A Priori Testing of Adaptive Sampling and Sparse PC representations for Ocean General Circulation Models*

J. Winokur[1], P. Conrad[2], I. Sraj[1], A. Alexanderian[1],

M. Iskandarani[3], A. Srinivasan[3], Y. Marzouk[2], O. Knio[4]

[1]Johns Hopkins University

[2]Massachusetts Institute of Technology

[3]University of Miami

[4]Duke University

*Work supported by: ONR, DOE/ASCR, DOE/SciDAC

Introduction

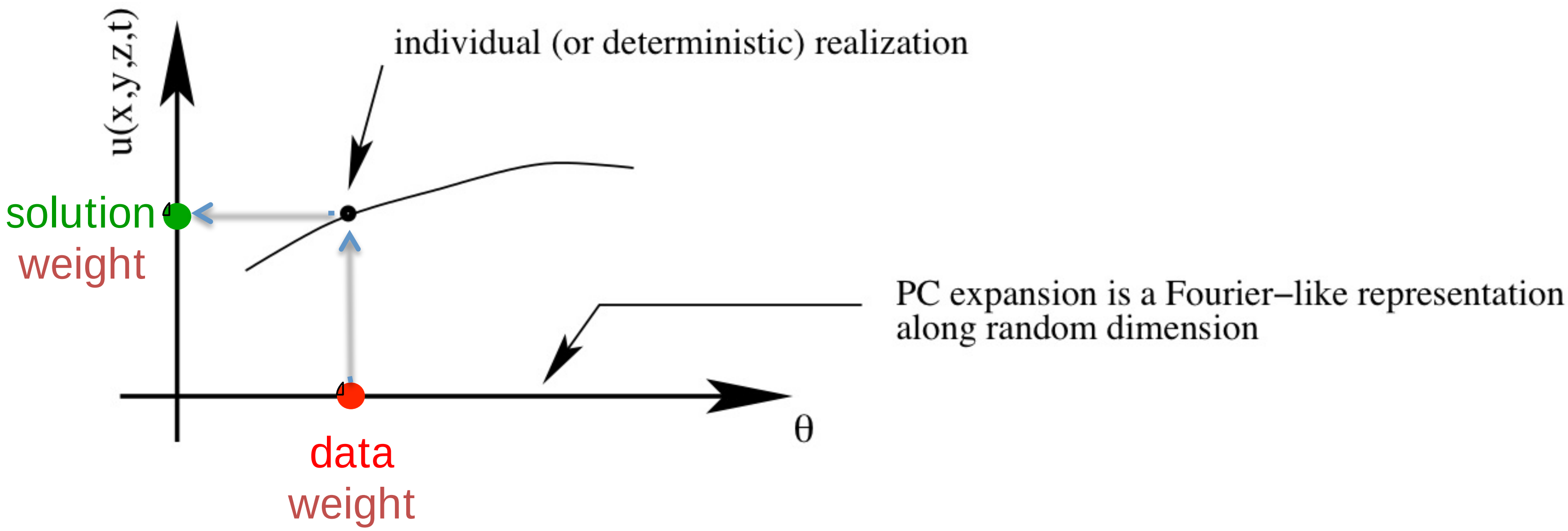
- OCGMs inherently rely on subgrid scale parametrizations that account for the impact of unresolved processes
- These parametrizations frequently involve uncertain inputs
- Assessment/calibration of uncertain inputs through systematic sampling is prohibitively expensive
- Objective of this work is to exploit existing database of model realizations to explore the potential of new sparse adaptive quadrature approaches in minimizing computational effort required to propagate and represent impact of uncertain model inputs

Outline

- Uncertainty representation and quantification framework
- Ocean circulation database:
  - isotropic sampling
  - solution behavior
- Adaptive sparse pseudospectral approach
- A priori tests:
  - error analysis
  - performance metrics
- Outlook

Spectral approach to UQ

- Probabilistic approach to UQ



- Stochastic solution is sought in a product space
  - probability space (x axis)
  - deterministic solution space (y axis)
  - both are assumed to have Hilbert space structure, with a countable orthogonal basis

Polynomial Chaos

- mean-square convergent expansion in the space of random variables:

$$u(x,t;\boldsymbol{\xi}) = \sum_k u_k(x,t)\Psi_k(\boldsymbol{\xi})$$

  - solution mode
  - vector of canonical random variables that are suitably used to parametrize the uncertain inputs
  - orthogonal basis in the space of square integrable random functions
  - solution mode (coordinate in Hilbert space) is unknown to be determined

Examples

- **Legendre:**

$$Le_n(x) = \frac{1}{2^n}\sum_{l=0}^{[n/2]}(-1)^l \begin{pmatrix} n \\ l \end{pmatrix}\begin{pmatrix} 2n-2l \\ n \end{pmatrix} x^{n-2l}$$

  - orthogonal over [-1,1] with uniform measure

- **Hermite:**

$$He_n(x) = n!\sum_{m=0}^{[n/2]}(-1)^m \frac{1}{m!2^m(n-2m)!}x^{n-2m}$$

  - orthogonal over (-∞,∞) with Gaussian measure

Non-Intrusive Spectral Projection (NISP)

- Alternative to Galerkin formalism in situations where modification of production or legacy codes is not feasible
- Essentially amounts to a collocation approach to the evaluation of probability integrals

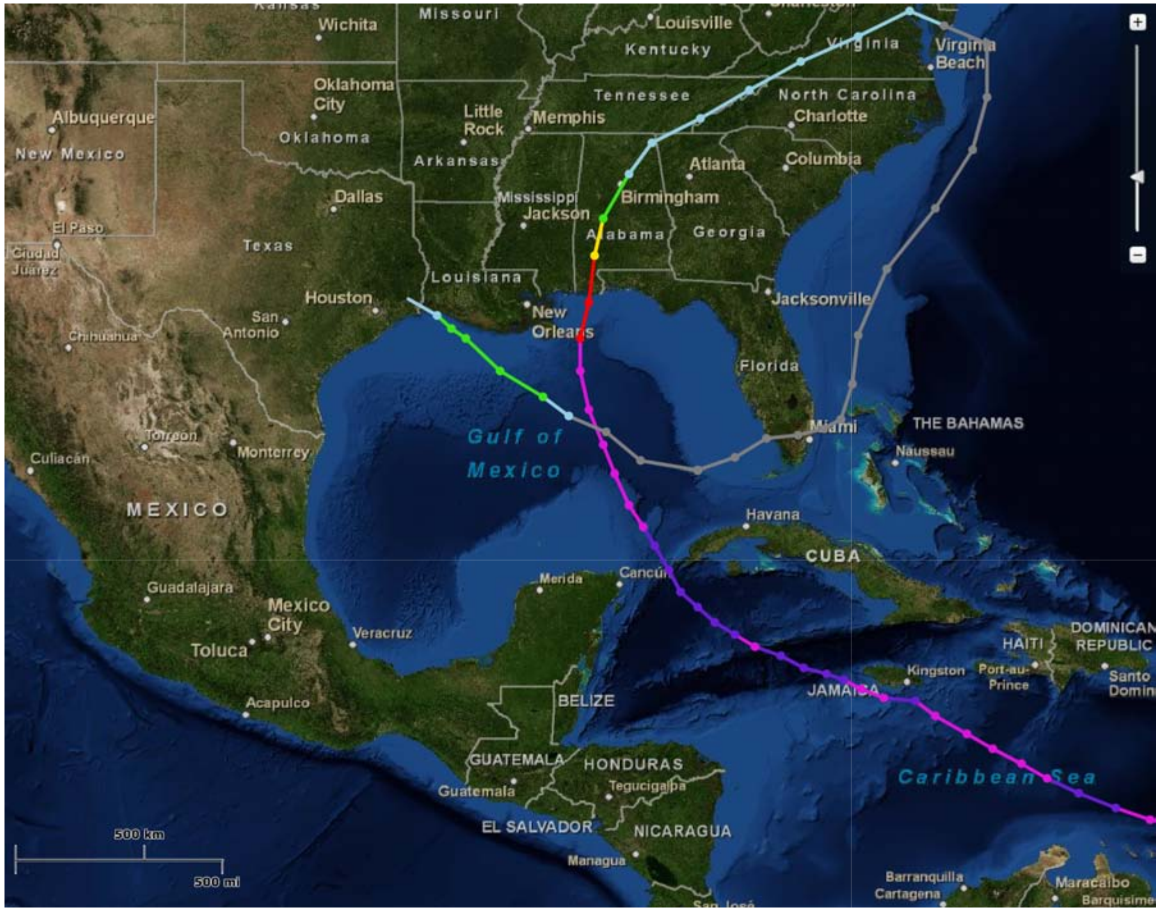
uncertain inputs → Model → quantity of interest

$$\langle \Psi_k^2 \rangle Q_k = \langle Q(\xi) \Psi_k(\xi) \rangle \approx \sum_{i=1}^{N} Q(x_i) w_i$$

Why spectral expansions?

- Efficiency:
  - provided, generally, that sufficient smoothness is present
- Probabilistic framework:
  - provides fundamental tools for analyzing stochastic errors, convergence, etc…
- Format:
  - key to the machinery of approximation theory (optimization, sensitivity analysis, inverse design, etc…)

Application focus



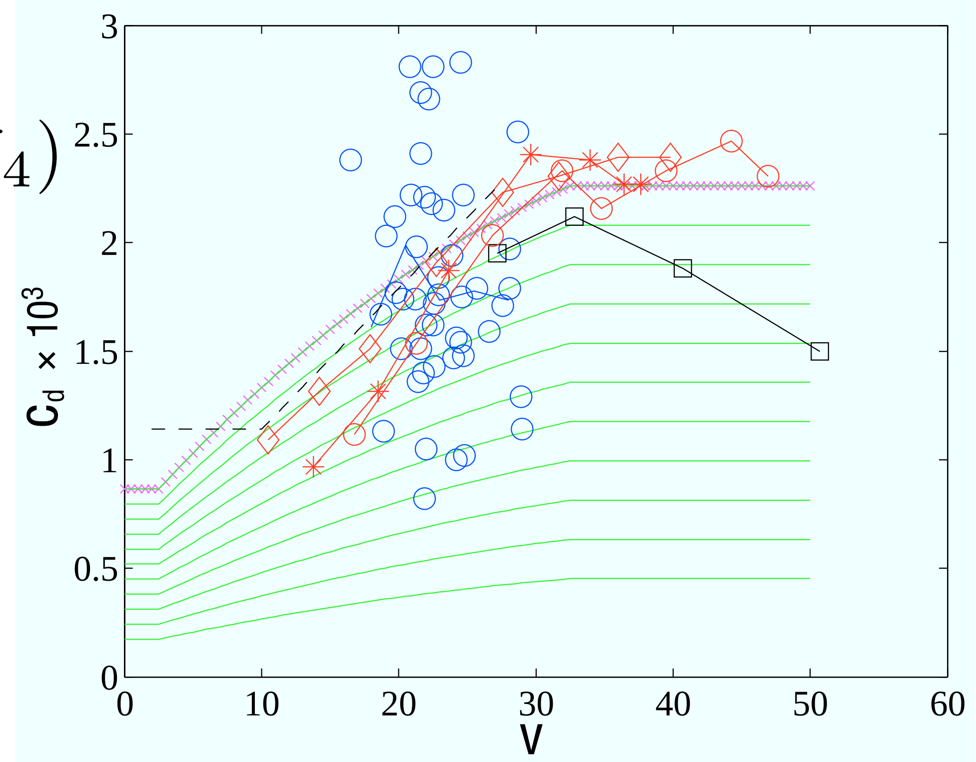
- ➢ Role of uncertainty in wind coupling parametrization and subgrid mixing during Hurricane Ivan (Alexanderian et al. 2012)

Uncertain inputs

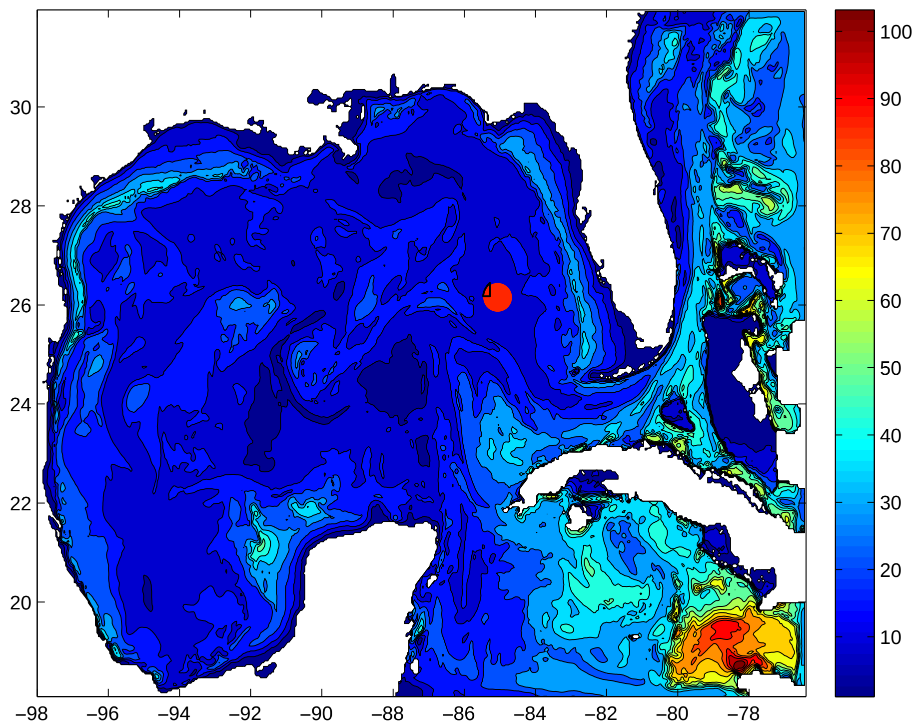
- Parameters of subgrid mixing model (KPP)
  - Richardson number $\mathcal{U}(0.1, 0.7) \quad (\xi_1)$
  - Background viscosity $\mathcal{U}(10^{-4}, 10^{-3}) \quad (\xi_2)$
  - Background diffusivity $\mathcal{U}(10^{-5}, 10^{-3}) \quad (\xi_3)$

- Wind drag coefficient (ξ_4)
  - simple scaling

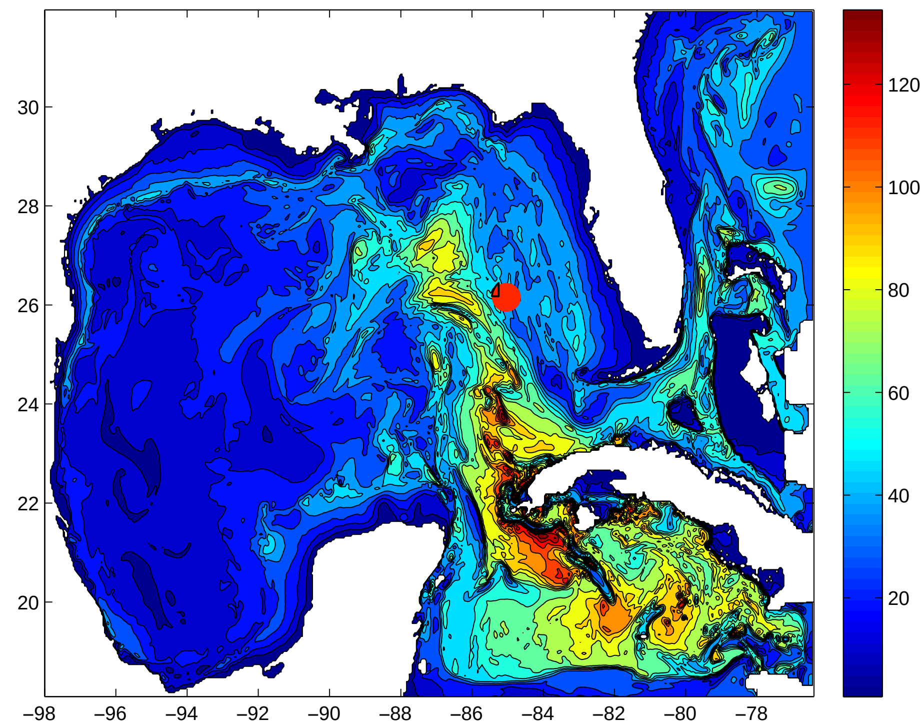
Blue circles: aircraft (French et al., 2007); Black: drop sonds (Powell et al., 2003); Dashed: drop sonds (Powell, 1981); Red: lab data (Donelan et al., 2004); Pink Line: HYCOM (Kara et al., 2000)



Illutration: mixed layer depth



t=69 hrs

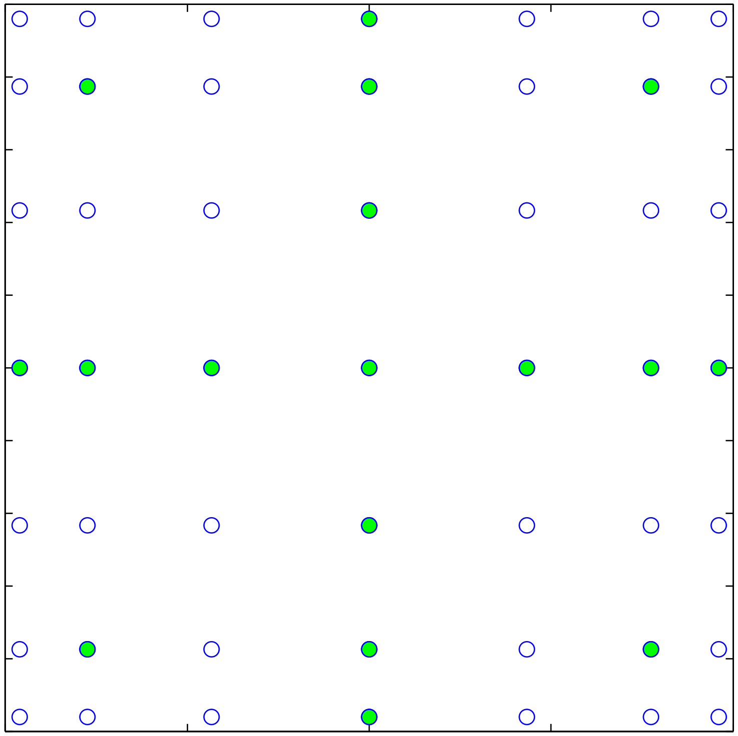


t=159 hrs

- High resolution HYCOM (1/25°) simulations
- Time period: Sep 1 – 10, 2004 (t=0 corresponds to 9/1)
- Track of Hurricane Ivan

Sparse quadrature database

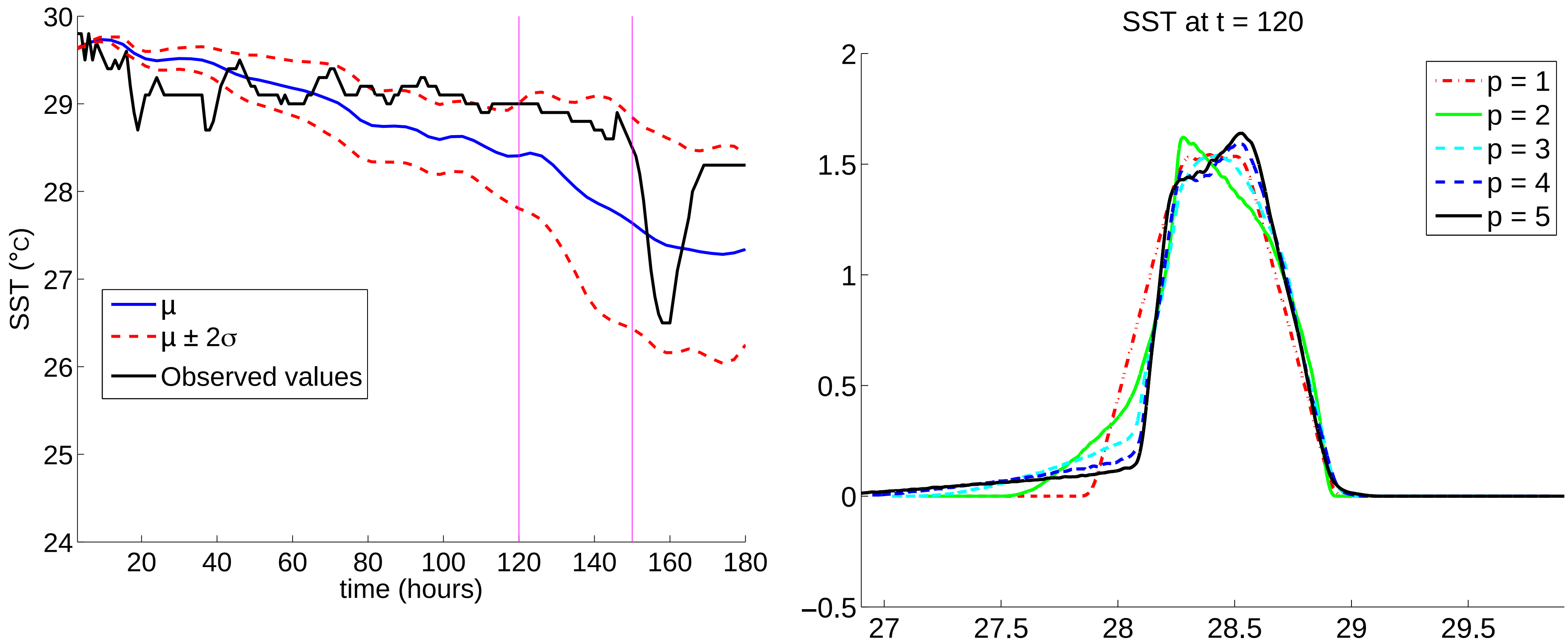
- Smolyak/Gauss Patterson
  - non-adaptive
  - nested grid
  - level p resolution ~ yields order p PC



- Adaptive extension
  - variance-based error indicators
  - "optimal" pseudo-spectral construction

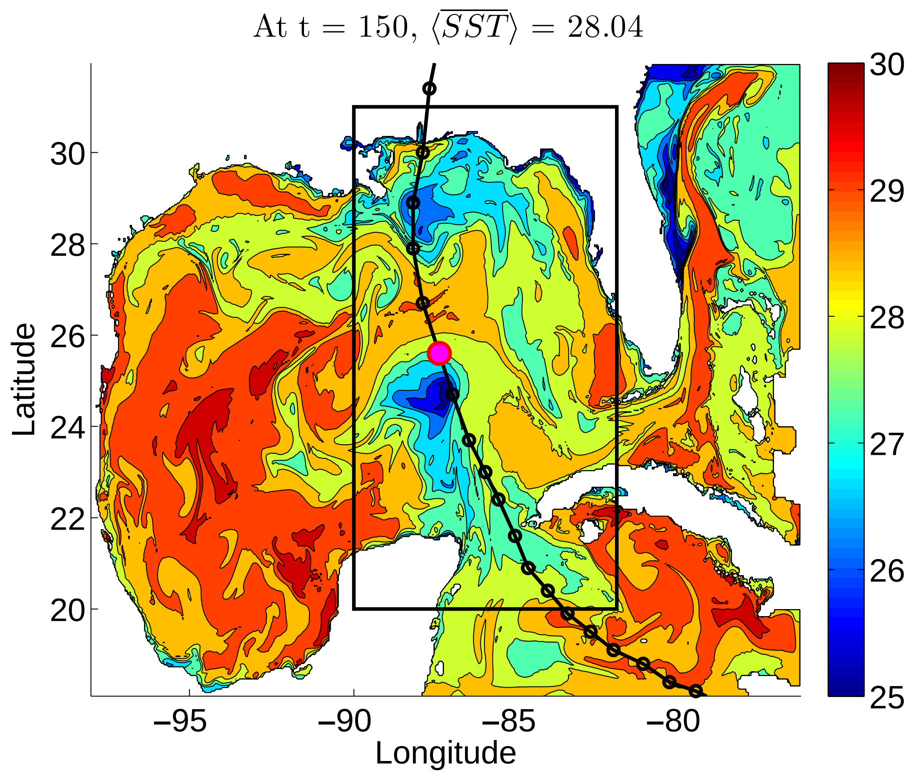
- Both approaches well suited to moderate dimensionality

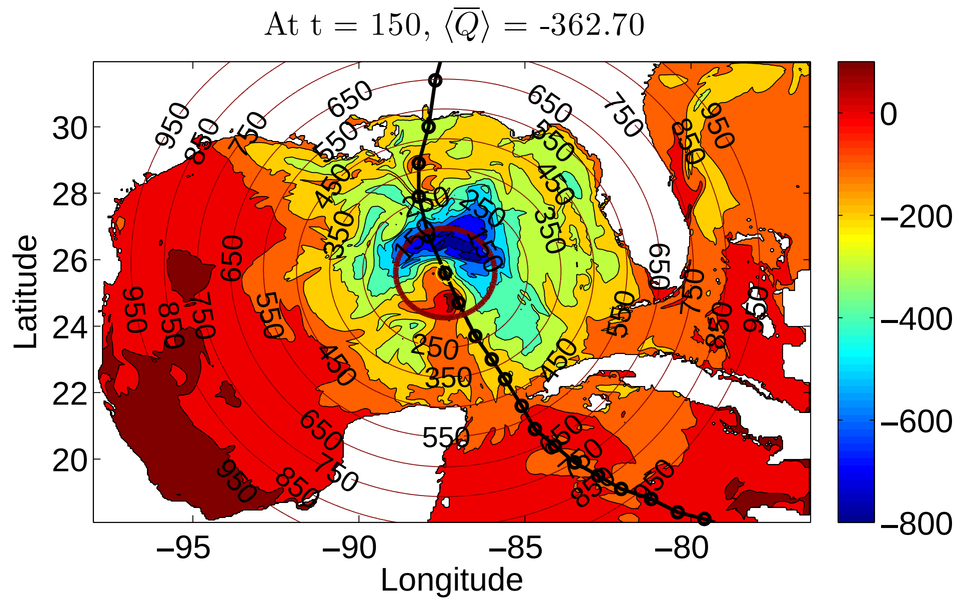
Non-adaptive sparse sampling



- Systematic refinement up to level 5 – total of 385 realizations ➔ reasonable stochastic resolution
- Of course, direct comparison of buoy measurement to OGCM prediction is not well founded

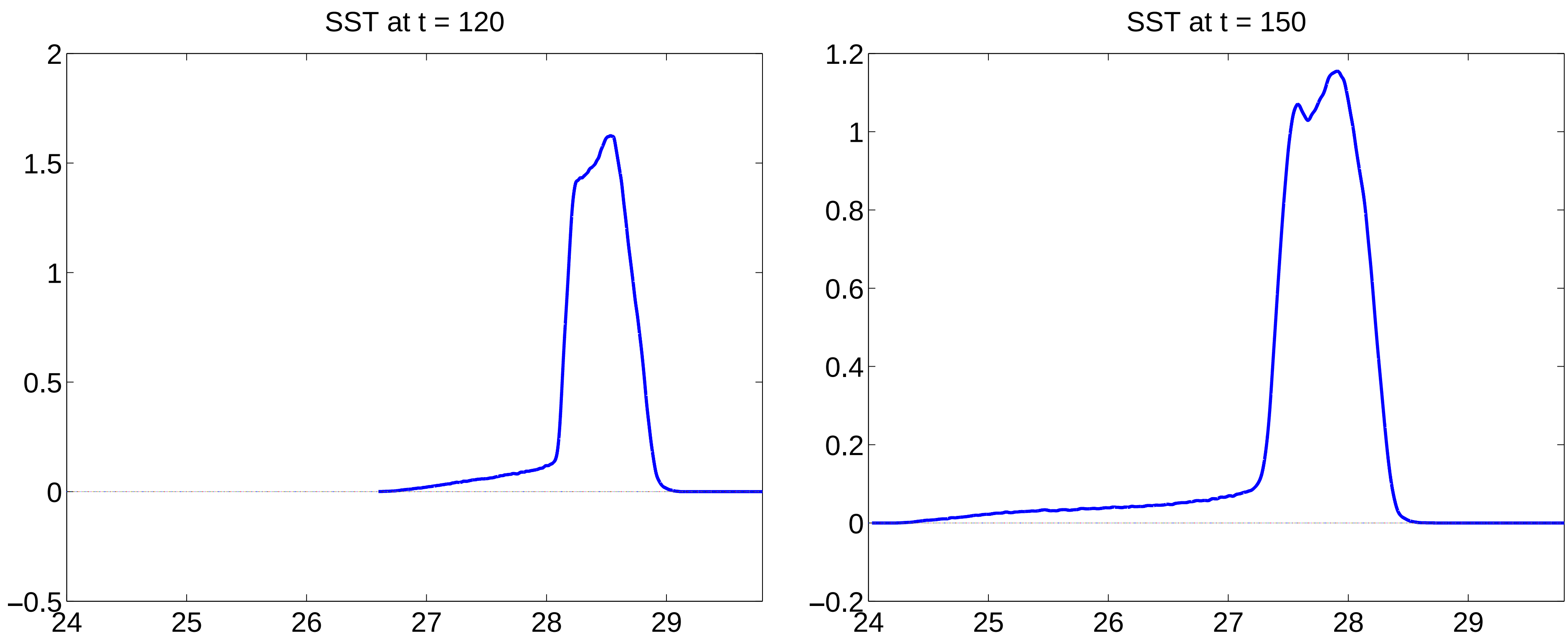
Variance analysis





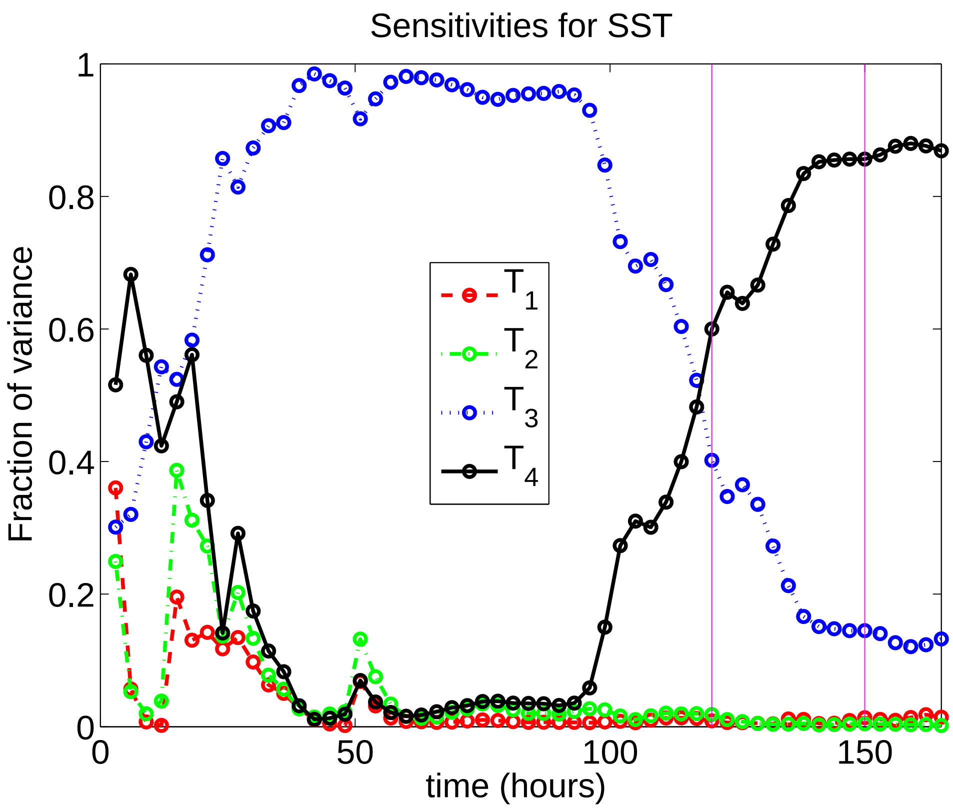
- Based on:
  - mean SST in "skill" region; and
  - heat flux (energy uptake) beneath the hurricane

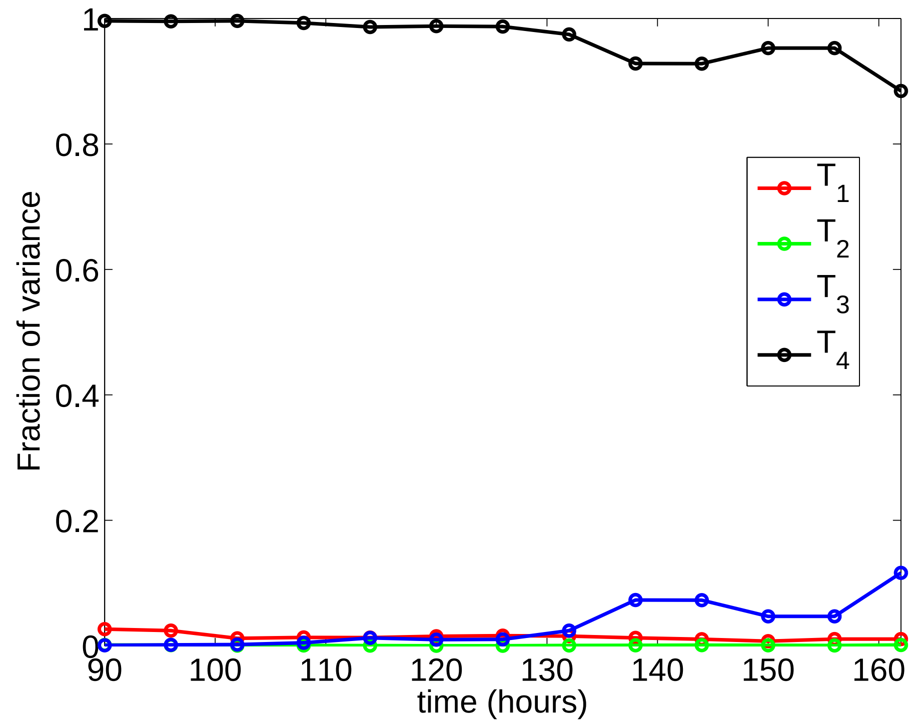
Average quantities



- can also show complex behavior:
  - extended tails observed locally also observed in integral quantities!

Total sensitivities





- contributions of individual components of random input vector to total variance can be directly obtained from functional representation
- wind coupling parameterization dominant at late times

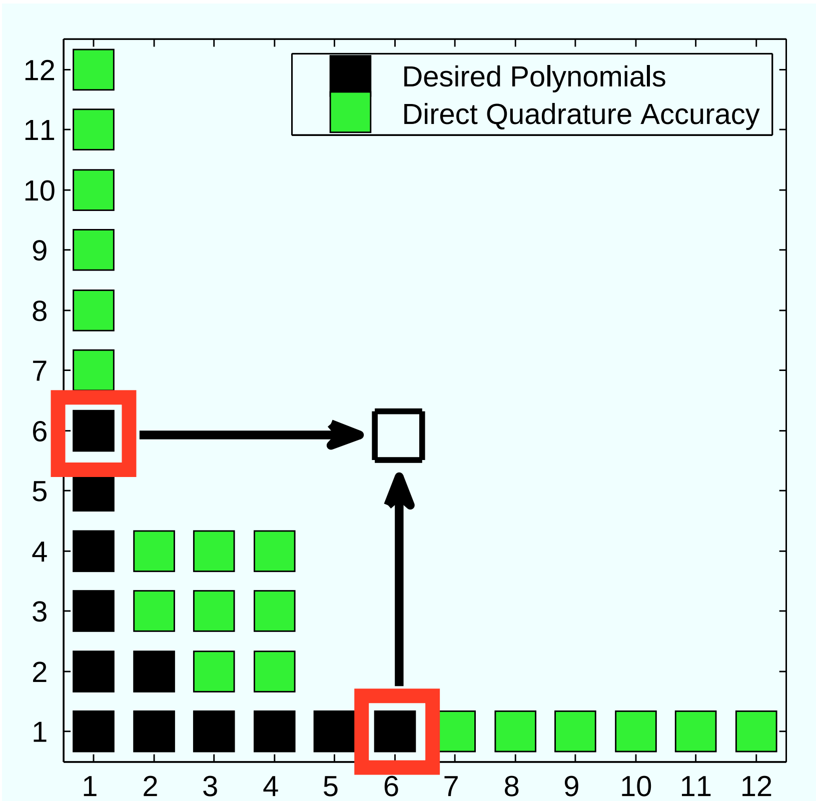
Isotropic Database

- Smolyak quadrature / Gauss-Patterson rule
  - level 5 – isotropic refinement
  - enriched to level 7 in the third (background diffusion) and fourth (wind drag) dimensions based on results of sensitivity analysis
- Exploit existing database to investigate further reduction of CPU and storage requirements:
  - a priori testing of adaptive quadratures
  - a priori testing of error indicators
- Driver: how can experiences gained from analysis of GOM Ivan response help us capitalize on ITOP experiment:
  - can relatively sparse ocean vertical structure observations help refine prior distributions?

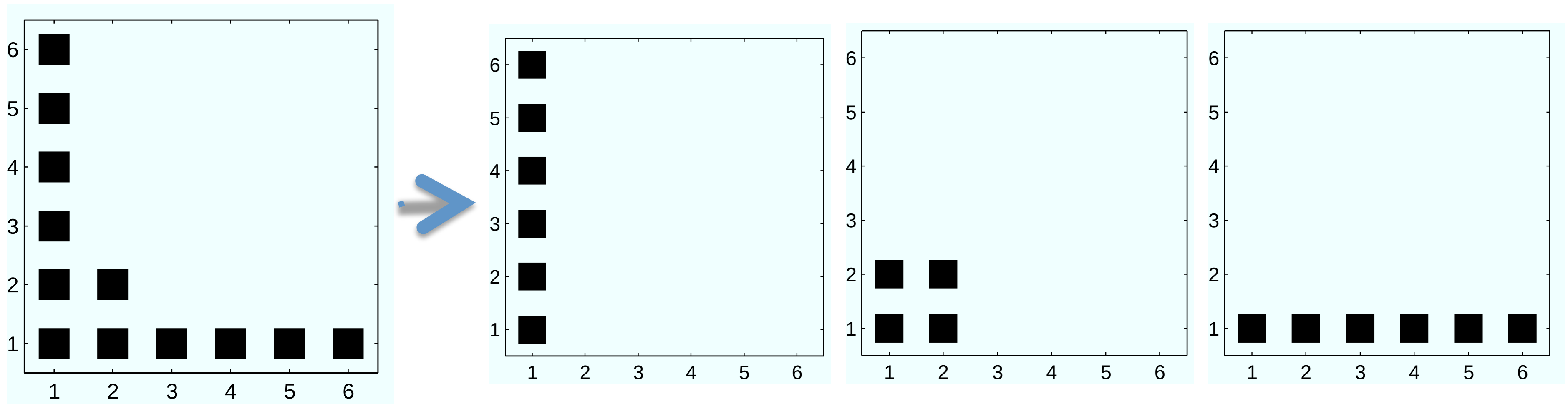
Smolyak pseudospectral construction – 1

- ➢ aims to include all polynomials on general sparse grid that can be computed without internal aliasing

  - With direct quadrature rule number of realizations needed to avoid internal aliasing (ensure discrete orthogonality) is substantially larger than index set one wishes to retain



Smolyak pseudospectral construction – 2



- New Smolyak pseudospectral approximation (Constantine et al., 2011; Conrad & Marzouk, 2012):
  - essentially a telescoping sum of pseudospectral projections (each internal-aliasing-free)
  - avoids explicit compatibility test
  - efficient use of data
  - adaptive variants avoid hand tuning

Smolyak pseudospectral construction – 3

- write multidimensional tensor product quadrature as sum of "surpluses" over (any admissible) index set:

$$\mathcal{S}_L^d = \sum_{\boldsymbol{k} \in \mathcal{K}} \left(\Delta_{k_1}^1 \otimes \cdots \otimes \Delta_{k_d}^1 \right)$$

$$\Delta_k^1 = \mathcal{S}_k^1 - \mathcal{S}_{k-1}^1$$

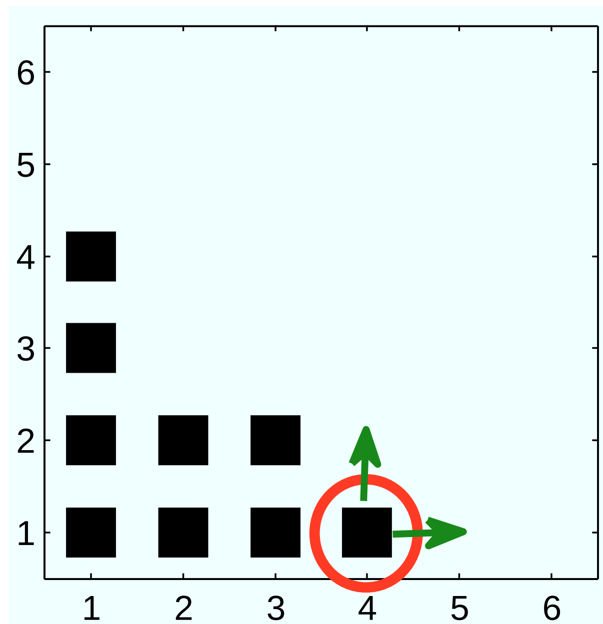
$$\mathcal{S}_L^1 f = \sum_{k(L)} w_k f(x_k)$$

- immediately leads to definition of (variance-based) local error indicator

$$\epsilon(\boldsymbol{k}) \equiv \| \Delta_{k_1} \otimes \cdots \otimes \Delta_{k_d} \|$$

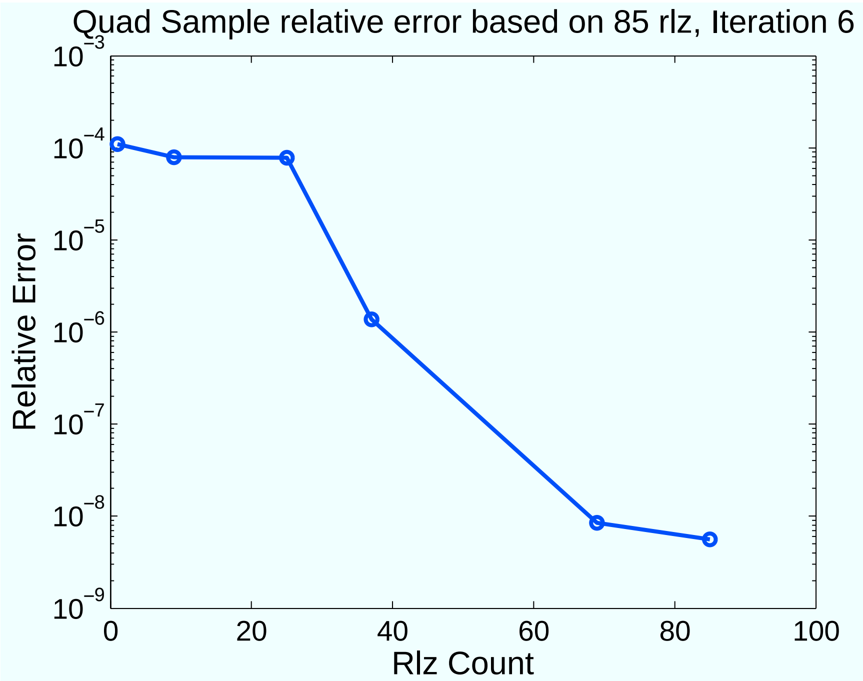
Adaptive refinement

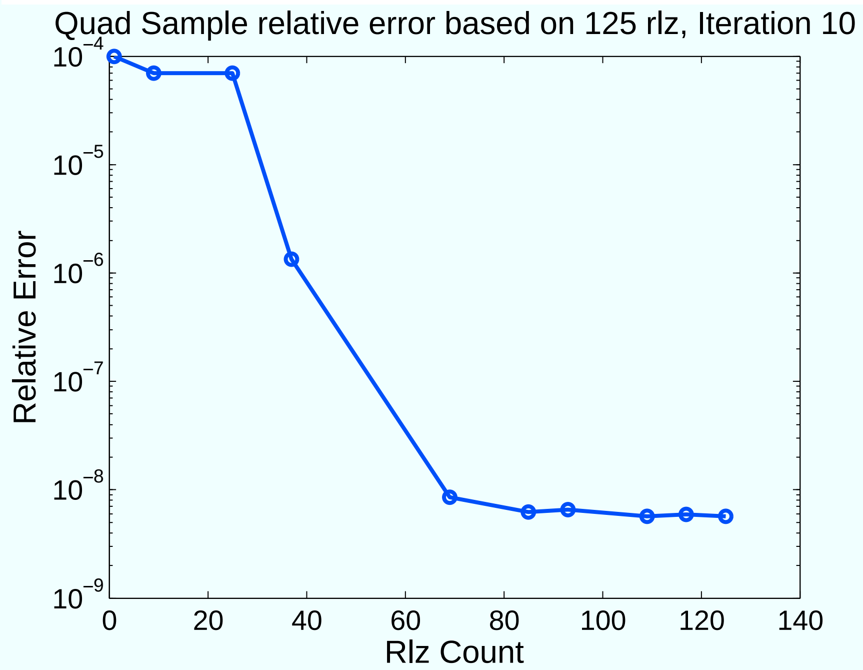
➢ Greedy algorithm that seeks terms that contribute most to $\|\tilde{f}\|_2^2$



- pick index with highest indicator
- add all admissible forward neighbors
- repeat until stopping criterion is reached

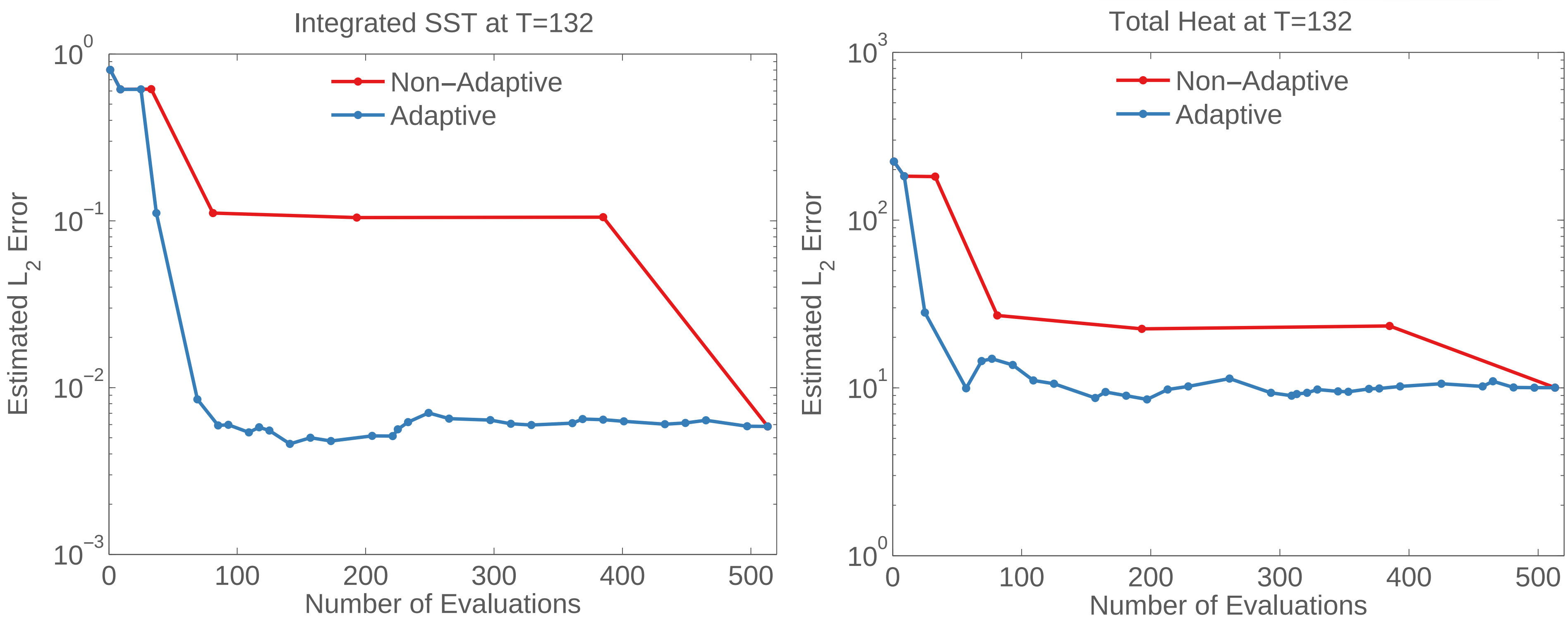
A priori test of adaptive quadrature – 1





- order-of-magnitude compression observed average SST(60hr)
- plateau in the error curves reached when adaptive refinement requires realizations not contained in the database

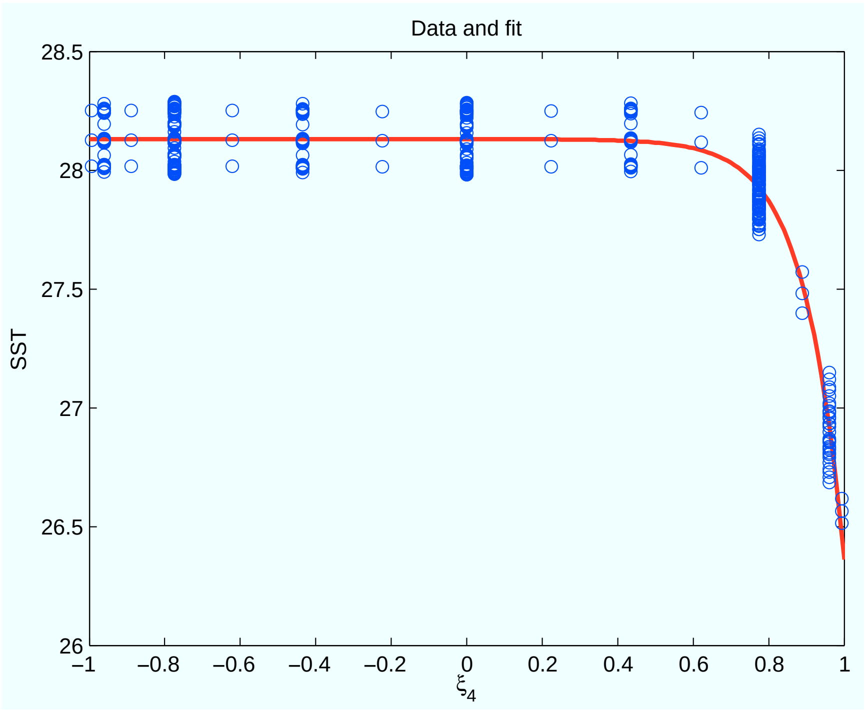
A priori test of adaptive quadrature – 2

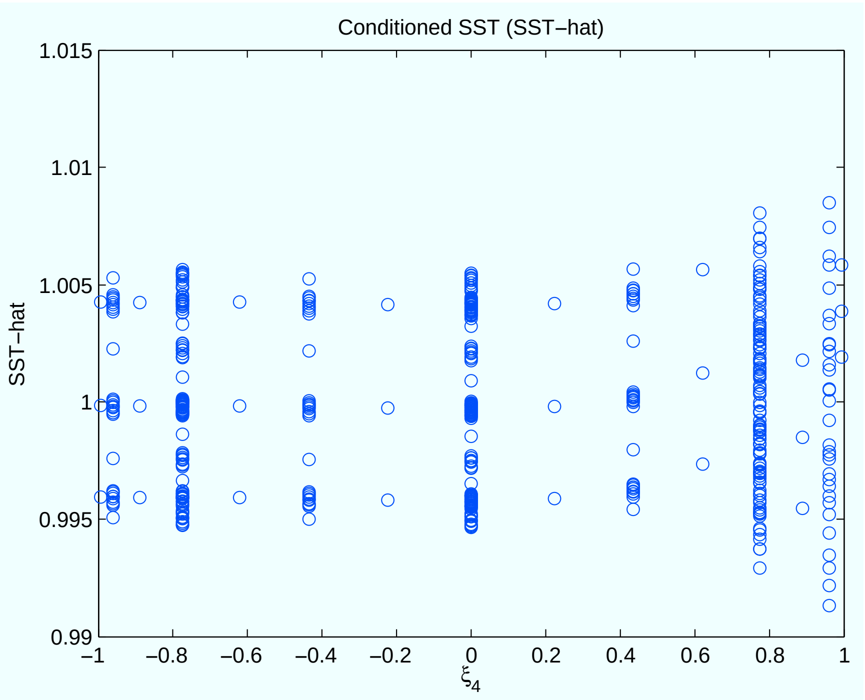


- errors measured with respect to independent LHS with 256 realizations
- results consistent with those obtained by measuring errors using on the realization grid

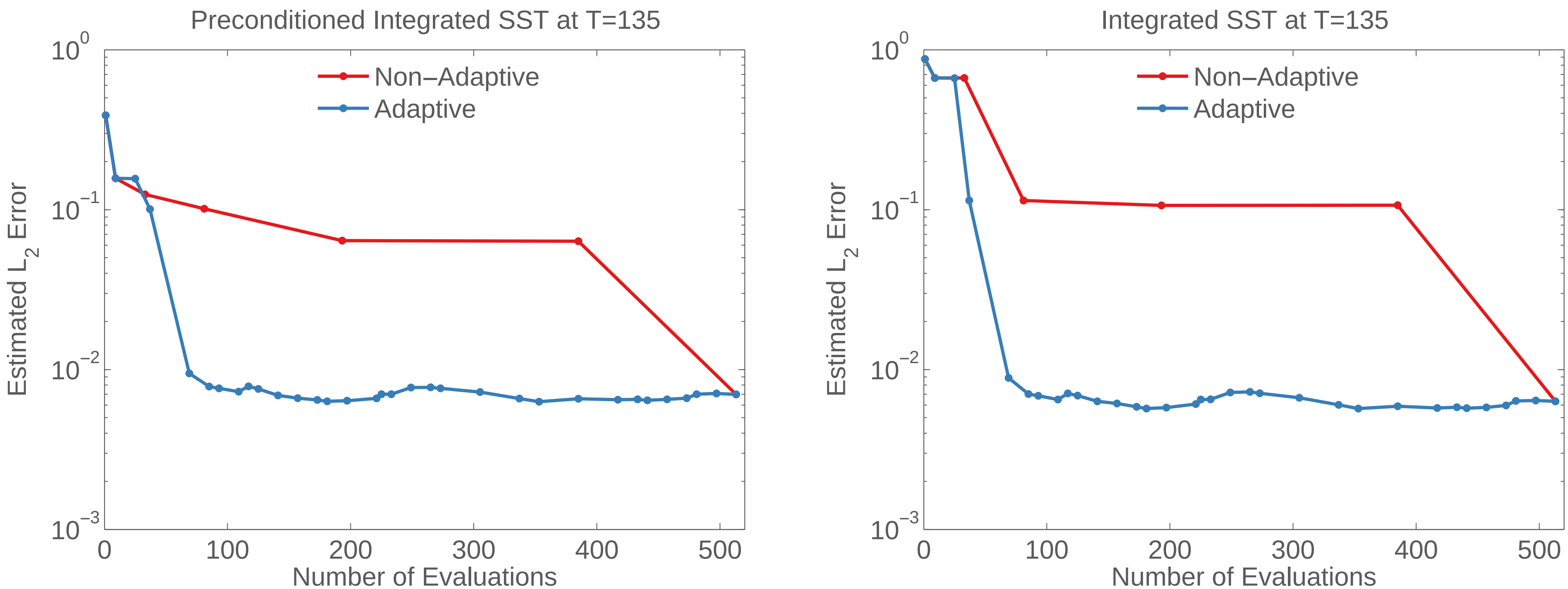
Preconditioned solution

- Since the fourth component is dominant, consider factorized construction: $\hat{f}(\xi) = f(\xi)/g(\xi_4)$



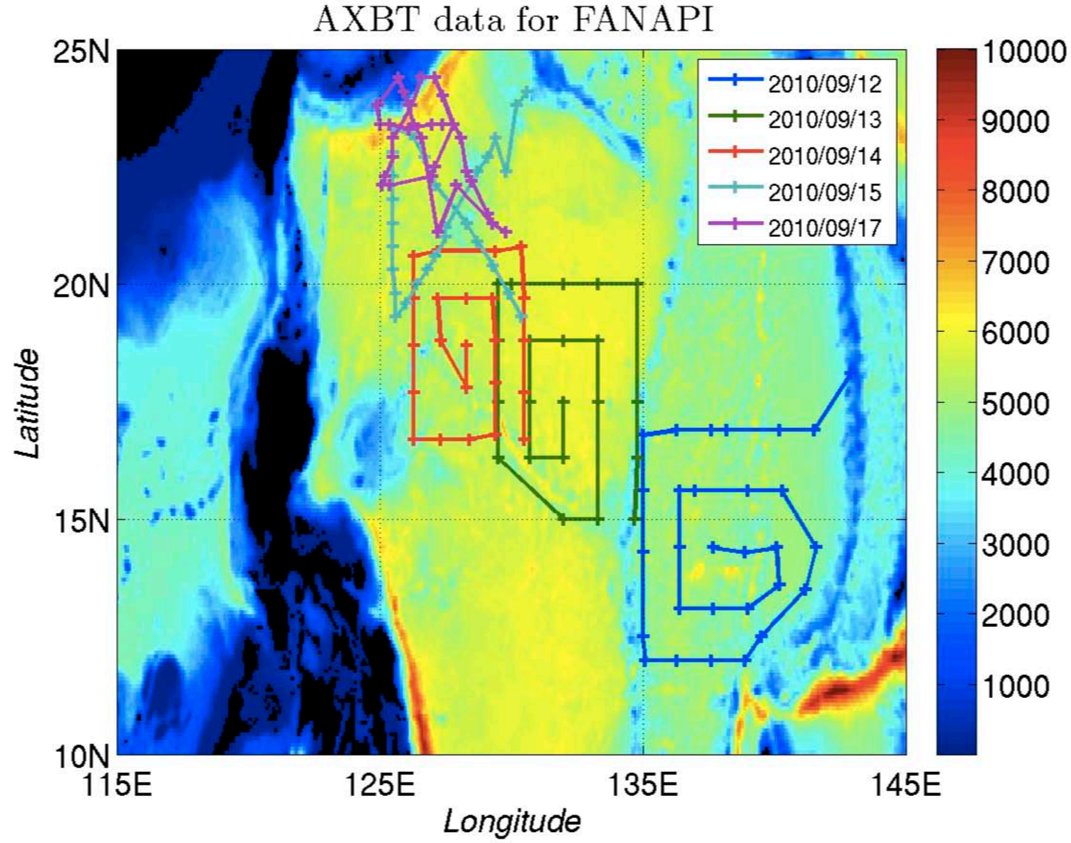


A priori test of preconditioned solution



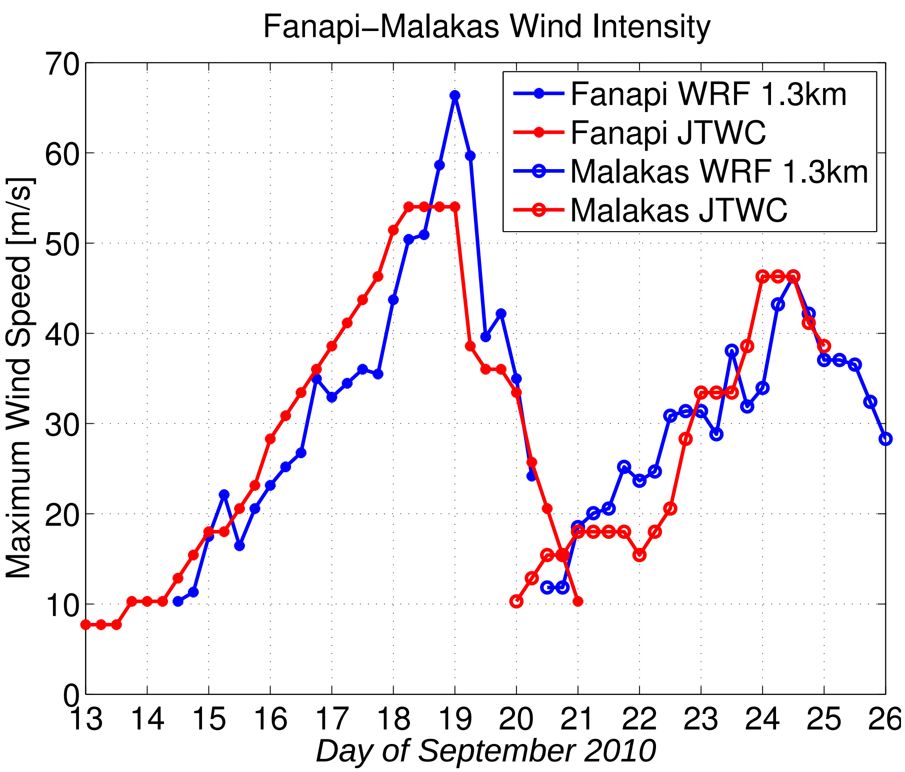
- preconditioned solution shows smaller initial errors and faster initial decau
- but saturates at approximately the same number of realizations

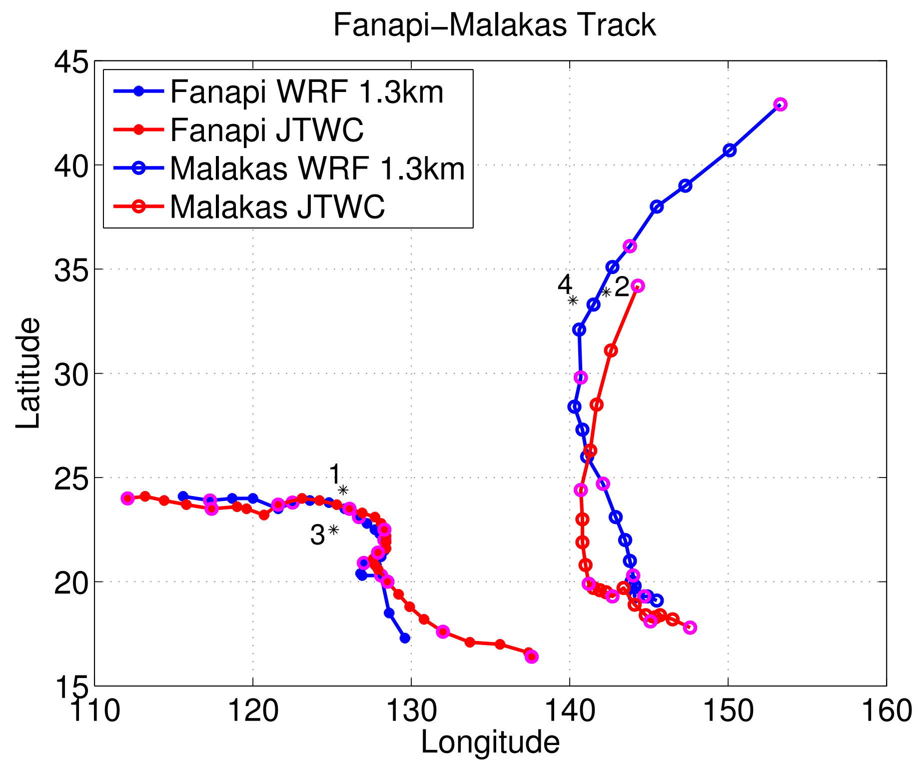
Driver: ITOP observations



- Availability of multiple AXBT probe measurements, particularly during Fanapi and Malakas

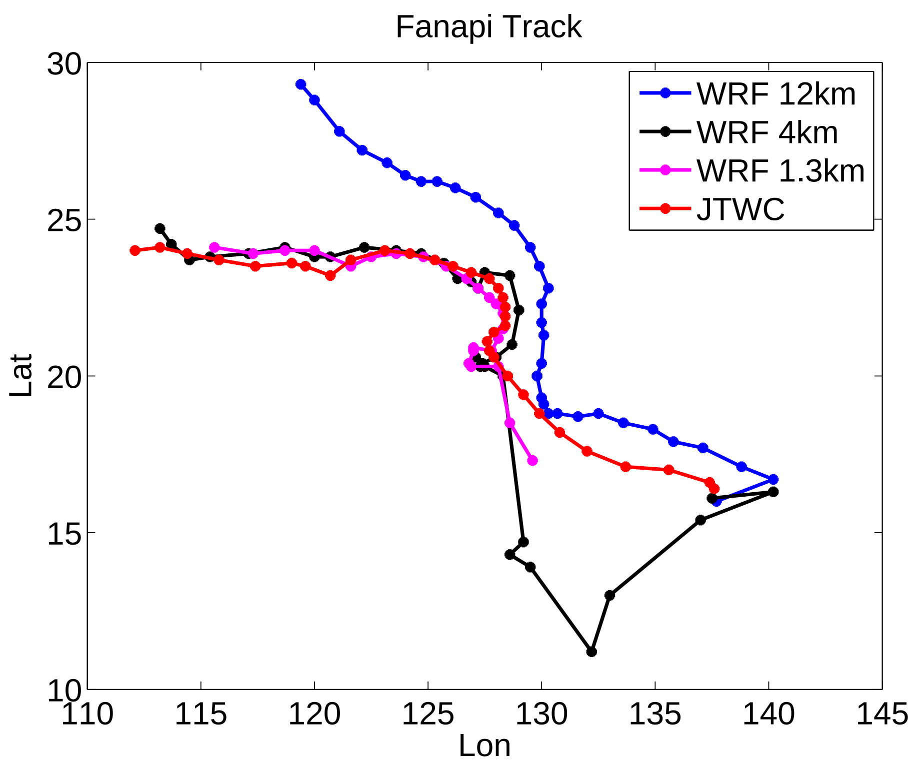
Wind forcing

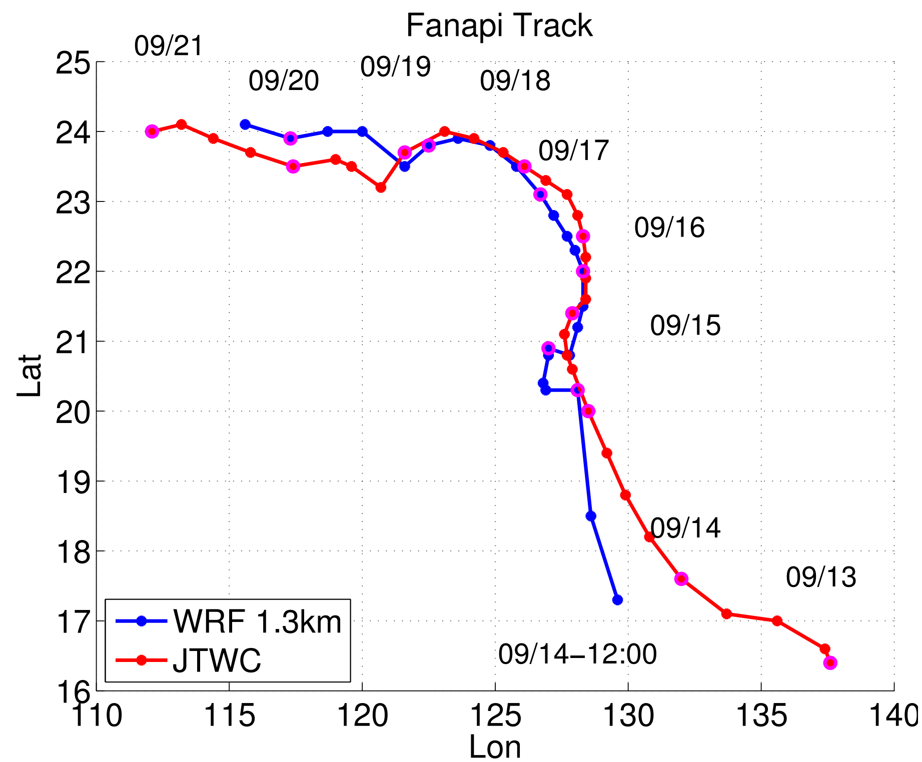




- ➢ detailed assimilation effort is needed to ensure wind forcing is adequate

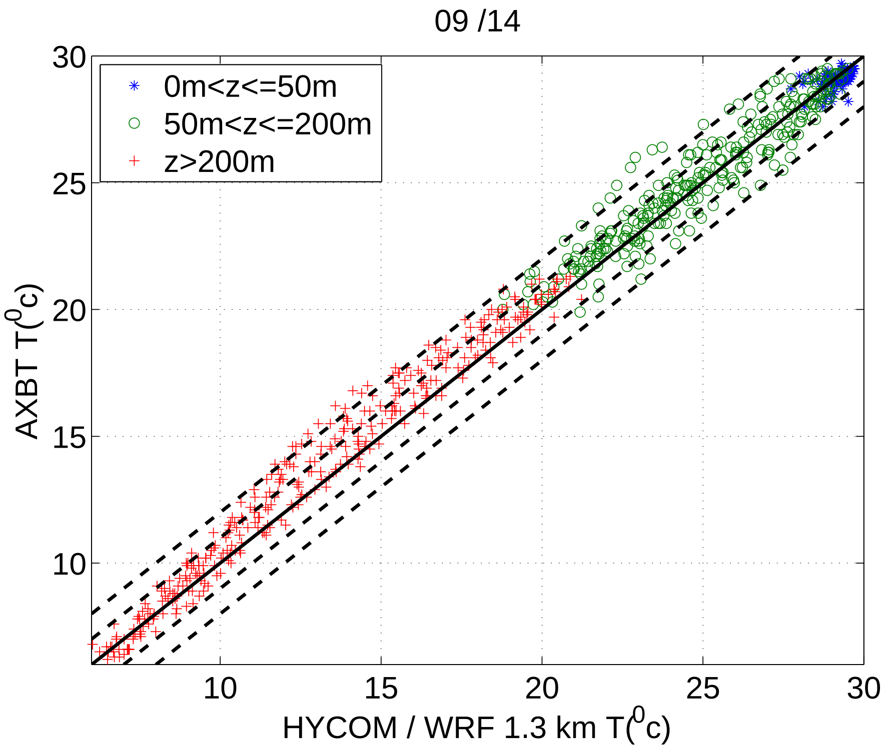
Impact of wind resolution

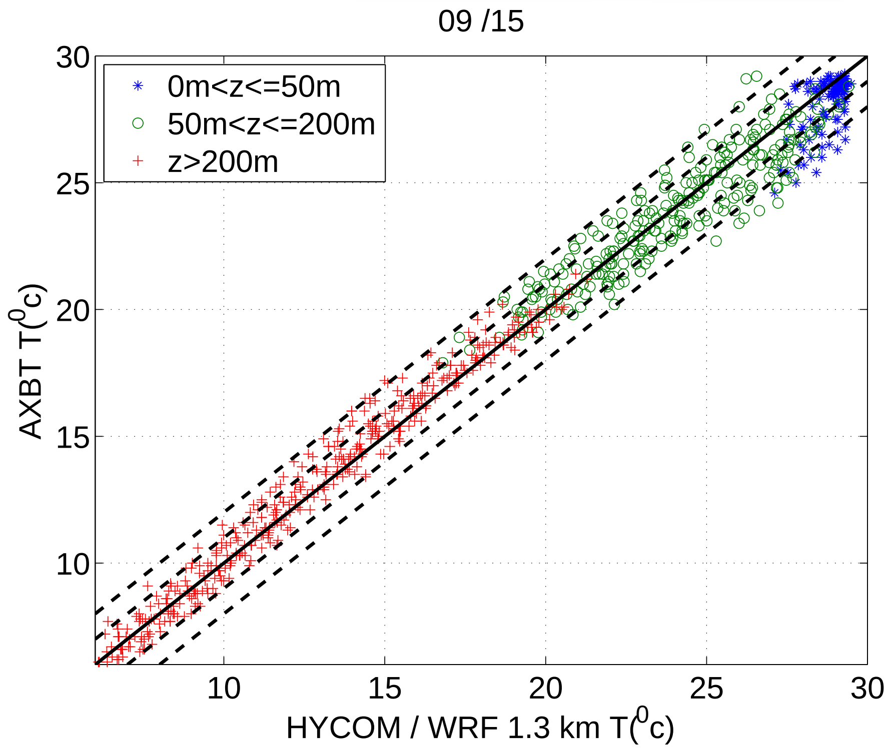




- ➢ systematic reduction of input discretization errors is essential prior to start of UQ analysis

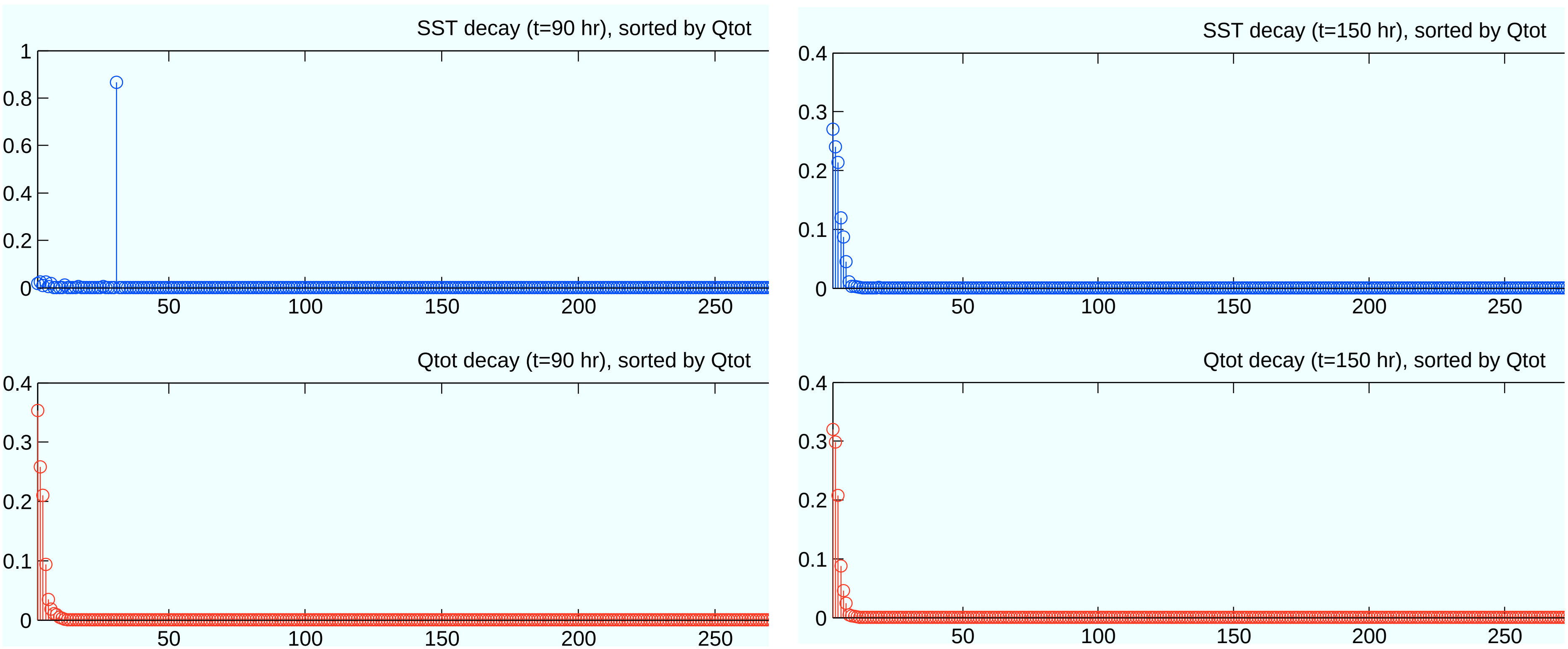
Preliminary results





- AXBT measurements are generally consistent with OGCM predictions
- Setup appear promising for:
  - refining priors
  - assisting in the design of field observations

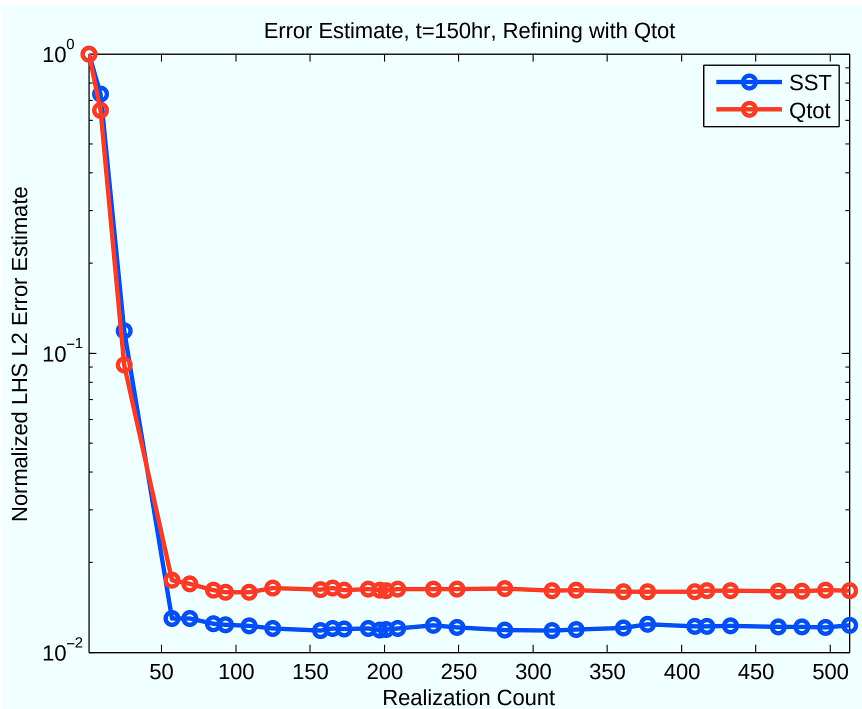
Multiple QoIs

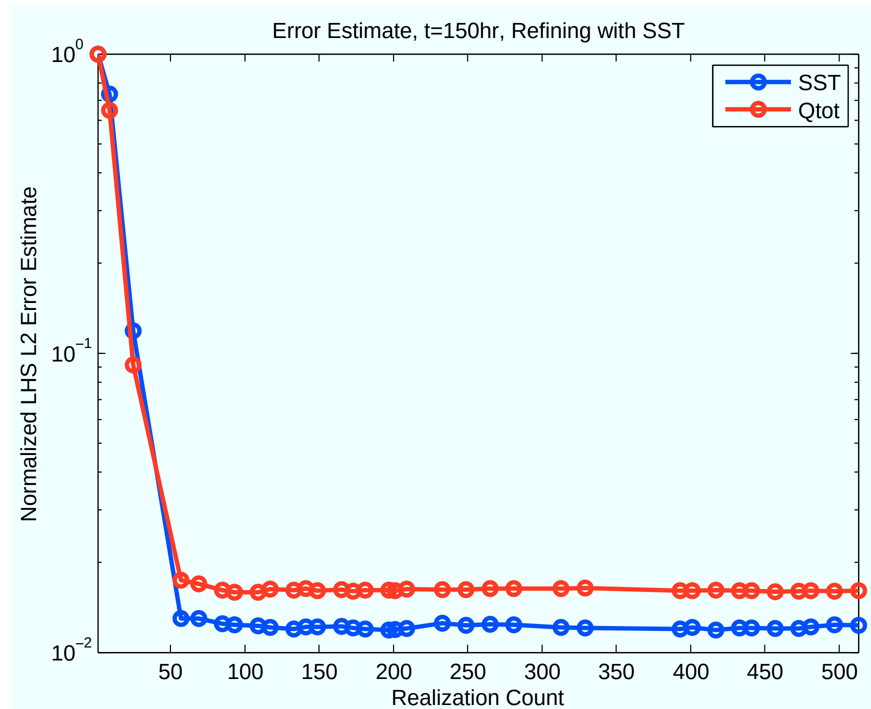


- careful analysis is required in cases involving multiple QoI's
- a priori test of composite indicators is ongoing

Composite error indicators

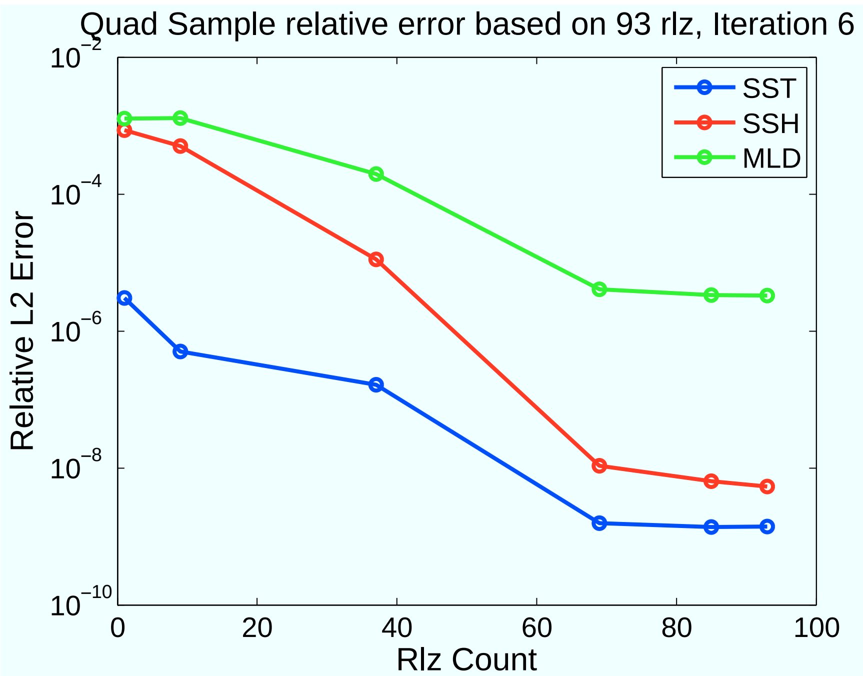
- define composite indicator based on a combination of individual error estimates

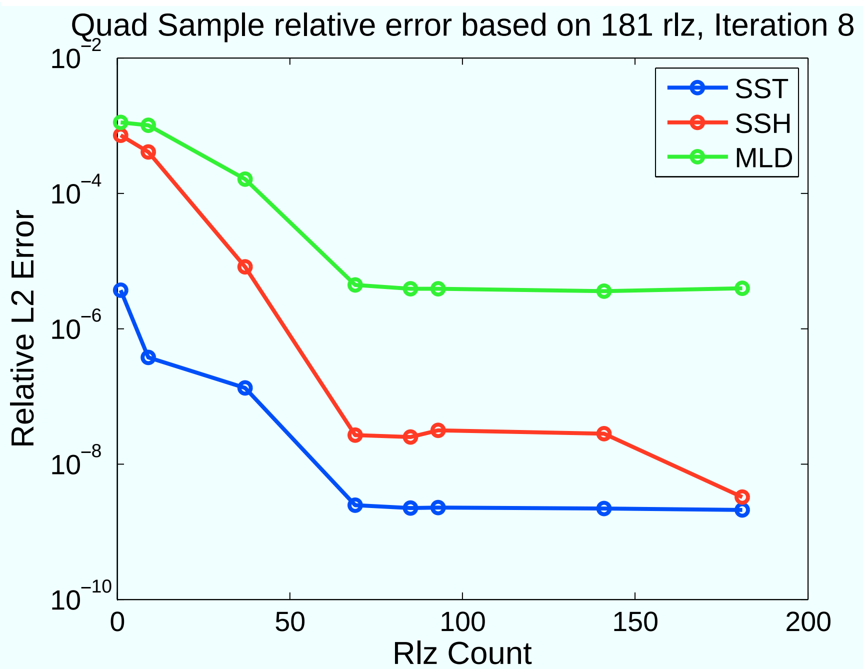




Multiple QoIs

- A possible approach is based on inclusion of union of "critical" indices for individual QoIs





  - errors curves exhibit similar trends and reach plateau at neighboring refinement stages

Summary

- Adaptive, sparse pseudospectral algorithms appear to be highly promising for efficient sampling of ocean circulation
- Ability to manage, treat, and exploit massive data is key
- Careful setup of forward problem, including suitable reduction of input uncertainties, is also essential