Atmospheric Compensation for a Landsat Land Surface Temperature Product

by

Monica J. Cook

B.S. Rochester Institute of Technology, 2010

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

22 October 2014

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

<u>CERTIFICATE OF APPROVAL</u>

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Monica J. Cook
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 R. Schott, Dissertation Advisor   Date

Dr. Carl Salvaggio

Dr. Emmett J. Ientilucci

Dr. Ernest Fokoue

Dr. Simon J. Hook

Atmospheric Compensation for a Landsat Land Surface Temperature Product

by

Monica J. Cook

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

Abstract

The Landsat series of satellites is the longest set of continuously acquired moderate resolution multispectral satellite imagery collected on a single maintained family of instruments. The data are very attractive because the entire archive has been radiometrically calibrated and characterized so that sensor reaching radiance values are well known. However, these values are not easily understood or applied, so this dataset has not been utilized to its fullest potential. This work focuses on atmospheric compensation at each Landsat pixel which will later be used with ASTER derived emissivity data from JPL to perform LST retrievals.

We develop a method to automatically generate the effective in band radiative transfer parameters transmission, upwelled radiance, and downwelled radiance for each pixel. We validate our methodology by comparing our predicted apparent temperatures to ground truth water temperatures derived from buoy data at a number of validation sites around the continental United States. Initial validation was performed using Landsat 5. Results show a mean error of -0.267 K and a standard deviation of 0.900 K for 259 cloud free scenes in the validation dataset. Based on the same validation dataset, our current best expectation for a confidence metric for the final product involves categorizing each pixel as cloudy, clouds in the vicinity, or cloud free based on the incorporation of a Landsat cloud product. The mean and standard deviation of the errors associated with each category will be included as a quantitative basis for each category.

To support future work, we explore the extension to a global dataset and possible improvements to the atmospheric compensation by more closely exploring the column water vapor contribution to error. Finally, we acknowledge the need for a more formal incorporation of the cloud product, and possibly improvements, in order to finalize the confidence metric for the atmospheric compensation component of the product.

Acknowledgements

To Jason, For always being someone I can aspire to be

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Land surface temperature (LST) is a valuable earth system data record most commonly applied in environmental endeavors, but it is useful in a large variety of applications. It is difficult to measure without altering the temperature of a surface, so large scale LST results are usually derived from satellite data. Deriving LST from satellite data requires multiple adjacent thermal bands to implement a split-window approach or a single thermal band with a well understood atmosphere and a surface emissivity.

The Landsat series of satellites is the longest set of continuously acquired moderate resolution multispectral satellite imagery collected on a single maintained family of instruments; they have historically collected a single thermal band. The data are very attractive, for both current analysis and historical research, because the entire archive has been radiometrically calibrated and characterized so that sensor reaching radiance values are well known. This, along with the spatial and temporal coverage of archived Landsat scenes, makes Landsat a very intriguing candidate for a wide scale LST product. However, radiance values from the thermal band are not quantities that can be intuitively applied to solve problems, so this dataset has not been utilized to its fullest potential. There are currently more than 4 million single band Landsat thermal images in the archive and between 990 and 1090 scenes acquired each day. Developing a widely accessible LST product for Landsat would broaden the usability of this vast database of imagery but requires atmospheric characterization and surface emissivity for each Landsat scene.

Recently, a high spatial resolution (100 m) gridded, surface emissivity database has been generated using emissivity information from the Advanced Spaceborne Thermal Emission and Reflection (ASTER) radiometer. Currently available from this source is the North American ASTER Land Surface Emissivity Database (NAALSED), but plans are underway to extend the dataset to global coverage [Hulley and Hook, 2009]. Assuming the availability of surface emissivity from the Jet Propulsion Laboratory (JPL), this work is focused on determining the necessary atmospheric characterization for each pixel in a Landsat scene.

The MODerate resolution atmospheric TRANsmission (MODTRAN) radiative transfer code is the standard moderate spectral resolution radiative transport model for the U.S. Air Force. With

1

the input of the appropriate atmospheric profiles, MODTRAN can be used to generate a characterization of the radiometric properties of the atmosphere at that time and place; we will use these characterizations to generate the atmospheric parameters necessary to solve for the temperature of the surface [Berk et al., 1999]. The atmospheric profiles required to execute MODTRAN are extracted from the North American Regional Reanalysis (NARR) dataset. The NARR dataset provides atmospheric variables at a fixed array of pressures on a fixed spatial grid in three-hour time intervals [Shafran, 2007].

The NARR data set is not sampled at the same resolution as the Landsat scenes, in time or space, and it is impractical to execute MODTRAN for every Landsat pixel. Because of these differences in temporal and spatial resolution, we need to interpolate both atmospheric profiles and radiative transfer parameters. Our goal in developing this methodology is to optimize data integration and interpolation to minimize errors in the final retrieved LST. This methodology is validated first for a North American product and then for a global product.

As with any scientific product, an evaluation of the error in the final quantity is required. Predicted temperatures were compared to actual water temperatures from moored buoys to calculate error values for the final retrieved LST. Using a validation dataset where the errors are known, various processes were investigated to assign either qualitative or quantitative error values or confidence metrics to each pixel in the Landsat scene based on atmospheric data, derived data, and cloud data as summarized in the results and appendices. Based on previous studies, such as that shown in Hook et al. (2007), in the thermal region we expect errors in the atmospheric compensation, rather than the development of the emissivity, to dominate uncertainty in the LST product. Therefore, the current best expectation for a confidence metric assigns each pixel to a cloud categorization with an associated mean and standard deviation of error.

Assuming the availability of an emissivity product, deliverables of this process include the Landsat thermal band radiance, elevation, transmission, upwelled radiance, and downwelled radiance of each pixel, along with the confidence metric categorization. With this information, only the emissivity is required to develop the complete land surface temperature product.

Chapter 2

Objectives

As alluded to in Chapter 1, generating a land surface temperature product is a complex and multi-step process. It is helpful to divide the project into separate tasks. Section 2.1 details the problem being approached. Section 2.2 outlines the project as four separate objectives and Section 2.3 expands on each of these tasks. The chapter closes with a summary of the contribution of the project to the field of remote sensing.

2.1 Problem Statement

The goal of our work is to develop a process to automatically generate the atmospheric parameters, along with an associated error or confidence metric due to atmospheric compensation, for a land surface temperature product for all current and historical Landsat scenes. Characterizing the atmosphere and generating the atmospheric parameters necessary to calculate the LST at each pixel requires knowledge of the atmospheric profile data at the necessary temporal and spatial resolution and use of these profiles in the radiative transfer code to predict the radiative transfer parameters.

2.2 Objectives

1. Determine a data source that can be integrated into MODTRAN with the appropriate atmospheric variables at reasonable temporal and spatial resolution and coverage for current and archived Landsat scenes in North America.

2. Develop an automated process to generate the appropriate radiative transfer parameters at each Landsat pixel.

3. Evaluate results and generate a confidence metric.

4. Extend this process to global coverage.

2.3 Tasks

1. **Determine a data source that can be integrated into MODTRAN with the appropriate atmospheric variables at reasonable temporal and spatial resolution and coverage for current and archived Landsat scenes in North America.**

   This process will require atmospheric profiles that accurately predict pressure, temperature, and one of a number of possible variables that predict humidity or water vapor at various heights. Generally, this is most accurately characterized with a radiosonde, a weather balloon that is released and rises through the atmosphere transmitting observations of various atmospheric parameters. However, because of the extent of spatial and temporal coverage and the resolution of the Landsat scenes, radiosonde data may not be the best solution for this work. We need to identify and evaluate a dataset that is accessible and provides the necessary variables over the desired time and space.

2. **Develop an automated process to generate radiative transfer parameters at each Landsat pixel.**

   Once the atmospheric profiles have been determined, MODTRAN can be used to generate the appropriate radiative transfer parameters: transmission, upwelled radiance, and downwelled radiance. There are various methods for generating such parameters using MODTRAN; each method needs to be evaluated for accuracy as well as computational efficiency. We need to develop an automated process to integrate the atmospheric profile data, execute the MODTRAN runs, and generate the parameters.

   The spatial and temporal resolution of the atmospheric data will not match that of the Landsat scenes. The computational requirements to use MODTRAN to generate radiative transfer parameters at every pixel are impractical. Therefore, studies will need to be conducted in order to evaluate the necessary number of MODTRAN runs and their optimal locations and elevations. The atmospheric profile data needs to be interpolated to the appropriate MODTRAN runs, and the resulting radiative transfer parameters need to be interpolated to the location and elevation of each pixel. Both the nature and number of interpolations need to be optimized to reduce errors in the final product.

3. **Evaluate results and generate a confidence metric.**

   The resulting product should include an estimation of uncertainty. Directly measuring the temperature of any land surface is difficult because the act of measuring often changes the observed temperature. However, the surface temperature of water can more easily be directly measured given the ability to submerge and acclimate an instrument. Buoy data from various geographical locations can be used to calculate the actual error associated with the predicted land surface temperature at a sample of water pixels; these error results will be used to analyze our methodology and to develop a process to assign a confidence estimation to each pixel in each Landsat scene.

4. Extend this process to global coverage.

This process will first be developed and validated for United States Landsat scenes. The decision is based on available atmospheric input data as well as ground truth validation sites. However, one of the benefits of Landsat is the global coverage. This product should utilize all scenes within the Landsat database, so the methodology and validation will be extended to a global product.

2.4 Contribution to Field

This product will make multiple unique contributions to the field of remote sensing.

Firstly, we develop a novel approach to large-scale single band land surface temperature retrieval. Most single band approaches require sounding data or other manual measurements to generate LST for only a single scene; our method of automatically integrating atmospheric data is unique.

Secondly, while there are other global temperature products, Landsat provides a unique combination of spatial and temporal resolution with at best 60 m pixels and a 16 day repeat cycle. This moderate spatial and temporal resolution lends itself to the applications that can best make use of a land surface temperature product.

Finally, this project puts to use a large dataset with unrealized potential. Landsat has been collecting thermal data singe 1982 but the potential of the such wide-reaching temporal and spatial coverage is unrealized because of the inability to intuitively apply radiance values. The Landsat LST product will make an already existing dataset, that is largely untapped, truly useful to, not only the remote sensing community, but other scientific communities with a need for this product.

Chapter 3

Background

Remote sensing can be described as the study from a distance of the interactions of electromagnetic energy with the object of interest and how this energy is measured by an imaging system. The purpose of this chapter is to provide the necessary background knowledge and science to understand our approach to LST retrieval through remote sensing. The chapter begins with a description of sensor reaching radiance and the thermal governing equation. Secondly, a brief history of Landsat and the capabilities of each sensor in the family are discussed to provide an understanding of the data being used. Thirdly, a description of applications that may utilize LST as well as both multiple and single band LST retrieval is discussed. Next, a description of the NARR dataset is provided and the capabilities and uses of MODTRAN are described. Finally, the chapter concludes with the explanation of data conversions used in the process. The goal of this chapter is to cover all terminology, data products, and tools that will be used in the methodology.

3.1   Sensor Reaching Radiance

This work only focuses on thermal energy reaching the sensor. Tables 3.1, 3.2, and 3.3 show that the Landsat thermal bands are sensitive from approximately 10.40 μm to 12.50 μm. Thermal energy in this portion of the electromagnetic spectrum is self-emitted from objects due to their temperature. The thermal energy paths that contribute to the sensor reaching radiance are shown in Figure 3.1.

The atmosphere above the target has some temperature, and therefore radiates energy, some of which reaches the sensor. Path A in Figure 3.1 is energy that is self-emitted from the atmosphere and scattered toward the sensor. This is often referred to as upwelled or path radiance. Self-emitted energy from the atmosphere can also be scattered toward and reflect off the target before reaching the sensor. This is path B in Figure 3.1 and referred to as the downwelled radiance. Path C is the radiation due to the temperature of the target of interest. This is ultimately the energy that we wish to quantify in order to determine the temperature of the target. Finally, objects surrounding
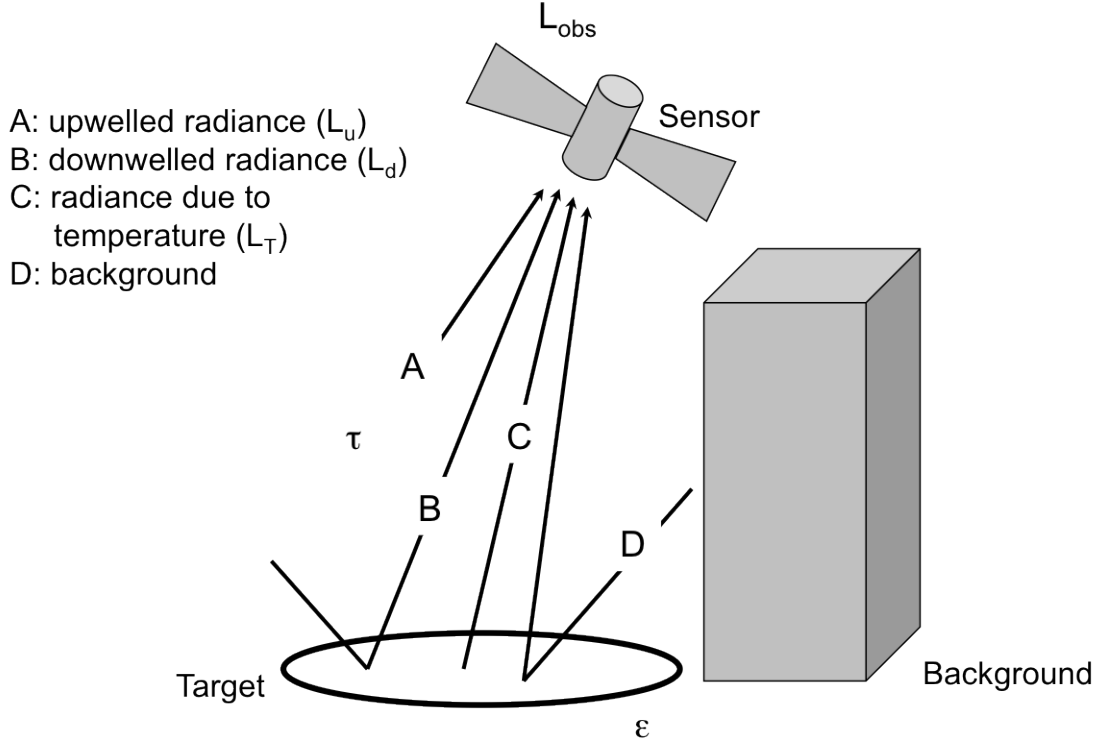
Figure 3.1: Thermal energy paths contributing to the sensor reaching radiance.

the target also have some temperature and therefore radiate energy. This energy can reflect off the target and also reach the sensor, as shown in Path D. When there are few background objects or they obscure only a small fraction of the sky, the photons from this energy path can often be considered negligible; we will make use of this assumption in our work [Schott, 2007].

The summation of paths A, B, and C compose the thermal sensor reaching radiance captured by the thermal band of the Landsat sensors. As will be described, separating and characterizing each energy path is critical to determining the temperature of the target.

3.2   Governing Equation

A governing equation includes all of the pertinent components of the sensor reaching radiance in the wavelength range of interest. The governing equation for a Landsat thermal band is shown in Equation 3.1

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

where $L_{obs\lambda eff}$ is the sensor reaching effective spectral radiance $[Wm^{-2}sr^{-1}\mu m^{-1}]$, $L_{T\lambda eff}$ is

the effective spectral radiance due to temperature $[Wm^{-2}sr^{-1}\mu m^{-1}]$ (path C in Figure 3.1), ϵ is the surface emissivity of the pixel of interest, τ is the transmission, $L_{u\lambda eff}$ is the upwelled effective spectral radiance $[Wm^{-2}sr^{-1}\mu m^{-1}]$ (path A in Figure 3.1), and $L_{d\lambda eff}$ is the downwelled effective spectral radiance $[Wm^{-2}sr^{-1}\mu m^{-1}]$ (path B in Figure 3.1). The radiance due to the temperature of the target pixel is the term that needs to be isolated; it can then be inverted to temperature of the surface using Equation 3.2, a variation of Planck's Equation, where $M_{\lambda eff}$ is the effective spectral exitance, h is Planck's constant, c is the speed of light, k is the Boltzman constant, and T is the desired temperature. The function $R(\lambda)$ is the spectral response function of the Landsat sensor; these spectral response functions vary for each sensor. The integration is over the wavelength range of interest, in this case, the sensitivity of the Landsat thermal band. Equation 3.2 cannot be directly solved for T, so we use a look up table (LUT) to determine T. To generate the LUT, we calculate an effective spectral radiance due to temperature value for every temperature (in 1 K increments) within a reasonable range of land surface temperatures; the integration is performed over the wavelength range of the thermal band. After we determine a $L_{T\lambda eff}$ value from Equation 3.1, we use a two point linear interpolation in the LUT to determine T.

$$L_{T\lambda eff} = \frac{M_{\lambda eff}}{\pi} = \frac{[\int 2hc^2\lambda^{-5}(e^{\frac{hc}{\lambda kT}} - 1)^{-1}]R(\lambda)d\lambda}{\int R(\lambda)d\lambda} \tag{3.2}$$

$L_{obs\lambda eff}$ is the effective spectral radiance reaching the sensor. This can be determined (in $Wm^{-2}sr^{-1}\mu m^{-1}$) from the calibrated digital number in the Landsat image file using Equation 3.3. In Equation 3.3, Q_{cal} is the quantized pixel value in digital counts, Q_{calmin} and Q_{calmax} are the minimum and maximum quantized calibrated pixel values (digital counts 0 and 255) corresponding to $LMIN_\lambda$ and $LMAX_\lambda$ respectively. Similarly, $LMIN_\lambda$ and $LMAX_\lambda$ are the effective spectral radiance values $(Wm^{-2}sr^{-1}\mu m^{-1})$ scaled to Q_{calmin} and Q_{calmax} respectively [Chander and Markham, 2003]. $LMIN_\lambda$, $LMAX_\lambda$, Q_{calmin}, and Q_{calmax} are given with the image metadata.

Because of the longevity of the Landsat program, and the efforts that have been put forth to maintain and re-establish calibration, for the purpose of this research, a calibrated instrument can be assumed. This means that the given calibration coefficients can be used to convert digital counts to radiance and this radiance value can be used and trusted without independent validation [Barsi et al., 2003], [Padula et al., 2010].

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

In these cases, the λeff indicates an effective spectral value of radiance per unit wavelength in units of $Wm^{-2}sr^{-1}\mu m^{-1}$, which means that the spectral response function has been incorporated as shown in Equation 3.4, where L could represent any type of radiance value (sensor reaching, upwelled, downwelled, due to temperature). We use all effective spectral values in this work and therefore will continue from this point forward without the explicit notation.

$$L_{\lambda eff} = \frac{\int L_\lambda R(\lambda)d\lambda}{\int R(\lambda)d\lambda} \qquad (3.4)$$

3.3 Landsat History

First conceived in the 1960s, Landsat is the longest set of continuously acquired moderate resolution satellite imagery. Although there were various weather satellites monitoring the Earth's atmosphere, there was little imagery that documented Earth's surface and the terrain. Initially, the idea of an Earth-observing satellite program was met with opposition due to concerns about cost, usability, and the political implications of capturing images of other countries. Despite these concerns, NASA began constructing the first Landsat satellite in 1970, known at the time as the Earth Resources Technology Satellite (ERTS).

Landsat, as the longest and only continuous record of the global land surface, can be applicable in work concerning agriculture, geology, forestry, mapping, and change detection among other applications. This dataset has been accessible to a wide community of users since all Landsat data became freely available in December 2009 [Irons and Rocchio, 2013].

A brief summary of all Landsat sensors is shown below; this tool is designed for single thermal band imagery (Landsats 4, 5, and 7). The only modification from one sensor to another is inserting a different spectral response function.

3.3.1 Landsats 1, 2, and 3

ERTS was launched in July 1972. This system, later renamed Landsat 1, operated until 1978, and exceeded expectations for both data quality and quantity. Landsat 2 launched in January 1975 and Landsat 3 in March 1978 [Irons and Rocchio, 2013]. The first three landsat instruments carried the same two sensors. The Return Beam Vidicon (RBV) and the Multi-Spectral Sensor (MSS) both had bands in the visible and near-infrared (NIR). Landsat 3 MSS actually had an additional thermal band in the long wave infrared, but this channel failed shortly after launch. Landsats 2 and 3 were decommissioned in July 1983 and September 1983 respectively [Irons and Rocchio, 2013]. These satellites cannot be used to calculate the land surface temperature because they did not capture a thermal band.

3.3.2 Landsat 4

Landsat 4 was launched in July 1982; this was the first Landsat to significantly differ from the original and the first to have a functioning long-wave thermal band (imagery that can be used to calculate land surface temperature). Landsat 4 carried the MSS instrument, identical to that on Landsats 1, 2 and 3, with four spectral bands in the visible and NIR that had 57 m and 79 m pixels. Landsat 4 also carried the Thematic Mapper (TM) with 7 spectral bands composed of 3 visible, 2 NIR, 1 short wave IR, and 1 thermal band. All bands had 30 m x 30 m pixels, except the thermal
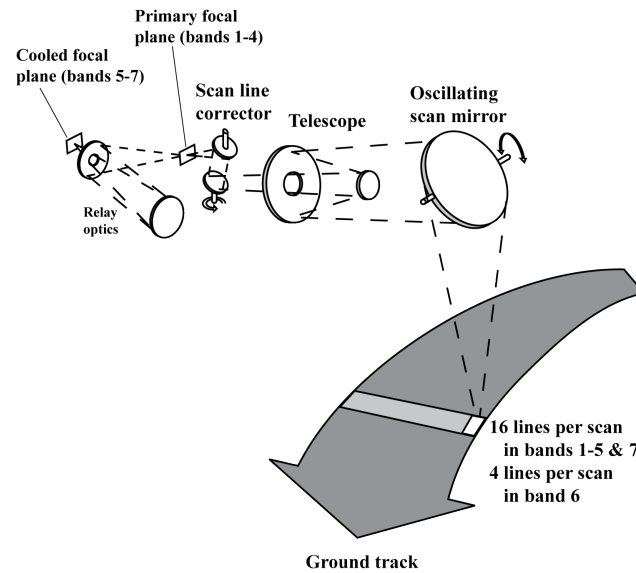
Figure 3.2: Optical layout of a line scanner [Schott, 2007].

band which had 120 m x 120 m pixels. The spectral bands of the TM instrument are shown in Table 3.1. Each scene captured was 170 km x 185 km [USGS, 2013]. Landsat 4 was not decommissioned until December 2001, although it stopped downlinking data in 1993 [Irons and Rocchio, 2013].

Table 3.1: Spectral bands of the Landsat Thematic Mapper [USGS, 2013].

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

The TM instrument is a line scanner and captures images using a Ritchey-Chretien telescope [Engel and Weinstein, 1983]. Line scanners, diagrammed in Figure 3.2, achieve along track motion by the advancement of the satellite and across track motion via an oscillating scan mirror. An oscillating scan mirror requires a scan-line corrector to align scans to eliminate data gaps; this process is illustrated in Figure 3.3.
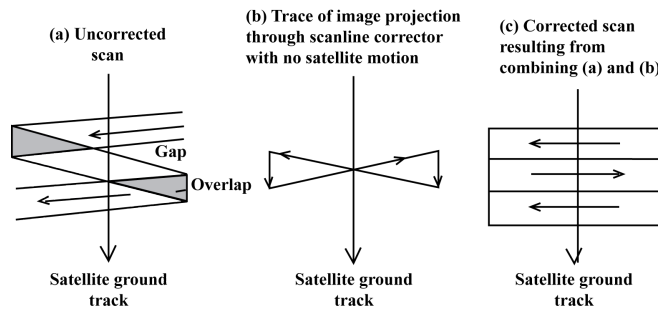
Figure 3.3: Operation of the scan line corrector on a line scanner [Schott, 2007].

3.3.3   Landsat 5

Landsat 5 launched in March 1984 carrying the same payload as Landsat 4. The system has experienced some problems with downlinking data, and the MSS was deactivated in 1995, but the TM operated for 29 years, 26 years longer than the minimum design life of 3 years. In November 2011, Landsat 5 stopped acquiring images due to degrading electrical components. Rather than operating until complete failure, the system was turned off for a period to investigate restorations or other options for image-to-ground transmission [Irons and Rocchio, 2013]. Landsat 5 was officially decommissioned 5 June 2013 and provides a tremendous amount of archived data.

Landsat 4 and Landsat 5 were privatized in 1984 and operated by a commercial vendor for 17 years. This led to a host of problems, including rising image prices, gaps in data collection, and lapses in system characterization and calibration. During this time of commercial operation, Landsat 6 was also built but failed to reach orbit. Landsats 4 and 5 were the only two of the series to ever be commercialized; operational control was returned to the federal government in 2001 [Irons and Rocchio, 2013].

3.3.4   Landsat 6

Privately owned, Landsat 6 failed to reach orbit during its October 1993 launch due to a ruptured rocket fuel chamber. The loss of Landsat 6 led to concerns about a data-gap, which would have been realized had Landsat 5 not long out-lasted its expected lifetime [Irons and Rocchio, 2013].

3.3.5   Landsat 7

Landsat 7 launched in April 1999 carrying the Enhanced Thematic Mapper Plus (ETM+), also a line scanner. The spectral bands on the ETM+ are shown in Table 3.2; improvements include better resolution for the thermal band and the addition of a higher resolution panchromatic band [USGS, 2013].

Table 3.2: Spectral bands of the Landsat Enhanced Thematic Mapper Plus [USGS, 2013].

Band	Spectrum Area	Response (μm)	Resolution
Band 1	Visible	0.45 - 0.52	30 m
Band 2	Visible	0.52 - 0.60	30 m
Band 3	Visible	0.63 - 0.69	30 m
Band 4	NIR	0.77 - 0.90	30 m
Band 5	NIR	1.55 - 1.75	30 m
Band 6	Thermal	10.40 - 12.50 (High and Low Gain Options)	60 m
Band 7	Mid-wave IR	2.08 - 2.35	30 m
Band 8	Panchromatic (PAN)	0.52 - 0.90	15 m

Landsat 7 was designed to have improved calibration from previous Landsat instruments and to be better for land cover monitoring, change detection and global mapping. Unfortunately, the scan line corrector on the ETM+ failed in May 2003 leaving gaps in each scene as illustrated in Figure 3.3(a). Better calibrated data is difficult to effectively utilize with so many missing pixels, but the instrument continues to collect and downlink imagery that is archived in "SLC-off mode," which captures approximately 75 percent of each scene. To deal with scan gaps as well as missing data due to clouds, the science community is increasingly looking to data composited from "good" pixels from multiple acquisitions. For this to be effective, the data often need to be in physical unity (reflectance, temperature, etc.) to be useful. Thus, while not a primary objective, this effort will provide support for compositing activities.

3.3.6   Landsat 8

The Landsat Data Continuity Mission (LDCM) launched 11 February 2013 and officially began normal operations, when operations transferred from NASA to USGS along with the name changing from LDCM to Landsat 8, on 30 May 2013. With Landsat 5 ailing and Landsat 7 operating in SLC-off mode, the launch of LDCM and operation of Landsat 8 is imperative to avoid a data gap. Landsat 8 carries the Operational Land Imager (OLI), with nine spectral bands, including a panchromatic band, and for the first time, a separate thermal instrument, the Thermal Infrared Sensor (TIRS), which has two thermal bands, as shown in Table 3.3 [USGS, 2013].

Table 3.3: Spectral bands of the Operational Land Imager and Thermal Infrared Sensor on board Landsat 8 [USGS, 2013].

Band	Spectrum Area	Response (μm)	Resolution
Operational Land Imager (OLI)			
Band 1	Visible	0.43 - 0.45	30 m
Band 2	Visible	0.450- 0.51	30 m
Band 3	Visible	0.53 - 0.59	30 m
Band 4	Red	0.64 - 0.67	30 m
Band 5	NIR	0.85 - 0.88	30 m
Band 6	SWIR	1.57 - 1.65 (High and Low Gain Options)	60 m
Band 7	SWIR	2.11 - 2.29	30 m
Band 8	Panchromatic (PAN)	0.50 - 0.68	15 m
Band 9	Panchromatic (PAN)	1.36 - 1.38	30 m
Thermal Infrared Sensor (TIRS)			
Band 10	TIRS 1	10.6 - 11.9	100 m
Band 11	TIRS 2	11.5 - 12.51	100 m

The LST methodology discussed in Chapter 4 could also be used in an identical fashion with a single band of Landsat 8. However, due to ongoing concerns with the calibration of Landsat 8, no effort to test against Landsat 8 was included in this study. This can be addressed when the Landsat 8 calibration knowledge is well understood and stable.

3.4 Land Surface Temperature

The land surface is the first solid surface between the lowest layer of the atmosphere and the Earth. From the point of view of thermal imaging from a satellite, this is generally considered to be a few millimeters thick and could consist of forest and shrubs, crops, grasslands, bodies of water, wetlands, ice or snow, barren or desert, urban, bare soil, bedrock, sand or sediments [Wan and Dozier, 1996]. Although not immediately obvious, the temperature of this land surface is individually an important data record as well as a tool used in obtaining and analyzing other variables. Identified by NASA as an important Earth System Data Record as part of the Earth Observing System (EOS) program [King, 1999], LST can be utilized in a number of fields for a wide variety of applications.

LST products from a number of different satellites are already produced and various methods exist for obtaining LST from a single image for a number of different instruments. We begin this section by summarizing various applications for a land surface temperature product. We then review both multiple channel methods and single channel methods as a basis for existing work and accuracy in the field. Finally, we discuss both validation methodologies and error analysis techniques that have been used for previous land surface temperature algorithms and products.

3.4.1 Applications

Land surface temperature results from interactions with and energy fluxes from the ground, making it a variable with far-reaching uses and applications in the physics of land surface processes [Sellers et al., 1988]. On large scales, it can be used to evaluate land surface energy balance [Diak and Whipple, 1993]. Because land surface temperature results from interactions between the atmosphere and the ground and is affected by various environmental variables, it is useful in terrestrial biosphere dynamics, change detection, hydrologic balances, and biogeochemistry of greenhouse gases. Hydrologic processes, such as evapotranspiration and snow melt, are sensitive to LST, as well as climate change and carbon cycles. Ecological processes, mostly associated with agriculture and the growing season, such as leaf phenology, photosynthesis, respiration and decomposition can all be affected by LST. Finally, LST not only affects land cover change but can also be used to delineate regional land cover classes [Running et al., 1994]. It can be applied in land cover and land cover change analysis [Ehrlich and Lambin, 1996].

LST is also invaluable in climate studies and meteorological research. There are various surface properties that are required to estimate energy, momentum, and moisture fluxes at Earth's surface used in numerical meteorological predictions. These surface properties, such as LST, can be estimated from satellite data, to be used in climatology and weather science [Price, 1982]. LST can also be used to study other weather patterns, such as the monsoon season in Asia, where the land-sea temperature contrast is a critical concept. Stronger summer monsoons, and lower land albedos, are associated with greater land-sea temperature contrast and warmer land temperature among other variables [Meehl, 1994].

On smaller scales, LST can be used to determine moisture availability and absorbed radiation used in estimating sensible and latent heat fluxes [Kimuru and Shimiru, 1994]. There are many applications in agriculture, specifically uses in canopy temperature or soil temperature. Canopy temperatures can be used to estimate sensible heat flux [Vining and Blad, 1992], evaluate water requirements [Jackson et al., 1977], and determine frosts in orange groves [Caselles and Sobrino, 1989]. Soil temperatures can be used to monitor and prepare response to drought [Feldhake et al., 1996]. In summary, uses of LST are highly variable and many are well-matched to the coverage and resolution of the Landsat data, giving us confidence in the purpose of our work.

3.4.2 Multiple Thermal Bands

The most commonly used LST algorithms require two adjacent spectral bands. Instead of utilizing atmospheric compensation that requires accurate atmospheric profiles like the single band methodology, the split-window techniques use differences in absorption in adjacent thermal bands to make the necessary corrections for atmospheric effects. The goal of this section is to provide a brief summary of the vast number of split-window algorithms and techniques. Although these cannot be applied to Landsat data, we aim for a general understanding of the types of satellites from which LST can be retrieved and the accuracy that is being achieved.

Wan and Dozier (1996) propose a generalized split-window technique using regression analysis

and radiative transfer simulations. They point out that the success of any LST algorithm depends on the atmospheric compensation, the characterization of the surface emissivity, and the quality of the thermal infrared data. With the Landsat and ASTER databases, the basis of this research assumes high radiometric data quality and surface emissivity, and therefore focuses solely on dealing with the atmospheric effects. This generalized split-window algorithm utilizes differential absorption in adjacent thermal bands. The split-window algorithms are less sensitive to errors in atmospheric optical properties because they do not rely on the absolute atmospheric transmission of a single band. They do, however, require radiative transfer simulations over a wide range of atmospheric and surface conditions in order to generate the necessary coefficients [Wan and Dozier, 1996].

Because of the requirement for both atmospheric compensation and emissivity data, many LST algorithms are not operational. Vazquez et al. (1997) compares four split window algorithms for the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) using directly measured 5 cm subsurface temperatures of soils in an area of low water vapor and low probability of cloudiness (good conditions) as ground based truth validation. However, all methods require a priori knowledge of the surface emissivity. For these four split-window algorithms, the root mean square deviations from ground based temperatures are 3.8 K, 3.0 K, 2.3 K, and 1.9 K and the mean bias deviations are 3.3 K, 1.8 K, 0.1 K, and 0.7 K respectively; all showed an overestimation of the high, summer morning temperatures and had maximum deviations between 4 K and 8 K [Vazquez et al., 1997]. This shows that while LST retrieval from AVHRR data is feasible, it requires accurate knowledge of surface emissivity and can be highly variable.

Qin et al. (2001) also discuss the derivation of a split window algorithm for NOAA AVHRR data. AVHRR sits on-board NOAA polar-orbiting meteorological satellites to monitor global meteorological change. While various algorithms have been explored to retrieve LST from NOAA AVHRR data, as shown above, this paper explores an algorithm that determines the necessary coefficients based on atmospheric transmittance, derived using LOWTRAN simulation, ground emissivity, estimated from other AVHRR bands as in [Sobrino et al., 2001], and viewing angle; they claim this algorithm has improved accuracy while requiring fewer parameters. Validation with atmospheric simulation indicates an accuracy of $0.25°$ C when transmittance and emissivity are perfectly known. With ground truth data, when atmospheric water vapor contents are not perfectly known, the algorithm has an accuracy of $1.75°$ C. Finally, for a ground truth data set with corresponding in situ water vapor contents, the accuracy of the algorithm is $0.24°$ C [Qin et al., 2001a].

Sun and Pinker (2003) discuss algorithms for use with the Geostationary Operational Environmental Satellite (GOES). While this satellite provides good spatial coverage and would allow for frequent estimates of LST, these algorithms are not implemented operationally due to the requirement of emissivity and the variation of coefficients with emissivity and atmospheric water vapor. They compare previously published generalized split-window algorithms, with and without water vapor correction, to a newly proposed algorithm and a three-channel approach. The generalized split-window algorithm has errors greater than 0.5 K, and greater than 1 K at temperatures over

290 K. The split-window algorithm with water vapor correction was an improvement with error less than 0.5 K for temperatures less than 290 K. The newly proposed algorithm was also a slight improvement over these results. The three-channel algorithm, for nighttime retrieval, shows improvement over the generalized split-window approach. Errors in the retrieved LST also differ with season, with an RMS error of 2.3 K in the summer but an RMS error as low as 1.38 K in the winter [Sun and Pinker, 2003].

Yu and Privette (2005) analyze a land surface temperature algorithm for NPOESS VIIRS data. VIIRS, in sun-synchronous orbit on board NPOESS, has 750 m pixels. NPOESS VIIRS LST algorithms include a daytime dual split window (DSW) algorithm, a nighttime DSW algorithm, and a backup split window algorithm. While the split window algorithms exploit differential absorption or atmospheric water vapor between adjacent bands, the DSW algorithms use two short-middle infrared bands and two thermal infrared bands to generate the necessary coefficients. All of these algorithms use linear brightness temperature as base estimators of land surface temperature, and the remaining terms provide corrections for atmospheric attenuation and viewing geometry. VIIRS system specifications require a precision error (standard deviation) of ≤ 0.5 K, an accuracy (mean bias) of ≤ 2.4 K and an uncertainty error (RMS) of ≤ 2.5 K. The VIIRS LST algorithms were applied to MODIS data and compared to MODIS products to generate a comparison of the difference between products, although this does not provide absolute accuracy. The results suggest product differences are ≤ 2 K and spatially distributed and that the split window algorithms are better than the DSW algorithms for both daytime and nighttime [Yu and Privette, 2005].

The Earth Observing System (EOS) Moderate Resolution Imaging Spectroradiometer (MODIS) instrument produces a daily land surface temperature product using multiple thermal bands. MODIS can provide global coverage, high spectral resolution, and accurately calibrated data. MODIS utilizes multiple bands in atmospheric windows for its LST retrieval; it implements a generalized split-window algorithm and a physics-based day/night algorithm. With seven available thermal infrared bands, this algorithm can adjust for uncertainties in temperature and water vapor profiles without simultaneous retrieval of surface data or atmospheric variable profiles. Emissivity is also immediately required for an operational product, so MODIS estimates classification-based emissivities from land-cover types using thermal infrared BRDFs and emissivity modeling. Over certain land cover types in the range of 263 K to 300 K, the MODIS LST can be better than 1 K, but can underestimate temperatures in semi-arid regions due to inaccuracies in the estimated surface emissivity [Wan et al., 2004]. MODIS does have lower spatial resolution than Landsat, which makes this product difficult to apply in certain applications that require LST, such as field specific agriculture or irrigation studies.

Gillespie et al. (1998) describe the three-part temperature and emissivity separation (TES) algorithm used for ASTER imagery. First, an iterative approach is used to estimate emissivities and temperature using the NEM Module, then the Ratio Module is used to calculate the emissivity band ratios, and finally surface temperature is recalculated using the atmospherically corrected radiance and new emissivity values. Numerical simulations and validation against simulated ASTER

imagery suggest the algorithm will perform within the 1.5 K specifications [Gillespie et al., 1998]. Although both MODIS and ASTER have LST products with validated performance, Hulley and Hook (2011) point out that they require trade offs in spatial or temporal resolution and are difficult to compare because of the different algorithms used to produce them. They propose a consistent land surface temperature and emissivity product by generating results using the TES algorithm [Gillespie et al., 1998] on both ASTER and MODIS data. This combines the higher resolution but infrequent revisit cycle (90 m pixels, every 16 days) of ASTER with the lower resolution but high temporal frequency (5 km, daily) of MODIS. Comparing LST retrieval over two areas at the Algodone Dunes and the Salton Sea, they found the products to match closely with differences of less that 1 K [Hulley and Hook, 2011].

Finally, Freitas et al. (2011) describe a study with the goal of generating near real time LST data based on a constellation of geostationary satellites. Although this study does not include a full validation of the LST products, it is particularly interesting for the fusion of products from different satellites and real time generation of LST data with error analysis. They consider the Meteosat Second Generation (MSG), Geostationary Operational Environmental Satellite (GOES), and Multifunction Transport Satellite (MTSAT). The generalized split window algorithm from Wan and Dozier (1996) is applied to SEVIRI data, a sensor on board MSG. The accuracy of these LST retrievals varied considerably with satellite view angle and atmospheric water vapor content. A dual algorithm, designed for sensors with one mid-infrared (MIR) band and one thermal infrared (TIR) band, is applied to GOES and MTSAT data. Because MIR data can introduce a large amount of uncertainty from contamination by solar radiation, the dual algorithm applies a two channel technique for nighttime data, and a mono-channel method for daytime data requiring a single TIR band. In both cases, coefficients are fitted to a calibration data set for a variety of total column water vapor, satellite zenith angle, and land-cover type. As expected, the two-channel method can produce results similar to the generalized split window technique, but curve-fitting coefficients for single channel retrieval is associated with considerable uncertainty increase. This provides evidence of the requirement of more than one channel for proper atmospheric correction using this technique. In order to select the proper coefficients in the dual algorithm, for one or two channels, TOA brightness temperature, forecasted total column water vapor, land-cover classification, and viewing angle are required. Comparison to other sources, and consistency in overlapping areas, suggest errors within the 2 K range for the generalized split window and two-channel approach, with larger errors for the mono-channel method [Freitas et al., 2011].

3.4.3 Single Thermal Band

As mentioned above, and noted by Sobrino et al. (2004), one thermal band limits the ability to retrieve LST due to the inability to apply a split-window algorithm or a temperature-emissivity separation.

As shown in Freitas et al. (2011), attempts to generate coefficients for a calibration data set with a single thermal band lead to an increase in uncertainty [Freitas et al., 2011]. Sun et al.

(2004) consider LST estimation for GOES satellites M-Q which will have only a single thermal band. They propose a single channel method that requires an estimation of total precipitable water and a two channel method, the second of which is a MIR channel. They found, like Freitas et al. (2011), that the two channel algorithm is comparable to the Wan and Dozier (1996) generalized split window algorithm, while the one channel method is less accurate. These algorithms are better for sea surface temperature derivation when the assumptions of surface emissivity are less restrictive [Sun et al., 2004].

Sobrino et al. (2004) compare three methods of temperature retrieval using Landsat thermal band data. The first uses radiosonde data and the radiative transfer equation, the second Qin et al.'s (2001) mono-window algorithm, and the third Jimenez-Munoz and Sobrino's (2003) single-channel method. They propose obtaining land surface emissivity using the NDVI method, which requires atmospheric compensation of Landsat bands 3 and 4 for the most accurate results. They use the LST derived from radiosonde data and in situ emissivity measurements as "truth," or a basis for comparison to all other methods, illustrating that these are considered the most accurate even though the radiative transfer equation still requires the use of MODTRAN to generate the necessary atmospheric parameters. Qin et al.'s (2001) mono-window algorithm requires not only emissivity, but also an estimate of water vapor content and near-surface temperature to calculate atmospheric transmissivity. Jimenez-Munoz's (2003) single-channel method also requires emissivity and an estimate of water vapor content [Sobrino et al., 2004], [Jimenez-Munoz and Sobrino, 2003], [Qin et al., 2001b].

They found that compared to using the radiosonde with the radiative transfer equation and in situ emissivity measurement, Qin et al.'s algorithm with emissivity from NDVI has a root mean square deviation (RMSD) of 2.2 K and Jimenez-Munoz and Sobrino's single channel method has a RMSD of 0.9 K [Sobrino et al., 2004].

Jimenez-Munoz and Sobrino (2004) provide a study of the error contributed to land surface temperature by several parameters based on using MODTRAN to simulate various conditions. Such parameters include atmospheric compensation, sensor noise, land surface emissivity, aerosols and other gaseous absorbers, angular effects, wavelength uncertainty and band-pass effects. Of particular interest to this work, it was found that atmospheric effects are the most important source of error and could introduce errors of 0.2 K if in situ data is used and 0.7 K if remote sensing data is used. Also, uncertainty in emissivity can lead to errors of 0.4 K, so a minimum error of 0.3 K can be obtained with all in situ data, and a minimum error of 0.8 K is expected when all remote sensing data is used [Jimenez-Munoz and Sobrino, 2004].

3.4.4 Validation Methodologies

For validation of our methodology, we will use water temperatures from buoys, corrected to the skin temperature, as ground truth data. As a basis for this decision, as well as possible future work, we explore here various validation methodologies used to evaluate results in the development and testing of other land surface temperature algorithms and products.

Vazquez et al. (1997) use images located near the Meteorological Office of the Air Force, consisting mostly of grassland with some patches of bare soil, to validate results for LST algorithms for NOAA-12 HRPT images. The Meteorological Offices measures subsurface soil temperatures at 5 cm with an integration time of 30 min; the collections closest to that of satellite overpass were used. They claim that the morning satellite overpasses correspond to the times of minimum difference between the 5 cm subsurface temperature and the skin temperature. They found RMS deviations of approximately 2 K to 4 K [Vazquez et al., 1997].

Sun and Pinker (2003) validate estimations of LST from GOES-8 using three different data sources. They utilize the Atmospheric Radiation Measurement (ARM) observations of surface skin temperature using infrared thermometers and outgoing long wave (LW) radiation from the Central Facility in Southwest Oklahoma, observations of soil temperature from soil probes and air temperature from automated weather stations from the North Carolina Agricultural Research Service (NCARS) Weather and Climate Network, and the Surface Radiation Network (SURFRAD) upwelling thermal infrared radiances measured by Precision Infrared Radiometers (PIR) at four different stations. Upwelling thermal irradiance from SURFRAD first needs to be converted to skin temperature, which requires an estimation of the surface emissivity. In all cases, they compare the ground truth to the proposed algorithm as well an established algorithm and found errors ranging from 0.5 K to slightly greater than 2 K, as discussed in Section 3.4.2. Using a variety of data sources against multiple algorithms allows them to study the difficulties with different types of validation data sets, as well as the performance of their algorithm [Sun and Pinker, 2003].

Yu et al. (2012) also validates the GOES-R ABI LST algorithm using SURFRAD data. They use statistical analysis to evaluate the error of both the satellite and ground truth data. Using more than 100 points from one year of SURFRAD data over each of six sites, they use a linear fitting model to estimate the precision of the two datasets. Rather than assuming the ground truth is correct data, this approach can determine the precision of the satellite and ground truth measurements and consider this precision in relation to the slope of the two datasets [Yu et al., 2012].

Wan et al. (2004) described the validation of the MODIS LST retrieval using a series of specific field campaigns as described in [Wan et al., 2002a] and [Wan et al., 2002b]. TIR radiometers measure lake surface kinetic temperatures at various locations, radiosonde balloons are launched from the lake shore, and winds speeds and air temperatures are recorded. Based on the surface kinetic temperature and this data using MODTRAN radiative transfer code, the lake surface temperature can be determined [Wan et al., 2002a]. Field campaigns using this same technique 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. With these extensive field campaigns, the accuracy of the measurements and estimation of ground truth data was thoroughly explored and they found they could retrieve temperature better than 1 K as discussed in Section 3.4.2 [Wan et al., 2004].

Gillespie et al. (1998) validate the ASTER TES algorithm using numerical simulations. These simulations allow validation against error free input radiances for a variety of atmospheric and

surface conditions and they found errors within the 1.5 K specifications [Gillespie et al., 1998]. Hulley and Hook (2011) validate their temperature and emissivity retrievals for both ASTER and MODIS using intercomparisons for two sites as well as laboratory spectra for the two sand dune sites and found they could validate temperature within 1 K [Hulley and Hook, 2011].

Yu and Privette (2005) use existing satellite data rather than ground truth data to validate the NPOESS VIIRS land surface temperature algorithm. The VIIRS LST algorithms were applied to MODIS radiance data and these LST retrieval results were compared to the MODIS Level 2 LST swath product. While this comparison reveals only the difference between the two products, rather than an absolute accuracy, it does create the ability to evaluate the performance of the algorithm over a large area in various atmospheric conditions [Yu and Privette, 2005].

3.4.5 Error Analysis

In this section we aim to summarize the type of error analysis information that is provided to users in current LST products from other sensors in order to better understand the type of information being used and the precision of uncertainty being predicted. These were used to inform the methods we explored for error analysis of this product. As we learned, traditional error analysis is difficult, in our process and many LST retrievals, because of the incorporation of various data sets and transfer codes.

The ASTER TES algorithm aims to capture the quality of their output in three eight-bit quality assurance (QA) data planes. Numerical simulation and execution on both lab and field simulated data are used to evaluate the algorithm; most information in the QA data planes is inferred from intermediate values in the algorithm and results of these evaluations. Numerical simulations indicate the performance of the algorithm with changing ground temperature and changing emissivity and the sensitivity to NEΔT, sky irradiance, and atmospheric compensation [Gillespie et al., 1998]. The first data quality plane is common to all ASTER products, although there are product specific bit patterns. The first four bits categorize data quality as "good," "bad," or "suspect" based on input data or algorithm completion. A "good" pixel would have no known defects. A "suspect" TES may have a bit pattern that indicates some output bands were out of range. A "bad" TES may have a bit pattern that indicates too few good bands or that the algorithm diverged rather than converged. The next two bits are the cloud mask; this indicates thick clouds, thin clouds, or clear conditions estimated using ASTER, VNIR, and SWIR data. The last two bits are the adjacency code which predicts what percentage of the radiance is uncorrected cloud irradiance. This is categorized as less than 10%, 10-20%, 20-30%, or may exceed 30%.

The second quality data plane has information on both temperature and emissivity. The first two bits specify scene conditions based on ϵ_{max}. They indicate possible error conditions when $\epsilon_{max} \leq 0.94$, silicate rocks when $0.94 < \epsilon_{max} < 0.96$, and that the default ϵ_{max} value is used when $0.96 < \epsilon_{max} < 0.98$. When $\epsilon_{max} \geq 0.98$, the scene is likely vegetation, snow, water, and some soils. The second two bits indicate fast to converge, nominal performance, or slow to converge based on the number of iterations (4, 5 or 6, more than 7). The next two bits use the ratio of the downwelling

atmospheric irradiance (normalized by π) to the measured land leaving radiance to draw conclusions about the atmospheric correction. If the ratio is ≤ 0.1, it is likely a high altitude scene and the correction is probably accurate, 0.1-0.3 are called nominal values and if the ratio is > 0.3 it is likely warm, humid air or cold land and the correction may be inaccurate. Finally, the last bit indicates if ϵ_{min} needed to be corrected or reduced in proportion to measurements [Leff, 1999]. An explicit description of bits in this plane make it more clear how intermediate values, in some cases related to numerical simulation results, can be predictors or indicators of quality.

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].

The MODIS LST product (at 1 km spatial resolution for a swath) is generated from MODIS calibrated and geolocated radiances using the generalized split window algorithm (Wan and Dozier, 1996); this product also utilizes MODIS geolocation, cloud mask, atmospheric profile, land cover and snow cover products [Wan, 2007]. As described in the algorithm theoretical basis document, there are instrument, algorithm, and emissivity contributions to uncertainty. These include calibration accuracy, spectral response function, optical and system noise equivalent temperature, and pointing knowledge and accuracy from the instrument, uncertainty in the generalized split window algorithm and in the day/night registration for the day/night algorithm, and some uncertainty in the emissivity associated with each land cover type in the emissivity knowledge base. A root sum square (RSS) of these uncertainties, varying with view angle and column water vapor, is calculated to estimate an error in the LST value [Wan, 1999].

The goal of the MODIS LST product is to provide algorithm and data quality that can be viewed in a spatial content to help determine how useful each result is for a particular user. This is achieved by presenting results in scientific data sets (SDS) with per pixel values along with accompanying local and global attributes. Of particular interest are the SDS for quality control (QC) and land surface temperature error. It is noted that the LST error is only an estimated value which is generally conservative for clear conditions and does not consider cloud contamination. The QC SDS consists of a 16-bit unsigned integer for each pixel. The first two bits are a quality assurance flag, indicating if the pixel was produced and if it is necessary to further investigate the QA. The next two bits indicate data quality based on input data and calibration, the next two whether the pixel is cloud free, thin cirrus, sub-pixel fraction of clouds, or affected by nearby clouds (not produced if cloudy), and the next four bits indicate which LST algorithm was used and in what capacity it was run. The final six bits deal with the emissivity; the first two indicate how the emissivity was defined (land cover, MODIS, default), the second two how the emissivity quality was checked, and the final two a categorization of the magnitude of the uncertainty of the two emissivities by land cover type. Unlike ASTER, MODIS provides an estimated value of the error [Wan, 2007].

The algorithm theoretical basis document for the GOES-R Advanced Baseline Imager (ABI)

LST product describes surface emissivity and atmospheric water vapor absorption as important error sources. They also discuss the effects of sensor view angle and expect largest errors with moist atmospheric conditions and a large local zenith angle. The LST product includes the LST values, product quality information and quality control flags; however, these simply provide information from within the algorithm, rather than providing an estimated error value or drawing conclusions about the effects on the usability of the output. The product quality information is defined for each pixel in 16 bits. The first byte defines availability (normal, bad data, missing data), surface type (land, snow/ice, in-land water, sea), and the cloud index (clear, probably clear, probably cloudy, cloudy), generated with the ABI cloud mask, an independent GOES-R product. The second byte defines the atmospheric condition (dry, moist, very moist) based on the water vapor, day/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). The quality control flags for each pixel contain a subset of the same information [Yu et al., 2010].

Similarly, the NPOESS VIIRS LST EDR software includes three bytes of quality information, generated from various inputs and intermediate values. The LST product is generated using either a 4-band dual split window or 2-band split window algorithm with different coefficients for each land cover type. VIIRS brightness temperature, optical thickness, cloud mask, and surface type are all used to generate the LST product. The first byte of the three byte quality data for each pixel includes the LST quality (high, medium, low, or no retrieval) and algorithm (4-band or 2-band), both determined using a logical combination of other LST flags, and the availability of the short wave and long wave IR brightness temperatures, determined from the corresponding brightness temperature products. It also includes bits for day/night, active fire (or not), and thin cirrus (or not), all determined from the VIIRS cloud mask. The second byte includes the clear measurement precision degradation, from the brightness temperature product, if the retrieved LST is within the acceptable range 213 K $\leq$ LST $\leq$ 343 K (or not), a confidence in cloudy or clear conditions from the cloud mask, if the aerosol optical thickness is considered within range ($\leq$ 1.0 or not), from the optical thickness product, the horizontal reporting interval based on the sensor zenith angle from the cloud mask, sun glint from the cloud mask, and terminator based on the solar zenith angle from the cloud mask. Finally, the third byte classifies the type of land/water background from the cloud mask and the surface type from the surface type product [Ip and Siebels, 2009]. Although some conclusions are drawn on quality effects, most are simple reproductions of information from inputs and no estimate of quantitative error is provided in the final product.

As indicated above, most operational algorithms provide quality information that categorize or summarize information from inputs or within the LST algorithm to indicate to the user the quality, usefulness, or trustworthiness of the LST data being provided. This is insightful for what information is used as inputs and what information is considered useful to users in the final product. In some cases conclusions on the effects are provided while in other cases information is simply presented, but it is rare to have a quantitative estimate of error and when it is provided it is a conservative estimate at best.

Freitas et al. (2011) describe the development of a LST product from multiple geostationary satellites. They attempt to generate LST in near real time and provide an error bar associated with each LST value. The error bar for each LST value considers input errors of sensor noise, uncertainty in emissivity (land cover) and statistics of total column water vapor forecast errors, as well as uncertainties in the retrieval algorithm, which is mostly influenced by the optical path between sensor and surface, dependent on viewing angle and column water vapor. They determine the sensitivity to each of these variables with radiative transfer simulations. By comparing the results from both exact and inaccurate inputs to a model, the sensitivity to each variable can be explored. With modeling to determine error from each input, the LST error bar is determined by assuming all errors sources are independent [Freitas et al., 2010] [Freitas et al., 2011].

Similarly, Hulley et al. (2012) describe the development of a temperature and emissivity uncertainty simulator. The goal is to accurately quantify land surface temperature and emissivity uncertainties for any sensor and algorithm combination under a wide range of atmospheric and surface conditions. They utilize MODTRAN radiative transfer code along with a global set of radiosonde profiles from Wyoming CLAR database and emissivity from the ASTER spectral library. To calculate various sources of error, simulations were run with actual and adjusted atmospheres (atmospheric noise), perfect and imperfect simulated TOA radiances (measurement noise), and perfect inputs for simulated LST to compare to retrieved LST (model error due to assumptions in LST algorithm). The root sum squared of these errors provides an estimate of total LST uncertainty as shown in Equation 3.5, where LST_A is the uncertainty contributed by atmospheric noise, LST_N is the uncertainty contributed from the measurement noise, and LST_M is the uncertainty contributed by the model noise.

$$\delta LST_{TES} = [\delta LST_A^2 + \delta LST_N^2 + \delta LST_M^2]^{\frac{1}{2}} \qquad (3.5)$$

For a specific sensor and algorithm, a least squares regression is performed between this simulated total LST uncertainty and a quadratic function of total column water vapor (TCW) (and other error contributors where applicable, sensor view angle (SVA) for example) as shown for MODIS in Equation 3.6.

$$\delta LST_{MODIS} = a_0 + a_1 TCW + a_2 SVA + a_3 TCW \cdot SVA + a_4 TCW^2 + a_5 SVA^2 \qquad (3.6)$$

Coefficients from this parameterization can then be applied to each pixel in a scene, given estimates of total column water vapor (and sensor view angle or other applicable error contributors). They found atmospheric errors were the largest source of error for both sensors they tested [Hulley et al., 2012].

Finally, Hook et al. (2007) also suggest a similar method of determining LST uncertainty by considering the contributions of the error in the atmospheric variables. By perturbing the atmosphere by the atmospheric uncertainty, and conducting simulations, the contributions to the

error in the final LST can be considered. Final errors in the LST were determined by calculating a nominal radiance, adjusting the profile and recalculating radiance, and comparing the corresponding at sensor temperature for the nominal and adjusted radiances. Adjustments were made to water vapor, air temperature, ozone, and visibility, as well as the path length and assuming incorrect emissivity. They found the largest uncertainty with changes in visibility and column water vapor [Hook et al., 2007]. These studies indicate that in the thermal region we expect errors from the atmospheric compensation (rather than emissivity) to dominate uncertainty in the LST product.

3.5 MODTRAN

MODTRAN radiative transfer code, created by Spectral Science Inc. and the United States Air Force, was developed from LOWTRAN, the original low resolution version of the program. MODTRAN radiative transfer code utilizes a propagation model that assumes the atmosphere is divided into a number of homogenous layers [Schott, 2007]. The user must input, or select pre-defined, vertical atmospheric profiles for parameters such as pressure, temperature, and humidity, and specify visibility, season, or time of year among other inputs. MODTRAN solves the radiative transfer equation to characterize molecular and particular absorption, emission, and scattering, as well as reflections, emissions and transmissions among other outputs [SSI, 2012].

The applications and uses of MODTRAN are far-reaching, but for the purpose of this work, atmospheric profiles of height, temperature, pressure and humidity were input in order to calculate from the MODTRAN output the transmission, upwelled radiance, and downwelled radiance, which are not explicitly given. There are various methods that can be used to calculate these three values from the spectral arrays contained in the output, as detailed in Appendix E, but contributions to error as well as processing time and memory constraints indicated the method detailed in Section 4.1 is optimal for the implementation of this operational algorithm.

3.6 NARR Data Set

Reanalysis, or retrospective-analysis, is the process of using observing systems with numerical models to generate, in a spatially and temporally consistent set, a set of variables that are not easily observed or measured [Rienecker and Gass, 2013]. Reanalysis data is a regional or global estimate to take empirical data from various inputs and re-estimate characteristics of the atmosphere on a regular spatial and temporal grid. The NARR dataset is produced by the National Center for Environmental Prediction (NCEP). This dataset is an extension of the NCEP global analysis, improving resolution and the number of variables over North America. Inputs to NARR include radiosondes, dropsondes, pibals, aircraft, surface data, and cloud drift winds. NARR provides data for sea level, surface level, specified pressure levels, specified heights above ground, hybrid level, and below surface as well as wind data, cloud data, tropopause, and atmospheric columns [Shafran, 2007].

This work uses variables provided at specific pressure levels. The NARR data includes geopotential height [gpm], temperature [K], specific humidity [kg/kg], pressure vertical velocity [Pa/s], u wind [m/s], v wind [m/s], cloud water [kg/kg], ice mixing ratio [kg/kg], and turbulent kinetic energy [J/kg] at pressure levels of 1000 hPa, 975 hPa, 950 hPa, 925 hPa, 900 hPa, 875 hPa, 850 hPa, 825 hPa, 800 hPa, 775 hPa, 750 hPa, 725 hPa, 700 hPa, 650 hPa, 600 hPa, 550 hPa, 500 hPa, 450 hPa, 400 hPa, 350 hPa, 300 hPa, 275 hPa, 250 hPa, 225 hPa, 200 hPa, 175 hPa, 150 hPa, 125 hPa, and 100 hPa. These data are provided in a 349 by 277 array on the Lambert Conformal Conic grid. This is roughly 0.3° or 32 km spacing at the lowest latitude. The corners of this spatial coverage are (12.2°N, 133.5°W), (54.5°N, 152.9°W), (57.3°N, 49.4°W), and (14.3°N, 65.1°W), covering North America. There are three different temporal resolutions: eight times daily, once daily, or once monthly depending on the variable and desired temporal coverage. NCEP currently provides this data in the original GRIB format starting 1 January 1979 with plans for continuing coverage and data provision. Data for this work was downloaded as GRIB files using the NOMADS data access FTP or HTTP site at `nomads.ncdc.noaa.gov/data.php?name=access#narr_datasets` [NOMADS, 2012].

The geopotential height, air temperature, and specific humidity will all be used at 29 pressure levels and eight times daily to input atmospheric profiles into MODTRAN. However, MODTRAN accepts only certain variables in certain units and therefore these data will need to be converted to the proper input variables.

3.7 Conversions

MODTRAN requires the temperature, pressure, and some humidity variable, each defined at the same heights, to characterize the atmospheric profiles. The heights can be at any spacing as long as they provide sufficient coverage to accurately characterize the desired atmospheric column and are the same for all provided variables. The atmospheric variables from the NARR data at specified pressure levels can be used to characterize the atmospheric profiles in MODTRAN.

The pressure and air temperature can be input into MODTRAN as given in the NARR data in hPa and K respectively. However, it is necessary to have a corresponding geometric height [km] and one humidity variable. MODTRAN accepts the volume mixing ratio [ppmv], number density [molecules/cm^3], mass mixing ratio [g/kg], mass density [g/m^3], partial pressure [mb], dew point temperature [K or °C], or relative humidity [%].

The geopotential height, provided by the NARR dataset, is the height of a given point in the atmosphere in units proportional to the potential of unit mass or geopotential at that height relative to sea level. There is an adjustment to the geometric height using the variation of gravity with latitude and elevation. The geometric height, the desired MODTRAN input, is simply the elevation above mean sea level. Therefore, the conversion from geopotential height to geometric height requires knowledge of the latitude of the location. Variables and constants for this conversion are summarized in Table 3.4.

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

Variable	Data	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]

The acceleration due to gravity, g [m/s^2], at the desired latitude must be computed as shown in Equation 3.7 and the gravity ratio, G, as shown in Equation 3.8.

$$g = 9.80616[1 - 0.002637cos(2\phi) + 0.0000059cos^2(2\phi)] \tag{3.7}$$

$$G = \frac{g}{g_0} \tag{3.8}$$

Finally, the radius of the Earth at the desired latitude, R_e [km], must be solved for from Equation 3.9.

$$R_e^2\left(\frac{cos^2(\phi)}{R_{max}^2} + \frac{sin^2(\phi)}{R_{min}^2}\right) = 1 \tag{3.9}$$

With these values, the desired geometric height, Z [m], can be calculated as shown in Equation 3.10 [Wright, 1997].

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

The humidity variable given in the NARR dataset, specific humidity, is not directly accepted by MODTRAN to characterize the atmospheric profiles. Through numerous intermediate steps, relative humidity can be calculated from specific humidity given the corresponding temperatures and pressures. Variables and constants for this conversion are summarized in Table 3.5.

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

Variable	Data	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_{H_2O}	molar mass of water	18.01534 [g/mol]
M_{dry}	molar mass of dry air	28.9644 [g/mol]

The Goff-Gratch equation, shown in Equation 3.11, was selected from various methods to calculate the saturation water vapor pressure, e [hPa] [Goff and Gratch, 1946].

$$
\begin{aligned}
\log(e) = & -7.90298\left(\frac{373.15}{T_K} - 1\right) \\
& + 5.02808 log\left(\frac{373.15}{T_K}\right) \\
& - 1.3816 \text{ x } 10^{-7}(10^{11.344(1-\frac{T_k}{373.15})} - 1) \\
& + 8.1328 \text{ x } 10^{-3}(10^{-3.49149(\frac{373.15}{T_K}-1)} - 1) \\
& + \log(1013.25)
\end{aligned}
\tag{3.11}
$$

The volume mixing ratio, X_{H_2O}, is required to calculate the partial pressure, P_{H_2O} [hPa] shown in Equations 3.12 and 3.13 respectively.

$$
X_{H_2O} = \frac{qM_{dry}}{M_{H_2O} - qM_{H_2O} + qM_{dry}}
\tag{3.12}
$$

$$
P_{H_2O} = pX_{H_2O}
\tag{3.13}
$$

And finally, the desired relative humidity, RH [%], can be calculated from the partial pressure and saturation water vapor pressure using Equation 3.14 [Kruger, 2010].

$$
RH = \frac{P_{H_2O}}{e} * 100
\tag{3.14}
$$

3.8 Concluding Remarks

Chapter 3 aimed to provide a summary of all of the information that will be utilized in our land surface temperature retrieval methodology. We began by presenting thermal radiance and the governing equation in Sections 3.1 and 3.2. This provides the basic physical and scientific theory that is the basis for our work. We then provided a brief history of Landsat and applications for and methods of LST retrieval in Sections 3.3 and 3.4. This historical information and investigation into already developed methods is important to understanding where our work fits in the field. Finally, in Sections 3.5, 3.6, and 3.7, we briefly discuss programs, datasets, and calculations with which we will assume familiarity throughout the rest of the work.

We use all of the above information to present our initial approach and methodology in Chapter 4. We explain each step in the process and briefly provide support or validation for each one. More comprehensive validation and verification is provided in Chapter 5. This includes both the errors in our methodology and methods for the assignment of confidence metrics. Finally, in Chapter 6, we detail future work, which includes extension to a global product and finalizing the confidence metric implementation.

Chapter 4

Methodology and Approach

Chapter 3 includes the basic scientific process and other information necessary for land surface temperature retrieval, but Chapter 4 explains the specifics of the approach for this work. Section 4.1 describes the process chosen for calculating the necessary radiative transfer parameters from the MODTRAN output. This procedure is central not only to the final product but also the development, validation, and verification of the process and will be used and referenced throughout. It is difficult to understand how each step contributes to the process without considering an end-to-end workflow; therefore, Section 4.2 gives a process overview and subsequent sections detail each individual step. Sections 4.3 and 4.4 consider the integration of the NARR data. The methods, and brief sensitivity studies validating the selections, for temporal and height interpolation are explored in Sections 4.5 and 4.6 respectively. We deal with NARR point selection in Section 4.7 before the final step in the process, interpolation of radiative transfer parameters in elevation and location, is explained and initially validated in Section 4.8.

4.1 Generating Radiative Transfer Parameters

The necessary effective in band radiative transfer parameters, transmission, upwelled radiance, and downwelled radiance, are not explicit outputs from MODTRAN. There are numerous ways to generate these parameters from the spectral outputs given in the MODTRAN output files, but both accuracy and computational efficiency are important when selecting a method for this process. Appendix E outlines each of the different methods considered and the sensitivity study performed to determine which is optimal; here we explain only the selected method.

The governing equation, summarized from Section 3.2,

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

reduces to

$$L_{obs} = L_T \tau + L_u$$

if $\epsilon = 1$. If L_{obs} is plotted against L_T, then the slope is equal to the transmission and the intercept is equal to the upwelled radiance as shown in Figure 4.1 where T_1 and T_2 are temperatures corresponding to L_T in Planck's equation from Equation 3.2 in Section 3.2.
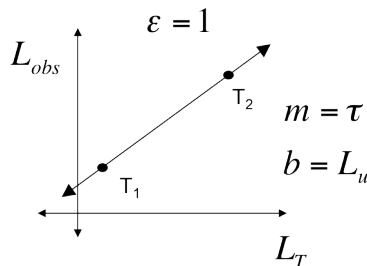
Figure 4.1: Illustration of the linear relationship between L_T and L_{obs} when ϵ=1.

Because transmission, upwelled radiance, and downwelled radiance are characteristics of the atmosphere, surface properties, such as emissivity and self-emitted radiance, can be varied in MODTRAN to determine the radiative transfer parameters. It is important to note that these emissivities and temperatures are simply specified as a tool in calculations and are different than the emissivity and surface temperature of the pixel where these radiative transfer parameters will be used to characterize the atmosphere for LST retrieval. In the long wave infrared (LWIR), characteristics of the atmosphere above a location are independent from the surface properties of that location; therefore we can model and adjust surface properties when characterizing the atmosphere. Temperature (directly related to L_T) and emissivity are MODTRAN inputs and the observed radiance (L_{obs}) can be derived from MODTRAN output (after incorporating the instrument spectral response function), so in two MODTRAN runs with two different surface temperatures and an emissivity equal to one the transmission and upwelled radiance can be determined using linear regression. The two temperatures used corresponding to the self-emitted radiance, 273 K and 310 K, were chosen to span a range that includes most land surface temperatures that will be encountered when generating this product. When the emissivity is not equal to one, the governing equation can be solved for downwelled radiance as shown in Equation 4.1.

$$L_d = \frac{\frac{L_{obs} - L_u}{\tau} - L_T \epsilon}{1 - \epsilon} \tag{4.1}$$

Therefore, a single MODTRAN run with $\epsilon < 1$ can generate the final parameter once the transmission and upwelled radiance are known. In MODTRAN, if the ground temperature is entered as '000', it specifies the air temperature of the initial atmospheric layer as the boundary temperature. Outside unusual circumstances, the air temperature of the lowest layer of the atmosphere and the surface temperature at a given location do not generally span a large range, so this should also be

characteristic of the LSTs the downwelled radiance will be used to determine. Therefore, the third MODTRAN run is executed with T = '000' and $\epsilon = 0.9$ to calculate downwelled radiance.

4.2 Process Overview

As described in Section 2.2, once an appropriate atmospheric characterization data set has been identified, the second objective is to develop an automated process to generate the three radiative transfer parameters at each Landsat pixel. Section 3.2 illustrates that with a calibrated sensor and known surface emissivity, transmission, upwelled radiance, and downwelled radiance (radiative transfer parameters) are required to generate land surface temperature. Section 3.6 describes the identified dataset (NARR) and this chapter describes how that dataset is utilized to generate the radiative transfer parameters. This process is developed with the assumptions that the instrument is calibrated and that both per pixel emissivity and elevation are provided.
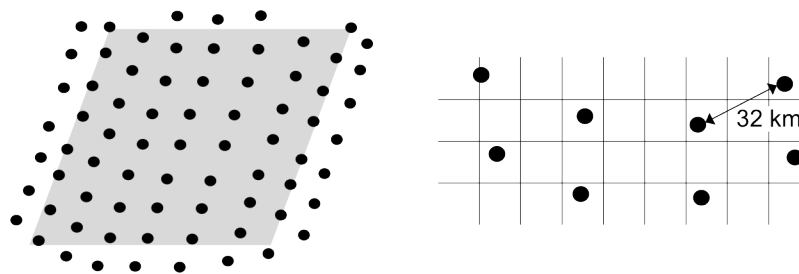
Figure 4.2: NARR points overlaid and subset to a Landsat scene. On the left, the gray box represents a Landsat scene and each black circle a NARR point. On the right, the grid represents the layout of the Landsat coordinate system in comparison to the NARR points. Note that both are for shape and layout and neither schematic is to scale.

Because the NARR dataset covers all of North America, when a particular Landsat scene has been identified, the first step is to spatially subset the NARR data based on the extent of the Landsat scene as shown in Figure 4.2. Points within and around the Landsat scene are selected from the entire NARR dataset. NARR data is on a fixed grid, so once the appropriate points have been identified, the relevant data (geometric height, temperature, and specific humidity at points pertinent to the current Landsat scene) at the samples before and after (NARR data is given on three hour intervals) the acquisition time of the Landsat scene are subset (described in Section 4.3). When the subset of NARR data has been extracted, it must be manipulated so that it can be used as atmospheric profiles in MODTRAN (described in Section 4.4).
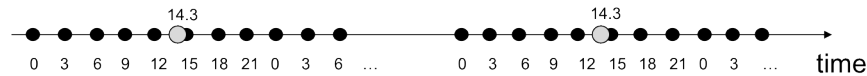
Figure 4.3: Illustration of temporal resolution. The black circles represents NARR points and the gray circles represent Landsat scenes in time. At a single location, each Landsat scene is collected at the same time each day. In this case, data from each NARR point at 12Z and 15Z would be interpolated to the Landsat collection time of 14.3Z.

The first data interpolation occurs in the temporal domain. The NARR data is given every three hours; geometric height, temperature, and relative humidity profiles from NARR samples before and after the Landsat acquisition time are linearly interpolated to the Landsat acquisition time (described in Section 4.5). This is illustrated in Figure 4.3. There are now atmospheric profiles for each necessary variable corresponding to the Landsat acquisition time at each NARR point location pertinent to the Landsat scene.

The necessary radiative transfer parameters are required at every pixel in order to generate a unique LST, which presents the issue of spatial resolution and the varying and unique elevation of each pixel. The ground altitude specified in MODTRAN, and the corresponding adjustment to the atmosphere, changes the resulting radiative transfer parameters, but it is unreasonable to run MODTRAN for every pixel in a scene.
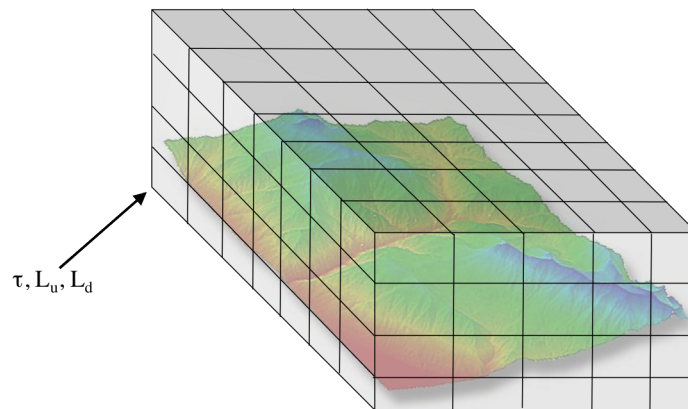
Figure 4.4: The data cube that is created for the Landsat scene by generating radiative transfer parameters at a set of elevations for each NARR point subset for the scene. Note that the cube is larger than the scene due to the selection of NARR points beyond the scene extent for interpolation. (Image from http://www.scisoft-gms.com).

The proposed solution generates the necessary radiative transfer parameters at set elevations at each NARR point location. The execution of MODTRAN at various elevations at the same location requires a linear interpolation of the atmospheric layers (described in Section 4.6). Generating the radiative transfer parameters at a set of elevations at each NARR point results in a three-dimensional (spatial and height) cube of data encompassing the entire Landsat scene as shown in Figure 4.4; radiative transfer parameters for each pixel will be interpolated from this cube. Once

this cube of data has been generated, the NARR points from which the radiative transfer parameters will be interpolated must be identified for every pixel in the scene (described in Section 4.7). The radiative transfer parameters are linearly interpolated to the appropriate elevation at each of these NARR locations, illustrated in Figure 4.5, and these resulting parameters are interpolated to the pixel location using Shepard's method, illustrated in Figure 4.6 (described in Section 4.8).
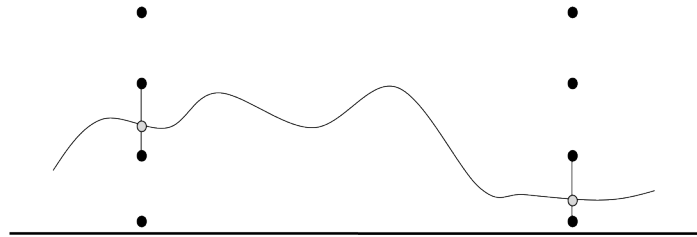
Figure 4.5: Illustration of interpolation in elevation. The black circles represent elevations at which the radiative transfer parameters were generated. For any NARR point, the radiative transfer parameters can be interpolated to the elevation of any pixel of interest, represented for two different pixels by the gray circles.
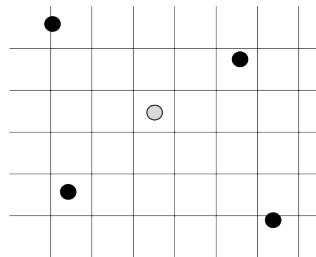
Figure 4.6: Illustration of spatial interpolation. The grid represents the layout of the Landsat pixels and the black circles the NARR points (not to scale). The radiative transfer parameters values at the four pertinent NARR points are interpolated to the location of the current pixel, represented by the gray circle.

The following sections describe the individual steps in the process, including how interpolators were chosen and the sensitivity studies performed to examine the error they can contribute to the final product. Note that the sensitivity studies in this chapter serve only as reasonableness tests for each process; it is difficult to isolate the error contributed by each interpolation and it is also difficult to find actual truth data for comparison. In many cases, the truth data presented in the sensitivity study is another best guess that does not account for differences in one or more dimensions, such as changes in time or space, correcting radiosondes to the appropriate location using surface weather data for example. We are using methods proven in other studies to give reliable results. Therefore, we use these tests to broadly evaluate the interpolations as reasonable to implement in the methodology and move forward. A more rigorous validation of the whole methodology and evaluation of results is presented in Chapter 5 by comparison to ground truth.

4.3   NARR Registration with Landsat

The NARR data is natively in a 349 by 277 point Lambert Conformal Conic grid. This grid is approximately evenly spaced at 32 km with each grid point specified by an (i,j) coordinate. The U.S. Climate Prediction Center provides a grid that gives the latitude and longitude of each (i,j) Lambert Conformal coordinate generated by bilinear interpolation [NOMADS, 2012].

Landsat data is natively given in Universal Transverse Mercator (UTM). In this coordinate system, the Earth is divided in 60 zones, each approximately 6° wide. A location is specified by a zone, an easting value, and a northing value. To avoid negative coordinates, each zone has a reference parallel and reference meridian. The equator is the reference parallel for all zones and assigned a false northing value of 10,000,000 m. Points in the southern hemisphere have northing values less than 10,000,000 m but greater than 0 m; points in the northern hemisphere have northing values greater than 10,000,000 m. Similarly, each zone has a false meridian (central meridian) that is assigned a false easting of 500,000 m. Points west of the central meridian have an easting value less than 500,000 m but greater than 0 m; points east of the central meridian have easting values greater than 500,000 m [USGS, 2007]. In the metadata of each Landsat scene, the zone of the upper left corner is specified; all UTM coordinates for a single scene are given in reference to the false parallel and central meridian of that zone, even if the scene spans more than one zone. The UTM coordinates and latitude and longitude for each corner of the image are all provided in the Landsat metadata. The UTM coordinates of any pixel can be determined using these corner values, the pixel indices within the scene, and the size of each pixel, also given in the metadata.

As described in Section 4.2, the NARR data must be spatially subset based on the extent of the Landsat scene, illustrated in Figure 4.2. The first step is to convert to a common coordinate system. The latitude and longitude coordinates are known for both the NARR points and the corners of the Landsat scene. Because the NARR data and Landsat scene are given in different native coordinate systems, when the NARR data is overlaid on the Landsat scene, the data no longer appears as a regularly spaced, linear grid. Using the corners of the Landsat scene, the NARR points that fall within the scene can be determined. Considering the interpolations that will need to be made later in the process, some NARR points beyond the edges of the Landsat scene need to be included. Therefore, the range determined by the corners of the Landsat scene is increased by approximately the spacing of the NARR points in degrees, in order to include at least one additional NARR point in each direction. The maximum and minimum i and j values are determined and the NARR points are subset based on these ranges of Lambert Conformal coordinates ($\min_i$:$\max_i$, $\min_j$:$\max_j$). This results in, when considered in their native grid, a rectangle that includes all NARR locations necessary to process the current Landsat scene. Figure 4.2 can be considered to be in the native grid of the Landsat scene where the Landsat pixels are regularly spaced and the NARR points are irregularly spaced. Figure 4.7 shows the same set of data, NARR points subset to the Landsat scene, in the NARR native Lambert Conformal coordinates. Here the Landsat scene has an irregular shape and the NARR points are in a regular grid. Note that the NARR points extend beyond the Landsat scene in all directions.
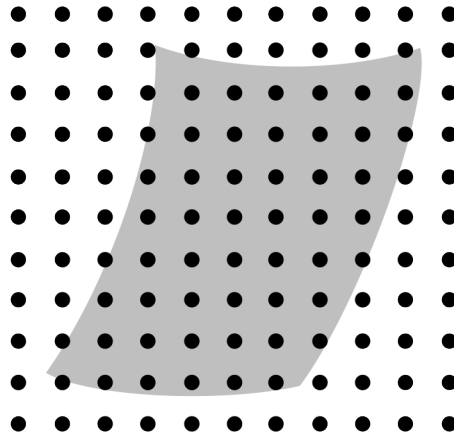
Figure 4.7: Illustration of NARR points overlaid on Landsat scene in NARR native Lambert Conformal coordinates.

4.4 NARR Data to MODTRAN

Various considerations need to be made before this NARR data can be used in MODTRAN. NARR includes air temperature, specific humidity, and geopotential height profiles at fixed pressure levels at each NARR location. As detailed in Section 3.7, MODTRAN requires corresponding points in pressure, air temperature, geometric height and one of the following humidity variables: volume mixing ratio [ppmv], number density [molecules/cm^3], mass mixing ratio [g/kg], mass density [g/m^3], partial pressure [mb], dew point temperature [K or $^\circ$C], or relative humidity [%].

By the methods detailed in Section 3.7, the geopotential height profiles are converted to geometric height profiles using the latitude of the NARR point and the specific humidity profiles are converted to relative humidity profiles using the temperature and pressure profiles of the NARR point.

Because the NARR data is provided at fixed pressure levels, the highest pressure level, corresponding to the lowest elevation, may correspond to a negative geometric height after the conversion. Any levels with negative geometric heights are cut from the profile in MODTRAN and the first pressure level with a positive geometric height becomes the lowest level of the profile.

It is also necessary for the atmospheric profiles to reach higher into the atmosphere than the height of the lowest NARR pressure level. Contents of the atmosphere at these heights are largely negligible when considering the generation of the radiative transfer parameters, but they are required to be present to execute MODTRAN. MODTRAN supplies various standard atmospheres, corresponding to different seasons and areas of the Earth. Therefore, MODTRAN's mid-latitude summer atmosphere is appended to the top of each NARR profile to run MODTRAN. To create a smooth transition between the two profiles, linear interpolations are made between the highest NARR point and the second closest standard atmosphere point. These points are chosen to avoid sudden variations in the profile. This procedure is performed for all four variables.
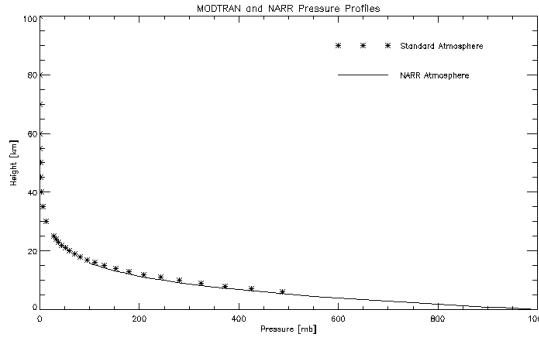
Figure 4.8: Plot of standard atmosphere and NARR pressure profiles.
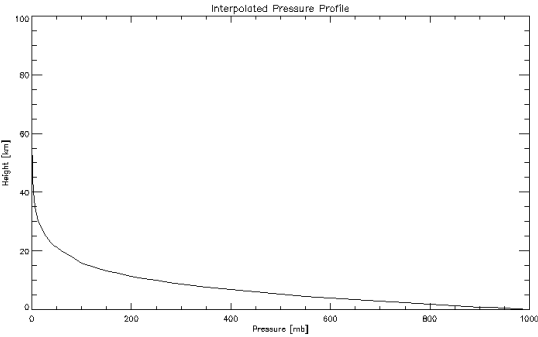
Figure 4.9: Plot of interpolated pressure profile for MODTRAN input.

Figure 4.8 shows the pressure levels for the NARR profiles and the MODTRAN provided standard atmosphere. Figure 4.9 shows where the standard atmosphere has been truncated, interpolated, and appended on the NARR profile. Figures 4.10 and 4.11 show the same for temperature and Figures 4.12 and 4.13 show the same for the relative humidity profiles. Note that the lowest NARR pressure level is 100 hPa, which is in the tens of kilometers for this particular location.
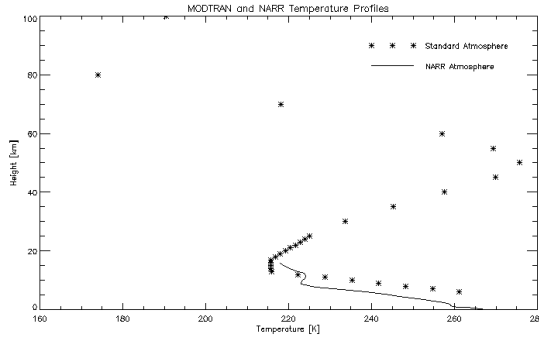
Figure 4.10: Plot of standard atmosphere and NARR temperature profiles.
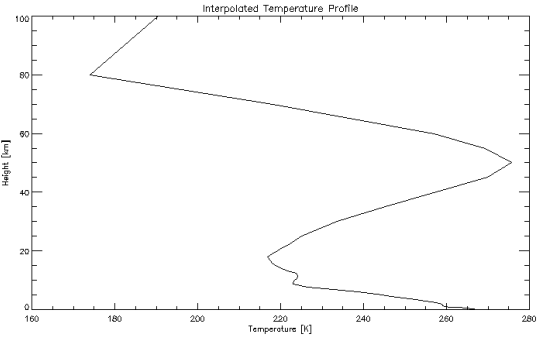
Figure 4.11: Plot of interpolated temperature profile for MODTRAN input.
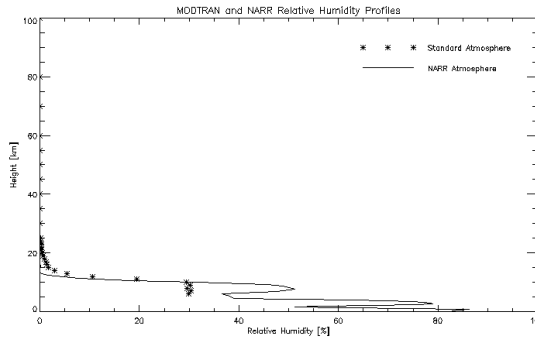
Figure 4.12: Plot of standard atmosphere and NARR relative humidity profiles.
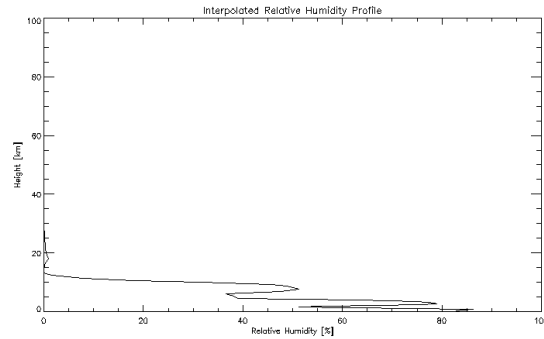
Figure 4.13: Plot of interpolated relative humidity profile for MODTRAN input.

4.5   Temporal Interpolation

All pixels in the Landsat scene will be assumed to be collected at the scene center scan time and all radiative transfer parameters need to be generated for this time.

The NARR profiles provide characterizations of the atmosphere to generate in MODTRAN the radiative transfer parameters necessary for calculating the land surface temperature. The NARR data is a 3-hourly product available at eight evenly spaced samples per day based on Greenwich Mean Time (GMT) or Zulu time (Z).

To determine the optimal temporal interpolation techniques, the structure of the NARR data was investigated. NARR data for a single date was chosen (2 August 2007) and points were selected in the northeast (42.809°N, 78.473°W) and southwest (32.303°N, 115.453°W) regions of the United States. The temperatures for pressure levels 1000 hPa, 875 hPa, 750 hPa, and 550 hPa at all eight samples throughout the day for both coordinates were plotted and are shown in Figures 4.14 and 4.15.
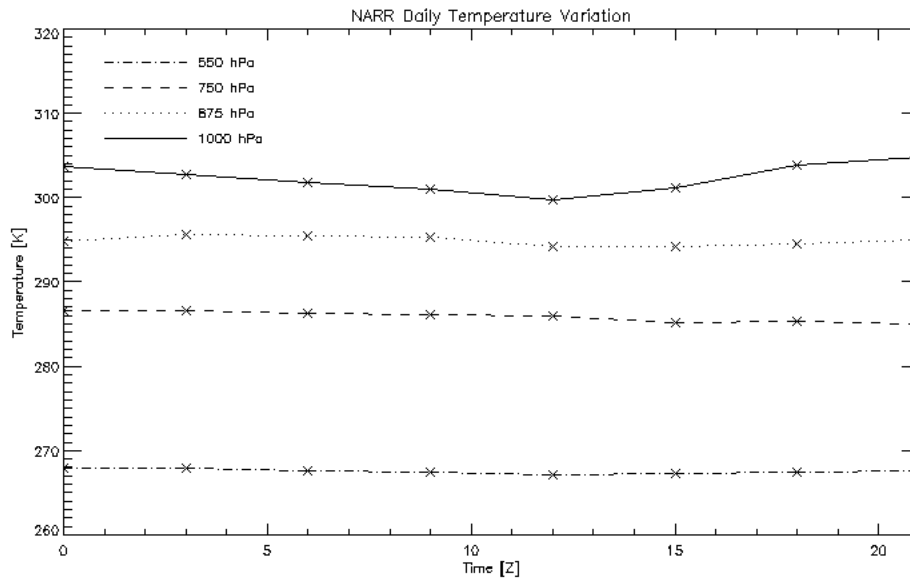
Figure 4.14: Temperature at four pressure levels plotted against time for a NARR point in the northeast region of the United States (42.809°N, 78.473°W).
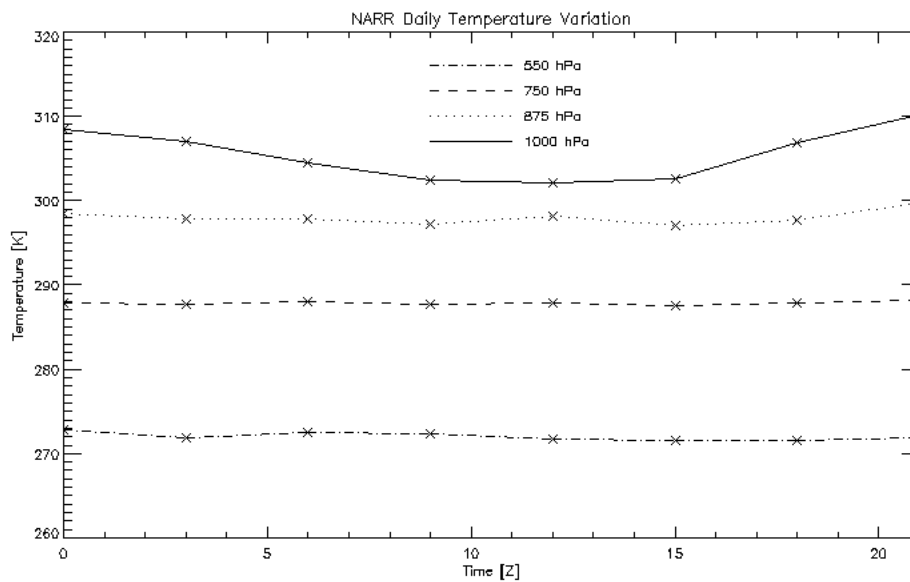
Figure 4.15: Temperature at four pressure levels plotted against time for a NARR point in the southwest region of the United States (32.303°N, 115.453°W).

Figures 4.14 and 4.15 do not show any obvious pattern in temperature as a function of time except for the low amplitude diurnal edge at the lowest level. However, the temperature range on each plot is fairly large, which could be diminishing our ability to observe finer patterns over

smaller ranges. It is important to consider that the layers at the highest pressure levels (lowest heights) have the largest affect on the MODTRAN results. For further investigation, temperatures at the five highest pressure levels (1000 hPa, 975 hPa, 950 hPa, 925 hPa, and 900 hPa) were plotted against time for the same two coordinates on the same date, shown in Figures 4.16 and 4.17.
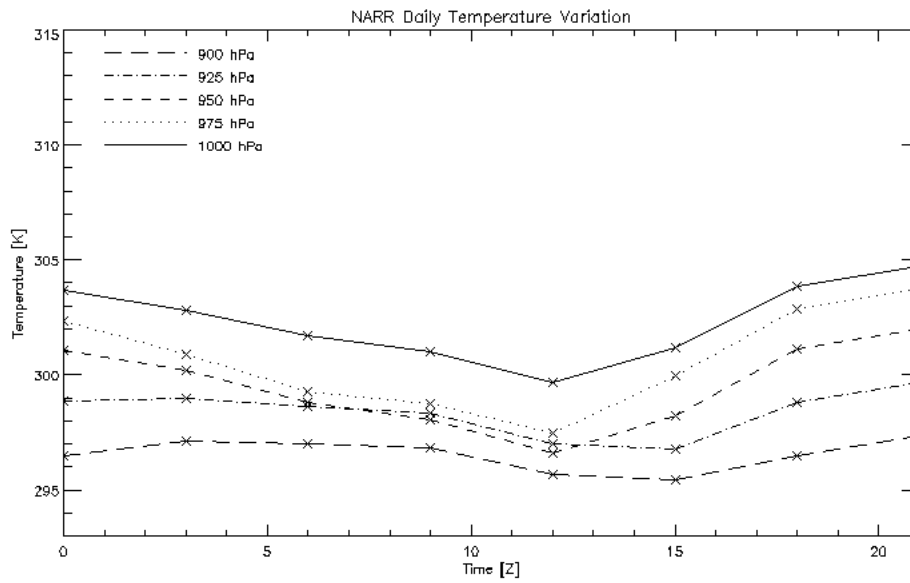
Figure 4.16: Temperature at five highest pressure levels plotted against time for a NARR point in the northeast region of the United States (42.809°N, 78.473°W).
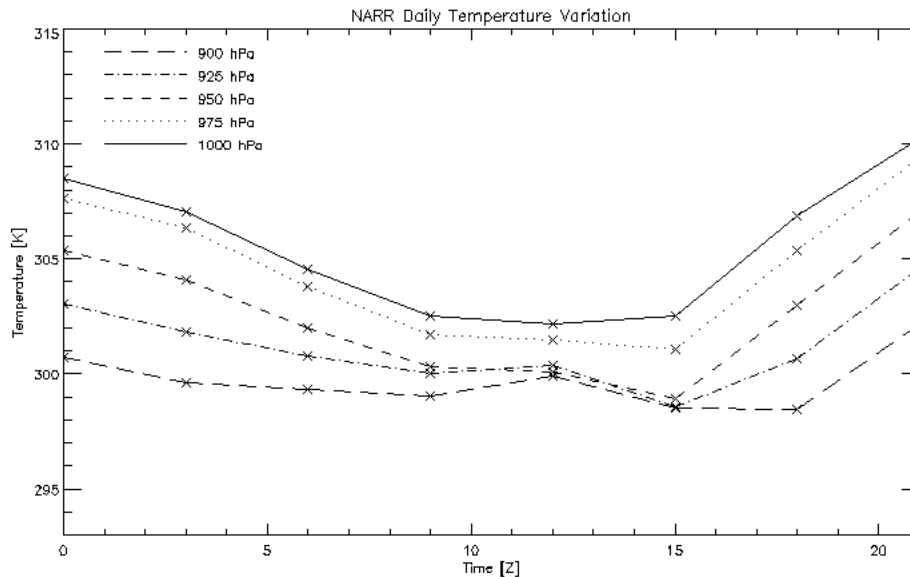
Figure 4.17: Temperature at five highest pressure levels plotted against time for a NARR point in the southwest region of the United States (32.303°N, 115.453°W).

These plots over a smaller temperature range show a greater variation in temperature with time. Sinusoidal, cubic spline, and nearest neighbor interpolators were all considered. However, one of these may best fit a single pressure level or location, but none were good or better fits to all points or pressure levels, and using a larger number of samples in time did not seem to improve the accuracy of the interpolation. A piecewise linear interpolation using one point before and one point after was implemented as an initial method.

The same analysis was conducted to investigate patterns in relative humidity. Figures 4.18 and 4.19 show relative humidity plotted as a function of time for four pressure levels throughout the range of the NARR data (1000 hPa, 875 hPa, 750 hPa, and 550 hPa) for the same northeast and southwest coordinates on the same date.
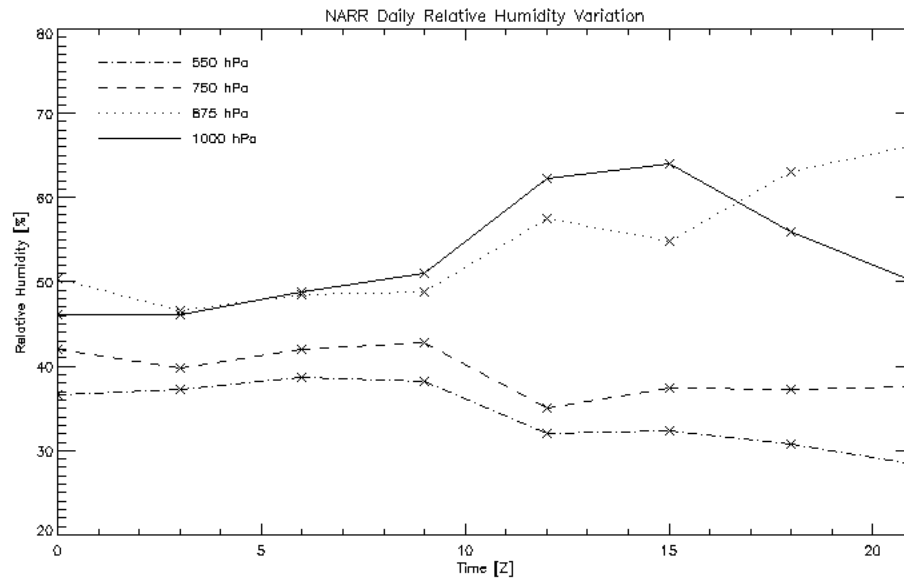
Figure 4.18: Relative humidity at four pressure levels plotted against time for a NARR point in the northeast region of the United States (42.809°N, 78.473°W).
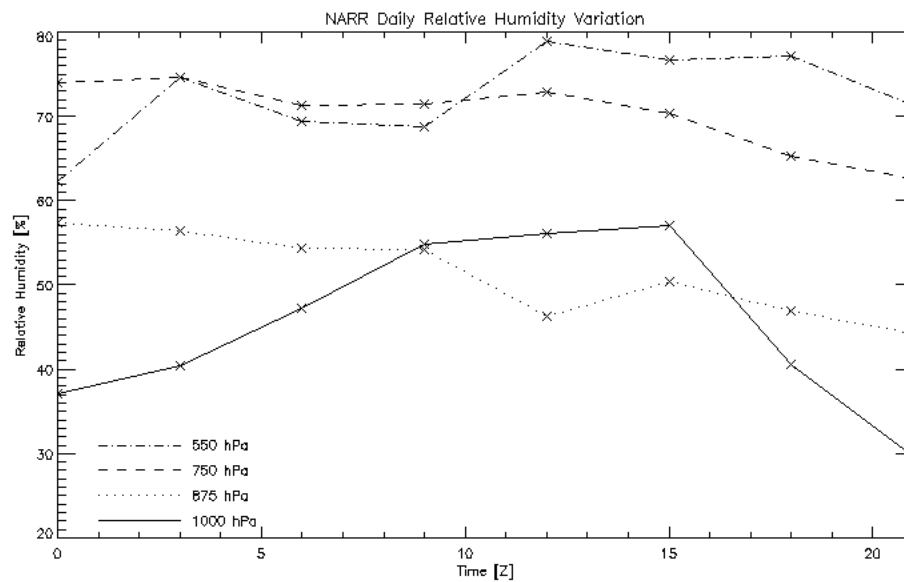
Figure 4.19: Relative humidity at four pressure levels plotted against time for a NARR point in the southwest region of the United States (32.303°N, 115.453°W).

These plots show more variation than was found with temperature but in a less uniform manner. The five highest pressure levels were also investigated, shown in Figures 4.20 and 4.21. Unlike temperature, these are not shown over a smaller range because each individual pressure level has

a large amount of variability. The shape and extent of this variability changes with pressure level
and location.
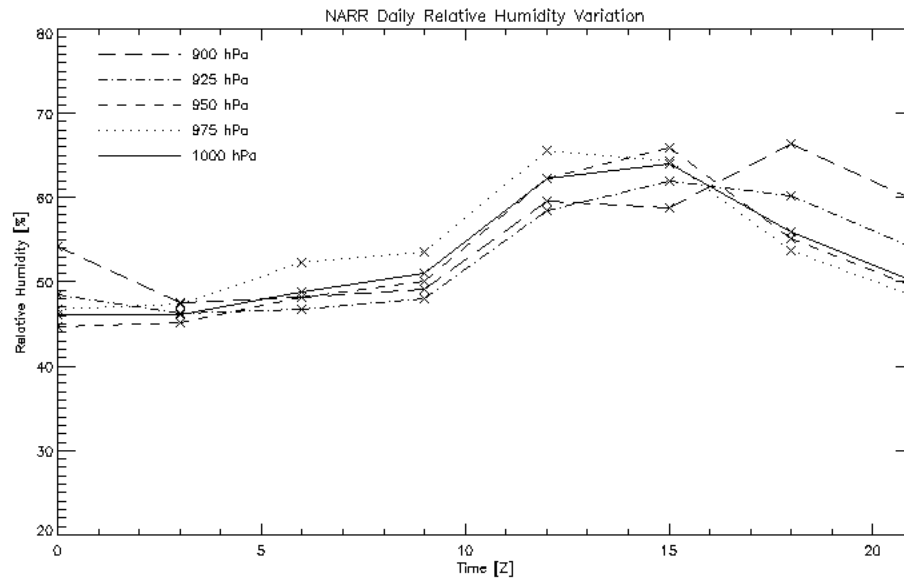
Figure 4.20: Relative humidity at five highest pressure levels plotted against time for a NARR point in
the northeast region of the United States (42.809°N, 78.473°W).
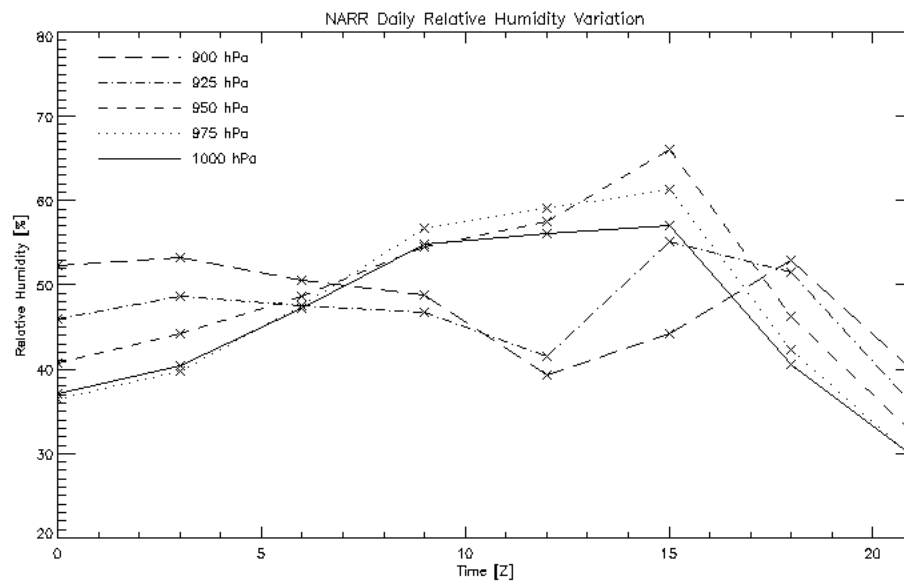
Figure 4.21: Relative humidity at five highest pressure levels plotted against time for a NARR point in
the southwest region of the United States (32.303°N, 115.453°W).

Due to the non-uniformity of the variability in relative humidity, a simple piecewise linear interpolator using one point before and one point after the Landsat acquisition time was selected.

There are also slight variations in geometric height with time at each fixed pressure level, so a piecewise linear interpolation of geometric height was also implemented for consistency. These temporal interpolations are performed at each pressure level in the atmospheric profile space. Because errors in the values of the individual radiative transfer parameters are difficult to interpret, our goal is to isolate the error in the final apparent temperature contributed only from these temporal interpolations.

In order to isolate the error contributed by the temporal interpolation, a ground temperature ($T_{MODTRAN}$) and a truth atmosphere are input into MODTRAN to generate a sensor reaching radiance, transmission, upwelled radiance, and downwelled radiance (L_{obs}, τ_{truth}, L_{u_truth}, and L_{d_truth}). This sensor reaching radiance and these radiative transfer parameters are used to calculate radiance due to temperature as shown in Equation 3.1; this is inverted to an apparent ground temperature using Equation 3.2 through the LUT. This apparent ground temperature from the truth atmosphere (T_{truth}) should be approximately the same as the temperature input into MODTRAN ($T_{MODTRAN}$). Any differences can be contributed to the method of generating radiative transfer parameters explained in Section 4.1. An interpolated NARR profile is then used to generate radiative transfer parameters (τ_{NARR}, L_{u_NARR}, and L_{d_NARR}) via the method described in Section 4.1. The same sensor reaching radiance value (L_{obs}) is used with this set of radiative transfer parameters to calculate the corresponding apparent ground temperature (T_{NARR}) using Equations 3.1 and 3.2. The error contributed by the method of temporal interpolation can then be calculated as the difference between T_{truth} and T_{NARR} as shown in Equation 4.2. Note that the absolute value operation is omitted; negative errors indicate the apparent ground temperature underestimated the actual temperature. Recall, as described in Section 4.2, the following studies use estimated rather than absolute truth. Larger errors than will be acceptable in our final results may be tolerated because this is a simple test of the implementation and reasonability of this interpolation.
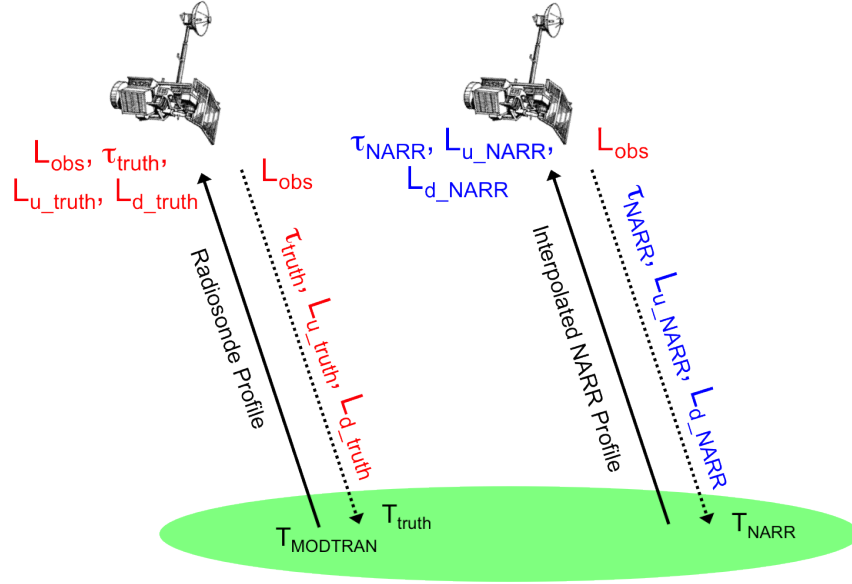
Figure 4.22: Method of determining error contributed by temporal interpolation of NARR profiles.

$$\text{Error} = T_{NARR} - T_{truth} \qquad (4.2)$$

As an initial worse case scenario, the truth atmospheric profile is a NARR profile at time 12 Z, and the NARR profiles for 9 Z and 15 Z are linearly interpolated to estimate this time for a day in August 2007 at a location in the northeast region of the United States (42.809°N, 78.473°W). These profiles, six hours apart and each three hours from the desired time of ground temperature prediction, cover a longer time span than the longest temporal interpolation that could be required in our operational process (two profiles three hours apart and each 1.5 hours from the time of ground temperature prediction). Ground temperatures ($T_{MODTRAN}$) of 273 K, 295 K, and 310 K are used to illustrate the effects over a range of temperatures. Predictions are also made at a range of ground altitudes to examine the effect of elevation.

Results of this investigation, presented graphically in Figure 4.23, give initial confidence in the implementation of the interpolation and showed errors that encourage further testing.
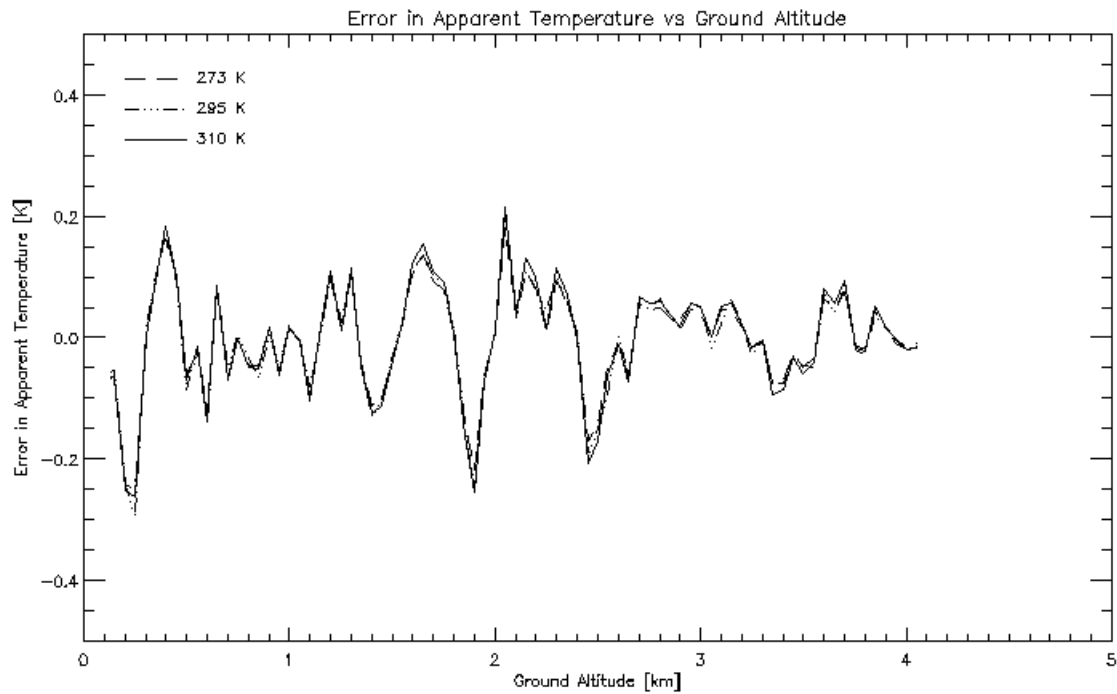
Figure 4.23: Error in apparent temperature contributed by linear interpolation of 9 Z and 15 Z NARR profiles; error computed in comparison to temperature predicted using 12 Z NARR profile.

Plotting the NARR profiles gives an idea of the variability of the atmospheric variables (temperature, relative humidity, and height) with time but not how this variability can affect the predicted ground temperature. To gain a better understanding of how atmospheric variability can alter the predicted LST, the 15 Z and 18 Z NARR profiles were interpolated to 16.5 Z. In respect to Equation 4.2, T_{NARR} was calculated with this interpolated profile but T_{truth} was calculated first with the 15 Z profile, results presented in Figure 4.24, and then the 18 Z profile, results presented in Figure 4.25. By knowingly comparing the retrieved temperature to a temperature from a different time, this error captures how much changes in the atmosphere with time can alter the apparent temperature. If we were to assume our bilinear interpolator is accurate, large errors in this investigation would be an argument against nearest neighbor interpolation.
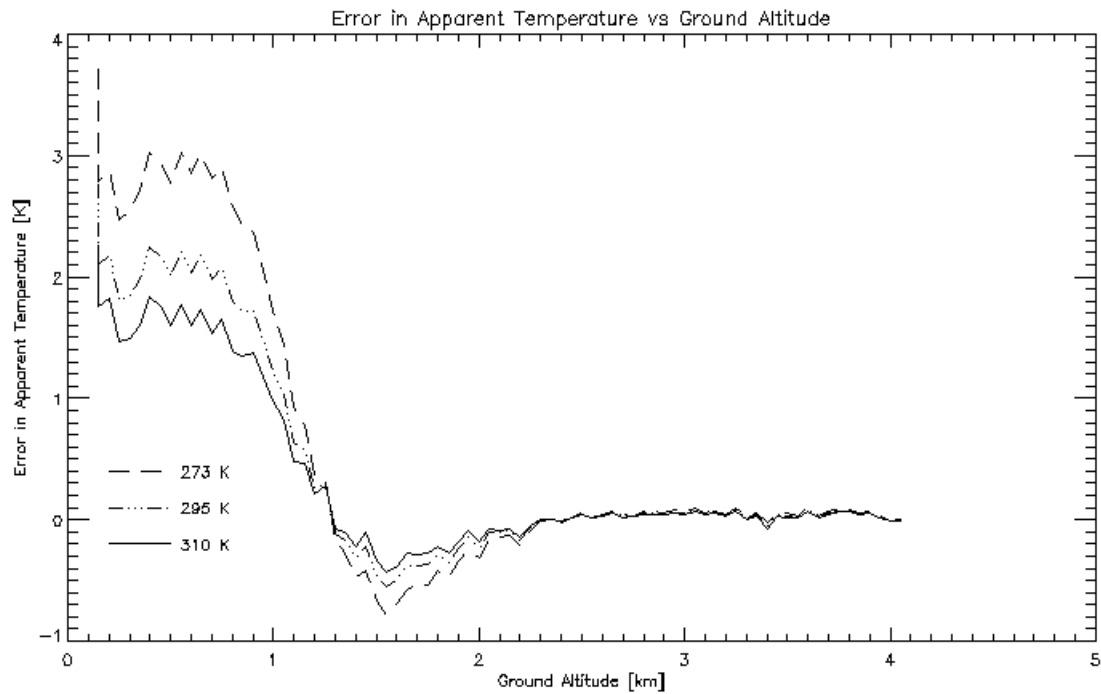
Figure 4.24: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to 15 Z NARR profile.

These results suggest that the atmospheric variability corresponds to 3 K or less of variability in apparent ground temperature; if our assumption of linearity is perfectly correct and we used nearest neighbor interpolation, our predicted error would be 3 K or less. Results in Figure 4.23 suggest our assumption of linearity is not unreasonable; the goal of this study was to assure ourselves that the variability shown in Figures 4.14 through 4.21 does not lead to widely variable predictions of ground truth temperature. The smaller the range of Figures 4.24 and 4.25, the smaller the magnitude of errors that can be introduced with any temporal interpolation.
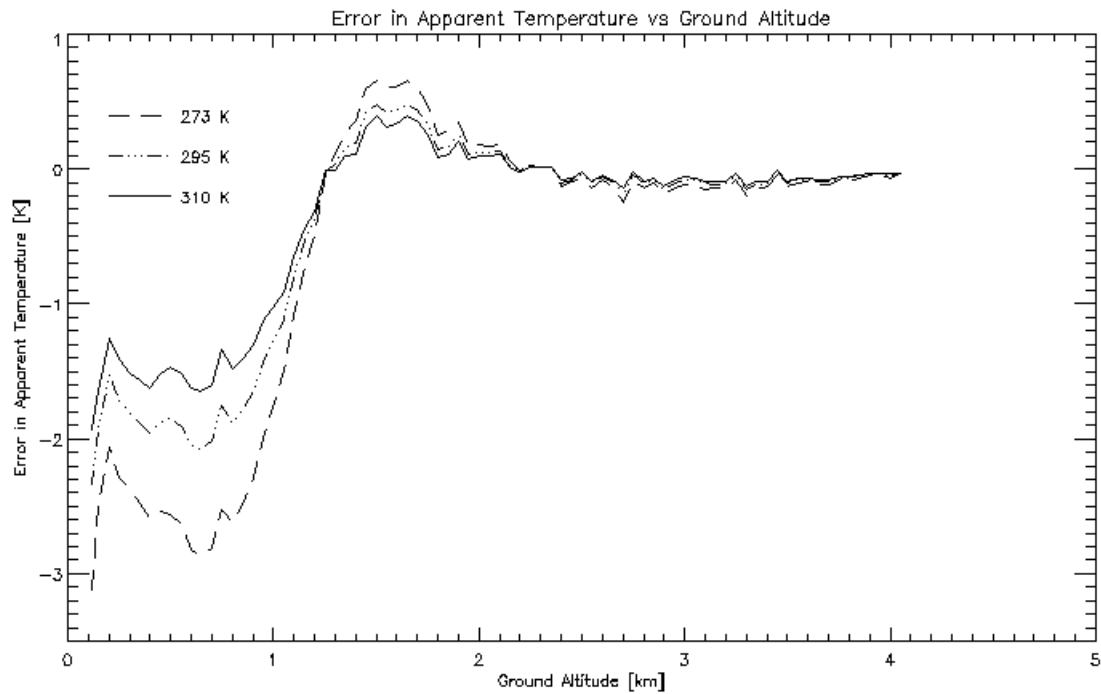
Figure 4.25: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to 18 Z NARR profile.

Finally, to generate a more dynamic set of test cases, radiosonde soundings are used as truth atmospheres. There are generally two radiosonde soundings available each day (00 Z and 12 Z); one sounding is chosen and then interpolated from the boundary layer to surface using surface weather at the time of interest. Refer to Padula's thesis for a complete description of this method [Padula, 2008]. In order to capture the longest time span for the linear interpolation that could be required in the proposed operational LST process, the radiosonde is corrected to surface weather from 16.5 Z and the 15 Z and 18 Z NARR profiles are linearly interpolated to that time. Although this method of radiosonde correction has proven reliable in other studies [Padula, 2008], this is still an estimate of truth rather than an absolute measurement. More importantly, the errors compared to the radiosonde profile do not account for the difference in location in the available radiosonde and surface weather and the nearest NARR point. While the final LST process includes a more involved spatial interpolation, for this simple comparison, nearest neighbor interpolation was used (the closest available radiosonde location and NARR point were compared). This is a very important consideration when analyzing results.

Time of year and season need to be considered in the accuracy of temperature retrieval. It is important to determine if errors are being caused by interpolation techniques or simply by the characteristics of the atmosphere inherent to that time and location. Atmospheric characteristics that can vary with season and location, or are inherent to certain climate or regions at different times of year, make temperature retrieval inherently more difficult. To consider this, an analysis of the

temporal interpolation was performed in each month. The 15 Z and 18 Z profiles were interpolated to 16.5 Z and temperature retrieval results were compared to those for a radiosonde corrected to 16.5 Z surface weather from the same day. The location of the NARR profiles, radiosonde data, and surface weather data were consistent for each month. For initial testing, a point in the northeast was selected based on the accessibility of radiosonde and surface weather data. The results for the error in apparent temperature between ground temperatures predicted using the interpolated NARR profile and the corrected radiosonde profile for each month are shown in Figures 4.26 through 4.37.
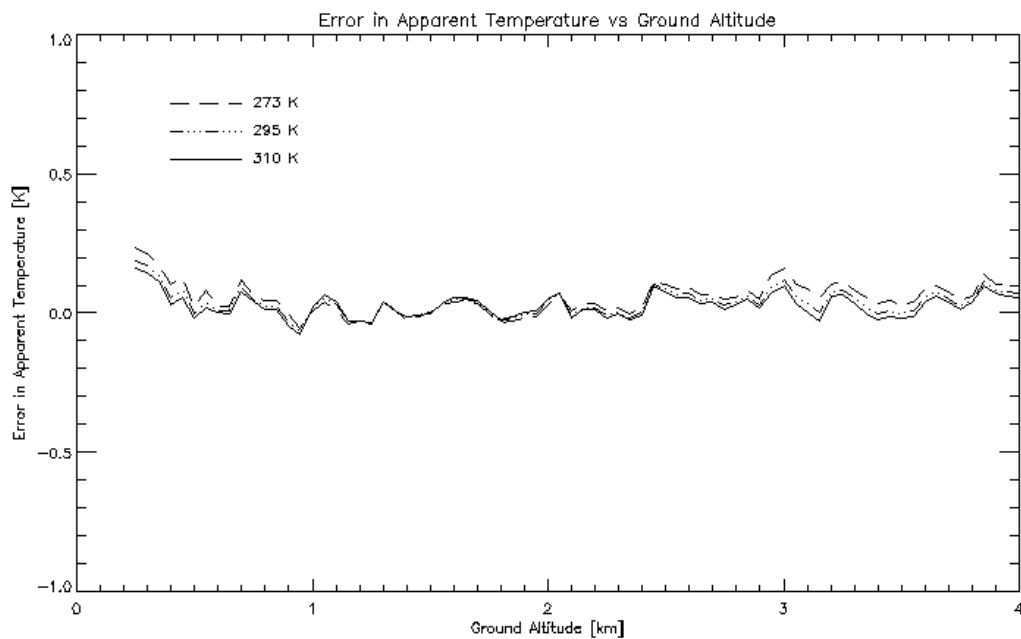
Figure 4.26: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 16 January 2009.
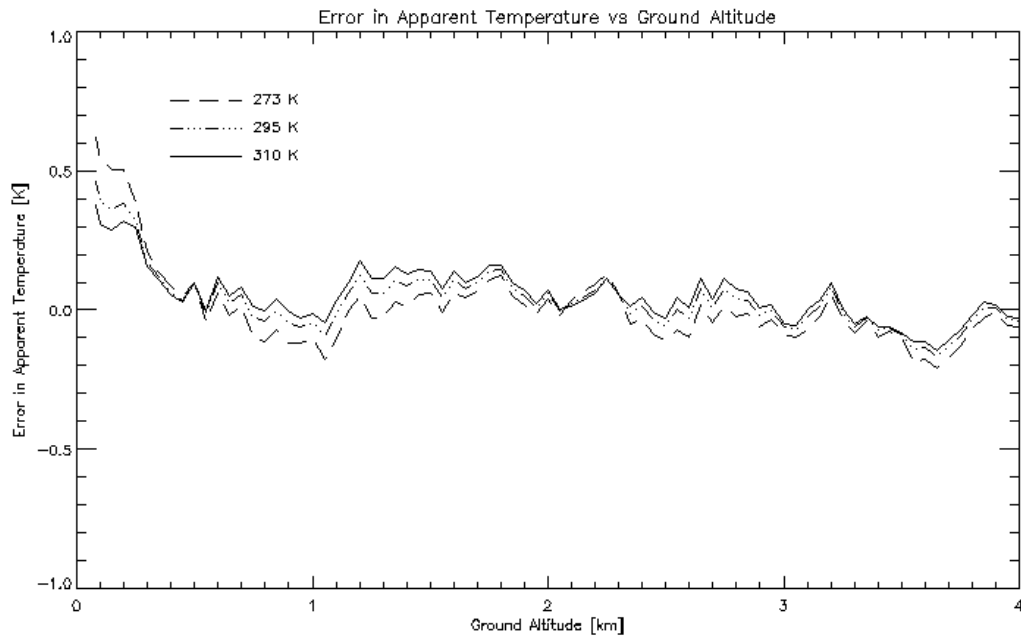
Figure 4.27: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 1 February 2007.
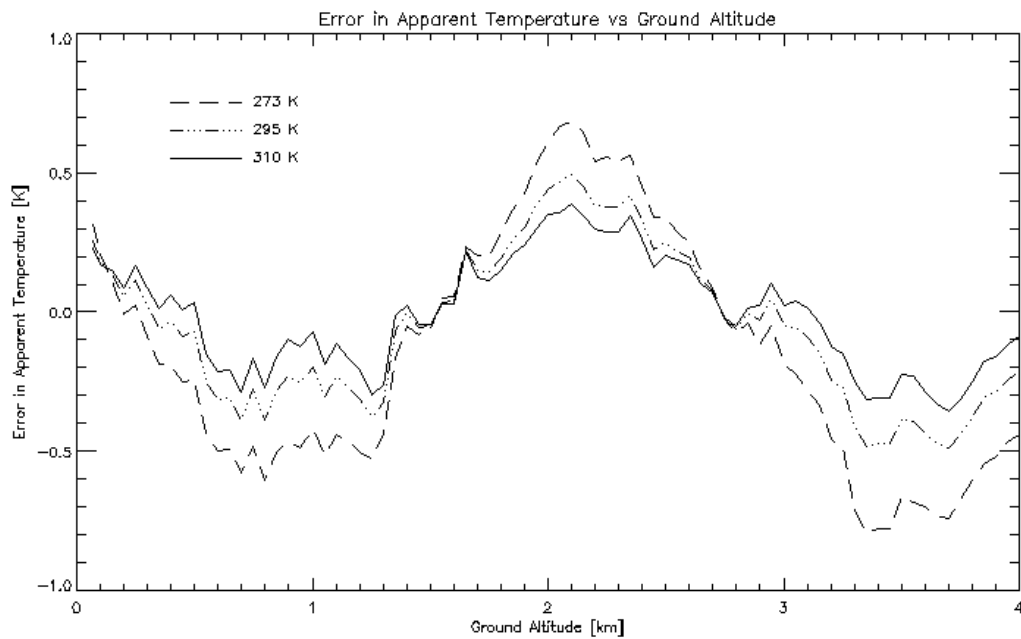
Figure 4.28: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 14 March 2008.
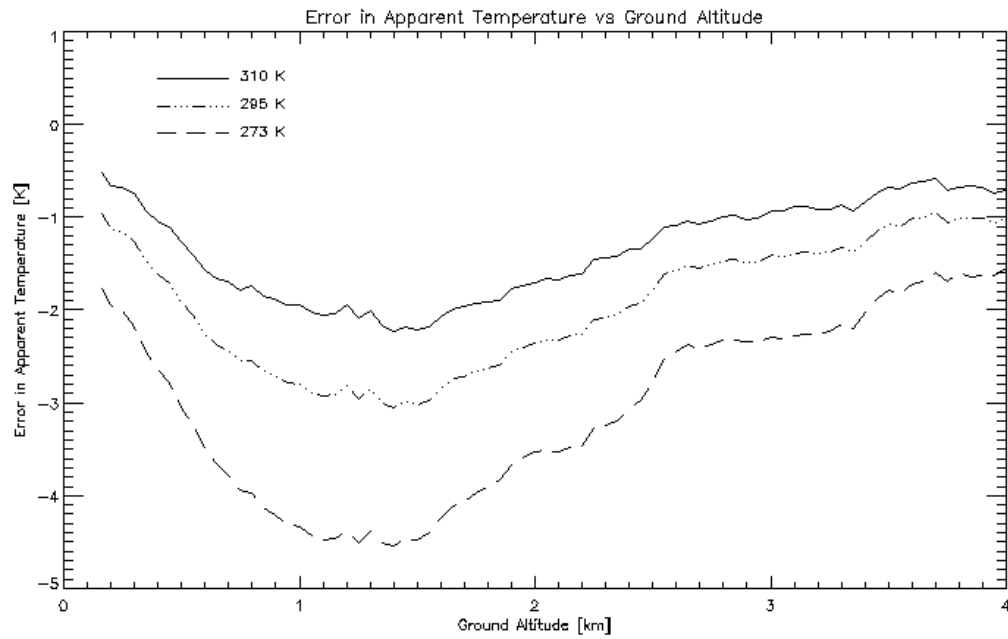
Figure 4.29: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 15 April 2009.
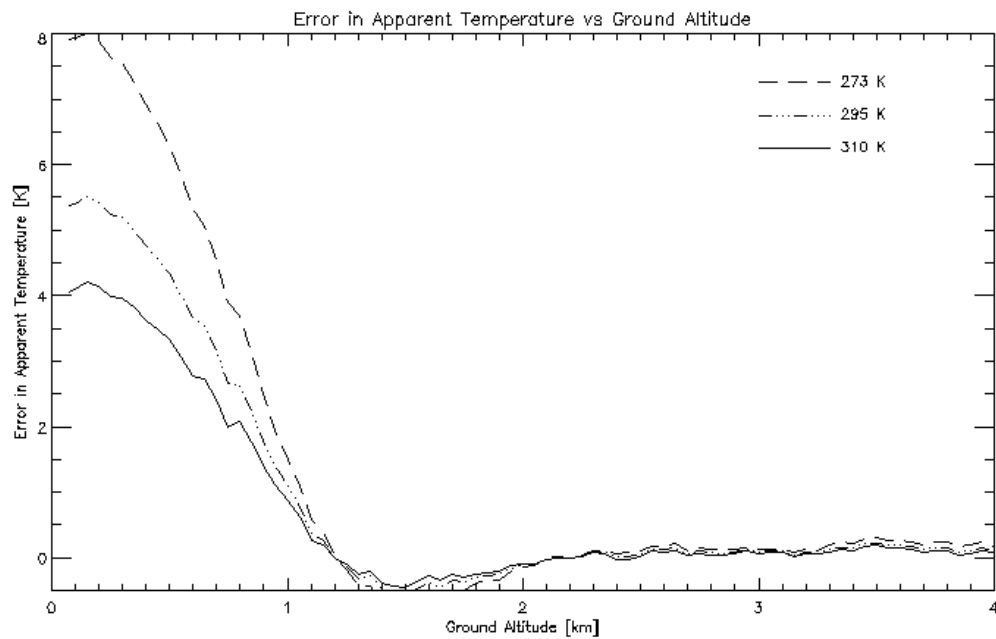
Figure 4.30: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 30 May 2009.
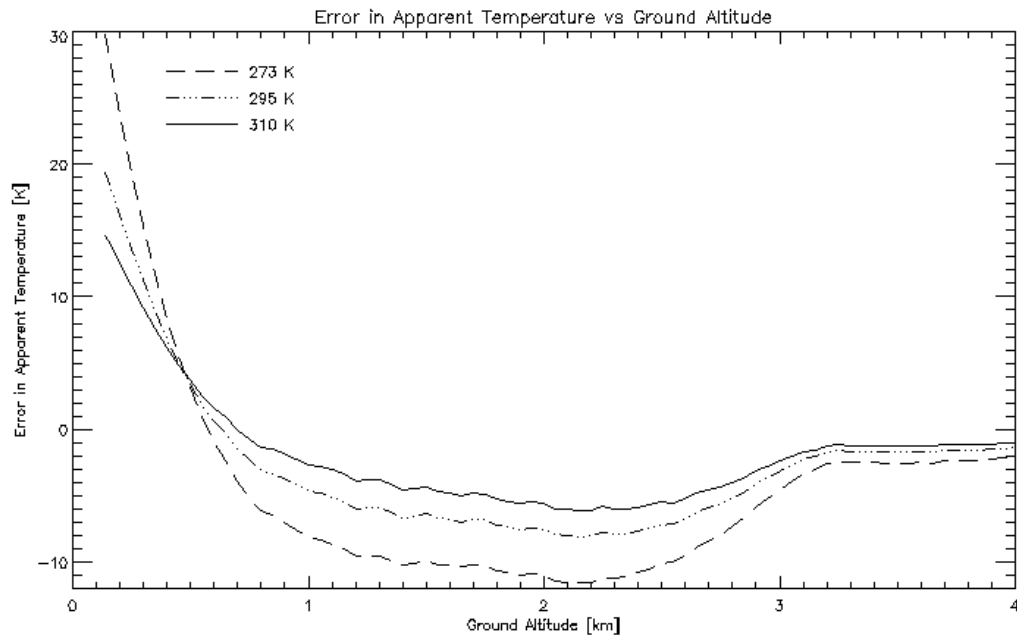
Figure 4.31: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 6 June 2008.
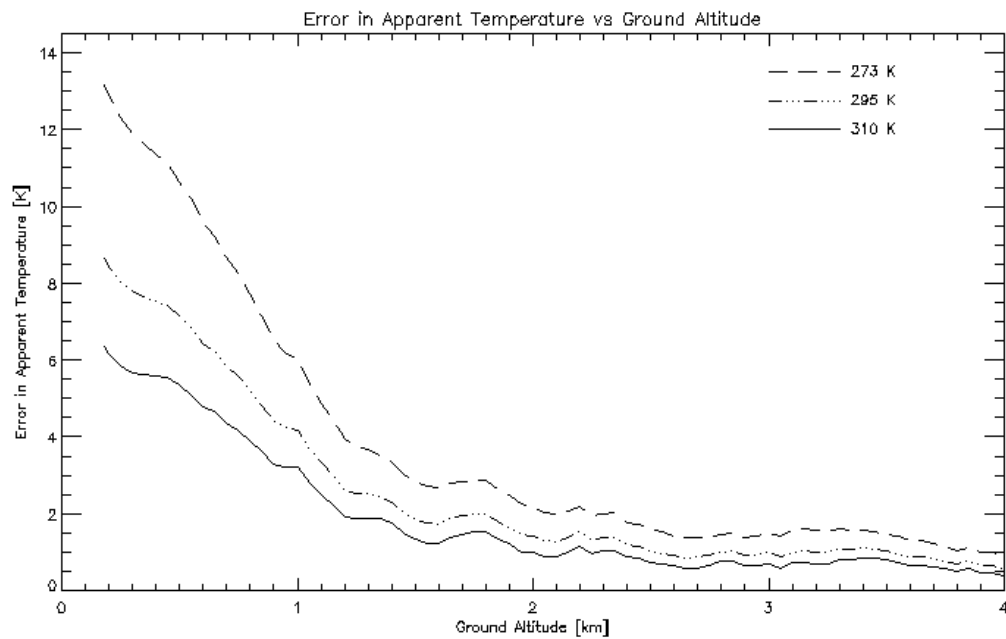
Figure 4.32: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 20 July 2009.
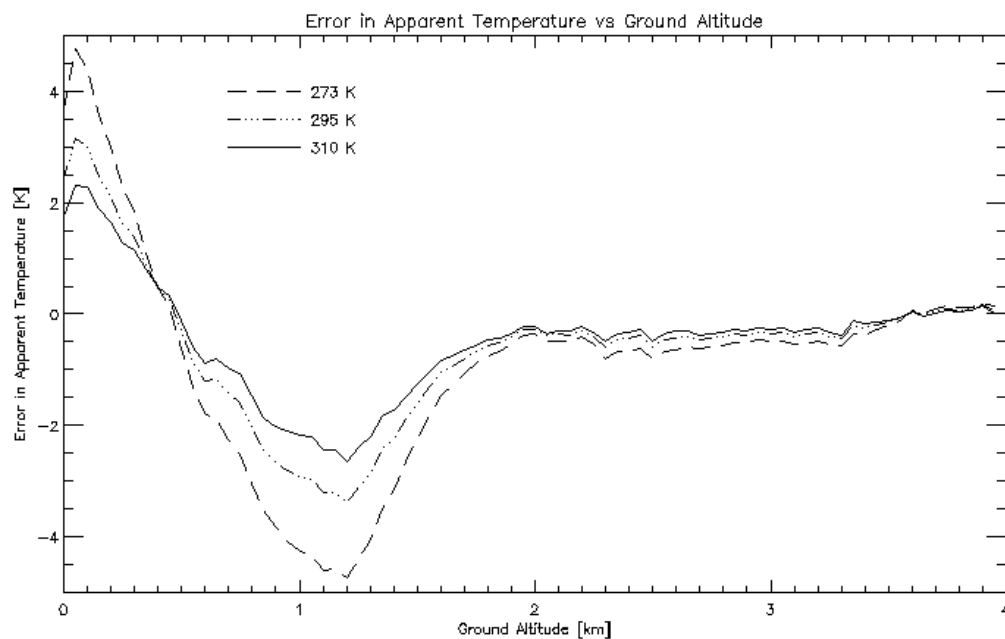
Figure 4.33: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 2 August 2007.
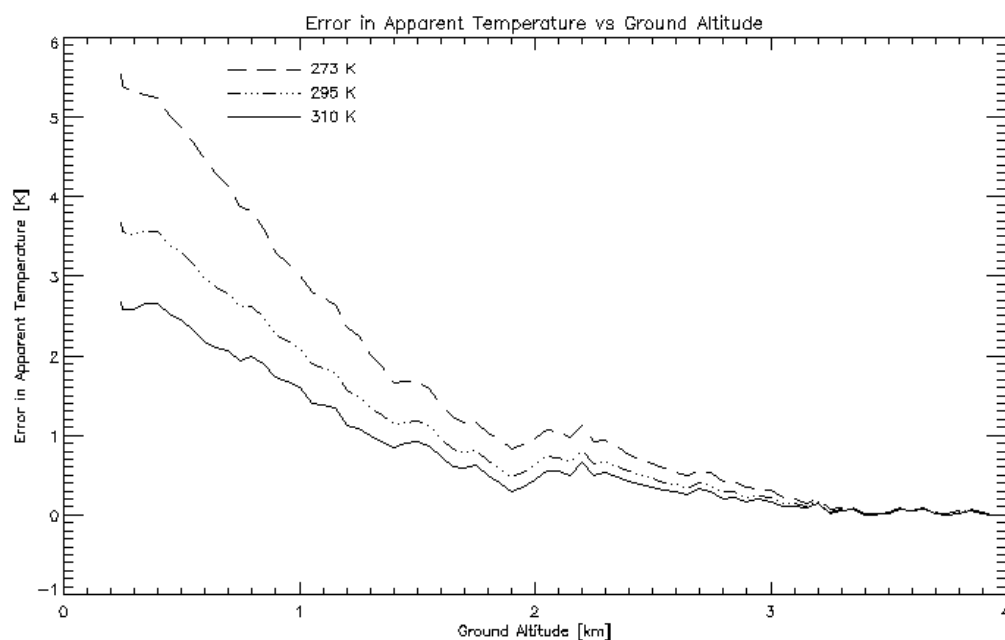
Figure 4.34: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 24 September 2008.
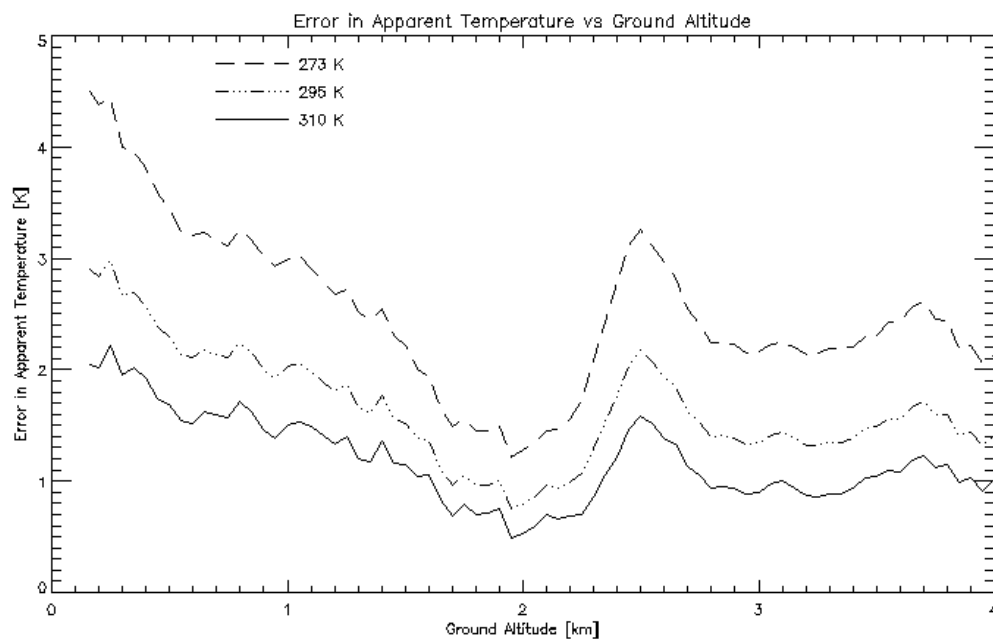
Figure 4.35: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 20 October 2009.
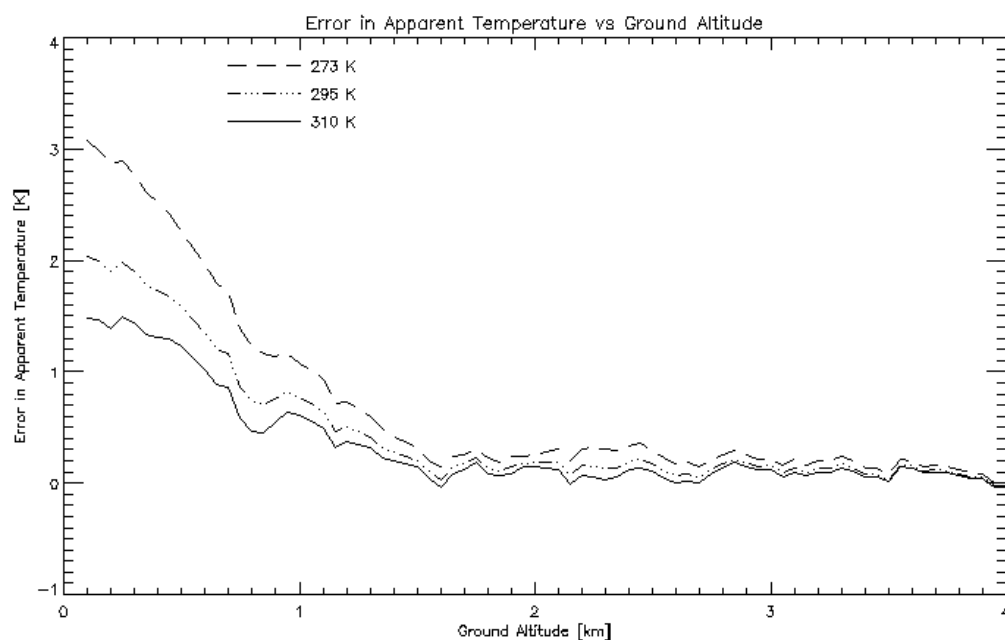
Figure 4.36: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 4 November 2007.
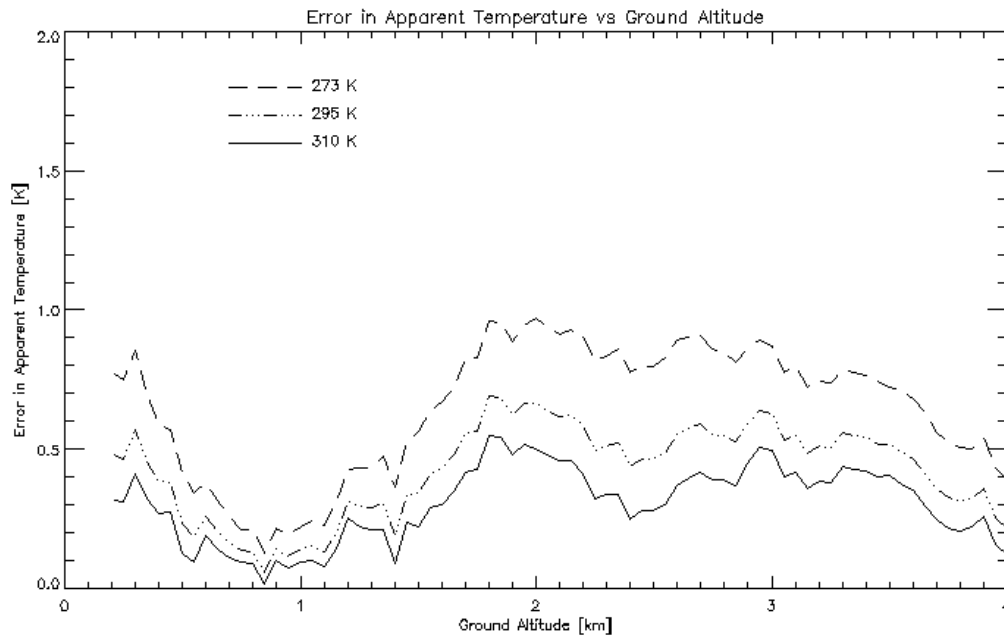
Figure 4.37: Error in apparent temperature contributed by linear interpolation of 15 Z and 18 Z NARR profiles; error computed in comparison to ground temperature predicted with radiosonde profile corrected to surface weather at 16.5 Z for 25 December 2008.

Higher errors at lower altitudes are expected due to the volume of atmosphere being considered. It is important to consider that arbitrary ground temperatures were used; errors may be higher when the input ground temperature is further from any actual LST that might be predicted at that location. Also, atmospheres at different times of year, particularly when it is warmer and more humid, can be more difficult to compensate for. Therefore, for this location near Buffalo, New York, LSTs predicted in the summer months tend to have higher errors than in the winter months. These factors, as well as the differences in location that are not considered, make it difficult to attribute all of the error in Figures 4.26 through 4.37 solely to temporal interpolation.

From Figures 4.26 through 4.37, results from 6 June 2008 and 20 July 2009 are particularly alarming and merit further investigation. As shown in Appendix G, we often expect larger errors with lower transmission and higher relative humidities. Figures 4.38 and 4.39 show the transmission curves from both the corrected radiosonde and the interpolated NARR profiles for 6 June 2008 and 20 July 2009. Compare these curves, particularly at lower altitudes, to Figure 4.40, the radiosonde and NARR transmission curve for 25 December 2008, which resulted in land surface temperature retrieval errors less than 1 K as shown in Figure 4.37. This gives us more confidence that the poor results in Figures 4.31 and 4.32 may be due to challenges presented by the composition or variability of the atmosphere, rather than the interpolation method. We also must realize that we are comparing radiosonde profiles to NARR profiles without appropriately considering the uncertainty associated with either. With these considerations, single digit errors in most results are reasonable enough to initially infer that temporal linear interpolation for the NARR profiles will be sufficient

for LST retrieval.
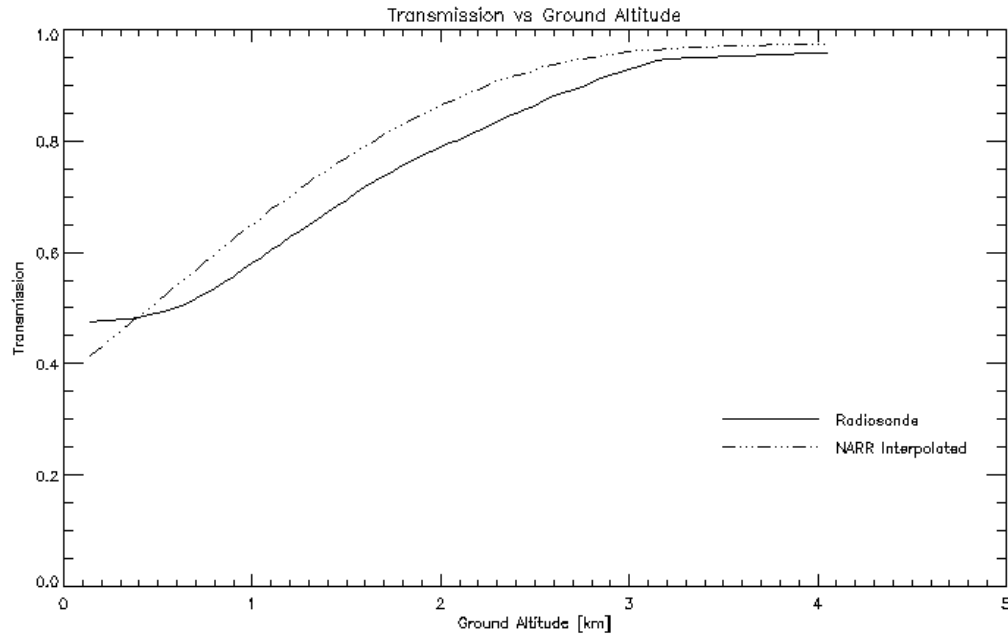
Figure 4.38: Transmission profiles for both the radiosonde and interpolated NARR profile from 6 June 2008. Notice transmission values are low, particularly at the lower altitudes, contributing to large errors in land surface temperature retrieval.
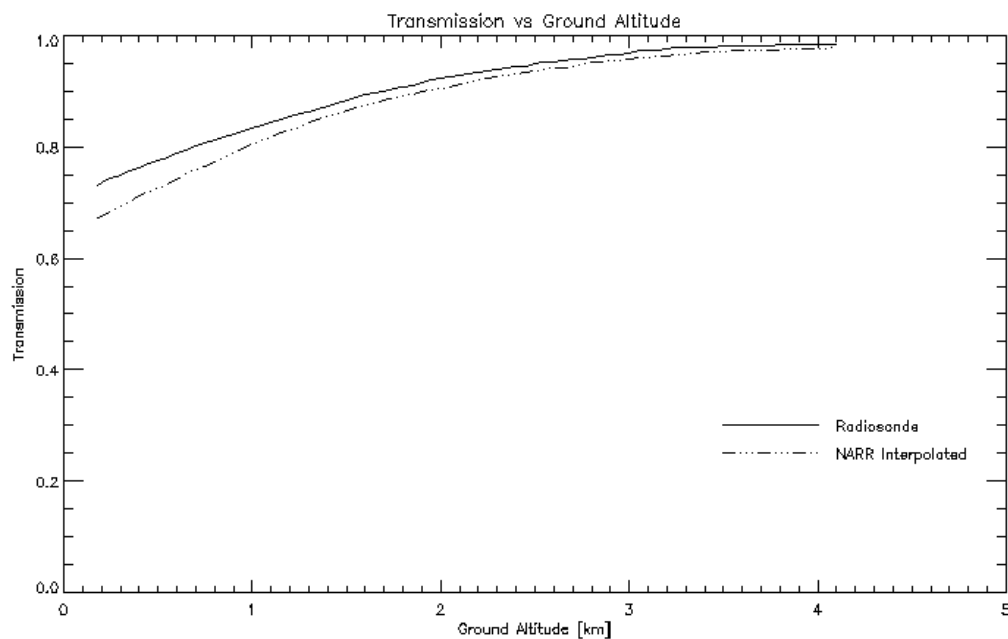
Figure 4.39: Transmission profiles for both the radiosonde and interpolated NARR profile from 20 July 2009. Notice transmission values are low, particularly at the lower altitudes, contributing to large errors in land surface temperature retrieval.
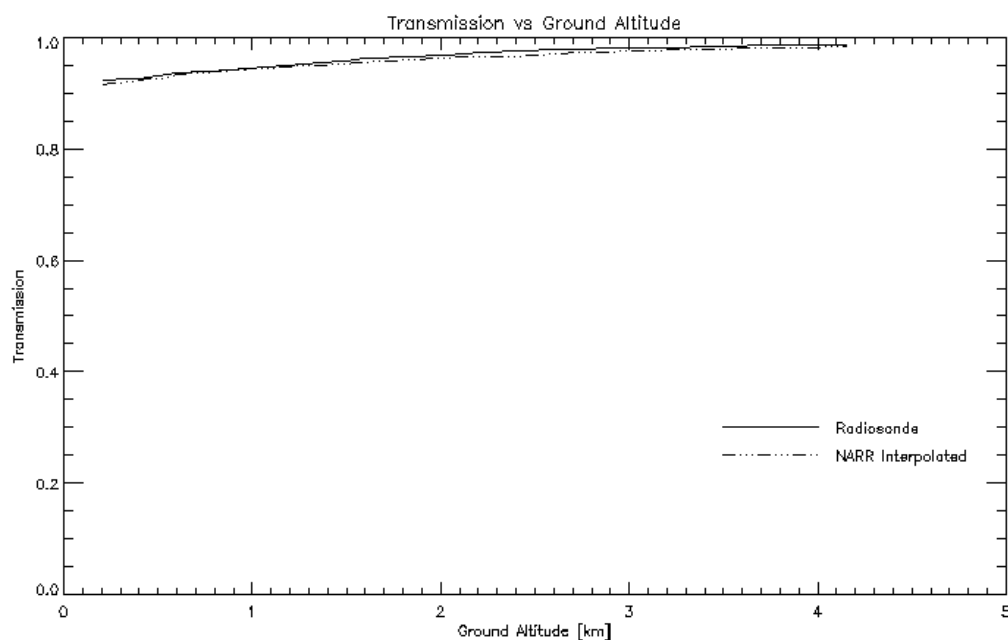
Figure 4.40: Transmission profiles for both the radiosonde and interpolated NARR profile from 25 December 2008. Notice, in comparison to Figures 4.38 and 4.39, transmission values are higher and land surface temperature retrieval results are better.

As shown, it is difficult to isolate errors contributed from individual interpolation techniques and determine how this error will contribute to the final LST product. Rather than attempting to optimize each step as the process is developed, we will make an initial guess at each interpolator with a brief sensitivity study and then implement an automated LST process in its entirety. By later considering errors in the final retrieved temperature with ground truth data and more rigorous testing, we can determine limiting factors and more closely investigate and improve each interpolator if necessary.

4.6   Height Interpolation

Each Landsat pixel has an associated elevation; this elevation affects the NARR parameters necessary for computing the LST. When considering the view from the satellite, generally upwelled and downwelled radiances decrease and transmission increases as the elevation of the ground increases because a smaller volume of atmosphere is being compensated for. However, it is unrealistic to execute MODTRAN for every pixel or every elevation in the image. By executing MODTRAN at a specific set of heights at each NARR location, and generating radiative transfer parameters at each of these heights, these parameters can later be interpolated to the appropriate elevation of each pixel. This requires determining the optimal number and values of the heights at which MODTRAN should be executed.

As an initial test, MODTRAN was executed at nine different heights evenly spaced between 0 km and 4 km. Although most pixels will have an elevation in the lower half of this range, 0 km to 4 km was selected to include most possible elevations around the globe. Nine was chosen as the initial number of heights as a computationally reasonable amount. Both of the selections can be reconsidered and optimized later. Figure 4.41 shows a typical atmospheric profile on the left with a closer look at the bottom of this profile on the right. The horizontal lines represent the elevations at which radiative transfer parameters are currently generated.
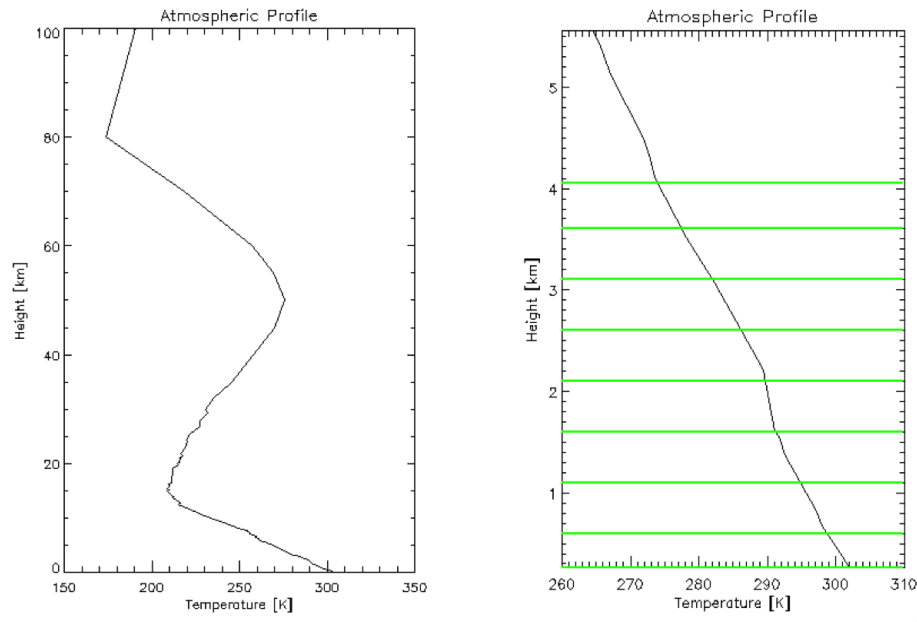
Figure 4.41: An example of an atmospheric profile on the left with a closer look at the lowest elevations on the right. The horizontal lines represent the elevations at which radiative transfer parameters are currently generated.

Figure 4.42: The top of a MODTRAN tape5 file (the MODTRAN input file) illustrating the linear interpolation of atmospheric profile layers.

In order to execute MODTRAN at a specific height, an interpolation within the atmospheric layers in required. While the ground altitude is an input to MODTRAN, this only appropriately affects the results if the atmosphere is modified accordingly. The lowest level of the input atmosphere must have a geometric height equal to the specified ground altitude for MODTRAN to produce the results we desire (MODTRAN must recognize that the atmosphere being compensated for begins at that ground altitude). The elevations of the levels in the atmospheric profiles are determined by the set pressure levels of the NARR data. To run MODTRAN at a particular elevation, the closest layer above and below the ground altitude are linearly interpolated to form an atmospheric layer at the desired ground altitude, after which the atmospheric layers below that ground altitude are

removed. Figure 4.42 shows the top of a typical tape5 file, the input file for a MODTRAN run (see Appendix D). The first four lines contain various inputs for the program while the fifth line (shown crossed out in Figure 4.42) is the first layer of the atmospheric profile. So, for example, if the MODTRAN run specified by the tape5 file in Figure 4.42 was to be executed at a ground altitude of 0.468 km, then 0.468 would be entered for the ground altitude (gdalt keyword). Using 0.358 km, 0.468 km, and 0.585 km as elevations for a linear interpolation, pressure, temperature, and relative humidity (the second, third, and fourth columns respectively) are linearly interpolated and this new atmospheric layer is inserted becoming the first atmospheric layer in the MODTRAN tape5 file after the lower layers (at heights 0.136 km and 0.368 km) are deleted. This basic method was developed and a simple study was implemented to explore the effects of using linear interpolation. While the decision and method of truncating atmospheric layers warrants further investigation and will be investigated later in this section, the choice of linear interpolation to fabricate the lowest atmospheric layer seems intuitive from Figure 4.41 and other investigations of NARR profiles. Linearly interpolating two layers to generate an atmospheric profile layer at an existing elevation (for example, linearly interpolating layers at 0.136 km and 0.585 km to get a layer at 0.358 km in Figure 4.42) allowed for comparisons that showed the effects were reasonable. Also considering the established use of linear interpolation in the recreation of the atmospheric column [Padula, 2008], this process was deemed adequate and implemented. It is difficult to generate a study to determine the total effect on the final apparent temperature because any error will be compounded with error from the following step.

The process described above was used to run MODTRAN at 9 heights, evenly spaced between 0 km and 4 km, as shown in Figure 4.41. The lowest height is always the first layer of the NARR profile and the next 8 are the same for every point (0.6 km, 1.1 km, 1.6 km, 2.1 km, 2.6 km, 3.1 km, 3.6 km, and 4.05 km). The radiative transfer parameters at these nine heights need to be interpolated to the elevation of each pixel, obtained from a digital elevation model (DEM). As an initial method, a simple piecewise linear interpolation using one point above and one point below the desired elevation is implemented. We believe this will provide more accuracy than nearest neighbor interpolation because we expect the radiative transfer parameters to vary monotonically with height but do not anticipate a more complicated predictable trend. To determine the error contributed by this interpolation, atmospheric parameters were generated at these nine heights, and then for the same NARR profile, atmospheric parameters were generated at eighty elevations between 0 km and 4 km. This is only used in validation because it is computationally unreasonable operationally. These finely sampled "truth" parameters were used, with observed radiances from input temperatures of 273 K, 295 K, and 310 K, to generate "truth" temperatures as described in Section 4.1. As with temporal interpolation, this is an estimation of truth rather than absolute measurements. Using only the elevation closest above and closest below, parameters from the sampled nine elevations were linearly interpolated to generate parameters at each of the eighty elevations. These linearly interpolated parameters are used to estimate an apparent land surface temperature using the same observed radiance, which can be compared to that generated from the

"truth" parameters. Figures 4.43 and 4.44 show the errors in apparent temperature for two dates in February and August respectively. The location of the NARR profile was the same as that used in the temporal interpolation study, near Buffalo, NY, at 42.809°N, 78.473°W.
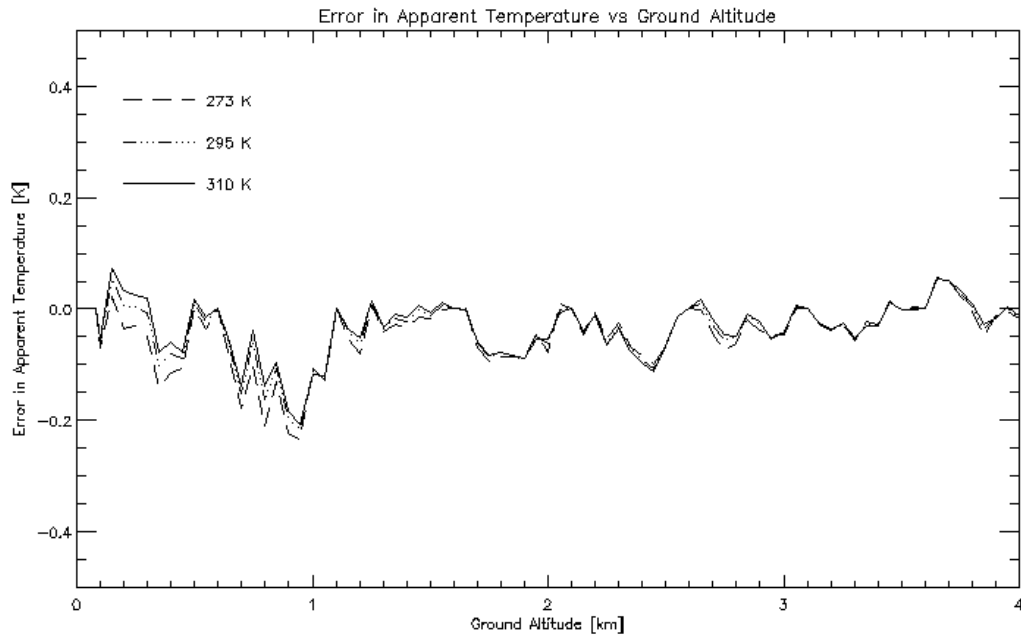
Figure 4.43: Errors in apparent temperature contributed by interpolation in elevation for a NARR location in the northeastern United States in February.
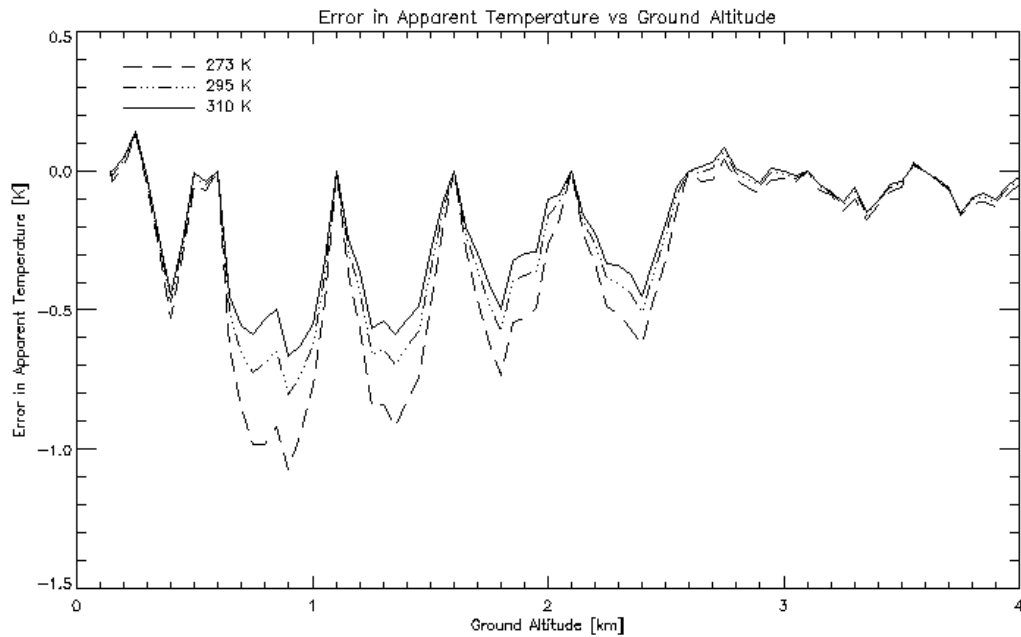
Figure 4.44: Errors in apparent temperature contributed by interpolation in elevation for a NARR location in the northeastern United States in August.

As expected, there are higher errors at lower altitudes from both dates. But even on a summer day where we expect that the warmer, more humid atmosphere may cause more difficult retrieval, the largest errors were not greater than 1 K. Because this is only an initial test and the goal is to determine the error in the final product, compounded by all interpolations and contributed from all datasets, this simple study was enough to implement this interpolation in the initial process.

Because the largest errors are at the lowest elevations and most pixels will have elevations of less than 2 km, we thought that it would be beneficial to investigate irregularly spacing the generation of the radiative transfer parameters with more samples between 0 km and 2 km. However, we found it difficult to determine the optimal spacing of samples because the distribution of water vapor will vary for different atmospheres. We considered the elevations corresponding to equal steps in water vapor for a number of atmospheres as shown in Figures 4.45, 4.46, and 4.47 and Tables 4.1, 4.2, and 4.3 respectively.
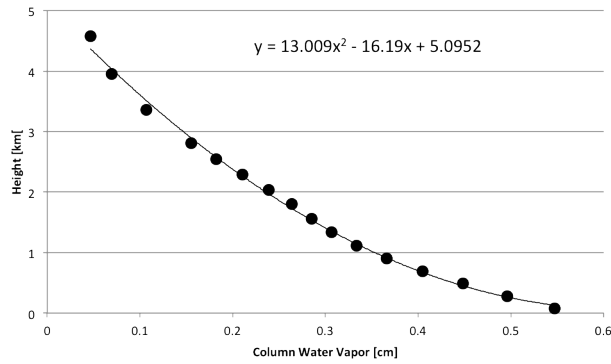
Figure 4.45: Distribution of heights for equal steps in water vapor and corresponding regression for a February atmosphere near Buffalo, NY.

CWV [cm]	Height [km]
0.05	4.3182
0.10	3.6063
0.15	2.9594
0.20	2.3776
0.25	1.8608
0.30	1.4090
0.35	1.0223
0.40	0.7006
0.45	0.4440
0.50	0.2525
0.55	0.1259
0.60	0.0644
0.65	0.0680
0.70	0.1366
0.75	0.2703
0.80	0.4690
0.85	0.7327

Table 4.1: Distribution of heights for equal steps in water vapor and corresponding regression for a February atmosphere near Buffalo, NY.
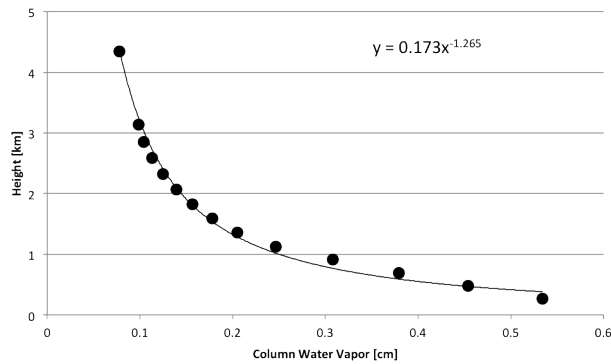
Figure 4.46: Distribution of heights for equal steps in water vapor and corresponding regression for a May atmosphere near Buffalo, NY.

CWV [cm]	Height [km]
0.05	7.6533
0.10	3.1845
0.15	1.9067
0.20	1.3251
0.25	0.9992
0.30	0.7934
0.35	0.6528
0.40	0.5514
0.45	0.4750
0.50	0.4158
0.55	0.3685
0.60	0.3301
0.65	0.2983
0.70	0.2716
0.75	0.2489
0.80	0.2294
0.85	0.2125

Table 4.2: Distribution of heights for equal steps in water vapor and corresponding regression for a May atmosphere in upstate New York.
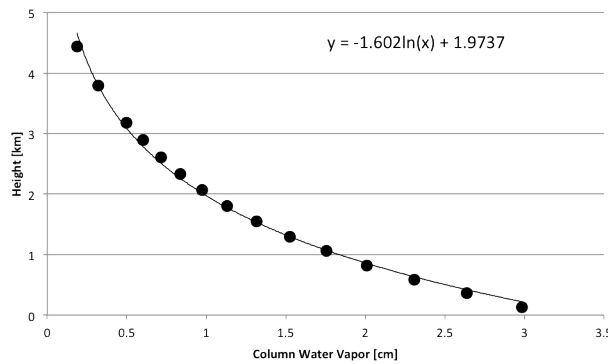
Figure 4.47: Distribution of heights for equal steps in water vapor and corresponding regression for an August atmosphere in upstate New York.

CWV [cm]	Height [km]
0.05	6.7729
0.10	5.6624
0.15	5.0129
0.20	4.5520
0.25	4.1945
0.30	3.9025
0.35	3.6555
0.40	3.4416
0.45	3.2529
0.50	3.0841
0.55	2.9314
0.60	2.7920
0.65	2.6638
0.70	2.5451
0.75	2.4546
0.80	2.3312
0.85	2.2341

Table 4.3: Distribution of heights for equal steps in water vapor and corresponding regression for an August atmosphere in upstate New York.

As shown above, the range and distribution of heights corresponding to equal steps in water vapor, and relationship between height and water vapor, varied greatly even for a single location at different times of the year. It would be difficult to optimize a distribution for a single location, and nearly impossible for a product across North America or the globe. Even the largest errors were less that 1 K for evenly distributed steps in height, so the process was implemented with evenly distributed steps in height. These nine heights were left as the height of the lowest atmospheric layer, 0.6 km, 1.1 km, 1.6 km, 2.1 km, 2.6 km, 3.1 km, 3.6 km, and 4.1 km for every NARR location.

As shown above, the distribution of column water vapor can vary greatly, even for a single location, and can greatly affect the temperature retrieval. We think that water vapor in the atmosphere will have a larger effect on retrievals than height interpolation, regardless of steps in height. However, we do consider that with the methods described above, we are simply truncating the column water vapor as we increase the ground altitude. That is, we remove layers below the ground altitude and are currently also removing any water vapor below that ground altitude from the atmosphere. Because we believe that water vapor in the atmosphere is one of our greatest obstacles in accurate temperature retrieval, we do a brief study on how we deal with this water vapor.

With the methodology described above, when the atmospheric layers are truncated, the column water vapor is removed. This study compares the effects of removing the column water vapor versus redistributing it through the remaining layers so that even though the ground altitude is higher, and the number of atmospheric layers is less, the total column water vapor remains the same. As an initial investigation, two atmospheres from the radiosonde near Buffalo, NY (42.809°N, 78.473°W), were selected. One in February that had a total column water vapor of 0.54718 cm and another from August with a total column water vapor of 2.998617 cm.

Figures 4.48 and 4.49 show the difference in retrieved temperature when the column water vapor is truncated (as the ground altitude is increased, more water vapor is removed) and when the column water vapor is maintained as the ground altitude is increased.
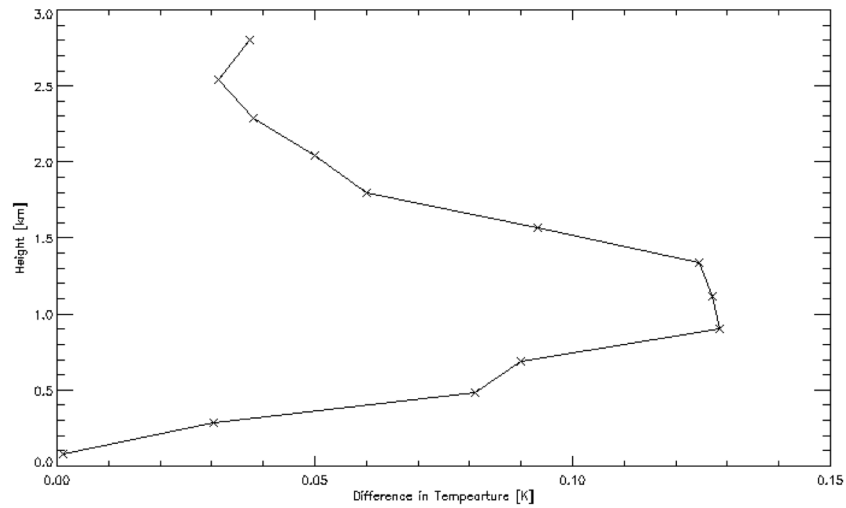
Figure 4.48: Difference in retrieved temperature when column water vapor is truncated and when column water vapor is maintained as ground altitude is increased for a February atmosphere near Buffalo, NY (where the initial total column water vapor is 0.547 cm).
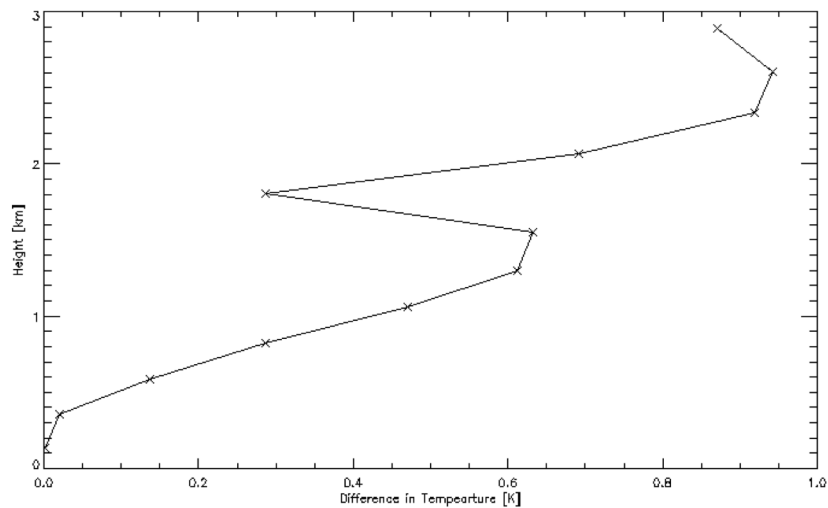
Figure 4.49: Difference in retrieved temperature when column water vapor is truncated and when column water vapor is maintained as ground altitude is increased for an August atmosphere near Buffalo, NY (where the initial total column water vapor is 2.999 cm).

Note that when the column water vapor is already small, the treatment has very small effects on the retrieved temperature as shown in Figure 4.48. When the column water vapor is considerably larger, the effects are greater, but still less than 1 K for all heights, and less than 0.5 K for ground altitudes less than 1 km. Note that the actual elevation of this location is approximately 0.18 km. As will be discussed in Chapter 5, we use water temperatures measured on buoys to validate our

results. Many of these are at sea level, so it is difficult to characterize this effect against truth data because we lack variability in elevation in our ground truth sites. However, as an initial study, we use points in Lake Huron and Lake Superior, described in Sections 5.1.8 and 5.1.9, because these are above sea level (0.177 km and 0.183 km respectively), so that we can characterize the effects of compressing the column water vapor at realistic heights in real scenes. A subset of atmospheres with corresponding ground truth (water temperatures measured on buoys described in Section 5.1) was selected and the retrieved temperatures were compared when truncating column water vapor and when compressing column water vapor. Figure 4.50 compares the difference between retrieved temperatures compressing and truncating column water vapor to the difference between the retrieved temperature by truncating column water vapor and the ground truth temperature. If we call truncating the column water vapor our original method, and the difference between truncated and truth our original error, the magnitude of difference between compressed and truncated are smaller than our original error and, more commonly than not, further underestimate the actual temperature. Note the difference in scale between the two axis.
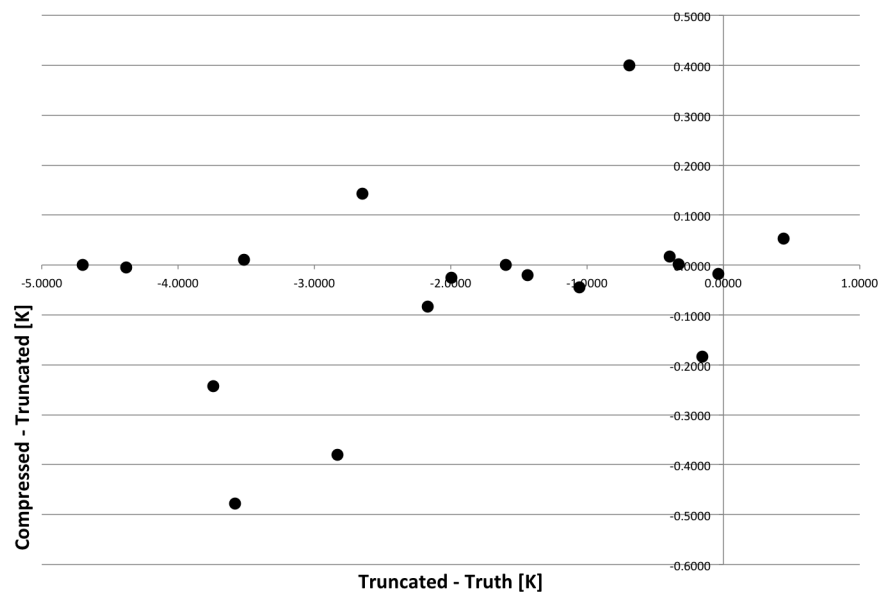
Figure 4.50: Difference in retrieved temperatures by compressing and truncating column water vapor compared to differences in retrieved temperature truncating column water vapor and ground truth temperatures for a subset of points over Lake Huron and Lake Superior.

As shown, the differences are all less 0.5 K, and there is no obvious relationship between differences to truth and differences between compressing and truncating column water vapor. This indicates that our errors to truth are not determined by how much water vapor we truncate. Also, as shown in Figure 4.51, we generally expected column water vapor to vary inversely with elevation. Therefore, the methodology will be implemented truncating layers of the atmosphere as the ground

altitude is increased as originally implemented.
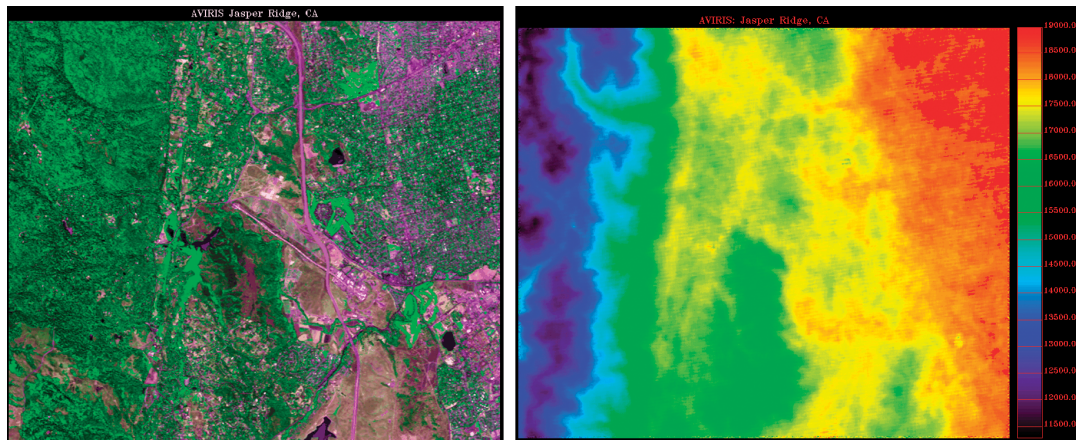
Figure 4.51: An AVIRIS image of Jasper Ridge on the left; the higher ridge is on the left and lower elevation on the right. The column water vapor data on the right (red being the highest values and purple being the lowest values) derived from the 940 nm absorption line data for the same image [Schott, 2007]. As expected, column water varies inversely with elevation.

4.7 Pixel Iteration and NARR Point Selection

To this point, the temporal interpolation of the NARR profiles and MODTRAN runs at each elevation, have been performed at the location of each NARR point on the native Lambert Conformal grid. However, the final steps in the generation of a complete LST product must involve pixel-wise operations. It was decided that four NARR points would be used in the generation of values for each pixel, and that the NARR points would be selected based on sampling in their native coordinate system to account for varying conditions in each direction. Because the Landsat pixels are natively in UTM coordinates and the NARR points are natively on a Lambert Conformal grid, the selection of these four points is not trivial. It is important to consider computational intensity in all of the remaining steps because the following operations are performed at every pixel and Landsat scenes are on the order of 7000 by 8000 pixels.

The latitude and longitude coordinates of each NARR point corresponding to the Lambert Conformal (i,j) coordinate are readily available and these latitude and longitude coordinates can be converted to UTM coordinates; in UTM coordinates, where locations are specified in meters, distances can be calculated with consideration for the curvature of the earth [NOMADS, 2012]. For each image, all UTM coordinates are computed in reference to the zone specified in the Landsat metadata.

Initially, distance calculations from every pixel to every NARR point pertinent to the scene were considered to find the four points nearest to each pixel. This introduced a number of problems including computation intensity and, because of the non-linearity of the coordinate systems, did

not always result in NARR points effectively characterizing the surround as desired. Instead, a systematic iteration through pixels from top to bottom and left to right was developed, so that the NARR points for each pixel are selected in quads from their native coordinate system.

Schematically represented as a grid, as they appear in the Lambert Conformal coordinates, although spacing is irregular in the UTM coordinate system, Figure 4.52 represents each NARR point using a letter. Each number represents a possible quad for a pixel to fall within. Quads are defined by the upper left NARR point. For example, quad three is defined by NARR point D in Figure 4.52.
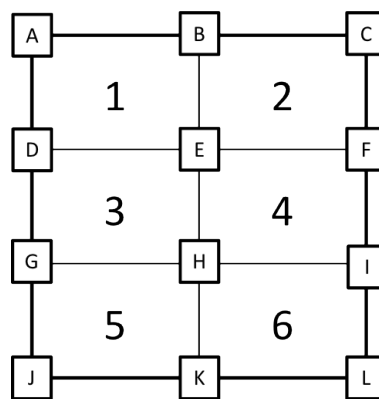
Figure 4.52: Schematic figure of NARR points and pixels for pixel interpolation and NARR point selection.

For the first pixel in each row, distance calculations are performed to every NARR point subset for the scene. The closest pixel above and left is found and the rest of the quad is defined using the Lambert Conformal NARR grid. If the closest point above and left is (i,j) in Lambert Conformal coordinates, the quad is (i,j), (i+1,j), (i,j+1), and (i+1, j+1). Continuing the example above, if the closest point above and left is point D, pixels in quad three are interpolated from NARR points D, E, G, and H. Once the quad for the first pixel in each row is determined, the quads for the rest of the pixels in the row, if iterating from left to right, can be determined using six distance calculations. Compared to using a method that finds the distance to every NARR point for each pixel, this can reduce the number of calculations by a factor of ten for every pixel.

For each successive pixel, assuming iteration from left to right in the image, there are six possible quads. The quad can remain the same or move right and the quad can remain the same or move up or down. Therefore, if the current pixel falls in quad three in Figure 4.52, the next pixel could fall in any of the six quads shown. This is based on the non-linearity between the UTM and Lambert Conformal coordinate systems and moving left to right across the image.

If the current pixel, r, falls in quad three, we want to determine the quad for the next pixel, s, moving from left to right. The distances from pixel s to NARR point D and NARR point F is calculated. If the distance to D is smaller, the quad does not move right, and if the distance to

NARR point F is smaller, the quad does move right. Similarly, the distances from the pixel s to NARR point B and NARR point H is calculated. If the distance to B is smaller, the quad moves up and if the distance to H is smaller, the quad does not move up. Finally, the distances from pixel s to NARR point E and NARR point K is calculated. If the distance to K is smaller, the quad moves down, and if the distance to E is smaller, the quad does not move down. This combination of move right or not, move up or not, and move down or not, determines the quad for pixel s. This requires six distance calculations and three logical operations to determine the four NARR points that will be used to interpolate the radiative transfer parameters for each pixel after the first pixel in each row.

4.8 Radiative Transfer Parameter Interpolation

Once the necessary NARR points have been determined, the last five pixel-wise operations can be performed. As discussed in Section 4.6, the radiative transfer parameters at each of the four NARR points in the quad are piecewise linearly interpolated to the elevation of the pixel specified in the DEM. This results in all three parameters at the appropriate elevation at the location of each NARR point in the quad. The final step is the spatial interpolation to the pixel location.

Various spatial interpolation methods were explored to optimally utilize the information available from the NARR points in the quad. The goal was to use an inverse distance weighting interpolation, weighting NARR points closer to the pixel of interest as more influential than those further from the pixel of interest. The chosen method, Shepard's method, is shown in Equations 4.3, 4.4, and 4.5.

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{4.3}$$

$$w_i = \frac{d_i^{-p}}{\sum_{j=1}^{n} d_j^{-p}} \tag{4.4}$$

$$F(x, y) = \sum_{i=1}^{n} w_i f_i \tag{4.5}$$
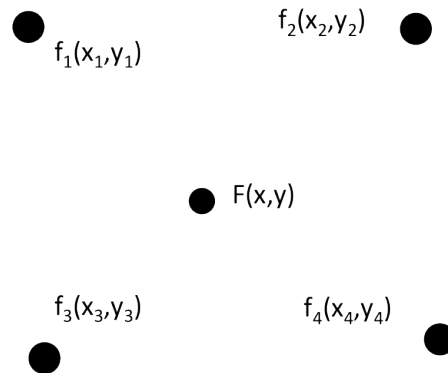
Figure 4.53: Schematic figure for points in Shepard's Method.

As shown in Figure 4.53, (x_i, y_i) are the coordinates of the points being interpolated, f_i are the values being interpolated, (x,y) is the coordinate being interpolated to, and F is the final interpolated value. In Equation 4.3, d_i are distance values from each NARR point to the pixel location; in Equation 4.4, n is the number of points being interpolated, in our case this is four, and w_i are weighting values for each NARR point, inversely related to the distance of the point from the interpolated value. Also, p is a weighting exponent. This is generally an arbitrary positive real number; the default value of 2 is used. Equation 4.5 calculates the final interpolated value as a weighted summation of the original values [Shepard, 1968].

As with both the temporal and height interpolations, we want to evaluate this step by isolating the error contributed by this interpolation. Like the height interpolations, we used radiosonde data as a truth profile for comparison. The nearest radiosonde profile will be corrected to the location of a surface weather station and this will be used as a truth profile. Using this profile, the apparent ground temperature was generated using the process described in Section 4.1 at the same nine heights given in Section 4.6. These are the "truth" apparent ground temperatures. The location of the surface weather station is then treated as the desired pixel location. The four nearest NARR points are selected as described in Section 4.7 and the radiative transfer parameters are generated at the same nine heights at each of these four NARR locations as described in Section 4.6. This utilizes the height interpolation within the tape5 file but not the height interpolation to the elevation of each individual pixel so as not to compound errors.

These radiative transfer parameters at each height are interpolated to the location of the surface weather station and these interpolated radiative transfer parameters are used to generate an array of apparent ground temperatures. These temperatures are compared to the array of "truth" temperatures generated from the radiosonde profile to analyze the error contributed by the spatial interpolation. In order to avoid compounding errors with temporal interpolation, the radiosonde profiles are corrected to surface weather at 15 Z and the 15 Z NARR profiles were used. This is another best estimate of truth and should be considered only a test of the reasonability of implementing Shepard's method.

Figure 4.54 and Figure 4.55 show the error contributed by the spatial interpolation for the same

surface weather location in the northeast used in Sections 4.5 and 4.6 for dates in February and August respectively. As with previous investigations, three temperatures are considered.
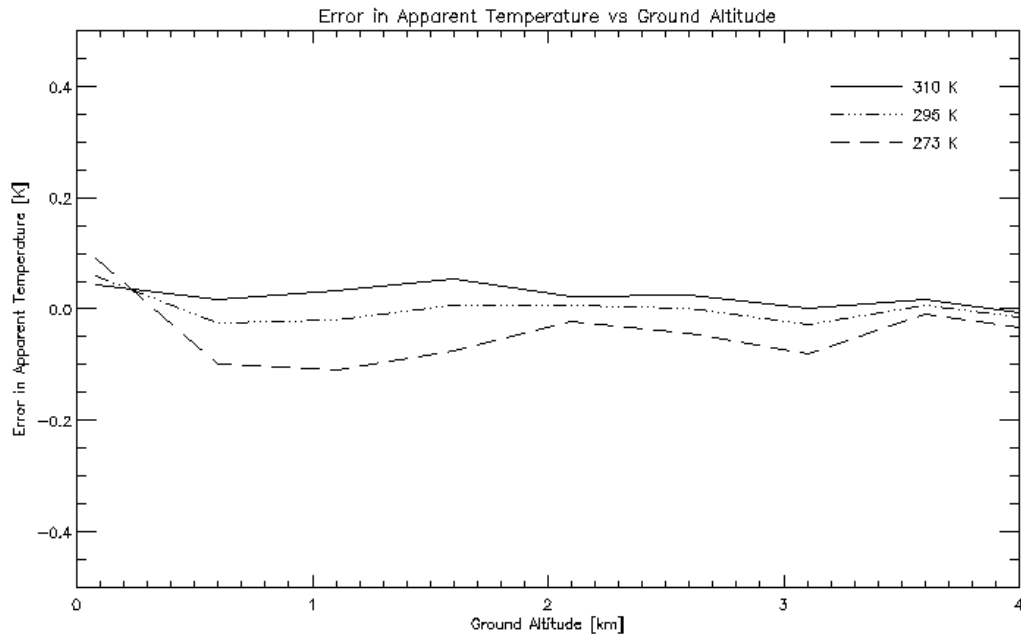
Figure 4.54: Errors in apparent temperature contributed by spatial interpolation to a location in the northeastern United States in February.
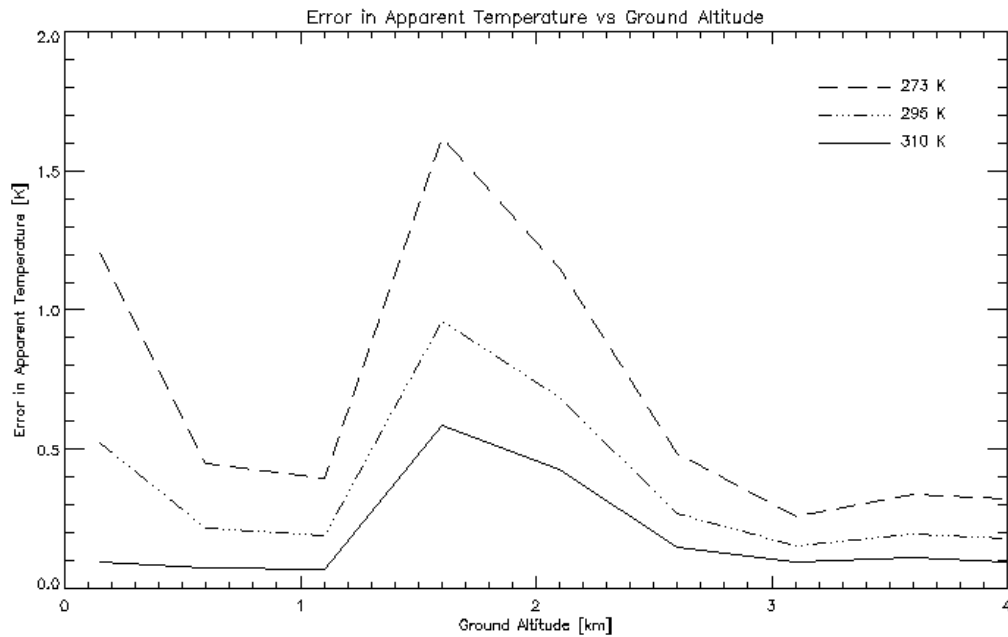
Figure 4.55: Errors in apparent temperature contributed by spatial interpolation to a location in the northeastern United States in August.

Similar to the results from the temporal interpolation, the results in the colder weather are much better than those in the warmer weather. Results in February are acceptable, even better than expected, and results from August also are reasonable enough to allow us to proceed with this method. While this is a brief and incomplete sensitivity study, as were most throughout the implementation of the methodology, we believe it is more worthwhile to implement an entire process and compare to ground truth to better investigate limiting factors before optimizing any one step.

4.9 Deliverables

Because the goal of this work is atmospheric compensation, the deliverable is not a single temperature value at each pixel. Rather, the deliverable includes all components necessary to determine the LST once the emissivity is known. For each Landsat scene, a five band geotiff image is produced, the same size as the original Landsat image. Table 4.4 details each band of this file. Explained further in Section 5.3.3, we also expect to extend this to a six band geotiff with some form of confidence metric or error estimate.

Table 4.4: Details for each layer of the deliverable five band geotiff file.

Band	Contents	Units
0	Landsat Thermal Band Radiance (L_{obs})	$Wm^{-2}sr^{-1}\mu m^{-1}$
1	Elevation	m
2	Transmission	
3	Upwelled Radiance	$Wm^{-2}sr^{-1}\mu m^{-1}$
4	Downwelled Radiance	$Wm^{-2}sr^{-1}\mu m^{-1}$

4.10   Concluding Remarks

This chapter aimed to explain each step in the automated process of LST prediction. We began by describing our selected method for generating radiative transfer parameters from MODTRAN output and a general overview of our entire process in Section 4.1 and 4.2. We then detailed each individual step in the process beginning with data manipulation in Sections 4.3 and 4.4. Sections 4.5 and 4.6 describe temporal and height interpolations respectively and Section 4.7 and 4.8 deal with the pixel-wise operations required to generate a complete operational LST product.

Most of the above steps included not only an explanation but a brief validation. However, a common theme was the need to implement and consider results from an entire process, to investigate feasibility and limiting factors, before revising or optimizing any one step. With an entire process implemented, such results can be generated; results are presented in Chapter 5. We discuss our ground truth sites, how we quantify our error, validation of our methodology at these ground truth sites, and how we develop a confidence metric. Based on these results, we outline future work in Chapter 6, which includes extension to a global dataset and improvements and finalization of the confidence metric.

Chapter 5

Results

This chapter aims to summarize our validation results. In Chapter 4, radiosonde profiles were used as truth profiles to evaluate each of the interpolation methods, usually aiming to quantify only error from that step in the final predicted temperature. Because these radiosonde profiles do not provide absolute truth, sensitivity studies in Chapter 4 only provided a check of the feasibility of our interpolation methods, not a rigorous evaluation of error. In this chapter, temperatures from buoys or instrument platforms will be used to evaluate the error in apparent temperature from the entire process. Because the emissivity of water is known, we can compute the temperature predicted by our process and compare this to the ground truth temperature observed from the buoy or platform. In Section 5.1 we describe each of the ground truth sites and in Section 5.2 we describe the quality of the results and magnitude of errors at each of these sites. Based on these results, we describe in Section 5.3 methods used in the development of and current expectations for a confidence metric to be included with the product.

5.1   Ground Truth Sites

From each of the sites described below, we were given or can obtain the ground truth land surface temperature of the water. We consider that different processes were used to obtain these measurements, and each has some associated uncertainty, but in all cases we call what we assume to be the accurate land surface temperature ground truth data in our validation. Here we first encounter the notation for specifying Landsat scenes. The second Worldwide Reference System (WRS-2) specifies a nominal scene center as an integer path and integer row in a grid that covers the globe [Science, 2014]. WRS-2 path, row is used only for scene specifications; scene centers and corners can vary and are obtained from the metadata.

Ground truth data at two sites was supplied by JPL. JPL maintains buoys in Lake Tahoe and a platform in the Salton Sea, with instrumentation including surface contact thermistors, nadir viewing calibrated radiometers, and weather stations. The water temperature is corrected for

the skin temperature using the radiometric temperature, surface contact temperature and down-welled radiance to account for the Landsat passband because the individual radiometers do not [Hook et al., 2004] and [Hook et al., 2007]. These corrections were made and the ground truth land surface temperature of the water was provided by JPL for the ground truth sites discussed below. It is difficult to estimate a single uncertainty value for the final ground truth temperature due to the number of steps in the process as well as the variation with time of day and windspeed. However, two types of radiometers are used, calibrated to ±0.2 K and ±0.1 K respectively, and the temperature sensors that capture the bulk water temperature have maximum errors of $\pm0.25°$C but were found to meet the typical errors of $\pm0.12°$C [Hook et al., 2002]. The rest of the truth data is derived from buoys owned and maintained by the National Data Buoy Center (NDBC). While there are many buoys, and likely multiple buoys even in the scenes specified, not all provide the necessary variables for deriving the land surface temperature. For all scenes below with buoys owned and maintained by the NDBC, the observed water temperatures for each buoy are obtained from the NDBC and corrected to the ground truth land surface temperature using the skin temperature method [Schott et al., 2012], [Padula and Schott, 2010]. An error propagation analysis of this method showed the expected error to be approximately 0.35 K, mostly contributed by uncertainties in the thermistor used to measure the bulk temperature [Padula and Schott, 2010]. Availability of appropriate meteorological data, watch radius, and reliable data availability are all important to choosing to use a particular buoy [NDBC, 2014a].

5.1.1 The Salton Sea

The Salton Sea is located in southeastern California. It is a saline lake that is 60 km long and 30 km wide and has an average depth of 9 m; the surface of the lake is approximately 70 m below sea level. Measurements are made from a platform located at 33.22532°N, 115.82425°W, which falls within WRS-2 path 39, row 37, shown in Figure 5.1. Ground truth data from this site is available from 2006 to present [Hook and Rivera, 2013b]. Ground truth data from this site is provided by JPL with corrections to surface temperature made as described in Section 5.1.
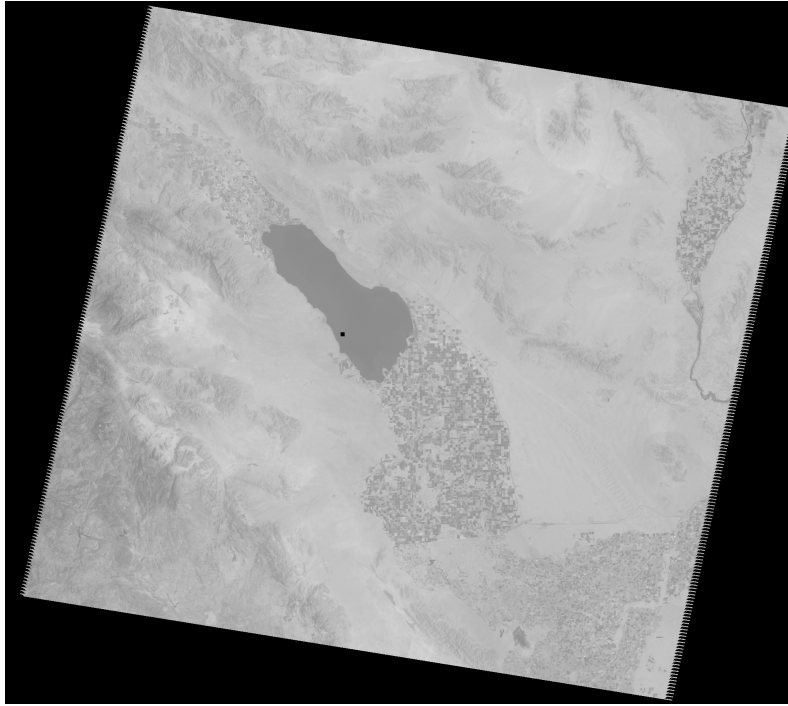
Figure 5.1: Landsat scene including the Salton Sea (path 39, row 37). The black square represents the approximate location of the platform.

5.1.2 Lake Tahoe

Lake Tahoe is located on the California-Nevada border approximately 1895 m above sea level. It has a surface area of 500 km^2, an average depth of 330 m and is known for its high water clarity. It does not freeze in the winter because of its large thermal mass. There are four permanently moored buoys in the lake that provide various observables. For the purposes of this work, we focused on a single buoy, referred to as TB4, located at 39.155°N, 120.0721667°W; this falls within WRS-2 path 43, row 33 and is shown in Figure 5.2. This buoy has data available beginning in May 1999 through present [Hook and Rivera, 2013a]. Corrections to surface temperature are made as described in Section 5.1 and this data is provided by JPL.
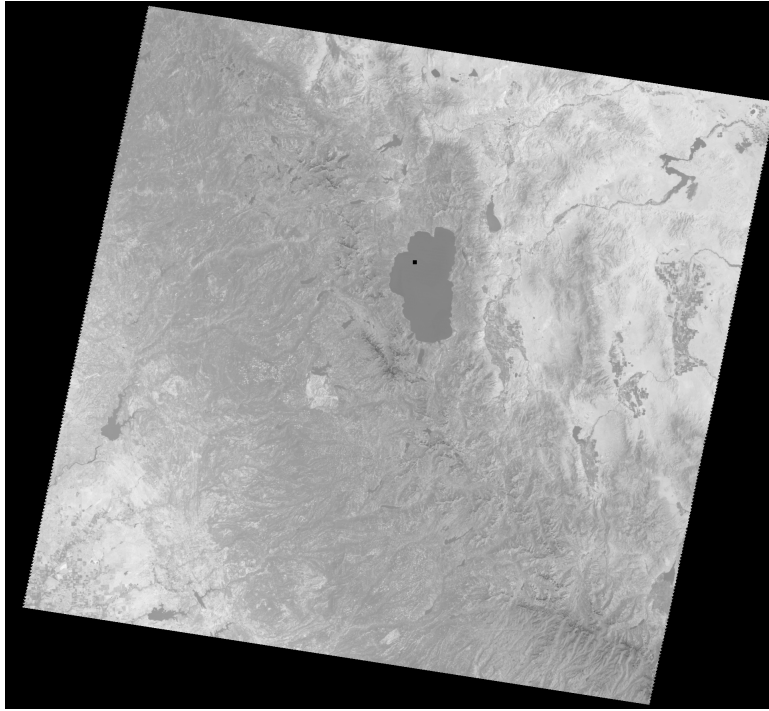
Figure 5.2: Landsat scene including Lake Tahoe (path 43, row 33). The black square represents the approximate location of the buoy.

5.1.3   Rochester

A 2.4 m foam hull buoy owned and maintained by the NDBC is located in eastern Lake Ontario northeast of Rochester at 43.619°N, 77.405°W (station 45012). This site is 74.7 m above sea level. In order to avoid problems with ice, this buoy is retrieved for a period of time in the winter and redeployed in the spring, so the data is only available for a portion of the year. Retrieval and redeployment dates can vary but data for some portion of the year is available for years 2002 to present. This buoy falls in the overlap between Landsat paths, so it falls within WRS-2 path 16, row 30 and path 17, row 30 as shown in Figures 5.3 and 5.4 [NDBC, 2013c].
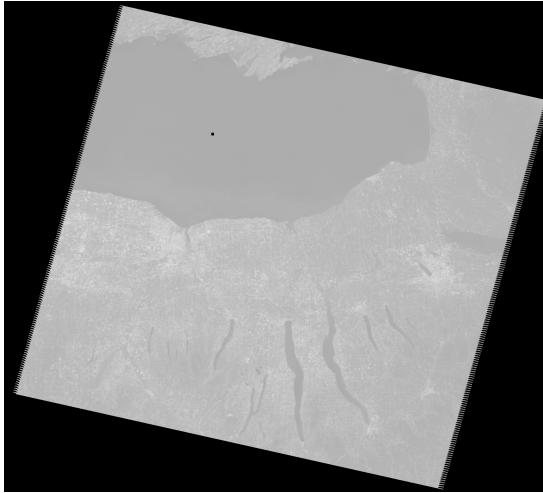
Figure 5.3: Landsat scene including Rochester, NY (path 16, row 30). The black square represents the approximate location of the buoy.
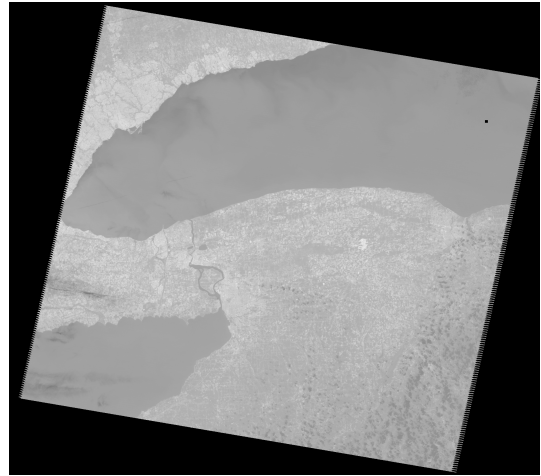
Figure 5.4: Landsat scene including Rochester, NY (path 17, row 30). The black square represents the approximate location of the buoy.

5.1.4   Delaware Bay (Delmar)

A 3 m discus buoy owned and maintained by the NDBC is located in Delaware Bay near the Delaware-Maryland line at 38.464°N, 74.702°W (station 44009). This buoy is located at sea level. Standard meteorological data is available as early as 1984 but most other forms of data from this buoy are available beginning in the late 1990s. This station stopped transmitting data in December 2012 but will be restored to regular functionality as soon as it can be serviced. This buoy also falls in the overlap of Landsat paths, so it provides ground truth data for WRS-2 path 13, row 33 and path 14, row 33 as shown in Figures 5.5 and 5.6 [NDBC, 2013b].
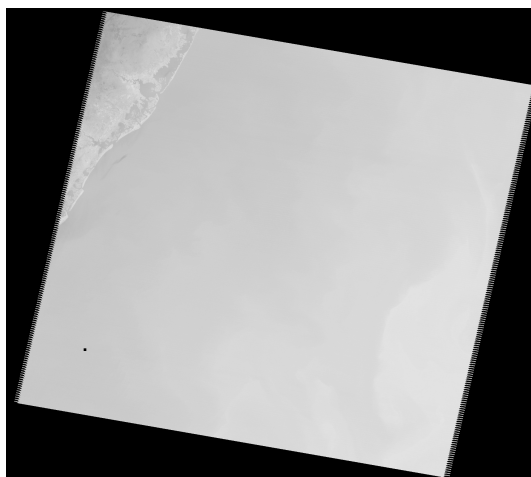
Figure 5.5: Landsat scene including Delaware Bay (path 13, row 33). The black square represents the approximate location of the buoy.
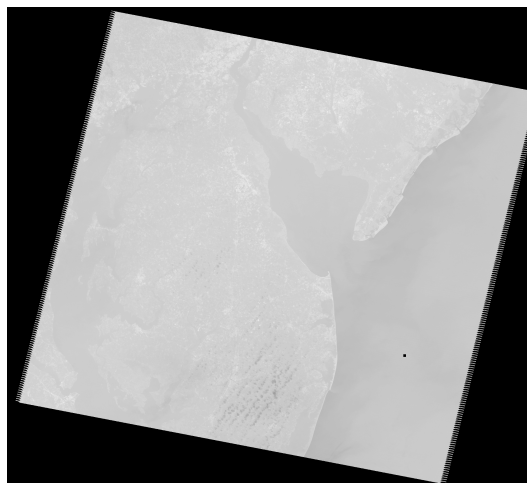
Figure 5.6: Landsat scene including Delaware Bay (path 14, row 33). The black square represents the approximate location of the buoy.

5.1.5   Georgia Coast

A 3 m discus buoy owned and maintained by the NDBC is located at sea level off the coast of Georgia, southeast of Savannah at 31.402°N, 80.869°W (station 41008). This buoy is located in Gray's Reef National Marine Sanctuary. Standard meteorological data is first available in 1988 and other forms of data become available in the late 1990s and mid 2000s; all are through present. This buoy falls within WRS-2 path 16, row 38 as shown in Figure 5.7 [NDBC, 2013a].
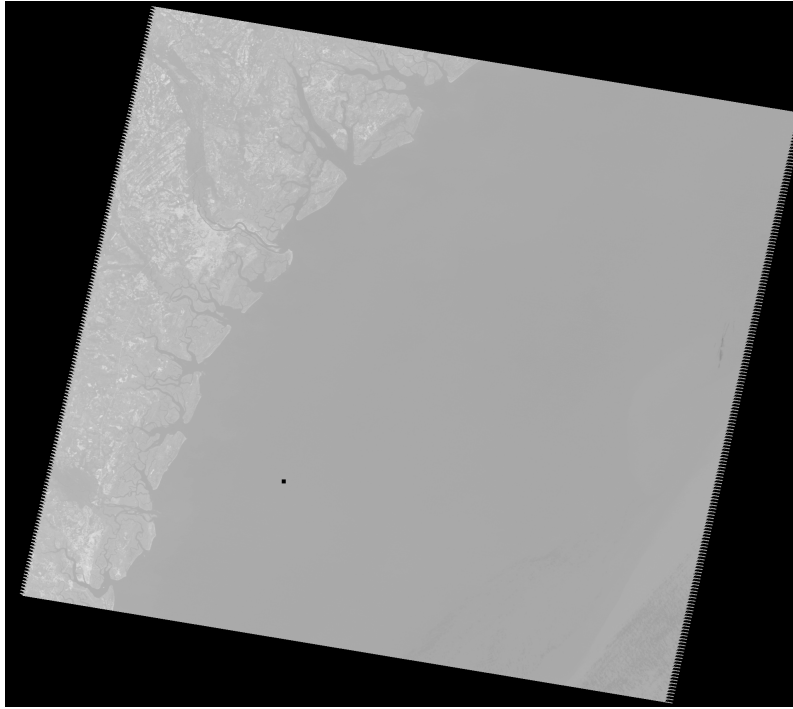
Figure 5.7: Landsat scene off the coast of Georgia (path 16, row 38). The black square represents the approximate location of the buoy.

5.1.6 California (Santa Maria)

A 3 m discus buoy owned and maintained by the NDBC is located at sea level off the coast of California, 21 NM northwest of Point Arguello, near Santa Maria, at 35.000°N, 120.992°W (station 46011). Some forms of meteorological data from this buoy are available as early as 1980, with all currently available variables beginning no later than the mid 2000s. The buoy falls within WRS-2 path 43, row 36, as shown in Figure 5.8 [NDBC, 2014f].
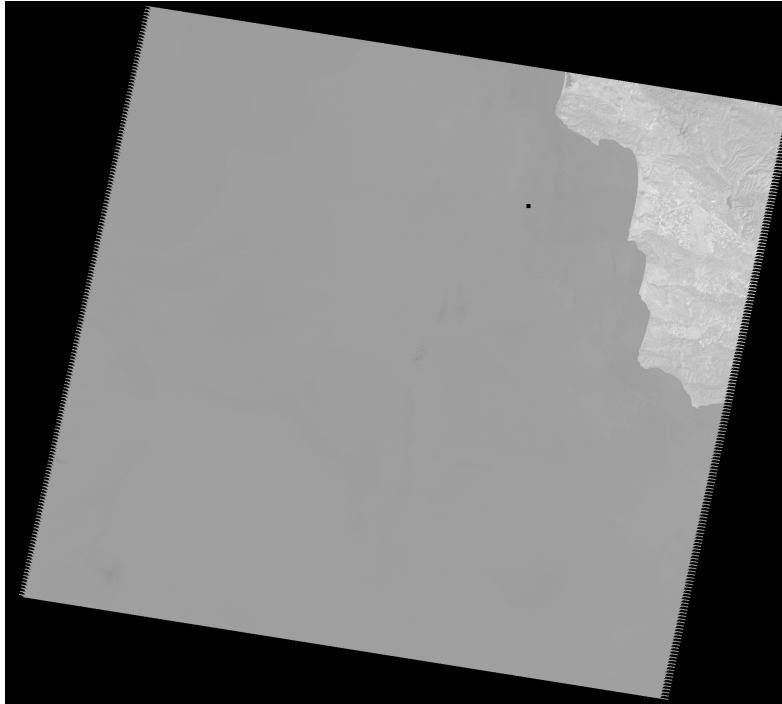
Figure 5.8: Landsat scene off the coast of California, near Santa Maria (path 43, row 36). The black square represents the approximate location of the buoy.

5.1.7 California (Santa Monica)

A 3 m discus buoy owned and maintained by the NDBC is located off the coast of California, 33 NM south west of Santa Monica at 33.749°N, 119.053°W (station 46025). This buoy is located near the Santa Monica Basin and falls within WRS-2 path 41, row 37 as shown in Figure 5.9 [NDBC, 2014g]. This buoy first started reporting meteorological data in 1982; some reported variables were added in the late 1990s and others in the late 2000s. All are reported through the present.
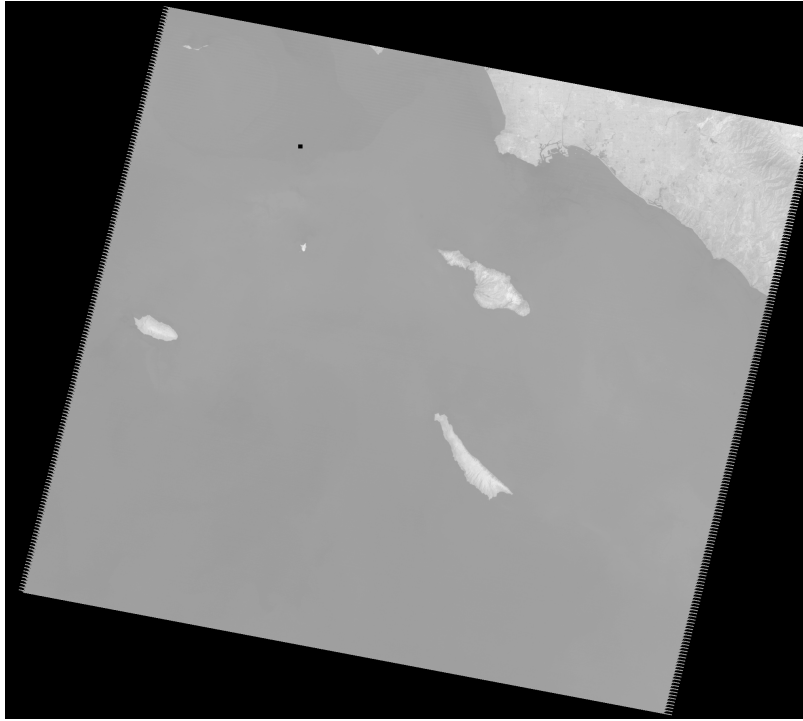
Figure 5.9: Landsat scene off the coast of California, near Santa Monica (path 41, row 37). The black square represents the approximate location of the buoy.

5.1.8   Lake Huron

Two buoys owned and maintained by the NDBC in Lake Huron, falling within WRS-2 path 20, row 29 were selected. A 2.4 m foam hull buoy in Northern Huron is located northeast of Alpena, Michigan at 45.351°N, 82.84°W (station 45003) [NDBC, 2014c], and a 3 m discus buoy in southern Huron is located east of Oscoda, MI at 44.283°N, 82.416°W (station 45008) [NDBC, 2014e]. Both buoys are at an elevation of 177 m above sea level. These buoys were launched and began reporting meteorological data in 1980 and 1981 respectively, with the rest of the currently reported variables being reported by the early 2000s. Ground truth data was obtained from both for validation. The locations of these buoys is shown in a Landsat scene in Figure 5.10.
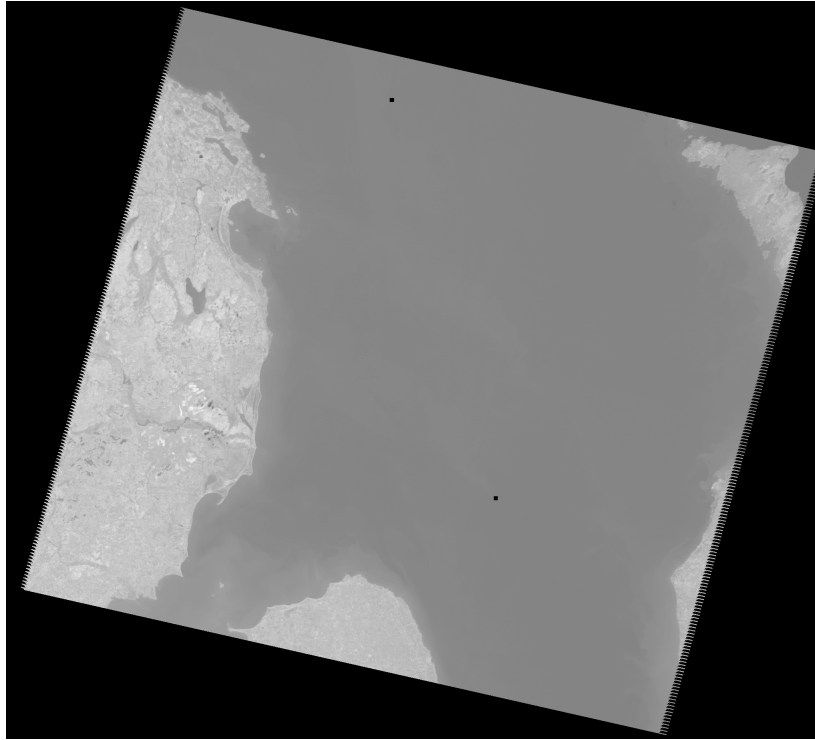
Figure 5.10: Landsat scene showing part of Lake Huron. The black squares represent the approximate locations of the buoys.

5.1.9   Lake Superior

Similarly, two buoys owned and maintained by the NDBC in Lake Superior, within WRS-2 path 24, row 27, were selected. Northeast of Hancock, mid Lake Superior, a 3 m discus foam buoy is located at 48.061°N, 87.793W (station 45001) [NDBC, 2014b], and another 3 m foam buoy is located in eastern Superior northeast of Marquette, MI at 47.584°N, 86.587°W (station 45004). These buoys, both 183 m above sea level, were launched and first began reporting data in 1979 and 1980 respectively, with various additional reportings added throughout the 2000s. The second buoy, northeast of Marquette, went adrift July 2014, and is no longer reporting data from the position above [NDBC, 2014d]. The location of both of these buoys within the Landsat scene is shown in Figure 5.11.
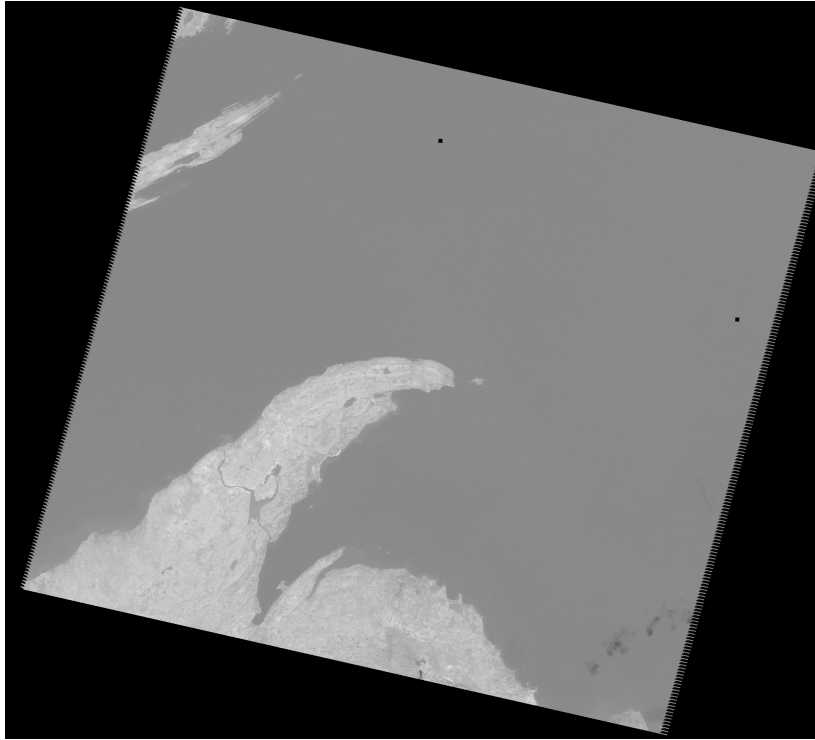
Figure 5.11: Landsat scene showing part of Lake Superior. The black squares represent the approximate locations of the buoys.

5.2   Validation of Methodology

Initial scenes to be processed were selected for spatial and temporal variation based on the location of the test sites described in Section 5.1 and the availability of ground truth data. The goal was to validate the methodology with a variety of atmospheres, including differences in location, weather, season, and elevation. Initially, all Landsat 5 scenes were processed for validation due to the large archive and convenience of all SLC-on scenes; only Landsat 5 validation is included in this section. Personal communication with the Landsat Calibration team indicated that Landsat 5 is currently miscalibrated with a bias of -0.33 K at 300 K. This bias correction is expected to be implemented, but has not yet been applied to scenes in the archive, so it was applied for this work post-download. Therefore, the mean calibration bias for all Landsat 5 scenes included in this validation is 0.0 K with a standard deviation of 0.73 K. A map of all the ground truth sites is shown in Figure 5.12. The predicted land surface temperatures for each site at the pixel location of each buoy are compared to the ground truth temperatures. It is important to consider scene variety and selection when considering the results. In order to most accurately validate the methodology described in Chapter 4, initially only cloud free scenes were considered. As mentioned, ground truth data for Lake Tahoe and the Salton Sea were provided by JPL; scenes from these sites were screened and only those that lent themselves to good temperature retrieval (clear skies, no clouds,

well-behaved atmospheres, etc.) were used. Ground truth data from the other sites is obtained from
the buoys and adjusted to the ground truth land surface temperature using the skin temperature
method [Schott et al., 2012], [Padula and Schott, 2010]. Scene selection at these locations was less
stringent than the selection requirements for Lake Tahoe and the Salton Sea. For all possible scenes
over a given time period, only scenes that were cloud free in the vicinity of the buoy based on a
visual analysis were used, but this was without consideration of other atmospheric variables. This
also leads to differing numbers of scenes per location.
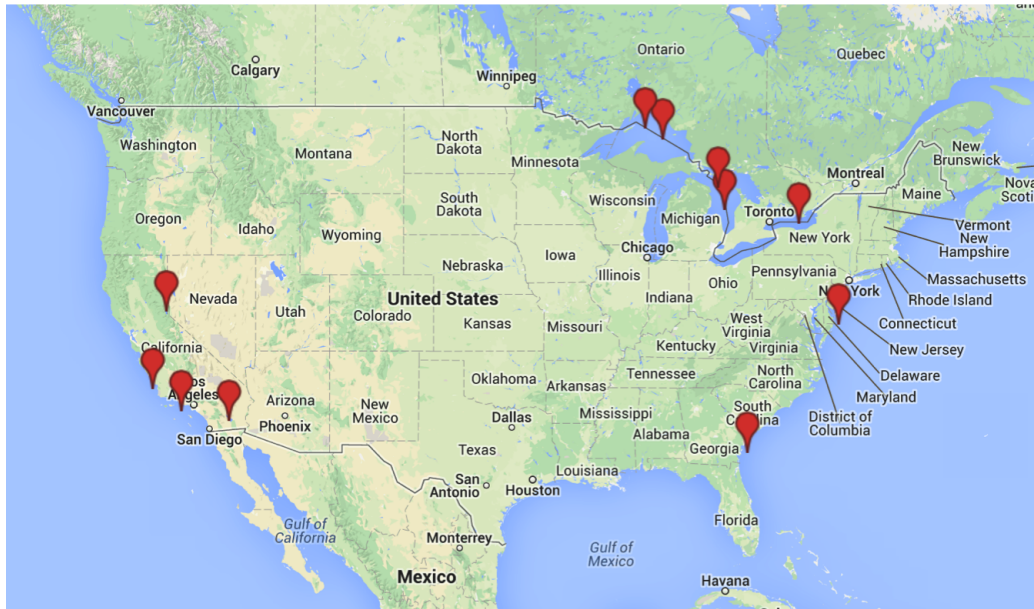
Figure 5.12: Location of all ground truth sites from Section 5.1 over the United States.

Error in the following sections is defined in Equation 5.1; a negative value indicates that the
land surface temperature retrieval process underestimated the temperature and a positive value
indicates that our process overestimated the temperature. We believe that for the applications
targeted with this product as discussed in Section 3.4.1, errors with magnitudes between 1 K and
2 K would be very acceptable.

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

5.2.1 The Salton Sea

Error Histogram - The Salton Sea
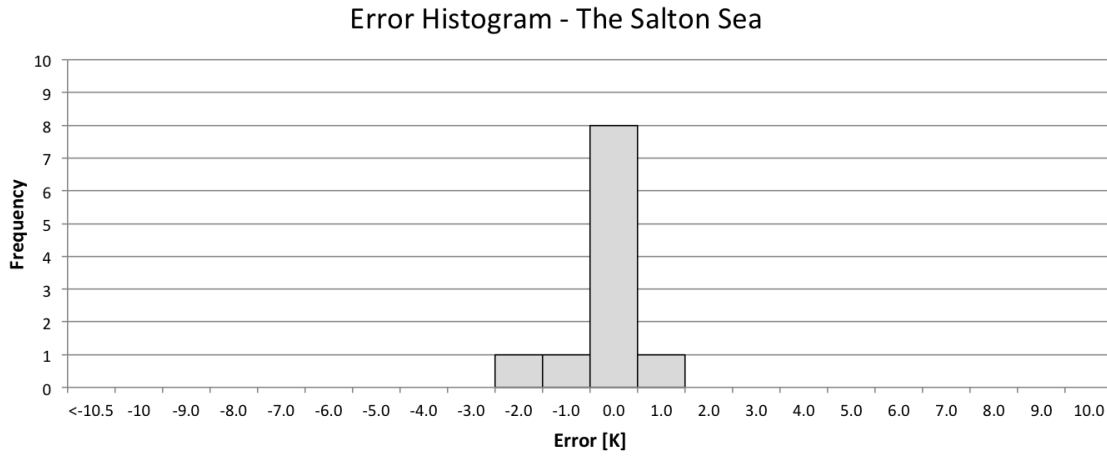
Figure 5.13: Histogram of error values for cloud free scenes over the Salton Sea.

Only 11 scenes were processed and compared to ground truth data for the Salton Sea validation site; this was the smallest dataset for any ground truth site. Figure 5.13 shows a histogram of the error values for all cloud free scenes over the Salton Sea. For these 11 scenes, the mean error value is -0.120 K and the standard deviation is 0.558 K. Although the dataset is small, recall that the elevation of the Salton Sea is below sea level. We expect the temperature to be more difficult to retrieve at lower elevations due to the large volume of atmosphere being compensated for; even for a standard elevation, these results are very encouraging.

5.2.2   Lake Tahoe

Error Histogram - Lake Tahoe
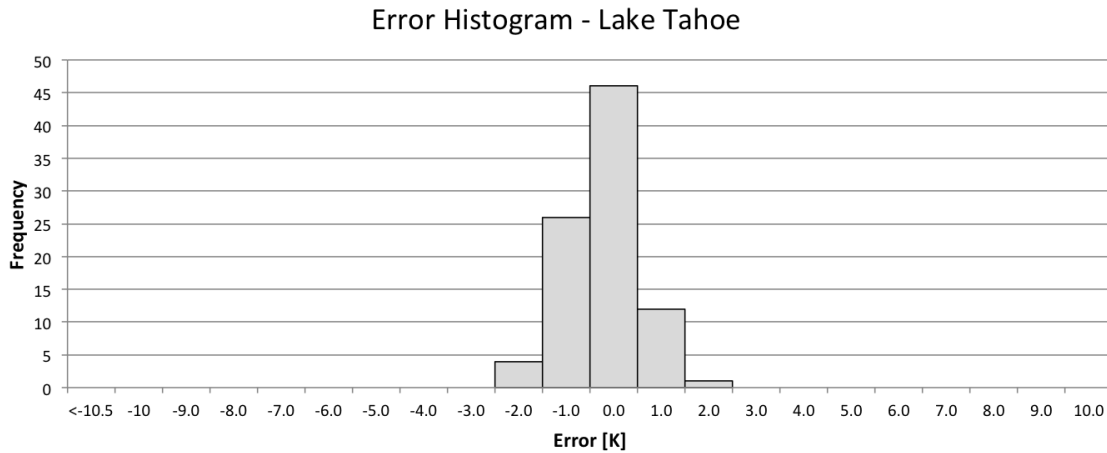
Figure 5.14: Histogram of error values for cloud free scenes over Lake Tahoe.

There were 89 cloud free scenes processed over Lake Tahoe; recall that this ground truth site had a much higher elevation than all other sites considered. A histogram of error values for these 89 comparisons to ground truth is shown in Figure 5.14. The mean error value for this dataset is -0.213 K and the standard deviation is 0.713 K. This was by far the largest cloud free dataset for any single location. With a greater number of scenes, the spread is slightly larger. The slight negative shift in the dataset, illustrated by both the mean error and the histogram, is also more apparent. However, 84 of 89 scenes fall within the center three bins of the histogram with errors [-1.5 K, 1.5 K]. These results are also very encouraging.

5.2.3 Rochester
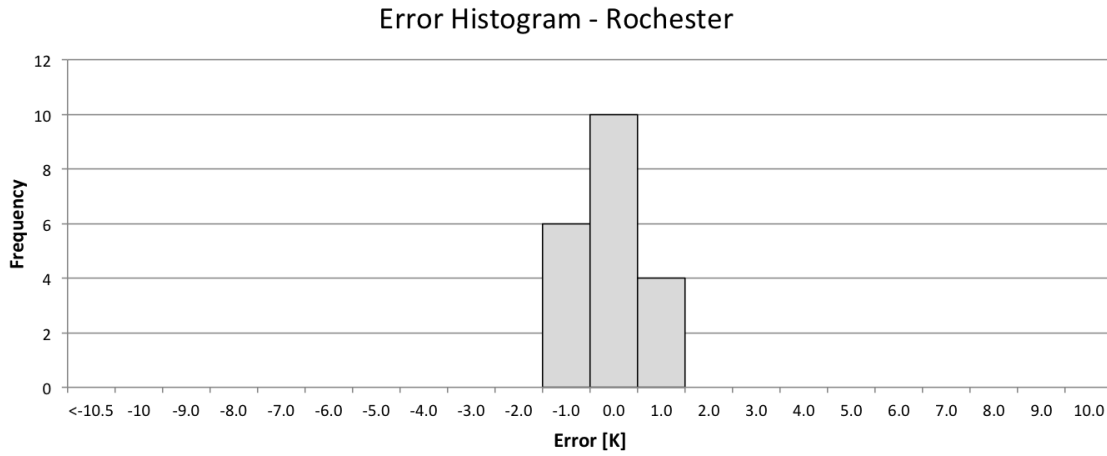
Figure 5.15: Histogram of error values for cloud free scenes over Rochester.

The cloud free dataset over Lake Ontario near Rochester, NY contained 20 scenes. The mean error for these 20 scenes is -0.068 K and the standard deviation is 0.639 K. A histogram of these errors is shown in Figure 5.15. The size of this dataset is comparable to many of the other ground truth sites, but has a slightly smaller standard deviation. Note that all cloud free scenes at this location fall within [-1.5 K, 1.5 K] on the histogram.

5.2.4   Delaware Bay (Delmar)
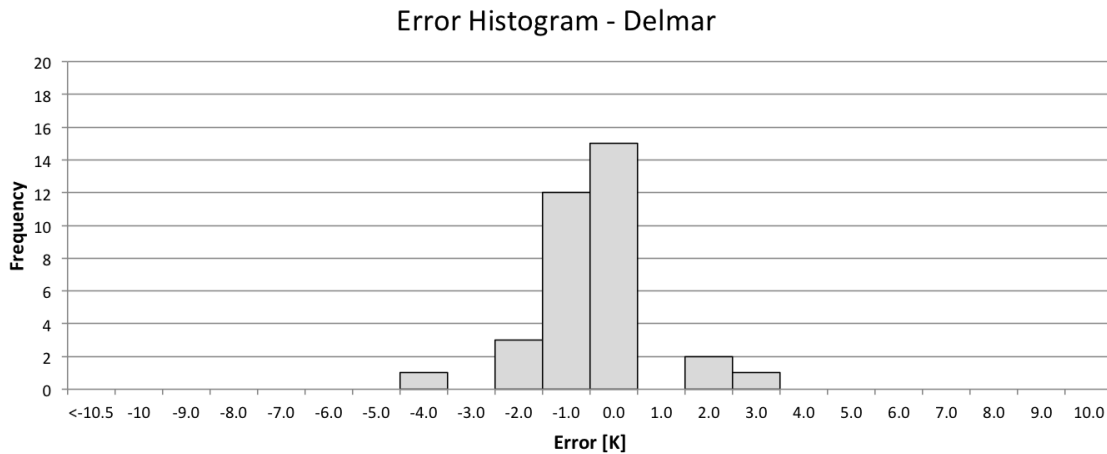
Figure 5.16: Histogram of error values for cloud free scenes over Delaware Bay.

There were 34 cloud free scenes processed off the Delaware Maryland coast. The mean error of this dataset is -0.447 K and the standard deviation is 1.179 K. The histogram of errors for these scenes is shown in Figure 5.16. This is the second largest dataset for any single ground truth site (with only fewer scenes than Lake Tahoe) and also the second largest standard deviation. Note that the negative shift in the data is also more apparent in the histogram, and while there is a large cluster of scenes in the center bins, there are also two scenes that have larger magnitude errors, such as [3K, 4K]. Both Delmar, and Georgia discussed below, are likely to have a greater number of hotter, more humid atmospheres where we expect the atmospheric compensation to be more difficult. Therefore, these results are not surprising, and the large number of scenes with accurate retrievals is still very encouraging.

5.2.5 Georgia Coast

Error Histogram - Georgia
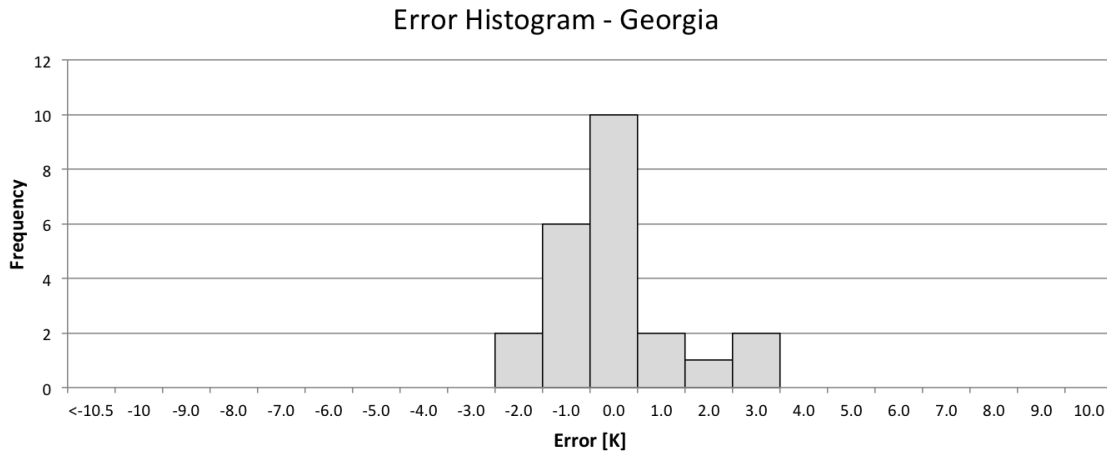
Figure 5.17: Histogram of error values for cloud free scenes off the Georgia coast.

The dataset off the coast of Georgia contained 23 cloud free scenes; a histogram of error values is shown in Figure 5.17. The mean error of this dataset is 0.041 K and the standard deviation 1.267 K. Note that this is the only site that does not have a negative mean error and that this is the largest standard deviation. While the magnitude of the mean error is close to zero, this is likely due to the combination of a larger number of small negative errors and the three larger (greater than 1.5 K) positive errors, as shown in the histogram. As described above for Delmar, this ground truth site was chosen because it was expected to have a larger number of days with hotter and more humid atmospheres that could be found at some of the other ground truth sites. These atmospheres are particularly difficult to compensate for because the water vapor in the atmosphere is both difficult to accurately quantify and has such a significant effect on the generation of the radiative transfer parameters.

5.2.6 California (Santa Maria)
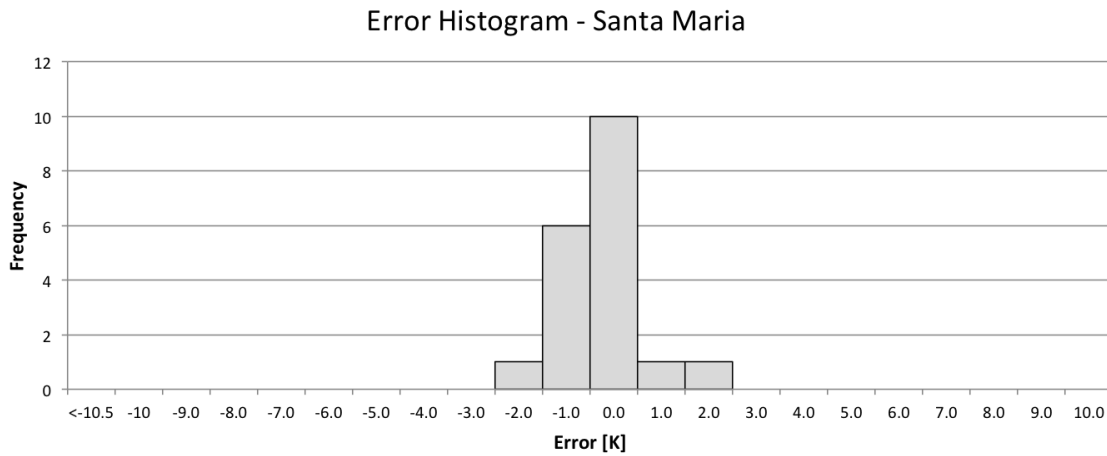
Figure 5.18: Histogram of error values for cloud free scenes off the California coast near Santa Maria.

The were 19 scenes processed off the coast of California near Santa Maria. The mean error of these scenes is -0.219 K and the standard deviation is 0.789 K. Only two scenes fall outside the error range of [-1.5 K, 1.5 K], although there is an apparent negative shift in the data as illustrated by both the mean and histogram.

5.2.7 California (Santa Monica)
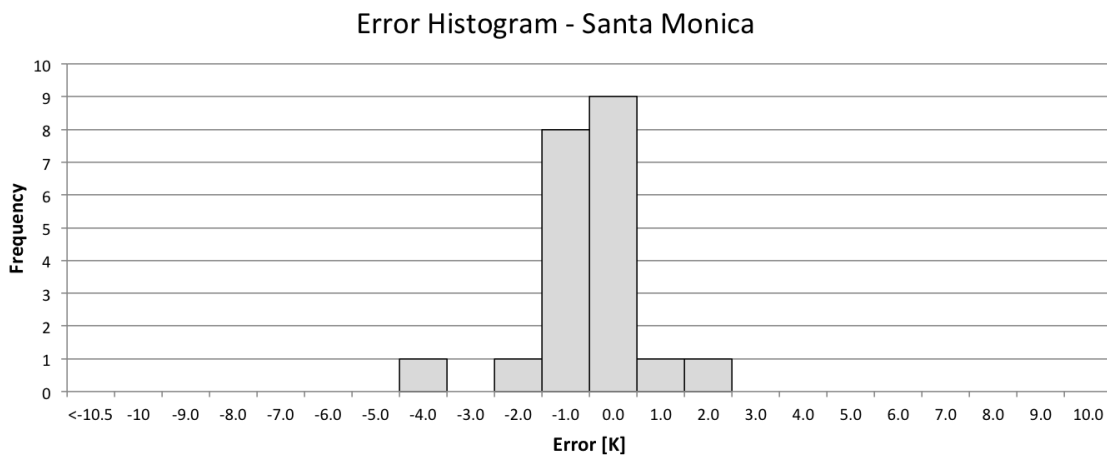
Figure 5.19: Histogram of error values for cloud free scenes off the California coast near Santa Monica.

There were 21 scenes in the cloud free data set off the coast of California near Santa Monica. A histogram of these errors is shown in Figure 5.19. The mean error for this dataset is -0.574 K and the standard deviation is 1.089 K. Similar to the scenes near Santa Maria, the negative shift in the data is apparent as the bin centered at -1 K is almost as large as the bin centered at 0 K. However, only three scenes fall outside [-1.5 K, 1.5 K] and the very accurate retrieval over a large portion of the scenes is still very encouraging.

5.2.8   Lake Huron
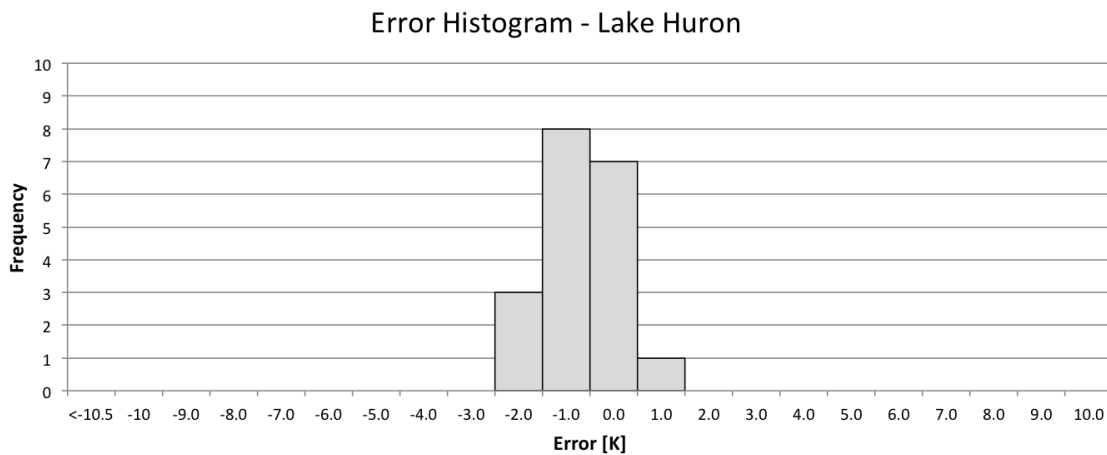
Figure 5.20: Histogram of error values for cloud free scenes over Lake Huron.

The data set over Lake Huron contained 19 cloud free scenes, with a mean error of -0.695 K and a standard deviation of 0.820 K. While it appears that there are more outliers at other locations, and only three scenes fall outside [-1.5 K, 1.5 K], the negative shift is very obvious here as the bin centered at -1 K is larger than the bin centered at 0 K. Even so, the magnitude of the mean error is still well less that 1 K and the distribution of the scenes is tightly clustered near zero, keeping the standard deviation small.

5.2.9 Lake Superior
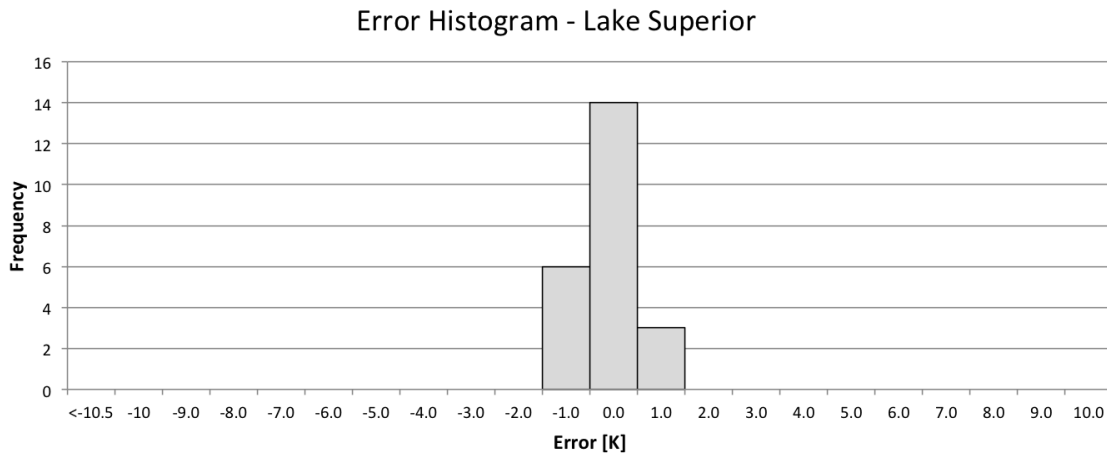
Figure 5.21: Histogram of error values for cloud free scenes over Lake Superior.

Finally, the dataset over Lake Superior contained 23 cloud free scenes. The mean error of this dataset is -0.167 K and the standard deviation 0.676 K. This location provided very good retrieval results; all scenes fall within [-1.5 K, 1.5 K] with the center bin larger than any others creating a small mean and a small standard deviation.

5.2.10 Summary of Initial Errors

Table 5.1 summarizes results for each individual location and all locations collectively, and Figure 5.22 shows a histogram of all cloud free scenes for all locations. There are 259 cloud free scenes across all eleven ground truth sites and the mean error is -0.267 K and the standard deviation 0.900 K; 90% of the scenes fall within [-1.5 K, 1.5 K]. With a mean error that has a magnitude well less than 0.5 K, and a standard deviation less than 1 K, these results are very encouraging. It validates our methodology, including the datasets, the tools within the process, and the various interpolations chosen and implemented. However, there is an observable negative shift in the histograms of cloud free scenes. Statistical tests indicate that mean errors for each individual location are not significantly different, as shown in Appendix A.4. This tells us that the small negative shift that we see in our data is not contributed from any one locations or group of locations. Therefore, we will investigate other contributing factors and methods of improving this negative bias. However, it is important to remember that the mean and standard deviation of errors in the validation dataset for cloud free scenes already provide very useful results for the applications at which this product is aimed.

Table 5.1: Summary of initial validation results for cloud free scenes for each individual location and all locations collectively.

Location	Mean [K]	Standard Deviation [K]	Number of Scenes
Salton	-0.12	0.558	11
Tahoe	-0.213	0.713	89
Rochester	-0.068	0.639	20
Delmar	-0.447	1.179	34
Georgia	0.041	1.267	23
Santa Maria	-0.219	0.789	19
Santa Monica	-0.574	1.089	21
Huron	-0.695	0.820	19
Superior	-0.167	0.676	23
Total	-0.267	0.900	259
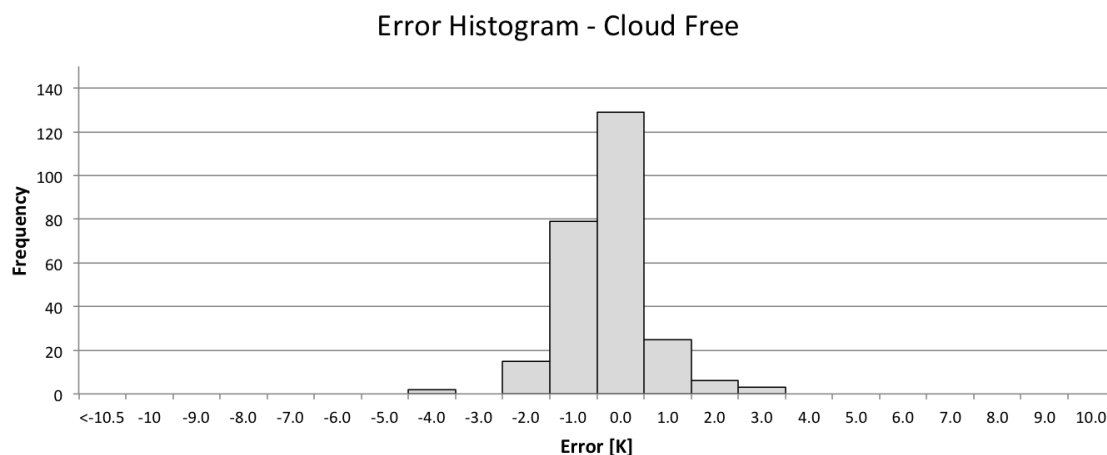
Figure 5.22: Histogram of error values for cloud free scenes for all ground truth sites.

5.3 Development of a Confidence Metric

An important aspect of any product is the generation of an error analysis, quality assurance band, or confidence metric. Because of the fusion of multiple data sources, use of radiative transfer and reanalysis code, and multiple interpolations, a traditional error analysis is difficult to implement. The goal of the error analysis is to provide the user with a metric, qualitative or quantitative, that describes how accurate the final predicted ground temperature is expected to be based on our atmospheric compensation. We want to be able to predict for users when they can have high confidence in high quality data, while still providing a best estimate solution for all pixels, even in lower confidence conditions with higher expected errors. Because this work does not include emissivity estimation, this metric only considers the error contributed from the atmospheric compensation and

not the emissivity component (which is being developed by JPL for inclusion in the final product).

A more traditional error propagation was first developed, by propagating error in the input atmospheres through the process in order to estimate an error value for each predicted temperature. This was first done for cloud free scenes to assess accuracy in best case scenarios. Because it is impractical that this would work in the presence of clouds, and the final product will be produced for all scenes, not just cloud free scenes, an analysis in the presence of clouds was also implemented in order to more closely investigate the effect of clouds on error in the final retrieved temperature. With this information, we make an initial suggestion for the confidence metric to be provided to the user.

5.3.1 Error Propagation

Using only cloud free scenes, the goal was to develop a process that propagated error from the input atmospheric profiles through the process to estimate the magnitude of the error in the final predicted temperature. Based on Hook et al. (2007), atmospheres were perturbed by the predicted atmospheric uncertainty and simulations were conducted to determine contributions to error in the final LST.

The error in the radiance due to temperature, defined by the governing equation expressed in Equation 3.1, can be written as shown in Equation 5.2.

$$
\begin{aligned}
S_{L_T} = \Bigg[&\left(\frac{\partial L_T}{\partial \tau} S_\tau \right)^2 \\
&+ \left(\frac{\partial L_T}{\partial L_u} S_{L_u} \right)^2 \\
&+ \left(\frac{\partial L_T}{\partial L_d} S_{L_d} \right)^2 \\
&+ \left(\frac{\partial L_T}{\partial L_{obs}} S_{L_{obs}} \right)^2 \\
&+ \left(\frac{\partial L_T}{\partial \epsilon} S_\epsilon \right)^2 \\
&+ 2\rho_{\tau L_u} \frac{\partial L_T}{\partial \tau} \frac{\partial L_T}{\partial L_u} S_\tau S_{L_u} \\
&+ 2\rho_{\tau L_d} \frac{\partial L_T}{\partial \tau} \frac{\partial L_T}{\partial L_d} S_\tau S_{L_d} \\
&+ 2\rho_{L_u L_d} \frac{\partial L_T}{\partial L_u} \frac{\partial L_T}{\partial L_d} S_{L_u} S_{L_d} \Bigg]^{\frac{1}{2}}
\end{aligned}
\tag{5.2}
$$

Although it would be important in the final product, we were initially interested in atmospheric errors and neglected the contributions of the observed radiance and emissivity. The partials in Equation 5.2, derived from Equation 3.1, are shown in Equations 5.3, 5.4, and 5.5.

$$\frac{\partial L_T}{\partial \tau} = \frac{L_u - L_{obs}}{\epsilon \tau^2} \tag{5.3}$$

$$\frac{\partial L_T}{\partial L_u} = \frac{-1}{\epsilon \tau} \tag{5.4}$$

$$\frac{\partial L_T}{\partial L_d} = \frac{\epsilon - 1}{\epsilon} \tag{5.5}$$

The error values are represented in Equations 5.6, 5.7, and 5.8.

$$S_\tau = \left[\left(\frac{\partial \tau}{\partial T} S_T \right)^2 + \left(\frac{\partial \tau}{\partial RH} S_{RH} \right)^2 + \left(\frac{\partial \tau}{\partial p} S_p \right)^2 + \left(\frac{\partial \tau}{\partial H} S_H \right)^2 \right]^{\frac{1}{2}} \tag{5.6}$$

Similar to effects from the observed radiance and emissivity, the pressure and height of the atmospheric profiles will contribute some error to the process, but we expect the largest error contributors to be the temperature and relative humidity and so first investigated only these.

$$S_{L_u} = \left[\left(\frac{\partial L_u}{\partial T} S_T \right)^2 + \left(\frac{\partial L_u}{\partial RH} S_{RH} \right)^2 \right]^{\frac{1}{2}} \tag{5.7}$$

$$S_{L_d} = \left[\left(\frac{\partial L_d}{\partial T} S_T \right)^2 + \left(\frac{\partial L_d}{\partial RH} S_{RH} \right)^2 \right]^{\frac{1}{2}} \tag{5.8}$$

The partial derivatives in Equations 5.6, 5.7, and 5.8 were determined using numerical simulations. For each atmosphere, the temperature profile and relative humidity profile were modified, in separate simulations, and the corresponding transmission, upwelled radiance, and downwelled radiance values were determined. The temperature profiles were both increased and decreased 5 K in steps of 1 K. The relative humidity profiles were increased and decreased 30% in steps of 5%. The relationship between the change in the profile and the radiative transfer parameters were acceptably linear ($R^2 > 0.95$), so the slope was calculated as the partial derivative for each atmosphere. However, this partial derivative value differed for the variety of scenes that were tested. Therefore, these partials were defined as functions of surface air temperature and column water vapor. As an initial test, linear relationships between each partial and atmospheric variable were defined. One example is shown in Figure 5.23. The correlations between transmission, upwelled radiance, and downwelled radiance ($\rho_{\tau L_u}$, $\rho_{\tau L_d}$, and $\rho_{L_u L_d}$) was also calculated from the simulations for each atmosphere. These were all large and similar (> 0.99), so an average value was used.
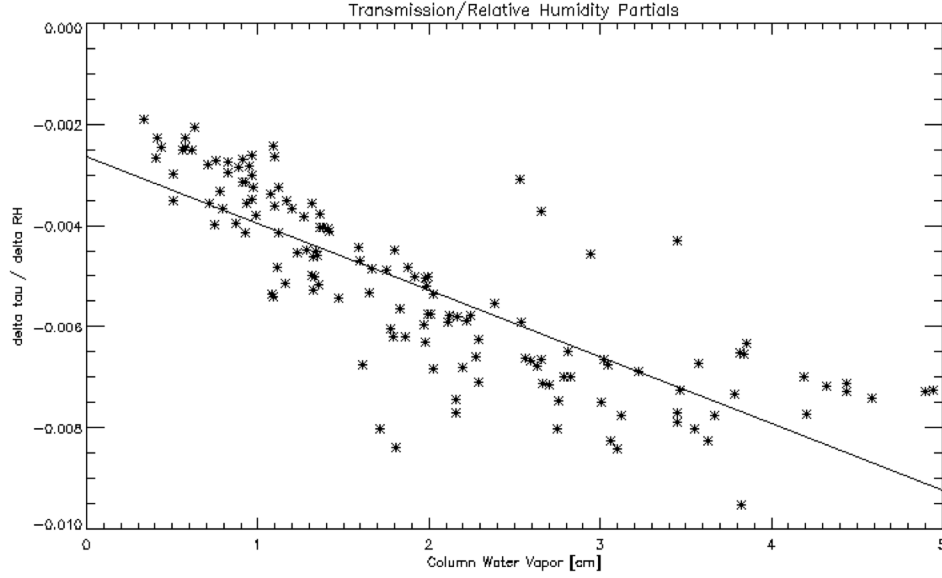
Figure 5.23: Relationship between transmission and relative humidity partial derivative and column water vapor.

Perhaps the most difficult values to accurately estimate are S_T and S_{RH}. These should be inherent to atmospheric profile data, in this case from the reanalysis product, but are often difficult to find or determine. As an initial investigation, we considered results for $S_T = 0.75$ K and $S_{RH} = 2\%$, based on the MODIS atmospheric profile retrievals [Seemann et al., 2006]. With these values and results, we can calculate an estimate for the error in the radiance due to temperature from atmospheric effects. Errors are mean centered since they are assumed to be random, so we use the average magnitude of the positive and negative errors to predict the error in LST. This prediction is only a magnitude and has no associated sign. This is summarized in Equation 5.9.

$$
\begin{aligned}
L_T &\to T \\
L_T - S_{L_T} &\to T_1 \\
L_T + S_{L_T} &\to T_2 \\
LST_{error} &= 0.5[(T_2 - T) + (T - T_1)]
\end{aligned}
\tag{5.9}
$$

In summary, the partial derivatives of Equation 3.1 are derived, the error in radiative transfer parameters are determined using numerical simulation and estimates of uncertainty in atmospheric variables, and the correlation coefficients are calculated using numerical simulations. Initial results for the error propagation are shown in Figure 5.24. This is a comparison of predicted errors generated using the error propagation described above and actual errors, calculated from Equation 5.1 and shown in Figure 5.22.
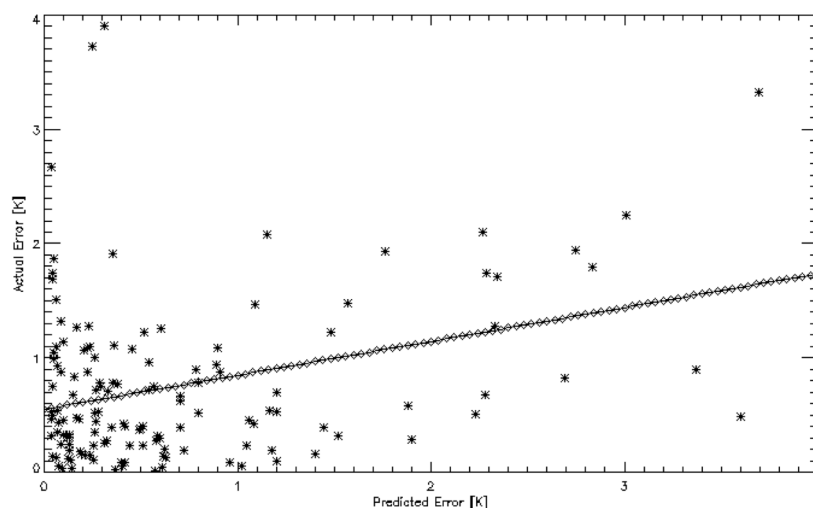
Figure 5.24: Comparison of actual to predicted error by perturbing atmosphere and using traditional error analysis.

The R^2 value for the fit shown in Figure 5.24 is 0.139. However, even though the fit is very poor, the RMS error is still only 0.677 K. This is because the errors are small to begin with. Although there is little correlation, the magnitude of almost all the errors in the predictions is small. In Figure 5.24, for anything below a one-to-one line, we would overestimate the error. There are still a large number of points where we overestimate the error. We think this is because the estimated input errors ($S_T = 0.75K$ and $S_{RH} = 2\%$) are too large for very well characterized atmospheres. For any points above a one-to-one line, we underestimate the error. As shown, there are still points where we underestimate the error by more than 2 K. We believe this is because if the water vapor in the atmosphere is not accurately captured in the original atmospheric profile, the error cannot be properly propagated through the process, no matter how accurate the estimated input error. For most cloud free points, as expected, both the predicated and actual errors are less than 1 K, although the relationship has little correlation, as shown in the plot. Therefore, this is hardly better than a qualitative assessment that cloud free scenes generally have small actual errors.

5.3.2   Cloud Analysis

Because we want to be able to both process and produce a quality metric for every pixel in every Landsat scene in the archive, it was important, after validating our methodology using cloud free data, to consider results and confidence metric development for a larger variety of scenes. It is important to consider how poor cloud classification or clouds in the surround might impact the accuracy of the LST retrievals, even at pixels originally classified as cloud free. Therefore, we explored how each pixel is affected by different types and amounts of clouds in the scene. Using the same ground truth sites discussed in Section 5.1, all scenes (available and with acquirable ground truth) for the years 2006 through 2011 were processed (except for the Salton Sea and Lake Tahoe, where select scenes were provided by JPL). This resulted in a complete validation

dataset containing 827 scenes. This section details the cloud categorization and the analysis for each individual location. The break down of mean and standard deviation by cloud category helps to illustrate how removing different types of clouds improves results and how removal of cloudy pixels, or clouds in the vicinity, can be used to build our confidence metric.

Cloud Categorization

The cloud free scenes analyzed in Section 5.2 were visually selected as having no clouds over the buoy, but also no clouds in the vicinity of the buoy. If there was any question of cloud influence, the scene was discarded. While this guarantees best results, we believe there are a large number of useable pixels, where the LST can still be reasonably accurately retrieved, with clouds in the vicinity. Therefore, we initially want to categorize pixels as cloudy, clouds in the vicinity, or cloud free. We also think that the type of cloud is important. Clouds are generally classified by both texture and height. For height classifications, the prefix cirro- means high and the prefix alto- means mid. For texture classifications, the prefix strato- means layer, meaning uniform or widespread, and the prefix cumulo- means heap, meaning cellular or individual elements. From the view of the satellite, we are more concerned with texture and less concerned with height. Therefore, we create two categories: cumulus clouds, which includes cirrocumulus, altocumulus, and other similar types, and stratus clouds, which includes cirrostratus, altostratus, nimbostratus, and other similar types. Finally, cirrus clouds are wispy, feathery clouds that are generally thinner. Because they do not have well defined edges, and are difficult to visually distinguish for an expert lacking expertise, we categorize them with stratus clouds. For each image in the validation dataset, a visual analysis of the buoy and its surrounding area was performed in order to classify the scene into one of six categories as shown in Table 5.2. The breakdown of the number of scenes in each category from the 827 image validation dataset is also shown in this table.

Table 5.2: Categories used in cloud analysis and breakdown of number and percentage of scenes in each category.

Category	Description	Number of Scenes	% of Scenes
0	Cloud Free	259	31.3 %
1	Cumulus in Vicinity of Buoy	98	11.9 %
2	Status 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 %

Note that cloud type and in the vicinity were subjective based on visual analysis but all scenes were categorized by the same analyst. An example of each is shown below. As with the images shown in Section 5.1, the black square in the Landsat scene represents the approximate location of the buoy. Figure 5.25 shows an image in category 0, which is cloud free, and Figure 5.26 shows an

image in category 5, a completely cloud covered image.
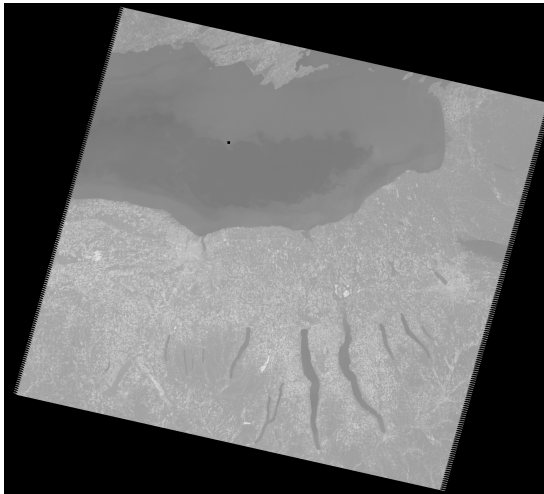
Figure 5.25: An example of a cloud free image (category 0). The black square indicates the approximate buoy location.
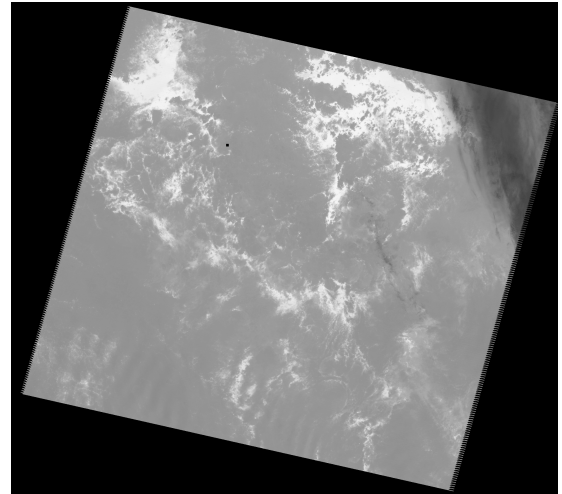
Figure 5.26: An example of a completely cloud covered image (category 5). The black square indicates the approximate buoy location.

Figure 5.27 shows an image with cumulus clouds in the vicinity of the buoy; in order to provide a better visual, a subset of the image showing only the surround of the buoy is shown in Figure 5.28. Similarly, Figure 5.29 shows an image with stratus or cirrus clouds in the vicinity, and Figure 5.30 shows a subset of the scene over the buoy location. Note that the size of the black square is not representative of the drift or possible watch radius of the buoy and the categorization of in the vicinity is based on the subjective opinion of the analyst of the possible influence of those clouds on the atmospheric compensation results.
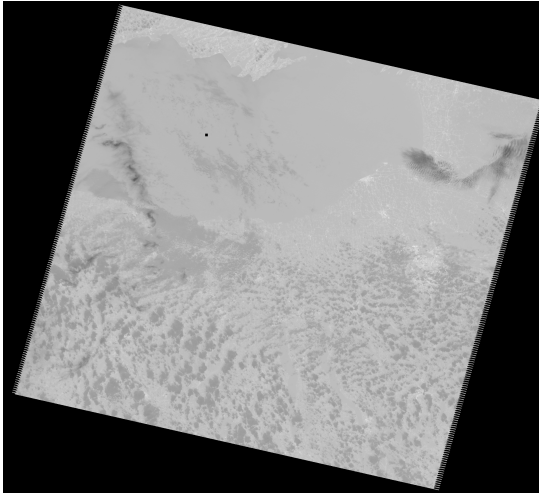
Figure 5.27: A Landsat scene with cumulus clouds in the vicinity of the buoy, represented by the black square.
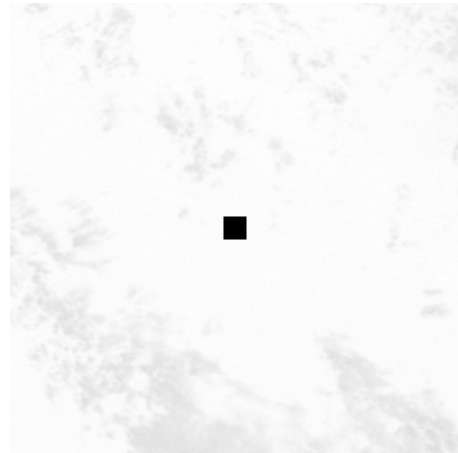
Figure 5.28: A subset of Figure 5.27, showing cumulus clouds in the vicinity of the buoy, represented by the black square (not to scale).
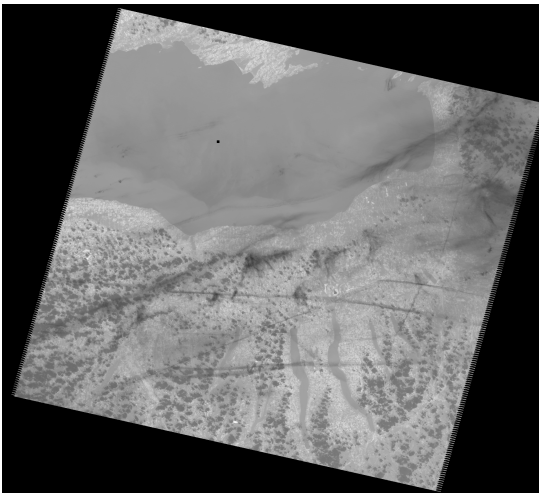
Figure 5.29: A Landsat scene with stratus or cirrus clouds in the vicinity of the buoy, represented by the black square.
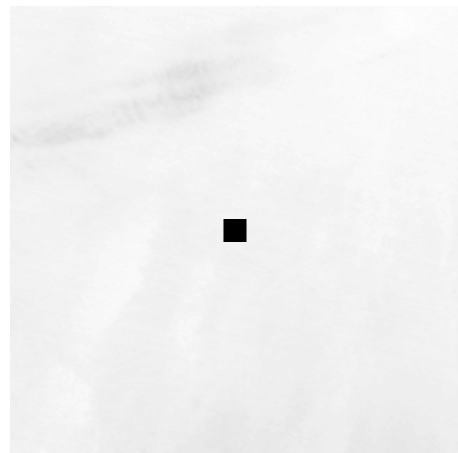
Figure 5.30: A subset of Figure 5.29, showing stratus or cirrus clouds in the vicinity of the buoy, represented by the black square (not to scale).

Figures 5.31 and 5.32 show a scene with cumulus clouds over the buoy and the subset of the scene surrounding the buoy. And similarly, Figures 5.33 and 5.34 show a scene with stratus clouds over the buoy and the subset of the scene containing the buoy.
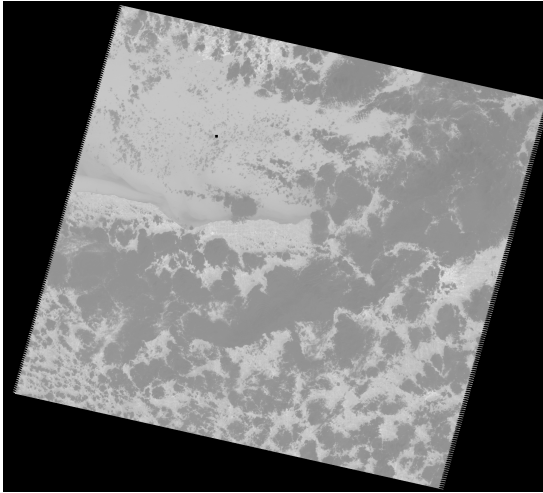
Figure 5.31: A Landsat scene with cumulus clouds over the buoy, represented by the black square.
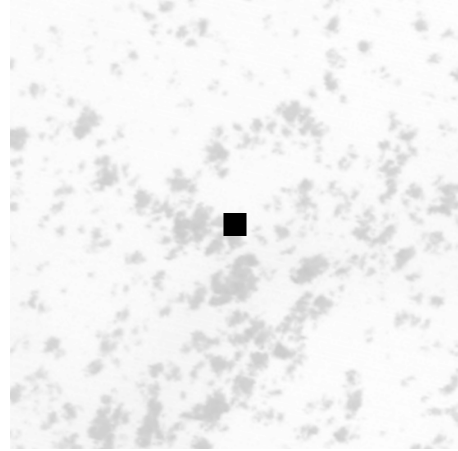
Figure 5.32: A subset of Figure 5.31, showing cumulus clouds over the buoy, represented by the black square (not to scale).
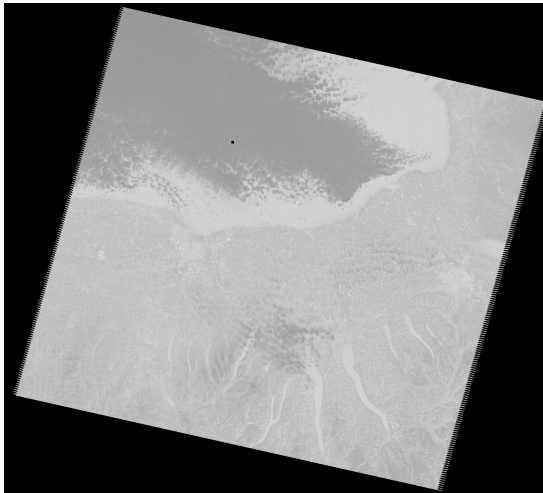
Figure 5.33: A Landsat scene with stratus clouds over the buoy, represented by the black square.
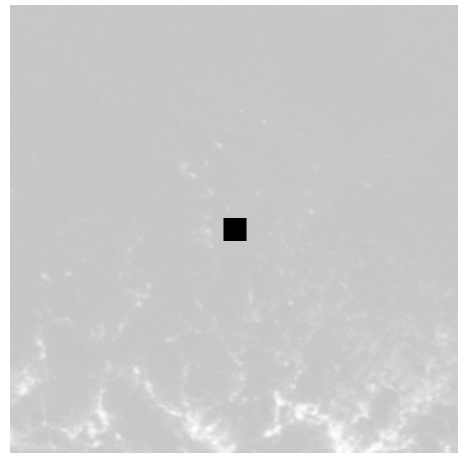
Figure 5.34: A subset of Figure 5.33, showing stratus clouds over the buoy, represented by the black square (not to scale).

Again, in all figures, the size of the black square indicates neither the size of the buoy nor the watch radius, and the difference between over the buoy versus in the vicinity, and similarly the difference between in the vicinity and cloud free, is not well defined but rather subjective to the visual analysis. While the was a productive way to analyze and understand the dataset, an automation of this process would need to better define these characteristics and categorizations. This is further discussed in Section 5.3.3.

The Salton Sea

Figures 5.35, 5.36 and 5.37 show histograms of error values for all scenes over the Salton Sea, scenes with only clouds in the vicinity or cloud free (0, 1, or 2) over the Salton Sea, and only cloud free scenes over the Salton Sea. These error values are calculated using Equation 5.1, just like those discussed in Section 5.2. Note that Figure 5.37 is the same as Figure 5.13. However, this histogram is repeated here for comparison to results with clouds included. As cloud restrictions are applied, these histograms become closer to zero-centered. The mean and standard deviation for results as cloud restrictions are applied, one category at a time, are shown in Table 5.3. This shows a decreasing mean and standard deviation as scenes with cloud contamination are removed. There are considerably fewer scenes, and fewer cloudy scenes, over the Salton Sea, but the same pattern of improvement is still shown. While it is obvious that cloudy pixels will have poor results and somehow need to be flagged accordingly, this analysis is also performed to understand how clouds in the vicinity affect our results and if the type of cloud in the vicinity (stratus or cumulus) matters. Both of these issues will be important in the development of a confidence metric, and conclusions about this will become more obvious with more scenes.

Figure 5.35: Histogram of error values for all scenes over the Salton Sea. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Figure 5.36: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over the Salton Sea. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Figure 5.37: Histogram of error values for cloud free scenes over the Salton Sea. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.3: Summary of results of means and standard deviations of errors for different cloud categorizations over the Salton Sea.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-0.392 K	1.092 K	14 [100%]
0,1,2,3	-0.392 K	1.092 K	14 [100%]
0,1,2	-0.136 K	0.548 K	13 [93%]
0,1	-0.136 K	0.548 K	13 [93%]
0	-0.120 K	0.558 K	10 [91%]

Lake Tahoe

Histograms for Lake Tahoe are shown in Figures 5.38, 5.39 and 5.40 for all scenes, clouds in the vicinity or cloud free, and only cloud free scenes respectively. The mean and standard deviation of errors for each distribution of cloud categorizations is shown in Table 5.4. While there are more

scenes over Lake Tahoe than Salton Sea, this location was also prescreened for good conditions, and therefore there is a considerably smaller fraction of scenes with clouds to consider. However, the mean and standard deviation still decrease as clouds are removed.
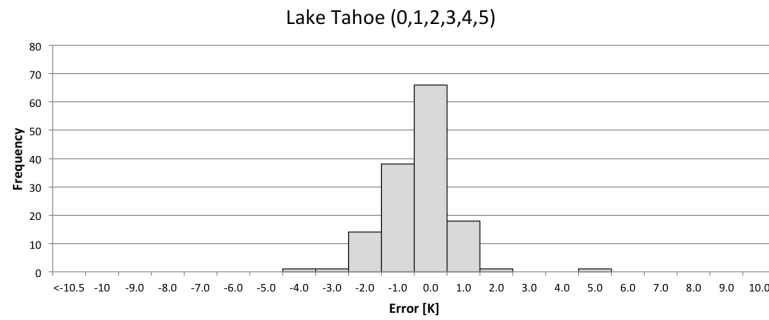
Figure 5.38: Histogram of error values for all scenes over Lake Tahoe. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
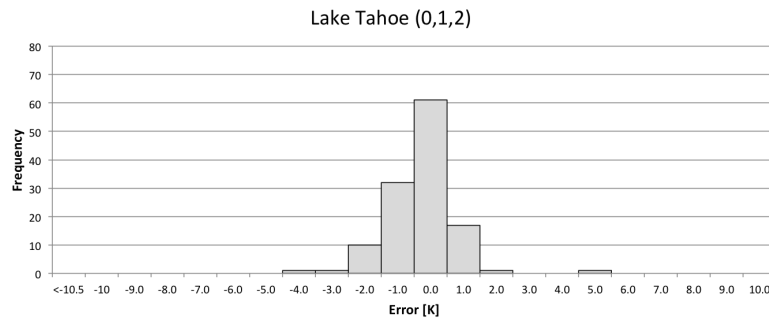
Figure 5.39: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Lake Tahoe. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
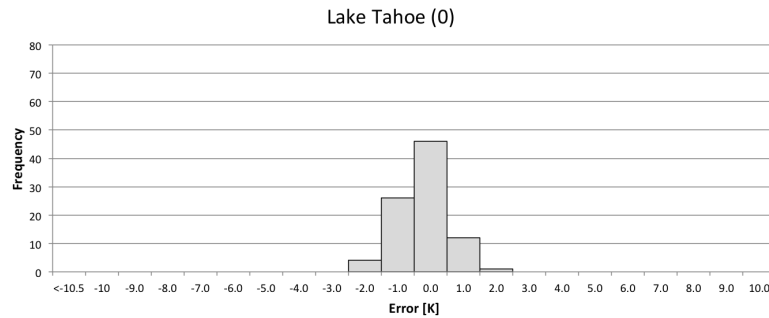
Figure 5.40: Histogram of error values for cloud free scenes over Lake Tahoe. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.4: Summary of results of means and standard deviations of errors for different cloud categorizations over Lake Tahoe.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-0.342 K	1.030 K	140 [100%]
0,1,2,3	-0.303 K	1.004 K	137 [98%]
0,1,2	-0.279 K	1.026 K	124 [89%]
0,1	-0.205 K	0.698 K	102 [73%]
0	-0.213 K	0.713 K	89 [64%]

Rochester

Figures 5.41, 5.42 and 5.43 show histograms of error results for scenes over Rochester, New York for all scenes, clouds in the vicinity or cloud free, and only cloud free scenes respectively. With a smaller fraction of the total number of scenes in this location being cloud free, the effect of removing cloudy scenes, or scenes with clouds in the vicinity of the buoy, is more easily observed. Note that from Figure 5.41 to Figure 5.42, the large, left-most bin is almost entirely eliminated, but there is a long left-hand tail when clouds in the vicinity remain in the histogram. In the Figure 5.43, the left hand tail is gone, but the number of scenes is also much lower, even in the center three bins, indicating that removing clouds in the vicinity removes both desired negative errors and undesired accurate scenes. The mean and standard deviations of each category are summarized in Table 5.5. This location illustrated a more drastic change when clouds over the buoy were removed; there is still a left hand tail in the histogram with clouds in the vicinity included, but the mean and standard deviation is much smaller, and fairly reasonable, depending on the application at which these results are aimed. This is important when considering how these types of pixels should be handled during the development of a confidence metric.
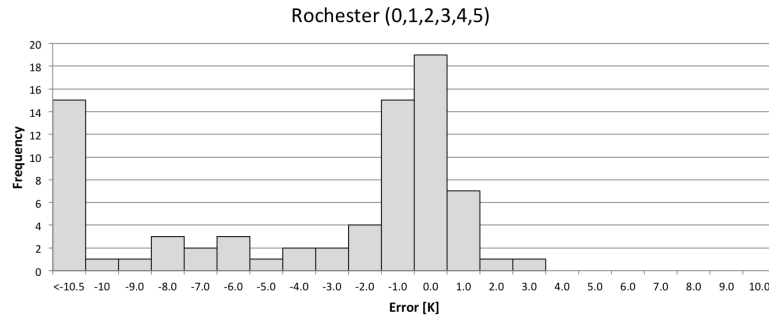
Figure 5.41: Histogram of error values for all scenes over Rochester. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
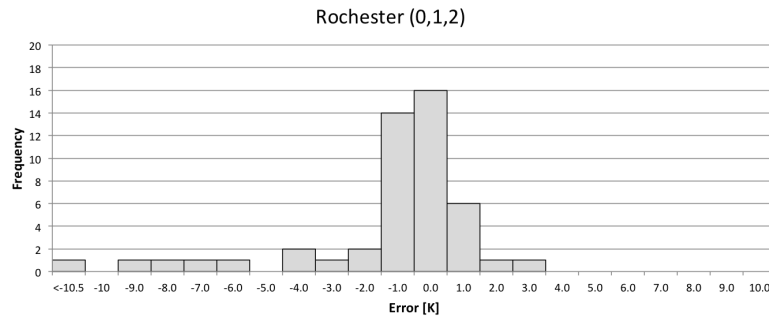
Figure 5.42: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Rochester. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
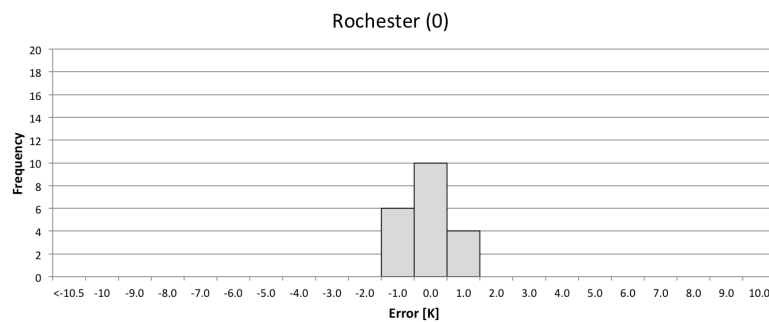
Figure 5.43: Histogram of error values for cloud free scenes over Rochester. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.5: Summary of results of means and standard deviations of errors for different cloud categorizations over Rochester.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-10.090 K	22.676 K	77 [100%]
0,1,2,3	-1.767 K	5.317 K	53 [69%]
0,1,2	-1.251 K	3.028 K	48 [62%]
0,1	-0.453 K	1.758 K	29 [38%]
0	-0.068 K	0.639 K	20 [26%]

Delaware Bay (Delmar)

Histograms for all scenes of Delmar, scenes with clouds in the vicinity or cloud free, and only cloud free scenes are shown in Figures 5.44, 5.45, and 5.46. Similar to Rochester, the largest negative errors are eliminated when scenes with clouds over the buoy are removed, but there still remains a negative tail when clouds in the vicinity are included. The largest of these negative errors are removed when scenes with clouds in the vicinity are removed, but the number of scenes is also greatly decreased. A summary of the errors is shown in Table 5.6. Similar to Rochester, the mean errors have magnitudes of less than 2 K even when clouds in the vicinity are included, although we know there is still a negative tail on the histogram.
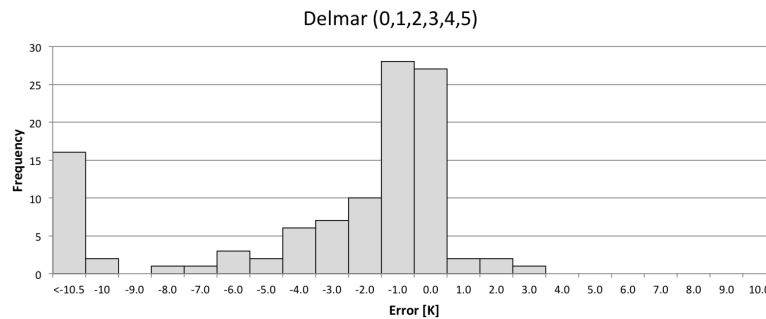
Figure 5.44: Histogram of error values for all scenes over Delmar. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
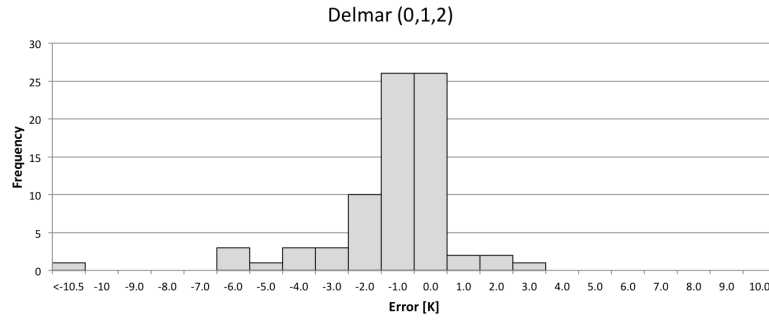
Figure 5.45: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Delmar. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
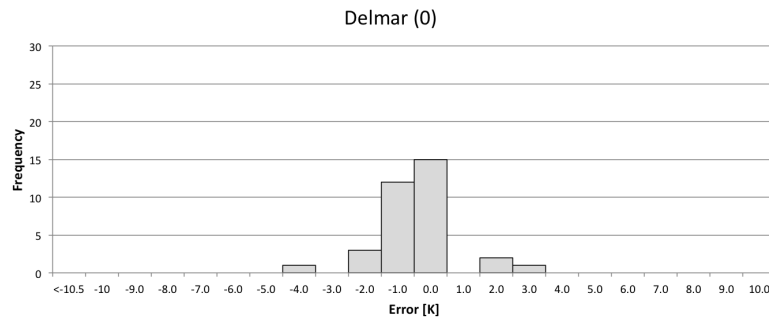
Figure 5.46: Histogram of error values for cloud free scenes over Delmar. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.6: Summary of results of means and standard deviations of errors for different cloud categorizations over Delmar.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-5.167 K	10.867 K	108 [100%]
0,1,2,3	-1.932 K	4.431 K	85 [79%]
0,1,2	-1.099 K	1.903 K	78 [72%]
0,1	-0.696 K	1.841 K	51 [47%]
0	-0.448 K	1.179 K	34 [31%]

Georgia Coast

Histograms for the scenes over the Georgia coast are shown in Figures 5.47, 5.48, and 5.49; these show error values for all scenes, errors for scenes categorized as 0, 1, or, 2, and only cloud free scenes respectively. Similar to other locations, the number of scenes is greatly decreased once

all scenes with possible cloud contamination are removed, which motivates us to investigate the usefulness of scenes with clouds in the vicinity. For this location, the mean error when clouds in the vicinity are included in still less that 1 K. This location is uniquely the only location with a positive mean error for cloud free scenes, and tends to have slightly larger standard deviations than other locations. This is not surprising considering the type of climate in Georgia, compared to other locations considered, which adds variability to our validation dataset. The mean errors and standard deviations of the cloud categorizations is summarized in Table 5.7.
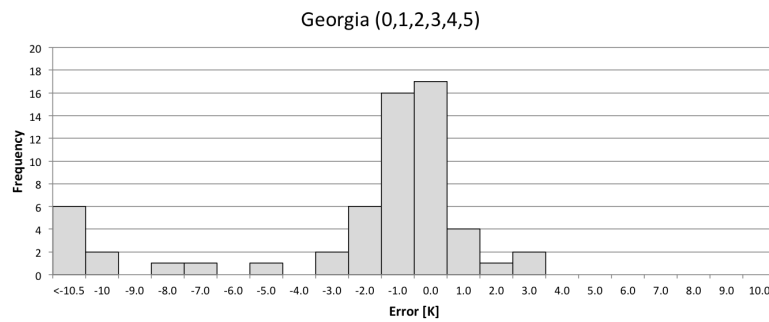
Figure 5.47: Histogram of error values for all scenes over Georgia. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
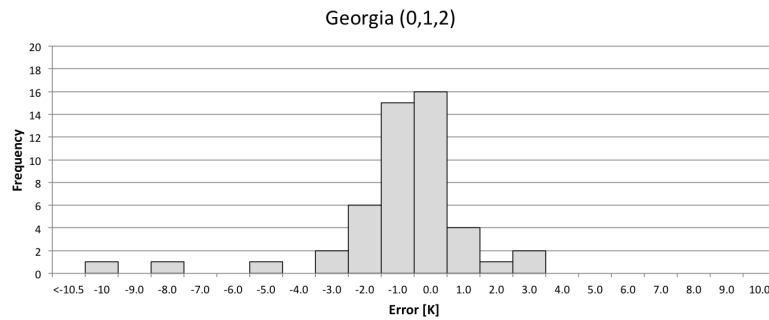
Figure 5.48: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Georgia. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
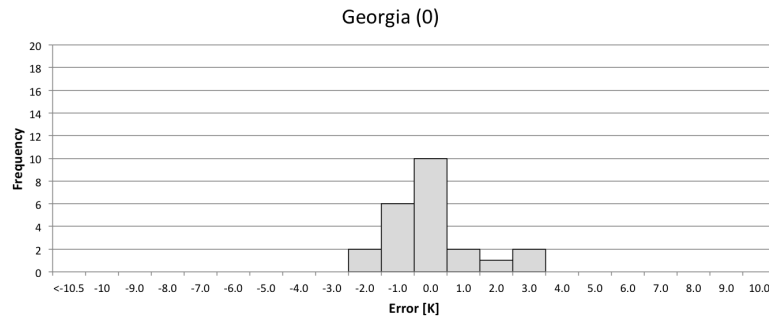
Figure 5.49: Histogram of error values for cloud free scenes over Georgia. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.7: Summary of results of means and standard deviations of errors for different cloud categorizations over Georgia.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-2.719 K	5.590 K	59 [100%]
0,1,2,3	-1.303 K	3.192 K	52 [88%]
0,1,2	-0.854 K	2.183 K	49 [83%]
0,1	-0.173 K	1.246 K	31 [53%]
0	-0.041 K	1.267 K	23 [39%]

California (Santa Maria)

Results for the cloud categorization for scenes off the coast of California near Santa Maria are shown in Figures 5.50, 5.51, and 5.52, showing all scenes, only scenes in cloud categories 0, 1, and 2, and only cloud free scenes respectively. Note that from including scenes with clouds in the vicinity to only cloud free scenes, the number of scenes in the center three bins (errors from -1.5 K to 1.5 K) is reduced by half. While removing clouds in the vicinity also removes a handful of scenes with moderate errors on the left and right tail of the histogram, this reinforces our thought that eliminating clouds in the vicinity removes a large number of good data points. A summary of the mean errors and standard deviations is shown in Table 5.8. For this location, the mean error when clouds in the vicinity are included in still less than 1 K in magnitude.
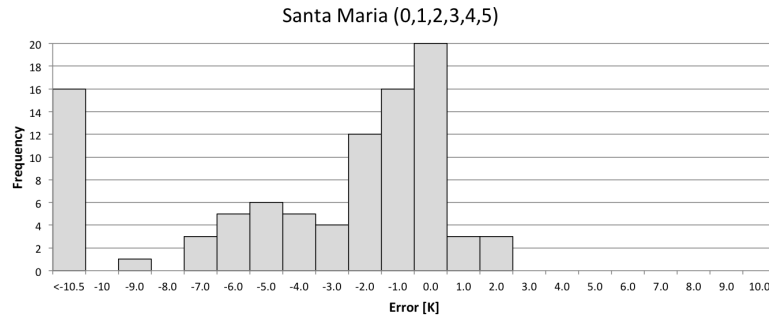
Figure 5.50: Histogram of error values for all scenes near Santa Maria. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
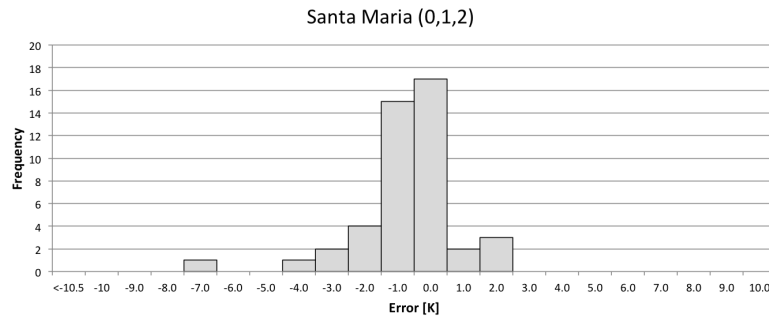
Figure 5.51: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes near Santa Maria. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Figure 5.52: Histogram of error values for cloud free scenes near Santa Maria. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.8: Summary of results of means and standard deviations of errors for different cloud categorizations near Santa Maria.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-5.232 K	9.636 K	94 [100%]
0,1,2,3	-1.170 K	2.872 K	49 [52%]
0,1,2	-0.639 K	1.509 K	45 [48%]
0,1	-0.175 K	1.042 K	33 [35%]
0	-0.219 K	0.789 K	19 [20%]

California (Santa Monica)

Figure 5.53 is a histogram for all scenes off the coast of California near Santa Monica, Figure 5.54 shows scenes at this locations with clouds in the vicinity or cloud free, and Figure 5.55 shows only errors for cloud free scenes at this location. Not surprisingly, this location behaves similarly to scenes off the coast of California near Santa Maria, although a larger portion of scenes in Figure 5.53 seem to be part of the negative tail with moderately large negative errors. Most of these scenes are removed when scenes with clouds over the buoy are removed, leaving a much smaller negative tail when only scenes with clouds in the vicinity and cloud free scenes are included. This is illustrated in Table 5.9; also note that the number of scenes included in only cloud free scenes (0) is less than half of the number of scenes included with clouds in the vicinity (0,1,2).

Figure 5.53: Histogram of error values for all scenes near Santa Monica. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Figure 5.54: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes near Santa Monica. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Figure 5.55: Histogram of error values for cloud free scenes near Santa Monica. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.9: Summary of results of means and standard deviations of errors for different cloud categorizations near Santa Monica.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-8.066 K	18.048 K	113 [100%]
0,1,2,3	-1.006 K	2.733 K	60 [53%]
0,1,2	-0.883 K	1.538 K	49 [43%]
0,1	-0.576 K	0.942 K	36 [32%]
0	-0.574 K	1.089 K	21 [19%]

Lake Huron

Finally, we consider scenes in the Great Lakes region, first with scenes over Lake Huron, with all scenes shown in Figure 5.56, scenes with clouds in the vicinity or no clouds shown in Figure 5.57, and only cloud free scenes shown in Figure 5.58. Also illustrated by the drastic change in mean

error shown in Table 5.10, note that the largest bin in Figure 5.56 is the leftmost bin, illustrating the number of cloudy scenes at this location. Even when clouds over the buoy are removed, a large negative outlier remains, and the negative tail is more pronounced when clouds in the vicinity are included as shown in Figure 5.57. This is reflected in the location statistics as this is the only location where the mean error when including clouds in the vicinity has a magnitude greater than 2 K. As with other locations, the number of scenes is greatly reduced when only cloud free scenes are included.
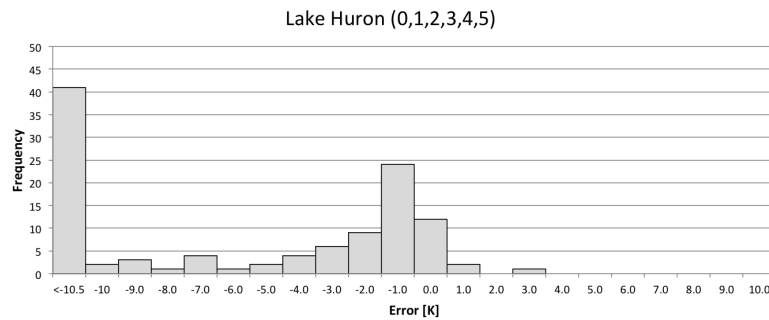
Figure 5.56: Histogram of error values for all scenes over Lake Huron. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
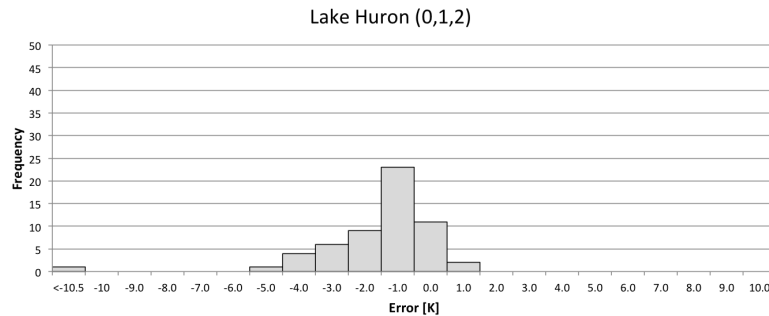
Figure 5.57: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Lake Huron. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
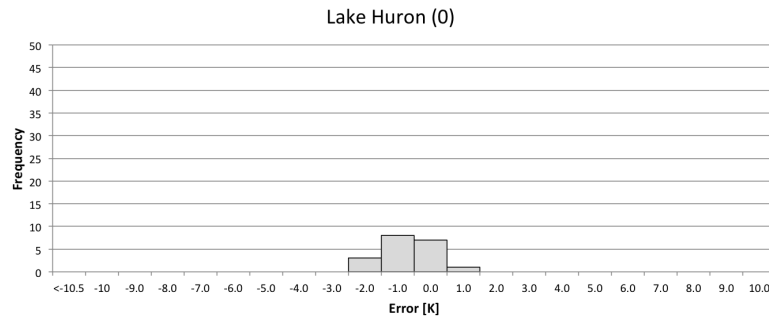
Figure 5.58: Histogram of error values for cloud free scenes over Lake Huron. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.10: Summary of results of means and standard deviations of errors for different cloud categorizations over Lake Huron.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-19.913 K	29.506 K	112 [100%]
0,1,2,3	-3.166 K	5.837 K	64 [57%]
0,1,2	-2.182 K	4.930 K	57 [51%]
0,1	-1.886 K	6.034 K	33 [29%]
0	-0.695 K	0.820 K	19 [17%]

Lake Superior

Histograms for scenes over Lake Superior are shown in Figures 5.59, 5.60, and 5.61. These show all scenes, scenes with clouds in the vicinity or cloud free, and only cloud free scenes respectively. A summary of mean errors and standard deviations is shown in Table 5.11. As with scenes over Lake Huron, the left most bin is by far the largest in Figure 5.59, and while the histogram of cloud free scenes is very well behaved, the number of scenes is also greatly reduced. The mean error when including both scenes with clouds in the vicinity and cloud free scenes is still less than 2 K in magnitude.
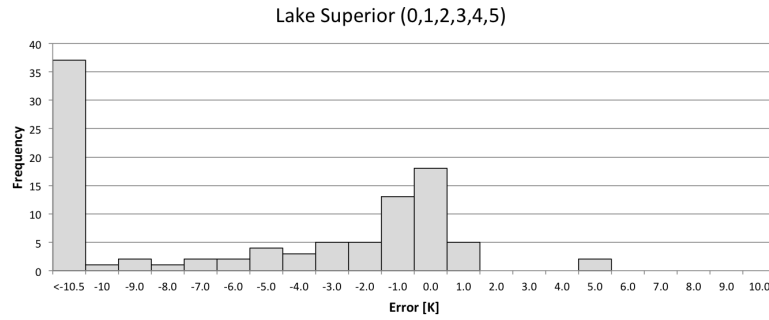
Figure 5.59: Histogram of error values for all scenes over Lake Superior. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
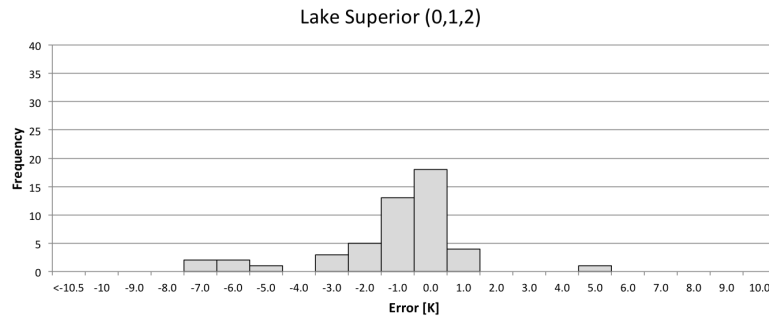
Figure 5.60: Histogram of error values for scenes with clouds in the vicinity and cloud free scenes over Lake Superior. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
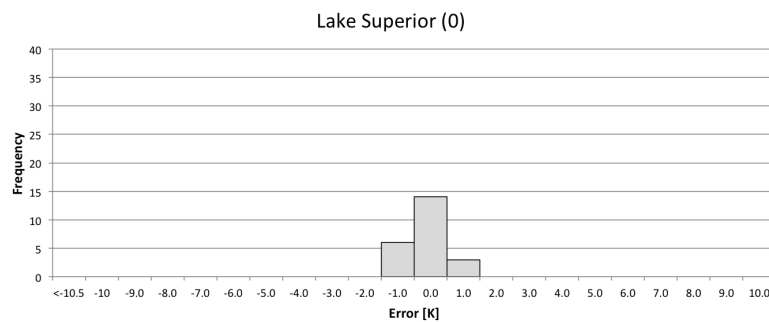
Figure 5.61: Histogram of error values for cloud free scenes over Lake Superior. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.11: Summary of results of means and standard deviations of errors for different cloud categorizations over Lake Superior.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-16.244 K	26.096 K	100 [100%]
0,1,2,3	-3.137 K	6.662 K	58 [58%]
0,1,2	-1.099 K	2.159 K	49 [49%]
0,1	-0.333 K	0.897 K	27 [27%]
0	-0.167 K	0.676 K	23 [23%]

Summary of Cloud Analysis

A similar analysis including all scenes at all locations is shown below. Figure 5.62 is a histogram including all scenes in the validation dataset, Figure 5.63 includes cloud free and scenes with clouds in the vicinity of the buoy for all locations, and Figure 5.64 is all cloud free scenes from the dataset. In the same pattern as illustrated at most individual locations, removing scenes with clouds over the buoy removes the largest of the errors. Removing scenes with clouds in the vicinity of the buoy leaves a very well behaved histogram. However, as shown in Table 5.12, the number of scenes included in the cloud free histogram is half the number of scenes included when clouds in the vicinity are included, while the mean error when clouds in the vicinity are included is still less than 1 K in magnitude. A larger standard deviation when including clouds in the vicinity does need to be taken into consideration. However, when providing data to the user, we believe that these data points with clouds in the vicinity may be useful to a number of users with the appropriate confidence considerations.
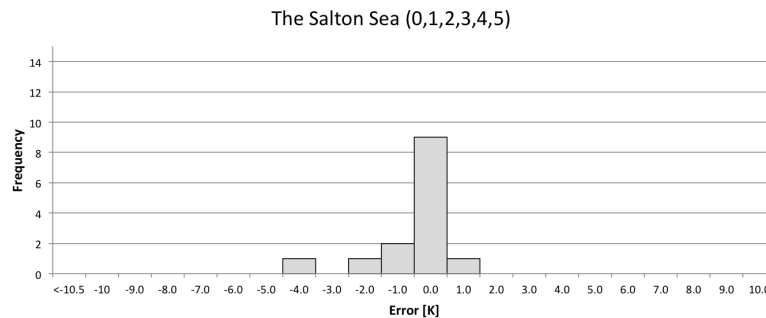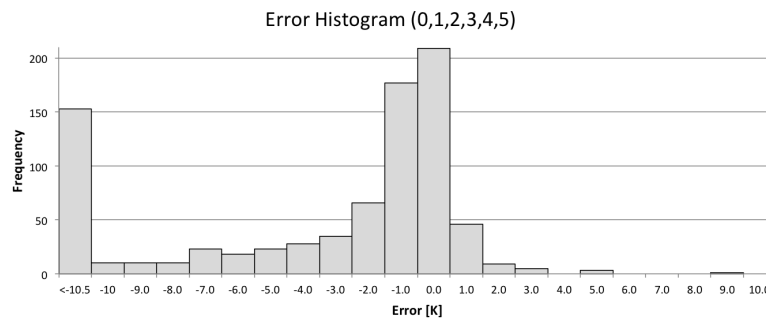

Figure 5.62: Histogram of error values for all scenes. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
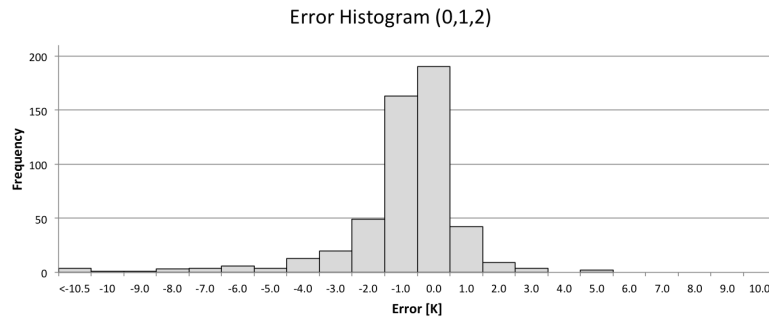
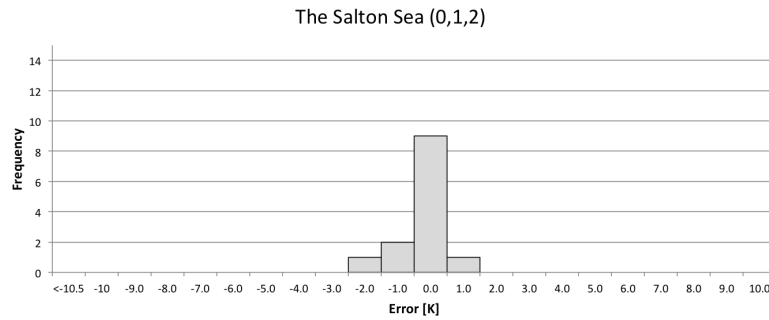**Figure 5.63:** Histogram of error values for scenes with clouds in the vicinity and cloud free scenes. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
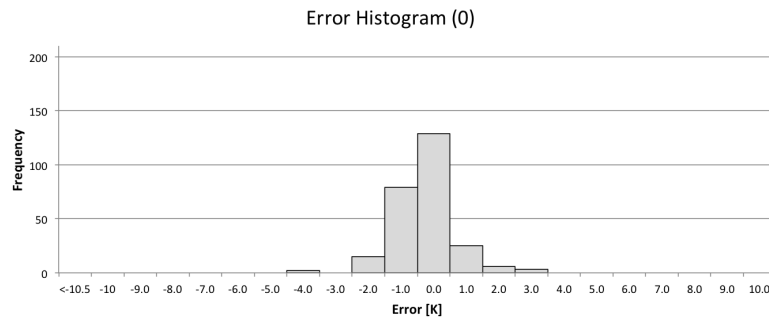
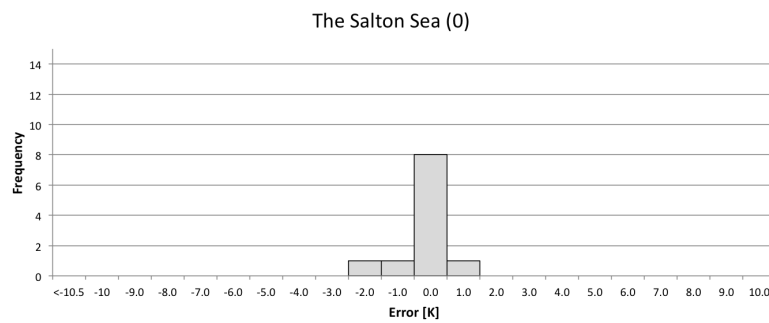**Figure 5.64:** Histogram of error values for cloud free scenes. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

Table 5.12: Summary of results of means and standard deviations of errors for different cloud categorizations for all locations.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-8.471 K	19.313 K	826 [100%]
0,1,2,3	-1.538 K	4.174 K	575 [70%]
0,1,2	-0.933 K	2.460 K	515 [62%]
0,1	-0.499 K	2.228 K	357 [43%]
0	-0.267 K	0.900 K	259 [31%]

5.3.3 Confidence Metric Expectations

As shown in Section 5.3.1, a quantitative estimation of error unique to each pixel is currently difficult to achieve. The validation of our methodology in Section 5.2 gives us confidence in our process, and the cloud analysis in Section 5.3.2 indicates cloud contamination is a larger contributor of error that anything in our implementation. Our breakdown by both cloud type and cloud

proximity illustrates the effect of various types of cloud contamination in our results. As shown above, the mean error for the validation dataset when including cloud free scenes and clouds in the vicinity is still less than 1 K. Even though the standard deviation is larger for this group of scenes, we believe that this data will still be useful to many users. With the confidence and validation of our process, our current and best expectation for a confidence metric, given the validation dataset shown here, is to account for and report the effects of cloud contamination, based on the cloud categories presented above.

As shown in Section 5.2.10, our current best data (cloud free) has a mean error of -0.267 K. As demonstrated in Appendix A.3, statistical testing based on current calibration data shows this negative shift to be significant. Based on familiarity with the data and observing the effects of cloud contamination, we wanted to look more closely at the effect of column water vapor on our retrievals. Comparing the NARR column water vapor values to a MODIS total precipitable column water vapor product, we found that NARR consistently underestimated column water vapor compared to the scene derived product. This study is detailed completely in Appendix B. This gives us more confidence that the negative bias is a result of our developed methodology and motivates some form of corrective action. Knowing the effects of cloud contamination, the uncertainty in cloud categorization, and the underestimation of water vapor illustrated by the water vapor comparison study, we explored the option of adding water vapor to the NARR atmospheric profiles. We performed another study that added varying amounts of water vapor, both distributed throughout the profile and at specific heights. Although we hypothesized that adding water vapor may lead to smaller standard deviations by increasing our underestimations and decreasing our overestimations, adding water vapor, in all cases but one, increased the retrieved temperature. This study is detailed completely in Appendix C.

Because the results were inconclusive, and the change in temperature when adding column water vapor was in a single direction, we believe the best option with the currently available information, is a bias shift to zero center the cloud free data and reduce the magnitude of expected error in the cloud contaminated pixels. A larger bias to shift pixels with clouds in the vicinity (based on average observed errors for this category of pixels) was also considered. However, because the influence of clouds in the vicinity is highly variable (some scenes have a larger negative error while other clouds in the vicinity have little effect on results), the magnitude of the necessary bias was difficult to determine. While a bias shift of this data would be better on average (the data would shift to zero mean), it would increase the errors of many already acceptable results. Therefore, to account for the observed significant negative bias, the retrieved temperature at each pixel was increased by 0.267 K. Figure 5.65 shows a histogram of errors for the cloud free biased data and Table 5.13 shows the updated mean errors (and the same standard deviations included for completeness).
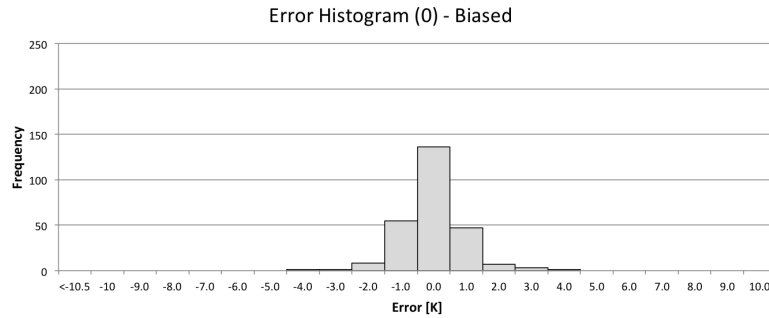
Figure 5.65: Histogram of error values for cloud free scenes after the bias shift of 0.267 K has been applied.

Table 5.13: Summary of results of means and standard deviations of errors for different cloud categorizations for all locations when the data has been biased by 0.267 K.

Cloud Category	Mean [K]	Standard Deviation [K]	Number of Scenes [% of Scenes]
0,1,2,3,4,5	-8.204 K	19.313 K	826 [100%]
0,1,2	-0.666 K	2.460 K	515 [62%]
0,1	-0.232 K	2.228 K	357 [43%]
0	0.0 K	0.900 K	259 [31%]

Although the cloud analysis documented above was subjective, revisiting results and estimating a distance threshold showed that the cloudy pixels (5,4,3) generally had clouds within 0.5 km of the buoy location (17 pixels). Pixels categorized as having clouds in the vicinity (2,1) had clouds more than 0.5 km but less than 5 km from the buoy location (between 17 pixels and 167 pixels), and cloud free pixels had no clouds within 5 km (167 pixels). Implementation of the confidence metric as described above relies on the incorporation of a cloud product and automation of the categorization as described. Clouds masks are currently available as an additional band of TM and ETM+ surface reflectance products. These cloud masks will be available with all Landsat products in the near future, and the finalization of this product relies on the availability of such a cloud masks. Based on the cloud pixels in the cloud mask, and the preliminary distances given above, each pixel can be placed in one of three cloud categories: cloudy, clouds in the vicinity, or cloud free. These cloud categorizations will be provided as an additional band and the users will be given expected errors for each category. An example of one such additional band is shown in Figure 5.66; red pixels are cloudy, blue pixels have clouds in the vicinity, and black pixels are cloud free. Figure 5.67 is a figure of the original cloud product for reference. Table 5.14 shows the expected mean and standard deviation of each category, calculated based on the validation dataset. This table also shows what percentage of this particular scene falls in each category. While this was generated as an initial example, without preprocessing, it is obvious that considerations need to be made for the edges of the image and how to deal with single pixels warrants further investigation. Further

adjustments and improvements to this confidence metric expectation will be discussed in Section 6.2.
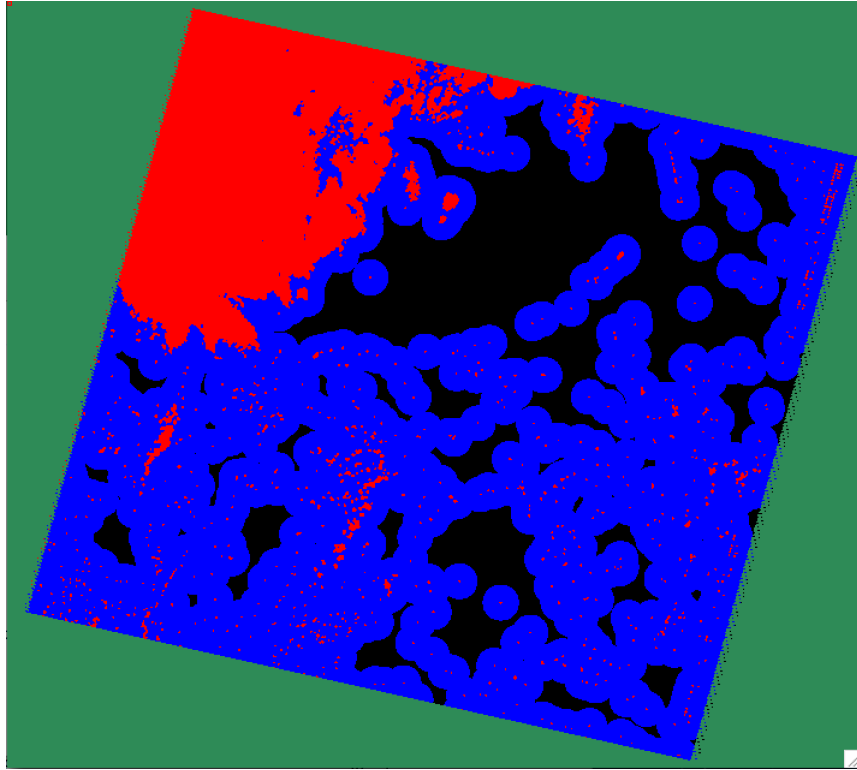
Figure 5.66: An example of the additional cloud categorization band that would be included based on the current confidence metric suggestion. Note that red pixels are cloudy, blue pixels have clouds in the vicinity, and black pixels are cloud free. Green represents the fill pixels around the scene.
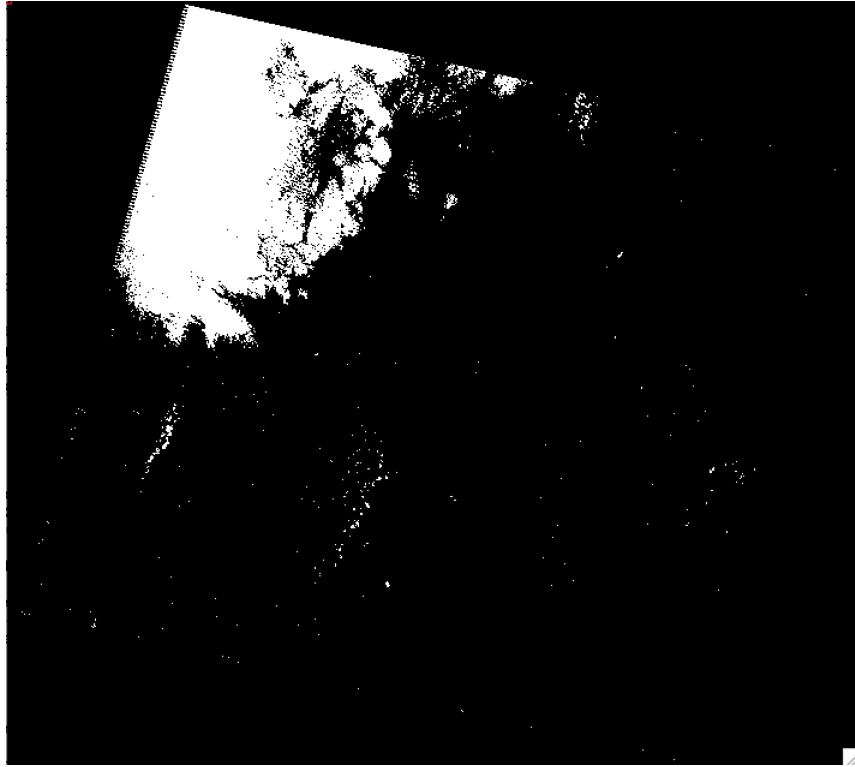
Figure 5.67: An image of the original cloud product from which Figure 5.66 was generated. Note that white pixels are clouds and black pixels (excluding fill pixels around scene) are cloud free.

Table 5.14: Summary of mean and standard deviations for cloud categorization example shown in Figure 5.66.

Pixel	Mean [K]	Standard Deviation [K]	% of Scene
Cloud Free (Black)	0.0	0.900	13.2%
Clouds in the Vicinity (Blue)	-1.340	3.239	65.8%
Black and Blue	-0.667	2.460	86.8%
Cloudy (Red)	Do Not Trust	NA	21.0%

5.4 Concluding Remarks

This chapter presented results obtained thus far with the methodology implemented as described in Chapter 4. We begin by introducing ground truth sites in Section 5.1 and validation of the methodology by considering results for cloud free scenes for each of these ground truth sites in Section 5.2. In Section 5.3 we present the development of our confidence metric through cloud analysis at each ground truth location. Overall, this accurately presents current performance and expectations, both of which are encouraging.

Although in Chapter 6 we present future work, including extension to a global product and further analysis of current confidence metric expectations, we should note here that validation of our current process shows that we can generate a product with low expected errors that would be very useful to the scientific community.

Chapter 6

Future Work

The goal of this chapter is to present any work that remains in the development of the atmospheric compensation component for a Landsat LST product and any initial progress or investigations that have already been made in this remaining work. As mentioned multiple times, the extension to a global atmospheric profile dataset for a global product is critical in utilizing much of the already existing Landsat archive as well as large area applications for which a LST product would be useful. As shown in Section 6.1, much initial work as been done to justify moving forward with a more rigorous validation for the proposed dataset. Then in Section 6.2, we consider methods to further investigate and improve our current confidence metric suggestion.

6.1   Extension to Global Dataset

As discussed in Section 2.2, once the process was sufficiently validated using the NARR data, we selected a global reanalysis dataset in order to extend the process to global coverage. Because we lack ground truth data at sites throughout the globe, we first aim to understand the behavior of the global reanalysis product over the United States. We then use this understanding to make initial relative comparisons at sites throughout the globe with varying climate and atmospheric conditions. This is a smaller, initial validation dataset in order to justify moving forward with a more extensive global development process.

6.1.1   MERRA Data Set

The modern-era retrospective reanalysis for research and applications (MERRA) dataset, uses a version of Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). Some inputs for the MERRA data product include radiosondes, wind profiles, aircraft data, dropsondes, and rain rates among other things. It provides 1.25° resolution around the globe 8-times daily at 42 pressure levels, with data available from 1979 to present [Rienecker and Gass, 2013]. The 1.25° resolution results in 288 points in longitude and 144 points in latitude. MERRA has the

123

same temporal resolution as the NARR dataset, with improved resolution in pressure but reduced spatial resolution. This dataset, like NARR, gives the geopotential height, specific humidity, and air temperature necessary for MODTRAN. However, MERRA also provides the relative humidity, eliminating the need for a specific humidity to relative humidity conversion. MERRA data for this work is downloaded as HDF files from the OPeNDAP data access from `http://goldsmr3.sci.gsfc.nasa.gov/opendap/MERRA/MAI3CPASM.5.2.0/contents.html` [OPeNDAP, 2013].

A comparison of the NARR and MERRA datasets is shown in Table 6.1.

Table 6.1: Comparison of North American and global datasets for atmospheric profiles.

	NARR	MERRA
Coverage	North America	Global
Spatial	32 km. spacing (0.3° at equator) 349 x 277	1.25° x 1.25° (140 km at equator) 288 x 144
Temporal	8x daily 3-hr intervals	8x daily 3-hr intervals
Pressure Levels	29 levels 1000 - 100 hPa	42 levels 1000 - 0.1 hPa

MERRA was selected over another global reanalysis dataset, as detailed in Appendix F.

6.1.2   Comparison to Ground Truth and NARR

A subset of the total validation dataset was selected and the same scenes were processed, using the same methodology, with the MERRA reanalysis data, rather than the NARR reanalysis data. A total of 397 scenes were selected, broken down by location as: 36 Rochester, 57 Delmar, 40 Georgia, 59 Santa Maria, 74 Santa Monica, 71 Lake Huron, and 60 Lake Superior. Figure 6.1 is a histogram of error values, for all 397 scenes calculated using Equation 5.1 comparing MERRA retrieved temperatures to ground truth temperatures. Figure 6.2 is the same dataset excluding scenes with clouds over the buoy, and Figure 6.3 includes only cloud free scenes. Note that the numbers in the title of each histogram indicate the cloud categorizations included in that histogram.
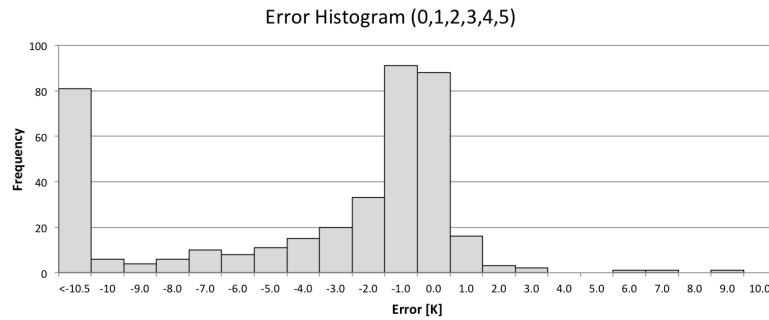
Figure 6.1: Histogram of error values for all scenes in MERRA dataset (397 total scenes). Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
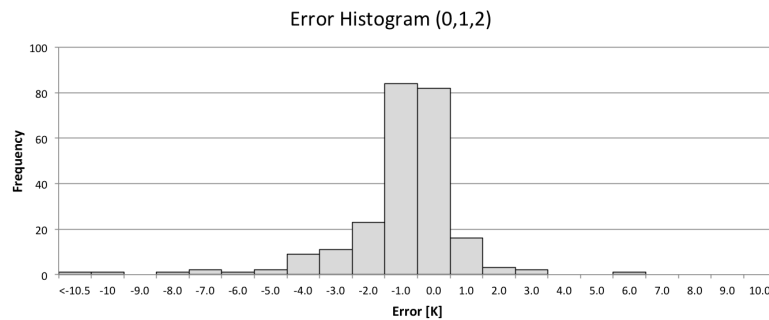
Figure 6.2: Histogram of error values for scenes in MERRA dataset with clouds in the vicinity and cloud free scenes. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.
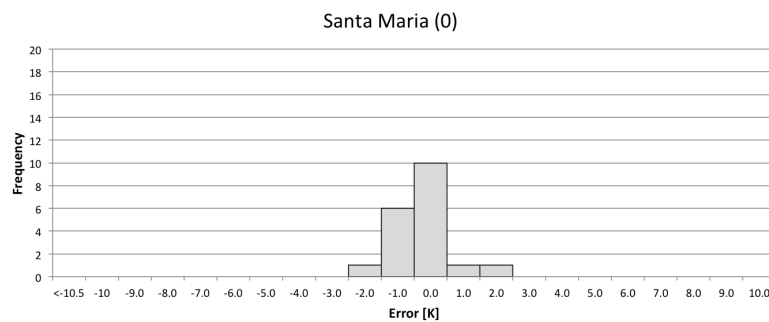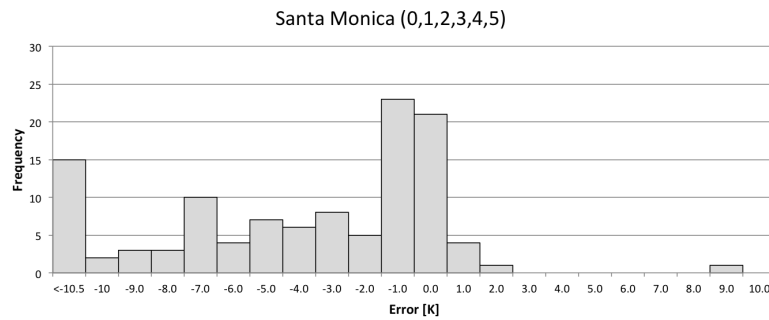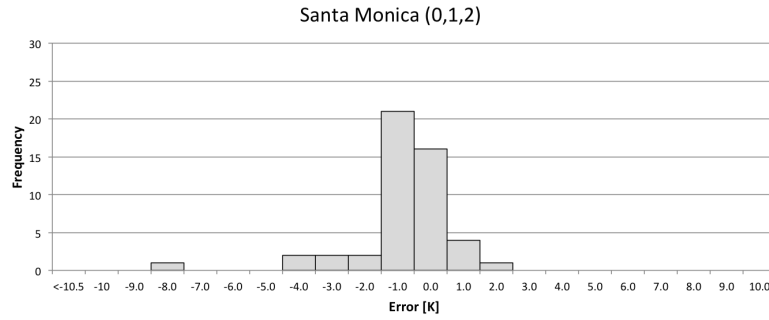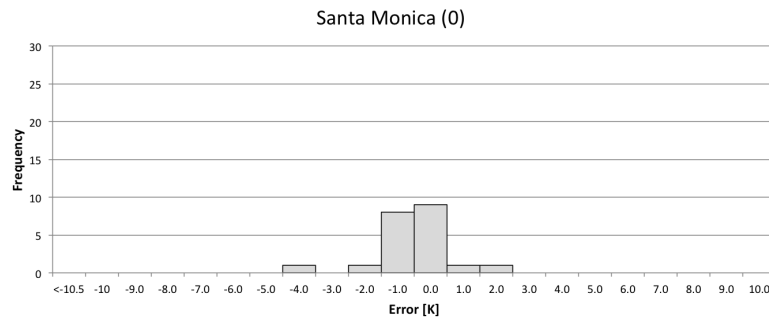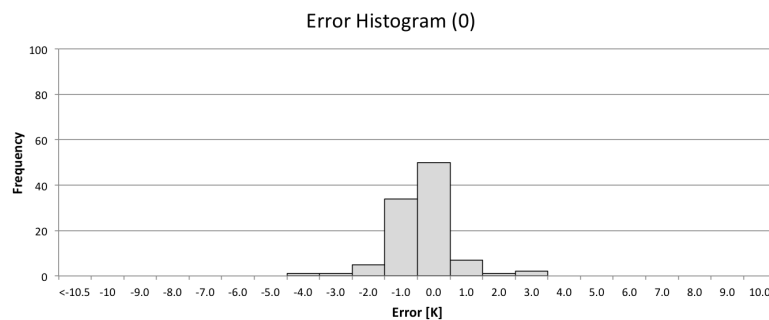

Figure 6.3: Histogram of error values for cloud free scenes in MERRA dataset. Note that the numbers in the title of the plot indicate the cloud categorizations included in the histogram.

These histograms have similar shape and show similar trends as those for the NARR dataset shown in Figures 5.62, 5.63, and 5.64. A summary of the mean and standard deviations of the

errors for the MERRA dataset is shown in Table 6.2; also shown in this table is the mean and standard deviation of errors for the same subset of scenes processed using the NARR dataset. This provides a direct comparison between MERRA and NARR results.

Table 6.2: Summary of mean and standard deviations of the errors for the same subset of scenes using MERRA and NARR datasets.

Cloud Category	Mean [K]		Standard Deviation [K]		Number of Scenes [% of Scenes]
	MERRA	NARR	MERRA	NARR	
0,1,2,3,4,5	-8.697	-8.470	18.520	18.190	397 [100%]
0,1,2,3	-1.474	-1.446	3.610	3.724	262 [66%]
0,1,2	-0.954	-1.002	1.846	2.139	239 [60%]
0,1	-0.513	-0.441	1.086	1.471	153 [39%]
0	-0.354	-0.235	0.911	0.921	101 [25%]

6.1.3 Comparison to MODIS SST

In Section 6.1.2, we compared MERRA retrievals to both ground truth and NARR retrievals. This initially shows the validity of the MERRA dataset at the same locations as the NARR dataset. Just because MERRA is available globally does not mean we can trust that performance will be consistent in other areas of the globe, where climates could differ or the density of input data could vary, among other factors. However, we lack both ground truth and NARR comparisons in other parts of the globe. Therefore, in order to validate MERRA retrievals in other parts of the world outside the United States, we look to compare against another currently available surface temperature product.

A MODIS TERRA product, aimed at different applications than Landsat because of its lower spatial but higher temporal and spectral resolutions, provides a good comparison because of the orbital similarity. MODIS TERRA has a 705 km, 10:30 a.m. descending node equatorial crossing, sun-synchronous, near-polar, circular orbit [Maccherone, 2014]. Landsat 7 also has a circular, sun-sun-synchrnous, near-polar orbit with equatorial crossing between 10:00 a.m. and 10:15 a.m. in descending node. Landsat 7 and TERRA have been collecting data in identical orbits since 1999 with acquisition times ideally 15 minutes apart [Tayolor, 2011]. Therefore, over a uniform area, MODIS TERRA provides a good comparison to Landast 7 because of the similarity in acquisition time.

MODIS TERRA produces a sea surface temperature (SST) product that we can use to compare to our retrieved land surface temperatures over water using MERRA reanalysis data. Ocean Level 2 products, including the MODIS TERRA SST product, are produced and distributed by the NASA Goddard Space Flight Centers Ocean Data Processing System (ODPS) [oce, 2010]. The MODIS TERRA SST product can be downloaded from the Ocean Color Browse on the Oceancolor Webpage [mod, 2014]. The HDF files were converted to Geotiff files and UTM coordinates using ENVI in order to select the same pixel location from both the MODIS image and corresponding Landsat 7

scenes. To average any slight non-uniformity in the area, a [3x3] pixel area of the MODIS product was averaged and compared to either a [5x5] pixel area of Landsat if the portion of the scene was unaffected or had a functioning scan line corrector, or an [11x11] pixel area with zero pixels removed if the portion of the Landsat scene was affected by the broken scan line corrector [lan, 2013]. The Ocean Level-2 products use a number of quality tests to set the quality level for each pixel; a small subset of all possible quality tests for all possible ocean products are used for the sea surface temperature product. The relevant quality tests are used to set the quality level for each pixel in the quality levels bands; each pixel has a valid range from 0, meaning best quality, to 3, meaning the pixel is invalid [oce, 2010]. The quality levels for the same [3x3] window were averaged, and if the average was greater than 0 (if the quality level for any pixel in the [3x3] window was greater than 0 or less than the best quality]), the pixel was removed from the analysis.

First, in order to justify comparing our Landsat MERRA retrieval to the MODIS SST product, a small set of Landsat 7 scenes over the already well understood ground truth sites were selected. A total of 71 Landsat 7 scenes were selected over the previously described Delmar, Georgia Coast, Santa Monica, and Lake Huron validation sites. As of 1 October 2013, Landsat 7 is calibrated with a mean bias of 0.0 K and a standard deviation of 0.42 K. Of these 71 scenes, 60 scenes had ground truth data available. A histogram of errors for all 60 scenes, comparing the retrievals using NARR reanalysis data and Landsat scenes to ground truth data, is shown in Figure 6.4. So that we consider best quality data, 16 scenes were removed because the MODIS quality level was greater than zero or a visual analysis of the Landsat scene showed clouds over or in the vicinity of the buoy; a histogram of the 44 remaining scenes is shown in Figure 6.5. Because this is the first time we have used Landsat 7, these plots are included to show that Landsat 7 for this small subset of scenes behaves in a similar fashion to our larger, more robust Landsat 5 validation dataset. This is also illustrated by the mean and standard deviation shown in Table 6.3. Similarly, MODIS scenes for the same 60 times and locations were compared to the same 60 ground truth data points, and the same 16 scenes were removed leaving the same 44 best quality scenes. Errors were calculated using Equation 6.1 and the error histograms for these datasets are shown in Figure 6.6 and 6.7 respectively. The mean and standard deviations of the errors for each dataset are summarized in Table 6.3.

$$\text{error} = \text{MODIS SST} - \text{Ground Truth Temperature} \qquad (6.1)$$
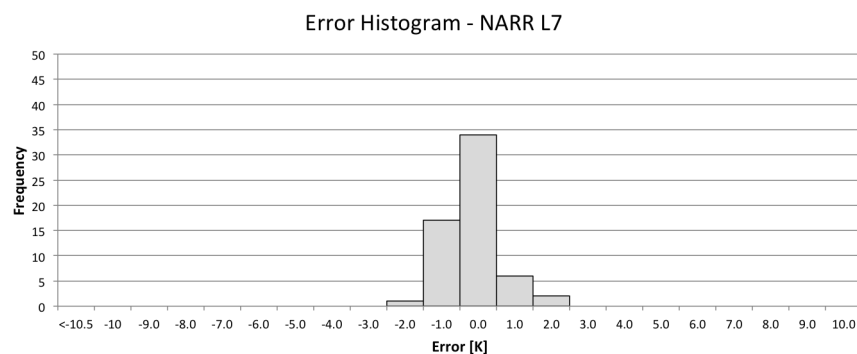
Figure 6.4: Histogram of error values comparing LST retrievals for Landsat 7 scenes using NARR reanalysis data to ground truth data.
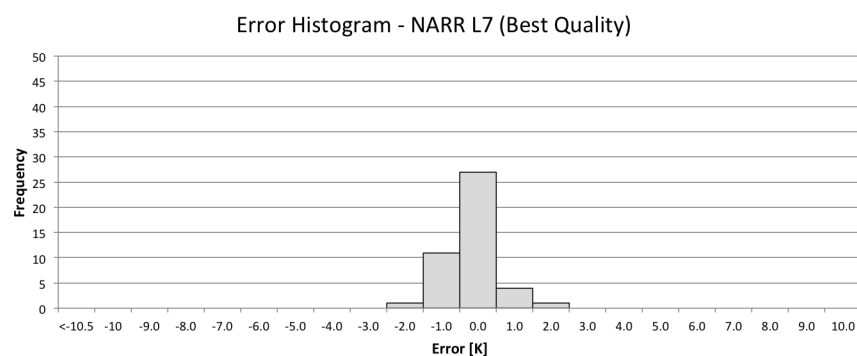
Figure 6.5: Histogram of error values comparing LST retrievals for Landsat 7 scenes using NARR reanalysis data to ground truth data for subset including best quality scenes.
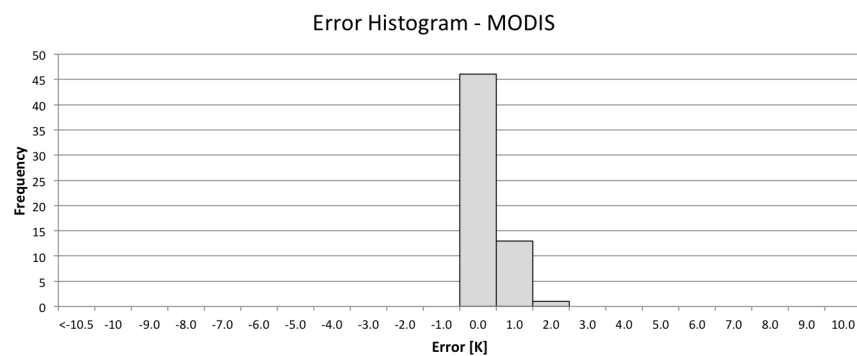
Figure 6.6: Histogram of error values comparing MODIS SST to ground truth data for scenes over North America.
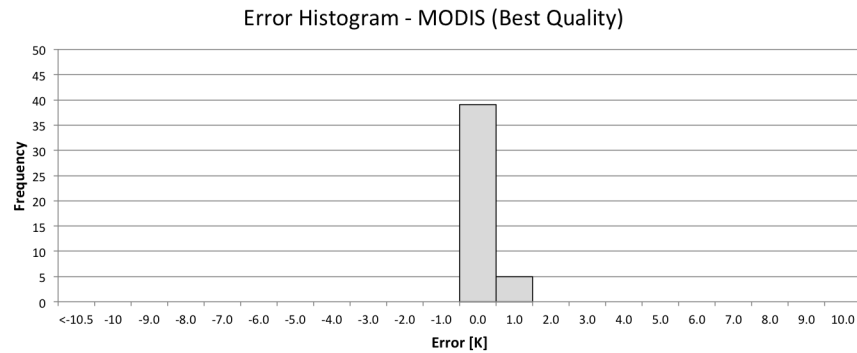
Figure 6.7: Histogram of error values comparing MODIS SST to ground truth data for subset including only best quality scenes over North America (same 44 scenes as used in Figure 6.5).

Table 6.3: Summary of NARR and MODIS SST errors compared to truth.

Dataset	Number of Scenes	Mean [K]	Standard Deviation [K]
Landsat and NARR	60	-0.20	0.68
Landsat and NARR, Best Quality	44	-0.20	0.63
MODIS SST	60	0.27	0.40
MODIS SST, Best Quality	44	0.20	0.27

It is interesting to note that the MODIS comparison to ground truth does not have the same negative bias as seen in the retrievals using the NARR dataset. The goal in this section is to justify using the MODIS SST product as a relative comparison for retrieval from our process generated with the MERRA dataset at sites throughout the globe. The MODIS to truth comparison yields mean errors less than 0.3 K and standard deviations less than 1 K. Figure 6.8 shows a comparison of NARR retrievals from Landsat 7 scenes to MODIS SST values; these values were calculated using Equation 6.2. Only the best quality scenes are shown for comparison (any scenes visually determined to be cloudy or with MODIS quality value greater than 0 were removed), but because ground truth data is not required, this histogram includes 53 scenes for comparison. The mean and standard deviation of the errors is -0.46 K and 0.69 K respectively. Because these means and standard deviations are sufficiently small, we move forward using MODIS as a tool for comparison to global retrievals using the MERRA data.
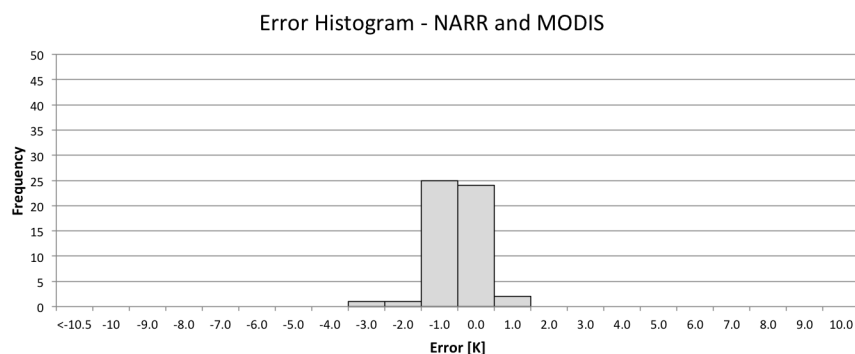
Figure 6.8: Histogram of error values comparing MODIS SST to LST retrievals from Landsat 7 scenes using NARR data for subset of best quality scenes.

Global sites were selected based on the density of radiosonde observations for MERRA reanalysis. Figure 6.9 shows one example of the density of inputs for a single MERRA reanalysis output. This particular plot is the 00Z radiosonde observations contributing to the specific humidity calculations for 17 July 2010. While radiosonde observations every 3 hours are more sparse, most maps of radiosonde observations every 12 hours look similar to the map shown. Based on these radiosonde observations, locations were selected with no radiosonde observations in the vicinity, sparse radiosonde observations in the vicinity and ample radiosonde observations in the vicinity. Points were selected over water within the center of the Landsat 7 scene where the operation of the scan line corrector is irrelevant. These locations are shown in Figure 6.10. The goal was to select locations that represented a variety of atmospheres (tropical, arctic, warm, cold, dry, humid, etc.) and a variety of input observation densities as an initial investigation into the feasibility of using MERRA data across the globe. Table 6.4 provides a description of the location and radiosonde availability for each of the locations shown on the map.
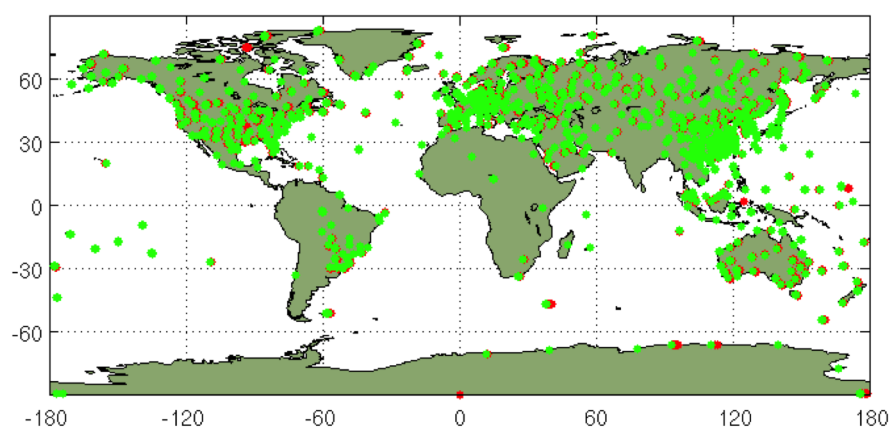
Figure 6.9: Density of radiosonde observations for MERRA calculations.

Figure 6.10: Points selected for MODIS and Landsat LST comparisons.

Table 6.4: Descriptions of locations and radiosonde densities for each point selected for global MODIS and Landsat LST comparisons. Note that the radiosonde description refers to the description of the density of input observations to the MERRA product provided in Figure 6.10.

Location	WRS-2 path_row	[lat, lon]	Radiosonde	Description
South America	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 - Nothern
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 - Nothern
South America	218_77	[-24.045, -45.18]	2	Low Lat - Southern
South America	233_93	[-47.88, -75.45]	1	Mid Lat - Southern

A total of 63 scenes across all 11 sites were processed and compared. Scenes were selected to be cloud free based on a visual analysis of the Landsat scene. A histogram of error results for all 63 scenes is shown in Figure 6.11. The errors in the histogram were calculated using Equation 6.2, such that negative errors still indicate an underestimation by our process. The mean of these errors is -0.05 K and the standard deviation is 1.15 K. Removing any scenes with less than the best MODIS quality, and a second pass at the visual cloud analysis to ensure only the best quality scenes were considered, left a dataset containing 48 scenes. A histogram of errors for these 48 scenes is shown in Figure 6.12. The mean and standard deviation of errors included in this histogram is -0.05 K and 0.76 K respectively. Table 6.5 shows the means and standard deviations of the error, and number

of scenes, for each individual location.

$$\text{error} = \text{Predicted LST} - \text{MODIS SST} \tag{6.2}$$
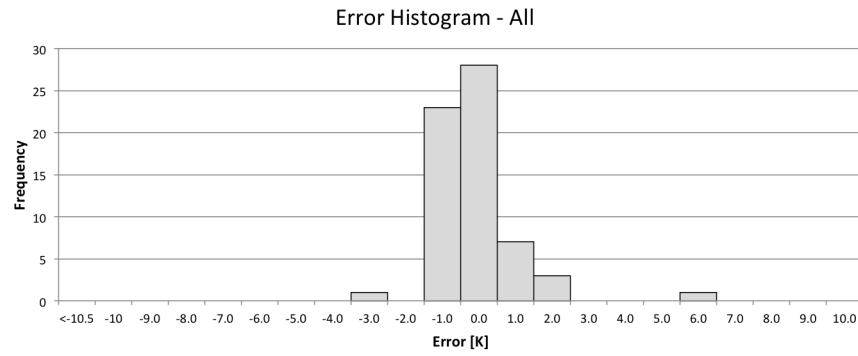
Figure 6.11: Error histogram for all MODIS SST and Landsat comparisons.
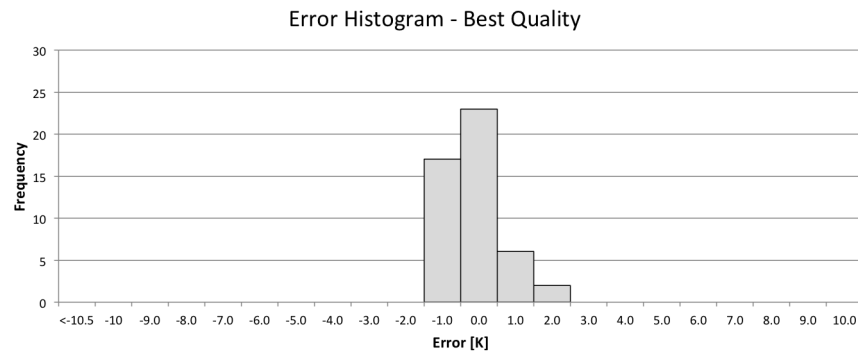
Figure 6.12: Error histogram for MODIS SST and Landsat comparisons for best quality scenes.

Table 6.5: Statistics of errors for each individual location for MODIS SST and Landsat comparisons. The error is calculated using Equation 6.2, n is the number of scenes analyzed for each location after removing scenes with less than the best quality, and SD is the standard deviation.

Location	Radiosonde	Description	n	Mean [K]	SD [K]
South America	1	Tropical	4	1.06	0.79
Mediterranean	2	Mid Lat - Northern	4	-0.27	0.28
Black Sea	2	Mid Lat - Northern	6	-0.41	0.45
India	0	Tropical	6	0.82	0.43
Hong Kong	2	Low Lat - Northern	4	0.09	1.31
Russia	1	High Lat - Northern	3	-0.60	0.32
Australia	2	Mid Lat - Southern	8	-0.32	0.43
Africa	0	Low Lat - Southern	4	-0.47	0.27
Greenland	1	High Lat - Nothern	1	-0.24	–
South America	2	Low Lat - Southern	3	0.58	0.29
South America	1	Mid Lat - Southern	5	-0.68	0.25
All	–	–	48	-0.05	0.76

Similar to the NARR comparison, this is an incomplete analysis but provides very encouraging initial results and justification enough to move forward with implementation and complete validation of a global product using the MERRA dataset. Providing variety in both location and input data density, retrievals consistently had mean errors less than 1 K and standard deviations typically less than 1 K. The overall mean error had a smaller magnitude than previous comparisons (NARR to ground truth or MERRA to ground truth), although MODIS to ground truth comparisons showed a small positive error. Analysis by location shows some larger positive errors, particularly in tropical atmospheres, conditions not evident from the analysis over the United States. This brings to light an area that deserves more attention with further development of the global product. Similarly, because of the small sample sizes used here as a first look, errors from location or atmosphere type would be confounded with density of input data. However, all results are reasonable for the targeted accuracy of the product. While a better understanding of the MERRA data is necessary in order to completely characterize a global product, this study validates that we can move forward with the implementation and validation of a global product using the MERRA reanalysis data.

6.2   Confidence Metric Improvements

As detailed in Section 5.3.3 our current best suggestion for a confidence metric for the user involves biasing the data by the expected error of our cloud free category and including an additional band in the product that provides the cloud categorization and expected error related to that classification. This section discusses work to be done before this confidence metric expectation can be implemented with certainty.

6.2.1 Discussion of Sources of Error

We begin with a more in depth consideration of all possible sources of error, in both the generation of atmospheric parameters and the analysis by comparison to ground truth, to ensure that we focus on components that can be and need to be improved. We consider both the categorical magnitude as well as whether the error is random or systematic. Systematic errors are predictable in some way, whether it is the precision associated with an input variable or a bias introduced by the process; randoms errors are unpredictable in our process. We also categorize each error as small, moderate, or considerable. Small errors we believe to be seemingly negligible in our process. Moderate errors are larger than small errors, but we do not believe that these errors are close to the limiting factor in the process. Considerable errors could be a limiting factor in our process and are worth improving. Each source of error is briefly discussed below and all are summarized in Table 6.6.

- **NARR Coordinates:** The NARR coordinates are natively given in Lambert Conformal coordinates, but the corresponding latitude and longitude coordinates, generated by bilinear interpolation, are provided when the data is obtained [NOMADS, 2012].

- **Height Conversion:** A tool was written for the process to convert the given geopotential height to the necessary geometric height for input into MODTRAN. This also utilizes the latitude of each point.

- **Humidity Conversion:** A tool was written for the process to convert the given specific humidity to the necessary relative humidity for input into MODTRAN. This also considers the temperature and pressure profiles at the given location.

- **Temporal Interpolation:** This is a linear interpolation between two samples three hours apart. After a sensitivity study to consider the data, we believe that a linear interpolation is the best option.

- **Appending Standard Atmosphere:** A MODTRAN standard atmosphere is appended above the NARR data for input into MODTRAN. A sensitivity study found that there were very small differences when using different standard atmospheres and therefore mid-latitude summer was used.

- **Layer Generation:** A linear interpolation between two existing layers is used to generate an atmospheric layer at the desired ground altitude for input into MODTRAN. A sensitivity study found linear interpolation to be sufficient.

- **Radiative Transfer Parameters:** Generating the radiative transfer parameters involves many steps, including generating a radiance due to temperature, calculating the satellite observed temperature (which incorporates the instrument noise), a linear regression to determine transmission and upwelled radiance, and solving for downwelled radiance. Sensitivty studies

found this method using three MOOTRAN runs to be sufficient when compared to methods using four or six MODTRAN runs.

- **Height Interpolation:** A linear interpolation between generated radiative transfer parameters over approximately 0.5 km or less. A sensitivity study found linear interpolation to be sufficient.

- **Location Interpolation:** An inverse distance weighting interpolation using four points approximately 32 km apart. A sensitivity study found this interpolation to be sufficient.

- **Location Determination:** This step involves registering the NARR data and the Landsat image and considers that the Landsat pixels are considered as single points within the scene. A conversion was written for the process to convert between UTM coordinates and latitude and longitude coordinates.

- **DEM:** This considers the uncertainty and precision associated with the digital elevation model used in the process.

- **Calibration:** This considers the uncertainty and precision associated with the current expectation for the calibration of the sensor [Schott et al., 2012].

- **Ground Truth:** Because our analysis involves comparisons to water temperatures, this considers any error associated with instrument noise on the buoy or platform and error associated with the surface to skin methodology [Padula, 2008], [Padula et al., 2010].

- **Emissivity:** Our analysis involves comparisons to water temperature, and we are confident in our emissivity estimations for water. Emissivity for the final product is being generated at the Jet Propulsion Laboratory, but we believe the uncertainty associated with the atmospheric compensation, not the uncertainty associated with the emissivity, to be the limiting factor in the final product [Hulley and Hook, 2009], [Hulley et al., 2012].

- **LUT:** A look up table is used to convert radiance due to temperature to temperature based on Plancks equation.

- **Atmospheric Profiles:** The atmospheric profiles input into MODTRAN (from NARR or MERRA in our analysis) have some associated uncertainty and precision for each value.

- **Errors in NARR:** From our analysis, we believe that in addition to the associated uncertainty and precision error for each NARR profile, due to the nature of the reanalysis, that the NARR data regularly misses localized occurrences of relative large amounts of column water vapor (sub visual cirrus clouds for example).

Table 6.6: Summary of all possible sources of error.

Source of Error	Random	Systematic	Small	Moderate	Considerable
NARR Coordinates		•	•		
Height Conversion		•	•		
Humidity Conversion		•	•		
Temporal Interpolation	•			•	
Appending Standard Atmosphere	•		•		
Layer Generation	•		•		
Radiative Transfer Parameters					
Calculate L_T		•	•		
Calculate L_{obs} (instrument noise)		•	•		
Linear Regression		•	•		
Solve for L_d		•	•		
Height Interpolation	•			•	
Location Interpolation	•			•	
Location Determination		•	•		
DEM		•	•		
Calibration		•			•
Ground Truth					
Instrument Noise		•			•
Surface to Skin		•			•
Emissivity		•	•		
LUT		•	•		
Atmospheric Profiles					
Specific Humidity		•			•
Temperature		•			•
Pressure		•	•		
Height		•	•		
Errors in NARR	•				•

6.2.2 Cloud Product Incorporation

The incorporation of the cloud product is discussed briefly in Section 5.3.3. There are a number of steps to automate the inclusion of an additional band with each pixel correctly categorized. Firstly, the cloud product must be provided and well understood so that it can be obtained and integrated into the automated process. Secondly, an algorithm to appropriately and automatically classify each pixel is necessary. This will require identifying cloudy pixels from the cloud product, and then all pixels within 0.5 km and 5 km. The subjective analysis in Section 5.3.2 should be compared to the objective analysis threshold results. The distance threshold may need to be adjusted and there may be more complicated determinations of cloud categories. Would it be beneficial to have a larger number of categories? Can we determine that the type of cloud in the vicinity is relevant and bias pixels with clouds in the vicinity based on cloud type or distance to cloudy pixel? How to grow cloudy pixels to categorize the surrounding area appropriately? As mentioned in Section 5.3.3, the edges of the image need to be considered and single pixels classified

as clouds also warrant further investigation. The bias applied to cloud free pixels, discussed further in Section 6.2.3 and Section 6.2.4, and the expected errors associated with each category, should be updated based on the objective and automated pixel categorization and results.

Previously, cloud free referred to cloud free in the vicinity of the buoy (allowing for clouds in other portions of the scene), based on subjective analysis. Another investigation was done to analyze only images without clouds in the entire scene. Figure 6.13 shows the error histogram for 95 scenes with no clouds in the whole image. While there is some improvement in the mean and standard deviation (-0.191 K and 0.871 K), the shape of the histogram is still similar to that for images only cloud free in the vicinity of the buoy. However, as shown in Appendix A.3, for images cloud free in the vicinity of the buoy, results are statistically different than the current expected calibration precision at alpha level 0.05, but for completely cloud free scenes, the mean error is not statistically different than the current expected calibration precision. This should also be considered when incorporating the cloud product.
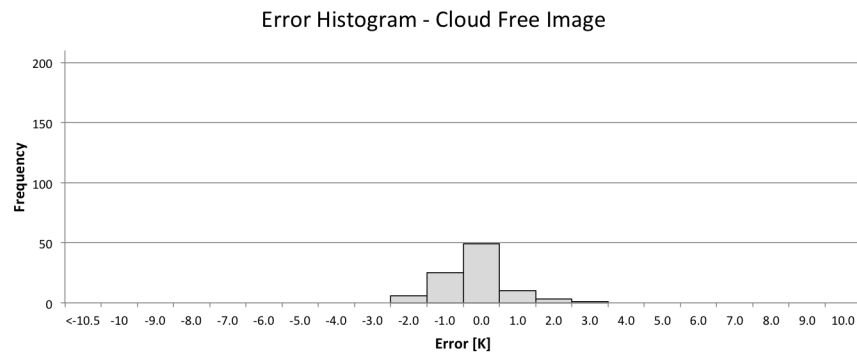
Figure 6.13: Error histogram for images without any clouds in the whole scene.

6.2.3   Column Water Vapor Consideration

The column water vapor comparison study, in Appendix B, and column water vapor adjustment study, in Appendix C, are both discussed in Section 5.3.3. These preliminary studies show that NARR underestimates column water vapor in comparison to a comparable in-scene derived product and that adding column water vapor to the atmosphere consistently increased the predicted land surface temperature. While we are confident that it would be too difficult to attempt to adjust the column water vapor at any particular height, because results are inconclusive that the water vapor is missing from the same height every time, we are not confident that the suggested bias shift is a better solution than the addition of some amount of column water vapor. While the bias shift will zero center our data and reduce expected errors for other cloud categorizations, we are not confident that adding some amount of column water vapor, distributed throughout the column, would not be a more appropriate adjustment to the retrieved temperatures. This requires further investigation of a larger dataset.

6.2.4 Calibration Consideration

Finally, we consider the assumption that the instrument is perfectly calibrated. For this analysis, we will only consider our original validation dataset of scenes that are cloud free in the vicinity of the buoy. As shown in Appendix A.3, the negative bias in the data is significant for Landsat 5 at alpha level 0.05 when considering 259 images cloud free in the vicinity of the buoy. However, this comparison is based solely on the precision of the calibration and does not allow for any uncertainty in the mean bias. While the uncertainty in the overall traditional calibration process has been investigated ([Schott et al., 1999] and [Padula, 2008]), it is difficult to assign an uncertainty value to the calibration of each instrument, and to couple this uncertainty with the precision given. With the current assumption that the calibration is zero centered, the LST and calibration results for Landsat 5 are significantly different at alpha levels 0.05, 0.01, and 0.005. However, if for example the mean bias in the calibration is actually currently -0.1 K, results would not be significantly different at alpha level 0.01, and if the mean bias in the calibration is actually currently -0.2 K, the results are not significantly different at alpha level 0.05. Furthermore, as shown in Appendix A.5, the LST retrievals are not significantly different from the current zero centered Landsat 7 calibration. Because the results are different for two different sensors, we question if the negative shift observed is in our process, or the instrument, and if a bias shift for all pixels, based on centering cloud free data, is appropriate. Note that the Landsat 7 dataset is significantly smaller. Therefore, a more in depth statistical analysis, considering precision and uncertainty, as well as larger datasets for Landsat 7 and Landsat 8, would help to determine if a bias shift is actually appropriate.

It is important to note that these discussions aim to improve already acceptable results. The current LST methodology and a confidence metric categorizing pixels based on a cloud mask, without implementing a bias shift of cloud free pixels, has low expected errors and would produce results that would be useful to most users.

6.3 Concluding Remarks

In this chapter, we outline work that could still be done in order to improve the Landsat LST product, including extension to a global product and improvements to the current expectation for a confidence metric. Although simply presented as initial investigations, results are encouraging for moving forward using the MERRA dataset in the development of a global product. While we feel confident that a confidence metric based on cloud categorizations will be helpful to the user, there are a number of possible improvements, including the automated inclusion of the cloud product and objective and more rigorous categorization of each pixel, as well as more in depth consideration of the suggested applied bias shift, based on both column water vapor consideration and instrument calibration. Current and above options would be better developed and understood if the validation dataset grew to also include comparable number of Landsat 7 and Landsat 8 scenes, as well as a larger variety of atmosphere and cloud types, even just within the United States.

It is important to recall that the current process and suggested confidence metric, based on the

Landsat 5 validation dataset over North America, have low expected errors and produce results that we believe would prove very useful to a large number of users.

Appendix A

Statistical Tests

A.1 Students t-test

Equation A.1 is the equation for the t statistic and Equation A.2 is the equation for the degrees of freedom for the students t-test comparing samples with unequal sample size and unequal variance. In these equations, $\bar{Y}$ is the sample mean, s^2 is the sample variance, and N is the sample size.

$$T = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{A.1}$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{A.2}$$

A.2 Analysis of Variance (ANOVA)

A one way ANOVA test can be used to analyze the difference between group means. We use a single factor ANOVA test with unequal sample sizes. Equation A.3 is the equation for the f statistic, where MST_r is the mean square for treatments, MSE is the mean square for error, SST is the total sum of squares, SST_r is the treatment sum of squares, and SSE is the error sum of squares. Equations for SST, SST_r, and SSE are shown in Equations A.4, A.5, and A.6 respectively, where I is the number of samples, J_i is the samples size of the Ith sample, $X_{i,j}$ is the jth measurement from the ith sample, $\overline{X}_{i\cdot}$ is the mean of the Ith sample, and $\overline{X}_{\cdot\cdot}$ is the grand mean or mean of all measurements in all samples.

$$f = \frac{MST_r}{MSE} = \frac{\frac{SST_r}{I-1}}{\frac{SSE}{n-I}} \tag{A.3}$$

$$SST = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{i,j} - \overline{X}_{..})^2 \qquad (A.4)$$

$$SST_r = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (\overline{X}_{i.} - \overline{X}_{..})^2 \qquad (A.5)$$

$$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{i,j} - \overline{X}_{i.})^2 = SST - SST_r \qquad (A.6)$$

A.3 Comparison of Landsat 5 LST Retrievals to Calibration Data

Using the students t-test described in Appendix A.1, sample 1 is the LST validation dataset (cloud free in the vicinity of the buoy) and sample 2 is the calibration dataset. The mean biases and standard deviations are values at the land surface for both datasets. Values for each sample are shown in Table A.1. Table A.2 shows the results of the t-test for alpha level 0.05.

Table A.1: Variable values for both the validation dataset (cloud free in the vicinity of the buoy) and calibration dataset.

Variable	Validation$_1$	Calibration$_2$
$\overline{Y}$	-0.267 K	0.0025 K
σ	0.8896 K	0.7266 K
s^2	0.791	0.528
N	259	162

Table A.2: Results of t-test.

ν	390.31
α	0.05
$t_{1-\frac{\alpha}{2},\nu}$	1.966
T	-3.395

Because $|T| > t$, we reject the null hypothesis that $\mu_1 = \mu_2$. That is, the population means, between our LST retrieval data and the calibration data, are statistically different at alpha level 0.05. This tells us that the negative bias that we observe in our LST results is statistically significant outside current expected calibration precision.

This test was repeated for the LST validation dataset when including only scenes that were cloud free throughout the whole image. This test is summarized in Tables A.3 and A.4.

Table A.3: Variable values for both the validation dataset (cloud free throughout the entire scene) and calibration dataset.

Variable	Validation$_1$	Calibration$_2$
$\overline{Y}$	-0.191 K	0.0025 K
σ	0.8714 K	0.7266 K
s^2	0.759	0.528
N	95	162

Table A.4: Results of t-test.

ν	169.80
α	0.05
$t_{1-\frac{\alpha}{2},\nu}$	1.974
T	-1.827

Because $|T| <$ t, we accept the null hypothesis that $\mu_1 = \mu_2$. That is, the population means, between our LST retrieval data and the calibration data, are not statistically different at alpha level 0.05 when only completely cloud free images are considered.

A.4 Comparison of Mean Errors for Each Ground Truth Site

We used an ANOVA test, described in Appendix A.2, to determine if the mean errors for cloud free scenes at each location are statistically different. Results for each location are summarized in Table A.5 and test statistics are summarized in Table A.6 where $\nu_1 =$ I - 1 and $\nu_2 =$ n - I. Because f $\leq$ F, we do not reject the null hypothesis that $\mu_1 = \mu_2 = ...\mu_9$. Therefore, we can conclude that the mean errors at each location are not statistically different.

Table A.5: Mean and standard deviations of the errors and number of scenes for cloud free scenes and each location.

Location	$\overline{X}_{i\cdot}$	s	J_i
Salton Sea	-0.120 K	0.558	11
Lake tahoe	-0.213 K	0.713	89
Rochester	-0.068 K	0.639	20
Delmar	-0.447 K	1.179	34
Georgia	0.041 K	1.267	23
Santa Maria	-0.219 K	0.789	19
Santa Monica	-0.574 K	1.089	21
Lake Huron	-0.695 K	0.82	19
Lake Superior	-0.167 K	0.676	23

Table A.6: Test statistics for ANOVA test to determine if location means are the same.

$\overline{X}_{..}$	-0.274
SST	204.179
SST_r	0.481
SSE	203.699
MST_r	0.060
MSE	0.790
alpha	0.05
ν_1	8
ν_2	250
F_{α,ν_1,ν_2}	1.98
f	0.076

This test has the assumption of equal variance; this assumption is generally assumed to be true if twice the minimum standard deviation is more than the maximum standard deviation. Because this Salton Sea violates this assumption (note that this location has far fewer samples), we also perform the test excluding results from the Salton Sea. Results are shown in Table A.7. We do not reject the null hypothesis.

Table A.7: Test statistics for ANOVA test to determine if location means are the same, excluding results from the Salton Sea.

$\overline{X}_{..}$	-0.293
SST	200.811
SST_r	0.457
SSE	200.354
MST_r	0.065
MSE	0.811
alpha	0.05
ν_1	7
ν_2	241
F_{α,ν_1,ν_2}	2.01
f	0.081

A.5   Comparison of Landsat 7 LST Retrievals to Calibration Data

Using the Students t-test described in Appendix A.1, sample 1 is the Landsat 7 LST retrieval dataset and sample 2 is the calibration dataset. The mean biases and standard deviations are values at the surface for both datasets. Values for each sample are shown in Table A.8. Table A.9 shows the results of the t-test for alpha level 0.05.

Table A.8: Variable values for both the Landsat 7 LST retrieval dataset and calibration dataset.

Variable	Validation$_1$	Calibration$_2$
$\overline{Y}$	-0.171 K	-0.047 K
σ	0.612 K	0.559 K
s^2	0.375	0.312
N	44	370

Table A.9: Results of t-test comparing Landsat LST retrieval data to Landsat 7 calibration data.

ν	51.89
α	0.05
$t_{1-\frac{\alpha}{2},\nu}$	2.0066
T	-1.279

Because $|T| <$ t, we accept the null hypothesis that $\mu_1 = \mu_2$. That is, the population means, between our LST retrieval data and the calibration data, are not statistically different for alpha level 0.05.

A.6 Confidence Intervals

Confidence intervals for the Landsat 5 calibration dataset and the Landsat 5 validation dataset (cloud free in the vicinity of the buoy) were calculated at the 95%, 99%, and 99.5% confidence levels. Confidence intervals were also calculated for the Landsat 7 calibration and the Landsat 7 validation dataset (cloud free in the vicinity of the buoy). All of these confidence intervals are summarized in Figure A.1. Note that confidence intervals can overlap and results can still be considered significantly different. In this graphic, only the Landsat 7 results are not significantly different.
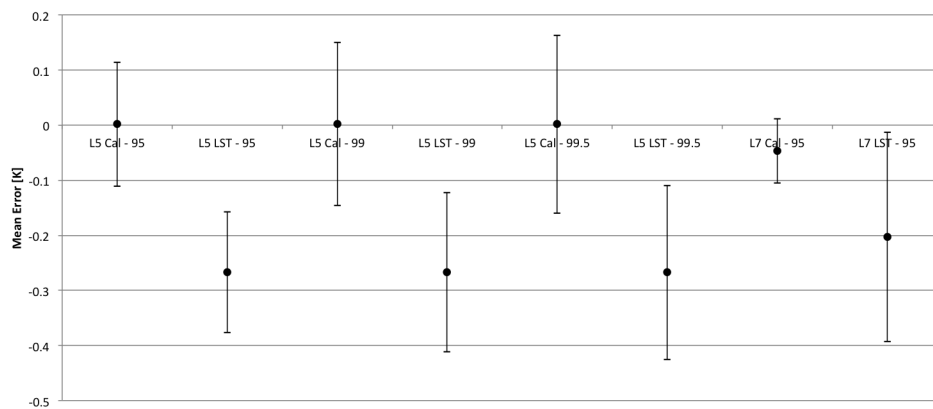
Figure A.1: Confidence intervals for Landsat 5 and Landsat 7 calibration and validation datasets (cloud free in the vicinity of the image).

Appendix B

Column Water Vapor Comparison Study

Based on familiarity with the data, the effects of cloud contamination, and the observed negative bias, we wanted to further investigate the column water vapor in the NARR product. This study compares the NARR column water vapor values to an in-scene derived column water vapor product.

We first wish to compare the column water vapor in the NARR atmospheric profiles to another reliable and concurrent source. Because we wish to make these comparisons relative to known temperature retrievals, and errors, we again turn to MODIS. As described in Section 6.1.3, because MODIS TERRA and Landsat 7 have acquisition times that differ by approximately 15 minutes, we can make comparisons between the MODIS TERRA SST product and LST retrievals using Landsat 7 scenes and the NARR data. Similarly, MODIS TERRA also has a total column precipitable water vapor product for the infrared retrieval. This gives the precipitable water vapor in the column in cm; the IR retrieval indicates a single column, rather than the total line of sight of water vapor. This is a better direct comparison to the NARR column water vapor. The total column precipitable water vapor IR retrieval also has a quality assurance band that provides a quality metric, indicating whether the pixel is useful or not useful, and whether it is fill or best quality. There are also indicators for the number of cloudy pixels in the [5x5] box, the number of clear pixels in the [5x5] box, the number of missing pixels in the [5x5] box, and the method of retrieval used (we wish to utilize the moisture profile integration) [Seemann et al., 2006].

Recall in Section 6.1.3, 71 scenes over Delmar, Georgia, Santa Monica, and Lake Huron were downloaded and processed and LST retrievals using NARR reanalysis profiles were compared to the MODIS SST product. A histogram of these temperature comparisons is shown in Figure 6.8. This histogram includes 53 of the 71 original scenes based on the MODIS SST quality and visual cloud analysis of Landsat scenes. The MODIS total column precipitable water vapor IR retrieval product (MOD07) was downloaded for the same 71 scenes and the NARR column water vapor was extracted for each of the same buoy points. A [3x3] pixel area was averaged for the MODIS

column water vapor values to account for any non-uniformity in the area. The NARR column water vapor is extracted from each of the four pertinent NARR points and spatially interpolated to the buoy location as described in Section 4.8. The MOD07 product passed all quality test for all 71 scenes, but only the 53 scenes that also passed the MODIS SST quality test and Landsat visual analysis were used in the column water vapor comparison. The MOD07 values were compared to the interpolated column water vapor values from the NARR profiles for these same 53 scenes. This comparison of total column precipitable water in cm is shown in Figure B.1.
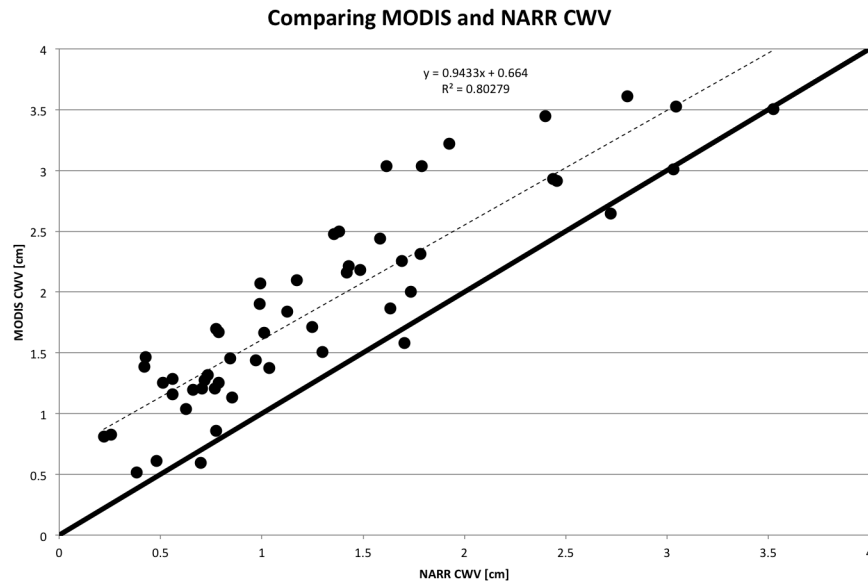
Figure B.1: Comparison of MODIS column water values from MOD07 and interpolated NARR column water vapor values. Note that the dotted line is a line of best fit, with the equation and R^2 values shown on the plot. The solid line is a one-to-one line for reference.

We can conclude from Figure B.1 that NARR underestimates column water vapor in comparison to MODIS, with some correlation as shown by the R^2 values on the line of best fit. The average difference, calculated by subtracting MODIS from NARR values such that negative values indicate underestimation by the NARR data, is -0.59 cm. This leads us to wonder if there is any correlation between difference in column water vapor and error in temperature. Figure B.2 plots the difference in retrieved temperature against the difference in column water vapor for NARR and MODIS.

Temperature and CWV Comparison
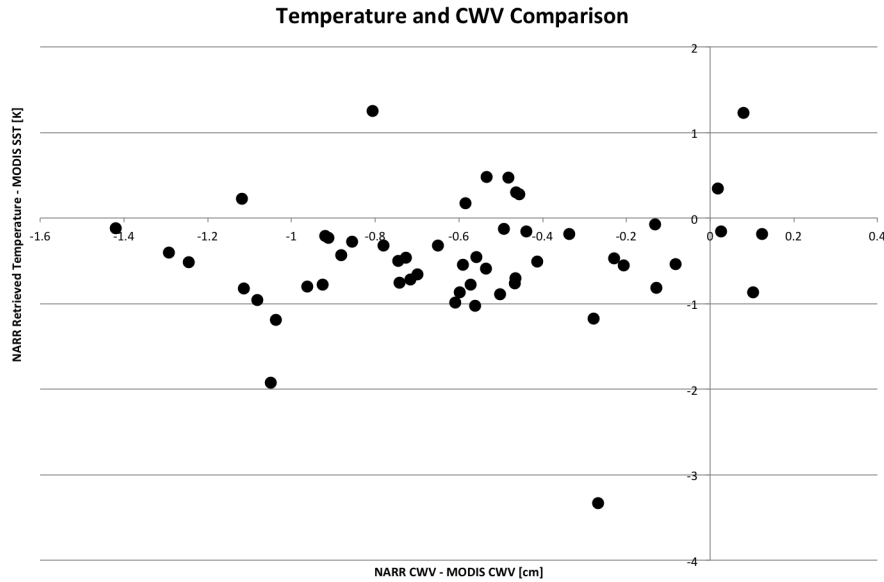
Figure B.2: Comparison of difference in temperature against difference in column water vapor for NARR and MODIS.

There is little or no pattern shown in Figure B.2. This indicates there is no relationship between error in temperature and difference in column water vapor; this suggests a constant difference in the temperatures retrieved, independent of column water vapor amount. The mean difference in temperature (average of values shown on Y axis) is -0.46 K. A similar plot was created for the difference between the retrieved temperature and ground truth data, compared against the difference in column water vapor. This plot showed a similar pattern and therefore was not included. From this we can conclude that although NARR regularly underestimates column water vapor in comparison to the MODIS product, there is no correlation between the magnitude of difference in column water vapor and the magnitude of difference in temperature estimation or error in temperature.

Although this is a relative comparison, we know that MODIS column water vapor is derived from in scene observations, rather than non-coincident measurements and numerical models like NARR. Therefore, we also compared the NARR column water vapor to MERRA column water vapor for the subset of 397 scenes in the MERRA validation dataset. This comparison is shown in Figure B.3. The solid line is a one-to-one line and the dotted line is a line of best fit, whose equation and correlation are shown on the plot. NARR also underestimates column water vapor in comparison to MERRA; the average difference is 0.107 cm.
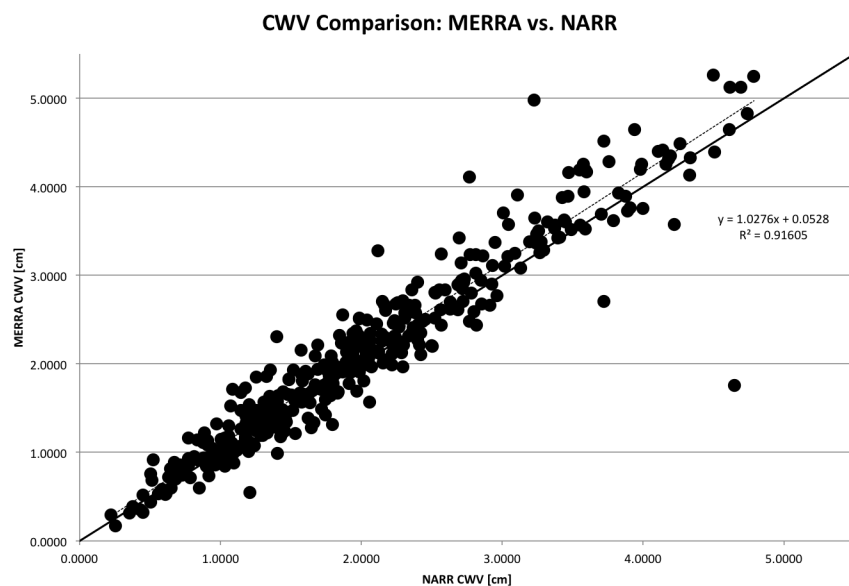
Figure B.3: Comparison of MERRA column water vapor and NARR column water vapor for 397 scenes over North America.

Appendix C

Column Water Vapor Adjustment Study

Based on the column water vapor comparison and observed negative bias, this study aims to investigate the effects of adding column water vapor to the atmospheric profiles by increasing the relative humidity values in the original NARR profiles.

As described above, comparison of MODIS and NARR column water vapor amounts indicated a constant underestimation of column water vapor in the NARR data. This section aims to explore the column water vapor contributions to the LST retrieval and the effects of adjusting the amount of column water vapor in the column.

As an initial investigation, the same 44 best quality scenes shown in Figure 6.5 were used. The column water vapor was increased by the average underestimation shown in the comparison in Appendix B. The addition of column water vapor is distributed throughout the profile. Figures C.1 and C.2 show one example of an original relative humidity and a new relative humidity profile with the total column precipitable water vapor increased by 0.59 cm.
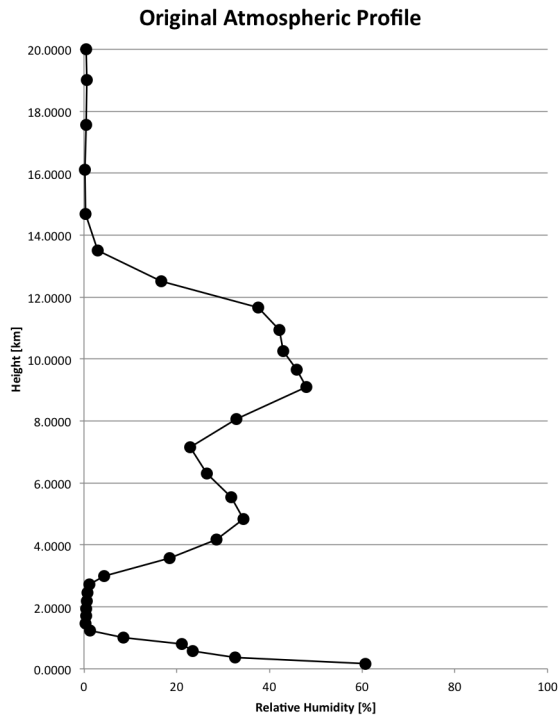
Figure C.1: An original relative humidity profile for a scene (Delmar) where the total column precipitable water is 0.24893 cm.
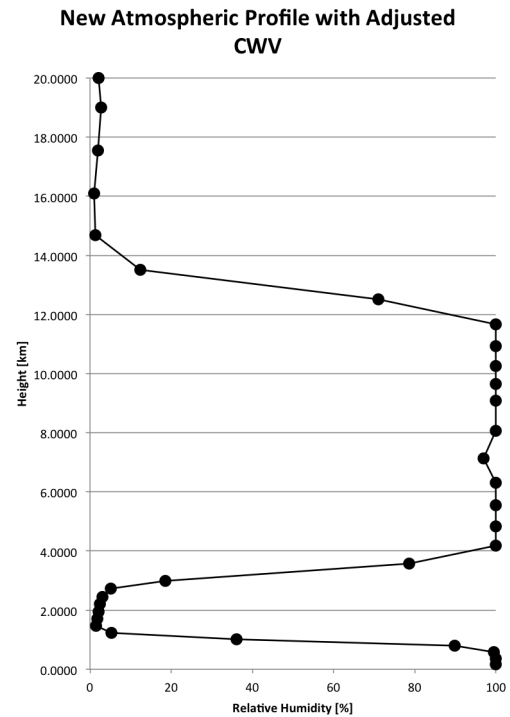
Figure C.2: The relative humidity profile for the same scene after the total column precipitable water vapor has been increased to 0.83893 cm.

Figure C.3 shows a histogram of the error results for LST retrieval (using Landsat 7 scene with NARR atmospheric data compared to ground truth retrievals). Note that this is the same as Figure 6.5, repeated here for convenience. Figure C.4 shows the errors results for LST retrieval after the total column water vapor for each atmosphere was increased by 0.59 cm.
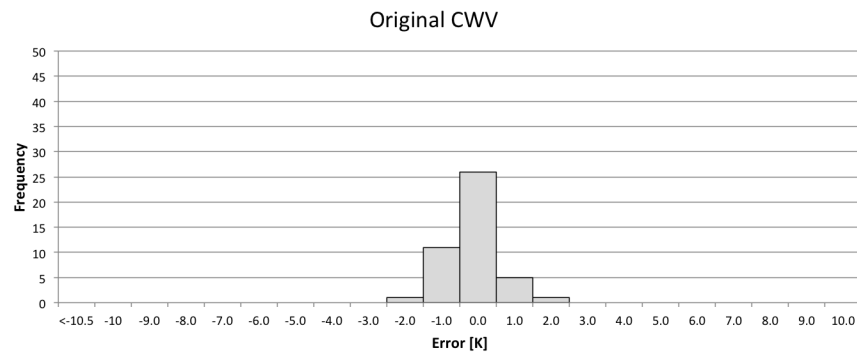
Figure C.3: Histogram of error results for 44 Landsat 7 scenes with the original atmospheric profiles.
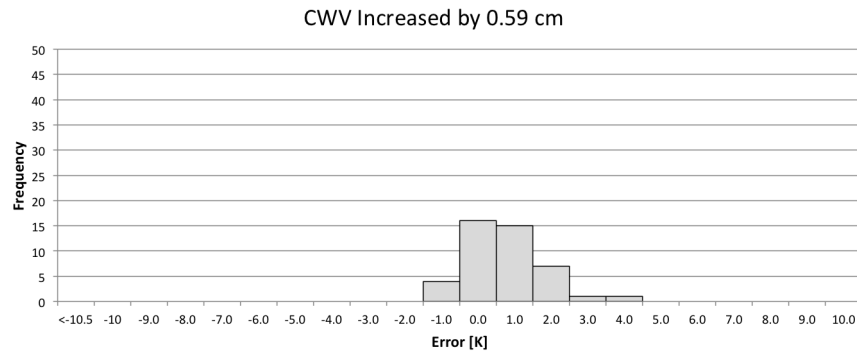
Figure C.4: Histogram of error results for 44 Landsat 7 scenes with the total column precipitable water vapor increased by 0.59 cm.

As shown, this adjustment in column water vapor appears to over compensate because the mean error for this small dataset increases from -0.171 K for the original atmospheres to 0.629 K for the adjusted atmospheres. Because we see in Appendix B that the MODIS SST, is on average, 0.46 K warmer than our LST retrievals (and the negative bias is in our dataset -0.267 K when compared to truth), we also explored results when the total column precipitable water vapor was increased by half the difference shown in the water vapor comparison, 0.295 cm. This histogram is shown in Figure C.5. Finally, we increase the column water vapor by 0.107 cm, the average difference between the NARR and MERRA column water vapor. This histogram is shown in Figure C.6. The mean and standard deviations for all adjustments are summarized in Table C.1. When the column water vapor is increased by 0.107 cm, the error results are not significantly different at alpha level 0.05 when compared to the current Landsat 5 calibration.
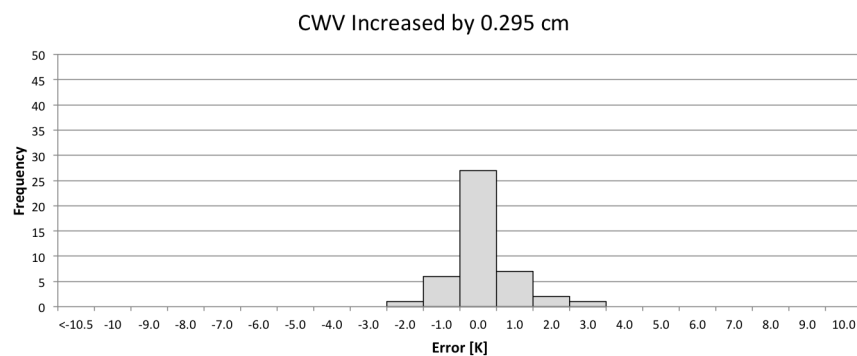
Figure C.5: Histogram of error results for 44 Landsat 7 scenes with the total column precipitable water vapor increased by 0.295 cm.
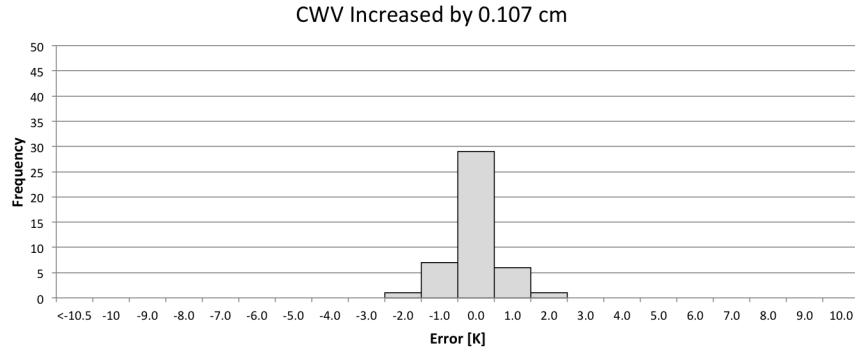
Figure C.6: Histogram of error results for 44 Landsat 7 scenes with the total column precipitable water vapor increased by 0.107 cm.

Table C.1: Summary of mean and standard deviation of errors for small Landsat 7 dataset with the original atmosphere, the total column precipitable water increased by 0.59 cm, and the total column precipitable water increased by 0.295 cm.

Atmosphere	Mean Error	SD of Error
Original	-0.171 K	0.612 K
CWV Increased by 0.59 cm	0.629 K	1.011 K
CWV Increased by 0.295 cm	0.189 K	0.782 K
CWV Increased by 0.107 cm	-0.048 K	0.672 K

As a final investigation, we explore adding the column water vapor to different levels of the atmosphere. For six atmospheres, selected for their variety in temperature and humidity, the relative humidity was increased by 20% or 40% in five atmospheric layers centered around 4 different heights (the first five layers, 1 km, 3 km, and 8 km). These relative humidity amounts were chosen to increase the total column precipitable water vapor enough to affect the retrieved LST but less than indicated above, and the heights were chosen as typical layers of low, middle, and high cloud formation. Because we saw only positive shifting when adding column water vapor throughout the profile, we were particularly interested in exploring the possibility of decreasing over estimations and increasing underestimations if the water vapor was added at a more appropriate height (if this additional water vapor is in fact a missed cloud influence) rather than distributed throughout the profile. One example of an adjusted atmosphere is shown in Figure C.7; five layers, centered around 3 km, were increased by 20% as shown in the Figure. The dotted blue line represents the original atmosphere and the black line the adjusted atmosphere. A new LST prediction was retrieved using each adjusted relative humidity profile and compared to that retrieved with the original profile.
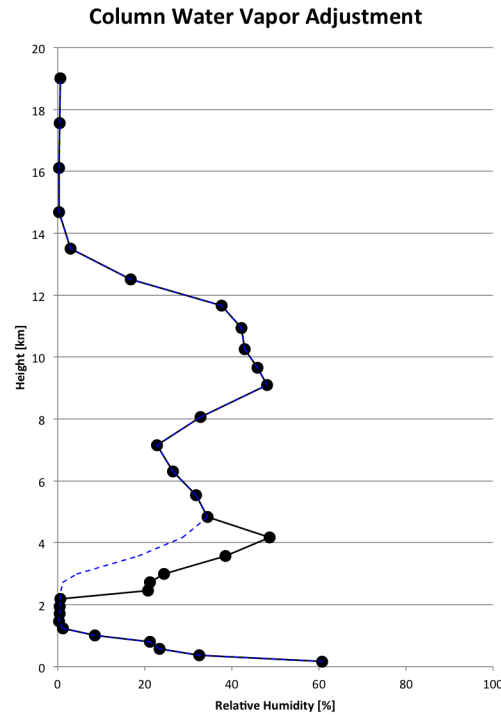
Figure C.7: An adjusted relative humidity atmosphere where 5 layers centered around 3 km were increased by 20% from the original atmosphere.

For all atmospheres except one, adding water vapor to the profile increased the retrieved temperature in all cases, even when the original profile overestimated in comparison to ground truth. For the one atmosphere where the added water vapor decreased the retrieved temperature, the original retrieval was already an underestimate in comparison to ground truth. Figure C.8 shows the change in column water vapor (different for each atmosphere based on how adjustments were implemented) compared to the change in retrieved temperature. This still shows results to be inconclusive but may be helpful in later consideration of the currently suggested bias shift.
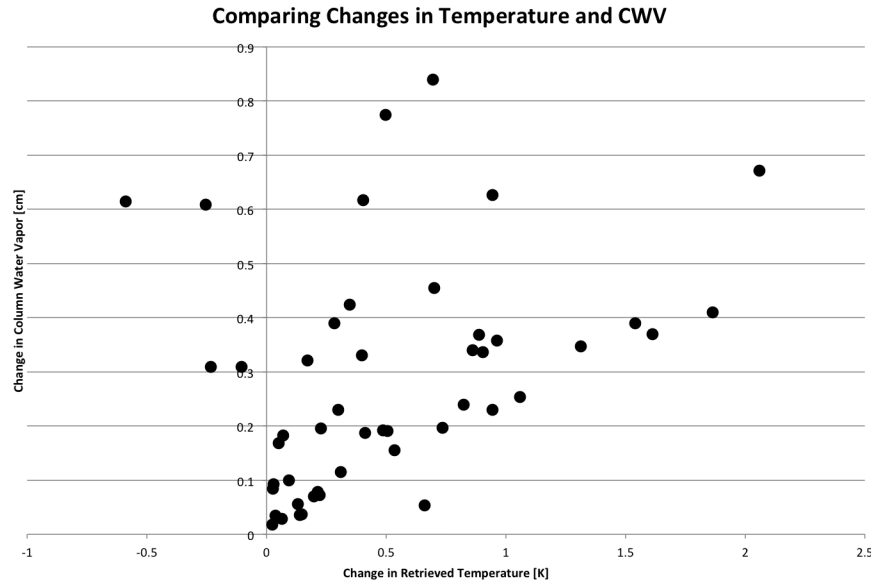
Figure C.8: Plot comparing the change in column water vapor against the change in temperature for column water vapor adjustments based on relative humidity.

Complete results from the study are included below, but conclusions from the column water vapor adjustment were inconclusive. There seemed to be no added benefits to adding water vapor at a particular height and no consistent cases of decreasing overestimations by adding water vapor. However, further investigation is warranted in the need to correct for the negative bias and the difference between adding column water vapor or simply applying a bias shift to the retrieved temperature results.

In the complete results of the column water vapor adjustment study shown below, dataset 1 is the original profile data and dataset 2 is the adjusted profile data. Note that dataset 1 is the same, regardless of how the atmosphere is adjusted.

Table C.2: Results for adding 20% relative humidity in the first five layers of the atmospheric profile.

Scene	Truth [K]	CWV_1 [cm]	CWV_2 [cm]	$Error_1$ [K]	$Error_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.33369	-0.64564	-0.62057
LE70130332010241EDC00	297.22846	1.96532	2.35487	-0.45692	-0.17294
LE70140332012270EDC00	293.23999	2.98713	3.30864	0.64671	0.81776
LE70160382010118EDC00	292.05833	1.0343	1.22613	0.49408	0.97993
LE70160382011137EDC00	296.25999	1.82008	2.05965	0.82195	1.64669
LE70410372009034EDC00	288.67137	0.70014	1.00947	-0.39131	-0.62426

Table C.3: Results for adding 40% relative humidity in the first five layers of the atmospheric profile.

Scene	Truth [K]	CWV$_1$ [cm]	CWV$_2$ [cm]	Error$_1$ [K]	Error$_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.41671	-0.64564	-0.59604
LE70130332010241EDC00	297.22846	1.96532	2.7401	-0.45692	0.04005
LE70140332012270EDC00	293.23999	2.98713	3.60464	0.64671	1.05094
LE70160382010118EDC00	292.05833	1.0343	1.37091	0.49408	1.39893
LE70160382011137EDC00	296.25999	1.82008	2.16675	0.82195	2.13512
LE70410372009034EDC00	288.67137	0.70014	1.31426	-0.39131	-0.98097

Table C.4: Results for adding 20% relative humidity in the five layers of the atmospheric profile centered at 1 km.

Scene	Truth [K]	CWV$_1$ [cm]	CWV$_2$ [cm]	Error$_1$ [K]	Error$_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.34186	-0.64564	-0.6176
LE70130332010241EDC00	297.22846	1.96532	2.38978	-0.45692	-0.1109
LE70140332012270EDC00	293.23999	2.98713	3.31711	0.64671	1.04452
LE70160382010118EDC00	292.05833	1.0343	1.22499	0.49408	0.99872
LE70160382011137EDC00	296.25999	1.82008	2.04998	0.82195	1.76592
LE70410372009034EDC00	288.67137	0.70014	1.00947	-0.39131	-0.49491

Table C.5: Results for adding 40% relative humidity in the five layers of the atmospheric profile centered at 1 km.

Scene	Truth [K]	CWV$_1$ [cm]	CWV$_2$ [cm]	Error$_1$ [K]	Error$_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.43093	-0.64564	-0.57612
LE70130332010241EDC00	297.22846	1.96532	2.80547	-0.45692	0.237
LE70140332012270EDC00	293.23999	2.98713	3.61341	0.64671	1.59182
LE70160382010118EDC00	292.05833	1.0343	1.39187	0.49408	1.45583
LE70160382011137EDC00	296.25999	1.82008	2.18978	0.82195	2.43584
LE70410372009034EDC00	288.67137	0.70014	1.30944	-0.39131	-0.64513

Table C.6: Results for adding 20% relative humidity in the five layers of the atmospheric profile centered at 3 km.

Scene	Truth [K]	CWV$_1$ [cm]	CWV$_2$ [cm]	Error$_1$ [K]	Error$_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.3487	-0.64564	-0.55242
LE70130332010241EDC00	297.22846	1.96532	2.30505	-0.45692	0.40303
LE70140332012270EDC00	293.23999	2.98713	3.24049	0.64671	1.70671
LE70160382010118EDC00	292.05833	1.0343	1.22127	0.49408	0.90469
LE70160382011137EDC00	296.25999	1.82008	2.01681	0.82195	1.55809
LE70410372009034EDC00	288.67137	0.70014	0.93054	-0.39131	-0.0911

Table C.7: Results for adding 40% relative humidity in the five layers of the atmospheric profile centered at 3 km.

Scene	Truth [K]	CWV_1 [cm]	CWV_2 [cm]	$Error_1$ [K]	$Error_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.44467	-0.64564	-0.41892
LE70130332010241EDC00	297.22846	1.96532	2.63654	-0.45692	1.60217
LE70140332012270EDC00	293.23999	2.98713	3.39717	0.64671	2.51055
LE70160382010118EDC00	292.05833	1.0343	1.40323	0.49408	1.38307
LE70160382011137EDC00	296.25999	1.82008	2.20994	0.82195	2.36217
LE70410372009034EDC00	288.67137	0.70014	1.15445	-0.39131	0.30804

Table C.8: Results for adding 20% relative humidity in the five layers of the atmospheric profile centered at 8 km.

Scene	Truth [K]	CWV_1 [cm]	CWV_2 [cm]	$Error_1$ [K]	$Error_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.26648	-0.64564	-0.62277
LE70130332010241EDC00	297.22846	1.96532	2.07996	-0.45692	-0.14725
LE70140332012270EDC00	293.23999	2.98713	3.06557	0.64671	0.86077
LE70160382010118EDC00	292.05833	1.0343	1.07154	0.49408	0.64034
LE70160382011137EDC00	296.25999	1.82008	1.85609	0.82195	0.96005
LE70410372009034EDC00	288.67137	0.70014	0.72866	-0.39131	-0.32691

Table C.9: Results for adding 40% relative humidity in the five layers of the atmospheric profile centered at 8 km.

Scene	Truth [K]	CWV_1 [cm]	CWV_2 [cm]	$Error_1$ [K]	$Error_2$ [K]
LE70130332010065EDC00	277.07999	0.24893	0.28334	-0.64564	-0.60752
LE70130332010241EDC00	297.22846	1.96532	2.018845	-0.45692	0.20363
LE70140332012270EDC00	293.23999	2.98713	3.14223	0.64671	1.18192
LE70160382010118EDC00	292.05833	1.0343	1.10715	0.49408	0.71576
LE70160382011137EDC00	296.25999	1.82008	1.88977	0.82195	1.01972
LE70410372009034EDC00	288.67137	0.70014	0.75581	-0.39131	-0.25938

Appendix D

MODTRAN Inputs and Outputs

This work uses MODTRAN4 Version 3 Revision 1 [Berk et al., 1999]. Because MODTRAN was developed in the punch card era, where programs and collections of data were stored on punch cards, input and output files are referred as the card deck and each file is called a tape file. The tape5 file is a precisely formatted input file and the execution of MODTRAN produces the tape6, tape7, tape7.scn, tape8 and the pltout and pltout.scn files.

MODTRAN is executed by inputting data into the tape5 file. Because of the historical punch card system, formatting of this file is extremely important. There have been various GUIs developed so that the user can specify inputs and a correctly formatted tape5 is generated for them. However, it is extremely important to understand each input to the program as many GUIs can have defaults that may or may not be desired. Each line in the tape5 file is referred to as a card; based on certain inputs, some cards are necessary while others can be omitted. A summary of the necessary cards and their corresponding inputs to MODTRAN for the runs completed in this work is summarized in Table D.1. Consult the MODTRAN manual for a more detailed description of the specific formatting [Berk et al., 1999].

Figure D.1 shows an example a default tape5 file for this work containing each of the inputs described in Table D.1. It has a complete NARR atmosphere but still contains flags for the temperature, surface albedo, and ground altitude.

Variables	Description	Input	Explanation
Card 1			
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	profile for temperature and pressure	0	
M2	profile for H_2O	0	JCHAR parameter in Card 2C1
M3	profile for O_3	0	supplies necessary profiles because
M4	profile for CH_4	0	user supplies model atmosphere
M5	profile for N_2O	0	
M6	profile for CO	0	
MDEF	CO_2, O_2, NO, SO_2, NO_2, NH_3, and HNO_3 profiles	1	default heavy species profiles
IM	read user input data	1	always read new user input data
NOPRNT	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
Card 1A			
DIS	select multiple scattering algorithm	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 recommend 8 streams in DISORT
LSUN	spectral resolution of irradiance	F	default solar 5 cm^{-1} spectral resolution irradiance
ISUN	FWHM of triangular scanning function	0	default values for FWHM
CO2MX	CO_2 mixing ratio in ppmv	360.00000	default value is 330 ppmv, recommended is closer to 365 ppmv
H20STR	vertical water vapor column character string	0	uses default water vapor column
03STR	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 aerosol 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 irradaince	0.000	do not scale TOA solar irradiance
Card 2			
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
Card 2C			
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
Card 2C1			

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)	defaults to M1 - M6 and MDEF	blank	MDEF = 1 specifies
JCHAR(5)	values when WMOL(2-3) are zero	blank	default profiles
JCHAR(6)		blank	
JCHAR(7)		blank	
JCHAR(8)		blank	
JCHAR(9)		blank	
JCHAR(10)	corresponds to WMOL(4-12)	blank	never read based on IRD1 in Card 2C
JCHAR(11)		blank	
JCHAR(12)		blank	
JCHAR(13)		blank	
JCHAR(14)		blank	
JCHARX	units for CFCs and other heavy molecules	blank	MDEF = 1 specifying default profiles
Card 3			
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 degrees) 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 but H1 and H2 (0-180 degrees)	0.000	sensor pointing directly at target
RO	radius of the Earth (km) at particular altitude of calculation	0.000	uses default mid-latitude 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 degrees)	0.000	

Card 3A1			
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
Card 3A2			
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
Card 4			
V1	initial frequency in wavenumber or wavelengths	9.000	wavelength in microns
V2	final frequency in wavenumber of wavelengths	14.000	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 no save current results

FLAGS(6:6)	degradation settings	blank	do not use saved results
FLAGS(7:7)	"spec flux" file	blank	do not write a spectral flux table
MLFLX	number of atmospheric levels of spectral flux values	blank	spectral flux values output at all atmospheric levels
Card 5			
IRPT	program execution setting	0	stop program

Table D.1: MODTRAN inputs to generate tape5 files for this work.

```
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 D.1: Image of an example of a tape5 file.

Appendix E

MODTRAN Run Study

A major portion of this work is the generation of the radiative transfer parameters. As implied in Section 4.1, there are a number of different methods to generate these parameters and the selection of the method is a balance between accuracy and computational intensity.

The most computationally intensive method uses six MODTRAN runs at combinations of three different temperatures and two different surface albedos. The governing equation

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

reduces to

$$L_{obs} = L_T \epsilon \tau + L_u$$

when ϵ=1. Therefore, we use three MODTRAN runs at three different temperatures ($T_1 = 273$ K, $T_2 = 295$ K, and $T_3 = 310$ K) with ϵ=1 to generate the transmission and upwelled radiance as shown in Figure E.1. We use an additional three MODTRAN runs, at the same three temperatures with ϵ=0.9, to perform a second linear regression. In this case, the slope and intercept can be rearranged, with the known transmission and upwelled radiance, to solve for the downwelled radiance using the equations shown in Figure E.1.
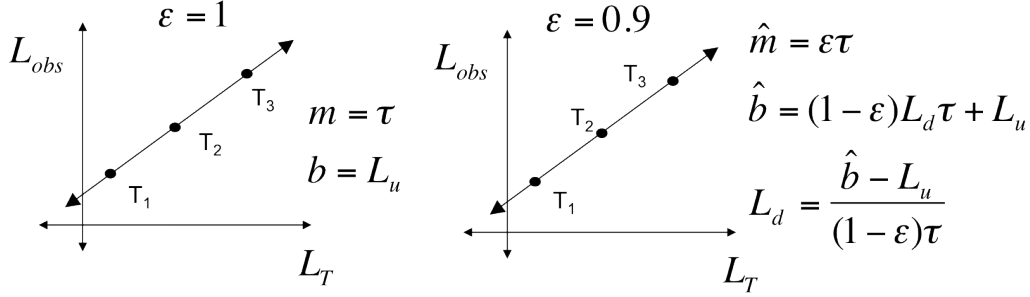
Figure E.1: Linear regressions and equations necessary to generate radiative transfer parameters with six MODTRAN runs.

This method requires the greatest number of MODTRAN runs (6) as well as two linear regressions.

Another method utilizes only four MODTRAN runs. The calculations for this method are the same as the six run method, but we perform both linear regressions with only two points rather than three. So, we use two MODTRAN runs at two different temperatures ($T_1 = 273$ K and $T_3 = 310$ K) with $\epsilon = 1$ and use a linear regression to solve for transmission and upwelled radiance. We then use two MODTRAN runs at the same two temperatures with $\epsilon = 0.9$ and a linear regression, with the known transmission and upwelled radiance, to solve for the downwelled radiance. This method is shown in Figure E.2.
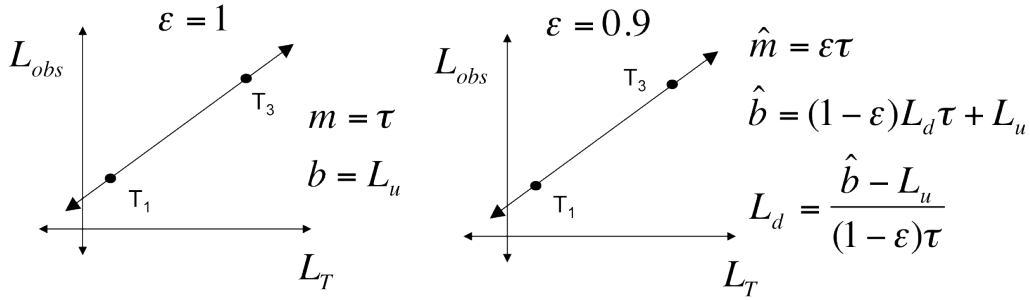
Figure E.2: Linear regressions and equations necessary to generate radiative transfer parameters with four MODTRAN runs.

This method requires the same two linear regressions but with fewer MODTRAN runs (4). The danger of using a two point linear regression is the error compounded by the error in a single point. This study should illustrate the difference, if any, in results between the three-point and two-point linear regressions.

There is another method that requires three MODTRAN runs, as described in Section 4.1. We use a two-point linear regression with two MODTRAN runs at two different temperatures ($T_1 = 273$ K and $T_3 = 310$ K) and $\epsilon = 1$ as shown in Figure E.3 to generate transmission and upwelled radiance, as with the four run method. Then, rather than using a two point linear regression to solve for the downwelled radiance, we rearrange the governing equation to solve for the downwelled

radiance and perform only one run with $\epsilon = 0.9$. Although the temperature of the MODTRAN run should not matter, because the linear relationship with real data is not perfect, we use T = '000'. This uses the temperature of the lowest atmospheric layer as the ground temperature for this run. This guarantees that we are using a ground temperature relatively close to the land surface temperature of the target. This is illustrated in Figure E.3.
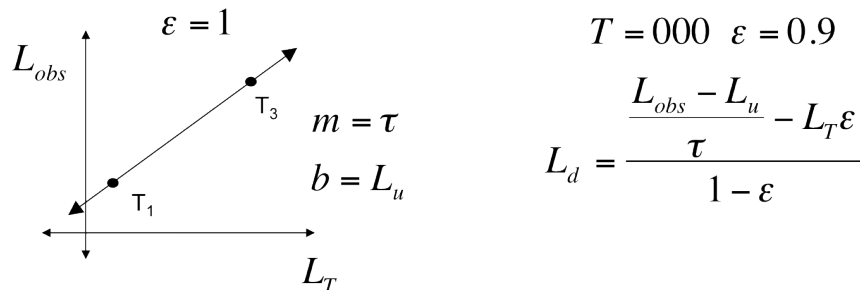
Figure E.3: Linear regression and equations necessary to generate radiative transfer parameters with three MODTRAN runs.

This requires one less MODTRAN run and one less linear regression if we find that solving for the downwelled radiance in this method is accurate. Similar to the two-point linear regression, this relies on a single MODTRAN run and any error from this single run could be compounded in our downwelled radiance results. In this case, unlike the six and four run methods, the retrieved radiative transfer parameters can vary slightly with the boundary temperature used in the calculations. Using the air temperature of the lowest layer of the atmospheric profile should guarantee the boundary temperature is reasonably close to the LST of the pixels where this downwelled radiance will be used.

In order to determine the optimal method, we generated radiative transfer parameters using all three methods with the same atmosphere. We used these radiative transfer parameters to generate the radiance due to the temperature and convert this to a temperature using a look up table as described in Section 4.1 using a range of boundary temperatures. In perfect results, we expect to retrieve these boundary temperatures. The retrieved temperature is plotted against height for test temperatures of 260 K, 280 K, 300 K, and 320 K is Figures E.4, E.5, E.6, and E.7 respectively.
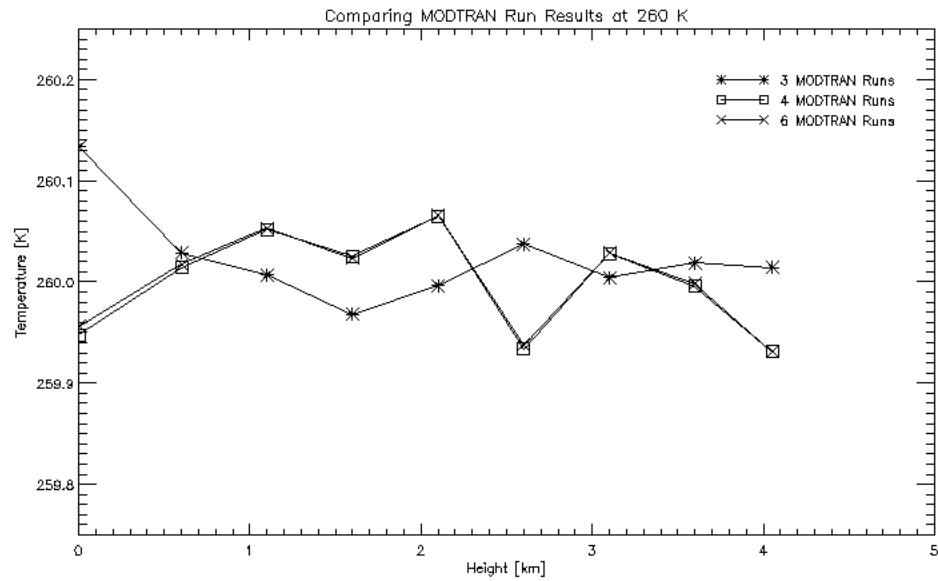
Figure E.4: Retrieved temperatures using three, four and six MODTRAN runs to generate radiative transfer parameters with a land surface temperature of 260 K.
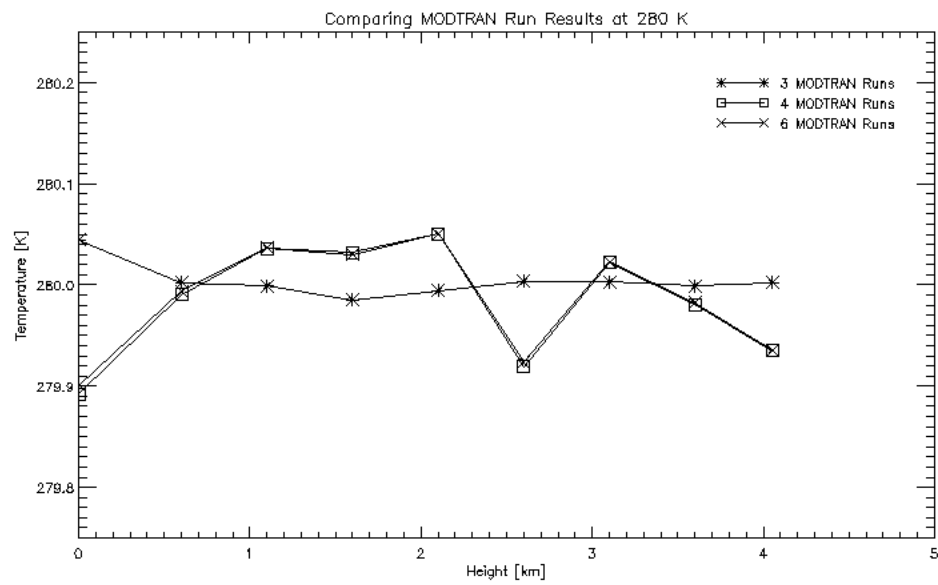
Figure E.5: Retrieved temperatures using three, four and six MODTRAN runs to generate radiative transfer parameters with a land surface temperature of 280 K.
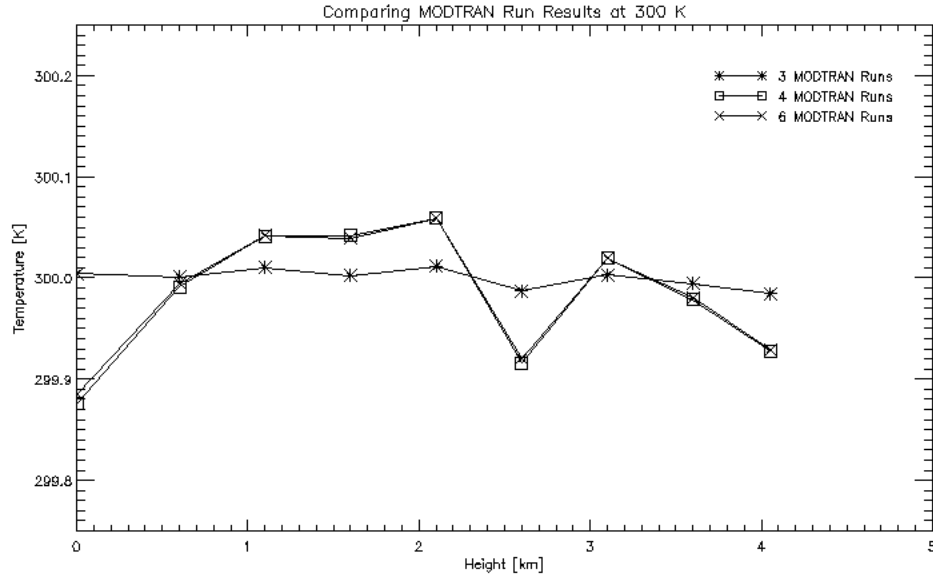
Figure E.6: Retrieved temperatures using three, four and six MODTRAN runs to generate radiative transfer parameters with a land surface temperature of 300 K.
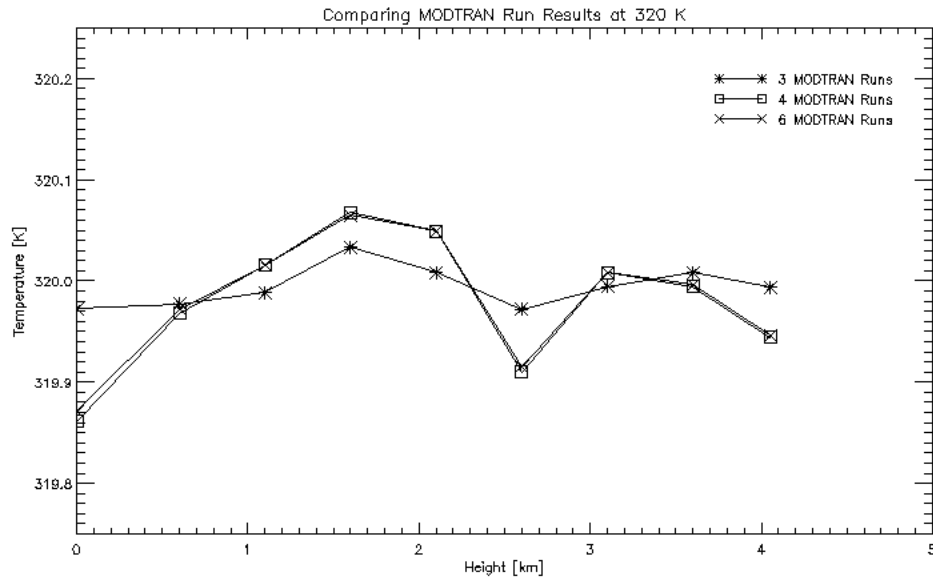
Figure E.7: Retrieved temperatures using three, four and six MODTRAN runs to generate radiative transfer parameters with a land surface temperature of 320 K.

Note the scale of each plot; the range spans only 0.5 K. Therefore, the differences in the retrieval results are small and all three methods are comparable. The four run and six run methods are almost identical, indicating there is negligible difference between the three-point and two-point

linear regression. Note that the three run method improves slightly as temperature increases; as discussed in Section 5.3, LST retrieval is more difficult with higher temperatures. While the four run and six run methods are constant for all temperatures, the three run method adapts slightly to this change, causing better results at higher temperatures. Therefore, because all three methods produce comparable results with three runs slightly better at higher temperatures and having the lowest computational intensity, the three run method was selected to be implemented in this work.

Appendix F

NCEP Dataset Consideration

NCEP/NCAR Reanalysis 1, referred to as the NCEP (National Centers for Environmental Prediction) dataset, provides data from 1840 to present. It also uses state of the art reanalysis to generate a consistent set of variables from inputs similar to other reanalysis such as radiosondes, pibals, and aircraft data. It is presented in a 2.5° by 2.5° global grid (144 by 73) 4-times daily; some variables are given at 17 pressure levels and others at 8 pressure levels. While air temperature and geopotential are given at 17 pressure levels, specific humidity and relative humidity are both given at 8 pressure levels, reducing the resolution of our process to 8 pressure levels [NOAA/ESRL/PSD, 2013]. This is reduced resolution in all dimensions. The NCEP data for this work was downloaded as packed NetCDF files from the FTP

`ftp.cdc.noaa.gov/Datasets/ncep.reanalysis/pressure`

[NCEP/NCAR, 2013].

Table F.1 provides a comparison between NARR, MERRA, and NCEP.

Table F.1: Comparison of datasets for atmospheric profiles.

	NARR	MERRA	NCEP
Coverage	North America	Global	Global
Spatial	32 km. spacing (0.3° at equator) 349 x 277	1.25° x 1.25° (140 km at equator) 288 x 144	2.5° x 2.5° (278 km at equator) 144 x 73
Temporal	8x daily 3-hr intervals	8x daily 3-hr intervals	4x daily 6-hr intervals
Pressure Levels	29 levels 1000 - 100 hPa	42 levels 1000 - 0.1 hPa	8 levels 1000 - 300 hPa

The methodology proposed in Chapter 4 was adapted using the same interpolations and methods for each of the above described datasets. A subset of 33 scenes with available ground truth data from Chapter 5 was selected. By using our global datasets on scenes over North America, we can

compare the results to both truth and results generated using NARR data. Figures F.1, F.2, and F.3 show error histograms generated using the same 33 scenes; all scenes selected were included regardless of clouds or atmospheric conditions. Errors were calculated using Equation 5.1. It should be noted that there are scenes selected from the Salton Sea (6), Lake Tahoe (11), Rochester (9) and Delmar (7).
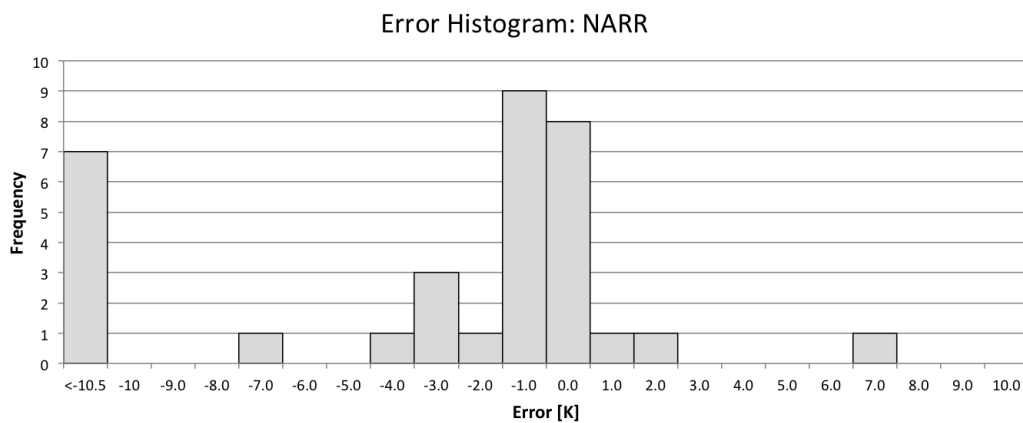
Figure F.1: Histogram of errors for 33 scenes using the NARR dataset for atmospheric profiles.
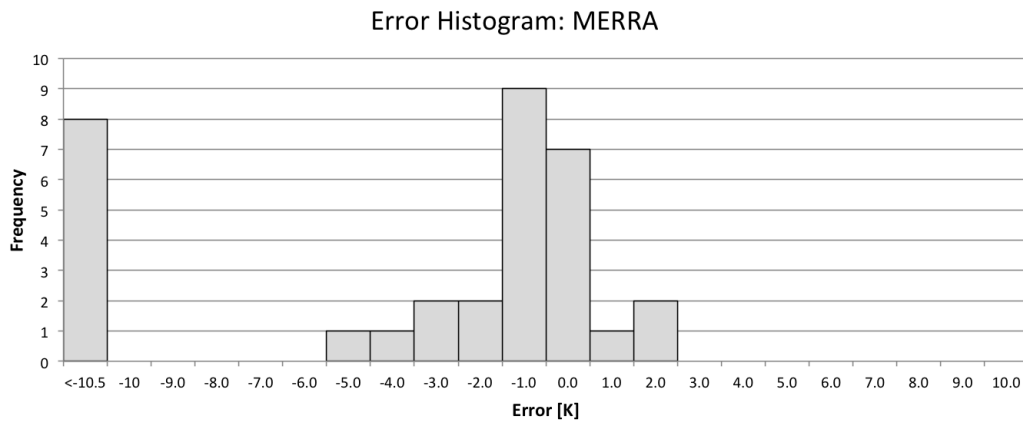
Figure F.2: Histogram of errors for 33 scenes using the MERRA dataset for atmospheric profiles.

Error Histogram: NCEP
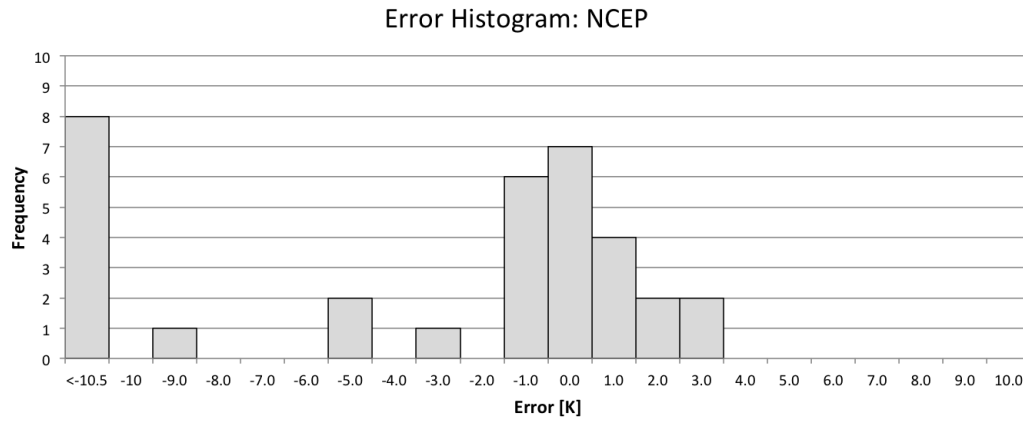
Figure F.3: Histogram of errors for 33 scenes using the NCEP dataset for atmospheric profiles.

Although there is some variability, no one dataset appears to have particularly different results for this small subset of scenes. Because of the reduced resolution of the NCEP dataset in all variables and dimensions, MERRA was chosen for initial investigation for a global LST product.

Appendix G

Previous Confidence Metric Investigations

Various methods for developing a confidence metric were also investigated before the error propagation and cloud analysis described in Section 5.3. Our investigations led us to believe that those methods would prove more accurate, or more reliable, than those described below. These are included here for completeness of work. As will be described below, these investigations were performed before the validation dataset was complete and before the visual cloud analysis. A simpler initial cloud detector was used, as described below, and only scenes from five buoy locations (Salton Sea, Lake Tahoe, Rochester, Delmar, and Georgia Coast) were used. Although evaluated with a smaller dataset, we believe that these methods of confidence metric estimation would not prove useful for the final product.

G.1 Metrics

Both methods of error analysis below are based on the theory that as certain atmospheric variables increase or decrease, the ability to accurately retrieve the land surface temperature improves or worsens. Rather than incorporate an additional data source, the confidence estimation will be made with metrics already contained within or generated from the atmospheric profile data and radiative transfer code. The five metrics that were chosen are described below.

The transmission, generated as described in Section 4.1, is one metric used to estimate the confidence in the predicted land surface temperature. Certain conditions do not lend themselves to accurate LST estimation, regardless of the performance of the method or process used. Low transmission of the atmosphere above the pixel of interest is one of these. With a low transmission, less radiance from the ground can reach the sensor. Therefore, the LST retrieval is less accurate with a lower atmospheric transmission and more accurate with a higher atmospheric transmission.

The second metric is the maximum air temperature in the atmospheric profile. This is retrieved

from the NARR atmospheric profile data. Because this is an input to MODTRAN and MODTRAN is not executed at every pixel, an atmospheric profile does not exist for every pixel. However, the same four NARR locations used to interpolate the radiative transfer parameters to that pixel location, as described in Section 4.8, are used to estimate the maximum air temperature in the atmospheric profile at a single pixel. The maximum air temperatures from each of these four NARR profiles are spatially interpolated using Shepard's method, as described in Section 4.8, to generate an estimation of the maximum air temperature in the atmospheric profile above the pixel of interest. This does not consider the height in the profile of the maximum temperature; we simply select the highest temperature from each profile and interpolate regardless of elevation. A higher maximum temperature tends to indicate a thicker, denser atmosphere, generating more path radiance. Upwelled radiance can be a large contributor to error, so a small error in the measurement of a hot temperature can lead to a large error in upwelled radiance. A similar magnitude error in a cold temperature would not cause as large an error in path radiance. Less intuitive but similar to a low transmission, it is less likely that the ground leaving radiance is accurately sensed from the satellite with higher air temperatures because the additive path radiance effect introduces error. Therefore, higher maximum air temperatures in the atmospheric profile indicate less accurate predicted land surface temperatures.

The last three metrics are all related to the amount of humidity in the atmosphere above the pixel of interest. Just like maximum air temperature, because MODTRAN is not executed at every pixel, an atmospheric profile does not exist for every pixel. The following three parameters are spatially interpolated from the four NARR points used to interpolate the radiative transfer parameters, like maximum temperature described above.

The relative humidity profile is one of the variables derived from the NARR data and used as an input profile in MODTRAN. This profile is a measure of the amount of water in the air at various heights at that location. Higher humidity levels hinder the amount of radiance reaching the sensor, and therefore a higher relative humidity means a less accurate land surface temperature retrieval. The maximum relative humidities from each of the four NARR points are interpolated to generate as estimate of the maximum relative humidity at the pixel of interest. As with maximum air temperature, elevation is not considered. Theoretically, the lower the maximum relative humidity the more accurate the land surface temperature retrieval.

The dew point depression is another measure of humidity in the atmosphere. The dew point temperature, directly related to relative humidity, is the temperature at which humidity in the air will condense to liquid water. The closer the air temperature is to the dew point temperature, the higher the water vapor content. The dew point depression is the difference between the dew point temperature and the air temperature. A smaller dew point depression indicates a higher level of humidity and therefore should correlate to less accurate land surface temperature prediction. This information is largely redundant to the maximum relative humidity level but both are considered; one may exhibit better confidence estimation due to differences in magnitude or interpretability.

The final metric is the total column water vapor. Also a measure of water in the profile, the

column water vapor can be interpreted in two ways. If all water molecules in the atmosphere were brought to the surface of the Earth at a pressure of one atmosphere and temperature of $0°C$, the water column would have some thickness in atmosphere-centimeters. Similarly, over each square centimeter of ground surface, the water molecules in that column of the atmosphere have some mass in grams per centimeter squared (gm/cm^2). Because the density of liquid water is 1 gm/cm^2, the mass of the water molecules in the column in grams per centimeter squared is equal to the centimeters of water on the ground if all of the water rained out of the atmosphere. Similar to relative humidity and dew point depression, more water in the atmosphere makes it more difficult for the sensor to accurately measure the radiance due to temperature leaving the ground; theoretically a larger column water vapor would lead to a less accurate predicted land surface temperature.

G.2   Initial Cloud Detection

The attempts at confidence metric estimation described below were performed with both a smaller dataset and before the cloud analysis described in Section 5.3.2. Therefore, a simpler, initial cloud detector was implemented. If the retrieved temperature is less than 275 K or if the retrieved temperature is more than 15 K below the air temperature of the lowest atmospheric layer in the profile, the pixel is classified as cloudy. These pixels are not used in the regressions described in Section G.3 and are also segmented as a separate category for the threshold analysis in Section G.4.

From our visual investigation of the results, this is a conservative cloud detector. While it may not classify all clouds as clouds, it has a very low instance of classifying non-clouds as clouds. Our main goal with this detector was to eliminate the most obvious clouds, the largest negative errors in the error histograms, from our confidence metric analysis, prior to a more in depth cloud analysis and the availability of a cloud detection algorithm or mask.

G.3   Regression Analysis

As described above, it is expected that the accuracy of the land surface temperature retrieval varies with the five chosen metrics. The mathematical model of this variability is unknown. Therefore, as an initial quantitative estimate, a linear regression was performed at each location individually and all locations together for each metric and the absolute values of the error from Equation 5.1. We will refer to this as the actual error. Also, multivariate linear regressions were performed for a subset of the variables and all of the variables together at each location individually and all of the locations together. While there is no particular reason to assume linear behavior, this initial analysis will give us a chance to visualize the data and determine if another form of regression would be more appropriate or if quantitative analysis is feasible. With these regression, given the actual value of a metric, we can predict the error in the land surface temperature at that pixel.

Each location is considered individually to investigate how such a regression behaves with dif-

fering datasets. The goal of the regression analysis is to be able to predict the possible error (in degrees kelvin) associated with the predicted LST value at each pixel. If we were to implement this method of error analysis, we would need a large database of ground truth and predicted temperatures from sufficiently variable conditions to build our regression models. Considering the linear regression model at individual locations gives us a better idea of the variability with location and the type and volume of data we would need to build such models.

The following sections summarize the regression analysis results for each metric at each location individually and all locations together. There are three methods presented to analyze the performance of our regression analysis.

Firstly, we can consider the shape of the data and if it appears to follow the linear relationship used to model it. In the following sections, for each metric for each location individually and all locations together, the error is plotted as a function of the metric and the linear regression model is shown on the same plot. This is one method of visual analysis.

$$\text{residual} = \text{ABS}(\text{Predicted Error} - \text{Actual Error}) \tag{G.1}$$

Secondly, we calculate the residual as shown in Equation G.1. This is the difference between the error predicted by the linear regression model and the absolute value of the error between the predicted land surface temperature and buoy temperature, which we refer to as the actual error. Smaller residuals indicate our regression analysis is accurately predicting the error associated with a pixel. In a perfect linear relationship, the residual would be zero for all scenes. However, we suspect the relationship may be linear only over some range of errors. Because they are the most difficult for us to understand, we are particularly concerned with the moderate errors, 5 K to 10 K, and if we can determine the accuracy of the land surface temperatures retrieved at these pixels. We perform a simple qualitative visual analysis by plotting the residual against the error for each point.

Finally, we calculate the mean and standard deviation of the residuals for each metric at each location. This is a quantitative analysis of the results of the regression analysis.

G.3.1 Transmission in Regression Analysis

Figures G.1, G.3, G.5, G.7, and G.9 show the transmission plotted as a function of the actual error for the Salton Sea, Lake Tahoe, Rochester, Delmar, and Georgia respectively. Also on each plot is the linear regression model. Figure G.11 shows the same plot with all locations and the line of best fit. All plots are shown with the same transmission and error ranges. Figures G.2, G.4, G.6, G.8, and G.10 show the residual computed as shown in Equation G.1 plotted against the actual error calculated as shown in Equation 5.1. All plots are shown with the same error and residual range. Note that these ranges were chosen to include most points and maintain some degree of detail. Some outliers may not be shown on the plot. The mean and standard deviation of the residuals for each location are summarized in Table G.1.
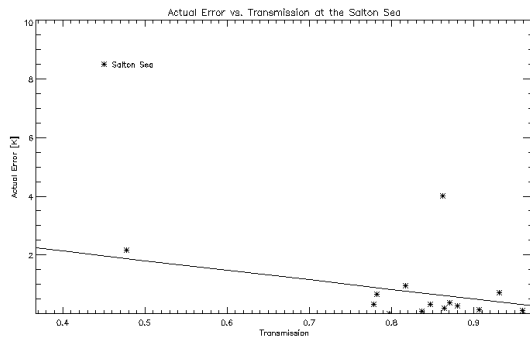
Figure G.1: Actual error vs. transmission for the Salton Sea with the line of best fit.
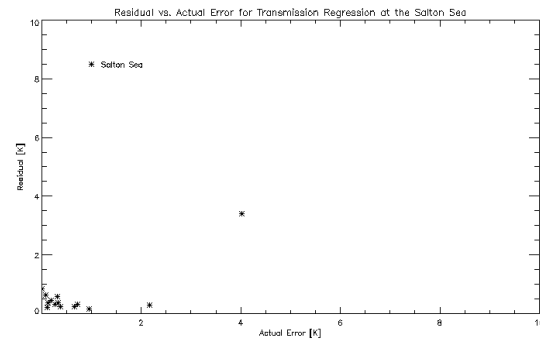
Figure G.2: Residual vs. actual error for transmission regression at the Salton Sea.
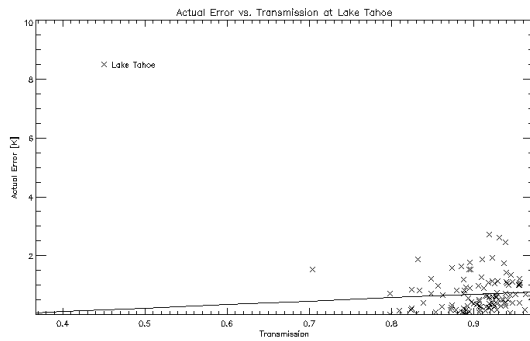
Figure G.3: Actual error vs. transmission for Lake Tahoe with the line of best fit.
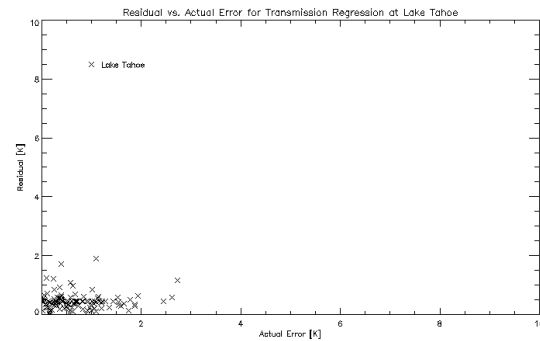
Figure G.4: Residual vs. actual error for transmission regression at Lake Tahoe.

Because the scenes over the Salton Sea and Lake Tahoe are well behaved, they can be modeled with a linear regression with relatively low residuals. Looking at Figures G.1 and G.3, the data is not necessarily linearly behaved but all have relatively high transmission values and the initial range of errors is small, so the residuals are small. Most points fit the linear model with only a handful of anomalies as shown in Figures G.2 and G.4. Notice that the Salton Sea regression has a positive slope while the Lake Tahoe regression has a negative slope, indicating that although the residuals are low, the data is not behaving as expected. We expect the error to decrease as the transmission increases, resulting in a negative gradient.
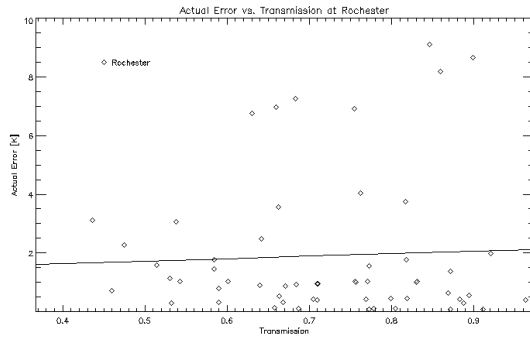
Figure G.5: Actual error vs. transmission for Rochester with the line of best fit.
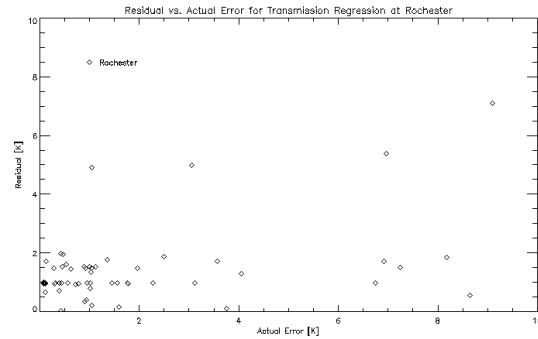
Figure G.6: Residual vs. actual error for transmission regression at Rochester.
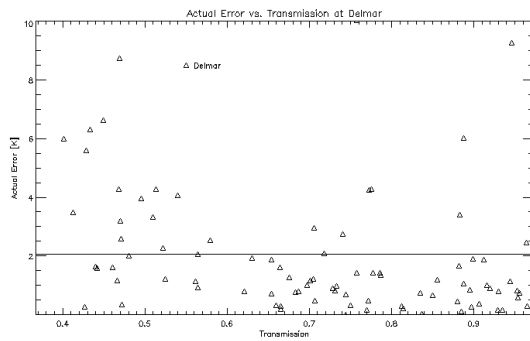
Figure G.7: Actual error vs. transmission for Delmar with the line of best fit.
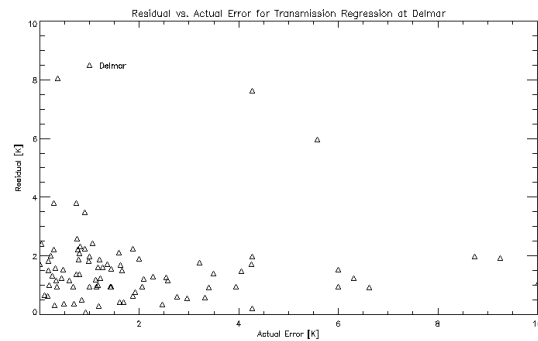
Figure G.8: Residual vs. actual error for transmission regression at Delmar.
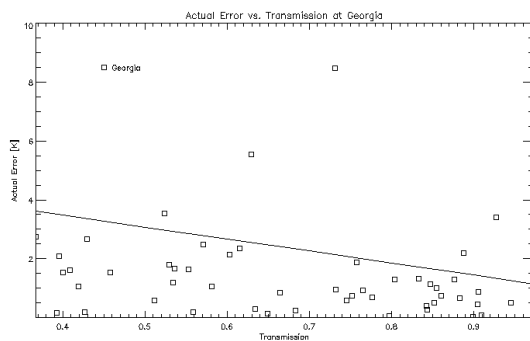
Figure G.9: Actual error vs. transmission for Georgia with the line of best fit.
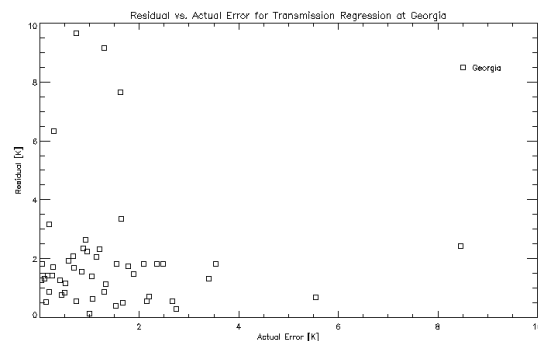
Figure G.10: Residual vs. actual error for transmission regression at Georgia.

Scenes over Rochester, Delmar, and Georgia are not as well-behaved, as shown in Figures G.5, G.7, and G.9. The range of transmission is larger for these three sites and the range of errors is also much larger. Visual analysis shows that these relationships are not linear, and this is also reflected in

the residuals in Figures G.6, G.8, and G.10. The regressions result in a positive, almost constant, and negative gradient for Rochester, Delmar and Georgia respectively, indicating that the data cannot be modeled as decreasing error with increasing transmission. Although the relationships do not appear to be linear, it is not easy to visually identify another model that would accurately represent the data.
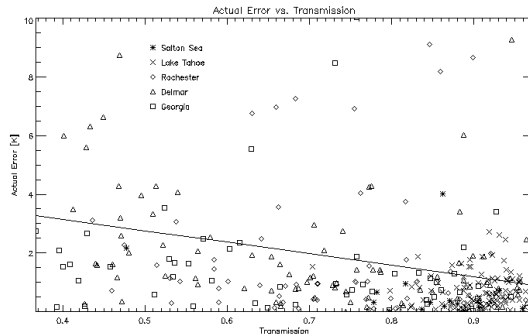
Figure G.11: Actual error vs. transmission with the line of best fit for all locations.
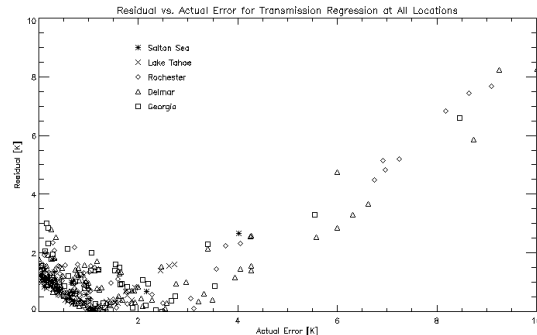
Figure G.12: Residual vs. actual error for transmission regression at all locations.

Finally, considering the linear regression for all locations together in Figure G.11, the data does not appear to be linear and this is reflected in Figure G.12. Because of the large number of scenes in the Lake Tahoe data set, and the relatively large proportion of scenes with errors between 1 K and 2 K, the linear regression model is largely influenced by these points and scenes with errors between 1 K and 2 K have the minimum residuals. Many scenes with errors between 5 K and 10 K have unacceptably high residuals. This is also reflected in Table G.1. The means and standard deviations of the residuals are low for Salton Sea and Lake Tahoe; although the means are not excessively large for the rest of the locations, the standard deviations are considerably larger. This indicates that the mean is influenced by the large portion of scenes with low errors and correspondingly low residuals; these have the largest influence on the linear regression model, but there are a handful of scenes with larger errors, and larger residuals, as indicated by the standard deviation. This does not give us confidence that our confidence metric will identify and predict scenes with moderate amounts of error.

Table G.1: Mean and standard deviation of residuals for transmission regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.595	0.826
Lake Tahoe	0.449	0.351
Rochester	1.746	1.618
Delmar	2.085	2.359
Georgia	2.177	2.609
All Locations	1.406	2.025

G.3.2 Relative Humidity in Regression Analysis

Figures G.13, G.15, G.17, G.19, and G.21 show the errors plotted as a function of relative humidity for each location individually with the linear regression model for that location included on the plot. Figures G.14, G.16, G.18, G.20, and G.22 show the residuals plotted against the error for these linear regression models. Figures G.23 and G.24 show the same for all locations together. The means and standard deviations of the residuals are summarized in Table G.2.
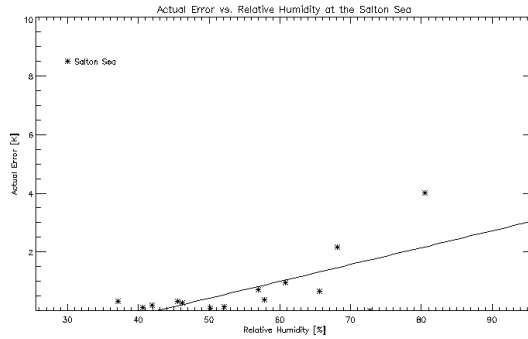
Figure G.13: Actual error vs. relative humidity for the Salton Sea with the line of best fit.
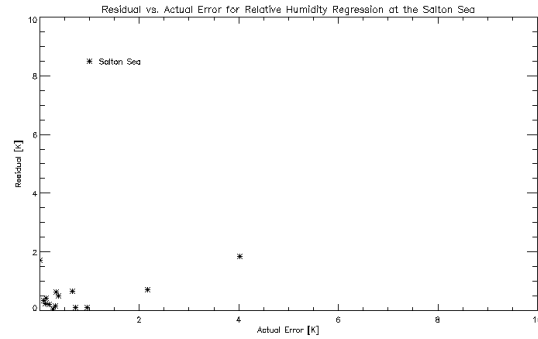
Figure G.14: Residual vs. actual error for relative humidity regression at the Salton Sea.
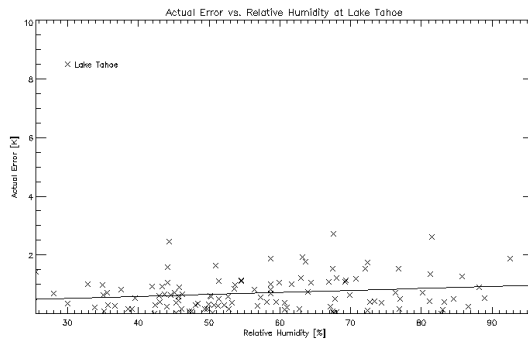
Figure G.15: Actual error vs. relative humidity for Lake Tahoe with the line of best fit.
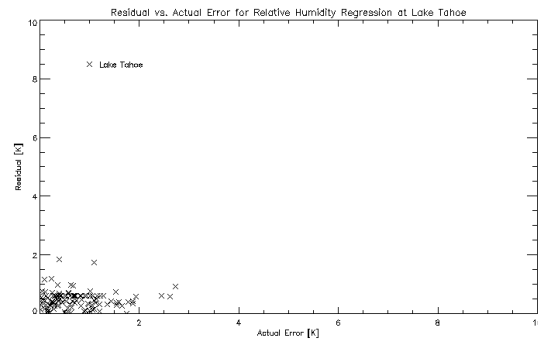
Figure G.16: Residual vs. actual error for relative humidity regression at Lake Tahoe.

Similar to the models for transmission, Salton Sea and Lake Tahoe provide the best results. Considering Figures G.13 and G.15 more closely, even though the scenes were chosen for good retrieval conditions, there is a large range of relative humidities present, and they appear to behave linearly with error. Although the range of errors is still small, there is a visual trend of increasing error with increasing relative humidity. This is reinforced by the residuals shown in Figures G.14 and G.16. Particularly at Lake Tahoe, residuals do not increase with error, as in many other cases, indicating that the linear relationship more accurately represents all scenes and is not only true over a small range of data.
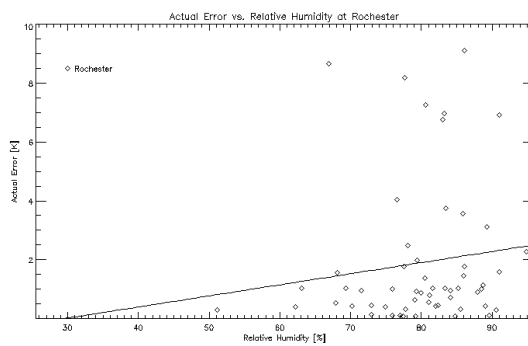
Figure G.17: Actual error vs. relative humidity for Rochester with the line of best fit.
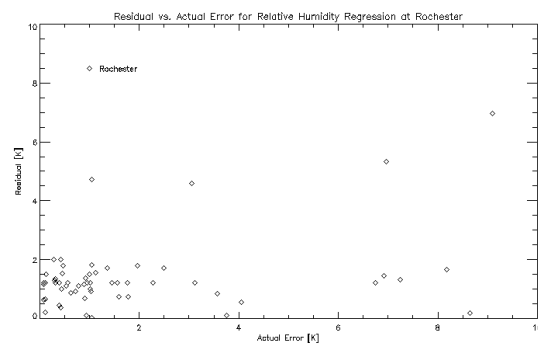
Figure G.18: Residual vs. actual error for relative humidity regression at Rochester.
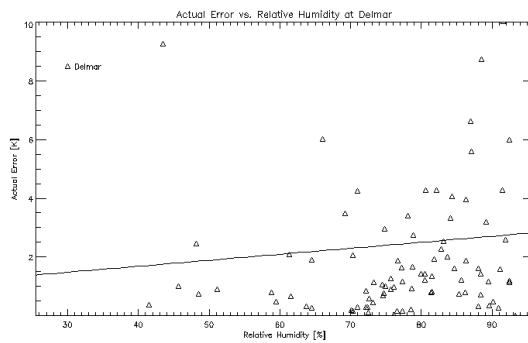
Figure G.19: Actual error vs. relative humidity for Delmar with the line of best fit.
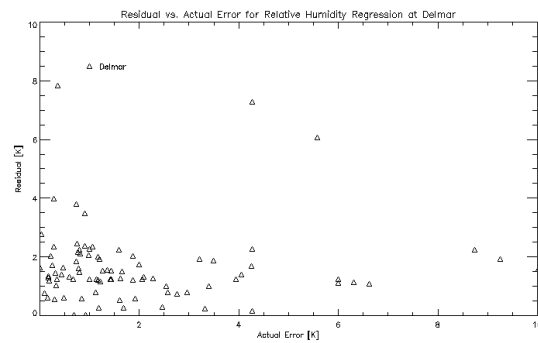
Figure G.20: Residual vs. actual error for relative humidity regression at Delmar.
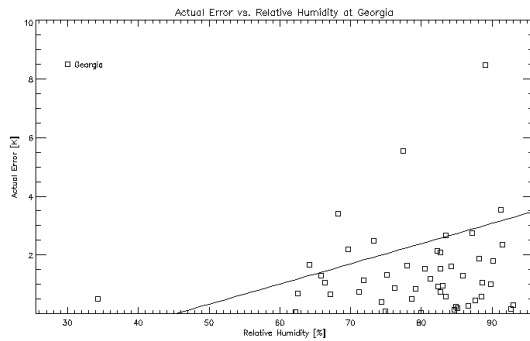
Figure G.21: Actual error vs. relative humidity for Georgia with the line of best fit.
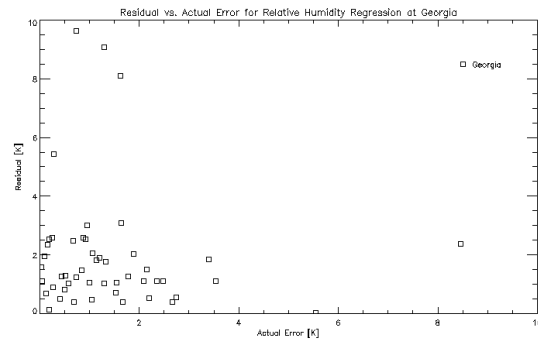
Figure G.22: Residual vs. actual error for relative humidity regression at Georgia.

Although the results for Rochester, Delmar and Georgia in Figures G.17, G.19, and G.21 appear to be more linearly behaved than transmission, and all have positive gradients as expected, there are a still number of anomalies that indicate that these would not model the data accurately enough for good error prediction. At all three locations, but particularly in Georgia, there are larger residuals even when the actual error is small in Figures G.18, G.20, and G.22.
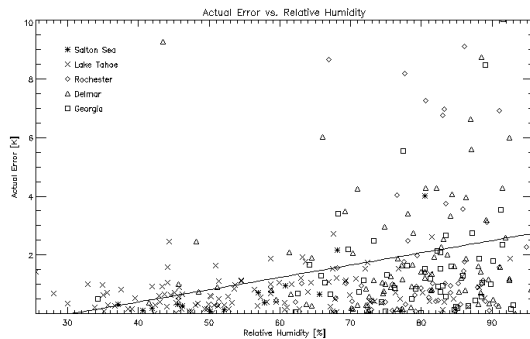
Figure G.23: Actual error vs. relative humidity with the line of best fit for all locations.
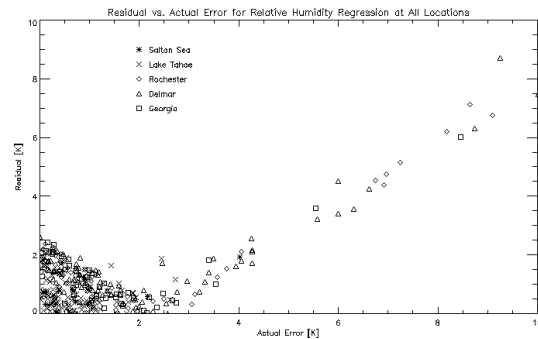
Figure G.24: Residual vs. actual error for relative humidity regression at all locations.

The results for the linear regression model for all locations are very similar to those for transmission. Largely influenced by the large number of scenes with 1 K to 2 K error, the minimum in the residuals is at this point. Larger residuals at the smallest errors and moderate errors are concerning. It appears that with this model we can only accurately predict the error over a small range. This is reinforced by the means and standard deviations as shown in Table G.2. Salton Sea and Lake Tahoe both have small means and standard deviations; the means for all other locations could be acceptably small but the larger standard deviations, as supported by the plots, indicate that there are a number of scenes that are not accurately modeled.

Table G.2: Mean and standard deviation of residuals for relative humidity regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.549	0.567
Lake Tahoe	0.444	0.347
Rochester	0.347	1.681
Delmar	1.660	2.109
Georgia	2.224	2.569
All Locations	1.439	1.981

G.3.3   Maximum Temperature in Regression Analysis

Actual error is plotted versus maximum temperature for the Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia in Figures G.25, G.27, G.29, G.31, and G.33 respectively with the linear regression models shown on the plots. Correspondingly, the residuals are plotted versus the actual error for the same locations in Figures G.26, G.28, G.30, G.32, and G.34. The same is shown for all locations in Figures G.35 and G.36.
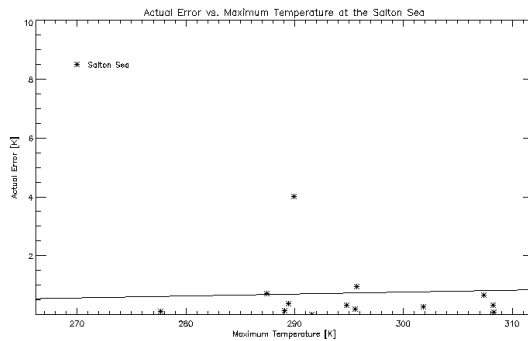
Figure G.25: Actual error vs. maximum temperature for the Salton Sea with the line of best fit.
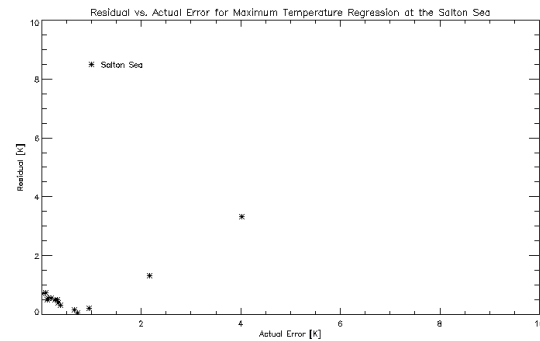
Figure G.26: Residual vs. actual error for maximum temperature regression at the Salton Sea.
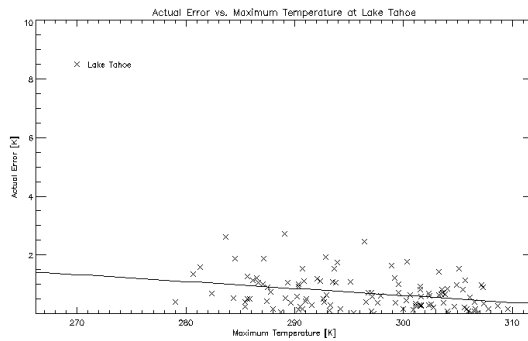
Figure G.27: Actual error vs. maximum temperature for Lake Tahoe with the line of best fit.
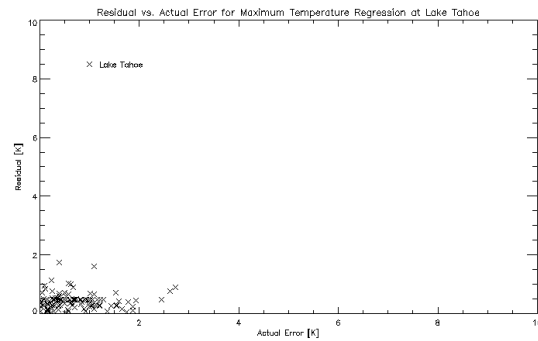
Figure G.28: Residual vs. actual error for maximum temperature regression at Lake Tahoe.

As with transmission, the scenes for the Salton Sea and Lake Tahoe have tendencies of linear behavior with maximum temperature and such a linear model would be reasonable for either single set data as shown in Figures G.25 and G.27. However, the slopes of the linear models are slight, with a slight positive gradient in the Salton Sea regression and a slight negative gradient in the Lake Tahoe regression. We would expect the error to increase with maximum temperature, unlike the trend shown in the Lake Tahoe data. With only one anomaly at the Salton Sea, all other residuals at both sites are less than 2 K in Figures G.26 and G.28. It is important to note that most actual errors for all scenes at both sites are below 4 K. Larger residuals are less likely with a smaller error range. Regardless of the accuracy of our linear regression, predicting any small error will result in a small residual, which is not helpful for this analysis of our error prediction.
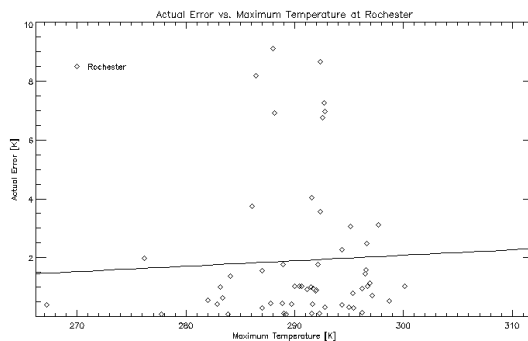
Figure G.29: Actual error vs. maximum temperature for Rochester with the line of best fit.
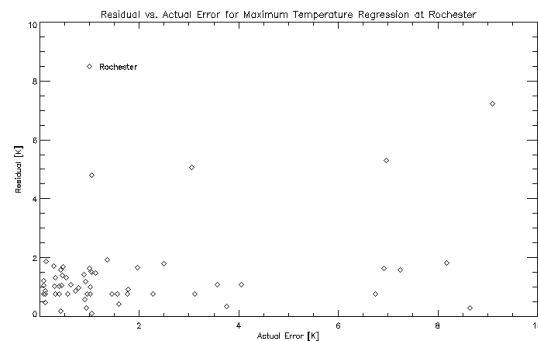
Figure G.30: Residual vs. actual error for maximum temperature regression at Rochester.
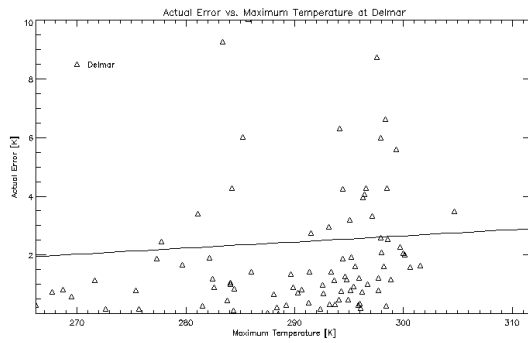
Figure G.31: Actual error vs. maximum temperature for Delmar with the line of best fit.
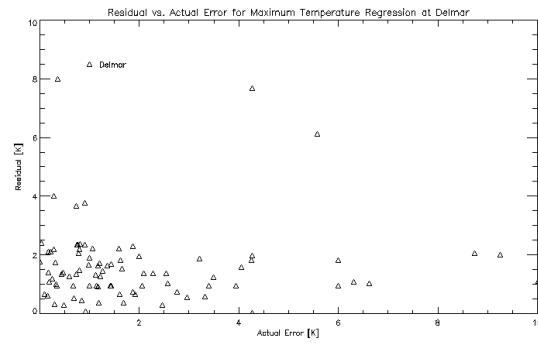
Figure G.32: Residual vs. actual error for maximum temperature regression at Delmar.
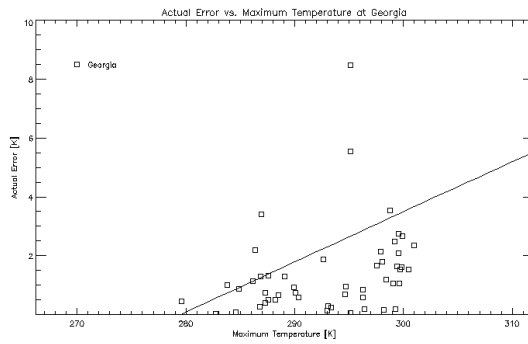
Figure G.33: Actual error vs. maximum temperature for Georgia with the line of best fit.
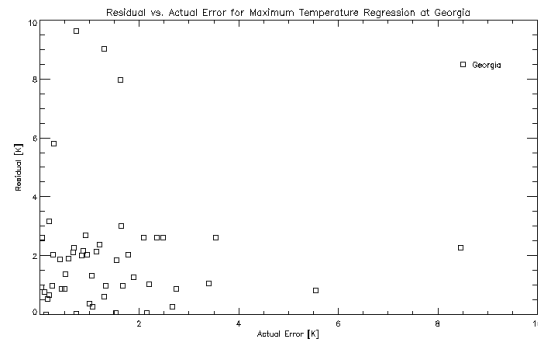
Figure G.34: Residual vs. actual error for maximum temperature regression at Georgia.

The data for maximum temperature at Rochester, Delmar, and Georgia do not have linear tendencies as a group as shown in Figures G.29, G.31, and G.33; all three regressions have positive gradients as expected, but there are enough anomalies to make the linear models unrealistic for predicting error for all scenes. Unlike some of the other metrics, where the model is decidedly more accurate over a small range of errors, regressions with maximum temperature at these three locations have a variety of residuals over the whole range shown in Figures G.30, G.32, and G.34.
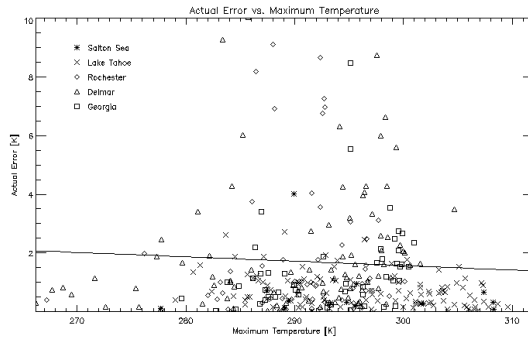
Figure G.35: Actual error vs. maximum temperature with the line of best fit for all locations.
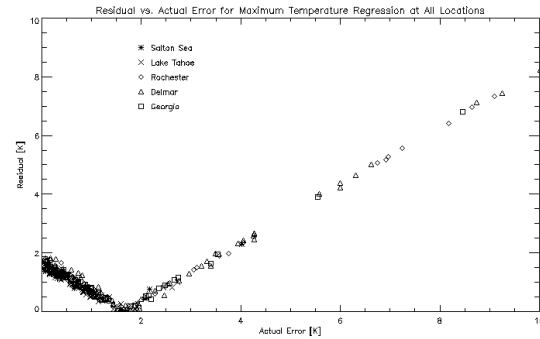
Figure G.36: Residual vs. actual error for maximum temperature regression at all locations.

Figure G.35 shows that all locations plotted together does not result in a linear data set. This can be observed visually and also by the linear regression model with an almost constant slope; in fact, the regression is influenced so much by the Lake Tahoe scenes that it has a slightly negative slope although we expect error to increase with increasing maximum temperature as it did with all individual locations except Lake Tahoe. This is further illustrated in Figure G.36; the pattern in the residuals shows a distinct minimum just below an error of 2 K, which is approximately equal to the almost constant value of the linear regression. Because of the constant nature of the model, there are larger residuals at error values smaller and larger than this constant. This suggests a linear model would not accurately predict errors for all of the locations, as supported by Table G.3. There are lower means and standard deviations for both the Salton Sea and Lake Tahoe and larger means and much larger standard deviations at all other locations, indicating that the linear regression models do not accurately predict the error for all points.

Table G.3: Mean and standard deviation of residuals for maximum temperature regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.702	0.816
Lake Tahoe	0.426	0.337
Rochester	1.733	1.632
Delmar	2.110	2.342
Georgia	2.143	2.553
All Locations	1.529	2.031

G.3.4 Dew Point Depression in Regression Analysis

Figures G.37, G.39, G.41, G.43 and G.45 show the actual error plotted against the dew point depression with the linear models and Figures G.38, G.40, G.42, G.44, and G.46 illustrate the corresponding residuals. Figures G.47 and G.48 show the same for all locations.
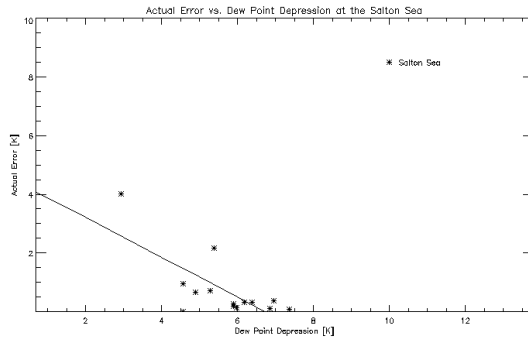
Figure G.37: Actual error vs. dew point depression for the Salton Sea with the line of best fit.
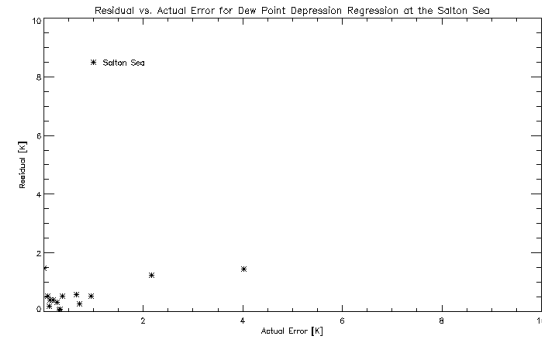
Figure G.38: Residual vs. actual error for dew point depression regression at the Salton Sea.
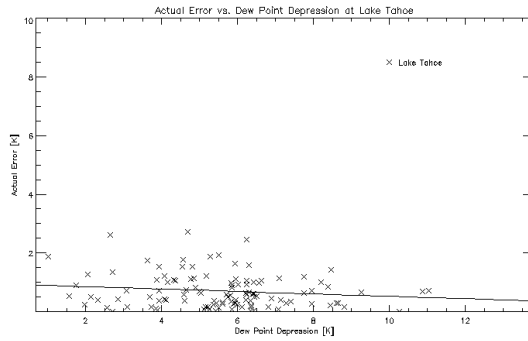
Figure G.39: Actual error vs. dew point depression for Lake Tahoe with the line of best fit.
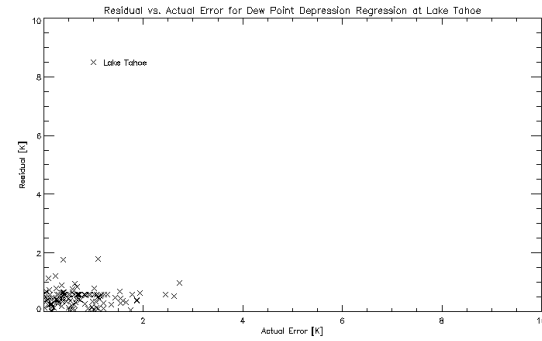
Figure G.40: Residual vs. actual error for dew point depression regression at Lake Tahoe.

Results with the dew point depression metric for the Salton Sea and Lake Tahoe in Figures G.37 and G.39 look similar to, but inverted from the relative humidity results. This is expected as the dew point depression is derived from and related to the relative humidity. It is considered as a different method of confidence metric determination only to investigate if another relationship provides more clarity but they appear to give redundant results.
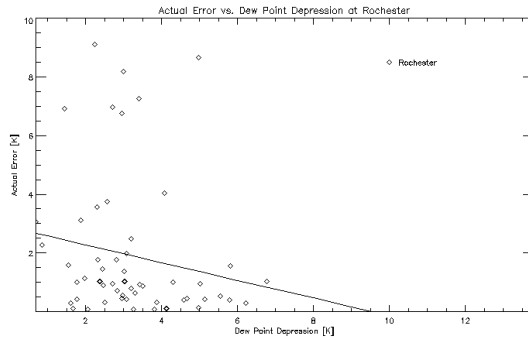
Figure G.41: Actual error vs. dew point depression for Rochester with the line of best fit.
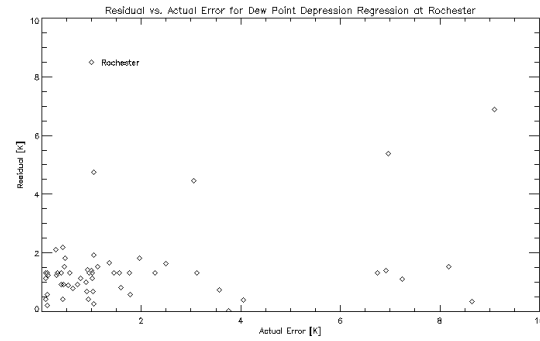
Figure G.42: Residual vs. actual error for dew point depression regression at Rochester.
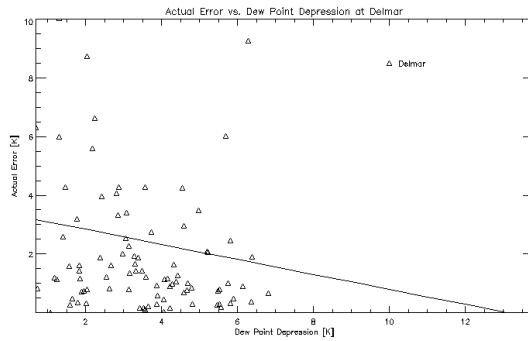
Figure G.43: Actual error vs. dew point depression for Delmar with the line of best fit.
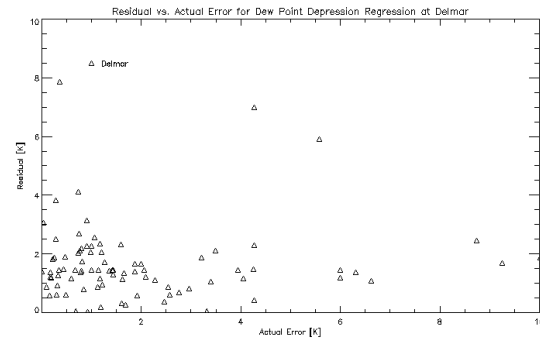
Figure G.44: Residual vs. actual error for dew point depression regression at Delmar.
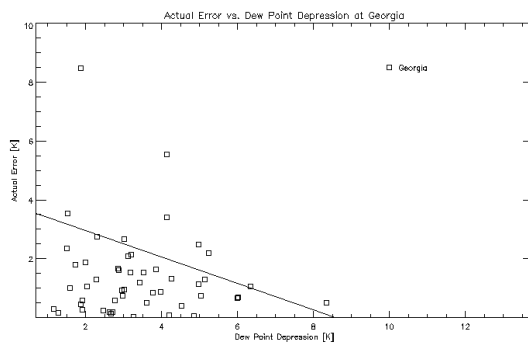
Figure G.45: Actual error vs. dew point depression for Georgia with the line of best fit.
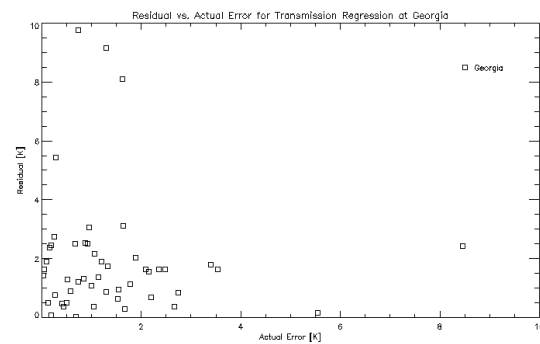
Figure G.46: Residual vs. actual error for dew point depression regression at Georgia.

Actual error plotted as a function of dew point depression in Figures G.41, G.43, and G.45 for Rochester, Delmar, and Georgia respectively also look similarly inverted from the relative humidity results at these locations. All have negative gradients as expected but the data does not appear

to behave linearly and this is confirmed by the residuals. There are large residuals for varying magnitudes of actual error in Figures G.42, G.44, and G.46.
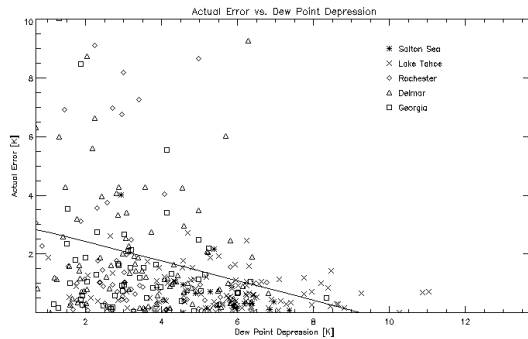
Figure G.47: Actual error vs. dew point depression with the line of best fit for all locations.
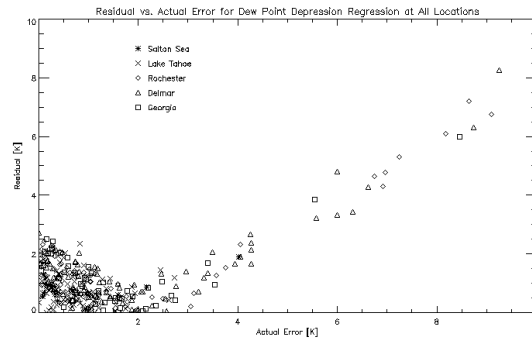
Figure G.48: Residual vs. actual error for dew point depression regression at all locations.

The plot for all locations together shown in Figure G.47 is inverted from the relative humidity plot shown in Figure G.23. As a general pattern, actual error decreases as dew point depression increases as expected and in Figure G.48 the residuals from the dew point depression regression are very similar to the residuals for the relative humidity regression. Residuals are smallest for scenes with errors between 1 K and 2 K due to the large number of scenes at these actual error values that largely influence the regression. However, residuals are larger for smaller errors and much larger for larger errors as shown in Figure G.48. This is concerning for predicting confidence in our best and worst results. Results for the means and standard deviations of the residuals for dew point depression, summarized in Table G.4, are very similar to those for relative humidity, shown in Table G.2. It does not appear, as we were hoping, that one of these metrics provides a different perspective on the error analysis and neither sufficiently predicts error for the entire dataset.

Table G.4: Mean and standard deviation of residuals for dew point depression regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.572	0.473
Lake Tahoe	0.451	0.343
Rochester	1.675	1.647
Delmar	2.090	2.334
Georgia	2.201	2.603
All Locations	1.460	1.975

G.3.5   Column Water Vapor in Regression Analysis

Figures G.49, G.51, G.53, G.55, and G.57 show the actual errors plotted as a function of the column water vapor for Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia respectively; the corresponding residuals are shown in Figures G.50, G.52, G.54, G.56, and G.58. The same is shown for all locations in Figures G.59 and G.60. Although column water vapor is also a measure of humidity, it is derived differently than the relative humidity and dew point depression, so it can provide new insight in our confidence metric determination.
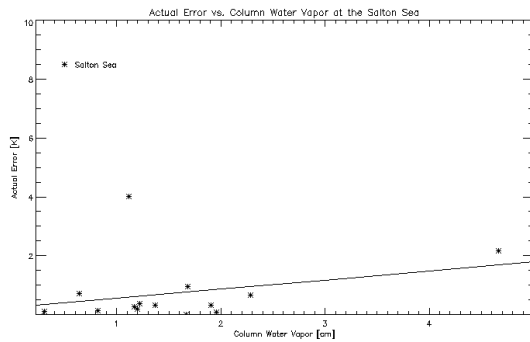
Figure G.49: Actual error vs. column water vapor for the Salton Sea with the line of best fit.
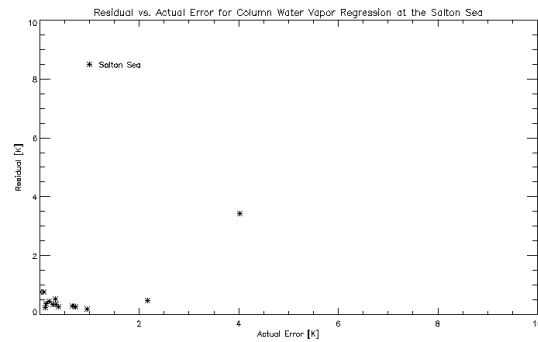
Figure G.50: Residual vs. actual error for column water vapor regression at the Salton Sea.
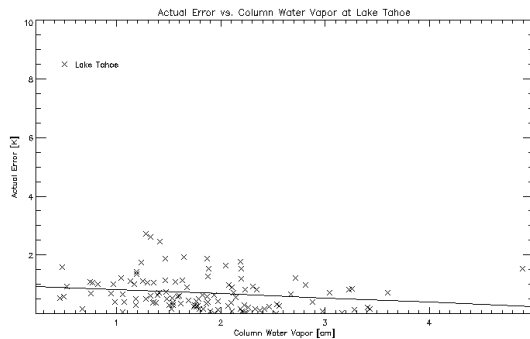
Figure G.51: Actual error vs. column water vapor for Lake Tahoe with the line of best fit.
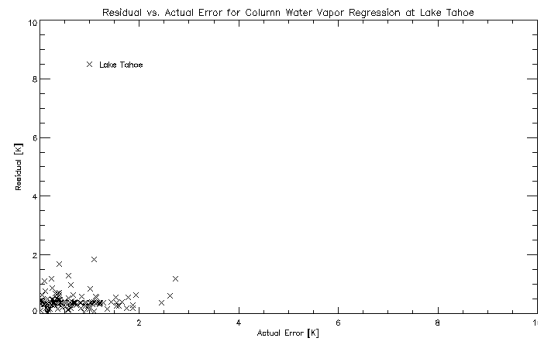
Figure G.52: Residual vs. actual error for column water vapor regression at Lake Tahoe.

Regression results for column water vapor look most similar to those for transmission. Most of the results are well-behaved but do not appear linear as shown in Figures G.49 and G.51. The Salton Sea regression has a positive gradient as expected because errors should increase with column water vapor, but the Lake Tahoe regression has a slight negative gradient. All actual errors and corresponding residuals are small, as shown in Figures G.49 and G.51, but the data does not follow the linearity relationship as expected or desired for error prediction.
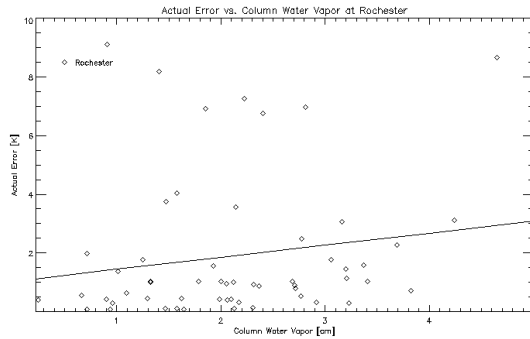
Figure G.53: Actual error vs. column water vapor for Rochester with the line of best fit.
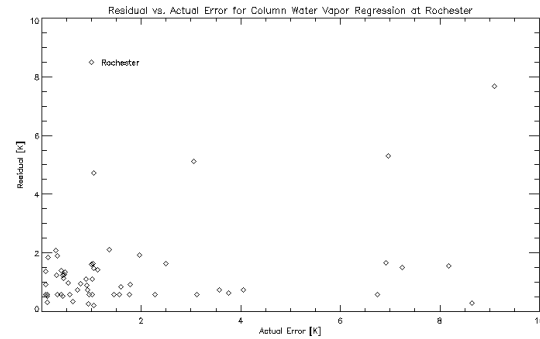
Figure G.54: Residual vs. actual error for column water vapor regression at Rochester.
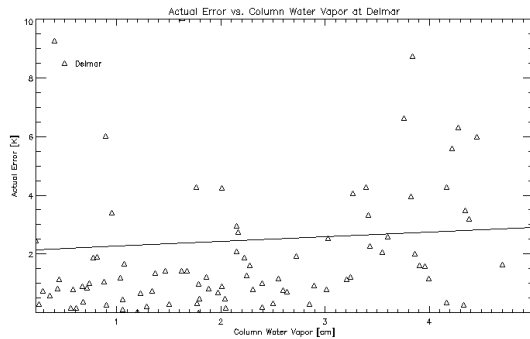
Figure G.55: Actual error vs. column water vapor for Delmar with the line of best fit.
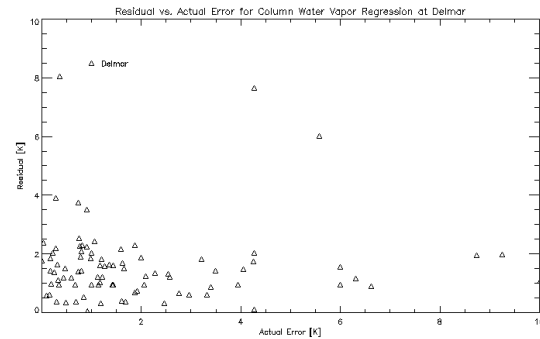
Figure G.56: Residual vs. actual error for column water vapor regression at Delmar.
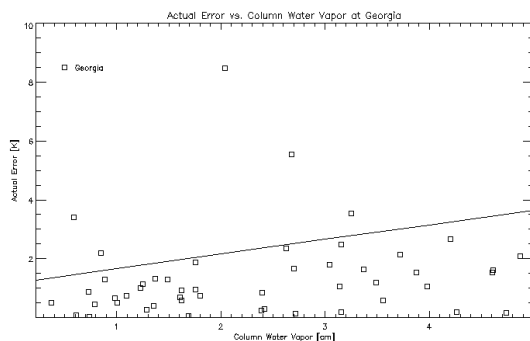
Figure G.57: Actual error vs. column water vapor for Georgia with the line of best fit.
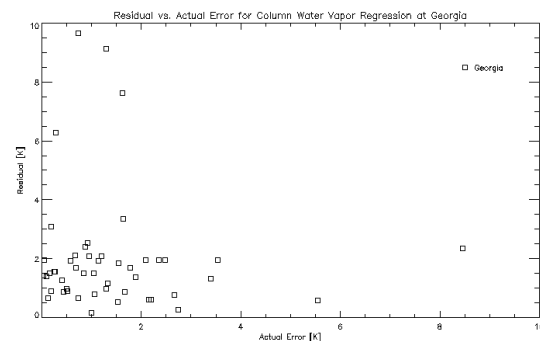
Figure G.58: Residual vs. actual error for column water vapor regression at Georgia.

Actual error is plotted as a function of column water vapor for Rochester, Delmar and Georgia in Figures G.53, G.55, and G.57. Also similar to transmission at these locations, these results all have regressions with positive gradients as expected, but the data is not linear nor well-behaved,

illustrated by the residuals in Figures G.54, G.56, and G.58. There is a large range of column water vapor values and a varied range of residual magnitudes over the entire range of actual error, rather than increasing residuals with increasing error as seen, although not desired, with some other metrics. These data do not appear to be modeled well with linear regressions and there is not another visually obvious model for the data.
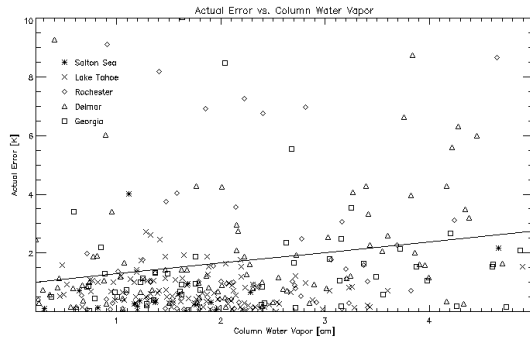
Figure G.59: Actual error vs. column water vapor with the line of best fit for all locations.
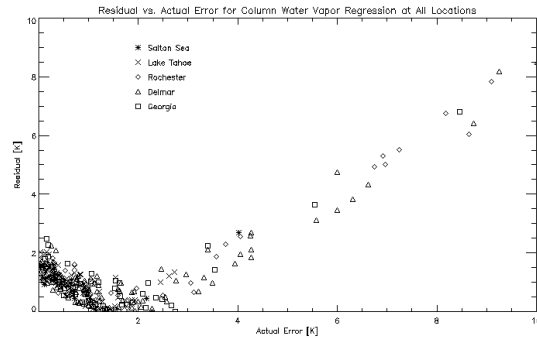
Figure G.60: Residual vs. actual error for column water vapor regression at all locations.

The regression model for all locations together is shown in Figure G.59 with the corresponding residuals shown in Figure G.60. Although the regression has a positive gradient as expected, similar to outcomes for most other metrics, the residuals are at a minimum for scenes with 1 K to 2 K error, indicating that the regression model is influenced by the large number of scenes with this range. Higher residuals for the smallest and largest errors are concerning. Table G.5 shows that although the means are reasonable, the higher standard deviations at Rochester, Delmar, Georgia and all locations together indicate that while the error for average scenes may be well predicted, errors for scenes with the smallest and largest errors are not predicted well.

Note that in Figure G.59, if we were to draw a vertical line at 3 cm and give a high confidence to anything below 3 cm and a low confidence to any result with a total column water vapor above 3 cm, we capture many of the poor performers in our low confidence category without eliminating very many good results. This is one indication we may have more success with a thresholding method as we will discuss in Section G.4.

Table G.5: Mean and standard deviation of residuals for column water vapor regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.622	0.826
Lake Tahoe	0.446	0.343
Rochester	1.692	1.632
Delmar	2.089	2.358
Georgia	2.187	2.617
All Locations	1.464	2.045

G.3.6 Three Metrics in Regression Analysis

Attempting to capture the effects of multiple metrics, a multivariate linear regression was performed with three of the five metrics. Because relative humidity, dew point depression, and column water vapor are all measures of humidity, a multivariate linear regression was performed with transmission, maximum temperature, and relative humidity. The goal is to capture the behavior of each metric and reduce the influence of anomalies in any one metric. Because the relationship between three metrics and the actual error is considered, plots like we have used to analyze all other metrics are four-dimensional and therefore cannot be visualized; the residuals are plotted for each location individually in Figures G.61, G.62, G.63, G.64, and G.65 respectively.
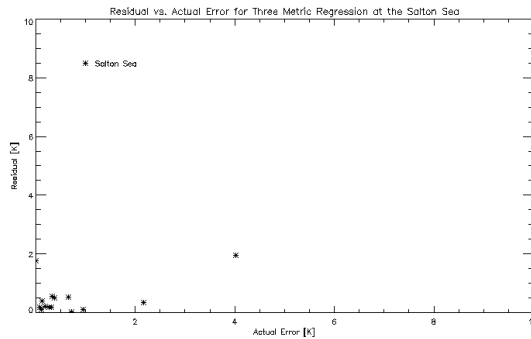
Figure G.61: Residual vs. actual error for three metric regression at the Salton Sea.
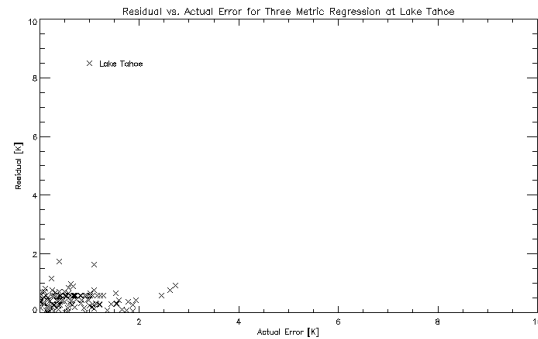
Figure G.62: Residual vs. actual error for three metric regression at Lake Tahoe.

Results for the Salton Sea and Lake Tahoe are difficult to analyze without being able to visualize the data. Because the data is so well-behaved for the samples that we have at these two sites, the error range is small and the residuals are small, shown in Figures G.61 and G.62, as they have been for most other regressions at these two sites. We are unable to conclude whether this is only due to small actual errors or more accurate error prediction.
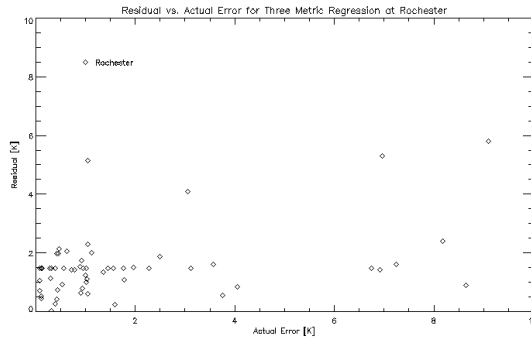
Figure G.63: Residual vs. actual error for three metric regression at Rochester.
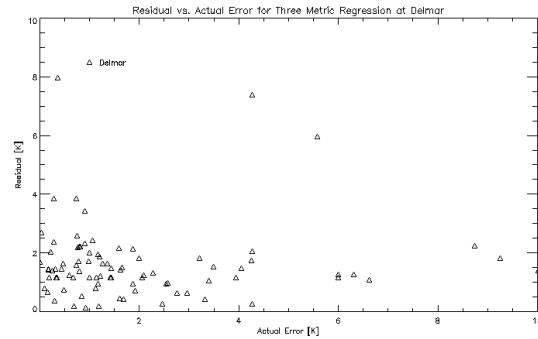
Figure G.64: Residual vs. actual error for three metric regression at Delmar.
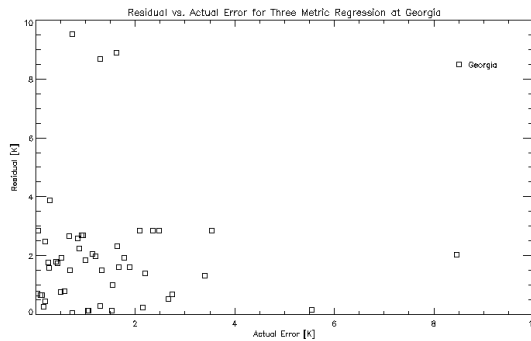
Figure G.65: Residual vs. actual error for three metric regression at Georgia.
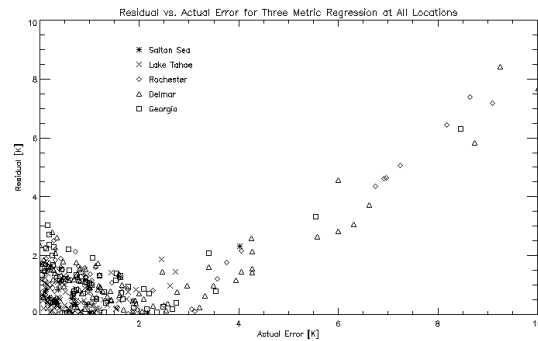
Figure G.66: Residual vs. actual error for three metric regression at all locations.

Residuals at Rochester, Delmar and Georgia, shown in Figures G.63, G.64, and G.65 are more informative than residuals at the Salton Sea and Lake Tahoe. The residuals are not excessively large and do not appear to have any particular pattern of behavior with error; that is, there are large and small residuals throughout the range of errors at all three locations. Most importantly, in a visual comparison to the single metric regressions, there is not a significant improvement in residuals.

The residuals for the three metric regression at all locations in Figure G.66 are also very similar to the patterns seen in the single metric regressions. Although there are more low residuals at the lowest errors, there is still a trend of the lowest residuals in the 1 K to 2 K error range and increasing residuals with increasing error. The regression is likely most influenced by the large set of scenes with 1 K to 2 K actual error, with the reoccurring theme of difficulty in predicting errors for the smallest and largest errors. This is reiterated by the means and standard deviations of the residuals as shown in Table G.6, very similar to most single metric results. The small means but larger standard deviations indicate that while the residuals are centered around a reasonable value, there is difficulty in predicting error for scenes different from the average.

Table G.6: Mean and standard deviation of residuals for three metric regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.503	0.597
Lake Tahoe	0.422	0.337
Rochester	1.736	1.474
Delmar	2.096	2.344
Georgia	2.147	2.400
All Locations	1.392	1.992

G.3.7 All Metrics in Regression Analysis

Finally, a linear regression was performed using all metrics. We realize it is shown above that relative humidity and dew point depression provide redundant information, and column water is also a measure of humidity, but a multivariate linear regression with all five metrics was calculated in an attempt to utilize all possible information. Because the data is six-dimensional, only the residuals can be plotted; this is shown in Figures G.67, G.68, G.69, G.70, and G.71 for each individual location.
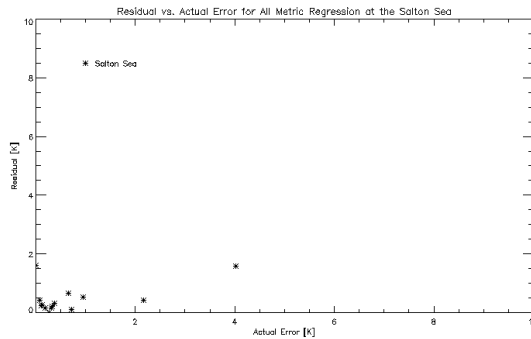
Figure G.67: Residual vs. actual error for all metric regression at the Salton Sea.
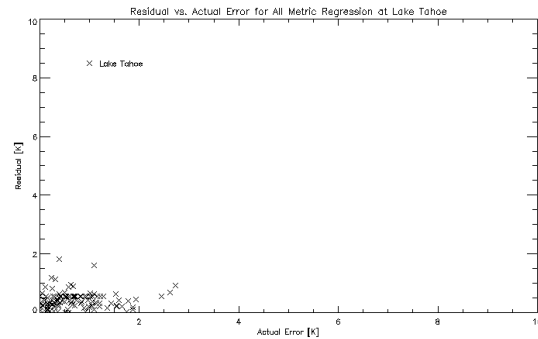
Figure G.68: Residual vs. actual error for all metric regression at Lake Tahoe.

As with the other regressions, data at the Salton Sea and Lake Tahoe are well-behaved in our sample set and it is difficult to analyze because the actual error range and the residuals are small as shown in Figures G.67 and G.68, so this tells us little about the abilities of the regression to predict error.
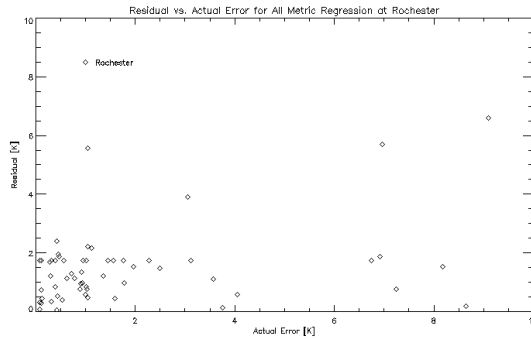
Figure G.69: Residual vs. actual error for all metric regression at Rochester.
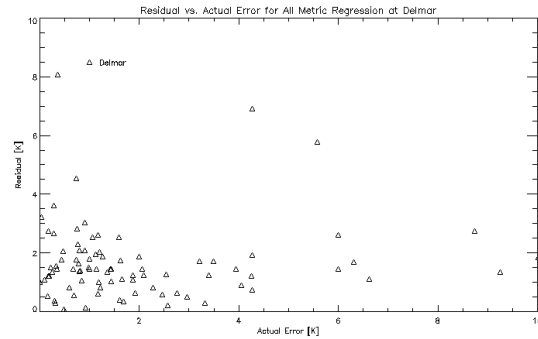
Figure G.70: Residual vs. actual error for all metric regression at Delmar.
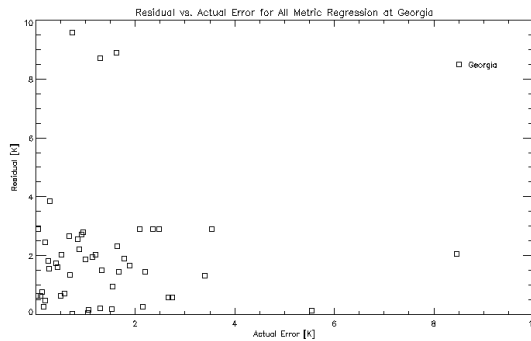
Figure G.71: Residual vs. actual error for all metric regression at Georgia.
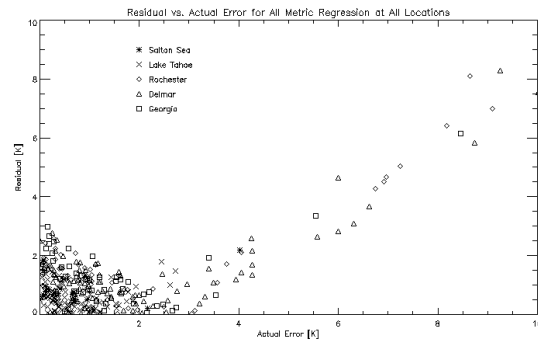
Figure G.72: Residual vs. actual error for all metric regression at all locations.

Residuals for Rochester, Delmar and Georgia in Figures G.69, G.70, and G.71 are, as expected, very similar to those for the three metric regression. It does not appear that the two additional metrics have any added value. There are anomalies and varying residuals throughout the range of actual errors, indicating that we cannot accurately predict error using this regression.

Finally, the residuals for the regression with all metrics at all locations is shown in Figure G.72. This looks similar the three metric regression results shown in Figure G.66. The means and standard deviations summarized in Table G.7 are also similar to the three metric regression shown in Table G.6. The all metric regression does not have more value added than the single or three metric regression. Like the others, it is largely influenced but the large number of scenes between 1 K and 2 K and does not appear to give accurate error analysis at the smallest and largest errors.

Table G.7: Mean and standard deviation of residuals for all metric regression.

Location	Mean [K]	Standard Deviation [K]
Salton Sea	0.477	0.501
Lake Tahoe	0.411	0.338
Rochester	1.501	1.474
Delmar	2.077	2.313
Georgia	2.133	2.412
All Locations	1.397	1.982

G.3.8   Concluding Remarks on Regression Analysis

Overall, regression for each metric or combination of metrics tends to yield similar results that indicate linear regression analysis is not a feasible method of confidence metric determination. We are aware that there are a number of obvious problems with our initial approach to this analysis. Perhaps most influential is the sample set we are using. Regression analysis for each individual location is useful for initial analysis but would be extremely difficult to implement operationally. This would require a set of truth data for each region to develop models for individual locations. The results for regressions including all locations show an obvious bias toward groups of scenes with an average amount of error. If the results were truly linear, this should not be the case. The influence by the group of average scenes is discouraging for the use of a linear regression. The selection of scenes for the regression data set would be extremely important. Finally, even if such a dataset can be built, analysis of the actual error plotted as a function of each metric does not indicate that a linear model is best for the data. Visual analysis does not lead to any obvious mathematical model for the data. Therefore, based on lack of accuracy and difficulty of operational implementation, regression analysis was eliminated as a possibility for confidence metric determination.

G.4   Threshold Analysis

Section G.1 discusses how we expect error to behave with each metric. Further investigation into a number of individual scenes with a range of errors validates these expectations. However, as shown in the plots in Section G.3, these relationship are neither linear, nor do they follow another obvious mathematical model, nor are they absolute. That is, while they can be true for most scenes, these factors are not absolute in influencing the errors in land surface temperature. Because of this, we began investigating a qualitative method of error analysis. Rather than attempting to predict the magnitude of error at every pixel, we attempt to give a qualitative evaluation of the likelihood of accuracy in each.

Using the same metrics described in Section G.1, a threshold for each metric was determined. Pixels are classified as high or low confidence based on the value of the metric for that pixel falling above or below the threshold. For example, a threshold for transmission is determined. Because we expect transmission to be inversely related to error, pixels with transmission above this threshold

are given a high confidence, meaning that we are confident in our LST prediction for this pixel, while scenes with transmission below this threshold are classified as low confidence, meaning we are unsure of our ability to accurately retrieve the land surface temperature for this pixel.

Analogous to residuals in the regression analysis, we evaluate this method of error analysis by quantifying errors of omission and errors of commission. We define an error of omission as a prediction of low confidence when the land surface temperature retrieval is sufficiently accurate and an error of commission as a prediction of high confidence when the land surface temperature is inaccurate. Percentages of the errors of omission or commission are calculated as the ratio of points categorized improperly to the total number of points in the data set. To gain an even better understanding we include the ratio of the number of points classified as errors of commission to the number of points assigned a high confidence. This provides an estimate of the percentage of points which we claim high confidence but have errors greater than the standard. We call this commission in high confidence only. Errors of omission cause us to distrust data that has acceptable errors, which results in eliminating good data. However, errors of commission would cause us to use and trust data that actually has large errors, which can cause considerably larger problems. Generally we wish to reduce errors of commission to minimize trusting data with poor results. This is repeated for each individual metric at each individual location and all locations together. Note that in regression analysis, including all locations in one analysis actually changed error predictions. In threshold analysis, the results for all locations together is simply a combination of results from each individual location, so the calculated percentages of error, means, and standard deviations may change, but the actual error prediction for any individual scene will not. The obvious flaw in this analysis is the definition of and dependence on a standard for sufficiently accurate land surface temperature retrieval and the sharp cutoff created by both the threshold and the standard.

As a first attempt, thresholds were chosen for each metric. These were chosen mostly arbitrarily based only on our experience with and manual investigations of the data. After a brief initial analysis, most of these threshold were found to be reasonable so an initial set of threshold analysis is presented below without further investigation into threshold determination. As a first look at this method of confidence metric determination, standards of 1 K and 2 K are used to estimate the errors of omission and commission (errors less than 1 K and 2 K are deemed sufficiently accurate). This needs to be further investigated in the community of users based on applications but is initially implying that users are willing to accept, and still use, 1 K or 2 K error in their LST product.

G.4.1   Transmission in Threshold Analysis

Figures G.73, G.74, G.75, G.76, and G.77 show transmission plotted as a function of actual error for the Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia respectively. Note the error range for these plots is 0 K to 5 K in order to capture the detail at the lower errors; some scenes with larger errors are not shown. Figure G.78 shows the error plotted against the transmission for all locations for an error range of 0 K to 10 K to capture a larger number of points; this conveniently also shows how many points were hidden in the previous plots. Each plot also shows

a line representing the initial transmission threshold of 0.8. Points above this line are assigned
a high confidence, as we expect higher transmission values to lead to more accurate land surface
temperature retrieval. Points below this line assigned a low confidence. Also shown in the plots are
vertical lines representing the 1 K and 2 K standards used to determine the percentage of points
that are considered errors of omission and the percentage of points that are considered errors of
commission. "Omission" and "Commission" are labeled on the plots; points in this quadrant,
above/below the threshold and left/right of the standard, are considered to be assigned inaccurate
confidence metrics. Points in the two opposing quadrants are considered to be assigned the correct
confidence. The corresponding tables, Tables G.8, G.9, G.10, G.11, and G.12, show the percentages
of points that are considered errors of omission and commission for both the 1 K and 2 K standard,
for the Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia respectively. These tables also
show the mean and standard deviation of the errors of the points at each location classified as
high confidence. Note that these values consider points that are both correctly classified as high
confidence and errors of commission. This would provide us an initial estimate of the error on
points that we assign a high confidence, and how sure we are of this error. Finally, commission in
high confidence is shown for both the 1 K and 2 K standard.

All points, regardless of the error range of the plots, are included in the calculations, unless
classified as cloudy by the cloud detection described in Section G.2. Cloudy scenes are not shown
in the plots nor used in calculations.

Figure G.73: Plot of transmission vs. actual
error for threshold analysis at the Salton Sea.

Table G.8: Statistics for transmission threshold
analysis at the Salton Sea.

Errors of Omission for 1 K Standard	21.43%
Errors of Commission for 1 K Standard	7.14%
Errors of Omission for 2 K Standard	21.43%
Errors of Commission for 2 K Standard	7.14%
Mean Error of High Confidence	0.719 K
Standard Deviation of High Confidence	1.194 K
Commission in High Confidence for 1 K	10.00%
Commission in High Confidence for 2 K	10.00%

Figure G.74: Plot of transmission vs. actual error for threshold analysis at Lake Tahoe.

Table G.9: Statistics for transmission threshold analysis at Lake Tahoe.

Errors of Omission for 1 K Standard	1.43%
Errors of Commission for 1 K Standard	20.71%
Errors of Omission for 2 K Standard	2.14%
Errors of Commission for 2 K Standard	2.14%
Mean Error of High Confidence	0.705 K
Standard Deviation of High Confidence	0.573 K
Commission in High Confidence for 1 K	26.13%
Commission in High Confidence for 2 K	2.70%

As with the regression analysis, results at the Salton Sea and Lake Tahoe are well-behaved as they were chosen on days good for land surface temperature retrieval. The threshold analysis is more consistent than regression analysis with changing locations. With these well-behaved datasets we investigate our ability to predict high confidence for good results. The Salton Sea is particularly difficult to analyze due to the small sample size. Just under 30% of the data are assigned the incorrect confidence as shown in Table G.8 but this is a total of four points and Figure G.73 shows that the errors of omission are very close to the threshold. The mean error is an acceptable magnitude but the standard deviation is quite large due to single error of commission with a large error as shown in the plot. At Lake Tahoe, most errors are due to commission so points with high transmission also had large errors. However, with a standard of of 2 K, the percent of points as errors of commissions decreases considerably as shown in Table G.9. The mean error of high confidence is encouraging as is the standard deviation. Considering the commission in high confidence, just over a quarter of the points that we claim have low errors actually have errors larger than acceptable at the 1 K. This decreases considerably at the 2 K standard.
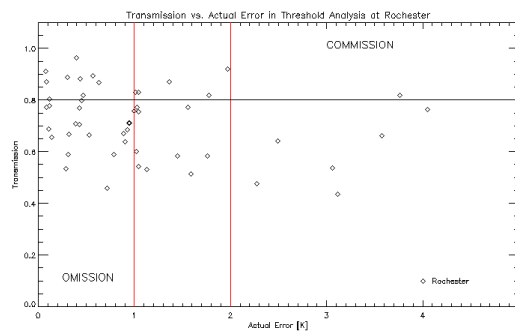
Figure G.75: Plot of transmission vs. actual error for threshold analysis at Rochester.

Table G.10: Statistics for transmission threshold analysis at Rochester.

Errors of Omission for 1 K Standard	25.97%
Errors of Commission for 1 K Standard	11.69%
Errors of Omission for 2 K Standard	37.66%
Errors of Commission for 2 K Standard	5.19%
Mean Error of High Confidence	2.218 K
Standard Deviation of High Confidence	3.091 K
Commission in High Confidence for 1 K	50.00%
Commission in High Confidence for 2 K	22.22%
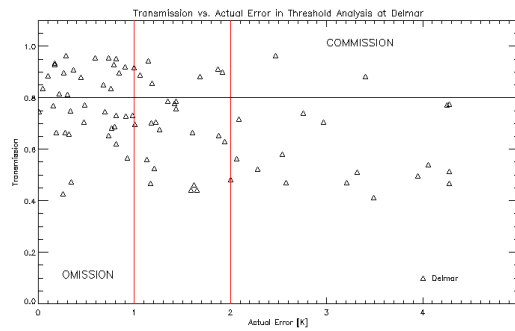
Figure G.76: Plot of transmission vs. actual error for threshold analysis at Delmar.

Table G.11: Statistics for transmission threshold analysis at Delmar.

Errors of Omission for 1 K Standard	17.59%
Errors of Commission for 1 K Standard	12.04%
Errors of Omission for 2 K Standard	34.26%
Errors of Commission for 2 K Standard	6.48%
Mean Error of High Confidence	2.512 K
Standard Deviation of High Confidence	4.166 K
Commission in High Confidence for 1 K	40.63%
Commission in High Confidence for 2 K	37.50%
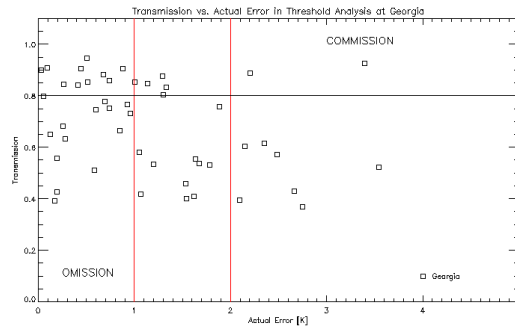
Figure G.77: Plot of transmission vs. actual error for threshold analysis at Georgia.

Table G.12: Statistics for transmission threshold analysis at Georgia.

Errors of Omission for 1 K Standard	23.73%
Errors of Commission for 1 K Standard	11.86%
Errors of Omission for 2 K Standard	40.68%
Errors of Commission for 2 K Standard	3.39%
Mean Error of High Confidence	0.953 K
Standard Deviation of High Confidence	0.833 K
Commission in High Confidence for 1 K	41.28%
Commission in High Confidence for 2 K	11.76%

As expected, errors are less well-behaved at Rochester, Delmar and Georgia as shown in Figures G.75, G.76, and G.77. There are more scenes with larger actual errors, but many fall below the 0.8 threshold line as expected. However, there are many scenes with lower transmission that still have small actual errors, leading to higher errors of omission as shown in Tables G.10, G.11, and G.12. At all three locations, with the 1 K standard, more than 30% of the data is incorrectly assigned a low confidence, and more than 40% at the 2 K standard. At Rochester and Delmar, the mean error of the high confidence scenes is between 2 K and 3 K, but more concerning are the large standard deviations indicating that these scenes are not concentrated at this mean but also contains scenes with larger errors as supported by the plots on Figures G.75, G.76, and G.77. The commission in high confidence is large for both of these locations as well, although decreases at the 2 K standard. The mean and standard deviation at Georgia is more encouraging but this site still has a high percentage of errors of omission; the commission in high confidence at the 2 K standard is lower at Georgia than at Rochester or Delmar.
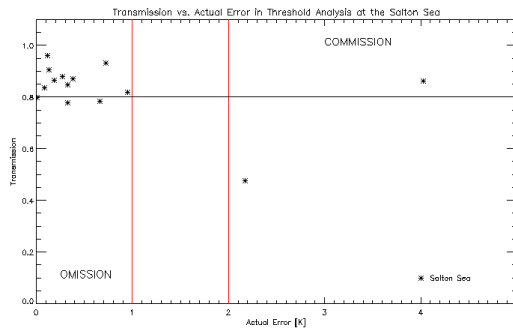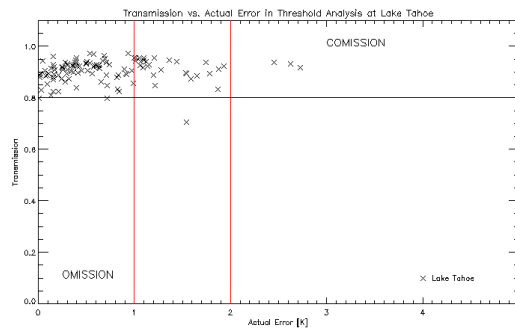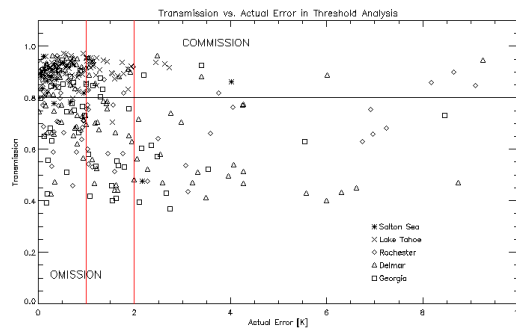
Figure G.78: Plot of transmission vs. actual error for threshold analysis at all locations.

Table G.13: Statistics for transmission threshold analysis at all locations.

Errors of Omission for 1 K Standard	14.57%
Errors of Commission for 1 K Standard	14.82%
Errors of Omission for 2 K Standard	24.12%
Errors of Commission for 2 K Standard	4.27%
Mean Error of High Confidence	1.181 K
Standard Deviation of High Confidence	2.150 K
Commission in High Confidence for 1 K	31.38%
Commission in High Confidence for 2 K	9.04%

Consdering all locations, close to 30% of the data points are incorrectly classified. However, the mean of the high confidence scenes from Table G.13 is 1.181 K with a standard devotion of 2.150 K. This is an encouraging result although possibly unfairly skewed by Lake Tahoe scenes and the percent of errors of omission is still high. The commission in high confidence at the 2 K standard is below 10% which is encouraging.

From a visual analysis of Figures G.73, G.74, G.75, G.76, G.77, and G.78, it does not appear a major adjustment to the initial transmission threshold would significantly improve results. At most locations, disregarding results at the Salton Sea and Lake Tahoe, there are large actual errors throughout the range of transmissions indicating that adjusting the transmission threshold would cause either the omission or commission errors to increase. As discussed above, the relationship of higher error with lower transmission is neither linear nor absolute.

G.4.2 Relative Humidity in Threshold Analysis

Relative humidity is plotted as a function of error for each validation site in Figures G.79, G.80, G.81, G.82 and G.83 for an error range of 0 K to 5 K. The initial relative humidity threshold of 70% is shown on the plots along with the 1 K and 2 K standards. Tables quantifying the errors of omission and commission as well as the mean and standard deviation of high confidence pixels are shown for corresponding locations in Tables G.14, G.15, G.16, G.17, and G.18. The same is shown for all locations over a larger range of errors in Figure G.84 with the corresponding calculations in Table G.19.

Figure G.79: Plot of relative humidity vs. actual error for threshold analysis at the Salton Sea.

Table G.14: Statistics for relative humidity threshold analysis at the Salton Sea.

Errors of Omission for 1 K Standard	7.14%
Errors of Commission for 1 K Standard	7.14%
Errors of Omission for 2 K Standard	7.14%
Errors of Commission for 2 K Standard	7.14%
Mean Error of High Confidence	0.527 K
Standard Deviation of High Confidence	0.584 K
Commission in High Confidence for 1 K	8.33%
Commission in High Confidence for 2 K	8.33%

Figure G.80: Plot of relative humidity vs. actual error for threshold analysis at Lake Tahoe.

Table G.15: Statistics for relative humidity threshold analysis at Lake Tahoe.

Errors of Omission for 1 K Standard	11.43%
Errors of Commission for 1 K Standard	15.71%
Errors of Omission for 2 K Standard	16.43%
Errors of Commission for 2 K Standard	1.43%
Mean Error of High Confidence	0.676 K
Standard Deviation of High Confidence	0.545 K
Commission in High Confidence for 1 K	24.44%
Commission in High Confidence for 2 K	2.22%

Only one scene is assigned an incorrect high confidence and an incorrect low confidence at the Salton Sea, as shown in Figure G.79, and both the mean and standard deviation of high confidence scenes is low, as shown in Table G.14. For Lake Tahoe, the percentage of scenes given incorrect confidences significantly decreases for the 2 K standard due to the decrease in errors of commission as can be observed in Figure G.80. There are still a handful of scenes with high relative humidities and accurate land surface temperature retrieval, but the mean and standard deviation of high confidence scenes is encouraging as shown in Table G.15. The commission in high confidence at the 2 K standard for both scenes is also encouraging.
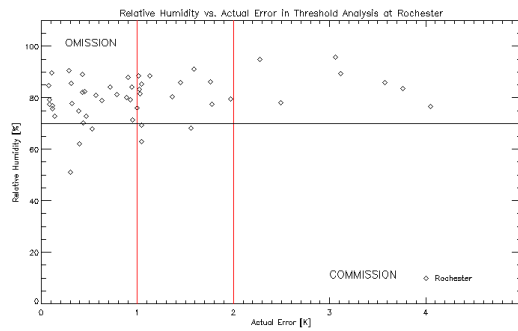
Figure G.81: Plot of relative humidity vs. actual error for threshold analysis at Rochester.

Table G.16: Statistics for relative humidity threshold analysis at Rochester.

Errors of Omission for 1 K Standard	33.77%
Errors of Commission for 1 K Standard	5.19%
Errors of Omission for 2 K Standard	48.05%
Errors of Commission for 2 K Standard	1.30%
Mean Error of High Confidence	1.931 K
Standard Deviation of High Confidence	2.992 K
Commission in High Confidence for 1 K	57.14%
Commission in High Confidence for 2 K	14.29%
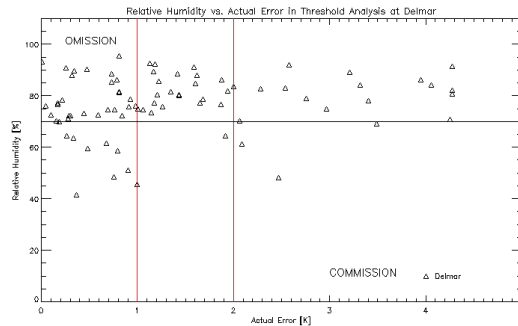
Figure G.82: Plot of relative humidity vs. actual error for threshold analysis at Delmar.

Table G.17: Statistics for relative humidity threshold analysis at Delmar.

Errors of Omission for 1 K Standard	26.85%
Errors of Commission for 1 K Standard	6.48%
Errors of Omission for 2 K Standard	48.15%
Errors of Commission for 2 K Standard	5.56%
Mean Error of High Confidence	2.776 K
Standard Deviation of High Confidence	3.779 K
Commission in High Confidence for 1 K	43.75%
Commission in High Confidence for 2 K	37.50%

For both Rochester and Delmar, errors of omission are considerably higher than errors of commission; a very large portion of the data would be classified as low confidence due to the large number of scenes with high relative humidities, as shown in Figures G.81 and G.82. This will be a problem in areas with generally humid conditions. Because few scenes would be classified as high confidence, there are only a handful of errors of commission to influence the high confidence mean and standard deviation, as shown in Tables G.16 and G.17. However, particularly at Delmar, there are a number of high confidence scenes with errors large enough to make the mean and standard deviation greater than is desirable. Georgia has similar results as shown in Figure G.83, but there are fewer scenes with low relative humidities and large errors, such that the results are improved at the 2 K standard, shown in Table G.18, and the mean and standard deviations are smaller than those for Rochester and Delmar. The commission in high confidence at the 2 K standard is lowest in Rochester but larger than is desirable at all three locations. So few scenes are assigned a high confidence in this case that a small number of errors can make this percentage of commission in high confidence large.
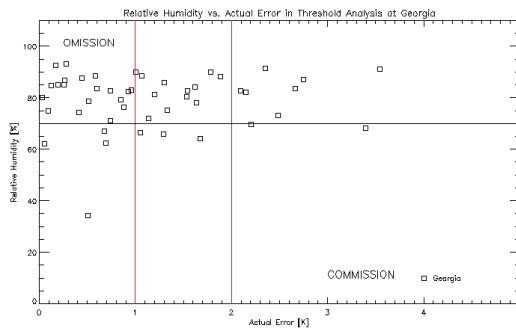
Figure G.83: Plot of relative humidity vs. actual error for threshold analysis at Georgia.

Table G.18: Statistics for relative humidity threshold analysis at Georgia.

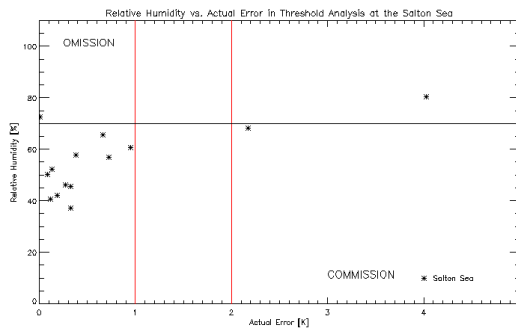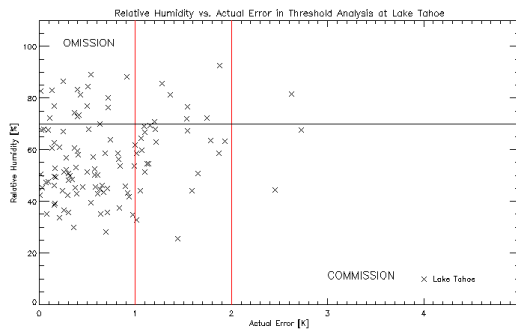|	
Errors of Omission for 1 K Standard	33.90%
Errors of Commission for 1 K Standard	8.47%
Errors of Omission for 2 K Standard	5.42%
Errors of Commission for 2 K Standard	3.39%
Mean Error of High Confidence	1.282 K
Standard Deviation of High Confidence	1.023 K
Commission in High Confidence for 1 K	55.56%
Commission in High Confidence for 2 K	22.22%
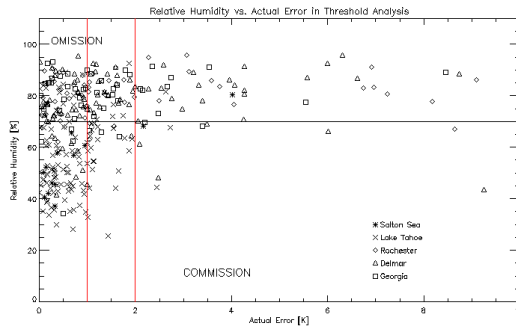
Figure G.84: Plot of relative humidity vs. actual error for threshold analysis at all locations.

Table G.19: Statistics for relative humidity threshold analysis at all locations.

Errors of Omission for 1 K Standard	23.12%
Errors of Commission for 1 K Standard	9.80%
Errors of Omission for 2 K Standard	36.43%
Errors of Commission for 2 K Standard	3.02%
Mean Error of High Confidence	1.282 K
Standard Deviation of High Confidence	1.023 K
Commission in High Confidence for 1 K	29.10%
Commission in High Confidence for 2 K	8.86%

Considering the results for all scenes shown in Figure G.84 and Table G.19, there are a large number of errors of omission due to the abundance of scenes with high relative humidities but low errors. However, there are few scenes with large error and low relative humidity, such that the errors of commission are low, the mean and standard deviation of high confidence scenes is acceptable, and the commission in high confidence at the 2 K standard is less than ten percent. This is greatly influenced by the number of Lake Tahoe scenes with low error and relative humidity. Because there are a large number of scenes with low error, increasing the relative humidity threshold would decrease the errors of omission; however, considering only scenes with larger than desired errors, the relative humidity threshold at 70% seems to be well matched to correctly predicting their low confidence.

G.4.3 Maximum Temperature in Threshold Analysis

Plots for maximum temperature as a function of error are shown in Figures G.85, G.86, G.87, G.88, and G.89 for the Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia respectively. Tables G.85, G.86, G.87, G.88, and G.89 show the values derived from these figures. Each plot also includes the threshold line, at 305 K for the maximum temperature and vertical lines for the 1 K and 2 K standards. Figure G.90 shows this same data for all locations over a larger error range with the

corresponding data in Table G.25.

Figure G.85: Plot of maximum temperature vs. actual error for threshold analysis at the Salton Sea.

Table G.20: Statistics for maximum temperature threshold analysis at the Salton Sea.

Errors of Omission for 1 K Standard	21.43%
Errors of Commission for 1 K Standard	7.14%
Errors of Omission for 2 K Standard	21.43%
Errors of Commission for 2 K Standard	7.14%
Mean Error of High Confidence	0.711 K
Standard Deviation of High Confidence	1.199 K
Commission in High Confidence for 1 K	10.00%
Commission in High Confidence for 2 K	10.00%

Figure G.86: Plot of maximum temperature vs. actual error for threshold analysis at Lake Tahoe.

Table G.21: Statistics for maximum temperature threshold analysis at Lake Tahoe.

Errors of Omission for 1 K Standard	11.43%
Errors of Commission for 1 K Standard	20.00%
Errors of Omission for 2 K Standard	12.86%
Errors of Commission for 2 K Standard	2.14%
Mean Error of High Confidence	0.755 K
Standard Deviation of High Confidence	0.585 K
Commission in High Confidence for 1 K	29.17%
Commission in High Confidence for 2 K	3.13%

Because the data at the Salton Sea and Lake Tahoe have a relatively small range of errors, the mean and standard deviation of the high confidence results are small in Tables G.20 and G.21. The commission in high confidence for both, particularly at the 2 K standard, are also low. However, there are a number of results with low errors that have high maximum temperatures, leading to errors of omission as shown in Figures G.85 and G.86. Climates are obviously variable with location and this is our first indication that the same threshold may not be optimal for all sites. Increasing the threshold would eliminate these errors of omission, but would result in all points falling below the threshold, or high confidence in all points. Therefore, a more diverse data set is needed to evaluate the utility of this metric in this capacity.
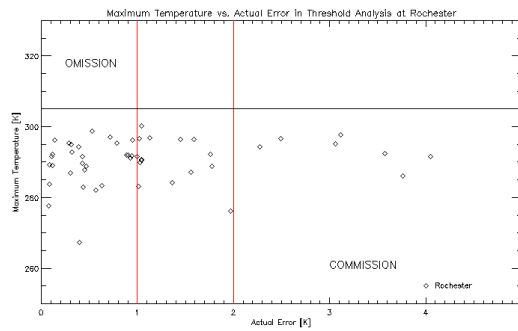
Figure G.87: Plot of maximum temperature vs. actual error for threshold analysis at Rochester.

Table G.22: Statistics for maximum temperature threshold analysis at Rochester.

Errors of Omission for 1 K Standard	0.00%
Errors of Commission for 1 K Standard	36.36%
Errors of Omission for 2 K Standard	0.00%
Errors of Commission for 2 K Standard	18.18%
Mean Error of High Confidence	1.907 K
Standard Deviation of High Confidence	2.395 K
Commission in High Confidence for 1 K	49.12%
Commission in High Confidence for 2 K	24.56%
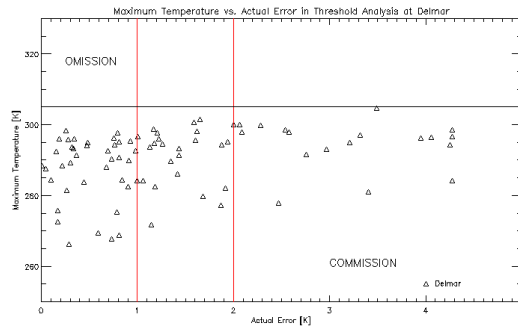
Figure G.88: Plot of maximum temperature vs. actual error for threshold analysis at Delmar.

Table G.23: Statistics for maximum temperature threshold analysis at Delmar.

Errors of Omission for 1 K Standard	0.00%
Errors of Commission for 1 K Standard	50.00%
Errors of Omission for 2 K Standard	0.00%
Errors of Commission for 2 K Standard	27.78%
Mean Error of High Confidence	2.449 K
Standard Deviation of High Confidence	3.164 K
Commission in High Confidence for 1 K	58.70%
Commission in High Confidence for 2 K	32.61%
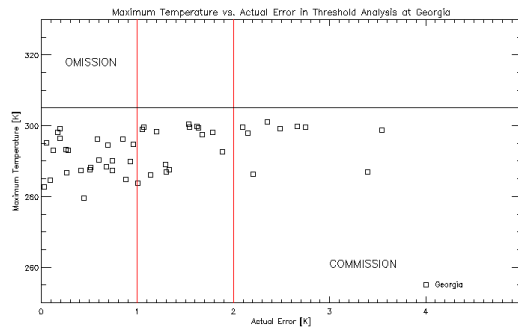
Figure G.89: Plot of maximum temperature vs. actual error for threshold analysis at Georgia.

Table G.24: Statistics for maximum temperature threshold analysis at Georgia.

Errors of Omission for 1 K Standard	0.00%
Errors of Commission for 1 K Standard	52.54%
Errors of Omission for 2 K Standard	0.00%
Errors of Commission for 2 K Standard	27.12%
Mean Error of High Confidence	2.379 K
Standard Deviation of High Confidence	3.485 K
Commission in High Confidence for 1 K	56.36%
Commission in High Confidence for 2 K	29.09%

At Rochester, Delmar and Georgia, in Figures G.87, G.88, and G.89, all points are below the threshold, resulting in high means and standard deviations for the high confidence points and no errors of omission in Tables G.22, G.23, and G.24. This would be assigning high confidence to every point and indicates that the threshold needs to be adjusted, which is also supported by the Salton Sea and Lake Tahoe results and the larger percentage of commission in high confidence. However,

considering the data visually, lowering the threshold to decrease the errors of commission would also result in errors of omission for all locations. Scenes with the largest maximum temperatures have a large spread of errors. There is a general trend of only low errors for the lowest maximum temperatures, but a large error range for all other maximum temperatures. This is the first indication that this metric may not be helpful in the threshold analysis. Results are the same when all locations are plotted together. While there are few points with maximum temperatures above the threshold, any decrease in the threshold would greatly increase the errors of omission in Figure G.90 and Table G.25. Also, the range of maximum temperatures is very different at each site. This may indicate that rather than behaving as expected, with error increasing with maximum temperature, we can only infer low errors from the lowest maximum temperatures but very little from any maximum temperature values.
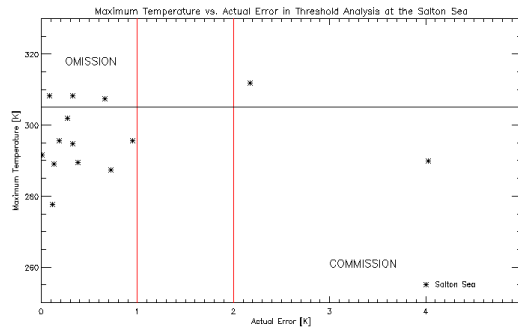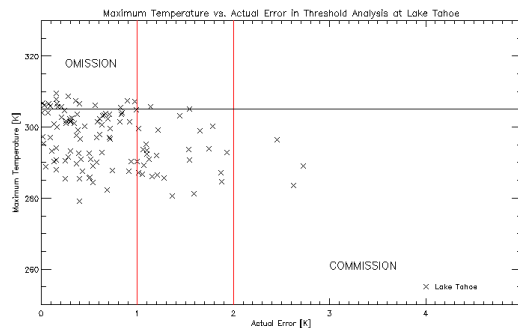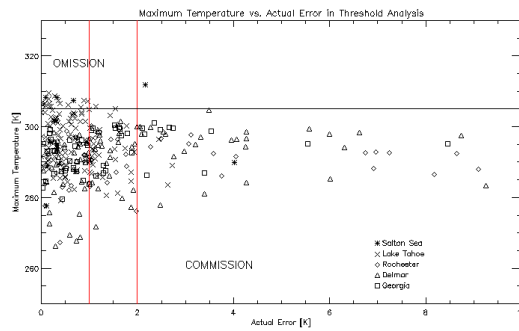

Figure G.90: Plot of maximum temperature vs. actual error for threshold analysis at all locations.

Table G.25: Statistics for maximum temperature threshold analysis at all locations.

Errors of Omission for 1 K Standard	4.77%
Errors of Commission for 1 K Standard	36.68%
Errors of Omission for 2 K Standard	5.28%
Errors of Commission for 2 K Standard	16.08%
Mean Error of High Confidence	1.756 K
Standard Deviation of High Confidence	2.612 K
Commission in High Confidence for 1 K	45.81%
Commission in High Confidence for 2 K	20.65%

G.4.4 Dew Point Depression in Threshold Analysis

Dew point depression is plotted as a function of error in Figures G.91, G.92, G.93, G.94, and G.95 for the Salton Sea, Lake Tahoe, Rocehster, Delmar, and Georgia respectively with related data in Tables G.26, G.27, G.28, G.29 and G.30. Each plot shows the original dew point depression threshold at 3.0 K and lines for both the 1 K and 2 K standards.

Figure G.91: Plot of dew point depression vs. actual error for threshold analysis at the Salton Sea.

Table G.26: Statistics for dew point depression threshold analysis at the Salton Sea.

Errors of Omission for 1 K Standard	0.00%
Errors of Commission for 1 K Standard	7.14%
Errors of Omission for 2 K Standard	0.00%
Errors of Commission for 2 K Standard	7.14%
Mean Error of High Confidence	0.487 K
Standard Deviation of High Confidence	0.577 K
Commission in High Confidence for 1 K	7.69%
Commission in High Confidence for 2 K	7.69%

Figure G.92: Plot of dew point depression vs. actual error for threshold analysis at Lake Tahoe.

Table G.27: Statistics for dew point depression threshold analysis at Lake Tahoe.

Errors of Omission for 1 K Standard	5.71%
Errors of Commission for 1 K Standard	18.57%
Errors of Omission for 2 K Standard	7.86%
Errors of Commission for 2 K Standard	1.43%
Mean Error of High Confidence	0.689 K
Standard Deviation of High Confidence	0.546 K
Commission in High Confidence for 1 K	25.49%
Commission in High Confidence for 2 K	1.86%

For the Salton Sea, there is no error of omission because all dew point depressions are larger than the threshold as shown in Figure G.91. The mean and standard deviation in Table G.26 are small by the nature of the actual errors at this site but the commission in high confidence is also low, particularly at the 2 K standard. However, increasing the threshold would increase the error of omission without decreasing error of commission. The sample is too small and the data too well-behaved to make any useful evaluations or conclusions. Lake Tahoe provides a little more variability in dew point depression but still a small range of errors. Especially at the 2 K standard, as shown in Figures G.92 and Table G.27, there are low errors of omission and commission and a large number of results successfully classified. The commission in high confidence is particularly low at the 2 K standard.

Figure G.93: Plot of dew point depression vs. actual error for threshold analysis at Rochester.

Table G.28: Statistics for dew point depression threshold analysis at Rochester.

Errors of Omission for 1 K Standard	12.99%
Errors of Commission for 1 K Standard	14.29%
Errors of Omission for 2 K Standard	23.38%
Errors of Commission for 2 K Standard	6.49%
Mean Error of High Confidence	1.587 K
Standard Deviation of High Confidence	2.339 K
Commission in High Confidence for 1 K	36.67%
Commission in High Confidence for 2 K	16.67%
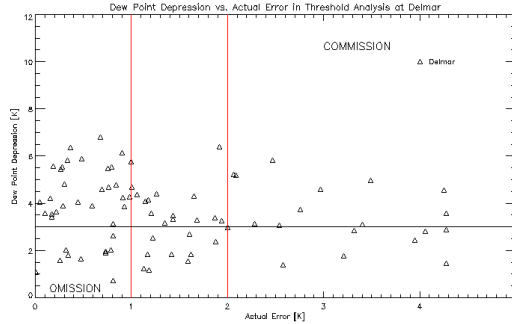
Figure G.94: Plot of dew point depression vs. actual error for threshold analysis at Delmar.

Table G.29: Statistics for dew point depression threshold analysis at Delmar.

Errors of Omission for 1 K Standard	9.26%
Errors of Commission for 1 K Standard	27.78%
Errors of Omission for 2 K Standard	18.52%
Errors of Commission for 2 K Standard	14.81%
Mean Error of High Confidence	2.127 K
Standard Deviation of High Confidence	3.198 K
Commission in High Confidence for 1 K	51.72%
Commission in High Confidence for 2 K	27.59%
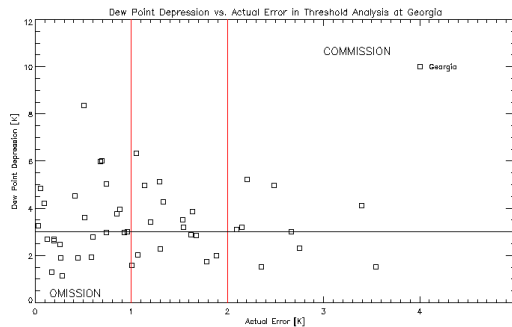
Figure G.95: Plot of dew point depression vs. actual error for threshold analysis at Georgia.

Table G.30: Statistics for dew point depression threshold analysis at Georgia.

Errors of Omission for 1 K Standard	20.34%
Errors of Commission for 1 K Standard	28.81%
Errors of Omission for 2 K Standard	32.20%
Errors of Commission for 2 K Standard	15.25%
Mean Error of High Confidence	2.066 K
Standard Deviation of High Confidence	2.764 K
Commission in High Confidence for 1 K	58.62%
Commission in High Confidence for 2 K	31.03%

Rochester, Delmar and Georgia have more variable results in Figures G.93, G.94 and G.95. Results for Rochester in Figure G.93 are well matched to the threshold and look as expected. There are a handful of errors of omission, but generally points with large errors have small dew points depressions and are correctly given low confidence. The mean error and standard deviation is still larger than desired for high confidence results as is the percent of commission in high confidence.

Although slightly lower for Rochester, the commission in high confidence is too large at all three sites. Points for Delmar and Georgia look less well-match to the threshold and more like results for maximum temperature; both the means and standard deviations are large as shown in Tables G.29 and G.30, but adjusting the threshold does not appear to improve results. There are a variety of errors throughout the range of dew point depressions. Finally, results for all locations are shown in Figure G.96. With the 2 K standard, around 25% of the data is classified incorrectly; the error and standard deviation of the high confidence results is larger than desired, as is the commission in high confidence, but adjusting the threshold does not appear to improve results. Most points with a small dew point depression and large errors are correctly given low confidence; increasing the threshold would increase the error of omission and there are already a large percentage of errors of omission due to the points with small dew point depressions and small actual errors.
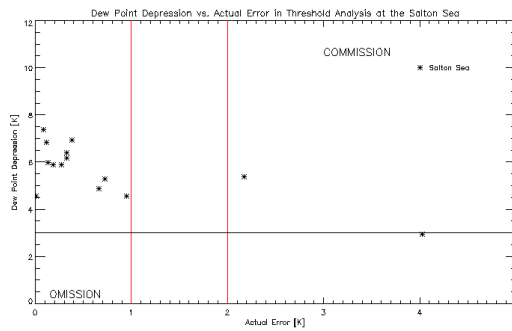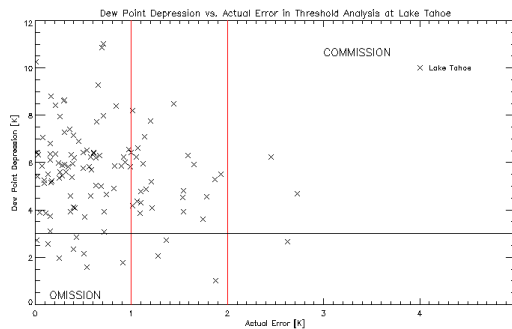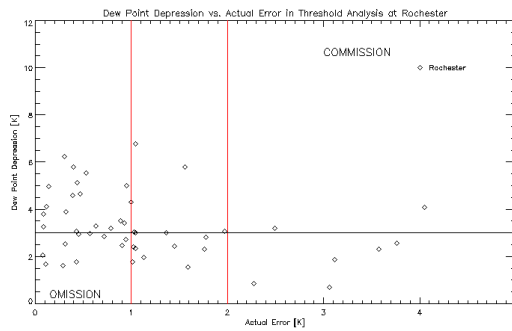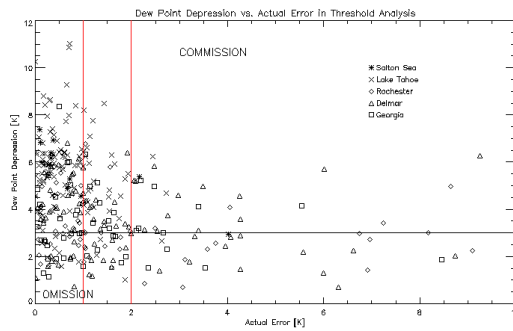

Figure G.96: Plot of dew point depression vs. actual error for threshold analysis at all locations.

Table G.31: Statistics for dew point depression threshold analysis at all locations.

Errors of Omission for 1 K Standard	10.05%
Errors of Commission for 1 K Standard	21.36%
Errors of Omission for 2 K Standard	17.09%
Errors of Commission for 2 K Standard	8.29%
Mean Error of High Confidence	1.325 K
Standard Deviation of High Confidence	2.178 K
Commission in High Confidence for 1 K	36.64%
Commission in High Confidence for 2 K	14.22%

G.4.5 Column Water Vapor in Threshold Analysis

Salton Sea, Lake Tahoe, Rochester, Delmar and Georgia column water vapor results are shown in Figures G.97, G.98, G.99, G.100, and G.101 with corresponding Tables G.32, G.33, G.34, G.35, and G.36. The initial column water vapor threshold is set at 2 cm; the 1 K and 2 K standards are also shown on the plots.

Figure G.97: Plot of column water vapor vs. actual error for threshold analysis at the Salton Sea.

Table G.32: Statistics for column water vapor threshold analysis at the Salton Sea.

Errors of Omission for 1 K Standard	7.14%
Errors of Commission for 1 K Standard	7.14%
Errors of Omission for 2 K Standard	7.14%
Errors of Commission for 2 K Standard	7.14%
Mean Error of High Confidence	0.626 K
Standard Deviation of High Confidence	1.104 K
Commission in High Confidence for 1 K	8.33%
Commission in High Confidence for 2 K	8.33%

Figure G.98: Plot of column water vapor vs. actual error for threshold analysis at Lake Tahoe.

Table G.33: Statistics for column water vapor threshold analysis at Lake Tahoe.

Errors of Omission for 1 K Standard	25.00%
Errors of Commission for 1 K Standard	17.14%
Errors of Omission for 2 K Standard	29.29%
Errors of Commission for 2 K Standard	2.14%
Mean Error of High Confidence	0.786 K
Standard Deviation of High Confidence	0.601 K
Commission in High Confidence for 1 K	32.88%
Commission in High Confidence for 2 K	4.11%

Illustrated in Figure G.97, the errors at the Salton Sea are too small to make a judgement on the column water vapor threshold. Almost all points are correctly given high confidence. Only one point has a larger than average column water vapor, correctly assigned a low confidence, and one point has a large error and low column water vapor, incorrectly assigned a high confidence. Lake Tahoe results provide a larger range of column water vapor values for a dataset with mostly small errors. Over 40% of the data is given incorrect confidence with the 1 K standard, and over 30% with the 2 K standard. The means for both Salton Sea and Lake Tahoe are small in Tables G.32 and G.33 respectively, due to the nature of the actual errors on these datasets; the standard deviation is larger than desired for Salton Sea but acceptably small for Lake Tahoe. The commission in high confidence is relatively low at the 2 K standard for both sites.

Figure G.99: Plot of column water vapor vs. actual error for threshold analysis at Rochester.

Table G.34: Statistics for column water vapor threshold analysis at Rochester.

Errors of Omission for 1 K Standard	20.78%
Errors of Commission for 1 K Standard	15.58%
Errors of Omission for 2 K Standard	29.87%
Errors of Commission for 2 K Standard	6.49%
Mean Error of High Confidence	1.837 K
Standard Deviation of High Confidence	2.583 K
Commission in High Confidence for 1 K	48.00%
Commission in High Confidence for 2 K	20.00%

Figure G.100: Plot of column water vapor vs. actual error for threshold analysis at Delmar.

Table G.35: Statistics for column water vapor threshold analysis at Delmar.

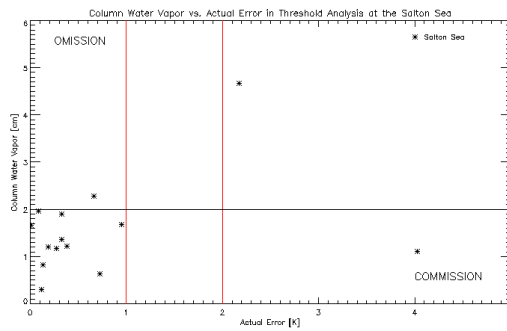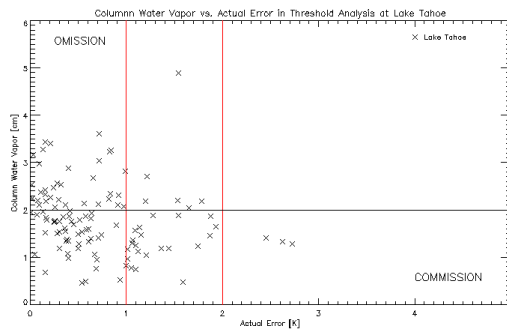|	
Errors of Omission for 1 K Standard	12.04%
Errors of Commission for 1 K Standard	19.44%
Errors of Omission for 2 K Standard	24.07%
Errors of Commission for 2 K Standard	9.26%
Mean Error of High Confidence	2.561 K
Standard Deviation of High Confidence	4.046 K
Commission in High Confidence for 1 K	45.65%
Commission in High Confidence for 2 K	21.74%

Figure G.101: Plot of column water vapor vs. actual error for threshold analysis at Georgia.

Table G.36: Statistics for column water vapor threshold analysis at Georgia.

Errors of Omission for 1 K Standard	13.56%
Errors of Commission for 1 K Standard	13.56%
Errors of Omission for 2 K Standard	28.81%
Errors of Commission for 2 K Standard	3.39%
Mean Error of High Confidence	0.919 K
Standard Deviation of High Confidence	0.752 K
Commission in High Confidence for 1 K	33.33%
Commission in High Confidence for 2 K	8.33%

Results for Rochester, Delmar and Georgia all have a large percentage of the data incorrectly classified, roughly 25% to 35% for both standards in all cases as described in Tables G.34, G.35 and G.36. However, the results with column water do appear to follow the trend of larger column water vapor and larger errors. There are still results with small errors and large column water vapors, but fewer results with large errors and small column water vapors. Unlike maximum temperature and

dew point depression in particular, which have little distinct shape in their results and a large range of errors within a small range of the metrics, there is shape to the column water vapor data. In these three locations, after visually analyzing Figures G.99, G.100, and G.101, it appears that a slight increase in the threshold could decrease the error of omission without significantly increasing the error of commission or the means and standard deviation of the high confidence results. Particularly at Delmar, but also Rochester, the mean and standard deviation of the high confidence results is still higher than other locations and higher than desirable. The commission in high confidence is below ten percent for the 2 K standard at Georgia, but still quite large at both Rochester and Delmar.
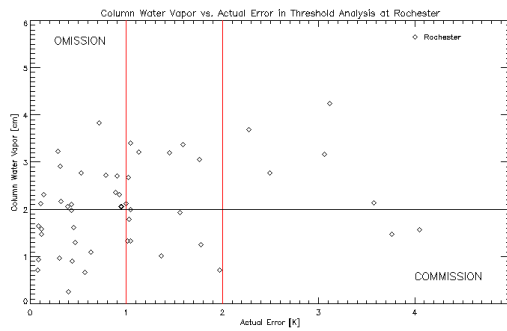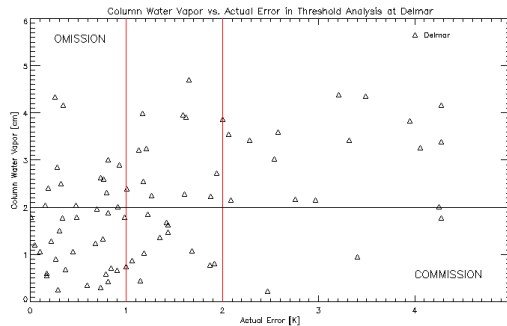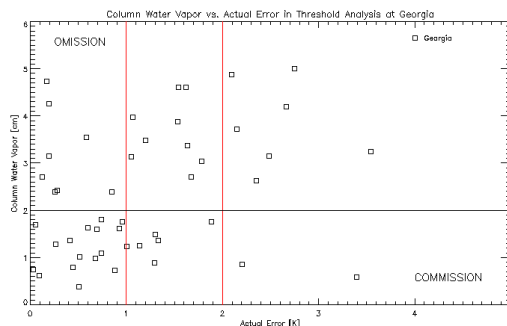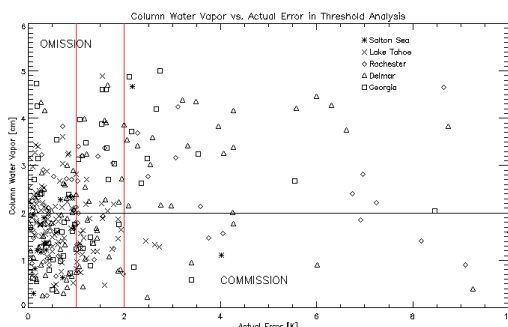

Figure G.102: Plot of column water vapor vs. actual error for threshold analysis at all locations.

Table G.37: Statistics for column water vapor threshold analysis at all locations.

Errors of Omission for 1 K Standard	18.34%
Errors of Commission for 1 K Standard	16.58%
Errors of Omission for 2 K Standard	27.14%
Errors of Commission for 2 K Standard	5.28%
Mean Error of High Confidence	1.393 K
Standard Deviation of High Confidence	2.429 K
Commission in High Confidence for 1 K	36.67%
Commission in High Confidence for 2 K	11.67%

Considering all locations together, there is a still a large percentage of the data incorrectly classified as shown in Table G.37. However, the data appears to have more shape than other metrics which is encouraging for future analysis. Although the threshold seems to be initially well predicted in Figure G.102, it appears that increasing the threshold slightly could significantly decrease the error of omission while only slightly increasing errors of commission. However, as discussed, it is likely we want to keep the errors of commission as small as possible. Currently the commission in high confidence for the 2 K standard is just above ten percent for all locations. The mean and standard deviations are reasonable in magnitude for this metric.

G.4.6   Concluding Remarks on Threshold Analysis

At an initial glance, it does not appear that the threshold analysis provides a better method of error estimation than regression analysis. In most cases, there is still a large percentage of points assigned incorrect confidence. There is potential for adjustments to thresholds based on a more robust validation dataset, but the patterns of the data are not encouraging for setting thresholds that would lead to accurate confidence metric predictions. While we can achieve fairly low errors of commission, this almost always leads to high errors of omission. We are more concerned with errors of commission, but still do not want to eliminate large volumes of good data. This is similar to the problem of excluding clouds in the vicinity; however, with the cloud analysis discussed in Section 5.3.2, there is an option for flagging clouds in the vicinity to be included with a lower level

of confidence. This currently does not exist in the threshold analysis. Therefore, while there is room for improvement, we believe that none would prove as useful as the methods discussed in Section 5.3, and therefore this method was also eliminated from consideration.

Bibliography

[oce, 2010] (2010). Ocean level-2 data products. http://oceancolor.gsfc.nasa.gov/DOCS/Ocean_Level-2_Data_Products.pdf.

[lan, 2013] (2013). Using landsat 7 data. http://landsat.usgs.gov/using_Landsat_7_data.php.

[mod, 2014] (2014). Ocean color browse. http://oceancolor.gsfc.nasa.gov/cgi/browse.pl?sen=tm.

[Barsi et al., 2003] Barsi, J., Schott, J., Palluconi, F., Helder, D., Hook, S., Markham, B., Chander, G., and O'Donnell, E. (2003). Landsat tm and etm+ thermal band calibration. *Can. J. Remote Sensing*, 29(2):141–253.

[Berk et al., 1999] Berk, A., Anderson, G., Acharya, P., Chetwynd, J., Bernstein, L., Shettle, E., Matthew, M., and Adler-Golden, S. (1999). *MODTRAN User's Manual*. Air Force Reserach Laboratory Space Vehicles Directorate, Hanscom AFB, MA.

[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.

[Diak and Whipple, 1993] Diak, G. R. and Whipple, M. S. (1993). Improvements to models and methods for evaluating the land surface energy balance and effective roughness using radiosonde reports and satellite-measured skin temperature data. *Agriculture and Forestry Meteorology*, 63:189–218.

[Ehrlich and Lambin, 1996] Ehrlich, D. and Lambin, E. F. (1996). The surface temperature-vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing*, 17:463–487.

[Engel and Weinstein, 1983] Engel, J. and Weinstein, O. (1983). The thematic mapper - an overview. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(3):258–263.

[Feldhake et al., 1996] Feldhake, C. M., Glenn, D. M., and Peterson, D. L. (1996). Pasture soil surface temperature repsonse to drought. *Agronomy Journal*, 88:652–656.

[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.

[Freitas et al., 2011] Freitas, 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.

[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., Abtahi, A., Palluconi, 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., 2002] Hook, S. J., Prata, F. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., and Palmarsson, S. O. (2002). Retrieval of lake bulk and skin temperautres using along-track scanning radiometer (atsr-2) data: A case study using lake tahoe, california. *Journal of Atmospheric and Oceanic Technology*, 20:534–548.

[Hook and Rivera, 2013a] Hook, S. J. and Rivera, G. (2013a). Lake tahoe validation. http://laketahoe.jpl.nasa.gov.

[Hook and Rivera, 2013b] Hook, S. J. and Rivera, G. (2013b). Salton sea validation. http://saltonsea.jpl.nasa.gov.

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

[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, One Space Park, Redondo Beach, California 90278.

[Irons and Rocchio, 2013] Irons, J. R. and Rocchio, L. (2013). From the beginning. http://landsat.gsfc.nasa.gov/about/history.html.

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

[Jimenez-Munoz and Sobrino, 2003] Jimenez-Munoz, J. C. and Sobrino, J. A. (2003). A generalized single-channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research*, 108.

[Jimenez-Munoz and Sobrino, 2004] Jimenez-Munoz, J. C. and Sobrino, J. A. (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.

[Kimuru and Shimiru, 1994] Kimuru, F. and Shimiru, P. (1994). Estimation of sensible and latent heat fluxes from soil surface temperature using a linear air-land heat transfer model. *Journal of Applied Meteorology*, 33(4):477–489.

[King, 1999] King, M. D. (1999). Eos science plan: The state of science in the eos. Technical report, NASA/Goddard Space Flight Center.

[Kruger, 2010] Kruger, M. (2010). Converter for humidity of air. http://www.cactus2000.de/uk/unit/masshum.shtml.

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

[Maccherone, 2014] Maccherone, B. (2014). Modis specifications. http://modis.gsfc.nasa.gov/about/specifications.php.

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

[NCEP/NCAR, 2013] NCEP/NCAR (2013). Ncep.reanalysis. ftp.cdc.noaa.gov/Dataset/ncep.reanalysis/pressure.

[NDBC, 2013a] NDBC (2013a). Station 41008 (llnr 833) - grays reef - 40 nm southeast of savannah, ga. http://www.ndbc.noaa.gov/station_page.php?station=41008.

[NDBC, 2013b] NDBC (2013b). Station 44009 (llnr 168) - delaware bay 26 nm southeast of cape may, nh. http://www.ndbc.noaa.gov/station_page.php?station=44009.

[NDBC, 2013c] NDBC (2013c). Station 45012 (llnr 2655) - east lake ontario - 20 nm north northeast of rochester, ny. http://www.ndbc.noaa.gov/station_page.php?station=45012.

[NDBC, 2014a] NDBC (2014a). National data buoy center. http://www.ndbc.noaa.gov.

[NDBC, 2014b] NDBC (2014b). Station 45001 (llnr 16880) - mid superior- 60nm north northeast hancock, mi. http://www.ndbc.noaa.gov/station_page.php?station=45001.

[NDBC, 2014c] NDBC (2014c). Station 45003 (llnr 11510) - north huron - 32nm northeast of alpena, mi. http://www.ndbc.noaa.gov/station_page.php?station=45003.

[NDBC, 2014d] NDBC (2014d). Station 45004 (llnr 16850) - east superior -70 nm ne marquette, mi. http://www.ndbc.noaa.gov/station_page.php?station=45004.

[NDBC, 2014e] NDBC (2014e). Station 45008 (llnr 10010) - south huron - 43nm east of oscoda, mi. http://www.ndbc.noaa.gov/station_page.php?station=45008.

[NDBC, 2014f] NDBC (2014f). Station 46011 (llnr 215) - santa maria - 21nm nw of point arguello, ca. http://www.ndbc.noaa.gov/station_page.php?station=46011.

[NDBC, 2014g] NDBC (2014g). Station 46025 (llnr 181) - santa monica basin - 33nm wsw of santa monica, ca image indicating link to rss feed of recent meteorological observations for station 46025 station 46025 (llnr 181) - santa monica basin - 33nm wsw of santa monica, ca image indicating link to rss feed of recent meteorological observations for station 46025 station 46025 (llnr 181) - santa monica basin - 33nm wsw of santa monica, ca. http://www.ndbc.noaa.gov/station_page.php?station=46025.

[NOAA/ESRL/PSD, 2013] NOAA/ESRL/PSD (2013). Ncep/ncar reanalysis 1: Pressure. http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.pressure.html.

[NOMADS, 2012] NOMADS (2012). Ncep north american regional reanalysis (32km, 25 years). http://nomads.ncdc.noaa.gov/data.php#narr_datasets.

[OPeNDAP, 2013] OPeNDAP (2013). Contents of /merra/mai3coasm.5.2.0. http://goldsmr3.sci.gsfc.nasa.gov/opendap/MERRA/MAI3CPASM.5.2.0/contents.html.

[Padula, 2008] Padula, F. P. (2008). Historic thermal calibration of landsat 5 tm through an improved physics based approach. Master's thesis, Rochester Institute of Technology.

[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.

[Padula et al., 2010] Padula, F. P., Schott, J. R., Barsi, J. A., Raqueno, N. G., and Hook, S. J. (2010). Calibration of landsat 5 thermal infrared channel: updated calibration history and assesment of the errors associated with the methodology. *Can. J. Remote Sensing*, 36(5):617–630.

[Price, 1982] Price, J. C. (1982). On the use of satellite data to infer surface fluxes at meteorological scales. *Journal of Applied Meteorology*, 21:1111–1122.

[Qin et al., 2001a] 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., 2001b] 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.

[Rienecker and Gass, 2013] Rienecker, M. and Gass, J. (2013). Merra: Modern-era retrospective analysis for research and applications. http://gmao.gsfc.nasa.gov/research/merra/.

[Running et al., 1994] Running, S. W., Justice, C., Salomonson, V., Hall, D., Barker, J., Kaufman, Y., Strahler, A., Huete, A., Mueller, J. P., Vanderbilt, V., Wan, Z., and Teillet, P. (1994). Terrestrial remote sensing science and algorithms planned for eos/modis. *International Journal of Remote Sensing*, 15(17):3587–3620.

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

[Schott et al., 1999] Schott, J. R., Brown, S. D., and Barsi, J. A. (1999). *Thermal Remote Sensing in Land Surface Processes*, chapter 10 Calibration of Thermal Infrared (TIR) Sensors. Ann Arbor Press, 1st edition.

[Schott et al., 2012] Schott, J. R., Hook, S. J., Barsi, J. A., Markham, B. L., Miller, J., Padula, F. P., and Raqueno, N. G. (2012). Thermal infrared radiometric calibration of the entire landsat 4, 5, and 7 archive (1982-2010). *Remote Sensing of Environment*, 122:41–49.

[Science, 2014] Science, L. (2014). The worldwide reference system. http://landsat.gsfc.nasa.gov/?p=3231.

[Seemann et al., 2006] Seemann, S. W., Borbas, E. E., Li, J., Menzel, P., and Gumley, L. E. (2006). Modis atmospheric profile retrieval algorithm theoretical basis document. Technical report, Co-operative Institute of Meteorological Satellite Studies, University of Wisconsin-Madison, 1225 W. Dayton St. Madison, WI 53706.

[Sellers et al., 1988] Sellers, P., Hall, F. G., Asrar, G., Strebel, D. E., and Murphy, R. E. (1988). The first islscp field experiment (fife). *Bulletin of American Meteorological Society*, 69:22–27.

[Shafran, 2007] Shafran, P. (2007). North american regional reanalysis homepage. http://www.emc.ncep.noaa.gov/mmb/rreanl/.

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

[SSI, 2012] SSI (2012). About modtran. http://modtran5.com/about/index.html.

[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 Meteorology*, 43:363–372.

[Tayolor, 2011] Tayolor, M. (2011). Landsat 7 orbit. http://landsathandbook.gsfc.nasa.gov/orbit_coverage/.

[USGS, 2007] USGS (2007). The universal transverse mercator (utm) grid. http://egsc.usgs.gov/isb/pubs/factsheets/fs07701.html.

[USGS, 2013] USGS (2013). Landsat missions timeline. http://landsat.usgs.gov/about_mission_history.php.

[Vazquez et al., 1997] Vazquez, 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.

[Vining and Blad, 1992] Vining, R. C. and Blad, B. C. (1992). Estimation of sensible heat flux from remotely sensed canopy tempeartures. *Journal of Geophysical Research*, 97(D17):18951–18954.

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

[Wan, 2007] Wan, Z. (2007). Collection-5 modis land surface temperature products users' guide. Technical report, ICESS, University of California, Santa Barbara.

[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., 2004] Wan, Z., Y., Zhang, 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., 2002a] Wan, Z., Zhang, Y., liang Li, Z., 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., 2002b] Wan, Z., Zhang, Y., Zhang, Q., and liang Li, Z. (2002b). Validation of the land-surface 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. http://www.ofcm.gov/fmh3/text/default.htm.

[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.