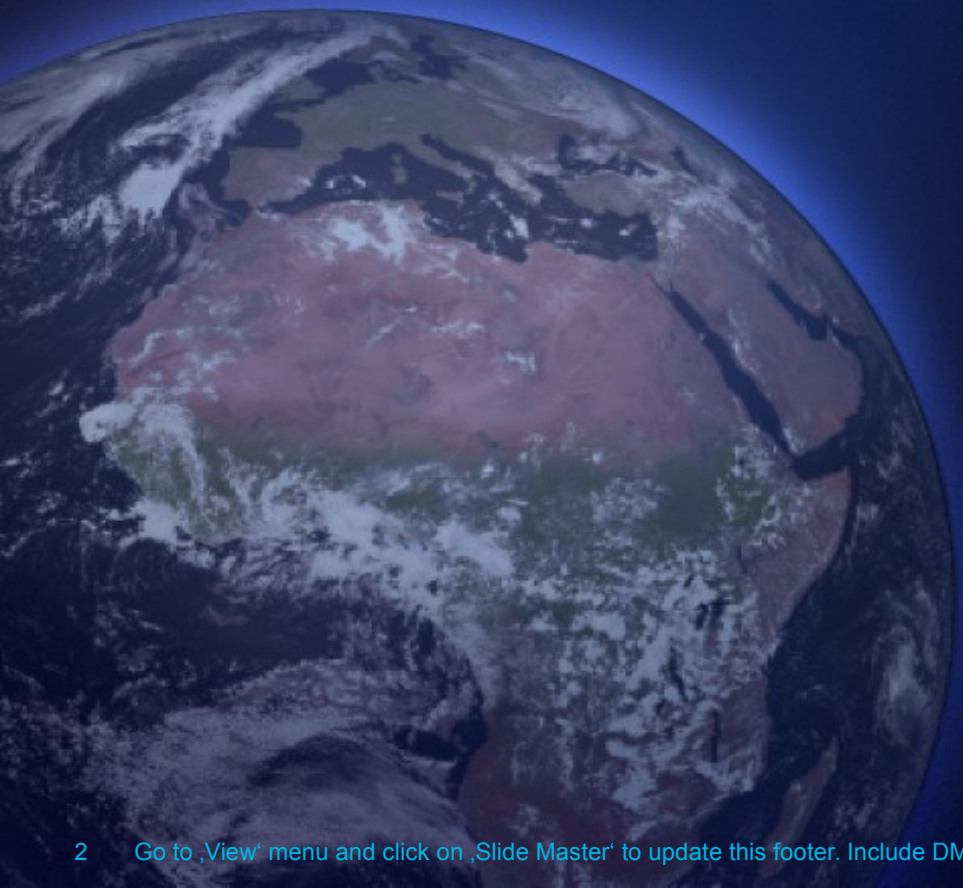
FELYX MATCH-UP DATABASES, IN SUPPORT TO SENTINEL-3 CAL/VAL

Jean-François Piollé (visiting scientist EUMETSAT, Ifremer) & OSI SAF team, Igor Tomazic, Anne O'Caroll

FELYX



Background

Intercomparison of different sources of data is a key asset when working with earth observations

- Validation (cal/val) against in situ or other sensors
- Algorithm development and improvement
- Combination of different parameters from different sources (synergy, ancillary data,…)
- Monitoring and detection of issues

Today’s sensor reach data **volume** and available **bandwidth** limitations of most users, plus **complexity** of managing multiple datastreams

Tools are required to extract the **relevant amount of information** only to perform the above tasks

felyx

- Intended for **satellite to in situ match-up extraction** and systematic data **extraction over user defined area or locations :**
  - Command line based – query through RESTful and python APIs.

- **Main functions**
  - Extraction of file subsets over static or moving locations
  - Extraction and indexing of metrics over the subsets for analytics
  - Assembling with in situ data

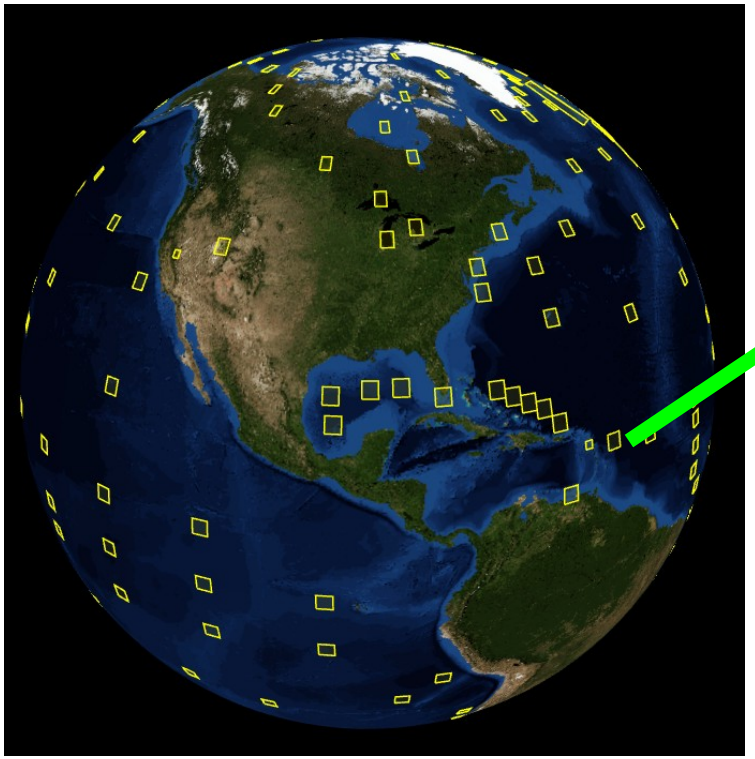
- **Main outputs**
  - Miniprods and metrics
  - Assembled multi-sensor match-up files
  - Display of metrics, alert detection through analytics tools

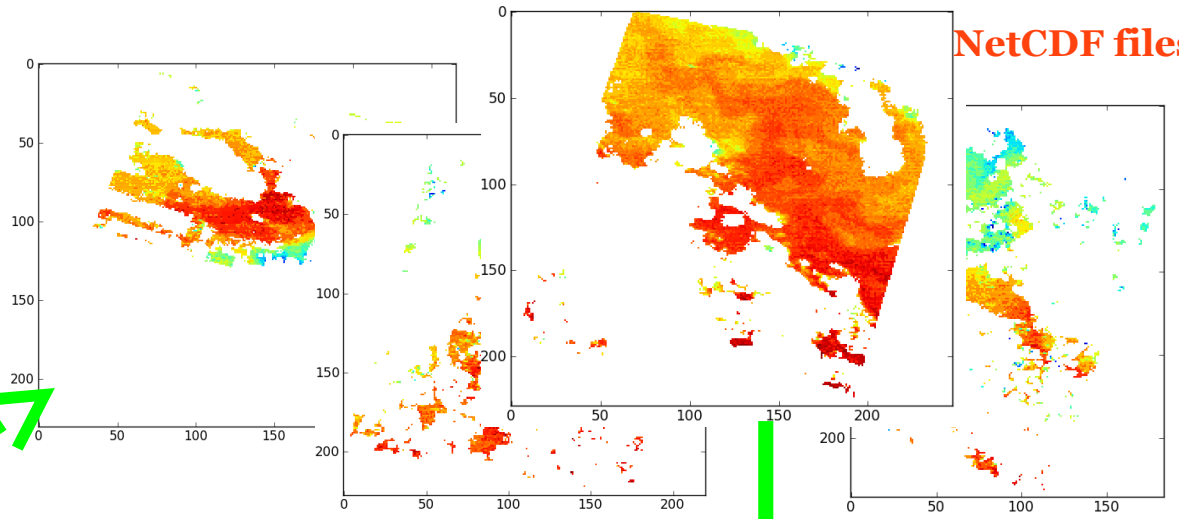
- **Implementation**
  - Open source software in python
  - Relies on existing open source frameworks for big data and distributed processing : **ElasticSearch**, **RabbitMQ**, **Celery**, ….

elasticsearch RabbitMQ

Felyx for MDB production

extract **miniprods** (subsets) over static and dynamic sites
process quantitative, qualitative, stat metrics over miniprods





NetCDF files


```
source: 20130101-IFR-L4_GHRSST-SSTfnd-ODYSSEA-GLOB_010-v2.0-fv1.0.nc
felyx_dataset_name: ifr-l4-sstfnd-odyssea-glob_010_v2_1
percentage_coverage_of_site_by_miniprod: 100.0
date_modified: 2014-04-18T10:30:21
felyx_site_identifier: ukm005
date_created: 2014-04-18T10:30:21
time_coverage_start: 2013-01-01T00:00:00
time_coverage_stop: 2013-0101T00:00:00

sst_standard_deviation : 1.34
mean_sst : 286.289
ice_presence: 0
cloud_presence": 46.80
day_or_night: "night"
mean_wind_speed: 4.8388
```

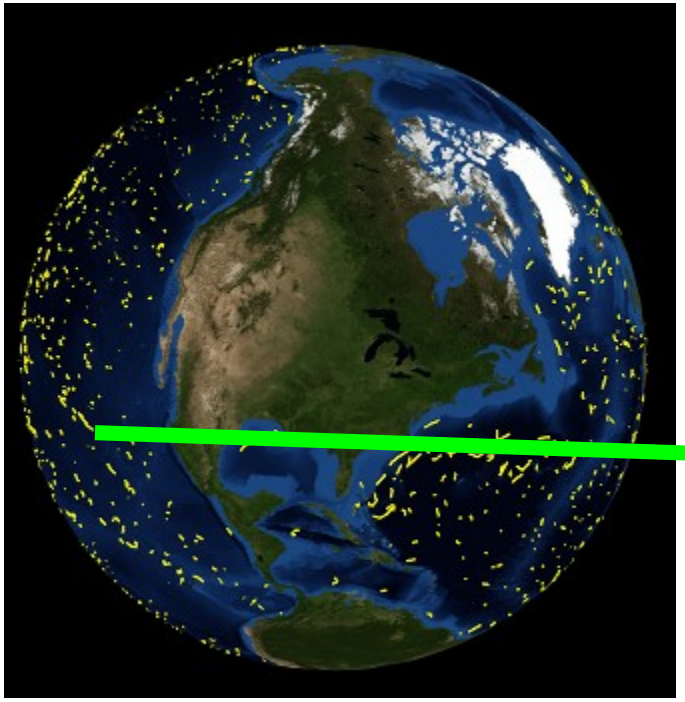

JSON files
indexed in a search
engine (ElasticSearch)

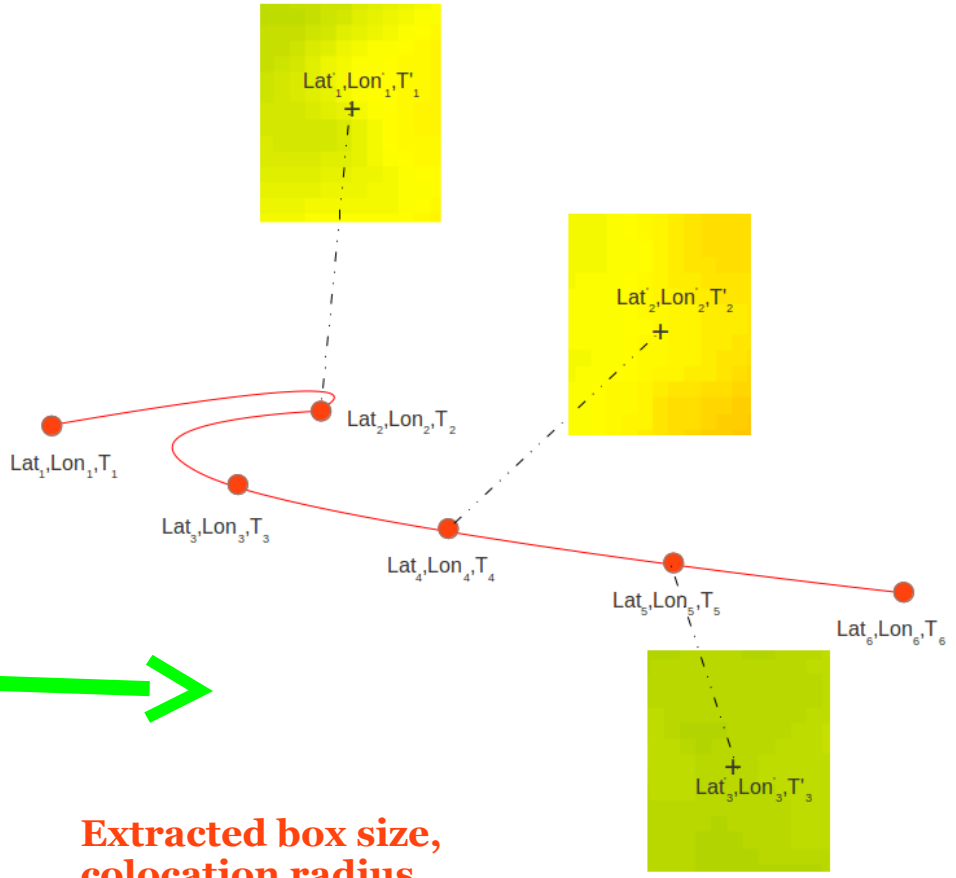
EUMETSAT

Felyx for MDB production

sites may be trajectories (buoys, cruise, hurricane)
MINIPROD's centred on trajectory locations closest in time
locations closest in time

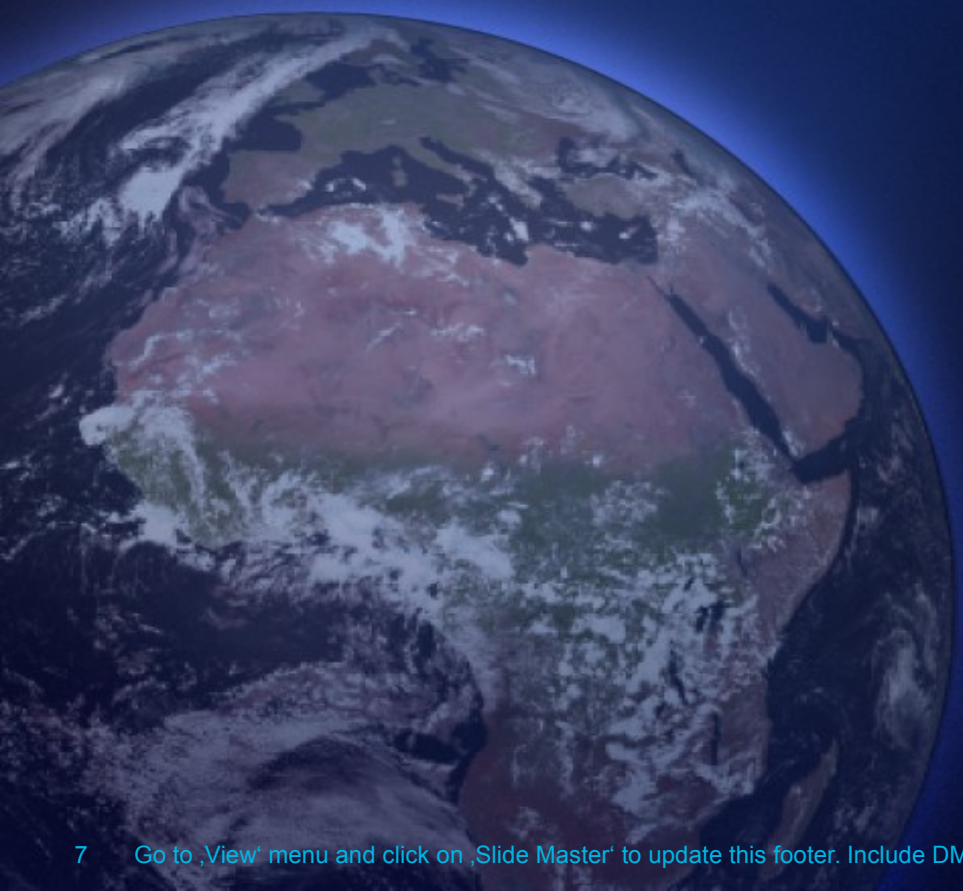
trajectory files ingested through import web service (CSV file)





Extracted box size, colocation radius, maximum temporal difference can be adjusted for each dataset

MATCH-UP DATABASES WITH FELYX



Input Sentinel-3A SLSTR marine products

Product type	Description	Level	Availability	Access	Timeliness
SL_1_RBT___	Brightness temperatures and radiances	Level 1	public	ODA / CODA	NRT/NTC
SL_2_WCT___	Sea Surface Temperatures (single view/ channels 2 and 3; dual view/ channels 2 and 3; aerosol-robust/ channel 3)	Level 2	Internal, available to S3VT	ODA	NRT/NTC
SL_2_WST___	Level 2P Sea Surface Temperature (GHRSST like), best SST, quality level and uncertainties	Level 2	Public by end June 2017	ODA/CODA/ EumetCAST	NRT/NTC

Different timeliness :

- **NRT** – Near Real-Time (less than 3 hours)
- **NTC** – None Time Critical (less than one month) – several reprocessing over specific time frames

Access :

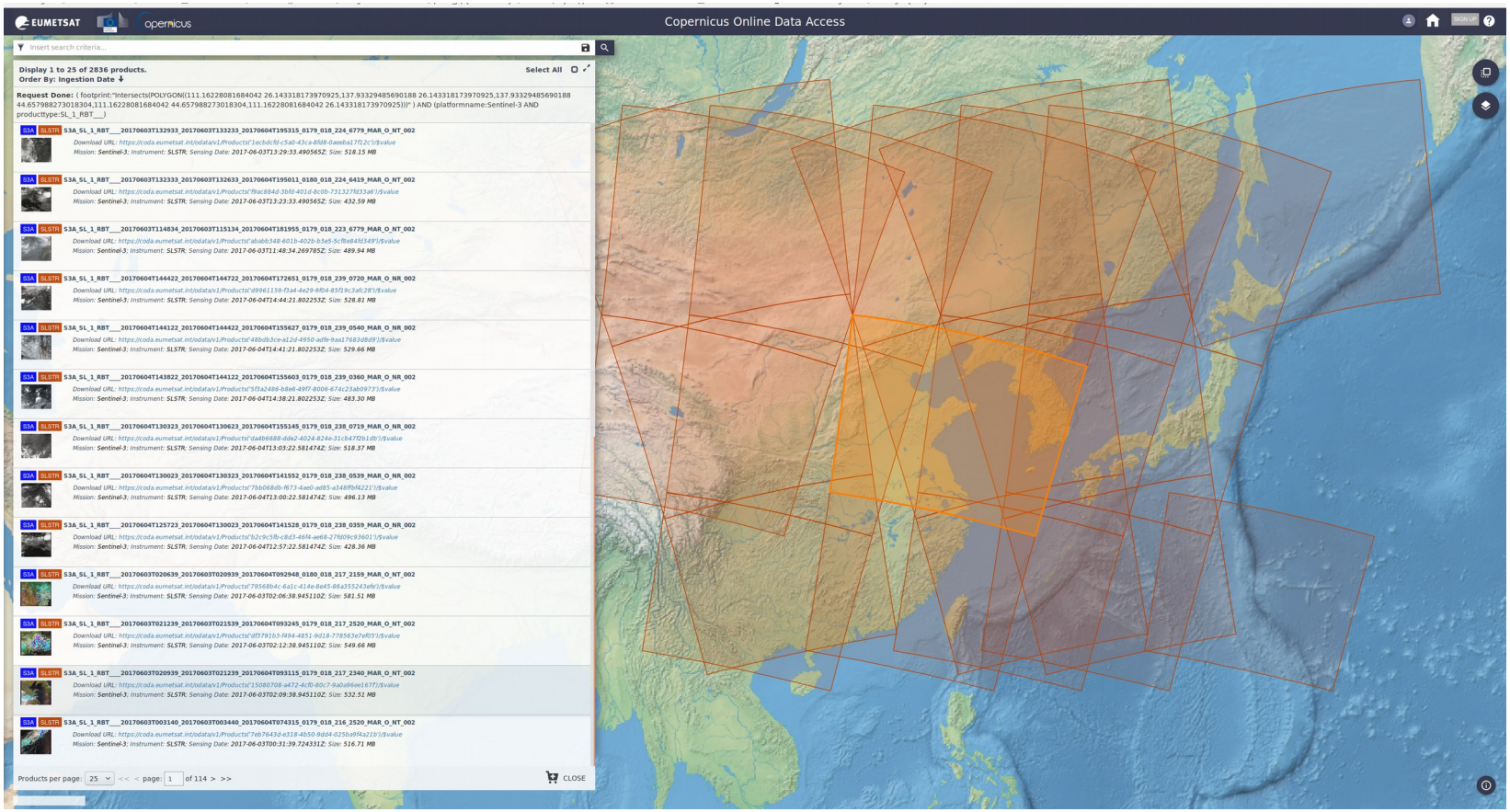
- **ODA** (ftp) : internal access for special users (CMEMS, S3VT, ...) (rolling archive of ~2 weeks)
- **CODA** (http, OpenData / OpenSearch query interfaces) : **http://coda.eumetsat.int** (rolling archive of ~1 year)
- **EumetCAST** (L2 WST only)

https://eoportal.eumetsat.int
Help desk : **ops@eumetsat.int**

See Anne O'Caroll poster

Copernicus Online Data Access (CODA)

http://coda.eumetsat.int



Based on DataHub open source software : http://sentineldatahub.github.io/DataHubSystem/about.html

EUMETSAT

In situ sources

- Benefit on general frameworks:
  - **CMEMS**
    - Integration with Copernicus/CMEMS service for the provision of **moored** and **drifting buoys** and **Argo** data : collection and availability of all data in the same format and quality control
    - Canadian & european GDACs for surface drifters being created
    - Expected improvements in quality control and metadata
  - **in situ radiometer**
    - http://www.shipborne-radiometer.org/
    - High quality data
    - Common format and content has been agreed
    - Shared repository will be soon available assembling all these data
    - Currently used in felyx : cruises from ABoM, NOC, RSMAS and DMI
- All these data formatted in felyx format and available on ftp for ingestion into other MDB (request jfpiolle@ifremer.fr)
- Felyx + in situ data : framework for consistent MDB production for each GHRSST product (which was the ambition in early GDS DPM)

Felyx match-up database workflow

Colocation window : 2h (12h for Argo), 5km

21 x 21 pixel boxes

+/- 6h of in situ data history

In situ data :

- **Copernicus/CMEMS (Coriolis)**
- **ISAR radiometer on opportunity ships (delayed-mode**

Sentinel-3 data :

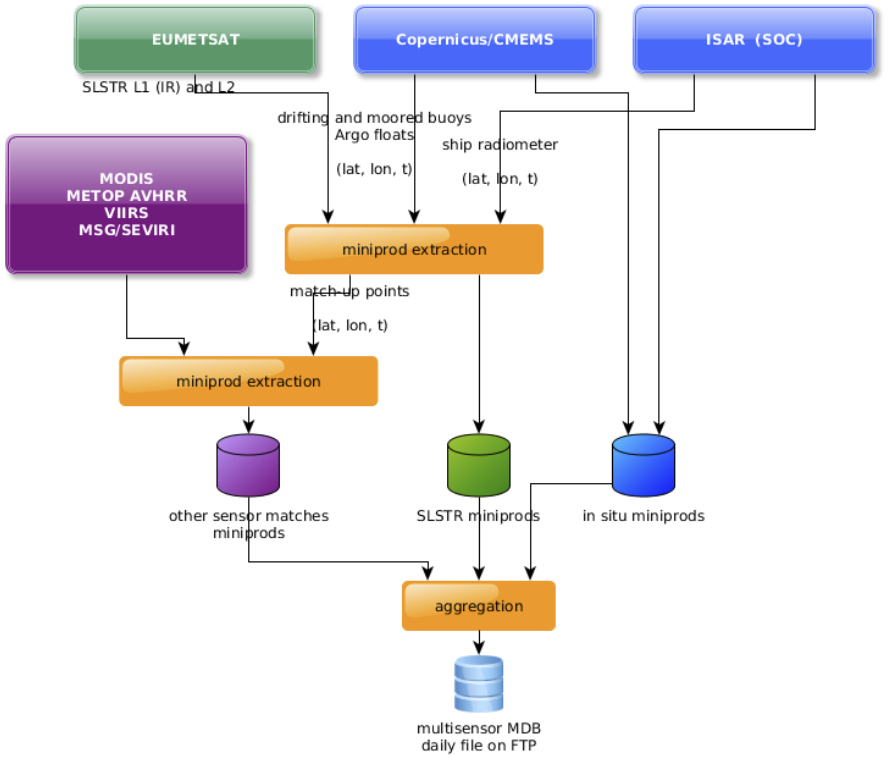
- **L1 infra-red channels**
- **L2 (SST) – all fields, incl. meteo and ancillary fields**

Other sensor data

- **Metop-B/AVHRR, MSG/SEVIRI, OLCI, (MODIS, VIIRS)**

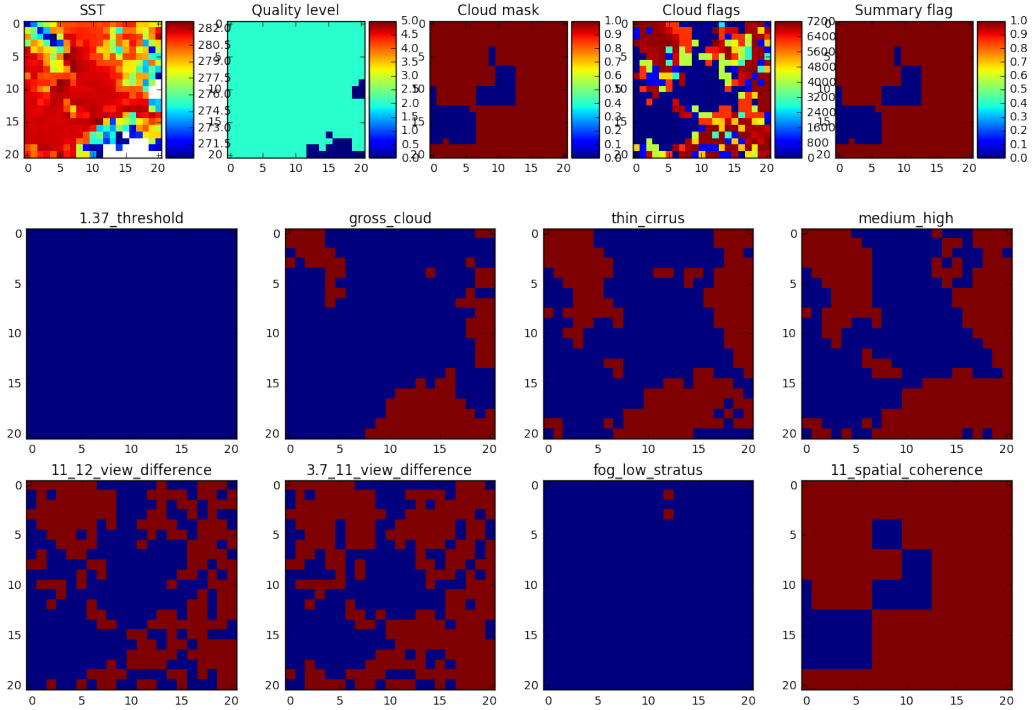
Resampling of all data to SLSTR grid

Daily aggregated match-up files on FTP : stack all matchups into a single file.



Content of match-up

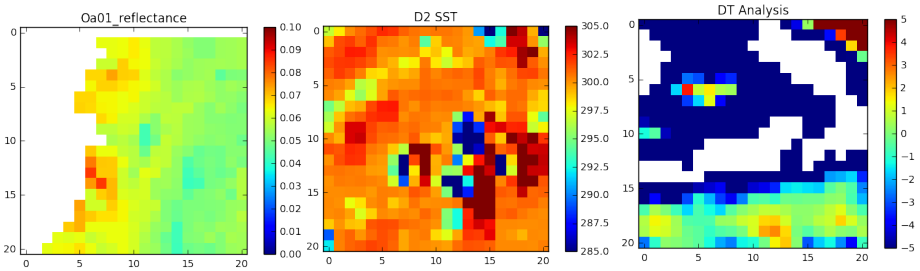
All fields from RBT (L1), WCT and WST (L2)



More than 600 variables from L1 to L2…..

21x21 boxes extracted with all fields for each match-up can be used to test and assess new algorithms or post-processing on a larger scale and time period in a fast way, with in situ information to directly estimate theimprovement.

All fields from cross-overs and complementary files

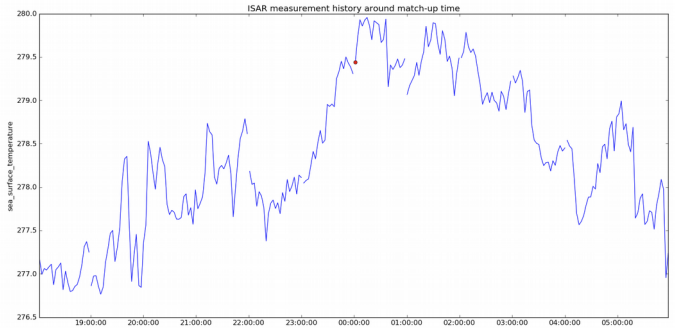


Cross-over fields from OLCI, METOP, VIIRS

Complementary files from post processing of match-ups (prototype SST, quality level, etc…)

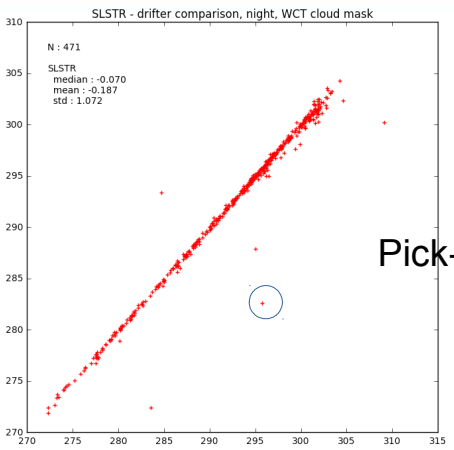
Ancillary fields (OSTIA dSST)

In situ buoy history centered on match-up



EUMETSAT

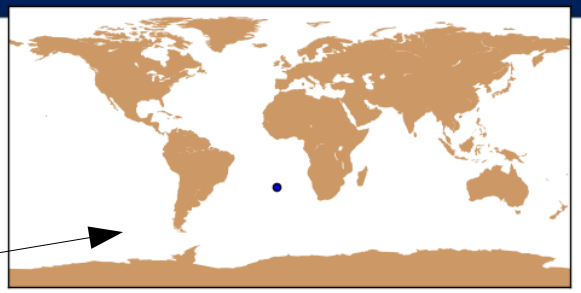
Traceability to source information

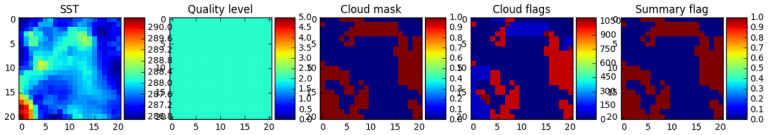


plotly

jupyter

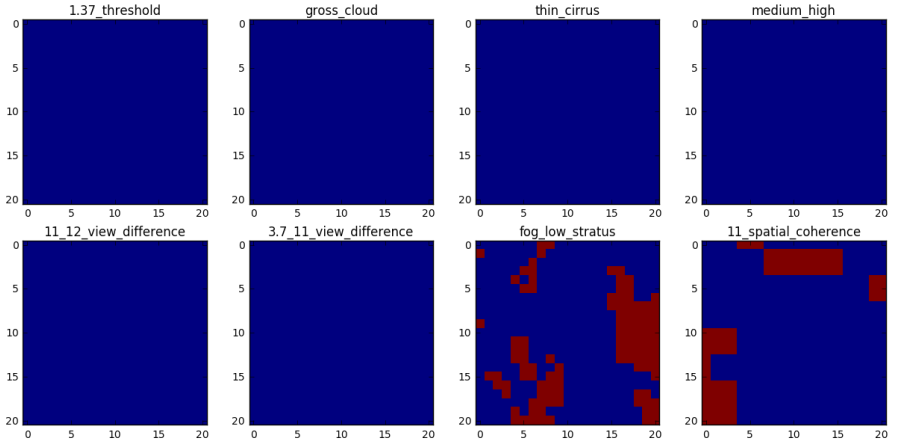
Pick-up outlier #match-up



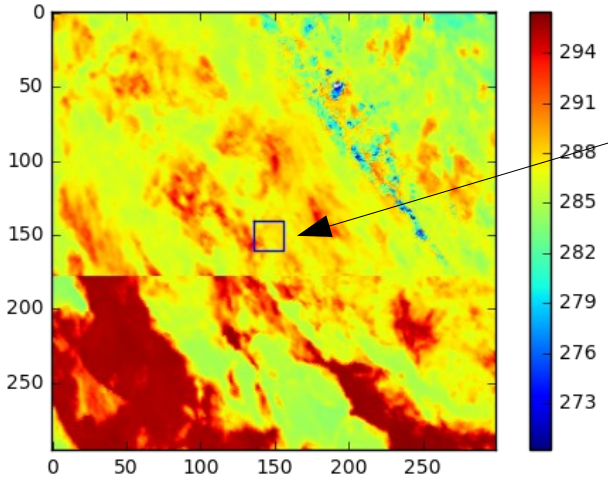


Investigate match-up, dynamic plotting to select individual match-ups

Match-up #reference to source file



Investigate match-up context (full boxes of neighbouring pixels extracted with each matchup)



Investigate original SLSTR file

Full traceability of information in match-up

Traceability analysis example with jupyter notebooks

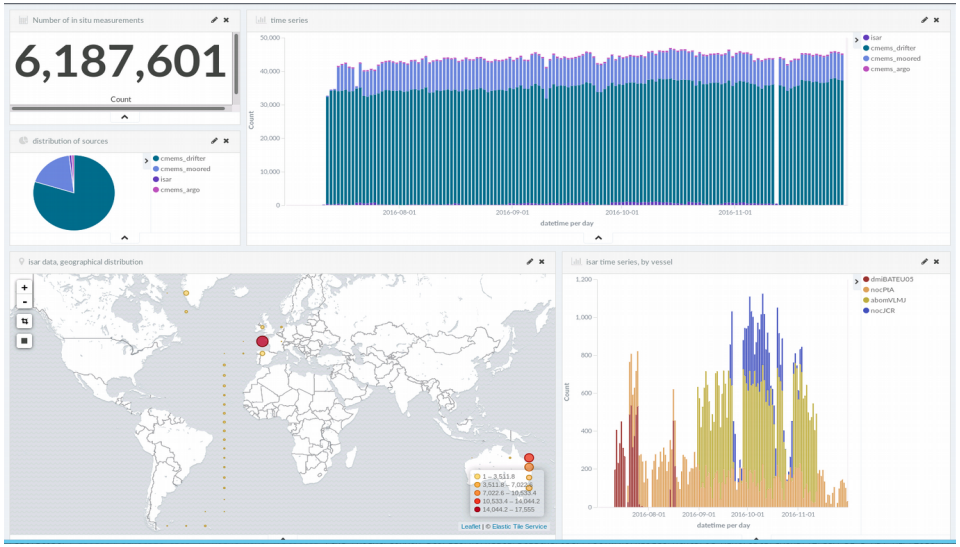
Existing match-up databases

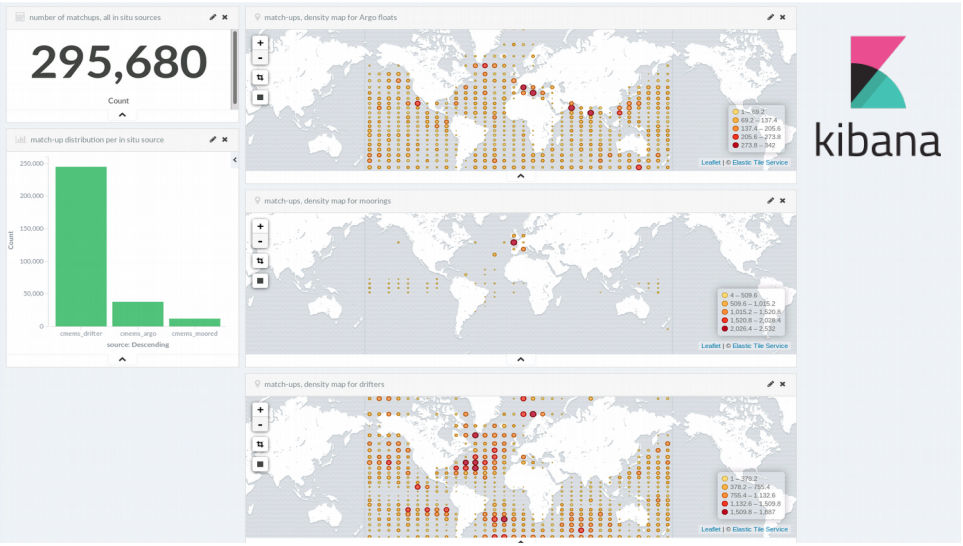
Match-up database	Input products	Complementary products	Availability
OSI SAF NRT **SLSTR** MDB	SLSTR NRT products	METOP (about 50%) SEVIRI (about 30%) SST prototype OSTIA	July 2016 - present
Eumetsat reprocessed **SLSTR** MDB	SLSTR REP v4	OLCI SST prototype OSTIA	July – Nov 2016
Eumetsat reprocessed **SLSTR** MDB	SLSTR REP v5	SST prototype OSTIA	Nov 2016 (*-April 2017)*
*Eumetsat **IASI** MDB*	*Eumetsat & OSI SAF L2P METOP-A and METOP-B IASI*		*June 2017 - onward*
*Eumetsat METOP-B **AVHRR** MDB*	*OSI SAF L2P METOP-B AVHRR*		*June 2017 – onward*
*Eumetsat **OLCI** MDB*	*EUMETSAT OLCI L2*		*TBD*

Planned improvements and changes

- **Data sources**:
  - Collection and ingestion of more in situ radiometer measurements (ISAR, M-AERI,…) though mostly delayed mode
  - HR drifters
- **Content**:
  - SLSTR : visible channels (probably in complementary files to preserve native resolution) + additional cross-over for ice temperature, other sensors
  - SST depth adjustment
  - Extended box size for ship measurements (shipborne radiometers)

Match-up content statistics





Typical match-up distribution for SLSTR, all weather conditions :

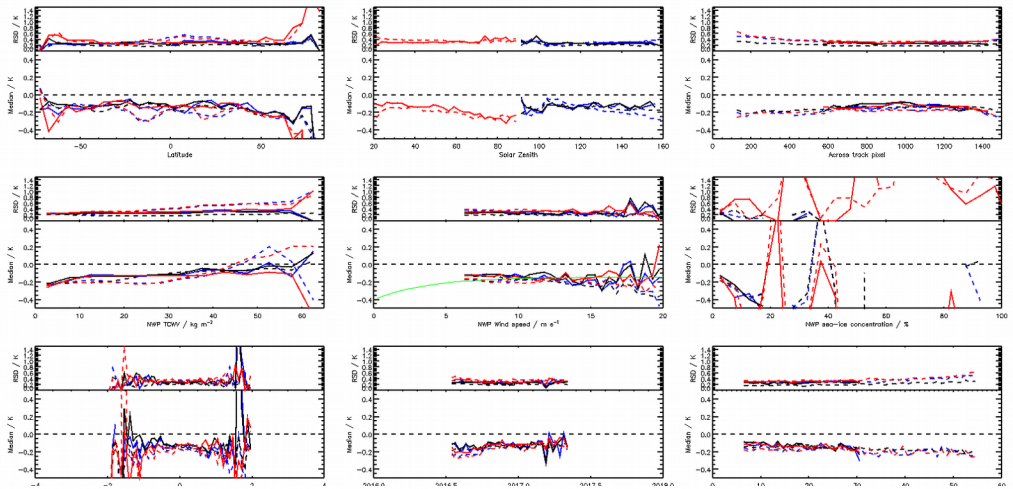
- more than 40.000 in situ measurements per day

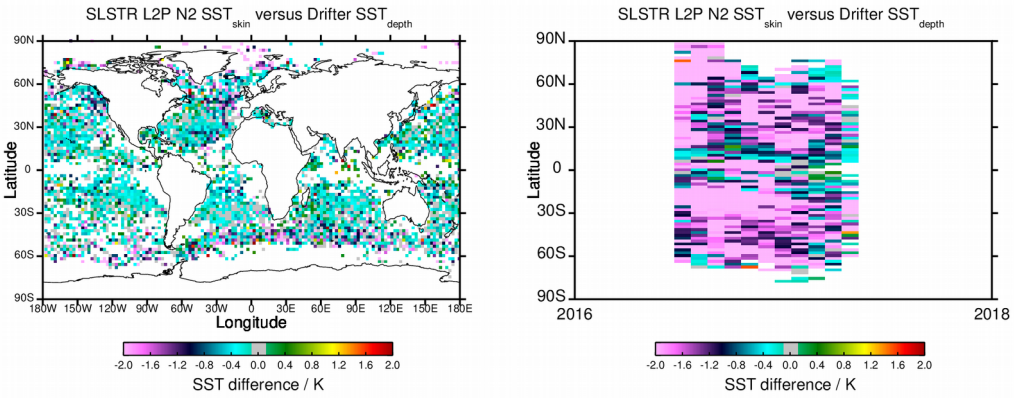
- ~2000 match-ups / day for buoys
- ~350 match-ups / day for moored buoys
- ~600 match-ups / day for argo floats

EUMETSAT

Application of SLSTR MDB(s)

- Used by different groups at Eumetsat, within S3VT and MPC Sentinel-3 for SLSTR
- major asset in:
  - L1 cloud screening validation (RAL)
  - L2 SST coefficient estimation (Univ. Of Reading)
  - L2 Quality level stratification and uncertainties estimation (Univ. Of Leicester)
  - SST validation : OSI SAF (Meto-France / DMI / MetNo), NOAA, Eumetsat
  - Metis intercomparison framework

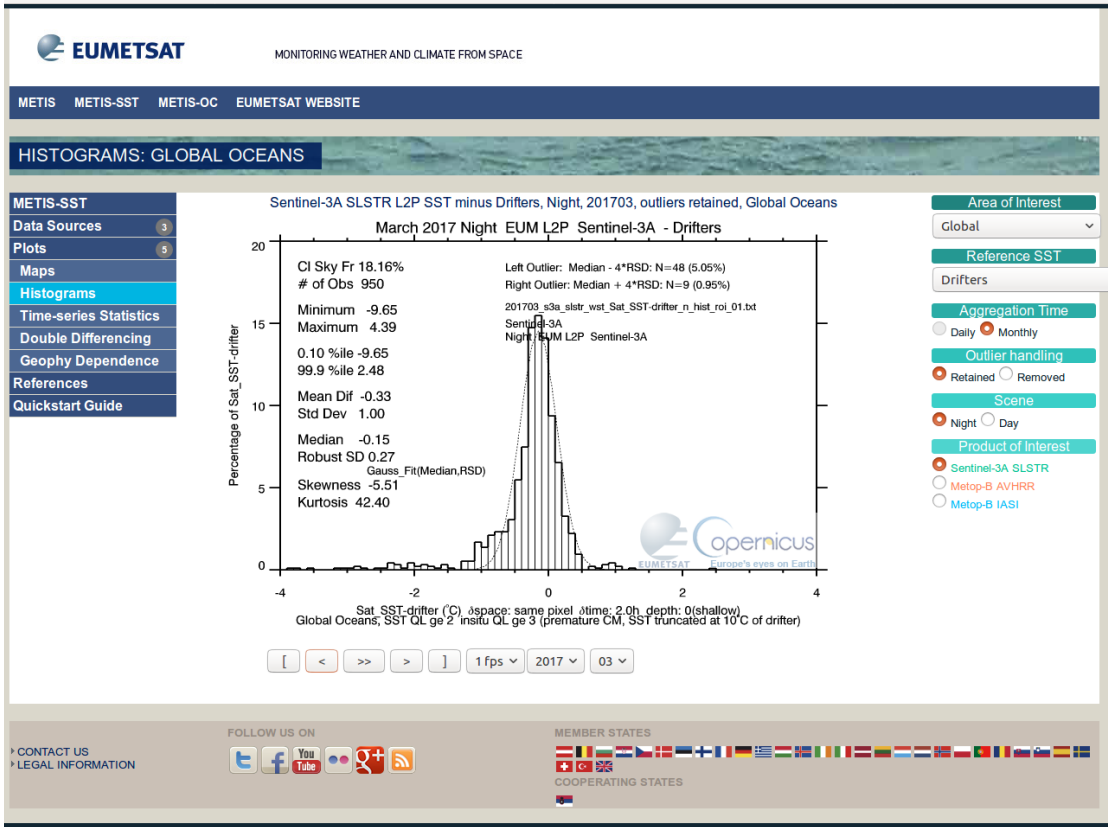




Courtesy: G.Corlett, Univ. Of Leicester

Quality monitoring statistics to be updated periodically for control and monitoring

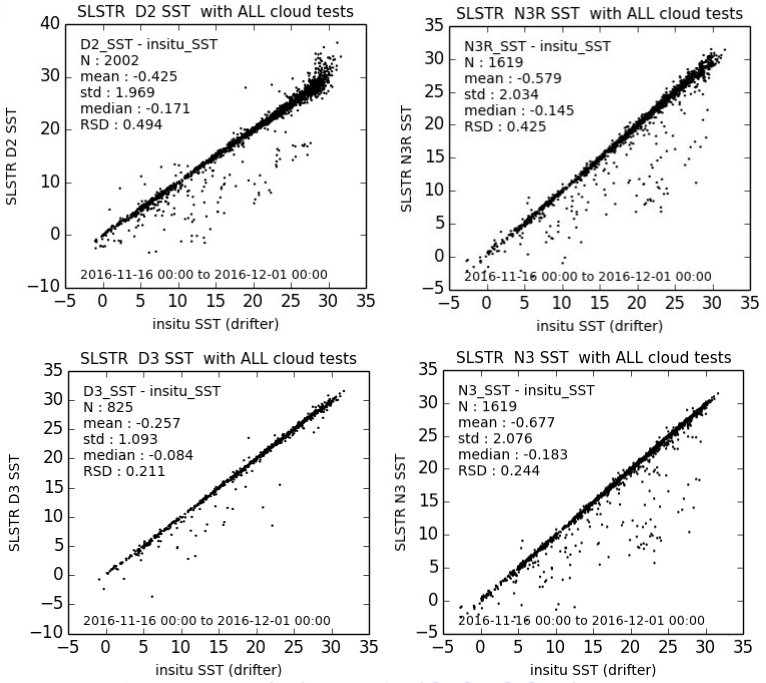
Metis monitoring interface



See Prasanjit Dash poster

OSI SAF SLSTR federated activity

- Funded by Eumetsat
- SST experts from Ocean & Sea ice SAF (Meteo-France, DMI and MetNo)
- Global assessment and specific on high latitudes with in situ data collection from ISAR in situ radiometer onboard arctic sea cruise and drifters + sea ice temperature
- Based on felyx generated match-up databases



(WCT_SST – insitu_SST) with cloud clearing REC

algo	N	mean	std	median	RSD
N2	5293	-1.520	3.449	-0.330	0.560
N3R	1911	-0.617	2.098	-0.146	0.430
N3	1911	-0.727	2.165	-0.192	0.245
D2	2653	-0.735	2.654	-0.188	0.561
D3	966	-0.294	1.179	-0.096	0.209

26586 cases, 5299 clear cases 19.9 %

(WCT_SST – insitu_SST) with cloud clearing ALL

algo	N	mean	std	median	RSD
N2	4130	-0.973	2.608	-0.296	0.479
N3R	1619	-0.579	2.034	-0.145	0.425
N3	1619	-0.677	2.076	-0.183	0.244
D2	2002	-0.425	1.969	-0.171	0.494
D3	825	-0.257	1.093	-0.084	0.211

26586 cases, 4133 clear cases 15.5 %

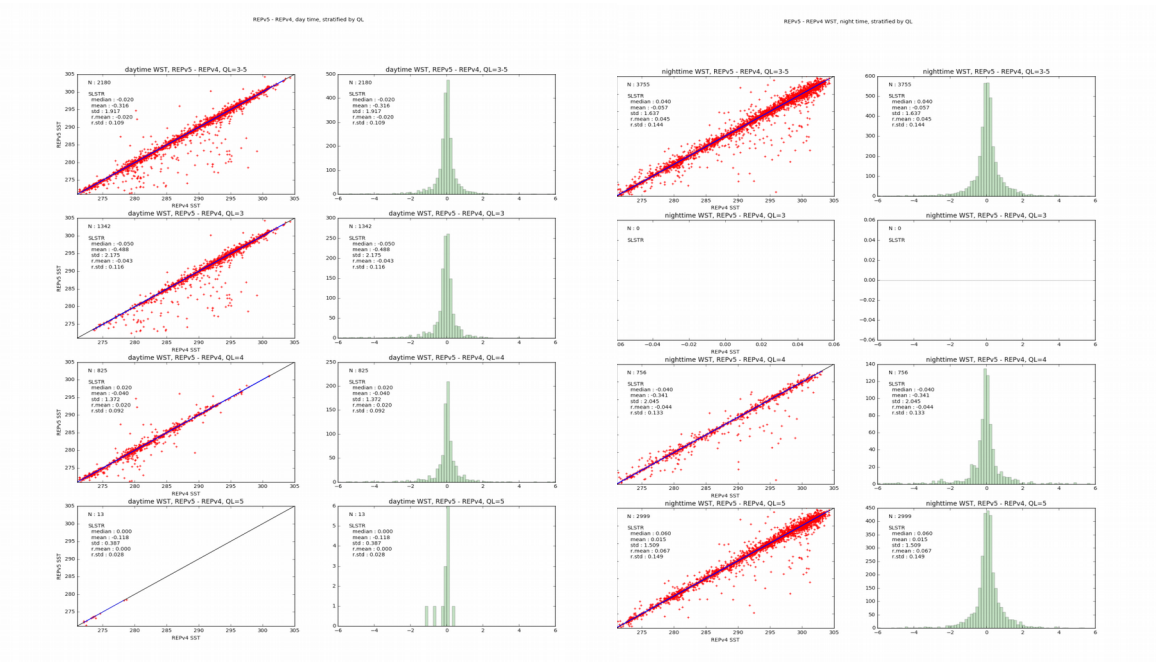
No correction : skin WCT SST vs bulk insitu SST

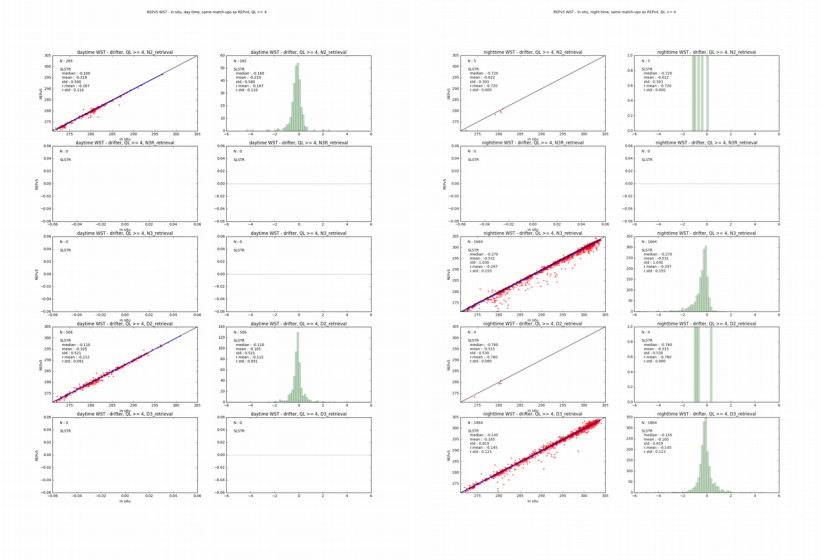
Courtesy: Anne Marsouin, Meteo-France

Intercomparison of MDBs

- Assessment of algorithm improvements

All match-ups are uniquely identified through buoy id and time and location : this makes easy to intercompare different versions of product with each other, through « match-ups of match-ups » (left) or respective comparison of each version to the same in situ values (right)

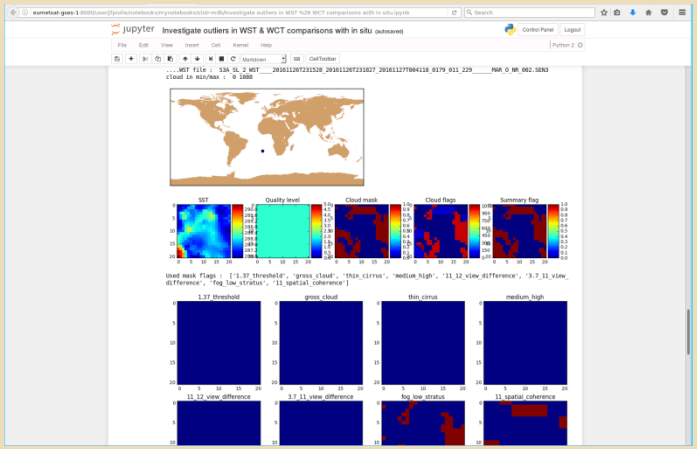




Comparison of reprocessing v5 vs v4 for SLSTR SST product

Integration with open source analytics

jupyter



Data and metrics extracted with felyx can be analysed locally or **remotely** with **jupyter** python notebooks (http://jupyter.org)

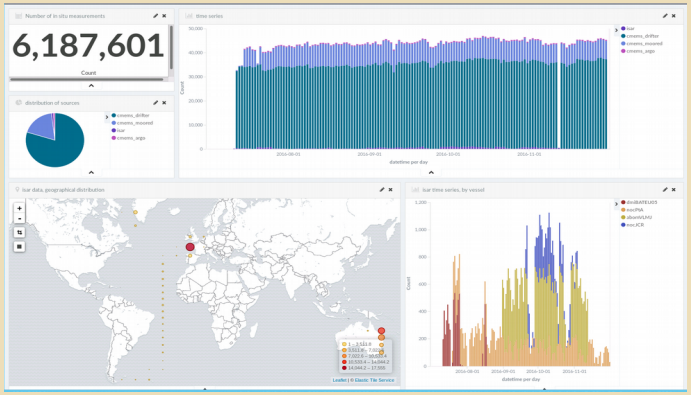
Code and images embedded in web pages

Dashboard pages capability

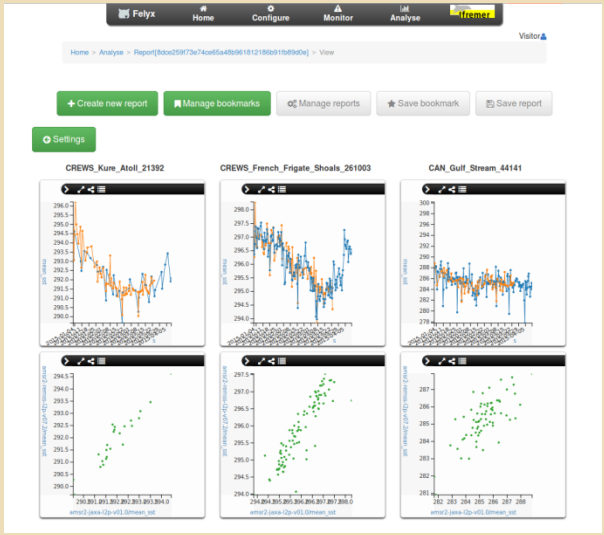
kibana

Grafana

x-pack



On-the-shelf analytics such as **X-pack**, **Kibana** or **Grafana** can be interfaced with felyx leveraging on the use of ElasticSearch for all metrics and metadata : visualization of statistics, dashboards, alerting



Felyx natively embeds a web front-end with plotting capabilities for the match-ups and miniprods

Ability to design reports, automate share them through a repository

Integration in processing environment

Deployment with **ansible** (https://www.ansible.com)

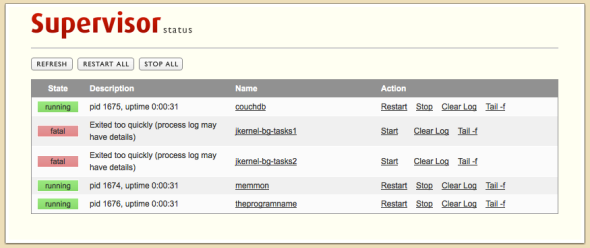
Ansible is an IT automation tool

Deployment procedure described in a playbook

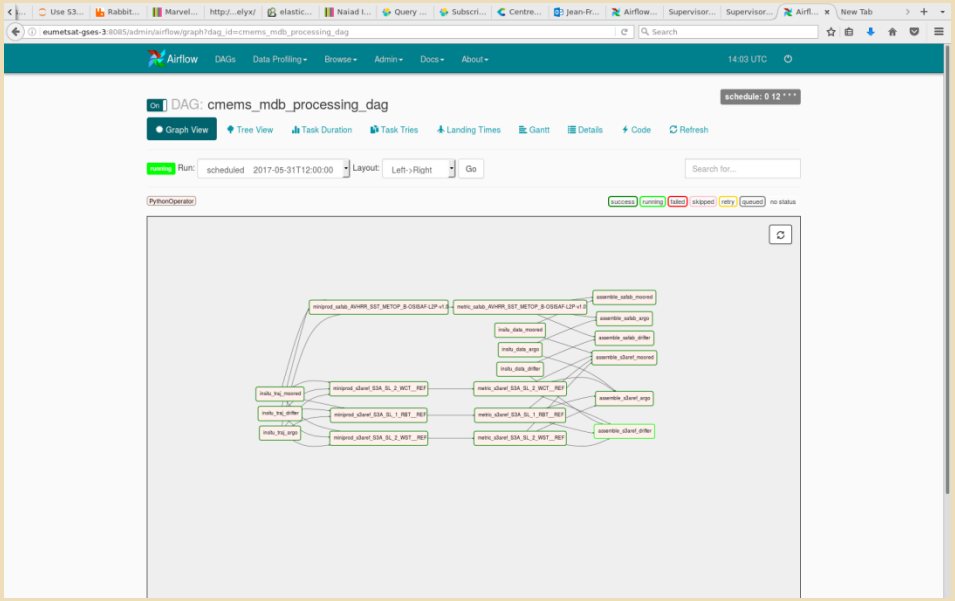
Description of processing framework in a configuration file (hosts, storage, roles of each host, etc…)

Automatic execution of installation

Playbook for ubuntu to be released this summer for easy felyx deployment



Felyx daemons supervised and run by **supervisor** (http://supervisord.org)



Airflow (https://airflow.incubator.apache.org) is a task scheduler

Processing workflow, from in situ data ingestion to match-up assembling, can be integrated in such system for automated MDB production

EUMETSAT

conclusion

- Felyx is an open framework for data intercomparison : open source but also integrating with other on-the-shelf tools and frameworks
- It is evolving with usages, some components more developed than others, growing collections of mappers for different product/formats
- Metrics and diagnostic part still to be further exploited

- Proved a major asset in Sentinel-3 cal/val activity

- Felyx + common in situ dataset + common diagnostics
  - Possible way forward to build consistent (format and content) match-up datasets in GHRSST
  - sharing of MDBs for better sensor intercomparison and improvement
  - Application to climate data record assessment (using CCI dataset)
- Possible group or task team on shared open source tools relevant to GHRSST data validation and usage

Thanks to ESA, EUMETSAT and Ifremer for supporting felyx development

Access

- **Contact**: Jean-François Piollé (jfpiolle@ifremer.fr)
- Web site: http://hrdds.ifremer.fr
- Documentation: http://felyx.readthedocs.io/
- Packages: https://felyx.cersat.fr/download/source/1.0.0/
- Source code: https://git.cersat.fr/groups/felyx
- Virtual machines for testing (virtualbox):
  - Bare configuration: ftp://ftp.ifremer.fr/ifremer/cersat/projects/felyx/download/vm/2016-09-19_felyx-1.0.0.ova
  - Pre-configured for OLCI & SLSTR datasets: ftp://ftp.ifremer.fr/ifremer/cersat/projects/felyx/download/vm/felyx_olci.ova

- Significant improvements to be released in July : deployment procedures (ansible), documentation and additional resources for match-up database generation
- **Contributions and shared developement is welcome and encouraged!**