

Statistical Inference for SPDEs: an overview

Igor Cialenco

Department of Applied Mathematics, Illinois Institute of Technology
10 W 32nd Str, Building E1, Room 208, Chicago, IL 60616, USA
cialenco@iit.edu, <http://math.iit.edu/~igor>

First Circulated: December 1, 2017

This Version: January 30, 2018

ABSTRACT: The aim of this work is to give an overview of the recent developments in the area of statistical inference for parabolic stochastic partial differential equations. Significant part of the paper is devoted to the spectral approach, which is the most studied sampling scheme under which the observations are done in the Fourier space over some finite time interval. We also discuss the practically important case of discrete sampling of the solution. Other relevant methodologies and some open problems are briefly addressed over the course of the manuscript.

KEYWORDS: parabolic SPDE, stochastic evolution equations, statistical inference for SPDEs, identification problems for SPDEs

MSC2010: 60H15, 65L09, 62M99

1 Introduction

The general analytical theory for (linear and nonlinear) Stochastic Partial Differential Equations (SPDEs) went through major advances during the past few decades and became a mature mathematical field. From practical point of view, SPDEs are used to describe and to model the evolution of dynamical systems in the presence of persistent spatial-temporal uncertainties, and are key ingredients in modeling various phenomena from fluid mechanics, oceanography, temperature anomalies, finance, economics, biological and ecological systems, and many other applied disciplines. We refer to the classical monographs [Roz90, DPZ92], and also to the textbooks [Cho07, Hai09, LR17], for an in depth discussion of the theory of SPDEs and their various applications. While the general form of a particular stochastic evolution equation is commonly derived from the fundamental properties of the underlying processes under study, frequently the parameters arising in the formulation need to be specified or determined on the basis of some empirical observations. Moreover, even if the parameters are known as part of the specification of the underlying model, the observer may need to test how well the empirical data fits the considered model. These inverse problems, naturally arising in practical applications, fall into realm of the well developed field of statistical inference for stochastic processes. Nevertheless, due to the infinite dimensional nature of SPDEs, some of the statistical models arising from SPDEs are fundamentally different from their counterpart in (finite dimensional) stochastic ordinary differential equations (SODEs). Albeit, there exists a good number of papers devoted to parameter estimation problems for SPDEs, it would be fair to say that the field of statistical inference for SPDEs is in its developing stage with many fundamental problems still open.

The aim of this paper is to give an overview of the main methodologies developed in the area of statistical inference for SPDEs. Due to the page limitation, the detailed proofs will be

omitted, and some results will be mentioned only briefly. We will also take as known most of the notions and fundamental results from infinite dimensional stochastic analysis. The paper is divided in (sub)sections, each of them being devoted to a general method or problem. All the relevant literature and the obtained results will be discussed within the corresponding subsection.

The equation. Throughout, we will assume that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ is a stochastic basis satisfying the usual assumptions, and let H be a separable Hilbert space endowed with the inner product (\cdot, \cdot) and the corresponding norm $\|\cdot\|$. The main object of interest is the following stochastic evolution equation

$$du(t) + (\theta A + B)u(t) dt = (Mu(t) + \sigma) dW^Q(t), \quad (1.1)$$

with the initial condition $u(0) = u_0 \in H$, and where A is a linear, positive defined, self-adjoint operator in H , B is a linear or nonlinear operator in H , M is an operator acting in a suitable Hilbert space, W^Q is a Q -cylindrical Brownian motion in H , σ and θ are some positive constants. If Q is the identity operator, then we will refer to W^Q as to the space-time white noise, and simple write W . Although most of the results presented below hold true for general evolution equations under fair general assumptions, practically speaking, usually A and M are some (pseudo)differential operators, hence SPDEs. Since most of the literature on statistical inference for SPDEs is devoted to parabolic equations, we will assume that (1.1) is parabolic. Moreover, we will make the standing assumption that the equation (1.1) has a unique solution in H (or more precisely in the appropriate triple of Hilbert spaces). The solution is either weak, mild or strong in the PDE sense, and strong in the probability sense. Of course, for each particular class of considered equations, one has to show that indeed the existence and uniqueness of the solution holds true; most of the examples considered here can be dealt within the abstract framework presented in [LR17, Section 4.4] and/or [Cho07, Hai09, Roz90, DPZ92].

The statistical problem. We will be mainly interested in statistical inference problems related to the model parameters θ and/or σ , assuming that all other quantities are known to the observer. We will call θ the drift or viscosity coefficient, and σ the volatility. If the only parameter of interest is θ , then all the obtained results can be easily adjusted to the case when σ is an adapted vector-valued function.

Statistical inference for SPDEs vs SODEs. Let us consider a finite dimensional diffusion process of the form $dX(t) = \mu a(X) dt + \eta b(X(t)) dw(t)$, with μ and η being the parameters of interest, and a, b some suitable, real valued functions. Assume that the solution is observed continuously over some finite interval of time. It is well known that estimating the volatility η is a singular problem, namely the measures $\{\mathbb{P}_{\eta, T}, \eta > 0\}$, on the space of trajectories $C([0, T]; \mathbb{R})$ generated by the solution X are singular to each other. This indicates that σ can be found exactly. For example, using the ‘quadratic variation argument’, we get $\eta = (\langle X \rangle_T / \int_0^T (b(X(t))^2 dt)^{\frac{1}{2}}$. Of course, if the data is observed discretely, then we can only estimate η , for example by approximating the above formula, or by using some statistical methods. Consistency of such estimators can be achieved by decreasing the frequency of the time step while keeping the time horizon T fixed, or increasing the observation time while step size is fixed, or both.

On the other hand, estimating μ usually is a regular problem, in the sense that the measures $\{\mathbb{P}_{\mu, T}, \mu \in \mathbb{R}\}$, generated by the solution X of the SODE with drift μ , are absolutely continuous with respect to each other, and one can employ the statistical inference theory for stochastic processes to find some estimators for μ , for example the maximum likelihood estimator (MLE). In this case to find μ one has to either increase the time (large time asymptotics $T \rightarrow \infty$), or decrease the noise (small noise asymptotics $\eta \rightarrow 0$), or both, increase time and decrease noise. Discrete time observations will not change the nature of the problem. Similar conclusions hold true for finite dimensional stochastic differential equations (SODEs), and there are known necessary and sufficient

conditions for measures $\mathbb{P}_{\mu,T}$ to be equivalent. We refer to classical monographs [LS00, LS78, Kut04] for details on statistical inference for SODEs, and to [Koz77, Koz78, KL85, Log84, MR94, MR02] for extension of Girsanov-type theorems and absolute continuity of measures to infinite dimensional spaces.

As far as estimating volatility σ for SPDEs, this is essentially the same as for SODEs. The key difference in parameter identification problems for SPDEs vs SODEs comes in estimating the drift θ - usually, if A is the leading order operator, the measures $\{\mathbb{P}_{\theta,T}, \theta > 0\}$, generated by the solution $u, t \in [0, T]$, are mutually singular. This indicates that the drift θ can be found exactly, if one path of the solution is observed as an element of the Hilbert space H and continuously over a finite time interval. As with any singular problem, there is no universal method to study it, and each problem has to be analyzed separately. In many situations, a singular model can be approximated by regular models, and for a large class of SPDEs this can be achieved by applying MLE methodology to a projection of the solution on an appropriate finite dimensional subspace of H ; see Section 2. Alternatively, for some particular classes of SPDEs, one can explore the singularity of measures directly and find explicitly the parameters of interest; see Section 2.3.

The observations. The first question to answer is what data is available to the observer. From practical point of view, usually, the observer will take measurements of the solution $u(t, x)$ at some *discrete time* points t_i and/or some *discrete spatial* points x_j , over one path/realization $\omega \in \Omega$. While practical, only few classes of linear SPDEs have been fully investigated under this sampling scheme; see Section 3.

Analogous to inverse problems for diffusion processes, it is reasonable to take the continuous time and/or space observation approach for SPDEs too. Most of the literature on this subject for SPDEs is devoted to the so-called *spectral approach*, when the experimenter observes continuously in time one trajectory of a finite number N of the Fourier modes over some finite interval of time. We discuss this sampling scheme in Section 2. Note that in this case, estimating volatility σ is a trivial task, and σ can be found exactly by using quadratic variation argument, similar to finite dimensional diffusions. Thus, in the spectral approach, we focus only on estimating θ .

In the context of the proposed observation schemes, there are several asymptotic regimes that one can study. Assuming spectral approach, fix N and study the long time asymptotics $T \rightarrow \infty$. This usually reduces the problem to a finite dimensional system of SODEs that has been thoroughly investigated in the existing literature; cf. [Kut04] and references therein. We will omit this analysis here, and for more details specific to SPDEs, we refer the reader to [LR17, CX15] for the spectral approach, to [KL85, Log84] for non-spectral approach, and to [GM02] for controlled SPDEs. As already mentioned, with the idea of approximating a singular model by regular models, in the spectral approach, we will take the large number of Fourier modes asymptotics, $N \rightarrow \infty$, which will be one of the main focuses of this paper; see Section 2. To some extent, this regime corresponds to the fine spacial resolution - at least from approximation point of view, the more sampling points available in the physical domain, the more and better Fourier modes can be approximated, and knowing the solution for all spacial points is equivalent to knowing all Fourier coefficients of the solution. Another natural asymptotics to consider in spectral approach is long time and large space resolution $T, N \rightarrow \infty$, which was not yet study in the existing literature.

Given the singular nature of the problem, in the discrete sampling regime it makes sense to study first the asymptotics in large number of spacial points (fine spacial resolution) and/or large number of time points (fine time-step size), while keeping the time horizon T fixed. We study these in Section 3. Alternatively, one can also study large time asymptotics, or a combination of all the above, which are still open problems.

Finally, a classical asymptotic regime for estimating θ is small noise asymptotics, i.e. the scalar

coefficient in front of the noise goes to zero. The literature on this regime is limited [IK98, IK99, IK00, Hue99, PR03]. These are interesting problems to investigate, while practically speaking the observer rarely deals with vanishing noises.

Note that all the above sampling scheme assume that only one realization or one path of the solution is observed. Sampling multiple paths is another possibility, and besides generalizing the methods developed for one path, one can use energy-type estimates of the solution to build statistical estimates for the parameters of interests.

Classes of SPDEs. It is well known that the structure of the noise of an SPDE, in our case the right hand side of (1.1), may affect significantly the properties of the solution, and more importantly, the technical difficulties in proving certain results. The two main classes of noise structures are: *additive noise*, when $M = 0$, i.e. the solution is not part of the noise, and *multiplicative noise* otherwise. Usually, dealing with multiplicative noise is more challenging, and statistical inference for SPDEs is not an exception in this regard. There are only few results on parameter estimation problems for SPDEs driven by multiplicative noise [PT07, CL09, Cia10, CH17], primarily due to the fact that the main methods of proofs exploit the diagonalizable structure of the additive noise.

The presence of the nonlinear operator B in (1.1) makes the SPDE nonlinear, and as one may expect, studying such equations usually involves techniques specific to the considered equation. In Section 2 we will present a general method how to construct and study estimators for θ in the presence of nonlinear term and driven by an additive noise.

Examples. We will present several examples of SPDEs that we will consider over the course of this work. Again, we will omit the discussion on the existence and uniqueness of the solutions, and we will only mention that all these equations are well-posed, in the sense that they do admit a unique solution in proper spaces under some technical assumptions that do agree with the assumptions encounter by the considered statistical problems and the proposed methods.

I) *Stochastic heat equation driven by additive noise*,

$$du(t, x) - \theta \Delta u(t, x) dt = \sigma dW^Q(t, x), \quad t > 0, \quad (1.2)$$

with zero initial condition $u(0, x) = 0$, and where $x \in G \subset \mathbb{R}^d$, $\Delta u = \sum_{k=1}^d u_{x_k x_k}$ denotes the Laplace operator, $\theta, \sigma \in \mathbb{R}_+$. As far as the domain G , we will consider two cases: I.a) the case of bounded domain, assuming that G is a smooth, bounded domain in \mathbb{R}^d , and equation (1.2) endowed with zero boundary condition; I.b) the whole space case, by taking $G = \mathbb{R}^d$.

II) $du(t, x) - (\Delta u(t, x) + \theta u(t, x)) dt = \sigma dW^Q(t, x)$, $t \in (0, T]$, $x \in (0, \pi)$, with zero initial and boundary conditions, and where $\theta \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$.

III) *2D Navier–Stokes Equations forced with additive noise.* Consider the equation that describes the flow of a viscous, incompressible fluid, known as Navier-Stokes equations, of the form,

$$\begin{aligned} du - \theta \Delta u dt + (u \cdot \nabla)u dt + \nabla P dt &= \sigma dW^Q(t), \\ \nabla \cdot U &= 0, \\ u(0) &= u_0, \end{aligned} \quad (1.3)$$

where u represents the velocity field and P the pressure. We consider either zero boundary conditions on some $G \subset \mathbb{R}^2$, or periodic boundary conditions.

IV) *Stochastic heat equation, simple multiplicative noise.* A variation of Example I,

$$du(t, x) - \theta \Delta u(t, x) dt = \sigma u(t, x) dw(t), \quad t > 0, \quad (1.4)$$

where $x \in [0, \pi]$, $u(0) = u_0$, and w is a one dimensional standard Brownian motion.

V) *Stochastic heat equation driven by space-time multiplicative noise*, which is the multiplicative noise counterpart of Example I,

$$du(t, x) - \theta \Delta u(t, x) dt = \sigma u(t, x) dW^Q(t, x), \quad t > 0, \quad (1.5)$$

with initial data $u(0) = u_0$, and either bounded domain $x \in G \subset \mathbb{R}^d$ with Dirichlet boundary conditions, or the whole space $x \in \mathbb{R}^d$.

2 Spectral approach

Most of the results on statistical inference for SPDEs are obtained within the so-called spectral approach. The first key assumption is that the equation (1.1) is ‘diagonalizable’. That is, we assume that the operators A and Q have pure point spectrum, and a common system of eigenfunctions $\{h_k\}_{k \in \mathbb{N}}$ that forms a complete orthonormal system in H . We will not make any assumptions on B yet. We denote by ν_k and q_k^2 the eigenvalues of A and Q , respectively, corresponding to the eigenfunction h_k , for $k \in \mathbb{N}$. Using this basis we define $H^N := \text{span}\{h_k : k = 1, \dots, N\}$. We denote by P^N the projection operator from H onto H^N , and we will write $u^N = P^N u = \sum_{k=1}^N u_k h_k$, where $u_k = (u, h_k)$, $k \in \mathbb{N}$, are the Fourier coefficients (or modes) of the solution u with respect to $\{h_k\}_{k \in \mathbb{N}}$. Note that in this case, W^Q can be formally written as

$$W^Q(t) = \sum_{k=1}^{\infty} q_k h_k w_k(t),$$

where w_k are independent standard Brownian motions. Also note that since $H^N \cong \mathbb{R}^N$, we can view u^N as an element of \mathbb{R}^N for any fixed $t \in [0, T]$.

Second key postulate in the spectral approach is to assume that we observe continuously in time one path of the Fourier modes $u_k(t)$, $k = 1, \dots, N$, over some finite interval of time $[0, T]$.

We are interested in estimating the parameter $\theta \in \Theta \subset \mathbb{R}_+$, and we will assume that the positive constant σ is known.

The class of equations to which the spectral approach can be applied is large and practically important. Usually, we are dealing with pseudo-differential operators on a bounded domain, for which there exist easy to check conditions that guarantee that the spectrum is discrete and the set of eigenfunctions form a basis. Moreover, the asymptotic behavior of the eigenvalues ν_k of A is tightly related to the order A (as a differential operator) and the dimension of the space \mathbb{R}^d , namely¹ $\nu_k \asymp k^{\text{ord}(A)/d}$. Thus, if $A = (-\Delta)^\beta$ is the fractional Laplace operator on some $G \subset \mathbb{R}^d$, then $\nu_k \asymp k^{2\beta/d}$. For this analysis, the particular form of the eigenfunctions is not important. Note that, the differential operator, and hence the SPDE, defined on the whole space will not have a pure point spectrum, and the spectral approach in principle can not be applied to these equations.

2.1 Modified MLEs: general methodology

We will consider equations driven by an additive noise, i.e. $M = 0$. By applying the projection operator P^N to (1.1) we obtain

$$du^N + (\theta A u^N(t) + \Psi^N) dt = \sigma dW^{Q,N}(t), \quad u^N(0) = P^N u_0, \quad (2.1)$$

¹As usual, for two sequences of positive numbers $\{a_n\}_{n \in \mathbb{N}}$ and $\{b_n\}_{n \in \mathbb{N}}$, we will write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 1$, and will write $a_n \asymp b_n$ if there exist universal constants $K_2 > K_1 > 0$, such that $K_1 b_n \leq a_n \leq K_2 b_n$ for large enough $n \in \mathbb{N}$.

where $\Psi^N = P^N B(u)$, and $W^{Q,N} = P^N W^Q$.

Note that (2.1) generally speaking is not an equation in u^N , unless P^N commutes with B . Nevertheless, let us pretend that Ψ^N is known, and denote by $\mathbb{P}_\theta^{T,N}$ the probability measure on $C([0, T]; H^N)$ generated by the solution u^N of (2.1). Under the *unjustified* assumption that the measures $\{\mathbb{P}_\theta^{T,N}(\cdot)\}_{\theta \in \Theta}$ are equivalent, we now fix a reference (true) parameter θ_0 and formally apply the Girsanov theorem (see e.g. [LS00, Section 7.6.4]) to (2.1), and we obtain the Radon-Nykodym derivative or Likelihood Ratio

$$\frac{d\mathbb{P}_\theta^N}{d\mathbb{P}_{\theta_0}^N}(u^N) = \exp\left(-(\theta - \theta_0)\sigma^{-2} \int_0^T (Q^{-1}Au^N, du^N) - \frac{1}{2}(\theta^2 + \theta_0^2)\sigma^{-2} \int_0^T \|Q^{-1/2}Au^N\|^2 dt - (\theta - \theta_0)\sigma^{-2} \int_0^T (Q^{-1}Au^N, \Psi^N) dt\right),$$

where with slight abuse of notations, some of the products are understood as dot products between vectors, or usual matrix multiplication. Next, we formally compute the maximum likelihood estimator for the parameter of interest θ by maximizing the log-likelihood ratio $\log(d\mathbb{P}_\theta^N / d\mathbb{P}_{\theta_0}^N(u^N))$ with respect to θ , and obtain the following estimator

$$\theta_N^\sharp = -\frac{\int_0^T (Q^{-1}Au^N, du^N) + \int_0^T (Q^{-1}Au^N, P^N B(u)) dt}{\int_0^T \|Q^{-1/2}Au^N\|^2 dt}. \quad (2.2)$$

Note that the estimator θ_N^\sharp uses the full spacial resolution of the solution u , unless B commutes with P^N which is the case for many linear equations. On the other hand, even if B is nonlinear, since u^N converges to u in some space, one would expect that replacing u by u^N would not change significantly the quality of the estimator θ_N^\sharp . With this in mind, taking (2.2) as an ansatz for our *modified maximum likelihood estimators* (mMLEs), we introduce two additional degrees of freedom, $\alpha, \rho \in \mathbb{R}$, and propose the following estimator for θ

$$\widehat{\theta}_N = -\frac{\int_0^T (Q^\rho A^\alpha u^N, du^N) + \int_0^T ((Q^\rho A^\alpha u^N, P^N B(u^N)) dt}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} u^N\|^2 dt}. \quad (2.3)$$

Several comments in order.

- (i) If the equation (1.1) is linear, and B has the same system of eigenfunctions $\{h_k\}_{k \in \mathbb{N}}$, and thus B commutes with A , and hence also with Q and P^N , then θ_N^\sharp is the true MLE of θ that satisfies all desired asymptotical properties as $N \rightarrow \infty$. There is no need to modify the MLE in this case.
- (ii) One has to be caution, since the true MLE is computed as maximum of the log-likelihood ratio over the domain Θ that supports θ . For the considered SPDEs, θ is assumed to be at least positive, and in some applications Θ clearly is a finite interval, while there is no guarantee that $\widehat{\theta}_N$ will belong to this domain. This aspect is studied into detail in [LR17]. However, one may argue that if the estimator is consistent, then eventually, $\widehat{\theta}_N \in \Theta$ for large N , and the observer has just to discard the estimates that fall outside of Θ .
- (iii) If (1.1) is linear, but B does not commute with A , then it is enough to modify θ_N^\sharp by projecting u to H^N , i.e. take in (2.3) $\alpha = 1$ and $\rho = -1$. Hence, there is no need to deal with the two additional parameters α, ρ , in order to prove consistency and asymptotical normality.

(iv) The additional parameters α, ρ are meant to take care of the nonlinear term while proving the asymptotic properties of the estimators. Usually, these free parameters are carefully chosen on case by case basis, and are dictated by the analytical properties of the solution. Hypothetically, one may even consider α and ρ dependent on N .

(v) For some nonlinear equations, dropping out the nonlinear term in (2.3), and considering

$$\hat{\theta}_N^l = -\frac{\int_0^T (Q^\rho A^\alpha u^N, du^N)}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} u^N\|^2 dt},$$

will still retain the consistency property. A lower computation complexity of the estimator $\hat{\theta}^l$ usually comes at the cost of a lower speed of convergence.

(vi) For some examples with nonlinear B it useful to truncate u in $B(u)$ at higher modes, and replace u^N with u^{N^p} for some $p \geq 1$.

To establish the asymptotic properties of the estimators $\hat{\theta}_N$ and $\hat{\theta}_N^l$, as usual, we write the estimators as θ + ‘error terms’, and prove that the error terms vanish. For example, using (2.1), the estimator $\hat{\theta}_N$ can be written as,

$$\hat{\theta}_N = \theta - \frac{\sigma \int_0^T (Q^\rho A^\alpha u^N, dW^{Q,N}) + \int_0^T ((Q^\rho A^\alpha u^N, P^N(B(u^N) - B(u)) dt)}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} u^N\|^2 dt}. \quad (2.4)$$

In the next sections we will sketch some proofs, starting with the linear case.

2.1.1 Linear Equations

More details on estimation of θ for linear equations discussed in this section can be found in the recent monograph [LR17, Chapter 6], and in the survey paper [Lot09].

Fully diagonalizable equations. Let us first focus on linear case and additive noise, i.e. $B = 0$, $M = 0$, and zero initial data. In this case, the equation (1.1) is fully diagonalizable, and each Fourier mode is an Ornstein–Uhlenbeck process,

$$du_k(t) + \theta \nu_k u_k dt = \sigma q_k dw_k(t), \quad k \geq 1. \quad (2.5)$$

The estimator $\hat{\theta}_N$ takes the form

$$\hat{\theta}_N = -\frac{\sum_{k=1}^N \int_0^T \nu_k^\alpha q_k^{2\rho} u_k du_k(t)}{\sum_{k=1}^N \int_0^T \nu_k^{1+\alpha} q_k^{2\rho} u_k^2(t) dt} = \theta - \frac{\sigma \sum_{k=1}^N \int_0^T \nu_k^\alpha q_k^{2\rho+1} u_k dw_k(t)}{\sum_{k=1}^N \int_0^T \nu_k^{1+\alpha} q_k^{2\rho} u_k^2(t) dt}. \quad (2.6)$$

Remark 2.1. A curious reader may ask why not to consider the MLE for θ by using individual Fourier modes (2.5) and then combine the first N of these estimators into one estimator, rather than taking the MLE for the first N Fourier modes at once. At intuitive level, since each Fourier mode is driven by an independent noise, N modes will contain more information than an individual one, and it makes sense to optimize over combined information rather than on individual piece of information and then combine/average them out. It turns out that indeed $\hat{\theta}_N$ has a higher rate of convergence, in the sense of the asymptotic normality property. For a detailed discussion, see [LR17, Chapter 6].

Theorem 2.2. Assume that, $\nu_k \rightarrow \infty$, $q_k \rightarrow 0$, and α, ρ are such that

$$\nu_n^{\alpha-1} q_n^{2\rho+2} \leq M, \quad n \geq 1, \quad \sum_{k \geq 1} \nu_k^\alpha q_k^{2\rho+2} = \infty, \quad (2.7)$$

for some $M \in \mathbb{R}$. Then, $\widehat{\theta}_N$ is a consistent estimator of θ , i.e. $\widehat{\theta}_N \rightarrow \theta$ with probability one. Moreover, if

$$\sum_{k \geq 1} \nu_k^{2\alpha-1} q_k^{4\rho+4} = \infty,$$

then, $\widehat{\theta}_N$ is also asymptotically normal

$$\frac{\sum_{k=1}^N \nu_k^\alpha q_k^{2\rho+2}}{\sqrt{\sum_{k=1}^N \nu_k^{2\alpha-1} q_k^{4\rho+4}}} (\widehat{\theta}_N - \theta) \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{2\theta}{T}\right).$$

Proof. Let us take

$$\xi_k = \int_0^T \nu_k^\alpha q_k^{\rho+1} u_k \, dw_k, \quad \eta_k = \int_0^T \nu_k^{1+\alpha} q_k^\rho u_k^2 \, dt, \quad b_n = \sum_{k=1}^n \mathbb{E}(\eta_k),$$

and write

$$\widehat{\theta}_N = \theta - \sigma \frac{\sum_{k=1}^N \xi_k}{b_N} \cdot \frac{\sum_{k=1}^N \mathbb{E}(\eta_k)}{\sum_{k=1}^N \eta_k} =: \theta - \sigma I_1 I_2. \quad (2.8)$$

Consistency follows from the (strong) law of large numbers (cf. [Shi96, Theorem IV.3.2], or [CGH11, Lemma 2.2] for a weaker version). The only non-trivial conditions to check are the convergence of the following series

$$\sum_{k \geq 1} \frac{\text{Var } \xi_n}{b_n^2} < \infty, \quad \sum_{k \geq 1} \frac{\text{Var } \eta_n}{b_n^2} < \infty, \quad (2.9)$$

which will imply, respectively, that $I_1 \rightarrow 0$ a.s., and $I_2 \rightarrow 1$ a.s., as $N \rightarrow \infty$, and by (2.8) the consistency of $\widehat{\theta}_N$ follows. To verify (2.9), one needs to establish precise asymptotic behavior of $\text{Var } \xi_n$, $\text{Var } \eta_n$ and b_n , which can be done by direct evaluations. For example, one can show that (see for instance [Lot09, Section 2] for details on these evaluations by several methods)

$$\begin{aligned} \mathbb{E} \int_0^T u_k^2(t) \, dt &= \frac{\sigma^2 q_k^2}{2\theta \nu_k} \left(T - \frac{(1 - e^{-2\theta \nu_k T})}{2\theta \nu_k} \right), \\ \text{Var} \int_0^T u_k^2 \, dt &= \frac{\sigma^4 q_k^4}{4\theta^2 \nu_k^2} \left(\frac{2T}{\theta \nu_k} + \frac{2e^{-2\theta \nu_k T}}{\theta^2 \nu_k^2} + \frac{e^{-4\theta \nu_k T}}{2\theta^2 \nu_k^2} + \frac{4Te^{-2\theta \nu_k T}}{\theta \nu_k} - \frac{5}{2\theta^2 \nu_k^2} \right). \end{aligned}$$

Using these, together with (2.7), it is straightforward to show that (2.9) are indeed satisfied.

One way to prove asymptotic normality is to apply the central limit theorem for martingales (cf. [LR17, Theorem 6.1.4] or [LS89, Theorem 5.5.4]). By similar arguments as above, using the law of large numbers, first one shows that

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \xi_k}{\sum_{k=1}^N \int_0^T \nu_k^{2\alpha} q_k^{2(\rho+1)} \mathbb{E}(u_k^2) \, dt} = 1,$$

with probability one, and hence by central limit theorem for martingales, we have that

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \xi_k}{\left(\sum_{k=1}^N \int_0^T \nu_k^{2\alpha} q_k^{2(\rho+1)} \mathbb{E}(u_k^2) \, dt \right)^{1/2}} \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

This, combined with (2.8), after some simple evaluations, implies the asymptotic normality. This concludes the proof. \square

By analogy, one can study linear diagonalizable equations (1.1) with A, B being linear differential operators that commute. This class of equations was studied in the seminal works [HKR93] and [HR95], where the spectral approach was first introduced; see also [Hue93]. In [HR95], the authors showed that the consistency of $\hat{\theta}_N$, as $N \rightarrow \infty$, depends on the order of the operators A and B , and it holds true, if and only if

$$\text{ord}(A) \geq \frac{1}{2}(\text{ord}(\theta A + B) - d). \quad (2.10)$$

This is dictated by the fact that the measures $\mathbb{P}_\theta, \theta > 0$, generated by the solution u of (1.1), are mutually singular if and only if (2.10) is satisfied. In [HR95, LR17] the authors also study the asymptotic efficiency of MLEs. If (2.10) does not hold true, then the measures $\mathbb{P}_\theta, \theta > 0$, are mutually absolutely continuous. This is the case with the Example II from Section 1, for $d=1$. To obtain consistency and asymptotic normality of $\hat{\theta}_N$ in this case, one has to consider large time asymptotics $T \rightarrow \infty$, or small noise asymptotics $\sigma \rightarrow 0$, or a combination of large times, small noise and large spectral resolution $N \rightarrow \infty$.

Using similar arguments and methods, one can study fully diagonalizable SPDEs with several unknown parameters (in front of A and B); cf. [Hue93, Hue97].

Almost diagonalizable equations. Analogous results to the fully diagonalizable equations discussed above can be obtained if B is linear, but does not commute with A . While the asymptotic behavior of $\hat{\theta}_N$ given by (2.3), with $\alpha = 1, \rho = -1$, remains the same, the proofs become much more technical. These classes of equations were studied in [Lot96, HLR97, Hue97, LR99, LR00, Lot03].

2.1.2 Nonlinear Equations

The never diagonalizable nonlinear equations (1.1), with B being the nonlinear part, and driven by an additive noise ($M = 0$), still can be investigated within the spectral approach. In this case, the key idea in proving the asymptotic properties, introduced in [CGH11], is to split the solution in its linear and nonlinear parts, a technique often used in studying PDEs and SPDEs. Namely, the solution u of equation (1.1) is written as $u = \bar{u} + v$, where \bar{u} solves the equation $d\bar{u} + \theta A \bar{u} dt = \sigma dW^Q(t)$, $\bar{u}(0) = u_0$, and v is the solution of

$$dv + \theta Av dt = -Bu dt, \quad t > 0, \quad v(0) = 0.$$

Recall that consistency is proved by showing that the second term in the right hand side of (2.4) vanishes. Using the splitting argument, this reduces to show that

$$\begin{aligned} J_1 &:= \frac{\int_0^T (Q^\rho A^\alpha \bar{u}^N, dW^{Q,N})}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} \bar{u}^N\|^2 dt} \rightarrow 0, & J_2 &:= \frac{\int_0^T (Q^\rho A^\alpha v^N, dW^{Q,N})}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} \bar{u}^N\|^2 dt} \rightarrow 0, \\ J_3 &:= \frac{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} (\bar{u}^N + v^N)\|^2 dt}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} \bar{u}^N\|^2 dt} \rightarrow 1, & J_4 &:= \frac{\int_0^T ((Q^\rho A^\alpha u^N, P^N(B(u^N) - B(u)))) dt}{\int_0^T \|Q^{\rho/2} A^{(1+\alpha)/2} \bar{u}^N\|^2 dt} \rightarrow 0, \end{aligned}$$

where the convergence is either in a.s. sense or in probability. Note that $J_1 \rightarrow 0$ is already covered by Theorem 2.2. Another key point in using the splitting method is that the nonlinear part v usually is slightly more regular than \bar{u} , and hence its Fourier modes will converge to zero faster.

This observation, used carefully, together with choosing the appropriate α and ρ , allows to show that $J_2 \rightarrow 0$, and $J_3 \rightarrow 1$. By akin arguments, and using the specific form of the nonlinear part, one shows that $J_4 \rightarrow 0$.

Despite the fact that the steps described above sound reasonable, each nonlinear equation has to be studied separately. To the best of our knowledge, the only nonlinear equations studied in the current literature are the 2D Navier–Stokes Equations, Example III, Section 1. For simplicity of writing, let us assume that $q_k = \nu_k^{-\gamma}$. Then, assuming that $\gamma > 1$ and $\alpha + 1 > 2\gamma(\rho + 1)$, it was proved in [CGH11] that $\hat{\theta}_N$ and $\hat{\theta}_N^l$ are consistent estimators for θ , as $N \rightarrow \infty$, in the context of (1.3). As one may expect, the proposed method, as well as the technical proves from the 2D Navier–Stokes Equations should care out for Burgers equation

$$du(t, x) + \left(-\theta u_{xx}(t, x) + \frac{1}{2} \partial_x(u^2(t, x)) \right) dt = \sigma dW^Q(t, x), \quad t > 0,$$

where $x \in [0, \pi]$, $u(0, x) = u_0$, and $u(t, 0) = u(t, \pi) = 0$. We leave the detailed analysis of this equation to future studies.

2.1.3 Fractional noise

A natural problem to consider is parameter estimation for SPDEs driven by a fractional Brownian noise. It turns out that the spectral approach, and the MLEs, can be successfully applied to such equations that are fully diagonalizable and driven by an additive noise. If $w^H(t), t \geq 0$, is a fractional Brownian motion with Hurst parameter H , then w^H is not a martingale, unless $H = 1/2$. However, it is well known that $\int_0^t k_H(t, s) dw^H(s)$, where $s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}$, is a martingale (cf. [Mis08]). Then, it is enough to integrate against k_H the Fourier modes u_k that follow the dynamics

$$du_k(t) + (\theta \nu_k + \mu_k) u_k(t) dt = \sigma q_k dw_k^H(t),$$

where $w_k^H, k \in \mathbb{N}$, are independent fractional Brownian motions, and $\mu_k, k \in \mathbb{N}$, are the eigenvalues of B . After this transformation, being in the martingale setup, in [CLP09] the authors followed the procedure of deriving the MLEs described above with $\alpha = 1, \rho = -1$, thanks to the Girsanov-type theorem developed in [KLB02, KLBR00]. Same problem but in the small noise asymptotic regime was studied in [PR04]. Since the kernel k^H is singular, proving that the error terms in (2.4) vanish is not an obvious task, and it remains an open problem for almost diagonalizable or nonlinear SPDEs.

2.2 Time dependent drift

Let us consider a fully diagonalizable parabolic SPDEs (1.1) driven by an additive space-time white noise ($q_k = 1$), and where θ is time-dependent. The non-parametric statistics for finite dimensional diffusion is a well developed field (cf. [Kut04]). Similar to MLEs some of the methods from SODEs can be adapted to the SPDE setup, again, using the spectral approach. A kernel based interpolation method was employed in [HL00a]. Let $R(t)$ be a compactly supported kernel of order $K \geq 1$, that is, R has compact support, $\int_{\mathbb{R}} R(t) dt = 1$, and $\int_{\mathbb{R}} t^j R(t) dt = 0$ for $j = 1, \dots, K$. Let $(v_n)_{n \geq 1}$, and $(a_n)_{n \geq 1}$ be two decreasing sequences of positive real numbers that converge to zero. Define the inverse cut-off function of u_k as follows

$$U_{k,N}(t) := \begin{cases} 1/u_k(t), & |u_k(t)| > v_N, \\ 1/v_N, & |u_k(t)| \leq v_N. \end{cases}$$

Then, the kernel based estimator for $\theta(t)$ is defined as

$$\tilde{\theta}_N(t) = \frac{1}{h_N \sum_{k=1}^N \nu_k} \sum_{k=1}^N \int_0^T R\left(\frac{s-t}{h_N}\right) U_{k,N}(s) (du_k(s) - \mu_k u_k(s) ds), \quad t \in [0, T], \quad N \geq 1.$$

In [HL00a] it was proved that $\tilde{\theta}_N(t)$ converges in mean-square sense to $\theta(t)$. Also within the spectral approach, in [HL00b] the authors propose a sieve estimator for time dependent viscosity coefficient $\theta(t)$.

2.3 Simple multiplicative noise

Let us focus on Example IV, where the noise is multiplicative, but driven just by a one dimensional Wiener process. The results of this subsection can be easily extended to finite dimensional noise. This type of equations are remarkably interesting from inference point of view, and underpin one more time the singular nature of these problems. On the one hand, (1.4) is diagonalizable, in the sense that the Fourier modes are decoupled, and follow the dynamics of a geometric Brownian motion

$$du_k(t) + \theta \nu_k u_k(t) dt = \sigma u_k(t) dw(t), \quad t > 0, u_k(0) = (u_0, h_k),$$

where in this case $h_k = \sqrt{2/\pi} \sin(kx)$ and $\nu_k = k^2$ are the eigenfunctions and the corresponding eigenvalues of the $-\Delta$ on $[0, \pi]$ with zero boundary conditions. Thus, this SPDE falls under the spectral approach. On the other hand, the general mMLE method described in Section 2.1 can not be applied directly, and even formally one can not apply Girsanov transformation to the first N Fourier modes considered together. However, we can use each individual Fourier mode to obtain an MLE for θ (in contrast to the Remark 2.1)

$$\check{\theta}_N = -\frac{1}{\nu_k T} \int_0^T \frac{du_k}{u_k} = \frac{1}{\nu_k T} \ln \frac{u_k(T)}{u_k(0)} + \frac{\sigma^2}{2\nu_k}.$$

It is easy to show that $\check{\theta}_N$ is consistent and asymptotically normal as $N \rightarrow \infty$.

What is even more interesting, since each Fourier mode contains the same noise factor $w(t)$, taking any two nontrivial modes to eliminate the noise altogether, and to solve for the unknown parameter θ explicitly

$$\theta = \frac{1}{T(\nu_m - \nu_k)} \ln \frac{u_k(T)u_m(0)}{u_k(0)u_m(T)}, \quad k \neq m.$$

Hence, once any two Fourier modes are observed at any time point T , the parameter can be found exactly, without any statistical procedure. In [CL09], we call such ‘estimators’ closed-form exact estimators. The fractional noise counterpart of this class of SPDEs is studied in [Cia10].

2.4 Trajectory Fitting Estimators

Note that as such, the spectral approach just reduces the original infinite-dimensional problem to a finite dimensional one, and hence, one can try to apply any available method from statistical inference for SODEs to the projected system, not necessarily based on MLEs. One such attempt, proposed in [CGH16], was to investigate the applicability of the so-called trajectory fitting estimators for ergodic processes first introduced by Kutoyants [Kut91] (see also [Kut04, Section 1.3 & Section 2.3]), that are an analog of the least squares estimators widely used in time-series analysis.

Let us consider the linear equation from Section 2.1.1, for which the Fourier modes follow the dynamics (2.5). In this case,

$$u_k^2(t) = u_k^2(0) + \int_0^t (\sigma^2 q_k^2 - 2\theta \nu_k u_k^2(s)) ds + 2\sigma q_k \int_0^t u_k(s) dw_k(s), \quad t \geq 0,$$

and we let

$$V_k(t; \theta) := u_k^2(0) + \int_0^t (\sigma^2 q_k^2 - 2\theta \nu_k u_k^2(s)) ds, \quad k \in \mathbb{N}, \quad t \in [0, T].$$

The Trajectory Fitting Estimator (TFE) for the unknown parameter θ is defined as

$$\tilde{\theta}_N^{\text{TFE}} := \arg \inf_{\theta \in \Theta} \sum_{k=1}^N \int_0^T (V_k(t; \theta) - u_k^2(t))^2 dt.$$

The maximizer of the last expression, with respect to θ , can be computed explicitly,

$$\tilde{\theta}_N^{\text{TFE}} = - \frac{\sum_{k=1}^N \nu_k \left(\frac{1}{2} \xi_k^2(T) - u_k^2(0) Y_k(T) - \sigma^2 q_k^2 X_k(T) \right)}{2 \sum_{k=1}^N \nu_k^2 Z_k(T)}. \quad (2.11)$$

where

$$\xi_k(t) := \int_0^t u_k^2(s) ds, \quad X_k(t) := \int_0^t s \xi_k(s) ds, \quad Y_k(t) := \int_0^t \xi_k(s) ds, \quad Z_k(t) := \int_0^t \xi_k^2(s) ds.$$

One advantage of the TFE is that it does not contain a stochastic integral. Assuming that $\nu_k \rightarrow \infty$, $q_k \rightarrow 0$ or q_k is constant, and that $\sum_{k=1}^{\infty} \nu_k^2 \mathbb{E}(Z_k) = \infty$, in [CGH16] the authors proved that $\tilde{\theta}_N^{\text{TFE}}$ is a consistent estimator of θ , as $N \rightarrow \infty$. Moreover, under some additional technical conditions $\tilde{\theta}_N^{\text{TFE}}$ is also asymptotically normal [CGH16, Theorem 3.6].

Numerical Example. We conclude this subsection by presenting a simple study of the convergence of MLE and TFE via a numerical simulations result. We consider the equation (1.2) on $[0, \pi]$,

$$du(t, x) - \theta \Delta u(t, x) dt = dW(t, x), \quad t > 0, x \in [0, \pi], \quad (2.12)$$

with zero boundary conditions, zero initial data, and driven by space-time white noise. We assume that true, but unknown, parameter $\theta = 1$. For illustration purposes, in Figure 1, we present one sample path of the solution over time interval $[0, 1]$; the left panel is the 3D plot and the right panel is the contour plot of the values of the solution $u(t, x)$. Since we assume the spectral approach, we used Euler–Maruyama method to simulate the Fourier coefficients (2.5) with time step equal to 10^{-3} , and where $\nu_k = k^2, q_k = 1, k \in \mathbb{N}$. The presented numerical values of the solution are computing by using first 200 Fourier modes. Assuming that these Fourier modes are observed, we construct the MLE (2.6), with $\alpha = 0, \rho = -1$, and TFE (2.11). The estimates are plotted in Figure 2. As expected, both MLE and TFE converge to the true value $\theta = 1$. The left panel corresponds to the first 60 Fourier modes. The right panel displays the estimates corresponding to the Fourier modes from 50 to 200, with a higher resolution in the vertical axis.

2.5 Hypothesis testing

Most of the literature on statistical inference for (parabolic) SPDEs concerns the parameter estimation problem in various setups and forms, with few exceptions [CX14, CX15] where the authors

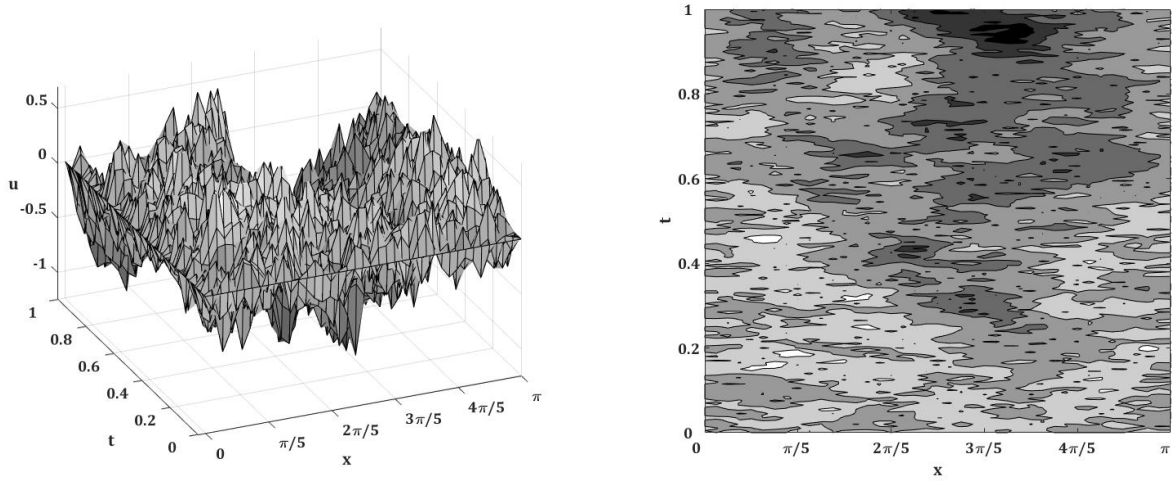


Figure 1: A simulated sample path of the solution of equation (2.12)

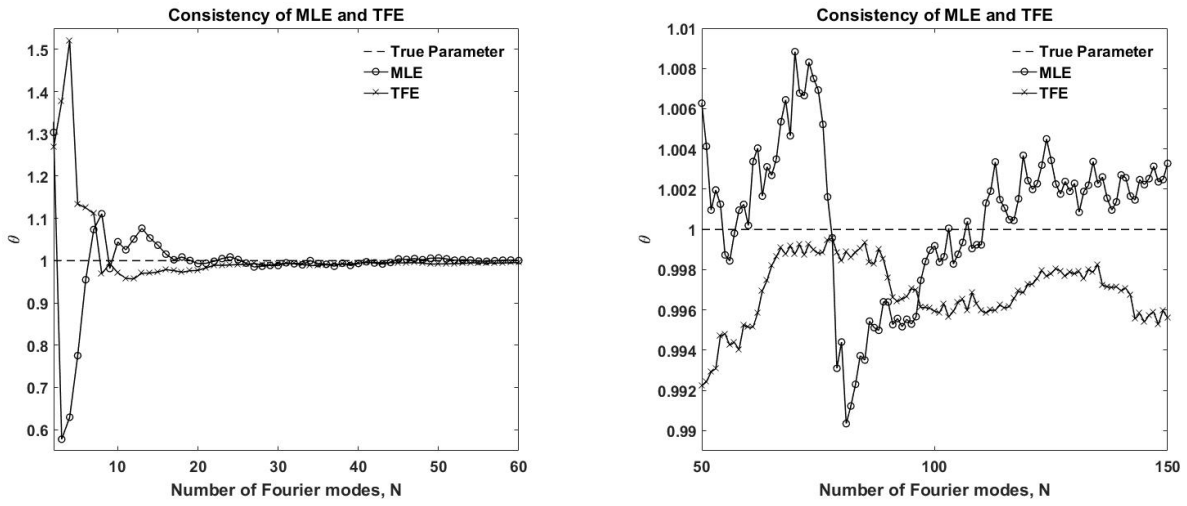


Figure 2: MLE and TFE for θ , equation (2.12)

study the simple hypothesis problems for the drift coefficient θ for the following SPDE

$$du(t, x) + \theta(-\Delta)^\beta u(t, x) dt = \sigma \sum_{k \in \mathbb{N}} \lambda_k^{-\gamma} h_k(x) dw_k(t), \quad t \in [0, T], \quad u(0, x) = u_0, \quad x \in G,$$

where $\theta > 0$, $\beta \geq 1$, $\gamma \geq 0$, $\sigma \in \mathbb{R}_+$, Δ is endowed with zero boundary conditions on the bounded domain $G \in \mathbb{R}^d$, and where λ_k are the eigenvalues of the $\sqrt{-\Delta}$. In this case, the likelihood ratio takes the form

$$L(\theta_0, \theta; u^N) = \exp \left(-\frac{\theta - \theta_0}{\sigma^2} \sum_{k=1}^N \lambda_k^{2\beta+2\gamma} \left(\int_0^T u_k(t) du_k(t) + \frac{1}{2}(\theta + \theta_0) \lambda_k^{2\beta} \int_0^T u_k^2(t) dt \right) \right),$$

that yields the (true) MLE for θ as

$$\widehat{\theta}_T^N = - \frac{\sum_{k=1}^N \lambda_k^{2\beta+2\gamma} \int_0^T u_k(t) du_k(t)}{\sum_{k=1}^N \lambda_k^{4\beta+2\gamma} \int_0^T u_k^2(t) dt}, \quad N \in \mathbb{N}.$$

Although in [CX14, CX15] both regimes, large times $T \rightarrow \infty$ and fine spacial resolution $N \rightarrow \infty$ are investigated, we will focus here only on the latter, $N \rightarrow \infty$. The estimator $\widehat{\theta}_N$ is consistent, and asymptotically normal with the rate of convergence $N^{\beta/d+1/2}$. Assume that θ can take only two values $\theta_0 < \theta_1$, and consider the simple hypothesis

$$\mathcal{H}_0 : \theta = \theta_0, \quad \text{vs} \quad \mathcal{H}_1 : \theta = \theta_1.$$

With MLE and its asymptotic properties at hand, it is easy to establish a Neyman-Pearson-type lemma. Indeed, if c_α is a real number such that $\mathbb{P}_{\theta_0}^{N,T}(L(\theta_0, \theta_1, u^N) \geq c_\alpha) = \alpha$, with $\alpha \in (0, 1)$ being the significance level, then $R^* := \{u^N : L(\theta_0, \theta_1, u^N) \geq c_\alpha\}$, is the most powerful rejection region in the class $\mathcal{K}_\alpha := \{R \in \mathcal{B}(C([0, T]; \mathbb{R}^N)) : \mathbb{P}_{\theta_0}^N(R) \leq \alpha\}$, i.e. $\mathbb{P}_{\theta_1}^{N,T}(R) \leq \mathbb{P}_{\theta_1}^{N,T}(R^*)$, for all $R \in \mathcal{K}_\alpha$. The problem is that the constant c_α can not be computed explicitly. To address this issue and to find a Likelihood Ratio-type test, we approximate c_α by an appropriately chosen sequence $c_\alpha(N)$, and apply the concept of asymptotically the most powerful test introduced in [CX14]; the rejection region $(\widehat{R}_N) \in \widetilde{\mathcal{K}}$ is *asymptotically the most powerful*, in the class $\widetilde{\mathcal{K}}$, as $N \rightarrow \infty$, if

$$\liminf_{N \rightarrow \infty} \frac{1 - \mathbb{P}_{\theta_1}^{N,T}(\widehat{R}_N)}{1 - \mathbb{P}_{\theta_1}^{N,T}(\widetilde{R}_N)} \geq 1, \quad \text{for all } (\widehat{R}_N) \in \widetilde{\mathcal{K}}.$$

To compensate for the asymptotic nature of this setup, one has to shrink the class \mathcal{K}_α , or more precisely its asymptotic version, by considering tests that have a certain rate of convergence of the Type I error. Namely, for $\delta \in \mathbb{R}$, we put $\widehat{\mathcal{K}}_\alpha(\delta) := \{(R_N) : \limsup_{N \rightarrow \infty} (\mathbb{P}_{\theta_0}^N(R_N) - \alpha) \sqrt{M} \leq \alpha_1(\delta)\}$, where $M = \sum_{k=1}^N \lambda_k^{2\beta}$, and $\alpha_1(\delta)$ is an explicitly computable constant (see [CX14], formula (3.15)), and

$$\widehat{R}_N^\delta = \left\{ u^N : L(\theta_0, \theta_1, u^N) \geq c_\alpha^\delta(N) \right\},$$

with

$$c_\alpha^\delta(N) = \exp \left(- \frac{(\theta_1 - \theta_0)^2 T M}{4\theta_0} + \frac{(\theta_1 - \theta_0)^2 N}{8\theta_0^2} - \frac{\sqrt{T M}(\theta_1^2 - \theta_0^2)}{\sqrt{8\theta_0^3}} q_\alpha - \frac{\sqrt{T}(\theta_1^2 - \theta_0^2)}{\sqrt{8\theta_0^3}} \delta \right).$$

Then, the rejection region (\widehat{R}_N^δ) is asymptotically the most powerful in the class $\widehat{\mathcal{K}}_\alpha(\delta)$. The main challenge is to identify the ‘right class’ of tests that requires an exact control of the power of the tests as $N \rightarrow \infty$. The proofs are rooted in the theory of sharp large deviations adapted to the case of large number of Fourier modes.

In [Mar03] the author also studies a simple hypothesis testing problem, in discrete sampling, by testing if an SPDE is parabolic or hyperbolic, also within the spectral approach.

It would be fair to say that this is just the first step towards hypothesis testing problems and goodness of fit test for SPDEs with many open problems left.

3 Discrete sampling

The literature on parameter estimation for discretely sampled SPDEs is limited, while in most applications, the observer will measure one realization of the solution u only at some discrete

points (t_j, x_k) in time and/or space. Of course, one way to deal with discretely sampled data, is to discretize or approximate the (m)MLEs using the available discrete data, and show that the statistical properties are preserved. We refer the reader to [PR02, PR03] for some results on this approach.

In [PR97] the authors develop MLE-type estimators for θ , within the spectral approach, by assuming that the Fourier modes are sampled at some discrete time points on a finite time interval, and under some additional technical assumptions show that these estimators are consistent and asymptotically normal as the mesh size of the time partition goes to zero.

Note that, if we assume that the solution itself is observed at some space-time grid points, one needs to approximate additionally the Fourier modes. To the best of our knowledge, a rigorous asymptotic analysis of this is still an open question. The closest to this approach is [Mar03] where the author considers a parabolic (or hyperbolic) SPDE on $[0, 1]$ driven by an additive noise, and constructs an approximated MLE, by assuming that the solution is observed at a finite and fixed number m of spacial points, and at discrete time points $t_k = \delta k$, $k = 1, \dots, n$, for some $\delta > 0$. The approximated MLE is constructed through the Fourier modes, and it is proved that these estimators are consistent and asymptotically normal as $n \rightarrow \infty$, while δ, m are fixed.

As already mentioned, by its nature, the spectral approach requires that the Fourier decomposition of the solution is performed with respect to the basis formed by the eigenfunctions of the operator A . If A is a differential operator, then essentially one has to deal with bounded domains.

Next, we will discuss some results related to discrete sampling of SPDE (1.1) that are obtained without assuming the spectral approach. It appears that for some classes of SPDEs, to estimate θ and σ , it is enough to observe the solution at one fixed time instant and at discrete grid points of a fixed and finite spacial interval, or just at one fixed spacial point and over a finite time-grid interval. Namely, we will focus on two sampling schemes

- (A) *Fixed time and discrete space.* For a fixed instant of time $t > 0$, and a given interval $[a, b] \subset G \subset \mathbb{R}$, the solution u is observed at points (t, x_j) , $j = 1, \dots, m$, with $x_j = a + (b-a)j/m$, $j = 0, 1, \dots, m$.
- (B) *Fixed space and discrete time.* For a fixed x from the interior of $G \subset \mathbb{R}$, and a given time interval $[c, d] \subset (0, +\infty)$, the solution u is observed at points $\{(t_i, x), i = 1, \dots, n\}$, where $t_i := c + (d - c)i/n$, $i = 0, 1, \dots, n$.

For simplicity of writing, we assume that the sampling points form a uniform grid, but generally speaking the results hold true assuming only that the mesh size of the grid goes to zero. We will use the notation $\Upsilon^m(a, b) = \{a_j \mid a_j = a + (b - a)j/m, j = 0, 1, \dots, m\}$ for the uniform partition of size m of a given interval $[a, b] \subset \mathbb{R}$. For a given stochastic process X on some interval $[a, b]$, and $p \geq 1$, we define the p -variation as

$$\mathbb{V}^p(X; [a, b]) := \lim_{m \rightarrow \infty} \mathbb{V}_m^p(X; [a, b]), \quad \mathbb{P} - \text{a.s.}, \quad \mathbb{V}_m^p(X; [a, b]) := \sum_{j=1}^m |X(t_j) - X(t_{j-1})|^p,$$

and $\mathbb{V}_{\mathbb{P}}^p(X; [a, b])$ will denote the p -variation when the limit is understood in probability sense; sometimes we will simply write $\mathbb{V}^p(X)$, and $\mathbb{V}_m^p(X)$.

The main idea behind this method is similar to the argument of estimating the volatility coefficient for finite dimensional diffusions through quadratic variation arguments. In fact, estimating σ for SPDEs (1.1) can be achieved by this approach. Estimating the drift θ is more delicate, and one has to find the correct, and exact, p -variation to be used.

In [PT07] (see also [Pos05]) the authors explore this idea for stochastic heat equation on whole real line

$$du(t, x) - \theta \Delta u(t, x) dt = \sigma(u(t, x)) dW(t, x), \quad t > 0, \quad x \in \mathbb{R},$$

with $u(0, x) = 0$, $x \in \mathbb{R}$, and where $\theta > 0$ is the parameter of interest, $\sigma : \mathbb{R} \rightarrow \mathbb{R}$, and W is the space-time white noise on $(0, \infty) \times \mathbb{R}$. Consider the sampling scheme (A), with the partition $\Psi^m(a, b)$ for some fixed $a, b \in \mathbb{R}$. It was proved in [PT07] that

$$\widehat{\theta}_{m,t} := \frac{b-a}{2m} \frac{\sum_{j=1}^m \sigma^2(u(t, x_j))}{\sum_{j=1}^m (u(t, x_j) - u(t, x_{j-1}))^2}$$

is a weakly consistent estimator² of θ as $m \rightarrow \infty$. Similarly, for some fixed $0 < c, d \leq T$ and $x \in \mathbb{R}$, assuming that the solution is sampled by scheme (B), the estimator

$$\widehat{\theta}_{n,x} := \frac{3(d-c)}{n\pi} \frac{\sum_{j=1}^n \sigma^4(u(t_j, x))}{\sum_{j=1}^n (u(t_j, x) - u(t_{j-1}, x))^4}$$

is weakly consistent as $n \rightarrow \infty$. As one may guess, $\widehat{\theta}_{m,t}$ comes from the computations of the quadratic variation, while $\widehat{\theta}_{n,x}$ is due to the fourth-variation of the solution. Moreover, assuming that $\sigma(u) = \sigma u$, for some positive constant σ (i.e. Example V, with $G = \mathbb{R}$, and driven by space-time white noise), and assuming that θ is known, then the estimators

$$\widehat{\sigma}_{m,t}^2 = \frac{2m\theta \sum_{j=1}^m (u(t, x_j) - u(t, x_{j-1}))^2}{(b-a) \sum_{j=1}^m u^2(t, x_j)}, \quad \widehat{\sigma}_{n,x}^2 := \sqrt{\frac{n\theta\pi \sum_{i=1}^n (u(t_i, x) - u(t_{i-1}, x))^4}{3(d-c) \sum_{i=1}^n u^4(t_i, x)}}$$

are weakly consistent estimators of σ , respectively when $m \rightarrow \infty$, and $n \rightarrow \infty$. The asymptotic normality of these estimators remains an open problem.

Two recent independent studies [CH17, BT17] were devoted to the stochastic heat equation driven by additive space-time white noise.

In [CH17] the authors further explore the quadratic variation method, starting with a simple and intuitively clear observation: the p -variation of a stochastic process is invariant with respect to smooth perturbations. Hence, if the p -variation of a process X can be computed by an explicit formula, and the parameter of interest enters non-trivially into this formula, one can derive consistent estimators of this parameter. However, since the p -variation of the perturbed process $X + Y$ remains the same, given that Y is smooth enough, then the same estimator remains consistent assuming that $X + Y$ is observed. Analogous arguments remain valid for asymptotic normality property; see [CH17, Proposition 2.1]. Hence, it remains to establish such representations for the solutions of the considered SPDEs. This can be done for equation (1.2), driven by space-time white noise, with G being either a finite interval or the whole real line ($d = 1$).

Assume that the solution u is sampled according to the sampling scheme (A), for some fixed $t > 0$. Then, one can show that the following estimator for θ (assuming σ is known), and σ^2 (assuming θ is known), respectively,

$$\widehat{\theta}_{m,t} := \frac{(b-a)\sigma^2}{2 \sum_{j=1}^m (u(t, x_j) - u(t, x_{j-1}))^2}, \quad \widehat{\sigma}_{m,t}^2 := \frac{2\theta}{b-a} \sum_{j=1}^m (u(t, x_j) - u(t, x_{j-1}))^2,$$

²An estimator is weakly consistent if it converges to the true parameter in probability.

is consistent and asymptotically normal with the rate of convergence \sqrt{m} .

Similarly, using the sampling scheme (B), the estimators

$$\widehat{\theta}_{n,x} := \frac{3(d-c)\sigma^4}{\pi \sum_{i=1}^n (u(t_i, x) - u(t_{i-1}, x))^4}, \quad \widehat{\sigma}_{n,x}^2 := \sqrt{\frac{\theta\pi}{3(d-c)} \sum_{i=1}^n (u(t_i, x) - u(t_{i-1}, x))^4},$$

are also consistent and asymptotically normal with the rate of convergence \sqrt{n} . It is interesting to note that the estimators and the rate of converge remain the same for both, bounded and unbounded domains, although the proof of asymptotic results differ.

Finally, let us consider a variation of equation (1.2), that was studied in [BT17]

$$\begin{aligned} du(t, x) &= \left(\theta_2 \frac{\partial^2}{\partial x^2} u(t, x) + \theta_1 \frac{\partial}{\partial x} u(t, x) + \theta_0 u(t, x) \right) dt + \sigma(t) dW(t, x), \quad t > 0, x \in [0, 1] \\ u_0 &= \xi. \end{aligned}$$

with zero boundary condition, and where σ is an α -Hölder continuous function with $\alpha \in (1/2, 1]$. Assume the sampling scheme (B) over $\Psi_n(0, T)$. In [BT17] the authors construct an estimator $RV_{n,x}$ of the integrated square volatility and showed that

$$RV_{n,x} := \frac{1}{\sqrt{nT}} \sum_{i=1}^n (u(t_i, x) - u(t_{i-1}, x))^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{e^{-x\theta_1/\theta_2}}{\sqrt{\theta_2\pi}} \int_0^T \sigma^2(t) dt,$$

and also established an asymptotic normality result for this estimator. In contrast to [CH17] where the authors use elements of Malliavin calculus, the methods used in [BT17] are rooted in the mixing theory for Gaussian time series.

4 Other methods and results

In this section we will briefly report on some other existing results relevant to the identification of the drift and volatility for SPDE.

Filtering. Generally speaking results on filtering in the context of SPDEs are rather scarce. Using spectral approach, and assuming that $\theta(t)$ follows an unobservable Ito diffusion, Lototsky [Lot04] derives an optimal filter for θ , which can be viewed as a generalization of Kalman-Bucy filter. Using different methods and techniques than those mentioned above, in a series of papers Aihara et al. [Aih91, Aih92, Aih98b, AS88, Aih98a, AB89] study a non-parametric estimation problem of a space dependent θ , combined with a filtering problem by assuming that the observations are $y(t) = \int_0^t F(u(s)) ds + w(t)$, where where u is the solution of the corresponding parabolic SPDE driven by an additive noise, F is an operator on H with finite-dimensional range, and w is a finite-dimensional Brownian motion.

Bayesian Inference for SPDEs is another area with few existing results [Bis02, Bis99, PR00] all within the spectral approach.

Large Time Asymptotics. As already mentioned, large time asymptotics in the spectral approach falls in the well established realm of statistical inference for finite dimensional systems of SODEs. Nevertheless, one can study such asymptotic regime by using the ergodicity property of the SPDE itself that leads to an interesting class of estimators. Such approach was proposed in [MP08] for SPDEs driven by an additive fractional noise. See also [Pos05, CLP09].

Concluding remarks. While this survey is dense and most of the proofs have been omitted, the overall statistical inference for SPDEs is yet to become a mature field with many practically important problems remaining open. Besides some open problems mentioned above, for example, there is little known about estimation of the drift of simplest SPDE, the stochastic heat equation, driven by a multiplicative space-time noise; equation (1.5), Example V. The inference for SPDEs driven by non-Gaussian noise is another open field for investigations. Problems beyond parameter estimation, such as hypothesis testing and goodness-of-fit tests, are areas of research with great potential, and even considering simple SPDEs may lead to interesting results that explore the singular nature of these statistical models.

Acknowledgments

The author is grateful to Yicong Huang for reading the initial draft of this paper and making a series of useful comments that improved greatly the final manuscript. The author would like to thank the anonymous referee and the editors for their helpful comments and suggestions which helped to improve the paper.

References

- [AB89] S. I. Aihara and A. Bagchi. Infinite-dimensional parameter identification for stochastic parabolic systems. *Statist. Probab. Lett.*, 8(3):279–287, 1989.
- [Aih91] S. I. Aihara. Parameter identification for stochastic parabolic systems. In *Systems and control*, pages 1–12. Mita, 1991.
- [Aih92] S. I. Aihara. Regularized maximum likelihood estimate for an infinite-dimensional parameter in stochastic parabolic systems. *SIAM J. Control Optim.*, 30(4):745–764, 1992.
- [Aih98a] S. I. Aihara. Consistency property of extended least-squares parameter estimation for stochastic diffusion equation. *Systems Control Lett.*, 34(5):249–256, 1998.
- [Aih98b] S. I. Aihara. Identification of a discontinuous parameter in stochastic parabolic systems. *Appl. Math. Optim.*, 37(1):43–69, 1998.
- [AS88] S. I. Aihara and Y. Sunahara. Identification of an infinite-dimensional parameter for stochastic diffusion equations. *SIAM J. Control Optim.*, 26(5):1062–1075, 1988.
- [Bis99] J. P. N. Bishwal. Bayes and sequential estimation in Hilbert space valued stochastic differential equations. *J. Korean Statist. Soc.*, 28(1):93–106, 1999.
- [Bis02] J. P. N. Bishwal. The Bernstein-von Mises theorem and spectral asymptotics of Bayes estimators for parabolic SPDEs. *J. Aust. Math. Soc.*, 72(2):287–298, 2002.
- [BT17] M. Bibinger and M. Trabs. Volatility estimation for stochastic PDEs using high-frequency observations. *Preprint, arXiv:1710.03519*, 2017.
- [CGH11] I. Cialenco and N. Glatt-Holtz. Parameter estimation for the stochastically perturbed Navier-Stokes equations. *Stochastic Process. Appl.*, 121(4):701–724, 2011.
- [CGH16] I. Cialenco, R. Gong, and Y. Huang. Trajectory fitting estimators for SPDEs driven by additive noise. *Forthcoming in Statistical Inference for Stochastic Processes DOI:10.1007/s1120*, 2016.
- [CH17] I. Cialenco and Y. Huang. A note on parameter estimation for discretely sampled SPDEs. *Preprint, arXiv:1710.01649*, 2017.
- [Cho07] P. Chow. *Stochastic partial differential equations*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2007.
- [Cia10] I. Cialenco. Parameter estimation for SPDEs with multiplicative fractional noise. *Stoch. Dyn.*, 10(4):561–576, 2010.
- [CL09] I. Cialenco and S. V. Lototsky. Parameter estimation in diagonalizable bilinear stochastic parabolic equations. *Stat. Inference Stoch. Process.*, 12(3):203–219, 2009.

- [CLP09] 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(2):169–185, 2009.
- [CX14] I. Cialenco and L. Xu. A note on error estimation for hypothesis testing problems for some linear SPDEs. *Stoch. Partial Differ. Equ. Anal. Comput.*, 2(3):408–431, 2014.
- [CX15] I. Cialenco and L. Xu. Hypothesis testing for stochastic PDEs driven by additive noise. *Stochastic Process. Appl.*, 125(3):819–866, March 2015.
- [DPZ92] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.
- [GM02] B. Goldys and B. Maslowski. Parameter estimation for controlled semilinear stochastic systems: identifiability and consistency. *J. Multivariate Anal.*, 80(2):322–343, 2002.
- [Hai09] M. Hairer. *Introduction to Stochastic PDEs*. Unpublished lecture notes, 2009.
- [HKR93] M. Huebner, R. Khasminskii, and B. L. Rozovskii. Two examples of parameter estimation for stochastic partial differential equations. In *Stochastic processes*, pages 149–160. Springer, New York, 1993.
- [HL00a] M. Huebner and S. V. Lototsky. Asymptotic analysis of a kernel estimator for parabolic SPDE’s with time-dependent coefficients. *Ann. Appl. Probab.*, 10(4):1246–1258, 2000.
- [HL00b] M. Huebner and S. V. Lototsky. Asymptotic analysis of the sieve estimator for a class of parabolic SPDEs. *Scand. J. Statist.*, 27(2):353–370, 2000.
- [HLR97] M. Huebner, S. V. Lototsky, and B. L. Rozovskii. Asymptotic properties of an approximate maximum likelihood estimator for stochastic PDEs. In *Statistics and control of stochastic processes (Moscow, 1995/1996)*, pages 139–155. World Sci. Publishing, 1997.
- [HR95] M. Huebner and B. L. Rozovskii. On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE’s. *Probab. Theory Related Fields*, 103(2):143–163, 1995.
- [Hue93] M. Huebner. *Parameter Estimation for SPDEs*. PhD thesis, University of Southern California, Los Angeles, USA, 1993.
- [Hue97] M. Huebner. A characterization of asymptotic behaviour of maximum likelihood estimators for stochastic PDE’s. *Math. Methods Statist.*, 6(4):395–415 (1998), 1997.
- [Hue99] M. Huebner. Asymptotic properties of the maximum likelihood estimator for stochastic PDEs disturbed by small noise. *Stat. Inference Stoch. Process.*, 2(1):57–68 (2000), 1999.
- [IK98] I. A. Ibragimov and R. Z. Khasminskii. Problems of estimating the coefficients of stochastic partial differential equations. I. *Teor. Veroyatnost. i Primenen.*, 43(3):417–438, 1998.
- [IK99] I. A. Ibragimov and R. Z. Khasminskii. Problems of estimating the coefficients of stochastic partial differential equations. II. *Teor. Veroyatnost. i Primenen.*, 44(3):526–554, 1999.
- [IK00] I. A. Ibragimov and R. Z. Khasminskii. Problems of estimating the coefficients of stochastic partial differential equations. III. *Teor. Veroyatnost. i Primenen.*, 45(2):209–235, 2000.
- [KL85] T. Koski and W. Loges. Asymptotic statistical inference for a stochastic heat flow problem. *Statist. Probab. Lett.*, 3(4):185–189, 1985.
- [KLB02] M. L. Kleptsyna and A. Le Breton. Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Stat. Inference Stoch. Process.*, 5(3):229–248, 2002.
- [KLBR00] M. L. Kleptsyna, A. Le Breton, and M.-C. Roubaud. Parameter estimation and optimal filtering for fractional type stochastic systems. *Stat. Inference Stoch. Process.*, 3(1-2):173–182, 2000. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).
- [Koz77] S. M. Kozlov. Equivalence of measures in Itô’s linear partial differential equations. *Vestnik Moskov. Univ. Ser. I Mat. Meh.*, (4):47–52, 1977.
- [Koz78] S. M. Kozlov. Some questions of stochastic partial differential equations. *Trudy Sem. Petrovsk.*, (4):147–172, 1978.
- [Kut91] Y. A. Kutoyants. Minimum-distance parameter estimation for diffusion-type observations. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(8):637–642, 1991.
- [Kut04] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.

- [Log84] W. Loges. Girsanov's theorem in Hilbert space and an application to the statistics of Hilbert space-valued stochastic differential equations. *Stochastic Process. Appl.*, 17(2):243–263, 1984.
- [Lot96] S. V. Lototsky. *Problems in statistics of stochastic differential equations*. PhD thesis, University of Southern California, Los Angeles, USA, 1996.
- [Lot03] S. V. Lototsky. Parameter estimation for stochastic parabolic equations: asymptotic properties of a two-dimensional projection-based estimator. *Stat. Inference Stoch. Process.*, 6(1):65–87, 2003.
- [Lot04] S. V. Lototsky. Optimal filtering of stochastic parabolic equations. In *Recent developments in stochastic analysis and related topics*, pages 330–353. World Sci. Publ., Hackensack, NJ, 2004.
- [Lot09] S. V. Lototsky. Statistical inference for stochastic parabolic equations: a spectral approach. *Publ. Mat.*, 53(1):3–45, 2009.
- [LR99] S. V. Lototsky and B. L. Rozovskii. Spectral asymptotics of some functionals arising in statistical inference for SPDEs. *Stochastic Process. Appl.*, 79(1):69–94, 1999.
- [LR00] S. V. Lototsky and B. L. Rozovskii. Parameter estimation for stochastic evolution equations with non-commuting operators. In *in Skorohod's Ideas in Probability Theory, V.Korolyuk, N.Portenko and H.Syta (editors)*, pages 271–280. Institute of Mathematics of National Academy of Sciences of Ukraine, Kiev, Ukraine, 2000.
- [LR17] S. V. Lototsky and B. L. Rozovsky. *Stochastic partial differential equations*. Universitext. Springer, Cham, 2017.
- [LS78] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. II*. Springer-Verlag, 1978.
- [LS89] R. S. Liptser and A. N. Shiryaev. *Theory of martingales*, volume 49 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, 1989.
- [LS00] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes I. General theory*. Springer-Verlag, New York, 2nd edition, 2000.
- [Mar03] B. Markussen. Likelihood inference for a discretely observed stochastic partial differential equation. *Bernoulli*, 9(5):745–762, 2003.
- [Mis08] Y. S. Mishura. *Stochastic calculus for fractional Brownian motion and related processes*, volume 1929 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2008.
- [MP08] B. Maslowski and J. Pospíšil. Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.*, 57(3):401–429, 2008.
- [MR94] R. Mikulevicius and B. L. Rozovskii. Uniqueness and absolute continuity of weak solutions for parabolic SPDEs. *Acta Appl. Math.*, 35(1-2):179–192, 1994.
- [MR02] R. Mikulevicius and B. L. Rozovskii. On martingale problem solutions for stochastic Navier-Stokes equation. In *Stochastic partial differential equations and applications (Trento, 2002)*, volume 227 of *Lecture Notes in Pure and Appl. Math.*, pages 405–415. Dekker, New York, 2002.
- [Pos05] J. Pospíšil. *On parameter estimates in stochastic evolution equations driven by fractional Brownian motion*. PhD thesis, University of West Bohemia, Plzen, 2005.
- [PR97] L. I. Piterbarg and B. L. Rozovskii. On asymptotic problems of parameter estimation in stochastic PDE's: discrete time sampling. *Math. Methods Statist.*, 6(2):200–223, 1997.
- [PR00] B. L. S. Prakasa Rao. Bayes estimation for some stochastic partial differential equations. *J. Statist. Plann. Inference*, 91(2):511–524, 2000. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998).
- [PR02] B. L. S. Prakasa Rao. Nonparametric inference for a class of stochastic partial differential equations based on discrete observations. *Sankhyā Ser. A*, 64(1):1–15, 2002.
- [PR03] B. L. S. Prakasa Rao. Estimation for some stochastic partial differential equations based on discrete observations. II. *Calcutta Statist. Assoc. Bull.*, 54(215-216):129–141, 2003.
- [PR04] B. L. S. Prakasa Rao. Parameter estimation for some stochastic partial differential equations driven by infinite dimensional fractional Brownian motion. *Theory Stoch. Process.*, 10(3-4):116–125, 2004.
- [PT07] J. Pospíšil and R. Tribe. Parameter estimates and exact variations for stochastic heat equations driven by space-time white noise. *Stoch. Anal. Appl.*, 25(3):593–611, 2007.
- [Roz90] B. L. Rozovskii. *Stochastic evolution systems*, volume 35 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990. Linear theory and applications to nonlinear filtering.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996.