
Zhao-Liang Li a,b,⁎, Bo-Hui Tang a, Hua Wu a, Huazhong Ren c, Guangjian Yan c, Zhengming Wan d, Isabel F. Trigo e,f, José A. Sobrino g

a State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Beijing 100101, China
b LSIIT, UDS, CNRS, Boulevard Sebastien Brant, BP10413, 67412 Illkirch, France
c State Key Laboratory of Remote Sensing Sciences, School of Geography, Beijing Normal University, Beijing 100875, China
d ERI, University of California, Santa Barbara, CA 93106, USA
e Instituto Português do Mar e da Atmosfera, Lisbon, Portugal
f Instituto Dom Luiz, University of Lisbon, Portugal
g Image Processing Laboratory, University of Valencia, Valencia 46071, Spain

⁎ Corresponding author at: LSIIT, UDS, CNRS, Boulevard Sebastien Brant, BP10413, 67412 Illkirch, France. Tel.: +33 368854516.
E-mail address: lzl@igsnrr.ac.cn (Z.-L. Li).

Article history:
Received 24 May 2012
Received in revised form 19 October 2012
Accepted 11 December 2012
Available online 9 January 2013

Abstract

Land surface temperature (LST) is one of the key parameters in the physics of land surface processes from local through global scales. The importance of LST is being increasingly recognized and there is a strong interest in developing methodologies to measure LST from space. However, retrieving LST is still a challenging task since the LST retrieval problem is ill-posed. This paper reviews the current status of selected remote sensing algorithms for estimating LST from thermal infrared (TIR) data. A brief theoretical background of the subject is presented along with a survey of the algorithms employed for obtaining LST from space-based TIR measurements. The discussion focuses on TIR data acquired from polar-orbiting satellites because of their widespread use, global applicability and higher spatial resolution compared to geostationary satellites. The theoretical framework and methodologies used to derive the LST from the data are reviewed followed by the methodologies for validating satellite-derived LST. Directions for future research to improve the accuracy of satellite-derived LST are then suggested.

© 2012 Elsevier Inc. Open access under CC BY-NC-ND license.

Keywords:
Land surface temperature
Land surface emissivity
Retrieval
Thermal infrared

Contents

1. Introduction ............................................................... 15
2. Basic theoretical background ........................................ 15
  2.1. Radiative transfer equation ........................................ 16
  2.2. Difficulties and problems in the retrieval of LST from space measurements ........................................ 17
3. Estimation of LST from space ........................................ 18
  3.1. LST retrieval with known LSEs .................................... 18
    3.1.1. Single-channel method ....................................... 18
    3.1.2. Multi-channel method ....................................... 19
    3.1.3. Multi-angle method ......................................... 21
  3.2. LST retrieval with unknown LSEs .................................. 22
    3.2.1. Stepwise retrieval methods .................................. 22
    3.2.2. Simultaneous LST and LSE retrieval methods with known atmospheric information .................................. 23
    3.2.3. Simultaneous retrieval of LST, LSEs, and atmospheric profiles .................................................. 28
  3.3. Comparison and analysis of different methods .................. 28
4. Validation of satellite derived LST ................................... 29
  4.1. Temperature-based method (T-based) ............................ 29
  4.2. Radiance-based method (R-based) ............................... 29
  4.3. Cross validation method .......................................... 30
5. Future development and perspectives .......................................................... 30
  5.1. Methodology to simultaneously derive LST, LSE, and atmospheric profiles (atmospheric quantities) from hyperspectral TIR data .......................................................................................................................... 30
  5.2. Methodology to simultaneously derive LST and LSE from the new generation of geostationary satellites with multispectral and multi-temporal data .................................................................................................................. 31
  5.3. Refinement of LST retrieval algorithms with the consideration of aerosol and cirrus effects .................................................................................................................................................................................. 31
  5.4. Retrieval of component temperatures in heterogeneous pixels .................................................................................................................................................................................. 31
  5.5. Methodology for retrieving LST from passive microwave data and for combining LSTs retrieved from TIR and passive microwave data .................................................................................................. 31
  5.6. Methodology for angular normalization of LST ........................................................................................................................................................................................................ 32
  5.7. Methodology for temporal (time) normalization of LST ................................. 32
  5.8. Concerns on the newly developed Hyperspectral Infrared Imager .................. 32
  5.9. Physical meaning of satellite-derived LST and its applications ....................... 33
  5.10. Validation of satellite-derived LST . .............................................................. 33

Acknowledgments ....................................................................................... 33

References ....................................................................................................... 33

1. Introduction

As the direct driving force in the exchange of long-wave radiation and turbulent heat fluxes at the surface–atmosphere interface, land surface temperature (LST) is one of the most important parameters in the physical processes of surface energy and water balance at local through global scales (Anderson et al., 2008, Brutsaert 2005; Kustas & Anderson, 2003; Kustas et al., 2010; Kustas & Anderson, 1998; Zhao et al., 2008). Knowledge of the LST provides information on the temporal and spatial variations of the surface equilibrium state and is of fundamental importance in many applications (Kerr et al., 2000). As such, the LST is widely used in a variety of fields including evapotranspiration, climate change, hydrological cycle, vegetation monitoring, urban climate and environmental studies, among others (Arnfield, 2003; Bastiaansen et al., 1998; Hansen et al., 2010; Kalma et al., 2008; Kogan, 2001; Su, 2002; Voogt & Oke, 2003; Weng, 2009; Weng et al., 2004) and has been recognized as one of the high-priority parameters of the International Geosphere and Biosphere Program (IGBP) (Townshend et al., 1994). Due to the strong heterogeneity of land surface characteristics such as vegetation, topography, and soil (Liu et al., 2006; Neteler, 2010), LST changes rapidly in space as well as in time (Prata et al., 1995; Vaulcin et al., 1982) and an adequate characterization of LST distribution and its temporal evolution, therefore, requires measurements with detailed spatial and temporal sampling. Given the complexity of surface temperature over land, ground measurements cannot practically provide values over wide areas. With the development of remote sensing from space, satellite data offer the only possibility for measuring LST over the entire globe with sufficiently high temporal resolution and with complete spatially averaged rather than point values.

Satellite-based thermal infrared (TIR) data is directly linked to the LST through the radiative transfer equation. The retrieval of the LST from remotely sensed TIR data has attracted much attention, and its history dates back to the 1970s (McMillin, 1975). To better understand the Earth system at the regional scale and to get the evapotranspiration with an accuracy better than 10%, LST must be retrieved at an accuracy of 1 K or better (Kustas & Norman, 1996; Moran & Jackson, 1991; Wan & Dozier, 1996). However, direct estimation of LST from the radiation emitted in the TIR spectral region is difficult to perform with that accuracy, since the radiances measured by the radiometers onboard satellites depend not only on surface parameters (temperature and emissivity) but also on atmospheric effects (Li & Becker, 1993; Ottlé & Stoll, 1993; Prata et al., 1995). Therefore, besides radiometric calibration and cloud screening, the determination of LSTs from space-based TIR measurements requires both emissivity and atmospheric corrections (Li & Becker, 1993; Vidal, 1991). Many studies have been carried out, and different approaches have been proposed to derive LSTs from satellite TIR data, using a variety of methods to deal with the emissivity and atmospheric effects (Becker & Li, 1990b; Gillespie et al., 1998; Hook et al., 1992; Jiménez-Muñoz & Sobrino, 2003; Kealy & Hook, 1993; Kerr et al., 1992; Pozo Vazquez et al., 1997; Price, 1983, 1984; Qin et al., 2001; Susskind et al., 1984; Tonooka, 2001; Wan & Dozier, 1996; Wan & Li, 1997). Consequently, there have been quite a large number of publications on LST retrieval algorithms and methods. It is important to present an overview of the state of the art in LST retrieval algorithms and to direct future research into improving the accuracy of satellite-derived LST. Although there have been earlier reviews on LST retrieval from space, presented by Prata et al. (1995) and Dash et al. (2002), since then there have been several new developments in LST retrieval algorithms and this review is intended to supplement those reviews with latest approaches. The objective of this paper is to review the progress in estimation of LST from TIR data primarily taken using sensors onboard polar-orbit satellites which have been acquiring data since the mid-eighties and to suggest directions for future research on the subject. Section 2 provides the theoretical basis for retrieving the LST from satellite TIR data and briefly discusses some major difficulties in LST retrieval from space measurements, including: (i) the coupling of the LST, the land surface emissivity (LSE) and the atmosphere; (ii) the physical meaning of satellite-derived LST; and (iii) validation problems related to satellite-derived LST. Section 3 presents an overview of a variety of methods and algorithms for estimating the LST. For each method or algorithm, the main theoretical basis and assumptions involved in the development of the model will be outlined, and the method’s advantages, drawbacks and potential will be highlighted. Section 4 reviews methods of validating satellite-derived LST. Finally, Section 5 suggests future developments and provides perspectives on retrieving LST from remotely sensed data.

2. Basic theoretical background

All objects with temperatures greater than absolute zero emit radiation, and the amount of radiation from a black body in thermal equilibrium at wavelength \( \lambda \) and temperature \( T \) is described by Planck’s law:

\[
B_\lambda(T) = \frac{C_1}{\lambda^5 \exp\left(\frac{C_2}{\lambda T}\right) - 1},
\]

where \( B_\lambda(T) \) is the spectral radiance \( (W \ m^{-2} \ \mu m^{-1} \ sr^{-1}) \) of a black body at temperature \( T \) (K) and wavelength \( \lambda \) (\( \mu m \)); \( C_1 \) and \( C_2 \) are physical constants \( (C_1 = 1.191 \times 10^5 \ \mu m^2 \ sr^{-1} \ m^{-2}, \ C_2 = 1.439 \times 10^4 \ \mu m \cdot K) \). Because most natural objects are non-black bodies, the emissivity \( \varepsilon \), which is defined as the ratio between the radiance of an object and that of a black body at the same temperature, must be taken into account. The spectral radiance of a non-black body is given by the spectral emissivity multiplied by Planck’s law as shown in Eq. (1). Obviously, if the atmosphere exerts no influence on the measured radiance, LST (i.e. \( T \)) can be retrieved by making temperature as the subject of Eq. (1) once the emitted radiance and emissivity are known. The wavelength \( \lambda_{\text{max}} \)
of the peak monochromatic radiance at a given temperature (T) is given by Wien’s displacement law:

\[ T\lambda_{\text{max}} = 2897.9 \ \text{K} \cdot \mu\text{m}. \]  

(2)

According to this equation, the wavelength \( \lambda_{\text{max}} \) at which maximum emission occurs varies roughly from 11.6 \( \mu\text{m} \) to 8.8 \( \mu\text{m} \) if the LST ranges from 250 K to 330 K with the average temperature of the Earth being approximately 288 K. The wavelength region between 8 and 13 \( \mu\text{m} \) coincides within a clear window in the atmosphere which is most transparent to TIR radiation. In cases where the temperature of the surface exceeds 330 K, the wavelength peak moves to shorter and shorter wavelengths, for example for a wildfire (about 800 K), the maximum emission is around 3.6 \( \mu\text{m} \) in the mid-infrared (MIR) region (3–5 \( \mu\text{m} \)) which also coincides with a clear window in the atmosphere.

### 2.1. Radiative transfer equation

An infrared sensor onboard a satellite viewing the Earth’s surface measures the radiation from the Earth and its atmosphere along the line of sight. Using the radiative transfer equation (RTE) and assuming a cloud-free atmosphere under local thermodynamic equilibrium, as illustrated in Fig. 1, the channel infrared radiance \( I_i \) received by a sensor at the top of the atmosphere (TOA) can be written as

\[ I_i(\theta, \varphi) = R_i(\theta, \varphi)\tau_i(\theta, \varphi) + R_{\text{at},i}(\theta, \varphi) + R_{\text{sc},i}(\theta, \varphi), \]

(3)

with \( R_i \) being the channel radiance observed in channel \( i \) at ground level given by

\[ R_i(\theta, \varphi) = \varepsilon_i(\theta, \varphi)B_i(T_s) + (1-\varepsilon_i(\theta, \varphi))R_{\text{at},i} + 1 - \varepsilon_i(\theta, \varphi)|R_{\text{sc},i}|. \]

Surface emission term \quad Surface reflected downwelling atmospheric emission term \quad Surface reflected downwelling atmospheric scattering term \quad Surface reflected downwelling solar beam term

(4)

in which \( \theta \) and \( \varphi \) represent the zenithal and azimuthal viewing angles. For simplicity, the zenithal and azimuthal viewing angles are ignored in the following expressions. \( \tau_i \) is the effective transmittance of the atmosphere in channel \( i \), \( R_i \) is the radiance observed at ground level attenuated by the atmosphere (path ① in Fig. 1). \( R_{\text{at},i} \) is the upward atmospheric thermal radiance (path ② in Fig. 1). \( R_{\text{sc},i} \) is the upward solar diffusion radiance resulting from atmospheric scattering of the solar radiation (path ③ in Fig. 1). \( \varepsilon_i \) and \( T_s \) are the effective surface emissivity and surface temperature in channel \( i \). \( \varepsilon_i B_i(T_s) \) represents the radiance emitted directly by the surface (path ④ in Fig. 1). \( R_{\text{at},i} \) and \( (1-\varepsilon_i)R_{\text{at},i} \) represent the downward atmospheric thermal radiance and solar diffusion radiance reflected by the surface (paths ⑤ and ⑥ in Fig. 1). \( \rho_{\text{bs}} \) is the bi-directional reflectivity of the surface, \( E_i \) is the solar irradiance at the TOA, \( \theta_i \) and \( \varphi_i \) are the solar zenithal and azimuthal angles, \( \rho_{\text{bs}} E_i \cos(\theta_i)\tau_i(\theta_i) \) is the direct solar radiance reflected by the surface (path ⑦ in Fig. 1). Because the contribution of solar radiation at the TOA is negligible in the 8–14 \( \mu\text{m} \) window during both day and night and in the 3–5 \( \mu\text{m} \) window at night, the

Fig. 1. Illustration of radiative transfer equation in infrared regions (see the text for the definitions of symbols). Here, \( I_i \) is the radiance measured by channel \( i \) at the top of atmosphere. Path ① represents the radiance observed at ground level attenuated by the atmosphere. Paths ② and ③ represent the upward atmospheric thermal radiance and the upward solar diffusion radiance, respectively. Path ④ represents the radiance emitted directly by the surface. Paths ⑤ and ⑥ represent the downward atmospheric thermal radiance and solar diffusion radiance reflected by the surface, respectively. Path ⑦ represents the direct solar radiance reflected by the surface.
solar-related items (paths 3, 6 and 7 in Fig. 1) in Eqs. (3) and (4) can be neglected without loss of accuracy.

For convenience and mathematical shorthand, the radiances $I_i$ and $R_i$ measured at the TOA and at ground level are generally expressed in terms of the brightness temperatures where the emissivity is fixed at 1.0. The TOA and ground level brightness temperatures $T_i$ and $T_g$ are defined respectively by

$$B_i(T_i) = I_i \quad \text{and} \quad B_i(T_g) = R_i$$

(5)

It is worth noting that all variables/parameters in Eqs. (3)–(5), except for the angles ($\theta, \phi, \theta_s$, and $\phi_s$), are channel-effective values. Most satellite sensors measure the outgoing radiation with a finite spectral-bandwidth, and the channel-effective quantities of interest are therefore a weighted average expressed by:

$$X_i = \frac{\int_{\lambda_1}^{\lambda_2} g_i(\lambda) X_d d\lambda}{\int_{\lambda_1}^{\lambda_2} g_i(\lambda) d\lambda}. \quad \text{(6)}$$

where $g_i(\lambda)$ is the spectral response function in channel $i$; $\lambda_1$ and $\lambda_2$ are the lower and upper boundaries of the wavelength in channel $i$; and $X$ stands for $B(T), I, R, R_{atm}, R_{sid}, R_{sid}, \theta_s, \phi, s, \tau$, or $\rho_d$.

Eqs. (3) and (4) are actually approximations to the theoretical RTE in which monochromatic quantities are replaced with channel-effective values, but these approximations or simplifications require several important preconditions:

- The integral of a product is assumed to be equal to the product of the integrals. This assumption is true only if the variables are constant within the integration limits, which is rarely the case. Fortunately, the bandwidth of the channel is generally narrow, and the various spectral quantities $X_d$ involved in Eq. (6) should not feature rapid variations. Therefore, the use of the weighted averages defined by Eq. (6) in Eqs. (3) and (4) is a good approximation to the RTE with monochromatic quantities.

- Either the surface is assumed to be Lambertian or the downward atmospheric and solar diffuse radiation are assumed to be isotropic in the calculation of the downward radiations reflected by the surface (paths 3 and 6 in Fig. 1). In practice, these conditions are never fulfilled. However, because the surface-reflected downward atmospheric thermal radiation term is much smaller than the surface thermal emission, and the surface-reflected diffuse solar radiation term is much smaller than the surface-reflected direct solar term, this simplification of Eqs. (3) and (4) is reasonable and does not introduce significant errors.

2.2. Difficulties and problems in the retrieval of LST from space measurements

As seen from Eqs. (3) and (4), estimating the LST from the radiance measured at the TOA requires corrections for both atmospheric and emissivity effects. Applying these corrections is not a simple task, and some key difficulties and problems involved in the retrieval of the LST must be overcome and resolved. These key difficulties and problems are the following:

(1) The retrieval of the LST from space is mathematically underdetermined and unsolvable (Hook et al., 1992; Kealy & Hook, 1993). The RTE described in Eqs. (3) and (4) shows that, if the radiance is measured in $N$ channels, there will always be $N + 1$ unknowns, corresponding to $N$ emissivities in each channel and an unknown LST for $N$ equations, even if quantities other than the emissivities and LST are known a priori. Such an ill-posed problem makes the solution of the RTE sets underdetermined at a ground level even if the atmospheric quantities involved in Eqs. (3) and (4) are accurately estimated. To make LST deterministic, one or more of the LSEs must be known, or the LST and LSEs have to be simultaneously solved with the aid of some assumptions or constraints on the LSEs (Dash et al., 2002; Gillespie et al., 1996; Hook et al., 1992; Kealy & Hook, 1993).

(2) Measurements in the TIR region are highly correlated, implying that instrumental noise and errors in the atmospheric corrections exert strong influences on the accuracy of the LST retrieval. This correlation represents a problem even if the LST is made solvable either by reducing the number of unknowns or by increasing the number of equations through reasonable assumptions or constraints on the LSEs (Gillespie et al., 1996; Li et al., 2013). These highly correlated measurements make LST retrieval unstable and have hampered the methodological development of LST retrieval.

(3) It is difficult to decouple the LST, the LSEs, and the downward atmospheric radiance in the measured radiances. As seen from Eq. (4), the downward atmospheric radiance and the surface emitted radiance are coupled together through LSEs. The non-unity LSE of a natural surface reduces the surface-emitted radiance while increasing the reflection of the downward atmospheric radiance back to the atmosphere, which compensates partly for the reduction in the surface-emitted radiance. This process can reduce or increase the total surface-leaving radiance depending on the atmospheric and surface conditions. This coupling of the reflected downward and surface-emitted radiation can be used to retrieve LST with the online/offline method but requires high spatial resolution data. However, in passively observed multispectral TIR data, it is impossible to separate, on a physical basis, the contributions of the LST from the contributions of the LSEs and the atmosphere in the observed radiance. For this reason, determining LST from space requires not only the atmospheric corrections but also the knowledge of the LSEs and vice versa.

(4) The atmospheric corrections are difficult to implement. The presence of the atmosphere between the surface and the sensors affects the radiances measured by a radiometer at the TOA. These radiances result primarily from emission/reflection at the surface modulated by the effects of the attenuation, and emission of the atmosphere. The atmospheric corrections thus consist of correcting the radiance measured by the sensors for the effects of atmospheric attenuation, emission and emission-reflection. Correcting for the atmospheric effects requires accurate knowledge of the vertical profiles of atmospheric water vapor and temperature both of which are highly variable (Perry & Moran, 1994).

(5) During the daytime, the reflected solar radiation is difficult to remove in the MIR measurements. As mentioned earlier, the highly correlated TIR measurements make LST retrieval unstable even if the solution of the RTE sets becomes deterministic through some assumptions and constraints on the LSEs. In the MIR since the direct solar irradiation reflected by the surface is on the same order of magnitude as the radiance directly emitted by the surface, if the surface albedo is about 0.1, the introduction of the MIR channels in LST retrieval significantly reduces the correlation of the RTE sets and greatly improves the accuracy of the estimated LST (Li et al., 2013). Additionally, MIR channels are less sensitivity to the water vapor in the atmosphere compared with TIR channels, and the LST retrieval from the MIR is only half as sensitive to errors in emissivity as that from the TIR (Mushkin et al., 2005). Therefore, LST retrieval with the MIR instead of the TIR sounds more appropriate. However, solar effects are difficult to eliminate in the MIR during the daytime because the separation of solar irradiation from the total energy measured in the MIR requires not only the accurate atmospheric information but also the knowledge of the bidirectional reflectivity of the surface. This information is typically unknown and affected by several factors (Adams et al., 1989; Mushkin et al., 2005), resulting in
large uncertainties that can lead to an even larger error on the LST retrieved from MIR measurements. Therefore, while the introduction of the MIR channels may benefit the retrieval of the LST in certain cases, it can also introduce even larger uncertainties in others.

(6) How to physically interpret the results of the LST measurement remains a crucial question. As noted by Prata et al. (1995), the definition of the surface temperature may depend strongly on the type of application and the method of measurement. Because the surface temperature $T_s$ in Eq. (4) is defined using the radiance emitted by a surface, this temperature is called the radiometric temperature (or the skin temperature) that corresponds to the radiation emitted from depths less than the penetration depth of a given wavelength (Becker & Li, 1995; Norman & Becker, 1995). The penetration depth is usually within a few millimeters in the TIR region (Wan, 1999). This radiometric temperature physically differs from other definition of temperatures, such as the thermodynamic temperature defined for a medium in thermal equilibrium and measured by a thermometer. For homogeneous and isothermal surfaces, the radiometric and thermodynamic temperatures are reported to be equivalent. As the thermodynamic temperature is actually hard to measure in reality, even for water, the radiometric temperature is often the only practical measure for the homogeneous and isothermal surface. However, most surfaces are not in equilibrium and for heterogeneous and non-isothermal surfaces, these two temperatures are different. Considering that the spatial resolution of the current onboard systems varies approximately from $10^{-2}$ to 10 km, there may be several surface types with different temperatures and emissivities within one pixel, which complicates the physical understanding of the LST values retrieved from space and the relation of the radiometric temperature at large scales to other temperatures used in different applications. To date, no consensus has been reached on the definition of the LST for heterogeneous and non-isothermal surfaces, but the definition by Becker and Li (1995), which depends on the distributions of the LST and the LSE within a pixel, is measurable from space and is thus recommended for LST retrieval from space.

(7) Validation of LST retrieved from spaceborne measurements at the scale of the sensor’s pixels is also challenging. Validation is problematic due to the difficulty of conducting in situ LST measurements, and in obtaining representative LST data at the scale of a single pixel. Generally, temperatures over the land surfaces vary strongly in space and time (Prata et al., 1995), and it is not unusual for the LST to vary by more than 10 K over just a few centimeters of distance or by more than 1 K in less than a minute over certain cover types. Appropriately scaling the satellite-derived LSTs to those measured at ground level, especially at large scale is also difficult. The strong spatial heterogeneity and temporal variation of the LST limits ground-based validation to several relatively homogeneous land surfaces, such as lakes, deserts, and densely vegetated land using direct comparisons of in situ temperature measurements with temperatures retrieved from the satellite data (e.g. Hook et al., 2005, 2003, 2007). A complementary approach is to use sites which are homogenous in terms of emissivity using the radiance based validation approach (Hulley et al., 2009). Sand dunes can be one example of this type of sites which are referred to as pseudo-invariant sites (Hulley et al., 2009). Furthermore, how to perform a representative measurement of the LST of a complex heterogeneous surface is also a critical question. Scaling methods should be developed to relate the measurements at different scales and help validate the retrieved LST (Liu et al., 2006; Wu & Li, 2009).

Besides the difficulties mentioned above in the retrieval of LST from space, the accuracy of the LST data also depends on the performance of the cloud mask used to isolate clear sky data and on the quality of the TIR data, i.e., the stability of the spectral response function $g(\lambda)$, the SNR and the accuracy of the radiometric calibration.

3. Estimation of LST from space

Over the past several decades, LST estimation from satellite TIR measurements has significantly improved. Many algorithms have been proposed to treat the characteristics of various sensors onboard different satellites and utilizing different assumptions and approximations for the RTE and LSEs. These algorithms can be roughly grouped into three categories: single-channel methods, multi-channel methods, and multi-angle methods, provided that the LSEs are known a priori. If the LSEs are not known, then the algorithms can be categorized into three types: stepwise retrieval method, simultaneous retrieval of LSEs and LST with known atmospheric information, and simultaneous retrieval with unknown atmospheric information.

3.1. LST retrieval with known LSEs

3.1.1. Single-channel method

The single TIR channel method, also called the model emissivity method (Hook et al., 1992), uses the radiance measured by the satellite sensor in a single channel, chosen within an atmospheric window, and corrects the radiance for residual atmospheric attenuation and emission using atmospheric transmittance/radiance code that requires input data on the atmospheric profiles. LST is then retrieved from the radiance measured in this channel by inverting the RTE given in Eqs. (3) and (4), provided that the LSE is well known or estimated in advance (Chédin et al., 1985; Hook et al., 1992; Li et al., 2004a; Mushkin et al., 2005; Ottlé & Vidal-Madjar, 1992; Price, 1983; Susskind et al., 1984). Accurate determination of the LST using this method requires high-quality atmospheric transmittance/radiance code to estimate the atmospheric quantities involved in Eqs. (3) and (4), good knowledge of the channel LSE, an accurate atmospheric profile, and a correct consideration of the topographic effects (Sobrino et al., 2004b).

Generally, the accuracy of atmospheric transmittance/radiance code is primarily limited by the radiative transfer model (RTM) used in the code and by the uncertainties in atmospheric molecular absorption coefficients and aerosol absorption/scattering coefficients (Wan, 1999). The most popular atmospheric RTMs, such as the series of MODTRAN (Berk et al., 2003) and 4A/OP (Chaumet et al., 2009), have been widely used to perform atmospheric corrections and/or to simulate satellite TIR data. A few studies have shown that the accuracy of the different RTMs ranges from 0.5% to 2% within known atmospheric windows, such as 3.4–4.1 μm and 8–13 μm, leading to uncertainties in the retrieved LST of 0.4 K to 1.5 K (Wan, 1999). It is worth noting that the incomplete characterizations of atmospheric profiles used in compensation for atmospheric absorption and path radiance constitute a serious problem even if the RTM itself is completely error-free (Gillespie et al., 2011). Studies have also demonstrated that an error of 1% in the LSE can give rise to an error in the LST ranging from 0.3 K for a hot and humid atmosphere to 0.7 K for a cold and dry atmosphere (Dash et al., 2002). As the single channel is usually chosen around 10 μm where the LSE for most land surfaces can be known within a few percent, the uncertainty in LSE may lead to an error of 1 K to 2 K in LST if the single-channel method is used. However, if the LSE is known at a wavelength within the TIR atmospheric window, then any error will be solely due to incorrect removal of the atmospheric contribution. Atmospheric profiles are generally obtained either from ground-based atmospheric radiosoundings, from satellite vertical sounders or from meteorological forecasting models. Due to the high spatial and temporal variability of the atmospheric water vapor, the use of ground-based radiosoundings far from the area of interest and/or from the time of satellite overpass may result in large errors in the LST (Cooper & Asrar, 1989). On top of these, radiosoundings reported measurement errors are of the order of 0.5 K.
for temperature and 10% for water vapor (Ma et al., 1999), and they are mostly associated to two types of uncertainties: 1) the actual sounding location and place where the balloon is set free may be several kilometers apart (up to 60 km); 2) the atmospheric status will change when the balloon rises which means the atmospheric profiles at different heights are measured at different times. Those uncertainties about the atmosphere will further propagate into the retrieved LST. Finally, as radiosoundings are not currently available with sufficient spatial density or in coincidence with the time of the satellite overpass, they can be only used occasionally for validation purposes at some special sites (Coll et al., 2005). In contrast, atmospheric profiles derived from satellite vertical sounders can, in theory, be used to retrieve the LST from concurrent TIR data in the atmospheric window using the single-channel method. Unfortunately, the accuracy of the retrieved atmospheric profiles near the surface is insufficient for the single-channel method (Chédin et al., 1985; Ottlé & Stoll, 1993; Susskind et al., 1984), and large errors in the LST retrieval can result. Today, the profiles provided as forecasts, analysis or reanalysis by numerical weather prediction centers, such as the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF), constitute a practical alternative to the use of radiosoundings. Coll et al. (2012a) and Jiménez-Muñoz et al. (2010) reported either atmospheric profile products or reanalysis may yield reasonable results to meet the required accuracy for many cases. Furthermore, Freitas et al. (2010) quantified the impact of ECMWF forecast errors of atmospheric humidity on LST retrievals, showing that it is generally less than 0.5 K. Nevertheless, the atmospheric profiles are often provided at spatial coarser than that of the satellite and it is therefore necessary to interpolate the atmospheric quantities in terms of the viewing zenith angle, the terrain altitude, space and time (Jiang et al., 2006; Schroeder et al., 2003; Tang & Li, 2008).

To reduce the dependence on radiosounding data, several single-channel algorithms have been proposed within the past decade to estimate the LST from satellite data assuming that the LSE is known. Qin et al. (2001) proposed a method to estimate the LST specifically from Landsat-5 (Thematic Mapper channel 6, TM6) data using only the near-surface air temperature and water vapor content instead of atmospheric profiles using empirical linear relationships between the atmospheric transmittance and the water vapor content and between the mean atmospheric temperature and the near-surface air temperature. Jiménez-Muñoz and Sobrino (2003) and Jiménez-Muñoz et al. (2009) developed a generalized single-channel algorithm for retrieving the LST from any TIR channel with a FWHM (full width at half maximum) of ∼1 μm, provided that the LSE and the total atmospheric water vapor content are known. This generalized single-channel algorithm requires the minimum input data and can be applied to different thermal sensors using the same equation and coefficient. Cristóbal et al. (2009) found that the inclusion of near-surface air temperature together with water vapor content in the single-channel method improves LST retrieval especially under intermediate and high atmospheric water vapor content conditions. Sobrino et al. (2004b), Sobrino and Jiménez-Muñoz (2005) and Jiménez-Muñoz and Sobrino (2010) analyzed and compared the aforementioned algorithms and pointed out that all of the single-channel algorithms that use empirical relationships provide poor results at high atmospheric water vapor contents because the relationships included in the algorithms become unstable at high water vapor concentrations.

It should be kept in mind that the single-channel methods involve a simple inversion of the RTE, provided that the LSE and the atmospheric profiles are known in advance. These methods can provide theoretically accurate LST retrieval, but LSE is rarely known with the necessary accuracy.

3.1.2. Multi-channel method

As highlighted in Section 3.1.1, the use of the single-channel method requires that LSE is known a-priori for each pixel as well as an accurate RTM and accurate knowledge of the atmospheric profiles over the study area at the satellite overpass. These requirements are difficult or even impossible to satisfy in most practical situations. To obtain the LST from satellite TIR data with sufficient accuracy on a global or regional scale, other methods must be developed. An alternate approach used over the oceans utilizes the differential atmospheric absorption in the two adjacent channels centered at 11 and 12 μm in the so-called split-window algorithm (SW) first proposed by McMillin (1975) which does not require any information about the atmospheric profiles at the time of the acquisition. Since then, a variety of SW techniques have been developed and modified to successfully retrieve the sea surface temperature (SST) (Barton et al., 1989; Deschamps & Phulpin, 1980; Franca & Carvalho, 2004; Llewellyn-Jones et al., 1984; McClain et al., 1985; Nicolòs et al., 2007; Sobrino et al., 1993).

Encouraged by the success of the SW method for estimating the SST from space measurements, many efforts have been made since the late 1980s to extend the SW method to retrieve the LST. With modifications to treat the spatio-temporal and spectral variations of the LSE, the large difference between the LST and the air temperature, the total column water vapor (WV) in the atmosphere, and the viewing zenith angle (VZA), a variety of SW algorithms for LST retrieval have been proposed. These algorithms assume that the LSEs in both SW channels are known a priori (Attiar & Sobrino, 2009; Becker, 1987; Becker & Li, 1990b; Coll et al., 1994; Prata, 1994a,b; Price, 1984; Sobrino et al., 1991, 1994, 1996; Tang et al., 2008; Ulivieri et al., 1994; Wan & Dozier, 1996). Below is a broad overview of the different SW algorithms found in the literature.

3.1.2.1. Linear split-window algorithm. A linear SW algorithm can be developed utilizing the differential absorption in two adjacent TIR channels i, j in the 10–12.5 μm linearizing the RTE with respect to the temperature or wavelength. This algorithm expresses the LST as a simple linear combination of the two brightness temperatures Ti and Tj measured in the two TIR channels (McMillin, 1975). A typical linear split-window algorithm can be written as

$$\text{LST} = a_0 + a_1T_1 + a_2(T_1 - T_j),$$

where $a_k$ (k = 0, 1, and 2) are coefficients that depend primarily on the spectral response function of the two channels $g_i(\lambda)$ and $g_j(\lambda)$, the two channel emissivities εi and ej, the WV, and the VZA. Thus:

$$a_k = f_k(\varepsilon_i, \varepsilon_j, T_i, T_j, \text{WV}, \text{VZA}).$$

It should be pointed out that the accuracy of this LST retrieval method is dependent on the correct choice of the coefficients $a_k$, which are pre-determined either by regressing the simulated satellite data with a set of atmospheres and surface parameters or empirically by comparing the satellite data against in situ LST measurements. It is extremely difficult to obtain a representative in situ LST at satellite pixel scale (a few km²) that is also synchronized with the satellite’s measurements over a wide range of surface types and atmospheric conditions. Therefore, the simulation of TOA brightness temperatures, using an RTM such as MODTRAN (Berk et al., 2003), represents the most efficient way to generate the data that allows pre-determining robust estimates of the coefficients $a_k$ by comparing simulated satellite data against the preset LST used in the simulation.

Over the past several decades, many linear SW algorithms have been developed in the literature, all of which have similar forms but different parameterizations of the coefficients $a_k$. These coefficients are parameterized either as linear or non-linear functions of various combinations of the LSEs, the WV, and the VZA.

(1) Parameterization of the coefficients $a_k$ with consideration of LSEs. Various parameterizations of the coefficients $a_k$ in the SW as functions of the LSEs have been developed, and all are somewhat
similar in formulation. Many are based on the linearization of various terms of the RTE. For instance, Becker and Li (1990b) developed a local SW algorithm and parameterized $a_k$ as functions of the mean ($\varepsilon = (\varepsilon_i + \varepsilon_j)/2$) and difference ($\Delta \varepsilon = \varepsilon_i - \varepsilon_j$) of the two channel LSEs

$$a_k = \text{constant},$$

$$a_k = A_{a_k} + A_{b_k} \frac{1 - \varepsilon}{\varepsilon} + A_{c_k} \frac{\Delta \varepsilon}{\varepsilon^2} \tag{k = 1, 2}$$

where the coefficients $A$ are constant and independent of the atmosphere if the atmosphere is relatively dry (WV $\leq 2.5$ g/cm$^2$). Some authors have proposed modifying the constant coefficient $a_0$ in Eq. (7) to correct for emissivity effects while keeping the other coefficients $a_1$ and $a_2$ independent of the LSEs as they are for SST retrieval. Other forms of $a_0$ have been proposed, including $a_0 = B_0 + B_1(1 - \varepsilon) + B_2 \Delta \varepsilon$ (Sobrino et al., 1994; Ulivieri & Cazzaniga, 1985; Ulivieri et al., 1994) or $a_0 = B_0 + B_1(1 - \varepsilon)/\varepsilon + B_2 \Delta \varepsilon/\varepsilon^2$ (Vidal, 1991), in which the coefficients $B$ are sensor dependent and atmospheric-independent. A general form of $a_0$ can be written as $a_0 = f_0(\varepsilon_i, \varepsilon_o, \Delta \varepsilon)$ where $f_0$ is a linear or nonlinear function. Some authors have proposed a general parameterization of $a_0$ as $a_0 = f_0(\varepsilon_i, \varepsilon_o, \Delta \varepsilon)$ (Sun & Pinker, 2003, 2005).

(2) Parameterization of the coefficients $a_k$ with consideration of LSEs and WV

The atmosphere-independent coefficients $a_k$ described above are only valid in relatively dry atmospheres (WV $\leq 2.5$ g/cm$^2$). To make the SW algorithm (Eq. (7)) applicable to more general atmospheric conditions, the WV content in the atmosphere has to be explicitly included in the coefficients $a_k$ as shown in Eq. (8). Many authors have proposed incorporating both the LSEs and the WV into the coefficients $a_k$, and all of them empirically described the coefficients $a_k$ as linear combinations of the LSEs and the WV (Becker & Li, 1995; Sobrino et al., 1991, 2004a, 2004b). François and Ottlé (1996) proposed another parameterization of $a_k$ in which the $a_k$ are expressed as quadratic functions of the WV at a given LSE, i.e.

$$a_k = C_{0_k} + C_{1_k} \text{WV} + C_{2_k} \text{WV}^2, \tag{10}$$

where the coefficients $C$ are constant for a given LSE.

(3) Parameterization of the coefficients $a_k$ with consideration of LSEs, WV, and VZA

Note that the atmospheric transmittance is expected to decrease at larger VZA due to the increased absorption path length. To achieve the required LST accuracy of $1 \text{K}$ over a wide range of atmospheric and surface conditions, error analysis demonstrates that the VZA must also be considered in the LST retrieval algorithm, particularly in hot and humid atmospheres. Many studies have shown that LST retrieval can be significantly improved at VZA larger than 50° by introducing the cosine of the VZA into the parameterization of $a_k$ (Becker & Li, 1995; Minnis & Khaiyer, 2000; Pinheiro et al., 2004). For instance, Becker and Li (1995) further modified their local SW algorithm to explicitly incorporate the cosine of VZA and the WV in the coefficients $a_k$ as in Eq. (8). The function $f_k$ in Eq. (8) is an empirically determined linear combination of LSEs, WV, and cos(VZA). Wan and Dozier (1996) developed a generalized split-window (GSW) algorithm, which uses a formula similar to that proposed by Becker and Li (1990b), to retrieve LST from the Moderate Resolution Imaging Spectroradiometer (MODIS) TIR channels 1 (channel 31) and J (channel 32). The GSW algorithm can be written as:

$$\text{LST} = b_0 + \left( b_1 + b_2 \frac{1 - \varepsilon}{\varepsilon} + b_3 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_i + T_o}{2} + \left( b_4 + b_5 \frac{1 - \varepsilon}{\varepsilon} + b_6 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_i - T_o}{2} \tag{11}$$

where $b_k$ ($k = 0\text{–}6$) are unknown coefficients that need to be determined for a given VZA and for given sub-ranges of $\varepsilon$, atmospheric WV and surface air temperature, i.e., the air temperature at the surface level ($T_o$) (Wan and Dozier, 1996) or LST (Tang et al., 2008). Using the same form as Eq. (7), it can be shown that:

$$a_0 = b_0, \quad a_1 = b_1 + b_2 \frac{1 - \varepsilon}{\varepsilon} + b_3 \frac{\Delta \varepsilon}{\varepsilon^2} \quad \text{and} \quad a_2 = b_4 + b_5 \frac{1 - \varepsilon}{\varepsilon} + b_6 \frac{\Delta \varepsilon}{\varepsilon^2}.$$

Similar to the use of piecewise linear functions to approximate nonlinear functions, in the operational algorithm, for each atmospheric WV, averaged emissivity $\varepsilon$, and $T_o$ or LST are divided into several tractable sub-ranges to improve the LST retrieval accuracy over a wide range of surface and atmospheric conditions. The WV is divided into sub-ranges up to 6.5 g/cm$^2$ with an overlap of 0.5 g/cm$^2$. The $\varepsilon$ is also separated into two groups, one varying from 0.90 to 0.96 and the other one varying from 0.94 to 1.0. The $T_o$ sub-ranges are divided by 273, 281, 289, 295, 300, 305, and 310 K. The LST varies within $T_o \pm 16$ K and the LST range of 32 K may be divided into four overlapping sub-ranges (Wan & Dozier, 1996). Instead of dividing $T_o$ into several sub-ranges, Tang et al. (2008) proposed to divide LST into several sub-ranges that overlap by 5 K, e.g., $\leq 280$ K, 275–295 K, 290–310 K, 305–325 K and $\geq 320$ K. For a given VZA and given sub-ranges of $\varepsilon$, atmospheric WV, and $T_o$ or LST, the coefficients $b_k$ ($k = 0\text{–}6$) are determined by minimizing Eq. (11) using radiative transfer (RT) simulated $T_i$ and $T_o$ data in ranges wide enough to cover the variations of surface and atmospheric conditions and are then saved in a set of multi-dimensional lookup tables (LUT). The coefficients $b_k$ can be linearly interpolated using the cosine of the VZA. Accordingly, the coefficients $b_k$ in the GSW algorithm vary with the LSEs, the VZA, the atmospheric WV and $T_o$ or LST. In practice, the LST is estimated in two steps. For Wan and Dozier (1996)’s algorithm, the approximate LST is first estimated with coefficients $b_k$ that cover the entire LST range of $T_o \pm 16$ K in a suitable WV sub-range, and then a more accurate LST is obtained by the coefficients $b_k$ determined by the difference between the approximate LST and $T_o$. For Tang et al. (2008)’s algorithm, the approximate LST is first estimated with coefficients $b_k$ that cover the entire LST range from 240 K to 330 K in a suitable WV sub-range, and then a more accurate LST is determined in terms of the coefficients $b_k$ for the LST sub-range in which the approximate LST is found.

Wan (1999) compared the viewing-angle-dependent LST algorithm to the viewing-angle-independent algorithm and pointed out that the latter has an LST error of up to 1.6 K if there is an uncertainty of 0.01 in the value of $\Delta \varepsilon/\varepsilon^2$ in the GSW algorithm. The major improvements of the GSW algorithm, as described by Wan et al. (2002, 2004), incorporated in the LST retrieval include: (1) VZA dependence, (2) WV dependence, and (3) $T_o$ or LST dependence. Validation results showed that an accuracy of LST retrieval better than 1 K in the range 263–322 K can be obtained using the GSW algorithm given in Eq. (11) for certain land covers in clear-sky conditions. One of the advantages of the GSW algorithm that should be highlighted is that LST retrieval is less sensitive to the uncertainties in the LSEs and the atmospheric conditions. Consequently, several SW formulas similar to the GSW algorithm have recently been developed to estimate LSTs from different satellite instruments, such as the Spinning Enhanced Visible and Infrared Imager (SEVIRI) and FengYun Meteorological satellite (FY-2C) instruments (Atitar & Sobrino, 2009; Jiang & Li, 2008b; Tang et al., 2008). However, the errors in the GSW retrieved LSTs may be larger in bare soil sites in cases where LSTs are larger than $T_o$ by more than 16 K (Wan & Li, 2008). A set of new coefficients developed for bare soil lands based on RT simulation data in a much wide LST range has been used in the new version of the
where $c_k$ ($k=0$–$3$) are coefficients pre-determined by regressing Eq. (12) on simulated satellite data with a set of atmospheres and surface parameters, similar to the $a_k$ in Eq. (7).

Many similar forms of non-linear SW algorithms have been developed in the literature in recent decades (Atitar & Sobrino, 2009; Coll & Caselles, 1997; François & Ottlé, 1996; Sobrino & Raissouni, 2000; Sobrino et al., 1994). Similar to the linear SW algorithms, some of these non-linear SW algorithms incorporate the LSEs into the coefficients $c_k$, some use both the LSE and the WV, and some also incorporate the VZA.

1. With consideration of LSEs

To account for the effect of the LSEs on LST retrieval, Sobrino et al. (1994), Coll and Caselles (1997) proposed a non-linear SW algorithm with a form similar to Eq. (12) with $c_0$ formulated as a linear function of $e$ and $\Delta e$:

$$
c_0 = D_0 + D_1(1-e) + D_2\Delta e, \quad (13)
$$

where the coefficients $D_k$ ($k=0$–$2$) are constants that are independent of the atmosphere. Sun and Pinker (2003) also proposed a non-linear SW algorithm in which the LSEs are implicitly considered by making each parameter $c_k$ in Eq. (12) dependent on the land surface type.

2. With consideration of LSEs and WV

To further improve the accuracy of and reduce the influence of wet atmospheric conditions on LST retrieval, Sobrino and Raissouni (2000) and Sobrino et al. (2004) developed a non-linear SW algorithm to retrieve the LST from the global land 1-km Advanced Very High Resolution Radiometer (AVHRR) data with an estimated error of 1.3 K compared with a set of three hundred LSTs measured in situ in two regions of Australia (Prata, 1994b). They parameterized $D_k$ ($k = 1, 2$) in $c_0$ (Eq. (13)) as linear functions of WV, i.e.,

$$
D_k = E_{0,k} + E_{1,k}WV \quad (k = 1, 2), \quad (14)
$$

in which the coefficients $E$ are sensor-dependent constants.

3. With consideration of LSEs, WV, and VZA

To improve the accuracy of LST retrieval from TIR data measured at larger VZA, the VZA has to be considered when developing the LST retrieval algorithm. Sobrino and Romaguera (2004) and Atitar and Sobrino (2009) proposed a physics-based non-linear SW algorithm for SEVIRI data in two TIR channels. They set $c_1 = 1$ and parameterized the coefficients $c_k$ ($k=2$, $3$) in Eq. (12) and $E$ in Eq. (14) as linear functions of the square of the secant of the VZA. It has been shown that this type of SW is capable of obtaining LST values with a root mean squared error (RMSE) of 1.3 K using SEVIRI data at VZAs lower than 50°.

3.1.2.3. Linear or non-linear multi-channel algorithms

When there are three or more TIR channels available, the LST can be retrieved from a linear or non-linear combination of the TOA brightness temperatures in those channels using methods similar to the SW algorithms described above (Sun & Pinker, 2003, 2005, 2007). For instance, Sun and Pinker (2003) developed a three-channel linear algorithm to retrieve nighttime LSTs from the Geostationary Operational Environmental Satellite (GOES) data, assuming that the LSEs in these three channels can be estimated from the land surface types. They proposed to use the characteristics of the MIR channel $i_1$ at 3.9 µm to improve the atmospheric correction at night, and the coefficients in the three-channel linear equation explicitly include the channel LSEs, but neglect the WV and the VZA:

$$
LST = d_0 + d_1 + d_2 \frac{1-e_i}{e_i} T_{i_1} + d_3 + d_4 \frac{1-e_j}{e_j} T_{i_j} + d_5 + d_6 \frac{1-e_k}{e_k} T_{i_k}, \quad (15)
$$

where the coefficients $d_k$ ($k=0$–$6$) are constants, independent of atmosphere and VZA, Comparison with some published SW algorithms (Becker & Li, 1990b; Wan & Dozier, 1996) demonstrated that the proposed three-channel algorithm obtained the best LST values, with an RMSE less than 1 K (Sun & Pinker, 2003). Furthermore, Sun and Pinker (2007) also proposed a four-channel non-linear algorithm to retrieve the LST from SEVIRI data, with the coefficients depending on land surface types to account for LSE effects. For LST retrieval at night-time, Eq. (16) is used

$$
LST = e_0 + e_1 T_{i_1} + e_2 (T_{i_1} - T_j) + e_3 (T_{i_2} - T_j) + e_4 (T_{i_3} - T_j) + e_5 (sec\text{VZA}-1), \quad (16)
$$

where the subscript $i$ represents the TIR channel at 8.7 µm and the coefficients $e_i$ ($k=0$–$5$) are dependent on the land surface type. To account for the solar radiation reflected by the Earth’s surface in the MIR channel $i_1$ during the daytime, a solar correction term $d_5 T_{ij} \text{cos} \theta_j$ was added to Eq. (16) (Sun & Pinker, 2007) or a solar correction must be performed to the $T_{ij}$ using the method proposed by Adams et al. (1989) and Mushkin et al. (2005). When evaluated against ground observations, the results showed that the LSTs retrieved using the four-channel algorithm were more accurate than those obtained using the GSW algorithm.

However, it should be noted that MIR data measured at the TOA during the daytime consists of a combination of reflected solar radiation and emitted radiation from both the surface and the atmosphere, and the error caused by the solar correction term can affect the accuracy of LST retrieval especially in arid and semi-arid regions with high reflectance in the MIR. In addition, introducing one more channel comes with the expense of increased measurement errors. The errors related to instrumental noise and other uncertainties, such as in surface emissivity of the 8.7 µm channel in arid regions, also influence the final LST retrieval accuracy. The range of emissivity values and respective uncertainty for natural or man-made surfaces is significantly higher for MIR and 8.7 µm channels than for the most commonly used SW channels (Trigo et al., 2008b), further limiting the wide spread use of those channels for operational purposes.

3.1.3. Multi-angle method

Similar to the rationale of SW method, the multi-angle method is based on differential atmospheric absorption due to the different path-lengths when the same object is observed in a given channel from different viewing angles (Chédin et al., 1982; Li et al., 2001; Prata, 1993, 1994a,b; Sobrino et al., 1996, 2004c).

This method was primarily developed based on the first sensor to operate in biangular-mode, the Along Track Scanning Radiometer (ATSR) onboard the first European Remote Sensing Satellite (ERS-1). ATSR can achieve a dual-angle observation of the same region of the Earth’s surface within about 2 min. One viewing angle is the nadir
with a zenithal angle from 0° to 21.6°, and the other is the forward view with a zenithal angle from 52° to 55°. Assuming that the LST and SST are independent on the VZA and that the atmosphere is horizontally uniform and stable over the observation time, Prata (1993, 1994a) derived a dual-angle method to retrieve the SST and the LST from ATS-R data. Sobrino et al. (1996) later proposed an improved dual-angle algorithm that incorporates the emissivity \( \varepsilon_n \) at nadir and the emissivity \( \varepsilon_f \) at forward view:

\[
\text{LST} = T_n + p_1 (T_n - T_f) + p_2 + p_3 (1 - \varepsilon_n) + p_4 (\varepsilon_n - \varepsilon_f),
\]

where \( p_i (k = 1 \sim 4) \) are coefficients related to the atmospheric transmittances and mean air equivalent temperatures in the nadir and forward views; \( T_n \) and \( T_f \) are the brightness temperatures measured in the nadir and forward views, respectively. This algorithm includes only emissivity dependence and has no explicit WV dependence. A non-linear dual-angle algorithm has been developed by Sobrino et al. (2004c) to reduce the influence of atmospheric WV on the LST retrieval:

\[
\text{LST} = T_n + q_1 (T_n - T_f) + q_2 (T_n - T_f)^2 + (q_3 + q_4 \text{WV})(1 - \varepsilon_n) + (q_5 + q_6 \text{WV}) \Delta A + q_0,
\]

where \( q_i (k = 0 \sim 6) \) are sensor-dependent constants that are pre-determined by the simulation method. Using simulated TIR data, Sobrino and Jiménez-Muñoz (2005) compared the LSTs retrieved using the dual-angle algorithm given in Eq. (18) and the non-linear SW algorithm incorporating LSEs, WV, and VZA described in 3.1.2.2. The results showed that the dual-angle algorithm is superior to the SW algorithm provided that the spectral and angular variations of the LSEs are well known.

However, it should be noted that although the multi-angle (dual-angle) algorithm provides better results than the SW algorithm, the dual-angle algorithm has several practical difficulties when applied to satellite data (Sobrino and Jiménez-Muñoz, 2005). A critical phenomenon in the multi-angle method is the angular dependence of the emissivity, as the angular behavior of natural surfaces such as soils and rocks is not well known at the scales of satellite’s spatial resolution (Sobrino & Cuenca, 1999). The angular dependence of the LST is also an issue. In addition to the requirement that the atmosphere is free of clouds and horizontally uniform, the multi-angular measurements must have significant difference in slant path-lengths. Otherwise, the measurements will be highly correlated, and the algorithm will be unstable and extremely sensitive to measurement noise (Prata, 1993, 1994a). Furthermore, “the same” object observed at different viewing angles can cover different sensor areas (in terms of pixels). Even though the same pixel size may be obtained, an object observed at different observation angles may also appear totally different because of the three-dimensional structure of the object. Finally, mis-registration of pixels under different viewing angles can result in drastic errors in the LST retrieval results. Consequently, multi-angle methods can only be applied to homogeneous areas (for example, the surface of the sea or densely vegetated forest) in ideal atmospheric conditions but not to heterogeneous areas.

3.2. LST retrieval with unknown LSEs

All of the methods mentioned above estimate the LST by assuming that the LSE is known. In practice, the heterogeneity of the surface and the angular and spectral variation of the LSE (Li et al., 2013) make it challenging to exactly determine the LSE at the satellite pixel scale in advance. The emissivity of land, unlike that of oceans, can differ significantly from unity and can vary with vegetation, surface moisture, roughness, and viewing angle (Salisbury & D’Aria, 1992). Therefore, the LSEs measured in the laboratory cannot be arbitrarily used at the pixel scale. Generally, an uncertainty of 1% in the LSE will result in about 0.5 K errors in the LST under normal conditions. To ensure the target of 1 K accuracy in LST retrieval, methods to estimate the LSE from space must also be developed. To date, there are at least three distinct methods to estimate the LST from space when the LSE is not known. The first is a stepwise retrieval method that determines the LSE and the LST separately. The LSE is estimated first, and then the LST is retrieved. The second is a simultaneous retrieval method that treats both the LST and the LSE as unknowns and resolves both of them from the atmospherically corrected radiances or with approximated atmospheric profiles. The third is a further development of the simultaneous retrieval method that simultaneously retrieves the atmospheric profiles (or atmospheric quantities in the RTE) with the LST and LSE.

3.2.1. Stepwise retrieval methods

This type of method retrieves the LST using two consecutive steps. First, the LSE is (semi-) empirically determined from visible/near-infrared (VNIR) measurements or physically estimated from pairs of atmospherically corrected MIR and TIR radiances at ground level. Then, the LST is estimated using any of the single, multi-channel (SW) or multi-angle (dual-angle) retrieval methods described in Section 3.1. Representative stepwise methods include the classification-based emissivity method (Peres & DaCamara, 2005; Snyder et al., 1998; Sun & Pinker, 2003), the normalized difference vegetation index (NDVI)-based emissivity method (Sobrino & Raissouni, 2000; Valor & Caselles, 1996; Van de Griend & Owe, 1993), and the temperature-independent spectral-indices method (Becker & Li, 1990a; Li & Becker, 1993; Li et al., 2000). Aside from the previously discussed advantages and disadvantages of the LST retrieval methods with known LSEs, stepwise retrieval methods present specific characteristics that will be briefly presented in the following sections.

3.2.1.1. Classification-based emissivity method (CBEM). This method assumes that similarly classified land covers exhibited very similar LSEs. The LSE can be assigned from look-up tables based on conventional land cover classification information. Snyder et al. (1998) reported that the LSE values of 70% of land cover can be estimated within sufficient accuracy (about 0.01) using this method by considering seasonal and dynamic state changes, thus meeting the goal of 1 K accuracy of LST retrieval, in theory.

Generally, the CBEM is the simplest method in terms of processing, and it can provide accurate LSEs for LST retrieval as long as the land surfaces are accurately classified and each class has well-known LSEs (Gillespie et al., 1996). Even for data at coarse resolutions, such as geostationary satellite data, the CBEM can be applicable using a linear mixing model (Peres & DaCamara, 2005; Sun & Pinker, 2003; Trigo et al., 2008b). However, the LSE will be less accurate because it is difficult to estimate the weights of each component (endmember) within a coarse pixel.

Once the LSE is obtained, the LST can be estimated directly from the methods described in Section 3.1. The LST retrieval accuracy is determined primarily by the accuracy of the LSEs. As described by Snyder et al. (1998), the accuracy of CBEM can degrade due to uncertainties in the soil moisture, the annual biophysical cycle of vegetation, and the appearance of snow and ice. In addition, classifications based on VNIR data are generally not well correlated to the LSE in the TIR channel. For example, estimating the LSE using the CBEM for geologic substrates is uncertain because the VNIR reflectivities used to classify the land surfaces respond mainly to OH- and Fe-oxides in the land surface, while the LSE in the TIR channels are mainly responsive to the Si-O bond (Gillespie et al., 1996). Furthermore, discontinuities in the classification will cause inappropriate discontinuities in the LSE map, which can appear seams or contoured (Gillespie et al., 1996). Finally, the CBEM is most suitable for spectral regions with low spectral contrast LSEs, for example around 11 μm and...
12 μm for most territorial surfaces. Otherwise, a large retrieval error may be caused by variations in the emissivity. All of these uncertainties may prevent an accurate estimation of the LSE from CBEM, thus degrading the accuracy of LST retrieval.

3.2.1.2. NDVI-based emissivity method (NBEM). This method is based on a statistical relationship between the NDVI derived from the VNIR bands and the LSE in the TIR channels. Van de Griend and Owe (1993) first found a very high correlation between the LSE in the TIR channels covering 8–14 μm and the logarithmic NDVI. Subsequently, Valor and Caselles (1996) applied this method to estimate the effective LSE of a rough row-distributed system. Starting from the method proposed by Valor and Caselles (1996), Sobrino and Raissouni (2000) reduced the complexity and formulated an operational NDVI threshold method to derive the LSE from space. This method assumes that: 1) the surface is only composed of soil and vegetation; 2) the emissivity of the bare soil can be linearly represented by the surface reflectivity in the red channel; 3) the LSE changes linearly with respect to the fraction of vegetation in a pixel. Therefore, the LSE of TIR channel i can be estimated using three linear functions corresponding to conditions in which a pixel is composed of full vegetation, full soil or of mixed soil/vegetation content.

Because of its simplicity, this method has already been applied to various sensors with access to VNIR data (Momeni & Saradjian, 2007; Sobrino & Raissouni, 2000; Sobrino et al., 2002, 2004b, 2008, 2009). Similar to the CBEM, an accurate atmospheric correction is unnecessary. However, the NDVI thresholds that indicate bare soil and full vegetation cover, the vegetation fraction, any caviencies, and the LSEs for bare soil and full vegetation must be known in advance. Since NDVI is used as a proxy for the fraction of vegetated surface within the pixel, it can be replaced by more accurate estimates of this variable (e.g., Trigo et al., 2008b). Nevertheless, one of the drawbacks of this method is the lack of continuity in the LSE values of regions transitioning from soil-type to vegetation-type, because the LSEs in such regions are calculated using different formulae (Sobrino et al., 2008). Using numerical analysis, Sobrino et al. (2008) found that this method can only provide acceptable results in the 10–12 μm interval, because the NDVI-LSE relationship for bare soil samples does not provide satisfactory results beyond these spectral intervals. In addition, this relationship may hold for soil and vegetation mixing areas, but surfaces like water, ice, snow and rocks do not provide satisfactory results beyond these spectral intervals. In general, the time-invariant LSE assumption appears to be reasonable in most situations. The LSEs will remain unchanged over several days unless rain and/or snow occur. Therefore, the value of the NDVI-LSE relationship may be worth trying to checking the relative humidity value in the low boundary layer to avoid heavy dew events becoming a serious problem (Wan, 1999). Therefore, this method is superior to the (semi-)empirical stepwise retrieval methods above, especially on bare and geologic substrates that exhibit contrast emissivities.

However, several requirements may limit the usage of this algorithm in LST retrieval from space. First of all, approximate atmospheric corrections and concurrence of both MIR and TIR data are required (Sobrino & Raissouni, 2000). Then, accurate image co-registration must be performed (Dash et al., 2005). Additionally, the surfaces must be observed under similar observation conditions, e.g., viewing angle, during both day and night (Dash et al., 2005).

3.2.2. Simultaneous LST and LSE retrieval methods with known atmospheric information

Because the accuracy of the retrieved LST is primarily dependent on the accuracy of the LSE, simultaneous determination of the LSE and the LST has been proposed to improve the retrieval accuracy. Many simultaneous LST and LSE retrieval methods with known atmospheric information have been developed since the 1990s. These methods can be roughly grouped into two categories: the multi-temporal and multi (hyper)-spectral retrieval methods. The multi-temporal retrieval methods primarily make use of measurements at different times to retrieve the LST and the LSE under the assumption that the LST is time-invariant. The representatives of these methods are the two-temperature method (Watson, 1992) and the physics-based day/night operational method (Watson & Li, 1997). The multi (hyper)-spectral retrieval methods rely on the intrinsic spectral behavior of the LSE rather than temporal information. The representatives of the multi (hyper)-spectral include the gray body emissivity method (Barducci & Pippi, 1996), the temperature emissivity separation (TES) method (Gillespie et al., 1996, 1998), the iterative spectrally smooth temperature emissivity separation method (Borel, 2008), and the linear emissivity constraint method (Wang et al., 2011). Based on some reasonable assumptions or constraints, these methods can retrieve the LST and LSE from the atmospherically corrected radiances at the ground level either by reducing the number of unknowns or by increasing the number of equations.

3.2.2.1. Two-temperature method (TTM). The idea underlying the TTM is the reduction of unknowns through multiple observations. Provided that accurate atmospheric corrections in the TIR channels have been performed and that the LSEs are time-invariant, there are 2 N measurements with N + 2 unknowns (N channel LSEs and two LSTs) if the land surface is observed by N channels at two different times. The N LSEs and the two LSTs can therefore be simultaneously determined from the 2 N equations if N ≥ 2 (Watson, 1992). Note
that the assumption of the time-invariant LSEs implicitly requires the surface to be homogeneous and have relatively stable soil moisture. The first restriction is to alleviate the LSE variation caused by pixel sizes and by viewing angles, while the second is to avoid the LSE changes with soil moisture, such as the occurrence of precipitation and dew.

The primary advantage of the TTM is that it makes no assumption about the spectral shape of the LSEs, only that they are time-invariant. This method has a simple and straightforward formulation; however, the retrieval accuracy is not always guaranteed because the 2 $N$ equations are highly correlated and their solutions may thus be unstable and very sensitive to instrument noise and errors in the atmospheric corrections (Caselles et al., 1997; Gillespie et al., 1996; Watson, 1992). Because accurate atmospheric corrections are difficult to perform without simultaneous atmospheric profile measurements, the use of approximate profiles could lead to large uncertainties in the LST and LSE retrievals. Peres and DaCamara (2004) found that increasing the number of observations and/or the temperature difference improved the retrieval accuracy, but this improvement is limited by the high correlation between TIR measurements.

In addition to the problems mentioned above, this method requires accurate geometric registration of images acquired at two different times (Gillespie et al., 1996; Watson, 1992). Similar to the day/night TISI based method, the impact of mis-registration on the LST and LSE errors is small for homogenous areas but large for heterogeneous areas (Wan, 1999). A change in the satellite VZA causes a change in the LSE, consequently violating the assumption of time-invariant LSEs and decreasing the accuracy of the TTM (Li et al., 2013).

3.2.2.2. Physics-based day/night operational method (D/N). Inspired by the day/night TISI based method and TTM method, Wan and Li (1997) further developed a physics-based D/N method to simultaneously retrieve LST and LSEs from a combined use of the day/night pairs of MIR and TIR data. This method assumes that the LSEs do not significantly change from day to night and that the angular form factor has very small variations (~2%) in the MIR wavelength range of interest to reduce the number of unknowns and make the retrieval more stable. To reduce the effect of the residual error of atmospheric corrections on the retrieval, two variables, the air temperature at the surface level ($T_a$) and the atmospheric WV, are introduced to modify the initial atmospheric profiles in the retrieval. With two measurements (day and night) in $N$ channels, the numbers of unknowns are $N + 7$ ($N$ channel LSEs, 2 LSTs, 2 $T_a$, 2 WV, and 1 angular form factor in the MIR channels). Thus, to make the equations determinant, $N$ must be equal to or greater than seven.

Generally, the physical D/N method is a development of the aforementioned TTM using two time observations. Compared with the TTM and TISI mentioned previously, this D/N method is highlighted by several facets:

1. The contribution of solar irradiation to the radiance of the MIR channels in daytime significantly decreases the correlations among the equations and makes the solution more stable and accurate. Unlike the TISI method that first obtains the bi-directional reflectivity of the pixel and then calculates LSE and LST separately, the D/N method retrieves simultaneously LST and LSE and avoids the propagation of error from stepwise retrieval. In addition, the D/N methods can accurately determine LSTs and LSEs even though the LSTs are equal at the two times (day and night times); while the TTM with only TIR measurements require significant difference in the temperatures.

2. The accuracy of the retrieved LSTs and LSEs is strongly improved by introducing two variables ($T_a$ and WV) to account for the uncertainties in the initial atmospheric profiles. As a result, the accuracy of the atmospheric correction is not required to be as high as that of TISI and TTM.

3. The D/N method does not require 12-hour interval measurements (day and night). As long as the surface emissivity does not change significantly, daytime and night-time data collected over several days are also appropriate.

However, similar to the other multi-temporal methods, the D/N method still suffers from the critical problems of geometry mis-registration and variations in the VZA. Wan (1999) aggregated the MODIS pixels to increase the scale from 1 km to 5 km or 6 km in order to overcome the mis-registration problems. Meanwhile, 16 VZA subranges are used to ensure quality of day and night VZA subranges (Wan & Li, 2010). A set of new refinements (Wan, 2008) were implemented to reject the worst solutions and for better LST retrievals even in less ideal conditions such as under the effects of near-by clouds and heavy aerosols, different surface emissivity values in the MIR and 8.75 μm channels during the day and night due to events of rain, snow and nighttime dew (given the relatively high emissivity values in bands 31 and 32 less affected by these events even in arid regions). The improvements include the combined use of Terra and Aqua MODIS data, setting larger weights on the daytime data, fully incorporating the viewing-angle dependent GSW method into the D/N algorithm as a close component and related constraints on LST differences, using the variables of emissivities in bands 31 and 32, WV and $T_a$ in the iterations of solution of the D/N algorithm and effectively increasing the weights on the highest quality data of bands 31 and 32. More details on the MODIS D/N method can be found in the literature (Wan, 2008; Wan & Li, 1997, 2010).

3.2.2.3. Gray body emissivity method (GBE). This method assumes that the LSE has a flat spectrum, i.e., the LSE is independent of the wavelength, for wavelengths larger than 10 μm to reduce the number of unknowns and stabilize the retrieval algorithm (Barducci & Pippi, 1996).

The main advantage of the GBE is that it has no additional assumption about the shape of the emissivity spectrum is required, except the assumption that it is flat in some wavelength interval. In theory, the LST and LSEs can be simultaneously retrieved as long as at least two channels have the same LSE (not necessarily the gray body) in the wavelength interval of interest. However, the limitations of the method are evident. The application of the GBE method to space-based measurements requires accurate atmospheric corrections in the TIR channels and at least two channels with the same LSE. Similar to the TTM, this method is very sensitive to instrument noise and errors in the atmospheric corrections because the TIR measurements are highly correlated. Moreover, requiring spectrally flat LSEs often hampers the use of the GBE in multispectral TIR data unless at least two channels with the same LSE can be identified. This problem can be more or less overcome with hyperspectral TIR data, because it is easier to find at least two channels with the same LSE in hyperspectral data than in multispectral data, and hundreds or even thousands of channels can further improve the retrieval accuracy. Therefore, the GBE method is thought to be more applicable to hyperspectral TIR data.

3.2.2.4. Temperature emissivity separation method (TES). This method was first developed by Gillespie et al. (1996) to separate the LST and the LSE using atmospherically corrected Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) TIR data. This method relies on an empirical relationship between the spectral contrast and the minimum emissivity to increase the number of equations (equivalent to reducing the number of unknowns) so that the underdetermined retrieval problem becomes deterministic. The TES method comprises three mature modules: the normalization emissivity method (NEM) (Gillespie, 1995), the spectral ratio (SR), and the maximum–minimum apparent emissivity difference method (MMD) (Matsunaga, 1994).
The NEM module is first used to estimate the initial LST and the normalized emissivities from the atmospherically corrected radiances at ground level (Gillespie et al., 1996). Subsequently, the SR module is employed to calculate the ratio of the normalized emissivities to their average. Although the SR module cannot directly obtain the actual LSE, it has been demonstrated to describe the shape of the emissivity spectrum well even if the surface temperature is roughly estimated by the NEM module. Finally, on the basis of the results of the SR module, the MMD module is utilized to find the spectral contrast (i.e., the MMD) in N channels, then to estimate the minimum LSE using the empirical relationship between the minimum LSE (LSE\text{min}) in N channels and the MMD. Once LSE\text{min} is estimated, the LSEs in the other channels can be straightforwardly derived from the SR, and then the LST can be refined and estimated (Gillespie et al., 1998).

The main advantage of the TES is that it combines attractive features of three precursors and uses an empirical relationship between the range of emissivities and the minimum emissivity in the N channels to retrieve the LST and LSEs. Consequently, it can be applied to any kind of natural surface without considering spectral variations in the emissivity, especially for surfaces with high spectral contrast emissivities such as rocks and soils (Gillespie et al., 1998; Sobrino et al., 2008). Numerical simulation and some field validations have demonstrated that the TES can retrieve the LST to within about ±1.5 K and the LSEs to within about ±0.015 when the atmospheric effects are accurately corrected (Gillespie et al., 1996, 1998; Sawabe et al., 2003). Besides, Hulley and Hook (2011) recently refined the relationship between LSE\text{min} and MMD to make TES algorithm available for MODIS’s three TIR channels (29, 31 and 32).

However, some reports have indicated that the TES method exhibited significant errors in the LST and LSEs of surfaces with low spectral contrast emissivity (e.g., water, snow, vegetation) and under hot and wet atmospheric conditions (Coll et al., 2007; Gillespie et al., 1996, 2011; Hulley & Hook, 2009b, 2011; Sawabe et al., 2003). Sabol et al. (2009) pointed out that the low emissivity contrast and high emissivity contrast have been treated differently in original version of TES. Consequently, the retrieved LSEs are too low and the LST is too high in the original version for the materials (such as soils, vegetation and water/snow) that are plotted above the regression line in the scatter plot of LSE\text{min} and MMD. That is why some studies have reported that inaccurate atmospheric corrections may produce LST errors of 2–4 K for bare soil (Dash et al., 2002). For warm and wet atmospheric conditions, the cause of significant errors is different. The uncertainties in the atmospheric corrections will result in a large apparent emissivity contrast. This effect is more serious over graybody surfaces (Hulley & Hook, 2011). To minimize atmospheric correction errors, Gillespie et al. (2011) improved the TES method by using a water vapor scaling (WVS) approach proposed by Tonooka (2005).

As shown by numerical simulations, the uncertainties on the LST and LSE retrievals increase when the number of channels is reduced, making the TES method inapplicable to most operational sensors (Sobrino et al., 2008). Moreover, sensor calibration errors and noise in the TIR channels also cause uncertainties in the retrieved LST and LSEs (Gillespie et al., 2011; Jiménez-Muñoz et al., 2006; Sobrino et al., 2008). In addition, TES scales low- and high-contrast surfaces differently, which leads to step discontinuities at the edges of graybody units such as water, forests, and crops (Sobrino et al., 2007). To overcome these problems, Sabol et al. (2009) recently replaced the power relationship of LSE\text{min} and MMD in the original TES method with a linear expression, and applied the new relationship available for all materials to alleviate such discontinuities. This revision was reported to reduce slightly the accuracy for both rock surfaces and graybodies but can improve the precision for near-graybody surfaces.

3.2.2.5. Iterative spectrally smooth temperature emissivity separation method (ISSTES). Hyperspectral TIR data provides much more detailed spectral information about the atmosphere and land surface. Borel (1997, 1998, 2008) reported that a typical emissivity spectrum is rather smooth compared with the spectral features introduced by the atmosphere. According to the RTE given in Eq. (4), if the LST is not accurately estimated, the corresponding LSE spectrum will exhibit the atmospheric spectral features, i.e., there will be sawteeth caused by the atmospheric absorption lines on the estimated LSE spectrum. The best estimates of the LST and LSE should be obtained when the spectral smoothness of the retrieved LSE is maximized. Based on this property, the iterative spectrally smooth temperature emissivity separation method (ISSTES) has been developed to iteratively retrieve the LST and LSEs from hyperspectral TIR data. Various smoothness criteria including the first and second derivative have been proposed (Borel, 2008; Cheng et al., 2010; Kanani et al., 2007; OuYang et al., 2010), though they all lead to the same statistical performance regardless of the details of the smoothness function.

Ingram and Muse (2001) analyzed the method’s sensitivity to smoothness assumptions and measurement noise and found that the retrieval accuracy caused by the assumptions is negligible for typical materials but is dependent on the SNR, i.e., high accuracy can be obtained with high SNR. Similar to most methods presented above, the atmospheric correction needs to be accurately performed, and its impact on the retrieval results is the greatest among all influences. The retrieval accuracy is also sensitive to shifts in the central wavelength and bandwidths of the TIR channels (Borel, 2008). In addition, Wang et al. (2011) reported that the occurrence of singular points may lead to difficulties in finding an acceptable solution when the LST is close to the effective temperature of the downward atmospheric radiance.

3.2.2.6. Linear emissivity constraint temperature emissivity separation method (LECTES). Inspired by the GBE initially proposed by Barducci and Pippi (1996), Wang et al. (2011) proposed a new TES method to retrieve simultaneously both LST and LSEs from atmospherically corrected hyperspectral TIR data. This method assumes that the emissivity spectrum can be divided into M segments and that the emissivity in each segment varies linearly with the wavelength. Thus, the emissivity spectrum can be reconstructed using a piecewise linear function with gains \(a_k\), and offsets \(b_k\) (\(k = 1, \ldots, M\)), and the LST and LSEs can be simultaneously obtained provided that the number of equations \(N\) corresponding to the N channel measurements is equal to or greater than the number of unknowns (2 \(M + 1\) corresponding to 1 LST, \(M a_k\), and \(M b_k\)). The requirement of \(N \geq 2M + 1\) is easily fulfilled for hyperspectral TIR data because a huge number of narrow channels are available in a hyperspectral TIR sensor.

Wang et al. (2011) carried out a series of sensitivity analyses and concluded that the errors introduced by the assumption of linear emissivity can be neglected if the segment length is well chosen. A segment length of about 10 cm\(^{-1}\) is suggested. Compared with ISSTES, this method produces less frequent singular points and is more resistant to both white noise and uncertainty in the downward atmospheric radiance. Because atmospheric spectral features are more significant under wet and warm atmospheric conditions, the LECTES method performs better in wet and warm atmospheres than in dry and cold atmospheres. Similar to the ISSTES, this method is only suitable for hyperspectral TIR data and requires accurate atmospheric corrections. However, because the number of unknowns can be greatly reduced with a piecewise linear function, this method exhibits great potential as a technique for simultaneously retrieving the LST, the LSE, and atmospheric profiles, as will be described in 3.2.3.2 and 5.1. Recently, Paul et al. (2012) developed a methodology for the simultaneous retrieval of LST and emissivity spectra from the Infrared Atmospheric Sounding Interferometer (IASI) hyperspectral data. In this case, the LST and LSE retrieval make use of a first guess for land emissivity, which is estimated from six MODIS channels interpolated to the IASI hyperspectral range using a non-linear statistical (neural network) scheme.
<table>
<thead>
<tr>
<th>Methods</th>
<th>Assumptions</th>
<th>Advantages</th>
<th>Limitations and disadvantages</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-channel algorithms</td>
<td>Different atmospheric absorptions in adjacent TIR channels</td>
<td>1) Simplicity and high efficiency</td>
<td>1) Require a priori knowledge of the emissivity in each TIR channel 2) Simplicity and high efficiency 3) Accurate atmospheric correction is not required 4) Approximate atmospheric corrections are sufficient 5) Solution more sensitive to instrument noise and errors in atmospheric corrections</td>
<td>McMillin (1975) Sobrino et al. (1994), Wan and Dozier (1996)</td>
</tr>
<tr>
<td>Multi-angle algorithms</td>
<td>1) LSTs are independent of the VZA 2) The atmosphere is horizontally uniform and stable over the observation time</td>
<td>1) Simplicity and high efficiency</td>
<td>1) Require a priori knowledge of the emissivity of the angular variation of the emissivity at satellite pixel scale 2) Require a significant difference in the slant path-lengths 3) Require accurate geometric registration 4) Applicable only to the homogeneous surfaces</td>
<td>Prata (1993) Sobrino et al. (1996) Sòria and Sobrino (2007)</td>
</tr>
<tr>
<td>Retrieval with unknown emissivity</td>
<td>Classification-based emissivity method (CBEM)</td>
<td>1) Simplicity</td>
<td>1) Require a priori knowledge of the emissivity of each class, as well as the corresponding classification map 2) Seasonal and dynamic states of surfaces may degrade the accuracy 3) Less accurate at coarse resolutions and less reliable for classes with contrast emissivities. 4) U</td>
<td>Peres and DaCamara (2005) Snyder et al. (1998)</td>
</tr>
<tr>
<td>NDVI-based emissivity methods (NBEM)</td>
<td>1) Surface is composed of soil and vegetation 2) Variation of LSE is linearly dependent on the fraction of vegetation in a pixel</td>
<td>1) Simplicity</td>
<td>1) Require approximate atmospheric corrections at different times 2) At least two channels, one in MIR and another in TIR atmospheric windows, are available 3) Require accurate geometric registration 4) Observations must be conducted at similar viewing angles</td>
<td>Valor and Caselles (1996) Van de Griend and Owe (1993) Sobrino and Raissouni (2000)</td>
</tr>
<tr>
<td>Day/night temperature independent spectral indices (TISI) based methods</td>
<td>Emissivity ratios are the same or do not change significantly between two times, i.e., day and night</td>
<td>1) Approximate atmospheric corrections are sufficient 2) Physical basis and suitable for various surfaces</td>
<td>1) Require approximate atmospheric corrections at different times 2) Requires multi-temporal TIR data and large temperature difference between different times 3) Require accurate geometric registration 4) Solution more sensitive to instrument noise and errors in atmospheric corrections 5) Observations must be conducted under similar viewing angles</td>
<td>Jiang et al. (2006) Li and Becker (1993) Li et al. (2000)</td>
</tr>
<tr>
<td>Two-temperature methods (TTM)</td>
<td>Emissivity is invariant at two times</td>
<td>1) More suitable for geostationary satellite data 2) Simultaneously retrieve LST and LSEs</td>
<td>1) Require accurate atmospheric corrections at different times 2) Requires multi-temporal TIR data and large temperature difference between different times 3) Require accurate geometric registration 4) Solution more sensitive to instrument noise and errors in atmospheric corrections 5) Observations must be conducted under similar viewing angles</td>
<td>Peres et al. (2010) Peres and DaCamara (2004) Watson (1992)</td>
</tr>
<tr>
<td>Method</td>
<td>Advantages</td>
<td>Disadvantages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Physics-based day/night operational methods (D/N) | 1) LSEs do not change significantly at day and night times  
2) Angular form factor has very small variation in MIR channels | 1) Require multi-temporal data in several channels in the MIR and TIR atmospheric windows  
2) Require accurate geometric registration  
3) Approximate shapes of the atmospheric profiles need to be given a priori  
4) Retrieval process is complicated, and initial guess values are required  
5) Observations must be conducted at similar viewing angles |
| Gray body emissivity methods (GBE)          | 1) Does not require a priori accurate atmospheric profiles  
2) Solutions become more stable and accurate by introducing the MIR channels  
3) Accuracy of the LSTs and LSEs is greatly improved by modifying the atmospheric profiles in the retrieval  
4) Accurately retrieve both LSTs and LSEs on a physical basis | 1) Require multi-temporal data in several channels in the MIR and TIR atmospheric windows  
2) Require accurate atmospheric corrections  
3) Assumption is not satisfied in most situations  
4) Difficult to find two channels with the same LSE in multi-spectral TIR sensors  
5) Solution more sensitive to instrument noise and errors in the atmospheric corrections |
| Temperature emissivity separation methods (TES) | 1) Multi-temporal data are no longer required  
2) Simultaneously retrieve the LST and LSEs | 1) Require accurate atmospheric corrections  
2) Require at least two channels with the same LSE  
3) Assumption is not satisfied in most situations  
4) Difficult to find two channels with the same LSE in multi-spectral TIR sensors  
5) Solution more sensitive to instrument noise and errors in the atmospheric corrections |
| Iterative spectrally smooth TES methods (ISSTES) | 1) Refines the values of the maximum LSE  
2) Does not require any assumptions about the shape of the LSE  
3) Simultaneously retrieve the LST and LSEs for any kind of surface | 1) Require accurate atmospheric corrections  
2) Require at least three TIR bands within atmospheric windows  
3) Accuracy depends on atmospheric compensation and the empirical relationship between the minimum LSE and the spectral contrast  
4) Uncertainty is more serious for gray bodies  
5) Step discontinuities at the edges of graybody may be caused due to the inaccurate inherent scaling behaviors of method |
| Linear emissivity constraint temperature emissivity separation methods (LECTES) | 1) Emissivity spectrum can be divided into M segments  
2) Emissivity in each segment changes linearly with wavelength | 1) Performance is insensitive to the choice of a smoothness function  
2) High accuracy can be obtained with high SNR  
3) Simultaneously retrieve the LST and LSEs  
4) Reducetion occurrence of singular points and resistance to white noise  
5) Perform well under wet and warm atmospheric conditions  
6) Simultaneously retrieve the LST and LSEs for any kind of surface | 1) Require accurate atmospheric corrections  
2) Only suitable for hyperspectral IR data  
3) Sensitive to spectral shifts and variations in the FWHM  
4) Occurrence of singular points leads to difficulties in finding an acceptable solution  
5) Step discontinuities at the edges of graybody may be caused due to the inaccurate inherent scaling behaviors of method |
| Retrieval with unknown emissivity and unknown atmospheric quantities | 1) Ability to learn from complex patterns  
2) Generalizable to noisy environments  
3) Incorporate knowledge and various physical constraints  
4) Simultaneously retrieve the LST, LSE, and the atmospheric profiles | 1) Highly dependent on the architecture of the ANN and the training data  
2) Difficult to determine the appropriate architectures and learning schemes and representative training data sets  
3) Retrieval process cannot be well controlled  
4) Difficult to interpret the weights assigned to each input and improve the output |
| Two-step physical retrieval methods (TSRM) | 1) Specular surface reflection and a constant angular form factor are used to simplify the RTE  
2) PCA can be used to reduce the number of unknowns without significant loss of accuracy | 1) Complexity  
2) Low computational efficiency limits applications  
3) Require adequate number of channels  
4) Require an initial guess for the LSEs, LST, and atmospheric temperature-humidity  
5) The solution is strongly dependent on the initial guess |
3.2.3. Simultaneous retrieval of LST, LSEs, and atmospheric profiles

Although the simultaneous LST and LSE retrieval methods reviewed above can accurately obtain the LST and LSEs if the atmospheric corrections are performed properly, accurate atmospheric profiles are usually unavailable synchronously with TIR measurements, and thus the accuracy of the retrieved LST and LSEs can be degraded. An ideal solution is to simultaneously retrieve the LST, LSEs, and atmospheric parameters (e.g., atmospheric profiles) (Ma et al., 2002). Because the narrow bandwidth offered by hyperspectral TIR sensors with thousands of channels can improve the vertical resolution and allow atmospheric profiles and surface parameters (LST and LSEs) to be obtained more accurately (Chahine et al., 2001), several methods have been proposed to retrieve simultaneously the surface and atmospheric parameters. The representatives of these methods are the artificial neural network (ANN) method (Wang et al., 2010) and the two-step physical retrieval method (Ma et al., 2002, 2000).

3.2.3.1. Artificial neural network (ANN) method. Because an ANN can robustly perform highly complex, non-linear, parallel computations, ANNs have become increasingly utilized by the remote sensing community (Mas & Flores, 2008). ANNs resemble the brain in two aspects: they acquire knowledge through a learning process, and store the acquired knowledge using interneuron connection strengths (Mas & Flores, 2008). Therefore, ANNs represent massively parallel distributed processors that can acquire experiential knowledge and make that knowledge available for use. The main advantages of ANN methods over conventional retrieval methods are their ability to learn complex patterns, generalization to noisy environments, and incorporation of both knowledge and physical constraints (Mas & Flores, 2008). Because of ANNs’ powerful non-linear retrieval abilities, a number of attempts have been made to develop neural networks to retrieve both the surface and atmospheric biophysical variables without exact knowledge of the complex physics mechanisms. For example, Mao et al. (2008) used an ANN to estimate the LST and LSE, while Aires et al. (2002b) and Blackwell (2005) used an ANN to retrieve atmospheric profiles. To reduce the effect of coupling between the surface and atmosphere on the retrieval accuracy, Aires et al. (2002a) proposed using an ANN to retrieve both the atmospheric and surface temperatures, and Wang et al. (2010) attempted to establish a neural network to simultaneously retrieve the LST, LSE, and atmospheric profiles from hyperspectral TIR data. The preliminary results demonstrated that ANNs can be used to simultaneously retrieve the LST, LSEs and atmospheric profiles from hyperspectral TIR data with acceptable accuracy for some applications. RMSEs of LST and temperature profiles in troposphere are about 1.6 K and 2 K, respectively; RMSE of WV is around 0.3 g/cm². RMSE of LSE is less than 0.01 in the spectral interval from 10 μm to 14 μm (Wang et al. 2013).

However, because ANNs perform like black boxes and can produce corresponding outputs from any given inputs, the retrieval process cannot be well controlled, and it is difficult to interpret the weights assigned to each input and improve the output due to the complex nature of the network. In addition, the implementation of an ANN depends largely on its architecture and the training data (Mas & Flores, 2008). It is difficult to determine the architectures and learning schemes for an ANN, which are directly related to its ability to learn and generalize. Although one or two hidden layers are recognized to be enough for most problems (Aires et al., 2002b; Mas & Flores, 2008; Sontag, 1992), a number of experiments are still required to determine what architecture-related parameters will improve the accuracy, such as the number of input and hidden nodes, the initial weight range, the activation functions, the learning rate and momentum, and the stopping criterion. Until now, no ANN architecture is universally accepted for a particular problem. The characteristics of the training data, such as the size and the representativeness, are also of considerable importance. The use of too few or unrepresentative training samples will result in a network that cannot accurately retrieve the outputs, while the use of too many training samples requires more time for learning. Because physical understanding is not required, ANN methods may be regarded as empirical methods. However, their results can be used to provide initial guesses for further improvements in the physical retrieval methods (Motteler et al., 1995). More detailed information about the application of ANNs can be found in the work of Mas and Flores (2008).

3.2.3.2. Two-step physical retrieval method (TSRM). Because the measured radiance at the TOA is a function of the surface and atmospheric parameters, the surface and atmospheric variables can theoretically be obtained by selecting appropriate channels even from multispectral data. Ma et al. (2000) made an initial attempt at simultaneously retrieving the LST and atmospheric profiles by assuming that the LSE is invariant within the MIR channels and also invariant within the TIR channels and by ignoring the solar contribution in MIR channels. However, these rough assumptions may lead to degraded accuracies in the troposphere. Along this line of reasoning, Ma et al. (2002) further considered the solar contribution and proposed an extended two-step physical retrieval method that simultaneously extracts the LST, the LSE, and the atmospheric profiles from MODIS data. The main idea underlying the TSRM inherits that of atmospheric profile retrieval. The first step is to tangent-linearize the atmospheric RTE with respect to the atmospheric temperature-humidity profiles, the LST, and the LSEs. Given initial guesses for those atmospheric and surface variables, a set of equations based on the tangent-linearized RTE can be derived using the remotely sensed measurements (Li et al., 1994; Ma et al., 1998; Smith, 1972). At the same time, the principle-component-analysis (PCA) technique and the Tikhonov regularization method are employed to reduce the number of unknowns and stabilize the ill-posed problem (Ma et al., 2000; Smith & Woolf, 1976), which makes the solution of these equations stable and deterministic. According to the statistical analysis in the work of Ma et al. (2000, 2002), only five temperature and three water vapor eigenvectors can explain all of the information of 40 atmospheric temperature and water vapor levels, respectively. In the second step, the Newtonian iteration algorithm is utilized with the regularized solution as the initial guess to obtain the final maximum likelihood solution of the atmospheric temperature-humidity profiles, LST, and LSEs.

There are at least three assumptions involved in this method: (1) the RTE can be tangent-linearized around an initial guess; (2) a constant angular form factor is used for the solar beam in the MIR region to simplify the RTE; (3) the PCA can be used to reduce the number of unknowns without significant loss of accuracy. These assumptions ensure the existence of stable and accurate solutions without a priori atmospheric corrections, as opposed to other conventional methods. However, Ma et al. (2002) found that the solutions are highly dependent on the initial guess. Therefore, one possible improvement to this type of method is to improve the initial guess. As pointed out earlier, the results of an ANN can be used as the initial guesses in the physical retrieval method. It is worth noting that the physical nature of the algorithm requires an adequate number of channels in each specific window, and its complex nature may lead to a low computational efficiency. These two properties make the use of this method difficult to apply.

3.3. Comparison and analysis of different methods

There is no universal method capable of always accurately retrieving LSTs from all satellite TIR data because the LST retrieval methods reviewed above were proposed for use under different conditions with different assumptions. It is meaningless to perform a comparison of these algorithms without considering those assumptions. Therefore, it is generally difficult to decide which algorithm is superior to others.
The optimal method to retrieve the LST from space in practice can be selected by considering the characteristics of the sensor, the availability of emissivity data and atmospheric information, the complexity of the method, and other considerations.

Because the SW algorithms are simple, effective and generally suitable for most sensors, many comparative studies evaluating the performance of these methods have been carried out. Vázquez et al. (1997) found that most SW algorithms are statistically indistinguishable. Kerr et al. (2000) performed an algorithm comparison and concluded that the selection of the best LST algorithm may depend on a priori knowledge of the water vapor content and the LSE. Sória and Sobrino (2007) reported that the RMSE of retrieved LST in SW algorithms generally decreases as more inputs parameters are explicitly included. However more inputs parameters will introduce more uncertainties and decrease the accuracy of the LST retrieval. Yu et al. (2008) have evaluated nine published SW algorithms to determine which are most capable of generating a consistent LST climate data record across satellite sensors and platforms. The results showed that the SW algorithms that depend on both the mean and the difference of channel emissivities are the most accurate and stable over a wide range of conditions if the emissivities are accurately known. However, Yu et al. (2009) reported that the use of both the mean and the difference of channel emissivities may be too sensitive to the emissivity uncertainty and should not be used in operational practice. As a compromise, the SW algorithms that only use the mean emissivities are recommended by Yu et al. (2009).

Because all of the assumptions and restrictions involved in the LST retrieval methods, such as the required number of TIR channels and the knowledge regarding emissivities or atmospheric quantities, cannot be met simultaneously, comparisons are seldom made except for SW algorithms. To provide a concise overview, the assumptions, advantages, and limitations of each of these methods are summarized in Table 1 to help users select the optimal method in practice.

4. Validation of satellite derived LST

Validation is a process of independently assessing the uncertainty of the data derived from the system outputs. Without validation, no methods, algorithms, or parameters derived from remotely sensed data can be used with confidence. As the retrieved LSTs from satellite TIR data involve corrections to the satellite-observed radiances to account for atmospheric effects and non-unity LSEs, it is necessary to assess the accuracy of the retrieval to provide potential LST users with reliable information regarding the quality of the LST product and to provide feedback to the developers of LST retrieval algorithms for future improvement. Although many algorithms have been proposed and developed over recent decades to retrieve the LST from satellite TIR data, far fewer studies have been undertaken to validate the satellite-derived LSTs due to the difficulty of making ground measurements of the LST that are representative at the satellite pixel scale and also due to the large spatio-temporal variations in the LST itself. In recent years, several studies have been performed to validate the LSTs derived from a variety of sensors, from largely homogenous targets. Sensor validated include TM/ETM+(Thermatic Mapper/Enhanced Thematic Mapper Plus), ASTER, AVHRR, AATSR (Advanced Along-Track Scanning Radiometer), MODIS and SEVIRI data (Coll et al., 2005, 2010, 2012b; Hook et al., 2005, 2003, 2007; Huiley & Hook, 2009a; Niclòs et al., 2011; Prata, 1994b; Sabol et al., 2009; Sawabe et al., 2003; Sobrino et al., 2007; Sória & Sobrino, 2007; Trigo et al., 2008a,b; Wan, 2008; Wan & Li, 2008; Wan et al., 2002, 2004; Wang & Liang, 2009). Three methods are generally utilized to validate LST values retrieved from space: the temperature-based method (T-based), the radiance-based method (R-based), and cross-validation. The following sections will introduce each validation method, discuss their advantages and disadvantages, and show that they should be regarded as complementary strategies for the assessment of satellite-based LST products.

4.1. Temperature-based method (T-based)

The T-based is a ground-based method that directly compares the satellite-derived LST with in situ LST measurements at the satellite overpass (Coll et al., 2005; Pinker et al., 2009; Prata, 1994b; Slater et al., 1996; Wan et al., 2002). However, performing LST measurements in the field is a complex and difficult task due to the difference of the scales probed by satellite pixels (a few km²), and field sensors (a few m² or cm²). Moreover, natural land cover and the corresponding LST and LSE values are quite variable at the scale of km². Snyder et al. (1997) pointed out that homogeneous and flat surfaces that can be easily instrumented and characterized, including inland water, sand, snow, and ice, can serve as validation sites (Coll et al., 2005; Guilevici et al., 2012; Sobrino et al., 2004c; Wan, 2008). The size of the area that needs to be viewed by the validation instrument depends on the inter-pixel variability of the surface and on how well measurements of several “endmembers” can be combined to obtain a representative value for the satellite pixel. This process remains challenging due to the difficulties inherent to finding adequate surfaces in the image and performing a representative thermal sampling on the ground.

Because most of the Earth’s surface is heterogeneous at the satellite pixel scale, high-quality ground LST validation data are scarce and are limited to a few homogeneous surface types such as lakes, silt playas, grasslands, and agricultural fields collected during dedicated campaigns (Coll et al., 2005, 2010, 2009; Wan et al., 2002, 2004). Once a thermally homogeneous area is identified, the average LST measured at several points by ground instruments within the validation site are considered to be the true LST and are compared with the satellite-derived LST at the pixel scale. Using this method, many authors have performed validation studies of the LST values produced using different sensors (Coll et al., 2005, 2010, 2009; Peres et al., 2008; Sabol et al., 2009; Wan, 2008; Wan et al., 2002).

The main advantage of the T-based method is that it provides a direct evaluation of the radiometric quality of the satellite sensor and the ability of the LST retrieval algorithm to correct for atmospheric and emissivity effects. However, the success of T-based validations depends crucially on the accuracy of the ground LST measurements and how well they represent the LST at the satellite pixel scale. Because the spatial and temporal variations of the LST at daytime might be 10 K or more over a few meters or over short time intervals depending on the nature of the surface, the solar irradiation level, and the local meteorological conditions, T-based validation activities are often restricted to nighttime and homogeneous surfaces such as lakes, dense grasslands, and vegetated regions. Moreover, even if ground-level LST measurements can be performed, there is still a problematic difficulty in scaling up from the ground point measurements to the pixel scale under the field of view of the satellite sensor, especially over heterogeneous surfaces (Wan et al., 2002). As a result, only a few surface types are suitable for T-based validation within an uncertainty of ±1 K for the ground measured LST at the pixel scale. The collection of in situ measurements is also a demanding task and is often limited to short-term, dedicated field campaigns. Therefore, the T-based method is not appropriate for global validation of satellite-derived LST measurements.

4.2. Radiance-based method (R-based)

The radiance-based method (R-based) is an advanced alternative method for validating space-based LST measurements (Coll et al., 2012b; Wan & Li, 2008). This method does not rely on ground-measured LST values but does require both LSE spectra, which can be measured in the field or estimated from land-cover types or from
other auxiliary data, and measured atmospheric profiles over the validation site at the time of the satellite overpass (Wan, 2008; Wan & Li, 2008). This method uses the satellite-derived LST and the aforementioned in situ atmospheric profiles and LSEs as initial input parameters to an atmospheric RTM that simulates the TOA radiance at the moment of the satellite overpass. Using the difference between the simulated TOA radiance and the measured radiance, the initial LST will be adjusted and the simulated radiance will be iteratively recalculated to match the satellite-measured radiance. The difference between the adjusted LST and the initial satellite-derived LST is the accuracy of the retrieved LST. More details about the R-based method are provided by Wan and Li (2008).

The R-based method does not require ground LST measurements, and it can therefore be applied to the surfaces on which ground LST measurements are unfeasible and extended to homogeneous and non-isothermal surfaces. The promising performance of the R-based method offers the possibility of validating satellite-derived LST values during the daytime and nighttime over homogeneous and non-isothermal surfaces. However, the strongest limitations of the R-based method are the use of measured or estimated LSEs representative at pixel scale, how to check the actual atmosphere really free of clouds, and how well the profiles used in simulations represent the actual atmosphere at the time of observations (Coll et al., 2012b). The success of the R-based method depends on the accuracies of the atmospheric RTM, the atmospheric profiles, and the LSEs at pixel scale.

4.3. Cross validation method

This method involves cross-validating the LST values retrieved by the method under test with well documented and validated LST values retrieved from other satellite data (Trigo et al., 2008a). This technique represents an alternative method for LST validation if there are no atmospheric profiles or ground LST measurements available or if the T- and R-based validations cannot be conducted.

The cross-validation method uses a well validated LST product as a reference and compares the satellite-derived LST to be validated with the referenced (well validated) LST derived from other satellites. Due to the large spatial and temporal variations in the LST, geographic coordinate matching, temporal matching, and VZA matching have to be performed before the two satellite-derived LST products can be compared (Qian et al., 2013; Trigo et al., 2008a). The main advantage of this method is that the LST can be validated without any ground measurements, and it can be used anywhere in the world if well validated LST products are available. As mentioned above, the accuracy of this method is sensitive to spatial and temporal mismatches of the two LST measurements. The observation time interval between the two measurements should be as short as possible. Considering that the LSE also depends on the viewing zenith angle and that the pixels of the two sensors cover different areas and contain different land surface information under different viewing angles, only pixels with the same or nearly same viewing zenith angles should be used for cross-validation.

5. Future development and perspectives

Accurately acquiring LSTs at the global scale is crucial to many fields of study including the Earth’s surface water and energy balances, material and energy exchange in terrestrial ecosystems and global climate change. Various methods have been developed to retrieve the LST from multispectral or multi-angular TIR data. Because of the limited spectral information provided in multispectral data, all of these methods rely on different approximations to the RTE and on different assumptions and constraints to solve the inherently ill-posed retrieval problem. Those approximations, assumptions, and constraints might not hold true under certain circumstances. Therefore, users must choose the optimal approach to estimate the LST from space by considering the sensor characteristics, the required accuracy, the computational time, the availability of atmospheric temperature and water vapor profiles, and the LSEs. Considering the significant progress made in recent decades in LST estimation from multispectral TIR data, there will be no significant further progress in LST retrieval from multispectral satellite data if there are no innovations in the acquisition of remotely sensed data. To overcome the shortage of multispectral data and to radically improve the accuracy of LST retrieval from space, it is necessary to explore new ideas and break new paths in remote sensing.

Undoubtedly, hyperspectral TIR sensors with thousands of channels are better able to extract atmospheric and land surface parameters than multispectral TIR sensors. A huge number of channels with narrow bandwidths can improve the vertical resolution of atmospheric soundings (Chahine et al., 2001) and extract the atmospheric quantities used in atmospheric corrections. The hyperspectral TIR data measured within the atmospheric window can provide more detailed land surface information, particularly the LSE spectrum rather than the discrete LSEs in multispectral data, as well as more reasonable assumptions or constraints used to radically separate the LST and the LSEs. These reasons have driven the development of quantitative remote sensing and other related disciplines. The exploration of hyperspectral TIR data for LST/LSE separation and the retrieval of atmospheric profiles or atmospheric quantities involved in atmospheric corrections will become one of the hotspots in quantitative remote sensing.

Progress can also be expected in the development of new methods for extracting the LST from a combination of multispectral and multi-temporal TIR data acquired from the multispectral sensors onboard the new generation of geostationary satellites, such as SEVIRI, GOES and the FY-2 series, which can provide diurnal coverage data and can scan the surface at least hourly with a fixed VZA. Except for the TTM, day/night TSI, and physics-based D/N methods in which data measured at two different times (one in daytime and the other in nighttime) are used, all of the methods developed to retrieve the LST from space are based on multispectral data but do not consider temporal information. It is therefore very attractive to utilize the multi-temporal information to derive the LST from multispectral, multi-temporal TIR data.

In addition, most of the current available LST methods retrieve the LST instantaneously from multispectral data acquired by polar-orbit satellites under clear-sky conditions. There are no long-term LST products derived under all weather conditions. Considering the complementarity of passive microwave and TIR data, a physics-based model for retrieving LSTs from passive microwave data and an effective model of combining LSTs retrieved from TIR and passive microwave satellite data must be developed in the future to produce LSTs with high spatial resolution under all weather conditions. Moreover, because of the intrinsic scanning property of the sensors onboard the polar-orbit satellites, LSTs retrieved at a given location from data acquired by the same polar-orbit satellite on different days or LSTs retrieved at different locations in the same day correspond to different local solar times of observation and different VZAs, let alone LSTs retrieved from different polar-orbit satellites. As the LST varies with both time and VZA, there is no comparability among LSTs of one pixel retrieved on different days or LSTs of different pixels on the same day, which significantly limits the applications of the LST products. To address these issues, a series of LST models, including angular normalization and temporal normalization, must be developed to produce a long-term, time- and angle-normalized consistent LST product under all weather conditions.

Future studies ought to focus on the following subjects to improve LST estimation from space-based measurements.

5.1. Methodology to simultaneously derive LST, LSE, and atmospheric profiles (atmospheric quantities) from hyperspectral TIR data

As stated by Li et al. (2013), the coupling of the surface-emitted radiance and the atmospheric absorption, diffusion and emission
complicates the separate retrieval of surface parameters (LST and LSEs) and atmospheric profiles. The determination of surface parameters from space requires knowledge of the atmospheric profiles and vice versa. It is therefore natural though challenging to develop a method that simultaneously retrieves the LST, LSEs, and atmospheric profiles (or atmospheric quantities used in the atmospheric corrections) without any a priori knowledge about the surface or atmosphere. Ma et al. (2000, 2002) made a first attempt at retrieving those parameters from multispectral TIR measurements. With the appearance of hyperspectral TIR sensors, the thousands of narrow bandwidth channels in TIR can supply enough vertical resolution to allow extraction of atmospheric information and can also provide more physical constraints to accurately separate the LST and the LSEs. Although a few studies have been conducted in recent years (Li et al., 2007; Wang et al., 2013), there are still at least two aspects that require increased attention in the future. First, rapid and accurate RTE models must be developed to meet the requirements of accuracy and speed in the retrieval process. Second, ANNs and physical retrieval methods should also be modified or developed to improve the retrieval accuracies. For example, more details should be considered in the ANNs, including the architectures and learning schemes, selection of representative training data, and the channels employed. At the same time, additional constraints, such as the linear emissivity constraint proposed by Wang et al. (2011), mathematical approaches and regularizations should be introduced into physically based retrieval methods to reduce the uncertainties related to the assumptions and stabilize the solutions. Combining ANNs and physics-based methods also represents an option in the near future, because the advantages of these two techniques can complement each other: ANNs can provide initial guesses for the LST, LSEs, and atmospheric profiles (or atmospheric quantities), and then physical retrieval methods can further improve these initial guesses.

5.2. Methodology to simultaneously derive LST and LSE from the new generation of geostationary satellites with multispectral and multi-temporal data

The new geostationary satellites are prevailing over the polar-orbit satellites in investigating the temporal evolution of land surface and atmospheric information because they provide high-frequency observations at fixed viewing angles over the same surface despite their coarser spatial resolutions. Efforts have focused on retrieving the LST from multispectral data but without considering multi-temporal information. It is therefore very attractive to develop a new method to simultaneously retrieve the LST and LSE by taking advantage of the multispectral and multi-temporal information provided by the geostationary satellites. With the geostationary satellite data, time- and angle-consistent LSTs can be directly produced using these new LST retrieval methods without needing to temporally or angularly normalize the LST.

5.3. Refinement of LST retrieval algorithms with the consideration of aerosol and cirrus effects

Atmospheric correction is one of the most important issues in the LST retrieval algorithms, and errors in atmospheric correction directly decrease the accuracy of the final derived LST. Because of the high transmittance of aerosol in the TIR channel (approximately 0.95–0.98 in MODIS TIR channels) (Wan, 1999) under normal clear-sky conditions and the lack of real-time aerosol estimates (aerosol loading, size distributions, types, and scattering phase functions), an average aerosol distribution and a constant aerosol loading have been used in the development of all of the LST retrieval algorithms reviewed in Section 3. The effect of aerosol on LST retrieval is relatively small compared with the effect of water vapor, but it cannot be ignored when aiming for highly accurate LSTs for use in certain special applications, especially in the presence of heavy aerosol loadings (Jiménez-Muñoz & Sobrino, 2006). To improve the accuracy of LST retrieval, existing LST retrieval algorithms must be refined, or new algorithms must be developed to correct for the aerosol effect, particularly in the case of heavy aerosol loading.

In addition, the effect of cirrus clouds on LST retrieval should also be considered. Cirrus clouds are always considered to be cloud contamination in many LST retrieval algorithms and the pixels covered by cirrus clouds are screened out in data preprocessing. Because thermal infrared wavelengths can penetrate cirrus layers, it is possible to obtain the LST under cirrus cover from TIR data. To this end, new LST retrieval algorithms should be developed to compensate for the effect of the cirrus clouds.

5.4. Retrieval of component temperatures in heterogeneous pixels

In a heterogeneous and non-isothermal pixel, the observed radiance is the ensemble radiance of several components (e.g., soil and vegetation). The pixel-average temperature does not reflect the real temperature of each component. If each component is assumed to be isothermal, the component temperature encapsulates more physical meaning than the pixel-average value and provides better parameterizations of the heat fluxes at the land-atmosphere interface. Therefore, the component temperatures of a mixed pixel are more important than the average values. However, the retrieval of component temperatures is difficult because more variables, including the component emissivities and atmospheric effects, must be known in advance. Several authors have attempted to retrieve component temperatures from multi-angular data (Jia et al., 2003; Li et al., 2001; Menenti et al., 2001; Shi, 2011). The methods that they have developed are far from satisfying and should be improved in the future. In addition, further investigations should focus on mining the auxiliary information provided by spatial, temporal, and spectral data. Because different VZAs may correspond to different pixel sizes, new algorithms are expected to use hyperspectral TIR data at a given VZA, as the information regarding the component temperatures within a mixed pixel is included in the hyperspectral TIR data.

5.5. Methodology for retrieving LST from passive microwave data and for combining LSTs retrieved from TIR and passive microwave data

The TIR data provides the LST with a fine spatial resolution (e.g., several kilometers), but it loses efficiency when the land surface is fully or partly covered by clouds. In contrast, microwaves can penetrate clouds, allowing for LST retrieval in all weather conditions but with a coarser spatial resolution (up to tens of kilometers) (Aires et al., 2004). TIR and microwave data can thus complement each other, and the combination of the two is a promising line of research for producing long-term LST products in all weather conditions with a spatial resolution as fine as that of TIR data. Future studies are advised to focus on the following subjects.

(1) Development of a new physics-based model for retrieving LST values from passive microwave data. Several techniques to retrieve the LST from passive microwave data have been proposed, including (semi)empirical statistical methods, neural networks, and physical models (Aires et al., 2001; Chen et al., 2011; Mao et al., 2007; McFarland et al., 1990; Njoku & Li, 1999; Weng & Grody, 1998). However, the physical mechanisms underlying those approaches are generally unclear, and their assumptions or simplifications regarding the LSE and atmospheric effects degrade both the feasibility and the accuracy of the derived LST. New physics-based model for LST retrieval from passive microwave data should be developed by focusing on both simplifying the parameterization of the RTM and developing the emissivity relationships between different frequencies and polarizations.
A satisfactory model is expected to retrieve the LST from a combination of brightness temperatures measured at different frequencies and polarization modes.

(2) Development of a model to derive the skin LST from passive microwave data. As well known, the LST retrieved from microwave data is different from that derived from TIR data. The former reflects an average value of the soil temperature from the land surface to a particular depth (depending on the frequency used to retrieve LST) underneath the surface, whereas the latter is the skin temperature with several microns of depth. To combine these two types of LST and extract the skin LST, a model must be developed to extract the skin LST from the LST derived from passive microwave data with the aid of LSTs derived from passive microwave data at different frequencies and of the thermal conductivity equation applied to soil.

(3) Development of a microwave-TIR fusion model. An effective method that allows the LSTs derived from TIR and passive microwave data must be developed in the future to produce high-resolution spatial LST data in all weather conditions. The key problem to be resolved is how to recover the LST at the spatial resolution of TIR data when a microwave pixel is fully or partly cloudy.

5.6. Methodology for angular normalization of LST

As reported by Lagouarde and Irvine (2008), Lagouarde et al. (1995, 2004), Chehbouni et al. (2001) and Li et al. (2004b), the LST varies with VZA, and its angular variation for three-dimensional surfaces results primarily from the angular variation of the pixel emissivity and the relative weights of different components (e.g., vegetation and background soil) with different temperatures in a non-isothermal pixel. The difference in the LST measured in nadir and off-nadir observations can be as large as 5 K for bare soils and even 10 K for urban areas. Because most polar-orbit satellites (e.g., MODIS, AVHRR) scan the land surface in the cross-track direction with different VZAs varying from −65° to +65°, angle-dependent variations in the retrieved LST are inevitable, making the LSTs of different pixels in the same orbit incomparable and causing erroneous results in application. This effect must also be considered for LSTs obtained from different sensors or at different times. Therefore, it is very crucial to normalize the satellite-derived instantaneous LSTs at various VZAs to a reference VZA (e.g., at nadir).

One method of performing angular normalization on satellite-derived LSTs is to simply attribute the angular variation of the measured effective temperature derived from area-weighted emitted radiances to the directional behavior of the pixel emissivity, as proposed by Li et al. (1999). However, the directional emissivity defined in this manner is usually not measurable from space, and the assumption that there is no downward environmental thermal radiation may cause some unexpected errors in the normalized result. Another technique for normalizing the satellite-derived LST relies on a simplified directional thermal RTM that considers the component temperatures and fractions within the pixel. New methods can begin by parameterizing the directional thermal RTM with the minimum number of unknowns based on the directional fraction of vegetation cover and the component temperatures of pixels. Then, the method should establish relationships between the directional radiative temperatures observed from different directions. The off-nadir LST can be normalized to a reference direction (e.g., at nadir) by determining the fractions of various components and the corresponding component temperatures or their ratios from multi-angle or multi-channel observations. The fraction of components under a specific viewing angle can be calculated using the bidirectional reflectance distribution function (BRDF) model in the visible and near infrared spectral regions. Although angular variations in the LST have been demonstrated or simulated at the pixel scale in the literature (Pinheiro et al., 2006, 2004; Rasmussen et al., 2011, 2010), there is no any practical way to perform angular normalization of satellite-derived LSTs due to the complexity of this normalization. This issue therefore requires further investigation in the future. To validate the normalization model, angular measurements of the thermal radiation at ground level must also be conducted.

5.7. Methodology for temporal (time) normalization of LST

It is well known that the acquisition times (UTC) of all the pixels in one image are nearly the same but the local solar times are much different. For instance, in a polar-orbit satellite image such as a MODIS image, the difference in the local solar time between the east and the west pixels along the scanning line can be up to 1.5 h, which means that the east pixels are exposed to solar irradiation approximately 1.5 h before the west ones if the sky is clear. As a result, the LST products derived from the same satellite cannot be compared if the differences in the local solar times of the pixels are significant. This phenomenon also affects LST products acquired by different satellites at different times and significantly limits the applicability of the instantaneous LST products. It is therefore necessary to temporally normalize the satellite-derived LSTs to the same local solar time. The diurnal temperature cycle (DTC) model shows promising ability to normalize the LST to any time of a cloud-free day. However, only DTC models with six parameters have been developed to describe the diurnal variation of the LST on cloud-free days (Göttsche & Olesen, 2001; Jiang et al., 2006; Schädlich et al., 2001). Because polar-orbit satellites generally pass a given location only once or twice per day (four times total for MODIS Terra and Aqua), either a new DTC model with a minimum number of unknown parameters (less than 4) or a combination of polar-orbit and geostationary satellites must be developed in the future to temporally normalize the polar-orbit satellite-derived LSTs. Considering that geostationary satellites observe the same location with high temporal frequency, LSTs derived from geostationary satellite data can be used to determine a typical DTC model. Assuming that the LSTs derived from polar-orbit satellite data exhibit the same diurnal pattern as those derived from geostationary data, the once- or twice-daily LSTs derived from polar-orbit satellite data can be interpolated to any time of a cloud-free day utilizing the DTC model developed using geostationary data. However, on a partly cloudy day (no cloud contamination when the polar-orbit satellite overpasses), these DTC models will not be applicable, and a local LST variation model should therefore be developed in the future with the aid of temporal information provided by geostationary satellites to normalize the instantaneous polar-orbit satellite-derived LSTs.

5.8. Concerns on the newly developed Hyperspectral Infrared Imager

The Hyperspectral Infrared Imager (HyspIRI), a combined visible, near infrared and shortwave infrared (VSWIR) imaging spectrometer and a multi-channel TIR radiometer, is one of the proposed missions of the “Decadal Survey” program in the National Research Council, U.S. This new system is scheduled to launch in 2013–2016 and designed to have a spatial resolution of 60 m at nadir, and the revisit times at the equator will be 19 and 5 days for the VSWIR and TIR instruments, respectively (Roberts et al., 2012). The objective of the HyspIRI is to continue the space-borne observation on the earth in the last decades and provide synergy analysis between VSWIR and TIR data for the land surface parameters. Up to now, three international HyspIRI symposiums have been carried out by Jet Propulsion Laboratory to discuss its potential applications. For example, Roberts et al. (Roberts et al., 2012) investigated some potential synergies in urban environment using a HyspIRI-like airborne dataset acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the MODIS/ASTER (MASTER) airborne simulator, and found that the atmospheric data retrieved from the VSWIR data provided the
atmospheric corrections of the TIR data, while the TIR radiance can be used to improve reflectance retrievals in the VSWIR especially when significant TIR aerosol absorption is present under high aerosol loadings. Moreover, this HyspIRI-like airborne system was also used to detect fire regions and retrieve their corresponding temperature (Dennison & Matheson, 2011; Matheson & Dennison, 2012). Additionally, HyspIRI is also prospected to be applied in the fire severity assessments, the monitoring of the volcanic eruption, the interpretation of snow properties and the retrieval of forest biophysical parameters, respectively (Dozier et al., 2009; Veraverbeke et al., 2012a, 2012b; Zhang et al., 2012). As for the retrieval of LST and LSE, HyspIRI’s sounding channels in VSWIR can provide atmospheric data for the atmospheric corrections of the MIR and TIR data, and its 8 spectral channels (7 between 7.5 and 12 μm and 1 at 4 μm) are suitable for several current algorithms in theory, such as the TES, TISI and SW algorithms, even available for developing new retrieval LST/LSE algorithms. Furthermore, because of its finer spatial resolution (60 m) and rapid revisit time (5 days), the LST/LSE retrieved from HyspIRI provides an opportunity to validate the LST/LSE products at current algorithms in theory, such as the TES, TISI and SW algorithms, even available for developing new retrieval LST/LSE algorithms. Furthermore, because of its finer spatial resolution (60 m) and rapid revisit time (5 days), the LST/LSE retrieved from HyspIRI provide an opportunity to validate the LST/LSE products at coarser resolutions, such as MODIS and (A)ATSR. Therefore, the study on HyspIRI will attract much attention in the future.

5.9. Physical meaning of satellite-derived LST and its applications

The LST must be physically defined with absolute certainty. However, no agreement has been reached on the definition of the LST because of the unclear physical meaning of the satellite-derived temperature, especially over heterogeneous and non-isothermal surfaces. The definition of the LST also depends on that of the LSE because the LST and LSE are coupled in the total radiation. There are currently several definitions of the LSE, such as the r-emissivity (Becker & Li, 1995), the e-emissivity (Norman & Becker, 1995) and the apparent emissivity (Li et al., 1999). These definitions are the same for homogeneous surfaces at thermal equilibrium, but because natural surfaces observed from space are usually heterogeneous, the assumptions of homogeneity and thermal equilibrium are often violated in reality, especially in measurements with low spatial resolution. Therefore, the differences between these definitions are evident in many cases. The r-emissivity definition is recommended for LST and LSE retrieval from space-based measurements because the r-emissivity is measurable from space.

Whatever definition of the LST is used, the satellite-derived LST, also known as the radiometric temperature or the skin temperature, can only capture the thermal radiation information from a very thin depth underneath the surface, and therefore cannot be directly substituted for the thermodynamic or aerodynamic temperatures in estimating surface fluxes or other relevant applications. Instead, a conversion must be made between these different temperatures. However, current studies seldom consider these differences and treat the skin temperature as the thermodynamic or aerodynamic temperature without any conversion. This simplification causes unexpected uncertainties in their results. Therefore, further attention should be paid to this problem in the future by considering the physical definition of different temperatures and the accuracy requirements of relevant applications.

5.10. Validation of satellite-derived LST

The most important problem in LST validation might be the accuracy and representativeness of the ground-truth LST at the satellite pixel scale. Although ground-based validation technique, measurements of the ground-truth LST are limited by the difficulty of finding a homogeneous region as large as the satellite pixel size and by the difficulty and associated costs of sampling over heterogeneous landscapes. The first difficulty might be solved by improving the spatial resolution of TIR data, while the second will require the development of a new sampling scheme for ground-level LST measurements, such as a wireless-net observation system or multiple-scale observation methods with corresponding new instruments. Attempts to use the LSTs predicted from land data assimilation systems or climate models to indirectly validate the LSTs retrieved from space must devote more attention to improving the accuracy of the output of these models and dealing with the scale mismatch issue in both space and time. Turning to cross-validation techniques, an appropriate scaling procedure to remove the effects of spatial, temporal, and angular effects on the LST must be developed. Therefore, LST validation is still an ongoing subject of research.

Acknowledgments

The authors would like to thank Professor A.R. Gillespie and other reviewers for their valuable and stimulating comments that have greatly improved the paper. The authors are greatly indebted to Dr. S.J. Hook for reading carefully the draft of this paper and for giving many useful suggestions for improving its presentation. This work was supported by the Hi-Tech Research and Development Program of China (863 Plan Program) under grant no. 2012AA12A304, by the National Natural Science Foundation of China under grant no. 41231170 and 41171287, and by the State Key Laboratory of Resources and Environmental System under grant 088RA800KA.

References


