# A model for generating atmosphere forcing perturbations

Isabelle Mirouze, Andrea Storto
CMCC

*DA-TT workshop, La Spezia, 11-13 October 2017*

# Plan

- Motivations
- Training Data
- Design of the Model
- Test results

# Generating an ensemble

- Running ensemble of reanalyses or forecasts has become prevalent despite the cost
  - Gives information on the background error structure
  - Introduces some flow-dependency in the system
  - OSEs and OSSEs (AtlantOS project)
- Running an ensemble requires generating perturbations to represent errors of the system
  - First guess: breeding methods, optimal perturbations, ...
  - Observations: random error drawn from a normal distribution and covariance $\mathbf{R}$
  - Model:
    - Equations and parameters: stochastic equations
    - Subsidiary conditions such as atmosphere forcing

# Atmosphere forcing perturbations

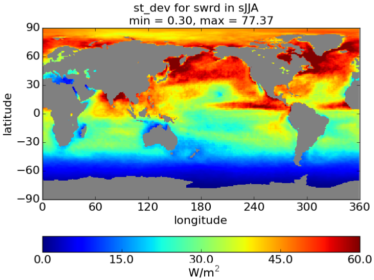
- Ocean models are generally forced by NWP products
  - C-GLORS: Short and long wave radiation fluxes, temperature, humidity, wind, precipitations provided by ERA-Interim
- How to estimate the error of these fields?
  - Compare to observations
    - Requires (at least) daily observations globally well distributed for all variables
    - Represents the error if the observations are accurate enough
  - Compare to another NWP product
    - Reflects the differences between the systems that produce them and is hence a good proxy of the error
- How to generate perturbations representing these errors?
  - Use directly the comparison
  - Design a model based on the comparison

# Ensemble at CMCC

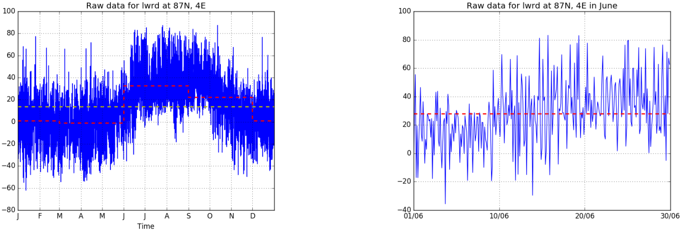
- CMCC operational ensemble systems
  - Seasonal forecasting system
- Generation of atmosphere forcing perturbations
  - Calculate differences between ERA-Interim and NCEP-R2 reanalyses from 1989 to 2014 on a $2 \times 2$ degree grid
  - Remove monthly mean and choose randomly one of the differences of the month
  - Introduce time correlation as an AR(1) process
- New method for generating the perturbations
  - Specifications
    - Perturbations should be unbiased and balanced
    - Perturbations structure should be controlled
    - Perturbations should be easy and quick to generate
  - Design a model based on linear regression

# Training data

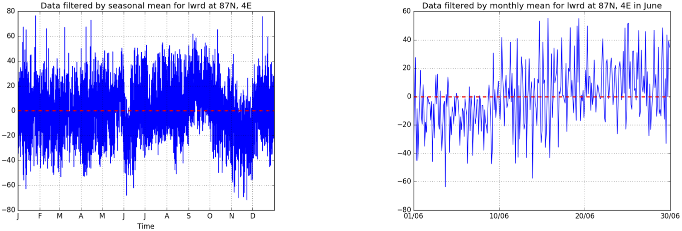
- Differences between ERA-Interim (ECMWF) and MERRA (NASA) reanalyses
- MERRA grid ($0.5 \times 0.667$ degree) interpolated on ERA-Interim grid ($0.75 \times 0.75$ degree)
- 10-year set from 2004 to 2013
- 3-hourly outputs averaged to daily outputs
  - temperature at 2 m
  - humidity at 2 m
  - u and v wind at 10 m
- Daily outputs:
  - short wave rad. flux
  - long wave rad. flux
  - precipitations
  - snow



# Unbiasing the data by removing the mean

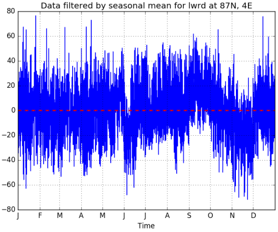


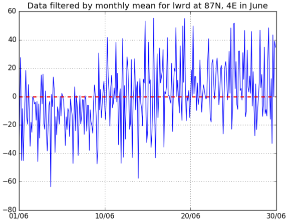

Long wave radiation flux differences at 87N, 4E and the different means



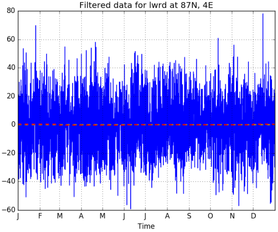

Long wave radiation flux differences at 87N, 4E filtered by the seasonal (left) or monthly (right) mean

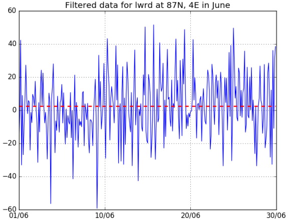
# Unbiasing the data with a high-pass filter




Long wave radiation flux differences at 87N, 4E filtered by the seasonal (left) or monthly (right) mean

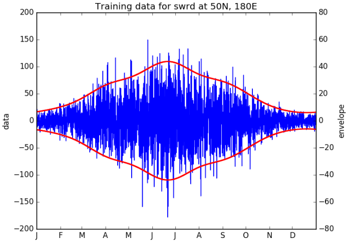



Long wave radiation flux differences at 87N, 4E filtered by a high-pass filter ×2 with 10-day cutoff

# Independent variable: swrd

- Time series for each grid point modelled by a normal distribution
  - Choice: short wave radiation flux
- The standard deviation can vary during the year
  - To redress the data
  - To low-pass filter $\times 2$ with 1-day cutoff
  - To split the wave envelope into amplitude categories
  - For each category
    - To test the normality ($-0.5 \leq$ skew, kurtosis $\leq 0.5$)
    - To possibly remove outliers (10% max)
    - To calculate standard dev.




Model: To define the category a particular day belongs to and to associate the standard deviation of this category

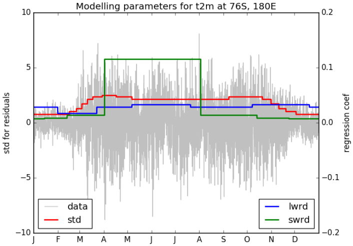
# Dependent variable

- To calculate the daily correlation between the dependent variable and the other variables
  - To choose 1 or 2 (independent) variables the dependent variable will be modelled from
- For each independent variable, the regression coefficient can vary during the year
  - To low-pass filter $\times 2$ with 1-day cutoff the correlation signal
  - To split the wave envelope into amplitude categories
  - To calculate a regression coefficient for each category
- To model the residual as explained previously for the independent variable

Model: To define the categories a particular day belongs to for each of its independent variables and its residual, and to associate the regression coefficients and the standard deviation of these categories

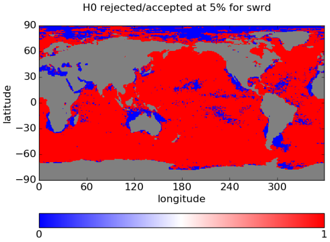
# Generation of one day perturbations

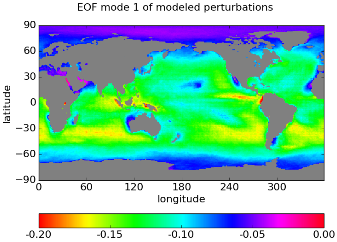
- swrd
  - To draw a random number from a normal distribution $\mathcal{N}(0,1)$ and apply the standard deviation of each grid point
- lwrd = f(swrd)
  - To apply the regression coefficients to swrd for each grid point
  - To add a residual by drawing a random number from a normal distribution $\mathcal{N}(0,1)$ and apply the standard deviation of each grid point
- u10m = f(lwrd)
- v10m = f(lwrd)
- t2m = f(lwrd,swrd)
- q2m = f(t2m,lwrd)
- precip = f(q2m,swrd)
- snow = f(precip)



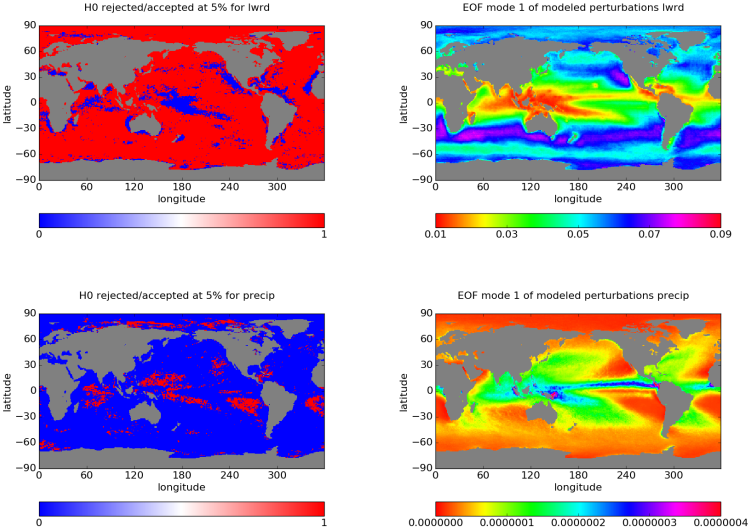
# Validation tests

- To generate 3 years of perturbations from 2001 to 2003
- To compare their distribution with the training data distribution → Kolmogorov-Smirnov test
- To control the structure → EOF test



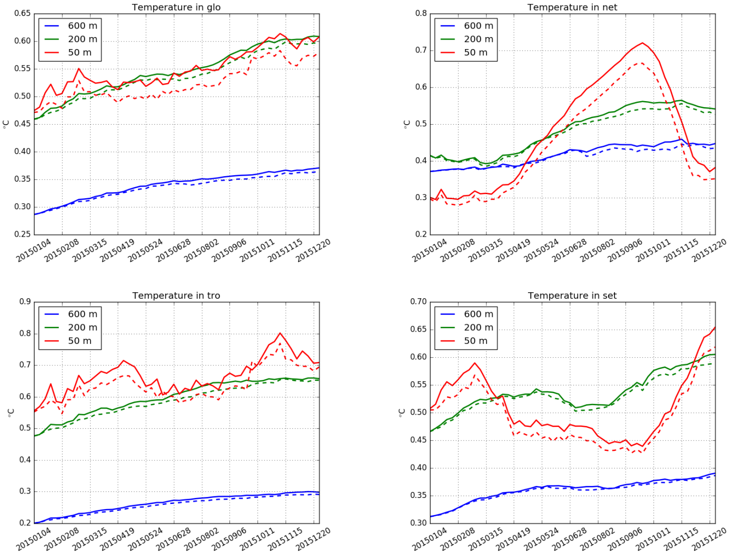
# Validation tests



# Reanalysis experiments

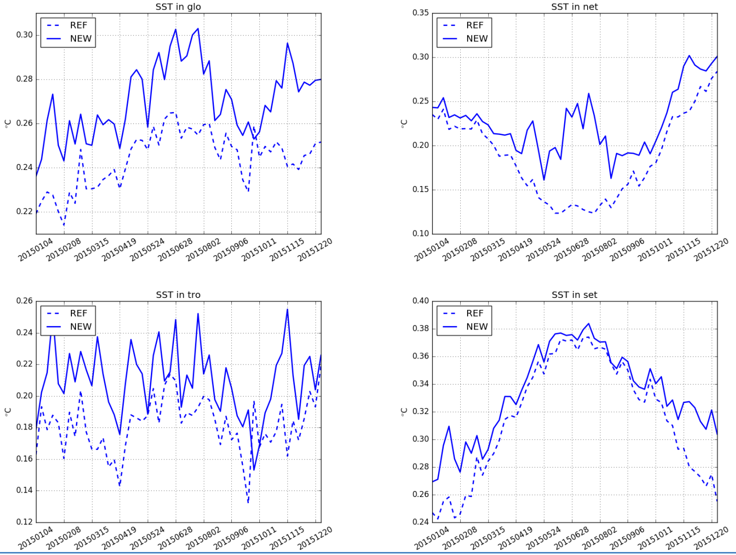
- C-GLORS framework for 1-year reanalysis (2015)
  - NEMO v3.6, global configuration at 1/4 degree
  - OceanVar: 3DVAR-FGAT, 7-day window
  - Observations: SST, temperature and salinity profiles, SLA
- Ensemble
  - 6 members
  - Observation perturbations from specified **R**
  - Equation of the state perturbations
  - 1-year spinup (2014)
- Experiments
  - Ref: no atmosphere forcing perturbations
  - New: atmosphere forcing perturbations from the model
  - Old: atmosphere forcing perturbations from the previous method (1 member only)

# Comparison of the spread for temperature

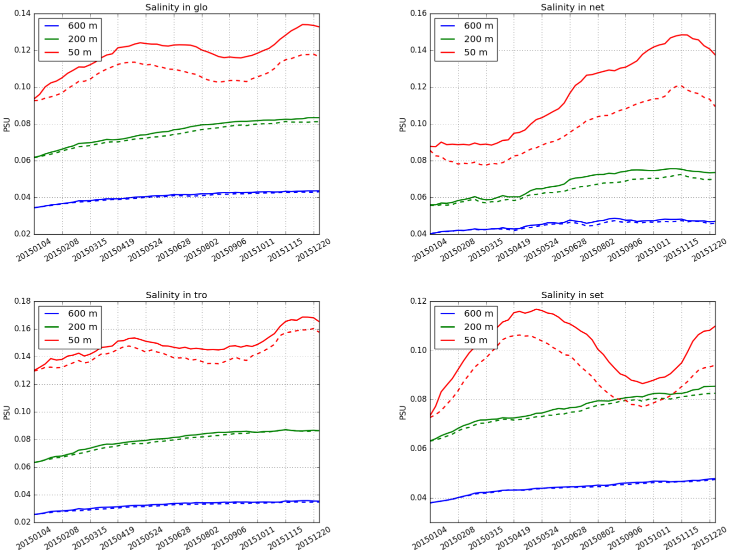



Ref: - - New: —

# Comparison of the spread for SST

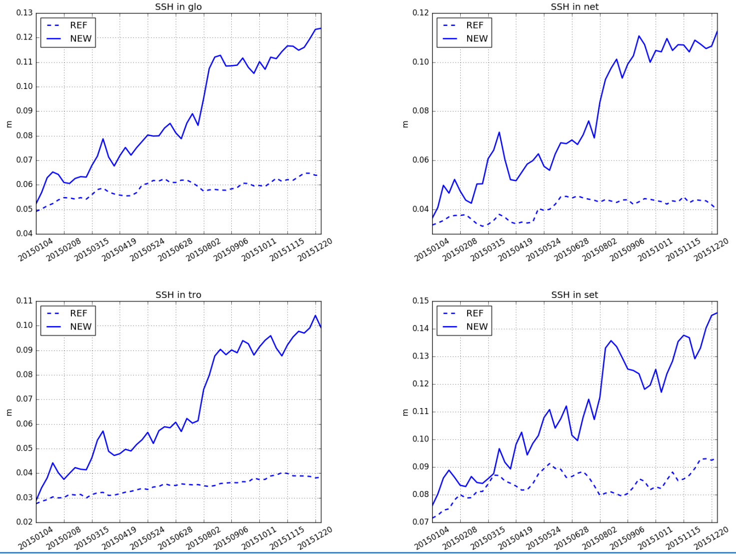


# Comparison of the spread for salinity

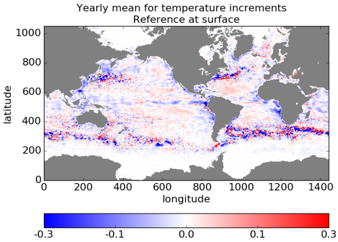


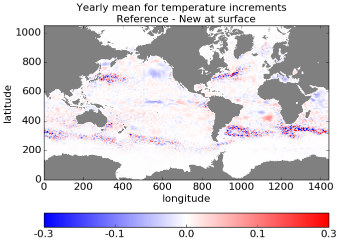

Ref: - -     New: —

# Comparison of the spread for SSH

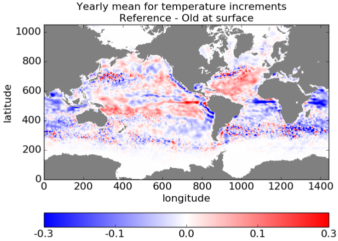


# Yearly mean for temperature increments at 0 m

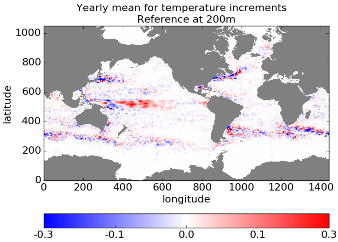


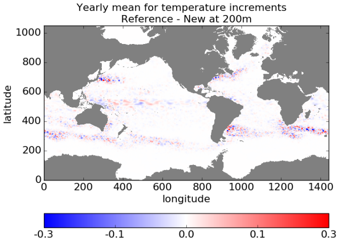




Presence of a bias in the Old experiment

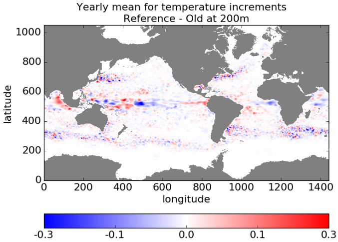


# Yearly mean for temperature increments at 200 m

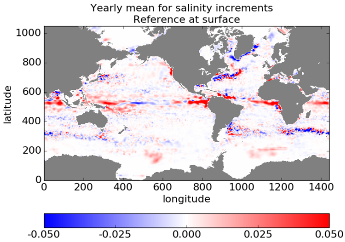


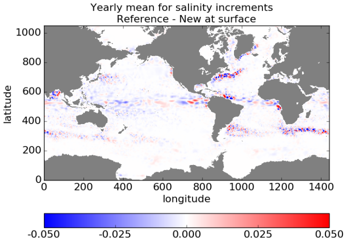




The bias in the Old experiment is fading

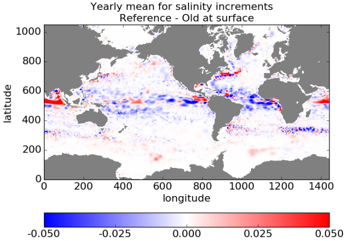


# Yearly mean for salinity increments at 0 m

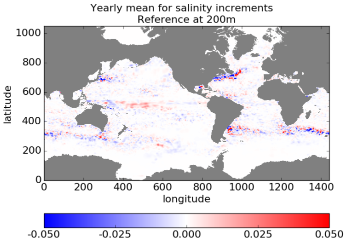


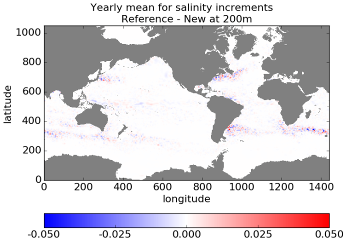




Presence of a bias in the Old experiment

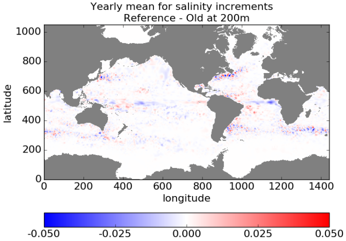


# Yearly mean for salinity increments at 200 m




The bias in the Old experiment is fading



# Summary

- New model to generate atmosphere forcing perturbations for ensemble simulations
  - Model based on linear regression
  - Training data: 10-year differences between ERA-Interim and MERRA reanalyses
  - Perturbations unbiased, balance relationship between variables, spatial structure controlled
- Kolmogorov-Smirnov test
  - General good match between the distributions for all variables except precipitations and snow
  - Difficulties where the training data show a high kurtosis for the period considered
  - Difficulties where the training data do not show any pattern from one year to another
- EOF test
  - First mode matches the standard deviation of the training data for all variables

# Summary

- 1-year ensemble reanalysis
  - Increase of the spread in the first hundred meters
  - Yearly-mean of the increments do not show any biases
- The model for generating atmosphere forcing perturbations is satisfying although some improvements could be done, especially for the precipitation and snow
- The model is currently used in the AtlantOS project for running OSSEs with ensemble reanalyses

# Thanks

AtlantOS

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici