Land Surface Air Temperature Retrieval from EOS-MODIS Images

Raquel Niclòs, José A. Valiente, Maria J. Barberà and Vicente Caselles

Abstract—The knowledge of the spatial and temporal patterns of Surface Air Temperature (SAT) is essential to monitor a region’s climate and meteorology, quantify surface exchange processes, improve climatic and meteorological model results, and study health and economic impacts. This work analyzed correlations between SAT and geophysical land surface variables, Land Surface Temperature (LST) mainly, to establish operative techniques to obtain spatially-continuous land SAT maps from satellite data, unlike data provided by meteorological station networks. The correlations were analyzed by using EOS-MODIS images, meteorological station network data, and geographical variables. Linear regressions with MODIS-retrieved LST data gave SAT with uncertainties higher than ±2K during daytime and of ±1.8K at night-time. Nevertheless, SAT uncertainties decreased up to ±1.2K when other satellite-retrieved surface parameters, i.e. vegetation index and albedo, together with meteorological and geographical data were considered as terms of multivariable regressions. The equations finally proposed were shown to work properly for different land covers.

Index Terms—EOS-MODIS, Geographical Data, Land Surface Temperature, Meteorological Data, Surface Air Temperature.

I. INTRODUCTION

Spatial and temporal patterns of Surface Air Temperature (SAT), defined at ~2 m above ground, are primary descriptors of terrestrial environmental conditions and describe a region’s climate and meteorology. SAT is a key variable to quantify surface exchange processes and to assess energy fluxes [1] and water stress indexes. The monitoring of SAT patterns can also improve meteorological model results by using them as input data. Additionally, extreme SATs have health consequences, increase electricity loads, and result in reduced crop yields [2]. SAT is usually measured by meteorological stations and thus SAT data are limited by the density and distribution of available station networks [3]. The cited applications require higher spatial resolutions than those given by station networks. The aim of this study is to establish relationships between ground-measured SAT, remote sensing variables, and geographical parameters to generate near-spatially-continuous SAT maps from satellite data, solving the discontinuity problem of ground measurements [3].

Studies to estimate SAT from remote sensing data are mostly based on statistical approaches. Goward et al. [4] proposed the Temperature-Vegetation Index (TVX) approach to derive SAT maps from least squares regressions of the Normalized Difference Vegetation Index (NDVI) on Land Surface Temperature (LST). It assumes that the LST of a fully vegetated canopy approximates SAT [8] and uses a mobile convolution matrix to obtain SAT that yields coarser SAT spatial resolutions than the satellite images used. The approach was tested in different parts of the world and yielded root-mean-square-error (RMSE) values from ±3K to ±4K [4]-[6]. Other methodologies establish linear regressions between daily or monthly mean, maximum or minimum SAT and LST at the spatial resolution of LST images. Vogt et al. [3] used NOAA-AVHRR LSTs for linear regressing daily maximum SAT on LST in Southern Spain (with SAT accuracies of ±2.5K). Green and Hay [7] used NOAA-AVHRR data for providing climatic monthly variables across Africa and Europe for epidemiological applications (with RMSE of ±2.4K).

The paper goal is to establish operative regressions to estimate SAT and generate instantaneous SAT maps from satellite images during the diurnal cycle. Statistical SAT data (daily or monthly means, maximums or minimums) could be obtained by using several satellite overpasses. EOS Moderate Resolution Imaging Spectroradiometer (MODIS) images were used to analyze SAT correlations with remote-sensing surface variables, together with geographical data and meteorological station measurements in Eastern Spain. This study region (Fig. 1) was selected due to the availability of high-quality data provided by a dense and well distributed station network, but also because of the high spatial variability of meteorological conditions. This variability is due to the region’s complex orography (with elevations from sea level up to 1560m) and the spatial heterogeneity of
land covers (Fig.1) [8], which involves a wide range of surface VIS-TIR response properties.

Fig. 1. 500-m land cover map (28 July - 12 August 2009) for the study region [12], together with the meteorological station locations.

II. DATA

A. Remote Sensing Data

The MODIS daily and 1-km MYD11A1 product was used to obtain LST [9]. It gives two LST retrievals per day: at 12-14 UTC and 1-3 UTC over the studied area, which are relatively close to the times of maximum and minimum SAT values in the diurnal cycle. The 16-day and 1-km MYD13A2 product was used to obtain the NDVI, with an accuracy of ±0.025 [10]. Additionally, 16-day and 500-m albedo (AL) values were obtained from the MCD43A3 product [11]. MODIS products have cloud contamination data and cloudy image pixels were excluded. Only pixel data identified as produced with good quality in the product quality flags were considered for the study. Finally, we used a regional land cover map (Fig. 1) generated by means of a maximum likelihood classification method based on spectral isotropic BRDF parameter images (bands 1-7) of the MODIS 16-day MCD43A1 product [12].

B. Ground Meteorological Data

Concurrent ground-measured data were collected by the CEAM-UMH’s meteorological station network (see locations in Fig. 1; www.ceam.es), with 36 stations in the Valencia region, Eastern Spain. Station elevations range from 10m to 1270m above sea level and the network covers different land covers (Fig. 1). The network systematically records 10-min data on humidity (Hr), SAT at 1.5-2m height, wind speed (U), rainfall, and solar irradiance (I), the latter at only 13 stations. Hr and SAT are measured by Vaisala HMP45A/50Y probes, which have PT-100 temperature probes with accuracies of ±0.2K. I is recorded by Kipp & Zonen CM3 pyranometers, with a ±5% accuracy, and U is measured by Gill Instruments WindSonic and R.M. Young Wind Sentry 03002 anemometers with accuracies of ±0.2m/s and ±0.5m/s. Apogee SI-111/121 8-14µm radiometers (with temperature accuracies of ±0.2K at 278K-303K [12]) were set up on two sites. Site A is a high-plain area of Mediterranean calcic thermophile shrubland (39.224°N, -0.903°E in WGS-84; site elevation of 800m). Site B is a flat homogeneous area of rice-crop fields (39.274°N, -0.317°E in WGS-84; at sea level), with full vegetation cover in summer. Ground LST data were obtained from the radiometer measurements as described in [12] and were used to analyze relationships between LST-SAT and I in [13]. An increase in LST-SAT was observed with I on both sites, but the degree of dependence was different for each one; only a slight dependence was observed at site B due to the low water stress and the full vegetation cover (i.e., the latent heat flux is the dominant term in the Energy Balance Equation, which keeps LST near SAT). The different behavior for each site suggested LST-SAT dependences on other surface properties that are analyzed here. Additionally, the ground-measured LSTs were used to validate MODIS-retrieved LSTs. An increase of MODIS LST uncertainties on MODIS-viewing angle was observed, with RMSEs > ±2K and ±3K at sites B and A respectively for viewing angles > 40°. RMSEs of ±0.6 and ±1.5 were obtained at B and
A for angles < 40°. Thus, only MODIS LSTs measured with angles < 40° were finally used.

C. Geographical Data

Ninyerola et al. [14] and Cristóbal et al. [15] pointed out SAT dependences on geographical variables, e.g., elevation (h), latitude (Lat), continentality and terrain curvature. Zaksek and Schroedter-Homscheidt [2] proposed a model to obtain SAT that considered solar incidence parameters and the terrain curvature, with a term defined as the difference between station elevation and mean elevation within 20 km vicinity (Δh). We computed topographic aspects and slopes (φ, s, a, and i) were obtained during the summer months of 2009 due to the prevalence of cloud-free conditions, which are essential to obtain LST, NDVI and AL from VIS-TIR satellite data. This condition also leads to a wide range of SAT-LST differences. After removing MODIS LSTs retrieved at angles > 40°, 845 sets of concurrent data were examined to analyze SAT dependences at night-time, and 285 sets at daytime due to the availability of I data. Table I shows statistics (mean, standard deviation (SD), minimum (MIN) and maximum (MAX) values) for the data included in the final databases. Statistics for the remote sensing variables were also estimated for the whole study region (see in Fig.1) and the similarity observed between the values for the whole area and those shown in Table I confirms the representativity of the data, at the station locations, included in the databases. SAT dependences were observed on: remote sensing LST (with the strongest dependence), AL and NDVI; ground-measured I, Hr and U; and geographical Lat, h, dist, Δh, φ, s, z, a, and i) were obtained during the summer months of 2009 due to the prevalence of cloud-free conditions, which are essential to obtain LST, NDVI and AL from VIS-TIR satellite data. This condition also leads to a wide range of SAT-LST differences. After removing MODIS LSTs retrieved at angles > 40°, 845 sets of concurrent data were examined to analyze SAT dependences at night-time, and 285 sets at daytime due to the availability of I data. Table I shows statistics (mean, standard deviation (SD), minimum (MIN) and maximum (MAX) values) for the data included in the final databases. Statistics for the remote sensing variables were also estimated for the whole study region (see in Fig.1) and the similarity observed between the values for the whole area and those shown in Table I confirms the representativity of the data, at the station locations, included in the databases.

Finally, a database of concurrent satellite-retrieved data (LST, NDVI, and AL), ground-measured meteorological variables (SAT, Hr, I and U) at the EOS-MODIS overpass times over the station locations, and geographical data (Lat, h, dist, Δh, φ, s, z, a, and i) were obtained during the summer months of 2009 due to the prevalence of cloud-free conditions, which are essential to obtain LST, NDVI and AL from VIS-TIR satellite data. This condition also leads to a wide range of SAT-LST differences. After removing MODIS LSTs retrieved at angles > 40°, 845 sets of concurrent data were examined to analyze SAT dependences at night-time, and 285 sets at daytime due to the availability of I data. Table I shows statistics (mean, standard deviation (SD), minimum (MIN) and maximum (MAX) values) for the data included in the final databases. Statistics for the remote sensing variables were also estimated for the whole study region (see in Fig.1) and the similarity observed between the values for the whole area and those shown in Table I confirms the representativity of the data, at the station locations, included in the databases.

III. METHODOLOGY

SAT dependences on the cited variables were analyzed both for daytime and night-time MODIS overpasses by using these databases. SAT dependences were observed on: remote sensing LST (with the strongest dependence), AL and NDVI; ground-measured I, Hr and U; and geographical Lat, h, dist, Δh, cosz, and a. The relationship between SAT and LST is fundamentally driven by the net radiation partitioning to latent or sensible heat fluxes through the Energy Balance Equation [1], [16]. I and AL are involved in the shortwave balance of the net radiation. As in [4], NDVI is used together with LST to estimate SAT because generally most of the net radiation is converted to latent heat over densely vegetated areas, which keeps LST near SAT. A higher level of latent heat exchange is found with more vegetated areas, while sensible heat exchange is more favored by sparsely vegetated areas [17]. The sensible heat flux is directly proportional to the LST-SAT difference and inversely proportional to aerodynamic resistances to heat transfer, which are in turn inversely dependent on U [1]. LST-SAT differences are usually smoothed for large Hr and thus Hr is a key term. The LST-SAT relationship is dependent on soil moisture (SM) for sites with low NDVIs, since wet soils exchange latent heat mainly, leaving LST similar to SAT, and dry soils give large LST-SAT differences. The relationship between LST and NDVI results in a triangular shape and the slope of the curve has been related to SM [18] and surface evapotranspiration. Unfortunately, neither SM remotely-sensed products nor ground data were available at the spatial resolution of the study. LST and NDVI variables, which some authors use for SM estimation [18], were already included in the SAT approach. Finally, Ninyerola et al. [14] considered geographical variables as important climate factors to include.

The available daytime and night-time databases were split into independent training and test datasets (with 80% and 20% of the data, respectively, to give more reliability to the regression procedure) with uniform sampling without replacement (MATLAB routine randsamp). Table II shows the regressions obtained between SAT and the studied variables for daytime and night-time with the training datasets. $R^2$ and $\sigma$ (SAT) are the coefficient of determination and the error of estimate. Differences between SATs calculated with the models for the test datasets and ground SATs were computed as a validation. Table II shows biases and standard deviations, $\sigma$, of these differences, together with RMSEs. Additionally, a bootstrapping technique (with 1000 iterations or bootstrap data samples) was used to evaluate the robustness of the models. The bootstrapping results showed similar or minor biases than the test results, and close $\sigma$ values, yielding the RMSEs, together with the confidence intervals, shown in the last column of Table II. First, linear regressions between SAT and LST were fitted to the training datasets. Then, a standard stepwise procedure [19], with entrance and exit tolerances of 0.01 and 0.10 respectively, was applied on the training datasets to select the statistically significant terms in multilinear regressions of all the cited variables, not just LST. Table II shows the regressions obtained by means of the stepwise procedure (D.2 and N.2) with the non-standardized regression coefficients. The standardized coefficients can be estimated from them by using the SDs included in Table I.
The D.2 and N.2 equations show the multilinear regressions of SAT against the 41% of the SAT variability, i.e., $R^2$ was 0.41 when only LST was used to estimate SAT. The second variable entered in D.2 is Hr, which increases $R^2$ in 0.09, i.e., the importance of Hr in the SAT prediction was 9%. A 20% increase of $h$, and test datasets. The agreement between the results obtained with both datasets separately proves the consistency of the models.

N.3-calculated against the ground SATs, again with a good agreement between the results for the training and test datasets. Nevertheless, the U dependence can be removed without increasing the error significantly (see N.3 results in Table II). Fig. 2b shows N.3 expression without Hr terms. Hr terms in equation. D.3 in Table II replaces the I and AL terms in D.2 with a common (1-AL)I term related to the surface absorption of I.

Table II shows the data statistics for the daytime and night-time databases.

### TABLE II

<table>
<thead>
<tr>
<th>Night-time Expressions</th>
<th>Daytime Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td><strong>Test</strong></td>
</tr>
<tr>
<td>R²</td>
<td>σSAT</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>(D.1) SAT=0.40LST+178.7</td>
<td>0.41</td>
</tr>
<tr>
<td>(D.2) SAT=0.26LST-12.4(1-NDVI)-5.9Hr-0.031Lat+0.88-0.081dist+27.0AL+260.1</td>
<td>0.80</td>
</tr>
<tr>
<td>(D.3) SAT=0.25LST-9.3(1-NDVI)-8.8(1-AL)I-0.029Lat+0.09Hr-0.029dist+264.6</td>
<td>0.79</td>
</tr>
<tr>
<td>(D.4) SAT=0.47LST-18.6(1-NDVI)-6.0I-0.040Lat+1.9Hr-0.031dist+44.4AL+195.4</td>
<td>0.67</td>
</tr>
<tr>
<td>(D.5) SAT=0.49LST-13.5(1-NDVI)-10.5(1-AL)I-0.037Lat-0.5h-0.056dist+194.6</td>
<td>0.63</td>
</tr>
<tr>
<td>(D.6) SAT=0.26LST-12.4(1-NDVI)-5.9Hr-0.031Lat-0.081dist=27.0AL+260.1</td>
<td>0.80</td>
</tr>
<tr>
<td>(D.7) SAT=0.52LST-152.7+8.6cosz(1-NDVI)-14a-4.1(1-AL)Icosz(1+sn/s)-0.2(1-AL)I</td>
<td>0.46</td>
</tr>
<tr>
<td>(D.8) SAT=0.52LST-152.8+8.5cosz(1-NDVI)-14a-4.1(1-AL)Icosz(1+sn/s)-0.2(1-AL)I</td>
<td>0.44</td>
</tr>
<tr>
<td>(N.1) SAT=0.51LST-9.5cosz(1-NDVI)-2.3I-1-ALIcosz(1+sn/s)-0.151Hr+0.033Lat-2.8h-0.059dist+270.4</td>
<td>0.76</td>
</tr>
<tr>
<td>(N.2) SAT=0.51LST-15.1cosz(1-NDVI)-3.1I-1-ALIcosz(1+sn/s)-0.044Lat+1.08h-0.059dist+188.5</td>
<td>0.61</td>
</tr>
</tbody>
</table>

**σ(SAT)**, bias, and RMSE are in K. Units: SAT and LST in K; I in 10⁻³ W/m²; Hr in %; Lat in km (UTM); h, dist and Δh in km; a and s in radians; U in m/s; NDVI, AL, cosz and cosi are dimensionless. (D.6) is from Zakssek and Schroedter-Homscheidt [2] and σ(SAT) is here RMSE over the training database.

### IV. RESULTS

σ(SAT) and RMSE values > ±2K and of ±1.8K were obtained when using SAT linear regressions just on LST at daytime and night-time, respectively (D.1 and N.1 in Table II). The D.2 and N.2 equations show the multilinear regressions of SAT against the statistically significant terms selected by the stepwise procedure. The stepwise procedure for D.2 showed that LST explained the 80% of the SAT variability, i.e., $R^2$ was 0.80 when only LST was used to estimate SAT. The second variable entered in D.2 was Hr, which increased $R^2$ in 0.09, i.e., the importance of Hr in the SAT prediction was 9%. A 20% was obtained for the h, and the rest of terms were less important, making the final D.2 model to attain a $R^2$ of 0.80. 71% and 14% were found for LST and Hr respectively in N.2, in contrast with the rest of variables. The LST importance in the SAT determination is explained by its implication in the net radiation partitioning into the different heat fluxes. Sensible heat flux is directly related to the LST - SAT ($R^2$ of 0.88 and 0.84 were obtained for them, respectively). Similar expressions without Hr terms were tested. Hr in daytime D.2 and N.2 (night-time N.2 and N.3) were not considered in D.4 and D.5 (N.4 and N.5) equations. However, minimum errors of ±1.7K were obtained for them. Additionally, the daytime model proposed in [2] was checked (D.6 in Table II) and gave RMSE > ±3K. Fig. 2a shows SATs calculated with D.6 and D.2 equations against the ground SATs. Much higher discrepancies are shown for D.6 [2] than for D.2. Better results were obtained after fitting the D.6 coefficients on our training dataset, but SAT
uncertainties $> \pm 2K$ were still shown (D.7). A stepwise procedure was also applied to evaluate the statistical significance of the D.7 terms. Two terms could be removed, leaving just the terms with LST, 1-NDVI and (1-AL)(L) (D.8). Other significant terms considered in D.2 were added to D.8 to improve the results. D.9 shows the resulting equation, which includes an Hr term, and D.10 is the same expression without it. Errors of $\pm 1.6K$ and $\pm 2K$ were obtained for D.9 and D.10, which did not achieve the accuracies given by D.2 and D.4.

Finally, D.2 and N.3 equations gave the lowest SAT uncertainties $(\pm 1.3K)$. D.2 and N.3 results were also analyzed against land cover. A land cover map (Fig. 1) was used to identify the land covers for the stations. The land covers were reclassified in 4 generic types (shrublands, forests, croplands and urban areas) for simplicity and for having a representative number of ground data to analyze possible model dependences on land cover. Fig. 3 shows the D.2 and N.3 results when they were applied on the training and test datasets split into the land cover types. There were no significant differences for the 4 covers, with similar RMSEs for all of them (see Fig. 3). The soundness of the D.2 and N.3 equations for the different land cover types is a consequence of the inclusion of terms related to surface properties, like AL and NDVI, in the expressions.

V. CONCLUSIONS

Operational techniques have been proposed to monitor SAT patterns. Linear regressions of SAT against MODIS LST provide SAT uncertainties $> \pm 1.8K$. Nevertheless, SAT uncertainties decrease up to $\pm 1.2K$ when other remote sensing inputs (NDVI and AL) and geographical and meteorological data (Lat, h, dist, I, and Hr) are considered (D.2 and N.3). The D.2 and N.3 equations are also shown to work properly for different land covers and to achieve better accuracies than the techniques previously published, described in Section I and [2]. This paper proves the soundness of combining remote sensing and geographical variables to obtain instantaneous and near-spatially-continuous SAT data (with a spatial resolution defined by the coarsest resolution of the input variable data, i.e., $1km \times 1km$), solving the discontinuity problem of ground data. The use of MODIS images restricts the SAT map generation to the two daily overpasses per polar EOS platform. The use of images given by geostationary satellites, like those of the MSG-SEVIRI, would be more appropriate to follow the SAT diurnal cycle, but with a
coarser spatial resolution (4km × 4.5km for the study region). The better knowledge of SAT patterns will allow new advances in a broad spread of scientific fields, such as air/land interactions, climate change, remote sensing, and health impacts.

REFERENCES


