Experiments for online estimation of model parameters for multidecadal climate reconstruction with the Community Earth System Model (CESM1.2)

Javier García-Pintado[1,*]

Adjoint workshop, Aveiro, 3rd July 2018

* and Pepijn Baker, André Paul, Matthias Prange and Michael Schulz

1. MARUM - Center for Marine Environmental Sciences, University of Bremen

Outline

Problem description: paleoclimate Data Assimilation

Assimilation methods

Synthetic test of observations at MARGO location

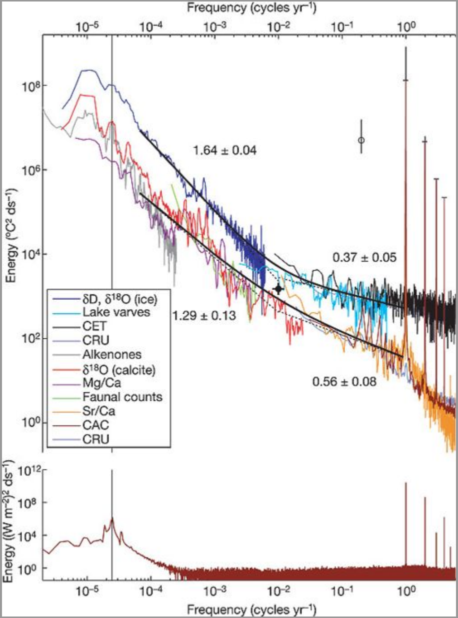
Conclusions

PAL MOD GERMAN CLIMATE MODELING INITIATIVE

Palmod is funded by the German Federal Ministry of Education and Science (BMBF) to understand climate system dynamics and variability during the last glacial cycle. Two specific topics are

- to identify and quantify the relative contributions of the fundamental processes which determined the Earth's climate trajectory and variability during the last glacial cycle
- to simulate with comprehensive Earth System Models (ESMs) the climate from the peak of the last interglacial (the Eemian warm period) up to the present, including the changes in the spectrum of variability

The second topic involves assimilation of paleoclimate proxy data [Eemian, Last Glacial Maximum, mid Holocene]



[Huybers and Curry (2006)]

Paleoclimate proxies: uncertainties

- **Proxy-observation locations** are sometimes uncertain [e.g. planktonic foraminifera].
- **Timestamps often long term means** often uncertain in width and time allocation [e.g. speleothems].
- **Proxy observation** $\sim$ **climate variable**: $\mathcal{H}(.)$ or Proxy system model (PSM). Deterministic? Statistical?

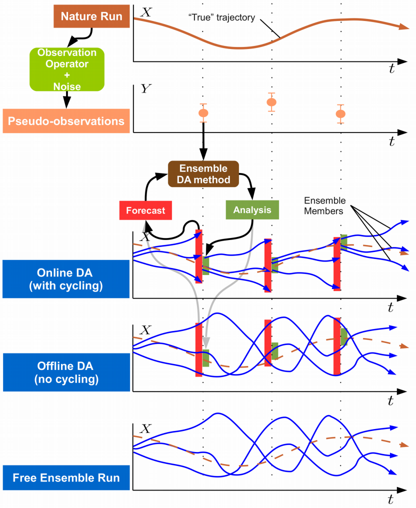
Paleoclimate modelling: uncertainties/DA computational constraints

- **Initial conditions** are uncertain. But it is generally assumed that for *long-term* paleoclimate evolution these (if reasonable) are a second-order effect with respect to other uncertainties
- **Boundary conditions** are also uncertain [e.g. ice sheets].
- **Model parameter [physics]** are uncertain to some degree, even for scientifically validated models
- **Adjoint** codes not available for ESMs
- **Computational requirements** are a strong constraint: CESM example [in HLRN3 HPC]:

```
 - b.e12.B1850C5CN.f19_g16.1850_equ.sta.000
Total PES active:  24*6=144
Overall Metrics:
Model Cost:  273.51 pe-hrs/simulated_year
Model Throughput:  12.64 simulated_years/day
```

  E.g.: $\{m = 50, \Delta t = 500 \text{ yr}\} \Rightarrow$ 7200 proc $\times$40 days

Paleoclimate assimilation: one current approach
Computational burden → the *offline* strategy



[Figure: Bijan Fallah]

Paleoclimate assimilation: a possible alternative approach
Uncertainty assumption for *online* DA in $\geq$ multidecadal paleoclimate reanalysis

- **Proposal:**
  1. Background error, after some time towards equilibrium, can be assumed to arise **mostly** from unknown dynamical parameters [memoryless PDF with respect to IC]. Use these as the only control variables into a **small** θ vector.
  2. θ normally would be the physics, but it can include parameterized initial condition uncertainty and forcings [including flux corrections], provided it is kept **small**.
- Thus the state vector is only this **small** set. Covariance can be explicitly dealt with.

Joint state-parameter estimation in the parameter space

- Within a DAW at a specific time t_k, $\theta_k \in \mathbb{R}^q$ is a vector of uncertain model parameters at, and we assume $\theta \equiv \theta_{k+1} = \theta_k$ within the DAW.
- For simplification, focusing on a t_k, we assume a state vector augmented with model parameters. The Kalman gain matrix rows $\mathbf{K}_{k_\theta}$, can be expressed e.g. as:

$$\begin{aligned} \mathbf{K}_{k\theta} &= \mathbf{P}_k \tilde{\mathbf{H}}_k^{\mathrm{T}} [\tilde{\mathbf{H}}_k \mathbf{P}_k (\tilde{\mathbf{H}}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1} \\ &= \mathbf{P}_{\theta\theta} \mathbf{G}_k^{\mathrm{T}} [\mathbf{G}_k \mathbf{P}_{\theta\theta} (\mathbf{G}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1}, \end{aligned} \tag{1}$$

Sensitivity estimation. What can we afford?

In (1) the covariance between a model parameter $\boldsymbol{\theta}_i$ and the observation y is expressed in each case as

$$\begin{aligned}(\mathbf{P}_k\tilde{\mathbf{H}}_k^{\mathrm{T}})[\theta_i, y] &= \sigma_{\mathbf{x}_{k_y}\boldsymbol{\theta}_i}\frac{\partial y}{\partial \mathbf{x}_{k_y}},\\ (\mathbf{P}_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathbf{G}_k^{\mathrm{T}})[\theta_i, y] &= \sum_{j=1}^{q}\sigma_{\boldsymbol{\theta}_j\boldsymbol{\theta}_i}\frac{\partial y}{\partial \boldsymbol{\theta}_j},\end{aligned} \qquad (2)$$

which, are identical as $\sigma_{\mathbf{x}_{k_y}\boldsymbol{\theta}_i} = \sum_{j=1}^{q}\sigma_{\boldsymbol{\theta}_j\boldsymbol{\theta}_i}\frac{\partial \mathbf{x}_{k_y}}{\partial \boldsymbol{\theta}_j}$.

- **ensemble sensitivity**: E.g., the **batch-EnRML** (Chen & Oliver, 2012), estimates and ensemble-based average sensitivity matrix $\bar{\mathbf{G}}_l$ at l iteration $(\Delta\mathbf{Y}^l = \bar{\mathbf{G}}_l)\Delta\boldsymbol{\theta}^l$;
- **OAT perturbation-based**: $\bar{\mathbf{G}}_l$ estimated from sensitivity experiments ($\boldsymbol{\theta}_i$ samples from conditional density)

Some options to deal with non-linearity

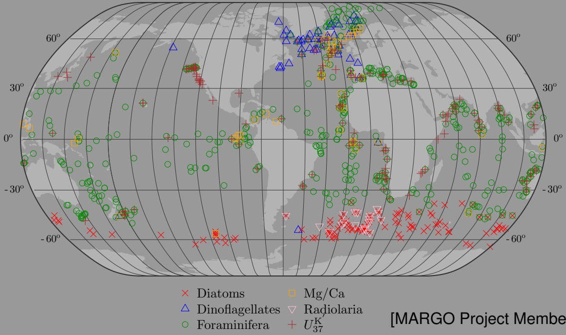
- Gaussian Anamorphosis [GA] (Simon & Bertino 2003)
- Iterations: **IKF** (Jazwinski 1970; Bell & Cathey, 1993), **IKS** (Bell 1994), **MLEF** (Zupanski 2004), **EnRML** (formulated in the parameter space; Gu & Oliver, 2007), **batch-EnRML** (Chen & Oliver, 2012), **IEnKF** (Sakov et al., 2012), **IEnKS** (Bocquet & Sakov, 2013,2014; the latter joint state+parameters), **IEnKF+Q** (Sakov et al. 2018)

Our simple "affordable" approach

- Assume model uncertainty encapsulated in a small θ
- Formulations as function of $\bar{\mathbf{G}}_l$, estimated form OAT perturbation experiments. Two (combinable) iterated schemes:
- **pIKS** perturbed-parameters Iterated Kalman Smoother. A Gauss-Newton akin to an asynchronous IKF
- **pMKS** perturbed-parameters multistep Kalman Smoother. Regularization based on deflation of $\mathbf{R}$.

Experimental setup

- CESM1.2.2: COMPSET B1850CN [ice sheets as BC]
- f45_g37 FV: [atm ∼4 deg reg | ocean/sea ice: displaced pole ∼3 deg]
- Preindustrial conditions in equilibrium [after 1200 yr spin up]
- Then run control [*truth*] for 100 yr
- Ensemble [m=60] with perturbed physics (clouds microphysics, ocean diffusivity, sea ice albedos. . .) branch from the same initial conditions as the control run. Then ETKF, ETKF-GA, pIKS, pMKS
- Observations: **MARGO**-like SST 20 yr mean at the end of the 100 yr



[MARGO Project Members, Nature (2009)]

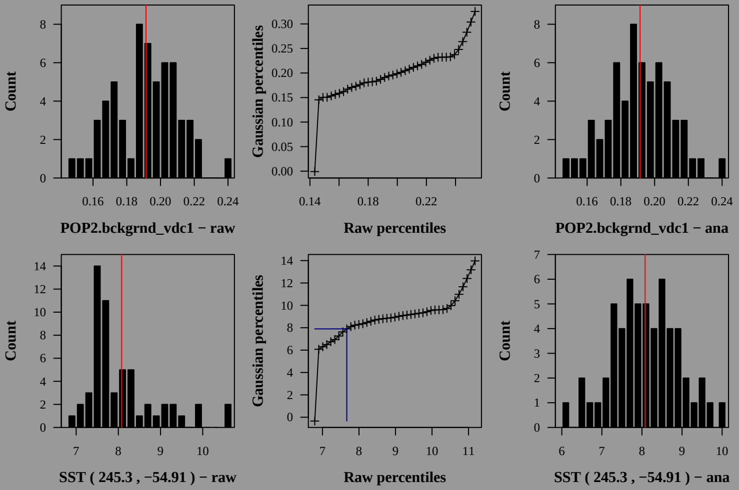
CESM experiments: parameter definition

Experimental setup

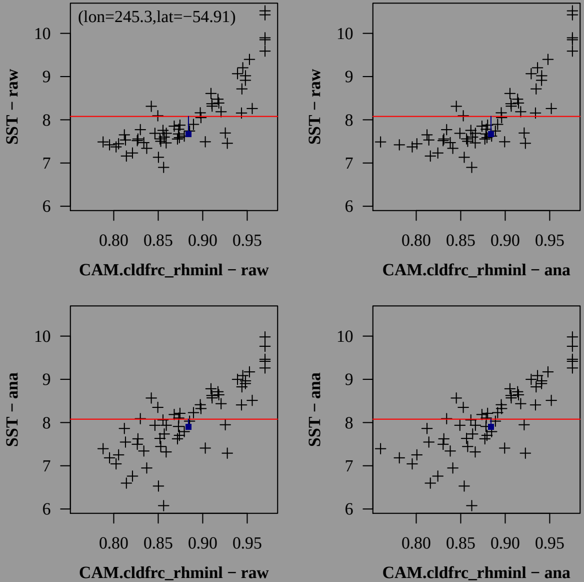
COMP.name[1]	*Description*	*Units*
CAM.cldfrc_rhminh	minimum relative humidity for high stable cloud formation	[-]
CAM.cldfrc_rhminl	minimum relative humidity for low stable cloud formation	[-]
CAM.ch4vmr	greenhouse gases, CH_4 volume mixing ratio	ppb
CAM.co2vmr	greenhouse gases, CO_2 volume mixing ratio	ppm
CAM.zmconv_c0_lnd	autoconversion coefficient over land in ZM deep convection	[-]
CAM.zmconv_c0_ocn	autoconversion coefficient over ocean in ZM deep convection	[-]
CAM.zmconv_ke	evaporation efficiency in ZM deep convection	[-]
POP2.bckgrnd_vdc1	KPP mixing: background vertical diffusivity (Ledwell)	cm^2s^{-1}
POP2.hmix_gm_nml.ah	Gent-Williams isopycnic tracer diffusion (Redi)[2]	cm^2s^{-1}
POP2.freshwater_gis	freshwater influx homogeneously distributed around Greenland	Sv

[1] *COMP:name* CESM component and parameter name.
[2] POP2.hmix_gm_nml.ah_bolus constrained to equal POP2.ah.hmix_gm_nml.ah.

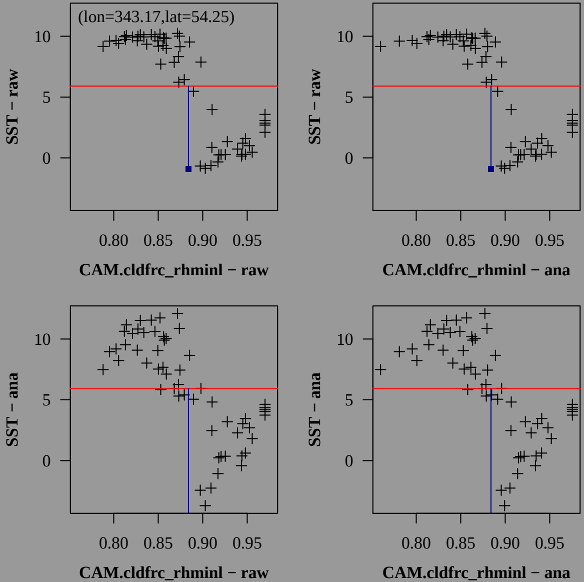
Gaussian Anamorphosis



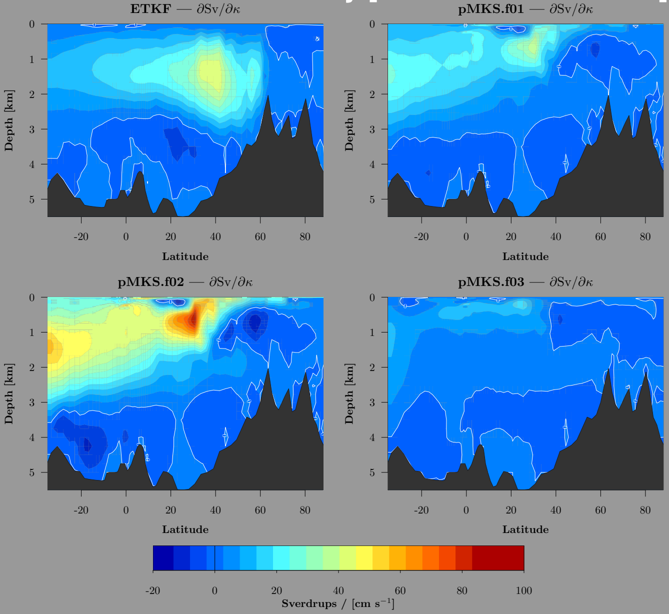
How nonlinear are multidecadal sensitivities?
Example observation at Equatorial Pacific



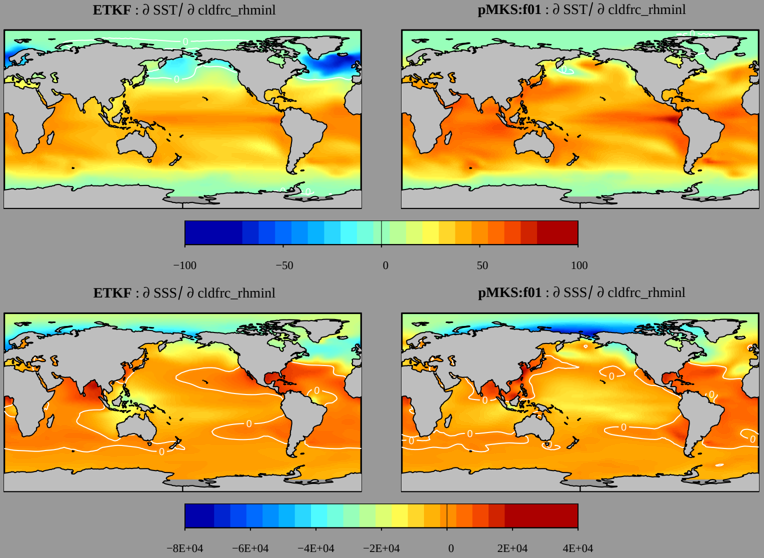
How nonlinear are multidecadal sensitivities?
Example observation at North Atlantic



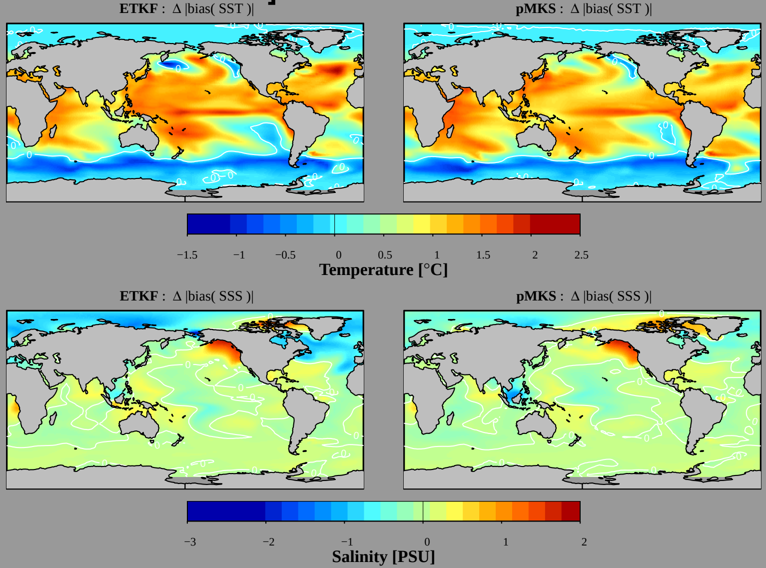
CESM: multidecadal sensitivity [AMOC ∼ ocean vdiff]



CESM: multidecadal sensitivity [SST,SSS $\sim$ moisture threshold for low clouds]



CESM: bias reduction given observations [SST & SSS moisture]



CESM: cost function values for MARGO synthetic experiment

COMP.name	$\mathbf{x}^t$	$\mathbf{x}^b$	$\mathbf{x}^a$ – $ETKF_{60}$	$\mathbf{x}^a$ – $ETKF_{60}$+GA.y	$\mathbf{x}^a$ – pEKS	$\mathbf{x}^a$ – pMKS.f03	pIKS.f03+br
$\mathcal{J}_y(\boldsymbol{\theta})$		373.39	64.95	61.23	91.81	50.24	48.88
$\mathcal{J}(\boldsymbol{\theta})$		373.39	66.43	66.83	93.13	51.85	55.20

[1] Units as described in Table 1.

Note: PIKS, pMKS required 34 integrations, ETKF required 61

Conclusions

- Most of the current work in paleoDA is actually on the forward operators (Paleoclimate proxy model [PSM]). Moreover, there is an open debate about how model and observations should be compared quantitatively, which is the basis for developing **H**.
- As models converge towards their climatology after some time, it seems the perturbed physics is a) a posible way of recovering the power spectrum showed by paleoproxy observations and b) a possible mechanism to create the background statistics needed for the assimilation
- Our tests indicate that $\geq$ multidecadal analysis of past climates (no background covariance at hand) the iterative methods based on OAT perturbations for each degree of freedom (assumed low-dimensional) beat ETKF (ensemble sensititivy) from a computational and statistical point of view.