EUSTACE System Technical Requirements

Joel R. Mitchelson

*Revisions*

|  |  |  |
| --- | --- | --- |
| **Date** | **Author** | **Comment** |
| 04/12/2015 | Joel Mitchelson | Created. |
| 09/12/2015 | Joel Mitchelson | Circulate to Met Office colleagues for comment. |
| 11/01/2016 | Joel Mitchelson | Add references and further definitions. |
| 12/01/2016 | Joel Mitchelson | Add sections from the text which defines each feature. |
| 13/01/2016 | Joel Mitchelson | Revise schematics and associated narrative. |
| 21/01/2016 | Joel Mitchelson | Revise system outline definitions, remove undecided requirements, remove ‘unknowns’ section, reference D4.1 user requirements, add detail to in-situ schematic and add satellite schematic, revise numbering of system requirements. |
| 11/04/2016 | Joel Mitchelson | Add PRDREQ015 (days defined using local solar time). |
| 08/02/2017 | Joel Mitchelson | Highlight processing blocks that may be removed. |
| 09/03/2017 | Joel Mitchelson | Simplify in-situ diagram by removing unused processing blocks, and add use of model covariate information for infilling. |

# Overview

This working document describes the known technical requirements of the EUSTACE end-to-end system for product generation; of the code produced; and of the products themselves. With reference to the description of action given in Annex 1 of the EUSTACE grant agreement , this document forms part of project deliverable D2.5: “Coded and tested system for product generation”, and was begun as part of project tasks T2.1 “Product Design” and T2.4 “System development”. The end-to-end system includes a step for checking quality, and this document provides input to the specifications for such a step.

# System Outline

## Definitions

For the purposes of this document and project work package 2 (WP2), a **product** is a collection of data files containing temperature data corresponding to a fixed time period. The **system** refers to the mechanism for creating products from input sources by execution of computer code; and the code itself.

An observational system is said to be **sustainable** if there is a defined way in which future temporal updates to input sources could be used by the system, so that the system has the potential to be used to generate up-to-date products for future time periods, using the same principles and quality standards as for the present.

A system with **appropriate** **maturity** for a given task is defined here as one whose maintainability, documentation, correctness, and usability meet the needs of that task – this is consistent with the definition used by CORE-CLIMAX .

An **automated** system is taken to be one which can be run by carrying out a defined set of instructions, with reproducible results.

**NetCDF** refers to the container file format described in public domain documentation , and **CF conventions** are standards for expressing climate and forecast data within NetCDF containers.

The **CEMS JASMIN** architecture refers to the computing cluster made available to the project by STFC .

The **CEDA archive** is the structure provided by STFC through which products can be disseminated .

Interoperability with the Earth System Grid Federation (**ESGF**) refers to the standardised cataloguing system which can serve as an index of the CEDA archive, together with the process of retrieving the data from the CEDA archive in a manner consistent with other ESGF datasets.

**CLIP-C** refers to an ongoing European project which also has the aim of storing climate data in a manner suitable for interoperability with ESGF. Interoperability with CLIP-C implies interoperability with ESGF together with the adoption of conventions used by CLIP-C where applicable.

## Products

The system produces the following two products. The descriptions are taken from the EUSTACE description of action .This document defines short names for the sub-parts of the system which produce each product, in order to facilitate concise documentation and computer code.

|  |  |
| --- | --- |
| **Description** | **System sub-part** |
| Surface air temperature estimates (with estimates of uncertainty) for all surfaces of Earth, derived from satellite surface skin temperature retrievals | **satstace** |
| Globally-complete daily analyses of surface air temperature (with estimates of uncertainty) for the whole of Earth since 1850, based on combined information from satellite and in situ data sources | **fullstace** |

# Schematics

Satellite sources

In situ sources

Relationship model parameters

Infill

Relationship models (satellite)

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

Globally-complete daily analyses of surface air temperature (with estimates of uncertainty) for the whole of Earth since 1850, based on combined information from satellite and in situ data sources

satstace quality check

fullstace quality check

Preprocess in situ

Preprocess satellite

Consistent set of air temperature fields (satellite)

Consistent set of air temperature fields  
(in situ)

Product generation

**satstace**

**fullstace**

Relationship model covariates

Figure : Top-level system schematic

Land surface air temperature

Global break detection  
(UBERN)

Global land station data with break detection (UBERN)

Sea surface air temperature

Pre-process sea surface air temperature  
(Met Office)

Consistent set of air temperature fields (in situ)

In situ sources

Quality-controlled sea surface air temperature (Met Office)

Satellite sources

Figure : Detail of in situ processing

Sea surface temperature from satellite

Append uncertainty information  
(UREAD)

Consistent set of air temperature fields (satellite)

Land surface temperature from satellite

Ice surface temperature from satellite

Lake surface temperature from satellite

Append uncertainty information  
(ULEIC)

Append uncertainty information  
(DMI)

Append uncertainty information  
(UREAD)

Relationship model  
 sea surface  
(Met Office)

Relationship model   
ice  
(DMI)

Relationship model  
lakes  
(UREAD)

Relationship model  
 land surface  
(Met Office)

Figure : Detail of satellite processing

A top-level system schematic is shown in Figure 1. To satisfy the requirements as set out in the EUSTACE grant agreement , the system takes as input a number of satellite and in situ sources and creates the products. In addition, a number of model parameters may be required as inputs, these having been generated during the earlier relationship-building steps in project work package 1 (WP1). Satellite data is pre-processed to include uncertainty information and any necessary reformatting or resampling. Together with model parameters, these are processed by a relationship model to produce air temperature fields derived from satellite data. satstace involves a step to produce products directly from these fields. fullstace applies pre-processing and relationship modelling to in situ sources and combines these with satellite-derived air temperature fields via an infilling step.

The ‘pre-process in situ’ and ‘relationship models (in situ)’ steps necessarily involve the detailed structure set out in Figure 2 in order for the work to be partitioned into the tasks set out in the description of action. Similarly, the requisite detailed structure for ‘pre-process satellite’ and ‘relationship models (satellite)’ is shown in Figure 3. The organisations from which the processing methods originate are indicated.

# Essential requirements

The following are essential technical features, cross-referenced against the documents which introduce them. DA refers to the Description of Action in Annex 1 of the EUSTACE grant agreement . MS30 refers to the project report on user requirements based on first user requirements workshop and literature review (WP4), and D4.1 refers to the resulting user requirement specification for product design. KO indicates minutes of the kick-off meeting. DR refers to the decision register maintained on the EUSTACE Wiki. GA1 indicates minutes of the first general assembly of the project.

## Product requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Feature** | **Reference** | **Defining text** |
| PRDREQ001 | Output variable is surface air temperature | DA | “...enable new Climate Data Records of the surface **air** temperature Essential Climate Variable (ECV) to be created,...” |
| PRDREQ002 | Temporal quantisation is daily or better | DA | “...for every day since 1850.” |
| PRDREQ003 | Spatial quantisation is 0.25 degrees or better (latitude, longitude) | D4.1 | “0.25 degrees latitude by longitude” |
| PRDREQ004 | Time period for fullstace output is from 01/01/1850. | DA | “...for every day since 1850.” |
| PRDREQ005 | Outputs include daily maximum, minimum, and mean surface air temperature | D4.1 | “Temperature variable: Mean, Minimum, Maximum” |
| PRDREQ006 | Uncertainty information must be provided | DA, MS30, D4.1 | DA: “report new, consistent, validated estimates of uncertainty...”  MS30: “All stakeholders recognize that information on uncertainties is important...”  D4.1: “Within box variability” |
| PRDREQ007 | Output in NetCDF format | DA | “The data will be stored in the Climate and Forecasts Conventions for NetCDF...” |
| PRDREQ008 | Compliant with CF-netCDF metadata conventions | DA | “The data will be stored in the Climate and Forecasts Conventions for NetCDF...” |
| PRDREQ009 | Interoperable with ESGF | DA | “The products will also be published to the Earth System Grid Federation (ESGF) federated data catalogue...” |
| ~~PRDREQ010~~ |  |  |  |
| PRDREQ011 | Format suitable for ingestion into CEMS-Academic archive by CEDA staff. | DA | “The final products, destined for wider dissemination, will be  ingested into the CEMS-Academic archive by CEDA staff.” |
| PRDREQ012 | Output is for all locations over all surfaces of Earth | DA | “...for all locations over all surfaces of Earth (i.e. land, ocean, ice and lakes),...” |
| PRDREQ013 | Output uses CF-netCDF standard name conventions. | DA | “The data will be stored...using the CF ‘standard name’ conventions to describe the phenomena being measured/simulated.” |
| PRDREQ014 | ESGF interoperability will be according to applicable conventions from CLIP-C | DA | “EUSTACE will make use of systems created under the FP7 CLIP-C project to  develop an approach to add observation-based climate data sets to the ESGF infrastructure.” |
| PRDREQ015 | The sampling period for daily records at each location is from midnight to midnight measured in local solar time for the location. | GA1 | *Minutes not yet available.* |

## System requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Requirement** | **Reference** | **Defining text** |
| SYSREQ001 | Executes on CEMS JASMIN architecture. | DA | “...all development work within EUSTACE will be undertaken on  the same IT platform (STFC’s CEMS platform)...” |
| SYSREQ002 | Runs sufficiently quickly to allow product generation and quality checks during the 12 months M30-M42 as described in task T2.5. | DA | See work package 2 tasks on page 17. |
| SYSREQ003 | Code is kept under version control. | DA,KO | DA: “...all code will be secured within a central version-  controlled code repository.”  KO: “people will develop code locally and then check code in and out.” |
| SYSREQ004 | Code is sufficiently correct for cross-validation checks given in DA | DA | “These verified and quality  assured products will be passed on to WP3 for validation.” |
| SYSREQ005 | Automated | DA | “EUSTACE will develop an automated system...” |
| SYSREQ006 | Sustainable | DA | “...that is capable of being sustained  after its lifetime.”  “...will have the potential to be used to  provide updates to the data set.” |
| SYSREQ007 | Appropriate maturity for potential production of products beyond the lifetime of the project. | DA | “...an appropriate level of maturity for the potential production of  the products beyond the lifetime of the project.” |

## Requirements for quality check

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Feature** | **Reference** | **Defining text** |
| QTYREQ001 | Includes a check that products (files and collections) comply with specification | DA | “Development, with detailed contribution from partner institutes, of a specification for quality checking key data sets  generated by the project. Building on work carried out for the CMIP5 and CORDEX projects, we will implement a  software suite that checks both files and collections comply with the quality specification.” |

# Desirable features

These features are known to be desirable by users though not absolute requirements.

|  |  |  |
| --- | --- | --- |
| **ID** | **Feature** | **Reference** |
| OPT001 | Code is suitable for open source release at the end of the project. | DA |
| OPT002 | Products are kept under version control | MS30 |
| OPT003 | Products are as small as possible to facilitate quick download | MS30 |
| OPT004 | System can be deployed in operational mode for regular update | MS30 |
| OPT005 | Products contain representation of variability | MS30 |
| OPT006 | Products contain representation of extremes | MS30 |
| OPT007 | Format facilitates extraction of sub-periods | MS30 |
| OPT008 | Format facilitates extraction of sub-regions | MS30 |
| OPT009 | Format facilitates extraction in ASCII format | MS30 |
| OPT010 | Format facilitates extraction to MS Excel | MS30 |
| OPT011 | Format suitable for delivery through Climate-Adapt | MS30 |
| OPT012 | Format suitable for delivery through Copernicus | MS30 |

# References

|  |  |
| --- | --- |
| [1] | European Commission, *Grant agreement number: 640171 — EUSTACE — H2020-EO-2014, Annex 1 - Description of Action,* 2014. |
| [2] | EUMETSAT, “CORE-CLIMAX Climate Data Record Assessment Instruction Manual,” Darmstadt, Germany, 2013. |
| [3] | University Corporation for Atmospheric Research, “NetCDF Documentation,” [Online]. Available: http://www.unidata.ucar.edu/software/netcdf/docs/index.html. [Accessed 06 10 2015]. |
| [4] | B. Eaton, J. Gregory, B. Drach, K. Taylor and S. Hankin, “NetCDF Climate and Forecast (CF) Metadata Conventions: Version 1.6,” 2011. [Online]. Available: http://cfconventions.org/Data/cf-conventions/cf-conventions-1.6/build/cf-conventions.pdf. [Accessed 6 10 2015]. |
| [5] | Science and Technology Facilities Council, Harwell, UK, “What is JASMIN,” [Online]. Available: http://www.jasmin.ac.uk/what-is-jasmin/. [Accessed 11 01 2016]. |
| [6] | Science and Technology Facilities Council, Harwell, UK, “CEDA Archive Users Guide,” [Online]. Available: http://www.ceda.ac.uk/help/users-guide/. [Accessed 11 01 2016]. |
| [7] | University Corporation for Atmospheric Research, Boulder, Colorado, USA, “About the Earth System Grid,” [Online]. Available: https://www.earthsystemgrid.org/about/overview.htm. [Accessed 11 01 2016]. |