Supplementary material for “Adaptive optimal estimation of irregular mean and covariance functions”

Steven Golovkine

MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland

E-mail: steven.golovkine@ul.ie

Nicolas Klutchnikoff

Univ. Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

E-mail: nicolas.klutchnikoff@UNIV-RENNES2.fr

Valentin Patilea

Univ. Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

E-mail: valentin.patilea@ensai.fr

In this Supplementary Material, we provide some additional technical arguments, proofs and simulation results. In Section A we provide details for proofs presented in the main text. In Section B, we provide details on some quantities and equations from the main text. In Section C, we prove Theorem 3. Additional simulation results are gathered in Section D.

A. Complements for the proofs

PROOF (COMPLEMENTS FOR THE PROOF OF THEOREM 2). Here, we provide a formal justification for the following properties: two constants $0 < \mathfrak{c}_1, \mathfrak{c}_2 < 1$ exist such that

$$\mathfrak{c}_1 N\{1 + o_{\mathbb{P}}(1)\} \leq \inf_{h\in\mathcal{H}_N} \frac{\mathcal{W}_N(t;h)}{\min\{1,\mathfrak{m}h\}} \leq \sup_{h\in\mathcal{H}_N} \frac{\mathcal{W}_N(t;h)}{\min\{1,\mathfrak{m}h\}} \leq \mathfrak{c}_1^{-1} N\{1 + o_{\mathbb{P}}(1)\}. \tag{SM.1}$$

and

$$\inf_{h\in\mathcal{H}_N} \frac{N\mathfrak{m}h}{\mathcal{N}_\mu(t;h)} \geq \mathfrak{c}_2\{1 + o_{\mathbb{P}}(1)\}, \tag{SM.2}$$

and of equation (A.4) in the main text. In reply to a Reviewer remark, we prove (SM.1) and (SM.2) in slightly more general framework. Let

$$\overline{M} = \frac{1}{N}\sum_{i=1}^{N} M_i,$$

such that $\mathfrak{m} = \mathbb{E}(\overline{M})$. The M_i are independent, but we do not need to impose them to have the same law. However, for simplicity, we still assume (23). For each $1 \leq i \leq N$, we denote by g_i the density of the independent variables $T_m^{(i)} \in \mathcal{T}$, $1 \leq m \leq M_i$. Moreover, the variables $T_m^{(i)}$ are drawn independently for each curve i. Assume that positive constants $C_{g,L}, C_{g,U} > 0$ exist such that

$$C_{g,L} \leq g_i(t) \leq C_{g,U}, \qquad \forall t \in \mathcal{T},\ \forall 1 \leq i \leq N, \tag{SM.3}$$

and all g_i are Hölder continuous on $\mathcal{T}$ with the same exponent and Hölder constant. We thus allow the observation times $T_m^{(i)}$ to be drawn independently with different distributions for different curves. Let

$$p_i(t;h) = \int_{t-h}^{t+h} g_i(u)du.$$

Under the conditions on the bandwidth range $\mathcal{H}_N$ and the g_i, we have $p_i(t;h) = 2hg_i(t)\{1+o(1)\}$, uniformly with respect to h and i.

To show (SM.2), recall that, with the NW estimator

$$\mathcal{N}_i(t;h) = \frac{w_i(t;h)}{\max_{1\leq m\leq M_i} |W_m^{(i)}(t;h)|} \quad \text{and} \quad \mathcal{N}_\mu(t;h)^{-1} = \frac{1}{\mathcal{W}_N^2(t;h)}\sum_{i=1}^{N}\frac{w_i(t;h)}{\mathcal{N}_i(t;h)}.$$

(Recall that the rule $0/0 = 0$ applies for $w_i/\mathcal{N}_i$). We simplify the notation in the following: $\mathcal{N}_\mu(t;h)$, $\mathcal{N}_i(t;h)$ and $w_i(t;h)$ become $\mathcal{N}_\mu$, $\mathcal{N}_i$ and w_i, respectively. Moreover, for simplicity, we assume the NW is built with the uniform kernel. The general case can be handled similarly using a positive lower bound for the kernel K on a sub-interval of $[-1,1]$. With a uniform kernel we have

$$\mathcal{N}_i = \sum_{m=1}^{M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\}.$$

By Cauchy-Schwarz inequality,

$$\frac{1}{\mathcal{N}_\mu} = \frac{1}{\mathcal{W}_N^2(t;h)}\sum_{i=1}^{N}\frac{w_i}{\mathcal{N}_i} \geq \frac{1}{S_N(t;h)} \quad \text{with} \quad S_N = \sum_{i=1}^{N}\mathcal{N}_i. \tag{SM.4}$$

Note that $S_N(t;h)$ is a sum of $M_1 + \ldots + M_N$ independent Bernoulli variables with parameters

$$\underbrace{p_1(t;h),\ldots,p_1(t;h)}_{M_1 \text{ times}},\ldots,\underbrace{p_N(t;h),\ldots,p_N(t;h)}_{M_N \text{ times}}.$$

We have

$$C_{g,L}N\mathfrak{m}h \times \{1+o(1)\} \leq \mathbb{E}[S_N(t;h)] = \sum_{i=1}^{N} p_i(t;h)\mathbb{E}(M_i) \leq C_{g,U}N\mathfrak{m}h \times \{1+o(1)\}. \tag{SM.5}$$

Recall that we impose $N\mathfrak{m} \times \min \mathcal{H}_N \to \infty$. By Chernoff's inequality, for any $0 \leq \delta < 1$,

$$\mathbb{P}\left(\left|\frac{S_N(t;h)}{\mathbb{E}[S_N(t;h)]} - 1\right| > \delta\right) \leq 2\exp(-\delta^2 C_{g,L}N\mathfrak{m}\min\mathcal{H}_N/3).$$

We can choose δ such that

$$\delta^2 = C_\delta \frac{\log(N\mathfrak{m})}{N\mathfrak{m}\min\mathcal{H}_N},$$

with C_δ some large constant. If $h_1,\ldots,h_J$ is an equidistant grid on $\mathcal{H}_N$ of J points, with $N\mathfrak{m} \leq J < N\mathfrak{m}+1$, we deduce

$$\mathbb{P}\left(\sup_{1\leq j\leq J}\left|\frac{S_N(t;h_j)}{\mathbb{E}[S_N(t;h_j)]} - 1\right| > \delta\right) \leq 2\exp\left[\log(N\mathfrak{m}) - \delta^2 C_{g,L}N\mathfrak{m}\min\mathcal{H}_N/3\right], \tag{SM.6}$$

and the exponential bound tends to zero when C_δ is sufficiently large. Next, the supremum over the grid can be extended over $\mathcal{H}_N$ using the the Lipschitz continuity of the map $h \mapsto \mathbb{E}[S_N(t;h)]$, and the monotonicity of the maps $h \mapsto S_N(t;h)$ and $h \mapsto \mathbb{E}[S_N(t;h)]$. Finally, by (SM.4), we write

$$\frac{N\mathfrak{m}h}{\mathcal{N}_\mu(t;h)} \geq \frac{N\mathfrak{m}h}{\mathbb{E}[S_N(t;h)]} \times \frac{\mathbb{E}[S_N(t;h)]}{S_N(t;h)} \times \frac{S_N(t;h)}{\mathcal{N}_\mu(t;h)} \geq \frac{N\mathfrak{m}h}{\mathbb{E}[S_N(t;h)]} \times \frac{\mathbb{E}[S_N(t;h)]}{S_N(t;h)},$$

and we deduce (SM.2) from (SM.5) and (SM.6).

Next, to show (SM.1), not that, given M_i, the indicator w_i is a Bernoulli variable with parameter, say,

$$\pi_i(t;h) = 1 - \{1 - p_i(t;h)\}^{M_i}. \tag{SM.7}$$

Let us notice that, for any $M > 0$,

$$-M\frac{u}{1-u} \leq \log(1-u)^M < -uM, \qquad \forall u \in (0,1).$$

Assuming, without loss of generality, that $p_i(t;h) \leq 1/2$, $\forall h \in \mathcal{H}_N$ and for all i, we deduce

$$1 - \exp(-M_i p_i(t;h)) \leq \pi_i(t;h) \leq 1 - \exp(-2M_i p_i(t;h)), \quad \forall h \in \mathcal{H}_N,\ 1 \leq i \leq N. \tag{SM.8}$$

By (SM.3), we have

$$2C_{g,L}h \leq p_i(t;h) \leq 2C_{g,U}h, \qquad \forall h \in \mathcal{H}_N,\ 1 \leq i \leq N.$$

From this and (23), we have

$$1 - \exp(-2C_{g,L}C_L\mathfrak{m}h) \leq \pi_i(t;h) \leq 1 - \exp(-4C_{g,U}C_U\mathfrak{m}h), \quad \forall h \in \mathcal{H}_N,\ 1 \leq i \leq N, \tag{SM.9}$$

from which we deduce

$$\begin{aligned}
&1 - \exp(-2C_{g,L}C_L\mathfrak{m}\min\mathcal{H}_N) \\
&\quad \leq 1 - \exp(-2C_{g,L}C_L\mathfrak{m}h) \leq \frac{\mathbb{E}[\mathcal{W}_N(t;h)]}{N} = \frac{1}{N}\sum_{i=1}^N \pi_i(t;h) \leq 1 - \exp(-4C_{g,U}C_U\mathfrak{m}h) \\
&\qquad\qquad \leq 1 - \exp(-4C_{g,U}C_U\mathfrak{m}\max\mathcal{H}_N), \quad \forall h \in \mathcal{H}_N.
\end{aligned} \tag{SM.10}$$

Condition (21) imposes $N\mathfrak{m}\min\mathcal{H}_N \to \infty$. Let us now consider the case $\mathfrak{m}\min\mathcal{H}_N \to 0$, the arguments for the case $\liminf\{\mathfrak{m}\min\mathcal{H}_N\} > 0$ being quite obvious. Since $1 - \exp(-x) = x\{1 + o(1)\}$ when x decreases to zero, we deduce (SM.1) with $\mathbb{E}[\mathcal{W}_N(t;h)]$ instead of $\mathcal{W}_N(t;h)$. Next, similarly to the justification of (SM.2), we use Chernoff's exponential bound and a grid on $\mathcal{H}_N$ to replace $\mathbb{E}[\mathcal{W}_N(t;h)]$ by $\mathcal{W}_N(t;h)$. The property (SM.1) follows, and we thus complete the proof of Theorem 2.

Finally, in order to justify (A.4), with a uniform kernel, let us note that by definition

$$\mathcal{W}_N(t;h)\mathcal{N}_\mu(t;h)^{-1} = \frac{1}{\mathcal{W}_N(t;h)}\sum_{i=1}^N w_i(t;h)\max_{1\leq m\leq M_i}|W_m^{(i)}(t;h)|$$

and that

$$w_i(t;h)\max_{1\leq m\leq M_i}|W_m^{(i)}(t;h)| = w_i(t;h)\min\{1,(M_ih)^{-1}\} \leq C_L^{-1}w_i(t;h).$$

(Up to a change of the constant, the last inequality can be also obtained with any common kernel which stays away from zero on a compact interval.) We deduce

$$\begin{aligned}\mathcal{N}_\mu(t;h)^{-1} \leq C_L^{-1}\min\{1,(\mathfrak{m}h)^{-1}\} \times \mathcal{W}_N(t;h)^{-1} \\ \leq (\mathfrak{c}_1 C_L)^{-1}\frac{\min\{1,(\mathfrak{m}h)^{-1}\}}{N\min\{1,\mathfrak{m}h\}}\{1+o_{\mathbb{P}}(1)\} \\ = (\mathfrak{c}_1 C_L)^{-1}\frac{1}{N\mathfrak{m}h}\{1+o_{\mathbb{P}}(1)\},\end{aligned}$$

with the $o_{\mathbb{P}}(1)$ rate uniform with respect to $h \in \mathcal{H}_N$. This and (SM.1) implies (A.4). Now, the justification of Theorem 2 is complete.

Proof (Complements for the proof of Lemma 1). Here, we provide a formal justification for the following property: for any t and $u, v \in \mathcal{O}_*(t)$, we have

$$\mathbb{E}\left[(X_u - X_v)^2\right] \approx \{A'(t)\}^{2H_t}|u-v|^{2H_t}.$$

The precise meaning of this approximation of the second order moments of the increments is described in (H2). First, let us notice that, a constant C exists, such that

$$0 \leq \frac{1}{2} - D(H_u, H_v) \leq C|H_t'|^2|u-v|^2, \qquad \forall u, v \in \mathcal{O}_*(t). \tag{SM.11}$$

To prove this double inequality, let us first note that the map $(x, y) \mapsto D(x, y)$ admits partial derivatives of any order on $(0,1)\times(0,1)$. Next, let

$$g(x) = \log(\Gamma(2x+1)) + \log(\sin(\pi x)) =: g_1(x) - g_2(x).$$

We notice that $g''(x) < 0$, for any $x \in (0,1)$. Indeed, using the expression of the derivative of the digamma function, cf. Abramowitz and Stegun (1964, page 260), we have

$$g''(x) = 4\sum_{k\geq 0}\frac{1}{(2x+1+k)^2} - \frac{\pi^2}{\sin^2(\pi x)} = g_1''(x) - g_2''(x).$$

We deduce that g'' is decreasing on $[1/2, 1)$ and, since $g''(0+) = -\infty$, the function g_1'' is decreasing on $(0, 1/2]$ with

$$g_1''(0) = \pi^2/6, \quad g_1''(1/4) = 4\{\pi^2/2 - 4\}, \quad g_2''(1/2) = 4\{\pi^2/6 - 1/4\},$$

and the function g_2'' is decreasing on $(0, 1/2]$ with

$$g_2''(0+) = \infty, \quad g_2''(1/4) = 2\pi^2, \quad g_2''(1/2) = \pi^2,$$

we conclude that $g'' < 0$ on $(0, 1]$. In other words, $x \mapsto g(x)$ is log-concave, and thus

$$2D(x,y) = \frac{\sqrt{\exp(g(x)) \times \exp(g(y))}}{\exp(g((x+y)/2))} < 1, \qquad \forall 0 < x \neq y \leq 1.$$

The left-hand side of (SM.11) now follows. Next, since, $2D(x, x) \equiv 1$, we deduce that, for any $x \in (0,1)$ the first order derivative of $y \mapsto D(x, y)$ is equal to zero at $y = x$. Then, by Taylor expansion, given a small value $r > 0$, a constant $C_{x,r}$ exists, depending on x and r, such that

$$\frac{1}{2} - D(x,y) = D(x,x) - D(x,y) \leq C_{x,r}|x-y|^2, \qquad \forall 0 \leq |x-y| \leq r.$$

Finally, use the fact that $|H_u - H_v| \approx |H'_u||u-v|$ when $u-v$ is close to zero, and deduce the right-hand side of (SM.11).

For any t and $u, v \in \mathcal{O}_*(t)$, let us now write

$$\begin{aligned}
\mathbb{E}\left[(X_u - X_v)^2\right] &= \mathbb{E}\left(X_u^2\right) + \mathbb{E}\left(X_v^2\right) - 2\mathbb{E}\left(X_u X_v\right) \\
&= A(u)^{2H_u} + A(v)^{2H_v} - 2D(H_u, H_v)\left[A(u)^{H_u+H_v} + A(v)^{H_u+H_v} - |A(v) - A(u)|^{H_u+H_v}\right] \\
&= \left\{A(u)^{2H_u} - 2D(H_u, H_v)A(u)^{H_u+H_v}\right\} + \left\{A(v)^{2H_v} - 2D(H_u, H_v)A(v)^{H_u+H_v}\right\} \\
&\quad + 2D(H_u, H_v)|A(v) - A(u)|^{H_u+H_v} =: D_1(u|v) + D_1(v|u) + 2D_2(u, v).
\end{aligned}$$

Next, let $\mathcal{T} \subset [0, \infty)$ be a compact interval, and for any real-valued function B defined on $\mathcal{T}$, let $\|B\|_{\mathcal{T},\infty} = \sup_{t\in\mathcal{T}} B(t)$. In the case $t > 0$, where A stays away from zero on $\mathcal{O}_*(t)$, we can write

$$D_1(u|v) = A(u)^{2H_u} - A(u)^{H_u+H_v} + R_1(u|v),$$

with

$$|R_1(u|v)| \leq \{1 - 2D(H_u, H_v)\} \|A^H\|^2_{\mathcal{T},\infty} \leq C\|A^H\|^2_{\mathcal{T},\infty}\|H'\|^2_{\mathcal{T},\infty}|u-v|^2 = O(|u-v|^2),$$

and

$$\begin{aligned}
A(u)^{2H_u} - A(u)^{H_u+H_v} &= A(u)^{2H_u}\left[1 - \exp\{(H_v - H_u)\log(A(u))\}\right] \\
&= A(u)^{2H_u}\left[-(H_v - H_u)\log(A(u)) + O(|u-v|^2)\right] \\
&= A(u)^{2H_u}\left[-H'_u(v-u)\log(A(u)) + O(|u-v|^2)\right].
\end{aligned}$$

The term $D_1(v|u)$ decomposed similarly, and we thus deduce

$$\begin{aligned}
D_1(u|v) + D_1(v|u) &= (v-u)\left[A(v)^{2H_v}H'_v\log(A(v)) - A(u)^{2H_u}H'_u\log(A(u))\right] + O(|u-v|^2) \\
&= O(|u-v|^2).
\end{aligned}$$

The last equality is due to the fact that, by assumptions, the map $v \mapsto A(v)^{2H_v}H'_v\log(A(v))$ is continuously differentiable on $\mathcal{O}_*(t)$. On the other hand, by (SM.11),

$$\begin{aligned}
D_2(u, v) &= |A(v) - A(u)|^{H_u+H_v} + \sup_{t\in\mathcal{T}}|H'_t| \times o(|u-v|^2) \\
&= \left|A'(t)(v-u) + O(|u-v|^2)\right|^{2H_t + 2H'_t(v-u) + O(|u-v|^2)} + o(|u-v|^2) \\
&= |A'(t)|^{2H_t}|v-u|^{2H_t} \times \left\{1 + O\left(|u-v|^{\min_{t\in\mathcal{T}} H_t}\right)\right\} + o(|u-v|^2).
\end{aligned}$$

Thus, (H2) is satisfied with any $0 < \beta \leq \min\{\min_{t\in\mathcal{T}} H_t, 2(1 - \max_{t\in\mathcal{T}} H_t)\}$.

In the case $t = 0$, we only have to investigate the case $A(0) = 0$. Whenever $A(0) > 0$, the previous arguments apply without any change. If $A(0) = 0$, we only have to revisit the arguments for bounding $D_1(u|v) + D_1(v|u)$. Now, the map

$$v \mapsto \zeta(v) = A(v)^{2H_v}H'_v\log(A(v)),$$

is no longer differentiable on $\mathcal{O}_*(0)$, if $H_v \leq 1/2$. However, this map is Hölder continuous, and we still have

$$D_1(u|v) + D_1(v|u) = O\left(|u-v|^{1+\gamma}\right), \quad \text{for any } 0 < \gamma < \min(1, 2\min_{v\in\mathcal{O}_*(0)} H_v),$$

and, given the regularity conditions imposed on the map $u \mapsto H_u$, this suffices to complete the proof of Lemma 1. With the rule $0\log(0) = 0$, the Hölder continuity we used is

$$\sup_{0 \leq u < v \leq \Delta_*/2} \frac{|\zeta(u) - \zeta(v)|}{|u - v|^{\gamma}} \leq C < \infty, \tag{SM.12}$$

for some C depending on γ, H and Δ_* and decreasing to zero with Δ_*. Indeed, under our assumptions, we have

$$\zeta(v) = [A'(0)v]^{2H_0} \times \zeta_1(v) \times H_0' \times \{\log(v) + \log(A'(0)\} \times \{1 + o(1)\},$$

with

$$\zeta_1(v) = v^{2H_0' v\{1+o(1)\}}.$$

All the $o(1)$ terms in the last display can be uniformly bounded, with respect to $v \in \mathcal{O}_*(0)$, by a constant times Δ_*, the constant depending only on the bounds of A', $|H'|$ and $|H''|$ near the origin. First, let us notice that for any $0 < \gamma < \min(1, 2H_0)$, a constant c exists such that

$$\sup_{0 \leq u < v \leq \Delta_*/2} \frac{\left|u^{2H_0}\log(u) - v^{2H_0}\log(v)\right|}{|u - v|^{\gamma}} \leq c < \infty.$$

Moreover, we have $\zeta_1(0+) = 1$, ζ_1 is bounded on $\mathcal{O}_*(0)$, and for $0 < u < v \leq \Delta_*/2$, we have

$$\begin{aligned}|\zeta_1(v) - \zeta_1(u)| &= |\exp(2H_0' v\log(v)\{1 + o(1)\}) - \exp(2H_0' u\log(u)\{1 + o(1)\}))\\ &\leq 2|H_0'||v\log(v) - u\log(u)| \leq 2c_1|H_0'||u - v|^{\gamma},\end{aligned}$$

for some constant c_1. Gathering facts, we deduce (SM.12). The justification of Lemma 1 is now complete.

B. Details on diverse quantities from the main text

B.1. Details on the approximation (16)

Recall that

$$c_i(t; h, \alpha) = \sum_{m=1}^{M_i} \left|(T_m^{(i)} - t)/h\right|^{\alpha} \left|W_m^{(i)}(t; h)\right|,$$

and

$$\overline{C}(t; h, \alpha) = \frac{1}{\mathcal{W}_N(t; h)} \sum_{i=1}^{N} w_i(t; h) c_i(t; h, \alpha).$$

When using the Nadaraya-Watson (NW) estimator, for each $1 \leq i \leq N$,

$$c_i(t; h, \alpha) = \frac{1}{\widehat{g}^{(i)}(t)} \frac{1}{M_i h} \sum_{m=1}^{M_i} \left|(T_m^{(i)} - t)/h\right|^{\alpha} K\left((T_m^{(i)} - t)/h\right),$$

with

$$\widehat{g}^{(i)}(t) = \frac{1}{M_i h} \sum_{m=1}^{M_i} K\left((T_m^{(i)} - t)/h\right) \approx g(t).$$

Here, g denotes the density of the $T_m^{(i)}$. By a standard change of variables,

$$\mathbb{E}[c_i(t;h,\alpha)\widehat{g}^{(i)}(t)] \approx g(t) \int |u|^\alpha K(u) du.$$

and this explains our proposal

$$\overline{C}(t;h,\alpha) \approx \int |u|^\alpha K(u) du, \tag{SM.13}$$

for the NW estimator. The same arguments apply for $\overline{\mathfrak{C}}(t|s;h,\alpha)$ used for estimating the covariance function. In the case of a local linear estimator, it suffices to use the equivalent kernels for local polynomial smoothing. Approximation (SM.13) could remain the same in the local linear case, but has to be changed for higher-order polynomials. See section 3.2.2 in Fan and Gijbels (1996).

B.2. Details on the definition (28)

Recall that $\widetilde{\gamma}_N(s,t) = N^{-1}\sum_{i=1}^N X_s^{(i)} X_t^{(i)}$. Here, $\mathcal{W}_N$ and w_i are short notations for $\mathcal{W}_N(s,t;h)$ and $w_i(s,t;h)$, respectively. Moreover, $\widehat{X}_t^{(i)} - X_t^{(i)} = B_t^{(i)} + V_t^{(i)}$, where $B_t^{(i)} := \mathbb{E}_i\left[\widehat{X}_t^{(i)}\right] - X_t^{(i)}$ and $V_t^{(i)} := \widehat{X}_t^{(i)} - \mathbb{E}_i\left[\widehat{X}_t^{(i)}\right]$. Let us define

$$\widetilde{\gamma}_W(s,t;h) = \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i X_s^{(i)} X_t^{(i)}.$$

To explain our empirical risk bound $\mathcal{R}_\Gamma(s|t;h)$ defined in (28), let us write

$$\begin{aligned}
\widehat{\gamma}_N(s,t;h) - \widetilde{\gamma}_W(s,t;h) &= \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\{\widehat{X}_s^{(i)} - X_s^{(i)}\}X_t^{(i)} + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i X_s^{(i)}\{\widehat{X}_t^{(i)} - X_t^{(i)}\} \\
&\quad + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\{\widehat{X}_s^{(i)} - X_s^{(i)}\}\{\widehat{X}_t^{(i)} - X_t^{(i)}\} \\
&= \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left\{B_s^{(i)} X_t^{(i)} + X_s^{(i)} B_t^{(i)}\right\} \\
&\quad + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left\{V_s^{(i)} X_t^{(i)} + X_s^{(i)} V_t^{(i)}\right\} \\
&\quad + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left\{B_s^{(i)} B_t^{(i)} + V_s^{(i)} V_t^{(i)}\right\} \\
&\quad + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left\{B_s^{(i)} V_t^{(i)} + V_s^{(i)} B_t^{(i)}\right\}.
\end{aligned}$$

By construction,

$$\mathbb{E}_{M,T}\left\{V_s^{(i)} B_t^{(j)}\right\} = \mathbb{E}_{M,T}\left\{B_s^{(i)} V_t^{(j)}\right\} = 0, \quad \forall 1 \le i,j \le N.$$

Moreover, whenever $h < |s-t|$, we have

$$\mathbb{E}_{M,T}\left\{V_s^{(i)}V_t^{(j)}\right\} = 0, \quad \forall 1 \leq i,j \leq N.$$

Using these properties, the inequality $(a+b)^2 \leq 2(a^2+b^2)$, and repeated application of Cauchy-Schwarz inequality to check the negligible terms, we deduce

$$\begin{aligned}
\mathbb{E}_{M,T}\left[\{\widehat{\gamma}_N(s,t;h) - \widetilde{\gamma}_W(s,t;h)\}^2\right] &= \mathbb{E}_{M,T}\left[\left\{\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left(B_s^{(i)}X_t^{(i)} + X_s^{(i)}B_t^{(i)}\right)\right\}^2\right] \\
&+ \mathbb{E}_{M,T}\left[\left\{\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left\{V_s^{(i)}X_t^{(i)} + X_s^{(i)}V_t^{(i)}\right\}\right\}^2\right] + \text{ negligible terms} \\
&\leq 2\mathbb{E}_{M,T}\left[\left\{\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i B_s^{(i)}X_t^{(i)}\right\}^2\right] + 2\mathbb{E}_{M,T}\left[\left\{\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i X_s^{(i)}B_t^{(i)}\right\}^2\right] \\
&+ \frac{1}{\mathcal{W}_N^2}\sum_{i=1}^N w_i\mathbb{E}_{M,T}\left[\left\{V_s^{(i)}X_t^{(i)}\right\}^2 + \left\{X_s^{(i)}V_t^{(i)}\right\}^2\right] + \text{ negligible terms} \\
&= \{G_1(s|t) + G_1(t|s) + G_2\}\{1 + o_{\mathbb{P}}(1)\}.
\end{aligned}$$

We can now write

$$\begin{aligned}
\mathbb{E}_{M,T}\left[\left\{X_s^{(i)}V_t^{(i)}\right\}^2\right] &= \mathbb{E}_{M,T}\left[\left\{X_s^{(i)}\right\}^2\left\{\sum_{m=1}^{M_i}\varepsilon_m^{(i)}W_m^{(i)}(t;h)\right\}^2\right] \\
&= \mathbb{E}_{M,T}\left[\left\{X_s^{(i)}\right\}^2\sum_{m=1}^{M_i}\mathbb{E}_i\left\{\left|\varepsilon_m^{(i)}\right|^2\right\}\left|W_m^{(i)}(t;h)\right|^2\right] \\
&\leq \sigma_{\max}^2 m_2(s)\left\{\max_m\left|W_m^{(i)}(t)\right| \times \sum_{m=1}^{M_i}\left|W_m^{(i)}(t;h)\right|\right\},
\end{aligned}$$

where $m_2(s) = \mathbb{E}\left[X_s^2\right]$ and $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot \mid M_i, \mathcal{T}_{obs}^{(i)}, X^{(i)})$. Let us recall that

$$\mathcal{N}_i(t|s;h) = \frac{w_i(s,t;h)}{\max_{1\leq m\leq M_i}|W_m^{(i)}(t;h)|} \quad \text{and} \quad \mathcal{N}_i(s|t;h) = \frac{w_i(s,t;h)}{\max_{1\leq m\leq M_i}|W_m^{(i)}(s;h)|}. \tag{SM.14}$$

We deduce

$$G_2 \leq \frac{\sigma_{\max}^2}{\mathcal{W}_N^2}\sum_{i=1}^N w_i\left[m_2(t)\frac{c_i(s;h)}{\mathcal{N}_i(s|t;h)} + m_2(s)\frac{c_i(t;h)}{\mathcal{N}_i(t|s;h)}\right],$$

where the $c_i(t;h)$ are defined by (15) and the $\mathcal{N}_i(s|t;h)$ and $\mathcal{N}_i(t|s;h)$ are defined using (SM.14).

To bound the terms related to the bias of $\widehat{X}_t^{(i)}$, moment assumptions, by the law of large numbers, dominated convergence theorem, we can write

$$
\begin{aligned}
G_1(s|t) + G_1(t|s) &\leq 2\mathbb{E}_{M,T}\left[\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i \left|B_t^{(i)}\right|^2 \times \left\{m_2(s) + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left(\left|X_s^{(i)}\right|^2 - m_2(s)\right)\right\}\right] \\
&+ 2\mathbb{E}_{M,T}\left[\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i \left|B_s^{(i)}\right|^2 \times \left\{m_2(t) + \frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i\left(\left|X_t^{(i)}\right|^2 - m_2(t)\right)\right\}\right] \\
&= 2\left[\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i \mathbb{E}_{M,T}\left\{\left|B_s^{(i)}\right|^2\right\}\right] m_2(t) + 2\left[\frac{1}{\mathcal{W}_N}\sum_{i=1}^N w_i \mathbb{E}_{M,T}\left\{\left|B_t^{(i)}\right|^2\right\}\right] m_2(s) \\
&\qquad + \text{ negligible terms} \\
&\leq \left\{2m_2(t)\overline{\mathfrak{C}}(s|t;h,2\widehat{H}_s)\hat{L}_s^2 + 2m_2(s)\overline{\mathfrak{C}}(t|s;h,2\widehat{H}_t)\hat{L}_t^2\right\}\{1+o_{\mathbb{P}}(1)\},
\end{aligned}
$$

where $\overline{\mathfrak{C}}(t|s;h,\cdot)$ is defined according to (29).

Gathering facts, we deduce that

$$
\begin{aligned}
&\mathbb{E}_{M,T}\left[\{\widehat{\gamma}_N(s,t;h) - \widetilde{\gamma}_W(s,t;h)\}^2\right] \\
&\quad \leq 2E^2(t)\overline{\mathfrak{C}}(s|t;h,2\widehat{H}_s)\hat{L}_s^2 + 2E^2(s)\overline{\mathfrak{C}}(t|s;h,2\widehat{H}_t)\hat{L}_t^2 \\
&\qquad + \frac{\sigma_{\max}^2}{\mathcal{W}_N^2}\sum_{i=1}^N w_i\left[m_2(t)\frac{c_i(s;h)}{\mathcal{N}_i(s|t;h)} + m_2(s)\frac{c_i(t;h)}{\mathcal{N}_i(t|s;h)}\right] + \text{ negligible terms.}
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\mathbb{E}_{M,T}\left[\{\widetilde{\gamma}_N(s,t) - \widetilde{\gamma}_W(s,t;h)\}^2\right] &= \frac{\operatorname{Var}(X_sX_t)}{\mathcal{W}_N^2}\sum_{i=1}^N\left\{w_i - \frac{\mathcal{W}_N}{N}\right\}^2 \\
&= \operatorname{Var}(X_sX_t)\left\{\frac{1}{\mathcal{W}_N} - \frac{1}{N}\right\}.
\end{aligned}
$$

It remains to note that

$$
\begin{aligned}
\mathbb{E}_{M,T}\left[\{\widehat{\gamma}_N(s,t;h) - \widetilde{\gamma}_N(s,t)\}^2\right] &\leq 2\mathbb{E}_{M,T}\left[\{\widehat{\gamma}_N(s,t;h) - \widetilde{\gamma}_W(s,t;h)\}^2\right] \\
&\quad + 2\mathbb{E}_{M,T}\left[\{\widetilde{\gamma}_W(s,t;h) - \widetilde{\gamma}_N(s,t)\}^2\right].
\end{aligned}
$$

C. Proof of Theorem 3

Below, $c, C, \mathfrak{c}\ldots$, are constants which may change from line to line, and are not necessarily the same in other proofs. For simplicity, we assume $\widehat{\Gamma}_N^*$ is built with the uniform kernel. Recall that $s \neq t$ are fixed and without loss of generality, we consider $\sup \mathcal{H}_N < |s-t|/2$. We can also assume $C_L\mathfrak{m} \geq 2$.

First, we prove that

$$
\frac{1}{\mathcal{W}_N(s,t;h)} - \frac{1}{N} \leq \min\left[\min\left\{h^{2H_s},\ \mathcal{N}_\Gamma^{-1}(s|t;h)\right\}, \min\left\{h^{2H_t},\ \mathcal{N}_\Gamma^{-1}(t|s;h)\right\}\right] O_{\mathbb{P}}(1), \quad \text{(SM.15)}
$$

uniformly with respect to $h \in \mathcal{H}_N$. For this purpose, we start by showing that there exists a constant $\mathfrak{c}_1 > 0$ such that

$$\inf_{h\in\mathcal{H}_N} \frac{N\min\{\mathfrak{m}h, (\mathfrak{m}h)^2\}}{\mathcal{N}_\Gamma(t|s;h)} \geq \mathfrak{c}_1\{1+o_{\mathbb{P}}(1)\} \quad \text{and} \quad \inf_{h\in\mathcal{H}_N} \frac{N\min\{\mathfrak{m}h, (\mathfrak{m}h)^2\}}{\mathcal{N}_\Gamma(s|t;h)} \geq \mathfrak{c}_1\{1+o_{\mathbb{P}}(1)\}.$$

Using the fact that the harmonic mean is less than or equal to the mean, we obtain

$$\frac{1}{\mathcal{N}_\Gamma(t|s;h)} \geq \frac{c_i(t;h)}{\sum_{i=1}^N w_i(s;h)w_i(t;h)\mathcal{N}_i(t|s;h)}, \tag{SM.16}$$

with $w_i(t;h)$, $c_i(t;h)$ and $\mathcal{N}_i(t|s;h)$ defined in (13), (15) and (SM.14), respectively. In the case we consider, for all i, we have $c_i \equiv 1$. To justify (SM.16), it suffices to prove that a positive constant $c_\mathcal{N}$ exists such that

$$\frac{\sum_{i=1}^N w_i(s;h)w_i(t;h)\mathcal{N}_i(t|s;h)}{N\min\{\mathfrak{m}h, (\mathfrak{m}h)^2\}} \leq c_\mathcal{N}\{1+o_{\mathbb{P}}(1)\}, \tag{SM.17}$$

with the $o_{\mathbb{P}}(1)$ uniform with respect to $h \in \mathcal{H}_N$. Let us notice that in the case of a NW estimator with a uniform kernel,

$$\sum_{i=1}^N w_i(s;h)w_i(t;h)\mathcal{N}_i(t|s;h) = \sum_{i=1}^N w_i(s;h)\sum_{m=1}^{M_i}\mathbf{1}\{|T_m^{(i)}-t|\leq h\} = \sum_{i=1}^N S^{(i)},$$

with

$$S^{(i)} = S^{(i)}(h) = w_i(s;h)\sum_{1\leq m\leq M_i}\mathbf{1}\{|T_m^{(i)}-t|\leq h\}.$$

We thus need to suitably bound from above the sum of $S^{(i)}(h)$. Let

$$\mathbb{P}_M, \quad \mathbb{E}_M, \quad \text{and} \quad \operatorname{Var}_M,$$

denote the conditional probability, expectation and variance, respectively, given $M_1, \dots, M_N$. By Cauchy-Schwarz inequality and (23), we have

$$\begin{aligned}
\mathbb{E}_M[S^{(i)}] &= \sum_{m=1}^{M_i}\mathbb{E}_M\left[w_i(s;h)\mathbf{1}\{|T_m^{(i)}-t|\leq h\}\right] \\
&= \sum_{m=1}^{M_i}\mathbb{E}_M\left[\mathbf{1}\left\{\sum_{1\leq m'\neq m\leq M_i}\mathbf{1}\{|T_{m'}^{(i)}-s|\leq h\}\geq 1\right\}\mathbf{1}\{|T_m^{(i)}-t|\leq h\}\right] \\
&= \sum_{m=1}^{M_i}\mathbb{E}_M\left[\mathbf{1}\left\{\sum_{1\leq m'\neq m\leq M_i}\mathbf{1}\{|T_{m'}^{(i)}-s|\leq h\}\geq 1\right\}\right]\times\mathbb{E}_M\left[\mathbf{1}\{|T_m^{(i)}-t|\leq h\}\right] \\
&= \left[1-\{1-p_i(t;h)\}^{M_i-1}\right]\times M_ip_i(t;h) \\
&= \{1+o(1)\}\times\pi_i(s;h)\times M_ip_i(t;h),
\end{aligned}$$

where $p_i(t;h) = \int_{t-h}^{t+h} g_i(u)du$ and $\pi(s;h)$ is defined as in (SM.7). The $o(1)$ term is uniform with respect to h. Moreover,

$$\{S^{(i)}\}^2 - S^{(i)} = w_i(s;h) \sum_{1\leq m'\neq m\leq M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\}\mathbf{1}\{|T_{m'}^{(i)} - t| \leq h\}$$

$$= \mathbf{1}\left\{ \sum_{1\leq m''\leq M_i, m''\notin\{m,m'\}} \mathbf{1}\{|T_{m'}^{(i)} - s| \leq h\} \geq 1 \right\} \sum_{1\leq m'\neq m\leq M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\}\mathbf{1}\{|T_{m'}^{(i)} - t| \leq h\},$$

and thus,

$$\begin{aligned}\mathbb{E}_M[\{S^{(i)}\}^2] &= \mathbb{E}_M[S^{(i)}] + \left[1 - \{1-p_i(t;h)\}^{M_i-2}\right] \times M_i(M_i-1)p_i^2(t;h) \\ &= \{1+o(1)\} \times \pi_i(s;h) \times M_i p_i(t;h) \times \{1 + M_i p_i(t;h)\}, \qquad \text{(SM.18)}\end{aligned}$$

with the $o(1)$ term uniform with respect to h. We deduce that

$$\begin{aligned}\mathrm{Var}_M[S^{(i)}] &= \mathbb{E}_M[\{S^{(i)}\}^2] - \mathbb{E}_M^2[S^{(i)}] \\ &= \{1+o(1)\} \times \pi_i(s;h)M_i p_i(t;h) \times [1 + M_i p_i(t;h) - \pi_i(s;h)M_i p_i(t;h)] \\ &= \{1+o(1)\} \times \pi_i(s;h)M_i p_i(t;h) + \{1+o(1)\} \times \pi_i(s;h)\{1-\pi_i(s;h)\}\{M_i p_i(t;h)\}^2.\end{aligned}$$

Let us introduce the following notation: given φ_1, φ_2, positive functions of M_i and h,

$$\varphi_1 \lesssim \varphi_2 \quad \Leftrightarrow \quad \exists C>0 \ \text{ a constant such that } \quad \varphi_1 \leq C\varphi_2,$$

and

$$\varphi_1 \asymp \varphi_2 \quad \Leftrightarrow \quad \varphi_1 \lesssim \varphi_2 \quad \text{and} \quad \varphi_2 \lesssim \varphi_1.$$

With this notation,

$$\mathbb{E}_M[S^{(i)}] \asymp \pi_i(s;h) \times \mathfrak{m}h,$$

and

$$\mathbb{E}_M\left[\sum_{i=1}^N S^{(i)}\right] \asymp \mathcal{W}_N(s;h) \times \mathfrak{m}h,$$

and thus, by (SM.10),

$$N\mathfrak{m}h\left\{1 - \exp(-2C_{g,L}C_L\mathfrak{m}h)\right\} \lesssim \mathbb{E}_M\left[\sum_{i=1}^N S^{(i)}\right] \lesssim N\mathfrak{m}h\left\{1 - \exp(-4C_{g,U}C_U\mathfrak{m}h)\right\},$$

$\forall h \in \mathcal{H}_N$. On the other hand,

$$\mathrm{Var}_M[S^{(i)}] \asymp \pi_i(s;h) \times \mathfrak{m}h + \pi_i(s;h)\{1-\pi_i(s;h)\} \times (\mathfrak{m}h)^2.$$

By (SM.9), we deduce

$$\begin{aligned}\{1 - \exp(-2C_{g,L}C_L\mathfrak{m}h)\}\, \mathfrak{m}h\{1 + \exp(-4C_{g,U}C_U\mathfrak{m}h)\mathfrak{m}h\} &\lesssim \mathrm{Var}_M[S^{(i)}] \\ &\lesssim \{1 - \exp(-4C_{g,U}C_U\mathfrak{m}h)\}\, \mathfrak{m}h\{1 + \exp(-2C_{g,L}C_L\mathfrak{m}h)\mathfrak{m}h\}.\end{aligned}$$

Since for any $c>0$, the map $x \mapsto x\exp(-cx)$, $x \geq 0$ is bounded, we deduce

$$\{1 - \exp(-2C_{g,L}C_L\mathfrak{m}h)\}\, \mathfrak{m}h \lesssim \mathrm{Var}_M[S^{(i)}] \lesssim \{1 - \exp(-4C_{g,U}C_U\mathfrak{m}h)\}\, \mathfrak{m}h.$$

Let us note that

$$\mathbb{E}_M\left[S^{(i)}\right] \asymp \operatorname{Var}_M[S^{(i)}] \asymp \mathfrak{m}h \times \min\{1, \mathfrak{m}h\}. \tag{SM.19}$$

It remains to show that the sum of $S^{(i)}(h)$ concentrates around a quantity which allows to deduce (SM.17). Let $A = A(h) > 0$ to be determined below, and let

$$\mathcal{A} = \mathcal{A}(h) = \left\{\max_{1\leq i\leq N} S^{(i)} \leq A\right\} \quad \text{and} \quad S_A^{(i)} = S_A^{(i)}(h) = S^{(i)}\mathbf{1}_{\mathcal{A}}.$$

Let

$$E_{M,A} = E_{M,A}(h) := \mathbb{E}_M\left[\sum_{i=1}^N S_A^{(i)}\right] \leq \mathbb{E}_M\left[\sum_{i=1}^N S^{(i)}\right] \quad \text{and} \quad V_{M,A} = V_{M,A}(h) := \operatorname{Var}_M\left[\sum_{i=1}^N S_A^{(i)}\right].$$

By definition,

$$V_{M,A} \lesssim \sum_{1=1}^N \operatorname{Var}_M[S^{(i)}] \asymp N\mathfrak{m}h \times \min\{1, \mathfrak{m}h\} =: \Omega_N(h) \to \infty.$$

Indeed, we have

$$\begin{aligned}\operatorname{Var}_M[S_A^{(i)}] = \mathbb{E}_M[\{S^{(i)}\}^2] - \mathbb{E}_M[\{S^{(i)}\}^2\mathbf{1}_{\overline{\mathcal{A}}}] - \left\{\mathbb{E}_M[S^{(i)}] - \mathbb{E}_M\left[S^{(i)}\mathbf{1}_{\overline{\mathcal{A}}}\right]\right\}^2 \\ \leq \operatorname{Var}_M[S^{(i)}] + 2\mathbb{E}_M[S^{(i)}]\mathbb{E}_M\left[S^{(i)}\mathbf{1}_{\overline{\mathcal{A}}}\right].\end{aligned}$$

Herein, for any set B, $\overline{B}$ denotes its complement. By (SM.19), we deduce

$$\operatorname{Var}_M[S_A^{(i)}] \lesssim \operatorname{Var}_M[S^{(i)}],$$

provided a constant exists such that $\mathbb{E}_M[S^{(i)}\mathbf{1}_{\overline{\mathcal{A}}}] \leq C$ for all $\mathfrak{m}$ and h. By Cauchy-Schwarz inequality and (SM.18),

$$\mathbb{E}_M\left[S^{(i)}(h)\mathbf{1}_{\overline{\mathcal{A}}(h)}\right] \leq \mathbb{E}_M^{1/2}[\{S^{(i)}(h)\}^2] \times \mathbb{P}(\overline{\mathcal{A}}(h)) \lesssim \mathfrak{m}h \times \mathbb{P}(\overline{\mathcal{A}}) \leq \mathfrak{m} \times \mathbb{P}(\overline{\mathcal{A}}(\min \mathcal{H}_N)) \to 0.$$

The convergence to zero follows from (SM.20) below. Next, by Bernstein inequality applied to the $S_A^{(i)}$'s, for each $h \in \mathcal{H}_N$,

$$\mathbb{P}_M\left[\sum_{i=1}^N S_A^{(i)}(h) > E_{M,A}(h) + \Omega_N(h)\right] \leq \exp\left(-\frac{\Omega_N(h)^2/2}{V_{M,A}(h) + A(h)\Omega_N(h)/3}\right).$$

To derive bounds for the concentration probability of the sum of $S^{(i)}(h)$, it suffices to take A such that

$$\sqrt{V_{M,A}(h)} \ll \Omega_N(h) \qquad \text{and} \qquad A(h) \ll \Omega_N(h).$$

Let

$$A(h) = \frac{\Omega_N(h)}{c_A \log(\mathfrak{m})},$$

with c_A some large constant. Consider $\mathcal{G}_N$ a uniform a grid in $\mathcal{H}_N$ with mesh of rate $1/N\mathfrak{m}$. By (23), and the taking c_A sufficiently large, we deduce that a constant $0 < C < c_A$ exists such that

$$\mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\frac{1}{\Omega_N(h)}\sum_{i=1}^N S_A^{(i)}(h) > C\right] \leq \mathbb{P}_M\left[\sup_{h\in\mathcal{G}_N}\frac{1}{\Omega_N(h)}\sum_{i=1}^N S_A^{(i)}(h) > C/2\right]$$
$$\leq \exp\left(\log(|\mathcal{G}_N|) - c_A\log(\mathfrak{m})\right) \leq \exp\left(-(c_A - C)\log(\mathfrak{m})\right) \to 0.$$

Here, $|\mathcal{G}_N|$ denotes the cardinal of $\mathcal{G}_N$. Finally, we have

$$\mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\frac{1}{\Omega_N(h)}\sum_{i=1}^N S^{(i)}(h) > C\right] = \mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\frac{1}{\Omega_N(h)}\sum_{i=1}^N S^{(i)}(h)\{\mathbf{1}_{\mathcal{A}(h)} + \mathbf{1}_{\overline{\mathcal{A}}(h)}\} > C\right]$$
$$\leq \mathbb{P}_M\left[\sup_{h\in\mathcal{G}_N}\frac{1}{\Omega_N(h)}\sum_{i=1}^N S_A^{(i)}(h) > C/4\right] + \mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\mathbf{1}_{\overline{\mathcal{A}}(h)} > 0\right]$$
$$\leq \exp\left(-(c_A - C)\log(\mathfrak{m})\right) + \mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\mathbf{1}_{\overline{\mathcal{A}}(h)} > 0\right].$$

Next, let $h_{j(h)}$, with $1 \leq j(h) \leq J$, be the point in the grid $\mathcal{G}_N$ such that $h_{j(h)-1} \leq h < h_{j(h)}$. Using the monotonicity of the $S^{(i)}(h)$ and $\Omega_N(h)$, with respect to h, we then have

$$\left\{\max_{1\leq i\leq N}\frac{S^{(i)}\left(h_{j(h)-1}\right)}{\Omega_N\left(h_{j(h)}\right)} \geq \frac{1}{c_A\log(\mathfrak{m})}\right\} \subset \overline{\mathcal{A}}(h) \subset \left\{\max_{1\leq i\leq N}\frac{S^{(i)}\left(h_{j(h)}\right)}{\Omega_N\left(h_{j(h)-1}\right)} \geq \frac{1}{c_A\log(\mathfrak{m})}\right\}.$$

This implies

$$\mathbb{P}_M\left[\sup_{h\in\mathcal{H}_N}\mathbf{1}_{\overline{\mathcal{A}}(h)} > 0\right] \leq \sum_{j=2}^J \mathbb{P}_M\left[\max_{1\leq i\leq N}\frac{S^{(i)}(h_j)}{\Omega_N\left(h_{j-1}\right)} \geq \frac{1}{c_A\log(\mathfrak{m})}\right]$$
$$\leq \sum_{j=2}^J\sum_{i=1}^N \mathbb{P}_M\left[S^{(i)}(h_j) \geq \frac{\Omega_N(h_{j-1})}{c_A\log(\mathfrak{m})}\right]$$
$$\leq \sum_{j=2}^J\sum_{i=1}^N \mathbb{P}_M\left[\sum_{m=1}^{M_i}\mathbf{1}\{|T_m^{(i)} - t| \leq h_j\} \geq M_i\mathbb{E}[\mathbf{1}\{|T_m^{(i)} - t| \leq h_j\}] \times (1+\delta_{ij})\right]$$
$$\leq J \times N \times \exp\left(-C_L C_{g,L}\min_{1\leq i\leq N}\min_{2\leq j\leq J}\left[\frac{\delta_{ij}}{2+\delta_{ij}} \times \delta_{ij}\mathfrak{m}h_j\right]\right), \quad \text{(SM.20)}$$

where for the last inequality, we used Chernoff's inequality, and C_L and $C_{g,L}$ are the constants in (23) and (SM.3), respectively. Here,

$$\delta_{ij} = \frac{\Omega_N(h_{j-1})/\{c_A\log(\mathfrak{m})\}}{M_i\mathbb{E}[\mathbf{1}\{|T_m^{(i)} - t| \leq h_j\}]} \geq C \times \frac{N\min\{1, \mathfrak{m}h_j\}}{c_A\log(\mathfrak{m})} \geq C\frac{N\min\{1, \mathfrak{m}\min\mathcal{H}_N\}}{c_A\log(\mathfrak{m})} \to \infty,$$

for some constant $C > 0$. Moreover, by the condition $N\{\mathfrak{m}\min\mathcal{H}_N\}^2/\log^2(N\mathfrak{m}) \to \infty$, we have

$$\delta_{ij}\mathfrak{m}h_j \geq C\frac{N\mathfrak{m}\min\mathcal{H}_N\min\{1, \mathfrak{m}\min\mathcal{H}_N\}}{c_A\log(\mathfrak{m})} \gg \log(JN).$$

This implies that the exponential bound in (SM.20) tends to zero. Gathering facts, we deduce (SM.17).

The next step is to prove that, a constant $c_{\mathcal{W}} \in (0,1)$ exists such that

$$\mathfrak{c}_{\mathcal{W}} N\{1+o_{\mathbb{P}}(1)\} \leq \inf_{h\in\mathcal{H}_N} \frac{\mathcal{W}_N(s,t;h)}{\min\{1,(\mathfrak{m}h)^2\}}, \tag{SM.21}$$

with the $o_{\mathbb{P}}(1)$ uniform with respect to $h\in\mathcal{H}_N$. For any $s\neq t$, $w_i(s;h)w_i(t;h)$ is a Bernoulli variable with parameter, say, $\pi_i(s,t;h)$. Let us note that in the case where the intervals $[t-h,t+h]$ and $[s-h,s+h]$ are disjoint, which is our case for each $h\in\mathcal{H}_N$, using the definition of the multinomial distribution, we have

$$\begin{aligned}\pi_i(s,t;h) = \sum_{l+l'=0}^{M_i-2} &\frac{M_i!}{(l+1)!(l'+1)!(M_i-2-(l+l'))!} \\ &\times p_i(s;h)^{l+1}p_i(t;h)^{l'+1}\{1-p_i(s;h)-p_i(t;h)\}^{M_i-2-(l+l')} \\ &\geq \frac{M_i!}{(M_i-2)!}p_i(s;h)p_i(t;h)\{1-p_i(s;h)-p_i(t;h)\}^{M_i-2} =: \underline{\pi}_i(s,t;h).\end{aligned}$$

Using bounds as in (SM.8), we can write

$$\begin{aligned}\min\{1,(\mathfrak{m}h)^2\} \asymp C_L^2(\mathfrak{m}-1)^2h^2 \times \exp(-4C_{g,L}C_L(\mathfrak{m}-2)h) \lesssim \underline{\pi}_i(s,t;h) \\ \lesssim C_U^2\mathfrak{m}^2h^2 \times \exp(-8C_{g,U}C_L(\mathfrak{m}-2)h) \asymp \min\{1,(\mathfrak{m}h)^2\}.\end{aligned}$$

and from this we deduce that, a constant $c_{\mathcal{W}}$ exists such that

$$\inf_{h\in\mathcal{H}_N} \frac{\mathbb{E}_M[\mathcal{W}_N(s,t;h)]}{\min\{1,(\mathfrak{m}h)^2\}} \geq \mathfrak{c}_{\mathcal{W}} N.$$

Since $\mathcal{W}_N(s,t;h)$ is a sum of independent Bernoulli variables, next, we proceed as above, applying Chernoff or Bernstein inequalities for a grid of bandwidths h, to derive exponential bounds for the concentration probability of $\mathcal{W}_N(s,t;h)$. Then, (SM.21) follows. The arguments have already been used above, and we thus omit the details.

Gathering facts, we deduce (SM.15). Finally, the proof of Theorem 3 can be completed as follows. Using the definition of $\mathcal{N}_\Gamma(t|s;h)$,

$$\begin{aligned}\mathcal{N}_\Gamma(t|s;h)^{-1} \leq C_L^{-1}\min\{1,(\mathfrak{m}h)^{-1}\}\times\mathcal{W}_N(s,t;h)^{-1} \leq (\mathfrak{c}_{\mathcal{W}}C_L)^{-1}\frac{\min\{1,(\mathfrak{m}h)^{-1}\}}{N\min\{1,(\mathfrak{m}h)^2\}}\{1+o_{\mathbb{P}}(1)\} \\ = (\mathfrak{c}_{\mathcal{W}}C_L)^{-1}\frac{1}{N\min\{\mathfrak{m}h,(\mathfrak{m}h)^2\}}\{1+o_{\mathbb{P}}(1)\},\end{aligned}$$

with the $o_{\mathbb{P}}(1)$ rate uniform with respect to $h\in\mathcal{H}_N$. Then, by arguments similar to those used for (A.4) in the main text (see also the end of the complements to Theorem 2 above), we obtain

$$\min\left\{h^{2H_t}+\mathcal{N}_\Gamma^{-1}(t;h)\right\} = O_{\mathbb{P}}\left(h^{2H_t}+(N\min\{\mathfrak{m}h,(\mathfrak{m}h)^2\})^{-1}\right).$$

Let us note that in the case of common design, we have

$$\begin{aligned}\widehat{\Gamma}_N^*(s,t)-\Gamma(s,t) = O_{\mathbb{P}}\left(\max\left\{(N\mathfrak{m}^2)^{-\frac{H(s,t)}{2\{H(s,t)+1\}}},(N\mathfrak{m})^{-\frac{H(s,t)}{2H(s,t)+1}},\mathfrak{m}^{-H(s,t)}\right\}+N^{-1/2}\right) \\ = O_{\mathbb{P}}\left(\mathfrak{m}^{-H(s,t)}+N^{-1/2}\right),\end{aligned}$$

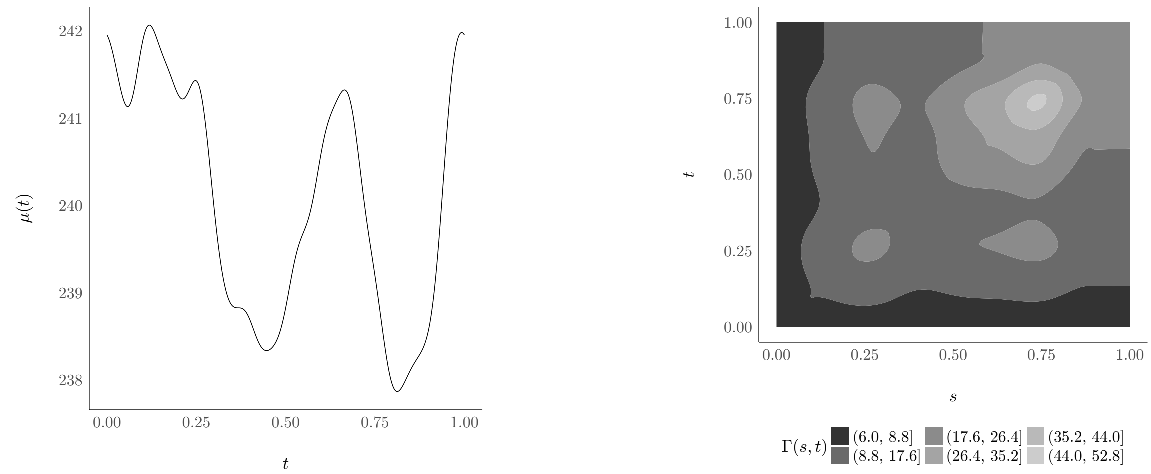
with the last equality implied by the fact that

$$\mathfrak{m}^{2H(s,t)} \ll N \quad \text{if and only if} \quad N^{-1/2} \ll (N\mathfrak{m})^{-\frac{H(s,t)}{2H(s,t)+1}} \ll (N\mathfrak{m}^2)^{-\frac{H(s,t)}{2\{H(s,t)+1\}}}.$$

D. Additional simulation results

Let us recall that we simulate datasets using the data generating process defined in Section 5.1 in the main text, with an Hurst index function H_t and a time deformation function A_t estimated on the Power Consumption dataset, to which we add a mean curve also fitted to the real dataset. The estimates $\widehat{H}_t$ and the estimates of the mean and covariance functions are obtained using the same data. That means we did not use a *learning sample* for $\widehat{H}_t$.

We consider eight experiments, each of them replicated 500 times. For each experiment, except specifically specified, we consider $N \in \{50, 100, 200\}$, $\mathfrak{m} \in \{20, 30, 40, 50\}$ and that the number of points per curve M_i has a Poisson distribution with mean $\mathfrak{m}$. In *Experiment 1*, we assume that the distribution of the sampling points is random uniform in $\mathcal{T}$, the standard deviation of the noise is $\sigma = 0.5$, the regularity of the mean function is $s = \exp(-6)$, the number of Fourier basis functions for the estimation of H_t and L_t is 9, and $\varpi = 2.5$. All the other experiments are designed starting from *Experiment 1* and modifying one parameter at a time. In *Experiment 2* and *Experiment 3*, we consider $\sigma = 0.25$ and $\sigma = 1$, respectively. We set $s = \exp(-3)$ for *Experiment 4* resulting in a smoother mean function μ (see Figure 1a). We used only 7 functions in the Fourier basis in *Experiment 5*, that is a smoother estimation of H_t and L_t and resulting in a smoother covariance surface Γ (see Figure 1b). For *Experiment 6*, the distribution of the sampling points is a mixture of beta distributions $0.5\mathcal{B}(1, 2) + 0.5\mathcal{B}(2, 1)$. For *Experiment 7*, we set $\varpi = 1$. Finally, in *Experiment 8*, we apply our approach to the case of differentiable trajectories that we obtain by integrating the sample paths generated as in *Experiment 1*.



(a) Mean curve $\mu(\cdot)$ for *Experiment 4*.

(b) Covariance surface $\Gamma(\cdot, \cdot)$ for *Experiment 5*.

Fig. 1: Description of the modification for *Experiment 4* and *5*.

The results from *Experiment 1*, with the ISE_0 criterion, are presented in the main text. Below we present the results from *Experiment 1*, with the $\text{ISE}_{0.05}$ criterion, and the results the other seven experiments. The results for the mean function are in Sectin D.1, while the results for the covariance function can be found in Section D.2.

D.1. Mean estimation

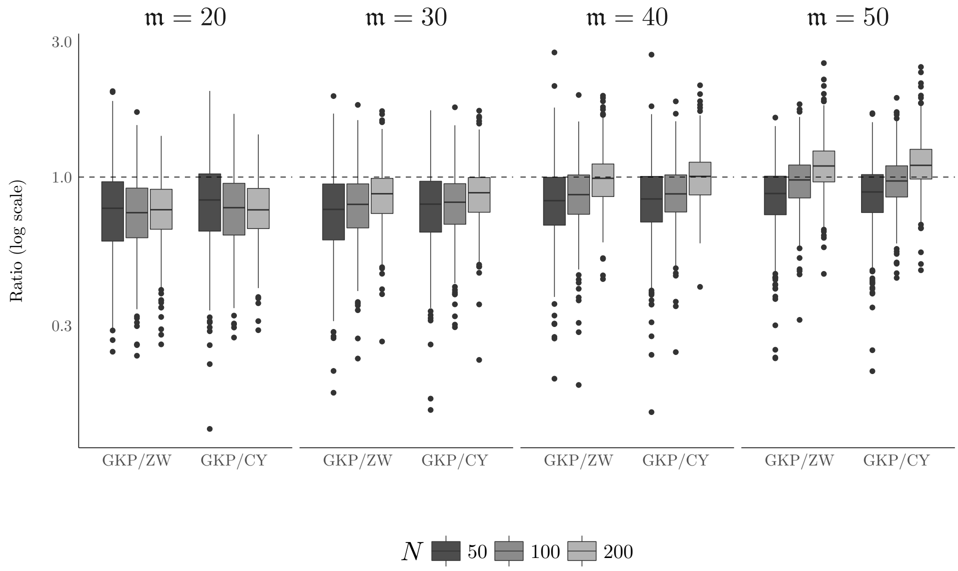


Fig. 2: Results for the estimation of μ for *Experiment 1*. The ratio are computed using $\mathrm{ISE}_{0.05}$.

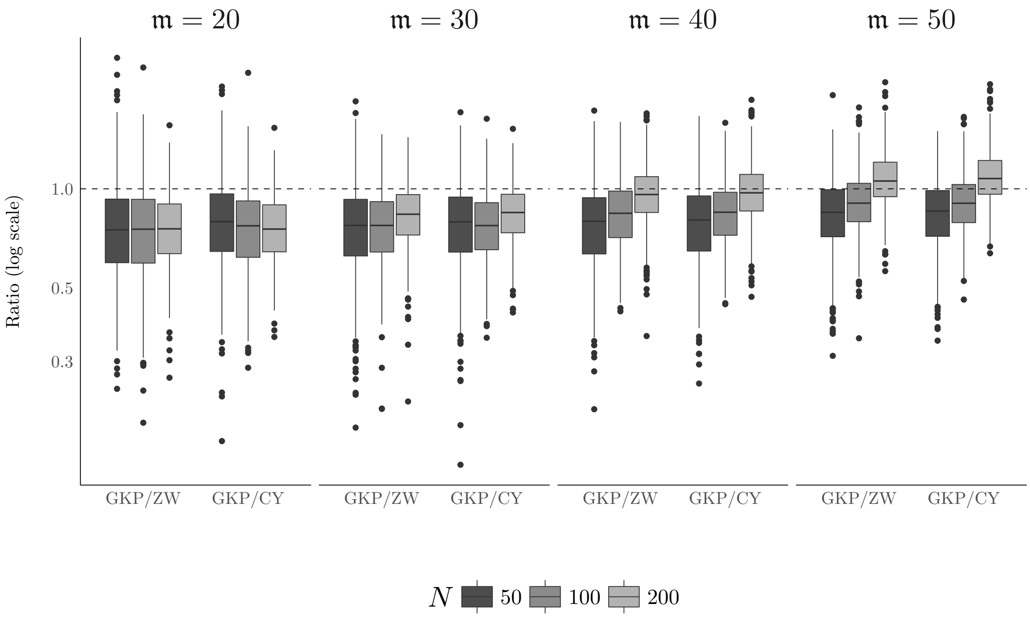


Fig. 3: Results for the estimation of μ for *Experiment 2* (noise std $\sigma = 0.25$). The ratio are computed using ISE_0.

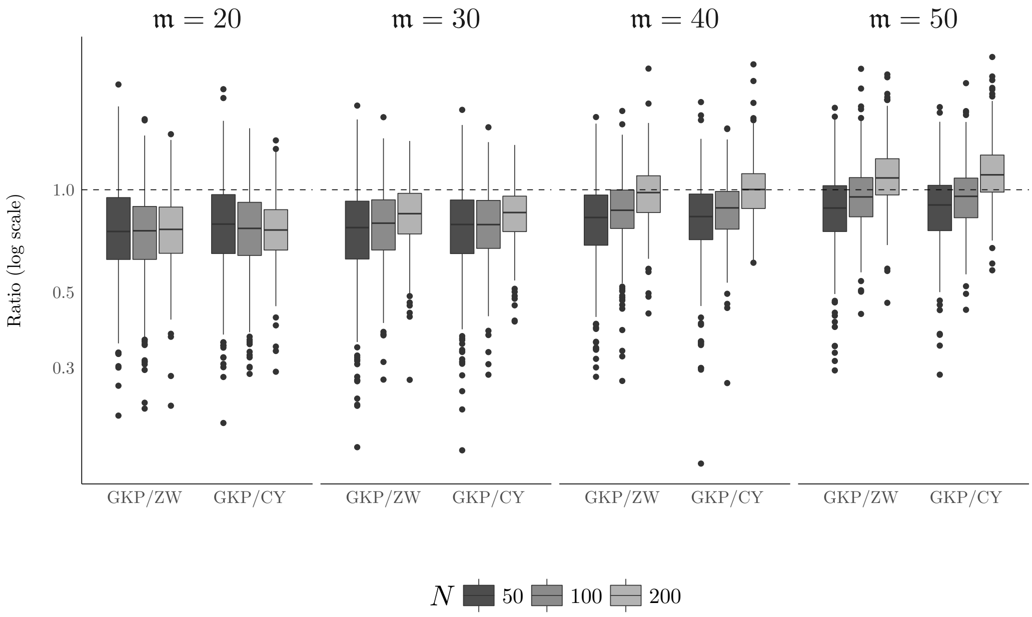


Fig. 4: Results for the estimation of μ for *Experiment 3* (noise std $\sigma = 1$). The ratio are computed using ISE_0.

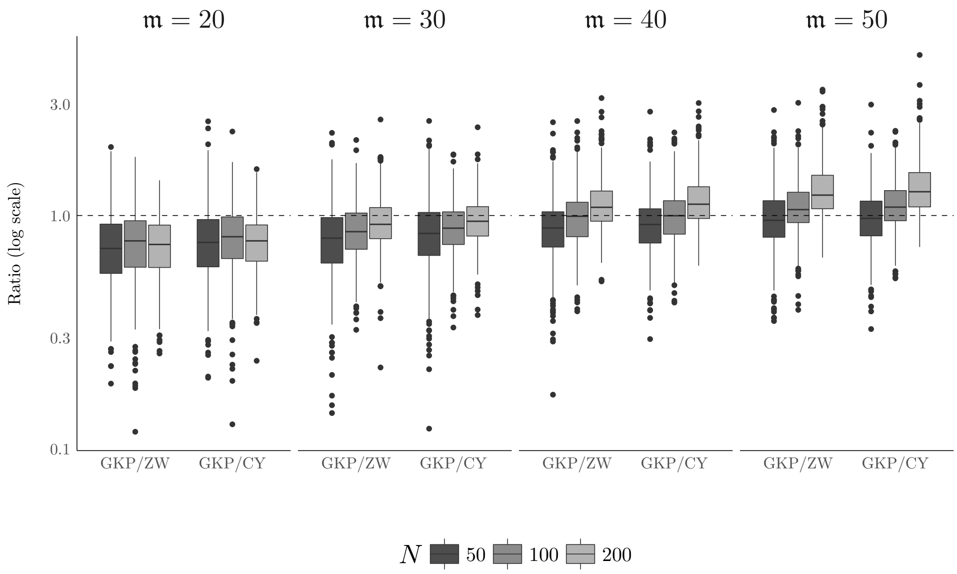


Fig. 5: Results for the estimation of μ for *Experiment 4* (smoother true mean curve μ). The ratio are computed using ISE_0.

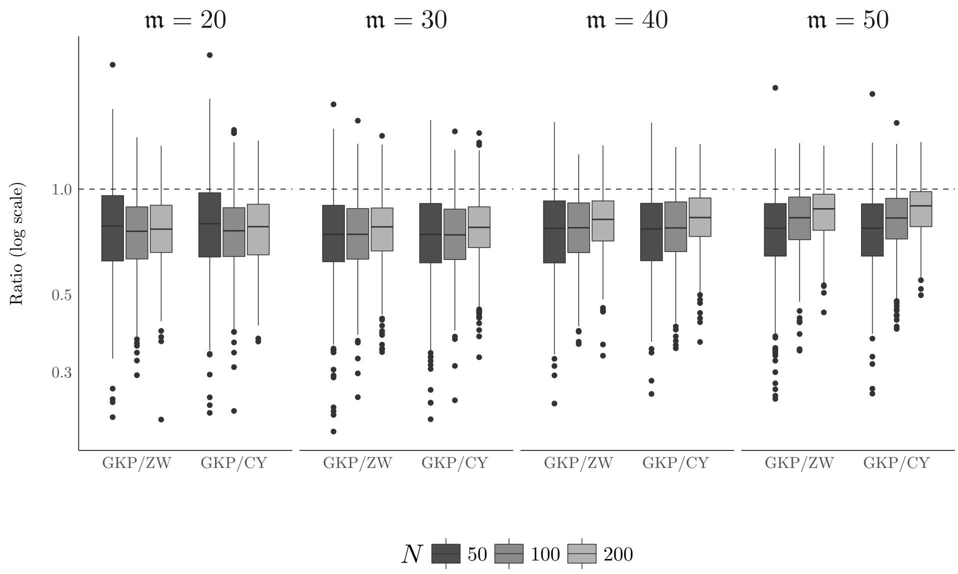


Fig. 6: Results for the estimation of μ for *Experiment 5* (smoother maps H and L). The ratio are computed using ISE_0.

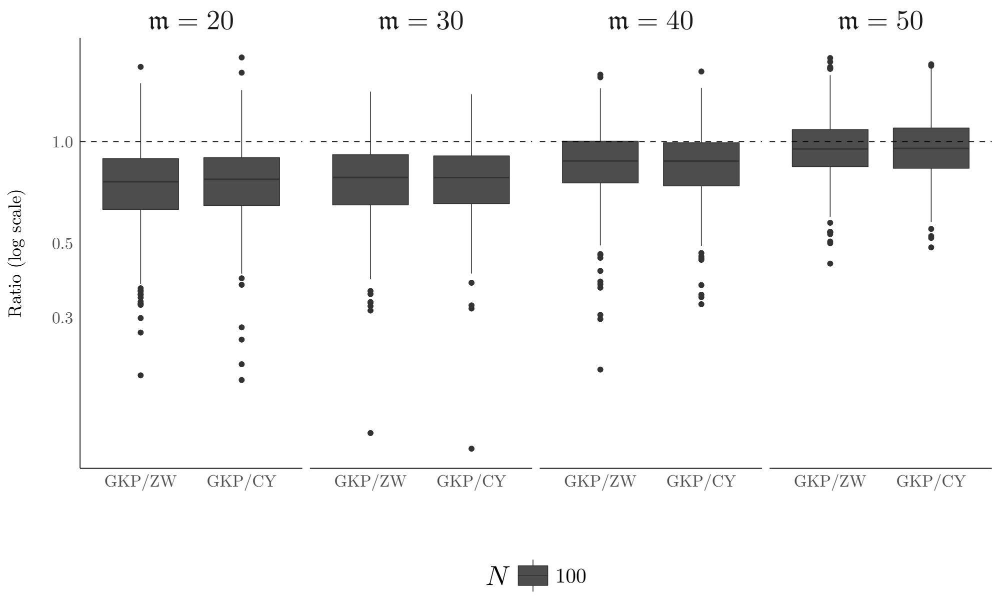


Fig. 7: Results for the estimation of μ for *Experiment 6* (the density of $T_m^{(i)}$ is a beta mixture). The ratio are computed using ISE_0.

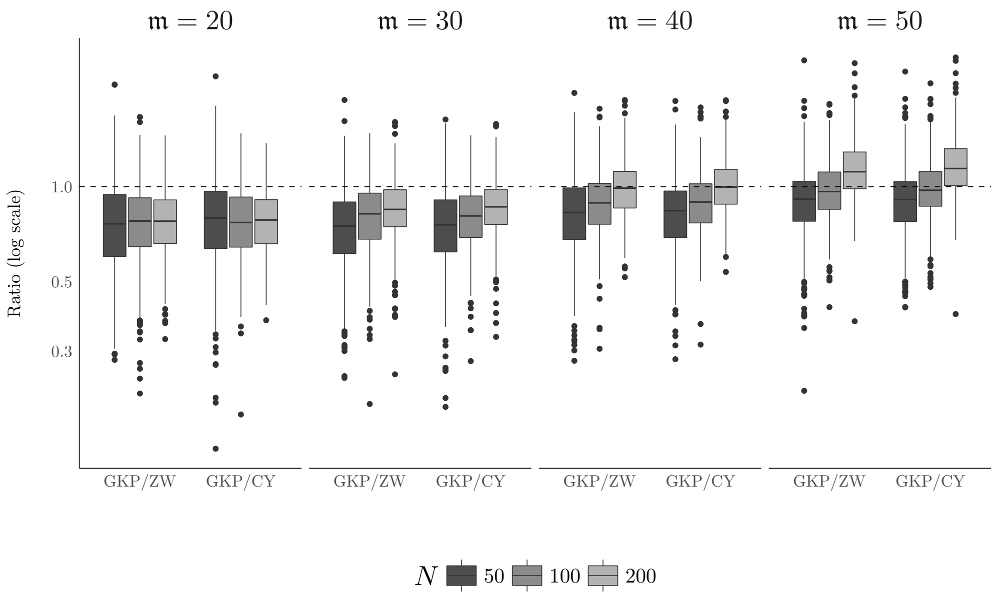


Fig. 8: Results for the estimation of μ for *Experiment 7* (std of $X(0)$ is $\varpi = 1$). The ratio are computed using ISE_0.

D.2. Covariance estimation

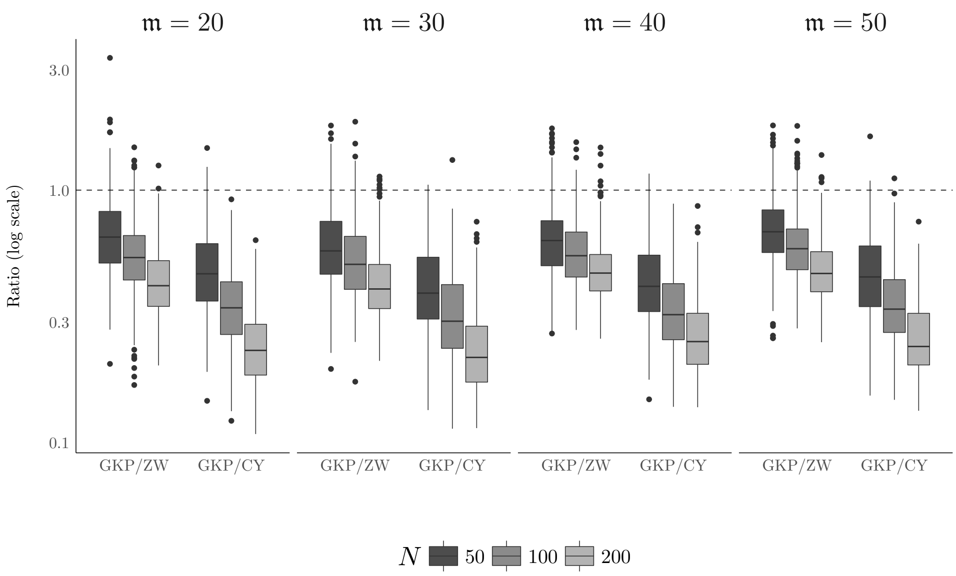


Fig. 9: Results for the estimation of Γ for *Experiment 1*. The ratio are computed using $\mathrm{ISE}_{0.05}$.

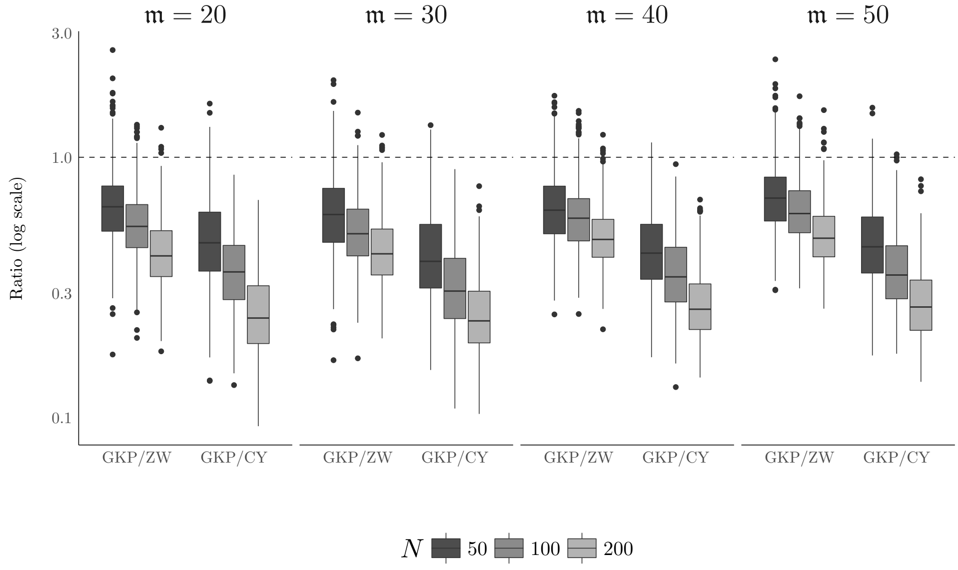


Fig. 10: Results for the estimation of Γ for *Experiment 2* (noise std $\sigma = 0.25$). The ratio are computed using ISE_0.

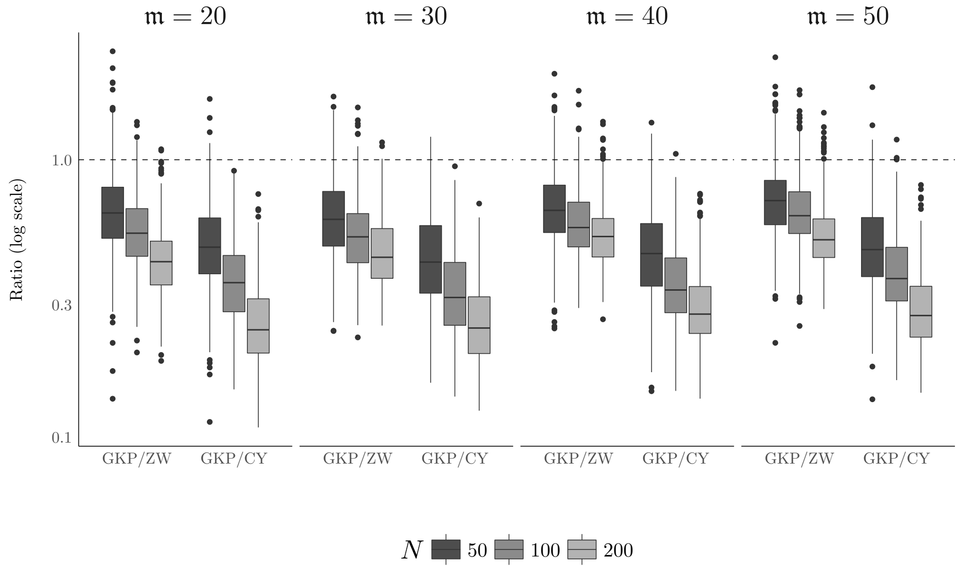


Fig. 11: Results for the estimation of Γ for *Experiment 3* (noise std $\sigma = 1$). The ratio are computed using ISE_0.

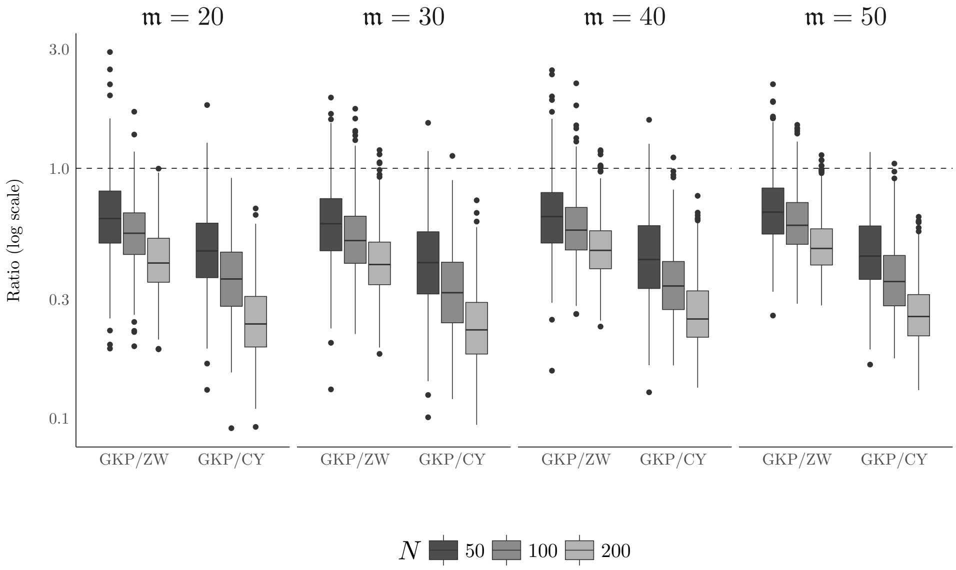


Fig. 12: Results for the estimation of Γ for *Experiment 4* (smoother true mean μ). The ratio are computed using ISE_0.

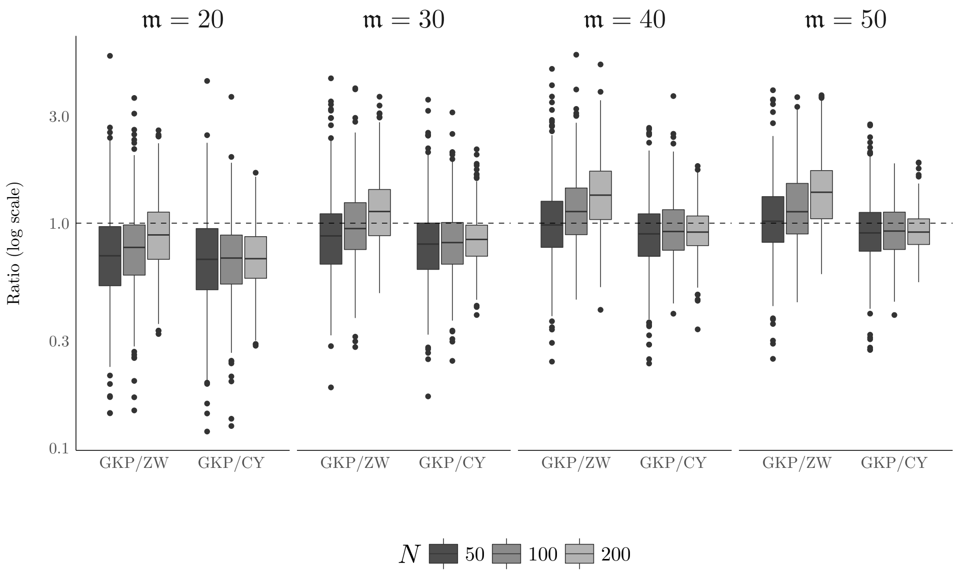


Fig. 13: Results for the estimation of Γ for *Experiment 5* (smoother maps H and L). The ratio are computed using ISE_0.

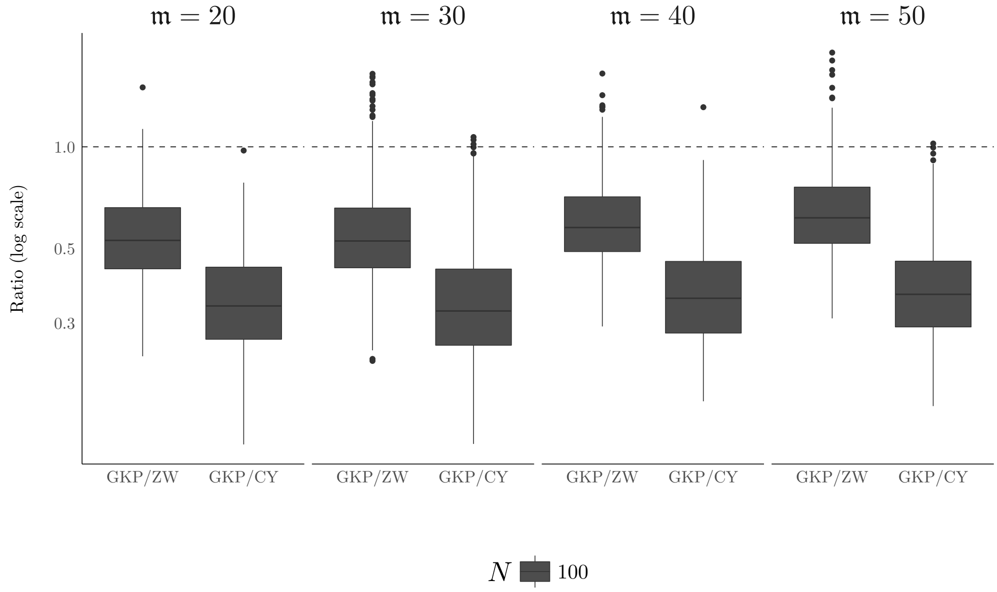


Fig. 14: Results for the estimation of Γ for *Experiment 6* (the density of the $T_m^{(i)}$ is a beta mixture). The ratio are computed using ISE_0.

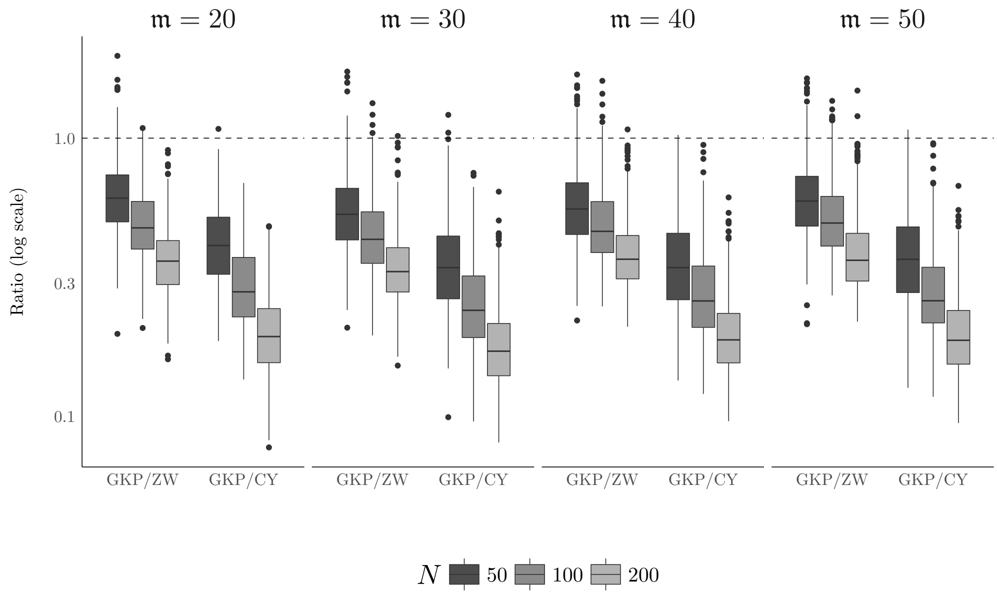
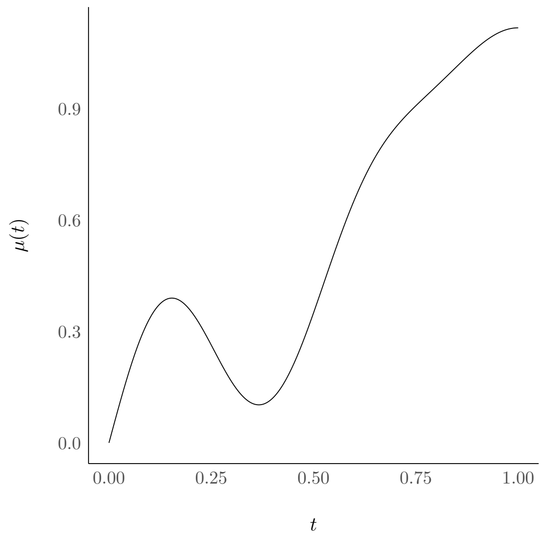


Fig. 15: Results for the estimation of Γ for *Experiment 7* (std of $X(0)$ is $\varpi = 1$). The ratio are computed using ISE_0.

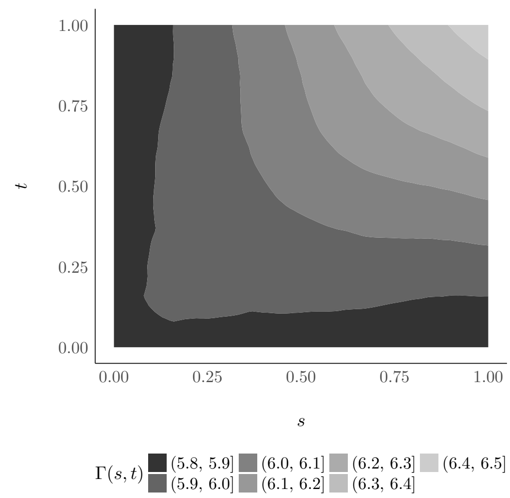
D.3. Case of differentiable curves

Let us note that, for any $d \geq 1$, we can use X as in (32) to define a process which, almost surely, has $d-$times differentiable sample paths and the derivatives of order d satisfy (H2). Indeed, it suffices to define

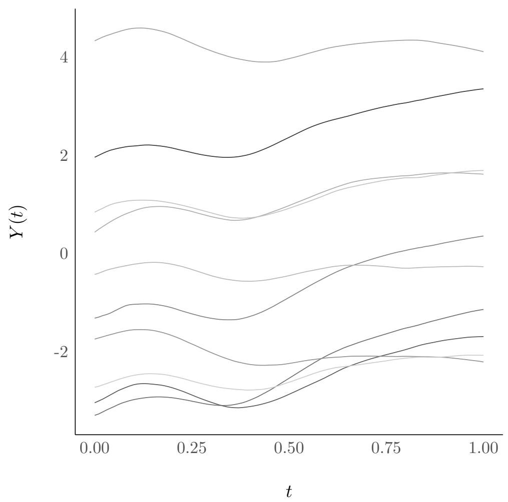
$$X(t) = \int_0^t \int_0^{s_1} \cdots \int_0^{s_{d-1}} X(s_d) ds_d \cdots ds_2 ds_1, \quad t \geq 0.$$



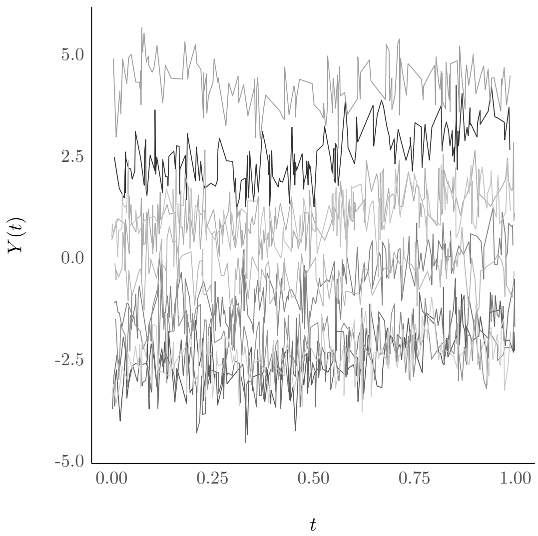
Mean curve $\mu(\cdot)$.



Covariance surface $\Gamma(\cdot, \cdot)$.



Curves $X^{(i)}$.



Noisy curves $Y^{(i)}$.

Fig. 16: Description of the simulated dataset with differentiable curves.

We consider the case of the estimation of the mean function for differentiable curves ($d = 1$), referred to as *Experiment 8*. More precisely, we generate curves as in *Experiment 1* and perform numerical integration such that the regularity of the curves is larger than one, and the Hurst index function H_t is defined on the sample path of the first derivative, for all $t \in [0, 1]$. See also Golovkine et al. (2022) for the formal definition of the local regularity for the case of differentiable sample paths. In this experiment, the mean curve is not learn from the Power Consumption dataset but generated as follows:

$$\mu(t) = \sqrt{2}\sum_{k=1}^{5} z_k \frac{\sin((k-1/2)\pi t)}{(k-1/2)\pi}, \quad (z_1, \dots, z_5) = (1.37, -0.56, 0.36, 0.63, 0.40).$$

The values z_k were obtained as random draws $\mathcal{N}(0,1)$.

We plot the mean curve $\mu(\cdot)$ on Figure 16a and the covariance matrix $\Gamma(\cdot,\cdot)$ on Figure 16b. A random sample of curves generated according to our simulation setup are plotted on Figure 16c without noise and on Figure 16d with noise.

As we assumed that the curves are differentiable, we first estimate their derivatives using local polynomials of degree 2 with bandwidth $3/\widehat{\mathfrak{m}}$. The estimation of the Hurst index function $\widehat{H}_t$ is then performed on the set of estimated derivative curves. Finally, our bandwidth selection methodology is run with $q_1^2 h^{2(1+\widehat{H}_t)}$ as the first term in the definition of $\mathcal{R}_\mu(t;h)$ in (18). The results are plotted in the Figure 17, on a logarithmic scale. The ratios are obtained using ISE_0. Our estimator outperforms the competitors for every pair $(N, \mathfrak{m})$.

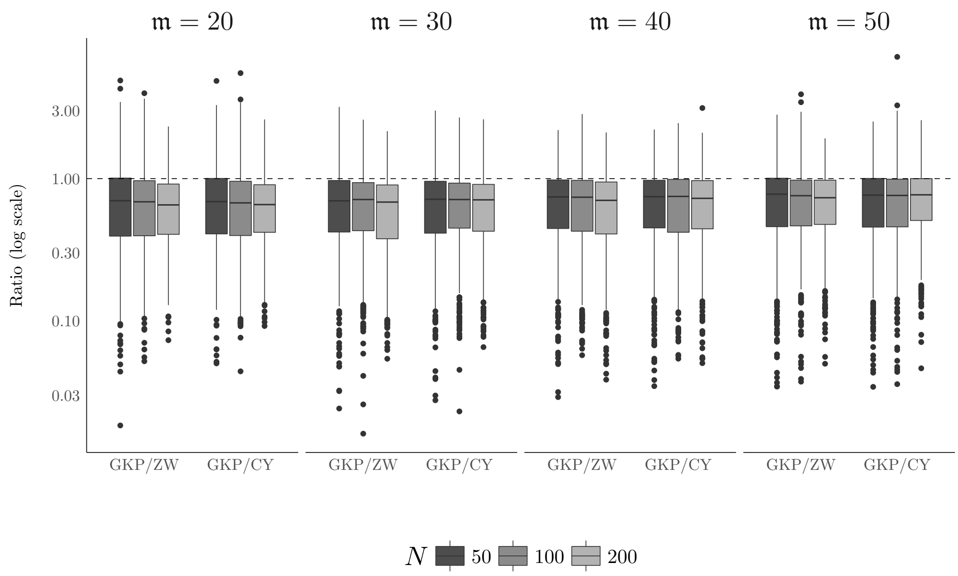


Fig. 17: Results for the estimation of μ for *Experiment 8*. The ratio are computed using ISE_0.

References

Abramowitz, M. and Stegun, I. A. (1964) *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* National Bureau of Standards Applied Mathematics Series, No. 55. U. S. Government Printing Office, Washington, D.C. For sale by the Superintendent of Documents.

Fan, J. and Gijbels, I. (1996) *Local polynomial modelling and its applications.* Monographs on statistics and applied probability. London: Chapman & Hall.

Golovkine, S., Klutchnikoff, N. and Patilea, V. (2022) Learning the smoothness of noisy curves with application to online curve estimation. *Electronic Journal of Statistics*, **16**.