Synergistic use of MERIS and AATSR as a proxy for estimating Land Surface Temperature from Sentinel-3 data

J.A. Sobrino [a,*], J.C. Jiménez-Muñoz [a], G. Sòria [a], A.B. Ruescas [b], O. Danne [b], C. Brockmann [b], D. Ghent [c], J. Remedios [c], P. North [d], C. Merchant [e], M. Berger [f], P.P. Mathieu [f], F.-M. Göttsche [g]

[a] *GCU/IPL, University of Valencia, Spain*
[b] *Brockmann Consult, GmbH, Germany*
[c] *University of Leicester, UK*
[d] *Swansea University, UK*
[e] *University of Reading, UK*
[f] *ESA/ESRIN, Italy*
[g] *Karlsruhe Institute of Technology, Germany*

ARTICLE INFO

ABSTRACT

Land Surface Temperature (LST) is one of the key parameters in the physics of land-surface processes on regional and global scales, combining the results of all surface-atmosphere interactions and energy fluxes between the surface and the atmosphere. With the advent of the ESA's Sentinel 3 (S3) satellite, accurate LST retrieval methodologies exploiting the synergy between OLCI and SLSTR instruments can be developed. In this paper we propose a candidate methodology for retrieving LST from data acquired with the forthcoming S3 instruments. The LST algorithm is based on the Split-Window (SW) technique with an explicit dependence on surface emissivity, in contrast to the AATSR level 2 algorithm with emissivity dependence embedded in the algorithm coefficients. Performance of the methodology is assessed by using MERIS/AATSR pairs (instruments with similar characteristics to OLCI and SLSTR, respectively). LST retrievals using different datasets of input emissivity are validated against in situ data measured along one year (2011) in five test sites and intercompared to the standard AATSR level 2 products. Validation results show that LST is retrieved with the proposed SW algorithm typically with RMSE below 2 K, providing slightly better results than the AATSR level 2 product. The main advantage of the proposed algorithm is that it allows for improvements in input emissivities to be directly translated into improved LST retrievals.

1. Introduction

The Sentinel satellite constellation series is developed by the European Space Agency (ESA) in order to support European operational services and the policy needs of the Copernicus programme. The first three Sentinel missions are planned to contribute to the understanding of the Earth System by detecting, monitoring and assessing changes in ocean, cryosphere, and land components (Malenovský et al., 2012; Berger, Moreno, Johannessen, Levelt, & Hanssen, 2012). In particular, Sentinel-2 (S2) and Sentinel-3 (S3) missions operate optical radiometers to provide multispectral data at high, medium and low resolution.

The S2 mission was designed i) to provide systematic global acquisitions of high-resolution multispectral imagery with a high revisit frequency, ii) to provide enhanced continuity of multi-spectral imagery provided by the Satellite Pour l'Obervation de la Terre (SPOT) series of satellites, and iii) to provide observations for the next generation of operational products such as land-cover maps, land change detection maps, and geophysical variables (Drusch et al., 2012). S2 platform includes the Multi Spectral Instrument (MSI) with 13 spectral bands covering the Visible and Near Infra-Red (VNIR) and the Short Wave Infra-Red (SWIR) spectral ranges, with a spatial resolution ranging from 10 m to 60 m depending on the band.

The S3 mission will provide continuity to ENVISAT's capabilities while contributing to a number of services related to ocean and land products (Donlon et al., 2012). The main instruments on board the S3 mission are the Ocean and Land Colour Imager (OLCI) and the Sea and Land Surface Temperature Radiometer (SLSTR). OLCI is a push-broom imaging spectrometer instrument building on the heritage of ENVISAT MERIS, with 21 spectral bands covering the 0.4–1 μm range at maximum 300 m spatial resolution. SLSTR is a dual view conical imaging radiometer building on the heritage of ENVISAT AATSR. It includes 9 spectral bands covering the 0.5–12 μm spectral range, with two additional bands for active fire detection, and with a spatial resolution of 500 m for the VNIR bands and 1 km for the thermal infrared (TIR) and fire bands.

 * Corresponding author at: Catedrático José Beltrán 2, 46980 Paterna, Valencia. Spain.
 E-mail address: sobrino@uv.es (J.A. Sobrino).

One of the main objectives of the S3 mission is to provide Europe continuity in ENVISAT type measurement capability for determining sea, ice and land surface temperature. In this context, ESA funded the "Synergistic Use of The Sentinel Missions For Estimating And Monitoring Land Surface Temperature (SEN4LST)" project, which had the main objective of fully utilizing the synergy between SLSTR and OLCI instruments to improve atmospheric correction (including cloud mask) and land surface emissivity characterization for a better estimation of land surface temperature (LST). The SEN4LST project also included an analysis of high resolution data obtained from S2/MSI to improve characterization of the subpixel heterogeneity of the land surface and to potentially use a land cover map derived from S2 for a better parameterization of emissivity retrieval. A general scheme of potential synergies between the different instruments (S2/MSI, S3/OLCI, S3/SLSTR) is presented in Fig. 1. This paper focuses on the performance of the proposed LST retrieval algorithm for the future S3 mission using pairs of MERIS/AATSR images as a proxy.

2. Review and selection of the candidate LST algorithm

2.1. Land surface and emissivity retrieval

The selection of the land surface temperature (LST) algorithm to be applied was based on a literature review and lessons learned from the development and validation of algorithms for ENVISAT MERIS/AATSR (e.g. Sòria & Sobrino, 2007). Operational and near-real time capabilities were also main drivers in the selection of the algorithm.

A number of algorithms for retrieving LST and land surface emissivity (LSE) from remote sensing data have been published in recent decades. An exhaustive review can be found in Li et al. (2013a, 2013b). Since the 1970s, operational LST algorithms for polar-orbiting and geostationary satellites have been mainly based on the split-window (SW) or two-channel (TC) technique. The split-window technique is based on the differential absorption concept (McMillin, 1975), in which the atmospheric effect over the measured signal is corrected using the difference signal measured in two TIR bands at two different wavelengths. This concept also applies to the signal measured by one TIR band (single wavelength) under two different view angles, which is the basis of the so-called dual-angle (DA) algorithms. Both SW and DA LST algorithms depend on LSE, which is assumed to be known a priori by means of other independent approaches. With the launch of the ASTER sensor, with multispectral capabilities in the TIR region, new techniques based on the Temperature and Emissivity Separation (TES) concept were adopted (e.g. Gillespie et al., 1998). The main advantage of the TES technique is that LST and LSE are simultaneously retrieved. Since the latter is also retrieved from TIR data, this implies that the retrieved LSE is sensitive to real changes on the surface and even more indicative of land cover change than traditional vegetation indices in some cases, especially when emissivities in the range 8–9 μm are used (French et al., 2008). Although currently the TES algorithm is only operationally applied to ASTER data, different initiatives are planned to provide LST and LSE products using this approach on other sensors, like the MODIS LST MOD21 product (Hulley & Hook, 2011) or a potentially new LST/ LSE product with MSG/SEVIRI data (Jiménez-Muñoz, Sobrino, Mattar, Hulley, & Gottsche, 2014a). Note that apart from the SW/TC, DA and TES algorithms, it is also possible to retrieve LST from only one TIR band under one single observation, the so-called mono-window or single-channel (SC) algorithm, a technique mainly used for the Landsat series (e.g. Jiménez-Muñoz et al., 2009a; Jiménez-Muñoz, Sobrino, Skokovic, Mattar, & Cristóbal, 2014b).

In terms of operational LSE retrieval from polar-orbiting and geostationary satellites, and because of the unavailability of multispectral TIR sensors, most methods have been based on Vegetation Indices (VIs) (e.g. Sobrino et al., 2008) and/or classification-based approaches (Peres & Da Camara, 2005; Snyder, Wan, Zhang, & Feng, 1998). Only since the launch of ASTER, operational LSE retrievals from multispectral TIR data with TES approaches have been possible.
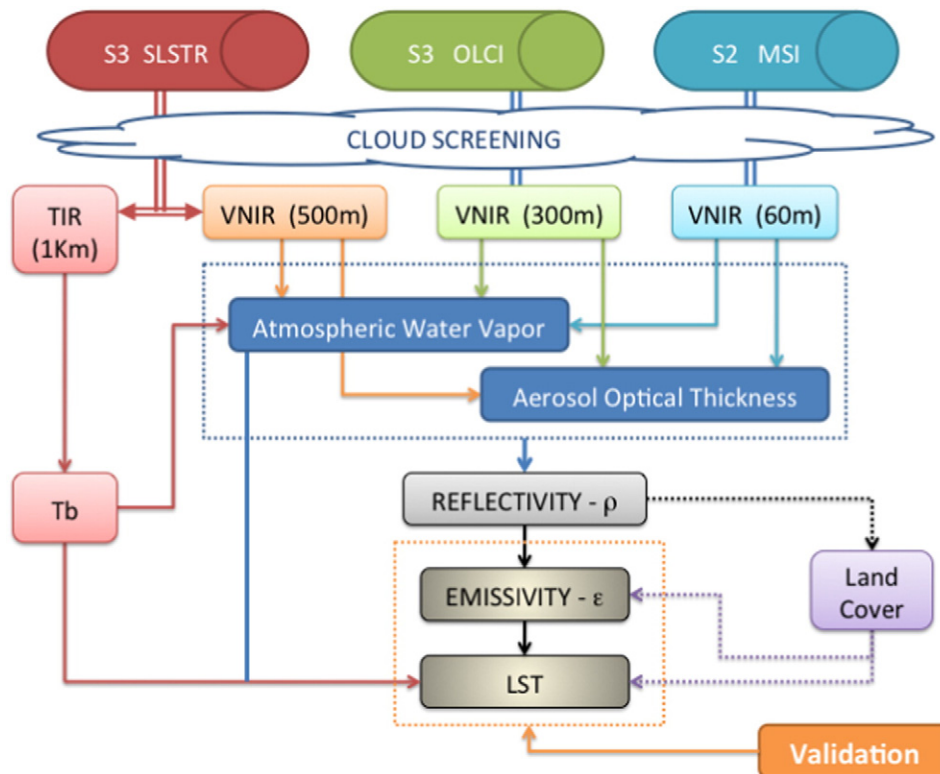

Fig. 1. Synergistic approach of the S2/MSI, S3/OLCI, and S3/SLSTR instruments with focus on a better characterization of the surface emissivity in order to improve land surface temperature retrievals.

Since the S3/SLSTR instrument only has two TIR bands, TES approaches had to be discarded from the list of candidate LST algorithms. Moreover, in order to allow for near-real time retrievals and to avoid the dependence on nighttime acquisitions, methods based on day/night pairs (Goïta & Royer, 1997; Wan & Li, 1997) were also discarded. Therefore, the analysis of algorithms focused on SW and DA algorithms. Among the different mathematical structures proposed for these algorithms, the expression proposed by Sobrino, Li, Stoll, and Becker (1996) was finally selected:

$$T_S = T_i + c_1(T_i - T_j) + c_2(T_i - T_j)^2 + c_0$$

$$+ (c_3 + c_4 W)(1 - \varepsilon) + (c_5 + c_6 W)\Delta\varepsilon \qquad (1)$$

where Ts is the LST (in K), $T_{i,j}$ are at-sensor brightness temperatures (in K), W is the atmospheric water vapor content (in $g \cdot cm^{-2}$ or cm), ε is the mean LSE $0.5 \cdot (\varepsilon_i + \varepsilon_j)$, and $\Delta\varepsilon$ is the LSE difference ($\varepsilon_i - \varepsilon_j$). Subindices 'i' and 'j' refer to two different TIR bands, thus leading to the SW algorithm, or to one TIR band but two different view angles (e.g. nadir 'n' and oblique 'o' views), thus leading to the DA algorithm. Coefficients c_0 to c_6 are obtained from statistical regressions performed over simulated data.

The main advantage of Eq. (1) is that this expression is physically-based (it can be directly derived from the radiative transfer equation after several assumptions), and it explicitly includes the atmospheric water vapor content and surface emissivity. In terms of operational retrieval, the algorithm given by Eq. (1) can easily incorporate external sources of W and LSE. Hence, whenever improved products of W and LSE are available, the potential LST product generated with this algorithm can be easily reprocessed to provide a new product version.

For input emissivities, we adopted the Normalized Difference Vegetation Index (NDVI) Thresholds Method (NDVI-THM; Sobrino et al., 2008) given by.

$$\begin{aligned} NDVI &< NDVIs : & \varepsilon &= a + b\rho_{red} \\ NDVIs &\leq NDVI \leq NDVIv : & \varepsilon &= \varepsilon_s(1 - Pv) + \varepsilon_v Pv + C \\ NDVIv &> NDVI : & \varepsilon &= 0.99 \end{aligned} \qquad (2)$$

where ρ_{red} is the reflectance at the red band, ε_s and ε_v are reference values of surface emissivity for soil and vegetation, respectively, C is a cavity term for rough surfaces which depends on geometrical factors (which can be neglected for operational purposes or assumed to be constant after estimation of a mean value for a number of typical geometries), and Pv is the fractional vegetation cover, which can be obtained from the scaled NDVI (Gutman & Ignatov, 1998):

$$Pv = \frac{NDVI - NDVIs}{NDVIv - NDVIs} \qquad (3)$$

where NDVI is the current value and NDVIs and NDVIv are reference values of NDVI for bare soil and fully vegetated surfaces, respectively. These reference NDVI values can be obtained from different techniques, although values of 0.15 and 0.9 for NDVIs and NDVIv, respectively, are considered representative at global conditions (Jiménez-Muñoz et al., 2009b). The NDVI-THM method was selected because it requires reduced computing time and allows for near-real time implementation.

For the input of emissivity data to the SW algorithm, Eq. (2) must be adapted to each sensor band because of the wavelength dependence of emissivity. For the DA algorithm, the angular dependence on emissivity should be also taken into account. Note also that NDVI-THM (Eq. (2)) applies only to a mixing of bare soil and vegetation surfaces. For snow/ice or water surfaces, these pixels should be identified in order to assign appropriate emissivity values to these two classes.

2.2. Atmospheric correction and cloud screening

In the framework of the SEN4LST project, a review and adaptation of atmospheric correction and cloud screening methodologies for potential application to the S3 mission was also performed. In terms of the atmospheric correction, total atmospheric water vapor content is a required input for the LST algorithm (Eq. (1)), whereas at-surface reflectances are also required for a correct estimation of the surface emissivity using approaches based on vegetation indices (e.g. NDVI-THM). It is not the scope of this paper to provide a detailed description of atmospheric correction and cloud screening methodologies, but in this section we summarize the main ideas behind this analysis based on the developments performed in the framework of the MERIS/AATSR Synergy project.

Accurate atmospheric correction requires information on aerosol scattering to obtain surface reflectance. In addition, independent information on atmospheric water vapor contents may be used to enhance the correction of atmospheric effects. While the effects of aerosols and absorbing gases may be modelled, the variability of aerosol scattering typically represents the greatest uncertainty in the derivation of surface reflectance over land and ocean from remotely-sensed data (Vermote & Kotchenova, 2008). The magnitude of the aerosol scattering increases with view angle. Therefore a robust correction is particularly important when making use of the AATSR or SLSTR off-nadir view (and also in high latitudes with nadir view), and will be of greater importance for Sentinel-3 than it was for ENVISAT due to the wider swath width employed. Use of multiple view-angle imagery, such as the one available on AATSR and SLSTR, allows for robust retrieval even over bright land surfaces. A multi-instrument synergistic approach to aerosol retrieval was recently developed to use both spectral and angular constraints, applicable to both ENVISAT and Sentinel-3. The proposed aerosol correction for this study was based on the Synergy method developed for MERIS and AATSR (North et al., 2008; North et al., 2010), which can be potentially applied to S3 SLSTR and OLCI (North & Heckel, 2012). The output is aerosol optical depth at a reference waveband, an estimate of aerosol model and Angstrom coefficient, and atmospherically corrected surface reflectances for all bands used.

In order to provide a cloud screening strategy for the SEN4LST project, three different methodologies were considered:

i. The operational AATSR cloud masking (Birks, 2007) comprises a series of threshold tests developed initially for cloud identification over the ocean (Zavody, Mutlow, & Llewellyn-Jones, 2000) based on an evolution of the APOLLO cloud clearing for AVHRR (Saunders, 1986; Saunders & Kriebel, 1988) and later adapted for use over land (Birks, 2007).

ii. SYNERGY and GlobAlbedo cloud masks generate a cloud probability from an ensemble of multilayer perceptron artificial neural networks using data from both the AATSR and the MERIS instruments (Gomez-Chova et al. 2007; Gomez-Chova et al. 2009a,b). The GlobAlbedo cloud-clearing algorithm can be applied to the AATSR, MERIS and Satellite Pour l'Observation de la Terre Vegetation (SPOT VGT) instruments. It is based on identifying a series of features useful for cloud detection. Each of these features is assigned a probabilistic value between 0 and 1, where 1 represents a feature that is definitely true, 0 a feature that is definitely false, and 0.5 that the status of the feature is unknown (Brockmann, Kruger, & Danne, 2011).

iii. The Bayesian approach (exploiting in this case only thermal infrared channels) (Merchant, Harris, Maturi, & MacCallum, 2005) calculates the probability of clear-sky conditions, given both the satellite observations and prior knowledge of the surface state and total column water vapor from numerical weather prediction (NWP) data. Previous studies give a detailed evaluation of the cloud detection algorithm performance over the ocean (Merchant et al., 2005), and some initial validation of the scheme applied to land imagery has

been undertaken using data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) (Mackie, Merchant, Old, Embury, & Francis, 2010).

The inter-comparison of the different methodologies using 25 scenes showed that the SYNERGY mask performs better, while the Bayesian performs remarkably well given its simple implementation with a limited number of channels. One of the areas of relative weakness of the SYNERGY mask was in detection of thin cirrus over land, which hopefully will be improved with the new bands of SLSTR.

2.3. Final selection of the LST algorithm

As discussed in Section 2.1, the mathematical structure given by Eq. (1) can be applied to both SW and DA algorithms. In the framework of the SEN4LST project, a combination between a SW and a DA algorithm (denoted as "Combined" algorithm, CO) was also considered. In the simulation procedure (detailed in Section 3.1), the uncertainty of the algorithm for the DA and CO cases was lower than the SW. This result suggested a priori improvement on the LST retrieval when the combination between nadir and oblique views are used. However, based on previous experiences (e.g. Sòria & Sobrino, 2007) and also on the validation of the DA and CO algorithms against in situ measurements during the SEN4LST project, it was observed that practical application of the combination between nadir and oblique acquisitions added significant uncertainty on the interpretation and retrieval of the LST because of the intrinsic thermal heterogeneity over land surfaces, as opposed to the thermal homogeneity over sea surfaces. For this reason, the DA and CO algorithms were discarded and a SW approach based on Eq. (1) was finally selected. A similar conclusion was drawn for the generation of the AATSR LST Standard Product (described in Section 3.2), in which the SW approach was preferred to the DA approach (whereas for the case of SST retrieval a DA approach was adopted). Therefore, results presented in the next sections will focus on a SW algorithm given by Eq. (1).

3. Data and processing

3.1. Simulated data

Algorithm coefficients involved in Eq. (1) are calculated after statistical fits (minimization) from a complete simulated database. For this purpose, at-sensor brightness temperatures T_i and T_j are obtained from forward simulations using the radiative transfer equation (RTE):

$$L_i^{sensor} = [\epsilon_i B_i(T_s) + (1-\epsilon_i)L_i^{atm\downarrow}]\tau_i + L_i^{atm\uparrow} \tag{4}$$

where values of ϵ are obtained from spectral libraries and atmospheric parameters (transmisivity, τ, and up-welling and down-welling atmospheric radiances, $L^{atm\uparrow}$ and $L^{atm\downarrow}$) are obtained from atmospheric profiles and MODTRANv4 radiative transfer code (Berk et al., 1999). All the spectral magnitudes are averaged using the spectral response function of the sensor TIR bands.

In this study, we used 108 emissivity spectra extracted from the ASTER spectral library (Baldridge, Hook, Grove, & Rivera, 2009) and 61 atmospheric profiles under five different surface temperatures. Emissivity spectra included natural samples of soils, vegetation, water, ice, and those rocks samples with sizes representative at remote sensing scales. The atmospheric profiles dataset is based on a selection of 61 atmospheric profiles extracted from the Thermodynamic Initial Guess Retrieval (TIGR) database. These atmospheric profiles represent a worldwide set of atmospheric situations, with 28 atm assigned to the tropical model, 12 to the midlatitude summer model, 12 to the subarctic winter, and 9 to the U. S. Standard. W values range from 0 to 6 cm, whereas temperatures at the first layer range from 240 to 330 K. In

the simulation, LST was chosen as $T_0 - 5$ K, T_0, $T_0 + 5$ K, $T_0 + 10$ K, and $T_0 + 20$ K, where T_0 is the temperature of the first layer of the atmospheric profile. The emissivity and atmospheric profiles datasets, as well as the simulation procedure are described in detail in Jiménez-Muñoz et al. (2009a) and Jiménez-Muñoz and Sobrino (2008).

3.2. MERIS and AATSR imagery

A set of MERIS (level 2; L2) and AATSR (level 1b; L1B) pairs were used as a proxy for the future S3 OLCI & SLSTR imagery. This dataset includes the necessary inputs in order to apply the selected algorithm and to intercompare the results obtained with the standard AATSR L2 LST products.

The AATSR L1B product (ATS_TOA_1P) includes TOA reflectances at bands 0.55 μm, 0.67 μm, 0.87 μm and 1.6 μm, and at-sensor brightness temperatures at bands 3.7 μm, 11 μm and 12 μm, both for nadir and forward views. The product also includes several metadata and cloud flags. MERIS L2 products (MER_RR_2P) include different geophysical variables related to both land and water studies. For this study, we used the at-surface reflectance product for computation of the NDVI using atmospherically corrected data, and the atmospheric water vapor product, a required input to the LST algorithms.

The standard AATSR LST product (level 2, ATS_NR_2P) was used as a reference for the assessment of performance of the prototype product. These are geophysical quantities, retrieved from single pass, single sensor data, used for regional or local studies and available a 1-km grid, Gridded Surface Temperature (GST). It includes the LST and other variables such as sea surface temperature (SST), cloud top temperature, cloud top height and NDVI, with several metadata and flags. The LST algorithm used in the AATSR level 2 products is given by

$$T_s = a_0 + b_0(T_i - T_j)^n + (b_0 + c_0)T_j \tag{5}$$

In comparison to the selected SW algorithm given by Eq. (1), the AATSR L2 algorithm does not include W and LSE in the formulation, but they are indirectly included in the coefficients a_0, b_0, c_0, which depend on the land cover, vegetation fraction, and water vapor (Prata, 2002).

3.3. In-situ data and validation strategy

Five different medium to large homogeneous areas are chosen as test sites in order to validate the LST retrievals, from North to South: 1) an arctic tundra area in Alaska (US), 2) a cropland area in Oklahoma (US), 3) a marshland area in Huelva (Spain) with natural vegetation and periodic flooding; 4) a soil area covered by seasonal grass and sparse trees in Dahra (Senegal); and 5) gravel plains covered by a mixture of gravel, sand, and sparse grass near the Gobabeb Research & Training Centre in the Namib Desert (Namibia). Location of the test sites are shown in Fig. 2, and they will be denoted as Barrow, Oklahoma, Doñana, Dahra, and Gobabeb, respectively.

Continuous measurements of radiometric temperature are performed in all the five sites (every minute or 10 min depending on the site), and values available for year 2011 were selected for the validation exercise. In situ values of emissivity required for the derivation of in situ LSTs over each site were obtained differently depending on the site. In the case of the Doñana site, measurements with a multiband radiometer (including also broadband) were performed in dedicated field campaigns to obtain the emissivities changes along the year (Sobrino, Skokovic, & Jiménez-Muñoz, 2015). The Temperature and Emissivity Separation (TES) algorithm was applied to ground-based measurements of thermal radiances in order to obtain the narrow- and broadband emissivities (Jiménez-Muñoz & Sobrino, 2007). In the case of Dahra and Gobabeb sites, surface emissivity was measured with the box method (Göttsche & Hulley, 2012; Xu, Yu, Tarpley, Göttsche, & Olesen, 2014). For the Barrow and Oklahoma sites, monthly values of
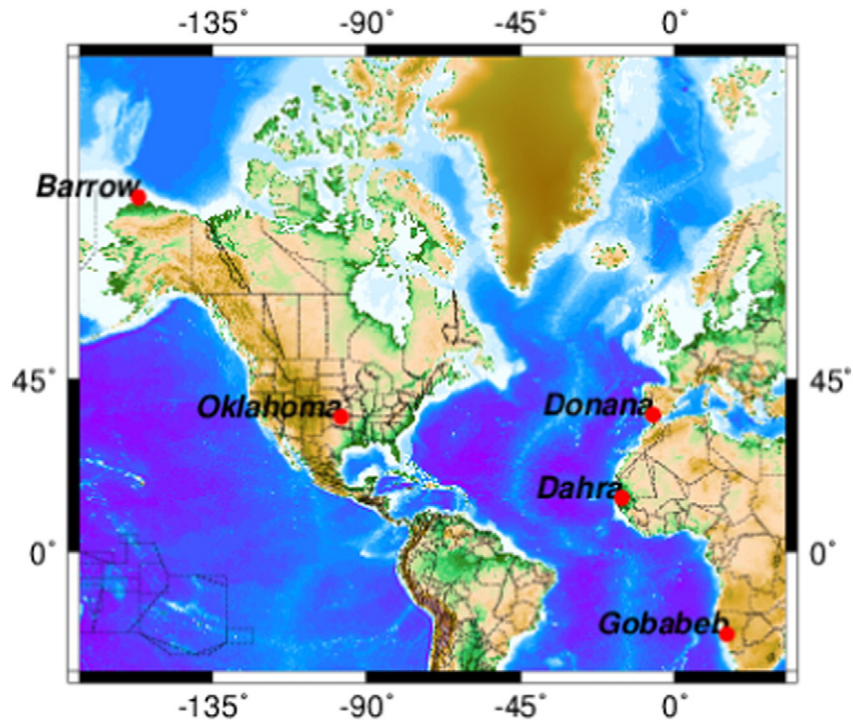
Fig. 2. Test sites used in the validation of the LST algorithms.

surface emissivity were extracted from the CIMSS Baseline Fit Emissivity Database (http://cimss.ssec.wisc.edu/iremis), which provides emissivity values at 0.05° spatial resolution at different wavelengths. Broadband emissivity was obtained after narrow-to-broadband conversion (Borbas & Ruston, 2010; Cheng, Liang, Yao, & Zhang, 2013). Uncertainty of estimated surface emissivities can vary from one site to another, but a mean uncertainty of around 1.5% can be considered for all the sites (see the detailed references). Assuming a surface temperature of 300 K and a surface emissivity of 0.96, this emissivity uncertainty leads to an uncertainty on the retrieved surface temperature of around 0.6 K (Jiménez-Muñoz & Sobrino, 2006).

In order to obtain robust statistics, it is important to define a proper outlier removal strategy, especially to identify pixel values contaminated by clouds. We used cloud flags included in the AATSR standard product (in which three different cloud mask schemes are used, versions, 1, 2 and 3), and also the cloud flags obtained from the MERIS/AATSR SYNERGY methodology. The differences between the cloud masks are analyzed in Section 4.3. Huge differences between satellite-derived and in-situ LSTs were observed during the validation process even when cloud flags were applied. Application of other common outlier removal methodologies such as 3σ criterion (based on the mean value of the difference) or 3σ Hampel identifier (based on the median of the difference) did not solve the problem either. Therefore, we adopted a 3σ criterion based on the theoretical uncertainty associated to the algorithm (provided in Section 4.1) in such a way that differences between satellite-derived and in-situ LSTs higher than three times the algorithm uncertainty were considered outliers and removed from the validation statistics computation.

3.4. Input data to the SW algorithm

Input data to the SW algorithm given by Eq. (1) are i) at-sensor brightness temperatures, ii) surface emissivity at bands 11 and 12 μm, and iii) atmospheric water vapor content. At-sensor brightness temperatures were extracted from the AATSR imagery (Level 1B), whereas atmospheric water vapor was extracted from the MERIS/AATSR

SYNERGY product. Input surface emissivities are the most critical parameter, since it accounts for the highest value of uncertainty on the retrieved LST (see sensitivity analysis in Section 4.1). In this study, different sources of surface emissivity were considered: i) estimations from the NDVI-THM (Eq. (2)), ii) the ASTER Global Emissivity Dataset (ASTER GED) version 4 (Hulley et al. (2015), and iii) the MODIS MOD11C3v5 product (Wan, 2008). For case i), results extracted from the MERIS/AATSR SYNERGY methodology (at-surface reflectance at red band and NDVI) were used. Cases ii) and iii) provide monthly values of emissivity at 5-km spatial resolution. It is important to remark that NDVI-THM can be only applied to daytime acquisitions because its dependence on the VNIR reflectance or NDVI. For nighttime acquisitions, the same value of the daytime emissivity estimation was considered.

3.5. Data extraction: the BEAM toolbox

BEAM is an open-source toolbox and development platform for viewing, analyzing and processing remote sensing raster data (http://www.brockmann-consult.de/cms/web/beam/). An "LST processor" was developed in BEAM in order to retrieve LST from the different algorithms and methodologies developed during the SEN4LST project to be potentially applied to the forthcoming S3 OLCI/SLSTR instruments. BEAM includes the 'MERIS/AATSR Synergy Toolbox', which is based on an atmospheric correction scheme together with a cloud-screening procedure, as explained in Section 2.2. BEAM also includes other useful tools for extracting pixel values for validation exercises or identification of the different quality flags. Results presented in this study were processed and extracted from this BEAM toolbox, as well as from dedicated programs developed in Interactive Data Language (IDL).

4. Results

In this section, we present the results obtained in the simulation procedure for the LST algorithm for both AATSR and SLSTR TIR bands, and for the LSE algorithm for both combinations MERIS/AATSR and SLSTR/OLCI. We also analyze the results obtained in the validation, where

satellite-derived LSTs are compared against in situ measurements over the five test sites using different sources of input emissivity, and we finish off the analysis with some discussion on the disparities observed in the different cloud masks.

4.1. Estimation of uncertainties

4.1.1. Coefficients for the LST SW algorithm and sensitivity analysis

The SW coefficients for Eq. (1) are given in Table 1 for the AATSR and SLSTR TIR bands. The values of the coefficients in these two cases are similar because the spectral configuration of both sensors is similar (Fig. 3). Table 1 also includes the different contributions to the total uncertainty of the LST algorithm. For this purpose, a reference case has been considered, with uncertainties in brightness temperatures, surface emissivity and water vapor of 0.1 K, 0.01 and 0.5 cm, respectively. These reference values of uncertainties have been justified and considered in other simulation exercises (e.g. Jiménez-Muñoz & Sobrino, 2008). The major contribution comes from the uncertainty in the surface emissivity, about 1.2 K. Other contributions come from the standard error of the algorithm (due to the statistical fit), 0.9 K, and the Noise Equivalent Delta Temperature (NEDT), 0.4 K. The uncertainty of the atmospheric water vapor typically has a minor contribution. The final total uncertainty of the algorithm is estimated at 1.5 K in this reference case.

However, in order to better characterize the different uncertainties of the LST algorithm, it is worth performing a more detailed analysis to include other reference uncertainties in water vapor and surface emissivity, as well as to assess the contribution to the algorithm uncertainty of the different SW coefficients. Fig. 4 shows the uncertainty in LST for different values of emissivity uncertainties (Fig. 4a) and water vapor uncertainties (Fig. 4b). As discussed before, the uncertainty in surface emissivity is a key factor for an accurate estimation of the LST. Uncertainties in emissivity above 2% lead to a contribution to the uncertainties in LST above 2.5 K. Overall, the uncertainty in the water vapor has a minor contribution, below 0.3 K, even for huge uncertainties in water vapor. However, this contribution depends on the terms $(1 - \varepsilon)$ and $\Delta\varepsilon$, so for low mean emissivities and/or high spectral variations, the contribution of the uncertainty in water vapor can be important. Hence, for the simulated database used in this study, the lowest mean emissivity value was 0.91, and the highest spectral variation was 0.018. For this extreme case, the contribution of the uncertainty in water vapor is 0.3 K, 0.5 K and 0.8 K for water vapor uncertainties of 0.5 cm, 1 cm, and 1.5 cm, respectively.

Table 1

Coefficients of the split-window algorithm given by Eq. (1) for ENVISAT/AATSR and Sentinel-3/SLSTR cases. Individual contributions to the total uncertainty in LST are also given: standard error of estimation (σ), Pearson linear correlation coefficient (r), uncertainty due to the Noise Equivalent Delta Temperature ($\delta_{NE\Delta T}$), contribution of the uncertainty in the emissivity ($\delta\varepsilon$), contribution of the uncertainty in the total atmospheric water vapor content (δw), and total uncertainty (u_{total}) of the LST algorithm. Uncertainties of 0.1 K, 0.010 and 0.5 cm for brightness temperatures, emissivity and water vapor have been considered, respectively.

Parameter	Units	AATSR	SLSTR
c_0	K	-0.268 ± 0.014	-0.268 ± 0.014
c_1	Unitless	1.029 ± 0.010	1.084 ± 0.010
c_2	K^{-1}	0.2679 ± 0.0017	0.2771 ± 0.0017
c_3	K	44.9 ± 0.7	45.1 ± 0.7
c_4	$K \cdot cm^{-1}$	-0.61 ± 0.19	-0.73 ± 0.19
c_5	K	-121.5 ± 1.7	-125.0 ± 1.7
c_6	$K \cdot cm^{-1}$	16.2 ± 0.5	16.7 ± 0.5
σ	K	0.9	0.9
r	Unitless	0.975	0.976
$\delta_{NE\Delta T}$	K	0.4	0.4
$\delta\varepsilon$	K	1.2	1.2
δw	K	0.08	0.08
u_{total}	K	1.5	1.6
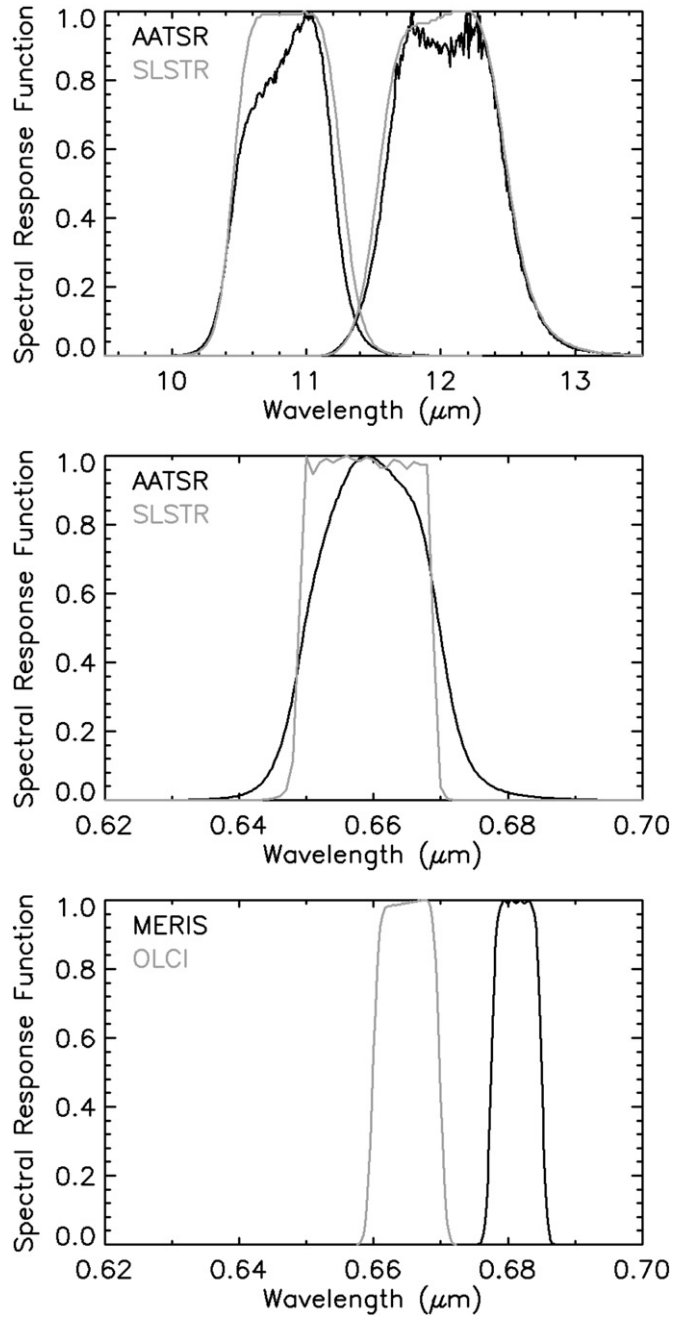
Fig. 3. Spectral Response Functions for AATSR and SLSTR TIR bands, and AATSR, SLSTR, MERIS and OLCI red bands.

It is worth to mention that this sensitivity analysis refers to nadir or near-nadir views. Even in the case of a SW algorithm, in which an oblique view is not used, the angular effect is playing a role for those pixels acquired near the border of the image because of the large swath sensor. In terms of atmospheric absorption, the water vapor content should be corrected for the view angle, since the atmospheric path (and then the atmospheric absorption) increase with increasing view angle. However, the main uncertainty may arise from the angular dependence of the surface emissivity, which is especially critical over land (e.g. Cuenca & Sobrino, 2004; Cuenca & Sobrino, 2004).

4.1.2. Estimation of surface emissivity using the NDVI-THM

The NDVI-THM is based on a linear approach between surface emissivity and red reflectance for bare soil pixels (Eq. (2)), $\varepsilon = a + b\rho_{red}$,
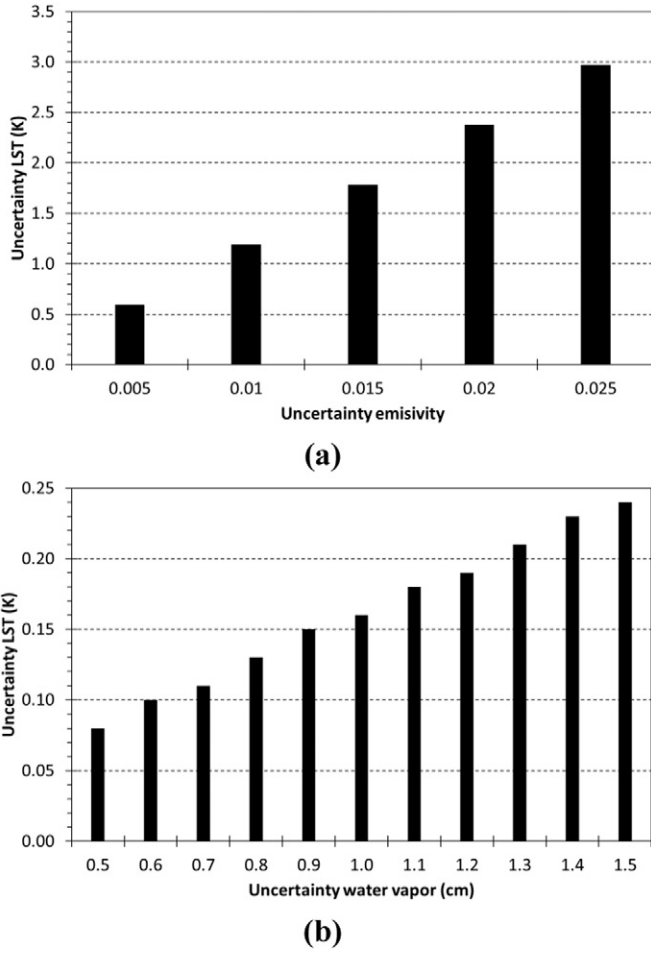
Fig. 4. Impact of the uncertainty in (a) emissivity and (b) atmospheric water vapor over the uncertainty in Land Surface Temperature (LST).

where coefficients a and b can be obtained from spectral libraries. Not all soils follow a linear relationship between ε and ρ_{red}, but it has been found that this linear equation provides acceptable estimations of surface emissivity for the most common soils (Sobrino, Raissouni, & Li, 2001). Table 2 shows the expressions for surface emissivities at AATSR and SLSTR TIR bands and AATSR/MERIS and SLSTR/OLCI red bands, using 27 soil samples included in the ASTER spectral library. The linear correlation coefficient is around 0.7, with standard errors of estimation of 0.003. The coefficients for a given TIR band are also similar because the spectral response functions of the red bands are similar (Fig. 3). Mean emissivity values for the 27 soil samples were 0.969 ± 0.005 for band 11, and 0.977 ± 0.003 for band 12 (with emissivity values ranging from a minimum value of 0.95 to a maximum value of 0.98). These values can be used as reference for ε_s in Eq. (2), whereas a constant value of 0.99 can be considered for ε_v.

4.2. Intercomparison between different sources of surface emissivity

As discussed in Section 3.4, we used three different sources of input emissivities to the SW algorithm: i) NDVI-THM, ii) ASTER-GEDv4, and iii) MOD11C3v5. Before discussing the LST validation results in next section, it is worth to analyze the differences observed in the three emissivity sources, since surface emissivity is a key input to the SW algorithm. Fig. 5 shows the monthly emissivity values (year 2011) for bands 11 and 12 μm obtained from the three datasets over the five test sites. It also includes the monthly NDVI extracted from ASTER-GEDv4 and MERIS. It should be noted that NDVI-THM and MOD11C3v5 emissivitiy refer to bands 11 and 12 μm, whereas ASTER-GEDv4 emissivitiy refer to

ASTER bands 13 and 14 (10.6 μm and 11.3 μm, respectively), which slightly differ from the 11 and 12 μm bands.

The main conclusion drawn from the intercomparison results presented in Fig. 5 is that both NDVI-THM and ASTER-GEDv4 datasets provide more realistic values of surface emissivity than the emissivity derived with MODIS, since they are in accordance with the seasonal variation on NDVI. In contrast, the MOD11C3v5 typically provides constant values of emissivity, which can be attributed to the classification-based approach used to obtain the emissivity in Collection 5. Except for the Barrow site, it is also observed that NDVI-THM emissivities are higher than ASTER-GEDv4 emissivities, especially for band 12. MOD11C3v5 provided the highest values of emissivity. Mean differences between NDVI-THM and ASTER-GEDv4 over the five sites were 0.006 ± 0.008 and 0.011 ± 0.011 for bands 11 and 12, respectively. These differences are within the average absolute band error of ~1% obtained in the validation of the ASTER-GED with lab spectra (Hulley et al., 2015).

4.3. Validation from in-situ measurements

The comparisons between LSTs extracted from the AATSR L2 and SEN4LST algorithms and the LSTs measured in situ are presented in Fig. 6 for the five test sites. Fig. 6 shows the results obtained for both daytime and nighttime acquisitions, and also the results obtained with the SW SEN4LST algorithm using the three emissivity datasets. When all the validation points are considered, the bias ($LST_{algorithm} - LST_{situ}$) obtained with the SEN4LST algorithm and emissivity estimated with the NDVI-THM (SEN4LSTndvi) is lower than the bias obtained with the AATSR L2 algorithm for all the test sites, except for the Gobabeb site, with similar standard deviations for the two algorithms. Consequently, the Root Mean Square Error (RMSE) for the SEN4LSTndvi algorithm is similar or lower than the RMSE for the AATSR L2 algorithm (1.8 K vs 1.8 K for Barrow, 1.8 K vs 3 K for Oklahoma, 2 K vs 2.1 K for Doñana, 2.3 K vs 2.6 K for Dahra), except for the Gobabeb site (2.4 K vs 1.8 K). When daytime and nighttime acquisitions are analyzed separately (Table 3), it is observed that both algorithms typically provide better results for nighttime acquisitions than for daytime acquisitions, with the SEN4LSTndvi providing similar or better results than the AATSR L2 algorithm. The better performance of LST algorithm for nighttime acquisitions can be explained because of the higher thermal homogeneity when the surface is not heated by direct light from the sun, and then the thermal contrast between sunlit spots and shadows is minimized. In the case of daytime acquisitions over the Gobabeb site, the SEN4LSTndvi algorithm provides a similar RMSE than the AATSR L2 algorithm (2.1 K vs 2.2 K), whereas for nighttime acquisitions the RMSE value provided by the SEN4LSTndvi algorithm is higher than the RMSE value provided by the AATSR L2 algorithm (2.2 K vs 1.4 K), mainly due to the higher bias obtained with the SEN4LSTndvi algorithm. An explanation for this result requires further analysis, but it may be attributed to the assumption that daytime emissivities remain unchanged during

Table 2
Coefficients for the linear relationship between surface emissivities (e-TIR) and red reflectances (ρ-RED) used in the NDVI Thresholds Method. Values are provided for the different combinations between AATSR TIR bands, and AATSR and MERIS red bands, as well as SLSTR TIR bands 11 and 12, and SLSTR and OLCI red bands. Pearson linear correlation coefficients (r) and standard error of estimation (σ) are also given.

ε-TIR = slope × ρ-RED + intercept					
ρ-RED	ε-TIR	Slope	Intercept	r	σ
aatsr_b2	aatsr_b11	−0.051 ± 0.009	0.979 ± 0.002	0.73	0.003
aatsr_b2	aatsr_b12	−0.031 ± 0.007	0.9830 ± 0.0016	0.65	0.003
meris_b8	aatsr_b11	−0.050 ± 0.009	0.980 ± 0.002	0.74	0.003
meris_b8	aatsr_b12	−0.030 ± 0.007	0.9832 ± 0.0016	0.65	0.003
slstr_b2	slstr_b11	−0.051 ± 0.009	0.979 ± 0.002	0.73	0.003
slstr_b2	slstr_b12	−0.032 ± 0.007	0.9829 ± 0.0016	0.65	0.003
olci_b8	slstr_b11	−0.051 ± 0.009	0.980 ± 0.002	0.73	0.003
olci_b8	slstr_b12	−0.032 ± 0.007	0.9830 ± 0.0016	0.65	0.003
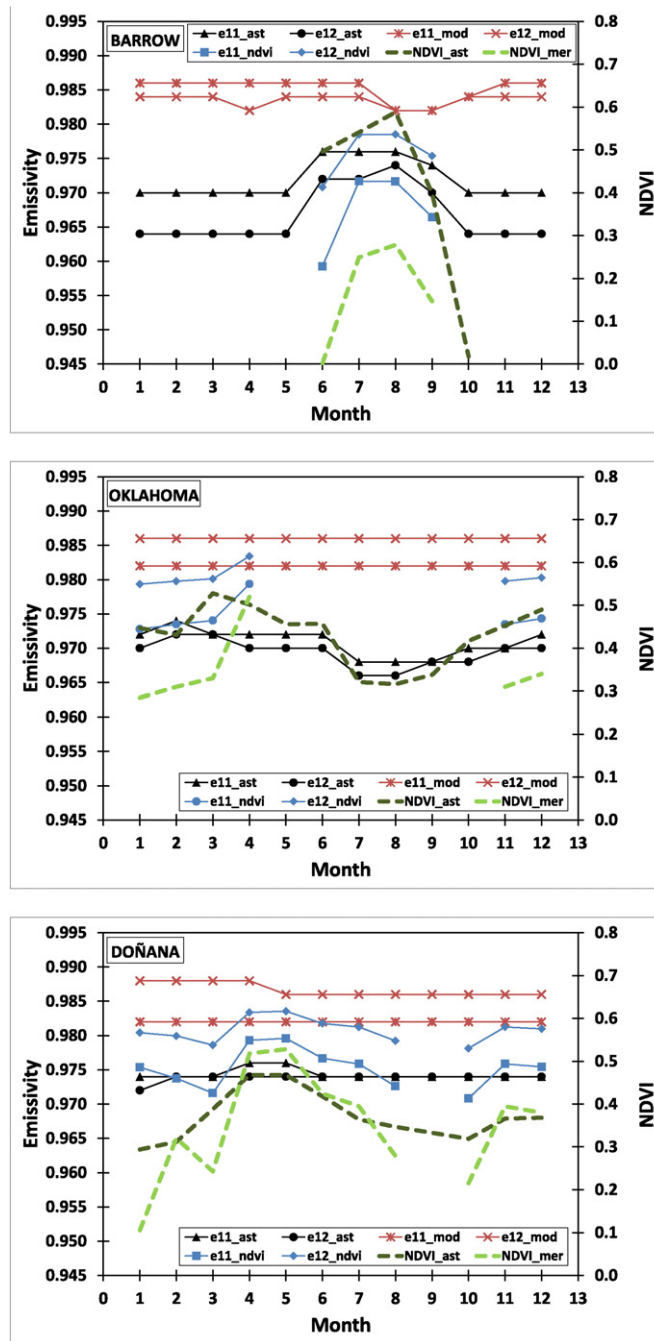
Fig. 5. Monthly values of surface emissivities for bands 11 and 12 μm obtained from the NDVI-THM (e11_ndvi, e12_ndvi), the ASTER-GEDv4 (e11_ast, e12_ast), and the MOD11C3v5 product (e11_mod, e12_mod) over the five test sites. Monthly values of NDVI extracted from ASTER-GEDv4 (NDVI_ast) and MERIS (NDVI_mer) are also provided. In the case of ASTER-GEDv4, e11 and e12 refer to ASTER bands 13 (10.6 μm) and 14 (11.3 μm), whereas in the case of MOD11C3v5, e11 and e12 refer to MODIS bands 31 (11 μm) and 32 (12 μm). Some monthly data are missing in the case of NDVI-THM because in-situ measurements of LST were not available at that particular month or pairs MERIS/AATSR were not available for the validation procedure. In the case of the Barrow site, the surface was covered by snow in month 1 to 5, and 11 to 12, so the NDVI was not computed.

nighttime, as considered for operational application of the NDVI-THM, which may be not accomplished over this particular site.

Fig. 6 also shows the validation results when input emissivities to the SW algorithm are extracted from the ASTER-GED (SEN4LSTaster) and MOD11C3 (SEN4LSTmodis) datasets. As a general rule, SEN4LSTaster provides slightly better results (lower RMSE values) than the SEN4LSTmodis, which agrees with a better emissivity characterization of the ASTER-GED dataset compared to the MOD11C3 product (as discussed in Section 4.2). When LST results obtained

from the three emissivity sources are intercompared (SEN4LSTndvi, SEN4LSTaster, SEN4LSTmodis), it is also typically observed that $RMSE_{aster} < RMSE_{ndvi} < RMSE_{modis}$, although the differences are only slight. These results also agree with the intercomparison among the three emissivity sources (Section 4.2), in which MOD11C3 provided the highest values of emissivity and ASTER-GED provide the lowest values of emissivity, with the NDVI-THM providing intermediate values. In the case of the Oklahoma site, SEN4LSTaster and SEN4LSTmodis provide a similar RMSE (2.3 K),
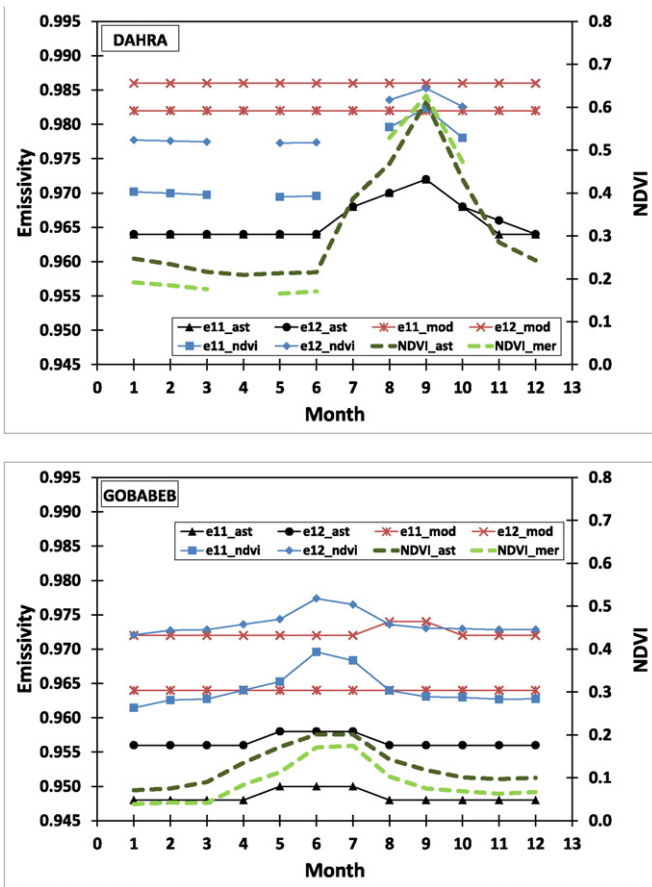
Fig. 5 (*continued*).

but the SEN4LSTndvi algorithm decrease the RMSE to 1.8 K. This improvement can be attributed to the land cover of the Oklahoma site, mainly croplands, in which methods for emissivity retrieval based on NDVI approaches provide better results than other Temperature/Emissivity Separation (TES) algorithms (Jiménez-Muñoz, Sobrino, Gillespie, Sabol, & Gustafson, 2006). In contrast, SEN4LSTndvi provides the worst results over the Gobabeb site (RMSE = 2.4 K), a desert area for which the NDVI-THM may be not applicable (see Table 4). In any case, the RMSE is similar to the value obtained using ASTER-GED emissivity (RMSE = 2.2 K). Note also that in this case the AATSR-L2 algorithm provided the lowest RMSE (1.8 K), although the emissivity is not explicitly present in the algorithm.

4.4. Inter-comparison among the different cloud masks

As explained in Section 3.3, the cloud flags included in both the AATSR-L2 and MERIS/AATSR SYN products did not completely filter out the values contaminated by clouds. However, it is worth analyzing the impact of the different cloud mask versions on the results. Table 5 provides the percentage of cloud/no cloud detections for the three versions of the AATSR-L2 cloud mask. Results are provided for all data points, and also separately for daytime and nighttime acquisitions. Doñana, Dahra and Gobabeb test sites were selected to illustrate the differences in the cloud mask. Percentages were calculated for the data points used in the validation (year 2011). When all data points are considered, version 2 provides a huge percentage of cloud detections (>75%). The other two versions provide similar percentage of cloud detection, with the exception of the Doñana test site where version 1 provides a higher detection of clouds than version 3 does. The percentage of cloud detections for version 2 is reduced for the case of daytime acquisitions, though it is still higher than the other versions for the Dahra and

Gobabeb test sites. However, cloud detection over the Doñana test site for daytime acquisitions is similar for versions 1 and 2 (around 60%). The overestimation of cloud detection for the cloud mask v2 is clearly observed for nighttime acquisitions, where the other versions provide a percentage of cloud detection between 15% and 30%, and version 2 provides 100% cloud detection. Table 5 also includes intercomparison of the different cloud mask versions using version 1 as a reference. The agreement between version 1 and version 3 is typically higher than 85%. For those cases where both versions provide different results, typically version 1 provides a "cloud detection" flag whereas version 3 identifies the values as "no cloud".

Table 6 provides a similar analysis as Table 5, but in this case the MERIS/AATSR SYN cloud mask is also included, and therefore the analysis is reduced to daytime acquisitions. Results show that SYN cloud mask provides the lowest percentage of cloud detection, with similar results to the AATSR-L2 versions 1 and 3 over the Dahra and Gobabeb test sites. In the case of the Doñana test site, version 3 provides a slightly lower percentage of cloud detections than SYN does, whereas versions 1 and 2 provide similar cloud detections percentages (and significantly higher cloud detections than v3 and SYN masks do).

5. Summary and conclusions

In the framework of the SEN4LST project, different LST algorithms, atmospheric correction and cloud screening methodologies were examined in order to fully exploit the synergy between S2/MSI, S3/OLCI, and S3/SLSTR instruments. In this paper we focused on the LST retrieval algorithm using the TIR bands of the SLSTR instrument, so the synergy approach is not directly linked to the LST algorithm itself but rather to the improvement of the input data to the algorithm, namely, cloud masking, surface emissivity characterization, and atmospheric water vapor content. Although the spectral configuration and high spatial resolution of S2/MSI has the potential to improve the land cover identification, and in turn the emissivity characterization, in this paper we focused only on the synergy between S3 instruments (OLCI and SLSTR), using ENVISAT MERIS and AATSR data as a proxy. Based on previous lessons learned in the development and validation of LST algorithms from AATSR data, where the DA algorithm provided poor results over heterogeneous areas, a SW approach was finally selected as a candidate algorithm for the future SLSTR instrument. The sensitivity analysis from simulated data shows that LST can be retrieved with the SW algorithm with uncertainties around 1.5 K, where the major uncertainty stems from surface emissivity (>1 K). The proposed SW algorithm and the standard AATSR LST product were validated against in situ measurements over five test sites (one site in Alaska, one site in the continental U.S., two African sites and one in Spain). Three different emissivity sources were also considered to extract the input emissivity to the SW algorithm (NDVI-THM, ASTER GED, MOD11C3). Among the three emissivity datasets, the NDVI-THM is actually a retrieval method that allows emissivity estimation from image only data (preferably atmospherically corrected values), whereas the other two datasets (ASTER GED and MOD11C3) are considered as external sources and therefore they are not directly linked to AATSR/MERIS (or future OLCI/SLSTR) acquisitions.

Validation results with in situ measurements over five test sites show a slightly better performance of the SEN4LST SW algorithm compared to the AATSR L2 product over four of the five sites (a better bias for the SEN4LST SW algorithm, and similar standard deviation for both algorithms). RMSE values for the SEN4LST SW algorithm are below 2 K for three sites, and around 2.4 K for the other two sites. The highest RMSE is obtained over Oklahoma and Dahra sites with the AATSR L2 algorithm, 3 K and 2.6 K, respectively. In the case of the Oklahoma site, the SEN4LST SW algorithm especially improves the accuracy of the retrieved LST, with a RMSE of 1.8 K. In contrast, the SEN4LST SW provides the highest RMSE value over the Gobabeb site (2.4 K), with the AATSR-L2 product providing a RMSE of 1.8 K. In this case the better performance of the AATSR L2 product is only observed in nighttime
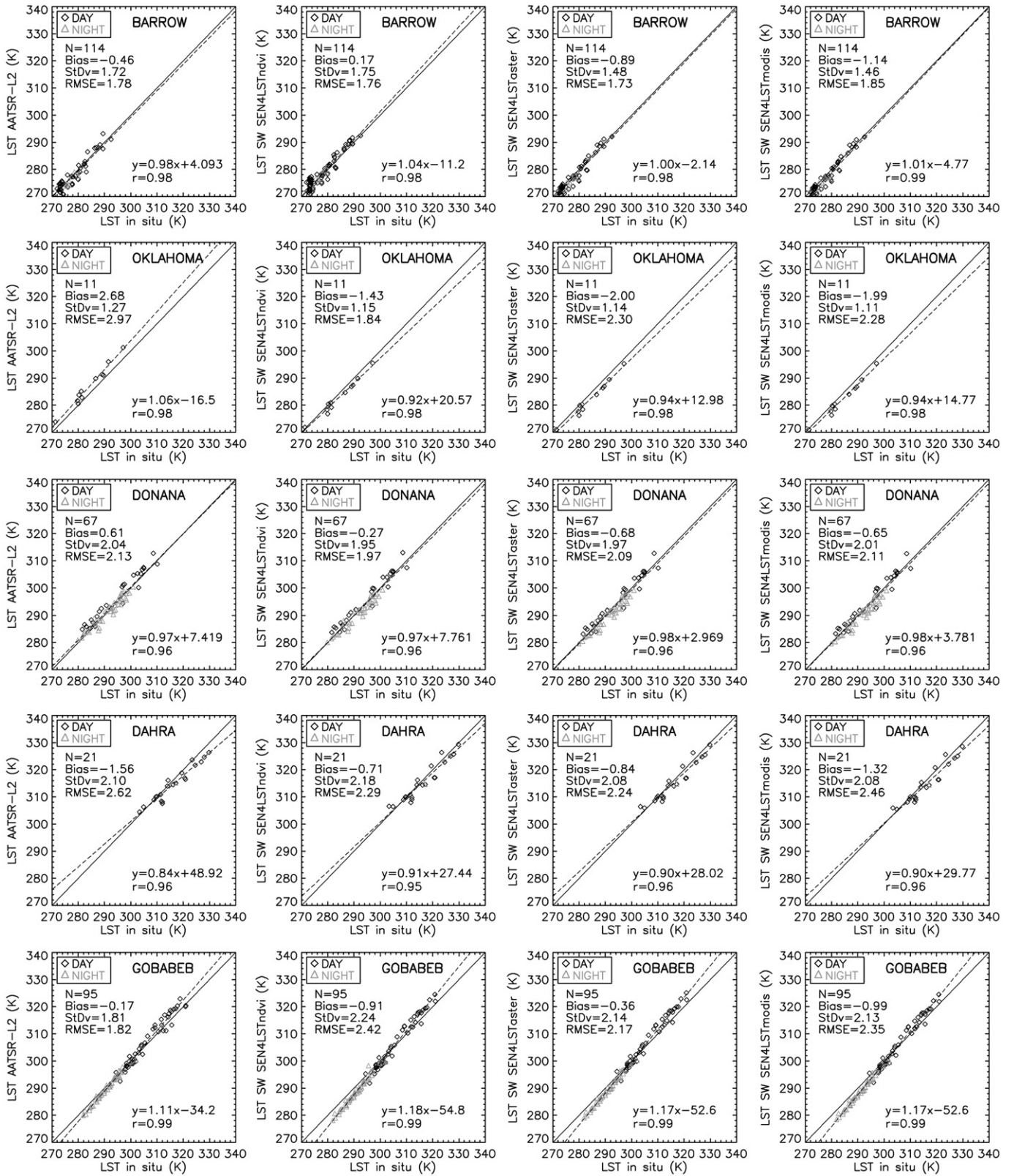
Fig. 6. Validation of the AATSR-L2 and SEN4LST products over the Barrow, Oklahoma, Doñana, Dahra and Gobabeb test sites. Input emissivities to the SW SEN4LST algorithm were extracted from the NDVI-THM (SEN4LSTndvi), the ASTER Global Emissivity Database (SEN4LSTaster), and the MODIS MOD11C3 product (SEN4LSTmodis). Statistics included in the graphs refer to all data points. Results obtained for daytime and nighttime acquisitions are provided in Table 3.

acquisitions (RMSE of 1.4 K for the AATSR L2 product and 2.2 K for the SEN4LST SW algorithm), since both algorithms provide a similar RMSE for daytime acquisitions (around 2.2 K).

It is important to remark that the proposed SEN4LST SW algorithm is an emissivity explicit algorithm (Eq. (1)), in contrast to the surface-type dependent algorithm used in the standard AATSR L2 LST product, in

Table 3

Statistics for the validation of the LST algorithms (SW given by Eq. (1) and the standard AATSR L2) over the different test sites. Input emissivities to the SW algorithm were extracted from the NDVI-THM (SEN4LSTndvi), the ASTER Global Emissivity Database (SEN4LSTaster), and the MODIS MOD11C3 product (SEN4LSTmodis). 'Bias' is the difference between LST derived from the algorithm and the LST measured in situ, 'SDev' is the 1-sigma standard deviation, and 'RMSE' is the Root Mean Square Error. Oklahoma and Dahra test sites are not included because nighttime data were filtered out, so only daytime acquisitions are available over these two sites (see Fig. 6).

Site	Algorithm	Day				Night			
		N	Bias	SDev	RMSE	N	Bias	SDev	RMSE
Barrow	AATSR-L2	98	−0.4	1.8	1.8	16	−1.1	1.1	1.6
	SEN4LSTndvi	98	0.3	1.8	1.8	16	−0.6	1.2	1.3
	SEN4LSTaster	98	−0.8	1.5	1.7	16	−1.4	1.1	1.8
	SEN4LSTmodis	98	−1.1	1.5	1.8	16	−1.6	1.1	2.0
Doñana	AATSR-L2	23	2.0	2.1	2.9	43	−0.2	1.5	1.5
	SEN4LSTndvi	23	1.2	1.9	2.2	43	−0.7	1.2	1.4
	SEN4LSTaster	23	0.6	2.0	2.1	43	−1.5	1.4	2.0
	SEN4LSTmodis	23	0.6	2.0	2.1	43	−1.4	1.5	2.1
Gobabeb	AATSR-L2	46	0.7	2.1	2.2	49	−1.0	0.9	1.4
	SEN4LSTndvi	46	1.1	1.8	2.1	49	−1.9	1.1	2.2
	SEN4LSTaster	46	1.2	2.0	2.3	49	−1.8	0.9	2.0
	SEN4LSTmodis	46	0.5	2.0	2.1	49	−2.4	0.9	2.6

which surface emissivity is embedded in the algorithm coefficients depending on land cover. It is recognized that significant uncertainty can be introduced into a surface-type dependent algorithm because of mis-classifications (e.g. assessment of the NPP-VIIRS LST algorithm performance; Yu, Liu, Yu, Wang, & Csiszar, 2014). Therefore, it is recommended the use of an emissivity explicit algorithm, which allows for improvements in input emissivity to be directly translated into improved LST retrievals. This fact is illustrated in the validation exercise, in which three different emissivity datasets were used as input to the SEN4LST SW algorithm, whereas it is not possible to include these emissivity sources into the AATSR L2 algorithm. The improvement on the LST retrieval with an emissivity explicit algorithm when input surface emissivities are improved can be demonstrated for the Gobabeb test case. The SEN4LST SW algorithm using emissivities estimated with the NDVI-THM provided a mean difference (bias) of (-0.9 ± 2.4) K, similar to the mean difference obtained when MOD11C3 emissivities are used. Results are slightly improved when ASTER-GED emissivities are considered, with a mean difference of (-0.4 ± 2.1) K. However, when in situ measurements of emissivity are introduced in the SEN4LST SW algorithm (results not shown), the mean difference is (0.4 ± 1.7) K. This example clearly illustrates the benefit of an emissivity explicit algorithm. It is worth to mention that validation results refer to the five test sites selected in this study. Although they are globally distributed and

Table 4

Values of NDVI, atmospheric water vapor (W) and surface emissivities at bands 11 (ε_{11}) and 12 (ε_{12}) over Doñana, Dahra and Gobabeb test sites. NDVI and W were extracted from the MERIS level 2 product, whereas ε_{11} and ε_{12} were calculated with the NDVI Thresholds Method. Mean, standard deviation (σ), minimum and maximum values refer to the dataset used in the validation along the year 2011.

Site	Parameter	Mean	σ	Min.	Max.
Doñana	NDVI	0.099	0.100	0.092	0.332
	W	1.851	0.669	0.717	3.313
	ε_{11}	0.974	0.001	0.973	0.976
	ε_{12}	0.980	0.001	0.979	0.981
Dahra	NDVI	0.196	0.070	0.163	0.475
	W	2.717	1.389	0.463	5.347
	ε_{11}	0.969	0.001	0.967	0.970
	ε_{12}	0.977	0.001	0.976	0.977
Gobabeb	NDVI	0.040	0.005	0.039	0.060
	W	1.363	0.924	0.390	3.783
	ε_{11}	0.968	0.003	0.961	0.969
	ε_{12}	0.976	0.002	0.972	0.977

Table 5

Percentage (%) of data points used in the validation identified as "cloud/no cloud" according to the AATSR level-2 cloud mask (versions 1, 2 and 3). The three last columns include the results obtained for the comparison of Versions 2 and 3 against Version 1. The value '−1' indicates that for a given version the value is identified as "no cloud" whereas version 1 identifies the pixel as "cloud"; the value "0" indicates that both versions provide the same results; the value "1" indicates that for a given version the value is identified as "cloud" whereas version 1 identifies the pixel as "no cloud".

Site	N	Mask	Cloud (%)	No cloud (%)	−1 (%)	0 (%)	1 (%)
All							
Doñana	125	v1	43.2	56.8	–	–	–
		v2	80.8	19.2	8.0	46.4	45.6
		v3	22.4	77.6	23.2	74.4	2.4
Dahra	60	v1	23.3	76.7	–	–	–
		v2	75.0	25.0	0.0	48.3	51.7
		v3	26.7	73.3	5.0	86.7	8.3
Gobabeb	132	v1	13.6	86.4	–	–	–
		v2	81.8	18.2	0.0	31.8	68.2
		v3	12.1	87.9	1.5	98.5	0.0
Day							
Doñana	62	v1	64.5	35.5	–	–	–
		v2	61.3	38.7	16.1	71.0	12.9
		v3	25.8	74.2	38.7	61.3	0.0
Dahra	33	v1	21.2	78.8	–	–	–
		v2	54.5	45.5	0.0	66.7	33.3
		v3	21.2	78.8	6.1	87.9	6.1
Gobabeb	64	v1	10.9	89.1	–	–	–
		v2	62.5	37.5	0.0	48.4	51.6
		v3	9.4	90.6	1.6	98.4	0.0
Night							
Doñana	63	v1	22.2	77.8	–	–	–
		v2	100.0	0.0	0.0	22.2	77.8
		v3	19.0	81.0	7.9	87.3	4.8
Dahra	27	v1	25.9	74.1	–	–	–
		v2	100.0	0.0	0.0	25.9	74.1
		v3	33.3	66.7	3.7	85.2	11.1
Gobabeb	68	v1	16.2	83.8	–	–	–
		v2	100.0	0.0	0.0	16.2	83.8
		v3	14.7	85.3	1.5	98.5	0.0

represent different land covers, validation over additional sites is recommended in order to better understand the performance of the LST algorithm over extreme atmospheric conditions (e.g. dry atmospheres in the Arctic or humid atmospheres in the tropics), where the accuracy of the LST retrievals is expected to be degraded. Note also that two of the sites (Oklahoma and Dahra) provide only a low number of available data, with the available data concentrated on daytime acquisitions and particular months, which may limit the validation results.

Another key factor for accurate LST retrieval and appropriate interpretation of the statistics obtained in validation exercises is the cloud

Table 6

Same as Table 5, but including also the MERIS/AATSR SYNERGY cloud mask. In this case, only statistics for daytime acquisitions are available.

Site	N	Mask	Cloud (%)	No cloud (%)	−1 (%)	0 (%)	1 (%)
Doñana	59	v1	64.4	35.6	23.7	42.4	33.9
		v2	59.3	40.7	22.0	50.8	27.1
		v3	25.4	74.6	30.5	67.8	1.7
		SYN	32.2	67.8	–	–	–
Dahra	33	v1	21.2	78.8	15.2	78.8	6.1
		v2	54.5	45.5	15.2	45.5	39.4
		v3	21.2	78.8	15.2	75.8	9.1
		SYN	15.2	84.8	–	–	–
Gobabeb	60	v1	10.0	90.0	1.7	90.0	8.3
		v2	61.7	38.3	1.7	38.3	60.0
		v3	8.3	91.7	1.7	91.7	6.7
		SYN	1.7	98.3	–	–	–

mask. It is not the scope of this paper to deal with cloud screening methodologies and validation of such methodologies, but the basic analysis presented in Section 4.4, in which different results were obtained for different versions and approaches to the cloud mask, illustrates the problem of the cloud mask in the LST retrieval procedure. Further improvements of the SEN4LST SW algorithm are also on the way, such as an adaptation of SW coefficients to day- and nighttime conditions, improvements on surface emissivity characterization using advanced land cover products, and validation over additional sites.

Acknowledgements

References

Baldridge, A. M., Hook, S. J., Grove, C. I., & Rivera, G. (2009). The ASTER spectral library version 2.0. *Remote Sensing of Environment, 113*, 711–715.

Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., & Hanssen, R. F. (2012). ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment, 120*, 84–90.

Berk, A., Anderson, G. P., Acharya, P. K., Chetwynd, J. H., Bernstein, L. S., Shettle, E. P., ... Adler-Golden, S. M. (1999). *MODTRAN4 User's Manual.* Hanscom, AFB, MA: Air Force Research Laboratory.

Birks, A. R. (2007). *Improvements to the AATSR IPF relating to Land Surface Temperature Retrieval and Cloud Clearing over Land.* Rutherford Appleton Laboratory, Chilton, UK: AATSR Technical Note.

Borbas, E. E., & Ruston, C. (2010). The RTTOV UWiremis IR land surface emissivity module. *Associate Scientist Mission Report Mission No. AS09-04 Document NWPSAF-MO-VS-042* Available on-line at: http://nwpsaf.eu/vs_reports/nwpsaf-mo-vs-042.pdf (last access: 4-march-2016)

Brockmann, C., Kruger, O., & Danne, O. (2011). *Algorithm Theoretical Basis Document — Pixel Classification.* In: GlobAlbedo Algorithm Theoretical Basis Document.

Cheng, J., Liang, S., Yao, Y., & Zhang, X. (2013). Estimating the optimal broadband emissivity spectral range for calculating surface longwave net radiation. *IEEE Geoscience and Remote Sensing Letters, 10*, 401–405.

Cuenca, J., & Sobrino, J. A. (2004). Experimental measurements for studying angular and spectral variation of thermal infrared emissivity. *Applied Optics, 43*(23), 4598–4602.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M. -H., Féménias, P., Frerick, J., ... Sciarra, R. (2012). The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment, 120*, 37–57.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... Bargellini, P. (2012). Sentinel-2: ESA's optical high-resoution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25–36.

French, A. N., Schmugge, T. J., Ritchie, J. C., Hsu, A., Jacob, F., & Ogawa, K. (2008). Detecting land cover change at the Jornada Experimental Range, New Mexico with ASTER emissivities. *Remote Sensing of Environment, 112*(4), 1730–1748.

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

Goïta, K., & Royer, A. (1997). Surface temperature and emissivity separability over land surface from combined TIR and SWIR AVHRR data. *IEEE Transactions on Geoscience and Remote Sensing, 35*(3), 718–733.

Gomez-Chova, L., Camps-Valls, G., Calpe, J., Munoz, J., & Moreno, J. (2009a). *MERIS/AATSR Synergy Algorithms for Cloud Screening, Aerosol Retrieval and Atmospheric Correction: Cloud Screening ATBD.* Spain: University of Valencia.

Gomez-Chova, L., Camps-Valls, G., Calpe-Maravilla, J., Guanter, L., & Moreno, J. (2007). Cloud-screening algorithm for ENVISAT/MERIS multispectral images. *IEEE Transactions on Geoscience and Remote Sensing, 45*(12), 4105–4118.

Gómez-Chova, L., Muñoz-Marí, J., Izquierdo-Verdiguier, E., Camps-Valls, G., Calpe-Maravilla, J., & Moreno, J. (2009b). Cloud screening with combined MERIS and AATSR images. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2009. IV.* (pp. 761–764).

Göttsche, F. -M., & Hulley, G. C. (2012). Validation of six satellite-retrieved land surface emissivity products over two land cover types in a hyper-arid region. *Remote Sensing of Environment, 124*, 149–158.

Gutman, G., & Ignatov, A. (1998). The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing, 19*(8), 1533–1543.

Hulley, G. C., & Hook, S. J. (2011). Generating consistent land surface temperature and emissivity products between ASTER and MODIS data for Earth science research. *IEEE Transactions on Geoscience and Remote Sensing, 49*, 1304–1315.

Hulley, G. C., Hook, S. J., Abbott, E., Malakar, N., Islam, T., & Abrams, M. (2015). The ASTER global emissivity dataset (ASTER GED): Mapping Earth's emissivity at 100 meter spatial scale. *Geophysical Research Letters, 42*, 7966–7976.

Jiménez-Muñoz, J. C., & Sobrino, J. A. (2006). Error sources on the land surface temperature retrieved from thermal infrared single channel remote sensing data. *International Journal of Remote Sensing, 27*(5), 999–1014.

Jiménez-Muñoz, J. C., & Sobrino, J. A. (2007). Emissivity spectra obtained from field laboratory measurements using the temperature and emissivity separation algorithm. *Applied Optics, 45*(27), 7104–7109.

Jiménez-Muñoz, J. C., & Sobrino, J. A. (2008). Split-window coefficients for land surface temperature retrieval from low-resolution thermal infrared sensors. *IEEE Geoscience and Remote Sensing Letters, 5*(4), 806–809.

Jiménez-Muñoz, J. C., Cristóbal, J., Sobrino, J. A., Sòria, G., Ninyerola, M., & Pons, X. (2009a). Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Transactions on Geoscience and Remote Sensing, 47*(1), 339–349.

Jiménez-Muñoz, J. C., Sobrino, J. A., Gillespie, A., Sabol, D., & Gustafson, W. T. (2006). Improved land surface emissivities over agricultural areas using ASTER NDVI. *Remote Sensing of Environment, 103*(4), 474–487.

Jiménez-Muñoz, J. C., Sobrino, J. A., Mattar, C., Hulley, G., & Gottsche, F. -M. (2014a). Temperature and emissivity separation from MSG/SEVIRI data. *IEEE Transactions on Geoscience and Remote Sensing, 52*(9), 5937–5951.

Jiménez-Muñoz, J. C., Sobrino, J. A., Plaza, A., Guanter, L., Moreno, J., & Martinez, P. (2009b). Comparison between fractional vegetation cover retrievals from vegetation indices and spectral mixture analysis: Case study of PROBA/CHRIS data over an agricultural area. *Sensors, 9*(2), 768–793.

Jiménez-Muñoz, J. C., Sobrino, J. A., Skokovic, D., Mattar, C., & Cristóbal, J. (2014b). Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Geoscience and Remote Sensing Letters, 11*(10), 1840–1843.

Li, Z. -L., Tang, B. -H., Wu, H., Ren, H., Yan, G., Wan, Z., ... Sobrino, J. A. (2013a). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment, 131*, 14–37.

Li, Z. -L., Wu, H., Wang, N., Qiu, S., Sobrino, J. A., Wan, Z., ... Yan, G. (2013b). Land surface emissivity retrieval from satellite data. *International Journal of Remote Sensing, 34*(9–10), 3084–3127.

Mackie, S., Merchant, C. J., Old, C., Embury, O., & Francis, P. (2010). Generalised Bayesian cloud detection for satellite imagery. Part 1: Technique and validation for nighttime imagery over land and sea. *International Journal of Remote Sensing, 31*(10), 2573–2594.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R., & Berger, M. (2012). *Remote Sensing of Environment, 120*, 91–101.

McMillin, L. M. (1975). Estimation of sea surface temperatures from two infrared window measurements with different absorption. *Journal of Geophysical Research, 80*(36), 5113–5117.

Merchant, C. J., Harris, A. R., Maturi, E., & MacCallum, S. (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quaterly Journal of the Royal Meteorological Society, 131*, 2735–2755.

North, P., & Heckel, A. (2012). Sentinel-3 L2 Products and Algorithm Definition: SYN Algorithm Theoretical Basis Document S3-L2-SD-03-S02-ATBD. available at: https://earth.esa.int/documents/247904/349589/SYN_L2-3_ATBD.pdf (last access: 13 July 2014).

North, P. R. J., Brockmann, C., Fischer, J., Gómez-Chova, L., Grey, W., Hecklel, A., ... Regner, P. (2008). MERIS/AATSR synergy algorithms for cloud screening, aerosol retrieval and atmospheric correction. *Proc. 2nd MERIS/AATSR User Workshop, ESA SP-666.* Noordwijk, The Netherlands: ESA Publications Division.

North, P., Grey, W., Heckel, A., Fischer, J., Preusker, R., & Brockmann, C. (2010). MERIS/AATSR Synergy: Land Aerosol and Surface Reflectance Algorithm Theoretical Basis Document (ATBD), available at: http://github.com/downloads/bcdev/beam-meris-aatsr-synergy/synergy-land_aerosol-atbd.pdf (last access: 13 July 2014).

Peres, L. F., & Da Camara, C. (2005). Emissivity maps to retrieve land-surface temperature from MSG/SEVIRI. *IEEE Transactions on Geoscience and Remote Sensing, 43*(8), 1834–1844.

Prata, F. (2002). *Land Surface Temperature Measurement From Space: AATSR Algorithm Theoretical Basis Document.* Aspendale, Australia: CSIRO Atmospheric Research.

Saunders, R. W. (1986). An automated scheme for the removal of cloud contamination from AVHRR radiances over western Europe. *International Journal of Remote Sensing, 7*, 867–886.

Saunders, R. W., & Kriebel, K. T. (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing, 9*, 123–150.

Snyder, W. C., Wan, Z., Zhang, Y., & Feng, Y. -Z. (1998). Classification-based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing, 19*(14), 2753–2774.

Sobrino, J. A., Jiménez-Muñoz, J. C., Sòria, G., Romaguera, M., Guanter, L., Moreno, J., ... Martínez, P. (2008). Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Transactions on Geoscience and Remote Sensing, 46*(2), 316–327.

Sobrino, J. A., Li, Z. -L, Stoll, M. P., & Becker, F. (1996). Multi-channel and multi-angle algorithms for estimating sea and land surface temperature with ATSR data. *International Journal of Remote Sensing, 17*(11), 2089–2114.

Sobrino, J. A., Raissouni, N., & Li, Z. -L. (2001). A comparative study of land surface emissivity retrieval from NOAA data. *Remote Sensing of Environment, 75*, 256–266.

Sobrino, J. A., Skokovic, D., & Jiménez-Muñoz, D. (2015). Spatial analysis of the homogeneity of the land temperature in three Spanish test sites. *International Journal of Remote Sensing, 36*(19–20), 4793–4807.

Sòria, G., & Sobrino, J. A. (2007). ENVISAT/AATSR derived land surface temperature over a heterogeneous region. *Remote Sensing of Environment, 111*, 409–422.

Vermote, E. F., & Kotchenova, S. (2008). Atmospheric correction for the monitoring of land surfaces. *Journal of Geophysical Research, 113*, D23S90. http://dx.doi.org/10.1029/2007JD009662.

Wan, Z. (2008). New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote Sensing of Environment, 112*, 59–74.

Wan, Z., & Li, Z. -L. (1997). A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote Sensing, 35*(4), 980–996.

Xu, H., Yu, Y., Tarpley, D., Göttsche, F. -M., & Olesen, F. -S. (2014). Evaluation of GOES-R land surface temperature algorithm using SEVIRI satellite retrievals with in situ measurements. *IEEE Transactions on Geoscience and Remote Sensing, 52*(7), 3812–3822.

Yu, Y., Liu, Y., Yu, P., Wang, Z., & Csiszar, I. (2014). Production of land surface temperature for JPSS mission. *In 4th International Symposium Recent Advances in Quantitative Remote Sensing (RAQRS-IV)* Torrent, Spain.

Zavody, A. M., Mutlow, C. T., & Llewellyn-Jones, D. T. (2000). Cloud clearing over the ocean in the processing of data from the Along-Track Scanning Radiometer (ATSR). *Journal of Atmospheric and Oceanic Technology, 17*, 595–615.