



# A tool for the quantitative assessment of long time series of satellite SST

## A prototype for CDAF framework implementation

Jean-François Piollé (Ifremer), Jon Mittaz (Univ. Of Reading), Prasanjit Dash (NOAA)

Processing of long time series of sea surface temperature (SST) is now undertaken by different agencies for a large range of sensors (AVHRR series, (A)ATSR series, MODIS) acquiring routine measurements of SST since the early 80s. One of the motivations of the GHRST Climate Data Assessment Framework (CDAF) was to define a set of common metrics to assess the **uncertainties, stability and sensitivity** of these datasets, evaluating their suitability for detecting changes and trends and therefore their use as Climate Data Records (CDR).

To achieve this assessment in a **simple, fair and consistent** way, we are designing an **open-source prototype** tool that take as input any GHRST formatted dataset to produce assessment diagnostics, as defined by various groups and projects (CDR, TRAC, CCI SST), based on comparisons with reference data (GTMA for temporal stability assessment, drifting buoys and Argo floats for spatial assessment) integrated with the tool. The current prototype of this tool is based on **felix** and **jupyter notebooks**, providing producers with a complete solution to evaluate the quality of a dataset with respect to these common metrics and to allow them to complete the CDAF (Climate Data Assessment Framework) in a consistent manner. We used here for demonstration AVHRR Pathfinder v3.5 as a candidate test dataset.

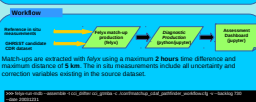
### The tool

The CDAF prototype tool was implemented by quickly assembling existing open-source solutions to deliver an end-to-end proof of concept of an integrated system (tools and needs) allowing a fair and objective assessment of different candidate CDR datasets.

**Scholarx**

**felix** (<http://mads.dtu.dk>) is an open-source framework for geophysical feature extraction and match-up production.

The **Jupyter Notebook** (<http://jupyter.org>) is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



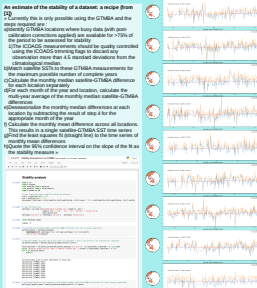
### The reference data

Several diagnostics in the CDAF are based on comparisons with in situ data, such as moored buoys (from GTMA), drifters and Argo floats. We used the **ESA SST CCI Independent Reference Data Set (IRDOS)** which covers the 1980-2016 time frame. The compiled data went through extensive quality check by the SST CCI team and we selected here only the measurements with quality level 5. All variables from the in situ dataset are retained in the final match-ups and can be used to refine the  $\kappa$  grid = match-up selection.



### Assessment diagnostics

Several diagnostics, described in above references, were defined by the GHRST Climate Data Record working group (CDR WG) addressing uncertainties, consistency, systematic and non systematic differences or stability. They can be implemented collaboratively as generic functions in **Jupyter notebooks**, and use raw match-up datasets processed at the previous step. For this prototype, we implemented the **stability diagnostics**, which can now be tested on different datasets, and be further refined.



### Application to Pathfinder dataset

For the demonstration, the CDAF prototype tool was run over NOAA Pathfinder v3.5 dataset, spanning from August 1980 to December 2014.



**Conclusions and perspectives**

We demonstrated a generic solution to perform a fair and shared assessment of GHRST CDR datasets. More precisely this can be done from satellite and in situ data, reference data and limited high level diagnostic frameworks for the analysis of datasets from various sensors, and a possible way forward for GHRST project to foster interactive and collaborative science, and investigate specific tasks with users.

- Improvements for a final open solution would include:
- Other diagnostics need now to be implemented, tested and refined, as well as the match-up filtering and in situ data correction (through Jupyter notebooks)
  - Usage of Jupyter notebooks for easy publication and interpretation of each CDR dataset evaluation and providing a space for sharing and publishing results
  - Some optimizations in the match-up processing (could be reduced by half)
  - Packaging a complete standalone solution on secure, scalable servers and in situ data or setting up a central service in a candidate host (software as a service infrastructure)
- We propose to iterate on this interactively and collaboratively working remotely on these preliminary Jupyter notebooks and pre-processed match-ups, hosted at Ifremer. Other CDR datasets can be added too. Match-up datasets and notebooks are also available for download for your local usage.