An Analysis of the Stability and Trends in the LST_cci Land Surface Temperature Datasets Over Europe

E. J. Good, F. M. Aldred, D. J. Ghent, K. L. Veal, C. Jimenez

AGU ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002317

Key Points:

- The stability of six satellite land surface temperature data sets is assessed using homogenized station air temperatures over Europe
- Only two data sets are stable; the other four show non-climatic discontinuities associated with a change in sensor and/or drift over time
- Significant trends in land surface temperature of ~0.65 K/decade over Europe are found for one of the stable data sets over 2002–2018

Correspondence to:
E. J. Good,
elizabeth.good@metoffice.gov.uk

Citation:
Good, E. J., Aldred, F. M., Ghent, D. J., Veal, K. L., & Jimenez, C. (2022). An analysis of the stability and trends in the LST_cci Land Surface Temperature datasets over Europe. *Earth and Space Science*, *9*, e2022EA002317. https://doi.org/10.1029/2022EA002317

Received 4 MAY 2022
Accepted 19 JUL 2022

Author Contributions:

Conceptualization: E. J. Good
Data curation: D. J. Ghent, K. L. Veal, C. Jimenez
Formal analysis: F. M. Aldred
Funding acquisition: E. J. Good, D. J. Ghent
Investigation: E. J. Good, F. M. Aldred
Methodology: E. J. Good, F. M. Aldred
Project Administration: E. J. Good, D. J. Ghent
Software: F. M. Aldred
Supervision: E. J. Good
Visualization: F. M. Aldred
Writing – original draft: E. J. Good

An Analysis of the Stability and Trends in the LST_cci Land Surface Temperature Datasets Over Europe

E. J. Good[1], F. M. Aldred[1], D. J. Ghent[2], K. L. Veal[2], and C. Jimenez[3]

[1]Met Office Hadley Centre, Exeter, UK, [2]Department of Physics and Astronomy, National Centre for Earth Observation (NCEO), University of Leicester, University Road, Leicester, UK, [3]Estellus, LERMA, Paris, France

Abstract Long-term satellite land surface temperature (LST) data are desirable to augment 2m air temperatures (T2m) measured in situ and as an independent measure of surface temperature change. However, previous studies show variable agreement between LST and T2m time series. The objective of this study is to assess the stability and trends in six new LST data sets from the European Space Agency's Climate Change Initiative for LST (LST_cci). LST anomalies are compared with homogenized station T2m anomalies over Europe, which verifies all six data sets are well coupled (LST vs T2m anomaly correlations and slopes: 0.6–0.9). The temporal stability of the LST_cci data is assessed through a comparison with the T2m anomaly time series. Only the LST_cci data sets for the MODerate resolution Imaging Spectroradiometer (MODIS) onboard Aqua and the Advanced Along-Track Scanning Radiometer (AATSR) appear stable; the MODIS/Terra, ATSR-2, and multisensor InfraRed and MicroWave data sets show non-climatic discontinuities associated with changes in sensor and/or drift over time. For MODIS/Aqua (2002–2018), significant trends in LST of 0.64–0.66 K/decade compare well with the equivalent T2m trends of 0.52–0.59 K/decade. The LST and T2m trends for AATSR (2002–2012) are found to be statistically insignificant, likely due to the comparatively short study period and specific years available for analysis. No evidence is found to suggest that trends calculated using cloud-free InfraRed observations are affected by clear-sky bias. This study suggests that satellite LST data can be used to assess warming trends over land and for other climate applications if the required homogeneity is assured.

Plain Language Summary Long-term changes in surface temperature due to climate change are usually studied using near-surface air temperatures (T2m) from weather stations. However, these data are not available everywhere and changes in T2m are poorly understood in some regions (e.g., parts of Africa). This study assesses the potential for satellite-observed land surface temperatures (LST) to help fill these gaps. The study finds that LST and T2m provide very similar information on surface temperatures and therefore LST can provide useful information where T2m data are unavailable. However, some of the LST data sets are found to be "unstable" and suffer from jumps and changes over time that result from for example, changes in satellite instrument or a change in the orbit of the satellite platform. These unstable LST data sets cannot be used to measure changes in surface temperature. However, the study finds that for "stable" LST data sets, which do not suffer from these non-climatic jumps and changes, the changes in LST over time are very similar to, or even the same as, the changes in T2m. Therefore, this work suggests that satellite LST data could be used to measure changes in surface temperature where there are not enough stations to make these measurements using T2m.

1. Introduction

Land surface temperature (LST) is a key parameter that drives surface processes, such as the exchange of energy and water between the land surface and atmosphere, and influences human, animal and plant health. It was defined as a Global Climate Observing System (GCOS) Essential Climate Variable (ECV) in the 2016 GCOS implementation plan (GCOS, 2016). LST represents how hot or cold the Earth's surface is to the touch and is not the same as near-surface air temperature measured at a height of about 2m above the surface (T2m), which has been measured at weather stations since the 1700s. Nonetheless, LST and T2m are very often well coupled and may even be the same when solar heating is low or absent, and/or wind speeds are moderate-to-high. However, differences of more than several kelvin are frequently observed when insolation is moderate-to-high and may even exceed 20 K in some situations (Good, 2016a).

Writing – review & editing: F. M. Aldred, D. J. Ghent, K. L. Veal, C. Jimenez

While LST can be measured at the surface using in situ radiometers (e.g., Good, 2016a), the primary source of LST data is from space-borne radiometers operating at InfraRed (IR) and MicroWave (MW) wavelengths. The last two decades have seen several satellite LST data sets become available to users, such as those from the MODerate resolution Imaging Spectroradiometer (MODIS) onboard the polar-orbiting Terra (2000–present) and Aqua (2002–present) platforms (Hulley et al., 2018), the Advanced Along-Track Scanning Radiometer (ATSR) onboard the Envisat satellite (Ghent et al., 2017) and the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the geostationary Meteosat Second Generation platforms (2005–present) (Freitas et al., 2010). As a result, observations of LST from satellites are now widely used in a variety of Earth Science applications, for example, to evaluate land surface models (Koch et al., 2016), diagnose surface response to dry spells (Folwell et al., 2016) and assess drought (Karnieli et al., 2010; Mühlbauer et al., 2016), to characterize urban heat islands (Bechtel et al., 2019; Cheval et al., 2020, 2022; Dousset et al., 2011), and to estimate T2m in the absence of meteorological stations (Chen et al., 2015; Good, 2015; Kilibarda et al., 2014).

In addition, satellite LST can provide a useful, independent measure of surface temperature change that has been traditionally assessed through the use of global weather station data (IPCC, 2021; Jones et al., 2012). Global station coverage is incomplete leading to spatial gaps in derived data sets, which in turn result in larger uncertainties on estimated trends (Stooksbury et al., 1999). Even when gap filling is carried out, which is often required for statistical analysis or for model input, this can contribute to uncertainties and biases in the long term trends (Begueria et al., 2019). Consequently, numerous studies have used satellite LST data sets to create more spatially complete and representative T2m data sets. For example, Kilibarda et al. (2014) use a spatio-temporal geostatistical model and spatio-temporal regression-kriging with MODIS LST to predict T2m at 1 km spatial resolution for the globe between 2001 and 2013. Chen et al. (2021) use MODIS LSTs to generate an all-sky 1 km daily T2m data set for China between 2003 and 2019. The recent EU Surface Temperature for All Corners of Earth (EUSTACE) project used satellite LST, together with satellite sea and ice surface temperatures, and in situ observations to create a global T2m analysis for every day since 1850 (Rayner et al., 2020).

In addition to supplementing T2m data, a few studies have also used satellite LST data directly to estimate changes in surface temperatures, particularly where station coverage is sparse. For example, Jiménez-Muñoz et al. (2013) detect anomalous warming in the Amazon dry season for the major drought years of 2005 and 2010 using 13 years of MODIS LST and ERA-Interim skin temperatures. Aguilar-Lome et al. (2019) use MODIS LST data to study changes in LST in the Andean region, obtaining a trend in winter daytime LST of 1.0 K/decade, although they find this has a clear dependence on elevation. MODIS LST data are also used by Zhou and Wang (2016) to estimate a warming trend of 0.25 K/decade from 2002 to 2015 for LSTs over global deserts. More recently, Sobrino et al. (2020) compared trends in global LST and sea surface temperatures (SST) from MODIS with trends in global T2m data between 2003 and 2016, finding linear trends of 0.18 K/decade and 0.21 K/decade, respectively.

However, such studies require the satellite LST data to be free from non-climatic discontinuities, which is not always the case. For example, Good et al. (2017) found that the 17-year GlobTemperature LST data set (http://www.globtemperature.info/) suffers from a discontinuity when the sensor changes from the ATSR-2 to the Advanced ATSR (AATSR). Discontinuties due to sensor changes are also visible in the LST data sets presented by Khorchani et al. (2018) and Ma et al. (2020), which are based on data from the Advanced Very High Resolution Radiometer (AVHRR) series (see their Figures 12 and 3 respectively). Although the majority of previous studies have looked at trends in single-sensor MODIS data, homogeneity in these data sets is not assured and should first be assessed in order to derive reliable trends. Moreover, it is useful to compare trends in different LST data sets as a measure of uncertainty and understand any other influential factors. Such results will be affected by the different methods used to derive the LST data sets, including differences in algorithm and cloud screening. This type of analysis is hindered by the limited availability of climate-quality LST data sets with sufficient homogeneity and length.

The European Space Agency's (ESA) Climate Change Initiative (CCI) for LST (LST_cci) addresses some of these issues through provision of a suite of LST data sets in a harmonized format with a more consistent approach to LST retrieval, cloud screening and uncertainty derivation. An objective of the project is to deliver a significant improvement on the capability of current satellite LST data records to meet the GCOS requirements for climate applications and realise the full potential of long-term LST data for climate science. The LST_cci project is now in its fourth year and outputs to date include new LST data sets derived from IR sensors in polar (or Low-Earth

Orbiting: LEO) and geostationary (GEO) orbits, and from MW sensors (only available from polar-orbiting platforms). The IR polar-orbiting data sets currently utilise data from the MODIS/Aqua and MODIS/Terra instruments, and from the ATSR-2, AATSR and their successor, the Sea and Land Surface Temperature Radiometer (SLSTR). For the IR geostationary sources, 3 years of the SEVIRI record is currently available, which has a field of view that encompasses Europe, Africa and South America (it should be noted that the SEVIRI LSTs for a longer time period are available from the LSA-SAF, e.g., but only 3 years of SEVIRI data are currently available through LST_cci). Other geostationary IR sensors are also being processed towards achieving the GEORING (i.e., the full circumference of the Earth observed by geostationary sensors), including the Imager on the Geostationary Operational Environmental Satellite (GOES) platforms and the Japanese Advanced Meteorological Imager (JAMI) on the Multi-functional Transport Satellite (MTSAT) platform.

Although separate data sets are provided for each individual satellite, the LST_cci project is also producing some merged and multisensor data sets. These include a multisensor IR data set based on the ATSR-2 and AATSR instruments, a merged GEO/LEO IR data set that is based on both polar-orbiting and geostationary data and a multisensor MW data set based on the Special Sensor Microwave-Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSMIS). Inclusion of MW and merged LST data sets within the LST_cci project is a major advancement over other LST development programmes, which have generally focused on IR LST data sets derived from a single sensor or succession of the same sensors: For example, the SEVIRI LST data set provided by EUMETSAT's Land Surface Analysis Satellite Application Facility (LSA-SAF), which comprises data from multiple, successive SEVIRI instruments. This is reflected in the peer-reviewed literature. However, these IR LST data sets include only cloud-free LST data as clouds are opaque at IR wavelengths, leading to large gaps in coverage that is problematic for many user applications. Using multiple IR sensors can therefore improve spatial and temporal coverage. Using MW LST observations improves coverage even further as these data are near "all sky." A major disadvantage of MW LSTs is that they have lower spatial resolution (12–25 km compared to $\leq$1 km for IR) and accuracy ($\sim$3 K compared with $\sim$1–2 K for IR) (Freitas et al., 2010; Ghent, 2012; Perry et al., 2020; Prigent et al., 2016). However, an important advantage of these MW LST data is that they do not suffer from the large errors due to cloud contamination that has been missed by the cloud screening process, which is often present in IR LST data sets (Ghent, 2012; Good, 2016b; Perry et al., 2020).

The objective of this study is to assess the stability and trends in the LST_cci data sets, and to compare them with trends in T2m. The presence of trends in T2m is well established and is considered one of the major indicators of anthropogenic climate change (IPCC, 2021). As noted above, results from previous studies show variable agreement between time series of LST and T2m. If it can be confirmed here that trends in these LST data sets are reliable and agree with "conventional" trends in T2m, then the two types of data can be used together to build a more complete picture of global temperature change. This is particularly important for studying regional temperature change in some parts of the globe where T2m observations are scarce. This can only benefit the wider effort in the global climate science community to understand and monitor efforts to mitigate global climate change. For LST, there is also an open question concerning how trends in IR LST are affected by using only cloud-free observations. This is the part of the so-called "clear-sky bias" effect and is also examined in this study.

The study utilises the six LST_cci data sets with sufficient temporal extent to perform a meaningful time series analysis. This excludes the LST_cci data sets based on SEVIRI ($\sim$3 years of data), SLSTR-A (<5 years of data), and SLSTR-B ($\sim$2 years of data), together with the merged GEO/LEO IR data set and AVHRR data sets, which were not available at the time the study was carried out. The LST data are first compared with homogenized station T2m data to verify that the relationship between LST and T2m that has been found in other studies is also evident using the LST_cci data sets. The temporal stability of the LST_cci data is then assessed by comparing the time series of LST with the homogenized T2m data, where any differences are assumed to indicate non-climatic discontinuities in the LST_cci data sets. Finally trends in both the LST and T2m data are calculated and compared. All the analysis presented here is performed using anomalies, that is, the deviation from a climatological "normal," which removes the annual cycle and therefore emphasises any other signals present in the data (e.g., long-term changes or non-climatic discontinuities). The analysis is performed using daily LST and T2m data to enable an exact match between the two data types, therefore facilitating a meaningful comparison.

The paper is organized as follows: The data sets used in the study are summarised in Section 2, while the methods are described in Section 3. Results are presented in Section 4, with a discussion on the "clear-sky bias" effect in Section 5. Finally, the conclusions of the study are described in Section 6.

Table 1
Summary of LST_cci Data Sets Used in This Study

Data set name	MODIS/Aqua	MODIS/Terra	ATSR-2	AATSR	Multisensor IR	Multisensor MW
Example product string	L3C-LST-MODISA-0.05deg_1 DAILY	L3C-LST-MODIST-0.05deg_1 DAILY	L3C-LST-ATSR_2–0.05deg_1 DAILY	L3C-LST-ATSR_3–0.05deg_1 DAILY	L3S-LST-ATSR_2–0.05deg_1 DAILY	L3C-LST-SSMI13–0.25deg_1 DAILY
Version	2.00	2.00	2.00	2.00	1.00	2.23
Instrument	MODIS	MODIS	ATSR-2	AATSR	ATSR-2, AATSR	SSM/I, SSMIS
Satellite	EOS Aqua	EOS Terra	ERS-2	Envisat	ERS-2, Envisat	DMSP-F13, DMSP-F17
Spatial resolution	0.05°	0.05°	0.05°	0.05°	0.05°	0.25°
Local time of ascending node	13:30	22:30	22:30	22:00	22:00 (ATSR-2 corrected)	18:00 (with correction)
Data availability	2002–2018	2000–2018	1995–2003	2002–2012	1995–2012	1996–2020
Number of Stations	571	550	221	475	651	279

Note. The bottom row of the table indicates the number of weather stations used in the analysis for each data set (see Section 4).

2. Data

2.1. LST_cci Data Sets

The LST_cci data sets used in this study are summarized in Table 1; all are Level 3 (L3) data sets, where data have been averaged over space and time onto a regular latitude-longitude grid. These include the InfraRed (IR) data sets from the Along-Track Scanning Radiometer (ATSR-2), Advanced Along-Track Scanning Radiometer (AATSR) and MODerate Resolution Imaging Spectroradiometers (MODIS) on board the Terra and Aqua platforms, and the IR and MicroWave (MW) multisensor data sets. The multisensor data sets comprise data from more than one platform merged together to form a long term climate data record (CDR). The multisensor IR data set is based on the ATSR-2 and AATSR, while the multisensor MW data set is based on the Special Sensor Microwave - Imager (SSM/I) and Special Sensor Microwave - Imager/Sounder (SSMIS) series. Descriptions of the LST_cci data sets used in this study are provided below. Further details can be found in Dodd et al. (2019). These descriptions reflect the data sets "as provided" by the project; any further processing applied to these data within this study is described in Section 3.

All the LST_cci IR data sets are daily with separate fields for daytime (LST_{day}) and night-time (LST_{night}) overpasses. The overpass time is different for each instrument (see Table 1) but is stable during its lifetime. For the multisensor IR data set, the LST at the overpass time for the ATSR-2 data (10:30 a.m./p.m.) has been corrected to be equivalent to the LST at the AATSR overpass time (10:00 a.m./p.m.) to ensure consistency across the CDR. For LST_{day}, this is essentially a cooling correction for ATSR-2, while for LST_{night}, the correction will usually have a small warming effect; these corrections typically range between approximately −1 K and +6 K, depending on the time of the year, time of the day and land cover/biome classification. The corrections have been derived using MODIS/Terra as a reference. Matchups have been produced from gridded Level-3 data using a sufficient time window, and are restricted to data where the difference in satellite zenith angle is <10°. These matchup LST differences have been binned to multi-dimensional histograms for day/night, observation time difference, latitude, and landcover class. The LST corrections applied to the data have been calculated from multi-year histograms for each calendar month. LST for all the IR data sets has been retrieved using observations from the split window (SW) instrument channels at ∼11 and ∼12 μm. Split-window retieval coefficients are defined separately for the different IR sensors (Ghent et al., 2017; Perry et al., 2020; Prata, 2002).

The multisensor MW data set is also daily, but with separate fields for ascending (LST_{asc}) and descending (LST_{desc}) orbits, rather than daytime and night-time. This is due to the nominal overpass time of 6:00 a.m./p.m. and orbital drift of the host platforms, Defense Meteorological Satellite Program (DMSP) F13 and F17. A correction is available within the data set to adjust the LSTs to an overpass time of 6:00 a.m./p.m. for both platforms to ensure a more stable LST record; this correction ranges between approximately −6 K and +6 K and depends

on the time of day and orbital drift of the satellite. The correction has been achieved by processing LST for two more SSMIS instruments and deriving the average ratio of the LST differences versus the time differences for the continental land surfaces. This has been done individually for each 0.25° cell of the retrieval grid, and corrects not only the local time drift, but also the changes in local time arising from the instrument scanning with a particular orbit and ground swath width. The errors associated with the correction are larger than for the IR, as MW LST is derived for both clear and cloudy skies, so the insolation conditions are more variable. MW LST also depends more on the surface emissivity compared with the IR, where the surface emissivity varies more substantially with changes in the surface conditions. This makes any attempt to derive an LST adjustment more uncertain as its value depends more on the specific local conditions of each particular day, with typical uncertaintities of the order of ∼1.5°C. Intercalibration of the underlying MW brightness temperatures (BT) is also assured through use of the Fundamental Climate Data Record of Microwave Imager Radiances (Fenning et al., 2020), where the BTs from the different SSM/I and SSMIS instruments have been inter-calibrated to reduce changes related to inter-sensor differences.

The algorithm used for retrieving LST from MW observations is based around a non-linear regression describing the relationship between LST and a combination of MW brightness temperatures covering the range 18–85 GHz. This is implemented by means of a neural network (Perry et al., 2020; Prigent et al., 2016). Together with the brightness temperatures, the algorithm also uses pre-calculated climatological estimates of MW emissivity as an input to help account for the large variation in surface emissivity in some environments. Atmospheric temperature and water vapor information are not used directly as inputs, but the presence of the 22 GHz channel as one of the inputs, which is close to a water vapor line, helps to account for changes in atmospheric conditions during the retrieval.

Data coverage varies between instruments. It is lowest for the ATSRs, which only achieve a revisit approximately every three days up to around 75° latitude because of its narrow swath (512 km), whereas MODIS and and the MW sensors achieve near-global coverage day and night for most of the globe due to their wider swaths (MODIS 2,330 km, SSM/I 1400 km, SSMIS 1707 km). At high latitudes and over the poles all instruments have multiple overpasses each day. Where this is the case, valid LSTs that are closest to the nominal overpass time of the satellite are selected for the single-sensor IR and MW data sets (Table 1). However, for the multisensor IR data set, all available observations are averaged. The approach is different for the multisensor IR data set as it is a v1.0 data set; the averaging approach was abandoned for v2.0 IR data sets as it was found to increase the noise in the data. Large gaps are also present in the IR data sets due to cloud cover, when the instrument cannot ““see” the Earth's surface. The multisensor MW data set is more spatially complete, although retrievals can be affected by precipitating or convective clouds. In addition, the MW LST estimates over locations partially covered by water bodies, such as coastal regions, or inundated areas, can be contaminated by the water presence. The ATSR-2 data sets have two temporal gaps, one of 6 months from January to June 1996 due to a scan mirror failure, and a second from January to June 2001 due to a gyro failure. These gaps are present in both the ATSR-2 and multisensor IR data sets.

Per-grid cell uncertainties are also provided for all LST_cci data sets following the approach originally developed by Bulgin et al. (2016) for sea surface temperatures from the ATSR, which has been adapted for LST as described by Ghent et al. (2019). Further details of all the data sets used in this study can be found in Dodd et al. (2020) and an in-depth report on the algorithms selected for the LST_cci project can be found in Dodd et al. (2019) and Perry et al. (2020). All the data used in this study have been obtained directly from the LST_cci project as beta products. Official versions of these data are now freely available via the ESA CCI data portal (https://climate.esa.int/en/odp/#/dashboard).

2.2. 2m-Air Temperature Data

The LST_cci data sets outlined above are compared with daily minimum (T_{min}), maximum (T_{max}) and mean (T_{mean}) near-surface air temperature measured at ∼2m above the surface (T2m) data provided by the EU Surface Temperature for All Corners of Earth (EUSTACE) project (Rayner et al., 2020) for meteorological stations within Europe and the Mediterranean (Squintu, Van Der Schrier, & Van Den Basselaar, 2019). This data set has been chosen because it is the only daily homogenized station data set in the public domain. Daily data are required in order to obtain an exact temporal match with the satellite data used in this study, which are also at daily temporal resolution. In addition, the data must also be free from non-climatic effects (homogenized) because a key part

of this study is to assess the temporal stability of the satellite LST data and calculate robust trends in T2m that can be compared directly with equivalent trends in LST. Use of a reanalysis data set such as ERA5 (Hersbach et al., 2020) was not considered in this study due to this requirement, as both the number and source of the observations assimilated in for example, ERA5, vary considerably through time (e.g., see Figure 3 and Table 4 in Hersbach et al. (2020)), which may introduce non-climatic discontinuities.

The daily minimum and maximum temperatures series, which originate from the European Climate Assessment and Data set (ECA& D, Klein Tank et al., 2002; Klok & Klein Tank, 2008), were homogenised using the quantile matching approach, which corrects for changes in station location or measurement equipment, for example, (Squintu, Van Der Schrier, Brugnara, et al., 2019). An automated method locates breaks in the series based on a comparison with surrounding series and applies adjustments which are estimated using homogeneous segments of surrounding series as a reference. In total, around 2,100 series have been adjusted. Data are provided across Europe which defines the region of this study to this area only. Data provision is up to 31/03/2019, however, comparisons are only made in this study up to the end of 2018 due to very few stations being available into 2019.

Although station T_{min} and T_{max}, which may occur at varying times during the day, are used here to assess the stability and trends in satellite LSTs with fixed overpass times, the comparisons are still meaningful despite any time differences between the different observation types. Both T_{min} and T_{max} have increased since pre-industrial times, for example, by 0.186 K/decade and 0.236 K/decade, respectively since the midtwentieth century (Thorne et al., 2016). However, the relative change between them (i.e., the change in the diurnal temperature range) is at least an order of magnitude smaller and may also be statistically insignificant over much of Europe during the satellite era (Sun et al., 2019; Thorne et al., 2016). Therefore differences in the temperature change signal due to observation time differences in this study are expected to be negligible in comparison to any signal in the overall temperature change.

3. Methods

All the MW LST data used in this study have been corrected to the 6 a.m./p.m. overpass time using the LST overpass time correction described in Section 2.

Quality control (QC) is carried out on all data sets using available quality information. For the EUSTACE T2m, data are excluded where the quality flag value is 1 (““suspect”). However, these quality flags are often missing and hence two other basic checks are carried out. First any value below −900 is excluded (−999 is used as the “fill” value in the data set). Second, it was found that some stations only report 0, so these records are also removed.

QC information provided with LST_cci data varies depending on the data set. Comprehensive quality flags are provided with the MW data. These quality flags inform the user where the retrieval uncertainty is expected to be increased. For this analysis, data with the following attributes are removed due to the associated high uncertainties in the data:

- Ground with large MW penetration depth
- Possibility of deep convection
- Coast
- Possibility of inundated land

Data are also flagged where snow and ice are present, as these retrievals can have a larger uncertainty. However these data are generally still useable and are therefore not excluded from this study, which in line with the approach also taken for the IR LSTs.

For the multisensor IR data set (v1.00) the following criteria must be met for the data to be included in the analysis:

- No water present in pixel
- Fraction of clear sky must be ≥ 0.9

For the v2.00 data (IR single sensor data sets), QC flags indicate if the observation time is near to the nominal overpass time and data are excluded where this is not the case. This is relevant where data are gridded at high latitudes and where there may be multiple observations for each grid cell each day. Grid cells containing cloud or water are excluded by default in the data set, and it is presumed that this is correct with no further screening carried out.

The QC'd, gridded IR LST_cci data sets are spatially collocated with EUSTACE station locations by taking the mean of a 3 × 3 block of grid cells centered on the grid cell containing the station location. The MW LST_cci data set is collocated using only the grid cell containing the station location due to the larger grid cell size (0.25° for the MW data set vs. 0.05° for the IR data sets). This results in time-series for LST and T2m at each station location, for each LST data set, which can be compared directly. Where a value is available for both LST_{min} and LST_{max} (or LST_{asc} and LST_{desc} for the multisensor MW data set) LST_{mean} is the mean of these two values. This can be compared with the T_{mean} timeseries.

Checks are carried out on each LST and T2m time series before further processing to ensure consistent data coverage over the periods of interest (i.e., the time period covered by each separate LST data set). The LST data sets generally have excellent data provision throughout the period of instrument operation (Table 1), with the exception of ATSR-2, which has two gaps of 6 months (January to June 1996 and 2001). The ATSR instruments generally have low data coverage due to the 3-day revisit times for lower latitudes. Because of this, a basic check is carried out to ensure each station-LST timeseries has some data available for each calendar year during the period of interest. Station-LST time series that do not meet this criteria are not used in the analysis. The EUSTACE T2m station data sets vary much more – some have long data gaps or simply do not provide data throughout the time period of interest. To ensure good temporal coverage, any T2m station which does not have some data available for each calendar month during the period of interest is not used. Different approaches to check for coverage are necessary for T2m and satellite LST data due to the different types of gaps in the data sets; in general the satellite data coverage is near complete apart from the known data gaps in the ATSR-2 record and missing data due to cloud in the IR data sets. If any of the data provision criteria described above are not met for day and night/ascending and descending (LST) or T_{min} and T_{max} (T2m) the data for that station are not processed further. Mean temperatures are not considered as part of the filtering as they are calculated directly from these data pairs, for example, the mean of LST_{night} and LST_{day} can only be calculated when LST_{night} and LST_{day} are both available. This results in fewer available data points for LST_{mean}.

All the analysis presented in this study is performed on anomalies (i.e., the deviation from a baseline daily climatology). This is to remove seasonal cycles in the data, which will dominate any temperature signal, and also to minimize the inherent differences that occur when comparing data from a point location (e.g., a station) to the areal average represented by the satellite data (i.e., 0.05°–0.25°). These point-to-areal-average differences may be further exaggerated where the position of the weather station is unrepresentative of the wider surroundings. Anomalies are frequently used in the analysis of surface temperature as they tend to be correlated over much larger distances than the actual temperatures observed (Jones et al., 1982).

The baseline climatologies used here are calculated for the LST and T2m time series separately, for each station location, based on an 11-day moving mean calculated for each calendar day of the year. This is further smoothed over a 31-day period using a convolution function. These smoothing windows were selected via visual inspection of data at several stations to achieve a smooth climatology that also captures the seasonal extremes. LST and T2m daily climatologies are created for the time periods covered by each respective LST data set (Table 1) to facilitate direct comparison between the two data types. For example, LST and T2m climatologies are calculated for 2002–2018 for the MODIS/Aqua comparison. For each data set, all available QC'd data are used within this time period; this approach in the study is discussed further in Section 3.1.

Anomalies for day d are calculated by subtracting the climatological value for that day (LST_{clim_d}, $T2m_{clim_d}$) from the observed data (LST_{obs_d}, $T2m_{obs_d}$) for each data set available (min, max and mean for T2m; day, night and mean where available for IR LST; ascending, descending and mean where available for MW LST):

$$LST_{anom_d} = LST_{obs_d} - LST_{clim_d} \quad (1)$$

$$T2m_{anom_d} = T2m_{obs_d} - T2m_{clim_d} \quad (2)$$

For the IR data sets, LST_{night} and LST_{day} are compared to T_{min} and T_{max} respectively. For the multisensor MW data set, the LST_{desc} (~6 a.m.) is compared to T_{min} and LST_{asc} (~6 p.m.) is compared to T_{max}. LST_{mean} and T_{mean} are also compared for all data sets. For simplicity, all these comparisons are hereon referred to using only the T2m terminology, that is, "T_{min} comparison," "T_{max} comparison," and "T_{mean} comparison."

The LST and T2m anomalies are then compared in several ways. First, the relationship between LST and T2m anomalies (LST anomaly vs T2m anomaly) is assessed at each station by calculating the Pearson correlation coefficient (r), and the slope (m) using the Median of PairWise (MPW) slopes (Sen, 1968). The MPW slopes method is recommended by Fernandes and Leblanc (2005) for remote sensing applications, and is chosen for this study due to its robustness to outliers and analytical estimates of confidence intervals. A value of unity for both these parameters indicates a perfect relationship at that location.

Second, the stability of the LST data sets is evaluated by comparing monthly mean anomalies for LST and T2m using only temporally-matched observations that are available in both data sets. Monthly rather than daily data are used for this analysis to ensure a clear signal over time that is not dwarfed by the "noise" that is present in a daily time series. The mean (spatial average) of all monthly means (LST_{anom_m}, $T2m_{anom_m}$) over all available stations is calculated and a time series of differences is calculated as:

$$\Delta T_{anom_m} = LST_{anom_m} - T2m_{anom_m} \quad (3)$$

The trend and confidence interval of this time series of differences is calculated as above using the MPW slopes. Should the confidence interval of this trend encompass zero, the slope is considered to be statistically insignificant, indicating agreement between the trends in LST and T2m. Therefore a confidence interval which does not encompass zero indicates a statistically significant difference between the trends in LST and T2m. The agreement between the monthly mean anomaly time series is also assessed by calculating the Perason correlation coefficient (r).

For the LST multisensor data sets that contain observations from multiple platforms, the trend, confidence interval, and correlation coefficient are also calculated separately for each platform in the series, together with a standard deviation of the difference from T2m anomalies. Any variations in these values between platforms is likely to be due to non-climatic discontinuities in the data sets, and indicates that they are not fully homogenised and may not be appropriate for climate trend analysis.

Finally, trends in the LST and T2m monthly anomaly time series are calculated and compared using the MPW slopes. Monthly data are used to calculate trends in preference to daily data as this not only provides a less noisy time series, but also a more consistent density of observations with time, that is, one data point per month compared to between one and 31 data points per month in a daily time series.

3.1. Choice of Data Used for the Calculation of Climatologies

The aim of this study is to evaluate the temporal stability of the LST_cci data sets and establish whether there is agreement between the anomalies trends in LST and T2m. The overall objective is to ascertain whether LST data can be used to provide consistent information on surface temperature changes where T2m observations are sparce or absent. T2m data used for climate monitoring are based on "all sky" observations from weather stations (IPCC, 2021; Jones et al., 2012; Klein Tank et al., 2002; Rayner et al., 2020; Squintu, Van Der Schrier, Brugnara, et al., 2019). Therefore to establish whether LST anomalies can capture the temperature change signals in the T2m data sets used for climate monitoring, the LST anomalies should be assessed using T2m anomalies calculated with respect to an all-sky T2m climatology. Therefore in this study, all available data are used to calculate the climatologies for each data set. This results in T2m, MW LST and IR LST climatologies representative of all conditions, convection-free, and clear-sky only conditions respectively, which enables a representative assessment of the LST data that are available in comparison to the all-weather T2m that are used for climate monitoring.

4. Results

Spatially co-locating in-situ weather stations with satellite observations for six new LST_cci data sets results in between 221 and 651 station comparisons per data set across Europe (Table 1 and Figure 1). The variation in the number of station comparisons is a result of the requirement to have both station and satellite data available for the same period of analysis, and data availability as a result of the different QC applied to each data set (see

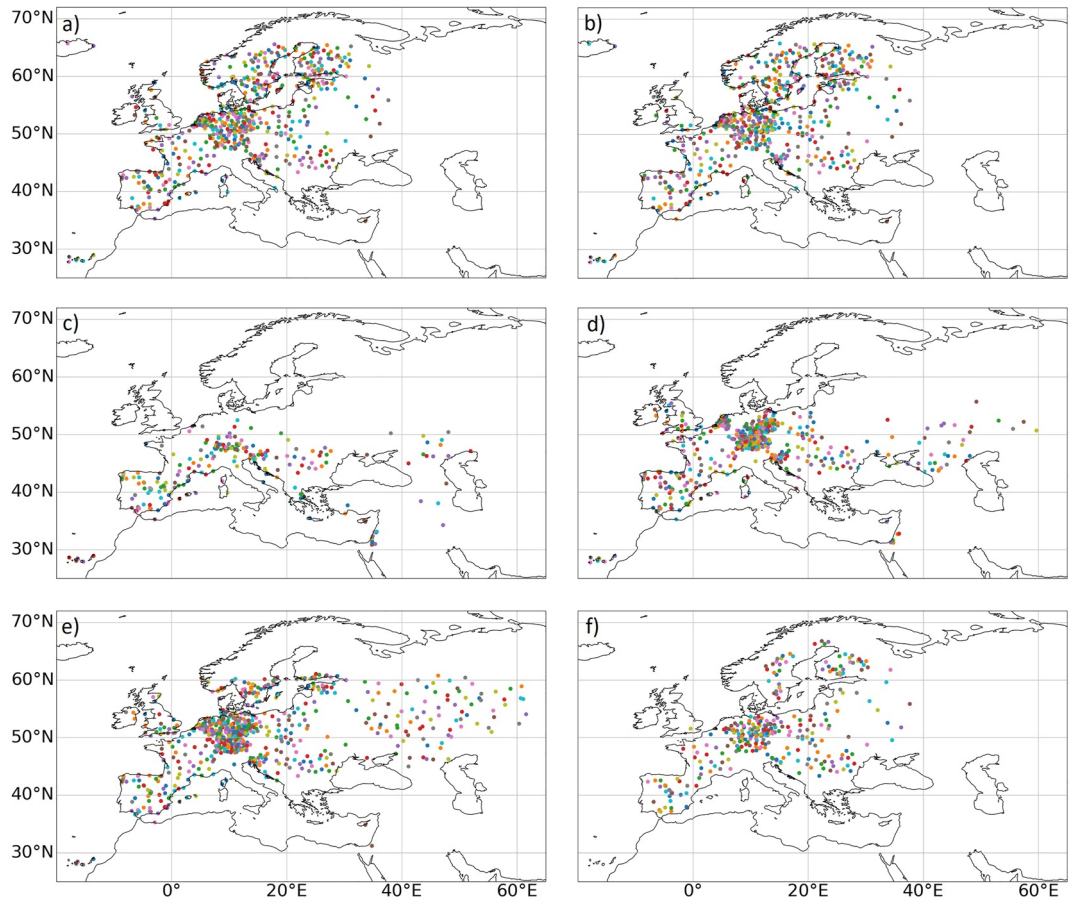

Figure 1. Locations of stations where data are available for both land surface temperature (LST) and 2m air temperature (T2m) for matches with the (a) MODerate resolution Imaging Spectroradiometer (MODIS)/Aqua, (b) MODIS/Terra, (c) Along-Track Scanning Radiometer (ATSR)-2, (d) Advanced ATSR (AATSR), (e) multisensor InfraRed, and (f) multisensor MicroWave data sets. The colors are used for visual purposes only and do not have any meaning.

Section 3). It is notable that the multisensor MW data set is matched with fewer stations, despite the long time series of data. This is due to a large number of MW data being rejected through the application of the QC information, mainly due to the loss of stations near the coast due to the coarser resolution of the MW data. This is evident through the results presented in the following sections (e.g., see Figures 2 and 3). The analysis reported in the following sub-sections is based on these station comparisons.

4.1. LST Versus T2m Relationship

The relationship between LST and T2m at each station location is assessed through the correlation coefficient (r) and the slope (m) of the daily LST anomalies versus T2m anomalies (see Section 3). Figures 2 and Figure 3 show the values of r and m for each station, organized by latitude, for MODIS/Aqua and the multisensor MW data sets, respectively. The results for MODIS/Aqua are typical for the other LST_cci IR data sets (not shown), and show a general increase in r with increasing latitude until ~50°N, with higher variance in the results for the T_{min} comparison compared with T_{max} and T_{mean} comparisons. There is a notable cluster of stations at $r \sim 0.8$ for the T_{max} and T_{mean} comparisons, with some outliers with much lower values of r. A few data points at low latitudes also have poor correlation ($r < 0.5$), which correspond to stations in the Canarias. By contrast, the slopes for MODIS/Aqua (Figure 2) and MODIS/Terra (not shown) show a slight decrease with increasing latitude, particularly for the T_{max} and T_{mean} comparisons. The slopes for ATSR-2, AATSR and the multisensor IR data sets are slightly more variable (not shown). There are also large latitudinal gaps in these T_{mean} pairs above ~40°N due to the longer revisit time for the ATSR sensors, and the requirement to have both LST_{day} and LST_{night} present to calculate LST_{mean}. The correlation and slope results for the multisensor MW data set are less variable with latitude than those for the IR

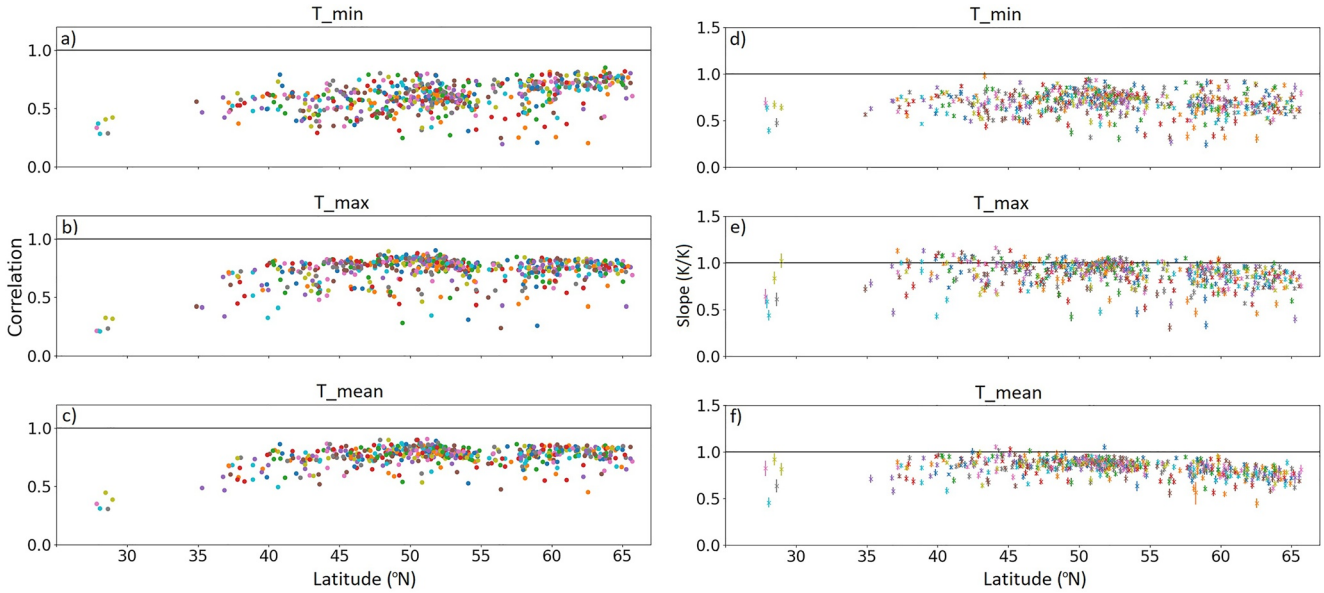

Figure 2. Land surface temperature (LST) versus 2m air temperature (T2m) correlation coefficients for each available station location for the MODerate resolution Imaging Spectroradiometer/Aqua comparison for (a) LST_{night} versus T_{min}, (b) LST_{day} versus T_{max}, (c) LST_{mean} versus T_{mean}, and LST versus T2m slope for (d) LST_{night} versus T_{min}, (e) LST_{day} versus T_{max}, (f) LST_{mean} versus T_{mean}. Each color represents a different station. The MPW slopes are shown as vertical lines to represent the 95% confidence interval for each station.

data sets with very few outliers (Figure 3). This may be a result of the larger footprint of this LST data set (0.25°) as it has previously been shown that correlations between T2m and LST peak where the LST footprint is ~25 km (Sohrabinia et al., 2014). In addition, the larger variability in the results for the IR data sets may reflect cloud contamination in these data, which does not affect the MW LST retrievals; this may be more prevalent at some station locations. Cloud contamination may also contribute to the lower correlations and slopes observed for the T_{min} comparisons, and larger variance in these results for the IR data sets (Table 2), as nighttime cloud-clearing can be less effective.

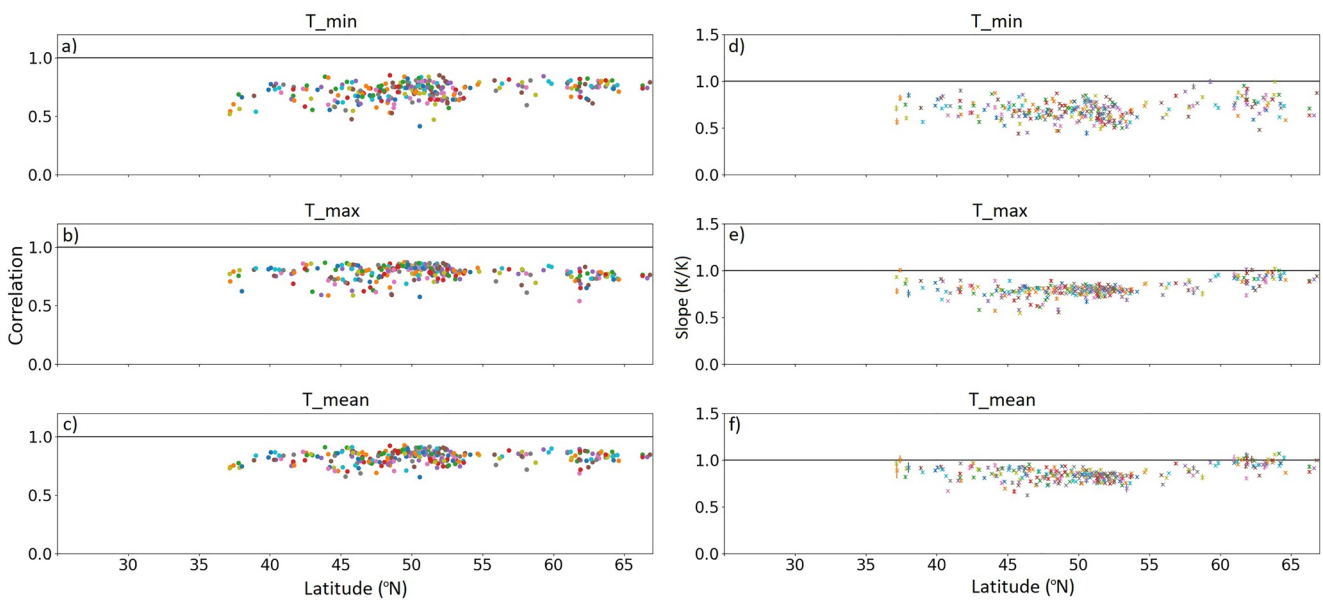

Figure 3. As for Figure 2 but for the multisensor MicroWave data set.

Table 2
Median Correlation Coefficient (r) and LST Versus T2m Slope (m) From All Available Stations for Each LST_cci Data Set Comparison

		MODIS/Aqua	MODIS/Terra	ATSR-2	AATSR	Multisensor IR	Multisensor MW
T_{min}	r	0.62	0.57	0.60	0.63	0.70	0.73
	m (K/K)	0.71	0.63	0.56	0.61	0.67	0.70
T_{max}	r	0.77	0.78	0.74	0.80	0.78	0.79
	m (K/K)	0.91	0.90	0.90	0.89	0.78	0.80
T_{mean}	r	0.78	0.80	0.67	0.72	0.84	0.84
	m (K/K)	0.85	0.88	0.87	0.86	0.79	0.85

It is notable that data coverage at high latitudes is worse for ATSR-2 and AATSR compared to the other IR data sets, which is attributed to the screening applied to these data before performing the collocation (Figure 1).

The correlation coefficients and slopes between daily LST and T2m anomalies generally indicate that the data compare well for T_{max} and T_{mean} with values typically ~0.8 (r) and ~0.9 K/K (m), while the T_{min} comparison values are closer to ~0.6 for both parameters (Table 2). Results are quite consistent across the different data sets and are similar to those reported by Good et al. (2017), who report correlation coefficients and slopes of ~0.7–0.9 for a comparison between T_{mean} and LST_{mean} anomalies over Europe. Slope values < 1 indicate that anomalies in LST are generally smaller in magnitude than those for T2m. The better agreement for the T_{max} comparison compared with the T_{min} comparison could be due to the larger range in the temperature anomalies for the T_{max} comparison compared with those for the T_{min} comparison. This is shown by the probability density functions (PDF) for LST and T2m anomalies from the MODIS/Aqua and multisensor MW comparisons shown in Figure 4. The distributions for T_{max}/LST_{day}/LST_{asc} are wider than those for T_{min}/LST_{night}/LST_{desc}, which could inflate the correlation for the T_{max} comparisons. For example, the 10th and 90th percentiles for the MODIS/Aqua LST distributions are −5.4 and 5.1 K, respectively, for LST_{night}, while for LST_{day}, these values are −6.0 and 5.9 K. Similarly for T2m, the 10th and 90th percentiles for T_{min} are −5.1 and 4.2 K compared with −4.1 and 5.9 K for T_{max}. Figure 4 also shows that the distributions of multisensor MW LST anomalies are much closer to the T2m distribution compared with the MODIS/Aqua analysis. This is also clear from the 10th and 90th percentiles, which are −4.9 and 4.8 K for LST_{desc} and −5.1 and 4.9 K for T_{min}, and −5.8 and 5.6 K for LST_{asc} and −5.7 and 5.8 K for T_{max}. The apparent slightly better agreement between the anomalies for the MW LSTs compared with the IR LSTs is also evident in some of the later results presented in this study (Section 5).

4.2. Stability Analysis

The results in the previous section demonstrate that spatially and temporally collocated LST anomalies calculated from the LST_cci data sets and station T2m anomalies are generally well correlated, which has also been found in studies using other LST data sets (e.g., Good et al., 2017). In this section, the stability of the LST_cci data sets is assessed. Monthly means are calculated for each station anomaly timeseries, where data are available for both LST and T2m (see Section 3). These monthly mean anomalies are averaged over all available stations for each of the six LST_cci data sets (Table 1). The correlation coefficient (r) and the p-value between LST and T2m monthly mean anomalies are calculated together with the mean, standard deviation and slope (trend) of the LST-minus-T2m difference time series. Coupling between the LST and T2m anomalies is greatest where the correlations are highest and standard deviations of differences are lowest. A non-zero mean difference indicates an offset between LST and T2m anomalies. Ideally the trend of the difference should be zero, indicating that the trends in LST and T2m anomalies are identical. Any non-zero trends are considered statistically insignificant here if the 95% confidence interval on the trend encompasses zero. As it is assumed that the station T2m time series are free from non-climatic effects, the LST_cci data set may not be stable over time where the trend of the differences is significant. A summary of the results is presented in Table 3.

The correlation coefficients shown in Table 3 are higher than for daily data (Table 2), with correlation coefficients generally >0.8 (and often >0.9). This is expected because day-to-day variabilty in temperatures is higher than month-to-month variability, which is why monthly averages have been used for this part of the study. With

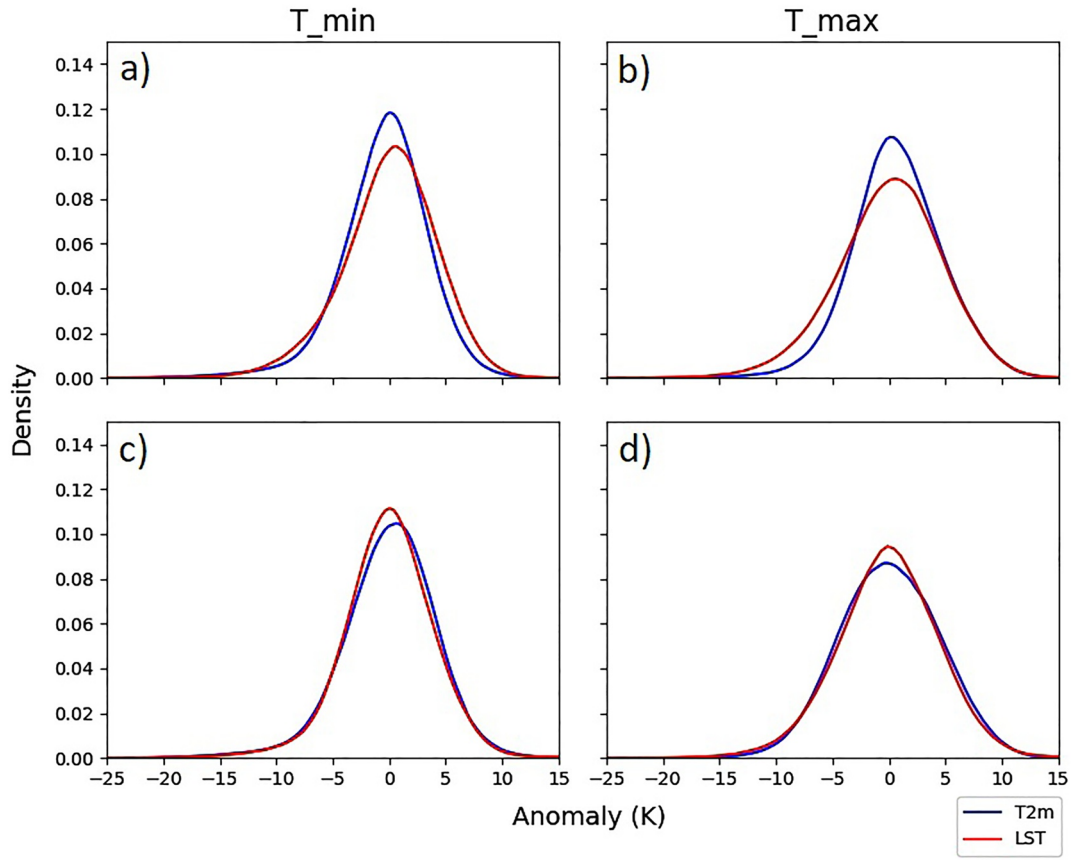

Figure 4. Probability density functions for all daily anomalies for the (a) MODerate resolution Imaging Spectroradiometer (MODIS)/Aqua land surface temperature (LST)$_{night}$ and T_{min}, (b) MODIS/Aqua LST$_{day}$ and T_{max}, (c) the multisensor MicroWave (MW) LST$_{desc}$ and T_{min}, and (d) the multisensor MW LST$_{asc}$ and T_{max}. Blue: T2m. Red: LST.

the exception of the multisensor MW data set, standard deviations of the anomaly differences are lower for the T_{min} comparisons compared with the T_{max} and T_{mean} comparisons. Again, this may be due to the larger range in daily day time temperature anomalies noted in Section 4.1. The mean anomaly difference for all IR data sets is positive for T_{min} comparisons and negative for T_{max}, however both signs occur for T_{mean} comaparisons. It is negative for all MW comparisons. This pattern in the mean differences is attributed to the requirement to use only temporally-matched LST and T2m data for the comparisons, and therefore the exclusion of some LST and T2m data. This is discussed further in Section 5 of this study.

Eleven of the 18 anomaly difference trends in Table 3 are statistically insignificant, suggesting there is no detectable difference between the trends in T2m anomalies and LST anomalies in these cases. However, MODIS/Aqua and AATSR are the only two data sets where all anomaly difference trends are insignificant, and therefore these may be the only LST_cci data sets examined here that are free from non-climatic discontinuities. Some of these discontinuities are clearly visible by eye. For example, Figures 5 and 6 show the time series for the multisensor IR and MW data sets, respectively. For the multisensor IR data set, a discontinuity can be seen when the sensor transitions from ATSR-2 to AATSR. In addition, for the ATSR-2 period the range of differences is visibly larger than for the AATSR, with a positive difference for the T_{min} comparison. The multisensor MW data set shows a discontinuity for T_{min} where the sensor changes from SSM/I to SSMIS in early 2009, however any discontinuity is less clear for T_{max}. Notably, no discontinuities are clear in the time series of anomalies (panels a and b in each figure), which demonstrates the value of examining time series of anomaly differences to assess the stability of these time series.

The differences between the instruments comprising the multisensor data sets are clear when the statistics are considered separately for each instrument (Table 4). The results for the multisensor IR data set show that there

Table 3
Correlation Coefficient (r) for the Relationship Between LST and T2m Monthly Means Averaged Spatially Over All Available Stations for Each LST_cci Data Set

		MODIS/Aqua	MODIS/Terra	ATSR-2	AATSR	Multisensor IR	Multisensor MW
T_{min}	r	0.94	0.93	0.91	0.91	0.91	0.88
	μ (*K*)	0.43	0.27	0.25	0.30	0.31	−0.08
	σ (*K*)	0.48	0.54	0.41	0.49	0.58	0.64
	t (*K*/decade)	0.03	*0.23*	0.03	−0.01	*−0.55*	*0.21*
	CI (*K*/decade)	−0.11–0.17	0.11–0.35	−0.32–0.45	−0.37–0.32	−0.71–−0.40	0.10–0.32
T_{max}	r	0.93	0.90	0.85	0.93	0.77	0.95
	μ (*K*)	−0.68	−0.61	−0.42	−0.57	−1.18	−0.11
	σ (*K*)	0.59	0.73	0.67	0.61	1.25	0.49
	t (*K*/decade)	0.00	0.15	*−1.21*	−0.04	−0.24	0.07
	CI (*K*/decade)	−0.19–0.19	−0.02–0.32	−1.80–−0.55	−0.38–0.32	−0.60–0.13	−0.00–0.15
T_{mean}	r	0.88	0.86	0.74	0.87	0.81	0.94
	μ (*K*)	0.05	0.14	0.06	−0.07	0.31	−0.14
	σ (*K*)	0.81	0.90	0.68	0.55	1.13	0.50
	t (*K*/decade)	0.03	*0.22*	*−1.18*	0.30	−0.08	*0.17*
	CI (*K*/decade)	−0.21–0.26	0.01–0.43	−1.79–−0.59	−0.05–0.68	−0.41–0.25	0.10–0.25

Note. P values are <0.001 for all values of r. Values of r are shown in the first row for T_{min}, T_{max} and T_{mean} respectively. The last four rows in each section relate to the LST-T2m anomaly difference time series, and present the mean (μ), standard deviation (σ), trend (t) and 95% confidence interval (CI) on the trend for each data set. Trends in italics are statistically significant.

are sizable differences between the means and standard deviations of the LST-T2m anomaly differences for ATSR-2 and AATSR, particularly for the T_{min} comparison. The larger mean LST_{night}-T_{min} anomaly difference for ATSR-2 (0.70 K) compared with AATSR (0.06 K) contributes to the overall negative trend in the differences for this data set and suggests a step change between the different sensors. The standard deviation in differences is notably higher for ATSR-2 than for AATSR. This reflects an increase in variance for this part of the data set that has probably resulted from the overpass time correction applied to the ATSR-2 data (Section 2.1). The trends in differences for the two sensors are also substantially different. In summary, Table 4 provides convincing

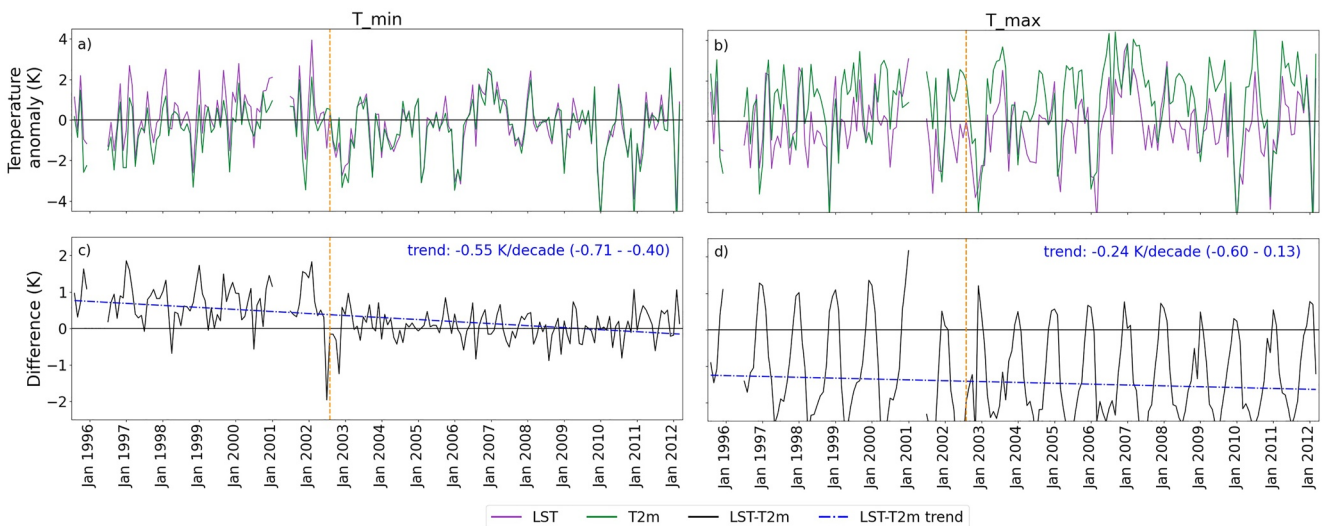

Figure 5. Monthly mean (a) land surface temperature $(LST)_{night}$ and T_{min} and (b) LST_{day} and T_{max} anomaly (K, relative to 1996–2012) averaged over all available stations for the multisensor InfraRed data set, with the respective differences in the time series shown in (c) and (d). The change in sensor from the Along-Track Scanning Radiometer (ATSR)-2 (early period) to the Advanced ATSR (AATSR) (later period) is shown by the vertical dashed line in each panel.

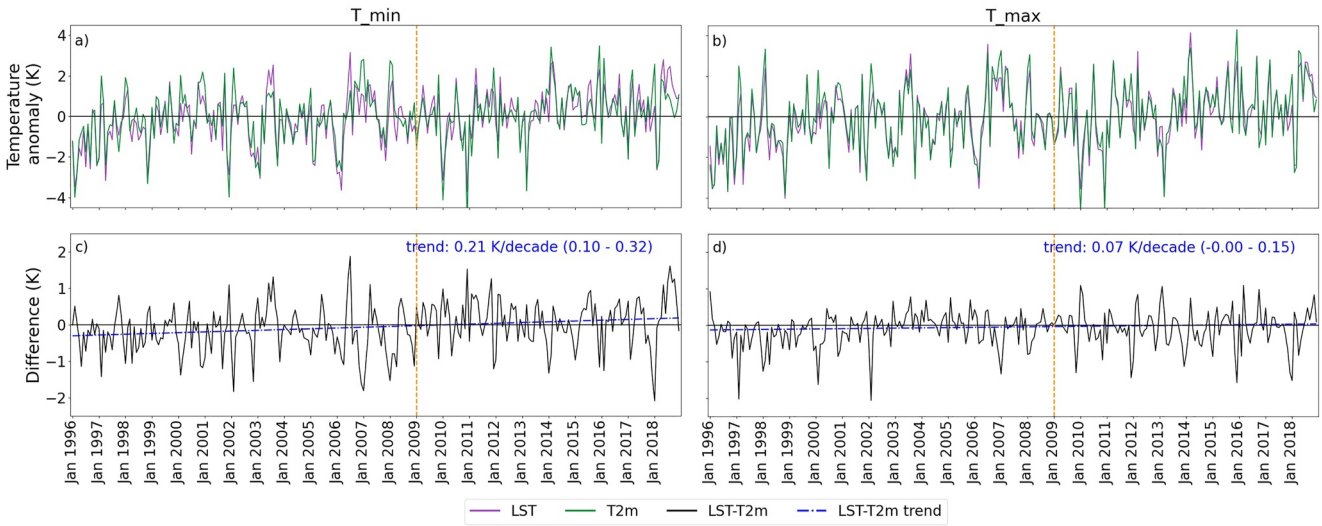

Figure 6. As for Figure 5 but for the multisensor MicroWave (MW) data set. Monthly mean (a) land surface temperature $(LST)_{desc}$ and T_{min} and (b) LST_{asc} and T_{max} anomaly (K, relative to 1996–2018) averaged over all available stations for the multisensor MW data set, with the respective differences in the time series shown in (c) and (d). The change in sensor from the Special Sensor Microwave/Imager (early period) to the Special Sensor Microwave Imager/Sounder (later period) is shown by the vertical dashed line in each panel.

evidence that the multisensor IR data set is inhomogeneous, which was also the finding of Good et al. (2017) for the GlobTemperature CDR based on the same sensors. However, it should be noted that as the trends in the LST-T2m differences are statistically insignificant, the results also suggest that the individual sensors comprising the multisenor IR data set are stable.

The same information for the multisensor MW data set is also shown in Table 4. Unlike for the multisensor IR data set, the correlations and standard deviations of the differences from in-situ T2m are similar for both sensors.

Table 4
As for Table 3 but With Statistics Shown for Each Sensor Comprising the Multisensor IR and Multisensor MW Data Sets in Addition to the Whole Data Set in Each Case

		Multisensor IR			Multisensor MW		
		All	ATSR-2	AATSR	All	SSM/I	SSMIS
T_{min}	r	0.91	0.89	0.96	0.88	0.89	0.88
	μ (K)	0.31	0.70	0.06	−0.08	−0.22	0.11
	σ (K)	0.58	0.60	0.40	0.64	0.62	0.63
	t (K/decade)	*−0.55*	−0.09	0.02	*0.21*	−0.05	−0.02
	CI (K/decade)	−0.71–−0.40	−0.77–0.64	−0.24–0.29	0.10–0.32	−0.30–0.19	−0.39–0.41
T_{max}	r	0.77	0.64	0.84	0.95	0.95	0.95
	μ (K)	−1.18	−1.02	−1.28	−0.11	−0.14	−0.08
	σ (K)	1.25	1.40	1.14	0.49	0.46	0.52
	t (K/decade)	−0.24	−1.66	0.18	0.07	*0.23*	0.02
	CI (K/decade)	−0.60–0.13	−3.17–0.02	−0.48–0.85	−0.00–0.15	0.07–0.38	−0.29–0.29
T_{mean}	r	0.81	0.51	0.90	0.94	0.94	0.93
	μ (K)	0.31	0.39	0.26	−0.14	−0.23	−0.01
	σ (K)	1.13	1.34	0.98	0.50	0.47	0.51
	t (K/decade)	−0.08	−1.18	0.19	*0.17*	0.12	0.08
	CI (Kdecade)	−0.41–0.25	−2.73–0.24	−0.44–0.92	0.10–0.25	−0.05–0.31	−0.23–0.36

Note. Trends in italics are statistically significant.

Although there is a significant positive trend in the differenced time series present for T_{min} and T_{mean} for the whole data set, the equivalent trends for the individual sensors are smaller and statistically insignificant suggesting that the time series for the individual sensors are stable. However, for all comparisons, there is a change in the mean of the differences for SSM/I and SSMIS (−0.22 to 0.11 K for T_{min} and −0.23 to −0.01 K for T_{mean}), which is visible in Figure 6 in the time series for T_{min} as noted earlier. This positive step change is the cause of the significant positive trend in the full timeseries for these comparisons. T_{max} shows an exception to this. The overall trend in the difference for the whole data set is insignificant, however the trend for SSMI is positive and significant, and the trend for SSMIS is insignificant. The change in the mean of the differences is also smaller (−0.14 to −0.08 K). In general, the multisensor MW data set shows slightly better agreement with the T2m anomalies than the multisensor IR data set, reflected in the smaller mean differences and standard deviations. However, as for the multisensor IR data set, Table 4 provides clear evidence that the multisensor MW data set is inhomogeneous over time, although the records for the individual sensors appear stable.

Figure 7 shows the time series of LST-minus-T2m monthly anomalies for the remaining four LST_cci data sets. The ATSR-2 time series appears reasonably stable for the T_{min} comparison, but there appears to be a jump in the T_{max} comparison after the second data gap in mid-2001, which is likely to be the cause of the significant negative trend in this differenced time series (−1.2 K/decade, Table 3). By contrast, the AATSR time series appears stable for all comparisons, which is also reflected in the statistics provided in Table 3. The MODIS/Terra T_{min} comparison has a positive trend in the differenced time series that is both visible and statistically significant (Table 3) and is further explored in Section 4.3 (Table 6). A significant trend is also present in the T_{mean} comparisons. However, the trend for the T_{max} comparison is not statistically significant. As noted previously, the MODIS/Aqua comparisons appear stable, with no significant trends detected in the anomaly difference time series.

Both the MODIS and ATSR analyses display an annual cycle in the LST-T2m differences, particularly for the T_{max} comparisons (Figure 7). The difference peaks in winter ($LST_{anom} > T2m_{anom}$). A candidate explanation for this could be that the climatologies do not fully capture minimum and maximum winter and summer temperatures leading to more extreme anomalies in one or both data sets. However, the smoothing windows used for the climatology calculation have been chosen with care to ensure that the more extreme temperatures are well represented, whilst retaining a reasonably smooth climatology (Section 3). Furthermore, this annual cycle does not occur in the multisensor MW analysis. A visual inspection of the data at individual stations confirms the climatologies represent the seasonal extremes quite well and therefore this hypothesis is rejected. Further analysis of the data indicates this annual cycle results from the requirement to use only temporally-matched LST and T2m data to facilitate a meaningful comparison, and therefore the exclusion of any cloudy T2m observations in the IR comparisons. This is discussed further in Section 5 of this study.

In summary, only the single-sensor AATSR and MODIS/Aqua data sets appear to be free from non-climatic discontinuities and are therefore the only data sets that are considered for reliable trend analysis in the following section.

4.3. Trends

The results of the stability analysis suggest that the AATSR and MODIS/Aqua are the only LST_cci data sets that are stable enough to be considered for trend analysis. Nonetheless, trends for all six LST_cci data sets are presented in Table 5 for completeness and to demonstrate the impact of using non-homogeneous time series for calculating trends.

This section focuses only on the trends labeled “LST” and “T2m” in Table 5, which are calculated from the monthly collocated anomaly time series of LST and T2m for each of the T_{min}, T_{max}, and T_{mean} comparisons used in the stability analysis (Section 4.2). These LST and T2m data are are matched in time, and therefore the time series exclude any days that are cloudy for the IR comparisons, and any other occasions where the LST data are missing (e.g., the two 6-month gaps in the ATSR-2 record). The trends reported in the ““T2mAll” rows in Table 5 are discussed in Section 5.

The “LST” and “T2m” trends show very good agreement for both the MODIS/Aqua and AATSR LST_cci data sets. For MODIS/Aqua, significant trends in LST of ∼0.64–0.66 K/decade are obtained for the T_{min}, T_{max} and T_{mean} comparisons. This compares well with the equivalent T2m trends of ∼0.52–0.59 K/decade. For the AATSR, although the trends are consistent between LST and T2m, none of these trends are statistically significant. In

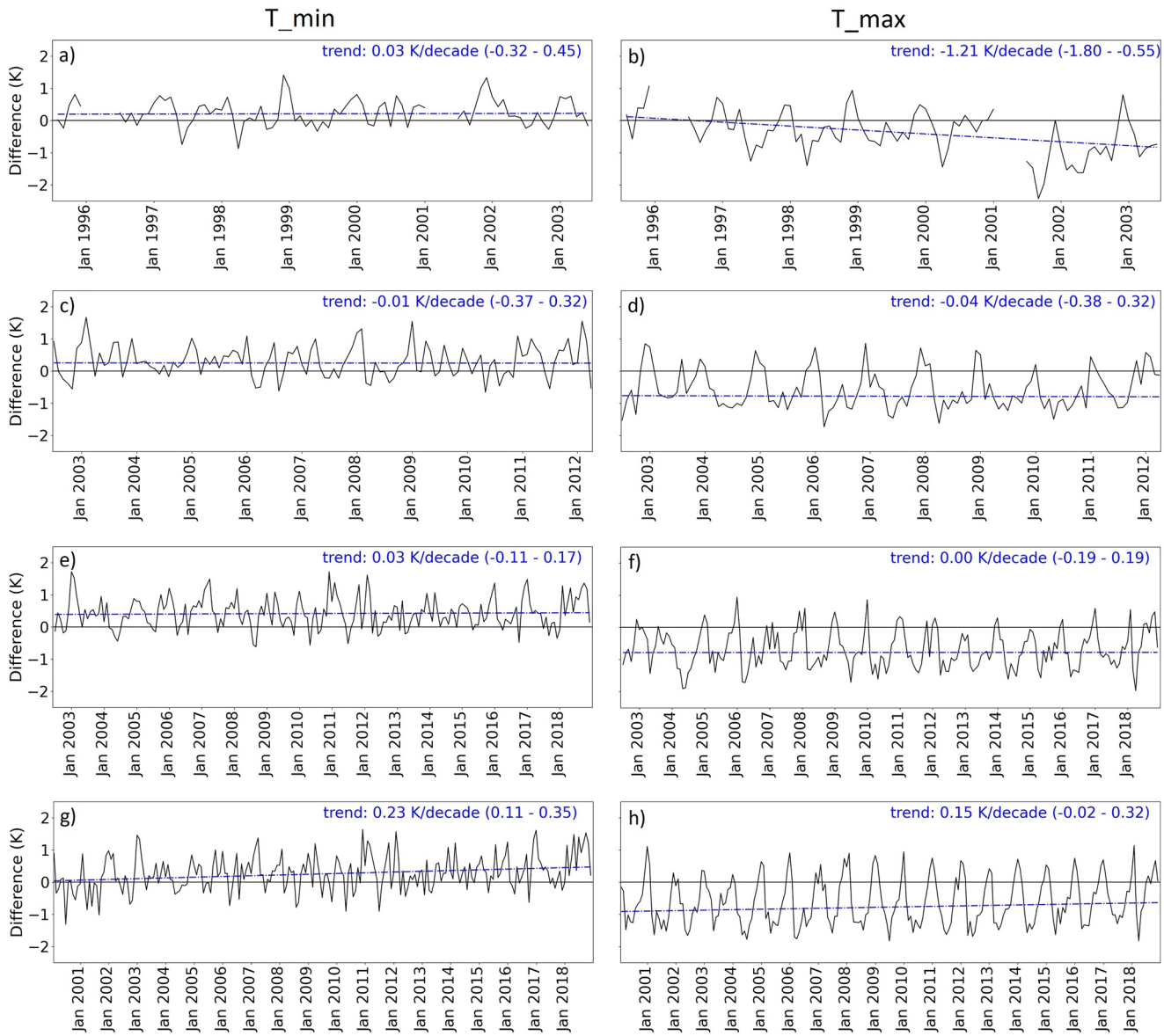

Figure 7. Land surface temperature (LST)-minus-2m air temperature (T2m) monthly anomaly time series for the (a) Along-Track Scanning Radiometer (ATSR)-2 T_{min}, (b) ATSR-2 T_{max}, (c) Advanced ATSR (AATSR) T_{min}, (d) AATSR T_{max}, (e) MODerate resolution Imaging Spectroradiometer (MODIS)/Aqua T_{min}, (f) MODIS/Aqua T_{max}, (g) MODIS/Terra T_{min}, and (h) MODIS/Terra T_{max} comparisons. The linear trend for each time series is shown in blue; the numerical value for this trend is also shown in each panel together with the 95% confidence interval on that trend.

particular, there is a negative trend for T2m for the T_{mean} comparison, although it should be noted that this is not statistically significant and is attributed to the small number of data points for T_{mean}, which requires both LST_{night} and LST_{day} to be available (there are 55 stations with T_{mean} data compared to 475 for T_{min} and T_{max} for AATSR). This sizable difference between trends calculated for MODIS/Aqua and AATSR is primarily due to the different time periods available for analysis. This is evident when looking at the statistically significant trends in T2m across all data sets, which differ considerably, ranging between 0.32 K/decade (MODIS/Terra period, T_{min}) and 1.02 K/decade (ATSR-2 period, T_{min}).

To illustrate this time-period effect, Table 6 shows the LST and T2m trends for both MODIS data sets but only for the MODIS/Aqua time period. A meaningful comparison has been performed by using only observations corresponding to stations that are available for both data sets and where observations are available for both LST and T2m. However, it should be noted this is not an exact like-for-like comparison as the LST-T2m observation

Table 5
Trend in Monthly Anomalies (K/Decade) Averaged Over all Stations for LST and T2m Over the Time Periods Specified for Each LST_cci Data Set

			MODIS/ Aqua (2002–2018)	MODIS/ Terra (2000–2018)	ATSR-2 (1995–2003)	AATSR (2002–2012)	Multisensor IR (1995–2012)	Multisensor MW (1996–2018)
T_{min}	*LST*	*t* (*K*/decade)	*0.64*	*0.57*	*0.96*	0.09	−0.41	*0.64*
		CI (*K*/decade)	0.24–1.05	0.24–0.92	0.01–1.96	−0.69–0.88	−0.83–0.01	0.42–0.85
		No. Obs.	1,896,413	2,219,831	102,545	311,755	598,276	1,862,226
	T2m	*t* (*K*/decade)	*0.52*	*0.32*	*1.02*	0.15	0.13	*0.41*
		CI (*K*/decade)	0.19–0.87	0.04–0.61	0.02–2.02	−0.51–0.84	−0.23–0.51	0.21–0.63
		No. Obs.	1,896,413	2,219,831	102,545	311,755	598,276	1,862,226
	T2mAll	*t* (*K*/decade)	*0.51*	0.27	*0.97*	0.07	0.28	*0.42*
		CI (*K*/decade)	0.16–0.83	−0.00–0.54	0.20–1.82	−0.62–0.69	−0.06–0.60	0.22–0.64
		No. Obs.	3,432,563	3,779,542	636,676	1,682,633	3,958,957	2,340,631
T_{max}	LST	*t* (*K*/decade)	*0.66*	*0.52*	−0.54	0.35	0.22	*0.69*
		CI (*K*/decade)	0.18–1.12	0.13–0.91	−1.65–0.71	−0.65–1.39	−0.27–0.70	0.42–0.95
		No. Obs.	1,639,676	1,738,545	98,369	278,089	455,171	1,866,153
	T2m	*t* (*K*/decade)	*0.59*	0.37	0.92	0.36	0.44	*0.63*
		CI (*K*/decade)	0.15–1.04	0.00–0.75	−0.25–2.13	−0.69–1.46	−0.14–1.03	0.36–0.92
		No. Obs.	1,639,676	1,738,545	98,369	278,089	455,171	1,866,153
	T2mAll	*t* (*K*/decade)	*0.58*	*0.34*	*1.38*	0.22	*0.52*	*0.60*
		CI (*K*/decade)	0.18–0.98	0.01–0.68	0.36–2.47	−0.75–1.12	0.09–0.96	0.33–0.89
		No. Obs.	3,433,456	3,780,312	636,619	1,682,774	3,959,369	2,340,922
T_{mean}	LST	*t* (*K*/decade)	*0.66*	*0.59*	−0.81	−0.14	0.38	*0.70*
		CI (*K*/decade)	0.23–1.11	0.23–0.97	−1.66–0.28	−0.89–0.64	−0.04–0.78	0.46–0.94
		No. Obs.	972,145	1,110,379	6,252	11,102	28,970	1,503,433
	T2m	*t* (*K*/decade)	*0.59*	*0.39*	0.51	−0.44	*0.59*	*0.55*
		CI (*K*/decade)	0.15–1.02	0.03–0.75	−0.38–1.49	−1.16–0.26	0.08–1.11	0.31–0.79
		No. Obs.	972,145	1,110,379	6,252	11,102	28,970	1,503,433
	T2mAll	*t* (*K*/decade)	*0.53*	*0.30*	*1.13*	0.15	*0.43*	*0.53*
		CI (*K*/decade)	0.16–0.93	0.00–0.60	0.20–2.14	−0.65–0.93	0.07–0.81	0.29–0.76
		No. Obs.	3,139,448	3,526,823	550,463	1,566,959	3,852,109	2,341,647

Note. The results for "T2m" are based on data where both LST and T2m data are available. The results for "T2mAll" use all available T2m observations; these results are discussed in Section 5. Trends that are statistically significant are shown in italics. "No. Obs." indicates the number of daily observations in each case.

matching has been carried out separately for each data set due to the different overpass times (and therefore differing data availability). (The requirement to use the same stations for both data sets also results in small differences in the MODIS/Aqua results between Tables 5 and 6.) The MODIS/Terra data are available for just over two additional years at the beginning of the record from February 2000 to July 2002 so this exercise essentially removes these additional two years of data from this data set. The results show that the T2m trends for this common time-period analysis are now almost identical (<0.03 K/decade difference) across both data sets. Prior to this the T2m trends were 0.52–0.59 K/decade and 0.32–0.39 K/decade for the MODIS/Aqua and MODIS/Terra time periods respectively. By contrast, the equivalent LST trends, also included in Table 6, still differ substantially. This difference in the LST trends is most likely due to the inhomogeneities in the MODIS/Terra LST data set observed in Section 4.2 and demonstrates why care should be taken to ensure only homogeneous data sets are used for trend analysis. Some small differences between the LST results for MODIS/Aqua and MODIS/Terra may also arise from inherent differences in the data sets, which includes small differences in the

Table 6
Trend (t, K/Decade) and Confidence Interval (CI) in Monthly Anomaly Data Averaged Over all Collocated Stations for LST and T2m for MODIS/Aqua and MODIS/Terra for the MODIS/Aqua Time Period (2002–2018)

			MODIS/Aqua	MODIS/Terra
T_{min}	LST	*t* (*K*/decade)	*0.65*	*0.82*
		CI (*K*/decade)	0.25–1.05	0.39–1.23
	T2m	*t* (*K*/decade)	*0.52*	*0.53*
		CI (*K*/decade)	0.19–0.88	0.19–0.88
T_{max}	LST	*t* (*K*/decade)	*0.67*	*0.80*
		CI (*K*/decade)	0.20–1.13	0.33–1.28
	T2m	*t* (*K*/decade)	*0.59*	*0.60*
		CI (*K*/decade)	0.16–1.04	0.14–1.07
T_{mean}	LST	*t* (*K*/decade)	*0.67*	*0.85*
		CI (*K*/decade)	0.23–1.12	0.40–1.13
	T2m	*t* (*K*/decade)	*0.59*	*0.62*
		CI (*K*/decade)	0.15–1.01	0.17–1.04

Note. The monthly averages are based on observations corresponding to stations that are available in both data sets and where observations are available for both LST and T2m. Trends in italics are statistically significant.

BT calibration and the cloud masking. However, it should be noted that the LST retrieval algorithms, the generation of coefficients and all other aspects of the data sets are identical so these data sets should be fundamentally the same.

The analysis presented here also demonstrates why it is preferable to use much longer data sets (e.g., >30 years) for calculating trends, despite using the MPW slopes method, which is more robust to outliers. In this case, removing the two additional years in the MODIS/Terra period has increased the overall trends in the T2m data by ∼0.2 K/decade. The need for longer data records for the detection of trends is also supported by the results for the AATSR data set, where no significant trend is detected despite the data set appearing to be quite stable over the 11-year record. However, it should be noted that even trends over longer periods are sensitive to what is happening in the climate over that period as climate change is not monotonic; interannual and decadal variability, as well as the specific start and end dates, can significantly affect trends regardless of the analysis period used.

Examination of the trends in LST for the other non-stable LST_cci data sets in Table 5 indicates the values can vary substantially. However, both the MODIS/Terra and multisensor MW data sets show some consistency between the trends for the different T2m comparisons in each case, with values of 0.64–0.70 K/decade and 0.52–0.59 K/decade, respectively. For the multisensor MW, all trends for LST and T2m are also significant and positive, which suggests some consistency between LST and T2m for this data set, despite its apparent instability identified in Section 4.2. Most trends for the other LST data sets are positive, but some are negative, which is not expected given all the T2m trends are positive. In some cases, the trends are quite different from the equivalent T2m trends. For example, for the ATSR-2 T_{max} comparison, the LST and T2m trends are −0.54 K/decade and 0.92 K/decade, respectively, although neither trend is statistically significant. These large differences and substantial variation in the calculated trends for these LST_cci data sets again shows the importance of ensuring homogeneous data sets are used for the analysis of trends, and that ideally this should also utilize multidecadal records.

Table 5 also presents some differences between significant trends in T_{min} and T_{max} that seem rather large. For example, for the multisensor MW period trends of 0.41 K/decade and 0.63 K/decade are obtained for T_{min} and T_{max}, respectively. These differences are attributed to natural variability in the calculated trends due to errors in the data, including missing data, and/or assumptions in the analysis, which is reflected in the confidence intervals. However, in all cases, the differences between the trends in T_{min} and T_{max} are small compared to the 95% confidence intervals on those trends (e.g., a difference of 0.22 K/decade compared with confidence interval width of 0.42 and 0.56 K/decade for T_{min} and T_{max}, respectively).

5. Discussion: The Effect of Using Only Clear-Sky Surface Temperatures in the Analysis of IR LSTs

The results presented in Section 4.2 show that the differences between the time series of monthly IR LST and T2m anomalies calculated in this study have a clear annual cycle. A particularly striking example can be seen in the MODIS/Terra T_{max} analysis shown in Figure 7, which shows that the annual cycle in the differences peaks in winter ($LST_{anom} > T2m_{anom}$) and troughs in summer ($LST_{anom} < T2m_{anom}$). However, this cycle is present in all the ATSR and MODIS T_{max} comparisons and to much lesser degree in the T_{min} comparisons. A similar annual cycle is not apparent in the equivalent multisensor MW analysis (Figure 7).

Further analysis of the data indicates this annual cycle results from the exclusion of some T2m observations during the process of temporally matching the daily LST and T2m data (described in Section 3). For the analysis of the IR LST data sets from MODIS and ATSR, the vast majority of these excluded T2m data correspond to cloudy days in the IR LST data, which can only be retrieved in cloud-free conditions. Figure 8 shows how the distribution of T_{max} and LST_{day} anomalies varies with the day of the year. The anomalies represent the full time

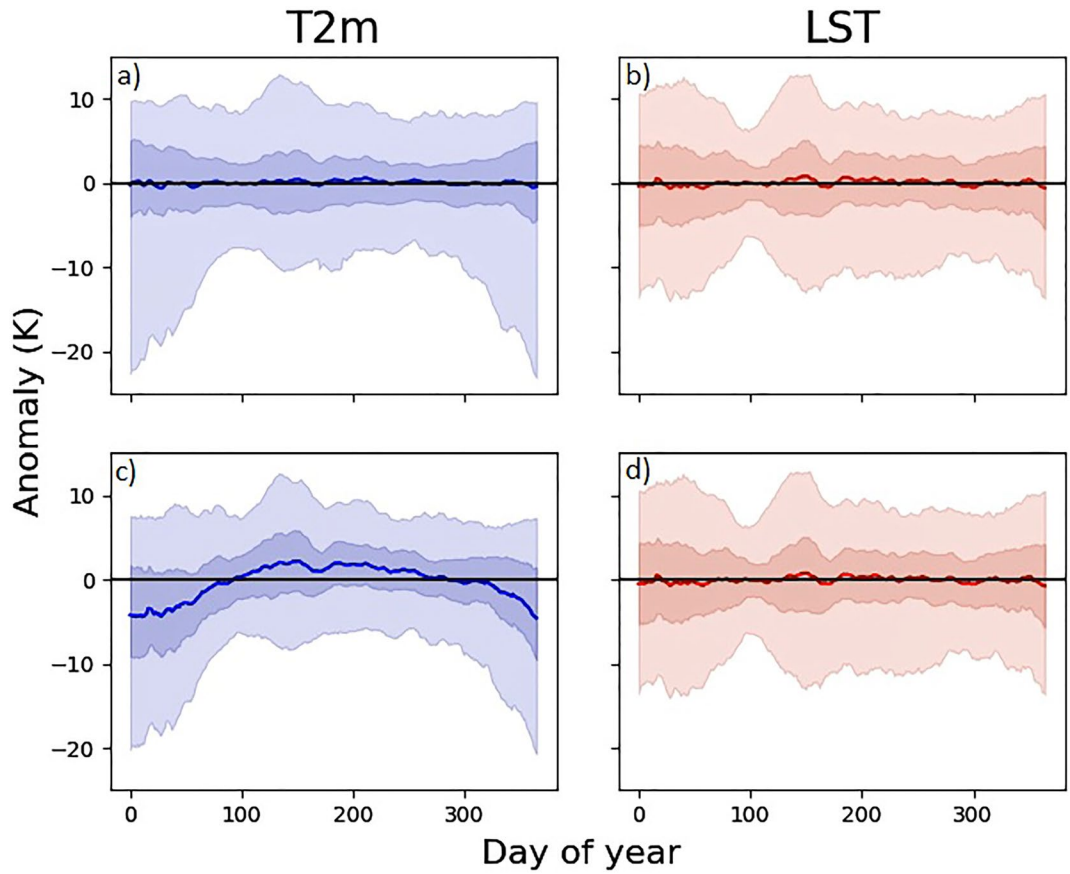

Figure 8. Percentiles of daily anomalies (K) for each day of the year averaged over five stations above 60°N for (a) T_{max} and (b) land surface temperature $(LST)_{day}$ before temporal matching between the 2m air temperature (T2m) and LST data is performed, and (c) T_{max} and (d) LST_{day} after temporal matching is performed. The solid line shows the mean, darker shading the inter-quartile range, and lighter shading the full range of data. The percentiles have been smoothed over a 21-day moving window using a convolution function to improve the visual clarity of the plot. Data are from the MODerate resolution Imaging Spectroradiometer/Terra comparison.

period of the MODIS/Terra record from five stations above 60°N (as an example). Anomalies are shown for each data set both before and after any data are excluded by temporal matching (i.e., requiring that both valid T_{max} and LST_{day} are present on the same day). As expected for the "before" analysis, the mean anomaly, shown by the solid line in the plots, is very close to zero throughout the year for both T2m and LST. There is a clear annual cycle in the T2m distribution with more extreme cold anomalies during winter and the anomalies during the middle of the year more symmetrically distributed about zero. For LST, the "before" pattern is more consistent throughout the year, although there is notable narrowing of the anomaly distribution in the spring. Examining the "after" analysis indicates that the distribution of LST anomalies is virtually unchanged. This is because the LST distribution is essentially always an "after" analysis, the cloudy data having been excluded from the start. This results in many more T2m data than LST being removed from the original pool of data in the "after" analysis. The consequence of this T2m data removal appears to cause an annual cycle in the mean T2m anomaly, whereby many warm anomalies have been removed in winter and cold anomalies have been removed in the summer. This pattern is expected as cloudy skies are typically associated with warmer days during the winter and colder days during the summer for this study area. The net effect of this T2m data removal therefore causes an annual cycle in the differenced anomaly time series, with a more positive LST-T2m anomaly difference in winter and a more negative LST-T2m difference in summer, exactly as observed in the analysis presented in Section 4.2.

This effect can be further verified by reproducing the anomaly difference time series without performing any temporal matching and simply using all available T2m in the calculation of the monthly T2m anomalies. This is shown in Figure 9 for MODIS/Aqua, where it can be seen that the annual cycle is completely removed through

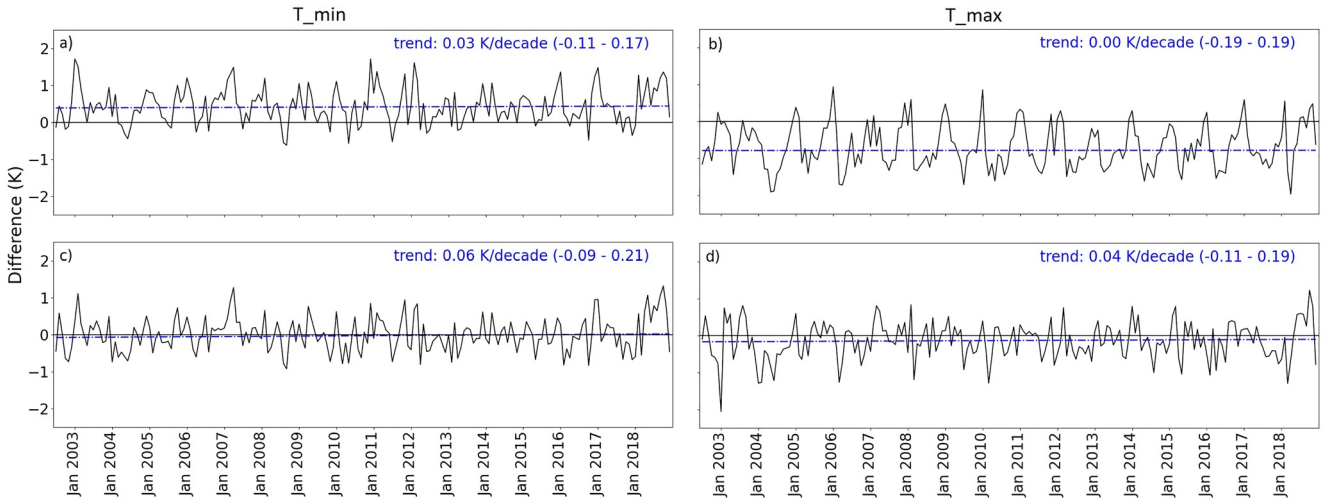

Figure 9. Time series of land surface temperature (LST) minus 2m air temperature (T2m) anomalies for the MODerate resolution Imaging Spectroradiometer/Aqua for the (a) T_{min} and (b) T_{max} comparison for temporally-matched LST/T2m data, and (c) T_{min} and (d) T_{max} comparison using all available T2m data in the monthly averages. Note that panels (a) and (b) are identical to panels (e) and (f) in Figure 7. The linear trend for each time series is shown in blue; the numerical value for this trend is also shown in each panel together with the 95% confidence interval on that trend.

this process. This also resembles the equivalent analysis performed by Good et al. (2017), who did not find any annual cycle in the time series of differences between an ATSR-2/AATSR CDR and the monthly, gridded T2m anomalies from the CRUTEM4 data set (Jones et al., 2012). This effect also causes the non-zero mean difference between the LST and T2m anomaly time series shown in Table 3 and Figure 7. Without the temporal matching, the time series shown in panels (c) and (d) of Figure 9 are now much closer to the zero difference line: mean LST-T2m differences are now 0.01 K (previously 0.43 K) and −0.15 K (previously −0.69 K) for the T_{min} and T_{max} comparison, respectively. For AATSR (not shown), the equivalent changes in mean differences using "matched" data and "all" data, respectively, are 0.30 to −0.02 K (T_{min} comparison) and −0.57 to −0.08 K (T_{max} comparison). (The remaining small non-zero differences are attributed to the monthly averaging process, whereas anomalies are calculated on a day-to-day basis.)

Having established there is a "clear-sky bias" effect on the comparison between IR LSTs and T2m observations, it is logical to establish whether this clear-sky bias affects the trends calculated in Section 4.3. Previous studies have shown that sampling only cloud-free conditions can affect trends in other geophysical variables (e.g., John et al., 2011). The impact of cloud-free sampling here is assessed by recalculating trends in T2m using all available or "all-sky" T2m observations, without any temporal matching to the IR LST observations. These results are shown in Table 5 in the sections labeled "T2mAll." Comparing these T2mAll trends with the original trends calculated from the temporally-matched T2m data (labeled ""T2m" in Table 5) shows that using all-sky T2m data results in a very small change in the trend, which reduces the trends by between 0.01 and 0.06 K/decade for the MODIS/Aqua period, with a similarly small change in the confidence intervals. Slightly larger changes in the trends are observed for the AATSR analysis, which are reduced by between 0.08 and 0.14 K/decade for T_{min} and T_{max} respectively. However, it should be remembered that none of the AATSR trends are statistically significant and that the period of analysis is very short for a climate trend analysis (<10 years for AATSR compared to ∼17 years for MODIS/Aqua) so these results should be interpreted with caution. Examining the significant trends in T2m and T2mAll for stations grouped by latitude (latitude <45 N, 45 ≤ latitude <52 N, latitude ≥52) for the MODIS/Aqua period suggests that the consistency between the T2m and T2mAll trends is not affected by varying station density and latitude. In conclusion, the results presented here do not offer any evidence that a clear-sky bias affects trends calculated using cloud-free IR observations.

6. Conclusions

This study assesses six new LST data sets available from ESA's LST_cci project based on data from the MODIS/Aqua (2002–2018), MODIS/Terra (2000–2018), ATSR-2 (1995–2003), AATSR (2002–2012), the combined ATSR-2-AATSR series (multisensor IR: 1995–2012) and combined SSM/I-SSMIS series (multisensor MW: 1996–2020). An initial comparison with T2m is performed using homogenized T2m from the EUSTACE project at a number of station locations across Europe. Co-locating satellite-observed LST data with the stations results in between 221 and 651 comparison locations for the six different LST_cci data sets. LST and T2m anomalies (LST vs T2m) are found to compare well with consistent slopes (m) and correlations (r). For comparisons with T_{max} and T_{mean} values of ∼0.8 and ∼0.9 K/K are obtained for r and m, respectively, while T_{min} comparison values are closer to ∼0.6 for both parameters. The T_{min} comparisons generally exhibit a higher variance than the T_{max} and T_{mean} comparisons. Slopes of <1 indicate that anomalies in LST are generally smaller in magnitude than those for T2m. These results are consistent with previous studies comparing LST and T2m anomalies. In addition, a general increase in r with latitude is observed for the IR data sets, but not for the multisensor MW data set, which appear to be less variable across different latitudes. Furthermore, the distribution of MW anomalies appears to match that for T2m more closely than for the IR data sets. This suggests that MW LST anomalies may provide a better proxy for T2m anomalies than those from IR LST data sets. This is likely to be because both the MW LST and T2m anomalies are all-sky, whereas the IR LST data are clear-sky only. Furthermore, the IR LSTs may be affected by cloud contamination resulting in erroneous LST values.

Having established a strong relationship between LST and T2m anomalies, the stability of the LST data sets is assessed by calculating the trend in the time series of monthly LST-minus-T2m anomalies averaged over the whole study region. As the T2m data are homogenised, a statistically significant trend in this LST-T2m difference is assumed to indicate that the LST data set is inhomogeneous. This is found to be the case for MODIS/Terra ($LST_{night,}$ LST_{mean}), ATSR-2 (LST_{day}, LST_{mean}), the multisensor IR data set ($LST_{night)}$ and the multisensor MW data set (LST_{desc} and LST_{mean}). These data sets are therefore considered inhomogeneous and unsuitable for use for robust climate trend analysis. However, the time series for the individual sensors comprising both multisensor data sets appear to be stable, suggesting the discontinuities in these records are due to the change in sensor. Overall, only the single-sensor MODIS/Aqua and AATSR data sets appear free from non-climatic discontinuities based on the approach to assess the stability of the LST_cci data sets that has been adopted in this study.

Trends in LST and equivalent T2m data are calculated using the median of pairwise slopes method, which is considered robust to outliers (Sen, 1968). For MODIS/Aqua, significant trends in LST of 0.64–0.66 K/decade are obtained, which compare well with the equivalent T2m trends of 0.52–0.59 K/decade. The trends for AATSR are found to be statistically insignificant, which is due to the specific years available for the analysis. This is demonstrated through a comparison of trends for T2m for the MODIS/Aqua and MODIS/Terra time periods, where the addition of just two additional years of data at the beginning of the MODIS/Terra period (2000–2001) is found to reduce the trends in T2m by ∼0.2 K/decade. This demonstrates the importance of using a long and homogeneous time series of data to calculate climate trends. Given satellites generally have an expected lifespan of a few years to a decade or so at the most, it requires the use of multiple sensors to generate data sets that are several decades in length. This study demonstrates the critical importance to ensure that transitions between sensors are homogenized to ensure these multisensor data sets are free from non-climatic discontinuities that affect the calculated trends.

Finally, this study has also demonstrated that the way in which LST and T2m data are compared can affect results. As the overall objective of this study is to assess whether LST anomalies are consistent with the T2m anomalies that are used routinely for climate monitoring, the LST anomalies are compared with T2m anomalies referenced to an all-sky T2m climatology. However, as the IR satellite LST data are are only available for cloud-free conditions it is found that excluding cloudy T2m observations from the comparison (whilst retaining these observations in the data used to calculate the T2m baseline climatology) introduces a “clear-sky bias” effect into the comparison. This results in an annual cycle and non-zero mean difference in the LST-minus-T2m anomaly time series, which are both removed when the T2m anomaly time series is regenerated from “all-sky” T2m observations. Nevertheless, when the trends in T2m are recalculated from the “all-sky” T2m data, they are still statistically significant and almost identical (0.01–0.06 K/decade difference) for the MODIS/Aqua analysis. A more substantial change in the T2m trends is found for the AATSR analysis when all-sky T2m data are used but these trends are all found to be statistically insignificant. Therefore, it is concluded that the results presented

in this study do not offer any evidence that a clear-sky bias affects trends calculated using cloud-free IR observations. The results of this study also suggest that T2mAll may be a better choice of reference data set for this type of analysis compared with T2m, even when using IR LST data sets. However, further analysis is required to verify this definitively, and on a global scale, and therefore IR LST data users are strongly encouraged to test for the presence and effect of clear-sky biases in their own studies.

All the results of this study summarized above suggest that satellite LST data have great potential to be used together with T2m data to assess warming trends over land and for other climate applications. Observations of land T2m are limited to the locations of weather stations, which currently leaves large parts of the Earth poorly observed (e.g., parts of Africa, Antarctica, Tibetan Plateau). Satellite LSTs can therefore be used to provide additional information where it is unavailable in situ, and independent analysis where in situ data are plentiful. However, creating a long-term, climate quality LST data sets poses significant challenges. Differences between LST data sets that provide observations with different coverage or at different times of the day can be substantial. The LST_cci project aims to overcome some of these challenges, by providing users with different LST data sets in the same format, and processed in a consistent way, for example, by using common LST retrieval algorithms and approaches to cloud screening. Significant progress toward these goals has already been achieved in Phase I of the LST_cci project (2018–2022). Phase II of the LST_cci project, which began in 2022, will improve these LST data sets further, in addition to providing new data sets that include the Advanced Very High Resolution Radiometers (AVHRR), the Visible Infrared Imaging Radiometer Suite (VIIRS) and Landsat. Other international initiatives are also making substantial progress toward creating long-term, stable LST records that can be used in climate applications. For example, in 2022 the Satellite Application Facility on Climate Monitoring (CM SAF) will release version 2.0 of its long-term LST data set based on observations from the Meteosat satellite series (Duguay-Tetzlaff et al., 2015). This data set will provide consistent and homogenized LSTs for Europe, Africa and part of South America for 1983 to 2020. The LST_cci project compliments these other initiatives, which together will enable satellite LST to play a more prominent role in understanding and monitoring Earth surface temperatures. Furthermore information about the LST_cci project, including all the project documentation, can be found on the project website https://climate.esa.int/en/projects/land-surface-temperature/.

Data Availability Statement

The satellite LST data used in the study are beta versions provided by the LST_cci project. Official versions of these datasets are freely available via the ESA Open Data Portal (https://climate.esa.int/en/odp/#/dashboard). The homogenized station T2m data provided by the EUSTACE project are freely available at https://www.ecad.eu/dailydata/predefinedseries.php for non-commercial purposes with further information provided at https://catalogue.ceda.ac.uk/uuid/81784e3642bd465aa69c7fd40ffe1b1b and in Squintu, Van Der Schrier, Brugnara, et al. (2019; doi:10.1002/joc.5874).

Acknowledgments
Elizabeth Good and Freya Aldred were supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra (GA01101). Darren Ghent, Karen Veal and Carlos Jimenez have received funding from the European Space Agency (ESA) within the framework of the Land Surface Temperature project under the Climate Change Initiative (LST_cci), contract number 4000123553/18/I-NB. The authors are grateful to the ESA for creating the CCI program which has strengthened the consistency of the many research communities related to developing, processing, qualifying, and using satellite CDRs. The authors are also grateful to Nick Rayner (Met Office Hadley Centre) and Anke Duguay-Tetzlaff (the reviewer appointed by the Earth and Space Science journal) for their extremely helpful and insightful reviews of this manuscript.

References

Aguilar-Lome, J., Espinoza-Villar, R., Espinoza, J.-C., Rojas-Acuña, J., Willems, B. L., & Leyva-Molina, W.-M. (2019). Elevation-dependent warming of land surface temperatures in the Andes assessed using MODIS LST time series (2000–2017). *International Journal of Applied Earth Observation and Geoinformation*, *77*, 119–128. https://doi.org/10.1016/j.jag.2018.12.013

Bechtel, B., Demuzere, M., Mills, G., Zhan, W., Sismanidis, P., Small, C., & Voogt, J. (2019). SUHI analysis using local climate zones—A comparison of 50 cities. *Urban Climate*, *28*, 100451. https://doi.org/10.1016/j.uclim.2019.01.005

Begueria, S., Tomas-Burguera, M., Serrano-Notivoli, R., Pena-Angulo, D., Vicente-Serrano, S. M., & Gonzalez-Hidalgo, J.-C. (2019). Gap filling of monthly temperature data and its effect on climatic variability and trends. *Americal Meteorological Society Journal of Climate*, *32*(22), 7797–7821. https://doi.org/10.1175/JCLI-D-19-0244.1

Bulgin, C. E., Embury, O., Corlett, G., & Merchant, C. J. (2016). Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the along-track scanning radiometer instruments. *Remote Sensing of Environment*, *178*, 213–222. https://doi.org/10.1016/j.rse.2016.02.022

Chen, F., Liu, Y., Liu, Q., & Qin, F. (2015). A statistical method based on remote sensing for the estimation of air temperature in China. *International Journal of Climatology*, *35*, 2131–2143. https://doi.org/10.1002/JOC.4113

Chen, Y., Liang, S., Ma, H., Li, B., He, T., & Wang, Q. (2021). An all-sky 1 km daily land surface air temperature product over mainland China for 2003–2019 from MODIS and ancillary data. *Earth System Science Data*, *13*(8), 4241–4261. https://doi.org/10.5194/essd-13-4241-2021

Cheval, S., Dumitrescu, A., & Amihaesei, V.-A. (2020). Exploratory analysis of urban climate using a gap-filled landsat 8 land surface temperature data set. *Sensors*, *20*(18), 5336. https://doi.org/10.3390/s20185336

Cheval, S., Dumitrescu, A., Irașoc, A., Paraschiva, M.-G., Perry, M., & Ghent, D. (2022). MODIS-based climatology of the surface urban heat island at country scale (Romania) (Vol. *41*, pp. 101056). https://doi.org/10.1016/j.uclim.2021.101056

Dodd, E., Ermida, S., Jimenez, C., & Ghent, D. (2020). *LST_cci product user guide v1.2*. European Space Agency. Retrieved from https://climate.esa.int/en/projects/land-surface-temperature/key-documents/

Dodd, E., Ghent, D., Jimenez, C., & Ermida, S. (2019). *LST_cci algorithm theoretical basis document v1.2*. European Space Agency. Retrieved from https://climate.esa.int/en/projects/land-surface-temperature/key-documents/

Dousset, B., Gourmelon, F., Laaidi, K., Zeghnoun, K., Giraudet, E., Bretin, P., et al. (2011). Satellite monitoring of summer heat waves in the Paris metropolitan area. *International Journal of Climatology*, *31*, 313–323. https://doi.org/10.1002/joc.2222

Duguay-Tetzlaff, A., Bento, V. A., Göttsche, F. M., Stöckli, R., Martins, J. P. A., Trigo, I., et al. (2015). Meteosat land surface temperature climate data record: Achievable accuracy and potential uncertainties. *Remote Sensing*, *7*, 13139–13156. https://doi.org/10.3390/rs71013139

Fenning, K., Schroder, M., Andersson, A., & Hollmann, R. (2020). A fundamental climate data record of SSMR, SSMI/S, and SSMIS brightness temperatures. *Earth System Science Data*, *12*(1), 647–681. https://doi.org/10.5194/essd-12-647-2020

Fernandes, R., & Leblanc, S. G. (2005). Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sensing of Environment*, *95*(3), 303–316. https://doi.org/10.1016/j.rse.2005.01.005

Folwell, S., Harris, P. P., & Taylor, C. M. (2016). Large-scale surface responses during European dry spells diagnosed from land surface temperature. *Journal of Hydrometeorology*, *17*(3), 975–993. https://doi.org/10.1175/JHM-D-15-0064.1

Freitas, S. C., Trigo, I. F., Bioucas-Dias, J. M., & Goettsche, F.-M. (2010). Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 523–534. https://doi.org/10.1109/TGRS.2009.2027697

GCOS. (2016). Implementation plan 2016. Retrieved from http://www.wmo.int/pages/prog/gcos/

Ghent, D. (2012). *Land surface temperature validation and algorithm verification*. European Space Agency. Retrieved from https://earth.esa.int/eogateway/documents/20142/37627/LST%20Validation%20and%20Algorithm%20Verification?text=Land+surface+temperature+validation+and+algorithm+verification

Ghent, D., Veal, K., Trent, T., Dodd, E., Sembhi, H., & Remedios, J. (2019). A new approach to defining uncertainties for MODIS land surface temperature. *Remote Sensing*, *11*(9), 1021. https://doi.org/10.3390/rs11091021

Ghent, D. J., Corlett, G. K., Gottsche, F.-M., & Remedios, J. J. (2017). Global land surface temperature from the along-track scanning radiometers. *Journal of Geophysical Research: Atmospheres*, *122*(22), 12167–12193. https://doi.org/10.1002/2017JD027161

Good, E. (2015). Daily minimum and maximum surface air temperatures from geostationary satellite data. *Journal of Geophysical Research: Atmospheres*, *120*(6), 2306–2324. https://doi.org/10.1002/2014JD022438

Good, E. J. (2016). An in situ-based analysis of the relationship between land surface "skin" and screen level air temperatures. *Journal of Geophysical Research: Atmospheres*, *121*(15), 8801–8819. https://doi.org/10.1002/2016JD025318

Good, E. J. (2016). Satellite observations of surface temperature during the march 2015 total solar eclipse. *Philosophical Transactions of the Royal Society A*, *374*, 20150219. https://doi.org/10.1098/rsta.2015.0219

Good, E. J., Ghent, D. J., Bulgin, C. E., & Remedios, J. J. (2017). A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series. *Journal of Geophysical Research: Atmospheres*, *122*(17), 9185–9210. https://doi.org/10.1002/2017JD026880

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Hulley, G. C., Malakar, N. K., Islam, T., & Freepartner, R. J. (2018). NASA's MODIS and VIIRS land surface temperature and emissivity products: A long-term and consistent Earth system data record. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(2), 522–535. https://doi.org/10.1109/JSTARS.2017.2779330

IPCC. (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. Retrieved from https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/

Jiménez-Muñoz, J. C., Sobrino, J. A., Mattar, C., & Malhi, Y. (2013). Spatial and temporal patterns of the recent warming of the Amazon forest. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5204–5215. https://doi.org/10.1002/jgrd.50456

John, V. O., Holl, G., Allan, R. P., Buehler, S. A., Parker, D. E., & Soden, B. J. (2011). Clear-sky biases in satellite infrared estimates of upper tropospheric humidity and its trends. *Journal of Geophysical Research*, *116*(D14). https://doi.org/10.1029/2010JD015355

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*, D05127. https://doi.org/10.1029/2011JD017139

Jones, P. D., Wigley, T. M. L., & Kelly, P. M. (1982). Variations in surface air temperatures: Part 1. Northern hemisphere, 1881–1980. *Monthly Weather Review*, *110*(2), 59–70. https://doi.org/10.1175/1520-0493(1982)110<0059:visatp>2.0.co;2

Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., et al. (2010). Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *Journal of Climate*, *23*(3), 618–633. https://doi.org/10.1175/2009JCLI2900.1

Khorchani, M., Vicente-Serrano, S. M., Azorin-Molina, C., Garcia, M., Martin-Hernandez, N., Peña-Gallardo, M., et al. (2018). Trends in LST over the peninsular Spain as derived from the AVHRR imagery data. *Global and Planetary Change*, *166*, 75–93. https://doi.org/10.1016/j.gloplacha.2018.04.006

Kilibarda, M., Hengl, T., Heuvelink, G. B. M., Gräler, B., Pebesma, E., Perčec Tadić, M., & Bajat, B. (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, *119*(5), 2294–2313. https://doi.org/10.1002/2013JD020803

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology*, *22*(12), 1441–1453. https://doi.org/10.1002/joc.773

Klok, E. J., & Klein Tank, A. M. G. (2008). Updated and extended European dataset of daily climate observations. *International Journal of Climatology*, *29*, 1182–1191. https://doi.org/10.1002/joc.1779

Koch, J., Siemann, A., Stisen, S., & Sheffield, J. (2016). Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5430–5452. https://doi.org/10.1002/2015JD024482

Ma, J., Zhou, J., Göttsche, F.-M., Liang, S., Wang, S., & Li, M. (2020). A global long-term (1981–2000) land surface temperature product for NOAA AVHRR. *Earth System Science Data*, *12*(4), 3247–3268. https://doi.org/10.5194/essd-12-3247-2020

Mühlbauer, S., Costa, A. C., & Caetano, M. (2016). A spatiotemporal analysis of droughts and the influence of North Atlantic Oscillation in the Iberian Peninsula based on MODIS imagery. *Theoretical and Applied Climatology*, *124*(3), 703–721. https://doi.org/10.1007/s00704-015-1451-9

Perry, M., Ghent, D. J., Jimenez, C., Dodd, E. M., Ermide, S. L., Trigo, I. F., & Veal, K. L. (2020). Multisensor thermal infrared and microwave land surface temperature algorithm intercomparison. *Remote Sensing*, *12*(24), 4164. https://doi.org/10.3390/rs12244164

Prata, F. (2002). *Land surface temperature measurement from space: AATSR algorithm theoretical basis document*. CSIRO Atmospheric Research.

Prigent, C., Jimanez, C., & Aires, F. (2016). Toward "all weather", long record, and real-time land surface temperature retrievals from microwave satellite observations. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5699–5717. https://doi.org/10.1002/2015JD024402

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., et al. (2020). The EUSTACE project: Delivering global, daily information on surface air temperature. *Bulletin of the American Meteorological Society*, *101*(11), E1924–E1947. https://doi.org/10.1175/BAMS-D-19-0095.1

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sobrino, J. A., Julien, Y., & García-Monteiro, S. (2020). Surface temperature of the planet Earth from satellite data. *Remote Sensing*, *12*, 218. https://doi.org/10.3390/rs12020218

Sohrabinia, M., Zawar-Reza, P., & Rack, W. (2014). Spatio-temporal analysis of the relationship between LST from MODIS and air temperature in New Zealand. *Theoretical and Applied Climatology*, *119*, 567–583. https://doi.org/10.1007/s00704-014-1106-2

Squintu, A. A., Van Der Schrier, G., Brugnara, Y., & Klein Tank, A. (2019). Homogenization of daily temperature series in the European climate assessment and dataset. *International Journal of Climatology*, *39*(3), 1243–1261. https://doi.org/10.1002/joc.5874

Squintu, A. A., Van Der Schrier, G., & Van Den Basselaar, E. J. (2019). *Eustace/ECA&D: European land station daily air temperature measurements, homogenised*. European Climate Assessment and Dataset., Retrieved from https://catalogue.ceda.ac.uk/uuid/81784e3642bd465aa69c7fd40ffe1b1b

Stooksbury, D. E., Idso, C. D., & Hubbard, K. G. (1999). The effects of data gaps on the calculated monthly mean maximum and minimum temperatures int he continental United States: A spatial and temporal study. *AMS Journal of Climate*, *12*(5), 1524–1533. https://doi.org/10.1175/1520-0442(1999)012<1524:teodgo>2.0.co;2

Sun, X., Ren, G., You, Q., Ren, Y., Xu, W., Xue, X., et al. (2019). Global diurnal temperature range (DTR) changes since 1901. *Climate Dynamics*, *52*, 3343–3356. https://doi.org/10.1007/s00382-018-4329-6

Thorne, P. W., Menne, M. J., Williams, C. N., Rennie, J. J., Lawrimore, J. H., Vose, R. S., et al. (2016). Reassessing changes in diurnal temperature range: A new data set and characterization of data biases. *Journal of Geophysical Research: Atmospheres*, *121*, 5115–5137. https://doi.org/10.1002/2015JD024583

Zhou, C., & Wang, K. (2016). Land surface temperature over global deserts: Means, variability, and trends. *Journal of Geophysical Research: Atmospheres*, *121*(24), 14344–14357. https://doi.org/10.1002/2016JD025410