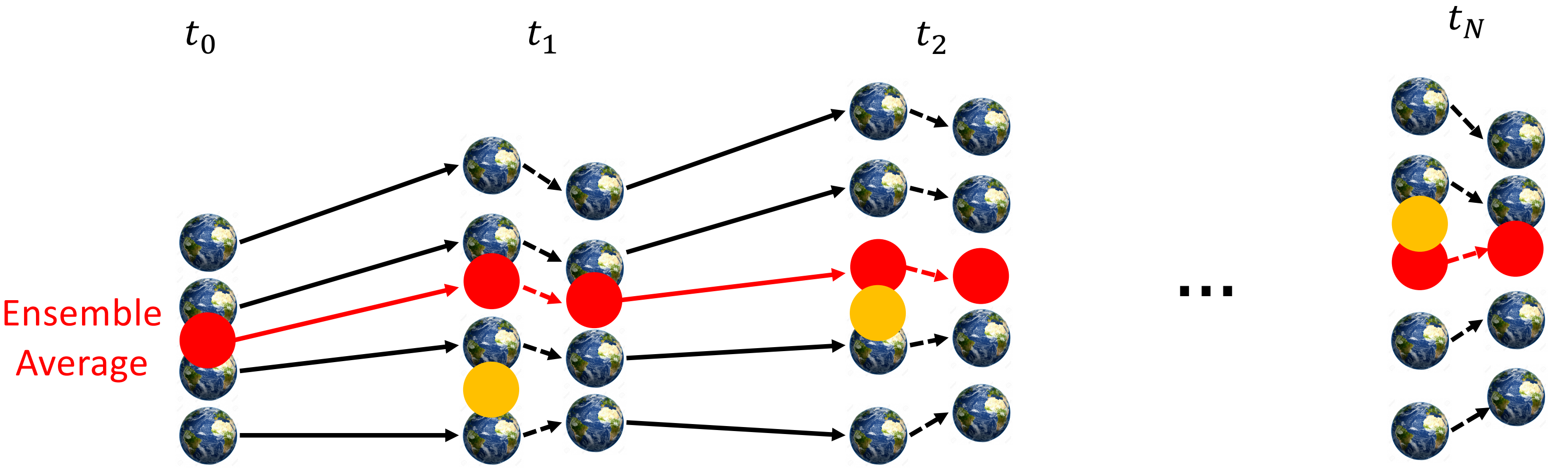


Ensemble Kalman filter parameter estimation of ocean optical properties for reduced biases in a coupled general circulation model

V. Kitsios, P. Sandery, T.J. O'Kane & R. Fiedler

Journal of Advances in Modeling the Earth System, Vol. 13, Issue 2, e2020MS002252

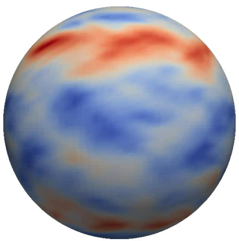
ACCESS Science Day, 10th-11th of December 2020, online

OCEANS AND ATMOSPHERE
www.csiro.au

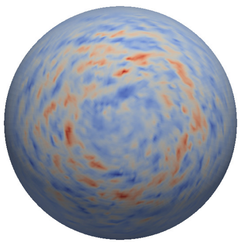
Acknowledge advice and feedback from Pavel Sakov.
Acknowledge funding from Decadal Climate Forecasting Project & Climate Resilience Enterprise Mission

Problem: Model Biases Limiting Predictability

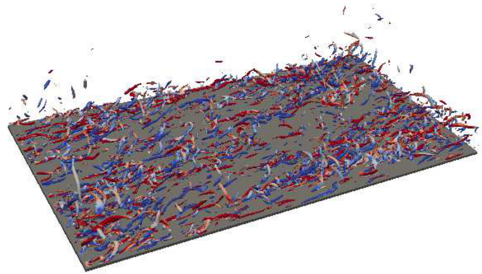
- In GCMs unresolved subgrid processes are parameterised. Many parameters are known with little precision, contributing to **model bias**.
- E.g. ENSO dynamics more difficult to represent in coupled GCMs with parameterised **air-sea fluxes** as opposed to prescribed BCs in O/A GCMs.
- **Typical approach:** Run various simulations, with manually tuned parameter values, and determine which produces the most realistic results.
- **Potential solution:** Calculate subgrid parameters from higher resolution *truth* simulation, such that lower resolution model matches the *true* stats.



Kitsios et. al. (2012, 2019, JAS)



Kitsios et. al. (2013, OM)

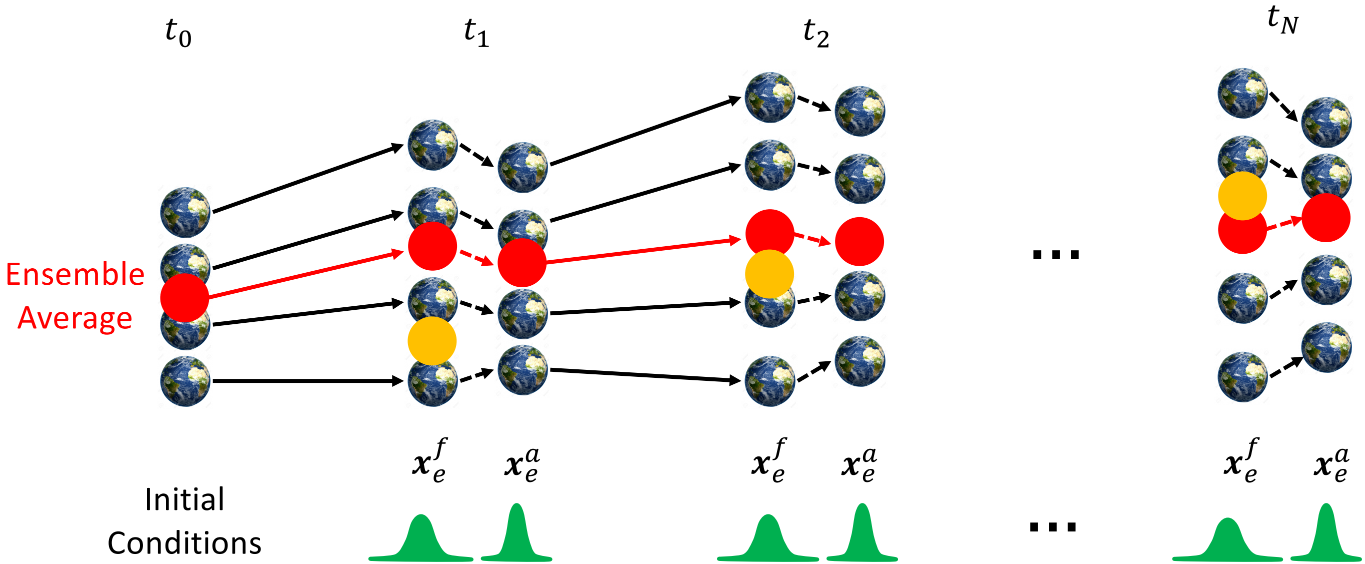


Kitsios et. al. (2017, C&F)

- What if a *truth* case is unavailable, or too computationally expensive?
- **Present approach:** Use the Ensemble Transform Kalman Filter, to estimate parameters governing the air-sea fluxes - specifically the ocean albedo (α), and the shortwave e-folding length scale in the ocean (L_{SW}).

Illustration of the Ensemble Kalman Filter

- Data assimilation (DA) modifies imperfect simulations of reality with a series of incomplete and possibly noisy measurements ($\mathbf{y}$), ideally resulting in a better representation of the true system state ($\mathbf{x}$).



$$\mathbf{x}_e^a(t) = \mathbf{x}_e^f(t) + \mathbf{K}\left[\mathbf{y}_e(t) - \mathbf{H}\mathbf{x}_e^f(t)\right]$$ for ensemble member e where

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left[\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right]^{-1}$$ with observational operator $\mathbf{H}$,

observational error covariance $\mathbf{R}$, and state error covariance $\mathbf{P}^f$.

Climate re-Analysis & Forecast Ensemble (CAFE) system

Data Assimilation System:

- ETKF designed for high dimensional geophysical models (enkf-c).
- $\mathbf{96}$ ensemble members (i.e. climate models).
- $\approx \mathbf{10^8}$ model states, with many parameters.
- $\approx \mathbf{10^4}$ CPUs required to simulate the GCM ensemble and solve EnKF.

Observations:

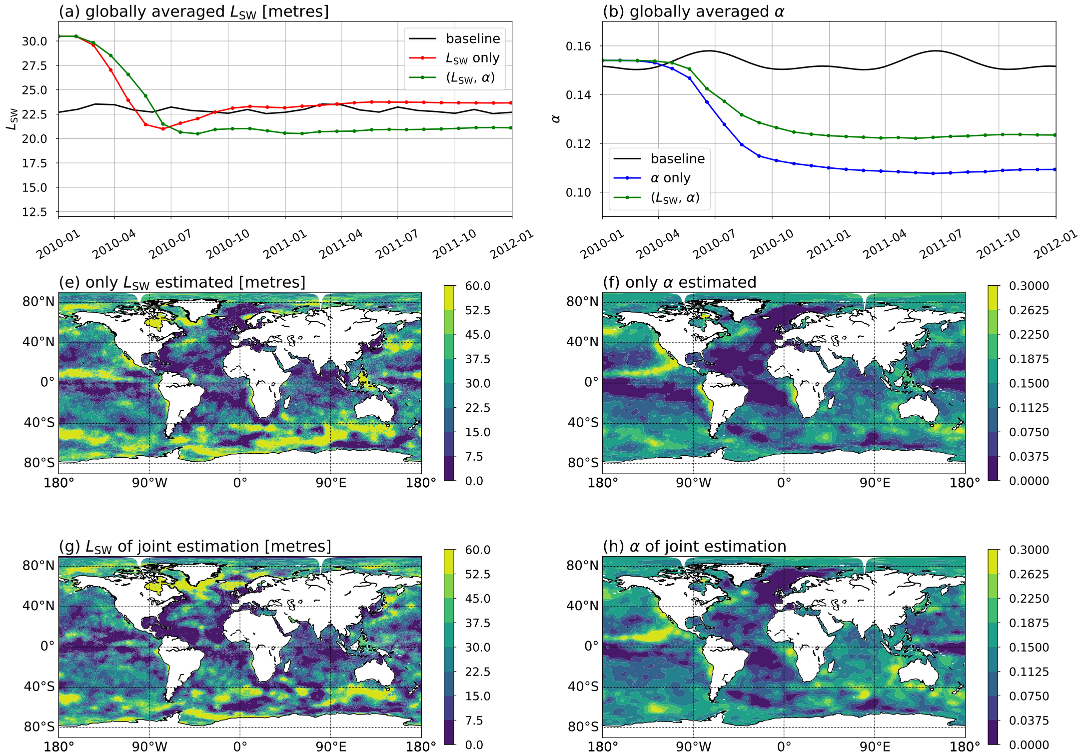
- Satellite and sub-surface observations of the ocean
- JRA55 reanalysis treated as atmospheric observations
- Satellite sea-ice concentration observations from Norwegian and Danish Meteorological Institutes.

General Circulation Model:

- Modified version of the Geophysical Fluid Dynamics Laboratory Coupled Model version 2.1
- Atmosphere: 2° latitude ; 2.5° longitude ; 24 vertical levels,
- Ocean: 1° longitude; finer latitudinal in specific regions; 50 vertical levels.
- Sea-ice: ocean horizontal resolution; 5 ice thickness categories.

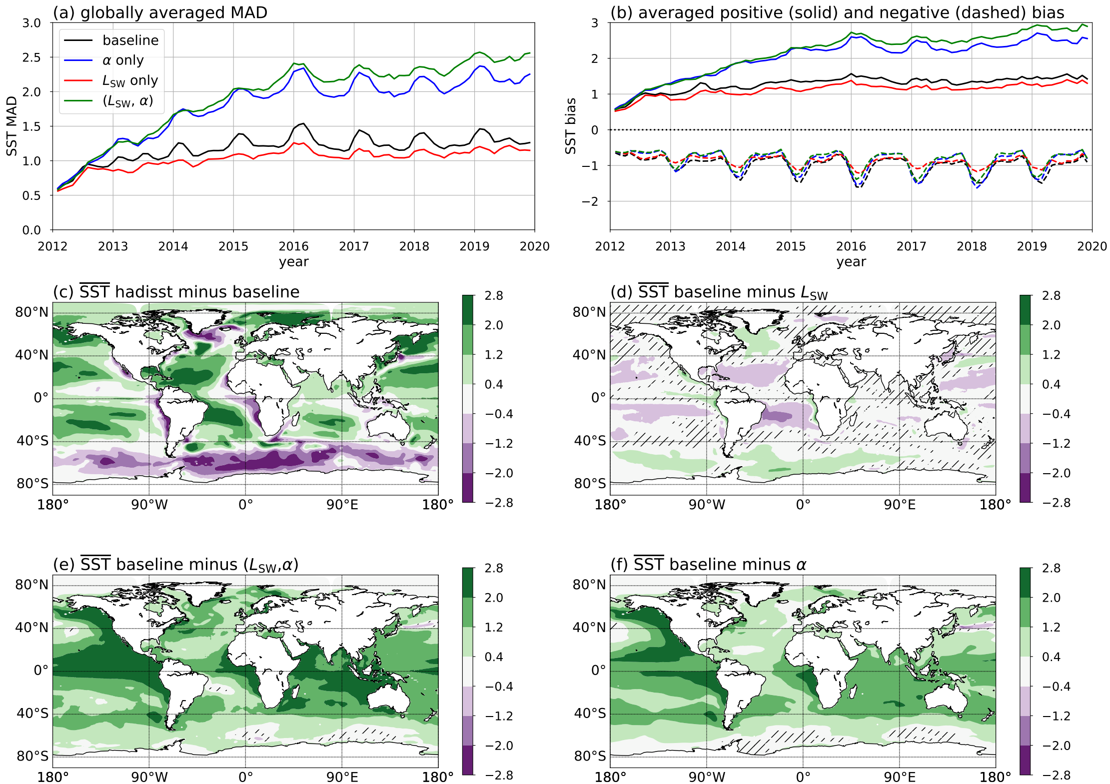
Spatio-temporal Properties of the Parameters

- DA undertaken from 1/1/2010 to 31/12/2011, on a 28-day cycle, with parameter estimation cases improving fit to observations.
- Parameters have a weak dependence upon time, with distribution more representative of biases than ocean optical properties.



Out-of-sample Multi-year Monthly Average SST Bias

- Only L_{SW} case has reduced globally averaged mean absolute deviation (MAD, $|\mathbf{y} - \mathbf{H}\langle \mathbf{x}^f \rangle|$) and positive and negative SST bias ($\mathbf{y} - \mathbf{H}\langle \mathbf{x}^f \rangle$) .
- Also true for sea-ice concentration and atmospheric jet position.

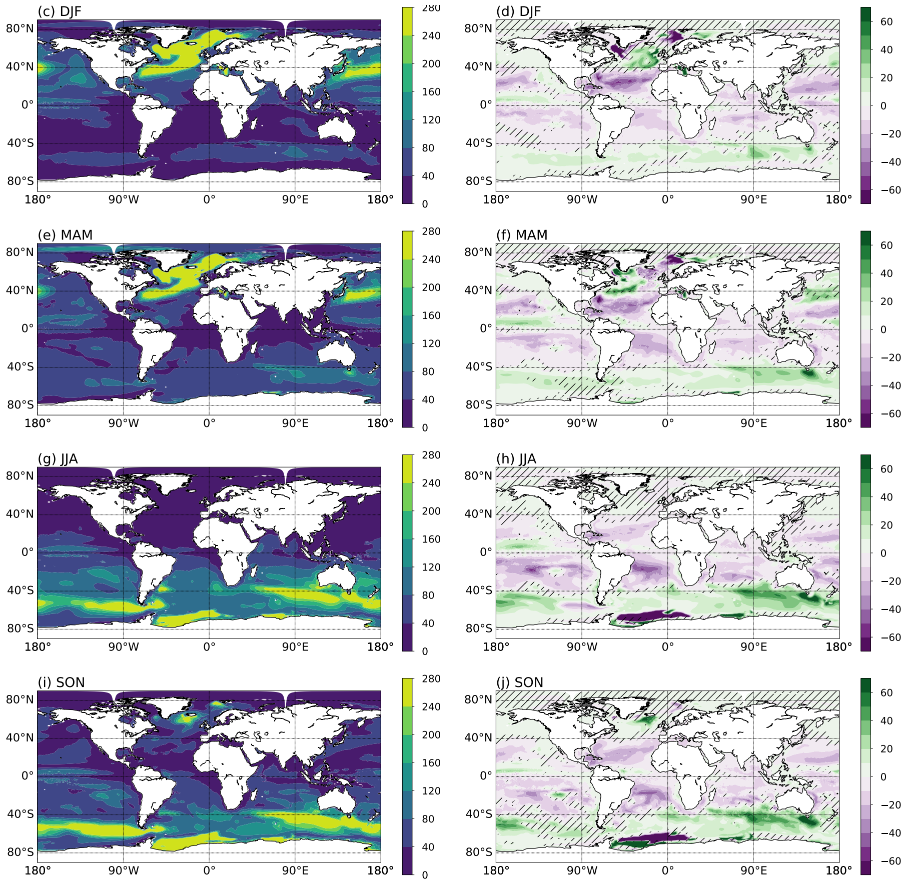


Concluding Remarks

- Ensemble Transform Kalman filter used to estimate spatio-temporally varying parameter maps of **ocean albedo** and the **ocean shortwave e-folding length scale** (L_{SW}) on the basis of short term 28-days forecasts of a coupled climate atmosphere / ocean / sea-ice GCM.
- Estimated parameter maps in these data assimilation experiments resemble known model biases, and **improve the in-sample fit of the coupled GCM** to a network of real world observations during 2010-2011 (see additional slides).
- However, only the multi-year forecast using the individually estimated map of L_{SW} systematically **improves out-of-sample skill** during 2012-2020.
- To our knowledge, this is the first attempt at undertaking parameter estimation on GCMs of this size and complexity.
- Whilst the parameter maps appear aphysical, there is a precedent for aphysical parameters in turbulence modelling (e.g. anisotropic and at times negative eddy viscosities), L_{SW} and α presumably not optimal, nor perhaps even an appropriate, selection of parameters.
- In current work we are estimating turbulence mixing parameters.

Changes to ocean mixed layer depth (MLD) per season

baseline $\quad$ L_{SW} case minus baseline



- MLD is seasonal because winter sea-ice formation increases salinity and hence density (left column).
- For L_{SW} case, parameters are time independent, but influence on ocean structure with respect to the baseline is seasonal (right column).
- Low (high) values of L_{SW} produce shallower (deeper) MLD - qualitatively similar to L_{SW} map.

Parameterisation of the Shortwave Flux

- Initial tests indicated that of the air-sea flux parameters, those governing the shortwave flux (F_{SW}) exhibited the greatest influence.
- Estimating parameters of the other fluxes were found to be either ineffective, or produced numerical instabilities - indicative of model form error.

Representation of shortwave radiation:

- In the atmosphere solved using an energy balance model.
- The air-sea flux at the interface governed by

$$F^{\mathrm{SW}} = F_0^{\mathrm{SW}}(1-\alpha), \text{ where ocean albedo } \alpha(\lambda,\phi,t) = \frac{0.037}{1.1\theta(\lambda,\phi,t)^{1.4}+0.15}$$

and θ is the zenith angle of the Sun.

- Shortwave flux at a given depth z of the ocean given by

$$\frac{F^{SW}(\lambda,\phi,z,t)}{F^{SW}(\lambda,\phi,z=0,t)} = \exp\left(\frac{-z}{L_{SW}(\lambda,\phi,t)}\right) \text{ where } L_{SW} \text{ is an e-folding length}$$

- Here we estimated α and L_{SW}, simultaneously and individually.

Ensemble Kalman Filter (EnKF) of Evensen (1996)

- Data assimilation (DA) modifies imperfect simulations of reality with a series of incomplete and possibly noisy measurements ($\mathbf{y}$), ideally resulting in a better representation of the true system state ($\mathbf{x}$).

System structure: Non-linear with Gaussian process and observational noise.

$$
\begin{aligned}
\mathbf{x}(t) &= \mathcal{F}\left(\mathbf{x}(t-1)\right) + \mathbf{w}(t), \text{ where } \mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(t)) \\
\mathbf{y}(t) &= \mathcal{H}\left(\mathbf{x}(t)\right) + \mathbf{v}(t), \text{ where } \mathbf{v}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(t))
\end{aligned}
$$

EnKF Forecast step: The ensemble of forecast states ($\mathbf{x}_e^f$) evolved in time and associated error covariance ($\mathbf{P}^f$) is sampled

$$
\begin{aligned}
\mathbf{x}_e^f(t) &= \mathcal{F}\left(\mathbf{x}_e^a(t-1)\right) \\
\mathbf{P}^f(t) &= \mathbf{A}\mathbf{A}^T/(m-1), \text{ columns of } \mathbf{A} \text{ contains } \mathbf{x}_e^f - \langle \mathbf{x}^f \rangle
\end{aligned}
$$

with m the number of ensemble members, and $\langle \cdot \rangle$ the ensemble average.

EnKF Analysis step: Sub-optimally minimises trace($\mathbf{P}^a$), where $\mathbf{P}^a$ is the error covariance of the analysed / corrected state ($\mathbf{x}_e^a$).

$$
\begin{aligned}
\mathbf{x}_e^a(t) &= \mathbf{x}_e^f(t) + \mathbf{K}\left[\mathbf{y}_e(t) - \mathcal{H}\left(\mathbf{x}_e^f(t)\right)\right] \text{ for ensemble member } e \\
\mathbf{K} &= \mathbf{P}^f\mathbf{H}^T\left[\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right]^{-1}, \text{ where } \mathbf{H} \text{ tangent linear operators of } \mathcal{H}
\end{aligned}
$$

Ensemble Kalman Filter for Parameter Estimation

- Redefine the state vector to include both the model states ($\mathbf{x}_S^f$) and parameters ($\mathbf{x}_\psi^f$), and associated observational operator (i.e. a satellite cannot observe a model parameter)

$$\mathbf{x}_e^f \equiv \begin{bmatrix} \mathbf{x}_S^f \\ \mathbf{x}_\psi^f \end{bmatrix}_e \text{, and } \mathbf{H} = [\mathbf{H}_S \ \mathbf{0}]$$

- Substituting above redefinitions

$$\langle \mathbf{x}^a \rangle = \langle \mathbf{x}^f \rangle + \mathbf{K}\left[\mathbf{y} - \mathbf{H}_S \langle \mathbf{x}_S^f \rangle\right] \text{, with } \mathbf{K} = \begin{bmatrix} \mathbf{P}_{SS}^f \mathbf{H}_S^T \\ \mathbf{P}_{\psi S}^f \mathbf{H}_S^T \end{bmatrix} \left[\mathbf{H}_S \mathbf{P}_{SS}^f \mathbf{H}_S^T + \mathbf{R}\right]^{-1}$$

- Kalman gain for the states is unchanged, but gain for the parameters is proportional to the covariance between the states and parameters $\mathbf{P}_{\psi S}^f$.
- For parameter estimation, the choice of forecast window ($t_1 - t_0$) is not obvious. In general, as the forecast window increases, $\mathbf{P}_{\psi S}^f$ decreases, associated $\mathbf{K}$ decreases, but the bias $\mathbf{y} - \mathbf{H}_S \langle \mathbf{x}_S^f \rangle$, increases.

In-sample DA error metrics

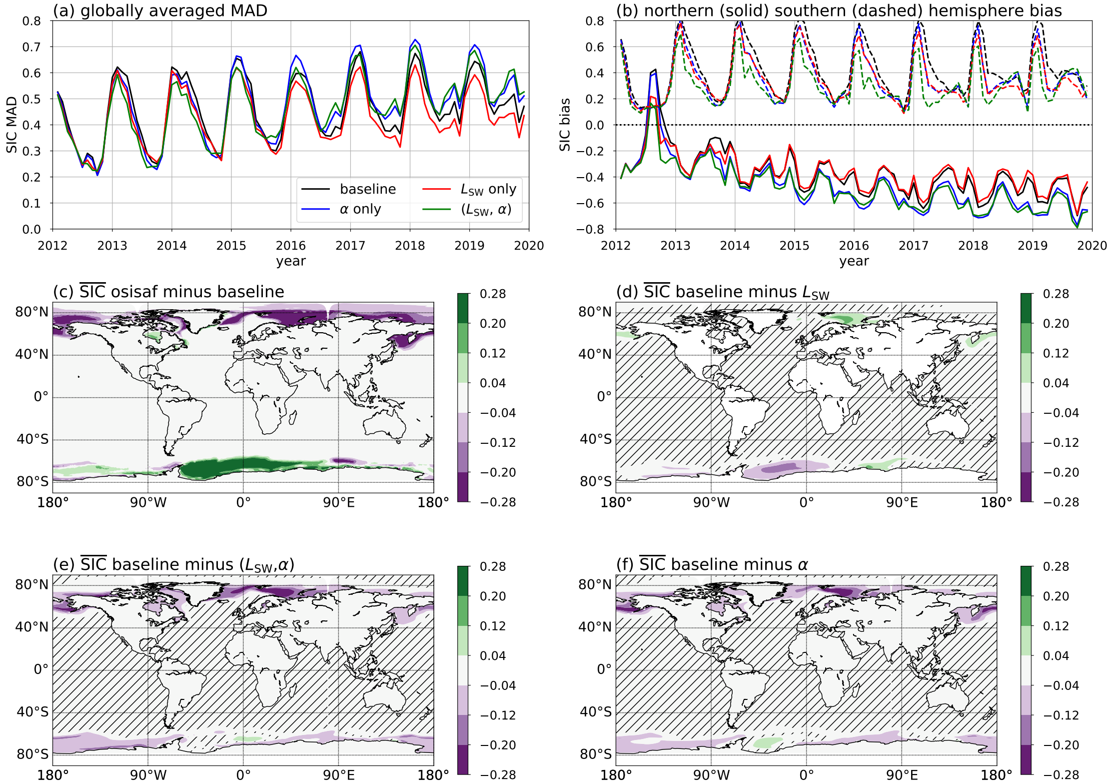
- Globally averaged mean absolute deviation (MAD, $|\mathbf{y} - \mathbf{H}\langle\mathbf{x}^f\rangle|$) and bias ($\mathbf{y} - \mathbf{H}\langle\mathbf{x}^f\rangle$) averaged over the second year of the DA experiments to remove "spin up" period.
- Note, the error is evaluated prior to the analysis step, therefore, it is a true measure of the forecast error.
- For selected key observations the estimated cases exhibit reduced in-sample error for the bold error measures.

		baseline	L_{SW} only	α only	$(L_{\mathrm{SW}}, \alpha)$
sea-ice concentration (m^2/m^2)	MAD	0.077	**0.071**	**0.075**	**0.072**
	bias	0.024	**0.021**	**0.022**	**0.017**
sea surface temperature (°K)	MAD	0.643	**0.634**	**0.634**	**0.634**
	bias	-0.022	**-0.014**	0.083	0.110
sub-surface ocean temperature (°K)	MAD	0.708	0.711	**0.702**	**0.694**
	bias	-0.106	**-0.098**	**-0.070**	**-0.041**

- However, since state and parameter estimation are intrinsically linked, one must assess the parameters in the freely running model during an out-of-sample period (i.e. 2012 to 2020).

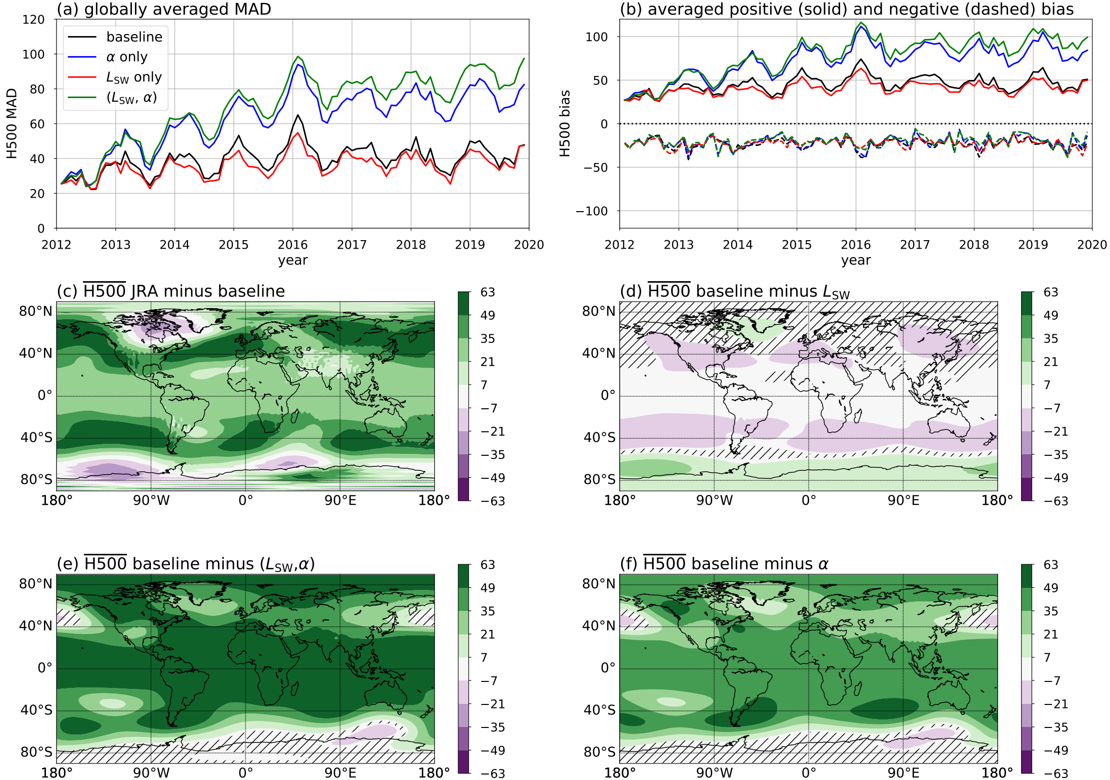
Out-of-sample Multi-year Monthly Average SIC Bias

- Only L_{SW} case has reduced MAD, and bias in both hemispheres.



Out-of-sample Multi-year Monthly Average H500 Bias

- Only L_{SW} case has systematically reduced MAD, and H500 bias.



Thank You

CSIRO Oceans and Atmosphere

Dr Vassili Kitsios
t
e Vassili.Kitsios@csiro.au
w research.csiro.au/dfp/