MOTECUSOMA

MOnitoring The Energy Cycle for a better UnderStanding Of cliMate chAnge

CCI Cross-ECV project

Fit-For-Purpose Report

	Name	Organisation	Date	Visa
Written by :	Consortium	Magellium	23/03/2025	
Checked by :	Michaël Ablain	Magellium	23/03/2025	
Approved by :	Joël Dorandeu	Magellium	23/03/2025	
Accepted by:	Sarah Connors	ESA		Digitally signed by Sarah Connors Date: 2025.04.14 15:48:30 +01'00'

Document reference :	MOTECUSOMA-DT-006-MAG_F4PR_D2.1/D2.2
Edition.Revision :	1.0
Date Issued:	23/03/2025
Customer :	ESA
Ref. Market, consultation :	ESA/AO/1-12062/23/I-NB

Distribution List

	Name	Organisation	No. copies
Sent to :	Sarah Connors	ESA	1 (digital copy)
Internal copy :	Project folder	Magellium	1 (digital copy)

Document evolution sheet

Ed.	Rev.	Date	Purpose evolution	Comments
1	0	23/10/2024	Creation of document	

Contents

1 Introduction

1.1 Scope and objective

This document is the report describing the results of the assessment of the input data (see ECV inventory (D1.2)) identified for the x-ECV application in terms of scientific quality, consistency, and usefulness for climate science and the description of the database of the project (D2.2).

1.2 Document structure

In addition to this introduction, the document is organised as follows:

- Section 2: Fit for Purpose analysis
- Section 3: Project DataBase description

1.3 Related documents

1.3.1 Applicable documents

Table 1 *List of applicable documents.*

Id.	Ref.	Description
[AD1]		ESA-EOP-SC-AMT-2023-21 : Call to tender CLIMATE-SPACE - THEME II: CROSS-ECV ACTIVITIES
[AD2]	MAG-24-PTF-023	Detailed MAGELLIUM offer in response to ESA/ECSAT Request for Quotation "CCI CROSS ECV" AO/1-12062/23/I-NB

1.3.2 References

- Brown, S., Willis, J., and Fournier, S.: Jason-3 Wet Path Delay Correction (Ver. F. PO.DAAC), https://doi.org/10.5067/J3L2G-PDCOR, 2023.
- Cropper, T. E., Berry, D. I., Cornes, R. C., and Kent, E. C.: Quantifying Daytime Heating Biases in Marine Air Temperature Observations from Ships, https://doi.org/10.1175/JTECH-D-22-0080.1, 2023.
- Embury, O., Merchant, C. J., Good, S. A., Rayner, N. A., Høyer, J. L., Atkinson, C., Block, T., Alerskans, E., Pearson, K. J., Worsfold, M., McCarroll, N., and Donlon, C.: Satellite-based time-series of sea-surface temperature since 1980 for climate applications, Sci. Data, 11, 326, https://doi.org/10.1038/s41597-024-03147-w, 2024.
- Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L., Gloeden, W., Ji, Z., Lawrimore, J., Rayner, N. A., Rosenhagen, G., and Smith, S. R.: ICOADS Release 3.0: a major update to the historical marine climate record, Int. J. Climatol., 37, 2211–2232, https://doi.org/10.1002/joc.4775, 2017.
- Legeais, J.-F., Meyssignac, B., Faugère, Y., Guerou, A., Ablain, M., Pujol, M.-I., Dufau, C., and Dibarboure, G.: Copernicus Sea Level Space Observations: A Basis for Assessing Mitigation and Developing Adaptation Strategies to Sea Level Rise, Front. Mar. Sci., 8, 2021.
- Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T.: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget, https://doi.org/10.1175/2008JCLI2637.1, 2009.
- Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N.-J., Amory, C., van den Broeke, M. R., Horwath, M., Joughin, I., King, M. D., Krinner, G., Nowicki, S., Payne, A. J., Rignot, E., Scambos, T., Simon, K. M., Smith, B. E., Sørensen, L. S., Velicogna, I., Whitehouse, P. L., A, G., Agosta, C., Ahlstrøm, A. P., Blazquez, A., Colgan, W., Engdahl, M. E., Fettweis, X., Forsberg, R., Gallée, H., Gardner, A., Gilbert, L., Gourmelen, N., Groh, A., Gunter, B. C., Harig, C., Helm, V., Khan, S. A., Kittel, C., Konrad, H., Langen, P. L., Lecavalier, B. S., Liang, C.-C., Loomis, B. D., McMillan, M., Melini, D., Mernild, S. H., Mottram, R., Mouginot, J., Nilsson, J., Noël, B., Pattle, M. E., Peltier, W. R., Pie, N., Roca, M., Sasgen, I., Save, H. V., Seo, K.-W., Scheuchl, B., Schrama, E. J. O., Schröder, L., Simonsen, S. B., Slater, T., Spada, G., Sutterley, T. C., Vishwakarma, B. D., van Wessem, J. M., Wiese, D., van der Wal, W., and Wouters, B.: Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020, Earth Syst. Sci. Data, 15, 1597–1616, https://doi.org/10.5194/essd-15-1597-2023, 2023a.
- Otosaka, I. N., Horwath, M., Mottram, R., and Nowicki, S.: Mass Balances of the Antarctic and Greenland Ice Sheets Monitored from Space, Surv. Geophys., 44, 1615–1652, https://doi.org/10.1007/s10712-023-09795-8, 2023b.
- Schröder, M., Danne, O., Falk, U., Niedorf, A., Preusker, R., Trent, T., Brockmann, C., Fischer, J., Hegglin, M., Hollmann, R., and Pinnock, S.: A combined high resolution global TCWV product from microwave and near infrared imagers - COMBI (3.1), https://doi.org/10.5676/EUM_SAF_CM/COMBI/V001, 2023.
- von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast, G., Adusumilli, S., Straneo, F., Ablain, M., Allan, R. P., Barker, P. M., Beltrami, H., Blazquez, A., Boyer, T., Cheng, L., Church, J., Desbruyeres, D., Dolman, H., Domingues, C. M., García-García, A., Giglio, D., Gilson, J. E., Gorfer, M., Haimberger, L., Hakuba, M. Z., Hendricks, S., Hosoda, S., Johnson, G. C., Killick, R., King, B., Kolodziejczyk, N., Korosov, A., Krinner, G., Kuusela, M., Landerer, F. W., Langer, M., Lavergne, T., Lawrence, I., Li, Y., Lyman, J., Marti, F., Marzeion, B., Mayer, M., MacDougall, A. H.,

McDougall, T., Monselesan, D. P., Nitzbon, J., Otosaka, I., Peng, J., Purkey, S., Roemmich, D., Sato, K., Sato, K., Savita, A., Schweiger, A., Shepherd, A., Seneviratne, S. I., Simons, L., Slater, D. A., Slater, T., Steiner, A. K., Suga, T., Szekely, T., Thiery, W., Timmermans, M.-L., Vanderkelen, I., Wjiffels, S. E., Wu, T., and Zemp, M.: Heat stored in the Earth system 1960–2020: where does the energy go?, Earth Syst. Sci. Data, 15, 1675–1709, https://doi.org/10.5194/essd-15-1675-2023, 2023.

1.4 Acronyms

Table 2 *List of abbreviations and acronyms.*

Synonym	Description
CEOS	Committee on Earth Observation Satellites
CL	Confidence Level
ECV	Essential Climate Variable
EEI	Earth's Energy Imbalance
EO	Earth Observation
FDRs	Fundamental Data Records
FIDUCEO	Fidelity and uncertainty in climate data records from Earth observation
FRMs	Fiducial Reference Measurements
GEOSS	Global Earth Observation System of Systems
GIOMAS	Global Ice-Ocean Modeling and Assimilation System
GMSL	Global Mean Sea Level
IPCC	Intergovernmental Panel on Climate Change
KPI	Key Performance Indicator
LRM	Low Resolution Mode
MSL	Mean Sea Level
NMI	National Metrology Institute
PDAP	Payload Data Acquisition and Processing
PDO	Pacific Decadal Oscillation
PTR	Point Target Response
QA4EO	Quality Assurance Framework for Earth Observation
SAR	Synthetic Aperture Radar
SI	International System of Units
SLA	Sea Level Anomaly

SLBC2	Phase 2 of the ESA Sea Level Budget Closure climate change initiative project
SLR	Sea Level Rise
SLR-SUB	Sea Level Rise Stability Uncertainty Budget
SROCC	Special Report on the Ocean and Cryosphere
SSB	Sea State Bias
SSH	Sea Surface Height
SWH	Significant Wave Height
S6-MF	Sentinel-6 Michael Freilich
S6-NG	Sentinel-6 Next Generation
S6PP	Sentinel-6 Processing Prototype
USO	Ultra Stable Oscillator

2 Fit for Purpose Analysis

2.1 Earth Energy Imbalance

2.1.1 Ocean Heat Content (OHC)

2.1.1.1 Absolute Sea Level

2.1.1.1.1 Dataset Strengths

Absolute Sea Level (SL) changes at regional scales are derived from the sea-level products operationally generated by the Copernicus Climate Change Service (C3S). This dataset, fully described in (Legeais et al., 2021) is dedicated to the sea level stability for climate applications. It provides daily sea-level anomalies grids based at any time on a reference altimeter mission (TopEx/Poseidon, Jason-1,2,3 and S6-MF) plus complementary missions (ERS-1,2, Envisat, Cryosat, SARAL/Altika, Sentinel-3A) to increase spatial coverage.

2.1.1.1.2 Dataset limitations

Dataset spatial coverage is limited up to 66° latitude since above this latitude, most of the ocean is under ice during several months and cannot be measured by altimetry. These data can also be affected by errors like any spatial measurements and some corrections need to be applied (correction of TOPEX-A, Jason-3 Wet Tropospheric Correction (WTC)).

2.1.1.1.3 Expected use within database

Thanks to the large spatial coverage of the global ocean and its high temporal resolution (monthly), it will allow to provide OHC estimates with high temporal and spatial resolution. Also, the current limitations identified such as the errors have already been identified by the community and some corrections are already available (e.g. correction to be applied to correct for the drift of the TOPEX-A altimeter provided by C3S or correction from Brown et al. (2023) for Jason-3 WTC drift).

2.1.1.1.4 Key interactions with other ECVs in the database

This dataset will be used combined with manometric sea level estimates from gravimetry to obtain OHC estimates. However, an effort will be made to homogenize the corrections that must be applied on both dataset (e.g. Glacial Isostatic Adjustment (GIA), correction due to atmosphere) do avoid any aliasing that could be coming from using different corrections from these 2 datasets.

2.1.1.2 Manometric sea level

2.1.1.2.1 Dataset Strengths

Manometric sea level change is derived from space gravimetry measurements of GRACE and GRACE-FO missions. Error sources in GRACE products are mainly due to ancillary data used to correct the geocenter motion, the glacial isostatic adjustment in response to the melting of paleo ice sheets, and the Earth oblateness (Blazquez et al., 2018). An ensemble approach made of different GRACE products is used to characterize these errors.

2.1.1.2.2 Dataset limitations

Gravimetry measurements provide estimates of the Manometric Sea Level with a spatial resolution around 300 kilometers and a temporal resolution on a monthly basis. It must be corrected from several errors that affect this observing system such as earthquakes, effect of leakage on the ocean on the gravimetry measurements which happens in areas of important ice melting.

2.1.1.2.3 Expected use within database

Thanks to the large spatial coverage of the global ocean and its high temporal resolution (monthly), it will allow to provide OHC estimates with high temporal and spatial resolution. Also, the corrections to make regarding dataset limitations will be implemented (leakage and earthquake corrections).

2.1.1.2.4 Key interactions with other ECVs in the database

This dataset will be used combined with absolute sea level estimates from altimetry to obtain OHC estimates. However, an effort will be made to homogenize the corrections that must be applied on both dataset (e.g. Glacial Isostatic Adjustment (GIA), correction due to atmosphere) do avoid any aliasing that could be coming from using different corrections from these 2 datasets.

2.1.2 Ice sheet

2.1.2.1 Dataset Strengths

The latest assessment of the ESA/NASA Ice-sheet Mass Balance Intercomparison Exercise (IMBIE-2_R2, (Otosaka et al., 2023b) compiled a total of 27 contributed mass balance assessment for GIS (23 for the AIS), comprising assessments from each of the major approach (altimetry, gravimetry, IOM). This aggregated assessment was an instrumental source for the IPCC AR6 assessment of ice sheets' sea-level contribution.

2.1.2.2 Dataset limitations

Some limitations are discussed by Otosaka et al. (2023a).

There is some ambiguity about the inclusion or exclusion of peripheral glaciers, which are typically excluded in altimetry and IOM assessments, but included, or poorly separated, in gravimetry assessments.
While the formal temporal resolution of the aggregated dataset is monthly, the effective temporal resolution of some of the input assessments (notably from altimetry) is coarser, so that the time series are unlikely to reflect the full month-to-month variability of mass balance.

2.1.2.3 Expected use within database

Thanks to the high temporal resolution (monthly), it will allow to provide OHC estimates with high spatial resolution combined with altimetry and gravimetry. The inclusion of peripheral glaciers will be assessed but has a negligible impact on the EEI estimation compared to the other components (OHC, AHC).

2.1.2.4 Key interactions with other ECVs in the database

This dataset will be used with OHC estimates based on space data (combination of altimetry and gravimetry) to estimate the total EEI (OHC represents 91% of the total EEI) while Ice Sheet contribution is close to 3% (von Schuckmann et al., 2023).

2.1.3 Atmosphere Heat Content: latent heat contribution (ESA CCI TCWV)

2.1.3.1 Dataset Strengths

The total column water vapour (TCWV) COMBI product by the Satellite Application Facility on Climate Monitoring (CM-SAF) and ESA-CCI is a dedicated satellite-based product to provide TCWV with global coverage. Specific care has been taken for best accuracy and temporal stability and as such it will serve as a benchmark for reanalysis data.

2.1.3.2 Dataset limitations

The product covers July 2002- December 2017 and thus less than the envisaged study period within the project. However, this period will be sufficient for intercomparison. The product cannot be viewed as independent reference for reanalyses as those assimilate at least some of the data used for the satellite TCWV product. The product quality documentation (Schröder et al., 2023) points out some (minor) break points which need to be kept in mind.

2.1.3.3 Expected use within database

ESA-CCI TCWV data will be used to evaluate TCWV in reanalyses, which have been shown to exhibit relatively larger inter-product spread compared to atmospheric total energy (von Schuckmann et al., 2023).

2.1.3.4 Key interactions with other ECVs in the database

This dataset will not have direct interaction with other ECVs in the database, but TCWV is an important contributor to atmospheric heat content.

2.1.4 Earth Energy Imbalance (CERES)

2.1.4.1 Dataset Strengths

Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data (Loeb et al., 2009) provide estimates of the global and regional Earth's energy imbalance at the top of the atmosphere (TOA). The CERES Energy Balanced and Filled (EBAF) product [14] is used as a reference for the EEI variability assessment because it is totally independent, and it is known to reproduce precisely the EEI variations with uncertainties of the order of a few tenths of W m−2.

2.1.4.2 Dataset limitations

CERES EEI mean value estimation can be used since it does not accurately estimate it regarding its design. Therefore, the CERES EEI mean value is anchored with an in situ product (Lyman and Johnson, 2014).

2.1.4.3 Expected use within database

CERES EBAF product will be used as a reference for comparisons with EEI estimated with space geodetic data (combination of altimetry and gravimetry) in terms of variability. Thanks to the high temporal resolution (monthly) of the EEI with CERES EBAF dataset, comparisons will be performed close to the annual time scale between the 2 products.

2.1.4.4 Key interactions with other ECVs in the database

This dataset will not have direct interaction with other ECVs in the database since it will be only used for comparisons activities as an purely independent estimate of the EEI.

2.2 Surface Air Temperature

2.2.1 Sea Surface Temperature

2.2.1.1 Sea surface temperature (CCI analysis product and L3)

2.2.1.1.1 Dataset Strengths

The SST CCI analysis is a daily gap-free analysis covering 1980 to 2021, and is extended in time with an interim climate data record to about 1 month behind the present (under Copernicus and UK national funding for the Earth Observation Climate Information Service). Observations from twenty infrared and two microwave radiometers are used, and, uniquely among SST products, these are adjusted for their differing times of day of measurement to avoid aliasing of diurnal variations into long-term trends, enhancing observational stability. The analysis has an explicit SST measurand (daily mean SST at 20 cm depth) which is compatible with use of SST CCI with centennial scale climate data records. It is extensively validated. Comparison with other multi-decadal SST analyses, particularly “OISST”, suggest the CCI SSTs are considerably more stable (Embury et al., 2024).

. The analysis stability and low uncertainty reflects the properties of the level 3 SST products (non gap-filled, but adjusted to SST depth and standardised local time) that are input to the analysis system.

2.2.1.1.2 Dataset limitations

The analysis has large uncertainties in certain periods during 1980 to 1982 when available satellite data are minimal or not-yet existent (these gaps will need to be addressed by exploiting non-traditional SST sensors in a future version). During the target period of this study, instrumental instability and problems causes some increased uncertainty and instability around the year 2000 (which will be addressed in a future version by including MODIS observations).

2.2.1.1.3 Expected use within database

The SST CCI analysis will be used to inform near-surface atmospheric stability over the ocean in combination with ERA5 winds and marine TAS estimates. The gap free nature of the product is exploited here to enable derivation of stability for every location and therefore every match of satellite SST and in situ used in relationship building.

Marine TAS inputs derived from SST-TAS relationships applied to SST CCI L3 supercollated SSTs (“L3S”). The supercollated (multisensor combined) L3 dataset will be constructed from the per-sensor SST CCI L3 datasets for this project specifically. This L3S will enable good marine TAS coverage, but will not be gap free. This is a change of choice compared to the originally proposed use of the SST CCI analysis, this change being made to avoid the TAS analysis ingesting already-analysed (gap-filled) data, which is necessary for the TAS analysis uncertainty estimates to be valid. In other words, only the TAS analysis system will do gap filling

2.2.1.1.4 Key interactions with other ECVs in the database

Marine TAS from this dataset will use relationships built against in situ air temperatures from ICOADS (see below).

2.2.2 Sea Ice Surface Temperature

2.2.2.1 Sea and sea-ice surface temperature (C3S analysis product)

2.2.2.1.1 Dataset Strengths

Sea ice surface temperatures are obtained from the global satellite-based sea and sea-ice surface temperature climate data record developed by DMI within C3S. One of the key strengths of the dataset is that it provides consistent surface temperatures over open ocean, sea ice and the marginal ice zone. The dataset is a daily gap-free (L4) dataset, which enables relationship building not only in clear-skies, but also in cloud covered regions, in contrast to existing L3 IST datasets such as AASTI/C3S and the LST_cci. The use of L4 data results in more in situ matchups enabling more robust relationships to be built. The validation results against in situ observations are stable and the accuracy similar to existing L3 IST datasets.

2.2.2.1.2 Dataset limitations

The satellite coverage over the Arctic sea ice suffers from frequent and persistent clouds in particular during the summer, which means that the L4 ISTs sometimes are based on very limited satellite observations. Another uncertainty in satellite-based IST datasets arises due to insufficient cloud masking. Cloud masking is very challenging over sea ice and in particular during winter, which usually leads to undetected clouds and a cold bias in the IST.

2.2.2.1.3 Expected use within database

The C3S L4 SST/IST data will be used as the main predictor of TAS over sea ice. The conversion from L4 IST to TAS will be based on a machine learning model trained against all-sky in situ observations. It is expected that the training against all-sky in situ observations will reduce the sampling uncertainties and cold-bias in the TAS (compared to the L4 IST) when also including information on cloud masking and time of the year.

2.2.2.1.4 Key interactions with other ECVs in the database

The C3S L4 SST/IST data used over sea ice and in the marginal ice zone meets the SST ECV over open ocean. This requires a fairly good consistency between the IST, SST and sea ice concentration fields that will most likely be used to identify the different surface types (sea ice, marginal ice zone and open ocean).

2.2.3 Ice Sheet Surface Temperature

2.2.3.0.1 Dataset Strengths

Surface temperatures for the Arctic and Antarctic Ice Sheets, as well as ice shelves, will be based on a gap-free (L4) IST product at around 0.01-0.02 degrees resolution. The underlying algorithm is an optimal interpolation scheme (similar to the one used to generate the L4 C3S ISTs over sea ice) that combines multiple satellite observations to provide daily gap-free fields of IST including uncertainties. The product will be generated during the first phase of MOTECUSOMA and the input data will be the LST_cci data (possibly in combination with the C3S/AASTI IST). The first step will be to compare the LST_cci and C3S/AASTI IST products in terms of coverage, stability and accuracy over the ice sheets. Similar to sea ice, the use of daily gap-free (L4) data enables relationship building in cloud covered regions as well when trained against all-sky in situ observations.

2.2.3.0.2 Dataset limitations

A key limitation is that the L4 datasets are not produced yet. The production will soon start for the Greenland Ice Sheet and will be followed for the Antarctic Ice Sheets and shelves. For the same reason, the coverage, accuracy and stability are not determined yet and a detailed analysis and assessment are required in order to do so.

2.2.3.0.3 Expected use within database

The L4 Ice Sheet IST datasets will be the baseline for the TAS developments over the Arctic and Antarctic ice sheets and ice shelves. The high resolution, expected to be 0.1 degrees (at least for the Greenland Ice Sheet), reduces the point to pixel uncertainty and allows a more realistic relationship to be built. The training against all-sky in situ observations is expected to reduce the uncertainties related to the sampling effect and the satellite effect.

2.2.3.0.4 Key interactions with other ECVs in the database

The L4 Ice Sheet IST datasets have shared boundaries with the land surface temperature (ice free regions of Greenland and Antarctica) and in the case of the ice shelves these have shared boundaries with ocean regions which can be both sea ice covered or ice free.

2.2.4 Land Surface Temperature

2.2.4.1 Land surface temperature (L3C)

2.2.0.1.1 Dataset Strengths

For all land surface temperatures including snow covered regions and lakes, the most appropriate ESA CCI land surface temperature (LST) products will be used. These data are global day/night L3C products with a nominal resolution of 0.01 degrees. The underlying algorithm is a generalised split window approach with a complete uncertainty breakdown by correlation type and length. Both Terra-MODIS and Aqua-MODIS LST_cci products will be used to provide sub-daily information on the diurnal cycle, with the LST_cci long-term Infrared Climate Data Record (IRCDR) used to provide data back into the 1990s from the ATSR series. Lake data will come from the LST_cci L2P product, as this is not retained through to L3C. A variant of the LST_cci L3C will actually be produced, which is identical to the standard product, but just with the inclusion of lakes data (which is currently only retained to L2P).

2.2.4.1.1 Dataset limitations

The main limitation of these datasets are that they are derived from thermal infrared data which are not able to penetrate clouds. All data have been cloud masked at the L2P stage and thus carried forward to the L3C output. Nevertheless, cloud contamination is still an issue even with the strictest of cloud masks. This can be more problematic during the night, where less information is available to determine the cloud mask.

2.2.4.1.2 Expected use within database

The LST datasets will be the baseline for the TAS developments over land and lakes. Relationships will be built at 0.05 degrees, which optimises the trade-off between reducing the point to pixel uncertainties and mining of huge data volumes. It is expected this will allow relationships to be built based on climatic and land cover classes amongst other predictors.

2.2.4.1.3 Key interactions with other ECVs in the database

The LST datasets have shared boundaries with the IST datasets over the Greenland and Antarctic ice sheets. While LST data over all surface types will be used in the relationship building, a shared mask will be applied which switches to the DMI L4 datasets over the ice sheets. There will also be shared boundaries with the sea-ice surface temperature datasets and sea surface temperature datasets at the coasts, whereby the land-sea mask will mark the boundary.

2.3 Models and re-analyses

2.3.1 ECMWF Reanalysis (ERA5)

2.3.1.0.1 Dataset Strengths

ERA5 offers highly resolved (~31 km and hourly) data with globally completed coverage. The reanalysis assimilates vast amounts of observational data from satellites, weather stations and buoys. A particular advantage of reanalyses such as ERA5 is the generation of a suite of variables that are dynamically consistent.

2.3.1.0.2 Dataset limitations

Tje main disadvantage of ERA5 in the context of this project is that the reanalysis does not directly assimilate marine TAS values. As such these values are derived from only one parameterisation of air-sea interaction. Furthermore, ERA5 may be susceptible to spatial/temporal variability in data coverage. It is noteworthy that u, v, T, q (used to compute AHC and TCWV) observations (away from the surface) are assimilated in ERA5 and thus are more strongly constrained by observations (depending on observation density).

2.3.1.0.3 Expected use within database

The wind speed and direction values will be used to derive the Marine TAS - SST relationships and to bias-adjust the in situ marine TAS values. Atmospheric state quantities u, v, T, q will be used to evaluate AHC and TCWV.

2.3.1.0.4 Key interactions with other ECVs in the database

Atmospheric stability estimates over the ocean will be developed using ERA5 wind data, L3 SST data (see section 2.2.1) and the in situ ICOADS marine TAS values (section 2.4.1). TCWV from reanalysis represents an important contributor to AHC, but there is no interaction with other ECVs.

2.3.2 Japanese Reanalysis for Three Quarters of a Century (JRA3Q)

2.3.2.0.1 Dataset Strengths

Similar to ERA5, except that JRA3Q exhibits slightly lower resolution (~40 km) but a newer vintage of the underlying model used to generate the reanalysis (JRA3Q: 2023, ERA5: 2020).

2.3.2.0.2 Dataset limitations

Similar to AHC/TCWV-related considerations for ERA5

2.3.2.0.3 Expected use within database

Atmospheric state quantities u, v, T, q will be used to evaluate AHC and TCWV.

2.3.2.0.4 Key interactions with other ECVs in the database

TCWV from reanalysis represents an important contributor to AHC, but there is no interaction with other ECVs.

2.3.3 Modern-Era Retrospective analysis for Research and Applications (MERRA-2)

2.3.3.0.1 Dataset Strengths

Similar to ERA5, except that MERRA-2 exhibits lower resolution (~50 km) and older vintage of the underlying model used to generate the reanalysis (2017). Next edition of MERRA (MERRA-21C) is expected to become available within project duration.

2.3.3.0.2 Dataset limitations

Similar to AHC/TCWV-related considerations for ERA5

2.3.3.0.3 Expected use within database

Atmospheric state quantities u, v, T, q will be used to evaluate AHC and TCWV. 2.3.3.0.4 Key interactions with other ECVs in the database

TCWV from reanalysis represents an important contributor to AHC, but there is no interaction with other ECVs.

2.4 In-Situ and in-situ only analyses

2.4.1 ICOADS ship/moored buoy data

2.4.1.0.1 Dataset Strengths

The International Comprehensive Ocean-Atmosphere Dataset (ICOADS; (Freeman et al., 2017)) is the most extensive global collection of surface marine data. This freely available dataset includes measurements from a wide range of platforms, including ships, moored and drifting buoys and coastal stations. Marine air temperature observations, primarily from ships and -- to a lesser extent -- buoys, provide the only direct measurements of this variable over the ocean. ICOADS data extend back to the late 17th century and continue to the present day, with monthly and daily updates—primarily sourced from values supplied via the Global Telecommunication System (GTS). Accompanying metadata, such as ship identification numbers and unique observation references (UIDs), support quality control and enable necessary data adjustments.

2.4.1.0.2 Dataset limitations

Marine TAS values in ICOADS are highly discontinuous across space and, to a lesser extent, over time. Ship-based observations are largely confined to major shipping routes, while the moored buoy network is concentrated in the tropics. Ship data require adjustments to account for differences in observing height, typically standardizing measurements to 10 meters above the sea surface. Additionally, a diurnal adjustment is necessary to correct for biases caused by artificial heating from the ship's superstructure. The ICOADS ship data also suffer from significant duplication and positioning errors, necessitating a complex quality-control system to ensure accuracy.

2.4.1.0.3 Expected use within database

The marine TAS data will be directly incorporated into the construction of the gap-free analysis and used to establish SST-TAS relationships over the ocean. To ensure consistency, the data will be adjusted to a common reference height using atmospheric stability estimates. Additionally, a correction based on (Cropper et al., 2023) will be applied to remove the diurnal heating bias. This adjustment is performed relative to a nighttime average and has the effect of also removing the true diurnal cycle. As such these values represent nighttime-equivalent values.

2.4.1.0.4 Key interactions with other ECVs in the database

Atmospheric stability estimates, used for height correction and the development of SST-TAS relationships, will be derived from ERA5 wind data and collocated SST CCI L3 data (see Section 2.2.1.1.3 above).

2.4.2 Sea-ice surface air temperature observations

2.4.2.0.1 Dataset Strengths

The main source of sea-ice surface air temperature observations are drifting buoys/stations and ice mass balance buoys. The availability of these is much larger in the northern hemisphere than the southern hemisphere. The northern hemisphere observation originates mainly from the North Pole drifting stations (established in 1937), a drifting buoy network provided by the International Arctic Buoy Programme, ice mass balance buoys from the US Army Cold Regions Research and Engineering Laboratory (CRREL) and Alfred Wegener Institute (AWI). For the southern hemisphere, the available buoys are mainly from the International Programme for Antarctic Buoys, which was initiated in 1994.

2.4.2.0.2 Dataset limitations

The data is definitely limited in coverage. Great efforts have already been put into collecting, formatting and quality controlling in situ observations over sea ice, and we will continue the efforts within this project to see if any more qualified observations are available. Another limitation is the poor knowledge of what the sensors are actually measuring.

2.4.2.0.3 Expected use within database

The data will be used to build the TAS relationships over sea ice. The data will be quality controlled to ensure that they meet the criteria for relationship building.

2.4.2.0.4 Key interactions with other ECVs in the database

-

2.4.3 Ice Sheet surface air temperature observations

2.4.3.0.1 Dataset Strengths

Historically, surface air temperatures over the ice sheets have been measured from automatic weather stations, usually providing the temperature of 1-3 m height (depending on the snow accumulation and melt). For the Greenland Ice Sheet the largest collection of observations are from the Greenland Climate Network (GC-NET) and Programme for Monitoring of the Greenland Ice Sheet (PROMICE). The observations cover the period from 1990 to present. The stations are year-round and designed to endure the extreme weather conditions. For the Antarctic Ice Sheet, the main source of observations are available from automatic weather stations distributed through the Antarctic Meteorological Research Center (AMRC) and the National Climatic Data Center (NCDC).

2.4.3.0.2 Dataset limitations

The main limitation of both the Greenland and Antarctic in situ datasets is definitely the sparse coverage. The sparse coverage leads to periods and regions that might not be very well represented. As an example, the PROMICE stations are mainly located in the ablation area and therefore not representing the conditions of the accumulation areas very well. The spatial coverage of the automatic weather stations also change over time as new stations are installed and other stations are dropped.

2.4.3.0.3 Expected use within database

The ice sheet in situ observations of TAS will be used along with other information (e.g. on their location and altitude) to best provide the relationship between the ice sheet surface temperatures and corresponding TAS.

2.4.3.0.4 Key interactions with other ECVs in the database

-

2.4.4 Land Surface Air Temperature

2.4.4.1 EUSTACE homogenised station data

2.4.4.1.1 Dataset Strengths

Homogenised time series of daily temperature observations for meteorological stations throughout Europe from the EUSTACE project are available AS daily maximum, minimum and mean temperatures. It uses an automated homogenization procedure, with a system to detect breaks, and a quantile matching technique in combination with a pairwise-comparison approach to determine adjustments.

2.4.4.1.2 Dataset limitations

The main limitation of this dataset is that it is only for Europe. It is though only complementary data to the global GHCN-D.

2.4.4.1.3 Expected use within database

The data will be used to complement the building of the relationships against the LST L3C input data. It is expected this will allow relationships to be built based on climatic and land cover classes amongst other predictors. Relationships will be built for these daily means, with relationships also tested for the daily minimums and maximums.

2.4.4.1.4 Key interactions with other ECVs in the database

The data is one of the relationship building input datasets and will be used in conjunction with the GHCH-D station data.

2.4.4.2 GHCN-d

2.4.4.2.1 Dataset Strengths

The Global Historical Climatology Network daily (GHCNd) is an integrated database of daily climate summaries from land surface stations across the globe. It is made up of daily climate records from numerous sources that have been integrated and subjected to a common suite of quality assurance reviews. The database contains records from 1000s of stations across the whole globe. These data are provided as daily maximum and minimum temperatures.

2.4.4.2.2 Dataset limitations

The main limitation of these datasets are that the station data do tend to be focussed in certain regions of the globe, specifically North America and Europe, with data sparsity over Africa and South America. This means some climate and biome classes may be less well covered.

2.4.4.2.3 Expected use within database

The data will be used to build the relationships against the LST L3C input data. It is expected this will allow relationships to be built based on climatic and land cover classes amongst other predictors. The station data will be converted to daily means, and relationships will be built for these daily means, with relationships also tested for the daily minimums and maximums.

2.4.4.2.4 Key interactions with other ECVs in the database

The data is one of the relationship building input datasets and will be used in conjunction with the EUSTACE station data.

2.4.4.3 Berkeley Global Temperature Data

2.4.4.3.1 Dataset Strengths

Berkeley Earth provides gridded temperature data, which go back to 1850. The newest generation of these products utilise machine learning techniques to improve spatial resolution. We specifically use the Berkeley Global Daily Land Gridded dataset Earth Surface Temperature Anomaly Field, which delivers daily maximum, minimum, and mean temperature anomalies at 1.0 degrees resolution.

2.4.4.3.2 Dataset limitations

The main limitation of these datasets are their relative coarse spatial resolution at 1 degrees, meaning the TAS outputs will need to be upscaled for validation.

2.4.4.3.3 Expected use within database

The data will be used primarily to validate the output of the TAS developments over land and lakes. The data from the relationship building will be re-gridded to the common grid of Berkeley and intercompared.

2.4.4.3.4 Key interactions with other ECVs in the database

The data is one of the validation datasets to assess the quality of the relationship building over land.

3 Project Database description

3.1 Context

The project database will comprise structured collections of data, primarily to facilitate collaborative efforts on the core science tasks of the project (Task-4 activities).

3.2 EEI database

The EEI database will be hosted on the Centre National des Etudes Spatiales (CNES) infrastructure. It provides a High Performance Computing (HPC) facility to store and process with high efficiency any type of data.

A specific project workspace will be available to store the EEI database. This database is composed of the data used to compute all the components of the EEI (OHC, AHC, etc…) such as:

- Absolute sea level data from Altimetry (C3S)
- Manometric sea level data from Gravimetry
- Temperature and Salinity profiles from Argo
- Water Vapour from ESA CCI dataset

- Greenland and Antarctica Ice Sheet from IMBIE dataset

3.3 TAS database

The TAS database will be hosted at the Centre for Environmental Data Analytics (CEDA). It will be a dedicated MOTECUSOMA group workspace containing copies of or (where already present on CEDA) symbolic links to all the datasets specified in Section 2.2.

Additionally, the TAS group workspace will also host new datasets:

- SST CCI L3S
- satellite-derived TAS estimates over each domain (i.e., inputs to TAS analysis)
- TAS analysis outputs (in initial and final versions).