A NOVEL CONFIDENCE METRIC APPROACH FOR A LANDSAT LAND SURFACE TEMPERATURE PRODUCT

Monica J. Cook and Dr. John R. Schott
Rochester Institute of Technology
Chester F. Carlson Center for Imaging Science
Digital Imaging and Remote Sensing Laboratory
54 Lomb Memorial Dr.
Rochester, NY 14623
mxc7441@rit.edu

ABSTRACT

The Landsat series is the longest body of continuously acquired, moderate resolution satellite imagery. The spatial and temporal resolution and coverage of Landsat make it an intriguing instrument for a land surface temperature product, which is an important earth system data record for a number of fields including climate, weather, and agriculture. Because current archived Landsat imagery has only a single thermal band, generation of a land surface temperature product requires an emissivity estimation and atmospheric compensation. This work, assuming imagery from a characterized and calibrated sensor and integration with ASTER derived emissivity data, focuses on the atmospheric compensation component by using reanalysis data and radiative transfer code to generate estimates of radiative transfer parameters. Along with an estimation of land surface temperature, the goal is to provide a confidence estimation for every pixel in a scene. Using water temperatures from buoy data, actual temperatures have been compared to predicted temperatures as validation of performance. These comparisons have shown acceptable performance when the atmosphere is well characterized, but larger errors when the atmosphere is not as well understood. The reanalysis data, radiative transfer code, bulk to skin temperature conversion, and lack of knowledge of atmospheric variation all complicate traditional error analysis. Various methods have been attempted, including error propagation based on perturbed atmospheres, regressions between metrics and error values, and thresholds based on atmospheric variables. Because of complications not faced by other large-scale products, a novel approach to error analysis will be developed by combining multiple approaches and data sources.

Keywords: Landsat, land surface temperature, thermal

INTRODUCTION

Land surface temperature (LST) is an independently useful Earth system data record, with applications focused mostly around environmental endeavors, including agriculture, climate studies and meteorological research among other fields. It is difficult to directly measure the temperature of a surface without altering that temperature, making remote sensing an optimal method to measure large scale surface temperatures. When there are multiple thermal bands available, it is common to use a split window algorithm that utilizes differential absorption in adjacent thermal bands (Wan and Dozier, 1996). However, with a single thermal band, retrieval of LST requires both an atmospheric compensation and an emissivity component.

Landsat is the longest body of continuously acquired, moderate resolution satellite imagery, with an archive that contains over 30 years of thermal imagery. However, until the launch of Landsat 8 in 2013, historical Landsat satellites have captured a single thermal band. The archived data is calibrated and characterized, but provided as radiance values, which are not immediately intuitively applied. Development of a land surface temperature product for the Landsat archive, from the available radiance data, requires atmospheric compensation and emissivity.

Recently, the Advanced Spaceborne Thermal Emission and Reflection (ASTER) radiometer has been used to develop a high spatial resolution (100 m) surface emissivity product. Currently available for North America, known as the North American ASTER Land Surface Emissivity Database (NAALSED), plans are underway to extend this dataset to global coverage (Hulley and Hook, 2009). Assuming the ability to combine with this product from the Jet Propulsion Laboratory, this work focuses on the generation and error analysis of the atmospheric compensation component.

The temporal resolution (16 day repeat) and spatial resolution (60-120 m resolution resampled to 30 m pixels) of the Landsat series of satellites is well matched to the applications of a long time-series, large scale LST product. There are endless environmental science applications of LST, including terrestrial biosphere dynamics, change detection and hydrologic balances, as well as multiple ecological processes, including leaf phenology, photosynthesis, respiration and decomposition (Running et al., 1994). LST is also important in determining a number of variables, such as energy, momentum, and moisture fluxes at the surface of the Earth, that are used in climate studies and numerical meteorological predictions (Price, 1982). Producing a LST product from the archived, and newly collected, calibrated radiance values, would make a largely under utilized dataset invaluable across a number of fields.

BACKGROUND

Landsat

The original Landsat was launched in July 1972, although Landsats 1, 2, and 3 did not in practice have thermal imaging capabilities. Landsat 4, launched in 1982, was the first to effectively capture thermal data; Landsat has since been continuously capturing thermal imagery. Although Landsat 6 failed to reach orbit, Landsat 5 long outlived its lifetime, taking the family of satellites through concerns of a data gap. Landsats 4, 5 and 7 all captured or capture a single thermal band, roughly within the range of 10 – 13 microns. Landsat 8, launched in February 2013, now captures two thermal bands over a similar spectral range. Over 30 years of data are largely under utilized due to the difficultly in interpreting radiance values.

MODTRAN

The MODerate resolution atmospheric TRANsmission (MODTRAN) radiative transfer code was created by Spectral Sciences Inc. and the United States Air Force. Given atmospheric profiles for a particular place and time, MODTRAN can be used to generate a characterization of the radiometric properties of the atmosphere (Berk et al., 1999); MODTRAN assumes the atmosphere is made up of a number of homogenous layers to make these calculations (Schott, 2007). For our purposes, we input atmospheric profiles for height, temperature, pressure, and relative humidity and use the spectral radiance output to calculate the radiative transfer parameters necessary to solve for LST, transmission, upwelled radiance, and downwelled radiance.

NARR

The necessary atmospheric profile inputs for MODTRAN are extracted from the North American Regional Reanalysis (NARR) data. Reanalysis, or retrospective analysis, inputs observables into numerical models to generate variables in a spatially and temporally regular and consistent dataset that are difficult to observe or measure (Rienecker and Gass, 2013). The NARR dataset was selected for its combination of temporal span (data available beginning 1 January 1979 with plans for continuing coverage) and resolution (eight times daily) and spatial span (all of North America) and resolution (0.3° or roughly 32 km at the lowest latitude). Consistent with the current NAALSED dataset, future work includes extending the same methodology to a global reanalysis dataset with larger coverage but likely lower resolution. Of the variables available in the NARR dataset, we use geopotential height [gpm], temperature [K], and specific humidity [kg/kg] at 29 pressure levels. These profiles are converted to geometric height [km] and relative humidity [%], and along with the unchanged temperature and pressure profiles, input into MODTRAN.

Radiometry and Governing Equation

A governing equation contains all of the components that are relevant to the sensor reaching radiance and is key to understanding how we approach solving the LST. The governing equation for the Landsat thermal band is shown in Equation (1), where $L_{obs\lambda eff}$ is the effective sensor reaching spectral radiance, $L_{T\lambda eff}$ is the effective spectral radiance due to temperature, ε is the surface emissivity of the pixel of interest, τ is the transmission, $L_{u\lambda eff}$ is the upwelled effective spectral radiance, and $L_{d\lambda eff}$ is the downwelled effective spectral radiance.

$$L_{obs\lambda eff} = \left(L_{T\lambda eff}\varepsilon + (1-\varepsilon)L_{d\lambda eff}\right)\tau + L_{u\lambda eff} \quad (1)$$

Effective spectral radiance indicates that the spectral response function of the respective sensor has been accounted for over the spectral range of sensitivity, giving the radiance value units of $Wm^{-2}sr^{-1}\mu m^{-1}$, as shown in Equation 2, where $R(\lambda)$ is the spectral response function of the sensor and the integration is over the spectral sensitivity of the thermal band. For this work, we will use all effective spectral values, and drop the explicit notation from here forward.

$$L = L_{\lambda eff} = \frac{\int L_\lambda R(\lambda) d\lambda}{\int R(\lambda) d\lambda} \tag{2}$$

Considering each variable in our governing equation, the observed radiance is generated from the digital number and calibration coefficients provided in the Landsat metadata, because we are assuming a characterized and calibrated sensor (Barsi et al., 2003), (Padula et al., 2010). We are also assuming incorporation with the ASTER derived emissivity. The radiance due to temperature is the variable we wish to solve for. This can be inverted to the desired temperature using Planck's equation, shown in Equation 3. Planck's equation is not directly invertible, so we use a look up table (LUT) approach to get from radiance to temperature values.

$$L_{T\lambda eff} = \frac{M_{\lambda eff}}{\pi} = \frac{\int [2hc^2\lambda^{-5}(e^{\frac{-hc}{\lambda kT}}-1)^{-1}] R(\lambda) d\lambda}{\int R(\lambda) d\lambda} \tag{3}$$

This leaves only the radiative transfer parameters, transmission, upwelled radiance, and downwelled radiance. These are the values we solve for using the outputs from MODTRAN. Generating these three values for each pixel in a Landsat scene is the problem around which our methodology is based.

METHODOLOGY

Temporal Interpolation

Profiles from the NARR data are provided eight times daily, on regular three-hour intervals. Before these profiles are used in MODTRAN, they are linearly interpolated to the acquisition time of the relevant Landsat scene. For example, the NARR profiles from 12 Z and 15 Z would be linearly interpolated for a Landsat acquisition time of 14.3 Z. These temporal interpolations are performed for each variable at each NARR grid point within the Landsat scene.

Generating Radiative Transfer Parameters

The necessary radiative transfer parameters, transmission, upwelled radiance, and downwelled radiance, are not explicit outputs of MODTRAN. For each MODTRAN run, we input a specific boundary temperature (or corresponding radiance due to temperature based on the spectral response function of the sensor, L_T) and emissivity (ε), and MODTRAN outputs the spectral observed radiance at the sensor for the atmosphere being characterized (L_{obs}). As shown in Equation 4, if $\varepsilon = 1$, the governing equation reduces to a linear equation where the slope is equal to the transmission and intercept is equal to the upwelled radiance.

$$L_{obs} = L_T\tau + L_u \tag{4}$$

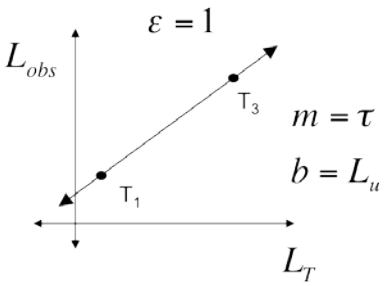


Figure 1. Illustration of regression used to generate transmission and upwelled radiance.

For the first two MODTRAN runs, the emissivity is set to 1 and boundary temperatures of 273 K and 310 K are used. A solution from the output observed radiances results in the effective in band transmission and upwelled radiance values for this particular atmosphere as shown in Figure 1. These temperatures were selected to ensure that the solution is over a large range of possible temperatures. Given the transmission and upwelled radiance, a third and final MODTRAN run with ε < 1 can be used to generate the downwelled radiance by solving the governing equation as shown in Equation 5.

$$L_d = \frac{\frac{L_{obs} - L_u}{\tau} - L_T \varepsilon}{1 - \varepsilon} \tag{5}$$

To ensure that this MODTRAN run is in the temperature range of the current atmosphere, the boundary temperature for this run is set to the air temperature of the lowest atmospheric layer.

These operations are performed at a specific NARR grid point for a specific ground altitude to generate the radiative transfer parameters for that location and elevation. This is repeated at nine separate ground altitudes for each NARR location to have radiative transfer parameters that cover all possible ground elevations in the Landsat scene.

Height Interpolation

At this point in the process, the necessary radiative transfer parameters have been generated at every NARR point within the scene at nine separate heights. It is then necessary to focus on each pixel individually. For each pixel, the four most relevant NARR locations are identified and the elevation of the current pixel is extracted from the digital elevation model (DEM). At each of these four NARR points, only the radiative transfer parameters from the height closest above and closest below the elevation of the current pixel are piecewise linearly interpolated to the height of the pixel. This results in radiative transfer parameters at the four closest NARR points associated with the appropriate elevation.

Spatial Interpolation

Finally, for each pixel, the radiative transfer parameters from the four relevant NARR points must be spatially interpolated to the appropriate location. Shepard's method, an inverse distance weighting function, is used; this function is expressed in Equations 6, 7, and 8.

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{6}$$

$$w_i = \frac{d_i^{-p}}{\sum_{j=1}^{n} d_j^{-p}} \tag{7}$$

$$F(x, y) = \sum_{i=1}^{n} w_i f_i \tag{8}$$

Equation 6 calculates d_i, distance values from each point (NARR location) to the interpolated point (pixel location). Equation 7 calculates a weighting value for each point based on these distances and the number of points, n=4. The weighting exponent, p, is an arbitrary positive real number and generally set to 2. Finally, Equation 8 calculates the interpolated value, F, as a weighted summation of the original values, f. This spatial interpolation is performed for each radiative transfer parameter at each pixel, resulting in transmission, upwelled radiance, and downwelled radiance values for each pixel.

Deliverables

In summary, this methodology generates transmission, upwelled radiance [$Wm^{-2}sr^{-1}\mu m^{-1}$], and downwelled radiance [$Wm^{-2}sr^{-1}\mu m^{-1}$], the three radiative transfer parameters necessary to calculate the surface temperature, at each pixel. We also report the Landsat observed radiance [$Wm^{-2}sr^{-1}\mu m^{-1}$] and elevation [m] for each pixel.

INITIAL RESULTS

Scenes for the initial validation dataset were selected for location, based on the ground truth sites as described below, without regard for atmospheric conditions or cloud cover. In order to capture a variety of atmospheres, and represent the world as realistically as possible, no filters were applied and all scenes were processed. Therefore, we expect a large variety of results. We expect some very large errors in cloudy scenes, some moderately large errors due to cloud contamination, and hopefully a significant number of good points from cloud free scenes. We expect a more robust cloud detection algorithm will be applied before the product is used; therefore, as described below, we use only a preliminary cloud detector and visual analysis to identify clouds and focus on results for clear scenes, where we hope to see a significant portion of good results and an ability to characterize the errors we do encounter.

Ground Truth Sites

As mentioned previously, it is difficult to directly measure LST without altering the temperature of the surface being measured. However, it is much easier to directly measure the temperature of water because tools can be submerged and acclimated to the temperature of the water. The emissivity of water is also well known. Therefore, multiple buoy temperatures were corrected to skin temperature and used as ground truth data as initial validation of this methodology (Schott et al., 2012), (Hook et al., 2007). Buoys in Lake Tahoe and Salton Sea operated by JPL as well as buoys operated by NOAA in Lake Ontario, Lake Huron and Lake Superior were used. NOAA buoys in the oceans off the coast Delaware/Maryland, Georgia, and California were also used. These were selected to provide a range of surface temperatures and atmospheric variability. The initial validation dataset includes 827 points.

Cloud Detection

It is not within the scope of this effort to develop a cloud detection algorithm, but it impossible to accurately retrieve LST when imaging a cloud. A rudimentary and preliminary cloud detector was applied in order to eliminate the most obvious of the cloudy results. If the retrieved temperature is less than 275 K or the difference between the retrieved temperature and air temperature at the lowest level of the atmosphere is greater than 15 K, the scene was discarded from these initial results as a cloud. A visual analysis of the remaining scenes is in progress to identify all cloudy scenes in order to more accurately characterize our initial results and error analysis. After removing obviously cloudy scenes with our preliminary cloud detector, there are 619 scenes left in the validation dataset.

Initial Results

For all scenes in the validation dataset, we calculated the radiance due to temperature by solving the governing equation in Equation 1 using the interpolated radiance transfer parameters for the Landsat pixel at the location of the selected buoy. This radiance due to temperature was inverted to a temperature value [K] and compared to the skin temperature from the buoy. For these initial comparisons, error is calculated as shown in Equation 9.

$$\text{Error [K]} = \text{Predicted LST [K]} - \text{Ground Truth Temperature [K]} \quad (9)$$

This results in a negative error value when the predicted value underestimates the ground truth temperature. These results are illustrated graphically in Figures 2 and 3. Figure 2 is a plot of the predicted apparent LST vs. the ground

truth temperatures. The red line is a one-to-one line, which would illustrate perfect results. As shown, a large number of results are clustered around this one-to-one line, although there are a handful of points with larger errors and more points below than above the line, indicating a slight negative bias.

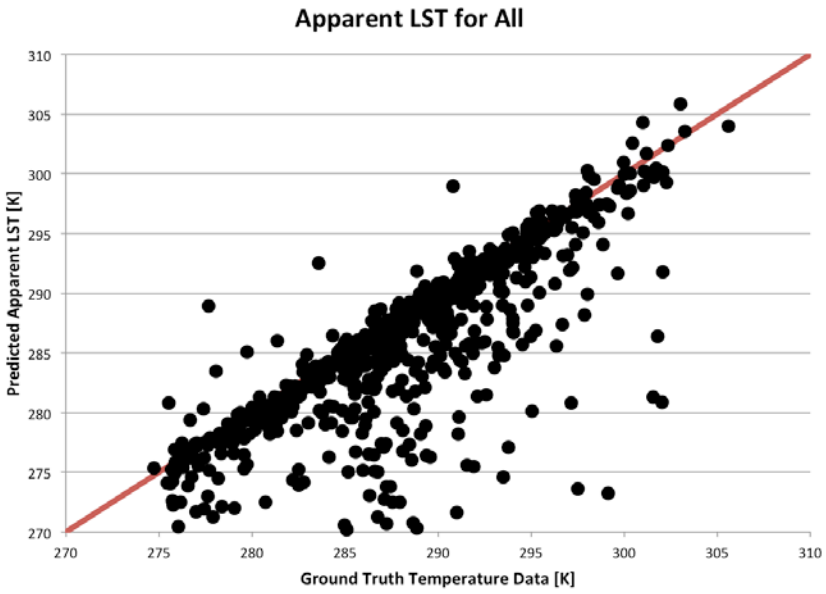


Figure 2. Plot of predicted LST vs. ground truth temperature for all scenes. The red line is a one-to-one line, representing perfect retrieval.

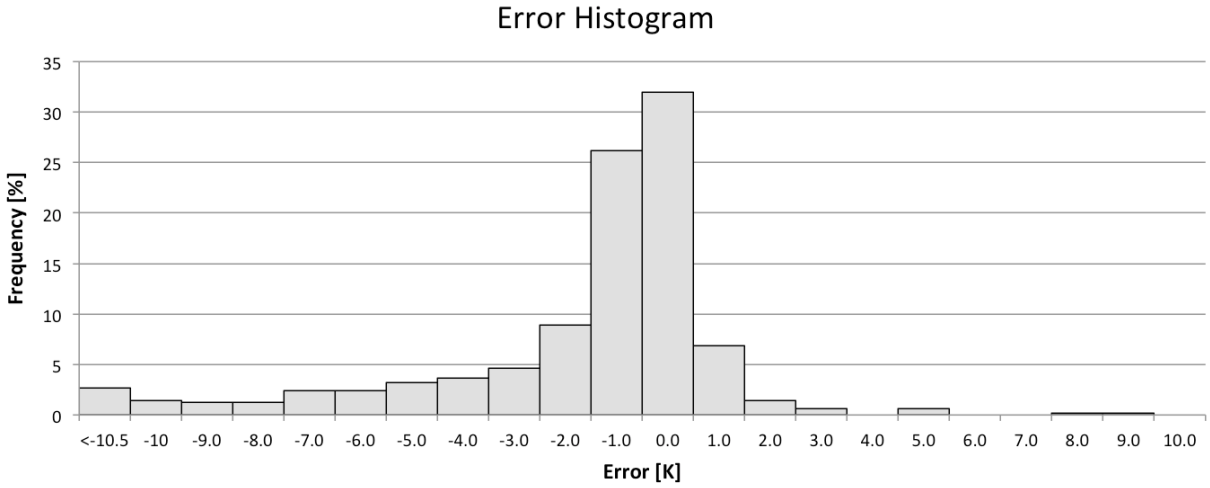


Figure 3. Histogram of error values in initial validation dataset with cloudy scenes removed.

Figure 3 is a histogram of error values with the cloudy scenes removed as determined with our preliminary cloud detector. These results are very encouraging. As shown, more than 65% of the scenes have errors in the range of [-1.5 K, 1.5 K]. These are very acceptable results for most applications of a LST product. We anticipate that some of the scenes with larger negative errors are cloud contaminated scenes; these should be identified during visual analysis and removal will give an even better characterization of our process. Of most concern are scenes with moderately large negative errors, in the range of approximately [-7 K, -3 K]. As also illustrated by the slight negative shift in the center of the histogram, there is water vapor in the atmosphere that is not captured by the reanalysis data. For the goal of this product, with a long time scale and large spatial coverage, we still feel that the

current data source is optimal, but using numerical predictions rather than observables can lead to such omissions of water vapor. Each process and interpolation in our methodology was analyzed with a sensitivity study and we do not feel that adjustments to this methodology will significantly reduce these errors. Because they are likely caused by water vapor not indicated in the profiles, our attention turns to understanding and predicting these errors based on the atmospheric data we have available.

INITIAL ERROR ANALYSIS

Challenges

The steps outlined in the methodology illustrate the various contributors of error in the process, along with any error associated with the input variables. Sensitivity studies were performed to examine the error contributed by each individual step of the methodology (each interpolation as well as generating radiative transfer parameters in MODTRAN), and all were found to be acceptably small. However, how these errors compound together, along with input errors, is difficult to predict.

Also, as discussed above, we predict the largest of our errors can be attributed to clouds, as we have illustrated with our initial cloud detector and intend to solidify with our visual analysis. And the majority of scenes in the center of the histogram have errors acceptable to the applications we are targeting. The errors we are most concerned with are the moderately large underestimations of temperature, which we think can be attributed not to systematic error in the input variables, but water vapor not accounted for in the input variables.

The following is a discussion of the implementation and success or failure of the error assessment methodologies we have considered thus far. The goal is to attribute for end users, qualitatively or quantitatively, a confidence in the atmospheric compensation to each pixel in the scene.

Traditional Error Propagation

We did consider a traditional error propagation, as discussed in Hook et al. (2007). The errors in the input variables, the NARR profiles, were estimated and propagated through the process by perturbing the atmosphere and conducting numerical simulations to determine the effect on the final predicted LST (Hook et al., 2007). The error in the radiance due to temperature, with contributions from each variable, is shown in Equation 10.

$$S_{L_T} = \sqrt{\left(\frac{\partial L_T}{\partial \tau} S_\tau\right)^2 + \left(\frac{\partial L_T}{\partial L_u} S_{L_u}\right)^2 + \left(\frac{\partial L_T}{\partial L_d} S_{L_d}\right)^2 + 2\rho_{\tau L_u} \frac{\partial L_T}{\partial \tau}\frac{\partial L_T}{\partial L_u} S_\tau S_{L_u} + 2\rho_{\tau L_d} \frac{\partial L_T}{\partial \tau}\frac{\partial L_T}{\partial L_d} S_\tau S_{L_d} + 2\rho_{L_d L_u} \frac{\partial L_T}{\partial L_d}\frac{\partial L_T}{\partial L_u} S_{L_d} S_{L_u}} \quad (10)$$

Note that error contributions from the observed radiance and emissivity will be important in the final product but are not considered here as they are not part of the atmospheric compensation. The partial derivatives in Equation 10 are determined by taking the partial derivatives of Equation 1. The error contribution from transmission (S_τ) is shown in Equation 11. Error contributions from upwelled radiance and downwelled radiance are determined in a similar fashion.

$$S_\tau = \sqrt{\left(\frac{\partial \tau}{\partial T} S_T\right)^2 + \left(\frac{\partial \tau}{\partial RH} S_{RH}\right)^2 + \left(\frac{\partial \tau}{\partial P} S_P\right)^2 + \left(\frac{\partial \tau}{\partial H} S_H\right)^2} \quad (11)$$

Contributions from pressure and height were not considered because of the fixed nature of the pressure levels and direct relation to the height. The error contributions from temperature and relative humidity (S_T and S_{RH}) are inherent to the NARR data, but were difficult to determine. These were estimated from error values in atmospheric profile products. Finally, the partial derivatives in Equation 11 were determined by perturbing each atmosphere in a carefully selected subset and using MODTRAN to determine how much each perturbation affected each radiative transfer parameter. For example, each temperature profile was shifted 5 K in 1 K increments; the change in transmission was noted each time. The relationship between the change in profile and change in radiative transfer parameter was acceptably linear for each single atmosphere, so the slope of this relationship was taken as the partial

derivative; however, this slope varied over the variability of atmospheres in the subset so these partials were defined as functions of air temperature and column water vapor. This sample of atmospheres was selected to cover a wide range of conditions. The correlations were also calculated from these numerical simulations.

Using all of these components, an error in radiance was predicted for each point in the validation dataset; this radiance was inverted to a predicted error in temperature. These were compared to the actual error values as calculated by comparison to the ground truth data. A plot of these comparisons is shown in Figure 4.

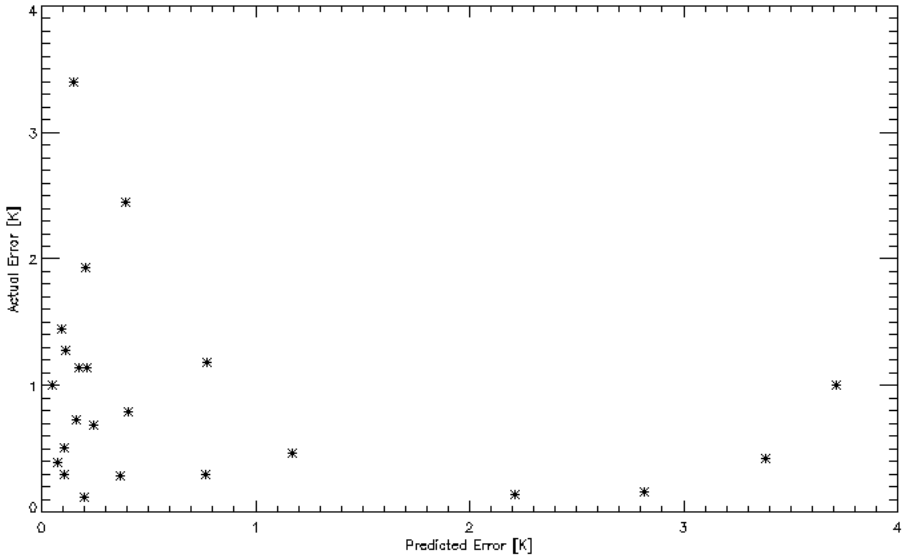


Figure 4. Comparison of actual and predicted error values from the traditional error analysis.

As shown, this method did not prove to adequately predict the larger errors in our process. There are a handful points that are reasonable when both the actual and predicted errors are small (< 1K), but we never accurately predict larger errors and we miss in both directions, sometimes inappropriately predicting large errors (when we are overly pessimistic about our knowledge of input variables) and more importantly missing larger actual errors with small predictions. We overestimate the errors in our input variables in an attempt capture larger errors in the process; this makes us pessimistic in the wrong cases and still unable to characterize the larger errors we are concerned about. The inability to propagate error from the input variables supports our prediction that when we largely underestimate the temperature, the sensor is imaging components of the atmosphere not accounted for in the NARR data.

Metrics

After concluding that traditional error propagation would not adequately predict error, we look to extract more information from atmospheric data we already use in the process. Although we believe that our errors are attributed to water vapor we are missing in our analysis, we hope that by considering the relationship between various individual metrics and error we can discern patterns that would allow us to identify those scenes where we may have difficulty predicting the LST.

We consider five variables from our current analysis. For each scene, we consider the calculated transmission. We expect that as transmission decreases, errors increase. We also consider the maximum air temperature in the atmospheric profile. A higher air temperature tends to indicate a thicker, denser atmosphere with high upwelled radiance, which can be a large contributor to error. Therefore, as air temperature increases, we expect higher errors. The final three metrics all deal with water vapor in the atmospheric profile. The relative humidity at each pixel was considered. We are less optimistic as this is directly related to the traditional error analysis as described above, but we expect higher errors with higher reported relative humidity values. The dew point depression, the difference between the dew point temperature and air temperature, is directly related to relative humidity but considered because it provides an inverse relationship. As dew point depression increases, we expect the error to decrease. Finally, we consider the column water vapor; over each square centimeter of ground surface, the water molecules in that column have some mass (gm/cm^2); because the density of liquid water is 1 gm/cm^2, this mass is equal to the centimeters of water on the ground if all of the water rained out of the atmosphere. This is the column water vapor

[cm], extracted from the MODTRAN output. As column water vapor increases, we expect the magnitude of our errors to increase. Each of these metrics were considered in the methods discussed below.

Regression Analysis

As discussed above, we expect a certain relationship between each metric and the error. The goal was to be able to explicitly determine this relationship and use regression to predict the error for each point. We plotted and considered the relationship between each metric and the actual errors. In general, a strong relationship was not illustrated and a regression over these errors did not yield accurate predictions of error. One example is shown below. Figure 5 is the actual error plotted against the maximum air temperature for each atmosphere over a subset of locations. As shown, there is not a strong relationship. The line shown represents a linear regression for all points.

No metric had a strong enough relationship to determine a solution that could accurately predict errors across a range of conditions. This failure is not surprising as this method characterizes and captures the same types of errors used in the error propagation method. We took a different approach, the threshold analysis described below, to attempt to identify bad conditions in order to avoid opportunities for larger errors.

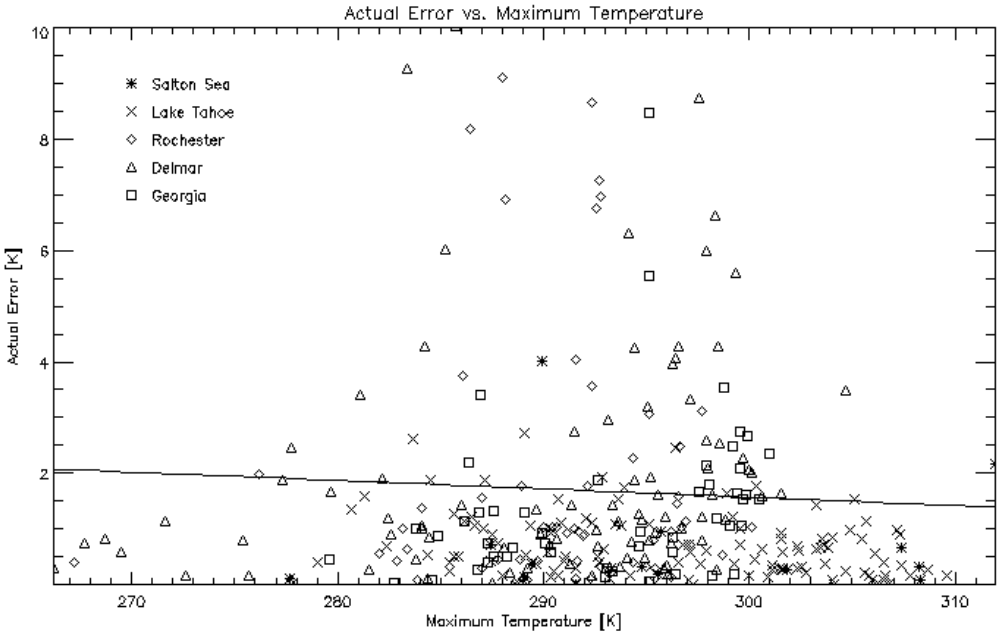


Figure 5. Actual error vs. maximum air temperature for a subset of locations. The line is a linear regression for these points.

Threshold Analysis

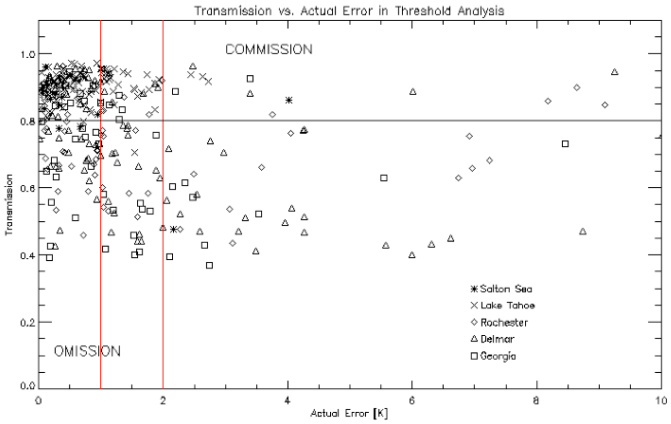
While no relationship was strong enough to provide a solution that could accurately predict errors, we considered similar relationships in a more qualitative sense. Rather than assigning each pixel a quantitative error value, we aimed to assign each pixel a confidence based on how well we think we can predict LST for that pixel. For each metric, we set a threshold above or below which we would say we did or did not have confidence in that pixel. The goal is to identify conditions where errors are more likely to be large.

Analyzing this error metric became more difficult. We can compare a quantitative error prediction to an actual error quantity, but we need to compare a qualitative confidence estimation to a qualitative analysis of each point. Therefore, we performed analysis with both 1 K and 2 K standards as acceptable error. For example, we set our transmission threshold to 0.8. We would predict high confidence for any scene with $\tau > 0.8$ and low confidence for any scene with $\tau < 0.8$. As shown in Table 1, assigning a high confidence to larger errors is an error of commission and assigning a low confidence to smaller errors is an error of omission, where large and small errors are determined by the standard. While Table 1 illustrates results for a standard of 1 K, we also did analysis with a standard of 2 K. We are more concerned with errors of commission as this results in trusting bad data. Figure 6 shows transmission plotted vs. actual error for a subset of locations with both the threshold and standards illustrated on the plot.

Table 1. Categories and results of threshold analysis for transmission.

TRANSMISSION	CONFIDENCE	ERROR	RESULT
$\tau > 0.8$	High	error < 1 K	Accurate
$\tau > 0.8$	High	error > 1 K	Commission
$\tau < 0.8$	Low	error > 1 K	Accurate
$\tau < 0.8$	Low	error < 1K	Omission

Figure 6 shows relatively small errors of commission, particularly with a 2 K standard. The average error for points assigned high confidence is also encouraging. Analysis of other metrics showed similar results, although transmission, relative humidity and dew point depression had the most promising results.



Omission, 1 K	14.57%
Commission, 1 K	14.82%
Omission, 2 K	24.12%
Commission, 2 K	4.27%
Mean Error in HC	1.181 K
SD in HC	2.150 K

Figure 6. Illustration of threshold analysis for transmission.

While this is not yet robust enough to provide a confidence analysis for our final process, these results are encouraging and provide a number of avenues for improvement, including a better method of threshold determination, tiered threshold levels with varying degrees of certainty, and combining information from multiple metrics.

FUTURE WORK

Current work focuses on extending this threshold analysis to a feasible method of confidence estimation for the final product. This requires acceptable and consistent results and a more complete understanding of the causes of these errors. Multiple thresholds for multiple levels of confidence is one aspect being considered. This is illustrated in Table 2. Considering only transmission, points with $\tau > 0.9$ are given high confidence, points where $0.8 < \tau < 0.9$ are given medium confidence, and when $\tau < 0.8$, points are given low confidence.

Table 2. Results for a tiered threshold approach with multiple levels of confidence.

	High	Medium	Low
Mean Error	0.8169 K	2.2679 K	2.8748 K
Standard Deviation	1.034 K	3.1721 K	3.4349 K
Percent Error	4.03% commission	9.03% commission	27.42% omission

Mean error increases as confidence decreases and then errors of commission are encouraging. Note there are still large errors of omission in the low confidence category. Adjustment of the locations of the threshold still needs to be considered. Combining multiple metrics in multiple levels is another option for utilizing more information. There are numerous combinations to be considered, including combinations of two, three, four or all five metrics, and adjustments to the thresholds for each metric. The current nomenclature is still rudimentary and could be made both more descriptive and more robust. The goal is to develop a methodology, using multiple metrics, that could assign not only a confidence value but also a predicted mean and standard deviation of errors for points in each

categorization. Without being able to assign a quantitative error value to each pixel, we want to be able to assign a qualitative categorization that has some expected error value associated with it.

Another step is to consider other tools that can give a better understanding of what is causing current errors. Although they may not span the lifetime of the Landsat series, comparing against other satellite data, such as MODIS data or even new and different data from Landsat 8, could provide a better understanding of the errors we are seeing and give insights for future aspects of the product.

ACKNOWLEDGEMENTS

We would like to thank NASA and JPL for funding this work.

REFERENCES

Barsi, J., J. Schott, F. Palluconi, D. Helder, S. Hook, B. Markhan, G. Chander, and E. O'Donnel, 2003. Landsat TM and ETM+ thermal band calibration. *Can. J. Remote Sensing*, 29(2):141-253.

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

Hook, S. J., G. Vaughan, H. Tonooka, and S. G. Schladow, 2007. Absolute radiometric in-flight validation of mid infrared and thermal infrared data from ASTER and MODIS on the TERRA spacecraft using Lake Tahoe, CA/NV, USA, automated validation site. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1798-1807.

Hulley, G. C. and S. J. Hook, 2009. The north American ASTER land surface emissivity database (NAALSED) version 2.0. *Remote Sensing of Environment*, 113:1967-1975.

Padula, F. and J. Schott, 2010. Historic calibration of the thermal infrared band of Landsat 5 TM. *Photogrammetric Engineering and Remote Sensing*, 76(11):1225 – 1238.

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

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

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

Schott, J. R., 2007. *Remote Sensing: The Image Chain Approach*. Oxford University Press, 2nd Edition.

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

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