

2016

## Sensitivity of evapotranspiration retrievals from the METRIC processing algorithm to improved radiometric resolution of Landsat 8 thermal data and to calibration bias in Landsat 7 and 8 surface temperature

Ayse Kilic

University of Nebraska-Lincoln, [akilic@unl.edu](mailto:akilic@unl.edu)

Richard G. Allen

University of Idaho, [RALLEN@uidaho.edu](mailto:RALLEN@uidaho.edu)

Ricardo Trezza

University of Idaho Research and Extension Center, Kimberly, [rtrezza@uidaho.edu](mailto:rtrezza@uidaho.edu)

Ian Ratcliffe

University of Nebraska-Lincoln, [iratcliffe2@unl.edu](mailto:iratcliffe2@unl.edu)

Baburao Kamble

University of Nebraska-Lincoln, [bkamble3@unl.edu](mailto:bkamble3@unl.edu)

Follow this and additional works at: <https://digitalcommons.unl.edu/natrespapers>



Part of the [Journal of the American Water Resources Association](#), [Conservation Commons](#), [Natural Resources Management and Policy Commons](#), and the [Other Environmental Sciences Commons](#)

---

Kilic, Ayse; Allen, Richard G.; Trezza, Ricardo; Ratcliffe, Ian; Kamble, Baburao; Robison, Clarence W.; and Ozturk, Doruk, "Sensitivity of evapotranspiration retrievals from the METRIC processing algorithm to improved radiometric resolution of Landsat 8 thermal data and to calibration bias in Landsat 7 and 8 surface temperature" (2016). *Papers in Natural Resources*. 1514.  
<https://digitalcommons.unl.edu/natrespapers/1514>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

## Authors

Ayşe Kilic, Richard G. Allen, Ricardo Trezza, Ian Ratcliffe, Baburao Kamble, Clarence W. Robison, and Doruk Ozturk

# Sensitivity of evapotranspiration retrievals from the METRIC processing algorithm to improved radiometric resolution of Landsat 8 thermal data and to calibration bias in Landsat 7 and 8 surface temperature

Ayşe Kilic,<sup>1</sup> Richard Allen,<sup>2</sup> Ricardo Trezza,<sup>2</sup> Ian Ratcliffe,<sup>3</sup>  
Baburao Kamble,<sup>3</sup> Clarence Robison,<sup>2</sup> & Doruk Ozturk<sup>3</sup>

<sup>1</sup> Dept. Civil Engineering and School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

<sup>2</sup> College of Agriculture and Life Sciences, University of Idaho, Kimberly, ID 83341, USA

<sup>3</sup> School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

*Corresponding author* — R. Allen, rallen@uidaho.edu

*E-mail addresses* — akilic@unl.edu (A. Kilic); rtrezza@uidaho.edu (R. Trezza); iratcliffe2@unlnotes.unl.edu (I. Ratcliffe); baburaokamble@gmail.com (B. Kamble); robison@uidaho.edu (C. Robison); dozturk89@gmail.com (D. Ozturk).

## Abstract

We made an assessment on the use of 12-bit resolution of Landsat 8 (L8) on evapotranspiration (ET) retrievals via the METRIC process as compared to using 8-bit resolution imagery of previous Landsat missions. METRIC (Mapping Evapotranspiration at high Resolution using Internalized Calibration) is an ET retrieval system commonly used in water and water rights management

---

Published in *Remote Sensing of Environment* 185 (2016), pp 198–209.

doi:10.1016/j.rse.2016.07.011

Copyright © 2016; published by Elsevier Inc. Used by permission.

Submitted 21 July 2015; revised 8 June 2016; accepted 7 July 2016; published 30 July 2016.

where the surface energy balance process is coupled with an extreme- and point calibration process to remove most impacts of systematic bias in remotely sensed inputs. We degraded L8 thermal images by grouping sequential digital numbers to reduce the apparent numerical resolution and then recomputed ET using METRIC and compared to nondegraded ET products. The use of 8-bit thermal data did not substantially impair the accuracy of ET retrievals derived from METRIC, as compared to the use of 12-bit thermal data. The largest error introduced into ET was  $<1\%$ . We also compared ET retrieved from images processed during the L8 and Landsat 7 (L7) March 2013 underfly to assess differences in ET caused by differences in signal to noise ratio (SNR) and scaling of the two systems.

We evaluated the impact of bias in land surface temperature (LST) retrievals on ET determination using the CIMEC calibration approach (Calibration using Inverse Modeling using Extreme Member Calibration) employed in METRIC by introducing globally systematic biases into LST retrievals from L7 and L8 and comparing to ET from non-biased retrievals. The impacts of the introduction of both additive and multiplicative biases into surface temperature on ET were small for the three regions of the US studied, and for both L7 and L8 satellite systems. An independent study showed that METRIC-produced ET compared to within 3% of measured ET for the California site. The study assessed the impact of the February 2014 recalibration of L8 thermal data that caused a 3 K downward shift in LST estimation and changed reflectance values by about 0.7%. We found that the use of the recalibrated LST and shortwave data sets in METRIC did not change the accuracy of ET retrievals due to the automatic compensation for systematic biases employed by METRIC.

## 1. Introduction

Many satellite-based evapotranspiration (ET) retrieval procedures estimate ET via a surface energy balance by keeping account of total net shortwave and longwave radiation at the vegetation or soil surface, the amount of heat conducted into soil, and the amount of heat convected into the air above the surface. The residual computed after differencing those three terms represents the amount of energy absorbed during the conversion of liquid water to vapor, which is ET. A utility of using energy balance over vegetation-based methods is that actual ET rather than potential ET is computed so that reductions in ET stemming from stresses caused by disease, salinity, or shortage of soil moisture are captured. Energy balance methods employ various

strategies for transforming measured surface temperature into the net radiation, soil heat flux and sensible heat flux components. Most satellite-based energy balance methods are highly influenced by the thermal signal from the land surface (Anderson, Allen, Morse, & Kustas, 2011).

The NASA Sustained Land Imaging (SLI) program was initiated in 2013 with the objective of reducing costs for future Landsat-like missions to support a sustainable, long-lived Landsat program at reduced cost. One variable impacting future costs is the number of data bits used to quantize, store and transmit sensed images. Another variable highly relevant to thermal-based ET retrievals is the absolute accuracy required for thermal imagers. The following investigations were initiated to provide information on accuracy requirements for future Landsat missions that may impact Evapotranspiration (ET) estimations used in water resources studies and water rights management. The investigations utilized the METRIC ET procedure that is commonly used in water and water rights management and has properties that are somewhat resilient to biases in thermal data.

## **2. The METRIC ET procedure**

METRIC (Mapping EvapoTranspiration at high Resolution with Internalized Calibration) is an evapotranspiration (ET) retrieval system developed at the University of Idaho that uses Landsat satellite data to compute and map ET (Allen, Tasumi, & Trezza, 2007a; Kilic (Irmak) et al., 2011a). METRIC calculates ET as a residual of the surface energy balance, where the energy balance is internally calibrated using ground-based reference ET. The internal calibration reduces computational biases inherent to remote sensing-based energy balance, including calibration biases in radiance or atmospheric correction (Allen et al., 2007a). The reference ET is based on local weather or gridded weather data sets. Slope and aspect functions and temperature lapsing are used for application to mountainous terrain (Allen, Trezza, Kilic, Tasumi, & Li, 2013). METRIC estimates near-surface temperature gradients used in heat convection using indexed functions of radiometric surface temperature, thereby eliminating the need for absolutely accurate surface temperature and the need for air-temperature measurements.

METRIC algorithms are designed for relatively routine application by trained professionals who possess a familiarity with energy balance and basic radiation physics. The primary inputs for the model are short-wave and long-wave thermal images from a satellite e.g., Landsat and MODIS, a digital elevation model (DEM), and ground-based weather data measured within or near the area of interest.

ET images via METRIC provide the means to quantify ET on a field-by-field basis in terms of both the rate and spatial distribution. In the decade since the introduction of METRIC, the application has been adopted for use in Montana, California, New Mexico, Utah, Wyoming, Texas, Nebraska (Allen, Tasumi, & Trezza, 2007b; Kilic et al., 2011b) and Colorado (Burkhalter et al., 2013), Nevada (Morton et al., 2013), and Oregon (Trezza, Allen, Robison, & Kjaersgaard, 2011). The mapping method has enabled those states to negotiate water rights with Native Americans, assess impacts of water transfers from agriculture to urban uses, manage aquifer depletions, monitor water rights compliance, and manage endangered species.

### **3. Landsat data characteristics**

Landsat 8 sensors were designed by NASA to achieve radiometric performance on the operational land imager (OLI) to an uncertainty of <5% in terms of absolute, at-aperture spectral radiance and to an uncertainty of <3% in terms of top-of-atmosphere (TOA) spectral reflectance for each of the spectral bands (Irons, Dwyer, & Barsi, 2012). The OLI signal-to-noise ratio (SNR) specifications were set higher than ETM+ of Landsat 7 and were based on results from the Advanced Land Imager (ALI) mission. Commensurate with the higher SNR ratios, OLI quantizes data to 12 bits as compared to the eight-bit data produced by the TM and ETM+ sensors. As a result, the OLI band SNR ratios exceed those achieved by the Landsat ETM+ by a factor of at least eight (Irons et al., 2012). The greater 12-bit quantization permits improved measurement of subtle variability in surface conditions and finer resolution under low reflectance conditions. The dynamic range of the OLI was improved compared to previous Landsat sensors, reducing band saturation over highly reflective surfaces such as snow or clouds (Roy et al., 2014).

Standard data products are processed by USGS EROS for Landsat 5, 7 and 8 using the Level 1 Product Generation System (LPGS) to produce GeoTIFF formatted images using Cubic Convolution (CC) resampling and 30-m(TM, ETM+) pixel size. Landsat scenes are processed to Standard Terrain Correction (Level 1T -precision and terrain correction) when sufficient clearness exists to observe ground control points (GCP). The L1 T products are defined in the Universal Transverse Mercator (UTM) map projection with WGS84 datum ([http://landsat.usgs.gov/Landsat\\_Processing\\_Details.php](http://landsat.usgs.gov/Landsat_Processing_Details.php) ).

Landsat 8 data are nominally processed into 185 km × 180 km Level 1 terrain-corrected (L1T) products. The 100m TIRS bands are resampled by cubic convolution to 30m and co-registered with the 30m OLI spectral bands. Calibration coefficients for all Landsat sensors are configured to globally maximize the range of land surface radiance in each spectral band (Markham, Goward, Arvidson, Barsi, & Scaramuzza, 2006).

The Landsat 8 L1T product has a 90% confidence level OLI to TIRS coregistration uncertainty requirement of <30 m and a circular geolocation error uncertainty requirement of <12 m (Irons et al., 2012). These geometry improvements enable more accurate multi-temporal change detection (Townshend, Justice, Gurney, & McManus, 1992).

Shortly after launch the Landsat 8 orbit was configured to under-fly Landsat 7 to provide a vicarious cross-calibration opportunity. The Landsat 7 ETM+ calibration is relatively well defined, with 5% absolute reflective band calibration uncertainty (Markham & Helder, 2012; Roy et al., 2014) and thermal band uncertainties of approximately 0.6 K when expressed as a change in apparent temperature of a 300 K surface (Schott et al., 2012; Roy et al., 2014). Initial results for OLI indicated absolute calibration to be within the pre-launch specification OLI requirements of 3% reflectance and 5% radiance (Czapla-Myers, Anderson, & Biggar, 2013 and Markham, Irons, & Storey, 2013). Currently the TIRS data show an approximate 2% (band 10) and 4% (band 11) bias in absolute radiance when compared to vicarious measurements from buoys in Lake Tahoe, the Salton Sea, and deep oceans (Reuter et al., 2015; Schott, 2015, personal communication). This bias translates to an approximately 2 K to 4 K over estimate when expressed as a change in the apparent temperature of a 300 K surface and exceeds the 2% accuracy requirements for TIRS. Stray light from

beyond the nominal 15° TIRS field of view has been identified as the cause of the bias and a correction approach is being developed by USGS and NASA. Until these studies produce a more comprehensive correction, the USGS produced an initial reprocessing of Landsat 8 TIRS data beginning February 2014 to subtract 0.29W/m<sup>2</sup>/sr/μm from TIRS band 10 and 0.51W/m<sup>2</sup>/sr/μm from TIRS band 11 to improve the accuracy of the data products for surface temperatures typical of mid-latitudes during the growing season. On-orbit assessments indicate that OLI and TIRS meet their geometric performance requirements with wide margins (Markham et al., 2013).

#### 4. Objectives

Objectives of this study were to:

1. **Evaluate the impact of using 8-bit vs. 12-bit numerical (radiometric) resolution.** We degraded the L8 thermal image by grouping sequential digital numbers (DN's) from L8 to reduce the apparent numerical resolution and then recomputed ET. The grouping simulated moving from a 12 bit L8 to 11 bit (by grouping two adjacent DN's) to 10 bit (grouping 4 adjacent DN's), to 9 bit (grouping 8 adjacent DN's) and to 8 bit (grouping 16 adjacent DN's).
2. **Evaluate the impact of systematic biases in surface temperature on ET retrievals.** We introduced globally systematic biases into LST retrievals from L7 and L8 during the March 2013 underfly, recomputed ET, and compared to ET from non-biased LST.
3. **Evaluate the impact of uncertainty in calibration accuracy of satellite data.** We compared relative ET retrievals from METRIC using the new calibration coefficients for the L8 thermal and shortwave bands (available in L8 products created after Feb. 3, 2014) with ET retrievals based on using the pre-Feb. 3, 2014 calibration coefficients.

## 5. General methodology

The METRIC model was applied to L7 and L8 images using shortwave and thermal band information contained in L1T packages downloaded from the USGS Earth Explorer web site. The ET values obtained from these images served as a baseline for subsequent sensitivity analyses on ET. ET was expressed in terms of the Fraction of alfalfa-reference evapotranspiration ( $ET_rF$ ), to facilitate showing relative differences in ET values. The alfalfa-reference ET represents a near maximum level of ET from a fully-covered, well-watered vegetated surface and was computed using the ASCE standardized Penman-Monteith equation (ASCE, 2005). The  $ET_rF$  is fully defined in Allen et al. (2007a) and scales the reference ET between 0 and approximately 1.0.  $ET_rF$  is a primary product of the METRIC process.

During the sensitivity analyses, METRIC was recalibrated and rerun using modified inputs to simulate differences in radiometric accuracy and resolution. During calibration of METRIC, identical assignments of  $ET_rF$  values for the two extreme conditions used in METRIC internal calibration,  $ET_{rF_{cold}}$  and  $ET_{rF_{hot}}$  were made, as were the calibration pixel locations, as compared to the baseline runs. This facilitated evaluation of changes in  $ET_rF$  caused solely by changes introduced into the Landsat sensor data sets.

A series of sample points were selected for creating the graphics and statistics. In most cases, these points were near field centers 1/3 of the way from boundaries of agricultural fields to avoid thermal contamination near field edges. Samples included an assortment of agricultural, rangeland, wetlands, and forest. Approximately 1000 locations were sampled for each study.

The METRIC model was first run with baseline data and then run a second time using adjusted data. The error statistics reported were:

- a) Scatter plots of  $ET_{rF}$  from adjusted data vs.  $ET_{rF}$  from original data.
- b) Statistics, in terms of  $ET_rF$  including:
  - Root Mean Squared Error (RMSE)
  - Mean absolute error (MAE)
  - Coefficient of determination ( $R^2$ )

The RMSE and MAE statistics are calculated as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{ET}_r\text{F}_s - \text{ET}_r\text{F}_b)^2}{n}} \quad (1)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |\text{ET}_r\text{F}_s - \text{ET}_r\text{F}_b|}{n} \quad (2)$$

where  $\text{ET}_r\text{F}_b$  represents the  $\text{ET}_r\text{F}$  values for baseline METRIC runs and  $\text{ET}_r\text{F}_s$  are the  $\text{ET}_r\text{F}$  values from subsequent runs where thermal bands were modified to simulate different radiometric biases and resolution accuracies.

### 5.1. Study areas

Three study areas were evaluated for one or more of the three tests. The SE California study area centered on the Palo Verde Irrigation District surrounding Blythe, CA and utilized L7 and L8 data sets for WRS Path 38 Row 37 collected on March 29, 2013. March 29 was during the “underfly” of L8 with L7 during the March 28–29, 2013 period. The Louisiana study area was contained in WRS Path 22 Row 39 and focused on agricultural areas north of New Orleans. The image processed was collected on March 29, 2013 during the L7 and L8 underfly. The Nebraska study area was contained in WRS Path 30 Row 32 and was used to evaluate Landsat 8 data acquired on 6/17/2013, 7/2013, and 8/20/2013 and Landsat 7 data acquired on 6/9/2013 and 7/11/2013.

For each image date, the Landsat data were run through the METRIC model to produce  $\text{ET}_r\text{F}$ . These model runs were used as baseline data for comparison with model runs with introduced LST bias or reduced radiometric resolution. The L1T data sets were obtained following the reprocessing of L8 data sets by the USGS to account for recalibration (post-February 3, 2014).

### 5.2. The CIMEC calibration process of METRIC

A widely used calibration approach in operational remote sensing models is the ‘CIMEC’ approach (“calibration using inverse modeling

of extreme conditions”) (Allen et al., 2007a). CIMEC is used to calibrate around uncertainties and systematic biases in satellite based energy balance components. CIMEC-based ET models include SEBAL (Bastiaanssen, Menenti, Feddes, & Holtslag, 1998; Bastiaanssen et al., 2005) and METRIC (Allen et al., 2007a, 2007b). The CIMEC process frees these models from systematic bias in the surface temperature and surface reflectance retrievals. Models that use absolute temperature and assumed air temperature fields are more susceptible to biases in each of those parameter fields.

METRIC directly estimates a near surface air temperature difference using two temperatures,  $T_1$  and  $T_2$ , assigned to two levels  $z_1$  and  $z_2$  via CIMEC-type inversion of the function for sensible heat flux,  $H$ , at two known evaporative conditions in the model. The function for  $H$  includes an aerodynamic resistance term between  $z_1$  and  $z_2$  that varies with estimated surface roughness and buoyancy caused by  $H$ , which, in turn, vary spatially within an image. The two evaporation conditions are referred to as the hot and cold conditions or as the wet and dry conditions. Detail on METRIC calibration is given in Allen et al. (2007a) and Kilic et al. (2011a).

In METRIC, the satellite-based energy balance is internally calibrated at the two extreme conditions (dry and wet) using alfalfa-based reference ET ( $ET_r$ ) computed from hourly weather data. Accuracy and dependability of the  $ET_r$  estimate has been established by lysimetric and other studies (ASCE, 2005). The internal calibration of the sensible heat computation within METRIC and the use of the indexed temperature gradient eliminate the need for atmospheric correction of reflectance and land surface temperature (LST) measurements using radiative transfer models (Tasumi et al., 2005). METRIC overcomes those biases by focusing internal calibration on  $H$  to absorb intermediate estimation and data biases. The CIMEC approach has been shown to be effective in removing systematic biases in input data created by radiometric bias and atmospheric attenuation (Allen et al., 2007a, 2007b). The internal calibration also reduces impacts of systematic biases in estimation of aerodynamic stability correction and surface roughness.

The calibration of the sensible heat process equations, and in essence the entire energy balance, to  $ET_r$  corrects the surface energy balance for lingering systematic computational biases associated

with empirical functions used to estimate some components and uncertainties in other estimates as summarized by Allen et al. (2007a, 2007b) and Kilic (Irmak) et al. (2011a). The end result is that systematic biases inherent to net radiation, soil heat flux and subcomponents of  $H$  are substantially reduced by the subtraction of a bias-canceling estimate for  $H$ . The biases include those caused by bias in LST.

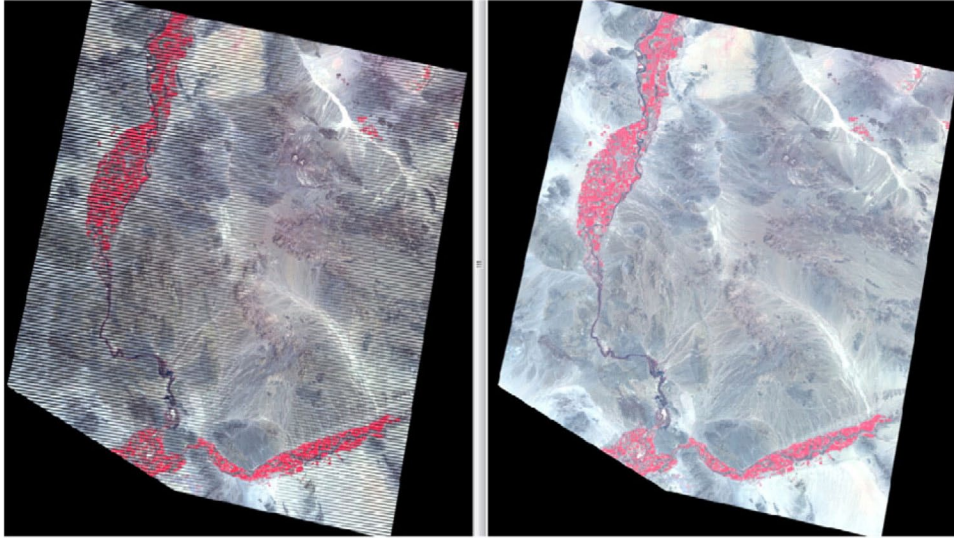
## 6. Comparison of $ET_rF$ during the L8/L7 underfly

A side-by-side comparison of  $ET_rF$  produced using METRIC was made using L8 and L7 satellite imagery during the March 29, 2013 underfly. On that date, images were collected only 6.5 min apart. The intent of the comparison was to quantify differences in  $ET_rF$  caused by differences in thermal sensor resolution, georegistration and data handling for the two Landsat platforms when the land surface characteristics, atmospheric, and solar conditions were essentially identical. Those differences were used as a basis for assessing the impacts of biases and radiometric resolution introduced in a later section.

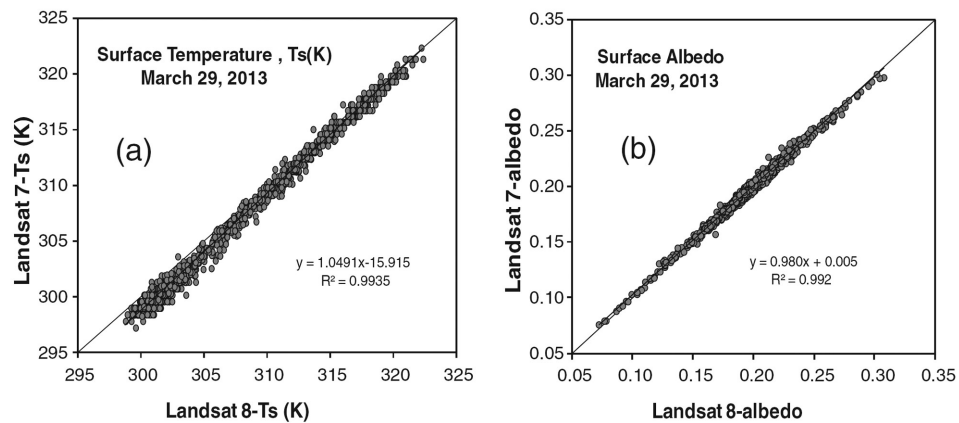
The satellite images from path 38 row 37 when L8 was copositioned with L7 provided near-coincident data collection from both satellites, with scene center Greenwich overpass times of 18:06:14.0 for L7 and 18:12:53.4 for L8 for the Palo Verde study area and 16:38:20 for L7 and 16:35:31 for L8 for the Louisiana study area. **Fig. 1** shows the Palo Verde irrigated study area as red color, located in the upper left of the L7 and L8 satellite scenes.

METRIC was applied to both L7 and L8 images. The assignment of  $ET_{rFcold}$  and  $ET_{rFhot}$  (in this case 1.05 and 0.1) to the cold and hot pixel locations was identical between the two images and produced the same  $ET_rF$  at the CIMEC endpoints that were selected manually. Hourly weather data from the Blythe CIMIS weather station were used to calculate alfalfa reference ET and to provide the weather information needed to set up and calibrate METRIC.

**Fig. 2** shows a comparison of LST and surface albedo calculated for the Landsat underflight of March 29, 2013. LST was computed using Band 10 for L8 and using Band 6 for L7 using coefficients provided in the metafiles. Thermal emissivity was estimated from normalized difference vegetation index. Path radiance in the 10.4–12.5  $\mu\text{m}$  region was estimated to be  $0.91\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$ , downward thermal



**Fig. 1.** Area of interest applied during processing of METRIC during the 03/29/2013 L8/L7 underfly for path 38 row 37. Left: false color composite (4,3,2) of Landsat 7, Right: false color composite (5,4,3) of Landsat 8 image for the same date and time.

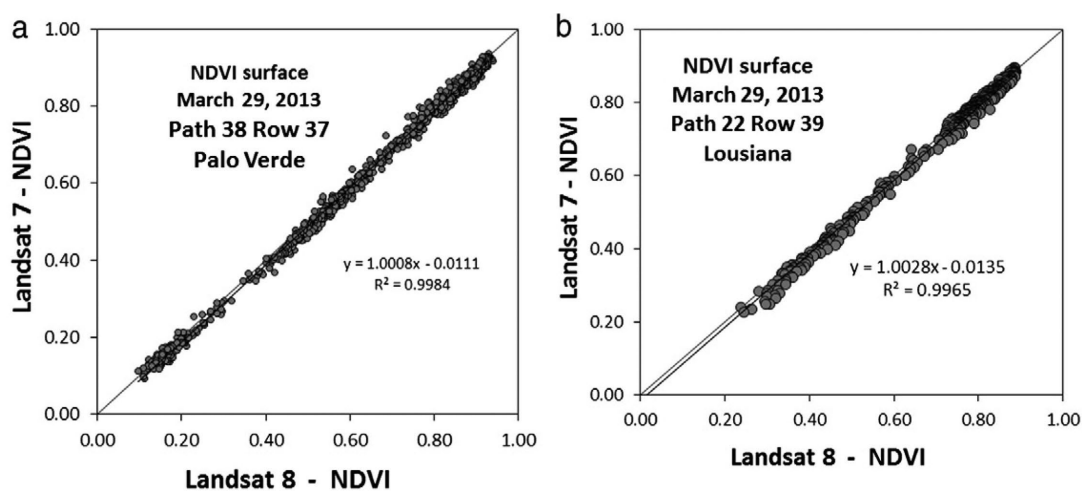


**Fig. 2.** Comparison between (a) surface temperature ( $T_s$ ) and (b) surface albedo for L8 and L7 images corresponding to March 29, 2013, path 38 row 37, as computed using the METRIC procedure.

radiation from a clear sky was estimated to be  $1.32 \text{ Wm}^{-2} \text{ sr}^{-1} \mu\text{m}^{-1}$ , and narrow band transmissivity was set to 0.866 following Allen et al. (2007a) for METRIC applications in the western US. Surface albedo was calculated using shortwave bands 2,3,4,5,6,7 for Landsat 8 and 1,2,3,4,5,7 for Landsat 7 using weighting coefficients by Tasumi et al. (2008) supplemented by new weighting coefficients for Landsat 8 by Trezza and Allen (2013).

Comparison of sampled LST from L7 to that from L8 shows close similarity, with about 1.5 K lower LST by L7 for temperatures around 300 K and nearly no differences, on average, for temperatures around 320 K (Fig. 2a). The RMSE was 1.06 K. Likewise, computed at-surface albedo was very similar between the two satellite platforms (Fig. 2b), with RMSE = 0.0027. The small differences in albedo were likely due to the different sensor response, differences in wavelength ranges, differences in radiometric and spatial resolution between the two platforms, differences in georegistration and differences in resampling and data handling. Differences in LST were additionally influenced by differences in spatial resolution of the thermal bands where L7 thermal data have 60 m native resolution and L8 have 100 m native resolution.

Normalized difference vegetation index (NDVI) is used in a number of satellite-based ET retrieval systems, including METRIC, to estimate aerodynamic roughness, soil heat flux, thermal emissivity and to identify water. Even though band widths and centers of the near infrared (NIR) band are substantially different between L7 and L8, NDVI calculated from surface reflectance was quite similar, when using the Tasumi et al. (2008) at-surface reflectance algorithms and Trezza and Allen (2013) Landsat-legacy coefficients for Landsat 8, as illustrated in **Fig. 3a** for the Palo Verde, California scene. The average ratio of NDVI from L8 to L7 was 1.017 and RMSE was 0.016 in NDVI units. The



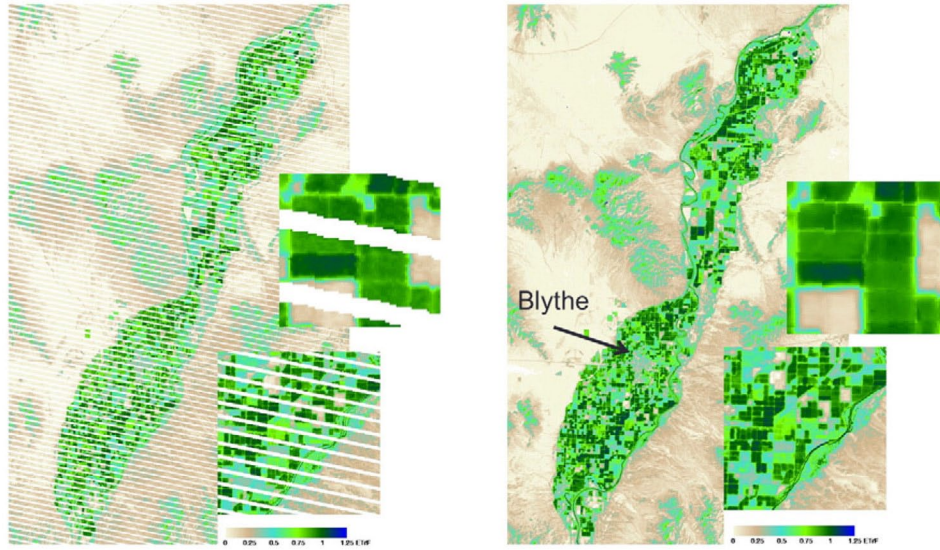
**Fig. 3.** Comparison between NDVI from L7 and NDVI from L8 corresponding to March 29, 2013 for **a)** path 38 row 37 and **b)** path 22 row 39 as computed using METRIC.

Landsat legacy coefficients of Trezza and Allen (2013) were designed to produce time-series of at-surface reflectance for Landsat 8 that produce NDVI values that are congruent with those produced from Landsat 5 and 7. That congruency in at-surface NDVI supports land-use change and disturbance analyses that span the long time spans of Landsat 5 through Landsat 8. The use of Landsat-legacy NDVI in METRIC does not impact values produced for ET, due to low sensitivity of algorithms for aerodynamic roughness, soil heat flux, etc. to NDVI, plus the use of CIMEC calibration in METRIC that offsets systematic biases in parameters. (See Fig. 4.)

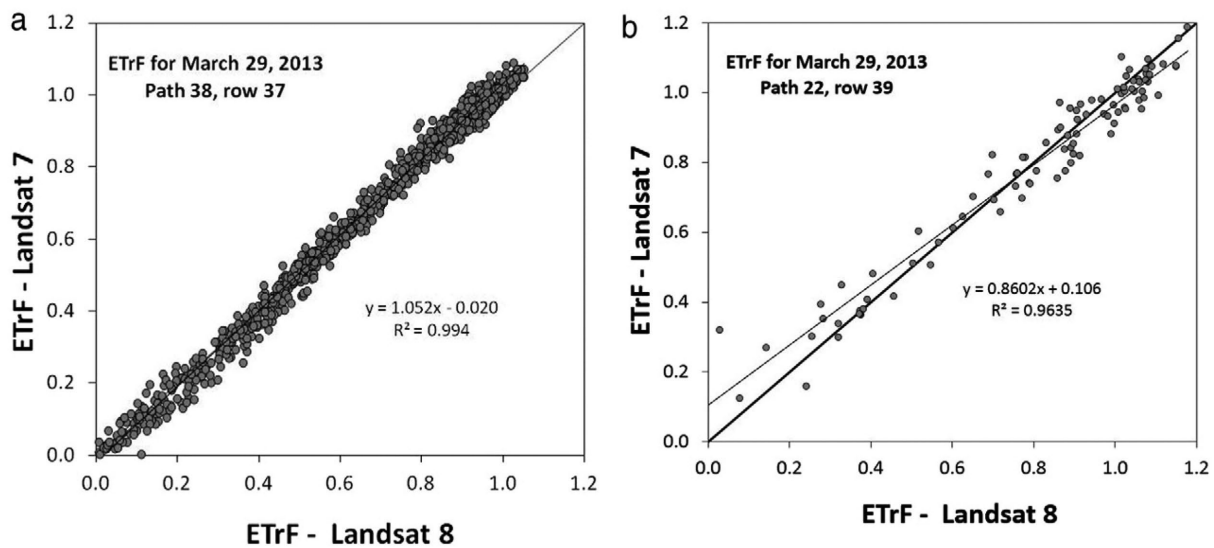
Differences in NDVI between L8 and L7 might be expected to be greater in more humid regions where the reduced sensitivity of the NIR band of L8 to water vapor absorption as compared to L7 would tend to produce higher values for NDVI than with L7. Atmospheric humidity is relatively low in the Southern California region. To test this, at-surface NDVI was computed for path 22 row 39 of southern Louisiana during the same March 29, 2013 underfly, where reported humidity was about triple that of Southern California. Results are plotted in Fig. 3b, where comparisons are similar to those for California. The ratio of NDVI of L8 vs. L7 did increase to an average of 1.018 for Louisiana, with RMSE of 0.020 ( $n = 540$ ). The ratio of Landsat-legacy NDVI of L8 to L7 was still close to 1.0 and would not significantly impact estimation of ET. Ratios of NDVI between L8 and L7 could be expected to be greater during summer when humidity levels and aerosol levels increase. The L8/L7 underfly only occurred during March 2013 when air humidity across the US was relatively low.

**Fig. 4** shows a colorized side-by-side comparison between  $ET_rF$  images obtained from L7 and from L8 for the Palo Verde application using the same METRIC model and calibration endpoints and calibration pixel locations. The spatial contrasts and magnitudes are very similar. The SLC-off gaps of L7 are apparent.

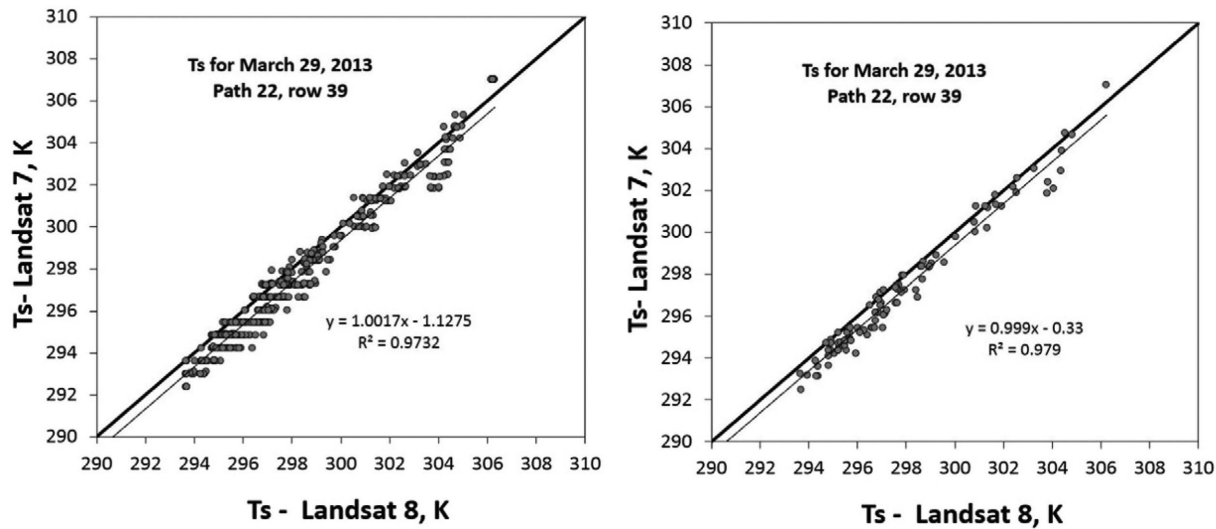
**Fig. 5** shows  $ET_rF$  for sampled areas in the Palo Verde region of path 38 row 37 and in the center region of path 22 row 39 in Louisiana for L7 vs. L8 images. The Louisiana data are means of five adjacent 30m pixels sampled in a cross-shaped pattern to reduce impacts of differences in  $T_s$  between L7 and L8 thermal images. The RMSE was 0.034 for Palo Verde and 0.21 for Louisiana in terms of  $ET_rF$  (unitless (0–1)). The variance of data about the 1:1 line of Fig. 5a mirrors that



**Fig. 4.** A progressive series of close-ups of ETrF maps created for the Palo Verde area of path 38 row 37 during the March 29, 2013 underfly with L7 (left) and L8 (right).



**Fig. 5.** Comparison between ETrF from L7 and L8 images corresponding to March 29, 2013 for **a)** the Palo Verde, CA area of path 38 row 37 and **b)** SE Louisiana for path 22 row 39 underfly scenes as computed using METRIC.



**Fig. 6.** Comparison between  $T_s$  from L7 and L8 images corresponding to the March 29, 2013 underfly for SE Louisiana, path 22 row 39 for **a)**  $n=515$  pixel locations and **b)**  $n=103$  locations where five adjacent pixels in a cross pattern were averaged.

of the surface temperature plot of Fig. 2, reflecting the strong influence of surface temperature on the ET retrievals. These results show high correspondence between the L7 and L8  $ET_F$  data sets as a result of the CIMEC calibration approach, with some differences due to differences in georegistration, resampling and thermal pixel size. Differences between L7 and L8  $ET_F$  retrievals were greater for the Louisiana example and were primarily due to larger scatter in  $T_s$  between the two satellites for the rainfed fields in Louisiana. Irrigated fields in Palo Verde, CA were quite uniform.

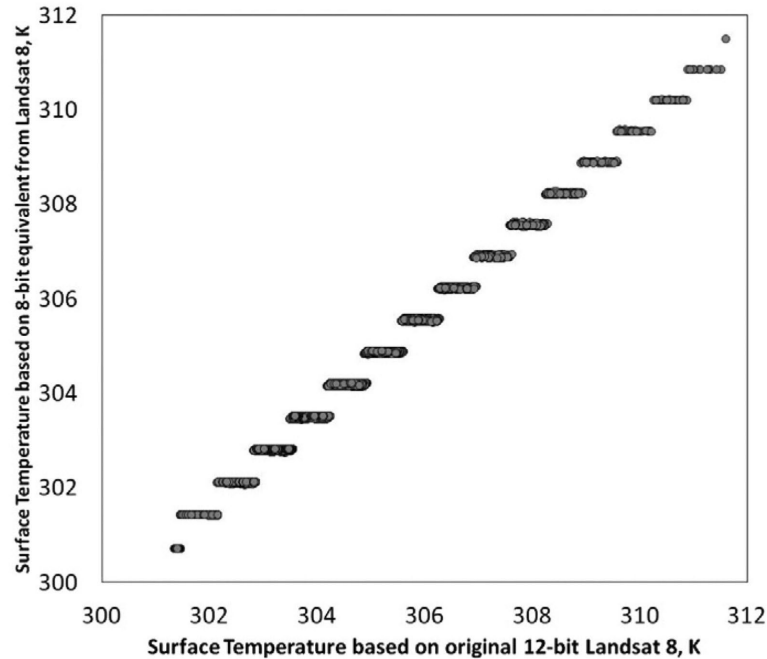
The larger disagreement between L7 and L8  $T_s$  and consequently  $ET_F$  for the Louisiana underfly is illustrated in **Fig. 6a** where  $T_s$  data from 515 30m pixels (resampled in EROS L1 T products from the native 60 and 100 m data) are plotted. The 512 pixels represent 103 locations where five adjacent pixels were sampled in a cross-shaped pattern. The 8-bit DN of L7 created noncontinuous steps of about 0.6 K in L7  $T_s$  as opposed to the more continuous  $T_s$  of L8 where quantization used 12-bits. Impacts of quantization resolution on  $ET_F$  are discussed in the following section. Fig. 6b shows means of the five-pixel groups that condensed some of the variation between  $T_s$  of L7 and L8. The RMSE for data in Fig. 6a and b was 0.84 K ( $n = 515$ ) and 0.81 K ( $n = 103$ ). The greater variation in the Louisiana  $T_s$  data set suggests

substantial variation in  $T_s$  at the 30 to 100m scale, possibly due to variation in soil properties or vegetation density that influenced water availability and therefore energy partitioning and  $T_s$ . Under these conditions, the impact of 60 m vs. 100 m pixel size of the two thermal sensing systems was stronger than for the uniform irrigated fields of Palo Verde, as well as the stronger impact of 8 vs. 12 bit quantization.

## 7. Impact of thermal radiometric resolution on estimation of ET

L8 thermal data are packaged as 16-bit digital number (DN) data in the L1 T products with native binary quantization resolution of 12-bit. L7 data are collected at 8-bit native binary quantization resolution and are handled and packaged as 8-bit. This section summarizes the evaluation of impacts of using 8-bit vs. 12-bit numerical resolution radiometric data on ET estimation. The impact of digital resolution is manifested primarily in terms of step changes in the ET and energy balance components per unit change in DN value. As is shown, native uncertainties in ET estimates are generally large relative to the unit change represented by one DN step for even 8-bit data so that, in applications with METRIC, the move to 12-bit is less essential. The evaluation of impacts of DN resolution was made by progressively degrading native 12-bit L8 thermal images by grouping sequential DN's reported in the original image to reduce apparent numerical resolution and recomputing ET. The grouping simulated moving from a 12 bit L8 to 11 bit (by grouping two adjacent DN's) to 10 bit (grouping 4 adjacent DN's), to 9 bit (grouping 8 adjacent DN's) and to 8 bit (grouping 16 adjacent DN's). The degradation calculations were done using the 'mod' function in ERDAS and ENVI. The degradation did not impact spatial fidelity of the original images, and was applied to TIRS data that had been resampled to 30 m by EROS using cubic convolution resampling.

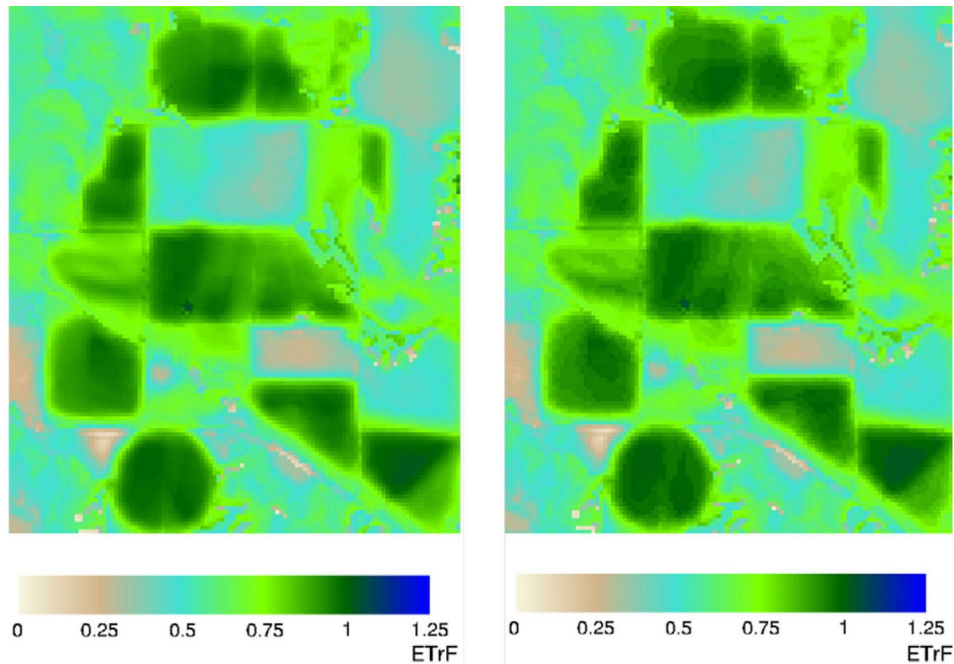
Non-degraded model runs were used as baselines for comparison with model runs made using degraded radiometric imagery for band 10. An example of resulting LST from the DN degrade is illustrated in **Fig. 7** where LST computed from DN degraded to 8 bit is plotted against LST computed from original 12-bit equivalent. LST in that figure was computed using a constant emissivity for illustrative



**Fig. 7.**  $T_s$  computed from DN degraded to 8 bit vs.  $T_s$  computed from original 12-bit equivalent using a constant emissivity for illustrative purposes.

purposes to prevent scattering in LST calculations caused by emissivity differences. The synthetic 8-bit data in the figure follow a characteristic stair-step pattern where the value for LST changes (jumps) only after the 12-bit data have progressively increased 16 times. The delta  $T$  per DN for the 8-bit was 0.7 K whereas the delta  $T$  per DN for the 12-bit was 0.04 K based reflecting a temperature range of about 160 K. The data in Fig. 7 are similar to actual L7 data plotted in Fig. 6a for southern Louisiana, where the delta LST per DN step was about 0.6 K, indicating the appropriateness and representativeness of the grouping approach. Data in Fig. 6a reflect impacts of variation in estimated thermal emissivity on retrieved LST.

$ET_F$  was computed at simulated radiometric resolutions of 11-bit, 10-bit, 9-bit, and 8-bit products. For each image date, METRIC model parameters were kept the same as the baseline METRIC run except for the impact of using degraded LST DN when assessing the hot and cold METRIC calibration endpoints. The calibration pixel locations for METRIC were kept the same as the baseline METRIC run for each date. A side-by-side comparison of  $ET_F$  using 12-bit vs.  $ET_F$  using 8-bit is



**Fig. 8.** Close up view of baseline  $ET_rF$  map (left) from 7/3/2013 for a section of agricultural fields near Lexington, NE produced using 12-bit original L8 data and, (Right) same view of  $ET_rF$  produced using an 8-bit equivalent thermal data set for the same date.

shown in **Fig. 8** for an approximately  $1.5 \text{ km} \times 2 \text{ km}$  area within Palo Verde Irrigation District. One can notice the terraced look in the fields for  $ET_rF$  derived using 8-bit thermal data due to the larger step size per DN.

About 1000 sample points were manually selected from agricultural areas and wetlands for Palo Verde and 500 sample points were manually selected from agricultural areas in Nebraska. Samples were selected approximately 1/3 in from field edges.

**Tables 1 & 2** show summary statistics for the March 29 underfly date in Palo Verde, CA and for three dates analyzed in the Central Platte area of Nebraska. The statistics use the baseline METRIC run as the basis of  $R^2$ , RMSE, MAE, and a and b of METRIC runs that used degraded DN.

Results were similar for both locations and for all dates evaluated, where the degradation of the radiometric resolution of the thermal band had small impact on the estimation of  $ET_rF$  using the METRIC

**Table 1**  $ET_rF$  statistics obtained from METRIC runs using degraded band 10 radiometric DN data for irrigated agriculture and wetland areas near Palo Verde (Blythe), CA March 29, 2013 underfly, path 38 row 37,  $n=1270$ . The  $a$  and  $b$  coefficients are for the relationship  $ET_{r,12\text{ bit}} = a ET_{r,x\text{ bit}} + b$  where  $x$  represents 8, 9, 10 or 11 bit quantization.

Case	Mean LST (K)	Mean $ET_rF$	RMSE	MAE	$R^2$	$a$	$b$
Original <sup>a</sup>	306.81	0.6899	0	0	1	1	0
8 bit	306.47	0.6966	0.0116	0.0093	0.9996	0.9866	0.01605
9 bit	306.80	0.6898	0.0003	0.0002	1.0000	1.0001	-0.00023
10 bit	306.81	0.6900	0.0002	0.0001	1.0000	1.0001	0.00005
11 bit	306.81	0.6899	0.0001	0.0001	1.0000	1.0001	-0.00009

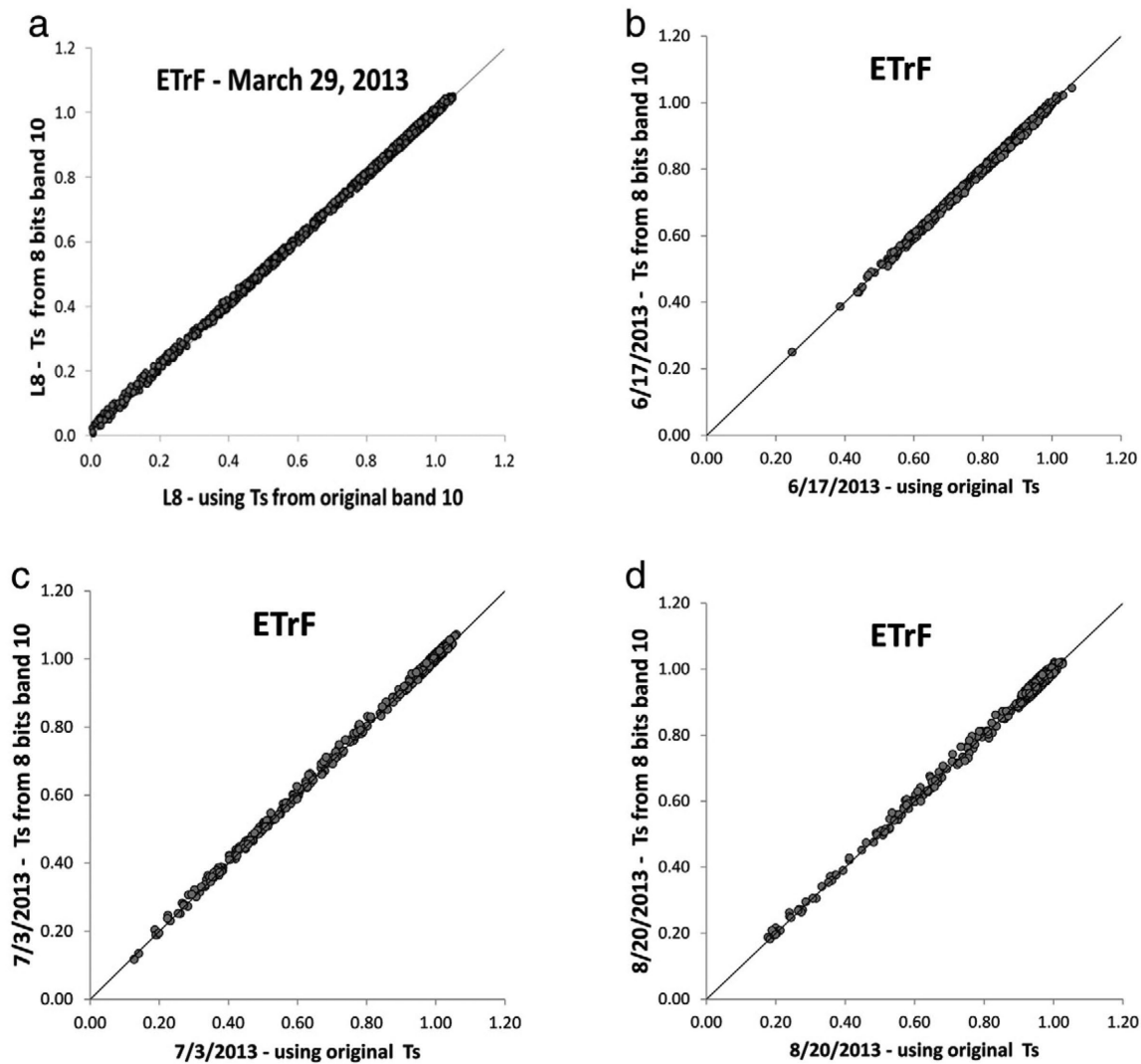
a. Calculated using original Landsat 8 band 10 with 12 bit radiometric resolution.

**Table 2**  $ET_rF$  statistics obtained from METRIC runs using band 10 radiometric DN data degraded to 8 bit resolution for irrigated and rain-fed agriculture and riparian areas near Grand Island, NE from path 30, row 32 for three dates,  $n = 500$ . The  $a$  and  $b$  coefficients are for the relationship  $ET_{r,12\text{ bit}} = a ET_{r,x\text{ bit}} + b$  where  $x$  represents 8, 9, 10 or 11 bit quantization.

Case	Mean LST of 12-bit original (K)	Mean $ET_rF$ of 12-bit original	Mean LST of 8-bit (K)	Mean $ET_rF$ of 8-bit	RMSE of $ET_rF$	MAE of $ET_rF$	$R^2$	$a$	$b$
June 17, 2013	306.05	0.7827	305.71	0.7811	0.00733	0.0062	0.9987	0.9914	0.0052
July 3, 2013	306.46	0.7985	306.12	0.8048	0.01029	0.0083	0.9995	1.0016	0.0050
August 20, 2013	303.81	0.8554	303.45	0.8578	0.01160	0.0098	0.9982	0.9956	0.0062

model. The degradation produced a maximum RMSE of only 0.012 of  $ET_rF$  using 8-bit equivalent degradation as shown by Tables 1 and 2, which translates into  $<1.5\%$  error in overall ET estimates. This amount of error is tolerable for essentially all ET applications and is well within the expected error of the METRIC process, where uncertainties in estimating components of the surface energy balance, for example, aerodynamic properties, soil heat flux, sensible heat flux and incoming thermal radiation, produce uncertainties in the ET estimation for individual pixels that are larger than the 1.5% error.

**Fig. 9** shows the high correspondence between the METRIC  $ET_rF$  calculated from the 12-bit band 10 of Landsat 8 vs. when band 10 was numerically degraded to 8 bits. Data were sampled on a 30 m basis. The variance between the  $ET_rF$  computed with degraded LST vs. that computed with original LST is consistently small across both locations



**Fig. 9.** Plot of ET<sub>r</sub>F values computed by METRIC using DN of LST (denoted as  $T_s$ ) degraded to 8-bit equivalent vs. ET<sub>r</sub>F computed using original 12-bit Landsat 8 LST for the Palo Verde March 29, 2013 underfly **(a)** and for three dates during summer of 2013 in Central, NE: **(b)** June 17, **(c)** July 3, and **(d)** August 20.

and dates and is well within error and uncertainty of the original ET<sub>r</sub>F estimates where typical expected error in ET<sub>r</sub>F retrievals by METRIC has been estimated to be 10% (Kilic (Irmak) et al., 2011a). The results indicate that the use of 8-bit, 9-bit, 10-bit, or 11-bit thermal data does not substantially degrade the accuracy of METRIC derived ET products, nor the model's performance.

## 8. Impacts of biased surface temperature on ET computation

This section investigates the impact that a systematic bias in land surface temperature (LST) retrieval may have on ET determined using a CIMEC-based ET process such as METRIC. This investigation provides some guidance on required accuracy of future LST retrievals and LST calibration, in an absolute sense, when producing estimates of ET. To test this, we introduced globally systematic biases into LST retrievals from both L7 and L8, recomputed  $ET_r$ , and compared to  $ET_r$  data from non-biased retrievals.

Global systematic LST biases were introduced into Landsat 8 and Landsat 7 LST data via the equations inside METRIC that apply the inverted Planck equation to estimate surface temperature from the narrowband radiance of band 10 for L8 and band 6 of L7. Systematic biases were created by adding 1 K and 3 K to all LST retrievals and by multiplying LST retrievals (in Kelvin) by 1.005, 1.01, 0.995, and 0.990.  $ET_r$  was recomputed using all LST bias parameters. METRIC model parameters were kept the same as the baseline METRIC run for each image date, except for the changing of LST for the hot and cold METRIC calibration pixels to reflect the biased values. The calibration pixel locations were kept the same.  $ET_r$  from biased model runs was compared with  $ET_r$  from non-biased model runs using sample point data. About 1000 sample points were manually selected from agricultural areas for the March 29, 2013 Palo Verde L8/L7 underfly images described earlier. For the Nebraska analyses, Landsat 8 imagery from path 30 row 32 for dates 6/17/2013, 7/2013, and 8/20/2013 and Landsat 7 imagery from path 30 row 32 for 6/9/2013 and 7/11/2013 were analyzed.

**Tables 3 & 4** summarize results for the L8 and L7 imagery from the March 29, 2013 underfly over Palo Verde, CA and **Tables 5 & 6** summarize results for June 17, 2013 L8 date and June 9, 2013 L7 date for path 30, row 32 in Nebraska. Values for  $R^2$ , RMSE, MAE, a and b, were computed using  $ET_r$  computed using original LST as a baseline. Results were similar for two additional sets of dates evaluated for Nebraska.

Impacts of the introduction of both additive and multiplicative biases into surface temperature on  $ET_r$  were very small for both locations, both satellite systems and over the range of dates evaluated.

**Table 3**  $ET_rF$  statistics obtained from METRIC runs using bias introductions to surface temperature of Landsat 8 for irrigated agriculture and wetland areas near Palo Verde (Blythe), CA March 29, 2013 underfly, path 38 row 37, n=1270. The a and b coefficients are for the relationship  $ET_{r\text{ original}} = a ET_{r\text{ with bias}} + b$ .

Case	Mean LST (K)	Mean $ET_rF$	RMSE	MAE	$R^2$	a	b
Original LST	306.81	0.690	0	0	1	1	0
Add 1 K	307.81	0.690	0.0002	0.0002	1.00000	1.0000	-0.0002
Add 3 K	309.81	0.689	0.0007	0.0005	1.00000	1.0001	-0.0005
Add - 3 K	303.81	0.690	0.0007	0.0005	1.00000	0.9999	0.0005
*1.01	309.88	0.689	0.0009	0.0007	1.00000	0.9999	-0.0005
*1.005	308.35	0.690	0.0004	0.0003	1.00000	1.0000	-0.0002
*0.990	303.75	0.690	0.0009	0.0006	1.00000	1.0001	0.0004
*0.995	305.28	0.690	0.0004	0.0003	1.00000	1.0000	0.0002

**Table 4**  $ET_rF$  statistics obtained from METRIC runs using bias introductions to surface temperature of Landsat 7 for irrigated agriculture and wetland areas near Palo Verde (Blythe), CA March 29, 2013 underfly, path 38 row 37, n=1270. The a and b coefficients are for the relationship  $ET_{r\text{ original}} = a ET_{r\text{ with bias}} + b$ .

Case	Average LST (K)	Average $ET_rF$	RMSE	MAE	$R^2$	a	b
Original	305.96	0.706	0	0	1	1	0
Add 1 K	306.96	0.706	0.0002	0.0001	1.00000	1.0001	-0.0001
Add 3 K	308.96	0.706	0.0006	0.0004	1.00000	1.0001	-0.0004
Add - 3 K	302.96	0.706	0.0006	0.0004	1.00000	0.9999	0.0003
*1.01	309.02	0.706	0.0008	0.0005	1.00000	0.9999	-0.0003
*1.005	307.49	0.706	0.0004	0.0003	1.00000	1.0000	-0.0002
*0.990	302.90	0.707	0.0007	0.0005	1.00000	1.0001	0.0003
*0.995	304.43	0.706	0.0004	0.0003	1.00000	1.0001	0.0001

The small effects are due to the internal calibration of METRIC using the CIMEC method and reassessment of assigned  $ET_rF$  at the two extreme calibration conditions (i.e., the cold and hot pixels). METRIC bases calculations of sensible heat flux, and thus ET computed as a residual of the surface energy balance, more on the differentials in surface temperature within an image relative to extreme conditions rather than on the absolute surface temperature. Therefore impacts of systematic bias in image parameters are substantially reduced.

Mean  $ET_rF$  produced by the biases was within 0.1% of the original values in all cases and MAE was <0.1% in all cases, indicating nearly complete compensation for a constant offset in surface temperature

**Table 5**  $ET_rF$  statistics obtained from METRIC runs using bias introductions to surface temperature of Landsat 8 for irrigated agriculture and riparian areas in Central Nebraska, June 17, 2013, path 30 row 32,  $n = 400$ . The  $a$  and  $b$  coefficients are for the relationship  $ET_{r, \text{original}} = a ET_{r, \text{with bias}} + b$ .

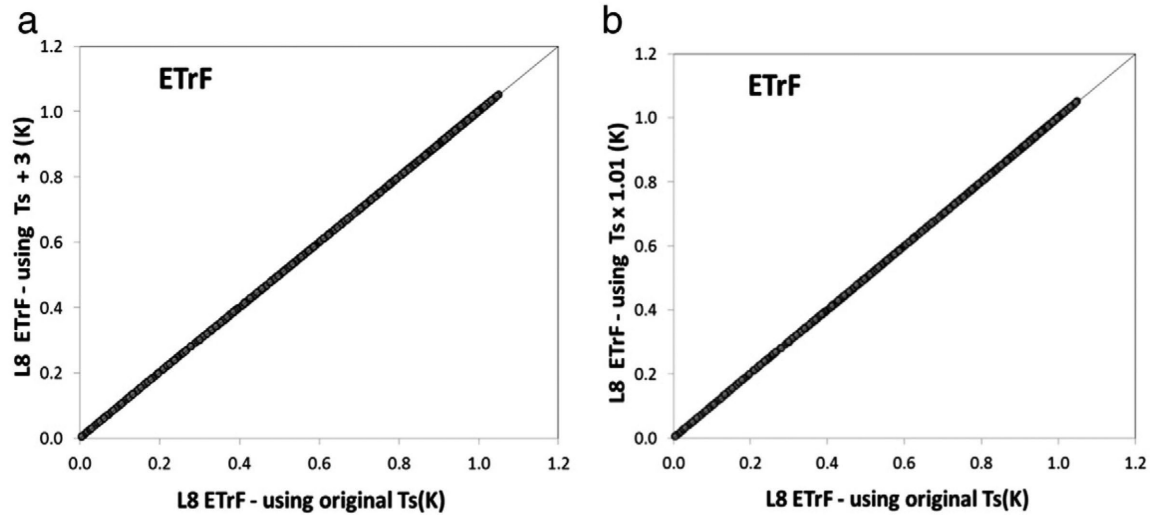
Case	Mean LST (K)	Mean $ET_rF$	RMSE	MAE	$R^2$	$a$	$b$
Original	306.26	0.781	0	0	1	1	0
Add 1 K	307.26	0.781	0.00023	0.00019	1.00000	1.0004	-0.0005
Add 3 K	309.26	0.781	0.00069	0.00059	1.00000	1.0013	-0.0016
*1.005	307.79	0.781	0.00049	0.00044	1.00000	1.0008	-0.0011
*1.01	309.32	0.781	0.00101	0.00091	1.00000	1.0017	-0.0023
*0.995	304.73	0.782	0.00049	0.00044	1.00000	0.9993	0.0010
*0.99	303.20	0.782	0.00096	0.00087	1.00000	0.9985	0.0020

**Table 6**  $ET_rF$  statistics obtained from METRIC runs using bias introductions to surface temperature of Landsat 7 for irrigated agriculture and riparian areas in Central Nebraska, June 9, 2013, path 30 row 32,  $n = 400$ . The  $a$  and  $b$  coefficients are for the relationship  $ET_{r, \text{original}} = a ET_{r, \text{with bias}} + b$ .

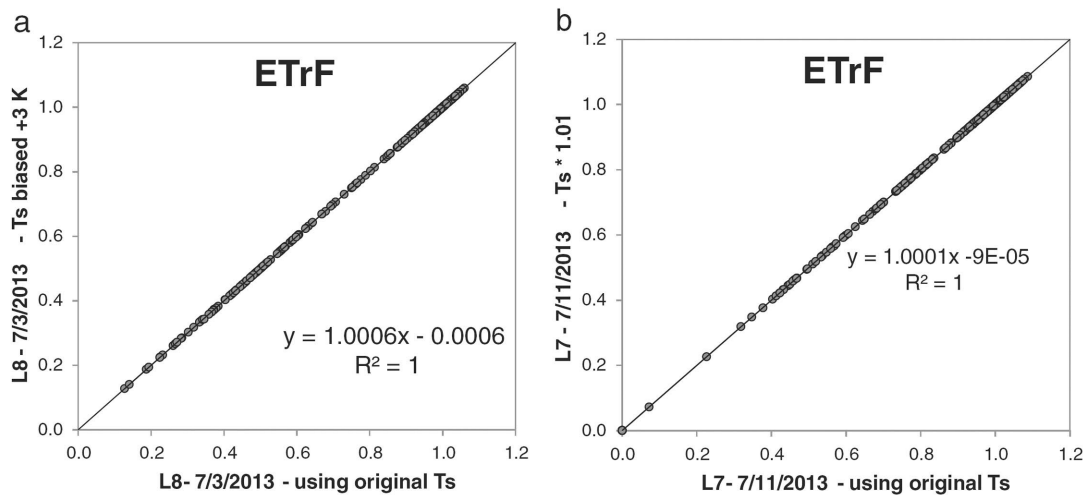
Case	Mean LST (K)	Mean $ET_rF$	RMSE	MAE	$R^2$	$a$	$b$
Original	288.89	0.324	0	0	1	1	0
Add 1 K	289.83	0.324	0.00016	0.00010	1.00000	0.9995	0.0001
Add 3 K	291.70	0.324	0.00048	0.00031	1.00000	0.9986	0.0002
*1.005	290.34	0.324	0.00033	0.00024	1.00000	0.9992	0.0000
*1.01	291.78	0.324	0.00067	0.00048	1.00000	0.9983	0.0001
*0.995	287.45	0.324	0.00032	0.00023	1.00000	1.0008	0.0000
*0.99	286.00	0.325	0.00063	0.00045	1.00000	1.0016	-0.0001

of up to 3 K or a constant multiplier (calibration coefficient) on surface temperature of 1%. The results are notable and expected, and suggest that future thermal sensors may not require highly accurate surface temperature retrieval for purposes of estimating evapotranspiration, provided that the sensors are consistent in their bias. It is the spatial consistency and consistency across the radiometric spectrum that is important to a CIMEC-type model such as METRIC.

**Figs. 10 & 11** show the high correspondence between the METRIC  $ET_rF$  results using original LST calculated from L8 band 10, and corresponding  $ET_rF$  results when using intentionally biased LST for the Palo Verde area (Fig. 10) and central Nebraska (Fig. 11). Results were nearly identical between with and without biases at both locations.



**Fig. 10.** ET<sub>r</sub>F produced by METRIC for sampled pixels computed following introduction of a + 3 K bias in surface temperature vs. ET<sub>r</sub>F from the original data set (a), and following introduction of a 1% bias in surface temperature (b) for Landsat 8 images during the March 29, 2013 underfly of path 38 row 37 near Palo Verde, CA.



**Fig. 11.** ET<sub>r</sub>F produced by METRIC for sampled pixels computed following introduction of a +3 K bias in surface temperature vs. ET<sub>r</sub>F from the original data set for a Landsat 8 image on July 3, 2013 of path 38 row 37 in central Nebraska (a), and following introduction of a 1% bias in surface temperature (b) for a Landsat 7 image on July 11, 2013 for the same path and row.

## 9. Impact of calibration accuracy of Landsat 8 on METRIC ET retrievals

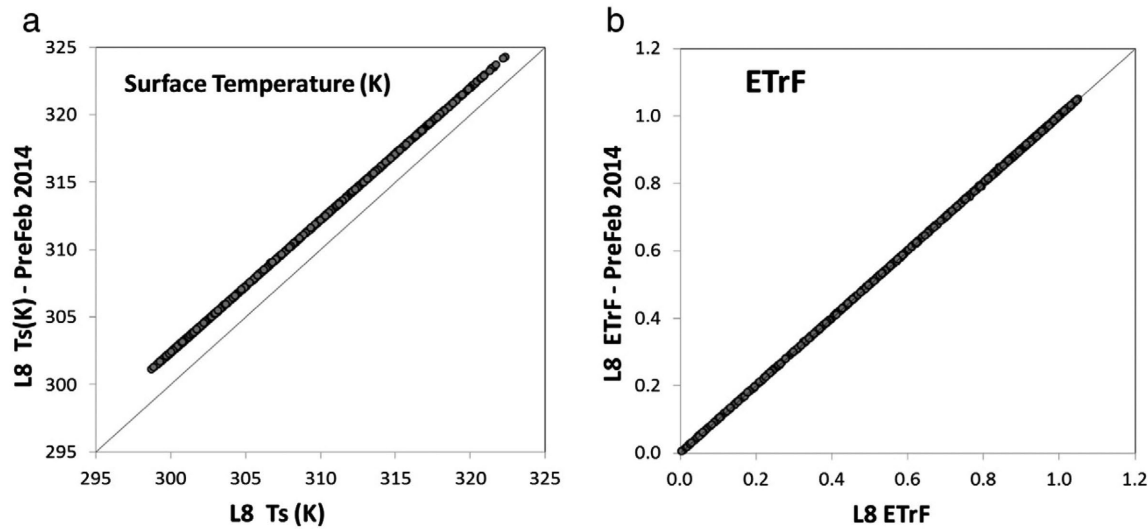
A final analysis compared ET calculated from imagery from the March 29, 2013 underfly in SE California that used the initial 2013 calibrations for the L8 shortwave and thermal bands with ET calculated following recalibration of the sensor in early 2014. The 2014 recalibration of L8 thermal imagery occurred on February 3, 2014, with the Landsat 8 archive reprocessed ([http://landsat.usgs.gov/calibration\\_notices.php](http://landsat.usgs.gov/calibration_notices.php)). According to the Landsat 8 website, the reprocessed data exhibit the following differences following recalibration:

- a) All OLI bands (excluding cirrus Band 9) have reflectance changes of up to 0.8% relative to products processed prior to February 3, 2014.
- b) The TIRS temperature offset removes about 2.1 K from Band 10 and about 4.4 K from Band 11, relative to products processed prior to February 3, 2014.

The purpose of the analysis was to explore the impact of applying METRIC to what would be considered to be 'very high quality' radiometric short wave and thermal data vs. applying METRIC to what might be considered to be 'standard quality' radiometric data. The latter data set refers to the use of satellite retrievals that utilize pre-launch calibrations over the life of the system with perhaps the occasional cross-calibration to other sources including ground measurements. The 'very high quality' term refers to the current Landsat system where a variety of means and efforts are expended to routinely test and update calibrations for all bands.

The Landsat 8 image corresponding to March 29, 2013 was processed with METRIC using the following two inputs:

- a) L8 image downloaded before February 3, 2014 that used pre-launch and near-post-launch calibrations. The image was generated on 06/25/2013.
- b) L8 image downloaded after February 3, 2014 that used recalibration coefficients determined post-launch. The image was generated on 02/14/2014.



**Fig. 12.** Comparison between (a) L8 LST (labeled  $T_s$ ) and (b)  $ET_rF$  using calibration coefficients available before February 2014 (L8\_PreFeb2014) vs. using revised thermal and shortwave calibration coefficients (L8) for Landsat 8 images during the March 29, 2013 underfly of path 38 row 37 near Palo Verde, CA.

**Fig. 12** shows a plot of L8 LST computed by METRIC using calibration coefficients available before February 2014 (L8\_PreFeb2014) vs. revised thermal calibration coefficients made available in February 2014. LST calculated from band 10 using old calibration coefficients was approximately 2.2–3 K warmer than the LST values calculated using revised calibration coefficients. The same figure includes  $ET_rF$  produced from METRIC using the old vs. revised calibration coefficients for both LST and for all reflectance bands. Even though LST was significantly impacted by the recalibration,  $ET_rF$  produced by METRIC was barely impacted due to the CIMEC-type of internal calibration of METRIC where fixed  $ET_rF$  values are assigned to extreme conditions (hot and cold pixels) of the image. RMSE in  $ET_rF$  was <0.1%.

## 10. Conclusions

We view the 12-bit quantization of thermal data in Landsat 8 as a major step forward, and one that is critical to many applications requiring a large signal to noise ratio. In regard to retrievals of evapotranspiration, however, our analyses indicate that the use of 8-bit thermal

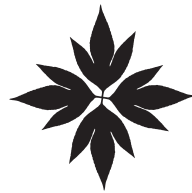
data does not substantially impact the accuracy of METRIC derived ET products, nor the model's performance as compared to using 12-bit data. RMSE of  $ET_F$  between 8-bit and 12-bit  $ET_F$  data was about 1%. Therefore, 8-bit thermal data, which is the radiometric resolution of previous Landsat platforms L7 and L5 would provide adequate performance and accuracy by a METRIC type of energy balance process, even though a better radiometric resolution would be preferable and probably critical to other investigations such as surface water quality monitoring and prediction.

Analysis of imagery collected during the March 29, 2013 L8/L7 underfly indicates that ET estimated by L8 imagery agrees closely with ET produced from L7 when using a CIMEC calibrated ET retrieval process such as METRIC. Surface temperature retrievals differed between L7 and L8 by as much as 2 K when using band 10 of L8. Some of these differences have been since mitigated by recalibration of band 10 of L8 and adjustment of image data for ghosting of the TIRS telescope (Schott et al., 2014, Montanaro, Gerace, Lunsford, & Reuter, 2014). NDVI from L8 produced using surface reflectance based on the Tsumi et al. (2008) algorithms and Landsat-legacy coefficients of Trezza and Allen (2013) agreed closely with NDVI from L7 during the underfly in both SE California, a dry climate, and in southern Louisiana, a humid climate.

Mean  $ET_F$  produced by introduction of biases into thermal images was within 0.1% of the original values in all cases and MAE was <0.1% in all cases, indicating nearly perfect compensation for a constant offset in surface temperature of up to 3 K or a constant multiplier (calibration coefficient) on surface temperature of 1%. The insensitivity of METRIC to a consistent bias in LST, is caused by the introduction of the bias into the internal calibration so that it basically cancels out when applied to pixels of an image. The results suggest that future thermal sensors may not require highly accurate surface temperature retrieval for purposes of estimating evapotranspiration, provided that the sensors are consistent in their bias. It is the spatial consistency and consistency across the radiometric spectrum that is important to a CIMEC-type model such as METRIC.

The recalibration of band 10 of Landsat 8 in February 2013 caused about a 3 K reduction in LST and changed reflectance values by about 0.7%. However, the use of recalibrated LST and shortwave data in

METRIC did not change the accuracy of ET retrievals due to the performance of the METRIC – CIMEC approach. These results support findings from the earlier section on bias introduction that systematic bias due to recalibration, provided it is linear, do not impact the evapotranspiration retrievals from process models such as METRIC that use an inversion process during internal calibration to compensate for those biases. These results suggest that, for purposes of producing accurate ET maps using CIMEC-driven process models such as METRIC, Landsat systems would not require frequent or rigorous recalibration of radiometric coefficients provided that biases are consistent across an image and are relatively linear across the radiometric spectrum.



**Acknowledgments** This research was supported by NASA Sustained Land Imaging Study Team (NNG14HX23P) (special acknowledgment and appreciation is given to Drs. Del Jenstrom and Jeff Masek for their support), the Landsat Science Team of the USGS and NASA (G12PC00068), by the Idaho Agricultural Experiment Station (CSREES 0226143), and by the University of Nebraska Agricultural Experiment Station (CSREES 0219182). The authors acknowledge the helpful reviews and suggestions by the journal reviewers and editors.

## References

- Allen, R.G., Tasumi, M., Trezza, R., 2007a. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) – Model. *ASCE Journal of Irrigation and Drainage Engineering* 133 (4), 380–394.
- Allen, R.G., Tasumi, M., Trezza, R., 2007b. Satellite-based energy balance for mapping evapotranspiration with internalized calibration METRIC—Applications. *ASCE Journal of Irrigation and Drainage Engineering* 133 (4), 395–406.
- Allen, R.G., Trezza, R., Kilic, A., Tasumi, M., Li, H., 2013. Sensitivity of Landsat-scale energy balance to aerodynamic algorithms in mountains and complex terrain. *Journal of the American Water Resources Association* 49 (3), 592–604.
- Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2011. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sensing of Environment* 122, 50–65.

- ASCE, 2005. In: Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R.L. (Eds.), *The ASCE Standardized Reference Evapotranspiration Equation*. Am. Soc. Civ. Engrs. - Environ. and Water Resources Instit. (Report 0-7844-0805-X, 69 pp. +App. A-F and Index).
- Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. *Journal of Hydrology* 212–213, 198–212.
- Bastiaanssen, W.G.M., Noordman, E.J.M., Pelgrum, H., Davids, G., Thoreson, B.P., Allen, R.G., 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *ASCE J. Irrig. and Drain. Engineering* 131 (1), 85–93.
- Burkhalter, J.P., Martin, T.C., Allen, R.G., Kjaersgaard, J., Wilson, E., Alvarado, R., Polly, J.S., 2013. Estimating crop water use via remote sensing techniques vs. conventional methods in the South Platte River basin, Colorado. *Journal of the American Water Resources Association* 49 (3), 498–517.
- Czapla-Myers, J.S., Anderson, N.J., Biggar, S.F., 2013. Early ground-based vicarious calibration results for Landsat 8 OLI. *Proceedings of SPIE* 8866.
- Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment* 122, 11–21.
- Kilic (Irmak), A., Allen, R.G., Kjaersgaard, J., Huntington, J., Kamble, B., Trezza, R., Ratcliffe, I., 2011a. Operational Remote Sensing of ET and Challenges Chapter in In: Irmak, A. (Ed.), *Evapotranspiration-Remote Sensing and Modeling*. InTech Publisher. (online).
- Kilic, A., Ratcliffe, I., Ranade, P., Hubbard, K.G., Singh, R.K., Kamble, B., Kjaersgaard, J., 2011b. Estimation of land surface evapotranspiration: A satellite remote sensing procedure. *Great Plains Research* 21 (1), 73–88.
- Markham, B.L., Helder, D.L., 2012. Forty-year calibrated record of earth-reflected radiance from Landsat: A review. *Remote Sensing of Environment* 122, 30–40.
- Markham, B.L., Goward, S., Arvidson, T., Barsi, J., Scaramuzza, P., 2006. Landsat-7 long-term acquisition plan radiometry — Evolution over time. *Photogrammetric Engineering and Remote Sensing* 72, 1129–1135.
- Markham, B.L., Irons, J.R., Storey, J.C., 2013. Landsat data continuity mission (LDCM) — Now Landsat-8: Six months on-orbit. *Proceedings of SPIE* 8866.
- Montanaro, M., Gerace, A., Lunsford, A., Reuter, D., 2014. Stray light artifacts in imagery from the Landsat 8 thermal infrared sensor. *Remote Sensing* 6 (11), 10435–10456.
- Morton, C.G., Huntington, J.L., Pohl, G.M., Allen, R.G., McGwire, K.C., Bassett, S.D., 2013. Assessing calibration uncertainty and automation for estimating evapotranspiration from agricultural areas using METRIC. *Journal of the American Water Resources Association* 49 (3), 549–562.
- Reuter, D.C., Richardson, C.M., Pellerano, F.A., Irons, J.R., Allen, R.G., Anderson, M., Jhabvala, M.D., Lunsford, A.W., Montanaro, M., Smith, R.L., Tesfaye, Z., 2015. The thermal infrared sensor (TIRS) on Landsat 8: Design overview and pre-launch characterization. *Remote Sensing* 7 (1), 1135–1153.

- Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., ... Kilic, A., ... Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment* 145, 154–172.
- Schott, J.R., Gerace, A., Raqueno, N., Ientilucci, E., Raqueno, R., Lunsford, A.W., 2014. Chasing the TIRS ghosts: Calibrating the Landsat 8 Thermal Bands. *Proc. of SPIE: Earth Observing Systems XIX* 9218.
- Schott, J.R., Hook, S.J., Barsi, J.A., Markham, B.L., Miller, J., Padula, F.P., Raqueno, N.G., 2012. Thermal infrared radiometric calibration of the entire Landsat 4, 5, and 7 archive (1982–2010). *Remote Sensing of Environment* 41–49.
- Tasumi, M., Allen, R.G., Trezza, R., 2008. Rapid operational estimation of at-surface reflectance and albedo from Landsat and MODIS satellites. *ASCE J. Hydrologic Engineering* 13 (2), 51–63.
- Tasumi, M., Trezza, R., Allen, R.G., Wright, J.L., 2005. Operational aspects of satellite-based energy balance models for irrigated crops in the semi-arid U.S. *J. Irrigation and Drainage Systems* 19, 355–376.
- Townshend, J.R.G., Justice, C.O., Gurney, C., McManus, J., 1992. The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing* 30, 1054–1060.
- Trezza, R., Allen, R.G., 2013. Report on developing albedo weighting coefficients and surface reflectance coefficients for the Landsat 8 operational land imager. Report by the University of Idaho Kimberly Research and Extension Center, Kimberly, Idaho (rev. 2016). 15 p. at: <http://extension.uidaho.edu/kimberly/2014/06/metric-mapping-evapotranspiration-at-high-resolution-and-internalized-calibration/>.
- Trezza, R., Allen, R.G., Robison, C.W., Kjaersgaard, J., 2011. Completion report on the production of satellite-based maps of evapotranspiration using the METRIC™ model for year 2010 for Landsat path/row45/30 in Oregon. Univ. Idaho Completion Report submitted to United States Forest Service, December 2011 (89 p).