

# Pathwise least-squares estimator for linear SPDEs with additive fractional noise<sup>\*</sup>

Pavel Kríž

*Charles University, Faculty of Mathematics and Physics, Sokolovská 83, Prague 8, 186 75, Czech Republic*

*e-mail:* [kriz@karlin.mff.cuni.cz](mailto:kriz@karlin.mff.cuni.cz)

*University of Chemistry and Technology, Prague, Department of Mathematics, Technická 5, Prague 6, Czech Republic*

and

Jana Šnupárková

*University of Chemistry and Technology, Prague, Department of Mathematics, Technická 5, Prague 6, Czech Republic*

*e-mail:* [jana.snuparkova@vscht.cz](mailto:jana.snuparkova@vscht.cz)

**Abstract:** This paper deals with the drift estimation in linear stochastic evolution equations (with emphasis on linear SPDEs) with additive fractional noise (with Hurst index ranging from 0 to 1) via least-squares procedure. Since the least-squares estimator contains stochastic integrals of divergence type, we address the problem of its pathwise (and robust to observation errors) evaluation by comparison with the pathwise integral of Stratonovich type and using its chain-rule property. The resulting pathwise LSE is then defined implicitly as a solution to a non-linear equation. We study its numerical properties (existence and uniqueness of the solution) as well as statistical properties (strong consistency and the speed of its convergence). The asymptotic properties are obtained assuming fixed time horizon and increasing number of the observed Fourier modes (space asymptotics). We also conjecture the asymptotic normality of the pathwise LSE.

**MSC2020 subject classifications:** Primary 62M09, 60H15, 60G22.

**Keywords and phrases:** Drift estimation, least squares, linear SPDEs, fractional Brownian motion, pathwise stochastic integration.

Received May 2021.

## 1. Introduction

Least-squares type estimators of an unknown drift parameter have recently become very popular in the setting of (semi)linear SPDEs driven by an infinite-di-

---

<sup>\*</sup>The work of PK was supported by the Czech Science Foundation project No. 19-07140S and the work of JŠ was supported by the grant LTAIN19007 Development of Advanced Computational Algorithms for Evaluating Post-surgery Rehabilitation.

mensional Brownian motion, because they have advantageous asymptotic properties (they basically coincide with MLE estimators in this case) and are relatively easy to implement. For recent theoretical works, the reader may check [27], [7], [2] or [3] to name just a few. The theory for models with Brownian motion has reached certain level of maturity so that it enables several interesting practical applications, such as [1] or [26].

Wide range of randomly evolving phenomena are forced with auto-correlated noise, which can be effectively modeled by a fractional Brownian motion (fBm). Stochastic evolution equations driven by a fBm appear in diverse fields such as biology, neuroscience, hydrology, climatology, finance and many others (see e.g. monographs [5] or [23] for discussion). Least-squares estimators of the drift parameter for continuously observed trajectories of solutions to one-dimensional linear equations with additive fractional noise have been studied in [13] for Hurst index  $H \geq 1/2$  and [14] for general  $H \in (0, 1)$ .

The available literature on parameter estimation for infinite-dimensional equations driven by an (infinite-dimensional) fBm is rather sparse. A least-squares estimator for a one-dimensional projection of the solution has been studied in long-span asymptotic regime in [22] assuming the regular case  $H \geq 1/2$ . Drift estimation in spectral setting (Fourier modes are observed) for infinite-dimensional equations with fractional noise has been studied in [8] (MLE approach,  $H \geq 1/2$ ) and [17] (ergodic-type estimator,  $H \in (0, 1)$ ). We are not aware of any study on LSE in spectral setting (Fourier modes are observed) for these models.

LSE-type estimators typically contain stochastic integrals, which makes them difficult to apply. For equations with Brownian noise, evaluations of Itô stochastic integrals may be sensitive to small perturbations of trajectories. A robust version of LSE can be obtained by application of Itô formula (for simple equations) or rough path theory (more complex equations), see [11]. The least-squares estimators for equations with fBm incorporate divergence-type integrals (sometimes called the Skorokhod integrals). These are not defined pathwise, but rather as the adjoint operators to Malliavin derivatives, which makes them extremely difficult (if not impossible) to evaluate from the observed trajectory in practice or in simulations, which prevents them from wider use. To avoid these complications, the authors in [14] switched to the ergodic-type estimator for one-dimensional fractional Ornstein-Uhlenbeck process, which contains only Riemann integrals. Estimators of ergodic type for more general equations driven by a finite-dimensional fBm were recently studied in [25]. Application of this ergodic approach in the spectral setting for infinite-dimensional equations corresponds to the weighted minimum-contrast estimator (weighted MCE) introduced in [17]. Such estimator is plausible for stationary processes, but fails for processes at far-from-stationary state and is difficult to generalize to more complex equations, because (in contrast to LSE) it requires ergodicity and precise knowledge of the limiting covariance operator.

In this paper, we start with derivation of a new spectral version of the least-squares estimator (theoretical LSE) for an unknown drift parameter in linear SPDEs driven by an additive infinite-dimensional fractional Brownian motion and prove its strong consistency in space (incl. speed of convergence) by the

tools of Malliavin calculus. We choose to work in spectral setting (observations in spectral space domain rather than in physical space domain are available) and to study asymptotic properties under increasing number of the observed Fourier modes while time horizon is fixed. Such approach makes it possible to consistently (with error approaching zero) estimate the unknown parameter based on observations in fixed time horizon, but with sufficient detail in space (number of Fourier modes). This is particularly useful for observations having limited time horizon but high space resolution/detail. It contrasts to more typical time asymptotics, which assumes increasing time horizon and which is not studied here. Moreover, severe complications with its numerical evaluation has led us to construct a pathwise version of the LSE, which is defined (and computable) from the observed trajectory and which is robust to small perturbations in the observations.

Our construction of the robust pathwise LSE is inspired by the work [11]. It is based on calculation of pathwise Stratonovich integral, which can be expressed explicitly in our case, and compensation for difference between Stratonovich and divergence-type integrals. In our setting with fBm, this leads to an implicitly defined estimator. We prove its existence and uniqueness, as well as its strong consistency in space. Surprisingly, if  $H > 1/2$ , it has slightly better performance than the theoretical LSE. To our best knowledge, this is the first attempt to define a pathwise robust LSE for equations driven by fBm. We believe that this pathwise least-squares approach is applicable also for different models with fBm (such as those in [22] or [14]) and may offer estimators applicable in practice or in simulations.

Main purpose of this paper is to make the first step in the development of a theoretical framework necessary for studying spectral asymptotic properties as well as performing effective enumeration of the least-squares estimators for stochastic evolution equations with infinite-dimensional auto-correlated noise. The assumed linearity of the operators ensures reasonable clarity and simplicity of the exposition. The next step would be to study this problem in the setting of semilinear equations with non-linear lower order perturbation with numerous practical applications. This, however, requires fine analysis of space regularity of solutions and is beyond the scope of this work.

The paper is organized as follows. In section 2 we introduce the studied problem in more detail. The two new estimators – theoretical and pathwise LSE – are derived in section 3. Existence and uniqueness of the implicitly defined pathwise LSE is studied in section 4. Section 5 is devoted to the strong consistency of the estimators. A simulation study is presented in section 6. Section 7 summarizes main results and findings of this paper. In Appendix, we briefly describe some elements of Malliavin calculus, which are useful for the asymptotic analysis and outline the connection between cylindrical fractional Brownian motion and Gaussian noise that is white in space and fractional in time.

## 2. Problem setup

### 2.1. The model

Consider the evolution equation

$$dX(t) = \lambda AX(t) dt + BX(t) dt + dB^H(t), \quad X(0) = X_0, \quad (1)$$

in a real separable Hilbert space  $V$  where  $A, B$  are densely defined linear operators,  $B^H$  is a cylindrical fractional Brownian motion on  $V$ ,  $X_0 \in V$  and  $\lambda$  is a positive unknown parameter, which we want to estimate.

Throughout this paper we assume that the equation (1) is diagonalizable, i.e. there exists an orthonormal basis  $\{e_k\}_{k=1}^\infty$  in  $V$  such that for any  $k \in \mathbb{N}$  we have

$$Ae_k = -\alpha_k e_k, \quad Be_k = -\beta_k e_k, \quad \langle B^H(t), e_k \rangle_V = B_k^H(t),$$

where  $\alpha_k \geq 0$  and  $\beta_k \geq 0$ . Furthermore, set

$$\mu_k = \lambda\alpha_k + \beta_k, \quad k \in \mathbb{N}.$$

*Remark 2.1.* For existence of the solution to (1) in an appropriate interpolation space, constructed via spectral decomposition, it suffices that for some  $\gamma \in \mathbb{R}$

$$\sum_{k=1}^{\infty} \frac{1}{(1 + \mu_k)^\gamma} < \infty \quad (2)$$

holds. For more details and proof, see e.g. Theorem 2.1. (and Remark 2.1.) in [17].

We assume (throughout the paper) that the existence condition (2) is satisfied. It also implies that

$$\lim_{k \rightarrow \infty} \mu_k = \infty. \quad (3)$$

Denote the Fourier modes (projections to eigenvectors) of the solution by  $x_k(t) = \langle X(t), e_k \rangle_V$ . By diagonality of the equation (1) these projections satisfy the system of independent one-dimensional equations

$$\begin{aligned} dx_k(t) &= -\mu_k x_k(t) dt + dB_k^H(t), \quad t > 0, \quad k \in \mathbb{N}, \\ x_k(0) &= \langle X_0, e_k \rangle_V, \end{aligned} \quad (4)$$

where  $\{B_k^H(t), t \geq 0\}$  are mutually independent real-valued fractional Brownian motions with the same Hurst index  $H \in (0, 1)$ .

The solutions to the equations (4) are mutually independent real-valued fractional Ornstein-Uhlenbeck processes given by a formula

$$x_k(t) = e^{-\mu_k t} x_k(0) + \int_0^t e^{-\mu_k(t-s)} dB_k^H(s), \quad t \geq 0. \quad (5)$$

for any  $k \in \mathbb{N}$ .

*Remark 2.2.* If the operators  $A, B$  from (1) are negative definite self-adjoint elliptic linear differential operators of even orders  $2m_1, 2m_2$  ( $m_1, m_2 \in \mathbb{N}$ ), respectively, acting on a compact  $d$ -dimensional manifold, then (see [30])

$$\alpha_k \sim k^{2m_1/d}, \text{ and } \beta_k \sim k^{2m_2/d}, \quad (6)$$

where

$$a_k \sim b_k \text{ means that } a_k/b_k \rightarrow C \in (0, \infty) \text{ as } k \rightarrow \infty. \quad (7)$$

Such operators are of great interest as they frequently appear in various stochastic PDEs.

## 2.2. Statistical problem

To estimate the value of  $\lambda$ , we have continuous sample trajectories of the first  $N$  Fourier modes observed on a fixed time-window  $[0, T]$  at our disposal, i.e. our data are  $\{x_k(t), t \in [0, T]\}$ ,  $k = 1, \dots, N$ . We aim at deriving an estimator that would be computable from this data and that would be consistent with increasing number of Fourier modes  $N \rightarrow \infty$  (the so-called space asymptotics), the time horizon  $T$  being fixed.

*Remark 2.3.* In this paper we assume that the Hurst parameter  $H$  is known. If not, it can be consistently determined from a single continuous trajectory of the process  $\{x_1(t), t \in [0, T]\}$  by some of various techniques developed for estimating  $H$  for real-valued processes sampled under in-fill asymptotics regime (discrete time observation with fixed time horizon and decreasing time step), cf. [15], [12], [9], or [29]) to name just a few. If, moreover, the noise processes for individual Fourier modes  $B_k^H$  have different intensities (volatilities, denote by  $\sigma_k$ ), we can simultaneously and consistently estimate  $\sigma_k$  and  $H$  from the observed modes under in-fill asymptotics regime using the powers of the second order variations (see [4], Chapter 3.3) and subsequently eliminate different noise intensities by appropriate rescaling of the observed processes  $x_k$ .

## 3. Derivation of the estimator

### 3.1. Theoretical LSE

We derive the estimator for the drift parameter by applying the least-squares concept to the SDEs understood as a system of linear regression models. In particular, we find  $\lambda$  which minimizes (for fixed time horizon  $T$ ) the formal sum

$$\sum_{k=1}^N \int_0^T (\dot{x}_k(t) + \mu_k x_k(t))^2 dt.$$

Note that this is only a heuristic technique, since time derivatives  $\dot{x}_k(t)$  do not exist (in the classical sense). However, we can rewrite the minimizer back in

terms of the well-defined stochastic integrals and Riemann integrals. It reads

$$\hat{\lambda}_N = -\frac{\sum_{k=1}^N \int_0^T \alpha_k x_k(t) dx_k(t) + \sum_{k=1}^N \int_0^T \alpha_k \beta_k x_k^2(t) dt}{\sum_{k=1}^N \int_0^T \alpha_k^2 x_k^2(t) dt}, \quad (8)$$

where the stochastic integral is understood in the Skorokhod sense (sometimes referred to as the divergence type integral). Applying the stochastic differential from the equation (4) we obtain

$$\hat{\lambda}_N = \lambda - \frac{\sum_{k=1}^N \alpha_k \int_0^T x_k(t) dB_k^H(t)}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt}. \quad (9)$$

This form is convenient for proving the asymptotic properties of the estimator  $\hat{\lambda}_N$  whereas the form (8) is better for derivation of the pathwise version of the LSE for  $\lambda$ .

*Remark 3.1.* We did not specify the type of stochastic integral with respect to the fBm in the equations (1) or (4), because all integration concepts coincide in case of additive noise. However, the choice of the appropriate type of stochastic integration in (9) (and in (8)) is critical, since the integrands are stochastic processes here. The choice of the Skorokhod type integral is motivated by the fact that its expectation is zero, which is advantageous for the error term in (9) and will be utilized in the sequel. If we choose the pathwise Stratonovich integral, for example, its nonzero expectation would generate unwanted bias in (9) and the corresponding estimator would not be consistent.

### 3.2. Robust pathwise LSE

Although the Skorokhod type integral in (8) has advantageous probabilistic properties, it is not constructed pathwise, but rather as the adjoint operator to Malliavin derivative. This imposes severe complications to numerical evaluation of the estimator  $\hat{\lambda}_N$  for the observed single trajectory. To overcome this issue, we derive below its robust pathwise version.

The strategy, inspired by the paper [11], is to express the estimator in terms of the pathwise Stratonovich integral (for definition, see [14]) with appropriate compensation for difference between Stratonovich and Skorokhod integrals.

$$\begin{aligned} \int_0^T x_k(t) dx_k(t) &= \int_0^T -\mu_k x_k^2(t) dt + \int_0^T x_k(t) dB_k^H(t) \\ &= \int_0^T -\mu_k x_k^2(t) dt + \left( \int_0^T x_k(t) \circ dB_k^H(t) - \Delta(\mu_k) \right) \\ &= \int_0^T x_k(t) \circ dx_k(t) - \Delta(\mu_k), \end{aligned}$$

where  $\int_0^T x_k(t) \circ dx_k(t)$  stands for the pathwise Stratonovich integral and  $\Delta(\mu_k)$  is the compensation. Since  $x_k$  is in the first Wiener chaos (it is Gaussian), the

compensation  $\Delta(\mu_k)$  takes rather simple form (see [14], formulas (3.4)-(3.6)):

$$\begin{aligned}\Delta(\mu_k) &= \mathbb{E} \left[ \int_0^T x_k(t) \circ dB_k^H(t) \right] \\ &= \frac{1}{2} T^{2H} (1 - \gamma(1, \mu_k T)) - \mu_k^{-2H} \gamma(2H + 1, \mu_k T) \left( H - \frac{1}{2} \right) \\ &\quad + TH \mu_k^{1-2H} \gamma(2H, \mu_k T),\end{aligned}\tag{10}$$

where

$$\gamma(h, T) = \int_0^T e^{-x} x^{h-1} dx.$$

Moreover, since Stratonovich integral satisfies the rules of the first order calculus, we can easily evaluate (without need for rough path lift):

$$\int_0^T x_k(t) \circ dx_k(t) = \frac{x_k^2(T) - x_k^2(0)}{2}.$$

This provides a simple pathwise formula for evaluation of the Skorokhod-type integral

$$\int_0^T x_k(t) dx_k(t) = \frac{x_k^2(T) - x_k^2(0)}{2} - \Delta(\mu_k),$$

and the corresponding formula for LSE

$$\hat{\lambda}_N = - \frac{\sum_{k=1}^N \alpha_k \left( \frac{x_k^2(T) - x_k^2(0)}{2} - \Delta(\mu_k) \right) + \sum_{k=1}^N \alpha_k \beta_k \int_0^T x_k^2(t) dt}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt}.$$

Unfortunately,  $\mu_k = \alpha_k \lambda + \beta_k$  in this evaluation formula depend on the unknown value of parameter  $\lambda$ . A natural workaround here is to define new estimator (the pathwise LSE) as a solution to the following equation (in unknown  $\Lambda$ ):

$$\Lambda = - \frac{\sum_{k=1}^N \alpha_k \left( \frac{x_k^2(T) - x_k^2(0)}{2} - \Delta(\alpha_k \Lambda + \beta_k) \right) + \sum_{k=1}^N \alpha_k \beta_k \int_0^T x_k^2(t) dt}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt}.\tag{11}$$

*Remark 3.2.* Although the theoretical LSE formula (8) does not explicitly contain the Hurst parameters  $H$ , its value is needed to evaluate the stochastic Skorokhod-type integral. Hence, it is present in defining equation for pathwise (computable) version of LSE. However, since we consider continuous-time setting, value of  $H$  can be determined exactly from continuous observation of a single trajectory  $\{x_k(t), t \in [0, T]\}$ , as discussed above.

*Remark 3.3.* If  $H = 1/2$  the pathwise LSE coincides with the theoretical LSE. This follows from the fact that the right-hand side of (11) is constant (in  $\Lambda$ ) in this case, which is shown below.

#### 4. Existence and uniqueness of pathwise LSE

We can not expect the existence and uniqueness of the positive solution to (11) in general. This can be seen on the simplest example of LSE for one-dimensional Ornstein-Uhlenbeck process (driven by a Wiener process)

$$\hat{\lambda} = -\frac{\frac{x^2(T)-x^2(0)}{2} - \frac{T}{2}}{\int_0^T x^2(t) dt},$$

which can provide negative estimates. We address this problem below by taking positive part of the solution and specifying, which one to choose if there are multiple solutions. To derive the conditions for the existence and uniqueness of the positive solution to equation (11), we need to understand the properties of the function on its right-hand side. Set

$$R_N(\Lambda) = -\frac{\sum_{k=1}^N \alpha_k \left( \frac{x_k^2(T)-x_k^2(0)}{2} - \Delta(\alpha_k \Lambda + \beta_k) \right) + \sum_{k=1}^N \alpha_k \beta_k \int_0^T x_k^2(t) dt}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt}, \quad (12)$$

with function  $\Delta$  defined in (10). Direct calculations lead to the following formula for its first derivative (after cancelling some terms)

$$\begin{aligned} \frac{d\Delta}{d\mu}(\mu) &= H(1-2H)(T\mu^{-2H}\gamma(2H, \mu T) - \mu^{-2H-1}\gamma(2H+1, \mu T)) \\ &= H(1-2H)\mu^{-2H-1} \int_0^{\mu T} e^{-s} s^{2H-1}(\mu T - s) ds \\ &\begin{cases} > 0 \dots H < 1/2 \\ = 0 \dots H = 1/2 \\ < 0 \dots H > 1/2 \end{cases}, \quad \forall \mu > 0, \end{aligned} \quad (13)$$

and the second derivative

$$\begin{aligned} \frac{d^2\Delta}{d\mu^2}(\mu) &= H(2H-1)\mu^{-2H-2} \int_0^{\mu T} e^{-s} s^{2H-1}(2H\mu T - (2H+1)s) ds \\ &\begin{cases} < 0 \dots H < 1/2 \\ = 0 \dots H = 1/2 \\ > 0 \dots H > 1/2 \end{cases}, \quad \forall \mu > 0, \end{aligned} \quad (14)$$

in view of

$$\begin{aligned} &\int_0^{\mu T} e^{-s} s^{2H-1}(2H\mu T - (2H+1)s) ds \\ &> \int_0^{\mu T 2H/(2H+1)} e^{-\mu T 2H/(2H+1)} s^{2H-1}(2H\mu T - (2H+1)s) ds \\ &+ \int_{\mu T 2H/(2H+1)}^{\mu T} e^{-\mu T 2H/(2H+1)} s^{2H-1}(2H\mu T - (2H+1)s) ds = 0. \end{aligned}$$



Derivatives of  $R$  are immediate consequence of (13) and (14)

$$\frac{dR_N}{d\Lambda}(\Lambda) = \frac{\sum_{k=1}^N \alpha_k^2 \frac{d\Delta}{d\mu}(\alpha_k \Lambda + \beta_k)}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt} \begin{cases} > 0 \dots H < 1/2 \\ = 0 \dots H = 1/2 \\ < 0 \dots H > 1/2 \end{cases}, \quad \forall \Lambda > 0, \quad (15)$$

and

$$\frac{d^2 R_N}{d\Lambda^2}(\Lambda) = \frac{\sum_{k=1}^N \alpha_k^3 \frac{d^2 \Delta}{d\mu^2}(\alpha_k \Lambda + \beta_k)}{\sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt} \begin{cases} < 0 \dots H < 1/2 \\ = 0 \dots H = 1/2 \\ > 0 \dots H > 1/2 \end{cases}, \quad \forall \Lambda > 0.$$

Moreover, since

$$\lim_{\mu \rightarrow 0_+} \Delta(\mu) = \frac{1}{2} T^{2H},$$

we obtain

$$\lim_{\Lambda \rightarrow 0_+} R_N(\Lambda) =: R_N(0) \in \mathbb{R}. \quad (16)$$

Direct calculations yield

$$\lim_{\Lambda \rightarrow \infty} R_N(\Lambda) = \begin{cases} \infty \dots H < 1/2 \\ R_N^\infty \in \mathbb{R} \dots H \geq 1/2 \end{cases}$$

and

$$\lim_{\Lambda \rightarrow \infty} R_N(\Lambda) - \Lambda = -\infty, \quad \forall H \in (0, 1).$$

The above properties of  $R_N$  lead us to the following conclusion:

**Theorem 4.1.** Fix  $N$  and consider the function  $R_N$  defined in (12) with  $R_N(0)$  defined in (16) and let  $H \in (0, 1)$ . Sufficient condition for the existence and uniqueness of the positive solution to equation (11), which can be rephrased as  $R_N(\Lambda) = \Lambda$ , is

$$R_N(0) > 0.$$

If  $H \geq 1/2$ , this condition is also necessary.

Note that the condition  $R_N(0) > 0$  can be easily verified from the data by direct evaluation of  $R_N$  at  $\Lambda = 0$  (if  $\beta_k$  are positive) or by calculating the limit for  $\Lambda \rightarrow 0_+$  (if  $\beta_k$  are zero). We are now in a position to define the pathwise LSE correctly:

$$\tilde{\lambda}_N := \begin{cases} \text{positive solution to (11), if it exists and is unique;} \\ 0, \text{ if there is no positive solution to (11);} \\ \text{the greater solution to (11), if there are two positive solutions.} \end{cases} \quad (17)$$

*Remark 4.1.* The choice of the greater solution in the third case ensures that  $\frac{dR_N}{d\Lambda}(\tilde{\lambda}_N) < 1$ , which corresponds to the limiting behavior of  $\frac{dR_N}{d\Lambda}$  in the neighborhood of the true value  $\lambda$ , detailed in Lemma 5.4.

*Remark 4.2.* Monotonicity and convexity (if  $H > 1/2$ ) or concavity (if  $H < 1/2$ ), respectively, ensure that the Newton-Raphson numerical method is well suited for numerical approximation of  $\tilde{\lambda}_N$ .

## 5. Strong consistency

### 5.1. Strong consistency of theoretical LSE

We start with strong consistency of the theoretical LSE  $\hat{\lambda}_N$  defined in (8), because it will be helpful to show strong consistency of the pathwise LSE in the sequel. For simplicity, we assume throughout this section that

$$x_k(0) = 0, \quad k = 1, 2, \dots$$

The self-similarity of a fBm enables us to study the effect of different drifts  $\mu_k$  on the distributions of the processes  $x_k$ . Its combination with the (long-span) asymptotic behavior of a real-valued fractional Ornstein-Uhlenbeck process, studied in [14], forms the strategy of the proof of the strong consistency (in space) of  $\hat{\lambda}_N$ .

Using integration-by-parts in the formula (5) it is possible to show that the solutions  $\{x_k(t), t \geq 0\}$  can be expressed as

$$x_k(t) = \left( B_k^H(t) - \mu_k e^{-\mu_k t} \int_0^t e^{\mu_k s} B_k^H(s) ds \right), \quad t \geq 0, \quad (18)$$

for any  $k \in \mathbb{N}$ . This can also be verified by direct substitution of (18) into the integral form of equation (4).

Let  $\{\tilde{x}_k(t), t \geq 0\}$  be the solution to the equation with unit drift

$$\begin{aligned} d\tilde{x}_k(t) &= -\tilde{x}_k(t) dt + dB_k^H(t), \quad t > 0, \quad k \in \mathbb{N}, \\ \tilde{x}_k(0) &= 0, \end{aligned} \quad (19)$$

given by the formula

$$\tilde{x}_k(t) = \int_0^t e^{-(t-s)} dB_k^H(s), \quad t \geq 0,$$

or

$$\tilde{x}_k(t) = \left( B_k^H(t) - \int_0^t e^{-(t-s)} B_k^H(s) ds \right), \quad t \geq 0, \quad (20)$$

for any  $k \in \mathbb{N}$ .

Using the self-similarity of a fractional Brownian motion in the expression (20) we get the equality of distributions

$$\{x_k(t), t \geq 0\} \stackrel{d}{=} \left\{ \frac{1}{\mu_k^H} \tilde{x}_k(\mu_k t), t \geq 0 \right\} \quad (21)$$

for any  $k \in \mathbb{N}$ .

Moreover, let  $\{y(t), t \geq 0\}$  be the stationary and ergodic solution to (19) (for  $k = 1$ ) given by a formula

$$y(t) = e^{-t}y_0 + \int_0^t e^{-(t-s)} dB_1^H(s), \quad t \geq 0.$$

Then

$$\mathbb{E} y(t) = 0, \quad \text{var}(y(t)) = H\Gamma(2H)$$

for any  $t \geq 0$  and

$$\frac{1}{T} \int_0^T y^2(t) dt \xrightarrow{T \rightarrow \infty} H\Gamma(2H) \quad \text{a.s. and in } L^1. \quad (22)$$

**Lemma 5.1.** *The following convergence*

$$\mathbb{E} \left[ \frac{1}{T} \int_0^T (\tilde{x}_1^2(t) - y^2(t)) dt \right] \xrightarrow{T \rightarrow \infty} 0$$

*holds.*

*Proof.* We have

$$\begin{aligned} & \mathbb{E} \left[ \frac{1}{T} \int_0^T (\tilde{x}_1^2(t) - y^2(t)) dt \right] \\ &= \mathbb{E} \left[ \frac{1}{T} \left( \int_0^T 2y(t)(\tilde{x}_1(t) - y(t)) dt + \int_0^T (\tilde{x}_1(t) - y(t))^2 dt \right) \right] \\ &= \frac{1}{T} \left( \int_0^T -2e^{-t} \mathbb{E}[y_0 y(t)] dt + \int_0^T e^{-2t} \mathbb{E}[y_0^2] dt \right) \\ &= \frac{1}{T} \int_0^1 -2e^{-t} \mathbb{E}[y_0 y(t)] dt + \frac{1}{T} \int_1^T -2e^{-t} \mathbb{E}[y_0 y(t)] dt \\ &\quad + H\Gamma(2H) \frac{1}{T} \int_0^T e^{-2t} dt. \end{aligned}$$

Obviously, the first and third term tend to zero as  $T$  goes to infinity. Also the second term tends to zero due to [16], Lemma 3.1, and the limit properties of

$$\mathbb{E}[y_0 y(t)] = \frac{1}{2} 2H(2H-1)t^{2H-2} + O(t^{2H-4}), \quad t \rightarrow \infty,$$

for  $H \neq 1/2$  (see [6], Theorem 2.3) and

$$\mathbb{E}[y_0 y(t)] = \frac{1}{2} e^{-t}, \quad t \geq 0,$$

for the Wiener case  $H = 1/2$ . □

Define the weak asymptotic equivalence  $\asymp$  for sequences  $\{a_N\}_{N \in \mathbb{N}}, \{b_N\}_{N \in \mathbb{N}}$  by the relation

$$\left(a_N \asymp b_N\right) \equiv \left(\exists 0 < l < L < \infty \text{ such that } l b_N \leq a_N \leq L b_N, \forall N \in \mathbb{N}\right). \quad (23)$$

Note that this relation is weaker than the equivalence relation  $\sim$  defined in (7).

Recall the formula (9) (suitable for probabilistic analysis) and denote

$$C_N = \sum_{k=1}^N \alpha_k \int_0^T x_k(t) \, dB_k^H(t),$$

and

$$D_N = \sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) \, dt,$$

so that

$$\hat{\lambda}_N = \lambda - \frac{C_N}{D_N}.$$

For the needs of Lemma A.1 we explore the limit behavior of the variance of  $\frac{C_N}{\mathbb{E} D_N}$  and  $\frac{D_N}{\mathbb{E} D_N}$ . Denote for a fixed  $T$

$$T_k := \mu_k T, \quad k \in \mathbb{N}.$$

**Proposition 5.1.** *The expectations of  $C_N$  and  $D_N$  satisfy*

$$\mathbb{E} C_N = 0 \quad \text{and} \quad \mathbb{E} D_N \asymp \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}}.$$

*Proof.* The first part is obvious. By Lemma 5.1

$$\mathbb{E} \left[ \frac{1}{T} \int_0^T \tilde{x}_1^2(t) \, dt \right] \xrightarrow{T \rightarrow \infty} H\Gamma(2H),$$

thus the sequence  $\left\{ 1/T_k \int_0^{T_k} \mathbb{E} \tilde{x}_1^2(t) \, dt \right\}_{k \in \mathbb{N}}$  is positive and bounded. Using (21) and substitution we get

$$\begin{aligned} \mathbb{E} D_N &= \sum_{k=1}^N \alpha_k^2 \int_0^T \mathbb{E} x_k^2(t) \, dt = \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \int_0^T \mathbb{E} \tilde{x}_k^2(\mu_k t) \, dt \\ &= \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \int_0^T \mathbb{E} \tilde{x}_1^2(\mu_k t) \, dt = \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} T \frac{1}{T_k} \int_0^{T_k} \mathbb{E} \tilde{x}_1^2(t) \, dt \asymp \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}}. \square \end{aligned}$$

To study the time asymptotics of  $\text{var} \left( \int_0^T \tilde{x}_1^2(t) \, dt \right)$ , we utilize the results from [14]. Denote

$$F_T := \int_0^T \tilde{x}_1(t) \, dB_1^H(t) = \int_0^T \left( \int_0^t e^{-(t-s)} \, dB_1^H(s) \right) \, dB_1^H(t).$$

By (3.9) in [14] this equals

$$F_T = \left[ \left( \frac{\tilde{x}_1^2(T)}{2} - \mathbb{E} \frac{\tilde{x}_1^2(T)}{2} \right) + \int_0^T (\tilde{x}_1^2(t) - \mathbb{E} \tilde{x}_1^2(t)) dt \right]. \quad (24)$$

Further define

$$Q_H(T) := \begin{cases} \frac{1}{T} & , \quad H < \frac{3}{4}, \\ \frac{1}{T \log T} & , \quad H = \frac{3}{4}, \\ \frac{1}{T^{4H-2}} & , \quad H > \frac{3}{4}. \end{cases}$$

Then by [14], Lemma 17, there exists a constant  $0 < C_H < \infty$  such that

$$\mathbb{E} [Q_H(T) F_T^2] \xrightarrow{T \rightarrow \infty} C_H. \quad (25)$$

As a consequence, we obtain:

**Lemma 5.2.**  $Q_H(T) \text{var} \left( \int_0^T \tilde{x}_1^2(t) dt \right) \xrightarrow{T \rightarrow \infty} C_H.$

*Proof.* In the first step we show that  $\text{var}(\tilde{x}_1^2(T)) \xrightarrow{T \rightarrow \infty} \text{var} y_0^2$ .

Since  $\tilde{x}_1(T) = y(T) - e^{-T} y_0$ , using the triangle inequality we obtain

$$\begin{aligned} \left( \sqrt{\mathbb{E} y^2(T)} - e^{-T} \sqrt{\mathbb{E} y_0^2} \right)^2 &\leq \mathbb{E} \tilde{x}_1^2(T) \leq \left( \sqrt{\mathbb{E} y^2(T)} + e^{-T} \sqrt{\mathbb{E} y_0^2} \right)^2 \\ H\Gamma(2H)(1 - e^{-T})^2 &\leq \mathbb{E} \tilde{x}_1^2(T) \leq H\Gamma(2H)(1 + e^{-T})^2. \end{aligned}$$

Similarly,

$$\mathbb{E} y_0^4 (1 - e^{-T})^4 \leq \mathbb{E} \tilde{x}_1^4(T) \leq \mathbb{E} y_0^4 (1 + e^{-T})^4.$$

Therefore

$$\begin{aligned} \overbrace{\mathbb{E} y_0^4 (1 - e^{-T})^4 - (H\Gamma(2H)(1 - e^{-T})^2)^2}^{\xrightarrow{T \rightarrow \infty} \text{var} y_0^2} &\leq \text{var} \tilde{x}_1^2(T) \\ &\leq \underbrace{\mathbb{E} y_0^4 (1 + e^{-T})^4 - (H\Gamma(2H)(1 - e^{-T})^2)^2}_{\xrightarrow{T \rightarrow \infty} \text{var} y_0^2}. \end{aligned}$$

In the second step we apply (25) and the first step to equality (24) and yield the statement.  $\square$

**Proposition 5.2.** *The variances of  $C_N$  and  $D_N$  satisfy*

$$\begin{cases} H < \frac{3}{4} & \Rightarrow \quad \text{var} C_N \asymp \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{4H-1}} & , \quad \text{var} D_N \asymp \sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^{4H+1}}, \\ H = \frac{3}{4} & \Rightarrow \quad \text{var} C_N \asymp \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^2} \log T_k & , \quad \text{var} D_N \asymp \sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^4} \log T_k, \\ H > \frac{3}{4} & \Rightarrow \quad \text{var} C_N \asymp \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^2} & , \quad \text{var} D_N \asymp \sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^4}. \end{cases}$$

*Proof.* Using again (3.9) in [14], (21) and the fact that  $\{\tilde{x}_k(t), t \geq 0\}$  have the same law for all  $k \in \mathbb{N}$ , we get equality of laws:

$$\int_0^T x_k(t) dB_k^H(t) \stackrel{d}{=} \frac{1}{\mu_k^{2H}} \int_0^{T_k} \tilde{x}_1(t) dB_1^H(t)$$

for any  $k \in \mathbb{N}$ . This implies that

$$\begin{aligned} \text{var } C_N &= \text{var} \left( \sum_{k=1}^N \alpha_k \int_0^T x_k(t) dB_k^H(t) \right) = \sum_{k=1}^N \alpha_k^2 \text{var} \left( \int_0^T x_k(t) dB_k^H(t) \right) \\ &= \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{4H}} \text{var} \left( \int_0^{T_k} \tilde{x}_1(t) dB_1^H(t) \right) = \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{4H}} \frac{1}{Q_H(T_k)} \mathbb{E} [Q_H(T_k) F_{T_k}^2], \end{aligned}$$

where the last equality follows from the definition of  $F_T$ . By (25) the sequence  $\{\mathbb{E} [Q_H(T_k) F_{T_k}^2]\}_{k \in \mathbb{N}}$  is bounded and together with the definition of  $Q_H$  the first part of the statements follows.

For the variance of  $D_N$ , apply (21) to get

$$\begin{aligned} \text{var } D_N &= \text{var} \left( \sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt \right) = \sum_{k=1}^N \alpha_k^4 \text{var} \left( \int_0^T x_k^2(t) dt \right) \\ &= \sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^{4H}} \text{var} \left( \int_0^{T_k} \tilde{x}_1^2(\mu_k t) dt \right) \\ &= \sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^{4H+2}} \frac{1}{Q_H(T_k)} Q_H(T_k) \text{var} \left( \int_0^{T_k} \tilde{x}_1^2(t) dt \right), \end{aligned}$$

and Lemma 5.2 to obtain that the sequence  $\left\{ Q_H(T_k) \text{var} \left( \int_0^{T_k} \tilde{x}_1^2(t) dt \right) \right\}_{k \in \mathbb{N}}$  is bounded. The second parts of the statements are again concluded by the definition of  $Q_H$ .  $\square$

Combining Proposition 5.1 and Proposition 5.2 we obtain the asymptotic formulas for variances of  $\frac{C_N}{\mathbb{E} D_N}$  and  $\frac{D_N}{\mathbb{E} D_N}$ .

$$\left\{ \begin{array}{ll} H < \frac{3}{4} : & \text{var } \frac{C_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{4H-1}}}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \right)^2}, \quad \text{var } \frac{D_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^{4H+1}}}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \right)^2}, \\ H = \frac{3}{4} : & \text{var } \frac{C_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^2} \log T_k}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{3/2}} \right)^2}, \quad \text{var } \frac{D_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^4} \log T_k}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{3/2}} \right)^2}, \\ H > \frac{3}{4} : & \text{var } \frac{C_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^2}}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \right)^2}, \quad \text{var } \frac{D_N}{\mathbb{E} D_N} \asymp \frac{\sum_{k=1}^N \frac{\alpha_k^4}{\mu_k^4}}{\left( \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} \right)^2}. \end{array} \right. \quad (26)$$

Note that if for some  $c, \delta > 0$  and for all  $N \in \mathbb{N}$

$$\text{var} \frac{C_N}{\mathbb{E} D_N} \leq c N^{-\delta} \quad \text{and} \quad \text{var} \frac{D_N}{\mathbb{E} D_N} \leq c N^{-\delta} \quad (27)$$

holds true then application of Lemma A.1 ( $C_N/\mathbb{E} D_N$  and  $(D_N - \mathbb{E} D_N)/\mathbb{E} D_N$  are in the second Wiener chaos, see (41) in Appendix) provides us with

$$\frac{C_N}{\mathbb{E} D_N} \xrightarrow{N \rightarrow \infty} 0 \quad \text{and} \quad \frac{D_N}{\mathbb{E} D_N} \xrightarrow{N \rightarrow \infty} 1 \quad \text{a.s.}$$

so that the strong consistency of the estimate

$$\hat{\lambda}_N = \lambda - \frac{\frac{C_N}{\mathbb{E} D_N}}{\frac{D_N}{\mathbb{E} D_N}} \xrightarrow{N \rightarrow \infty} \lambda \quad \text{a.s.}$$

follows.

It is difficult to verify the sufficient conditions (27) in general and it should be done on case-by-case basis. However, if we adopt assumption (6), often satisfied for parabolic SPDEs, the calculations of asymptotic variances simplify substantially.

**Theorem 5.1.** *Let  $B \equiv 0$ , i.e.  $\beta_k = 0$  for all  $k \in \mathbb{N}$ , and the power growth (6) is satisfied for  $\alpha_k$ . Then  $\hat{\lambda}_N$  is strongly consistent for any  $m_1, d \in \mathbb{N}$ .*

*Proof.* Denote the growth exponent of  $\alpha_k$  by  $M_1 := 2m_1/d$ . Then using formula (26) we obtain

$$\begin{aligned} \text{var} \frac{C_N}{\mathbb{E} D_N} &\asymp \text{var} \frac{D_N}{\mathbb{E} D_N} \\ &\asymp \begin{cases} \frac{\sum_{k=1}^N k^{M_1(3-4H)}}{\left(\sum_{k=1}^N k^{2M_1(1-H)}\right)^2} \asymp N^{-(M_1+1)} & , \quad H < \frac{3}{4}, \\ \frac{\sum_{k=1}^N \log(k^{M_1} T)}{\left(\sum_{k=1}^N k^{1/2M_1}\right)^2} \asymp N^{-(M_1+1)} \log N & , \quad H = \frac{3}{4}, \\ \frac{N}{\left(\sum_{k=1}^N k^{2M_1(1-H)}\right)^2} \asymp N^{-(4M_1(1-H)+1)} & , \quad H > \frac{3}{4}. \end{cases} \quad (28) \end{aligned}$$

Thus (27) is satisfied for any  $m_1, d$ , so the conclusion follows from Lemma A.1.  $\square$

**Theorem 5.2.** *Assume the power growth of eigenvalues  $\alpha_k$  and  $\beta_k$  given in (6):*

$$\alpha_k \sim k^{2m_1/d} \quad \text{and} \quad \beta_k \sim k^{2m_2/d}.$$

(i) *If  $m_1 > m_2$  then  $\hat{\lambda}_N$  is strongly consistent for any  $m_1, m_2, d \in \mathbb{N}$ ,*

(ii) If  $m_1 < m_2$  and

$$m_1 > \begin{cases} -\frac{d}{4} + \frac{1}{2}m_2 & , \quad H < \frac{3}{4}, \\ -\frac{d}{4} + m_2(2H - 1) & , \quad H \geq \frac{3}{4}, \end{cases}$$

then  $\hat{\lambda}_N$  is strongly consistent.

*Proof.* (i) Applying again formula (26) we obtain the same estimates as in the proof of Theorem 5.1 thus the conclusion follows.

(ii) For conciseness, we consider the canonical case (all the involved series have power growth) only. The special cases, where some series in the expectation of  $D_N$  or the variances of  $C_N, D_N$  have logarithmic growth, have to be analyzed separately. Nevertheless, they are covered by (ii) as well.

For  $\text{var } \frac{D_N}{\mathbb{E} D_N}$  we get the same estimates as in the proof of Theorem 5.1 but with  $M_1$  substituted by  $M_2 := 2m_2/d$ . For  $\text{var } \frac{C_N}{\mathbb{E} D_N}$  formula (26) yields

$$\text{var } \frac{C_N}{\mathbb{E} D_N} \asymp \begin{cases} \frac{\sum_{k=1}^N k^{2M_1 - M_2(4H-1)}}{\left(\sum_{k=1}^N k^{2(M_1 - HM_2)}\right)^2} \asymp N^{-2M_1 + M_2 - 1} & , \quad H < \frac{3}{4}, \\ \frac{\sum_{k=1}^N k^{2M_1 + 2M_2} \log(k^{M_2} T)}{\left(\sum_{k=1}^N k^{2M_1 - 3/2 M_2}\right)^2} \asymp N^{-2M_1 + M_2 - 1} \log N, & H = \frac{3}{4}, \\ \frac{\sum_{k=1}^N k^{2(M_1 - M_2)}}{\left(\sum_{k=1}^N k^{2(M_1 - HM_2)}\right)^2} \asymp N^{-2(M_1 + M_2(1-2H)) - 1} & , \quad H > \frac{3}{4}. \end{cases} \quad (29)$$

These estimates give us the required relationship between  $m_1, m_2$  and  $d$  to meet (27). The rest follows from Lemma A.1.  $\square$

*Remark 5.1.* The case  $m_1 = m_2$  is reduced to the case of Theorem 5.1.

*Remark 5.2.* The expressions for  $\text{var } \frac{C_N}{\mathbb{E} D_N}$  given in formulas (28) and (29) show the speed of the convergence of  $\hat{\lambda}_N$  to  $\lambda$ .

## 5.2. Strong consistency of pathwise LSE

In this subsection, strong consistency (as  $N \rightarrow \infty$ ) of  $\tilde{\lambda}_N$  is demonstrated. We continue to denote the true value of the unknown parameter by  $\lambda$ . Recall that the theoretical LSE satisfies  $\hat{\lambda}_N = R_N(\lambda)$ . Hence, if  $\hat{\lambda}_N$  is strongly consistent (e.g. if the conditions of Theorem 5.1 or Theorem 5.2 are satisfied), we have:

$$\lim_{N \rightarrow \infty} R_N(\lambda) = \lambda.$$



To show that the pathwise LSE  $\tilde{\lambda}_N = R_N(\tilde{\lambda}_N)$  is arbitrarily close to  $\lambda$  for sufficiently large  $N$ , the limiting behavior of derivatives of  $R_N$  is studied. The following elementary observation will be useful in the sequel:

**Lemma 5.3.** *Let  $a_k, b_k \geq 0$  for  $k = 1, 2, \dots$ ,  $\lim_{k \rightarrow \infty} b_k = b \in (0, \infty)$ , and  $\sum_{k=1}^{\infty} a_k = \infty$ . Then*

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N a_k b_k}{\sum_{k=1}^N a_k b} = 1.$$

*Proof.* It is a simple consequence of the Stolz–Cesaro theorem.  $\square$

Next lemma bounds the derivatives of  $R_N$  on a compact neighborhood of  $\lambda$  from above by a constant smaller than one (for sufficiently large  $N$ ):

**Lemma 5.4.** *Let the power growth condition (6) be satisfied and  $H \neq 1/2$ . Further assume:*

$$\lim_{k \rightarrow \infty} \frac{\beta_k}{\alpha_k} = c \in [0, \infty], \quad (30)$$

$$\sum_{k=1}^{\infty} \frac{\alpha_k^2}{(\alpha_k + \beta_k)^{2H}} = \infty. \quad (31)$$

*Then, almost surely, if we take arbitrary  $0 < \Lambda_L < \Lambda_U < \infty$  we have*

$$\begin{aligned} \lim_{N \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ \frac{1}{1-2H} \frac{dR_N}{d\Lambda}(\Lambda) \right] &= \frac{(\lambda + c)^{2H}}{(\Lambda_L + c)^{2H}}, \\ \lim_{N \rightarrow \infty} \inf_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ \frac{1}{1-2H} \frac{dR_N}{d\Lambda}(\Lambda) \right] &= \frac{(\lambda + c)^{2H}}{(\Lambda_U + c)^{2H}}, \end{aligned} \quad (32)$$

*where the right-hand sides equal one if  $c = \infty$ .*

*Proof.* Recall the formula for the derivative of  $R_N$  in (15) and the notation from previous subsection  $D_N = \sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt$ . It follows from the proof of Proposition 5.1 that

$$\mathbb{E}D_N = \sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} T\varphi(\mu_k T),$$

where

$$\varphi(\tau) = \frac{1}{\tau} \int_0^\tau \mathbb{E} \tilde{x}_1^2(t) dt \xrightarrow{\tau \rightarrow \infty} H\Gamma(2H).$$

Use (15) to write (for any  $\Lambda > 0$ ):

$$\frac{dR_N}{d\Lambda}(\Lambda) = \frac{\sum_{k=1}^N \alpha_k^2 \frac{d\Delta}{d\mu}(\alpha_k \Lambda + \beta_k)}{D_N} = \frac{\mathbb{E}D_N}{D_N} \frac{\sum_{k=1}^N \alpha_k^2 \frac{d\Delta}{d\mu}(\alpha_k \Lambda + \beta_k)}{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} T\varphi(\mu_k T)}. \quad (33)$$

Formula (13) together with  $\lim_{T \rightarrow \infty} \gamma(h, T) = \Gamma(h)$  yield

$$\lim_{\mu \rightarrow \infty} \mu^{2H} \frac{d\Delta}{d\mu}(\mu) = TH(1-2H)\Gamma(2H). \quad (34)$$

Consequently, we can factorize

$$\frac{dR_N}{d\Lambda}(\Lambda) = \frac{\mathbb{E}D_N}{D_N} U_N V_N(\Lambda) W_N(\Lambda),$$

with

$$\begin{aligned} U_N &= \frac{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} TH\Gamma(2H)}{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} T\varphi(\mu_k T)}, \\ V_N(\Lambda) &= \frac{\sum_{k=1}^N \alpha_k^2 \frac{d\Delta}{d\mu}(\alpha_k \Lambda + \beta_k)}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda + \beta_k)^{2H}} TH(1-2H)\Gamma(2H)}, \\ W_N(\Lambda) &= \frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda + \beta_k)^{2H}} TH(1-2H)\Gamma(2H)}{\sum_{k=1}^N \frac{\alpha_k^2}{\mu_k^{2H}} TH\Gamma(2H)}. \end{aligned}$$

Recall that condition (6) implies (cf. proofs of Theorem 5.1 and Theorem 5.2)

$$\lim_{N \rightarrow \infty} \frac{D_N}{\mathbb{E}D_N} = 1 \quad \text{almost surely.}$$

Assumptions of Lemma 5.3 with

$$a_k = \frac{\alpha_k^2}{\mu_k^{2H}}, \quad b_k = T\varphi(\mu_k T), \quad \text{and } b = TH\Gamma(2H),$$

are guaranteed by conditions (31) and (3). In result,

$$\lim_{N \rightarrow \infty} U_N = 1.$$

Denote (motivated by Lemma 5.3)

$$a_k(\Lambda) = \frac{\alpha_k^2}{(\alpha_k \Lambda + \beta_k)^{2H}}, \quad b_k(\Lambda) = (\alpha_k \Lambda + \beta_k)^{2H} \frac{d\Delta}{d\mu}(\alpha_k \Lambda + \beta_k)$$

and

$$b = TH(1-2H)\Gamma(2H).$$

Fix arbitrary  $0 < \Lambda_L < \Lambda_U < \infty$ . Obviously, due to (3) and (34),

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |b_k(\Lambda) - b| \xrightarrow{k \rightarrow \infty} 0,$$

and, due to (31)

$$\inf_{\Lambda \in [\Lambda_L, \Lambda_U]} \sum_{k=1}^N a_k(\Lambda) = \sum_{k=1}^N a_k(\Lambda_U) \xrightarrow{N \rightarrow \infty} \infty.$$

Similarly to proof of Lemma 5.3 we can show

$$\lim_{N \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |V_N(\Lambda) - 1| = \lim_{N \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{\sum_{k=1}^N a_k(\Lambda) b_k(\Lambda)}{\sum_{k=1}^N a_k(\Lambda) b} - 1 \right| = 0.$$

For the last term we can use the upper bound:

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ \frac{W_N(\Lambda)}{1 - 2H} \right] = \frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda_L + \beta_k)^{2H}}}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \lambda + \beta_k)^{2H}}}.$$

Assume first  $\lim_{k \rightarrow \infty} \frac{\beta_k}{\alpha_k} = c \in [0, \infty)$ . Then

$$\frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda_L + \beta_k)^{2H}}}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \lambda + \beta_k)^{2H}}} = \frac{\sum_{k=1}^N \alpha_k^{2-2H} (\Lambda_L + \frac{\beta_k}{\alpha_k})^{-2H}}{\sum_{k=1}^N \alpha_k^{2-2H} (\lambda + \frac{\beta_k}{\alpha_k})^{-2H}}.$$

Application of Lemma 5.3 with

$$a_k = \alpha_k^{2-2H}, \quad b_k = \left( \Lambda + \frac{\beta_k}{\alpha_k} \right)^{-2H}, \quad b = (\Lambda + c)^{-2H},$$

and  $\Lambda = \Lambda_L$  or  $\Lambda = \lambda$ , respectively, and where

$$\sum_{k=1}^{\infty} \alpha_k^{2-2H} \geq \sum_{k=1}^{\infty} \frac{\alpha_k^2}{(\alpha_k + \beta_k)^{2H}} = \infty,$$

results in

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda_L + \beta_k)^{2H}}}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \lambda + \beta_k)^{2H}}} = \lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \alpha_k^{2-2H} (\Lambda_L + c)^{-2H}}{\sum_{k=1}^N \alpha_k^{2-2H} (\lambda + c)^{-2H}} = \frac{(\lambda + c)^{2H}}{(\Lambda_L + c)^{2H}}.$$

Next, assume instead  $c = \infty$ , i.e.  $\lim_{k \rightarrow \infty} \frac{\alpha_k}{\beta_k} = 0$ . We can proceed similarly to previous case

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda_L + \beta_k)^{2H}}}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \lambda + \beta_k)^{2H}}} = \lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \frac{\alpha_k^2}{\beta_k^{2H}} \left( \frac{\alpha_k}{\beta_k} \Lambda_L + 1 \right)^{-2H}}{\sum_{k=1}^N \frac{\alpha_k^2}{\beta_k^{2H}} \left( \frac{\alpha_k}{\beta_k} \lambda + 1 \right)^{-2H}} = 1.$$

To summarize, we have

$$\lim_{N \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ \frac{W_N(\Lambda)}{1 - 2H} \right] = \begin{cases} \frac{(\lambda + c)^{2H}}{(\Lambda_L + c)^{2H}} & \text{if } c \in [0, \infty), \\ 1 & \text{if } c = \infty. \end{cases}$$

The limit for infimum can be obtained accordingly, since

$$\inf_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ \frac{W_N(\Lambda)}{1 - 2H} \right] = \frac{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \Lambda_U + \beta_k)^{2H}}}{\sum_{k=1}^N \frac{\alpha_k^2}{(\alpha_k \lambda + \beta_k)^{2H}}}.$$

□

**Theorem 5.3.** *Assume that the eigenvalues  $\alpha_k$  and  $\beta_k$  satisfy the power growth conditions (6) and the asymptotic conditions (30), (31) and (3). Further assume that the theoretical LSE  $\hat{\lambda}_N$  is strongly consistent (see e.g. Theorem 5.1 and Theorem 5.2). Then the pathwise LSE is strongly consistent as well:*

$$\lim_{N \rightarrow \infty} \tilde{\lambda}_N = \lambda \quad \text{almost surely.} \quad (35)$$

Moreover,  $\tilde{\lambda}_N$  and  $\hat{\lambda}_N$  have the same speed of convergence (up to a constant):

$$\lim_{N \rightarrow \infty} \frac{|\lambda - \hat{\lambda}_N|}{|\lambda - \tilde{\lambda}_N|} = 2H \quad \text{almost surely.} \quad (36)$$

*Proof.* Recall that if  $H = 1/2$ , the theorem holds true trivially, because  $\tilde{\lambda}_N = \hat{\lambda}_N$  in this case. In the rest of the proof, consider  $H \neq 1/2$ . First, take  $\Lambda_L < \lambda$  so that

$$(1 - 2H) \frac{(\lambda + c)^{2H}}{(\Lambda_L + c)^{2H}} < 1.$$

and arbitrary  $\Lambda_U > \lambda$ . Lemma 5.4 guarantees that, for almost all  $\omega$  and some  $\Psi > 0$ ,

$$\inf_{\Lambda \in [\Lambda_L, \Lambda_U]} \left[ 1 - \frac{dR_N}{d\Lambda}(\Lambda) \right] > \Psi,$$

for sufficiently large  $N$ . Fix  $\epsilon > 0$  such that  $\Lambda_L < \lambda - \epsilon < \lambda + \epsilon < \Lambda_U$  and denote  $\delta = \epsilon\Psi$ . The almost sure convergence  $\lim_{N \rightarrow \infty} R_N(\lambda) = \lambda$ , which is guaranteed by strong consistency of the theoretical LSE, enables us to take  $N$  large enough in order to

$$|\lambda - R_N(\lambda)| < \delta.$$

Combination of the lower bound on derivative of  $\Lambda - R_N(\Lambda)$  and its closeness to zero at  $\lambda$  implies existence of its root inside  $(\lambda - \epsilon, \lambda + \epsilon)$ , which is unique therein and it equals  $\tilde{\lambda}_N$ . This proves (35).

To demonstrate (36), note that for sufficiently large  $N$

$$\frac{|\lambda - \hat{\lambda}_N|}{|\lambda - \tilde{\lambda}_N|} = \left| 1 - \frac{dR_N}{d\Lambda}(\Lambda_N^0) \right|,$$

where  $\Lambda_N^0$  is between  $\tilde{\lambda}_N$  and  $\lambda$  and therefore converges to  $\lambda$  almost surely. Next, fix arbitrary  $\epsilon > 0$  and take  $\Lambda_L < \lambda < \Lambda_U$  so that the right-hand sides of (32) are close enough to one and

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{dR_N}{d\Lambda}(\Lambda) - (1 - 2H) \right| < \epsilon,$$

for sufficiently large  $N$ . In particular, for  $N$  large enough we have almost surely

$$\left| \frac{dR_N}{d\Lambda}(\Lambda_N^0) - (1 - 2H) \right| < \epsilon.$$

This proves the desired almost sure convergence

$$\lim_{N \rightarrow \infty} \frac{|\lambda - \hat{\lambda}_N|}{|\lambda - \tilde{\lambda}_N|} = \lim_{N \rightarrow \infty} \left| 1 - \frac{dR_N}{d\Lambda}(\Lambda_N^0) \right| = |1 - (1 - 2H)| = 2H. \quad \square$$

*Remark 5.3.* Immediate, and possibly surprising, consequence of (36) is that the pathwise LSE performs asymptotically better than the theoretical LSE whenever  $H > 1/2$ .

### 5.3. Example

As an example of the use of Theorem 5.3, consider a stochastic heat equation with additional mean reversion, with a distributed noise autocorrelated in time and with the Dirichlet boundary condition. Formally, we can write this equation as

$$\begin{aligned} \frac{\partial u}{\partial t}(t, \xi) &= \lambda_1 \Delta_\xi u(t, \xi) - \lambda_2 u(t, \xi) + \eta^H(t, \xi), \quad \text{for } (t, \xi) \in \mathbb{R}_+ \times \mathcal{O}, \\ u(t, \xi) &= 0, \quad \text{for } (t, \xi) \in \mathbb{R}_+ \times \partial\mathcal{O}, \\ u(0, \xi) &= 0, \quad \text{for } \xi \in \mathcal{O}, \end{aligned} \quad (37)$$

where  $\Delta_\xi$  is Laplace operator in variable  $\xi$ ,  $\mathcal{O} \subset \mathbb{R}^d$  is a  $d$ -dimensional bounded domain with smooth boundary  $\partial\mathcal{O}$ ,  $\lambda_1 \geq 0$  is the diffusivity parameter determining the speed of diffusion of a modeled substance along the domain  $\mathcal{O}$ ,  $\lambda_2 \geq 0$  is the rate of mean reversion related to the speed of reversion of the amount of substance to the zero mean. The process  $\{\eta^H(t, \xi); t \geq 0, \xi \in \mathcal{O}\}$  is a Gaussian noise, which is fractional (autocorrelated) in time with Hurst parameter  $H \in (0, 1)$  and white in space. Formally it can be modelled as the differential of a cylindrical fractional Brownian motion  $dB^H(t)$  (see Appendix B for details).

This formal equation can be rewritten rigorously as a stochastic evolution equation (1) in  $V = L^2(\mathcal{O})$ , with  $A = \Delta|_{\text{Dom}(A)}$ , where  $\text{Dom}(A) = H^2(\mathcal{O}) \cap H_0^1(\mathcal{O})$ , being Dirichlet Laplace operator defined on a standard Sobolev space (cf. [31]) and with  $B$  being the identity operator on  $L^2(\mathcal{O})$ , should  $\lambda_1$  be the estimated parameter (or vice-versa, should we estimate  $\lambda_2$ ). Consequently, this equation is diagonalizable with eigenvalues satisfying the power growth condition (6).

Firstly, assume we want to estimate the diffusivity parameter  $\lambda_1$ . If  $\lambda_2 = 0$ , we have a stochastic heat equation with  $\alpha_k \sim k^{2/d}$ , and  $\beta_k = 0$ . Theorem 5.1 guarantees strong consistency of the theoretical LSE  $\hat{\lambda}_N$ . If  $\lambda_2 > 0$ , there is an additional mean reversion in the process and we have  $\alpha_k \sim k^{2/d}$ , and  $\beta_k \sim 1$ . We can apply Theorem 5.2 (i) to get strong consistency of the theoretical LSE  $\hat{\lambda}_N$ . In both cases, Theorem 5.3 implies strong consistency of the pathwise LSE

$\tilde{\lambda}_N$  with the speed of convergence (in terms of RMSE) given by

$$\sqrt{\mathbb{E}(\tilde{\lambda}_N - \lambda)^2} \asymp \begin{cases} N^{-(1/d+1/2)} & , \quad H < \frac{3}{4}, \\ N^{-(1/d+1/2)} \sqrt{\log N} & , \quad H = \frac{3}{4}, \\ N^{-(4(1-H)/d+1/2)} & , \quad H > \frac{3}{4}. \end{cases} \quad (38)$$

Secondly, let  $\lambda_2$  be the parameter of interest. If  $\lambda_1 = 0$  we have  $\alpha_k \sim 1$  and  $\beta_k = 0$ , and the existence condition (2) is violated. Hence, we consider the case  $\lambda_1 > 0$ , when the process combines diffusion and mean reversion. For the eigenvalues, we have  $\alpha_k \sim 1$ , and  $\beta_k \sim k^{2/d}$ . Theorem 5.2 (ii) provides us with strong consistency of  $\tilde{\lambda}_N$  if  $d > 2$  for  $H < 3/4$  and  $d > 4(2H - 1)$  otherwise. To show strong consistency of the pathwise LSE by Theorem 5.3, we need in addition that  $d \geq 4H$  so that (31) is satisfied. If the conditions for consistency are satisfied, the speed of convergence is

$$\sqrt{\mathbb{E}(\tilde{\lambda}_N - \lambda)^2} \asymp \begin{cases} N^{1/d-1/2} & , \quad H < \frac{3}{4}, \\ N^{1/d-1/2} \sqrt{\log N} & , \quad H = \frac{3}{4}, \\ N^{(4H-2)/d-1/2} & , \quad H > \frac{3}{4}. \end{cases} \quad (39)$$

## 6. Simulation study

To illustrate the actual performance of the studied pathwise LSE we perform a Monte Carlo analysis for two specific equations – a 1D heat equation (the simplest setting) and a 2D heat equation with coupling with surroundings (a model popular e.g. in oceanography). All simulations were performed in statistical software R.

### 6.1. 1D stochastic heat equation

This equation has very simple structure and serves as an easy-to-understand toy-model suitable for illustration of the theory above. We simulate a stochastic heat equation with zero boundary condition from subsection 5.3 on the line segment  $(0,1)$ , with true diffusivity  $\lambda_1 = 1$  being a parameter to be estimated ( $\lambda = \lambda_1$ ) and without additional mean reversion ( $\lambda_2 = 0$ ). In particular, we set

- $d = 1$ ,  $\mathcal{O} = (0, 1)$  and time horizon  $T = 1$ ,
- $(Af)(\xi) = \frac{\partial^2 f}{\partial \xi^2}$ ,  $\xi \in (0, 1)$ ,  $B \equiv 0$ ,
- $e_k(\xi) = \sin(k\pi\xi)$ ,  $\alpha_k = (k\pi)^2$ ,  $\beta_k = 0 \quad \forall k \in \mathbb{N}$ ,
- two scenarios for Hurst index:  $H \in \{0.2, 0.8\}$ ,
- two scenarios for initial condition: (i)  $x_k(0) = 0$  for all  $k \geq 1$  (zero IC), and (ii)  $x_1(0) = 10$ ,  $x_2(0) = 5$ ,  $x_3(0) = 2$ ,  $x_k(0) = 0$ ,  $k \geq 4$  (non-zero IC).

The observed Fourier modes take the following form

$$x_k(t) = \int_0^1 u(t, \xi) \sin(k\pi\xi) d\xi.$$

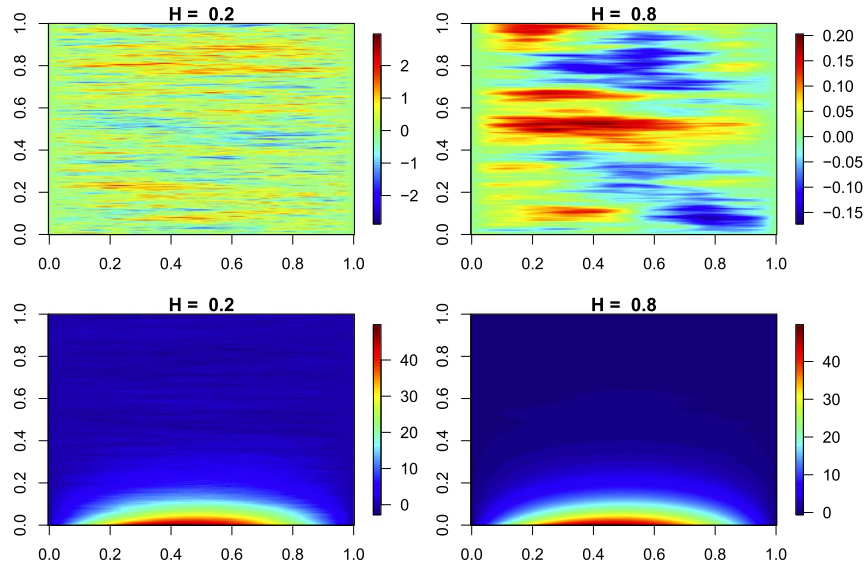


FIG 1. Heat maps of sample solutions of 1D stochastic heat equation with zero (the upper row) and non-zero (the lower row) initial conditions. Note different colour scales in the two rows. Horizontal axis represents the physical space domain and vertical axis the time domain.

For simulations we used the Spectral Galerkin Method (see [20], chapter 10.7.), which is very efficient in our setting (space discretization provides system of independent linear stochastic ODEs, which, in addition, can be solved explicitly). Heat maps of four simulated sample solutions for various scenarios are shown in Figure 1.

We test performance of the pathwise LSE, defined in (17), and compare it with the weighted MCE, an alternative estimator defined and studied in [17], cf. formulas (3.21)–(3.23) and Theorem 3.2 therein. Recall that the weighted MCE is strongly consistent and asymptotically normal with increasing number of observed modes.

Simulation results were obtained from 200 runs for each scenarios producing sample of 200 estimates for each setting. Figure 2 illustrates the convergence of the estimates to the true value of the parameter with increasing number of Fourier modes by showing selected sample quantiles. It confirms the convergence for all settings. In addition, non-zero initial condition significantly improves the performance of the pathwise LSE. Figure 3 shows convergence of the root mean square error (RMSE) for samples of the two types of estimators (pathwise LSE and weighted MCE) and its comparison with the theoretical speed of convergence for RMSE from (38) (applicable only for solutions with zero initial condition). In case of zero initial condition, RMSE for the pathwise LSE and the weighted MCE practically coincide and their speed of convergence is consistent with the theoretical one. For solutions with far-from-zero initial conditions RMSE for the pathwise LSE is close to zero even for small number of

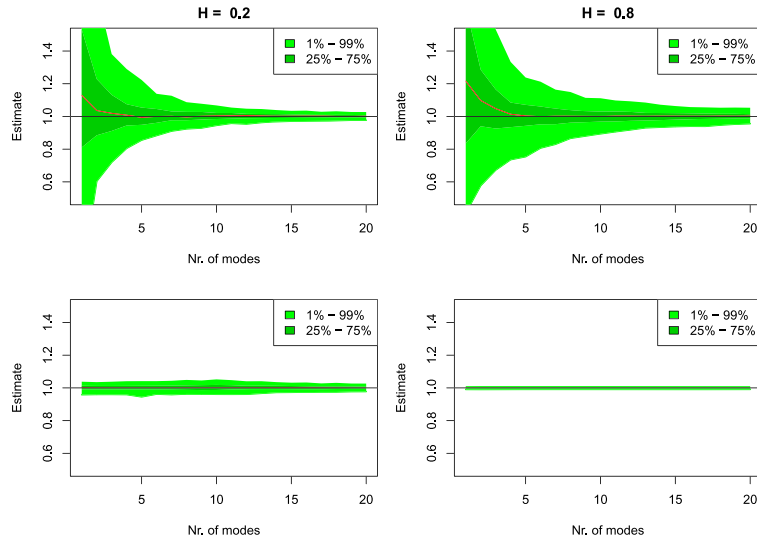


FIG 2. Sample quantile ranges (green areas) and median (red line) of the simulated estimates (200 samples), true value is 1. Left column: Hurst index = 0.2 (rough case); right column: Hurst index = 0.8 (smooth case). First row: solutions with zero init. cond.; second row: solutions with non-zero init. cond.

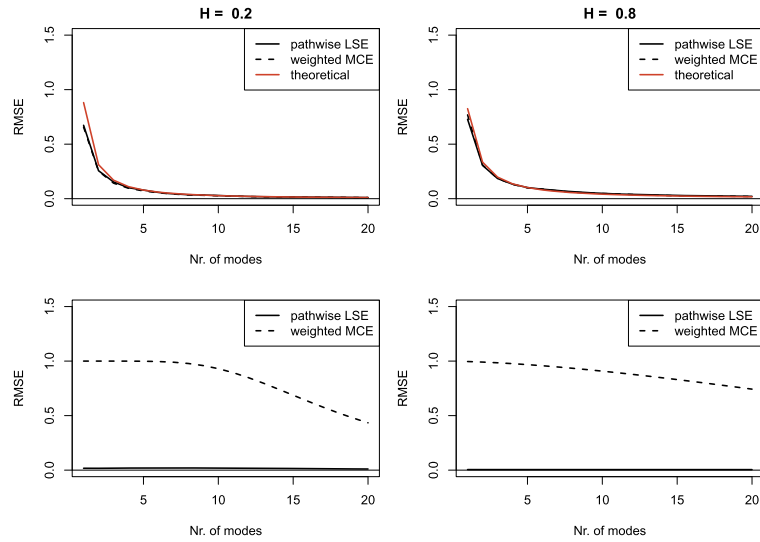


FIG 3. Root mean square error (RMSE) for pathwise LSE and weighted MCE and its theoretical speed of convergence from (38) (applicable only for zero initial condition). Left column: Hurst index = 0.2 (rough case); right column: Hurst index = 0.8 (smooth case). First row: solutions with zero init. cond.; second row: solutions with non-zero init. cond.



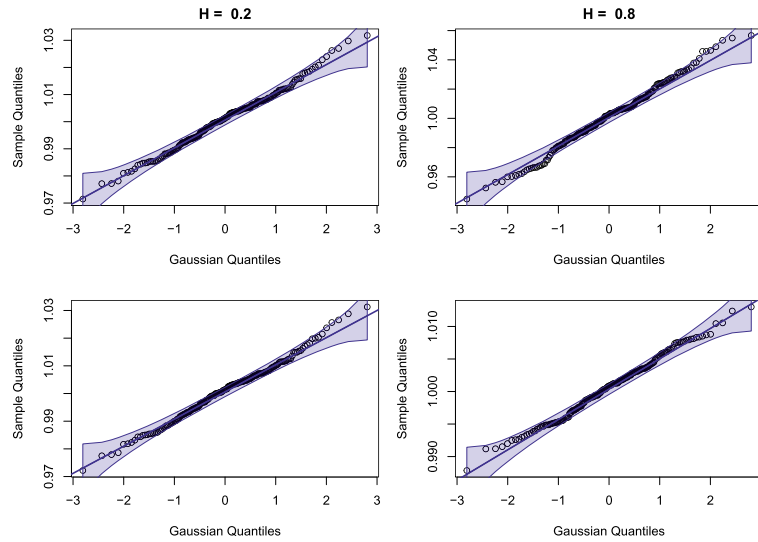


FIG 4.  $Q$ - $Q$  plot for sample of 200 pathwise LSE estimates (calculated from 40 Fourier modes) including 95% confidence envelope for normal distribution. Left column: Hurst index = 0.2 (rough case); right column: Hurst index = 0.8 (smooth case). First row: solutions with zero init. cond.; second row: solutions with non-zero init. cond.

Fourier modes, whereas the weighted MCE converges significantly slower. Figure 4 shows comparison of sample quantiles of pathwise LSE (40 Fourier modes considered) with quantiles of normal distribution including 95% confidence envelope. Although not proved theoretically, these  $Q$ - $Q$  plots suggest asymptotic normality of the pathwise LSE for all settings considered.

## 6.2. 2D stochastic heat equation with coupling with surroundings

Equations of this type are popular (not only) in physical oceanography, where they model fluctuations of the temperature of the top layer of the ocean around its long-time average, see [28] or [21]. These equations fall into the setting of Example in 5.3 with  $d = 2$  (a two-dimensional domain being the surface of the ocean), where the 2D Laplace operator models thermal conduction along the water surface, the mean-reversion term reflects atmosphere and ocean coupling and the noise term represents random forcing (e.g. vertical heat fluxes due to turbulent environment) – for more details see [28]. The authors in [28] and [21] (Section 6.1.1) discuss estimation of parameters in these equations based on observations in spectral domain under the assumption that the noise process is white in time and space (generated by a standard cylindrical Brownian motion). In contrast, we study parameter estimation assuming more general noise, which can be autocorrelated in time (generated by a cylindrical fractional Brownian motion).

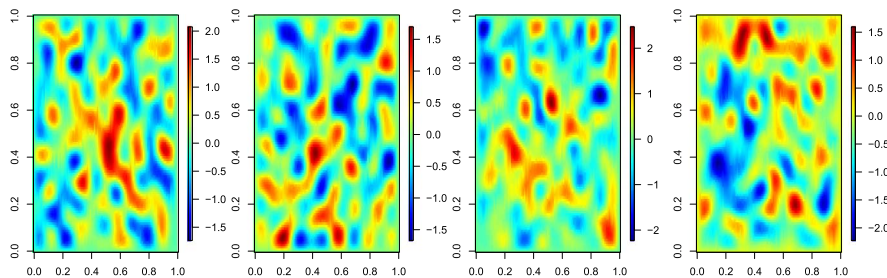


FIG 5. Heat maps of snapshots of a sample solution of 2D stochastic heat equation with mean reversion on a rectangular physical space domain taken at four time instants (from left to right:  $t=0.2, 0.4, 0.6$  and  $0.8$ ).

In particular, we simulate equation (37) with the following setting:

- $d = 2$ ,  $\mathcal{O} = (0, 1)^2$  and time horizon  $T = 1$ ,
- $\lambda_1 = 1$  (the parameter to be estimated),  $\lambda_2 = 1$ ,
- Hurst index  $H = 0.3$ , zero initial condition.

We used again the Spectral Galerkin Method with eigenfunctions  $\sin(k_1\pi\xi_1) \times \sin(k_2\pi\xi_2)$  and corresponding eigenvalues  $1 + (k_1^2 + k_2^2)\pi^2$ , where  $k_1, k_2 = 0, 1, 2, \dots$  and  $(\xi_1, \xi_2) \in (0, 1)^2$ .

The simulated distribution of the heat along the water surface at specific time instants is shown in Figure 5. Results of the Monte Carlo study of the behavior of the pathwise LSE obtained from 200 runs are depicted in Figure 6. They confirm our theoretical findings on the almost-sure convergence of the pathwise LSE and its speed of convergence (in terms of RMSE). Moreover, Q-Q plot indicate its asymptotic normality, although we do not provide theoretical reasoning (for conciseness).

### 6.3. Discussion

Results of the simulations confirm our findings on strong consistency of the pathwise LSE and on the speed of the convergence. In case of zero initial condition, it behaves similarly to the weighted MCE. However, for non-zero initial condition, pathwise LSE significantly outperforms the weighted MCE. This is in accordance with similar findings from finite-dimensional models (cf. [18] or [19]), where LSE-type estimators improve with far-from-stationary initial conditions (the drift part dominates the noise part in the dynamics of the process in such case), whereas the ergodic-type estimators (such as the weighted MCE) reflecting long-term stationary behavior are ruined.

The theory for the weighted MCE, developed in [17], covers only single-operator equations, so it is not directly applicable for the second example (2D stochastic heat equation with coupling with surroundings). To our best knowledge, the pathwise LSE is the first efficient tool for parameter estimation in this

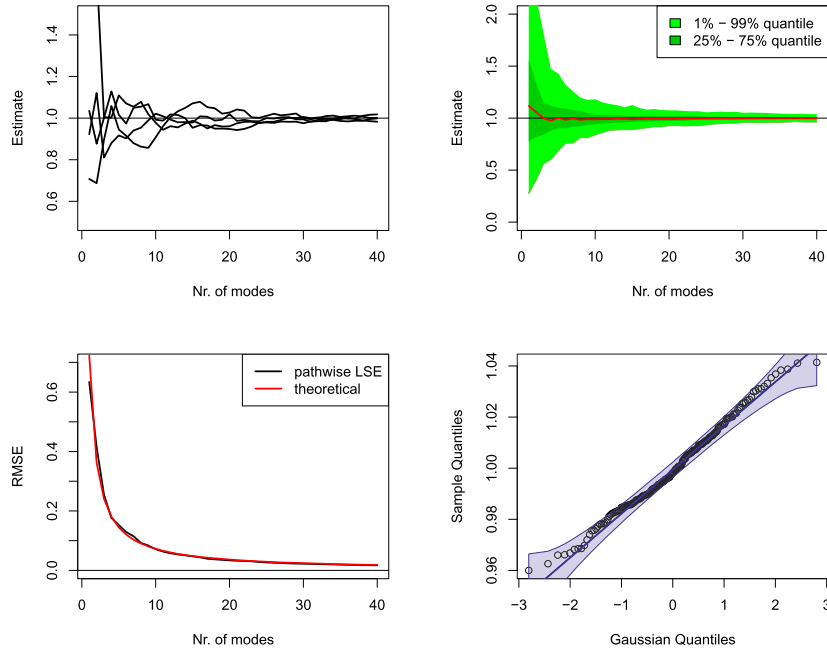


FIG 6. Characteristics of the convergence of pathwise LSE for 2D heat equation with mean reversion. 200 runs, Zero IC and  $H=0.3$ . Top-left: Four trajectories of estimates calculated from four different samples (realizations). Top-right: Sample quantile ranges (green areas) and median (red line), true value is 1. Bottom-left: Convergence of sample RMSE and theoretical RMSE, see (38). Bottom-right: Q-Q plot for sample of estimates (from 40 Fourier modes) including 95% confidence envelope for normal distribution.

equation (important in oceanography, for example) when the noise is generated by a cylindrical fractional Brownian motion with general Hurst parameter  $0 < H < 1$ .

Although we did not address the asymptotic normality of the pathwise LSE in this paper (to keep it concise), we conjecture that it holds. Our conjecture is based on simulation results and asymptotic similarity with the weighted MCE, whose asymptotic normality has been proven in [17].

## 7. Conclusions

Using the least-squares formalism we have derived a spectral version of the least-squares estimator of the drift parameter in linear SPDEs with additive fractional noise for arbitrary  $H \in (0, 1)$ . This estimator is strongly consistent in space, however, it can not be directly implemented pathwise, since it contains divergence-type stochastic integral w.r.t. a fractional process. This fact prevents its simulations and its use in practice. We addressed this issue by eliminating the stochastic integral getting a novel pathwise LSE. This estimator is constructed

trully pathwise and, moreover, is robust to minor perturbations of the observed process (it does not contain differentiation of the observed trajectory), which makes it particularly useful in practice.

Using the standard tools from classical probability and analysis, as well as modern tools from Malliavin calculus, we demonstrated that the newly developed pathwise LSE is strongly consistent in space and we conjecture that it is also asymptotically normal. Simulations reveal that it can efficiently utilize information about the unknown parameter from both non-stationary and stationary parts to the observed process.

The presented pathwise least-squares estimation procedure appears to be promising approach for various equations driven by fractional Brownian motion. The potentially interesting directions of further research may include (but are not limited to) proving asymptotic normality, studying different observation/sampling schemes, different asymptotic regimes, or generalization to semi-linear stochastic equations driven by fractional Brownian motion.

## Appendix A: Elements from Malliavin calculus

In the Appendix, we describe some basic constructions from Malliavin calculus in the setting of our problem. These constructions form a key ingredient for the study of asymptotic behavior of our estimators.

### A.1. Construction of an infinite-dimensional isonormal Gaussian process

Let  $W^{(k)} = \{W^{(k)}(h^{(k)}), h^{(k)} \in \mathcal{H}^{(k)}\}$  be an isonormal Gaussian process generated by a fractional Brownian motion  $\{B_k^H(t), t \geq 0\}$  for any  $k \in \mathbb{N}$  and defined over a separable Hilbert space  $\mathcal{H}^{(k)}$ . The processes  $W^{(k)}$  are mutually independent.

Define

$$\mathcal{H} = \bigoplus_{n=1}^{\infty} \mathcal{H}^{(n)}$$

as a direct sum of  $\mathcal{H}^{(n)}$ 's, i.e.

$$\mathcal{H} = \{(h^{(1)}, h^{(2)}, \dots), h^{(n)} \in \mathcal{H}^{(n)} \wedge \sum_{n=1}^{\infty} \|h^{(n)}\|_{\mathcal{H}^{(n)}}^2 < \infty\},$$

with the scalar product

$$\langle (h^{(1)}, h^{(2)}, \dots), (g^{(1)}, g^{(2)}, \dots) \rangle_{\mathcal{H}} = \sum_{n=1}^{\infty} \langle h^{(n)}, g^{(n)} \rangle_{\mathcal{H}^{(n)}}.$$

Then  $\mathcal{H}$  is a separable Hilbert space.

For  $h = (h^{(1)}, h^{(2)}, \dots) \in \mathcal{H}$  define

$$W(h) = \lim_{N \rightarrow \infty} \sum_{n=1}^N W^{(n)}(h^{(n)}).$$

Note that this limit, understood in  $L^2$ -sense, is well-defined because

$$\mathbb{E} \left( \sum_{n=N}^{\infty} W^{(n)}(h^{(n)}) \right)^2 = \sum_{n=N}^{\infty} \mathbb{E} (W^{(n)}(h^{(n)}))^2 = \sum_{n=N}^{\infty} \|h^{(n)}\|_{\mathcal{H}^{(n)}}^2 \xrightarrow{N \rightarrow \infty} 0.$$

Since

$$\begin{aligned} \mathbb{E} (W(h)W(g)) &= \mathbb{E} \left( \sum_{n=1}^{\infty} W^{(n)}(h^{(n)}) \right) \left( \sum_{n=1}^{\infty} W^{(n)}(g^{(n)}) \right) \\ &= \sum_{n=1}^{\infty} \mathbb{E} (W^{(n)}(h^{(n)})W^{(n)}(g^{(n)})) = \sum_{n=1}^{\infty} \langle h^{(n)}, g^{(n)} \rangle_{\mathcal{H}^{(n)}} = \langle h, g \rangle_{\mathcal{H}} \end{aligned}$$

for any  $h = (h^{(1)}, h^{(2)}, \dots), g = (g^{(1)}, g^{(2)}, \dots) \in \mathcal{H}$  the process  $\{W(h), h \in \mathcal{H}\}$  is an isonormal Gaussian.

Denote by  $CH_n^{(k)}$  the  $n$ -th Wiener chaos associated with  $W^{(k)}$  for any  $k \in \mathbb{N}$ , i.e.  $CH_n^{(k)}$  is the closed (in  $L^2(\Omega)$ ) linear subspace generated by

$$\{H_n(W^{(k)}(h^{(k)})), \|h^{(k)}\|_{\mathcal{H}^{(k)}} = 1\},$$

where  $H_n$  is the  $n$ -th Hermite polynomial. Similarly, let  $CH_n$  be the  $n$ -th Wiener chaos associated with  $W$ , i.e.  $CH_n$  is the closed linear subspace generated by

$$\{H_n(W(h)), \|h\|_{\mathcal{H}} = 1\}.$$

Fix  $k \in \mathbb{N}$ , take arbitrary  $h^{(k)} \in \mathcal{H}^{(k)}$  such that  $\|h^{(k)}\|_{\mathcal{H}^{(k)}} = 1$  and define

$$h = (0, \dots, 0, \underset{\uparrow k\text{th}}{h^{(k)}}, 0, \dots, 0).$$

Then

$$\|h\|_{\mathcal{H}}^2 = \|h^{(k)}\|_{\mathcal{H}^{(k)}}^2 = 1$$

and

$$W(h) = \sum_{n=1}^{\infty} W^{(n)}(h^{(n)}) = W^{(k)}(h^{(k)}).$$

Consider  $X = H_n(W^{(k)}(h^{(k)})) \in CH_n^{(k)}$ . Then  $X = H_n(W(h))$  and  $X \in CH_n$ . By the linearity and  $L^2(\Omega)$ -closedness of  $CH_n$  it follows that

$$CH_n^{(k)} \subset CH_n \quad \forall k \in \mathbb{N}. \quad (40)$$

### A.2. Selected functionals of the solution in the second chaos

It can be shown (see e.g. [14]) that

$$\int_0^T x_k(t) dB_k^H(t) \in CH_2^{(k)}, \quad \int_0^T x_k^2(t) dt - \mathbb{E} \left[ \int_0^T x_k^2(t) dt \right] \in CH_2^{(k)}.$$

Using the construction of the infinite-dimensional isonormal Gaussian process above and the chaos embedding (40), we easily get for

$$C_N = \sum_{k=1}^N \alpha_k \int_0^T x_k(t) dB_k^H(t),$$

and

$$D_N = \sum_{k=1}^N \alpha_k^2 \int_0^T x_k^2(t) dt,$$

that

$$C_N \in CH_2 \quad \text{and} \quad D_N - \mathbb{E} D_N \in CH_2 \quad \forall N \in \mathbb{N}. \quad (41)$$

### A.3. Almost sure convergence on a fixed chaos

The following lemma is an important tool to demonstrate strong consistency of our estimators. It turns the  $L^2$ -convergence of a sequence of variables on a fixed chaos into the almost sure convergence. This idea, based on the combination of hypercontractivity with Borel–Cantelli 0-1 law, has already been used before, but we decided to formulate it as a separate lemma, because we believe it is interesting on its own.

**Lemma A.1.** *Let  $F_N, N \in \mathbb{N}$ , be random variables in a fixed Wiener chaos of an isonormal Gaussian process such that for some constants  $c > 0$  and  $\delta > 0$  the inequality*

$$\mathbb{E} F_N^2 \leq \frac{c}{N^\delta}$$

*holds for any  $N \in \mathbb{N}$ . Then  $F_N \xrightarrow[N \rightarrow \infty]{} 0$  a.s.*

*Proof.* Consider parameters  $\zeta$  and  $\eta$  so that  $0 < \zeta < \delta/2$  and  $\eta > \frac{1}{\delta/2 - \zeta}$ . Denote by  $C$  a positive constant (independent of  $N$ ), which may change from line to line and calculate

$$\mathbb{P}(|F_N| > N^{-\zeta}) \leq \frac{\mathbb{E}|F_N|^\eta}{N^{-\zeta\eta}} \leq C \frac{(\mathbb{E} F_N^2)^{\eta/2}}{N^{\zeta\eta}} \leq C \frac{1}{N^{\eta(\delta/2 - \zeta)}},$$

where Chebyshev's inequality and hypercontractivity property on the fixed Wiener chaos (see e.g. [24], Theorem 2.7.2) were used. Application of Borel–Cantelli lemma yields the almost sure convergence.  $\square$

## Appendix B: Cylindrical fractional Brownian motion and fractional Gaussian noise

The use of cylindrical fractional Brownian motion as a rigorous model for fractional Gaussian noise  $\{\eta^H(t, \xi); t > 0, \xi \in \mathcal{O}\}$  from equation (37) can be found in several publications. More detailed exposition can be found e.g. in [10]. In this Appendix, we briefly outline the connection between the two objects.

Let us begin with white noise in spatial dimension  $\mathcal{O} \subset \mathbb{R}^d$  (no time coordinate considered). A natural space associated with (37) is  $V = L^2(\mathcal{O})$  with an orthonormal basis  $\{e_k\}_{k=1}^\infty$ . Take a sequence of i.i.d. real-valued random variables  $\{U_k\}_{k=1}^\infty$  having standard Gaussian distribution. Now we can formally model the white noise in space as

$$\eta(\xi) = \sum_{k=1}^{\infty} e_k(\xi) U_k, \quad \xi \in \mathcal{O}. \quad (42)$$

Indeed, take arbitrary  $f, g \in L^2(\mathcal{O})$  and calculate (again formally)

$$\int_{\mathcal{O}} f(\xi) \eta(\xi) \, d\xi = \sum_{k=1}^{\infty} U_k \left( \int_{\mathcal{O}} f(\xi) e_k(\xi) \, d\xi \right),$$

which is a centered Gaussian random variable with variance  $\|f\|_V^2$ , and

$$\begin{aligned} \mathbb{E} \left( \int_{\mathcal{O}} f(\xi) \eta(\xi) \, d\xi \right) \left( \int_{\mathcal{O}} g(\xi) \eta(\xi) \, d\xi \right) &= \mathbb{E} \left( \sum_{k=1}^{\infty} U_k \langle f, e_k \rangle_V \right) \left( \sum_{k=1}^{\infty} U_k \langle g, e_k \rangle_V \right) \\ &= \sum_{k=1}^{\infty} \langle f, e_k \rangle_V \langle g, e_k \rangle_V = \int_{\mathcal{O}} f(\xi) g(\xi) \, d\xi. \end{aligned}$$

In particular, by taking  $f$  and  $g$  the indicator functions, we get that the noise is uncorrelated on arbitrary two disjoint subsets of  $\mathcal{O}$  and it has constant intensity (variance is proportional to the volume of the subset).

*Remark B.1.* Although the series in (42) does not converge in the  $L^2(\Omega)$  sense in the space  $V$ , it does converge in some larger Hilbert space  $\tilde{V}$  so that there exists a Hilbert-Schmidt embedding of  $V$  into  $\tilde{V}$  (this corresponds to the fact that white noise can not be considered as a proper function, but rather as a distribution).

Continue with modeling of the fractional (autocorrelated) noise in time domain. Recall that real-valued fractional Brownian motion  $\{b^H(t), t \geq 0\}$  with Hurst index  $H \in (0, 1)$  is a centered Gaussian process with continuous trajectories starting from zero and with autocovariance function

$$\mathbb{E} b^H(t) b^H(s) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}).$$

As a consequence, it has stationary Gaussian increments. If  $H = 1/2$  we obtain standard Wiener process (with independent increments), if  $H < 1/2$  we have

negatively correlated increments and for  $H > 1/2$  we have positively correlated increments. Formal derivative of fBm (which exists only in distributional sense) thus represents a good model for fractional noise in time.

To conclude, take the family of independent real-valued fractional Brownian motions  $\{B_k^H(t), t \geq 0\}, k = 1, 2, \dots$  and construct the cylindrical fBm on  $V = L^2(\mathcal{O})$  using an orthonormal basis  $\{e_k\}_{k=1}^\infty$  as a formal sum

$$B^H(t) = \sum_{k=1}^{\infty} e_k B_k^H(t).$$

Increments (differentials) of this cylindrical fBm can serve as a model for the noise that is white in space and fractional in time. Indeed, for a fixed  $0 < s < t$  we have

$$B^H(t) - B^H(s) = \sum_{k=1}^{\infty} e_k (B_k^H(t) - B_k^H(s)),$$

which is the (scaled) white-in-space noise. Next, for a fixed measurable  $D \subset \mathcal{O}$  with the finite volume take

$$\begin{aligned} \langle B^H(t), 1_D \rangle_V &= \sum_{k=1}^{\infty} \left( \int_{\mathcal{O}} 1_D(\xi) e_k(\xi) \, d\xi \right) B_k^H(t) \\ &= \sum_{k=1}^{\infty} w_k B_k^H(t) = \beta^H(t), \end{aligned}$$

which itself is a (scaled) fractional Brownian motion, since  $\sum_{k=1}^{\infty} w_k^2 = \|1_D\|_V^2 < \infty$  and so it represents an integrated fractional-in-time noise.

## Acknowledgments

We would like to thank to the anonymous referee for valuable comments and suggestions that helped to improve clarity and readability of this paper.

## References

- [1] R. Altmeyer, T. Bretschneider, J. Janák and M. Reiß. *Parameter Estimation in an SPDE Model for Cell Repolarisation*. Preprint arxiv:[2010.06340](#), 2020. [MR4376300](#)
- [2] R. Altmeyer and M. Reiß. *Nonparametric estimation for linear SPDEs from local measurements*. Ann. Appl. Probab. **31** (2021), no. 1, 1–38. [MR4254472](#)
- [3] R. Altmeyer, I. Cialenco and G. Pasemann. *Parameter estimation for semi-linear SPDEs from local measurements*. Preprint arXiv:[2004.14728v2](#), 2020.
- [4] C. Berzin, A. Latour and J. R. León. *Inference on the Hurst parameter and variance of diffusions driven by fractional Brownian motion*, Lectures notes in statistics 216. Springer, 2014. [MR3289986](#)



- [5] F. Biagini, Y. Hu, B. Øksendal and T. Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*; Springer: London, UK, 2008. [MR2387368](#)
- [6] P. Cheridito, H. Kawaguchi and M. Maejima. *Fractional Ornstein-Uhlenbeck processes*. Electron. J. Probab. **8** (2003), 14 p. [MR1961165](#)
- [7] I. Cialenco, F. Delgado-Vences, and H.-J. Kim. *Drift estimation for discretely sampled SPDEs*. Stoch PDE: Anal Comp **8** (2020), 895–920. [MR4174073](#)
- [8] I. Cialenco, S. V. Lototsky and J. Pospíšil. *Asymptotic properties of the maximum likelihood estimator for stochastic parabolic equations with additive fractional Brownian motion*. Stoch. Dyn. **9** (2009), no. 02, 169–185. [MR2531625](#)
- [9] J. F. Coeurjolly. *Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles*. Ann. Statist. **36** (2008), 1404–1434. [MR2418662](#)
- [10] T. Duncan, B. Maslowski, and B. Pasik-Duncan. *Semilinear Stochastic Equations in a Hilbert Space with a Fractional Brownian Motion*. SIAM J. Math. Analysis. **40** (2009), 2286–2315. [MR2481295](#)
- [11] J. Diehl, P. Friz and H. Mai. *Pathwise stability of likelihood estimators for diffusions via rough paths*. Ann. Appl. Probab. **26** (2016), no. 4, 2169–2192. [MR3543893](#)
- [12] A. Gloter and M. Hoffmann. *Estimation of the Hurst parameter from discrete noisy data*. Ann. Statist. **35** (2007), 1947–1974. [MR2363959](#)
- [13] Y. Hu and D. Nualart. *Parameter estimation for fractional Ornstein-Uhlenbeck processes*. Stat. Probab. Lett. **80** (2010), no. 11, 1030–1038. [MR2638974](#)
- [14] Y. Hu, D. Nualart and H. Zhou. *Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter*. Stat. Inference Stoch. Process **22** (2019), 111–142. [MR3918739](#)
- [15] J. Istas and G. Lang. *Quadratic variations and estimation of the local Hölder index of a Gaussian process*. Ann. Inst. H. Poincaré (Probab. et Stat.) **33** (1997), 407–436. [MR1465796](#)
- [16] P. Kříž and B. Maslowski. *Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations*. Stochastics **91** (2019), 1109–1140. [MR4023510](#)
- [17] P. Kříž. *A space-consistent version of the minimum-contrast estimator for linear stochastic evolution equations*. Stochastics and Dynamics **20** (2020). [MR4101086](#)
- [18] P. Kříž and L. Szała. *Least-Squares Estimators of Drift Parameter for Discretely Observed Fractional Ornstein–Uhlenbeck Processes*. Mathematics **8** (2020), no. 5.
- [19] P. Kříž and L. Szała. *The Combined Estimator for Stochastic Equations on Graphs with Fractional Noise*. Mathematics **8** (2020), no. 10.
- [20] G. J. Lord, C. E. Powell and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2014. [MR3308418](#)

- [21] S. V. Lototsky and B. L. Rozovsky. *Stochastic Partial Differential Equations*. Universitext. Springer, Cham, 2017. [MR3674586](#)
- [22] B. Maslowski and C. A. Tudor. *Drift parameter estimation for infinite-dimensional fractional Ornstein-Uhlenbeck process*. Bull. Sci. math. **137** (2013) no. 7, 880–901. [MR3116218](#)
- [23] Y. Mishura. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Springer: Berlin/Heidelberg, Germany, 2008. [MR2378138](#)
- [24] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 2012. [MR2962301](#)
- [25] F. Panloup, S. Tindel and M. Varvenne. *A general drift estimation procedure for stochastic differential equations with additive fractional noise*. Electron. J. Statist. **14** (2020), no. 1, 1075–1136. [MR4068875](#)
- [26] G. Pasemann, S. Flemming, S. Alonso, C. Beta and W. Stannat. *Diffusivity Estimation for Activator-Inhibitor Models: Theory and Application to Intracellular Dynamics of the Actin Cytoskeleton*. J. Nonlinear Sci. **31** (2021), no. 59, 1432–1467. [MR4255684](#)
- [27] G. Pasemann and W. Stannat. *Drift estimation for stochastic reaction-diffusion systems*. Electron. J. Statist. **14** (2020), no. 1, 547–579. [MR4056266](#)
- [28] L. Piterbarg and B. Rozovsky. *Maximum likelihood estimators in the equations of physical oceanography*, in Stochastic Modelling in Physical Oceanography, ed. by R. J. Adler, P. Müller, B. Rozovsky. Progress in Probability, vol. 39 (Birkhäuser Boston, Boston, 1996), 397–421. [MR1383880](#)
- [29] M. Rosenbaum. *Estimation of the volatility persistence in a discretely observed diffusion model*. Stochastic Process. Appl. **118** (2008), 1434–1462. [MR2427046](#)
- [30] Y. Safarov and D. Vassiliev. *The Asymptotic Distribution of Eigenvalues of Partial Differential Operators* volume 155 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, 1997. [MR1414899](#)
- [31] M. A. Shubin. *Pseudodifferential operators and spectral theory*. Springer-Verlag, Berlin, 2nd edition, 2001. [MR1852334](#)