Ocean Data Assimilation and analysis

DA training course 2020

Hao Zuo

With inputs from M Balmaseda, K Mogensen, M Chrust, P Browne, E de Boisseson, R Buizza and many others

hao.zuo@ecmwf.int
Outline

- Ocean system and ocean observations
- NEMOVAR ocean data assimilation system
- Bias correction in ODA
- Assimilation of sea-level data
- Assimilation of SST data
- Assimilation of sea-ice data
- Ocean (re)analysis system and its applications
Why do we do Ocean DA?

- **Forecasting:** initialization of coupled models
  - NWP, monthly, seasonal, decadal
  - Seasonal forecasts need calibration

- **Towards coupled DA system (weakly -> quasi-strong -> strong …)**
  - See Phil’s presentation

- **Climate application:** reconstruct & monitor the ocean (re-analysis)

- **Verification/evaluation of Global Ocean observing network (OSE/OSSE)**

- **Other applications**
  - Commercial applications (oil rigs, ship route …), safety and rescue, environmental (algii blooms, spills)
Ocean versus Atmosphere

**Spatial/time scales** The radius of deformation in the ocean is small (~30km) compared to the atmosphere (~3000km). Time scales vary from hours (mixing) to decades (overturning).

**Ocean is a data sparse system**, in-situ observation is limited and mostly covers upper ocean only, satellite observation only covers ocean surface.

**The ocean is forced at the surface and land boundary**, by the wind/waves, heating/cooling and fresh-water fluxes

  Uncertainty in forcing fluxes contributes to uncertainty in model results.

**The ocean is strongly stratified in the vertical**, although deep convection also occurs

  Density is determined by Temperature and Salinity

**The ocean has continental boundaries**: dealing with them is not trivial in data assimilation
Ocean time scales: from hours to centuries

Wind Driven: Gyres, Western Boundary Currents, Upwelling regions (coastal, equatorial), Ekman pumping and subduction

Density Driven: Thermohaline Circulation

Ocean is a system with much longer memory but slow response compared to the Atmosphere
Ocean spatial scales

Satellite image of SST in the North Atlantic Ocean (from NOAA)

1/50 degree Ocean surface relative vorticity (CHASSIGNET and Xu, 2017)

Ocean variables with various spatial scales: from hundred meters to hundreds of km
Ocean is forced by external forcings

Impact of Atmospheric forcing
SST spread due to atmospheric forcing perturbation

Impact of land freshwater input
SSS spread due to land freshwater input perturbation

de Boisséson et al., 2020

Zuo et al., 2020
Ocean is a data sparse system

Ocean observation is about 1/1000 to 1/10000 smaller than Atmospheric observation
Ocean in-situ observations

New observations types are emerging: gliders, Deep Argo, BioArgo, drifter, saildrone …
The Global Ocean Observing System (GOOS)

Ocean in-situ observations used in OCEAN5 (5-days in Feb 2019)

Total observations = 4144
Impact on ocean data assimilation system

Maps of normalized RMSD of Temperature (upper 700m) in OSEs

a) NoMooring
b) NoShip

c) NoArgo
d) NoInsitu

Moored buoys
Ship-based (CTD/XBT/MBT)

Argo Floats
All in-situ

Zuo et al., 2019
Satellite ocean surface observations

Sea-ice concentration

Sea-ice thickness

Ocean model

3DVar

Nudging

Sea-Level Anomaly (Altimeter)

SST (IR, PMW)

Satellite ocean surface observations

Ocean model

Sea-ice concentration

Sea-ice thickness

3DVar

Nudging

Sea-Level Anomaly (Altimeter)

SST (IR, PMW)
Observations impact on the ocean state estimation

Observations are essential for improving initialization.

Assimilation of ocean in-situ observations helps to constrain the 3D ocean, therefore providing better estimation of the ocean initial condition for the coupled forecasting system.
Impact of Ocean DA in Seasonal Forecast

A proper initialisation played a key role in seasonal forecasts

SEAS5 is the new ECMWF seasonal forecasting systems (Johnson et al 2018, GMD)
SEAS5 initialized by Ocean Reanalyses ORAS5 (Zuo et al, 2018)

SEAS5-NoOObs is initialized by an “Ocean Simulation” where Ocean observations have are not assimilated (Only winds and SST)
• Ocean system and ocean observations
• **NEMOVAR Ocean data assimilation system**
• Bias correction in ODA
• Assimilation of sea-level data
• Assimilation of SST data
• Assimilation of sea-ice data
• Ocean (re)analysis system and its applications
Ocean DA at ECMWF: NEMOVAR

- The “NEMOVAR” assimilation system used in ECMWF.
  - Variational DA system as a collaborative project among CERFACS, ECMWF, INRIA and the Met Office for assimilation into the NEMO ocean model.
    - Solves a linearized version of the full non-linear cost function.
    - Incremental 3D-Var FGAT running operational, 4D-Var in research model
  - Background correlation model based diffusion operators
  - Background errors are correlated between different variables through balance operator
- To avoid initialization shock increments are typically applied via Incremental Analysis Update (IAU) which applies the increments as a forcing term over a period of time.
Figure 1: Schematic illustration of the procedure used to cycle 3D-Var. On each cycle $c$, the model is integrated from $t_0$ to $t_N$ starting from a background initial condition $x^b_c(t_0)$ (grey dots) to produce the background trajectory $x^b_c(t_i)$ (black solid curve). The difference between the observations $y^o_{c,i}$ (black dots) and their background counterpart ($H_{c,i}x^b_c(t_i)$) is computed (represented by the vertical thin dotted lines) for use in the 3D-Var FGAT minimization. After minimization, the model integration is repeated from the same initial condition ($x^b_c(t_0)$) but with the analysis increment applied using IAU. This produces the analysis trajectory $x^a_c(t_i)$ (grey dashed curve). The updated model state $x^a_c(t_N)$ at the end of cycle $c$ is then used as the background initial condition for the next cycle $c+1$ (grey dots).
**NEMOVAR: Linearized Cost function**

\[
J[\delta w] = \frac{1}{2} \delta w^T B^{-1} \delta w + \frac{1}{2} (G \delta w - d)^T R^{-1} (G \delta w - d)
\]

\[
y^o = [(y^o_0)^T \cdots (y^o_i)^T \cdots (y^o_N)^T]^T \quad \rightarrow \text{4D observation array}
\]

\[
\delta w = w - w^b \quad \rightarrow \text{w is the control vector}
\]

\[
d = y^o - G(w^b) \quad \rightarrow \text{Departure vector}
\]

- **Balance operator**: convert to w space, B becomes block diagonal, representing the spatial covariance model.
- **Diffusion operator**: The spatial covariances is specified by diffusion operator (Weaver and Courtier 2001)

\[
G(w) = \begin{pmatrix}
    \vdots \\
    G_i(w) \\
    \vdots
\end{pmatrix} = \begin{pmatrix}
    \vdots \\
    H_i[M(t_i, t_0)[K(w)]] \\
    \vdots
\end{pmatrix}
\]

\[
\delta w^a \approx B G^T (GBG^T + R)^{-1} d.
\]

\[
\delta x^a = K(w^b + \delta w^a) - K(w^b) \approx K \delta w^a
\]

\[
x^a(t_i) = M(t_i, t_{i-1})[x^a(t_{i-1}), F_i \delta x^a]
\]

**Solution:**

IAU, Bloom et al 1996

Weaver et al 2003, 2005
Daget et al 2009
Mogensen et al 2012
Balmaseda et al 2013
Define the balance operator symbolically by the sequence of equations

\[
\begin{align*}
\delta T & = \delta T \\
\delta S & = K_{S,T}^{b} \delta T + \delta S_{U} = \delta S_{B} + \delta S_{U} \\
\delta \eta & = K_{\eta,\rho} \delta \rho + \delta \eta_{U} = \delta \eta_{B} + \delta \eta_{U} \\
\delta u & = K_{u,\rho} \delta \rho + \delta u_{U} = \delta u_{B} + \delta u_{U} \\
\delta v & = K_{v,\rho} \delta \rho + \delta v_{U} = \delta v_{B} + \delta v_{U} \\
\delta \rho & = K_{\rho,T}^{b} \delta T + K_{\rho,S}^{b} \delta S \\
\delta p & = K_{p,\rho} \delta \rho + K_{p,\eta} \delta \eta
\end{align*}
\]

Treated as approximately mutually independent without cross correlations.

Weaver et al., 2005, QJRMS
Salinity balance (approx. T-S conservation)

\[ \partial S^k_B = \gamma^{k-1} \left( \frac{\partial S}{\partial z} \right)_{S=S^{k-1}} \left( \frac{\partial z}{\partial T} \right)_{T=T^{k-1}} \partial T^k \]

T/S/SSH balance: vertical displacement of the profile.

Ricci et al. 2005
NEMOVAR: BGE covariance parameters

General B formulation in NEMOVAR

\[ B = \alpha B_m + \beta B_e \]
\[ B_m = K_b D_{m}^{1/2} C_{m} D_{m}^{1/2} K_{b}^{T} \]

\( B_m \) is covariance model for each variable, \( C_m \) is correlation matrix (including diffusion operator), and \( D_m \) is a diagonal matrix of variance (block-diagonal).

Horizontal correlation length-scales used in ORAP5

\[ C_{x}^{1/2} = \Gamma_{x}^{1/2} L_{x}^{1/2} W_{x}^{-1/2} \]

Representation of diffusion operator that requires de-correlation length-scales

Rossby radius of deformation, phase speed = 2.7 m/s

Zuo et al., 2015
NEMOVAR: BGE covariance parameters

ORAS5 background error (BGE) standard deviations for unbalanced variables at 100m

\[ \sigma_T^b = \begin{cases} \max \left\{ \sigma_{T}^b, \sigma_{T}^{ml} \right\} & \text{if } z \geq -D_{ml} \\ \max \left\{ \sigma_{T}^b, \sigma_{T}^{do} \right\} & \text{if } z < -D_{ml} \end{cases} \]

\[ \hat{\sigma}_T^b = \min \left\{ \left| \frac{\partial T^b}{\partial z} \right| \delta z, \sigma_T^{max} \right\}, \]

\[ \sigma_{SU}^b = \begin{cases} \sigma_{SU}^{max} & \text{if } z \geq -D_{TS} \\ \max \left\{ \hat{\sigma}_{SU}^b, \sigma_{SU}^{do} \right\} & \text{if } z < -D_{TS} \end{cases} \]

\[ D_{TS} \text{ is the depth where } \max ds/dt \text{ happens} \]

\[ \hat{\sigma}_{SU}^b = B(z) \sigma_{SU}^{max} \]

\[ B(z) = 0.1 + 0.45 \times \left\{ 1 - \tanh \left( 2 \ln \left( \frac{z}{D_{TS}} \right) \right) \right\} \]

Mogensen et al., 2012
Cooper and Haines, 1996

Zuo et al., 2019

\( D_{TS} \) is the depth where max ds/dt happens
NEMOVAR: Horizontal cross-correlation of T at 100m

- From single observation of temperature experiment.
- The specific background determines the shape due to the balance relations.
- S, U, V, SSH increments are from balance with T only.
On-going development in ocean DA

- Flow-dependent B ($B_m + B_e$, hybrid ...)
- Multi-grid capability
- Multiple spatial scales
- Enhanced perturbation scheme (SPP, SPPT SKEB)
- 4DVar (tangent linear and adjoint model)

BGE std dev for unbalanced variables
- parametrized with analytical function
- ensemble estimation from ORAS5

Temperature
- temperature BKG err SDV(Rs) 341 75–200m

Salinity
- salinity BKG err SDV(Rs) 341 75–200m
• Ocean system and ocean observations
• NEMOVAR ocean data assimilation system
• **Bias correction in ODA**
• Assimilation of sea-level data
• Assimilation of SST data
• Assimilation of sea-ice data
• Ocean (re)analysis system and its applications
Why do we need bias correction in ODA

- To correct systematic errors in models/forcing and bias introduced by DA
- To mitigate changes in the observing system
Bias Correction scheme

Bias term include two parts, (a) a prior bias for systematic errors, and (b) a temporal variable bias for slow evolving signals (Balmaseda et al, 2007)

$$b^f_k = \bar{b}_k + b'_k$$

- **Seasonal term**, estimated offline from rich-data (Argo) Period
- **Slow varying term**, estimated online from assimilation increments $d_k$

$$b'_k = \alpha b'_{k-1} + A(y) \beta d_k$$

$A(y)$: Partition of bias into T/S and pressure gradient.

A prior bias ($\bar{b}_k$) estimation (ORAS5)

(Zuo et al 2018)
Temporal variable bias term

The latitude dependent partition coefficients determine the proportion of online bias corrections applied directly on T/S, and on pressure term. These values ensure that at low latitude the dominant bias term is pressure correction.

$$A = \begin{bmatrix} a^{r,T} & 0 & 0 & 0 & 0 \\ 0 & a^{r,S} & 0 & 0 & 0 \\ 0 & 0 & a^{p,T} & 0 & 0 \\ 0 & 0 & 0 & a^{p,S} & 0 \end{bmatrix}$$

A: Partition matrix, The coefficients in A is latitude dependent in NEMOVAR

**Zuo et al., 2015**
Bias correction in ODA is essential, and in particular important for mitigating spurious signals introduced due to changes in the observing system.
• Ocean system and ocean observations
• NEMOVAR Ocean data assimilation system
• Bias correction in ODA
• **Assimilation of sea-level data**
• Assimilation of SST data
• Assimilation of sea-ice data
• Ocean (re)analysis system and its applications
**Assimilation of Sea Surface Height (SSH)**

**Altimeter measures SSH** (respect reference ellipsoid)

**Model represents** $\eta$ (ssh referred to the Geoid)

**SSH-Geoid** = $\eta$

Geoid was poorly known (until recent years) and changes in time

---

Alternative: Assimilate Sea Level Anomalies (SLA) respect a time mean

**Obs:** SSH anomalies = $\text{SSH-MSSH} = \text{Obs SLA}$

**Mod:** $\eta$ anomalies = $\eta – \text{MDT} = \text{Mod SLA}$

Where: $\text{MSSH} =$ Temporal Mean SSH ;

$\text{MDT} =$ Temporal Mean of model SL Mean Dynamic Topography

**MSSH – Geoid = MDT**
Assimilation of SSH: MDT

- $MDT_m$: model MDT as mean($SSH_m$), mean model biases not corrected (Balmaseda et al., 2013)
- $MDT_o$: observation MDT as mean($SSH_o$), observation bias not corrected (Waters et al., 2015 and Lellouche et al., 2018)
- bias corr. $MDT_o$: observation biases corrected (Lea et al., 2008)

$MDT_m - MDT_o$ (in m)

$T \Delta RMSE (O-B): bias corr. MDT_o - MDT_m$

Temperature RMS error 242 0–50m
Assimilation of SSH: pre-process obs

- The SLA along track data has very high spatial (9-14km) resolution for the operational ocean assimilation systems.
  - Features in the data which the model can not represent
  - “Overfitting” to SLA obs
- This can be dealt with in different ways:
  - Inflate the observation error
  - Construction of “superobs” or thinning

Inflate OBE std dev of SSH

Total observations = 24579

Thinning of SLA obs

ERS-2: 13005  Envisat: 0
Jason-1 N: 0  Jason-2: 0
Envisat N: 0  T/P: 13219
Assimilation of SSH: impact on ocean states

Assimilation of SSH improves simulated ocean states

Temporal correlation (monthly) to AVISO data

Zuo et al., 2018
• Ocean system and ocean observations
• NEMOVAR Ocean data assimilation system
• Bias correction in ODA
• Assimilation of sea-level data
• Assimilation of SST data
• Assimilation of sea-ice data
• Ocean (re)analysis system and its applications
Assimilation of SST: nudging

A simple nudging scheme to L4 objective analysis data (e.g. OSTIA)

\[ Q_{ns} = Q_{ns}^o + \frac{dQ}{dT}(SST_{MODEL} - SST_{TARGET}) \]

Haney 1917

- Non-solar total heat flux
- Fixed negative feedback coefficient

\[ SST(\text{model}) = T_{\text{layer}}(x) \]

\[ SST(\text{target}) = SST_{\text{find}} \]

ESA CCI2 SST data (Jan 2016)
Impact of SST nudging

- Overall very effective except for some areas with weak vertical stratification
- Not accounting complicated error characteristics in the L4 SST analysis
- Not accounting vertical correlation when apply SST constrain in the surface

Global mean SST

SST bias: free run - OSTIA

SST bias: ORAS5 - OSTIA
Flow-dependent SST nudging

\[ Q_{ns}^{sst} = \frac{dQ}{dT} (SST_{model} - SST_{target}) \]

A flow-dependent relaxation coefficient can be obtained by filtering of \( Q_{ns}^{sst} \) in order to avoid introducing spurious convection in regions where surface stratification is very weak.

Anomaly correlation map of T2M, verified against ERA-int

Johnson et al., 2019
Assimilation of L2P SST with 3DVAR

• In general 3DVar DA does a better job than SST nudging
• Pre-processing including bias correction to SST observation is critical
• Flow-dependent vertical correlation length-scale is vital in determining vertical propagation of the T increment

Daily L2P SST obs (24 Nov 2010)

Temperature increments

SST Nudging

SST assimilation

\~1e6/day
• Ocean system and ocean observations
• NEMOVAR Ocean data assimilation system
• Bias correction in ODA
• Assimilation of sea-level data
• Assimilation of SST data
• **Assimilation of sea-ice data**
• Ocean (re)analysis system and its applications
Assimilation of SIC: L4 gridded data

Sea-Ice Concentration data from L4 analysis is assimilated through 3DVar scheme

- Treated as univariate
- Pre-thinned via regular or stratified random sampling
- Assimilated through outer-loop coupling in NEMO-LIM2
Assimilation of SIC: L4 gridded data

SIC bias (1980-2016)
Ref data: OSI-SAF 430

Positive impact on both SIC and SIT state

SIT bias (2011-2016)
Ref data: CS2SMOS merged data
Assimilation of SIC: L3 data

Daily SIC on 20130118

L4 SIC analysis

L3 SIC data

L4 analysis: with filtering, masking, infilling to produce a gap-free product

with 10km resolution there is \(~1\) milion obs per day from L3 OSI-SAF, with no infilling created observation
Assimilation of sea-ice thickness (SIT): nudging scheme

\[
SIT^n = SIT^m - \left( \frac{\Delta t}{\tau} (SIT^m - SIT^o) \right)
\]

where \(SIT^n\) is the nudged thickness, \(SIT^m\) is the modelled thickness, \(SIT^o\) is the observed thickness, \(\tau\) is the nudging coefficient.

Difference in forecast Integrated Ice Edge Error (2011-2016, verified against OSI-401b)

with SIT nudging – No SIT nudging

Sea-ice forecasts

Balan Sarojini, et al. 2019
• Ocean system and ocean observations
• NEMOVAR Ocean data assimilation system
• Bias correction in ODA
• Assimilation of sea-level data
• Assimilation of SST data
• Assimilation of sea-ice data

• **Ocean (re)analysis system and its applications**
OCEAN5 is the 5th generation of ECMWF ocean and sea-ice ensemble reanalysis-analysis system (Zuo et al., 2018, 2019).

- Ocean: NEMOv3.4
- Sea-ice: LIM2
- Resolution: 1/4 degree with 75 levels
- Assimilation: 3DVAR-FGAT
- 5 ensemble member
- Forcing: ERA-int
Application of Ocean analysis: coupled forecasts

OCEAN5 provides ocean and sea-ice initial conditions for all ECMWF coupled forecasting system: (ENS, HRES, Seasonal). OCEAN5 also provides SST and SIC conditions for the ECMWF atmospheric analysis system (Browne et al., 2018)

- Gain about 2 months in ENSO prediction
- Without Ocean observation and DA, we would lose about 15 years of progress.
Application of Ocean reanalysis: calibration and reforecasts

- Correcting model error
- Extreme Events
- Tailored products (health, energy, agriculture)

**Ocean/Atmosphere reanalyses**

- Hindcasts, needed to estimate climatological PDF, require a historical ocean and atmospheric reanalyses
- Consistency between historical and real-time initial conditions is required.

**Real time Probabilistic Coupled Forecast**
Application of Ocean ReAnalysis: climate monitoring

ORAS4 suggests that there is more heat absorbed by the deeper ocean after 2004.

---

**ocean heat content changes**

Balmaseda et al., 2013

Sea-ice extent anomalies

Zuo & Lien, 2018
Application of ocean analysis: RT monitoring

Real-Time monitoring of ENSO state Ref: 1981-2010

Contribution to the ORIP-RT project
• Update on the 1st day each month
• Compare the latest mean ocean state with 8 other RT Ocean analysis products
Summary

• **Data assimilation in the ocean serves a variety of purposes**, from climate monitoring to initialization of coupled model forecasts and ocean mesoscale prediction.

• This lecture dealt mainly with ocean DA for **initialization of coupled forecasts and reanalyses**, with a global ocean model in climate resolution and use **NEMOVAR** as an example.

• ECMWF NEMOVAR uses a incremental **3DVar-FGAT** configuration and linearized cost function. The BGE covariance is modelled use **balance** operator and **diffusion** operator.

• Compared to the atmosphere, **ocean observations are sparse**. The main source of information are temperature and salinity profiles, sea level from altimeter, SST/SIC/SIT from satellite and in-situ.

• **Assimilation of ocean observations reduces the large uncertainty due to both model and forcing errors**. It improves the initialization of coupled forecasts in NWP, and provides calibration and initialization for reforecast for seasonal forecasts and decadal forecasts.

• **Data assimilation changes the ocean mean state**. consistent ocean reanalysis requires an explicit treatment of the bias.
Further Readings

Ocean Data assimilation


Ocean DA and Reanalysis

