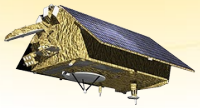
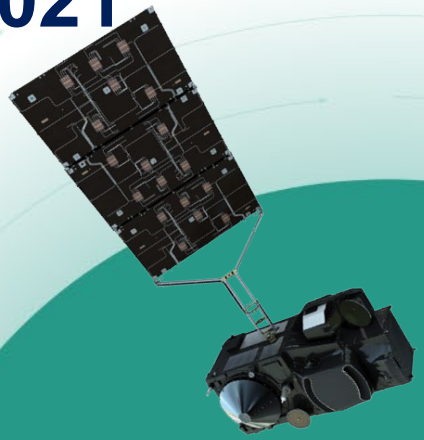
Matchup Database (MDB) Task Team Report #2

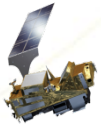
GHRSST XXII, 07/06/2021

Igor Tomazic
Jean-Francois Piolle





GHRSST
GROUP FOR HIGH RESOLUTION SEA SURFACE TEMPERATURE



Ifremer

MDB TT

- Task Team initiated at GHRSST XX, 2019, Frascati, Italy.

- Objective: **Suggest the way forward towards common SST MDB production method and assessment metrics and protocols**

- **Co-chairs**: Igor Tomazic, Jean-Francois Piolle
- **Participants**: Edward Armstrong, Gary Corlett, Owen Embury, Chelle Gentemann, Jacob Hoeyer, Alexander Ignatov, Stephane Saux Picart, Jorge Vazquez, Werenfrid Wimmer

MDB TT – timeline

- GHRSST XX: Task Team initiated, Frascati, Italy.
- GHRSST XXI: Tasks drafted, Report #1
- 2021/05: Progress meeting
- GHRSST XXII: Report #2 + discussion

Tasks

- T1. Collect journal papers explaining different MDB criteria's, tools, metrics in SST analysis ← Ongoing
- T2. MDB tools
  - T2.1. List of MDB tools for comparison
  - T2.2. List of MDB functionalities for matrix comparison
  - T2.3. Create matrix between MDB functionalities

  ← Almost completed: felyx, OSI-SAF MDB, DMI+FI MDB, NOAA ACSPO MDB, CCI MDB, *NASA SDAP MDB (next slide)*
- T3. Metrics and protocols
  - T3.1. Identify a list of validation and referent types of data and datasets
  - T3.2. Identify current metrics and protocols in SST analysis
  - T3.3. Group metrics and protocols by validation (i.e. satellite) and referent (i.e. in situ) data

  ← In progress (next slides)
- T4 Round robin validation SST MDB intercomparison
  - T4.1 Define round-robin intercomparison
  - T4.2 Perform round-robin validation intercomparison on test MDB

  ← Next GHRSST XXIII (in preparation)
- T5. Create whitepaper + online information

  ← GHRSST XXIII/XXIV

MDB TT – progress meeting 05/2021

Participants: Igor/Jean-Francois/Ed/Gary/Jorge/Sasha/Stephane/Werenfrid

- currently defined MDB TT activities are leading to white paper with overview of currently used SST MDB tools (Task 2) and the focus on metrics and protocols to be used in MDB analysis (T3) based on the outcome of the round robin exercise (T4)
  - there was discussion on defining GDS for MDB, but this is currently not in the scope (could be next phase)
  - JFP raised that match-up methodology and format for MDB was initially part of GDS v1.x and that the future white paper shall revisit and expand it

- Confirmed list of main MDB tools and functionalities (Task 2)
  - Completed MDBs: EUM FELYX, OSI SAF, DMI+FI, NOAA ACSPO, ESA CCI
  - Pending info on NASA SDAP MDB (discussion with COVERAGE PI)

- identify currently used metrics and methods in MDB productions and analysis (T3)
  - cover L2P, L3 and L4 (IR and MW, GEO and LEO) and different data types (e.g. drifters, Argo, moored, radiometers, saildrones)
  - Template on MDB generation parameters prepared (Gary) for others to fill
  - Template on MDB analysis in preparation

- different methods and metrics will be applied during the round robin activity (T4)
  - most probably this activity could be carried in WEkEO platform (presented concept), but this will be still confirmed
  - further discussion if single tool (e.g. felyx) could be used to produce all needed MDB datasets and perform comparison using different methods

T2.2/2.3 MDB functionalities

FELYX, OSI SAF, DMI+FI, NOAA ACSPO MDB, CCI MDB

Name	Organisation	development	Active evolution	Open source	web link	comment	contacts
FELYX	EUMETSAT, Ifremer	Ifremer	yes	yes		new version in development to simplify installation and further improve realibility, performance, traceability to in situ data and more	Igor Tomazic (operational Sentinel-3 MDB) Jean-Francois Piolle (development)
SST CCI MDB	SST CCI/UoR/Brockmann	Brockmann					
NOAA ACSPO MDB	NOAA	NOAA					Alexander Ignatov
NASA SDAP	NASA	NASA				There is a new 2020 NASA activity funded to bring matchup capability to the cloud. A follow-on to the DOMS project. I could invite the PI to give a presentation.	Nga Chung (PI) and Varids Tsontos are the JPL contacts
OSI-SAF MDB	OSI-SAF	OSI-SAF	yes	yes		The code is not easily portable	Stéphane Saux Picart
DMI+FI MDB	DMI + Farallon Institute	DMI+FI	yes	yes		Code is simple and portable, has been used for SLSTR L2 validation against Saildrone observations	Jacob L. Høyer and Chelle Gentemann

Category	Name	Description	FELYX	OSI-SAF MDB	DMI+FI MDB	CCI MDB
General	Intended users	operational, scientific, reanalysis	operational + reanalysis	operational, scientific, reanalysis	scientific, reanalysis	Project team
General	development status	mature or in development	v1 released + new changes in v2	mature	v1 working	Mature
General	costs	free or with costs	open source	free	free	
General	tool availability	location	gitlab	internal, on demand	internal on demand	
General	running context	nrt/analyis/ad hoc	NRT or analysis	nrt/analyis/ad hoc	nrt	Ad hoc
Software	license	code license	GPL v3.0			
Software	resources	minimum resources for running	VM 4 cores + 16 GB RAM			
Software	installation	simple/complex/containers	complex (evolution to docker containers and lighter standalone version for limited processing resources or evaluation)	complex	simple	
Software	programing language	python, ...	python 3	python 3	Python 3	Java
Software	API	rest/API/ only internal				
Software	COTS	additional of the shelf software needed to run	ElasticSearch, RabbitMQ, Celery, (Airflow, optional). These dependencies will be optional in the new lighter version.			
Software	running context	within scheduler/orchestrator or simple command line or cron, ...	airflow or command lines	command line/cron	command line	command line
Scientific	multi mission/sensor/ parameter/level	possibility to process different missions (e.g. S3A, S3B) and different parameters (not only SST) and different levels (e.g. L1, L2, L3, ..)	configurable. The tool can be used for any MDB (not SST oriented) or even match-ups with cruises, hurricane tracks,...)	only process OSI SAF internal workfiles	The core of the script can be used for matching any product with insitu, but additional coding is needed to use for specific satellite product	
Scientific	matchup criterias	time, distance	time and distance, configurable per pair (EO dataset, in situ collection)	time and distance, configurable	time and distance, configurable	time and distance, configurable
Scientific	extraction criteria	window size, history size	configurable window size per defined extraction	window size configurable	window size configurable	window size configurable
Scientific	matchup method	one to many, one to one	single matchup within selected window. For a (granule, in situ buoy) par, only one match-up is retrieved: the one with minimal time difference	one to one	one to many and one to one	one to one
Scientific	satelite product tracebility	how traceability to source satellite products are ensured	ensured through satellite product filename written in the final output. Contains also the indices of the extracted subset.	ensured through satellite product filename written in the final output. Contains also the indices of the extracted subset.	product name in final output	Product name in MMDB file
Scientific	insitu data traceability	how traceability to source in situ data are ensured	not ensured but planned in next release (source filename and metadata attributes)	in situ measurement ID, time and position written in the final output	in situ measurement ID, time and position written in the final output	Unique ID per in situ measuement in MMDB file

- To complete:
  - NASA SDAP MDB: to be discussed with COVERAGE team
  - NOAA MDB: Sasha to complete information on functionality
    - already provided detailed information on the production criterias

Outputs	output type	dataset, database, ...	dataset in netCDF files + defined metrics inside DB	dataset in netCDF format (daily)	dataset in netCDF format (daily)	Weekly dataset in NetCDF format
Outputs	output format	netcd/hdf/ascii/..	netcdf (v4)	netcdf (v4)	netcdf (v4)	netcdf (v4)
Outputs	granularity	per insitu type, satellite	per insitu type and satellite product	per satellite product (daily)	per insitu type and satellite product	per insitu type and sensor
Outputs	output content		full extraction from satellite data and selected information from the in situ data with history of in situ measurements within configurable time window		full extraction of satellite data, selected information from insitu data	Extraction of satellite data accoring to window, auxiliary MET fields and in situ history
Visualisation	results	tools for visualising results	internal jupyter notebooks, matlab + IDL scripts	internal scripts	any tool that can visualize netcdf4 files	Internal python and IDL tools
Visualisation	insitu data	tools for visualising input insitu data	ES/Kibana for position&time of individual in situ measurement (to be revised in next release)	internal scripts	any tool that can visualize netcdf4 files	Internal python and IDL tools
Visualisation	satellite data	tools for visualising input satellite data	N/A	internal scripts	any tool that can visualize netcdf4 files	Internal python and IDL tools
Visualisation	matchup data	tools for visualising matchup data	ES/Kibana for position&time (to be revised in next release)	internal scripts	any tool that can visualize netcdf4 files	Internal python and IDL tools
monitoring	monitoring	monitoring types	ES/Kibana for monitoring of in situ and satellite inputs and matchup outputs (to be revised in next release)	internal scripts		

T3 – metrics/protocols

- Collect currently used methods and criteria's in matchup analysis
  - Template in preparation (Gary) to circulate within ST
  - Split between MDB generation/production and analysis/methods
- Template for MDB generation criteria's (just finalized)
  - Prefilled for EUM FELYX, ESA CCI and NOAA ACSPO:

MDB name	**EUMETSAT FELYX**	**ESA SST CCI**				**ACSPO**		
Product level	L2P	L2P	L3U	L3C	L4	L2P	L3C	L4
Products covered	SLSTR; AVHRR; IASI; VIIRS	ATSR; AVHRR; SLSTR	As per L2P	As per L3U	SST CCI L4	VIIRS	As per L2P	As per L3C
Product depth	SSTskin	SSTskin and SST20	As per L2P	As per L3U	SST20	SSTsubskin	As per L2P	As per L3C
in situ data	Drifters; Argo; moorings; radiometers;	Drifters; Argo; GTMBA; coastal moorings; VOS; CBT, MBT, XBT; animal; radiometers; bottles	As per L2P	As per L3U	As per L3C	Drifters; Argo; GTMBA	As per L2P	As per L3C
In situ data source	CERSAT In situ (derived from CMEMS in situ TAC); ships4sst; TRUSTED	HadIOD; ships4sst	As per L2P	As per L3U	As per L3C	iQuam	As per L2P	As per L3C
Match-up criteria	Single pixel within +/- 2 hours Argo: +/- 12 hours	Single pixel within +/- 2 hours	As per L2P	Single cell; L3C assumed to be 10:30 AM or 10:30 PM	Single cell; L4 assumed to be 10:30 AM or 10:30 PM	All pixels within 10 km and 30 minutes	Single cell within 10 km and 30 minutes	As per L3C
Additional QC applied prior to match-up generation	None	None	As per L2P	As per L3U	As per L3C	None	As per L2P	As per L3C
MDB product extract	21 x 21 pixels centred on match-up location radiometers: 401x401 IASI: 5x5 pixels	9 x 9 (1 km) or 5 x 5 (GAC) pixels centred on match-up location	Single pixel	As per L3U	As per L3C	Single pixel	As per L2P	As per L3C
MDB in situ extract	+/- 12 hours centred on match-up time	+/- 12 hours centred on match-up time	Single measurement	As per L3U	As per L3C	Single measurement	As per L2P	As per L3C
Auxiliary fields in MDB not in product or in situ	Additional ECMWF fluxes for FKC model; FKC model adjustments; RTTOV simulations; TRUSTED flags	Additional ECMWF fluxes for FKC model; FKC model adjustments; RTTOV simulations	As per L2P	As per L3U	As per L3C	None	As per L2P	As per L3C

T3 – metrics/protocols

- Template for MDB analysis in preparation
  - L2P/L3/L4, IR/MW, LEO/GEO and for different in situ types
  - On-to-one/one-to-many
  - Statistics: mean/median; SD/RSD; …
  - Type of analysis:
    - Bias dependence: WV, time difference, wind, date, latitude, solar zenith angle, satellite zenith angle, across-track, SST, …
    - Histograms, uncertainty plots, …
    - Gradient analysis

- Reference: Optical radiometry for ocean climate measurements, Chapter 6.2 Assessment of Long-term Satellite Derived SST records, Corlett et al.,

- Revisit status of GHRSST Validation Protocol Document (VPD)?

Next

- In progress
  - T3: complete template (questionnaire) on matchup methods and analysis and circulate within ST
  - T2: complete functionality for other MDB tools (NASA SDAP MDB)
  - Update Moodle pages
- Q3&Q4/2021:
  - T3 – identify currently used matchup methods in MDB analysis to be used in round robin (T4)
- Q2/2022:
  - T4: round robin on different methods applied to different datasets (L2P/L3/L4; IR/MW; LEO/GEO vs. in situ types)
- Q3/2023: Whitepaper

Discussion points

- Round robin (core activity):
  - Focus on comparison of different MDB methods/analysis and not tools itself
    - Identify all MDB types of analysis
  - Single tool (e.g. felyx) applied consistently to all defined satellite/in situ datasets
    - Pros: Consistent formats with datasets available to everyone
    - Cons: Challenging to produce all possible matchup datasets
  - Different tools produce different datasets
    - Issue could be with slightly different inputs sources (e.g. in situ) Assumption is that all tools produce consistent Not analyzing tools but methods
  - Or combination of different matchup datasets produced by several tools
  - Common platform (accessible to MDB TT) to use for performing RR activity (e.g. WEkEO - TBC)
    - Available analysis (e.g. through jupyter notebooks)

- Datasets to use in round robin activity
  - Duration: 1 month or longer? Overlap between datasets?
  - Satellite data: IR: LEO:SLSTR, VIIRS, AVHRR, MODIS?; GEO: SEVIRI/AHI/ABI?; MW: AMSR-E; levels: L2, L3, L4 (L1?)
  - In situ data: types: Drifters, moorings, Argo, radiometers, saildrone; Sources: CMEMS In situ, iQuam, …

Questions & Discussion