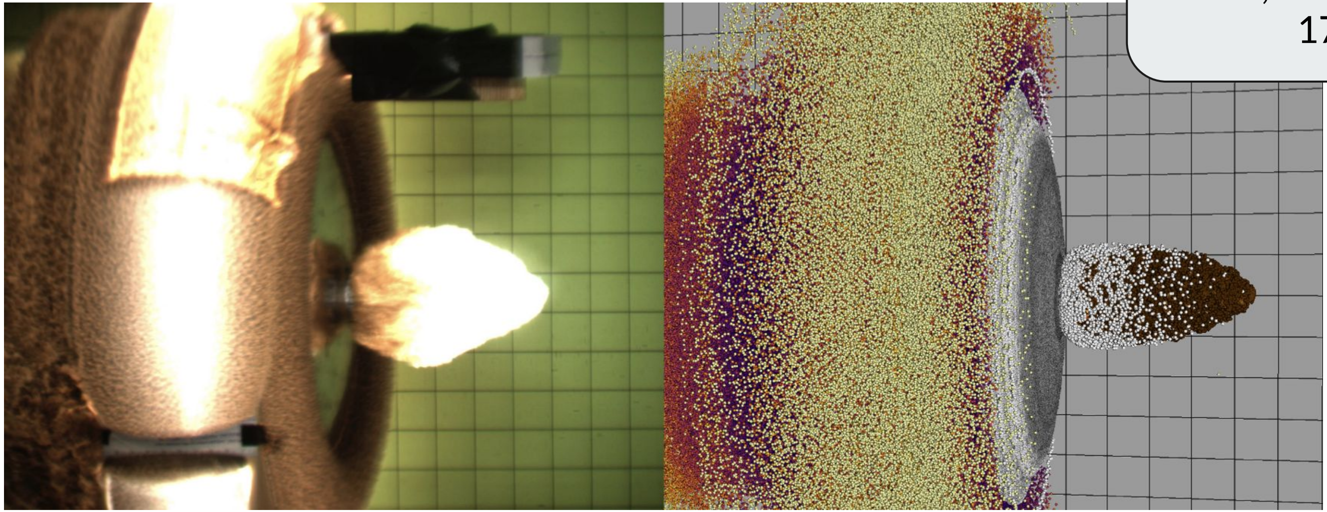
Multi-Fidelity and Reduced Order Modeling with Focus on Extremes and Naval Applications

2024. 04.26 Minji Kim

work with Prof. Vladas Pipiras

Exploring Statistical Approaches in Physics Simulation



One forward simulation:
10.1 million particles
1,440 processors
178 hours

Contents

Part1. Application Setting

- Where 'Multi-fidelity' model comes from
- Idea of Reduced Order Modeling
- Data-driven Physics Simulation

Part2. Statistical Methodologies

- Dealing with multi-fidelity data
- Direction 1:
  aided by specially sampled lo-fi data
- Direction 2:
  incorporate extra observation from lo-fi

Application Setting

Part 1: Data-Driven Physics Modeling

Motivation : Modeling Ship Motions/Loads

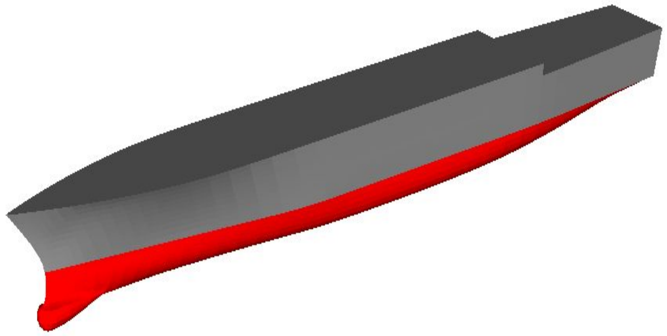


Fig: ONR topsides flared geometry.

- Head seas: ship pointing into the waves
- Modeling heave(ζ_g) and pitch (θ) motions.

The governing physics is the Newton's equation $F = ma$:

$$\begin{cases} m\ddot{\zeta}_g = F_{3,fkhs} + F_{3,hd} =: F_3, \\ I_Y\ddot{\theta} = F_{5,fkhs} + F_{5,hd} =: F_5, \end{cases}$$

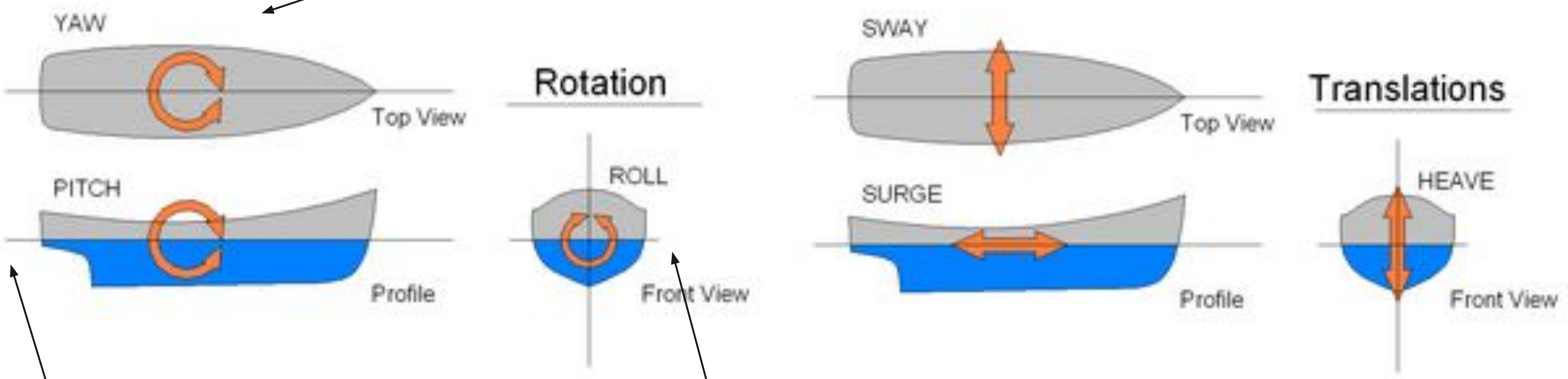
where m: ship mass, I_Y : pitch mass moment,

F_sth : respective forces..

Wikipedia “Ship motions”

Motions

Z-axis through the center of mass

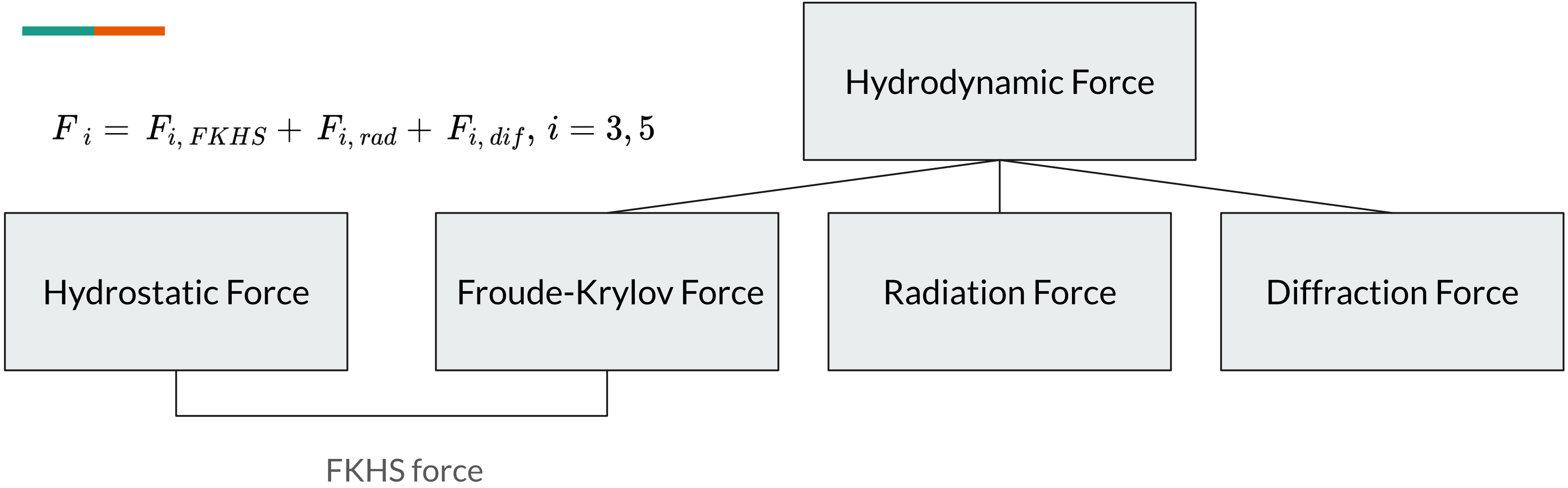


Y-axis

X-axis

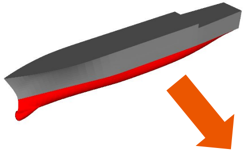
Forces

$$F_i = F_{i,\,FKHS} + F_{i,\,rad} + F_{i,\,dif},\ i = 3, 5$$



Forces

$$F_i = F_{i,\,FKHS} + F_{i,\,rad} + F_{i,\,dif},\; i = 3, 5$$



Radiation force

Hydrodynamic Force

Hydrostatic Force

Froude-Krylov Force

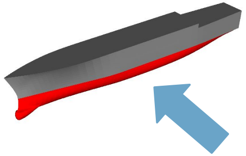
Radiation Force

Diffraction Force

FKHS force

Results from the oscillation of the ship. Causes it to act as a wave maker.

Results from the diffraction of the incident waves by the presence of the ship.



Incident wave

Wave exciting forces. Results from integrating the incident wave - pressure forces in the absence of the ship, over the surface of the immersed ship.

Reduced Order Modeling

$$F_i = F_{i,\,FKHS} + F_{i,\,rad} + F_{i,\,dif},\ i = 3, 5$$

Radiation Force

$$F_{i,rad} \sim -\left(A_{i3}\ddot{\zeta}_g + A_{i5}\ddot{\theta} + B_{i3}\dot{\zeta}_g + B_{i5}\dot{\theta}\right), \quad i = 3, 5.$$

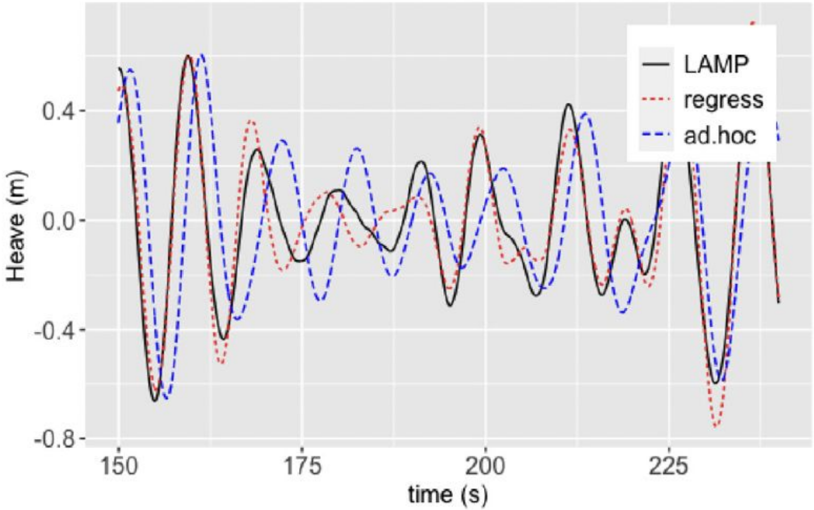
=> Simplifies the effect of radiation forces to a set of constant **added mass** (in phase with acc.) and **damping** (in phase with vel.) coefficients

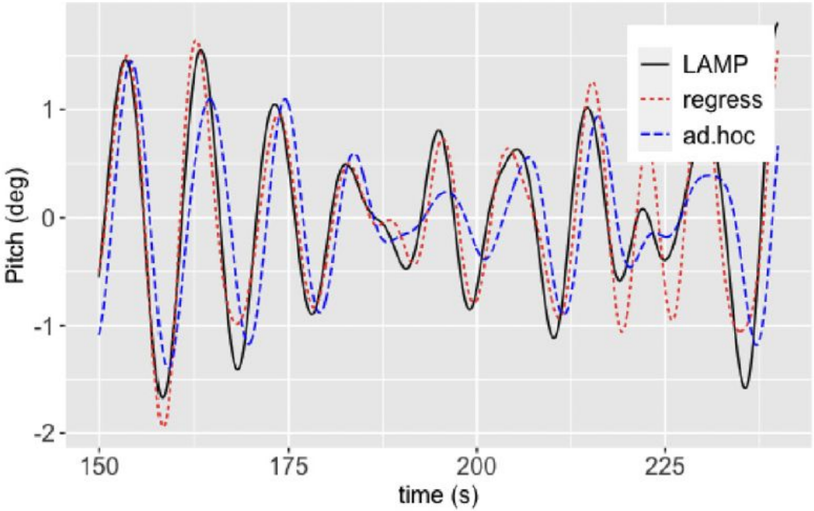
Diffraction Force

$$F_{i,dif} \sim \sum_{n=1}^{N_\omega} F_{c,i,n} \cos\left(\omega_n t + \phi_{0,n}\right) + F_{s,i,n} \sin\left(\omega_n t + \phi_{0,n}\right), \quad i = 3, 5$$

=> Penalized regression involving frequencies and phases of the underlying wave elevation models.

Reduced Order Modeling





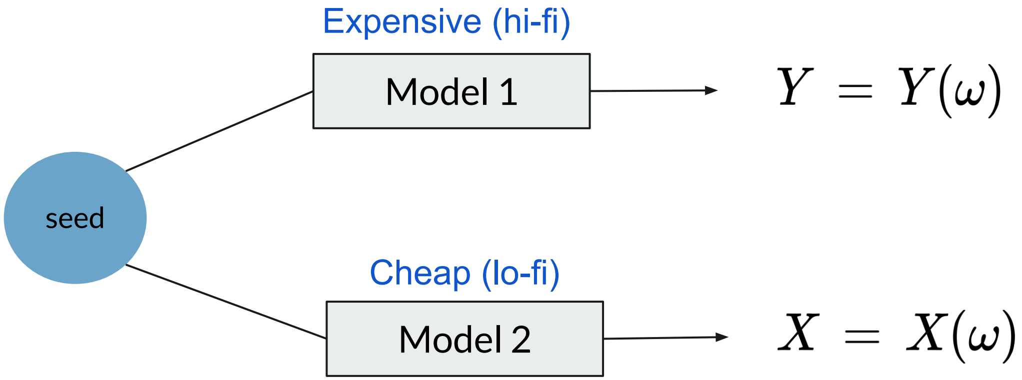
$$\begin{cases} (m + A_{33})\ddot{\zeta}_g + A_{35}\ddot{\theta} + B_{33}\dot{\zeta}_g + B_{35}\dot{\theta} + C_{33}\zeta_g + C_{35}\theta = F_{3k}(t), \\ A_{53}\ddot{\zeta}_g + (I_Y + A_{55})\ddot{\theta} + B_{53}\dot{\zeta}_g + B_{55}\dot{\theta} + C_{53}\zeta_g + C_{55}\theta = F_{5k}(t). \end{cases}$$

Before the regression approach, constants were set manually based on simulation experience.

Statistical Perspective

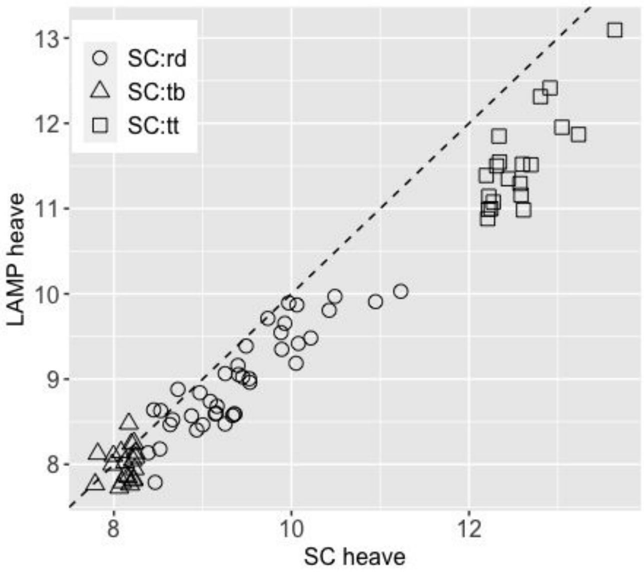
Part 2: Statistical Methodologies, No Physics

Motivation



Q: How can we leverage the low- fidelity outputs to enhance the estimation of high-fidelity quantities?

Problem 1



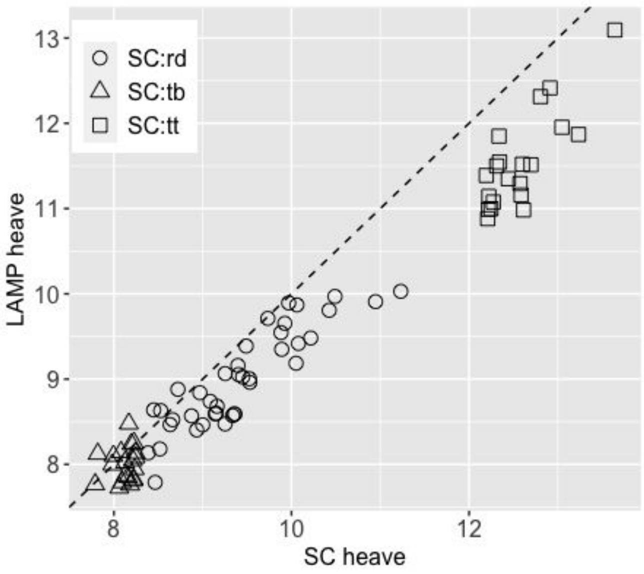
Objective

Let Y be the record maxima from 30 min run.

The distribution of Y, especially the behavior of its tails (for extremal values) is of interest.

Q: How can we leverage the low- fidelity outputs to enhance the estimation of high-fidelity quantities?

Problem 1



Let Y be the record maxima from 30 min run.

The distribution of Y, especially the behavior of its tails (for **extremal** values) is of interest.

One can obtain samples, but they are expensive.

Issue

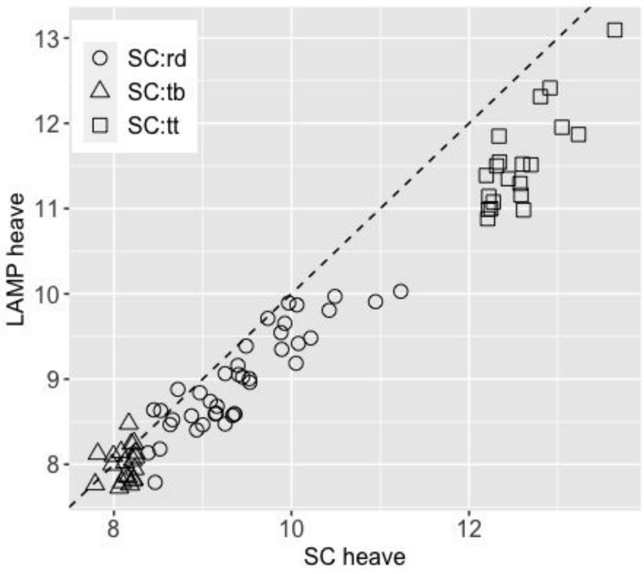
One can generate samples for X easily.

Once generate a lot of data from X, we can look at the large values of X and its seed.

Naive Idea

Q: How can we leverage the low- fidelity outputs to enhance the estimation of high-fidelity quantities?

Problem 1



Let Y be the record maxima from 30 min run.

The distribution of Y, especially the behavior of its tails (for **extremal** values) is of interest.

One can obtain samples, but they are expensive.

Issue

One can generate samples for X easily.

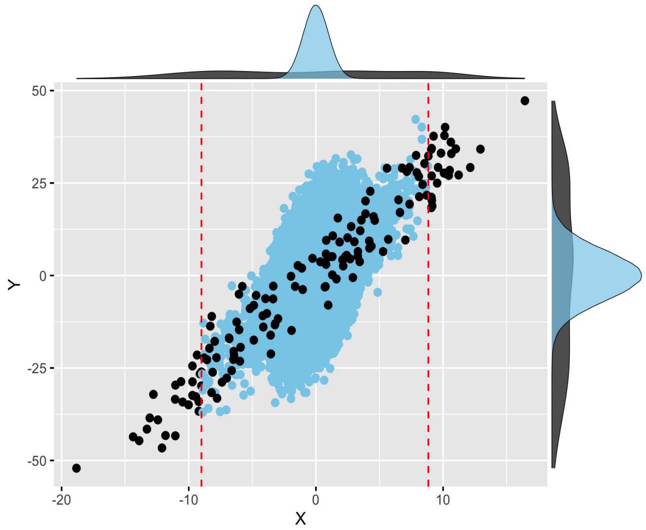
Once generate a lot of data from X, we can look at the large values of X and its seed.

Naive Idea

Q: How can we leverage the low- fidelity outputs to enhance the estimation of high-fidelity quantities?

Direction1: Specially sampled X can help!

Direction 1: How to sample lo-fi output



(m1) monotone

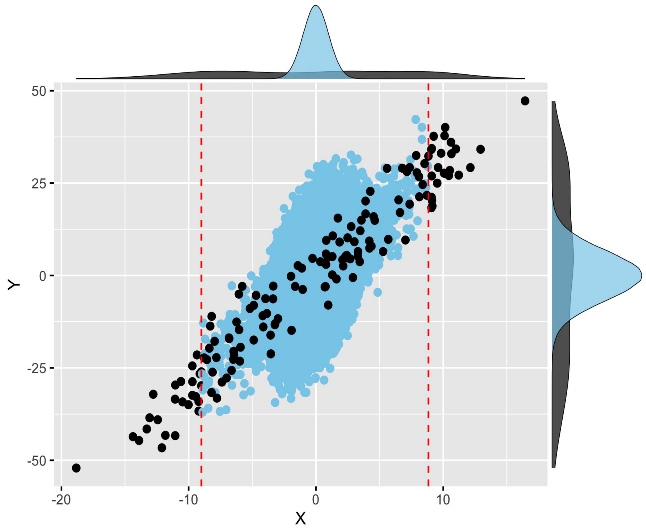
Setting: $Y = m(X) + \epsilon$

Once you have generated a lot of values from fX ...

Want to sample with putting more weights on extreme values of X

=> **Importance sampling** is employed.

Direction 1: How to sample lo-fi output



(m1) monotone

Setting: $Y = m(X) + \epsilon$

Once you have generated a lot of values from fX ...

Q1. What is the natural estimate for fY?

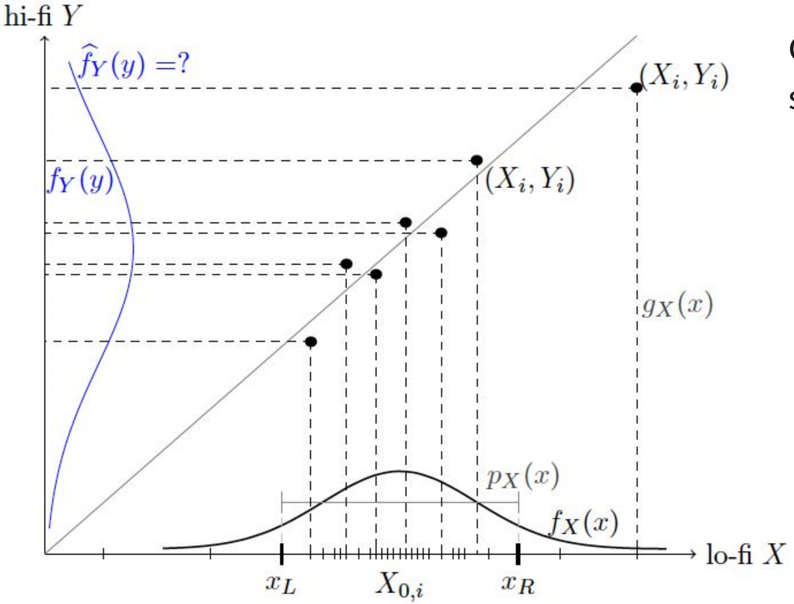
Kernel-based estimator of target PDF

$$\widehat{f}_Y(y) = \frac{1}{N}\sum_{i=1}^{N} K_h(y - Y_i)w(X_i), \quad w(X_i) = \frac{f_X(X_i)}{g_X(X_i)},$$

where $h > 0$ is a bandwidth, $K_h(u) = h^{-1}K(h^{-1}u)$ and K is a kernel function.

In order to calculate the appropriate weight, fX should be known

Direction 1: How to sample lo-fi output



Q1. What is the natural estimate?

Q2. What is the setting for sampling distribution, with special considerations for the distribution tails?

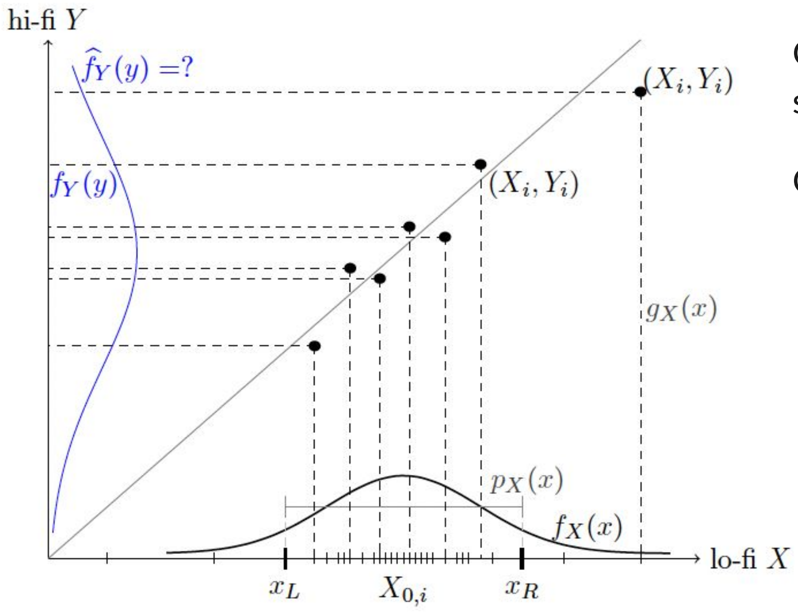
Proposal PDF

$$g_X(x) = \begin{cases} c_L\, f_X(x|X \le x_L), & \text{if } x \le x_L, \\ c_0\, p_X(x), & \text{o.w. }, \\ c_R\, f_X(x|X \ge x_R), & \text{if } x \ge x_R, \end{cases} \qquad w(x) = \begin{cases} \frac{1}{c_L}\mathbb{P}(X \le x_L), & \text{if } x \le x_L, \\ \frac{1}{c_0}\frac{f_X(x)}{p_X(x)}, & \text{o.w.}, \\ \frac{1}{c_R}\mathbb{P}(X \ge x_R), & \text{if } x \ge x_R. \end{cases}$$

E.g. for $x \ge x_R$, this ensures that all $X_{0,i} \ge x_R$ can be selected in the sampled $X_1, \ldots, X_N$.

pX should be determined

Direction 1: How to sample lo-fi output



Q1. What is the natural estimate?

Q2. What is the setting for sampling distribution, with special considerations for the distribution tails?

Q3. What is the optimal sampling distribution?

Optimality criteria for the proposal PDF

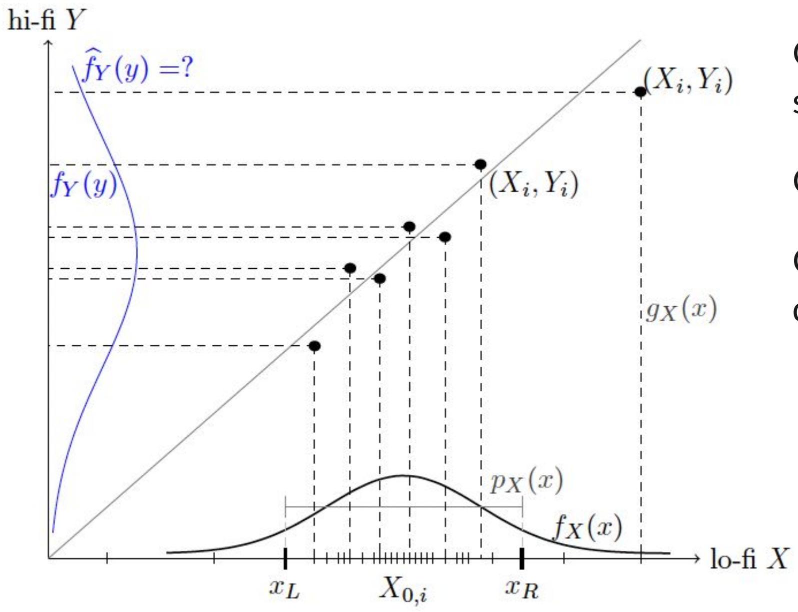
$$\text{Optimality}: \frac{N\text{Var}(\widehat{f}_Y(y))}{f_Y(y)^2} \simeq \text{const.} \qquad (\star)$$

For monotone m, the optimality $(\star)$ translates into

$$p_X(x) \propto m'(x), \quad x_L < x < x_R.$$

[with O’Connor, Pipiras, Sapsis]

Direction 1: How to sample lo-fi output



Q1. What is the natural estimate?

Q2. What is the setting for sampling distribution, with special considerations for the distribution tails?

Q3. What is the optimal sampling distribution?

Q4. Separate treatment in the tails where less or no data are available, based on extreme value theory.

Modification in the tails

$$\widehat{f}_Y^{(m)}(y) = \begin{cases} g_{\widehat{\xi}_R, \widehat{\delta}_R}(y - y_R), & \text{if } y \geq y_R, \\ \widehat{f}_Y(y), & \text{if } y_L < y < y_R, \\ g_{\widehat{\xi}_L, \widehat{\delta}_L}(-(y - y_L)), & \text{if } y \leq y_L, \end{cases}$$

where $g_{\xi,\delta}(u)$ is the PDF of the generalized Pareto distribution (GPD).

[with O'Connor, Pipiras, Sapsis]

Direction 1: How to sample lo-fi output

Case study $Y = m(X) + \epsilon$

Extended to

- piecewise-monotone m
- Homoscedastic noise
- Heteroscedastic noise

In practice, we do not know the relationship Y=m(X)+e. We apply piecewise linear regression to estimate m.

Algorithm 2 Adaptive Sampling Incorprating m Estimation

Input: PDF f_X, thresholds x_L and x_R

sample (X_i, Y_i) where $X_i \sim \mathrm{Unif}(x_L, x_R)$ for $i = -1, \dots, -n_0$

construct $D^{(0)} = \{(X_i, Y_i),\ i = -1, \dots, -n_0\}$

fit piecewise linear regression (PLR) to $D^{(0)}$ to obtain the initial estimate $\widehat{m}^{(0)}$ and its monotone components $\{\widehat{m}_{j,0}, j \in \mathcal{J}^{(0)}\}$

for $t = 1, \dots, \tilde{N}$ **do**

  $\widehat{f}_Y^{(t)}(y) \leftarrow \sum_{j \in \mathcal{J}^{(t-1)}} \frac{f_X(\widehat{m}_{j,t-1}^{-1}(y))}{|\widehat{m}'_{j,t-1}(\widehat{m}_{j,t-1}^{-1}(y))|} \mathbb{1}(y \in \widehat{m}^{(t)}(A_j))$

  $\widehat{p}_X^{(t)}(x) \leftarrow \frac{f_X(x)}{\widehat{f}_Y^{(t)}(\widehat{m}^{(t-1)}(x))}$ ▷ construct $\widehat{p}_X$

  normalize $\widehat{p}_X^{(t)}$ on $x_L < x < x_R$

  sample (X_t, Y_t) where $X_t \sim \widehat{p}_X^{(t)}$ ▷ sample new point

  $w(X_t) \leftarrow \frac{f_X(X_t)}{\widehat{p}_X^{(t)}(X_t)}$ ▷ update weights

  update $D^{(t)} = \{(X_{-n_0}, Y_{-n_0}), \dots, (X_{-1}, Y_{-1}), (X_1, Y_1), \dots, (X_t, Y_t)\}$

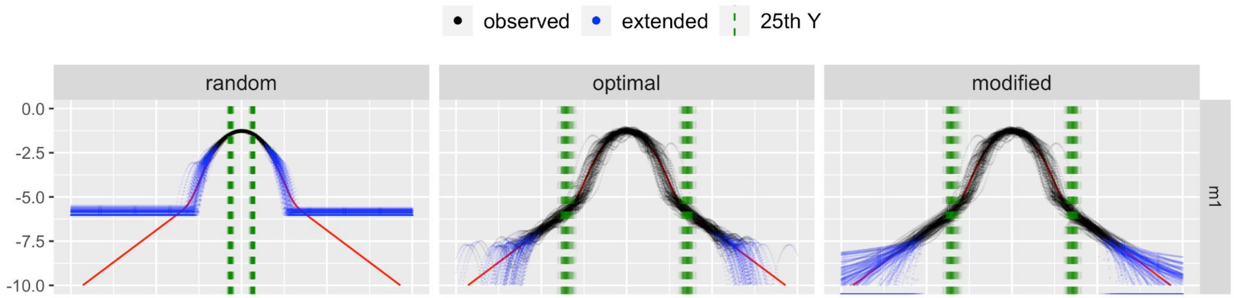
  fit PLR to $D^{(t)}$ to obtain $\widehat{m}^{(t)}$ and its monotone components $\{\widehat{m}_j^{(t)}, j \in \mathcal{J}^{(t)}\}$ ▷ update $\widehat{m}$

end for

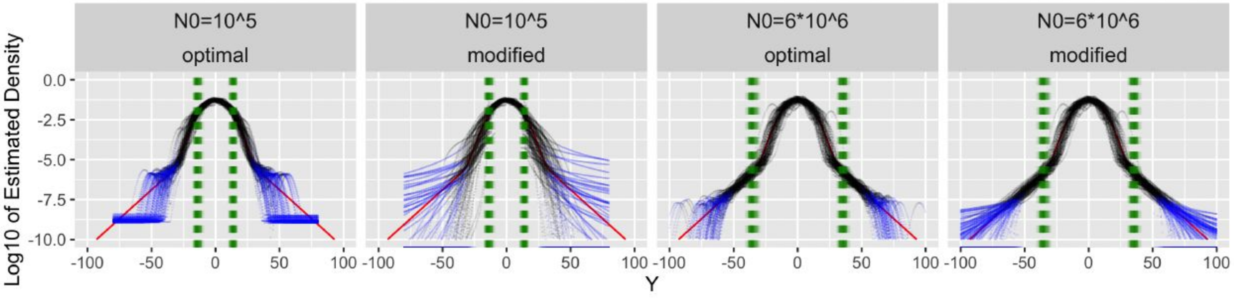
Output: Sample $(X_1, Y_1), \dots, (X_{\tilde{N}}, Y_{\tilde{N}})$.

Direction 1: How to sample lo-fi output

Case study



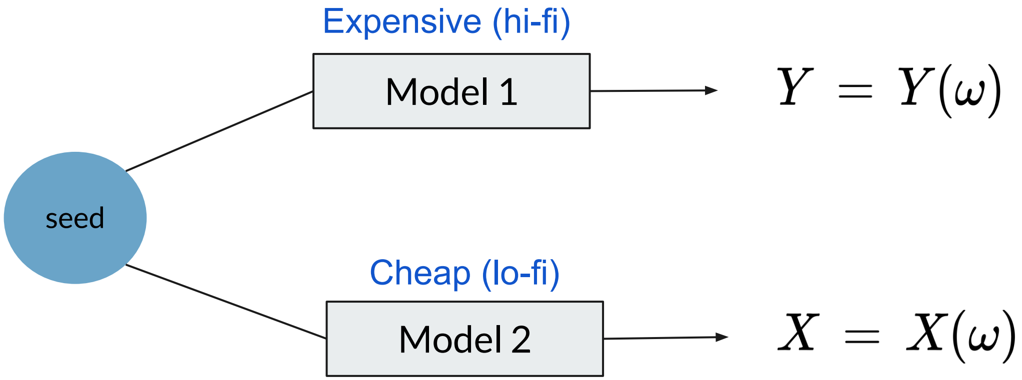
Role of threshold for GPD



Statistical Perspective

Part 2-2: Direction 2

Problem 2



Q: How can we leverage the low- fidelity outputs to enhance the estimation of high-fidelity quantities?

Direction 2. To get a better hi-fi estimate, devise an estimator with low uncertainty by incorporating (a larger amount of) lo-fi observation.

Direction 2: Incorporate larger amount of lo-fi data

Literature

Let the Quantity of Interest (QoI) be the unknown mean $\mathbb{E}Y$,

Assume we have observed $(X_1, Y_1), \; \dots (X_n, Y_n), \; X_{n+1}, \; \dots X_{n+m}.$

Given additional data for low-fidelity outputs, Multifidelity Monte Carlo (MFMC) estimator is defined as:

$$\hat{\mu}_{mf} = \bar{Y}_n + \alpha\big(\bar{X}_{n+m} - \bar{X}_n\big)$$

Direction 2: Incorporate larger amount of lo-fi data

Let the Quantity of Interest (QoI) be the unknown mean $\mathbb{E}Y$,

Assume we have observed $(X_1, Y_1),\ \ldots (X_n, Y_n),\ X_{n+1},\ \ldots X_{n+m}.$

Given additional data for low-fidelity outputs, Multifidelity Monte Carlo (MFMC) estimator is defined as:

$$\hat{\mu}_{mf} = \bar{Y}_n + \alpha\left(\bar{X}_{n+m} - \bar{X}_n\right)$$

Rationale

When X and Y are correlated, then the uncertainty of $\bar{Y}_n - \alpha\bar{X}_n$ is expected to be smaller than that of $\bar{Y}_n$

Direction 2: Incorporate larger amount of lo-fi data

Let the Quantity of Interest (QoI) be the unknown mean $\mathbb{E}Y$,

Assume we have observed $(X_1, Y_1),\ \ldots (X_n, Y_n),\ X_{n+1},\ \ldots X_{n+m}.$

Note that this is the multi-fidelity data of interest!

Given additional data for low-fidelity outputs, Multifidelity Monte Carlo (MFMC) estimator is defined as:

$$\hat{\mu}_{mf} = \bar{Y}_n + \alpha\left(\bar{X}_{n+m} - \bar{X}_n\right)$$

When X and Y are correlated, then the uncertainty of $\bar{Y}_n - \alpha \bar{X}_n$ is expected to be smaller than that of $\bar{Y}_n$

Q. Optimal choice of alpha and the optimal variance?

With $\alpha^* = \arg\min_\alpha Var\left(\bar{Y}_n - \alpha\bar{X}_n\right) = \frac{Cov(X,Y)}{Var(X)},\quad Var\left(\hat{\mu}_{mf}\,(\alpha^*)\right) < Var\left(\bar{Y}_n\right)$ holds

Direction 2: Incorporate larger amount of lo-fi data

Motivation: When QoI is about the extremal quantity, e.g. $P(Y > a)$, it would be better to fit parametric distribution.

Approach: Assume parametric model for X and Y.

Devise a MFMC approach for estimation of the hi-fi **parameter**.

Direction 2: Incorporate larger amount of lo-fi data

Motivation: When QoI is about the extremal quantity, e.g. $P(Y > a)$, it would be better to fit parametric distribution.

Approach: Assume parametric model for X and Y.

Devise a MFMC approach for estimation of the hi-fi **parameter**.

Joint ML estimator

$$(\hat{\theta}_{1,ml}, \hat{\theta}_{2,ml}, \hat{\theta}_{1,2,ml}) = \underset{(\theta_1,\theta_2,\theta_{1,2})}{\arg\max} \prod_{i=1}^{n} f_{(\theta_1,\theta_2,\theta_{1,2})}(X_i, Y_i) \prod_{i=n+1}^{n+m} f_{\theta_2}(X_i)$$

marginal ML MF estimator

$$\hat{\theta}_{1,mml} = (\hat{\theta}_{1,ml})_n + \beta \odot \left((\hat{\theta}_{2,ml})_{n+m} - (\hat{\theta}_{2,ml})_n \right)$$

moment MF estimator

$$\hat{\theta}_{1,mom} = g\left(\overline{h(Y)}_n + \alpha \odot \left(\overline{h(X)}_{n+m} - \overline{h(X)}_n \right) \right), \text{ where } \theta_1 = g(\mathbb{E}h(Y))$$

We compare these MF estimators to baseline estimators that use only high-fidelity data.

Direction 2: Incorporate larger amount of lo-fi data

Joint ML estimator

$$(\hat{\theta}_{1,ml}, \hat{\theta}_{2,ml}, \hat{\theta}_{1,2,ml}) = \underset{(\theta_1,\theta_2,\theta_{1,2})}{\arg\max} \prod_{i=1}^{n} f_{(\theta_1,\theta_2,\theta_{1,2})}(X_i, Y_i) \prod_{i=n+1}^{n+m} f_{\theta_2}(X_i)$$

marginal ML MF estimator

$$\hat{\theta}_{1,mml} = (\hat{\theta}_{1,ml})_n + \beta \odot \left((\hat{\theta}_{2,ml})_{n+m} - (\hat{\theta}_{2,ml})_n \right)$$

moment MF estimator

$$\hat{\theta}_{1,mom} = g\left(\overline{h(Y)}_n + \alpha \odot \left(\overline{h(X)}_{n+m} - \overline{h(X)}_n \right) \right), \text{ where } \theta_1 = g(\mathbb{E}h(Y))$$

These (multi-fidelity) estimators are devised to work with partially labeled datasets.

Question: What are these estimators for various parametric distributions?

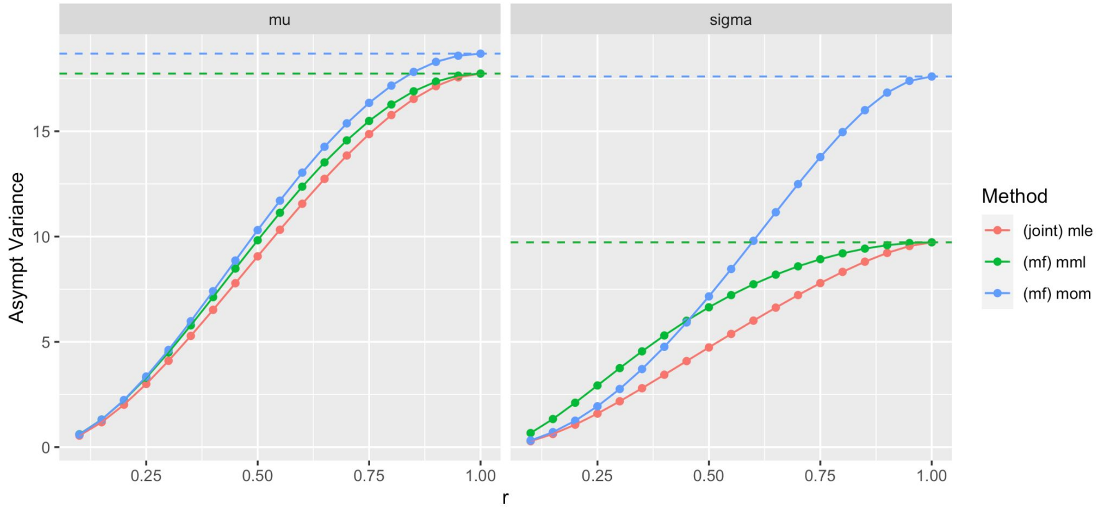
Question: What are the theoretical efficiencies of these estimators?

Question: How can we compare the efficiencies?

Direction 2: Incorporate larger amount of lo-fi data

Result1 : Marginal ML MF estimator is *optimal* for bivariate Gaussian distribution!

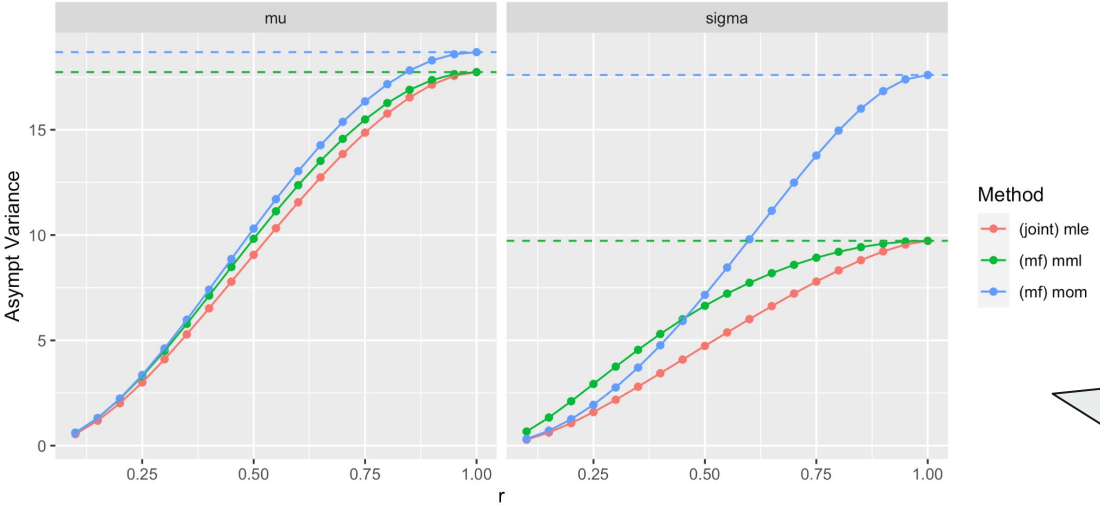
Result2 : Compare the 'uncertainties' across different approaches, e.g., for Gumbel distribution



Direction 2: Incorporate larger amount of lo-fi data

Result1 : Marginal ML MF estimator is *optimal* for bivariate Gaussian distribution!

Result2 : Compare the 'uncertainties' across different approaches, e.g., for Gumbel distribution



Dashed lines denote baseline performance. The MF estimators show better performance as the dependence increases.

References

- K. Griffin, K. Wang, E. Brugger, C. Harrison, T. Rehagen, M. Larson, (2020), Smoothed Particle Hydrodynamics modeling of a jet penetration experiment, *The International Conference for High Performance Computing, Networking, Storage, and Analysis*.
- M. Kim, V. Pipiras, A. Reed, K. Weems, (2023), Calibration of low-fidelity ship motion programs through regressions of high-fidelity forces, *Ocean Engineering*, 290, 116321
- M. Kim, K. O'Connor, V. Pipiras, T. Sapsis, (2024a), Sampling low-fidelity outputs for estimation of high-fidelity density and its tails, *SIAM/ASA Journal on Uncertainty Quantification*, to appear
- M. Kim, B. Brown, V. Pipiras, (2024b), Parametric multi-fidelity Monte Carlo estimation with applications to extremes, Preprint.

Thank you!