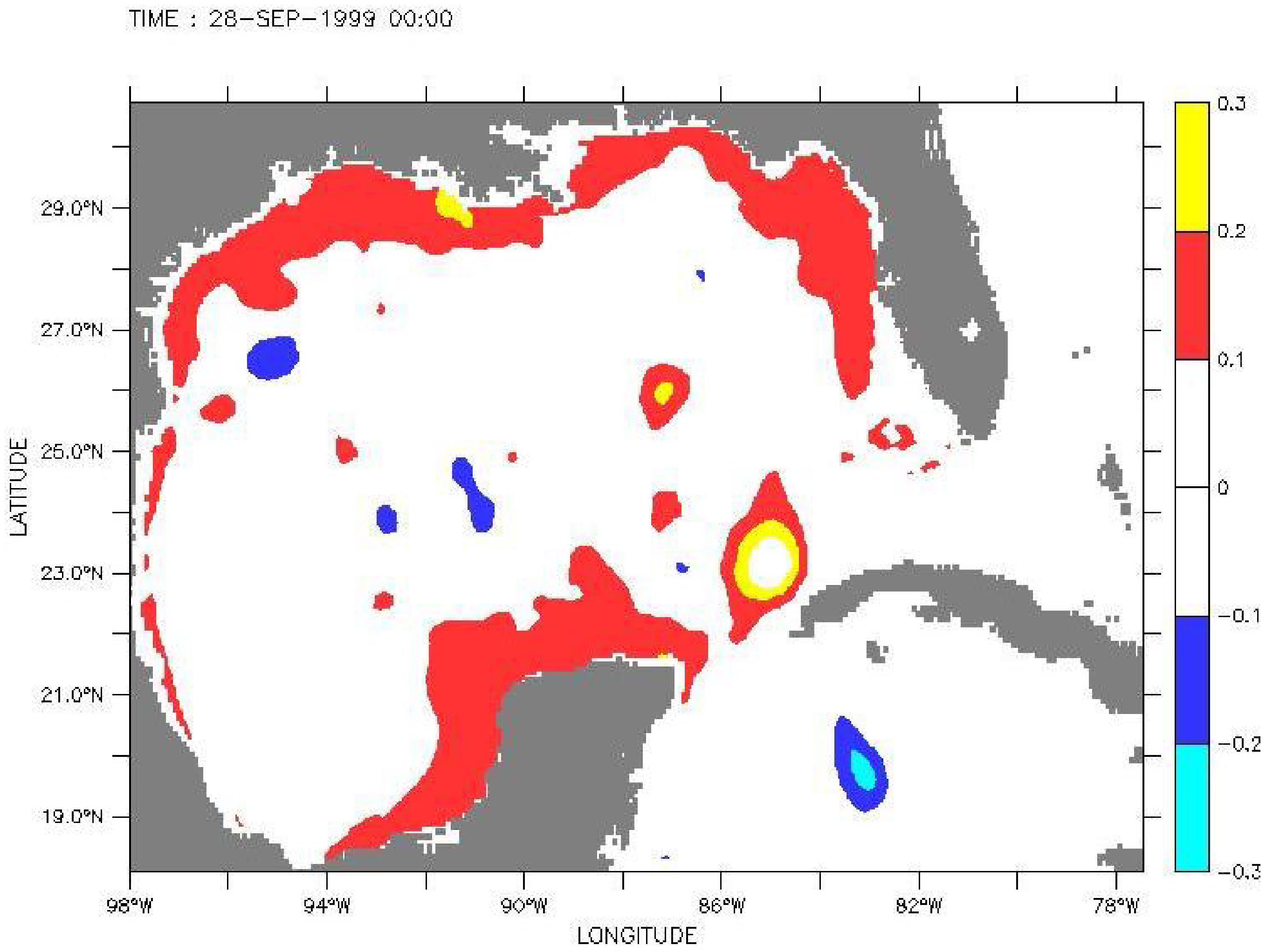
On “objective” (optimization-based) formulation of hybrid coordinate gridding

T. Mike Chin, RSMAS/U. Miami

Acknowledgement:
Mohamed Iskandarani, Carlisle Thacker

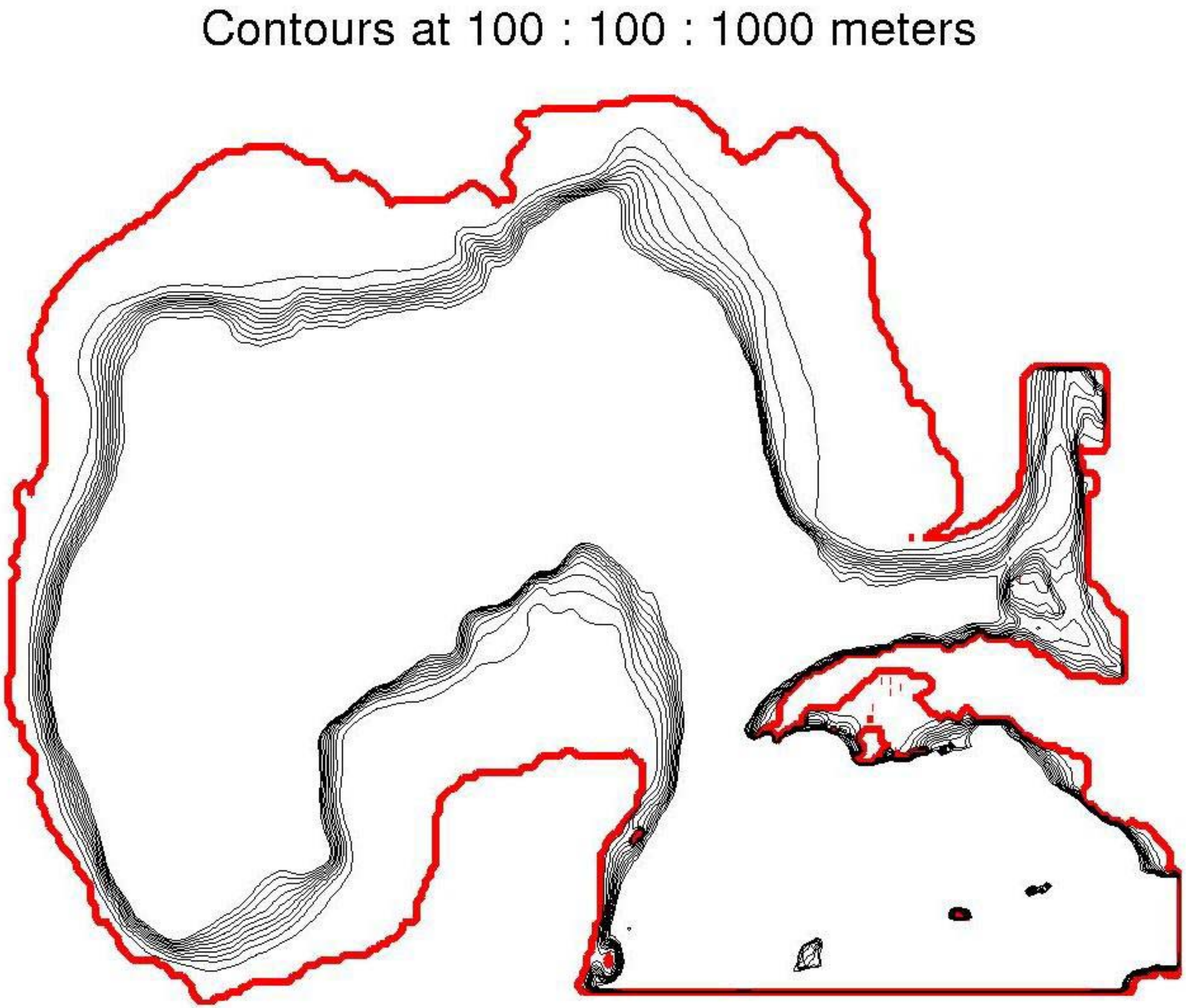
- Hybrid coordinate generation (HybGen), or *regridding for heterogeneous vertical coordinates at each timestep*, is a highly nonlinear procedure.
- Typical data assimilation procedure is a least-squares of linear terms (equivalently, Bayesian analysis on assumed Gaussian densities).
- How can data assimilation be fixed to be (more) consistent with HybGen?

Motivation: Residual SSH over Gulf of Mexico (1/12° HYCOM data assimilation experiment)



ENROIF − (Truth − Forecast)

Errors over shallow areas where a *terrain-following* (depth based) coordinate is used.



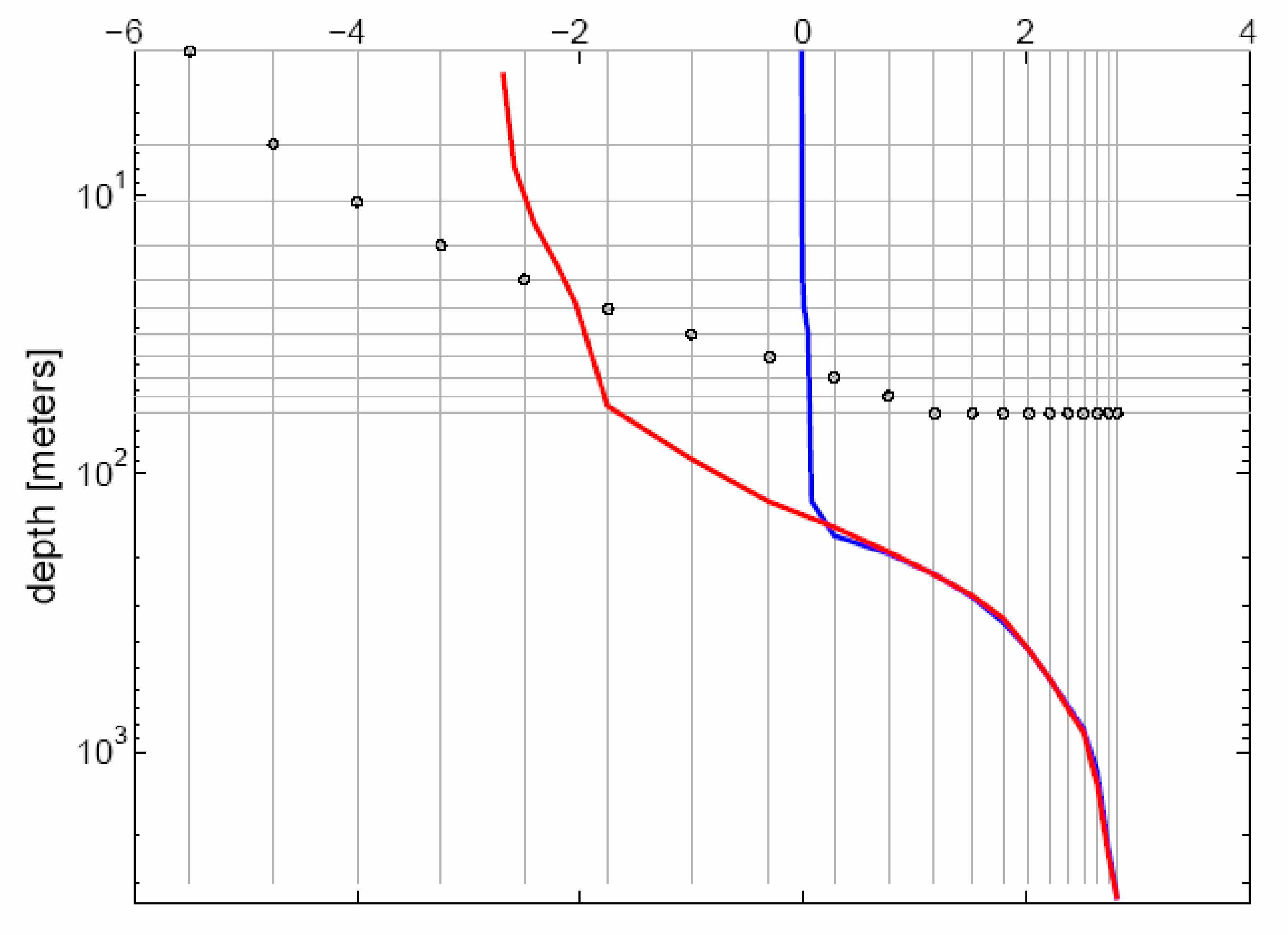
Hybrid coordinate specifications (simplified):

- $L = 20$ layers, indexed by $l = 1, 2, \ldots, L$.
- Minimum layer thickness: h_l,
  e.g., {3.0, 3.5, 4.0, 4.5, … meters}.
  (terrain-based values in shallow areas.)
- Target densities (isopycnal densities): ρ_l.
  (the preferred coordinate if minimum h_l is met.)
- HYCOM: $(\rho, z) \longrightarrow (\alpha\,, p)$

$$\alpha \equiv 1/\rho,\ \ p \equiv \rho\, g\, z,\ \ h \longrightarrow \delta p$$

Mid-domain Gulf of Mexico density profiles: summer & winter

density-1025 [kg/m^3]



- Problem with homogeneous vertical propagation of data:

$$\Delta(SSH) \rightarrow \Delta(\delta p_l),\ \ l = 1,\ldots,L-1$$

- Cannot avoid $\delta p_l < 0$.
- Either δp_l or α_l is constant (target value):

$$\Delta(SSH) = \sum_{l=1}^{L} \Delta(\alpha_l \delta p_l)$$

  - Positivity (minimum value): $\delta p_l \geq \delta p_l^* \geq 0$.
  - Mass conservation: $\int \rho dz$ or $\int dp$ fixed/known.
  - Volume conservation: $\int dz$ or $\int \alpha\, d(\delta p)$ fixed/known.

"HybGen" specifications for data assimilation

To update model state $\mathbf{x}$ by data

1. *Coordinate regimes* (fixed values) need to be known.

2. *Minimum layer-thickness* (inequality) constraint: $\mathbf{Sx} \geq \mathbf{c}$.

3. *Mass/volume conservation (equality)* constraints: $\mathbf{Ax} = \mathbf{b}$.

Sequential assimilation:
(background $_b$ is the model forecast; y is data)

$$\min_{\mathrm{x}} \left\| \mathrm{x} - \mathrm{x}_b \right\|^2_{\mathrm{P}_b^{-1}} + \left\| \mathrm{y} - \mathrm{Hx} \right\|^2_{\mathrm{R}^{-1}}$$

becomes a *constrained* minimization.

Applying constrained optimization

- Increased computational cost:
  - Equality constraints: 2 more variables per profile.
  - Inequality constraints: combinatorial trials, 2^L cases!
- An approximation …
  - Build around existing assimilation routine.

Steps in an assimilation cycle (SSH)

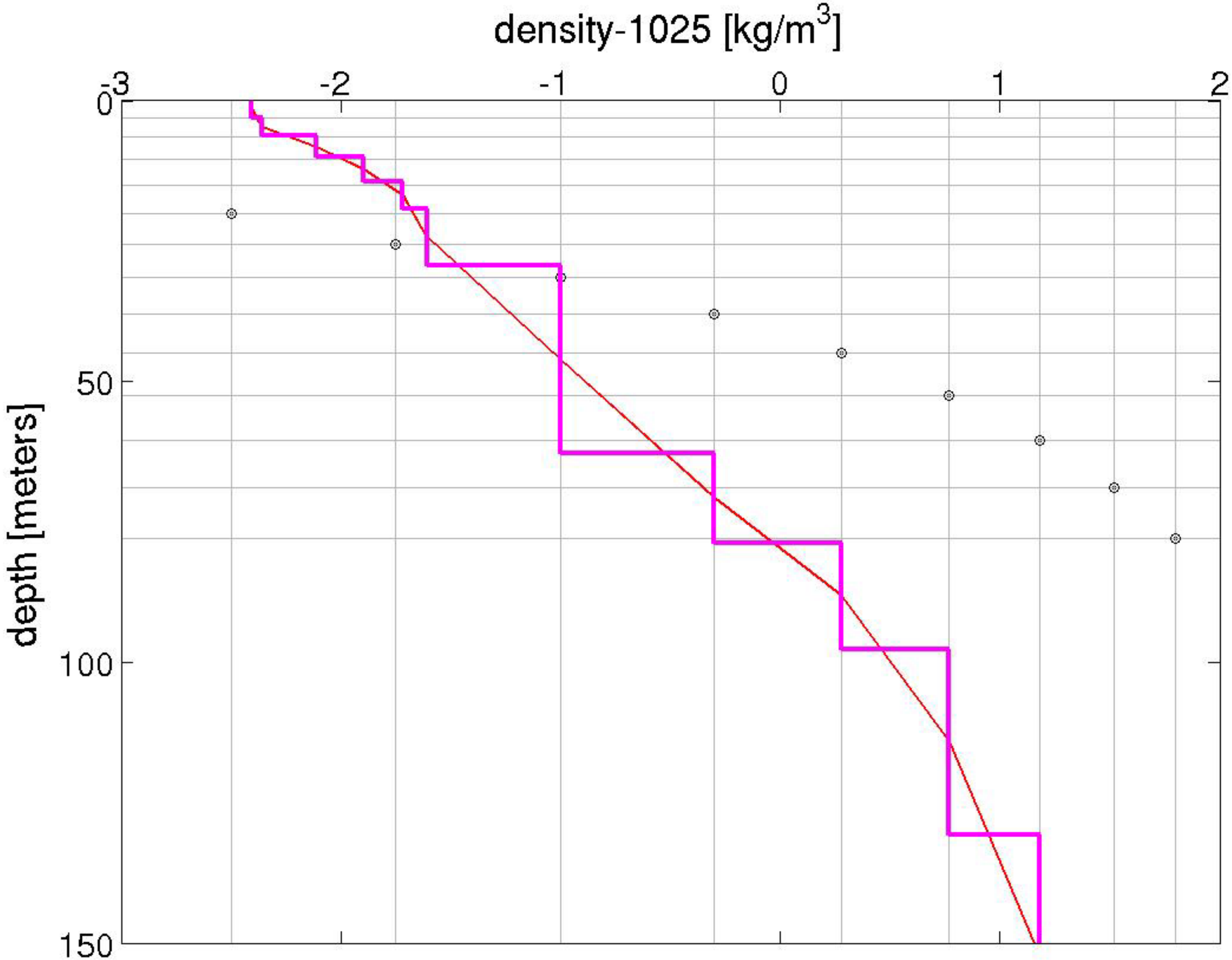
1. Identify the coordinate regimes.

2. Perform (typical) data assimilation steps;
   project the results onto the target coordinate systems.
   $\rightarrow \bar{x}$ (unconstrained state values);

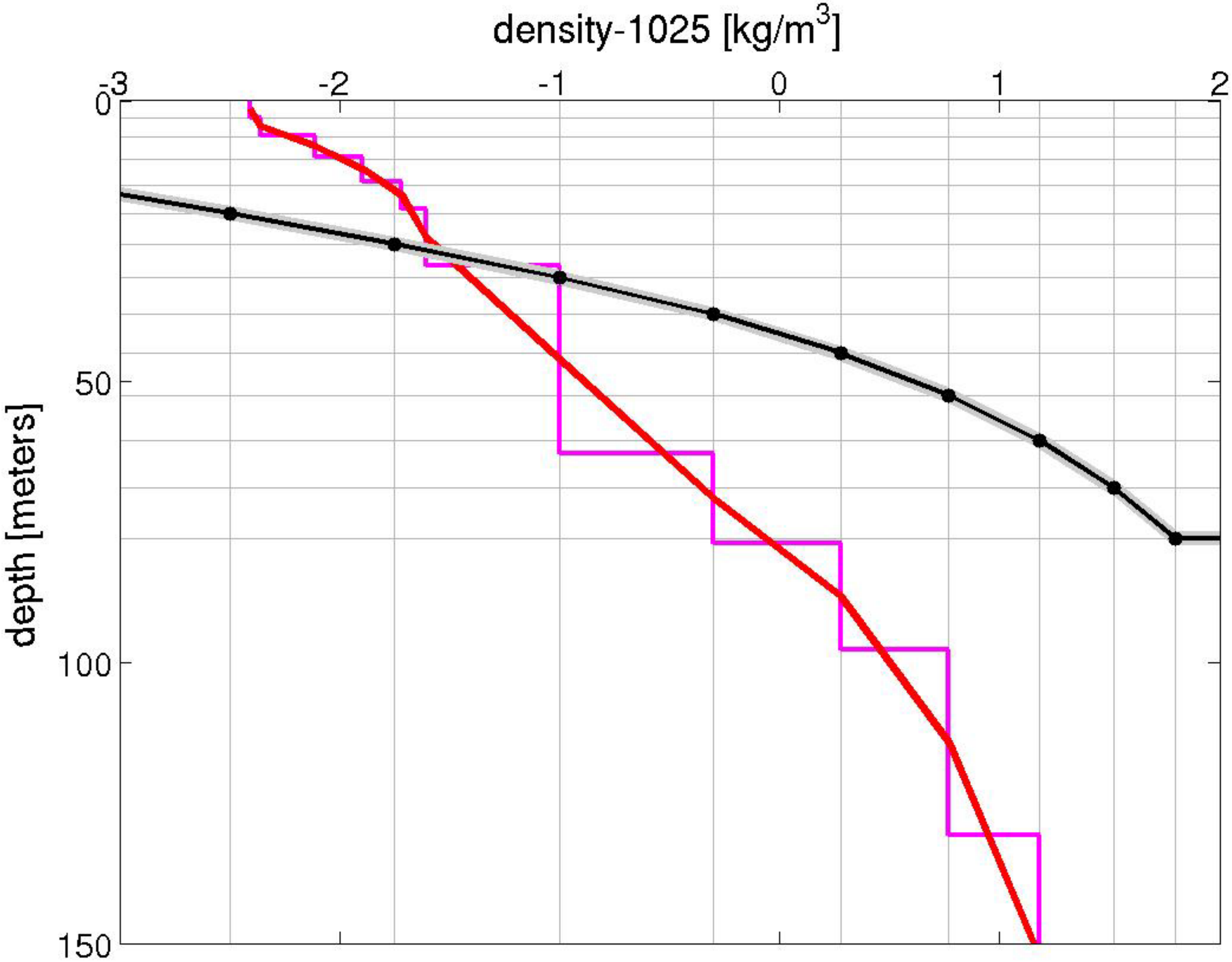
3. Apply equality (mass/volume conservation) constraints:

$$\min_{\mathbf{x}} \; \|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \text{regularization}(\mathbf{x})$$

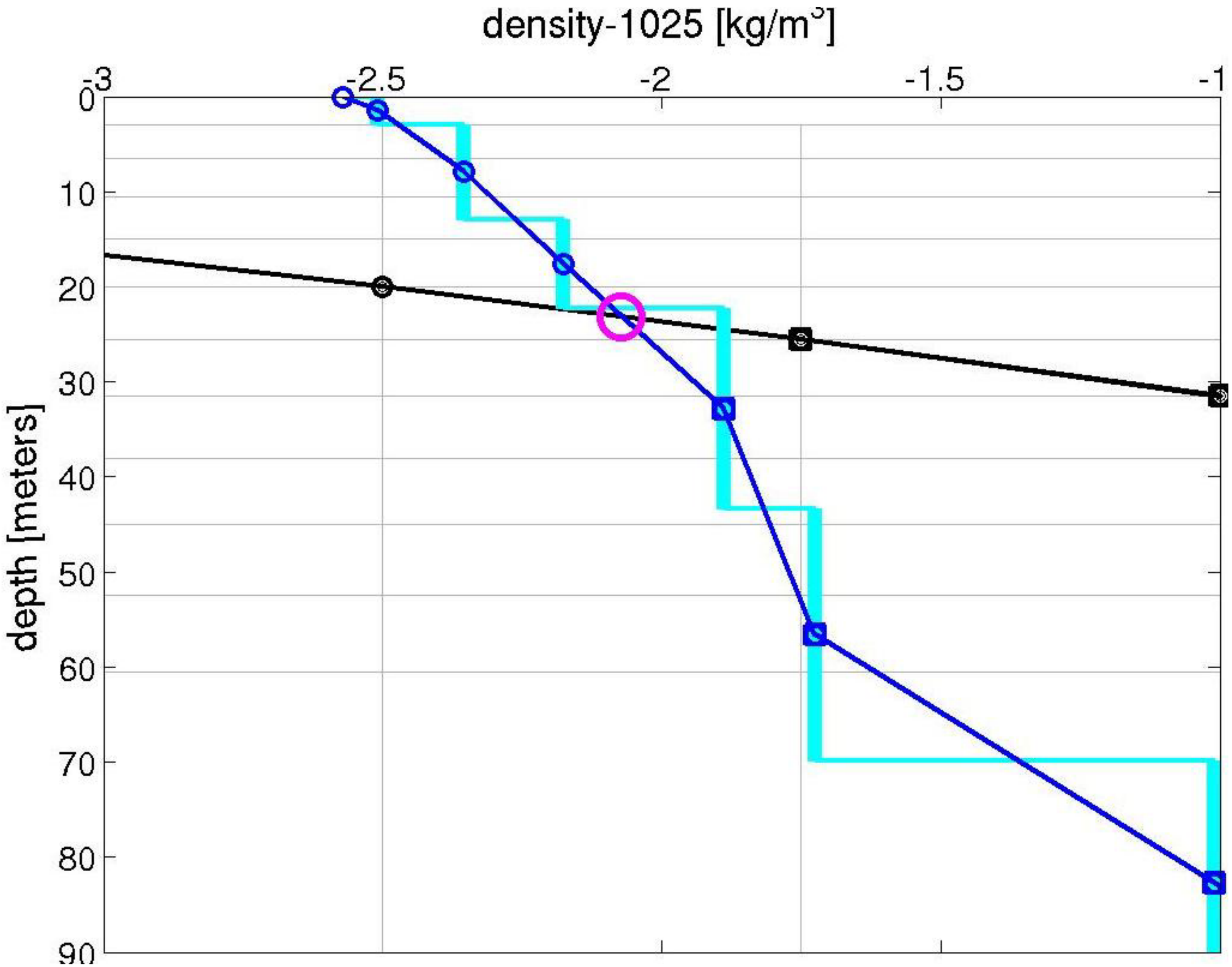
Identify Coordinate Regimes



Coordinate Transition Line



Uncertainty/Ambiguity in Regime Transition



➔ **use several "regime maps".**

Steps in an assimilation cycle (SSH)

Repeated trials for regime-boundaries & inequalities:

1. Identify the coordinate regimes.

1a. Make alternative regime maps by perturbing the boundaries.

2. Perform (classical) data assimilation steps; project the results onto the target coordinate systems.

$\rightarrow \bar{x}$ (unconstrained state values);

2a. Repeat 2 with all regime maps; choose $\bar{\mathbf{x}}$ that minimizes $\|\bar{\mathbf{x}} - \mathbf{x}_b\|^2$.

Steps in an assimilation cycle (SSH) (cont')

3. Apply equality (mass/volume conservation) constraints:

$$\min_{\mathbf{x}} \ \|\mathbf{x}-\bar{\mathbf{x}}\|^2 + \mathrm{regularization}(\mathbf{x})$$

3a. If any δp_l violates its minimum, set δp_l to the minimum and repeat 3.

Application of mass & volume (equality) constraints

- Normalized layer thickness $\delta p_\ell / p_L$ would form a *probability mass function* (PMF).

  $\rightarrow$ $\delta p_\ell \geq 0$, always.

- Mass = 0th order statistical moment;
  Volume = 1st order statistical moment (the mean).

  (In ρ-z system, mass = 1st order and volume = 0th.)

- Existing formulas/procedures in applied probability?
  - Moment matching (0th to (L-1)st)

    $\rightarrow$ regularization needed!
  - "*Histopolation*" (area-matching interpolation).

    $\rightarrow$ provision for $\delta p_\ell \geq 0$?

Experiments:

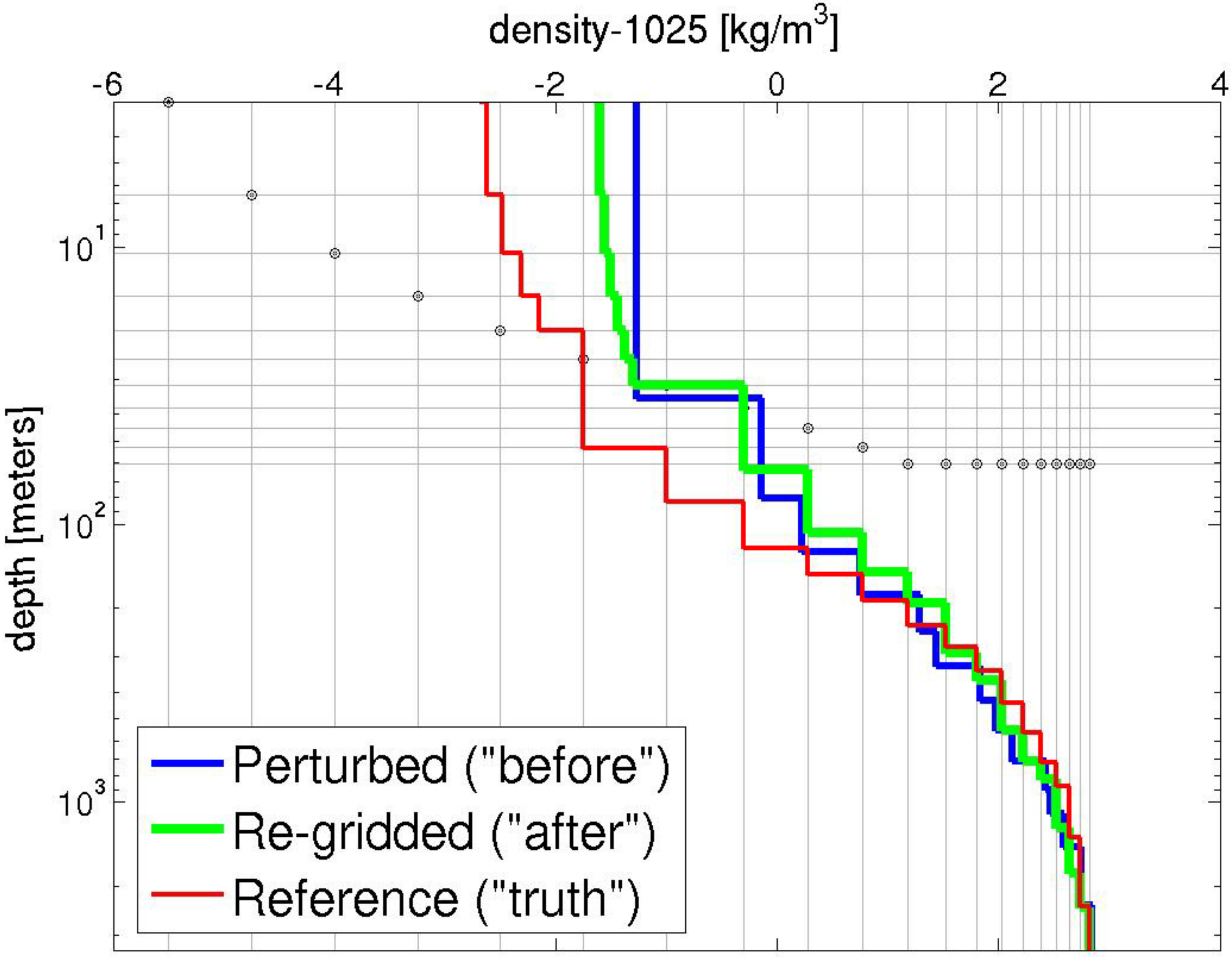
- Given $p_L = \sum_{\ell=1}^{L} \delta p_\ell$ and $\mathsf{SSH} = \sum_{\ell=1}^{L} \alpha_\ell \, \delta p_\ell$

$$\min \sum_{\ell \in \text{isobars}} (\alpha_\ell - \bar{\alpha}_\ell)^2 + \sum_{\ell \in \text{isopycnal}} (\delta p_\ell - \overline{\delta p}_\ell)^2$$

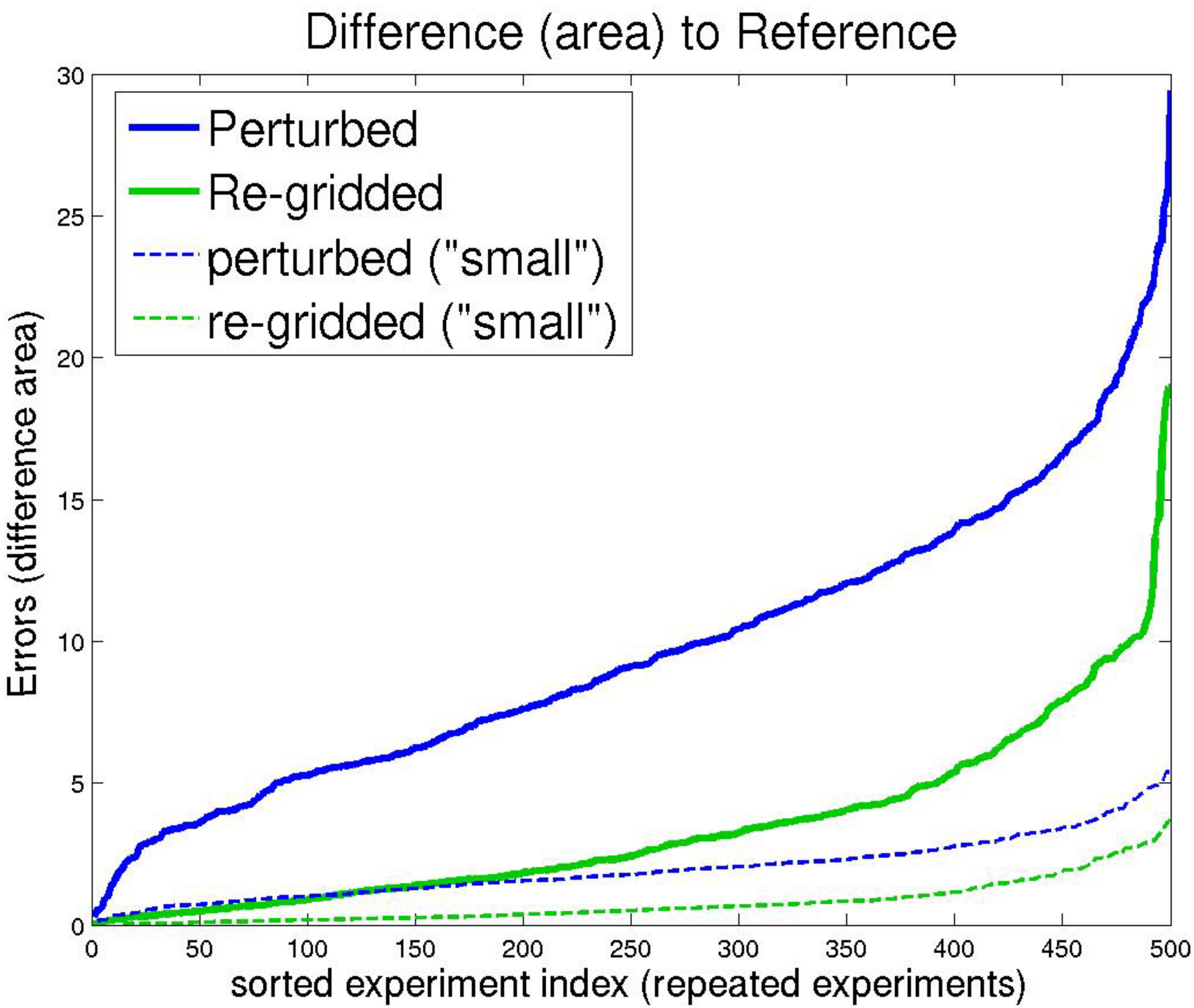
1. A reference (α, p) profile is perturbed.
2. Perturbed profile is re-gridded using the mass (p_L) and volume (SSH) values from the reference.

- Is the re-gridding robust/unique?
- How does it respond to Δ(SSH)?

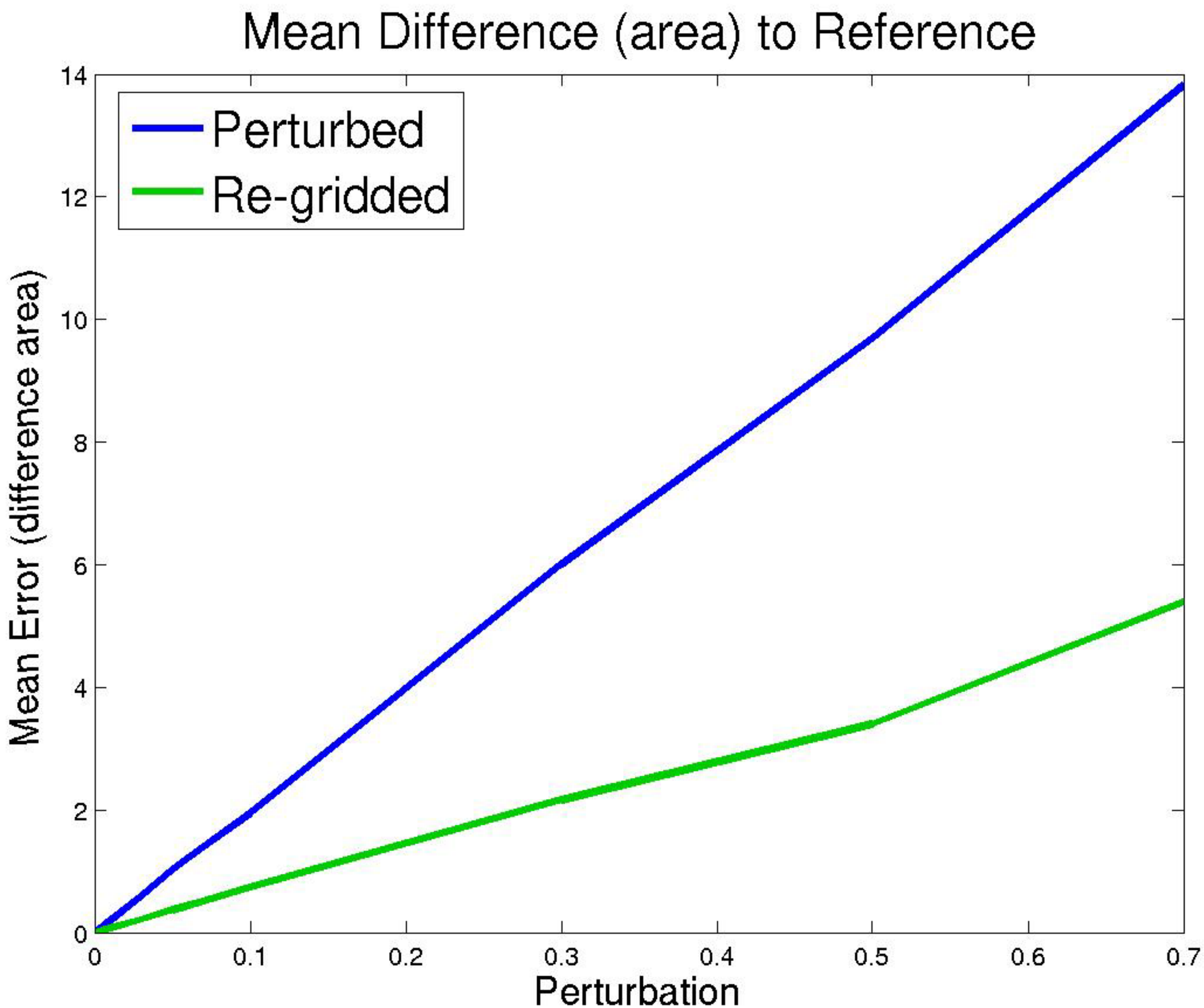
Example: re-gridding of perturbed density profile



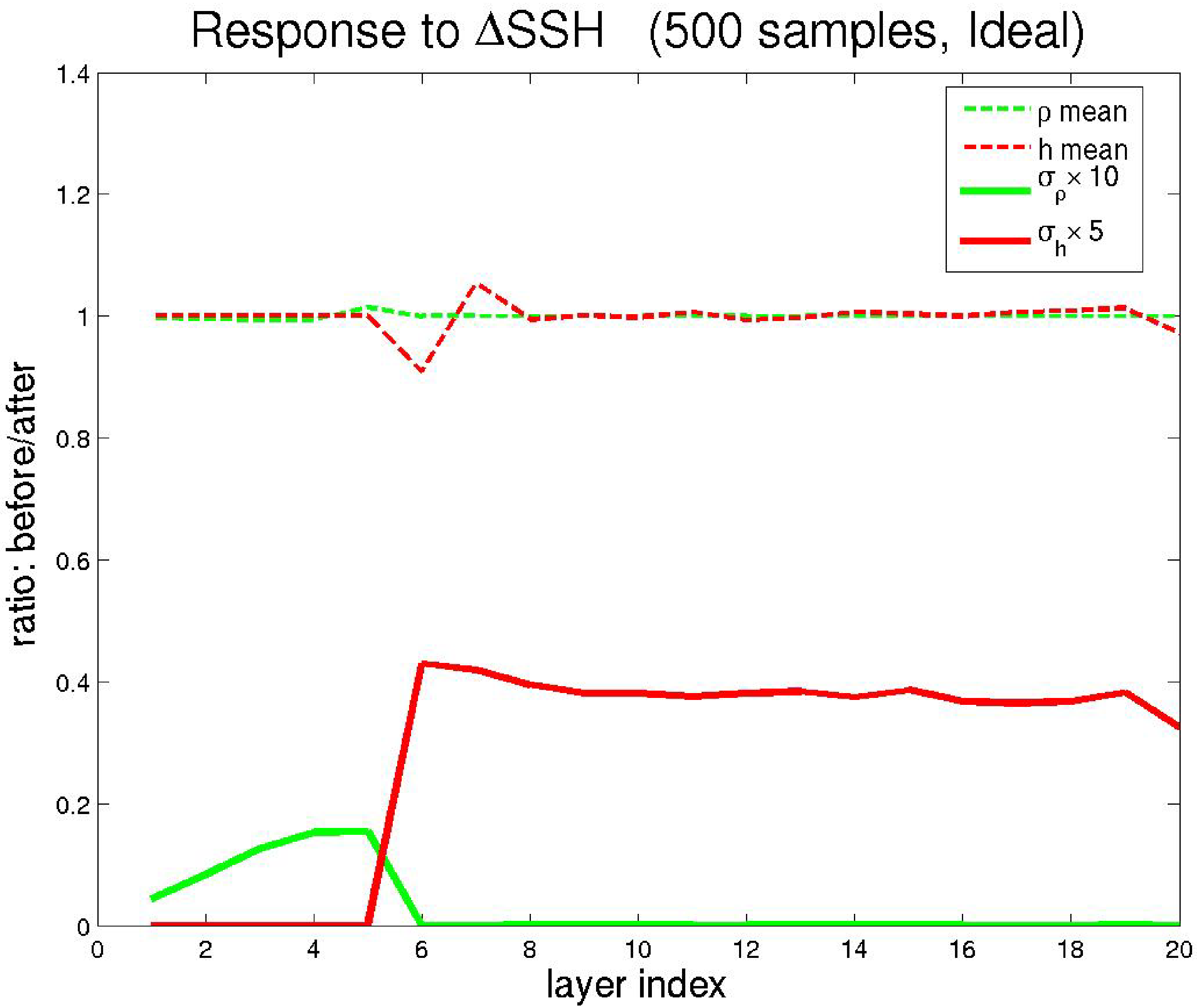
"Errors" before & after re-gridding



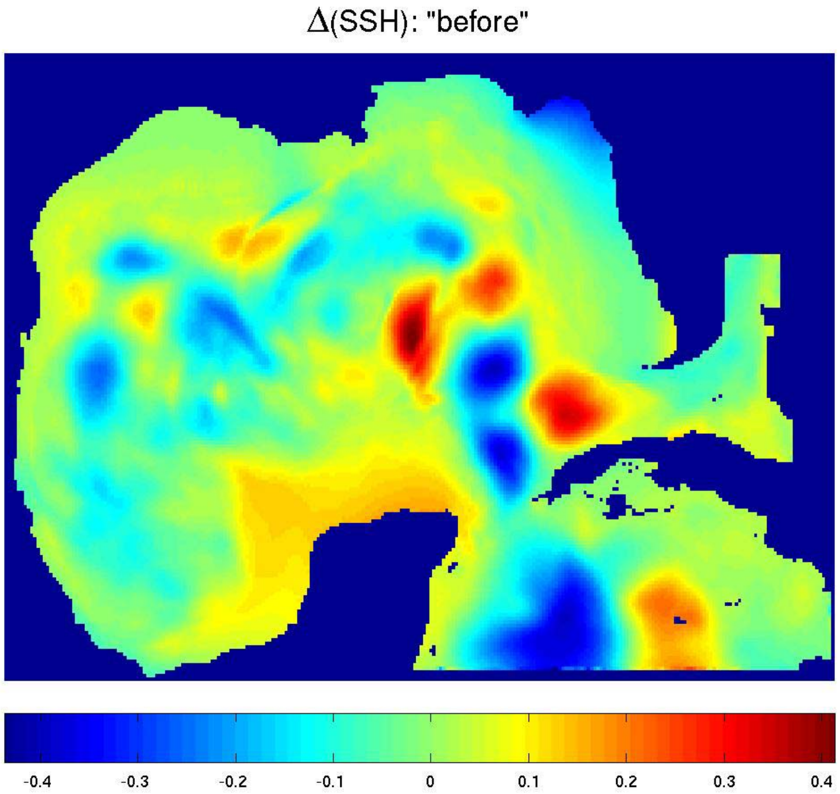
“Error” as a function of perturbation size

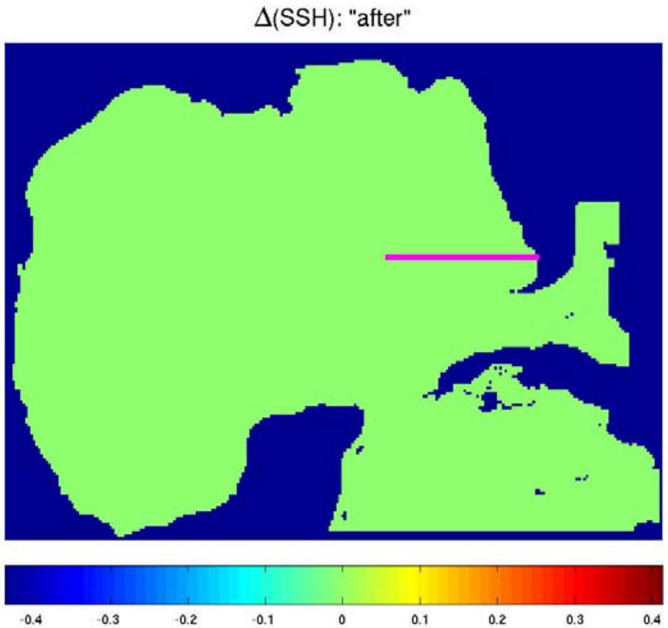


Random variations in SSH: vertical distribution

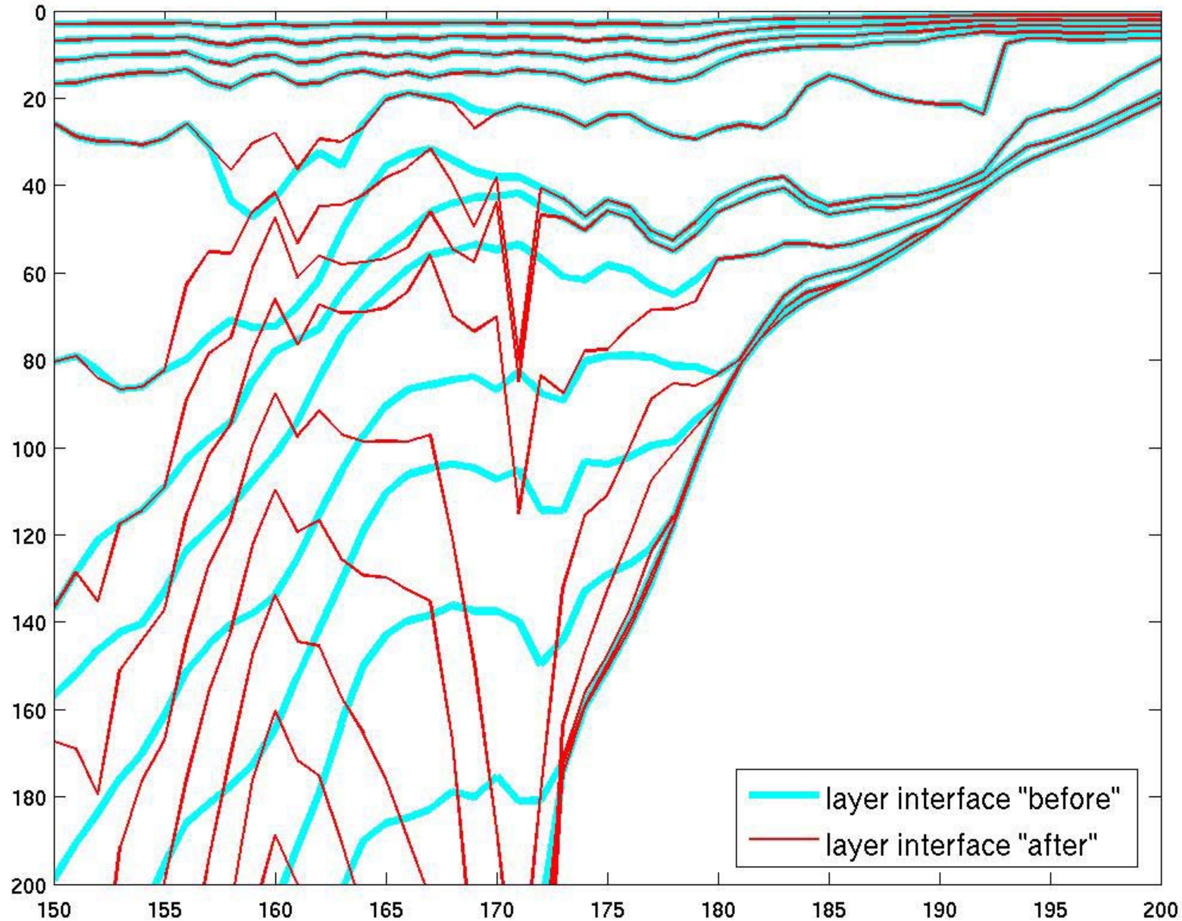


SSH-constrained regridding over Gulf-of-Mexico (HYCOM)





Re-gridded layer-interfaces over Florida Shelf



Summary

- A modest (computationally etc) scheme for post-assimilation adjustments:
  - *Coordinate regime identification*
  - *Lagrange multipliers*

    $\rightarrow$ being applied to SSH (and T profile) data sets

- A more general formulation: functional/continous representations

  $\alpha(s),\ p(s),$ etc.
  - *"Histopolation"* (area-preserving spline)
  - Area-preserving sampling (!)

- Other approaches
  - *Ensemble filter:* constraints applied during ensemble generation.
  - *4D-Var:* HybGen-conscious adjoint (linearized backward) model.