D5_5 Product User Guide – Ensemble Sea and Lake Surface Temperature

Christopher J. Merchant, Jonathan P. D. Mittaz, Michael Taylor and Claire Bulgin

University of Reading

26 August 2019

FIDUCEO has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no. 638822

1 Contents

2 Introduction

2.1 Scope

This document describes the Ensemble Sea and Lake Surface Temperature (ST) climate data record (CDR) created by the FIDUCEO project in August 2019 with version designation v0.20. The released data record is based on the MetOp-A Advanced Very High Resolution Radiometer (AVHRR) “easy” FCDR (fundamental CDR) version 1.0, with additional ST optimisation of the brightness temperature calibration coefficients. The ST CDR is a gridded dataset (level 3 uncollated) at 0.05° latitude-longitude resolution. This product user guide gives:

1. An overview of the specifications of the ST CDR;
2. A high-level description of the implementation of the retrieval processing chain;
3. Information on limitations of this current version of the data record;
4. Technical details on the format and on how to access the data.

2.2 Version Control

Version	Reason	Reviewer	Date of Issue
1.0	Initial version	Phipps	27 August 2019

2.3 Applicable and Reference Documents

- FIDUCEO website, http://www.fiduceo.eu/ or https://research.reading.ac.uk/fiduceo/
- D2.4-e, SST CDR Uncertainty report (see website)
- CF-standards version 1.7, http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html
- AVHRR FCDR PUG, Product user guide, latest version (see website)
- Giering, R., Quast, R., Mittaz, J., Hunt, S., Harris, P., Wooliams, E. and Merchant, C. (2019) A novel framework to harmonise satellite data series for climate applications. *Remote Sensing*.
- Merchant, C. J., Le Borgne, P., Marsouin, A. and Roquet, H. (2008) Optimal estimation of sea surface temperature from split-window observations. Remote Sensing of Environment, 112 (5). pp. 2469-2484. ISSN 0034-4257 doi: https://doi.org/10.1016/j.rse.2007.11.011
- Merchant, C J, S Saux-Picart and J Waller (2019 in review) Bias Correction and Covariance Parameters for Optimal Estimation by Exploiting Matched In-situ References, *Remote Sensing of Environment.*

2.4 Glossary

BT	Brightness Temperature
CDR	Climate Data Record
CEDA	Centre for Environmental Data Archiving
FCDR	Fundamental Climate Data Record
FIDUCEO	Fidelity and Uncertainty in Climate data records for Earth Observation

TCWV	Total Column Water Vapour
NOAA	National Oceanic and Atmospheric Administration
AVHRR	Advanced Very High Resolution Radiometer
ST	Sea and Lake Surface Temperature

3 CDR overview characteristics

General	CDR name	FIDUCEO Ensemble Sea and Lake Surface Temperature CDR
	CDR reference	ENSEMBLE2.0-v02.0-fv01.0
	CDR digital identifier(s)	10.5285/dd63f6f7239f4c1da830950c6e58cfdd
	CDR description	Climate Data Record containing grid-cell instantaneous averages of retrieved surface temperature over ice-free oceans and 300 large lakes
	CDR type	Level 3 Uncollated ST CDR
	CDR period	Dec 2006 – Dec 2018
	CDR satellites	• MetOp-A MetOp-A files in a sun-synchronous near-polar orbit.
	CDR content	For each 0.05° latitude longitude cell the main content is: • Skin surface temperature estimate (sea or lake ST, best estimate) • Depth temperature estimate (20 cm below surface) • ST uncertainty decomposed by correlation properties • ST quality flag (use of QL = 5 is recommended) • Ensemble of 10 perturbations of the ST reflecting uncertainty • Ensemble of 10 perturbations of the QL (mostly no change) • Miscellaneous other auxiliary fields
Instrument	Instrument name	Advanced Very High Resolution Radiometer
	Instrument description	AVHRR is a scanning infra-red radiometer calibrated using an internal calibration target and cold space
Data	Input data	• AVHRR L1 data processed with the FIDUCEO FCDR processor (with minor adaptations for the intermediate radiance ensemble step and ST-optimised calibration coefficients) • ST-optimised calibration coefficients derived to ensure good surface temperature retrieval using method of Merchant et al. (2019) adapted to AVHRR-drifting buoy matches: effective SST reference is therefore the global drifting buoy network

	Output data	Files per orbit or semi-orbit of level 3 uncollated data – i.e., of gridded swath data. Data include: • Surface temperature estimates at skin and 20 cm • Quality level information • Perturbations to ST estimates for each of 10 ensemble members • Perturbations to quality level information for each member
	Format	The data are provided in NetCDF4 format, using the file format conventions of the ESA SST Climate Change Initiative that include standardised latitude, longitude and time information.
Access	CEDA	The data are hosted by CEDA at : https://catalogue.ceda.ac.uk/uuid/dd63f6f7239f4c1da830950c6e58cfdd
	Delivery	Available through CEDA
Resolution	Horizontal	Average of best-available-quality ST in latitude-longitude cells of 0.05° x 0.05° resolution.
	Vertical	Skin temperature and (model-informed) estimate at depth of 20 cm
	Temporal	Instantaneous
Physical Content	CDR physical quantity	The core retrieved quantity is the skin (radiometric) temperature of the Earth's water surfaces (sea and large lakes). This is provided as a best estimate, plus an ensemble of 10 perturbations capturing known uncertainties
	CDR physical description	Around 25 semi-orbital files per record day, each of order 20 MB.
Uncertainty target	Accuracy	Metrologically traceable uncertainty estimates are provided for each grid cell average, plus ensemble members sample the estimated distribution of uncertainty across multiple scales.
	Precision	ST is stored in kelvin with a precision of 0.01 K. Uncertainties are stored with a precision of 0.001 K.
	Stability	Stability of global mean ST is expected to be ~0.05 K/decade. Larger (up to 0.2 K) instability is expected regionally and seasonally.
	Known problems	The uncertainties included in the ST CDR uncertainty and perturbation values are: propagated uncertainty in instrument counts, propagated calibration effects and choice of clear-sky probability threshold for cloud detection.

Data record characteristics	Users interested principally in sea surface temperature without the need for Monte Carlo propagation of uncertainty in their application are recommended to use ESA SST Climate Change Initiative products v2.1. The ST perturbations provided have an ensemble member index from 1 – 10. The ensemble members should be used for the purpose of uncertainty assessment in cases where error correlation cannot be neglected, because of the large spatio-temporal scales of application and/or non-linearity in the downstream processing.

4 Description of AVHRR

Metop-A AVHRR is an AVHRR/3 instrument. NOAA have described the instrument (NOAA-L brochure, text quoted from www.esa.int/our_activities/observing_the_earth/meteorological_missions/metop/about_avhrr_3.

> *The Advanced Very High Resolution Radiometer (AVHRR/3) is one of the complement of American instruments provided by the National Oceanic and Atmospheric Administration (NOAA) to fly on MetOp-A, B and C.*
>
> *The AVHRR/3 scans the Earth surface in six spectral bands in the range of 0.58 - 12.5 microns. It provides day and night imaging of land, water and clouds, measures sea surface temperature, ice, snow and vegetation cover.*
>
> *The AVHRR/3 is a six-channel imaging radiometer that detects energy in the visible and infrared (IR) portions of the electromagnetic spectrum. The instrument measures reflected solar (visible and near-IR) energy and radiated thermal energy from land, sea, clouds, and the intervening atmosphere. The instrument has an instantaneous field-of-view (IFOV) of 1.3 milliradians providing a nominal spatial resolution of 1.1 km (0.69 mi) at nadir. A continuously rotating elliptical scan mirror provides the cross-track scan, scanning the Earth from ± 55.4° from nadir. The mirror scans at six revolutions per second to provide continuous coverage.*
>
> *The instrument provides spectral and gain improvements to the solar visible channels that provide low light energy detection. Channel 3A, at 1.6 microns, provides snow, ice, and cloud discrimination. Channel 3A will be time-shared with the 3.7-micron channel, designated 3B, to provide five channels of continuous data. An external sun shield and an internal baffle have been added to reduce sunlight impingement into the instrument's optical cavity and detectors.*

5 Differences with existing products

The principal alternative CDR for sea surface temperature (SST) is from the ESA SST CCI, which has also generated L3U SST using the same cloud detection and SST retrieval methodology.

The FIDUCEO ST CDR differs that CDRs in the following points:

- The calibration of the brightness temperatures used is revised for the FIDUCEO ST CDR. The first step in this has been multi-sensor harmonisation to obtain baseline calibration coefficients (Giering et al., 2019). For specific ST application, these coefficients were adjusted such that SSTs had lower bias, using a method of cross-referencing to matched drifting buoys (Merchant et al., 2019).

- Perturbations to the obtained ST and quality level determination are provided for an ensemble of 10 members, for the purpose of propagating uncertainty in ST in complex (large scale, non-linear) applications.
- The FIDUCEO ST CDR includes retrievals over the world's 300 largest lakes, unlike the SST-only product. (Lakes, including much smaller lakes, are addressed in other CDRs requiring significantly different methods to cope with the difficulties of small target water bodies.)

6 ST processing chain methods

This section provides an overview of the method used to retrieve ST for this product.

Cloud detection and retrieval are based on the physics of radiative transfer using the model RTTOV version 11.3 for calculating and integrating clear-sky absorption and (for infrared) emission of channel radiance. Clouds absorb radiance emitted from the sea surface and emit radiance at the cloud top temperature. ST retrieval under the assumption of cloud-free conditions is therefore erroneous if pixels are in fact fully or partially cloud filled. Cloud detection is applied to the satellite imagery to minimise cloud biases in STs. We calculate the probability of clear-sky given the radiances and the prior atmospheric and surface state using a simplified formulation of Bayes' theorem as follows:

$$P(c|\mathbf{y},\mathbf{x}) = \frac{P(\mathbf{y}|\mathbf{x},c)P(c)}{P(\mathbf{y}|\mathbf{x})}$$

where: c is the condition of being clear-sky over ice-free ocean; $\mathbf{y}$ is the observation vector, here containing the brightness temperatures (BTs) of thermal channels, the reflectances (for day-lit scenes) of reflectance channels and a local standard deviation of BT over 3-by-3 pixels of a further channel; and $\mathbf{x}$ is the state vector, listing variables describing the prior understanding, from NWP, of the surface temperature, surface wind speed, atmospheric temperature profile and atmospheric humidity profile. For $P(c)$, the NWP local cloud fraction is used, although constrained to the range 0.05 and 0.5 so as not to determine the outcome too strongly from that prior. $P(\mathbf{y}|\mathbf{x},c)$ is calculated on-the-fly by radiative transfer simulation, accounting for the uncertainty in $\mathbf{x}$, noise in observations and uncertainty in forward modelling. $P(\mathbf{y}|\mathbf{x},\bar{c})$ is evaluated from look-up tables, obtained iteratively by accumulating the reflectance, brightness temperature and spatial coherence properties of cloud-flagged areas over several years of orbits in a prior pass of cloud detection. SSTs are evaluated for those pixels for which the posterior probability of clear sky, $P(c|\mathbf{y},\mathbf{x})$, exceeds a threshold.

STs from the AVHRR are derived using a reduced-state-space "optimal estimation" (OE; Merchant et al, 2008). Designating the simulated BTs as $\mathbf{F}(\mathbf{x})$, the OE is

$$\mathbf{z} = \mathbf{z}_a + \mathbf{S}_a\mathbf{K}^\mathrm{T}\left(\mathbf{K}\mathbf{S}_a\mathbf{K}^\mathrm{T} + \mathbf{S}_\epsilon\right)^{-1}\left(\mathbf{y} - \mathbf{F}(\mathbf{x}_a)\right) = \mathbf{z}_a + \mathbf{G}\left(\mathbf{y} - \mathbf{F}(\mathbf{x}_a)\right)$$

where $\mathbf{x}_a$ is both a prior estimate of the state and point of linearization for forward modelling; $\mathbf{z}_a$ is the reduced equivalent to $\mathbf{x}_a$; $\mathbf{S}$ variables are error covariance matrices, $\mathbf{S}_\epsilon$ being that of the measurement-relative-to-forward-model errors, and $\mathbf{S}_a$ being that of the reduced prior state errors; $\mathbf{K}$ comprises the derivatives of the observations in $\mathbf{y}$ with respect to the reduced state variables, which are outputs of RTTOV.

Estimates of standard uncertainty (which may be considered as the standard deviation of the estimated error distribution) are provided for the core ST retrieval. The components of uncertainty are designated by their error correlation structure (uncorrelated, synoptically correlated and large-scale correlated). Errors that are independent (uncorrelated) between observations arise from the instrumental noise in the satellite observations of brightness temperature. The uncorrelated component of uncertainty is estimated therefore by propagating models of instrumental noise through the retrieval process. The component of uncertainty labelled as synoptically correlated refers to errors that are largely in common (nearly perfectly correlated) between SST observations that are adjacent and simultaneous, and become randomised (uncorrelated) as spatio-temporal distance between observations increases. In OE, the error covariance matrix of the retrieval is a standard quantity that is calculated, and extracting the component corresponding to the propagation of $\mathbf{S}_a$ through the retrieval provides an estimate of the SST synoptically correlated uncertainty. The systematic component in the SST uncertainty covers all effects that may be described as biases, whether in the sensors' calibrations, radiative transfer models or physical assumptions made in retrieval (for example, in relation to the loading of atmospheric aerosol).

Uncertainty from the ST impact of cloud-affected pixels that nonetheless pass the cloud-detection procedures is not accounted for in the tri-partite uncertainty attached to the best-estimate ST (but is addressed in the ensemble generation process).

A confidence level on a scale 0 to 5 is provided for each ST as a quality indicator, following international convention. Five (5) indicates the highest confidence. Quality levels 4 and 5 should be used for climate applications where absolute accuracy of ST is important.

The primary retrieved quantity is the skin ST estimate made at the satellite overpass time. Many users seek an estimate of SST at depths of order tens of centimetres. For this reason, the products include a model-based adjustment to estimate also ST at a depth of 20 cm. The model is a near-surface turbulence closure model validated on ocean conditions, with less validity for inland waters (e.g., because of lower salinity).

The product is gridded on a spatial grid of 0.05° in latitude and longitude. This is done from the full imagery by averaging only the STs of the highest available quality level within the cell. Simple averaging is used. The quality level of the gridded value is the quality level of the data used to form the average.

When averaging n L2P SSTs to make daily 0.05° gridded L3 products, the uncertainty from random errors decreases from "$1/\sqrt{n}$" averaging, whereas the uncertainty from the other two components does not. (When gridding L2P data to larger and/or longer scales, averaging down of the correlated errors would occur, but this is negligible for one pass on a scale of the grid cell.) The SST of a 0.05° cell is often calculated from pixels that do not fill the cell, because of cloud cover, but users typically treat the gridded SST as a value representative of the cell as a whole, and therefore the sub-sampling is another source of uncertainty. The uncertainty is parameterised effectively in terms of the fraction of the cell observed and the variability in SST in the observed part of the cell. There is no correlation of this effect between cells, so this contributes to the uncorrelated component of uncertainty in the L3U SST products.

The ensemble of perturbations is obtained via the following steps:

- Probabilistic simulation (using information from the on-board calibration cycles) of random perturbations to the instrument counts (including calibration cycle counts as well as Earth view) of the correct distribution.
- Use of perturbed counts in the processing chain: counts -> radiance -> brightness temperature -> cloud detection -> retrieval -> gridding with the consequences of perturbations arising at each step.
- Use of perturbed calibration coefficients for counts -> radiance conversion. The perturbations to the calibration coefficients are derived from sampling the estimated error covariance matrix for the coefficients obtained from the harmonisation processes. A single set of perturbed calibration coefficients is used for each ensemble member.
- Use of perturbed cloud detection thresholds, spanning the plausible range of daytime and nighttime probability thresholds evenly. A consistent pair of day and night probability thresholds is used for a given ensemble member throughout the dataset.
- Given the above, a quality level and ST retrieval are obtained for each ensemble member and pixel. The ensemble of perturbations in the products is then found by subtracting from each perturbed result the corresponding best-estimate result.

7 Product definition

The FIDUCEO ST CDR is generated by a version of the ESA SST CCI processing chain adapted to enable the ensemble aspects. Therefore, the CDR product definition is an adapted product of the SST CCI CDR. For this reason, the naming of variables within the product etc is designed for an SST only product, and thus both sea and lake STs appear in variables named for SST. “SSTs” in lake locations are therefore lake surface water temperature estimates in reality.

The ESA SST CCI product specification document gives comprehensive information of all aspects of the product definition, and is available in the relevant version at http://www.esa-sst-cci.org/PUG/pdf/SST_CCI-PSD-UKMO-201-Issue-2-signed.pdf. Here, only an overview and statement of differences is required.

The products are netCDF4. The following table gives the metadata differences from the SST CCI specification, all other metadata being as in SST CCI v2.0 products:

Global Metadata Field	Value
Conventions	"CF-1.6"
title	"FIDUCEO SST Ensemble Member"
summary	""
references	"CDF_CDR_File_Spec"
acknowledgement	"Funded by the European Commission under Grant Agreement 638822"
project	"Fidelity and Uncertainty in Climate Data Records from Earth Observation (FIDUCEO), European Commission, Grant Agreement: 638822"
license	"This dataset is released for use under CC-BY licence (https://creativecommons.org/licenses/by/4.0/) and was developed in the EC FIDUCEO project \"Fidelity and Uncertainty in Climate Data Records from Earth Observations\". Grant Agreement: 638822"
creator_name	"FIDUCEO project"
creator_email	"fiduceo-coordinator@lists.reading.ac.uk"
creator_url	"www.fiduceo.eu"
institution	"University of Reading"
source	""
naming_authority	"Centre for Environmental Data Archival (CEDA)"
file_quality_level	""
creator_processing_institution	"These data were produced on the JASMIN infrastructure at STFC as part of the FIDUCEO project"
id	""
Metadata_Conventions	None
gds_version_id	None
spatial_resolution	None
metadata_link	None
publisher_name	None
publisher_url	None
publisher_email	None
comment	None
product_specification_version	None

The variable dimensions, coordinates and variable names are as follows:

```
Dimensions:         (bnds: 2, ensemble_number: 10, lat: 3600, lon: 7200, time: 1)
Coordinates:
  * lat                                 (lat) float32 -89.975 ... 89.975
  * lon                                 (lon) float32 -179.975 ... 179.975
  * time                                (time) datetime64[ns] YYYY-MM-DDTHH:MM:SS
  * ensemble_number                     (ensemble_number) int8 1 2 3 ... 8 9 10
Dimensions without coordinates: bnds
Data variables:
    lat_bnds                            (lat, bnds) float32 ...
    lon_bnds                            (lon, bnds) float32 ...
    time_bnds                           (time, bnds) datetime64[ns] ...
    sea_surface_temperature             (time, lat, lon) float32 ...
    sea_surface_temperature_depth       (time, lat, lon) float32 ...
    sst_dtime                           (time, lat, lon) timedelta64[ns] ...
    sst_depth_dtime                     (time, lat, lon) timedelta64[ns] ...
    sses_bias                           (time, lat, lon) float32 ...
    sses_standard_deviation             (time, lat, lon) float32 ...
    sst_depth_total_uncertainty         (time, lat, lon) float32 ...
    l2p_flags                           (time, lat, lon) float32 ...
    quality_level                       (time, lat, lon) float32 ...
    wind_speed                          (time, lat, lon) float32 ...
    large_scale_correlated_uncertainty  (time, lat, lon) float32 ...
    synoptically_correlated_uncertainty (time, lat, lon) float32 ...
    uncorrelated_uncertainty            (time, lat, lon) float32 ...
    adjustment_uncertainty              (time, lat, lon) float32 ...
    aerosol_dynamic_indicator           (time, lat, lon) float32 ...
    sensitivity                         (time, lat, lon) float32 ...
    sea_surface_temperature_delta       (time, ensemble_number, lat, lon) float32 ...
    quality_level_delta                 (time, ensemble_number, lat, lon) float32 ...
```

8 Example contents

To introduce the contents to users, an iPython notebook has been prepared. This notebook shows codes and outputs for a sample day of data (2009/03/01) comprising 26 L3U files. Key file variables are identified and plotted. It is shown how to "flatten" the 26 files into a single daily file (level 3 collated) of the recommended quality levels. The statistical characteristics of the perturbations are also plotted, illustrating that the perturbations both modify SST values and the quality level designation (as expected). This latter aspect of uncertainty has not been assessable prior to creation of this dataset.

See Appendix C for the workbook.

A. Future plans

The dataset will be investigated within the ESA SST CCI and ESA Lakes CCI projects for the new insights it will yield into surface temperature uncertainty in products, particularly in relation to the interaction of pixel uncertainties and quality level designation. It is possible that ESA SST CCI may adopt an ensemble based approach for its v4.0 product (expected release 2024), depending on the results of the above investigation. (The design of the v3.0 release is already fixed.) This will involve reprocessing and replacing the current ensemble dataset, since various aspects of the processing chain (prior NWP fields, version and capabilities of radiative transfer model, etc) will have evolved.

There is no doubt that the Ensemble ST CDR will be very informative scientifically. Whether the concept of ensemble production is used in future full-scale production will depend partly on user interest and demand. Users who find useful potential in the approach (for example, if scaled up to a full timeseries production) are encouraged to liaise with the FIDUCEO / CCI team: c.j.merchant@reading.ac.uk and j.mittaz@reading.ac.uk.

B. Known problems

The Ensemble ST CDR spans a decade derived from a single sensor, which is short compared to the requirement for many climate applications.

C. Appendix: Workbook

The PDF output of the workbook is reproduced in the following pages.

Workbook illustrating the content and use of FIDUCEO Ens ST CDR files

In [137]:
```
import numpy as np
import xarray as xr
import matplotlib.pylab as plt
import os
```

Concatenate data for a 1-day sample of files

... and list the content of the dask arrays

In [138]:
```
# Open the file(s)
path = "/Users/chris/Projects/FIDUCEO/SST-ensemble/"
ds = xr.open_mfdataset(path+'*ENSEMBLE2.0-v02.0-fv01.0.nc')
ds
```

Out[138]:
```
<xarray.Dataset>
Dimensions:                          (bnds: 2, ensemble_number: 10, l
at: 3600, lon: 7200, time: 26)
Coordinates:
  * lat                              (lat) float32 -89.975 ... 89.975
  * lon                              (lon) float32 -179.975 ... 179.9
75
  * ensemble_number                  (ensemble_number) int8 1 2 3 ...
8 9 10
  * time                             (time) datetime64[ns] 2009-03-01
T00:42:29 ... 2009-03-01T21:24:40
Dimensions without coordinates: bnds
Data variables:
    lat_bnds                         (time, lat, bnds) float32 dask.a
rray<shape=(26, 3600, 2), chunksize=(1, 3600, 2)>
    lon_bnds                         (time, lon, bnds) float32 dask.a
rray<shape=(26, 7200, 2), chunksize=(1, 7200, 2)>
    time_bnds                        (time, bnds) datetime64[ns] das
k.array<shape=(26, 2), chunksize=(1, 2)>
    sea_surface_temperature          (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sea_surface_temperature_depth    (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sst_dtime                        (time, lat, lon) timedelta64[ns]
dask.array<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sst_depth_dtime                  (time, lat, lon) timedelta64[ns]
dask.array<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sses_bias                        (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sses_standard_deviation          (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sst_depth_total_uncertainty      (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    l2p_flags                        (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    quality_level                    (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    wind_speed                       (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    large_scale_correlated_uncertainty   (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    synoptically_correlated_uncertainty  (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    uncorrelated_uncertainty         (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    adjustment_uncertainty           (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    aerosol_dynamic_indicator        (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
    sensitivity                      (time, lat, lon) float32 dask.ar
ray<shape=(26, 3600, 7200), chunksize=(1, 3600, 7200)>
```

```
    sea_surface_temperature_delta       (time, ensemble_number, lat, lo
n) float32 dask.array<shape=(26, 10, 3600, 7200), chunksize=(1, 10, 3600,
7200)>
    quality_level_delta                 (time, ensemble_number, lat, lo
n) float32 dask.array<shape=(26, 10, 3600, 7200), chunksize=(1, 10, 3600,
7200)>
Attributes:
    Conventions:                     CF-1.6
    title:                           FIDUCEO SST Ensemble Member
    summary:
    references:                      CDF_CDR_File_spec
    institution:                     University of Reading
    history:                         Created using GBCS library v2.11.0-2
-g07...
    license:                         This dataset is released for use und
er C...
    id:
    naming_authority:                Centre for Environmental Data Archiv
al (...
    product_version:                 2.0
    uuid:                            8da32ba8-be97-11e9-b71b-3b268b69f2b4
    tracking_id:                     8da32ba8-be97-11e9-b71b-3b268b69f2b4
    netcdf_version_id:               4.7.0 of May 17 2019 18:14:24
    date_created:                    20190814T132933Z
    file_quality_level:
    start_time:                      20090301T004229Z
    time_coverage_start:             20090301T004229Z
    stop_time:                       20090301T022351Z
    time_coverage_end:               20090301T022351Z
    time_coverage_duration:          P0DT01H41M21S
    time_coverage_resolution:        P0DT1H40M00S
    source:
    platform:                        MetOpA
    sensor:                          AVHRR_GAC
    keywords:                        Oceans > Ocean Temperature > Sea Sur
face...
    keywords_vocabulary:             NASA Global change Master Directory
(GCM...
    standard_name_vocabulary:        NetCDF Climate and Forecast (CF) Met
adat...
    geospatial_lat_units:            degrees_north
    geospatial_lat_resolution:       0.05
    geospatial_lon_units:            degrees_east
    geospatial_lon_resolution:       0.05
    geospatial_vertical_min:         -0.2
    geospatial_vertical_max:         -1e-05
    acknowledgment:                  Funded by the European Commission un
der ...
    creator_name:                    FIDUCEO project
    creator_email:                   fiduceo-coordinator@lists.reading.a
c.uk
    creator_url:                     www.fiduceo.eu
    creator_processing_institution:  These data were produced on the JASM
IN i...
    project:                         Fidelity and Uncertainty in Climate
Data...
```

```
    northernmost_latitude:            90.0
    southernmost_latitude:            -90.0
    easternmost_longitude:            180.0
    westernmost_longitude:            -180.0
    geospatial_lat_min:               -90.0
    geospatial_lat_max:               90.0
    geospatial_lon_min:               -180.0
    geospatial_lon_max:               180.0
    processing_level:                 L3U
    cdm_data_type:                    grid
    source_file:                      FIDUCEO_FCDR_L1C_AVHRR_MTAC3A_200903
0100...
    source_uuid:                      1090c1b4-5098-40b1-8f43-b8ae034a3683
```

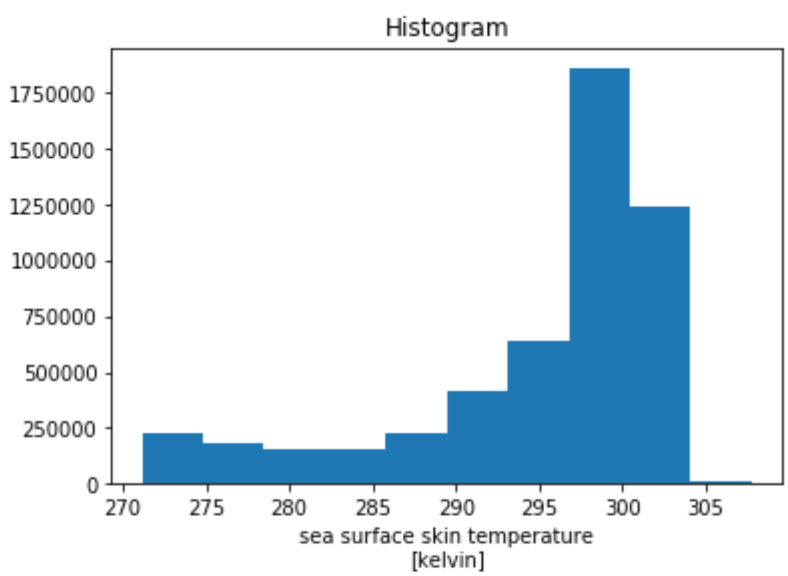
The time dimension is along the times associated with each of the 26 semi-orbital files during the day.

Look at some basic content: surface temperature, deltas, quality levels

The principal variable is labelled sea_surface_temperature and contains the skin surface temperature estimate for oceans and large inland waters (lakes)

In [139]:
```
ds.sea_surface_temperature.plot.hist()
```

Out[139]:
```
(array([ 223519.,  185337.,  152353.,  151606.,  229543.,  419070.,
         640744., 1860830., 1238633.,   11667.]),
 array([271.15   , 274.811  , 278.472  , 282.133  , 285.794  , 289.45502,
        293.116  , 296.777  , 300.43802, 304.099  , 307.76   ],
       dtype=float32),
 <a list of 10 Patch objects>)
```



The SST distribution dominates the histogram, and warmer waters are more common globally.

As well as the baseline (best estimate, default processing) ST data above, the product includes an ensemble of perturbations to that baseline result. These are constructed to capture the uncertainty from various sources with realistic error covariance (i.e., realistic standard deviation of errors and error correlation properties). The intention of the ensemble is strictly in application for uncertainty assessment in downstream uses of the ST product.

What are the properties of the perturbations?

In [140]:
```
mdel = ds.sea_surface_temperature_delta.mean(axis=(0,2,3)).values
print(mdel)
```

```
[-0.11520482  0.03553322 -0.09873942 -0.05916727  0.10415475 -0.02281259
  0.16591626 -0.00414346  0.02532599 -0.01538843]
```

In [141]:
```
sdel = ds.sea_surface_temperature_delta.std(axis=(0,2,3)).values
print(sdel)
```

```
[0.3838479  0.26443893 0.3292664  0.3146555  0.2821481  0.35439718
 0.34243074 0.37238365 0.26211488 0.35600147]
```

In [142]:
```
ndel = ds.sea_surface_temperature_delta.count(axis=(0,2,3)).values
print(ndel)
```

```
[4874309 5000996 4943935 4949442 4948465 4959176 4986870 4964162 4994210
 4975565]
```

The number of data in the ensemble members differs because after perturbation the quality assessment changes, and some are no longer valid at the quality level we are selecting. This is an effect (source of uncertainty) that influences the product surface temperature estimates, but which cannot be propagated through the retrieval conventionally.

In [143]:
```
print(np.mean(mdel), np.std(mdel), np.sqrt(np.mean(sdel**2)))
print(np.std(ndel)/np.mean(ndel))
```

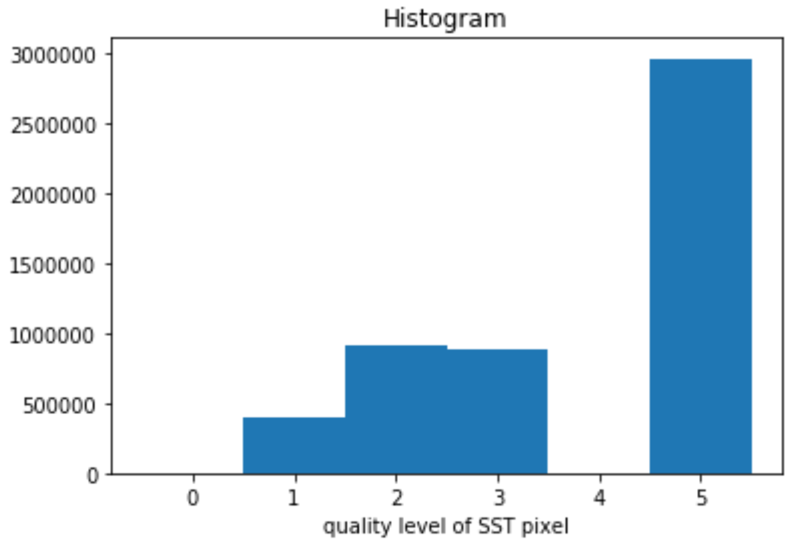
```
0.0015474218 0.08219079 0.32882613
0.006890134058311235
```

The mean across all the perturbations is close to zero. The standard deviation across the mean of each ensemble member is ~0.1, represents more systematic effects between ensemble members. The mean standard deviation within each ensemble member is ~0.3 K ro ~0.4 K, representing the faster-varying error sources. These uncertainties are consistent with results we get in validation comparison with drifting buoys.

The other key variable is the quality level attributed to each datum.

In [144]:
```
ds.quality_level.plot.hist(bins = [i-0.5 for i in range(0,7)])
```

Out[144]:
```
(array([      0.,  399288.,  915805.,  878851.,       0., 2963843.]),
 array([-0.5,  0.5,  1.5,  2.5,  3.5,  4.5,  5.5]),
 <a list of 6 Patch objects>)
```

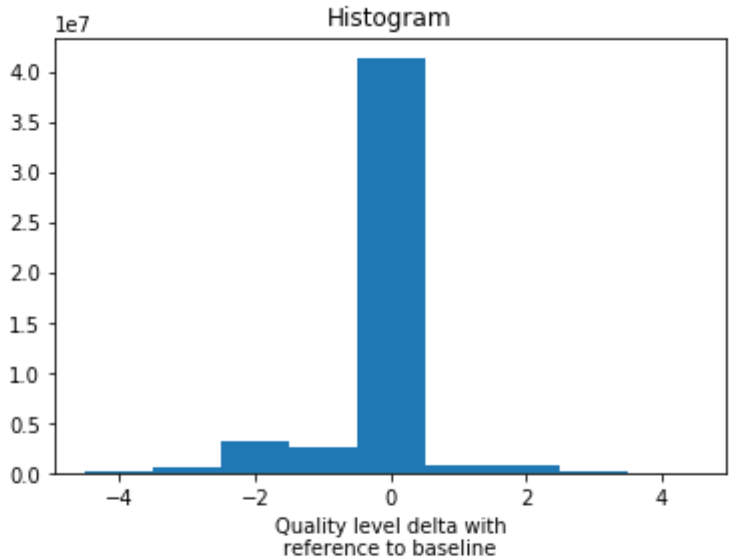


There are no QL = zero, instead they are NaN. QL = 5 is most the common and is recommended for use. The absence of QL = 4 is expected for periods (such as this) when there are no major stratospheric aerosol events.

Since the brightness temperatures are used in QL assessment and are perturbed as part of the ensemble generation process, there are also perturbations on the best-estimate QL field.

In [145]:
```
ds.quality_level_delta.plot.hist(bins = [i-5.5 for i in range(1,11)])
```

Out[145]:
```
(array([3.2272400e+05, 6.5948800e+05, 3.2675230e+06, 2.6381360e+06,
        4.1352218e+07, 9.2185500e+05, 7.6007200e+05, 1.5289900e+05,
        2.5030000e+04]),
 array([-4.5, -3.5, -2.5, -1.5, -0.5,  0.5,  1.5,  2.5,  3.5,  4.5]),
 <a list of 9 Patch objects>)
```



In [146]:
```
np.sum(ds.quality_level_delta.values==0)/np.sum(np.isfinite(ds.quality_leve
```

Out[146]:
```
0.825394478975975
```

So, in this sample, about one fifth of the baseline quality levels get revised when the brightness temperatures are realistically perturbed. The implication of this is that there are significant fractions of marginally clear-sky cells, and the uncertainty in resulting surface temperature products from the uncertain decision to include them or not for ST retrieval is a form of uncertainty not captured by non-ensemble methods.

Create daily ST QL5 ensemble from the best-estimate + deltas

In [147]:
```
# Make a list of the filenames for the day
fl = np.sort([f for f in os.listdir(path) if f.endswith('ENSEMBLE2.0-v02.0-
```

In [148]:
```
ds = xr.open_dataset(path+fl[0]) # open the first of the files
ds['time'] = ds.indexes['time'].normalize() # only interested in the day no
```

In [149]:
```
edQL = ds.quality_level_delta
edQL = edQL.where(np.isfinite(edQL), 0)

bQL = ds.quality_level
bQL = bQL.where(np.isfinite(bQL), 0)

eQL = edQL.copy(deep=True)
eQL.load()
for e in range(10): eQL[:,e,:,:] = bQL + edQL[:,e,:,:]
```

eQL now holds the quality level for each ensemble member and bQL holds the quality level for the best estimate. For climate applications, the recommendation is that only QL 4 & 5 are used for further computations. Some users whose requirement for absolute accuracy is less critical may find utility in QL 3.

In [150]:
```
edST = ds.sea_surface_temperature_delta
eST = edST.copy(deep=True)
eST.load()
for e in range(10): eST[:,e,:,:] = ds.sea_surface_temperature + edST[:,e,:,
```

In [151]:
```
print(ds.sea_surface_temperature[0,2596,6010].values, edST[0,:,2596,6010].v
```

```
272.83 [ 0.          -0.04        0.21000001 -0.15        -0.24000001  0.1
 -0.18        0.35000002  0.15        0.08000001] [272.83    272.78998 27
3.03998 272.68    272.59    272.93    272.65
 273.18    272.97998 272.90997]
```

These are the best estimate ST and the ensemble of perturbations for an illustrative cell from the first semi-orbital file.

Going to use only the QL 4 & 5 (in practice, 5)

In [152]:
```
eST = eST.where(eQL > 3,0)
eST = eST.where(np.isfinite(eST.values), 0)

nST = np.zeros(np.shape(eST.values))
nST[eST.values>0] += 1

beST = ds.sea_surface_temperature.where(ds.quality_level > 3, 0)
beST = beST.where(np.isfinite(beST.values), 0)

nbeST = np.zeros(np.shape(beST.values))
nbeST[beST.values>0] += 1
```

In [153]:

```
# Loop over the other orbit files of the day and add in any QL 5 SSTs (simp
# The loop sums valid ST values and divides by the corresponding number to
for f in fl[1:]:
    ds = xr.open_dataset(path+f)
    ds['time'] = ds.indexes['time'].normalize() # only interested in the da
    edQL = ds.quality_level_delta
    edQL = edQL.where(np.isfinite(edQL), 0)
    bQL = ds.quality_level
    bQL = bQL.where(np.isfinite(bQL), 0)
    for e in range(10): eQL[:,e,:,:] = bQL + edQL[:,e,:,:]
    edST = ds.sea_surface_temperature_delta
    edST.load()
    for e in range(10): edST[:,e,:,:] += ds.sea_surface_temperature
    edST = edST.where(eQL > 3,0)
    edST = edST.where(np.isfinite(edST.values), 0)
    nST[edST.values>0] += 1
    eST += edST
    buse = np.isfinite(ds.sea_surface_temperature.values)
    beST += ds.sea_surface_temperature.where(buse,0 )
    nbeST[buse] += 1
```

In [154]:

```
# Form the average
nST[nST==0]=-1 # avoid div 0
eST2 = eST.values/nST

nbeST[nbeST==0]=-1
beST2 = beST.values/nbeST
```

In [155]:

```
eSTs = eST.to_dataset(name='surface_temperature')
eSTs.surface_temperature.attrs['long_name'] = "skin surface temperature of
eSTs.surface_temperature.values = eST2
eSTs.to_netcdf(path+'ST-QL5-ensemble-daily.nc') # writing out the dataset s

beSTs = beST.to_dataset(name='surface_temperature')
beSTs.surface_temperature.attrs['long_name'] = "skin surface temperature of
beSTs.surface_temperature.values = beST2
beSTs.to_netcdf(path+'ST-QL5-best-daily.nc') # writing out the dataset so a
```

In [156]:

```
eSTs
```

Out[156]:

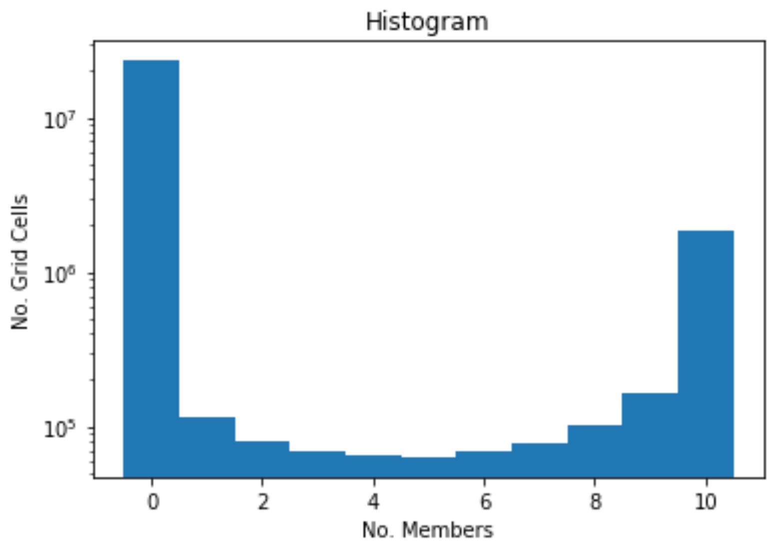
```
<xarray.Dataset>
Dimensions:              (ensemble_number: 10, lat: 3600, lon: 7200, tim
e: 1)
Coordinates:
  * lat                  (lat) float32 -89.975 -89.925 -89.875 ... 89.925
89.975
  * lon                  (lon) float32 -179.975 -179.925 ... 179.925 179.
975
  * time                 (time) datetime64[ns] 2009-03-01
  * ensemble_number      (ensemble_number) int8 1 2 3 4 5 6 7 8 9 10
Data variables:
    surface_temperature  (time, ensemble_number, lat, lon) float64 -0.0
... -0.0
```

In [157]:
```
sdST = eSTs.surface_temperature.where(eSTs.surface_temperature>0).std(axis=
```

In [158]:
```
neST = eSTs.surface_temperature.where(eSTs.surface_temperature>0).count(axi
```

In [168]:
```
nhist = np.array(neST.plot.hist(yscale='log',bins = [i-0.5 for i in range(0
plt.xlabel('No. Members') # How many cells have how many ensemble members
plt.ylabel('No. Grid Cells')
```

Out[168]:
```
Text(0, 0.5, 'No. Grid Cells')
```



In [160]:
```
print('Ensemble mean clear-sky cell rate over water (approx)', np.sum(nhist
```

```
Ensemble mean clear-sky cell rate over water (approx) 0.14682115299823634
```

In [161]:
```
print('Ensemble estimate of mean uncertainty per grid cell ST', sdST.where(
```

```
Ensemble estimate of mean uncertainty per grid cell ST 0.2755325173700267
6
```

In [162]:
```
print('Global mean ST for day of observations', beSTs.surface_temperature.w
```

```
Global mean ST for day of observations <xarray.DataArray 'surface_tempera
ture' ()>
array(294.838827)
```