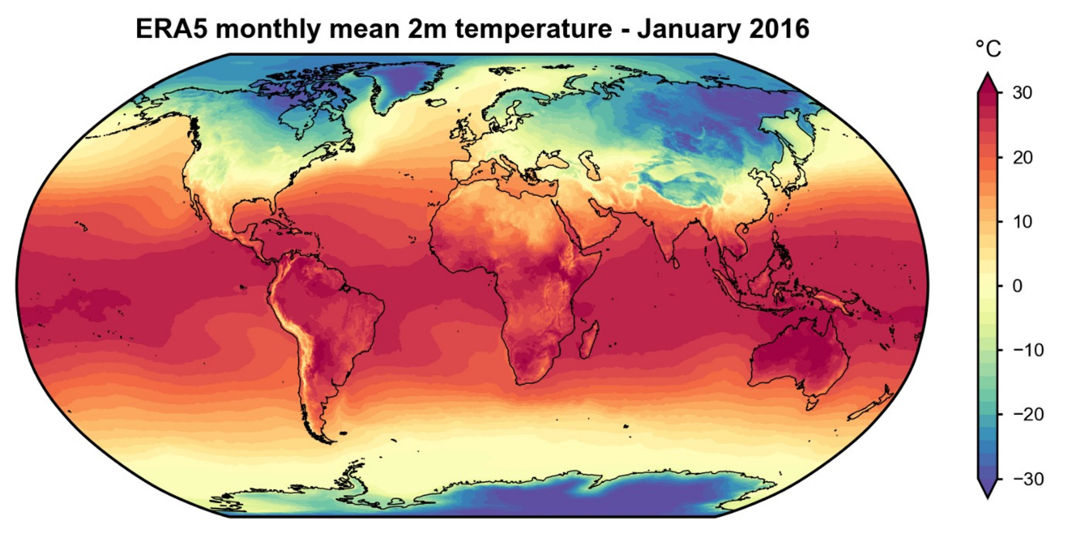
# Improving climate reanalysis with the offline ensemble Kalman smoother

Yiguo Wang[1,2], François Counillon[1,2,3], Yue Ying[1,2] and Sébastien Barthélémy[2,3]

1. Nansen Environmental and Remote Sensing Center, Norway
2. Bjerknes Centre for Climate Research, Norway
3. Geophysical Institute, University of Bergen, Norway

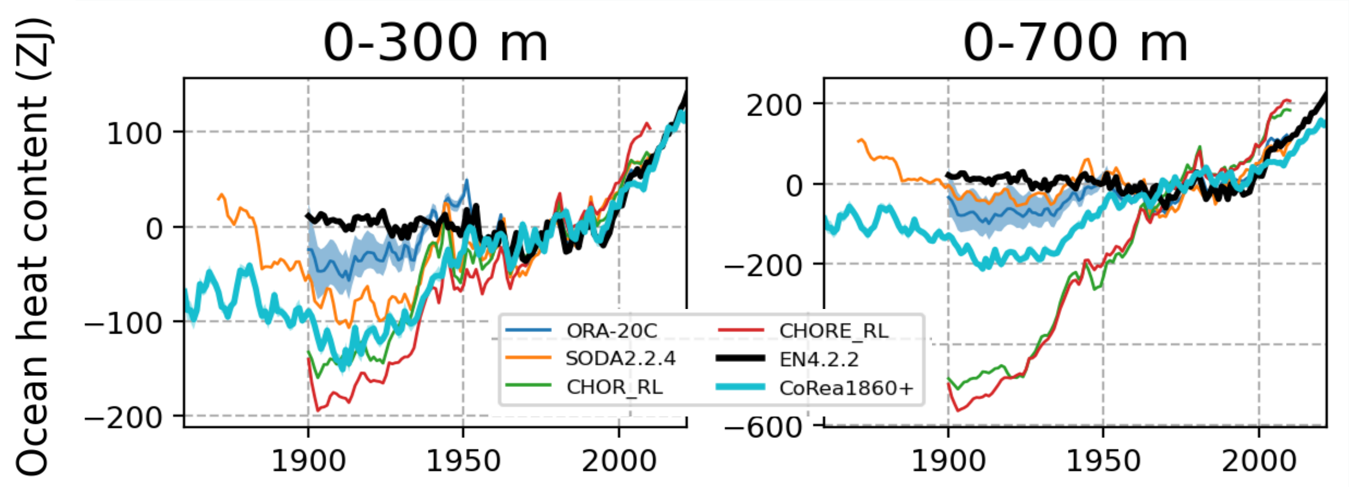
# What’s climate reanalysis?

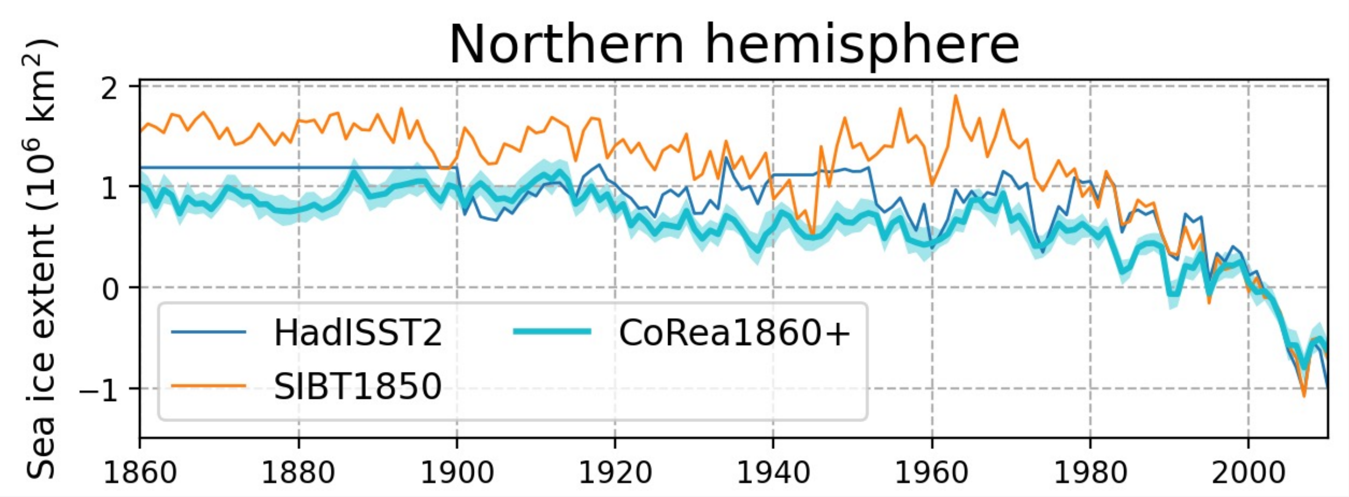
- Dynamically consistent reconstruction of the climate system, atmospheric reanalyses (ERA5, Hersbach et al., 2023), ocean reanalyses (ORAS5, Zuo et al., 2019), coupled reanalyses (Laloyaux et al., 2018, O’Kane et al., 2021)
- Reanalysis = dynamical model + observations + data assimilation
- Understanding anthropologically driven global warming
- Studying climate variability and teleconnections
- Initialising climate predictions



# What do we do in CoRea?

- Developed and tested new DA techniques
- produced a long-coupled reanalysis over 1860-2022




**Improving climate reanalysis with the offline ensemble Kalman smoother**


Yiguo Wang[1], François Counillon[1,2], Yue Ying[1], and Sébastien Barthélémy[2]

[1]Nansen Environmental and Remote Sensing Centre and Bjerknes Centre for Climate Research, Bergen, Norway
[2]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway



ORIGINAL RESEARCH article

Front. Clim., 15 December 2022
Sec. Predictions and Projections
Volume 4 - 2022 | https://doi.org/10.3389/fclim.2022.918572

This article is part of the Research Topic
Recent Advances in Climate Reanalysis
View all 7 Articles >


**Benefit of vertical localization for sea surface temperature assimilation in isopycnal coordinate model**


Yiguo Wang[1*] François Counillon[1,2] Sébastien Barthélémy[2] Alexander Barth[3]

[1] Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, Bergen, Norway
[2] Geophysical Institute and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway
[3] GeoHydrodynamics and Environment Research (GHER), Department of Astrophysics, Geophysics and Oceanography, University of Liège, Liège, Belgium



JAMES | Journal of Advances in Modeling Earth Systems*

Research Article | Open Access


**Adaptive Covariance Hybridization for the Assimilation of SST Observations Within a Coupled Earth System Reanalysis**


Sébastien Barthélémy, François Counillon, Yiguo Wang



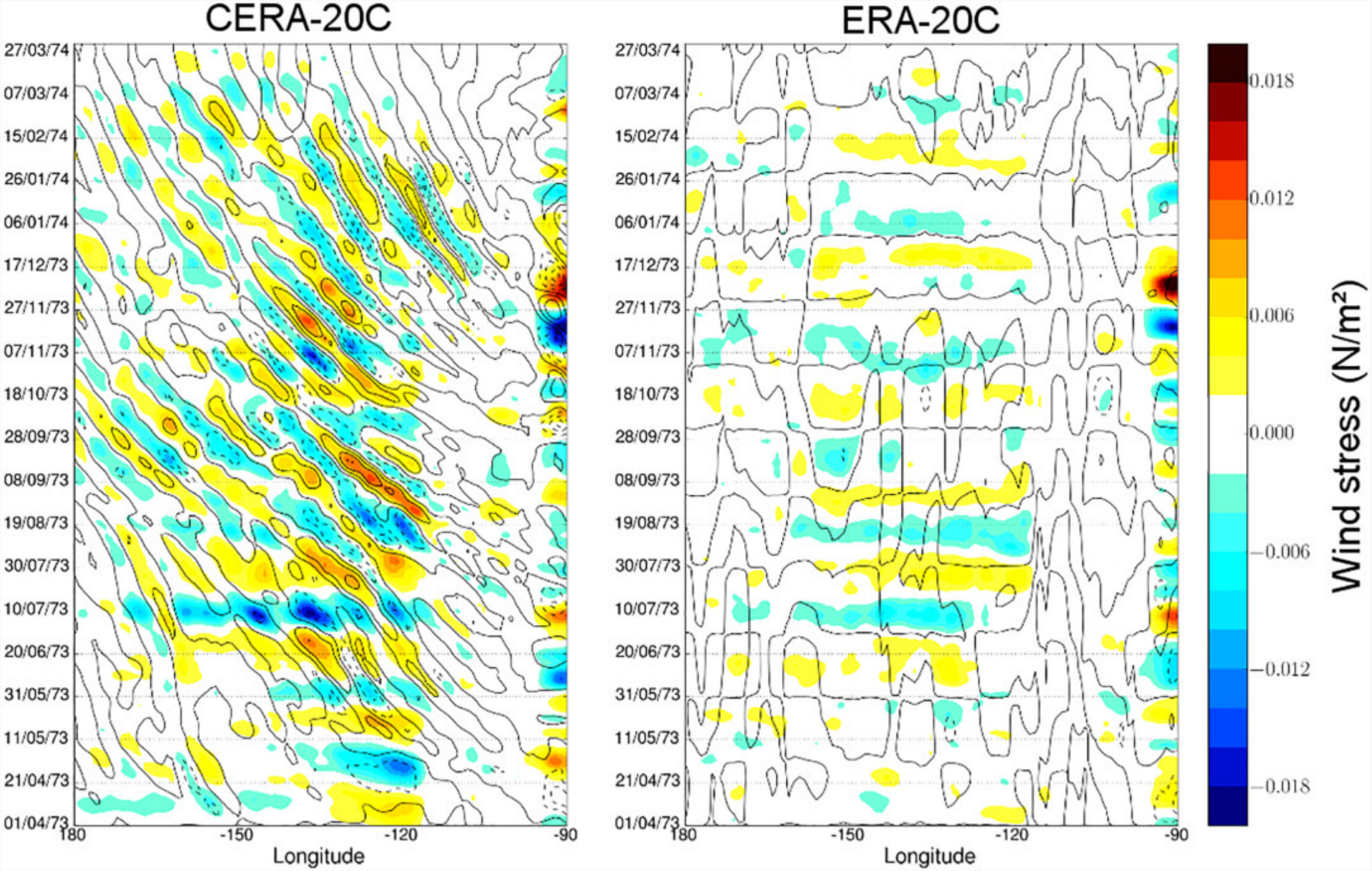
First published: 14 June 2024 | https://doi.org/10.1029/2023MS003888

# Why do we need yet another reanalysis?

NERSC

| Dataset | Component | Description | Category | Reference |
|---|---|---|---|---|
| 20CRv2c | Atmosphere/Land | Produced by NOAA–CIRES. Available globally from 1851 to 2014. | Surface input | Compo et al. (2011) |
| 20CRv3 | Atmosphere/Land | Produced by NOAA–CIRES. Available globally from 1806 to 2015. | Surface input | Slivinski et al. (2021) |
| ERA-20C | Atmosphere/Land | Produced by ECMWF. Available globally from 1900 to 2010. | Surface input | Poli et al. (2016) |
| ORA-20C | Ocean/Sea ice | Produced by ECMWF. Available globally from 1900 to 2009. | Surface and subsurface input | de Boisseson et al., (2018) |
| SODA2.2.4 | Ocean | Produced by U. Maryland. Available globally from 1871 to 2010. | Surface and subsurface input | Giese et al. (2016) |
| CHOR | Ocean | Produced by CMCC. Available globally from 1900 to 2010. | Surface and subsurface input | Yang et al. (2017) |
| EN4.2.2 | Ocean | Produced by Met Office. Available globally from 1900 to the present. | Subsurface input | Good et al. (2013) |
| CERA-20 | Atmosphere/Land/ Ocean/Sea ice | Produced by ECMWF. Available globally from 1901 to 2010. | Surface and subsurface input | Laloyaux et al. (2018) |

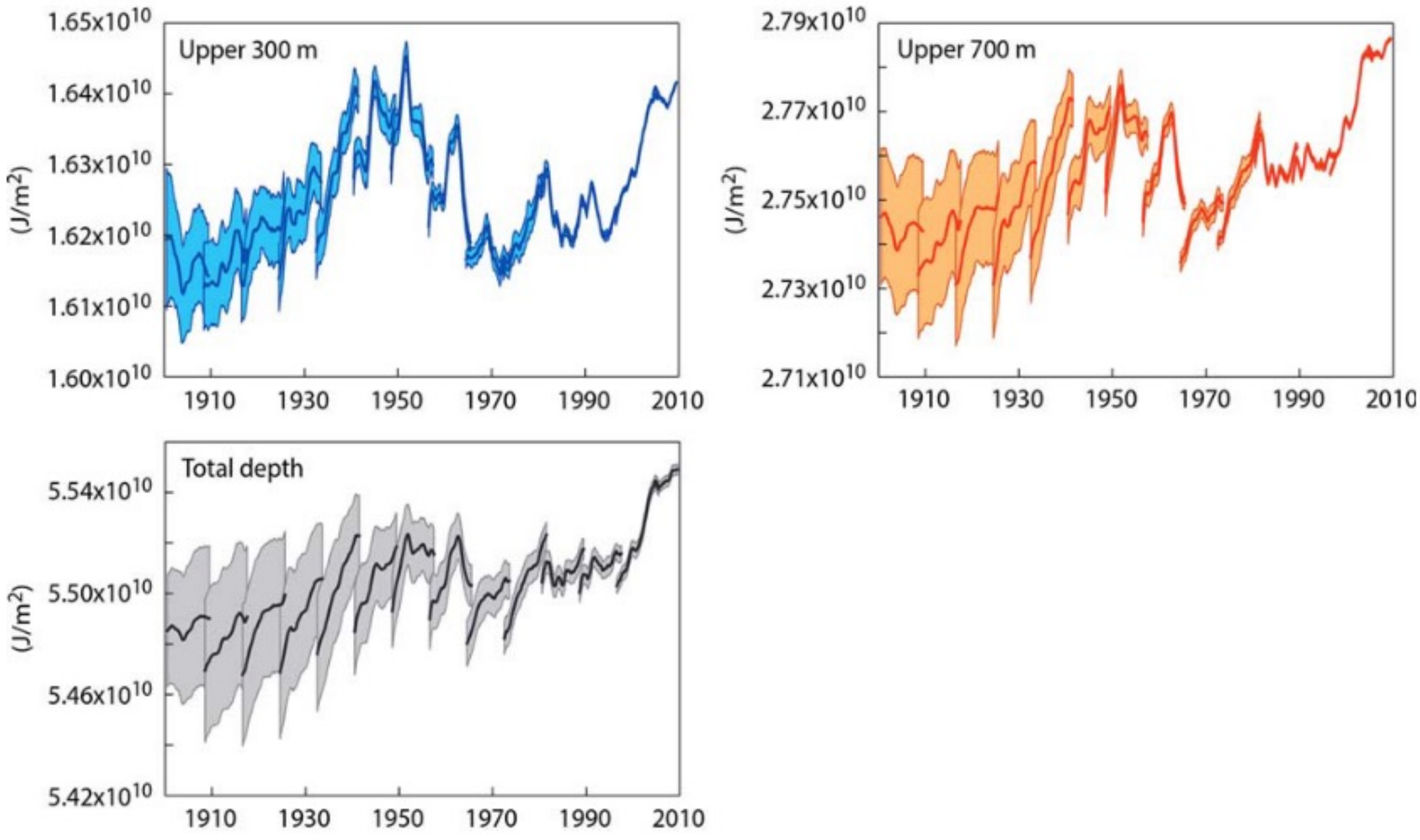
# 1st motivation: coupled processes




Hovmoller time series of spatially high-passed filtered SST (contours) and wind stress (shading) at 1°N in the eastern Pacific from April 1973 to April 1974.

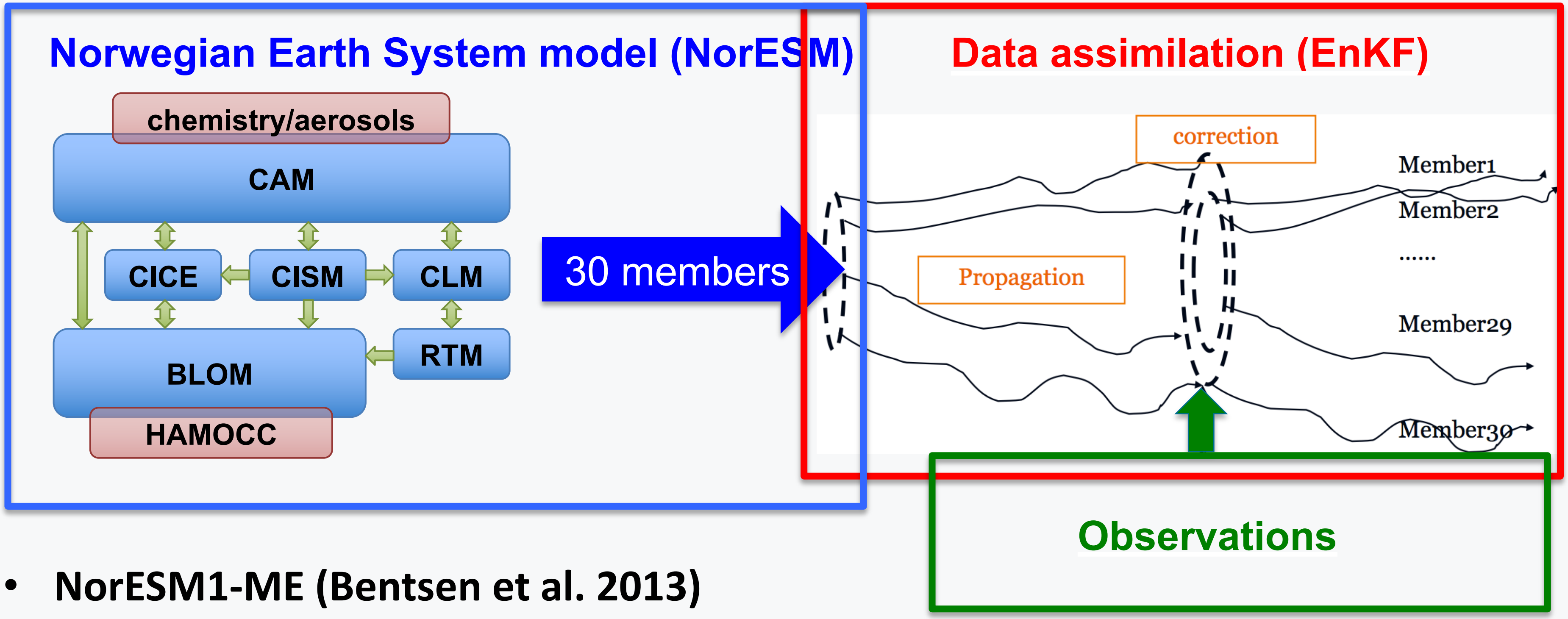
(Laloyaux et al. 2018)

# 2nd motivation: continuous reconstruction




(Laloyaux et al. 2018)

# Norwegian Climate Prediction Model (NorCPM)

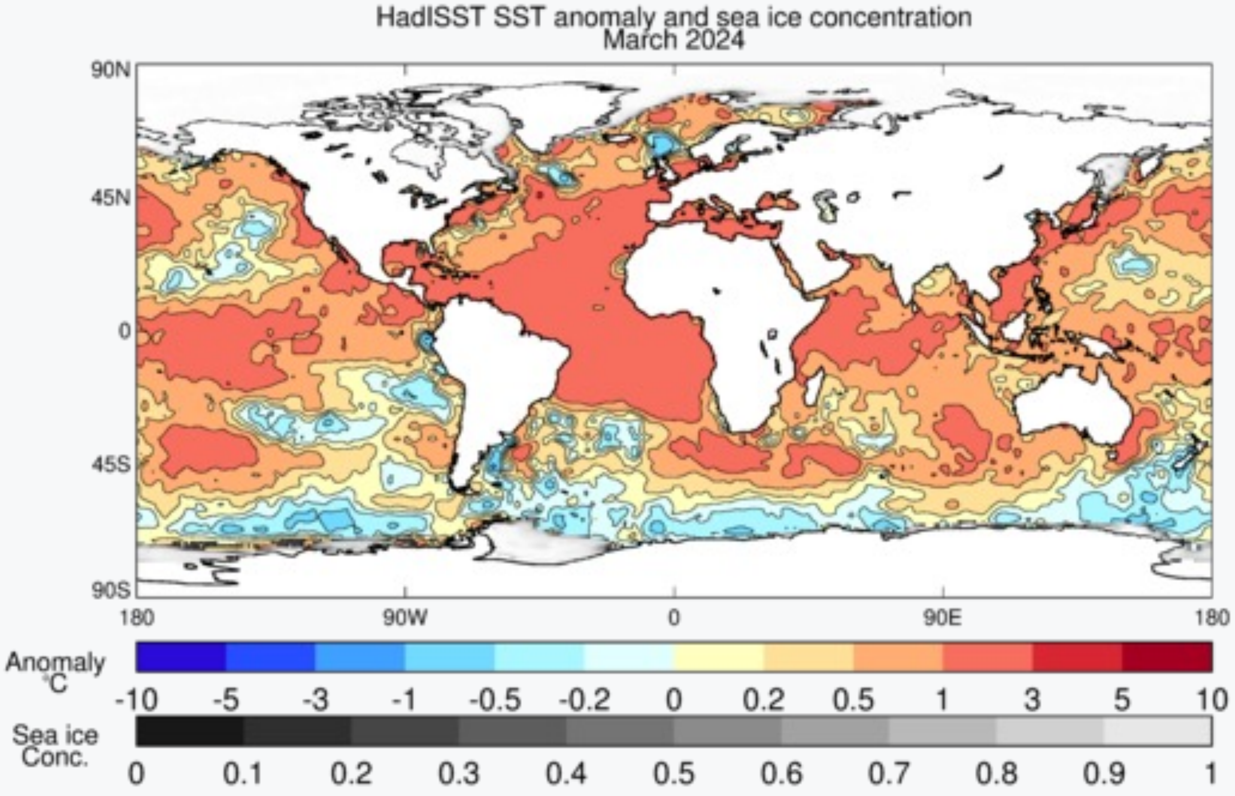


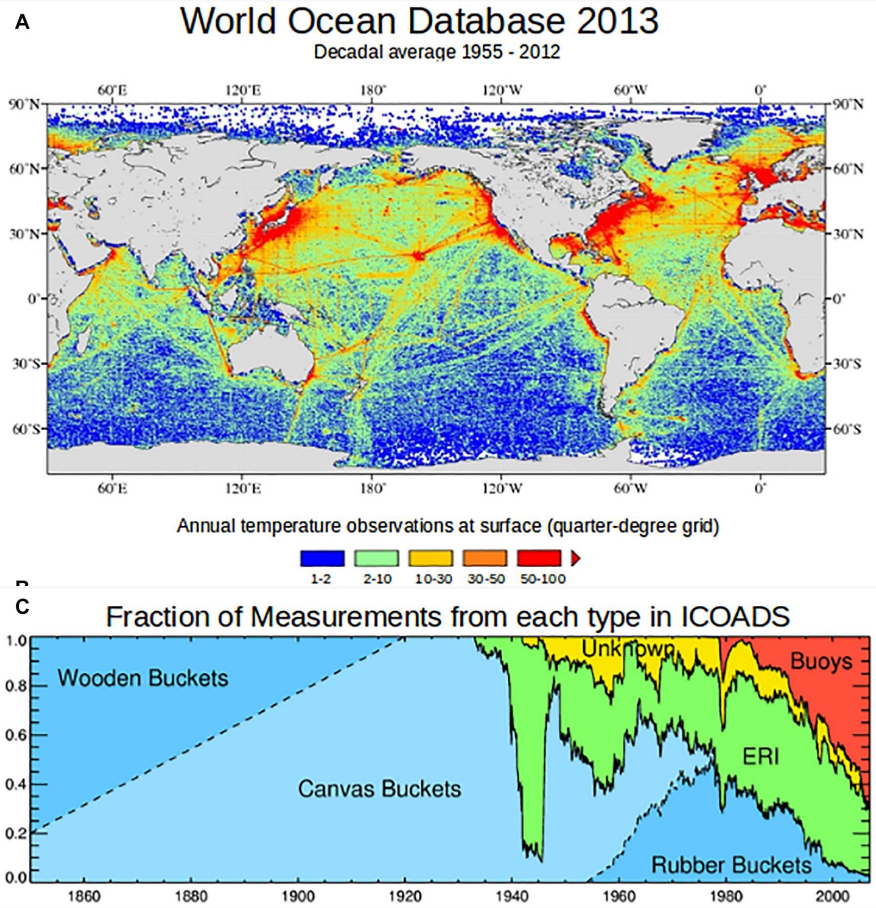

- NorESM1-ME (Bentsen et al. 2013)
- CMIP5 forcings
- 2° for atmosphere and land components
- 1° for ocean and ice components
- 30 ensemble members

(Counillon et al., 2016)

# Which data did we assimilate in CoRea1860+?

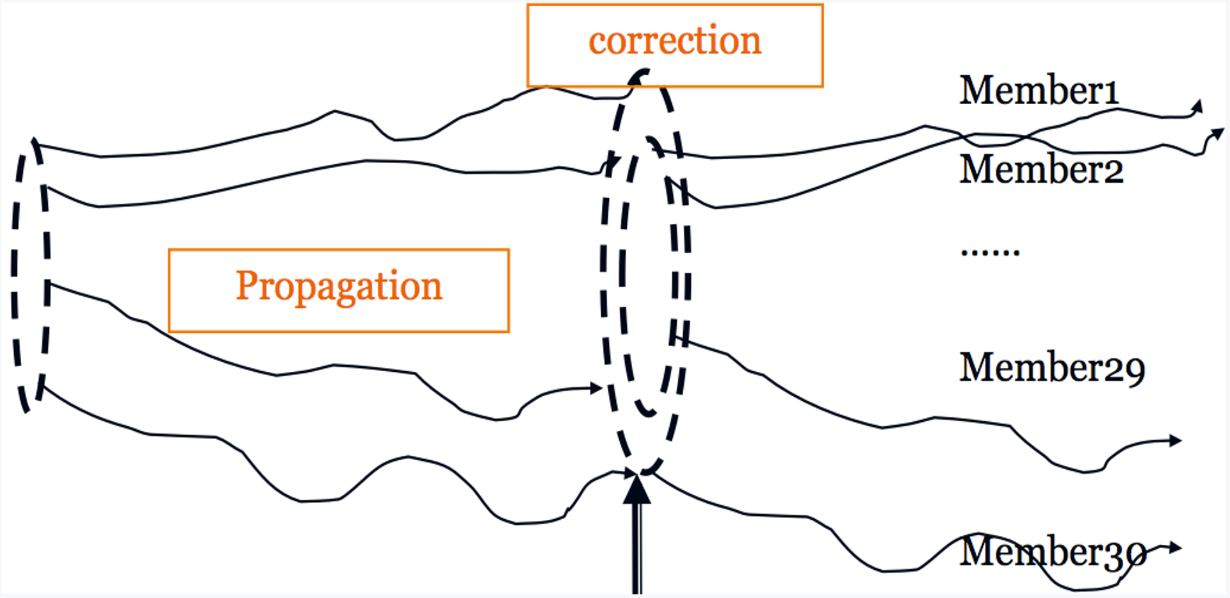
- The HadISST2 product (accounts for bias between measurements and provide a good uncertainty estimate)
- SST data is the most primary instrumental data prior to satellite era
- The use of new data in the course of the reanalysis introduces discontinuity

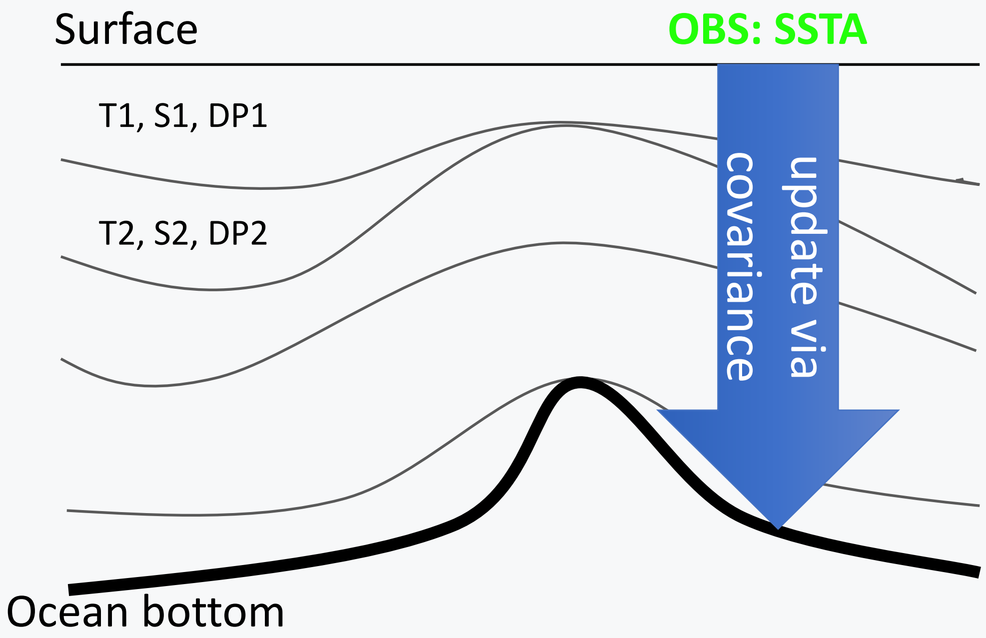



(O'Carroll et al. 2019)

# How did it work?

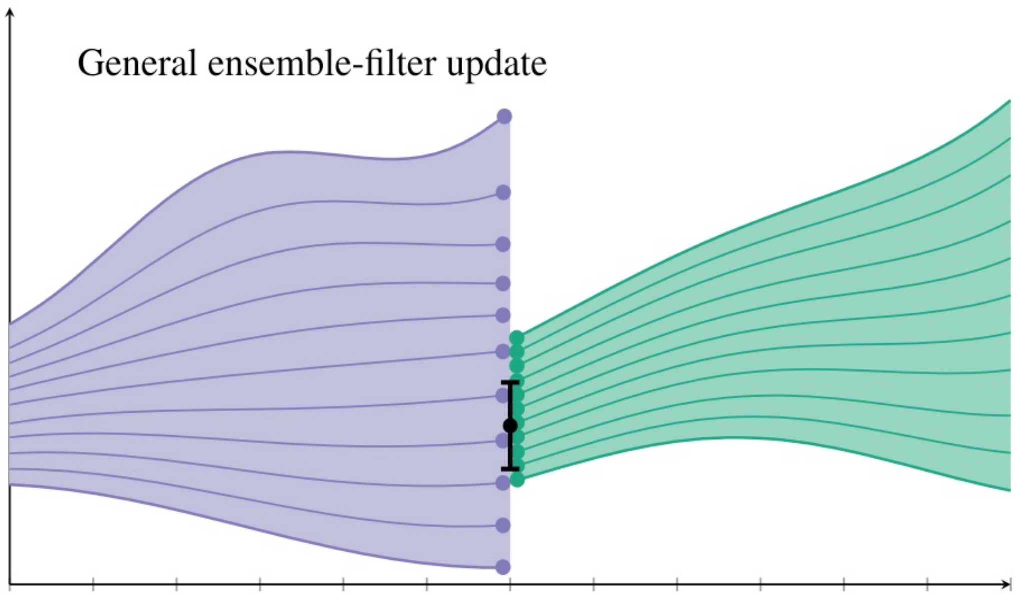


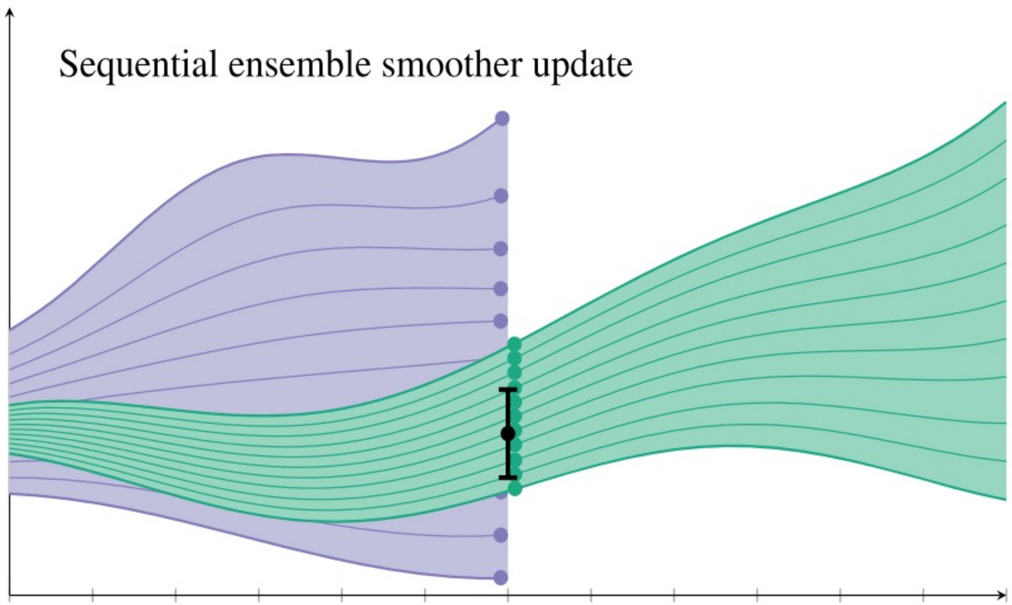




**Key features:**

- Anomaly assimilation (Carrassi et al., 2014), limiting the emergence of bias in the deep ocean
- Flow-dependent covariance (i.e., changing with the climate regime)
- Conserving heat/salt content (Wang et al., 2017)
- Isopycnal assimilation efferently use surface observations (Gavart and De Mey,1997)
- Small ensemble size (vertical localisation, Wang et al., 2022)

# Ensemble Kalman filter (EnKF)



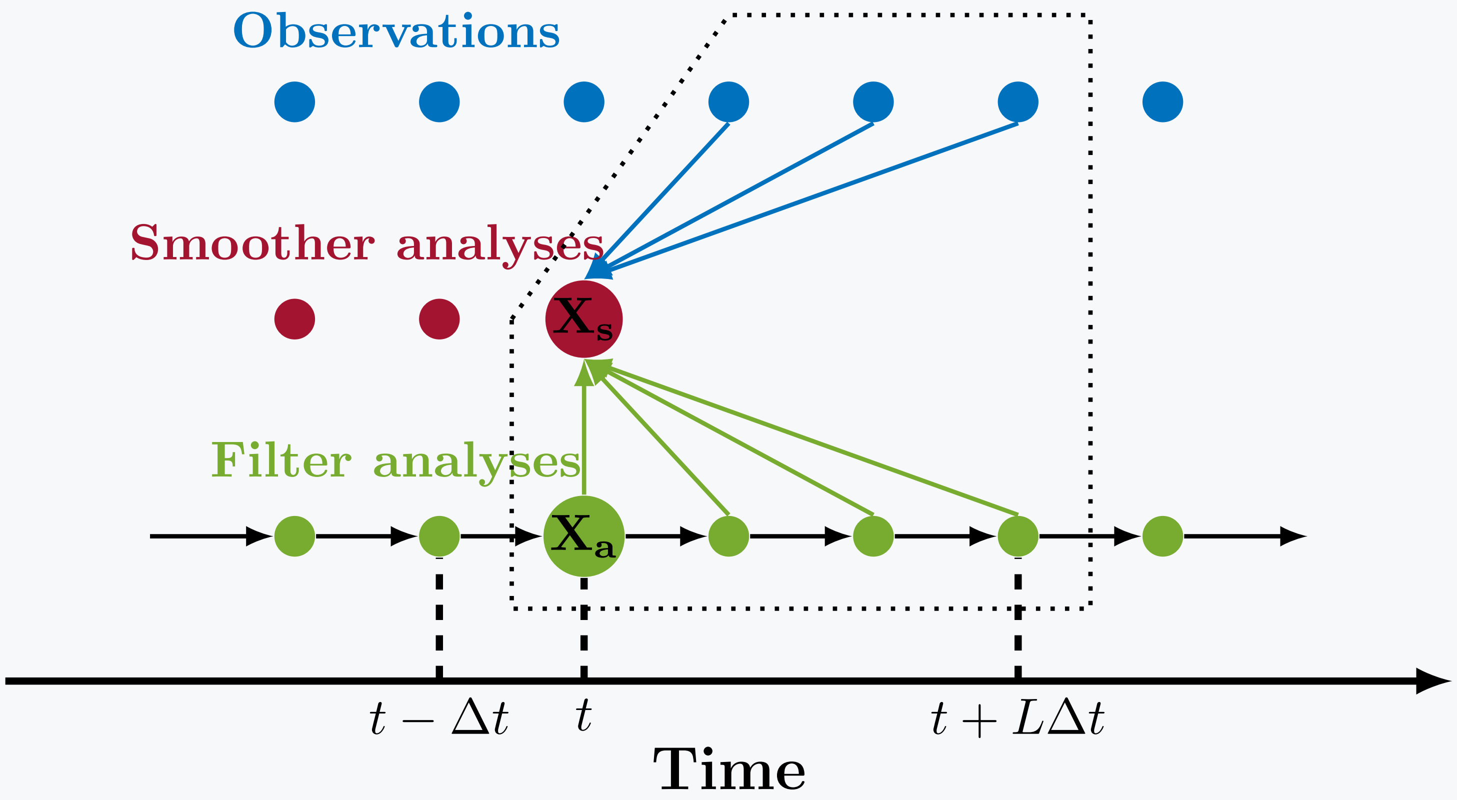




Probabilistic characteristic:

- uncertainty quantification through Monte-Carlo ensembles
- flow-dependent forecast error covariances

(Evensen et al., 2022)

# Offline EnKS




We define the concatenated observations as

$$\mathbf{y}_\mathrm{s} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_L^T\right]^T, \tag{6}$$

where $\mathbf{y}_k$ are observations at $k$ time steps ahead of the existing analysis $\mathbf{X}_a$ at time $t$ (Eq. 3) and $1 \leq k \leq L$. Note that $\mathbf{y}_0$ (i.e., $\mathbf{y}$ in Eq. 1) has been assimilated during the production of the filter analysis $\mathbf{X}_a$ and are not included in $\mathbf{y}_s$ for the smoother analysis. The concentrated observation error covariance matrix becomes a block diagonal matrix as follows:

$$\mathbf{R}_\mathrm{s} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{R}_L \end{bmatrix}, \tag{7}$$

where observations from different times are assumed to be uncorrelated. The ensemble mean of the filter analyses within $L$ time steps ahead of the model state $\mathbf{X}_\mathrm{a}$ mapped to the observation space is written as

$$H_\mathrm{s}(\overline{\mathbf{x}}_{\mathrm{a},1:L}) = \left[H_1(\overline{\mathbf{x}}_{\mathrm{a},1})^T, H_2(\overline{\mathbf{x}}_{\mathrm{a},2})^T, \ldots, H_L(\overline{\mathbf{x}}_{\mathrm{a},L})^T\right]^T. \tag{8}$$

The tangent linear operator of the operator $H_s$ is

$$\mathbf{H}_\mathrm{s} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{H}_L \end{bmatrix}. \tag{9}$$

The concatenated ensemble anomaly is defined as

$$\mathbf{A}_{1:L} = \left[\mathbf{A}_1^T, \mathbf{A}_2^T, \ldots, \mathbf{A}_L^T\right]^T. \tag{10}$$

The Kalman gain of the offline EnKS is written as

$$\mathbf{K}_\mathrm{s} = \mathbf{A}_\mathrm{a}\mathbf{A}_{1:L}^T\mathbf{H}_\mathrm{s}^T\left(\mathbf{H}_\mathrm{s}\mathbf{A}_{1:L}\mathbf{A}_{1:L}^T\mathbf{H}_\mathrm{s}^T + (m-1)\mathbf{R}_\mathrm{s}\right)^{-1}. \tag{11}$$

The offline EnKS can be written as follows:

$$\overline{\mathbf{x}}_\mathrm{s} = \overline{\mathbf{x}}_\mathrm{a} + \mathbf{K}_\mathrm{s}\left(\mathbf{y}_\mathrm{s} - H_\mathrm{s}\left(\overline{\mathbf{x}}_{\mathrm{a},1:L}\right)\right), \tag{12}$$

$$\mathbf{A}_\mathrm{s} = \mathbf{A}_\mathrm{a} - \frac{1}{2}\mathbf{K}_\mathrm{s}\mathbf{H}_\mathrm{s}\mathbf{A}_{1:L}, \tag{13}$$

$$\mathbf{X}_\mathrm{s} = \overline{\mathbf{x}}_\mathrm{s}\mathbf{1}_m + \mathbf{A}_\mathrm{s}. \tag{14}$$

For simplification, we can rewrite Eq. 14 as follows:

$$\mathbf{X}_\mathrm{s} = \mathbf{X}_\mathrm{a}\mathbf{T}, \tag{15}$$

$$\mathbf{T} = \mathbf{I} + \mathbf{A}_{1:L}^T\mathbf{H}_\mathrm{s}\left(\mathbf{H}_\mathrm{s}\mathbf{A}_{1:L}\mathbf{A}_{1:L}^T\mathbf{H}_\mathrm{s}^T + (m-1)\mathbf{R}_\mathrm{s}\right)^{-1}\left(\left(\mathbf{y}_\mathrm{s} - H_\mathrm{s}\left(\overline{\mathbf{x}}_{\mathrm{a},1:L}\right)\right)\mathbf{1}_m - \frac{1}{2}\mathbf{K}_\mathrm{s}\mathbf{H}_\mathrm{s}\mathbf{A}_{1:L}\right), \tag{16}$$

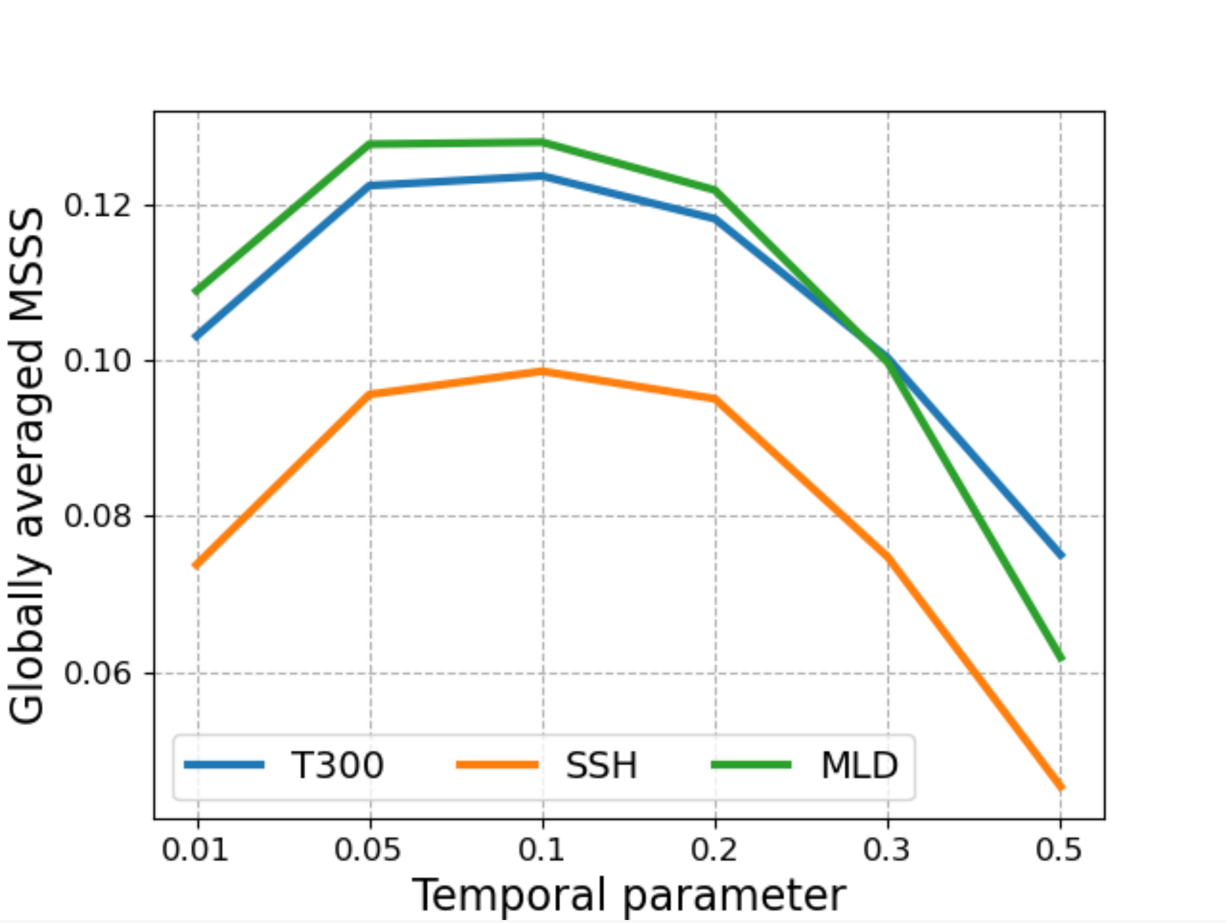
- Past and present observations as ensemble filter
- Future observations via cross-time error covariances
- Numerically cheap
- Tested in Lorenz 1963 (Dong et al., 2023)
- Investigate whether the offline EnKS as post-processing approach can improve long-term climate reanalyses

# Experiments in NorCPM

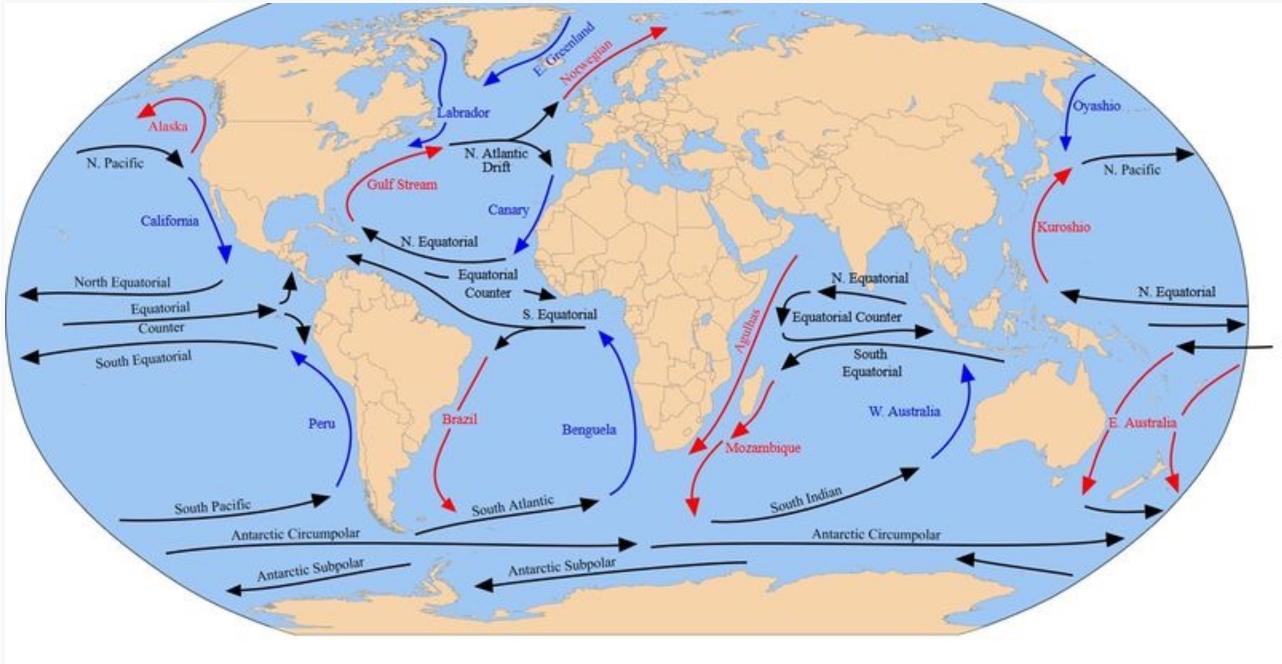
| Experiment | Idealised twin | Real-world |
|---|---|---|
| Pre-existing reanalysis | 1980-2010 (Wang et al., 2022) | 1950-2010 (Counillon et al., 2016) |
| Assimilated data | SST observations in three months in 'future' | |
| R Inflation factor | 1 | 1, 2, 4, 6, 8, 10, 16 or 25 |
| Spatial localization | Gaspari and Cohn, (1999) and Wang et al., (2017) | |
| Temporal localization | $\gamma^n$, where $\gamma$ signifies the decay rate per month and n denotes the time lag (Dong et al., 2021;2023), $\gamma$ = 0.01, 0.05, 0.1, 0.2, 0.3, and 0.5 | |
| Validation dataset | **Truth** | **Independent** datasets: EN4.2.1, ARMOR-3D L4, ORAS5 |
| Metrics | Mean squared skill score (MSSS) | |

$$\mathbf{MSSS} = 1 - \frac{\mathrm{MSE_s}}{\mathrm{MSE_f}} = \frac{\mathrm{MSE_f} - \mathrm{MSE_s}}{\mathrm{MSE_f}},$$

# Twin experiments: Globally averaged MSSS (monthly)



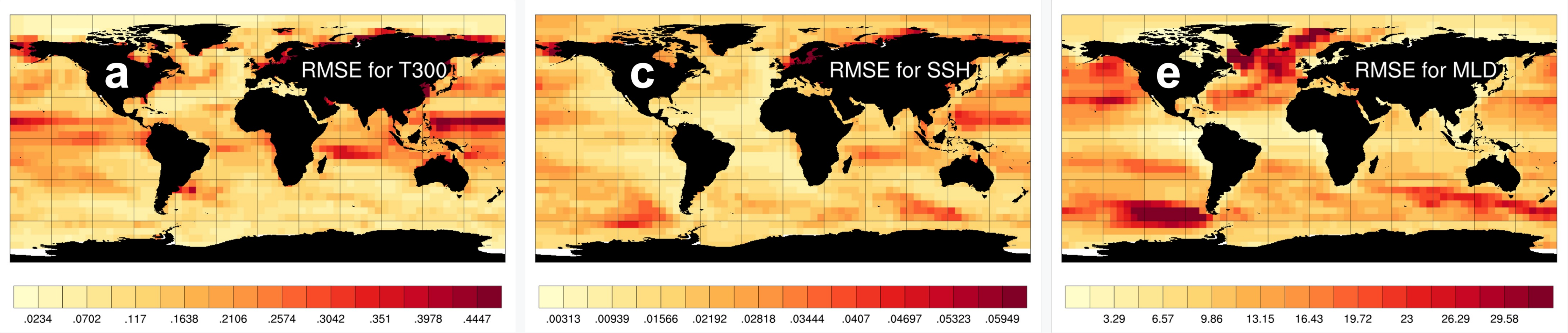

Global ocean circulations



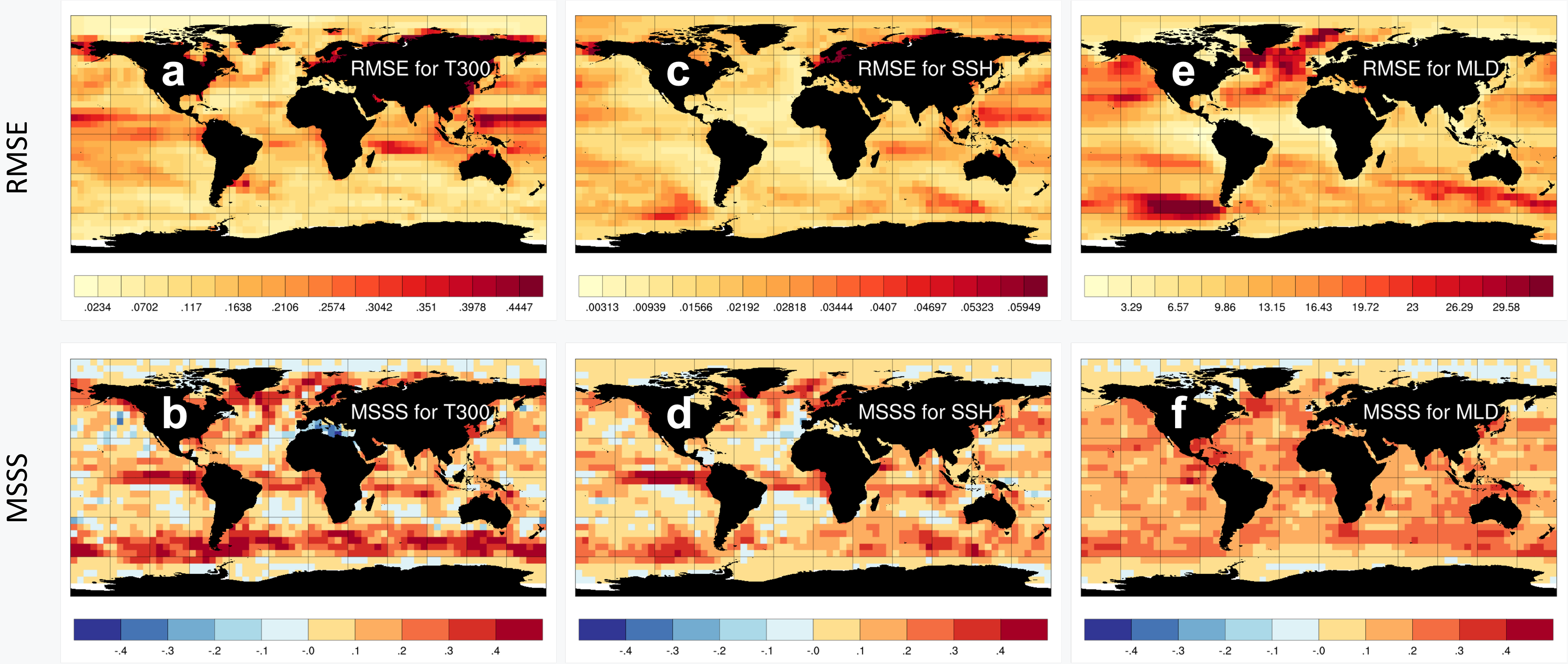

- Optimal results are obtained with the temporal parameter of 0.1 (e-folding time: 13.0 days).
- Assimilating future SST data is more efficient for MLD and T300 than SSH.

# Twin experiments: monthly T300, SSH and MLD

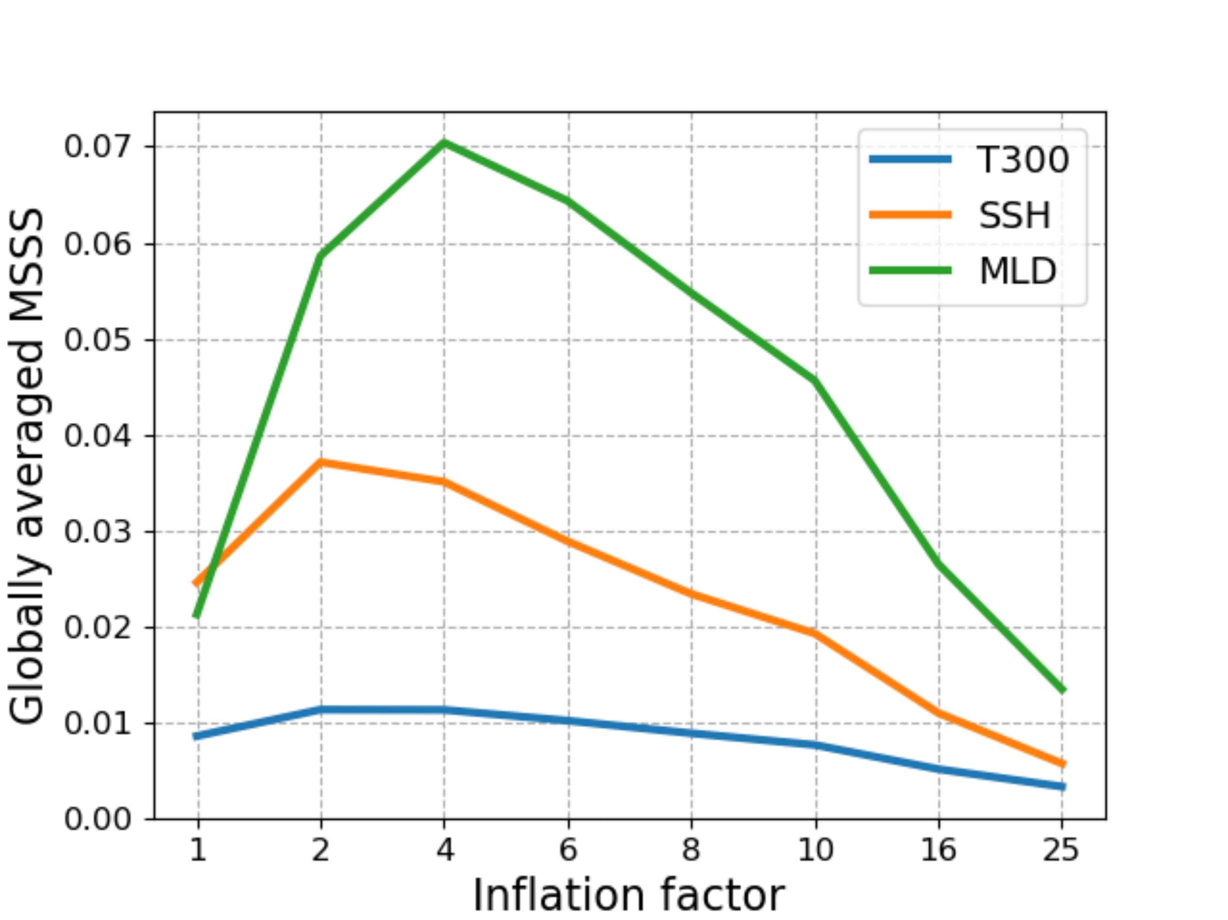
RMSE



# Twin experiments: monthly T300, SSH and MLD

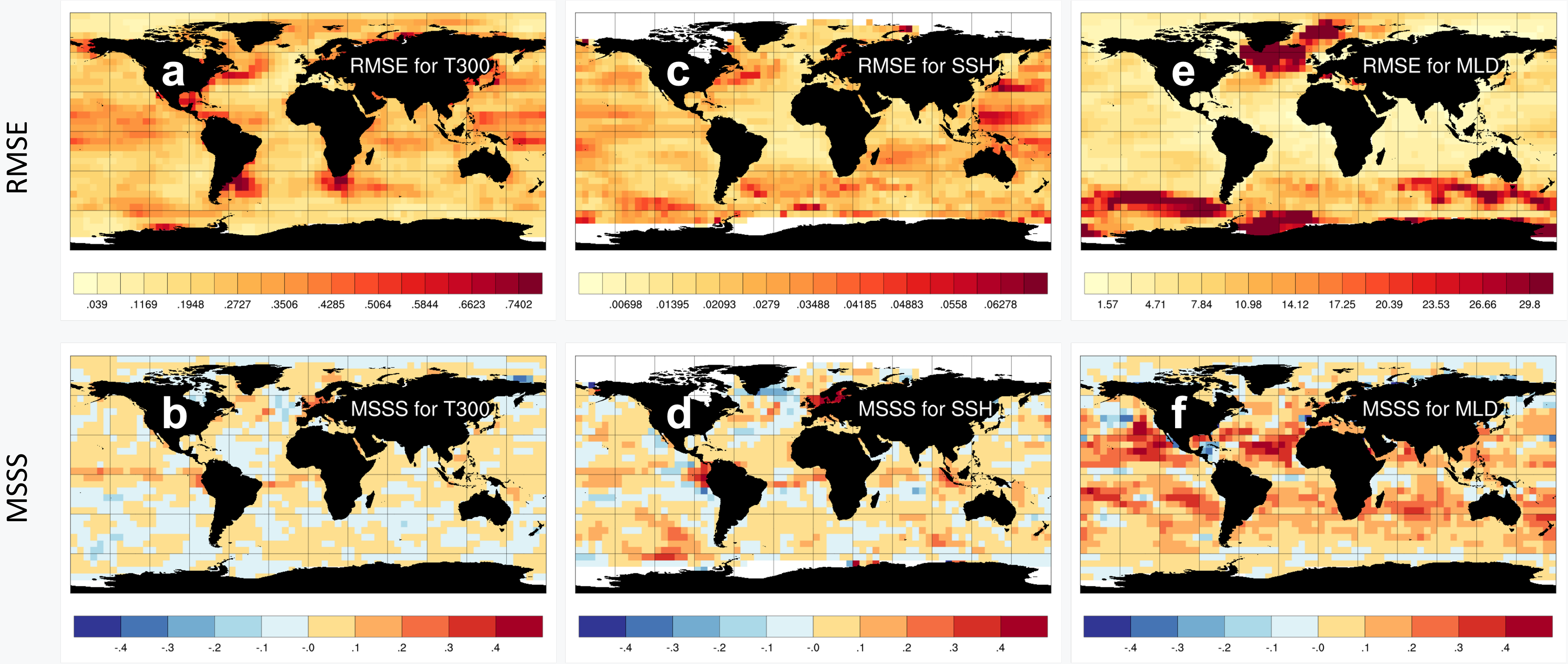


# Real experiments: globally averaged MSSS (yearly)

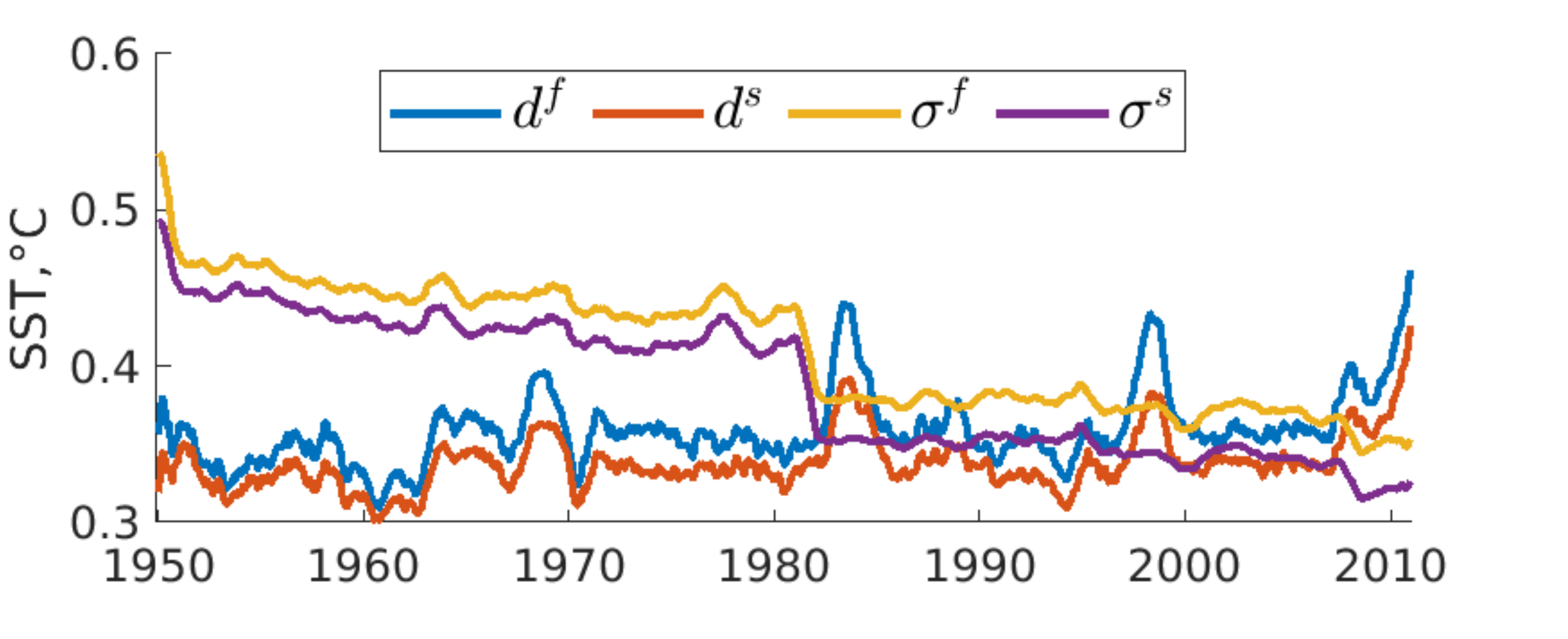



- Optimal results are obtained with the R inflation factor of 4 (when $\gamma = 0.1$)
- Assimilating future SST data is more efficient for MLD and SSH than T300

# Real experiments: Yearly T300, SSH and MLD



# Real experiments: reliability (Desroziers et al. , 2005)




$$d = \sqrt{\sum_i w_i d_i^2},$$

$$\overline{\sigma^r} = \sqrt{\sum_i w_i \left(\sigma_i^r\right)^2},$$

$$\overline{\sigma^o} = \sqrt{\sum_i w_i \left(\sigma_i^o\right)^2},$$

$$\sigma = \sqrt{\left(\overline{\sigma^r}\right)^2 + \left(\overline{\sigma^o}\right)^2},$$

- **Smoother reduces both RMSE and total error (i.e., combination of obs and background errors)**
- **Reliability is not significantly change**

# Take-home messages

NERSC

- Tested an offline ensemble Kalman smoother technique for climate reanalysis.
- Developed a methodology to fine temporal localisation parameters and R inflation factors.
- The results seem promising to improve the accuracy of the reanalysis.
- Different performance in twin experiments and real applications, due to many factors
  - Truth is unknown
  - Inconsistency between datasets
  - Model biases
- Application to deterministic reanalyses
  - Refer to the EnOI, a simplified variant of the EnKF (Evensen, 2003, Counillon et al., 2009)