Bayesian Learning of Multifidelity Surrogate Networks (MFNets): MCMC and Variational Inference Approaches

SIAM Conference on Uncertainty Quantification
April 12, 2022, Atlanta, Georgia

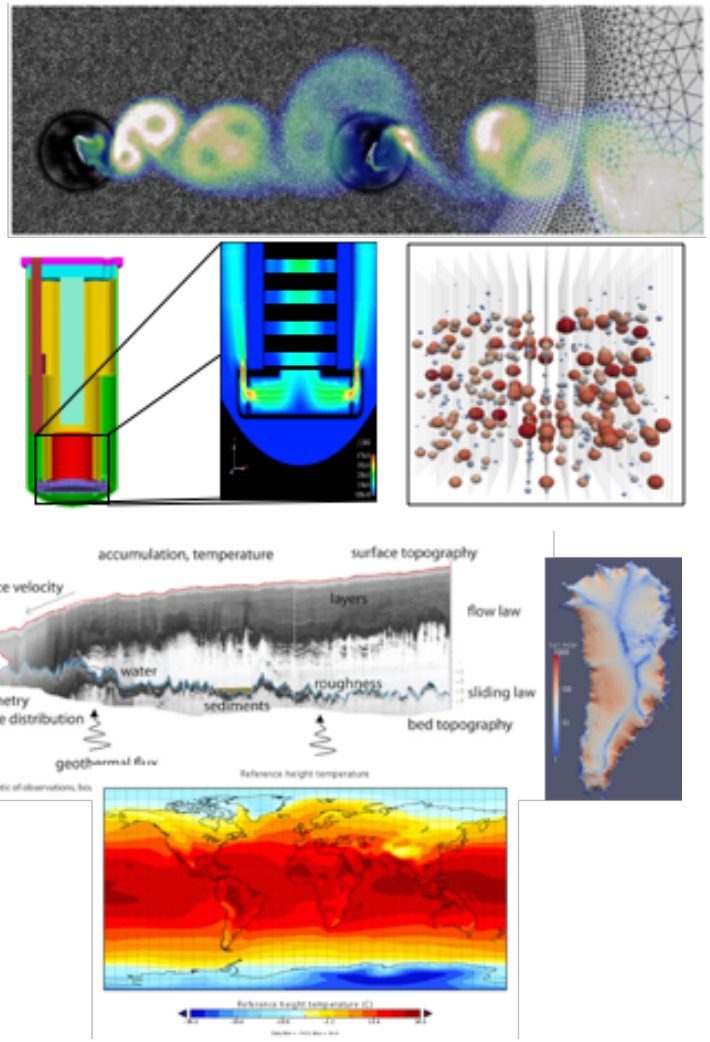
Alex Gorodetsky, Trung Pham, and John Jakeman

Aerospace Engineering, University of Michigan and
Sandia National Laboratories

UNIVERSITY OF MICHIGAN

Uncertainty quantification for high-fidelity models

- Determine uncertainty on HF models
  - Expectation, variance
  - Probability of failures
- Severe simulation budget constraints
  - High dimensional PDEs
  - Large-scale computing resources
- Model complexity increases dimensionality
- HF models $\rightarrow$ UQ more important
  - Less validation, study, and analysis
  - Greater exploitation of nonlinearities
  - Uncertainty due to model form, initial conditions, and operating conditions

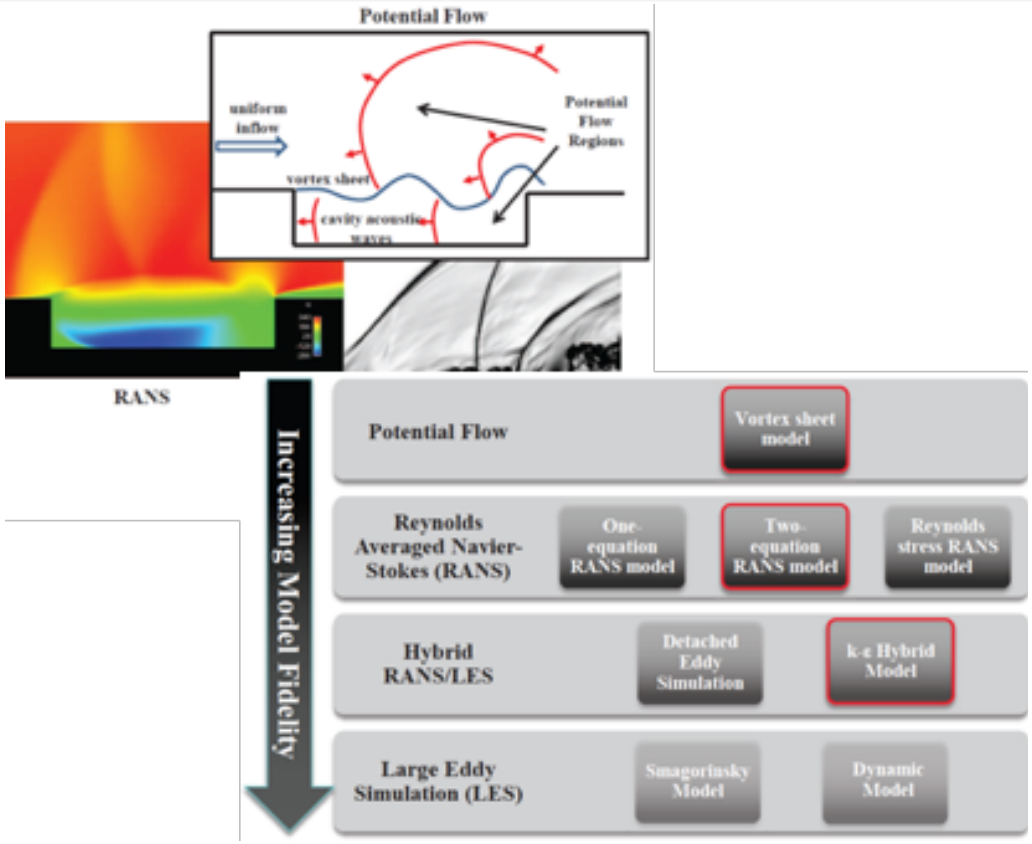


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data

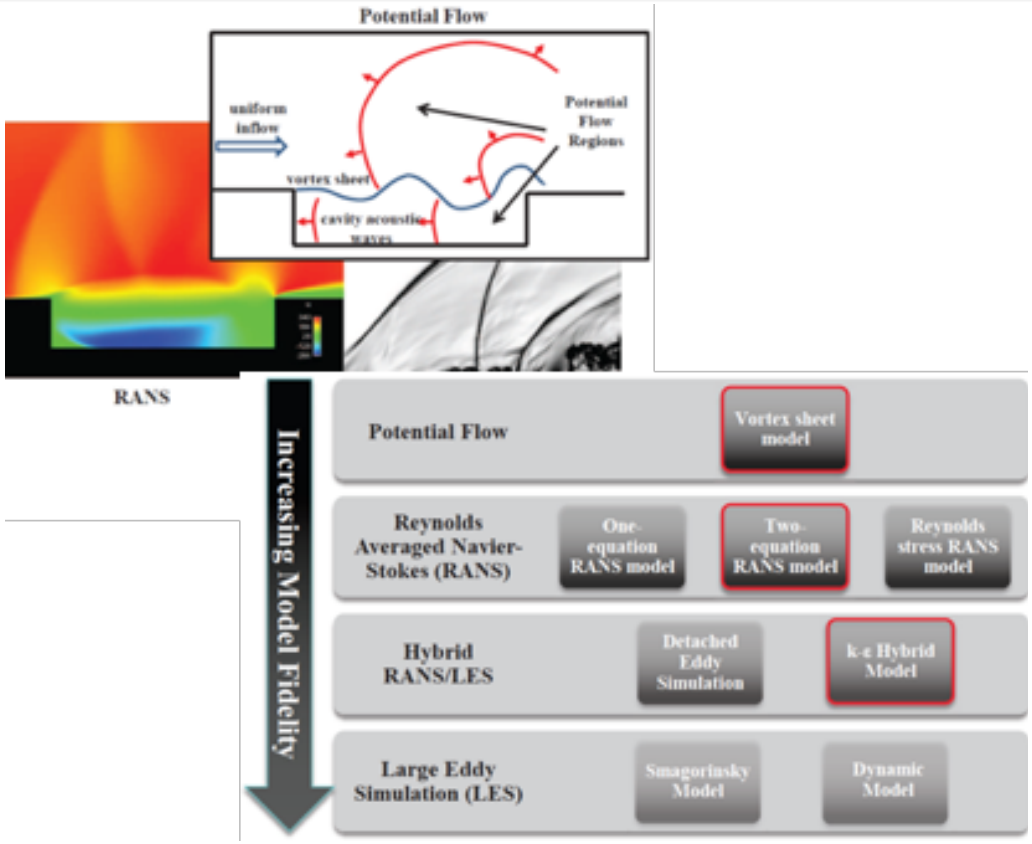


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data
- Challenges

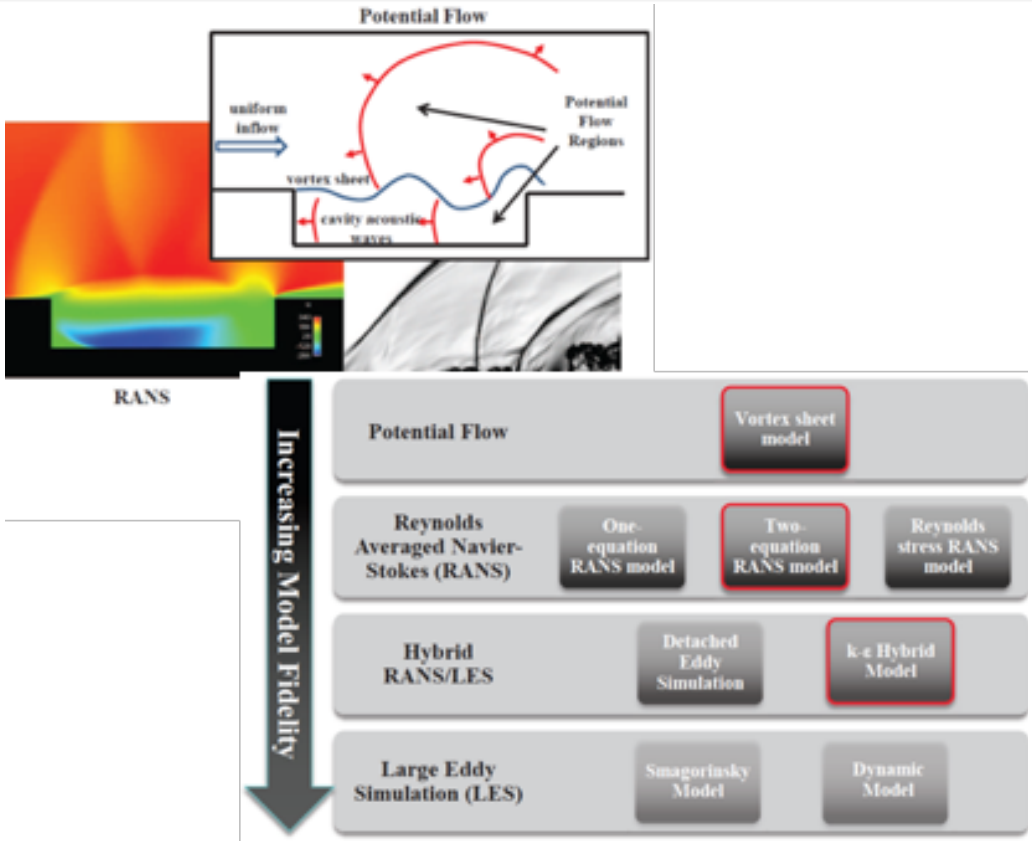


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data
- Challenges
  - Models with varying inputs/outputs

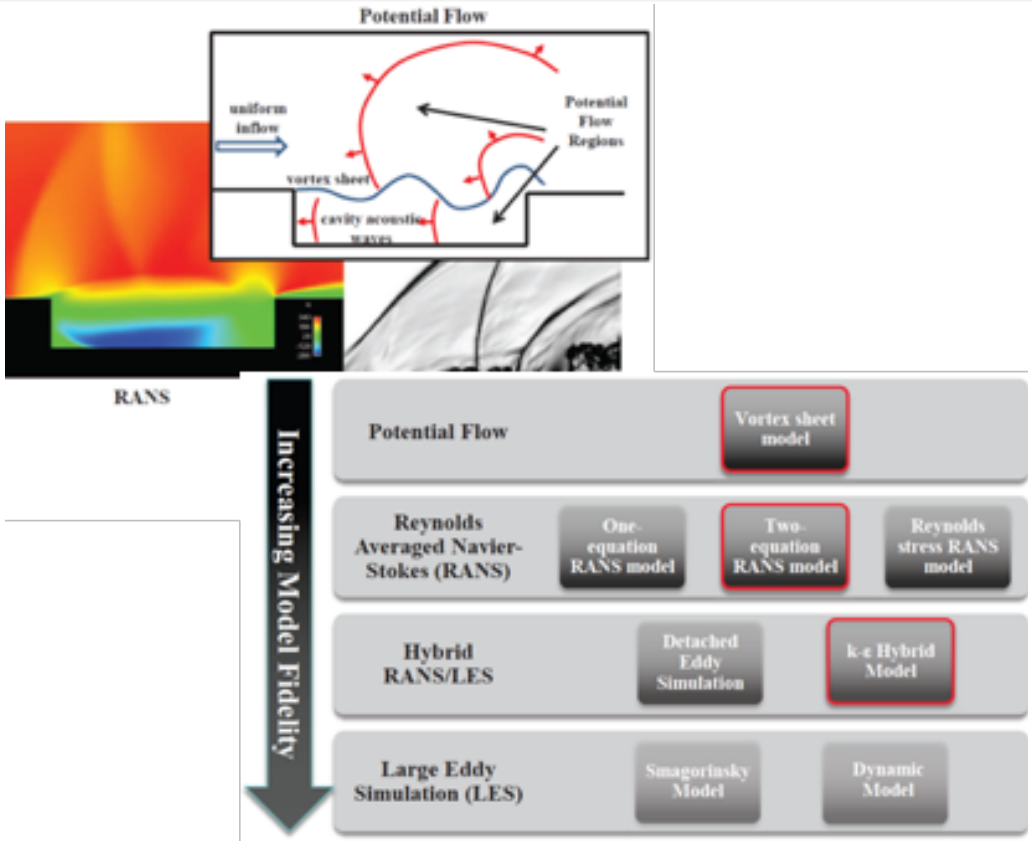


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data
- Challenges
  - Models with varying inputs/outputs
  - Legacy data collected separately

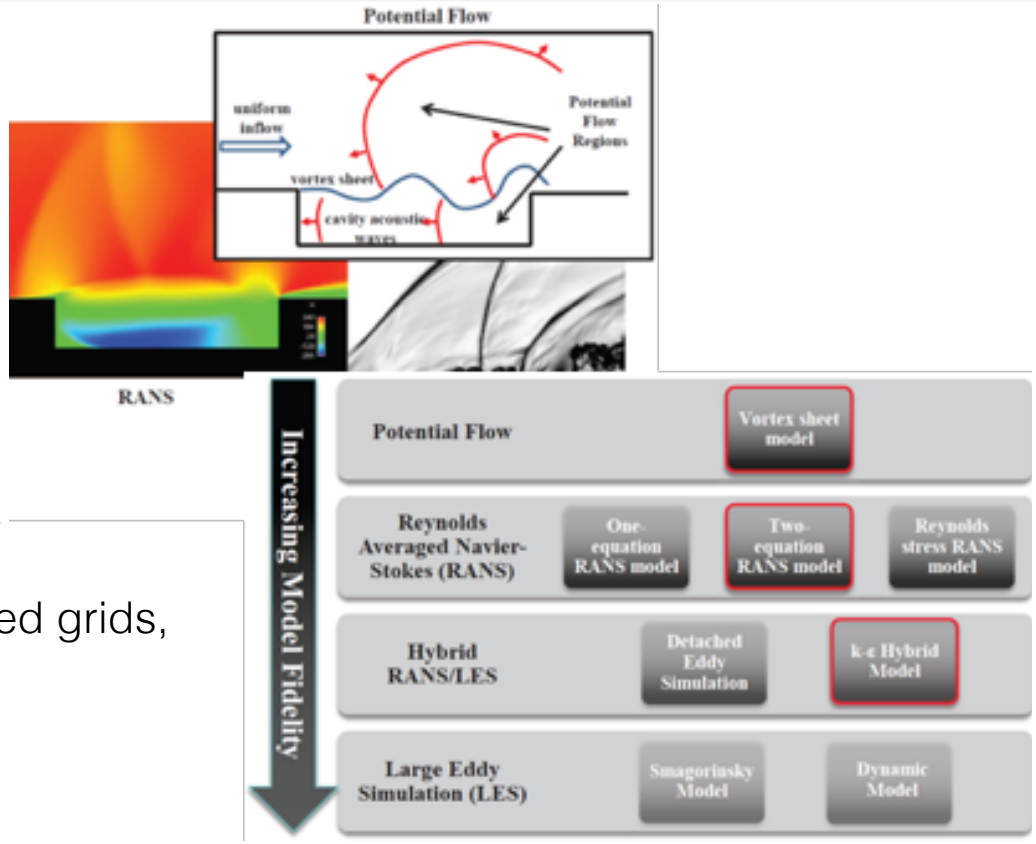


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data
- Challenges
  - Models with varying inputs/outputs
  - Legacy data collected separately
  - Corrupted evaluations: unconverged grids, unexplored parameters

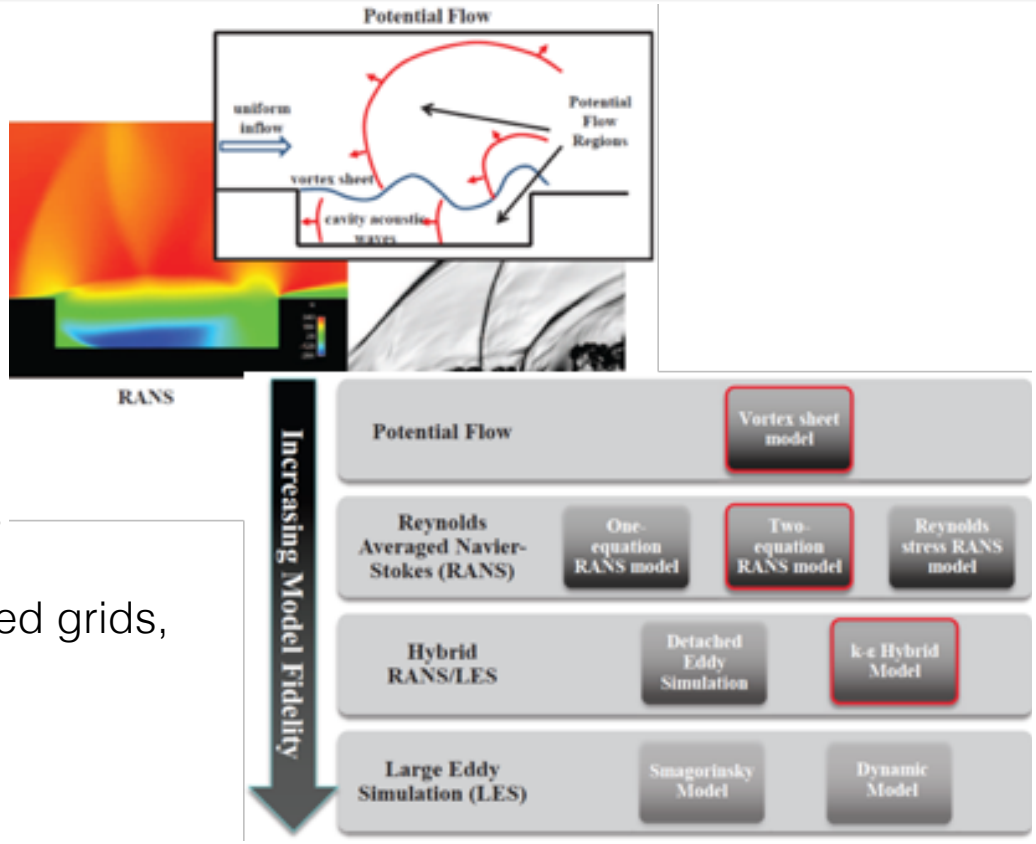


Leveraging multiple models/simulation sources

Key idea

We almost never have just ***one*** model, instead we develop many over the course of a study/analysis. Can we leverage these?

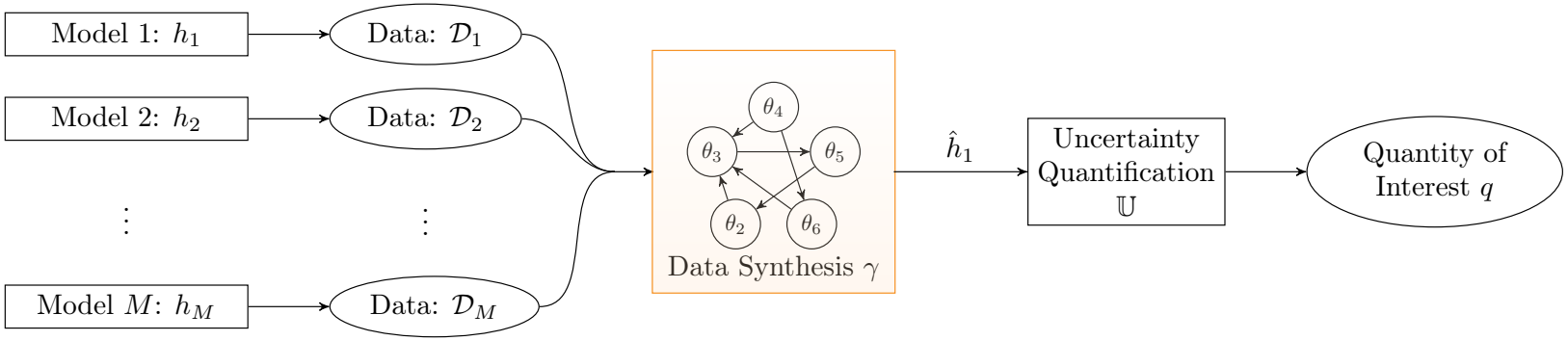
- Sources
  - Hierarchy of fidelities
  - Ensemble of peer models
  - Discretization levels
  - Experimental data
- Challenges
  - Models with varying inputs/outputs
  - Legacy data collected separately
  - Corrupted evaluations: unconverged grids, unexplored parameters
  - Lack of assumed relationships



Contribution in this talk

Problem: *A-priori* structural assumptions (hierarchical or otherwise) between models can lack robustness and limit efficiency

Contribution: A new modeling framework for multi-fidelity surrogate models that can flexibly adapt to variety of situations



Outline

- Multifidelity methods for enhancing surrogate approaches
  - Overview of existing approaches
  - A graph-based formulation: MFNets
  - Examples
- Probabilistic MFNets

The standard approach: autoregressive (recursive) modeling

Two-model case

1. A single node represents a surrogate for some information source

The standard approach: autoregressive (recursive) modeling

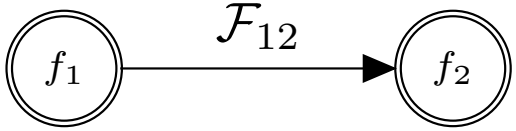
Two-model case

1. A single node represents a surrogate for some information source
2. A new (higher fidelity) information source is introduced

The standard approach: autoregressive (recursive) modeling

Two-model case

1. A single node represents a surrogate for some information source
2. A new (higher fidelity) information source is introduced
3. A recursive relationship is established, by passing f_1 through some functional $\mathcal{F}_{12}$

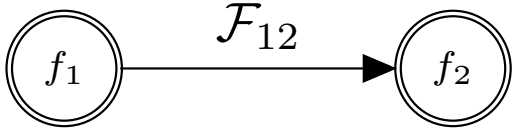


$$f_2 = \mathcal{F}_{12}[f_1] + \delta_2$$

The standard approach: autoregressive (recursive) modeling

Two-model case

1. A single node represents a surrogate for some information source
2. A new (higher fidelity) information source is introduced
3. A recursive relationship is established, by passing f_1 through some functional $\mathcal{F}_{12}$
4. The discrepancy is denoted by δ_2, hope is that it is "simple"
   - Recursive co-kriging: Kennedy and O'Hagan, Gratiet and Garnier

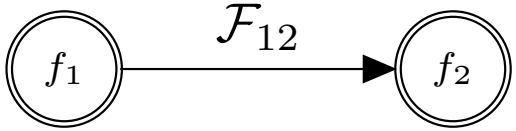


$$\delta_2 = f_2 - \mathcal{F}_{12}[f_1]$$

The standard approach: autoregressive (recursive) modeling

Two-model case

1. A single node represents a surrogate for some information source
2. A new (higher fidelity) information source is introduced
3. A recursive relationship is established, by passing f_1 through some functional $\mathcal{F}_{12}$
4. The discrepancy is denoted by δ_2, hope is that it is "simple"
   - Recursive co-kriging: Kennedy and O'Hagan, Gratiet and Garnier



$$\delta_2 = f_2 - \mathcal{F}_{12}[f_1]$$

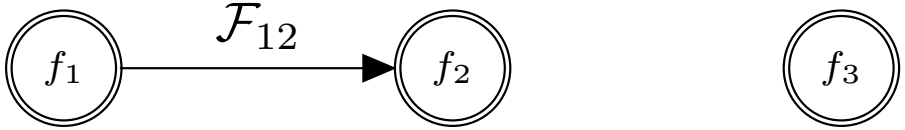
Learning Requirements

- Things to learn: f_1, $\mathcal{F}_{12}$, δ_2
- If any are nonlinearly parameterized, then not a Gaussian process!

The standard approach: autoregressive (recursive) modeling

General case

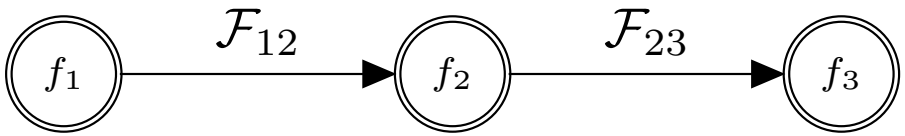
1. Introduce a third (higher fidelity) information source



The standard approach: autoregressive (recursive) modeling

General case

1. Introduce a third (higher fidelity) information source
2. Again establish a recursive relationship

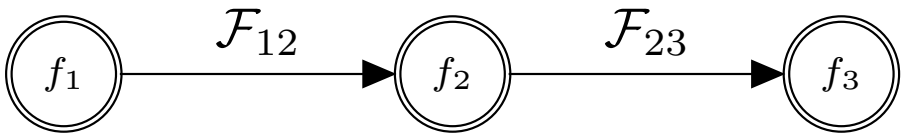


$$f_3 = \mathcal{F}_{23}[f_2] + \delta_3 = \mathcal{F}_{23}\left[\mathcal{F}_{12} f_1 + \delta_2\right] + \delta_3$$

The standard approach: autoregressive (recursive) modeling

General case

1. Introduce a third (higher fidelity) information source
2. Again establish a recursive relationship
3. Again learn the discrepancy

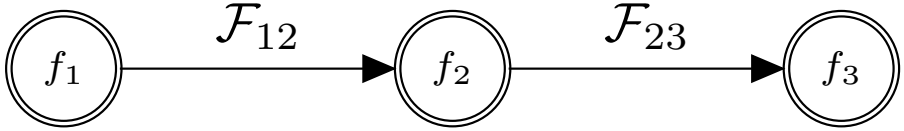


$$\delta_3 = f_3 - \mathcal{F}_{23}[f_{23}]$$

The standard approach: autoregressive (recursive) modeling

General case

1. Introduce a third (higher fidelity) information source
2. Again establish a recursive relationship
3. Again learn the discrepancy



$$\delta_3 = f_3 - \mathcal{F}_{23}[f_{23}]$$

Challenges

- Many existing techniques require nested sample points and step-wise optimization [Gratiet 2014, Perdikaris 2017]
- Many existing techniques that assume separable $\mathcal{F}_{ij}[f_i] = \rho_{ij}(x) f_i(x)$ treat parameters of ρ_{ij} as hyperparameters that are optimized separately to retain Gaussianity.

Why do we need another paradigm?

1. Not all information sources are obviously hierarchical
   - Depends on QoI
   - Peer models are common
   - Discretizations and numerical issues affect hierarchies

Why do we need another paradigm?

1. Not all information sources are obviously hierarchical
   - Depends on QoI
   - Peer models are common
   - Discretizations and numerical issues affect hierarchies
2. In the sampling context, moving beyond hierarchical has shown significant benefits to extract more correlations
   - MIMC [Haji-Ali 2016] vs. MLMC [Giles 2008]
   - Generalized ACV [Gorodetsky 2020] vs. recursive control variate approaches (MFMC) [Peherstorfer 2016]
   - Multilevel Blue [Schaden and Ullman 2020]

Why do we need another paradigm?

1. Not all information sources are obviously hierarchical
   - Depends on QoI
   - Peer models are common
   - Discretizations and numerical issues affect hierarchies
2. In the sampling context, moving beyond hierarchical has shown significant benefits to extract more correlations
   - MIMC [Haji-Ali 2016] vs. MLMC [Giles 2008]
   - Generalized ACV [Gorodetsky 2020] vs. recursive control variate approaches (MFMC) [Peherstorfer 2016]
   - Multilevel Blue [Schaden and Ullman 2020]
3. Need more training algorithms to consider non-GP cases and non-nested sampling

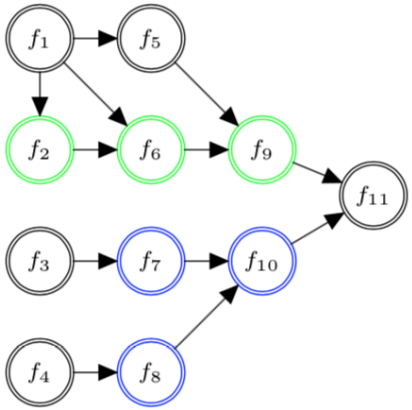
Multifidelity Networked Surrogate Models (MFNets)

UNIVERSITY OF MICHIGAN

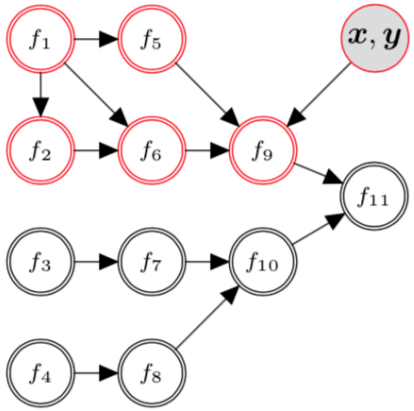
Main idea: model relationships between surrogate outputs

An MFNet is a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes representing information sources $\mathcal{V} = \{f_1, \ldots, f_M\}$ and directed edges $\mathcal{E} = \{(j \to i)\}$ representing their relationships

$$f_i(x) = \sum_{j \in \text{pa}(i)} \mathcal{F}_{ji}[f_j(x)] + \delta_i(x).$$



(a) Sample structure of a multifidelity surrogate.



(b) Evaluating (10) for $k = 9$ requires data from f_9 and traversing the ancestors of f_9 (depicted in red).

Figure 1: An example DAG used to define a multifidelity surrogate. This structure exhibits a complicated relationship between each function and the high-fidelity f_{11}. Both hierarchical and peer relationships are exhibited within these networks. For instance the left panel shows an example of hierarchical structure $(f_2 \to f_6 \to f_9)$ in green and example of peer structure $(f_7 \to f_{10}, f_8 \to f_{10})$ in blue.

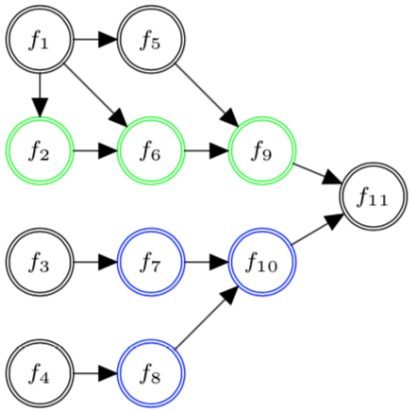
Multifidelity Networked Surrogate Models (MFNets)

UNIVERSITY OF MICHIGAN

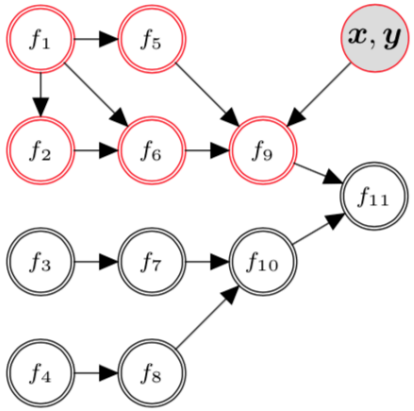
Main idea: model relationships between surrogate outputs

An MFNet is a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes representing information sources $\mathcal{V} = \{f_1, \ldots, f_M\}$ and directed edges $\mathcal{E} = \{(j \to i)\}$ representing their relationships

$$f_i(x) = \sum_{j \in \text{pa}(i)} \rho_{ji}(x) f_j(x) + \delta_i(x).$$



(a) Sample structure of a multifidelity surrogate.



(b) Evaluating (10) for $k = 9$ requires data from f_9 and traversing the ancestors of f_9 (depicted in red).

Figure 1: An example DAG used to define a multifidelity surrogate. This structure exhibits a complicated relationship between each function and the high-fidelity f_{11}. Both hierarchical and peer relationships are exhibited within these networks. For instance the left panel shows an example of hierarchical structure $(f_2 \to f_6 \to f_9)$ in green and example of peer structure $(f_7 \to f_{10}, f_8 \to f_{10})$ in blue.

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\,\beta_k\right\|_2^2$$

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\,\beta_k\right\|_2^2$$

  - Edge approximation $\rho_{jk} = W_{jk}^T(x)\alpha_{jk}$

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\,\beta_k\right\|_2^2$$

  - Edge approximation $\rho_{jk} = W_{jk}^T(x)\alpha_{jk}$
  - Ancestral parameters $\gamma_j = \{\alpha_{\ell i} : \ell, i \in \text{anc}(j)\} \cup \{\beta_\ell : \ell \in \text{anc}(j)\}$

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\,\beta_k\right\|_2^2$$

  - Edge approximation $\rho_{jk} = W_{jk}^T(x)\alpha_{jk}$
  - Ancestral parameters $\gamma_j = \{\alpha_{\ell i} : \ell, i \in \text{anc}(j)\} \cup \{\beta_\ell : \ell \in \text{anc}(j)\}$

$$-p_k\left(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k, \{\alpha_{jk}, \gamma_j; j \in \text{pa}(k)\}\right) = \frac{n}{2}\log 2\pi + n\log\sigma_k +$$

$$\frac{1}{2\sigma_k^2}\sum_{i=1}^{n}\left(y_k^{(i)} - \left(\sum_{j\in \text{pa}(k)} f_j(x^{(i)};\gamma_j)W_{jk}^T(x^{(i)})\alpha_{jk} + V_k^T\left(x^{(i)}\right)\beta_k\right)\right)^2$$

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\,\beta_k\right\|_2^2$$

  - Edge approximation $\rho_{jk} = W_{jk}^T(x)\alpha_{jk}$
  - Ancestral parameters $\gamma_j = \{\alpha_{\ell i} : \ell, i \in \text{anc}(j)\} \cup \{\beta_\ell : \ell \in \text{anc}(j)\}$

$$-p_k\left(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k, \{\alpha_{jk}, \gamma_j; j \in \text{pa}(k)\}\right) = \frac{n}{2}\log 2\pi + n\log\sigma_k +$$

$$\frac{1}{2\sigma_k^2}\sum_{i=1}^{n}\left(y_k^{(i)} - \left(\sum_{j\in\text{pa}(k)} f_j(x^{(i)};\gamma_j)W_{jk}^T(x^{(i)})\alpha_{jk} + V_k^T\left(x^{(i)}\right)\beta_k\right)\right)^2$$

- Independent exponential family prior $p(\theta) = g(\theta)\exp(\phi^T(\theta)\nu)$

Training: setting up the log posterior

- Log likelihood:
  - Root node approximation $f_k = V_k^T(x)\beta_k$

$$-\log p(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k) = \frac{n}{2}\log 2\pi + n\log\sigma_k + \frac{1}{2\sigma_k^2}\left\|\boldsymbol{y}_k - V_k^T(\boldsymbol{x})\beta_k\right\|_2^2$$

  - Edge approximation $\rho_{jk} = W_{jk}^T(x)\alpha_{jk}$
  - Ancestral parameters $\gamma_j = \{\alpha_{\ell i} : \ell, i \in \text{anc}(j)\} \cup \{\beta_\ell : \ell \in \text{anc}(j)\}$

$$-p_k\left(\boldsymbol{y}_k \mid \boldsymbol{x}; \beta_k, \{\alpha_{jk}, \gamma_j; j \in \text{pa}(k)\}\right) = \frac{n}{2}\log 2\pi + n\log\sigma_k +$$

$$\frac{1}{2\sigma_k^2}\sum_{i=1}^{n}\left(y_k^{(i)} - \left(\sum_{j\in\text{pa}(k)} f_j(x^{(i)};\gamma_j)W_{jk}^T(x^{(i)})\alpha_{jk} + V_k^T\left(x^{(i)}\right)\beta_k\right)\right)^2$$

- Independent exponential family prior $p(\theta) = g(\theta)\exp(\phi^T(\theta)\nu)$
- DAG structure enables fast and efficient backpropagation for gradient computation

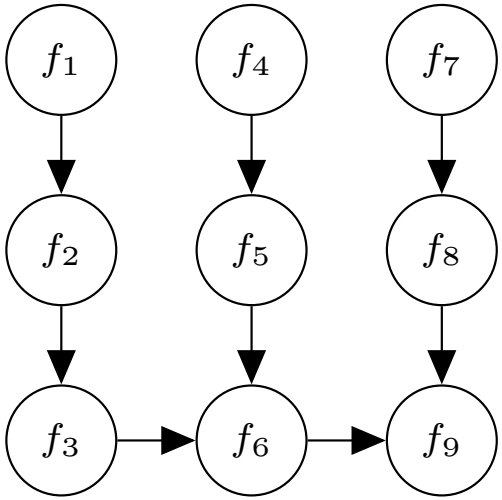
Synthetic example with stochastic models

$$f_k(x) = \left(2 + (2x_1^5 + 2x_2^5)\Delta_1 + 3x_1x_2 + (x_1^2 + x_2^2 + 5x_1^2x_2^2)\Delta_2 + 0.5x_1 + 0.5x_2\right)(1 + \mathbb{E}[\mathcal{N}(0,1)]),$$

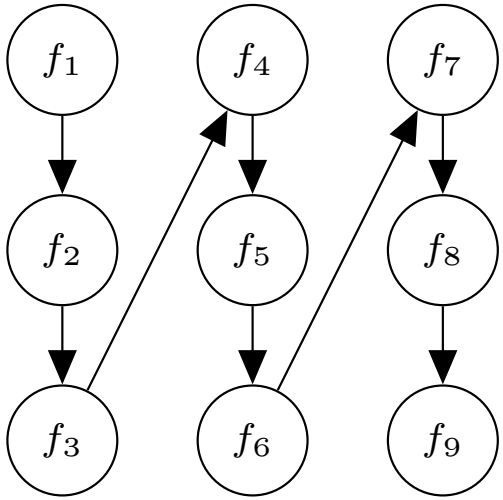
	Δ_1	Δ_2	N (Monte Carlo)
f_1	0	0	5
f_2	0	0	10
f_3	0	0	100
f_4	0	1	5
f_5	0	1	10
f_6	0	1	100
f_7	1	1	5
f_8	1	1	10
f_9	1	1	100

Synthetic example with stochastic models

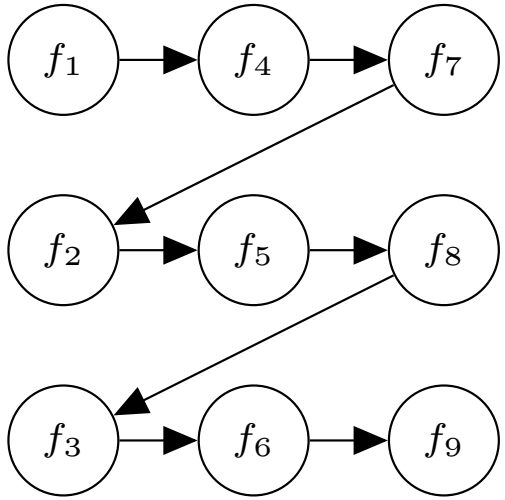
$$f_k(x) = \left(2 + (2x_1^5 + 2x_2^5)\Delta_1 + 3x_1x_2 + (x_1^2 + x_2^2 + 5x_1^2x_2^2)\Delta_2 + 0.5x_1 + 0.5x_2\right)(1 + \mathbb{E}[\mathcal{N}(0,1)]),$$



(a) Natural ordering



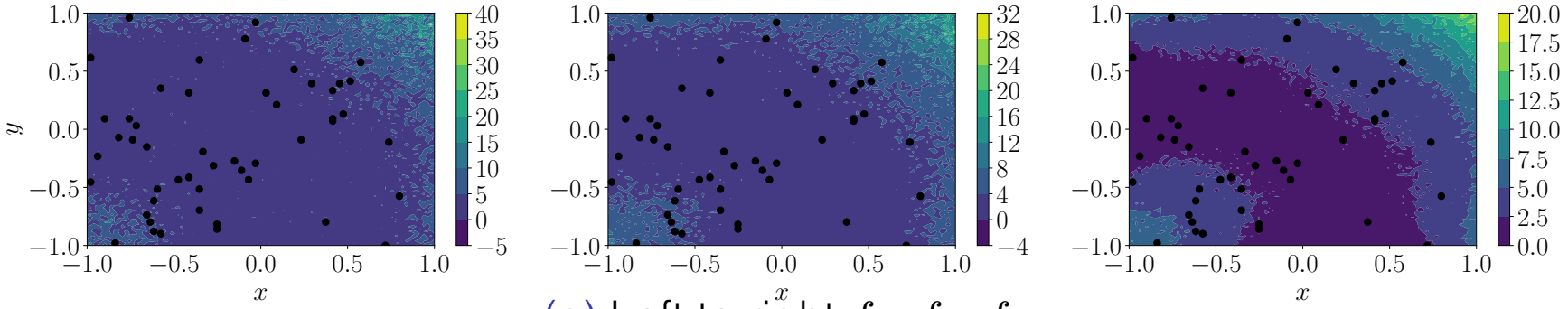
(b) Hierarchical ordering by model fidelity Δ first



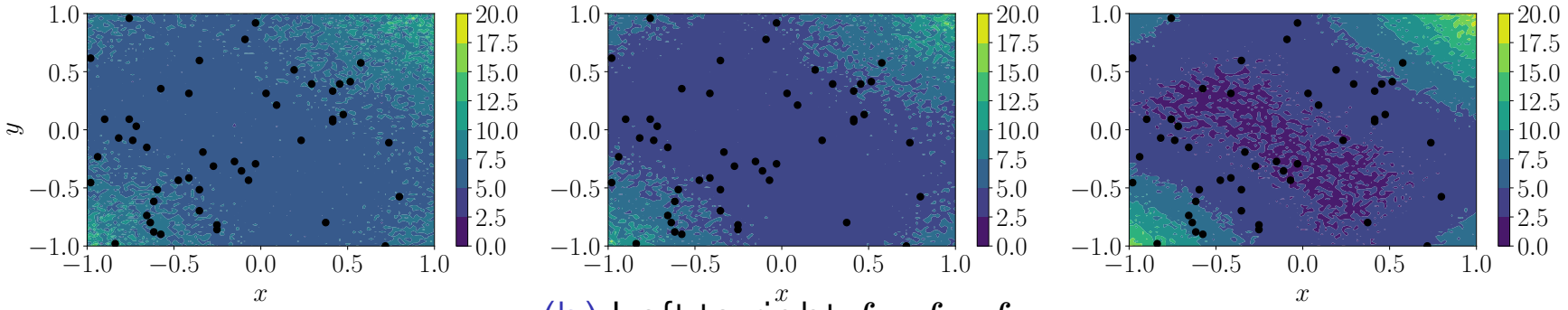
(c) Alternate Hierarchical ordering by noise fidelity N first.

Figure: Analytical noise test case: models' natural structure versus two candidate hierarchical orderings.

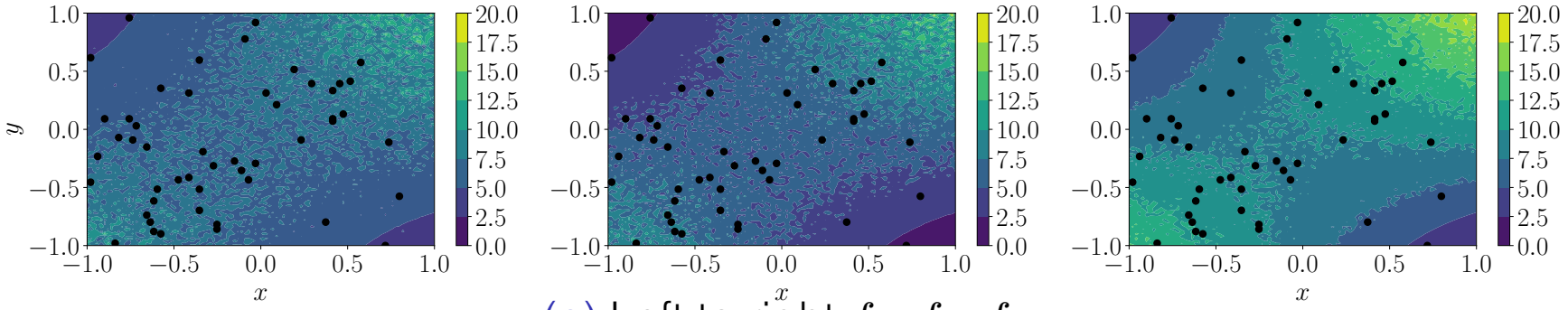
Synthetic response surfaces



(a) Left to right f_7, f_8, f_9



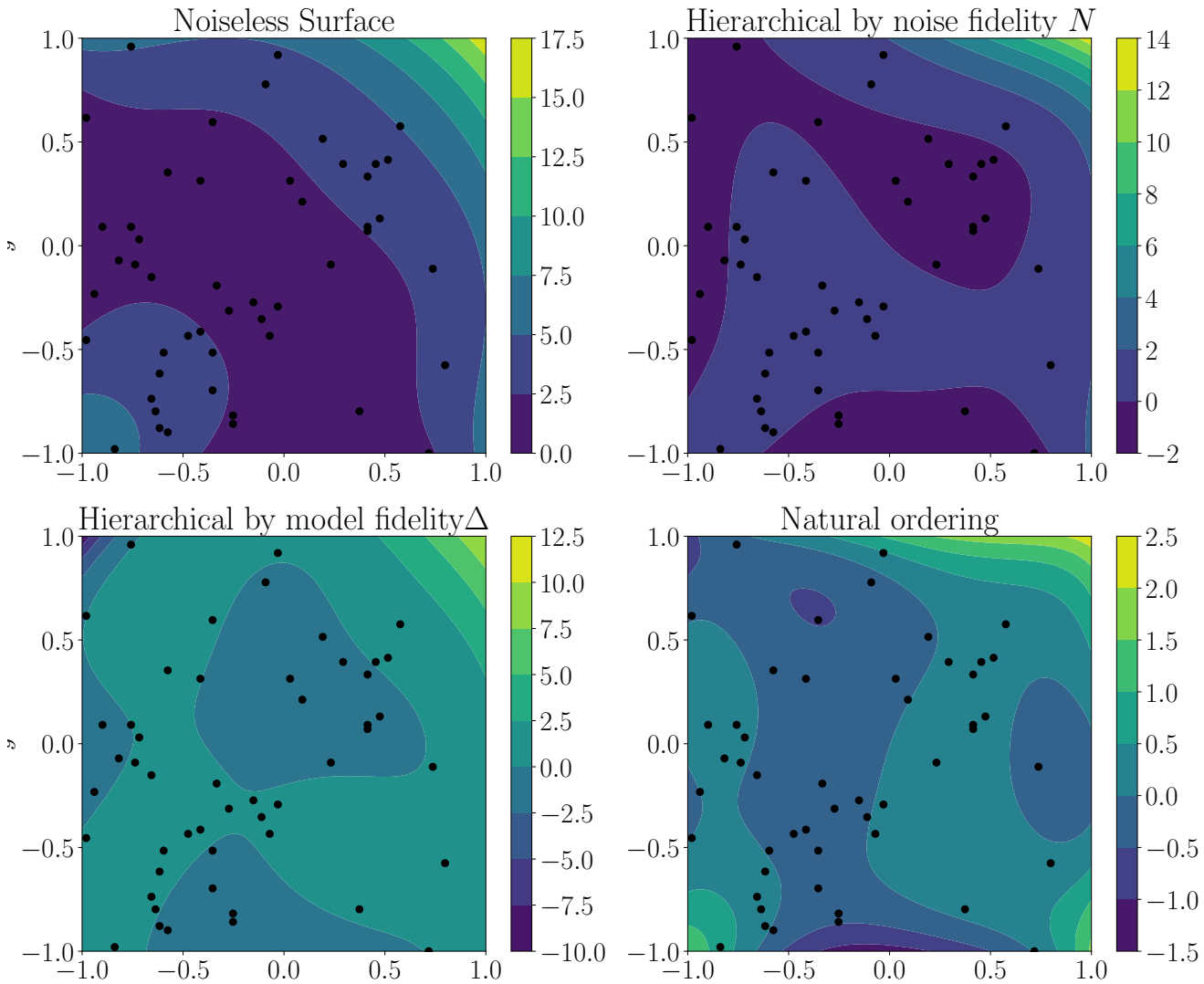
(b) Left to right f_4, f_5, f_6



(c) Left to right f_1, f_2, f_3

Highest fidelity model and pointwise errors

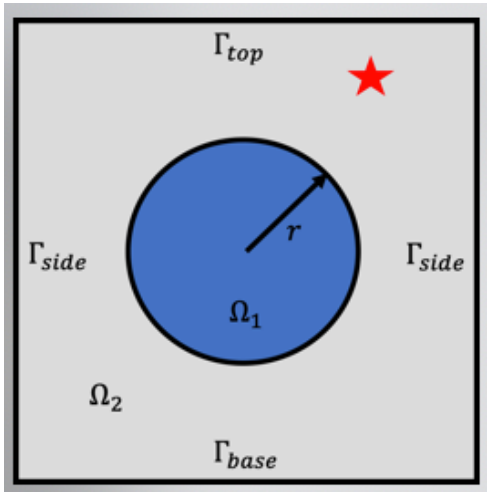
Natural ordering obtains order of magnitude lower errors



Thermal block

$$\begin{cases} -\text{div}(\kappa(\mu_0)\nabla u(\boldsymbol{\mu})) = 0 & \text{in } \Omega, \\ u(\boldsymbol{\mu}) = 0 & \text{on } \Gamma_{top}, \\ \kappa(\mu_0)\nabla u(\boldsymbol{\mu}) \cdot \mathsf{n} = 0 & \text{on } \Gamma_{side}, \\ \kappa(\mu_0)\nabla u(\boldsymbol{\mu}) \cdot \mathsf{n} = \mu_1 & \text{on } \Gamma_{base}. \end{cases}$$

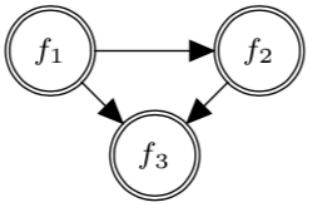
$$\kappa(\mu_0) = \begin{cases} \mu_0 & \text{in } \Omega_1, \\ 1 & \text{in } \Omega_2, \end{cases}$$



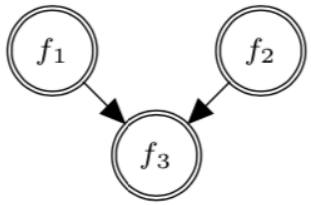
1. Three models: two finite element models with different mesh resolution, and reduced order model.
2. Constructed using RBniCS [Hesthaven, Rozza, Samm, 2015]
3. Predict temperature at (0.5, 0.8)

Thermal block results

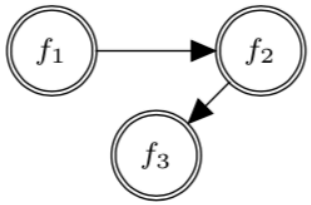
Non-recursive solution is best over 85% realizations of training data



(a) Full

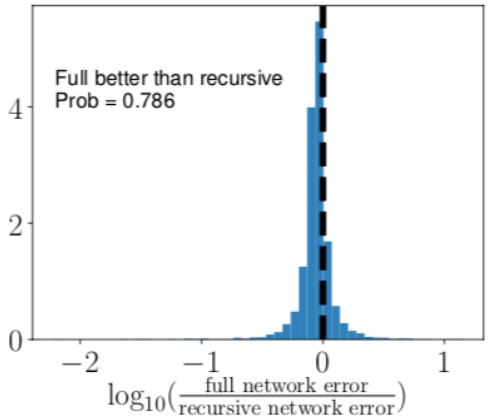


(b) Peer

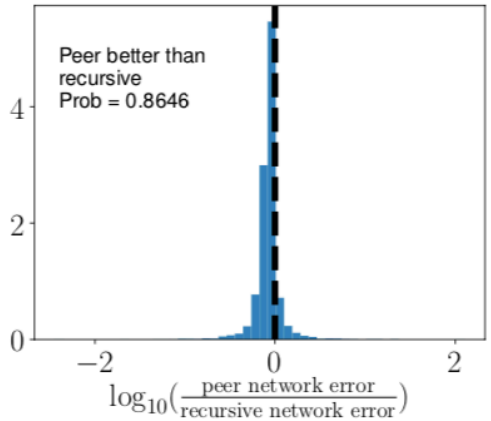


(c) Hierarchical

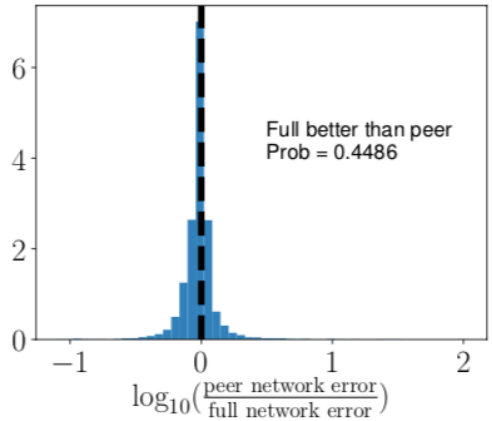
Figure 9: Thermal block model network structures.



(a) Full vs. hierarchical/recursive.



(b) Peer vs. hierarchical/recursive.



(c) Peer vs. full.

Probabilistic approaches

Questions:

Probabilistic approaches

Questions:

1. Can we extend our approaches to probabilistic learning?

Probabilistic approaches

Questions:

1. Can we extend our approaches to probabilistic learning?
2. Is the split into hyperparameters and parameters typically done to retain Gaussianity required?

Probabilistic approaches

Questions:

1. Can we extend our approaches to probabilistic learning?
2. Is the split into hyperparameters and parameters typically done to retain Gaussianity required?
3. Does this modeling framework pose difficulty for inference methodologies that are fully Bayesian?

Probabilistic approaches

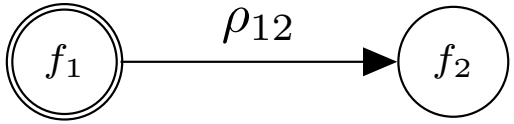
Questions:

1. Can we extend our approaches to probabilistic learning?
2. Is the split into hyperparameters and parameters typically done to retain Gaussianity required?
3. Does this modeling framework pose difficulty for inference methodologies that are fully Bayesian?

Next: A description and visualization of challenges

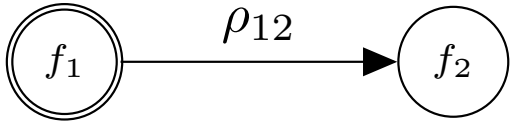
Non-identifiability in the posterior
Analytic expansion of two-model case

- Consider the two-function graph with linear parameterizations



Non-identifiability in the posterior
Analytic expansion of two-model case

- Consider the two-function graph with linear parameterizations

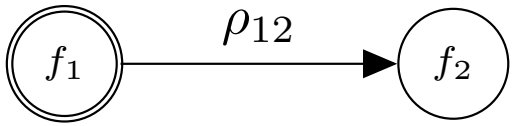


- The model for f_1 is

$$f_1(x) = V_1^T(x)\beta_1$$

Non-identifiability in the posterior
Analytic expansion of two-model case

- Consider the two-function graph with linear parameterizations



- The model for f_1 is

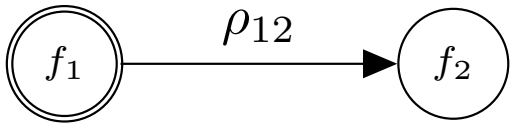
$$f_1(x) = V_1^T(x)\beta_1$$

- The model for f_2 is

$$\begin{aligned} f_2(x) &= \rho_{12}(x) f_1(x) + \delta_2(x) \\ &= \left(W_{12}^T(x)\alpha_{12}\right) V_1^T(x)\beta_1 + V_2^T(x)\beta_2 \\ &= \left(W_{12}^T \otimes V_1^T\right)(x)\left(\alpha_{12} \otimes \beta_1\right) + V_2^T(x)\beta_2 \end{aligned}$$

Non-identifiability in the posterior
Analytic expansion of two-model case

- Consider the two-function graph with linear parameterizations



- The model for f_1 is

$$f_1(x) = V_1^T(x)\beta_1$$

- The model for f_2 is

$$\begin{aligned} f_2(x) &= \rho_{12}(x) f_1(x) + \delta_2(x) \\ &= \left(W_{12}^T(x)\alpha_{12}\right) V_1^T(x)\beta_1 + V_2^T(x)\beta_2 \\ &= \left(W_{12}^T \otimes V_1^T\right)(x)\left(\alpha_{12} \otimes \beta_1\right) + V_2^T(x)\beta_2 \end{aligned}$$

- Explicit example with linear 1D models

$$\begin{aligned} f_2(x) &= \left(\alpha_{12,0} + \alpha_{12,1}x\right)\left(\beta_{1,0} + \beta_{1,1}x\right) + \beta_{2,0} + \beta_{2,1}x \\ &= \alpha_{12,1}\beta_{1,1}x^2 + \left(\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1}\right)x + \left(\beta_{2,0} + \alpha_{12,0}\beta_{1,0}\right) \end{aligned}$$

Non-identifiability in the posterior
Manifold of equal probabilities

$$f_1(x) = \beta_{1,0} + \beta_{1,1}x$$
$$f_2(x) = \alpha_{12,1}\beta_{1,1}x^2 + (\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1})\,x + (\beta_{2,0} + \alpha_{12,0}\beta_{1,0})$$

1. $\beta_{1,0}$ and $\beta_{1,1}$ can be identified ***from low fidelity data*** if it sufficient

Non-identifiability in the posterior
Manifold of equal probabilities

$$f_1(x) = \beta_{1,0} + \beta_{1,1}x$$
$$f_2(x) = \alpha_{12,1}\beta_{1,1}x^2 + (\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1})\,x + (\beta_{2,0} + \alpha_{12,0}\beta_{1,0})$$

1. $\beta_{1,0}$ and $\beta_{1,1}$ can be identified ***from low fidelity data*** if it sufficient
2. $\alpha_{12,1}$ can be uniquely identified because it is multiplied by known $\beta_{1,1}$

Non-identifiability in the posterior
Manifold of equal probabilities

$$f_1(x) = \beta_{1,0} + \beta_{1,1}x$$
$$f_2(x) = \alpha_{12,1}\beta_{1,1}x^2 + (\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1})\,x + (\beta_{2,0} + \alpha_{12,0}\beta_{1,0})$$

1. $\beta_{1,0}$ and $\beta_{1,1}$ can be identified ***from low fidelity data*** if it sufficient
2. $\alpha_{12,1}$ can be uniquely identified because it is multiplied by known $\beta_{1,1}$
3. Remaining parameters per term:
   1. $(\beta_{2,0}, \alpha_{12,0}) \rightarrow \beta_{2,0} + \alpha_{12,0}\beta_{1,0} = C$

Non-identifiability in the posterior
Manifold of equal probabilities

$$f_1(x) = \beta_{1,0} + \beta_{1,1}x$$
$$f_2(x) = \alpha_{12,1}\beta_{1,1}x^2 + (\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1})x + (\beta_{2,0} + \alpha_{12,0}\beta_{1,0})$$

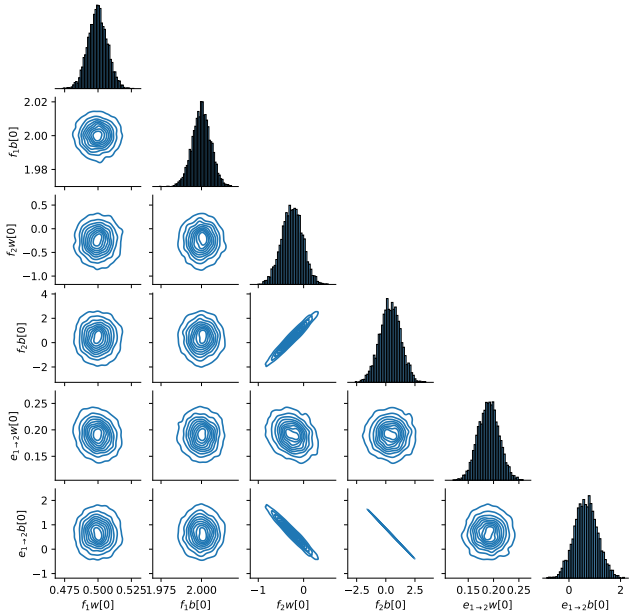
1. $\beta_{1,0}$ and $\beta_{1,1}$ can be identified ***from low fidelity data*** if it sufficient
2. $\alpha_{12,1}$ can be uniquely identified because it is multiplied by known $\beta_{1,1}$
3. Remaining parameters per term:
   1. $(\beta_{2,0}, \alpha_{12,0}) \rightarrow \beta_{2,0} + \alpha_{12,0}\beta_{1,0} = C$
   2. $(\alpha_{12,0}, \beta_{2,1}) \rightarrow \alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{1,0} + \beta_{2,1} = D$

Non-identifiability in the posterior
Manifold of equal probabilities

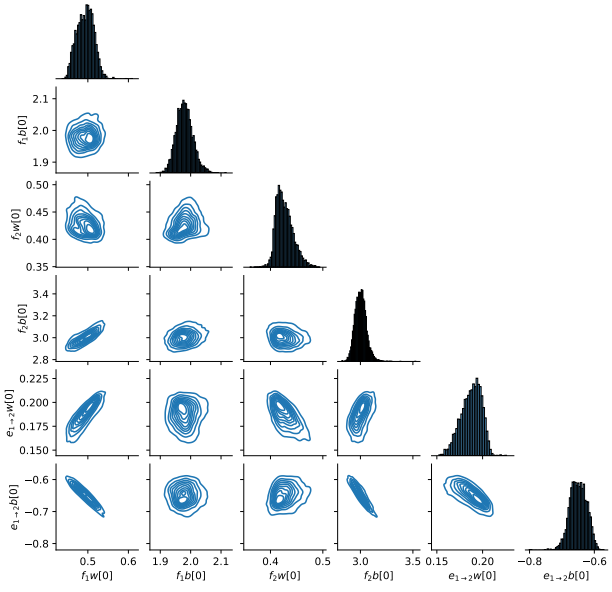
$$f_1(x) = \beta_{1,0} + \beta_{1,1}x$$
$$f_2(x) = \alpha_{12,1}\beta_{1,1}x^2 + (\alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{10} + \beta_{2,1})x + (\beta_{2,0} + \alpha_{12,0}\beta_{1,0})$$

1. $\beta_{1,0}$ and $\beta_{1,1}$ can be identified ***from low fidelity data*** if it sufficient
2. $\alpha_{12,1}$ can be uniquely identified because it is multiplied by known $\beta_{1,1}$
3. Remaining parameters per term:
   1. $(\beta_{2,0}, \alpha_{12,0}) \rightarrow \beta_{2,0} + \alpha_{12,0}\beta_{1,0} = C$
   2. $(\alpha_{12,0}, \beta_{2,1}) \rightarrow \alpha_{12,0}\beta_{1,1} + \alpha_{12,1}\beta_{1,0} + \beta_{2,1} = D$
4. 3 unknowns and 2 equations $\rightarrow$ 1 dimensional null space

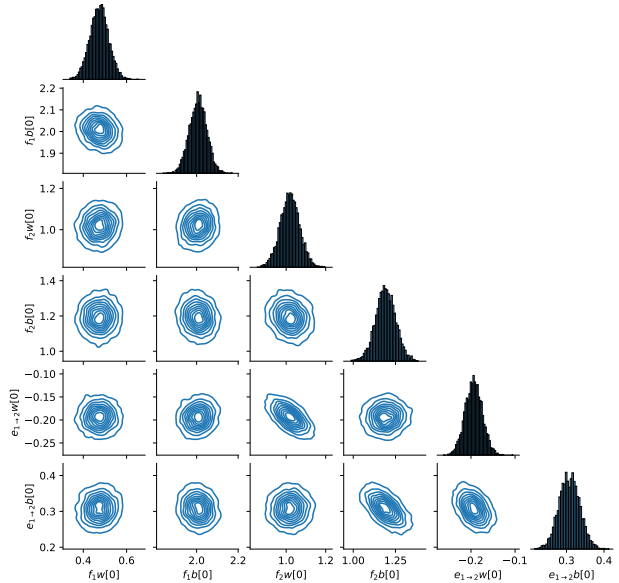
Non-identifiability in the posterior
HMC vs. Inverse Autoregressive Normalizing Flow vs. Normal



(a) HMC



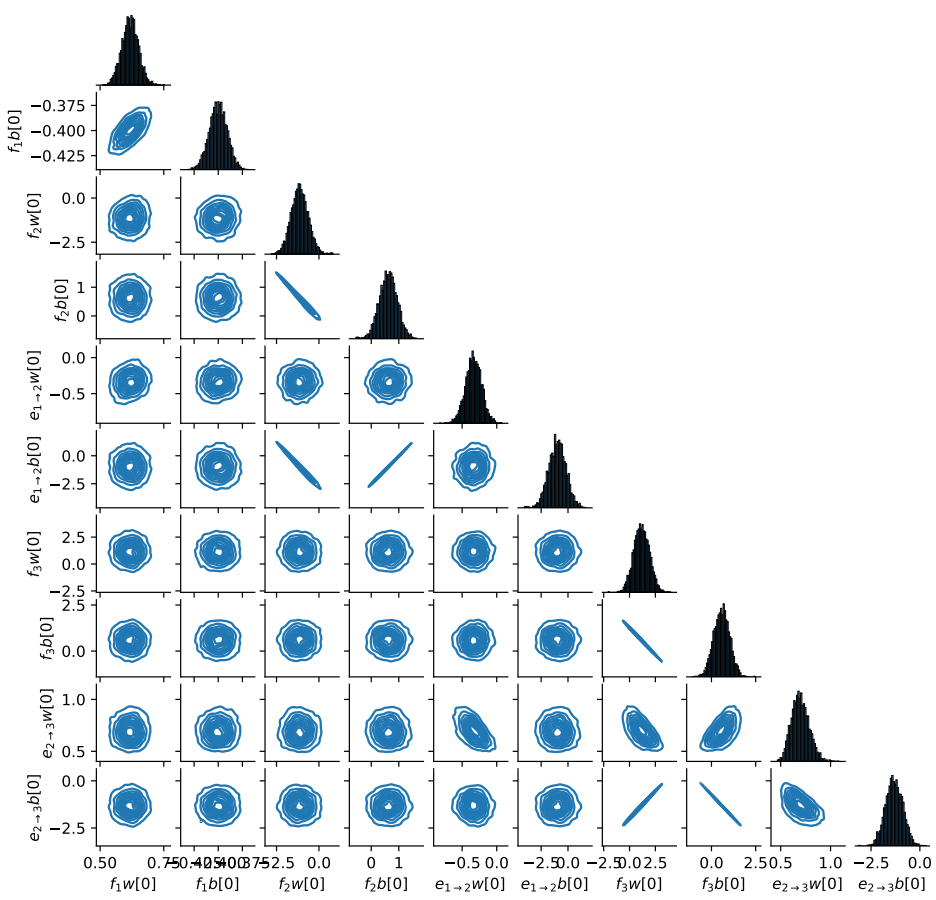
(b) IAF



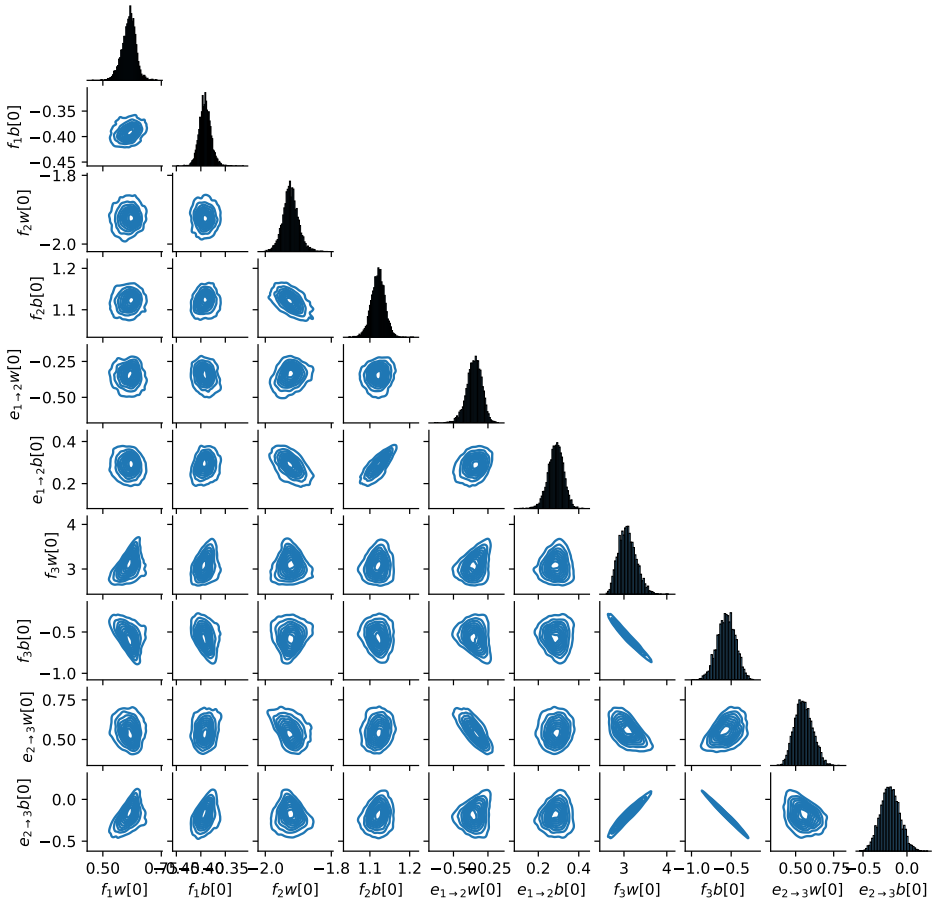
(c) Multivariate Normal

Non-identifiability in the posterior

3D case: variational inference misses manifold



(a) HMC



(b) IAF

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem
- Data can be used to update and inform the modeling structure only if this possibility is encoded.

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem
- Data can be used to update and inform the modeling structure only if this possibility is encoded.
- MFNets can increase accuracy compared to standard recursive approaches

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem
- Data can be used to update and inform the modeling structure only if this possibility is encoded.
- MFNets can increase accuracy compared to standard recursive approaches
- Challenges for Bayesian inference due to manifolds of equal probability
  - Investigating better priors to collapse the manifold
  - Investigating approximation quality of variational inference

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem
- Data can be used to update and inform the modeling structure only if this possibility is encoded.
- MFNets can increase accuracy compared to standard recursive approaches
- Challenges for Bayesian inference due to manifolds of equal probability
  - Investigating better priors to collapse the manifold
  - Investigating approximation quality of variational inference
- References

  - Gorodetsky, A. A., Jakeman, J. D., and Geraci, G. "MFNets: data efficient all-at-once learning of multifidelity surrogates as directed networks of information sources." *Computational Mechanics* 68:4 (2021): 741-758.
  - Gorodetsky, A. A., Jakeman, J.D., Geraci, G., Eldred, M.S., "MFNETS: multifidelity data-driven networks for Bayesian learning and prediction." *International Journal of Uncertainty Quantification* 10:6 (2020) : 595-622

Summary

- Determining how to efficiently fuse data from different simulations remains a *modeling* problem
- Data can be used to update and inform the modeling structure only if this possibility is encoded.
- MFNets can increase accuracy compared to standard recursive approaches
- Challenges for Bayesian inference due to manifolds of equal probability
  - Investigating better priors to collapse the manifold
  - Investigating approximation quality of variational inference
- References
  - Gorodetsky, A. A., Jakeman, J. D., and Geraci, G. "MFNets: data efficient all-at-once learning of multifidelity surrogates as directed networks of information sources." *Computational Mechanics* 68:4 (2021): 741-758.
  - Gorodetsky, A. A., Jakeman, J.D., Geraci, G., Eldred, M.S., "MFNETS: multifidelity data-driven networks for Bayesian learning and prediction." *International Journal of Uncertainty Quantification* 10:6 (2020) : 595-622
- Code: `http://github.com/goroda/MFNetsSurrogates`

Acknowledgements

This project was funded by the Laboratory Directed Research Development (LDRD) program at Sandia National Laboratories. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.