Intercomparison and Validation Task Team (IV-TT): Overview and Goals

Greg Smith, Fabrice Hernandez, Jinshan Xu, Prasanth Divakaran, Andy Ryan, Charly Regnier, Avichal Mehra, Todd Spindler, Gary Brassington, Pat Hogan, *Pam Posey*, Huier Mo, Ziqing Zu, Fraser Davidson, Katja Fennel, Hal Ritchie, & members of the IV-TT

1- Concise overview and goals of your task team
2- Summary of achievements/progress over the last 8 years
3- Short-term future and experienced/foreseen challenges in near term (next 2 years)
4- Your TT vision and strategic outlook for the next 10 years and how it contributes to overall Operational Oceanography advancement

GOV-ST-8, Bergen, Norway, November 6-10, 2017

IV-TT Overview and Goals

Intercomparison and Validation Task Team (IV-TT)

GODAE OceanView

Co
int

- rests
- MM
- e

C

Ro

IV-TT first international workshop – Montreal – ECCC – September 2016

Validation of Ocean Forecasting Systems (IV-TT)

GODAE OceanView

Scientific Challenges:

- Take full advantage of available obs. datasets and define relevant metrics for validating operational products
- Extend multi-model validation approaches to better characterize hindcast/forecast errors of individual OOFS
- Move from "classical" statistics to more adapted metrics
  - contingency tables, ocean process oriented metrics….
- Up-stream approach: from user oriented metrics, derive metrics for OOFS
  - eg Ecosystem metrics → key variables → impact metrics
- Extension to polar regions and probabilistic forecasting

Summary of achievements/progress over the last 8 years

- Main outcomes from the IV-TT:
  - Class 4 metrics, forecast skill monitored in near real time among 5+ operational centres and growing…
  - Class 1 metrics and ensemble and multi-model approach
  - Participation to the CLIVAR/GSOP Intercomparison project
- Advances in real time monitoring and validation methodology implemented in operational oceanography centres
- How operational validation is provided to users in term of Product Quality

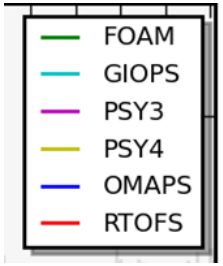
IV-TT Class4 Intercomparison

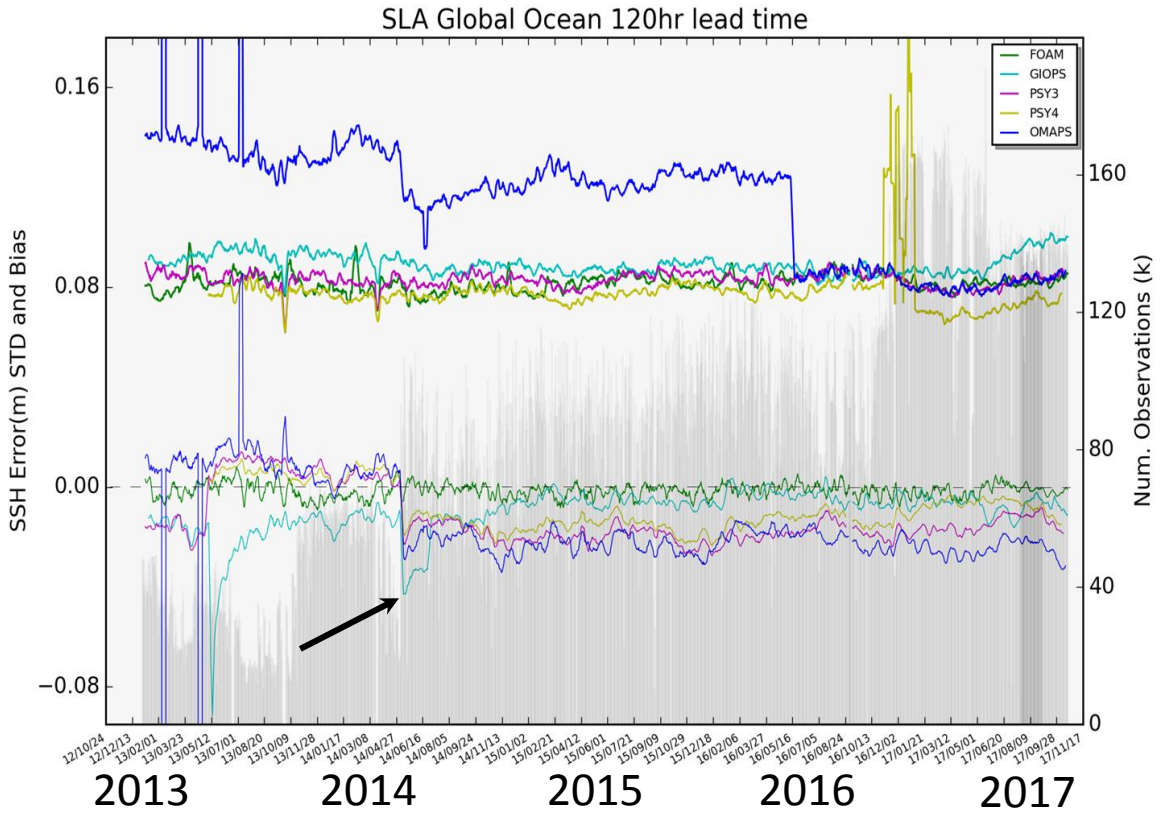
Brief History:

- UK Met began providing SLA, SST and in situ T&S profile files on US GODAE ftp server in late 2012
- January 2013 various groups started providing model equivalents
- Several papers published demonstrating skill of GOV prediction systems for 2013 and providing a basis for the intercomparison:
  - Ryan et al. (2015); Divakaran et al. (2015), Hernandez et al. (2015)
- ECCC added AMSR-2 sea ice concentration in mid-2014
  - Mercator and UK Met now contributing
- Various other centres have shown interest (e.g. NMEFC, MRI, NRL-GOFS, TOPAZ, DMI)

Global Sea Level Anomaly
5 day lead time

- Notable feature is change to reference dataset (AVISO)
- Update to AVISO included:
  - reduced filtering and decimation (i.e. increase in # obs)
  - 20yr reference MDT (instead of 7yr).
- Use of reference period in MDT and SLA obs needs to be the same. Lack of consistency is some feeds caused mean offsets.
- Overall, a convergence of GOV prediction systems from 2013 to 2016 can be seen.
- Statistics fairly consistent for 2015/16 with some divergence in 2017.

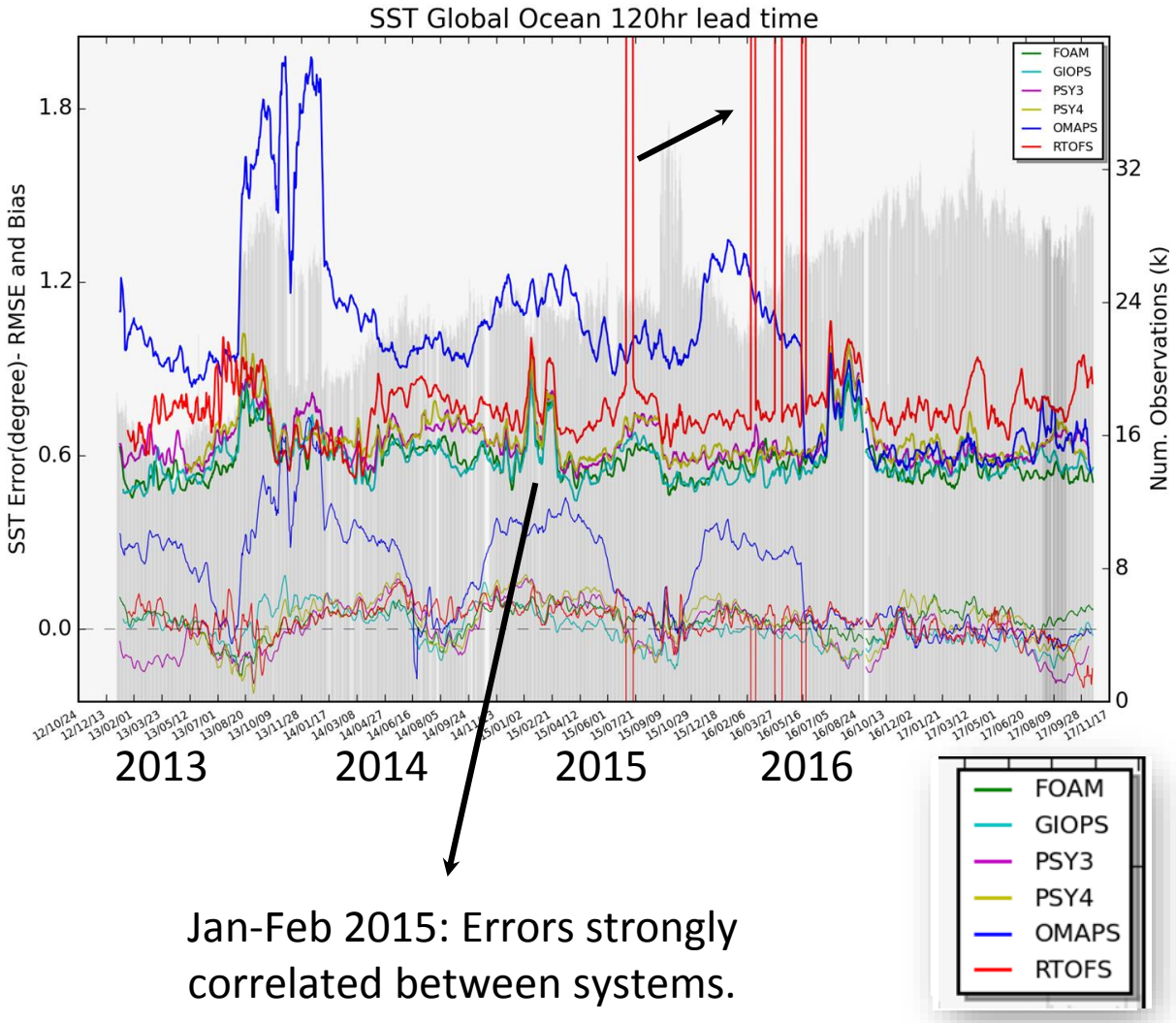




May 2014: Updated AVISO feed

Global Sea Surface Temperature
5 day lead time

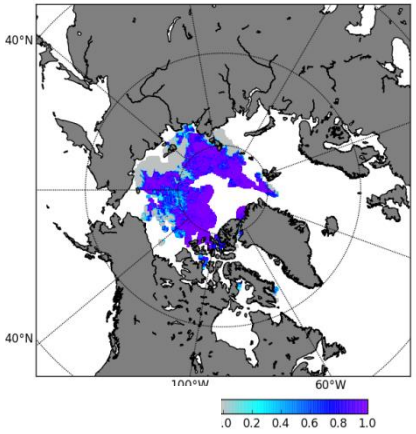
- Strong seasonality of bias and RMSE can be seen.
- Strong sensitivity to spatial variability in observation coverage
- Several interesting anomalies with strongly correlated errors
  - E.g. Jan-Feb 2015
- Large episodic errors in RTOFS since July 2015 symptomatic of processing issues experienced by various groups.
  - ➢ Improved quality assurance procedures and robustness required for operationalization.
- Scores fairly stable since 2016



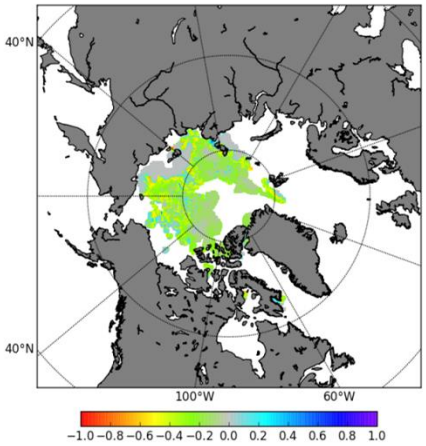
Sea Ice Concentration: Differences from AMSR-2

2016-08-08, Best Estimates (Obs-Model)

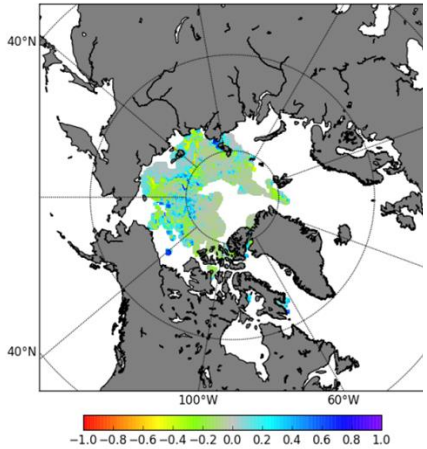
AMSR-2



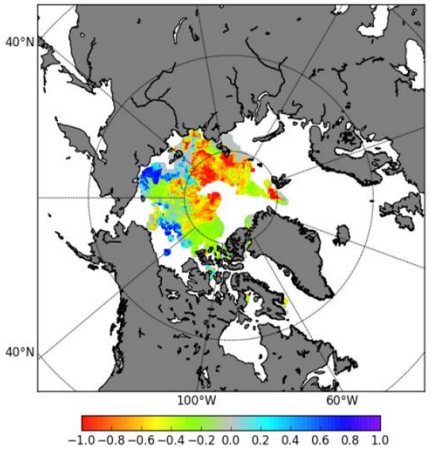
FOAM



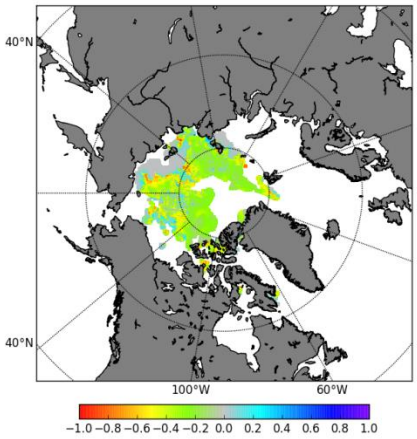
GIOPS

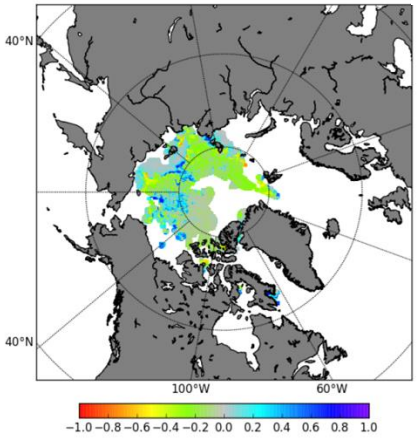


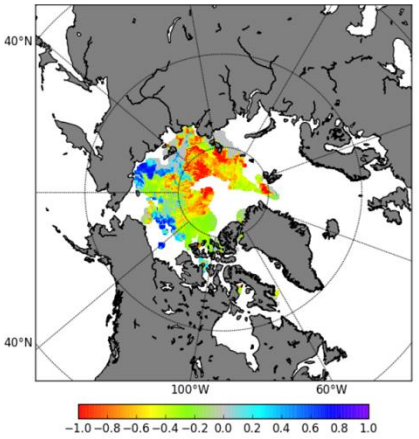
PSY4



120hr forecast







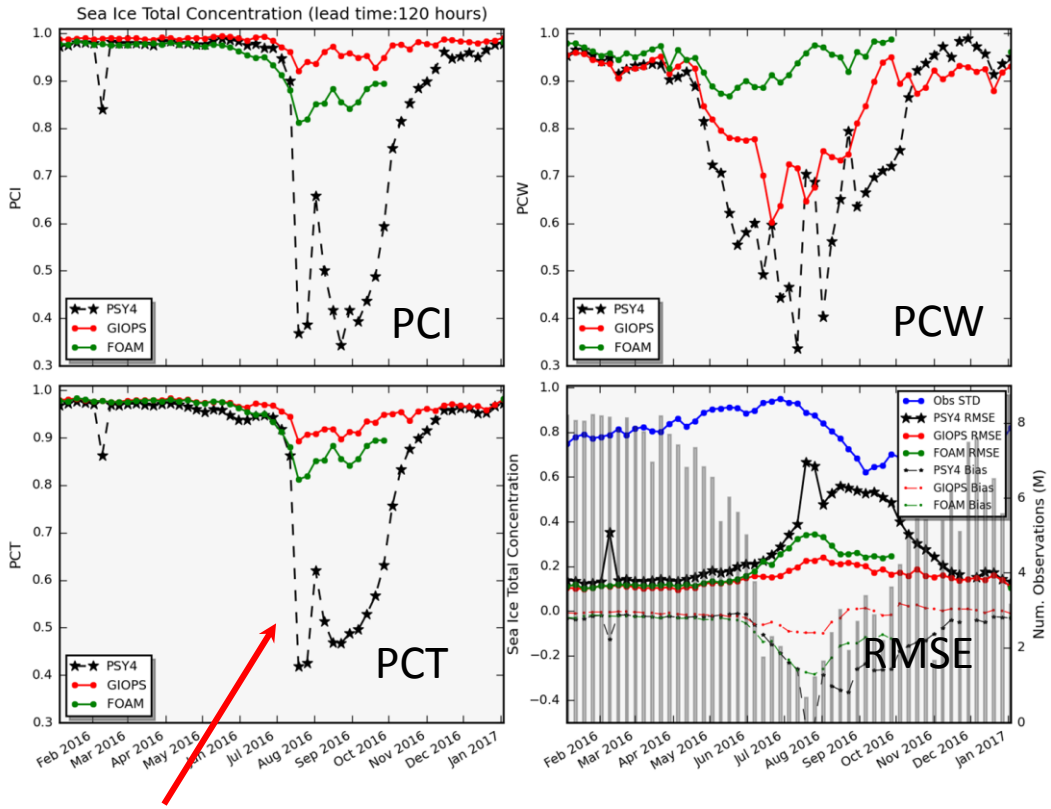
No Assimilation

Sea ice concentration (AMSR-2)
5 day lead

	AMSR2 Ice	AMSR2 Water
Forecast Ice	Hit ice (a)	False Alarm (b)
Forecast Water	Miss (c)	Hit water (d)

- Feed in place since mid-2014.
  - Initially only GIOPS and Mercator (PSY4) contributing.
  - UK Met (FOAM) joined in 2016
  - Valuable contribution for Year of Polar Prediction (2017-2019)
- Scores calculated using contingency table metrics:
  - **Proportion correct total:** PCT=(a+d)/n
  - Proportion correct Ice: PCI=a/(a+c)
  - Proportion correct Water: PCW=d/(b+d)
  - Range [0,1]; 1 is perfect score

Northern Hemisphere 2016

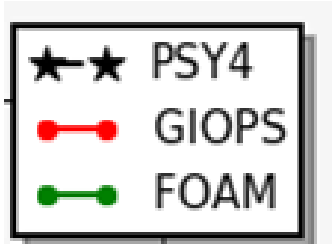


No sea ice assimilation

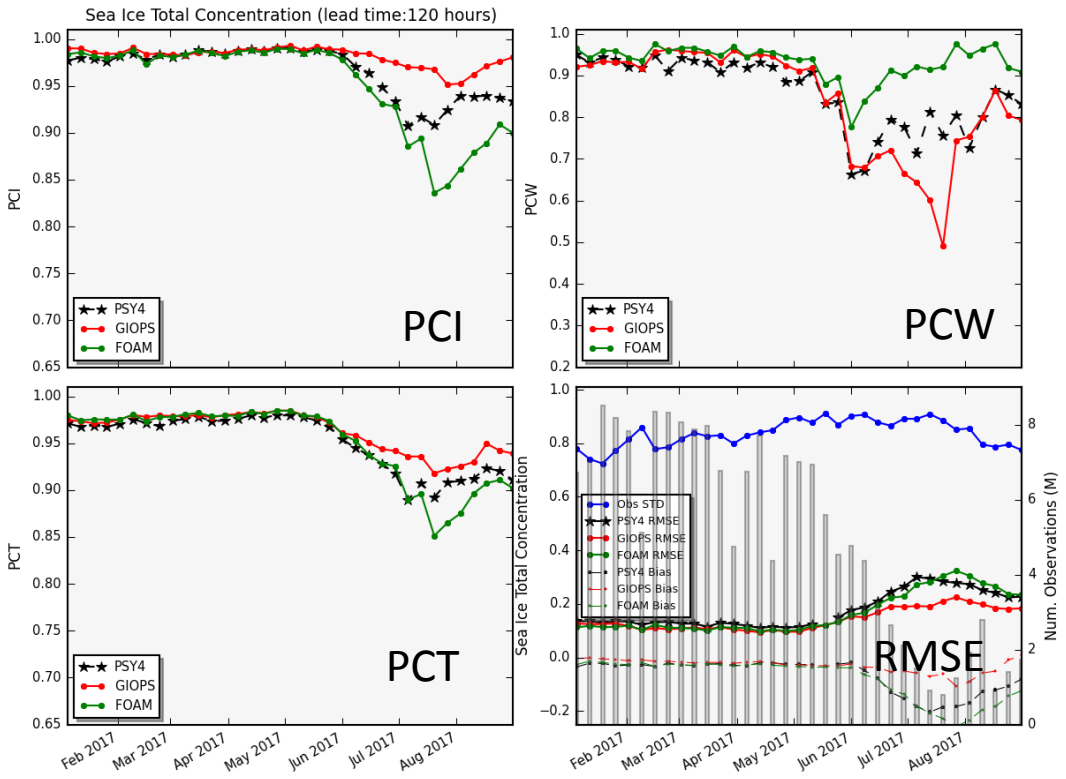
Sea ice concentration (AMSR-2)
5 day lead

	AMSR2 Ice	AMSR2 Water
Forecast Ice	Hit ice (a)	False Alarm (b)
Forecast Water	Miss (c)	Hit water (d)

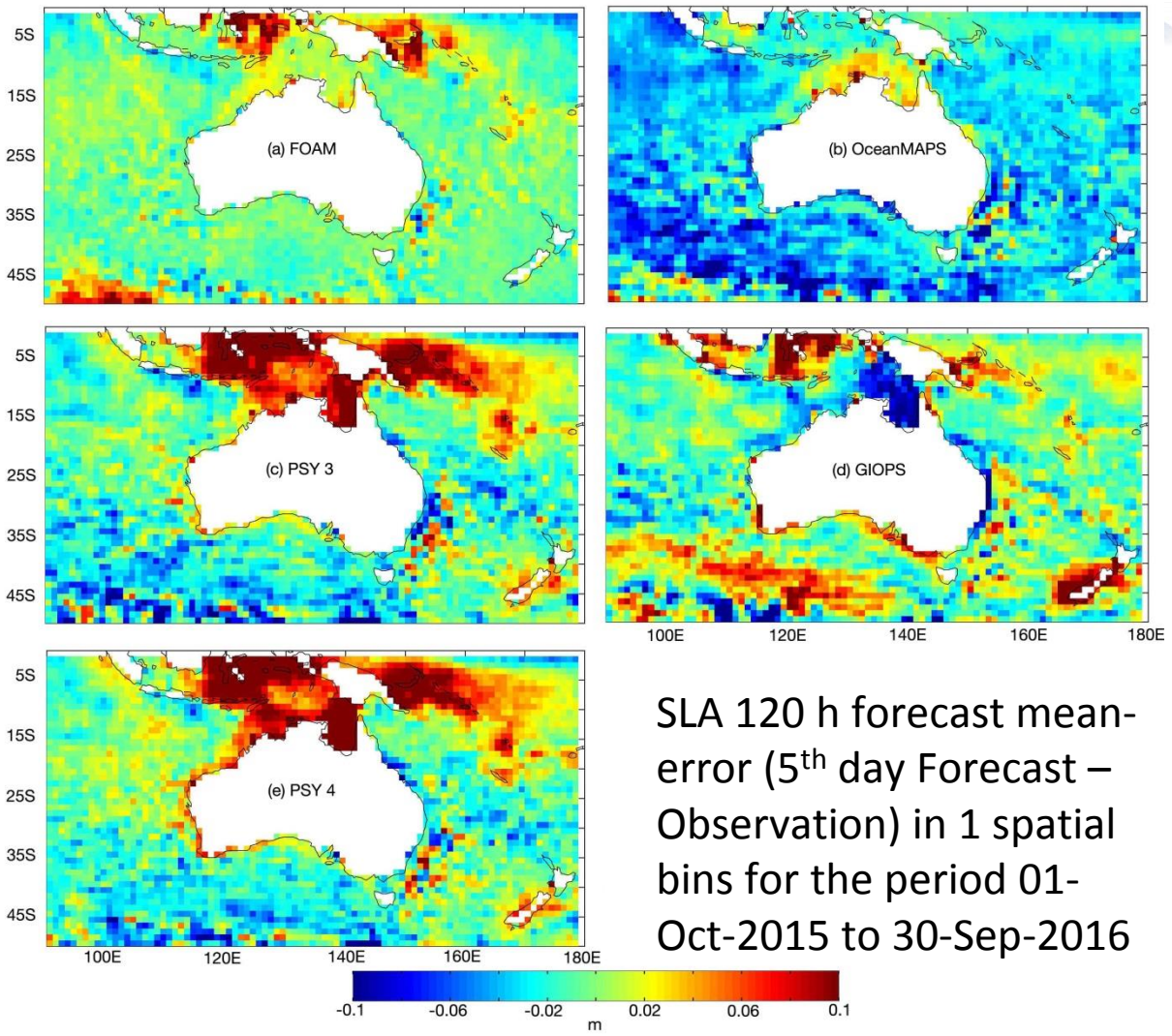
- Feed in place since mid-2014.
  - Initially only GIOPS and Mercator (PSY4) contributing.
  - UK Met (FOAM) joined in 2016
  - Valuable contribution for Year of Polar Prediction (2017-2019)
- Scores calculated using contingency table metrics:
  - **Proportion correct total**: PCT=(a+d)/n
  - Proportion correct Ice: PCI=a/(a+c)
  - Proportion correct Water: PCW=d/(b+d)
  - Range [0,1]; 1 is perfect score



Northern Hemisphere 2017

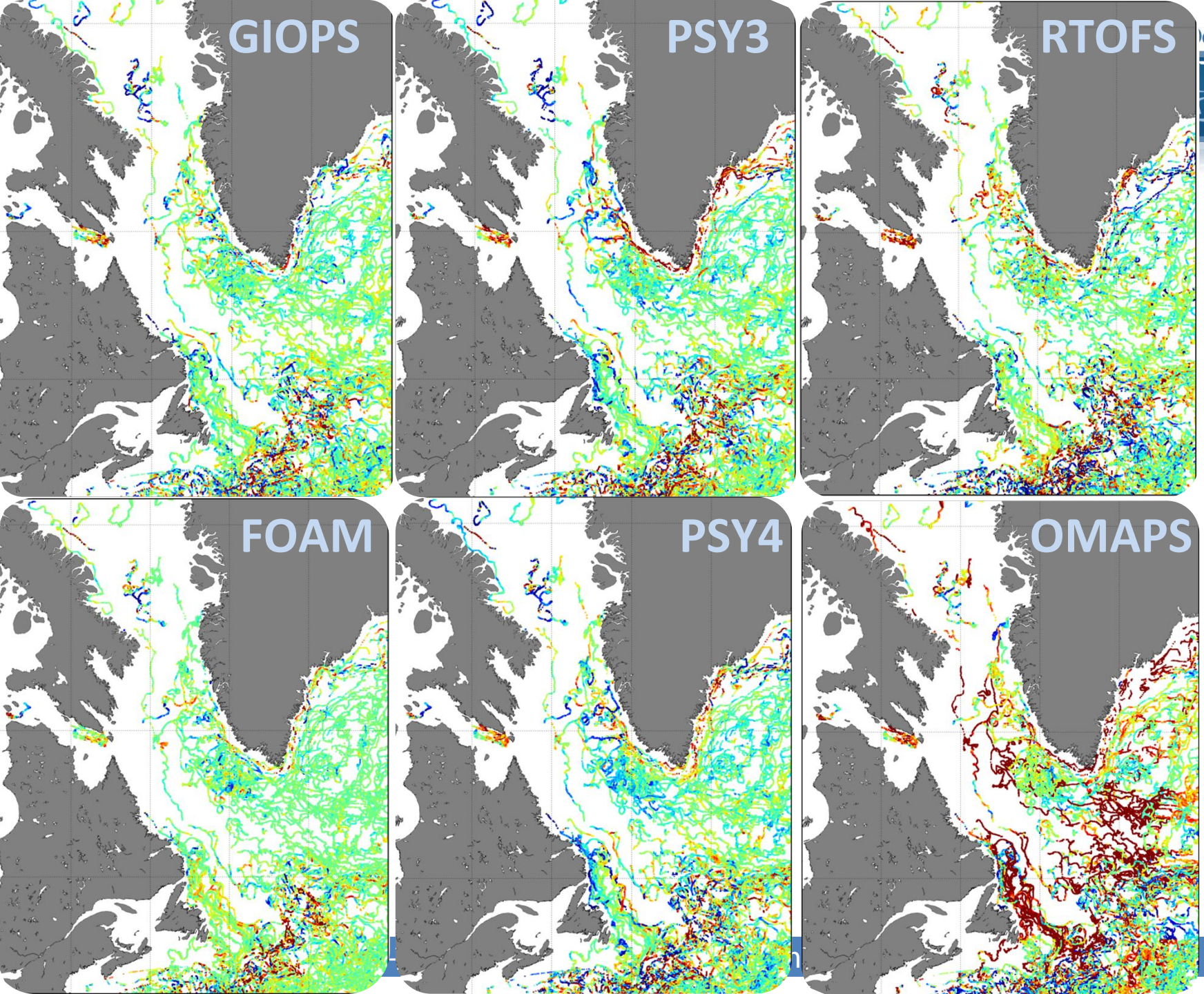


Regional Class4 evaluations

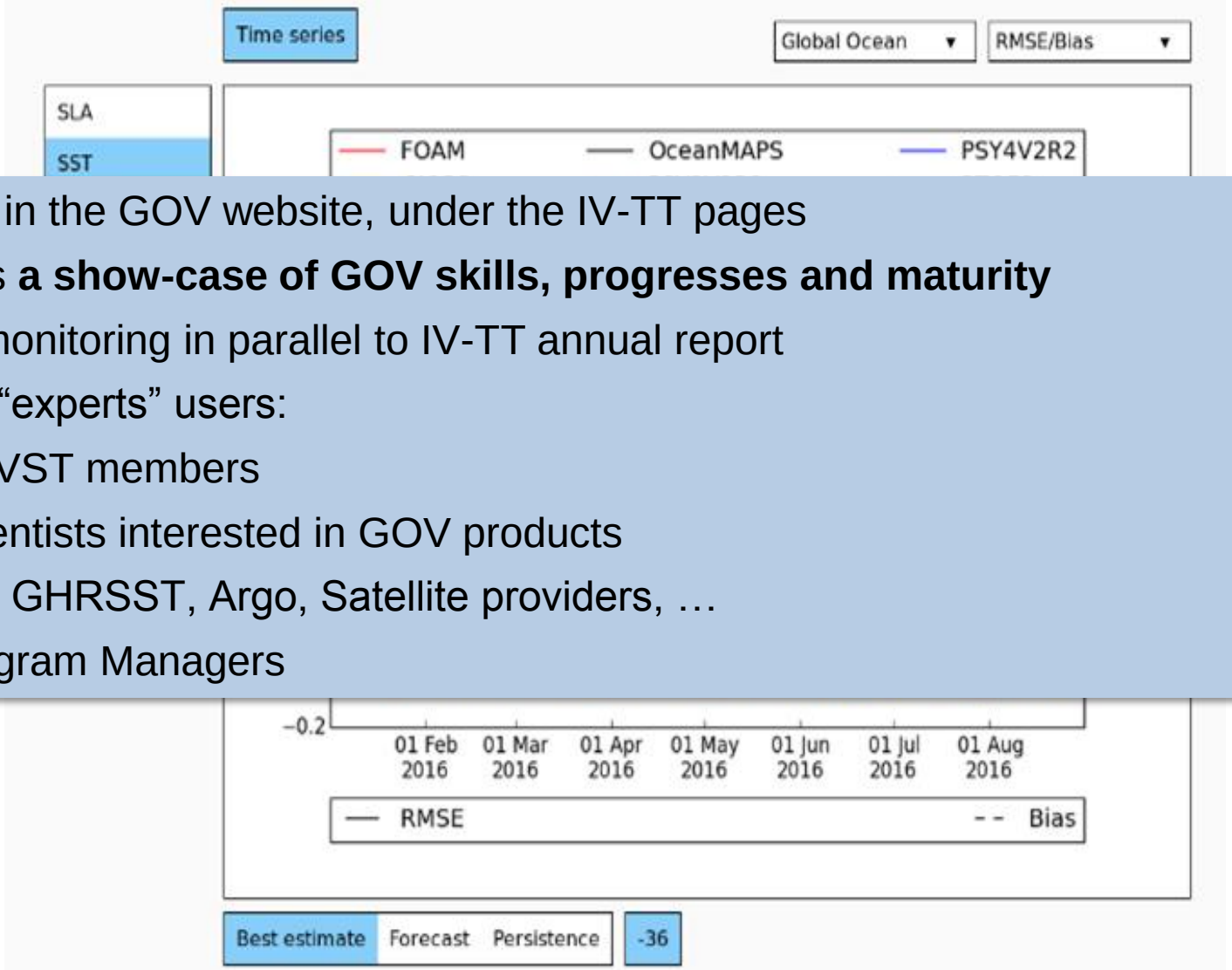


SLA 120 h forecast mean-error (5th day Forecast – Observation) in 1 spatial bins for the period 01-Oct-2015 to 30-Sep-2016

Contribution from Prasanth Divakaran


GIOPS
PSY3
RTOFS
FOAM
PSY4
OMAPS
2.0
1.6
1.2
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.6
−2.0
SST diff.

Web site demonstration

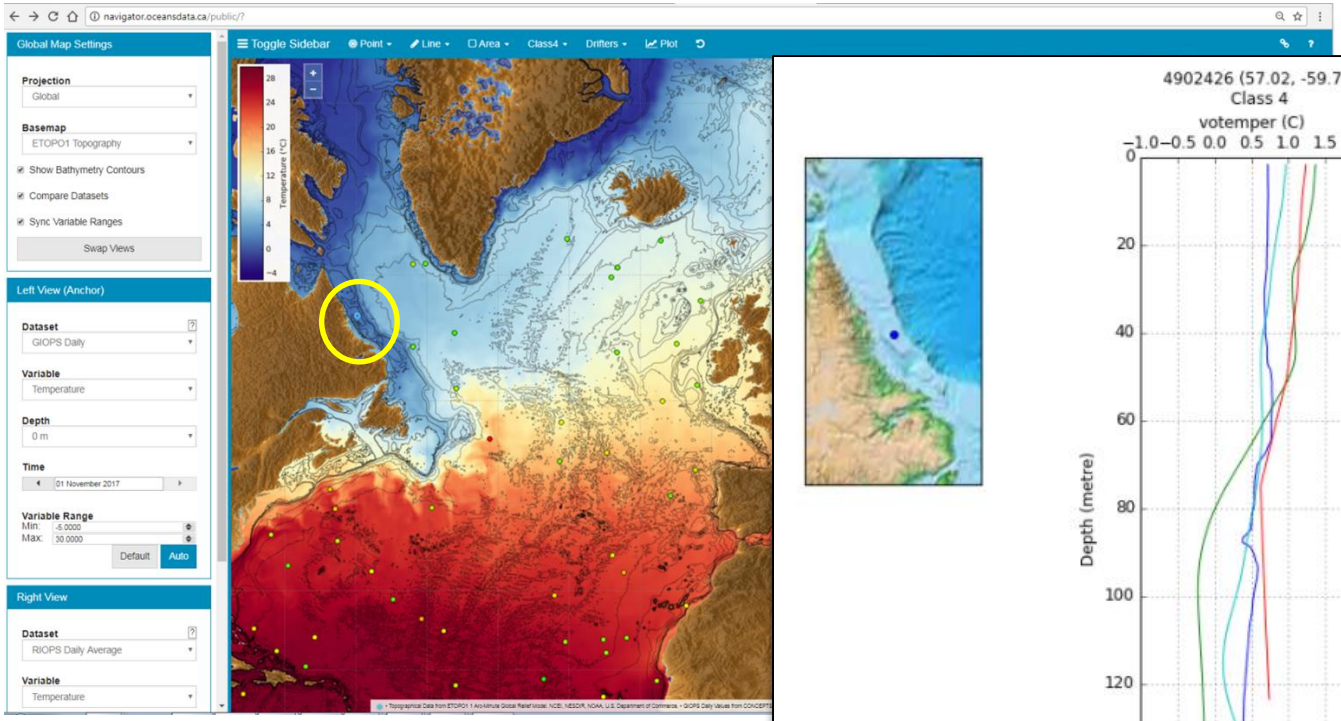


- Located in the GOV website, under the IV-TT pages
- Provides **a show-case of GOV skills, progresses and maturity**
- Allows monitoring in parallel to IV-TT annual report
- Targets “experts” users:
  - GOVST members
  - Scientists interested in GOV products
    - GHRSST, Argo, Satellite providers, …
  - Program Managers

Andy Ryan’s contribution

Dynamic web viewers

CONCEPTS OceanNavigator: http://navigator.oceansdata.ca



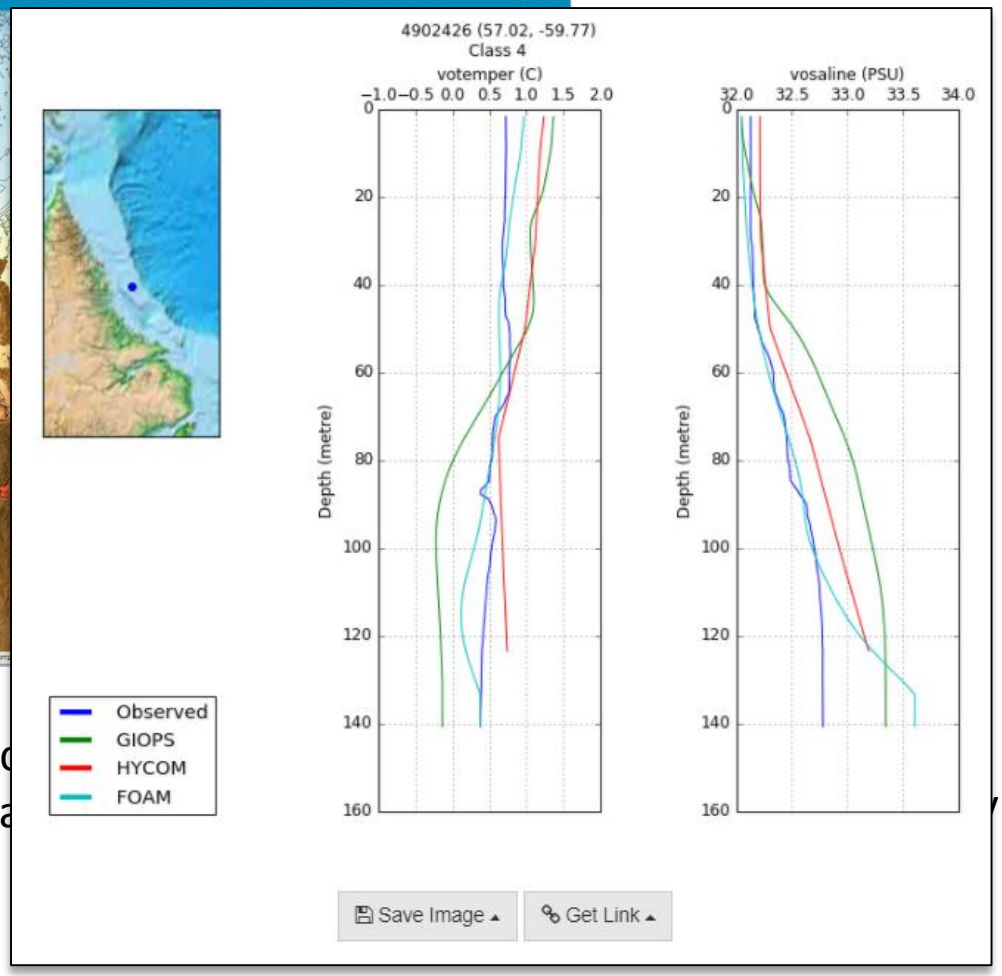
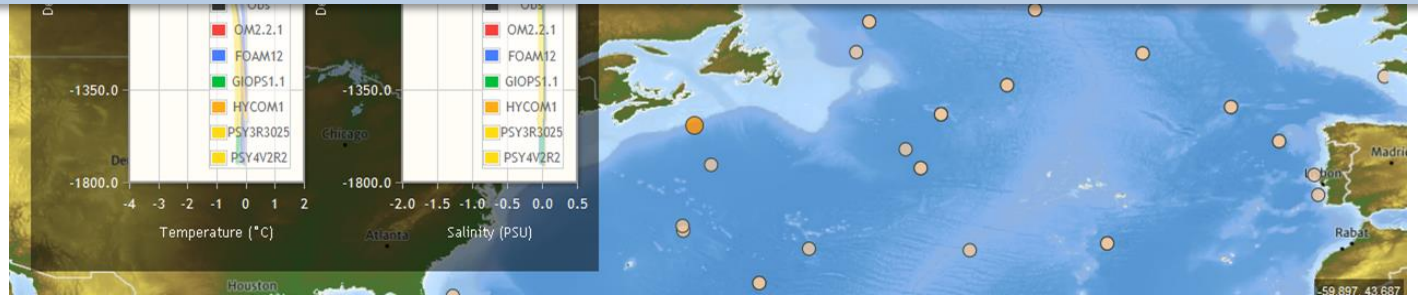


Figure: The above figure shows surfac
(left) and RIOPS 1/12°. Dots on the ma
class 4 IV-TT files.

BoM Website

- http://130.56.244.252/monitoring/index.php?pg=class4

- Allows **specific monitoring** in parallel to IV-TT annual report or GOV monitoring website
  - *Do not replace operational postprocessing of each OOFS*
- Foster coordination among GOV task teams
- Targets experts audience: GOVST members
  - *Broader audience subject to policy decisions from contributing OOFS*



Prasanth Divakaran's contribution

Short-term future and experienced/foreseen challenges in near term (next 2 years)

- Evolution of class4 system
  - Consolidation (QC, monitoring, …)
  - Extension to new observation types and metrics
  - New contributors (NMEFC, TOPAZ, DMI, …)
  - Endorsement by ET-OOFS?
  - Broader distribution for YOPP
- New metrics
  - Engagement with JWGFVR, S2S, CLIVAR, GOV-TTs, …
  - Workshop 2018
- Re-invigorating Class1 effort
  - Linkages to S2S and new PRCC effort

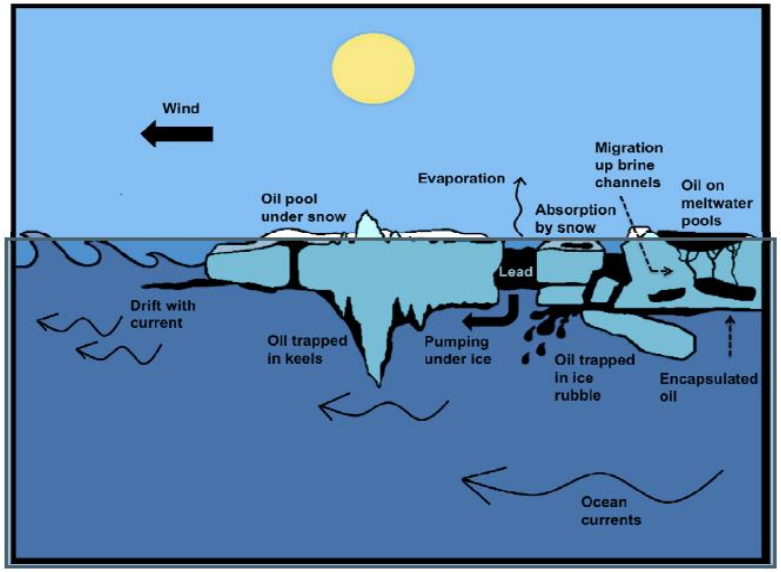
GODAE OceanView

Year of Polar Prediction (YOPP)

Year of Polar Prediction (YOPP)

- Period of intense observation and modelling in polar regions
- Opportunity/challenge for GOV
- Demonstrate maturity (Class4)
- Strong user-driven needs
  - E.g. shipping, northern communities
- Need for probabilistic forecasts
- Fundamental issues:
  - Sea ice predictability
  - Lack of Argo
  - Importance of fine-scale features

Oil spill in ice: Is GOV ready?



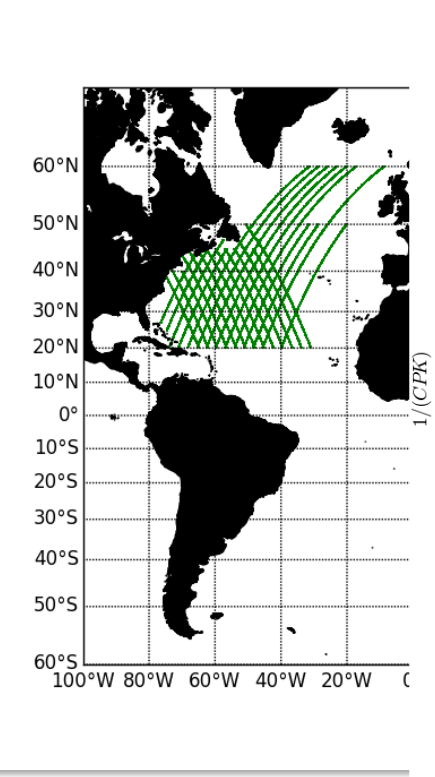
GODAE OceanView

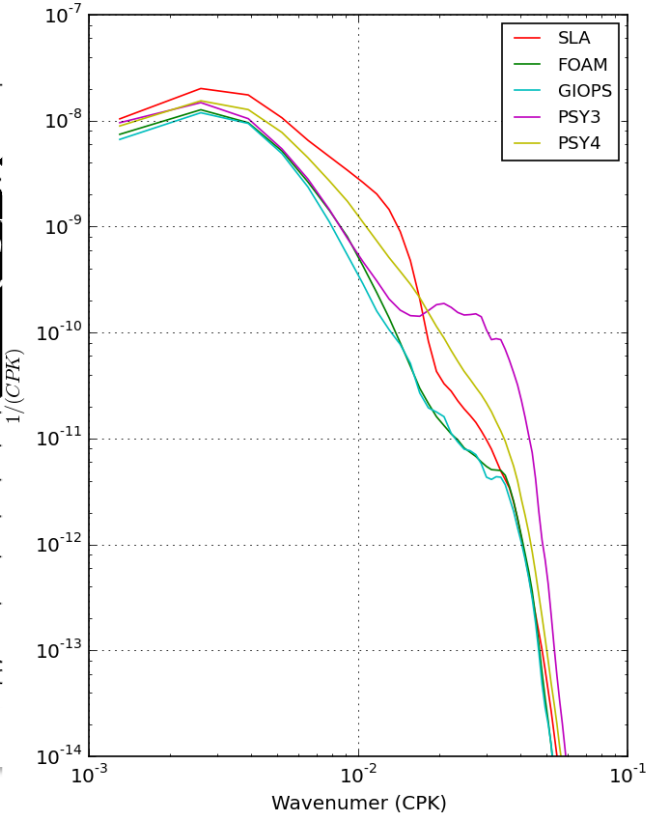
CLASS 4: ENHANCED METRICS

Class 4 metrics task

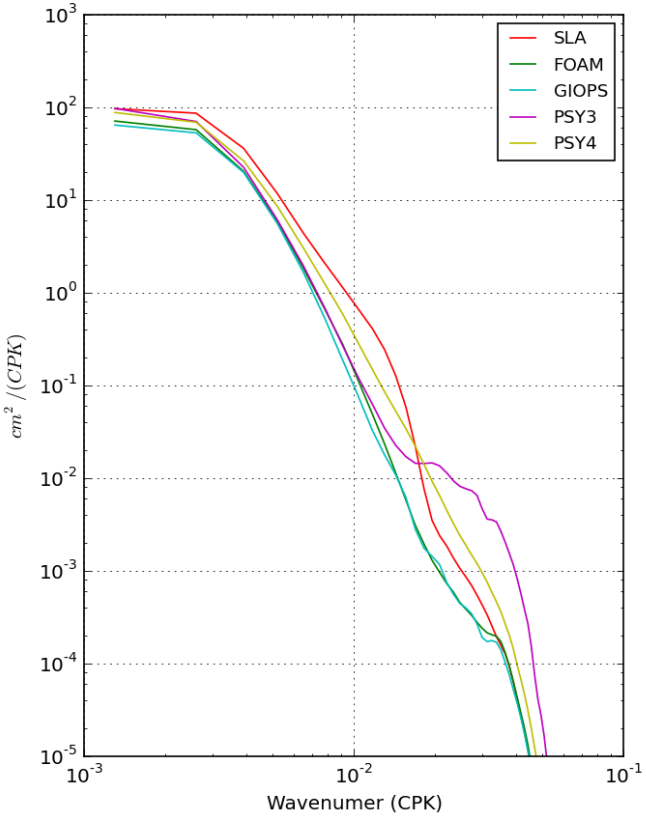
- **Provided by CLS in the EU CMEMS framework**
- **Global, all altimeters**
- **Filtered along track, tidal corr., Dyn. Atm. Corr**
- ***Allows global vs regional OOFS assessment***

Regional Class4 Evaluations: SLA power spectra over NW Atlantic





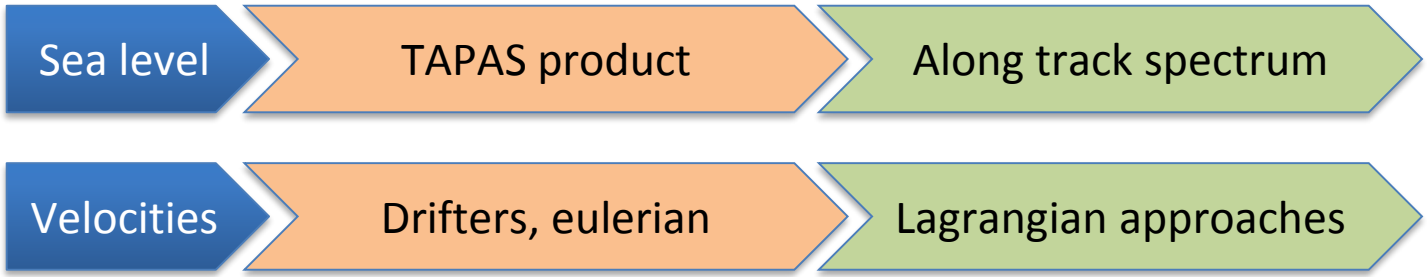
Spectrum_sla



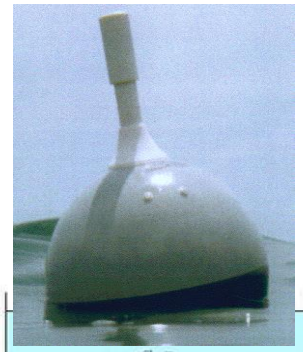
Spectrum_slope

Jinshan Xu's contribution

Class 4 metrics task



- **Prepared by Mercator using the EU CMEMS framework**
- **SVP 15-m drogued real time dataset**
- **Dedicated corrections to cope with undrogued drifters**
- ***Mostly for global OOFS velocity assessment***



Global Drifter Program (GDP)

GODAE OceanView

STATUS OF GLOBAL DRIFTER ARRAY

September 12, 2016

All drifters have SST

SST only: 558
SLP: 819
WIND: 0
SAL: 21
ARGOS: 893
IRIDIUM: 505

buoys= 1398

(M. Pazos and R. Lumpkin)

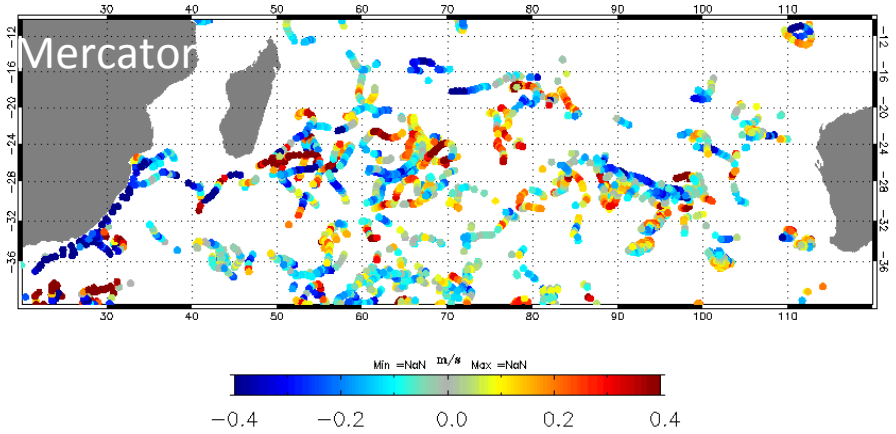
80°N 60°N 40°N 20°N 0° 20°S 40°S 60°S 80°S

30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W 0° 30°E

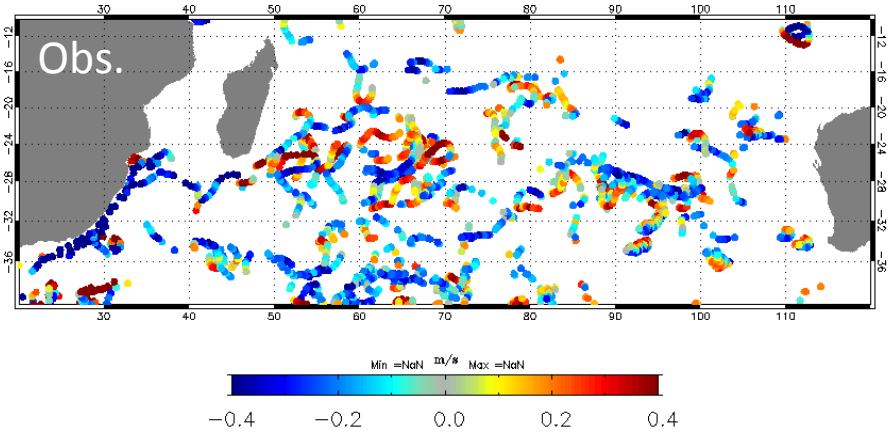
Eulerian approach

Derive daily mean velocity from filtered drifter track

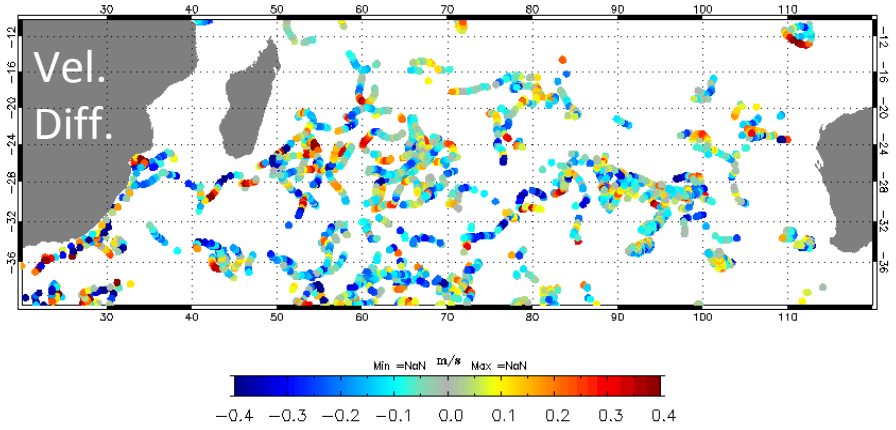
Zonal velocity of model PSY4V2R2 U drifts in 2016_JFM



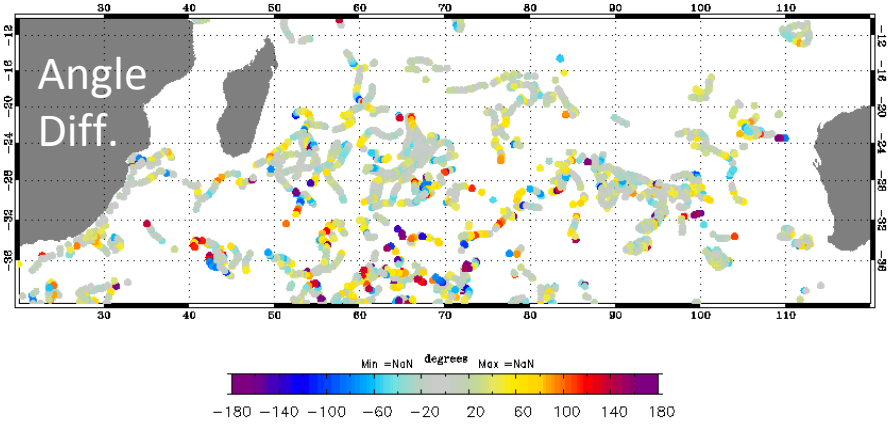
Zonal velocity of obs U drifts in 2016_JFM



Zonal velocity : U drifts − model PSY4V2R2 in 2016_JFM

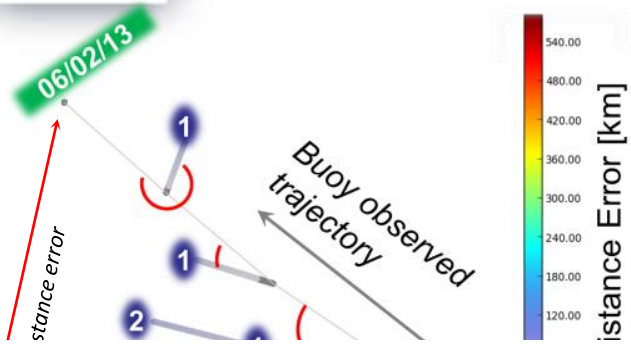


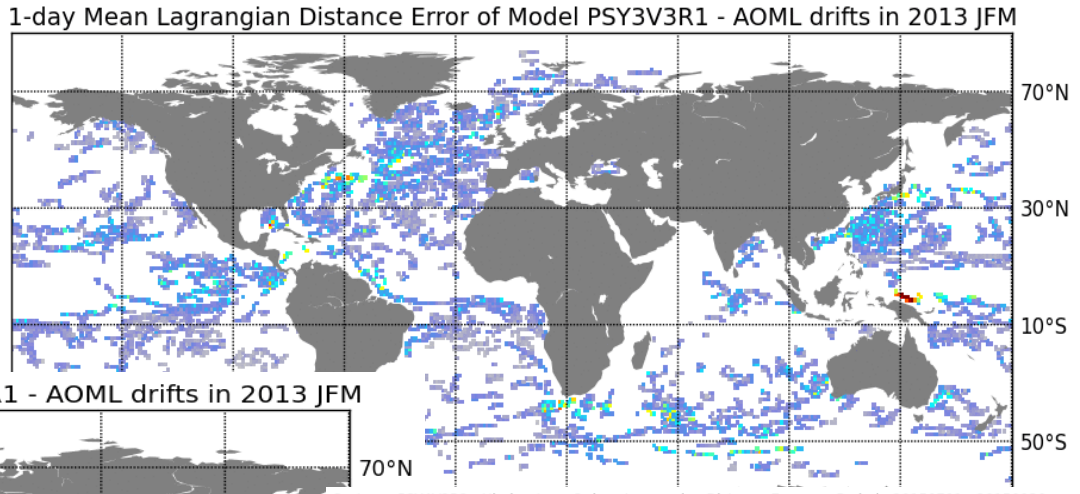
Angle of observation−model drift in 2016_JFM

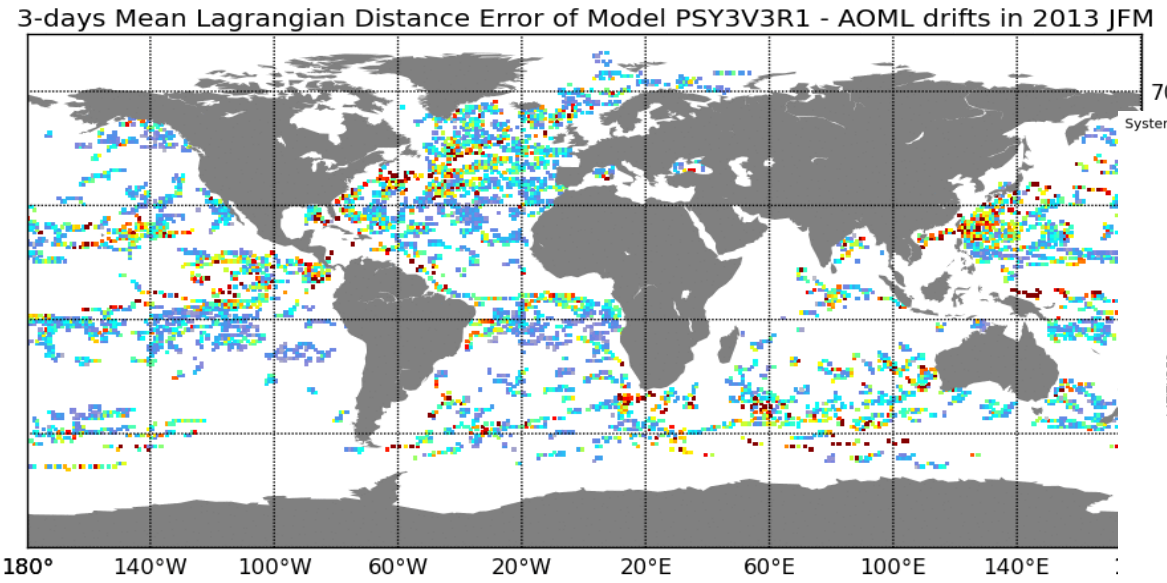


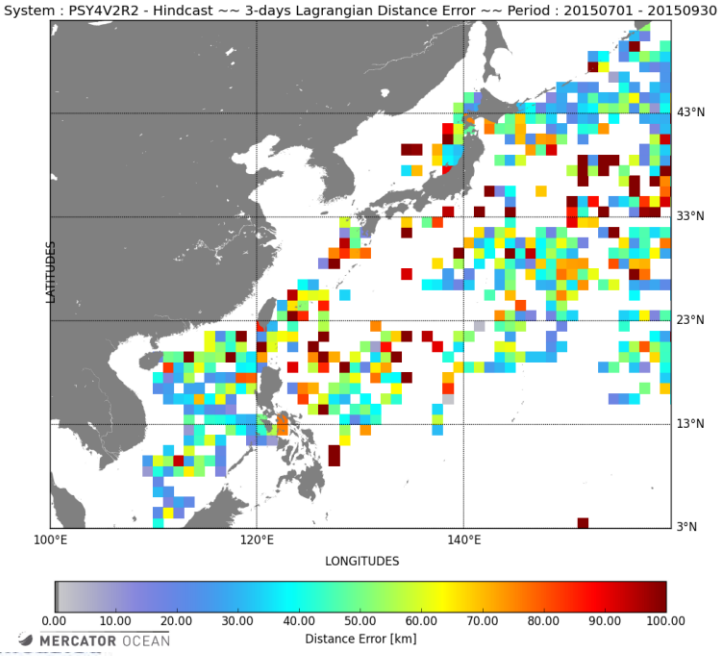
Lagrangian Approach

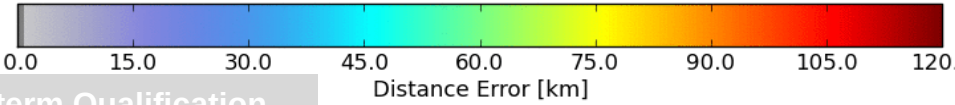
Produce drift estimates using model currents





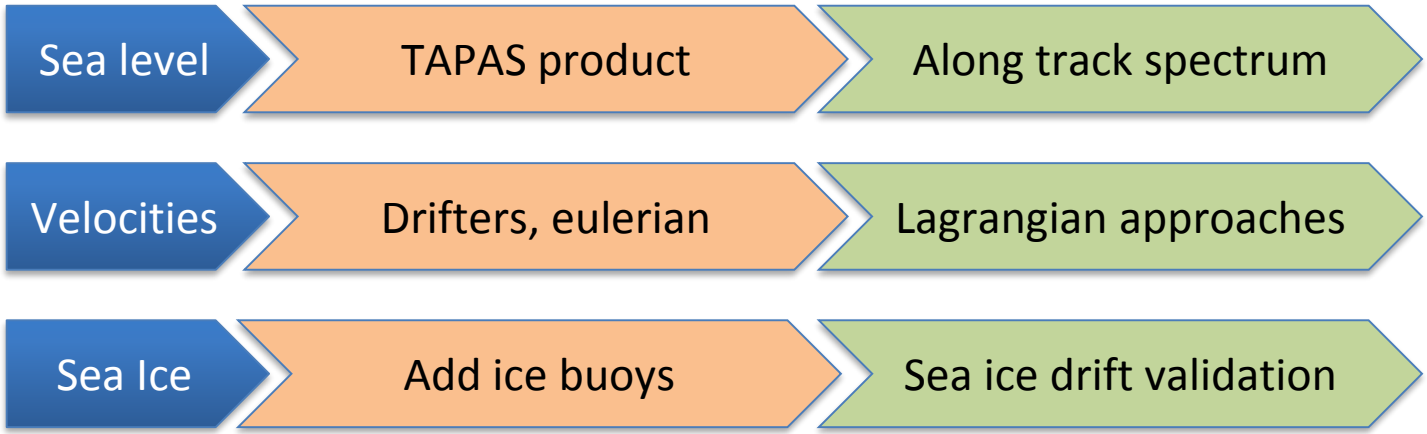


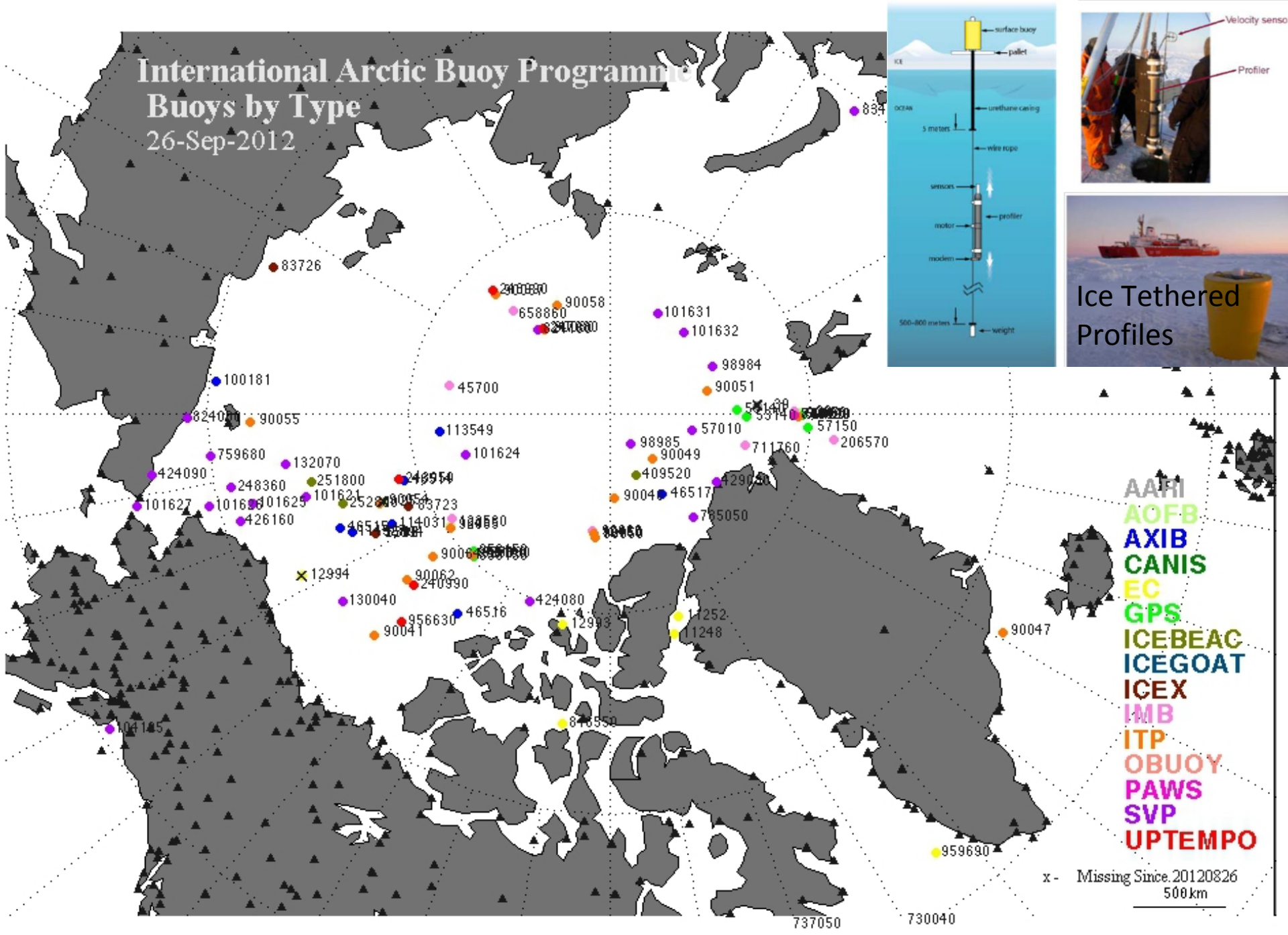




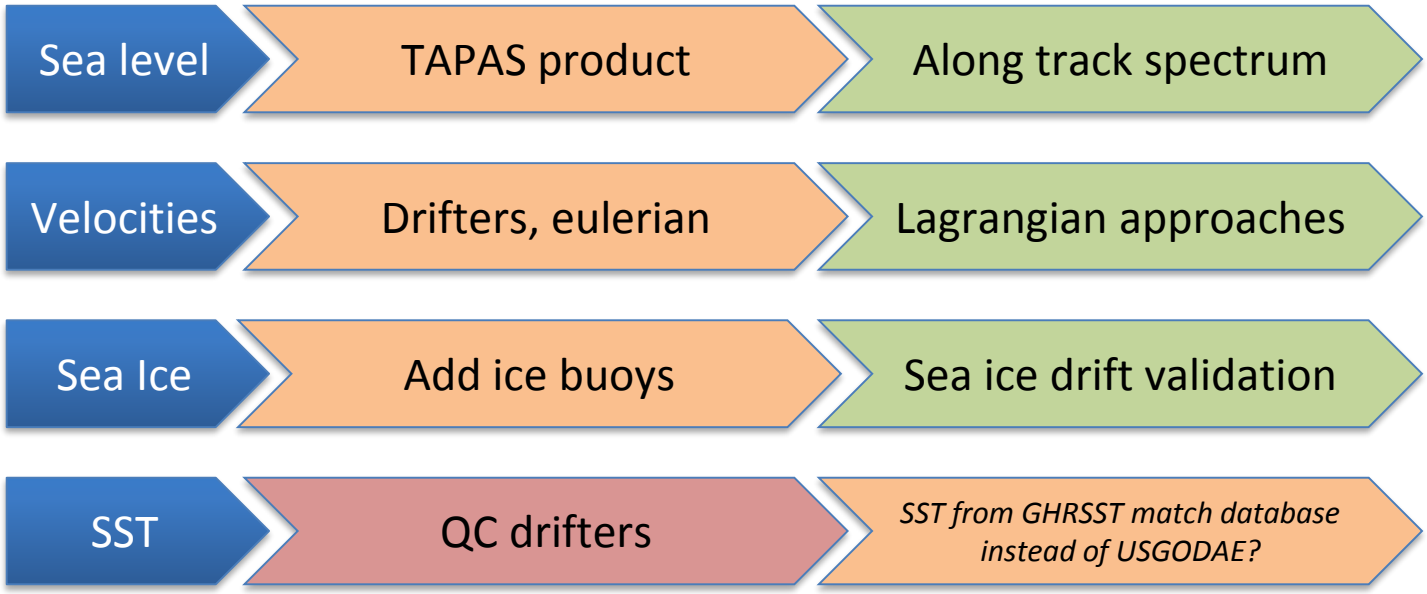
Charly Regnier's contribution

Class 4 metrics task




International Arctic Buoy Programme
Buoys by Type
26-Sep-2012
AARI
AOFB
AXIB
CANIS
EC
GPS
ICEBEAC
ICEGOAT
ICEX
IMB
ITP
OBUOY
PAWS
SVP
UPTEMPO
x - Missing Since 20120826
500 km
surface buoy
pallet
urethane casing
5 meters
wire rope
sensors
profiler
motor
modem
500–800 meters
weight
Velocity sensor
Profiler
Ice Tethered Profiles

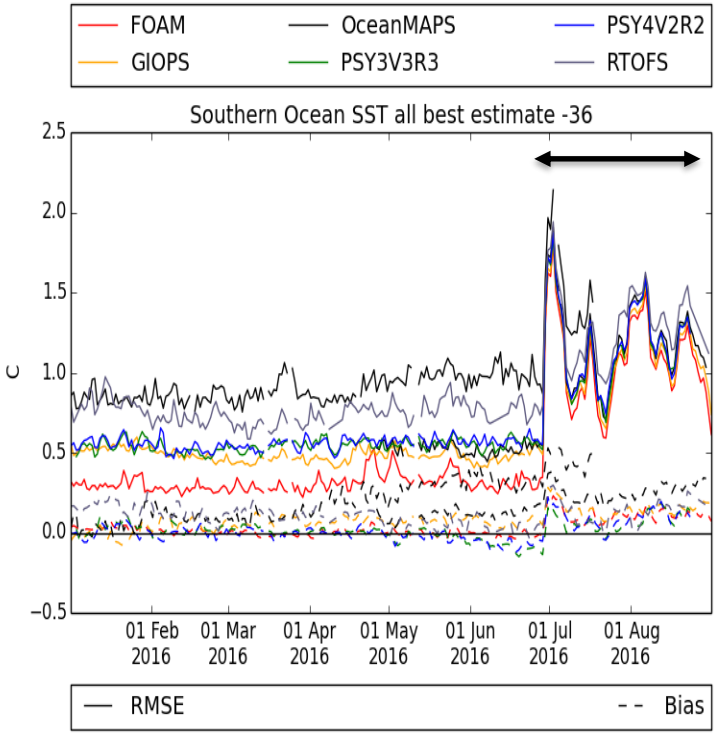
Class 4 metrics task

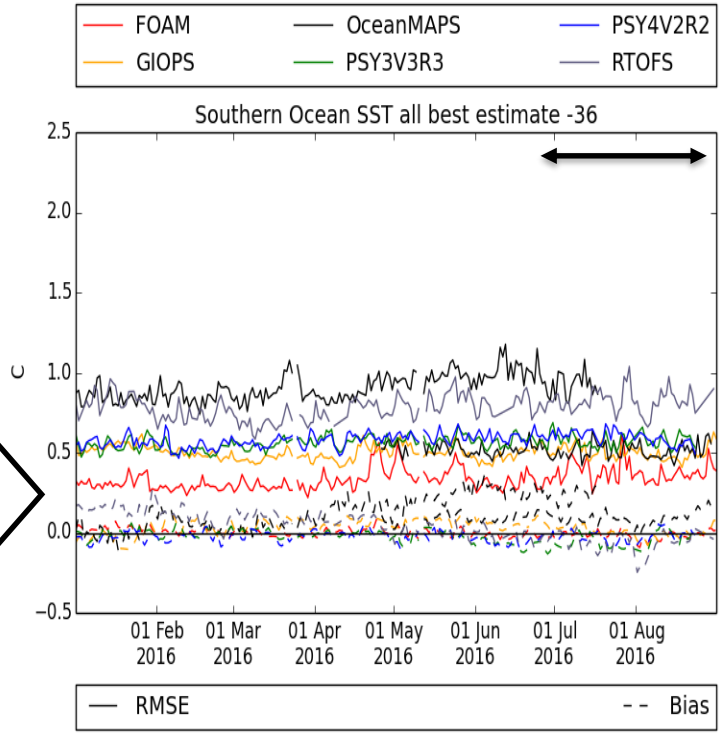


Cleansed Southern Ocean SST errors

Before -2C screening

After screening

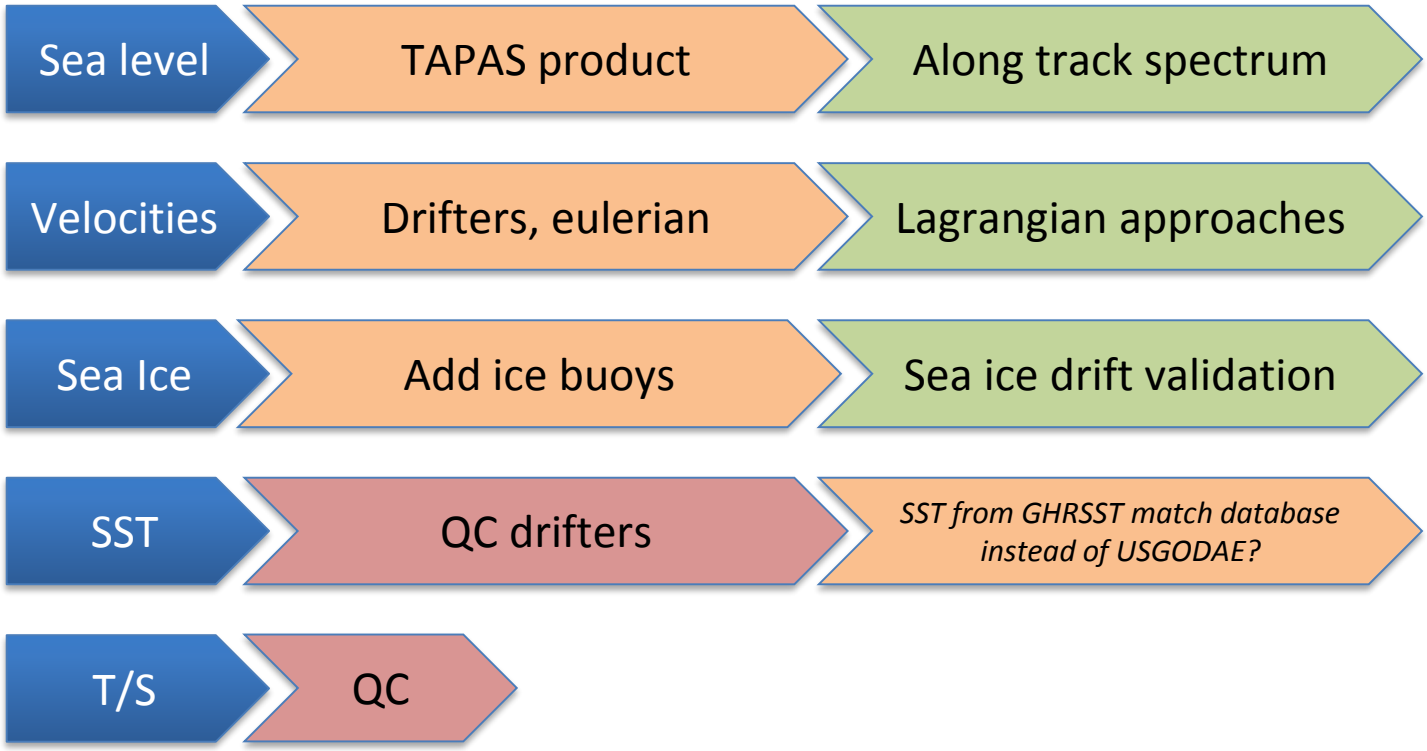




Andy Ryan's contribution

© Crown Copyright 2016, Met Office

Class 4 metrics task



Problematic Argo floats

(Manually detected)

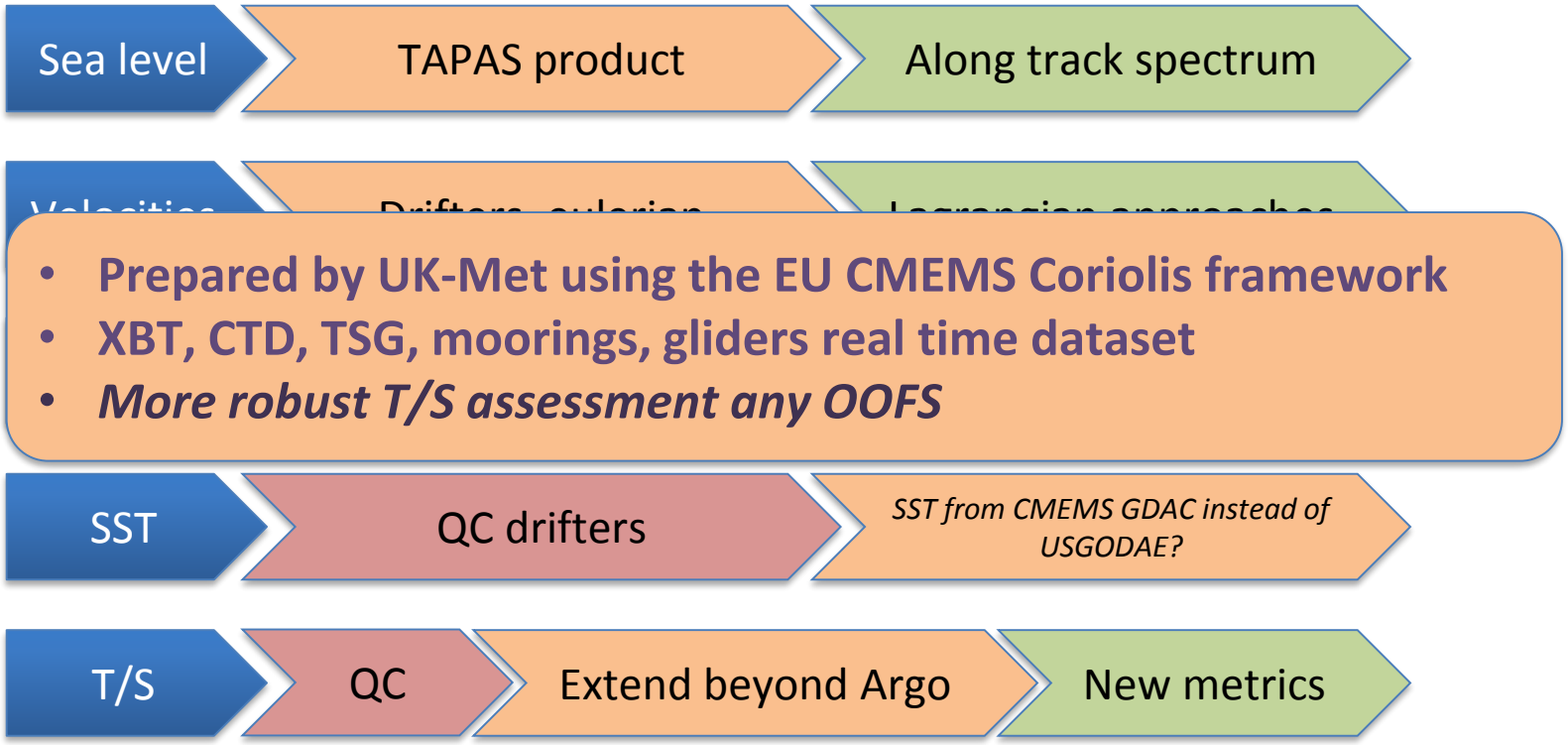
ID	Region	Symptom
4901707	North Atlantic	Step function 10/35 PSU
5904466	Indian Ocean	Large difference from GODAE partners and
6900320	Mediterranean Sea	
6901846	Mediterranean Sea	

- **All OOFS adopt common QC when processing Class 4 files**
- ***Intent blacklisting with Coriolis GDAC***

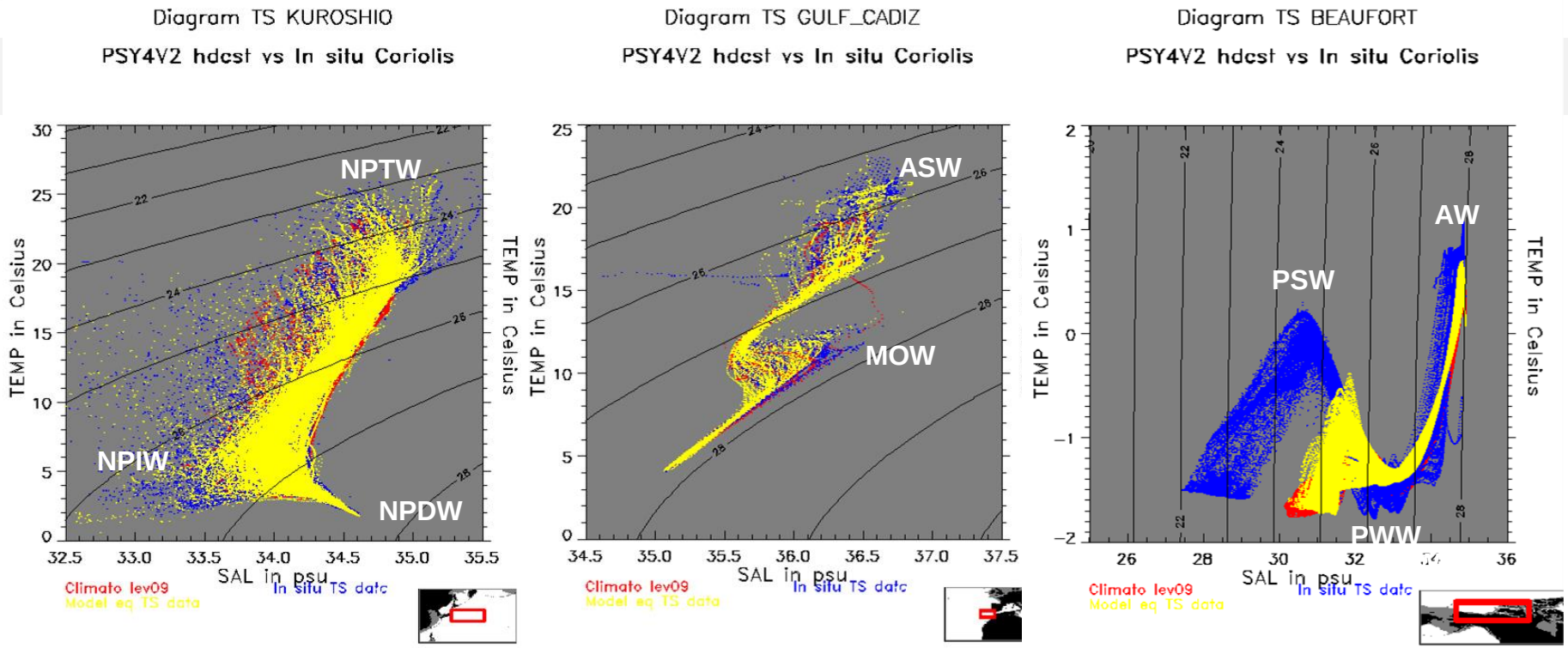
Andy Ryan's contribution

© Crown Copyright 2016, Met Office

Class 4 metrics task



T/S : control of Water masses



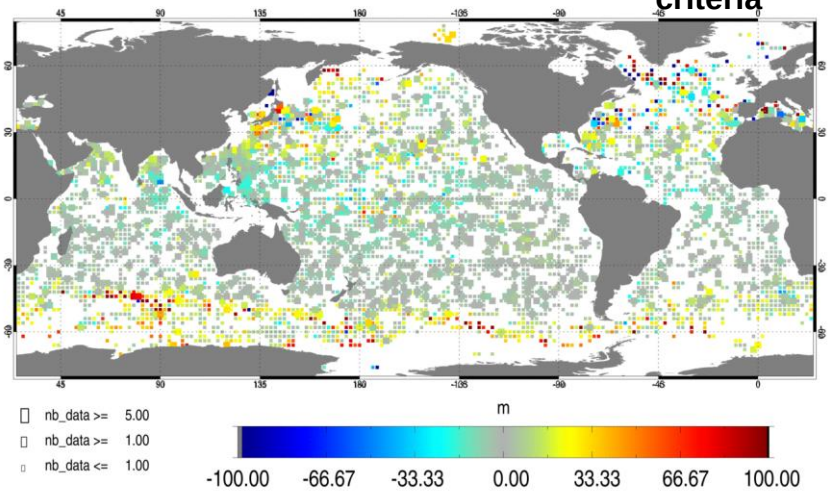
- T/S diagrams of well known water masses
- Identify some missing or not well represented water masses in operational systems

Charly Regnier's contribution

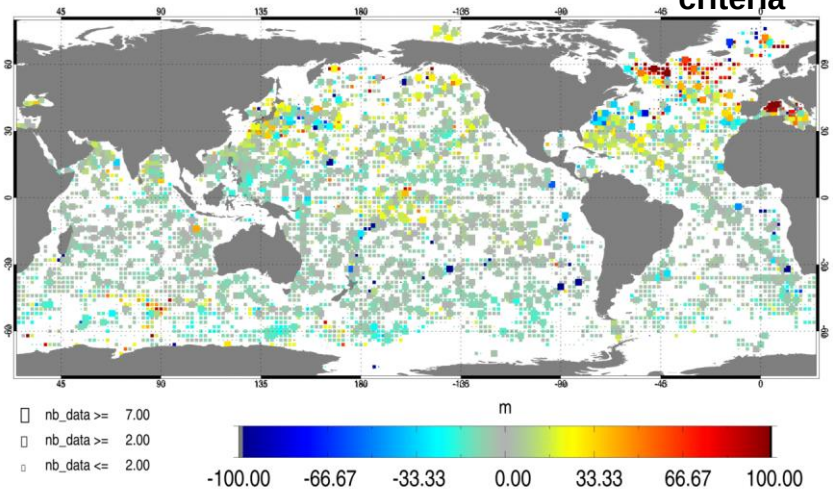
Evaluation with derived quantities

We can compute the MLD in the Argo profiles and in the systems

Density criteria



Temperature criteria



Compare with basics 2D maps or pdf by regions

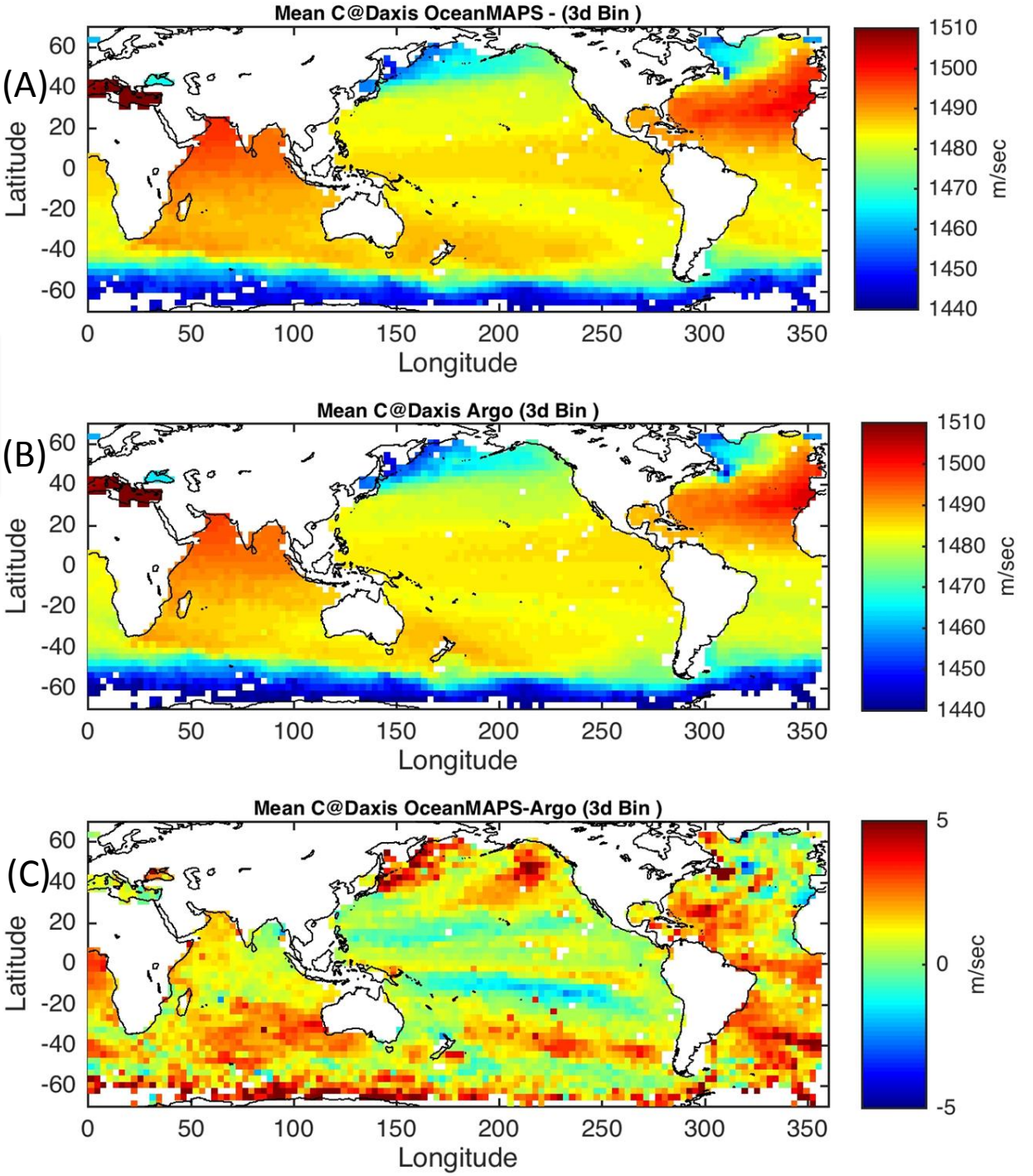
Charly Regnier's contribution

Evaluation of derived quantitites:

Sound Speed at Deep Sound Channel depth

(A) OceanMAPS
(B) Observation
(C) OceanMAPS - Observation

Class 4 metrics can be used to measure the errors in sound properties from ocean forecasting systems, such as OceanMAPS, FOAM, GIOPS, PSY etc.



Prasanth Divakaran

ANNUAL REVIEW

- Provide a **demonstration** based on available material (at present Class 4 results, additional assessment highlighted year after year)
- **Monitoring and assessing** the reliability of GOV OOFS every year, with regard to **observing system changes** or **OOFS upgrades (progresses)**
- September to September synthesis,
- Prepared **by experts of the IV-TT** contributing OOFS for their own benefit: this is also a **coordination tool**,
- Published in February the next year

Issues:

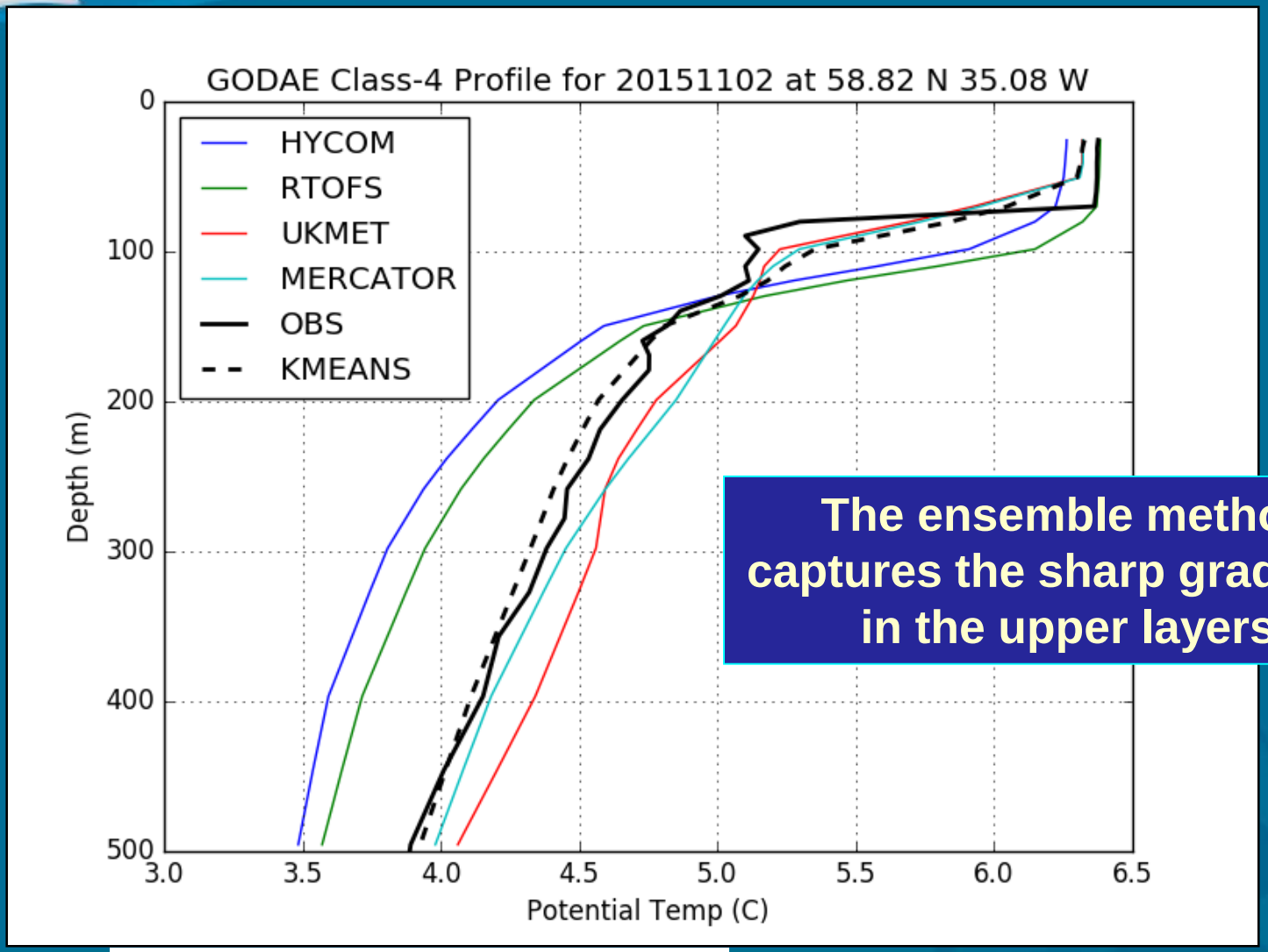
- Subject to available resources
- More of an “operational” than a “research” task.
  - ➢ Role for JCOMM ET-OOFS?

CLASS 1: Ensemble Approaches & Consensus Forecast

- Started in 2013, **as a limited effort activity**
- in order to offer **real time opportunities** for participants, able to enhance emergency forecast tasks/decisions with complementary estimated
- Focusing **on surface parameters**, daily forecast
- Every OOFS participant asked to maintain a 2-3 month rolling archive on ftp
- Scientific task by Todd Splinder, NOAA/NCEP: testing ensemble computations

- **Model state** assessment, complementary to the Class 4 observation space approaches
- Decision to **standardize on common grids** for at "eddy resolving scales" (1/12°)
- Extend **to full depth parameters**, **short term forecast**
- Repository issues

Class 1 Ensemble and Class-4 ARGO Profiles



The ensemble method captures the sharp gradients in the upper layers

TT vision and strategic outlook for the next 10 years and how it contributes to overall Operational Oceanography advancement

- Gradually reduce focus on operational aspects of Class4 and Class1 activities to focus more on relevant scientific questions
  - New metrics, e.g.:
    - Additional observed quantities and derived variables
    - Probabilistic metrics and exploring sources of uncertainty in forecasts and sources of error growth ,
    - Mesoscale activity and feature-based studies
- Greater engagement and collaboration with external partners

Possible envisaged collaborations

- The monitoring tools (BoM, CONCEPTS) allows access by other GOV experts to Class 4 metrics assessment
- Global versus regional Class 1, Class 4 assessment: This is an important aspect that will strengthen common activities with **COSS-TT**
- **CP-TT**: common assessment of surface parameters using Class 4, Class 1, including surface forcings
- **MEAP-TT**: Share experience on Class 4 management for bio-argo floats
- Links with **CLIVAR/GSOP**: Polar-ORA-IP started, in parallel to YOPP efforts, need for coordinations.
- YOPP: sharing Class 4, Class 1 in the Arctic would be valuable. Opening of results to larger community risks limiting by-in by new partners.
- Liaise with WMO JWG FVR , WWRP S2S

Taking benefit of the Joint Working Group on Forecast Verification Research (WMO/JWGFVR) advances (B. Casati, ECCC):

GODAE OceanView

Existing verification techniques:

Traditional (point-by-point) methods:

1. graphical summary (scatter-plot, box-plot)
2. Continuous scores (MSE, correlation)
3. Categorical scores from contingency tables (FBI,HSS,PC).
4. Probabilistic verification (Brier, CRPS, rank histograms)

Extreme dependency scores : Ferro and Stephenson (2011) Wea&For (EDI,SEDI)

Spatial methods:

1. Scale-separation
2. Neighbourhood
3. Field-deformation
4. Feature-based
5. Distance metrics for binary images

There is no single technique which fully describes the complex observation-forecast relationship

Key Q1: what do we wish to know from our verification ? (end-user / verification purpose / questions addressed / attributes of interest)

Key Q2: What are the (statistical) characteristics of the variable and forecast verified ? What are the available obs?

Taking benefit of the Joint Working Group on Forecast Verification Research (WMO/JWGFVR) advances (B. Casati, ECCC):

GODAE OceanView

Spatial verification techniques:

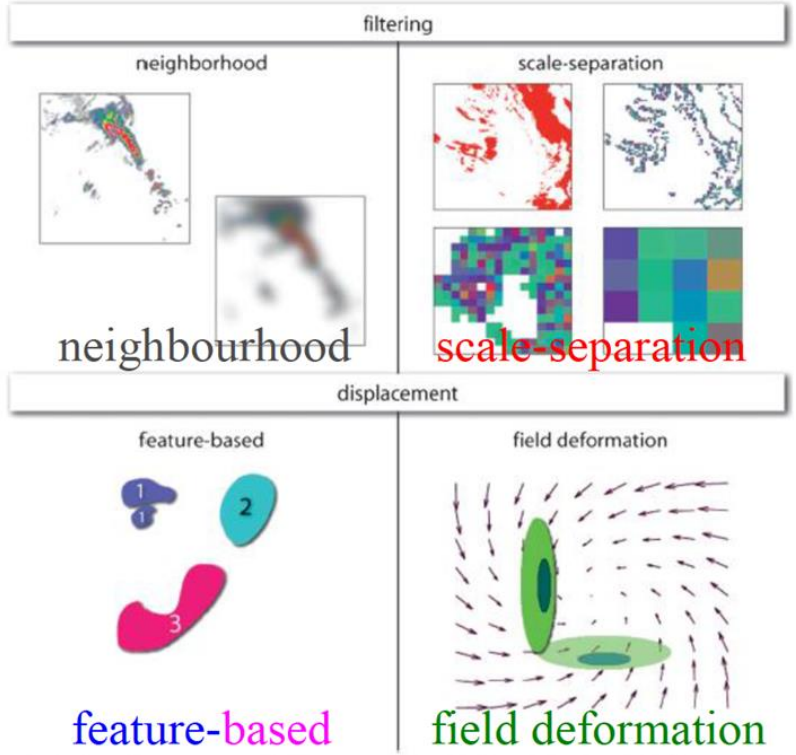


FIG. 1. Schematic representations of the four categories of verification methods reviewed in this paper. (top) The neighborhood and scale-separation methods can both be considered "filtering" approaches while (bottom) the feature-based and field deformation methods fall under the "displacement" category.

The **Spatial Verification Inter-Comparison Project** (ICP), is entering its second phase (MesoVICT).

http://www.ral.ucar.edu/projects/icp

Includes an impressive list of references for spatial verification studies (>200).

Review article: Gilleland et al (2009): *Inter-comparison of Spatial Forecast Verification Methods*. Wea&For, **24**

Spatial verification techniques **aim** to:

- account for field spatial structure and the presence of features
- provide information on error in physical terms (meaningful verification)
- Assess location and timing errors (separate from intensity error)
- account for small time-space uncertainties (avoid double-penalty issues)

Summary

- Class 4 Intercomparison is well-underway and showing improving robustness of data feeds and GOV prediction system skill:
  - Can be used for demonstration, results highlight relative strengths and weaknesses of various systems and allow a measure of progress
  - Operational use (adopted by BoM, NCEP, CONCEPTS)
  - Improving “every day” activity is ongoing
  - Extensions to new metrics, and new parameters (velocity, sea ice)
- Class 1 Intercomparison to be strengthened
  - Model state approaches
  - Standardisation and common grid
- Growing IV-TT community and new possibilities inside and outside GOV-ST
- Growing feeling IV-TT needs to focus more on research questions and less on operational issues

Thank you

Extras

Managing IV-TT activities: decisions

- A wiki and forum to :
  - **Store documentation** on producing Class 4 (and others)
  - **Provide tools** (interpolation…), methods to perform the assessment (QC recom…)
  - Discuss on **new metrics**
  - **Track changes** on OOFS, anomalies on assessment

Managing IV-TT activities: decisions

- A wiki and forum to :
  - Store documentation on producing Class 4 (and others)
  - Provide tools (interpolation…), methods to perform the assessment (QC recom…)
  - Discuss on new metrics
  - Track changes on OOFS, anomalies on assessment
- Monitoring and demonstration (at present from Class 4 assessment) on GOV IV-TT pages

Managing IV-TT activities: decisions

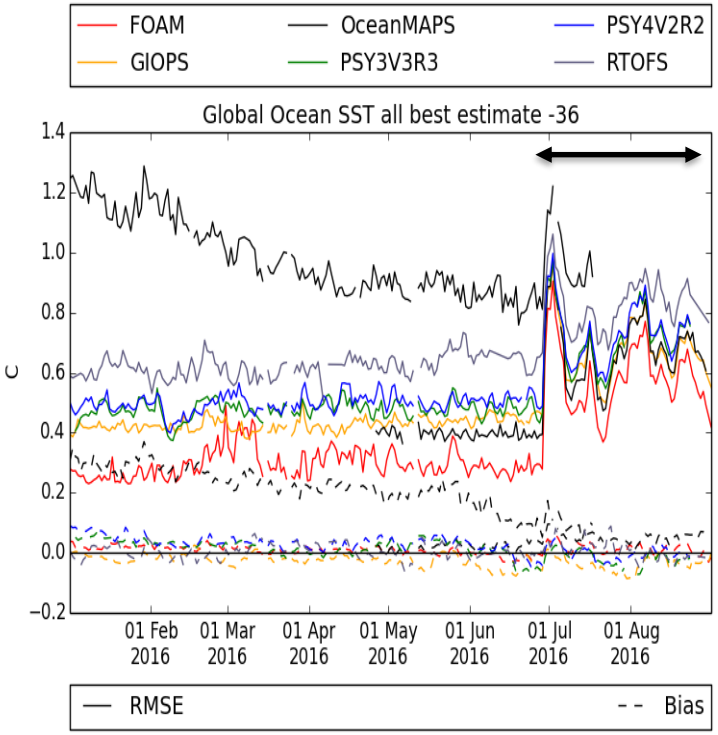
- A wiki and forum to :
  - Store documentation on producing Class 4 (and others)
  - Provide tools (interpolation…), methods to perform the assessment (QC recom…)
  - Discuss on new metrics
  - Track changes on OOFS, anomalies on assessment
- Monitoring and demonstration (at present from Class 4 assessment) on GOV IV-TT pages
- Scheduling the annual reporting each year
- Offering visualisation to other GOV-ST task teams (BoM initiative)

More unexpected SST signals

SST errors for all systems have grown dramatically in the last few months.

Is this a global trend or is it regional?

Is it a true reflection of GODAE system accuracy?

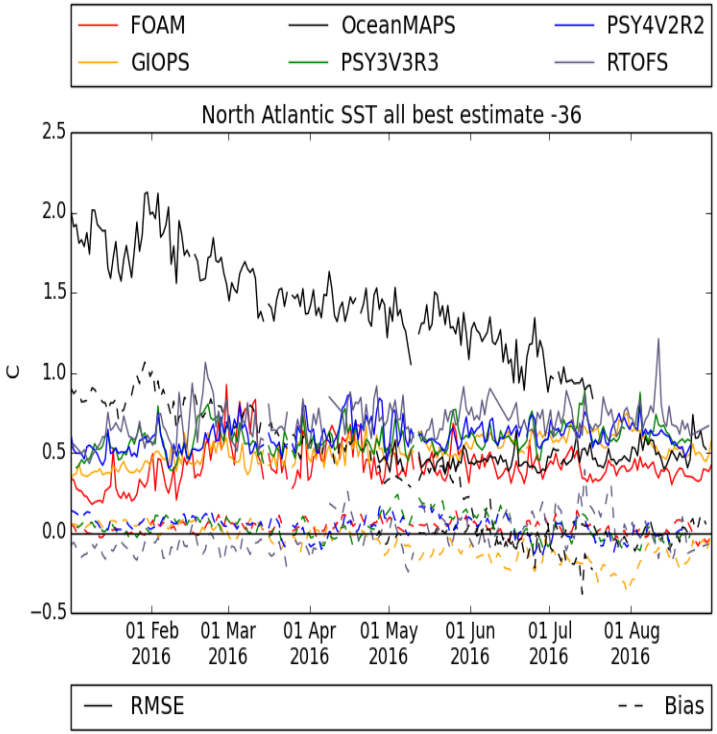


Andy Ryan's contribution

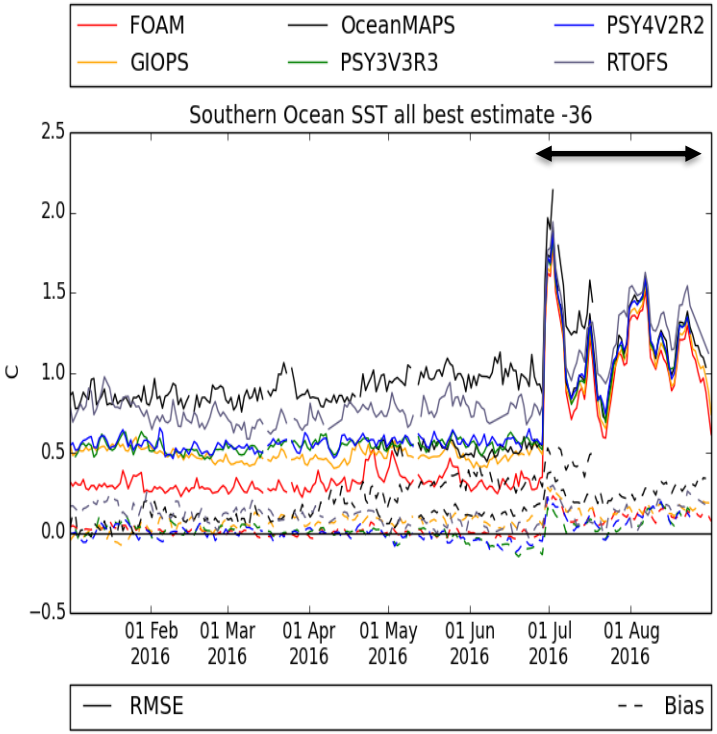
© Crown Copyright 2016, Met Office

Where are the errors located?

North Atlantic



Southern Ocean



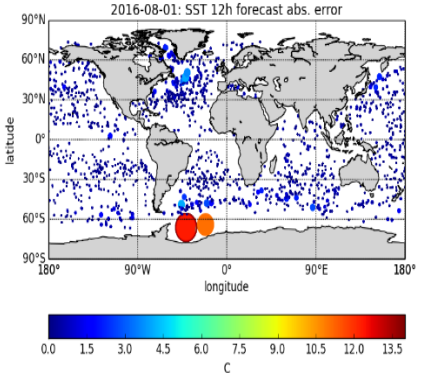
Andy Ryan's contribution

© Crown Copyright 2016, Met Office

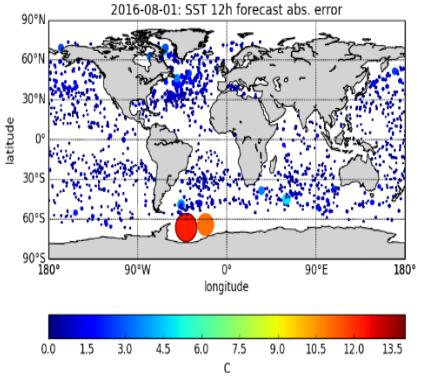
Southern Ocean SST errors

Across GODAE partners: greater than 10C error 2016/08/01

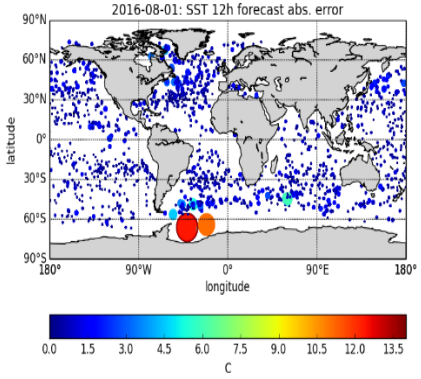
FOAM



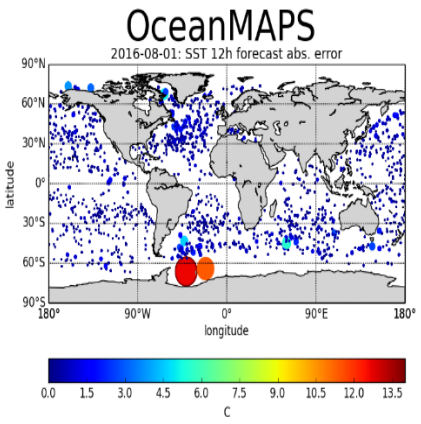
GIOPS



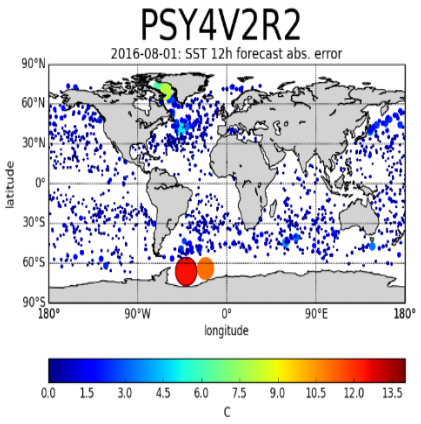
RTOFS



OceanMAPS



PSY4V2R2



Are all ocean forecasts wrong by **14C** in the Southern Ocean?

Andy Ryan's contribution

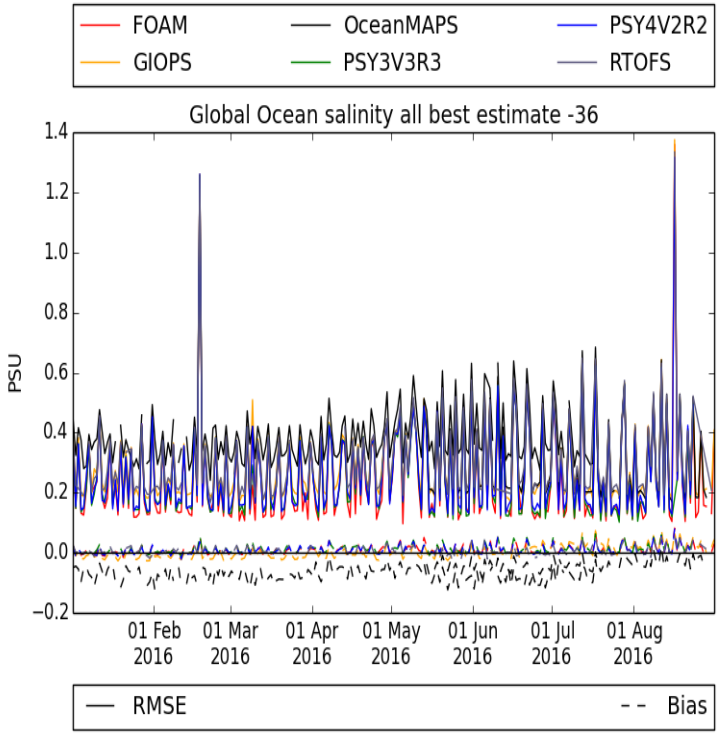
© Crown Copyright 2016, Met Office

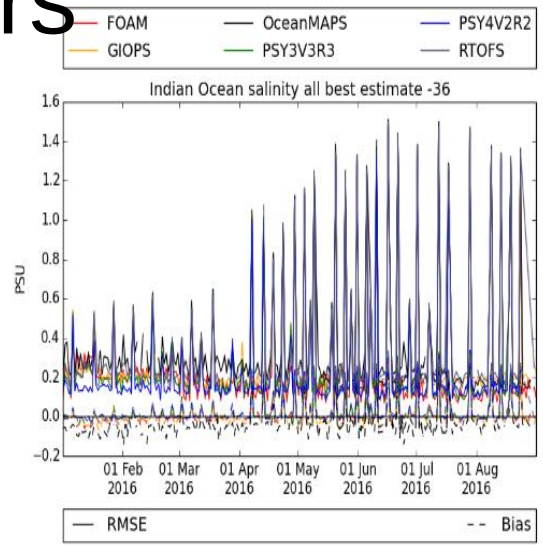
Spurious salinity errors

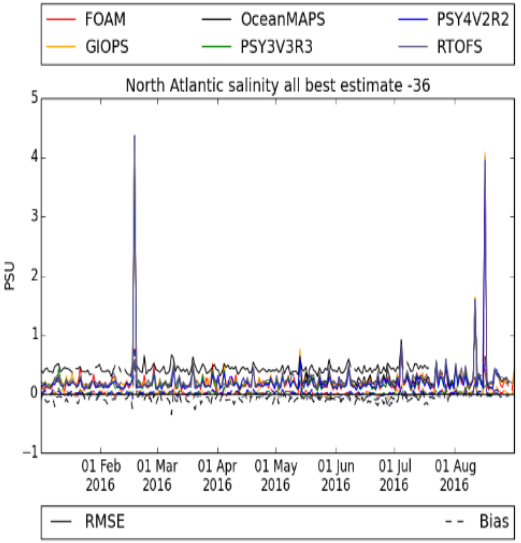
Are they real?

Are they processing faults?

Can we understand their causes?







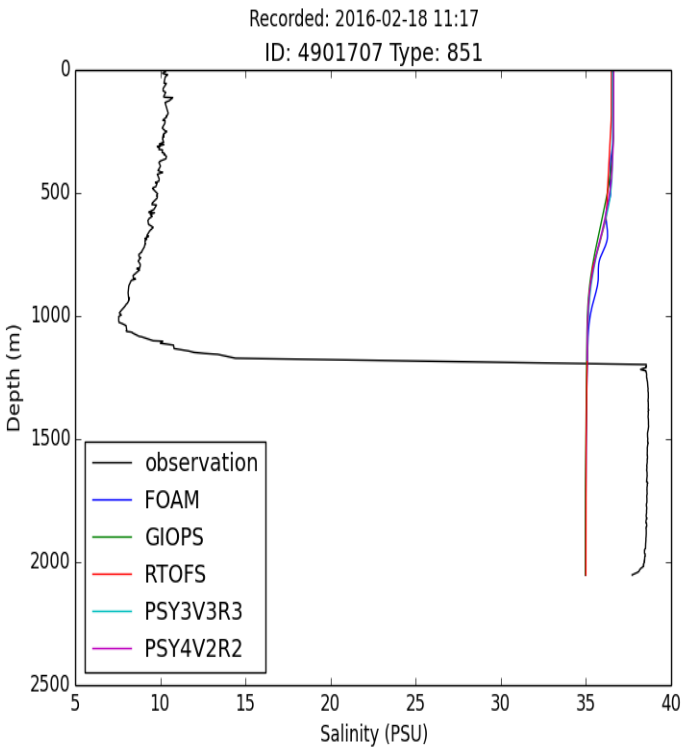
Andy Ryan's contribution

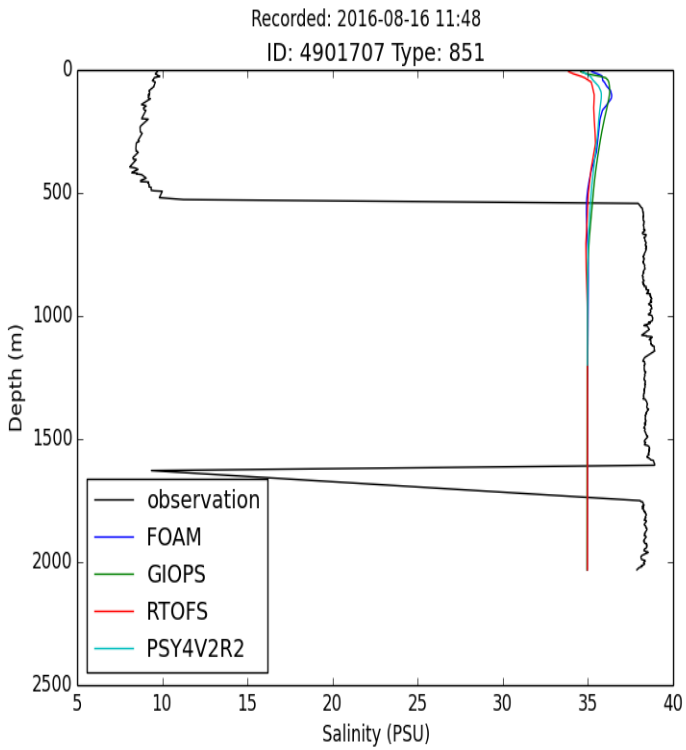
© Crown Copyright 2016, Met Office

Are these realistic profiles?

North Atlantic float with step function between 5 and 38 PSU

GODAE systems suggest ~35PSU throughout water column





Andy Ryan's contribution

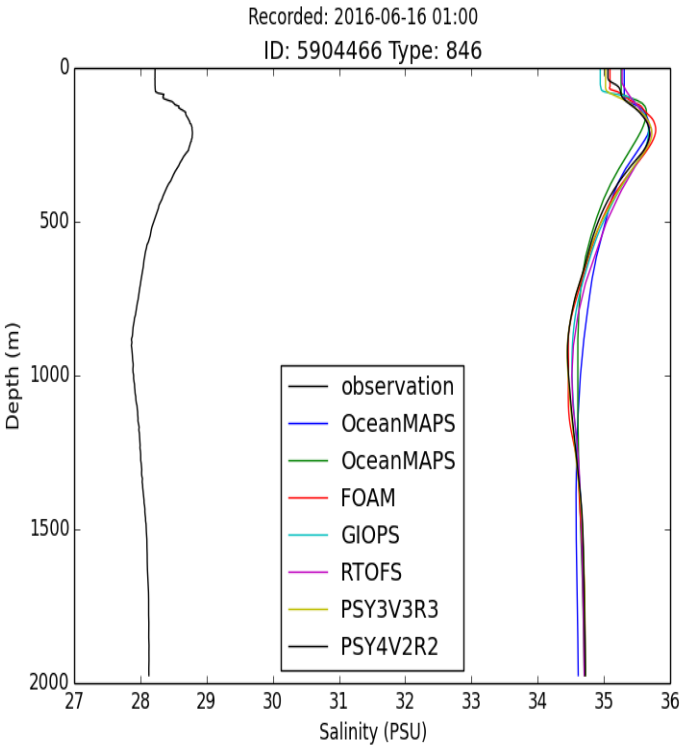
www.metoffice.gov.uk

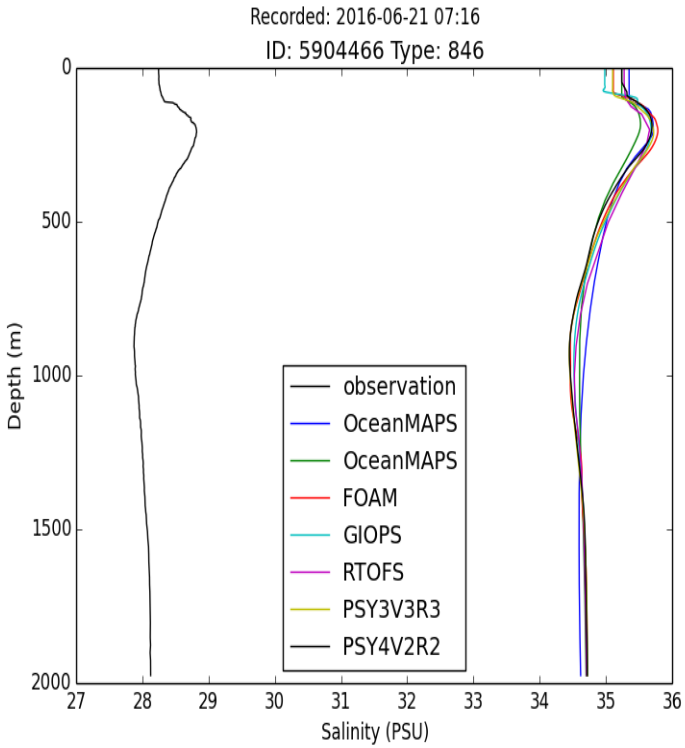
© Crown Copyright 2016, Met Office

Are these realistic profiles?

Indian Ocean float consistently different from nearby Argo floats

GODAE systems suggest profile should be +6PSU





Andy Ryan's contribution

© Crown Copyright 2016, Met Office

The Ensemble Components

Member Models:

- UK Met Office Global Seasonal forecast system (GloSea5, 1/4°)
- Operational Mercator Ocean Global Ocean forecasting system (1/12°)
- US Navy HYCOM + NCODA Global Analysis (1/12°)
- NCEP Operational Global Real Time Ocean Forecasting System (1/12°)

Reference Data (Daily Observations):

- SST, SSS and the North Atlantic Regional Ensemble: FNMOC High Resolution Ocean Analysis for GODAE (1/6°)
- U, V: SSALTO/DUACS MADT Near Real Time Absolute Geostrophic velocities from multi-satellite observations (1/4°)

Climatologies:

- SST, SSS: World Ocean Atlas 2013 V2
- U, V: GLORYS2V3 22-year reanalysis bootstrapped as a proxy

Ensemble Workflow

Acquire Member Data

↓

Interpolate to reference grid

↓

Calculate Clusters

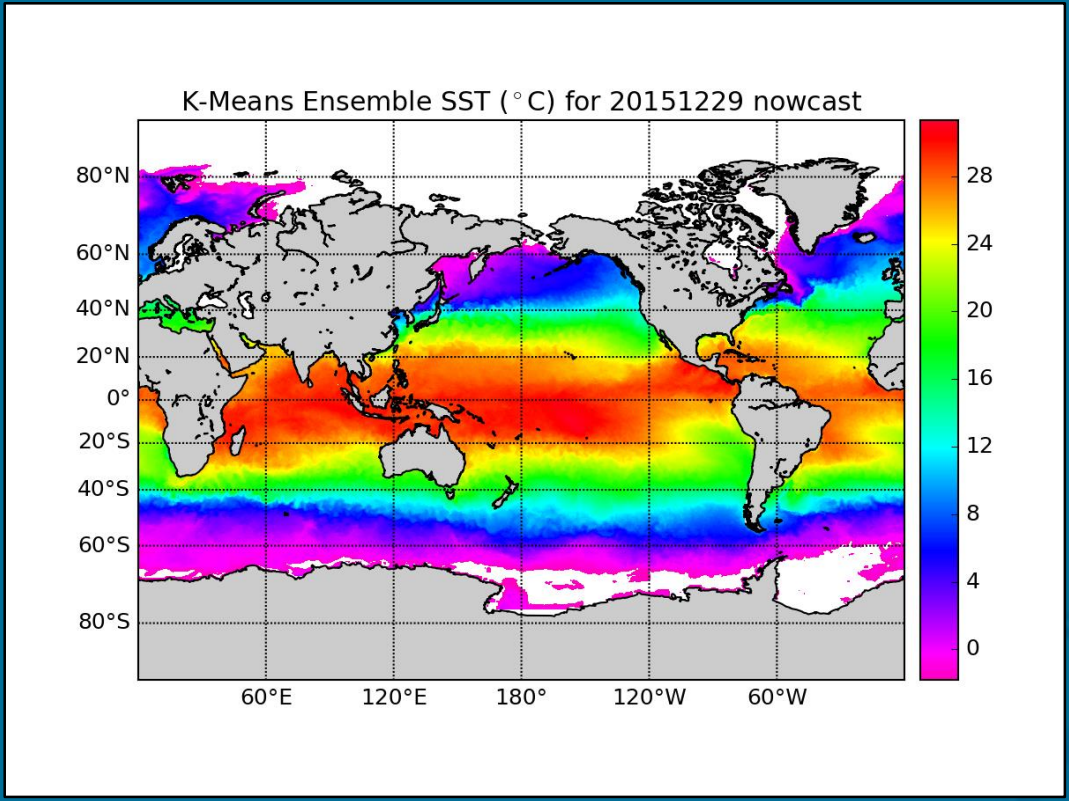
↓

Calculate weights

↓

Combine all members

Inverse weighted averages using anomaly RMSE within each cluster

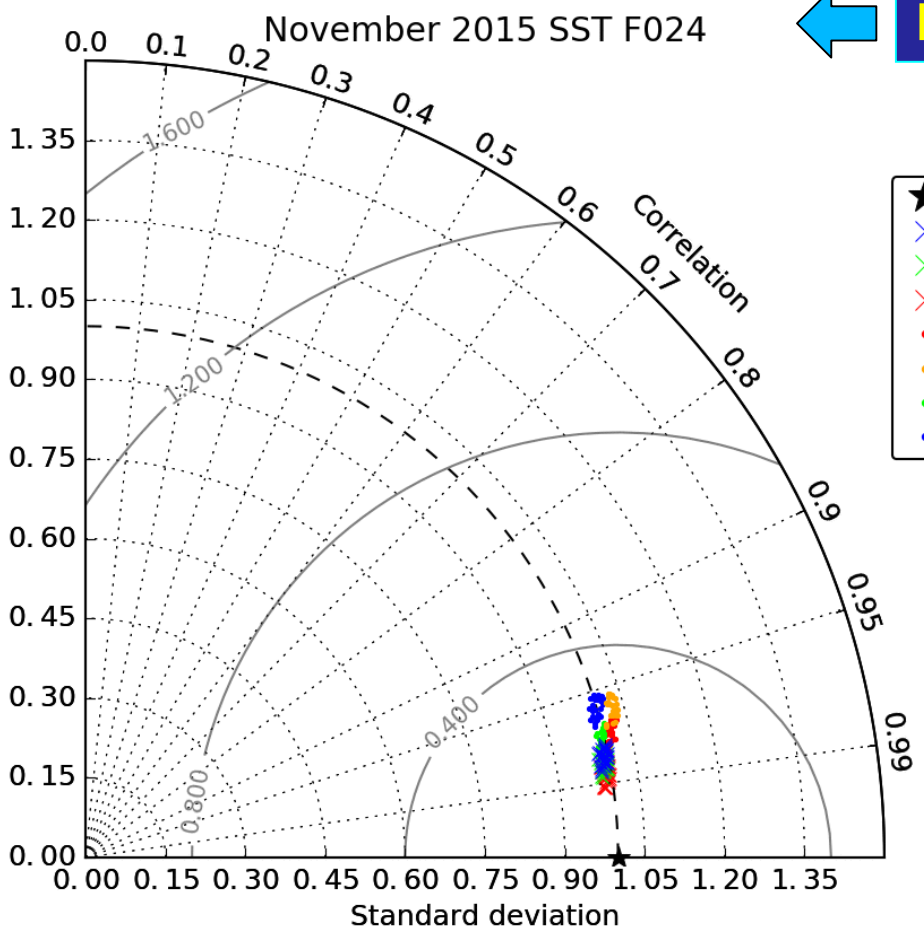


- FNMOC[1] or MADT[2] as reference field for calculating weights
- WOA 2013 V2 or GLORYS2V3 as climatology
- 1/12° resolution
- 4 Members
- 7 Clusters

[1] FNMOC High Resolution Analysis for GODAE (SST, SSS)
[2] SSALTO/DUACS MADT Near Real Time Geostrophic Surface Currents

Todd Spindler's contribution

Class 1 Ensemble and Class 4 SST Validation

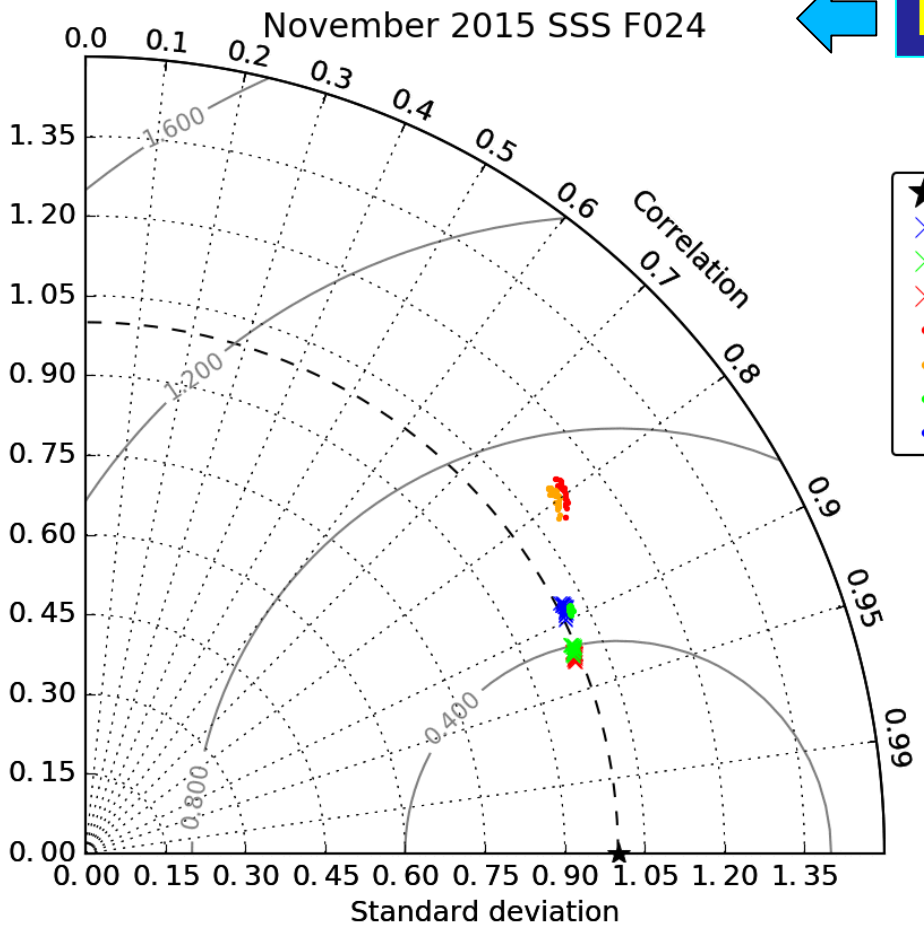


Forecast loop

26 days of data

All points normalized by the standard deviation of the observations

Class 1 Ensemble and Class 4 SSS Validation

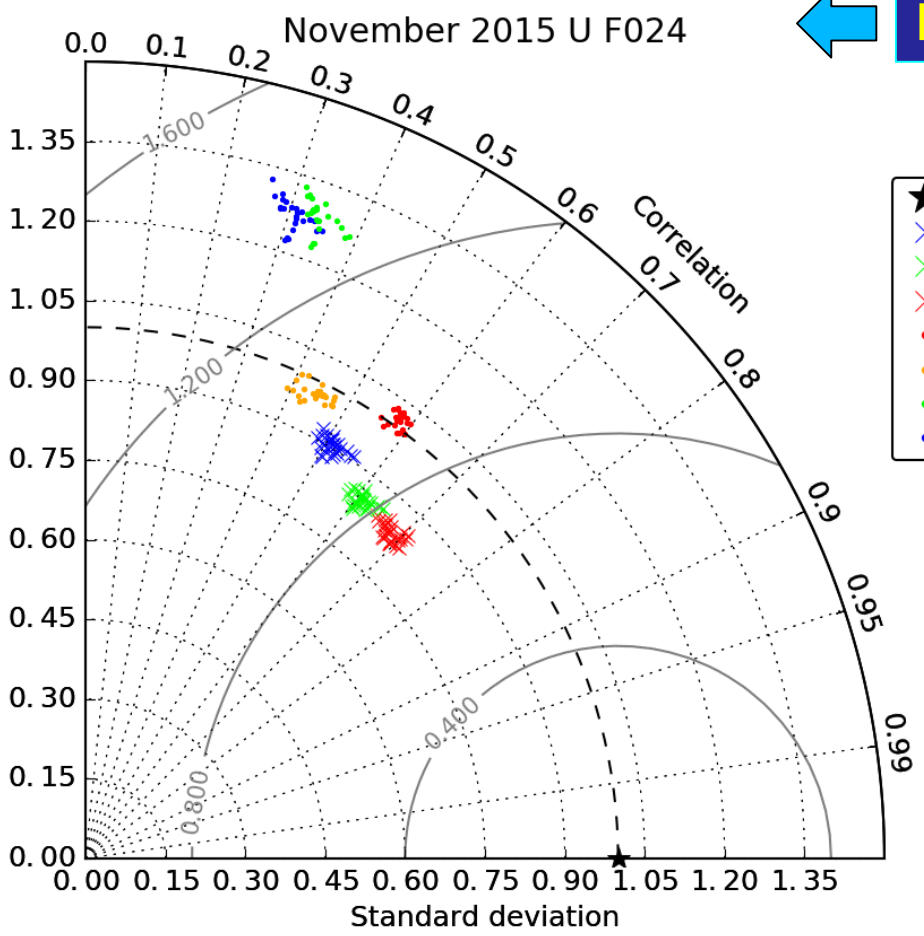


Forecast loop

26 days of data

All points normalized by the standard deviation of the observations

Class 1 Ensemble and Class 4 U Validation

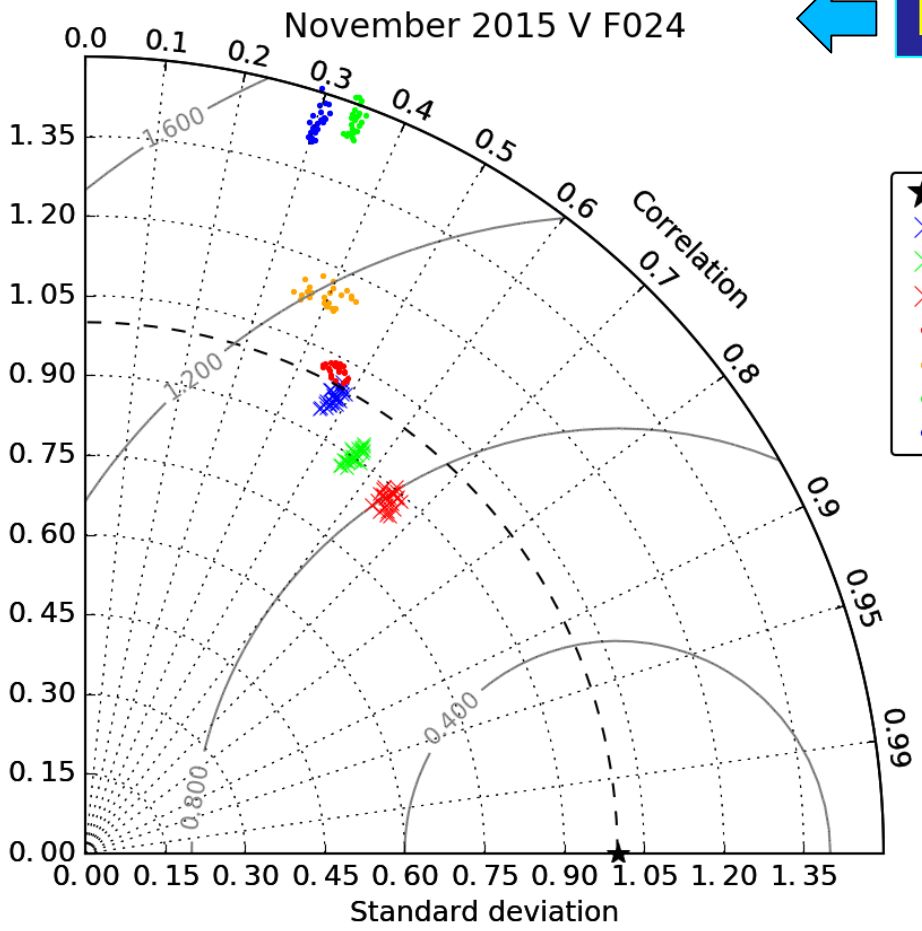


Forecast loop

26 days of data

All points normalized by the standard deviation of the observations

Class 1 Ensemble and Class 4 V Validation

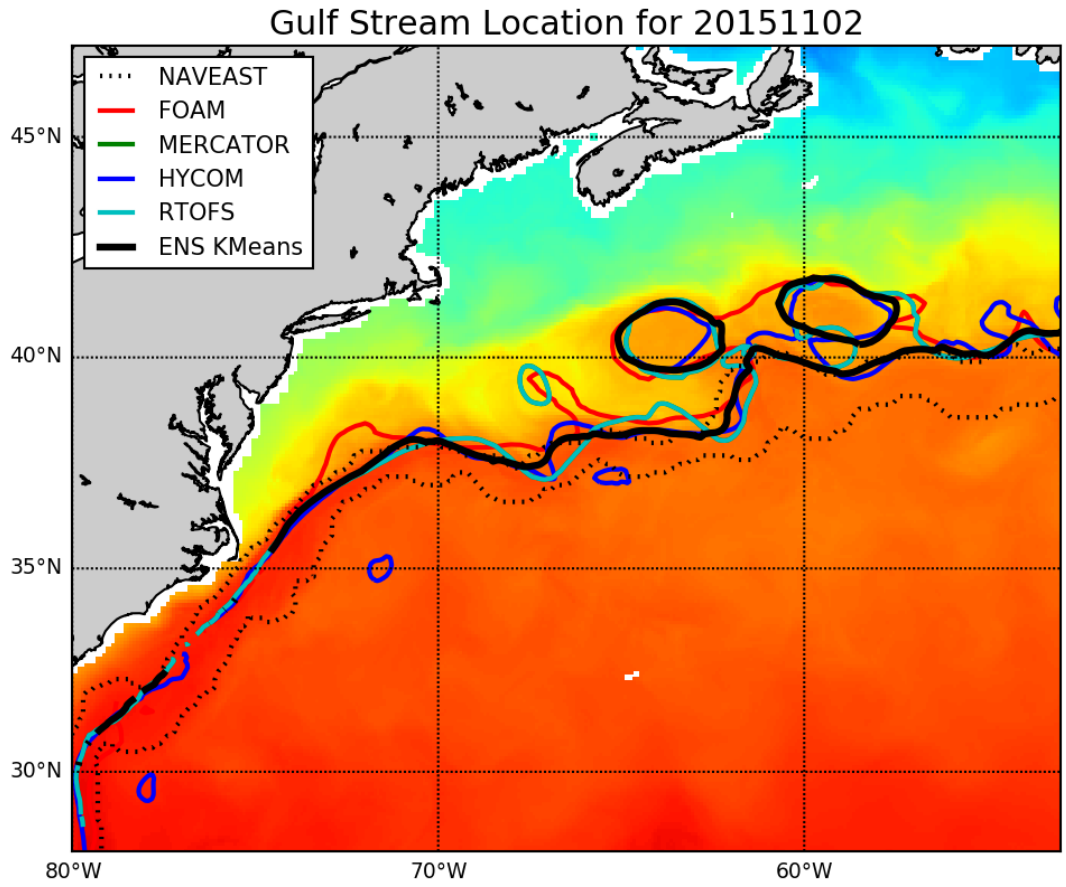


Forecast loop

26 days of data

All points normalized by the standard deviation of the observations

Gulf Stream Location



Front location defined as the intersection of the 12°C isotherm and the 400m isobath

(Halkin and Rossby, 1985)

Gulf Stream north and south walls from the Naval Eastern Ocean Center