EUSTACE

Horizon 2020

H2020-EO-2014 New ideas for Earth-relevant Space Applications

EUSTACE
(Grant Agreement 640171)

EU Surface Temperature for All Corners of Earth
Deliverable D4.9/Milestone MS36

EUSTACE
Report on user requirements: results from second round of user consultations

Deliverable Title	EUSTACE: Report on user requirements: results from second round of user consultations	
Brief Description	The results of the user consultation workshops, inventory and interviews are described in this report, including feedback on mock-ups of the data files. Where possible, recommendations for the design of the EUSTACE products are given.	
WP number	4	
Lead Beneficiary	Janette Bessembinder, KNMI	
Contributors	Nick Rayner, Darren Ghent, Kristine Madsen, Joel Mitchelson	
Creation Date **Version Number** **Version Date**	15/12/16 1 05/03/17	
Deliverable Due Date **Actual Delivery Date**	31/12/16 06/03/17	
Nature of the Deliverable	X	R – Report
		DEM – Demonstrator, Pilot, Prototype
		DEC – Dissemination, Exploitation or Communication
		O – Other
Dissemination Level/ Audience	X	PU – Public
		CO - Confidential, only for members of the consortium, including the Commission services

Version	Date	Modified by	Comments
1	15-12-16		
	05-03-17	Nick Rayner	Minor editorial suggestions

EUSTACE

Table of Contents

Executive Summary

The conclusions from this 2016 user consultation are generally in line with those from 2015, but provide more or more detailed information. Below the most important additional information is presented compared to the report from 2015.

Data sources, time period, area and resolutions

- Relatively more people indicated this year that they would prefer to have long datasets (starting before 1950).
- More information was collected on which data sources are used now (see figure presented in Section 3.1). It is good to compare this list with the list of data sources that are proposed for the intercomparison in WP3 (validation and inter-comparison) to ensure we provide comparisons which are meaningful to users.

Data format

After the user consultation in 2015 it was decided to use NetCDF as the format in which the EUSTACE products will be provided. Since we know that not everyone is very familiar with the format, questions were included to get an idea what kind of support is needed.

- Most people, also those familiar with NetCDF, would like to have support ranging from: example scripts; providing (links to) tools for conversion (esp. ASCII and Excel) and processing/analysing; tools for extraction of data; and detailed examples on how to access the data, how to process and interpret the data (see also comments below under "user guides"). Links to some tools will be provided within this project in the user guide, although it is impossible to be complete. Where available, experiences of users will be added.
- Some tools/packages (or versions thereof) do not work well with NetCDF, some examples are mentioned (ArcMAP, ArcGIS, Quantum GIS, Excel, R-packages, some Python modules). Information on issues/problems with certain tools/packages will be provided in the user guides, as far as information can be obtained. New problems may appear later on. It would be good to think of "locations" where these issues could be reported by users to inform others also when the project is finished[1].
- Depending on the format of the NetCDF files, the data either can or cannot be used easily with standard open source tools. The data file mock-ups presented in July 2016 worked well for most of the potential users who trialled them, although some remarks for improved were given. Mock-up A for presenting local solar time (local solar time ancillary variable) worked best and was preferred by the limited number of users that could give feedback.
- To make the EUSTACE data useful to a wide range of users, it is advised to give air temperature estimates for all surfaces present in grid cells containing more than one type of surface.

User guides

No real new subjects were proposed for the scientific and product user guide compared with our proposed set-up earlier in 2016.

- A quick start guide was requested, since few people read user guides completely before using a data set. It should be short and contain among others file format and content, access information, quick overview (resolution, spatial coverage, period, etc.; can partly be given also in the NetCDF files), known limitations/issues, link to further information, and some information on the applications (e.g. how to interprete, do's and don'ts).

[1] This may be easier if there were to be a follow-on project.

- Probably as important as a quick start guide is a clear structure of the user guide to make it easy to find required information and options to navigate quickly through it.
- Links to existing tools and scripts/readers are useful, especially for software packages regularly used (e.g. for C, Python, R) or provide links, or lines of code for extracting data. NetCDF has problems in some situations. For this it would be helpful to have an idea on the most used software packages. Also tools for visualizing, extracting data and processing are considered useful.
- Clear and consistent use of terms and definitions is very important throughout the user guides (definitions referring to some standard).
- Examples (especially with open source tools) that allow "copy and paste" e.g. to read data are regularly mentioned as useful. Also examples that show "how working with EUSTACE data worked out" were considered interesting.
- Tools for visualizing (e.g. Panoply, ADAGUC, KML quick look for Google earth) are helpful, also for users not familiar with NetCDF, to see whether the data are useful for them. They hide the complexity of NetCDF and usually they can be used on a windows platform.
- Those not familiar with NetCDF require some introduction to it and probably more detailed examples
- Only a limited number of people is familiar with the ESGF portal, from where we plan to make the EUSTACE products available. From this we can conclude that the user guide should certainly contain information/guidance on how to access and download data from here.

Several of the requests/suggestions mentioned can be included in the user guides, however, it is impossible to provide support for all available packages. Therefore it is recommended to focus especially on the most used packages such as Python, R, and Matlab.

Uncertainties

- The term uncertainty is interpreted differently by different people and it was mentioned a few times that information on uncertainties should be easy to understand and easy to use. This indicates that a clear description of the types of uncertainties is required in the user guides, and also examples on how the information provided can be used or interpreted.
- Many indicated that besides a "best estimate", they would desire to have a measure of the spread (standard deviations, percentiles (not indicated which) or maximum and minimum values). Although as one of the respondents indicates "everything can be useful", users do not want to be overloaded with information. It was suggested also to produce quality flags to make it easier to choose "best quality" data. For the selection of which type of uncertainty information will be provided and how, it should be considered that some climate models require certain information in order to ingest the data.
- A few people indicated that they would use more than one ensemble member. These persons all have a lot of experience in the use of climate data.
- Most of the respondents considered information on the separate sources of uncertainties useful or thought that it could be useful, but at the same time they think that most users will only use the overall uncertainty and don't want to be overloaded with information. E.g. in GlobTemp, the total uncertainty is given as default and components as optional extra data set. Many of these respondents have a background in climate science. As a project we will consider whether this information can be made available.
- There is also a need to describe how well the ensemble conveys the uncertainties in the products – is the ensemble a good reflection of the true uncertainty in the product?

Updating EUSTACE data sets?

There is no financing yet for updating the EUSTACE dataset after the project has finished, but when looking for a follow-on it is good to know what frequency of updating is desired by potential users. Once per year and once per month are mentioned most regularly as desired updating frequencies. In case of updates, most people are interested in receiving alerts by e-mail.

EUSTACE

Project Objectives

With this milestone, 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: https://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/Paul van der Linden
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 results of the second user consultation activities in April-July and September-December 2016. This section starts with the general aim of EUSTACE, potential stakeholders and users[2] of the products developed in this project and how we tried to reach them during this user consultation round.

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[3].

1.2 Potential users/audiences

The potential audiences of EUSTACE are in the first place the (potential) users of historical temperature data time series and related tools/methods. Since temperature data are used in many sectors and applications, there are potentially many other stakeholder groups. Besides these stakeholders, many people only use information based on temperature time series (derived data; impact researchers, policy makers, public), that could be based on EUSTACE products (indirect users) (see also Communication and Dissemination Plan, 2016 (update), D4.2).

Aims of user consultation:

- As EUSTACE gets better insight into the potential use of temperature data in various sectors, the existing network of potentially interested people can be broadened;
- Collecting information on user requirements can be used for either increasing the relevance and usability of EUSTACE products (also input to the design of the products) or to be able to have more targeted communication to potential users (give information that better fits their interests, better understand the users and their requirements, improve the two-way interaction).

In the project plan user consultation workshops are mentioned to get direct feedback from stakeholders. However, it appears rather difficult to reach a diverse group of stakeholders with a limited number of workshops (very diverse groups and therefore

[2] 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.

[3] 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.

more difficult to get people to one place[4]). Therefore, the inventory of user requirements results from several sub-activities, as in the first user consultation round in 2015.

Timing of user consultation: throughout the project with largest effort in the first six months (report on user requirements: MS30). However, after these first six months new information on user requirements can become available and the overview can be updated.
Frequencies:

- Literature review: in the first six months of the project;
- User consultation meetings: during three phases of the project. First round took place around month 4, the second round (reported on here) by months 16-20 and the third round by month 30 (June 2017);
- Questionnaire/interviews: in the first six months and later on during the project to get feedback on e.g. the website.

Whenever possible, members from the product development team are invited to the stakeholder workshops to allow direct communication between them and the stakeholders.

Expected result of user consultation: better insight into the requirements of various stakeholder groups, documented in this report. Whenever possible this will give feedback to the project team on aspects relevant for the design of products (e.g. the format of the datasets, the form and content of the user guide).

1.3 Methods used during this user consultation round

Splinter session at the European Geophysical Union (EGU) annual meeting, Vienna, April 2016
This time a longer session of a whole morning was organized, compared to the meeting in 2015, to enable more thorough discussion of results and to get more feedback. As before, the session consisted of informing the audience about the project and its progress. Also a lot of time was reserved for questions from the audience and for discussion with the audience on their requirements.

To reach more people, before the conference all sessions that had some relation with the use of climate data (presentations and posters) were checked. Presenters from relevant subjects were all invited via an e-mail before the conference. Mainly (climate) researchers were reached this way.

Virtual user consultation workshop
For those that could not visit the splinter session at the EGU conference a virtual user consultation workshop was organized. This way also researchers from other disciplines not present at the EGU could be reached together with people not working in research. The advantage of a virtual meeting over a meeting in a physical location is that it takes less time (2 hours in this case), incurs no traveling costs and people can participate from all corners of the Earth. As during the session at the EGU, the virtual session consisted of informing the audience about the project and its progress (using a summary of what was presented at the EGU), and a lot of time was allocated for questions and discussion with the audience. All people on the list for the newsletter were informed about the virtual meeting and also those that did not visit the splinter session at the EGU.

[4] When many multiple day workshops are organized for many projects, it is difficult to get people to participate in such a separate user workshop.

A few people that could not attend the virtual meeting were informed and interviewed afterwards by phone/skype. This gave the opportunity to discuss in more detail their specific needs and situations.

Questionnaire/inventory
An online questionnaire also gives the possibility to get feedback from potential users/stakeholders. The advantage can be that reactions to the same questions from a larger group of people can be collected. This time Survey Monkey was used as a tool. This tool made it easy to include some information on the project and to refer to the website with the presentations of the EGU-meeting and other information. In Appendix 2 the text and questions in the questionnaire are presented. An invitation was sent to the same group that received the newsletter and to the presenters at the EGU that were not present during the workshops. The questionnaire was open from about May 11 until July 10, 2016.

Request for feedback on some mock-ups of EUSTACE data products
Some mock-ups of EUSTACE data products have been produced by WP2 (dataset construction): first round of these consultations was performed in July 2016, and the second round at the end of November 2016. These mock-ups do not yet contain real data, but they were used to check whether or not the proposed format works well with the systems/tools that potential users use at this moment. In July an e-mail was sent out to a selected number of people (the trailblazer users among others) with the request to give feedback on the mock-ups. The mock-ups were made available behind a pass word. Due to summer holidays, the feedback was collected in September. In December a few trailblazer users were asked to give feedback on new mock-ups using different ways to present the local solar time, since this is not commonly done in climate products and no established standards exist.

2 User consultation workshops and interviews

2.1 Information about the meetings and interviews

Splinter session at the EGU, April 19 2016, Vienna

Programme

8:30-8:40 Welcome and aim of the meeting (Janette Bessembinder)
8:40-9:05 Introduction to the project: background and products (Nick Rayner)
9:05-9:50 Data sources and homogenization (Nick Rayner)
9:50-10:00 Uncertainties: types included (Darren Ghent)
10:00-10:30 Coffee-break
10:30-10:50 Uncertainties: questions/discussion
10:50-11:30 Methods: deriving air temperature data and infilling (Kristine Madsen)
11:30-11:50 User guides and tools for using NetCDF files (Janette Bessembinder)
11:50-12:00 Wrap-up and follow-up

Presentations are available at: https://www.eustaceproject.eu/users/getting-informed-and-participating/user-consultation-meetings/user-consultation-meeting-april-2016/.

Participants: people from NASA (MODIS), University of Edinburgh (Geosciences), University of Leuven (Geography and tourism), Instituto Português do Mar e da Atmosfera, China National Centre for Climate change, Institute of Atmospheric Physics (Chinese Academy of Sciences), Fondazione Edmund Mach, India (organization unknown), NOAA, EUSTACE team (Nick Rayner, Darren Ghent, Kristine Madsen, Janette Bessembinder)

Virtual user consultation meeting, May 24 2016

Programme: May 24, 2016, Webex meeting from 14:00-16:00 CEST. After a presentation by Nick Rayner (condensed presentation based on the presentation at the EGU splinter session), questions from the participants were answered and user requirements/wishes were discussed.

Participants: one person from ZAMG working on phenology (also at poster session EGU), 2 people from CCIAM (Faculty of Sciences from the University of Lisbon), EUSTACE team (Nick Rayner, Janette Bessembinder

Interviews

A few interviews were held after this virtual meeting (with J. Bessembinder):

- The first one (June 21) was a consultant who works for TEC-conseil especially in the fields of tourism and transport in relation to climate change.
- The second interview (July 12) was with someone who contacted us through the website and who is working for Arizona State University, especially in the field of health in relation to climate and climate change.

2.2 Summary of the discussions

Below a summary is given of the points discussion during the meetings and interviews. Only sometimes, when relevant, it is indicated during which meeting/interview the comment was made or when a certain question was posed.

Data sources, time period and resolutions

- The participants in the splinter meeting were especially interested in data for the Himalayas, desert regions and China. The ones in the virtual meeting were focussing more on Europe.
- Currently they use among others: CRU3.2.2 and IMD (Indian Met Department; not publicly available), (individual) station data, E-OBS, GISS and reanalysis data, other sources (with temperature for all sky conditions; Daymet data). For China, Chinese station measurements are used for T_{max} and T_{min} and MERRA reanalysis (hourly data). One of the participants mentioned that MERRA uses satellite data to estimate LST and air temperature.
- The time period used depends a lot on the purpose of the use. If people want to look at trends, often datasets going back as far as possible are used. However, for tourism it was mentioned that most often the past 30 years are used.
- The required spatial resolution depends on the application. In general, for applications where the spatial differences on small scales are important, very high spatial resolution are requested, e.g. tourism in mountainous areas, health in urban areas.

Uncertainty information

During the discussions several things related to the use of uncertainty information were mentioned:

- Important to give a clear explanation about the types of uncertainties
- Consider also the way information on uncertainties is communicated!
- Some methodological aspects such as the fact that the emissivity-related uncertainties in the satellite retrievals would likely be filtered out in the regression and that we could do better than use split window retrievals for MODIS, since uncertainty is increased by using these.
- A participant wants to use uncertainty as an extra source of information to assess the significance of trends. When validating LST also uncertainty of in situ information will be used.
- Some climate models require information on uncertainties in order to ingest the data.
- Provision of the components of the total uncertainty allows users to recompute it, or recombine it in a better way for their application.
- In GlobTemp, the total uncertainty is given as default and components as optional extra data set. This seems a good approach, because users shouldn't be overwhelmed with information that they will not use.
- It is important to communicate that the air temperature is dependent on the skin temperature and that there are co-varying uncertainties.
- Participants differed in the format of the uncertainties that they would use: some mention "best estimate plus estimate of the spread" (or relatively simple representation), others may want to select ensemble members (depending on how the members are created). Many use information on uncertainties especially to check the quality of the data
- There is also a need to describe how well the ensemble conveys the uncertainties in the products – is the ensemble a good reflection of the true uncertainty in the product?
- It was suggested to produce quality flags, e.g flagging different ranges of uncertainties as well as uncertainties. In that case a user can either choose "best quality" data or to use the uncertainties directly for different applications.

Relationship building and analysis methods

During the discussions some of the methodological aspects were discussed: different components of the analysis (for the different surfaces), can they be provided separately (could be important for coastal research) and the climatology fraction (should be explained better), how to process overpassing time of the satellite, use of daytime and

night time data separately or together for the relationship building, and on the impact of clear sky bias in the satellite data.

User guides

During the discussion in the splinter meeting on the user guides it was suggested to make a "Quick Start Guide". If a user guide is very long/has many pages it is often not read or only partly. A quick start guide could give the most important information and if needed, more detailed information can be found in the more elaborate user guide. When asking what the quick start guide should contain the following things were mentioned: basic information on what is available and in what format (e.g. resolution, detailed information can be provided in the NetCDF files), where to find the data and how to get access to the data, and a few pages on the applications (e.g. how to interpret, do's and don'ts).

As an example of a "bad" user guide, the one for EOS data products was given, especially concerning definitions. It was recommended to have a set of standard definitions referencing to some standard, e.g. now view angle is defined differently for different sensors on the same platform; now the same name used for different things. It is important to use a standard nomenclature.

The participants in the splinter meeting were all familiar with NetCDF and also several in the other meetings. Those familiar with NetCDF did not need explanation about this format, however, they could give some examples/suggestions related to this format:

- One participant mentioned that he had seen two kinds of NetCDF: with lats and longs as sub data sets, with actual data in sub data sets. This was not easy to use and it was difficult to use standard open source tools.
- Some others suggested to make readers available for the NetCDF files, or provide links, or lines of code for extracting. If a lot of project specific metadata is added, specific readers may be needed to cope with this (and if we introduce new metadata standards – need to try to get this into next CF standards).
- Useful to provide scripts or links to scripts for opening the file in different languages, e.g. C, python, etc. NetCDF has problems in some cases. For this it would be helpful to have an idea on the range of software packages used, so we can make sure the files are useable in a range of packages used regularly.
- Examples for various open source tools can be very useful (describing and when possible showing images on how to read the data).
- Important to have all information in one file, or give clear instructions on how to combine them.
- Some tools were presented to quickly visualize data or to subtract data for specific regions. Besides the ones shown (Panoply and ADAGUC) also KML quick look for Google earth was suggested to see quickly the data for a country/region of interest. The tools are considered useful, also for users not familiar working with NetCDF. The tools are a way of visualising the data and to see whether they are of use to them. They hide the complexity of NetCDF. Usually they can be used on a windows platform.
- There were also some questions about the tools: Can they show spatial averages, as well as just single points? Can the plot data be downloaded as a csv rather than downloading the images?
- It was also suggested to provide examples in such a way that it allows users to copy and paste and use this to read the data – this is the most practical way to learn to use a new data set.
- Examples that show "how working with EUSTACE data worked out" were considered interesting.
- Some indicated that they would be interested to give feedback on e.g. mock-ups of the datasets. These were sent information later on and their feedback is included in Section 4.

Additional information

At the EGU also a session was organized by EUSTACE partners with some others on: Taking the temperature of Earth: Variability, trends and applications of observed surface temperature data across all domains of Earth's surface (oral presentations: April 18, 13:30–15:00; posters: April 18, 17:30–19:00). Several contributions from EUSTACE were included (oral presentation and some posters). During the poster session we discussed with several people how they used temperature data and whether EUSTACE could be useful for them, e.g. for pollen forecasting, growth of trees. Those people were added to the list for the newsletter or other mailings.

3 Questionnaire

Thirty two persons filled in the questionnaire (a few only partly). Some of the questions were repeated from the first user consultation round in 2015 to see whether we have a different composition of potential stakeholders now and whether that would give us a different picture.

The answers are presented below. The number of respondents this year was slightly higher than in 2015, however it is still limited, therefore the results may give a biased picture.

3.1 General information

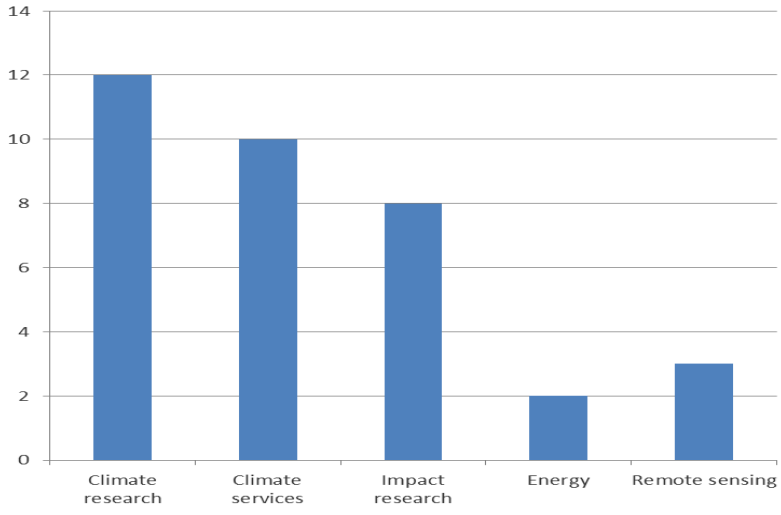
The results of the general questions (also included in the inventory of 2015) do not give a clearly different picture of user requirements:

- Spatial resolution as high as possible
- Daily temporal resolution in most cases sufficient, certainly if minimum and maximum temperatures are provided
- Many respondents also use other climate variables such as precipitation, wind, radiation, humidity

They do give some more detail on certain points:

- Which data sets and sources are used currently

In which sector are you working?
Most of the respondents are working in climate research and climate services (sometimes both). The respondents working on impact research were working on agriculture, ecosystems, hydrology or combinations of these. In the inventory of 2015 these groups were also most represented.

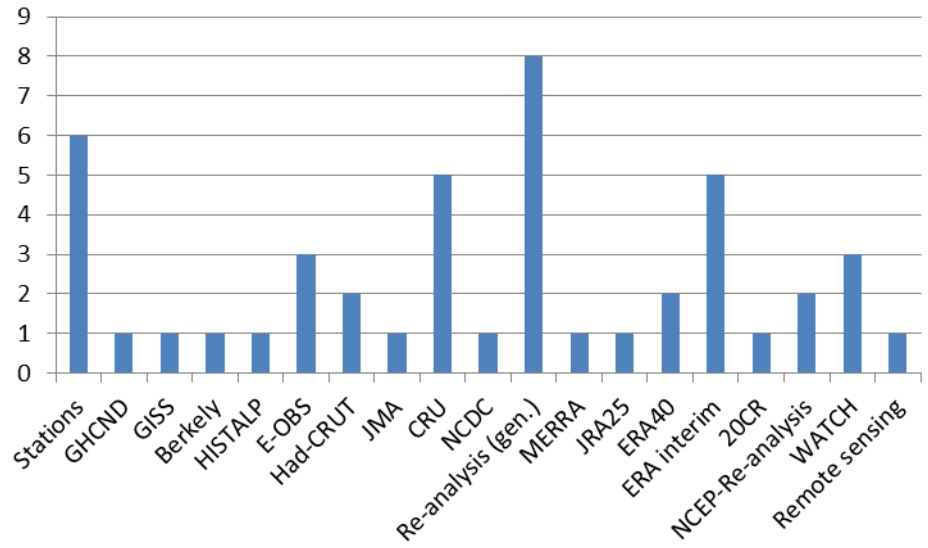


(32 respondents)

For which purposes, if any, do you use temperature data currently?
The answers to this question gave some more specification on what the respondents use the temperature information for (see Appendix 2). This ranges for climate research from model validation, skill assessment, bias correction, statistical downscaling, algorithm development/ improvement to showing global warming. Some examples for impact assessments and climate services are: modelling the influence of temperature on electricity demand; for calculating heatwave (and also cold spell) indices; for forcing and

validating a vegetation model; and to evaluate and forecast climatic drivers of grapevine culture, production and quality.

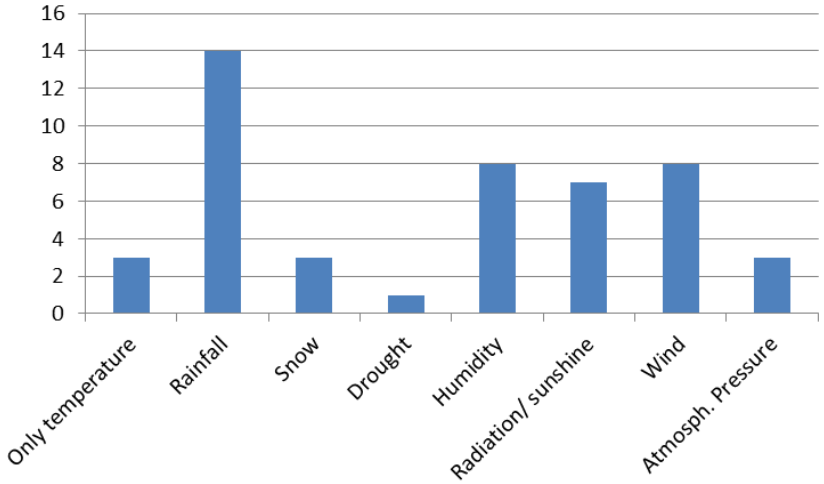
Which data set(s) do you use now for these purposes?
The figure below summarizes which datasets are used now by the respondents. As can be seen station data, gridded data, re-analysis and remote sensing data are used. This inventory gives some more information, compared to the inventory in 2015, on which datasets/sources are used, however again station data, E-OBS, Had-CRUT, CRU and re-analysis data are mentioned most often.



(26 respondents)

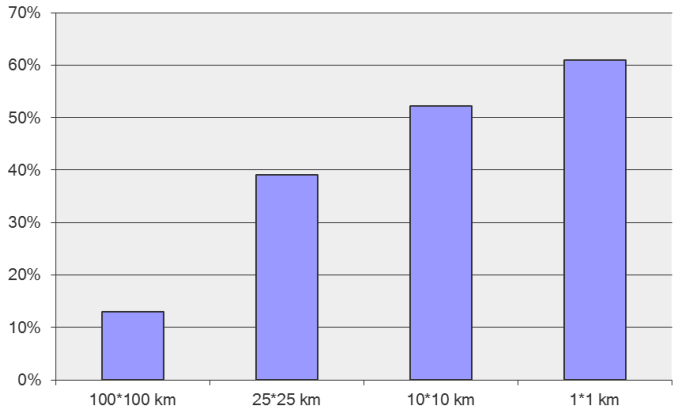
Note: not always the exact data source was mentioned, but e.g. only that re-analysis was used, station data or that datasets from CRU or NCDC were used.

Do you also need data on other climate variables besides (average, minimum, maximum) temperature? If so, please specify.
As in 2015, most respondents indicate that they also use other climate variables. Most often mentioned are precipitation/rainfall, humidity, radiation/sunshine and wind, as was also the case in 2015.



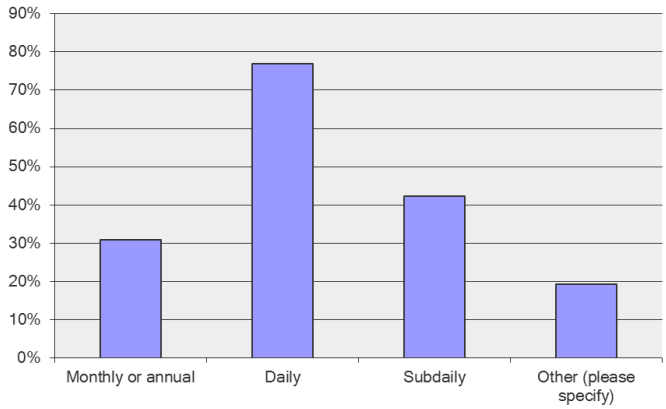
(24 respondents)

Which spatial resolution do you need?
The respondents often prefer a spatial resolution that is as high as possible (as in 2015). Although several indicated that they prefer a resolution of 1 by 1 km, the same persons also indicated that they could use a resolution of 10 by 10 km, etc. This is also indicated by several of the 12 persons that gave additional information. For global applications a coarser resolution is more accepted than for regional applications. For certain regions with high spatial heterogeneity also high spatial resolutions are preferred. Some request country/region-wide data (not clear whether this means averages for country/region).



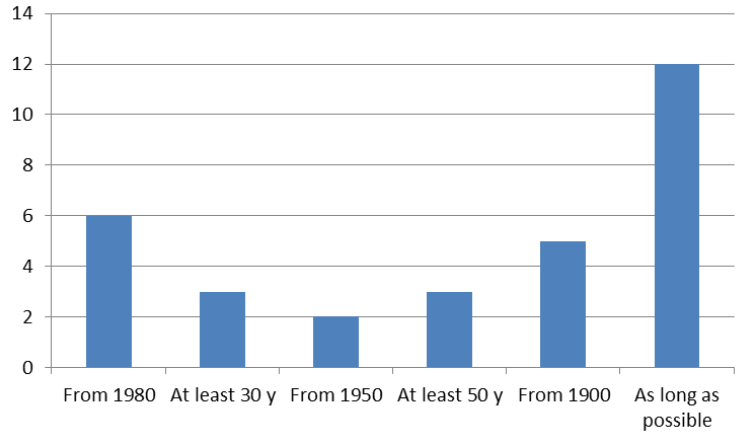
(23 respondents)

Which temporal resolution do you need?
A daily resolution is sufficient for most respondents, and in the case sub daily data are requested, minimum and maximum temperature is regularly sufficient (as also indicated in 2015).



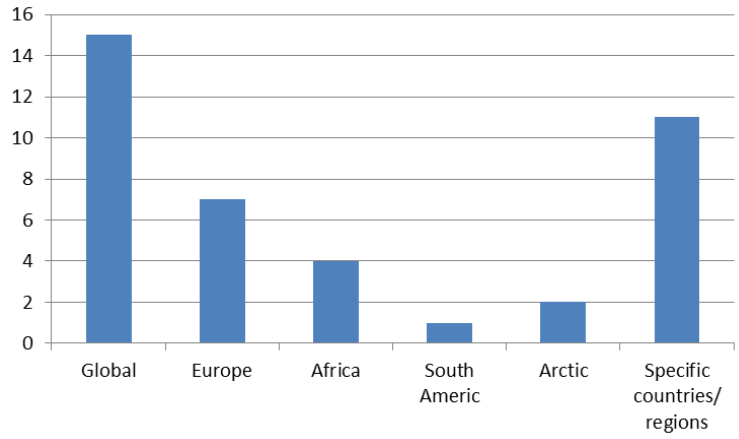
(26 respondents)

For which period in the past do you need data (length of the data set)?
In this group of respondents more often data sets of 100 years or starting around 1900 are requested than compared to the respondents in 2015. As can be seen several prefer datasets that are as long as possible.



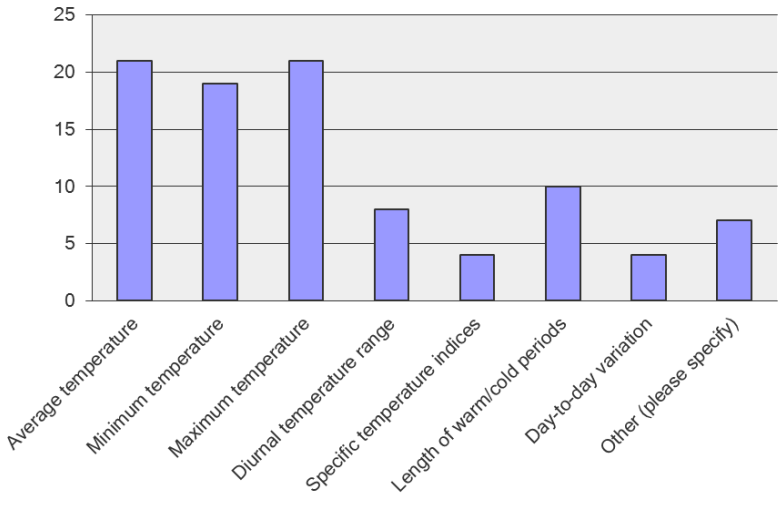
(26 respondents)

For which continent/region/country do you need data (spatial coverage)?
As in 2015 the area for which data are requested varies considerably per respondent. In this group of respondents most are working at the global level and Europe, but also a considerable number is working on specific regions or countries.



(26 respondents)

Do you need information about: time series, indices, etc.?
Most of the respondents indicate that they would like to have minimum, maximum and average daily temperatures. Some indicate that they would like to have temperature indices, although some of the respondents mention that they can calculate the indices themselves from the time series. This was also indicated in 2015. In this inventory there are fewer requests for day-to-day variation. One of the respondents explicitly asked for trends and anomalies, another asked for joint CCl/WCRP/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) indices.

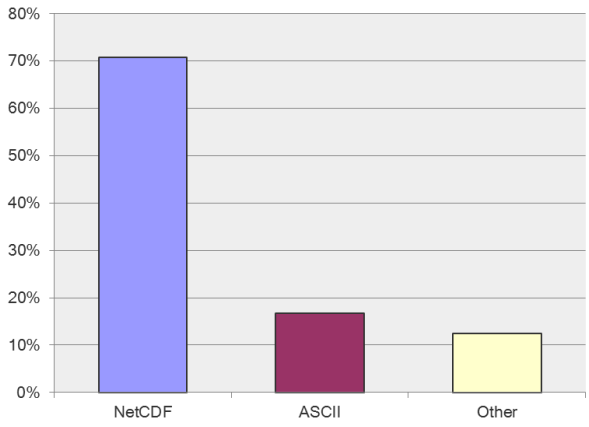


(25 respondents)

3.2 Data format

In 2015 it was decided to provide the EUSTACE datasets in NetCDF format. However, not all people are familiar with this format, especially not people from outside climate science. Besides that, there are many different tools/packages that may not work with this format or that require specific things. With this inventory already some more detailed information was collected. Also some mock-ups were prepared and sent out to a limited number of people (8-10), to collect feedback on the proposed data file format (see also Section 4).

Which data format do you use currently? E.g. NetCDF, ASCII.
In this group of respondents, many are familiar with NetCDF, as could be expected since many work in climate research. As other formats GRIB and xls (and txt, csv) are mentioned. Outside climate research NetCDF is often not known. Some of those using "other" formats do have some experience working with NetCDF files.



(24 respondents)

Do you need support for using the NetCDF format? If yes, please specify what kind of support?
About one third of the respondents (6 out of 19) to this question indicate that they would like to have some support. The following types of support are mentioned:

- ncd-coding
- Example of e.g R scripts to read and process information

- Useful to provide basic guidance on NetCDF for the different users - including reference to guide the use of the different tools
- Conversion to Excel-readable files and ESRI-compatible point shapefiles
- (in the python developer world modules are good enough to make NetCDF easy to use)

Two respondents indicate that they need support, but they do not know what exactly, since they do not know the format.

Several tools exist for visualizing NetCDF data and for extracting data for certain periods and regions (e.g. Panoply, ADAGUC). Is referring to these existing tools and their user guides sufficient?

More than 50% of the respondents indicates that it is useful to refer to existing tools, only 1 person indicates it is not useful (out of 25). Several remarks are made on other tools and tools for conversion to other formats or for statistical analysis:

- Besides Panoply for initial data exploration, I use MatLab or R for analysis/visualisation
- I use IDL and GDAL and this works fine.
- Panoply is useful for visualisation; extraction of data is really limited and not possible as a batch for a number of locations or files. Not familiar with ADAGUC.
- Ideally, a direct converter for Excel should be provided that would produce as output data for a certain region and timeframe. Same for ESRI-compliant shapefiles.
- Necessary, but no sufficient
- Using data is far more than just visualising them. Especially for climate services making statistical analysis is essential. So also reference to R or Python tools would be good
- Conditional yes, if an example on the data is included.

Are there tools/packages you work with, that do not work well with NetCDF? If so, please specify.

A few persons gave examples:

- ArcMAP and ENVI cannot read it. This is a problem for many students.
- ArcGIS
- Excel, QuantumGIS
- R-packages
- (python modules are not always up to date to each other and maintaining modules should get more resources and attention to ease all our work)

In the former question R was mentioned as a good tool for analysis, but here it is mentioned that some problems were experienced.

Other remarks or suggestions related to the use of NetCDF?

- What might help is to let users know that software packages such as "R", "IDL", "Python", and "Matlab", which are all quite common programming languages, support NetCDF files - partly on a different level of maturity. Also, Panoply is fine for a quick glimpse but not helpful if one aims to carry out an analysis.
- Ensure accompanying metadata is of a good standard wherever possible and only missing if absolutely necessary.
- I have tried to work with NetCDF but the format needs specific training for people not versed in these gridded formats. It is unthinkable to foresee even technical people who are not trained in climate disciplines to work with currently available tools.
- Providing tools for conversion to other formats (e.g. GDAL capabilities for GeoTIFF, etc.) would be useful to many users I work with beyond the academic research community
- Format is less the issue than the tools to utilize the data!

3.3 User guides

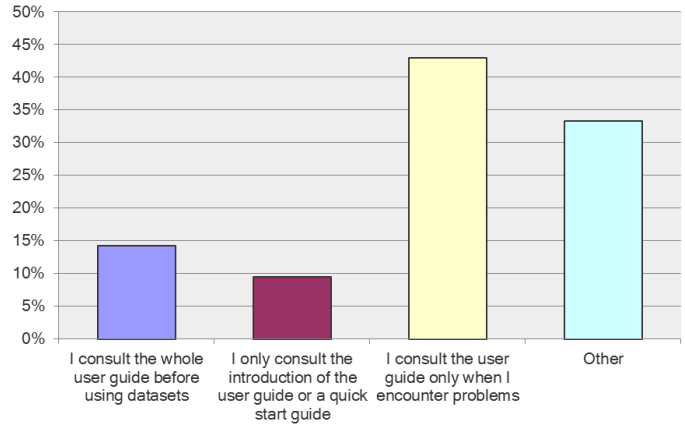
This year a set-up for the user guide(s) was made. The proposed set-up was presented at the user consultation meetings. One of the suggestions was to include also a quick start guide. To get some feedback on this suggestion, and what is desired for the other parts of the user guide, some questions related to this were included in the questionnaire.

Based on the information from these questions we can at least conclude the following:

- Most people don't read a user guide completely before using a dataset
- It should be easy to find the information that is required or that people are interested in. Therefore a good table of content is very important.
- Some important information can also be given in the NetCDF files
- A quick start guide should be short and contain among others file format and content, access information, quick overview (resolution, spatial coverage, period, etc.), known limitations/issues, link to further information
- No real new things were proposed for the scientific and product user guide compared with our proposed set-up. Some respondents didn't find the distinction between scientific and product use guide very clear.

How do you use user guides at this moment?

To write a useful user guide it is good to know how people use it. Below some information is presented on that. Hardly anyone reads the user guide completely before using a data set. Often this is also not needed. Those familiar with the format will probably only consult the user guide when they encounter problems.



(21 respondents)

Remarks on the use of user guides:

- This depends on the complexity of the data set. Some data sets contain a lot of useful information already in the NetCDF files themselves and if they are simple this might be sufficient - together with a key publication where the essential elements of data processing and evaluation are described. For other data sets it won't be sufficient to read the product user guide but reading the ATBD and several other documents are a must to understand what's in the files.
- Typically use a quick start guide to obtain the data and start working with it. Refer to the full documentation when necessary, e.g. when analysing or writing up results.
- I will have a brief look through at the start, concentrating on specific areas of interest.
- I use the table of contents to go the topic I need information about.

- For datasets that I know well, I consult the user guide only when I encounter problems. For new datasets, I consult the quick start guide, if available, and follow up detailed explanations/documentation as needed.
- First " I only consult the introduction of the user guide or a quick start guide" then "I consult the user guide only when I encounter problems"

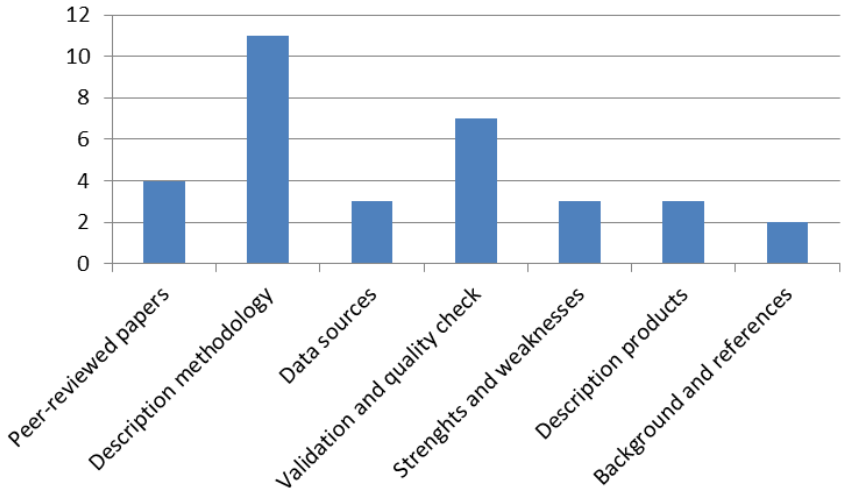
What should be included in the quick start guide (at minimum)?

(19 respondents) The following suggestions were made for the quick start guide:

- I would not vote for a quick start guide. This bears the danger that the user will not look at the Product user guide. If EUSTACE aims for maximum CORRECT usage and INTERPRETATION of the data sets to be created then a scientific user guide (kind of an ATBD) and a product user guide should do it. Also, NetCDF files have the potential to contain a lot of useful information about the data and its processing (background) already. If one lists key documents already in the NetCDF files then there is no need for a quick start guide.
- Metadata structure of dataset dimensions and units
- Caveats on the use of the dataset, description on the structure of data files
- Resolution, how to identify missing values, enough to read in NetCDF, period of coverage
- Key points, limitations to be aware of, etc.
- Brief overview, how to use the data and where to find it, links to further information if needed.
- format of files, period quick information on data processing
- Description of: 1. files content; 2. known issues users should be aware of when using the data.
- Example on how to get the data into selected tools
- Short data presentation : datasets and methodologies used, length, main issues
- file name, content, how to open and read it and/or extract desired geographical area
- Access information, technical information (data formats), some quick-look option, basic information about the data (periods, coverage, parameters, resolution, ...)
- How to choose spatial and temporal coverage, how to download desired observations and indices, how to convert files, how to save previous options or create a regular user profile. Also a FAQ and glossary (this last item is of essence).
- General overview of the data and metadata - spatial and temporal extent, coverage, resolution -- as well as some basic information about the inputs used and the methods employed to produce the data, and information on how to find and download the data.
- How to setup from scratch
- Jargon-busting intro & user-friendly overview
- Best guides are empty, so if the information on the download page is enough - no guides needed ;)

What subjects or documents should be included in the scientific user guide (at minimum)?

In the inventory a description is given of the proposed content. No real new things for the scientific user guide were proposed.



(18 respondents)

Some additional remarks:

- An introduction stating clearly the areas of application and the limitations of the data set.
- Discussion on strengths and weaknesses.
- Methodologies used to derive the uncertainties of the data set offered and information about the different kinds of uncertainties included (physical, retrieval, smearing, etc.)
- Infilling/interpolation, gridding method, underlying data, long-term homogeneity and change in data over time
- Deep description of algorithms and methods. Assessment/ validation of the dataset.
- Results from own evaluation (with clear critical statements about where and how the evaluation data were obtained from or retrieved and about their quality)
- Results from external evaluation (if existing)
- A summary (at the end) with a bullet list of recommendations / limitations involved when using the data set.
- Comparisons with other sources (re-analysis, station time series, ...)
- Processing workflow, validation and quality checking, estimating uncertainty. Scientific overview

What subjects or documents should be included in the product user guide (at minimum)?

(17 respondents) In the inventory a description is given of the proposed content. No real new things for the product user guide were proposed. Some respondents didn't find the distinction between scientific and product use guide very clear:

- An introduction stating clearly the areas of application and the limitations of the data set. Information about the data format and content of the data layers in the data product. Key references leading to the production and evaluation.
- Overview of products, including differences with other datasets, where possible include explanations about the consequences for e.g. accuracy/uncertainty (e.g. in data sparse regions); assumptions and do's and don'ts (interpretation),
- How to obtain the data
- Brief description of the product and methods; know issues and summary of product assessment; practical examples or applications.
- More detailed examples on the potential use of the data, and its limitations.
- The resolution, the grid, the covered period and time step, how to get the land/sea mask
- I'm not sure, if a separation between "scientific" and "product" is useful.

- How to find/get/download EUSTACE products, background information for less experienced users (including links to useful processing tools, etc.), case studies, product details: metadata, including links to the scientific user guide, data structures and formats.
- Overview of products, comparison with other datasets, user implications of accuracy/uncertainty, links to the scientific user guide, FAQs
- Background information
- To introduce new users it is always helpful to make example scripts for postproduction of data to visualise or to regrid or to statistically analyse them.
- Descriptions and background

Can you give examples of good or bad user guides? Please provide link or title.
Besides information on what user guides should contain, examples of what are considered good or bad user guides often are very useful.

Remarks on what are good/bad user guides:

- They need to be short so they are actually used, with reference to a publication for people who want to know more
- The datasets usually lack "User guides", but often have good scientific documentation.

Bad example:

- http://www.ecmwf.int/en/forecasts/documentation-and-support/long-range/seasonal-forecast-documentation/eurosip-user-guide/eurosip-graphical

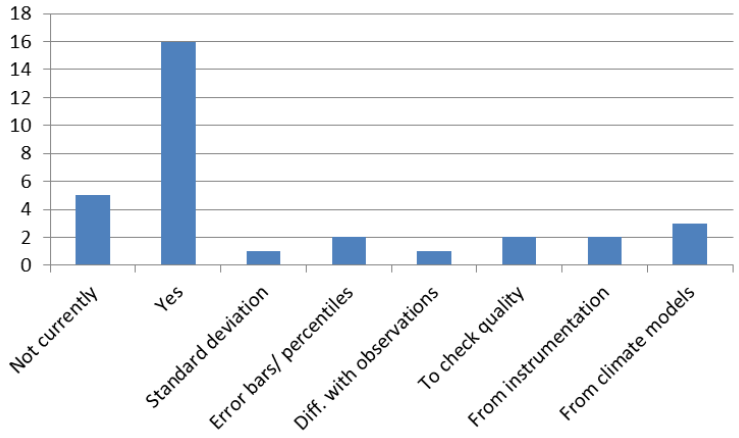
Good examples:

- CHIRPS data from UCSB: http://chg.geog.ucsb.edu/data/chirps/
- https://iridl.ldeo.columbia.edu/maproom/Global/Forecasts/Flexible_Forecasts/temperature.html?bbox=bb%3A-20%3A35%3A40%3A75%3Abb#tabs-1
- it depends on the data set. I'd tend to say that most of the ESA-CCI ECV data set product guides are not too bad.

3.4 Information on uncertainties

In the first user consultation round in 2015 some questions about uncertainty were included. This time we tried to get some more information on the use of ensemble members and which other information on uncertainties would be desired.

Do you also use information about uncertainties in climate variables? If yes, please specify what uncertainty information you use and how.
Most respondents use in some way or another uncertainty information. Those that do not, often indicate that it would be useful/interesting to use it in the future (very similar to 2015). What kind of uncertainty information is used and how differs greatly among the respondents as can be seen below.



(25 respondents)

The term uncertainty is interpreted differently:

- as a measure of likeliness/probability (in observations and climate model data)
- as a measure for quality e.g. for use in error/uncertainty propagation ("uncertainty information related to the quality and reliability of the input data would be really useful as the quality and accuracy of gridded datasets depend a lot on the input data and their characteristics of spatial and temporal coverage")

Two respondents indicated here already that they are interested in the total uncertainty, but also in individual components:

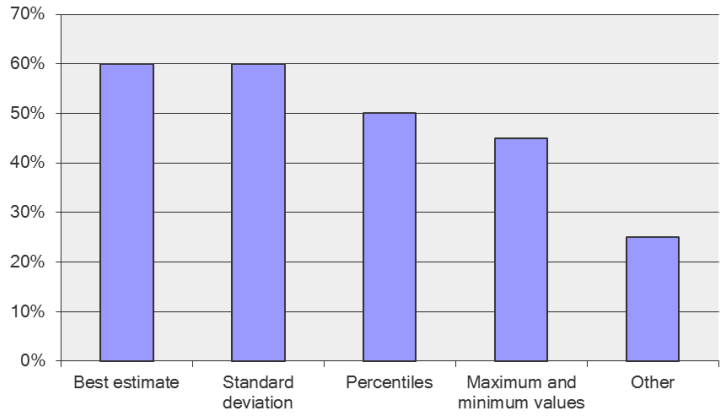
- Currently I use various different uncertainty components and total uncertainty
- Error bars, giving REALISTIC information about TOTAL data uncertainty (not only some error specification of some technical instrument, but an estimate about the total error including measurement error, interpolation error, and everything that adds to the total error).

A few remarks on the "format" of the uncertainty information are given too:

- A data set which comes with easy-to-understand, easy-to-use transparently derived uncertainty estimates will be given preference over a data set with complicated uncertainty information, uncertainty information in cryptic difficult-to-use formats, or no uncertainty information.
- We use the uncertainties in HadCRUT and find them very useful.

If we summarise uncertainty, what is useful?

Some possibilities to summarize the uncertainty information were given. As can be seen below, many indicated that besides a best estimate they would desire to have standard deviations, percentiles (not indicated which) and maximum and minimum values. No clear conclusion can be drawn on what should be provided, but as one of the respondents indicates "everything can be useful".



(20 respondents)

Some remarks on summarizing uncertainties:

- The samples provided with HadCRUT are the only kind of thing that works for us, as the time-space structure of uncertainty and its correlation is important for us
- Not familiar in sufficient detail with how uncertainty estimates are usefully summarised for users to comment on this.
- Everything can be useful :-)

Would you use one or more members of the ensemble? If yes, please specify which and how.

(17 respondents) Less than half of the people (7) that answered this question indicated that they would use more than 1 ensemble member. These persons all have a lot of experience with climate data. Also several persons working in climate research indicated that they do not know whether they will use more than one ensemble member (even though they acknowledge the importance of information on uncertainties).

Some of the remarks on the use of ensemble members from those that will use them:

- Some extreme events are not captured in observations directly, but using extremer values can help to explain for example icy rain disasters, etc.
- Yes, to calculate spread
- All the members to have a better estimate of the uncertainty
- Probably yes, to calibrate probabilistic models and to sample the uncertainty
- Using a single member of an ensemble is generally not appropriate!
- Simulation on standard deviation to assess uncertainties on model uncertainties related to input uncertainties
- Ensemble mean and spread: 1. to assess the robustness of results obtained with nominal product (if such ""nominal"" dataset is to be provided); 2. ensemble standard dev can also be used as a measure of uncertainty.

Some remarks of those that may use but are not sure:

- Depends on the use I'm focusing on.
- Probably only in rare cases
- Not until I understand their differences towards my goals
- I am a bit hesitant to use / offer ensembles from Earth observation data. Either the data are of sufficient quality that they can be used properly and also allow to derive a proper uncertainty estimate. Or they are not and in this case one might either not extend the time series beyond limits and simply allow gaps (modellers can handle these) and/or allow apparently unphysical large uncertainty values.
- Would depend on what was available

Is information on the separate sources of uncertainty useful for you? If so, please indicate how you could use this information?
(18 respondents) Uncertainties from different sources will be determined during the EUSTACE project. In principle this information can also be made available for potential users. To check whether this is considered useful information this question was include. Most of the respondents considered this information useful or thought that it could be useful. Many of these respondents have a background in climate science.

Some remarks on information on the separate sources of uncertainties:

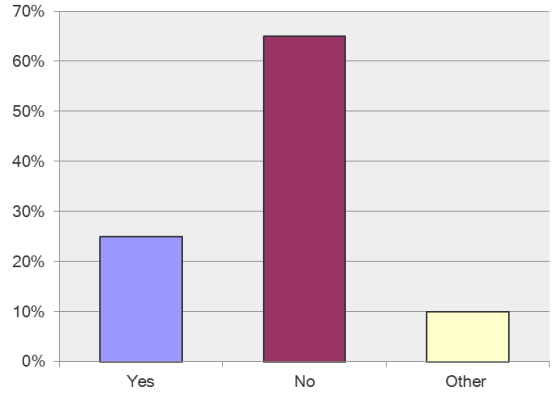
- I believe that splitting the uncertainties into different contributions is essential because these may vary between surface types, measurement type, and retrieval method.
- It would be useful to evaluate the uncertainty of the whole information flow for a specific application
- Yes as the time-space covariance changes with uncertainty source
- To understand the sources of uncertainty.
- It might be useful to examine outliers and anomalies.
- As indicator of main sources of uncertainty per estimate.
- To select product values according to the uncertainty associated to a specific factor.
- Systematic errors can be used to correct the data, random errors for error propagation
- Yes, depending on the category, uncertainty will be used at different scale or for different type of use
- It can be used either as information or for sensitivity experiments
- Useful! In particular the separation of systematic and random is prerequisite for many further statistical analyses or applications. Systematic errors may be mitigated by bias correction methods, random errors are accessible to data assimilation
- Indeed. Would use this info to choose the relevant dataset to work with in case same data are available from a different source.
- Yes, comparison to ensemble output

Other remarks or suggestions related to uncertainties

- Uncertainties should be easy to understand and easy to use. Otherwise they are not used.
- User guide has to be very explicit what uncertainties are included and which ones aren't.

3.5 Earth System Grid Federation (ESGF) portal

The products will be made available through ESGF catalogue portals. Are you familiar with obtaining data sets through these portals?
From the graph below it is clear that only a limited number of people is familiar with the ESGF portal, even though a considerable number of people in this inventory is familiar with NetCDF. From this we can conclude that the user guide should certainly contain information/guidance on how to get data from the ESGF portal.



(20 respondents)

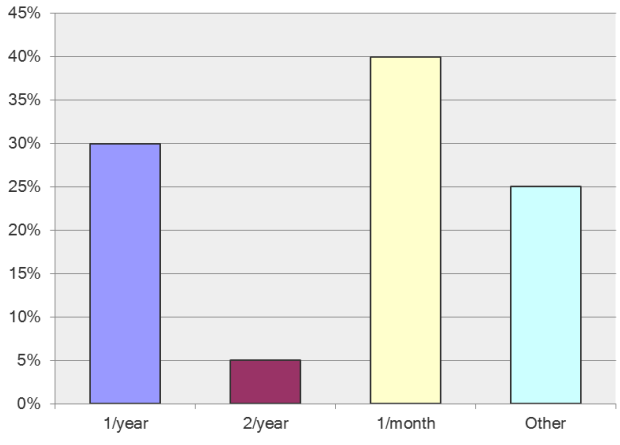
Remarks on ESGF portal:

- Not familiar myself but enough expertise in-house
- I already used it, but find it not very convenient (generation of scripts to run on the machine with a certificate)

3.6 The period after the EUSTACE project

Data sets such as developed by EUSTACE remain valuable when they are updated regularly. The EUSTACE project itself does not include regular updating or prolonging the dataset, however, the aim is to develop a method that can be sustained after the project. To get an idea about the preferred updating frequency, a question about his was included in this questionnaire.

How often would you like the air temperature dataset to be updated?
As can be seen below, once per year and once per month are mentioned most regularly (and of course when improvements in the algorithms are made an update is desired). However, the question is perhaps somewhat ambiguous and may have been interpreted differently by different people; there is evidence that some have answered it as we intended, but that others have interpreted it to mean "how often should new versions of the data be made available?".



(20 respondents)

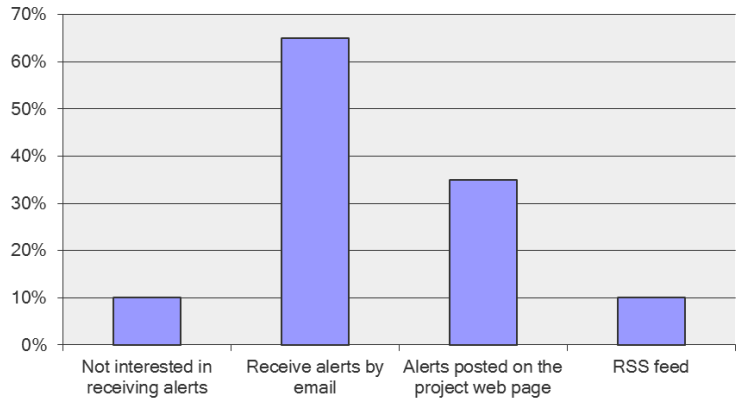
Remarks on frequency of update/prolonging:

- At scientifically reasonable and production-wise realistic intervals.
- Once per month would be great but fewer is still helpful.

- Updating the historical air temperature dataset would be relevant as soon as an improved algorithm is tested, evaluated, etc. but if this refers to product latency, i.e. how soon after observations are collected the data is available in near-real time, then once per month would be required.
- When necessary

Would you wish to receive alerts about problems/new releases of the EUSTACE datasets and by what means?

Another question relating to the period after the EUSTACE project referred to alerts. Most people are interested in receiving alerts by e-mail. This is probably true especially for those requesting updates once per year.



(20 respondents)

4 Feedback on mock-up data files

Eleven people were contacted around July-September 2016 to give feedback on the mock-ups of the data files. Eight persons gave feedback. Appendix 3 and 4 show which information was sent to them. In Appendix 5 the feedback from individual users is presented. Here a summary of the feedback is given. In the internal EUSTACE Milestone reports MS12[5] and MS15[6] and the document on the systems technical requirements[7], information can be found about the design of the data products and on how (indirectly) information from this user consultation is used (several times reference is made to input from users).

4.1 Mock-ups of the data files in NetCDF (July-Sept 2016)

All potential users contacted for this feedback round were familiar with NetCDF, also those that were not working primarily as climate scientists. The packages used to open and sometimes process the mock-up data files were:

- CDO (3x)
- ODV
- Ferret
- Grads
- Panoply
- ADAGUC
- Ncdump (2x)
- Ncview (2X)
- R
- NCL

None had problems opening or reading the files made available in July 2016. A few persons mentioned some problems when processing the data. Further remarks concerning the data file mock-ups were:

- A few persons mentioned that it would be more convenient to have more time steps than just one day per file. One person indicated that many CORDEX files have 4 years of data per file, but shorter periods of e.g. a year could also be more convenient than files with information for just one day.
- The kind of standard deviation should be explicit: "sample standard deviation" or "population standard deviation"?
- If both variables tasmax and tasmin were in the same netcdf file, it would simplify the processing for the end user (fewer files to download/manage)
- The names of variables could be more compact. Instead of "tasstandarddeviation" it would be more simple to use "tas_stddev", for instance.
- Having files in a similar format to current reanalyses/climate model output would be ideal when comparing the data with reanalysis/climate model data.
- Some software appreciates an attribute "axis" on longitude ("X") and latitude ("Y").
- Why are the units duplicated in long_name? This is not adjusted by software that operates on the data and hence will be misleading after processing, e.g. by taking monthly means. conversion from K to Celsius,

[5] MS 12: Recommendations on data formats and management. Joel R. Mitchelson, John Kennedy, Colin Morice, Finn Lindgren, Ag Stephens, Alison Waterfall
[6] MS15: Product design for satstace and fullstace system outputs. Joel R. Mitchelson
[7] WP2 EUSTACE System Technical Requirements. Joel R. Mitchelson

- One person indicates that he appreciates a contact e-mail address and/or web link in the global metadata.
- A "cell_methods:mean" in the time axis would be appreciated for daily means.

4.2 Mock-ups on consistent air temperature fields from satellite data (SATSTACE) (Sept 2016)

In September 2016 the information in Appendix 4 was sent to the same potential users contacted in July 2016. Information on air temperatures above the various types of surfaces (land, ice and ocean) in the same grid cell could either be presented in separate data files or in the same file. For grid cells with more than one surface type, air temperature estimates could be presented in different ways. We asked the potential users for their preferences and possible advantages and disadvantages for different types of users. The reactions were very diverse on the preference of the various methods, therefore it was good to know why people had a certain preference. The following conclusions can be drawn:

- Most people will only use the land data, therefore a good representation of the information available over land is important to have in the data files for all grid cells that have at least a small percentage of land.
- Since there are various ways of dealing with the data in grid cells with more than one surface type, it is probably best not to make assumptions on how to estimate the average air temperature for grid cells with more than one surface type. To make the dataset relevant for as large a group of users as possible it would be best to present the estimates for all surface types in a grid cell separately.
- The data for the various surface types can be presented in different data files or in one data file with separate "variables" (for the various surface types). Both have their advantages (easier to visualize and process in certain situations) and disadvantages (larger data sets). Based on this inventory, it is not possible to give advice on what is the best option.

- Daily mean air temperature will be estimated over all surfaces and daily minimum and maximum temperature only over land. One person indicated that it would be easier for processing to have the minimum and maximum temperature in one file. Since several users will use mean temperature or minimum and maximum temperature, it might be a good idea to have separate files for mean air temperature and other files in which minimum and maximum temperature are presented.

- Most people will only use total uncertainty (if they are able to use it), although several recognize that it would be good to have the information on the separate sources of uncertainty available for expert users.
- It would be good to have the air temperature estimates and the related (total) uncertainty in the same file (as separate variables). This might promote the use of it (really use it in analysis, or easier to consult it to get a general idea of the "quality" of the data)

- People do not want to be overloaded with data. In the case they want to use the total uncertainty, they do not want to have to download the data on the separate sources of uncertainties too. Therefore, a few suggested to have files with only the total uncertainty for most users ("common case", "aggregated data", "operational dataset") and separate files with more detailed information for expert users (if possible).
- In case people only use data over land, they do not want to have to download also the data over the oceans and ice. However, for visualization and processing

of data it may be more useful to have the data for land, oceans and ice in one file (as separate variables).

4.3 Mock-ups for representation of local solar time

By the end of November 2016 some new mock-ups were prepared with two different ways for presenting the local solar time (see Appendix 3). Only about five persons were contacted to get feedback on these mock-ups, due to the limited time available to collect feedback.

All three climate researchers that reacted within a week did not have problems with Mock-up A (among others used with Grads, Ferret, Ncview), but they all had problems with mock-up B. Based on this very short feedback round, it can be concluded that Mock-up A is probably the best one to use.

5 Discussion and recommendations for design

In Sections 2 and 3, detailed information is given on the results of the 2016 user consultation. The conclusions from this user consultation are generally in line with those from 2015, but more or more detailed information is provided now. In this section the most important things are summarized and compared with the conclusions from 2015 (conclusions that were drawn in 2015 are presented in italics). Also some information is given on what can be done with this information.

"General" Information

In the 2016 user consultation relatively more climate researchers or climate service researchers were included than in 2015. Conclusions that were drawn also in 2015:

- *Temperature variables: many stakeholders indicate that they are interested in minimum and maximum temperatures, besides average temperatures.*
- *Several stakeholders use indices, however, regularly these are determined by themselves*
- *Many stakeholders indicate that they generally use more climate variables than just temperature.*

Data sources, time period, area and resolutions

User consultation 2015:

- *A dataset with a daily resolution is often sufficient for the stakeholders consulted. Some also would like to have sub daily data, however, regularly the supply of minimum and maximum temperatures per day is sufficient for them;*
- *The requested spatial resolution varies with the stakeholders (from very local to coarser resolution). In general stakeholders want a high spatial resolution.*
- *Many of the stakeholders would like to have datasets starting around 1950 up to 1981 (often for describing the current climate), only a few need much longer datasets (100 year or longer for trend analysis);*
- *The area for which data are requested also varies largely per stakeholder. Some are currently working on a global level or in Europe (from specific places/regions to the continental level), but some work at the local level.*
- *The data mentioned by the stakeholders are very divers, from local meteorological station observations to datasets with continental or global coverage such as E-OBS(ECA&D), GHCN, HadCRUT, re-analysis*.

User consultation 2016:

- No real new insights in user needs compared to 2015, although now relatively more people indicated that they would prefer to have long datasets (starting before 1950).
- More information was collected on which data sources are used now (see figure presented in Section 3.1). This is a large number of data sources. The EUSTACE data can only be compared with a limited number of these.

Data format

User consultation 2015:

- *Most important for all potential users is the easy accessibility of the dataset. However, the interpretation may differ between various stakeholder groups. Stakeholders familiar with large climate data sets often mention the NETCDF is the preferred format and that they would like to have scripts for downloading. Also tools for customizing would be very useful (e.g. selecting specific areas and time periods and visualizing);*
- *Other stakeholders also would like to have tools for customizing, but also for processing the data to other formats, to calculate indices or statistics or to*

aggregate the data for specific regions. These stakeholders more often request also other data formats such as ASCII or Excel;
- *Supply through other portals such as Climate-Adapt or Copernicus was mentioned a few times.*

User consultation 2016:
After the user consultation in 2015 it was decided to use NetCDF as the format in which the EUSTACE products would be provided. Since we know from the 2015 consultation that not everyone is very familiar with the format, questions were included to get an idea what kind of support is needed (more information under "user guides").

- In the 2016 consultation, relatively a lot of participants had experience with NetCDF. Many working in climate research and services are familiar with NetCDF, but those working on impacts of climate (change) often not.
- Most people, also those familiar with NetCDF, would like to have support. This support ranges from example scripts, providing (links to) tools for conversion (especially ASCII and Excel are mentioned) and processing/analysing, tools for extraction of data, to detailed examples on how to access the data, how to process and interpret the data (see also under "user guides"). Links to some tools will be provided within this project in the user guide, although it is impossible to be complete. Where available some experiences of users can be added.
- When asked which tools/packages do not work well with NetCDF, some examples are mentioned (ArcMAP, ArcGIS, Quantum GIS, Excel, R-packages, some Python modules). What works well or not may also depend on e.g. versions, since others mention that they do use e.g. R for working with NetCDF. Information on known issues/problems will be provided in the user guides, however, new problems may appear later on. It might be an option to have e.g. a web page/wiki where these issues can be reported by users to inform others also when the project is finished.
- Depending on the format of the NetCDF files the data can or cannot be used easily with standard open source tools. Mock-ups of the data files were sent to a limited number of users, and the feedback on this was described in Section 4.
- The data file mock-ups presented in July 2016 worked well for most of the potential users, although some remarks for improved were given. Mock-up A for presenting local solar time (local solar time ancillary variable) worked best and was preferred by the limited number of users that could give feedback.
- To make the EUSTACE data useful to a wide range of users, it is advised to give the air temperature estimates for all surfaces in grid cells containing more than one type of surface.

User guides
User consultation 2015:

- *Climate scientists request especially a good description of the methods (including information on how the uncertainties are determined) and assumptions used, technical information (including things like metadata, file structure) and comparison with data sets. Generally, they have enough background knowledge to know how to interpret the data.*
- *Other stakeholders more often request information on how to process and interpret the data, examples and/or good practices, or they request guidance in a more general sense. Often requested is information on uncertainties: understandable information on uncertainties, managing and dealing with uncertainties, importance of the wording used. From this it can be concluded that some tailoring for different stakeholder groups is needed when writing the user guidelines. The tailoring also concerns how is communicated about uncertainties;*
- *Comparison with other datasets (e.g. E-OBS, HadCrut, Re-analysis) is needed to determine the quality and/or to be able to judge which dataset is most suitable for a particular project;*

- *Further needed: reference for the dataset and the methodologies, description of the data policy and version management.*

User consultation 2016:

- A quick start guide could be useful, since few people read user guides completely before using a data set. However, it should be short and contain among others file format and content, access information, quick overview (resolution, spatial coverage, period, etc.), known limitations/issues[8], link to further information, and a few pages on the applications (e.g. how to interprete, do's and don'ts).
- Probably as important as a quick start guide is a clear structure of the user guide to make it easy to find the information that the various users need. Therefore, a good table of content is very important and options to navigate quickly through the user guide. Also good cross referencing in the user guides may be important.
- Some important information can also be given in the NetCDF files.
- No real new subjects were proposed for the scientific and product user guide compared with our proposed set-up. Some respondents didn't find the distinction between scientific and product use guide very clear.
- Links to existing tools and scripts/readers are considered useful, especially for software packages regularly used (e.g. for C, Python, R) or provide links, or lines of code for extracting. NetCDF has problems in some cases. For this it would be helpful to have an idea on the most used software packages. Also tools for visualizing, extracting data and processing are considered useful.
- Clear and consistent use of terms and definitions is very important throughout the user guides (definitions referring to some standard).
- Examples (especially with open source tools) that allow "copy and paste" e.g. to read data are regularly mentioned as useful. Also examples that show "how working with EUSTACE data worked out" were considered interesting.
- Tools for visualizing (e.g. Panoply, ADAGUC, KML quick look for Google earth) are considered useful, also for users not used to working with NetCDF, to see whether they are of use to them. They hide the complexity of NetCDF. Usually they can be used on a windows platform.
- Those not familiar with NetCDF require some introduction to it and probably more detailed examples
- Only a limited number of people is familiar with the ESGF portal. From this we can conclude that the user guide should certainly contain information/guidance on how to access and download data from the ESGF portal.

Many of the requests/suggestions mentioned can be included in the user guides, however, it is impossible to provide support for all available packages. Therefore it is recommended to focus especially on the most used packages such as Python, R, Matlab.

Uncertainties

User consultation 2015:

- *All stakeholders recognize that information on uncertainties is important, however, although not mentioned explicitly, many have difficulties using it or use it mainly to get an idea of the quality of the data;*
- *Providing the information on uncertainties also in another format than just an ensemble would make it easier for several stakeholders to use the information;*
- *The term "uncertainty" is interpreted in diverse ways by stakeholders. Good communication and framing is important for correct interpretation of the uncertainty information provided by EUSTACE.*

[8] Some of the problems with tools/packages can be described in the user guides, however, new problems may appear later on. It might be an option to have a web page/wiki where these issues can be reported by users to inform others.

User consultation 2016:

The results of this user consultation provide similar results and some more detailed information on uncertainties.

- Very similar to 2015, most respondents use in some way or another uncertainty information, ranging from general use to check the quality of the data to the use of ensembles/percentiles. Those that do not, often indicate that it would be useful/interesting to use it in the future. What kind of uncertainty information is used and how differs greatly among the respondents.
- The term uncertainty is interpreted differently: 1. as a measure of likeliness/probability and 2. as a measure for quality. It is mentioned also a few times that the information on uncertainties should be easy to understand and easy to use. This indicates that a clear description of the types of uncertainties is required in the user guides, and also examples on how the provided information can be used or interpreted.
- Many indicated that besides a "best estimate", they would desire to have a measure of the spread (standard deviations, percentiles (not indicated which) or maximum and minimum values). Although as one of the respondents indicates "everything can be useful", users do not want to be overloaded with information. It was suggested also to produce quality flags to make it easier to choose "best quality" data. For the selection of which uncertainty information will be provided and how, it should be considered that some climate models require certain information in order to ingest the data.
- A few people indicated that they would use more than one ensemble member. These persons all have a lot of experience in the use of climate data.
- Most of the respondents considered information on the separate sources of uncertainties useful or thought that it could be useful, but at the same time they think that most users will only use the overall uncertainty and don't want to be overloaded with information. E.g. in GlobTemp, the total uncertainty is given as default and components as optional extra data set. Many of these respondents have a background in climate science. As project we will consider whether this information can be made available, but also check whether and how it will be used.
- There is also a need to describe how well the ensemble conveys the uncertainties in the products – is the ensemble a good reflection of the true uncertainty in the product?

Updating EUSTACE data sets?

Although there is no financing yet for updating the EUSTACE dataset, it was indicated already in 2015 that the dataset would lose its added value compared with other datasets that are updated regularly. When looking for a follow-on it is good to know what frequency of updating is desired by potential users. In this questionnaire a question was included on the desired updating frequency. Once per year and once per month are mentioned most regularly as desired updating frequencies (and of course when improvements in the algorithms are made an update is desired). In case of updates, most people are interested in receiving alerts by e-mail.

Appendix 1 Flyer on user consultations

EUSTACE USER CONSULTATION WORKSHOPS

EU Surface Temperature for All Corners of Earth

You are cordially invited to participate in one of our user consultation workshops:

- **Splinter meeting at EGU (SPM15):** Tuesday April 19, 2015, 8:30-12:00, room 2.96, Vienna, Austria (EGU-conference, max. 45 persons).
- **Co-organized session at EGU:** Taking the temperature of Earth: Variability, trends and applications of observed surface temperature data across all domains of Earth's surface. Oral presentations: Monday, 18 Apr, 13:30–15:00, Room 2.47. Posters: Authors in Attendance: Monday, 18 Apr, 17:30–19:00.
- **(Virtual) user consultation workshop:** May-June, 2016. Time and location to be determined.

What is EUSTACE?

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.

Why EUSTACE?

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

- Health and well-being: it can cause cold stress or heat stress
- Food security: link between surface air temperature and crop growth and animal health
- Energy: influence on the demand for heating or cooling, high temperatures compromise the efficiency of solar panels
- Commerce: it affects the sales of a large variety of products
- Tourism: it affects the attractiveness of a region, risk for bush fires
- 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. To achieve this, we must develop 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 models.

How?

The EUSTACE project will use new statistical techniques to provide information on higher spatial and temporal scales than currently available, making optimum use of the information in data-rich eras. EUSTACE will undertake this work between January 2015 and June 2018. The final and intermediate products (e.g. overview of current data sets on temperature, satellite skin temperature retrievals over all domains with consistent uncertainty estimates; station time series with discontinuities identified; information on the relationship between skin and air temperature over different domains and in different seasons) of EUSTACE will be interesting for many applications.

Differences from other surface temperature data:

- globally complete daily dataset from 1850 on
- designed in collaboration with users
- validated information on the certainty of each daily value as an integral component
- air temperature over the ocean, rather than sea surface temperature
- use information from satellites to estimate air temperature over all surfaces of Earth
- new statistical techniques to create complete pictures of air temperature everywhere

Potential users and their requirements?

To increase the usability of the EUSTACE products, we would like to involve potential users from the early start. Potential users can get involved at different levels:

- Regularly getting information on the progress and products of the project;
- Participating in user workshops or questionnaires giving feedback on requirements and preliminary products;
- As trail blazer users testing preliminary results from EUSTACE (limited number only).

Interested?

If you are interested, please come to our EGU splinter meeting, session or poster session. If you want to participate in the (virtual) workshops in May-June, or if you just want to be kept informed please send an e-mail to Janette Bessembinder (bessembi@knmi.nl). You may also give this information to anyone else who is potentially interested.

Or look at our website: https://www.eustaceproject.eu/.

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

Appendix 2 Questionnaire

Below the text and questions in the questionnaire are presented.

General information

What is EUSTACE?

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.

More information about the project can be found at our website: https://www.eustaceproject.eu/about/.

Potential users and their requirements?

To increase the usability of the EUSTACE products, we would like to involve potential users from the early start. In 2015 a first user consultation was held and results were used in the first stage of the design of products (https://www.eustaceproject.eu/about/eustace-deliverables/). This year we have a second round of user consultations. Below you find a questionnaire about user requirements.

We would be pleased if you could help us by filling in this questionnaire. Your answers will be treated anonymously, however we would like to have your name and e-mail address to contact you in case of any questions.

Those that participated in the survey in 2015, do not have to fill in again the information on the current use of climate data.

Q1: What is your name?
Q2: What is your e-mail address?
Q3: In which sector are your working? E.g. climate science, health, climate services.

Your current use of temperature data

Q4: For which purposes, if any, do you use temperature data currently?
Q5: Which data set(s) do you use now for these purposes?
Q6: Do you also need data on other climate variables besides (average, minimum, maximum) temperature? If so, please specify
Q7: Which spatial resolution do you need?
Q8: Which temporal resolution do you need?
Q9: For which period in the past do you need data (length of the data set)?
Q10: For which continent/region/country do you need data (spatial coverage)?

Familiarity with NetCDF format

Q11: Do you need information about: Average temperature, Minimum temperature, Maximum temperature, Diurnal temperature range, Specific temperature indices, Length of warm/cold periods, Other (please specify)
Q12: Do you also use information about uncertainties in climate variables? If yes, please specify what uncertainty information you use and how.
Q13: Which data format do you use currently? E.g. NetCDF, ASCII.
Q14: Are you familiar working with NetCDF-files?
Q15: Do you need support for using the NetCDF format? If yes, please specify what kind of support?

Q16: Several tools exist for visualizing NetCDF data and for extracting data for certain periods and regions (e.g. Panoply, ADAGUC). Is referring to these existing tools and their user guides sufficient?
Q17: Are there tools/packages you work with, that do not work well with NetCDF? If so, please specify.
Q18: Other remarks or suggestions related to the use of NetCDF?

User guides

We currently proposed to make three user guides:

- Quick start guide
- Scientific user guide (e.g. scientific overview: peer-reviewed articles and additional information, datasets/material used, processing workflow, validation and quality checking, estimating uncertainty, comparisons with other datasets, future directions)
- Product user guide (e.g. overview of products, including differences with other datasets, where possible include explanations about the consequences for e.g. accuracy/uncertainty (e.g. in data sparse regions), product details: metadata, including links to the scientific user guide, data structures and formats, working with products (different levels of detail or background info for novice or experienced users), how to find/get/download EUSTACE products, working with uncertainty estimates, assumptions and do's and don'ts (interpretation), FAQ's, background information for less experienced users (including links to useful processing tools, etc.), case studies

Q19: What should be included in the quick start guide (at minimum)?
Q20: What subjects or documents should be included in the scientific user guide (at minimum)?
Q21: What subjects or documents should be included in the product user guide (at minimum)?
Q22: How do you use user guides at this moment? Other (please specify)
Q23: Can you give examples of good or bad user guides? Please provide link or title.

Uncertainties

EUSTACE will produce consistent uncertainty estimates for satellite skin temperature retrievals over all surfaces (land, ocean, ice and lakes), so we know how far to trust the estimates everywhere.

Uncertainties are categorized by effects whose errors have distinct correlation properties (for more info: Presentation D. Ghent):

- Random (level 1 NEdTs, Geolocation)
- Locally systematic (atmospheric effects, emissivity)
- (large-scale) Systematic

An ensemble will be created, but also summarized information on uncertainties can be presented. We would like to know how you might use the uncertainty information and what would be the most suitable format to present it.

Q24: Is information on the separate sources of uncertainty useful for you? If so, please indicate how you could use this information?
Q25: If we summarise uncertainty, what is useful?
Q26: Would you use one or more members of the ensemble? If yes, please specify which and how.
Q27: Other remarks or suggestions related to uncertainties?

EUSTACE products

EUSTACE will produce the following products:

- Station series and E-OBS update: Global data set of daily weather station air temperature measurements (Tmax and Tmin) with non-climatic breaks identified – Station time series and gridded for Europe (Dec 2016)
- Satellite skin temperature retrievals: Daily satellite skin temperature estimates for all surfaces of Earth with consistent uncertainty estimates – Gridded or along satellite's track (Apr-May 2016)
- Skin/air temperature relationships: Understanding of the relationship between surface skin and surface air temperature over all surfaces of Earth and in different seasons – report (Dec 2016)
- Air temperature estimates from satellites: Daily estimates of surface air temperature from skin temperature retrievals - Gridded or along satellite's track (TBD; Jun 2017)
- Globally complete air temperature fields: Globally-complete daily fields of surface air temperature over all corners of Earth since 1850 – Gridded (0.25° lat/lon) perhaps an ensemble. (Maximum and minimum temperature over land, average temperature elsewhere; May 2018)
- Derived products: e.g. global means and climatologies (May 2018)

Q28: The products will be made available through ESGF catalogue portals. Are you familiar with obtaining data sets through these portals?
Q29: Would you wish to receive alerts about problems/new releases of the EUSTACE datasets and by what means?
Q30 How often would you like the air temperature dataset to be updated?

Appendix 3 Information on data file mock-ups

A first mail to a limited number of potential users was sent in July 2016, with a link to the website where they could find the mock-ups (see below). In the beginning of September 2016 a mail was sent again. Hardly anyone had reacted on the first one, probably due to the summer holidays. Some additional feedback was requested on what should be included in the data files and how (see Appendix 4 on Mock-up presentations for Consistent air temperature data from satellite (SATSTACE)).

Below the information is presented that the potential user could find when logging in at the following site: http://gws-access.ceda.ac.uk/public/eustace/mockup/

EUSTACE File Format Mockups (version July 2016)

The following NetCDF files demonstrate the proposed NetCDF format for the EUSTACE infilled product. The mock data contained in these files is entirely fictitious. These files do not in any way represent the coverage or quality of data which will later be made available by the EUSTACE project. As such, the files should not be used for anything other than evaluating the suitability of the file format. The files should not be redistributed in any way, and should not be used in the creation of any presentations or publications of any kind.

User Guide

Coming soon.

Downloads of the user guide document will appear here when available.
Meanwhile the following information gives a brief outline of the file format:

- The data is stored in NetCDF format.
- There is one file per day of data.
- In this mockup we present just 5 days, whereas in the real project output there will be ~60000 days, beginning on 01/01/1850.
- Each file contains a global field of values for a surface air temperature variable, together with information about uncertainty.
- In this mockup we have a variable named **tas** which is the name to be used for mean daily surface air temperature.
- The **tasstandarddeviation** field indicates the total uncertainty.
- The **tasclimatologyfraction** field indicates the extent to which the results rely on inference from more distant points in space and time, rather than local weather observations.
- Similar separate files will be used to present **tasmin** (minimum daily surface air temperature) and **tasmax** (maximum daily surface air temperature).
- An ensemble of possible realisations will also be presented in the final project output. Where the climatology fraction is high the ensemble should be used to get a full picture of uncertainty structure.

Compressed Archive of NetCDF Files

eustace.mockup.R000262.tar.gz [5.6MB]

Individual NetCDF Files

2010/eustace.mockup.R000262.tas.20100601.nc [6.0MB]
2010/eustace.mockup.R000262.tas.20100602.nc [6.0MB]
2010/eustace.mockup.R000262.tas.20100603.nc [6.0MB]
2010/eustace.mockup.R000262.tas.20100604.nc [6.0MB]
2010/eustace.mockup.R000262.tas.20100605.nc [6.0MB] Friday 1st July 2016

EUSTACE File Format Mockups (version end of November 2016)

The following NetCDF files demonstrate the proposed NetCDF format for the EUSTACE infilled product. The mock data contained in these files is entirely fictitious. These files do not in any way represent the coverage or quality of data which will later be made available by the EUSTACE project. As such, the files should not be used for anything other than evaluating the suitability of the file format. The files should not be redistributed in any way, and should not be used in the creation of any presentations or publications of any kind.

User Guide

Coming soon.
Downloads of the user guide document will appear here when available.

Meanwhile the following information gives a brief outline of the file format:

- The data is stored in NetCDF format.
- There is one file per day of data.
- In this mockup we present just 5 days, whereas in the real project output there will be ~60000 days, beginning on 01/01/1850.
- Each file contains a global field of values for a surface air temperature variable, together with information about uncertainty.
- In this mockup we have a variable named tas which is the name to be used for mean daily surface air temperature.
- The tasstandarddeviation field indicates the total uncertainty.
- The tasclimatologyfraction field indicates the extent to which the results rely on inference from more distant points in space and time, rather than local weather observations.
- Similar separate files will be used to present tasmin (minimum daily surface air temperature) and tasmax (maximum daily surface air temperature).
- An ensemble of possible realisations will also be presented in the final project output. Where the climatology fraction is high the ensemble should be used to get a full picture of uncertainty structure.

As of November 2016 mockups show that temperatures are expressed with respect to local solar time, as will be the case in the final system design. The UTC sample period for each grid box therefore depends on longitude. This can be expressed either as an ancillary variable describing the temporal offset (Mockup A) or more explicitly using an additional main variable to describe the observation time (Mockup B). Only one of these will be selected for inclusion in the final design.

Mockup A (local solar time ancillary variable)

Compressed Archive of NetCDF Files [November 2016]
eustace_mockupancillarytime_R000492.tar.gz [5.6MB]

Individual NetCDF Files [November 2016]
2010/tas_mockupancillarytime_eustace_0_20100601.nc [6.0MB]
2010/tas_mockupancillarytime_eustace_0_20100602.nc [6.0MB]
2010/tas_mockupancillarytime_eustace_0_20100603.nc [6.0MB]
2010/tas_mockupancillarytime_eustace_0_20100604.nc [6.0MB]
2010/tas_mockupancillarytime_eustace_0_20100605.nc [6.0MB]

Mockup B (local solar time as main variable)

Compressed Archive of NetCDF Files [November 2016]

eustace_mockuplocalsolar_R000492.tar.gz [5.6MB]

Individual NetCDF Files [November 2016]

2010/tas_mockuplocalsolar_eustace_0_20100601.nc [6.0MB]
2010/tas_mockuplocalsolar_eustace_0_20100602.nc [6.0MB]
2010/tas_mockuplocalsolar_eustace_0_20100603.nc [6.0MB]
2010/tas_mockuplocalsolar_eustace_0_20100604.nc [6.0MB]
2010/tas_mockuplocalsolar_eustace_0_20100605.nc [6.0MB]

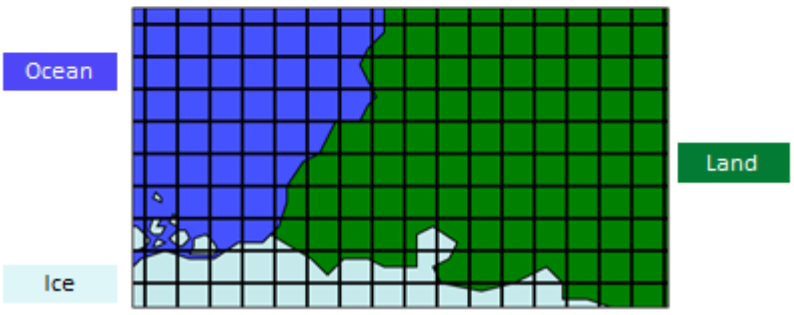
Appendix 4 Mock-up presentation for consistent air temperature data from satellite (SATSTACE)

Basic Mock-ups for SATSTACE

What is the most convenient way of providing the air temperature estimates and the attached uncertainties?

Product description

- Surface air temperature estimates (with estimates of uncertainty) for all surfaces of Earth, derived from satellite surface skin temperature retrievals;

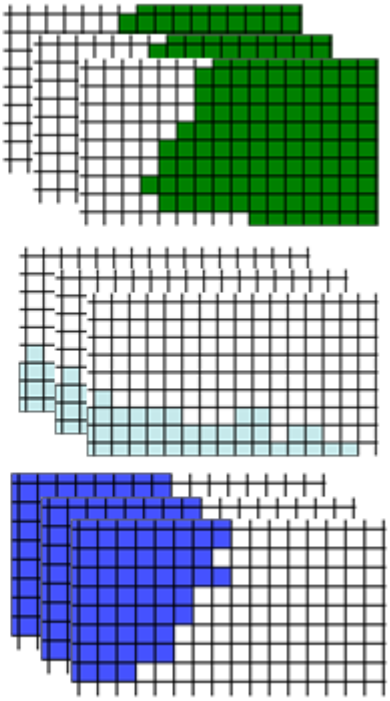


There are a number of ways that we could provide this information. Schematics of those methods are shown on the next few slides. We would like to have feedback from (potential) users on what would be the most convenient way to present the information. Also potential advantages or disadvantages of the various ways of presenting from the perspective of users would be interesting information for the design of the final EUSTACE products.

We have a estimates of air temperature derived from satellite-retrieved skin temperatures. The surfaces are Land, Ocean and Ice, for which different methods are developed to estimate air temperature from surface temperature from satellites.
For each surface there may be multiple satellite sensors from which air temperatures have been derived.
The data will have gaps, due to the limited field of view of the satellite as well as things like cloud cover and aerosol contamination.

Detailed uncertainty information will be provided with each set of fields. Uncertainty components will be broken down into random, structured random and correlated components. There may be more than one of each type of uncertainty component for example if two error sources have rather different correlation scales.

This is potentially a large amount of information



For more information about the products of EUSTACE: https://www.eustaceproject.eu/eustace/static/media/uploads/eustaceuserconsultations_2016_intro_nick_v1.pdf (slide "Eustace products) or see slide below.
For more information about the various types of uncertainties: https://www.eustaceproject.eu/eustace/static/media/uploads/eustace_egu_splinter_session_20160419_dg.pdf

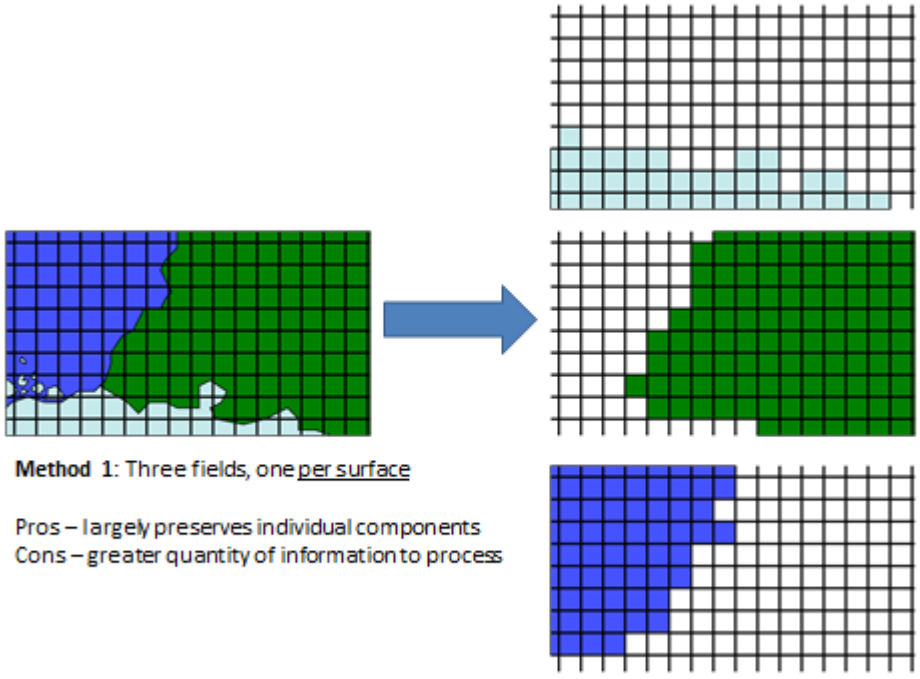
EUSTACE PRODUCTS

Product	Description	Date
Station series and E-OBS update	Global data set of daily weather station air temperature measurements (Tmax and Tmin) with non-climatic breaks identified – Station time series and gridded for Europe	Dec 2016
Satellite skin temperature retrievals	Daily satellite skin temperature estimates for all surfaces of Earth with consistent uncertainty estimates – Gridded or along satellite's track	Apr-May 2016
Skin/air temperature relationships	Understanding of the relationship between surface skin and surface air temperature over all surfaces of Earth and in different seasons – A report	Dec 2016
Air temperature estimates from satellites	**Daily estimates of surface air temperature from skin temperature retrievals - Gridded or along satellite's track (TBD)**	Jun 2017
Globally complete air temperature fields	**Globally-complete daily fields of surface air temperature over all corners of Earth since 1850 – Gridded (0.25° lat/lon) perhaps an ensemble. (Tmax and Tmin over land, Tmean elsewhere.)**	May 2018
Derived products	For example, global means and climatologies	May 2018

The issues described above and below are relevant for the "Air temperature estimates form satellites" and for the "Globally complete air temperature fields" (where the gaps mentioned in the former slide have been filled in and where estimates of air temperature before the satellite period have been added).

Questions

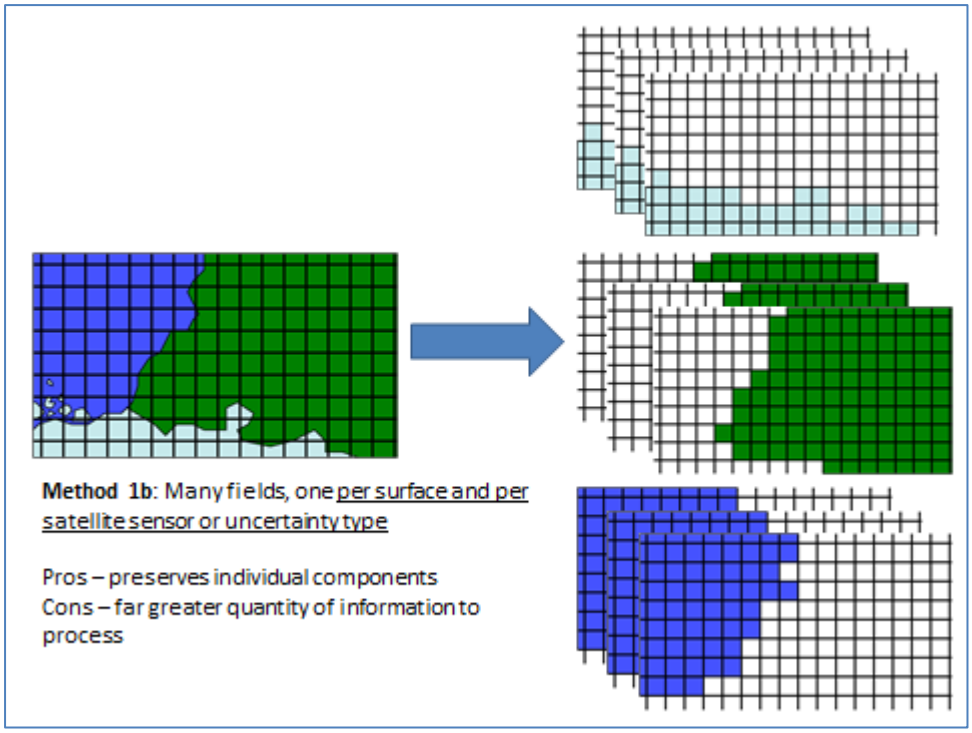
- There are a number of ways that we could provide this information. Schematics of 3 options on how we might do this are shown on the next few slides.
- How would you use the data in your analysis?
- How would you use the uncertainties in your analysis?
- Would any of these options prevent you from using the data (and why)?
- How would you prefer the data to be presented?
- Do you have other suggestions on how to present the information?



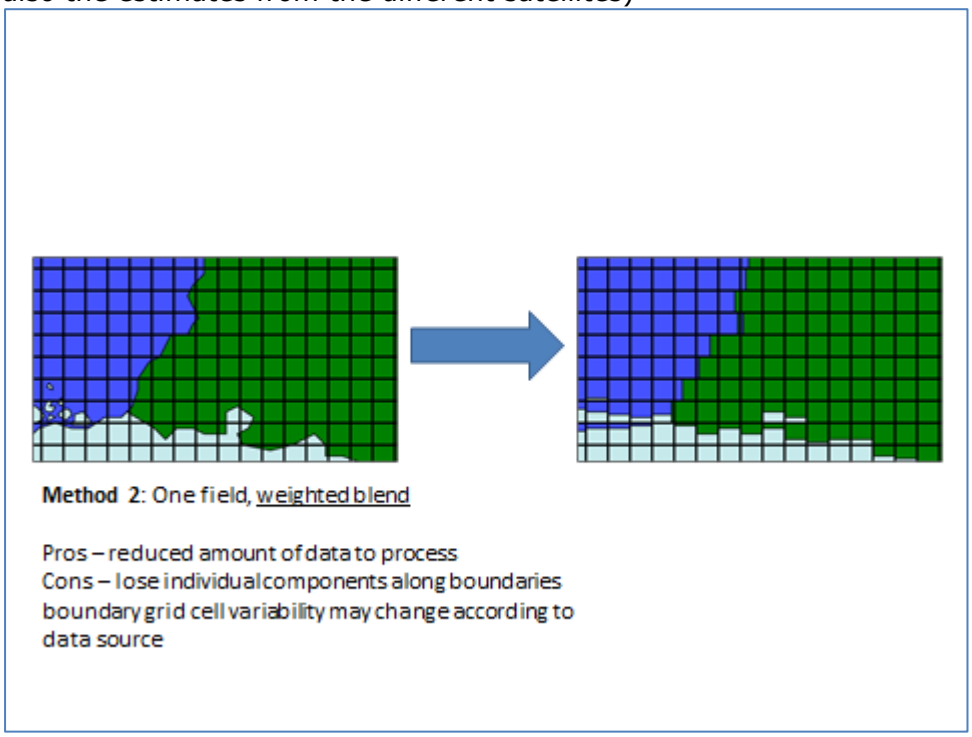
Method 1: Three fields, one per surface

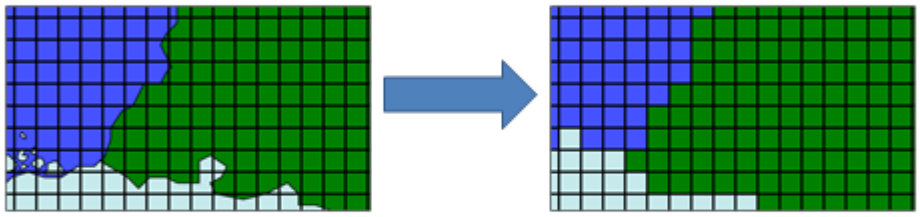
Pros – largely preserves individual components
Cons – greater quantity of information to process

The method can be used for the estimated air temperatures as well as for the related uncertainties (total uncertainty)



As for method 1, but now the separate types of uncertainty are presented (and possible also the estimates from the different satellites)





Method 3: One field, hierarchical blend
Land is used in preference to ice or sea
Ice is used in preference to sea

Pros – gives greater prominence to land data
Cons – gives greater prominence to land data
lose individual components along boundaries
boundary grid cell variability may change according to data source

Other options?

Appendix 5 Feedback from individual users

KNMI, climate researcher involved in project PRIMAVERA

(description of what was discussed during an interview):
The EUSTACE datasets may be very useful for them for validation of the high resolution modelling in this project.

- He indicated that he would have a clear preference for method 1 (or 1b) with separate data files for the air temperature estimates over land, ocean or ice. He explained that the masks for land-water used in climate models are often not the same as the real "masks". This has to do especially with the modelling resolution (small island may not be explicitly included, Scandinavia may be connected with Denmark, etc.). Therefore using the other options may cause difficulties when trying to validate temperatures for locations on the "borders" of different land surfaces. Having different data files for the different surfaces would allow them to use their own specific masks and therefore to have a better validation.
- He also indicated that he wasn't sure whether he would like to have (or use) data files with the air temperature estimates from the different satellites of data files with the separate uncertainty types. When I mentioned that during the user consultation several indicated that probably most users don't want to be bothered with a lot of data, but that for some specific users the information on the separate satellites or uncertainty types could be useful, he indicated that he agreed very much with that. However, 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 climate model has or does not have a (large) bias. He more or less said that he would prefer to leave it to EUSTACE to make a good estimate of the total uncertainty (and therefore not having to decide himself how to use the uncertainty information on the separate types of uncertainty).
- Validation of climate model data can be done in two really different ways: 1. Compare the statistics produced by the climate model with the statistics of the "observations"; and 2. Compare the model estimates for specific days with the "observations" for specific days. In the second case information on (total) uncertainty per day is needed, but in the first option some derived (total) uncertainty is required. Something like this was mentioned before (also the uncertainty not just for a grid cell but for a larger area).

Detailed feedback on test of mock-ups was also later provided by email.

CIIFEN, Ecuador, person involved in setting up similar system as ECA&D:

We didn´t find any problem in the structure, it works fine, we use panoply, ODV and CDO to open them and works fine with these software.
About the question on method:

1. I think, method 3 will simplify manipulation of data for most users, maybe some of them would like to have different files for everything, but operationally is better to have one file, giving prominence to land, where most of users are interested in.
2. Uncertainties the same, one file, which the message, Is this grid point useful or not, some kind of flag alerting about the uncertainty and better alerting about how useful is that grid point of the dataset.
3. Maybe to have two types of datasets, if this is not costly. For example, "operational dataset" (one file) and "expert dataset" (many files as needed). I think most of the users are going to be operational.

KNMI, researcher, involved in developing e.g. ADAGUC, Climate4impact

(description of what was discussed during an interview):
Much of the services developed for this Climate4impact portal are also used for the CLIP-C portal. It contains many options to visualize, process, make subsets, download data (also some measure of uncertainty can be projected over the maps, averages, percentiles, max/min can be calculated, etc). All datasets that are open data in the ESGF-portal can be used in this Climate4impact portal (automatically).

He had one remark related to the data file mock-ups: he saw that the data files were per day. He suggested that it would be better to make files with data for 1 year of more (most CORDEX data sets that can be used in Climate4impact have files with 4 years such that the size is about 1 TB). This would also make it easier to use all the possibilities of the climate4impact portal. He estimated that the size of 4 year files with the proposed structure would be around 2 TB. This is of course rather large for some users, but since the climate4impact portal has the option to "cut out" certain areas and time periods, this is maybe not such a problem.

I also talked with him about the other question (whether to present the data for land, oceans and ice, etc) in separate files. His first reaction was more in favor of method 2 and providing e.g. a land (and ice?) fraction for those grids with more than land surface (is done regularly in gridded data files). A version of method 2 could be more interesting in combination with e.g. the climate4impact tool (he said it was difficult to integrate files for different surfaces), although he also understood the preference of others for method 1. In principle it is possible to present the information for the different surfaces in one file (different "columns" for the different surfaces), however this would make the files much larger. This problem is only relevant for Tas, for Tasmin and Tasmax we will only have data over land anyway.

Further detailed feedback on testing of mock-ups was also received by email later.
"I down loaded the data. The data work well in ADAGUC. What struck me was that there is only one time step stored in a netcdf file, you could choose to put more in one file. I did this for the five files for tas.
The original extracted data and aggregate data are here in the autowms:
http://bhw485.knmi.nl:8080/autowms/autowms.php?dir=/eustace

- These mockups are in NetCDF format. You are familiar with this format, but which software packages or tools do you use to visualize, download and/or analyze the data?(I know you developed ADAGUC, but do you forsee any problems with other tools/software?) --> I use ADAGUC and ncview, both work fine with these EUSTACE NetCDF files.
- Do you sometimes use other tools/software, and if so do you expect problems using them? --> NCVIEW, works good with the datasets
- Were problems encountered? --> No.
- Are there things about this format that you would prefer to be presented differently? --> Please aggregate the individual files in one larger file with multiple time steps, this allows for easy time series visualization for a location."

University of Lisbon, climate change impact researchers

"Some impressions about the mockup data:
In general I think the organization of the data is already well done. Some details that might be improved:

- "There is one file per day of data." It might be useful to have other temporal intervals per file (example: 1 file per month).

- "tasstandarddeviation" field indicates the total uncertainty. The kind of standard deviation should be explicit: "sample standard deviation" or "population standard deviation"?)
- "Similar separate files will be used to present tasmin (minimum daily surface air temperature) and tasmax (maximum daily surface air temperature)." If both variables tasmax and tasmin were in the same netcdf file, it would simplify the processing for the end user (less files to download/manage)
- "These mockups are in NetCDF format. Do you already use NetCDF in your work, and if so with which software packages or tools to visualize, download and/or analyze the data?" I do use NetCDF and I think it is a good choice for the EUSTACE. To analyze the data I usually convert it to a Postgres database because I'm more familiar with SQL (and from there it's easy to visualize with GIS tools like QGIS)
- "Are you able to read the mock temperature fields contained within these files with the software that you usually use / or would like to use?" Yes, I had no problems to read the files (I used the 'ncdump' command line utility)
- "Were problems encountered?" No problems
- "Are there things about this format that you would prefer to be presented differently?" The names of variables could be more compact. Instead of "tasstandarddeviation" it would be more simple to use "tas_stddev", for instance. This will be useful when using the data with R or other programming environment."

"As for the way data is organized ("separate data files with estimates for land, oceans or ice vs all surfaces in one file"), I'm not sure which one is best but my instinct tells me that the more granular the data is, the best. That is, if you have all the components separated, the end user has more possibilities to work with it ("extract-transform-load"). On the other hand, it also makes sense to have the common case already ready to use (whatever that common case is). This could be done by having different files: one with the common cases (aggregated data), other with the separated components."

"Other improvement to make the outputs more accessible is to have the data in other (more common) formats, such as shape files, and even plain images (just for visualization purposes)."

CCRC (Climate Change Research Centre), climate researchers, interested esp. in climate extremes

Provided the following feedback:

- "We're mostly interested in a sensitivity analysis of extremes based on input dataset (that is, we would ultimately like to use EUSTACE in tandem with e.g. other in situ-based and reanalyses datasets). Having files in a similar format to current reanalyses/climate model output would be ideal, but for our particular application, we're interested in using daily maximum, minimum and mean temperatures for the globe, although in our current work we would only use the land-only values. We will use standard software to perform analysis (e.g. NCL, R, Fortran, Python etc) and are used to working with NetCDF format. Generally, the layout of the mock-up files is good, and no problems were encountered. We have used both R and NCL, as well as tools like ncdump, ncview and CDO to examine the files and their contents.
- It would be good to split the different types of uncertainties into separate files (or to set the NetCDF files up in such a way as that uncertainties could be extracted as 'variables'). But for us one estimate of the total uncertainty is also okay as long as it is clear in the file and easy to understand and interpret. In terms of air temperature fields in each file, we would be fine with all estimates from land,

ocean and sea-ice to be in the one file as long as appropriate masks were also included.
- While we would prefer all of the above, we could work with any of the various options you provide in the 'Basic Mockup for SATSTACE as long as they were CF-compliant.

We hope this feedback is useful to the team and please do get back to us if there is anything else we can help with."

KNMI, climate researcher, works a.o. on attribution of climate extremes

"I had a look at the sample files. Please note that my knowledge of data formats is out of date and mainly limited by having to deal with whatever people throw at me for inclusion in the Climate Explorer web site. On the whole, it looks useful and as far as I can tell CF-compliant (but please ask experts for that!). Some minor comments and one major one:

- Some software appreciates an attribute "axis" on longitude ("X") and latitude ("Y").
- Why are the units duplicated in long_name? This is not adjusted by software that operates on the data and hence will be misleading after processing, eg by taking monthly means. conversion from K to Celsius, ... Please leave out.
- I always appreciate a contact e-mail address and/or web link in the global metadata.
- A "cell_methods:mean" in the time axis would be appreciated for daily means."

"I have no experience in dealing with descriptions of uncertainty yet. Although I strongly support their use I cannot handle them at the moment except for ensembles (and that only very partially)."

KNMI, researcher, works especially on ECA&D and has contact with CIIFEN

(description of what was discussed during an interview)

- Tried to upload the data-files through ADAGUC and the Climate4impact portal (also uses ADAGUC) and had no problems with that.
- Often uses CDO or own C-modules and sometimes Ferret, but only for visualizing
- CDO seems to have some problems with using rotated grids, but here a regular grid is used
- Because of the command structure of CDO (CDO command inputfile_1 inputfile_2 output file) it would be easier to have the variables (minimum, maximum and average T) in separate files, otherwise one will have to make separate files first.
- Also thought it would be better to have the air temperature estimates separate for land, ocean and ice surfaces (so no some kind of weighted average or hierarchy for grid with more than one surface type)
- Since there are relatively few people that will use the temperature data for land, oceans an water at the same time, would advise to have separate files for them. Those that do need the air temperature for all surfaces often also know how to combine the datasets.
- Indicated that she would prefer to have data files with longer time periods. For EOBS they seem to have data files for the whole time period per climate variable, but also files with 15 year periods.

eLTER (Long Term Ecosystems Research)

"We use the NetCDF format routinely, so this should be no problem at all."

Appendix 6 User stories: what do users do with temperature data?

As a form of support for the EUSTACE partners, especially those working on the design of the data products, some examples were given on how potential users of climate data will or could use temperature data. Below examples are given for different sectors. Per sector there may be different types of users (e.g. scientists, consultants, policy makers)

Climate

Some examples of how climate scientists could use temperature data:

- Temperature data can be used to calculate temperature indices (climatologies, trends, etc.) as is done e.g. for Europe at the ECA&D website (www.ecad.eu/). For that purpose time series from stations[9] are used as the basis. Information on uncertainties is not used currently in this application, although information can be obtained on inhomogeneities or "suspected" time series. When uncertainties would only be given as standard deviations or percentiles, it would not be possible to calculate the uncertainties of the derived temperature indices (it would only give some information on the quality of the underlying information). To calculate the uncertainty of the indices an ensemble of daily time series is needed that represents the uncertainties in the daily values well: the indices would then be calculated for each ensemble member and the uncertainty of the derived indices can then be calculated from the ensemble of the values of the indices (e.g. std or percentiles)
- The European Environment Agency also has several indicators that are calculated with temperature. One of them is the Global annual average temperature (http://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature-3/assessment). Another one is the Heating Degree Days (see under "energy"). These are calculated for the EEA by other (research) institutes. For the Global annual average temperature ranges in the trends are presented, but it is not clear how these ranges are determined; it should be possible to use an ensemble of EUSTACE data to generate the same indicators and their uncertainties.
- Validation of climate model simulations (determining the bias) for the past climate can be done in two really different ways: 1. Compare the statistics produced by the climate model with the statistics of the "observations"; and 2. Compare the model estimates for specific days with the "observations" for specific days. The second method is used for reanalysis. In this second method information on (total) uncertainty per day is needed. With this information it can be checked whether the climate model simulation is really significantly different from the "observations" (sometimes the difference between the model simulation and the "observation" may be considerable, but if the uncertainty of the observation is large, the conclusion still may be that the differences are not significant). In the first method some derived (total) uncertainty is required for the statistics. However, until now often no information is used on uncertainties in the "observations" to determine the quality/bias of a climate model run.
- Temperature observations can be used for statistical downscaling of global or regional climate model data. However, the spatial resolution of the EUSTACE dataset is more or less the same as the spatial resolution of many regional climate model data and therefore not interesting for spatial downscaling of many RCMs. In other regions of the world where hardly any RCM-simulations are available, the EUSTACE dataset may be used for statistical downscaling of GCM's

[9] In principle also gridded data can be used for this.

(although the spatial resolution of GCMs is also increasing as in the PRIMAVERA-project where a resolution of 0.25° is used globally)

Policy makers:

- These users will often use the indices or indicators mentioned above. In general they will not calculate/derive them themselves. This means that they will be indirect users. They may require background information on the quality of the used data (e.g. as is the case for the indicators of the EEA).

Water sector

In the water sector in most cases, besides temperature also data/information on precipitation is needed. Temperature is used to estimate evapotranspiration, melting of ice/snow, etc. There are few situations where only temperature data are used. An example could be the following:

- Water temperature of large European rivers and lakes (indicator of the EEA): http://www.eea.europa.eu/data-and-maps/indicators/water-temperature-1/assessment to determine whether the change in water temperature is due to an increase of air temperature or also due to other factors.

For most applications in the water sector the EUSTACE-dataset will not be the first choice, unless it can be combined with datasets with the other climate variables (but that will pose challenges due to differences in methods and therefore possible inconsistencies).
For operational purposes (water management of water boards, flood protection) information from the ensemble of weather forecasts is used, but no examples are currently known where uncertainties in measurements are used (uncertainties in climate model projections are used to a limited amount).

Agriculture

Researchers:

- Temperature is important for plant growth, since it determines the speed of the development of crops (in a relatively cold year it takes more time for wheat from germination to harvest, than in a relatively warm year). To estimate crop production in a region crop growth simulation models can be used. There are different types of crop growth models, but most use at least minimum temperature, maximum temperature, radiation and precipitation, but often also wind and humidity (to estimate evapotranspiration; http://www.eea.europa.eu/data-and-maps/indicators/crop-yield-variability-1/assessment-1) are used. Often data on daily basis are used. For (most) crop growth simulation models the EUSTACE dataset will not be useful, unless it can be combined with datasets with the other climate variables (but that will pose challenges due to differences in methods and therefore possible inconsistencies at the daily level). Uncertainty information on climate data is not used commonly in crop growth simulation, although it would be relatively easy to use various ensemble members. For the crop growth forecasting the ECMWF's Ensemble Prediction System (ENS and ENSextended) and Seasonal forecast model (SEAS) is used as input (http://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Meteorological_data_from_ECMWF_models)
- Phenological aspects are sometimes linked directly to temperature, but often other climate variables also affect phenology. Possible indicators based only on temperature: length of growing season, number of frost free days, etc. http://www.eea.europa.eu/data-and-maps/indicators/growing-season-for-agricultural-crops-1/assessment
- For animal husbandry temperature is important, because it affects the well-being of animals (and possibly the heating or cooling requirements). Information on

average daily temperature is not enough; at least minimum and maximum temperature are needed to detect whether there are periods during the day that are uncomfortable. Often also information on radiation, humidity and wind is needed since it is the "perceived temperature" that is important and not just the air temperature.

Policy makers:

- These users will often use the indices or indicators mentioned above. In general they will not calculate/derive them themselves. This means that they will be indirect users. They may require background information on the quality of the used data (e.g. as is the case for the indicators of the EEA).

Health

Several examples exist where health problems/risks are linked to extreme temperatures:

- After the extreme heat in 2003 mortality due to heat stress got much more attention and studies were published where mortality rates were linked to daily maximum and average temperature (e.g http://www.eea.europa.eu/data-and-maps/indicators/heat-and-health-1/assessment-1)
- Warmer winters may facilitate the expansion of Lyme borreliosis to higher latitudes and altitudes, particularly in northern Europe, but it would decrease in parts of Europe projected to experience increased droughts. Mosquito- borne diseases: Various studies have found that warm seasonal and annual temperature and sufficient rainfall provide favourable climatic conditions for A. albopictus in Europe[xxiii]. http://www.eea.europa.eu/data-and-maps/indicators/vector-borne-diseases-1/assessment

In these studies it is realized that completeness and reliability of observations (also for climate) may differ between regions and/or institutions, but it is not made explicit at the above websites whether information on uncertainties in climate observations is used.

Tourism climate indices (TCI) are commonly used to describe the climate conditions suitable for tourism activities, from the planning, investment or daily operations perspectives. They often use more than just temperature information, but some simple ones mainly use only temperature. The following article gives an example where climate data form the ENSEMBLES project were used to research the effect of uncertainties: https://earth-perspectives.springeropen.com/articles/10.1186/s40322-016-0034-y. In principle the same could be done using an ensemble for the current climate.

Nature/ecosystems

Some examples:

- Distribution and abundance of animal species: The Community Temperature Index (CTI) is a measure for the rate of change in community composition in response to temperature change http://www.eea.europa.eu/data-and-maps/indicators/distribution-of-animal-species-1/assessment
- Distribution of plant species: Several European plant species have shifted their distribution northward and uphill. These changes have been linked to observed climate change, in particular to milder winters. The variety of modelling approaches and results do not make clear statements as to where ecosystems and their services are at greatest risk from climate change. Furthermore, most ecological studies assess climate change (or just temperature change) in isolation from concurrent processes, such as increasing atmospheric CO2 concentration, soil water availability or land-use changes.http://www.eea.europa.eu/data-and-maps/indicators/distribution-of-plant-species-1/assessment
- Climate envelope modeling: relies on statistical correlations between existing species distributions and environmental variables to define a species' tolerance. Envelopes of tolerance are then drawn around existing ranges. By predicting

future levels of factors such as temperature, rainfall, and salinity, new range boundaries are then predicted. These methods are good for examining large numbers of species, but are likely not a good means of predicting effects at fine scales. In the following article also precipitation was used besides minimum and maximum temperature (on monthly basis) from the WorldClim database): http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2486.2006.01256.x/full.

  - As for agriculture often more climate variables than temperature are used/needed. When this is the case, the EUSTACE dataset may be of limited use, unless it can be combined with other datasets.
  - Since ecosystems are often complex, it is more difficult to determine the relation between a climate variable and the effect. Therefore, more often than for agriculture monthly of annual averages are used and also more often relations with a limited number of climate variables
  - WorldClim is regularly used in ecosystems research. It is often mentioned that this dataset has a spatial resolution of approximately 1 km (created by interpolation using a thin-plate smoothing spline of observed climate at weather stations, with latitude, longitude, and elevation as independent variables), however, it has a grid of 1 km, which is not the same as spatial resolution, but several users are not aware of this.

- In the Netherlands volunteers collect phenological data on many plant and animal species. Regularly these are linked to e.g. temperature data and this information is also used to get an idea of the effect of climate change on plant and animal species. E.g. relation between a certain tree species and average temperature in the months March-May https://www.naturetoday.com/intl/nl/observations/natuurkalender/about-de-natuurkalender

Energy

Some examples:

- Eurostat calculates Heating Degree Days (HDD) as (18 °C - Tm) x d if Tm is lower than or equal to 15 °C (heating threshold) and zero if Tm is greater than 15 °C, where Tm is the mean (Tmin + Tmax / 2) outdoor temperature over a given period of d days (http://www.eea.europa.eu/data-and-maps/indicators/heating-degree-days-1/assessment). No information on the uncertainty in input data is apparently used ("Data for calculation of HDD have been collected by Eurostat for decades; this indicator can therefore be considered as very reliable").
- Outside air temperature affects the process of extracting water from natural gas. Too high temperatures have a negative effect and therefore negatively affect the possibility to replete stocks for winter again. In the Netherlands there is a legal obligation for the gas company to have sufficient gas in stock to supply to households in extremely cold winters (up to a certain return time). Analyses are made to see whether there is a risk that they cannot supply sufficient gas in these extreme winters due to high temperatures in summer. Daily data are used (maximum and minimum temperature) or hourly data. For the current climate no information on uncertainty in the climate data is used, but for the future several climate scenarios are used.