Path-space information metrics for uncertainty quantification of molecular systems

Markos Katsoulakis
Mathematics & Statistics
University of Massachusetts Amherst

NIST Workshop on Quantification of Uncertainties in Materials Science
January 2016

Acknowledgments

- Paul Dupuis (Brown University)
- Petr Plechac (University of Delaware)
- Luc Rey-Bellet (University of Massachusetts, Amherst)
- Dion Vlachos (Chem. Engineering & EFRC, University of Delaware)

- Georgios Arampatzis (ETH, Zurich)
- Yannis Pantazis (University of Massachusetts, Amherst)
- Eric Hall (University of Massachusetts, Amherst)
- Kostis Gourgoulias (University of Massachusetts, Amherst)
- Jie Wang (University of Massachusetts, Amherst)

DARPA

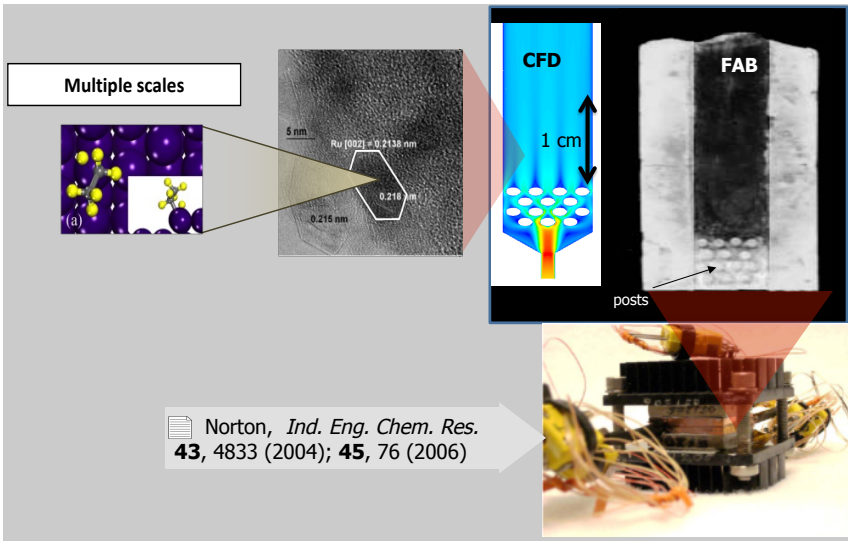
DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA ·

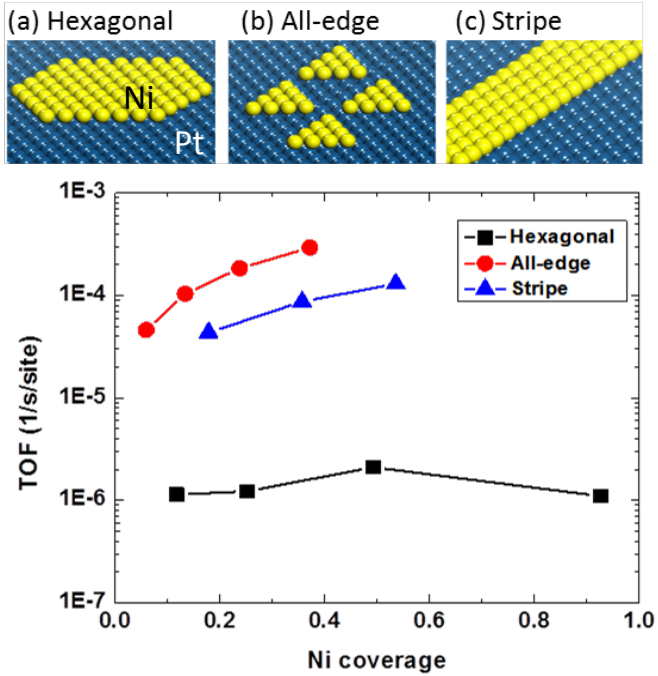
NSF

Overview

- **Main mathematical topics**
  1. Sensitivity analysis, screening, uncertainty quantification, coarse-graining and model selection based on information inequalities
  2. Model Robustness, Parameter Identifiability
  3. Sensitivity analysis, screening and uncertainty quantification based on information metrics for rare event problems
  4. Focus on dynamics and non-equilibrium, driven molecular systems
- **Main application area**
  - Predictive materials design. Catalysis and Chemical Kinetics.
  - The main models in this connection are complex chemical reaction networks–which are high (often very high) dimensional stochastic systems.
  - UQ/Sensitivity at mesoscales: kinetic Monte Carlo, complex reaction networks, Langevin dynamics,...

- Mean field, PDE-based models are inadequate due to the role of microstructure. Need (at least) Kinetic Monte Carlo-level resolution.
- Need UQ methods for possibly stochastic dynamics, scaling for $T \gg 1$, $N \gg 1$ (long times, high dimensions). Multi-scale problems.



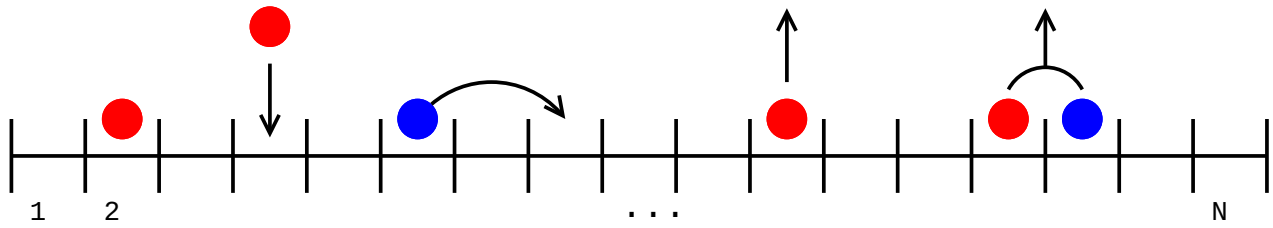


(Right) Patched bimetallic catalysts for ammonia decomposition: W. Guo, D. G. Vlachos, Nature Communications (2015)

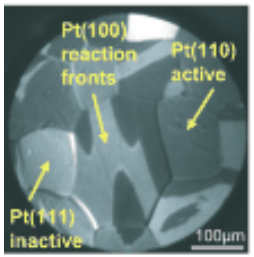
Lattice Dynamics and Kinetic Monte Carlo (KMC)

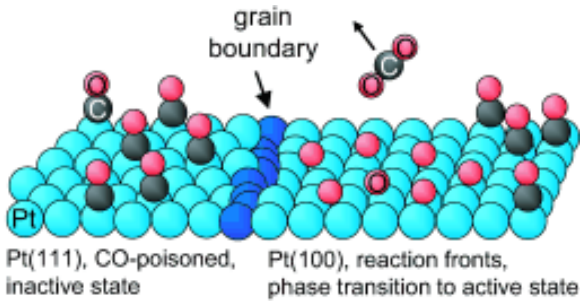
Dynamics

- Adsorption/Desorption/Reactions/Surface diffusion



- **Continuous Time Markov Chain** modeling with state space $\Sigma =$ all configurations σ
- Multi-site updates or most systems, e.g.





Suchorski et al ChemPhysChem (2010)

Main mathematical ideas

Use of information metrics (relative entropy, related divergences) as a set of analytically and computationally tractable measures of *measures of distance* or *metrics* between models.

Information measures we use are not true metrics, but have excellent quantitative and qualitative properties. We use these measures to develop practical and broadly applicable techniques for UQ with respect to ordinary observables, as well as those that are determined by rare events.

- Use of large deviation ideas for the UQ and sensitivity analysis for rare event problems.

  Large deviations is the part of probability theory that estimates the probabilities of rare events and characterizes the most likely pathway of the system leading to the rare event. Well known as a theoretical tool, our use is in the design and analysis of computational methods for UQ.

- Information metrics on "path space" (i.e. all pertinent time series)
  - Non-equilibrium Steady States (NESS): $NESS \neq Z^{-1}e^{-\beta H(\sigma)}$
  - Relative Entropy Rate: "pseudo-metric" on path-space.
  - Path-Space Fisher Information Matrix

- Observables and UQ Information Inequalities for Sensitivity Screening that: scale with system-size and time.

Why information-based methods?

- Pseudo-distance between probabilistic models P, Q:

$$\mathcal{R}_{KL}(P \,|\, Q) := \int \log\left(\frac{P}{Q}\right) dP$$

- Properties: (i) $\mathcal{R}_{KL}(P \,|\, Q) \geq 0$ and
  (ii) $\mathcal{R}_{KL}(P \,|\, Q) = 0$ iff $P = Q$ a.e.
- Other probability metrics and divergences:

  Total Variation, Hellinger, χ^2, F-divergence, etc. e.g.

$$\chi^2(P \,||\, Q) = \int \left(\frac{P}{Q} - 1\right)^2 dQ.$$

- Is relative entropy special? see later.
- Drawbacks:
  - need absolute continuity, i.e. some probability models cannot be compared. Need other methods (stochastic coupling, Malliavin calculus, etc)
  - What is the connection with observables of interest?

Information metrics in inference, model selection and UQ

- Variational inference methods in machine learning
- Variational inference for building coarse-grained models in materials
- Information metrics for UQ and sensitivity analysis of stochastic models
- Information metrics for quantifying predictive skill in model selection/reduction

Variational inference and coarse-graining

Loss of Information in Coarse-Graining, CG error, model fidelity : K., Vlachos J. Chem. Phys. (2003), Majda, Abramov (2006), K., Plechac, Rey-Bellet, Tsagkarogiannis (2014), Chen, Tong, Majda (2014) ...

$$\mathcal{R}_{KL}\left(P \,|\, Q\right) = N \times \mathcal{O}(\epsilon^p)\,, \quad N = \text{system size}, \quad \epsilon = \text{tolerance}$$

CG Parametrizations via Variational Inference: Shell (2008, 2012), Noid et al (2011), Espanol, Zuninga (2011), Bilionis, Koutsourelakis (2012), Bilionis, Zabaras (2013), K., Plechac (2013) ...

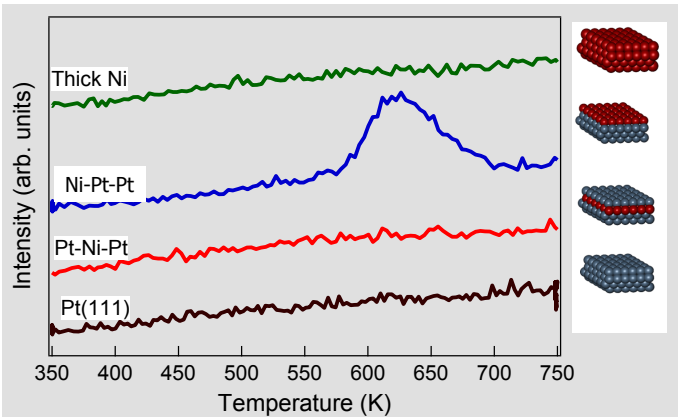
$$\min_{\theta} \mathcal{R}_{KL}\left(P \,|\, Q(\theta)\right)\,.$$

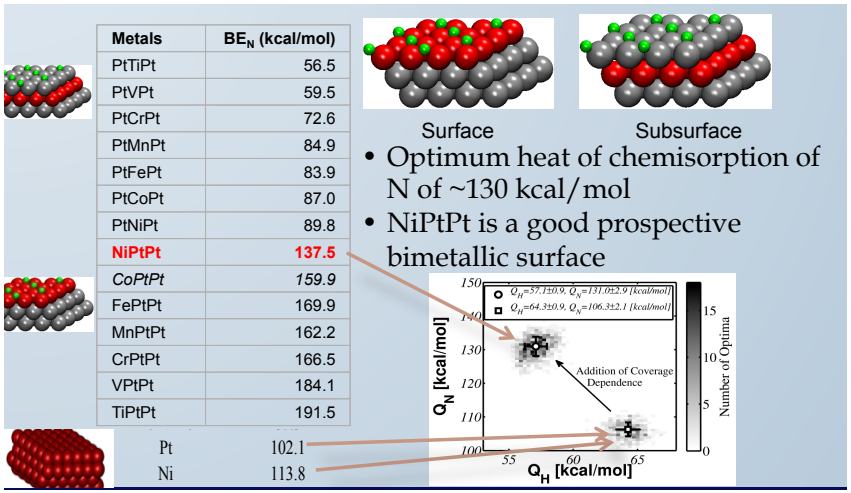
Machine/Statistical Learning via Variational inference: Amari (1998), Jordan et al (1999), Bottou (2003), Wainright, Jordan (2008), Hoffman, Blei, Wang, Paisley (2013)...

Example: Sensitivity analysis – Applications

Identifying Bimetallic Catalysts

Hansgen, Chen, Vlachos, *Nature Chem.* 2010

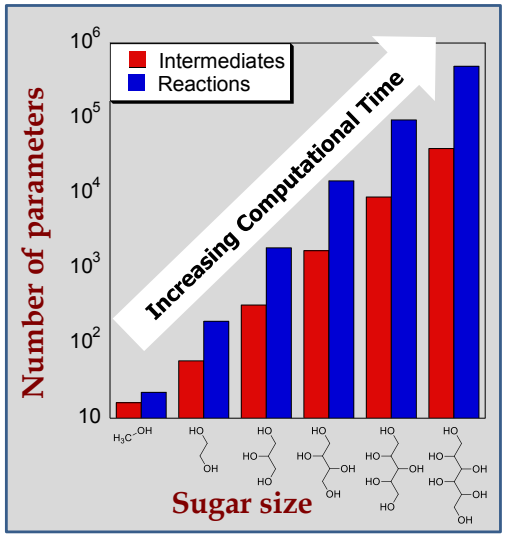


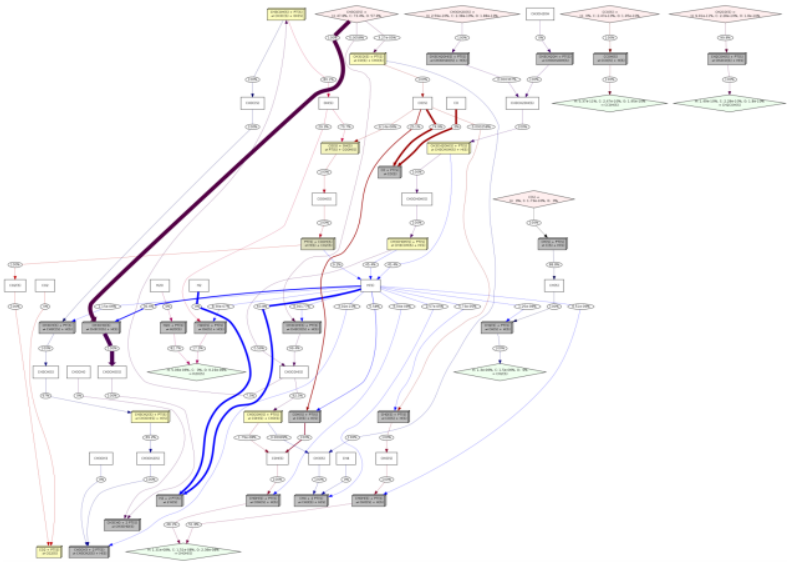


- Model sensitivity to lateral interactions: Parameter Sensitivity Analysis

  Use costly DFT simulation and estimation only for sensitive parameters or important mechanism of the KMC.
- Strong dependence on some molecular architecture (but insensitive to most): Uncertainty Quantification
- Have we included all reactions? Model-form Uncertainty

Complex Reaction Networks





Left: Number of parameters in metal-catalyzed upgrade of small biomass derivatives for the production of renewable fuels and chemicals vs. molecular size. The reaction network even of typical sugars, such as glucose, entails nearly a million of parameters. Right: Example of reaction network of a small oxygenate (ethanol on platinum catalyst); the thickness of lines indicates the reaction flux.

J. E. Sutton and D. G. Vlachos. Chem. Engin. Sci. 2015.

Sensitivity Analysis - some challenges

- Perform SA on stochastic (Langevin) and possibly discrete systems, non-gaussian (KMC, reaction-networks).
- Transient and long-time integration (numerical/approximation challenges).
- Need to tackle non-equilibrium stochastic processes:
  - $NESS \neq Ce^{-\beta H(\sigma)}$
- SA for molecular (and multi-physics) models with a very large number of parameters. Sensitivity analysis and bounds that can be used for screening.
- Parameters can be strongly correlated (e.g. due to thermodynamic constraints, DFT calculations, etc) $\Longrightarrow$ type of "global" SA.
- Uncertainties related to rare events and induced multiscale aspects.

---

[1] Pantazis, K., *J. Chem. Phys.* '13; K., Y. Pantazis, D. Vlachos, *BMC Bioinformatics*, '13, Pantazis et al., *J. Chem. Phys.* '15.

[2] Arampatzis, Pantazis, Katsoulakis *PLOS 1*, (2015); Dupuis, Katsoulakis, Pantazis, Plechac *SIAM UQ*, (2016)

[3] Sutton, Guo, Katsoulakis, and Vlachos, *Nature Chemistry*, to appear (2016)

[4] P. Dupuis, M. Katsoulakis, Y. Pantazis and L. Rey-Bellet (2016)

Deterministic Sensitivity Analysis - Local SA

- System of ODEs:

$$\dot{y} = f(y;\theta) \quad , \quad y(0) = y_0 \in \mathbb{R}^N$$

  - Goal: Perform SA on the model parameters $\theta \in \mathbb{R}^K$.
- Define sensitivity indices:

$$s_k = \frac{\partial y}{\partial \theta_k}$$

- A new system of ODEs is derived and augmented to the previous:

$$\dot{s}_k = \frac{\partial f}{\partial y} s_k + \frac{\partial f}{\partial \theta_k}, \ k = 1, ..., K$$

  - need to solve $K \times N$ additional equations.

Stochastic Sensitivity Analysis - Background

- Observable-based Methods:
  - Finite-differencing:

$$S(\theta,t)=\frac{\partial}{\partial\theta}\mathbb{E}_{P_t^\theta}[f(x)]=\frac{\mathbb{E}_{P_t^{\theta+\epsilon}}[f(x)]-\mathbb{E}_{P_t^\theta}[f(x)]}{\epsilon}+O(\epsilon)$$

  - Likelihood ratio method [P. Glynn, *Comm. ACM* (1990)]:

$$S(\theta,t)=\frac{\partial}{\partial\theta}\mathbb{E}_{P_t^\theta}[f(x)]=\int f(x)\partial_\theta P_t^\theta(x)dx=\mathbb{E}_{P_t^\theta}[f(x)\partial_\theta \log P_t^\theta(x)]$$

  - Linear Response methods in (primarily equilibrium) Stat. Mechanics.
  - Pathwise derivatives, Malliavin Calculus methods.
  - Vast literature on gradient estimation for discrete event systems in Queuing Theory and Operation Research.

Stochastic SA - Observable-based

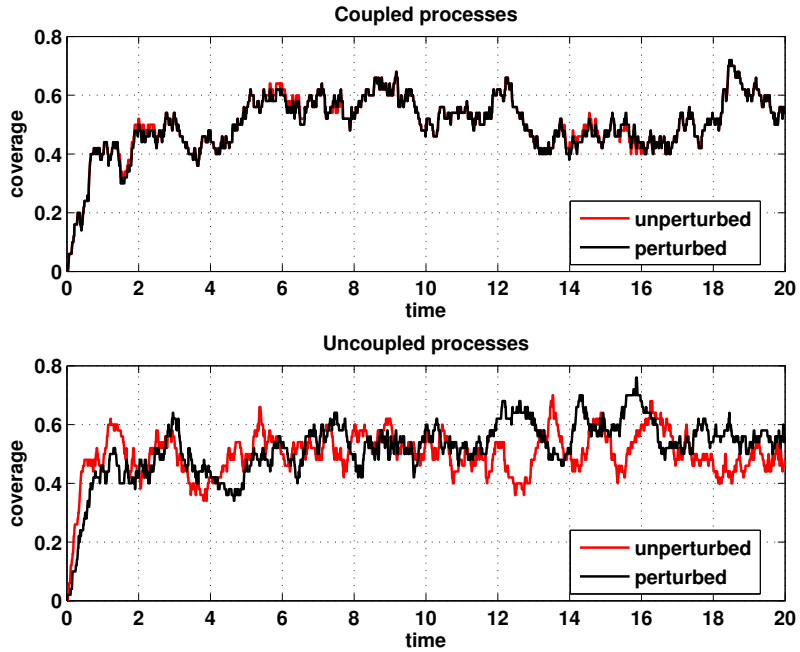
- Finite differencing:

$$S(\theta,t) \approx \frac{\mathbb{E}_{P_t^{\theta+\epsilon}}[f(x)] - \mathbb{E}_{P_t^{\theta}}[f(x)]}{\epsilon} + O(\epsilon)$$

- Variance of the estimator controlled by:

$$\begin{aligned}\mathrm{var}\left(f(x_t^{\theta+\epsilon}) - f(x_t^{\theta})\right) = \mathrm{var}\left(f(x_t^{\theta+\epsilon})\right) + \mathrm{var}\left(f(x_t^{\theta})\right) \\ - 2\mathrm{cov}\left(f(x_t^{\theta+\epsilon}), f(x_t^{\theta})\right)\end{aligned}$$

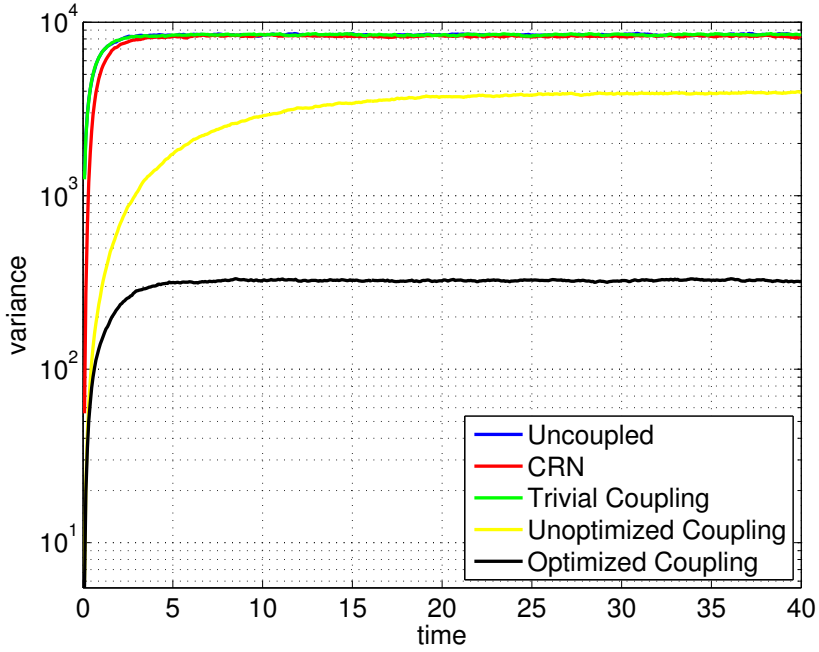
High Variance due to sampling of two different stochastic processes. Can we "couple" the two processes?



Stochastic Coupling Methods

Simulate joint process $(x_t^{\theta+\epsilon}, x_t^{\theta})$ with constrained marginals:

- Common Random Number/Common Reaction Path approach: [Rathinam, Sheppard, Khammash, J. Chem. Phys., 2010]
- Markov $(x_t^{\theta+\epsilon}, x_t^{\theta})$ for well–mixed systems: [Anderson, SIAM Numerical Analysis, 2012], [Srivastava, Anderson, Rawlings, J. Chem. Phys., 2013]
- Markov $(x_t^{\theta+\epsilon}, x_t^{\theta})$ for spatial KMC - couplings based on coupling optimization principle: [Arampatzis, K. , J.Chem.Phys., 2014]
- Efficient methods, but impractical for systems with a large # of parameters



Variance for different coupling methods in Lattice KMC:

[Arampatzis, K., J.Chem.Phys., '14]

Sensitivity analysis - Information Theory Methods

$$\mathcal{R}_{KL}\left(P^{\theta} \,|\, P^{\theta+\epsilon}\right) = \text{Loss of Information due to perturbation by } \epsilon$$

A. J. Majda and B. Gershgorin, *Proc. Natl. Acad. Sci.* (2010) - Maximum Entropy principle

M. Komorowski et al, *Proc. Natl. Acad. Sci.* (2011) - Linear Noise Approximation models or reaction networks

- Here the PDF is known, e.g. a Gibbs equilibrium $\mu \sim Ce^{-\beta H(\sigma)}$ or Gaussian fluctuations: allows for explicit calculations on $\mathcal{R}$:

$$\mathcal{R}_{KL}\left(\mu \,|\, \mu^{(0)}\right) \sim \mathbb{E}_{\mu}[\beta(H^{(0)} - H)] + \log \frac{Z^{(0)}}{Z}$$

- Gradients of $\mathcal{R}_{KL}\left(\mu \,|\, \mu^{(0)}\right)$ become observables: stochastic optimization methods.
- However, typically

$$\mu \neq Ce^{-\beta H(\sigma)}$$

in dynamics, non-equilibrium systems, non-gaussian fluctuations, etc.

Sensitivity Analysis - Relative Entropy Rate (RER)

$\mathcal{R}_{KL}\left(Q^{\theta}_{0,T} \,|\, Q^{\theta+\epsilon}_{0,T}\right)$ = Loss of Information (in time-series) due to perturbation by ϵ

- For long times $T >> 1$, RER is viewed as measure of parameter sensitivity:

$$\mathcal{R}_{KL}\left(Q^{\theta}_{0,T} \,|\, Q^{\theta+\epsilon}_{0,T}\right) = T\mathcal{H}\left(Q^{\theta}_{0,T} \,|\, Q^{\theta+\epsilon}_{0,T}\right) + \mathcal{R}_{KL}\left(\mu^{\theta} \,|\, \mu^{\theta+\epsilon}\right)$$

  - θ: parameter vector (for local sensitivity it is fixed)
  - ϵ: parameter vector perturbation

- (RER): $\mathcal{H}\left(Q^{\theta}_{0,T} \,|\, Q^{\theta+\epsilon}_{0,T}\right) = \mathbb{E}_{\mu^{\theta}}\left[\int p^{\theta}(\sigma,\sigma') \log \frac{p^{\theta}(\sigma,\sigma')}{p^{\theta+\epsilon}(\sigma,\sigma')} d\,\sigma'\right]$

  - $p(\sigma,\sigma')$: transition probabilities (local dynamics).
  - $\mathbb{E}_{\mu^{\theta}}$.... steady-state sampling: $\bar{\mathcal{H}}^{(n)}_2 = \frac{1}{n}\sum_{i=0}^{n-1} \log \frac{p^{\theta}(\sigma_i,\sigma_{i+1})}{p^{\theta+\epsilon}(\sigma_i,\sigma_{i+1})}$

Path-space Fisher Information Matrix (FIM)

Under a smoothness assumption on θ, (checkable, on the rates only!)

$$\mathcal{H}\left(Q^{\theta}_{0,M} \mid Q^{\theta+\epsilon}_{0,M}\right) = \frac{1}{2}\epsilon^{T}\mathbf{F}_{\mathcal{H}}\left(Q^{\theta}_{0,M}\right)\epsilon + O(|\epsilon|^{3})$$

where the Fisher Information Matrix is defined as

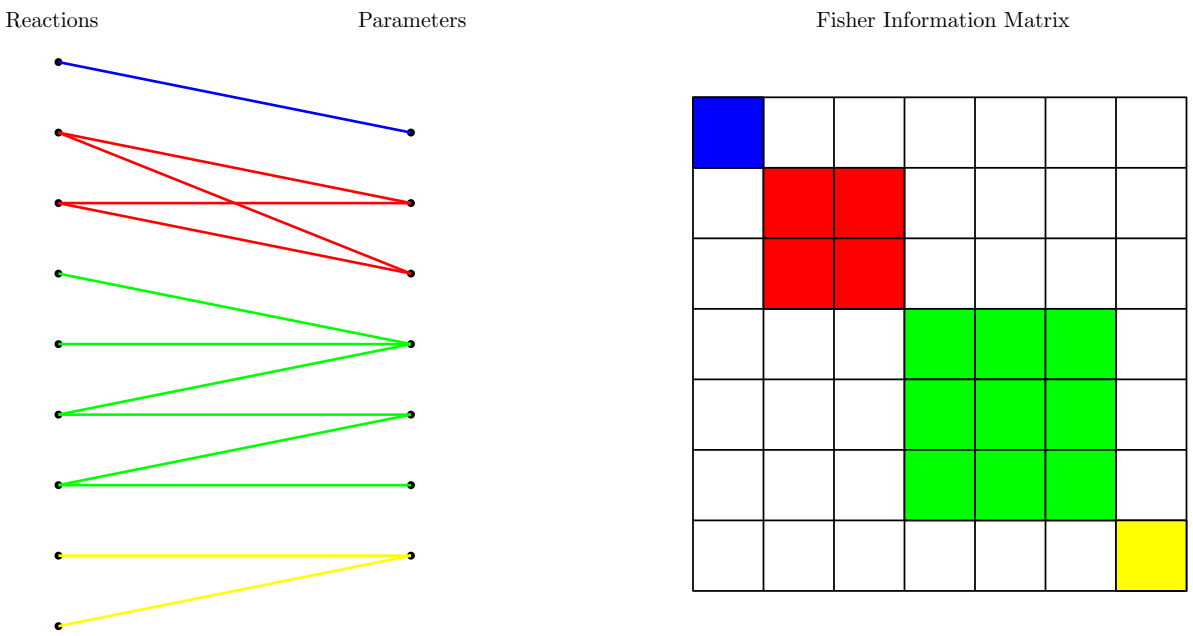
$$\mathbf{F}_{\mathcal{H}}\left(Q^{\theta}_{0,M}\right) = \mathbb{E}_{\mu^{\theta}}\left[\int_{E} p^{\theta}(\sigma,\sigma')\nabla_{\theta}\log p^{\theta}(\sigma,\sigma')\nabla_{\theta}\log p^{\theta}(\sigma,\sigma')^{T}\, d\,\sigma'\right]$$

- Spectral analysis of FIM gives the most/least sensitive directions.
- Sparse structure of the path FIM-see examples below.
- Derivative-free sensitivity analysis method.
- Characterizes robustness on simultaneous parameter perturbations.
- Determines parameter identifiability, e.g. Cramer-Rao Thms

[1] Pantazis, K., *J. Chem. Phys.* (2013); K., Y. Pantazis, D. Vlachos, *BMC Bioinformatics*, (2013), Arampatzis, K. Pantazis *PLOS 1* (2015).

Path-space FIM - Sparsity

- For reaction networks, Langevin, etc. we typically have a block diagonal structure in the FIM
- Scalable computations - FIM **scales linearly** in the number of parameters.
- Contains key information:
  - Graph structure
  - Dynamics on the graph: reaction rates and their functional form



How "predictive"? Information Inequalities and QoIs

- Observables/QoIs f: Csiszar-Kullback-Pinsker inequality:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq ||\phi||_\infty \sqrt{2\mathcal{R}_{KL}\left(P \mid Q\right)}$$

[1] K., L. Rey-Bellet, J. Wang, '16; Dupuis, M. K., Pantazis, Plechac*SIAM UQ* '16

How "predictive"? Information Inequalities and QoIs

- Observables/QoIs f: Csiszar-Kullback-Pinsker inequality:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq ||\phi||_\infty \sqrt{2\mathcal{R}_{KL}\left(P \mid Q\right)}$$

- Does not scale with system size N, time T.
- Same with χ^2 bounds (Chapman Robbins inequality).

[1] K., L. Rey-Bellet, J. Wang, '16; Dupuis, M. K., Pantazis, Plechac*SIAM UQ* '16

How "predictive"? Information Inequalities and QoIs

- Observables/QoIs f: Csiszar-Kullback-Pinsker inequality:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq ||\phi||_\infty \sqrt{2\mathcal{R}_{KL}(P \mid Q)}$$

- Does not scale with system size N, time T.
- Same with χ^2 bounds (Chapman Robbins inequality).

UQ Information Inequalities[1]**:**

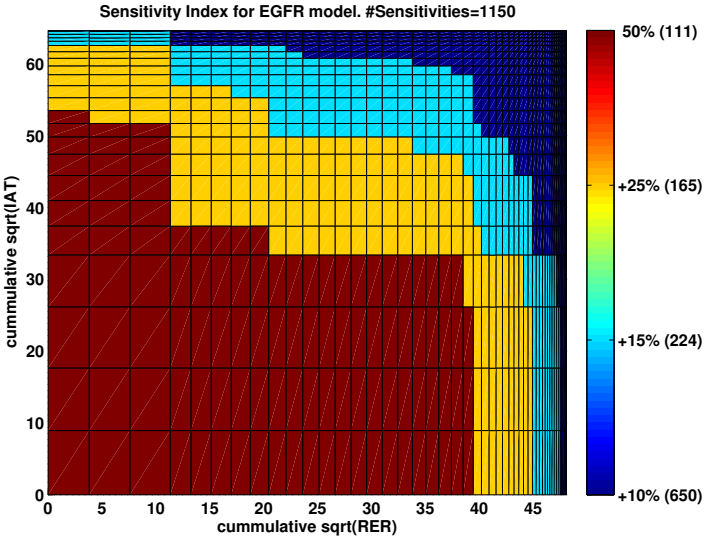
$$\Xi_-(P, Q; f) \leq \mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \Xi_+(P, Q; f),$$

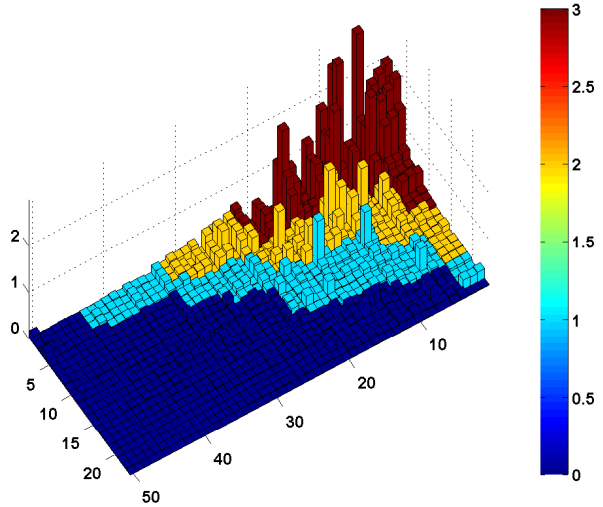
- **Goal-oriented divergence** $\Xi_\pm(P, Q; f)$:
  - $\Xi_\pm(P, Q; f) \geq 0$, (resp. $\Xi_-(P, Q; f) \leq 0$)
  - $\Xi_\pm(P, Q; f) = 0$ if and only if $P = Q$ a.s. or f is deterministic $P-$a.s.
  - Divergence contains information on QoI f, e.g.
  - Linearization – SA bounds:
    $\Xi_+(P, Q; f) = \sqrt{\mathrm{Var}_P[f]}\sqrt{2\mathcal{R}_{KL}(Q \mid P)} + \mathcal{O}(\mathcal{R}_{KL}(Q \mid P))$
- Variational representation: $\Xi_+(P, Q : f) = \inf_{\alpha>0}\left\{\frac{1}{\alpha}\Lambda_{P,f}(\alpha) + \frac{1}{\alpha}\mathcal{R}_{KL}(P \mid Q)\right\}$
- To date the only bound scaling with system size N, time T.

[1] K., L. Rey-Bellet, J. Wang, '16; Dupuis, M. K., Pantazis, Plechac *SIAM UQ* '16

Screening Strategy - Information Metrics & Observables

- Biological network describing Epidermal Growth Factor Receptor. *[Kholodenko et.al., J. Biol. Chem., 1999]*
  Data from: `http://www.ebi.ac.uk/biomodels-main/BIOMD0000000048`
- 47 reactions, 23 species, 23 observables, 50 parameters, $23 \times 50 = 1150$ sensitivities





Screening model sensitivities based on the path-FIM upper bound:

$$|S_f(\theta_k)| \leq \sqrt{\sum_{i=-\infty}^{\infty} A_f(i) \sqrt{\mathbf{F}_{\mathcal{H}}(Q^\theta)_{k,k}}}$$

Cramer-Rao inequality on path space

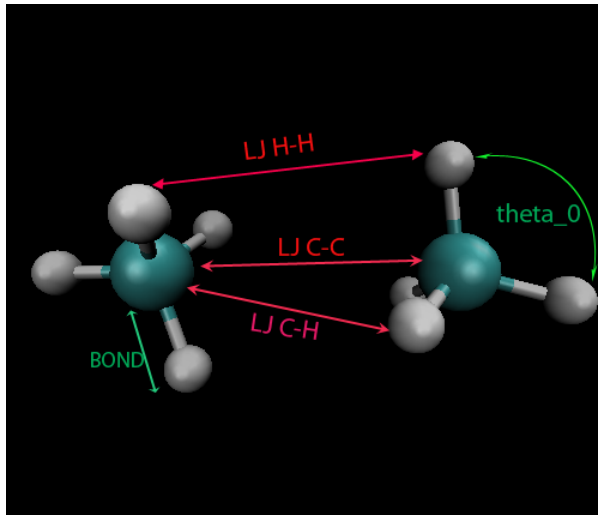
[1] Arampatzis, Pantazis, Katsoulakis *PLOS 1*, (2015); Dupuis, Katsoulakis, Pantazis, Plechac *SIAM UQ*, (2016)

Sensitivity for Langevin dynamics

- Potential: $V(q) = V_{angle}(q) + V_{bond}(q) + V_{LJ}(q)$
- Bond potential: $V_{bond}(r_{ij}) = \frac{1}{2}K_b(r_0 - r_{ij})^2$
- Angle potential: $V_{angle}(\theta_{ijk}) = \frac{1}{2}K_\theta(\theta_0 - \theta_{ijk})^2$
- Interaction potential:

$$V_{LJ}(r_{ij}) = \begin{cases} 4\epsilon_{LJ}\left[\left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{6}\right] & \text{if } r_{ij} < r_{cut}. \\ 0 & \text{otherwise.} \end{cases}$$

- Parameter vector:
$\theta = [\epsilon_{LJ}^{C-C}, \sigma_{LJ}^{C-C}, \epsilon_{LJ}^{C-H}, \sigma_{LJ}^{C-H}, \epsilon_{LJ}^{H-H}, \sigma_{LJ}^{H-H}, K_b, r_0, K_\theta, r_\theta]^T$



- Relative Entropy Rate: (related to force-matching in coarse-graining)

$$\mathcal{H}(Q^\theta | Q^{\theta+\epsilon}) = \frac{1}{2}\mathbb{E}_{\mu^\theta}[(F^{\theta+\epsilon}(q) - F^\theta(q))^T(\sigma\sigma^T)^{-1}(F^{\theta+\epsilon}(q) - F^\theta(q))]$$

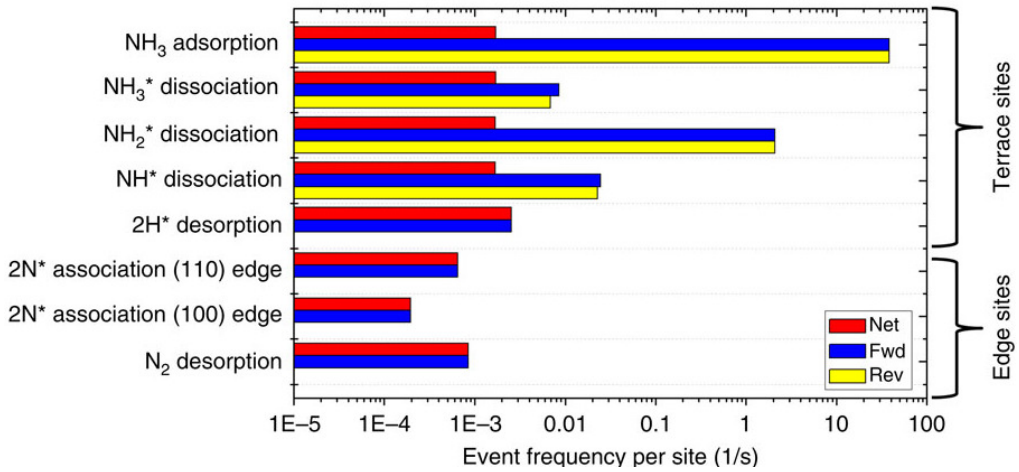
- Fisher Information Matrix:

$$F_{\mathcal{H}}(Q^\theta) = \mathbb{E}_{\mu^\theta}[\nabla_\theta F^\theta(q)^T(\sigma\sigma^T)^{-1}\nabla_\theta F^\theta(q)] \ ,$$

[1] V. Harmandaris, A. Tsourtis, M.K., Pantazis *J. Chem. Phys.* (2015)

Rare events and multi-scale problems

- Patched bimetallic catalysts for ammonia decomposition
- Ammonia decomposition: $2NH_3 \rightarrow N_2 + 3H_2$
- Multiple time scales:



Steady-state statistics indicating the frequency of elementary reactions and the flow of reaction flux. The net, forward and reverse rates in red, blue and yellow bars, respectively. Reactions with equal forward and reverse bars (for example, NH3 adsorption/desorption) are in partial equilibrium. Diffusion steps are not depicted, because of being fast and equilibrated, and dehydrogenations at edges are left out from the graph due to their negligible contribution.

W. Guo, D. G. Vlachos, Nature Communications (2015)

Rare events and multi-scale problems

An elementary example:

$$dX^\epsilon = b[X^\epsilon, Y^\epsilon]dt\,, \quad dY^\epsilon = \frac{1}{\epsilon}B(X^\epsilon, Y^\epsilon)dt + \frac{1}{\epsilon^{1/2}}\Sigma(X^\epsilon, Y^\epsilon)dW_t\,,$$

As $\epsilon \to 0$, $X^\epsilon \to \bar{X}$ in $[0, T]$:

$$d\bar{X}_t = \bar{b}[\bar{X}_t]dt\,, \quad \text{where} \quad \bar{b}[x] = \int b[x, y]\mu_x(dy)\,,$$

- $\mu_x(dy)$: invariant measure of fast equation after fixing the value of $X_t^\epsilon = x$.
- The averaged equation is deterministic and is derived as a limit in fixed time intervals $[0, T]$.
- For long time integration *rare events* appear which are not captured by the deterministic limit $\bar{X}$.

Sensitivity analysis for rare events

- An event A is rare if $P(A) \ll 1$. We usually consider $\log P(A)$.
  - Other applications to: reliability analysis, queueing theory, operation research, insurance, statistical mechanics, etc.

- Sensitivity analysis for rare events:

$$S_A(P^\theta) := \partial_\theta \log P^\theta(A) = \frac{\partial_\theta P^\theta(A)}{P^\theta(A)}$$

- Relative entropy is NOT the most appropriate divergence:
- Rare event bounds based on Renyi divergence:

$$\log Q(A) - \log P(A) \sim \mathcal{R}_\alpha (Q \,|\, P)$$

- Renyi divergence: $\mathcal{R}_\alpha (Q \,|\, P) := \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{dQ}{dP} \right)^\alpha \right]$

[1] Atar, Chowdhary and Dupuis, SIAM UQ '14

Sensitivity analysis for rare events

- An event A is rare if $P(A) \ll 1$. We usually consider $\log P(A)$.
  - Other applications to: reliability analysis, queueing theory, operation research, insurance, statistical mechanics, etc.

- Sensitivity analysis for rare events:

$$S_A(P^\theta) := \partial_\theta \log P^\theta(A) = \frac{\partial_\theta P^\theta(A)}{P^\theta(A)}$$

- Relative entropy is NOT the most appropriate divergence:
- Rare event bounds based on Renyi divergence:

$$\log Q(A) - \log P(A) \sim \mathcal{R}_\alpha(Q \mid P)$$

- Renyi divergence: $\mathcal{R}_\alpha(Q \mid P) := \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{dQ}{dP} \right)^\alpha \right]$

$$\frac{1}{\alpha - 1} \log Q(A) - \frac{1}{\alpha} \mathcal{R}_\alpha(Q \mid P) \leq \frac{1}{\alpha} \log P(A) \leq \frac{1}{\alpha + 1} \log Q(A) + \frac{1}{\alpha + 1} \mathcal{R}_{\alpha+1}(P \mid Q)$$

[1] Atar, Chowdhary and Dupuis, SIAM UQ '14

Sensitivity analysis for rare events

- Sensitivity bounds ($Q = P^{\theta+\epsilon}$, $P = P^{\theta}$ and $\alpha = \frac{1}{\epsilon}(\alpha_0 + O(\epsilon))$):

$$\sup_{\alpha>0} -\frac{\bar{H}(-\alpha) - \log P^{\theta}(A)}{\alpha} \leq S_A(P^{\theta}) \leq \inf_{\alpha>0} \frac{\bar{H}(\alpha) - \log P^{\theta}(A)}{\alpha}$$

- $\bar{H}(\alpha) := \log \mathbb{E}_{P^{\theta}}[\exp\{\alpha \partial_{\theta} \log P^{\theta}\}]$
  Renyi-like quantity: cumulant generating function of $\partial_{\theta} \log P^{\theta}$.
- Sensitivity screening bounds involve the calculation of a rare event.
- Extensions to bound rate function derivatives (Large Deviations, Moderate Deviations, etc).
- General **Risk-Sensitive** observables for characterizing rare events:

$$\log \int_{\mathcal{X}} e^{F} dP$$

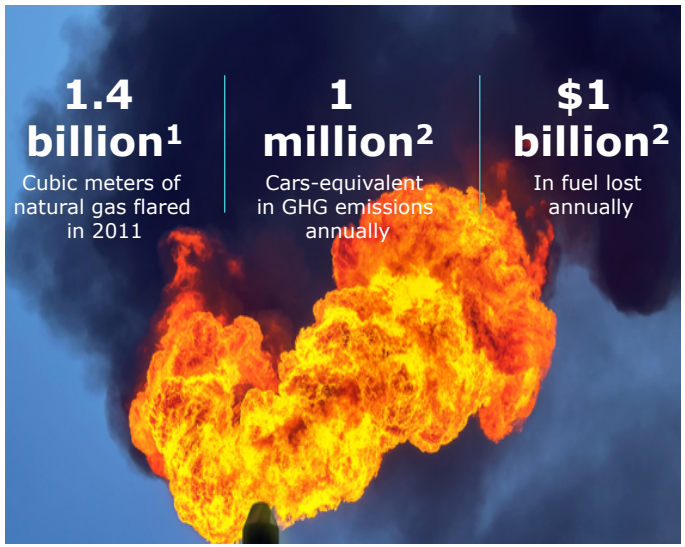
[1] P. Dupuis, M. Katsoulakis, Y. Pantazis and L. Rey-Bellet (2016)

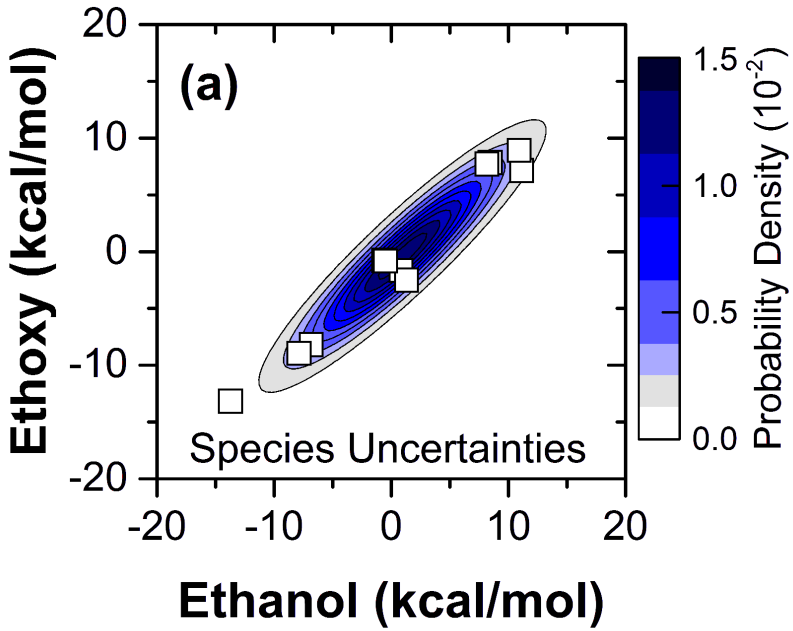
Uncertainty and Correlations in DFT Calculations

- DFT uncertainties
  - Arise from approximate exchange-correlation functional; vary significantly with functional
- Chemical kinetics parameters are correlated
  - Introduce correlative UQ for reaction networks
- Employ Hierarchical Bayesian methods to create probabilistic model.

Example: Ethanol Steam Reforming (ESR) for hydrogen production

Steam reforming is a catalysis-based method for producing hydrogen or other useful products from fuels such as natural gas. E.g. can convert flared ("stranded") gas such as methane into a more practical source of energy.





Correlative SA & UQ - Mathematical Formulation

- Necessary to consider ensembles of models (Bayesian formulation):

$$P(X|\theta)p(\theta)\,,$$

  - $P(X|\theta)$: Forward model, i.e. pdf of state X for fixed parameter θ (reaction network, KMC, etc).

    Observables: $f(\theta) = \int h(X)P(X|\theta)dX\,,$

  - $p(\theta)$: distribution of θ.

- Correlations between the parameter sets $\theta = (\theta_1, \theta_2)$:

$$p(\theta_1, \theta_2) = p(\theta_2|\theta_1)p(\theta_1)\,, \qquad p(\theta_1)\text{: marginal}$$

- $\nabla_1 F(\theta_1)$: Local correlative sensitivity index
  Fix θ_1 and consider correlations with θ_2:

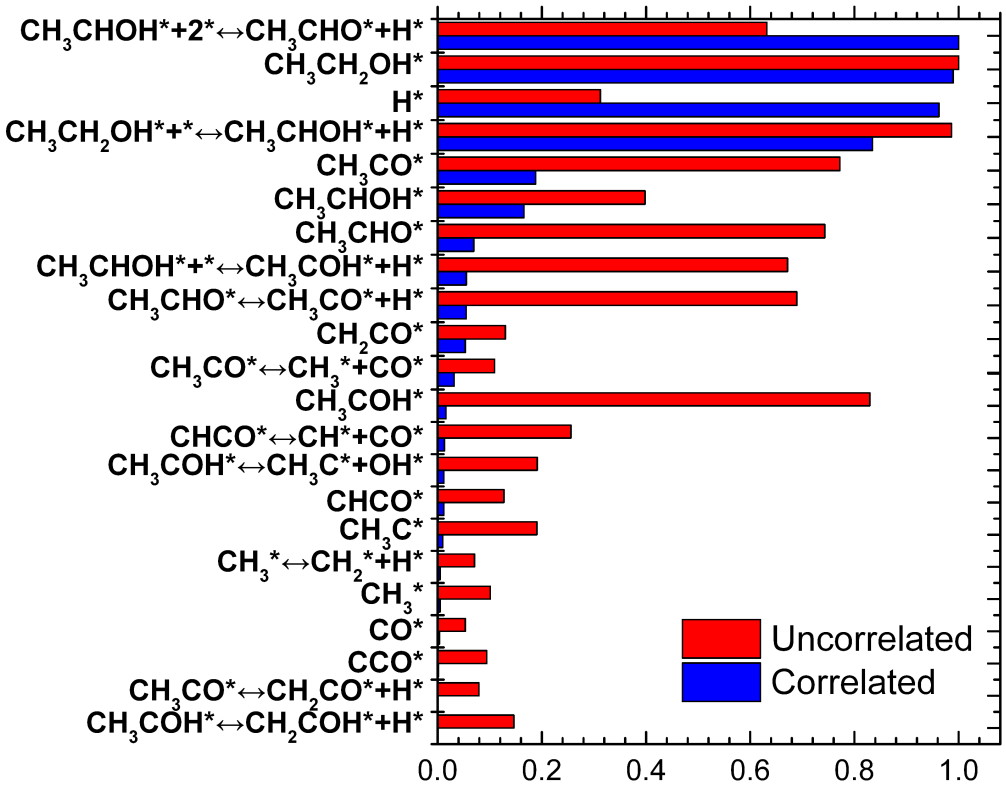
$$F(\theta_1) = \int f(\theta_1, \theta_2)p(\theta_2|\theta_1)d\theta_2$$

- Global correlative SI: $\int_{\Theta 1} |\nabla_1 F(\theta_1)|p(\theta_1)d\theta_1$
- We build on the mathematical tools developed earlier: UQ Information Inequalities, pFIM, etc.

[1] Sutton, Guo, Katsoulakis, and Vlachos, *Nature Chemistry*, to appear (2016)

Correlations Strongly Impact Predictions

- Apply the methodology to a complex reaction network of ethanol steam reforming on a Pt/Al_2O_3 catalyst with 67 species and 160 reactions.
- Correlations reduce significantly the number of important parameters
- Sensitivity of key parameters differ when correlations are included



[1] Sutton, Guo, Katsoulakis, and Vlachos, *Nature Chemistry*, to appear (2016)

Conclusions

- SA methods for mesoscale models: stochastic (Langevin) and possibly discrete systems (KMC, reaction-networks).
- Sensitivity screening for molecular models with a very large number of parameters
- Tackle non-equilibrium stochastic processes:
  - Coupled mechanisms (reaction-diffusion), reaction networks, driven systems, multi-physics models, etc.
- Methods for parameters which are strongly correlated (e.g. due to thermodynamic constraints, DFT calculations, etc) $\Longrightarrow$ type of "global" SA.
- Transient and long-time integration (numerical/approximation challenges).
- Some information-based methods for UQ and sensitivity analysis of rare events; Renyi methods

References

- **Sensitivity Analysis in Path Space (information-theoretic, goal-oriented)**

  A Relative Entropy Rate Method for Path Space Sensitivity Analysis of Stationary Complex Stochastic Dynamics, Y. Pantazis, M.K., *J. Chem. Phys.* (2013).

  Parametric Sensitivity Analysis for Biochemical Reaction Networks based on Pathwise Information Theory, M. K., D. Vlachos, Y. Pantazis, *BMC Bioinformatics*, (2013).

  Measuring the irreversibility of numerical schemes for reversible stochastic differential equations, M. K. Y. Pantazis, L. Rey-Bellet., *ESAIM: Math. Model. Num. Analysis*, (2014).

  Goal-oriented sensitivity analysis for lattice kinetic Monte Carlo simulations, M. K. and G. Arampatzis, *J. Chem. Phys.*, (2014).

  Accelerated Sensitivity Analysis for High-Dimensional Stochastic Reaction Networks, G. Arampatzis, Y. Pantazis, M. K. *PLOS1* (2015).

  Sensitivity Bounds and Error Estimates for Stochastic Models, P. Dupuis, M. K., P. Plechac, Y. Pantazis *SIAM UQ* (under revision) (2015).

  Parametric Sensitivity Analysis for Stochastic Molecular Systems using Pathwise Information Metrics, A. Tsourtis, V. Harmandaris, M. K., Y.Pantazis *J. Chem. Phys.* (2015).

- **Coarse-graining (path-space, multi-body effects, non-equilibrium)**

  Information-theoretic tools for parametrized coarse-graining of non-equilibrium extended systems, M. K., P. Plechac, *J. Chem. Phys.* (2013).

  Coarse-graining schemes for stochastic lattice systems with short and long-range interactions, M. K., P. Plechac, L. Rey-Bellet and D. Tsagkarogiannis, *Math. Comp.*, (2014).

  Spatial two-level interacting particle simulations and information theory-based error quantification, E. Kalligiannaki, M. K. , P. Plechac *SIAM Sci. Comp.*, 36, A634A667 (2014).

  Measuring the irreversibility of numerical schemes for reversible stochastic differential equations, M. K. Y. Pantazis, L. Rey-Bellet., *ESAIM: Math. Model. Num. Analysis*, (2014).

  Path-space variational inference for non-equilibrium coarse-grained systems, V. Harmandaris, E. Kalligiannaki, M. K. and P. Plechac, *J. Comp. Phys.*, submitted, (2015).

See also: Markos Katsoulakis' Homepage, ResearchGate Profile