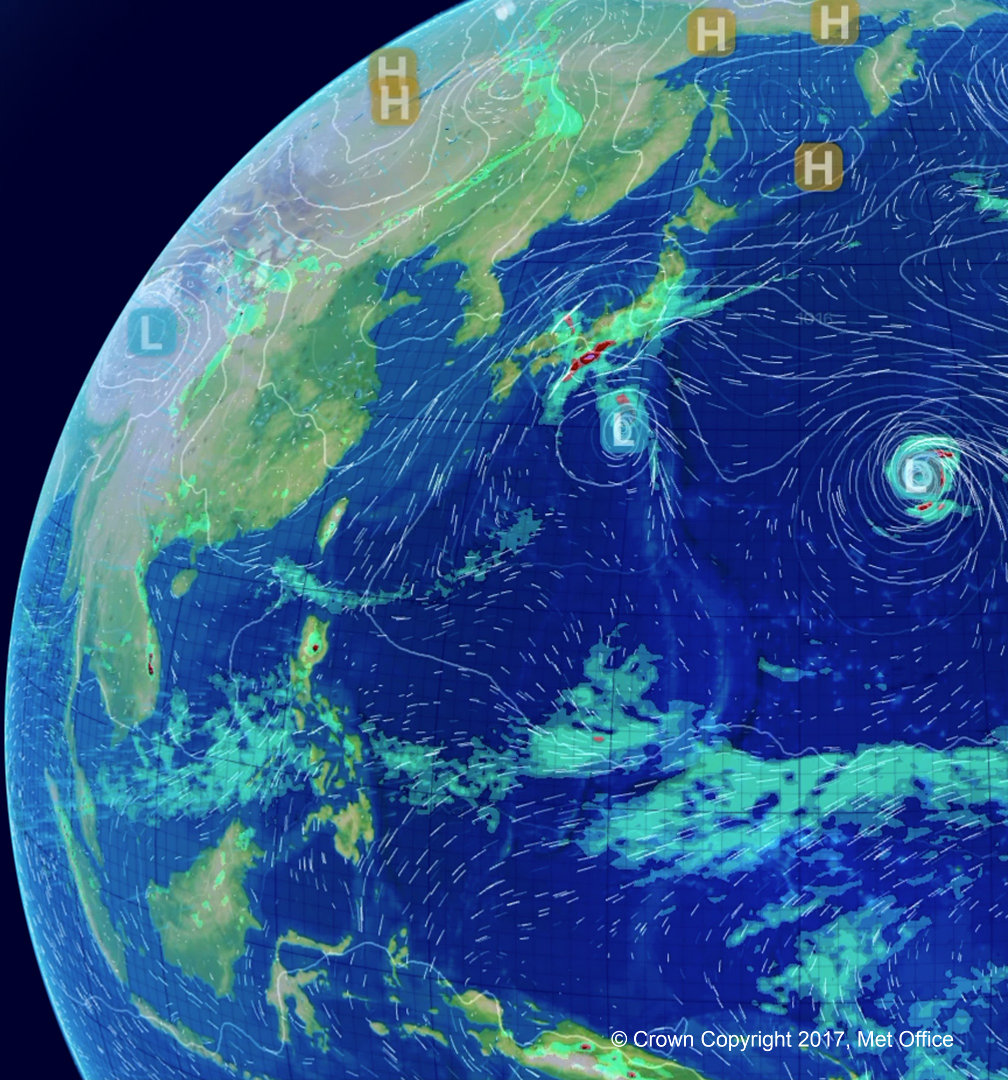
ERA-CLIM2 WP2 achievements

Matt Martin.

ERA-CLIM2 review, University of Bern,

15th December 2017.



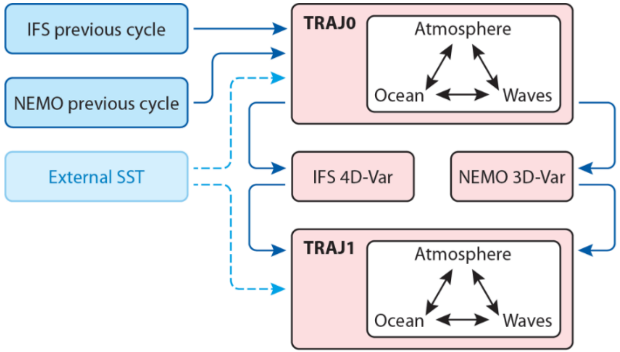
WP2 objectives
Future coupling methods

Research and development in coupled data assimilation for climate reanalysis, and work on development of the carbon component.

Developments will be available for implementation in the CERA (Coupled ECMWF Reanalysis) framework developed at ECMWF.

The work package addressed the special requirements for the pre-satellite data-sparse era and the requirement to maintain a consistent climate signal throughout the entire reanalysis period.

- T2.1: Coordination and management
- T2.2: To include SST and sea-ice assimilation in NEMOVAR
- T2.3: To improve the ocean analysis component including use of ensembles and 4D-VAR
- T2.4: Development of the carbon component of coupled earth system reanalysis
- T2.5: Towards development of fully coupled data assimilation



WP2 status of deliverables

Deliverable number	Deliverable title	Delivery date	Type
D2.1	Assimilation of sea-surface temperature observations [METO]	27 => 39	Code + documented results
D2.2	Assimilation of sea-ice observations [MERCO]	27 => 39	Code + documented results
D2.3	Ensemble-based covariance estimates [CERFACS]	34 => 46	Code + documented results
D2.4	Ensemble-based covariances in coupled data assimilation [CMCC]	24 => 36	Report
D2.5	4D-Var in NEMOVAR [INRIA]	27 => 39	Report
D2.6	Optimised model parameters for the carbon cycle [UVSQ]	34 => 46	Report
D2.7	Alternatives for coupling ocean biogeochemistry [MERCO]	34 => 46	Report
D2.8	Weakly coupled assimilation methods [UREAD]	18	Report
D2.9	Covariances from weakly coupled data assimilation [METO]	18	Report
D2.10	Coupled-model drift [UREAD]	34 => 46	Report
D2.11	Fully coupled data assimilation [INRIA]	34 => 46	Report
D2.12	Status report WP2 [METO]	8	Report

All deliverables completed, reviewed and submitted, except for D2.6 for which some minor revision is being made (will be submitted on Monday 18th Dec).

Code developments:

- All relevant code developments have been made available in the NEMOVAR code repository hosted at ECMWF.
- A new version of the NEMOVAR code (v5), containing all the ocean DA developments made in ERA-CLIM2 is about to be released.

WP2 papers

13 papers have been published, submitted or in preparation so far:

1. Feng, X., et al., 2017 Coupling of surface air and sea surface temperatures in the CERA-20C reanalysis, Quart. J. Roy. Met. Soc.
2. Feng, X., and K. Haines, 2017 Atmospheric response and feedback to sea surface temperatures in coupled and uncoupled ECMWF reanalyses, In preparation.
3. Lea, D. J., et al., 2015: Assessing a New Coupled Data Assimilation System Based on the Met Office Coupled Atmosphere-Land-Ocean-Sea Ice Model. Monthly Weather Review, 143, 4678-4694, doi: 10.1175/MWR-D-15-0174.1.
4. Mulholland, D. P., P. Laloyaux, K. Haines and M.-A. Balmaseda. Origin and impact of initialisation shocks in coupled atmosphere-ocean forecasts. Mon. Wea. Review.
5. Mulholland, D. P., Haines, K. and Balmaseda, M. A. (2016), Improving seasonal forecasting through tropical ocean bias corrections. Q.J.R. Meteorol. Soc., 142: 2797-2807.
6. Pellerej, R., et al, 2016. Toward variational data assimilation for coupled models: first experiments on a diffusion problem.. CARI 2016, Oct 2016, Tunis, Tunisia. 2016
7. Peylin, P., et al.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, Geosci. Model Dev., 9, 3321-3346.
8. Storto, A., et al. Strongly coupled data assimilation experiments with linearized ocean-atmosphere balance relationships, submitted to MWR.
9. Storto, A., et al., 2017. Constraining the global ocean heat content through assimilation of CERES-derived TOA energy imbalance estimates. Geophysical Research Letters, 44.
10. Storto, A., et al., 2016, Sensitivity of global ocean heat content from reanalyses to the atmospheric reanalysis forcing: A comparative study, Geophys. Res. Lett., 43, 5261–5270.
11. Weaver AT, et al., 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. Q. J. Roy. Meteorol. Soc., 142: 455-471.
12. Weaver A. T., et al. 2017. "Time"-parallel diffusion-based correlation operators. Technical Memorandum 808, ECMWF, Reading, UK.
13. While, J., M.J. Martin, 2017. Variational bias correction of satellite sea surface temperature data incorporating direct observations of the bias. In preparation.

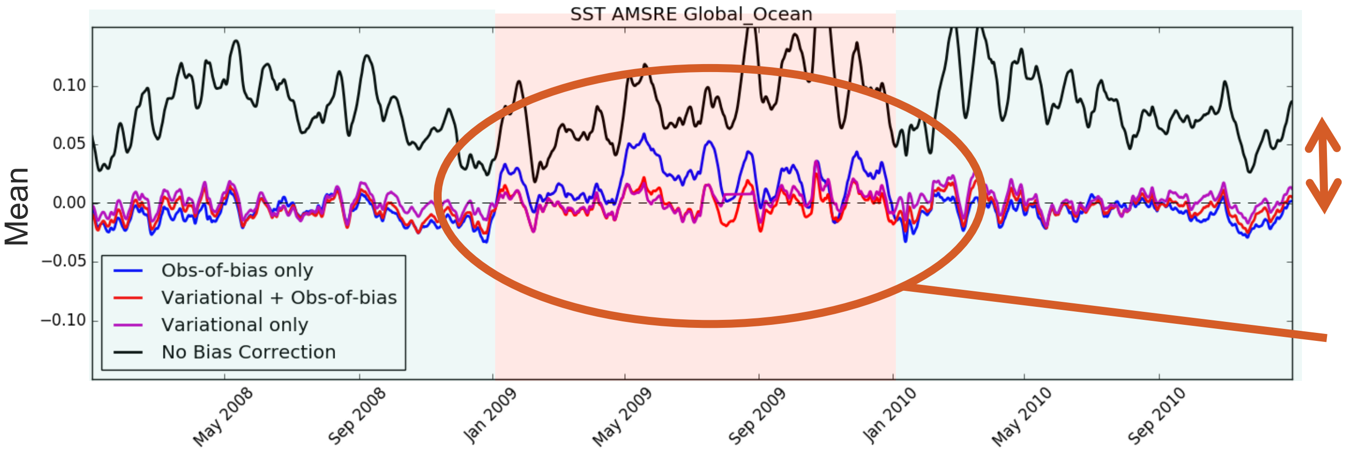
Main scientific achievements

D2.1 SST assimilation: bias correction [METO]

Aim: To develop a bias correction scheme for SST data that will give consistent results across the entire observing period.

Achievements:

- A variational bias correction scheme that uses observations-of-bias. Code implemented in the NEMOVAR system and available from the central Git repository at ECMWF.
- The scheme has been tested using a simplified model and using a full ocean model tested over a three year period (2008-2010)



↕ The overall bias is much reduced

In the period with fewer reference observations, the old MO system (blue line) does not do as well

🟦 AATSR Data used as reference

🟥 AATSR Data not used as reference,

The plots show the difference between AMSRE data and a 1 day forecast of the model.

D2.1: EOF data assimilation [METO]

Aim: To make better use of sparse historical data by using large-scale error covariances defined by EOFs

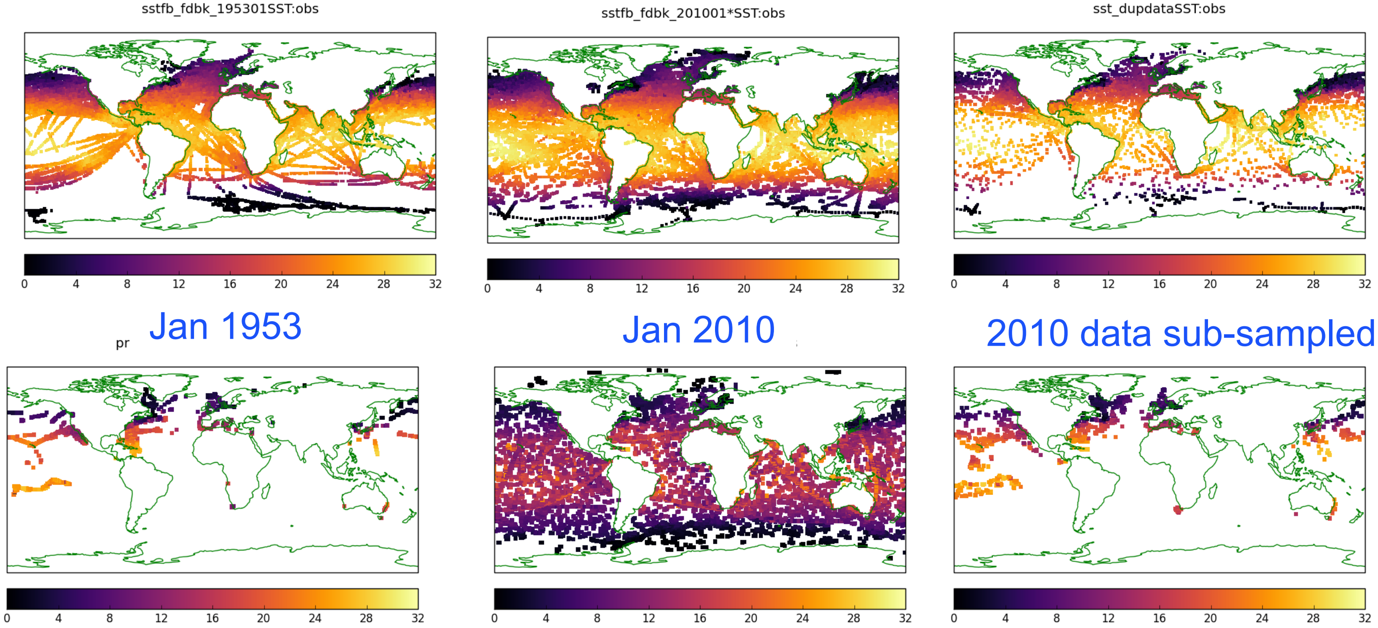
Achievements:

- Developed a method for EOF data assimilation in NEMOVAR and the code has been made available in the central Git repository at ECMWF.
- Tested EOF method in observing system experiments (withholding some of the present day obs to mimic an historic period).

In-situ SST

Data from HadIOD

Profile T

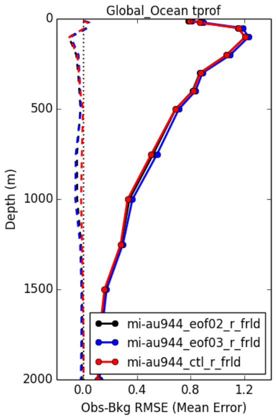


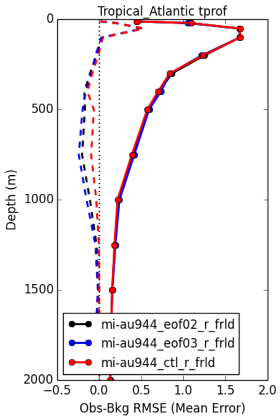
D2.1 EOF data assimilation [METO]

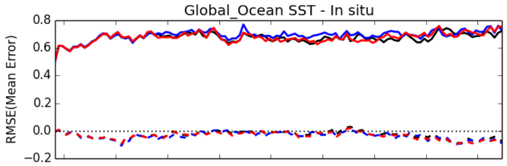
- In monthly objective analysis tests (just based on observaitons, no model) the EOF assimilation improves results compared to the standard DA method used in NEMOVAR.
- Tests in a cycling ocean reanalysis framework show some impact, although model bias appears to reduce the positive impact seen in the previous tests.

Std	Standard FOAM system
EOF02	Hybrid 1% EOF
EOF03	Hybrid 5% EOF

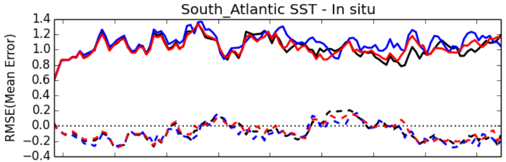
- EOF02 improved SST stats
- Issue with sub-surface T bias
- Not much impact on S







1 Jan 30 April



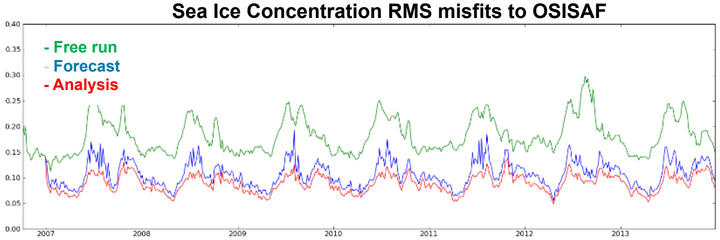
1 Jan 30 April

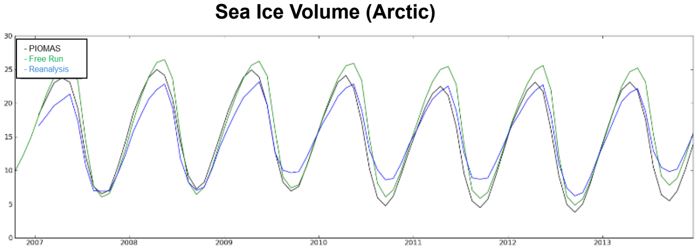
D2.2 Sea-ice data assimilation [MERCO]

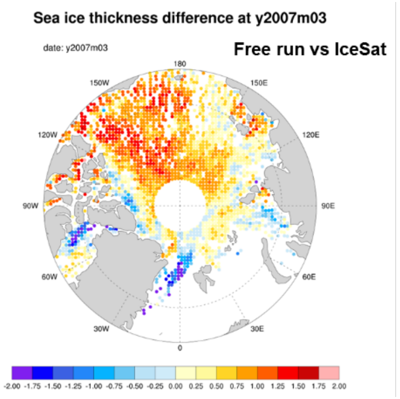
Aim: To improve sea-ice concentration assimilation by investigating multi-variate assimilation to adjust thickness, and by testing anamorphosis transformations to deal better with non-Gaussianity.

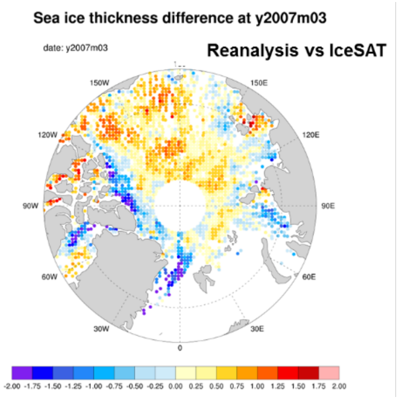
Reference multivariate Sea Ice Reanalysis :

- Regional Arctic configuration : CREG4/NEM3.6/LIM3
- Weakly-coupled DA system (Ice + ocean)
- Sea Ice Model update from analysis : [Concentration, Concentr. and Vol. per category]


Sea Ice Concentration RMS misfits to OSISAF


Sea Ice Volume (Arctic)


Sea ice thickness difference at y2007m03


Sea ice thickness difference at y2007m03

D2.2 Sea-ice data assimilation [MERCO]

Impact of the Gaussian anamorphic transformation on the sea-ice data assimilation problem.

Idealized exercises :

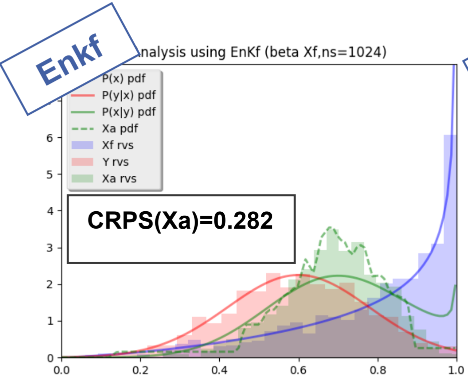
- modelling uncertainties in the sea-ice concentration by a Beta distributed Prior as background error
- Analysis using Enkf with or without Anamorphosis

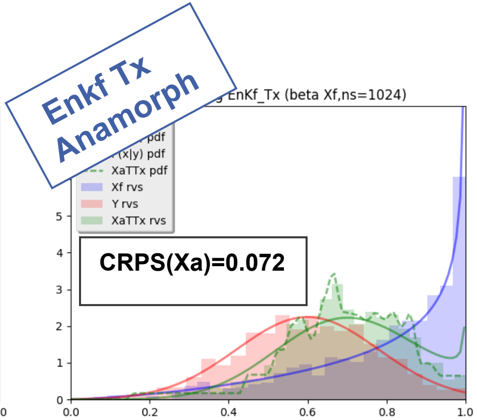
- **How does the EnKF solution compares to the Bayesian solution?**
- **Can Gaussian anamorphosis help to partially restore bayesianity?**

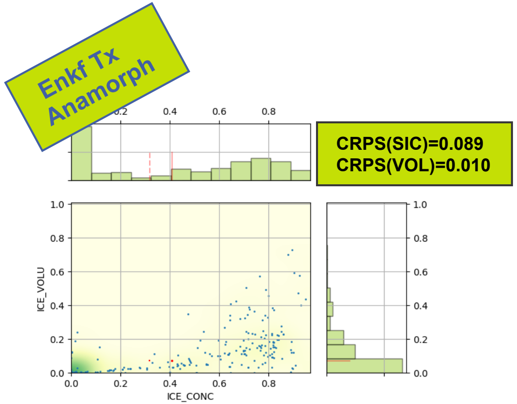
Realistic test case :

- CREG4/NEMO3.6/LIM3 model
- 256 members
- 1D multivariate (SIC,VOL) sea ice analyses
- SIC observation SIC provided by a nature run

We compare the joint (SIC,VOL) posterior distribution from different assimilation scheme (EnKf with or without Anamorphosis , a Particle filter taking as a reference solution)







EnKf using Anamorphosis improves the posterior distribution => closer to the Bayesian solution than the EnKf

EnKf using Anamorphosis improves the posterior distribution ⇒ closer to the Pf solution than the EnKf

D2.3: Using ensemble-estimated background error variances and correlation scales in NEMOVAR [CERFACS]

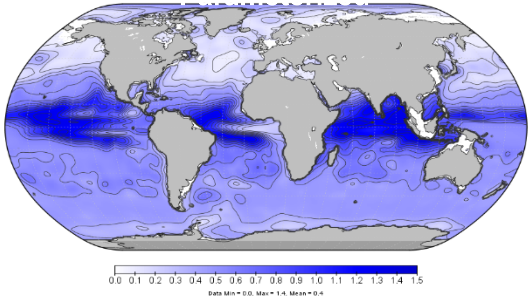
Two methods have been developed to use ensemble perturbations to define the background error covariance matrix (**B**).

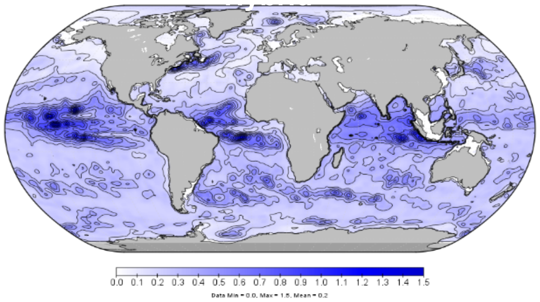
1. Estimate parameters (variances and correlation length scales) of the covariance model.
2. Define a localized, low-rank sample estimate of the covariance matrix.

Hybrid formulations of both 1 and 2 have also been developed in which the ensemble component is linearly combined with a parameterized component.

1 and 2 include optimally-based algorithms for filtering parameters and estimating hybridization weights and localization scales.

Example of parameterized and hybrid **temperature error standard deviations** at 100m, estimated from the ECMWF 11-member ensemble of ocean reanalyses.





D2.3: Using ensemble-estimated background error variances and correlation scales in NEMOVAR [CERFACS]

The correlation operator, localization operator and parameter filter are based on an algorithm that involves solving an implicitly formulated diffusion equation. The diffusion model has been completely revised to make it more general, to eliminate numerical artefacts near complex boundaries, and to improve computational efficiency and scalability on high-performance computers.

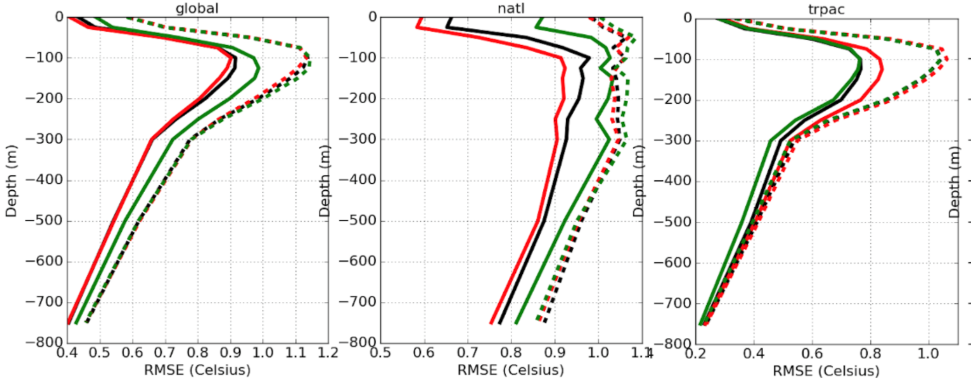
Details have been documented in an article for the Quarterly Journal (Weaver *et al.,* 2016) and in an ECMWF technical memorandum (Weaver *et al.,* 2017).

All methods have been integrated into a new version of NEMOVAR (v5) that is available in the central code repository at ECMWF.

The operational scripts have been adapted to run NEMOVAR v5 in an Ensemble of Data Assimilations (EDA) framework.

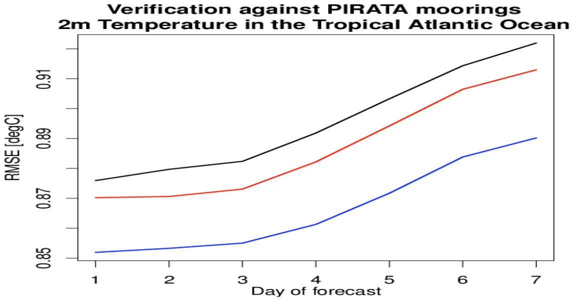
Preliminary experiments testing ensemble and hybrid (parameterized + ensemble) variances show positive results compared to parameterized-alone variances.

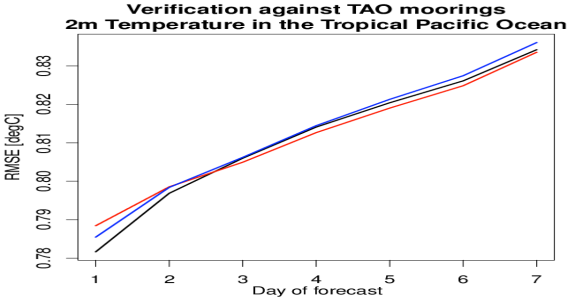
RMS of Obs-Ana (solid curves) and Obs-Bkg (dashed curves) for temperature with **parameterized-only**, **ensemble-only** **and hybrid** variances in a 1/4° global model.

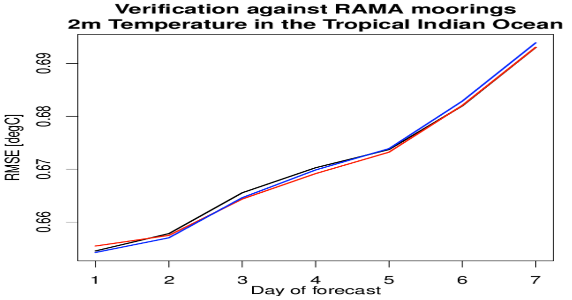


D2.4 A simplified air-sea balance operator [CMCC]

- To couple the sea-surface variables with 2m atmospheric variables, balances might be thought either purely statistical, or purely analytical, or mixed (balanced + unbalanced components)
- We introduce a balance operator that maps the increments of SST onto those of (**T_{2m}, Q_{2m}**) and uses tangent-linear version of CORE bulk formulas (Large & Yeager, 2007)
- Results were compared to ensemble estimates of the air-sea relationships


Verification against PIRATA moorings
2m Temperature in the Tropical Atlantic Ocean


Verification against TAO moorings
2m Temperature in the Tropical Pacific Ocean


Verification against RAMA moorings
2m Temperature in the Tropical Indian Ocean

Weakly Coupled

Strongly Coupled (air-sea balance)

Strongly coupled (statistics)

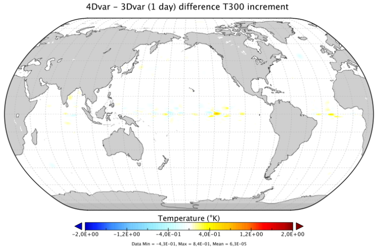
- **Negligible Impact in the Extra-Tropics (probably due to the themodynamical coupling and not dynamical)**
- **Persistent impact through the forecast length in the Atlantic. In other basins emerges later.**
- **Positive everywhere, although significant improvements only in the Atlantic Ocean**

D2.5: 3D-Var or 4D-Var for the ocean component [INRIA]

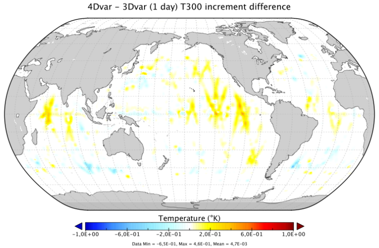
Is it worth replacing 3DVar-FGAT by 4Dvar in CERA's ocean component?

- For ORCA1, assimilating onto T and S profiler data… **no!** One can barely notice the difference. (for CERA-20C)
- Both assimilating SSH and increasing resolution induce a noticeable impact switching to 4DVar, so… yes!
(for CERA-SAT)
- However, at high resolution the cost of the ocean analysis becomes dominant and increasing further would limit the achievable length of CERA-SAT.

Orca1 T/S only



Orca025 T/S+SSH



Two options were tested in order to reduce the cost (both made available in the NEMOVAR repository):

- Lower resolution in the inner loop (not trivial transfer operators due to complex geometry)
- Drastic simplification of the inner model equations
- In ¼ degree model, multi-incremental 4D-Var can be made as quick as 3D-Var.

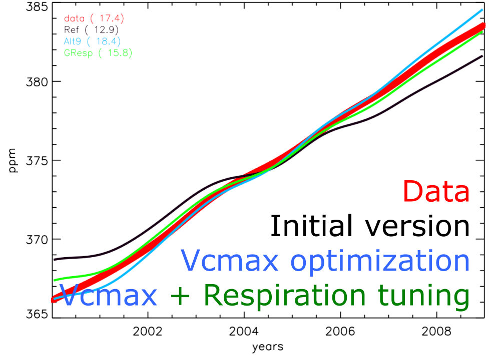
D2.6: optimized terrestrial model parameters and carbon fluxes for the 20th century [UVSQ/LSCE]

LSCE produced an updated variational data assimilation system to optimize ORCHIDEE model parameters.

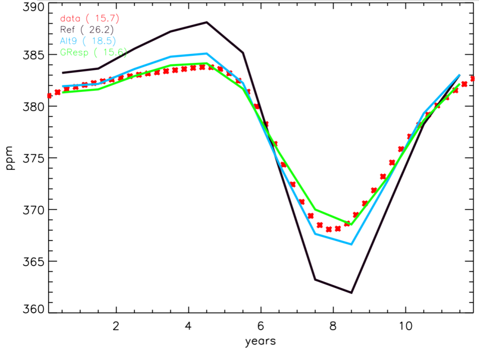
Main achievements:

- ➢ Update of the model with most recent version (CMIP6)
- ➢ Chain to get automatically the Tangent Linear model
- ➢ Assessment of the benefit of different optimization strategy (genetic algorithm vs gradient method)
- ➢ Evaluation of simultaneous vs stepwise optimization
- ➢ Evaluation of the optimization performances
- ➢ Assimilation of new data streams (Vegetation fluorescence, atm. [COS])

Trend (SPO)



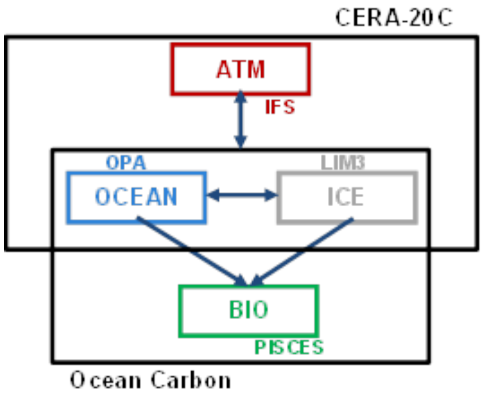
Mean seas. Cycle (ALT)

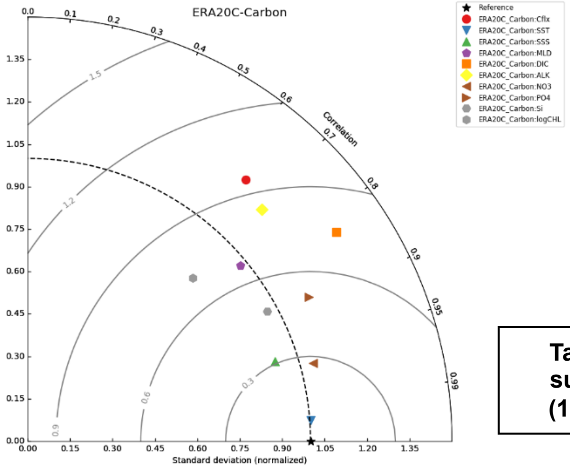


D2.7: Alternatives for coupling ocean biogeochemistry [MERCO]

Development of the configuration of CERA-20C/ ocean carbon

- Many sensitivity tests to single out the best initial condition, NEMO version, parameter settings
- Choice of the coupling strategy with the coupled ocean – atmosphere reanalysis CERA-20C
- Run of a first 20th century experiment ERA-20C/ocean carbon forced by the output of previous ERA-CLIM project ERA-20C
- Assessment of this long experiment (main biogeochemical variables and the carbon flux)





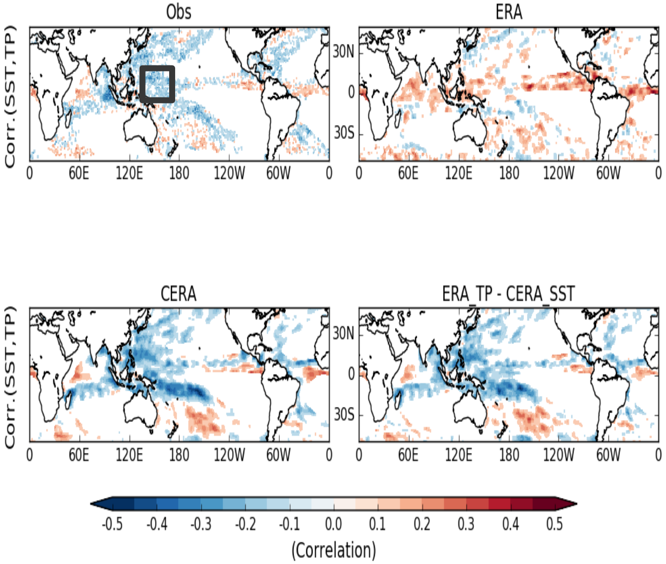
Data sources:
- GLODAPv2 for DIC, ALK
- Landschützer for Cflx = CO_2 flux
- Globcolour for logCHL
- WOA 2013 for NO_3, O_2, PO_4, Si, SST, SSS
- De Boyer-Montégut for MLD

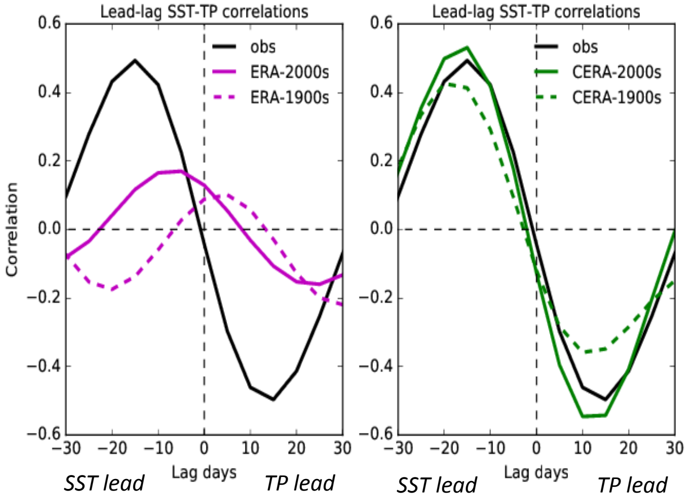
Taylor diagram computed with surface monthly climatologies (1998 – 2009 years if available)

D2.8: Strengths of weakly coupled assimilation methods [UREAD]

- SST-total precipitation (TP) intra-seasonal relationships are better represented in CERA-20C than in ERA-20C, mainly due to coupled modelling.

- Lead-lag plots demonstrate both importance of coupled model when there are few observations (green dashed line vs purple dashed line) and the assimilation of oocean/atmosphere observations (green solid line vs dashed line)

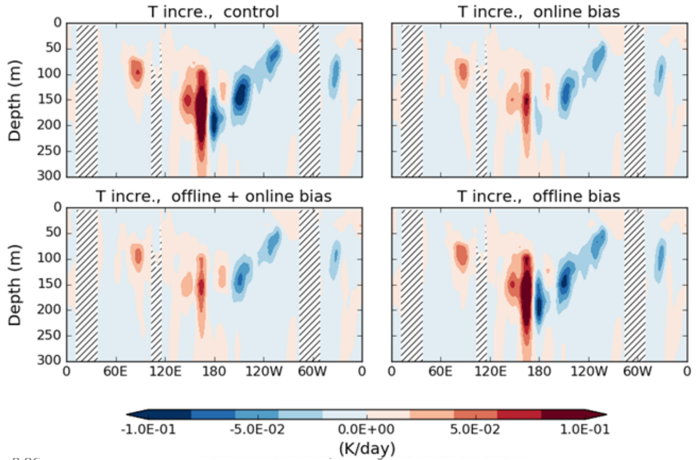


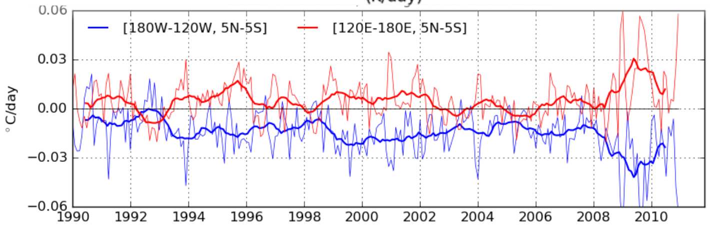


Lead-lag correlation in the black box (left)

D2.10: Coupled-model drift and bias correction [UREAD]

- Large ocean bias increments are diagnosed in CERA-20C, especially in the tropics. No ocean bias correction in CERA-20C. Strong temporal variations, indicating the 'offline' bias correction may not represent such features of ocean bias
- Tested online and offline bias correction in recent year in re-runs of CERA-20C for 2009.

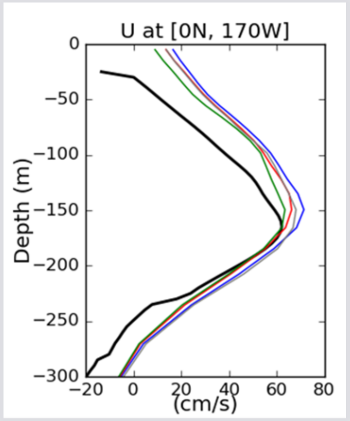


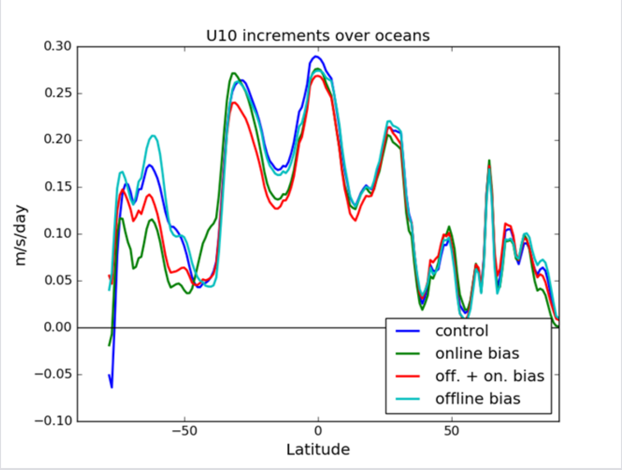


Spatially averaged T increments

- T increments are reduced, with 'online' correction having the largest impact. Ocean analyses are improved as well, e.g. vertical and horizontal velocities in the tropics.
- Impact on atmosphere – reduces wind increments in the Tropics.

Ocean velocity at TAO mooring





D2.11 : Report on fully coupled data assimilation in simplified systems [INRIA]

- Common tractable coupling algorithms lead to flux inconsistency (asynchronicity), and can be damaging to the system behaviour.
- Can we improve the ocean-atmosphere flux consistency through data assimilation?

Developments:

- A stand alone single column ocean-atmosphere model was developed and interfaced with OOPS.
- A collection of 4DVar cost functions were proposed, penalising the flux consistency and/or controlling the ocean/atmosphere interface conditions.
- Convergence of minimisation of said algorithms (including CERA) was studied.

Outcome:

- Flux consistency can indeed be improved (moderately at a small additional cost or significantly at a huge additional cost).
- Global (outer) convergence can also be improved compared to CERA, so more benefit can be expected from the first outer iterations

Summary

Summary

- WP2 has delivered many developments which could be included in future reanalyses.
- ***Ocean data assimilation*** developments have been incorporated into a new version of the NEMOVAR code (hosted at ECMWF):
  - SST bias correction; EOF error covariances; hybrid ensemble-variational DA; 4DVar.
  - It is now much closer in terms of complexity to the atmospheric DA used in CERA.
- ***Coupled data assimilation*** research has led to some useful ideas for improving future versions of CERA:
  - Improved understanding of methods to increase the coupling in the DA either through linearised air-sea balance or methods used to improve coupling in models.
  - Improved understanding of the ocean bias correction in coupled system.
- Improvements have been made to the ***carbon component*** (ocean and land) of the reanalysis.
- Funding for implementing these developments into the next reanalyses, and continued research, is required.

Thankyou for listening

A simplified air-sea balance operator

To couple the sea-surface variables with 2m atmospheric variables, balances might be thought either purely statistical, or purely analytical, or mixed (balanced + unbalanced components)
We introduce a balance operator that maps the increments of SST onto those of ($\mathbf{T}_{2m}$, $\mathbf{Q}_{2m}$) and uses tangent-linear version of CORE bulk formulas (Large & Yeager, 2007)

- $\delta \mathbf{T}_{2m} = \Delta t\ \ [\delta\, \mathbf{Q}_{LW}\, (\delta \mathbf{SST}) + \delta\, \mathbf{Q}_{SEN}\, (\delta \mathbf{SST})] / [\rho_A\, cp_A\, \mathbf{H}_{ABL}]$
  (no condensation in ABL)

- $\delta \mathbf{q}_{2m} = \Delta t\ \ [\delta\, \mathbf{E}\, (\delta \mathbf{SST})] / [\rho_A\, \mathbf{H}_{ABL}]$

TL model of air-sea thermodynamics

Where the transfer coefficients (**Ce, Ch** for Evaporation and Sensible heat, respectively) are assumed not to depend on **SST** and taken from the fully non-linear model. *(Might be relaxed with simple parametric formulations)*

Physical space (T,S,η,T2m,Q2m)

$$\delta \mathbf{x} = [\ \mathbf{V}_A\ \mathbf{V}_\eta\ \mathbf{V}_H\ \mathbf{V}_V\ ]\ \mathbf{v}$$

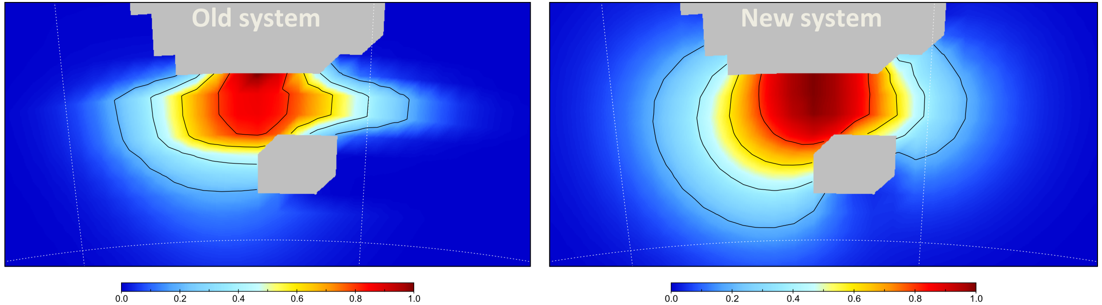
Control Variable

Air-Sea Balance Operator

D2.3: Using ensemble-estimated background error variances and correlation scales in NEMOVAR

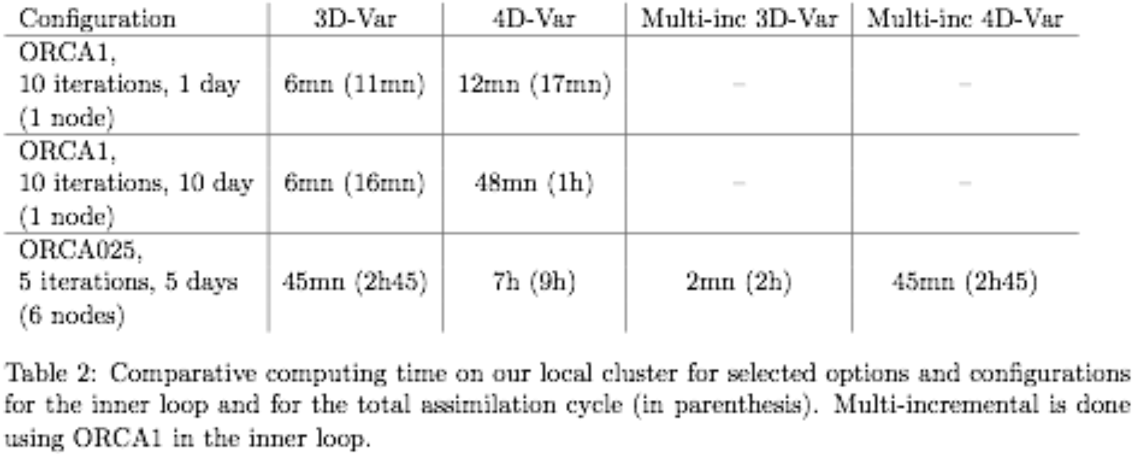
- The correlation operator, localization operator and parameter filter are based on an algorithm that involves solving an implicitly formulated diffusion equation.
- The diffusion model has been completely revised to make it more general, to eliminate numerical artefacts near complex boundaries, and to improve computational efficiency and scalability on high-performance computers.
- Details have been documented in an article for the Quarterly Journal (Weaver *et al.,* 2016) and in an ECMWF technical memorandum (Weaver *et al.,* 2017).

Example of improved representation of 2D correlations near complex boundaries.



D2.5: 3D-Var or 4D-Var for the ocean component

- Two options were tested in order to reduce the cost (both made available in the NEMOVAR git repository):
  - Lower resolution in the inner loop (not trivial transfer operators due to complex geometry)
  - Drastic simplification of the inner model equations
  - In ¼ degree model, multi-incremental 4D-Var can be made as quick as 3D-Var.



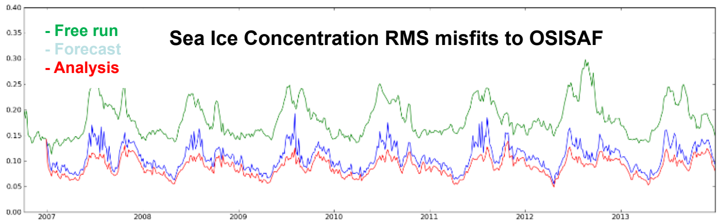
Configuration	3D-Var	4D-Var	Multi-inc 3D-Var	Multi-inc 4D-Var
ORCA1, 10 iterations, 1 day (1 node)	6mn (11mn)	12mn (17mn)	–	–
ORCA1, 10 iterations, 10 day (1 node)	6mn (16mn)	48mn (1h)	–	–
ORCA025, 5 iterations, 5 days (6 nodes)	45mn (2h45)	7h (9h)	2mn (2h)	45mn (2h45)

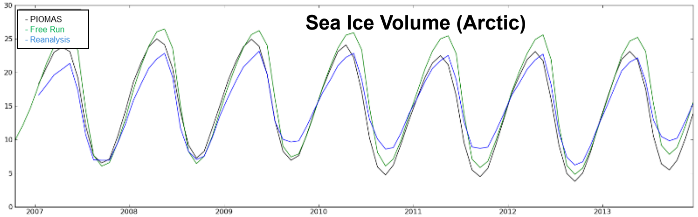
Table 2: Comparative computing time on our local cluster for selected options and configurations for the inner loop and for the total assimilation cycle (in parenthesis). Multi-incremental is done using ORCA1 in the inner loop.

Production of a Multivariate Sea Ice Reanalysis
Regional Arctic reanalysis assimilating OSI-SAF SIC Observations

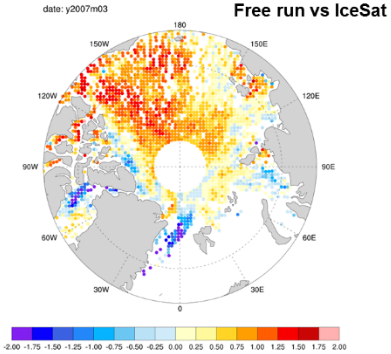
Reference multivariate Sea Ice Reanalysis :

- Regional Arctic configuration : CREG4/NEM3.6/LIM3
- Weakly-coupled DA system (Ice + ocean)
- Sea Ice Model update from analysis : [Concentration, Concentr. and Vol. per category]

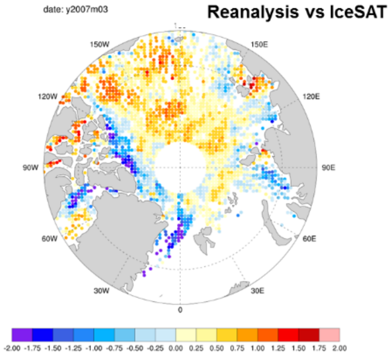

Sea Ice Concentration RMS misfits to OSISAF


Sea Ice Volume (Arctic)

Sea ice thickness difference at y2007m03


Free run vs IceSat

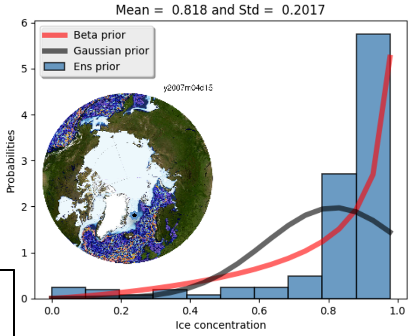
Sea ice thickness difference at y2007m03


Reanalysis vs IceSAT

Impact of the Gaussian anamorphic transformation on the sea-ice data assimilation problem

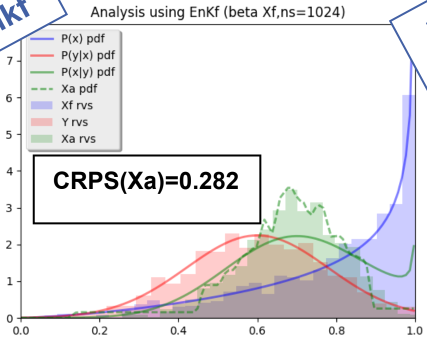
Idealized exercises :

- modelling uncertainties in the sea-ice concentration by a Beta distributed Prior as background error
- Analysis using Enkf with or without Anamorphosis



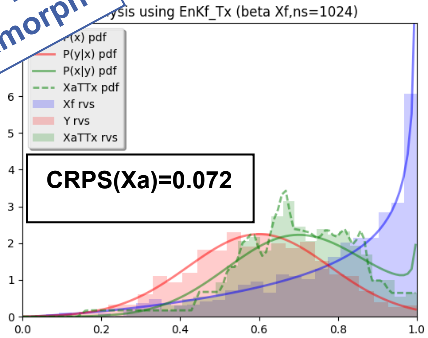
- **How does the EnKF solution compares to the Bayesian solution?**
- **Can Gaussian anamorphosis help to partially restore bayesianity?**

Enkf



CRPS(Xa)=0.282

Enkf Tx Anamorph



CRPS(Xa)=0.072

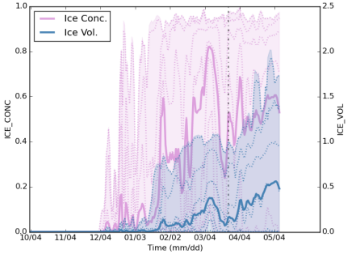
$$CRPS = \sum_{i=1}^{nbins} (cdf^B - cdf^F)^2$$

EnKf using Anamorphosis improves the posterior distribution
=> closer to the Bayesian solution than the EnKf

Impact of the Gaussian anamorphic transformation on the sea-ice data assimilation problem

Realistic test case :

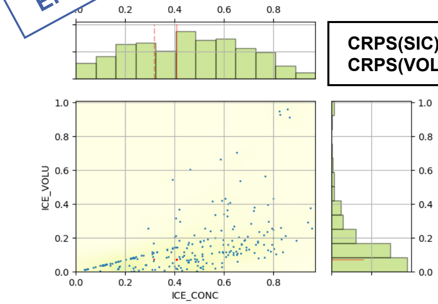
- CREG4/NEMO3.6/LIM3 model
- 256 members
- 1D multivariate (SIC,VOL) sea ice analyses
- SIC observation SIC provided by a nature run



$$CRPS = \sum_{i=1}^{nbins} (cdf^B - cdf^F)^2$$

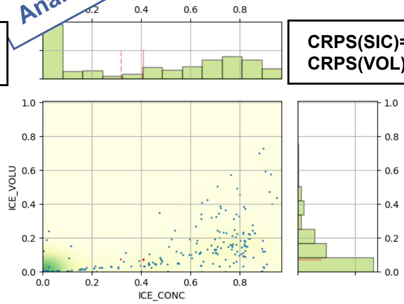
We compare the joint (SIC,VOL) posterior distribution from different assimilation scheme (EnKf with or without Anamorphosis , a Particle filter taking as a reference solution)

Enkf



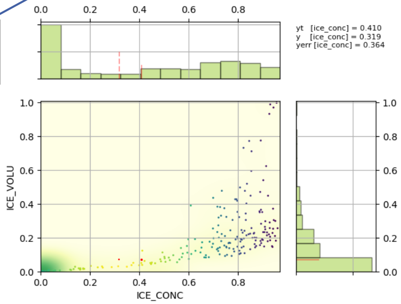
CRPS(SIC)=3.232
CRPS(VOL)=0.346

Enkf Tx Anamorph



CRPS(SIC)=0.089
CRPS(VOL)=0.010

Pf



EnKf using Anamorphosis improves the posterior distribution
$\Rightarrow$ closer to the Pf solution than the EnKf

CMCC activities

- **Strongly Coupled DA experiments**
  - Configuration: intermediate complexity experiments
  - Idealized results (single-obs tests)
  - Real-world results

- **Other activities relevant to ERA-CLIM2**
  - Constraining the global ocean heat budget through CERES data
  - Sampling-aware verification methods for reanalyses
  - Sensitivity of GOHC in reanalyses to atmospheric forcing and other datasets
  - Comparing advanced DA methods

Deliverable already sent in Jan 2017
Manuscript on “Strongly Coupled DA experiments” in review for MWR

Toward a couple Carbon – Climate reanalysis of the 20th Century

D2.6 : Report on optimized terrestrial model parameters and carbon fluxes for the 20th century, including requirements for coupling land carbon biogeochemistry in future Earth system reanalyses

Philippe Peylin, Nicolas Vuichard, Vladislav Bastrikov
Natasha MacBean, Fabienne Maignan, Cedric Bacour, Sauveur
Belviso, Catherine Ottle, & the ORCHIDEE project team

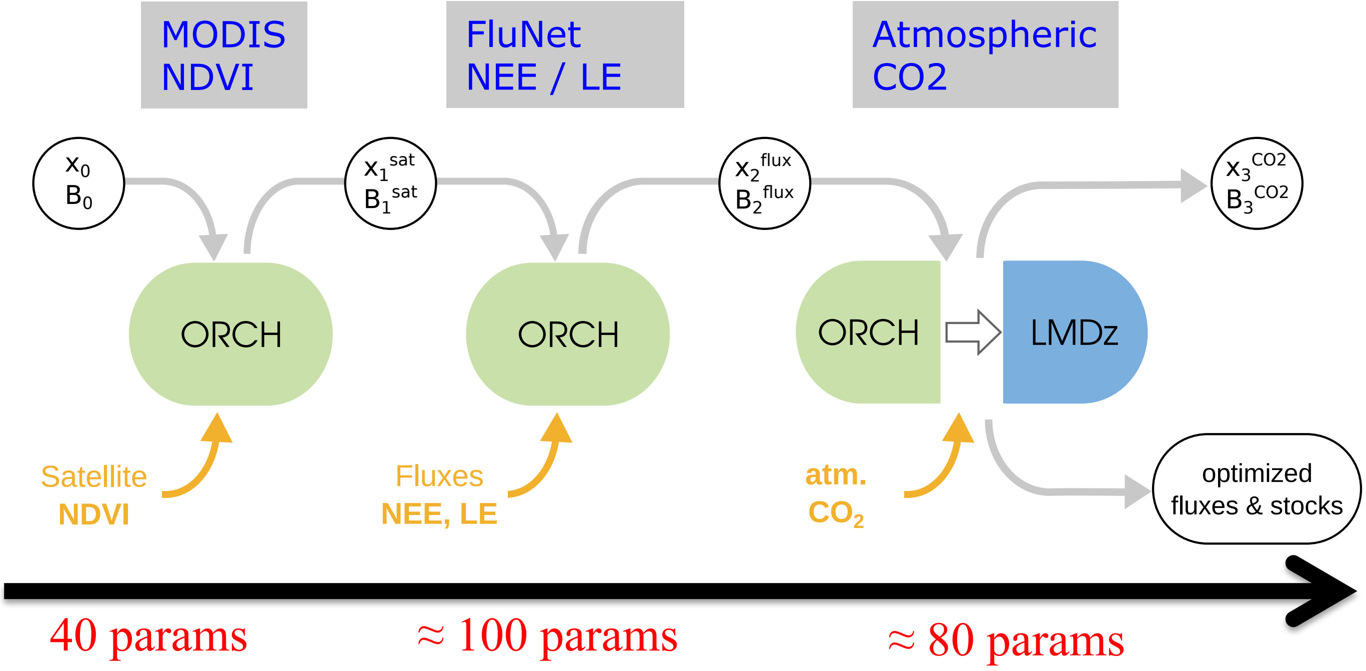
Main achievements

➔ LSCE updated his variation data assimilation system to optimize ORCHIDEE model parameters

- ➢ Update of the model with most recent version (CMIP6)
- ➢ Chain to get automatically the Tangent Linear model
- ➢ Assessment of the benefit of different optimization strategy (genetic algorithm vs gradient method)
- ➢ Evaluation of simultaneous vs stepwise optimization
- ➢ Evaluation of the optimization performances
- ➢ Assimilation of new data streams (Vegetation fluorescence, atm. [COS]

Step wise data assimilation system

$$J(x) = \underbrace{\tfrac{1}{2}(\mathbf{H}.x-y)^T \mathbf{R}^{-1}(\mathbf{H}.x-y)}_{\text{Observation term}} + \underbrace{\tfrac{1}{2}(x-x_b)^T \mathbf{B}^{-1}(x-x_b)}_{\substack{\text{Prior parameter term} \\ \text{(from previous step)}}}$$



MODIS NDVI

FluNet NEE / LE

Atmospheric CO2

x_0 B_0

x_1^{sat} B_1^{sat}

x_2^{flux} B_2^{flux}

x_3^{CO2} B_3^{CO2}

ORCH

ORCH

ORCH

LMDz

Satellite **NDVI**

Fluxes **NEE, LE**

atm. CO_2

optimized fluxes & stocks

40 params

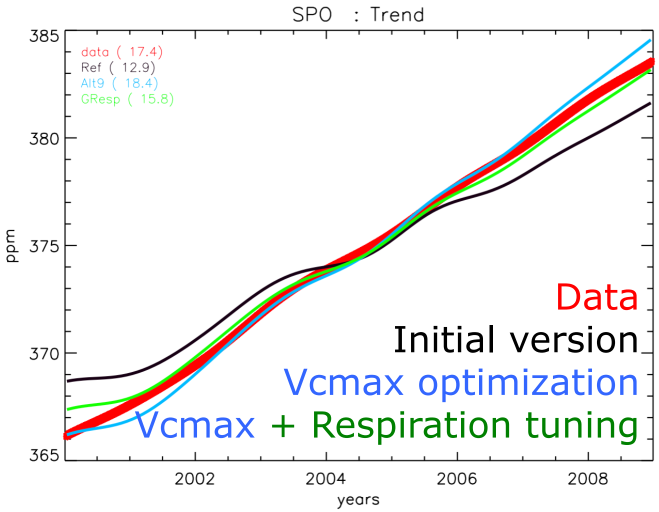
≈ 100 params

≈ 80 params

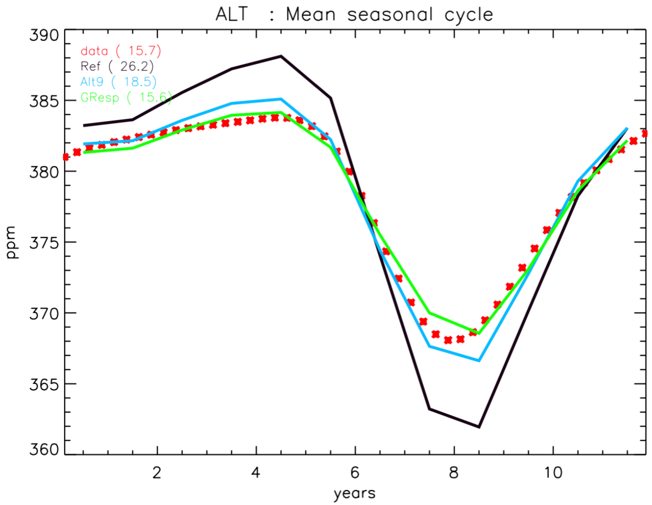
Optimization against atm. [CO2]

- ➔ **Optimization** of key parameters (Vcmax, LAImax,…) using Bayesian optimizations with FluxNet data & atm. [CO2]
- ➔ Evaluation of the trend and seasonal cycle

Trend (SPO)

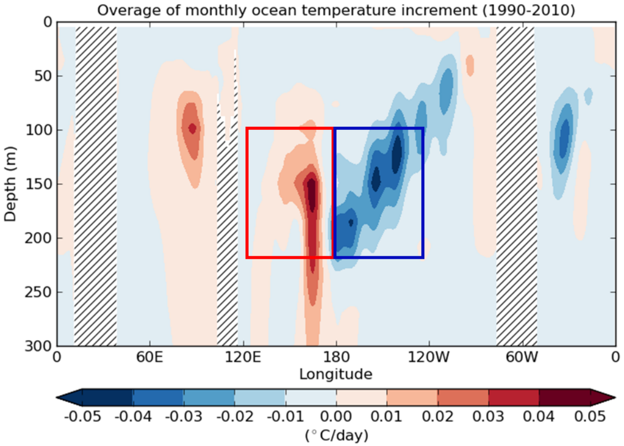


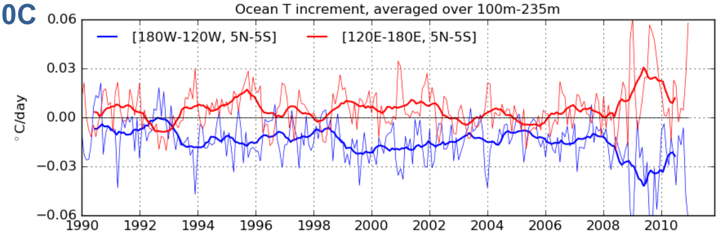
Mean seas. Cycle (ALT)

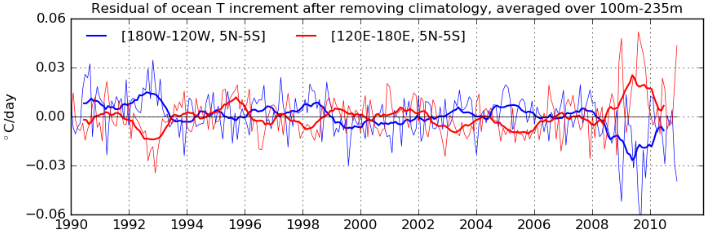


2. Assessment of coupled-model drift and approaches for obtaining consistent ocean and atmospheric bias corrections.

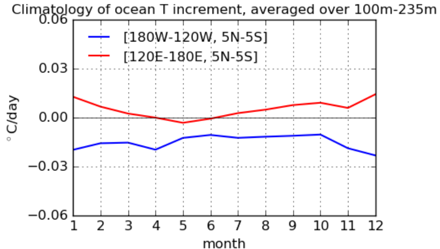
Diagnostics of ocean biases (5S-5N) in CERA-20C







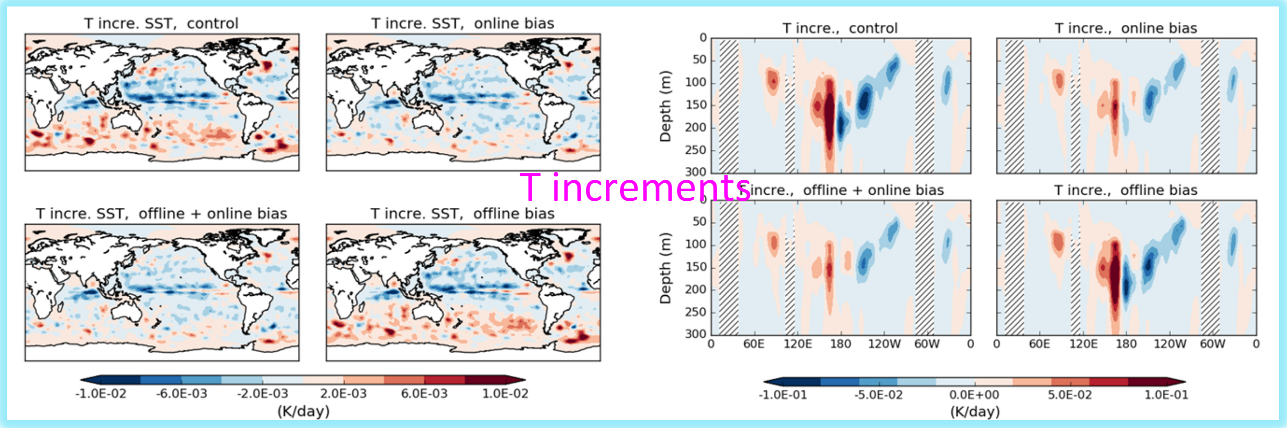
- Large ocean bias increments are diagnosed in CERA-20C, especially in the tropics
- Strong temporal variations, indicating the ‘offline’ bias correction may not represent such features of ocean bias



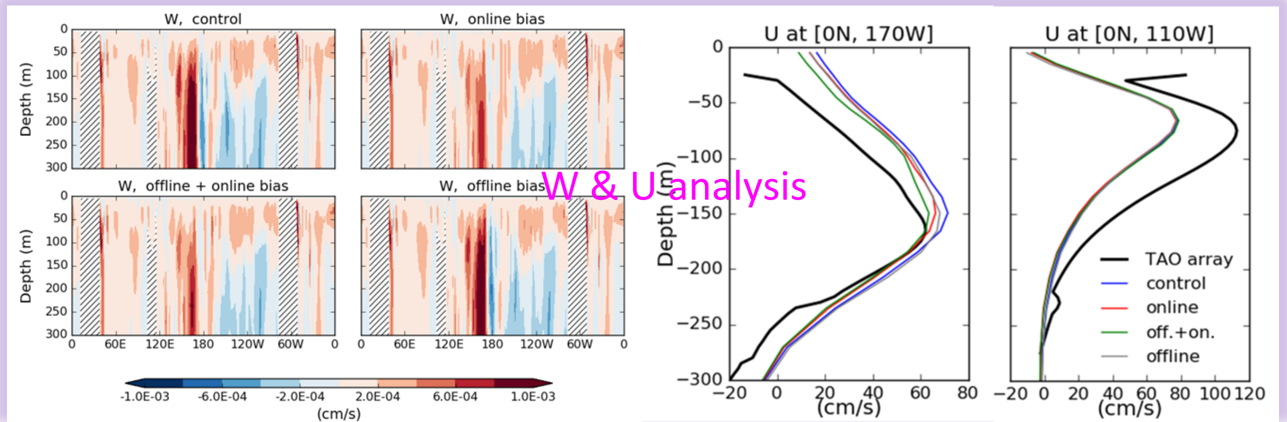
D2.10: Assessment of coupled-model drift and approaches for obtaining consistent ocean and atmospheric bias corrections.

Tests for bias correction in CERA (applying climatology 'offline'-only, 'online'-only and 'off.+onl.' corrections, for one year 2009)

- T increments are reduced, with 'online' correction having the largest impact
- Ocean analyses are improved as well, e.g. W&U in the tropics

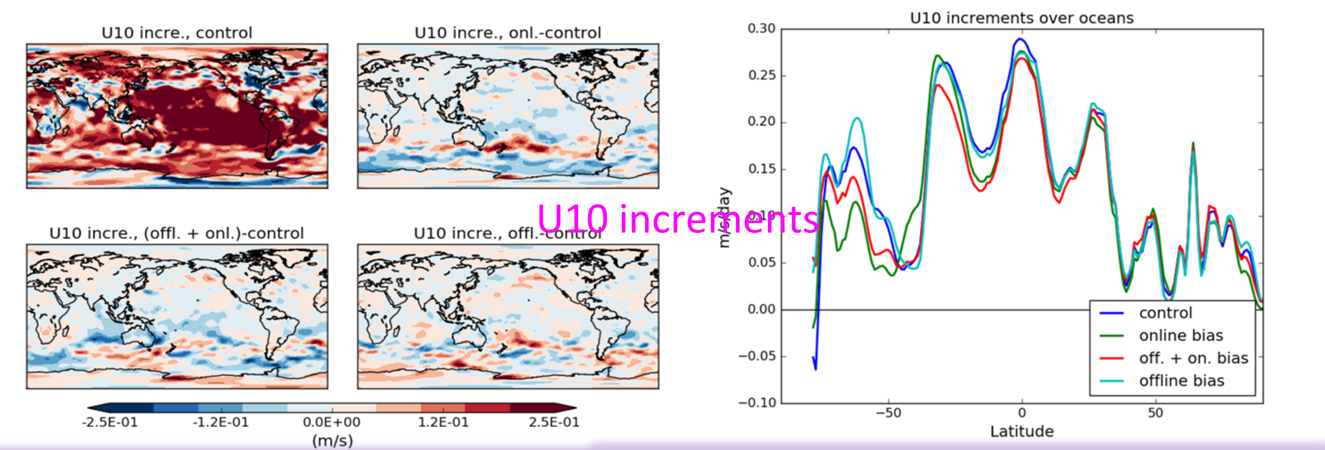


T increments

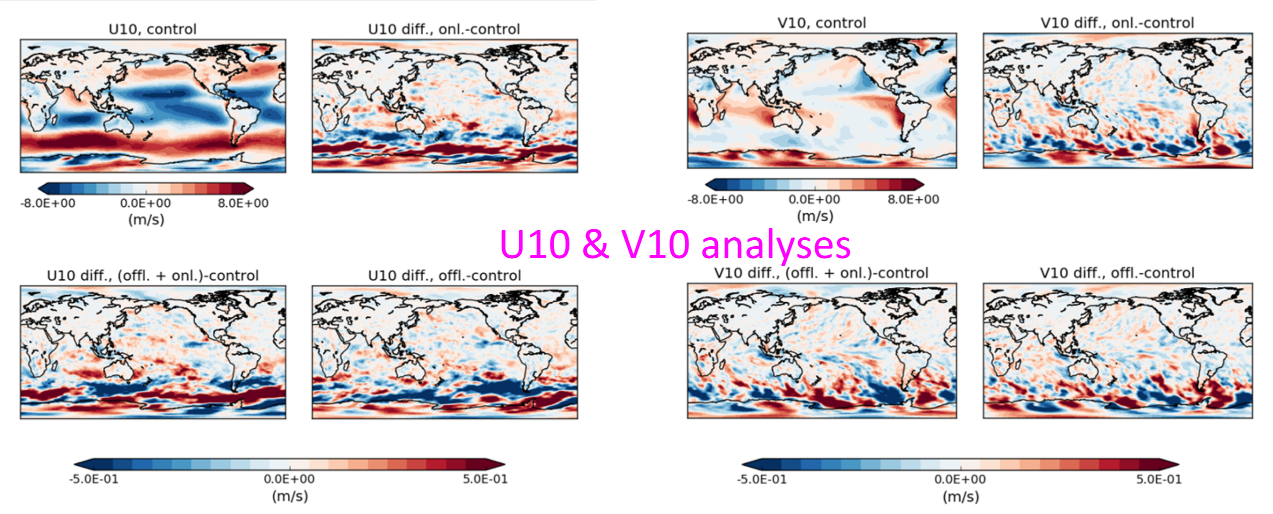


W & U analysis

Ocean bias correction has impacts on the atmosphere analysis



U10 increments



U10 & V10 analyses