Product User Guide and quick start guide

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640171

Version number	Date updated	Changes made	By whom
Draft v0.8	19-03-19	Draft for first release	A. Waterfall
Draft v1.0	08-04-2019	Updated with remaining EUSTACE products	A. Waterfall
Version 1.0	31/05/2019	Released version	A. Waterfall

Draft, v1.0

Table of Contents

1 Introduction

1.1 Purpose of this document

This **Product User Guide** (PUG) is meant as a practical guide for the data products of the EUSTACE project (see section 1.2) and to facilitate (potential) users in their exploitation of the datasets. It contains:

- descriptions of the datasets
- instructions on how to access and use the data,
- advice on what can and cannot (or should not) be done with the data,
- example use cases,
- links to useful tools, etc
- examples of how to visualize the data, and process or extract data

For an overview of the methods used to produce the datasets, the interested user is referred to the scientific user guide (SUG) (under development).

This document is structured in such a way that you can easily navigate through it to the parts that are of interest for your specific use case, without having to scroll through the whole document. For this purpose in many places you will find links to go directly to the table of contents or to related content in other parts of the document or even to the scientific user guide (SUG). The links in the table of contents are clickable and bring you directly to the chapter or section of interest. At the bottom of each page you can find a link to take you back to the Table of contents.

A Quick Start Guide is provided in Chapter 3 of this Product User Guide. This is intended to act as a quick reference guide to where to find all the information required to get started with access and making use of the data. Experienced users of climate datasets in NetCDF format may only need this and the specific relevant product description in Chapter 4. However, we hope that this user guide also helps less experienced users and even those with no experience at all with the NetCDF-format to profit from the wealth of information in the EUSTACE datasets. To this end we have added example use cases (Chapter 5) and useful information on tools for extracting and visualising the data (Annex 2 and the Step-by-step guide for Climate4Impact).

1.2 The EUSTACE project

1.2.1 Aim

EUSTACE (EU Surface Temperature for All Corners of Earth) produced publicly available daily estimates of surface air temperature since 1850 across the globe with consistent uncertainty estimates for the first time by combining surface and satellite data using novel statistical techniques. It also produced many other useful products along the way towards developing the global analysis and these are described here.

1.2.2 Why EUSTACE?

Day-to-day variations in surface air temperature affect society in many ways:

- Health and well-being: they can cause cold stress or heat stress.
- Food security: there is a link between surface air temperature and crop growth and animal health.
- Energy: they influence the demand for heating or cooling; high temperatures compromise the efficiency of solar panels.
- Commerce: they affect the sales of a large variety of products.
- Tourism: surface air temperature affects the attractiveness of a region, and the risk of bushfires.
- Infrastructure: extremes affect the functioning of bridges, railways.

However, daily surface air temperature measurements are not available everywhere. Satellite data can be used to estimate temperatures at locations where no ground (or in situ) observations are available. However, satellites measure surface skin temperatures and not surface air temperatures. To estimate air temperatures at locations where no air temperatures are measured, we need an understanding of the relationships between traditional (land and marine) surface air temperature measurements and satellite measurements, i.e. land surface temperature, ice surface temperature, sea surface temperature and lake surface water temperature. These relationships can be derived either empirically, or with the help of physical understanding.

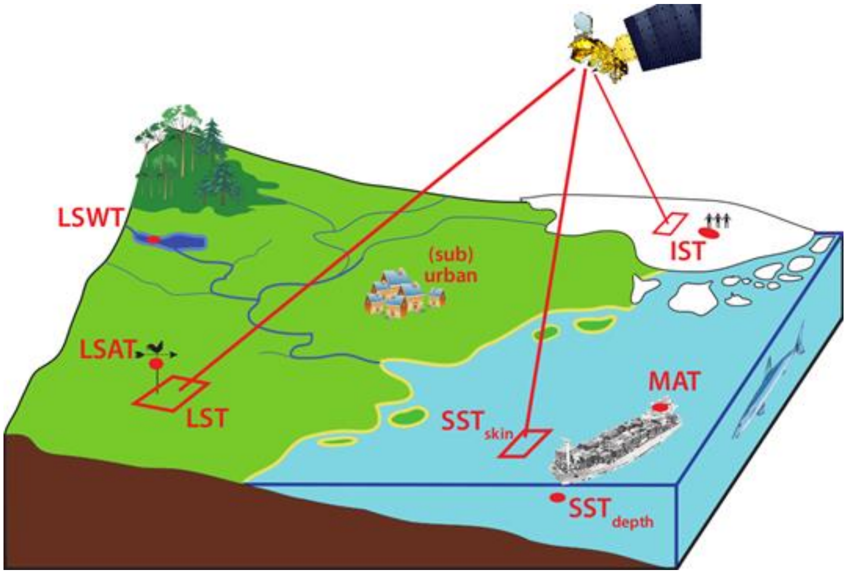


Figure 1-1: Different surface temperature measurements used in EUSTACE. Skin temperature measurements from satellite are: land surface temperature (LST); sea surface temperature (SST); ice surface temperature (IST); and lake surface water temperature (LSWT). Near-surface air temperature measurements, measured in situ over land and ocean are: land surface air temperature (LSAT) and marine air temperature (MAT). From Merchant et al., 2013 community paper and roadmap: http://www.geosci-instrum-method-data-syst.net/2/305/2013/gi-2-305-2013.html

1.2.3 How does EUSTACE achieve this?

The EUSTACE project used new statistical techniques to provide information on higher spatial and temporal scales than have been available before, making optimum use of the information in data-rich eras. The final and intermediate products from EUSTACE below could be used in many applications (Figure 1-2):

- Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015 (section 4.2)
- European land station daily air temperature measurements, homogenised (section 4)
- Gridded European surface air temperature based on homogenised land station records since 1950: dataset of daily air temperatures from 1950 onwards (E-OBS, by interpolation; Section 4.4)
- Global clear-sky land, ice, and sea surface temperatures from satellites with estimates of uncertainty components (section 4.1)
- Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016 (section 4.5)
- Information on the relationship between skin and air temperature over different domains and different seasons (see the Scientific User Guide)
- Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015 (Section 4.6)

- Coincident daily air temperature estimates and reference measurements, for validation, 1850-2015 (Section 4.7; Validation matchup database)

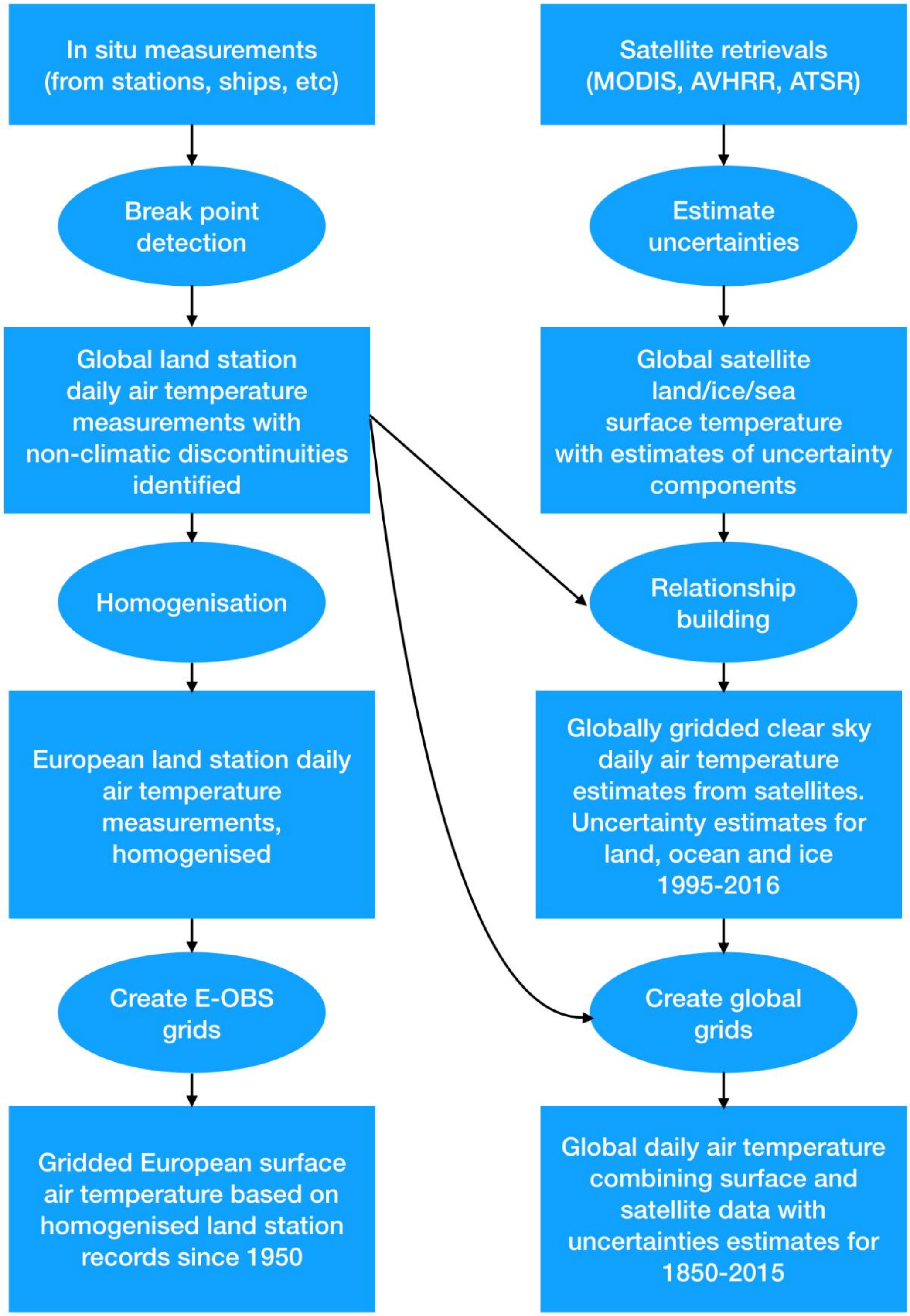


Figure 1-2: Links between the different surface temperature measurements and EUSTACE products.

The various steps in the EUSTACE processing chain are described briefly below and shown in Figure 1-2:

1. Break-point detection: daily series of measurements of air temperature (made *in situ* over land at weather stations) have to be checked for discontinuities in time due to changes of instruments, re-locations of stations, etc. (non-climatic factors). Identification of these discontinuities or break points is needed to make spatially and temporally consistent estimates of air temperatures (see section 4.2). The data are also quality controlled and new information on the reliability of the break-point detection is included.
2. Homogenisation: With the help of the knowledge of break points, the air temperature time series can be homogenized. This means that the original time series are adjusted in such a way that the effect of the non-climatic factors is removed. This can result in an increase or decrease of observed trends in temperature. This homogenization has been performed directly for the European station air temperature time series (see Section 4.3), which is new in a pan-European data set. For the rest of the globe this is included in the creation of the global grids in step 6.
3. Create E-OBS grids: the homogenised station air temperature time series for Europe are interpolated to create a gridded data set for Europe, the homogenised E-OBS dataset (see Section 4.4)
4. Estimate uncertainties of the satellite skin temperatures: satellite skin temperatures (see section 2.1.1) contain various types of uncertainties, which need to be propagated through the air temperature estimation in step 5. In this step the uncertainties are estimated for all surfaces in a consistent way, breaking down those uncertainties into components arising from errors with different structures (see section 4.1). This consistency across surfaces is new.
5. Relationship building: functions are determined with which the air temperature can be estimated from the satellite skin temperature over different surfaces and during the various periods of the year. This is done with the help of a selection of the station air temperature data and the satellite skin temperatures. Subsequently for all locations and days where satellite skin temperatures are available the air temperatures are estimated (see section 4.5)
6. Create global daily grids: in this last step statistical methods are developed to fill in the gaps for the days and locations where no satellite skin temperatures nor direct air temperature measurements were available (partly due to clouds in the case of the satellite data) and to estimate air temperatures more completely before the time period where satellite data are available (see section 4.6). This combination of satellite data and measurements made *in situ,* in a daily analysis over such a long period is new. For validation of this final dataset, an independent match-up data base is used (see section 4.7).

More information on EUSTACE products and methodologies can be found in this Product User guide in the indicated sections and in the Scientific User guide.

2 Product Overview

In Section 1.2.3 a list of EUSTACE products was given. In this chapter a short overview is given of the data in these EUSTACE products. Section 2.1 describes the types of temperatures that are provided in the EUSTACE datasets, and also includes some definitions of other variables; some terms are used in different ways in different projects or situations, and so to avoid confusion we explain which definitions we use here (see also Glossary of terms and acronyms). In Section 2.2 the datasets are described with references. The methods used are described in the Scientific User Guide. Along with these datasets, EUSTACE also produced other non-data products, such as the relationships between air and skin temperatures. These are described in Section 2.3.

2.1 Available temperature variables and related variables in the EUSTACE datasets

2.1.1 Temperature variables

Satellites measure skin temperatures. This is the temperature of the surface of the earth which can be estimated (or "retrieved") from the radiation given off by Earth and seen from space. This can be the temperature of the top of the canopy of a forest or the soil surface in an area with bare soils. At meteorological measuring stations, air temperature at a height of about 2 meters (also called surface air temperature), is measured by thermometers. The ultimate aim of EUSTACE is to construct a global data set of daily surface air temperature, using satellite skin temperatures and surface air temperatures from measuring stations.

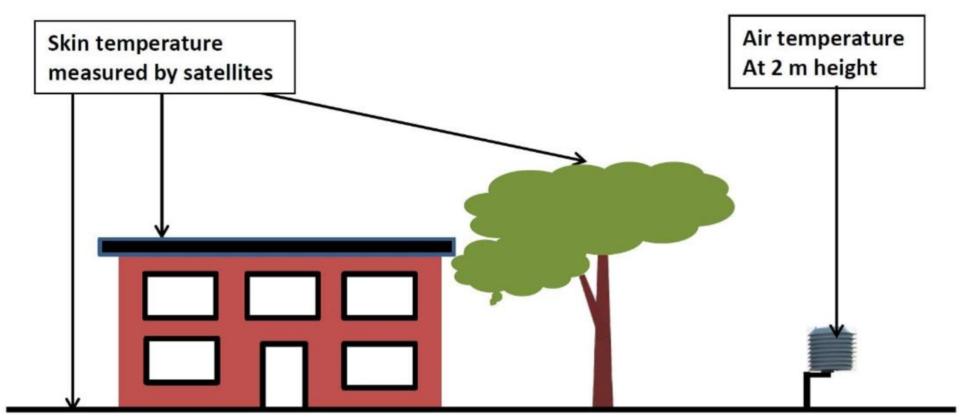


Figure 2-1: Schematic presentation of the difference between air temperature and skin temperature.

In EUSTACE some datasets mentioned below (under "Satellite skin temperature retrievals") contain surface skin temperatures. Abbreviations used for skin temperatures are (see Figure 1-1):

- LST = Land Surface (skin) Temperature
- IST = Ice Surface (skin) Temperature
- LSWT = Lake Surface (skin) Water Temperature
- SST_{skin} and SST_{depth} = Sea Surface (skin) Temperature[1] and Sea Surface Temperature at a depth of about 20 cm

Skin temperatures always represent an area average to some extent and are not point data, unlike measurements made *in situ* at weather stations or by ships.

The other datasets mentioned in Section 2.2 contain daily surface air temperature information. These use the following definitions:

- minimum daily air temperature: lowest temperature measured during a day, often measured in the early morning just before sunrise.
- maximum daily air temperature: highest temperature measured during a day, often measured in the early afternoon.
- mean daily air temperature: this is defined in this project as the average of the minimum and maximum daily temperature (over land) for the estimates in the final dataset based on satellite and station data. In station data files often the average temperature based on hourly temperature data is presented. The average and mean temperatures may differ.

Minimum and maximum temperatures are only provided over land, and not for sea/ocean, large lakes and ice.

Abbreviations used for air temperatures are (see Figure 1-1):

- LSAT = Land Surface Air Temperature
- MAT = Marine Air Temperature

Air temperatures in EUSTACE products can be either point data (from weather stations) or area average data (estimated from satellite data or using statistical methods). The spatial resolution of the final global gridded air temperature dataset based on satellite data and station data is 0.25 degrees in latitude and longitude.

2.1.2 Related variables

2.1.2.1 Uncertainty

The terms 'error' and 'uncertainty' are often confused. Careful usage (following international standards; VIM, 2012) brings clarity to thinking about uncertainty information. Terms with precise definitions that need careful usage appear in italic in the next paragraph.

---

[1] In practice, here we use satellite sea surface temperature estimated at a depth of 20cm below the surface, rather than the skin temperature. We do this to provide consistency with the long record of sea surface temperatures made *in situ*, which we use to understand the relationship between surface temperature and air temperature over the oceans.

A *measured value* results from *measurement* of a target quantity, called the *measurand*. It is only an estimate of the measurand, because various effects introduce errors into the process of measurement. These errors are unknown, although their distributional properties may be able to be characterized. *Uncertainty* information characterizes the distribution of values it is reasonable to attribute to the measurand, given the measured value and our characterization of effects causing error.

In short, the error is the 'wrongness' of the measured value. The uncertainty describes the 'doubt' we have about the measurand's value, given the measured value and our understanding of effects causing errors.

Uncertainty arises from and propagates through every physical and data transformation involved in creating a dataset.

For the purpose of estimating consistent uncertainties in skin temperature retrievals across all surfaces of Earth in EUSTACE, uncertainty is modelled as three components[2]:

- "random" -- meaning errors that are both random and independent between all data
- "locally systematic" – meaning errors that are highly correlated across short separations in time and distance; statisticians may refer to this case as "structured random"
- "[large-scale] systematic" – meaning errors that have a structure that is persistent in time and space; this includes but is not limited to "biases"

Thus, the components are distinguished by their error correlation length scales. In truth, the division between locally systematic and systematic cases is somewhat artificial, and how best to decompose effects with a range of types of correlation into these components is a matter of judgement (see Merchant et al 2015 for more detail). Information on the various uncertainty components is provided, but also a measure of total uncertainty.

These components of uncertainty are propagated separately through the gridding process and through the process of estimating air temperature from satellite skin temperature using the relationships understood to exist between skin and air temperature over the different surfaces. As well as taking into account uncertainties in the skin temperature used, this process propagates uncertainties in other information such as fractional vegetation cover (over land) and the mask used to exclude retrievals affected by the presence of clouds (over ice).

The above uncertainty components are propagated, where possible, through the process of creating global or European fields which have been completed using statistical methods of infilling. In this case, uncertainty is communicated by provision of an ensemble of gridded fields for each day which are all consistent with our understanding of uncertainties in the input data and the uncertainty in the infilling process itself. In this way, the correlation structure of the uncertainties is encoded, where possible, in the structure of the differences between the ensemble members and provides a convenient way for users to

[2] For some datasets the local systematic uncertainty is subdivided into two

explore the impact of the uncertainties on their application of the EUSTACE data. Besides the ensemble, a best-estimate of the daily air temperature and a measure of the total uncertainty is also provided.

2.1.2.2 Observation Influence

The observation influence is a dimensionless quantity in the range 0-1 indicating the extent to which the daily temperatures in the analysis are influenced by observations. Where the value is close to 0, there are no nearby observations, or they are very uncertain, and the analysis is poorly constrained. The analysis temperature will revert to be close to the climatological average. Where it is close to 1, the value is well constrained by observations. The ensemble spread will be large when the observation influence is zero and small where the observation influence is one.

2.1.3 Local time or UTC?

A day can be defined in various ways, e.g. by local time/time zone, or 00:00-24:00 UTC. Countries regularly use different definitions for their observations (Van den Besselaar et al., 2012). The most appropriate definition may differ per type of user. In EUSTACE we use local time, since we assume this makes it easier for certain groups of potentially new users to use our data (based on feedback from users). Experienced users generally will know how to transform the local time into the Universal Time.

2.1.4 Daily time series and derived data

Some users, e.g. climate scientists, may want to use the data set with the daily data, but others may want to use derived data/information, such as climatologies, trends, etc. These derived products can be calculated with the help of the produced daily time series. Additional to this Product User Guide some examples are given on how to process the EUSTACE datasets[3] into derived data using the Climate4Impact portal. This includes processing into derived data, but also e.g. selecting data for a certain region or period (Step-by-step guide for Climate4Impact).

2.2 Overview of datasets produced

The table below contains a summary of the final output products from the EUSTACE project which are described in more some more detail just below the table and in more detail in Chapter 4.

[3] The examples do not include EUSTACE datasets, since they weren't available when the -step-by-step guide was made.

Table 2-1: EUSTACE datasets with access links and details of any access constraints

Short name	Descriptive Name	Dataset Link	Licence
		Satellite[4] skin temperatures	
Global satellite land surface temperature, v2.1	EUSTACE / GlobTemperature: Global clear-sky land surface temperature from MODIS Aqua on the satellite swath with estimates of uncertainty components, v2.1, 2002-2016	http://catalogue.ceda.ac.uk/uuid/0f1a958a130547febd40057f5ec1c837	Open
	EUSTACE / GlobTemperature: Global clear-sky land surface temperature from MODIS Terra on the satellite swath with estimates of uncertainty components, v2.1, 2000-2016	http://catalogue.ceda.ac.uk/uuid/655866af94cd4fa6af67809657b275c3	Open
Global satellite ice surface temperature, v1.1	EUSTACE / AASTI: Global clear-sky ice surface temperature from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009	https://catalogue.ceda.ac.uk/uuid/60b820fa10804fca9c3f1ddfa5ef42a1	Open
Global satellite sea surface temperature, v1.2	EUSTACE / CCI: Global clear-sky sea surface temperature from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012	https://catalogue.ceda.ac.uk/uuid/b8285969426a4e00b74814342	Open
		Surface air temperatures from stations or in-situ data	
European station measure-ments	EUSTACE/ECA&D: European land station daily air temperature measurements, homogenised	https://catalogue.ceda.ac.uk/uuid/81784e3642bd465aa69c7fd40ffe1b1b	Non-commercial use only
Global Station Measure-ments	EUSTACE: Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015	http://catalogue.ceda.ac.uk/uuid/7925ded722d743fa8259a93acc7073f2	Non commercial use only

[4] Skin temperature estimates based on different satellites are available.

Validation match up database, v1.0	EUSTACE: coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/4b34a2c6890f4e518cacc88911193354	Non-commercial use only
E-OBS	EUSTACE / E-OBS: Gridded European surface air temperature based on homogenised land station records since 1950	https://catalogue.ceda.ac.uk/uuid/b2670fb9d6e14733b303865c85c2065d	Non commercial use only
Surface air temperature estimates from statistical analysis			
Air temperature estimates from satellite	EUSTACE: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016	https://catalogue.ceda.ac.uk/uuid/f883e197594f4fbaae6edebafb3fddb3	Open
Surface air temperature estimates from statistical analysis			
Global air temperature estimates, v1.0	EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/468abcf18372425791a31d15a41348d9	Open

Some further explanation of the above mentioned datasets is given in the following text:

Satellite skin temperatures:

- Satellite skin temperature retrievals: contains the skin temperatures for all surfaces of Earth as obtained in other projects (ESA DUE GlobTemperature, ESA Climate Change Initiative SST and FP7 NACLIM), but with new, consistent total uncertainty estimates (and estimates of uncertainty components split according to their correlation structure) across surfaces for cloud free days. Skin temperatures are obtained from different satellites (section 4.1)

Surface air temperatures from stations or in-situ data:

- European station measurements: This dataset contains homogenized and publicly available station data on air temperatures in Europe (minimum, average and maximum temperature). This is a dataset that is updated regularly with new and more recent measurements (Figure 2-4; section 4.3)
- Global station measurements: This dataset contains all publicly available station data on air temperatures with inhomogeneities identified where possible. It gives an indication in which year there are breaks and it provides a likelihood measure of the identified breaks (Figure 2-2;Figure 2-3; Section 4.2).

- Validation match up database: A surface air temperature match up data base for validation of the EUSTACE global gridded products (section 4.7).
- E-OBS: An in-filled analysis of European surface air temperature based on homogenised meteorological station records since 1950. The dataset interpolates the station data, taking into account differences in altitudes. This dataset is produced as part of the European Climate Assessment and Dataset project (ECA&D), and work undertaken during EUSTACE contributed to updates to this dataset. From version 20 on the data set will contain the homogenized temperature data. The dataset also contains an ensemble to describe the uncertainties (section 4.4).

Surface air temperature estimates

- Air temperature estimates from satellite: Surface air temperature estimates (with estimates of uncertainty) for all surfaces of Earth, derived from satellite surface skin temperature retrievals (for days and periods where satellite skin temperatures are available, i.e. clear sky days; section 4.5).
- Global air temperature estimates: These contain daily analyses of surface air temperature (with estimates of uncertainty) since 1850, based on combined information from satellite and in situ data sources. This dataset also contains air temperature estimates for days with clouds and for the period before satellite data were available (section 4.6).

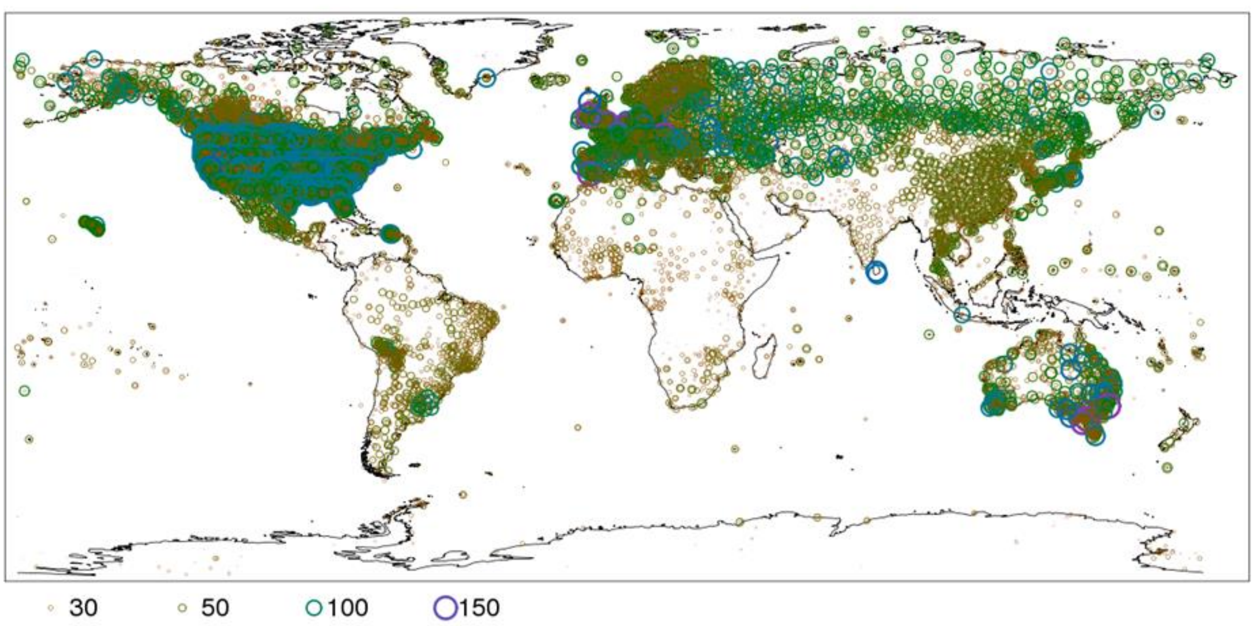


Figure 2-2: Overview of regional spread of available stations in the EUSTACE global data set of surface air temperature measurements from meteorological stations. The size of the circles indicates the length of data available in a particular location. (Brugnara et. al, 2019)

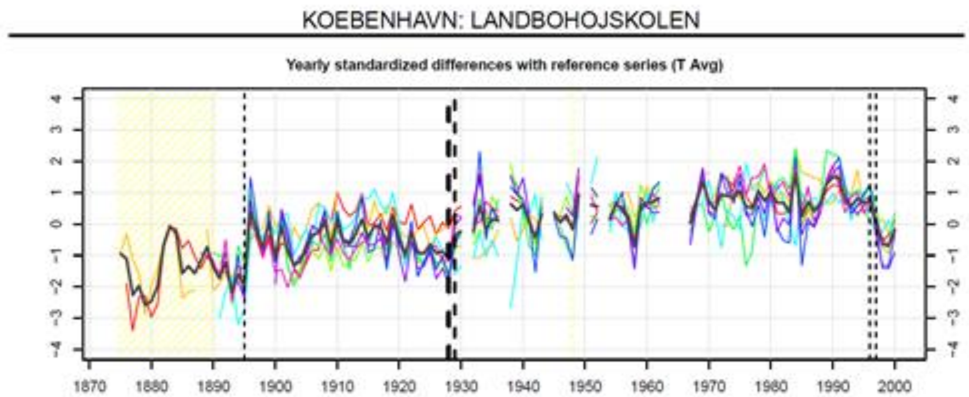


Figure 2-3: Example of breakpoint detection in the EUSTACE global station dataset for Copenhagen. A breakpoint is a point/year where the temperature changes for non-climatic reasons. This can be due to e.g. a change in location, change in instrument or a change in environment . (Brugnara et. al, 2019).

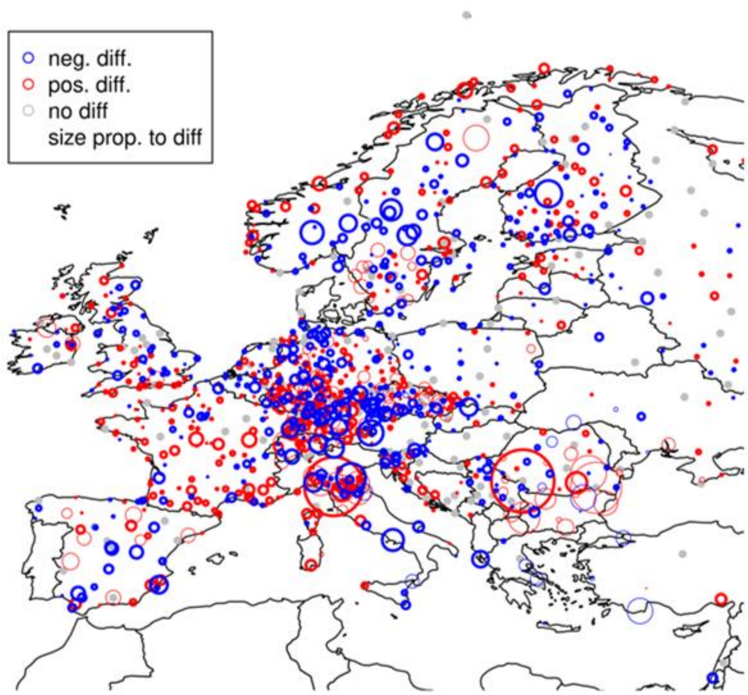


Figure 2-4: Change in trends in annual mean values of minimum temperatures after homogenization. Red circles indicate an increase in trend after homogenisation and blue circles indicate a decrease in trend after homogenisation. The size of the circles indicates the relative sizes of the changes in trend (Squintu et al, 2018).

Table 2-2: Some characteristics of the datasets in the EUSTACE project

Short name	Further detail	Surface	Period	Spatial resolu-tion	Tempo-ral reso-lution	Variables
			Satellite[5] skin temperatures			
Global satellite land surface temperature, v2.1	MODIS-Aqua	Land	2002-2016	on satellite swath	on satellite swath	Tskin
	MODIS Terra	Land	2000-2016	on satellite swath	on satellite swath	Tskin
Global satellite ice surface temperature, v1.1	AVHRR	Ice	2000-2009	on satellite swath	on satellite swath	Tskin
Global satellite sea surface temperature, v1.2	ATSR-series	Ocean	1991-2012	0.25	daily	Tskin(depth)
			Surface air temperatures from stations or in-situ data			
European station measurements	Homogenized	land	1850-2019	stations	Daily	Tn, Tx, Tg[6].
Global Station Measurements	With break-point locations	land	1850-2015	stations	daily	Tmin, Tmax, Tmean
Gridded European station data	Based on homogenized station data, +100 member ensemble	land	1950-2018	0.25 & 0.1 degree	daily	Tn, Tx, Tg[6]
Validation match up database, v1.0	Selection stations for validation	Land and ocean	1850-2015	stations	daily	Tmin, Tmax
			Surface air temperature estimates from satellites			
Air temperature estimates from satellite, v1.0		Land, ocean, ice	1995-2016	0.25	Daily	Tmin, Tmax, Tmean
			Surface air temperature estimates from statistical analysis			
Global air temperature estimates, v1.0	+ 10 member ensemble	Land, ocean, ice, lakes	1850-2015	0.25	daily	Tmean,

[5] Skin temperature estimates based on different satellites are available.

[6] Tn, Tx, Tg are used in these products for the daily minimum, daily maximum and daily averaged temperatures respectively. The daily averaged temperature may be based on the average of the minimum and maximum temperature, but it can also be based on e.g, the average of the hourly temperatures. This has changed over time, and differs between countries.

2.3 Other EUSTACE products

In addition EUSTACE also produced the following products:

- Relationships between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice, ice sheets, and lakes. How these relationships are derived and which relationships are used is described in several articles;
- "Infilling" methods; The methods used are described in articles (in preparation);
- Coded and tested system for product generation;
- User guides for obtaining and using the dataset (this document and the Scientific User Guide). The Scientific User Guide contains information about the methods used to produce the various products. This document, the Product User Guide, gives more practical information about the datasets.

2.4 Format of the EUSTACE data sets

Please see the sections of this Product User Guide (Chapter 4) relevant to each EUSTACE product to find a description of their format.

3 Quick start guide

This section provides a quick reference section on obtaining and making use of the EUSTACE datasets. It is intended to be read both in conjunction with the Product User Guide (PUG), and as an independent reference guide for users starting to use the EUSTACE data; as such it contains pointers to the more detailed information available elsewhere in the EUSTACE product user guide.

3.1 General information on EUSTACE datasets

The final product of the EUSTACE project is:

- Global daily air temperature combining surface and satellite data with uncertainty estimates, for 1850-2015.

Alongside this also other related products are made available:

- global clear-sky land, ice and sea surface temperature from satellite with estimates of uncertainty components;
- globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice;
- global land station daily air temperature measurements with non-climatic discontinuities identified;
- European land station daily air temperature measurements, homogenised;
- Gridded European surface air temperature based on homogenised land station records since 1950;
- coincident daily air temperature estimates and reference measurements, for validation.

Where possible, these data are made freely available to all users with no restriction. Due to limitations of use on some of the station datasets used as input in EUSTACE, the EUSTACE land station surface temperature products are only available for non-commercial use (see Table 2-1 and Table 2-2 in Section 2.2 of the EUSTACE Product User Guide for licensing information and characteristics of individual datasets).

3.2 How to obtain EUSTACE products

The EUSTACE datasets are archived and made available from the Centre for Environmental Data Analysis (CEDA), with the exception of the European Station data and E-OBS dataset (see below). Links to the source of the individual data products are given in Section 2.2 and in the relevant dataset sections of the EUSTACE Product User Guide. They can also be found through the EUSTACE project page in the CEDA catalogue at http://catalogue.ceda.ac.uk/uuid/a52b2cc065a847b8a77a93896880349f.

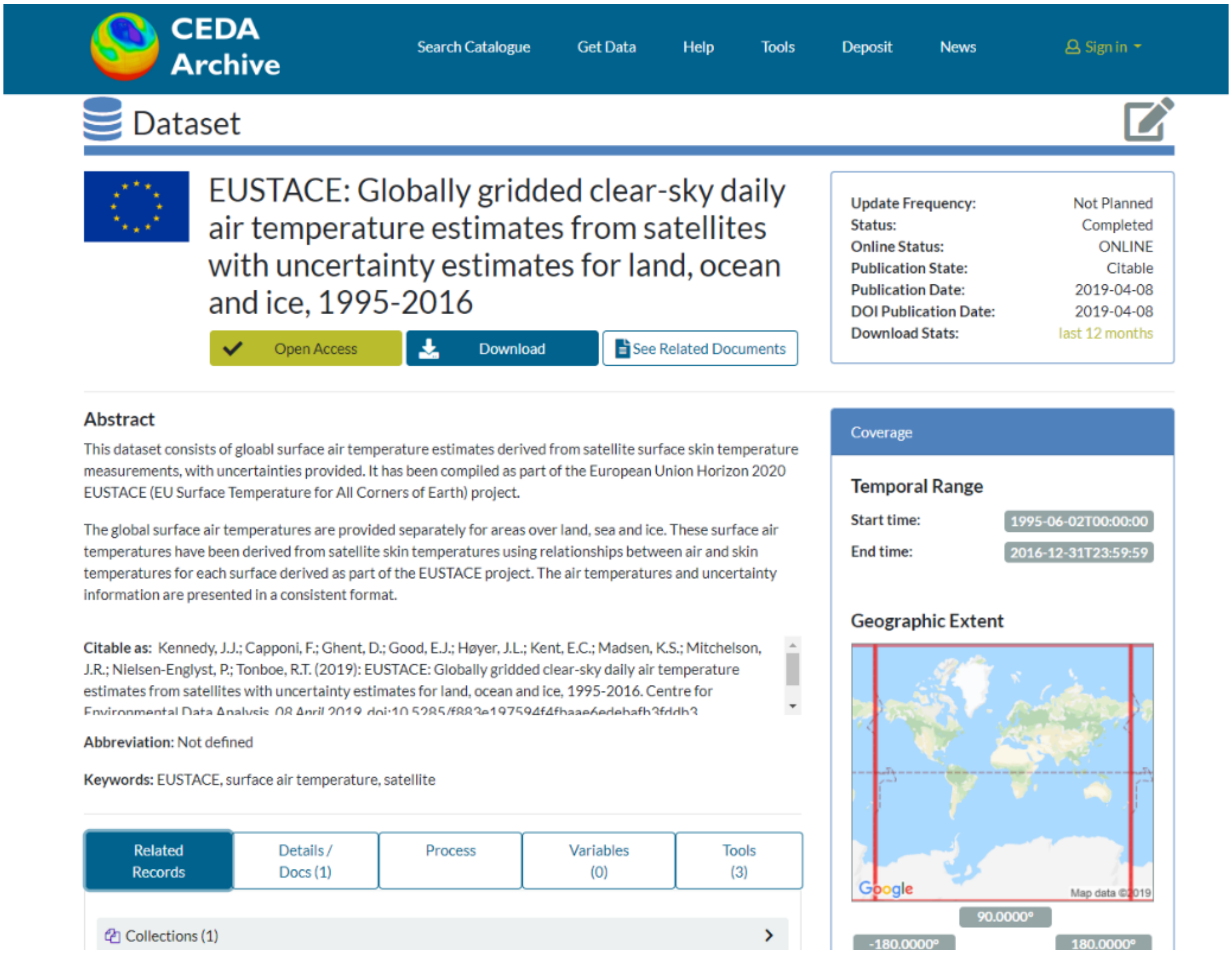


Figure 3-1: An example of a EUSTACE data product that can be obtained via the CEDA catalogue. Further information on the data product, licence and related documentation can be found under the 'details' tab. Data can be downloaded via the 'Download' link at the top of the page.

Data can be downloaded from CEDA via a number of methods as described below:

Web download from CEDA

- From the individual dataset catalogue pages follow the 'Get Data' link at the top of the page.
- Individual files can be downloaded by clicking on the appropriate link, whilst multiple files can be downloaded using the 'Download multiple files' option towards the top of the page. To download large numbers of multiple files (amounting to more than 2-3GB in total), it is recommended to use the FTP download option instead (see FTP section below).
- A subset of the data in the files can be downloaded using the 'subset' link on the web page (see OPeNDAP section below).

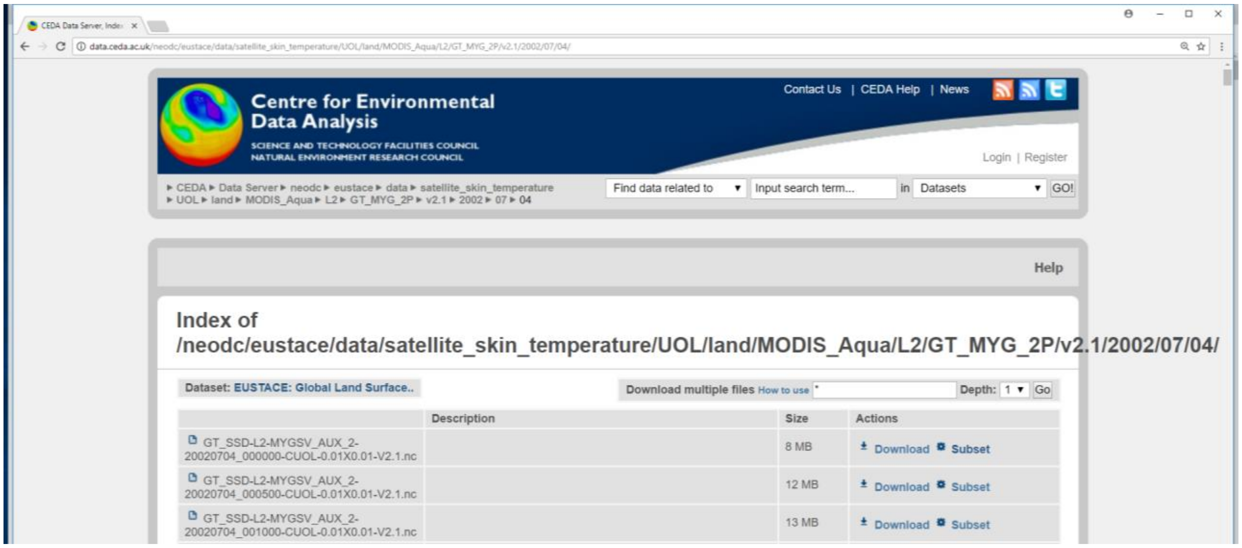


Figure 3-2: Example of a screen where the various available data files for satellite skin temperatures from the EUSTACE project can be downloaded.

FTP:

- Access is available via the CEDA ftp site: ftp.ceda.ac.uk
- Note that a CEDA user account is required to access data via ftp. Users can sign up for this at www.ceda.ac.uk

OPeNDAP:

Data can be downloaded and subset via OPeNDAP. For information on downloading data via OPeNDAP from the CEDA archive see https://help.ceda.ac.uk/article/4712-reading-netcdf-with-python-opendap

E-OBS and European station measurement data:

- E-OBS gridded data and European station data will also soon be available through ECA&D. Links to both these products are also contained in the relevant CEDA page.

Other data access methods: In the future, it may also be possible for some data products to be accessed and visualised directly through some external portals (e.g. the climate4impact portal). For a step-by step example see the Step-by-step guide for Climate4Impact.

3.3 Format and how to read the data products

The data is supplied in NetCDF format and follows the CF-convention. It can be read in by any standard NetCDF reading tools. The exact format for each product is described in more detail in the individual product sections later in this document.

Some suggested common tools to browse NetCDF data are given in Annex 2

3.4 Working with Uncertainty Estimates

A key aim of EUSTACE is to provide consistent uncertainty estimates across all surfaces of the Earth. These uncertainties can be expressed differently per product type, and it is important to understand the types of uncertainties described with the data and how to use them correctly. A brief summary is given in Section 2.1.2.1 of the PUG, and for more information please see the relevant sections for the individual products.

3.5 Things to be aware of when using the data (Assumptions and Do's and Don'ts)

For a detailed description of each data product, we recommend that you read the individual product descriptions (chapter 4). However, there are some general things to be aware of when using the data, which are briefly summarised in this section.

- The station data products made available in our downloadable products only include those stations which are publically available for non-commercial uses. In the processing of the downstream datasets, a larger number of stations have been included in the analysis; unfortunately, we are not able to redistribute some of these stations. Due to the original data restrictions, the EUSTACE station data products are only made available for non-commercial use.
- For some products, data points are given in local solar time; in these cases the definition of a EUSTACE 'day' is local mid-night to mid-night. (See section 2.1.3)
- If averaging data, care needs to be taken to do this properly taking into account the uncertainties and any correlations between data points. For satellite data, averaging needs to account for missing days.
- We would not recommend using the skin and air temperature products to study the skin/air temperature difference, as their derivations are too interconnected. The relationship between how the two variables are determined are described further in the Scientific User Guide.

3.6 How to acknowledge / reference the data

If using this dataset please cite the data as detailed on the relevant CEDA catalogue record, and given in the table in Chapter 7 of the Product User Guide. Please also cite the EUSTACE overview paper from Rayner et al. (2019, in prep).

To refer to articles/reports describing the methodology and datasets look in the scientific user guide and in the list of references (Section 6).

3.7 Where to go for help and further information

For further help on using EUSTACE products please go to sections per product in Chapter 4 where this information is given. Help can also be obtained from the CEDA helpdesk at support@ceda.ac.uk.

For further details on the processing of the EUSTACE products please also see the EUSTACE Scientific User Guide, and published EUSTACE papers.

4 Sections for individual data products with further information

The following sections provide more detailed information on the individual output products produced by EUSTACE.

4.1 Satellite Skin Temperature Retrievals with added uncertainties

The satellite skin temperature retrievals for all surfaces of the Earth have been derived in the context of other projects (e.g. GlobTemperature, CCI SST, FP7 NACLIM), but within EUSTACE, new consistent uncertainty estimates across all surfaces have been added. The following sections describe the details of each of these sub-products, with further scientific information also available in the Scientific User Guide.

4.1.1 Global clear-sky land surface temperature from MODIS on the satellite swath with estimates of uncertainty components, v2.1, 2002-2016

4.1.1.1 Summary

Land Surface Temperature data from the MODIS satellite instruments on NASA's Terra (EOS-AM1) and Aqua (EOS-PM1) satellites have been produced by the University of Leicester under EUSTACE and the ESA DUE GlobTemperature project [http://www.globtemperature.info/] with support from the National Centre for Earth Observation (NCEO) [https://www.nceo.ac.uk/].

The EUSTACE MODIS products provide retrieved Land Surface Temperature on the original satellite measurement grid (i.e. Level 2 products). The Aqua and Terra datasets are archived as two separate data products, with each providing global coverage for the time periods from 04/07/2002 -31/12/2016 and 05/03/2000 – 31/12/2016 respectively. Data is provided on the MODIS swath with a pixel resolution of 1km at nadir.

For further information on the derivation of this product see the EUSTACE Scientific User Guide.

4.1.1.2 Heritage of the data product

The first version of the product (Version 2.0) was produced by the University of Leicester and distributed by the ESA DUE GlobTemperature project as the GlobTemperature GT_MYG_2P (Aqua) and GT_MOG_2P (Terra) products. The LST retrieval algorithm was developed under the ESA DUE GlobTemperature with support from the National Centre for Earth Observation (NCEO), whilst the method for estimating the components of the LST uncertainty was developed under EUSTACE. In Version 2.0 the uncertainty estimates are separated into three components based on correlation scales: a random component, a component that is correlated on local length and time scales, and a large-scale correlated component.

The EUSTACE dataset is Version 2.1 and further separates the locally correlated uncertainty into atmosphere and surface components.

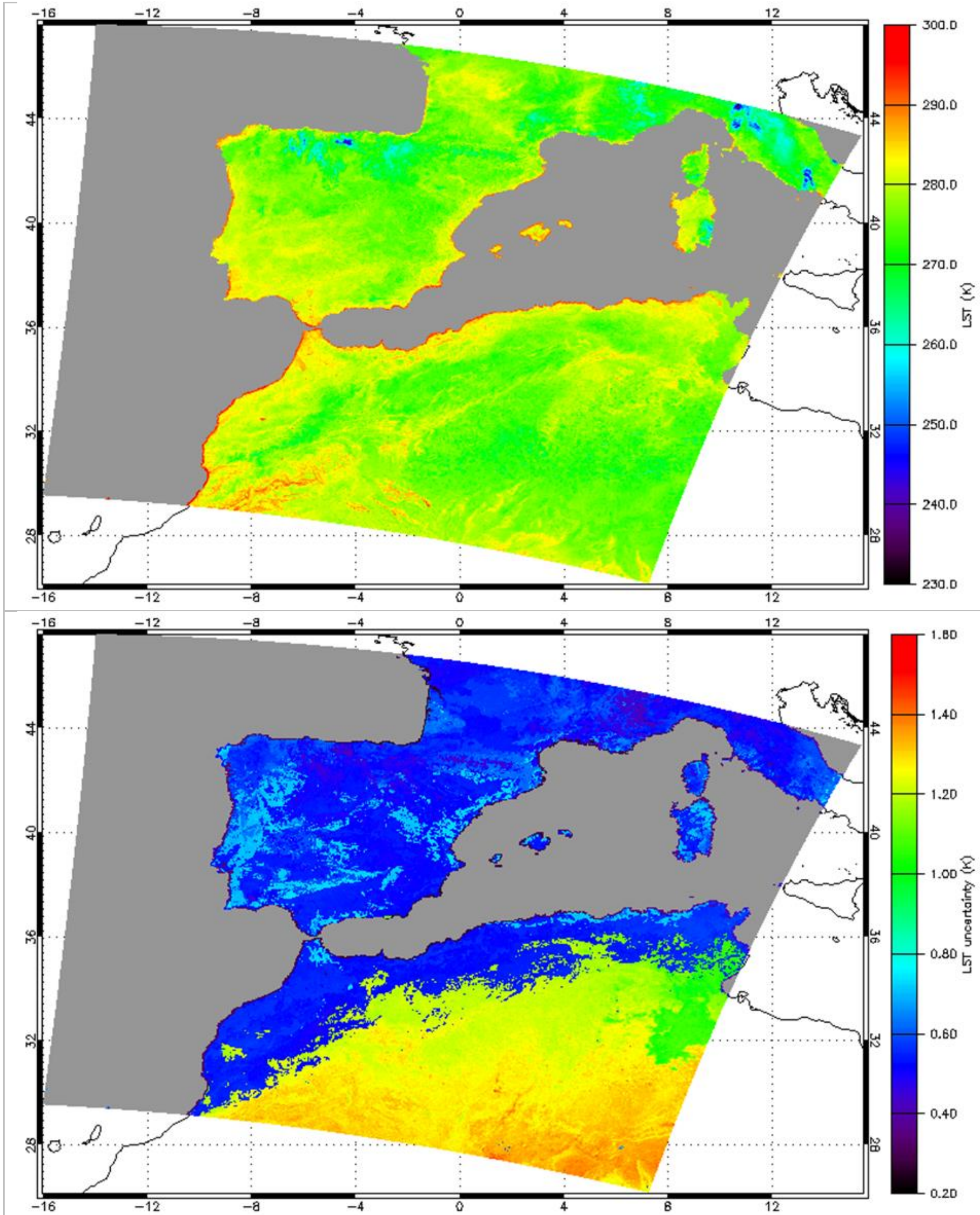


Figure 4-1: An example of the MODIS LST from the Aqua instrument (top) and the associated LST uncertainty field (below). Data plotted are from file GT_SSD-L2-MYGSV_LST_2-20110114_020500-CUOL-0.01X0.01-V2.1.nc

In both versions the LST estimates are retrieved from MODIS Collection 6 radiances downloaded from the NASA Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) [https://ladsweb.modaps.eosdis.nasa.gov/]. Further information on the processing is provided in the EUSTACE Scientific User Guide.

4.1.1.3 Access to Data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. Data are made freely available for use under the Open Government Licence.

Table 4-1: Links to the EUSTACE/GlobTemperature global clear-sky land surface skin temperature data from MODIS with estimates of uncertainty components.

Data Product	CEDA catalogue record	Download links
EUSTACE/GlobTemperature: Global clear-sky land surface temperature data from MODIS Aqua on the satellite swath with estimates of uncertainty components, v2.1, 2002-2016	http://catalogue.ceda.ac.uk/uuid/0f1a958a130547febd40057f5ec1c837	http://data.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/UOL/land/MODIS_Aqua/L2/GT_MYG_2P/v2.1/ ftp://ftp.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/UOL/land/MODIS_Aqua/L2/GT_MYG_2P/v2.1/
EUSTACE/GlobTemperature: Global clear-sky land surface temperature data from MODIS Terra on the satellite swath with estimates of uncertainty components, v2.1, 2000-2016	http://catalogue.ceda.ac.uk/uuid/655866af94cd4fa6af67809657b275c3	http://data.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/UOL/land/MODIS_Terra/L2/GT_MOG_2P/v2.1/ ftp://ftp.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/UOL/land/MODIS_Terra/L2/GT_MOG_2P/v2.1/

The data are in NetCDF format and follow the NetCDF Climate and Forecast (CF)-convention (http://cfconventions.org/). The detailed format specification is given in Section 4.1.1.6.

4.1.1.4 Working with the data

This dataset comprises Land Surface Temperature data retrieved on the satellite orbit track. The data is at Level 2 (i.e. it is not gridded). There are two file types within the above directories:

1) ….LST…: This contains the primary data product of LST and its associated total uncertainty (example field is plotted in Figure 4-1), along with geolocation, viewing geometry and quality flag information.
2) ….AUX..: This contains the more detailed breakdown of the LST uncertainty estimates into four components (see below) as well as further auxiliary information that has been used in the LST retrieval, such as surface emissivity.

The data is split into 5 minute granule files consistent with those provided for the MODIS operational Level-1b and Level-2 data.

Using the uncertainties

A pixel level uncertainty analysis is provided in each AUX file, which takes account of the expected performance of the retrieval algorithm under varying surface and atmospheric conditions. These uncertainties have been categorised into four components: random; locally correlated (atmosphere); locally correlated (surface); and systematic (Figure 4-2).

Table 4-2: Uncertainty components in the global surface skin temperature data files from EUSTACE

Uncertainty component	Definition	Contributions
Random (LST_unc_ran)	Uncorrelated in space and time.	Radiometric noise of the instrument and sub-pixel variations in the emissivity
Locally Correlated – surface (LST_unc_loc_sfc)	Correlated within same land cover type but appear random across different land cover types	Pixel-to-pixel uncertainty in emissivity across each landcover class
Locally Correlated – atmosphere (LST_unc_loc_atm)	Correlated on synoptic scales, but appear random on larger scales	Uncertainty in atmospheric state (primarily water vapour)
Systematic (LST_unc_sys)	Fully correlated across space and time	Uncertainty in the bias of the satellite surface temperatures relative to other data sources of temperature once all known residual biases have been corrected for

A total uncertainty, accounting for all components, is available in the LST file and is simply, for each pixel, the sum in quadrature of the uncertainty components assuming these are independent. Uncertainty components are provided so that the user can take account of error covariances when averaging the data to coarser resolutions (temporal and/or spatial). Thus, the total uncertainty on a spatial or temporal average (σ_{av}) over a sample of pixels of size N is given by:

$$\sigma_{av}^2 = \frac{1}{N^2} \sum_{k=1}^{K} \left[\sum_{i=1}^{N} \sigma_{ik}^2 + 2 \sum_{i=j+1}^{N} \sum_{i=1}^{N-1} R_{ijk} \sigma_{ik} \sigma_{jk} \right]$$

Here, N is the number of data points contributing to the average, σ_{ik} is the k^{th} uncertainty component on the i^{th} pixel and R_{ijk} is the coefficient of correlation of the k^{th} uncertainty component at the i^{th} and with k^{th} uncertainty component at the j^{th} pixel. Commonly, it is assumed that $R = 1$ for pixel separations less than some correlation scale and that $R = 0$ elsewhere. For the randomly correlated component the correlation scale is zero and $R = 0$ for all pixel pairs. For the systematic component the correlation scale is assumed infinite and $R = 1$ for all pixel pairs. In the case of the locally correlated components the correlation scales must be estimated. It is likely that the errors in the LST due to errors in estimating the emissivity of a particular land cover type are correlated for pixels of the same land cover type although this may not be the case for similar land cover types on different continents (for example, deciduous woodland is likely to include different species of trees in different regions). Regarding LST errors due to errors in atmospheric characterization, these are likely to be correlated with atmospheric water vapour which has correlation scales of less than 10 km and less than 1 hour [Steinke et al., 2015; Vogelmann et al., 2015].

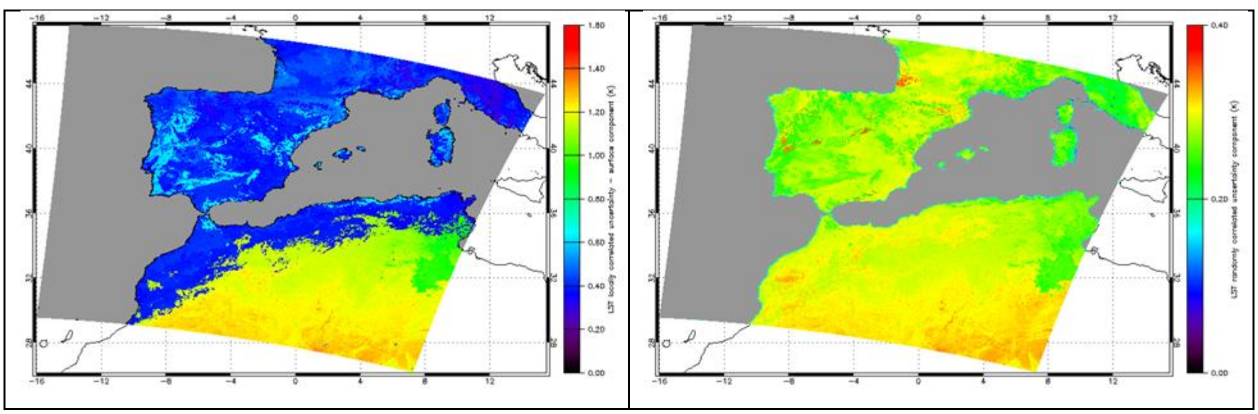


Figure 4-2: The surface locally correlated (left) and random (right) components of the LST uncertainty field shown in Figure 4-1

Quality Flags

The MODIS LST files contain a quality control variable (QC). This is a 4 byte integer with the bits corresponding to the following cases:

Table 4-3: Explanation of the Quality control variable in the MODIS LST files

Bit	Flag
0	Day / Night flag (day = 0, night = 1) • Night is set where the solar zenith angle is ninety degrees or more
1	Consolidated cloud flag (0 = no cloud, 1 = cloud) • The best estimate is where primary (bit 6), secondary (bit 7) and thin cirrus (bit 9) are set to 0
2	Consolidated confidence flag (0 = good confidence, 1 = low confidence) • This is set where either of the input split-window brightness temperatures have missing values or are suspect
3	Aerosol information available (1 = yes) • This bit is set to 1 since the information is acquired by the MODIS Cloud Product
4	Aerosol detected (1 = yes) • Information acquired from MODIS Cloud Product
5	Spare bit for potential future evolution
6	Primary cloud mask – definitely cloudy (from MODIS Cloud Product)
7	Secondary cloud mask – probably cloudy (from MODIS Cloud Product)
8	Tertiary cloud mask – possibly cloudy (from MODIS Cloud Product)
9	Thin cirrus cloud mask (from MODIS Cloud Product)
10	Shadow flag (from MODIS Cloud Product)
11	Sun glint flag (from MODIS Cloud Product)
12	Non-optimum confidence flag where the pixel is produced but the quality may be suspect
13-31	Not used

To select only data of the best quality, it is recommended to use those values indicated as cloud free by the Consolidated Cloud Mask (QC Bit 1) and of good quality by the Consolidated Confidence Flag (QC bit 2). This accounts for both the best knowledge of cloud contamination and invalid MODIS radiance pixels.

Values where the total 'LST_uncertainty' field is greater than 2.0K should also be treated with more caution.

Known limitations of the data

When using the data, users should be aware of the following:

- Users should take account of the quality flag (as described above).
- The uncertainty in the LST due to failure of the cloud detection algorithm is not included in the uncertainty budget.

For more information on the data quality and validation methods refer to the ESA DUE GlobTemperature Validation Report [Martin and Goettsche et al 2017].

4.1.1.5 File naming convention

Product filenames follow the GlobTemperature file naming conventions (see GlobTemperature Product User Guide available from http://www.globtemperature.info)

4.1.1.6 Data format specification

The data are provided in NetCDF-4 format, and compliant with the CF-1.6 metadata convention. The data follow the GlobTemperature harmonised data format (see the GlobTemperature Product User Guide available from http://www.globtemperature.info/).

Two separate files are provided: the primary LST file and an auxiliary file.

1) The primary LST file holds those variables specified as essential by GlobTemperature, with the size kept to a minimum. Geolocation coordinates are specified to 4 decimal places (Table 4-4).

Table 4-4: GlobTemperature primary LST harmonised format (from GlobTemperature PUG)

Dimensions	Name				
	time				
	nj				
	ni				
Variables	**Name**	**Type**	**Dimensions**	**Units**	**Comment**
	jul_date	double	time	days	Reference time at start of datafile in Julian Days (days since 12h Jan1 , 4713BC)
	lat	float	nj,ni	degrees_north	Pixel centre latitude in decimal degrees north
	lon	float	nj,ni	degrees_east	Pixel degrees longitude in decimal degrees east
	dtime	int	nj,ni	seconds	Time difference from reference time in seconds with an accuracy of 0.1s
	LST	short	nj,ni	K	LST
	LST_uncertainty	short	nj,ni	K	total LST uncertainty
	QC	Int	nj,ni	unitless	Quality control flags
	satze	short	nj,ni	degree	Satellite zenith viewing angle
	sataz	short	nj,ni	degree	Satellite azimuth viewing angle

2) The auxiliary (AUX) file includes extra variables, outside of the mandatory GlobTemperature variables (Table 4-5).

Table 4-5 Format of the auxiliary AUX files

Dimensions	Name				
	Channel				
	Nj				
	Ni				
	Systematic				
	ch_length				
Variables	**Name**	**Type**	**Dimensions**	**Units**	**Comments**
	LST_unc_loc_atm	short	nj,ni	K	Uncertainty due to locally correlated atmospheric effects
	LST_unc_loc_sfc	short	nj,ni	K	Uncertainty due to locally correlated surface effects
	LST_unc_ran	short	nj,ni	K	Uncertainty due to random effects
	LST_unc_sys	short	systematic	K	Uncertainty due to systematic effects
	channel	char	channel, ch_length	Unitless	channel description
	elevation	short	nj,ni	m	Elevation of land surface
	emis	short	channel,nj,ni	unitless	Channel emissivity
	lwm	short	nj,ni	1	Land water mask*
	solaz	short	nj,ni	degree	Solar azimuth angle
	solze	short	nj,ni	degree	Solar zenith angle
	tcwv	short	nj,ni	kg m-2	Total column water vapour

*Land water mask flag values and meanings:

(0) *Shallow Ocean* – ocean less than 5k from coast *or* more than 50m deep

(1) *Land* - not anything else

(2) *Ocean Coastlines and Lake Shorelines*

(3) *Shallow Inland Water* – inland water less than 5km from shore *or* more than 50m deep

(4) *Ephemeral* – intermittent water

(5) *Deep Inland Water* – inland water more than 5km from shoreline *and* more than 50m deep

(6) *Moderate or Continental Ocean* - ocean more than 5km from coast *and* more than 50m deep *and* less than 500m deep

(7) *Deep Ocean* – ocean more than 500m deep

4.1.1.7 Where to go for further Information

- Eustace Scientific User Guide (under development)
- GlobTemperature [http://www.globtemperature.info/]
- GlobTemperature Product User Guide available from http://www.globtemperature.info/

Points of contact for further questions: Darren Ghent (djg20@le.ac.uk) and Karen Veal (klv3@le.ac.uk)

4.1.1.8 References for 4.1.1

Ghent, D., I. Trigo, A. Pires, O. Sardou, J. Bruniquel, F. Göttsche, M. Martin, C. Prigent, and C. Jimenez (2018), ESA DUE GlobTemperature Product User Guide, Rep. GlobT-WP3-DEL-11, University of Leicester. [http://www.globtemperature.info/index.php/public-documentation/deliverables-1/108-globtemperature-product-user-guide]

Steinke, S., Eikenberg, S., Löhnert, U., Dick, G., Klocke, D., Di Girolamo, P., & Crewell, S. (2015). Assessment of small-scale integrated water vapour variability during HOPE. Atmospheric Chemistry and Physics, 15(5), 2675–2692. https://doi.org/10.5194/acp-15-2675-2015

Vogelmann, H., Sussmann, R., Trickl, T., & Reichert, A. 2015). Spatiotemporal variability of water vapor investigated using lidar and FTIR vertical soundings above the Zugspitze. Atmospheric Chemistry and Physics, 15(6), 3135–3148. https://doi.org/10.5194/acp-15-3135-2015

Martin, M., and Goettsche, F.M. (2017) GlobTemperature Satellite LST Validation Report (Report to ESA: GlobTemp_DEL-12_i3r0) [http://www.globtemperature.info/index.php/public-documentation/deliverables-1/117-validation-report-del12]

4.1.2 Global clear-sky sea surface temperature from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012

4.1.2.1 Summary

The dataset archived here has been produced using methods from the ESA Climate Change Initiative (CCI) Sea Surface Temperature (SST) project; it is an experimental v1.2 (A)ATSR Level 3C data product, which was made available to EUSTACE as representing the state of the art at the time of the project. However, this was not made publically available via the ESA CCI, and has now been superseded by their v2.0 product. The EUSTACE dataset made available here, is provided primarily for traceability, and it is recommended for scientific studies to use the subsequent CCI products from http://cci.esa.int/data.

The files provided here are Level 3C files, where the data has been gridded daily onto a regular 0.25 degree latitude/longitude grid.

4.1.2.2 Heritage of the Data Product

The dataset relates closely to others produced as part of the ESA Climate Change Initiative (CCI) Sea Surface Temperature (SST) project (http://www.esa-sst-cci.org/).

This particular product comprises Level 3C gridded sea surface temperatures from the Along Track Scanning Radiometer (ATSR) series of instruments (ATSR-1, ATSR-2, AATSR). The data has been gridded to 0.25 degrees, and covers the period 1992-2012.

4.1.2.3 Access to Data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. Data are made freely available for use under the Open Government Licence.

Table 4-6: Links to the EUSTACE/CCI global clear-sky sea surface temperature data from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012.

Data Product	CEDA catalogue record	Download links
EUSTACE / CCI: Global clear-sky sea surface temperature from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012	https://catalogue.ceda.ac.uk/uuid/b8285969426a4e00b7481434291ad603	http://data.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/UOR/ocean/ATSR/L3C/v1.2/

The data are in NetCDF format and follow the NetCDF Climate and Forecast (CF)-convention (http://cfconventions.org/). The detailed format specification is given in Section 4.1.2.5.

New versions of the CCI datasets are now available which supersede this product; it is provided here primarily for traceability but for scientific studies we recommend using the subsequent CCI products from http://cci.esa.int/data.

4.1.2.4 File naming Convention

The file names have the format:

<Date><Time>-<RDAC>-<Level>_GHRSST-<Product>-<Dataset>-v02.0-fv01.0.nc

The variable components within braces: *<component>* are summarised in Table 4-7 and detailed in the following sections. Note:

- *<Dataset>* and *<Extra>* correspond to the GHRSST *<Additional Segregator>* component and are therefore separated by an underscore ("_") rather than a dash ("-").
- The fixed version string "v02.0" indicates that the file is a GDS 2.0 format file. The Climate Data Record version is indicated by the *<Dataset>* string.

Table 4-7: GDS 2.0 File naming convention components

Component	Definition	Description
<Date>	YYYYMMDD	The identifying date for this file in ISO8601 basic format
<Time>	HHMMSS	The identifying time for this file in ISO8601 basic format
<RDAC>	ESACCI	The RDAC where the file was created
<Level>	L3C	The data processing level
<Product>	GHRSST product string e.g. AATSR	Indicates the source data for this product. See Table 4-8 for full list
<Dataset>	EXP1.2	Indicates this is an experimental CCI version (1.2)

Date

The identifying date for this file, using the ISO8601 basic format: YYYYMMDD.

Time

The identifying time for this file in UTC, using the ISO8601 basic format: HHMMSS. This corresponds to the centre time of collation window (1200 for daily files)

RDAC

GHRSST Regional Data Assembly Centre (RDAC) where the dataset was generated. Here this is ESACCI:ESA Climate Change Initiative

Level

GHRSST Processing level for this product = L3C.

Product

GHRSST Product string indicating the source sensor for this dataset:

Table 4-8: GDS 2.0 Product strings

Name	Description
AATSR	Advanced Along Track Scanning Radiometer (AATSR) on Envisat satellite
ATSR<X>	Along Track Scanning Radiometer (ATSR) on ERS-1 or -2

Dataset

Indicates which (Interim) Climate Data Record this file belongs to = EXP1.2

NOTE – the <*Dataset*> element correspond to the GHRSST <Additional Segregator> component and are therefore separated by an underscore ("_") rather than a dash ("-").

4.1.2.5 Data format Specification

The data files are in netCDF-4 format and are CF-compliant [D4], following the GDS [D1] SST-CCI [D2] standards.

NetCDF variable Attributes

Variables in the netCDF files will include the standard metadata attributes listed in Table 4-9 below. These are recognised by most tools and utilities for working with netCDF files.

Table 4-9: Standard variable attributes

Attribute name	Description
_FillValue	The number put into the data arrays where there are no valid data (before applying the scale_factor and add_offset attributes).
long_name	A descriptive name for the data
standard_name	A unique descriptive name for the data, taken from the CF conventions [D4]
units	The units of the data after applying the scale_factor and add_offset conversion
add_offset	After applying scale_factor below, add this to obtain the data in the units specified in the units attribute
scale_factor	Multiply the data stored in the netCDF file by this number
valid_min	The minimum valid value of the data (before applying scale_factor and add_offset).
valid_max	The maximum valid value of the data (before applying scale_factor and add_offset).

comment	Miscellaneous information
references	References that provide more information about the data
source	A list of data sources used for the data in this variable
depth	The effective depth for SST data. Must be one of [skin, depth_20, depth_100]. The default value is 'skin'

Coordinate Grids

The coordinate variables are listed in Table Table 4-10 and discussed in the following sections.

Table 4-10: Coordinate variables

Variable name	Description
lat	Central latitude of each grid cell
lon	Central longitude of each grid cell
time	Reference time of SST file

Time coordinate

All SST files include time as a dimension and coordinate variable to represent the reference time of the SST data array. The reference time for L3C files is the centre time of collation window (midday for daily files)

Regular latitude/longitude grid

The files are stored on a global regular latitude/longitude grid and variables have the following dimensions:

time: UNLIMITED (1)
lat: Number of latitude points (720)
lon: Number of longitude points (1440)

The resolution used for the products is 0.25°. The time dimension is specified as **unlimited**, allowing standard netCDF tools to easily concatenate and manipulate multiple files, but each file is distributed with a single time slice (corresponding to a day).

SST Data Variables

The data files contain two SST variables: the primary satellite measurement is the temperature of the skin at the time the satellite observes it; the adjusted SST is provided for a standard depth (0.2 m) and time representative of the daily mean to allow comparison with *in situ* measurements and prevent aliasing the diurnal cycle into the data due to satellite drift.

Table 4-11: SST data variables

Variable name	Description
sea_surface_temperature	Best estimate of SST_{skin} as observed by the satellite
sea_surface_temperature_depth	Best estimate of $SST_{0.2m}$ at standard time representing daily mean

Quality Indicators

Each pixel has an associated `quality_level` which indicates the general quality of that pixel – higher values being better. Quantitative analyses should use the higher quality levels (4 or 5). Quality levels 2 and 3 may be useful for qualitative analyses, but the pixels have an increased change of being cloud contaminated.

Table 4-12: Quality indicators

Variable name	Description
quality_level	Quality level of the SST: 0 no data 1 bad data 2 worst usable data 3 low quality 4 good quality 5 best quality

Auxiliary variables and uncertainties

There are several auxiliary variables and SST uncertainties listed in Table 4-13below.

Table 4-13: Auxiliary variables and uncertainties

Variable name	Description
sses_standard_deviation	Total uncertainty in SST_{skin}
sst_depth_total_uncertainty	Total uncertainty in $SST_{0.2m}$
large_scale_correlated_uncertainty	Systematic uncertainty that is highly correlated between pixels over large scales
synoptically_correlated_uncertainty	Systematic uncertainty that is highly correlated between pixels over synoptic scales only
uncorrelated_uncertainty	Non-systematic uncertainty (uncorrelated or weakly correlated between pixels)

4.1.2.6 Where to go for further information

See the ESA CCI SST project pages at http://www.esa-sst-cci.org/

For access to the latest versions of the cci data go to cci.esa.int/data.

4.1.3 Global clear-sky ice surface temperature from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.0, 2000-2009

4.1.3.1 Summary

Arctic and Antarctic ice Surface Temperatures from thermal Infrared satellite sensors (AASTI), covering high latitudes Seas, Sea Ice and Ice Cap surface temperatures based on satellite infrared measurements. AASTI is a supplement to existing data sets, with lower resolution, but with longer temporal coverage.

The primary data input are the satellite radiometer infrared brightness temperature (Tb) data used to calculate surface temperatures. These data are derived from a 28-year Global Area Coverage – Advanced Very High-Resolution Radiometer (GAC-AVHRR) data set. The areas of interest are the Polar Regions, above and below latitudes 40 degree North and South, respectively. Data at mid and low latitudes are masked out.

The AASTI temperature algorithm uses a suite of algorithms selected by surface temperature and sun-elevation, where the primary choice of algorithm lies in the distinction between sea and ice surfaces.

Calibration of each infrared instrument is done from simulated Top-Of-Atmosphere Tb's, calculated by a radiative transfer model.

4.1.3.2 *Heritage of the data product*

The first version of the AASTI data was generated by Met Norway and DMI within the NORMAPP and the NACLIM projects (Dybkjaer et al (2014)). The derivation of uncertainties in the three components was subsequently done within EUSTACE (see Høyer et al., 2018 for a description of this) and was added to version 1.1. Nothing changed in the actual IST field from version 1 to 1.1, except the addition of uncertainty components.

Since then, AASTI v2 has been produced, also including the uncertainty components developed within EUSTACE. The operational OSI-SAF dataset Dybkjaer et al (2018) continues the processing forwards in time.

4.1.3.3 Access to the data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. Data are made freely available for use under the [Open Government Licence](#).

Table 4-14: Links to the EUSTACE/AASTI global clear-sky ice surface temperature data from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009 1991-2012.

Data Product	CEDA catalogue record	Download links
EUSTACE / AASTI: Global clear-sky ice surface temperature from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009	https://catalogue.ceda.ac.uk/uuid/60b820fa10804fca9c3f1ddfa5ef42a1	http://data.ceda.ac.uk/neodc/eustace/data/satellite_skin_temperature/DMI/ice/AASTI/L2/v1.1/

4.1.3.4 Working with the data

This dataset comprises Sea Surface Temperature, Ice Surface Temperature and Marginal Ice Zone surface temperatures on the satellite orbit track. The data is at Level 2 (i.e. it is not gridded). There is one file per orbit for each satellite, similar to the Level 1b input data from CLARA A1.

Using the uncertainties

A pixel level uncertainty analysis is provided, which takes account of the expected performance of the retrieval algorithm under varying surface and atmospheric conditions. These uncertainties have been categorised into three components: random, locally correlated and systematic.

Table 4-15: Uncertainty components and their definitions

Uncertainty component	Definition	Contributions
Random (uncorrelated_uncertainty)	Uncorrelated in space and time	Radiometric noise of the instrument and sub-pixel variations in the emissivity
Locally Correlated: (synoptically_correlated_uncertainty)	Correlated within synoptic scales, I.e. 500-1000 km spatial and 2-5 days	Changes in emissivity and uncertainty in atmospheric state (primarily water vapour)
Systematic (large_scale_correlated_uncertainty)	correlated across space and time	Uncertainty in the bias of the satellite surface temperatures relative to

	other data sources of temperature once all known residual biases have been corrected for

The total uncertainty, accounting for all components, can be derived for each pixel, as the square root of the sum in quadrature of the uncertainty components, assuming these are independent.

4.1.3.5 Data format specification

Variable description:

- **lat:** Latitude.
- **lon:** Longitude.
- **mask:** sea, icecap and land mask, from NOAA NESDIS.
- **sea_ice_fraction;** Sea ice concentration, from OSISAF.
- **surface_temperature:** The integrated sea and ice skin temperature.
- **sea_surface_temperature**: Sea surface skin temperature in 10-20 microns depth. This field is a subset of the surface temperature field, above, and identical to the SST field in the GDS format description (GDS) from the GHRSST community (GHRSST, 2014-9).
- **st_dtime**: a line-wise time offset in seconds, from the Indicative date/time (reference time). I.e. “Indicative date/time + st_dtime” is the line-wise time of data acquisition.
- **l2p_flags**: A 2-BYTE bit-field including miscellaneous information. Bit no.0-5. Landmask – bit0, algorithm bit 1-7, Cloudmask bit 8-15:
    - 0. Land: Land - from the input GAC data set
    - 1. No_algorithm_applied: Algorithm check failed.
    - 2. Sea_Surface_Temperature_day_algorithm
    - 3. Sea_Surface_Temperature_night_algorithm
    - 4. Sea_Surface_Temperature_twilight_algorithm
    - 5. Ice_Surface_Temperature_algorithm.
    - 6. Marginal_Ice_Zone_Temperature_algorithm_sstday_ist.
    - 7. Marginal_Ice_Zone_Temperature_algorithm_sstnight_ist.
    - 8. Cloudmask_quality_high: Quality indicator from PPS cloud mask software (PPScloud, 2014-09).
    - 9. Cloudmask_not_processed: No cloud mask data or corrupted data (PPScloud, 2014-09).
    - 10. Cloud_free: no contamination by snow/ice covered surface, no contamination by clouds but contamination by thin dust/volcanic clouds not checked (PPScloud, 2014-09).
    - 11. Cloud_contaminated: Partly cloudy or semitransparent. May also include dust clouds or volcanic plumes (PPScloud, 2014-09).

- 12. Cloud_filled: opaque clouds completely filling the FOV. May also include thick dust clouds or volcanic plumes (PPScloud, 2014-09).
- 13. Snow_ice_contaminated: (PPScloud, 2014-09).
- 14. Undefined: Cloud Mask has been processed but not classified due to separability problems (PPScloud, 2014-09).
- 15. Not in use.

- **quality_level:** A 1BYTE bit field of relative quality indicators (GHRSST, 2007-09). Bit no.:
  - 0. no_data
  - 1. bad data
  - 2. worst_quality
  - 3. low_quality
  - 4. acceptable_quality
  - 5. best_quality
  - 6. not in use
  - 7. not in use
- **sses_standard_deviation**: Single Sensor error statistics (GHRSST, 2007-09).
- **sses_bias:** Single Sensor bias statistics (GHRSST, 2007-09).
- **satellite_zenith_angle:** Also called view angle.
- **probability_of_water:** Baysian approach - will be implemented in AASTI version2
- **probability_of_ice:** Baysian approach - will be implemented in AASTI version2
- **wind:** ERA-INTERIM reanalysis (ERA_interim, 2014-09)
- **t2m:** 2m air Temperature from ERA-INTERIM reanalysis (ERA_interim, 2014-09)
- **solar_zenith_angle:** Also called Sun elevation angle.

4.1.3.6 File naming convention

The AASTIs file name convention follows the GDS file format description (GDS, 2014-09):

<Indicative Date><Indicative Time>-<RDAC>-<Processing Level>_GHRSST-<SST Type>- <Product String>-<Additional Segregator>-v<GDS Version>-fv<File Version>.<File Type>

Where:

- Indicative Date: The data set, acquisition start date
- Indicative Time: The data set, acquisition start date
- RDAC: Place of creation
- Processing Level: Processing level code
- SST Type: Type of surface temperature
- Product String: The data set identification
- Additional Segregator: optional text – here describing the area of interest.
- GDS Version: Version 2
- File Version: The version of the output data set

- File Type: Type of output file.

An example of a AASTIs file name:

20070315092700-DMI_METNO-L2P_GHRSST-STskin-GAC_polar_SST_IST-noaa18_00000_10589-v02.0-fv01.0.nc

4.1.3.7 Where to go for further information
See references below and the EUSTACE Scientific User Guide for further information.

4.1.3.8 References
Dybkjaer et al (2014) Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST, NACLIM Deliverable D32.28, https://naclim.cen.uni-hamburg.de/fileadmin/user_upload/naclim/Archive/Deliverables/D32.28-SUBMITTED-2014-10_PU.docx

Høyer, J. L., E. Good, P. Nielsen-Englyst, K. S. Madsen, I. Woolway, J. Kennedy, 2018: Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes, EUSTACE Deliverable 1.5, https://www.eustaceproject.org/eustace/static/media/uploads/d1.5_revised.pdf

Dybkjaer et al (2018) Algorithm theoretical basis document for the OSI SAF Sea and Sea Ice Surface Temperature L2 processing chain, http://osisaf.met.no/docs/osisaf_cdop3_ss2_atbd_hl-l2-sst-ist_v1p4.pdf

4.2 Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015

4.2.1 Summary

The land station dataset is a global collection of daily maximum and minimum temperature series (Tmax and Tmin, respectively) covering the period 1850-2015. The series available to the user were obtained from public databases (hence, there is no data in this collection that cannot be found elsewhere).

The advantage over other similar collections is that the data have been "polished" by removing duplicates and unreliable data sources, and they come with a large amount of additional information on quality, homogeneity, and resolution.

The ideal user of this dataset is someone that is more concerned about the quality of the data than the quantity. Nevertheless, users have more than 30 thousand pairs of maximum and minimum temperature series to choose from (Figure 4-3).

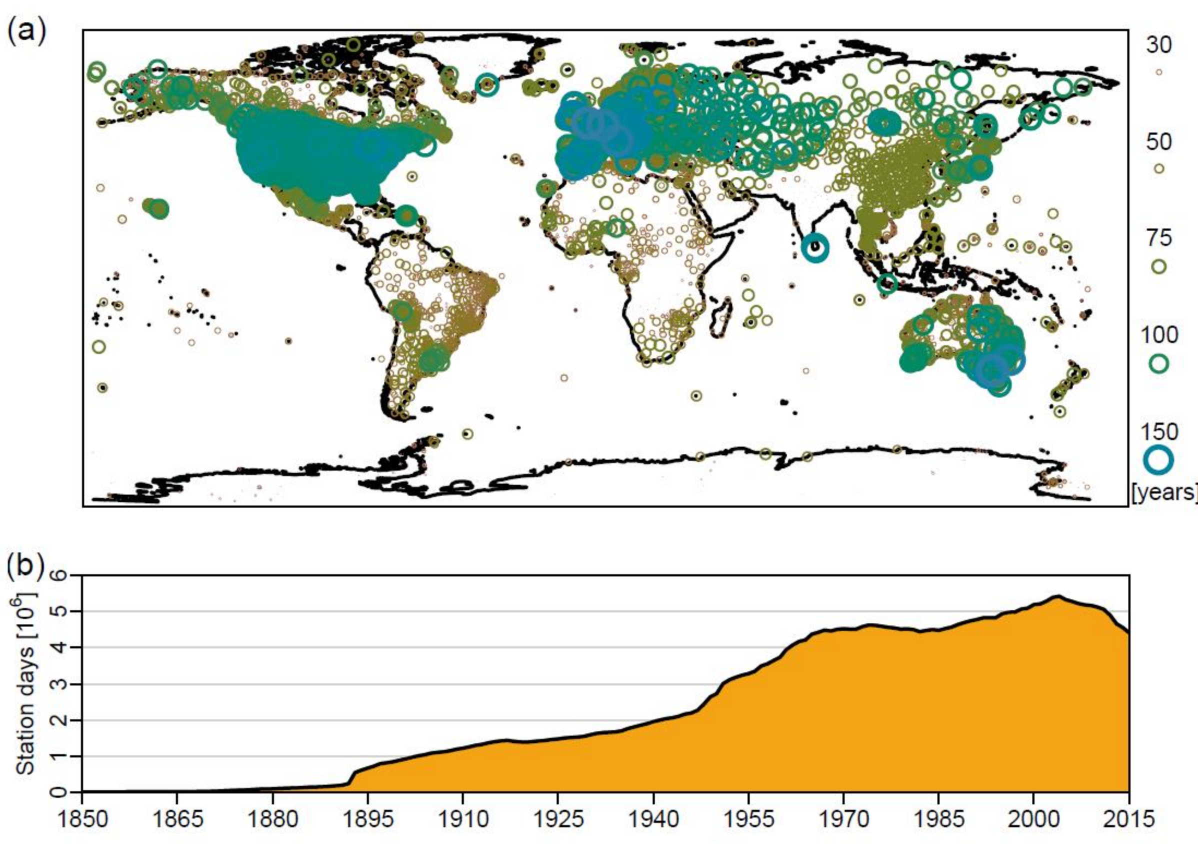


Figure 4-3 a) Maps of the stations included in the EUSTACE land station dataset, where the size and color of the points depend on the length of the series. b) Temporal evolution of the total amount of data. Both public and non-public data are shown.

4.2.2 Data sources

This dataset is built on three pivotal public data collections:

- Global Historical Climatology Network (GHCN)-Daily (Menne et al., 2012)
- European Climate Assessment & Dataset (ECA&D) (Klein-Tank et al., 2002)
- International Surface Temperature Initiative (ISTI) (Rennie et al., 2014)

In addition, EUSTACE is among the first projects to include the DECADE dataset for the Central Plateau in South America (Hunziker et al., 2017), as well as the surface observations digitised during the ERA-CLIM project (Stickler et al., 2014).

4.2.3 Access to data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. It is made available for non-commercial purposes under the Non-Commerical Government Licence, due to non-commercial restrictions on the input in-situ datasets.

Table 4-16: Links to the EUSTACE: Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015.

Data Product	CEDA catalogue record	Download links
EUSTACE: Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015	http://catalogue.ceda.ac.uk/uuid/7925ded722d743fa8259a93acc7073f2	http://catalogue.ceda.ac.uk/uuid/7925ded722d743fa8259a93acc7073f2

4.2.4 Quality assurance

All data series underwent the 14 quality tests that are applied operatively to the GHCN-Daily dataset, as described in Durre et al. (2010). These include checks on basic integrity, outliers, as well as internal, temporal and spatial consistency. If a certain observation fails one of the tests, a quality flag is assigned to that observation (see table in section 4.2.8).

In addition to these 14 flags (one for each test), two more flags indicate data obtained from the Global Summary of the Day (GSOD) database and duplicates (see next section).

Geographical coordinates were verified semi-automatically by means of a Digital Elevation Model and a land mask; more than 100 sets of coordinates were corrected.

4.2.5 GSOD and duplicates

Global Summary of the Day constitutes an important source of GHCN-Daily. Daily data in GSOD are derived from synoptic/hourly observations, in some cases from only four synoptic observations per day. Hence, they are not actual daily extremes, but they have the advantage of being published in near real time and for countries that do not provide daily data. However, GSOD series are hampered by very large fractions of missing values, particularly in the early years (the oldest series dates back to 1929). Figure 4-4 provides an example of the difference between a GSOD minimum temperature series and a "real" daily series (note this station is not in the open release).

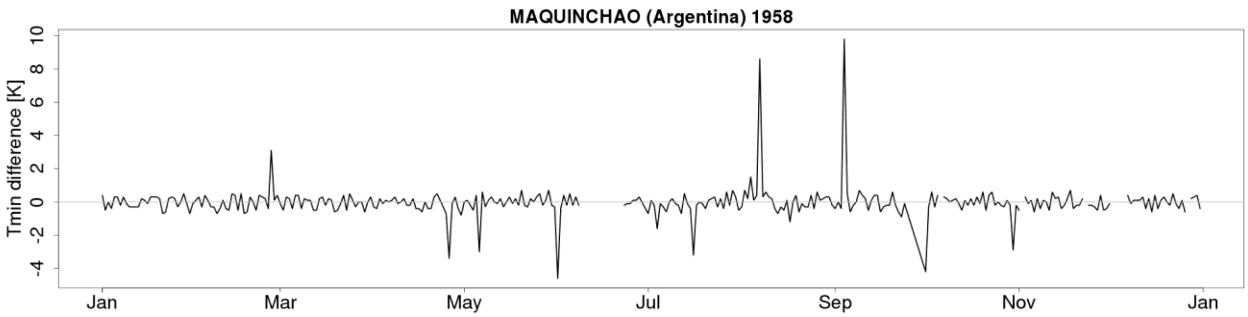


Figure 4-4: Difference between the minimum temperature series of Maquinchao in 1958 provided by the national weather service of Argentina and the minimum temperature series for the same station from the GSOD database. Data gaps are due to missing GSOD data. Note that GSOD data have a resolution of 1°F.

Duplicates are observations measured by the same instrument that are reported in two or more different series, sometime for different locations. This happens because the measurement and distribution of climate data is highly fragmented and disorganised, meaning that in most cases there is at least one intermediary between the station manager (i.e., national weather services and the like) and the global public collection. This also causes duplicate series to be slightly different from each other, depending on the time when they were acquired from the station manager and on the manipulation that they were subjected to before ending up in a certain global database. For the same reason, the measurement procedure (what instrumentation is used, time of observation, etc.) is usually unknown or very uncertain.

The EUSTACE station dataset does not include series that arise entirely from GSOD or are entirely duplicates of a longer series. When only part of a series is from GSOD or is a duplicate, then the entire series is retained and the respective quality flags are used for the affected part (see previous section).

4.2.6 Homogeneity assessment

A climate series is defined to be homogeneous when it contains only a signal related to climate variability. In reality, this is nearly impossible to achieve: the instruments, the observers, the procedures, the surroundings, and even the location of a station change over time. Hence, several artificial signals are usually part of a climate series. In most cases, these artificial signals are step-like functions caused by instantaneous changes in the station, so-called "breakpoints".

Even if "perfect" homogeneity cannot be achieved, small inhomogeneities are tolerable for most applications. Large inhomogeneities, on the other hand, can significantly affect results, in particular for trend analysis. The homogeneity assessment is the procedure used to estimate the degree of inhomogeneity of a climate series.

Breakpoints can be detected statistically by looking for sudden changes in some statistical property of the target series (usually the mean). This can be done in two ways: by analysing the target series directly (absolute test), or by analysing its deviations from a homogeneous reference series (relative test). The former strategy is risky, because climate itself can cause step-like changes that could be interpreted as breakpoints; the latter requires a suitable reference series (e.g., from a nearby station), which is not always available.

Since the degree of homogeneity of a reference series is in general also not known, the best way to perform a relative test is to repeat it several times using different reference series. Assuming that breakpoints occur randomly in every series, the agreement in the results will determine which breakpoints come from the target series (high agreement) and which from one of the reference series (low agreement).

EUSTACE brought this procedure a step further by looking at the agreement between different types of statistical tests, different variables, and different seasons.

Three different relative tests are applied to up to 8 reference series. A breakpoint is considered significant if at least 3 reference series agree on it. If not enough suitable reference series can be found, then an additional absolute test is applied.

Annual and semi-annual (October to March and April to September) means of Tmin, Tmax, Tmean, and the Daily Temperature Range (DTR) are analysed. Therefore, in the best case scenario 36 sets of breakpoints are provided for a certain series (3 tests X 3 temporal aggregations X 4 variables), or 12 sets when only the absolute test is applied.

Two additional sets (one for Tmin and one for Tmax) are obtained from "hints" for breakpoints in the data, namely changes in the reporting resolution, changes of source, and large gaps.

The user can combine the sets in any way that best suits his/her needs. However, a "recommended" merged set is also provided. The merged set combines all available sets to give the most likely position of the breakpoints (based on local maxima of detections). A measure of the relevance of the breakpoints, called "likelihood" index, is calculated from the number of detections from all sets within one year of a breakpoint. Figure 4-5 illustrates the merging with an example.

The likelihood index not only depends on the size of a breakpoint, but also on the ability of the method to detect it, given the information that is available (in particular, the reference series). An additional index, the "detectability" index, gives a measure for each year of the quantity and quality of reference series used. The detectability index is defined as the sum of the correlation coefficients between the reference series and the candidate series, hence its maximum value is 8. The results of the absolute test are used when the detectability index is lower than 4.

The breakpoints are given with annual resolution and their dates are all set on the 1st of January. As a consequence, an obvious limitation is that two breakpoints occurring in the same year or in consecutive years cannot be distinguished from each other. A finer detection of the breakpoints, for example by using visual methods, is recommended when attempting to apply corrections for the inhomogeneities.

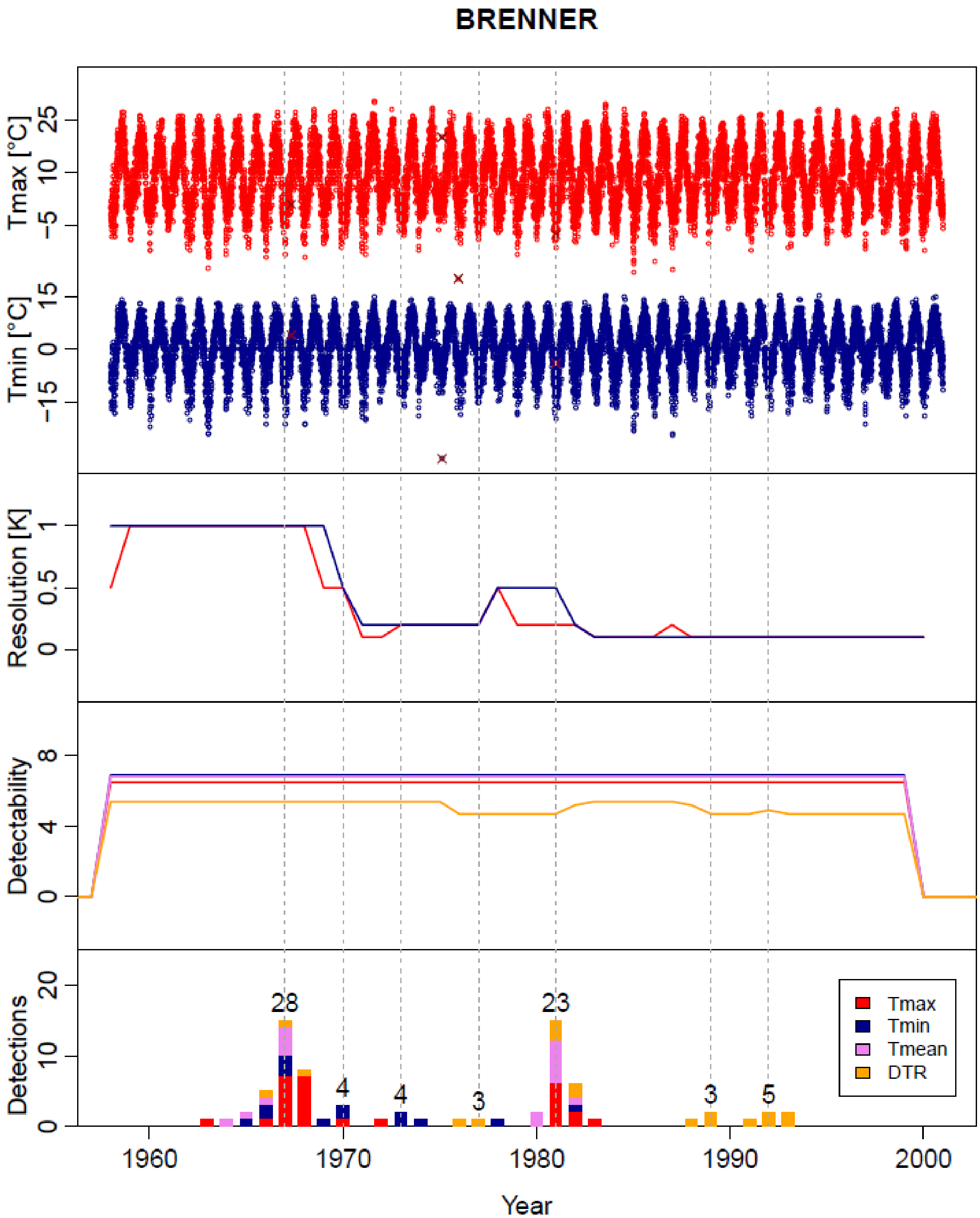


Figure 4-5: Overview of the information provided for the station of Brenner, Austria. From top to bottom: raw daily Tmax series, raw daily Tmin series, reporting resolution, annual detectability index, number of detections. Crossed values in the raw data series indicate observations flagged by the quality control. Vertical dashed lines indicate the position of the merged breakpoints and the numbers on top of the bars indicate their respective likelihood index.

4.2.7 Estimation of reporting resolution

Temperature data are usually provided with a resolution of 0.1 Kelvin. However, in most cases the resolution of the thermometer that measured the data was coarser. The actual reporting resolution is relevant in some applications, in particular when calculating indices based on percentiles, and random white noise should be added to the part of a series that has too coarse resolution.

The EUSTACE dataset provides a rough estimation of the reporting resolution for each year of each series. In complicated cases the reporting resolution can change on a daily basis (due, e.g., to different observers working at the same station): what is provided in the dataset is the estimation of the coarser resolution that was used during the target year. For instance, if the data have a resolution of 1 K in January and of 0.5 K in the rest of the year, the estimated resolution will be 1 K. Insufficient data in a certain year results in a missing value for the resolution.

Typically, the resolution is coarser at the beginning of a long series and improves over time (see example in Figure 4-5), especially where electronic thermometers have been introduced.

4.2.8 Detailed format specifications

The dataset is organised in 166 temperature files (one per year) containing the observations and the quality flags, plus one "status" file for all years containing additional information on the series (such as the breakpoints).

TEMPERATURE FILES

Table 4-17: Dimensions used in the temperature data files for the global land station data

Dimensions	Explanation
Time	days since 01-01-1850
Station	station progressive number (starting from 1, identical for every year)
name_strlen	length of station name (max 50 characters)
code_strlen	length of station code (max 50 characters)

Table 4-18: Variable names used in the temperature data files for the global land station data

Variable name	Explanation
station_name (station, name_strlen)	Name of the station (not necessarily unique)
station_code (station, code_strlen)	Official station ID from source (unique)
latitude (station)	Latitude in degrees North
longitude (station)	Longitude in degrees East
elevation (station)	Elevation above mean sea level in m
data_source (station)	Source ID: ***0. GHCN-DAILY VERSION 3.22***

	1. ***ISTI version 1.00*** 2. ***ECA&D (non-blended, updated October 2016)*** 3. ***ERA-CLIM and ERA-CLIM 2 projects*** 4. ***DECADE project*** 5. ***Homogenised series from Brugnara et al. (2016)*** 6. ***Data provided by National Weather Service of Argentina (SMN)***
data_policy (station)	Is the data re-distributable? (0=Yes, 1=No)
tasmax (time, station)	Daily Tmax in Kelvin
tasmin (time, station)	Daily Tmin in Kelvin
tasmax_qc (time, station)	QC flag, the following options are possible: **0 (ok):** Did not fail any quality assurance test **1-14 (capital letters):** QC flags as defined in GHCN-D (see Durre et al., 2010): D = failed duplicate check G = failed gap check I = failed internal consistency check K = failed streak/frequent-value check L = failed check on length of multiday period M = failed megaconsistency check N = failed naught check O = failed climatological outlier check R = failed lagged range check S = failed spatial consistency check T = failed temporal consistency check W = temperature too warm for snow X = failed bounds check Z = flagged as a result of an official Datzilla investigation **15 (GSOD):** Observation obtained from the Global Summary Of the Day (often unreliable) **16 (duplicate):** Observation already included in another (longer) series; used for partial duplicates
tasmin_qc (time, station)	See tasmax_qc

tasmax_definition (time, station)	Flag for how Tmax is defined, the following options are possible: **0-14 (TX--):** Definition codes from ECA&D: TX1: Maximum temperature unknown interval TX2: Maximum temperature 18-18 UT TX3: Maximum temperature 0-0 UT TX5: Maximum temperature morning previous day 06, 07, 08 until morning today (shifted 1 day back by ECA staff) TX6: Maximum temperature morning today 06, 07, 08 until morning next day TX7: Maximum temperature between 06 and 18 UT today TX8: Maximum temperature 21-21 CET TX9: Maximum temperature morning previous day 09 h GMT until morning today (shifted 1 day back by ECA staff) TX10: Maximum temperature from 2130 previous day until 2130 CET TX11: Maximum temperature morning today 9 UTC until morning next day TX12: Maximum temperature 19-19 UTC TX13: Maximum temperature within 00-24, 12-12 or 06-06 TX14: Maximum temperature 0-0 LT based on hourly intervals TX15: Maximum temperature 17-17 CET TX16: Maximum temperature 22-22 or 23-23 UT **15-20 (ISTI_---):** Definition codes from ISTI: ISTI_101: Daily value original ISTI_102: Daily value calculated from main standard synoptic observations (00, 06, 12, 18 UTC) ISTI_103: Daily value calculated from main and intermediate synoptic observations (00, 03, 06, 09, 12, 15, 18, 21 UTC) ISTI_104: Daily value calculated from other sub-daily observations (at least 3 obs available) ISTI_105: Daily value calculated from other sub-daily observations (at least 20 obs available) ISTI_999: Missing/Unknown/Not Applicable **-128 (missing value):** No information available on definition

tasmin_definition (time, station)	Flag for how Tmin is defined, the following options are possible: **0-14 (TN--):** Definition codes from ECA&D: TN1: Minimum temperature unknown interval TN2: Minimum temperature 18-18 UT TN3: Minimum temperature 0-0 UT TN5: Minimum temperature morning previous day 06, 07, 08 until morning day TN6: Minimum temperature between 18 UT previous day and 06 UT today TN8: Minimum temperature 21-21 CET TN9: Minimum temperature morning previous day 09 h GMT until morning today TN10: Minimum temperature from 2130 previous day until 2130 CET TN11: Minimum temperature 19-19 UTC TN12: Minimum temperature within 00-24, 12-12 or 18-18 TN13: Minimum temperature 0-0 LT based on hourly intervals TN14: Minimum temperature 17-17 CET TN15: Minimum temperature 22-22 or 23-23 UT **15-20 (ISTI_---):** Definition codes from ISTI: ISTI_101: Daily value original ISTI_102: Daily value calculated from main standard synoptic observations (00, 06, 12, 18 UTC) ISTI_103: Daily value calculated from main and intermediate synoptic observations (00, 03, 06, 09, 12, 15, 18, 21 UTC) ISTI_104: Daily value calculated from other sub-daily observations (at least 3 obs available) ISTI_105: Daily value calculated from other sub-daily observations (at least 20 obs available) ISTI_999: Missing/Unknown/Not Applicable **-128 (missing value):** No information available on definition

STATUS FILE

Table 4-19: Dimensions used in the status files for the global land station data

Dimensions	**Explanation**
detection_time	time dimension (annual resolution) for the detection score (days since 01-01-1850 of the first day of the year)
resolution_time	time dimension (annual resolution) for the reporting resolution (days since 01-01-1850 of the first day of the year)
Station	station progressive number (starting from 1, identical to the temperature files)
name_strlen	length of station name (max 50 characters)
code_strlen	length of station code (max 50 characters)
tas_break	breakpoint progressive number for Tmean (starting from 1)
tasmax_break	breakpoint progressive number for Tmax (starting from 1)
tasmin_break	breakpoint progressive number for Tmin (starting from 1)
tasdtr_break	breakpoint progressive number for DTR (starting from 1)
merged_break	merged breakpoint progressive number (starting from 1)

Table 4-20: Variable names used in the status files for the global land station data

Variable names	Explanation
station_name (station, name_strlen)	Name of the station (not necessarily unique)
station_code (station, code_strlen)	Official station ID from source (unique)
latitude (station)	Latitude in degrees North
longitude (station)	Longitude in degrees East
elevation (station)	Elevation above mean sea level in m
data_source (station)	Source ID: as above
tasmax_reporting_resolution (resolution_time, station)	Worst resolution of Tmax observations within a year
tasmin_reporting_resolution (resolution_time, station)	Worst resolution of Tmin observations within a year
tas_break_time (tas_break)*	Year in which a breakpoint in Tmean was detected (days since 01-01-1850 of the first day of the year)
tas_break_station (tas_break)*	Station at which a breakpoint in Tmean was detected (station number)
tas_detectability (detection_time, station)*	Detectability index for Tmean
tas_break_type (tas_break)*	Type of breakpoint for Tmean (0=from the relative tests, 1=from the absolute test, 2=from metadata)
tas_break_season (tas_break)*	Temporal aggregation used to detect a breakpoint for Tmean (0=annual means, 1=Oct-Mar, 2=Apr-Sep)
tas_break_count (tas_break)*	Number of relative tests that detected a breakpoint for Tmean (1-3)
merged_break_time (merged_break)	Year of a merged breakpoint (days since 01-01-1850 of the first day of the year)
merged_break_station (merged_break)	Station at which a merged breakpoint was detected (station number)
merged_break_likelihood (merged_break)	Likelihood index for a merged breakpoint
detection_feasibility (station)	Feasibility of detection at a certain station (0=not possible, 1=only absolute test, 2=all tests)

* These variables are repeated for tasdtr (DTR), tasmax (Tmax) and tasmin (Tmin).

To extract breakpoints information for a certain series, the breakpoints IDs must be obtained from the desired *_break_station variable. Breakpoints affecting a certain series have adjacent ID (hence, the ID of the first breakpoint and the number of breakpoints are sufficient to extract the breakpoints information). What follows is an example of R code to extract the dates of the merged breakpoints shown in Figure 4-5

```
# load NetCDF library
library(ncdf4)

# target station code (see inventory file)
station <- "TX_SOUID103851"

# open NetCDF status file
nc <- nc_open("<path-to-the-status-file>")

# get station id
st_id <- which(ncvar_get(nc, "station_code") == station)
```

```
# get breakpoints ids
bp_id <- which(ncvar_get(nc, "merged_break_station") == id)

# get breakpoints dates using first id and number of breakpoints
breakpoints <- ncvar_get(nc, "merged_break_time", start = bp_id[1], count =
length(bp_id))

# convert to date format
breakpoints <- as.Date(breakpoints, origin=”1850-01-01”)
```

4.2.9 Guidance on the use of flags

Observations that have a quality flag different from zero should be discarded by most users. Some quality issues can be overlooked by the automatic quality control: when working on single series or a smal l subset, manual inspection of the series is recommended.

Some users might choose to ignore the flag 3 (I = failed internal consistency check), because this flag can arise from a non-standard definition of the maximum or minimum temperature (for instance, when the minimum temperature refers only to night time and the maximum to the daytime). In fact, the quality assurance algorithm assumes that both maximum and minimum temperature refer to the same 24-hour period. Hence, a series that was produced using a different definition is more likely to fail the internal consistency check. When this is the case, the flagged data might still be useful for some applications such as extreme analysis. However, it is worth remarking that the observations that have failed the internal consistency check did not undergo most of the other quality tests (see Durre et al., 2010, for details on the quality assurance algorithm).

The exact definition of the observations is provided for only a minority of series (from ECA&D) and is not always reliable, particularly for long series (for which the definition might have changed over time). When this information is relevant, it should be requested from the data owner, indicated in the documentation of the source collections. In some cases, however, tracing the data owner might not be possible: the best bet is then to contact the national weather service of the interested country.

GSOD data (quality flag 15) should be used carefully. It is not recommended to use these data when the entire dataset or a large subset is analysed. Duplicated data (quality flag 16) might be useful when analysing single series; however, the flags from the quality assurance are not provided for these data, therefore an additional quality control is necessary.

4.2.10 Known limitations

Metadata, the information on where and how the data was measured, is very scarce due to the way climate data have been shared so far. Some of the data (most of those measured before the 1950s) were not measured following modern WMO standards, therefore the absolute values might not be directly comparable with those measured by modern weather stations. Besides, the precision of the geographical coordinates of the stations is often very poor and some erroneous coordinates have certainly been overlooked by the quality tests.

The quality assurance is made using a fully automatic algorithm that is "tuned" to be rather conservative (i.e., to avoid false alarms). Hence, many erroneous values are not detected by the algorithm.

The automatic detection of breakpoints is not perfect either, even when good reference series are available. For instance, breakpoints located at the edges of a series are less likely to be detected than those occurring in the middle. Moreover, as already mentioned, breakpoints occurring a few months from each other cannot be distinguished. For details on the performance of the breakpoint detection the reader is referred to the scientific user guide.

4.2.11 **Example of use of breakpoints**

A typical use of daily temperature data is the analysis of trends in extreme indices. Here we show an example of how the reliability of these trends can be assessed using the likelihood index.

The upper panel of Figure 4-6 shows the trends of the number of hot days (days with Tmax above the 90th percentile or "tx90p" index as calculated with the R climdex.pcic package (https://cran.r-project.org/web/packages/climdex.pcic/index.html) in all series in eastern Asia that are long (at least 48 years) and had sufficient reference series for the relative breakpoint detection. Generally the trends are positive (red colour), with some exceptions particularly in eastern China and south western Japan.

In the bottom panel we show only the trends for those series that do not contain merged breakpoints with a likelihood index greater or equal to 10 (note that this value does not have a particular meaning, higher or lower thresholds would give similar results), which we can consider "quasi-homogeneous". Negative trends in China are mostly still there, while those in Japan have all disappeared. We can conclude that weak negative trends in hot days are locally observed in China, while in Japan they are the result of inhomogeneities in some stations.

The lower the likelihood threshold, the higher the probability that the series are homogeneous. However, a too low threshold will discard the large majority of the series, leaving not enough of them for a meaningful analysis of spatial patterns.

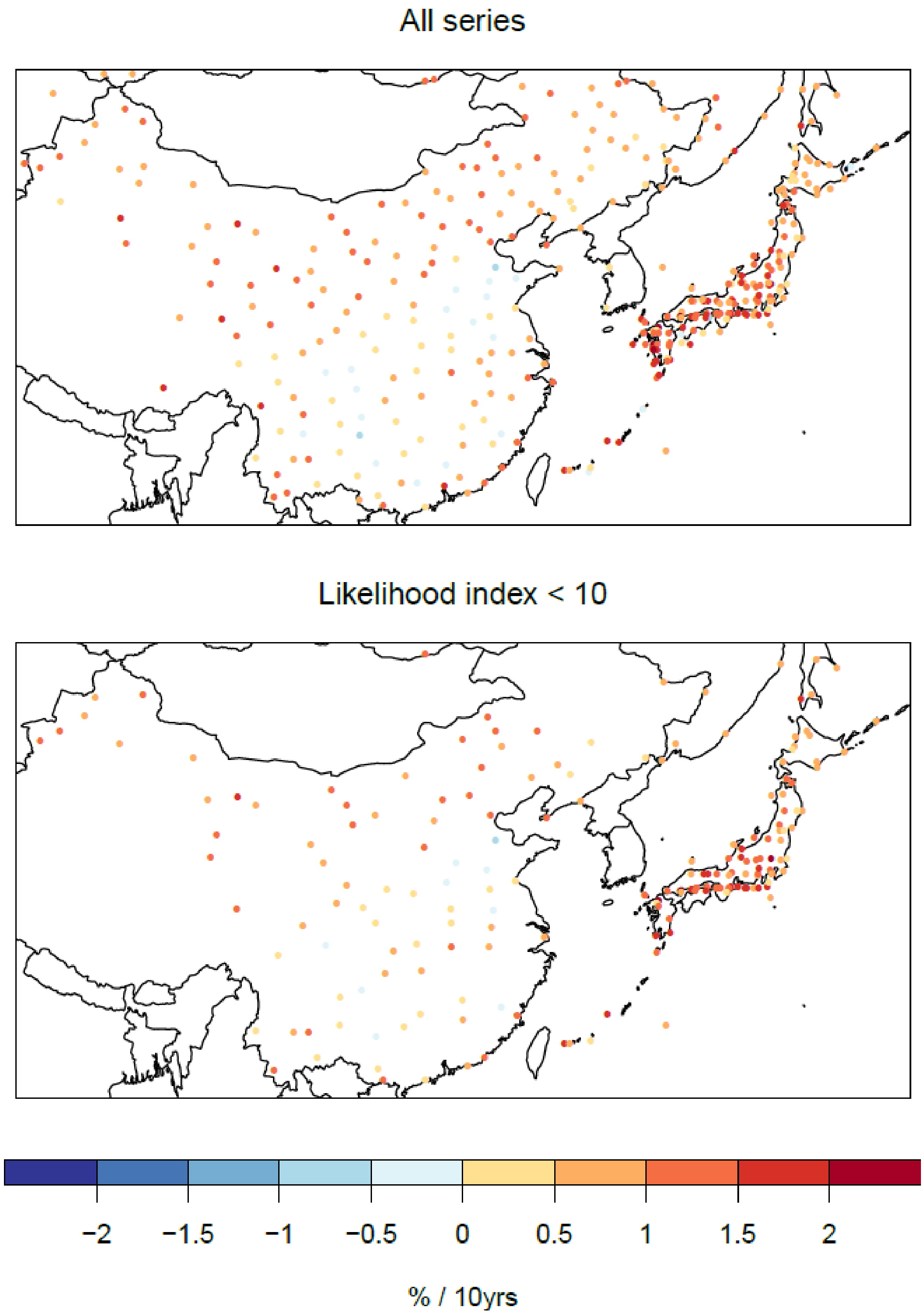


Figure 4-6: Trends in the number of hot days over the period 1951-2010 for all series that underwent the breakpoint detection with relative tests (top panel) and for those that do not have inhomogeneities with a likelihood index larger or equal to 10 (bottom panel).

4.2.12 Where to go for further information

- Scientific User Guide (under development)
- GHCN-Daily website: https://www.ncdc.noaa.gov/ghcn-daily-description
- ISTI website: http://www.surfacetemperatures.org/
- ECA&D website: https://www.ecad.eu/
- DECADE page: http://www.geography.unibe.ch/research/climatology_group/research_projects/decade/index_eng.html

4.2.13 References

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010). Comprehensive automated quality assurance of daily surface observations. Journal of Applied Meteorology and Climatology, 49, 1615-1633.

Hunziker, S., Gubler, S., Calle Fernandez, J. M., Moreno, I., Andrade, M.F., Velarde, F., Carrasco, G., Castellón, Y., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., & Brönnimann, S. (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. International Journal of Climatology, 37, 4131-4143.

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, 1441-1453.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. Journal of Atmospheric and Oceanic Technology, 29, 897-910.

Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., et al. (2014). The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. Geoscience Data Journal, 1, 75-102.

Stickler, A., Brönnimann, S., Valente, M. A., Bethke, J., Sterin, A., Jourdain, S., et al. (2014). ERA-CLIM: historical surface and upper-air data for future reanalyses. Bulletin of the American Meteorological Society, 95, 1419-1430.

4.3 European land station daily air temperature measurements, homogenised

4.3.1 Summary

Within the EUSTACE project, the daily temperature data from stations in the European Climate Assessment & Dataset (ECA&D, www.ecad.eu, Klein Tank et al. 2002) have been homogenized to adjust for e.g. changes in location of the stations or changes in measurement equipment. These data are made available as a separate dataset https://www.ecad.eu//dailydata/predefinedseries.php (also other station data files are provided through the same website: check carefully that you have the homogenized temperature).

4.3.2 History and background

The dataset from ECA&D is a set of data provided by the European National Meteorological Services and other data holding institutes. ECA&D was initiated by the European Climate Support Network (ECSN) in 1998 as part of an EUMETNET activity, where EUMETNET is the grouping of 31 European National Meteorological Services (NMSs) that provides a framework to organise co-operative programmes between its members in the various fields of basic meteorological activities and has received financial support from the EUMETNET and the European Commission.

ECA&D is receiving data from 69 participants for 63 countries and the ECA dataset contains 55783 series of observations for 14 elements at 15778 meteorological stations throughout Europe and the Mediterranean at the time of writing (March 2019). 80% of these daily series can be downloaded from this website for non-commercial research and education. ECA&D forms the backbone of the climate data node in the Regional Climate Centre (RCC) for WMO Region VI (Europe and the Middle East) since 2010. The data and information products contribute to the Global Framework for Climate Services (GFCS).

The coverage of temperature stations (Figure 4-7) is spatially inhomogeneous. Many NMSs provide their complete networks to ECA&D while others provide a subset. The sharing of data for scientific research is made possible by WMO Resolutions 40 and 60. Nevertheless, Figure 4-7 shows a distinction between stations for which daily data can be shared with the scientific community and stations for which only derived data can be shared. Derived data are e.g. station-based monthly mean values or gridded data.

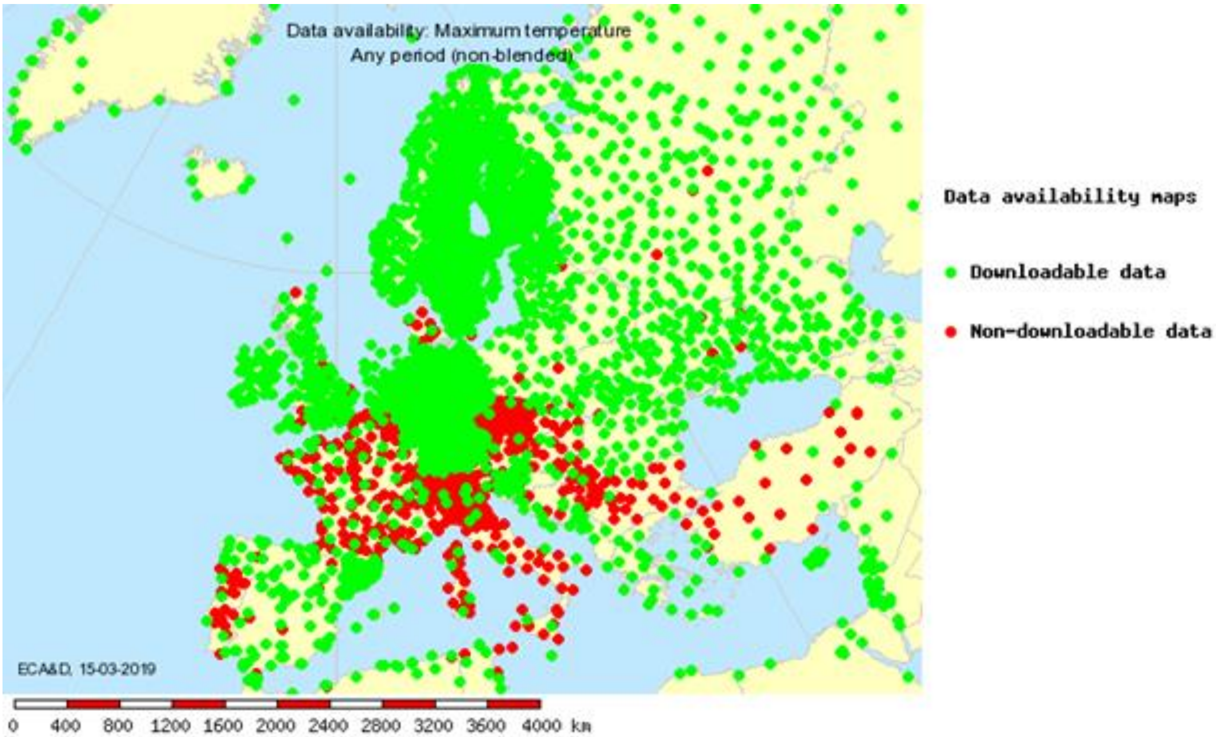


Figure 4-7: Coverage of stations which provide daily values for maximum temperature in ECA&D. The colour coding relates to data policy issues of the data owner, often the National Meteorological Institute. points which are green: daily data can be shared for non-commercial research and education. Red, only derived data can be shared.

4.3.3 Data access

The data are available from ECA&D at www.ecad.eu using the link in the table below. Data are made available for non-commercial purposes only.

Table 4-21: Links to the EUSTACE / ECA&D: European land station daily air temperature measurements, homogenised

Data Product	CEDA catalogue record	Download links
EUSTACE / ECA&D: European land station daily air temperature measurements, homogenised	https://catalogue.ceda.ac.uk/uuid/81784e3642bd465aa69c7fd40ffe1b1b	https://www.ecad.eu//dailydata/predefinedseries.php (Homogeneous blended ECA dataset section)

4.3.4 Detailed format specifications of the data files and metadata

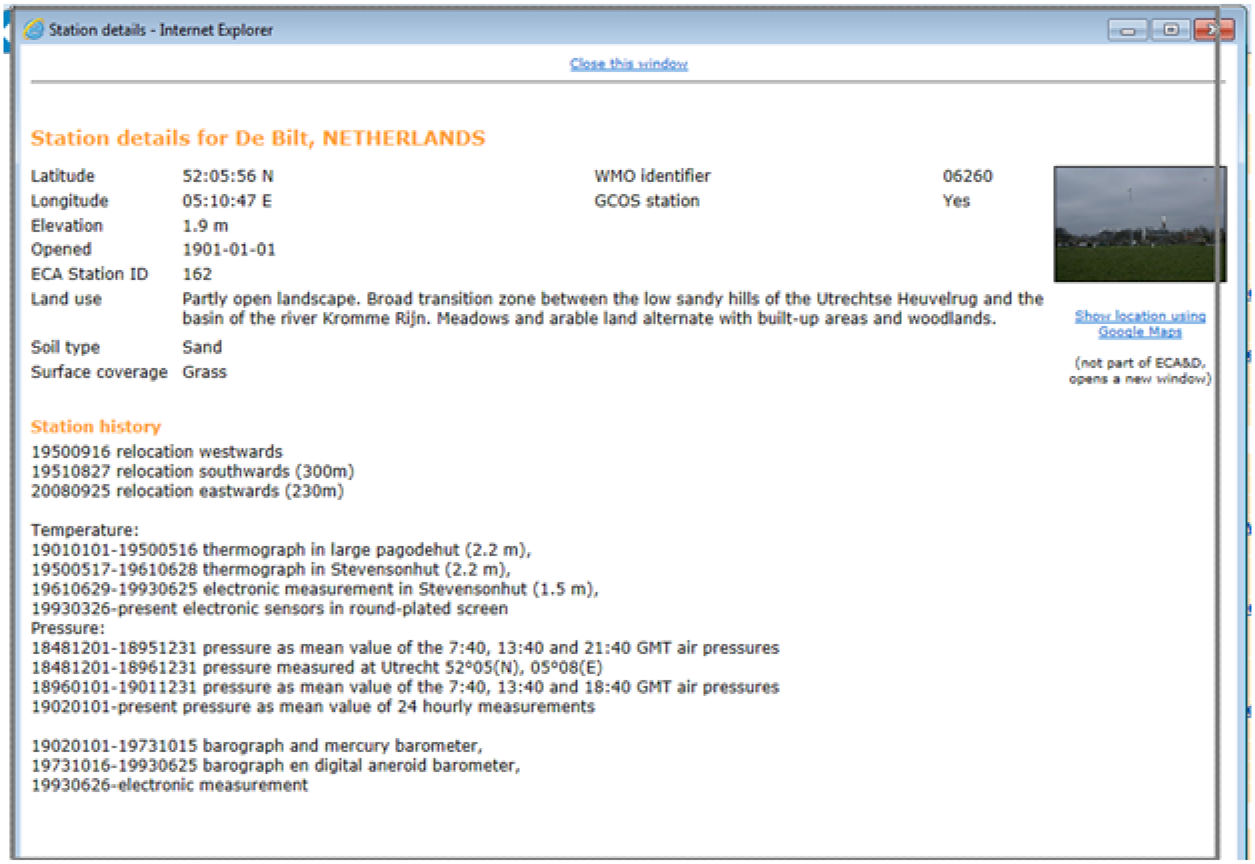
The ECA&D dataset developed within EUSTACE is provided with one file per station per element. In the header of the file, the metadata of the station (location, name etc.) and the measurement (time period for which the measurement is valid etc.) are added, a description of the columns and the preferred reference to use when citing the data (Figure 4-8). When the whole dataset is downloaded, separate files providing metadata on all stations and elements are included as well.

In addition, a file is provided with information on the timing of the detected breaks.

Although the breaks in the series are detected by a statistical procedure, there is metadata available relating to changes in the location of the station or other types of changes. These metadata records are provided directly by the NMSs and can be accessed from the ECA&D website. One example is given for De Bilt (the Netherlands) in Figure 4-9, which can be accessed through: https://www.ecad.eu/utils/stationdetail.php?stationid=162.

```
TX_STAID000001 - Notepad
File  Edit  Format  View  Help
EUROPEAN CLIMATE ASSESSMENT & DATASET (ECA&D), file created on: 10-03-2019
THESE DATA CAN BE USED FOR NON-COMMERCIAL RESEARCH AND EDUCATION PROVIDED THAT THE FOLLOWING SOURCE IS ACKNOWLEDGED

Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface
air temperature and precipitation series for the European Climate Assessment.
Int. J. of Climatol., 22, 1441-1453.
Data and metadata available at http://www.ecad.eu

FILE FORMAT (MISSING VALUE CODE = -9999):

01-06 STAID: Station identifier
08-13 SOUID: Source identifier
15-22 DATE : Date YYYYMMDD
24-28 TX   : Maximum temperature in 0.1 &#176;C
30-34 Q_TX : quality code for TX (0='valid'; 1='suspect'; 9='missing')

This is the blended series of station VAEXJOE, SWEDEN (STAID: 1)
Blended and updated with sources:2 35382
See files sources.txt and stations.txt for more info.

STAID, SOUID,    DATE,   TX, Q_TX
     1,      2,19180101,    0,    0
     1,      2,19180102,  -40,    0
     1,      2,19180103,  -75,    0
     1,      2,19180104,  -55,    0
     1,      2,19180105,   25,    0
     1,      2,19180106,   22,    0
     1,      2,19180107,    5,    0
     1,      2,19180108,   -5,    0
     1,      2,19180109, -120,    0
     1,      2,19180110,  -40,    0
```

Figure 4-8: Screen shot of the setup of a data file (not a homogenized one in this case) with the described information in the heading and the daily data below it. The setup of the homogenized data files may be slightly different.


Station details - Internet Explorer

Close this window

Station details for De Bilt, NETHERLANDS

Latitude	52:05:56 N	WMO identifier	06260
Longitude	05:10:47 E	GCOS station	Yes
Elevation	1.9 m		
Opened	1901-01-01		
ECA Station ID	162		
Land use	Partly open landscape. Broad transition zone between the low sandy hills of the Utrechtse Heuvelrug and the basin of the river Kromme Rijn. Meadows and arable land alternate with built-up areas and woodlands.		
Soil type	Sand		
Surface coverage	Grass		

Show location using Google Maps

(not part of ECA&D, opens a new window)

Station history

19500916 relocation westwards
19510827 relocation southwards (300m)
20080925 relocation eastwards (230m)

Temperature:
19010101-19500516 thermograph in large pagodehut (2.2 m),
19500517-19610628 thermograph in Stevensonhut (2.2 m),
19610629-19930625 electronic measurement in Stevensonhut (1.5 m),
19930326-present electronic sensors in round-plated screen
Pressure:
18481201-18951231 pressure as mean value of the 7:40, 13:40 and 21:40 GMT air pressures
18481201-18961231 pressure measured at Utrecht 52°05(N), 05°08(E)
18960101-19011231 pressure as mean value of the 7:40, 13:40 and 18:40 GMT air pressures
19020101-present pressure as mean value of 24 hourly measurements

19020101-19731015 barograph and mercury barometer,
19731016-19930625 barograph en digital aneroid barometer,
19930626-electronic measurement

Figure 4-9: Screen dump of the webpage showing metadata for station De Bilt which can be matched with the statistically detected breaks (often indicating relocations of the station etc.).

4.3.5 Explanation of uncertainties and quality

The station data in ECA&D are subjected to a basic Quality Control (QC) procedure (ECA&D, 2012). There is an inherent uncertainty in observed data. Several sources of uncertainty are discussed below.

4.3.5.1 Location of the station

A poorly positioned station will record temperatures which are relevant for the immediate surroundings of the thermometer but not for the wider surroundings. This is the motivation for the guidance of the World Meteorological Organisation to locate meteorological stations in an open field, away from disturbances (WMO, 2011). Relevant to this issue is the Urban Heat Island effect, since this may give an additional non-climate related warming to the measurements.

There are several estimates of the urbanization effects. For the Netherlands, Brandsma et al. (2003) find an urbanization trend in station De Bilt of approx. 0.1˚C per century while for Central London, Jones & Lister (2009) find that warming trends are not significantly different from those at rural sites. Estimates for larger areas vary likewise, with an increase of 0.05˚C from 1900 to 1990, as suggested to be present in the HadCRUT data set (Folland et al., 2001), while Parker (2004) concludes that large-scale warming due to urban effects is too small to detect.

4.3.5.2 Measurement error

The random error in a single thermometer reading has been estimated at 0.2˚C (Folland et al. 2001) which is a realistic estimate of the measurement error in the daily maximum (Tx) and minimum (Tn) temperatures. Daily averaged temperatures are based at least on two measurements (Tx and Tn), which would make the error in this measurement smaller (by a factor 1/sqrt(2)).

4.3.5.3 Calculation and reporting error

Station data are quality checked before inclusion in the database but not adjusted. Data are flagged when they fail the quality test and discarded for further use. However, it is possible that in reporting the data to ECA&D, errors are introduced, like reporting 29.1˚C instead of 19.1˚C.

Many of these errors will be identified by the quality control. Those that remain, are likely to be small and because they are also uncorrelated in time and space, will have a negligible effect on the conclusions based on the complete dataset. Following Brohan et al. (2006) this error will not be considered further.

4.3.5.4 Uncertainty in the homogenization and break detection

Squintu et al. (2019a) identify the issue with uncertainties in the statistical detection of break related to the timing of the breaks. The uncertainty in the adjustment of the break is discussed as well. There are no established methods to quantify uncertainties in these steps.

4.3.6 **Guidance on the homogenised station data and known limitations**

The daily maximum and minimum temperature series of the European Climate Assessment & Dataset are homogenised using the quantile matching approach. As the dataset is large and the detail of metadata is generally missing, 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 reference. A total of 6,500 series have been processed and after removing duplicates and short series, about 2,100 series have been adjusted. This process is documented by Squintu et al. (2019a).

For climatological purposes, it is worthwhile to have long and homogeneous series. Relocation of measuring equipment from a station to another, like from the city centre to the rural area or a nearby airport, is one of the causes of discontinuities in the temperature measurements which may affect trend estimates. Long series suitable for long-term climatological studies can be built by gathering and combining data from nearby stations. For the EUSTACE dataset, an updated procedure for the composition of long series is used documented by Squintu et al. (2019b).

These homogenisation procedures have the drawback in areas where the station density is low, that an insufficient number of neighbouring reference stations can be found for a reliable break detection and homogeneity adjustment. This holds in particular for the early part of the records, when the number of stations is low.

Break detection is done using annual averages. This may mean that possible breaks which are apparent only in a few months or one season may go unnoticed. We do see that the amplitude of the homogeneity adjustment has a seasonal cycle, and possibly the use of annual averages hides smal ler discontinuities from the detection.

Finally, measurement practices have changed over time and they change between countries. This means that e.g. the daily mean temperature may be based on 8 3-hourly values for one period in time and on 24 1-hourly measurements for a more recent period. When such changes are picked-up by the break detection, a homogeneity adjustment is made but it is likely that these more nuanced changes go unnoticed for the break detection.

4.3.7 Further information

A general source of information for ECA&D is the website (www.ecad.eu) and staff members can be contacted at eca@knmi.nl for questions relating to the data.

4.3.8 References

Brandsma, T., Können, G. P. and Wessels, H. R. A. (2003) Emperical estimation of the effect of urban heat advection on the temperature series of De Bilt (The Netherlands) Intern. J. Climatol. 23:829-845, doi:10.1002/joc.902

Brohan, P., Kennedy, J. J., d Harris, I. , Tett, S. F. B. and Jones, P. D. (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J. Geophys. Res. (Atmospheres)_ 111:D12106, doi:10.1029/2005JD006548

ECA&D Project Team (2012) European Climate Assessment & Dataset Algorithm Theoretical Basis Document (ATBD), Royal Netherlands Meteorological Institute KNMI, De Bilt, NL version 10.5

Folland, C. K., Rayner, N. A., Brown, S. J. , Smith, T. M., Shen, S. S. P. , Parker, D. E., Macadam, I., Jones, P. D., Jones, R. N., Nicholls, N. and Sexton, D. M. H. (2001) Global temperature change and its uncertainties since 1861. Geophys. Res. Lett 28:2621-2624, doi:10.1029/2001GL012877

Jones, P. D. and Lister, D. H. (2009) The urban heat island in Central London and urban-related warming trends in Central London since 1900, Weather 64:323-327, doi:10.1002/wea.432

Klein Tank et al., 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. Climatol. 22: 1441–1453. https://doi.org/10.1002/joc.773.

Parker, D. E. (2004) Large scale warming is not urban. Nature 432:290

Rayner, N. A. et al., (2019) The EUSTACE project: delivering global, daily information on surface air temperature, in preparation for Bulletin of the American Meteorological Society

Squintu, A.A., G. van der Schrier, Yuri Brugnara, Albert Klein Tank, 2019a: Homogenization of daily temperature series in the European Climate Assessment & Dataset, International Journal of Climatology, 39:1243-1261. doi.org/10.1002/joc.5874

Squintu, A.A., G. van der Schrier, E. J. M. van den Besselaar, R. C. Cornes, A. Klein Tank, 2019b. Building long homogeneous temperature series across Europe: a new approach for the blending of neighboring series, submitted to the Journal of Applied Meteorology and Climatology

World Meteorological Organization (2011) Guide on Climatological Practices, World Meteorological Organization, Geneva, Switzerland, WMO-No. 100

4.4 Gridded European surface air temperature based on homogenised meteorological station records since 1950

4.4.1 Summary

E-OBS is a daily gridded observational dataset for precipitation, temperature and sea level pressure in Europe. The input data are the time series from the blended time series in ECA&D (https://www.ecad.eu). The ensemble version (indicated with a 'e' after the version number) is for 5 elements (daily mean temperature TG, daily minimum temperature TN, daily maximum temperature TX, daily precipitation sum RR and daily averaged sea level pressure PP). The data files are in NetCDF-4 format. The Global 30 Arc-Second Elevation Data Set (GTOPO30), a global raster Digital Elevation Model (DEM) with a horizontal grid spacing of 30 arc seconds (approximately 1 kilometre) developed by USGS is used for the elevation file as well.

The ensemble dataset is constructed through a conditional simulation procedure. For each of the 100 members of the ensemble a spatially correlated random field is produced using a pre-calculated spatial correlation function. The mean across the 100 members is calculated and is provided as the "best-guess" fields. The spread is calculated as the difference between the 5th and 95th percentiles over the ensemble to provide a 90% uncertainty range.

For the EUSTACE product, a preliminary version of the E-OBS using the homogenized station discussed in section 4.3 has been calculated. This version, EOBSv19.0eHOM, is made available through the ECA&D website (www.ecad.eu) and the homogenized data will be used in future versions of EOBS from version 20 onward.

4.4.2 History and background

In the EU-FP6 project ENSEMBLES, the E-OBS was initially developed as a pan-European gridded dataset for validation and calibration of Regional Climate Models (Van der Linden and Mitchell, 2009). Since then, the E-OBS has grown to a position with widespread use in the scientific community, with over 1800 citations in the literature to this dataset (status December 2018). The E-OBS is now also routinely used for Climate Monitoring purposes (Figure 4-10).The initial code, developed by Haylock et al. (2008) was recently replaced by one where extremes are preserved a little better and the uncertainty in the gridding is captured in terms of an ensemble (Cornes et al. 2018).

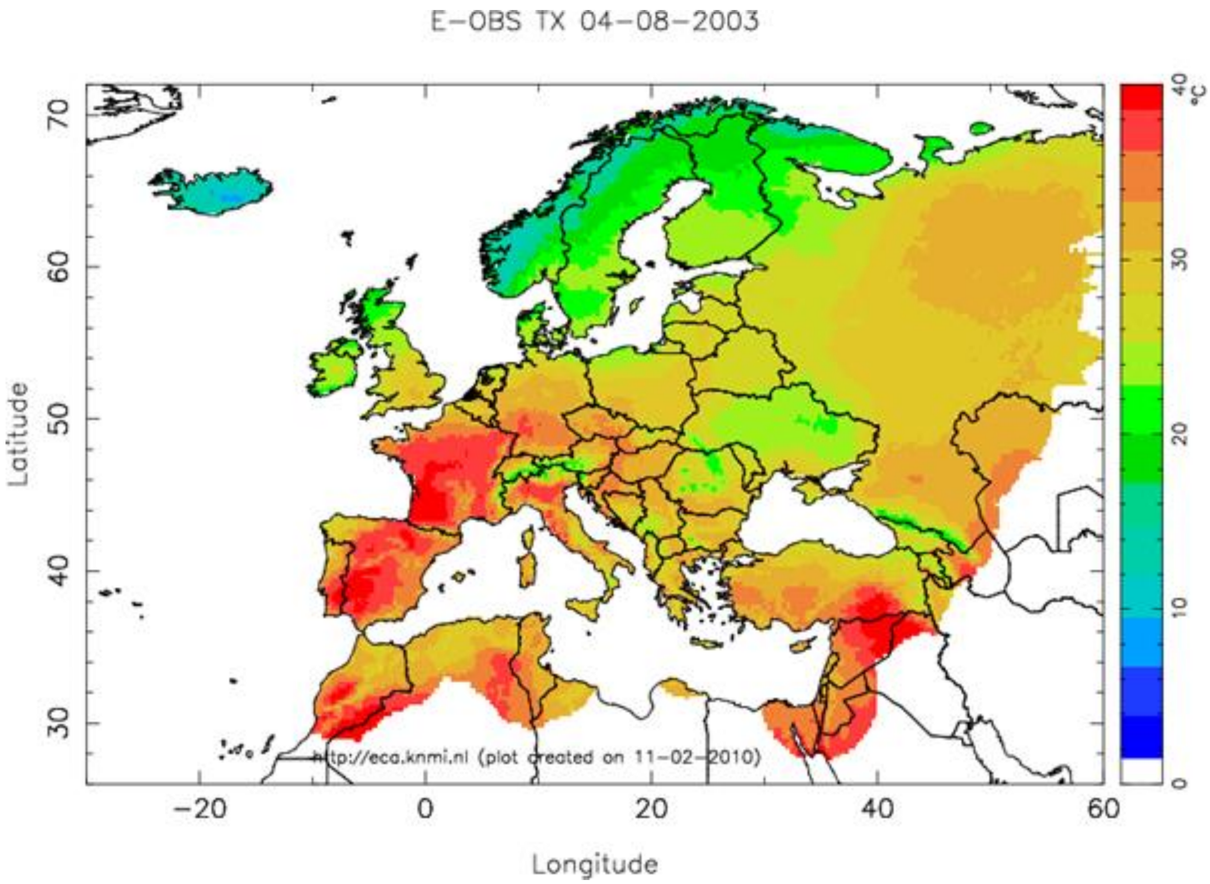


Figure 4-10: An example of the E-OBS, showing the daily maximum temperature on August 4 2003 in the middle of the 2003 heat wave.

4.4.3 Access to data

The data are made available through the ECA&D website at www.ecad.eu for use for non-commercial purposes.

Table 4-22: Links to the EUSTACE / ECA&D: European land station daily air temperature measurements, homogenised

Data Product	CEDA catalogue record	Download links
EUSTACE / E-OBS: Gridded European surface air temperature based on homogenised land station records since 1950	https://catalogue.ceda.ac.uk/uuid/b2670fb9d6e14733b303865c85c2065d	https://www.ecad.eu/download/ensembles/download.php (EUSTACE version is 19.0eHOM)

4.4.4 Data file format

All E-OBS data files are in NetCDF-4 format. These data files contain the necessary metadata for further processing in graphical routines or other types of software. Each file contains information on one variable, for the complete time period (from January 1, 1950 onward) at the daily temporal resolution and with a spatial resolution of 0.1˚. For EUSTACE, daily values of maximum, mean and minimum temperatures are provided and relate to the 'best guess' estimate.

The uncertainty information is provided in an ensemble which consists of 100 realizations of each daily field. For EUSTACE, the spread in the ensemble is made available as a measure of the standard uncertainty. This should be interpreted as a measure of uncertainty of the best-guess value. The best guess and the standard uncertainty are provided in separate files.

4.4.5 Construction of the E-OBS ensemble

An elaborate guidance document of the ensemble E-OBS can be accessed at: http://surfobs.climate.copernicus.eu/userguidance/use_ensembles.php.

Ensemble datasets are climate datasets that consist of a number of equally probable realizations, and relate to data in gridded format. The ensemble aims to give a measure of uncertainty in the data field; such datasets are widely used in a number of areas of climate change science. The two most widely known examples are:

- Perturbed physics simulations of climate models with different values of key parameters (e.g. climateprediction.net, but undertaken now by most modelling centres);
- Numerical weather predictions, where it is known as ensemble forecasting. This area is extending to seasonal-to-interannual predictions.

Other examples exist in other fields. The idea behind the above two examples is to consider the effects of slight differences in boundary conditions and/or in the values of key parameters within the physics simulations.

The term "ensemble" is also used when referring to multiple realizations from gridded observed datasets. Such datasets are formed from the interpolation of station values. Although the aim is the same as ensembles calculated for model simulations - to quantify uncertainty in the data - the generation and hence interpretation of the realizations is quite different. With gridded observational datasets, decisions have to be made for a number of features that affect the final gridded field of the key parameters involved in the gridding algorithm. These could include:

- Search radius for inclusion of stations influencing a grid box;
- Estimates of the impacts of homogeneity issues on the quality of the input station data;
- Impact of a number of possible co-variates used in the gridding (e.g. latitude, longitude, elevation and distance from coast or inland water body)

- Including station data where the base period (e.g. 1961-90 or 1981-2010) is less precisely known.

The usual approach in the production of a gridded dataset is to determine the best or most likely values for the key parameters to interpolate station values to values on a regular grid. With an ensemble version of the dataset, a number of these key parameters can be varied producing a range of possible gridded datasets (referred to as the ensemble). An example of this approach has been produced for the gridded global temperature dataset (HadCRUT4) by Morice et al. (2012), where 100 ensemble members were developed. The advantage of producing a dataset this way is that with the ensemble it can be easier to determine the errors of the estimate per grid-box, regional and hemispheric averages. Using standard statistical approaches this can be difficult as a number of the error components have spatial and temporal structures which are difficult to model.

Although all the examples given use the term "ensemble dataset", the most analogous to what we propose is HadCRUT4 or the EUSTACE Global air temperature estimates but the way the key parameters are varied is not the same.

4.4.6 Guidance on the use of E-OBS dataset

The E-OBS dataset consists of daily interpolations of temperature (maximum, mean and minimum daily values), precipitation and Mean Sea-level Pressure. Based on 100 realizations of each daily field, the mean across the ensemble and the standard error across the realizations is provided. The ensemble mean provides a "best guess" value, and the standard error provides a measure of uncertainty of the best-guess value.

Where users require a single measure of the interpolated daily fields, then the "best guess" values should be used. However, the standard error should always be consulted as the uncertainty of the gridded field varies across the domain, and is ultimately determined by the variations in station coverage.

The individual ensemble members are mainly intended for users who require the uncertainty in the gridded fields to propagate through to various other applications. For example, in Figure 4-11 the first four ensemble members of the E-OBS precipitation dataset are displayed for the heavy rainfall event of 1st June 2013 across central Europe. If a user requires this rainfall data for hydrological modelling then each of the ensemble members could be fed into the hydrological model. In this way the uncertainty in the rainfall interpolation would propagate through to the hydrological model output (this is also true for temperature data).

When one wants to calculate the uncertainty for a derived variable e.g. the number of days with a maximum temperature higher than 25 °C, one also has to use the ensemble to determine the standard deviation for the calculated derived index. In this case the uncertainty cannot be calculated using only the "best guess" and the daily standard deviations.

Although the E-OBS is based on point measurements, this dataset provides area-averaged values. This makes it easier to compare against satellite data, where skin temperatures are also more an area-averaged quantity than a point value (see Sect. 2.1.1.) or model data (see Sect. 5.1).

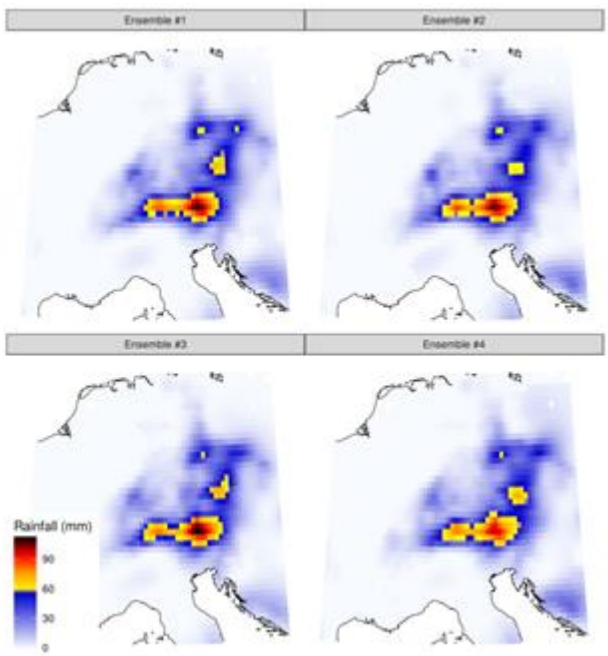


Figure 4-11: The first four ensemble members from the E-OBS ensemble precipitation dataset showing the heavy rainfall event on 1st June 2013.

The method used to generate the ensemble for E-OBS is based on the work of Hutchinson & Gessler (1994), and relates to the model uncertainty. It is closely related to the residuals of the statistical model used to construct the dataset, i.e. larger model residuals relate to a larger spread across the ensemble. In the case of precipitation the uncertainty scales with the interpolated values, i.e. higher precipitation values are associated with greater uncertainty, and is also related to station density. With temperature the uncertainty does not scale with the magnitude of the interpolated values, but it is ultimately determined by the density of stations. The standard error calculated from across the ensemble members is broadly consistent with the standard error value calculated in the original version of E-OBS. Hence the members represent samples from within that standard error range.

4.4.7 Further information

A general source of information for ECA&D is the website (www.ecad.eu) and staff members can be contacted at eca@knmi.nl for questions relating to the data

4.4.8 References

Cornes, R. C.,van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D. (2018) An Ensemble Version of the E-OBS Temperature and Precipitation Datasets. J. Geophys. Res. (Atmospheres) 123:9391–9409, doi:10.1029/2017JD028200

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D. and New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. J. Geophys. Res. (Atmospheres) 113, D20119, doi:10.1029/2008JD010201

Hutchinson, M. F., and P. E. Gessler. "Splines—more than just a smooth interpolator." *Geoderma* 62.1-3 (1994): 45-67.

Morice, C.P., J.J. Kennedy, N.A. Rayner & P.D. Jones, 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. J. Geoph. Res., VOL. 117, D08101, doi:10.1029/2011JD017187

Van der Linden, P. and Mitchell, J. F. B. (2009) ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK. 120pp.

4.5 Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016

4.5.1 Summary and general introduction

EUSTACE has developed an understanding of relationships between satellite surface temperature observations and surface air temperature over different surfaces of Earth, in order to use satellite retrievals to derive air temperature fields and thus increase the density of surface air temperature information globally (see Høyer et al. (2018) for details). This product contains those estimated air temperature fields.

Empirical relationships have been determined using measurements made *in situ*, satellite observations and auxiliary information. The relationships have been determined separately for each surface type to account for the different physical conditions present. The land and ice regions use the satellite skin observations directly in combination with auxiliary information, whereas the ocean approach applies a spatially and seasonally varying offset to the observed sea surface temperatures. The final results, however, can be used to estimate air temperature on the same global grid for all surfaces with the same definition of a day.

The data are provided in a set of NetCDF files and the format is described in Section 4.5.4. In short, for each surface type there is one file per day for each surface type containing a field, or fields, of air temperature estimates with corresponding estimates of the total uncertainty (see Figure 4-12). Data are formatted in a consistent way. A more-complete breakdown of uncertainty information is provided in an ancillary file. Temperature and uncertainty information is provided split according to correlation structure as used throughout the EUSTACE project and a consistent nomenclature has been used for all surfaces.

Key things to bear in mind for air temperature estimates over each surface are:

- To estimate marine air temperatures (MAT), the offset was calculated from sea surface temperature (SST) and MAT measurements made by ships between 1963 and 2000.
- For the land, the main challenges for estimating land surface air temperature (LSAT) from satellite observed land surface (skin) temperatures (LST), are the heterogeneity of the surfaces and residual cloud effects.
- The high latitude ice-covered regions, are challenging due to residual cloud contamination in the satellite observations and data sparsity.
- The surface air temperature estimates from satellite validate well against independent reference data over ocean: there is a small positive median discrepancy (Robust Standard Deviation, RSD) against a withheld subset of data input to HadNMAT2 of +0.25 K (1.19 K) with the highest discrepancies occurring in the mid to high northern latitudes. Note there were few matchups south of 20° N and none south of 50° S (see Veal, 2019 for details).
- Although the global median discrepancies (RSDs) of the estimated air temperatures over land against withheld stations from the EUSTACE Global Station Data set are small: -0.23 K (2.95 K) for

Tmin and +0.21 K (3.37) for Tmax, there is considerable regional and seasonal variation in median discrepancy (see Veal, 2019 for details).

- Over ice, the air temperature estimates have regional median discrepancies of +0.02 to +0.98 K and RSDs of 2.73 to 3.64 K depending on region and surface type (land or sea-ice, see Veal, 2019 for details).
- The uncertainty estimates vary in quality depending on domain. Over ocean, the uncertainty estimates are accurate, over land for both Tmin and Tmax the uncertainties are accurate for the smaller uncertainties but underestimated on data with higher uncertainties, and over ice the uncertainties are underestimated.

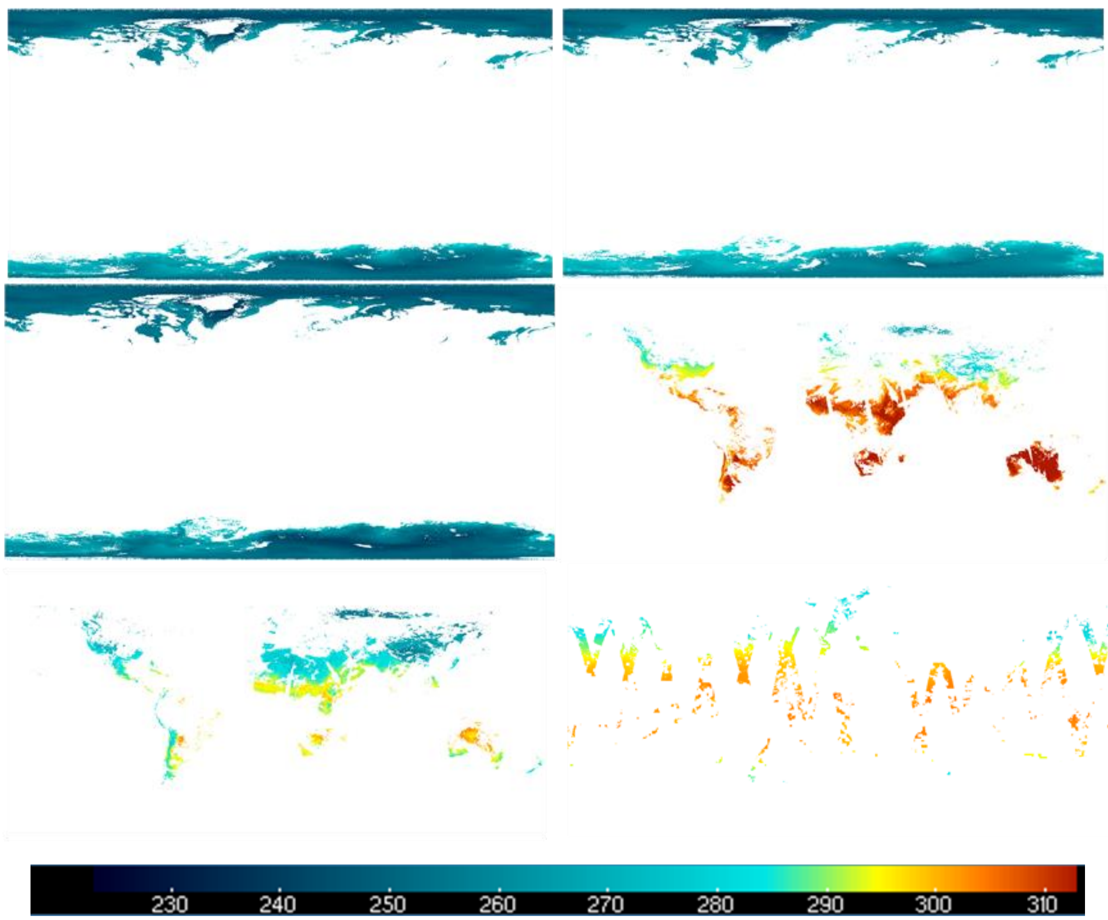


Figure 4-12: Example estimated air temperature fields (K) for 4th February 2003. (from top left): Tavg Ice surface air temperature, Tmax ice surface air temperature, Tmin ice surface air temperature, Tmax land surface air temperature, Tmin land surface air temperature and Tavg marine surface air temperature. All images use the same temperature scale

The following sections summarise the process of air temperature estimation for land, ice and ocean. Høyer et al. (2018) details the development of the regression relationships, here only their implementation is discussed.

4.5.2 Estimating air temperature

4.5.2.1 Over land

Table 4-23 provides an overview of the data sets used in the prediction of LSAT using the empirical statistical model.

Table 4-23: Data sources used. The uncertainty information provided for each data set is also shown; 'full breakdown' indicates that the information is partitioned into uncertainties with different correlation properties

Variable	Source	Native spatial resolution	Native temporal resolution	Uncertainty information
LSTday and LSTngt	EUSTACE/GlobTemperature global satellite land surface temperature, v2.1 from MODIS Aqua (Ghent 2019)	~ 1 km	Daily	Full breakdown
FVC Fraction of vegetation	Copernicus/Geoland-2 global land service FCOVER (*Baret et al.*, 2013; *Camacho et al.* 2013)	1/112°	10-day	Total only
Snow fraction	NASA MODIS Aqua (Hall et al., 2006)	0.05°	Daily	Quality and confidence (0-100) indices only

The method for estimating LSAT from satellite data is based on the approach detailed in Good (2015) and references therein, which was developed for data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) over Europe. The model is essentially a multiple linear regression used to predict daily *Tmin* and *Tmax* from daily satellite LST and other predictors, such as vegetation.

The generalised equations for predicting *Tmin* and *Tmax* are:

$$Tmax = \alpha_0 + \alpha_1.LST_{day} + \alpha_2.LST_{ngt} + \alpha_3.FVC + \alpha_4.SZA_{noon} + \alpha_5.Snow + \varepsilon_{Tmax.} \qquad \text{(Eq. 4.2.1)}$$
$$Tmin = \beta_0 + \beta_1.LST_{day} + \beta_2.LST_{ngt} + \beta_3.FVC + \beta_4.SZA_{noon} + \beta_5.Snow + \varepsilon_{Tmin} \qquad \text{(Eq. 4.2.2)}$$

Where α and β are the multiple linear regression coefficients. LST_{day} and LST_{ngt} are the retrieved daytime and night-time satellite LSTs (in °C), respectively. *FVC* is the fractional vegetation cover (0-1), *Snow* is the

% snow cover and SZA_{noon} is the solar zenith angle at local solar noon (in degrees). ε is the error on the model. All quantities correspond to the mean value for each grid cell at the specified spatial resolution. The regression coefficients are given in Table 4-24, together with the valid range of data for each predictor. Model error terms are discussed further in Section 4.5.5.1.

A global approach is adopted, whereby one set of coefficients is used for every spatial point at all times. Two model variants are used, which require different sets of predictors. The primary model for both temperature predictions, $Tmin_1$ and $Tmax_1$, requires both LST_{day} and LST_{ngt} in each case. The secondary model - $Tmin_2$ and $Tmax_2$ – requires only one LST to be present, thereby enabling greater spatial coverage of the LSAT predictions because LST_{day} and/or LST_{ngt} may be unavailable on a particular day, e.g. due to cloud or failed observations. The primary model is used wherever possible as these are the most accurate. The secondary model is used for other locations; it should be noted that these are mutually exclusive, so are not applied in any preferential order. The global fields of *Tmin* and *Tmax* are therefore a 'best guess' field and may include data from both models.

Table 4-24: Model regression coefficients and the valid data range for each predictor, determined physically (FVC, Snow, SZA) or from the input data (LST). 'StDev' gives the standard deviation of the model residuals (i.e. predicted satellite-derived LSAT minus observed station LSAT).

Model	Offset (°C) α_0 / β_0	LSTday α_1 / β_1	LSTngt α_2 / β_2	FVC α_3 / β_3	SZAnoon α_5 / β_5	Snow α_6 / β_6	StDev (°C)
Tmin1	-1.513	0.032	0.835	0.765	0.000	0.000	2.84
Tmax1	7.092	0.388	0.432	1.516	0.000	-0.011	3.02
Tmin2	0.184	0.000	0.850	0.595	-0.021	0.000	2.84
Tmax2	5.042	0.594	0.000	2.956	0.000	-0.022	3.65
$Range_{min}$	-	-80.0 °C	-80.0°C	0.0	0.0°	0	-
$Range_{max}$	-	65.0 °C	40.0°C	1.0	90.0°	100	-

While the model coefficients are derived using data at 0.05°, the model is implemented at 0.25° to match the EUSTACE target grid. The implementation of the model is straightforward. For each 0.25° cell, the predictor variables are aggregated and uncertainties are propagated (where possible). The coefficients in Table 4-24 are then applied to each predictor variable to give an estimate of Tmin and Tmax for that cell, according to Equations 4.2.1 and 4.2.2. The primary model – Model 1 – is implemented wherever possible, followed by Model 2. LSAT is not estimated for grid cells where the percentage of cloud-free 1 km pixel LST observations within that grid cell is less than 20%, or where the sampling uncertainty is greater than 3 °C. This is to reduce the amount of residual cloud contamination in the satellite LSAT estimates, as discussed above. The choice of these thresholds is a compromise between successfully rejecting poor-quality data while still preserving good data coverage. LSAT is also not estimated for cells classified as land or sea ice; near-surface air temperatures for these cells are provided in the ice air temperature products described in Section 4.5.2.2.

4.5.2.2 Over ice

The satellite data used are from the EUSTACE/AASTI Global satellite ice surface temperature, v1.0 (itself a development of the Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors (AASTI; Dybkjær et al., 2014) data set), covering high latitude seas, sea ice, and ice cap surface temperatures based on satellite infrared measurements. The data set is based on one of the longest existing satellite records from the Advanced Very High Resolution Radiometer (AVHRR) instruments aboard a long series of NOAA satellites. The first version of the AASTI product, used here, is available from 2000 to 2009 in approximately 5 km spatial resolution. The different AVHRR satellites have been orbiting 14 times per day, providing approximately bi-hourly coverage of the Polar Regions. An increasing number of satellites has been in orbit from 2000 to 2009 (Figure 4-13).

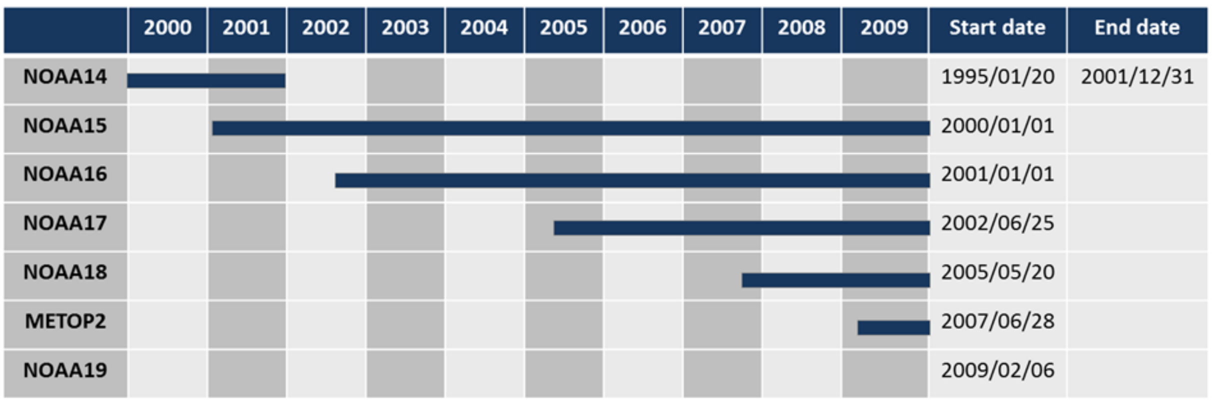


Figure 4-13: NOAA and Metop satellites carrying the AVHRR sensor, used for AASTI version 1.

The satellite swath observations were aggregated into daily averages of ice surface temperatures on a fixed 0.25 by 0.25 degree grid, which is common for all surfaces. EUSTACE is developing products for days defined by local solar time. The data in the aggregated files thus contains observations from 00 to 24 local solar time. The main aggregated surface variables are the daily average, minimum and maximum surface temperatures, but 3-hourly averages of surface temperatures and the number of observations were also calculated for the eight intervals during each day. These 3-hourly averages were used for estimating the satellite sampling throughout the day and to gain confidence in the daily cycle estimates. In the aggregation, all satellite observations with a quality flag of 4 or 5 were used. The aggregated surface skin data set is available to users upon request.

In order to best resolve the diurnal cycle with satellite information without losing too many observations we require data during both night (6pm-6am) and day (6am-6pm) in order to calculate the daily mean temperature (Tmean). Here, Tmean is calculated by averaging all available observations during the day. To calculate the daily minimum (Tmin) we require observations in at least one of the night bins and for daily maximum (Tmax) we require observations in at least one of the day bins.

The regression model is based on multiple linear regression analysis using least squares. As explaining factors, latitude, downward shortwave radiation not considering clouds (theoretical), seasonal cycle, and wind (ERA-Interim reanalysis) were tested. These factors were selected for testing based on current knowledge from the literature and an analysis of parameters that influence the relationship between IST and IAT at PROMICE stations and field experiments in Qaanaaq, limited by the available data.

The best correlation against training data was found using a model where IAT is predicted from daily satellite IST combined with a seasonal variation assumed to have the shape of a cosine function $A \cdot cos\left(\frac{time \cdot 2\pi}{1\ year} - \varphi\right)$ where A is the amplitude and ϕ is the phase. This can be rewritten in the form:

$$T_{mean} = \alpha_0 + \alpha_1 \cdot IST_{avg} + \alpha_2 \cdot cos\left(\frac{time \cdot 2\pi}{1\ year}\right) + \alpha_3 \cdot sin\left(\frac{time \cdot 2\pi}{1\ year}\right) + \varepsilon \qquad \textit{(Eq. 4.2.3)}$$

suitable for linear regression. The training data have been used to calculate the regression coefficients for each regression model pertaining to land and sea ice in both hemispheres (Table 4-25), giving a model fit with correlation of 94-96% and standard deviations of 3.2-3.3°C against training data for all four ice surfaces (Table 4-26). Similar regression coefficients have been derived for Tmin ($\beta_0 to \beta_3$) and Tmax ($\gamma_0 to \gamma_3$).

Table 4-25: Model regression coefficients

		Offset (°C) α_0	IST factor α_1	Cos amplitude (°C) α_2	Sin amplitude (°C) α_3
Land Ice	Northern Hemisphere	4.20	1.06	2.14	-0.74
	Southern Hemisphere	5.70	1.04	-0.42	-0.22
Sea Ice	Northern Hemisphere	1.46	0.89	-1.34	-1.24
	Southern Hemisphere	1.41	0.87	0.96	0.76

Table 4-26. Statistics on the relationship between observed and modelled IAT for the training data

	Number of observations	Correlation (%)	Standard deviation (°C)	Minimum difference (°C)	Maximum difference (°C)
Land ice N Hemisphere	13792	96.3	3.28	-12.9	12.2
Land ice S Hemisphere	15122	96.5	3.26	-10.9	9.9
Sea ice N Hemisphere	15010	96.2	3.25	-11.4	11.9
Sea ice S Hemisphere	430	94.4	3.16	-11.5	8.6

The derived coefficients have been used to estimate IAT over land and sea ice from satellite IST during the period 2000-2009. Four datasets are developed covering the Northern Hemisphere sea ice, Northern Hemisphere land ice, Southern Hemisphere sea ice, and Southern Hemisphere land ice. Each dataset provides daily estimates of mean air temperature (Tmean), daily maximum (Tmax) and daily minimum (Tmin). The data sets contain global IAT estimates on a 0.25 degree regular latitude-longitude grid. Each temperature estimate is associated with three categories of uncertainty: random uncertainties on the 0.25 degree daily scale; synoptic scale correlated uncertainty; and globally correlated uncertainty, including and excluding uncertainties related to the masking of clouds. The three types of uncertainties are also gathered in two total uncertainty estimates: one which includes the cloud mask uncertainty, and one which does not.

4.5.2.3 Over ocean

Modelled relationships derived from *in situ* data were used to estimate air temperatures based on satellite-retrieved SSTs. The satellite-retrieved SSTs were from the EUSTACE/CCI Global satellite sea surface temperature, v1.2 (Merchant et al. 2014) on daily 0.25° latitude/longitude resolution. This data set is based on the Along Track Scanning Radiometer (ATSR) series of instruments and is provided with uncertainty information, which is decomposed into three components associated with errors arising from: uncorrelated effects, locally correlated effects and systematic effects. The locally correlated error component is assumed to have a temporal scale of around 1 day and a length scale of 100km. The systematic error is assumed to be correlated perfectly in space and time, but is different for each of the satellite sensors of which two were used here: ATSR2 and AATSR. ATSR1 was not used owing to remaining difficulties with retrievals from the ATSR1 instrument.

Existing data sets of air-sea temperature difference (Parker et al. 1995) show that the air temperature is, on average, 1 to 2°C cooler than the sea-surface but with significant geographical and seasonal variations. In some areas, such as at the continental margins, there are significant deviations from the climatological average. Therefore, the modelled relationships between sea-surface temperature and air temperature developed here are based on an assessment of the climatological average difference (used to infer MAT from SST) and on the climatological variance of the difference (used to estimate the uncertainty in the MAT).

Given an estimate of SST the MAT is modelled as:

$$MAT(s,t) = SST(s,t) + \delta(s,t) + error \qquad \text{(Eq. 4.2.4)}$$

where δ is the climatological average offset modelled as a Fourier series,

$$\delta(s,t) = a0(s) + a1(s)Sin(2\pi d/365) + a2(s)Cos(2\pi d/365) + a3(s)Sin(4\pi d/365) + a4(s)Cos(4\pi d/365)$$

(Eq 4.2.5)

where d is the day of the year (from 0 to 365), the ai are the Fourier coefficients estimated without interpolation.

SST retrievals from the ATSR-2 and AATSR instruments were used to estimate MAT using the above method. The adjustments were bilinearly interpolated to the target 0.25° grid. We estimate Tmean and ten separate uncertainty fields:

1. Uncertainty from random uncorrelated errors from the satellite retrievals.
2. Uncertainty from locally-systematic errors in the satellite SST retrievals which are correlated on approximately synoptic scales i.e. 100km and 1 day.
3. Uncertainty from global systematic errors in the satellite SST retrievals, which are completely correlated in space and time.
4. Uncertainty from locally-systematic errors arising from estimating MAT from SST, which are correlated with a length scale of approximately 1000km.
5. Uncertainty from global systematic errors in the in situ data used to construct the climatology of AST difference. These are correlated in time and strongly correlated in space.
6. -10. Uncertainty in the offset and Fourier components used to smooth the climatology. These are correlated in time (as modulated by the Fourier components), but uncorrelated in space.

Coverage is typically lower than that of the satellite data because relationships between SST and MAT are not specified everywhere. Uncertainties are dominated by the locally correlated error component (number 4 in the list).

4.5.3 How to access the data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. It is made freely available for all purposes under the Open Government Licence.

Table 4-27: How to access the Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016

Data Product	CEDA catalogue record	Download links
EUSTACE: 4.5: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016	https://catalogue.ceda.ac.uk/uuid/f883e197594f4fbaae6edebafb3fddb3	http://data.ceda.ac.uk/neodc/eustace/data/satellitederived/mohc/eustace/v1.0/day/0/0/R001336/20190111/

4.5.4 Detailed format specifications and metadata

The dataset comprises two files per surface type (land, ice and ocean) per day. One of these is the main product file and contains fields of air temperature and total uncertainty. The other is an ancillary file that contains additional information that is specific to each surface, such as a detailed breakdown of the uncertainty components. The format is given in the NetCDF file and reproduced below.

EUSTACE products are intended to be representative of a day defined as running from midnight to midnight local solar time. This is represented in the NetCDF files by a combination of a "time" variable, defined as the UTC time at zero longitude and a "timeoffset" variable, which gives the local time offset from UTC for each longitude. UTC values specified by the "time" variable are therefore only correct at zero longitude and this is indicated in the long name. Combining the "time" variable and the "timeoffset" field, gives the local time for each longitude in UTC.

Main product file

Table 4-28: Dimensions in the main product file for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016

time	UNLIMITED; // (1 currently)
bounds	2
latitude	720
longitude	1440

Table 4-29: Variables in the main product file for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016. (note, not all surfaces have all temperature types: max,min,mean)

Variable (dimensions)	Standard name	Description
time(time)	time	Time at zero longitude in days since 1850-01-01T00:00:00Z
timebounds(time, bounds)		
latitude(latitude)	latitude	Latitude in degrees north
longitude(longitude)	longitude	Longitude in degrees east
timeoffset(longitude)	none	Local time offset from UTC (days)
tas(time, latitude, longitude)	air_temperature	Mean daily surface air temperature in K
tasmin(time, latitude, longitude)	air_temperature	Minimum daily surface air temperature in K
tasmax(time, latitude, longitude)	air_temperature	Maximum daily surface air temperature in K
tasuncertainty(time, latitude, longitude)	none	Total uncertainty in mean daily surface air temperature in K
tasminuncertainty(time, latitude, longitude)	none	Total uncertainty in minimum daily surface air temperature in K
tasmaxuncertainty(time, latitude, longitude)	none	Total uncertainty in maximum daily surface air temperature in K

Ancillary file – Land

Table 4-30: Dimensions in the ancillary files for land for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, 1995-2016

time	UNLIMITED; // (1 currently)
bounds	2
latitude	720
longitude	1440

Table 4-31: Variables in the ancillary files for land for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, 1995-2016

Variable (dimensions)	Standard name	Description	length scale	time scale
time(time)	time	Time at zero longitude in days since 1850-01-01T00:00:00Z		
timebounds(time, bounds)				
latitude(latitude)	latitude	Latitude in degrees north		
longitude(longitude)	longitude	Longitude in degrees east		
timeoffset(longitude)	none	Local time offset from UTC (days)		
tasmin_unc_rand(time, latitude, longitude)	none	Random uncertainty on minimum daily surface air temperature in K		
tasmin_unc_corr_atm(time, latitude, longitude)	none	Locally correlated atmospheric uncertainty on minimum daily surface air temperature in K	unknown	unknown
tasmin_unc_corr_sfc(time, latitude, longitude)	none	Locally correlated surface uncertainty on minimum daily surface air temperature in K	unknown	30 days
tasmin_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on minimum daily surface air temperature in K		
tasmax_unc_rand(time, latitude, longitude)	none	Random uncertainty on maximum daily surface air temperature in K		
tasmax_unc_corr_atm(time, latitude, longitude)	none	Locally correlated atmospheric uncertainty on	unknown	unknown

		maximum daily surface air temperature in K		
tasmax_unc_corr_sfc(time, latitude, longitude)	none	Locally correlated surface uncertainty on maximum daily surface air temperature in K	unknown	30 days
tasmax_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on maximum daily surface air temperature in K		
tasmin_model_number(time, latitude, longitude)	none	Model number used for estimating Tmin from satellite data Model 1: include both LST-day and LST-night to predict tasmin Model 2: include LST-night to predict tasmin Model 3: include LST-day to predict tasmin		
tasmax_model_number(time, latitude, longitude)	none	Model number used for estimating Tmax from satellite data Model 1: include both LST-day and LST-night to predict tasmax Model 2: include LST-day to predict tasmax Model 3: include LST-night to predict tasmax		

Ancillary file – Ice

Table 4-32: Dimensions in the ancillary files for ice for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for ice, 1995-2016

time	UNLIMITED; // (1 currently)
bounds	2
latitude	720
longitude	1440

Table 4-33: Variables in the ancillary files for ice for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for ice, 1995-2016

Variable (dimensions)	Standard name	Description	length scale	time scale
time(time)	time	Time at zero longitude in days since 1850-01-01T00:00:00Z		
timebounds(time, bounds)				
latitude(latitude)	latitude	Latitude in degrees north		
longitude(longitude)	longitude	Longitude in degrees east		
timeoffset(longitude)	none	Local time offset from UTC (days)		
tas_unc_no_cloud(time, latitude, longitude)	none	Total uncertainty excluding cloud on average daily surface air temperature in K		
tas_unc_rand(time, latitude, longitude)	none	Random uncertainty on average daily surface air temperature in K		
tas_unc_corr_local(time, latitude, longitude)	none	Locally correlated uncertainty on average daily surface air temperature in K	500 km	5 days
tas_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on average daily surface air temperature in K		
tas_unc_cloud(time, latitude, longitude)	none	Cloud component of uncertainty on average daily surface air temperature		
tasmin_unc_no_cloud(time, latitude, longitude)	none	Total uncertainty excluding cloud on minimum daily surface air temperature in K		

tasmin_unc_rand(time, latitude, longitude)	none	Random uncertainty on minimum daily surface air temperature in K		
tasmin_unc_corr_local(time, latitude, longitude)	none	Locally correlated uncertainty on minimum daily surface air temperature in K	500 km	5 days
tasmin_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on minimum daily surface air temperature in K		
tasmin_unc_cloud(time, latitude, longitude)	none	Cloud component of uncertainty on minimum daily surface air temperature		
tasmax_unc_no_cloud(time, latitude, longitude)	none	Total uncertainty excluding cloud on maximum daily surface air temperature in K		
tasmax_unc_rand(time, latitude, longitude)	none	Random uncertainty on maximum daily surface air temperature in K		
tasmax_unc_corr_local(time, latitude, longitude)	none	Locally correlated uncertainty on maximum daily surface air temperature in K	500 km	5 days
tasmax_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on maximum daily surface air temperature in K		
tasmax_unc_cloud(time, latitude, longitude)	none	Cloud component of uncertainty on maximum daily surface air temperature		

Ancillary file – Ocean

Table 4-34: Dimensions in the ancillary files for ocean for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for ocean, 1995-2016

time	UNLIMITED; // (1 currently)
bounds	2
latitude	720
longitude	1440

Table 4-35: Variables in the ancillary files for ocean for the globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for ocean, 1995-2016

Variable (dimensions)	Standard name	Description	length scale	time scale
time(time)	time	Time at zero longitude in days since 1850-01-01T00:00:00Z		
timebounds(time, bounds)				
latitude(latitude)	latitude	Latitude in degrees north		
longitude(longitude)	longitude	Longitude in degrees east		
timeoffset(longitude)	none	Local time offset from UTC (days)		
tas_unc_rand(time, latitude, longitude)	none	Random uncertainty on average daily surface air temperature in K		
tas_unc_corr_sat(time, latitude, longitude)	none	Locally correlated uncertainty (from satellite retrieval) on average daily surface air temperature in K	100 km	1 day
tas_unc_sys(time, latitude, longitude)	none	Systematic uncertainty on average daily surface air temperature in K		
tas_unc_corr_mod(time, latitude, longitude)	none	Locally correlated uncertainty (from surface-air model) on average daily surface air temperature in K	1000 km	3 days
tas_unc_sys_mod(time, latitude, longitude)	none	Systematic uncertainty (from surface-air model) on average daily surface air temperature		
tas_unc_parameter_0(time, latitude, longitude)	none	Systematic uncertainty mean offset on average daily surface air temperature		

tas_unc_parameter_1(time, latitude, longitude)	none	Systematic uncertainty first fourier component on average daily surface air temperature
tas_unc_parameter_2(time, latitude, longitude)	none	Systematic uncertainty second fourier component on average daily surface air temperature
tas_unc_parameter_3(time, latitude, longitude)	none	Systematic uncertainty third fourier component on average daily surface air temperature
tas_unc_parameter_4(time, latitude, longitude)	none	Systematic uncertainty fourth fourier component on average daily surface air temperature

4.5.5 Uncertainties

4.5.5.1 Estimating uncertainties

Land

Uncertainties are estimated for each predictor variable based on the uncertainty information available for each native data set used. For the LST data – which is the only predictor to have complete characterisation of uncertainties - each uncertainty component (random, locally correlated, and systematic) is propagated through the gridding process according to EUSTACE Deliverable 1.2 (Merchant et al., 2015). For the *FVC* data set, the mean total uncertainty within a particular grid cell is used as an estimate of the locally correlated uncertainty for that grid cell, while the difference between the maximum and the mean uncertainties within each cell is used as an estimate of the random uncertainty. The quoted precision of the FVC estimates, 0.004, is assumed to be the systematic uncertainty on a pixel-by-pixel basis. At present, no uncertainty information has been estimated for the *snow* estimates. As the dominant contribution to the satellite-derived LSAT estimates is from the LST data, omitting the uncertainty contribution from the snow data should have only a very small impact.

Uncertainty estimates for the satellite-derived LSAT data are estimated so that for a particular grid cell at a particular point in time, the random uncertainty component is given by:

$$Tmax_{random} = (\alpha_1^2.LST_{day\_random}^2 + \alpha_2^2.LST_{ngt\_random}^2 + \alpha_3^2.FVC_{random}^2)^{½} \quad \text{(Eq. 4.2.6)}$$
$$Tmin_{random} = (\beta_1^2.LST_{day\_random}^2 + \beta_2^2.LST_{ngt\_random}^2 + \beta_3^2.FVC_{random}^2)^{½} \quad \text{(Eq. 4.2.7)}$$

Where LST_{day_random}, LST_{day_random} and FVC_{random} are the random uncertainty components from the input data sets for the same grid cell at the same point in time for LST_{day}, LST_{night} and *FVC*, respectively. The coefficients α and β are the model coefficients from Table 4-24 so for models that do not use a particular predictor, that term in equations 4.2.6 and 4.2.7 will be zero. Similarly, the locally correlated uncertainty components are given by:

$Tmax_{atm}= (\alpha_1^2.LST_{day\_local\_atm}^2 + \alpha_2^2.LST_{ngt\_loca\_atml}^2+ \sigma_{Tmax}^2)^{½}$ (Eq. 4.2.8)
$Tmin_{atm}= (\beta_1^2.LST_{day\_local\_atm}^2 + \beta_2^2.LST_{ngt\_local\_atm}^2+ \sigma_{Tmin}^2)^{½}$ (Eq. 4.2.9)
$Tmax_{surf}= (\alpha_1^2.LST_{day\_local\_surf}^2 + \alpha_2^2.LST_{ngt\_local\_surf}^2+ \alpha_3^2.FVC_{local}^2)^{½}$ (Eq. 4.2.10)
$Tmin_{local}= (\beta_1^2.LST_{day\_local\_surf}^2 + \beta_2^2.LST_{ngt\_local\_surf}^2+ \beta_3^2.FVC_{local}^2)^{½}$ (Eq. 4.2.11)

Where $LST_{day\_local\_atm}$, $LST_{ngt\_local\_atm}$, $LST_{day\_local\_surf}$, $LST_{ngt\_local\_surf}$ and FVC_{local}, are the locally correlated uncertainty components from the input data sets for the same grid cell at the same point in time for LST_{day}, LST_{night} and FVC, respectively. (Note: The EUSTACE MODIS LST products provide separate uncertainty estimates due to atmospheric (atm) and surface (surf) effects on the LST retrievals, and it is these separate components that are used here.)

The terms σ_{Tmin} *and* σ_{Tmax} represent the standard deviation of the residuals, i.e. the satellite-predicted Tmin/Tmax minus the station Tmin/Tmax, calculated during the model training process ('StDev' in Table 4-24). The grid-cell systematic uncertainties are assumed to be 0.1 °C, which is consistent with the systematic uncertainty estimated for the EUSTACE LST data.

Ice

First, we propagate the different uncertainty components for the IST retrievals onto the 0.25° latitude by 0.25° longitude grid used here.

The random uncertainty component of the gridded satellite IST data is given by:

$$RU_{sat} = \sqrt{RU_{instrument}^2 + RU_{geolocation}^2} \qquad \textit{(Eq. 3.4.2)}$$

where $RU_{instrument}$ and $RU_{geolocation}$ are the uncertainty components due to instrument noise and geolocation errors in the marginal ice zone, respectively, from the same grid cell at the same point in time.

The locally correlated uncertainties are given by:

$$SSU_{sat} = \sqrt{SSU_{emissivity}^2 + SSU_{atmosphere}^2} \qquad \textit{(Eq. 3.4.3)}$$

where $SSU_{emissivity}$ and $SSU_{atmosphere}$ are the synoptic scale uncertainty components of the IST data due emissivity errors and atmospheric corrections.

The grid-cell systematic uncertainties (LSU_{sat}) are based on expert judgements and provided as a fixed value of 0.2°C.

The cloud mask uncertainty is calculated from the cloud mask quality level of the grid cell, using

$$CU_{sat,avg} = \alpha_1 \cdot (\delta_{avg} + 0.5 \cdot (5 - QL)) \qquad \textit{(Eq. 3.4.4)}$$

where α_1 is the regression coefficient (see Table 4-25) and δ_{avg} is an overall uncertainty level for CU_{avg}, set to 0.8°C. The uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1 and δ_{avg} with δ_{min}=2.8°C and δ_{max}=0.8°C. The higher δ-value for Tmin occurs because undetected clouds will appear as artificial cold temperatures, especially affecting the Tmin field because they are harder to distinguish from the coldest surface temperatures.

Then we estimate uncertainties in our air temperature estimate. The random uncertainty component for Tmean is given by:

$$RU_{avg} = \sqrt{(\alpha_1 \cdot RU_{sat})^2 + RU_{sampling}^2} \qquad \textit{(Eq. 3.4.5)}$$

where $RU_{sampling}$ is the uncertainty relating to sampling errors in space and time (see section 3.4.5.3). The random uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

Similarly, the locally correlated (synoptic scale) uncertainty component is given by:

$$SSU_{avg} = \sqrt{(\alpha_1 \cdot SSU_{sat})^2 + SSU_{relation}{}^2} \qquad \textit{(Eq. 3.4.6)}$$

The term $SSU_{relation}$ represents the standard deviation of the residuals calculated at in situ stations where both skin and air temperatures are available, i.e. the skin temperature-predicted Tavg minus the observed Tmean, calculated during the model training process of these in situ only data. Again, the uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

Table 4-36. Uncertainty estimates for the relationship error

	T_{mean} (°C)	T_{min} (°C)	T_{max} (°C)
Land ice	1.5	1.8	2.0
Sea ice	1.7	1.8	2.1

Three forms of systematic uncertainties are provided: the covariance matrix of the model coefficients, an estimate of the systematic uncertainty in IST for each grid cell and the cloud mask uncertainty component. The grid-cell systematic uncertainties are $LSU_{avg} = \alpha_1 \cdot LSU_{sat}$ and cloud mask uncertainty $CU_{avg} = \alpha_1 \cdot CU_{sat}$, and as above, the uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

The sampling uncertainty is determined indirectly by closing the average RMS error budget when validating against the independent in situ air temperature measurements, taking the in situ air temperature uncertainty into account. This approach was used due to limited time and limited validation data, but has the disadvantage that other unaccounted uncertainty sources are artificially grouped with the sampling uncertainty. The sampling uncertainty is determined individually for the average, minimum and maximum temperatures, and for each surface type, see Table 4-33.

Table 4-33. Sampling uncertainty (°C) for Tmean, Tmin and Tmax.

	T_{mean} (°C)	T_{min} (°C)	T_{max} (°C)
Land ice Northern Hemisphere	1.6	2.2	2.1
Sea ice Northern Hemisphere	0.0[7]	2.9	2.0
Land ice Southern Hemisphere	1.6	1.5	3.8
Sea ice Southern Hemisphere	1.7	5.7	2.9

Ocean

The error term in Eq. 4.2.4 has components arising from:

1. SST measurement error which, in the case of satellite data, will itself have uncorrelated, locally correlated and systematic components.
2. The climatological variance of the air-sea temperature difference, which is assumed to be normally distributed with mean zero and variance *σvar2(s,d)*.

$$\sigma var2(s,d) = b0(s) + b1(s)Sin(2\pi d/365) + b2(s)Cos(2\pi d/365) + b3(s)Sin(4\pi d/365) + b4(s)Cos(4\pi d/365)$$

3. Systematic uncertainty in the bias adjustments applied to the in situ data, estimated to be around 0.1°C and correlated in space and time.

The uncertainty in the Fourier coefficients, *a0, a1, a2, a3* and *a4*, are described by the error covariance matrix *cov(ai, aj)* and the square roots of the diagonal terms are given in the NetCDF file as. The errors in the Fourier components are assumed to be uncorrelated in space.

4.5.5.2 Using uncertainty estimates

The main product files contain a "total uncertainty" in the estimated temperature, which combines all the different sources of quantified uncertainty.

However, care is needed when using the uncertainty information in calculations as the estimate of total uncertainty does not provide sufficient information on its own to correctly propagate the uncertainties. The total uncertainty combines uncertainties associated with several different error sources each of which has different correlation structures. The ancillary files for each surface contain those individual uncertainty components.

The individual uncertainty components in the ancillary files fall into three categories: random, locally correlated and systematic. Random uncertainties are associated with uncorrelated, independent errors. Locally-correlated uncertainties are associated with errors that are correlated on relatively short space and time scales. Systematic uncertainties are associated with systematic errors that are strongly correlated on large space and time scales.

[7] This zero-value is likely caused by a slight overestimation of some of the other uncertainty quantities.

Each component should be propagated separately through any calculation and, if a total uncertainty is required, combined at the end assuming they are independent from each other. In order to propagate the uncertainties, it is necessary to know the correlations structures and the propagation of uncertainty formula.

Propagation of uncertainty formula – for a function f computed from uncertain values x_i with uncertainties σ_{x_i}, the uncertainty in $f(x_1 \dots x_n)$ for uncorrelated errors is given by

$$\sigma_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2$$

For correlated errors, the more general formula is

$$\sigma_f^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \sigma_{x_i} \sigma_{x_j} \rho_{x_i x_j}$$

Where ρ is the error correlation.

Random uncertainties – these arise from errors that are uncorrelated and independent. These uncertainties can therefore be propagated using the familiar formula for propagation of uncorrelated errors.

Locally-correlated uncertainties – these arise from errors that are locally-correlated. The correlation *r* of the errors is given by an exponentially decaying function:

$$r = e^{-(\frac{\delta d}{\lambda_d} + \frac{\delta t}{\lambda_t})}$$

Where δd is the distance between the grid cell centres, δt is the time separation of the grid cells and λ_d is the characteristic length scale and λ_t the characteristic time scale (given in the attributes of the variable in the NetCDF files). For some of the land components the characteristic length and time scales are not known.

In order to propagate these uncertainties, the version of the propagation of errors formula which includes correlated effects should be used.

An alternative approach, which might be appropriate for regridding the data, is to assume that, within the characteristic space and time scales, the errors are fully correlated. This is an approximate, but simple method.

Systematic uncertainties – these arise from errors which are strongly correlated on large space and time scales. In propagating these uncertainties, assume that the errors are completely correlated.

Parameter uncertainties – for the oceans, there is another category of systematic uncertainties, which is associated with errors in the estimated seasonal cycle of air sea temperature differences. The seasonal cycle is represented by Fourier components (see equation 4.2.5). In this case, the uncertainty should be propagated as if it were uncorrelated in space, but systematic in time with one caveat. The Fourier components can take both negative and positive values, but the uncertainties in the files are all positive. To deal with this, the correlation should be set to ±1 in the propagation formula. Whether it is +1 or -1 is determined by the signs of the appropriate Fourier component for the two days of the year under consideration. If the signs are the same, the correlation is +1, if they are different it is -1.

4.5.6 Known limitations

When using the data, users should be aware of the following:

- Some of the satellite data used to estimate air temperatures suffer from "cloud contamination". As far as possible, cloud effects have been removed from the satellite data, but in some conditions clouds are very difficult to detect reliably. Grid cells affected by cloud contamination can have a very different temperature from their neighbours. However, cloud contamination can also be more subtle leading to skewed error distributions.
- Coverage of the air temperatures over ocean derived from satellites is limited to well-established shipping routes, where there are sufficient data to build the relationships between surface and air temperatures. Consequently, the coverage is very sparse in the tropics and southern hemisphere.
- The errors in the data do not conform precisely to the nice breakdown of uncertainty components. In practice there are uncertainties with large space and time scales and more complex structures than allowed for in the simple error model. Information about the likely scale and structure of these can be found in the validation (Veal, 2019).
- In the case of the Tmin estimates over ice, for uncertainty estimates above 3.0K, the median discrepancy between estimated Tmin and independent reference data is negative and increases in magnitude with increasing uncertainty estimate. EUSTACE uncertainty estimates are underestimated where they exceed 3.0K. Users may find it appropriate to discard data with an uncertainty greater than 3.0 K.
- Inputs to EUSTACE are not all defined from midnight to midnight local time. For example, some stations report max and min temperatures read at 0900 local time other report at other times of day. Consequently, the daily temperatures may be locally offset from the time defined in the files. This can affect, for example, the precise timing of the arrival of a cold front. An alternative approach using a longitude-dependent time variable was tested with users, but found to cause problems in some software packages.

4.5.7 Where to go for further information

Detailed methods of deriving relationships between skin and air temperature can be found in the following references:

Høyer, J., E. Good, P. Nielsen-Englyst, K. S. Madsen, R. I. Woolway, J. Kennedy (2018) Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes, EUSTACE Deliverable 1.5

Nielsen-Englyst, P., J. L. Høyer and K. S. Madsen (2019) Deriving Arctic 2 m air temperatures from satellite, in preparation

Good, E. J. and D. J. Ghent (2019) Estimating satellite land surface air temperatures from MODIS for the EUSTACE global analysis, in preparation

Kennedy, J. J. and E. C. Kent (2019) Estimating the climatological mean and spatial covariance of air-sea temperature differences, in preparation

4.5.8 References

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., & Smets, B. (2013). GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. Remote Sensing of Environment, 137, 299-309

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M. (2013). GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. Remote Sensing of Environment, 137, 310-329

Dybkjær, G., J. L. Høyer, R. Tonboe, S. M. Olsen (2014). Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST. NACLIM Deliverable D32.28.

Ghent D., K. Veal, T. Trent, E. Dodd, and J. Remedios, 2019: A new approach to defining uncertainties for MODIS land surface temperature, submitted to Remote Sensing of Environment
Hall et al. (2006)

Good, E. (2015), Daily minimum and maximum surface air temperatures from geostationary satellite data, JGR-Atmospheres, doi:10.1002/2014JD022438.

Hall, D. K., V. V. Salomonson, and G. A. Riggs (2006), MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG, Version 5. [MYD10.C1]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. http://dx.doi.org/10.5067/EI5HGLM2NNHN.

Høyer, J., E. Good, P. Nielsen-Englyst, K. S. Madsen, R. I. Woolway, J. Kennedy (2018) Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes, EUSTACE Deliverable 1.5, https://www.eustaceproject.org/eustace/static/media/uploads/d1.5_revised.pdf

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S. and Donlon, C. (2014), Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). Geoscience Data Journal. doi: 10.1002/gdj3.20

Merchant, C. J., Ghent, D., Kennedy, J., Good, E. and Høyer, J. (2015) Common approach to providing uncertainty estimates across all surfaces, EUSTACE Deliverable 1.2, https://www.eustaceproject.org/eustace/static/media/uploads/Deliverables/eustace_d1-2.pdf

Nielsen-Englyst, P., J. L. Høyer and K. S. Madsen (2019) Deriving Arctic 2 m air temperatures from satellite, in preparation

Parker, D.E., C.K. Folland and M. Jackson (1995) Marine Surface Temperature: Observed variations and data requirements, Climatic Change 31: 559-600, 1995

Veal, K. L. (2019) Validation report for the final in-filled EUSTACE surface air temperature product, EUSTACE Deliverable 3.5

4.6 Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0

4.6.1 Summary

This product is a global daily mean air temperature data set combining surface and satellite data. It is presented on an equirectangular (regular lat-lon) grid with a grid spacing of 0.25°. The data set is daily from 1850 to 2015. The data are presented in a set of NetCDF files, which use a format which is consistent with the satellite derived air temperatures from Section 4.5 and uses CF conventions for describing the variables.

Users of the product have access to mean temperature data and uncertainty estimates that are consistent across a broad range of space and time scales from daily 0.25° to multidecadal global averages. The coverage is significantly better than is available from station data alone, and covers land, ocean and ice areas.

Uncertainty in the product is represented by a "total" uncertainty which gives a point estimation of the uncertainty in the central estimate. In addition there is an ensemble of 10 samples drawn from the posterior distribution of the statistical model. The spread of the ensemble captures the full range of covariance in the statistical model after the observations have been assimilated.

A statistical method has been used to estimate air temperatures at all places and times. It takes into account uncertainty in the input data sets covering errors in the in situ measurements, land station homogenisation (from breaks identified in Section 4.2) and errors in the air temperatures estimated from satellite data (Section 4.5). Although the statistical model estimates temperatures at all locations, the product is not globally complete, as areas with too few data to provide a reliable air temperature estimate have been masked out.

4.6.1 How to access the data

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. It is made freely available for all purposes under the Open Government Licence.

Table 4-37: How to access the dataset: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015

Data Product	CEDA catalogue record	Download links
EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/468abcf18372425791a31d15a41348d9	http://data.ceda.ac.uk/neodc/eustace/data/combined/mohc/eustace/v1.0/day/0/0/R001400/20190326/global/

Detailed format specifications are given in section 4.6.5 below.

4.6.2 Input data sets and pre-processing

Three principal input data sources are used to create the infilled product. These are the EUSTACE Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016 (Section 4.5), EUSTACE Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015 (Section 4.2) and quality-controlled in situ marine air temperature data from ships, as used in the HadNMAT2 data set (Kent et al., 2013).

The total uncertainty provided with the satellite-derived air temperatures is used and treated as if the errors were uncorrelated. This is an approximation to the full error model.

In situ station data and locations in time of station breaks were taken from the EUSTACE Global land station daily air temperature measurements with non-climatic discontinuities identified (Section 4.2). Data to be used in the validation were excluded. An uncertainty of 0.3°C was assumed for each measurement, corresponding to rounding to the nearest whole number of degrees. Homogenisation errors, corresponding to changes in the mean temperature at a station between station breaks, were assumed to have a mean of zero and a prior uncertainty of 0.3°C; the statistical model re-estimated the size of each break.

In situ ocean data were taken from ICOADS 2.5 (Woodruff et al. 2011). Individual observations were bias adjusted using the method of (Kent et al. 2013) to a standard reference height of 2m. Only data from platforms identified as ships were used. An uncertainty of 1K was assumed for each measurement, with the errors assumed to be uncorrelated. At least 10% of the data for each month were excluded from the analysis for use in validation

During the Second World War, it is thought that the protocol for making marine air temperature measurements was changed (Kent et al. 2013). Measurements were made at different times of the day and night. In addition it is possible that measurements were not made on deck, but thermometers were taken indoors for reading. This leads to a warm bias in the marine data between approximately January 1942 and February 1946, which affects the analysis over the oceans.

4.6.3 Statistical model

A statistical model was used to produce a global temperature analysis from the input data. The statistical model comprises three components which correspond to temperature variations over three time scales: the climatology component; the large scale component and the local component. The climatology component includes very slow changes in temperature on time scales greater than about 10 years and the seasonal cycle. The large-scale component describes year-to-year variability and the local component is related to daily weather. Different types of errors in the input measurements are associated with individual component where they are most relevant. For example, station biases arising from non-climatic discontinuities are associated with the large-scale component because breaks in the station series are identified at an annual resolution. Each component is further split into elements which are summarised in Table 4-38, Table 4-39 and Table 4-40.

Each of the components is fit to the data in turn starting with the climatology, followed by the large-scale and then the local, moving from the broadest and slowest scales, to the shortest and fastest. The process is then repeated starting again with the climatology component, followed by the large-scale and local. The process was repeated five times with each iteration improving the overall fit of the model to the data.

The statistical model represents temperatures on the globe using a triangular mesh. The triangles can be further subdivided to provide a higher resolution analysis. For example, the local component uses a grid with approximately 0.5° resolution, whereas the large-scale component has fewer triangles and hence a lower resolution of around 5°. Table 4-38 gives the resolution of the different climatology components. Smoothly-varying patterns of temperature can be represented by adding together simple basis functions that sit at each node in the mesh. The more triangles (and nodes) there are, the smaller the details that can be resolved on the mesh. The relationships between the weights for the basis functions are determined by solving a "stochastic partial differential equation" or SPDE (see Lindgren et al., 2011).

The SPDE used in the local component to represent daily weather is a simple spatial SPDE with the assumption that the weather on one day is independent of the weather on the days before and after. In the large scale and climatology components, space-time SPDEs are used such that there are interactions between one time step and the next. As with the nodes on the triangular mesh, the separation of the time steps determines the level of temporal detail that the SPDE allows. Space-time SPDEs require a lot of computing resources to solve, so the resolution is necessarily more limited.

Table 4-38: Sub-components of the climatology component. Prior means are zero for all sub-components.

Sub-Component	Description	Model resolution	Prior parameters	Comments
Seasonal	Space-Time SPDE	1° spatial resolution triangulation, monthly time resolution	Standard deviation = 5 K; Spatial length scale = 25°; Time length scale = 2 months	Nodes placed at the start of each calendar month, with SPDE linking 12th month to the 1st.
Very long scale	Space-Time SPDE	5° spatial resolution triangulation, 4 yearly time resolution	Standard deviation = 0.5 K; Spatial length scale = 30°; Time length scale = 10 years	Models very slow and climatic changes
Latitude	1D SPDE	0.5 degree latitudinal resolution	Standard deviation = 12 K; length scale = 30°	
Altitude	Covariate	Covariate for 0.25° gridded digital elevation map	Standard deviation = 6.5 K per km	

Water fraction	Covariate	Covariate for 0.25° gridded areal fraction of water	Standard deviation = 10 K at 100% water	Water fraction derived from CCI land cover
Grand mean	Covariate	Covariate for analysis mean temperature for all space and time	Standard deviation = 546.3 K	Loose prior set at double the triple point of water.

Table 4-39: Sub-components of the large-scale component. Prior means are zero for all sub-components.

Sub-Component	Description	Model resolution	Prior parameters	Comments
Mid-scale	Space-Time SPDE	3 monthly x 5°	Standard deviation = 1 K; Spatial length scale = 15°; Time length scale = 4.5 months	
Station biases	Bias estimate for each station adjustment	One covariate per homogeneous segment	Standard deviation = 0.9 K	Based on breakpoint assessment in EUSTACE Global station data set (see Section 4.2). Most recent homogeneous period considered unbiased and not included.

Table 4-40: Sub-components of the local daily component, to be estimated independently for each day of the analysis. Prior means are zero for all sub-components.

Sub-Component	Description	Model resolution	Prior parameters	Comments
Local	Spatial SPDE	0.5° spatial resolution triangulation	Standard deviation = 2 K; Spatial length scale = 5°	Spatial field that models daily temperature variations
Satellite-derived marine air temperature bias	Global covariate	Single covariate for data source for whole global	Standard deviation = 2 K	Daily estimates of mean bias relative to other data sources

Satellite-derived ice air temperature bias	Hemispheric covariate	Two covariates for data source, one for each hemisphere	Standard deviation = 2 K	Daily estimates of mean bias relative to other data sources for each hemisphere
Satellite-derived land air temperature bias	Spatial SPDE	1° spatial triangulation for data source	Standard deviation = 1 K; Spatial length scale = 25°	Daily estimates of spatial bias field relative to other data sources

As well as providing an assessment of the temperatures across the globe, the statistical method also provides an estimate of the uncertainty in the field. Where there are many high-quality observations, the uncertainty will be small. Where there are few observations, the uncertainty will be larger. In some regions where there have historically been very few observations, such as the Southern Ocean, even the long-term average can be uncertain.

In the Global air temperature estimates, v1.0 product, the uncertainty is represented in two ways. First, every grid cell has an estimated uncertainty. This gives an indication of how reliable the grid cell value is relative to others. However, the grid cell uncertainty does not tell you anything about how errors in neighbouring grid cells might be related. The second way that uncertainty is represented is by drawing samples from the posterior distribution of the statistical model. The samples that make up the ensemble are consistent with the available measurements, but where there are no measurements each ensemble member will do something different. The spread between the samples is indicative of the uncertainty.

4.6.4 Post-processing

The Global air temperature estimates, v1.0 product is provided on a regular latitude-longitude grid with a grid cell spacing of 0.25° in latitude and longitude. The conversion from the triangular mesh to the equirectangular grid is performed by evaluating the value of the SPDE at the centre of the gridcell. This value will be a combination of the basis functions on nearby nodes of the triangular mesh.

The uncertainty is evaluated by drawing 30 samples from the posterior of the statistical model (as for the ensemble) and calculating the standard deviation of them. Note that more samples are used to calculate the grid cell uncertainty than are available in the ensemble, but using a finite number of samples means that the uncertainties are correct to within around 10%.

The uncertainty combines all components of uncertainty in the analysis, including uncertainty in the climatology, large-scale and local components of the statistical model. In addition the files contain a variable, "observation influence", running from zero to one, which gives the degree to which a grid cell value for the local component is constrained by nearby observations. This is calculated by subtracting the ratio of posterior and prior variances from 1. Values close to 1 give an indication of when daily variability in the analysis is constrained and will be of use to those using the daily temperatures in their analysis. Values close to 0 indicate that the daily local analysis is unconstrained or weakly constrained by observations. Note that the large-scale and climatology components can still provide useful information on slower changes in these areas as they aggregate information across a wider area and time scale.

In some times and places there are insufficient data to constrain the analysis, so these areas have been masked. In addition, in a few limited areas the statistical model produced extreme climatological values; these were also masked. The following criteria were used to mask the infilled analysis:

- Values outside the range -80°C to 60°C were masked
- Values exceeding four times the interquartile range of the regional distribution (after adjustment for altitude) were masked
- Areas where the uncertainty in the climatology component exceeded 0.7°C were masked
- Areas where the uncertainty in the large-scale component exceeded 0.7°C were masked
- After 2012, all ocean areas were masked as there were no marine data inputs after 31 December 2012.

Consequently, the fields are not globally complete.

4.6.5 Detailed format specification and metadata

The dataset comprises one file per day. The format is given in the NetCDF file and reproduced below.

Table 4-41 Dimensions

time	UNLIMITED; // (1 currently)
bounds	2
latitude	720
longitude	1440

Table 4-42: Variables

Variable (dimensions)	Standard name	Description
time(time)	time	Time at zero longitude in days since 1850-01-01T00:00:00Z
timebounds(time, bounds)		
latitude(latitude)	latitude	Latitude in degrees north
longitude(longitude)	longitude	Longitude in degrees east
timeoffset(longitude)	none	Local time offset from UTC (days)
tas(time, latitude, longitude)	air_temperature	Mean daily surface air temperature in K
tasuncertainty(time, latitude, longitude)	none	Total uncertainty in average daily surface air temperature
tasobservationinfluence(time, latitude, longitude)	none	Observation influence for average daily surface air temperature
tasensemble_0(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 0

tasensemble_1(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 1
tasensemble_2(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 2
tasensemble_3(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 3
tasensemble_4(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 4
tasensemble_5(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 5
tasensemble_6(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 6
tasensemble_7(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 7
tasensemble_8(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 8
tasensemble_9(time, latitude, longitude)	none	Average daily surface air temperature ensemble member 9

4.6.6 Understanding and using uncertainty estimates

The main product file contains a total uncertainty. This is estimated from 30 samples drawn from the posterior of the statistical distribution. The estimated uncertainty should be within 10% or so of the true value. It is not possible to propagate this total uncertainty as this is a point value without additional information about the covariance, i.e. about how the errors vary in space and time.

In addition to the total uncertainty provided at each point, there is an ensemble of 10 samples drawn from the posterior of the statistical distribution. The samples represent the covariance described by the statistical model.

Propagation of uncertainty using the ensemble is simple. A user can calculate their required diagnostic, for example the area-weighted mean temperature over Europe, separately for each of the 10 ensemble members. The mean of the 10 members will provide a central, or best estimate, and the standard deviation will provide an estimate of the uncertainty (or standard error). Some care needs to be taken in interpreting this standard deviation as the ensemble is relatively small.

4.6.7 Where to go for further information

Further details of the statistical infilling are provided in the scientific user guide and in the papers describing the data set and statistical methods.

The code used to run the analysis will be made publicly available.

4.6.8 Known limitations

When using the data, users should be aware of the following:

- The data set is not globally complete. There are some areas, represented by the missing data flag in the data files, where there are insufficient data to make a reliable estimate of the air temperature. The trimming of the fields is based on the assessed uncertainty and extreme value checks.
- Although the product is relatively high resolution for global daily temperature product, at 0.25° in latitude and longitude, temperature variability within a grid cell means that there might not be a perfect correspondence between a grid cell value and a temperature measured by a station within that grid cell.
- The uncertainty information indicates values that are more or less reliable. Validation of the uncertainty information can be found in Veal (2019). However, some things are worth bearing in mind. First, the uncertainty is estimated from only 30 samples which is a relatively small number. The uncertainty should typically be accurate to around 10% of its value. Second, far away from observations, the uncertainty is determined by the prior values of the statistical model (summarised in Table 4-38, Table 4-39, and Table 4-40). Although, the prior values are reasonable, they have not been explicitly fitted to the data and globally representative values are used. This means, for example, that uncertainty might be overestimated over unobserved ocean regions because the variability of ocean areas is lower than the global mean.
- The data set provides information for Europe which is consistent with other long-term surface temperature data sets on the continental scale from 1895 onwards. For North America, it is consistent with other long-term surface temperature data sets on the continental scale from 1870 onwards.

There are some known problems with the Global air temperature estimates, v1.0 analysis.

- There is a warm bias in ocean temperatures during the Second World War (1942-1946) and in the 1850s. This is due to well-documented biases in the marine air temperature measurements at these times. During the Second World War, measurements were thought to have been made inside rather than out on deck. In the 1850s, measurements were made at local noon, and few night time measurements (as used at other times for the analysis) were available. The data were retained because the alternative would be no data over the ocean at these times and they still may provide useful information to some users despite the biases.
- A cold bias affects Africa and parts of southern Asia starting in 2000. This is associated with residual biases from the EUSTACE satellite derived air temperatures over land. The statistical analysis calculates a daily, spatially-varying bias adjustment for the satellite-derived air temperatures, but over Africa and southern Asia there are too few daily in situ data to effectively adjust the biases, which therefore reverts to its prior value of zero bias, albeit with large uncertainty.
- Although the data have been masked in unconstrained regions, there are still areas early in the record where the analysis takes unrealistic high values. These are close to the northern and

southern limits of the available data and do not affect data rich areas. Parts of South America are also affected between 1945 and 1957.

- Trends over Australia are not accurately represented prior to 1955.
- Also see the validation results section below.

4.6.9 Validation Results

- Over ocean, the EUSTACE Global air temperature estimates, v1.0 performs well with a global median discrepancy of 0.00 K (Robust Standard Deviation (RSD) 1.15 K) against an independent subset of HadNMAT2 inputs over 1850-2012; also demonstrating good stability across different latitudes and with time.
- The EUSTACE Global air temperature estimates, v1.0 also performs well in most land regions with a global median discrepancy (RSD) against a withheld subset of data from the EUSTACE Global Station Dataset of -0.13K (1.76K) over 2002-2015 and -0.23K (1.76 K) over 1850-2015; the negative discrepancy arises largely over some regions of Africa and the west of North America. However seasonal median discrepancies over central Asia can be high, 6-10 K in DJF at some stations (these most erroneous values are likely to have been masked out of the product). Comparison to the validation results of the EUSTACE Air temperature estimates from satellite v1.0 demonstrates that significant seasonal biases have successfully been removed by the analysis method over land in most locations.
- The EUSTACE Global air temperature estimates, v1.0 displays regional median discrepancies (RSDs) over land-ice (including the Antarctic ice-shelf) against independent station data, 2001-2009, of +0.57 K (3.57 K) in the Northern Hemisphere and +0.27 K (2.38 K) in the Antarctic. Over 1890-2015, the EUSTACE Global air temperature estimates, v1.0 shows discrepancies of +0.37 K (4.04 K) over Northern Hemisphere land ice and +0.47 K (2.68 K) in the Antarctic.
- Over sea ice, the EUSTACE Global air temperature estimates, v1.0 shows regional median discrepancies (RSDs) against ice buoy data of +0.25 K (3.17 K) in the Arctic and -0.29 K (4.14 K) in the Antarctic sea ice regions. These results represent a reduction in bias compared to those seen in the EUSTACE Air temperature estimates from satellite v1.0 in these regions. Median discrepancies and RSDs for sea-ice are larger in the years after 2009, when the satellite air temperature estimates end in these regions. Over 1890-2015, the EUSTACE Global air temperature estimates, v1.0 shows discrepancies of +1.19 K (4.60 K) over Northern Hemisphere sea ice and +4.76 K (6.81 K) over Southern Hemisphere sea ice. The increase in positive bias when the whole period is considered arises from a drift in the EUSTACE Global air temperature estimates, v1.0 over the Poles prior to about 1960; these erroneous data have not been released in these regions, as they do not provide useful information.

References

Kent, E. C., Rayner, N. A., Berry, D. I., Saunby, M., Moat, B. I., Kennedy, J. J., and Parker, D. E. (2013), Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set, J. Geophys. Res. Atmos., 118, 1281– 1298, doi:10.1002/jgrd.50152.

Lindgren, F., et al. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion), J. Roy. Statist. Soc. B, 73(4), 423–498.

Veal, K. L. (2019) Validation report for the final in-filled EUSTACE surface air temperature product, EUSTACE Deliverable 3.5

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z. , Eric Freeman, J. , Berry, D. I., Brohan, P. , Kent, E. C., Reynolds, R. W., Smith, S. R. and Wilkinson, C. (2011), ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol., 31: 951-967. doi:10.1002/joc.2103

4.7 Coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0

4.7.1 Summary

This dataset contains matched in situ reference and EUSTACE temperature estimates used in the validation of EUSTACE Globally gridded clear-sky daily air temperature estimates from satellites product v1.0 (Section 4.5) and the EUSTACE Global daily air temperature product v1.0 (Section 4.6). The data are made available to allow traceability of the validation. Some of the ice datasets used in validation do not allow onward redistribution, so these data are not being released.

These data have been matched to the full analysis fields produced in the creation of the EUSTACE Global daily air temperature combining surface and satellite data product, not the masked fields that are released in the final version. The location of the matchups means this is not likely to make much practical difference to the results, but some matchup values may been seen in areas that are masked in the final product; these are likely to be more extreme values.

Table 4-43: List of in situ reference data used for validation of EUSTACE product

	EUSTACE test product			
	EUSTACE Global air temperature estimates, v1.0	EUSTACE Air temperature estimates from satellite, v1.0		
	Global	Ocean	Land	Ice[8]
In situ Reference				
Global Tropical Moored Buoy Array (GTMBA)	✓	✓	✗	✗
HadNMAT2 inputs (HADNMAT2)	✓	✓	✗	✗
EUSTACE Global Station Dataset (EGSD)	✓	✗	✓	✗
University of Leicester in situ Matchup DataBase (ULMDB)	✓	✗	✓	✗
DMI Quality Controlled Station data (DMIQC)	✓	✗	✗	✓[8]
DMI Quality Controlled Ice Buoy Data (DMIQC_SEA)	✓	✗	✗	✓[8]

[8] As permission for onward distribution was not received for some validation data over ice regions, the ice validation matchups have not been publicly released.

The matchups were produced separately for each test-reference dataset pairing. Therefore each file contains only matchups of one EUSTACE product to one in situ dataset.

The matchup process varied between platforms and as a consequence the filename format, file format and variables contained within each file depend on the matched datasets. Broadly the files can be divided into two types: those containing matchups to in situ station data and those containing matchups to data obtained from in situ mobile platforms.

4.7.2 Data Access

The data are archived at the Centre for Environmental Data Analysis and can be accessed and downloaded as described in Section 3.2 above. It is made available for non-commercial purposes under the Non-Commercial Government Licence, due to non-commercial restrictions on the input in-situ datasets.

Table 4-44: How to access the dataset: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015

Data Product	CEDA catalogue record	Download links
EUSTACE: coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/4b34a2c6890f4e518cacc88911193354	http://data.ceda.ac.uk/neodc/eustace/data/validation/v1.0/

4.7.3 File name format

In the case of the files containing matchups to in situ station data, the filename has the following format:

> [domain]_[test eustace product]_[eustace code revision]_matchups_with_[reference in situ dataset]_insitu_[year].nc

An example is

> land_eustace_0_R001336_matchups_with_eust2_insitu_2007.nc

For matchups to in situ mobile platform data, the date format is slightly different:

> [domain]_[test eustace product]_[eustace code revision]_matchups_with_[reference in situ dataset]_insitu_[year][month].nc

For example,

> ocean_eustace_analysis_R001400_matchups_with_hadnmat2_insitu_201212.nc

In the above the domain is 'land' or 'ocean'; the test EUSTACE product is 'eustace_0' for EUSTACE air temperatures from satellite or 'eustace_analysis' for the EUSTACE Global daily air temperature combining surface and satellite data. The reference in situ dataset codes are given below.

Table 4-45: Reference in situ dataset codes for those datasets that have been released.

reference in situ dataset codes	Explanation
eust2	EUSTACE Global Station Dataset stations withheld from both relationship building and in-filled analysis
gtmba	Global Tropical Moored Buoy Array
hadnmat2	Hadley Centre / National Oceanography Centre Nighttime Marine Air Temperature inputs corrected to 2m
ulmdb	University of Leicester Matchup Database

4.7.4 File format

4.7.4.1 Stationary Platforms

In the case of ULMDB, GTMBA and DMIQC stations, daily statistics were calculated from reference observations and matched to the EUSTACE grid box estimate of daily temperature. The temperatures in the EGSD data are already in the form of daily statistics (Tmin and Tmax).

Each file contains one year of matched data.

Table 4-46: Variables in the stationary platforms data files

Variable	Type	Dimension	Units	Comments
station_code	Char	Station		Station ID
time	Double	Time	days	Time (Julian date) at zero longitude
test_timeoffset	Double	Station	days	Local time offset from UTC at centre of gridbox
ref_timeoffset	Double	Station	days	Local time offset from UTC at in situ station
xc	Short	Station	1	x-coordinate on EUSTACE grid
yc	Short	Station	1	y-coordinate on EUSTACE grid
longitude	Float	Station	degrees_east	Longitude coordinate of reference station

latitude	Float	Station	Degrees_north	Latitude coordinate of reference station
test_tas	Float	Time , Station	K	EUSTACE gridded daily mean temperature
test_tas_unc	Float	Time , Station	K	Uncertainty estimate associated with EUSTACE temprature
test_tas_observationinfluence	Float	Time , Station	1	Observation influence for EUSTACE infilled average daily surface air temperature
ref_tasmin	Float	Time , Station	K	Daily Tmin at in situ station
ref_tasmin_qc	Byte	Time , Station		Quality control flags (see GHCN-D documentation)
ref_tasmin_def	Byte	Time , Station		Definition flags
ref_tasmax	Float	Time , Station	K	Daily Tmax at in situ station
ref_tasmax_qc	Byte	Time , Station		Quality control flags (see GHCN-D documentation)
ref_tasmax_def	Byte	Time , Station		Definition flags
data_source	short	Station		Data source flag

4.7.4.2 Mobile platforms

In the case of HADNMAT2 ships and DMIQC buoys each in situ observation is matched to the corresponding EUSTACE daily temperature.

Each file contains one month of matched data.

Table 4-47: Variables in the mobile platform data files

Variable	Type	Dimension	Units	Comments
id_code	Char	record		Platform ID
ref_time	Double	Record	days	Julian date of reference observation
ref_lon	Float	Record	Degrees East	longitude of reference observation

ref_lat	Float	Record	Degrees North	latitude of reference observation
ref_tas	Float	Record	K	in situ air temperature
test_time	Double	Record	days	Julian date of test estimate at start of day at Greenwich Meridian"
test_xcoord	Float	Record	1	test grid x coordinate
test_ycoord	Float	Record	1	test grid y coordinate
test_tas	Float	Record	K	daily mean surface temperature on test grid
test_tas_unc	Float	Record	K	uncertainty of daily mean surface temperature on test grid
test_tas_observationinfluence	float	record	1	Observation influence for EUSTACE infilled average daily surface air temperature

4.7.5 Where to go for further Information

Veal (2019), Validation report for the final in-filled EUSTACE surface air temperature product, EUSTACE, Deliverable D3.5

5 Use case examples

This section provides some examples of applications that the EUSTACE products might be suitable for. The EUSTACE products have not yet been used for these applications, but the following sections provide guidance on considerations that should be borne in mind when thinking of using the EUSTACE products for such applications.

5.1 Using EUSTACE temperature data for calibration or validation in climate model research

5.1.1 Introduction/background

Climate models are always run for a historical period and (often) also for the future to get an idea of the climate change that we can expect. The quality of the projections for the future can only be determined indirectly, since we, of course, have no observations for the future. For the past we do have observations, therefore the climate model runs for the past are compared with the historical observations. The Global daily air temperature combining surface and satellite data (described in Section 4.6) can also be used for the historical temperature observations.

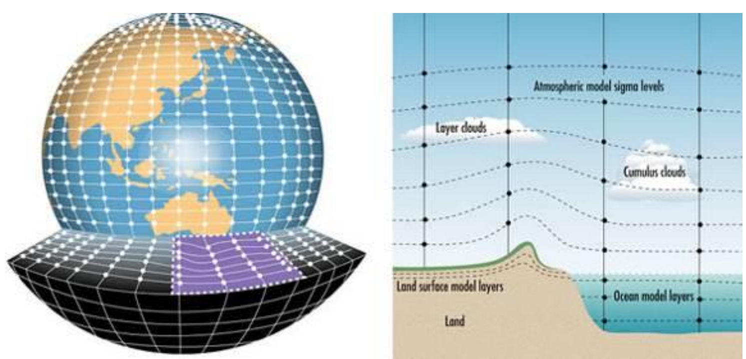


Figure 5-1 Schematic presentation of a climate model: the earth is subdivided in many grids in the horizontal and vertical directions: the climate model simulates the climate for each grid.

Climate models calculate the average of the climate variables over a grid (Figure 5-1). Grid size varies between climate models (Figure 5-2). In global models the spatial resolution is relatively coarse, although now also high resolution global simulations are under development with grid sizes of about 25 km (e.g. in the PRIMAVERA project[9]). Observational station data cannot be compared directly with the climate model data, since the station data give point measurements and the climate model give area average data. This

[9] For more info on the project look at the website: https://www.primavera-h2020.eu/.

is especially a problem for precipitation data (since some rainfall is very local), but also for temperature in a spatially heterogeneous area this may cause difficulties.

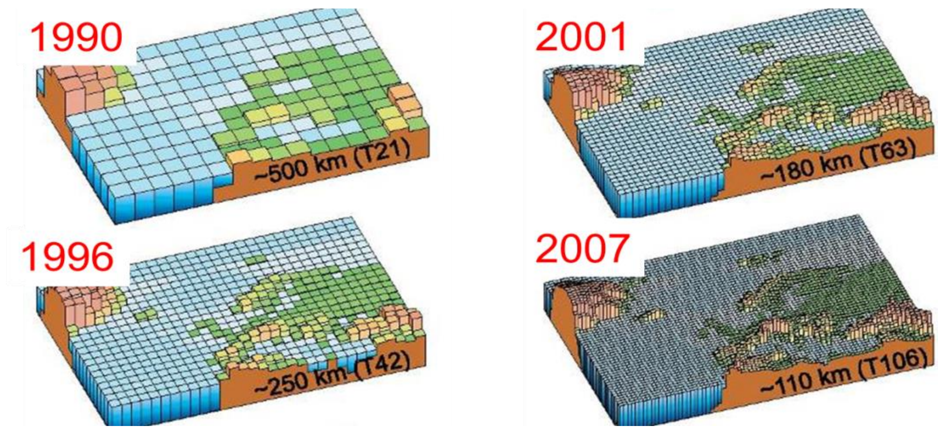


Figure 5-2 Evolution of spatial resolution in Global Climate Models used for IPCC assessment reports (Source: www.wmo.int/pages/themes/climate/climate_models.php).

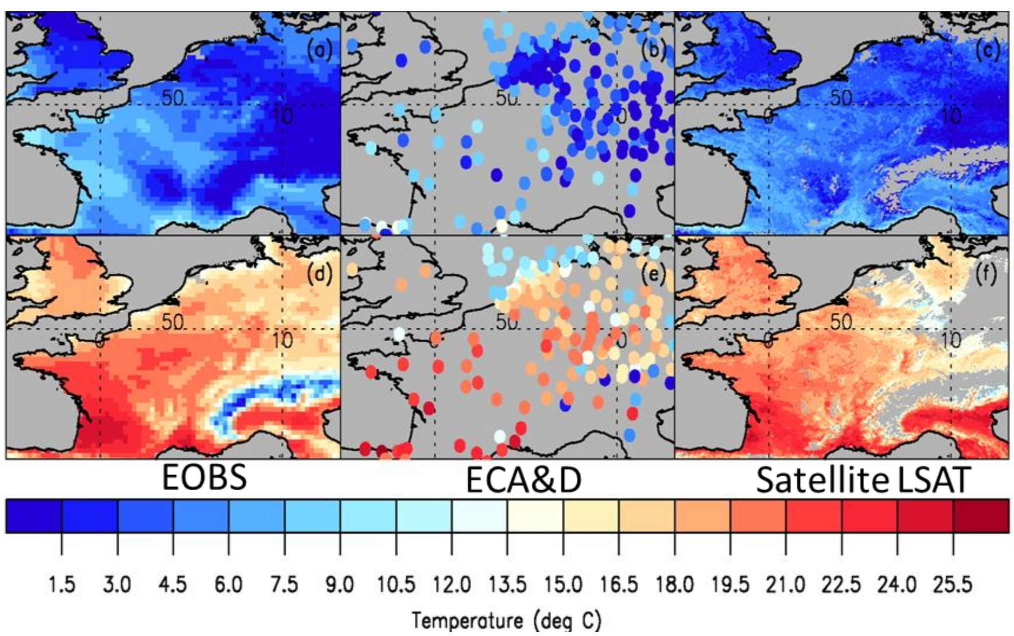


Figure 5-**3 Differences in spatial resolution based on different data sources (E-OBS: based on interpolation of station data from ECA&D taking into account height) (Source: E. Good)**

Station data are only available for a limited number of locations, whereas climate models simulate the climate for a large region or the whole globe. To overcome this, the following approaches can be followed:

- Station data can be interpolated in a sophisticated way (e.g. taking into account the effect of height on temperature, or the effect of the presence of the sea or large water bodies). The resulting gridded datasets give estimates of the e.g. area average temperature in the past at locations where no observations took place. An example of such a dataset is the E-OBS data set (see also section 4.4).

- Another way to create a dataset with a better spatial coverage is the use of re-analysis. In reanalysis the climate in the past is simulated with a weather model integrating as much observational data as possible.
- Use of satellite data. With satellites several climate variables can be measured, directly or indirectly. The advantage is the good spatial coverage and higher resolution (Figure 5-3). When measured indirectly the measured variable should be translated to the required climate variable. EUSTACE has developed such a data set, where the measured surface or "skin" temperature is translated to the air temperature.

In climate models every grid is assigned a certain "surface": land, sea, ice, etc. To assign a grid to a certain surface a "mask" is used (Figure **5-4**). Although the grid size has decreased considerably over time, it is still not exactly the same as in reality. In reality grids may contain two or more types of surfaces (e.g. land and ocean). The satellite observations used in EUSTACE are based on reality and in the project different methods are used to estimate the air temperature from the skin temperature above different surfaces. For a grid that contains in reality more than one surface, estimates of the air temperature for the various surfaces within that grid are made within EUSTACE. This is possible since the spatial resolution of the used satellite data is clearly higher than the final spatial resolution in the EUSTACE data set (about 0.25˚).

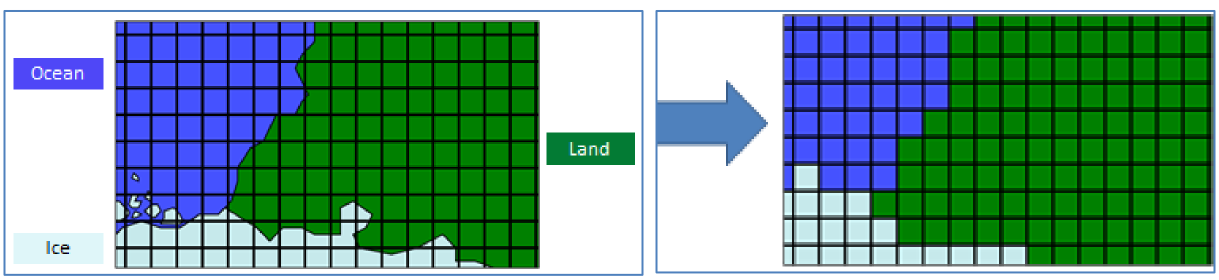


Figure 5-4: Left: hypothetical distribution of the "surfaces" ocean, land and ice over different grids. Right: a way to create a "mask" where only one surface is assigned to a grid cell (in this case land (green) is used in preference to ice (light blue) and ocean (blue), and ice is used in preference to ocean).

5.1.2 How can the EUSTACE dataset be used in PRIMAVERA?

PRIMAVERA is working on high resolution global climate modelling. Several models will be run on a 0.25° spatial resolution and will be compared with coarser resolutions used until now. The idea behind this is that several processes may be simulated better due to the higher resolution. As part of the process of climate model development the model results are always validated. The currently available sources of data for the past-current climate do have some disadvantages:

- Interpolated station data: the E-OBS dataset is available from 1950 on, but it is available only for land and only for Europe. The dataset is available on a 0.25° or 0.10° spatial resolution and for determining the temperature per grid only the land surface is taken into account. Therefore, grids with sea and land only give the average temperature for the area over land. (An advantage is the availability of precipitation and sea level pressure)
- Re-analysis: currently available global re-analyses have a relatively coarse spatial resolution (https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables)[10]
- Satellite based observational data sets: Until now these datasets were available for a limited period, the data were only available for a limited area and/or data were missing for days/periods with no measurements due to clouds. The EUSTACE project tries to overcome these shortcomings of satellite based data by creating a data set for all surfaces and using a sophisticated statistical method for filling in the days/period without data and for extrapolating to the past.

During the development of the data file structure for the EUSTACE products, someone from the PRIMAVERA project was interviewed. The final EUSTACE dataset (section 4.6) is interesting to them since it will be a global dataset covering all surfaces, and since it has a high resolution (similar to the one they use for their highest resolution runs):

- **Structure of the dataset:** Masks for land-water used in climate models are not the same as the real "masks". This has to do with the modelling resolution (small islands may not be explicitly included, Scandinavia may be connected with Denmark, etc.). Having estimates for air temperature for the different surfaces separately would allow them to make a dataset for validation with air temperatures for the same surfaces per grid cell as used in the climate model runs. This would allow them to make a better or "fairer" validation.
- **Use of uncertainty information:** For validation of climate model data it is important to have an idea of the uncertainty in "observations": If there is wide variation in air surface "observations" (or estimates) then the simulated air temperature by the climate model may still be in the range of the "observations" although at first sight they may seem to differ a lot. This may clearly affect the conclusion about whether the climate model run has or does not have a (large) bias.

Validation of climate model projections is done by comparing the statistics produced by the climate model with the statistics of the "observations", e.g. for the average temperature in a month, or for the highest temperature reached per year. In the final EUSTACE dataset information is given on the uncertainty per day, but also an ensemble is produced to describe the uncertainty. If one wants to take into account the uncertainty in the EUSTACE dataset for validation of climate model projections, the ensemble can be used best.

5.2 Using EUSTACE temperature data for calculating temperature indices

5.2.1 Introduction/background

[10] ERA-5 has a clearly higher spatial resolution, almost comparable with EUSTACE data.

On the basis of time series of air temperature (minimum, average and maximum; daily values) many different temperature indices can be derived. These indices are often used in specific sectors or used for a more general public. The indices give an (indirect) indication of the impacts of the climate or climate change. The (change) in e.g. the number of ice days (maximum temperature below 0 ˚C) gives an indication of the accessibility of unpaved roads in Scandinavia during winter or the number of tropical nights (minimum temperature of 20 ˚C or higher) can be an indication of heat stress. At the website of ECA&D (European Climate Assessment Database; www.ecad.eu/) a large number of indices are generated (http://www.ecad.eu/indicesextremes/indicesdictionary.php). Similar databases are set up also for part of South America (LACA&D; http://lacad.ciifen.org/), and South East Asia (SACA&D; http://sacad.database.bmkg.go.id/).

Station data are only available for a limited number of locations, but people may also want to have data for their own particular location/village. To overcome this, station data can be interpolated in a sophisticated way (e.g. taking into account the effect of height on temperature, or the effect of the presence of the sea or large water bodies). The resulting gridded datasets give estimates of the e.g. area average temperature in the past in locations where no observations took place. An example of such a dataset is the E-OBS data set (section 4.4). However, the spatial resolution and accuracy depends very much on the density of stations (see Figure 5-5). For temperature, also satellite data can be used. The advantage is the good spatial coverage and higher resolution (Figure 5-3). Satellites measure "skin temperature", although more often air temperature is used for indices. However, skin temperatures can be used to estimate air temperatures. EUSTACE has developed an example of such a data set, where the measured surface or "skin" temperature is translated to the air temperature by determining the relationships for several surfaces over land, oceans, ice and lakes.

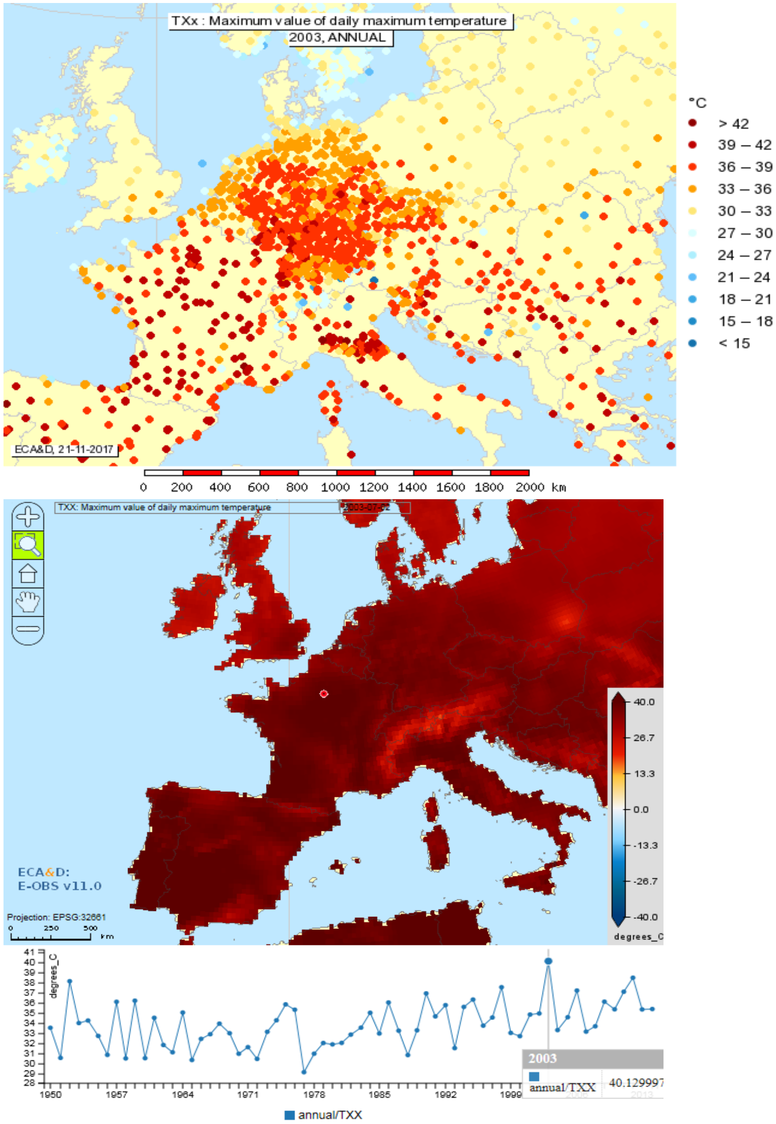


Figure 5-5 Example from the ECA&D website for the maximum value of the maximum temperature in the year 2003 (high temperatures in a large part of Europe). On top: information based on station data; Bottom: information based on interpolated station data.

5.2.2 How can the EUSTACE dataset be used for LACA&D?

Where limited station data are available, the satellite data can be of added value for determining more local temperature indices, since they have complete spatial coverage (at least the infilled EUSTACE dataset, 4.6) at relatively high spatial resolution. For the land area where the Latin American version of ECA&D, LACA&D, is working, less station data are available than for Europe in ECA&D. The length of the time series may in general also be shorter. Besides, the EUSTACE dataset will also provide average daily air temperatures above other surface than land (e.g. the ocean). With the EUSTACE data various indices can be calculated at a higher spatial resolution than possible until now.

The uncertainty information that will be provided with the EUSTACE dataset, makes it possible to calculate the uncertainty in the indices. For many indices the uncertainty cannot be calculated by just using the uncertainty measure that is given in the file with the “best estimate”. The value of the index depends very much on the absolute daily value: in the case of an estimated minimum temperature of -0.5 ˚C with a standard error (measure for the spread of the ensemble) of about 1 ˚C the values minus or plus 1 standard deviation would range from –0.5 to 1.5˚ C. When calculating the number of frost days (minimum temperature of 0 ˚C or lower), on one side of the range the day would be considered a frost day, whereas on the other side of the range not. The use of an ensemble of time series, together describing well the total uncertainty would in that case be the most appropriate way to determine the uncertainty of the index “frost days”.

5.3 Attribution of climate change

5.3.1 Introduction and background

Detection of climate change is demonstrating whether a climate has changed significantly, or in other words showing that the change is more than can be explained with natural variability alone. Attribution of climate change is the effort to scientifically determine what could be the cause of observed climate change. In most cases this focusses on determining whether the change could be due to the increase in Greenhouse gases (GHG) and not by natural variability alone. IPCC assessment reports have paid attention to attribution of global and regional temperature increases.

When extreme weather events occur, often the question arises whether the event is due to climate change. A single event cannot be attributed to climate change, but we might be able to indicate whether the return time of certain extreme events has changed significantly due to climate change or whether there is a significant trend in the likelihood of extreme events. One of the first studies on this so-called *event attribution* showed that the risk for a European heat wave such as in 2003 had clearly increased due to climate change (Stott et al., 2004). Methodologies for attribution of extreme events have been developed in several projects, among others in the EUropean CLimate and weather Events: Interpretation and Attribution (EUCLEIA[11]) and the World Weather Attribution (WWA[12]) projects (follow up of EUCLEIA). The projects try to provide information on attribution relatively fast after the extreme weather events have occurred, when there is a need and still interest from decision makers and the public. In both projects several types of extreme weather events are considered, including extreme heat and cold. Below we will indicate how the EUSTACE dataset can contribute potentially to the attribution of extreme warm and cold weather events. In between the description of the methods used for event attribution we describe in italics the potential use of the EUSTACE dataset and the aspects that people have to take into account when using it.

5.3.2 Methods for extreme weather event attribution

To determine whether the probability of extremes has increased, long time series are needed and different types of climate data are combined (e.g. observations and climate models).

The first step in the attribution studies is to select appropriate climate indices. This can be a few-days-extreme (three-day maximum temperature average as used for the analysis of the European heat waves in 2017 and 2018; Figure 5-6) or multiple-weeks-extreme (as used in the analysis of record June temperatures in western Europe in 2017, or the two-week cold outbreak in northern North America). In some of these analyses, station data are used (for one location) and in others the average temperatures over regions are used.

[11] https://eucleia.eu/

[12] https://www.worldweatherattribution.org/

If EUSTACE data are to be used for such analysis, this requires good representation of the extremes (maximum, minimum and average temperatures) at daily, weekly and monthly scales and good spatial representation (e.g. spatial extend of heat waves or cold spells). In the main EUSTACE dataset the best estimate is presented and also the "observation influence" is given. When little local information is available e.g. from nearby stations for a particular period, the observation influence will be high and the estimate of the daily temperature will be "smoothed" to the longer term average. The EUSTACE ensemble that is also produced will contain a better estimate of the extreme high or low temperatures[13].

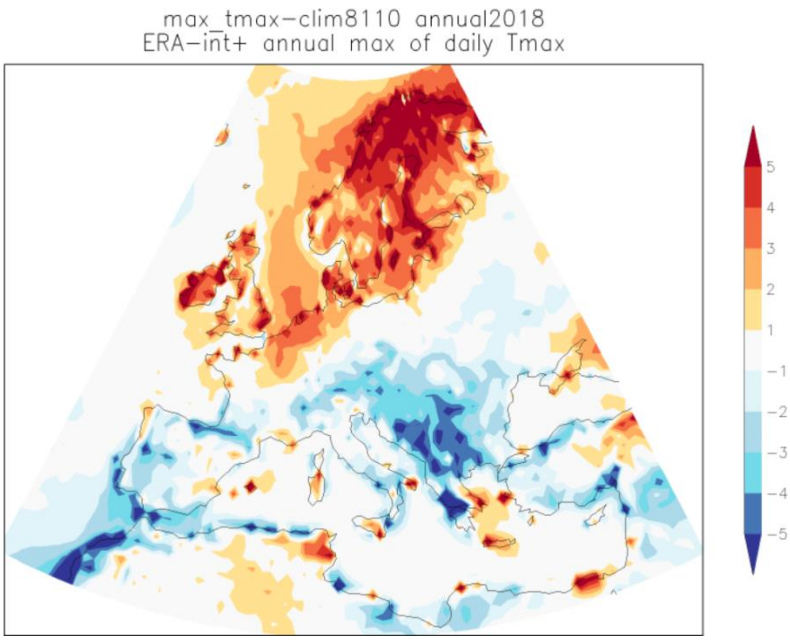


Figure 5-6: The hottest 3-day average of the maximum temperature in 2018 (ECMWF analyses up to 24 July, forecasts up to 31 July) compared to the highest 3-day maximum temperature in the period 1981-2010 that is currently the "normal" period (based on ERA-interim re-analysis). Along coasts there are artefacts from comparing the high-resolution ECMWF analyses with the lower-resolution ERA-interim reanalysis (Source: https://www.worldweatherattribution.org/attribution-of-the-2018-heat-in-northern-europe/).

reanalysis (Source: https://www.worldweatherattribution.org/attribution-of-the-2018-heat-in-northern-europe/).

The second step involves determining the probability of the extreme event in the recent climate (in the "normal" period: Figure 5-6 above) and its probability in the past when the concentration of greenhouse

[13] In some areas the total uncertainty is under- or overestimated. In case of underestimation of the uncertainty the range of potential values in the ensemble will also be underestimated. Information about this is available in the user guide under validation.

gases (GHGs) was much lower. Historical observations are regularly not available for a long enough period to enable such an evaluation[14] (NAS, 2016).

For determining the probabilities, observational data are used and a Generalized Extreme Value Distribution (GEV) or other statistical distributions are fitted and confidence intervals are determined (Figure 5-7 below). After this it is determined whether there are significant changes or trends in the occurrence of these extreme events (detection whether the climate has changed).

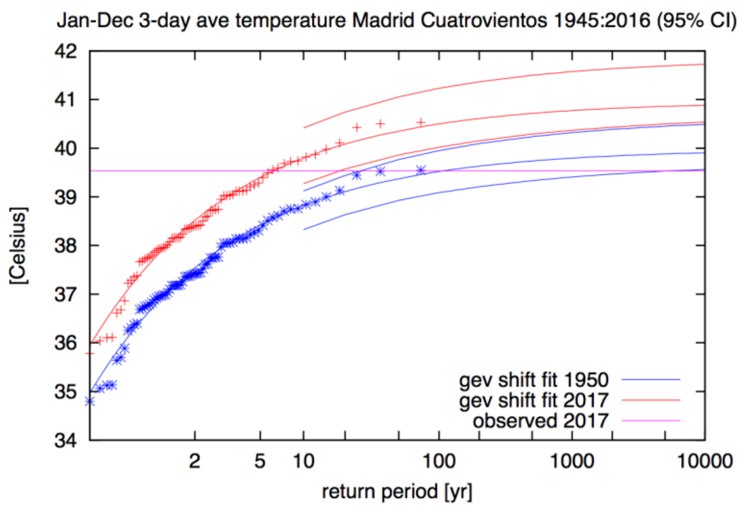


Figure 5-7: Example of the determination of the 3 day mean maximum temperature for station Madrid Cuatrovientos in the current climate (2017) and around 1950, for an event attribution study on the heat wave in southern Europe in 2017.

The EUSTACE dataset is a reconstruction of the weather in the past using observations, satellite data and the spatial differences between locations. The dataset represents an alternative to station data or re-analysis data for determining how extreme a certain event is in the current climate (often 1981-2010 is used). The EUSTACE dataset provides a clearly higher spatial resolution than the station data in most regions in the world. The newest re-analyses have a similar spatial resolution to the EUSTACE dataset, but they cannot always represent extremes very well. Especially in the period with satellite data, the EUSTACE dataset may have added value since spatial differences and extremes may be represented better.

The EUSTACE dataset provides much longer time series and for many more locations in the world than with station data alone. As such the dataset may make it easier to determine how extreme a certain event is in the current climate and whether the probability has changed compared to periods with clearly lower GHG concentrations. However, in regions with relatively few observations in the past, the estimates of the daily temperatures will contain larger uncertainties and extremes are represented potentially less well.

[14] In model based approaches simulated weather and climate phenomena are compared under different input conditions: for instance, with and without human-caused changes in GHGs. Also combined approaches exist (NAS, 2016).

When determining the confidence intervals for the probabilities one has to take into account the uncertainties due to natural variability, but also the uncertainties in the estimates of the temperatures (with the help of the ensemble or the other uncertainty information). In such regions the combined uncertainty will be much larger and it will be more difficult to find significant differences in probabilities, especially for extremes such as the above mentioned 3-day average maximum temperature. This is also true for periods in the 19th century and in the early 20th century with relatively few station data.

During the generation of the EUSTACE data set inhomogeneities (or break points) in the underlying observational data sets are corrected for. Therefore, it is less probable[15] *that changes in probabilities in the EUSTACE datasets are caused by changes in station locations, changes in instruments, etc. (*Figure 5-8 *below for example of inhomogeneity).*

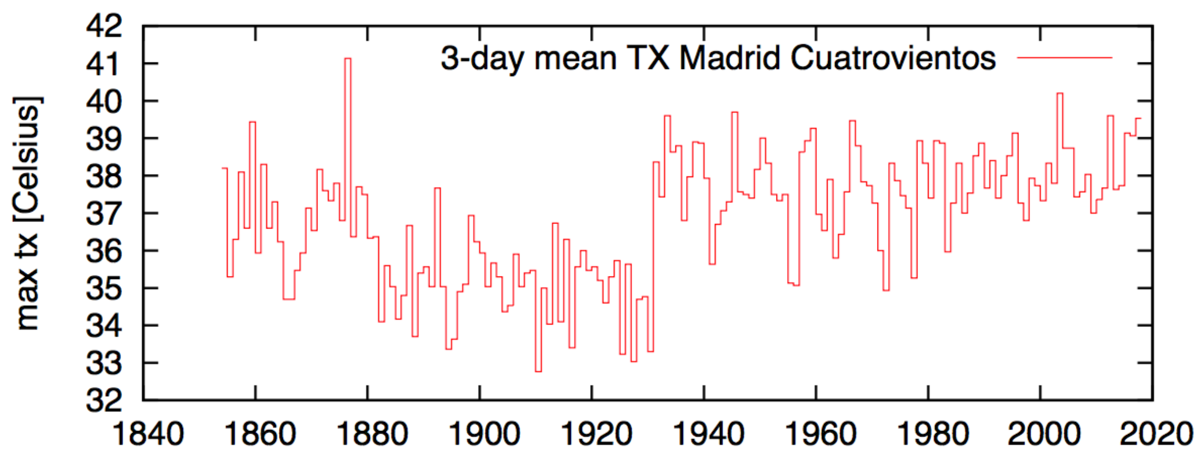


Figure 5-8: Inhomogeneity in the 3-day mean maximum temperature for station Madrid Cuatrovientos in the early 1930's. With these inhomogeneities it is not possible to determine changes in probabilities

[15] As explained in section 4.2 not all break points can be detected, but the most obvious ones are detected, and are corrected for in the EUSTACE method.

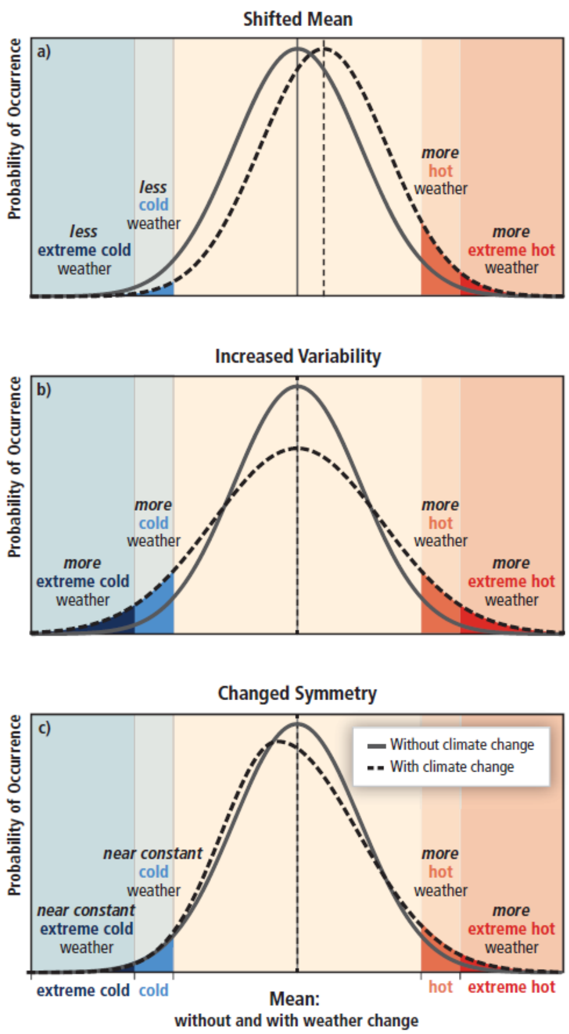


Figure 5-9: Probability density functions showing different ways in which the climate can change, all resulting in changing probabilities of extremes (Source IPCC, 2012).

In the **third step** it is determined whether the change can be due to the increase of GHGs. When climate change is detected in step 2, it is not clear yet whether there is a relationship with the increase of GHGs. For the attribution, a link is sought with a factor that is clearly influenced by the increase of GHGs. The most obvious climate variable is temperature. For this in the statistical approach, global warming is factored in by allowing the fit to the distribution to be a function of the global mean surface temperature (GMST). Thus the probability density function is shifted up or down with GMST but does not change shape (Figure 5-9 above). In this way, it results in a distribution that varies continuously with GMST. This distribution can be evaluated for a GMST in the past (e.g., 1950 or 1900) and for the current GMST. If there is a clear relation between the climate index under study and the GMST (and thus with the increase of GHGs) the probability of the climate index under study will have changed over time.

In the methods used by EUCLEIA and WWA the observations are also compared with climate models that are available and suitable for the climate variables and regions under study. This allows to study better what is the cause of the observed change in climate. This requires that in the used models the variability of the studied extremes is compatible with the observed variability.

The EUSTACE dataset could be used to determine the Global mean surface air temperature and its change or the mean surface air temperature for a region.

The EUSTACE dataset can also be potentially used for determining whether the variability of the studied temperature extremes in the climate models is compatible with those in observations. This can only be done for periods and regions where we know that the day-to-day variability (and thus the extremes) is represented well (low observation influence in final EUSTACE dataset). This is probably the case for the period with satellite data and periods and regions with sufficient station data. Important for this comparison is also that the dataset is homogeneous. During the generation of the EUSTACE data set inhomogeneities (or break points) in the underlying observational data sets are corrected for as much as possible. Changes in probabilities in the EUSTACE datasets will, therefore, probably not be caused by changes in station locations, changes in instruments, etc.

5.3.3 References

IPCC, 2012. Attribution of Extreme Weather Events in the Context of Climate Change. (Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (Eds.)). https://doi.org/10.17226/21852.

NAS, 2016. Attribution of Extreme Weather Events in the Context of Climate Change. Washington, DC: The National Academies Press. https://doi.org/10.17226/21852.

Stott, P.A., D.A. Stone & M.R. Allen, 2004. Human Contribution to the European Heatwave of 003. Nature 432(7017):610-4. DOI: · 10.1038/nature03089

Otto, F. E. L., G. J. van Oldenborgh, J. Eden, P.A. Stott, D. J. Karoly, M.R. Allen. “The Attribution question”. *Nature Climate Change 6* (2016): 813 – 816. doi:10.1038/nclimate3089. https://www.nature.com/articles/nclimate3089#affil-auth

5.4 Analysis of extremes

Besides the attribution of extreme weather events, the EUSTACE data may also be used for other analyses of extremes.

5.4.1 Summer of 2010 in Europe

In Summer 2010, Russia was hit by record temperatures, meagre rainfall amounts and subsequent crop loss, peat and forest fires. The extreme situation was mostly due to an exceptionally strong and persistent area of high pressure that lasted for weeks, favouring southerly flow, ample sunshine and little rain. This "blocking pattern", which effectively prevents the normal east-to-west movement of weather systems that typically bring cooler temperatures and rain, began in early July and lasted through mid-August. According to NOAA, it was the most extreme and longest lasting blocking pattern since 1920. In the E-OBS anomaly map below, one can see that for vast areas of Western Russia and Eastern Europe, mean temperatures were between 4 and 8°C above normal during July and the first two weeks of August 2010.

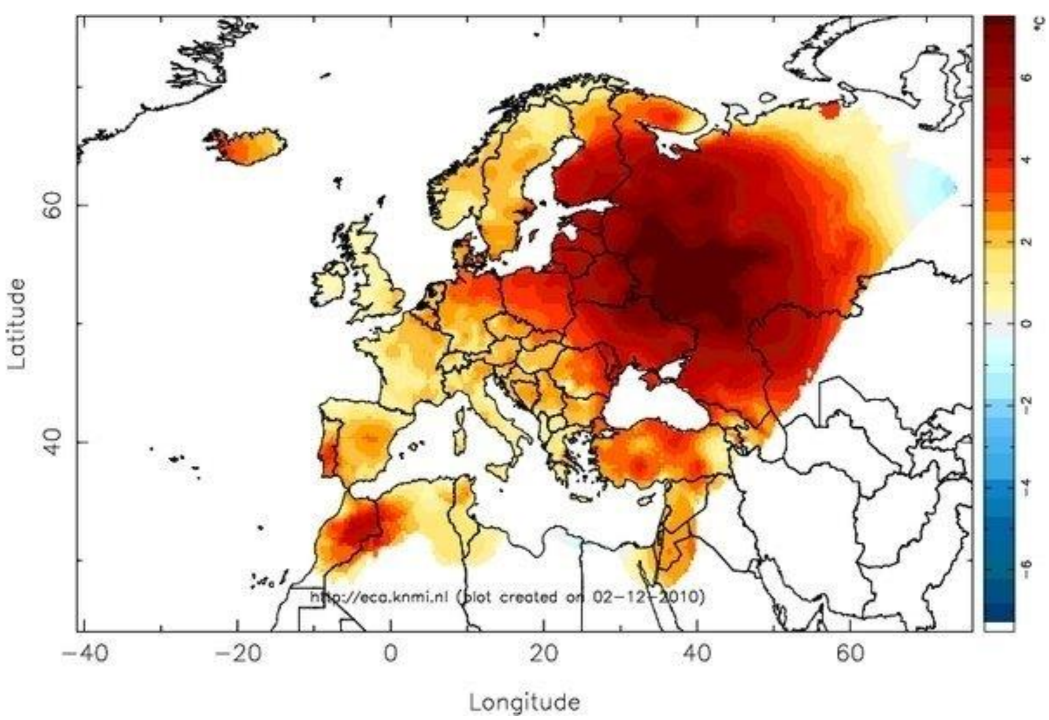


Figure 5-10: E-OBS anomalies for mean temperature during July and the first two weeks of August 2010 compared to the normal period 1961-1990.

The number of summer days, SU, where the maximum temperature is above 25°C was also well above normal. The following ECA&D anomaly map shows that for many stations, there were at least 10, and up to 30, more summer days than normal during July 2010. In Moscow, the number of summer days in July was 31; the normal is 9.5 in a year.

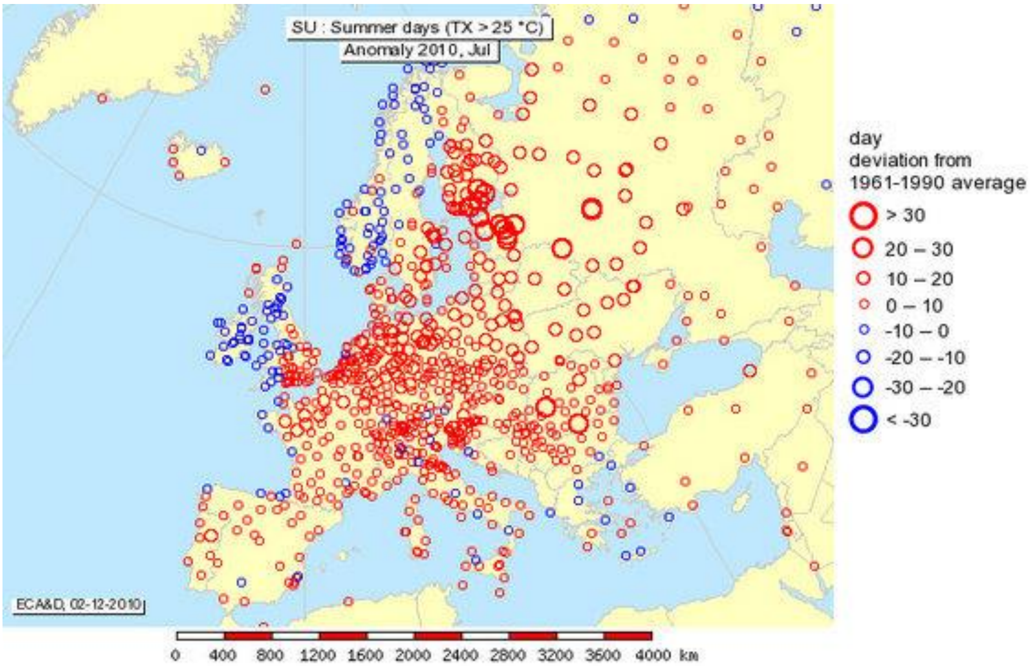


Figure 5-11: ECA&D anomaly map of the number of summer days, where the maximum temperature is above 25°C, during July 2010 compared to the normal period 1961-1990. Please click on the map for its most recent version.

The heat wave also resulted in many maximum temperature records being broken. In Moscow, the old record maximum temperature of 35.6°C, set on 11 July 1996 according to ECA&D data which begins in 1936 for this station, was broken each day from 23 July through 29 July, culminating in the city's highest temperature ever recorded: 38.2°C. Although this day was indeed extremely hot for Moscow, it was not the hottest location in the country. As seen below, several ECA&D stations southeast of Moscow reported temperatures of over 40°C on 29 July, including 41.5°C at the station Alexandrov Gaj in the province of Saratov near the border of Kazakhstan.

5.4.2 Winter of 2010 in Europe

The winter of 2010 was unusually cold and snowy for most of Europe. On most days between mid-December and mid-March, the mean temperature was below normal across much of the continent. Below are E-OBS mean temperature anomalies for December 2009 through February 2010.

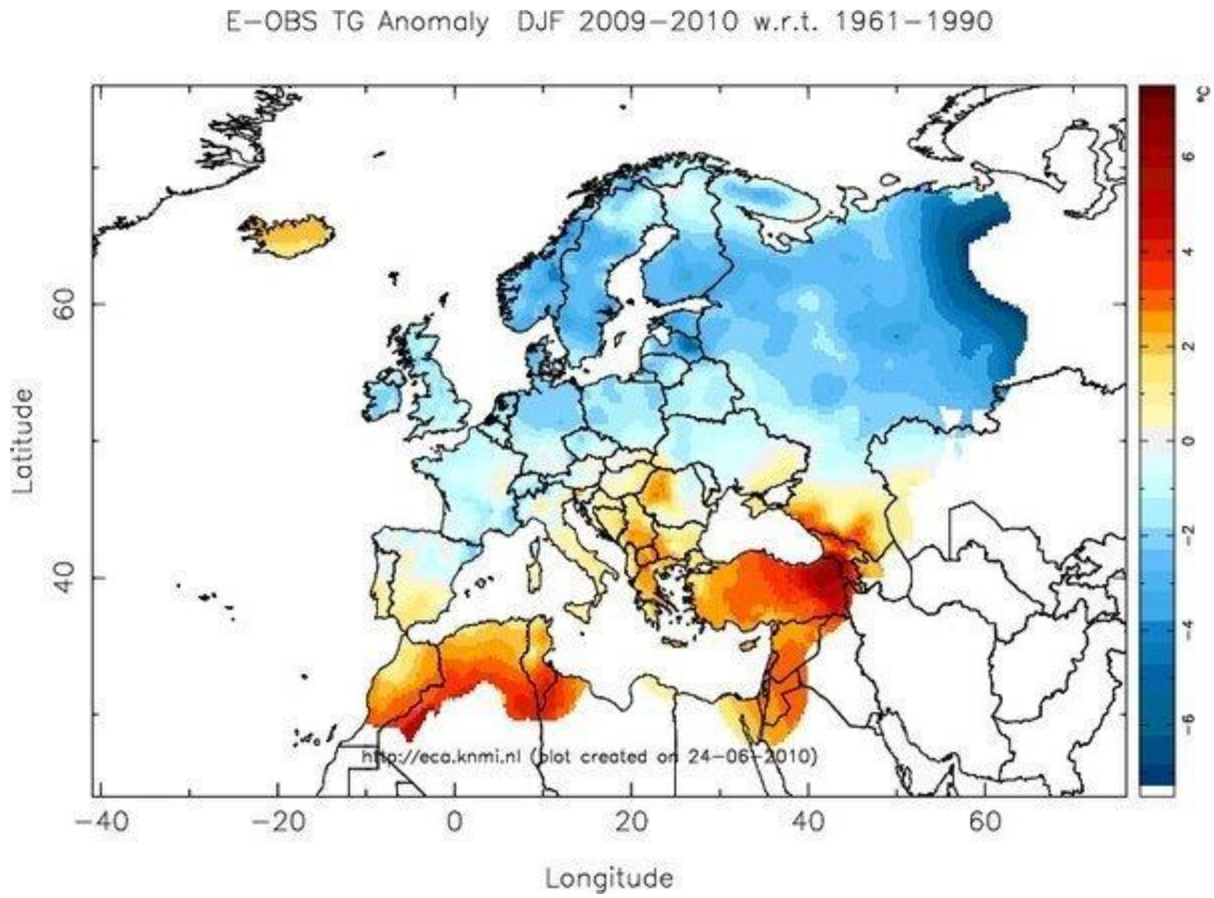


Figure 5-12: ***E-OBS anomalies for mean temperature for December 2009 through February 2010 compared to the normal period 1961-1990.***

The anomalously cold temperatures contributed to significantly more ice days, where the maximum temperature is below freezing, than normal. Most areas saw between 10 and 20 additional ice days from December through February, but some areas in Southern Scandinavia had over 30 more than normal (See the ECA&D anomaly map below).

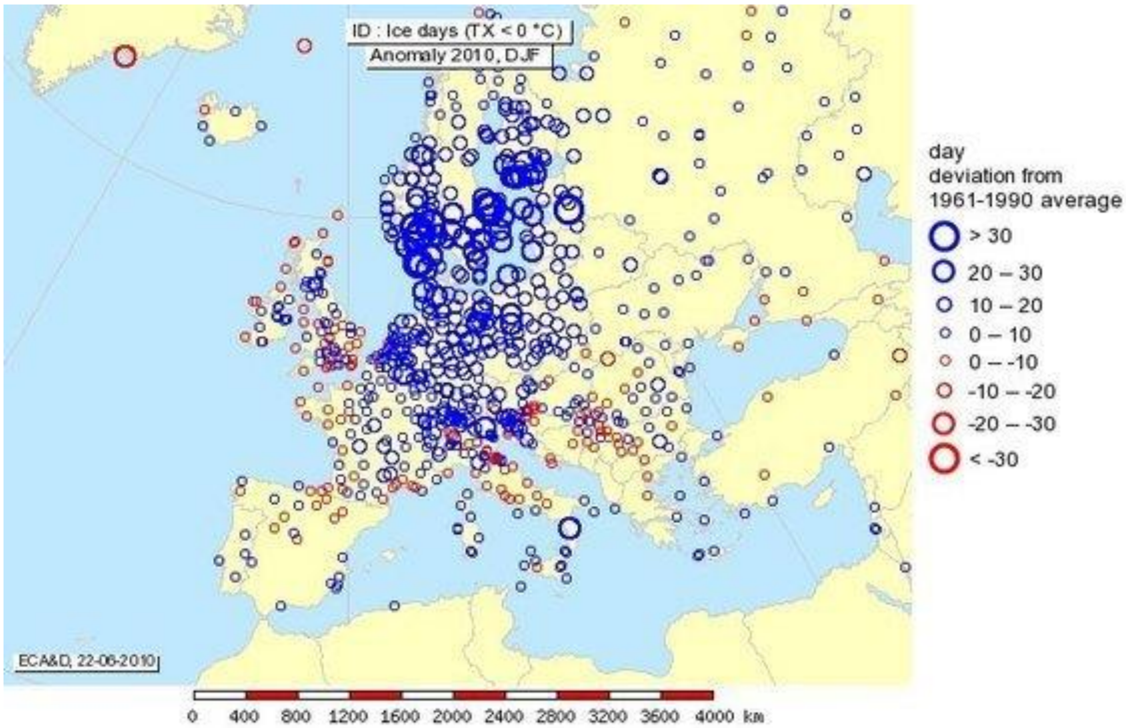


Figure 5-13: ECA&D anomaly map of the number of ice days, where the maximum temperature is below 0°C, for December 2009 through February 2010 when compared to the normal period 1961-1990. Please click on the map for its most recent version.

One of the coldest days was 19 December 2009. This day was extremely cold across most of Europe with large areas experiencing minimum temperatures below -20°C, including -44°C at the North-western Russian Federation station Ust Tzilma, -34°C at Drevsjo in Norway, and -27°C on the Zugspitze summit in Germany.

5.5 Use of breakpoints for reliability of trends

In Section 4.2.10 an example is described on how the information on break points in the dataset "Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015" can be used to check the reliability of trends in the time series of the land station data.

The dataset "European land station daily air temperature measurements, homogenised" has been corrected for inhomogeneities. Therefore, trends in these time series are considered relatively reliable. The homogenisation procedures have the drawback in areas where the station density is low, that an insufficient number of neighbouring reference stations can be found for a reliable break detection and homogeneity adjustment. This holds in particular for the early part of the records, when the number of stations is low. Therefore, one should be more careful with the interpretation of trends in regions with few stations and in the case of long time series (Section 4.3.5).

6 References

Brugnara, Y. et al., 2019. The EUSTACE global land station daily air temperature dataset, submitted to Geoscience Data Journal

Dybkjaer et al (2014) Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST, NACLIM Deliverable D32.28, https://naclim.cen.uni-hamburg.de/fileadmin/user_upload/naclim/Archive/Deliverables/D32.28-SUBMITTED-2014-10_PU.docx

Dybkjaer et al (2018) Algorithm theoretical basis document for the OSI SAF Sea and Sea Ice Surface Temperature L2 processing chain, http://osisaf.met.no/docs/osisaf_cdop3_ss2_atbd_hl-l2-sst-ist_v1p4.pdf

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010). Comprehensive automated quality assurance of daily surface observations. Journal of Applied Meteorology and Climatology, 49, 1615-1633.

Ghent, D., I. Trigo, A. Pires, O. Sardou, J. Bruniquel, F. Göttsche, M. Martin, C. Prigent, and C. Jimenez (2018), ESA DUE GlobTemperature Product User Guide, Rep. GlobT-WP3-DEL-11, University of Leicester. [http://www.globtemperature.info/index.php/public-documentation/deliverables-1/108-globtemperature-product-user-guide]

Høyer, J. L., E. Good, P. Nielsen-Englyst, K. S. Madsen, I. Woolway, J. Kennedy, 2018: Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes, EUSTACE Deliverable 1.5, https://www.eustaceproject.org/eustace/static/media/uploads/d1.5_revised.pdf

Hunziker, S., Gubler, S., Calle Fernandez, J. M., Moreno, I., Andrade, M.F., Velarde, F., Carrasco, G., Castellón, Y., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., & Brönnimann, S. (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. International Journal of Climatology, 37, 4131-4143.

IPCC, 2012. Attribution of Extreme Weather Events in the Context of Climate Change. (Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (Eds.)). https://doi.org/10.17226/21852.

Kent, E. C., Rayner, N. A., Berry, D. I., Saunby, M., Moat, B. I., Kennedy, J. J., and Parker, D. E. (2013), Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set, J. Geophys. Res. Atmos., 118, 1281– 1298, doi:10.1002/jgrd.50152.

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, 1441-1453.

Lindgren, F., et al. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion), J. Roy. Statist. Soc. B, 73(4), 423–498.

Martin, M., and Goettsche, F.M. (2017) GlobTemperature Satellite LST Validation Report (Report to ESA: GlobTemp_DEL-12_i3r0) [http://www.globtemperature.info/index.php/public-documentation/deliverables-1/117-validation-report-del12]

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. Journal of Atmospheric and Oceanic Technology, 29, 897-910.

Merchant, C., Ghent, D., Kennedy, J., Good, E., Hoeyer, J., 'Common approach to providing uncertainty estimates across all surfaces', EUSTACE Deliverable D1.2, 2015, https://www.eustaceproject.org/eustace/static/media/uploads/Deliverables/eustace_d1-2.pdf

NAS, 2016. Attribution of Extreme Weather Events in the Context of Climate Change. Washington, DC: The National Academies Press. https://doi.org/10.17226/21852.

Otto, F. E. L., G. J. van Oldenborgh, J. Eden, P.A. Stott, D. J. Karoly, M.R. Allen. "The Attribution question". *Nature Climate Change 6* (2016): 813 – 816. doi:10.1038/nclimate3089. https://www.nature.com/articles/nclimate3089#affil-auth

Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., et al. (2014). The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. Geoscience Data Journal, 1, 75-102.

Stickler, A., Brönnimann, S., Valente, M. A., Bethke, J., Sterin, A., Jourdain, S., et al. (2014). ERA-CLIM: historical surface and upper-air data for future reanalyses. Bulletin of the American Meteorological Society, 95, 1419-1430.012

Squintu, A.A., G. van der Schrier, Yuri Brugnara, Albert Klein Tank, 2019a: Homogenization of daily temperature series in the European Climate Assessment & Dataset, International Journal of Climatology, 39:1243-1261. doi.org/10.1002/joc.5874

Squintu, A.A., G. van der Schrier, E. J. M. van den Besselaar, R. C. Cornes, A. Klein Tank, 2019b. Building long homogeneous temperature series across Europe: a new approach for the blending of neighboring series, submitted to the Journal of Applied Meteorology and Climatology

Steinke, S., Eikenberg, S., Löhnert, U., Dick, G., Klocke, D., Di Girolamo, P., & Crewell, S. (2015). Assessment of small-scale integrated water vapour variability during HOPE. Atmospheric Chemistry and Physics, 15(5), 2675–2692. https://doi.org/10.5194/acp-15-2675-2015

Stott, P.A., D.A. Stone & M.R. Allen, 2004. Human Contribution to the European Heatwave of 003. Nature 432(7017):610-4. DOI: · 10.1038/nature03089

van den Besselaar, E.J.M, A.M.G. Klein Tank, G. van der Schrier, P.D. Jones, 2012, Synoptic messages to extend climate data records, J. Geophys. Res., 117, D07101, doi:10.1029/2011JD016687

Veal, K. L. (2019) Validation report for the final in-filled EUSTACE surface air temperature product, EUSTACE Deliverable 3.5

VIM (2012), International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition) JCGM 200:2012 (JCGM 200:2008 with minor corrections)

Vogelmann, H., Sussmann, R., Trickl, T., & Reichert, A. 2015). Spatiotemporal variability of water vapor investigated using lidar and FTIR vertical soundings above the Zugspitze. Atmospheric Chemistry and Physics, 15(6), 3135–3148. https://doi.org/10.5194/acp-15-3135-2015

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z. , Eric Freeman, J. , Berry, D. I., Brohan, P. , Kent, E. C., Reynolds, R. W., Smith, S. R. and Wilkinson, C. (2011), ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol., 31: 951-967. doi:10.1002/joc.2103

7 How to Cite the Datasets

Dataset	Citation
Global clear-sky land surface temperature data from MODIS Terra on the satellite swath with estimates of uncertainty components, v2.1, 2000-2016	Ghent, D.; Veal, K.; Dodd, E. (2019): EUSTACE/GlobTemperature: Global clear-sky land surface temperature data from MODIS Terra on the satellite swath with estimates of uncertainty components, v2.1, 2000-2016. Centre for Environmental Data Analysis, *18 March 2019*. doi:10.5285/655866af94cd4fa6af67809657b275c3. http://dx.doi.org/10.5285/655866af94cd4fa6af67809657b275c3
Global clear-sky land surface temperature from MODIS Aqua on the satellite swath with estimates of uncertainty components, v2.1, 2002-2016	Ghent, D.; Veal, K.; Dodd, E. (2019): EUSTACE/GlobTemperature: Global clear-sky land surface temperature from MODIS Aqua on the satellite swath with estimates of uncertainty components, v2.1, 2002-2016. Centre for Environmental Data Analysis, *18 March 2019*. doi:10.5285/0f1a958a130547febd40057f5ec1c837. http://dx.doi.org/10.5285/0f1a958a130547febd40057f5ec1c837
Global clear-sky sea surface temperature from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012	Embury, O.; Carrea, L.; Woolway, R.I. (2019): EUSTACE / CCI: Global clear-sky sea surface temperature from the (A)ATSR series at 0.25 degrees with estimates of uncertainty components, v1.2, 1991-2012. Centre for Environmental Data Analysis, *08 April 2019*. doi:10.5285/b8285969426a4e00b7481434291ad603. http://dx.doi.org/10.5285/b8285969426a4e00b7481434291ad603
Global clear-sky ice surface temperature from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009	Høyer, J.L.; Dybkjær, G.; Eastwood, S.; Madsen, K.S. (9999): EUSTACE/AASTI: Global clear-sky ice surface temperature data from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009. Centre for Environmental Data Analysis, *date of citation*. http://catalogue.ceda.ac.uk/uuid/60b820fa10804fca9c3f1ddfa5ef42a1
Global land station daily air temperature measurements with non-climatic discontinuities	Brugnara, Y.; Brönnimann, S.; Good, E.J.; Squintu, A.; van der Schrier, G. (2019): EUSTACE: Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015. Centre for Environmental Data Analysis, *22 February 2019*. doi:10.5285/7925ded722d743fa8259a93acc7073f2. http://dx.doi.org/10.5285/7925ded722d743fa8259a93acc7073f2

identified, for 1850-2015.	
European land station daily air temperature measurements, homogenised	Klein Tank, A. M. G., 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.
Gridded European surface air temperature based on homogenised land station records since 1950	Cornes, R. C., van der Schrier, G., Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. /Journal of Geophysical Research: Atmospheres/, 123, 9391– 9409. https://doi.org/10.1029/2017JD028200
Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016	Kennedy, J.J.; Capponi, F.; Ghent, D.; Good, E.J.; Høyer, J.L.; Kent, E.C.; Madsen, K.S.; Mitchelson, J.R.; Nielsen-Englyst, P.; Tonboe, R.T. (2019): EUSTACE: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016. Centre for Environmental Data Analysis, *08 April 2019*. doi:10.5285/f883e197594f4fbaae6edebafb3fddb3. http://dx.doi.org/10.5285/f883e197594f4fbaae6edebafb3fddb3.
Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0	Morice, C.P.; Capponi, F.; Kennedy, J.J.; Killick, R.E.; Lindgren, F.; Mitchelson, J.R.; Rayner, N.A.; Winn, J.P. (2019): EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0. Centre for Environmental Data Analysis, *30 May 2019*. doi:10.5285/468abcf18372425791a31d15a41348d9. http://dx.doi.org/10.5285/468abcf18372425791a31d15a41348d9
Coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0	Veal, K.; Ghent, D. (2019): EUSTACE: Coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0. Centre for Environmental Data Analysis, *31 May 2019.* doi: 10.5285/4b34a2c6890f4e518cacc88911193354. http://dx.doi.org/10.5285/4b34a2c6890f4e518cacc88911193354

Please also cite the EUSTACE overview paper:Rayner et al. The EUSTACE project: delivering global, daily information on surface air temperature, in preparation

Annex 1 Glossary of terms and acronyms

A	
AASTI	Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors
AIRS	Atmospheric InfraRed Sounder
Air temperature	See section 2.1.1
ARM	Atmospheric Radiation Measurement programme
Average temperature	See section 2.1.1
C	
CCI	Climate Change Initiative
CDG	Climate Data Guide (https://climatedataguide.ucar.edu/)
CMSAF	EUMETSAT's Climate and Monitoring, Satellite Application Facility
CEDA	Centre for Environmental Data Analysis
CEMS	Climate and Environmental Monitoring from Space
D	
DEM	Digital Elevation Model
DMI	Danish Meteorological Institute
DOI	Digital Object Identifiers
E	
EC	European Commission
ECA&D	European Climate Assessment Database (https://www.ecad.eu/)
ECV	Essential Climate Variables
E-OBS	gridded dataset based on data from ECA&D (https://www.ecad.eu/download/ensembles/ensembles.php)
ERA	ECMWF Reanalysis
ESA	European Space Agency
ESA CCI	European Space Agency Climate Change Initiative
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EUSTACE	EU Surface Temperature for All Corners of Earth
F	
FVC	Fraction of Vegetation Cover
G	
GHCN-D	Global Historical Climate Network – Daily
GlobTemp	GlobTemperature project (http://www.globtemperature.info/)
H	
HadISD	Met Office Hadley Centre Integrated Surface Dataset
I	
IAT	Near surface air temperatures over ice
ICA&D	International Climate Assessment Database (https://www.ecad.eu/icad.php)
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IPCC	Intergovernmental Panel on Climate Change

IPCC DDC	IPCC Data Distribution Centre
IST	Ice Surface Temperature
J	
JULES	Joint UK Land Environment Simulator
K	
KNMI	Royal Dutch Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut)
L	
Locally systematic uncertainty	See section 2.1.2
LSA-SAF	EUMETSAT's Land Surface Analysis Satellite Application Facility
LSAT	Land Surface Air Temperature
LST	Land Surface Temperature
LSTday	Daytime LST
LSTngt	Night time LST
LSWT	Lake Surface Water Temperature
M	
MAD	Mean Absolute Difference
MAT	Marine Air Temperature
Mean temperature	See section 2.1.1
MDB	Matchup DataBase
MODIS	MODerate resolution Imaging Spectroradiometer (satellite)
N	
NACLIM	North Atlantic CLIMate (project)
NASA	National Aeronautics and Space Administration
NH	Northern Hemisphere
NMAT	Night time Marine Air Temperature
NWP	Numerical Weather Prediction
O	
OSI-SAF	EUMETSAT's Ocean and Sea Ice Satellite Application Facility
P	
PROMICE	PROGramme for Monitoring of the Greenland ICE Sheet
R	
Random uncertainty	See section 2.1.2
RMSD	Root Mean Square Difference
S	
SAF	Satellite Application Facilities
SAT	Surface Air Temperature
SEVIRI	Spinning Enhanced Visible and InfraRed Imager (satellite)
SH	Southern Hemisphere
Skin temperature	See section 2.1.1
SST	Sea Surface Temperature
STFC	Science & Technology Facilities Council

Surface air temperature	See section 2.1.1
Surface skin temperature	See section 2.1.1
Systematic uncertainty	See section 2.1.2
SZA	Solar Zenith Angle
T	
T2m	Screen-level air temperature (at 2 m height, normally measured at in situ station)
Tavg	Averaged daily T2m
Tmax	Maximum daily T2m
Tmin	Minimum daily T2m
Tskin	Skin temperature (measured with satellite)
U	
Uncertainty	See section 2.1.2
W	
WP	Work package

Annex 2 Links to some useful websites and tools

Below a selection of useful links is presented, based on suggestions from potential users and the knowledge of the partners within the EUSTACE project. On other sites also useful links, tools, etc. may be available. If you have more useful links that you want to share with others, send an e-mail to janette.bessembinder@knmi.nl and we may include these in an update of this user guide or make them available through our website.

Portals to temperature datasets:

- Climate Data Guide: https://climatedataguide.ucar.edu/climate-data/global-temperature-data-sets-overview-comparison-table. With summaries, metadata, a comparison table and links to a large number of temperature datasets.

- ECA&D and ICA&D: European and International Climate Assessment Databases: https://www.ecad.eu/. Collects information on station observations in different parts of the world. Where the daily observations are not available, often derived indices and trends are freely available (for non-commercial use). Part of the functionalities is moved to Copernicus Climate Change Services websites.

- E-OBS: gridded dataset bases on data from ECA&D: http://eca.knmi.nl/download/ensembles/ensembles.php. Also a version with the homogenized temperature station data will be made available later on.

- Copernicus Climate Change Services (C3S) Climate Data Store: https://cds.climate.copernicus.eu/#!/home. Through this website a large number of climate data sets will be made available or is already available. The connected Toolbox offers a variety of processing and visualizing tools (or will provide this in the future).

- IPCC DCC: Data Distribution Centre: http://www.ipcc-data.org/index.html. With observational datasets and climate model simulations used for the various Assessment reports. Also guidance material available (http://www.ipcc-data.org/guidelines/index.html)

- Climate Explorer: https://climexp.knmi.nl/start.cgi. Website where many observational, re-analysis and climate model data can be accessed and processed (especially for climate researchers)

Portals with tools to visualize, process, check datasets, etc.

- Climate data guide: https://climatedataguide.ucar.edu/climate-data-tools-and-analysis. With a variety of tools for climate data processing.

- Climate data guide, Common Climate Data Formats: Overview: https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/common-climate-data-formats-overview. Also with some example codes to read, write or change data files

- Panoply: https://www.giss.nasa.gov/tools/panoply/. For viewing and processing of NetCDF, HDF and GRIB data sets (also mentioned in the Climate Data Guide).

- Climate4Impact portal: https://climate4impact.eu/impactportal/general/index.jsp and https://climate4impact.eu/impactportal/data/esgfsearch.jsp. Portal to access, visualize and processing climate data. The ADAGUC viewer is used within this tool. A step-by-step guide with some examples on how to use the portal is included in Annex2 of this Product User Guide.

- KML tool: https://developers.google.com/kml/documentation/. Tool from Google to visualize and process data.

Background information:

- IPCC Glossary: https://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_Glossary.pdf. With many terms related to climate data, climate and climate change.

- General information about climate data. C3S User Learning Services: https://uls.climate.copernicus.eu/login. Portal for on-line learning about many aspects of climate data. Freely available, only registration needed. Related to the Climate Data Store of C3S. For those with very little knowledge about climate data the lesson/resource on "Introduction to climate data" may be a good introduction.

- Common Climate Data Formats: Overview. Climate Data Guide: https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/common-climate-data-formats-overview. Also with some example codes to read, write or change data files

- NetCDF: https://www.unidata.ucar.edu/software/netcdf/ Potential users mentioned that there are some issues with NetCDF data used in certain packages . Check on internet whether this is the case when you encounter problems in a specific package.

Other related projects or initiatives:

- GlobTemperature: http://www.globtemperature.info/. Also with access to other LST (land Surface Temperature) products.
- ESA Climate Change Initiative (CCI): http://cci.esa.int/

Annex 3 Frequently Asked Questions

Which products should I use for x...?

We recommend using several products if possible, as all datasets have pros and cons.

I'm interested in using station data over Europe, which dataset should I use?

For trend analysis, we would recommend the European land station daily air temperature measurements, homogenised (Section 4.3). The homogenised data for a station has been adjusted to remove non-climatic effects. This has been achieved by using neighbouring stations as a reference, so the individual stations may not be independent. The data should have the correct climatic trend, but there will be uncertainty in the absolute values (generally set by reference to the last period). The Global land station daily air temperature measurements with non-climatic discontinuities identified (Section 4.2) dataset has information on likely break points, but is not homogenised.

Which air temperature product should I use?

- Use Global daily air temperature combining surface and satellite data (Section 4.6) and/or Gridded European surface air temperature based on homogenised meteorological station records (Section 4.4) for a gap free product over Europe. Other regions may be more or less complete in the Global daily air temperature combining surface and satellite data product, depending on the time period of interest.
- If the application requires data further back in time than 1950, then use Global daily air temperature combining surface and satellite data (Section 4.6).
- We do not recommend using satellite air temperature products to study the air / skin temperature difference without understanding the product's construction first (since a regression model was used to estimate air temperature and so findings will be influenced by that).

When should I use skin / air temperature?

- Air temperature is the temperature of the air ~2m about the earth surface, and is the temperature reported in weather forecasting.
- Surface skin temperature is the temperature of the upper most surface of the earth, the top few microns of sea, bare land, and probably most corresponds to the temperature of the canopy over vegetation.
- Surface skin temperature is very strongly influenced by solar radiation, so during the day can be much higher or lower than air temperature.
- For different applications either or both could be relevant.

What do we do at the edges between sea, land, ice, lakes? What does the Global daily air temperature combining surface and satellite data product represent at these edgelands?

It will represent a combination of whichever surfaces are in the 0.25 degree latitude by longitude grid box. See Section 4.6 for information on how the Global daily air temperature combining surface and satellite data product is constructed.

How do I average the data?

Averaging of the temperature values can be undertaken as needed by the application. However, please see the relevant sections for each product for advice on propagating uncertainty information when averaging.

Why are satellite air temperatures not spatially complete?

Satellite skin temperature is only retrieved from infrared measurements where there are no clouds between the surface and the sensor.

What does Tmean , Tmin, Tmax actually mean?

- Tmax is the maximum temperature on any given day
- Tmin is the minimum temperature on any given day
- We define in EUSTACE Tmean to be Tmin+Tmax / 2

How can you get a max/min temp from a satellite measuring at a specific time?

For land, the daytime and nighttime skin temperature retrievals were matched with Tmax and Tmin to derive the regression relationships used to estimate Tmax and Tmin. This accounts for difference in time between the measurements.

Why is there no max/min temperature information over the oceans?

Daytime air temp observations are more affected by the observing platform e.g. solar heating of a ships deck, and so aren't representative of the true air temperature. Therefore we used only nighttime ship measurements to derive the regression relationship between sea-surface temperature and marine air temperature. We assume this relationship holds when applied to sea-surface temperature representative of the daily mean to create estimates of Tmean.

Why are there no satellite air temperature over the oceans in the southern oceans?

There are no ship measurements available in this region, so it has not been possible to construct a model of the skin/air temperature relationship there.

How can we get daily global data from 1850?

See Section 4.6 regarding the Global daily air temperature combining surface and satellite data product.

How do I identify which points are well-constrained by observations in the Global daily air temperature combining surface and satellite data?

This product includes an indicator of the observation influence, which can be used to identify points where more measurement information has been used to inform the local component of the analysis.

What are the ensembles? How should they be used?

See Sections 4.4 and 4.6 for descriptions and advice regarding the two ensemble products.