Initialization and Forecast Strategies for Seamless Prediction

- Setting the scene
- Initialization shock, forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Ensemble generation for ocean initial conditions

Magdalena A. Balmaseda

The basis of forecasts beyond weather time scales

System with multiple time scales

From the fast component (atmospheric) point of view is a boundary problem

Predictability of the second kind or "loaded dice"

Forcing exterted by boundary conditions changes the atmospheric circulation, modifying the large scale patterns of temperature and rainfall, so that the probability of occurrence of certain events deviates significantly from climatology.

Which boundary conditions?:

SST, soil moisture, snow, sea-ice, radiative forcing, stratosphere

In Earth System models these slow components are prognostic –no longer boundary.

For the slow component perspective, S2S prediction is an initial value problem

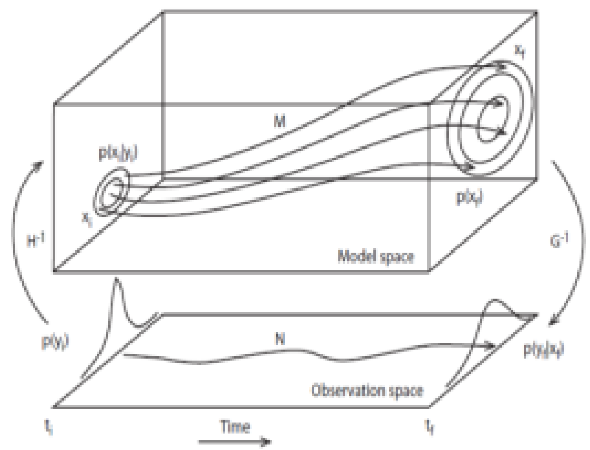
Predictability of the first kind: The slow components need to be initialized

Initialization: an essential stage in the forecasting process

2) Propagating information and uncertainty into the future: *Forecast model*

1)Initialization
Data Assimilation

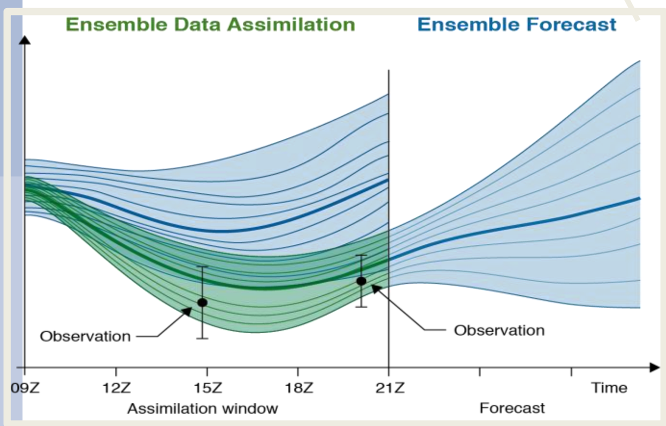
$$p(x_i|y_i) = \frac{p(y_i|x_i)p(x_i)}{p(y_i)}$$



3) Calibration
Forecast
Assimilation

$$p(y_f|x_f) = \frac{p(x_f|y_f)p(y_f)}{p(x_f)}$$

$$J_{x|y} = (x - x_b)^T B^{-1}(x - x_b) + (y - Hx)^T R^{-1}(y - Hx)..$$

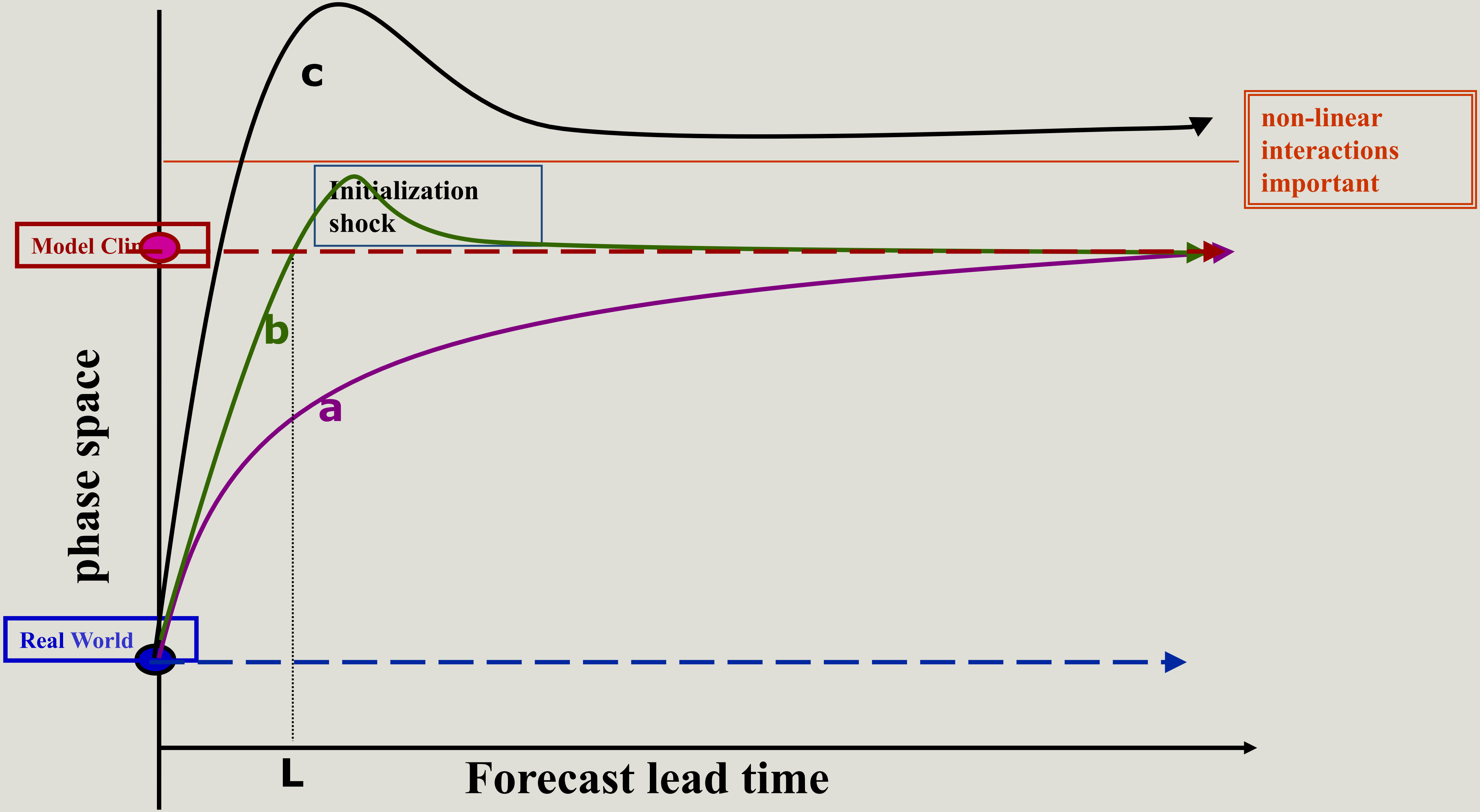


Initialization distinguishes a forecast from a simulation/projection.

But it may not be perfect:

- Models are not perfect
- Observations are insufficient
- The data assimilation (translator) has deficiencies

Initialization shock – drift - skill



What causes initialization shock?

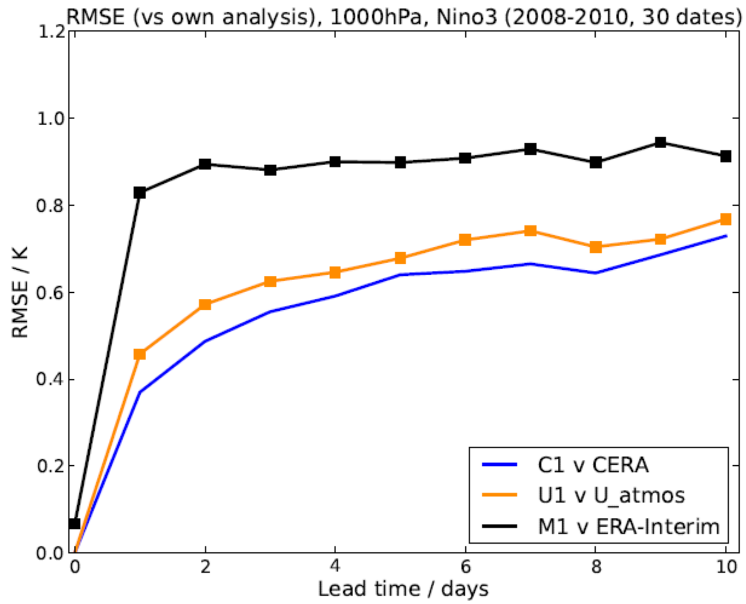
Initialization shock implies that the data assimilation process has created imbalances in the initial condition, not supported by the model physical constrains. The observation information is rapidly lost via adjustment processes that deteriorate skill.

Possible reasons for initialization shock

1. **Deficient data assimilation**
   - Example: Insufficient physical constrains
   - Example: Data assimilation forces scales that the model is not able to represent.
   - Example: Too much weight to observations and poor quality control leads to erroneous observations being assimilated.

2. **Initial conditions produced with a different model than the used for the forecast.**
   - Separate initialization of ocean and atmosphere
   - Different model cycles

Initialization shock: forecast error growth depends on initialization



Experiment: same forecast model with 3 different initialization

Initialization

1) Uncoupled: different models

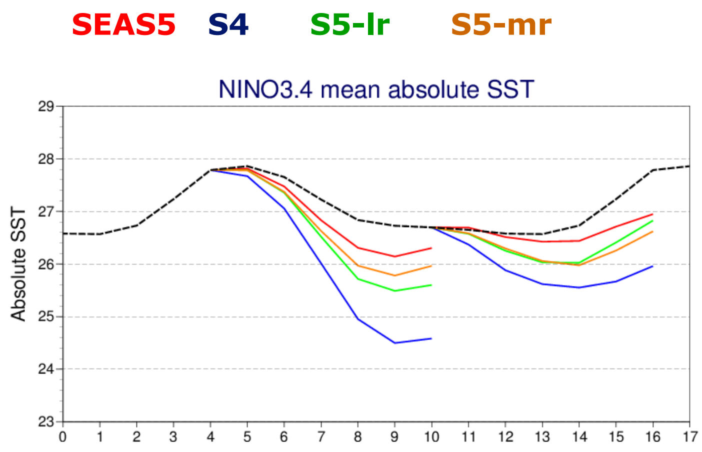
2) Uncoupled: Same models

3) Coupled

Slowest Forecast Error Growth: coupled initialization
Fastest Forecast Error Growth: Ini Model .ne. FC. Model and uncoupled initialization

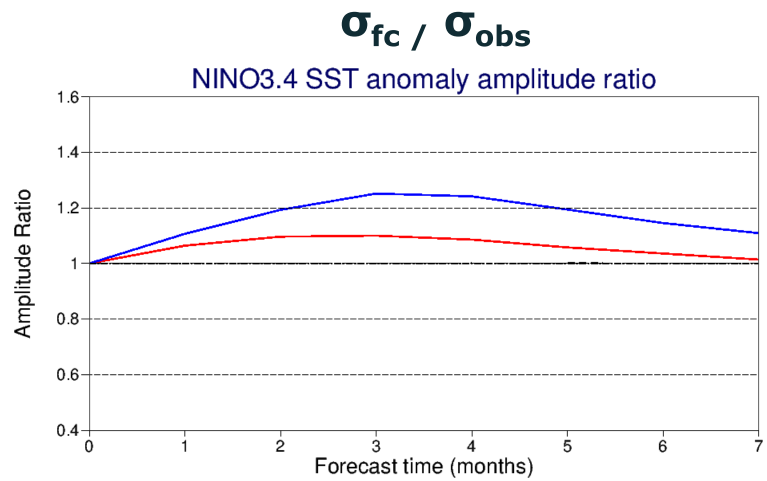
From Mulholland et al, 2015 MWR

Drift and bias: a seasonal forecast example



Fc drift in the mean: first moment of distribution (bias)

- Bias depends on model resolution
- Bias depends of lead time
- Bias depends on the phase of seasonal cycle



Fc drift in the variance (the second moment)

- The interannual variability is affected
- Non linearity: links mean-variability

Note 1: basic a-posteriori bias correction only valid if Biases are stationary and system is linear

Note 2: One common perception is the drift only depends on the model. But it also depends on initialization (e.g. ini. shock)

Initialization Problem: Production of Optimal I.C.

Optimal Initial Conditions: those that produce the best forecast.

Need of a metric: lead time, variable, region (i.e. subjective choice)

In 4D-var the metric are the atmospheric forecasts errors at short lead time (6-12h)

This does not guarantee optimal forecast at the extended or seasonal range.

There is not criteria to optimize the other Earth System Components: ocean, land, …

Initial conditions should represent accurately **the state of the real world and project into the model attractor**, so the model is able to evolve them.

Difficult in the presence of model error

Initialization Shock and forecast drift

Practical requirements arising from calibration:

Stationary forecast errors

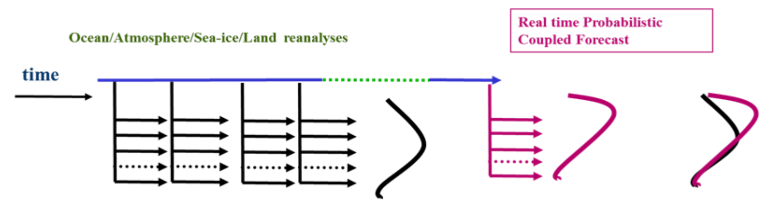
Consistency between re-forecasts and real time fc ***Need for historical reanalysis***

Adequate representation of uncertainty

Additional requirement: Reanalyses to initialize Reforecasts

Applications

- Calibration of forecasts from days to decades
- Detection and prediction of extreme events
- Skill assessment
- Reanalyses used for monitoring
- Predictability and evaluation studies



Reforecast errors should be consistent with the errors of the real time forecast.

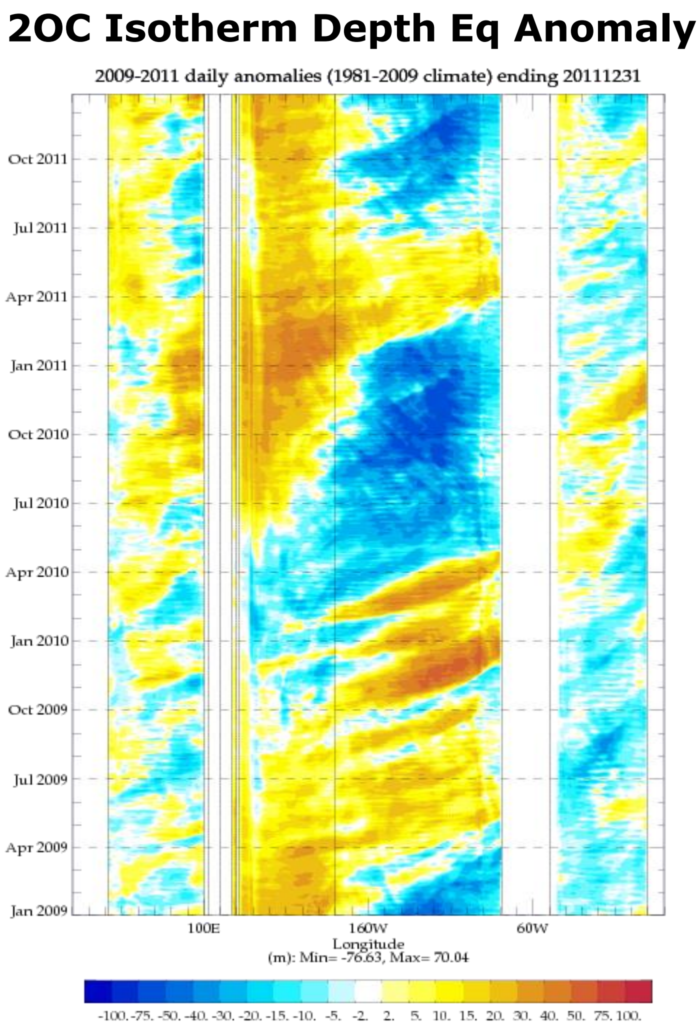
Requirements

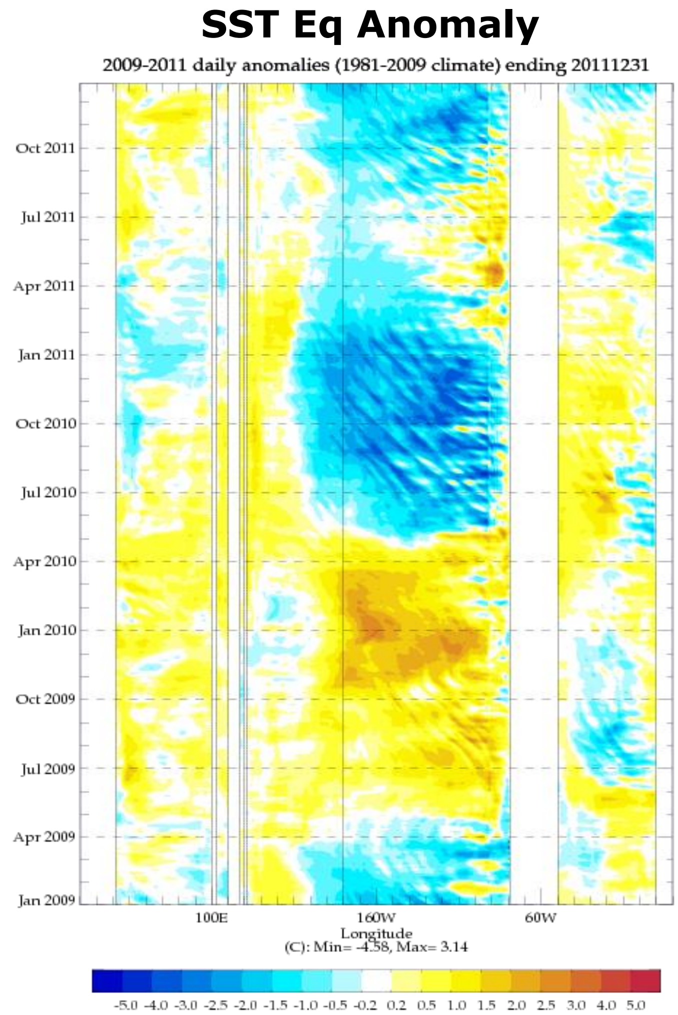
- **Consistence with real time forecasts, so calibration makes sense**
- **Temporal consistency and faithful representation of a wide range of time scales**:
  diurnal cycle-intraseasonal-seasonal-interannual-decadal variability –trends
  This is challenging in the presence of model error and a changing observing system
- **Accurate and physically balanced estimate estimation and associated uncertainty.**
  - So observational information can be propagated into the forecasts.
  - So relevant processes can be reliably quantified
- **As far back as possible**

ECMWF

- Setting the scene
- Initialization shock, forecast drift and calibration
- **Example: initialization of the ocean**
- Approaches to initialize Earth System predictions
- Ensemble generation for ocean initial conditions

Need to Initialize the slow components: The ocean example





Emphasis on the thermal structure of the upper ocean
Predictability is due to higher heat capacity and **predictable dynamics**

Information needed to initialize the ocean

Ocean model + Atmospheric fluxes from atmospheric reanalysis

AND

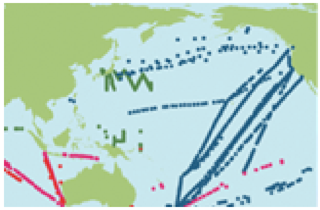
Ocean Observations **+** **Data Assimilation methods**

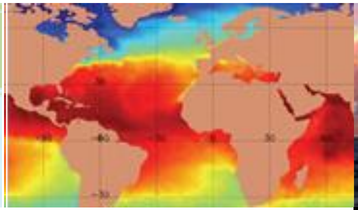
- SST
- Subsurface ocean information

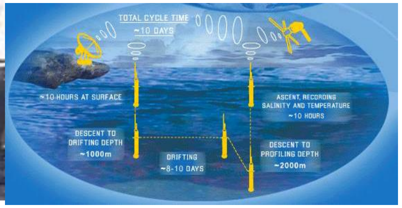
Time evolution of the Ocean Observing System

1982 **1993** **2001**

XBT's 60's **Satellite SST** **Moorings/Altimeter** **ARGO**





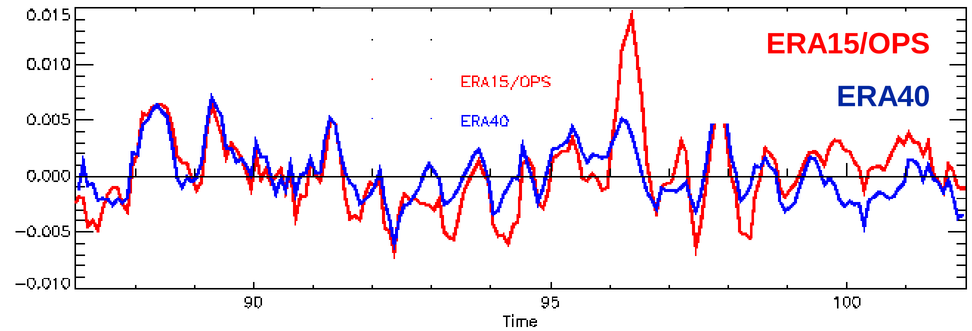


Need for data assimilation: Uncertainty in Surface Fluxes

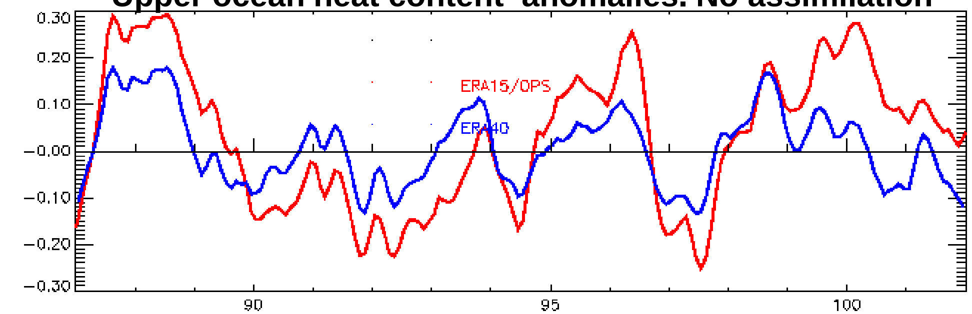
- Large uncertainty in wind products lead to large uncertainty in the ocean subsurface
- The possibility is to use additional information from ocean data (temperature, others…)

Equatorial Atlantic

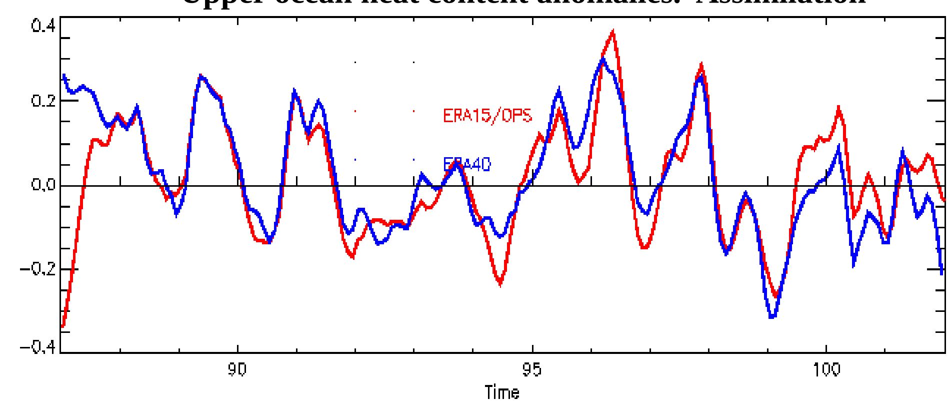
Taux anomalies



Upper ocean heat content anomalies. No assimilation



Upper ocean heat content anomalies. Assimilation

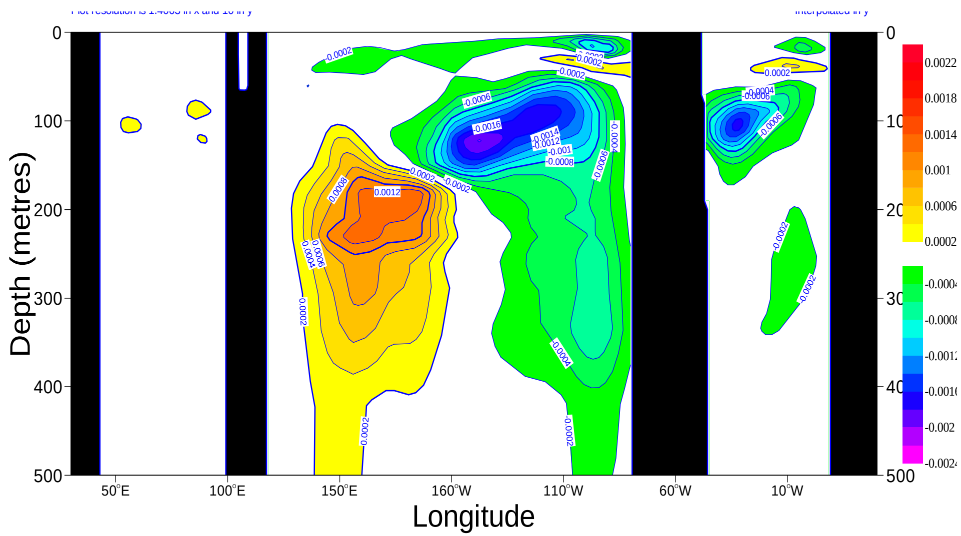


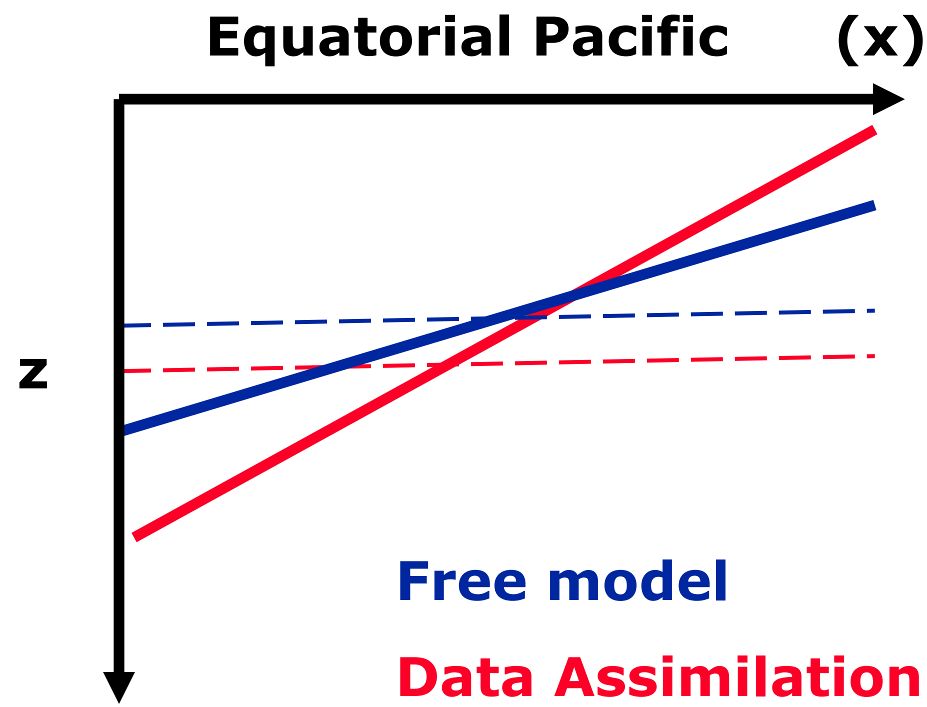
- **Questions:**

  1. Does assimilation of ocean data constrain the ocean state? **YES**
  2. Does the assimilation of ocean data improve the ocean estimate? **YES**
  3. Does the assimilation of ocean data improve the seasonal forecasts. **YES**

Need for data assimilation: Correction of model error

Mean Assimation Temperature Increment

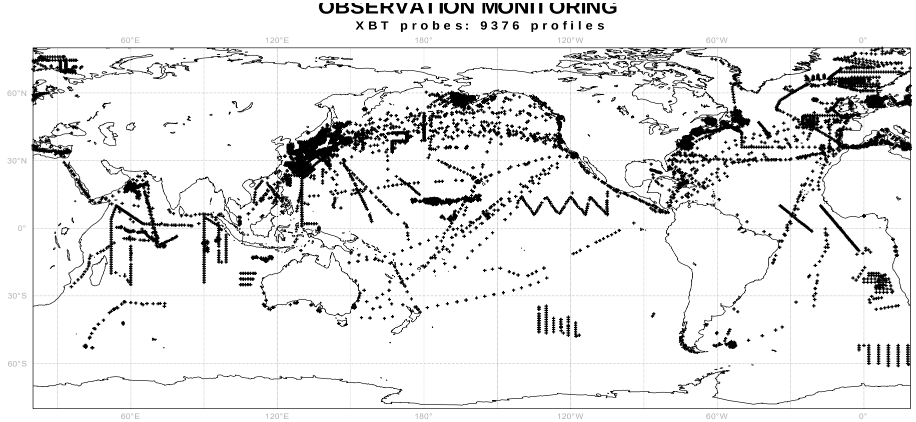




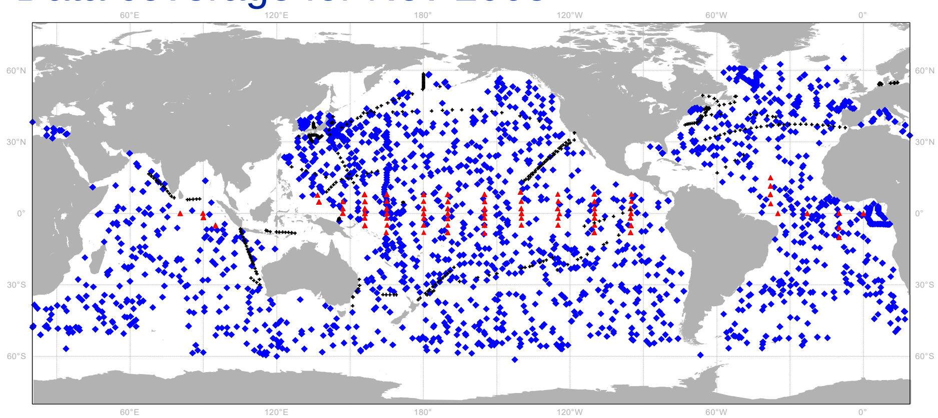
Data assimilation corrects the slope and mean depth of the equatorial thermocline

The changing observing system

Data coverage for June 1982

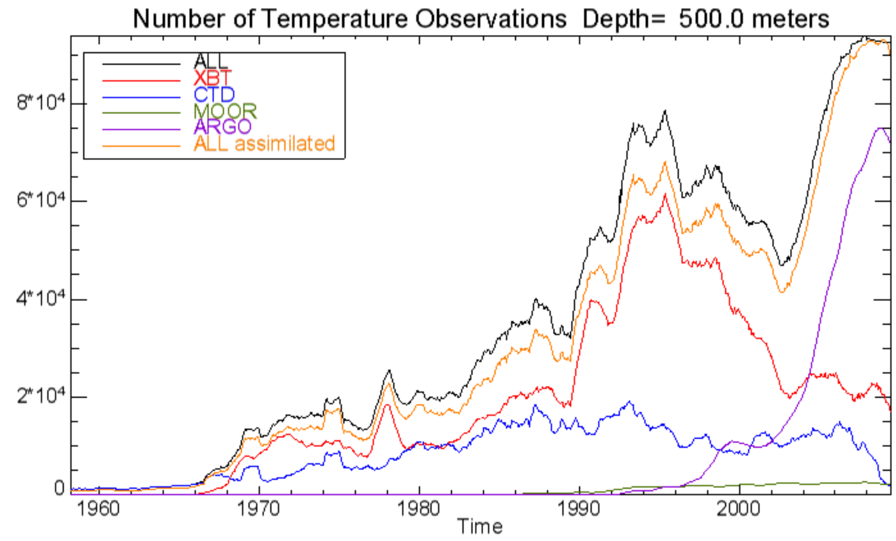


Data coverage for Nov 2005



Changing observing system is a challenge for consistent reanalysis

especially in the presence of systematic model error

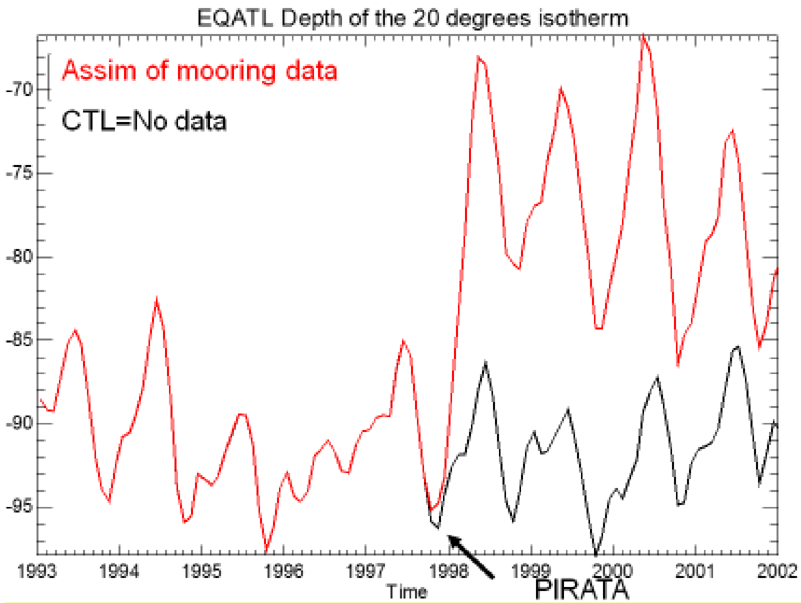


▲Moorings: SubsurfaceTemperature

◊ ARGO floats: Subsurface Temperature and Salinity

+ XBT : Subsurface Temperature

Changes in observing system induce spurious variability in ocean reanalyses



Ocean had too deep thermocline. The assimilation of PIRATA observations corrects this error, but results on spurious variability in the time series

This can be alleviated by including a bias correction term to the model tendencies, which extrapolates the observational information into the past

$$\mathbf{b}_c = \bar{\mathbf{b}} + \mathbf{b}'_c$$
$$\mathbf{b}'^f_c = \alpha \mathbf{b}'^a_{c-1} + \boldsymbol{\varepsilon}$$
$$\mathbf{b}'^a_c = \alpha \mathbf{b}'^a_{c-1} + \mathbf{A}\delta x_c$$

The bias correction has two terms

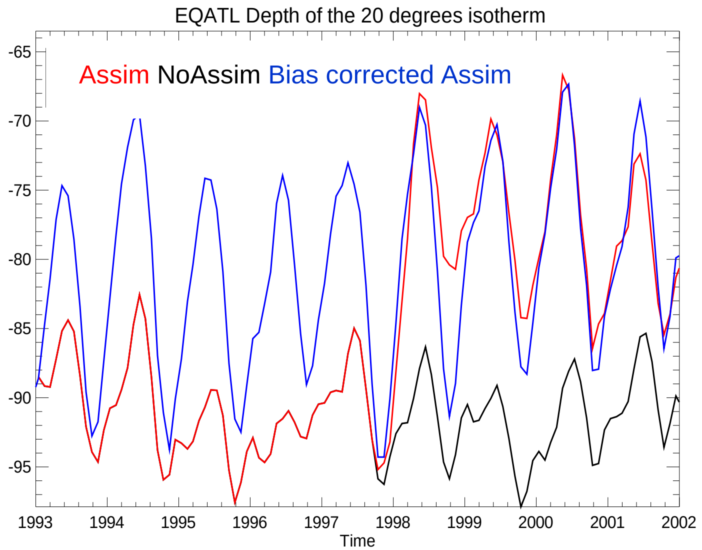
Б estimated offline from the well observed period
b’ estimated online from assimilation increments

The explicit treatment of model bias in assimilation also allows imposing different dynamical balances for the increment and bias,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{b}^f + \mathbf{K}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{b}^f)]$$
$$\mathbf{b}^a = \mathbf{b}^f + \mathbf{L}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{b}^f)]$$

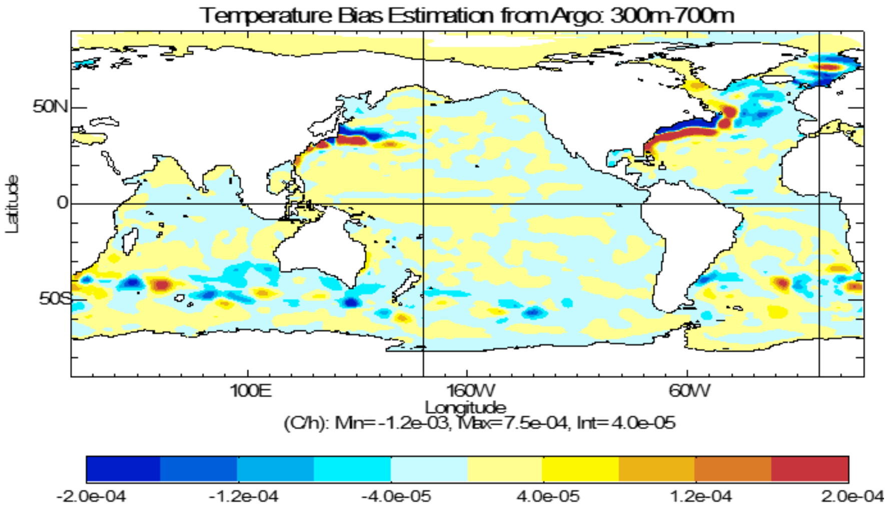
Balmaseda et al 2007

Importance of treatment of model error in ocean data assimilation



The extrapolation to the past of the PIRATA information alleviates the problem of spurious temporal variability

The observing system will always be changing:
To achieve temporal reliable reanalyses it is important to extrapolate the observation information into the past.

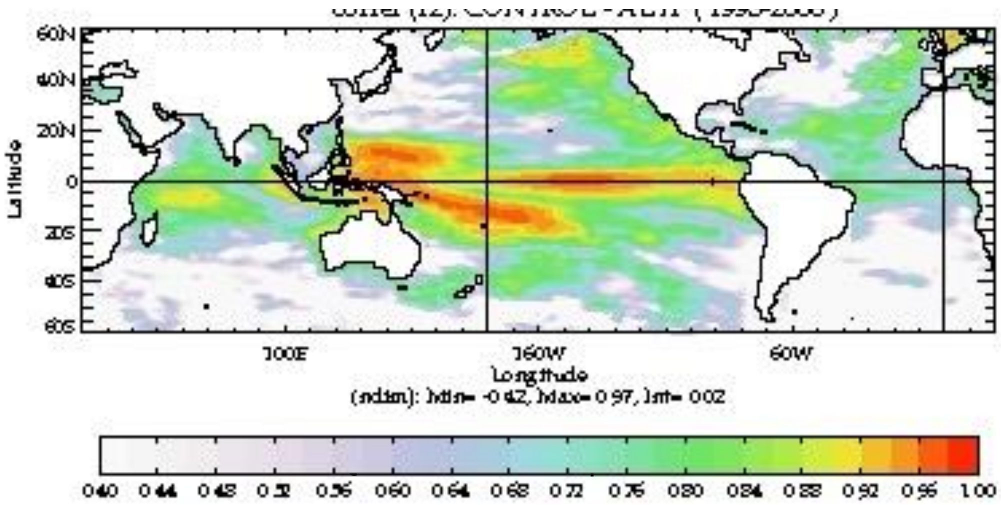


This is an important difference with respect to the atmos data assimilation, where FG is assumed unbiased

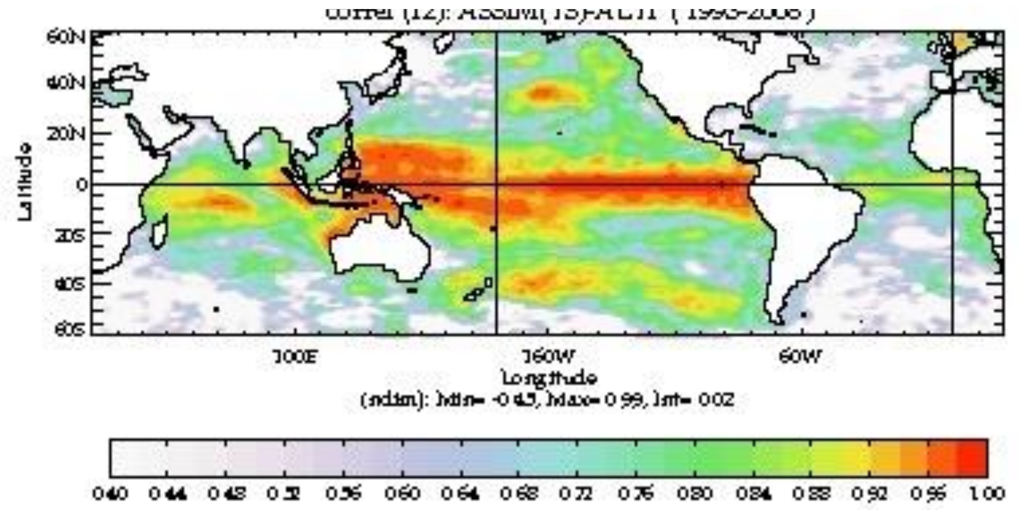
Balmaseda et al 2007

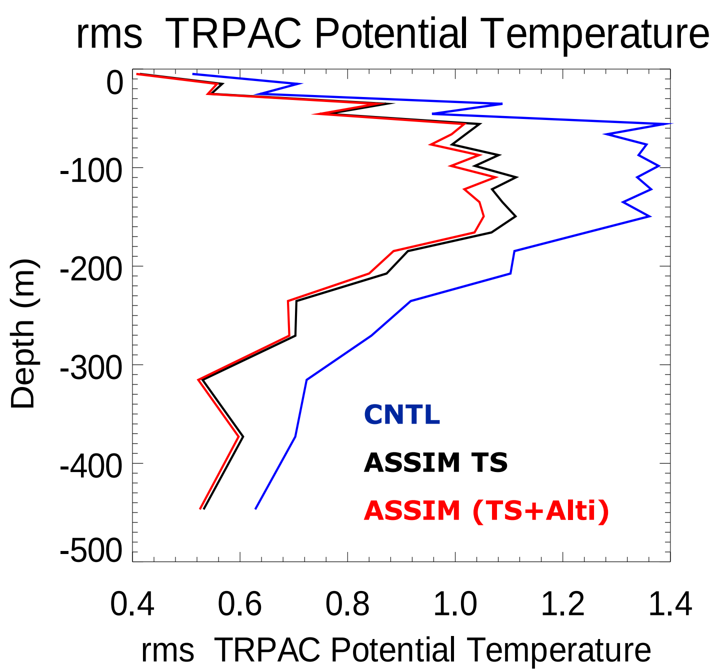
DA+bias correction: Improved temporal variability

CNTL: NoObs

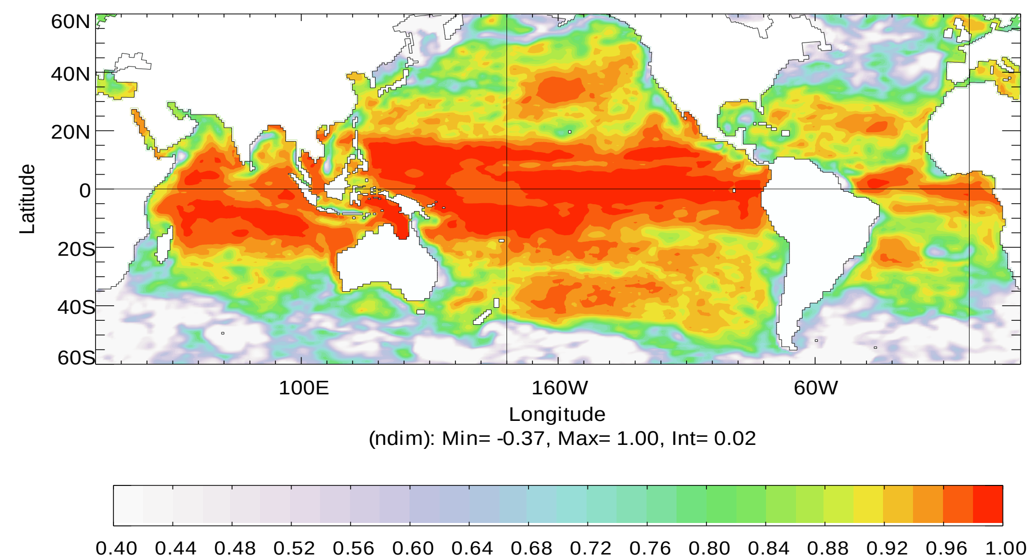


Assimilation of T+S

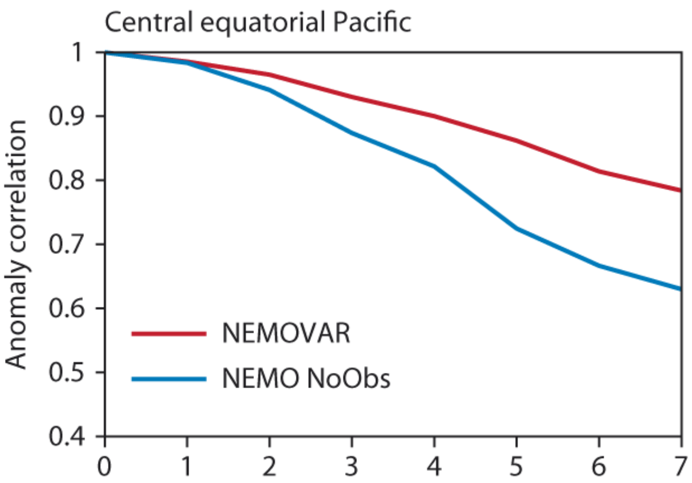




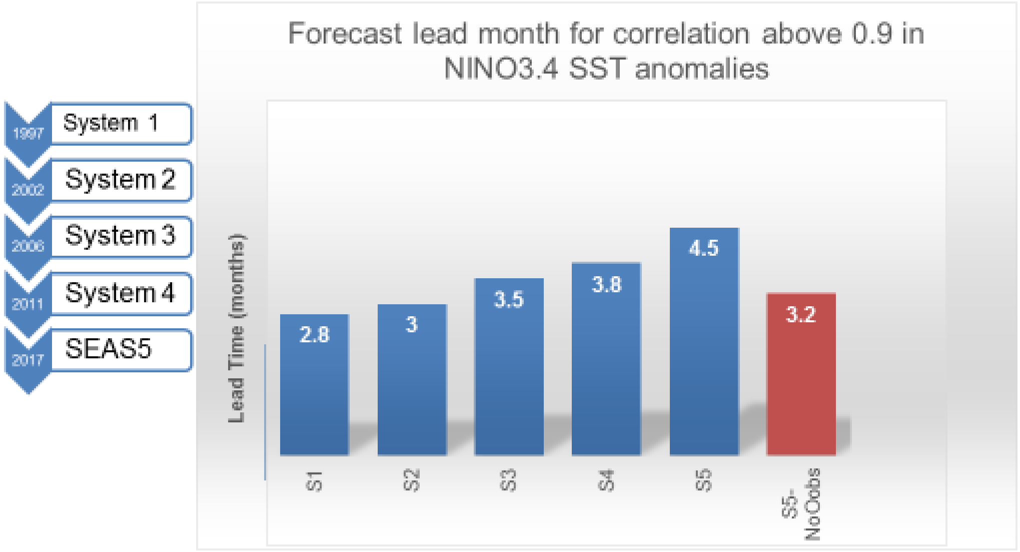
Assimilation of T+S+Alti



Data Assimilation improves the forecast skill



Contribution of Ocean Data Assimilation to 20 years of Progress on ENSO prediction at ECMWF



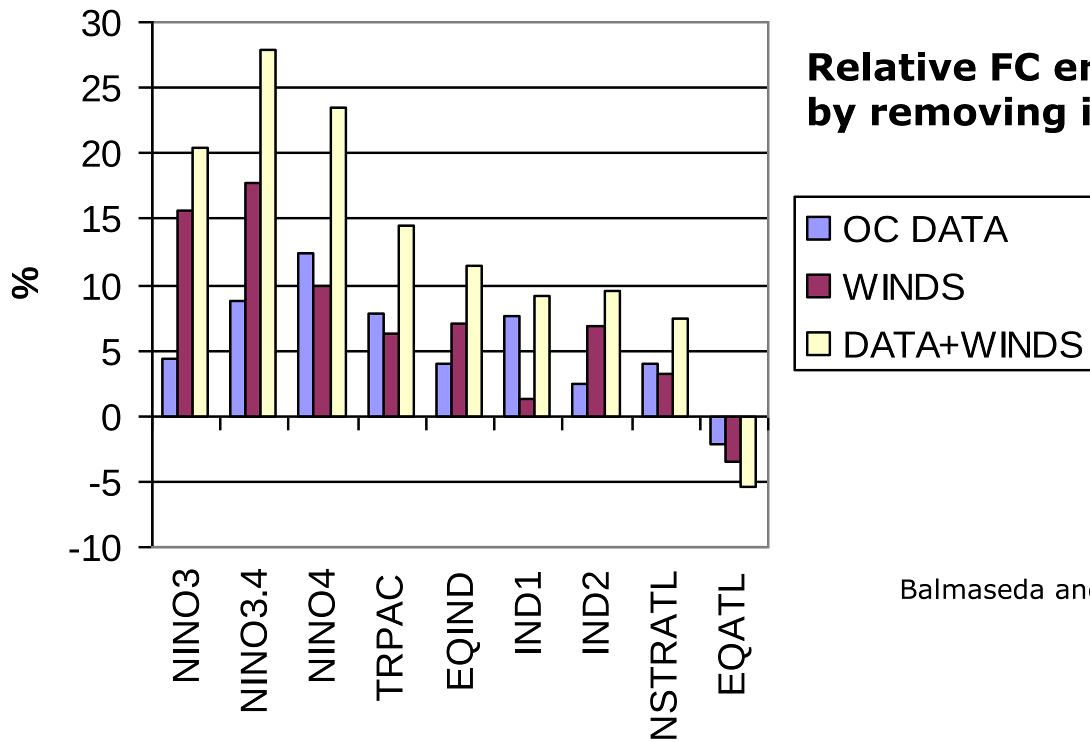
Quantifying the value of observational information

Experiments

SST (SYNTEX System Luo et al 2005, Decadal Forecasting Keenlyside et al, 2008)

SST+ Atmos observations (fluxes from atmos reanalysis)

SST+ Atmos observations+ Ocean Observations (ocean reanalysis)



Relative FC error increase on by removing information

Balmaseda and Anderson 2009, GRL

The outcome may depend on the coupled system

In a good system information may be redundant, but not detrimental.

- Setting the scene
- Initialization shock, forecast drift and calibration
- Example: initialization of the ocean
- **Approaches to initialize Earth System predictions**
- Ensemble generation for ocean initial conditions

Perceived Paradigm for initialization of coupled forecasts

Real world **Model world**

<u>Medium range</u>

Full initialization: Being close to the real world is perceived as advantageous. Model slowly drift to its own mean state.

<u>Seasonal?</u>

<u>Decadal or longer</u>

Anomaly initialization: Avoid forecast drift by initializing around the model mean state

<u>At first sight, this paradigm would not allow a seamless prediction system.</u>

Anomaly initialization is not the same as model attractor initialization

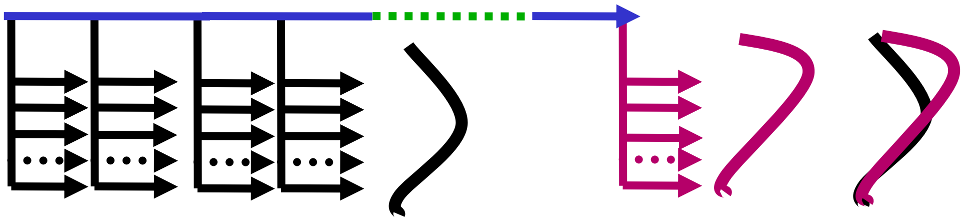
So far we have seen some Caveats of Full Initialization:

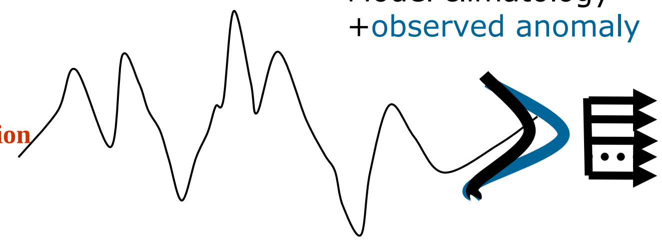
Initialization shock resulting from unbalanced states

Non-linearities and non-stationarity can sometimes render the a-posteriori calibration invalid

Full Initialization

Anomaly Initialization





As Medium range but:
<u>Ocean Model bias taken into account during DA.</u>

A posteriori calibration of forecast is needed. Calibration depends on lead time.

If uncoupled initialization: the model during the initialization is different from the forecast model. Bias correction estimated during initialization can not be applied during the forecasts

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{b}^f + \mathbf{K}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{b}^f)]$$

$$\mathbf{b}^a = \mathbf{b}^f + \mathbf{L}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{b}^f)]$$

Original purpose: to avoid expensive calibration. The model climatology does not depend of forecast lead time. Cheaper in principle than reforecasts in full initialization.

But reforecasts are still needed for skill estimation.
And calibration is still needed in practice.

Definition of the anomaly? It needs an existing reanalyses from which compute anomalies

Acknowledgment of existence of model error during initialization.

Model error is not corrected ("bias blind algorithm"):

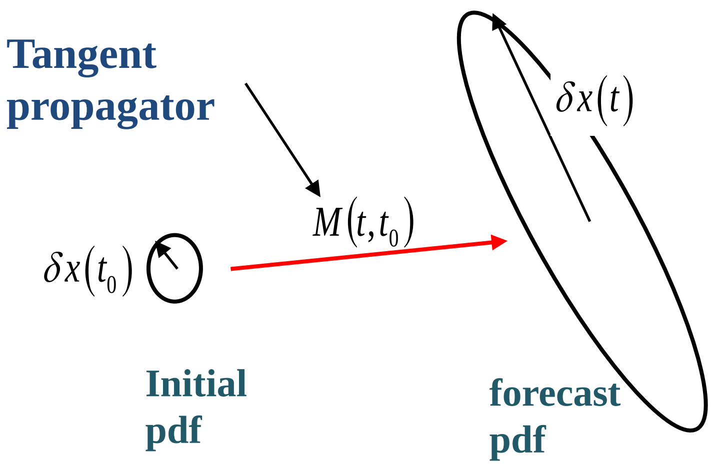
$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}[(\mathbf{y} - \overline{\mathbf{y}}) - \mathbf{H}(\mathbf{x}^f - \overline{\mathbf{x}})]$$

Balmaseda, JMR, 2017

- Setting the scene
- Initialization shock, forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- **Ensemble generation for ocean initial conditions**

Are Singular Vectors a valid approach for operational seasonal forecasts?

Medium Range: Singular Vectors



$$\Rightarrow M^{*}M\,\delta x(t_0) = \lambda\,\delta x(t_0)$$

We need the TL& Adjoint of the full coupled model is required.

BUT…

The linear assumption would <u>fails </u>for the atmosphere at lead times relevant for seasonal (~>1month).

Alternatives

1. **Other approaches for optimal sampling of initial condition uncertainty:**
   - **Breeding Vectors (NASA, BoM. Not shown here)**
   - **SV using Generalized Linear Propagators**
2. **Sample known i.c. uncertainties, without considering optimality**

Uncertainty in initial conditions may not be the dominant source of error

Generalized Singular Vector Problem (I)

Generalized Linearized Propagator (not necessary tangent linear)

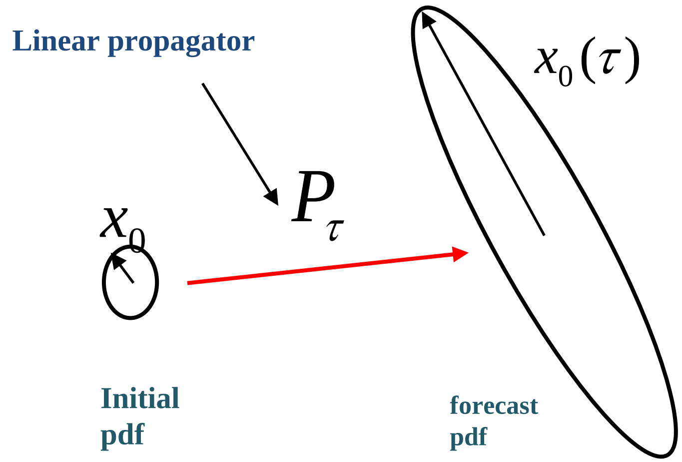
$$\mathbf{x}\ (\tau) = \mathbf{P}_{\tau}\mathbf{x}_0$$

Given a final **N** and initial norm **L**, the growth in **x** can be measured by

$$A(\tau) = \frac{\mathbf{x}(\tau)^{\mathrm{T}}\mathbf{N}\mathbf{x}(\tau)}{\mathbf{x}_0^{\mathrm{T}}\mathbf{L}\mathbf{x}_0} = \frac{\mathbf{x}_0^{\mathrm{T}}\mathbf{P}_{\tau}^{\mathrm{T}}\mathbf{N}\mathbf{P}_{\tau}\mathbf{x}_0}{\mathbf{x}_0^{\mathrm{T}}\mathbf{L}\mathbf{x}_0},$$

Optimal perturbations are those that maximize **λ**

$$\mathbf{P}_{\tau}^{\mathrm{T}}\mathbf{N}\mathbf{P}_{\tau}\mathbf{x}_0 = \lambda\mathbf{L}\mathbf{x}_0$$



Different ways of estimating the Linear Propagator P(τ)

I. Empirical (or Inverse modelling): basically a regression
II. A simplified linear dynamical model (equilibrium atmosphere rather than tangent linear)
III. A hybrid system: Ocean GCM coupled to a simplified atmosphere

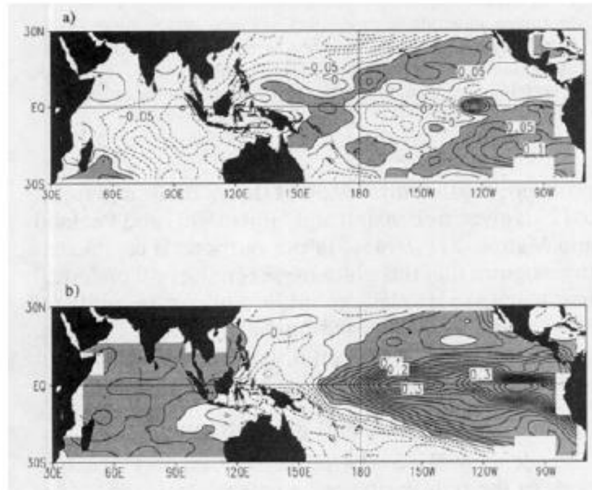
Generalized Singular Vector Problem (II)

Linear Propagator estimated empirically via regression model (Inverse modelling)

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \xi,$$

- **From temporal records of observations**
  von Storch and Xu 1990 MJO (POPs Principal Oscillation Patterns)
  Blumenthal 1991 ENSO
  Penland and Sadershmuck 1995, ENSO (inverse modelling)

- **From temporal records of model evolution**
  Xue et al 1997a,b; Fan et al 1999 ENSO
  Hawkins and Sutton 2009 Decadal Prediction AMOC



Initial SST

Final SST

Penland and Sadershmukh 1995

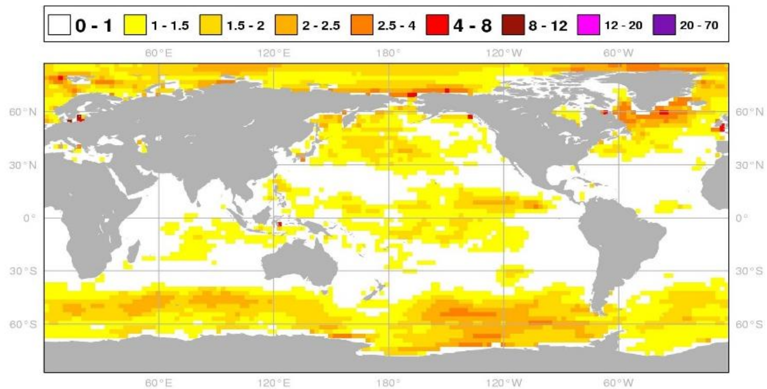
This approach is based on temporal sampling of existing timeseries: Difficult to capture flow dependence or errors of the day.
Judgement: not appropriate for ensemble generation in operational systems.

These are powerful tools for a-posteriori diagnostics of ensemble statistics for evaluation of forecasts;. Ensemble Sensitivity. Magnusson 2017 QJRMS

Representing Known Ocean Analysis Uncertainties at ECMWF

2002 Uncertainty in wind stress and SST

SDV of Wind Stress Perturbation

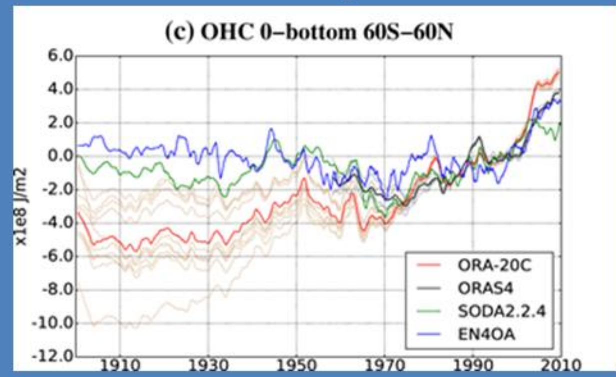


- Create data base with errors in the monthly anomalies of wind stress, arranged by calendar month:
- Random draw of monthly perturbations, applied during the ocean analyses.
- **Create a centered ensemble of 5 reanalysis is constructed symmetric wind perturbations -P2 -P1 0 P1 P2**

2011 Uncertainty on ocean reanalyses spin-up
Uncertainty in missing processes
(sea ice in SEAS4 by analogues)

2016 Other surface fluxes
Observation representative errors

Sampling Spin-Up uncertainty in ocean reanalyses



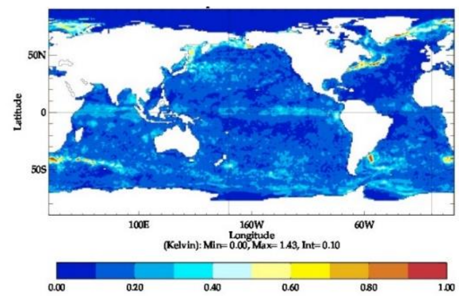
From de Boisseson et al 2017

Uncertainty representation in ORAS5

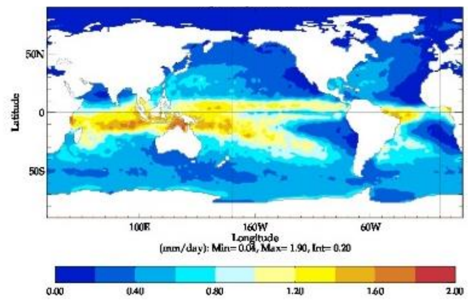
Multivariate - Updated data sets – 2 temporal scales – Multiple uncertainty sources

Still conservative: it does not sample error in the mean.

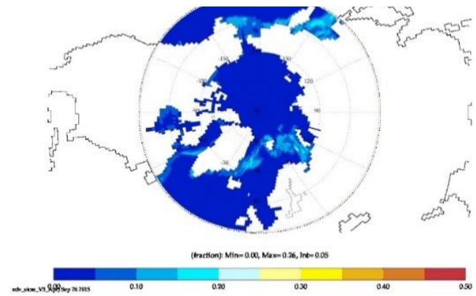
SDV SST SE New (V3)



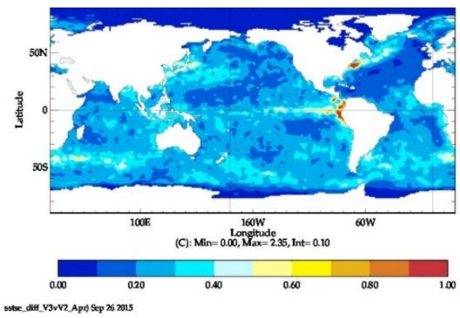
SDV PME



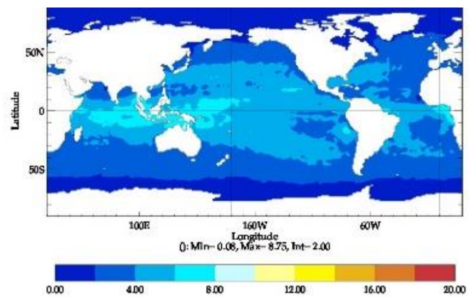
SDV SIC Apr



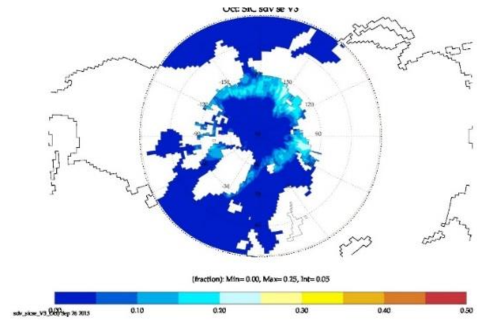
SDV SST SE OLD (V2)



SDV Abs Solar



SDV SIC Oct



Zuo et al 2017, Hirahara et al 2016

Perturbing the Observations

Representativeness error

1) Profile displacement and stretching

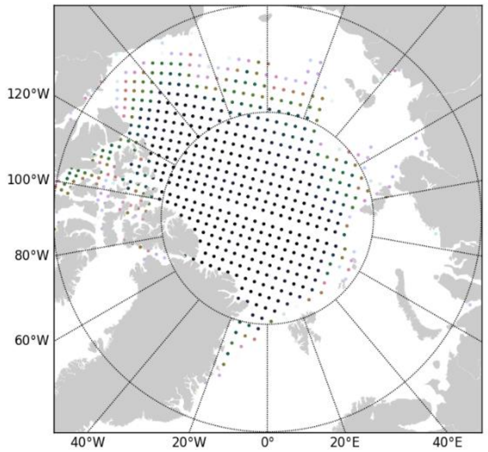
2) Thinning with random seed in different ensemble members:

More observations are used in the ensemble

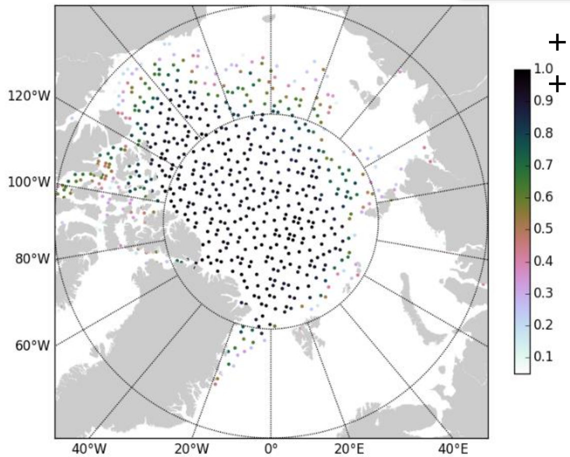
Used in ensembles of the ocean reanalyses.

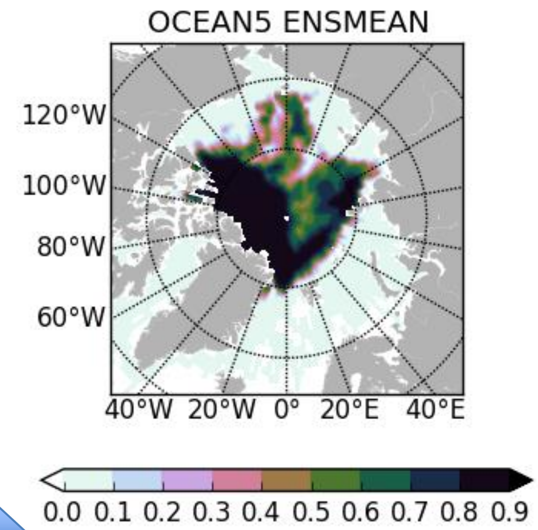
Thinning of Sea Ice Concentration Observations

Regular thinning

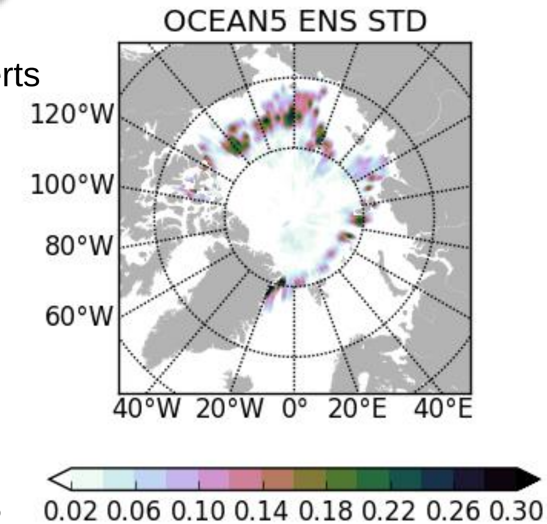


Random sampling





+ forcing perts
+ other obs perts



Zuo et al 2017, Tech. Memo 795

What about the ensemble spread in **coupled** data assimilation?

Observation uncertainties + **Model uncertainties** + **Model forcing uncertainties** = **Reanalysis uncertainties**

Compare ensemble spread of CERA-20C with equivalent uncoupled ocean reanalysis.

Uncoupled: Forcing and SST perturbations . By design, only capture seasonal dependence
Coupled: Spread generated by coupling. SST from HadISST.

same observations, same data assimilation, same observation perturbations

We diagnose the flow dependence of the spread: Decadal, interannual, intraseasonal

Decadal variations of spread

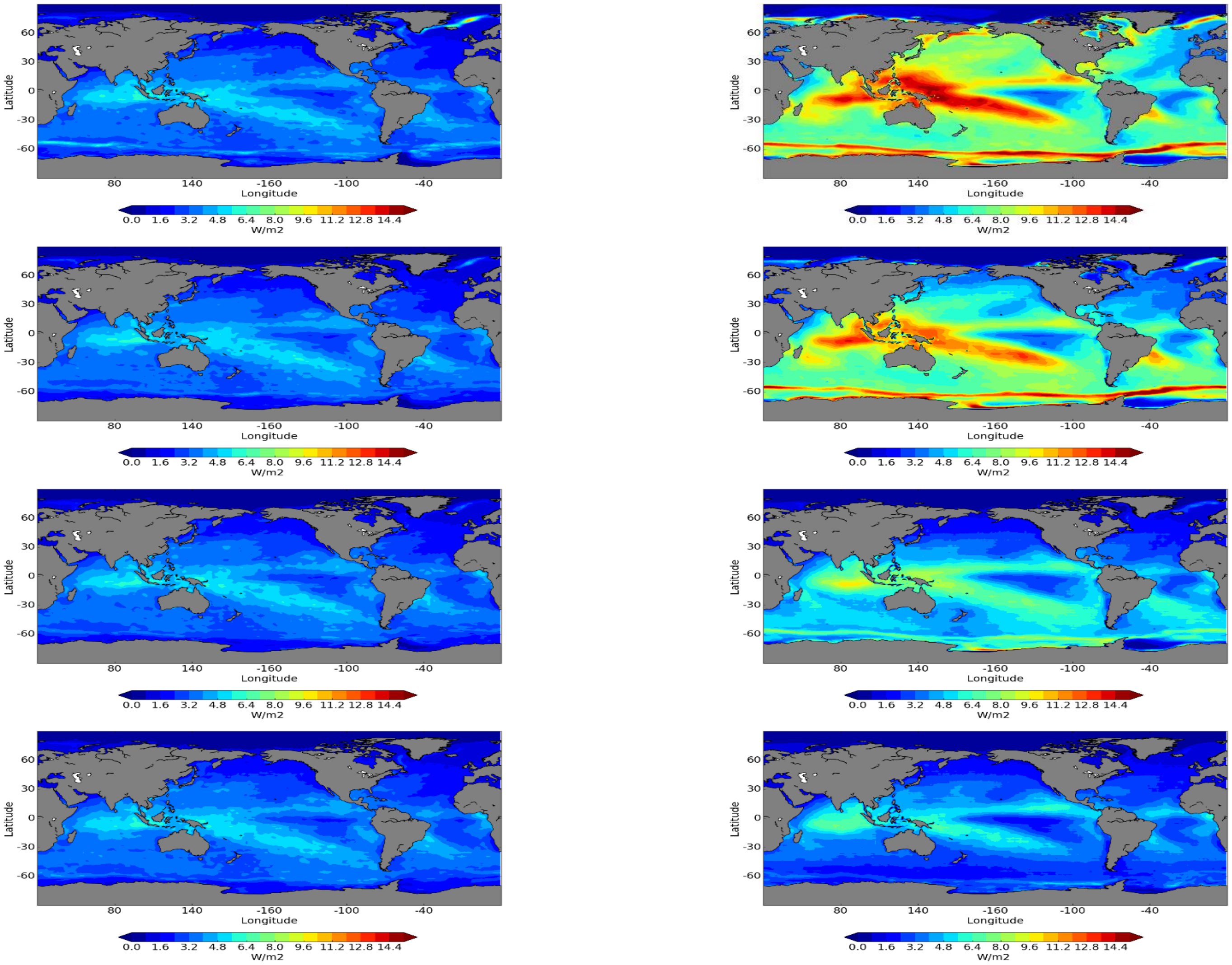
ORA-20C | Solar radiation | CERA-20C

1900s

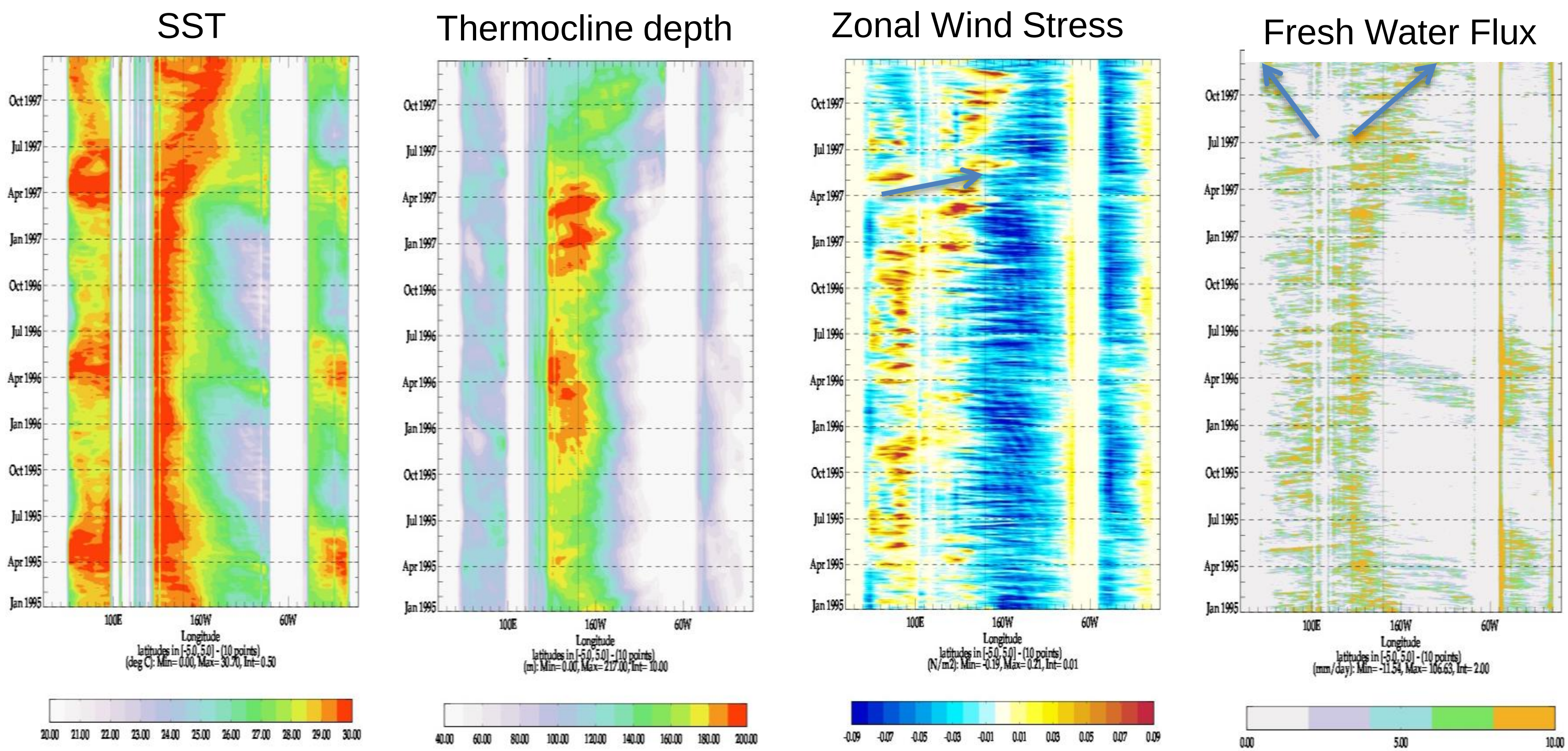
1940s

1970s

2000s



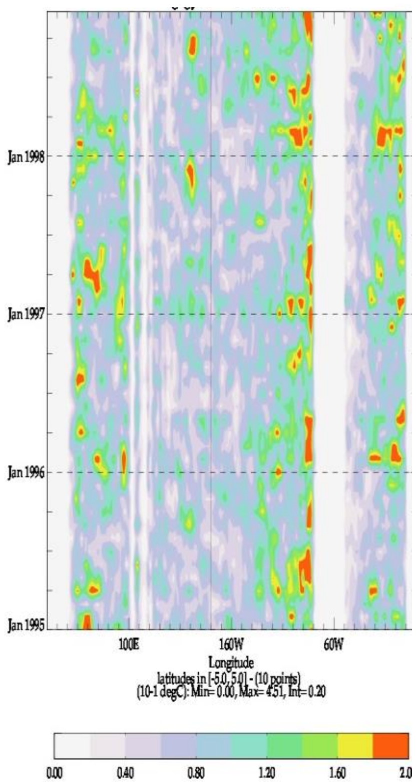
Zoom on 1996-1997: Onset of El Nino
Equatorial daily time series of actual reanalysis fields



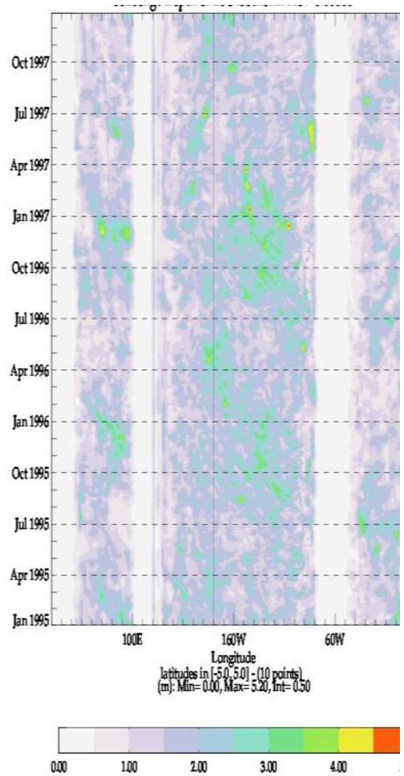
Coherent behaviour among variables SST-Precipation-Wind and thermocline response
Seasonal cycle, intraseasonal variability and onset of El Nino can be appreciated

Zoom on 1996-1997: Onset of El Nino
Equatorial daily time series of UNCOUPLED ensemble spread

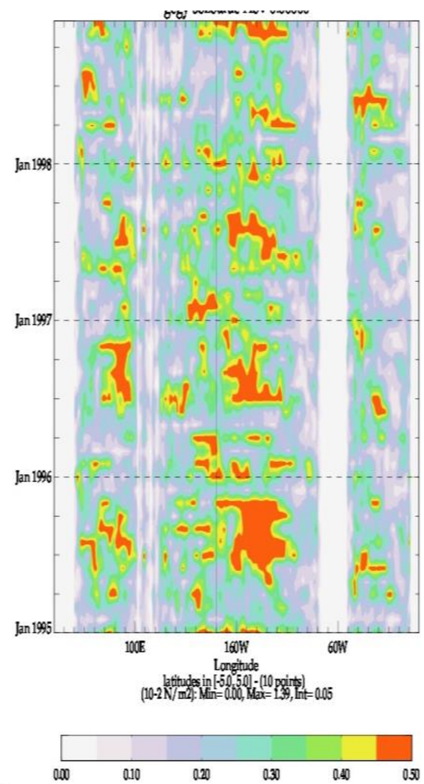
SST



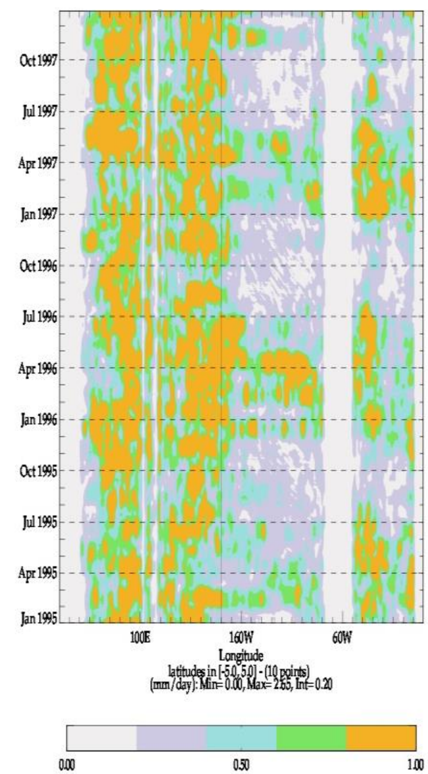
Thermocline depth



Zonal Wind Stress

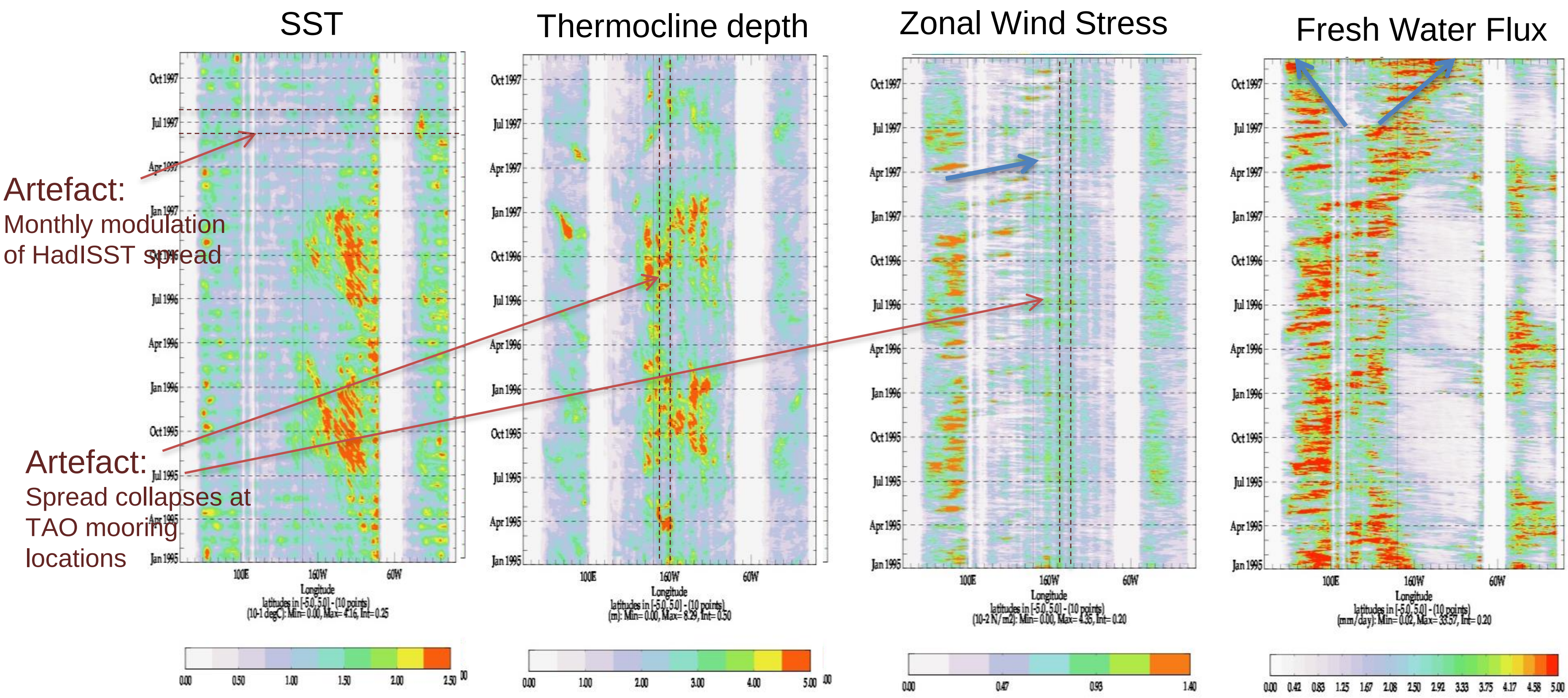


Fresh Water Flux



Coherent spread between ocean and atmopheric variables only at seasonal time scales (by design)
Ocean variables -SST and Thermocline depth- spread show intraseasonal –TIWs- and interannual modulation

Zoom on 1996-1997: Onset of El Nino
Equatorial daily time series of COUPLED ensemble spread



Coherent behaviour among variables SST-Precipitation-Wind and thermocline at seasonal-intraseasonal-interannual time scales

Summary Initialization

- Criteria to design a good Initialization of Earth System:
    - Reduce initialization shock: coupled DA contributes to more balance I.C.
    - Drift and calibration: Historical and stable records of initial conditions consistent with real time needed for calibration: bias correction, reanalyses
    - Important to exploit observational information and deal with the non stationary observing system
- Initialization of the ocean (focus on seasonal forecasting)
    - Important to initialize the dynamical and thermodynamic process
    - Data assimilation changes the ocean mean state. Therefore, consistent ocean reanalysis requires an explicit treatment of the bias
    - Assimilation of ocean observations reduces the large uncertainty (error) due to the forcing fluxes. Initialization of Seasonal Forecasts needs SST, subsurface temperature, salinity and altimeter derived sea level anomalies.
- Different approaches to initialization: full versus anomaly initialization
- Ensemble generation for ocean initial conditions:
    - sampling known uncertainty. Next step is to sample model error in ocean.
    - Coupled reanalysis should represent better the flow dependent uncertainty