# Treating Modeling Uncertainty in Multiscale Methods.

Roger Ghanem

Viterbi School of Engnineering
University of Southern California
Los Angeles, CA,

WORKSHOP ON UNCERTAINTY QUANTIFICATION FOR MULTISCALE STOCHASTIC SYSTEMS AND APPLICATIONS

IPAM, LOS ANGELES, JANUARY 19-22, 2016.

## Acknowledgment

- Suppport from NSF, US Department of Energy under SciDAC, Vehicle Technology and Ultra Deepsea Drilling Programs.
- Jianyu Li (Tianjin University)
- Loujaine Mehrez and Charanraj Thimmmisetty (USC)
- Venkat Aitharaju and William Rodgers (GM Motors Inc.)
- Han Yu and Praveen Pasupuleti (ESI Group)
- Colin McAuliffe and Robert Crouch (Altair Inc.)
- Jacob Fish (Columbia University and Altair Inc.)

# Outline

# Error Budget

$$U = \hat{U}|_{h,d,p,m} + \underbrace{\epsilon_h|_{d,p,m} + \epsilon_p|_{d,m} + \epsilon_d|_m}_{\text{Limits on Predictability: Must be quantified}} + \epsilon_m$$

$\epsilon_h|_{d,p,m}$ : can be reduced through better numerics.
$\epsilon_p|_{d,m}$ : can be reduced through better statistics.
$\epsilon_d|_m$ : can be reduced through better data.
$\epsilon_m$ : can be reduced through better models.

# Observations from elementary statistics

CLT: average out the noise

- $X \sim \chi^2_d$

$$X = \sum_{i=1}^{d} \xi_i^2, \qquad \xi_i \sim N(0,1)$$

- $T \sim t_d$

$$T = \frac{1}{d} \sum_{i=1}^{d} \xi_i, \qquad \xi_i \sim N(0,1)$$

Reverse-engineer CLT:
Features matter: away from mean-field theories

Given coarse observable, construct a functional model from the finer scales:

$$X = f(\xi_1, \cdots, \xi_d) = f(\boldsymbol{\xi})$$

# Rethink Data, Models and Risks

**Models: Degrees of freedom back in fashion.**

Reverse perspectives on Central Limit Theorems:

$$X \longleftrightarrow \{\xi\} \qquad\qquad T \longleftrightarrow \{\xi\}$$

Given observations and physics, find the statistical DOF.

**Polynomial Chaos**

First: decide/identify the root cause of uncertainty.
Second: relate everything to this root cause.

**Risk**

Is dynamic, function of data, and mathematical and statistical models.

# Rethink Data, Models and Risks

## Models: Degrees of freedom back in fashion.

Reverse perspectives on Central Limit Theorems:

$$X \longleftrightarrow \{\xi\} \qquad\qquad T \longleftrightarrow \{\xi\}$$

Given observations and physics, find the statistical DOF.

## Polynomial Chaos

First: decide/identify the root cause of uncertainty.
Second: relate everything to this root cause.

## Risk

Is dynamic, function of data, and mathematical and statistical models.

# Rethink Data, Models and Risks

## Models: Degrees of freedom back in fashion.

Reverse perspectives on Central Limit Theorems:

$$X \longleftrightarrow \{\xi\} \qquad\qquad T \longleftrightarrow \{\xi\}$$

Given observations and physics, find the statistical DOF.

## Polynomial Chaos

First: decide/identify the root cause of uncertainty.
Second: relate everything to this root cause.

## Risk

Is dynamic, function of data, and mathematical and statistical models.

# Probabilistic Models

## Construct Probabilistic Models

- Assume some model for the measure of $K$, parameterized by $\omega$:

$$f_K(k;\omega)$$

.

- Invoke enough constraints from data to evaluate $\omega$.
  - Kernel density estimation
  - Maximum Likelihood
  - Moment matching
  - Maximum Entropy
  - Bayes rule

## Common Approach; constraints from data

Ignores additional knowledge (from physics ?) about $K$.

# Probabilistic Models

## Construct Probabilistic Models

- Assume some model for the measure of $K$, parameterized by $\omega$:

$$f_K(k;\omega)$$

.

- Invoke enough constraints from data to evaluate $\omega$.
  - Kernel density estimation
  - Maximum Likelihood
  - Moment matching
  - Maximum Entropy
  - Bayes rule

## Common Approach; constraints from data

Ignores additional knowledge (from physics ?) about $K$.

# Probabilistic Models

### I/O models with random dynamics

$K$ is a function of fine scale behavior, which could be characterized, independently, through experiments.

### Postulate dependence of $K$ on fine scale random parameters:

$$K = f(\xi_1, \cdots, \xi_d)$$

### We embed ourselves in a $d$-dimensional space.

- The functional form of $f$ must be estimated from observations of $K$.
- Joint probability measure of $\xi$ must be estimated.
- We have accomplished a hierarchical decomposition.

# Probabilistic Models

**I/O models with random dynamics**

$K$ is a function of fine scale behavior, which could be characterized, independently, through experiments.

**Postulate dependence of $K$ on fine scale random parameters:**

$$K = f(\xi_1, \cdots, \xi_d)$$

**We embed ourselves in a $d$-dimensional space.**

- The functional form of $f$ must be estimated from observations of $K$.
- Joint probability measure of $\xi$ must be estimated.
- We have accomplished a hierarchical decomposition.

# Probabilistic Models

**I/O models with random dynamics**

$K$ is a function of fine scale behavior, which could be characterized, independently, through experiments.

**Postulate dependence of $K$ on fine scale random parameters:**

$$K = f(\xi_1, \cdots, \xi_d)$$

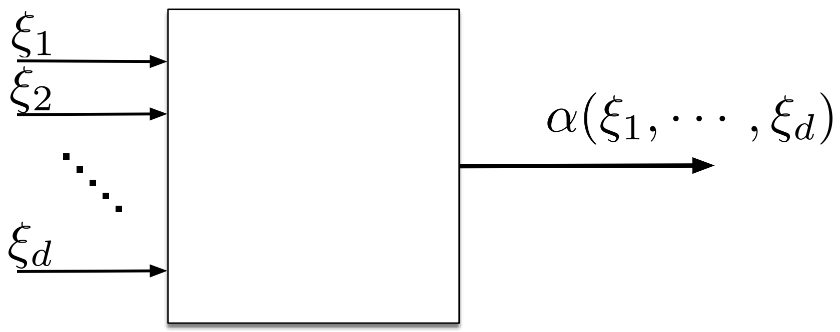
**We embed ourselves in a $d$-dimensional space.**

- The functional form of $f$ must be estimated from observations of $K$.
- Joint probability measure of $\boldsymbol{\xi}$ must be estimated.
- We have accomplished a hierarchical decomposition.

# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ :

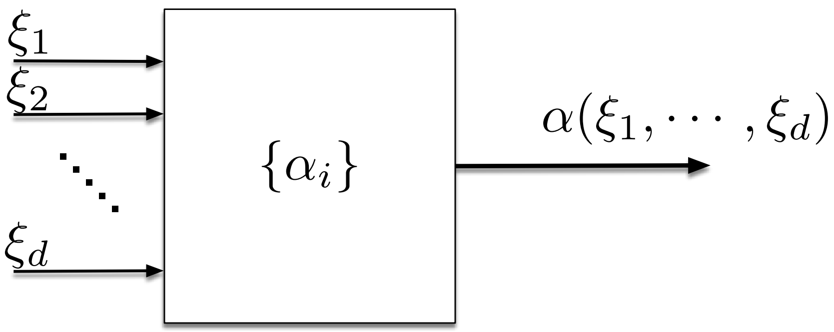
$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x)\Psi_i(\boldsymbol{\xi}(\omega))$$



# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ :

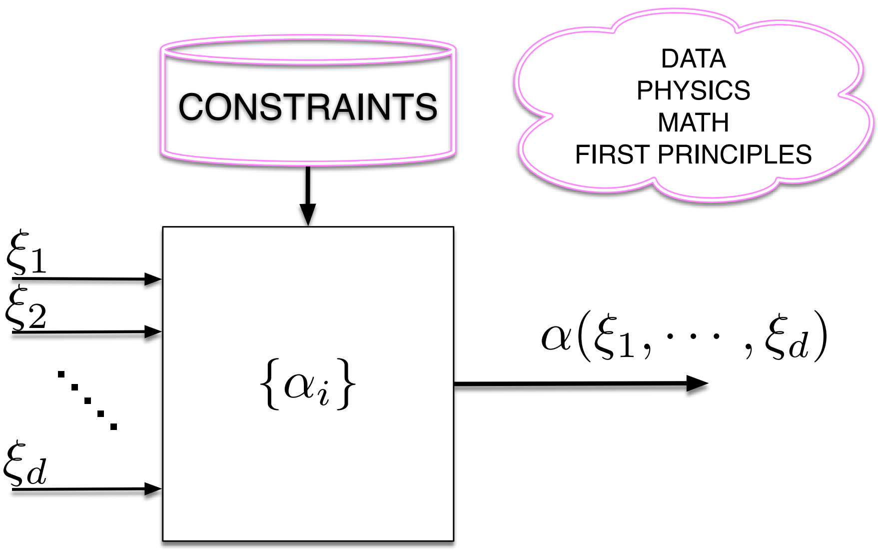
$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x)\Psi_i(\boldsymbol{\xi}(\omega))$$



# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ :

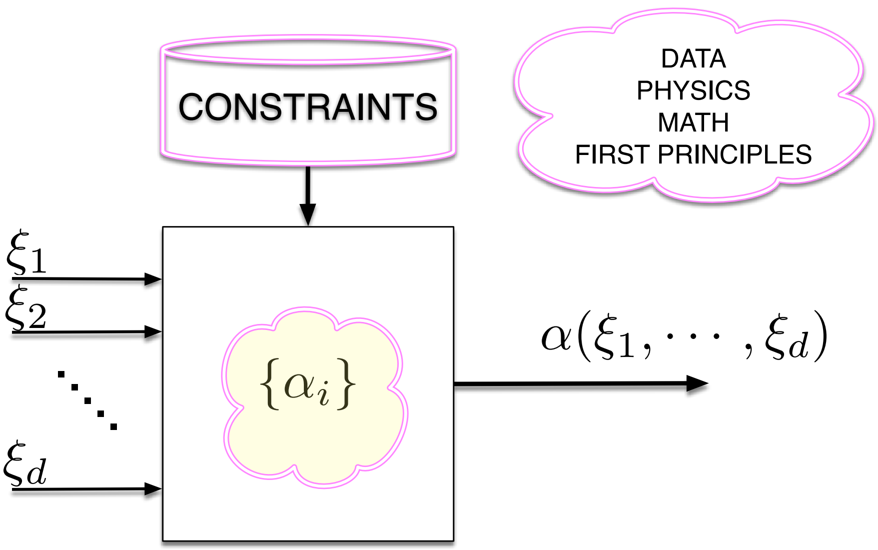
$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x)\Psi_i(\boldsymbol{\xi}(\omega))$$



# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ :

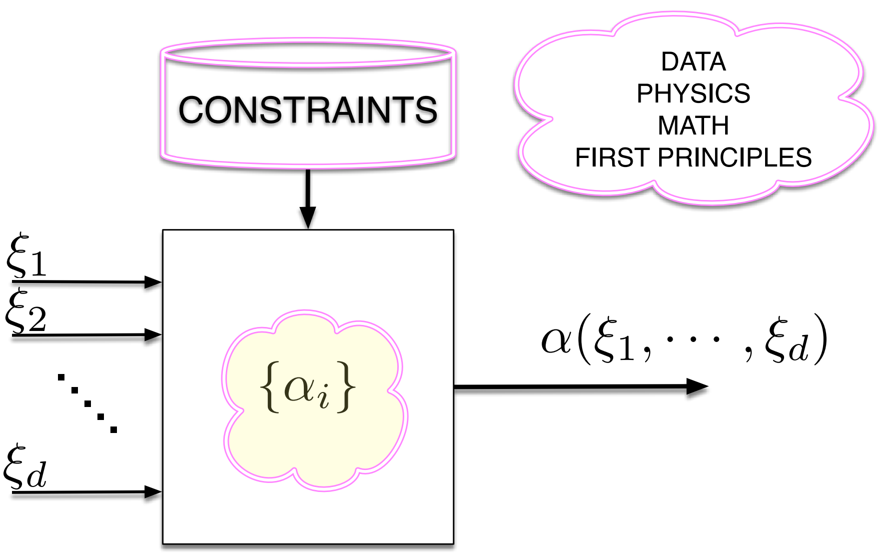
$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x)\Psi_i(\boldsymbol{\xi}(\omega))$$



# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$:

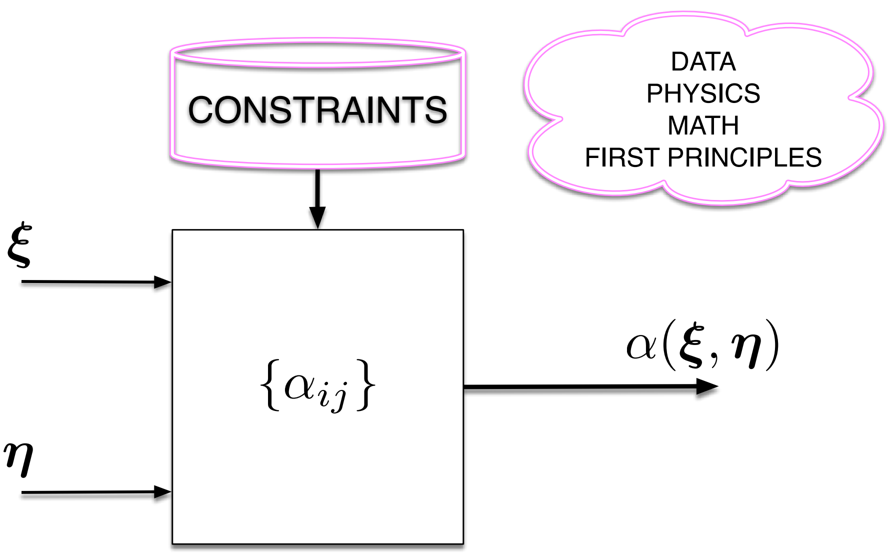
$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x,\boldsymbol{\eta})\Psi_i(\boldsymbol{\xi}(\omega))$$



# Polynomial Chaos

Blame uncertainty in $\alpha$ on $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ :

$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_{ij}(x)\Psi_i(\boldsymbol{\xi}(\omega))\Psi_j(\boldsymbol{\eta}(\omega))$$



### Versatile Representations:

By changing

- dimension
- order
- coefficient values

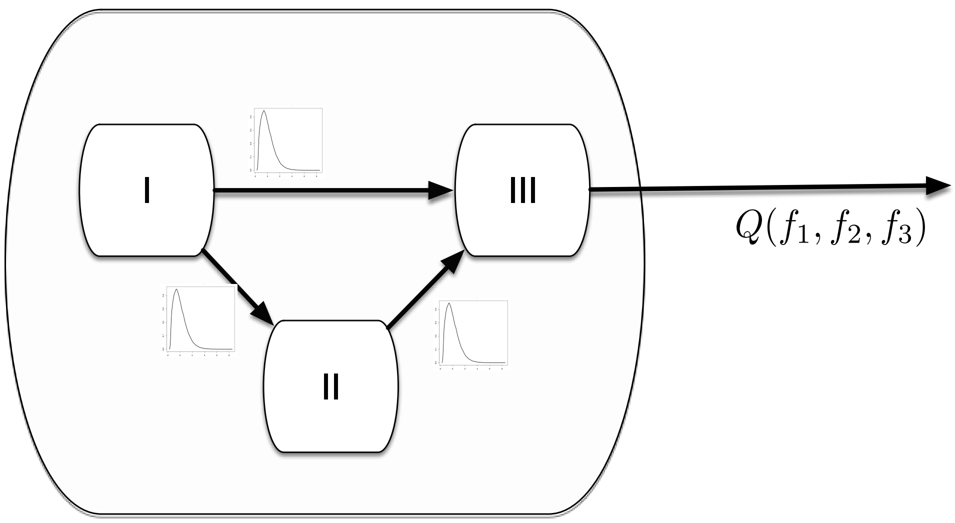
can come very close to any probability measure.

### Stable representations:

Continuous dependence on parameters

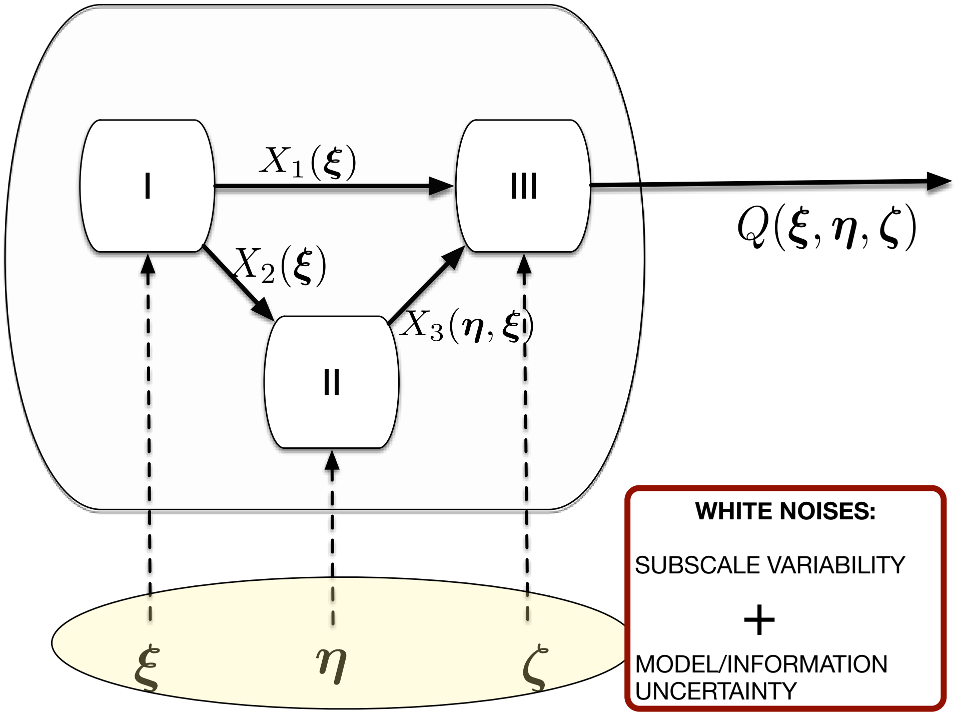
# Coupled Systems

## PDF at Interfaces



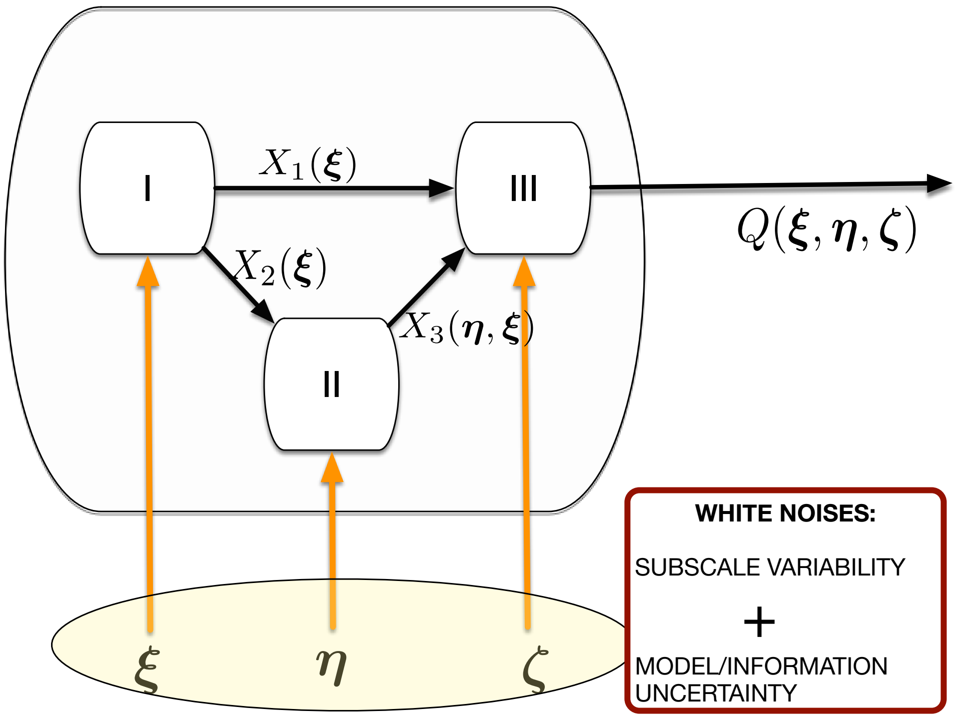
# Coupled Systems

## PCE at Interfaces



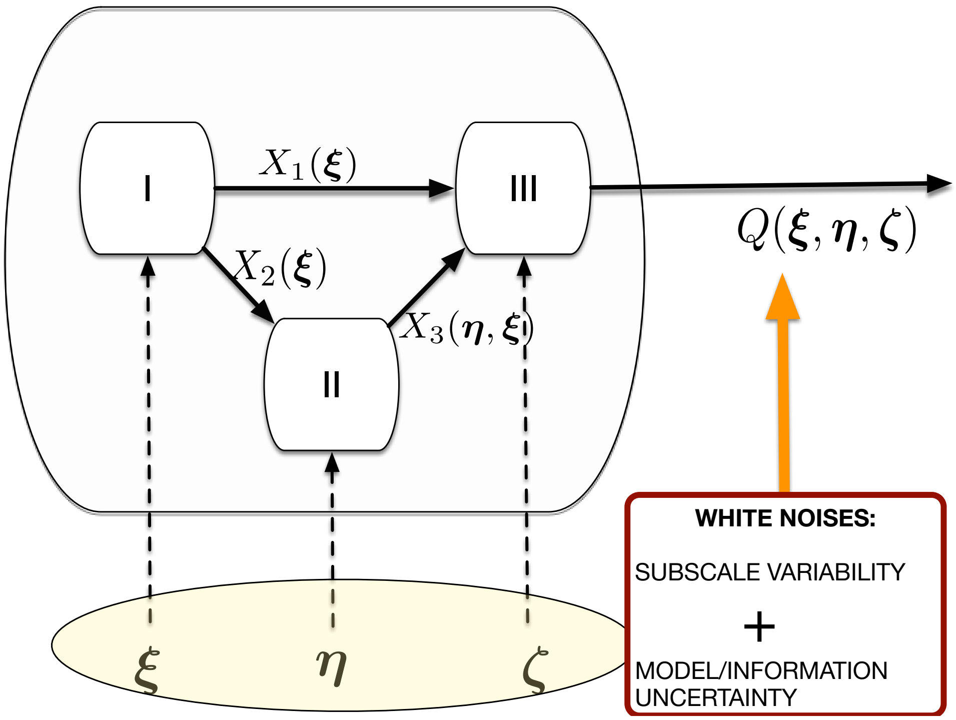
# Coupled Systems

## Updating Probabilistic Models of Subsystems



# Coupled Systems

## Updating Probabilistic Models of QoI



# Polynomial Chaos

$$\alpha(x,\omega) = \sum_{i=0}^{\infty} \alpha_i(x)\, \Psi_i(\boldsymbol{\xi}(\omega))$$

## Note

- Must estimate $\alpha_i$ constrained by information:
  - experimental constraints:
    -$\boldsymbol{\xi}$ captures endogenous sources of uncertainty.
  - physics constraints:
    -$\alpha$ depends on $\boldsymbol{\xi}$ through a conservation law that must be honored.
- Dimension of $\boldsymbol{\xi}$ reflects complexity of the process $\alpha$.
- Probability measure of $\boldsymbol{\xi}$ determines the geometry in which analysis and approximation are carried out.

# We embed the problem in a high-dimensional space

### Mean field approaches:

First order approximations smears out fluctuations.
Sacrificing reality for simplicity.

### Modern approaches

Consistent with modern sensors and computers.
Challenges are computational, mathematical, and logical.

# Outline

# Lightweight Vehicle

## Design of Composite Car

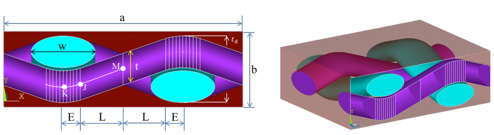



Figure 7: Unit cell RVE model of interlaced yarns in the woven fabric composite with (a) geometric characteristic parameters (b) 3D RVE model for permeability calculations.

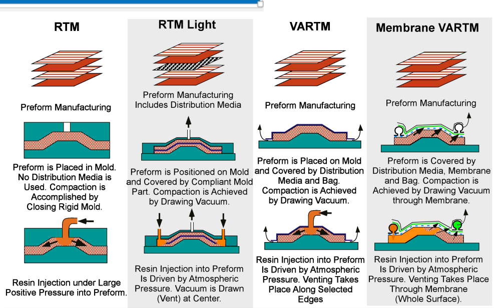



**Figure1.** Schematic of manufacturing steps for common LCM processes (a) Resin Transfer Molding (RTM) (b) RTM Light (c) Vacuum Assisted Resin Transfer Molding (VARTM) and (d) membrane VARTM.

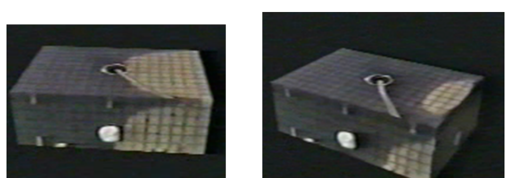
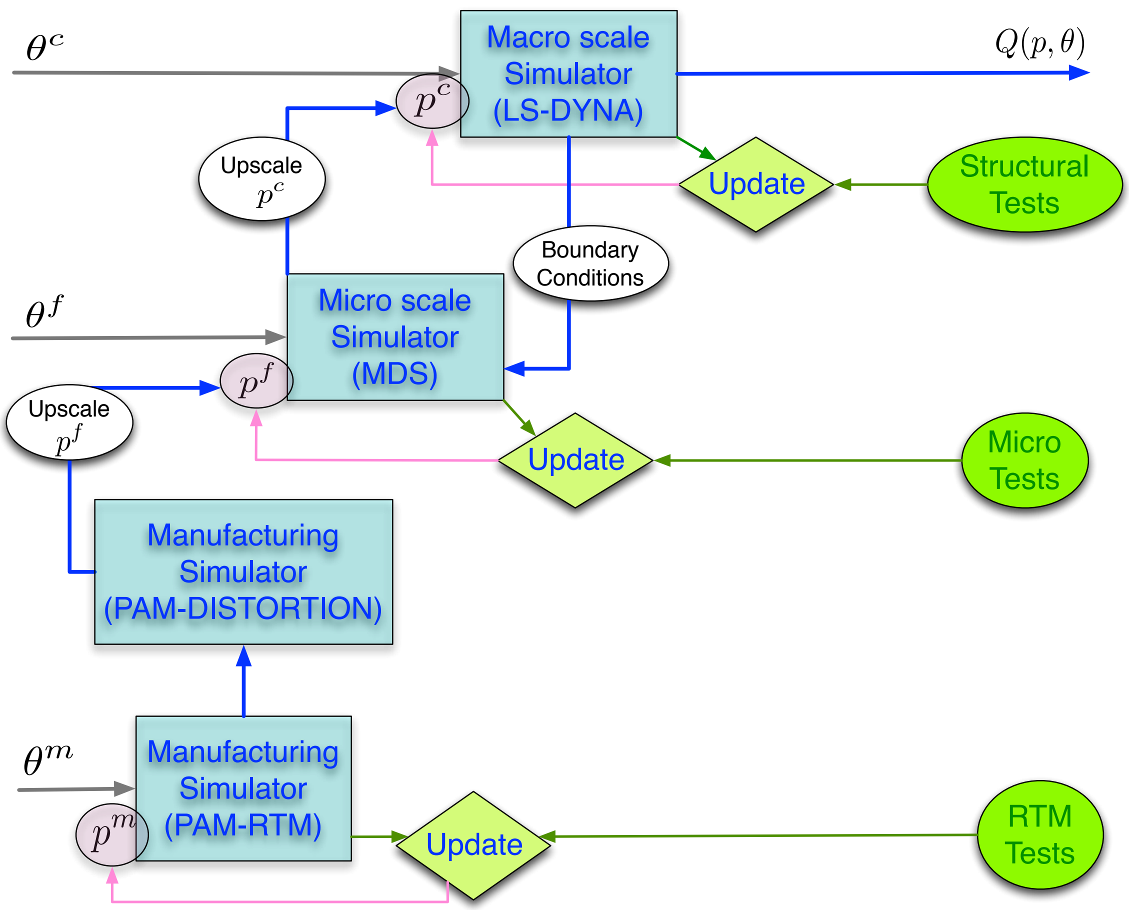
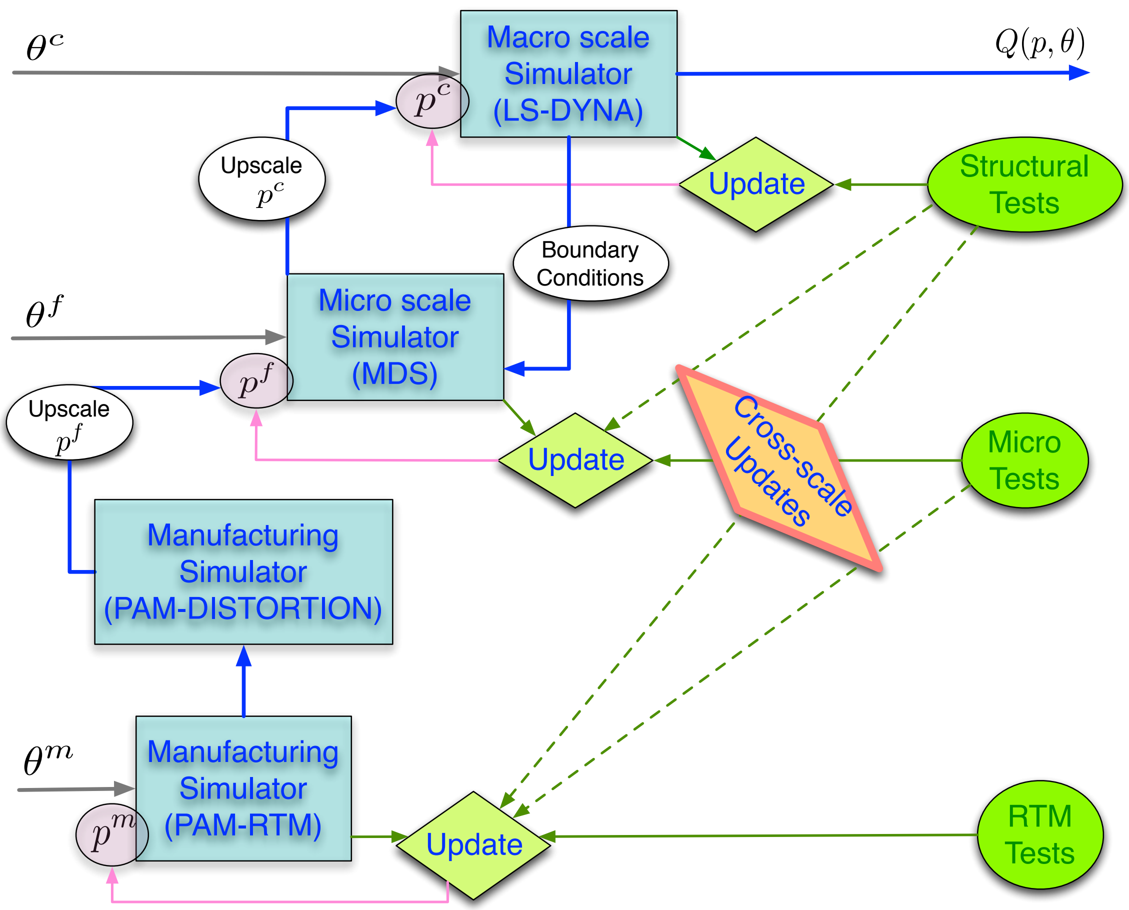


**Figure 2**: Racetracking as observed experimentally. The infused box images are assembled from experimental videos of individual sides. Resin "races" through channels formed in sharp corners by gaps between reinforcement and mold

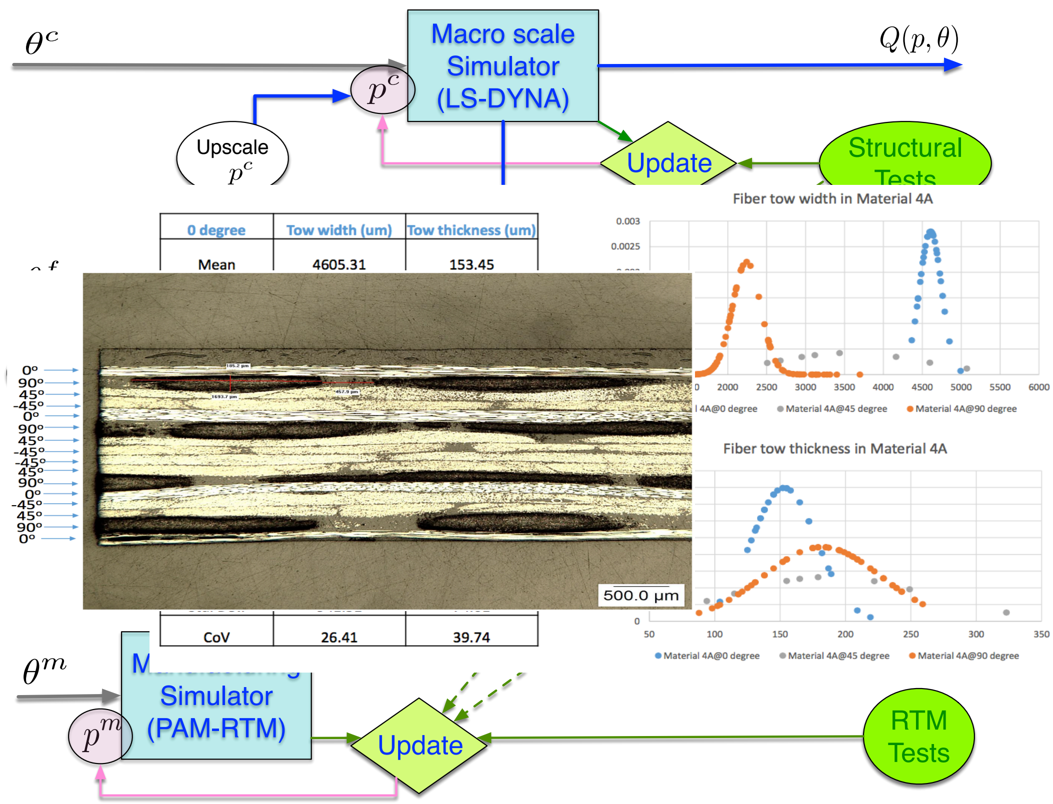
# Multiple Scales



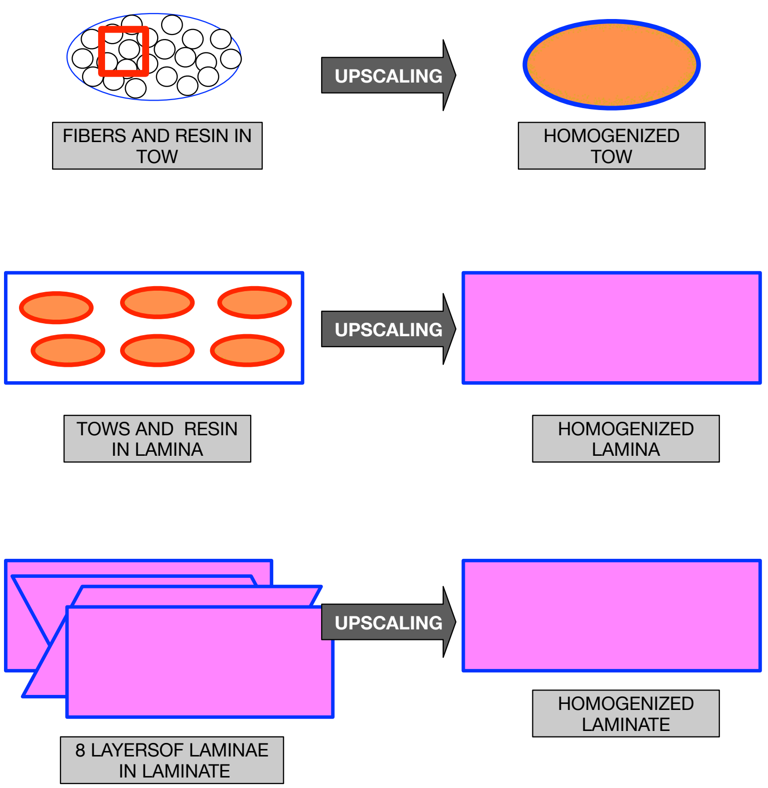
# Multiple Scales

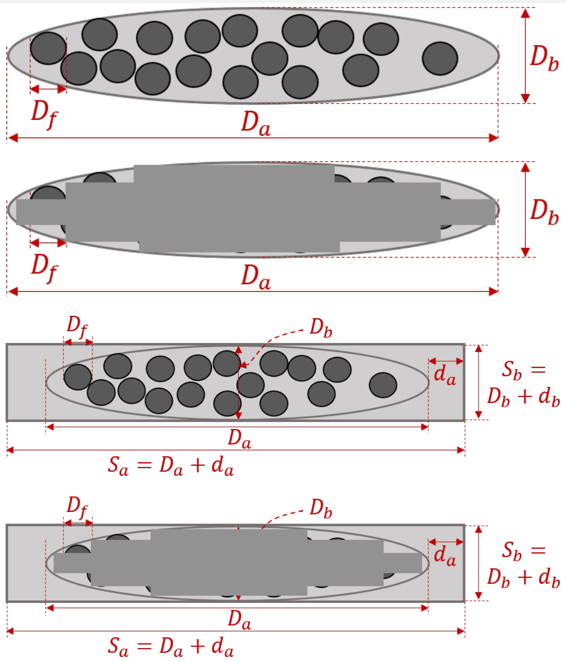


# Multiple Scales

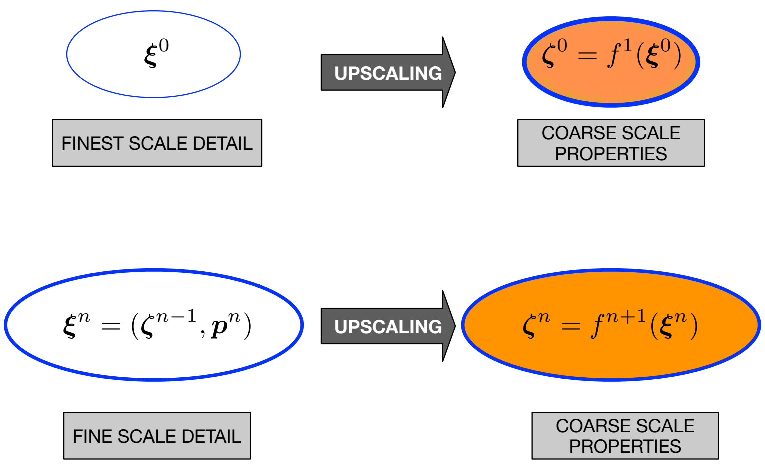


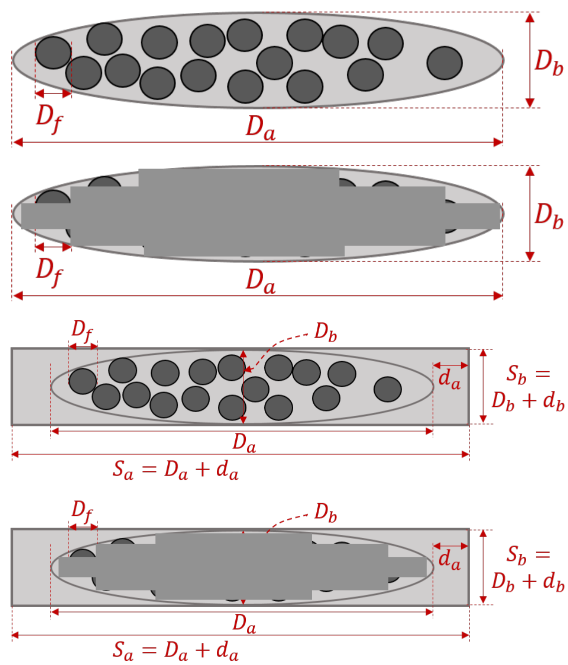
# Multiple Scales



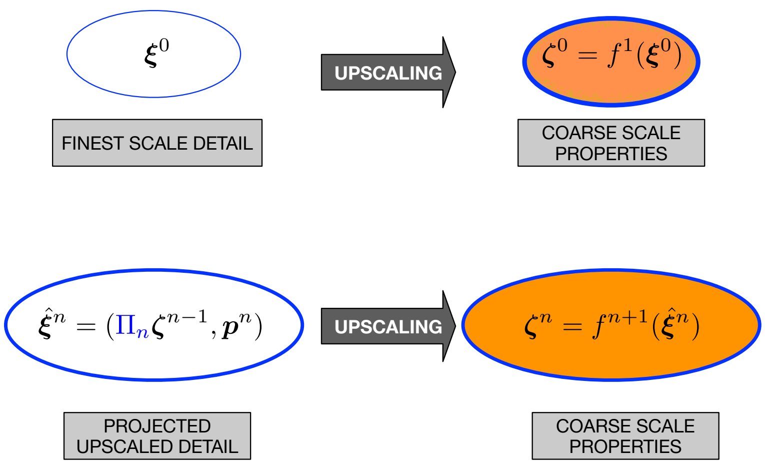



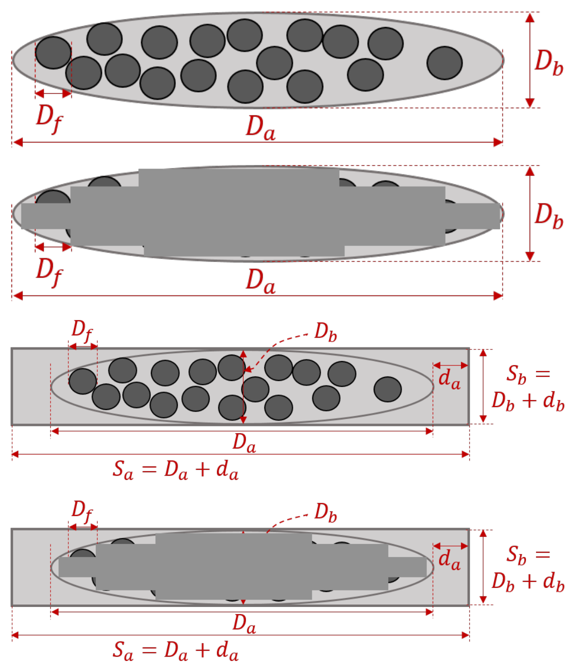
# Multiple Scales



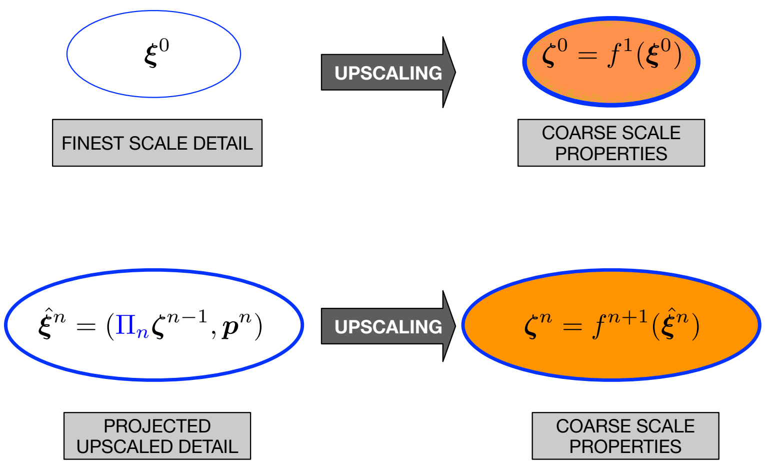



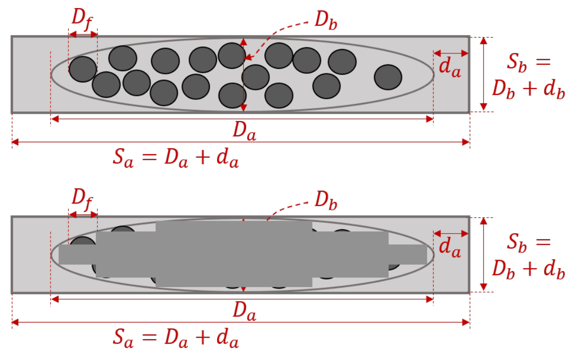
# Multiple Scales



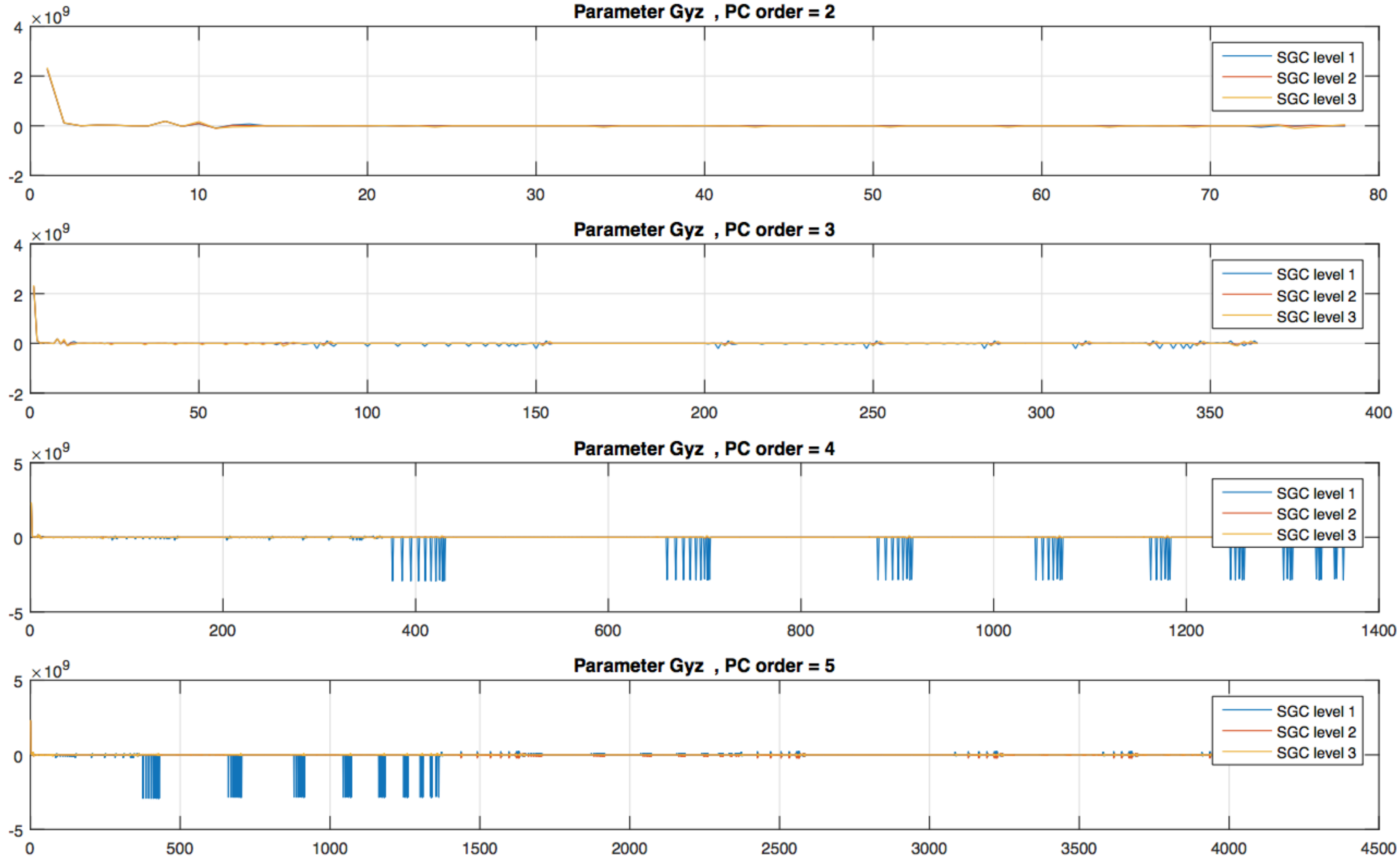
# Multiple Scales



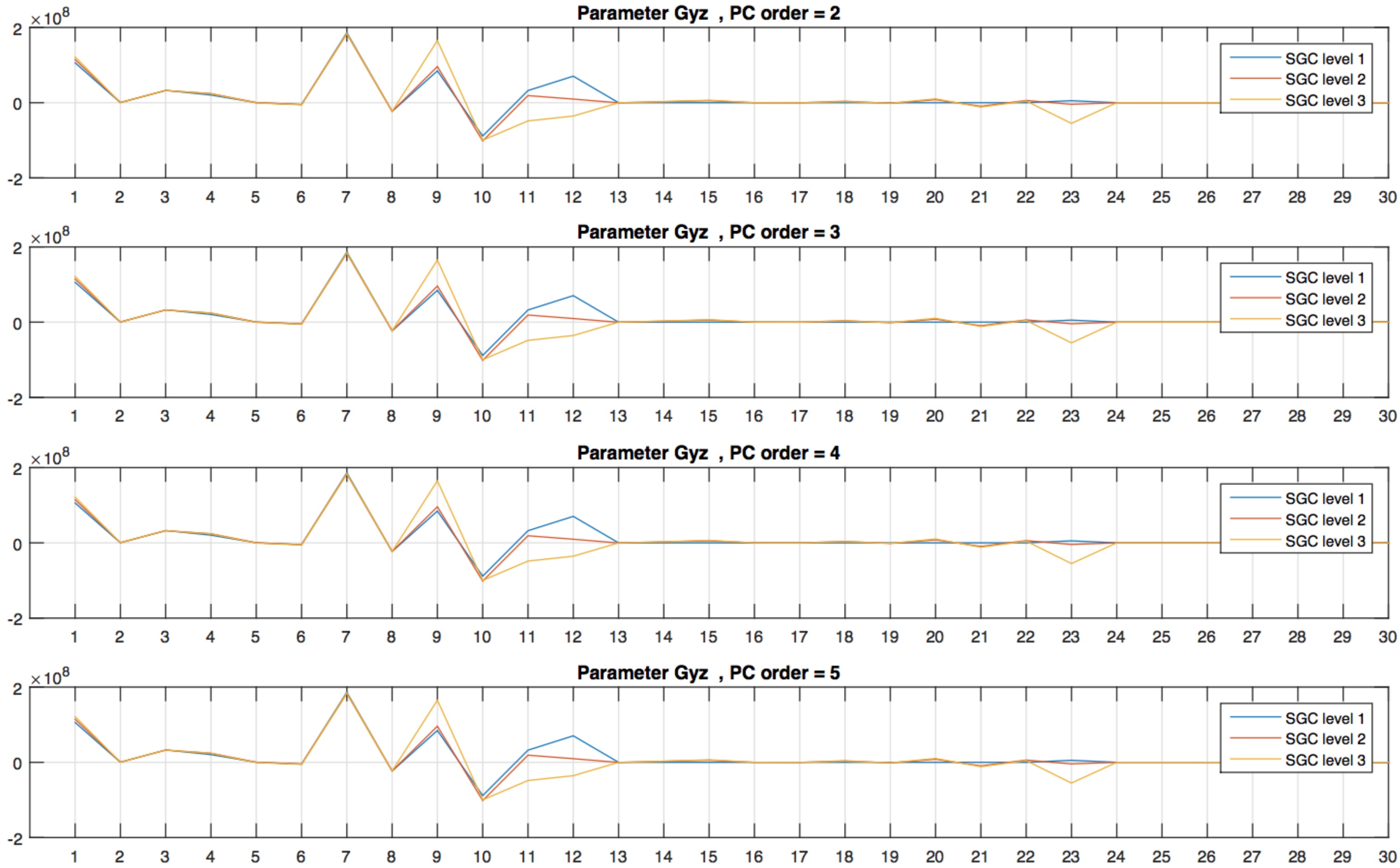




| Random variable | Mean value | Lower limit | Upper limit | C.O.V. |
|---|---|---|---|---|
| $V_f$ % | 83.4 | 76.4 | 90.4 | 4 % |
| $E_{fA}$ | 230.0e9 | 200e9 | 240e9 | 2.51 % |
| $E_{fT}$ | 20.0e9 | 14.8e9 | 25.2e9 | 15 % |
| $G_{fA}$ | 25.0e9 | 20.7e9 | 29.3e9 | 10 % |
| $\nu_{fA}$ | 0.016 | 0.013 | 0.019 | 10 % |
| $\nu_{fT}$ | 0.400 | 0.331 | 0.469 | 10 % |
| $E_m$ | 3.40e9 | 2.81e9 | 3.99e9 | 10 % |
| $\nu_m$ | 0.335 | 0.294 | 0.376 | 7 % |
| $D_a$ | 3.47 | 0.74 | 6.2 | 45.4 % |
| $d_a$ | 0.505 | 0.01 | 1.0 | 56.6 % |
| $D_b$ | 0.27 | 0.12 | 0.42 | 32.1 % |

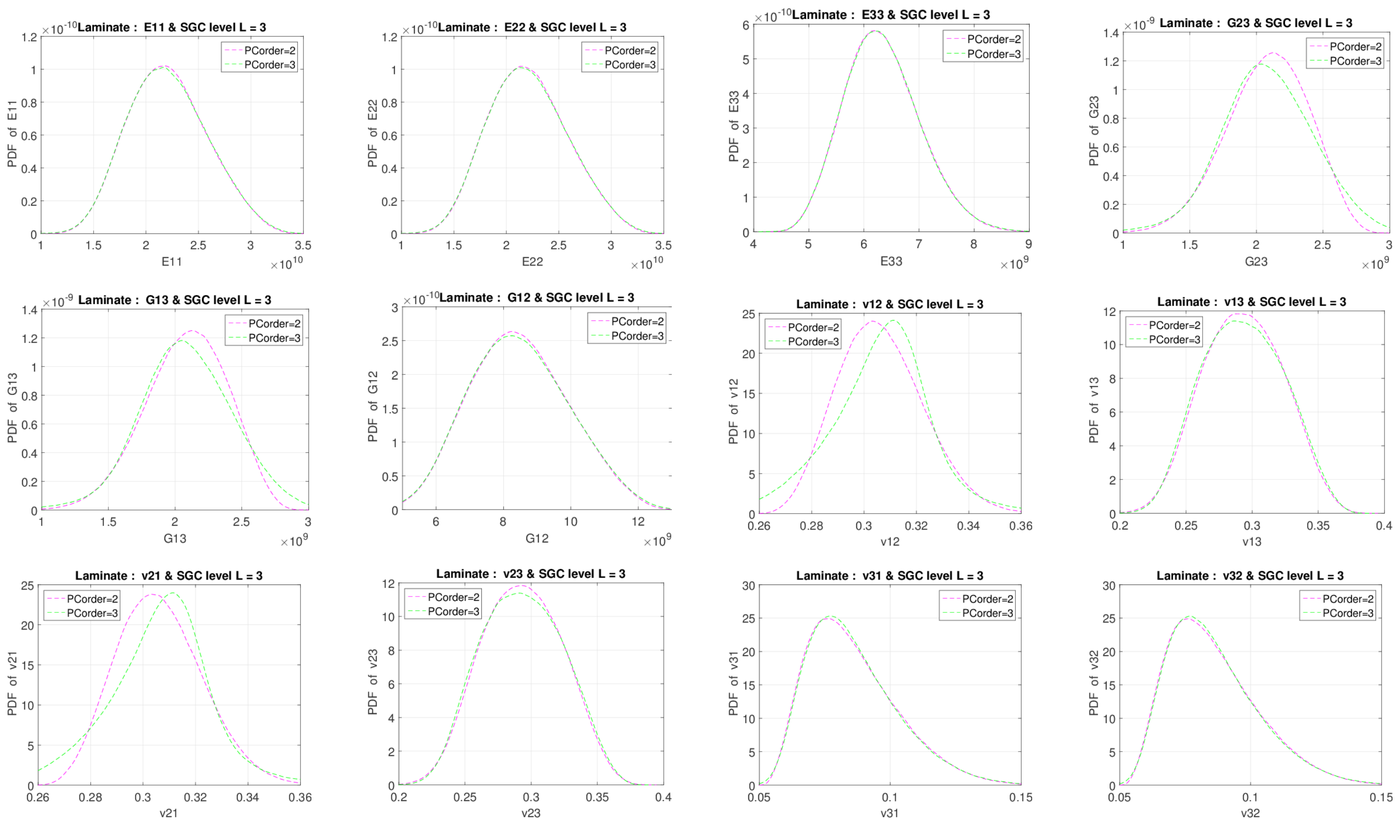
# PC of coarse scale in terms of fine scale



# PC of coarse scale in terms of fine scale



# Example of upscaling the elasticity matrix of the laminate



# Uncertainty in material processing stage (RTM)

## Uncertainty propagation in 13D

Uncertainty in ($K_1^{(0)}$,$K_2^{(0)}$, $K_3^{(0)}$, $K_{cha}^{(0)}$, $\phi$, $T_1$, $T_2$, $T_3$, $T_4$, $T_5$, $T_6$, $T_7$, $T_8$).

# Results

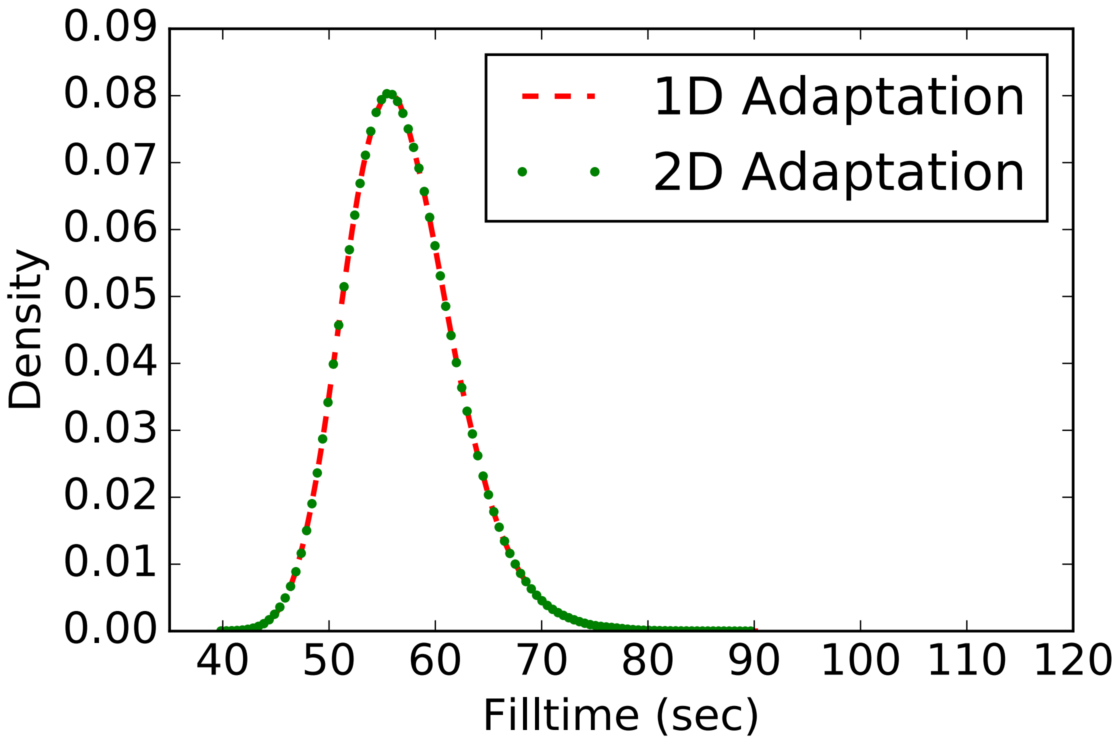



Figure: pdf of the fill time.

# Adaptation: $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\xi}$

- Numerical RTM model (PAM-RTM) evaluated 57 times; Gaussian part computation (27 times) + 1D adaptation computation (9 times) + 2D adaptation computation (21 times)
- Dominant $\eta$ direction is

$$\begin{aligned}\eta(\boldsymbol{\xi}) = -0.95226259\, K_1^{(0)}(\xi_1) - 0.21606623\, K_2^{(0)}(\xi_2) \\ -0.00269772\, K_3^{(0)}(\xi_3) - 0.00377367\, K_{cha}^{(0)}(\xi_4) \\ -0.00640321\, \phi(\xi_5) - 0.00641398\ T_1(\xi_6) \\ -0.02025733\ T_2(\xi_7) - 0.02158714\ T_3(\xi_8) \\ +0.04207637\ T_4(\xi_9) - 0.04635585\ T_5(\xi_{10}) + \\ 0.0593152\ T_6(\xi_{11}) + 0.06031331\ T_7(\xi_{12}) \\ -0.18562338\ T_8(\xi_{13}).\end{aligned}$$

# Prediction vs Measurements

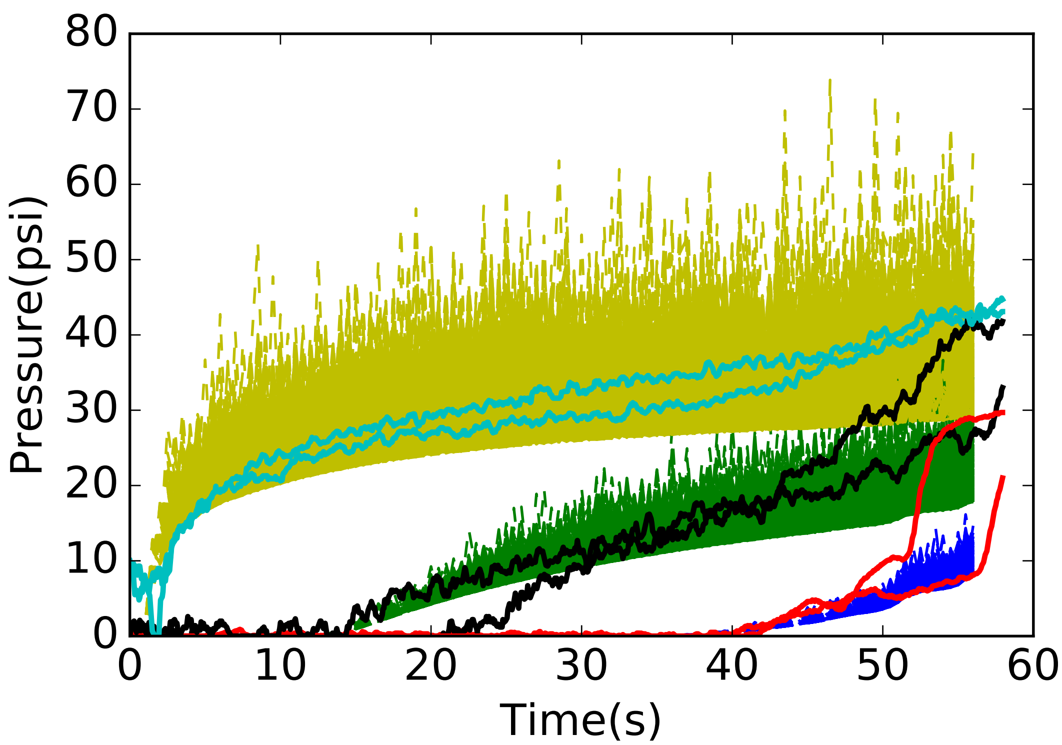



Figure: Realizations of the pressure at the sensors.

# Adapted Bases: $\eta = A\xi$

## Leveraging structure in high-dimensional geometry

### Fill time of resin transfer molding

| | Method | Number of function evaluations | Comments |
|---|---|---|---|
| 1 | Monte Carlo | Computationally prohibitive | PDF, not functional map. |
| 2 | PCE with quadrature | $n^d(2^{18})$ | Functional map from input to output. |
| 3 | PCE with Sparse Grid | $< n^d$ (10,597 for level 3) | Functional map from input to output. |
| 4 | PCE with adapted basis | 46 | Adapted to specific QoI. |

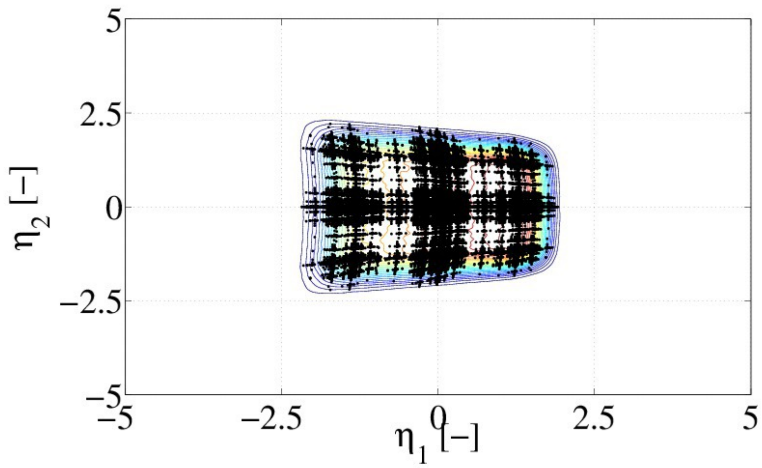
# Embedded Quadrature

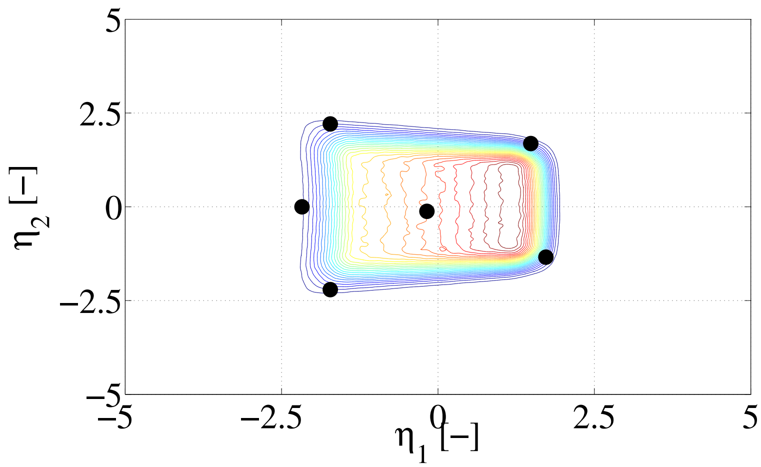
## Leveraging structure in high-dimensional geometry

**Probabilities of events are integrals over the basic random variables:**

Downselect the gauss quadratures from the high-dimensional space using $l_1$ regularization on the weights.

**Reduced Quadrature in High Dimensions**



# What to do with evidence ?

## Data can be acquired on several scales, often separately.

- Data can be used to calibrate the finer scales.
- Data can be used to valdate the upscaling process.

## Validating Upscaling

- We use models of random tensors to capture theoretical constraints on bounds of tensors.
- Statistical comparisons of these models against models upscaled through models.
- Stochastic sensitivity to understand the discrepancy.

# Bounded Matrix

Constrain the probabilistic model of the heterogeneous material with this information: $\mathcal{G} = \{G \in \mathbb{M}_n^+ : G_l < G < G_u\}$

$$\begin{cases} \int_{\mathcal{G}} p_{\mathcal{G}}(G) dG & = & 1 \\ \int_{\mathcal{G}} \ln\left[\det(G - G_l)\right] p_{\mathcal{G}}(G) dG & = & g_l \\ \int_{\mathcal{G}} \ln\left[\det(G_u - G)\right] p_{\mathcal{G}}(G) dG & = & g_u \end{cases}$$

## MaxEnt Distribution of Bounded Random Matrix:

Using Lagrange Multipliers, we maximize the following Lagrangian:

$$
\begin{aligned}
\mathcal{L}(p_{\mathsf{C}}, \lambda_l, \lambda_u) &= -H(p_{\mathsf{G}}) + (\lambda_0 - 1)\Big[\int_{\mathcal{G}} p_{\mathsf{G}}(G)\, dG - 1\Big] \\
&+ \lambda_l \Big[\int_{\mathcal{G}} \ln[\det(G - G_l)] \times p_{\mathsf{G}}(G)\, dG - g_l\Big] \\
&+ \lambda_u \Big[\int_{\mathcal{C}} \ln[\det(G_u - G)]\, p_{\mathsf{G}}(G)\, dG - g_u\Big],
\end{aligned}
$$

MaxEnt Distribution of Bounded Random Matrix:

$$p_{\mathbf{G}}(G) = \frac{\det(G - G_l)^{a-(N+1)/2} \det(G_u - G)^{b+(N+1)/2}}{\beta_N(a, b) \det(G_u - G_l)^{(a+b)-(N+1)/2}}$$

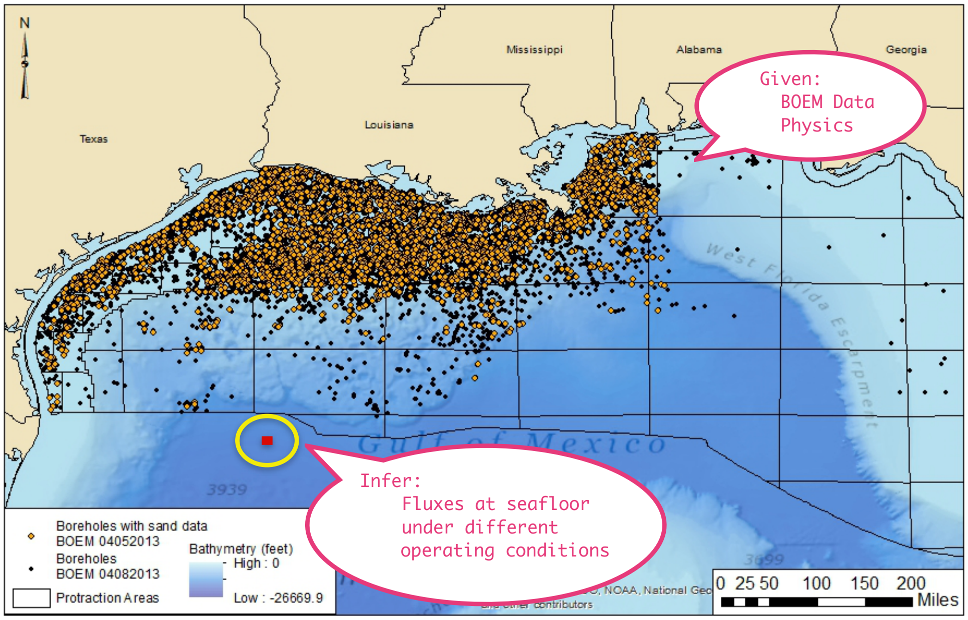
Notes:

- $a$, $b$ are obtained from the MaxEnt optimization.
- efficient sampling algorithms have been developed for this distribution.
- each realization $G$ of the random matrix is obtained through an inverse analysis from experimental measurements of the mechanical field.

# Outline

# Objective

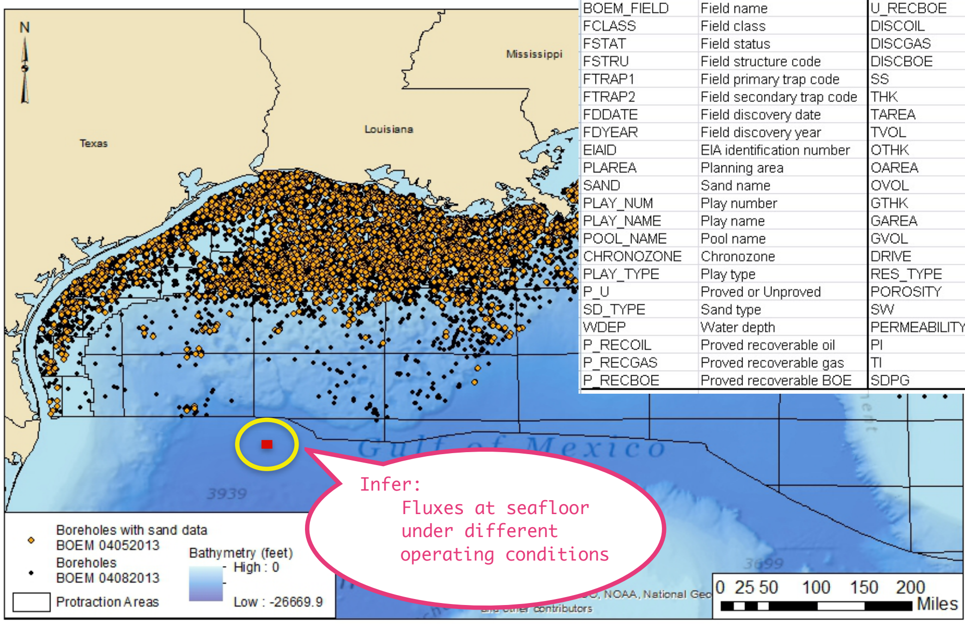
## Data from BOEM: Bureau of Ocean Data Management



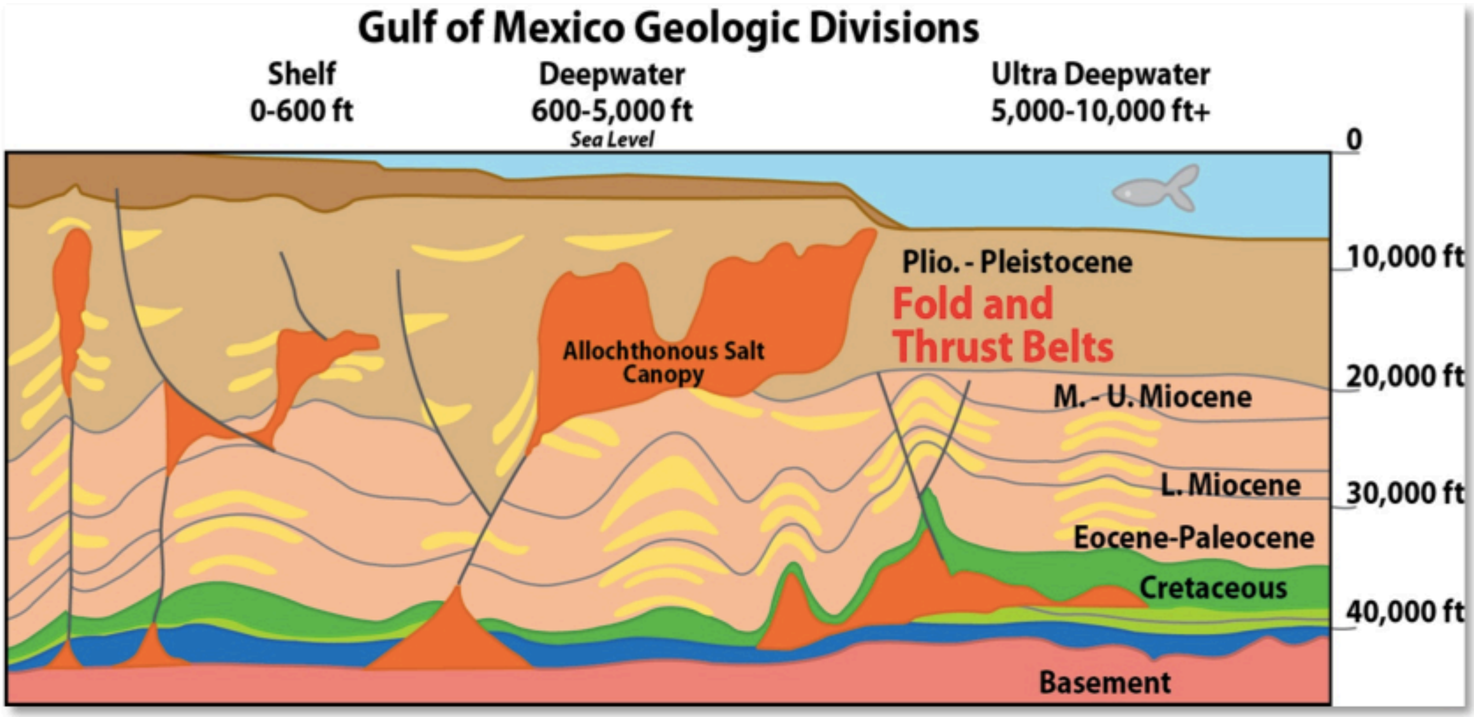
# Objective

## Data from BOEM: Bureau of Ocean Data Management

| Name | Alias | Name | Alias | Name | Alias |
|---|---|---|---|---|---|
| SN_FORMSAND | Sand sequence number | P_CUMOIL | Cumulative oil produced | SDTG | Sand temperature gradient |
| SAND_NAME | Sand name | P_CUMGAS | Cumulative gas produced | RSI | Initial solution gas-oil ratio |
| ASSESSED | Assessed | P_CUMBOE | Cumulative BOE produced | YIELD | Yield |
| SDDATE | Sand discovery date | P_REMOIL | Proved remaining recoverable oil | PROP | Proportion oil |
| SDYEAR | Sand discovery year | P_REMGAS | Proved remaining recoverable gas | GOR | Gas-oil ratio |
| SDDATEH | Sand discovery date high | P_REMBOE | Proved remaining recoverable BOE | SPGR | Gas specific gravity |
| SDYEAREH | Sand discovery year high | U_RECOIL | Unproved recoverable oil | API | Oil API gravity |
| WELLAPI | Discovery well | U_RECGAS | Unproved recoverable gas | BGI | Initial gas formation volume f |
| BOEM_FIELD | Field name | U_RECBOE | Unproved recoverable BOE | BOI | Initial oil formation volume fac |
| FCLASS | Field class | DISCOIL | Discovered oil | RECO_AF | Recoverable oil per acre-foot |
| FSTAT | Field status | DISCGAS | Discovered gas | RECG_AF | Recoverable gas per acre-fo |
| FSTRU | Field structure code | DISCBOE | Discovered BOE | OIP | Oil in place |
| FTRAP1 | Field primary trap code | SS | Subsea depth | GIP | Gas in place |
| FTRAP2 | Field secondary trap code | THK | Total average net thickness | ORF | Oil recovery factor |
| FDDATE | Field discovery date | TAREA | Total area | ORECO | Oil reservoirs' recoverable o |
| FDYEAR | Field discovery year | TVOL | Total volume | ORECG | Oil reservoirs' recoverable ga |
| EIAID | EIA identification number | OTHK | Oil average net thickness | ORP | Produced GOR for oil reserv |
| PLAREA | Planning area | OAREA | Oil total area | GRF | Gas recovery factor |
| SAND | Sand name | OVOL | Oil total volume | GRECO | Gas reservoirs' recoverable |
| PLAY_NUM | Play number | GTHK | Gas average net thickness | GRECG | Gas reservoirs' recoverable |
| PLAY_NAME | Play name | GAREA | Gas total area | GRP | Produced GOR for gas reser |
| POOL_NAME | Pool name | GVOL | Gas total volume | NCNT | Count of nonassociated gas |
| CHRONOZONE | Chronozone | DRIVE | Drive type | UCNT | Count of undersaturated oil r |
| PLAY_TYPE | Play type | RES_TYPE | Reservoir type | SCNT | Count of saturated oil reserv |
| P_U | Proved or Unproved | POROSITY | Porosity | TRCNT | Count of total reservoirs in th |
| SD_TYPE | Sand type | SW | Water saturation | | |
| WDEP | Water depth | PERMEABILITY | Permeability | | |
| P_RECOIL | Proved recoverable oil | PI | Initial pressure | | |
| P_RECGAS | Proved recoverable gas | TI | Initial temperature | | |
| P_RECBOE | Proved recoverable BOE | SDPG | Sand pressure gradient | | |

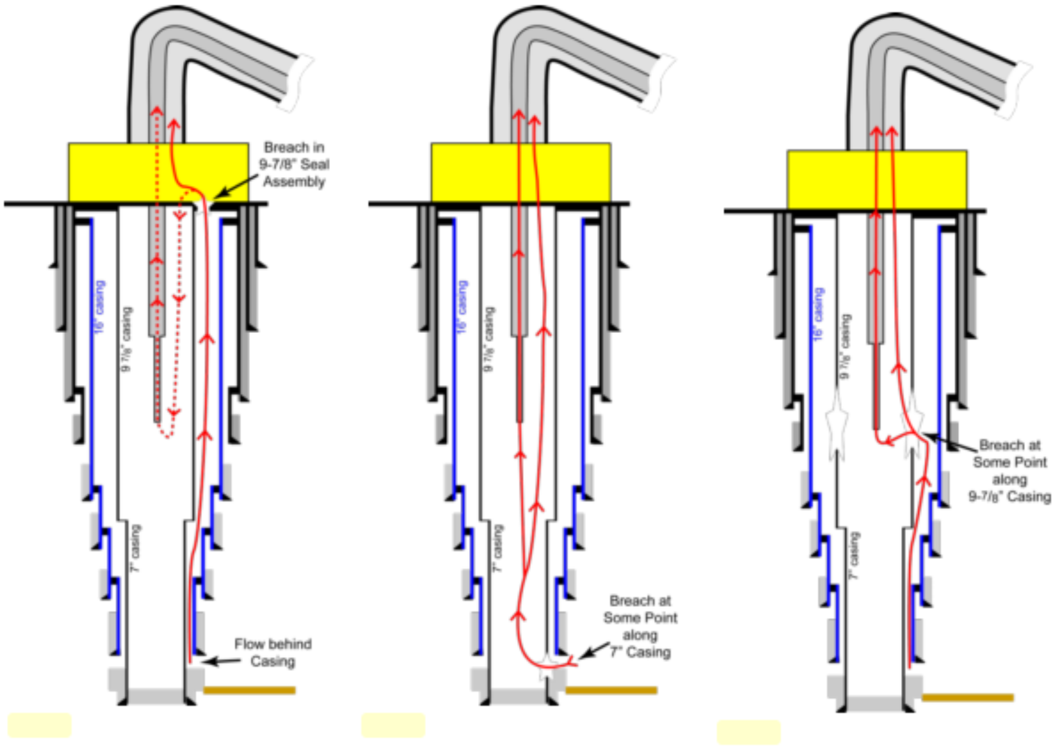


# Complexity



Gulf of Mexico Geologic Divisions
Shelf
0-600 ft
Deepwater
600-5,000 ft
Sea Level
Ultra Deepwater
5,000-10,000 ft+
0
10,000 ft
20,000 ft
30,000 ft
40,000 ft
Plio.-Pleistocene
Fold and
Thrust Belts
Allochthonous Salt
Canopy
M.-U. Miocene
L. Miocene
Eocene-Paleocene
Cretaceous
Basement

# Complexity


Breach in
9-7/8" Seal
Assembly
Flow behind
Casing
Breach at
Some Point
along
7" Casing
Breach at
Some Point
along
9-7/8" Casing

# Multiscale Knowledge

### Physics Constraints

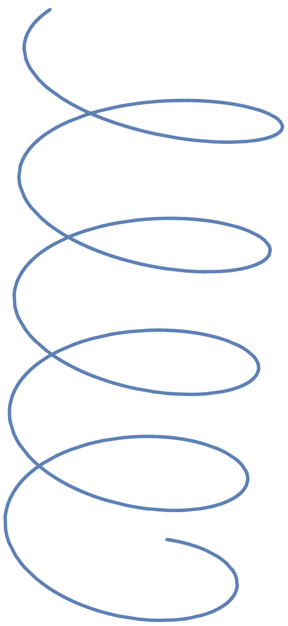
- At scale of each reservoir: Multiphase flow in porous medium.
- At scale of wellbore at location of interest

### Poorly understood constraints

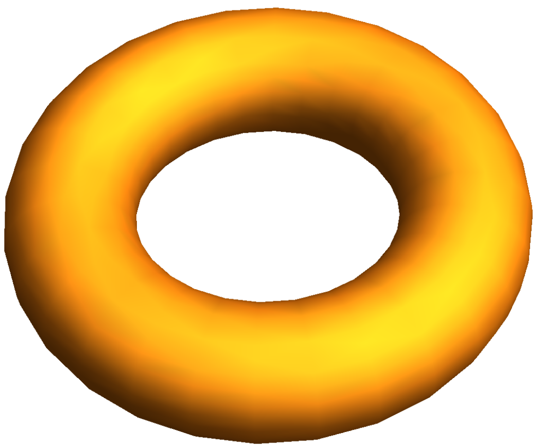
- local geological fluctuations
- global geological fluctuations

# Diffusion Metric

A Markov chain with controlled step evolving on the data set will eventually discover intrinsic distances between data points.



Spiral has intrinsic dimension of 1.



Torus has intrinsic dimension of 2.

## Decomposition of uncertainty:

$$\alpha(\omega) = f(\underbrace{\xi_1, \cdots, \xi_n}_{\text{Aleatoric Uncertainty}}, \underbrace{\xi_{n+1}, \cdots, \xi_m}_{\text{Model/Data Uncertainty}})$$

$$= \sum_{ijkl} \alpha_{ijkl}(x)\, \psi_i(\boldsymbol{\xi}_1)\, \psi_j(\boldsymbol{\xi}_2)\, \psi_k(\boldsymbol{\xi}_3)\, \psi_l(\boldsymbol{\xi}_4)$$

- $\xi_1$ uncertainty at scale of reservoir (polynomial chaos).
- $\xi_2$ uncertainty at scale of field (lease) (Gaussian process interpolation).
- $\xi_3$ uncertainty at scale of Gulf of Mexico (Gaussian Process with Diffusion metric from data).
- $\xi_4$ uncertainty in wellbore performance at target site.

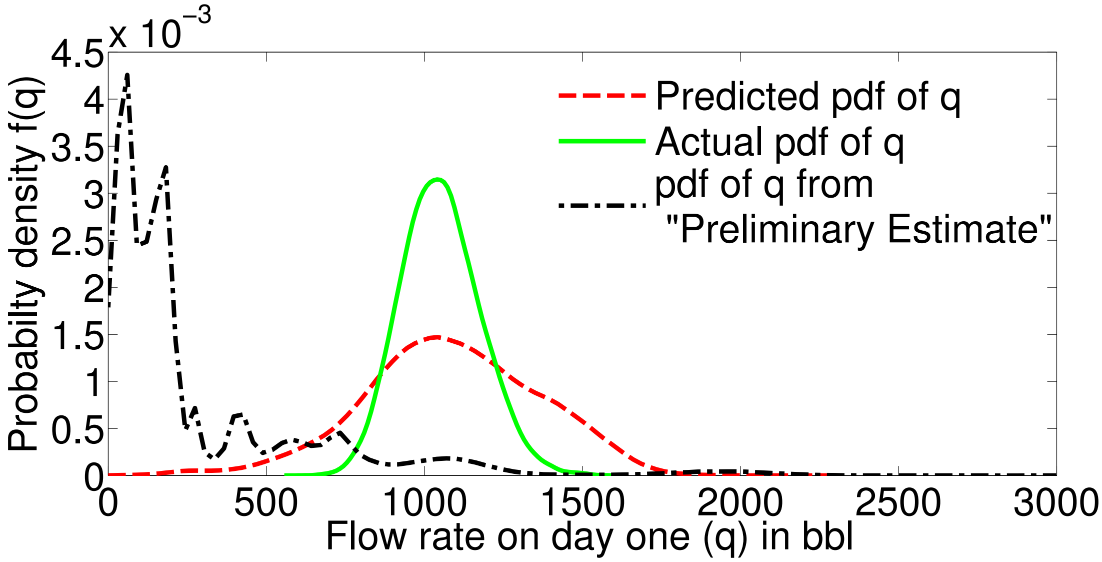
## Decomposition of uncertainty:

$$\alpha(\omega) = f(\underbrace{\xi_1, \cdots, \xi_n}_{\text{Aleatoric Uncertainty}}, \underbrace{\xi_{n+1}, \cdots, \xi_m}_{\text{Model/Data Uncertainty}})$$

$$= \sum_{ijkl} \alpha_{ijkl}(x)\, \psi_i(\boldsymbol{\xi}_1)\, \psi_j(\boldsymbol{\xi}_2)\, \psi_k(\boldsymbol{\xi}_3)\, \psi_l(\boldsymbol{\xi}_4)$$

- $\boldsymbol{\xi}_1$ uncertainty at scale of reservoir (polynomial chaos).
- $\boldsymbol{\xi}_2$ uncertainty at scale of field (lease) (Gaussian process interpolation).
- $\boldsymbol{\xi}_3$ uncertainty at scale of Gulf of Mexico (Gaussian Process with Diffusion metric from data).
- $\boldsymbol{\xi}_4$ uncertainty in wellbore performance at target site.

# Sample Results




]

# Outline

# Comments: Constraints are building blocks UQ models

Mathematical constraints:

- sum of rows equal 1.
- satisfy a governing equation

Physics constraints:

- bounds on elasticity
- symmetry and conservation laws
- behavior across interfaces

Data constraints:

- Statistics
- Manifolds

# References


- Tipirreddy, R. and Ghanem, R. (2014) Basis adaptation in homogeneous chaos spaces, *Journal of Computational Physics*, **259**:304-317.

- Arnst, M., Ghanem, R., Phipps, E., and Red-Horse, (2012) J., Measure transformation and efficient quadrature in reduced-dimensional stochastic modeling of coupled problems, *International Journal for Numerical Methods in Engineering,* **92**: 1044-1080.

- Soize, C. and Ghanem, R., (2015) Data-driven probability concentration and sampling on manifolds, submitted to *Journal of Computational Physics.*

- Thimmisetty, C., Khodabakhshnejad, A., Jabbari, N., Aminzadeh, F., Ghanem, R., Rose, K., Disenhof, C., and Bauer, J., (2015) Multiscale Stochastic Representation in High-Dimensional Data using Gaussian Processes with Implicit Diffusion Metrics, *Lecture Notes in Computer Science, Vol. 8964,* Springer-Verlag; DOI: 10.1007/978-3-319-25138-7 15 .

- Das, S. and Ghanem, R. (2009) A Bounded random matrix approach for stochastic upscaling, *SIAM Journal on Multiscale Modeling & Simulation*, **8**(1):296-325.