Surrogate modeling method for Bayesian uncertainty quantification with applications to Large Eddy Simulation turbulence models

Hoang Tran [1] Clayton Webster [1] Guannan Zhang [1]

[1]Computer Science and Mathematics Division
Oak Ridge National Laboratory, USA

SIAM UQ 2014 at Savannah, GA
April 2, 2014

Funding support: **DOD** (AFOSR), **DOE** (ASCR).

Overview

There are uncertainties in turbulence modeling coming from:

- user-chosen, application specific parameters
- model bias

Bayesian analysis of turbulence models is extremely expensive

- large number of forward executions
- each execution is costly

Overview

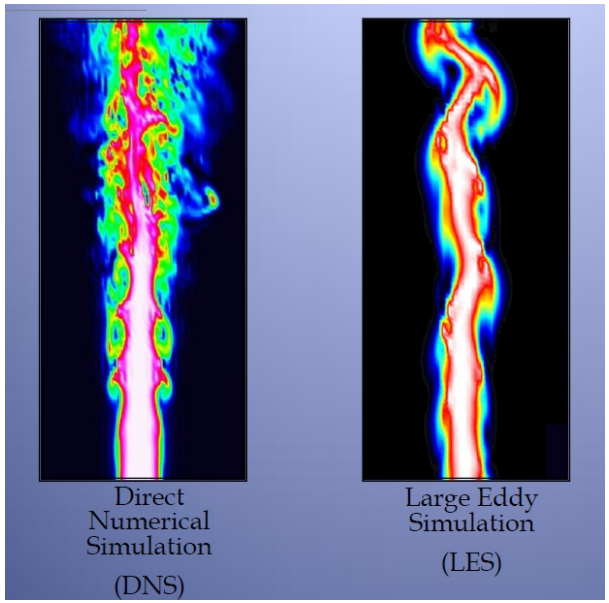
A brief history:

- Cheung, Oliver, Prudencio, Prudhomme, Moser (2011), Oliver, Moser (2011): Bayesian uncertainty analysis for RANS models.
- Ma and Zabaras (2009), Zeng, Shi, Zhang, Wu (2012), Zhang, Lu, Ye, Gunzburger, Webster (2013): surrogate modeling for geophysical and groundwater models.

Our objective: applying surrogate-based MCMC approach to Bayesian inference of LES models to reduce the cost of exploring the posterior distribution. We justify the accuracy and efficiency of surrogate modeling method in this talk.

Turbulence

Courtesy of Katsushika Hokusai



Courtesy of Pittsburgh Supercomputing Center

Smagorinsky model

Incompressible Navier-Stokes equation

$$u_t + \nabla \cdot (u \otimes u) - \nu \Delta u + \nabla p = f,$$
$$\nabla \cdot u = 0, \;\; u(x,0) = u_0(x).$$

Space Filtered Navier-Stokes equation

$$\overline{u}_t + \nabla \cdot (\overline{u} \otimes \overline{u}) - \nu \Delta \overline{u} + \nabla \overline{p} + \nabla \cdot \underbrace{(\overline{u \otimes u} - \overline{u} \otimes \overline{u})}_{\mathbf{R}(u,u)} = \overline{f}.$$

Smagorinsky model (von Neumann, Richtmyer (1950), Smagorinsky (1963)):

$$\nabla \cdot \mathbf{R}(u,u) \simeq -\nabla \cdot \left(2 \left(C_S \delta \left(1 - e^{-y^+/A^+} \right)^p \right)^2 |\nabla^s \overline{u}| \nabla^s \overline{u} \right).$$

Bayesian inference

Bayesian inference problem:

$$\mathbf{d} = \mathbf{f}(\theta) + \varepsilon$$

- $\mathbf{d} = (d_1, \ldots, d_{N_d})$: vector of N_d measurement data,
- $\theta = (\theta_1, \ldots, \theta_{N_\theta})$: vector of N_θ model parameters,
- $\mathbf{f}(\theta)$: forward model,
- ε: measurement, parametric and structural errors.

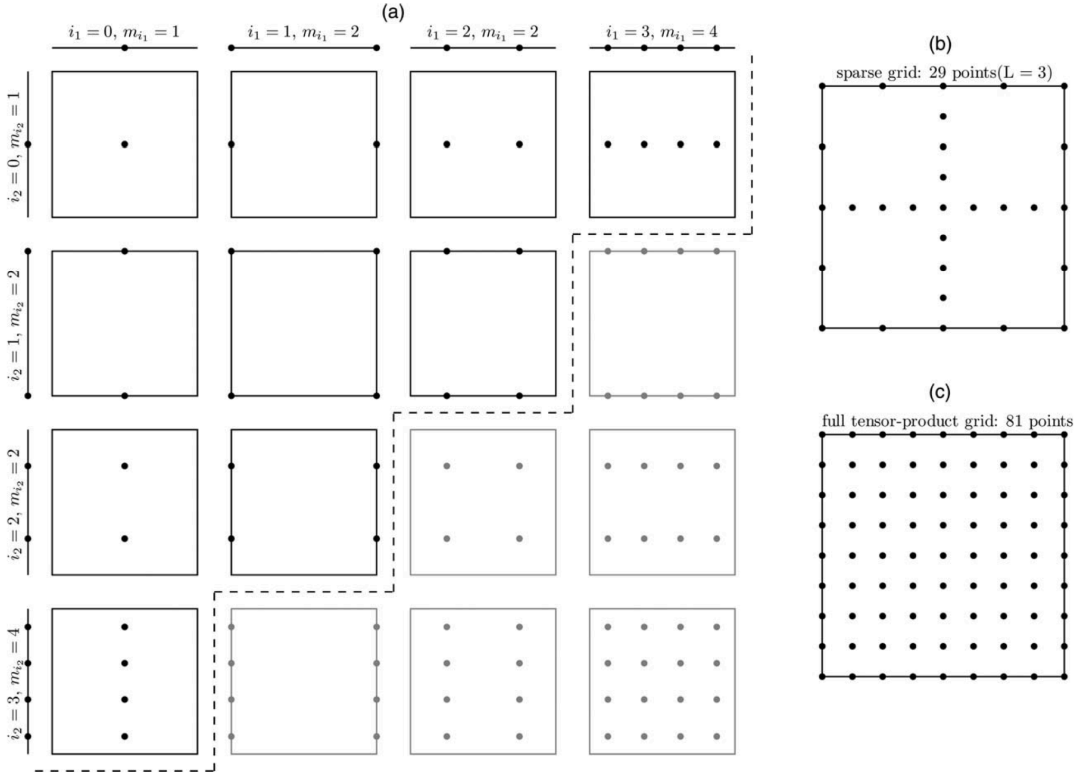
Bayes' theorem

$$p(\theta|\mathbf{d}) = \frac{L(\theta|\mathbf{d})p(\theta)}{\int L(\theta|\mathbf{d})p(\theta)d\theta}$$

Example: multivariate normal likelihood function

$$L(\theta|\mathbf{d}) = \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{f}(\theta))^\top \Sigma^{-1}(\mathbf{d} - \mathbf{f}(\theta))\right).$$

Isotropic sparse grid



Number of interpolation points is:

$$\begin{cases} m_0 = 1, \\ m_1 = 2, \\ m_i = 2^{i-1} \quad \text{if} \quad i \geq 2. \end{cases}$$

The grid abcissas is defined by:

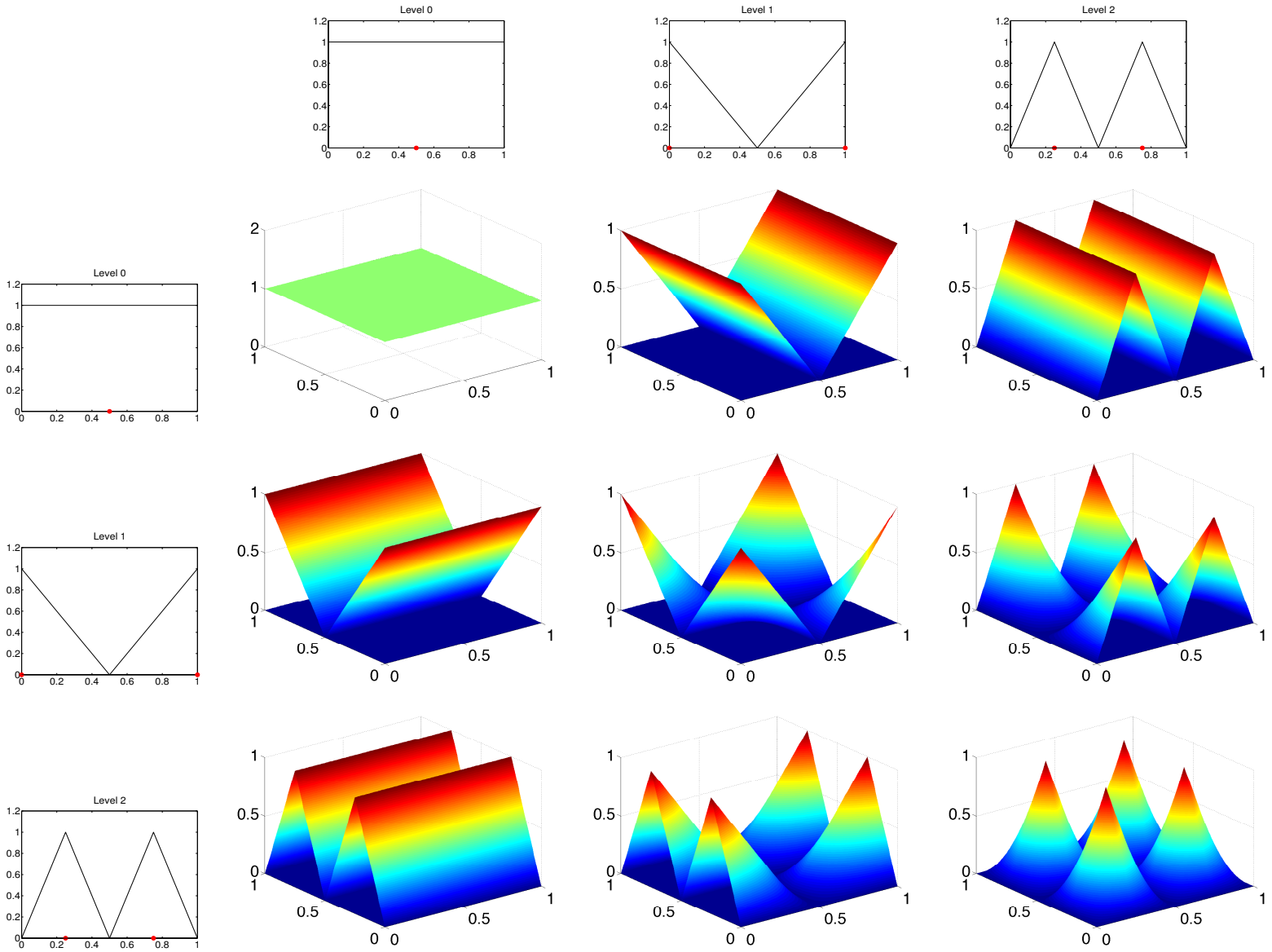
$$\theta_1^0 = 0.5 \qquad \text{for } i = 0,$$

$$\theta_1^1 = 0,\ \theta_2^1 = 1 \qquad \text{for } i = 1,$$

$$\theta_j^i = \frac{2j-1}{\sum_{k=0}^{i} m_k - 1} \qquad j = \overline{1, m_i},\ i \geq 2.$$

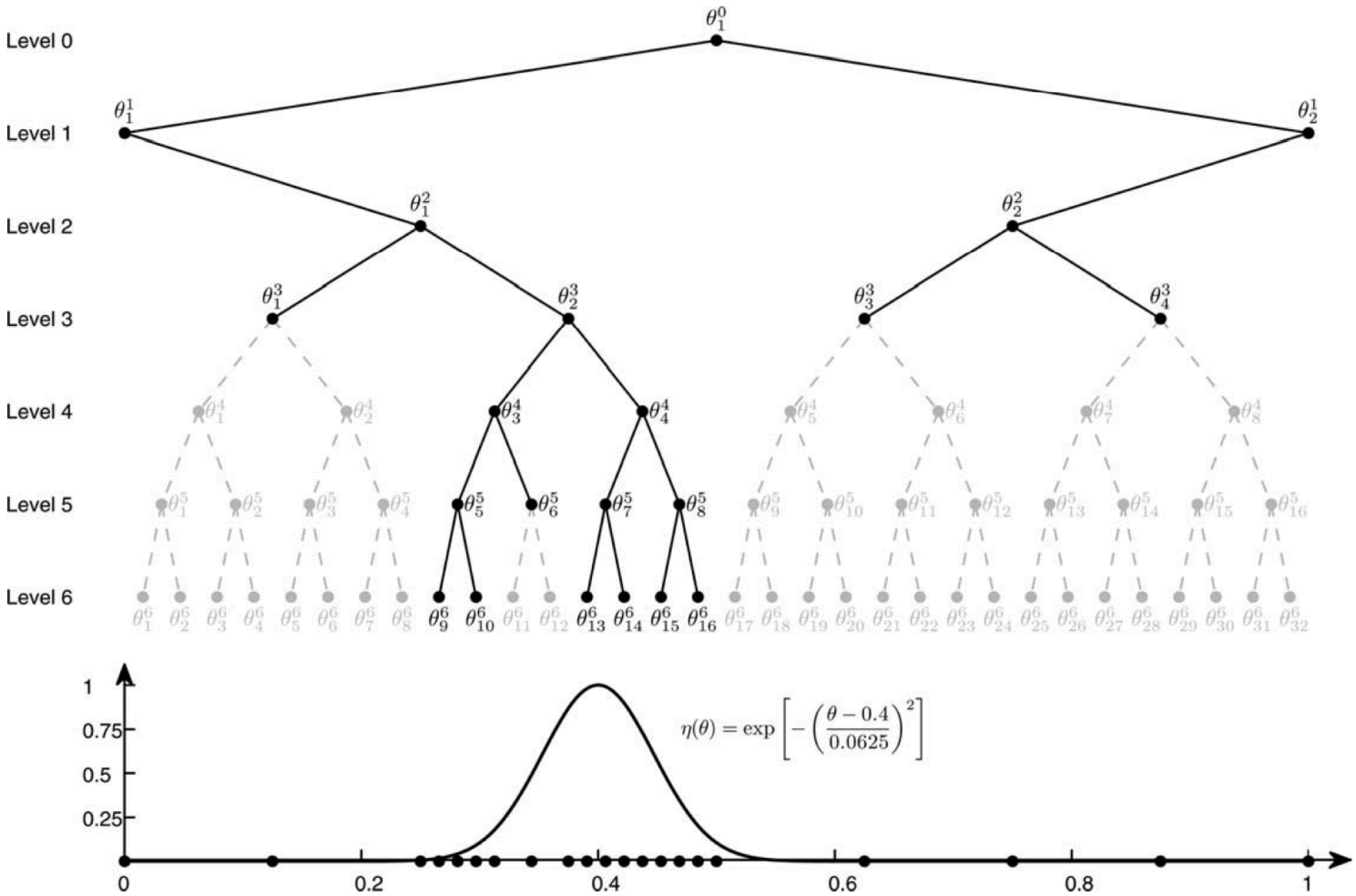
- Full tensor-product grid: $\lambda(\mathbf{i}) = \max_{n=1,\ldots,N_\theta} i_n$,
- Isotropic sparse grid: $\lambda(\mathbf{i}) = |\mathbf{i}| = i_1 + \cdots + i_{N_\theta}$.

Linear basis functions



Adaptive sparse grid

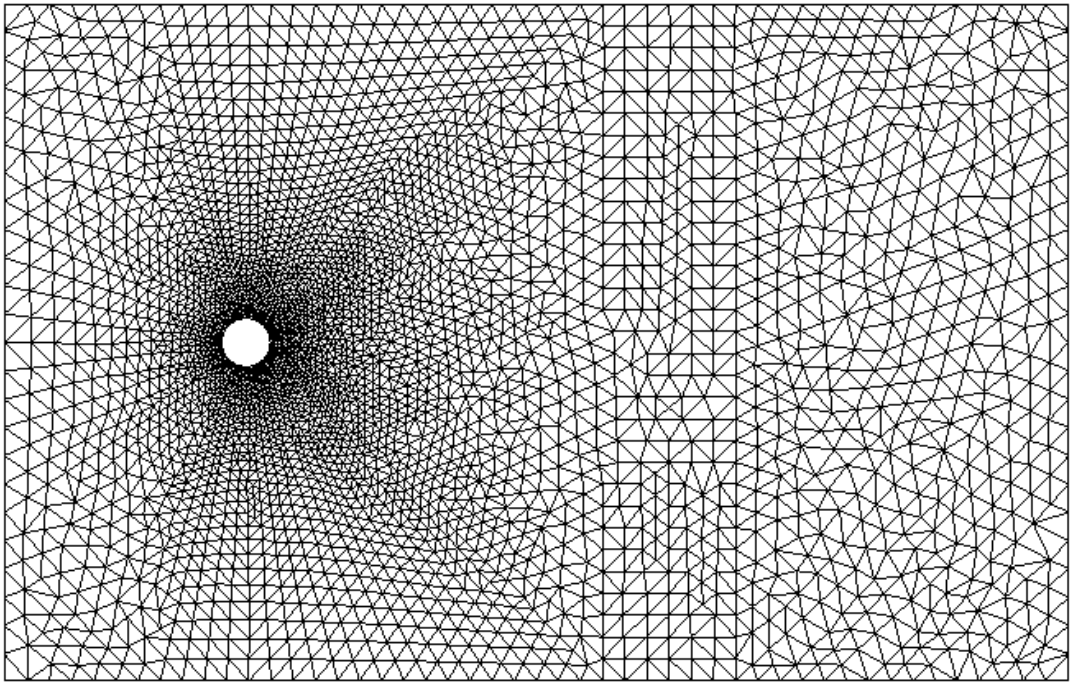
$$\mathcal{I}^{L,N_\theta}(\eta)(\theta) = \sum_{\lambda(\mathbf{i})\le L} \Delta\mathcal{I}^{\mathbf{i},N_\theta}(\eta)(\theta) = \sum_{\lambda(\mathbf{i})\le L}\sum_{\mathbf{j}\in B_{\mathbf{i}}} c^{\mathbf{i}}_{\mathbf{j}}\phi^{\mathbf{i}}_{\mathbf{j}}(\theta)$$



aSG-SC surrogate modeling

1. Build the surrogate model for $\mathbf{f}(\theta)$ using adaptive sparse-grid stochastic collocation (aSG-SC) method. The forward problem is solved only for the sparse parameter samples.
2. The surrogate model for $p(\theta|\mathbf{d})$ can be obtained from that for $\mathbf{f}(\theta)$ by simple and cheap calculations.
3. Evaluate $p(\theta|\mathbf{d})$ by MCMC sampling the surrogate model directly without forward execution.

Example: 2D flow around a circular cylinder



- $\Omega = (0, 2.2) \times (0, 1.4)$, $Re = 500$.
- $P2$-$P1$ finite element with polygonal boundary approximation.
- Crank-Nicolson time discretization with fixed point iteration.
- $\Delta t = 1/1000$, $T = 12$.
- Dirichlet inflow and outflow, no-slip top and bottom boundary.

Example: 2D flow around a circular cylinder

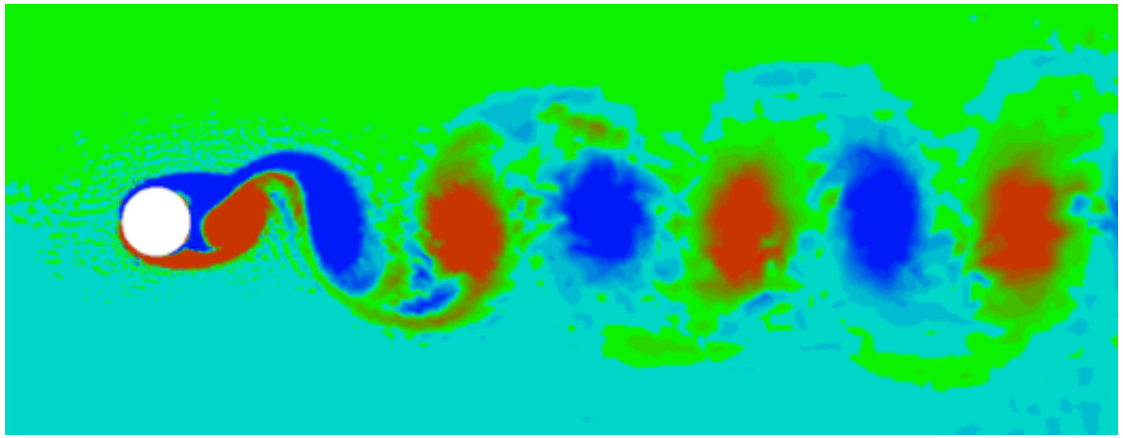


Figure : $C_S = 2$, $p = 0$, $\delta \simeq \pi/480$.

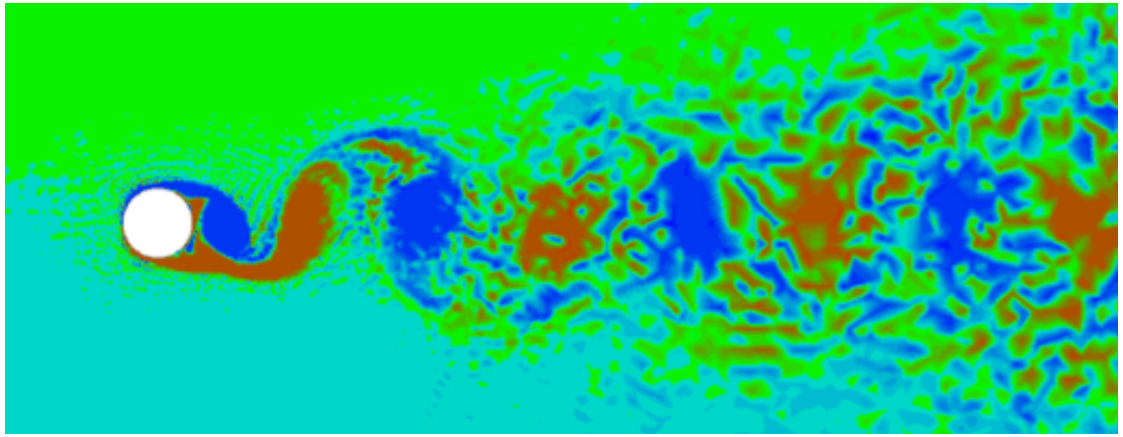
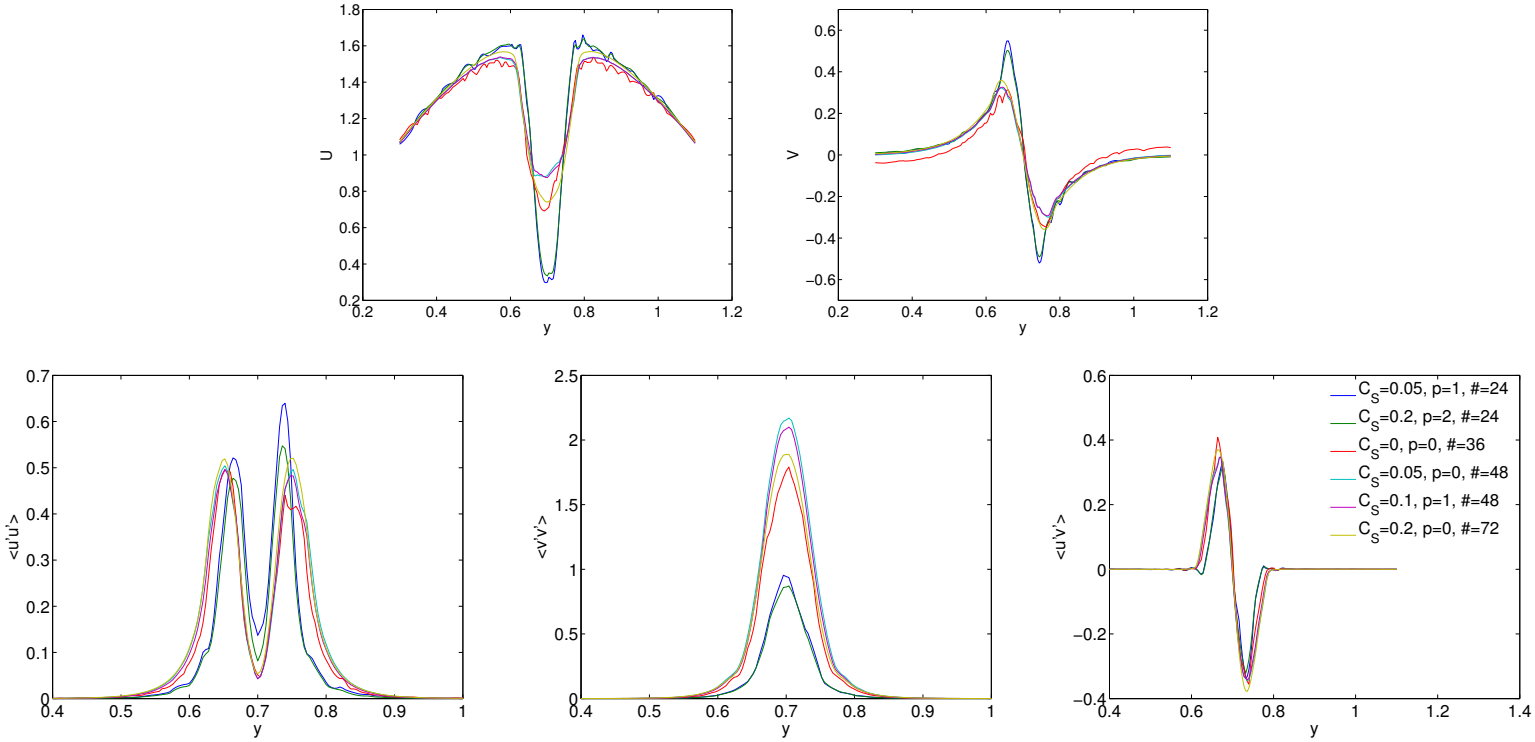


Figure : $C_S = 0.5$, $p = 0$, $\delta \simeq \pi/720$.

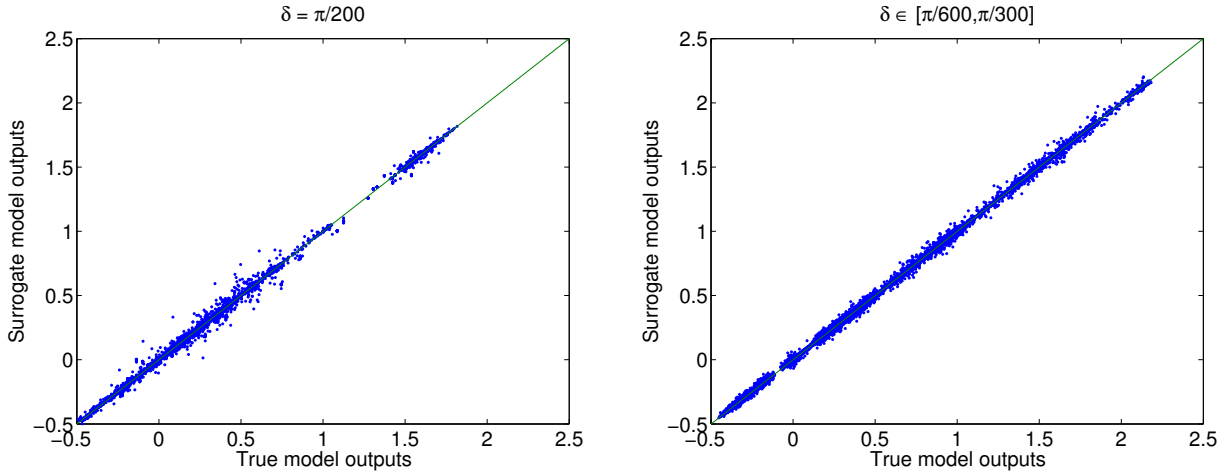
Example: 2D flow around a circular cylinder

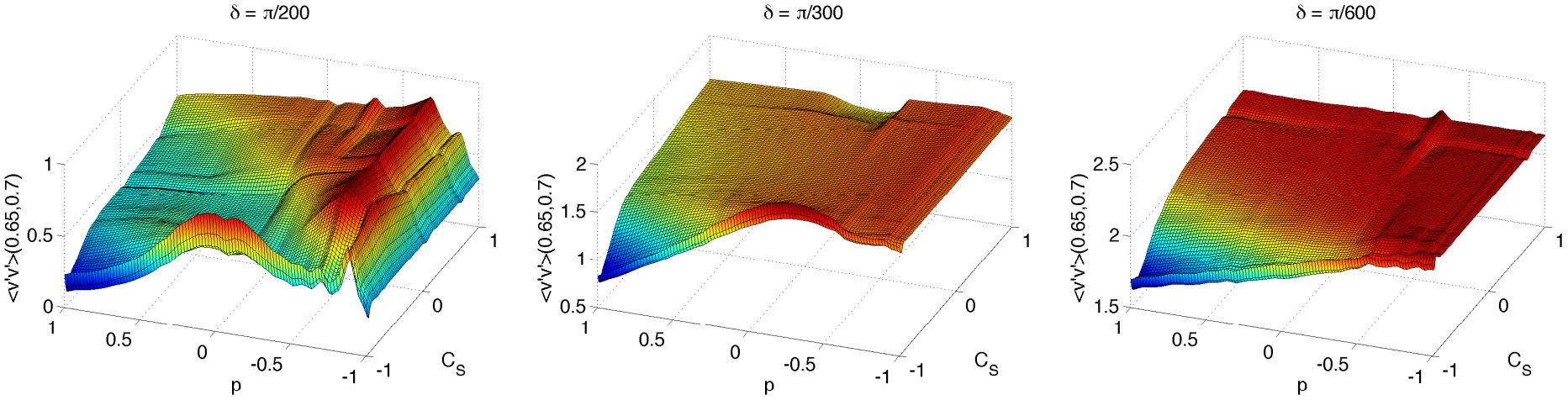
Calibration data: synthetic data of mean velocity and Reynold stresses at 11 points at $x/D = 1$ generated by solving Smagorinsky models.



Example: 2D flow around a circular cylinder

C_S	p	δ
$[0, 0.2]$	$[0, 2]$	$[\pi/600, \pi/200]$

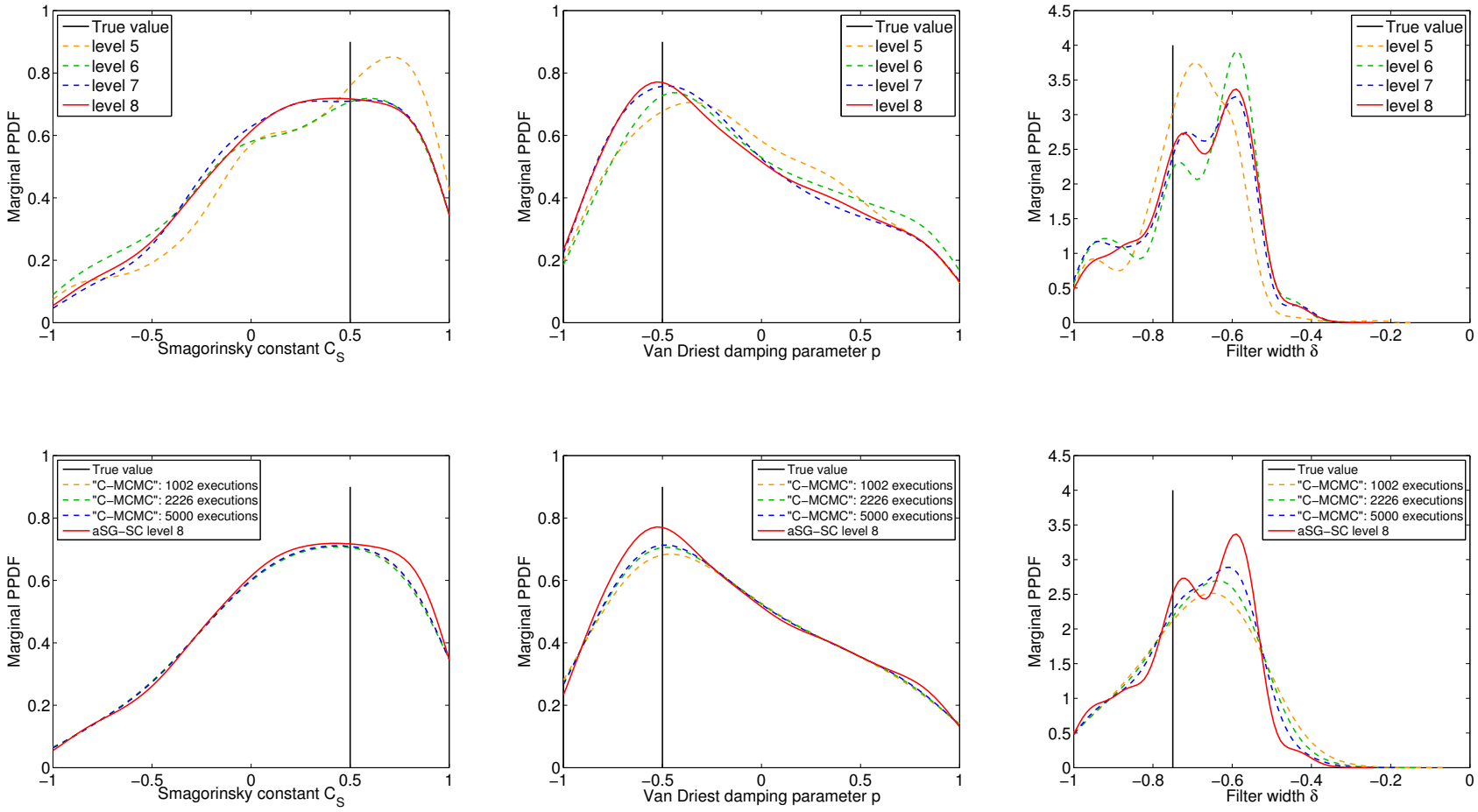




Example: 2D flow around a circular cylinder

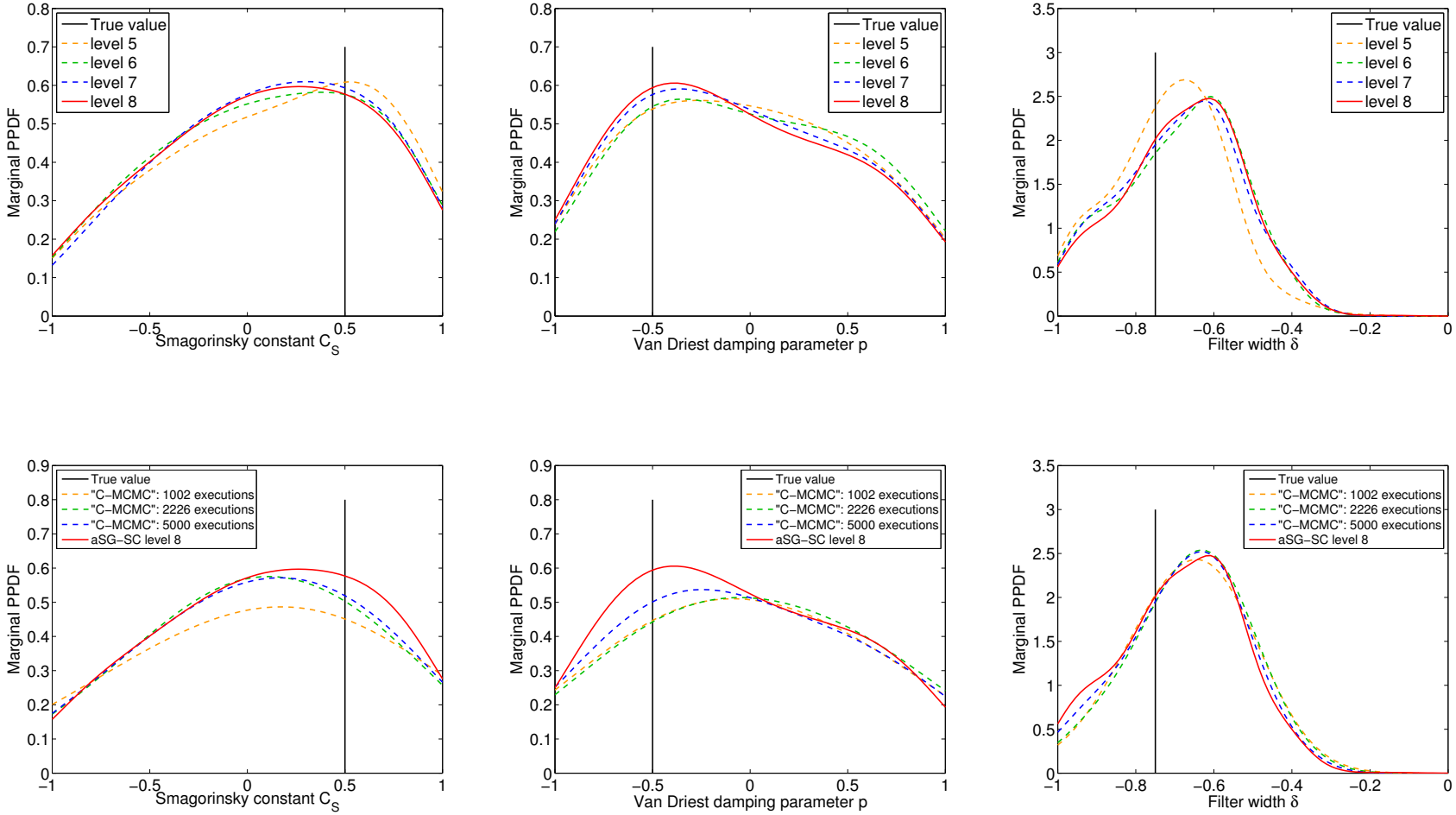
Likelihood function: $\exp(-\frac{1}{2}(\mathbf{d}-\mathbf{f})^\top \Sigma^{-1}(\mathbf{d}-\mathbf{f}))$ where $\Sigma = (0.1)^2$.

True value: $C_S = 0.5$, $p = -0.5$, $\delta = -0.75$.



Example: 2D flow around a circular cylinder

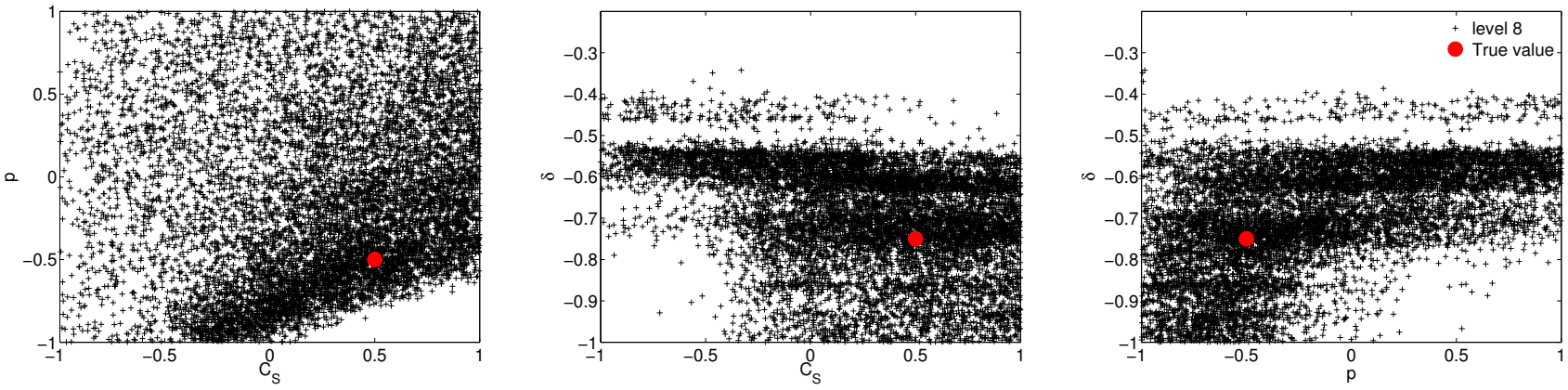
$$\text{Likelihood function: } \exp\left(-500 \cdot \frac{\sum_{i=1}^{N_d}\left((d_i - f_i) - (\overline{\mathbf{d}} - \overline{\mathbf{f}})\right)^2}{\sum_{i=1}^{N_d}\left(d_i - \overline{\mathbf{d}}\right)^2}\right).$$



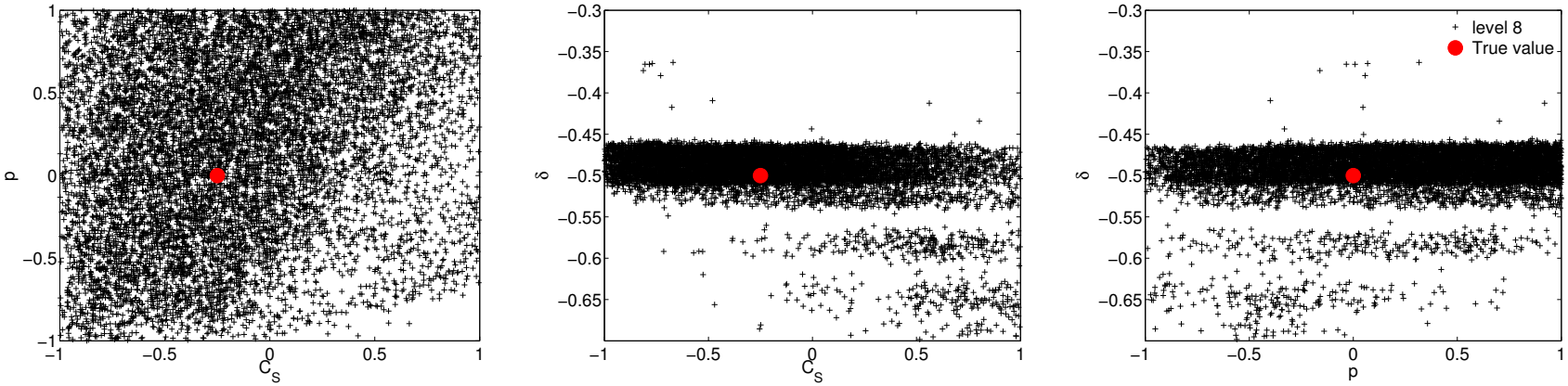
Example: 2D flow around a circular cylinder

Likelihood function: $\exp(-\frac{1}{2}(\mathbf{d}-\mathbf{f})^\top \Sigma^{-1}(\mathbf{d}-\mathbf{f}))$ where $\Sigma = (0.1)^2$.

True value: $C_S = 0.5$, $p = -0.5$, $\delta = -0.75$.



True value: $C_S = -0.25$, $p = 0$, $\delta = -0.5$.



Conclusions

In this work, we introduce a surrogate modeling approach to Bayesian uncertainty analysis of Large Eddy Simulation. For the numerical example of 2D flow around a cylinder, this approach has been shown to:

- significantly reduce the number of costly LES executions,
- be reliable with *reasonable* search regions of filter width,
- be reflexible for different observations and likelihood functions.

It is worth to study surrogate-based MCMC sampling for engineering turbulent flow models.

Thank you for your attention!!!
Questions?