Horizon 2020

H2020-EO-1-2014 NEW IDEAS FOR EARTH-RELEVANT SPACE APPLICATIONS

EUSTACE

(GRANT AGREEMENT 640171)

EU SURFACE TEMPERATURE FOR ALL CORNERS OF EARTH

DELIVERABLE 4.1/ MILESTONE 33

EUSTACE
User requirement specification for product design

Version number	Date updated	Changes made	By whom
0.1	21-9-15		Janette Bessembinder
0.2	28-9-15	Comments Ag Stephens included	Janette Bessembinder
0.3	10-12-15	Comments Nick Rayner/Katie Herring included	Janette Bessembinder
1.0	18-12-15	Comments Darren Ghent, John Kennedy, Finn Lindgren included	Janette Bessembinder

Milestone Title	*User requirement specification for product design (MS33)*	
Deliverable Title	*User requirement specification for other WPs (D4.1)*	
Brief Description	*Discussion and translation of user requirements (gathered through (trailblazer) user engagement and literature review) for aspects relevant for product design*	
WP number	4	
Lead Beneficiary	*KNMI*	
Contributors	*Janette Bessembinder (KNMI)*	
Creation Date Version Number Version Date	21/09/15 1.0 18/12/15	
Milestone Due Date Deliverable Due Date Actual Delivery Date	31/08/15 30/09/15 18/12/15	
Nature of the Deliverable	X	*R – Report*
		DEM – Demonstrator, Pilot, Prototype
		DEC – Dissemination, Exploitation or Communication
		O – Other
Dissemination Level/ Audience		*PU – Public*
	X	*CO - Confidential, only for members of the consortium, including the Commission services*

Contents

Executive summary

For EUSTACE to be able to design and produce relevant and usable products, it was identified at the proposal stage that information on user requirements is needed to guide certain aspects of product design, and that this information is required early on in the product development phase. A report on user requirements was provided in MS30 which gave the first indication of aspects that would be important for product design.

In this document for D4.1/MS33 we provide an overview of EUSTACE products, and we indicate the user requirements that we have chosen to fulfill at this stage and how these are related to information that was gathered from users. It explains where user requirements have been included or excluded in product design, and also provides information on which design criteria will be further detailed at a later stage of the project, once more user feedback has been obtained.

Table 2.1 summarizes the requirements for product design which have been identified and are very important to know at the product design stage. In the right-hand column of this table either the decision on what will be used is specified or further actions to be taken are detailed.

Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (EUSTACE Description of Action, Section B1.1).

No.	Objective	Yes	No
1	Intensively develop the hitherto immature use of Earth Observation estimates of Earth's surface **skin** temperature to enable new Climate Data Records of the surface **air** temperature Essential Climate Variable (ECV) to be created, for all locations over all surfaces of Earth (i.e. land, ocean, ice and lakes), for every day since 1850. EUSTACE will achieve this by: combining information estimated from multiple satellites with surface air temperature measurements made *in situ* and **creating complete analyses** of surface air temperature, through the application of novel statistical in-filling methods.		X
2	Integrate these new daily surface air temperature Climate Data Records into a range of applications in Earth System Science and Climate Services and research, amongst others. EUSTACE will achieve this via the active and continuous engagement of trail-blazer users, and the provision of products through already-existing user community data portals and service mechanisms, in standard formats.	X	
3	Undertake and report detailed research into the relationships between surface skin temperature estimated from Earth Observation satellite measurements and surface air temperature observed *in situ* by conventional measurements, over all surfaces of the Earth, including the polar regions. This is likely to provide information useful for refining coupling in Earth system models.		X
4	Create a sustainable, automated system at an appropriate level of maturity for the potential production of the products beyond the lifetime of the project. To enable this, EUSTACE will also identify Earth Observation and conventional data streams that could be used to update the surface air temperature Climate Data Records in the future, including those from Sentinel missions.		X
5	Extensively validate the new surface air temperature Climate Data Records against independent, surface-based reference data, sourced by the project for this purpose.		X
6	Develop and report new, consistent, validated estimates of uncertainty both in already-existing Earth Observation surface skin temperature estimates and in the new surface air temperature Climate Data Records, at all locations and times across the Earth's surface.		X
7	Develop links with related activities within Europe and beyond to help to ensure the execution of a joined-up work programme, the Copernicus Services and to enable the provision of requirements for the future surface skin temperature and surface air temperature observing system.		X
8	Other – not directly linked to one of the above objectives	X	

Project Information

Project Name	EUSTACE
Description of Work	The aim of EUSTACE is to produce a fully-global daily analysis (or ensemble of analyses) of surface air temperature since 1850, integrating different ground-based and satellite-borne data types. Project Website: www.eustaceproject.eu
Grant Reference(s)	640171 (H2020)
Principle Investigator	Nick Rayner
Start/End Dates	Jan. 1, 2015 to June 30, 2018

Organisation

Communication and dissemination contact and lead	Janette Bessembinder Ag Stephens Nick Rayner/Katie Herring
Project Contact	Nick Rayner (PI) – Met Office
Any other team members with responsibility for communication and dissemination	WP 1: Observation Integration: (Jacob Hoyer, DMI) WP 2: Dataset construction (John Kennedy, Met Office) WP 3: Validation and Intercomparison (Darren Ghent, University of Leicester) WP 4: Janette Bessembinder (KNMI) & Ag Stephens (STFC) WP 5: Scientific Coordination of EUSTACE (Nick Rayner, MO) WP 6: Project management (Katie Herring, MO)

1 Introduction

This report describes the discussions that have taken place between WP4 and the rest of the EUSTACE project team, on the conclusions and other useful information that could be drawn from the report on user requirements (MS30), how this can inform the product design and what information on requirements is still missing for the product design. This is the first version of this report. Discussions on product design will continue throughout the project, as more feedback is collected from the trail blazer users and other potential stakeholders and users[1] of EUSTACE products.

For reference, this chapter starts with the general aim of EUSTACE and a short description of the products to be designed in this project.

1.1 General aims and objective of EUSTACE

EUSTACE addresses the key transformational challenge in utilizing Earth Observation surface skin temperature data, namely the derivation of surface air temperature consistently across all surfaces of Earth from land and lakes to ocean and ice. Surface air temperature is not observed directly by satellite instruments and therefore innovative advances are needed to derive it. The value of such a step forward would be in the availability of novel data sets that Earth System models, including climate models, can utilize directly (EUSTACE project plan, 2014), but also many other users.

This overall aim is translated into the following key-message:

> EUSTACE will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques[2].

1.2 EUSTACE products

EUSTACE will generate Climate Data Records of surface air temperature for the whole of Earth and provide improved Climate Data Records of surface skin temperature:

- Satellite skin temperature retrievals (from other projects) for all surfaces of Earth but with new, consistent uncertainty estimates across surfaces;
- Homogenized meteorological station records of surface air temperature measurements for Europe;
- Global data set of surface air temperature measurements from meteorological stations with discontinuities in each station record identified, where possible;
- An in-filled analysis of European surface air temperature based on homogenized meteorological station records since 1951;
- Surface air temperature estimates (with estimates of uncertainty) for all surfaces of Earth, derived from satellite surface skin temperature retrievals; and
- Globally-complete daily analyses of surface air temperature (with estimates of uncertainty) for the whole of Earth since 1850, based on combined information from satellite and in situ data sources.

In addition EUSTACE will produce:

- A surface air temperature match-up data base for intercomparison of different surface temperature Climate Data Records;
- Relationships between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice, ice sheets, and lakes[3];

[1] Stakeholders include all people potentially interested. They can also be indirect users of the final results. Users are those that will use the results of the project themselves.

[2] This short description of the objective of EUSTACE for a broader public was formulated during the kick-off meeting, based on the project plan. The formulation may be slightly adapted later on in the project.

[3] For these products there may not be specific user requirements. However, traceability from the EUSTACE products to journal articles and the reassurance of a tested system are necessary to score highly on the CORE CLIMAX maturity matrix (and also on other measures of system quality). Users often do not personally need these things, but they are

- "Infilling" methods;
- Coded and tested system for product generation[4];
- User guides for obtaining and using the dataset and further documentation[5];
- Connection of final products to data services (CEDA/ESGF).

important in an indirect way since users may find it desirable to have a datasets which meets certain quality assurance standards (without wanting to know the details).

[4] See foot note 3.

[5] See foot note 3

2 User requirements and product design

In this section we discuss the various user requirements that have been gathered and any possible dilemmas that they pose. We have also indicated why specific decisions to include or exclude certain user wishes were made and which design criteria will be further detailed at a later stage, when more user feedback is obtained. The basis for this is the document on the user requirements (MS30)[6]

When is requirements information needed for product design?
For product design some choices have to be made early on in the development process (e.g. whether minimum and maximum temperatures have to be generated because these would be useful to users). Other decisions about product design can be made at a later stage (e.g. whether a derived product will be generated by EUSTACE, or whether information will be given in the user guides on how users can generate the derived products themselves). The first set of design requirements has to be identified as soon as possible to enable effective project planning and product design (see table 2.1).

Some of the information on user requirements for product design will only become available at a later stage of the project, following feedback from the (trail blazer) users. In particular this includes aspects related to presentation, communication about the products and the details of information required in the product user guides. These types of requirements do not need to be known in detail in the early stages of the project and can be further specified later, but ahead of when user guides and dissemination/communication materials will be produced.

Changes in user requirements?
The use of information on user requirements in the product design is an iterative process. From experience of tailoring climate data/information,[7] we know that the specification of user requirements may change during a project due to a number of reasons, for example:

- Several users (and stakeholders) have difficulty fully articulating their requirements;
- Incorrect interpretation of the expressed user requirements by researchers (e.g. due to differences in language, bias in interpretation);
- New insights from users/stakeholders that only become apparent when they begin working with the data/information.

Therefore, continued contact with users and stakeholders is needed to ensure that the products are relevant and usable. The planned workshops, the trail blazer users and the communication/dissemination activities, are all intended to help gather feedback during the project and detect possible changes in user requirements[8].

Which user requirements to include?
The potential user groups of EUSTACE products may be very diverse, ranging from experienced users of climate data, to people with hardly any knowledge of climate data and no experience with working with large gridded datasets. Within the EUSTACE project there are insufficient resources to tailor the products individually for all potential user groups. Therefore, choices will have to be made about which user requirements can be taken into account and which cannot be.

[6] The report on user requirements also built upon recent work led by KNMI for JPI-Climate (see http://www.climate-services.org/sites/default/files/jpi2012_assessment_of_users_requirements.pdf) in which over 70 user requirements documents were analyzed from several national inventories and EU-projects (ECLISE, CLIMRUN, EURO4M, EUPORIAS, IS-ENES2, ADAM, PROMOTE, MEDIATION, COST-VALUE, International Conference on Climate Service (ICCS) and international inventories (EUMETNET, GFCS) and experience from other projects.
[7] Experience of the author of this document, and from several colleagues in Europe (as discussed during workshops on Climate Services in the past years).
[8] Partners in EUSTACE will be informed about the user feedback in various ways: by participating in user consultation meetings, by reports on user workshops and meetings with trail blazer users, during the monthly WP-leaders teleconferences and/or through updates of the documents on user requirements.

Table 2.1 When is certain information relevant for product design needed?

Requirement for product design	Req'd Before M9	Relevant for*	Product Specification or further actions[9]
Temperature variable: mean, min, max?	x	DS	Mean, Minimum, Maximum Discuss most appropriate definition for day (January 2016)
Which aspects of daily temperature time series especially important?	x	DS/DP	Besides averages, also extreme daily values, length of warm/cold periods, variability
Which derived temperature variables?		DP	To be decided later on in user workshops and/or with trail blazer users
Spatial resolution	X	DS	0.25
Spatial resolution: "Within box variability"	X	DS	Will probably be produced as a by-product of the infilling work (is needed to inform us how far a station might deviate from the grid box average)
Data format	x	DS	CF-NETCDF[10]
Tool for formatting data format		DP	Inventory of existing tools, tools will not be developed within EUSTACE
Tool for processing data sets		DP	Inventory of existing tools, tools will not be developed within EUSTACE
Validation		OP	Against in situ data, information about the quality of the data
Comparison with other datasets		OP	List of datasets to be included to be decided later[11]
Description of "uncertainties"	(x)	OP	For user guides
Uncertainties: std or percentiles from ensemble and best estimate		DS/DP	Which percentiles to be determined later
Uncertainties of derived products		DP	Discuss what is possible. Will probably depend on the type of derived product whether it is possible to generate this. Not planned to provide this, but maybe some guidelines can be given on how to generate this if possible.
Temporal variability	x	DS	Depends on methods and datasets used
Content of user guides		OP	To be decided later after feedback users. Probably different user guides for different user groups needed
Documentation and testing of tools/software		OP	Related to infilling methods, relationships and product generation [12], including some statements about post-project maintenance Contact Earth System Modelers for possible additional requirements
Connection to other data portals		DS/OP	To be decided later after feedback users "Connection" most often will consist of a link or information on another website/portal about the dataset(s) or derived data Connections may also consist of links to websites with tools for processing

* DS = Data sets; DP = Derived products; OP = Other products.

[9] When indicated that decisions will be taken later on, on most cases this will be before months 18-20. In these month several milestones on draft and final versions for the product design (MS 34 and MS15) and user guides (MS35) are planned

[10] See also MS12 on recommendations for data formats

[11] See footnote 6. The report on the Intercomparison with other datasets is foreseen for month 40 (D3.4)

[12] See footnote 3

2.1 Temperature variables

During the EUSTACE kick-off meeting in January 2015, a discussion has taken place regarding whether minimum and maximum temperature datasets could be generated by EUSTACE, in addition to mean daily temperatures. In many sectors minimum and maximum temperatures are required and used much more than mean daily temperatures (e.g. Bessembinder, 2013; Themessl, 2011, COST-VALUE). Since after discussion with the EUSTACE team, it seemed possible to estimate minimum and maximum temperatures over land, it was ultimately decided that these would also be generated.

Many users also ask for good representation of the extremes, length of warm/cold periods and variability (day-to-day or year-to-year) in the data set.

A day can be defined in various ways, e.g local time/time zone, 0:00-24:00 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.

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. Within the project there is limited time to generate such derived products, but some can be generated. As pointed out above, users are often interested in extremes (high and low values, periods with low/high values (absolute temperatures), number of days crossing certain thresholds, distribution), more than in averages. Therefore, it seems more logical to invest some time in generating climatological information on extremes. Which extremes can be decided later on, after more feedback from (trail blazer) users is obtained. One of the trail blazer users (CIIFEN/LACA&D), will also use the data set to generate climate indicators for other users.

2.2 Other climate variables

Although several potential users indicate that they use several other climate variables (e.g. precipitation, amongst others), the generation of these other climate data is beyond the scope of the EUSTACE project.

In the user guides some information could be included on the possibilities, difficulties, possible inconsistencies, etc. of combining the EUSTACE dataset with other data sources. How much information can be given on this will be decided upon at a later stage in the project, and will depend on the feedback of users/stakeholders.

2.3 Spatial resolution

During and after the first General Assembly meeting, the team discussed what spatial resolution the EUSTACE global dataset could be produced at. It was decided that a suitable and usable resolution would be 0.25° for all surface of earth.

From previous research, we know that that several potential users also have requirements for data at a higher spatial resolution (e.g. Bessembinder, 2015; 2013; UKCIP, 2006). The methods used to generate the EUSTACE dataset will be made publicly available. It was suggested by one of the project team members that the methods could be used by others to generate datasets with higher resolution for certain regions of the world (where enough information is available e.g. on ground cover, altitude, etc.). This would increase the impact potential of the EUSTACE work and potentially make it more useful to a wider range of users, but providing higher spatial resolution data for some regions in the world is outside the scope of this project.

Something that may also assist those users who would have a use for data at a higher spatial resolution, is the option of instead of providing averages per grid box, providing information on "within box variability", which would give another level of detail which could be useful. The

ability for EUSTACE to provide this information however also relies on sufficient information being available on variables such as topography and ground cover.

2.4 Comparison with other datasets

Currently several other global or continental datasets are used when data on past temperature are needed. To indicate the advantages and the quality of the EUSTACE products, the EUSTACE dataset should be compared with the most used/most important datasets currently used. This will give essential information to potential users and will determine whether they will consider using the EUSTACE data set or not.

As described in the Description of Action comparisons will be made with CRUTEM4, HadCRUT4 so as to compare to data that uses SST anomalies as a proxy for MAT anomalies, GISTEMP, Berkeley, and GHCN-M gridded anomalies. Comparisons will be made also with ERACLIM2 global, ERA-Interim, MERRA, NCEP CFS, 20CR, and UERRA high resolution products for Europe. Based on user feedback it will be decided whether also other datasets will be included in the comparison.

Many potential users use more local datasets or station data and will be interested in how EUSTACE data compare with station data. During the validation, EUSTACE data will be compared with station data.

2.5 Period and area

During the project planning and proposal preparation phase, the project partners made the decision to generate temperature data from 1850 onwards, for the whole globe.

Previous user engagement has shown that users and stakeholders largely use datasets starting from 1950 up to 1981, often for describing the current climate. Only a few need much longer datasets of 100 years or longer, for example for trend analysis. In the case of climatological information, it seems logical to generate derived climate variables, especially for the period after 1950.

2.6 Uncertainties and variability

The term “uncertainty” and related terms (variability, accuracy, precision, etc.) are interpreted in diverse ways by users and stakeholders (and sometimes even within the project team). Therefore it is important to clearly describe and define what is meant by “uncertainties”, and to describe the various uncertainties clearly in the user guides (see also MS11). Consistency in the use of the various terms related to uncertainties is also important. WP1 has produced D1.3: Uncertainties added to satellite data in CEMS, which will need to be included in all thoughts and approaches to describing uncertainty.

All users and stakeholders recognise that information on uncertainties is important, however many have difficulties using it. Within EUSTACE an ensemble will be produced to describe the uncertainties, but from the user consultation for EUSTACE we knows that potential users are also likely to want to use derived statistics or summary statistics (Bessembinder, 2015). If possible the following products will be generated:

- A “best estimate”;
- Standard deviation (only useful when the ensemble has an (almost) Gaussian distribution);
- Certain percentiles (e.g. the 10^{th} and 90^{th} percentile of the distribution (the exact percentiles can be determined after feedback from users).

Several potential users will use derived products (e.g. climatologies, trends, averaging over time or area) and might want to have uncertainty estimates for these derived products.

As alluded to above, variability is sometimes seen as a form of uncertainty, however often it is also considered separately. For many potential users a good representation of the variability on various scales (on temporal scales from day-to-day to year-to-year, and spatial variability) is useful. This is discussed above in Section 2.1 and 2.3.

2.7 Data format and supply portal

The standard format for the final EUSTACE products (for gridded datasets) is NETCDF which is the format that was described in the proposal (see also MS12 on the recommendations for data formats).

Some users have asked for other formats in addition to this, however, EUSTACE does not have the resources or the scope to develop tools to transform the datasets into other formats. Some tools do already exist which can transform NETCDF files into other data formats (or that need little modification to do so), and that can extract data for certain areas and/or time periods. These options will be included in the user guides. The same is true for requests from users and stakeholders for other processing tools; again this is not within the scope of the EUSTACE project, but we will refer to existing tools within the documentation (perhaps with the exception of scripts for downloading the datasets).

Most important for all potential users is the easy accessibility of the dataset. No separate portal will be constructed, but the datasets will be supplied through the STFC Centre for Environmental Data Archival (CEDA). Whenever possible links will be made with other existing portals or the dataset may also be uploaded onto other existing portals.

2.8 Content of user guides

The report on user requirements (MS30) highlights several aspects that should be treated in the user guides. It is clear that the potential users/stakeholders vary enormously, therefore user guides with different levels of detail and guidance have to be developed. At the least, a distinction has to be made between those users who are used to working with large gridded climate datasets (who may require more technical information about the dataset, but less guidance in using it) and users who do not have this experience. User feedback (from the trail blazer users amongst others) will be important to determine which information has to be included in the user guides.

The Updated Dissemination and Communication Plan (D4.2) contains information on how all products and the project will be effectively disseminated and communicated to users, stakeholders and wider audiences.

Reference

Bessembinder, J., 2015. EUSTACE Report on user requirements: results from first user consultation workshops, questionnaires and literature review. Milestone 30.

Bessembinder, J. (Ed.), 2013. Mapping users' requirements: What do we know and what not? JPI-Climate, Report of FTA 2.1,Version November 18, 2013.

COST-VALUE, 2014. White paper on climate change data for end-users. COST ES1102-VALUE project. http://www.value-cost.eu/sites/default/files/value-wg1-whitepaper_04-03-2014.pdf

Themeßl, M., 2011. Empirical-statistical downscaling and error correction of temperature, precipitation and derived extremes in Europe (http://www.uni-graz.at/igam7www-report_45_mtheme_l_june11.pdf), PhD thesis.

UKCIP, 2006. Expressed preferences for the next package of UK climate change information. Final report on the user consultation, December 2006.

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