Remote Sensing of Environment 200 (2017) 206–219

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Disaggregation of remotely sensed land surface temperature: A simple yet flexible index (SIFI) to assess method performances

Lun Gao[a], Wenfeng Zhan[a,b,*], Fan Huang[a], Xiaolin Zhu[c], Ji Zhou[d], Jinling Quan[e], Peijun Du[f], Manchun Li[f]

[a] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, International Institute for Earth System Science, Nanjing University, Nanjing, Jiangsu 210023, China
[b] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[c] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic, Hong Kong 999077, China
[d] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[e] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[f] Department of Geographic Information Science, Nanjing University, Nanjing, Jiangsu 210023, China

A R T I C L E I N F O

Keywords:
Thermal remote sensing
Land surface temperature
Disaggregation
Model performance
Accuracy assessment

A B S T R A C T

Disaggregation of land surface temperature (DLST), the aim of which is to generate LSTs with fine resolution, has been attracting increasing attention since the 1980s. The past three decades have been witness to the emergence of DLST methods in large numbers, the accuracies of which were often assessed by comparing the disaggregated with fine spatial resolution LSTs using error indexes such as the root mean square error (RMSE). However, the majority of previous error indexes are, by their nature, insufficient for assessing the performances of DLST methods. This insufficiency is due in part to their lower competence at distinguishing the DLST error from LST retrieval errors and in part to their inability to remove the process controls resulting from different thermal contrasts, temperature units, and resolution ratios among different scenarios in which DLST is conducted. This is also because they are unable to denote the sharpening statuses of the DLST results (e.g., under- or over-sharpening). This *status quo* has made the evaluation of method performances challenging and sometimes unreliable.

To better assess DLST method performances under diversified scenarios, we formulated five protocols, through which a simple yet flexible index (SIFI) was subsequently designed. The establishment of an SIFI includes the following four steps: (1) a detail-based evaluation, which is designed primarily to exclude the impacts of systematic deviations on estimated LSTs; (2) a Gaussian normalization, which is primarily intended to remove the differences in temperature units and thermal contrasts; (3) a triple comparison, with the aim of attenuating the influence of the difference in the resolution ratio in comparisons of method performances; and (4) a piecewise comparison, which is primarily scheduled to distinguish among the three sharpening statuses, under-sharpening, acceptable over-sharpening, and unacceptable over-sharpening. The evaluation ability of SIFI was compared with those of the RMSE, Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), and image quality index (Q) using simulation tests and actual thermal data. The results illustrate that SIFI generally outperforms the other indexes; it is able to mitigate the impacts from process errors and controls during evaluation and is able to indicate the sharpening statuses accurately. We believe this new index will likely promote the design of future DLST algorithms and procedures.

1. Introduction

The large-scale monitoring of the thermal status of land surfaces was difficult and even impossible until the advent of satellite thermal infrared remote sensing. Thermal sensors enable the generation of the land surface temperature (LST) products, which are instrumental to research in many disciplines (Anderson et al., 2012; Bisht et al., 2005; Jiménez-Muñoz et al., 2016; Sandholt et al., 2002; Sobrino et al., 2007, 2012; Teggi, 2012). However, spaceborne sensors are subject to a tradeoff between spatial and temporal resolutions (Zhan et al., 2013), and the spatial resolution of thermal spaceborne-derived LST maps is too coarse for many applications. This challenge has encouraged research

* Corresponding author at: Nanjing University, Xianlin Campus, No.163 Xianlin Avenue, Qixia District, Nanjing, Jiangsu 210023, China.
E-mail addresses: gaolun724@foxmail.com (L. Gao), zhanwenfeng@nju.edu.cn (W. Zhan), nju_huangfan@163.com (F. Huang), zhuxiaolin55@gmail.com (X. Zhu), jzhou233@uestc.edu.cn (J. Zhou), quanjinlin_@126.com (J. Quan), dupjrs@126.com (P. Du), limanchun@nju.edu.cn (M. Li).

http://dx.doi.org/10.1016/j.rse.2017.08.003
Received 10 January 2016; Received in revised form 28 July 2017; Accepted 3 August 2017

on the spatial disaggregation of LST (DLST), which is able to generate LST images with high spatial and temporal resolutions.

A quick literature survey shows that DLST has experienced phenomenal growth in the past three decades, and more methods have been proposed, particularly since the 2000s (Zhan et al., 2013). Most of them tried to reconstruct thermal details with the aid of finer resolution data sets (e.g., data in other bands, classification maps, or designed scaling factors), transform these details into thermal ones by statistical inference, and finally add them to the coarse resolution LSTs (Chen et al., 2014). In considering the fast dynamics of LSTs, recent developments in DLST have been focused on the simultaneous disaggregation of the spatial and temporal resolutions (Addesso et al., 2015; Mechri et al., 2014; Moosavi et al., 2015; Weng et al., 2014; Wu et al., 2015; Zhan et al., 2016; Gao et al., 2017). It is anticipated that DLST will continue to be a research focus in the foreseeable future because the resolutions of the current and planned satellite thermal sensors remain far from satisfactory for the relevant applications (Anderson et al., 2012; Lagouarde et al., 2013; Roberts et al., 2012; Teggi and Despini, 2014; Zhou et al., 2013).

As more and more DLST methods are being proposed, an index that is more appropriate for assessing their performances under various scenarios is urgently required. Early studies on DLST often employ conventional indexes that measure the similarity between disaggregated and fine resolution LSTs, such as the root mean square error (RMSE) (Agam et al., 2007) and the mean absolute error (MAE) (Nishii et al., 1996; Stathopoulou and Cartalis, 2009). Subsequent studies also use the Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), which is able to eliminate the resolution difference between pre- and post-disaggregation LSTs (Gevaert and García-Haro, 2015; Pardo-Igúzquiza et al., 2006). In considering the documented similarity between DLST and optical image fusion (OIF) since 2010, a few researchers have resorted to the indexes that were designed to evaluate the OIF algorithms. These indexes include the universal image quality index (Q) (Mukherjee et al., 2014; Zhou et al., 2016) and the structural similarity index (SSIM) (Rodriguez-Galiano et al., 2012), among others.

Practitioners may also turn to other advanced indexes in OIF that were recently developed for evaluations, such as the four bands multispectral images fusion index (Q4), the Quality with No Reference (QNR) (Vivone et al., 2014), or the combination of various indexes (Despini et al., 2014). However, although it inherited some traits from the OIF, DLST has differed from its counterpart in the following two regards. First, the DLST process strictly requires that the thermal radiance of a single pixel block at the coarse resolution should be equal to the mean thermal radiance of the corresponding disaggregated fine resolution pixels (Liang, 2005) because its applications are primarily quantitative, whereas spectral distortion is occasionally tolerable in OIF (Pohl and Van Genderen, 1998). Second, fine resolution LSTs, which are either obtained by using thermal data from a different sensor or directly produced by the aggregation-and-then-disaggregation strategy, are indispensable for validation (Zhan et al., 2011). Although references from a different sensor can also help in the evaluation of OIF techniques, they are frequently assessed by comparison with the coarse resolution multispectral images in terms of spectral distortion and with the fine resolution panchromatic images with respect to spatial details (Vivone et al., 2014).

Although previous indexes can be used to assess the performances of DLST methods, they are intrinsically flawed in the following three regards. First, their values depend on multiple errors, including those from disaggregation methods but also those due to the preprocessing of LST, which is unrelated to the model performance (e.g., the temperature retrieval error; more clarifications are given in Section 2.1). Any comparison that disregards the errors due to LST preprocessing would no longer be related to the method performances alone. Second, their values depend on multiple controls, including that from the disaggregation method but also those related to the thermal contrast difference and resolution gap. Finally, their values are mostly not indicative of the sharpening statuses including under-sharpening, acceptable over-sharpening, and unacceptable over-sharpening (more clarifications are given in Section 2.3).

To address these issues, this work designed a new index able to better assess DLST method performances. Followed by the clarifications of background (Section 2) and the five protocols (Section 3) that an index should comply with, Section 4 provides the definition of this index. Sections 5 and 6 exhibit the experiment, the results and discussion, respectively. The conclusions are finally drawn in Section 7.

2. Background

An accurate evaluation of the method performances requires researchers to first identify all the possible errors/controls that may affect the associated evaluation. Generally, the overall errors of disaggregated LSTs (given as err_{overall}) can be expressed as the function of the temperature retrieval errors (given as e_{LST}, including the errors from both the original low-resolution and the reference fine resolution LST images), the image co-registration error (given as e_{cr}), and the DLST error (given as e_{DLST}). In other words, err_{overall} can be expressed as follows:

$$err_{\text{overall}} = q_1\left(\overbrace{e_{\text{LST}},\, e_{\text{cr}}}^{\text{process error}},\ \overbrace{e_{\text{DLST}}}^{\text{DLST error}}\right) \tag{1}$$

where q_1 is the function between err_{overall} and the three types of errors. Hereafter, we refer to the combination of e_{LST} and e_{cr} as the 'process errors' because they primarily stem from the pre-processes that are performed before DLST is conducted (more clarifications are given in Section 2.1).

Nevertheless, it remains unsuitable to use e_{DLST} to represent the performances of the DLST methods because e_{DLST} is also dependent on several other controls in addition to the performance control (given as c_{pm}). These controls are involved in scenarios under which the method performance can be distorted; they include scenarios with different thermal contrasts (given as c_{tc}), temperature units (given as c_{tu}), and resolution ratios (given as c_{rr}). Therefore, the DLST error can be given by the following:

$$e_{\text{DLST}} = q_2\left(\overbrace{c_{\text{tc}},\, c_{\text{tu}},\, c_{\text{rr}}}^{\text{process control}},\ \overbrace{c_{\text{pm}}}^{\text{performance}}\right) \tag{2}$$

where q_2 is the function between the DLST error and the associated controls. Hereafter we refer to the combination of c_{tc}, c_{tu}, and c_{rr} as the 'process controls' (refer to Section 2.2 for more details).

The above analysis indicates that the evaluation of method performances should be conducted by c_{pm} rather than by err_{overall}. In other words, the impacts from the process errors and controls should be excluded before the precise evaluation of method performances. In addition, the performance evaluations would be further improved, once the sharpening statuses, including the under-sharpening, acceptable over-sharpening, and unacceptable over-sharpening, is determined. Elaborate interpretations of this issue are presented in Section 2.3.

2.1. Process errors

As indicated by Eq. (1) and graphically represented in Fig. 1, the overall errors for disaggregated LSTs include both the DLST and process errors. The process errors can be divided into temperature retrieval and image registration errors.

The remote retrieval of the surface temperature is a complex process (Fig. 1). Temperature retrieval errors may be directly due to noise-equivalent temperature differences (NEΔT) (Gillespie et al., 1998), or due to inaccuracies/differences in the conversion from the digital number (DN) to the thermal radiance (i.e., the radiometric calibration process) and then to the brightness temperature (BT) (see Fig. 1). These