Weak Constraint 4D-Var in ROMS using a Saddle-Point Algorithm

Andy Moore, UC Santa Cruz
Hernan Arango, Rutgers University
John Wilkin, Rutgers University
Chris Edwards, UC Santa Cruz

Supported by the National Oceanographic Partnership Program (NOPP), NOAA and ONR.

Outline

- **Motivation – accelerating the performance of 4D-Var**
- **Saddle-point 4D-Var in ROMS – parallel in time**
- **Applications to the California Current System**
- **Mixed-resolution and mixed-precision**
- **Summary**

Incremental Weak Constraint 4D-Var

Fisher and Gürol (2017)

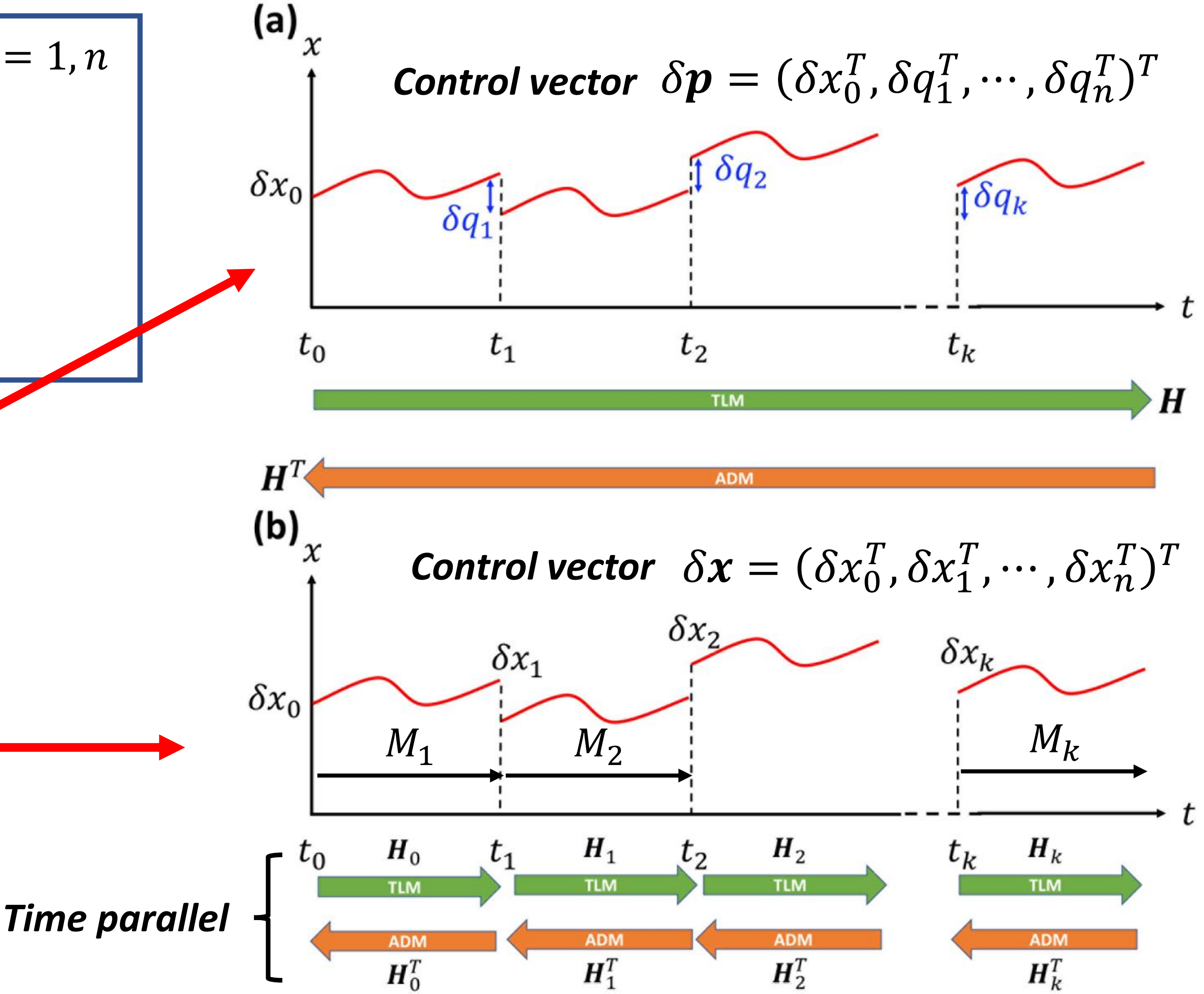
x =ocean state-vector q =correction for model error → a "forcing" term

Nonlinear model	$x_k = \underbrace{\mathcal{M}_k}_{NLM}(x_{k-1}) + q_k \quad k = 1, n$
State vector increment	$\delta x_k = x_k - (x_b)_k$
Model error increment	$\delta q_k = q_k - (q_b)_k$
Tangent linear model	$\delta x_k = \underbrace{M_k}_{TLM} \delta x_{k-1} + \delta q_k$

A forced problem

A series of initial value problems

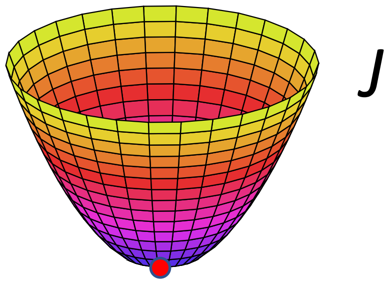
$$\delta q_k = \delta x_k - M_k \delta x_{k-1}$$



Incremental Weak Constraint 4D-Var

Fisher and Gürol (2017)

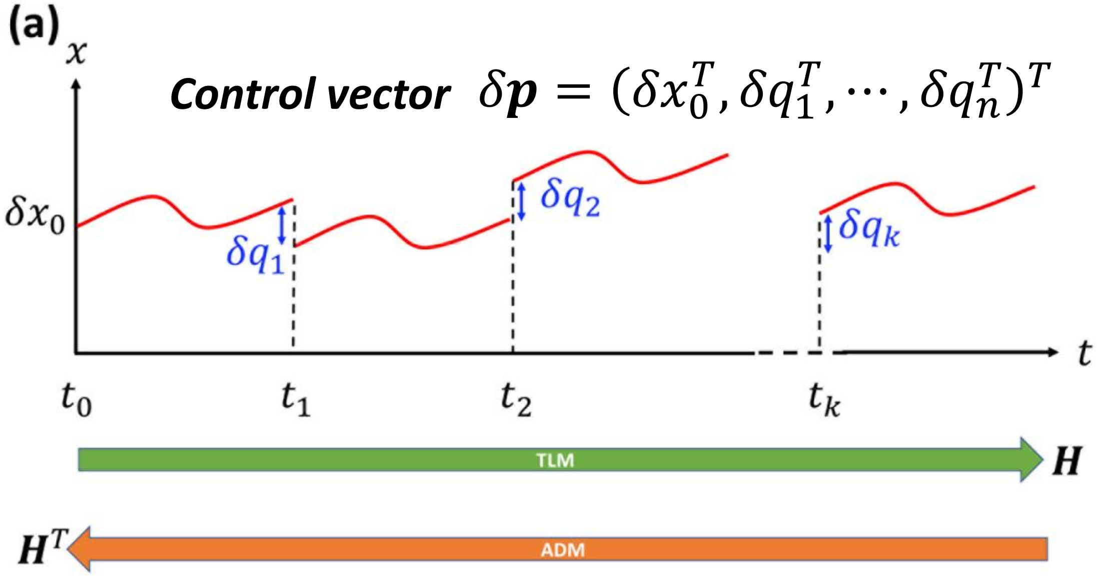
Cost function J



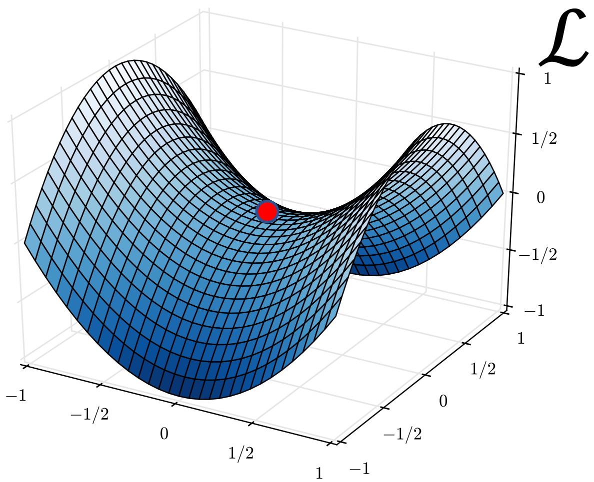
Forcing Formulation

↓

Minimize the usual cost function J



Lagrange function $\mathcal{L}$

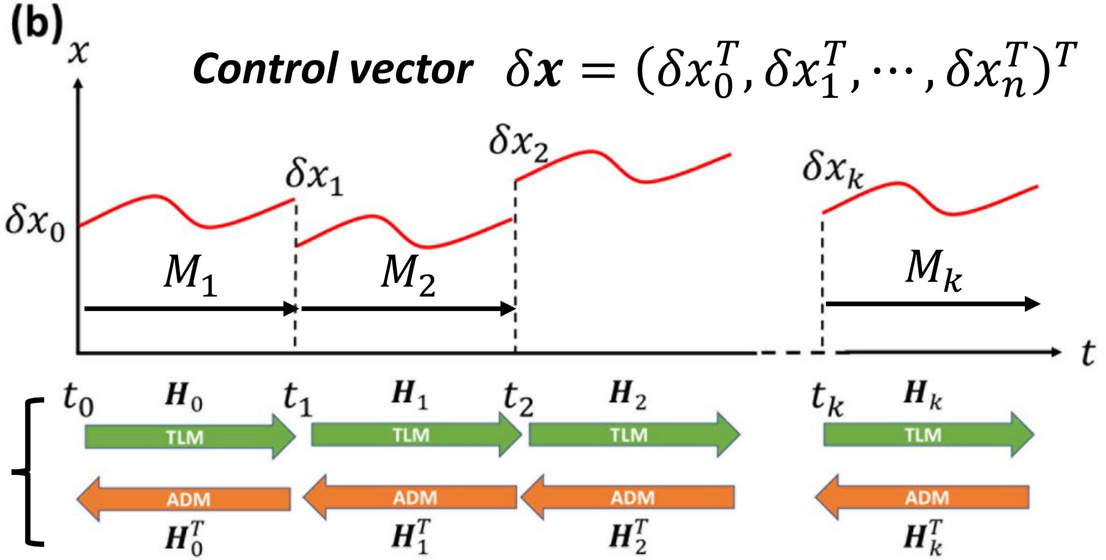


Saddle-point Formulation

↓

Minimize J subject to additional constraints → Lagrange function

Time parallel



Saddle-Point 4D-Var in the California Current System

Two ROMS configurations:

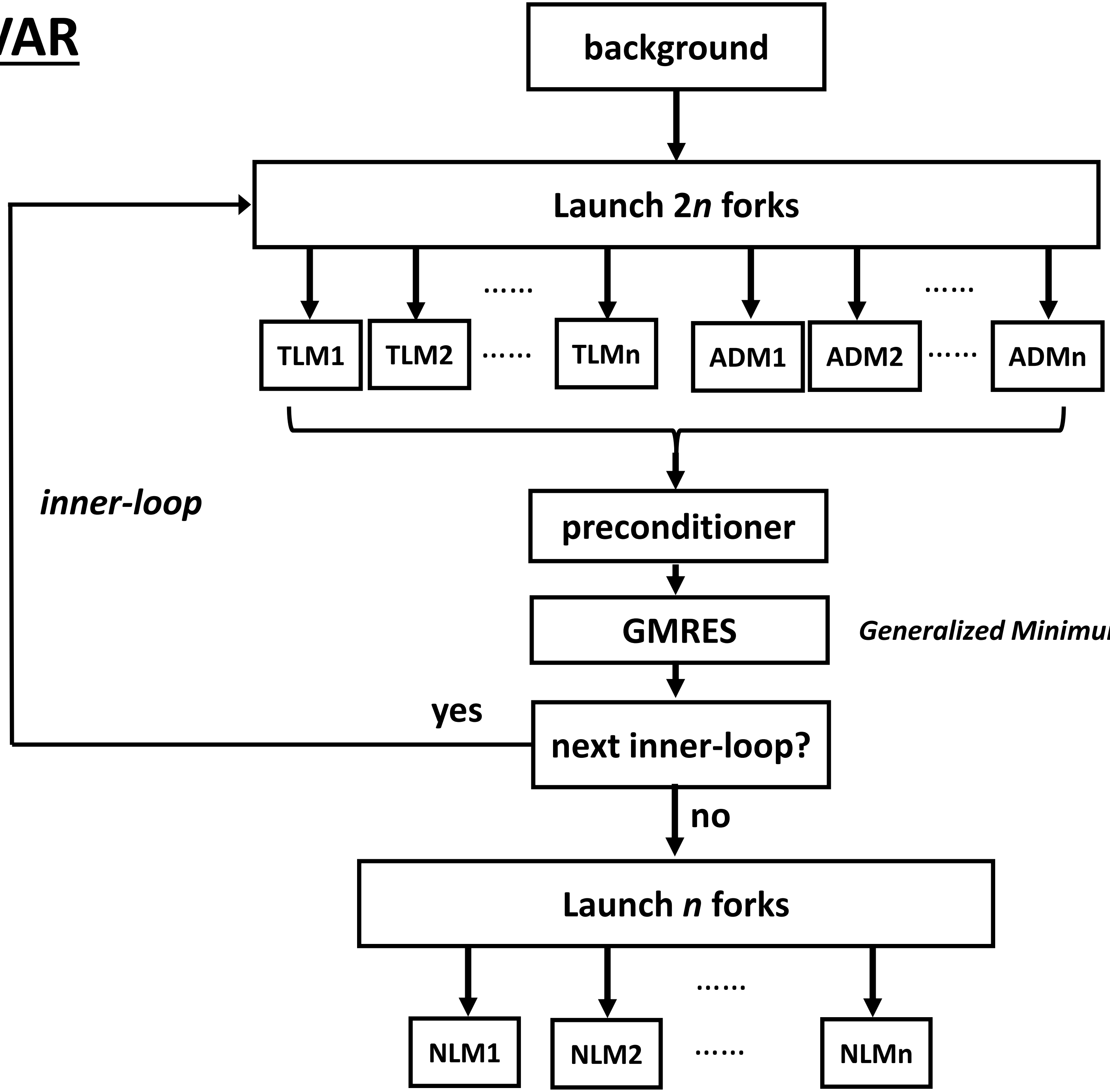
- 1/3rd degree resolution, 42 σ-levels
- COAMPS surface forcing
- ECCO open boundary conditions
- Observations:
  - satellite SST
  - Aviso altimetry
  - Argo profiling floats
- 4-day 4D-Var windows
- Standard test case (WC13)

- 1/10th degree resolution, 42 σ-levels
- ERA surface forcing
- Global HYCOM open boundary conditions
- Observations:
  - satellite SST
  - Aviso altimetry
  - Argo profiling floats
- 8-day 4D-Var windows

Forcing formulation: RBCG (Restricted B-preconditioned CG)

Saddle-point formulation: SP4DVAR

ROMS SP4DVAR



n = number of sub-intervals within the assimilation window.

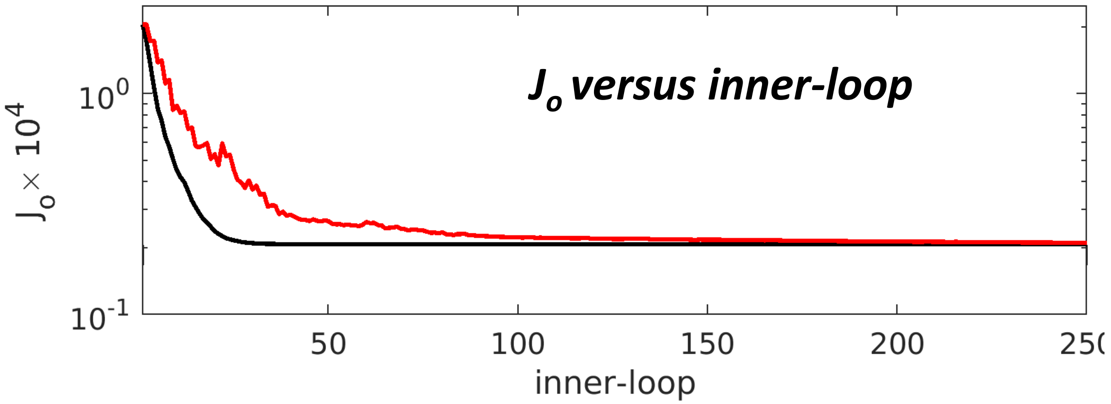
Generalized Minimum Residual algorithm

30km resolution, 4-day assimilation window, 3-7 Jan 2004

1 outer-loop, 4-day cycle, *n=8, Q=0.2B*

8 sub-intervals

n=8

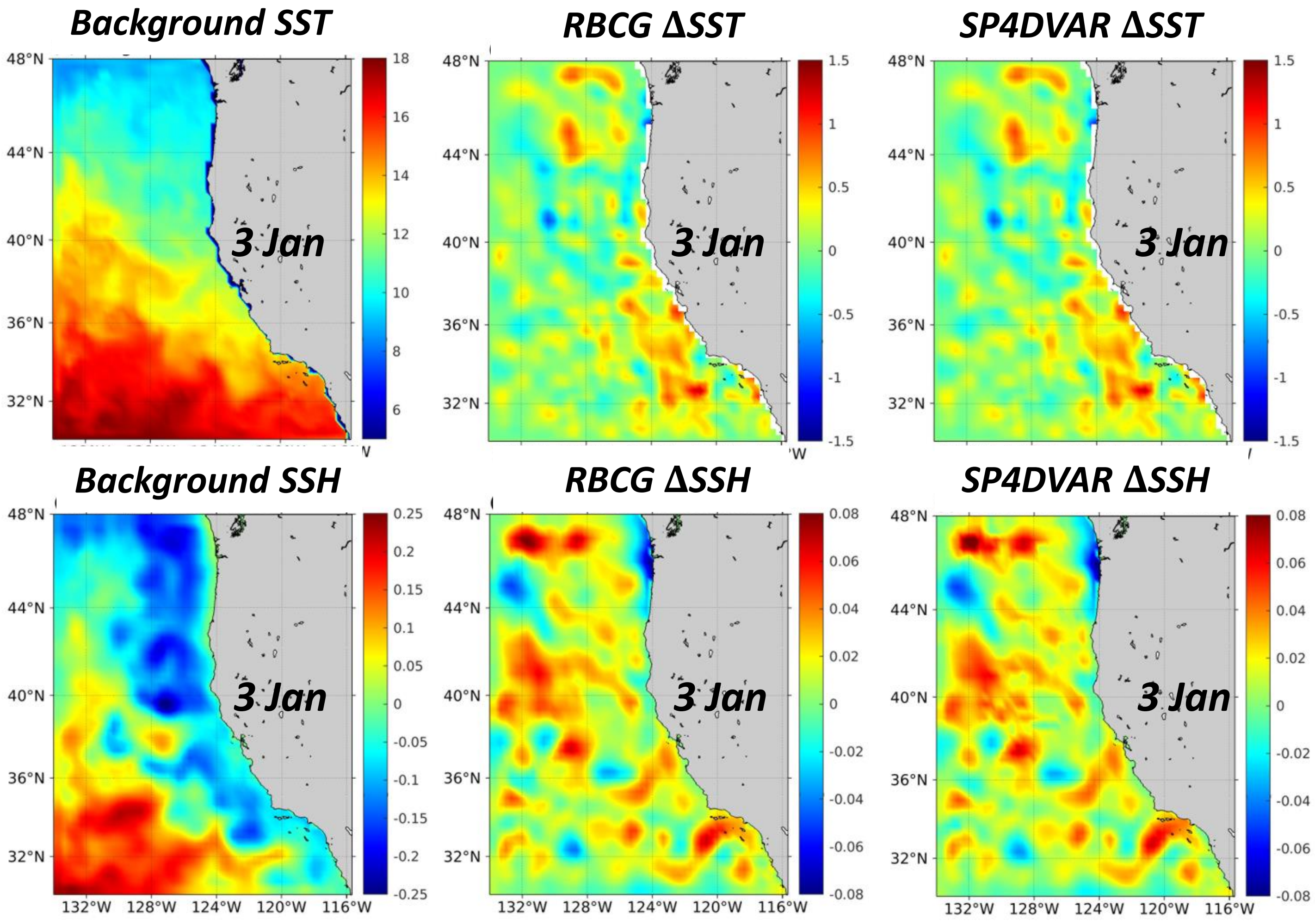


$$J_o = \left(y - H(z)\right)^T R^{-1} \left(y - H(z)\right)$$

RBCG (forcing formulation of 4D-Var)

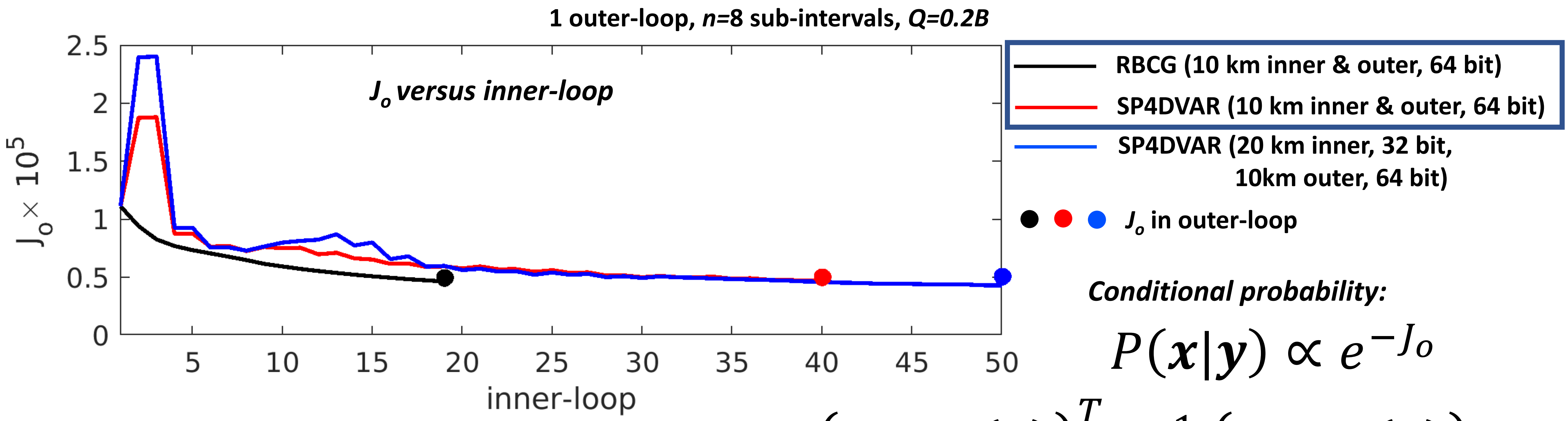
SP4DVAR (saddle-point formulation)

SST & SSH increments: RBCG vs SP4DVAR



Performance

10km resolution, 8-day assimilation window, 3-11 Jan 2004



Conditional probability:

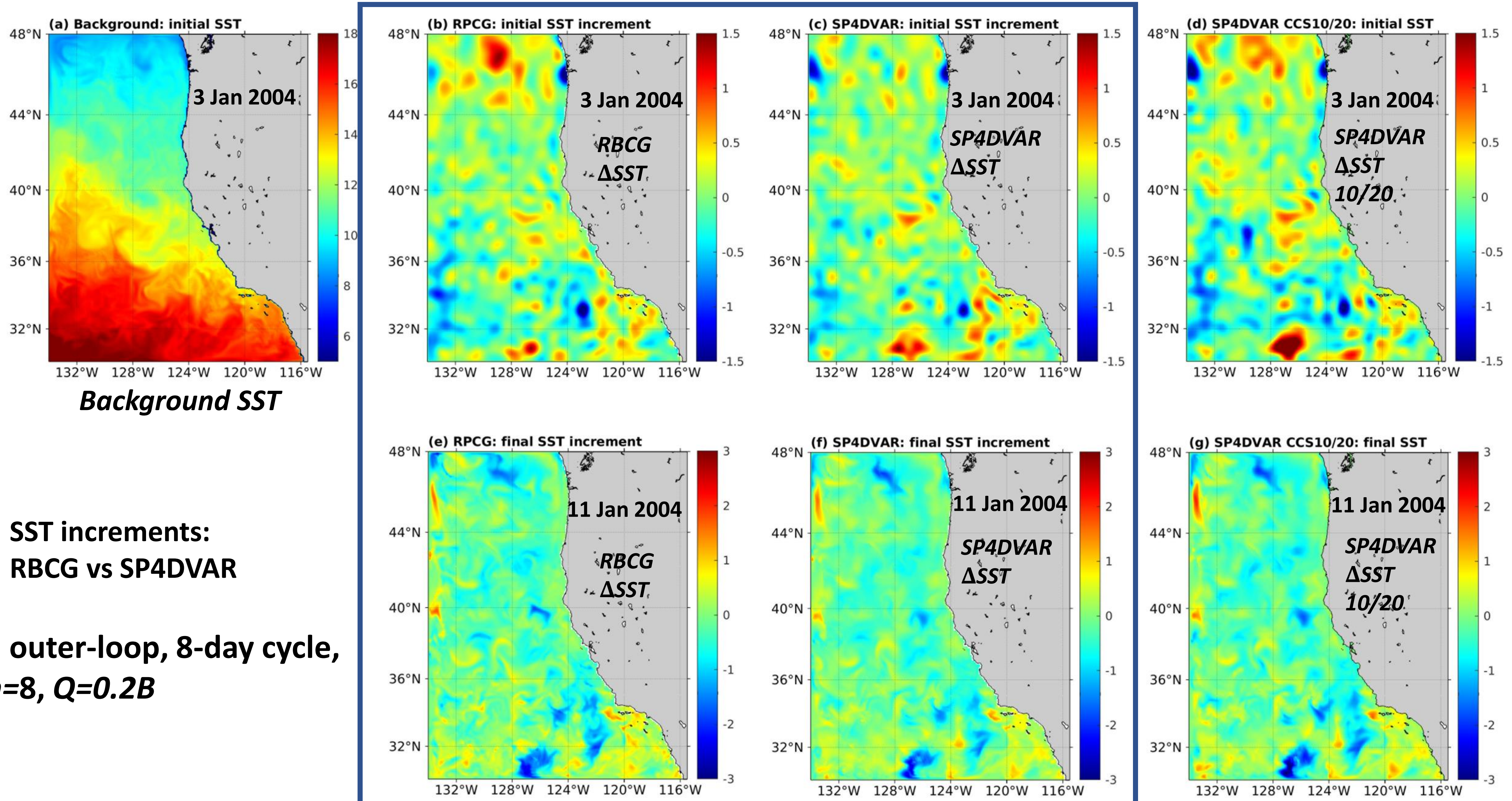
$$P(\boldsymbol{x}|\boldsymbol{y}) \propto e^{-J_o}$$

$$J_o = \left(\boldsymbol{y} - H(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1} \left(\boldsymbol{y} - H(\boldsymbol{x})\right)$$

J_o	RBCG	SP4DVAR	SP4DVAR-20km
Temperature	4.34×10^4	4.45×10^4	4.53×10^4
SSH	6.29×10^3	5.03×10^3	5.10×10^3
Salinity	6.75×10^2	7.00×10^2	6.53×10^2
TOTAL:	5.03×10^4	5.04×10^4	5.10×10^4

Obs: SST (MODIS, AVHRR, GOES, AMSR), SSH (Aviso), in situ T & S (XBT, CTD, Argo)

SST 4D-Var Increments: 10km resolution, 3-11 Jan 2004



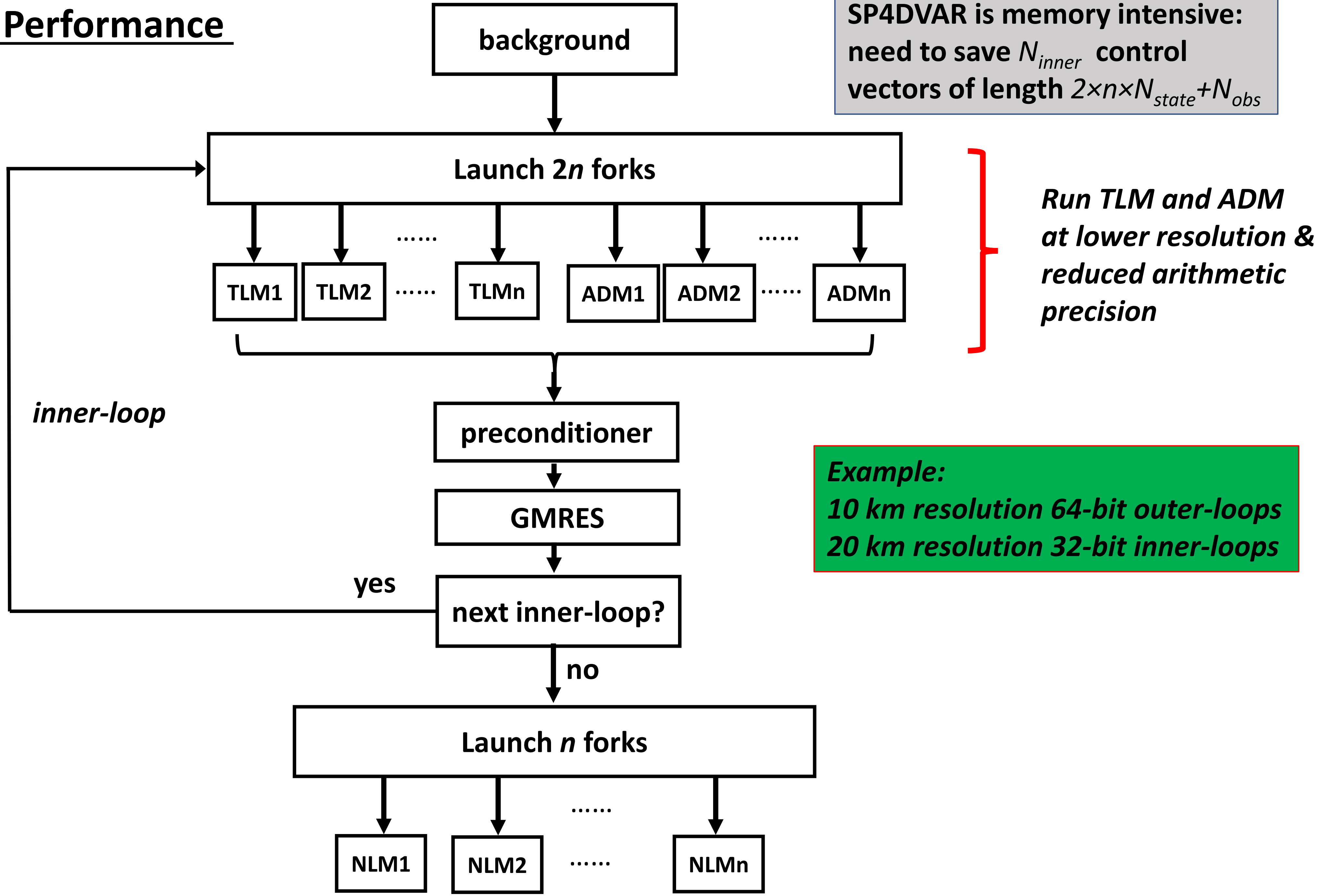
$J_o = 5.03 \times 10^4$ $J_o = 5.04 \times 10^4$ $J_o = 5.10 \times 10^4$

SST increments:
RBCG vs SP4DVAR

1 outer-loop, 8-day cycle,
***n*=8, *Q*=0.2B**

Accelerating the Performance of SP4DVAR

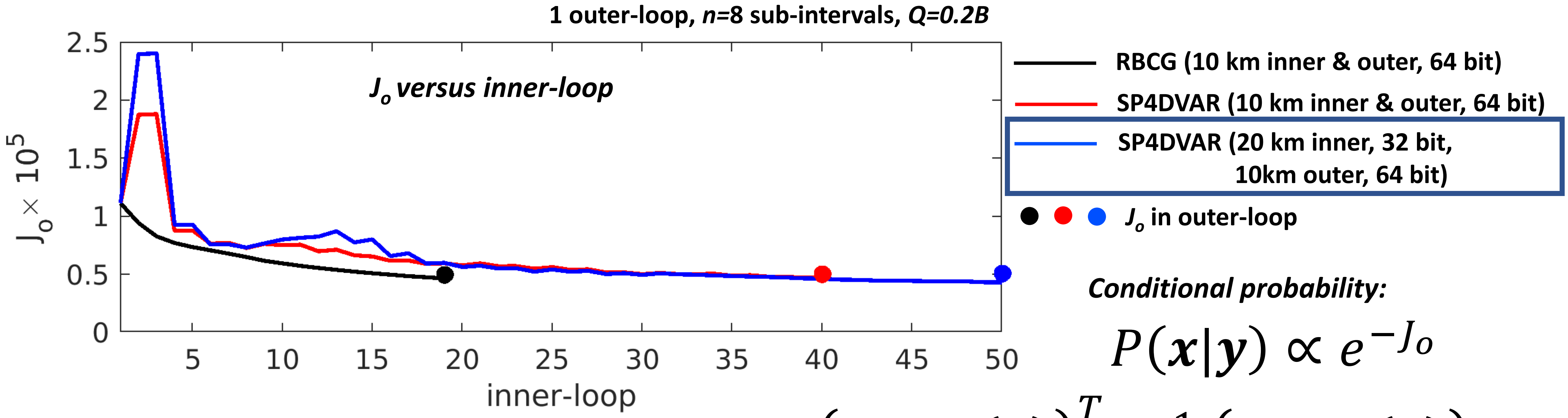
SP4DVAR is memory intensive: need to save N_{inner} control vectors of length $2 \times n \times N_{state} + N_{obs}$



Example:
10 km resolution 64-bit outer-loops
20 km resolution 32-bit inner-loops

Performance

10km resolution, 8-day assimilation window, 3-11 Jan 2004



Conditional probability:

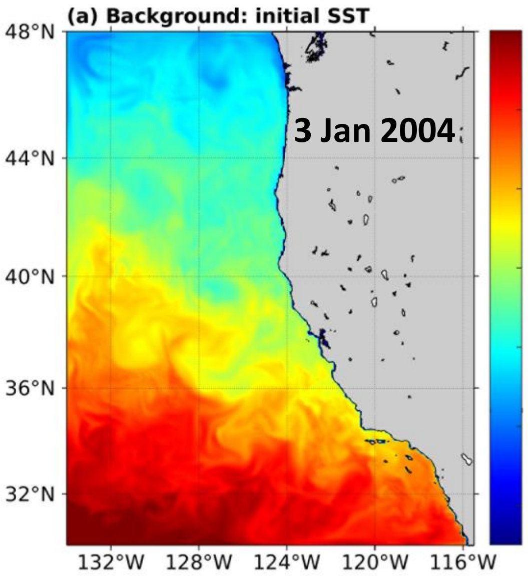
$$P(\boldsymbol{x}|\boldsymbol{y}) \propto e^{-J_o}$$

$$J_o = \left(\boldsymbol{y} - H(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1} \left(\boldsymbol{y} - H(\boldsymbol{x})\right)$$

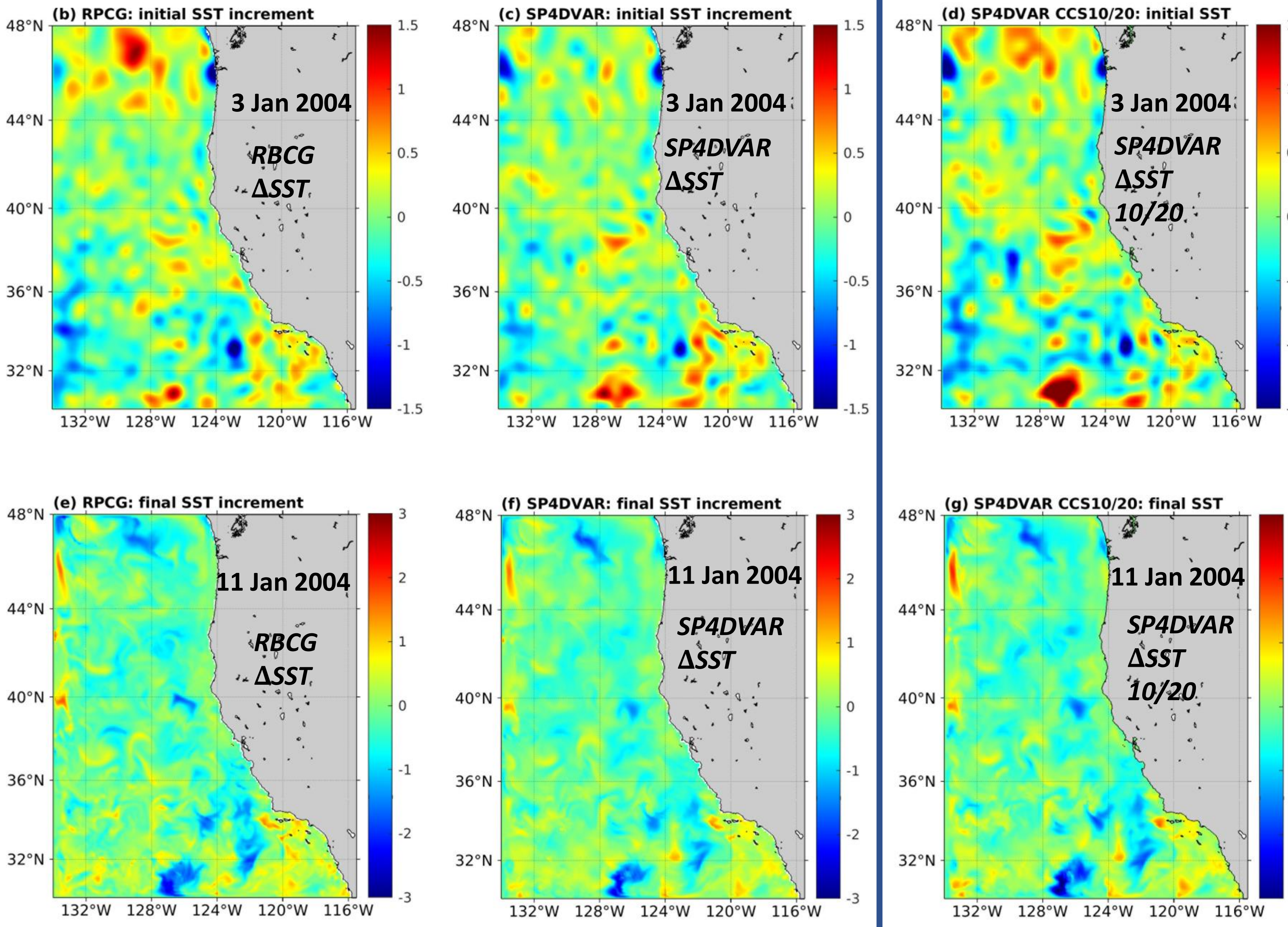
J_o	RBCG	SP4DVAR	SP4DVAR-20km
Temperature	4.34×10^4	4.45×10^4	4.53×10^4
SSH	6.29×10^3	5.03×10^3	5.10×10^3
Salinity	6.75×10^2	7.00×10^2	6.53×10^2
TOTAL:	5.03×10^4	5.04×10^4	5.10×10^4

Obs: SST (MODIS, AVHRR, GOES, AMSR), SSH (Aviso), in situ T & S (XBT, CTD, Argo)

SST 4D-Var Increments: 10km resolution, 3-11 Jan 2004



Background SST



$J_o = 5.03 \times 10^4$ $J_o = 5.04 \times 10^4$ $J_o = 5.10 \times 10^4$

SST increments:
RBCG vs SP4DVAR

1 outer-loop, 8-day cycle,
n*=8, *Q*=0.2*B

Computation Time: SP4DVAR vs RBCG

1 outer-loop, 20 inner-loops, 8-day cycle
***n*=8 sub-intervals**
Q=0.2B

Experiment	Relative time per inner-loop
RBCG, 10km outer & inner, r64	100%
SP4DVAR, 10km outer & inner, r64	12%
SP4DVAR, 10km outer, 20km inner, r64	2.4%
SP4DVAR, 10km outer, 20km inner, r32	1.6%

Scales as ~n^{-1}

- **SP4DVAR performance is still sub-optimal:**
  - **- ADM 40% slower than TLM**
  - **- solution assembly & GMRES overhead**
- **Remedy:**
  - **- optimize ADM**
  - **- further parallelize GMRES**

- **Single precision yields modest gain in CPU but a factor or 2 reduction in memory => can use larger *n***

Summary and Conclusions

- Saddle-point 4D-Var has the potential to be a game-changer!
- Saddle-point 4D-Var will run ***much*** faster than RBCG on very large HPC systems
- Outstanding performance issues in ROMS-SP4DVAR:
  - improve efficiency of adjoint model
  - solution assembly & GMRES overhead
  - preconditioning
- Ongoing work:
  - multiple outer-loops
  - specification of model error covariances, ***Q*** (ML?)

- Topic 1: How best to transition saddle-point to an operational environment (*e.g.* NOAA NOS)?
- Topic 2: Promote the development of efficient, hybrid DA methods that exploit parallelization in time.