Fast ab initio uncertainty quantification

Mengyang Gu

Department of Statistics and Applied Probability
University of California, Santa Barbara

Fast ab initio uncertainty quantification

Mengyang Gu

Department of Statistics and Applied Probability
University of California, Santa Barbara

"The Department of Statistics helps students acquire the **conceptual**, **computational**, and **mathematical tools** for **quantifying uncertainty** and **making sense of complex data** arising from many applications."

Fast ab initio uncertainty quantification

Mengyang Gu

Department of Statistics and Applied Probability
University of California, Santa Barbara

"The Department of Statistics helps students acquire the **conceptual**, **computational**, and **mathematical tools** for **quantifying uncertainty** and **making sense of complex data** arising from many applications."
–From the front page of the Department of Statistics at Harvard:
https://statistics.fas.harvard.edu/

From loss minimization to generative models

Parameter estimation often involves minimizing a selected loss and statistics

$$\boldsymbol{\theta}_{est} = \text{argmin}_{\boldsymbol{\theta}} \text{Loss}(\mathbf{D}, \mathbf{D}_m(\boldsymbol{\theta})) \tag{1}$$

where $\mathbf{D}$ is the observed statistics and $\mathbf{D}_m(\boldsymbol{\theta})$ is the output from a model.

Limitation: (1) Choice of loss and statistics may not be optimal; (2) Lack of uncertainty quantification; (3) Scalability and stability of optimization.

From loss minimization to generative models

Parameter estimation often involves minimizing a selected loss and statistics

$$\boldsymbol{\theta}_{est} = \text{argmin}_{\boldsymbol{\theta}} \text{Loss}(\mathbf{D}, \mathbf{D}_m(\boldsymbol{\theta})) \tag{1}$$

where $\mathbf{D}$ is the observed statistics and $\mathbf{D}_m(\boldsymbol{\theta})$ is the output from a model.

Limitation: (1) Choice of loss and statistics may not be optimal; (2) Lack of uncertainty quantification; (3) Scalability and stability of optimization.

We propose to improve the loss minimization in following a two-step principle.

- Build a (large) probabilistic model of **untransformed data** with parameters that can be identified from data. Show **equivalence** between estimation in loss minimization and a statistical estimator of the probabilistic model.
- Derive a more efficient estimator by marginalizing out random components.

Fast algorithms can be developed to accelerate the computation. An approach that defines generative models and propagates the uncertainty from the beginning of data processing, an *ab initio* uncertainty quantification (AIUQ) approach [1].

[1] Gu, M., He, Y., Liu, X., & Luo, Y. (2023). Ab initio uncertainty quantification in scattering analysis of microscopy. *Under review, Physical Review X*, arXiv preprint arXiv:2309.02468.

Outline

1. Example: AIUQ in scattering analysis of dynamics
   - Scattering analysis and differential dynamic microscopy
   - A latent factor model as the data generative model
   - Acceleration by the generalized Schur method for Toeplitz covariances
   - Three sets of simulated studies
   - Three sets of real experiments
2. Marginalization and generative models of a few other examples
3. Calibration of imperfect geophysical models by satellite interferograms with measurement bias
   - Calibration of imperfect models
   - Imperfect model calibration by imperfect data

Outline

From time-lapse videos to dynamical information

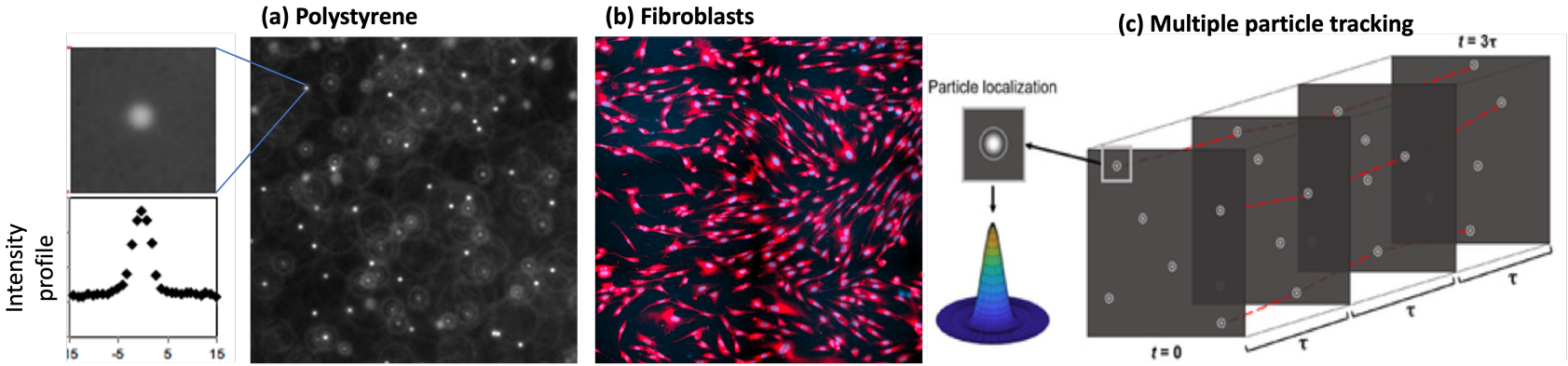


Figure 1: Multiple particle tracking (MPT) for (a) Polystyrene and (b) guiding fibroblast toward alignment [Gu et al., 2023b].

Two different approaches for analyzing time-lapse microscopy videos:

- Localizing particle intensity profile and linking trajectory by multiple particle tracking (MPT) [Crocker and Grier, 1996] shown in Fig 1. Parameters needs to be tuned in MPT case-by-case. Poor quality in e.g. (1) optically dense systems, (2) systems with non-elliptical objects, (3) not eligible for placing florescent probes.
- Basis decomposition method without tracking individual particle (to be discussed).

Outline

Scattering analysis of dynamics

Denote the normalized summation of 2D positions of M particles in the reciprocal space $\psi(\mathbf{q},t)=\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\exp(-i\mathbf{q}\cdot\mathbf{x}_m(t))$ where $\mathbf{x}_m(t)$ is the position of the mth particle. Particles are often assumed to have no or weak interaction and the intermediate scattering function (ISF) follows:

$$f_{\boldsymbol{\theta}}(\mathbf{q},\Delta t)=\mathrm{Cov}(\psi(\mathbf{q},t),\psi^*(\mathbf{q},t+\Delta t))=\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}\exp\left(i\mathbf{q}\cdot\Delta\mathbf{x}_m(t,\Delta t)\right)\right],$$

Scattering analysis of dynamics

Denote the normalized summation of 2D positions of M particles in the reciprocal space $\psi(\mathbf{q},t)=\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\exp(-i\mathbf{q}\cdot\mathbf{x}_m(t))$ where $\mathbf{x}_m(t)$ is the position of the mth particle. Particles are often assumed to have no or weak interaction and the intermediate scattering function (ISF) follows:

$$f_{\boldsymbol{\theta}}(\mathbf{q},\Delta t)=\mathrm{Cov}(\psi(\mathbf{q},t),\psi^*(\mathbf{q},t+\Delta t))=\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}\exp\left(i\mathbf{q}\cdot\Delta\mathbf{x}_m(t,\Delta t)\right)\right],$$

- (**Parametric models**). E.g., for 2D diffusive processes, the ISF follows

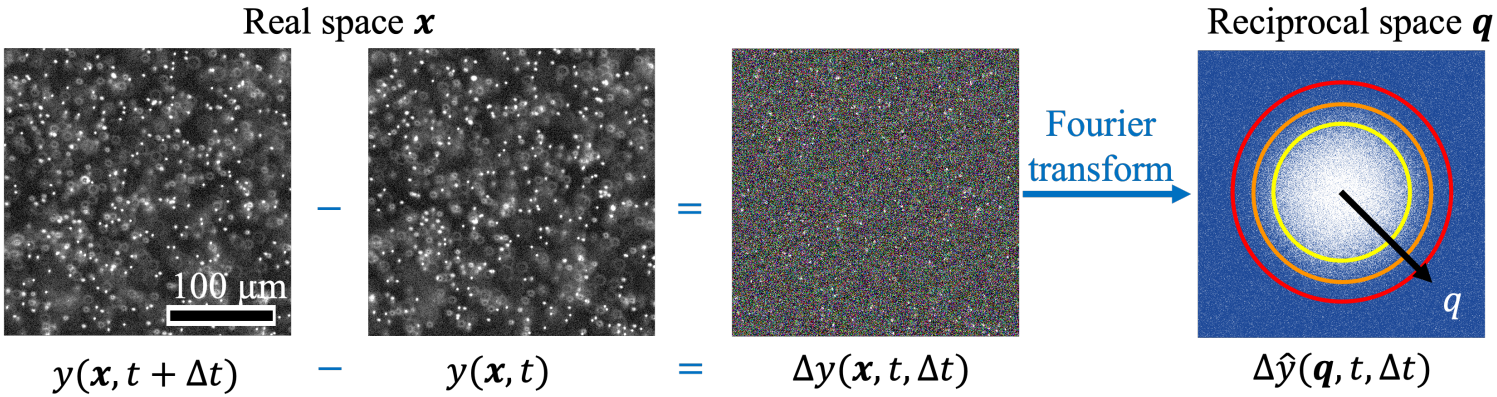
$$f_{\boldsymbol{\theta}}(\mathbf{q},\Delta t)=\exp\left(-q^2\theta\Delta t\right),$$

  where $q=\sqrt{q_1^2+q_2^2}$. Here θ is the diffusion coefficient, from which one can estimate viscosity or particle sizes from Stokes-Einstein relation: $\theta=\frac{k_BT}{6\pi\eta r}$.
- (**Nonparametric models**). Approximate ISF by mean squared displacement

$$f_{\boldsymbol{\theta}}(\mathbf{q},\Delta t)\approx\exp\left(-\frac{q^2\langle\Delta x^2(\Delta t)\rangle)}{4}\right),\qquad\text{(by cumulant theorem)}$$

  where $\theta(\Delta t)=\langle\Delta x^2(\Delta t)\rangle$. The storage and loss modulus can be obtained by the generalized Stokes-Einstein relation [Mason, 2000].

Differential dynamic microscopy



In *differential dynamic microscopy* (DDM) [Cerbino and Trappe, 2008], one compute the image structure function $D(\mathbf{q}, \Delta t) = \sum_{i=1}^{n} \Delta\hat{y}(\mathbf{q}, t_i, \Delta t)^2/n$

$$D(\mathbf{q}, \Delta t) = A(\mathbf{q})(1 - f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t)) + B, \tag{2}$$

where $A(\mathbf{q})$ is the real-valued scalar of amplitude parameter for wavevector $\mathbf{q}$, $f_{\boldsymbol{\theta}}$ is the ISF and B denotes a random noise with mean parameter $\bar{B}$. Suppose we have J distinct ISFs for each Δt, and n time frames. One fits the image structure function with respect to a loss function (such as the L_2 loss):

$$(\boldsymbol{\theta}_{est}, \mathbf{A}_{est,1:J}, \bar{B}_{est}) = \text{argmin}_{\boldsymbol{\theta}, \mathbf{A}_{1:J}, \bar{B}} \text{Loss}(\mathbf{D}, \mathbf{D}_m), \tag{3}$$

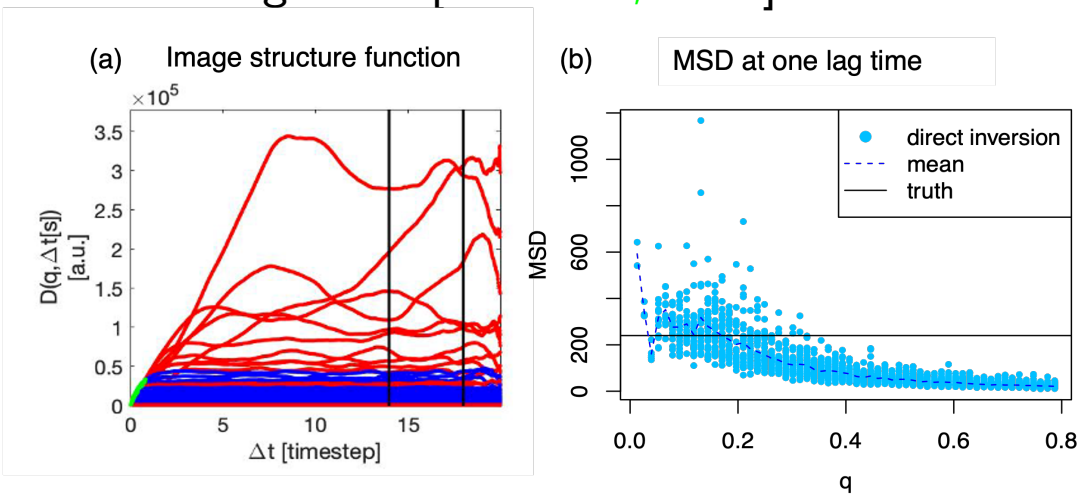
where the entry of $\mathbf{D}_m$ and $\mathbf{D}$ is $D_m(q_j, \Delta t_k) = A(q_j)(1 - f_{\boldsymbol{\theta}}(q_j, \Delta t_k)) + \bar{B}$ and $D(q_j, \Delta t_k)$, respectively, for $j = 1, ..., J$ and $k = 1, ..., n$.

Wide applications and challenges in automated analysis

- In DDM, **each pixel** in microscopic images is analogous to **a point detector in dynamic light scattering** [Berne and Pecora, 2000].
- Complement to particle tracking: applicable to nonspherical particles, optically dense, and fast dynamics, etc, **without tracking individual particles**.
- DDM has wide applications: e.g. bacteria [Wilson et al., 2011], concentrated particle suspensions [Lu et al., 2012], phase separating colloidal gels [Gao et al., 2015], and actin-microtubule network [Lee et al., 2021].

Wide applications and challenges in automated analysis

- In DDM, **each pixel** in microscopic images is analogous to **a point detector in dynamic light scattering** [Berne and Pecora, 2000].
- Complement to particle tracking: applicable to nonspherical particles, optically dense, and fast dynamics, etc, **without tracking individual particles**.
- DDM has wide applications: e.g. bacteria [Wilson et al., 2011], concentrated particle suspensions [Lu et al., 2012], phase separating colloidal gels [Gao et al., 2015], and actin-microtubule network [Lee et al., 2021].
- **Three challenges**: (1) A range of wavevectors needs to be pre-selected for aggregating the information; (2) Lack of uncertainty quantification; (3) computational and storage cost [Gu et al., 2021].



A question to be answered: What is the probabilistic model implicitly assumed for the real-space image intensity in DDM?

Outline

A latent factor model as the generative model

Consider a latent factor model of a square view of image intensities at N pixels $\mathbf{y}(t) = [y(\mathbf{x}_1, t), ..., y(\mathbf{x}_N, t)]^T$ [1]:

$$\mathbf{y}(t) = \frac{1}{\sqrt{N}} \mathbf{W}^* \mathbf{z}(t) + \boldsymbol{\epsilon}(t); \tag{4}$$

- $\mathbf{W}^*$ is an $N \times N$ complex conjugate of the 2D discrete Fourier basis.
- Noises follow $\boldsymbol{\epsilon}(t) \sim \text{MN}(\mathbf{0}, \sigma_0^2 \mathbf{I}_N)$ with variance $\sigma_0^2 = \frac{\bar{B}}{2}$.
- The complex latent factor follows $\mathbf{z}(t)$: $\mathbf{z}(t) = \mathbf{z}_{re}(t) + i\mathbf{z}_{im}(t)$, where $\mathbf{z}_{re}(t)$ and $\mathbf{z}_{im}(t)$ are both N-dimensional real-valued zero-mean Gaussian vectors.
- Each latent factor at n time frames follows $\mathbf{z}_{j',re} \sim \text{MN}(0, \frac{A_j}{4}\mathbf{R}_j)$ and $\mathbf{z}_{j',im} \sim \text{MN}(0, \frac{A_j}{4}\mathbf{R}_j)$ for indices $\mathbf{j}' \in \mathcal{S}_j = \{(j_1', j_2') : q_{j_1',1}^2 + q_{j_2',2}^2 = q_j^2\}$, which contains indices of 'ring' j, for $j = 1, ..., J$.
- **Key**: $\mathbf{R}_j$ is formed by ISF: $R_j(k_1, k_2) = f_{\boldsymbol{\theta}}(q_j, \Delta t_k)$ with $\Delta t_k = |k_2 - k_1| \Delta t_{min}$ with Δt_{min} being the length of consecutive frames.

[1] Gu, M., He, Y., Liu, X., & Luo, Y. (2023). Ab initio uncertainty quantification in scattering analysis of microscopy. arXiv preprint arXiv:2309.02468.

Connection between physics and statistics

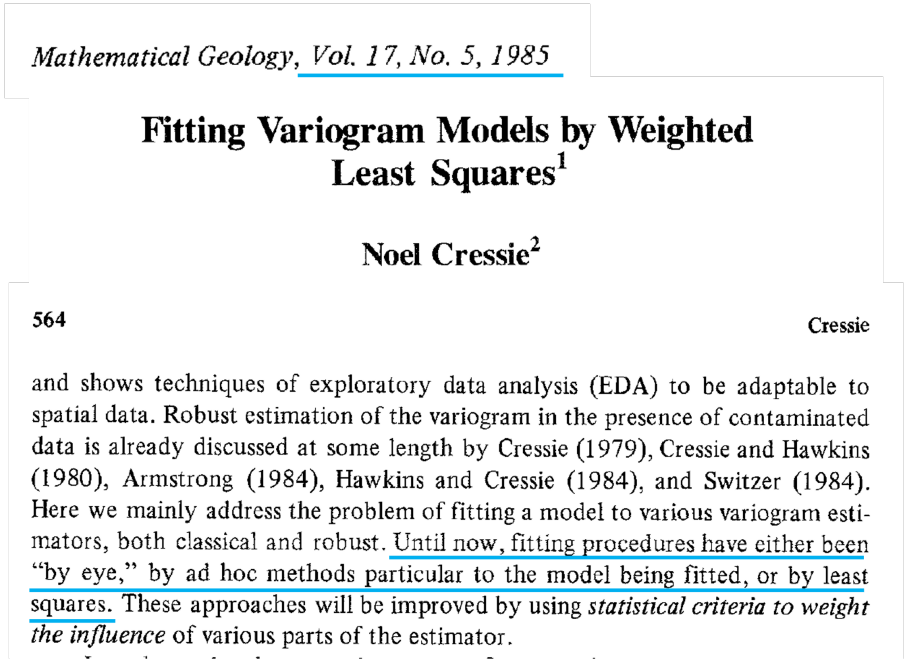
Let $\hat{\mathbf{y}}(t) = \frac{\mathbf{W}\mathbf{y}(t)}{\sqrt{N}} = \hat{\mathbf{y}}_{re}(t) + i\hat{\mathbf{y}}_{im}(t)$ be the normalized Fourier transformed quantities. For any wavevector $\mathbf{q} = (q_1, q_2)$, the image structure function in DDM **equals to** the temporal variogram in the reciprocal space.

$$\mathbb{E}\left[(\hat{y}_{\mathbf{q}}(t+\Delta t) - \hat{y}_{\mathbf{q}}(t))(\hat{y}^*_{\mathbf{q}}(t+\Delta t) - \hat{y}^*_{\mathbf{q}}(t))\right] = A(\mathbf{q})(1 - f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t)) + \bar{B},$$

Connection between physics and statistics

Let $\hat{\mathbf{y}}(t) = \frac{\mathbf{W}\mathbf{y}(t)}{\sqrt{N}} = \hat{\mathbf{y}}_{re}(t) + i\hat{\mathbf{y}}_{im}(t)$ be the normalized Fourier transformed quantities. For any wavevector $\mathbf{q} = (q_1, q_2)$, the image structure function in DDM **equals to** the temporal variogram in the reciprocal space.

$$\mathbb{E}\left[(\hat{y}_{\mathbf{q}}(t+\Delta t) - \hat{y}_{\mathbf{q}}(t))(\hat{y}^*_{\mathbf{q}}(t+\Delta t) - \hat{y}^*_{\mathbf{q}}(t))\right] = A(\mathbf{q})(1 - f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t)) + \bar{B},$$



Mathematical Geology, Vol. 17, No. 5, 1985

Fitting Variogram Models by Weighted Least Squares[1]

Noel Cressie[2]

564 Cressie

and shows techniques of exploratory data analysis (EDA) to be adaptable to spatial data. Robust estimation of the variogram in the presence of contaminated data is already discussed at some length by Cressie (1979), Cressie and Hawkins (1980), Armstrong (1984), Hawkins and Cressie (1984), and Switzer (1984). Here we mainly address the problem of fitting a model to various variogram estimators, both classical and robust. Until now, fitting procedures have either been "by eye," by ad hoc methods particular to the model being fitted, or by least squares. These approaches will be improved by using *statistical criteria to weight the influence* of various parts of the estimator.

Summary: DDM is fitting the temporal variogram at each wavevector in the reciprocal space by the generative model in Eq. (4), whereas aggregating the estimators is notoriously hard.

Efficient estimator, no need to specify a wavevector range

- Integrating out the random factor processes $p(\mathbf{Y} \mid \boldsymbol{\theta}, \boldsymbol{A}_{1:J}, \bar{B}) = \int p(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{A}_{1:J}, \bar{B}) p(\mathbf{Z} \mid \boldsymbol{\theta}, \boldsymbol{A}_{1:J}, \bar{B}) d\mathbf{Z}$, the marginal likelihood of J rings of Fourier transformed quantity follows:

$$\mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{A}_{1:J}, \bar{B}\right) = \prod_{j=1}^{J} \prod_{j' \in \mathcal{S}_j} p_{MN}\left(\hat{\mathbf{y}}_{re,j'};\, \mathbf{0},\, \boldsymbol{\Sigma}_j\right) \times p_{MN}\left(\hat{\mathbf{y}}_{im,j'};\, \mathbf{0},\, \boldsymbol{\Sigma}_j\right),$$

  where $\boldsymbol{\Sigma}_j = \frac{A_j}{4}\mathbf{R}_j + \frac{\bar{B}}{4}\mathbf{I}_n$, $p_{MN}(\cdot)$ is a multivariate normal density, and $\hat{\mathbf{y}}_{re,j'} = [y_{re,j'}(t_1), ..., y_{re,j'}(t_n)]^T$ and $\hat{\mathbf{y}}_{im,j'} = [y_{im,j'}(t_1), ..., y_{im,j'}(t_n)]^T$ are the real and imaginary parts of Fourier transformed intensities at $\mathbf{j}'$.

- Directly maximizing the marginal likelihood needs to search a large parameter space. For any $\bar{B}$, an unbiased estimator of A_j follows

$$A_{est,j} = \frac{2}{S_j n} \sum_{j' \in \mathcal{S}_j} \sum_{k=1}^{n} |\hat{y}_{j'}(t_k)|^2 - \bar{B}, \tag{5}$$

  for $j = 1, ..., J$. Absolute value is used to ensure $A_{est,j}$ nonnegative. Then

$$(\boldsymbol{\theta}_{est}, \bar{B}_{est}) = \text{argmax}_{\boldsymbol{\theta}, \bar{B}} \mathcal{L}\left(\boldsymbol{\theta}, \mathbf{A}_{est,1:J}, \bar{B}\right). \tag{6}$$

Extension to anisotropic processes

- We can split the intermediate function to two coordinates: $f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t) = f_{\boldsymbol{\theta}_1}(q_1, \Delta t) f_{\boldsymbol{\theta}_2}(q_2, \Delta t)$, where $f_{\boldsymbol{\theta}_l}(q_l, \Delta t)$ is an intermediate scattering function for the lth coordinate, and the parameters can be split to $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ being the parameters of $f_{\boldsymbol{\theta}_l}(q_l, \Delta t)$ for $l = 1, 2$.
- From the cumulant approximation, the anisotropic ISF follows

$$f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t) \approx \exp\left(-\frac{q_1^2 \langle \Delta x_1^2(\Delta t)\rangle + q_2^2 \langle \Delta x_2^2(\Delta t)\rangle}{2}\right),$$

where $\langle \Delta x_1^2(\Delta t)\rangle$ and $\langle \Delta x_2^2(\Delta t)\rangle$ are MSD at Δt along the two coordinates, respectively. As the process is anisotropic, the ISF is different for each wavevector in general. Then one can compute the maximum marginal likelihood estimator in Eq. (6) with the anisotropic ISF. Since the amplitude A_j depends on the transformed intensity at zero lag time and image noise [Giavazzi et al., 2009, Nixon-Luke et al., 2022].

Outline

Accelerating the computation by generalized Schur

- Denote no. of indices in each ring $S_j = \#\mathcal{S}_j$, and $\tilde{N} = \sum_{j=1}^{J} S_j$. Directly computing the likelihood T times requires $\mathcal{O}(TJn^3) + \mathcal{O}(T\tilde{N}n^2)$ and $\mathcal{O}(T\tilde{N}n^3)$ operations for isotropic and anisotropic processes, respectively, making $\sim 10^{15}$ operations in computing MMLE for a $500 \times 500 \times 500$ video.
- Video microscopy is equally spaced in time, making $\mathbf{\Sigma}_j$ a symmetric Toeplitz matrix, where the $(k, k+\Delta k)$th entry follows

$$\mathbf{\Sigma}_j(k, k+\Delta k) = A_j f_{\boldsymbol{\theta}}(q_j, \Delta t_k) + \frac{\bar{B}}{4} 1_{\Delta k = 0},$$

  with $\Delta t_k = \Delta k \Delta t_{min}$ with Δt_{min} is the interval between consecutive time frames.
- The inverse and determinant of any $n \times n$ Toeplitz matrix $\mathbf{\Sigma}$ follows

$$\mathbf{\Sigma}^{-1} = \frac{1}{\delta_{n-1}} \left(\mathbf{L}_1 \mathbf{L}_1^T - \mathbf{L}_0 \mathbf{L}_0^T \right), \quad \text{and} \quad |\mathbf{\Sigma}| = \Sigma_{1,1} \prod_{k=1}^{n-1} \delta_k,$$

  where $\mathbf{L}_0$, $\mathbf{L}_1$ (lower triangular Toeplitz matrices) and δ_k can be obtained by the generalized Schur algorithm [Ammar and Gragg, 1987, 1988] reducing the cost from $\mathcal{O}(n^3)$ to $\mathcal{O}(n(\log(n))^2)$. In R, see the "SuperGauss" package [Ling, 2019].
- One can further reduce dimensions by selecting J_0: $\sum_{j=1}^{J_0} A_j / \sum_{j=1}^{J} A_j \geq 1 - \varepsilon$.
- Together we reduce the cost more than $\mathbf{10^5}$ times with nearly no approximation.

Automated analysis without selecting wavevectors

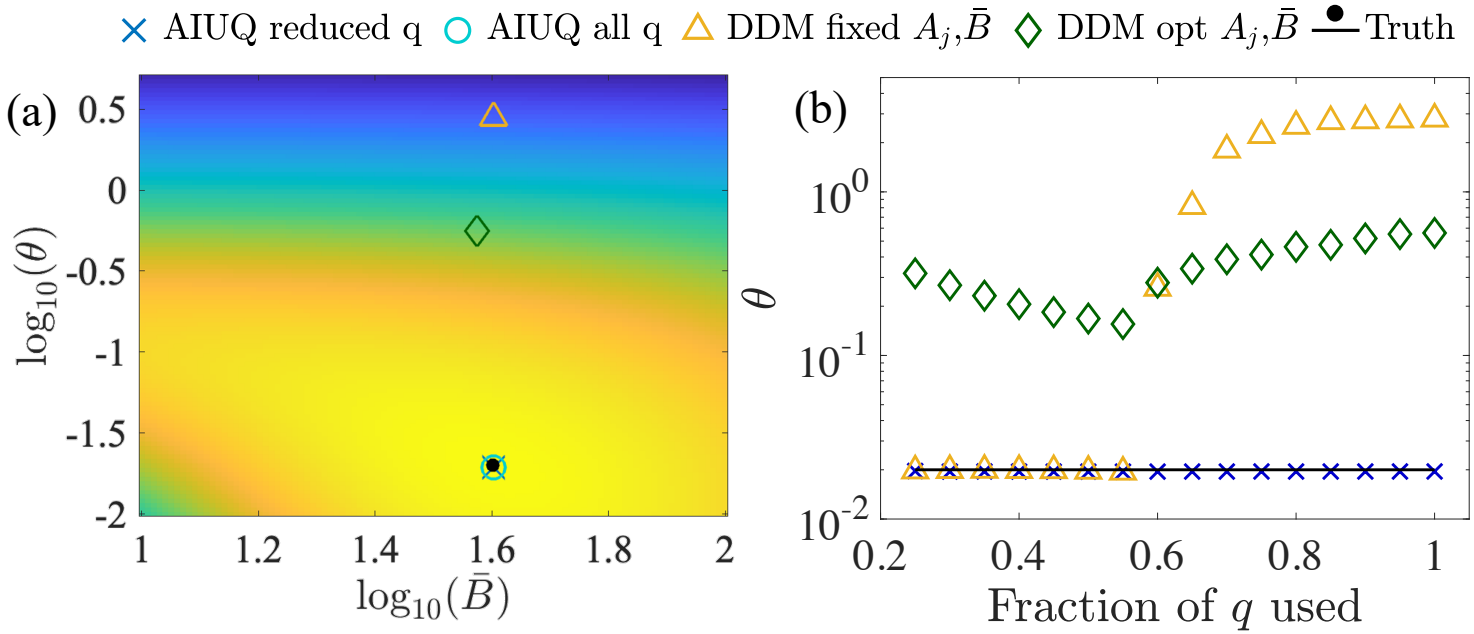


Figure 2: (a) B, D estimates in likelihood space using different methods: AIUQ (x), AIUQ with all q (o), DDM with fixed A, B (△), DDM with optimized A, B (⋄), whereas the solid dot denotes the truth. (b) The D estimate using different methods with the fraction of q used in the estimate and the solid line dots the truth. Each video is $500 \times 500 \times 500$, generated by simulating a Brownian motion of 50 particles with slow dynamics.

Confidence interval and model selection

Interval and model selection was not available in DDM but they are available because of the probabilistic representation through the generative model.

- **Uncertainty from parameter estimation**. We use central limit theory (CLT) [Mardia and Marshall, 1984] to approximate the parameter estimation uncertainty through an asymptotically normal distribution.
  - Information comes from M trajectories ($M \approx$ 50-200) of particles instead of $N \approx 10^5$ pixels at each frame. Discount the likelihood by a factor of M/N seems needed in either CLT, bootstrap or Bayesian sampling for this problem.
- **Uncertainty in discretization of pixels** may be considered by re-estimating $(\boldsymbol{\theta}, \bar{B})$ by letting the associated amplitude of the wavevector q_j to be $q_j - \Delta q_{min}$ and $q_j - \Delta q_{max}$ for $j = 1, ..., J$ separately.
- **Model selection** can be achieved by information criteria, such as Akaike information criterion (AIC): $\text{AIC} = 2p - 2\log(\mathcal{L}_{max})$, or Bayesian information criteria (BIC): $\text{AIC} = p\log(n_{obs}) - 2\log(\mathcal{L}_{max})$, where $\mathcal{L}_{max}$ is the maximum likelihood, p and n_{obs} are the no. of parameter different and obs., respectively. Or one can select the model by the predictive error.

Outline

Setting for simul study 1

- Simulate 6 processes: Brownian motion (BM) with fast and slow dynamics, fractional Brownian motion (FBM) with sub-diffusion and super-diffusion, one OrnsteinUhlenbeck (OU) and a mixture of OU process and FBM.
- Small video ($100 \times 100 \times 100$) and regular videos ($500 \times 500 \times 500$) are simulated.
- Compare AIUQ with MPT and DDM with (i) fixed or pre-estimated A_j and $\bar{B}$, or (ii) optimize parameters at each set of waveveVtors and then average.
- AIUQ with all and reduced wavevectors range are shown. The first 25% of rings of wavevetors typically explains 99% variability and the first 50% explains more than 99.9% variability. Estimation of $\boldsymbol{\theta}$ by different reduced wavevectors are almost exactly the same but estimation of noise can be slightly different.

Parametric	ISF, $f_{\boldsymbol{\theta}}(q, \Delta t)$
BM	$\exp\left(-q^2\sigma^2_{BM}\Delta t/4\right)$
FBM	$\exp\left(-q^2\sigma^2_{FBM}\Delta t^{\alpha}/4\right)$
OU	$\exp\left(-q^2\sigma^2_{OU}(1-\rho^{\Delta t})/4\right)$
OU+FBM	$\exp\left(-q^2(\sigma^2_1\Delta t^{\alpha}+\sigma^2_2(1-\rho^{\Delta t}))/4\right)$
Nonparametric	ISF, $f_{\boldsymbol{\theta}}(q, \Delta t)$
Cumulant approx.	$\exp\left(-q^2\langle\Delta x^2(\Delta t)\rangle/4\right)$

Table 1: A list of models of the intermediate scattering function (ISF).

Processes, true parameters	Regular video			
	DDM fixed	DDM opt	AIUQ reduced q	AIUQ all q
BM, $\sigma^2_{BM} = .020$	2.8	.56	.019 [.019,.020]	**.020** [.019,.020]
BM, $\sigma^2_{BM} = 2.0$	3.6	2.8	**2.0** [2.0, 2.1]	**2.0** [2.0, 2.1]
FBM,	6.7	3.8	**8.1** [7.8,8.5)	**8.1** [7.8,8.5)
$\alpha = .60$	.87	.88	**.59** [.58,.61)	**.59** [.58,.61)
FBM, $\sigma^2_{FBM} = .50$	3.0	.74	**.50** [.47,.54]	**.50** [.46,.55]
$\alpha = 1.4$	1.3	1.0	**1.4** [1.4,1.4]	**1.4** [1.3,1.4]
OU, $\sigma^2_{OU} = 64$	7.0×10^4	22	**61** (44,86)	**61** [35,110)
$\rho = .95$	.71	.57	**.95** (.93,.96]	**.95** (.91,.97)
OU+FBM, $\sigma^2_1 = 2.0$	1.9	2.5	**2.0** (1.6,2.4)	**2.0** (1.4,2.7)
	1.2	.76	**.44** (.41,.48]	**.44** (.38,.51)
$\sigma^2_2 = 9.0$	7.6	2.1	**9.7** (7.1,13]	**9.7**(5.4,17]
$\rho = .85$	.62	.50	**.85** (.81,.89)	**.85** (.77,.91]

Table 2: Parameter estimation for regular videos have contains 500×500 pixels over 500 time frames. The brackets give 95% confidence intervals by AIUQ approaches. All intervals by AIUQ approaches cover the true parameters given in the first column.

Simul Eg 1: Mean squared displacements (MSD)

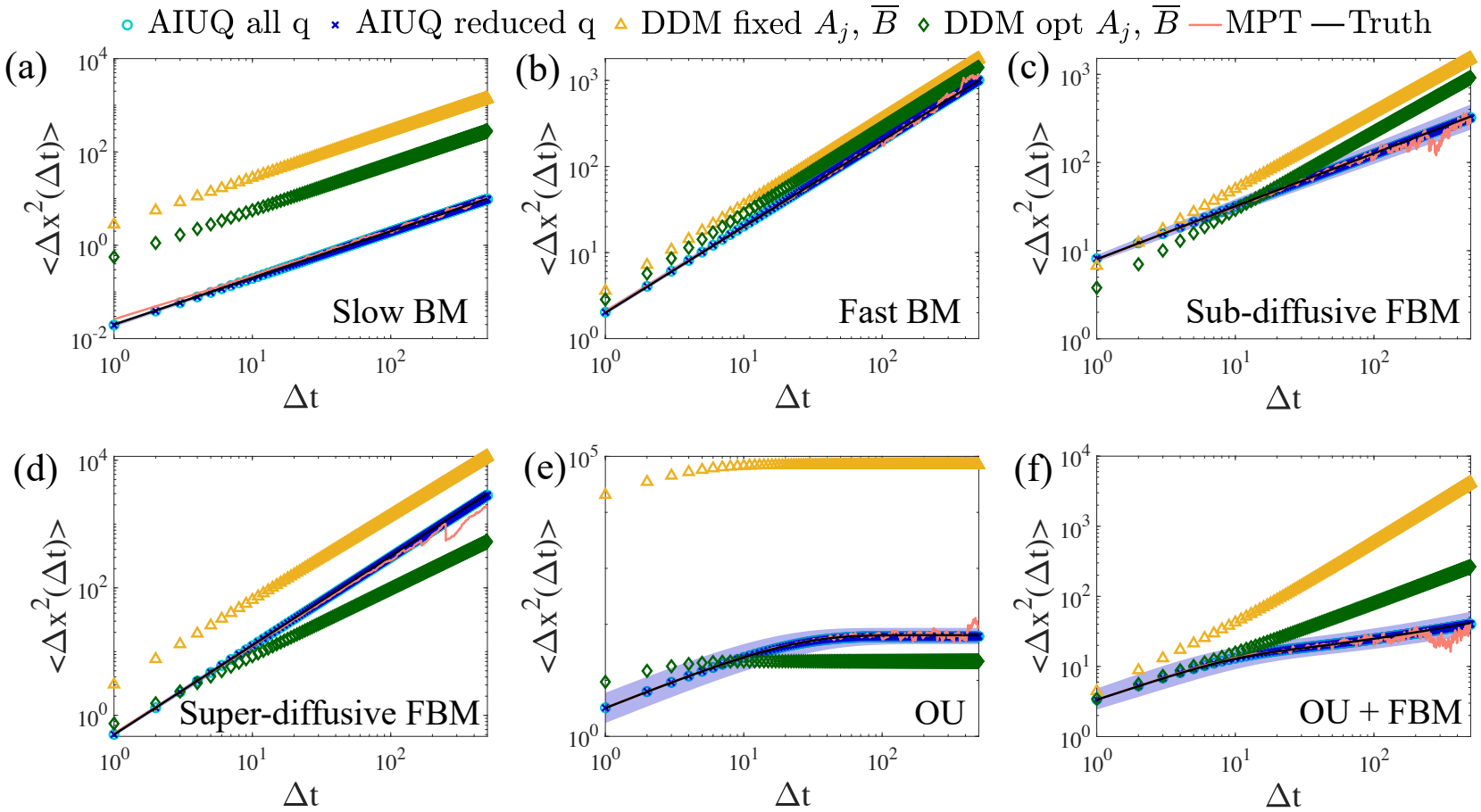


Figure 3: Estimating mean-squared displacements (MSDs) versus lag time for simulated videos containing the trajectories of $M = 50$ simulated particles. The shaded area denotes the 95% interval for AIUQ with reduced wavevectors. The truth overlaps with both AIUQ approaches.

Estimated MSD by model-free methods

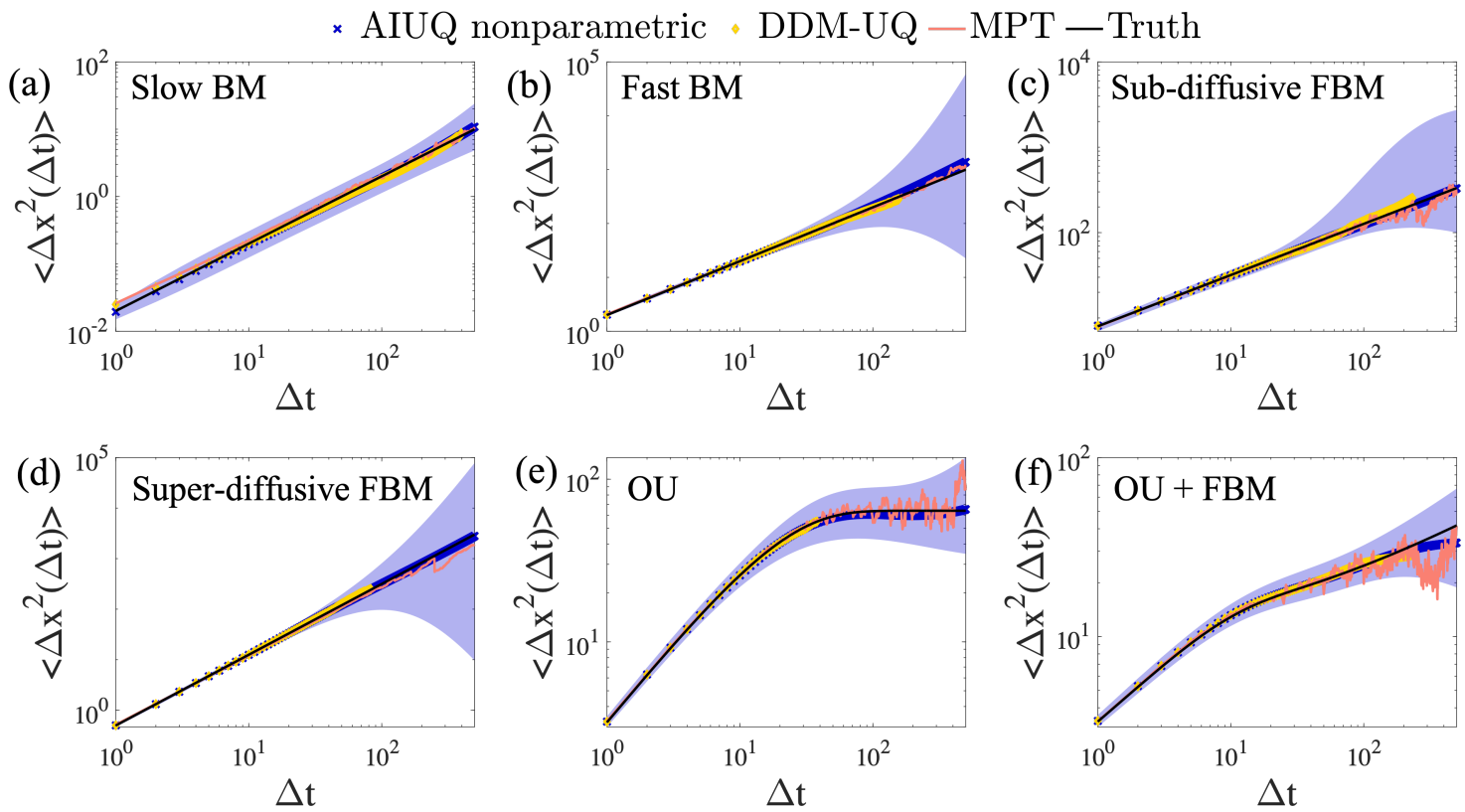


Figure 4: Estimated MSDs by approximating ISF through cumulant theorem $f_{\boldsymbol{\theta}}(\mathbf{q}, \Delta t) \approx \exp\left(-q^2\langle\Delta x^2(\Delta t)\rangle/4\right)$. The AIUQ nonparametric approach is the current approach [1] and DDM-UQ os our previous approach [2].

[1] Gu, M., He, Y., Liu, X., & Luo, Y. (2023). Ab initio uncertainty quantification in scattering analysis of microscopy. arXiv preprint arXiv:2309.02468.
[2] Gu, M., Luo, Y., He, Y., Helgeson, M. E., & Valentine, M. T. (2021). Uncertainty quantification and estimation in differential dynamic microscopy. *Physical Review E*. Matlab package: https://github.com/UncertaintyQuantification/DDM-UQ.

Simul Eg 2: Model selection

- We simulate processes by BM, FBM and OU and fit different model.
- BM is a special case of FBM with $\alpha = 1$, and BM can be well-approximated by OU as $\text{MSD}_{OU} = \sigma_{OU}^2(1-\rho^{\Delta t}) \approx \sigma_{OU}^2(1-\rho)\Delta t$, when $1-\rho$ is close to zero.

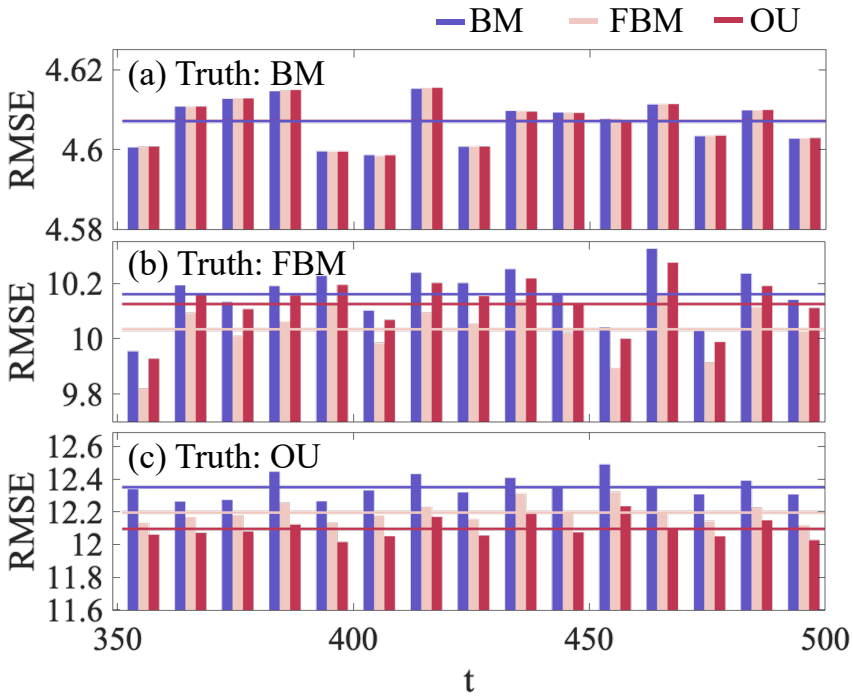


Figure 5: Predictive root Mean Square Error (RMSE) for (a) Brownian Motion ($\sigma_{BM}^2 = 0.02$) (b) Fractional Brownian Motion ($\sigma_{FBM}^2 = 8$, $\alpha = 0.6$) and (c) OrnsteinUhlenbeck process ($\sigma_{OU}^2 = 64$, $\rho = 0.5$) trained on the first 300 lag times. The horizontal lines are the averages. AIC also shows similar results.

Simul Eg 3: Anisotropic systems

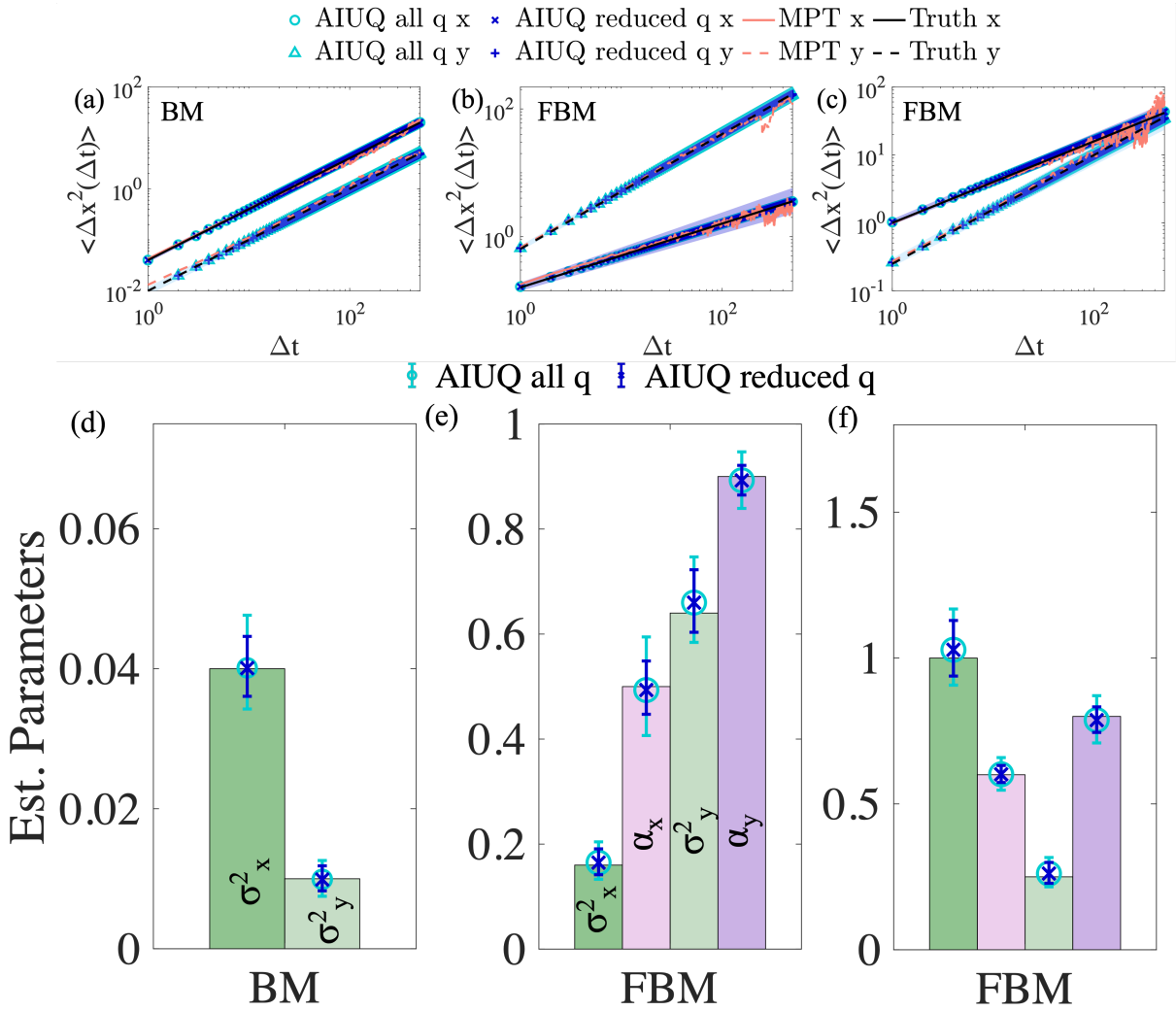


Figure 6: (a)-(c): Estimated MSD by AIUQ, MPT and truth from anisotropic BM and two FBMs. (d)-(f): true parameters (histogram), estimated parameters and 95% intervals by AIUQ approaches.

Outline

Real Eg 1: Beyond diffraction limit in optically dense samples

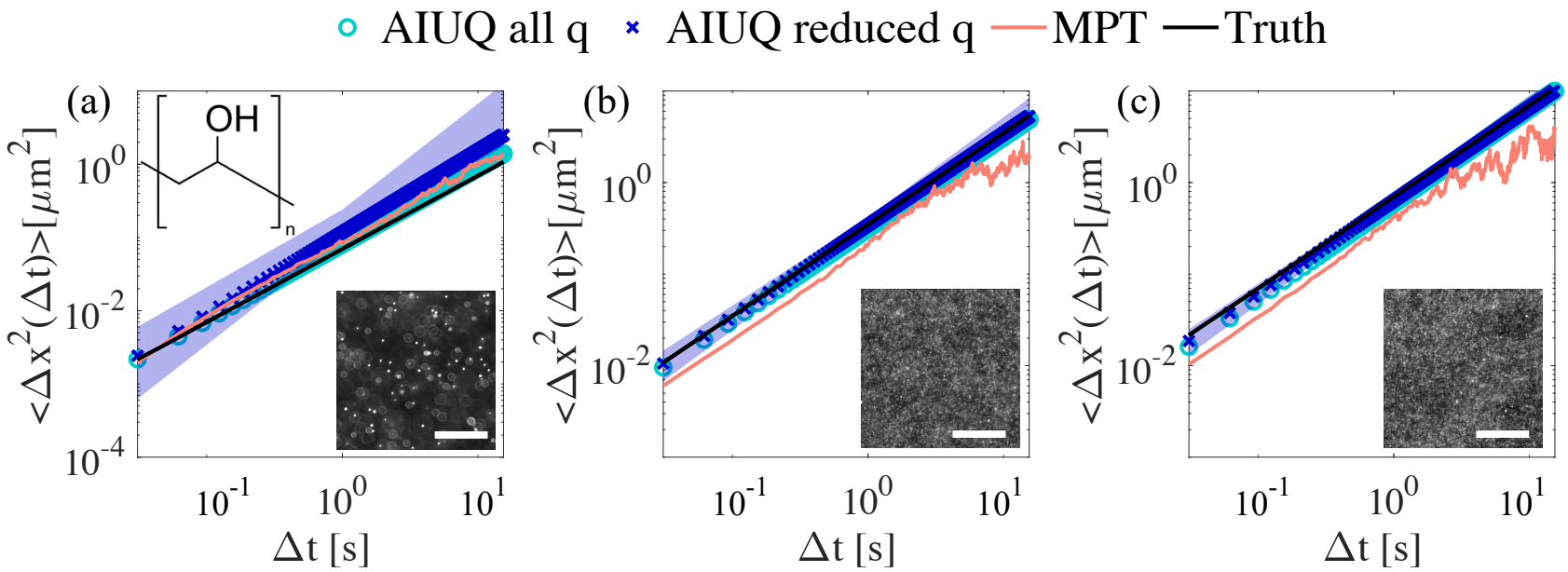


Figure 7: The diagram shows the mean-squared displacement of fluorescent probes in a 4 wt% PVA solution, the molecular structure is also shown in the inset in (a). (a-c) The mean squared displacement against the lag time is plotted for different embedded particle sizes: (a) $2a = 1$ μm, (b) $2a = 200$ nm and (c) $2a = 100$ nm. The thin solid black line denotes literature values, where viscosity $\eta \approx 0.025$ Pa·s. The blue shaded region denotes the confidence interval using AIUQ with a reduced wavevector range.

Real Eg 2: Automated determination of the gelling point

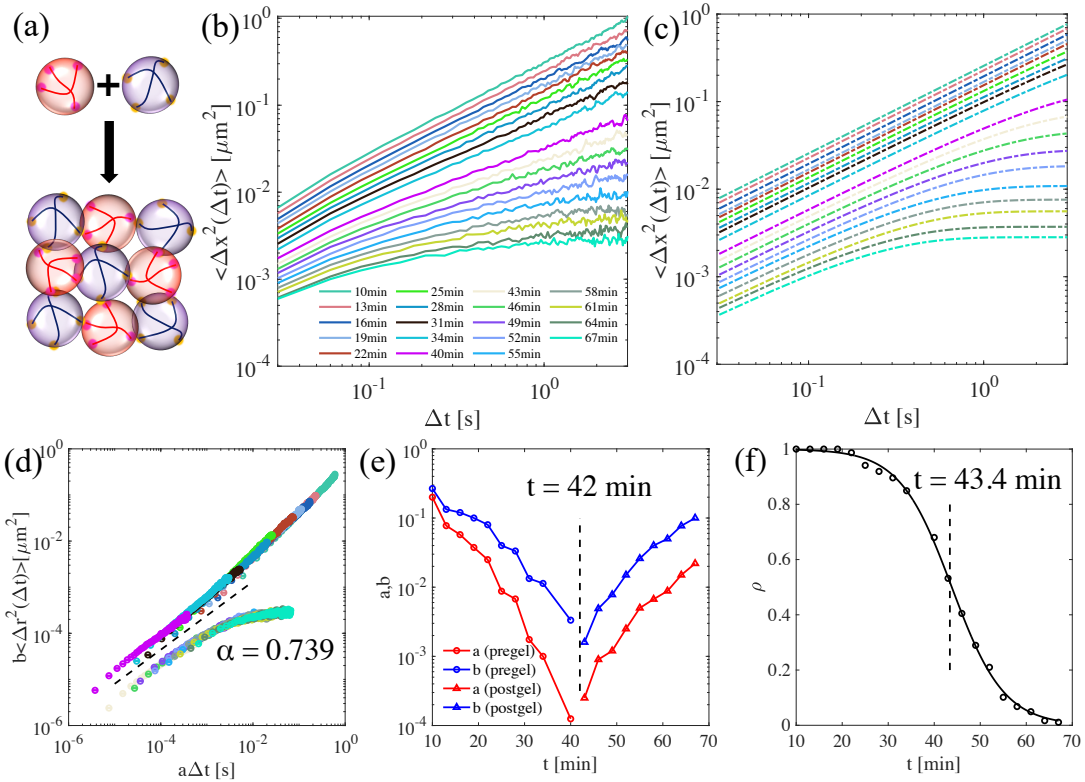


Figure 8: (a) Reaction between tetraPEG-SG and tetraPEGNH2. (b) Estimated MSD by MPT of the probes embedded in the mixed tetra-functional PEG. (c) Estimated MSD by AIUQ. (d) Superposition in (b) using MPT data. (e) Shift factors a (red) and b (blue) for pregel (circles), and postgel (triangles). (f) Estimated ρ from OU from video at different time denoted by black circles. The black solid line denotes generalized logistic fit. The literature value is $\sim$ 44min.

Real Eg 3: Anisotropic systems

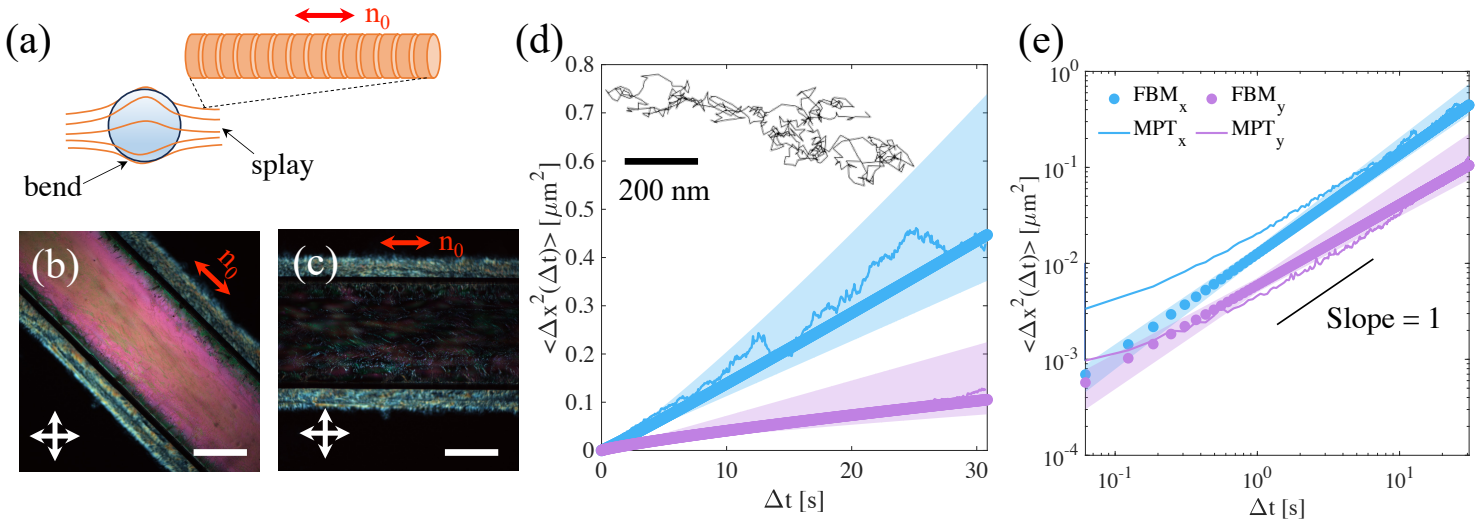


Figure 9: Anisotropic diffusion in lyotropic LC disodium cromoglycate (DSCG), where LC molecules self-assemble into rod-like structure. (a) Schematics of a probe particle moving amongst assembled stacks of DSCG, not drawn to scale. (b-c) Crossed-polarizer images when the channel is aligned either 45^o in (b) or parallel to the alignment direction of the channel in (c). The double-sided arrows denote the direction of the polarizer and analyzer. (d-e) Method comparisons between AIUQ fitting (filled circles) and MPT tracking (solid lines) of the image stack are presented in (d) lin-lin plot and (e) log-log plot. The inset in (d) shows an example particle trajectory. The shaded region denotes 95% confidence interval estimated using AIUQ reduced q.

Outline

Principles of parameter estimation and marginalization

Minimization a loss function can be improved in two steps:

- Build a (large) probabilistic generative model of **untransformed data**.
- Show **equivalence** to a statistical estimator of the generative model, and derive a more efficient estimator by **marginalizing out** random components.

Principles of parameter estimation and marginalization

Minimization a loss function can be improved in two steps:

- Build a (large) probabilistic generative model of **untransformed data**.
- Show **equivalence** to a statistical estimator of the generative model, and derive a more efficient estimator by **marginalizing out** random components.

Marginalization of random components is an iconic feature of Bayesian analysis.

- The art of marginalization includes De Finetti's theorem [De Finetti, 1937], which states that $\{Y_i\}_{i=1}^{\infty}$ is exchangeable, if and only if there exists a random variable $\theta \in \Theta$ with probability distribution $\pi(\theta)$:
$$p(y_1, ..., y_n) = \int \prod_{i=1}^{n} p(y_i|\theta)\pi(\theta)d\theta.$$
- Bayesian model selection [Berger and Pericchi, 1996] and Bayesian model averaging (BMA) [Raftery et al., 1997], are two examples of marginalization. For spatially correlated data, defining an objective prior according to the marginal likelihood can make posterior proper [Berger et al., 2001].
- Marginalization is a natural way for propagating the uncertainty [Lakshminarayanan et al., 2017, Wilson and Izmailov, 2020] and avoids over-fitting.

Eg 1: Generalized probabilistic principal component analysis (GPPCA)

- Consider the generative model $\mathbf{Y} = \mathbf{AZ} + \mathbf{E}$, where $\mathbf{Z} \sim \text{MN}(\mathbf{0}, \mathbf{I}_{dN})$ and $\mathbf{E} \sim \text{MN}(\mathbf{0}, \sigma_0^2 \mathbf{I}_{Nn})$ The estimation by **principal component analysis** of $\mathbf{A}$ has the same linear subspace to the MMLE of this model [Tipping and Bishop, 1999]:

$$\hat{\mathbf{A}} = \text{argmax}_{\mathbf{A}} \int p(\mathbf{Y} \mid \mathbf{A}, \mathbf{Z}) p(\mathbf{Z}) d(\mathbf{Z}),$$

Eg 1: Generalized probabilistic principal component analysis (GPPCA)

- Consider the generative model $\mathbf{Y} = \mathbf{AZ} + \mathbf{E}$, where $\mathbf{Z} \sim \text{MN}(\mathbf{0}, \mathbf{I}_{dN})$ and $\mathbf{E} \sim \text{MN}(\mathbf{0}, \sigma_0^2 \mathbf{I}_{Nn})$ The estimation by **principal component analysis** of $\mathbf{A}$ has the same linear subspace to the MMLE of this model [Tipping and Bishop, 1999]:

$$\hat{\mathbf{A}} = \text{argmax}_{\mathbf{A}} \int p(\mathbf{Y} \mid \mathbf{A}, \mathbf{Z}) p(\mathbf{Z}) d(\mathbf{Z}),$$

- Consider outputs $\mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}), ..., y_N(\mathbf{x}))^T$ modeled by:

$$\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{z}(\mathbf{x}) + \boldsymbol{\epsilon},$$

where $\mathbf{A}_{[N \times d]} = [\mathbf{a}_1, ..., \mathbf{a}_d]$ and factor processes $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), ..., z_d(\mathbf{x}))^T$, with $d \leq N$ and $\boldsymbol{\epsilon} \sim MN(\mathbf{0}, \sigma_0^2 \mathbf{I}_N)$. Denote $\mathbf{Z}_l^T = (z_l(\mathbf{x}_1), ..., z_l(\mathbf{x}_n))^T \sim \text{MN}(\mathbf{0}, \boldsymbol{\Sigma}_l)$. Assume $\mathbf{A}^T\mathbf{A} = \mathbf{I}_d$, marginalizing out $\mathbf{Z}_l^T$ for $l = 1, 2, ..., d$ [Gu and Shen, 2020]:

  - If $\boldsymbol{\Sigma}_1 = ... = \boldsymbol{\Sigma}_d = \boldsymbol{\Sigma}$, the marginal likelihood is maximized when

$$\hat{\mathbf{A}} = \mathbf{UR},$$

    where $\mathbf{U}$ is a $N \times d$ matrix of the first d principal eigenvectors of $\mathbf{G} = \mathbf{Y}(\sigma_0^2 \boldsymbol{\Sigma}^{-1} + \mathbf{I}_n)^{-1}\mathbf{Y}^T$ and $\mathbf{R}$ is a $d \times d$ rotation matrix;
  - If any $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$, denoting $\mathbf{G}_l = \mathbf{Y}(\sigma_0^2 \boldsymbol{\Sigma}_l^{-1} + \mathbf{I}_n)^{-1}\mathbf{Y}^T$, we have:

$$\hat{\mathbf{A}} = \text{argmax}_{\mathbf{A}} \sum_{l=1}^{d} \mathbf{a}_l^T \mathbf{G}_l \mathbf{a}_l, \quad \text{s.t. } \mathbf{A}^T\mathbf{A} = \mathbf{I}_d.$$

Eg 2: Kalman filter is a fast way of marginalization

For a continuous-time linear state space model:

$$y(t_i) = \mathbf{F}(t_i)\mathbf{z}(t_i) + \epsilon_i$$
$$\mathbf{z}(t_i) = \mathbf{G}(t_i)\mathbf{z}(t_{i-1}) + \mathbf{w}(t_i),\ \mathbf{w}(t_i) \sim N(0, \mathbf{W}(t_i))$$

Write $\mathbf{F}(t_i) = \mathbf{F}_i$ $\mathbf{G}(t_i) = \mathbf{G}_i$, $\mathbf{W}(t_i) = \mathbf{W}_i$ and $\mathbf{z}(t_i) = \mathbf{z}_i$. Conditional on $(\sigma^2, \sigma_0^2, \gamma)$:

Eg 2: Kalman filter is a fast way of marginalization

For a continuous-time linear state space model:

$$y(t_i) = \mathbf{F}(t_i)\mathbf{z}(t_i) + \epsilon_i$$
$$\mathbf{z}(t_i) = \mathbf{G}(t_i)\mathbf{z}(t_{i-1}) + \mathbf{w}(t_i),\ \mathbf{w}(t_i) \sim N(0, \mathbf{W}(t_i))$$

Write $\mathbf{F}(t_i) = \mathbf{F}_i$ $\mathbf{G}(t_i) = \mathbf{G}_i$, $\mathbf{W}(t_i) = \mathbf{W}_i$ and $\mathbf{z}(t_i) = \mathbf{z}_i$. Conditional on $(\sigma^2, \sigma_0^2, \gamma)$:
(**Kalman Filter** [Kalman, 1960]) Let $\mathbf{z}_{i-1}|\mathbf{y}_{1:i-1} \sim \text{MN}(\mathbf{m}_{i-1}, \mathbf{C}_{i-1})$. For $i = 2, ..., n$:

(i) The one-step-ahead predictive distribution of $\mathbf{z}_i$ given $\mathbf{y}_{1:i-1}$ is $\mathbf{z}_i|\mathbf{y}_{1:i-1} \sim \text{MN}(\mathbf{b}_i, \mathbf{B}_i)$. with $\mathbf{b}_i = \mathbf{G}_i\mathbf{m}_{i-1}$ and $\mathbf{B}_i = \mathbf{G}_i\mathbf{C}_{i-1}\mathbf{G}_i^T + \mathbf{W}_i$.

(ii) The one-step-ahead predictive distribution of Y_i given $\mathbf{y}_{1:i-1}$ is $Y_i|\mathbf{y}_{1:i-1} \sim N(f_i, Q_i)$, with $f_i = \mathbf{F}_i\mathbf{a}_i$, and $Q_i = \mathbf{F}_i\mathbf{B}_i\mathbf{F}_i^T + \sigma_0^2$.

(iii) The filtering distribution of $\mathbf{z}_i$ given $\mathbf{y}_{1:i}$ is $\mathbf{z}_i|\mathbf{y}_{1:i} \sim \text{MN}(\mathbf{m}_i, \mathbf{C}_i)$, with $\mathbf{m}_i = \mathbf{b}_i + \mathbf{B}_i\mathbf{F}_i^T Q_i^{-1}(y_i - f_i)$ and $\mathbf{C}_i = \mathbf{B}_i - \mathbf{B}_i\mathbf{F}_i^T Q_i^{-1}\mathbf{F}_i\mathbf{B}_i$.

The **likelihood** $p(\mathbf{y}_{1:n}) = p(y_1)p(y_2|y_1)...p(y_n|y_1, ..., y_{n-1})$ follows
$p(\mathbf{y}_{1:n}) = (2\pi)^{-n/2}\prod_{i=1}^n Q_i^{-1/2}\exp\left\{-\sum_{i=1}^n \frac{(y_i - f_i)^2}{Q_i}\right\}$.

Remark [Gu et al., 2022a]:

- Latent states $\mathbf{z}_{1:n}$ are marginalized (integrated) out.
- Computing the likelihood only takes $\mathcal{O}(n)$ instead of $\mathcal{O}(n^3)$ operations.

Eg 2: Kalman filter is a fast way of marginalization

For a continuous-time linear state space model:

$$y(t_i) = \mathbf{F}(t_i)\mathbf{z}(t_i) + \epsilon_i$$
$$\mathbf{z}(t_i) = \mathbf{G}(t_i)\mathbf{z}(t_{i-1}) + \mathbf{w}(t_i),\ \mathbf{w}(t_i) \sim N(0, \mathbf{W}(t_i))$$

Write $\mathbf{F}(t_i) = \mathbf{F}_i$ $\mathbf{G}(t_i) = \mathbf{G}_i$, $\mathbf{W}(t_i) = \mathbf{W}_i$ and $\mathbf{z}(t_i) = \mathbf{z}_i$. Conditional on $(\sigma^2, \sigma_0^2, \gamma)$:
(**Kalman Filter** [Kalman, 1960]) Let $\mathbf{z}_{i-1}|\mathbf{y}_{1:i-1} \sim \text{MN}(\mathbf{m}_{i-1}, \mathbf{C}_{i-1})$. For $i = 2, ..., n$:

(i) The one-step-ahead predictive distribution of $\mathbf{z}_i$ given $\mathbf{y}_{1:i-1}$ is $\mathbf{z}_i|\mathbf{y}_{1:i-1} \sim \text{MN}(\mathbf{b}_i, \mathbf{B}_i)$. with $\mathbf{b}_i = \mathbf{G}_i\mathbf{m}_{i-1}$ and $\mathbf{B}_i = \mathbf{G}_i\mathbf{C}_{i-1}\mathbf{G}_i^T + \mathbf{W}_i$.

(ii) The one-step-ahead predictive distribution of Y_i given $\mathbf{y}_{1:i-1}$ is $Y_i|\mathbf{y}_{1:i-1} \sim N(f_i, Q_i)$, with $f_i = \mathbf{F}_i\mathbf{a}_i$, and $Q_i = \mathbf{F}_i\mathbf{B}_i\mathbf{F}_i^T + \sigma_0^2$.

(iii) The filtering distribution of $\mathbf{z}_i$ given $\mathbf{y}_{1:i}$ is $\mathbf{z}_i|\mathbf{y}_{1:i} \sim \text{MN}(\mathbf{m}_i, \mathbf{C}_i)$, with $\mathbf{m}_i = \mathbf{b}_i + \mathbf{B}_i\mathbf{F}_i^T Q_i^{-1}(y_i - f_i)$ and $\mathbf{C}_i = \mathbf{B}_i - \mathbf{B}_i\mathbf{F}_i^T Q_i^{-1}\mathbf{F}_i\mathbf{B}_i$.

The **likelihood** $p(\mathbf{y}_{1:n}) = p(y_1)p(y_2|y_1)...p(y_n|y_1, ..., y_{n-1})$ follows
$p(\mathbf{y}_{1:n}) = (2\pi)^{-n/2}\prod_{i=1}^n Q_i^{-1/2}\exp\left\{-\sum_{i=1}^n \frac{(y_i - f_i)^2}{Q_i}\right\}$.

Remark [Gu et al., 2022a]:

- Latent states $\mathbf{z}_{1:n}$ are marginalized (integrated) out.
- Computing the likelihood only takes $\mathcal{O}(n)$ instead of $\mathcal{O}(n^3)$ operations.

Is Kalman a Bayesian? Or perhaps he thinks like a Bayesian?

Eg 3: A generative model of DMD

- Let us consider $N \times n$ matrix $\mathbf{Y}$ at n time points, where the $\mathbf{Y}_{1:n-1}$ and $\mathbf{Y}_{2:n}$. Dynamic mode decomposition (DMD) [Schmid, 2010, Tu et al., 2014] reconstructs the input-output pair, $\mathbf{y}(\mathbf{x}_{t+1}) \approx \mathbf{A}\mathbf{y}(\mathbf{x}_t)$, by

$$\hat{\mathbf{A}} = \text{argmin}_{\mathbf{A}} \|\mathbf{Y}_{2:n} - \mathbf{A}\mathbf{Y}_{1:n-1}\| = \mathbf{Y}_{2:n}(\mathbf{Y}_{1:n-1})^{+}, \tag{7}$$

  which produces discretized approximation of Koopman modes and eigenvalues [Rowley et al., 2009, Brunton et al., 2022].

- (**A generative model for DMD**). In [Gu et al., 2023c], we found $\hat{\mathbf{A}}$ in Eq. (7) is the MLE of $\mathbf{A}$ of the model:

$$\mathbf{y}(\mathbf{x}_{t+1}) = \mathbf{A}\mathbf{y}(\mathbf{x}_t) + \boldsymbol{\varepsilon}_{t+1}, \tag{8}$$

  where $\boldsymbol{\varepsilon}_{t+1} \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ is a vector of Gaussian distributions with a positive definite covariance matrix $\boldsymbol{\Sigma}_\varepsilon$, for any $t = 1, 2, ..., n$.

- The **predictive distribution** for DMD follows [Gu et al., 2023c]:

$$\left(\mathbf{y}(\mathbf{x}_{t^*}) \mid \mathbf{Y}, \hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}_\varepsilon\right) \sim \mathcal{MN}\left(\hat{\mathbf{y}}(\mathbf{x}_{t^*}), \sum_{i=0}^{t^*-n-1} \hat{\mathbf{A}}^i \hat{\boldsymbol{\Sigma}}_\varepsilon (\hat{\mathbf{A}}^T)^i\right), \tag{9}$$

  where $\hat{\mathbf{y}}(\mathbf{x}_{t^*}) = \hat{\mathbf{A}}^{t^*-n}\mathbf{y}(\mathbf{x}_n)$ for any $t^* > n$.

Limitation of DMD: (1) Noise is not modeled; (2) The process is non-differentiable.

Outline

Geological hazard quantification

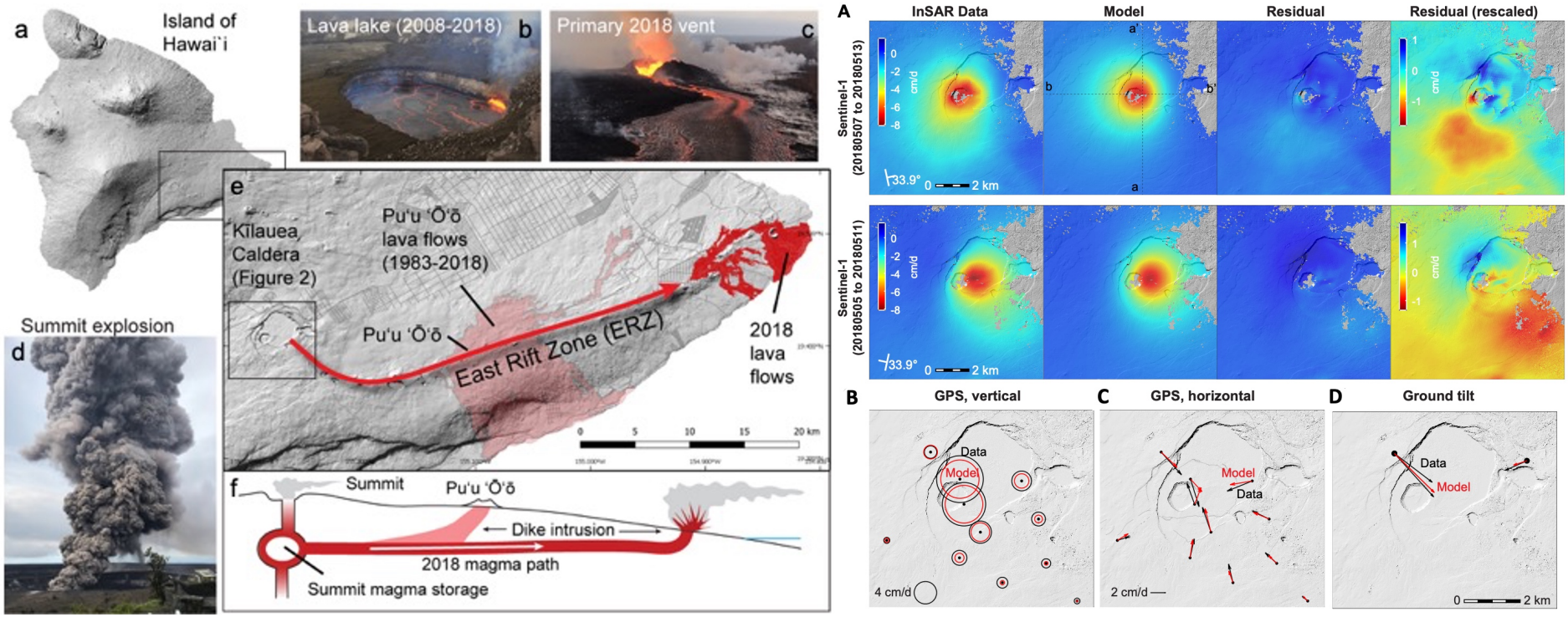


Figure 10: 1 Kīlauea Volcano eruption in 2018, ground deformation observations and model fit.

[3] Anderson, K., Johanson, I., Patrick, M., Gu, M., Segall, P., Poland M., Montgomery-Brown E. and Miklius, A. (2019). Magma reservoir failure and the onset of caldera collapse at Kīlauea volcano in 2018. *Science*, 366(6470).
[4] Gu, M., Palomo, J., & Berger, J. O. (2019). RobustGaSP: Robust Gaussian stochastic process emulation in R. The R Journal. MATLAB version: doi:10.5281/zenodo.3370575. Python version: https://github.com/UncertaintyQuantification/P_RobustGP. Fast prediction and UQ for expensive computer models with $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^N$ where N can be e.g. 10^5. It was also used routinely for nonparametric regression [Edwards et al., 2021].

Outline

Calibrating imperfect mathematical models

- Denote $f^M(\mathbf{x}, \boldsymbol{\theta})$ a mathematical model at observable input $\mathbf{x}$ and unobservable inputs $\boldsymbol{\theta}$. When the mathematical model is *imperfect*, it is usual to model

$$y^F(\mathbf{x}) = \underbrace{f^M(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x})}_{y^R(\mathbf{x})} + \epsilon,$$

  where $\delta(\mathbf{x})$ models the discrepancy and $\epsilon \sim \mathrm{N}(0, \sigma_0^2)$ is a zero-mean independent Gaussian noise. $y^R(\mathbf{x})$ denotes the reality at $\mathbf{x} \in \mathcal{X}$.

- Assume the trend and intercept are properly defined in f^M. [Kennedy and O'Hagan, 2001] modeled δ via a Gaussian process (GP) or Gaussian stochastic process (GaSP), such that any marginal distribution follows

$$(\delta(\mathbf{x}_1), ..., \delta(\mathbf{x}_n))^T \sim \mathrm{MN}(\mathbf{0}, \sigma^2 \mathbf{R}),$$

  where $R_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$, with $K(\cdot, \cdot)$ being a kernel function. This approach is used in many studies (e.g. [Bayarri et al., 2007, Higdon et al., 2008, Liu et al., 2009, Paulo et al., 2012, Arendt et al., 2012]).

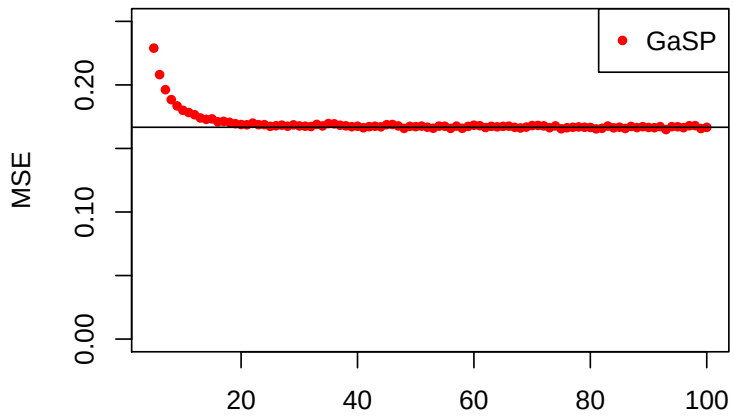
Identifiability and inconsistency

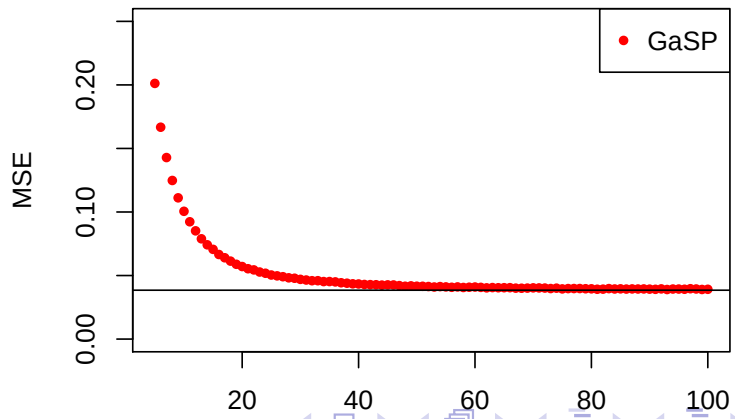
Example 1 ([Gu et al., 2023a])

Assume noise-free data: $y^F(x) = f^M(x, \theta) + \delta(x)$ and $\delta(\cdot) \sim GaSP(0, \sigma^2 K(\cdot, \cdot))$, where $f^M(x, \theta) = \theta$ and $K(x_i, x_j) = \exp(-|x_i - x_j|/\gamma)$. $y^F(x_i)$ is equally spaced at $x_i \in [0, 1]$, for $i = 1, ..., n$. When the sample size $n \to \infty$, the maximum likelihood estimator (MLE) $\hat{\theta}_{MLE} = (\mathbf{1}_n^T \mathbf{R}^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^T \mathbf{R}^{-1} \mathbf{y}^F$ has the limiting distribution

$$\hat{\theta}_{MLE} \sim N\left(\theta, \frac{2\sigma^2\gamma}{2\gamma + 1}\right).$$

Mean squared error (MSE) of MLE of θ (red curve) and theoretical bound (black curve) for $\gamma = 0.1$ (left) and for $\gamma = 0.02$ (right) are shown above.





L_2 calibration and LS calibration

- **The L_2 calibration** in [Tuo and Wu, 2015, 2016].
  - Step 1, use an estimator $\hat{y}^R(\cdot)$ of the reality using a nonparametric regression model without a computer model.
  - Step 2, estimate $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}}_{L_2} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \int_{\mathbf{x}\in\mathcal{X}} (\hat{y}^R(\mathbf{x}) - f^M(\mathbf{x}, \boldsymbol{\theta}))^2 d\mathbf{x}.$$

  - In step 1, The L_2 approach does not use the mathematical model to predict the reality, while mathematical model is useful for prediction.
- **The LS calibration** in [Wong et al., 2017]:
  - Step 1, use a least squared estimator for calibration parameter $\hat{\boldsymbol{\theta}}_{LS} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^n (y^F(\mathbf{x}_i) - f^M(\mathbf{x}_i, \boldsymbol{\theta}))^2$.
  - Step 2, estimate the model discrepancy by a nonparameteric regression by fitting the residual $y^F(\mathbf{x}_i) - f^M(\mathbf{x}_i, \hat{\boldsymbol{\theta}}_{LS})$.

The scaled Gaussian stochastic process

We build the scaled Gaussian stochastic process (S-GaSP) to model the discrepancy $\delta_z(\cdot)$ [Gu and Wang, 2018]:

$$\begin{aligned}\delta_z(\mathbf{x}) &= \left\{\delta(\mathbf{x}) \mid \int_{\boldsymbol{\xi}\in\mathcal{X}} \delta^2(\boldsymbol{\xi})d\boldsymbol{\xi} = Z\right\},\\ \delta(\cdot) &\sim \text{GaSP}(0, \sigma^2 K(\cdot,\cdot)),\end{aligned} \tag{10}$$

Given $Z = z$, the S-GaSP becomes a GaSP constrained at the space $\int_{\mathbf{x}\in\mathcal{X}} \delta^2(\mathbf{x})d\mathbf{x} = z$. Let $p_\delta(Z = z)$ be the density of Z induced by GaSP. Conditional on all the parameters, the default choice of $p_Z(\cdot)$ is

$$p_Z(z) = \frac{g_Z(z)\, p_\delta(Z = z)}{\int_0^\infty g_Z(t)\, p_\delta(Z = t)\, dt}, \tag{11}$$

with the scaling function $g_Z(z) = \frac{\lambda_z}{2\sigma^2}\exp\left(-\frac{\lambda_z z}{2\sigma^2}\right)$.

Remark: We showed any GaSP (or Gaussian Markov random field) is a special case of S-GaSP with $\lambda_z = 0$. We will discuss the default choice of λ_z.

Properties of S-GaSP: orthogonal sequence representation

Let ρ_k and $\phi_k(\cdot)$ be the ordered eigenvalues and eigenfunctions of kernel $K(\cdot,\cdot)$.

Lemma 1 ([Gu et al., 2022b])

- *Any marginal distribution of the S-GaSP is given below:*

$$[\delta_z(\mathbf{x}_1),...,\delta_z(\mathbf{x}_n) \mid \sigma^2\mathbf{R}_z] \sim \mathrm{MN}(\mathbf{0}, \sigma^2\mathbf{R}_z),$$

where the (i, j) entry of $\mathbf{R}_z$ *is* $K_z(\mathbf{x}_i,\mathbf{x}_j)$ *below*

$$K_z(\mathbf{x}_i,\mathbf{x}_j) = \sum_{k=1}^{\infty} \frac{\rho_k}{1+\lambda_z\rho_k}\phi_k(\mathbf{x}_i)\phi_k(\mathbf{x}_j).$$

- *For any* $\mathbf{x} \in \mathcal{X}$, *the S-GaSP* $\delta_z(\cdot)$ *can be written as*

$$\delta_z(\mathbf{x}) = \sigma\sum_{k=1}^{\infty} \frac{\rho_k}{1+\lambda_z\rho_k} Z_k\phi_k(\mathbf{x}),$$

where $Z_k \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$.

Remark: These two results can be viewed as the Mercer's theorem and Karhunen-Loève theorem for S-GaSP.

Eigenvalues in GaSP and the associated S-GaSP

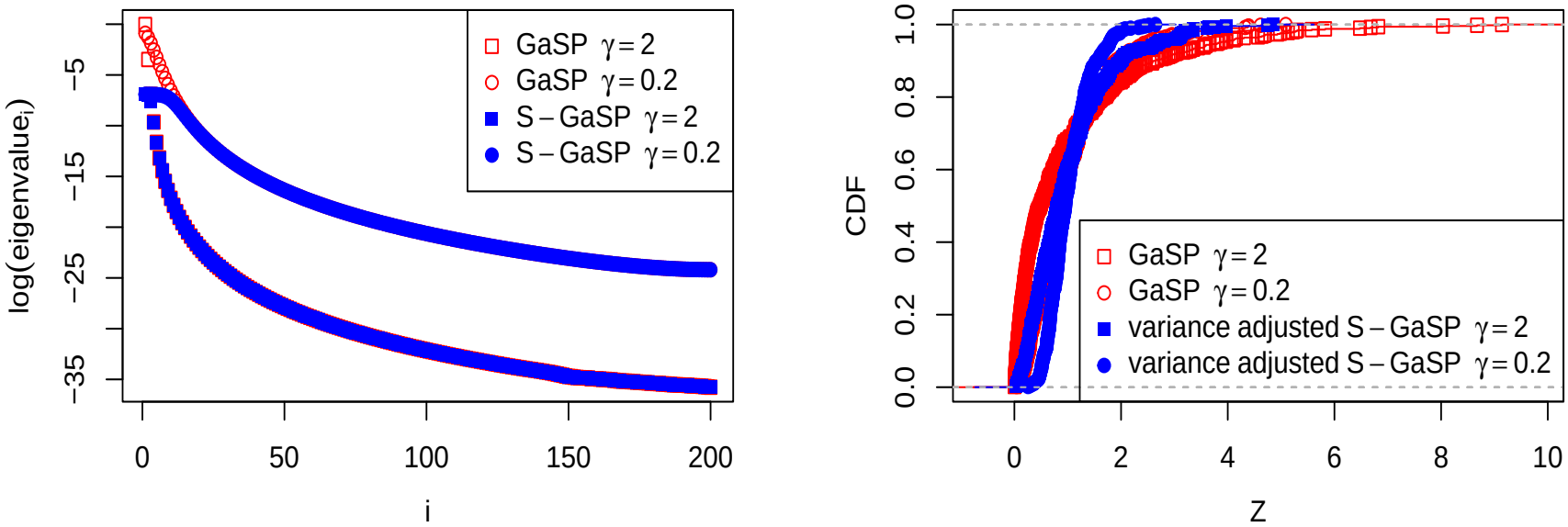


Figure 11: Log eigenvalues and cumulative distribution function (CDF) of the L_2 loss based on $n = 200$ inputs equally spaced from $[0, 1]$. In the left panel, red symbols give the logarithm of empirical eigenvalues $\tilde{\rho}_i/n$ of the covariance matrix parameterized by the unit variance Matérn kernel in (14) with two range parameters. The scaled eigenvalues $\frac{\tilde{\rho}_i/n}{\tilde{\rho}_i/n+\lambda_z}$ by S-GaSP, with $\lambda_z = 10^3$, are graphed by blue symbols. In the right panel, the empirical CDF of the approximated L_2 loss $Z = \int \delta^2(x)dx$ is shown. The variance parameter is adjusted such that the summation of the eigenvalues to be the same.

MLE as the penalized kernel ridge regression in S-GaSP

Denote $\mathcal{L}_z(\boldsymbol{\theta})$ the likelihood for $\boldsymbol{\theta}$ in the S-GaSP calibration in (10) after marginalizing out $\boldsymbol{\delta}_z$.

Lemma 2

The **MLE** $\hat{\boldsymbol{\theta}}_z := \text{argmax}_{\boldsymbol{\theta}} \mathcal{L}_z(\boldsymbol{\theta})$ *and* **predictive mean** $\hat{\delta}_z(\cdot) := \mathbb{E}[\delta_z(\cdot) \mid \mathbf{y}^F, \hat{\boldsymbol{\theta}}_z, \lambda, \lambda_z]$ *are equivalent to the estimator of the penalized kernel ridge regression below*

$$(\hat{\boldsymbol{\theta}}_z, \hat{\delta}_z(\cdot)) = \underset{\delta(\cdot)\in\mathcal{H}, \boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmin}} \ \ell_z(\boldsymbol{\theta}, \delta), \qquad \textit{where}$$

$$\ell_z(\boldsymbol{\theta}, \delta) = \left[\frac{1}{n}\sum_{i=1}^{n}(y^F(\mathbf{x}_i) - f^M(\mathbf{x}_i, \boldsymbol{\theta}) - \delta(\mathbf{x}_i))^2 + \lambda||\delta||^2_{\mathcal{H}_z}\right],$$

with $||\delta||^2_{\mathcal{H}_z} := ||\delta||^2_{\mathcal{H}} + \lambda_z||\delta||^2_{L_2(\mathcal{X})}$.

Remark: both $||\delta||^2_{\mathcal{H}}$ and $||\delta||^2_{L_2(\mathcal{X})}$ are in the penalty term.

Theorem 1

Define the estimator for the reality in S-GaSP for any $\mathbf{x} \in \mathcal{X}$

$$\hat{y}_z^R(\mathbf{x}, \hat{\boldsymbol{\theta}}_z) := f^M(\mathbf{x}, \hat{\boldsymbol{\theta}}_z) + \hat{\delta}_z(\mathbf{x}).$$

i. *(Convergence to the reality.) Assume the reality resides in the* p *dimensional Sobolev space with order* $m > p/2$*, we have*

$$\|y^R(\cdot) - \hat{y}_z^R(\cdot, \hat{\boldsymbol{\theta}}_z)\|_{L_2(\mathcal{X})} = O_p(n^{-\frac{m}{2m+p}}),$$

by choosing $\lambda = O(n^{-\frac{2m}{2m+p}})$ *and* $\lambda_z = O(\lambda^{-1/2})$ *and the* j*th ordered eigenvalue of* $K(\cdot,\cdot)$ *satisfies* $c_\rho j^{-2m/p} \leq \rho_j \leq C_\rho j^{-2m/p}$ *for some* $c_\rho > 0$ *and* $C_\rho > 0$*.*

ii. *(Parameter convergence.) Under some more conditions, one has*

$$\hat{\boldsymbol{\theta}}_z = \boldsymbol{\theta}_{L_2} + O_p(n^{-\frac{m}{2m+p}}),$$

with the same choice of $K(\cdot,\cdot)$*,* λ *and* λ_z*.*

Remark: (1) The second result does not hold for GaSP calibration. (2) The proof is not trivial, as $||\delta||^2_{\mathcal{H}_z} = ||\delta||^2_{\mathcal{H}} + \lambda_z||\delta||^2_{L_2(\mathcal{X})}$ is not bounded when $n \to \infty$, whereas $||\delta||^2_{\mathcal{H}}$ is typically bounded.

Theoretical bound and numerical bound

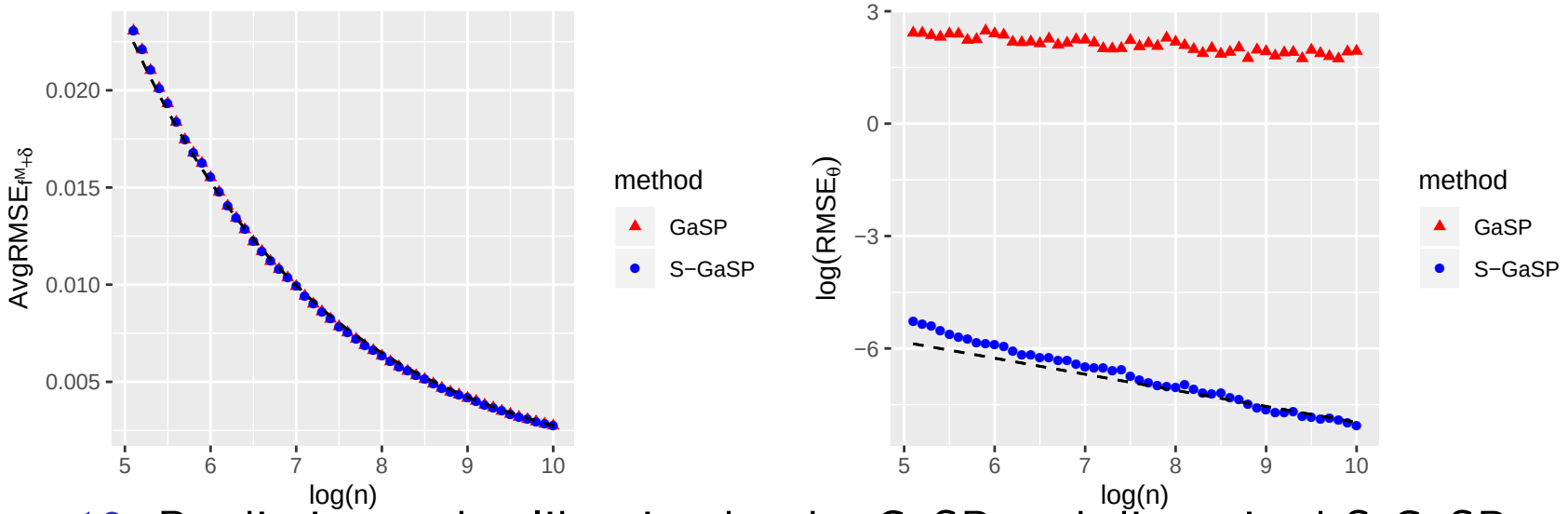


Figure 12: Prediction and calibration by the GaSP and discretized S-GaSP calibration models based on a function in the Soblev space of order 3. In the left panel, the average RMSE of predicting the reality by the GaSP calibration and that of the discretized S-GaSP calibration are graphed as the red triangles and blue dots, respectively; the black curve is $n^{-m/(2m+p)}/5$, representing the theoretical upper bound (up to a constant). In the right panel, the natural logarithm of the RMSE of the calibration parameter of the GaSP and discretized S-GaSP calibration are graphed as the red triangles and blue circles, respectively; the black line is $\log(n^{-m/(2m+p)}/40)$, the theoretical upper bound (up to a constant). $\lambda = n^{-2m/(2m+p)} \times 10^{-4}$ and $\lambda_z = \lambda^{-1/2}$ are assumed.

The discretized scaled Gaussian stochastic process

- W.l.o.g., let the discretization points be the observed inputs, i.e. $\mathbf{x}_i^C = \mathbf{x}_i$ for $i = 1, ..., N_C$ and $N_C = n$. The discretized S-GaSP is to replace δ_z in Equation (10) by [Gu and Wang, 2018]:

$$\delta_{z_d}(\mathbf{x}) = \left\{ \delta(\mathbf{x}) \mid \frac{1}{n}\sum_{i=1}^{n} \delta(\mathbf{x}_i)^2 = Z_d \right\}. \tag{12}$$

- Marginalizing out Z_d, $\delta_{z_d}(\cdot)$ is a zero-mean GaSP with the covariance function

$$\sigma^2 K_{z_d}(\mathbf{x}_a, \mathbf{x}_b) = \sigma^2 (K(\mathbf{x}_a, \mathbf{x}_b) - \mathbf{r}^T(\mathbf{x}_a)\tilde{\mathbf{R}}^{-1}\mathbf{r}(\mathbf{x}_b)) \tag{13}$$

  for any $\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}$, where $\tilde{\mathbf{R}} := \mathbf{R} + n\mathbf{I}_n/\lambda_z$.
- Denote the ith largest eigenvalues of $\mathbf{R}$ and $\tilde{\mathbf{R}}$ by $\tilde{\rho}_i$ and $\tilde{\rho}_{z,i}$, respectively. Note $\frac{\tilde{\rho}_{z,i}}{n} = \frac{\frac{\tilde{\rho}_i}{n}}{(1+\lambda_z \frac{\rho_i}{n})}$, which coincides with the shrinkage of eigenvalues of a non-discretized S-GaSP, by using $\tilde{\rho}_i/n$ as approximation of ρ_i.
- Default choice $\lambda_z = (c_z \lambda ||\tilde{\boldsymbol{\gamma}}||)^{-1/2}$, where by default $c_z = 1$, $\lambda = \sigma_0^2/(\sigma^2 n)$ and $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, ..., \tilde{\gamma}_p)^T$, with $\tilde{\gamma}_i$ being the range parameter normalized by the input length, as it satisfies the convergence conditions in [Gu et al., 2022b]. Besides, only the ratio of the range and variance parameters can be estimated consistently [Zhang, 2004].

Marginal distributions

Recall $\lambda = \sigma_0^2/(n\sigma^2)$. We have the following predictive distribution of field observations.

Theorem 2

The predictive distribution of the field data at any $\mathbf{x} \in \mathcal{X}$ by the discretized S-GaSP calibration model in (12) is a multivariate normal distribution

$$y^F(\mathbf{x}) \mid \mathbf{y}^F, \boldsymbol{\theta}, \sigma_0^2, \lambda, \lambda_z \sim MN(\hat{\mu}_{z_d}(\mathbf{x}), \sigma_0^2((n\lambda)^{-1}K^*_{z_d}(\mathbf{x},\mathbf{x})+1)),$$

where $\hat{\mu}_{z_d}(\mathbf{x}) = f^M(\mathbf{x},\boldsymbol{\theta}) + \frac{\mathbf{r}^T(\mathbf{x})}{1+\lambda\lambda_z}\left(\mathbf{R}+\frac{n\lambda}{1+\lambda\lambda_z}\mathbf{I}_n\right)^{-1}\left(\mathbf{y}^F - \mathbf{f}^M_{\boldsymbol{\theta}}\right)$, *and*

$K^*_{z_d}(\mathbf{x},\mathbf{x}) = K(\mathbf{x},\mathbf{x}) - \mathbf{r}^T(\mathbf{x})\left[\mathbf{I}_n + \left(\mathbf{R}+\frac{n\lambda}{1+\lambda\lambda_z}\mathbf{I}_n\right)^{-1}\frac{n}{(1+\lambda\lambda_z)\lambda_z}\right]\tilde{\mathbf{R}}^{-1}\mathbf{r}(\mathbf{x})$, *with*

$\mathbf{r}(\mathbf{x}) = (K(\mathbf{x},\mathbf{x}_1),...,K(\mathbf{x},\mathbf{x}_n))^T$ *and* $\tilde{\mathbf{R}} = \mathbf{R} + \frac{n}{\lambda_z}\mathbf{I}_n$ *with the* (i,j) *entry of* $\mathbf{R}$ *being* $K(\mathbf{x}_i,\mathbf{x}_j)$.

For any $\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}$, one may assume $K(\mathbf{x}_a,\mathbf{x}_b) = \prod_{i=1}^p K_i(d_i)$, where $K_i(\cdot)$ is a kernel for $d_i = |x_{ai} - x_{bi}|$. For example, the Matérn kernel with smoothness $\nu_i = 5/2$ is

$$K_i(d_i) = \left(1 + \frac{\sqrt{5}d_i}{\gamma_i} + \frac{5d_i^2}{3\gamma_i^2}\right)\exp\left(-\frac{\sqrt{5}d_i}{\gamma_i}\right). \tag{14}$$

Example 2 (Estimators from coherent generative models are better)

Let $y^F(x) = y^R(x) + \epsilon$, where $\epsilon \sim N(0, 0.05^2)$ independently and $y^R(x) = g_1(x) + g_2(x)$, with $g_1(x) = \sum_{j=1}^{100} j^{-1} \cos(5\pi(j-0.5)x)\sin(5j)$ and $g_2(x) = \sum_{j=1}^{100} j^{-6} \cos(5\pi(j-0.5)x)\sin(5j)$. Let the mathematical model be $f^M(x) = g_1\theta$. The goal is to predict $y^R(\cdot)$ and estimate θ.

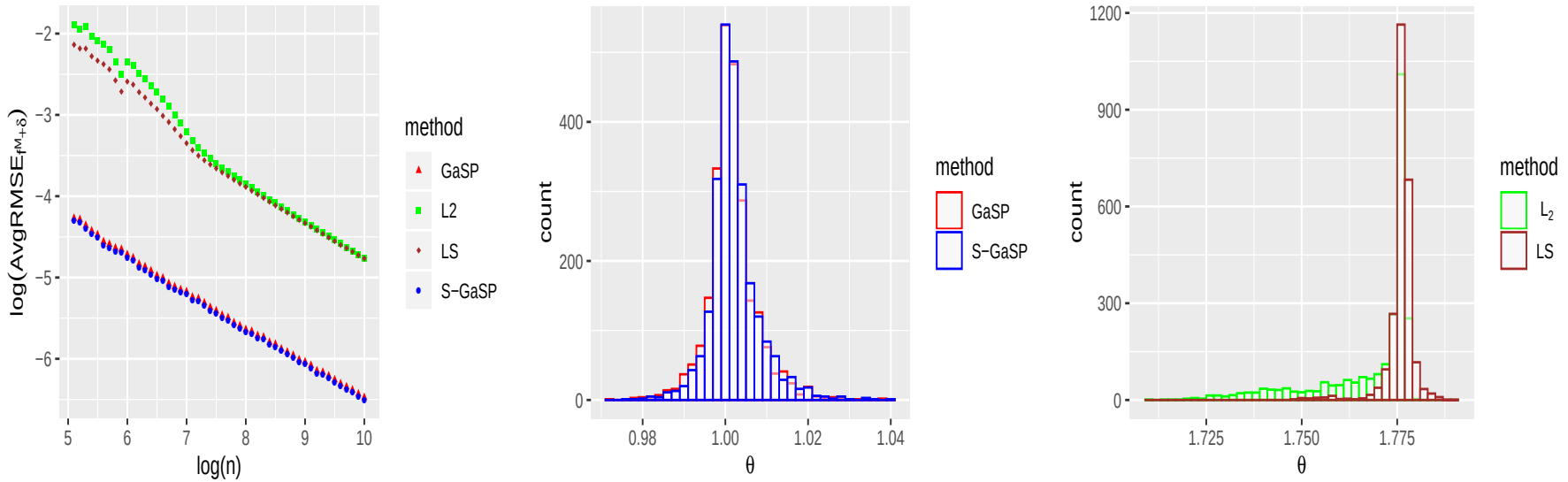


Figure 13: Comparison of different approaches in Example 2. In the left panel, the logarithm of the AvgRMSE$_{f^M+\delta}$ of four calibration approaches are graphed at the logarithm of different sample sizes. The histogram of the estimated calibration parameter of each experiment of different approaches are given in the middle panel and the right panel.

Outline

Interferograms processing

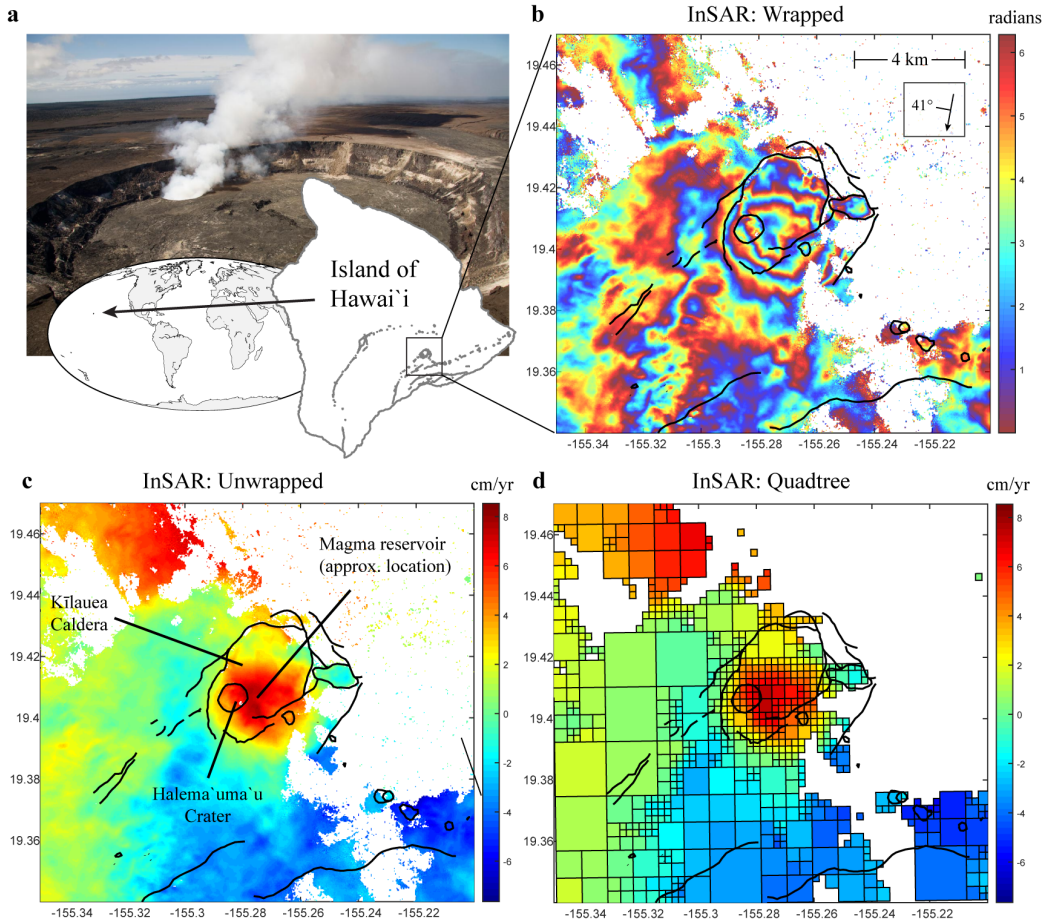


Figure 14: (b) Wrapped InSAR interferogram from the COSMO-SkyMed satellite, spanning 20 Oct 2011 to 15 May 2012. The inset box shows the flight path of the satellite (arrow) and the downward look direction of the satellite at 41°. White areas have no data due to radar decorrelation. Number of data points is around 1.5×10^5. (c) Same data as in (b), but unwrapped. (d) Quadtree-processed interferogram.

A model with correlated atmospheric error

The calibration model by multiple sources of data with measurement bias (correlated atmospheric error) can be written as

$$\mathbf{y}_l^F(\mathbf{x}) = f^M(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \delta_l(\mathbf{x}) + \mu_l + \epsilon_l, \tag{15}$$

for each source l, $l = 1, ..., k$, and any $\mathbf{x} \in \mathcal{X}$, where μ_l is an unknown mean parameter and $\delta_l(\cdot)$ models the measurement bias in the source l, $l = 1, ..., k$, independent of the discrepancy function.

We model $\delta_{z_d}(\cdot)$ and $\delta_l(\cdot)$ by the discretized S-GaSP and GaSP respectively. i.e.

$$\boldsymbol{\delta}_{z_d} \sim \mathrm{MN}(\mathbf{0}, \sigma^2 \mathbf{R}_{z_d}), \qquad \boldsymbol{\delta}_l \sim \mathrm{MN}(\mathbf{0}, \sigma_l^2 \mathbf{R}_l),$$

for $l = 1, ..., k$, where $\mathbf{R}_{z_d} = (\mathbf{R}^{-1} + \frac{\lambda_z}{n}\mathbf{I}_n)^{-1}$. The connection between the individual level data and aggregated data (stacked image) is also studied.

Connection between individual data and aggregated data

- When this is no measurement bias,

$$y_l^F(\mathbf{x}) = f^M(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \mu + \epsilon_l(\mathbf{x}), \tag{16}$$

  where the independent noise follows $\epsilon_l(\mathbf{x}) \sim N(0, \sigma_0^2)$ for each $\mathbf{x}$, and $l = 1, ..., k$.

- Directly averaging the data gives

$$\bar{y}^F(\mathbf{x}) = f^M(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \mu + \bar{\epsilon}(\mathbf{x}), \tag{17}$$

  where the noise independently follows $\bar{\epsilon}(\mathbf{x}) \sim N(0, \sigma_0^2/k)$ for each $\mathbf{x}$.

- Integrating out $\boldsymbol{\delta} \sim \text{MN}(\mathbf{0}, \tau^2\mathbf{R})$, the log marginal likelihood in model (16) follows

$$\ell(\boldsymbol{\theta}, \mu, \sigma_0^2, \tau, \mathbf{R}) = c_{\sigma_0^2} + \bar{\ell}(\boldsymbol{\theta}, \mu, \sigma_0^2, \tau, \mathbf{R}), \tag{18}$$

  where $\bar{\ell}(\boldsymbol{\theta}, \mu, \sigma_0^2, \tau, \mathbf{R})$ is log marginal likelihood from (17) [Gu et al., 2023a].

Remark: When σ_0^2 is known, two models are equivalent; when σ_0^2 is unknown, sufficient statistics are $\bar{\mathbf{y}}^F$ and s^2, where $s^2 = \sum_{l=1}^k \sum_{i=1}^n (y_l^F(\mathbf{x}_i) - \bar{y}^F(\mathbf{x}_i))^2$. The sample variance of σ_0^2 by aggregate data is $k\sum_{i=1}^n (\bar{y}^F(\mathbf{x}_i) - y^R(\mathbf{x}_i))^2/(n-1)$, which has a variance of $2n\sigma_0^4/(n-1)^2$, which is larger than the variance of the sample variance $s^2/(n(k-1))$ based on full data: $2\sigma_0^4/(n(k-1))$. When there is measurement bias, typically aggregated data (stacked image) is less efficient.

The RobustCalibration package

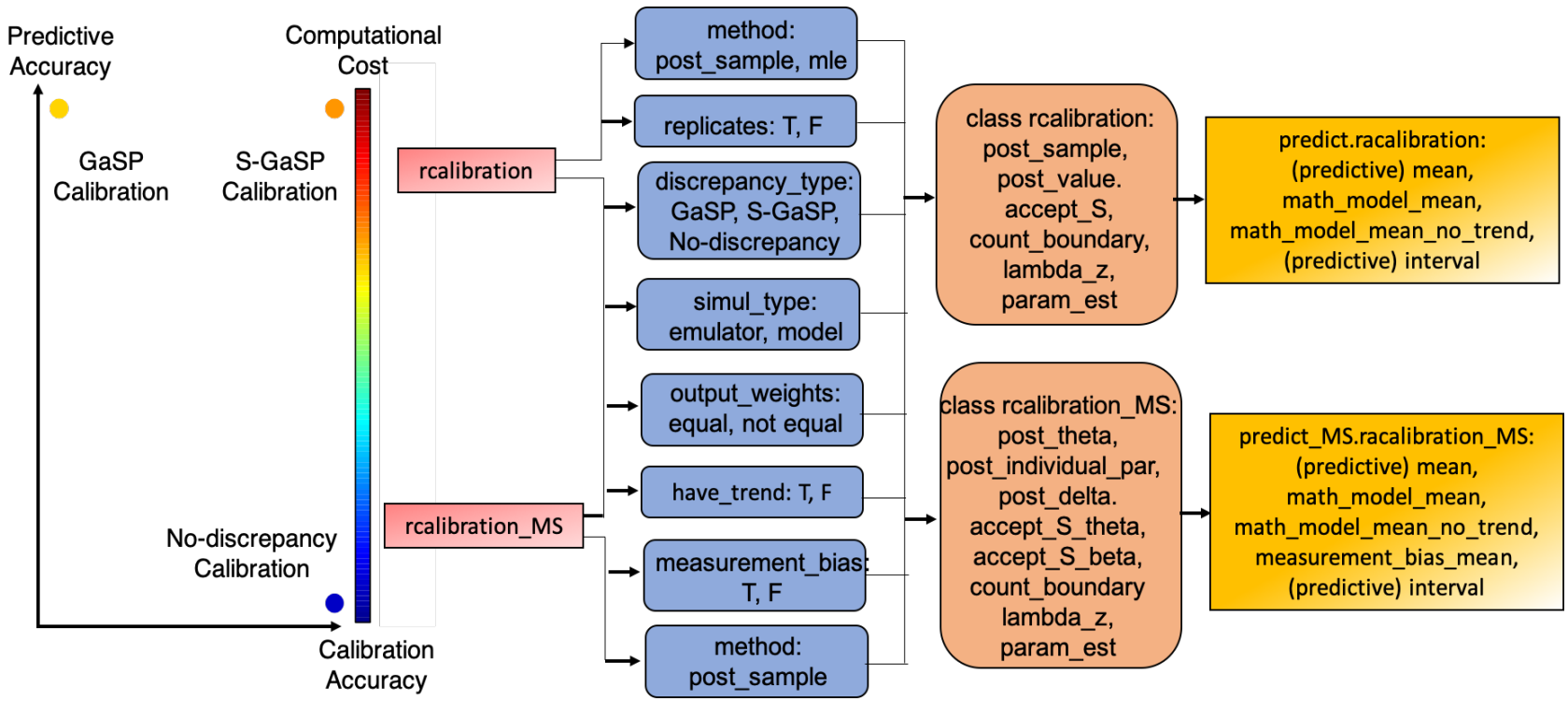


Figure 15: Schematic overview of the **RobustCalibration** package available on CRAN [Gu, 2022]. All three type of calibration models in the left panel is implemented. The right panel give an overview of the major functions and classes. The scalar-valued and vector-valued emulator from the **RobustGaSP** package available on CRAN [Gu et al., 2019] can be called for approximating expensive computer models. Finally, the object classes can be used for predicting the reality.

Simulated study

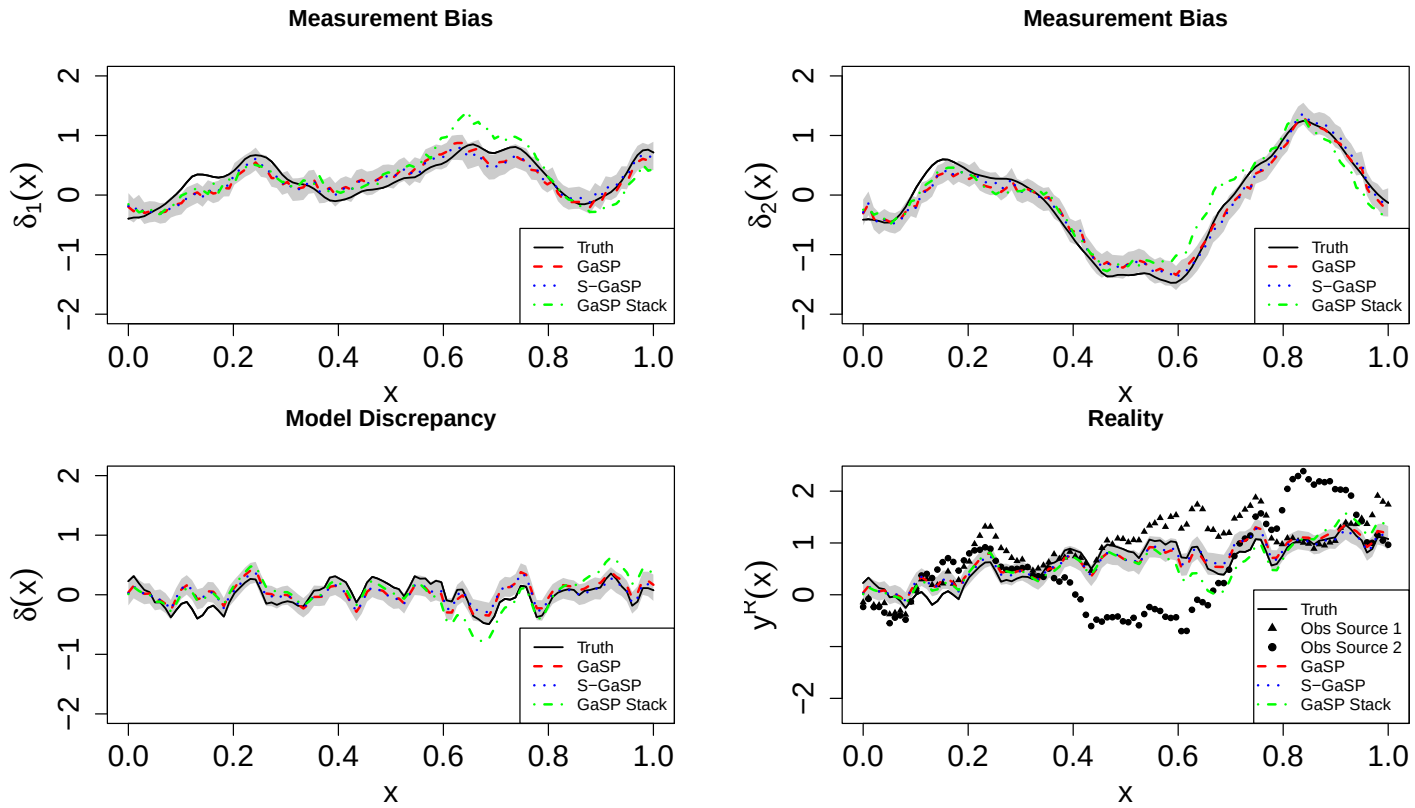


Figure 16: Simulated study: $y_l^F(\mathbf{x}) = f^M(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \delta_l(\mathbf{x}) + \epsilon_l$, where $f^M(x, \theta) = \sin(\theta x)$ with the true $\theta = \pi/2$. The model discrepancy and measurement bias are both assumed to follow GaSPs with Matérn kernel with roughness parameter 5/2. The 95% posterior credible intervals from the S-GaSP calibration is graphed as the shaded area. The true $\delta(\cdot)$ is from a GaSP but estimation from S-GaSP is also reasonably well. Stacking reduces information.

Satellite interferograms

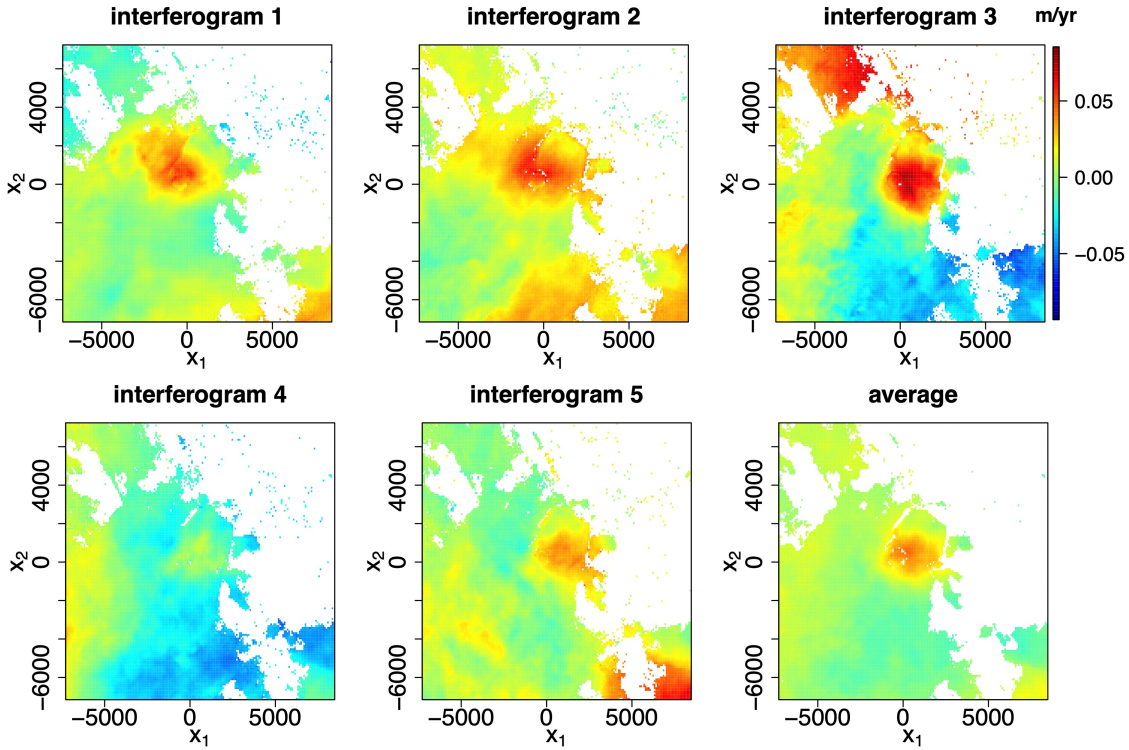


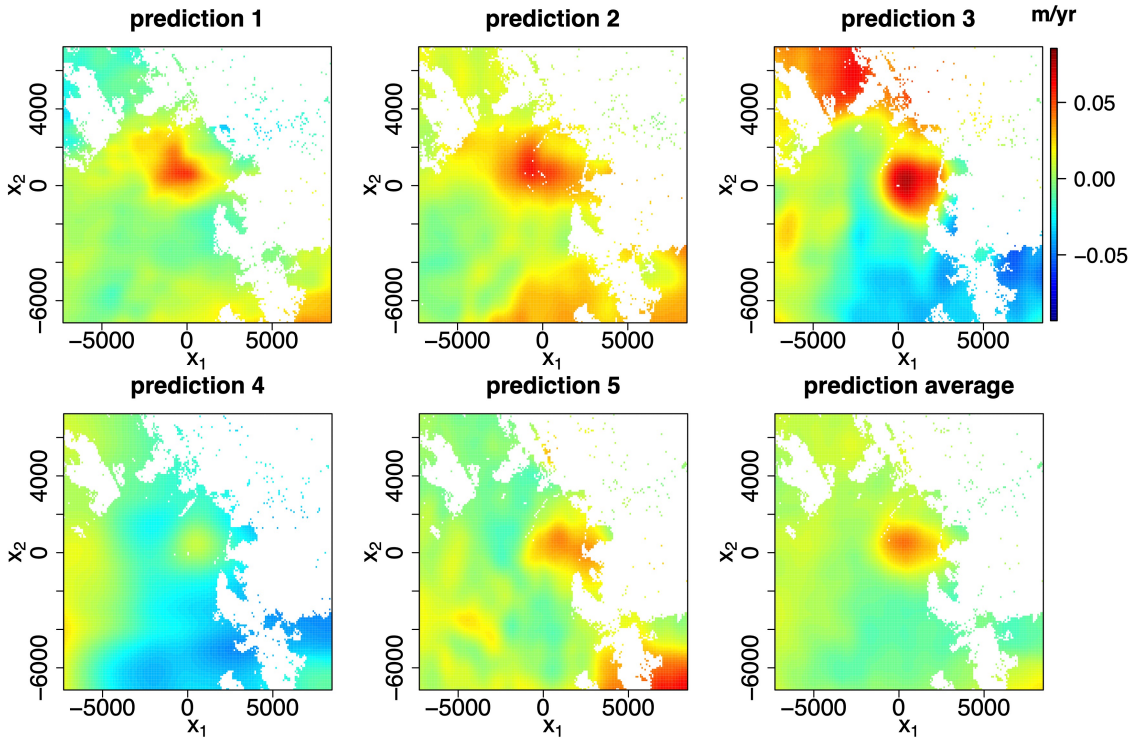
Figure 17: Five COSMO-SkyMed satellite interferograms spanning from: 1) 17 Oct 2011 - 04 May 2012; 2) 21 Oct 2011 - 16 May 2012; 3); 20 Oct 2011 to 15 May 2012; 4) 28 Oct 2011 to 11 May 2012; 5) 12 Oct 2011 - 07 May 2012. Interferograms 1 and 2 have an ascending-mode look angle, while the rest are descending-mode. Horizontal position is in meters relative to a chosen point in Kīlauea Caldera; vertical scale is m/yr. The last figure shows the stack (average) of 6 images.

Input variables and calibration parameters

Table 3: Input variables, calibration parameters of the geophysical model in calibration.

Input variables ($\mathbf{x}$)	Description
x_1	East-west spatial coordinate
x_2	South-north spatial coordinate
Calibration parameters ($\boldsymbol{\theta}$)	**Description**
$\theta_1 \in [-2000, 3000]$	Spatial coordinate of east-west chamber (m)
$\theta_2 \in [-2000, 5000]$	Spatial coordinate of north-south chamber (m)
$\theta_3 \in [500, 6000]$	Depth of the chamber (m)
$\theta_4 \in [0, 0.15]$	Volume change rate of the reservoir (m^3/s)
$\theta_5 \in [0.25, 0.33]$	Poisson's ratio (host rock property)

Predictions by S-GaSP



MSE_{fM}	image 1	image 2	image 3	image 4	image 5
GaSP	1.26	1.63	7.80	4.33	1.97
S-GaSP	**1.21**	**1.45**	**7.66**	**4.05**	**1.76**
$\text{MSE}_{fM+\delta+\delta_l}$	image 1	image 2	image 3	image 4	image 5
GaSP	0.116	0.115	**0.264**	0.134	**0.120**
S-GaSP	**0.109**	**0.112**	0.267	**0.131**	0.123

MSE in predicting each interferogram by 400 samples. The number is by 10^{-4}.

Posterior samples of the parameters

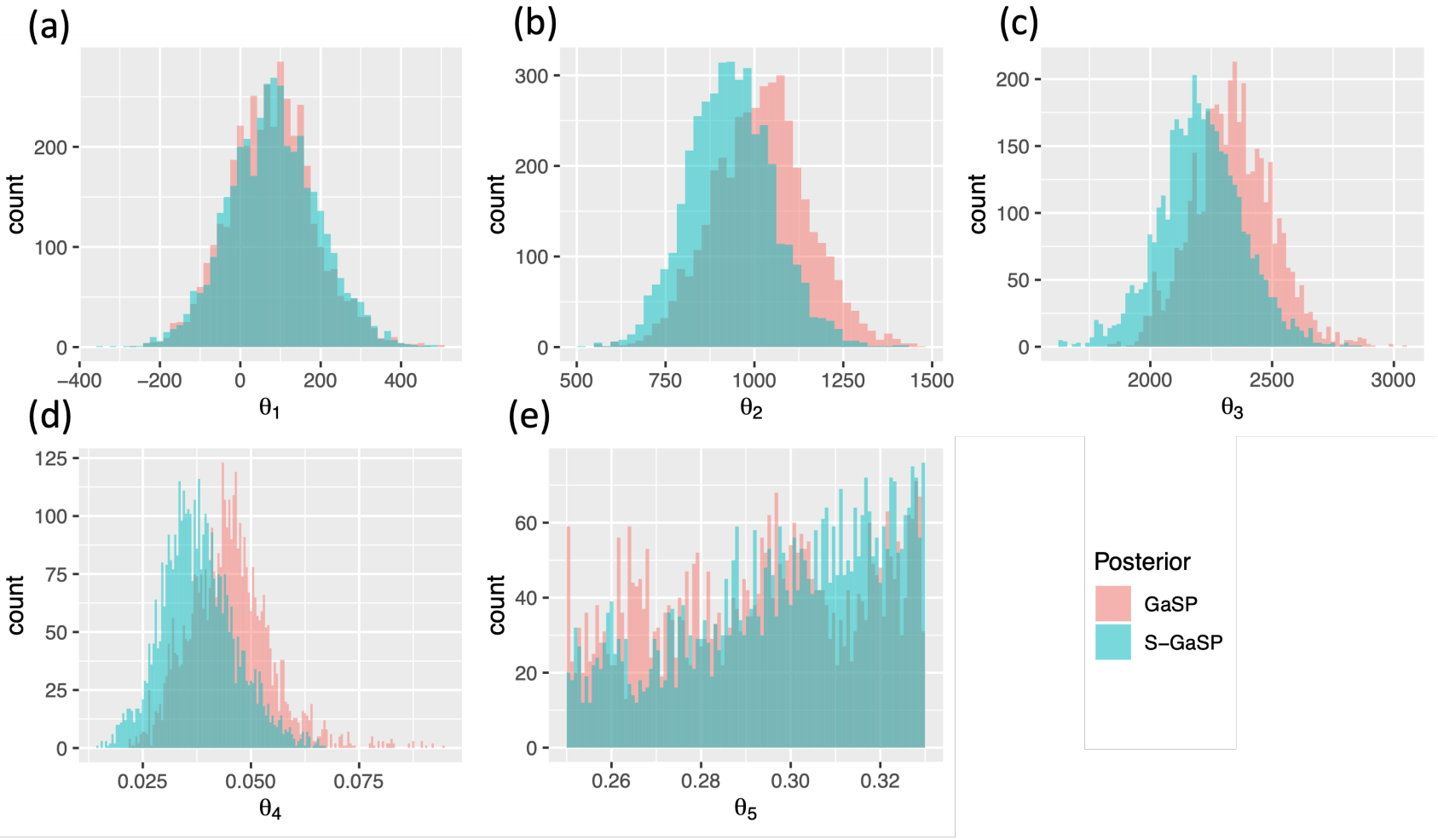
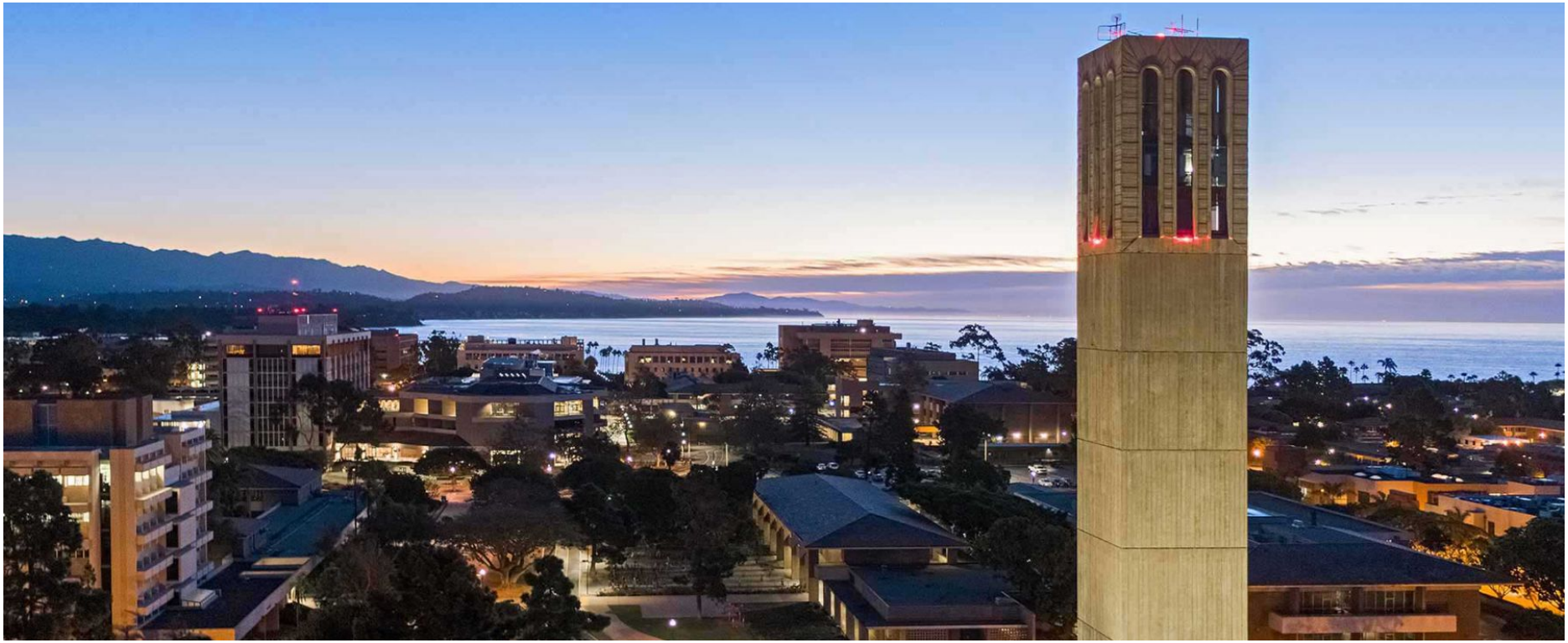


Figure 18: The posterior samples of $\boldsymbol{\theta}$ in the GaSP and S-GaSP calibrations. The range of the parameter θ_5 (Poisson's ratio) is consistent with many rock types, but the geophysical model is relatively insensitive to this parameter [Mogi, 1958].

Summary and Acknowledgement

- We discussed a two-step principle to improve parameter estimation from loss-minimization:
  - Build a (large) probabilistic generative model of **untransformed data**.
  - Show **equivalence** to a statistical estimator of the generative model, and derive a more efficient estimator by **marginalizing out** random components.
- Latent factor models are generative models of many physical approaches of images and functions. Physical models can be encoded in correlation.
- Fast and exact approaches, such as fast Fourier transform, generalized Schur algorithm, and Kalman filter can be used for computing the likelihood at pseudo-linear complexity without approximation.
- We thank all graduate students and co-authors: Xinyi Fang, Yue He, Yizi Lin, Xubo Liu, Erika McPhillips, Matt Helgeson, Megan Valentine (UCSB), Victor Lee, Yimin Luo, Diana Qiu (Yale) and Weining Shen (UCI).
- We acknowledge the support from NSF CDS&E program DMS-2053423 and NSF BioPACIFIC Material Innovation Platform (MIP) DMR-1933487.

Welcome to apply to the PhD program in Statistics and Applied Probability at UCSB



Visit the website:

`https://www.pstat.ucsb.edu/graduate/prospective`

Thanks!

Gregory S Ammar and William B Gragg. The generalized schur algorithm for the superfast solution of toeplitz systems. In Rational Approximation and its Applications in Mathematics and Physics: Proceedings, Łańcut 1985, pages 315–330. Springer, 1987.

Gregory S Ammar and William B Gragg. Superfast solution of real positive definite toeplitz systems. SIAM Journal on Matrix Analysis and Applications, 9(1):61–76, 1988.

Paul D Arendt, Daniel W Apley, and Wei Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. Journal of Mechanical Design, 134(10):100908, 2012.

Maria J Bayarri, James O Berger, Rui Paulo, Jerry Sacks, John A Cafeo, James Cavendish, Chin-Hsu Lin, and Jian Tu. A framework for validation of computer models. Technometrics, 49(2):138–154, 2007.

James O Berger and Luis R Pericchi. The intrinsic Bayes factor for model selection and prediction. Journal of the American Statistical Association, 91(433):109–122, 1996.

James O Berger, Victor De Oliveira, and Bruno Sansó. Objective Bayesian

analysis of spatially correlated data. Journal of the American Statistical Association, 96(456):1361–1374, 2001.

Bruce J Berne and Robert Pecora. Dynamic light scattering: with applications to chemistry, biology, and physics. Courier Corporation, 2000.

Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern Koopman theory for dynamical systems. SIAM Review, 64(2): 229–340, 2022. doi: 10.1137/21M1401243.

Roberto Cerbino and Veronique Trappe. Differential dynamic microscopy: probing wave vector dependent dynamics with a microscope. Physical Review Letters, 100(18):188102, 2008.

John C Crocker and David G Grier. Methods of digital video microscopy for colloidal studies. J. Colloid Interface Sci., 179(1):298–310, 1996.

Bruno De Finetti. La prévision: ses lois logiques, ses sources subjectives. In Annales de l'institut Henri Poincaré, volume 7, pages 1–68, 1937.

Tamsin L. Edwards, Sophie Nowicki, Ben Marzeion, Regine Hock, Heiko Goelzer, Hlne Seroussi, Nicolas C. Jourdain, Donald A. Slater, Fiona E.

Turner, Christopher J. Smith, Christine M. McKenna, Erika Simon, Ayako Abe-Ouchi, Jonathan M. Gregory, Eric Larour, William H. Lipscomb, Antony J. Payne, Andrew Shepherd, Ccile Agosta, Patrick Alexander, Torsten Albrecht, Brian Anderson, Xylar Asay-Davis, Andy Aschwanden, Alice Barthel, Andrew Bliss, Reinhard Calov, Christopher Chambers, Nicolas Champollion, Youngmin Choi, Richard Cullather, Joshua Cuzzone, Christophe Dumas, Denis Felikson, Xavier Fettweis, Koji Fujita, Benjamin K. Galton-Fenzi, Rupert Gladstone, Nicholas R. Golledge, Ralf Greve, Tore Hattermann, Matthew J. Hoffman, Angelika Humbert, Matthias Huss, Philippe Huybrechts, Walter Immerzeel, Thomas Kleiner, Philip Kraaijenbrink, Sbastien Le clech, Victoria Lee, Gunter R. Leguy, Christopher M. Little, Daniel P. Lowry, Jan-Hendrik Malles, Daniel F. Martin, Fabien Maussion, Mathieu Morlighem, James F. ONeill, Isabel Nias, Frank Pattyn, Tyler Pelle, Stephen F. Price, Aurlien Quiquet, Valentina Radi, Ronja Reese, David R. Rounce, Martin Rckamp, Akiko Sakai, Courtney Shafer, Nicole-Jeanne Schlegel, Sarah Shannon, Robin S. Smith, Fiammetta Straneo, Sainan Sun, Lev Tarasov, Luke D. Trusel, Jonas Van Breedam, Roderik van de Wal,

Michiel van den Broeke, Ricarda Winkelmann, Harry Zekollari, Chen Zhao, Tong Zhang, and Thomas Zwinger. Projected land ice contributions to twenty-first-century sea level rise. Nature, 593(7857): 74–82, 2021.

Yongxiang Gao, Juntae Kim, and Matthew E Helgeson. Microdynamics and arrest of coarsening during spinodal decomposition in thermoreversible colloidal gels. Soft Matter, 11(32):6360–6370, 2015.

Fabio Giavazzi, Doriano Brogioli, Veronique Trappe, Tommaso Bellini, and Roberto Cerbino. Scattering information obtained by optical microscopy: differential dynamic microscopy and beyond. Physical Review E, 80(3):031403, 2009.

Mengyang Gu. RobustCalibration: Robust calibration of computer models in R. arXiv preprint arXiv:2201.01476, 2022.

Mengyang Gu and Weining Shen. Generalized probabilistic principal component analysis of correlated data. Journal of Machine Learning Research, 21(13), 2020.

Mengyang Gu and Long Wang. Scaled Gaussian stochastic process for

computer model calibration and prediction. SIAM/ASA Journal on Uncertainty Quantification, 6(4):1555–1583, 2018.

Mengyang Gu, Jesus Palomo, and James O. Berger. RobustGaSP: Robust Gaussian Stochastic Process Emulation in R. The R Journal, 11(1): 112–136, 2019. doi: 10.32614/RJ-2019-011.

Mengyang Gu, Yimin Luo, Yue He, Matthew E Helgeson, and Megan T Valentine. Uncertainty quantification and estimation in differential dynamic microscopy. Physical Review E, 104(3):034610, 2021.

Mengyang Gu, Xubo Liu, Xinyi Fang, and Sui Tang. Scalable marginalization of correlated latent variables with applications to learning particle interaction kernels. The New England Journal of Statistics in Data Science, In Press, 2022a.

Mengyang Gu, Fangzheng Xie, and Long Wang. A theoretical framework of the scaled Gaussian stochastic process in prediction and calibration. SIAM/ASA Journal on Uncertainty Quantification, 10(4):1435–1460, 2022b. doi: 10.1137/21M1409949.

Mengyang Gu, Kyle Anderson, and Erika McPhillips. Calibration of imperfect geophysical models by multiple satellite interferograms with

measurement bias. Technometrics, 0(0):1–12, 2023a. doi: 10.1080/00401706.2023.2182365.

Mengyang Gu, Xinyi Fang, and Yimin Luo. Data-driven model construction for anisotropic dynamics of active matter. PRX Life, 1(1): 013009, 2023b.

Mengyang Gu, Yizi Lin, Victor Chang Lee, and Diana Qiu. Probabilistic forecast of nonlinear dynamical systems with uncertainty quantification. Revised version submitted to Physica D: Nonlinear Phenomena, arXiv preprint arXiv:2305.08942, 2023c.

Dave Higdon, James Gattiker, Brian Williams, and Maria Rightley. Computer model calibration using high-dimensional output. Journal of the American Statistical Association, 103(482):570–583, 2008.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1):35–45, 1960.

Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3):425–464, 2001.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems, 30, 2017.

Gloria Lee, Gregor Leech, Michael J Rust, Moumita Das, Ryan J McGorty, Jennifer L Ross, and Rae M Robertson-Anderson. Myosin-driven actin-microtubule networks exhibit self-organized contractile dynamics. Sci. Adv., 7(6):eabe4334, 2021.

Yun Ling. Superfast Inference for Stationary Gaussian Processes in Particle Tracking Microrheology. PhD thesis, University of Waterloo, 2019.

Fei Liu, MJ Bayarri, and JO Berger. Modularization in Bayesian analysis, with emphasis on analysis of computer models. Bayesian Analysis, 4(1): 119–150, 2009.

Peter J Lu, Fabio Giavazzi, Thomas E Angelini, Emanuela Zaccarelli, Frank Jargstorff, Andrew B Schofield, James N Wilking, Mark B Romanowsky, David A Weitz, and Roberto Cerbino. Characterizing concentrated, multiply scattering, and actively driven fluorescent systems with confocal differential dynamic microscopy. Phys. Rev. Lett., 108(21):218103, 2012.

Kanti V Mardia and Roger J Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71(1): 135–146, 1984.

Thomas G Mason. Estimating the viscoelastic moduli of complex fluids using the generalized Stokes–Einstein equation. Rheol. Acta, 39(4): 371–378, 2000.

K Mogi. Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them. Bulletin of the Earthquake Research Institute, University of Tokyo, 36:99–134, 1958.

Reece Nixon-Luke, Jochen Arlt, Wilson CK Poon, Gary Bryant, and Vincent A Martinez. Probing the dynamics of turbid colloidal suspensions using differential dynamic microscopy. Soft Matter, 18(9): 1858–1867, 2022.

Rui Paulo, Gonzalo García-Donato, and Jesús Palomo. Calibration of computer models with multivariate output. Computational Statistics and Data Analysis, 56(12):3959–3974, 2012.

Adrian E Raftery, David Madigan, and Jennifer A Hoeting. Bayesian

model averaging for linear regression models. Journal of the American Statistical Association, 92(437):179–191, 1997.

Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. Journal of fluid mechanics, 641:115–127, 2009.

Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. Journal of fluid mechanics, 656:5–28, 2010.

Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3):611–622, 1999.

Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. Journal of Computational Dynamics, 1(2): 391–421, 2014.

Rui Tuo and CF Jeff Wu. Efficient calibration for imperfect computer models. The Annals of Statistics, 43(6):2331–2352, 2015.

Rui Tuo and CF Jeff Wu. A theoretical framework for calibration in computer models: parametrization, estimation and convergence

properties. SIAM/ASA Journal on Uncertainty Quantification, 4(1): 767–795, 2016.

Andrew G Wilson and Pavel Izmailov. Bayesian deep learning and a probabilistic perspective of generalization. Advances in neural information processing systems, 33:4697–4708, 2020.

Laurence G Wilson, Vincent A Martinez, Jana Schwarz-Linek, J Tailleur, Gary Bryant, PN Pusey, and Wilson CK Poon. Differential dynamic microscopy of bacterial motility. Physical Review Letters, 106(1): 018101, 2011.

Raymond KW Wong, Curtis B Storlie, and Thomas Lee. A frequentist approach to computer model calibration. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79:635–648, 2017.

Hao Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association, 99(465):250–261, 2004.