Initialization and Forecast Strategies for Seamless Forecasting

Magdalena A. Balmaseda

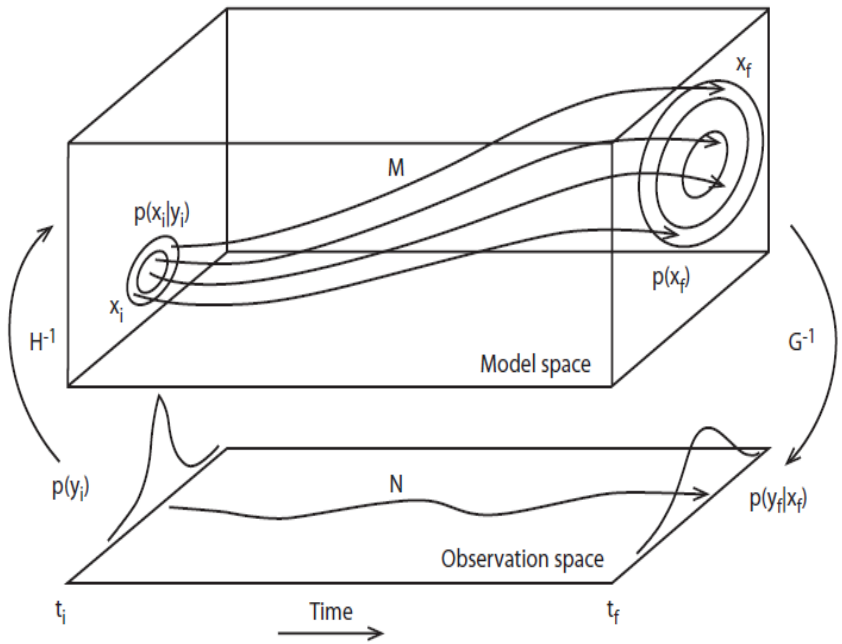
ECMWF

End to end Initialized Probabilistic Forecasting System

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

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

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

Stephenson et al 2005

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 is the energy norm of the atmosphere 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

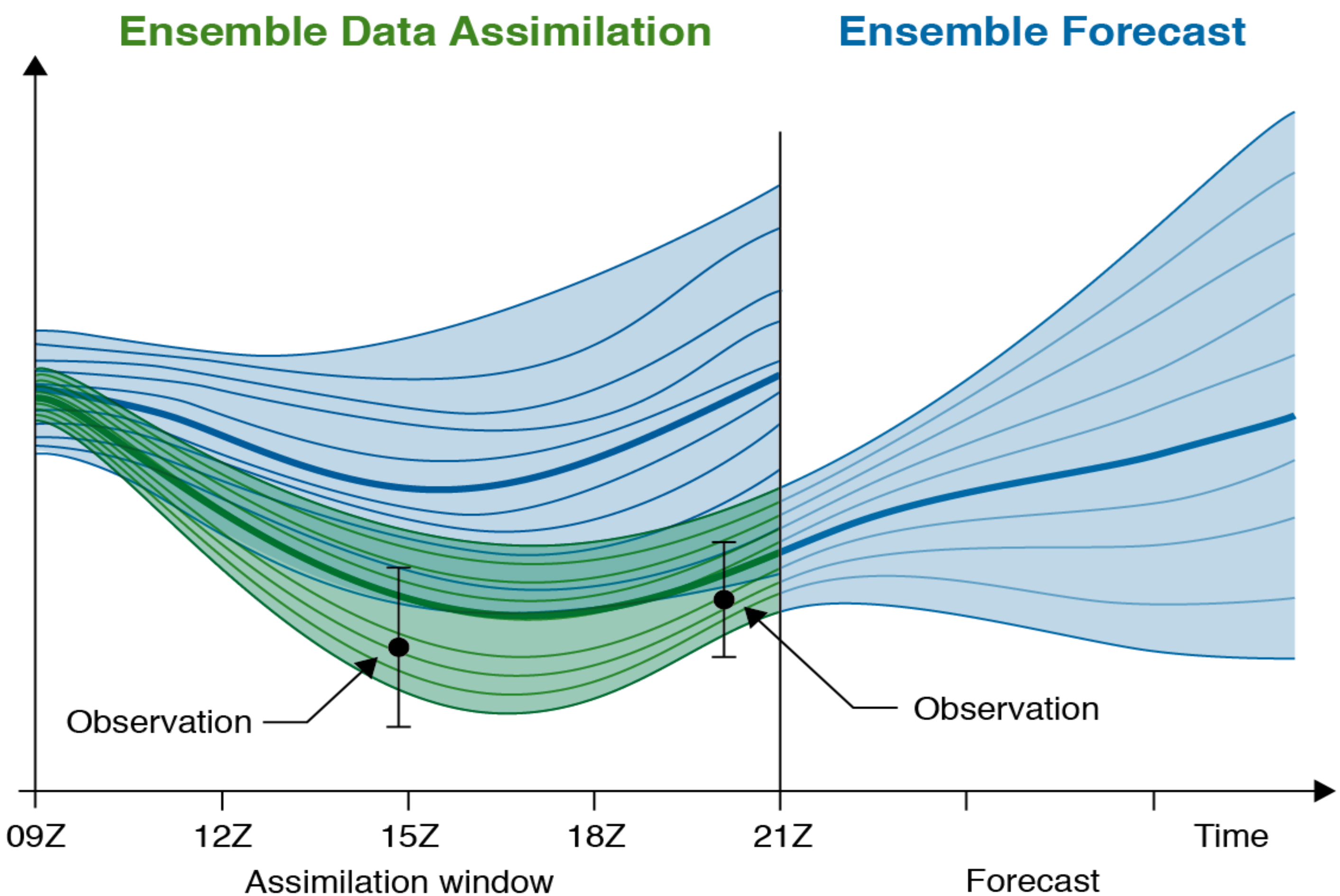
- **Practical requirements arising from calibration**:
  - Stationary forecast errors
  - Consistency between re-forecasts and real time fc

  Need for historical reanalysis

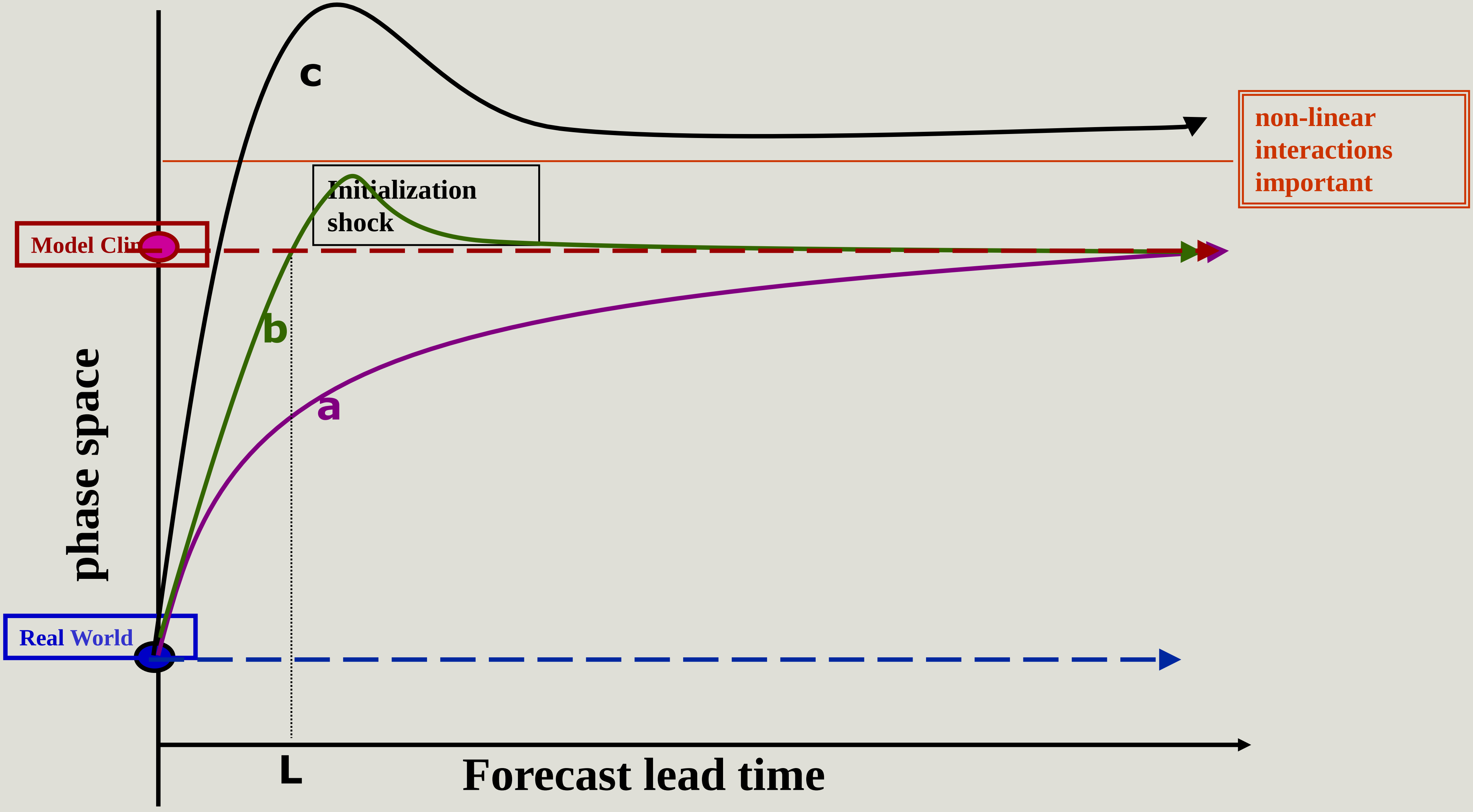
OUTLINE

- Initialization Shock
- Forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Initialization in the context of forecasting strategies:
    - Dealing with model error

The Data assimilation process



Initialization Shock and 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.

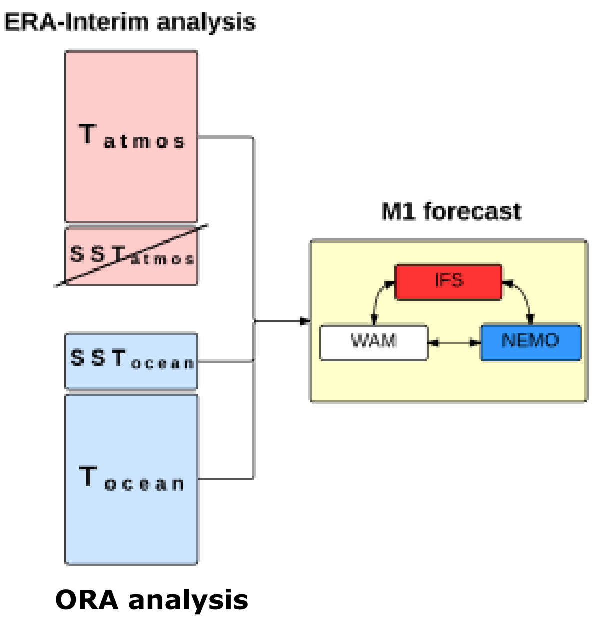
Possible reasons for initialization shock

1. **Data assimilation does not preserve model physical constrains**
    - 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

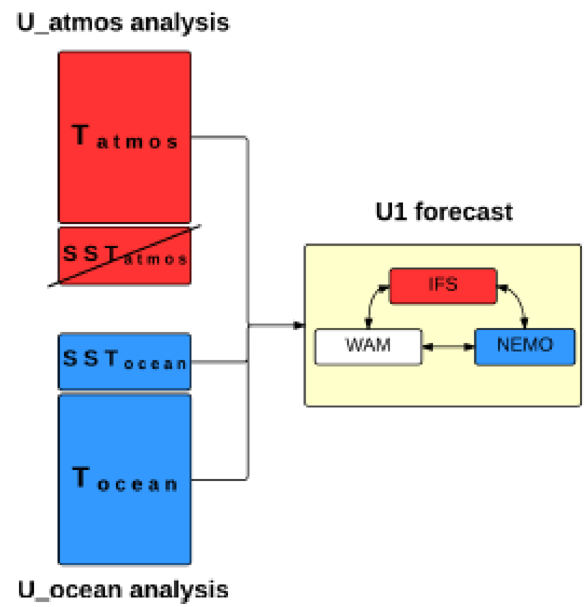
Coupled Initialization and Forecast Shock

M1
Uncoupled Ini
AN mod .ne. FC mod



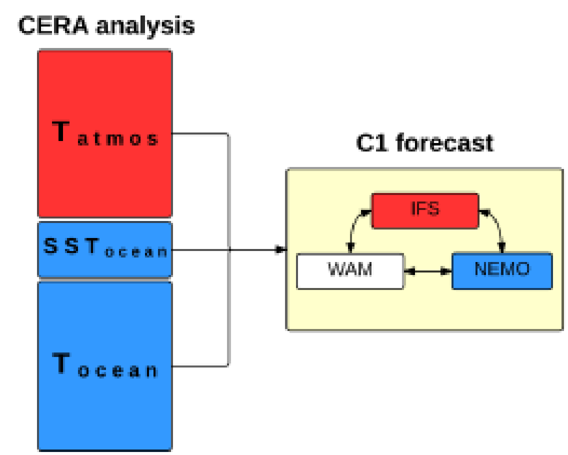
Approach for Coupled reforecast

U1
Uncoupled Ini
AN mod = FC mod



Approach for real time fc

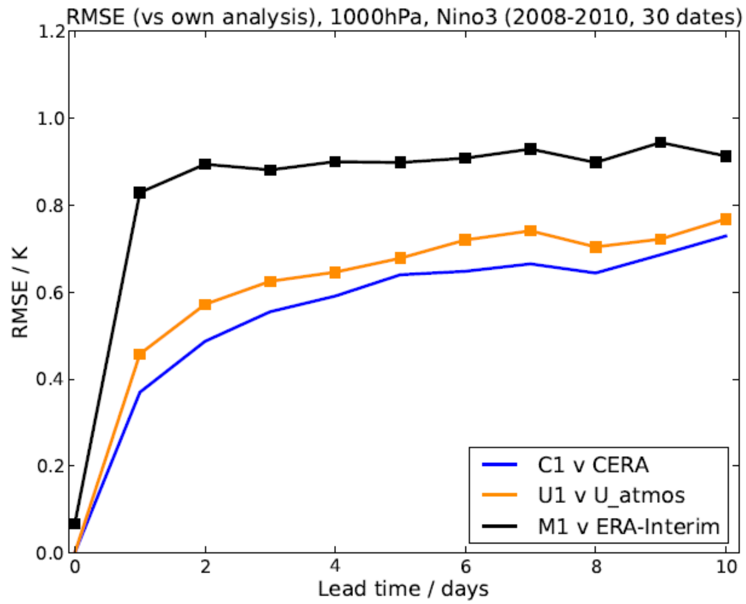
C1
Coupled Ini (CERA)
AN mod = FC mod



Future systems

Laloyaux et al, 2014 QJ, Mulholland et al, 2015MWR

Initialization shock:
forecast error growth depends on Initialization



Uncoupled: different models

Uncoupled: Same models

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

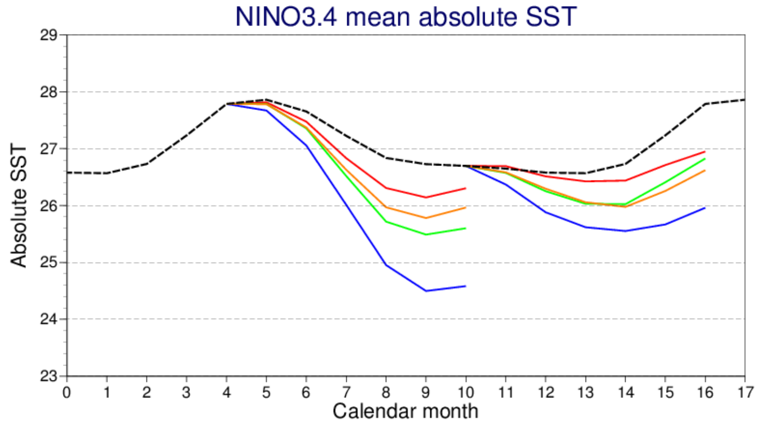
OUTLINE

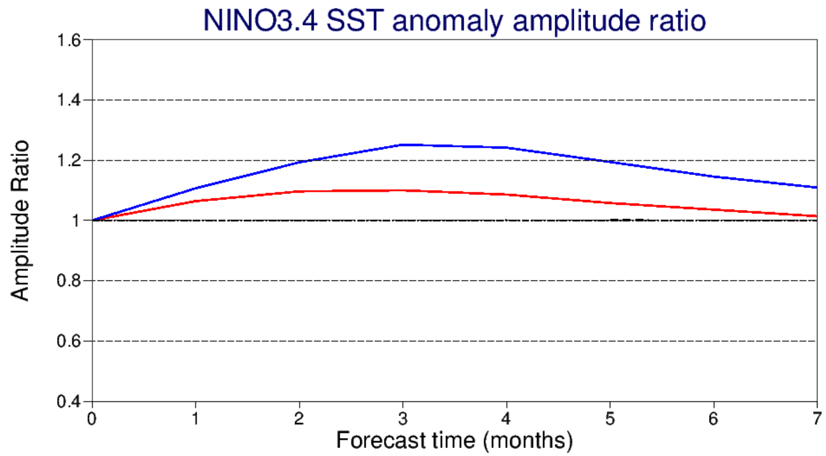
- Initialization Shock
- Forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Initialization in the context of forecasting strategies:
    - Dealing with model error

Example of fc drift in Seasonal Forecast

FC drift depends on the model

SEAS5 S4 S5-lr S5-mr





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

- Bias depends on model (not on the initialization)
  - 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
- The figure shows the ratio model/obs variability.

SEAS5: Current ECMWF Seasonal Forecasting System
S4 : Previous " " " "
S5-lr : As SEAS5, with low resolution ocean and atmos
S5-mr: As SEAS5, with low resolution ocean, high res atmos

Stockdale et al 2018. ECMWF Tech Memo 835
Johnson et al 2019, GMD

Forecast calibration

At the medium range, *a-posteriori* fc calibration is not used

- Forecast drift exist, but *a-posteriori* calibration is considered not worth it.
  - Timeliness of the forecasts is critical.
  - *A-posteriori* calibration would slow the production.
  - Large amount of data to process.
  - Instead, the ensemble is tuned a-priori to achieve reliable forecast

At the extended/seasonal range *a-posteriori* fc calibration is needed

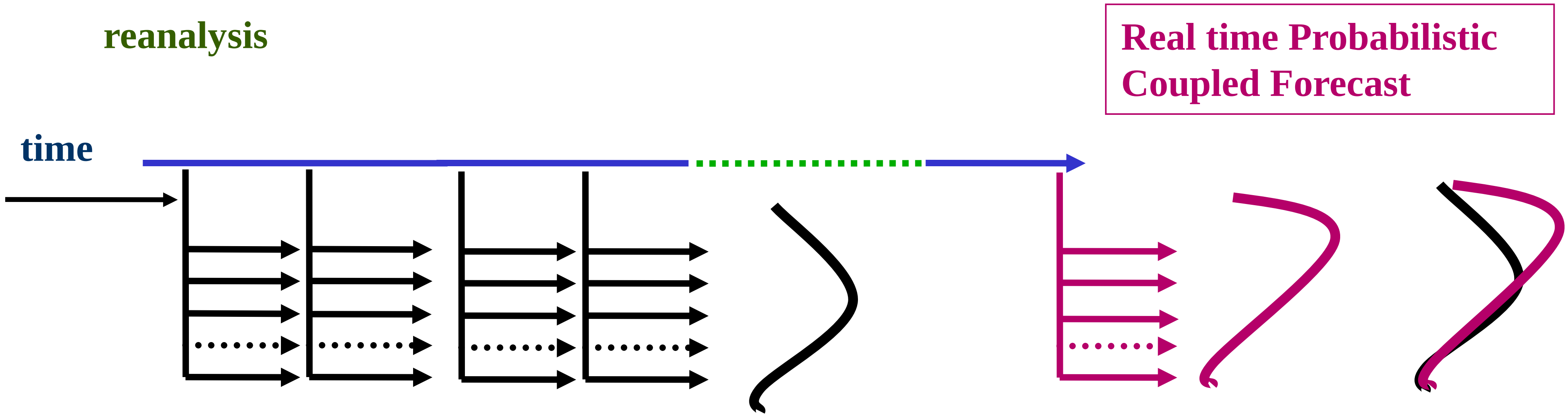
$$\tilde{x} = \bar{y} + \mathbf{K}(x - \bar{x}) + \mathbf{F}\varepsilon_x$$

Bias correction ($\bar{x} \neq \bar{y}$)

K: linear transformation of anomalies

F: Adjustment of ensemble spread

Dealing with model error: Reforecast



Coupled reforecasts, needed to estimate climatological PDF, require a historical reanalysis

Consistency between historical and real-time initial conditions is required.

Reforecasts are also needed for skill estimation

OUTLINE

- Initialization Shock
- Forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Initialization in the context of forecasting strategies:
    - Dealing with model error

The basis for extended range forecasts

•Forcing 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.

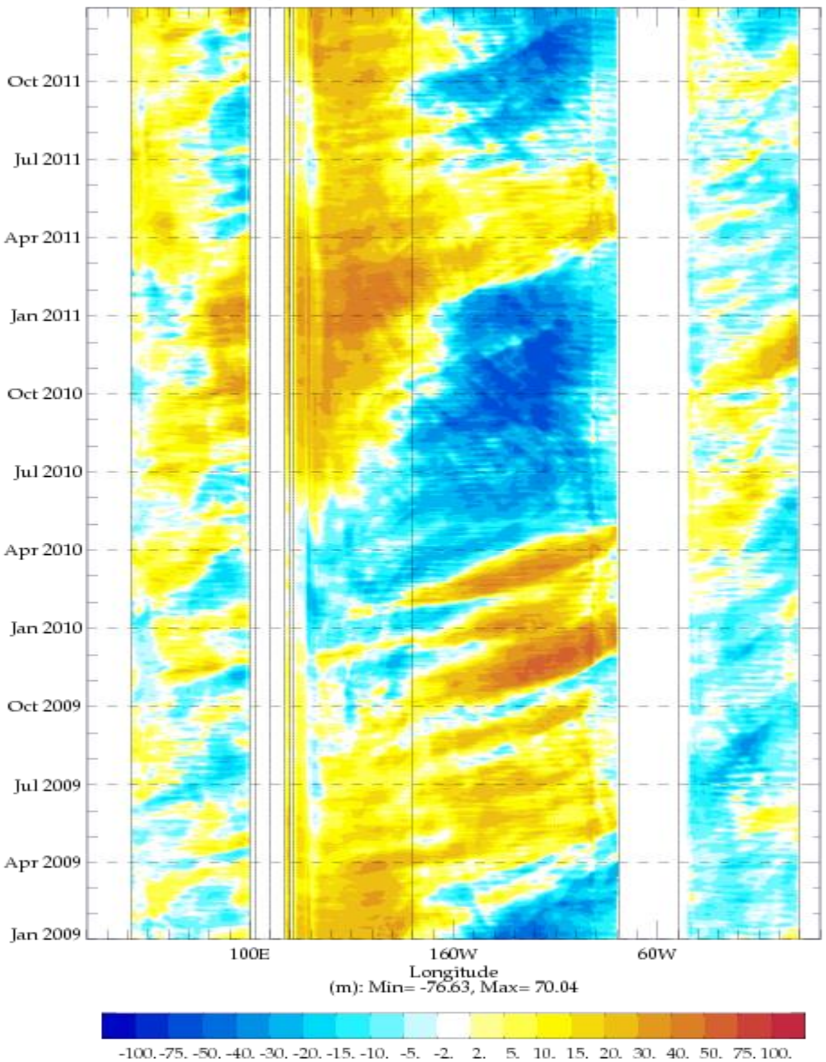
- Important to bear in mind the probabilistic nature of SF

•The boundary conditions have longer memory, thus contributing to the predictability. Important boundary forcing:

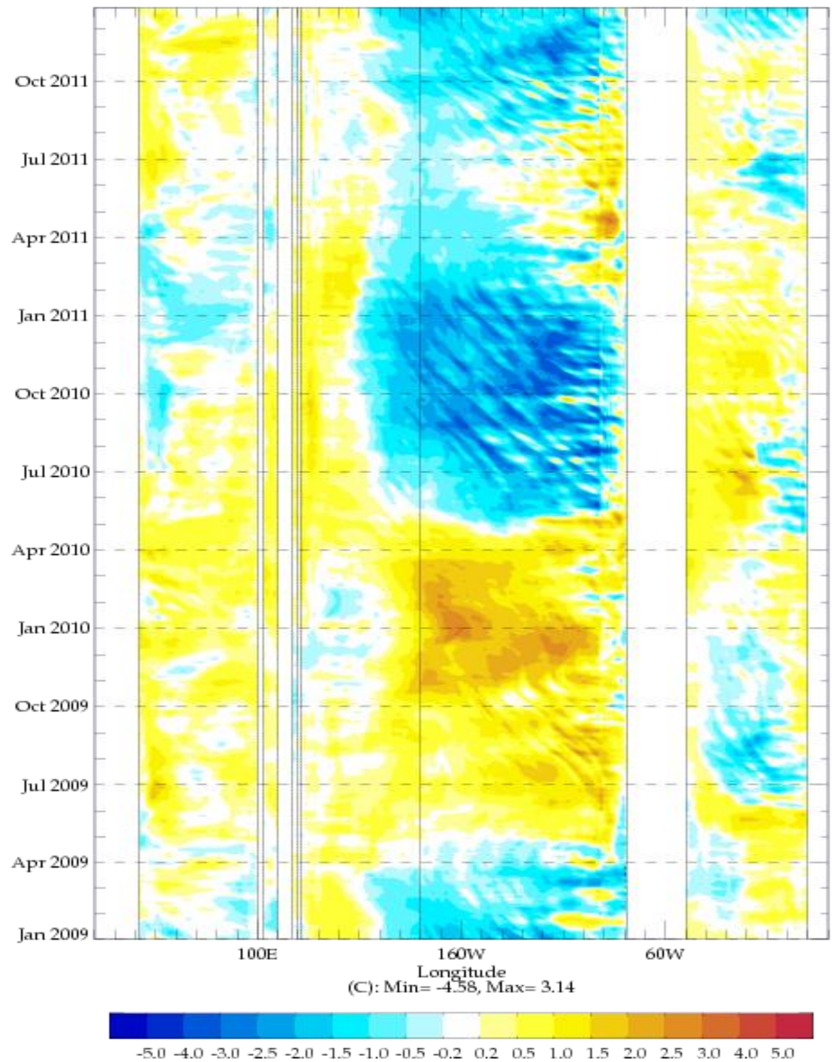
- **Tropical SST: ENSO, Indian Ocean Dipole, Atlantic SST**
- Land: snow depth, soil moisture
- Sea-Ice
- Mid-Latitude SST
- Atmospheric composition: green house gases, aerosols,…
- Stratosphere

Need to Initialize the subsurface of the ocean

20C Isotherm Depth Eq Anomaly

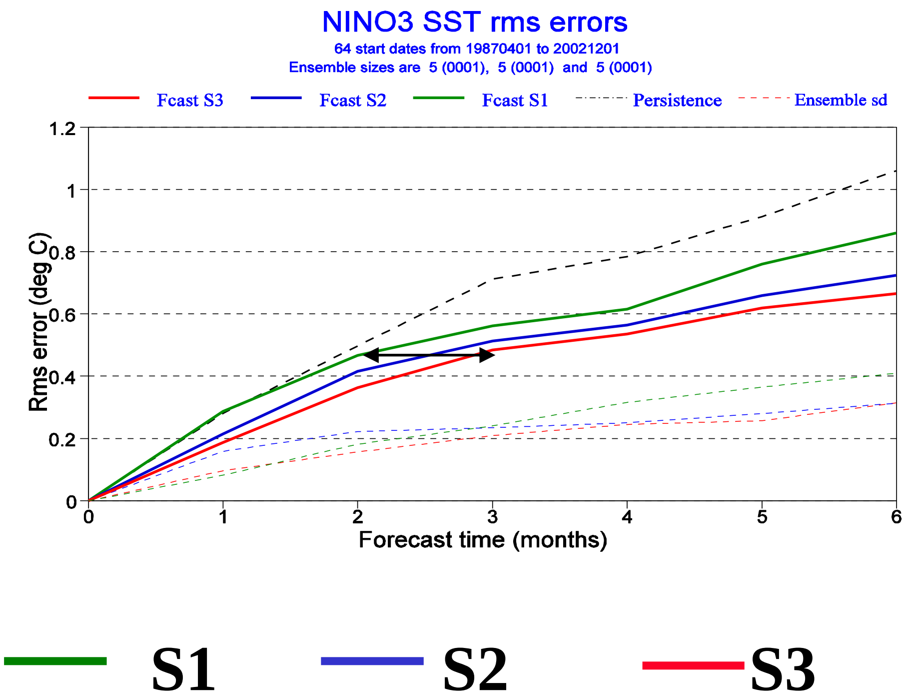


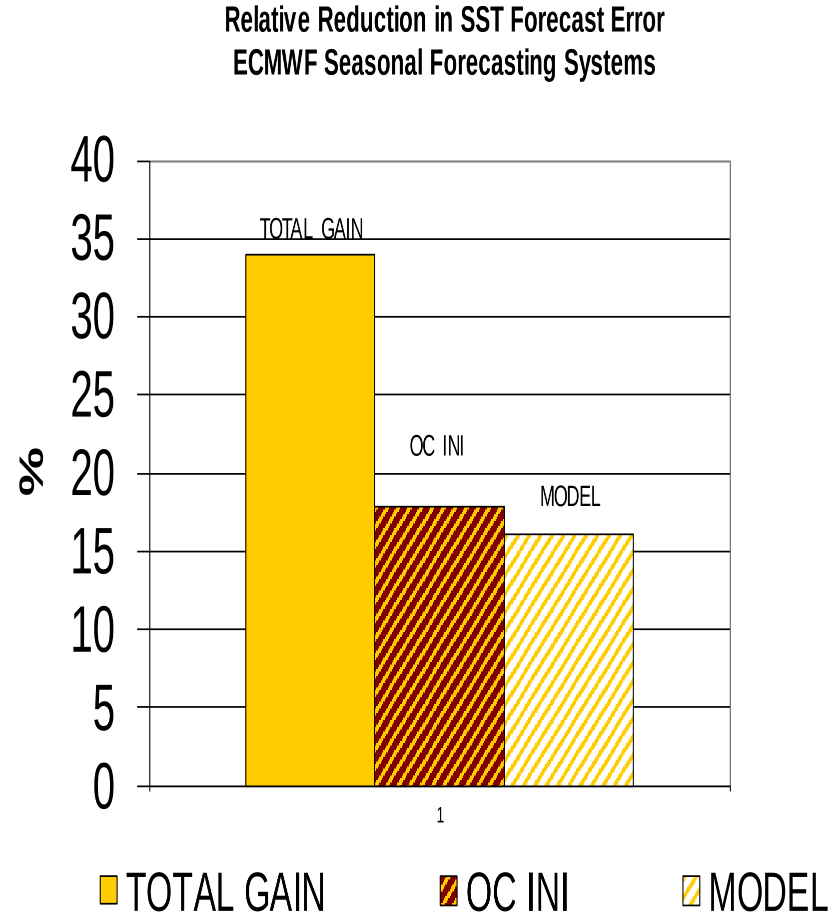
SST Eq Anomaly



Initialization into Context

Progress on ENSO prediction





- •Steady progress: ~1 month/decade skill gain
- •How much is due to the initialization, how much to model development?

Half of the gain on forecast skill is due to improved ocean initialization

Information to initialize the ocean

- Ocean model Plus:

  SST

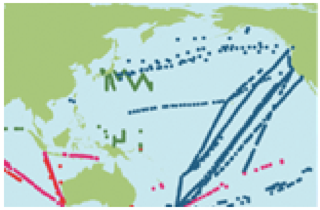
  Atmospheric fluxes from atmospheric reanalysis

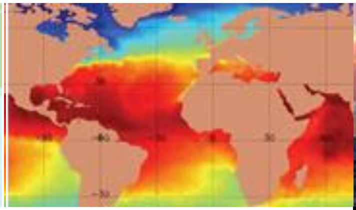
  Subsurface ocean information

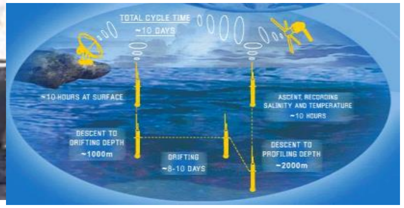
Time evolution of the Ocean Observing System

1982 1993 2001

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





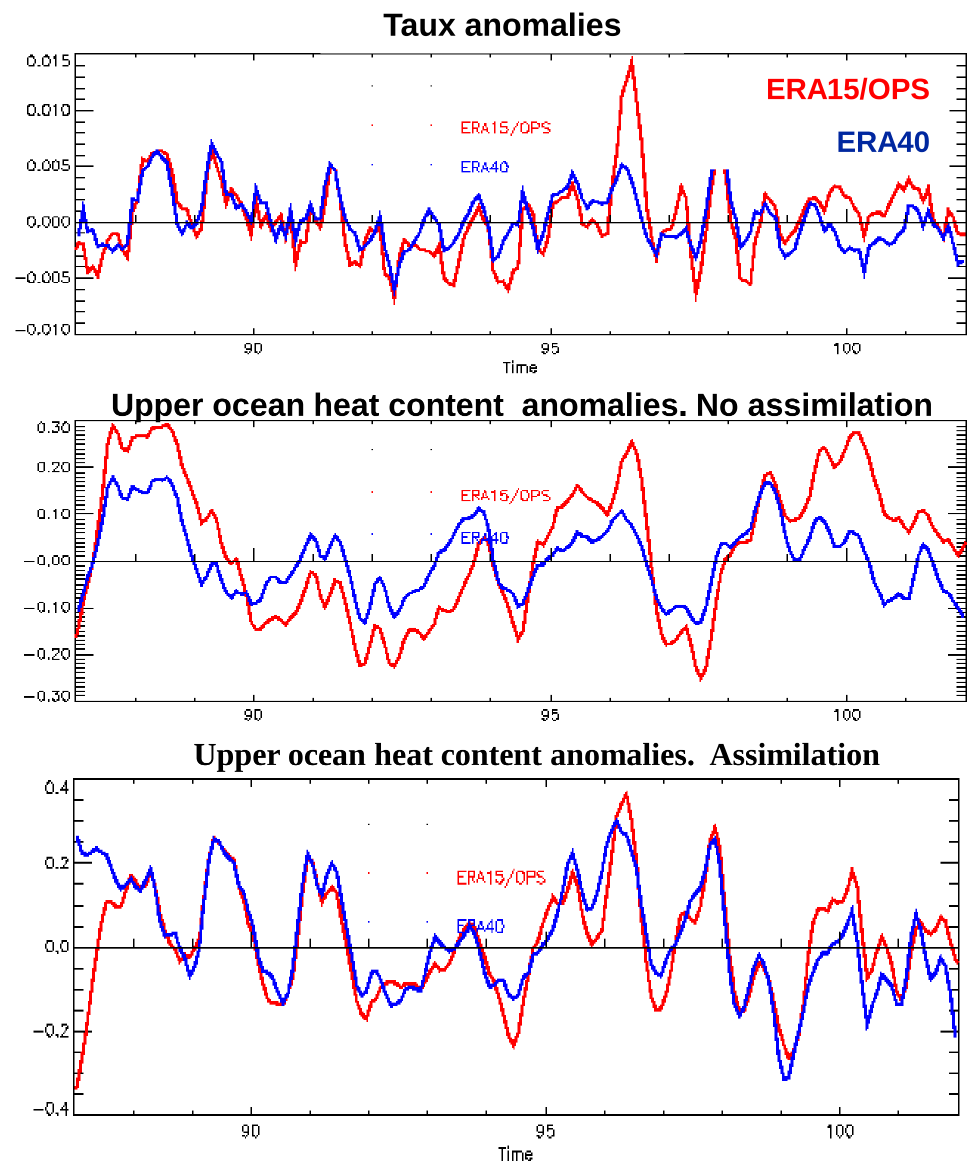


Uncertainty in Surface Fluxes:

Need for Data Assimilation

- 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

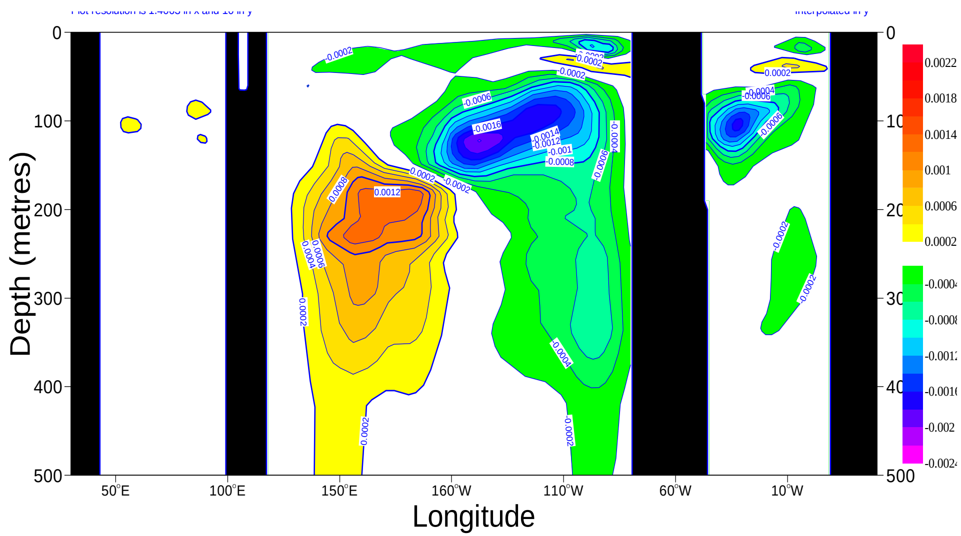


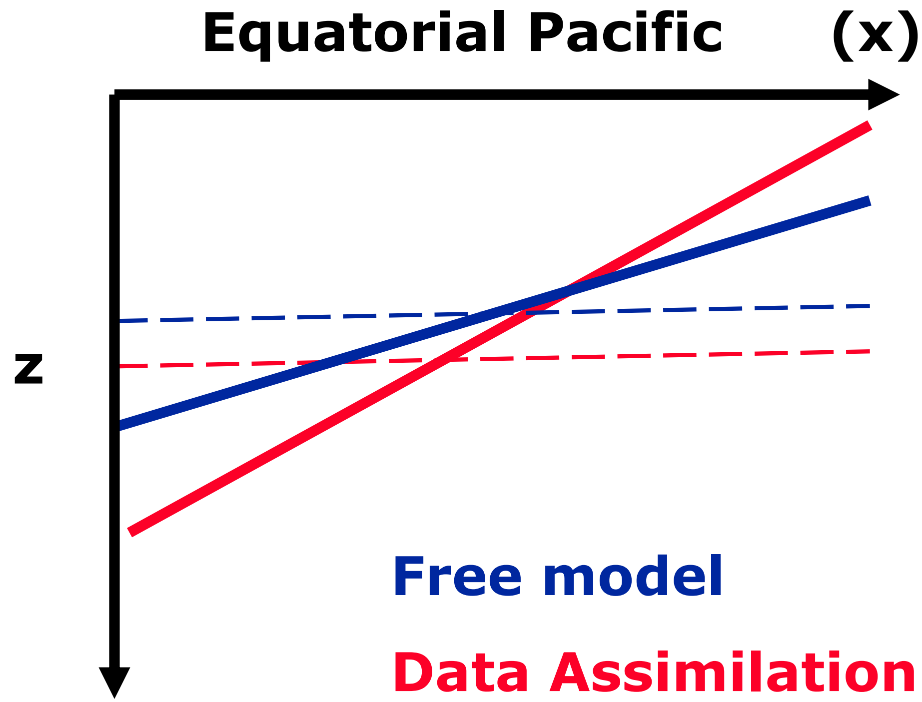
- Questions:

  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**

The Assimilation corrects the ocean mean state

Mean Assimation Temperature Increment

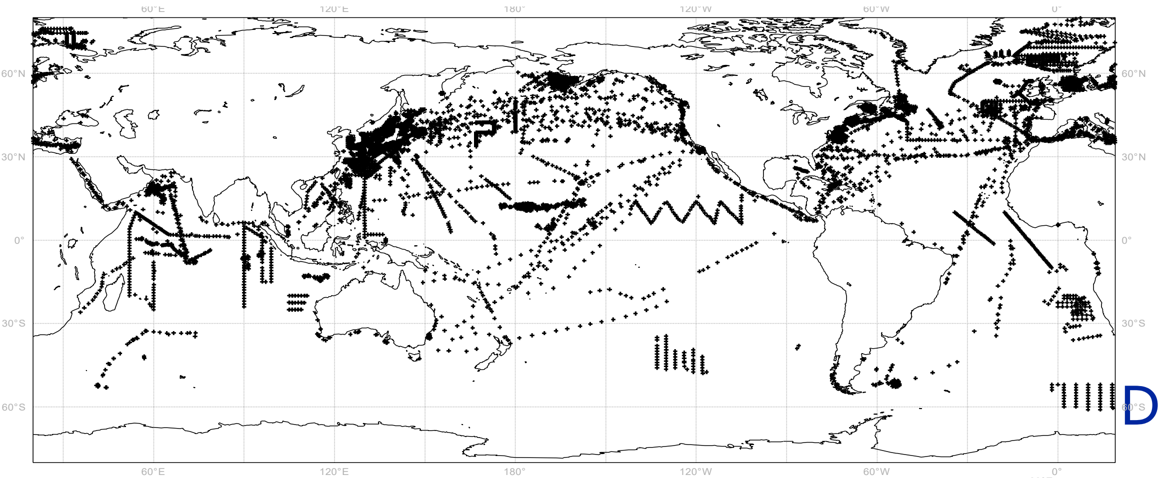




Data assimilation corrects the slope and mean depth of the equatorial thermocline

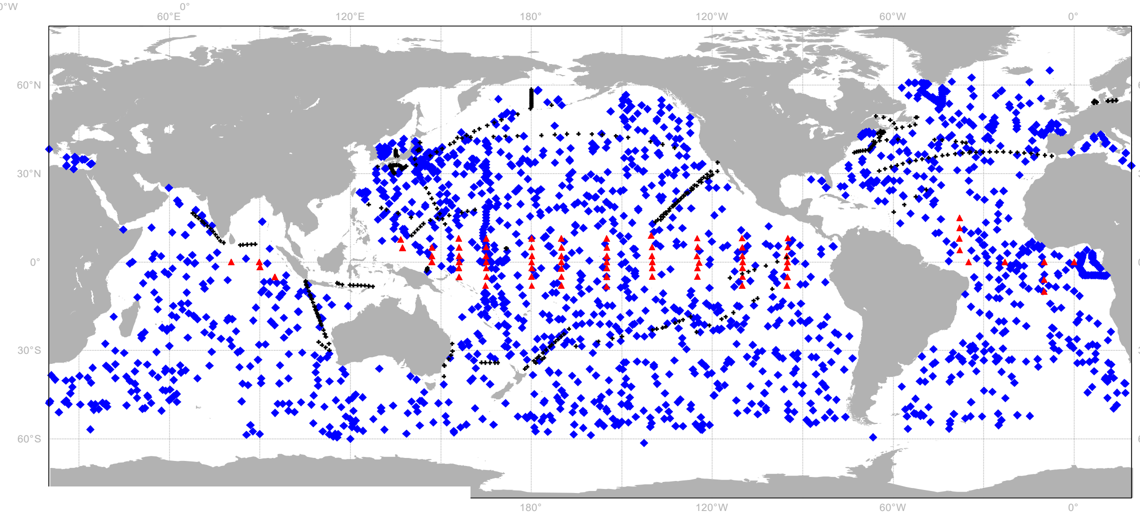
Ocean Observing System

Data coverage for June 1982



Changing observing system is a challenge for consistent reanalysis

Data coverage for Nov 2005



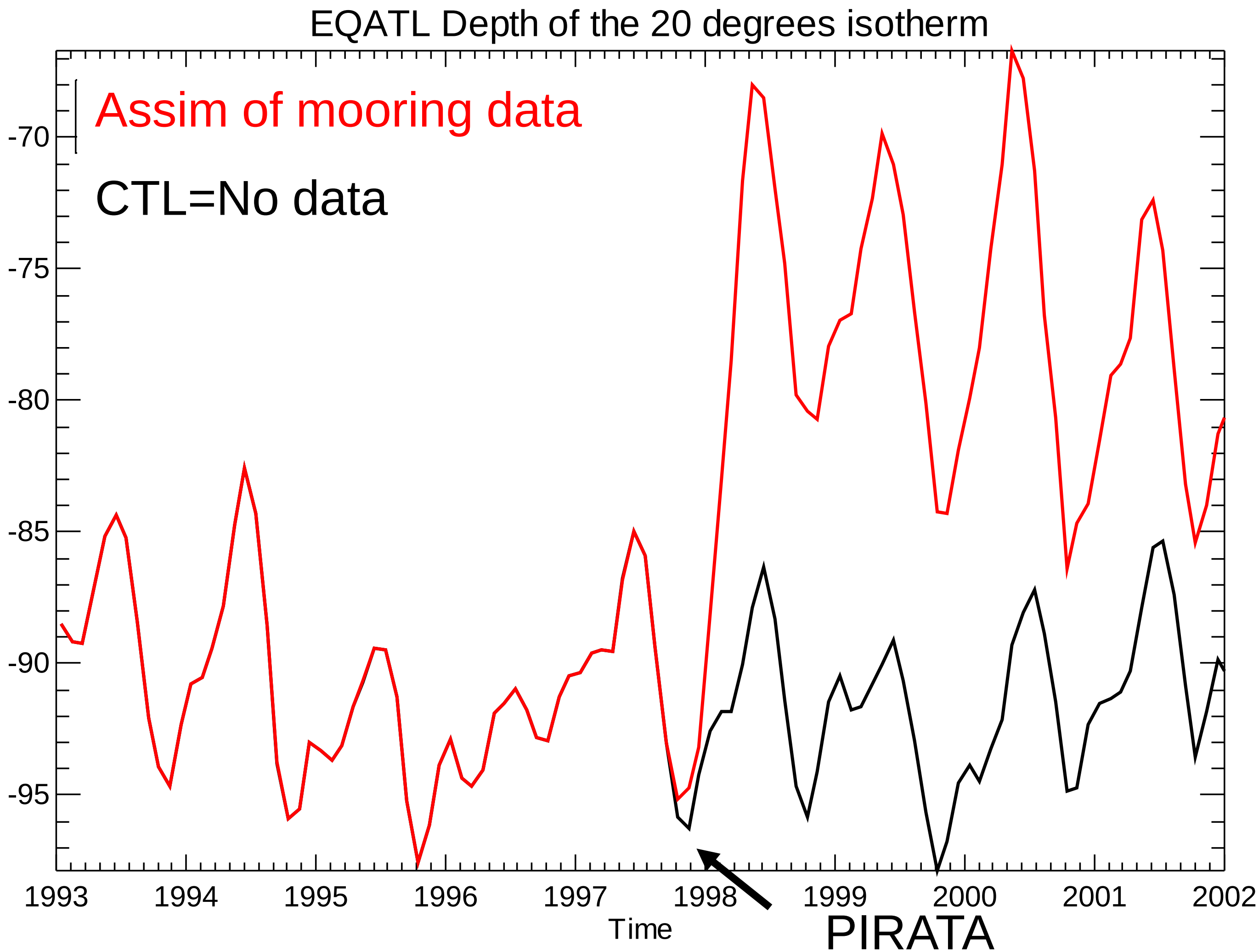
Today's Observations will be used in years to come

▲Moorings: SubsurfaceTemperature

◊ ARGO floats: Subsurface Temperature and Salinity

+ XBT : Subsurface Temperature

Impact of data assimilation on the mean



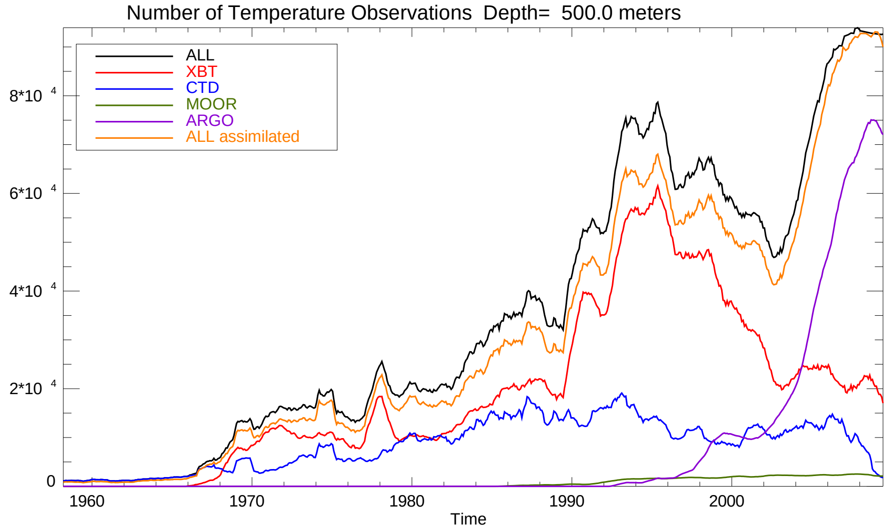
Large impact of data in the mean state leading to spurious variability

This is largely solved by the introduction of bias correction

Need to correct model bias during assimilation

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

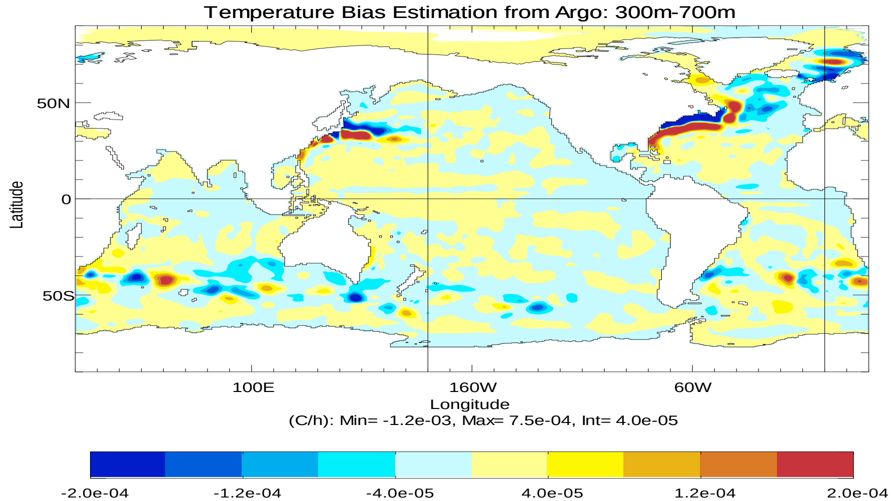
$$\mathbf{b}^a = \mathbf{b}^f + \mathbf{L}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{b}^f)]$$



There is a model for the time evolution of the bias

$$\mathbf{b}^f{}_k = \overline{\mathbf{b}}_k + \mathbf{b}'^f_k$$

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

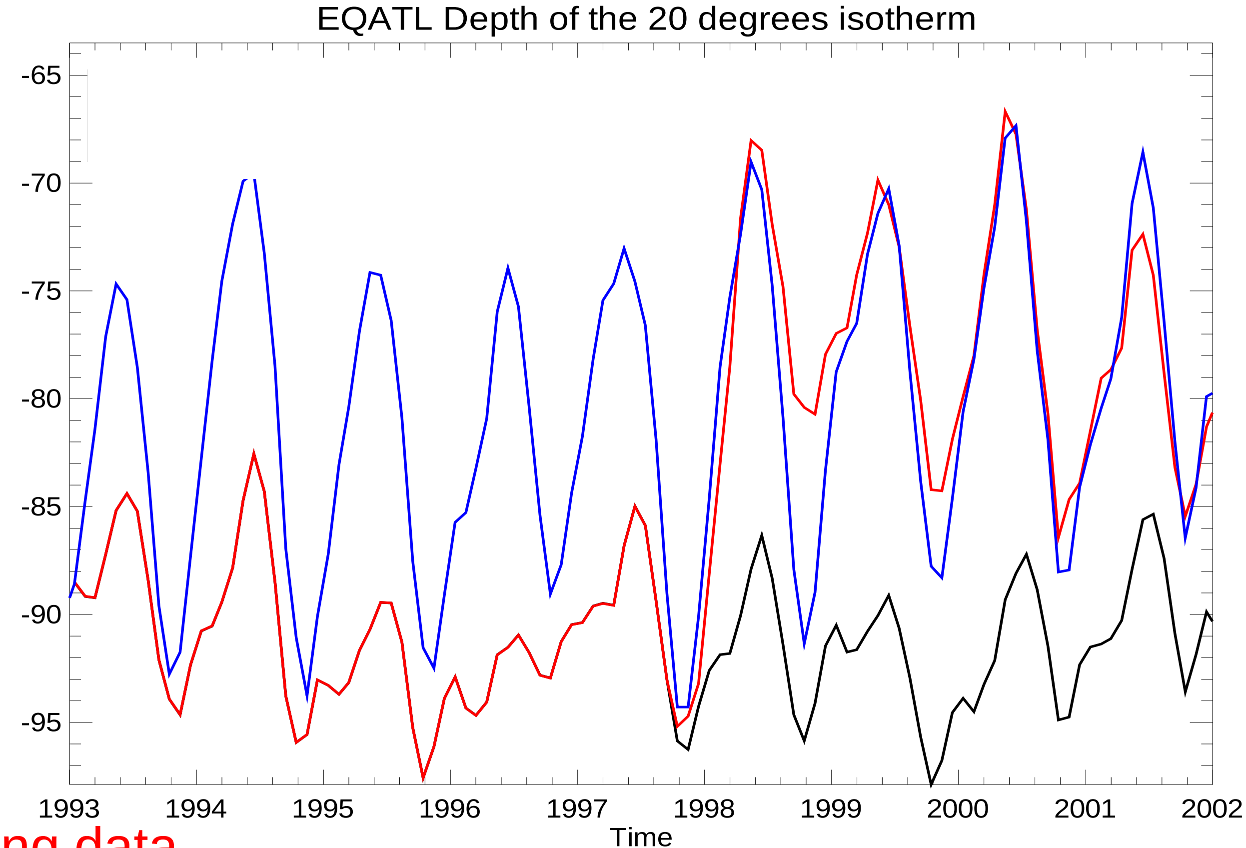


Without explicit bias correction changes in the observing system can induce

- Spurious signals in the ocean reanalysis
- Non-stationarity of the forecast bias, leading to forecast errors.

Ideally, the bias information should be propagated during the forecast (for this the FG model and FC model should be the same, e.i. coupled model)

Effect of bias correction on the time-evolution



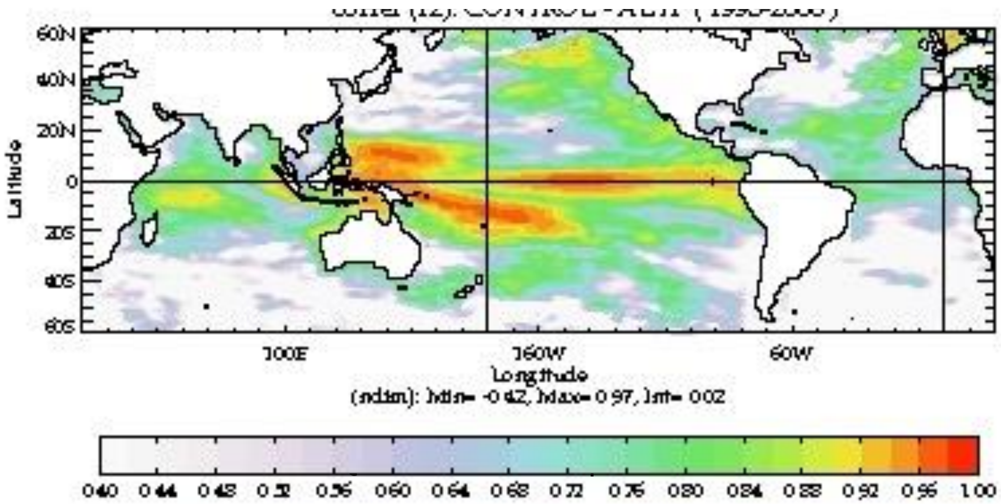
Assim of mooring data

CTL=No data

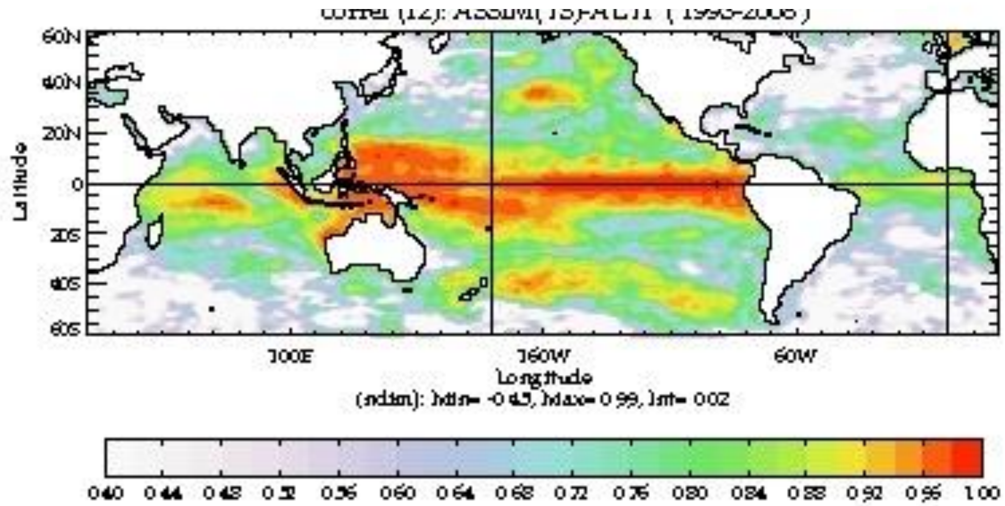
Bias corrected Assim

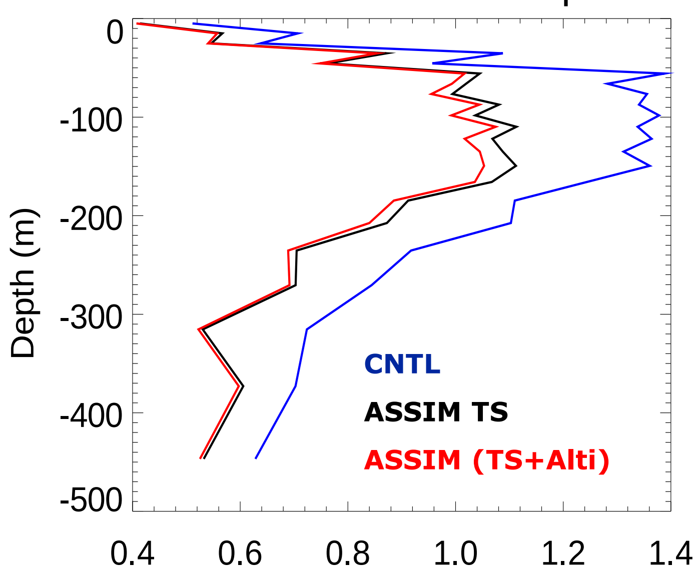
Time correlation with altimeter SL product

CNTL: NoObs

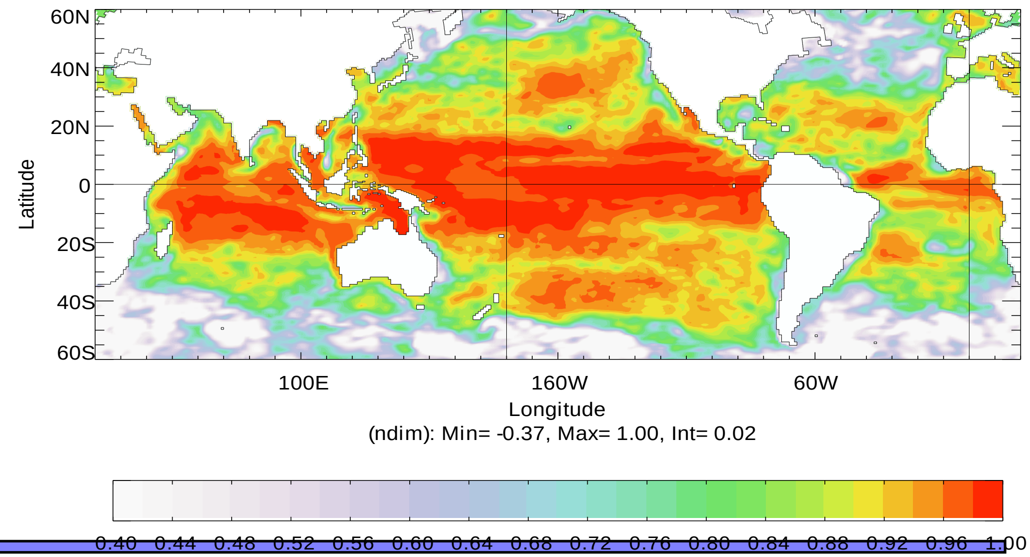


Assimilation of T+S



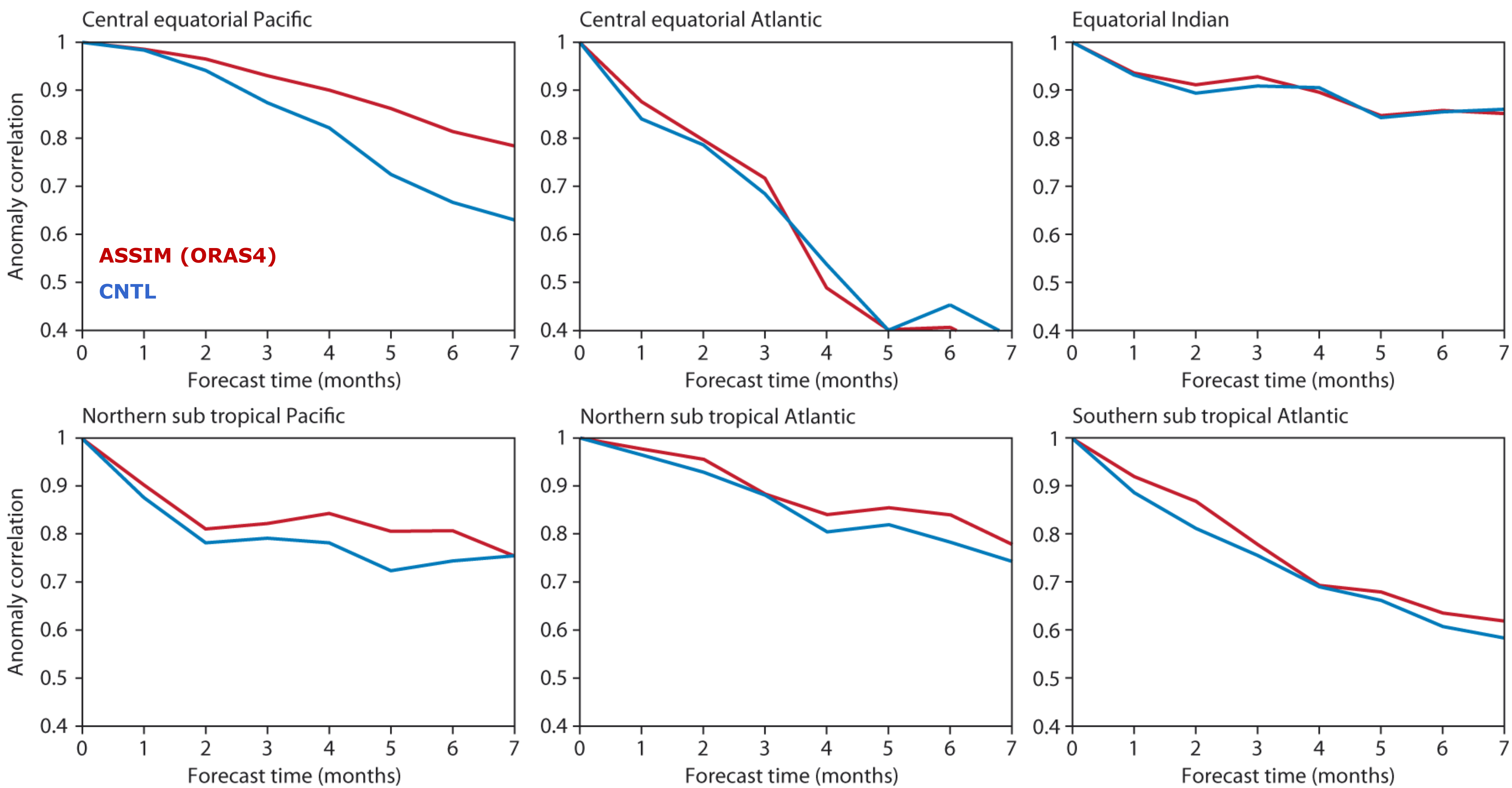


Assimilation of T+S+Alti



Impact on Seasonal Forecast Skill

Consistent Improvement everywhere. Even in the Atlantic, traditionally challenging area



Quantifying the value of observational information

- The outcome may depend on the coupled system
- In a good system information may be redundant, but not detrimental.

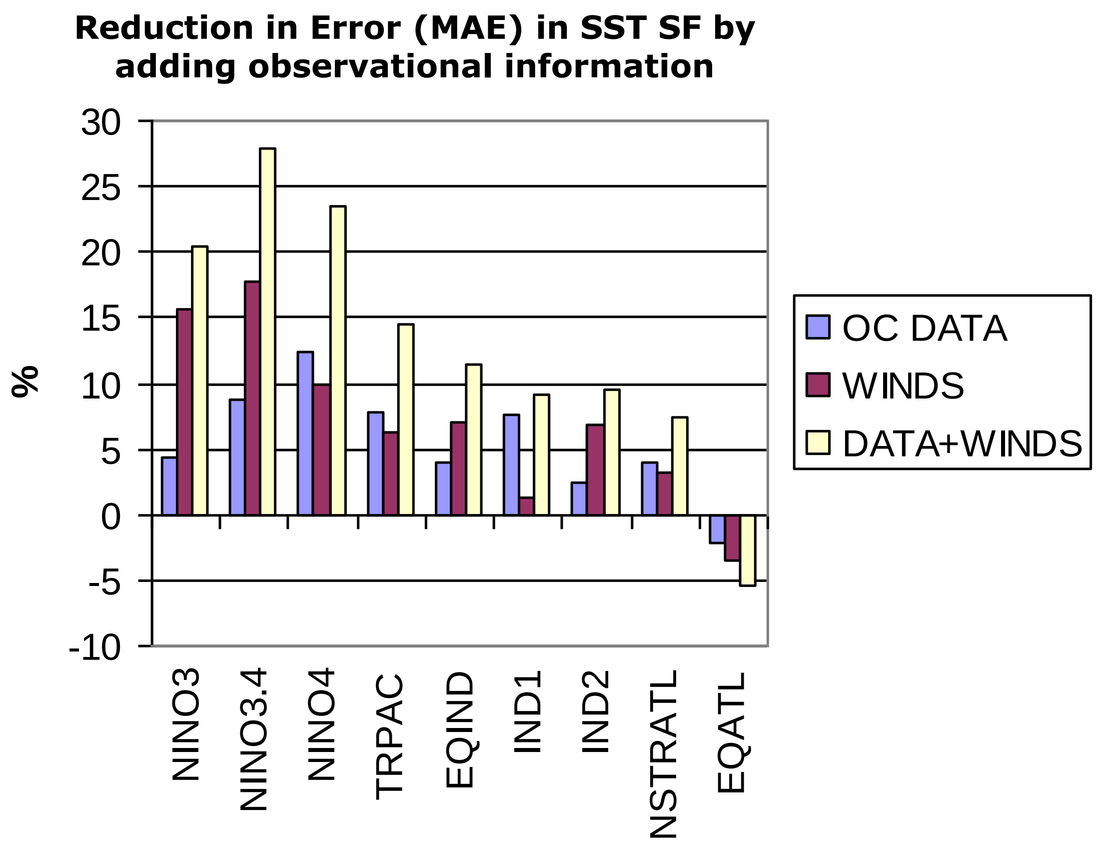
  If adding more information degrades the results, there is something wrong with the methodology (coupled/assim system)

- Experiments conducted with the ECMWF S3

  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)

Balmaseda and Anderson 2009, GRL

Impact of “real world” information on skill:

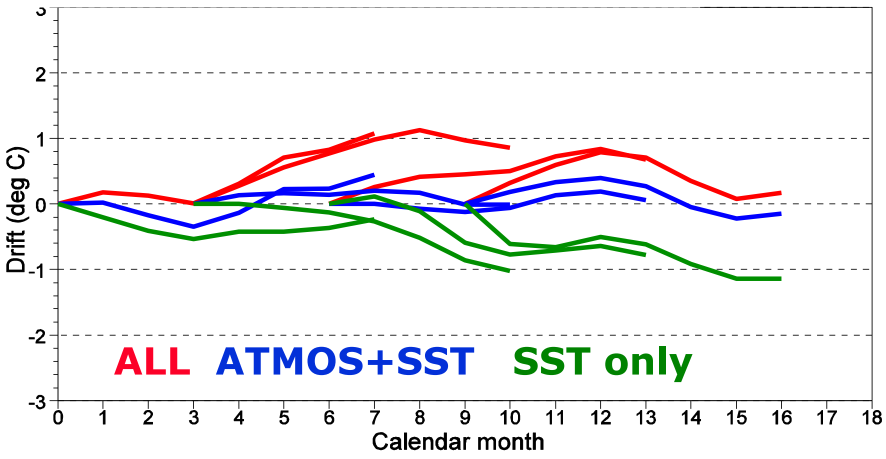


The additional information about the real world improved the forecast skill, except in the Equatorial Atlantic

Optimal use of the observations needs more sophisticated assimilation techniques and better models, to reduced initialization shock

Initialization and forecast drift and shock

NINO3 mean SST drift



Different initializations produce different drift in the same coupled model.

Warm drift in **ALL** caused by Kelvin Wave, triggered by the slackening of coupled model equatorial winds

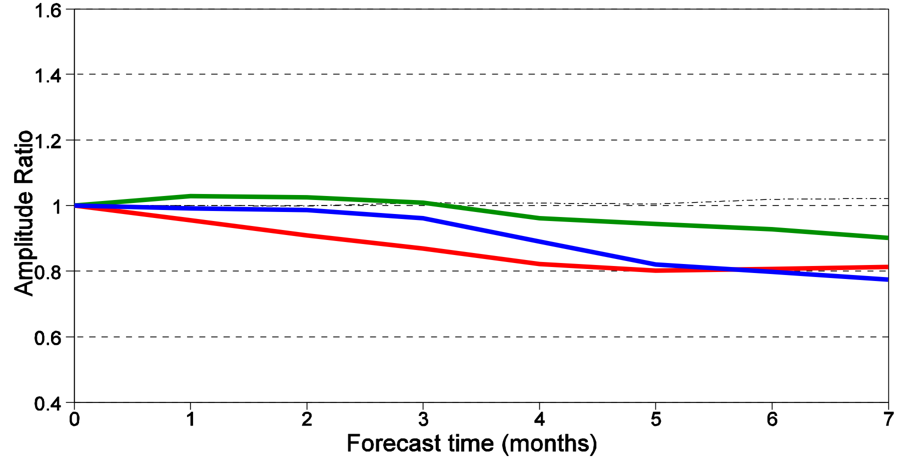
SST only has very little equatorial heat content, and the SST cool s down very quickly.

SST+ATMOS seems balanced in this region. Not in others

Sign of non linearity:

The drift in the mean affects the variability

NINO3 SST anomaly amplitude ratio



OUTLINE

- Initialization Shock
- Forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Initialization in the context of forecasting strategies:
    - Dealing with model error

Perceived Paradigm for initialization of coupled forecasts

Real world **Model world**

Medium range

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

Seasonal?

Decadal or longer

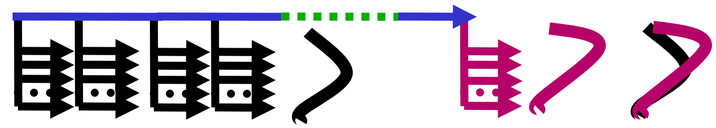
Anomaly initialization: Avoid forecast drift by initializing around the model mean state

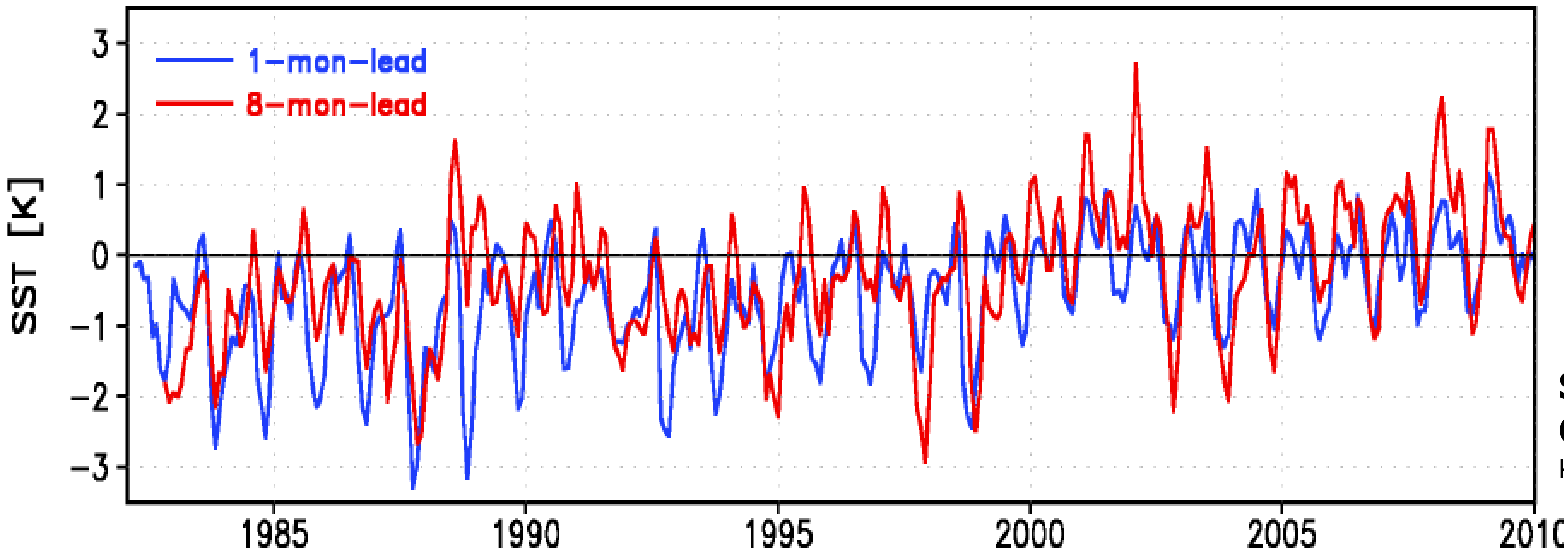
At first sight, this paradigm would not allow a seamless prediction system.

Anomaly initialization is not the same as model attractor initialization

Anomaly Initialization (decadal forecasts, Smith et al 2007)
Full initialization with coupled models of the slow component only
Other more sophisticated (EnKF, coupled DA, weakly coupled DA)

Full Initialization Approach
Some Known Caveats





SST FC bias
CFS.v2
Kumar et al 2011 MWR

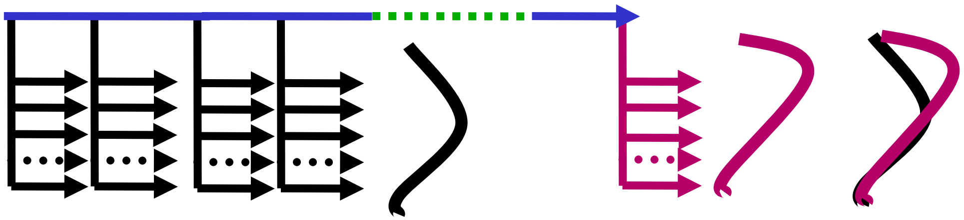
- **Non-stationary model error.**
  - Seasonal cycle dependence, which is known and catered for.
  - Other unknown dependences not considered: trends, changes in observing system
  - Drift depends of lead time. A large number of hindcasts is needed. This is even more costly in decadal forecasts.

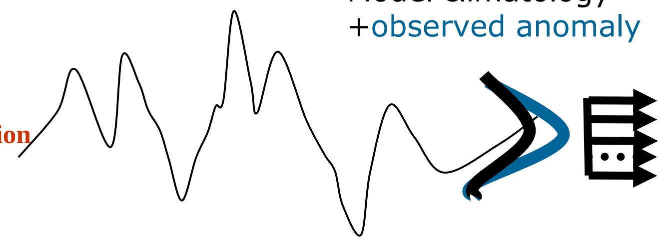
- **Initialization shock** can be larger than model bias

Non-linearities and non-stationarity can sometimes render the aposteriori 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: 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.

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

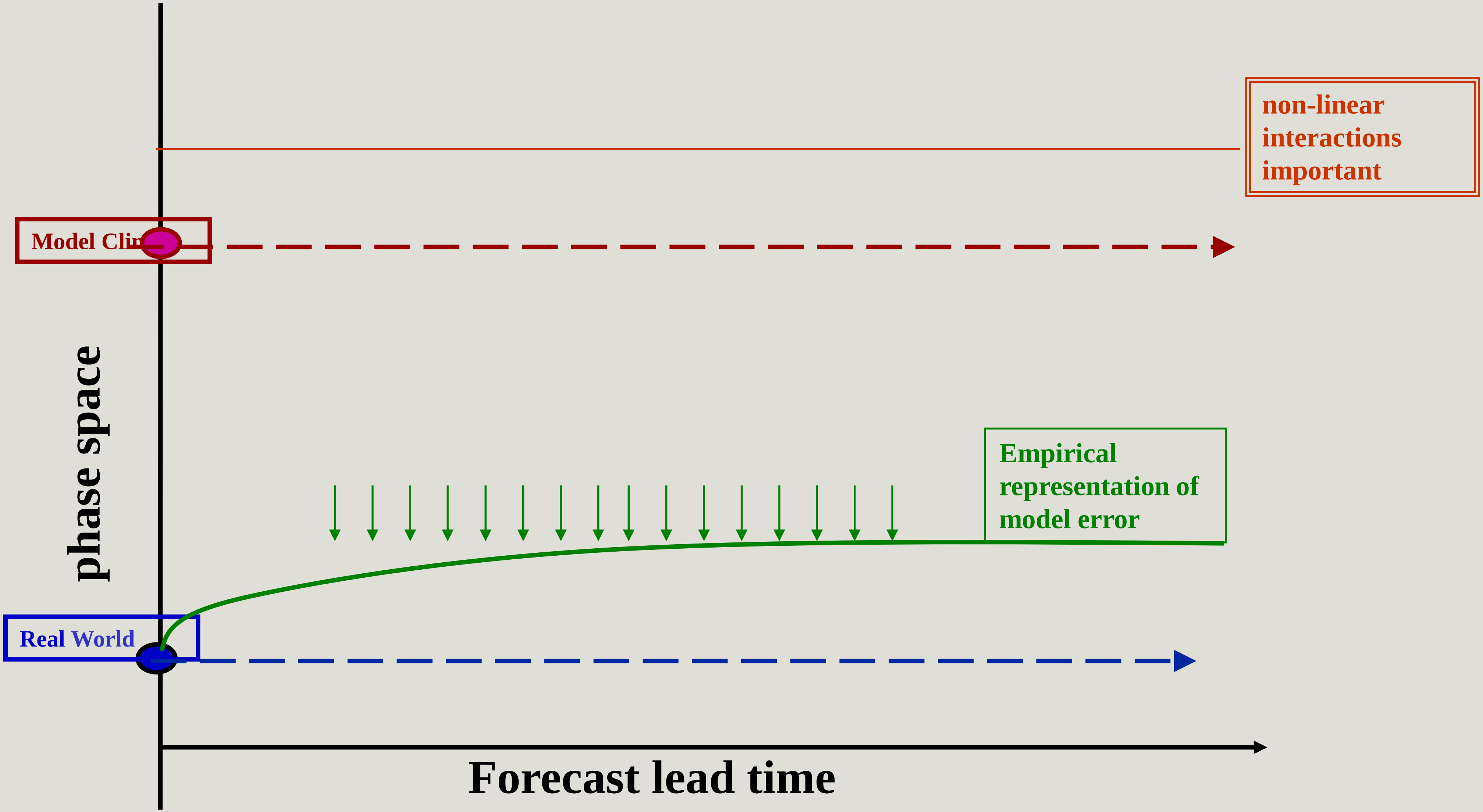
Acknowledgment of existence of model error during initialization.

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

Shock, Model error and non linearities



OUTLINE

- Initialization Shock
- Forecast drift and calibration
- Example: initialization of the ocean
- Approaches to initialize Earth System predictions
- Initialization in the context of forecasting strategies:
    - Dealing with model error

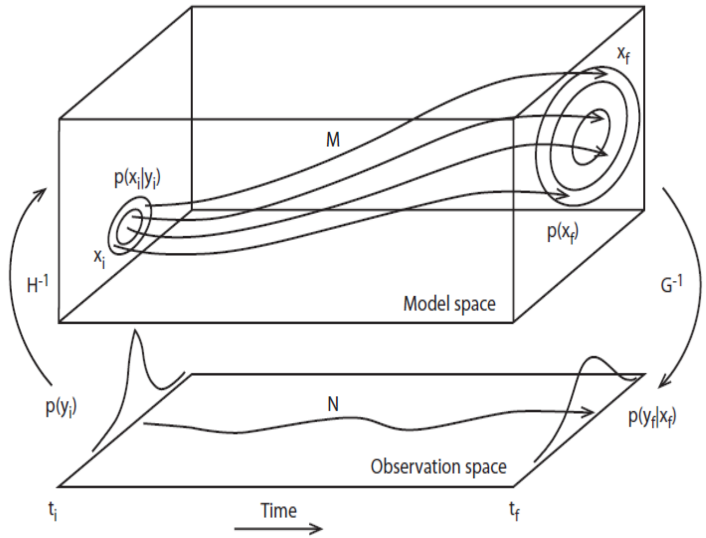
Treatment of system errors

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

✓ Stochastic parameterizations for sub-grid processes.
X Other missing processes and earth system components not represented
X Model bias is not targeted

1) Initialization *Data Assimilation*

✓ Initial uncertainty considered.
✓ Model uncertainty starts being considered.
✓ Observation uncertainty considered
✓ Observation bias considered
X Model bias ignored in atmosphere.



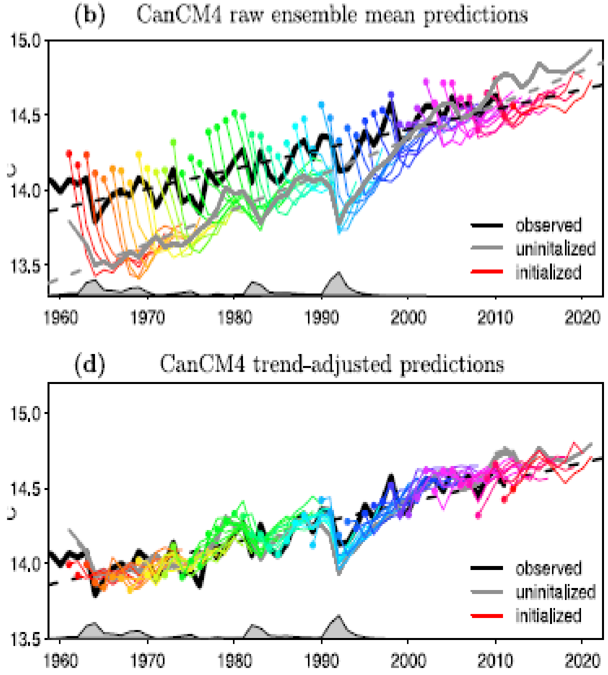
Model: $x = \bar{x} + \acute{x} + \varepsilon_x$
Observations: $y = \bar{y} + \acute{y} + \varepsilon_y$

3) Calibration *Forecast Assimilation*

- **Model Bias accounted for: removed a posterirori.** Stockdale et al 1997
- Model uncertainty considered (ensemble)
- Observation error neglected *
- Residuals are non stationary, non gaussian.
  Limitation to forecast skill
  calibration is more difficult

ECMWF

Calibration is complex if errors are non stationary



From Kharin eta la 2012

$$\tilde{x} = \bar{y} + \mathbf{K}(x - \bar{x}) + \mathbf{F}\varepsilon_x + \mathbf{T}(t) + \mathbf{G}(y_0)$$

Bias correction $(\bar{x} \neq \bar{y})$

K: linear transformation of anomalies

F: Adjustment of ensemble spread

T: detrending

G: other flow dependent corrections

Error in mean state errors degrades variability and forecast skill, making forecast errors non stationary and calibration difficult. Too many parameters

Can model error be treated more explicitly during the forecast process?

Stephenson et al 2005
Kharin et al 2012
Fukar et al 2014

Mean state error influencing model fidelity and skill

Correcting model biases leads to better representation of variability (or model fidelity) :

(several papers: D'Andrea and Vautard 2000, Balmaseda et al 2010, Scaife 2011, ….)

Correcting bias in tropical SST improves seasonal forecast skill of ENSO, tropical cyclones…

Magnusson et al 2012, Vecchi et al 2014:

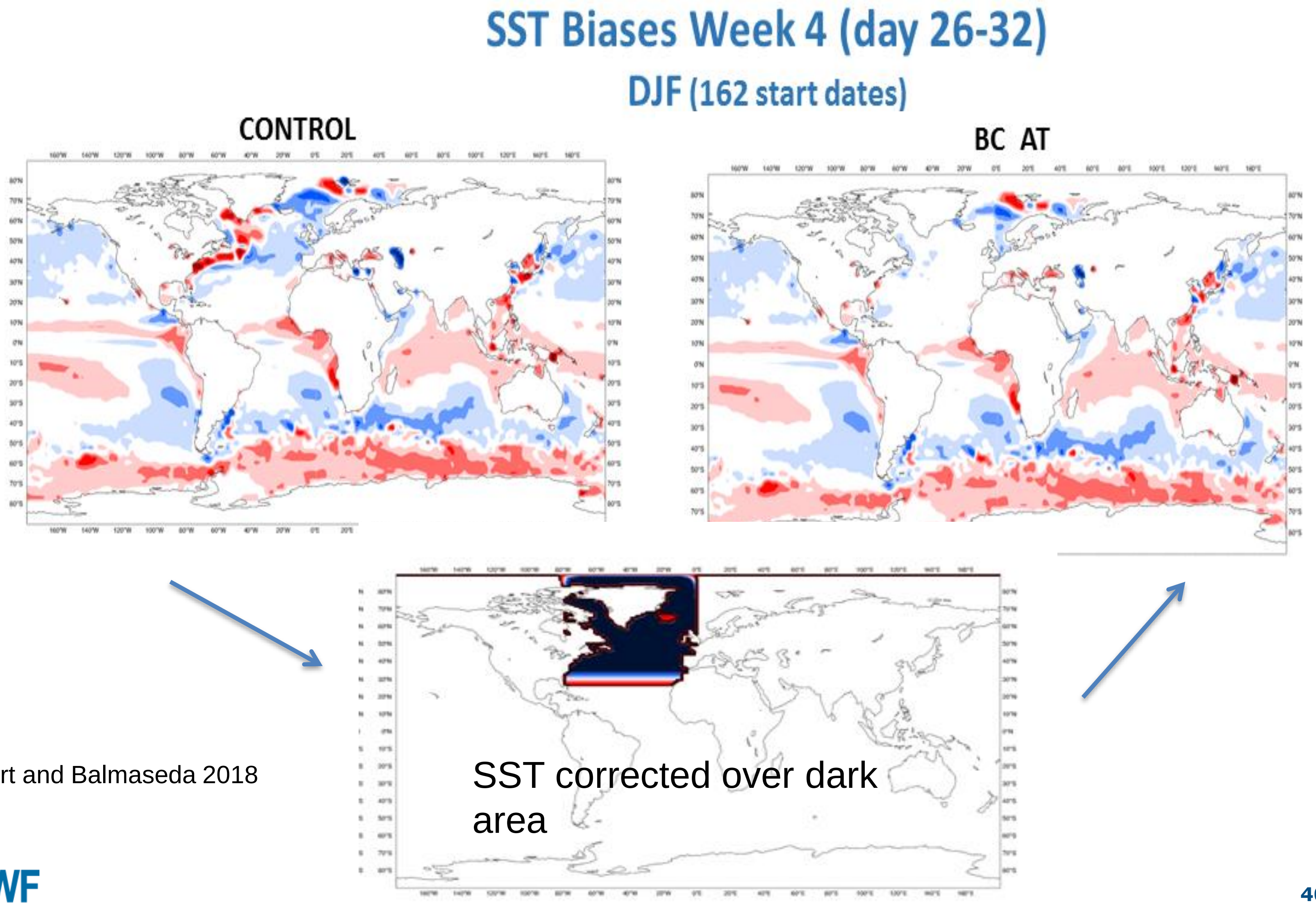
Correcting biases in atmosphere improves seasonal atmospheric predictability:

Kharin and Scinocca 2012

Correcting North Atlantic SST bias improves subseasonal skill over North Atlantic and Europe

Vitart and Balmaseda 2018

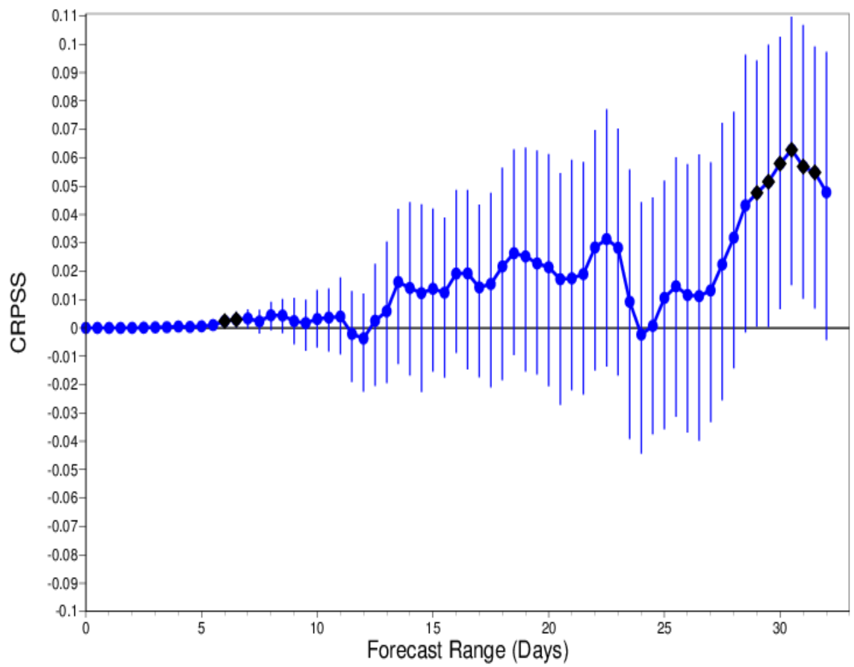
Non linear interactions: North Atlantic SST mean errors impact subseasonal forecast skill



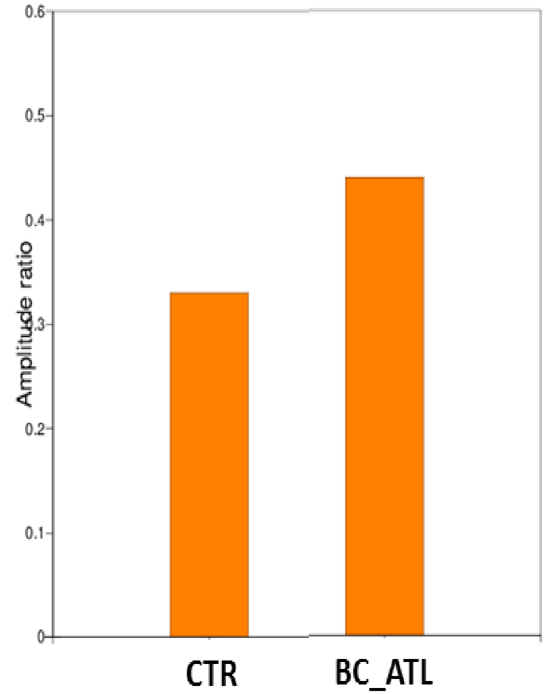
From Vitart and Balmaseda 2018

Non linear interactions: North Atlantic SST errors impact subseasonal forecast skill

NAO CRPSS differences
+ve is improvement



3rd pentad after MJO Phase 7



Main impact is on MJO/ NAO –ve teleconnections

From Vitart and Balmaseda 2018

Treatment of model error during the forecast

A. Stochastic parameterizations of sub-grid scale processes

- SPPT, SPP, SKEB, intrinsic stochastic parameterizations. See Berner et al 2017 for a review.
- They increase the ensemble spread (Leutbecher et al 2018). Important for tropical convection and ENSO (Weisheimer et al 2014).
- They do not tackle model bias explicitly, but change model climate (Christensen et al 2017, Berner et al 2018)
- Choice of parameters: tuned to calibrate ensemble spread or first principles
- Do not use optimal control based on observations

B. Model error estimation based on observational "optimal" control (or approximations): data assimilation to estimate model error

– D'Andrea and Vautard 2000

$$\mathbf{x}_i = M_i(\mathbf{x}_{i-1}) + \eta_i$$

– Piccolo and Cullen 2016

$$J(\eta) = \frac{1}{2}\sum_{i=1}^{n} \eta_i^T \mathbf{Q}^{-1} \eta_i + \frac{1}{2}\sum_{i=1}^{n} (H_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}^{-1} (H_i(\mathbf{x}_i) - \mathbf{y}_i)$$

-Proxi: nugdging terms as in Kharin and Scinocca 2012…

Data assimilation can be used to estimate model error:

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

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

Step 1) Analysis of assimilation increments to derive an empirical stochastic model for model error.

$$\mathbf{b}^f{}_k = \overline{\mathbf{b}}_k + \mathbf{b}'^f_k$$

Step2) Apply that empirical model during the coupled model integration

Adaptive term; AR1 process, constrained by observations

$$\boldsymbol{b}'^f{}_k = \mathbf{A}\, \boldsymbol{b}'^f{}_{k-1} + \boldsymbol{\varepsilon}_k$$

$$\boldsymbol{b}'^a{}_k = \mathbf{A}\, \boldsymbol{b}'^f{}_{k-1} + \mathbf{L}\,[\mathbf{y}\text{-}\mathbf{H}(\mathbf{x} + \boldsymbol{b}'^f{}_k\ )]$$

For the above to make sense, the data assimilation should be coupled.

Dee and DaSilva 1998
Dee and Todling 2001
Tremolet 2007
Balmaseda et al 2007

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
- Objective of a forecasting system: stationary forecast errors, so they are easy to calibrate.