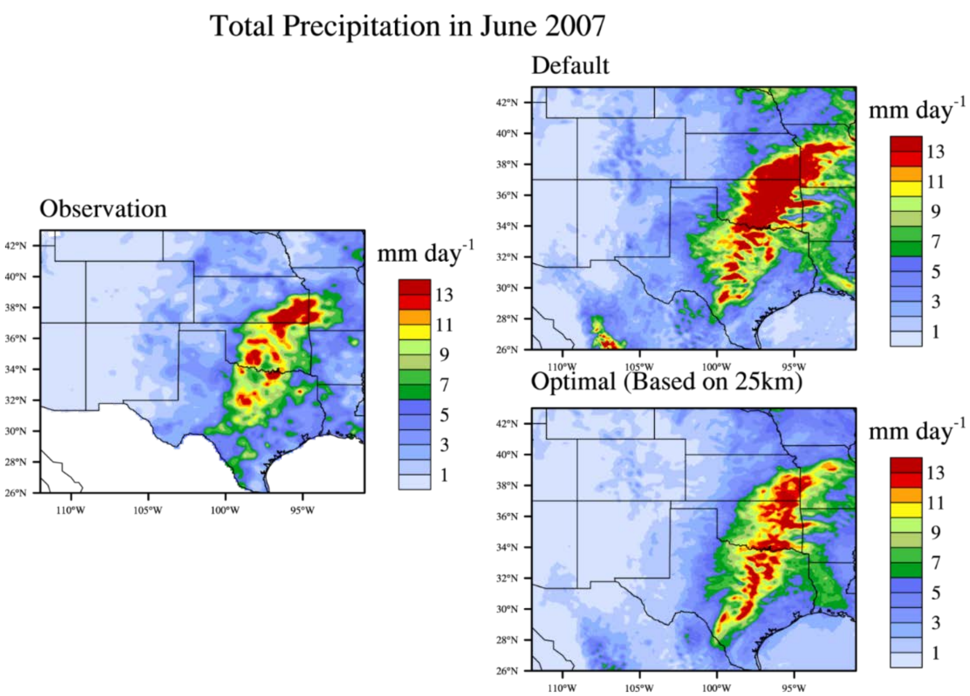
Uncertainty Quantification and Scientific Machine Learning for Complex Engineering Systems

Guang Lin,

Director, Data Science Consulting Service, Departments of Mathematics, Statistics, Mechanical Engineering, Earth, Atmospheric, and Planetary Sciences, Purdue University

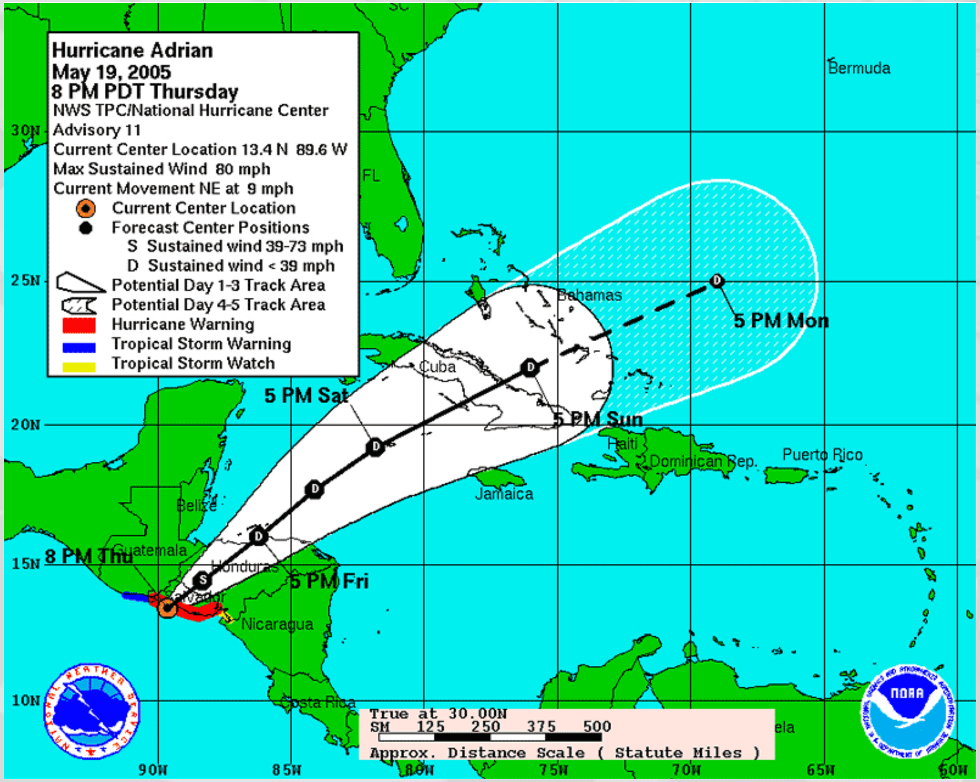


BNC Faculty Seminar Series., Feb. 13, 2020

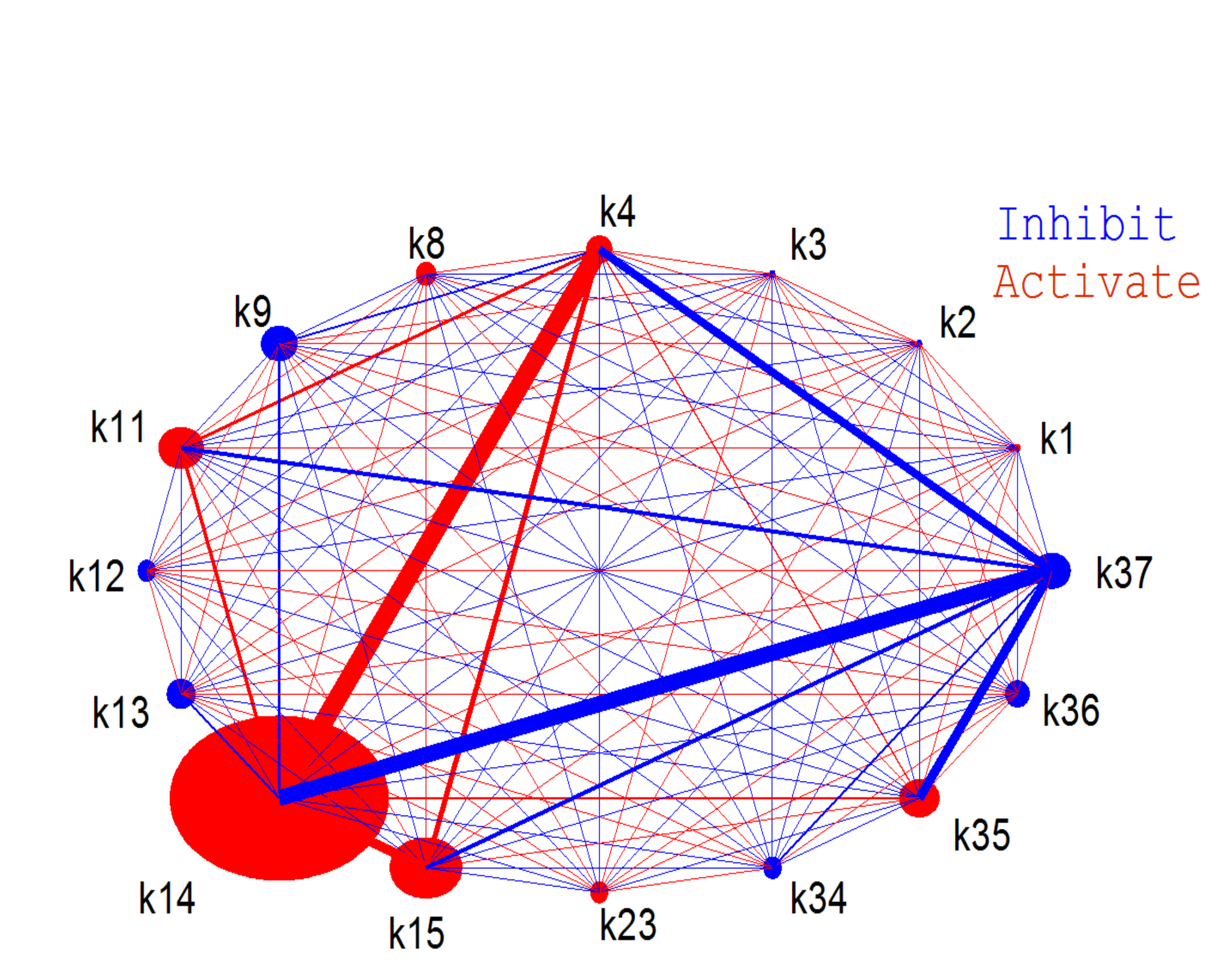
Why Uncertainty Quantification?

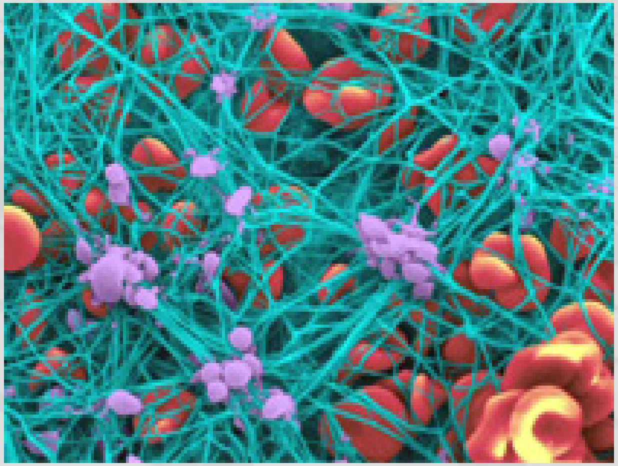
-Use available data to improve high-fidelity model's predictive capability to enable new scientific discovery and make critical decision

UQ for Decision Making: Hurricane Forecasting



Sensitivity Analysis of Reaction Networks Related to Tissue Factor Pathway of Blood Coagulation



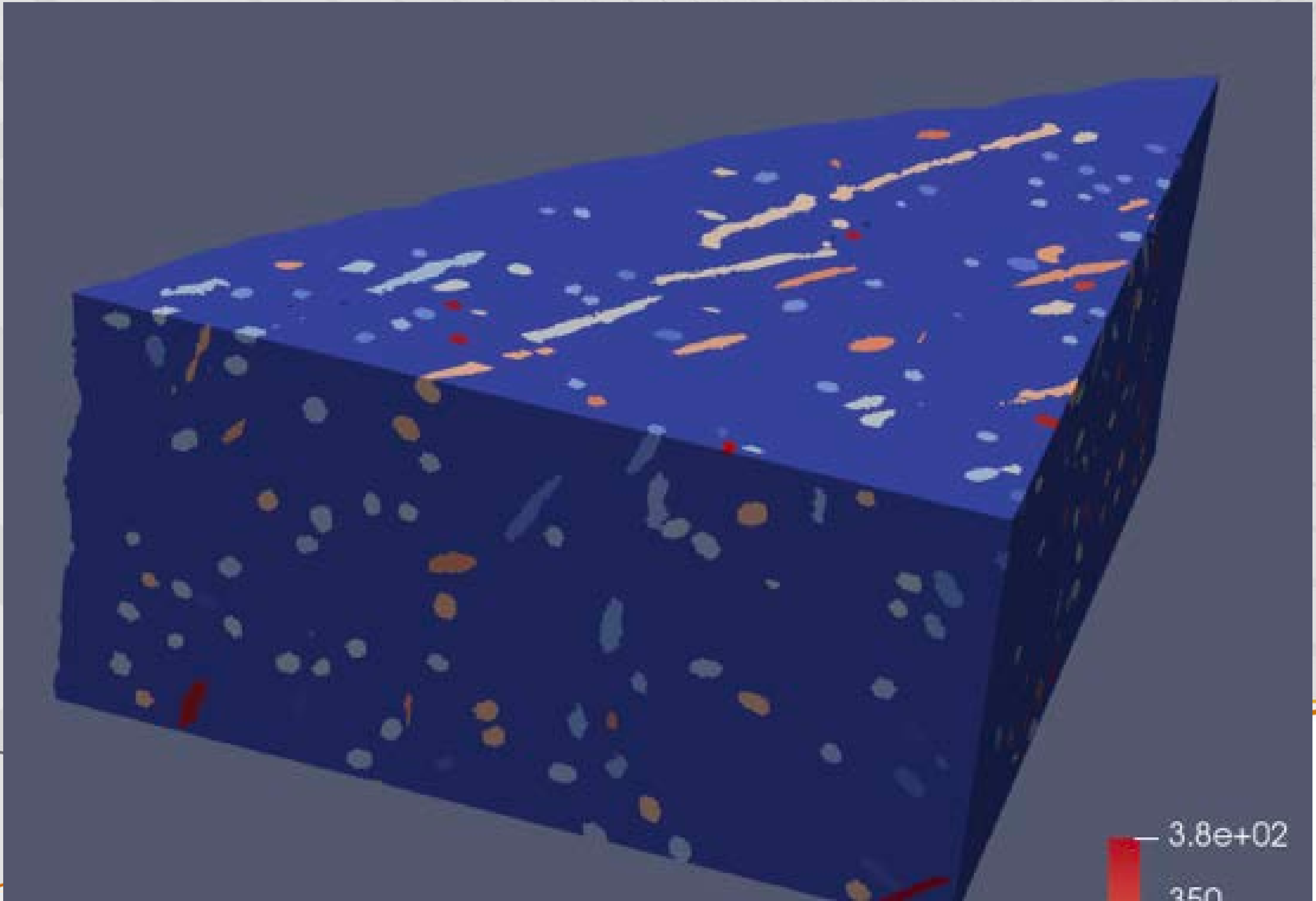


Identify the important coagulation factors (reaction rates) and their interactions in blood coagulation with respect to total thrombin

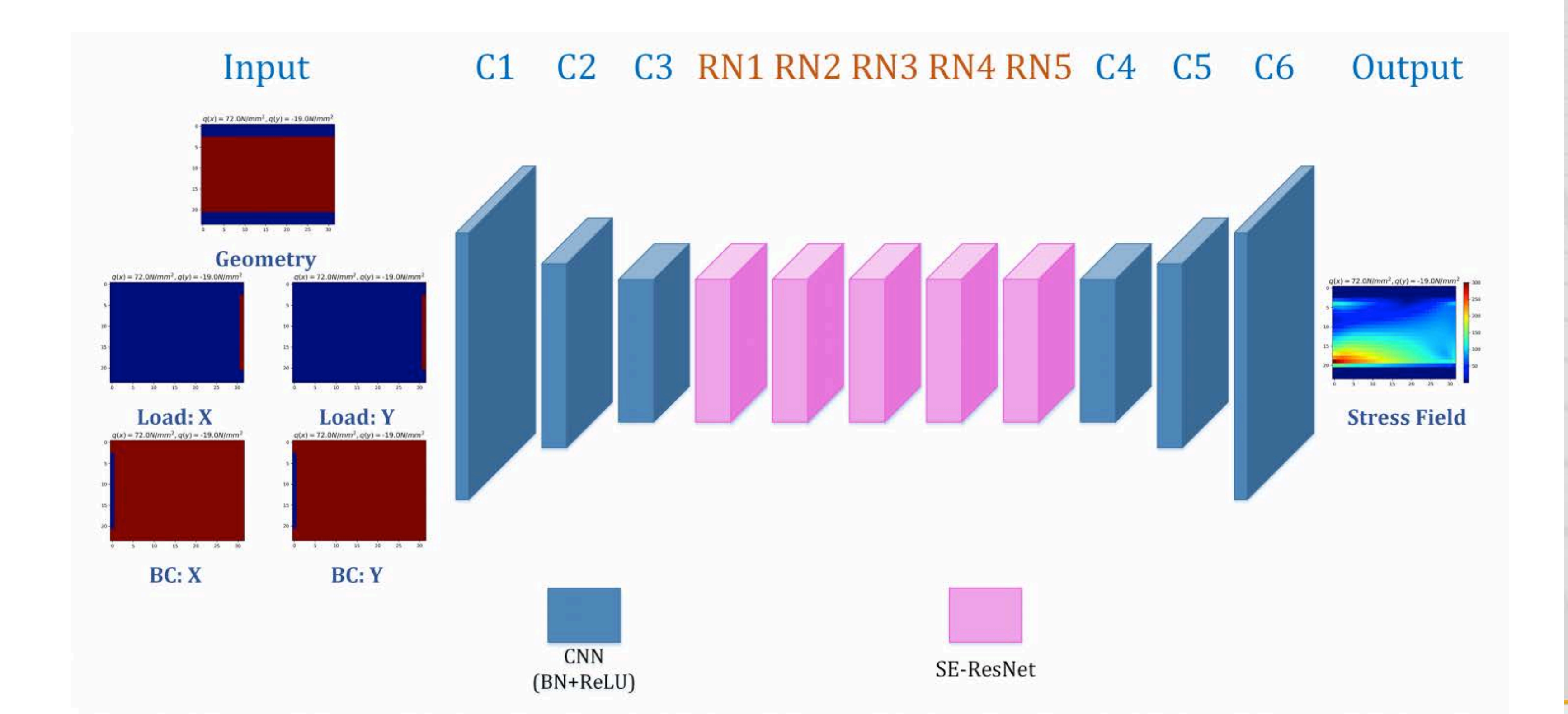
Sensitivity Analysis of Reaction Networks Related to Tissue Factor Pathway of Blood Coagulation

Identify the sensitivity shift and their interactions in blood coagulation with respect to total thrombin when blocking K14 and K15

Deep Learning for Material Science: DNN-based Processing-Structure-Performance Map for Fibre Reinforced Polymer



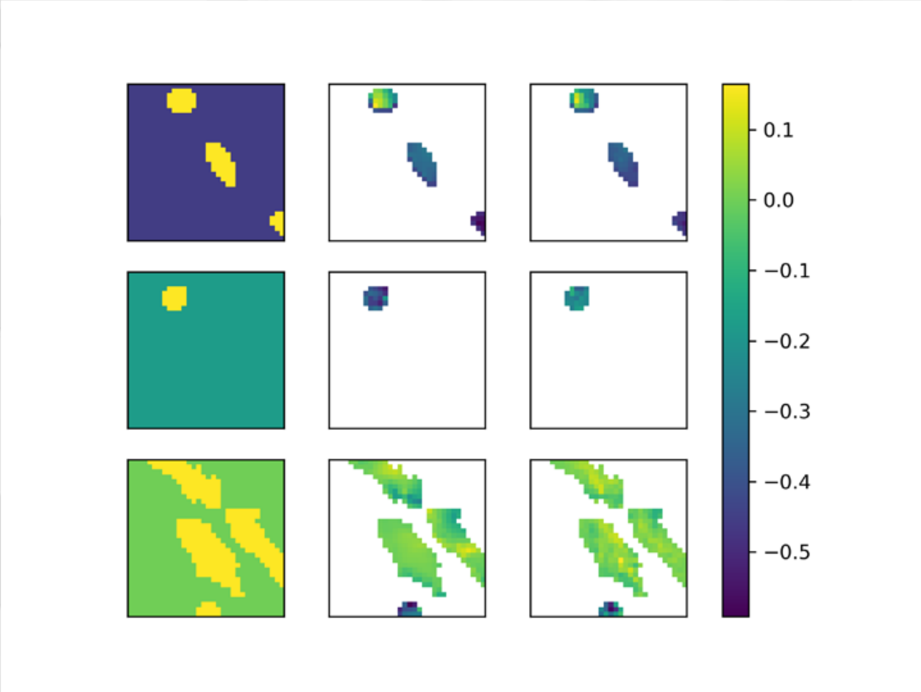
Deep Learning for Material Science: DNN-based Processing-Structure-Performance Map for Fibre Reinforced Polymer

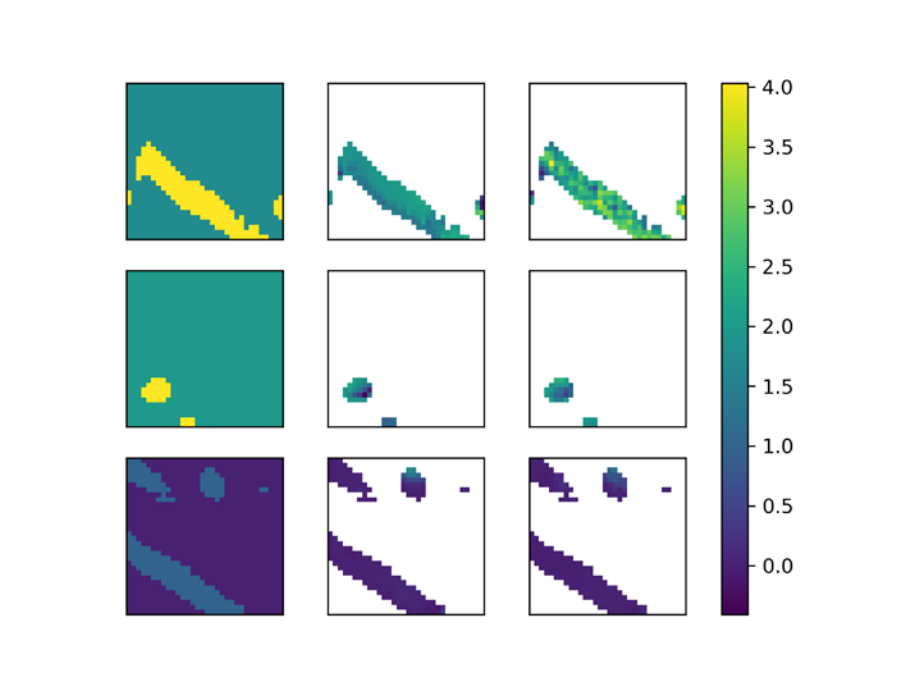


Results

Models	
Hydraulic Stress Field	0.33

Deep Learning for Material Science: DNN-based Processing-Structure-Performance Map





Three columns are: microstructure, true stress field, predicted stress field

Pacific Northwest NATIONAL LABORATORY

Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

- Uncertainty Quantification for Deep Learning

Generalized Polynomial Chaos - gPC

$$T(x,t;\omega)=\sum_{j=0}^{\infty}T_j(x,t)\Phi_j(\xi(\omega))$$

- Polynomials of random variable $\xi(\omega)$
- Orthogonality : $\langle \Phi_i \Phi_j \rangle = \langle \Phi_i^2 \rangle \delta_{ij}$

$$\langle \mathrm{f}(\xi)\mathrm{g}(\xi)\rangle = \int \mathrm{f}(\xi)\mathrm{g}(\xi)\mathrm{W}(\xi)\mathrm{d}\xi$$

$$\langle f(\xi)g(\xi)\rangle = \sum_i f(\xi_i)g(\xi_i)w(\xi_i)$$

- Weight function determines underlying random variable *(not necessarily Gaussian)*
- Complete basis from *Askey scheme*
- Each set of basis converges in L^2 sense

Continuous Cases

Polynomials	Distribution of
Hermite	Gaussian
Laguerre	Gamma
Jacobi	Beta
Legendre	Uniform

Discrete Cases

Polynomials	Distribution of
Charlier	Poisson
Krawtchwouk	Binomial
Meixner	Negative binomial
Hahn	Hypergeometric

Xiu & Karniadakis SIAM J. Sci. Comput. 24(2) (2002)

Implementation of gPC method

$$L(x,u;\xi) = f(x)$$

- *Galerkin Projection:*
  - PC expansion: $u = \sum_{|\alpha|=0}^{p} u_\alpha \phi_\alpha$
  - Residual: $R(\xi) = L(x, \sum_{|\alpha|=0}^{p} u_\alpha \phi_\alpha) - f(x)$
  - Deterministic system of u_α: $\mathbf{E}[R(\xi)\phi_\beta(\xi)] = 0,\ |\beta| \le p$

- *Collocation Projection:*
  - Interpolation operator: $\{\xi^{(i)}\}_{i=1}^{Ng}$ a set of grid points in parameter space.
  - Deterministic system on grid points: $L(x,u;\xi^{(i)}) = f(x)$
  - Choices of grid points: *full tensor products of Gauss quadrature points – $O(N^M)$; sparse grids – $O(N\log(N)^{M-1})$*

Computational Speed-Up

Lucor & Karniadakis, Generalized Polynomial Chaos and Random Oscillators
Int. J. Num. Meth. Eng., vol. 60, 2004

PDF	Error (mean)	Monte-Carlo: M	GPC: (P+1)	**Speed-Up**
Gaussian	2% 0.8% 0.2%	350 2,150 33,200	56 120 220	6.25 18 151
Uniform	0.2% 0.018% 0.001%	13,000 1,580,000 610,000,000	10 20 35	1,300 79,000 17,430,000

Advantage of gPC

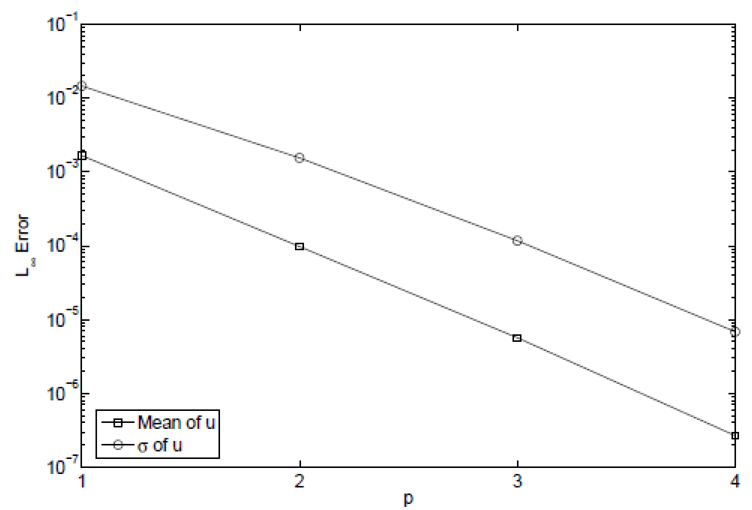
- Fast convergence due to spectral expansion.
- Efficiency due to orthogonality.

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u &= -\nabla p + \nu (1 + \delta \xi) \nabla^2 u \\ \nabla \cdot u &= 0 \end{aligned}$$

<u>Kovasznay Flow:</u>

$$\begin{aligned} u &= 1 - \mathrm{e}^{\lambda x} \cos 2\pi y \\ v &= \frac{\lambda}{2\pi} \mathrm{e}^{\lambda x} \sin 2\pi x \\ \lambda &= \frac{Re(\xi)}{2} - \left(\frac{Re^2(\xi)}{4} + 4\pi\right)^{1/2} \end{aligned}$$

ξ : random variable of Beta(1,1).

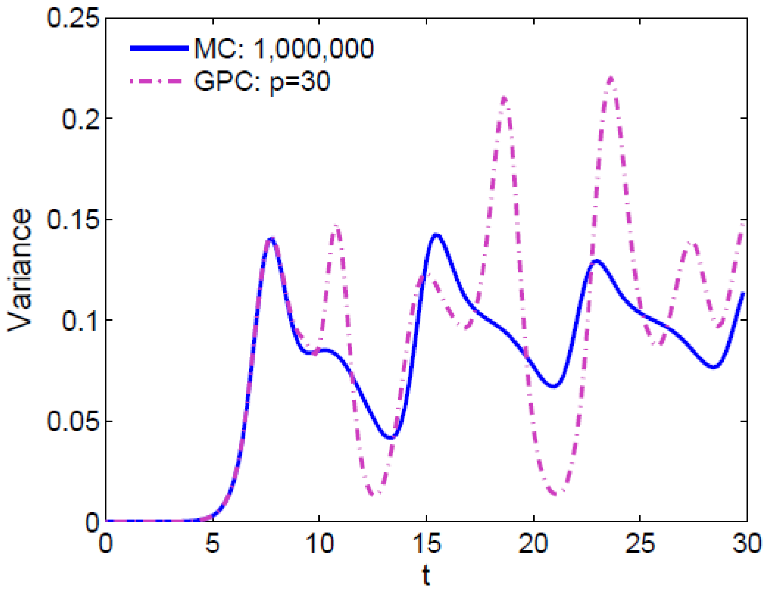


Limitations of gPC

- Inefficient for problems with low regularity in the parametric space.
- May diverge for long-time integrations.

Kraichnan-Orszag three-mode model:

$$
\begin{cases}
\frac{\mathrm{d}Y_1}{\mathrm{d}t} = Y_2 Y_3 \\
\frac{\mathrm{d}Y_2}{\mathrm{d}t} = Y_1 Y_3 \\
\frac{\mathrm{d}Y_3}{\mathrm{d}t} = -2 Y_2 Y_3 \\
\text{random intial conditions.}
\end{cases}
$$



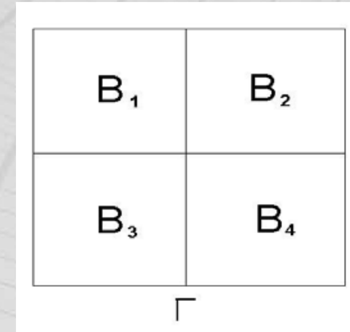
Outline:

- UQ for Complex systems: its challenge and open issues
  - **UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)**
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

- Uncertainty Quantification for Deep Learning

Open Issue 1: Parametric Discontinuities/Bifurcations - Multi-Element Probabilistic Collocation Method (ME-PCM)

- Decompose Γ into non-overlapping elements B^i

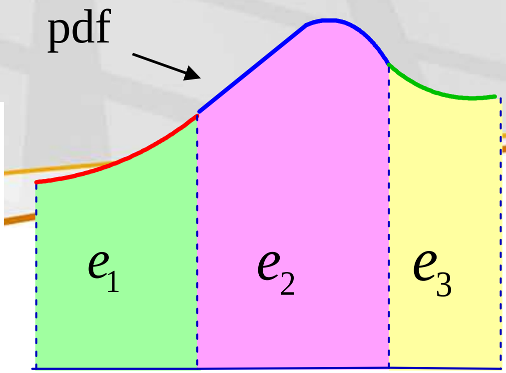


- Define $A_k = \mathbf{Y}^{-1}(B^k)$

- Define new random variable $\eta_k : A_k \to B_k$ on the restricted space $(A_k, \mathcal{F} \cap A_k, P(\cdot|A_k))$ with conditional PDF $\hat{\rho}(x|A_k) = \dfrac{\rho(x)}{P(A_k)}$

- Numerically reconstruct local polynomial chaos basis on each element, orthogonal with respect to $\hat{\rho}$

- Perform PCM on each element. No C^0 requirement on boundaries (measure 0).



$$\tilde{u}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N_e} \mathcal{I}_{B^i} u_k(\boldsymbol{x}, \boldsymbol{y}) \mathbb{I}_{\{\boldsymbol{y} \in B^i\}} \quad \forall \boldsymbol{x} \in \overline{D},\ \forall \boldsymbol{y} \in \Gamma$$

choose spatial discretization method

Lin et al., CiSE, 9(2) 21-29, 2007

Outline:

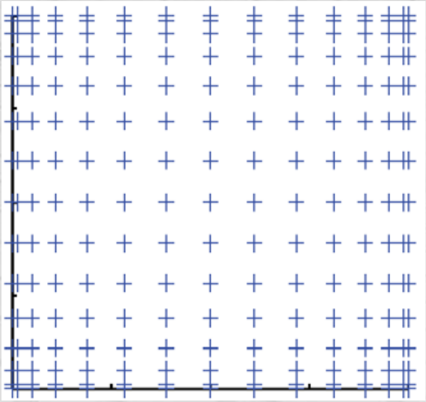
- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - **UQ open issue 2: Curse of Dimensionalities** (Sparse grid, Adaptive ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

- Uncertainty Quantification for Deep Learning

Open Issue 2: Curse of Dimensionality – Multi-Element Probabilistic Collocation Method (ME-PCM)

Choice of N-dimensional approximation operator: $\mathcal{I}_{B^i}$

tensor product



- Tensor product Lagrangian interpolation

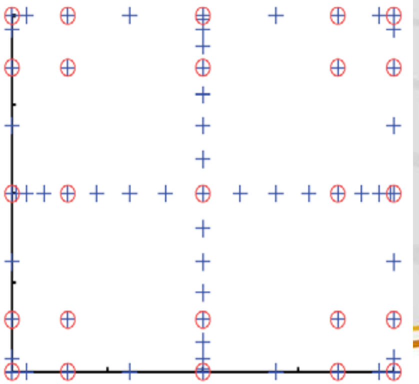
interpolation orders

$$L^{\mathbf{P}}_{B^i} u_k(\boldsymbol{x}, \boldsymbol{y}) = \sum_{m=1}^{r} u_k(\boldsymbol{x}, \boldsymbol{q}_m) \cdot l_m(\boldsymbol{y})$$

interpolation points

- Smolyak sparse grid approximation (Smolyak, 1963)

sparse



$$\mathcal{V}^i_j(v) = \sum_{m=1}^{n_i} v(y^i_m) \cdot a^i_m$$

1D interp. rule in dimension i

$$\mathcal{S}_{B^i}(s) = \sum_{s-N+1 \le |\mathbf{i}| \le s} (-1)^{s-|\mathbf{i}|} \begin{pmatrix} N-1 \\ s-|\mathbf{i}| \end{pmatrix} \cdot (\mathcal{V}^{i_1}_1 \otimes \cdots \otimes \mathcal{V}^{i_N}_N)$$

Pacific Northwest
NATIONAL LABORATORY

interpolatory for nested 1D rules

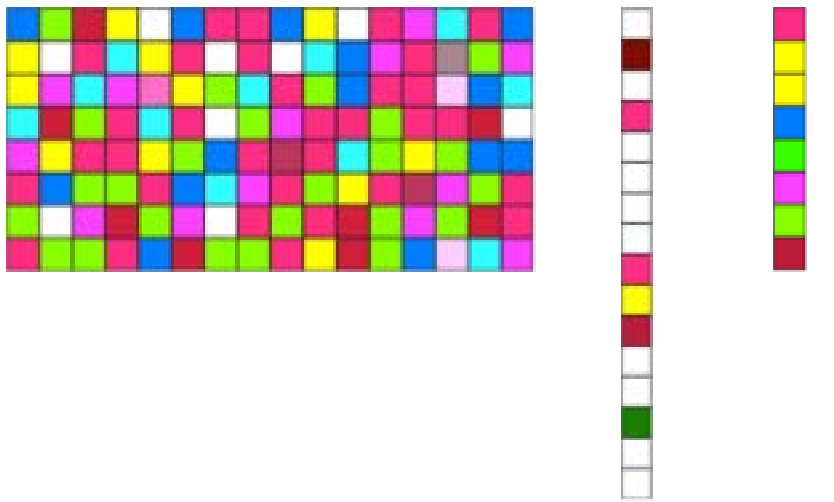
Compressive sensing for gPC expansion

$$\boxed{M < N}$$

measurement matrix *signal* *observation*

$\boldsymbol{\Psi}$ $\boldsymbol{c}$ $\boldsymbol{u}$

$M \times N$ $N \times 1$ $M \times 1$



$$(P_{h,\epsilon}): \quad \arg\min_{\hat{\boldsymbol{c}}} \|\hat{\boldsymbol{c}}\|_h, \quad \text{subject to} \quad \|\boldsymbol{\Psi}\hat{\boldsymbol{c}} - \boldsymbol{u}\|_2 \le \epsilon$$

- H. Lei, X. Yang, B. Zheng, **G. Lin**, N. Baker, Constructing Surrogate Models of Complex Systems with Enhanced Sparsity: Quantifying the Influence of Conformational Uncertainty in Biomolecular Solvation, *SIAM Multiscale Modeling and Simulation*, 13(4): 1327-1353, 2016.
- X. Yang, H. Lei, N. Baker, **G. Lin***, Enhancing sparsity of Hermite polynomial expansions by iterative rotations, Journal of Computational Physics, 307: 94-09, 2016.

Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - **UQ open issue 3: Heterogeneous big data & Computational Expensive Models** - Bayesian parameter estimation in large-scale regional and global climate models

Pacific Northwest
NATIONAL LABORATORY

- Uncertainty Quantification for Deep Learning

Uncertainty Quantification and Bayesian Parameter Estimation in Convective Cloud scheme using Large-Scale, Heterogeneous Data

- Downdraft Rate
- Entrainment Rate
- CAPE Consumption Time
- TKE for Shallow Convection
- Starting Height of Downdraft

▶Five Parameters

GPCP daily precipitation → (UW) Daily 1/8-degree gridded meteorological data (Maurer et al., 2002)

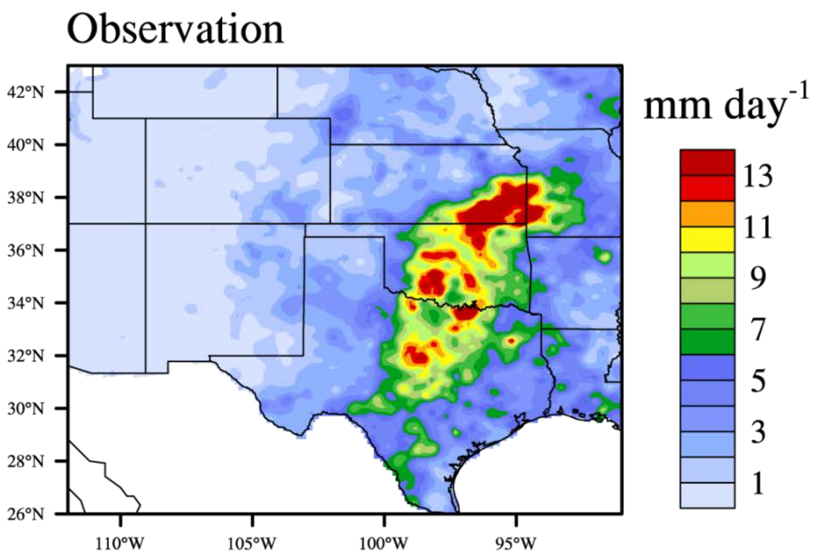
Impact of Optimization on Other Variables (added 2-m air temperature, 10-m wind speed)

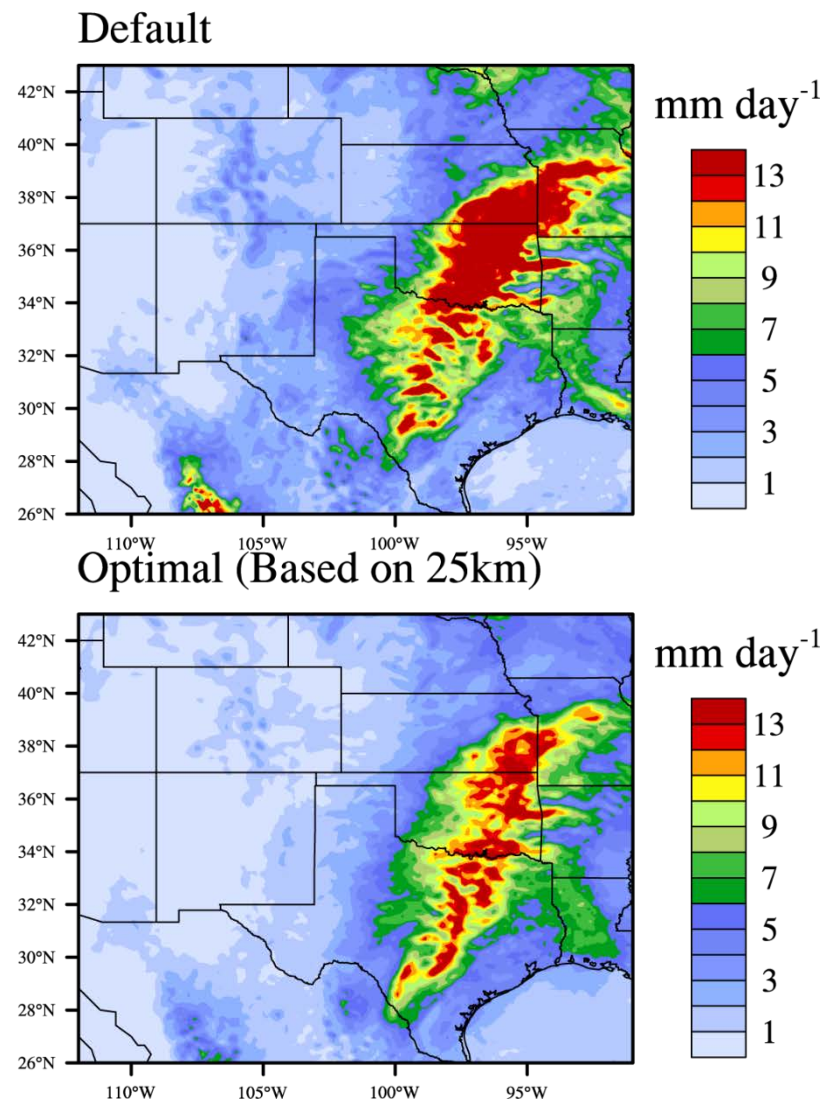
▶Observational Constrain

▶ Yang, B., Qian, Y., Lin, G., Leung, R., and Zhang, Y., Atmos. Chem. Phys. 11, 31769-31817, doi:10.5194/acpd-11-31769-2011, 2011.

Bayesian Parameter Estimation of Convection Scheme on Regional Climate Model using SAA

Total Precipitation in June 2007

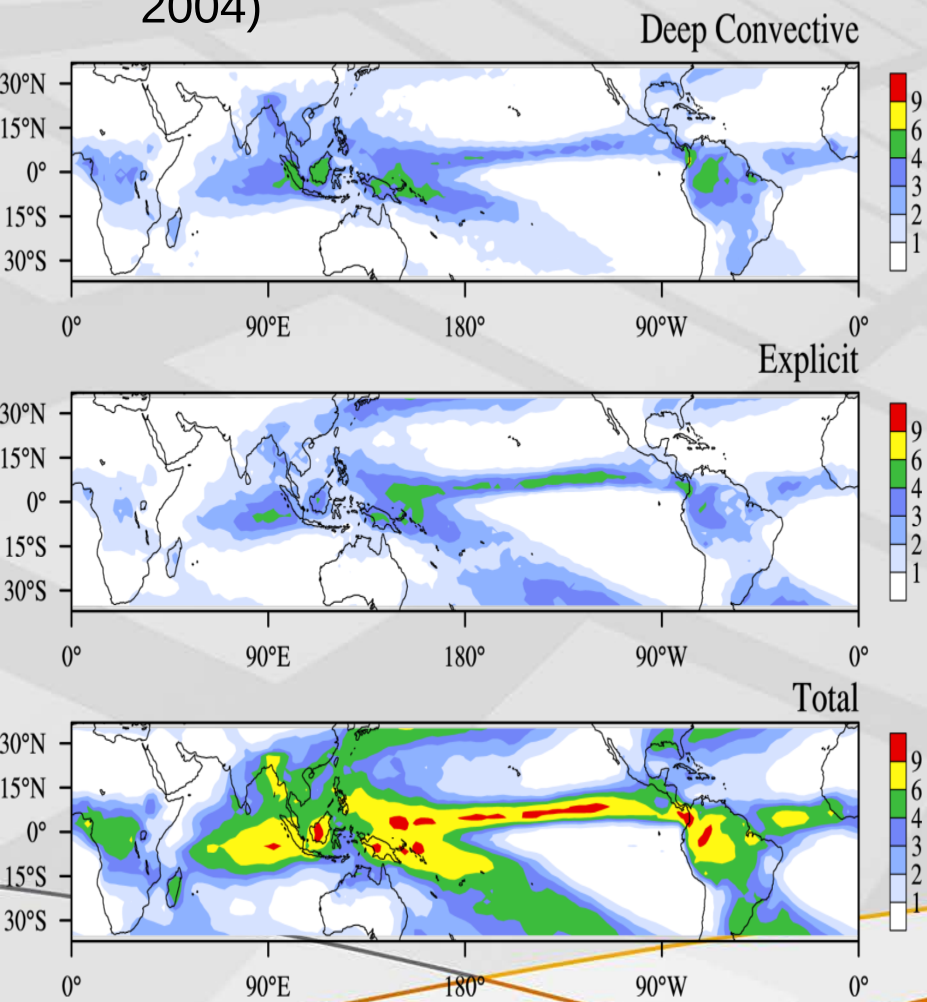




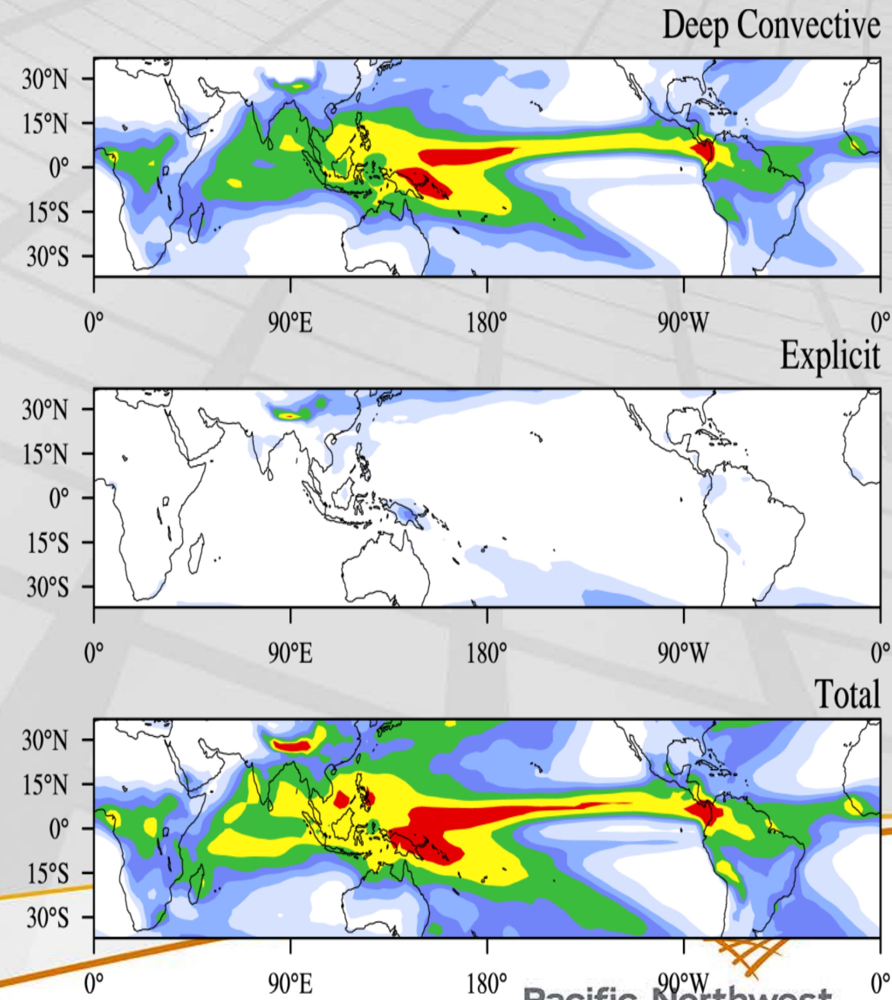
Atmos. Chem. Phys, doi:10.5194/acp-12-2409-2012, 2012

Optimal Parameter Estimation in Community Atmosphere Models - Motivation on Parameter Tuning in Deep Convective Precipitation

Observed Precipitation (TRMM & GPCP, 2001-2004)



Simulated precipitation (Standard CAM5.1)



Methodology: Selected 12 parameters in ZM scheme

Blue: Ocean Red: Land

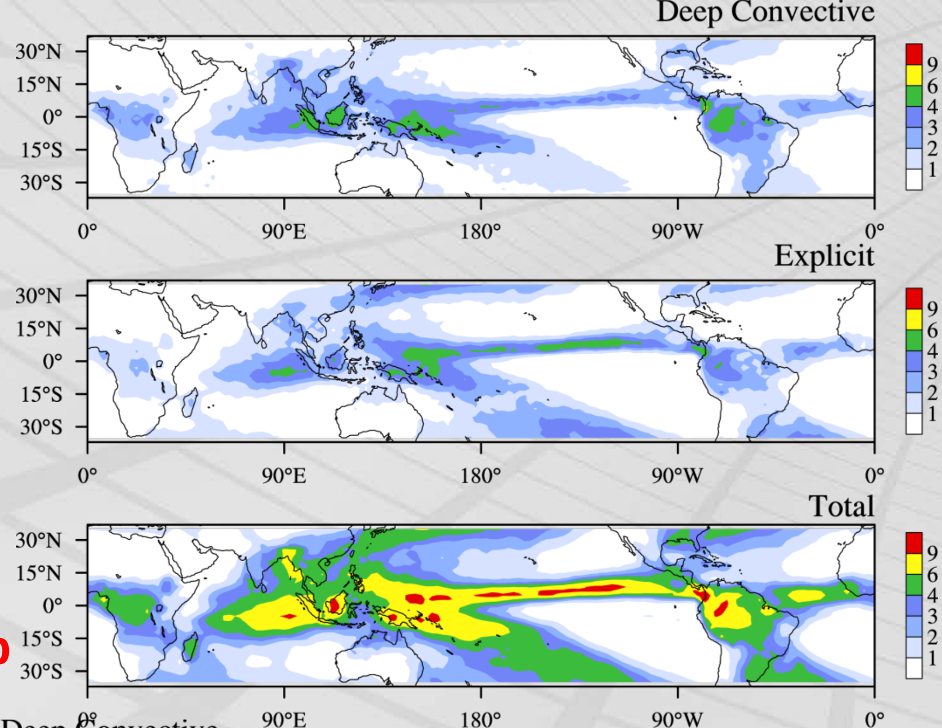
Parameter	Default	Minimum	Maximum	Description
c0	45E-3 5.9E-3	1E-3 1E-3	20E-3 20E-3	Deep convection precipitation efficiency
dmpdz	-1.0E-3 -1.0E-3	-2.0E-3 -2.0E-3	0 0	Parcel fractional mass entrainment rate
Tau	3600 3600	1800 1800	14400 14400	Consumption time scale
Capelmt	70	20	200	Threshold value for CAPE for deep convection
ke	1.0E-6	0.5E-6	10E-6	Evaporation efficiency parameter
alfa	0.1	0.05	0.6	Initial cloud downdraft mass flux
edratio	2	1	3	Ratio of downdraft entrainment to updraft
dsliq	8	4	24	Radius of detrained liquid from convection
dsice	25	10	50	Radius of detrained ice from convection

Bayesian Parameter Estimation of Deep Convection Scheme in Global Climate Model

2001-2004

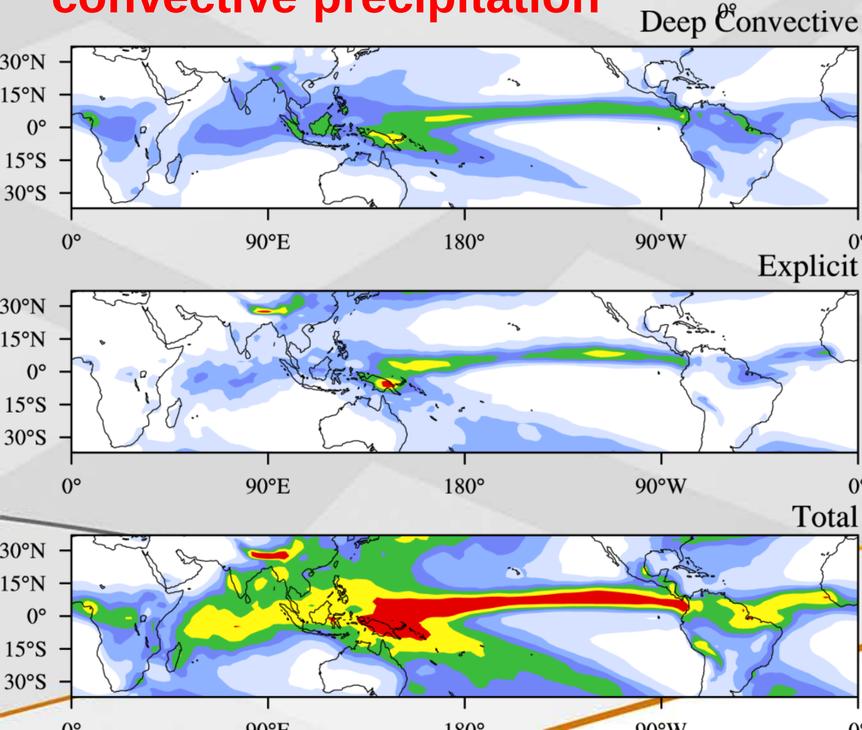
J. Geophys. Res.,
doi:10.1029/2012JD018213

TRMM/GPCP

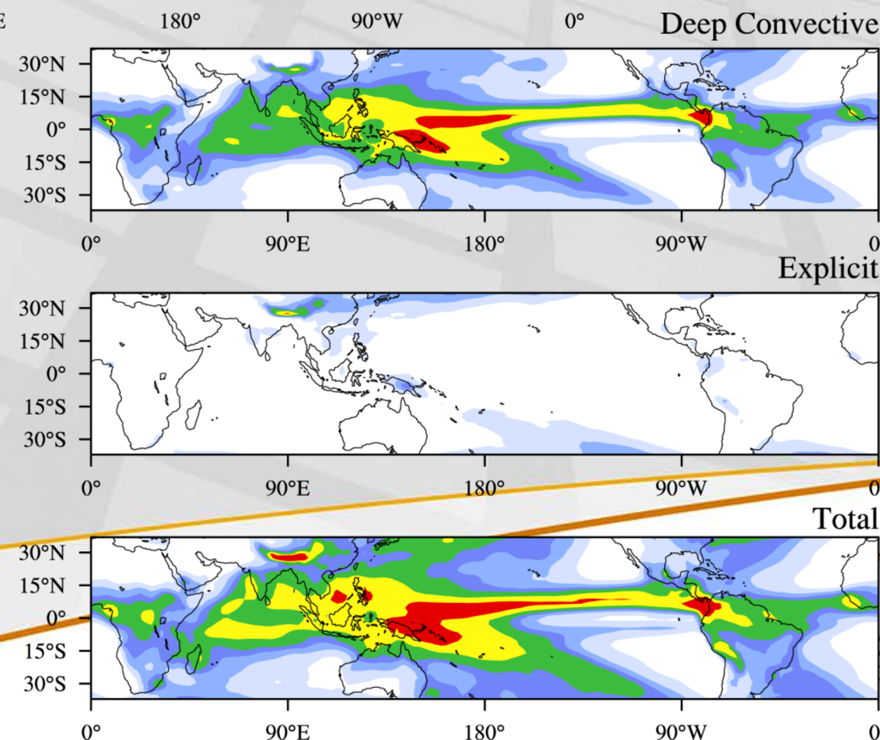


4-year mean precipitation (mm/day, 2001-2004)

Optimal by matching deep convective precipitation



Default



Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

Uncertainty Quantification for Deep Learning

ConvPDE-UQ: Quantify the uncertainties in deep learning of PDE solutions in arbitrary domain

Goal: Replace FEM solver and achieve real-time prediction of PDE solutions and quantify the uncertainty

N. Winovich, K. Ramani, **G. Lin***, ConvPDE-UQ: Fast convolutional encoder-decoder networks with quantified uncertainty for heterogeneous elliptic partial differential equations on varied domains, Journal of Computational Physics, in press, 2019.

Overview of Problem Setups

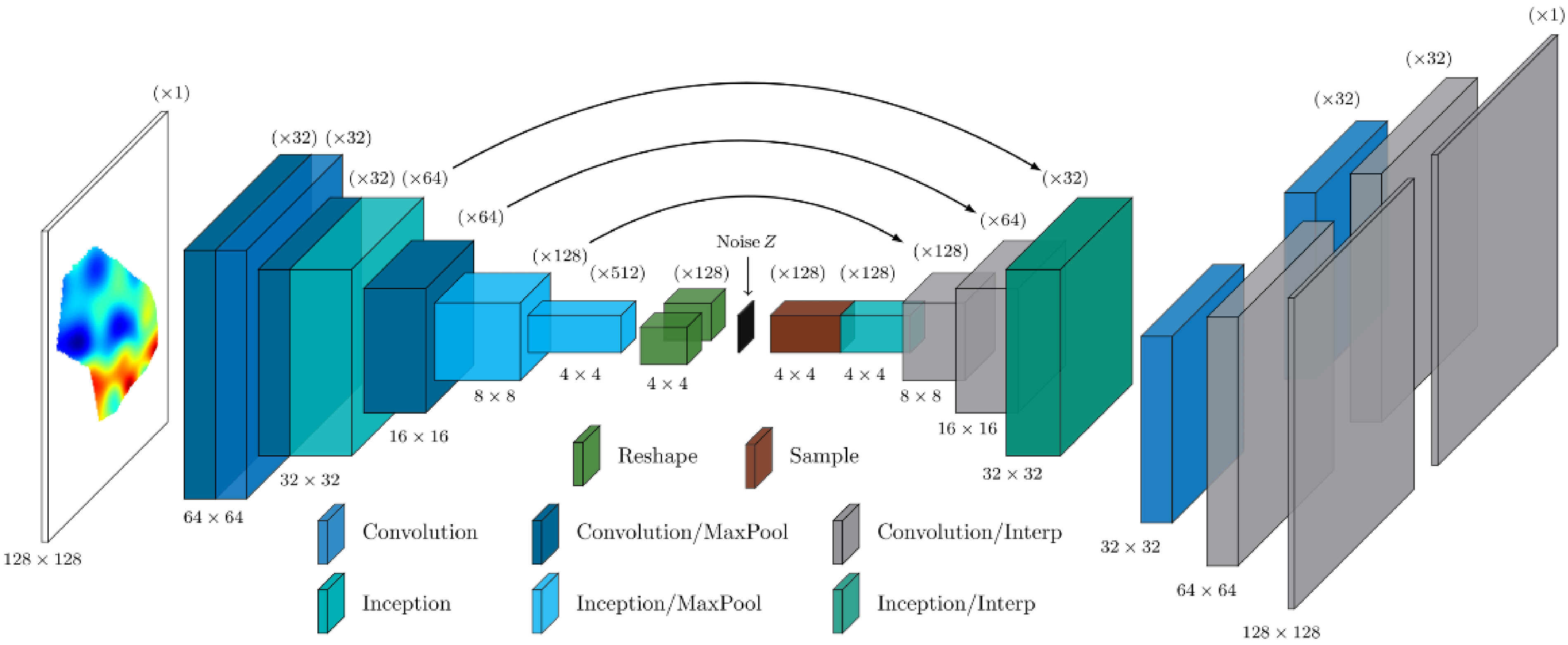
Poisson on Circle	Varied Domain	Nonlinear Poisson
$\Delta u = f$ on fixed disk D	$\Delta u = f$ on varying domain Ω	$\operatorname{div}\left(\left(1+\lvert u\rvert^2\right)\cdot\nabla u\right) = f$ on varying domain Ω

- Homogeneous Dirichlet boundary condition: $u = 0$ on $\partial\Omega$

$$\begin{cases} \mathcal{L}u = f & \text{in} \quad \Omega \\ u = 0 & \text{on} \quad \partial\Omega \end{cases}$$

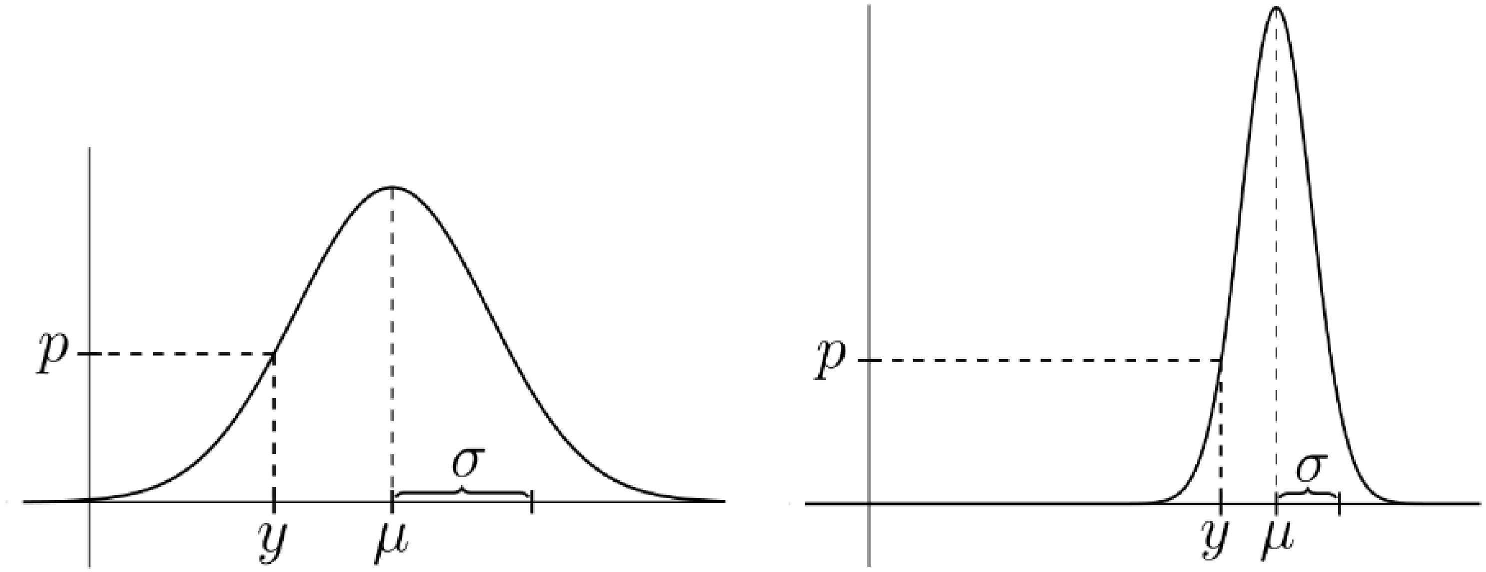
Goal: Find the solution u given a domain Ω and a *source term* f.

Network Architecture



- The output of the network consists of two channels: one channel for the pointwise mean predictions $\mu[i,j]$ and another channel for the predicted pointwise log standard deviations $\log \sigma[i,j]$.

Probabilistic Predictions

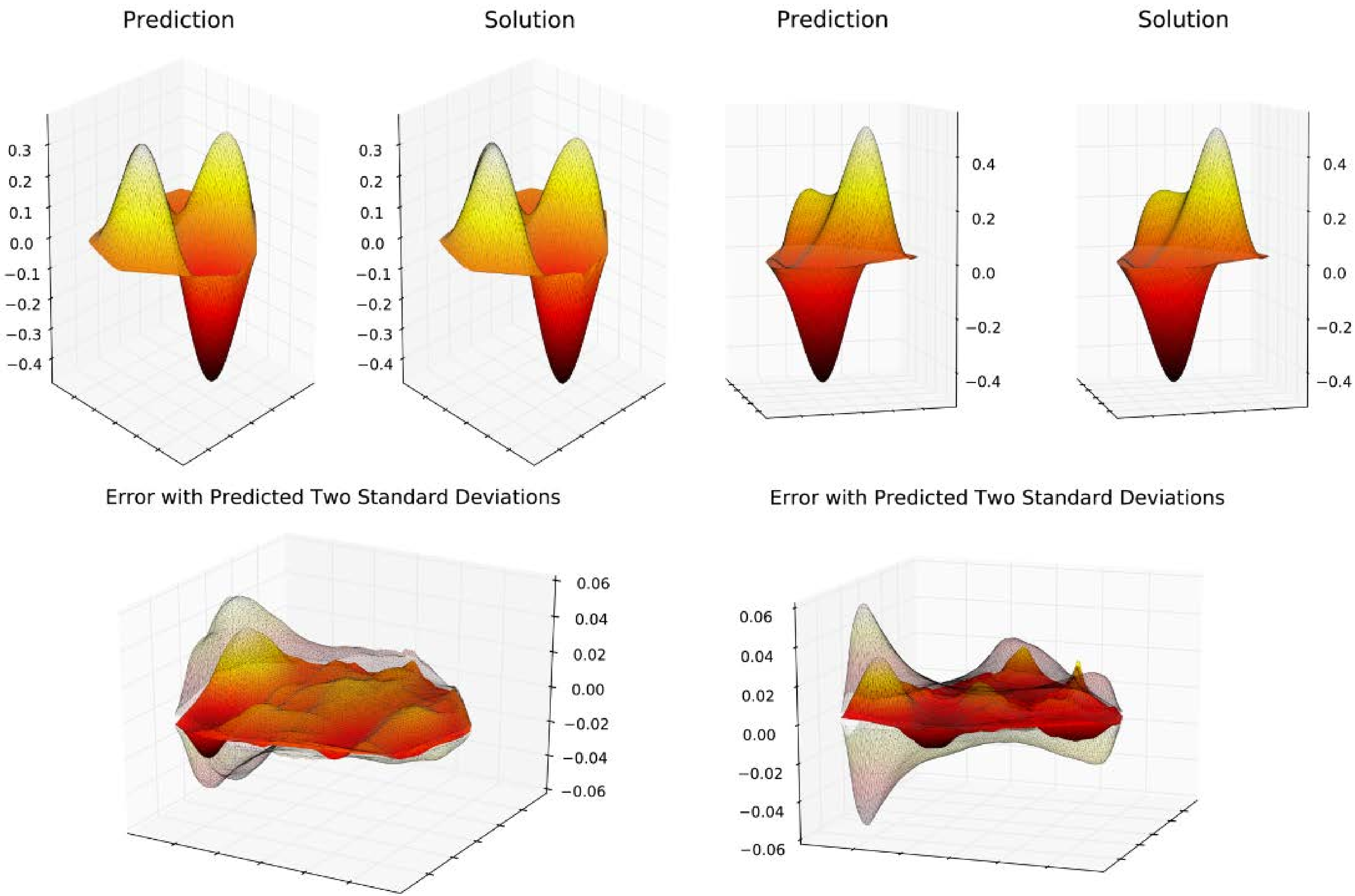


- The probabilistic prediction framework allows the network to begin with coarse, low-confidence predictions (left) and to gradually build confidence by lowering the predicted standard deviations.
- High confidence predictions (right) allow the network to attain far lower losses when correct, but have steep drop-offs which severely penalize any inaccuracy in the network's predictions.

Numerical Results

Problem	Model	L^1 Relative Error		L^2 Relative Error	
Poisson on Circle	**Probability**	$\mathbf{9.19e{-}3}$	$\mathbf{1.00e{-}2}$	$\mathbf{1.18e{-}4}$	$\mathbf{1.50e{-}4}$
	MSE ($\lambda = 0.1$)	$1.23e{-}2$	$1.28e - 2$	$2.60e{-}4$	$3.06e{-}4$
	MSE ($\lambda = 0.0$)	$1.23e{-}2$	$1.29e - 2$	$2.48e{-}4$	$2.90e{-}4$
Varying Domain	**Probability**	$\mathbf{1.82e{-}2}$	$\mathbf{2.11e{-}2}$	$\mathbf{1.21e{-}3}$	$\mathbf{1.45e{-}3}$
	MSE ($\lambda = 0.1$)	$3.43e{-}2$	$3.57e{-}2$	$2.25e{-}3$	$2.62e{-}3$
	MSE ($\lambda = 0.0$)	$3.60e{-}2$	$3.75e{-}2$	$2.43e{-}3$	$2.86e{-}3$
Nonlinear Poisson	**Probability**	$\mathbf{1.94e{-}2}$	$\mathbf{2.24e{-}2}$	$\mathbf{1.32e{-}3}$	$\mathbf{1.58e{-}3}$
	MSE ($\lambda = 0.1$)	$3.21e{-}2$	$3.46e{-}2$	$1.84e{-}3$	$2.46e{-}3$
	MSE ($\lambda = 0.0$)	$3.37e{-}2$	$3.61e{-}2$	$2.09e{-}3$	$2.69e{-}3$

Qualitative Results



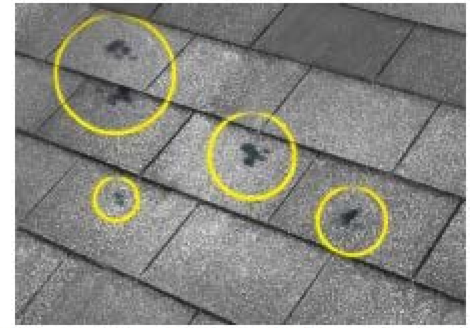
Peri-Net: Deep learning of material failure and fracture propagation

Goal: Replace computational expensive fracture mechanics solver and achieve real-time prediction of material failure and fracture propagation

M. Kim, N. Winovich, **G. Lin***, W. Jeong, Peri-Net: Analysis of crack patterns using deep neural networks, Journal of peridynamics and nonlocal modeling, in press, 2019.

Why do we need this study?

Disaster (Hail damage)



(Image: www.wikihow.com)

Every day life (Cell phone damage)

(Image: www.instructables.com)

Structure (Bridge damage)

Bridge

Cantilever section

(Image: www.californiabeat.org)

What is the objective of this study?

Take a photo (Crack/Damage)

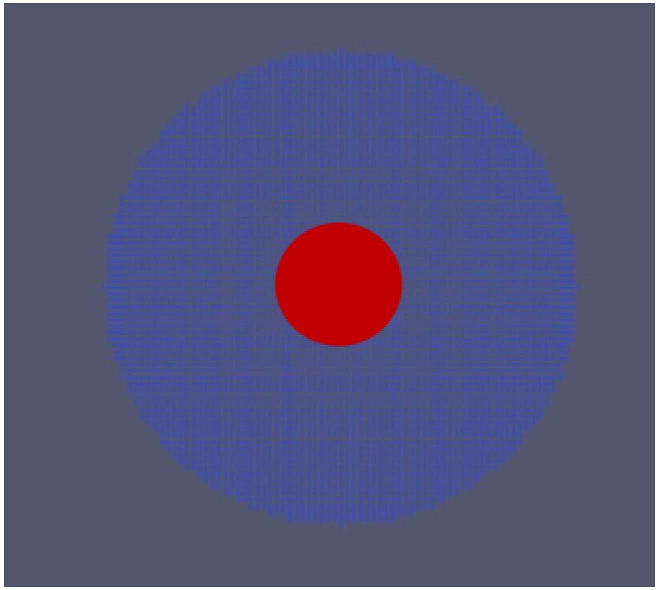

1) Inverse problem

Image → indenter size
velocity, angle etc…

2) Forward problem

Image → prediction
(Crack pattern)

Set up for damage in LAMMPS (Forward problem)



Structure: Disk

Radius: 0.037m

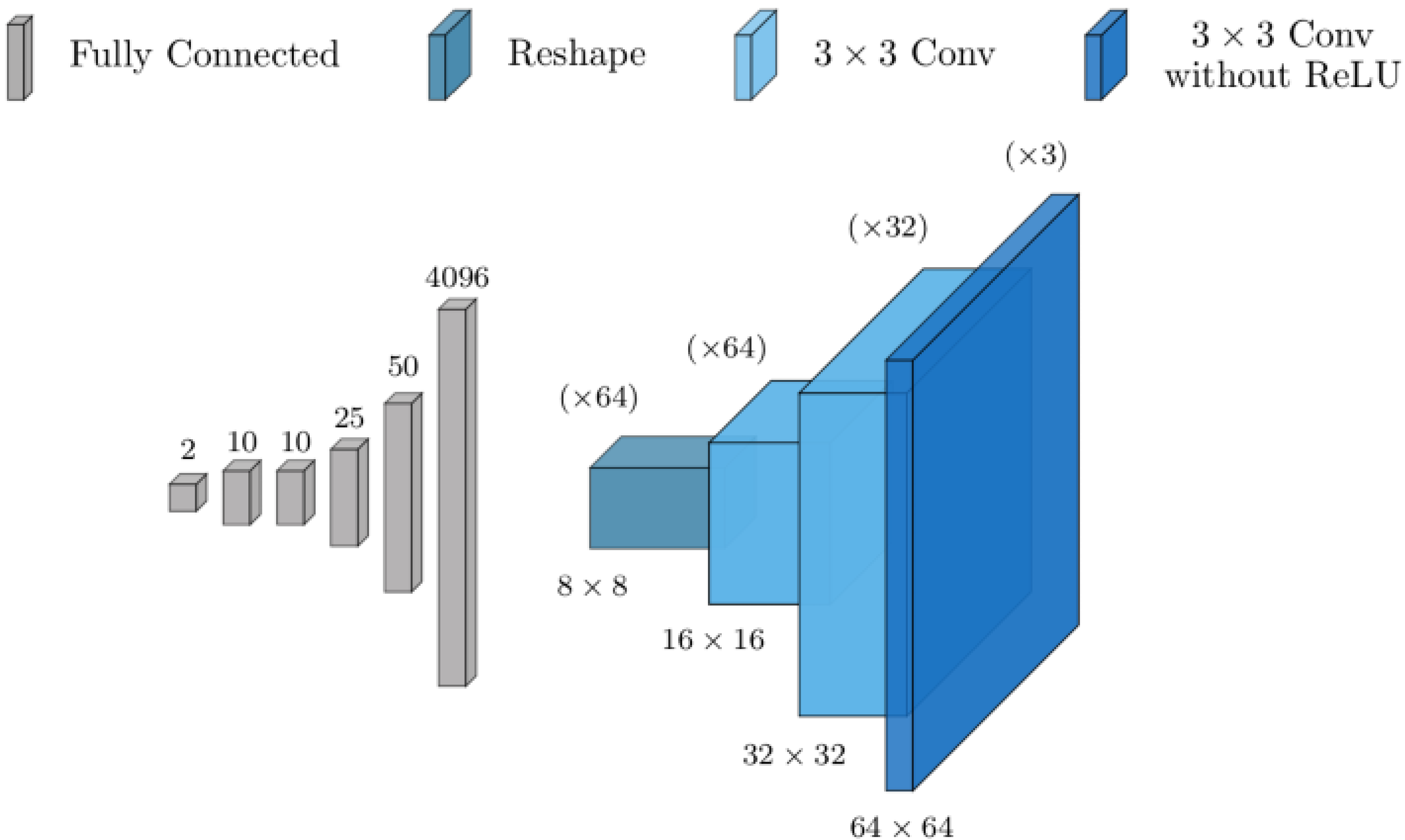
Depth: 0.0025m

1 Variables 2 fixed
1) Radius of indenter (0.0050m) fixed
2) Velocity of indenter (100m/s) fixed
3) Hitting location of disk (1000 hitting points)

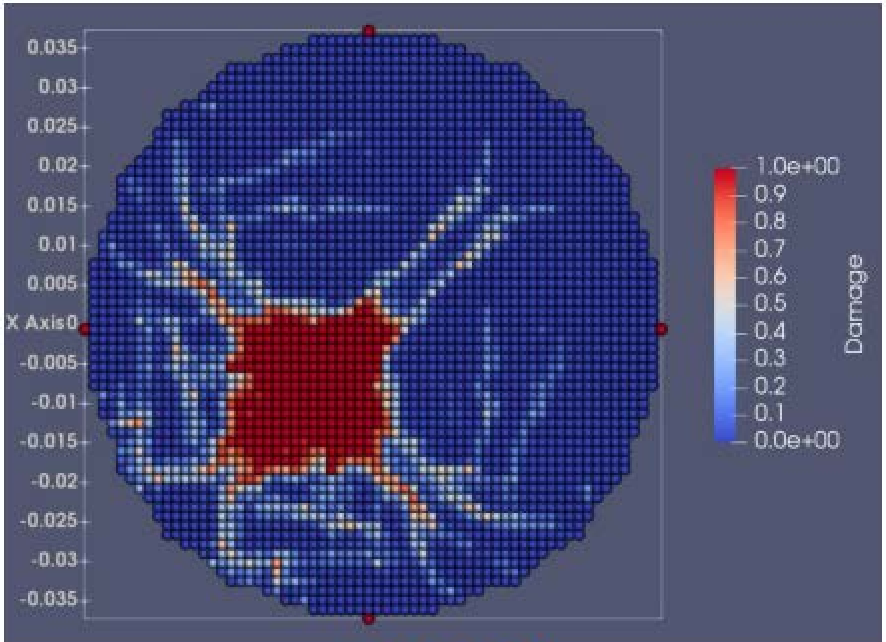
Total data: 1x1x1000=1000 data

Architecture of CNN

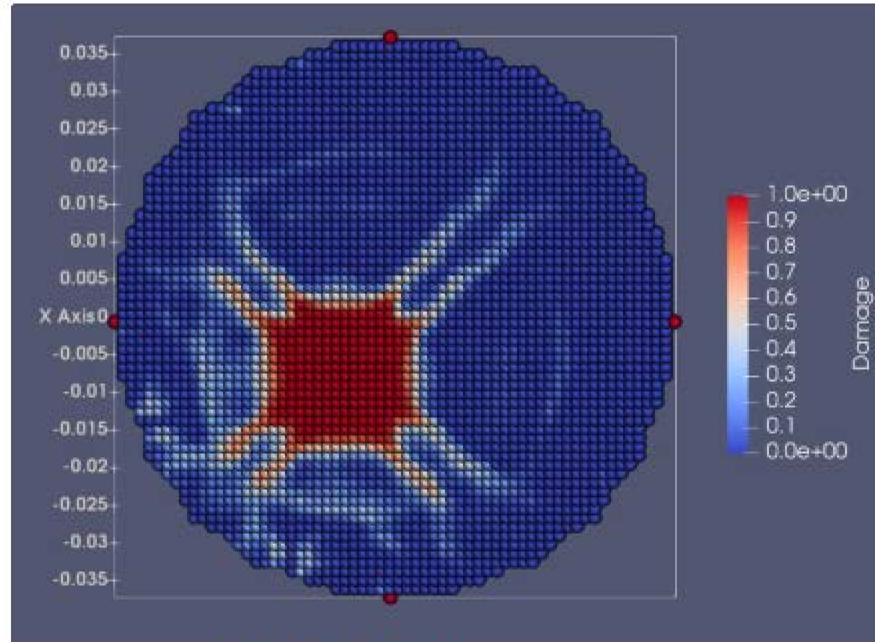
(Train Error (MSE) = 0.0639, Test Error (MSE) = 0.0158)



Result



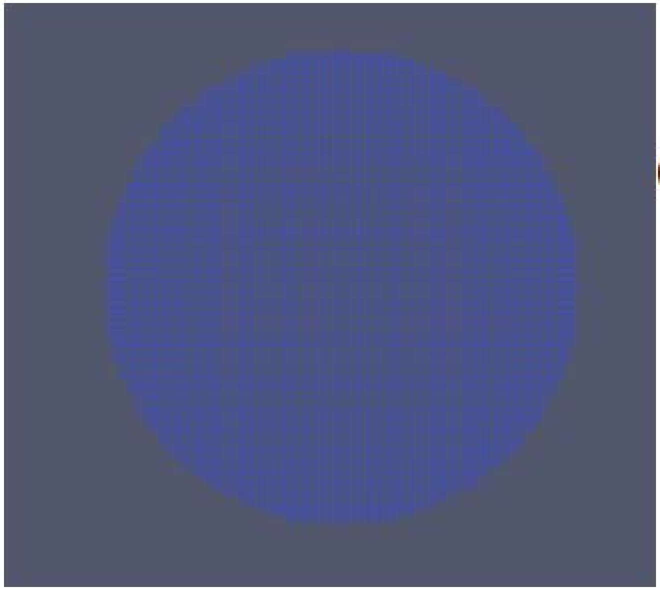
Solution



Prediction

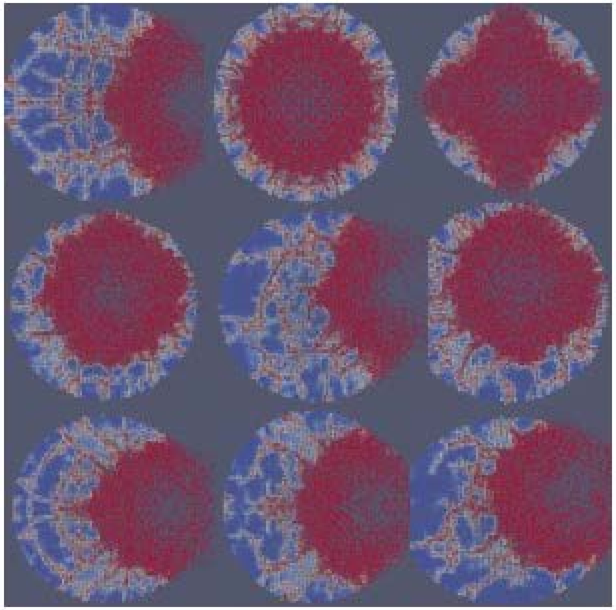
Computation time per one simulation	Peridynamics	CNN
	7.5sec	0.00313sec

Getting Data (Forward problem)



Change 3 variables

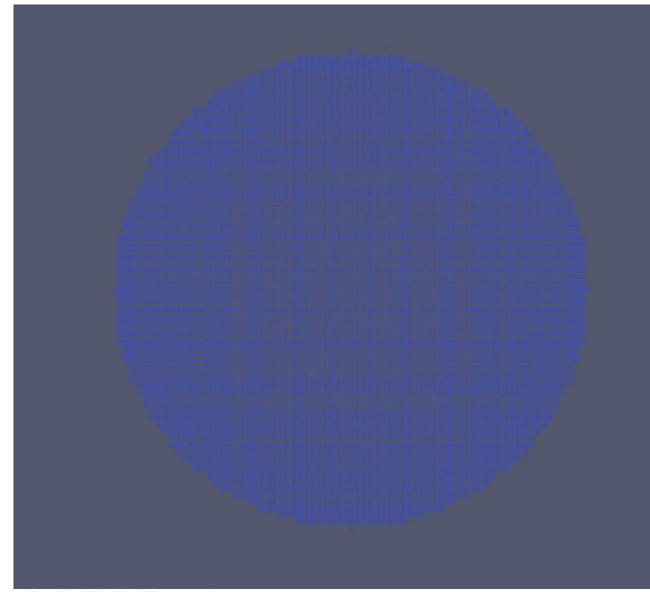
Peridynamics in LAMMPS



.
.
.

15625 data

Set up for damage in LAMMPS (Forward problem)



Structure: Disk

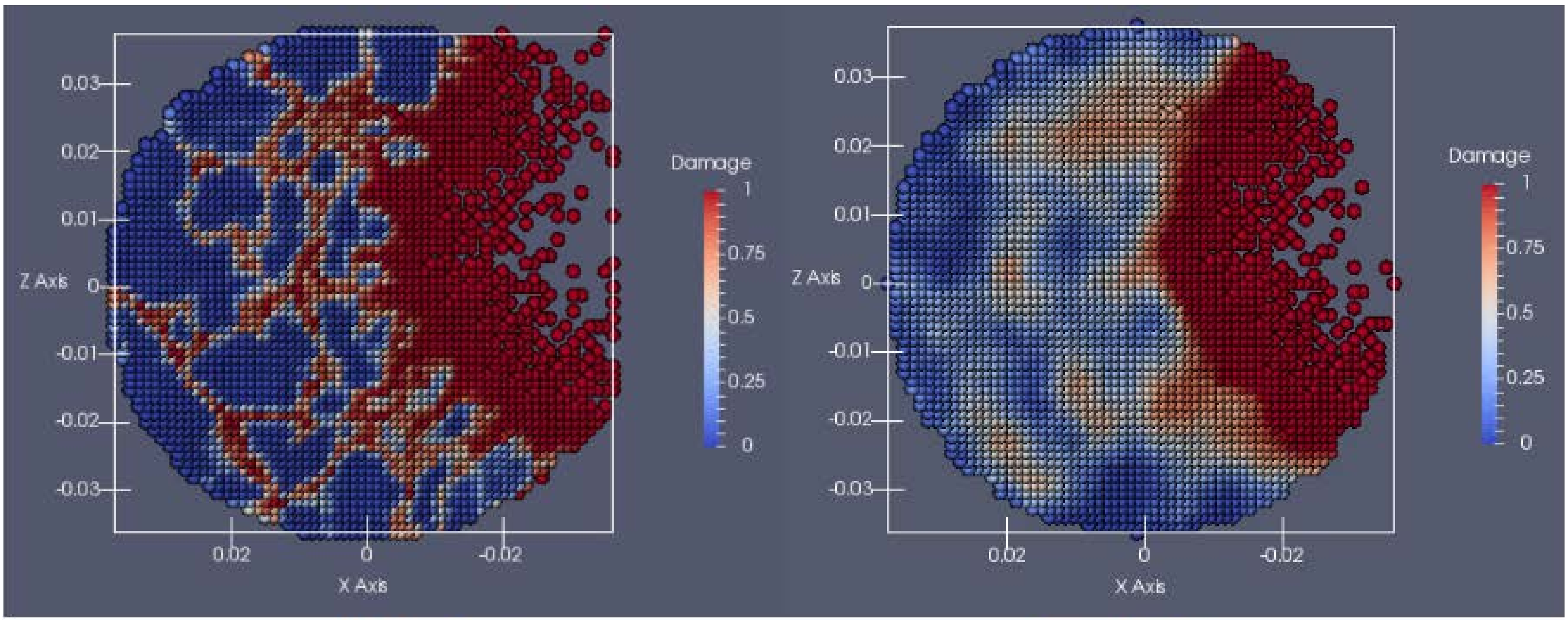
Radius: 0.037m

Depth: 0.0025m

3 Variables
1) Radius of indenter (25 data, 0.0050m - 0.00525m)
2) Velocity of indenter (25 data, 100m/s -102.5m/s)
3) Hitting location of disk (25 hitting points)

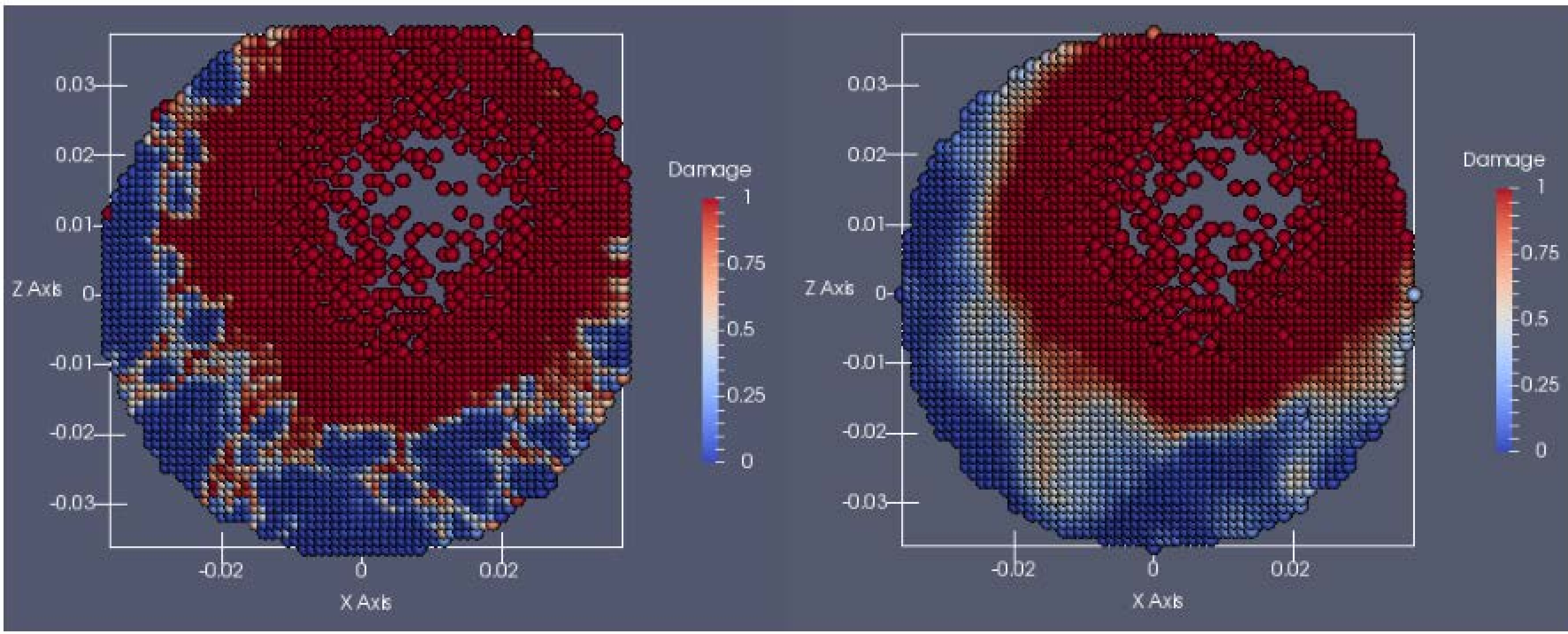
Total data: 25x25x25=15625 data

Results



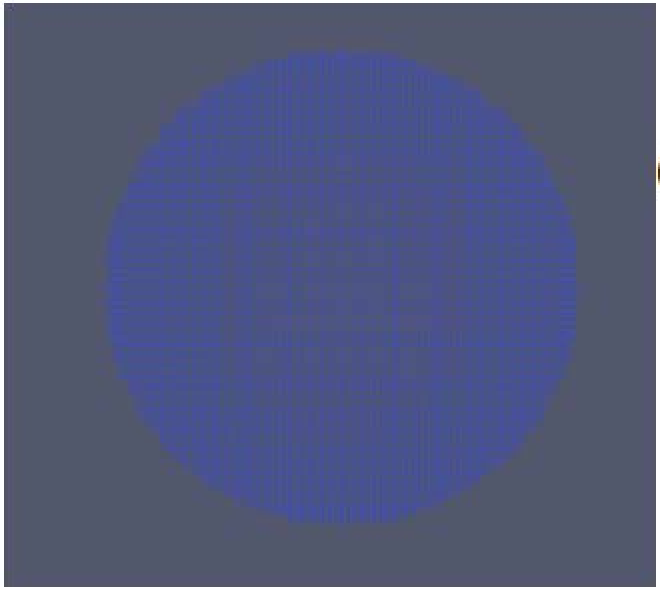
Solution **Prediction**

Results



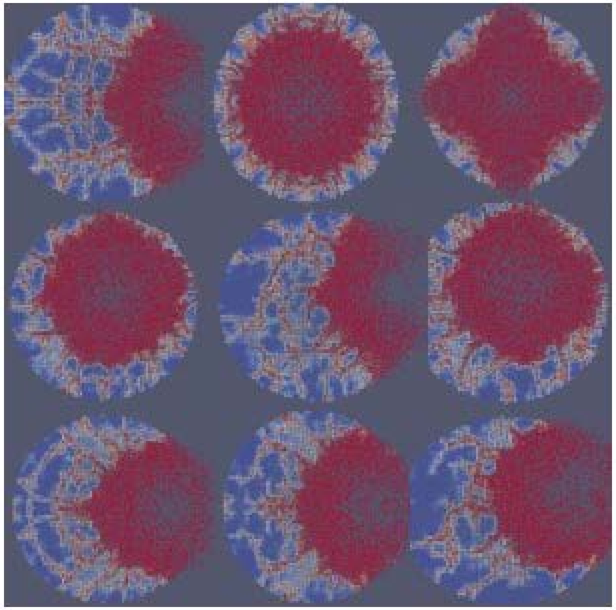
Solution **Prediction**

Getting Data (Inverse problem)



Change 4 variables

Peridynamics
in LAMMPS



.
.
.

7200 data

Set up for damage in LAMMPS (Inverse problem)

- Structure: Disk
- Radius: 0.037m
- Depth: 0.0025m

- 4 Variables

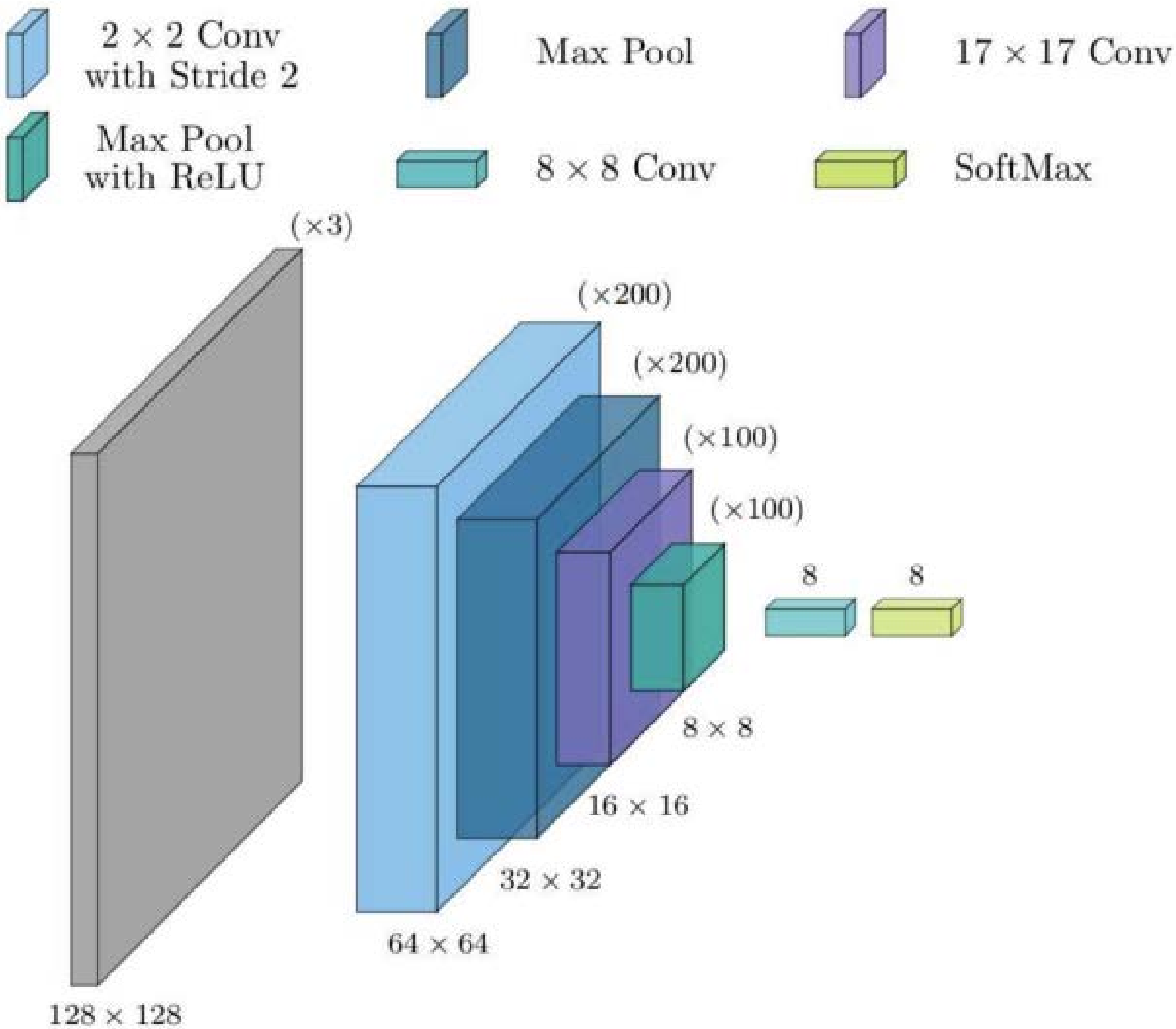
1) Radius of indenter (0.007m, 0.008m)
2) Velocity of indenter (100m/s, 100.1m/s)
3) Angle of indenter (0°, 45°)
4) Hitting location of disk (900 hitting points)

- **Data (labeled all of data)**

  8 Mode (each mode 900data – Training data = 7200, Test data = 1200)

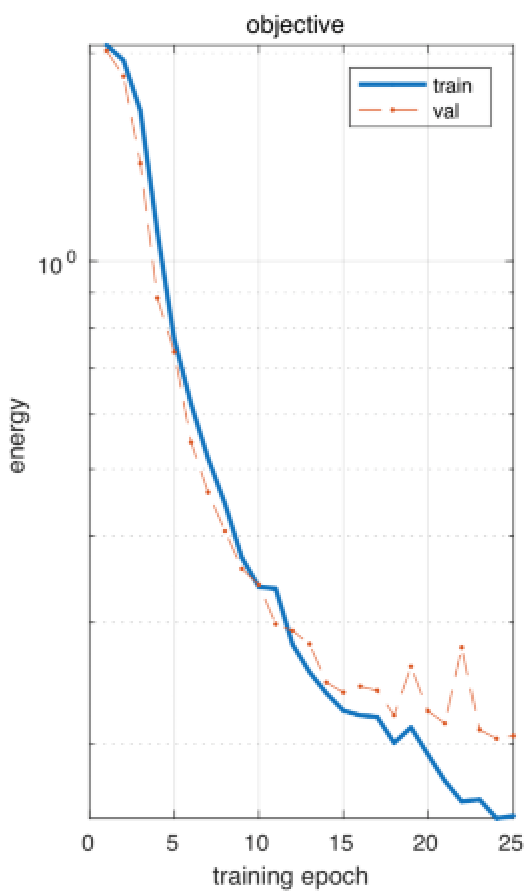
Mode1	**r=0.007m, v=100m/s, 0°**
Mode2	**r=0.007m, v=100.1m/s, 0°**
Mode3	**r=0.008m, v=100m/s, 0°**
Mode4	**r=0.008m, v=100.1m/s, 0°**
Mode5	**r=0.007m, v=100m/s, 45°**
Mode6	**r=0.007m, v=100.1m/s, 45°**
Mode7	**r=0.008m, v=100m/s, 45°**
Mode8	**r=0.008m, v=100.1m/s, 45°**

Architecture of CNN



Results

Success rate: 95.6%
(Train Error (MSE) = 0.0657, Test Error (MSE) = 0.044)



Test image number : Label (Mode)	CNN prediction results (Mode)
Test image 1 : 5 (Mode)	5 (Mode)
Test image 100 : 6 (Mode)	6 (Mode)
Test image 200 : 7 (Mode)	7 (Mode)
Test image 300 : 2 (Mode)	2 (Mode)
Test image 400 : 8 (Mode)	8 (Mode)
Test image 500 : 7 (Mode)	7 (Mode)
Test image 600 : 2 (Mode)	1 (Mode)
Test image 700 : 3 (Mode)	3 (Mode)
Test image 800 : 8 (Mode)	8 (Mode)
Test image 900 : 6 (Mode)	6 (Mode)
Test image 1000 : 7 (Mode)	7 (Mode)
Test image 1100 : 7 (Mode)	7 (Mode)
Test image 1200 : 5 (Mode)	5 (Mode)

Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

Pacific Northwest NATIONAL LABORATORY

- Uncertainty Quantification for Deep Learning

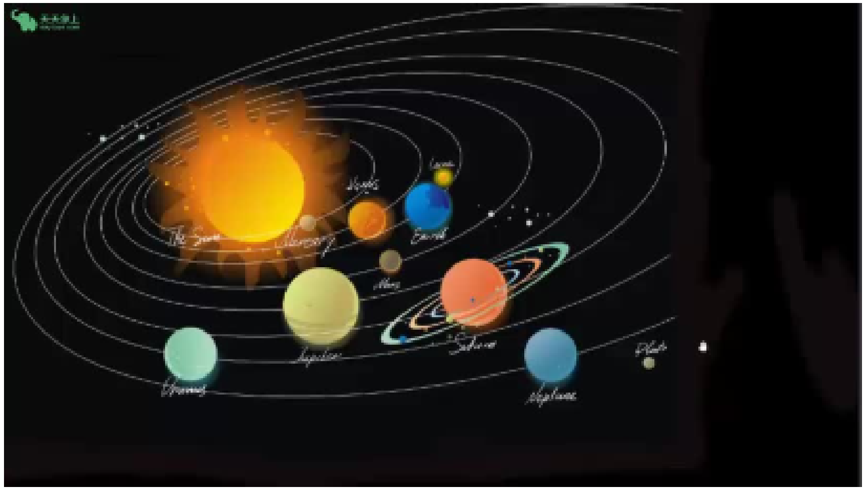
Question: Can we use available observation data to discover the physical laws?

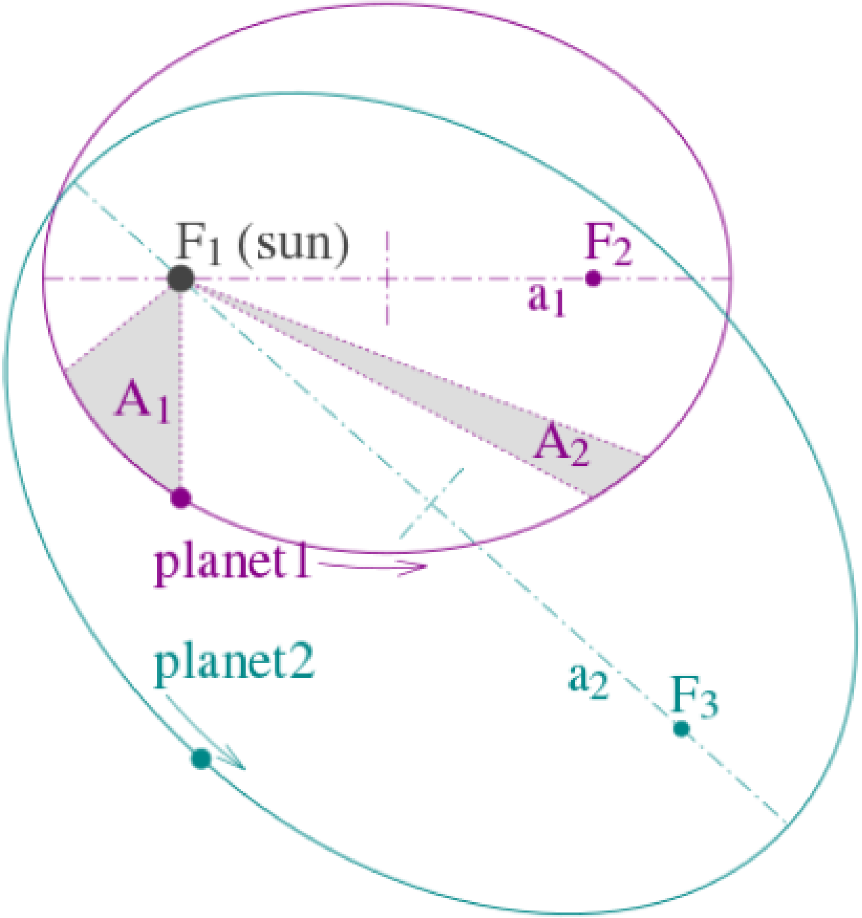
Goal: Enable Data-driven Scientific Discovery?

S. Zhang, **G. Lin**, Robust data-driven discovery of governing physical laws with error bars, Proceedings of the Royal Society of London. Series A, mathematical, physical and engineering sciences, in press, 2018.

Motivation

- For example, Kepler discovered the laws of planetary motion by analyzing observational data.





- Now, we will develop an algorithm that discovers the laws automatically.

General version

- Suppose the basis functions are

$$\mathbf{f} = [f_1(x, y, y', \ldots, y^{(k)}), \ldots, f_M(x, y, y', \ldots, y^{(k)})].$$

- The problem is to find a sparse weight vector $\mathbf{w} = [w_1, \ldots, w_M]^T$ satisfying

$$0 = \mathbf{f}\mathbf{w}.$$

- After collecting the data, we construct the matrix $\mathbf{F}$:

$$\mathbf{F}_{ij} = f_j(x_i, y_i, y_i', \ldots, y_i^{(k)}),$$

and solve the following sparse regression problem

$$\mathbf{0} = \mathbf{F}\mathbf{w} + \epsilon,$$

where ϵ is the error. Each data induces one line of the matrix.

General version

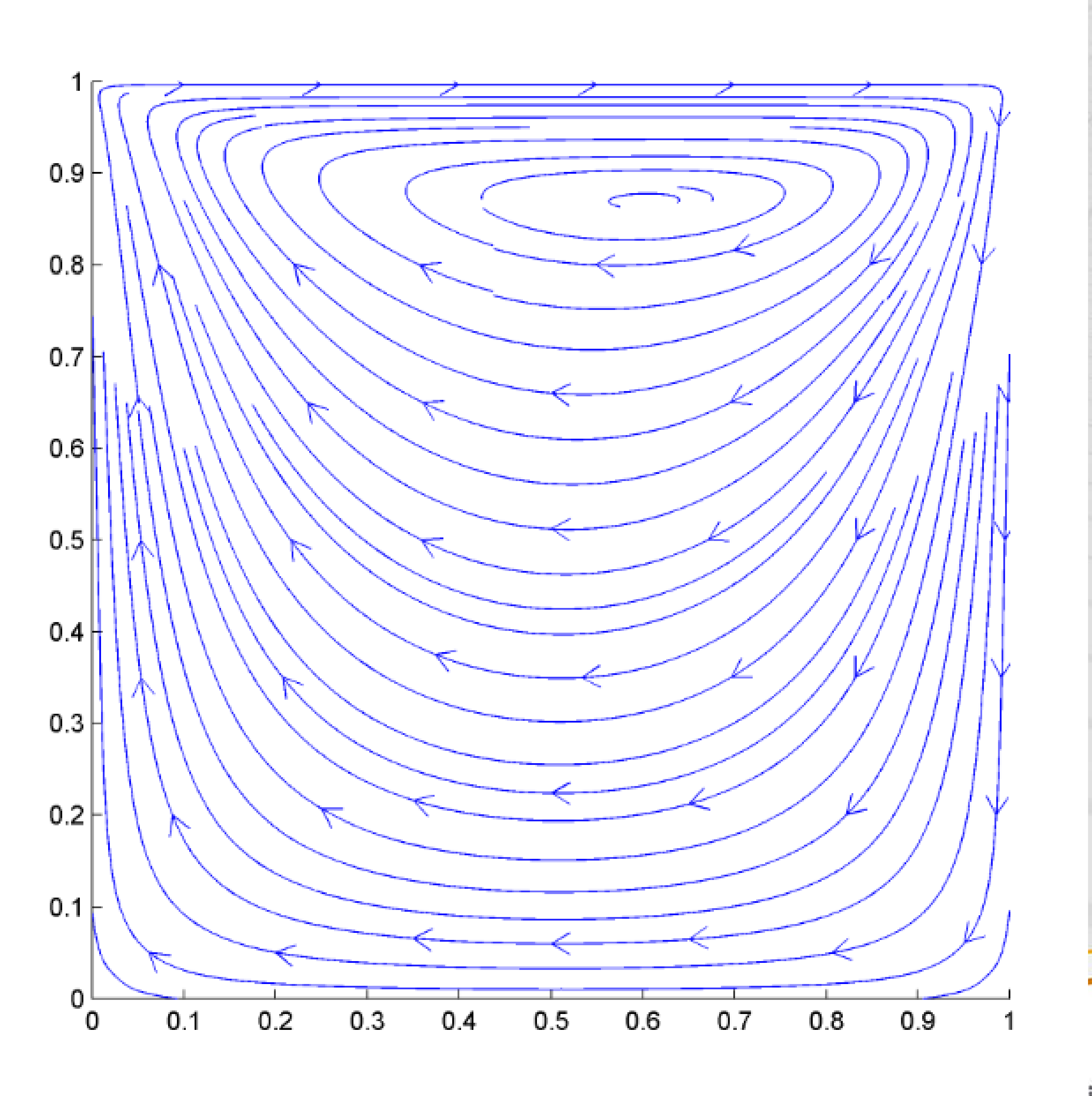
$$\begin{aligned} \mathbf{0} &= \mathbf{F}\mathbf{w} + \epsilon \\ \mathbf{0} &= \mathbf{F}_{\cdot 1} w_1 + \cdots + \mathbf{F}_{\cdot M} w_M + \epsilon \end{aligned}$$

- If we use sparse regression directly, we will probably get $\mathbf{w} = \mathbf{0}$.
- As a result, we fix one of the weight component as 1 to prevent getting trivial solution.
- For each $j \in \{1, \ldots, M\}$, fix $w_j = 1$ and move $\mathbf{F}_{\cdot j}$ to left hand side of the equation.

- Consider the following two dimensional incompressible Navier-Stokes equations:

$$\frac{\partial \vec{\mathbf{u}}}{\partial t} + [\vec{\mathbf{u}} \cdot \nabla]\vec{\mathbf{u}} - \nu \triangle \vec{\mathbf{u}} = -\nabla(p/\rho).$$

- $\vec{\mathbf{u}}$: flow velocity.
  ν: kinematic viscosity.
  p: pressure.
  ρ: density.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

- If we write $\vec{\mathbf{u}} = (u, v)$, then the equations are

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + \nu\frac{\partial^2 u}{\partial x^2} + \nu\frac{\partial^2 u}{\partial y^2} - \frac{\partial(p/\rho)}{\partial x}\\ \frac{\partial v}{\partial t} &= -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} + \nu\frac{\partial^2 v}{\partial x^2} + \nu\frac{\partial^2 v}{\partial y^2} - \frac{\partial(p/\rho)}{\partial y}.\end{aligned}$$

- We choose the basis functions based on dimensional analysis.
- Then we have the following result

$$\begin{aligned}\frac{\partial u}{\partial t} &= -0.980(\pm 0.002)u\frac{\partial u}{\partial x} - 0.986(\pm 0.001)v\frac{\partial u}{\partial y} + 0.973(\pm 0.002)\nu\frac{\partial^2 u}{\partial x^2}\\ &\quad +0.998(\pm 0.001)\nu\frac{\partial^2 u}{\partial y^2} - 0.997(\pm 0.001)\frac{\partial(p/\rho)}{\partial x}\\ \frac{\partial v}{\partial t} &= -0.986(\pm 0.001)u\frac{\partial v}{\partial x} - 1.011(\pm 0.001)v\frac{\partial v}{\partial y} + 0.995(\pm 0.001)\nu\frac{\partial^2 v}{\partial x^2}\\ &\quad +1.004(\pm 0.002)\nu\frac{\partial^2 v}{\partial y^2} - 0.997(\pm 0.000)\frac{\partial(p/\rho)}{\partial y}\\ 0 &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}.\end{aligned}$$

Result Num.	1	2	3	4	5	6	7
$u\partial u/\partial x$	**1**			**1.000**		-0.399	
$u\partial u/\partial y$		1					
$u\partial v/\partial x$		2.271	1			-0.421	0.123
$u\partial v/\partial y$	**1.000**	-0.909		**1**		0.880	-0.275
$v\partial u/\partial x$		1.746			**1**		
$v\partial u/\partial y$		-0.608				1	-0.216
$v\partial v/\partial x$		4.997	0.250			-1.396	1
$v\partial v/\partial y$					**1.000**	-0.462	
$\nu\partial^2 u/\partial x^2$		-3.389	-0.555			0.728	-0.555
$\nu\partial^2 u/\partial y^2$		-1.366					
$\nu\partial^2 v/\partial x^2$		2.455				-0.774	0.543
$\nu\partial^2 v/\partial y^2$		-7.156	-0.604			2.844	-1.338
$\partial(p/\rho)/\partial x$			-0.147			0.261	
$\partial(p/\rho)/\partial y$		-1.008				0.103	-0.197
Error bar $\times 10^3$	**0.000**	1110.350	94.758	**0.000**	**0.000**	3160.293	203.762

Merits

- Our method doesn't need to know much about the form of the equations in advance.
- Our method doesn't require the data to come from the same initial or boundary conditions.
- Our method can find invariant formulas automatically.
- Sparse Baysian regression has better sparsity and robustness than sequential threshold least squares and lasso.
- Our method generates a confidence interval for each parameter and an error bar for the whole equation.
- Random sub-sampling can be done parallelly.

Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

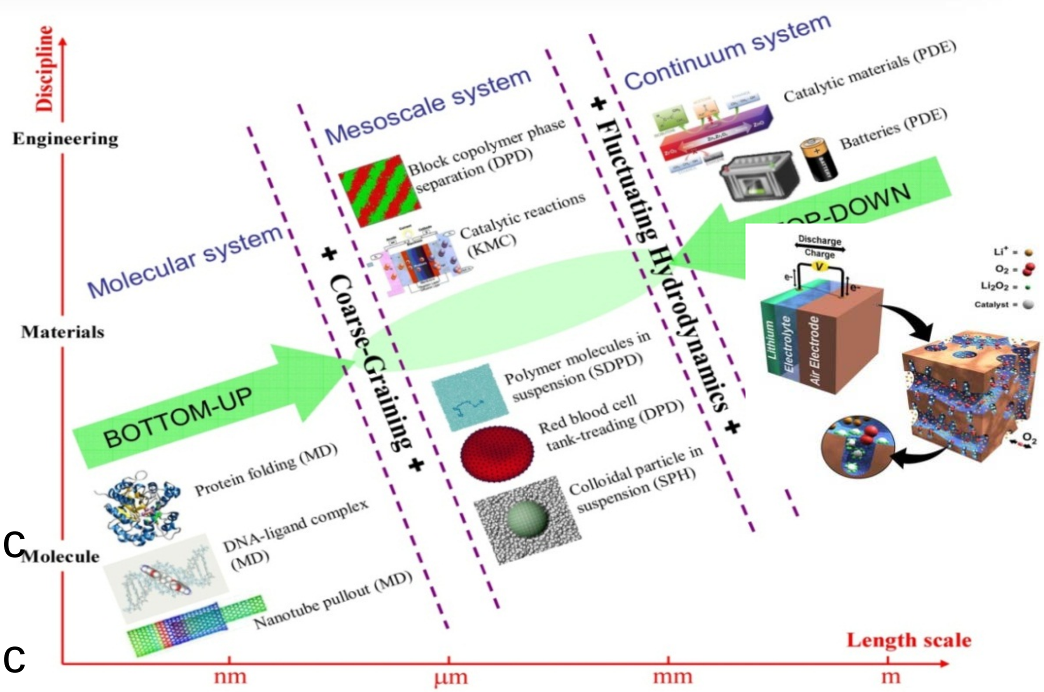
- Uncertainty Quantification for Deep Learning

Data Science & Modeling Challenges in Mesoscale Science

- Center for Mathematics for Mesoscopic Modeling of Materials (CM^4) – Guang Lin (Funded by DOE)

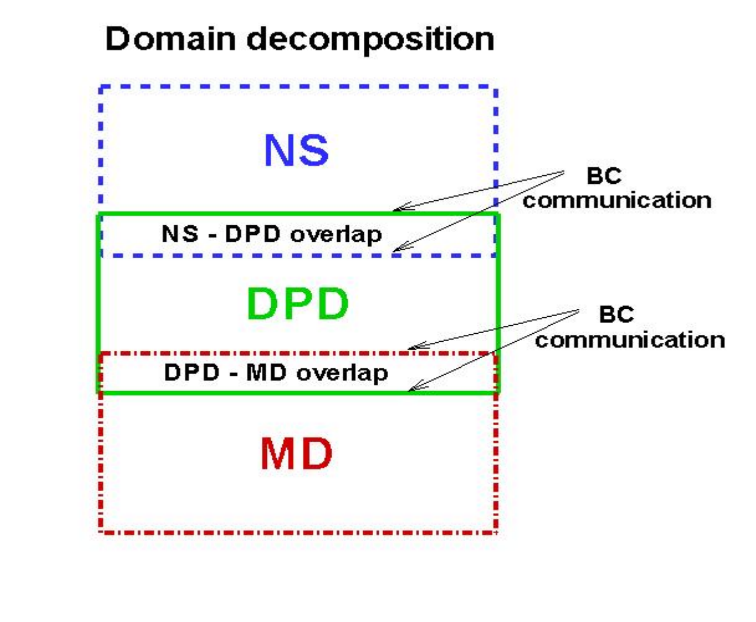
Traditional Approach	*Limitation*
- Mean field theory and analysis	- Non-equilibrium and dynamic processes.
- Continuum equation approximation	- Inhomogeneous systems, resolving microscopic details
- Ab-initio / Molecular Dynamics	- Prohibitive expensive computation for large scale systems

- **Objective**: Understand and control the design of multiscale complex systems with desired composition and structure.
- **Approach:** *Mesoscopic* model to bridge the gap between microscopic and macroscopic levels.
- **Difficulties:** Simply bottom-up / top-down scaling fails in general; Scale ambiguity across multiple regimes; Propagation of long-range microscopic interaction.
- **Goal:** Developing efficient mesoscopic simulation *methods*, *algorithms* and *models* applicable to complex physical systems across multiple regimes.



Sketch illustrating the range of mesoscale phenomena and their connection to molecular, mesoscale and continuum-based description.

Data-Driven Stochastic Multiscale Challenges in Mesoscale Science

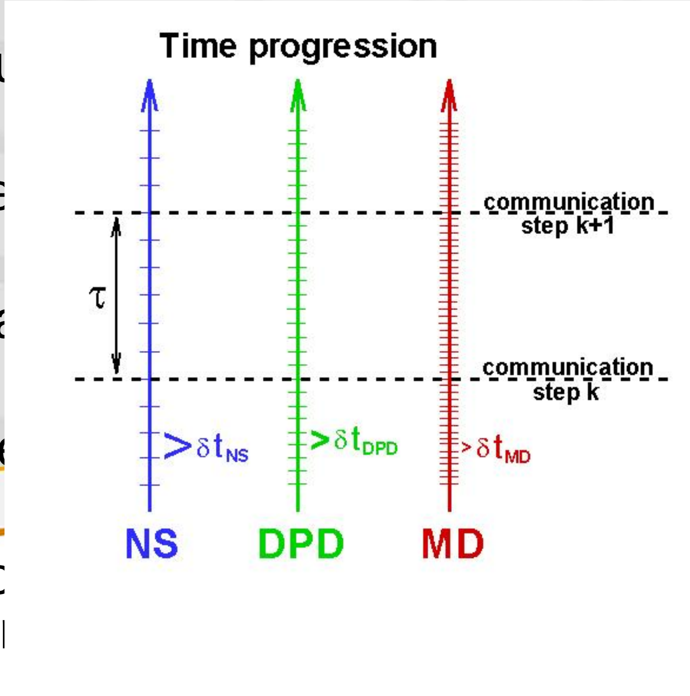


►Sensitivity Analysis, Error Control & Uncertainty Quantification

- Sensitivity analysis - identify what is the u bridge scales

Example: identifying the sensitive paramete estimating using DFT

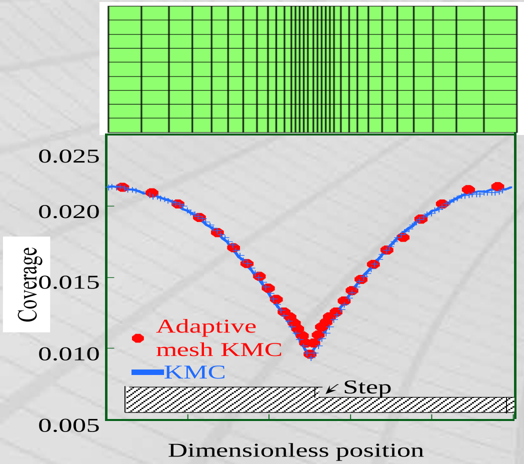
- UQ – provides a way to economically cha information across scales
- Error control, UQ and Bayesian paramete grained models
- Employ Bayesian inference framework to uncertainty using limited number of accu



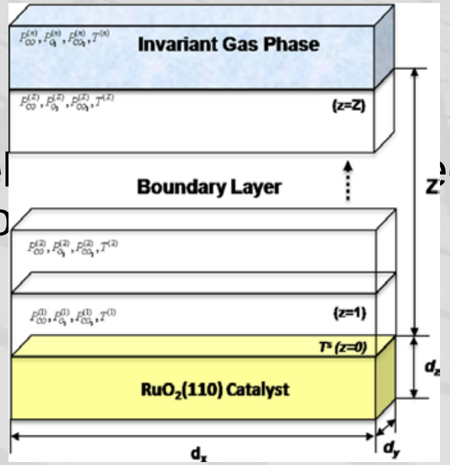
Data-Driven Stochastic Multiscale Challenges in Mesoscale Science

- **Separation of Time Scales:**
  - Reduced manifold techniques for overcoming stiffness
  - *Multiple-event execution in spatial KMC*
  - *Equation-Free method*
  - Time parareal
- **Separation of Length Scales:**
  - *Multiple length scale KMC*
  - *Adaptive Mesh/Model Refinement to overcome the problem of multiple length scales in realistic KMC simulations.*

- **Coupling KMC and stochastic continuum model** ects of heat and mass transfer on the kinetics of CO oxidatio catalyst.



AMR



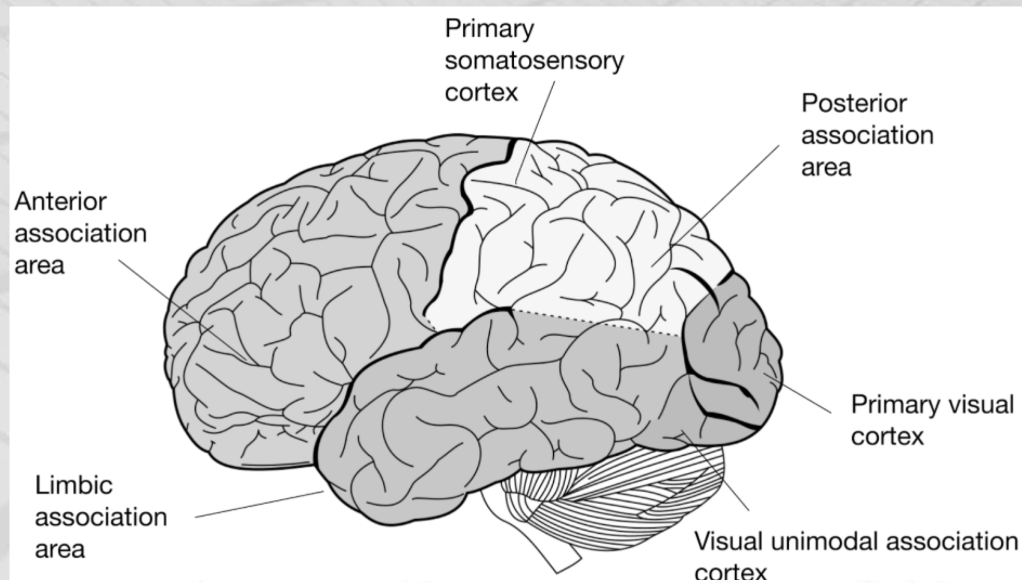
Schematic diagram of the multiscale model to investigate the effects of mass and heat transfer on the heterogeneous reaction kinetics

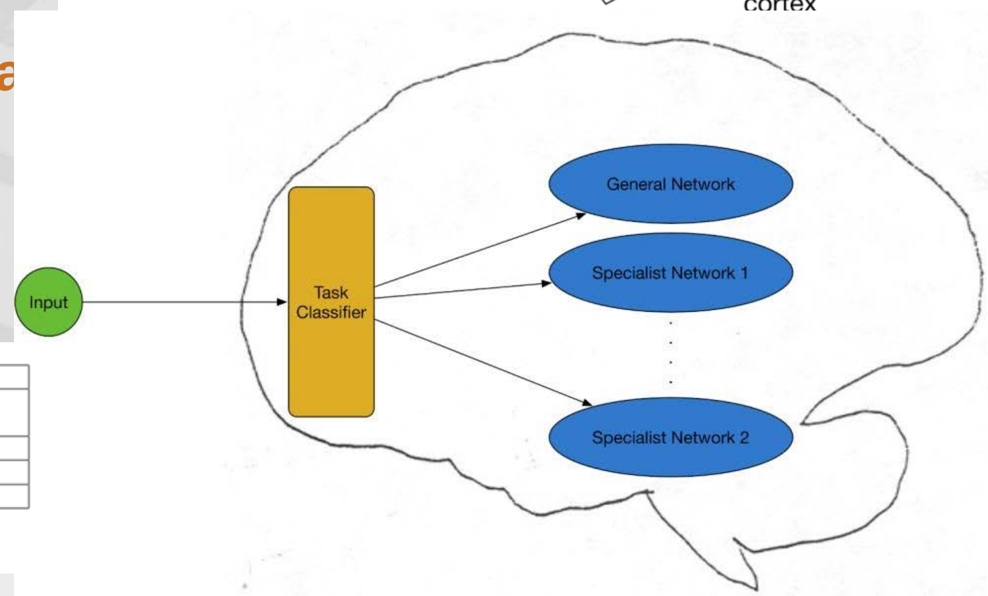
Mei, D.H. and G. Lin, *Effects of heat and mass transfer on the kinetics of CO oxidation over RuO2(1 1 0) catalyst.* Catalysis Today, 2011. 165(1): 56-63.

Cognitive Science
- Cortex & Brain Architecture
- Different areas in our brain will be activated for different tasks

CNNG
Base Networks (General, Specialist)
Task Classifier

Deep Learning with Colla





Preformance of CNNG and SingleNN					
Task	CNNG Accuracy (*)	SingleNN Accuracy	CNN accuracy	Adaboost Accuracy	CNNE Accuracy
MNIST	**99.65%**	97.84%	96.35%	79%	98.6%
EMNIST	**90.88%**	82.32%	83.5%	65%	89.3%
Fashion-MNIST	**98.81%**	85.56%	91.9%	93.6%	92.22%

Table 1: Preformance of CNNG and SingleNN

L. Gao, H. Wang, **G. Lin**, Reflective neural network ensembles, 2019 International Joint Conference on Artificial Intelligence, August 10, 2019, Macao, P.R. China.

Outline:

- UQ for Complex systems: its challenge and open issues
  - UQ open issue 1: Discontinuities (ME-gPC, ME-PCM, et. al)
  - UQ open issue 2: Curse of Dimensionalities (Sparse grid, Adpative ANOVA, compressive sensing algorithm with basis rotation, et. al)
  - UQ open issue 3: Heterogeneous big data & Computational Expensive Models - Bayesian parameter estimation in large-scale regional and global climate models

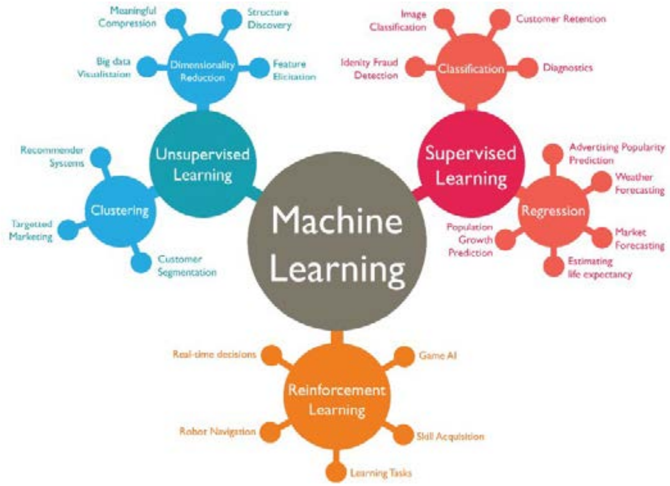
Pacific Northwest NATIONAL LABORATORY

- Uncertainty Quantification for Deep Learning

DATA SCIENCE CONSULTING SERVICE

As part of Purdue Integrative Data Science Initiative (IDSI), Data Science Consulting Service will provide hands-on consulting support for data analysis and business analytics in all areas to overcome data science challenges arising in research, education, and business and organization management. Our consultants have advanced degrees and years of experience in deep machine learning, data mining, big data analysis, artificial intelligence, business analytics and computational statistics.

Mission Statement



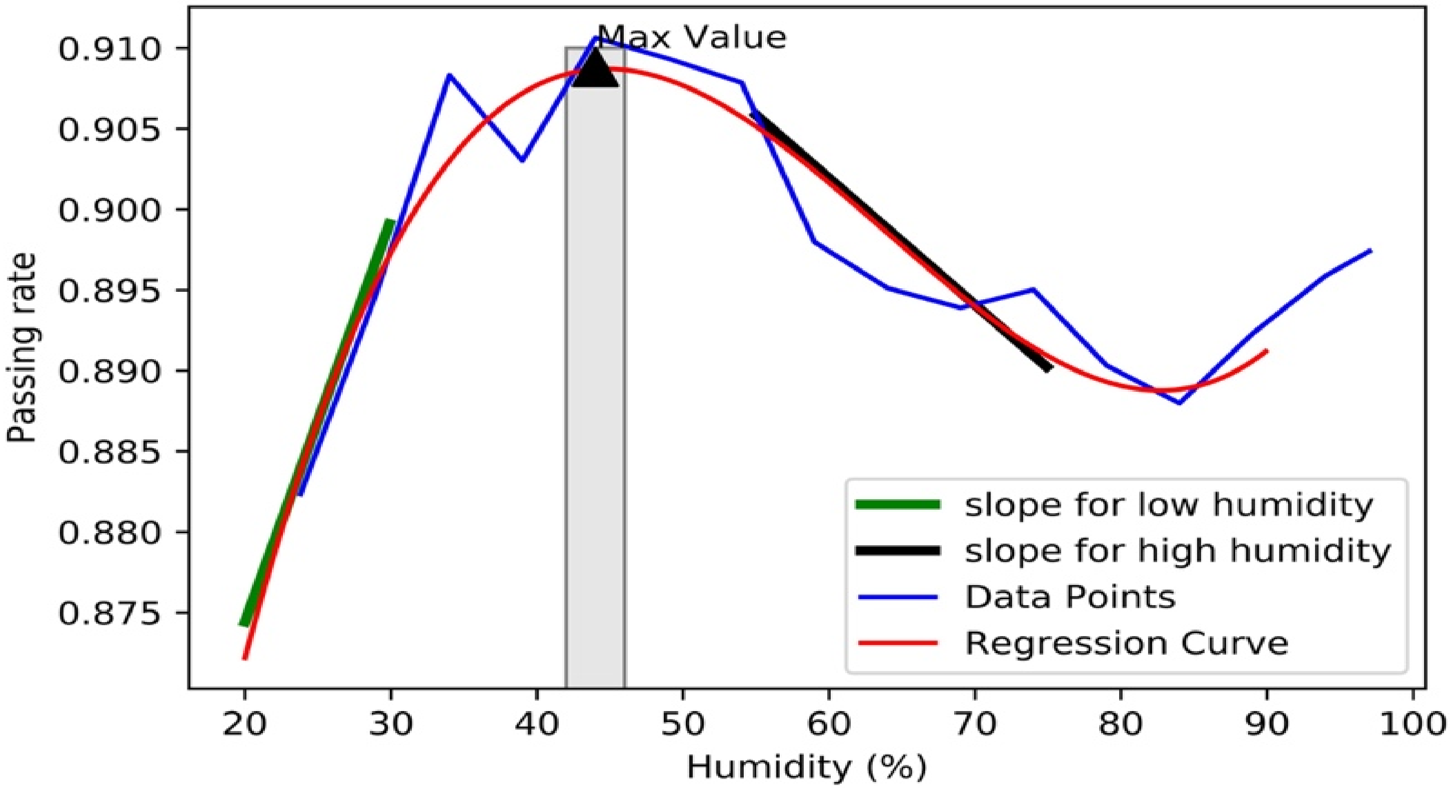
- Establish a leading role in data science consulting for research, education and industry clients
- Provide self-sustainable, efficient data science consulting service
- Provide a data-science focal point for federal, state and private industry to engage
- Develop a consulting platform for Purdue faculties with different expertise to collaborate and provide a unified consulting service for industry clients

Data Science Consulting Expertise

- **Business analytics and business intelligence.** DSCS staff and consultants have expertise in using advanced statistical analysis, data mining, machine learning and artificial intelligence tools to explore the client's data in support of data-driven decision-making.
- **Data and information management.** DSCS staff and consultants have experience in using advanced database and data processing tools to manage big, and unstructured data for analysis using a variety of scripting languages and tools.
- **Advanced methodology for data science.** DSCS staff and consultants have expertise in methodological aspects of data science including statistical data analysis, machine learning, artificial intelligence, uncertainty quantification, and sensitivity analyses.
- **Data exploration.** DSCS staff and consultants have experience in data visualization, interpretation, and hypothesis-generating research.
- **High-performance data processing.** DSCS staff and consultants have expertise in optimization of code for data processing in CPU and GPU environment.

Case Studies

IMPROVING THE QUALITY OF CHRYSLER CROSSMEMBER CASTINGS



Summary:
A crossmember is a structural component that undergoes strict X-ray inspection to ensure its quality. The optimal environmental and operational parameter settings are identified for making quality CHRYSLER crossmember castings through a novel optimization algorithm.

Y. Sun, **G. Lin**, Q. Han, D. Yang, C. Vian, Exploratory data analysis for achieving optimal environmental and operational parameter settings for making quality crossmember castings, Die Casting Congress & Exposition, 1, 2019.

Case Study: Deep Learning for Electron Cryo-Microscopy (Cryo-EM) Images

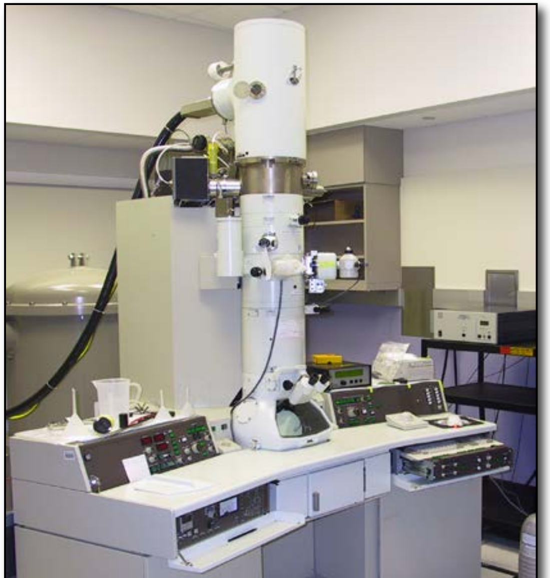
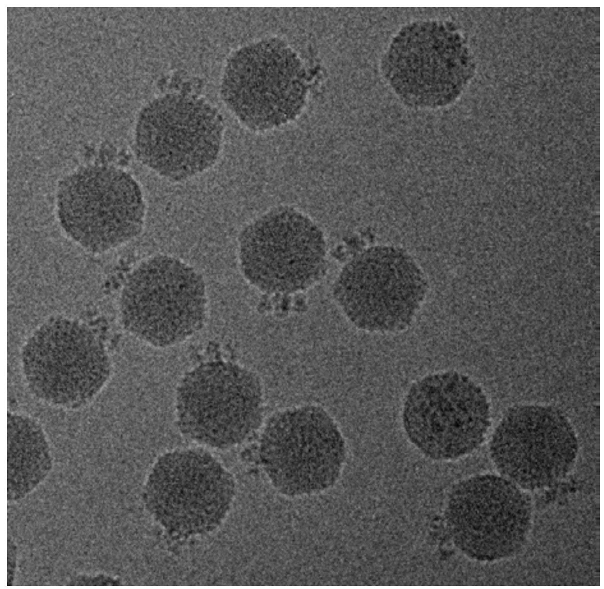
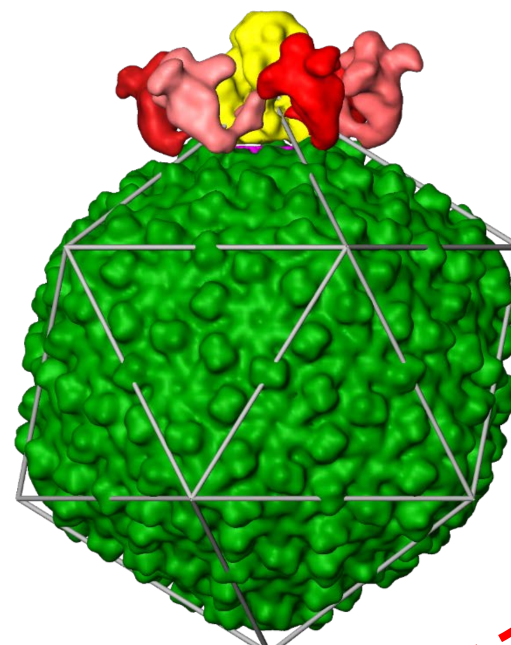


image collection
(3D→2D)

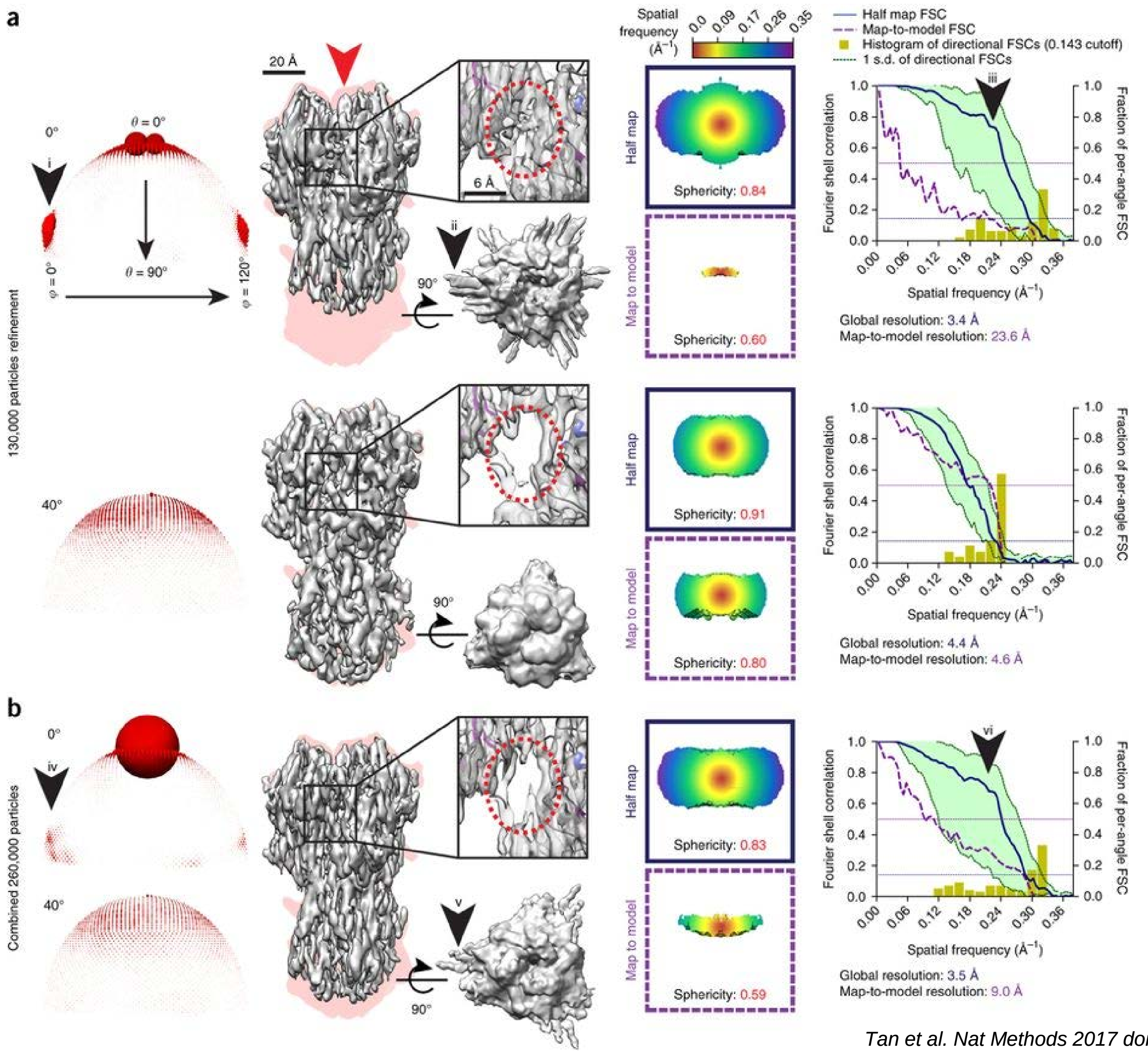


3-D reconstruction
(2D→3D)



Jiang et al. Nature 2006

Case Studies: Frequent Problems-Preferred Views

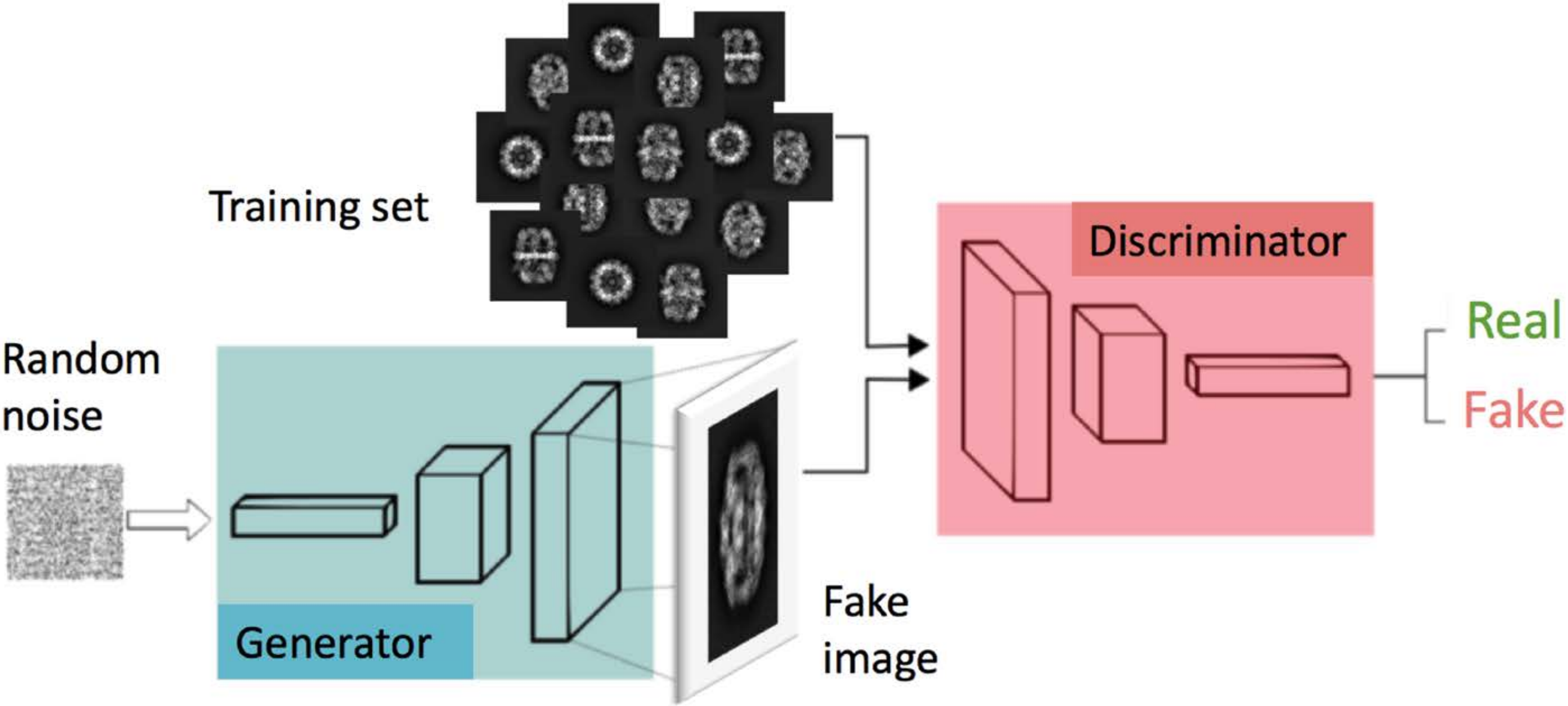


Tan et al. Nat Methods 2017 doi:10.1038/nmeth.4347

Case Studies: Deep Generative Models

Generative Adversarial Networks (GAN)

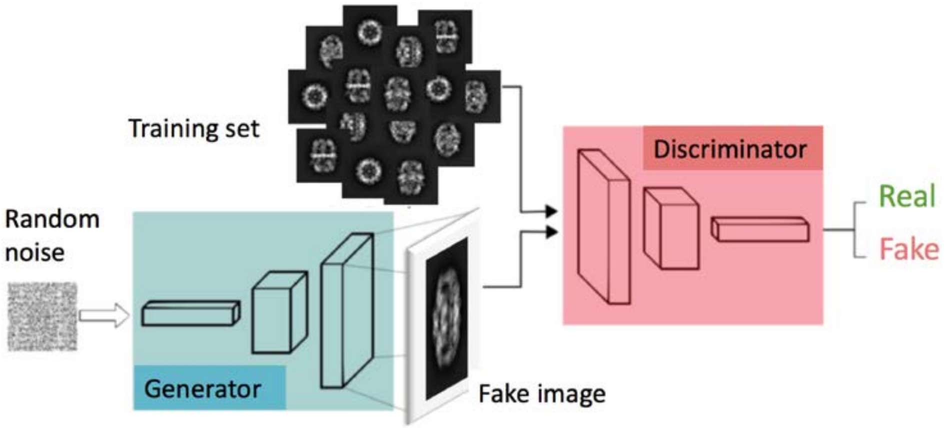
- Model structure: generator, discriminator
- Input: projections from various angles of a bio-structure

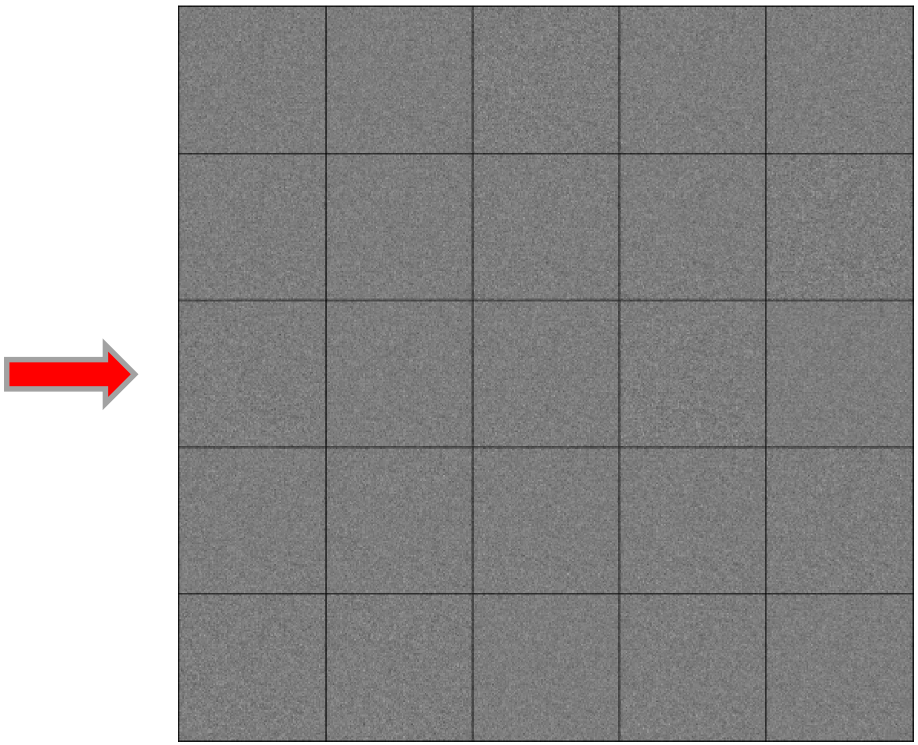


Case Studies: Deep Generative Models

Preliminary Results

Generative Adversarial Networks (GAN)





Uncertainty Quantification and Scientific Machine Learning for Complex Engineering Systems

"…Because I had worked in the closest possible ways with physicists and engineers, I knew that our data can never be precise…"

Norbert Wiener