,863E+00 0,000E+00 0,000E+00AAH
     2,292 7,750E+02 2,778E+02 2,863E+00 0,000E+00 0,000E+00AAH
     2,559 7,500E+02 2,766E+02 6,646E-01 0,000E+00 0,000E+00AAH
     2,833 7,250E+02 2,749E+02 5,281E-01 0,000E+00 0,000E+00AAH
     3,115 7,000E+02 2,729E+02 5,516E+00 0,000E+00 0,000E+00AAH
     3,704 6,500E+02 2,687E+02 3,156E+01 0,000E+00 0,000E+00AAH
     4,330 6,000E+02 2,646E+02 4,607E+01 0,000E+00 0,000E+00AAH
     5,001 5,500E+02 2,604E+02 3,914E+01 0,000E+00 0,000E+00AAH
     5,724 5,000E+02 2,558E+02 2,591E+01 0,000E+00 0,000E+00AAH
     6,507 4,500E+02 2,504E+02 2,251E+01 0,000E+00 0,000E+00AAH
     7,362 4,000E+02 2,438E+02 2,257E+01 0,000E+00 0,000E+00AAH
     8,305 3,500E+02 2,369E+02 1,936E+01 0,000E+00 0,000E+00AAH
     9,361 3,000E+02 2,288E+02 2,641E+01 0,000E+00 0,000E+00AAH
     9,940 2,750E+02 2,242E+02 3,781E+01 0,000E+00 0,000E+00AAH
    10,562 2,500E+02 2,198E+02 4,911E+01 0,000E+00 0,000E+00AAH
    11,234 2,250E+02 2,161E+02 4,796E+01 0,000E+00 0,000E+00AAH
    11,985 2,000E+02 2,178E+02 2,989E+01 0,000E+00 0,000E+00AAH
    12,847 1,750E+02 2,193E+02 1,677E+01 0,000E+00 0,000E+00AAH
    13,838 1,500E+02 2,172E+02 1,124E+01 0,000E+00 0,000E+00AAH
    14,993 1,250E+02 2,138E+02 6,343E+00 0,000E+00 0,000E+00AAH
    16,387 1,000E+02 2,104E+02 6,692E+00 0,000E+00 0,000E+00AAH
    17,693 8,475E+01 2,142E+02 3,671E+00 0,000E+00 0,000E+00AAH
    19,000 6,950E+01 2,179E+02 6,500E-01 0,000E+00 0,000E+00AAH
    20,000 5,950E+01 2,192E+02 4,900E-01 0,000E+00 0,000E+00AAH
    21,000 5,100E+01 2,204E+02 3,800E-01 0,000E+00 0,000E+00AAH
    22,000 4,370E+01 2,216E+02 3,000E-01 0,000E+00 0,000E+00AAH
    23,000 3,760E+01 2,228E+02 2,400E-01 0,000E+00 0,000E+00AAH
    24,000 3,220E+01 2,239E+02 1,900E-01 0,000E+00 0,000E+00AAH
    25,000 2,770E+01 2,251E+02 1,500E-01 0,000E+00 0,000E+00AAH
    30,000 1,320E+01 2,337E+02 3,000E-02 0,000E+00 0,000E+00AAH
    35,000 6,520E+00 2,452E+02 1,000E-02 0,000E+00 0,000E+00AAH
    40,000 3,330E+00 2,575E+02 0,000E+00 0,000E+00 0,000E+00AAH
    45,000 1,760E+00 2,699E+02 0,000E+00 0,000E+00 0,000E+00AAH
    50,000 9,510E-01 2,757E+02 0,000E+00 0,000E+00 0,000E+00AAH
    55,000 5,150E-01 2,693E+02 0,000E+00 0,000E+00 0,000E+00AAH
    60,000 2,720E-01 2,571E+02 0,000E+00 0,000E+00 0,000E+00AAH
    70,000 6,700E-02 2,181E+02 0,000E+00 0,000E+00 0,000E+00AAH
    80,000 1,200E-02 1,741E+02 7,000E-02 0,000E+00 0,000E+00AAH
   100,000 1,000E-05 1,905E+02 0,000E+00 0,000E+00 0,000E+00AAH
   100,000     0,000   180,000     0,000     0,000     0,000    0          0,000
    1    2  348    0
    40,732    75,859     0,000     0,000    12,000     0,000     0,000     0,000
    10,000    13,000     0,050     0,050RM       M  A
    0
```

Figure A.1: Example of a tape5 file to use with MODTRAN 4v3r1 [Cook, 2014].

Table A.1: Card 1 input descriptions and settings used for the LST process.

Card 1

Variable	Description	Input	Explanation
MODTRN	band model algorithm for radiative transport	T	using MODTRAN band model
SPEED	correlated k-option	M	'medium' speed correlated-k option
MODEL	geographical/seasonal atmosphere	7	user-specified model atmosphere
ITYPE	atmospheric line-of-sight	3	vertical or slant path to space or ground
IEMSCT	mode of execution	2	spectral thermal and solar/lunar radiance
IMULT	multiple scattering	1	execute with multiple scattering
M1 M2 M3 M4 M5 M6	profile for temp & pressure profile for H_2O profile for O_3 profile for CH_4 profile for CO	0 0 0 0 0 0	JCHAR parameter in Card 2C1 supplies necessary profiles because user supplies model atmosphere
MDEF	$CO_2, O_2, NO, SO_2, NO_2$ $H_3 and HNO_3$ profiles	1	default heavy species profiles
IM	read user input data	1	always read new user input data
NOPRINT	controls output	0	normal tape6 output
TPTEMP	boundary temperature	tmp.000	boundary temperature input based on current MODTRAN run
SURREF	albedo of the Earth	alb0	surface albedo input based on current MODTRAN run

Table A.2: Card 1A input descriptions and settings used for the LST process.

Card 1A

Variable	Description	Input	Explanation
DIS	select multiple scattering algorithms	T	activate discrete ordinate multiple scattering algorithm (slower and more accurate) DISORT
DISAZM	azimuth dependence flag	blank	excludes azimuth dependence
NSTR	number of streams in scattering algorithm	8	uses recommended 8 streams in DISORT
LSUN	spectral resolution of irradiance	F	default solar 5 cm^{-1} spectral resolution irradiance
ISUN	FWHM of triangular scanning function	F	default values for FWHM
CO2MX	CO_2 mixing ratio in ppmv	360.000	default value is 330 ppmv, recommended is closer to 365 ppmv
H2OSTR	vertical water vapor column character string	0	uses default water vapor column
O3STR	vertical ozone column character string	0	default ozone column used
LSUNFL	reading solar radiance data	F	use default solar radiance data
LBMNAM	read band model parameter data	F	default band model (1 cm^{-1} bin) database
LFLTNM	read file for user-defined instrument filter	F	no user-defined instrument filter function
H2OAER	relating aerosal properties and relative humidity	blank	fixed H_2O properties even though water amount has changed
LDATDR	reading MODTRAN data files	blank	data files assumed to be in directory in DATA/
SOLCON	scaling TOA irradiance	0.000	do not scale TOA solar irradiance

Table A.3: Card 2 input descriptions and settings used for the LST process.

Card 2

Variable	Description	Input	Explanation
APLUS	aerosol profiles	blank	default aerosol profiles
IHAZE	type of extinction and meteorological range	1	rural extinction, default VIS = 23 km
CNOVAM	aerosol model	blank	default aerosol model
ISEASN	appropriate seasonal aerosol profile for tropospheric and stratospheric aerosols	0	season determined by model, spring-summer when model = 7
ARUSS	defining aerosol optical properties	blank	default aerosol optical properties
IVULCN	stratospheric aerosols and extinction	0	background stratospheric profile and extinction
ICSTL	air mass character where 1 = open ocean, 10 = strong continental influence	0	uses default air mass character = 3
ICLD	cloud and rain models	0	no clouds or rain
IVSA	army vertical structure algorithm	0	does not use army vertical structure algorithm for aerosols in boundary layer
VIS	surface meteorological range	0.000	uses default meteorological range set by IHAZE
WSS	current wind speed (m/s)	0.000	only used with IHAZE = 3 or IHAZE = 10
WHH	24-hour average wind speed	0.000	only used with IHAZE = 3
RAINRT	specifies the rain rate	0.000	default is 0 for no rain
GNDALT	altitude of the surface relative to sea level (km)	gdalt	altitude input based on current MODTRAN run

Table A.4: Card 2C input descriptions and settings used for the LST process.

Card 2C

Variable	Description	Input	Explanation
ML	number of atmospheric levels	mml	number of levels in profile determined based on current MODTRAN run
IRD1	reading of Card 2C2	0	no reading of Card 2C2
IRD2	reading of Card 2C3	0	no reading of Card 2C3
HMODEL	identification of new model atmosphere	blank	no new model atmosphere identified
REE	earth radius in kilometers	blank	only read in model = 8

Table A.5: Card 2C1 input descriptions and settings used for the LST process.

Card 2C1

Variable	Description	Input	Explanation
ZM	altitude of layer boundary		input for each atmospheric layer in the current MODTRAN run
P	pressure of layer boundary		input for each atmospheric layer in the current MODTRAN run
T	temperature of layer boundary		input for each atmospheric layer in the current MODTRAN run
WMOL(1)	water vapor		input for each atmospheric layer in the current MODTRAN run
WMOL(2)	carbon dioxide	0.000e+00	not specified for any layer
WMOL(3)	ozone	0.000e+00	not specified for any layer
JCHAR(1)	units of pressure at layer boundary	A	specifies pressure in mb
JCHAR(2)	units of temperature at layer boundary	A	specified temperature in K
JCHAR(3)	specifies which water vapor	H	specified water vapor as relative humidity in %
JCHAR(4) JCHAR(5)	defaults to M1-M6 and MDEF values when WMOL(2-3) are zero	blank blank	MDEF = 1 specifies default profiles
JCHAR(6) JCHAR(7) JCHAR(8) JCHAR(9) JCHAR(10) JCHAR(11) JCHAR(12) JCHAR(13) JCHAR(14)	corresponds to WMOL(4-12)	blank blank blank blank blank blank blank blank blank	never read based on IRD1 in Card 2C
JCHARX	units for CFC's and other heavy molecules	blank	MDEF = 1 specifying default profiles

Table A.6: Card 3 input descriptions and settings used for the LST process.

Card 3

Variable	Description	Input	Explanation
H1	initial altitude (km)	100.000	observer/sensor altitude of 100 km
H2	tangent height (km)	0.000	target on the ground
ANGLE	initial zenith angle (0-180°) as measured from H1	180.000	sensor looking at the ground
RANGE	path length (km)	0.000	path length from sensor to ground
BETA	earth center angle subtended by H1 and H2 (0-180°)	0.000	sensor pointing directly at target
RO	radius of the earth (km) at particular altitude of calculation	0.000	uses default mid-altitude value of 6371.23 km for MODEL = 7
LENN	path length specification	0	short path length
PHI	zenith angle as measured from H2 towards H1 (0-180°)	0.000	phi does not need to be specified since H1, H2, and ANGLE are defined

Table A.7: Card 3A1 input descriptions and settings used for the LST process.

Card 3A1

Variable	Description	Input	Explanation
IPARM	method of specifying lunar/solar geometry on Card 3A2	1	see Card 3A2 inputs
IPH	specification of phase function	2	mid-generated internal database of aerosol phase functions for MODTRAN models
IDAY	day of year from 1 to 365 to specify sun's locations	jay	day of year input from current MODTRAN run
ISOURC	extraterrestrial source	0	extraterrestrial source is the sun

Table A.8: Card 3A2 input descriptions and settings used for the LST process.

Card 3A2

Variable	Description	Input	Explanation
PARM1	observer latitude (-90° to +90°)	latitu	latitude input from current MODTRAN run
PARM2	observer longitude (0° to 360° West of Greenwich)	longit	longitude input from current MODTRAN run
PARM3	sun latitude	0.000	not required for IPARM = 1
PARM4	sun longitude	0.000	not required for IPARM = 1
TIME	Greenwich time	12.000	12 Z used for all MODTRAN runs
PSIPSO	true path azimuth from H1 to H2	0.000	degrees East of true North
ANGLEM	phase angle of the moon	0.000	not required in our settings
G	asymmetry factor	0.000	not required in our settings

Table A.9: Card 4 input descriptions and settings used for the LST process.

Card 4

Variable	Description	Input	Explanation
V1	initial frequency in wavenumber or wavelength	10.000	wavelength in microns
V2	final frequency in wavenumber of wavelengths	13.000	wavelength in microns
DV	frequency or wavelength increment used for spectral outputs	0.050	wavelength increment in microns
FWHM	slit function full width at half maximum	0.050	FWHM of slit function in microns
YFLAG	values in output files	R	output radiances (rather than transmittances)
XFLAG	units of values in output files	M	spectral wavelength in microns
DLIMIT	separate output from repeat in MODTRAN runs	blank	not necessary in our settings, no repeat
FLAGS	seven character string	see below	
FLAGS(1:1)	spectral units	M	spectral units in microns
FLAGS(2:2)	slit function	blank	default slit function
FLAGS(3:3)	FWHM characteristics	blank	FWHM is absolute
FLAGS(4:4)	degradation of results	A	degrade all radiance and transmittance components
FLAGS(5:5)	degradation settings	blank	do not save current results
FLAGS(6:6)	degradation settings	blank	do not use saved results
FLAGS(7:7)	”spec flux” file	blank	spectral flux values output at all atmospheric levels

Table A.10: Card 5 input descriptions and settings used for the LST process.

Card 5

Variable	Description	Input	Explanation
IRPT	program execution setting	0	stop program

Appendix B

Comparing MODIS SST and Landsat LST

: As discussed in Section 5.4.2, the MODIS Sea Surface Temperature (SST) product is an adequate source of ground truth for validating Landsat 7 on a global level. In order to assess how well the SST product represents the truth, we took a set of Landsat 7 scenes with buoy truth available, and obtained the corresponding MODIS SST scenes. Then we compared the predicted temperature from the LST process at the buoy location to the buoy truth, and we compared the MODIS SST product at the buoy location to the buoy truth.

When the global validation of Landsat 7 was performed, we continued to compare LST to SST, but we were no longer able to also obtain comparison points at buoy locations, because they are not available globally. Instead, we were free to select several points within each Landsat scene for comparison. There are a few main steps leading up to the comparison process that we will discuss, such as download and georeferening MODIS SST imagery, and then we will also go into detail about how we designed the selection process for finding comparison points.

B.1 Downloading MODIS SST

Given a set of Landsat 7 scenes containing buoys, the corresponding MODIS scenes need to be obtained. MODIS products can be downloaded from the Ocean Color website (http://oceancolor.nasa.gov/cms), where there is a browser interface as well as a direct access page. Note that only MODIS scenes from the TERRA satellite were used, because of its similar orbit to Landsat 7. Below is a description of the format for MODIS SST scene identifiers:

iyyyydddhhmmss.L2_rrr_ppp

i	=	Sensor (A for AQUA, T for TERRA)
yyyy	=	Year
ddd	=	Day
hh	=	Hour
mm	=	Minute
ss	=	Second
L2	=	Level 2 product
rrr	=	Resolution (GAC for subsampled, LAC for full resolution)
ppp	=	Product identifier (SST for sea surface temperature)

Since the Landsat scene identifiers contain similar information to MODIS scene identifiers, it is a fairly simple matter to identify the MODIS scene that images over the same area on the same day. Considering as an example the Landsat scene LE70130332010065EDC00, the MODIS file name would be T2010065155000.L2_LAC_SST. In order to automate the download process for MODIS scenes, we need to be able to predict the MODIS scene identifier based on a given Landsat identifier.

MODIS scenes are captured anywhere between 15 and 30 minutes after their corresponding Landsat scenes, but there is no obvious way to predict the file name that needs to be downloaded. After a MODIS file is downloaded, however, one can identify the corner latitude and longitude values, and check if the Landsat scene falls within that range. This way, we can simply download MODIS scenes until the right one is found. Although this is not an ideal solution because multiple downloads are required, the correct scene is typically found after three attempts, which is much faster than selecting MODIS scenes manually through the Ocean Color website.

B.2 Georeferencing and Subsetting MODIS

The MODIS SST files downloaded from the Ocean Color website are in NetCDF format. Among the information provided in the SST files are scan time and location, latitude/longitude control points, sea surface temperature, and quality levels. The desired entries such as the sea surface temperature and quality levels can be read and manipulated using ENVI Classic, an image analysis software that is especially useful for geospatial applications. ENVI can be used to georeference the MODIS scenes so that they are in the same coordinate system as Landsat, which is needed to perform the desired comparison. The MODIS scenes can also be reduced to a subset around an area of interest, such as a buoy. The benefit of this is that the subset can be saved as a relatively small file, which reduces computation time when comparing with Landsat scenes. Once again, there is a clear desire to automate this process, so we used IDL to run ENVI application control routines that can georeference and subset a whole set of MODIS NetCDF files.

The following steps are applied to both the SST image and the quality levels image for a given MODIS file. The georeferencing is done using ENVI application control routines that build Geographic Lookup Tables (GLTs) and apply them to the SST and quality images so that they have a UTM projection type. For the subsetting step, the Landsat corner latitude/longitude values are converted to UTM and then to pixel locations within the georeferenced MODIS image, which are used to create the subset. Finally, the subset image is saved as a geotiff image.

B.3 Comparing MODIS and Landsat LST at Specific Points

For the Landsat 7 global validation study, we came up with several rules for an automatic process to find adequate comparison points between Landsat and SST. The program uses a 10 x 10 window to search through the georeferenced and subset SST image, although it is only checking points at the center of the window. The following rules/guidelines describes the necessary circumstances for a comparison point to be found.

1. If any pixels within window are labeled as land, then move a window's width forward
2. If any pixels within window are outside the Landsat scene, move a window's width forward
3. If the pixel at the center of the window has at less than 5 best quality pixels in the surrounding 3 x 3, check four more pixels within the same window and the n move on if necessary
4. If the standard deviation of the surrounding 3 x 3 is greater than 0.5, then move on

5. If a pixel meets all the requirements, data is recorded and the window is moved forward until the whole scene has been searched or until seven points have been identified

Appendix C

Details of Standard Error Propagation Terms

This appendix will expand on the details of the error propagation equations used to evaluate the error associated with the surface leaving radiance, S_{L_T}. Equations C.1 and C.2 serve as a reminder of the various terms that contribute to S_{L_T}, and they were first introduced in Section 4.6.

$$S_{L_T} = \sqrt{S_A^2 + S_I^2 + S_E^2 + S_P} \tag{C.1}$$

$$
\begin{aligned}
S_A^2 &= \left(\frac{\delta L_T}{\delta \tau} S_\tau\right)^2 + \left(\frac{\delta L_T}{\delta L_u} S_{L_u}\right)^2 + \left(\frac{\delta L_T}{\delta L_d} S_{L_d}\right)^2 \\
S_I^2 &= \left(\frac{\delta L_T}{\delta L_{obs}} S_{L_{obs}}\right)^2 \\
S_E^2 &= \left(\frac{\delta L_T}{\delta \epsilon} S_\epsilon\right)^2 \\
S_P &= 2\rho_{\tau L_u} \frac{\delta L_T}{\delta \tau} \frac{\delta L_T}{\delta L_u} S_\tau S_{L_u} \\
&\quad + 2\rho_{\tau L_d} \frac{\delta L_T}{\delta \tau} \frac{\delta L_T}{\delta L_d} S_\tau S_{L_d} \\
&\quad + 2\rho_{L_u L_d} \frac{\delta L_T}{\delta L_u} \frac{\delta L_T}{\delta L_d} S_{L_u} S_{L_d}
\end{aligned}
\tag{C.2}
$$

Error due to the atmosphere

This section will discuss the components that make up the error due to the atmosphere, represented by S_A. This term refers to the error that is introduced by using atmospheric profile inputs that are provided by reanalysis products. The definition for S_A can be found in Equation C.3, and all further definitions of terms within that equation are listed in Equation C.4.

$$
S_A^2 = \left(\frac{\delta L_T}{\delta \tau} S_\tau\right)^2 + \left(\frac{\delta L_T}{\delta L_u} S_{L_u}\right)^2 + \left(\frac{\delta L_T}{\delta L_d} S_{L_d}\right)^2 \tag{C.3}
$$

$$\begin{aligned}
\frac{\delta L_T}{\delta \tau} &= \frac{\delta L_T}{\delta \tau}\left(\frac{L_{obs} - L_u}{\tau\epsilon} + \frac{(\epsilon - 1)\, L_d}{\epsilon}\right) = \frac{L_u - L_{obs}}{\epsilon\tau^2} \\
\frac{\delta L_T}{\delta L_u} &= \frac{\delta L_T}{\delta L_u}\left(\frac{L_{obs}}{\tau\epsilon} - \frac{L_u}{\tau\epsilon} + \frac{(\epsilon - 1)\, L_d}{\epsilon}\right) = -\frac{1}{\tau\epsilon} \\
\frac{\delta L_T}{\delta L_d} &= \frac{\delta L_T}{\delta L_d}\left(\frac{L_{obs} - L_u}{\tau\epsilon} + \frac{(\epsilon - 1)\, L_d}{\epsilon}\right) = \frac{\epsilon - 1}{\epsilon} \\
S_\tau &= \left[\left(\frac{\delta\tau}{\delta T} S_T\right)^2 + \left(\frac{\delta\tau}{\delta RH} S_{RH}\right)^2\right]^{\frac{1}{2}} \\
S_{L_u} &= \left[\left(\frac{\delta L_u}{\delta T} S_T\right)^2 + \left(\frac{\delta L_u}{\delta RH} S_{RH}\right)^2\right]^{\frac{1}{2}} \\
S_{L_d} &= \left[\left(\frac{\delta L_d}{\delta T} S_T\right)^2 + \left(\frac{\delta L_d}{\delta RH} S_{RH}\right)^2\right]^{\frac{1}{2}}
\end{aligned} \tag{C.4}$$

Our approach to evaluating this source of error involves several radiative transfer simulations that will let us estimate the value of the partials as well as the error in each of the atmospheric parameters S_τ, S_{L_u}, and S_{L_d}. This will begin by taking temperature and relative humidity profiles from MERRA and perturbing them to see how transmission, upwelled radiance, and downwelled radiance change so that ultimately the effective error in L_T can be found. The goal is to be able to create a lookup table of S_τ, S_{L_u}, and S_{L_d} values that will correspond to some error in L_T that is due to the use of the reanalysis profiles.

Cross correlation terms

This section will discuss the cross correlation components from taking partial derivatives, represented by S_P. There may be some amount of correlation between trasmission, upwelled radiance, and downwelled radiance, so it is important to include these terms in the overall calculation of S_{L_T}. The definition for S_P can be found in Equation C.5, and all further definitions of terms within that equation are listed in Equation C.6. There are some terms which are not explicitly defined here, because they were already defined in Equation 3.9.

$$\begin{aligned} S_P = 2\rho_{\tau L_u}\frac{\delta L_T}{\delta\tau}\frac{\delta L_T}{\delta L_u}S_\tau S_{L_u} \\ + 2\rho_{\tau L_d}\frac{\delta L_T}{\delta\tau}\frac{\delta L_T}{\delta L_d}S_\tau S_{L_d} \\ + 2\rho_{L_u L_d}\frac{\delta L_T}{\delta L_u}\frac{\delta L_T}{\delta L_d}S_{L_u}S_{L_d} \end{aligned} \tag{C.5}$$

$$\begin{aligned} \rho_{\tau L_u} &= \text{correlation coefficient for} \tau \text{ and } L_u \\ \rho_{\tau L_d} &= \text{correlation coefficient for} \tau \text{ and } L_d \\ \rho_{L_u L_d} &= \text{correlation coefficient for} L_d \text{ and } L_d \end{aligned} \tag{C.6}$$

The values for the cross correlation terms can be obtained through the process that is used to determine S_A, the error due to the atmosphere. The mathematical expressions for the partial derivatives as well as S_τ, S_{L_u}, and S_{L_d} were already defined in Equation 3.9 because they contributed to the error due to the atmosphere; therefore, the values of these will be calculated from the simulations. We are then left with the task of obtaining the correlation coefficients described in Equation C.6. These correlations can be observed simply by plotting τ, L_u, and L_d against each other using the simulation values and calculating the correlation coefficients.

Error due to the Landsat instrument

This section will discuss the components that make up the error due to the Landsat instrument, represented by S_I. This term refers to the uncertainty introduced by the instrument as it captures images. The definition for S_I can be found in Equation C.7, and all further definitions of terms within that equation are listed in Equation C.8.

$$S_I^2 = \left(\frac{\delta L_T}{\delta L_{obs}} S_{L_{obs}} \right)^2 \tag{C.7}$$

Imaging systems always have some amount of measurement uncertainty associated with the signal they capture, the value of which depends of factors such as read noise, shot noise, and dark current. For each Landsat sensor, there is a corresponding estimate of this total measurement uncertainty in terms of radiance. These values will be obtained from Landsat calibration papers.

$$\begin{aligned} \frac{\delta L_T}{\delta L_{obs}} &= \frac{\delta L_T}{\delta L_{obs}} \left(\frac{L_{obs}}{\tau\epsilon} - \frac{L_u}{\tau\epsilon} + \frac{(\epsilon - 1)\, L_d}{\epsilon} \right) = \frac{1}{\tau\epsilon} \\ S_{L_{obs}} &= \text{scalar value associated with each Landsat sensor} \end{aligned} \tag{C.8}$$

Error due to emissivity

This section will discuss the components that make up the error due to the emissivity product, represented by S_E. This term refers to the error that is introduced by using the ASTER Global Emissivity Database to provide us with knowledge of surface emissivities. The definition for S_E can be found in Equation C.9, and all further definitions of terms within that equation are listed in Equation C.10.

$$S_E^2 = \left(\frac{\delta L_T}{\delta \epsilon} S_\epsilon\right)^2 \tag{C.9}$$

$$\frac{\delta L_T}{\delta \epsilon} = \frac{\delta L_T}{\delta \epsilon}\left(\frac{L_{obs} - L_u}{\tau \epsilon} + L_d - \frac{L_d}{\epsilon}\right) = \frac{L_u - L_{obs} + L_d \tau}{\tau \epsilon^2}$$
$$S_\epsilon = \text{standard deviation values from the ASTER GED} \tag{C.10}$$

It is important to clarify that the term S_E represents the amount of radiance that the emissivity product contributes to the uncertainty of L_T, while the term S_ϵ refers to the uncertainty in the product itself. We would first like to see if there is a relationship between the emissivity uncertainty and δL_T, so that we can determine if S_E can be approximated by an average value or if it needs to be calculated every time. We anticipate performing a small study where a variety of Landsat pixels are selected to represent a range of temperatures, and the corresponding emissivity standard deviations are recorded. The atmospheric paramters can be determined by MODTRAN, so Equation C.10 can be easily solved for δL_T. Plotting δL_T versus the emissivity standard deviation will reveal what if any relationship exists between the two.

Bibliography

[Aber, 2013] Aber, J. S. (2013). Landsat technical characteristics.

[AFRL, 2015] AFRL (2015). About modtran.

[Anderson and Kustas, 2008] Anderson, M. C. and Kustas, W. P. (2008). Thermal remote sensing of drought and evapotranspiration. *EOS Transactions*, 89(26):233–240.

[Barsi et al., 2005] Barsi, J., Barker, J., and Schott, J. (2005). Validation of a web-based atmospheric correction tool for single thermal band instruments. *SPIE*, 5882.

[Barsi et al., 2010] Barsi, J., Schott, J., Palluconi, F., Helder, D., Hook, S., Markham, B., Chander, G., and O'Donnell, E. (2010). Landsat tm and etm+ thermal band calibration. *Canadian Journal of Remote Sensing*, 36(5):617–630.

[Berk et al., 2003] Berk, A., Anderson, G., Acharya, P., Hoke, M., Chetwynd, J., Bernstein, L., Shettle, E., Matthew, M., and Adler-Golden, S. (2003). *MODTRAN4 Version 3 Revision 1 User's Manual*. Air Force Research Laboratory.

[Caselles and Sobrino, 1989] Caselles, V. and Sobrino, J. A. (1989). Determining of frosts in orange groves from noaa-9 avhrr data. *Remote Sensing of Environment*, 29:135–146.

[Chander and Markham, 2003] Chander, G. and Markham, B. (2003). Revised landsat-5 tm radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2674–2677.

[Cook, 2014] Cook, M. (2014). *Atmospheric Compensation for a Landsat Land Surface Temperature Product*. Phd, Rochester Institute of Techmology.

[Dee et al., 2015] Dee, D., Fasullo, J., Shea, D., and Walsh, J. (2015). Atmospheric reanalysis: Overview and comparison tables.

[Feitas et al., 2011] Feitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., and Perdigao, R. (2011). Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, 34(9-10):3051–3068.

[Freitas et al., 2010] Freitas, S. C., Trigo, I. F., Bioucas-Dias, J. M., and Gottsche, F. M. (2010). Quantifying the uncertainty of land surface temperature retrievals from seviri/meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):523–534.

[Gillespie et al., 1998] Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, J. S., Hook, S., and Kahle, A. B. (1998). A temperature and emissivity separation algorithm for advanced spaceborn thermal emission and reflection radiometer (aster) images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4):1113–1126.

[Goff and Gratch, 1946] Goff, J. A. and Gratch, S. (1946). Low-pressure properties of water from -160 to 212 f. *Transactions of the American Society of Heating and Ventilating*, pages 95–122.

[Hook et al., 2004] Hook, S. J., Chander, G., Barsi, J. A., Alley, R. E., Palluconi, A. A. F. D., Markham, B. L., Richards, R. C., Schladow, S. G., and Helder, D. L. (2004).

In-flight validation and recovery of water surface temperature with landsat-5 thermal infrared data using an automated high-altitude lake validation site at lake tahoe. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12):2767–2776.

[Hook et al., 2007] Hook, S. J., Vaughan, G., Tonooka, H., and Schladow, S. G. (2007). Absolute radiometric in-flight validation of mid infrared and thermal infrared data from aster and modis on the terra spacecradt using the lake tahoe, ca/nv, usa, automated validation site. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1798–1807.

[Hulley et al., 2015] Hulley, G., Hook, S., Malakar, E., Islam, T., and Abrams, M. (2015). The aster global emissivity database (aster ged): Mapping earths emissivity at 100 meter spatial resolution. *Geophysical Research Letters*.

[Hulley and Hook, 2009a] Hulley, G. C. and Hook, S. J. (2009a). The north american aster land surface emissivity database (naalsed) version 2.0. *Remote Sensing Environment*, 113(9):1967–1975.

[Hulley and Hook, 2009b] Hulley, G. C. and Hook, S. J. (2009b). The north american aster land surface emissivity database (naalsed) version 2.0. *Geophysical Research Letters*, 36(22).

[Hulley and Hook, 2011] Hulley, G. C. and Hook, S. J. (2011). Generating consistent land surface temperature and emissivity products between aster and modis data for earth science research. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9):1304–1315.

[Hulley et al., 2012] Hulley, G. C., Hughes, C. G., and Hook, S. J. (2012). Quantifying uncertainties in land surface temperature and emissivity retrievals from aster and modis thermal infrared data. *Journal of Geophysical Research*, 117(D23113).

[Ip and Siebels, 2009] Ip, J. and Siebels, P. D. (2009). *Operational algorithm description document for viirs land surface temperature (lst) edr software.* Technical Report D38714 Rev A, Northrop Grumman Space and Missions Systems Corporation.

[Irons and Rocchio, 2015a] Irons, J. and Rocchio, L. (2015a). History.

[Irons and Rocchio, 2015b] Irons, J. and Rocchio, L. (2015b). Landsat applications.

[Irons and Rocchio, 2015c] Irons, J. and Rocchio, L. (2015c). The worldwide reference system.

[Jackson et al., 1977] Jackson, R., Reginato, R. J., and Idso, S. B. (1977). Wheat canopy temperature: A practical tool for evaluating water requirements. *Journal of Geophysical Research*, 13:651–656.

[Jimenez-Munoz and Sobrino, 2004] Jimenez-Munoz, J. C. and Sobrino, J. (2004). Error sources on the land surface temperature retrieved from thermal infrared single channel remote sensing data. *International Journal of Remote Sensing*, 27(5):999–1014.

[Kempler, 2009] Kempler, S. (2009). Merra data holdings.

[Kruger, 2010] Kruger, M. (2010). Converter for humidity of air.

[Leff, 1999] Leff, C. (1999). *Aster higher-level data product quality assessment plan.* Technical Report, Jet Propulsion Laboratory.

[Meehl, 1994] Meehl, G. A. (1994). Influence of land surface in the asian summer monsoon: External conditions versus internal feedback. *Journal of Climate*, 7:1033–1049.

[Montanaro et al., 2014] Montanaro, M., Gerace, A., Lunsford, A., and Reuter, D. (2014). Stray light artifacts in imagery from the landsat 8 thermal infrared sensor. *Journal of Remote Sensing*, 6(11):10435–10456.

[NDBC, 2009] NDBC (2009). Handbook of automated data quality control checks and procedures.

[NOAA, 2015] NOAA (2015). Ten basic cloud types.

[NOMADS, 2015] NOMADS (2015). Nomads data access.

[Padula et al., 2003] Padula, F., Schott, J., Barsi, J., Raqueno, N., and Hook, S. (2003). Landsat tm and etm+ thermal band calibration. *Canadian Journal of Remote Sensing*, 29(2):141–253.

[Padula and Schott, 2010] Padula, F. P. and Schott, J. R. (2010). Historic calibration of the thermal infrared band of landsat-5 tm. *Photogrammetric Engineering and Remote Sensing*, 76(11):1225–1238.

[Qin et al., 001a] Qin, Z., Dall'Olmo, G., and Karnieli, A. (2001a). Derivation of split window algorithm and its sensitivity analysis for retrieving land surface temperature from noaa-advanced very high resolution radiometer data. *Journal of Geophysical Research*, 106(D19):22655–22670.

[Qin et al., 001b] Qin, Z., Karnieli, A., and Berliner, P. (2001b). A mono-window algorithm for retrieving land surface temperature from landsat tm data and its application to the israel-egypt border region. *International Journal of Remote Sensing*, 22(18):3719–3746.

[Rubel and Kottek, 2010] Rubel, F. and Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the kppen-geiger climate classification. *Meteorol. Z.*, 19.

[Schneider et al., 2009] Schneider, P., Hook, S. J., Radocinski, R. G., Corlett, G. K., Hulley, G. C., Schladow, S. G., and Steissberg, T. E. (2009). Satellite observations indicate rapid warming trend for lakes in california and nevada. *Remote Sensing of Environment*, 113:1967–1975.

[Schott, 2007] Schott, J. (2007). *Remote Sensing: The Image Chain Approach*. Oxford University Press, 2nd edition.

[Schott, 1986] Schott, J. R. (1986). The role of remotely sensed data in the studies of the thermal bar. *Remote Sensing Reviews*, 2:341–358.

[Schott et al., 2001] Schott, J. R., Barsi, J., Nordgren, B., Raqueno, N., and de Alwis, D. (2001). Calibration of landsat thermal data and application to water resource studies. *Remote Sensing Environment*, 78:108–117.

[Schott et al., 2012] Schott, J. R., Hook, J. S., Barsi, J. A., Markham, B. L., Miller, J., Padula, F. P., and Raqueno, N. (2012). Thermal infrared radiometric calibration of the entire landsat 4, 5, and 7 archive. *Remote Sensing of Environment*, 122:41–49.

[Seemann et al., 2006] Seemann, S. W., Borbas, E. E., Li, J., Menzel, P., and Gumleyl, L. E. (2006). Modis atmospheric profile retrieval algorithm theoretical basis document.

[Shafran, 2007] Shafran, P. (2007). North american regional reanalysis homepage.

[Shepard, 1968] Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *ACM National Conference*, pages 517–524.

[Sobrino et al., 2004] Sobrino, J., Jimenez-Munoz, J. C., and Paolini, L. (2004). Land surface temperature retrieval from landsat tm 5. *Remote Sensing of Environment*, 90:434–440.

[Sobrino et al., 2001] Sobrino, J., Raissouni, N., and Li, Z. L. (2001). A comparative study of land surface emissivity retrieval from noaa data. *Remote Sensing of Environment*, 75(2):256–266.

[Sun and Pinker, 2003] Sun, D. and Pinker, R. T. (2003). Estimation of land surface temperature from a geostationary operational environmental satellite (goes-8). *Journal of Geophysical Research*, 108(D11).

[Sun et al., 2004] Sun, D., Pinker, R. T., and Basara, J. B. (2004). Land surface temperature estimation from the next generation of geostationary operational environmental satllites: Goes m-q. *Journal of Applied Metrology*, 43:363–372.

[USGS, 2001] USGS (2001). The universal transverse mercator (utm) grid.

[USGS, 2013] USGS (2013). Landsat: A global land-imaging mission.

[USGS, 2014] USGS (2014). What are the band designations for the landsat satellites?

[USGS, 2015] USGS (2015). Product guide: Landsat 4-7 climate data record (cdr) surface reflectance.

[Vasquez et al., 1997] Vasquez, D. P., Reyes, F. J. O., and Arboledas, L. A. (1997). A comparative study of algorithms for estimating land surface temperature from avhrr data. *Remote Sensing of Environment*, 62:215–222.

[Wan,] Wan, Z. *Collection-5 modis land surface temperature products users guide.*

[Wan, 1999] Wan, Z. (1999). *Modis land-surface temperature algorithm theoretical basis document.* Technical report, Institute for Computational Earth System Science, University of California.

[Wan and Dozier, 1989] Wan, Z. and Dozier, J. (1989). Land-surface temperature measurement from space: physical principles and inverse modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 27(3):268–278.

[Wan and Dozier, 1996] Wan, Z. and Dozier, J. (1996). A generalized split window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4):892–905.

[Wan et al., 2002a] Wan, Z., Zhang, Y., Li, Z. L., Wang, R., Salomonson, V. V., Yves, A., Bosseno, R., and Hancocq, J. F. (2002a). Preliminary estimate of calibration of the moderate resolution imaging spectroradiometer thermal infrared data using lake titicaca. *Remote Sensing of Environment*, 80:497–515.

[Wan et al., 2004] Wan, Z., Zhang, Y., Zhang, Q., and Li, Z. (2004). Quality assessment and validation of the modis global land surface temperature. *International Journal of Remote Sensing*, 25(1):261–274.

[Wan et al., 2002b] Wan, Z., Zhang, Y., Zhang, Q., and Li, Z. L. (2002b). Validation of the landsurface temperature products retrieved from terra moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 83:163–180.

[Wright, 1997] Wright, J. M. (1997). Federal meteorological handbook no. 3.

[Yu and Privette, 2005] Yu, Y. and Privette, J. L. (2005). Analysis of the npoess viirs land surface temperature algorithm using modis data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(10):2340–2350.

[Yu et al., 2012] Yu, Y., Tarpley, D., Privette, J. L., Flynn, L. E., Xu, H., Chen, M., Vinnikov, K. Y., Sun, D., and Tian, Y. (2012). Validation of geos-r satellite land surface temperature algorithm using surfrad ground measurements and statistical estimates of error properties. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):704–713.

[Yu et al., 2010] Yu, Y., Tarpley, D., Xu, H., and Chen, M. (2010). *Goes-r advanced baseline imager (abi) algorithm theoretical basis document for land surface temperature*. Technical Report Version 2.0, NOAA NESDIS Center for Satellite Applications and Research.

[Zhu and Woodcock, 2012] Zhu, Z. and Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94.