The concept of Interim Climate Data Records and its Pros & Cons

Jörg Schulz[1], Jaap Onderwaater[1], Marco Marquard[2], Oliver Sus[1], Mike Grant[1], Timo Hanschmann[1], Marie Doutriaux-Boucher[1]

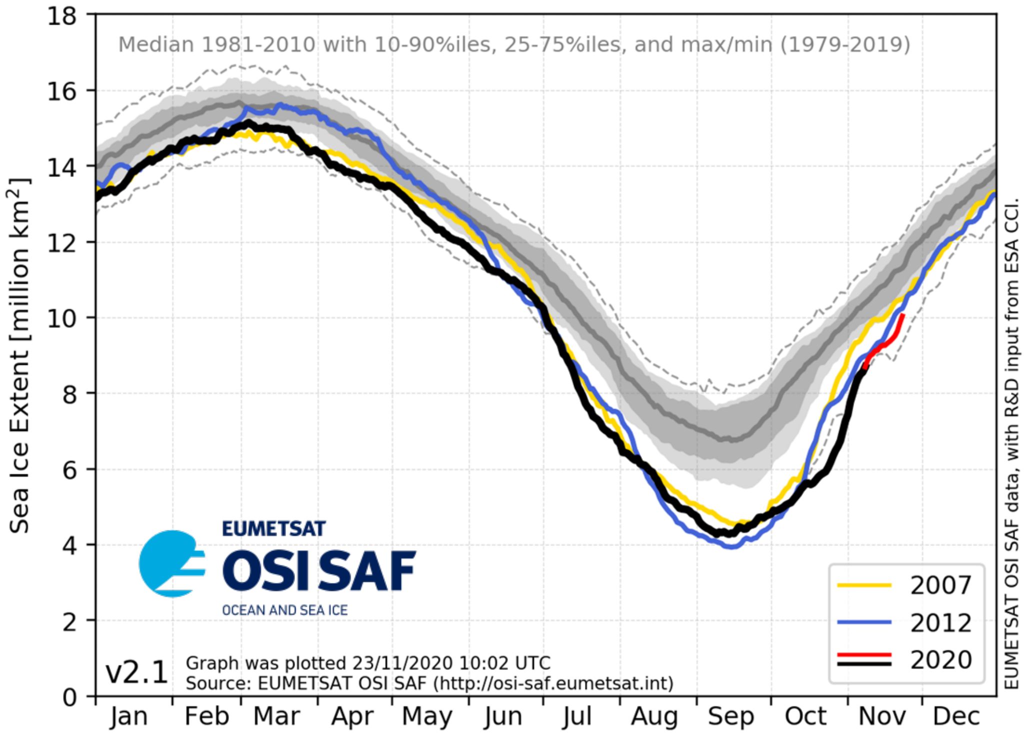
[1]EUMETSAT, [2]Deutscher Wetterdienst

EUMETSAT Contribution to Climate Monitoring

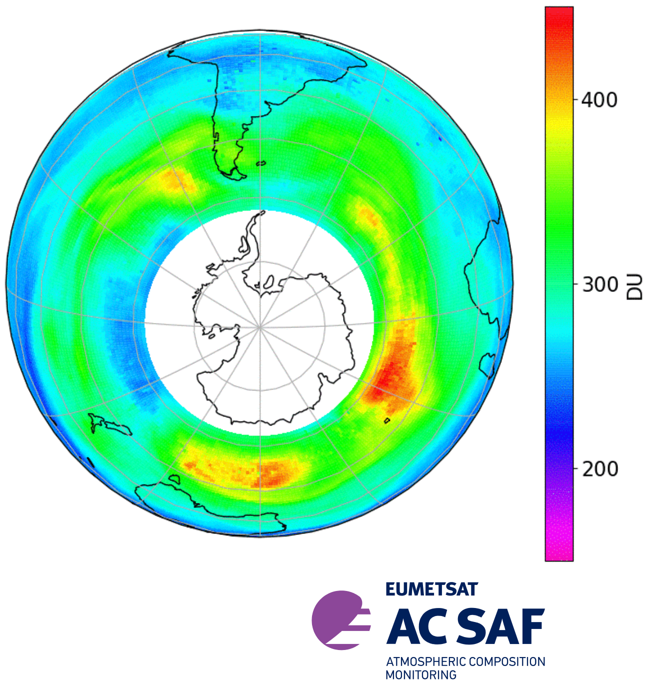
- Long term, multi-satellite programmes, with service continuity
- Continuous improvement, expansion of portfolio of observations
- Unique patrimonial archive: decades of observations
- Data rescue (historic satellite observations) aiming at going back in time as far as possible;
- Recalibration and production of climate records identifying and preparing satellite data of best possible quality:
  - Physical parameters directly observed by satellites: level 1 (mostly done at EUM HQ)
  - Geophysical parameters: GCOS ECVs (ocean, atmosphere, land) (mostly done by EUM SAF)
  - Estimation of uncertainties
- Data access
- Cooperation with users: validation, research, applications
- Training, support to climate-related capacity building initiatives

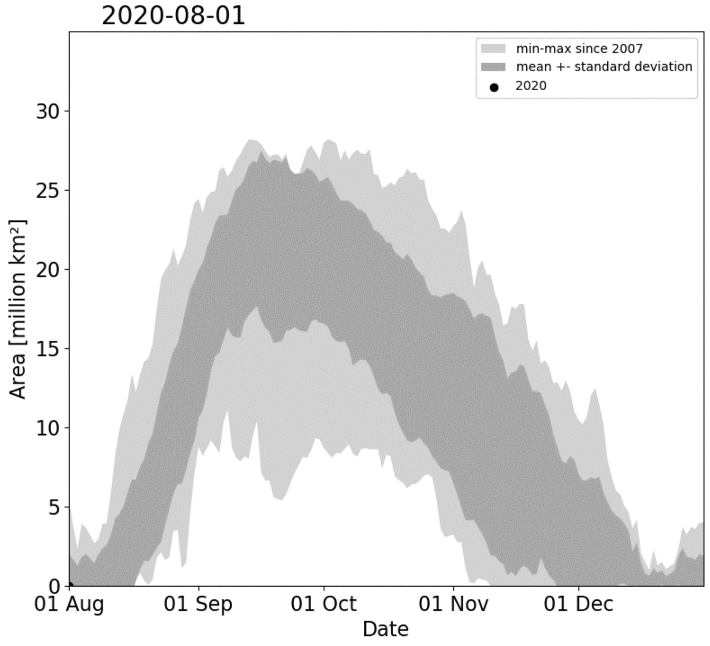
Real time climate monitoring à la EUMETSAT

Arctic Sea Ice Extent



Ozone Hole versus climatology





Ozone hole: Total ozone column <220 DU, which is about 30% below average.

EUMETSAT

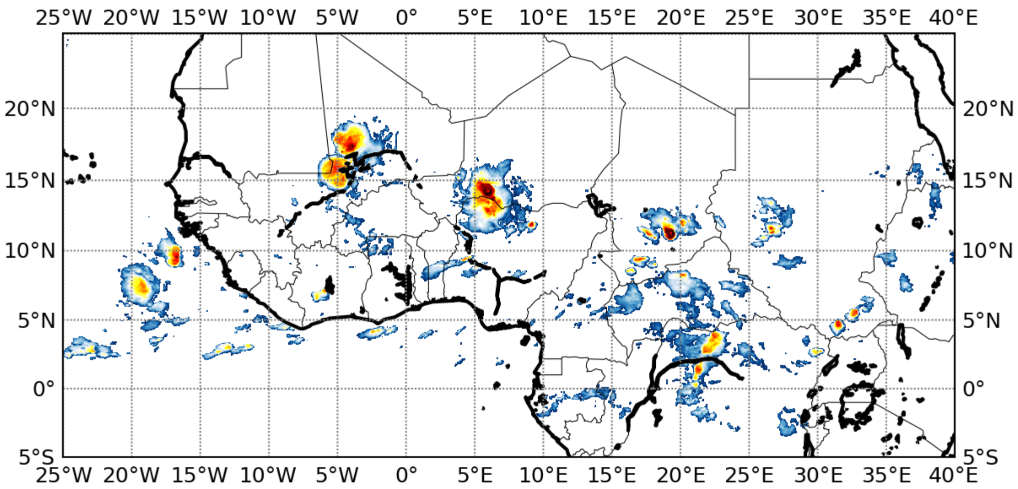
Use cases: Climatology for convective systems from geostationary satellite data

- Observation of the high cloud clusters from geostationary infrared data
- Application of the TOOCAN algorithm on the 31-year homogenised METEOSAT IR data accomplished
- Can run in ICDR mode as well

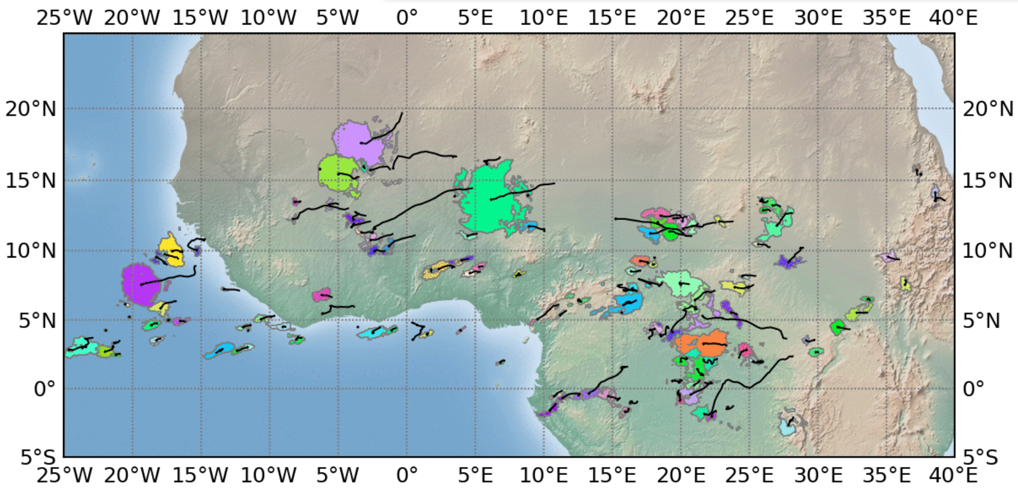
LEGOS TOOCAN cloud tracking algorithm

- pattern recognition and tracking algorithm
- Detection of high cold cloud cover by applying a 235K threshold

1999/07/10-01



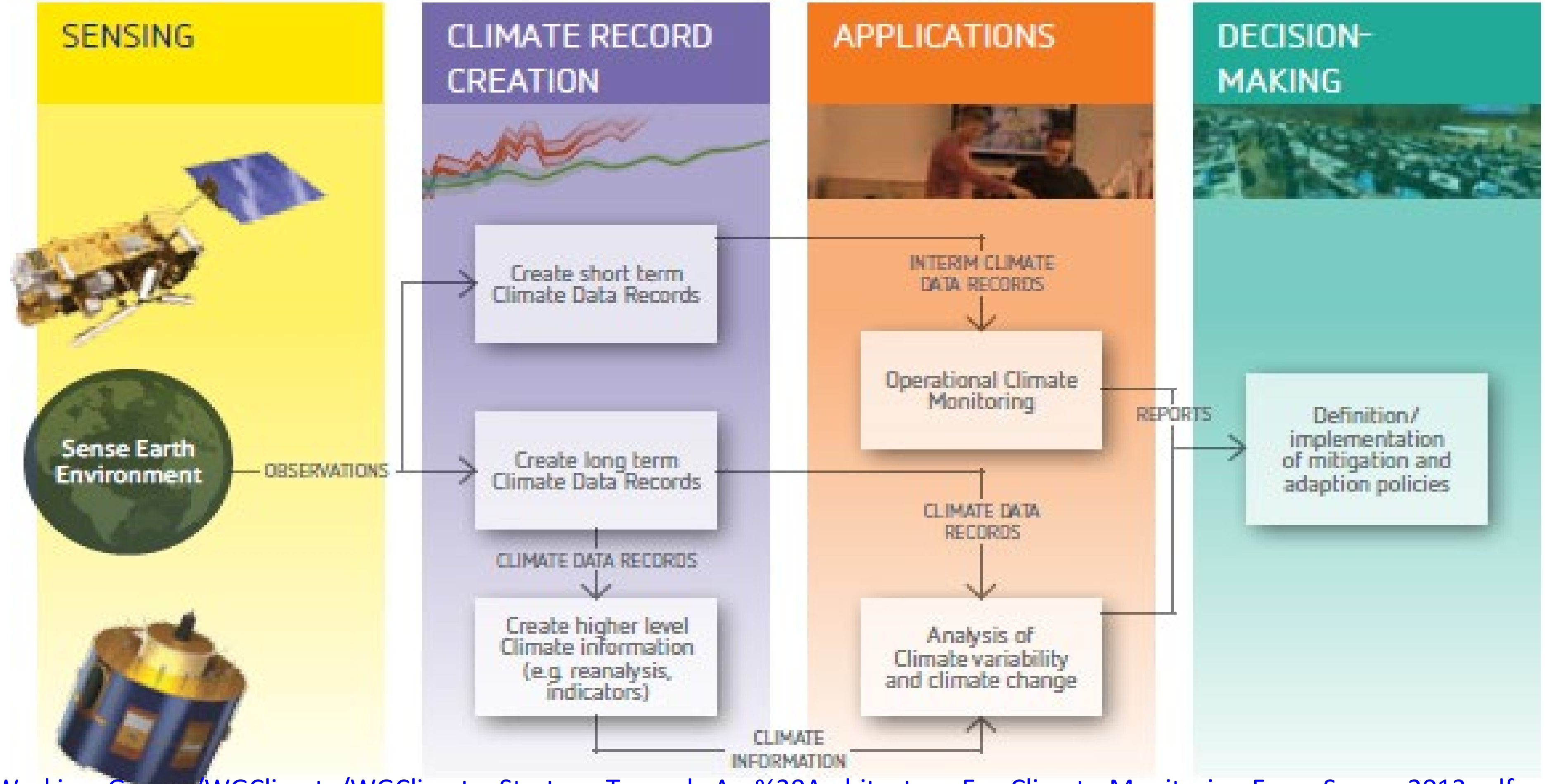
Brightness temperatures



Content courtesy of Thomas Fiolleau, LEGOS

EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

The Architecture for Climate Monitoring from Space



http://ceos.org/document_management/Working_Groups/WGClimate/WGClimate_Strategy-Towards-An-%20Architecture-For-Climate-Monitoring-From-Space_2013.pdf

WGClimate: Proposed Definitions for Data Records

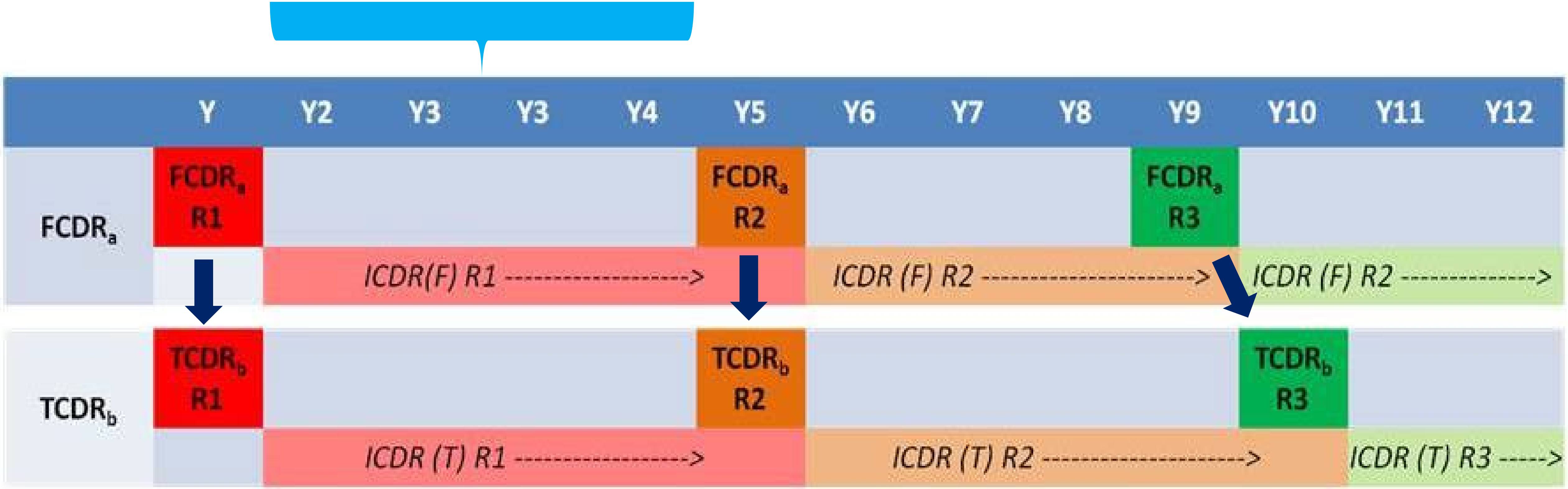
- An FCDR consists of a consistently-processed time series of uncertainty-quantified sensor observations calibrated to physical units, located in time and space, and of sufficient length and quality to be useful for climate science or applications.

- A CDR consists of a consistently-processed time series of uncertainty-quantified retrieved values of a geophysical variable or related indicator, located in time and space, and of sufficient length and quality to be useful for climate science or applications.

- An Interim Climate Data Record (ICDR) is consistently-processed times series of uncertainty-quantified estimates of CDR values produced at lower latency [*or higher timeliness*] than, but otherwise minimizing differences with, the estimated CDR values.

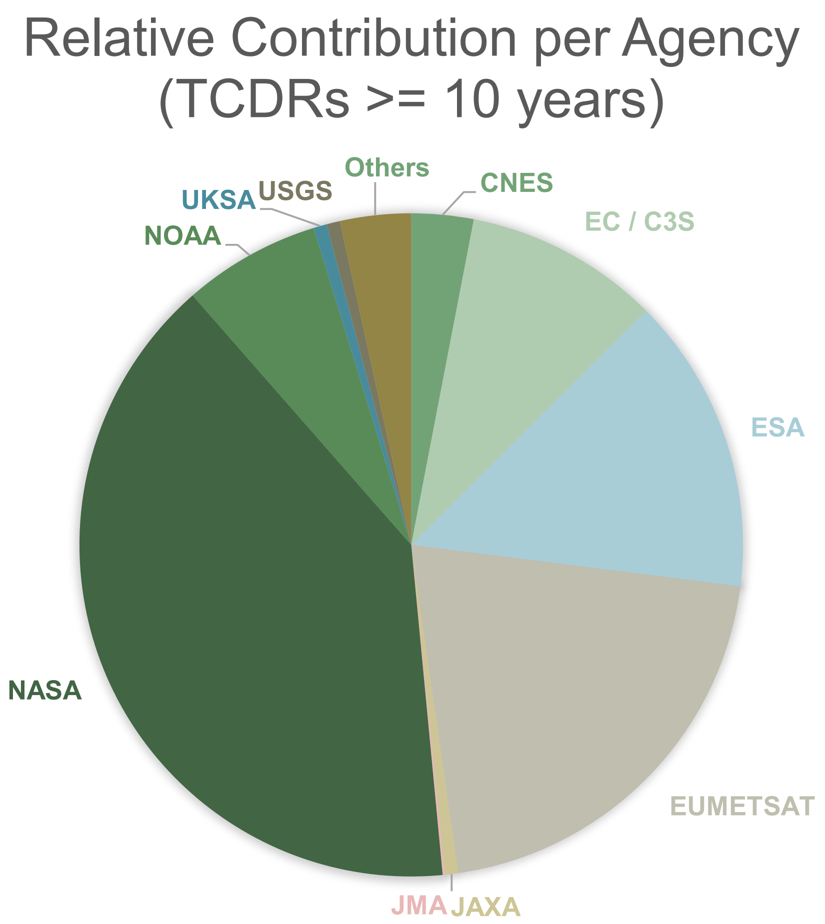
CDR/ICDR Interplay

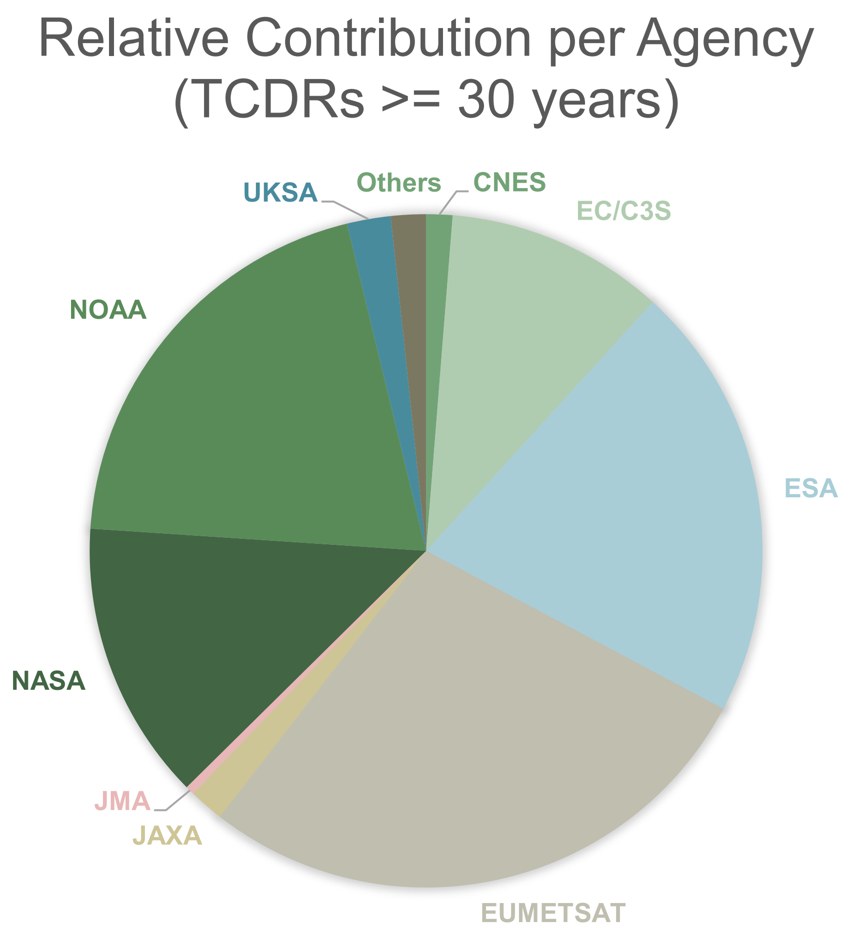
How many years can we maintain an ICDR?



EUMETSAT

ECV CDR Inventory and Gap Analysis





ECV Inventory contains:

- 72% CDRs
- 28% ICDRs

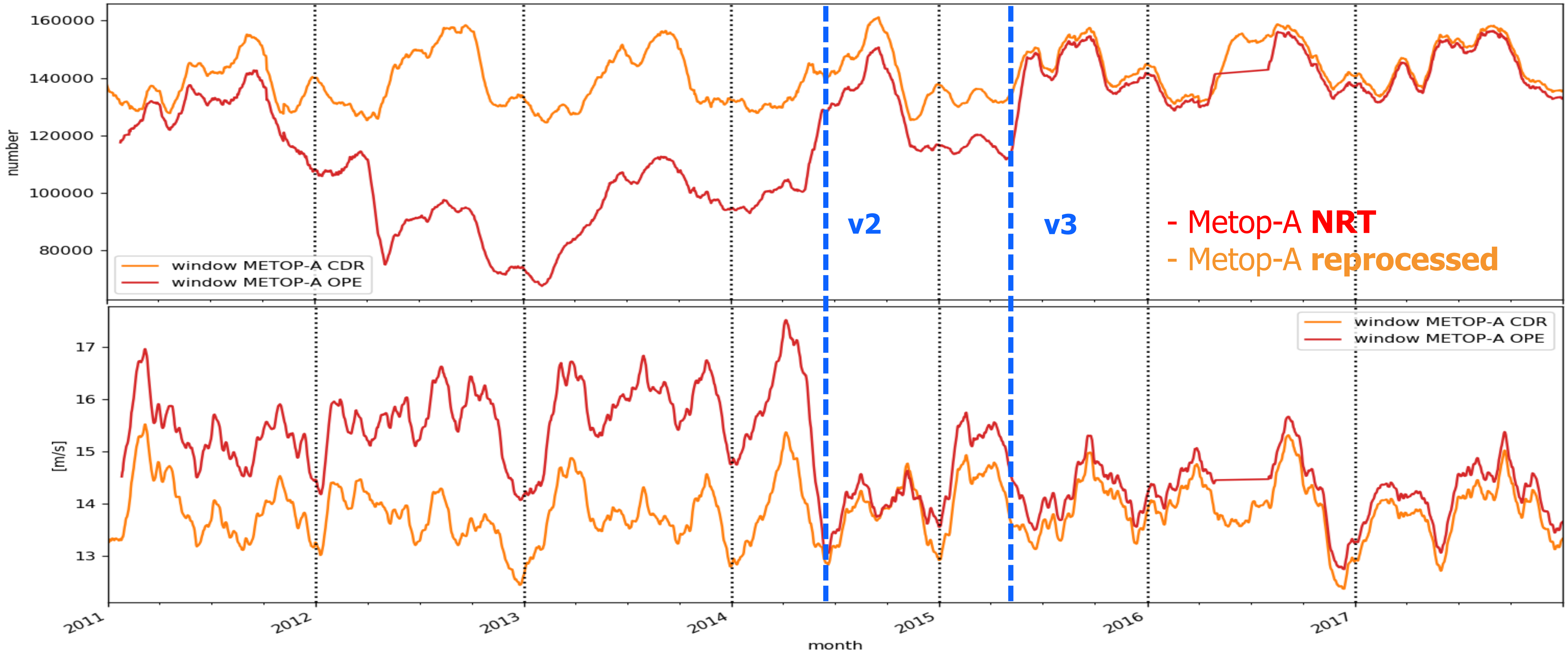
CDR length:

- 27% < 10 years
- 73% >= 10 years
- 33% >= 20 years
- 18% >= 30 years

Maximum length possible: ~50 years.

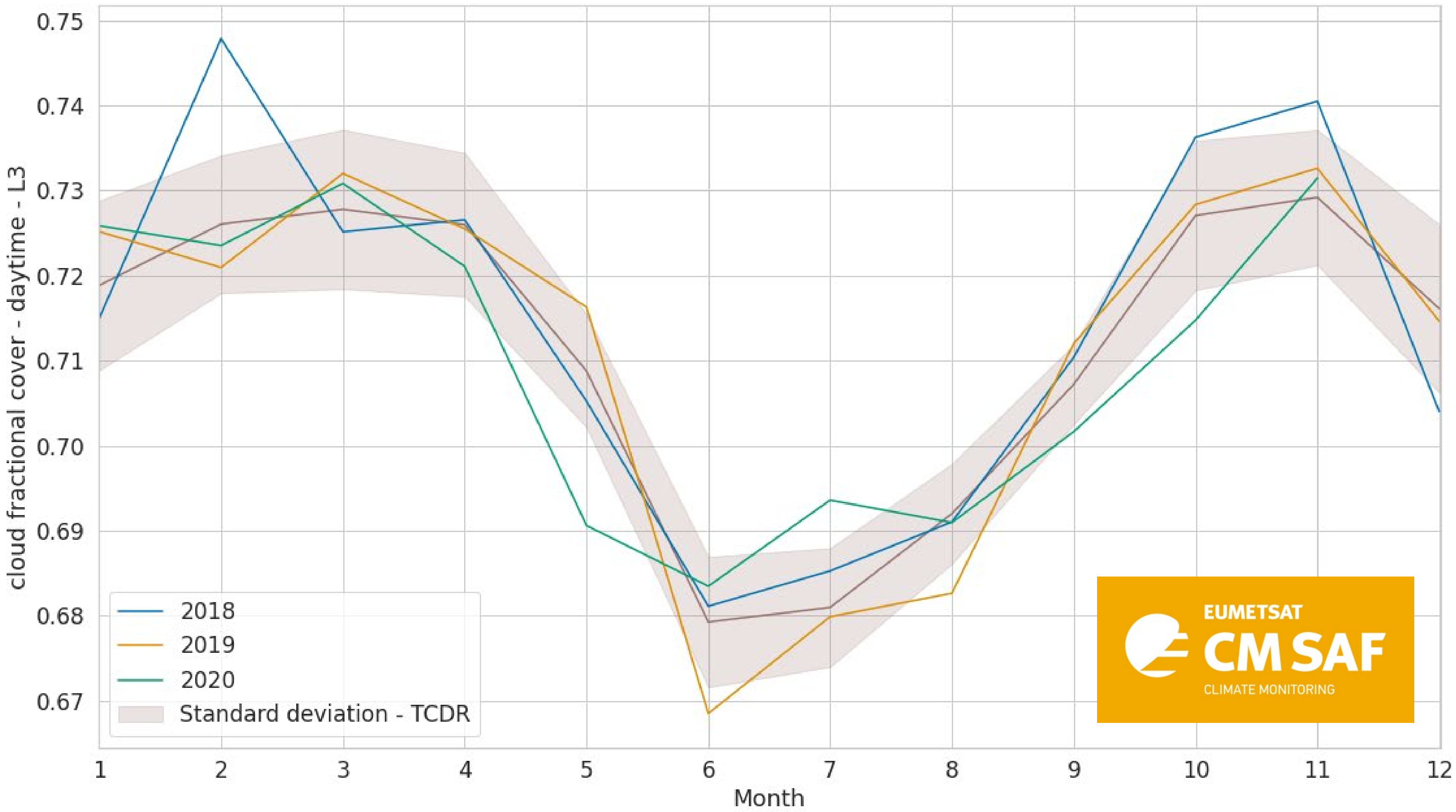
Do we need an ICDR?

Metop A/B Atmospheric Motion Vectors over polar areas derived from AVHRR



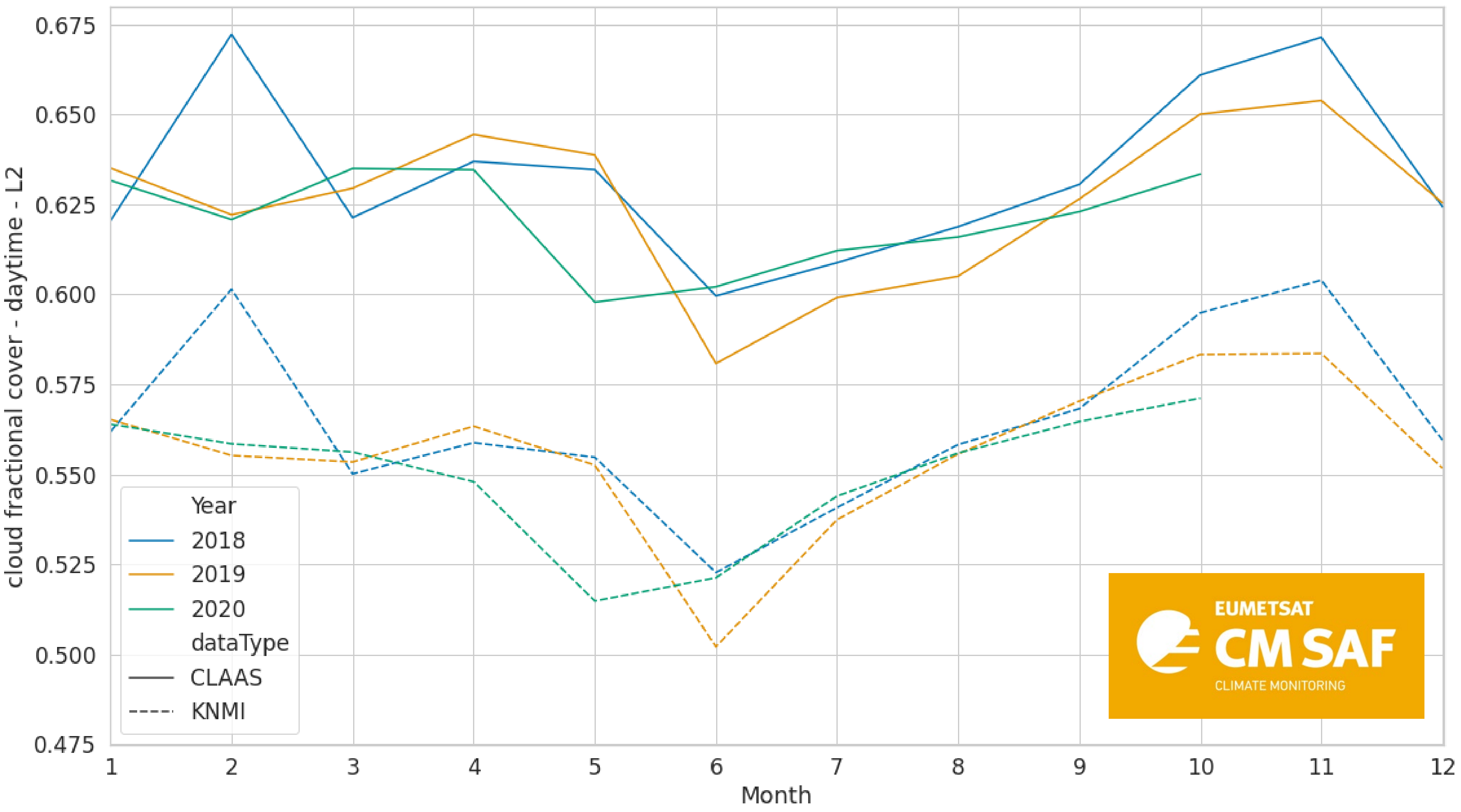
EUMETSAT

CM SAF CLAAS CDR vs. ICDR – Cloud fraction



EUMETSAT

CM SAF CLAAS CDR vs. KNMI EDR NRT – Cloud fraction



What is an ICDR technically?

- A CDR is for a set time period
- An interim CDR is an ongoing incremental **continuation** of a CDR
- “Identical” code, aux data, processing
    - Minor changes that don’t break continuity are always acceptable
    - Major changes to the underlying conditions should cause the ICDR to be stopped
    - Intermediate changes (e.g. ongoing sensor degradation) need careful judgement
- Why is this useful? How does it compare to operational ground segment outputs?

What’s involved in running an ICDR?

- Timeliness constraints must be hit
  - Or explained if not!
- Automated production
  - Must be totally robust (no unnoticed errors or missing data)
  - Including human intervention, when (any)thing goes wrong
- Change control to ensure continuity / reproducibility
- Monitoring systems
  - e.g. batch processing system breaks, disk fills up
- Quality assurance and automated validation before delivery
  - Are there products when they’re expected?
  - Are the products plausible?
  - Are the products good?
  - Is there a continuity break?
- Automated reporting and “call-outs” for non-nominal situations

This is how system requirements look like

NRT (OPE) that generates SAF LEO Products and Log (Reports) while the (VAL) instance is used for validation purposes. The same SW is deployed on the RealClimate LEO Backlog and configured to generate Products and Logs in case of failures or processing interruption as defined in the JOP/OICD.

The RealClimate LEO processing functions as input data: AVHRR, TLE, Forecast Data and Ice Concentration Data to generate SAF LEO Products and Logs.

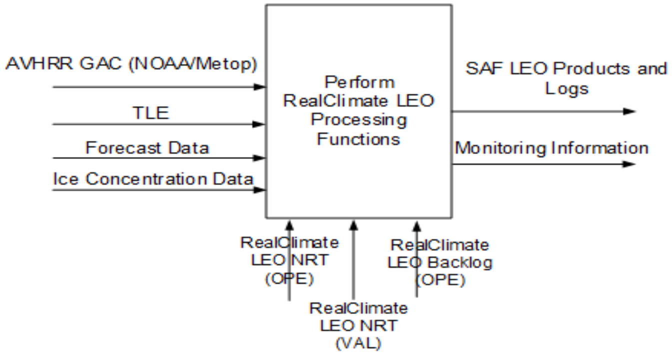


Figure 4 Perform RealClimate LEO Processing Functions

REQ ID	Title	Requirement Text	Requirement Link
RCL-FUNC-001	Product Generation	RealClimate LEO shall generate climate data records (ICDR) listed in Appendix C based on AVHRR data. These products are referred to as SAF LEO Products and Logs.	SR/TOP-0001
RCL-FUNC-002	Input Data Access	RealClimate LEO shall be able to access permanently AVHRR Data and Auxiliary Data.	SR/TOP-0010
RCL-FUNC-003	3 Month Storage on Rolling Archive	RealClimate LEO shall store all level 2 and level 3 Products on a rolling archive for three months. Note: The list of files can be found in Appendix C.3	SR/TOP-0010, SR/TOP-0095
RCL-FUNC-004	Assessment of Timeliness	MMDS shall check the timeliness of the AVHRR and Auxiliary Data for the RealClimate LEO.	SR/TOP-0015
RCL-FUNC-005	Completeness Check of Input Data	The RealClimate LEO Data Availability Manager or the Data Processing Software shall check the completeness, file format and corruption of the AVHRR and Auxiliary Data.	SR/TOP-0015
RCL-FUNC-006	Quality Control	CM SAF shall be responsible for quality control of the SAF LEO Products.	SR/TOP-0015
RCL-		RealClimate LEO shall be resilient to failure or erroneous input and	SR/TOP-

5.2 RealClimate LEO Processing Functional Requirements

The RealClimate LEO Processing Functions are realized by 3 different Sub-systems that differ only for the role they play. The primary operational service is realized by the RealClimate LEO

EUMETSAT
EUM/SEP/REQ/19/1051923
v1E, 23 November 2020
Real Climate LEO - System Requirements Specification Document

REQ ID	Title	Requirement Text	Requirement Link
		Note: Errors are handled by the system as defined in the Error Code convention.	
RCL-FUNC-009	Re-Process Data	RealClimate LEO shall be capable to re-start the generation of products with modified or reprocessed level of input data or intermediate data.	SR/TOP-0070
RCL-FUNC-010	Products Compliance	RealClimate LEO shall generate SAF LEO products in compliance to the CM SAF defined in Appendix C.	SR/TOP-0075
RCL-FUNC-011	Products Naming Convention Compliance	RealClimate LEO shall generate products in compliance with the CM SAF Naming Convention in Appendix C.	SR/TOP-0095
RCL-FUNC-012	Product Generation List	RealClimate LEO shall generate the following products: a) AVHRR GAC Fractional Cloud Cover EDR (CFC_AVHRR_Global, CM-6010) b) AVHRR GAC Cloud Top Level EDR (CTO_AVHRR_Global, CM-6030) c) AVHRR GAC Cloud Phase EDR (CPH_AVHRR_Global, CM-6040) d) AVHRR GAC Liquid Water Path EDR (LWP_AVHRR_Global, CM-6050) e) AVHRR GAC Ice Water Path EDR (IWP_AVHRR_Global, CM-6060) f) AVHRR GAC Surface incoming shortwave radiation EDR (SIS_AVHRR_Global, CM-6210) g) AVHRR GAC Surface albedo EDR (SAL_AVHRR_Global, CM-6220)	SR/ICDR_AVHRR-0010
RCL-FUNC-013	Matrix Product and SW	RealClimate LEO product generation shall be based on the following initial software packages; a) for the cloud parameters CFC, CTP, CPH,LWP and IWP -> NWC SAF PPS v2014-patch20150327; b) for SIS, SAL, L2 and L3 -> algorithms developed by CM SAF; c) for the processing of AVHRR GAC -> pyGAC version 0.1.1-clara2-ext-201810250.1.0. Note: these are the initial SW versions of the Systems and are subject to change during the life-time of the system, only after co-ordination with CM SAF.	SR/ICDR_AVHRR-0025
RCL-FUNC-014	Product Specification Validation	RealClimate LEO shall generate products that can be validated by CM SAF against the following product specifications: (see table below)	SR/ICDR_AVHRR-0020

	Spatial coverage	Spatial resolution	Temporal resolution
CFC_AVHRR_Global	a)Global b)Northern/Southern polar regions	a)$(0.25°)^2$ b)$(25km)^2$	a) + b) daily mean, monthly mean
CTO_AVHRR_Global	a)Global b)Northern/Southern polar regions	a)$(0.25°)^2$ b)$(25km)^2$	a) + b) daily mean, monthly mean
CPH_AVHRR_Global	Global	$(0.25°)^2$	daily mean, monthly mean
LWP_AVHRR_Global	Global	$(0.25°)^2$	daily mean, monthly mean
IWP_AVHRR_Global	Global	$(0.25°)^2$	daily mean, monthly mean
SIS_AVHRR_Global	Global	$(0.25°)^2$	daily mean,

Timeliness, yes there is variety

1. Near real time "NRT" (as fast as possible)
2. Soft NRT (within 1 day)
3. Short time critical "STC" (within 3-5 days)
4. Non-time critical "NTC" (within 30 days)
5. Regular update (within 3-6 months)
6. CDR Release (annual or more)

Harder

Easier

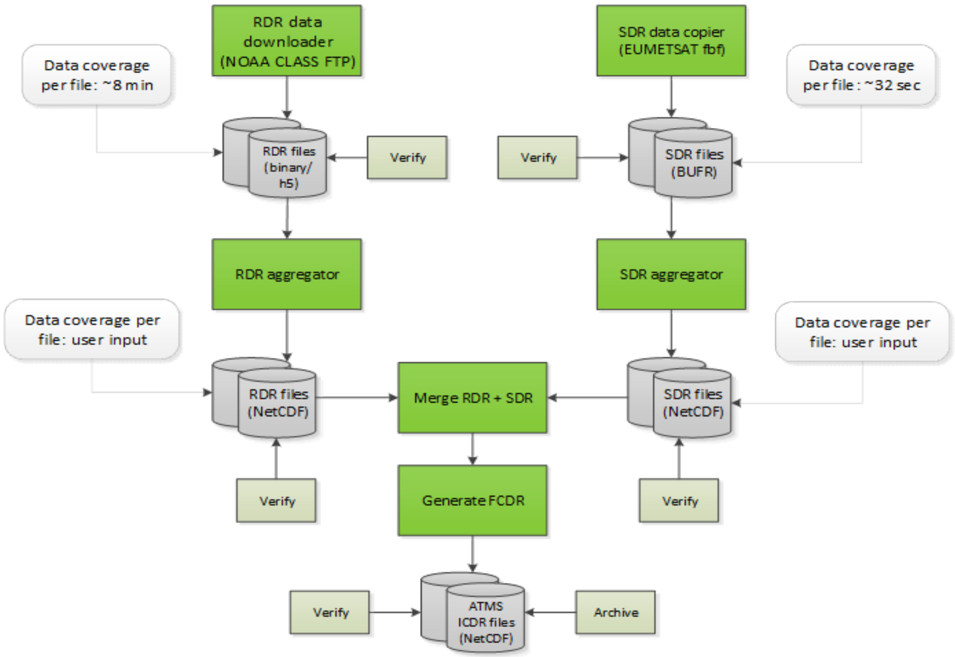
- We have two ICDR prototypes (between 2. and 3.)
  - ATMS brightness temperatures – relatively simple chain (almost data in → data out)
  - GEO AMV – high complexity, multi-stage, temporally dependent

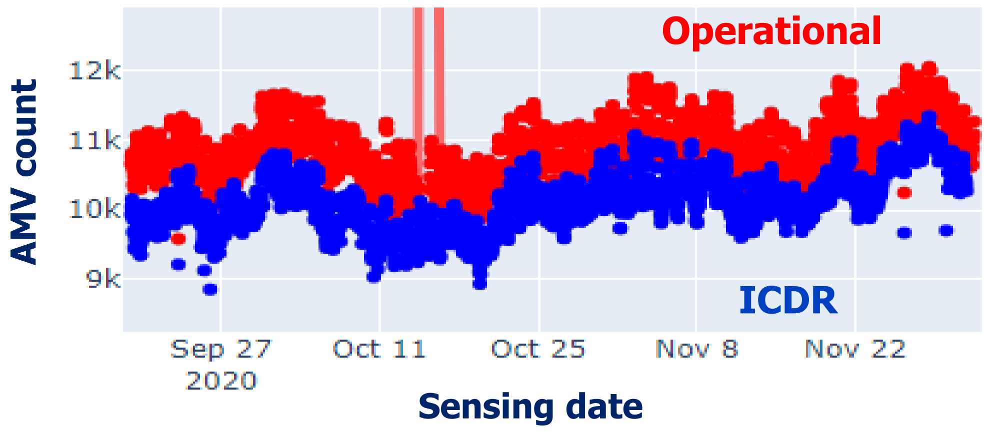
- Both ICDRs aim at a **timeliness of 18h** per target user requirement for assimilation into ERA5T[1].

[1]ERA5 continues to be extended forward in time, with daily updates being available 5 days behind real time. Initial release data, i.e. data no more than three months behind real time, is called ERA5T.

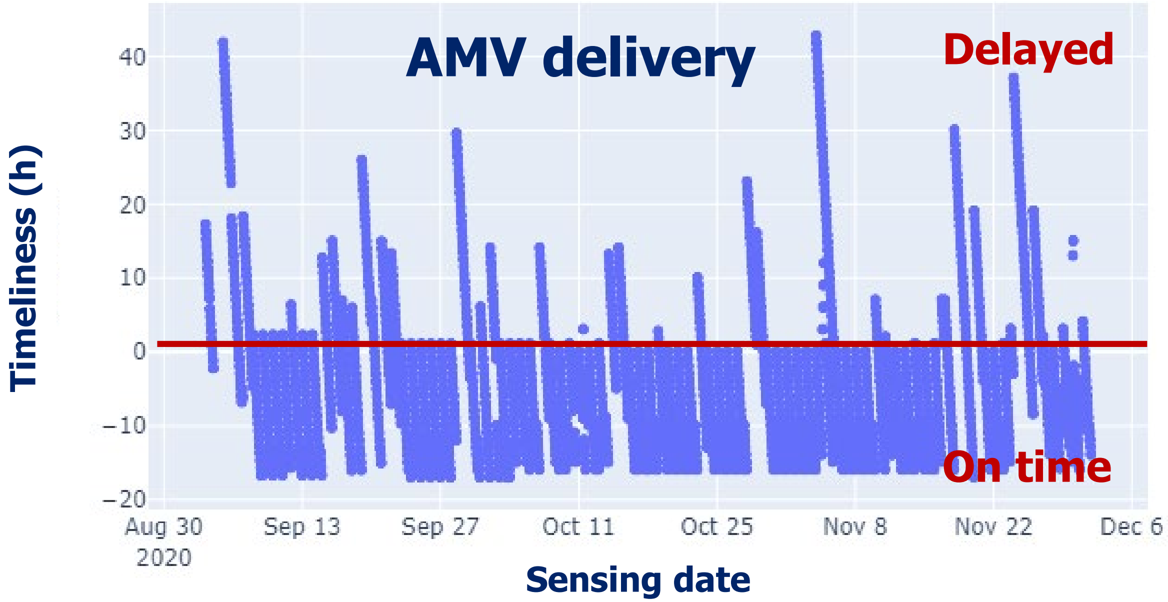
ICDR prototypes at EUMETSAT

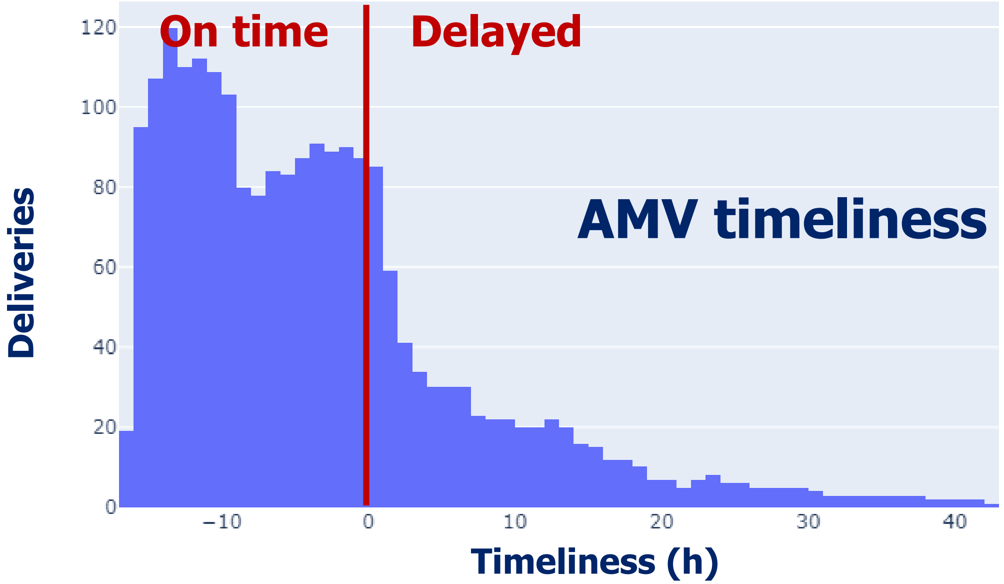
- Individual processing blocks are chained with a workflow manager (Luigi)
- Dispatch tasks (also Luigi) analyze what data should be/can be/was processed, and…
- …send processing jobs to the HTCondor batch processing system
- Sanity checks on input, intermediate, and final data raise warnings
- Errors and warnings are communicated through chat feeds and email to maintainers for follow-up
- A web dashboard is available for product inspection by experts.





AMV ICDR – Timeliness so far (initial deployment)



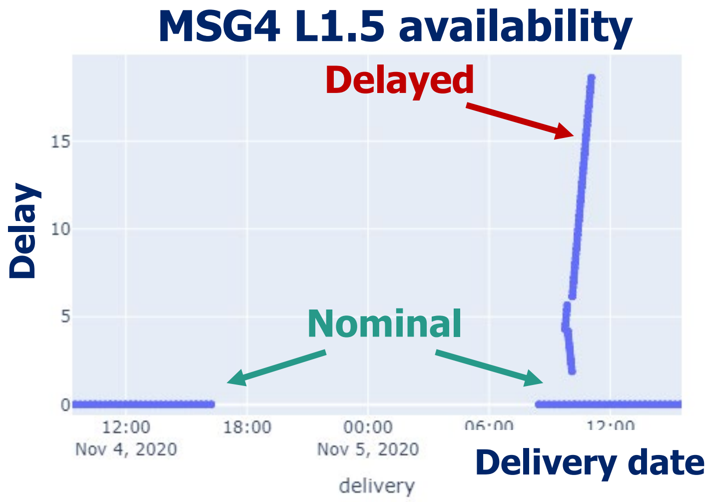


Reasons for delayed delivery:

- Short-term availability of input data (forecast) for part of the day
- Delays in the input data availability (right)
- Human intervention required outside office hours (non-operational demo)
- Resource competition (no dedicated resources)

Mitigations:

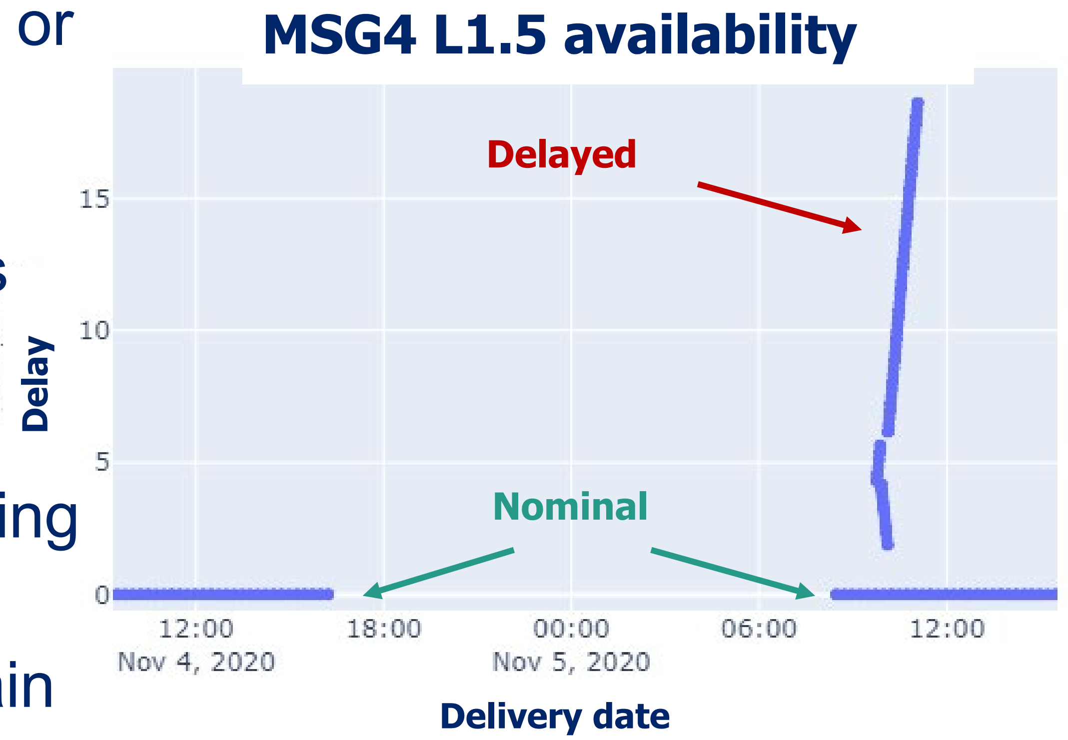
- Reduce the amount of time to run through the processing chain
- Automate recovery procedures to reduce the need for human intervention



EUMETSAT

ICDR lessons learned so far

- Critical to define a strategy that accounts for occasional delays in input data delivery (right) or re-delivery of improved input data
  - Be clear on your needs for completeness/correctness vs timeliness vs computing resources
  - Particularly complicated if processing timeslots are inter-dependent and processing is resource intensive
- Raise warnings as early in the processing chain as possible
- Competition for computing resources with other projects requires adequate scheduling management



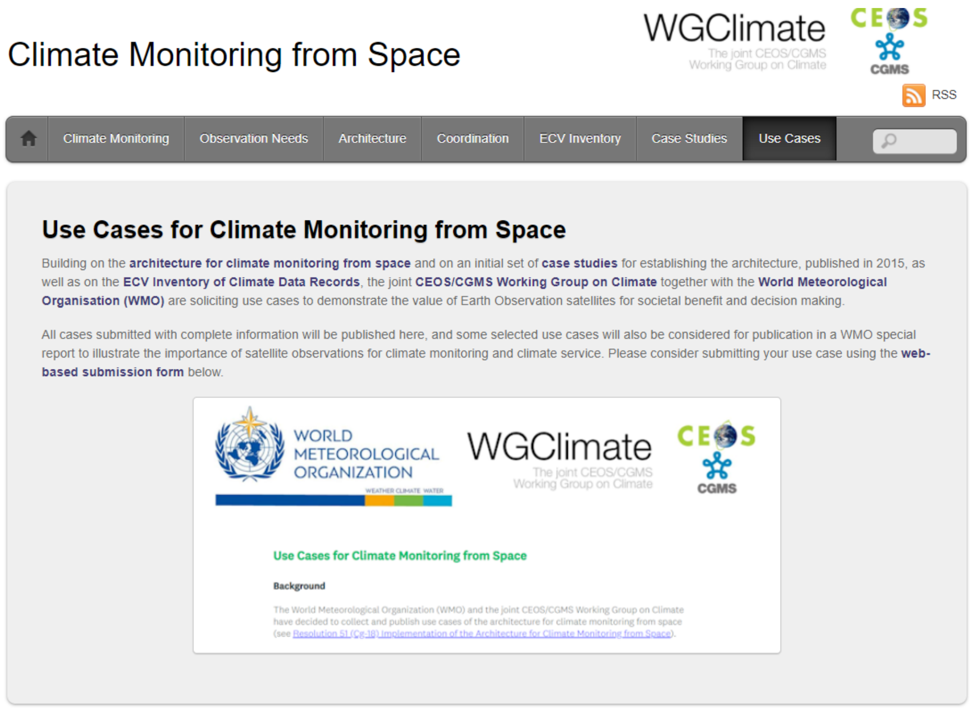
EUMETSAT

What next?

- Improve understanding of the diverse needs for ICDRs;
- Further analyse which outputs from processing systems (EDR/ICDR) satisfy needs:
  - How different is the ICDR from a well maintained EDR (homogeneity, uncertainty, etc.)?
  - How does an EDR impacts an application?
- Analyse cost that dramatically increase with tougher timeliness requirements, variety of input data needs, and complex processors
- Use results to develop solutions for ICDR processing schemes:
  - Are there more intelligent solutions than just another processing chain?
  - How can we deal with issues in input data if it impacts applications?
  - Etc.

Use Cases for Climate Data Records

- WGClimate#12 in May 2020 decided that a new routine activity on collecting use cases for climate data records will be started
- Use Case gathering tool has been integrated into climate "Use Cases" web page (https://climatemonitoring.info/use-cases) open for submission on July 27, 2020 with widespread distribution on social media, A47.14 can now be executed.



- Major Objectives:
  - Demonstrate value of climate data records for decision/policy making, e.g., usage of satellite data in Paris Agreement Global Stocktakes by demonstrating usage in a use case with UNFCCC Parties
  - Optimise the use of climate data records in applications relevant for climate services and science
  - Learn about needs of applications to foster requirement engineering by GCOS
  - Validate the top-down architecture for climate monitoring from space with a down-top approach ensuring traceability from usage to space-based observing system
  - Support capacity building by providing/receiving use cases for/from training activities, e.g., for developing countries (link to CGMS and CEOS capacity building activities)