# Data assimilation in coupled chaotic dynamics and its combination to machine learning

**Alberto Carrassi**

(1) Dept of Meteorology & National Centre for Earth Observation, U of Reading - UK

(2) Mathematical Institute, U of Utrecht - NL

University of Reading | National Centre for Earth Observation, Natural Environment Research Council | Universiteit Utrecht

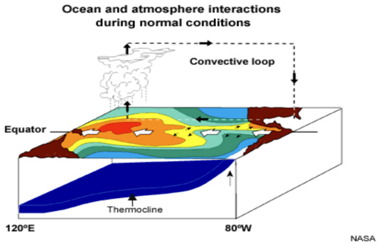
With: **L Bertino, M Bocquet, J Brajard, J Demaeyer, M Tondeur and S Vannitsem**

*Workshop: Multi-annual to Decadal Climate Predictability in the North Atlantic-Arctic Sector*

21st September 2021

# Coupled data assimilation - Issues and terminology

- The scale separation makes difficult to carry out the **uncertainty quantification** necessary to propagate the data content across compartments.
- If the scale separation is not very large, one can still rely upon uncoupled DA that operates on each compartment independently, and then use the full coupled model to forecast between successive observations $\Longrightarrow$ **weakly CDA**.
- In wCDA the effect of the coupling manifests indirectly via the model integration, yet the cross-compartment correlations (if any) are not exploited in the analysis update $\Longrightarrow$ *wCDA* may still be prone to produce imbalances.
- Full CDA, where the analysis update is across compartments, is called **strongly CDA**.



# The nature of the problem (focus on time-scale separation)

$$
\begin{aligned}
\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \epsilon\mathbf{f}(\mathbf{x},\mathbf{z}) \qquad \textit{Ocean} \\
\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} &= \mathbf{g}(\mathbf{x},\mathbf{z}) \qquad \textit{Atmosphere}
\end{aligned} \tag{1}
$$

- $\epsilon \ll 1$ accounts for the temporal scales difference between the processes $\mathbf{f}$ and $\mathbf{g}$
- We assume that $\mathbf{f}$ and $\mathbf{g}$ are both bounded from above as $\mathcal{O}(1)$, and that the characteristic spatial scales of $\mathbf{x}$ and $\mathbf{z}$ are similar.
- The model sub-components will grow differently: within the time interval $t_k - t_{k-1} = \mathcal{O}(1)$ the slow scale changes such as $\mathcal{O}(\mathbf{x}_k) = \mathbf{x}_{k-1} + \mathcal{O}(\epsilon)$, while the fast scale as $\mathcal{O}(\mathbf{z}_k) = \mathbf{z}_{k-1} + \mathcal{O}(1)$.
- *Remark*: In realistic coupled atmosphere-ocean system, $\mathbf{x}$ and $\mathbf{z}$ may also have different amplitudes bounds and different spatial scales.

▶ In order for the observations of the slow scale to monitor its variability, we need to have $\Delta t^{\mathrm{x}} \leq \mathcal{O}(\epsilon^{-1})$.

▶ When $\Delta t^{\mathrm{z}} = \mathcal{O}(1)$ (*i.e.* frequent obs) the solution of the slow system, $\mathbf{x}(t)$, is approximately constant in the interval $t \in [t_k, t_k + \Delta t^{\mathrm{z}}] \Longrightarrow$ A good scenario for *wCDA*.

# The nature of the problem (a linear analysis)

The linearised error evolution between two subsequent analyses reads

$$\begin{aligned} \Delta\mathbf{x}^{\mathrm{f}}_{k} &\approx \epsilon\mathbf{F}_{\mathbf{x}}\Delta\mathbf{x}^{\mathrm{a}}_{k-1} + \epsilon\mathbf{F}_{\mathbf{z}}\Delta\mathbf{z}^{\mathrm{a}}_{k-1}, \\ \Delta\mathbf{z}^{\mathrm{f}}_{k} &\approx \mathbf{G}_{\mathbf{x}}\Delta\mathbf{x}^{\mathrm{a}}_{k-1} + \mathbf{G}_{\mathbf{z}}\Delta\mathbf{z}^{\mathrm{a}}_{k-1}, \end{aligned} \tag{2}$$

- The cross terms describe the error impact across compartments: $\mathbf{F}_{\mathbf{z}}\Delta\mathbf{z}^{\mathrm{a}}_{k-1}$ for the *Fast*-to-*Slow* dependence, $\mathbf{G}_{\mathbf{x}}\Delta\mathbf{x}^{\mathrm{a}}_{k-1}$ for the *Slow*-to-*Fast*.
- They encode, *e.g.*, the *Atmosphere-Ocean* mechanical transfer of kinetic energy and heat transfer.
- By inserting the error order in (2) and taking the norm of both sides we have

**1-st order error dynamics - Error bounds**

$$\begin{aligned} \mathcal{O}(\Delta\mathbf{x}^{\mathrm{f}}_{k}) &\approx \epsilon[\|\mathbf{F}_{\mathbf{x}} + \mathbf{F}_{\mathbf{z}}\|]\mathcal{O}(1) \leq \epsilon[\|\mathbf{F}_{\mathbf{x}}\| + \|\mathbf{F}_{\mathbf{z}}\|]\mathcal{O}(\Delta\mathbf{x}^{\mathrm{a}}_{k-1}), \\ \mathcal{O}(\Delta\mathbf{z}^{\mathrm{f}}_{k}) &\approx [\|\mathbf{G}_{\mathbf{x}} + \mathbf{G}_{\mathbf{z}}\|]\mathcal{O}(1) \leq [\|\mathbf{G}_{\mathbf{x}}\| + \|\mathbf{G}_{\mathbf{z}}\|]\mathcal{O}(\Delta\mathbf{z}^{\mathrm{a}}_{k-1}). \end{aligned} \tag{3}$$

## Sensitivity bounds: *how to choose where to observe...*

- Suppose (very lucky situation!) one wants to have $\mathcal{O}(\Delta\mathbf{x}^{\mathrm{f}}_{k}) = \mathcal{O}(\Delta\mathbf{x}^{\mathrm{a}}_{k-1})$ and $\mathcal{O}(\Delta\mathbf{z}^{\mathrm{f}}_{k}) = \mathcal{O}(\Delta\mathbf{z}^{\mathrm{a}}_{k-1})$, then

**1-st order sensitivity of the Slow Scale**

$$\begin{aligned} \mathcal{O}(\|\mathbf{F}_{\mathbf{x}}\|) &\leq \mathcal{O}(\epsilon^{-1}) \quad \textit{Slow} \mapsto \textit{Slow} \text{ sensitivity},\\ \mathcal{O}(\|\mathbf{F}_{\mathbf{z}}\|) &\leq \mathcal{O}(\epsilon^{-1}) \quad \textit{Fast} \mapsto \textit{Slow} \text{ sensitivity}, \end{aligned} \tag{4}$$

**1-st order sensitivity of the Fast Scale**

$$\begin{aligned} \mathcal{O}(\|\mathbf{G}_{\mathbf{x}}\|) &\leq \mathcal{O}(1) \quad \textit{Slow} \mapsto \textit{Fast} \text{ sensitivity},\\ \mathcal{O}(\|\mathbf{G}_{\mathbf{z}}\|) &\leq \mathcal{O}(1) \quad \textit{Fast} \mapsto \textit{Fast} \text{ sensitivity}. \end{aligned} \tag{5}$$

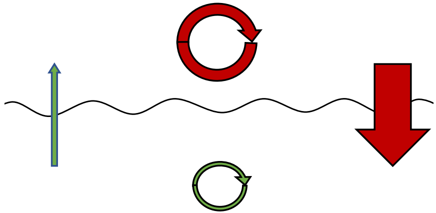
- The fast scale bounds are much smaller $\Longrightarrow$ analysis error must be kept within $\mathcal{O}(1)$ otherwise a "locally" large sensitivity $\|\mathbf{G}_{\mathbf{z}}\|$, beyond $\mathcal{O}(1)$, will lead the forecast error to grow over $\mathcal{O}(1)$.

- *Remark*: When doing CDA for state estimate, we do not act on the TLMs directly, but only on the trajectory upon which the Jacobian is evaluated $\Longrightarrow$

CDA for **coupling parameter** estimation can be more effective (it acts directly on $\|\mathbf{G}\|$ and $\|\mathbf{F}\|$).

# Coupled DA: conclusions from the linear framework

- **Cross compartments effects are generally stronger in the direction from the slow to the fast scale**.
- But, the **inter compartments effects are stronger in the fast scale** which requires to be controlled by observations.
- The rate at which the fast scale error is bound to grow (if left uncontrolled) is such that it can quickly reach high level and thus affecting the slow scale.

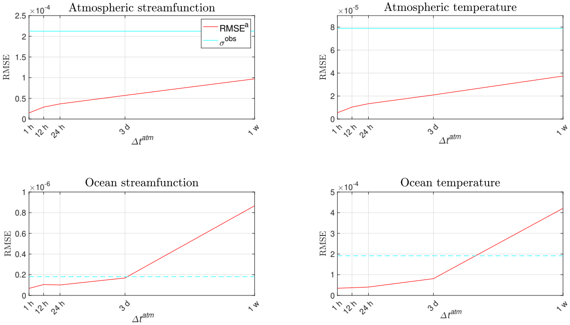


**Observing the fast scale** is thus of paramount importance.

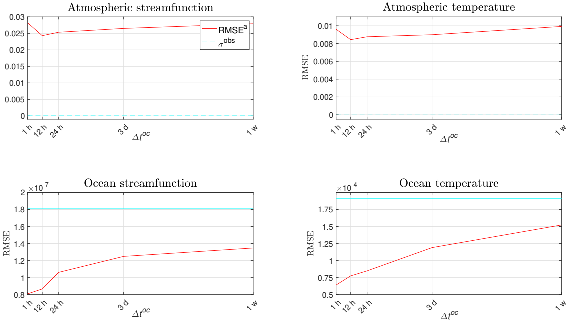
- *Remark*: The above is an effect of the temporal scale difference alone.

# Strongly coupled EnKF with atmosphere-ocean model MAOOAM

## Observations in the atmosphere




## Observations in the ocean

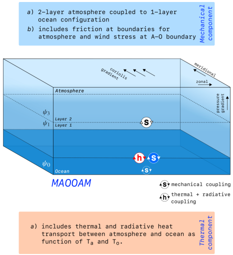



- Atmosphere practically insensitive to the removal of the ocean observations.
- For $1h \leq \Delta t^{\text{atm}} \leq 3d$ ocean controlled by atmospheric data only.

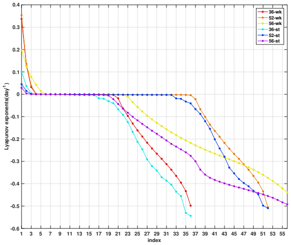
- Atmosphere is not controlled by ocean data: error above the observational level.
- Ocean error slightly lower than with only atmospheric data but not as good as when both ocean and atmosphere is observed.

# Coupling and instablities: *Do they relate?*

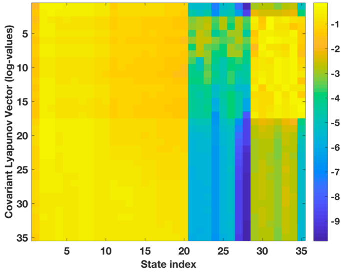
### MAOOAM coupled model




### LEs spectr




### Projection of Cov Lyap vectors



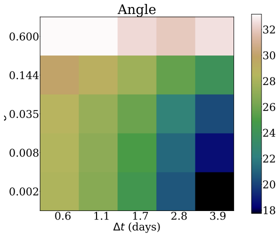

Tondeur, *et al*, 2020

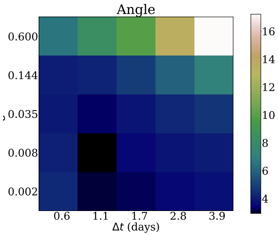
- ▶ The coupling leads to the appearence of **“quasi-neutral”** modes.
- ▶ Unstable and Stable CLVs show a transition in projections $\Longrightarrow$ Instabilities are either originated in the atmosphere or in the ocean.
- ▶ The “quasi-neutral” CLVs show comparable projections on both atmosphere and ocean $\Longrightarrow$ **They are a manifestation of the coupling**.

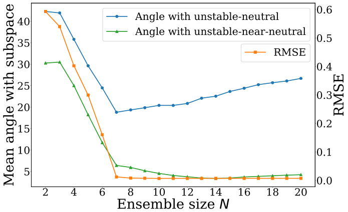
**Coupled DA should rely on CLVs to propagate information across model components**

# **Strongly coupled EnKF**: instabilities tracking & minimum ensemble

- Angle ensemble span with unstable-neutral (**left**) and unstable plus quasi-neutral modes (**center**).




**In coupled DA, all "quasi-neutral" modes – related to the coupling – must be taken into account**

Carrassi *et al*, 2021

# Combined CDA-ML to infer unresolved scales parametrizations

**The objective is to produce a hybrid (physical/data-driven) model**

$$\mathbf{x}(t+\delta t) = \mathcal{M}^{\varphi}[\mathbf{x}(t)] + \mathcal{M}^{\mathrm{UN}}[\mathbf{x}(t)],$$

where:

- $\mathbf{x}(t)$ is the state of the dynamical system
- $\mathcal{M}^{\varphi}$ is the physical model (assumed to be known a priori)
- $\mathcal{M}^{\mathrm{UN}}$ is the unresolved component of the dynamics to be inferred from data
- $\delta t$ is the integration time step

$\mathcal{M}^{\mathrm{UN}}$ is approximated by a **data-driven model** represented under the form of a neural network whose parameters are $\boldsymbol{\theta}$: $\mathcal{M}_{\boldsymbol{\theta}}[\mathbf{x}(t)]$

# Proposed approach

Simplified description of the algorithm:

1. Estimating the state $\mathbf{x}^{\mathrm{a}}_{1:K}$. At each time $t_k$, we calculate a forecast $\mathbf{x}^{\mathrm{f}}$:

$$\mathbf{x}^{\mathrm{f}}_{k+1} = \mathbf{x}^{\mathrm{f}}(t_k + \Delta t) = (\mathcal{M}^{\varphi})^{N_c}(\mathbf{x}^{\mathrm{a}}_k)$$

   An observation $\mathbf{y}_{k+1}$ is assimilated with **strongly coupled EnKF** to produce an analysis $\mathbf{x}^{\mathrm{a}}_{k+1}$

2. Determining an estimation of the unknown part of the model. We assume that:
   - $\mathbf{x}(t+\Delta t) \approx (\mathcal{M}^{\varphi})^{N_c}(\mathbf{x}(t)) + N_c \cdot \mathcal{M}^{\mathrm{UN}}[\mathbf{x}(t)]$
   - $\mathbf{x}(t) \approx \mathbf{x}^{\mathrm{a}}(t)$

   We consider that $\mathcal{M}^{\mathrm{UN}}(\mathbf{x}_k) \approx \mathbf{z}_{k+1} = 1/N_c \cdot \left(\mathbf{x}^{\mathrm{a}}_{k+1} - \mathbf{x}^{\mathrm{f}}_{k+1}\right) \Longrightarrow$ The "target" (*i.e.* the model error) is estimated using the *analysis increments* (Carrassi and Vannitsem, 2011).

3. Training a neural network $\mathcal{M}_{\boldsymbol{\theta}}$ by minimising the loss $L(\boldsymbol{\theta}) = \sum_{k=0}^{K-1} ||\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x}^{\mathrm{a}}_k) - \mathbf{z}_{k+1}||^2$

4. Using the hybrid model $\mathcal{M}^{\varphi} + \mathcal{M}_{\boldsymbol{\theta}}$ to produce new simulations (*e.g.* to make forecasts).

Brajard *et al*, 2021

# Experiments with MAOOAM

1. **Truth**: $n_{\mathrm{a}} = 20$ and $n_{\mathrm{o}} = 8$ modes for atmosphere and ocean. **Total dimension** $N_x = 56$.
2. **Truncated**: $n_{\mathrm{a}} = 10$ and $n_{\mathrm{o}} = 8$ modes for atmosphere and ocean. **Total dimension** $N_x = 36$.

- The truncated model is **missing 20 high-order atmospheric variables**
- There is not locality in spectral space so the NN is made of 3 layers multi-layer perceptrons

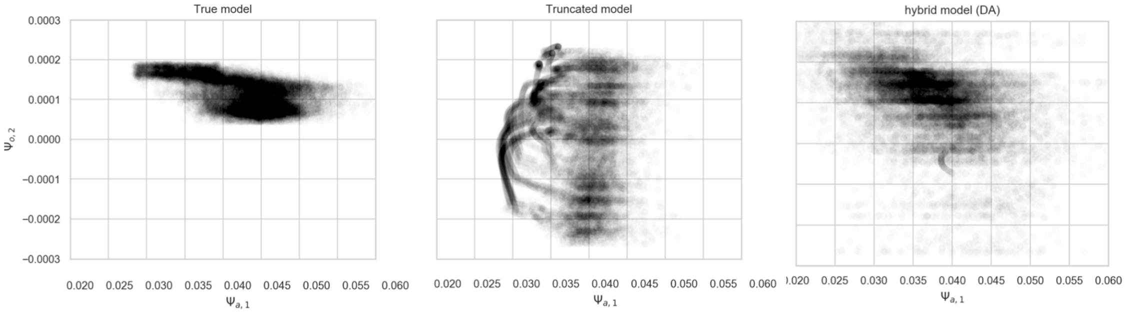
**RMSE-f of hybrid and truncated MAOOAM models**

| RMSE-f(lead time $\tau$) | $\psi_{o,2}$(2 years) | $\theta_{o,2}$(2 years) | $\psi_{a,1}$(1 day) |
|---|---|---|---|
| Truncated | 0.23 | 0.21 | 0.36 |
| **Coupled DA-ML hybrid** | **0.10** | **0.06** | **0.28** |

- The hybrid models have superior skill to the truncated model.
- The improvement is larger for the ocean that is fully resolved $\Longrightarrow$ **Enhanced representation of the atmosphere-ocean coupling processes thanks to coupled DA**.
- The atmosphere is improved less: the hybrid is not very good in representing the fast processes.

# Numerical experiments: atmosphere-ocean model MAOOAM

## Reconstruction of the model attractor




- ▶ The truncated model visits areas of the phase space that are not admitted in the real dynamics.
- ▶ Discrepancies are reduced by the hybrid models.

Brajard *et al*, 2021

# Conclusions and paths

- We have deduced the 1-st order **dynamical mechanisms for error transmission** within and across scales.
- **Cross-compartments effects** are stronger in the *Slow*-to-*Fast* direction, but **inter-compartment effects** are much larger in the fast scale $\Longrightarrow$ **Fast scale error must be controlled** before they affect the full system.
- Results confirmed numerically using MAOOAM and a strongly coupled EnKF.
- **Quasi-neutral modes are connected to the coupling** and must be accounted for in the coupled DA process.
- How would this picture change in the presence of spatial scale and amplitude difference?
- We develop a **combined DA-ML method** to build a **hybrid model made of a physics-based + data-driven surrogate of the unresolved scales**.
- The use of coupled DA is pivotal to the success as the coupled analysis used to learn the NN embodies essential coupling mechanisms.

# Bibliography

[1] Brajard J, A Carrassi, M Bocquet and L Bertino. *Combining data assimilation and machine learning to infer unresolved scale parametrisation.*, *Phil. Trans A of the Roy Soc.*, 2021. 379 (2194)

[2] Carrassi A, M Bocquet, J Demayer, C Grudzien, P Raanes, and S Vannitsem. *Data assimilation for chaotic dynamics*, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV), edited by Park, S. K. and Xu, L., Springer International Publishing, Switzerland, 2021

[3] Tondeur, M, A. Carrassi, S. Vannitsem and M. Bocquet: *On temporal scale separation in coupled data assimilation with the ensemble Kalman filter. J. Stat. Phys.*, **44**. 101171, 2020