Maximum likelihood estimation for stochastic differential equations using sequential Gaussian-process-based optimization

Peter F. Craigmile

The Ohio State University department of STATISTICS

http://www.stat.osu.edu/~pfc/

The Joint Statistical Meetings, Vancouver, Canada

July 29, 2018

Joint research with Grant Schneider and Radu Herbei

Supported in part by the US National Science Foundation

(DMS-1209142, DMS-1407604, and SES-1424481)

Stochastic differential equations (SDEs)

$$dX_t = \mu(X_t, \boldsymbol{\theta})\, dt + \sigma(X_t, \boldsymbol{\theta})\, dW_t, \quad 0 \le t \le T,$$

where $X_0 = x_0$ is the initial value of the process and $\{W_t\}$ is a standard Brownian motion (BM).

Assume the **drift function** $\mu(\cdot, \cdot)$ and **diffusion function** $\sigma(\cdot, \cdot)$ are known up to the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is some compact set in $\mathbb{R}^p$.

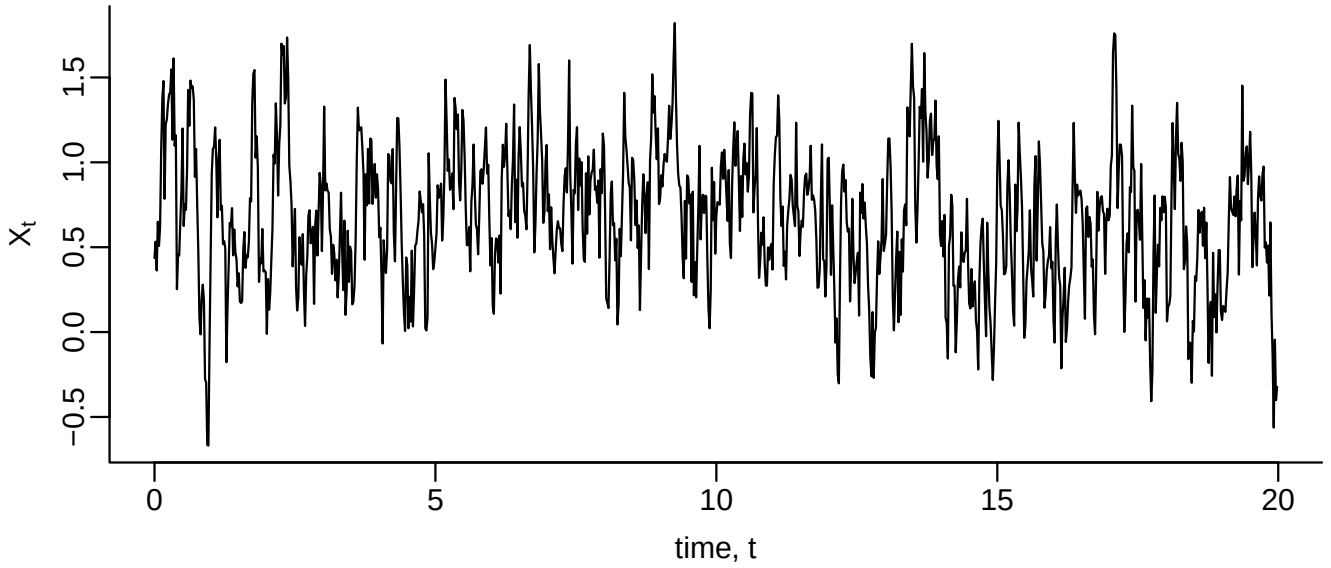
(Also assume locally Lipschitz with linear growth bounds so that a weakly unique solution exists.)

Our inference problem: **maximum likelihood estimation** of $\boldsymbol{\theta}$ when $\{X_t\}$ is observed at N time points $\{t_i : i = 1, \ldots, N\}$.

Example: Ornstein-Uhlenbeck (OU) process

$$dX_t = (\theta_0 + \theta_1 X_t)\, dt + dW_t, \quad 0 \le t \le T,$$

where $X_0 = x_0$ is the initial value, $\theta_0 \in \mathbb{R}, \theta_1 < 0$, and W_t is a std. BM.



The likelihood function

Treating $X_0 = x_0$ as fixed, we use the **Markov property** to write the **likelihood** as the product of individual **transition densities**:

$$L(\boldsymbol{\theta}|\mathbf{X}) = \prod_{i=1}^{N} p(X_{t_i}|X_{t_{i-1}}, \boldsymbol{\theta}).$$

As the transition density **does not exist in closed-form** except for a handful of cases, approximations are typically necessary.

The most commonly used **Euler approximation** is

$$\xi(X_\Delta|X_0, \boldsymbol{\theta}) = n(X_\Delta; X_0 + \mu(X_0, \boldsymbol{\theta})\Delta, \sigma^2(X_0, \boldsymbol{\theta})\Delta).$$

Can do well for small Δ, but there are better approximations ...

Approximating the transition density with importance sampling

1. Partition $[0, \Delta)$ into K subintervals of width Δ/K with endpoints

   $0 = \tau_0 < \tau_1 < \ldots < \tau_K = \Delta.$

2. The **discretized transition density** [Kloeden and Platen, 1992] is

$$p^{(K)}\left(X_\Delta | X_0, \boldsymbol{\theta}\right) = \int \prod_{k=1}^{K} \xi\left(X_{\tau_k} | X_{\tau_{k-1}}, \boldsymbol{\theta}\right) \lambda(d\mathbf{X}_\tau)\,,$$

   where λ denotes the Lebesgue measure.

3. We approximate this density by **importance sampling**, using M random samples from an importance density $q(\cdot)$.

As a compromise between accuracy and computational efficiency of approximating the transition density, we use the **modified Brownian bridge sampler** for $q(\cdot)$ [Durham and Gallant, 2002].

Obtaining the MLE for the SDE process parameters

Limited discussion in the literature about exploring $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Fine-scale grid based methods are very computationally intensive and not efficient.

Gradient based methods suffer from Monte Carlo variability, and are also computationally intensive (require prohibitively large sample sizes, M).

Using methodology from computer experiments

We believe the underlying discretized log-likelihood function

$$l^{(K)}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log p^{(K)}\left(X_{t_i}; X_{t_{i-1}}, \boldsymbol{\theta}\right)$$

is smooth in $\boldsymbol{\theta}$, but our estimates are:

1. Subject to Monte Carlo variability.

2. Expensive to make – $O(KMN)$.

The statistical methodology for **computer experiments** deals with estimation and prediction of expensive-to-evaluate functions (here, measured under uncertainty). A good fit for what we want to do!

A Gaussian process (GP) model

Start with estimates at n **initial parameter values** $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)^T$ chosen based on some **space-filling design**.

Letting $Y(\boldsymbol{\theta}_i)$ denote the estimates of $l^{(K)}(\boldsymbol{\theta}_i)$, assume

$$Y(\boldsymbol{\theta}_i) = l^{(K)}(\boldsymbol{\theta}_i) + \epsilon(\boldsymbol{\theta}_i), \quad i = 1, \ldots, n,$$

where $\{\epsilon(\boldsymbol{\theta}_i) : i = 1, \ldots, n\}$ is a set of independent of N(0, σ^2) errors.

Model $\{l^{(K)}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ using a GP with mean function $\mu_L(\boldsymbol{\theta}; \beta)$ and some valid covariance function $C_L(\boldsymbol{\theta}, \boldsymbol{\theta}'; \zeta)$, where β, ζ are unknown parameters.

Predicting the log-likelihood function

By Gaussianity of the GP and data model, the distribution of $l^{(K)}(\boldsymbol{\theta}^*)$ given the data $\boldsymbol{Y}_n = (Y(\boldsymbol{\theta}_i) : i = 1, \ldots, n)^T$ is normal with a conditional mean of

$$\eta_{L,n}(\boldsymbol{\theta}^*) \;=\; \mu_{L,n}(\boldsymbol{\theta}^*) + \boldsymbol{c}_{L,n}^T(\boldsymbol{\Sigma}_{L,n} + \sigma^2 \boldsymbol{I}_n)^{-1}[\boldsymbol{Y}_n - \boldsymbol{\mu}_{L,n}],$$

and conditional variance given by

$$v_{L,n}^2(\boldsymbol{\theta}^*) \;=\; C_{L,n}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) - \boldsymbol{c}_{L,n}^T(\boldsymbol{\Sigma}_{L,n} + \sigma^2 \boldsymbol{I}_n)^{-1}\boldsymbol{c}_{L,n}.$$

$\boldsymbol{\mu}_{L,n}$ is a mean vector of length n with ith element $\mu_{L,n}(\boldsymbol{\theta}_i; \boldsymbol{\beta})$,

$\boldsymbol{c}_{L,n}$ is a covariance vector of length n with ith element $C_{L,n}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_i; \boldsymbol{\zeta})$, and

$\boldsymbol{\Sigma}_{L,n}$ is the $n \times n$ covariance matrix with (i, j) element $C_{L,n}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j; \boldsymbol{\zeta})$.

Maximizing the likelihood using expected improvement

Let $\widetilde{\eta}_{L,n} = \max_{i=1,\ldots,n} \eta_L(\boldsymbol{\theta}_i)$ denote the maximum of the conditional mean over the explored n values of $\boldsymbol{\theta}$.

The **expected improvement** at parameter value $\boldsymbol{\theta}^*$ is [Jones et al., 1998]

$$\left[\eta_{L,n}\left(\boldsymbol{\theta}^*\right) - \widetilde{\eta}_{L,n}\right] \Phi\left(\frac{\eta_{L,n}\left(\boldsymbol{\theta}^*\right) - \widetilde{\eta}_{L,n}}{v_{L,n}\left(\boldsymbol{\theta}^*\right)}\right) + v_{L,n}(\boldsymbol{\theta}^*)\ \phi\left(\frac{\eta_{L,n}\left(\boldsymbol{\theta}^*\right) - \widetilde{\eta}_{L,n}}{v_{L,n}\left(\boldsymbol{\theta}^*\right)}\right),$$

where $\Phi(\cdot)/\phi(\cdot)$ is the standard Gaussian cdf/pdf.

The expected improvement **balances** the need to

maximize the discretized log-likelihood (the first term)

while cognizant of

the uncertainty in estimating the log-likelihood (the second term).

Sequential Gaussian-Process-Based Optimization (SGBO)

We add the parameter value $\boldsymbol{\theta}^*$ that maximizes the expected improvement.

- Estimate the discretized log-likelihood at that value, yielding $Y(\boldsymbol{\theta}^*)$.

From the new data vector and vector of estimated log-likelihoods, we update:

1. GP parameter estimates (using an empirical Bayes technique);

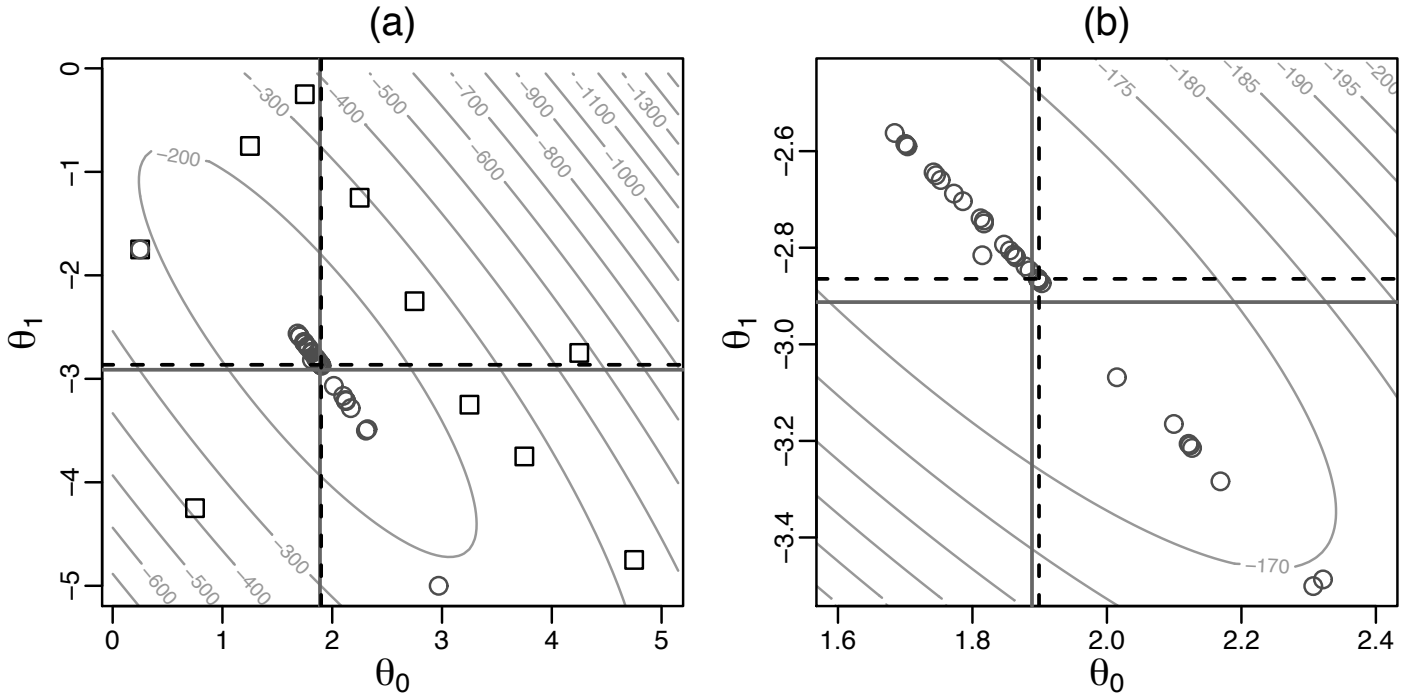
2. The conditional mean and variance for the BLUP.

Repeat until some stopping criteria is met (little change in estimated MLEs).

After n steps, straightforward to obtain the estimated MLE,

$$\widehat{\boldsymbol{\theta}} = \arg \max_{i=1,\ldots,n} \eta_{L,n}(\boldsymbol{\theta}_i).$$

An example SGBO path



(a) A contour plot of the discretized log-likelihood for an OU process with $\theta_0 = 2$ and $\theta_1 = -3$ ($\theta_2 = 1$ is fixed). The solid horizontal and vertical lines denote the exact MLEs of θ_0 and θ_1 respectively, and the dashed lines denote the SGBO-based estimate. For the SGBO method, the squares indicate the initial parameter values, and the circles denote the values added sequentially. (b) is a zoomed in version of (a).

Remarks on the SGBO method

- We can obtain approximate $(1-\alpha)\%$ joint confidence regions for $\boldsymbol{\theta}$ directly from the conditional mean using a **likelihood ratio test**.
- We carry out **simulations** to estimate $\boldsymbol{\theta}$ for OU and Generalized Cox-Ingersoll-Ross (GCIR, Chan et al. [1992]) processes:
  - For OU processes there is no appreciable difference between the SGBO-based estimator and the exact MLE.
  - SGBO clearly outperforms grid methods.
  - SGBO with $K = 10$ outperforms $K = 5$.
  - Methods improve with M with a loss in computational efficiency; it helps to start with more initial points.

Estimating Deep Flow in the South Atlantic Ocean

Estimating the **state** of the world's oceans is a fundamental problem.

Ocean circulation cannot be measured directly.

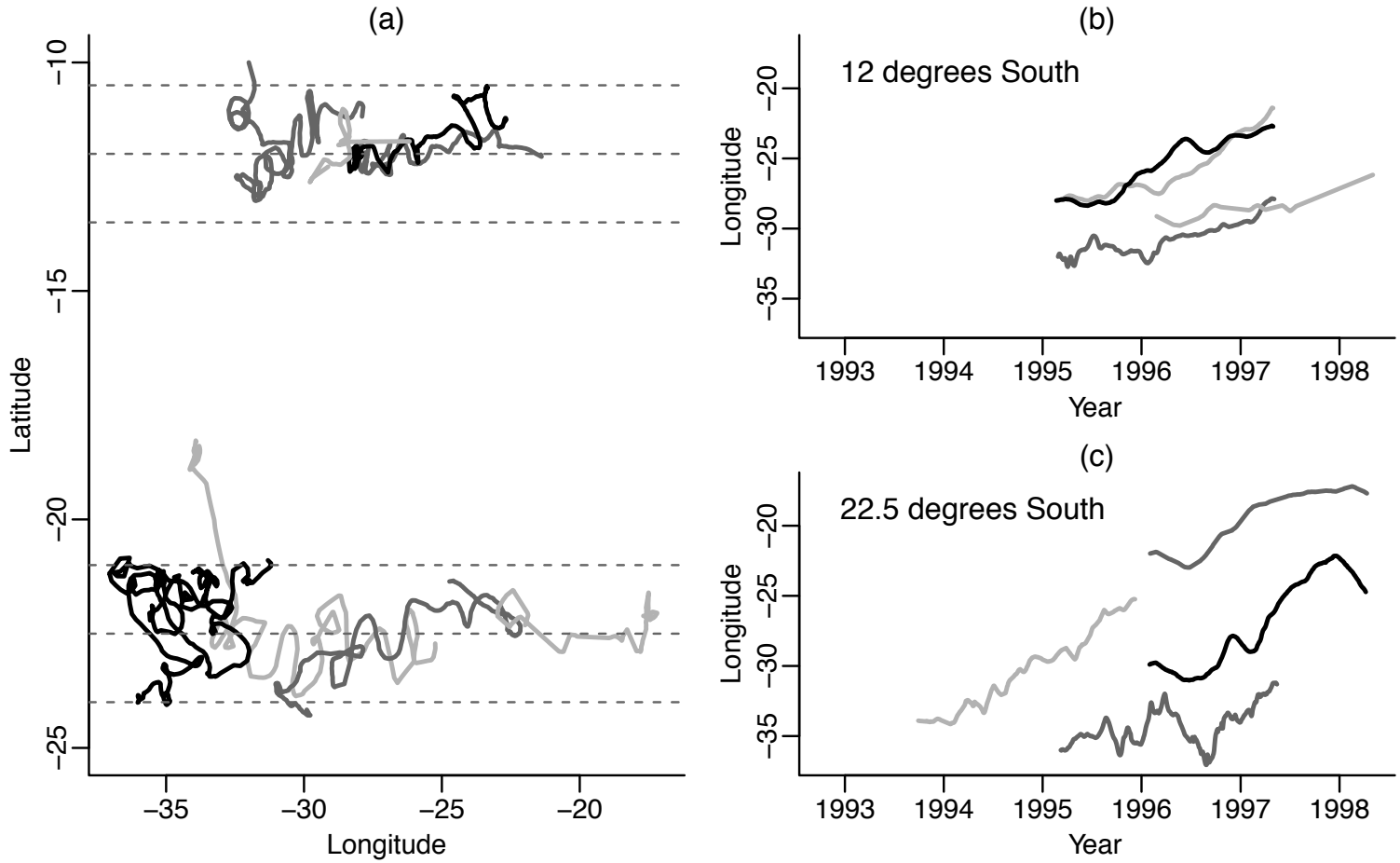
It is inferred from other physical and chemical properties of the ocean, such as measurements of water temperature, salinity, silica, etc.

This is an **inverse problem** [Wunsch, 1996].

We focus on estimating **water velocities**.

Using **float data** [Hogg and Owens, 1999], we estimate the **deep flow** ($2500m$) in two latitude bands (12^oS and $22.5\ ^oS$) in the western South Atlantic Ocean.

Estimating Deep Flow in the South Atlantic Ocean, cont.



In this area the circulation structure is dominated by strong alternating zonal jets [Hogg and Owens, 1999, McKeague et al., 2005].

Estimating Deep Flow in the South Atlantic Ocean, cont.

Let $\{X_t^{(i)} : t \in [0,T]\}$ denote the underlying (continuous) longitude process for float i in a specific latitude band. Data are available every two days.

Assume $\{X_t^{(i)}\}$ satisfies

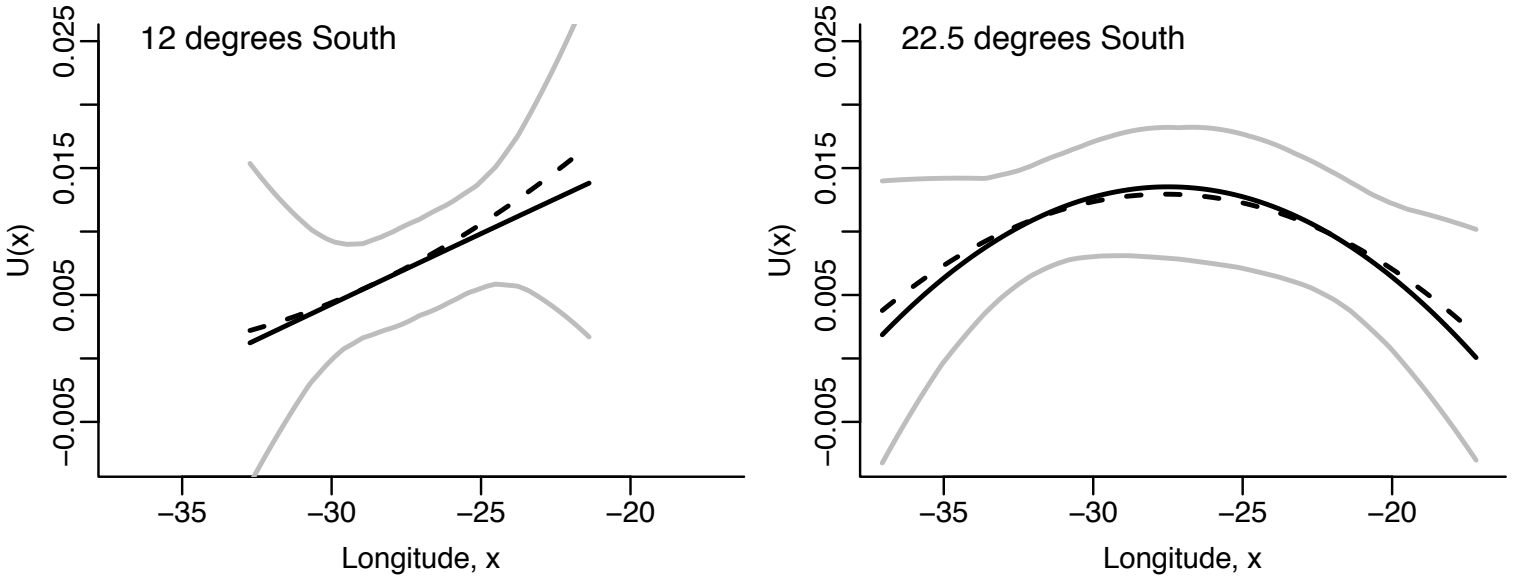
$$dX_t^{(i)} = U(X_t^{(i)})\, dt + \sigma\, dW_t^{(i)} \quad X_0^{(i)} = x_0^{(i)}, \quad t \in [0,T], \tag{1}$$

where $U(\cdot)$ is the **zonal velocity of interest** assumed common for each series in a given latitude band, σ is the diffusion coefficient, and $\{W_t\}$ is a standard BM. We assume conditional independence over i.

Estimating Deep Flow in the South Atlantic Ocean, cont.

We estimated the three parameters of a quadratic model for the common zonal velocity function $U(x)$ using the SGBO method.

$K = 10$, $M = K^2$ and $n = 20 \times 3 = 60$ initial points.



(Dashed line: Euler approximation)

Estimating Deep Flow in the South Atlantic Ocean, cont.

This preliminary analysis provides motivation that the zonal velocity function varies by latitude.

Also demonstrates the utility of picking up more accurate velocity structures using the SGBO method, compared to naive Euler-based approximations.

Further research will investigate methods to estimate the parameters of a spatially varying SDE model that can estimate the **zonal and meridonal velocities jointly**, using more floats deployed in the Atlantic Ocean.

Discussion and further work

Our methods can be extended to observing SDEs with noise.

Can use calibration methods [e.g., Kennedy and O'Hagan, 2001] to minimize the bias introduced from the IS-estimate of the log-likelihood through an adaptive selection of the values of K and M in the SGBO algorithm.

- Another solution: lognormal kriging [e.g. Cressie and Pavlicová, 2005].

Also considering other ways to ameliorate the mild, but present, under-coverage of confidence regions.

References

K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47:1209–1227, 1992.

N. Cressie and M. Pavlicová. Lognormal kriging: bias adjustment and kriging variances. In *Geostatistics Banff 2004*, pages 1027–1036. Springer, New York, NY, 2005.

G. B. Durham and A. R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20:297–316, 2002.

N. G. Hogg and W. B. Owens. Direct measurement of the deep circulation within the Brazil Basin. *Deep Sea Research Part II: Topical Studies in Oceanography*, 46:335–353, 1999.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13: 455–492, 1998.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, 63:425–464, 2001.

P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, New York, NY, 1992.

I. W. McKeague, G. Nicholls, K. Speer, and R. Herbei. Statistical inversion of South Atlantic circulation in an abyssal neutral density layer. *Journal of Marine Research*, 63:683–704, 2005.

C. Wunsch. *The Ocean Circulation Inverse Problem*. Cambridge University Press, Cambridge, England, 1996.