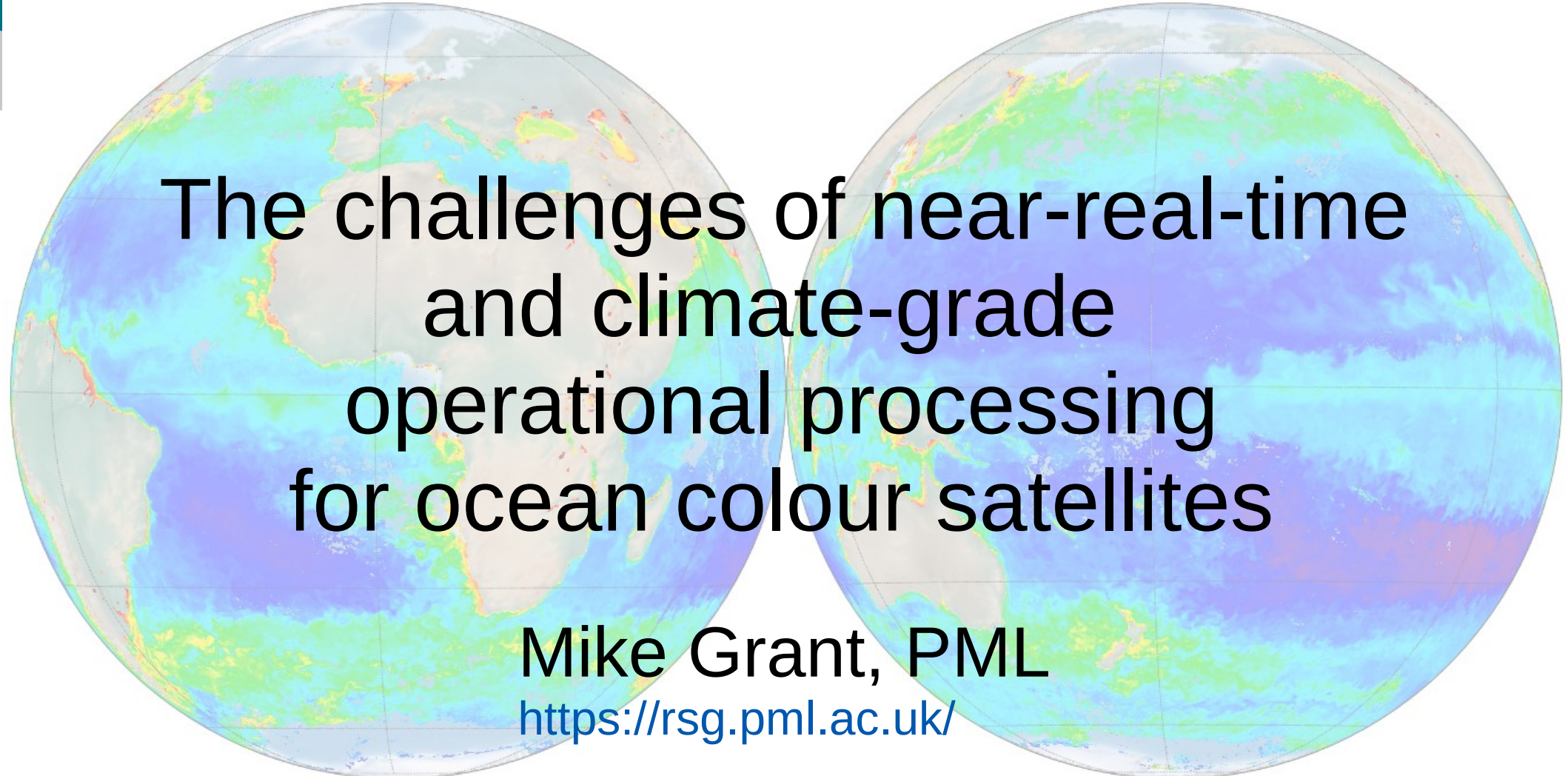


**PML**

Plymouth Marine  
Laboratory

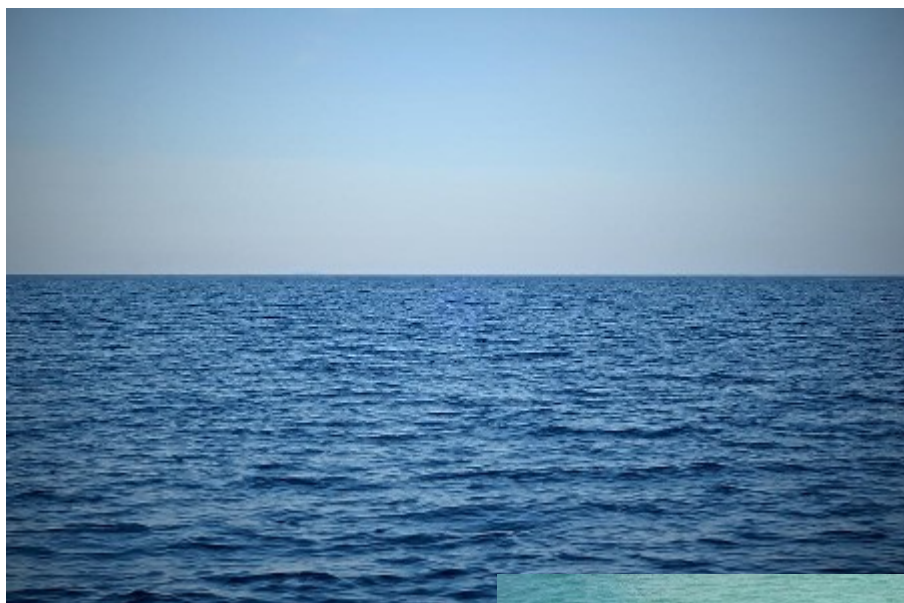


The challenges of near-real-time  
and climate-grade  
operational processing  
for ocean colour satellites

Mike Grant, PML

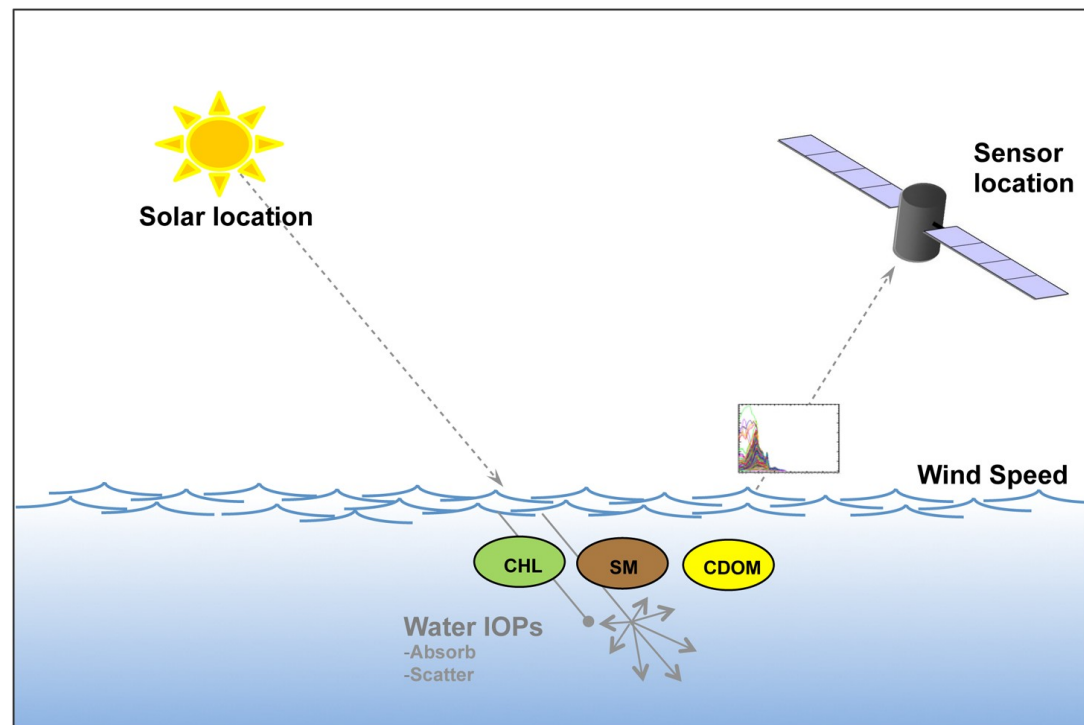
<https://rsg.pml.ac.uk/>

# What is Ocean Colour?



# What is ocean colour?

- Light emitted from the sun interacts with the seawater and its constituents before being captured by the remote sensor
  - Looking at the difference constituents make vs “pure” seawater
  - Case 1 = open ocean, CHL is the main contributor





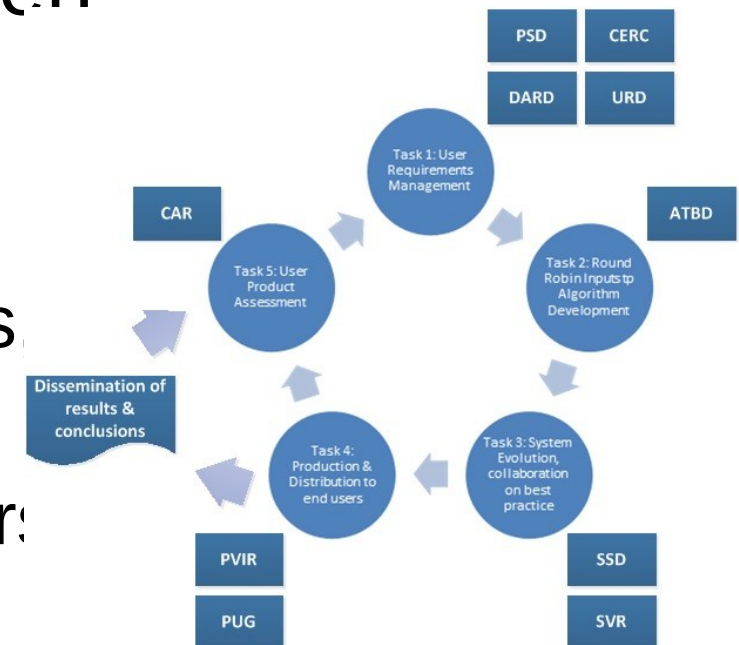
- PML hosts the largest marine EO group in the UK: ~30 staff and 12 students & visitors
  - EO research, applications, service provision & support, leading major projects (ESA OC CCI)
  - Largely CR funded (NERC, H2020, ESA), with some NC funding for EO services
- PML's modelling group is ~15 people and one of the UK's leaders, coordinating a number of H2020 and NERC projects



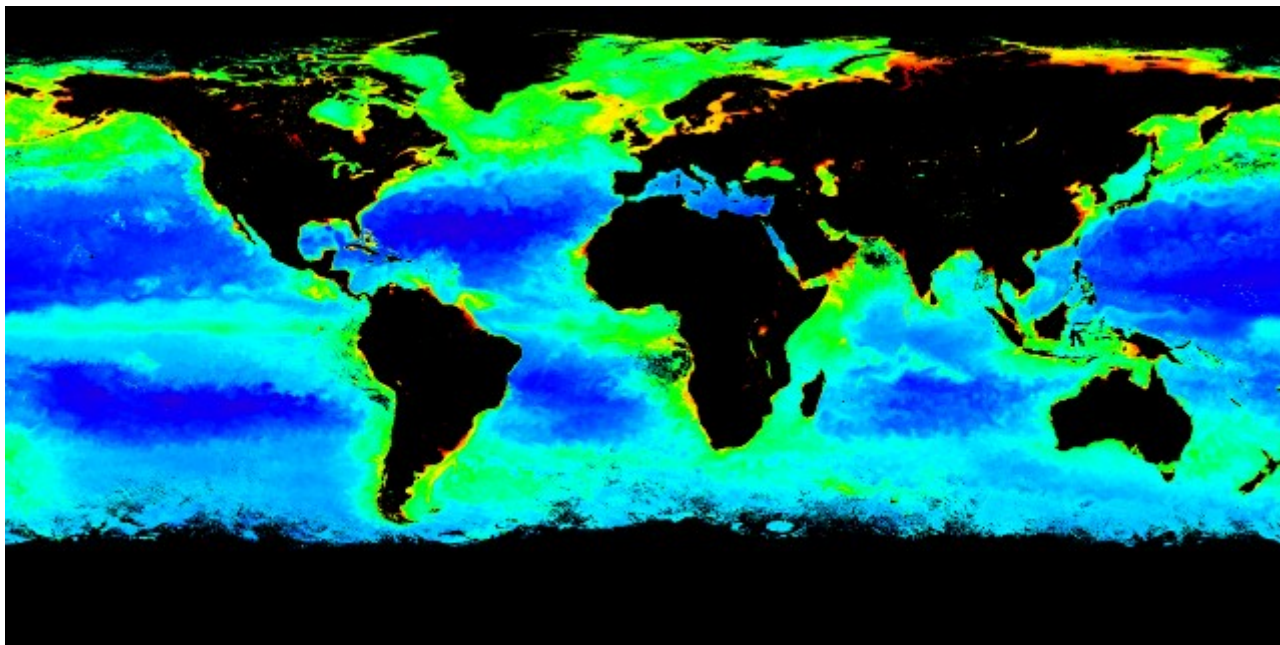
- Operational/NRT ocean colour
- Climate quality data reprocessing
- Scientific advancement / R&D
- Thermal fronts, HABs,
- Integration of in-situ (buoy, ship), airborne, satellite, modelling
  - S2 & S3 VT
- End-user data access and use
- Many other activities in PML; EO about 25%

- Operational/NRT ocean colour
  - PU in CMEMS OCTAC (ATL & ARC)
  - Partner in Copernicus land (inland waters / lakes)
  - UK NEODAAS (cruise support, academic data preparation and analysis)
  - Commercial interests
- Climate quality data reprocessing
  - ESA Climate Change Initiative programme
- *End-user data access and use*
  - *Web-based visualisation and collaboration*
  - *Big data services*

- 6 year ESA programme aiming to produce climate-grade global datasets for ~13 ECVs
  - OC nearly complete
- First 3 years to establish best approach and release v1.0, then annual updates
  - Now on v3.0
  - Updates include new algorithms, retuning, etc
  - Essentially repeating first 3 years each year
  - Aiming for operational status



- Open and robust evaluation mechanism
  - Best in class A/C and in-water product algorithms
- Multi-sensor merged product, with bias correction
- Per pixel uncertainties, based on water type and in-situ comparison





- Copernicus Marine Environment Monitoring Service; descendent of MyOcean
- Essentially level 3 products, with regional tuning, quality control, training and support
- OCTAC produces REP & DT/NRT datasets for Copernicus Marine areas
  - Focus is on NRT
  - “Bump” between REP and DT/NRT
- OC-CCI is one of the two 4km GLO REP datasets, and feeds into the REP 1km ATL, ARC, BS and MED regions

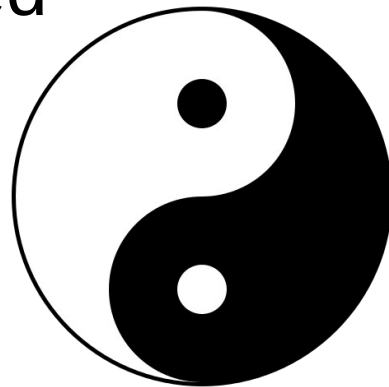


- Near-real-time = as quick as one can
  - This can vary a lot! (hours to months)
  - Quick consumption and reactive use
  - e.g. research cruises, fish farms
- Climate grade = highest quality achievable
  - Frequently lags months or years behind NRT
  - Trend analysis, data mining, model inputs
- Operational
  - Highly automated, robust, controlled processing and validation
  - Often associated with NRT, but principles apply to cyclical updates of climate data



## NRT:

- Speed / throughput
- Continuous stream, no pauses
- Well-controlled system [but must be reactive!]
- Robustness [failures should be clear]
- Highly automated



## Climate grade:

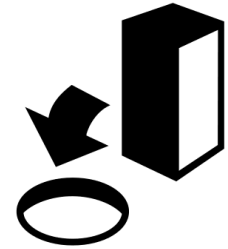
- Quality before timeliness
- Large input datasets, large outputs
- More advanced techniques, likely “science” code
- Completeness [no failures]
- Potentially large manual component


# Convergent challenges

- How to bring the computing power to bear
- Data in and out; tracking upstream
- Maintaining and updating a complex system
- Monitoring for deviations while producing data
- Quality appropriate to the needs
- Reducing human error: automation
- Organising and focussing a team
- Clear communication, internal and external

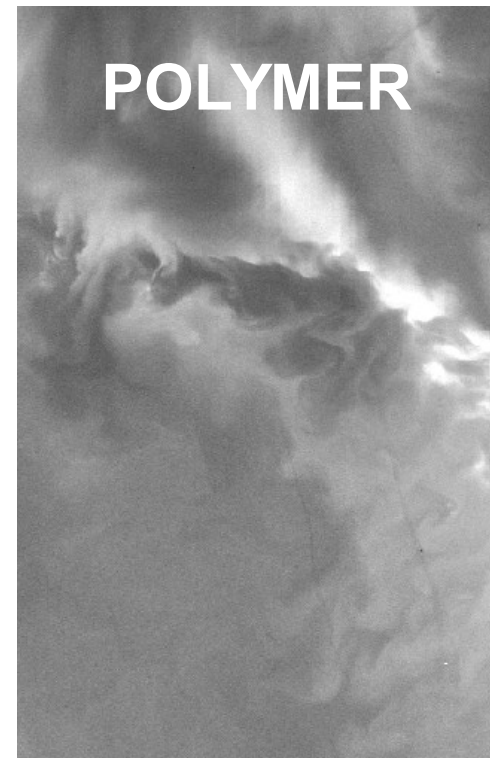
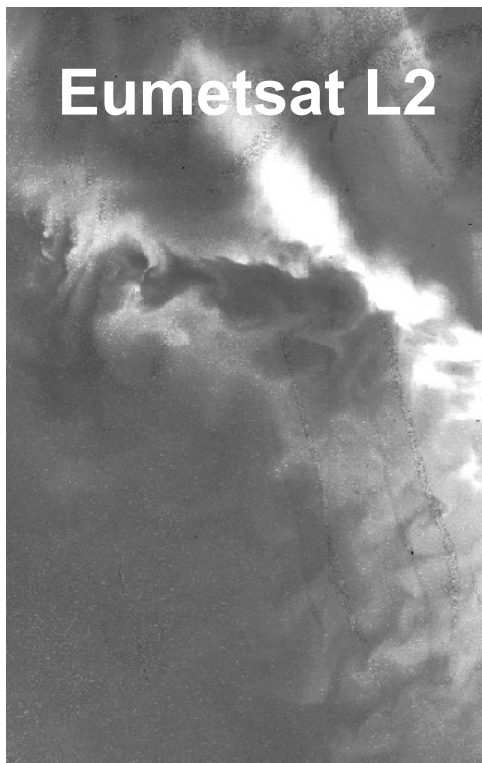


- NRT is explicitly “buyer beware”
- Climate-grade is “highest possible quality”
- Are these incompatible?
- Climate-grade datasets may have varying levels of trustworthiness
  - OC-CCI v3.0 advises caution >2012, and warns against drawing strong conclusions from data newer than 2014

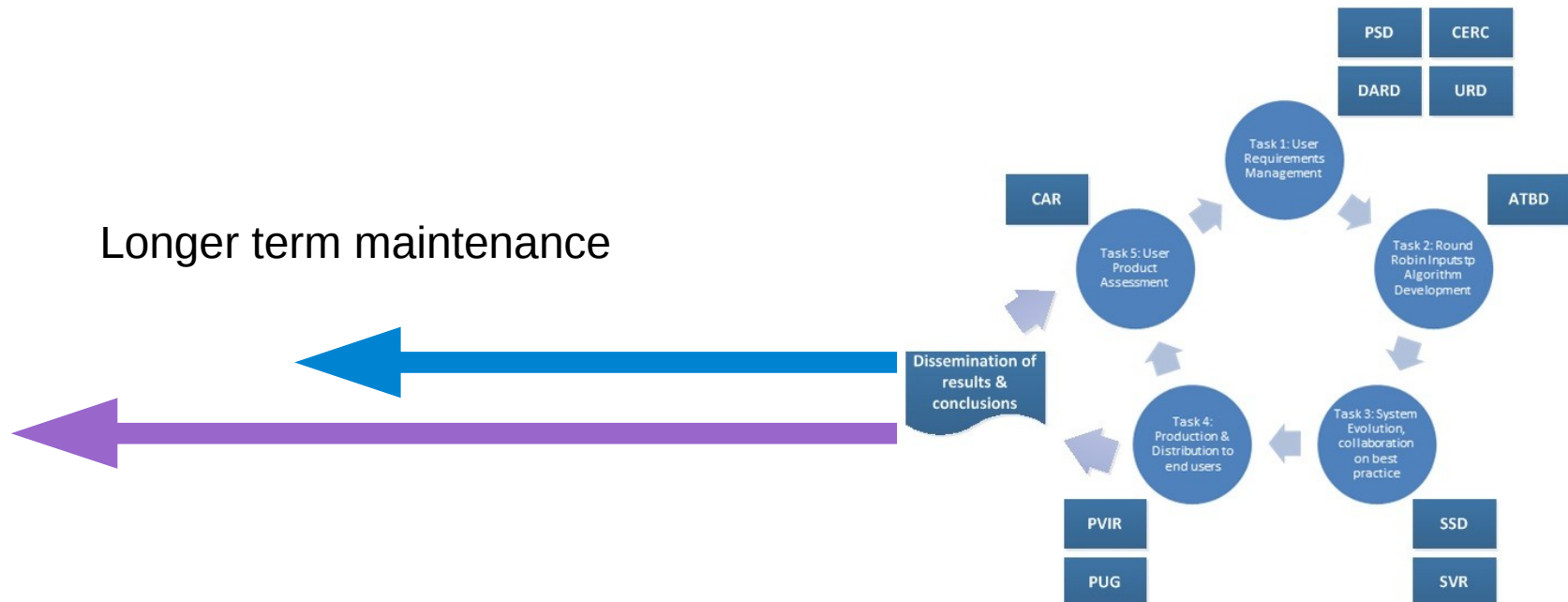


- Take a more flexible view on quality?
    - Allows a “core” climate-grade dataset with well-known characteristics
    - Extensions or “interim” records, that preserve the same approach and usage characteristics
    - Could even bring up to NRT timeliness
- 
- Cyclical reprocessing to incorporate improvements into the whole time series
    - OC-CCI has found that its current 1 year cycle is actually too fast for publications!
    - However, lagging behind more than a year is also too much

- Incorporating OLCI into v4.0 (Q2 2017) is unlikely to be achieved due to delays
- v4.0 will be “OLCI ready” but become more experimental
  - A/C, bands included, bias correction, resolution

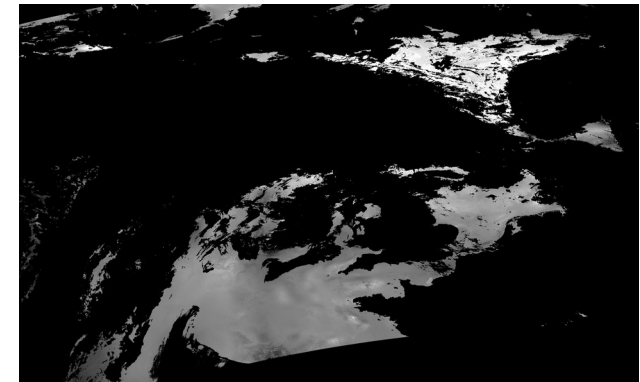


- OC-CCI will start a “v3.0-sustained” stream
  - Maintain the v3.0 approach for at least 1 year
  - Continuous extension (~2 weeks “NRT”)
  - Fully compatible climate->NRT transition
  - **Could** be used to verify continuation of a trend.. with caution!





- CCI multi-sensor approach will be in next year's ATL and ARC, running in DT (“NRT”)
- OC-CCI v3.0-sustained stream potentially available for GLO REP “continuation”
  - How to deal with upstream changes?
- Annually-produced Ocean State Report
  - Unreliable trend analysis with NRT or REP that stops too soon
  - Now can use an 'interim' dataset compatible with the climate-grade dataset

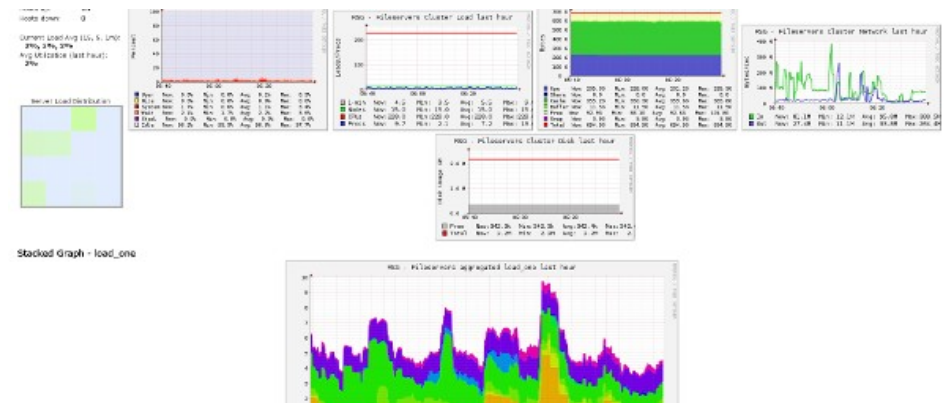
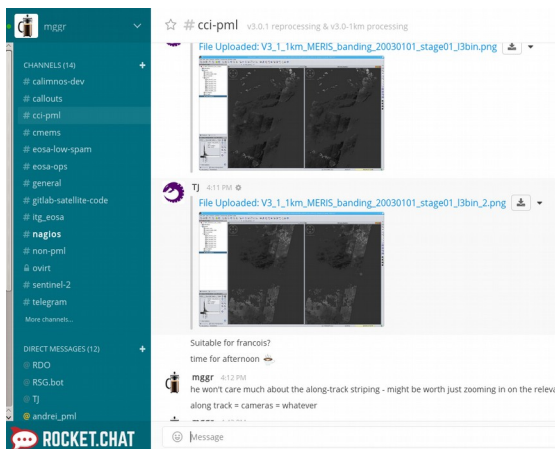


# Convergent challenges

- How to bring the computing power to bear
- Data in and out; tracking upstream
- Maintaining and updating a complex system
- Monitoring for deviations while producing data
- Quality appropriate to the needs
- Reducing human error: automation
- Organising and focussing a team
- Clear communication, internal and external

- PML's EO group is a mix of oceanographers, Earth Observation scientists and computer scientists
- Bi-directional flow from pure science, to targeted/contract R&D, to operations
- Science code 'hardened' by operational experience and computing expertise
- Large scale processing quality-controlled and refined through direct involvement of scientists

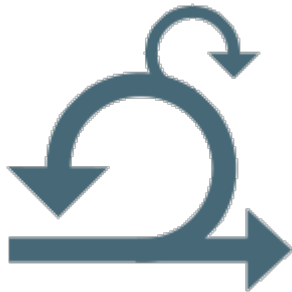
- Understanding and control of infrastructure
  - Cooperate closely with IT dept
  - Can achieve a lot with a little
- Easy to trial new technologies
- Supportive electronic communication
  - Rocket.chat [Slack clone]
  - Gitlab (source control, ticketing, code review)
  - nagios/ganglia/zabbix (monitoring)





# Integration

- Easy mixing encouraged (coffee breaks, social, etc); works well up to about 50 people
- Small subgroups, using agile approaches
- Control appropriate to the project
  - Operational work uses “harder” rules
  - Transition requires review and following guidelines
- Dedicated Satellite Services manager, to keep things coordinated

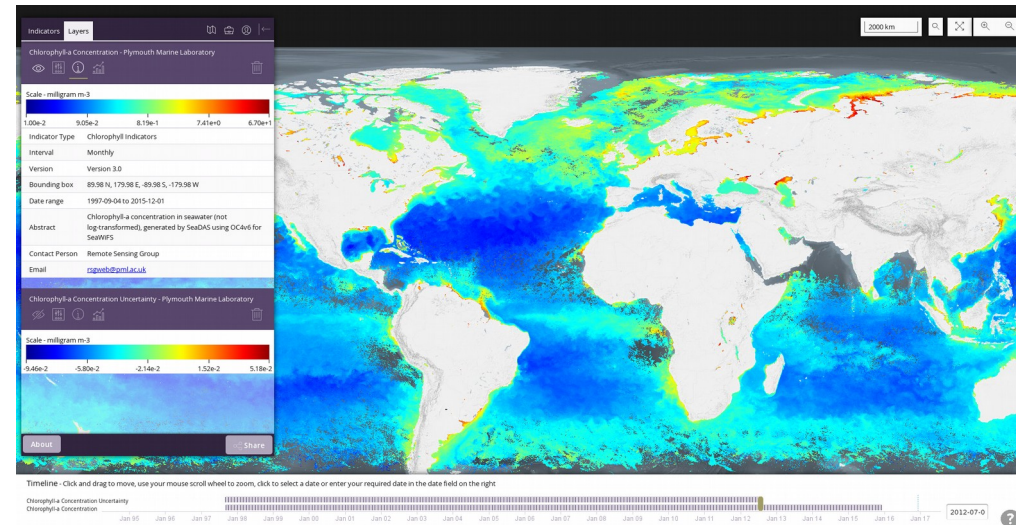


- Copernicus; global leader in climate monitoring
  - OC highly reliant on OLCI (S3a & S3b)
  - Other satellites from other countries may help, but none have the commitment of Copernicus
- OCTAC aims to have L3 products available “ASAP” (~6 weeks?) after OLCI L2 released
  - Likely to be subsequent rapid iterations for tuning
  - Derived products to follow
- CCI “OLCI ready” in 2017
  - “Climate grade” is hard to quantify for several years
  - Funding after 2017 is uncertain; CCI+? C3S?
  - PML will actively maintain CCI until at least 2019

# Next: operational challenges

- Resolution jump: CCI (4km) – CMEMS (1km) – OLCI (300m)
  - Experimental 1km-global CCI products are 20GB per day, compressed
  - Blemishes more apparent; new techniques?
- Computing challenges
  - Efficient, robust, quick processing in the face of huge volumes & multiple satellites
  - Maintaining complex systems and cultures
  - Are 2D daily maps the right model?
- Cascading quality
  - NRT with live validation → Climate grade assurance
  - How to communicate this spectrum?

- Getting the data used more
  - Fast, streamlined portals, easy access
  - Ease of use (data formats, code, documentation)
- Handling the size; as a provider and as a user
  - Web-based subsetting/processing services?
  - Full cloud processing?
- Online collaboration
  - Portals for collaboration and e-learning
  - Interactive work (e.g. Jupyter notebooks)



- Copernicus offers a step change, globally
- Computing challenges
  - Primarily volume
- Human challenges
  - Maintaining innovation while increasing consistency
- Tapered climate-grade → NRT transition is perhaps the best general approach?
- Ease of access and usage is key
  - No point solving the problems above, if the dataset isn't used