Statistical inference for SPDEs under spatial ergodicity

Randolf Altmeyer

Imperial College London

Based on joint work with S. Gaudlitz

AHIDI2024, Università di Verona, November 8, 2024

IMPERIAL

Outline

1. Motivation
2. LAN for local measurements
3. Inference on nonlinearity for small noise

Outline

Semi-linear SPDEs

Consider the equation

$$\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y)) + \sigma \partial_t W(t,y), \quad (t,y) \in [0,T] \times \Lambda,$$

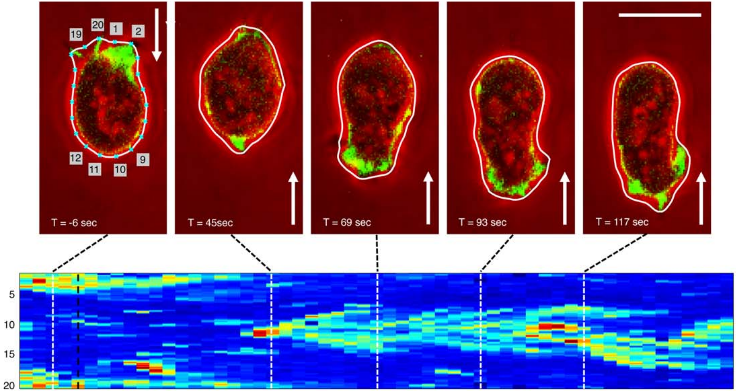
on a smooth bounded domain $\Lambda \subset \mathbb{R}^d$ with $\Delta = \sum_{i=1}^d \partial_{ii}^2$, some initial value $X(0,\cdot)$ and boundary conditions.

- ϑ diffusivity, $\sigma > 0$ noise level, $\partial_t W$ is space-time white noise.
- $f : \mathbb{R}^d \to \mathbb{R}$ Lipschitz continuous.
- $\partial_t W(t,y)$ is dynamic/intrinsic noise, NOT measurement noise.

We want to validate the SPDE model (i.e. ϑ, f) on data.

Case study: Cell repolarisation

A particular aspect of chemotaxis is to understand how a cell changes direction when exposed to an external signal. This process is related to a change in actine concentration along the cell boundary.

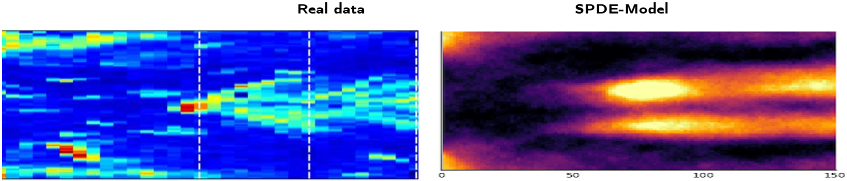


Case study: Cell repolarisation

In order to describe the change in actine concentration, *A., Bretschneider, Janak, Reiß ('22)* introduce a stochastic reaction diffusion model on the circle $\Lambda = [0, 2\pi]$, extending a classical activator-inhibotor model:

$$\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y), Z(t,y), y) + \sigma \partial_t W(t,y)$$
$$\partial_t Z(t,y) = \gamma \Delta Z(t,y) + g(X(t,y), Z(t,y), y).$$

Only X (actine concentration) can be measured.



Dynamic noise speeds up repolarisation and makes front-splitting more likely.

Two statistical models

Recall: $\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y)) + \sigma \partial_t W(t,y)$ on $[0,T] \times \Lambda$.

1. **Inference on ϑ**

- From observations

$$X_\delta = (X_{\delta,k})_{k=1}^M \in L^2([0,T];\mathbb{R}^M), \quad X_{\delta,k}(t) = \langle X(t), K_{\delta,x_k} \rangle_{L^2(\Lambda)},$$

at $x_1,\ldots,x_M \in \mathbb{R}^d$ with $K_{\delta,x_k}(x) = \delta^{-d/2} K(\delta^{-1}(x - x_k))$.

(A., Reiß ('21), Reiß, Strauch, Trottner ('23), ...)

- **Spatial ergodicity/asymptotics:**

$$\langle X(t), K_{\delta,x_k} \rangle_{L^2(\Lambda)} \overset{d}{=} \langle \delta Y(\delta^{-2} t), K \rangle_{L^2(\delta^{-1}(\Lambda - x_k))} + \text{rem},$$

$$\partial_t Y(t,y) = \vartheta \Delta Y(t,y) + \sigma \partial_t W(t,y) \text{ on } [0, \delta^{-2} T] \times \delta^{-1}(\Lambda - x_k).$$

- **A., Tiepner, Wahl ('24):** *as $\delta \to 0$, $M \to \infty$, T fixed, have*

$$M^{1/2} \delta^{-1} (\hat{\vartheta}_\delta - \vartheta) \xrightarrow{d} N(0, \Sigma_\vartheta).$$

Two statistical models

Recall: $\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y)) + \sigma \partial_t W(t,y)$ on $[0,T] \times \Lambda$.

2. Inference on f

- From $(X(t,y))_{0 \leq t \leq T, y \in \Lambda}$ *(Gaudlitz ('23), Gaudlitz, Reiß ('23), Ibragimov ('03)).*
- **Spatial ergodicity/asymptotics:** $\sigma = \vartheta^{1/2}$,

$$X(t,y) \overset{d}{=} Y(t, \vartheta^{-1/2} y),$$

$$\partial_t Y(t,y) = \Delta Y(t,y) + f(Y(t,y)) + \partial_t W(t,y) \text{ on } [0,T] \times \vartheta^{-1/2}(\Lambda - x_k).$$

- Gaudlitz ('23): nonparametric rates for f as $\vartheta \to 0$, $\sigma = \vartheta^{1/2}$, T fixed.

Outline

Local measurements

Let $\sigma = 1$, $f = 0$, $\|K\|_{L^2(\mathbb{R}^d)} = 1$, K compactly supported.

$$\partial_t X(t,y) = \vartheta \Delta X(t,y) + \partial_t W(t,y).$$

$$\Rightarrow \partial_t \langle X(t), K_{\delta,x_k}\rangle = \vartheta \langle X(t), \Delta K_{\delta,x_k}\rangle + \partial_t \langle W(t), K_{\delta,x_k}\rangle.$$

- $\langle X(t), K_{\delta,x_k}\rangle$ and $\langle X(t), K_{\delta,x_j}\rangle$ are correlated (not independent)
- $t \mapsto \langle X(t), K_{\delta,x_k}\rangle$ not Markov processes.

A., Tiepner, Wahl ('24): M local measurements, minimax-optimal rate $h_\delta = \delta^{-1} M^{1/2}$ as $\delta \to 0$, $M \to \infty$.

What is the minimal variance for estimating ϑ from local measurements?

The log-likelihood process

Goal: Show that the log-likelihood process is locally asymptotically normal

$$\log \frac{d\mathbb{P}_{\vartheta + r_\delta}}{d\mathbb{P}_\vartheta}(X_\delta) = Z - \frac{1}{2}\mathcal{I}_\vartheta + o_{\mathbb{P}_\vartheta}(1), \quad \delta \to 0, M \to \infty$$

with $Z \sim N(0, \mathcal{I}_\vartheta)$ and Fisher information $\mathcal{I}_\vartheta > 0$.

Then by the local asymptotic minimax theorem

$$\lim_{\delta \to 0, M \to \infty} \inf_{\hat{\psi}} \sup_{\vartheta_\delta \in [\vartheta, \vartheta + r_\delta]} {r_\delta}^2 \mathbb{E}_{\vartheta_\delta}\left[(\hat{\psi} - \vartheta_\delta)^2\right] \geq \mathcal{I}_\vartheta^{-1}.$$

The likelihood

Notation:

$$X_\delta(t) = (\langle X(t), K_{\delta,x_k}\rangle_{L^2(\mathbb{R}^d)})_{k=1}^M,$$
$$X_\delta^\Delta(t) = (\langle X(t), \Delta K_{\delta,x_k}\rangle_{L^2(\mathbb{R}^d)})_{k=1}^M.$$

Markov projection: $m_\vartheta(t) = \mathbb{E}_\vartheta[\vartheta X_\delta^\Delta(t) | (X_\delta(s))_{0\leq s\leq t}]$.

$$\Rightarrow dX_\delta(t) = m_\vartheta(t)dt + d\bar{w}(t), \quad \bar{w} \ M\text{-dim. BM.}$$

Log-likelihood process: By Girsanov's thm

$$\log\left(\frac{d\mathbb{P}_{\vartheta+r_\delta}}{d\mathbb{P}_\vartheta}(X_\delta)\right) = \log\left(\frac{d\mathbb{P}_{\vartheta+r_\delta}}{d\mathbb{P}_0}(X_\delta)\right) + \log\left(\frac{d\mathbb{P}_0}{d\mathbb{P}_\vartheta}(X_\delta)\right)$$
$$= \sum_{k=1}^M \int_0^T \left(m_{\vartheta+r_\delta,k}(t) - m_{\vartheta,k}(t)\right) d\bar{w}_k - \frac{1}{2}\sum_{k=1}^M \int_0^T \left(m_{\vartheta+r_\delta,k}(t) - m_{\vartheta,k}(t)\right)^2 dt.$$

Two links

Partial observations of OU-process:

- d-dimensional OU process $dX_t = A_\vartheta X_t dt + dW_t$.
- Observe $X_t^T \varphi$ (e.g., projection on first coordinate).
- 'usual' rates of convergence (for ϑ, as $T \to \infty$), loss in information?

Kalman-Bucy method for filtering:

- Hidden OU-process:
$$dX_t = A_\vartheta X_t dt + dW_t,$$
$$dY_t = X_t dt + dB_t.$$
- State estimation X_t from $m(t) = \mathbb{E}[X_t | (Y_s)_{0 \le s \le t}]$.
- ODEs for $m(t)$, $\gamma(t) = \mathbb{E}[(X_t - m(t))^2]$.
- Based on $m(t) = \int_0^t g(t,s) dY_s$ where with $c(t,s) = \mathbb{E}[X_t X_s]$

$$g(t,s) + \int_0^t c(s,s') g(t,s') ds' = c(t,s).$$

Reduction to M independent measurements

Notation:

- Let C_ϑ be covariance operator of $\mathbb{P}_\vartheta$ on $L^2([0,T],\mathbb{R}^M)$, so

$$\mathbb{E}_\vartheta\left[\int_0^T X_\delta(t)\cdot f(t)dt\int_0^T X_\delta(t)\cdot g(t)dt\right]=\int_0^T C_\vartheta f(t)\cdot g(t)dt.$$

- Let $\tilde{C}_\vartheta$ be covariance operator of the law $\tilde{\mathbb{P}}_\vartheta$ of M independent local measurements (so $\tilde{C}_\vartheta$ is diagonal).

By Feldman-Hajek-Thm

$$\mathbb{P}_\vartheta\left(\left|\sqrt{\frac{d\tilde{\mathbb{P}}_\vartheta}{d\mathbb{P}_\vartheta}(X_\delta)}-1\right|>\varepsilon\right)\leq\varepsilon^{-2}\int\left(\sqrt{\frac{d\tilde{\mathbb{P}}_\vartheta}{dQ}}-\sqrt{\frac{d\mathbb{P}_\vartheta}{dQ}}\right)^2dQ$$

$$=\varepsilon^{-2}H^2(\tilde{\mathbb{P}}_\vartheta,\mathbb{P}_\vartheta)\leq 2\varepsilon^{-2}\|\tilde{C}_\vartheta^{-1/2}\left(C_\vartheta-\tilde{C}_\vartheta\right)\tilde{C}_\vartheta^{-1/2}\|_{HS}^2$$

with Hellinger distance H^2.

Reduction to M independent measurements

A., Tiepner, Wahl ('24): for ρ-separated locations $\inf_{k\neq j}|x_k - x_j| \geq \rho$ and $K = \Delta\tilde{K}$ have

$$\|\tilde{C}_\vartheta^{-1/2}\left(C_\vartheta - \tilde{C}_\vartheta\right)\tilde{C}_\vartheta^{-1/2}\|_{HS}^2 \lesssim M\delta^{-2}\frac{\delta^{2+2d}}{\rho^{2+2d}}.$$

$\Rightarrow$ Effect of correlated measurements in likelihood expansion can be ignored if $M\delta^{-2}(\delta/\rho)^{2+2d} \to 0$, e.g. when $M \leq \rho^{-d}$ and $\rho = \delta^{\frac{2d}{2+3d}}\log\delta^{-1}$.

Therefore suppose in the following $M = 1$.

The 'filter'

Proposition

We have the representation

$$m_\vartheta(t) = \mathbb{E}_\vartheta[\vartheta X_\delta^\Delta(t) | (X_\delta(s))_{0\le s\le t}] = \delta^{-1}\int_0^{\delta^{-2}t} g_{\vartheta,\delta}(\delta^{-2}t,s)dY(s)$$

where $\partial_t Y(t,y) = \vartheta\Delta Y(t,y) + \partial_t W(t,y)$ *on* $[0,\delta^{-2}T]\times\delta^{-1}(\Lambda - x_k)$

and where $g_{\vartheta,\delta}$ *is solution to Wiener-Hopf integral equation*

$$g_{\vartheta,\delta}(\delta^{-2}t,s) - \int_0^{\delta^{-2}t} c''_{\vartheta,\delta}(|s-s'|)g_{\vartheta,\delta}(\delta^{-2}t,s')ds' = -c''_{\vartheta,\delta}(s).$$

- Uses Gaussian correlation and mapping properties of cov. op.
- LHS in WH-equation is cov. op. of GP $g \mapsto \delta^{-1}\int_0^{\delta^{-2}t} g(s)dY(s)$.
- Computing Fisher information as $\delta\to 0$ requires $g_\vartheta(s) = \lim_{\delta\to 0} g_{\vartheta,\delta}(\delta^{-2}t,s) \in L^2(\mathbb{R}^+)$.

Wiener-Hopf integral equation

Expect $g_\vartheta(s) = \lim_{\delta\to 0} g_{\vartheta,\delta}(\delta^{-2}t, s)$ with

$$g_\vartheta(s) - \int_0^\infty c''_\vartheta(|s-s'|) g_\vartheta(s') ds' = -c''_\vartheta(s).$$

Note: $g - \int_0^\infty c''_\vartheta(|\cdot - s'|) g(s') ds'$ is *not* compact on $L^2(\mathbb{R}^+)$ nor on $C_0(\mathbb{R}^+)$, so Fredholm theory does not apply and solution of WH-equation above may not exist (Wiener, Hopf (1931), Anselone, Sloan (1985)).

Instead:

- Rewrite as Volterra equation

$$g_{\vartheta,\delta}(\delta^{-2}t, s) = \int_0^s 2c''_{\vartheta,\delta}(|s-s'|) g_{\vartheta,\delta}(\delta^{-2}t, s') ds' + \text{rem}.$$

- Obtain unique solution in $C(\mathbb{R}^+)$ as $\delta \to 0$ on all intervals $[0, a]$.

Wiener-Hopf integral equation

- Note $g_{\vartheta,\delta}(t,s) = (-C^t_{\vartheta,\delta}{}'')^{-1} c''_{\vartheta,\delta} = -(C^t_{\vartheta,\delta})^{-1} c_{\vartheta,\delta} + \text{rem}$.
- Using properties of RKHS have

$$\begin{aligned}
&\| \int_0^{\delta^{-2}t} c''_{\vartheta,\delta}(|\cdot - s'|) g_{\vartheta,\delta}(\delta^{-2}t, s') ds' \|_{L^2([0,\delta^{-2}t])} \\
&\quad = \| U g_{\vartheta,\delta}(\delta^{-2}t, \cdot) \|_{L^2([0,\delta^{-2}t])} \\
&\quad = \sup_f \langle Uf, g_{\vartheta,\delta}(\delta^{-2}t, \cdot) \rangle_{L^2([0,\delta^{-2}t])} \\
&\quad \leq \sup_f \| (C^t_{\vartheta,\delta})^{-1/2} Uf \|_{L^2([0,\delta^{-2}t])} \left(\| (C^t_{\vartheta,\delta})^{-1/2} c_{\vartheta,\delta} \|_{L^2([0,\delta^{-2}t])} + \text{rem} \right) \\
&\quad < \infty! \text{ (uniformly in } \delta, 0 \leq t \leq T).
\end{aligned}$$

- So $g_\vartheta \in L^2(\mathbb{R}^+)$ after all.

Rescaling

Using Gaussian concentration (Borell-Sudakov-Tirelson) have

$$\int_0^T \left(m_{\vartheta+h_\delta}(t) - m_\vartheta(t)\right)^2 dt$$

$$= o_{\mathbb{P}_\vartheta}(1) + \mathbb{E}_\vartheta\left[\int_0^T \left(m_{\vartheta+h_\delta}(t) - m_\vartheta(t)\right)^2\right] dt$$

$$= o_{\mathbb{P}_\vartheta}(1) + \delta^{-2}\int_0^T \mathbb{E}_\vartheta\left[\left(\int_0^{\delta^{-2}t}\left(g_{\vartheta+h_\delta,\delta}(\delta^{-2}t,s) - g_{\vartheta,\delta}(\delta^{-2}t,s)\right) dY(s)\right)^2\right] dt$$

$$\xrightarrow{\mathbb{P}_\vartheta} \vartheta^{-1}T\mathcal{I}.$$

Main result

Theorem

Let $Z \sim N(0, \mathcal{I})$. Under regularity assumptions on K, $M \leq \rho^{-d}$ and $\rho = \delta^{\frac{2d}{2+3d}} \log \delta^{-1}$, we find

$$\log \frac{d\mathbb{P}_{\vartheta + M^{-1/2}\delta}}{d\mathbb{P}_{\vartheta}}(X_\delta) = \vartheta^{-1/2} Z - \frac{1}{2}\vartheta^{-1} T\mathcal{I} + o_{\mathbb{P}_\vartheta}(1), \quad \delta \to 0, M \to \infty.$$

- inverse Fisher information is linear in ϑ.
- $\vartheta T^{-1}\mathcal{I}^{-1}$ is minimal variance for estimating ϑ from local measurements

$$X_\delta = (\langle X(t), K_{\delta,x_k} \rangle, 0 \leq t \leq T, k = 1, \ldots, M).$$

- The result transfers to nonparametric diffusivity ϑ.
- Eventually we want to do Bayesian inference on ϑ ...

Outline

Tractable likelihood

Now: $\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y)) + \sigma \partial_t W(t,y)$ on $[0,T] \times \Lambda$.

Observe

$$(X(t,y))_{0 \le t \le T, y \in \Lambda},$$

so this is a regression problem:

$$\underbrace{\partial_t X(t,y) - \vartheta \Delta X(t,y)}_{\text{observable}} = f(X(t,y)) + \sigma \partial_t W(t,y).$$

Log-likelihood by Girsanov's thm:

$$\ell(f) = \frac{1}{\sigma^2} \int_0^T \langle f(X(t)), dX(t) - \vartheta \Delta X(t) dt \rangle_{L^2(\Lambda)} - \frac{1}{2\sigma^2} \int_0^T \| f(X(t)) \|_{L^2(\Lambda)}^2 dt.$$

Posterior is Gaussian

Suppose Π is an M-dim. Gaussian process prior on $f \in \Theta \subset C^3(\mathbb{R})$.

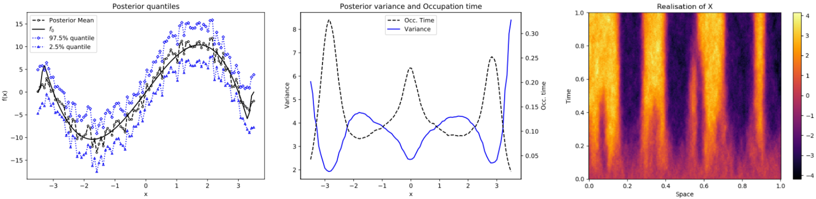
Given an ONB (e_k) of $L^2(U)$, $\mathrm{supp} f \subset U \subset \mathbb{R}$ compact, then

$$\Pi(f|X) = N(\mu, Q)$$

with

$$Q_{jk}^{-1} = \frac{1}{\sigma^2}\int_0^T \langle e_k(X(t)), e_j(X(t))\rangle_{L^2(\Lambda)}\, dt + \Sigma_{jk}^{-1},$$

$$\mu = Q\left(\frac{1}{\sigma^2}\int_0^T \langle e_k(X(t)), dX(t) - \vartheta\Delta X(t)dt\rangle_{L^2(\Lambda)}\right)_{k=1}^{M}.$$



A concentration result

Let $\vartheta = \sigma^{-1/4}$. Define

$$h_\sigma^2(f,g) := \int_0^T \|f(X(t)) - g(X(t))\|_{L^2(\Lambda)}^2 dt.$$

Theorem

If $f_0 \in C^3(\mathbb{R})$, $f, g \in L^2(\mathbb{R})$, then for $x \geq 0$

$$\mathbb{P}_{f_0}\left(\left|h_\sigma^2(f,g) - \mathbb{E}_{f_0}\left[h_\sigma^2(f,g)\right]\right| \geq \sigma x\right) \leq 2\exp\left(-C\frac{x^2}{\sigma^2\|f-g\|_{L^2(\mathbb{R})}^2}\right).$$

- $y \mapsto X(t,y)$ is hard to control (generally no Itô formula, no local time).
- Proof based on Malliavin calculus and fine density bounds for law of $X(t,y)$.

Posterior contraction

Recall: $h_\sigma^2(f,g) := \int_0^T \|f(X(t)) - g(X(t))\|_{L^2(\Lambda)}^2 dt$.

For $\alpha > 3$ suppose $f_0 \in C^\alpha(U) \cap H^\alpha(U)$, $r_\sigma = (\sigma^2)^{\frac{\alpha}{2\alpha+1}} \log(\sigma^{-2})$.

Theorem

Let Π be a (finite-dim.) Gaussian wavelet prior on $L^2(U)$ (cut-off $J \approx r_\sigma^{-1/\alpha}$) with RKHS $H^{\alpha+1/2}(U)$. Then for every $M_\sigma \to \infty$

$$\Pi\left(f : \mathbb{E}_{f_0}\left[h_\sigma^2(f,f_0)\right] \geq M_\sigma r_\sigma^2\right) \xrightarrow[\sigma\to 0]{\mathbb{P}_{f_0}} 0.$$

- proof based on v.d.Meulen, v.d. Vaart, v. Zanten ('06), see also Nickl, Ray ('20).
- As long as posterior is supported on compact $\bar{U} \subset U$, we have

$$\|f - f_0\|_{L^2(U)}^2 \lesssim \mathbb{E}_{f_0}\left[h_\sigma^2(f,f_0)\right] \lesssim \|f - f_0\|_{L^2(U)}^2.$$

Summary

- Statistical inference for semilinear SPDEs

$$\partial_t X(t,y) = \vartheta \Delta X(t,y) + f(X(t,y)) + \sigma \partial_t W(t,y).$$

- LAN property for ϑ from local measurements as $\delta \to 0$.
- Posterior contraction with nonparametric rates for f from full observations on $[0,T] \times \Lambda$ as $\sigma = \vartheta^{1/4} \to 0$.
- Ongoing work: nonparametric BvM in both settings (à la Nickl, Ray ('18)).