Scalable parallel solution and uncertainty quantification techniques for variational data assimilation

Vishwas Rao and Adrian Sandu

Adjoint Workshop 2015, June 5, 2015

Computational Science Laboratory (CSL)
Department of Computer Science
Virginia Tech
visrao@cs.vt.edu

Outline

1. Data assimilation in one slide
2. Parallel 4D-Var
3. A-posteriori error estimates for 4D-Var
4. Conclusions

Data assimilation in one figure

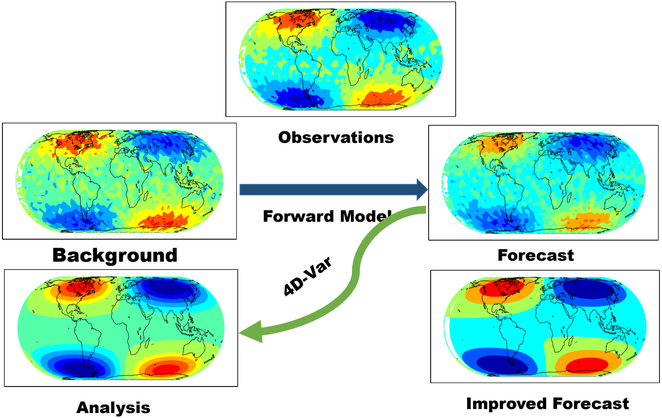

Figure: Data assimilation

4D-Var formulation

- The model:

$$\mathcal{A} := \mathbf{x}_{k+1} - \mathcal{M}_{k,k+1}(\mathbf{x}_k, \theta) = 0, \quad k = 0, \dots, N-1\,, \quad \mathbf{x}_0 = \mathbf{x}_0(\theta)\,. \tag{1}$$

- 4D-Var cost function:

$$\begin{aligned} \mathcal{J}(\mathbf{x}_0) &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \\ &\quad + \frac{1}{2}\sum_{k=0}^{N} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)\,, \end{aligned} \tag{2}$$

- Inverse problem:

$$\begin{aligned} \mathbf{x}_0^{\mathrm{a}} = &\arg\min_{\mathbf{x}_0 \in \mathbb{R}^n} \quad \mathcal{J}(\mathbf{x}_0) \\ &\text{subject to} \quad \mathcal{A} \end{aligned} \tag{3}$$

Serial 4D-Var

The Lagrangian function associated with the inverse problem is

$$\begin{aligned}\mathcal{L} \;=\;& \frac{1}{2}\left(\mathbf{x}_0-\mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}_0^{-1}\left(\mathbf{x}_0-\mathbf{x}_0^{\mathrm{b}}\right) \\ &+\frac{1}{2}\sum_{k=0}^{N}\left(\mathcal{H}_k(\mathbf{x}_k)-\mathbf{y}_k\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathcal{H}_k(\mathbf{x}_k)-\mathbf{y}_k\right) \\ &-\sum_{k=0}^{N-1}\underbrace{\lambda_{k+1}^{\mathrm{T}}}_{\text{L M}}\cdot\overbrace{\left(\mathbf{x}_{k+1}-\mathcal{M}_{k,k+1}(\mathbf{x}_k,\theta)\right)}^{\text{Model Constraints}}-\lambda_0^{\mathrm{T}}\cdot\left(\mathbf{x}_0-\mathbf{x}_0(\theta)\right)\end{aligned} \tag{4}$$

- Requires several forward and adjoint computations which are inherently serial.
- Can we reformulate the problem so that we end up solving small and independent pieces of adjoint and forward models

Parallel 4D-Var–Augmented Lagrangian I

- Divide the assimilation window to multiple sub-intervals.
- Propagate the background state to get an initial guess at the beginning of each of these subintervals.
- The augmented Lagrangian associated with 4D-Var cost function and model constraints is given by

$$\begin{aligned}
\mathcal{L}(\mathbf{X}) \quad = \quad & \frac{1}{2}\left(\widetilde{\mathbf{x}}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^T \mathbf{B}_0^{-1}\left(\widetilde{\mathbf{x}}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \\
& + \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}}\left(\mathcal{H}\left(\widetilde{\mathbf{x}}_{\mathrm{k}}\right) - \mathbf{y}_{\mathrm{k}}\right)^T \mathbf{R}_{\mathrm{k}}^{-1}\left(\mathcal{H}\left(\widetilde{\mathbf{x}}_{\mathrm{k}}\right) - \mathbf{y}_{\mathrm{k}}\right) \\
& - \sum_{\mathrm{k}=0}^{\mathrm{N}-1}\lambda_{\mathrm{k}+1}^T\left(\widetilde{\mathbf{x}}_{\mathrm{k}+1} - \mathcal{M}_{\mathrm{k},\mathrm{k}+1}\left(\widetilde{\mathbf{x}}_{\mathrm{k}}\right)\right) \\
& \underbrace{+ \frac{\mu}{2}\sum_{\mathrm{k}=0}^{\mathrm{N}-1}\left(\widetilde{\mathbf{x}}_{\mathrm{k}+1} - \mathcal{M}_{\mathrm{k},\mathrm{k}+1}\left(\widetilde{\mathbf{x}}_{\mathrm{k}}\right)\right)^T \mathbf{P}_{\mathrm{k}}^{-1}\left(\widetilde{\mathbf{x}}_{\mathrm{k}+1} - \mathcal{M}_{\mathrm{k},\mathrm{k}+1}\left(\widetilde{\mathbf{x}}_{\mathrm{k}}\right)\right)}_{\text{Penalty term}},
\end{aligned}$$

$$\mathbf{X} \quad := \quad [\widetilde{\mathbf{x}}_0, \cdots, \widetilde{\mathbf{x}}_N]$$

Parallel 4D-Var–Augmented Lagrangian II

- Gradient computation:

$$\nabla_{\widetilde{\mathbf{x}}_0}\mathcal{L} = \mathbf{B}_0^{-1}\left(\widetilde{\mathbf{x}}_0 - \mathbf{x}_0^b\right) - \mathbf{M}^{\mathrm{T}}\mathbf{P}_\mathbf{1}^{-\mathbf{1}}\left(\widetilde{\mathbf{x}}_1 - \mathcal{M}\left(\widetilde{\mathbf{x}}_0\right)\right), \quad (5)$$

$$\begin{aligned}\nabla_{\widetilde{\mathbf{x}}_k}\mathcal{L} = {} & \mathbf{H}_\mathbf{k}\mathbf{R}_\mathbf{k}^{-\mathbf{1}}\left(\mathcal{H}\left(\widetilde{\mathbf{x}}_k\right) - \mathbf{y}_k\right) + \mathbf{M}^{\mathrm{T}}\lambda_{k+1} \\ & - \mu\mathbf{M}^{\mathrm{T}}\mathbf{P}_\mathbf{k}^{-\mathbf{1}}\left(\widetilde{\mathbf{x}}_{k+1} - \mathcal{M}\left(\widetilde{\mathbf{x}}_k\right)\right) \\ & + \mu\mathbf{P}_\mathbf{k}^{-\mathbf{1}}\left(\widetilde{\mathbf{x}}_k - \mathcal{M}\left(\widetilde{\mathbf{x}}_{k-1}\right)\right) - \lambda_k, \quad k = 1, \ldots, N-1,\end{aligned} \quad (6)$$

$$\nabla_{\widetilde{\mathbf{x}}_N}\mathcal{L} = \mathbf{H}_\mathbf{N}\mathbf{R}_\mathbf{N}^{-\mathbf{1}}\left(\mathcal{H}\left(\widetilde{\mathbf{x}}_N\right) - \mathbf{y}_N\right) + \mu\mathbf{P}_\mathbf{N}^{-\mathbf{1}}\left(\widetilde{\mathbf{x}}_N - \mathcal{M}\left(\widetilde{\mathbf{x}}_{N-1}\right)\right). \quad (7)$$

- The gradients can be evaluated in parallel!!!

Numerical Results - Lorenz 96 Model I

- Lorenz-96 model is given by:

$$\frac{dx_i}{dt} = x_{i-1}\left(x_{i+1} - x_{i-2}\right) - x_i + F \ , \tag{8}$$

  where $\mathbf{x} = (x_1, x_2, \ldots, x_{40})^T \in \mathbb{R}^{40}$ is the state vector, and $F = 8$ is the forcing term.
- Synthetic observations with background errors $\sim$ 8% and observation errors $\sim$ 5% are generated.
- Observations: Equally spaced in the temporal direction.
- Weak scaling – each processor approximately does the same amount of work.

Numerical Results - Lorenz 96 Model II

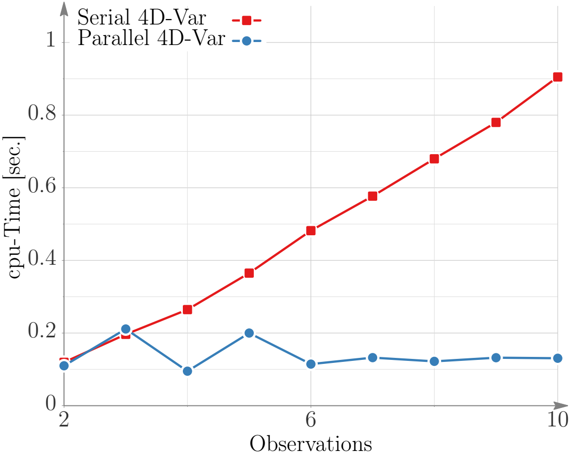

Figure: Scalability of cost function evaluations.

Numerical Results - Lorenz 96 Model III

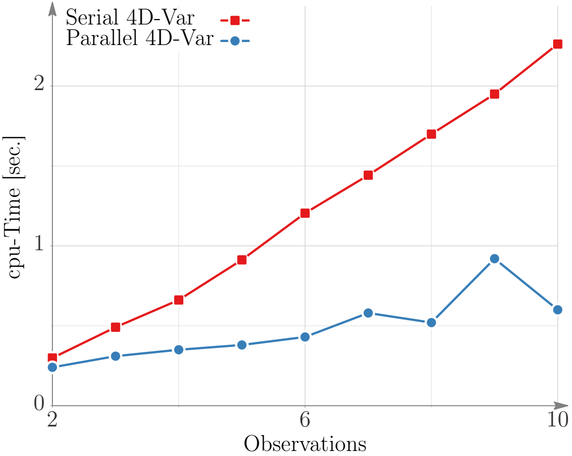

Figure: Scalability of gradient evaluations.

Numerical Results - Lorenz 96 Model IV

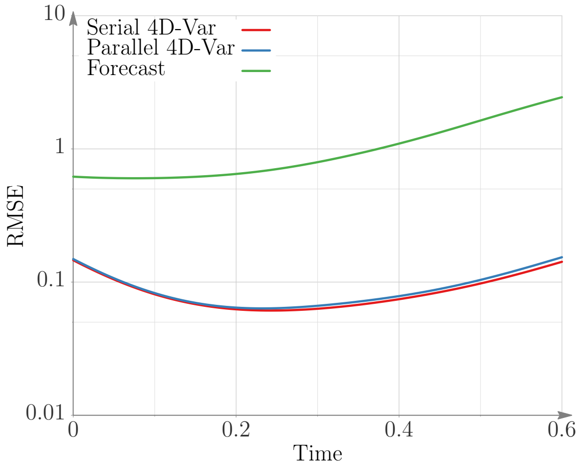

Figure: RMSE Comparisons between serial and 4D-var for Lorenz model.

Numerical Results - Lorenz 96 Model V

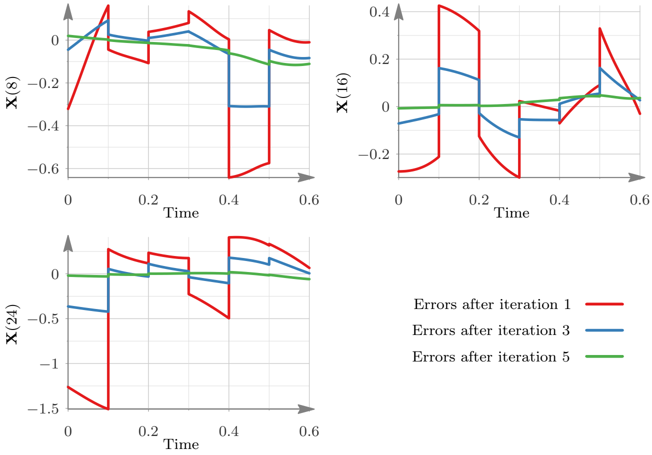

Figure: Errors at different stages: Lorenz model.

Numerical Results - Lorenz 96 Model VI

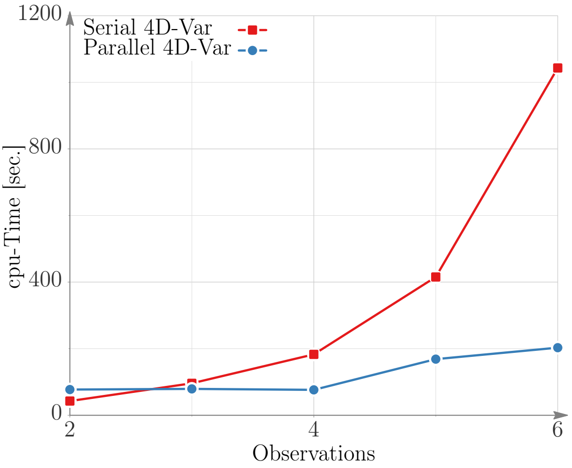

Figure: Timing Comparisons between serial and 4D-Var for Lorenz model.

A-posteriori error estimates for 4D-Var

- The perfect model:

$$\mathcal{A} := \mathbf{x}_{k+1} - \mathcal{M}_{k,k+1}(\mathbf{x}_k, \theta) = 0, \quad k = 0, \ldots, N-1\,, \quad \mathbf{x}_0 = \mathbf{x}_0(\theta)\,. \qquad (9)$$

- Ideal 4D-Var cost function:

$$\begin{aligned} \mathcal{J}(\mathbf{x}_0) \;=\; & \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \qquad (10)\\ & + \frac{1}{2}\sum_{k=0}^{N} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k^{\mathrm{true}}\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k^{\mathrm{true}}\right)\,, \end{aligned}$$

- Inverse problem:

$$\begin{aligned} \mathbf{x}_0^{\mathrm{a}} = & \arg\min_{\mathbf{x}_0 \in \mathbb{R}^n} \quad \mathcal{J}(\mathbf{x}_0) \\ & \text{subject to} \quad \mathcal{A} \end{aligned} \qquad (11)$$

‘Imperfect’ 4D-Var

- Imperfect cost function:

$$\widehat{\mathcal{J}}(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \tag{12}$$

$$+\frac{1}{2}\sum_{k=0}^{N}\left(\mathcal{H}_k(\widehat{\mathbf{x}}_k) - \mathbf{y}_k^{\mathrm{true}} - \Delta\mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1}\left(\mathcal{H}_k(\widehat{\mathbf{x}}_k) - \mathbf{y}_k^{\mathrm{true}} - \Delta\mathbf{y}_k\right).$$

- The perturbed strongly constrained 4D-Var analysis problem solved in reality is

$$\widehat{\mathbf{x}}_0^{\mathrm{a}} = \underset{\mathbf{x}_0 \in \mathbb{R}^n}{\arg\min}\ \widehat{\mathcal{J}}(\mathbf{x}_0) \quad \text{subject to } \widehat{\mathbf{x}}_{k+1} - \mathcal{M}_{k,k+1}(\widehat{\mathbf{x}}_k, \theta) - \Delta\widehat{\mathbf{x}}_{k+1}(\widehat{\mathbf{x}}_k, \theta). \tag{13}$$

Ideal super-Lagrangian

- QoI is $\mathcal{E}\left(\mathbf{x}_0^{\mathrm{a}}\right)$
- We are interested in estimating $\mathcal{E}\left(\widehat{\mathbf{x}}_0^{\mathrm{a}}\right) - \mathcal{E}\left(\mathbf{x}_0^{\mathrm{a}}\right)$.
- Super-Lagrangian with Ideal KKT:

$$
\begin{aligned}
\mathcal{L}^{\mathcal{E}} \;=\;& \mathcal{E}(\mathbf{x}_0) - \sum_{k=0}^{N-1} \nu_{k+1}^T \cdot \overbrace{\left(\mathbf{x}_{k+1} - \mathcal{M}_{k,k+1}(\mathbf{x}_k)\right)}^{\text{'Ideal' Forward Model}} \\
& - \mu_N^T \cdot \left(\lambda_N - \mathbf{H}_N^{\mathrm{T}} \mathbf{R}_N^{-1} \left(\mathcal{H}_N(\mathbf{x}_N) - \mathbf{y}_N^{\mathrm{true}}\right)\right) \\
& - \sum_{k=0}^{N-1} \mu_k^T \cdot \overbrace{\left(\lambda_k - \mathbf{M}_{k,k+1}^{\mathrm{T}}\, \lambda_{k+1} - \mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k^{\mathrm{true}}\right)\right)}^{\text{'Ideal' Adjoint model}} \\
& \underbrace{- \zeta^T \mathbf{B}_0^{-1} (\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}) - \zeta^T \lambda_0}_{\text{'Ideal' Optimality}} \,.
\end{aligned}
\tag{14}
$$

Perturbed super-Lagrangian

$$
\begin{aligned}
\widehat{\mathcal{L}}^{\mathcal{E}} \;=\; & \mathcal{E}(\widehat{\mathbf{x}}_0) - \sum_{k=0}^{N-1} \nu_{k+1}^{\mathrm{T}} \cdot \overbrace{\left(\widehat{\mathbf{x}}_{k+1} - \mathcal{M}_{k,k+1}(\widehat{\mathbf{x}}_k) - \Delta\widehat{\mathbf{x}}_{k+1}\right)}^{\text{Perturbed Forward Model}} \\
& - \mu_N^{\mathrm{T}} \cdot \left(\widehat{\lambda}_N - \mathbf{H}_N^{\mathrm{T}} \mathbf{R}_N^{-1} \left(\mathcal{H}_N(\widehat{\mathbf{x}}_N) - \mathbf{y}_N^{\mathrm{true}}\right) + \mathbf{H}_N^{\mathrm{T}} \mathbf{R}_N^{-1} \Delta \mathbf{y}_N\right) \\
& - \sum_{k=0}^{N-1} \mu_k^{\mathrm{T}} \cdot \overbrace{\left(\widehat{\lambda}_k - \left(\mathbf{M}_{k,k+1}^{\mathrm{T}} + \left(\Delta\widehat{\mathbf{x}}_{k+1}\right)_{\widehat{\mathbf{x}}_k}^{\mathrm{T}}\right) \widehat{\lambda}_{k+1}\right)}^{\text{Perturbed Adjoint Model}} \\
& - \sum_{k=0}^{N-1} \mu_k^{\mathrm{T}} \cdot \overbrace{\left(\mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \Delta \mathbf{y}_k - \mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathcal{H}_k(\widehat{\mathbf{x}}_k) - \mathbf{y}_k^{\mathrm{true}}\right)\right)}^{\text{Perturbed Adjoint model}} \\
& - \zeta^{\mathrm{T}} \cdot \underbrace{\left(\mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}) + \widehat{\lambda}_0 + \sum_{k=0}^{N-1} \left(\Delta\widehat{\mathbf{x}}_{k+1}\right)_{\theta}^{\mathrm{T}} \widehat{\lambda}_{k+1}\right)}_{\text{Perturbed Optimality}} .
\end{aligned}
\tag{15}
$$

Error estimate

- At $(\mathbf{x}_0^{\mathrm{a}}, \mathbf{x}, \lambda, \mu, \nu, \zeta)$, super-Lagrangian is stationary. Hence we have:

$$\Delta\mathcal{L}^{\mathcal{E}} = \mathcal{L}^{\mathcal{E}}(\widehat{\mathbf{x}}_0^{\mathrm{a}}, \widehat{\mathbf{x}}, \widehat{\lambda}, \mu, \nu, \zeta) - \mathcal{L}^{\mathcal{E}}(\mathbf{x}_0^{\mathrm{a}}, \mathbf{x}, \lambda, \mu, \nu, \zeta) \approx 0. \tag{16}$$

- The estimate can be obtained by subtracting the ideal super-Lagrangian from the perturbed one:

$$\begin{aligned} 0 \;\approx\;& \Delta\mathcal{E} - \sum_{k=0}^{N-1} \nu_{k+1}^{\mathrm{T}} \cdot \left(-\Delta\widehat{\mathbf{x}}_{k+1}\right) - \mu_N^{\mathrm{T}} \cdot \left(\mathbf{H}_N^{\mathrm{T}}\mathbf{R}_N^{-1}\Delta\mathbf{y}_N\right) \\ & - \sum_{k=0}^{N-1} \mu_k^{\mathrm{T}} \cdot \left(\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\Delta\mathbf{y}_k - \left(\Delta\widehat{\mathbf{x}}_{k+1}\right)_{\widehat{\mathbf{x}}_k}^{\mathrm{T}} \widehat{\lambda}_{k+1}\right) \\ & - \zeta^{\mathrm{T}} \cdot \left(\sum_{k=0}^{N-1} \left(\Delta\widehat{\mathbf{x}}_{k+1}\right)_{\mathbf{x}_0}^{\mathrm{T}} \widehat{\lambda}_{k+1}\right) . \end{aligned}$$

Evaluating the super-Lagrange parameters

- ζ can be obtained by solving the linear system:

$$(\nabla^2_{\mathbf{x}_0,\mathbf{x}_0} j)(\mathbf{x}_0^{\mathrm{a}}) \cdot \zeta = \mathcal{E}_{\mathbf{x}_0}^T \,. \tag{17}$$

- μ_k is obtained by the TLM initialized with ζ.
- ν_k requires a solution of the second order adjoint system.

Experimental Settings

- Experiments using shallow water model on the sphere.
- Hourly observation for 9 (24) hours.
- Synthetic observations with mean = 0 and std. deviation = 2% for Height and 10% for velocity.
- For model errors, observations are collected on a fine grid. But the optimization is performed on a coarse grid.
- Experiments performed with dense and sparse observation grids.

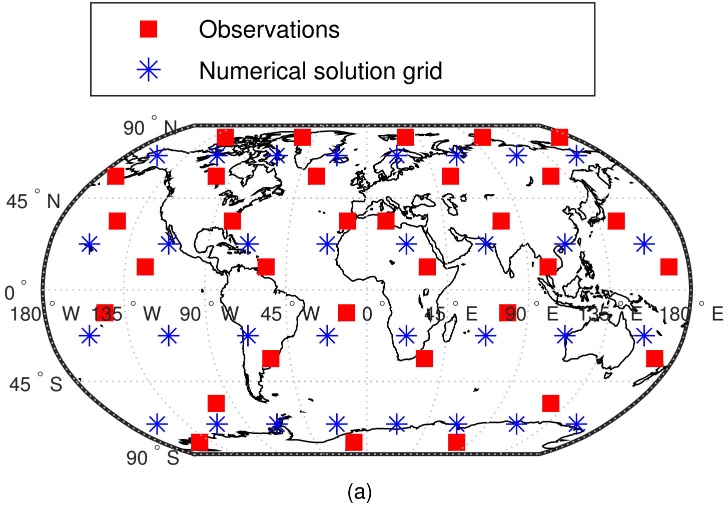

Figure: Selected coarse grid points and sensor locations

Deterministic validation

$$\Delta\mathcal{E}_{\mathrm{actual}} = \mathcal{E}(\widehat{\mathbf{x}}_0^{\mathrm{a}}) - \mathcal{E}(\mathbf{x}_0^{\mathrm{actual}})$$

	$\Delta\mathcal{E}_{\mathrm{actual}}$	$\Delta\mathcal{E}_{\mathrm{est}}$
Data Errors	54.70	57.26
Model Errors (Discrete)	1.9278	2.9683

Table: The comparison between actual error and the a posteriori error estimates for the dense observation network.

$\Delta\mathcal{E}^{\mathrm{actual}}$	$\Delta\mathcal{E}^{\mathrm{est}}$	Contributions (Data Errors)	Contributions (Model Errors)
284.321	581.883	624.772	- 42.889

Table: Comparison between actual errors in the QOI and the a-posteriori error estimates for the shallow water model for the sparse observation network.

Data error contributions

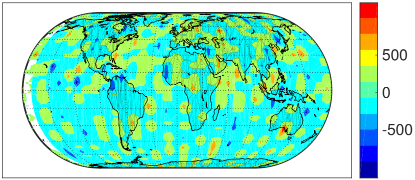

Figure: *Sparse observation network scenario: Data errors and its impact on the Height component*.

Model error contributions

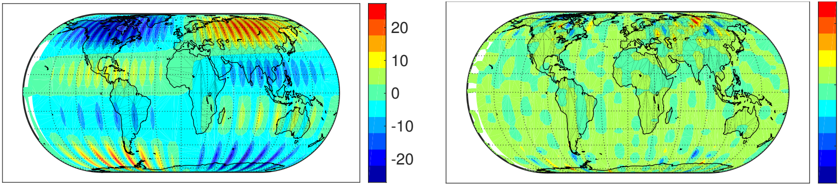

Figure: *Sparse observation network scenario :Model errors and its impact on the Height component.*

Conclusions

- Augmented Lagrangian framework is promising and can give real speedups
- A-posteriori error estimates can prove useful in optimal sensor locations, mesh refinement.
- Has to be tested on WRF.