



isds



Statistical test for validating Gaussian process regression covariance: A new way to detect non-stationarity

Théo SYLVESTRE (CEA/DM2S/SGLS/LIAD, Laboratoire de Mathématiques d'Avignon (LMA))

PhD director: Amandine MARREL (CEA Cadarache, LMA)

Supervisors: Guillaume DAMBLIN, Raksmy NOP (CEA Paris-Saclay)

Laboratory: Laboratoire d'intelligence artificielle et de science des données/ LMA

Doctoral school: ED 536

Context

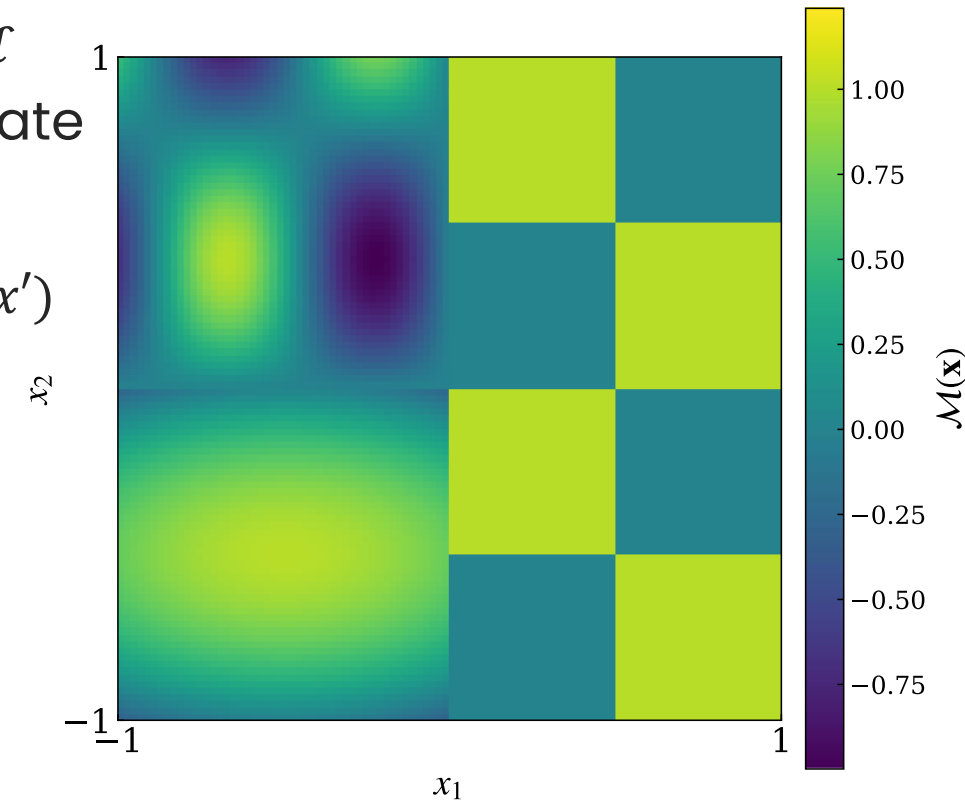
- Surrogate modelling for expensive computer codes $\mathcal{M}$
 - **Popular solution:** Gaussian Process ($\mathcal{GP}$) surrogate

Context

- Surrogate modelling for expensive computer codes $\mathcal{M}$
 - **Popular solution:** Gaussian Process ($\mathcal{GP}$) surrogate
- $\mathcal{GP}$ defined by **mean** $m(\cdot)$ and **covariance** $k(\cdot, \cdot)$
 - Common **stationary** covariance $k(x, x') = k(x - x')$

Context

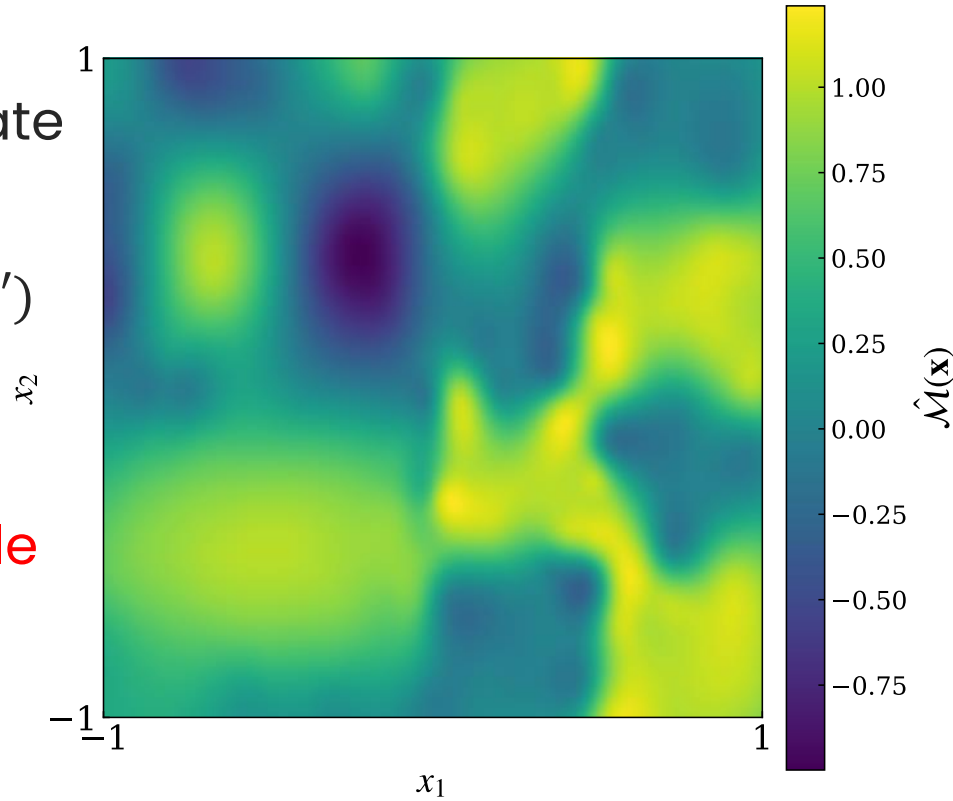
- Surrogate modelling for expensive computer codes $\mathcal{M}$
 - **Popular solution:** Gaussian Process ($\mathcal{GP}$) surrogate
- $\mathcal{GP}$ defined by **mean** $m(\cdot)$ and **covariance** $k(\cdot, \cdot)$
 - Common **stationary** covariance $k(x, x') = k(x - x')$
 - ↳ **Problem:** Inadequate when $\mathcal{M}$ exhibits distinct behaviours



Manhattan function [MS23]

Context

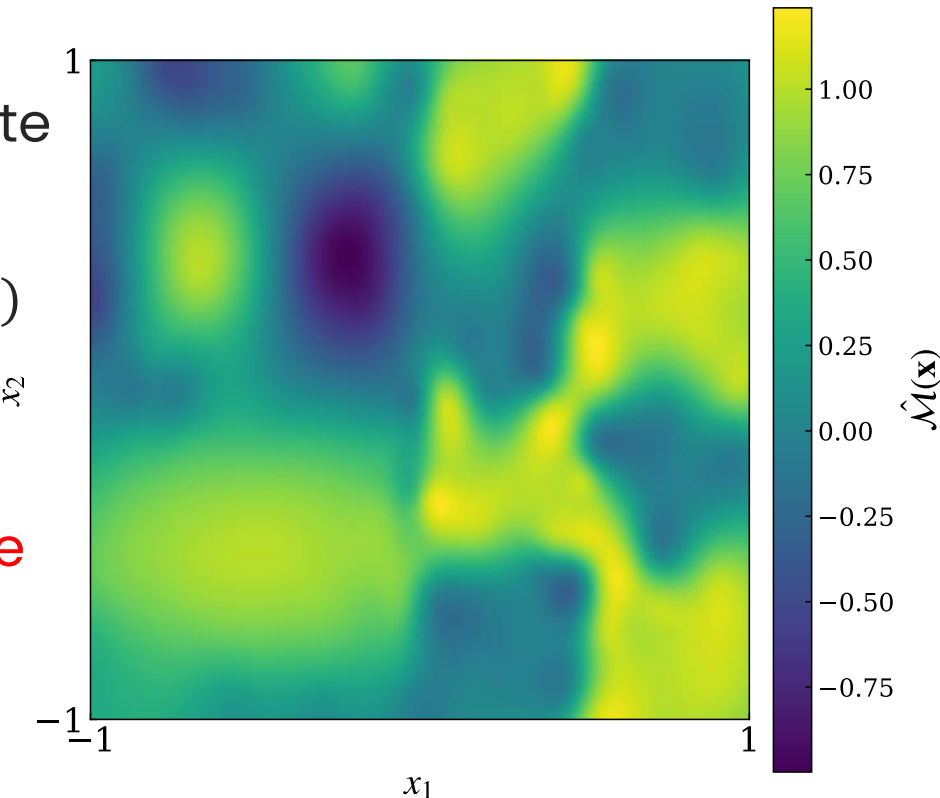
- Surrogate modelling for expensive computer codes $\mathcal{M}$
 - **Popular solution:** Gaussian Process ($\mathcal{GP}$) surrogate
- $\mathcal{GP}$ defined by **mean** $m(\cdot)$ and **covariance** $k(\cdot, \cdot)$
 - Common **stationary** covariance $k(x, x') = k(x - x')$
 - ↳ **Problem:** Inadequate when $\mathcal{M}$ exhibits distinct behaviours
- Non-stationarity ignored → **bad predictions /unreliable uncertainties**



Predictions from a stationary $\mathcal{GP}$

Context

- Surrogate modelling for expensive computer codes $\mathcal{M}$
 - **Popular solution:** Gaussian Process ($\mathcal{GP}$) surrogate
- $\mathcal{GP}$ defined by **mean** $m(\cdot)$ and **covariance** $k(\cdot, \cdot)$
 - Common **stationary** covariance $k(x, x') = k(x - x')$
 - ↳ **Problem:** Inadequate when $\mathcal{M}$ exhibits distinct behaviours
- Non-stationarity ignored → **bad predictions /unreliable uncertainties**
 - **Essential to:**
 - **Test** the adequacy between stationary model and data
 - **Develop** an appropriate non-stationary model when needed



Predictions from a stationary $\mathcal{GP}$



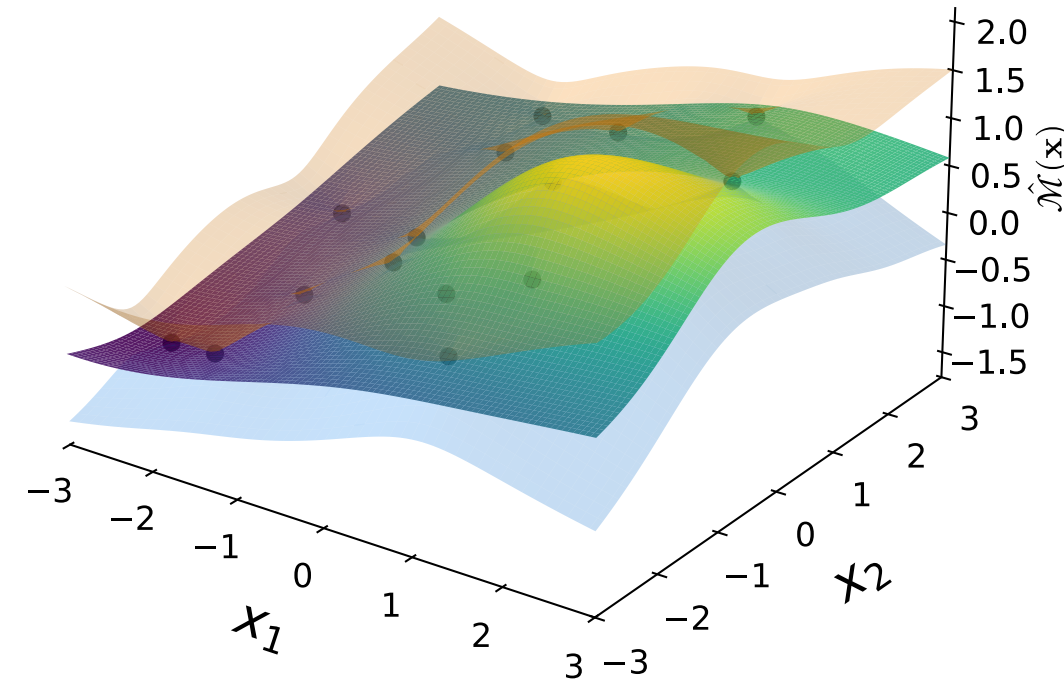
Gaussian process regression and non-stationarity

When standard covariance kernels fail

Gaussian process regression [RW06]

- Model a function: $y = \mathcal{M}(x), y \in \mathbb{R}, x \in \mathbb{R}^d$
- Learning set: $B_l = (x_l, y_l)$
- Prior distribution: $\mathcal{M} \sim \mathcal{GP}(0, k(\cdot, \cdot))$
- Posterior distribution: $\mathcal{M} | B_l \sim \mathcal{GP}(\bar{m}(\cdot, \cdot), \bar{k}(\cdot, \cdot))$
 - **Predictor:** $x \mapsto \bar{m}(x)$
 - **Posterior variance:** $x \mapsto \bar{k}(x, x)$

- GP predictive mean
- Upper 95% predictive bound
- Lower 95% predictive bound
- Training data ($n = 15$)

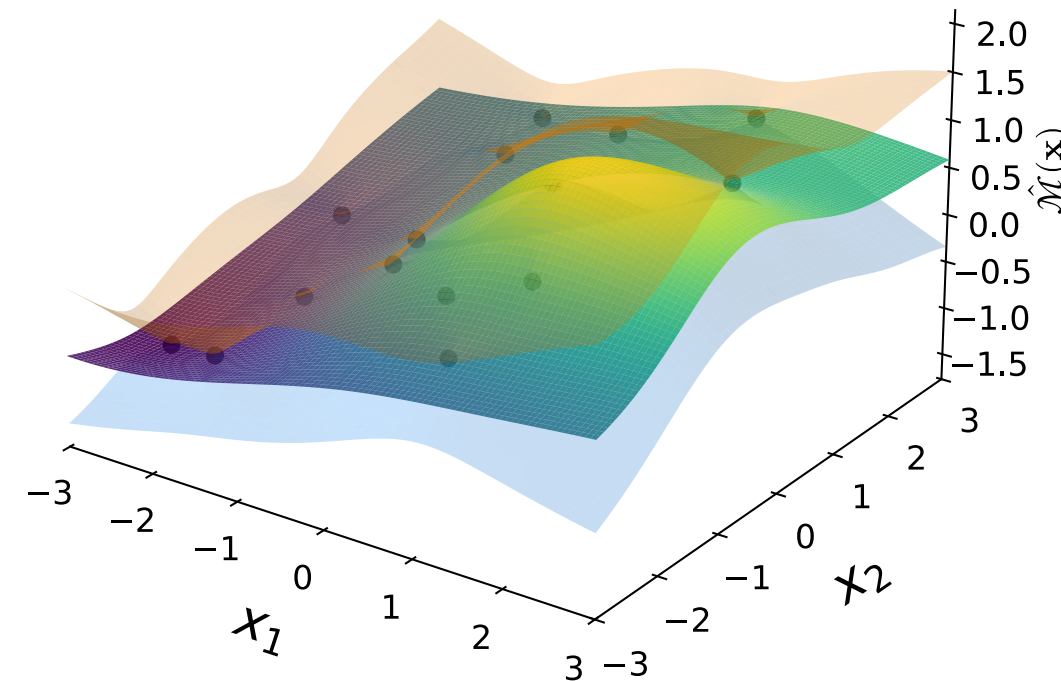


2D Gaussian process regression

Gaussian process regression [RW06]

- Model a function: $y = \mathcal{M}(x), y \in \mathbb{R}, x \in \mathbb{R}^d$
- Learning set: $B_l = (x_l, y_l)$
- Prior distribution: $\mathcal{M} \sim \mathcal{GP}(0, k(\cdot, \cdot))$
- Posterior distribution: $\mathcal{M} | B_l \sim \mathcal{GP}(\bar{m}(\cdot, \cdot), \bar{k}(\cdot, \cdot))$
 - **Predictor:** $x \mapsto \bar{m}(x)$
 - **Posterior variance:** $x \mapsto \bar{k}(x, x)$
- Common covariance parametrised by hyperparameters (HPs) θ :
 - ν – Matérn $_{\theta}$, $\nu \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$
 - $\theta = (\sigma^2, \mathbf{l} = (l_i)_{i \in \llbracket 1, d \rrbracket}) \in \Theta \subset (\mathbb{R}^{+*})^{d+1}$
 - θ estimated by Maximum Likelihood

- GP predictive mean
- Upper 95% predictive bound
- Lower 95% predictive bound
- Training data ($n = 15$)



2D Gaussian process regression

Validation Metrics

- Validation set: $\mathbf{B}_v = (\mathbf{x}_v, \mathbf{y}_v) = (x_{vi}, y_{vi})_{i \in \llbracket 1, n_v \rrbracket}$

Validation Metrics

- Validation set: $\mathbf{B}_v = (\mathbf{x}_v, \mathbf{y}_v) = (x_{vi}, y_{vi})_{i \in \llbracket 1, n_v \rrbracket}$

- Prediction accuracy metric:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_v} (y_{vi} - \overline{m}_{\theta}(y_{vi}))^2}{\sum_{i=1}^{n_v} (y_{vi} - \overline{y}_v)^2} \in] - \infty, 1]$$

- Good predictions: Q^2 close enough to 1

Validation Metrics

- Validation set: $\mathbf{B}_v = (\mathbf{x}_v, \mathbf{y}_v) = (x_{vi}, y_{vi})_{i \in \llbracket 1, n_v \rrbracket}$

- Prediction accuracy metric:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_v} (y_{vi} - \bar{m}_{\theta}(y_{vi}))^2}{\sum_{i=1}^{n_v} (y_{vi} - \bar{y}_v)^2} \in] - \infty, 1]$$

- Good predictions: Q^2 close enough to 1

- Posterior variance diagnostic:

$$IAE = \int |\widehat{IC}(\omega) - \omega| d\omega \in [0, 0.5]$$

- $\widehat{IC}(\omega)$: ω -level empirical coverage of confidence intervals
 - Reliable predictions and calibrated uncertainty: Q^2 close to 1, IAE close to 0

Evidence of non-stationarity in the data

- How to assess evidence of non-stationarity?
 - Usually assessed by inspecting the predictive distribution of stationary GPs [BO09]
 - ↳ **Limitation:** Stationarity primarily concerns the prior covariance kernel
 - **One Solution: Test** the suitability of several stationary kernels (i.e prior kernels)
 - ↳ **No valid** stationary kernel → strongly suggests **non-stationarity**
- Key idea for testing:
 - Compare observed validation metrics to **values expected** if data were generated by the tested model
 - ↳ **Inconsistent** with expected values
→ Kernel **rejected**
 - ↳ **Consistent** with expected values
→ Kernel **not rejected**



■ Statistical test overview

Statistical test in a nutshell

- H_0 : **Null** hypothesis VS H_1 : **Alternative** hypothesis
- Significance level α : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true})$,
 - Empirical rejection rate $\hat{\alpha} = \alpha \rightarrow$ Test **well calibrated**
- Power β : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ false})$, $\beta \geq 0,8 \rightarrow$ Test **powerful**

Statistical test in a nutshell

- H_0 : **Null** hypothesis VS H_1 : **Alternative** hypothesis
- Significance level α : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true})$,
 - Empirical rejection rate $\hat{\alpha} = \alpha \rightarrow$ Test **well calibrated**
- Power β : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ false})$, $\beta \geq 0,8 \rightarrow$ Test **powerful**
- Test statistic T :
 - Null distribution **known** or **simulated accurately** $\rightarrow$ control α
 - **Distinct** values under H_0 and $H_1 \rightarrow$ maximise β

Statistical test in a nutshell

- H_0 : **Null** hypothesis VS H_1 : **Alternative** hypothesis
- Significance level α : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true})$,
 - Empirical rejection rate $\hat{\alpha} = \alpha \rightarrow$ Test **well calibrated**
- Power β : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ false})$, $\beta \geq 0,8 \rightarrow$ Test **powerful**
- Test statistic T :
 - Null distribution **known** or **simulated accurately** $\rightarrow$ control α
 - **Distinct** values under H_0 and $H_1 \rightarrow$ maximise β
- General steps of statistical testing:
 1. Choose α
 2. Compute T on observed data, noted t_{obs}
 3. Simulate the null distribution of T
 4. Test decision $\rightarrow$ Reject H_0 if t_{obs} is **unlikely** under H_0

Statistical test in a nutshell

- H_0 : **Null** hypothesis VS H_1 : **Alternative** hypothesis
- Significance level α : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true})$,
 - Empirical rejection rate $\hat{\alpha} = \alpha \rightarrow$ Test **well calibrated**
- Power β : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ false})$, $\beta \geq 0,8 \rightarrow$ Test **powerful**
- Test statistic T :
 - Null distribution **known** or **simulated accurately** $\rightarrow$ control α
 - **Distinct** values under H_0 and $H_1 \rightarrow$ maximise β
- General steps of statistical testing:
 1. Choose α
 2. Compute T on observed data, noted t_{obs}
 3. Simulate the null distribution of T
 4. Test decision $\rightarrow$ Reject H_0 if t_{obs} is **unlikely** under $H_0 \rightarrow$ **p-value**

The p -value

- p -value: Probability of obtaining a T at least **as extreme as** t_{obs} under H_0
- For a level $\alpha \in [0,1]$:
 - $p\text{-value} \leq \alpha \rightarrow$ **significant** evidence against $H_0 \rightarrow$ **Reject H_0**
 - $p\text{-value} > \alpha \rightarrow$ **non-significant** evidence against $H_0 \rightarrow$ **Do not Reject H_0**

The p -value

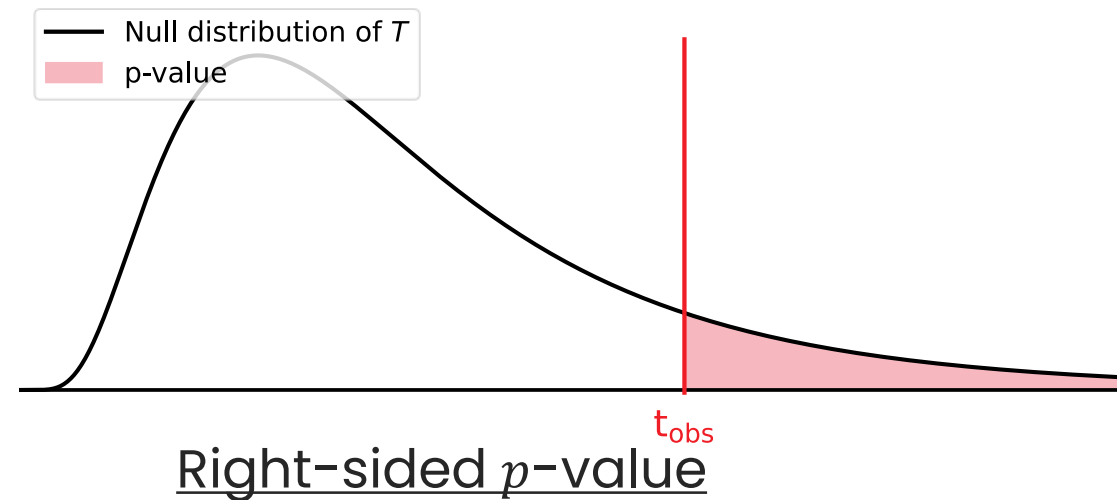
- **p -value**: Probability of obtaining a T at least **as extreme as** t_{obs} under H_0
- For a level $\alpha \in [0,1]$:
 - $p\text{-value} \leq \alpha \rightarrow$ **significant** evidence against $H_0 \rightarrow$ **Reject H_0**
 - $p\text{-value} > \alpha \rightarrow$ **non-significant** evidence against $H_0 \rightarrow$ **Do not Reject H_0**
- Two types of p -value:

The p -value

- **p -value**: Probability of obtaining a T at least **as extreme as** t_{obs} under H_0
- For a level $\alpha \in [0,1]$:
 - $p\text{-value} \leq \alpha \rightarrow$ **significant** evidence against $H_0 \rightarrow$ **Reject H_0**
 - $p\text{-value} > \alpha \rightarrow$ **non-significant** evidence against $H_0 \rightarrow$ **Do not Reject H_0**
- Two types of p -value:

- **Right sided** p -value:

$$\mathbb{P}_{H_0}(T \geq t_{obs})$$

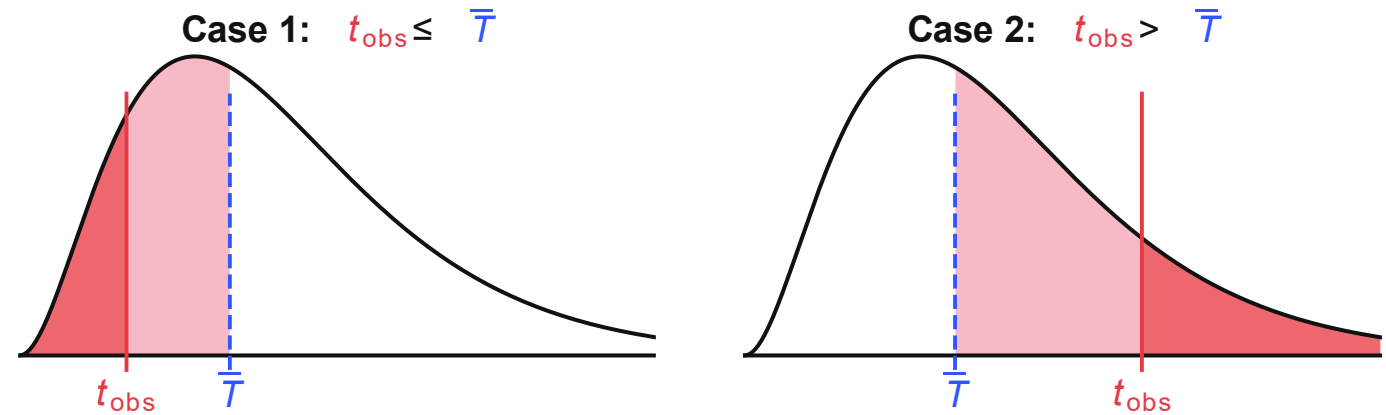


The p -value

- **p -value**: Probability of obtaining a T at least **as extreme as** t_{obs} under H_0
- For a level $\alpha \in [0,1]$:
 - $p\text{-value} \leq \alpha \rightarrow$ **significant** evidence against $H_0 \rightarrow$ **Reject H_0**
 - $p\text{-value} > \alpha \rightarrow$ **non-significant** evidence against $H_0 \rightarrow$ **Do not Reject H_0**
- Two types of p -value:
 - **Bilateral** p -value centered at the mean $\bar{T}$ for **asymmetric** null distribution: [K08]

$$\hookrightarrow \frac{\mathbb{P}(T \leq t_{obs})}{\mathbb{P}(T \leq \bar{T})} \text{ if } t_{obs} \leq \bar{T}$$

$$\hookrightarrow \frac{\mathbb{P}(T \geq t_{obs})}{\mathbb{P}(T \geq \bar{T})} \text{ if } t_{obs} > \bar{T}$$



Bilateral p -value for asymmetric null distribution



■ Statistical test to assess covariance adequacy

A new way for validating covariance kernel

Formulation of the covariance testing problem

- The two functional hypotheses:

$$H_{0,f} : \exists \boldsymbol{\theta} \in \Theta, \mathcal{M} \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot)) \quad \text{vs} \quad H_{1,f} : \forall \boldsymbol{\theta} \in \Theta, \mathcal{M} \not\sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot))$$

Formulation of the covariance testing problem

- The two functional hypotheses:

$$H_{0,f} : \exists \boldsymbol{\theta} \in \Theta, \mathcal{M} \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot)) \quad \text{vs} \quad H_{1,f} : \forall \boldsymbol{\theta} \in \Theta, \mathcal{M} \not\sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot))$$

- With single learning set $\mathbf{B}_l = (\mathbf{X}_l, \mathbf{Y}_l)$ and validation set $\mathbf{B}_v = (\mathbf{X}_v, \mathbf{Y}_v)$:

$$H_0 : \exists \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}}) \quad \text{vs} \quad H_1 : \forall \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \not\sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}})$$

➤ $\mathbf{K}_{\boldsymbol{\theta}}$ the prior covariance matrix

Formulation of the covariance testing problem

- The two functional hypotheses:

$$H_{0,f} : \exists \boldsymbol{\theta} \in \Theta, \mathcal{M} \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot)) \quad \text{vs} \quad H_{1,f} : \forall \boldsymbol{\theta} \in \Theta, \mathcal{M} \not\sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot))$$

- With single learning set $\mathbf{B}_l = (\mathbf{X}_l, \mathbf{Y}_l)$ and validation set $\mathbf{B}_v = (\mathbf{X}_v, \mathbf{Y}_v)$:

$$H_0 : \exists \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}}) \quad \text{vs} \quad H_1 : \forall \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \not\sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}})$$

- $\mathbf{K}_{\boldsymbol{\theta}}$ the prior covariance matrix
- Unknown HPs $\boldsymbol{\theta} \rightarrow H_0$ composite $\rightarrow H_0$ distribution is not uniquely defined
 - $\hat{\boldsymbol{\theta}}_{\text{obs}} := \hat{\boldsymbol{\theta}}(\mathbf{y}_l)$ used to simulate the chosen H_0 distribution

Formulation of the covariance testing problem

- The two functional hypotheses:

$$H_{0,f} : \exists \boldsymbol{\theta} \in \Theta, \mathcal{M} \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot)) \quad \text{vs} \quad H_{1,f} : \forall \boldsymbol{\theta} \in \Theta, \mathcal{M} \not\sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(\cdot, \cdot))$$

- With single learning set $\mathbf{B}_l = (\mathbf{X}_l, \mathbf{Y}_l)$ and validation set $\mathbf{B}_v = (\mathbf{X}_v, \mathbf{Y}_v)$:

$$H_0 : \exists \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}}) \quad \text{vs} \quad H_1 : \forall \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \not\sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}})$$

- $\mathbf{K}_{\boldsymbol{\theta}}$ the prior covariance matrix
- Unknown HPs $\boldsymbol{\theta} \rightarrow H_0$ **composite** $\rightarrow H_0$ distribution is not uniquely defined
 - $\hat{\boldsymbol{\theta}}_{\text{obs}} := \hat{\boldsymbol{\theta}}(\mathbf{y}_l)$ used to simulate the chosen H_0 distribution
- **Which T to choose?**
 - $Q^2, \text{IAE} \rightarrow$ Where to reject H_0 ?

Where to reject H_0

- How $T(Q^2, IAE)$ behaves under H_0 and H_1 ?

Simulation procedure

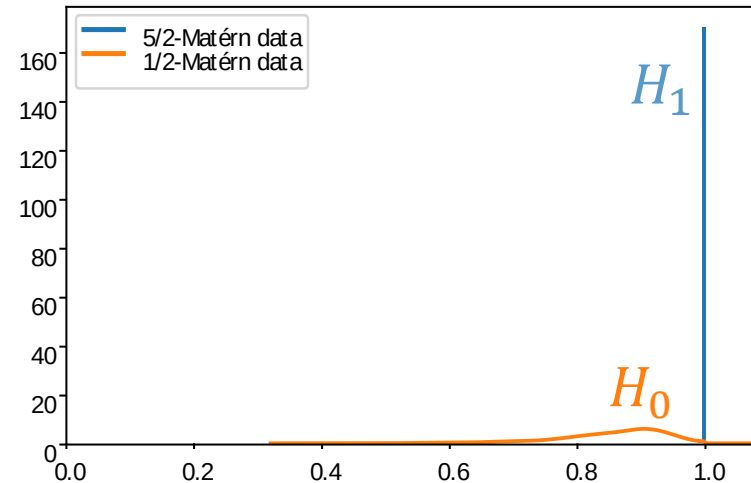
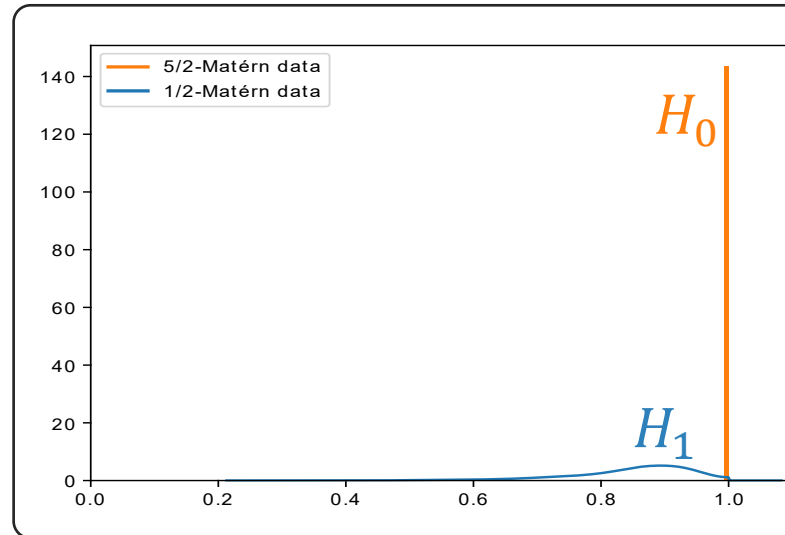
1. Generate sample paths from $\mathcal{GP}\left(0, \frac{1}{2} - \text{Matérn}_\theta\right)$ and $\mathcal{GP}\left(0, \frac{5}{2} - \text{Matérn}_\theta\right)$ with θ
 2. Condition a regression model $\mathcal{GP}\left(0, \frac{1}{2} - \text{Matérn}_\theta\right)$ (or $\mathcal{GP}\left(0, \frac{5}{2} - \text{Matérn}_\theta\right)$) with θ on each sample path
 3. Compute T for each regression model
- Same (resp. Distinct) covariance in generating/regression models: H_0 (resp. H_1) distribution

Where to reject H_0

Regressed $k_\theta = 5/2$ -Matérn

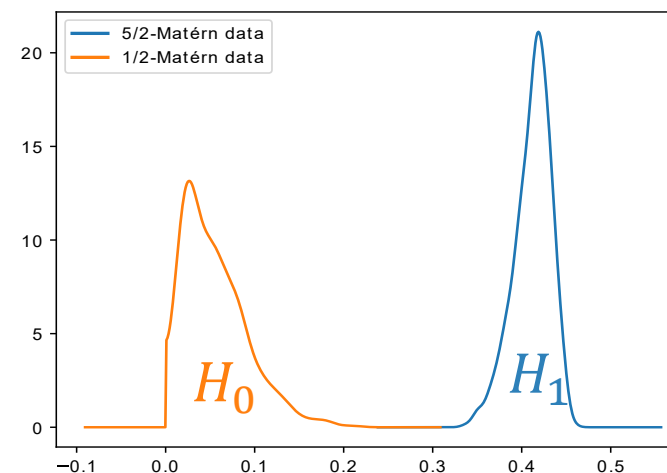
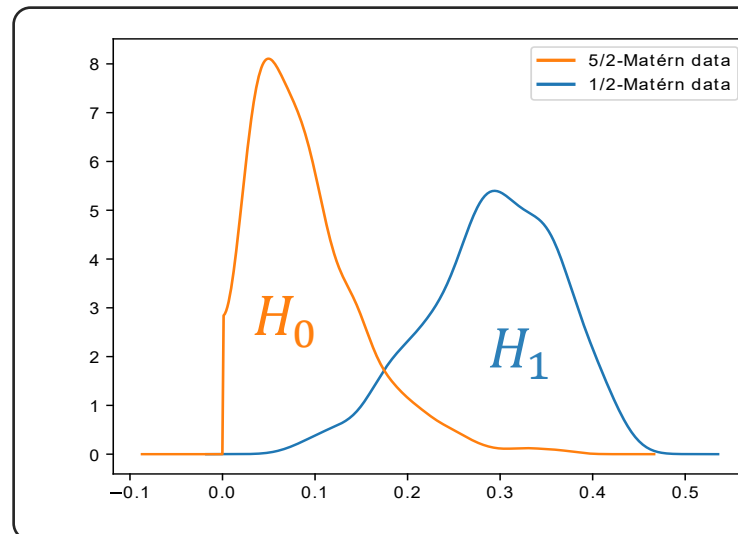
Regressed $k_\theta = 1/2$ -Matérn

Q^2 densities



➤ Bilateral test

IAE densities



➤ Right-sided test

Test procedure

Algorithm: With HPs estimate

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l , compute T on B_v



Step 2: Simulation of the null distribution of T



Step 3: Test decision

Test procedure

Algorithm: With HPs estimate

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l , compute T on B_v

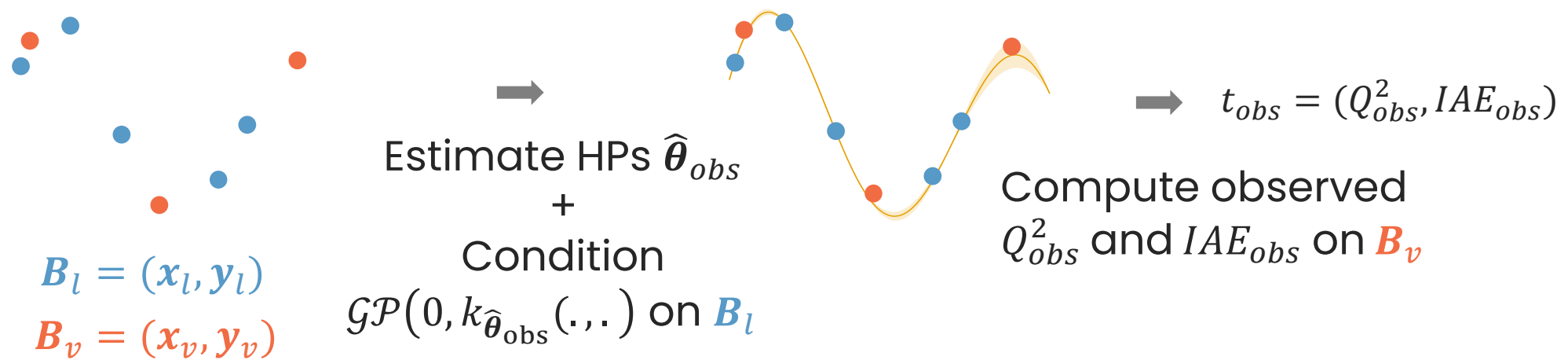


Step 2: Simulation of the null distribution of T



Step 3: Test decision

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l ,
compute T on B_v



Test procedure

Algorithm: With HPs estimate

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l , compute T on B_v

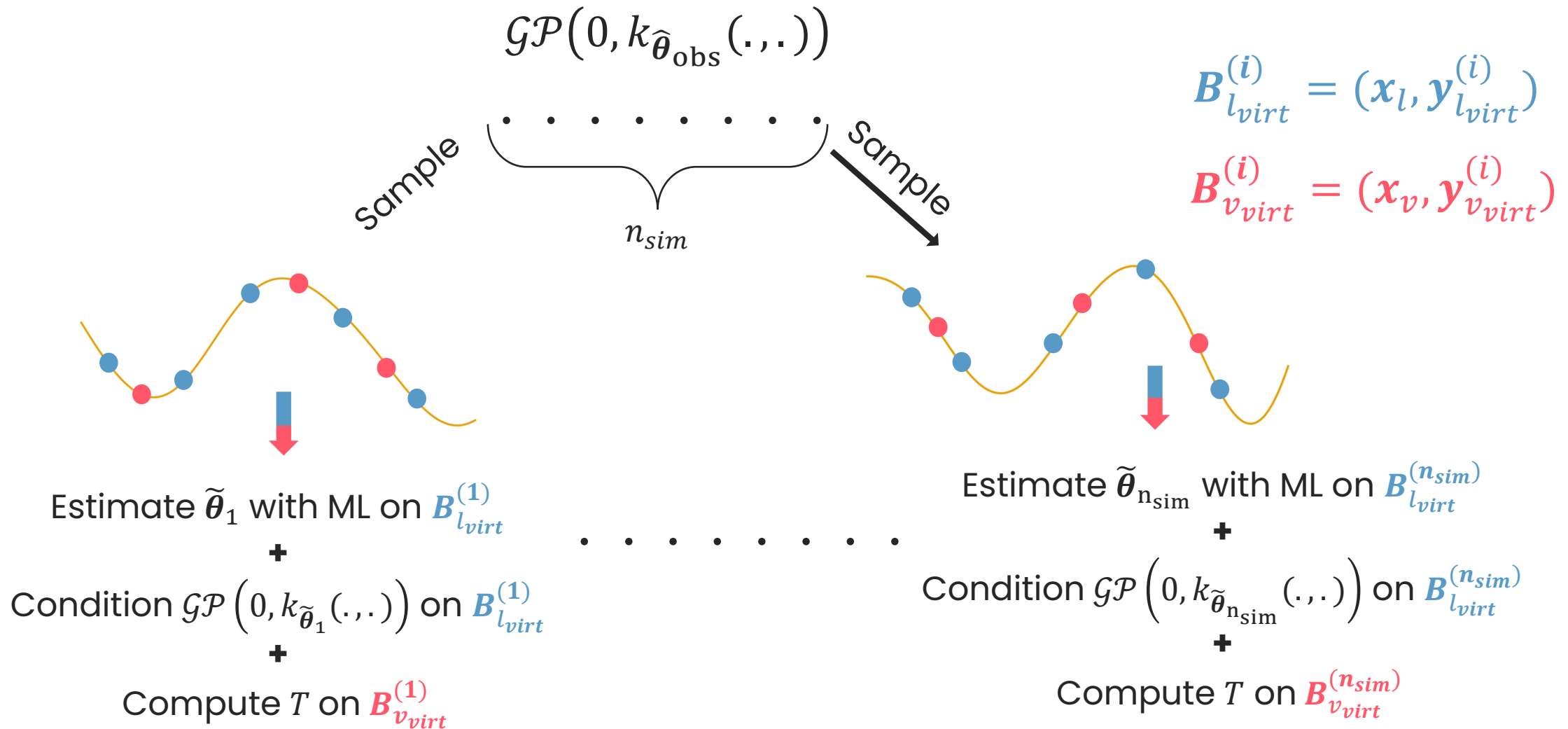


Step 2: Simulation of the null distribution of T



Step 3: Test decision

Step 2: Simulation of the null distribution [BR04]



Test procedure

Algorithm: With HPs estimate

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l , compute T on B_v



Step 2: Simulation of the null distribution of T



Step 3: Test decision

Test procedure

Algorithm: With HPs estimate

Step 1: Estimate $\hat{\theta}_{\text{obs}}$, condition $\mathcal{GP}(0, k_{\hat{\theta}_{\text{obs}}}(\cdot, \cdot))$ on B_l , compute T on B_v



Step 2: Simulation of the null distribution of T



Step 3: Test decision

- What is the calibration of the test? $\hat{\alpha} \approx \alpha$?

Calibration of the proposed test

Simulation parameters

- Target significance level $\alpha = 5\%$
- $n_l \in [40, 80, 120], n_v = \frac{n_l}{3}$
- θ fixed for all the data-generating $\mathcal{GP}$
- Input dimension $d = 3$, tested covariances ($\frac{1}{2}$ -Matérn, $\frac{3}{2}$ -Matérn, $\frac{5}{2}$ -Matérn)
- Input DoEs sampled using mLHS (maximin LHS) with $(\mathbf{x}_l, \mathbf{x}_v) \in [0, 1]^d$
- **Algorithm 2** is applied to several sample paths generated with the same data-generating kernel to estimate $\hat{\alpha}$

Calibration of the proposed test

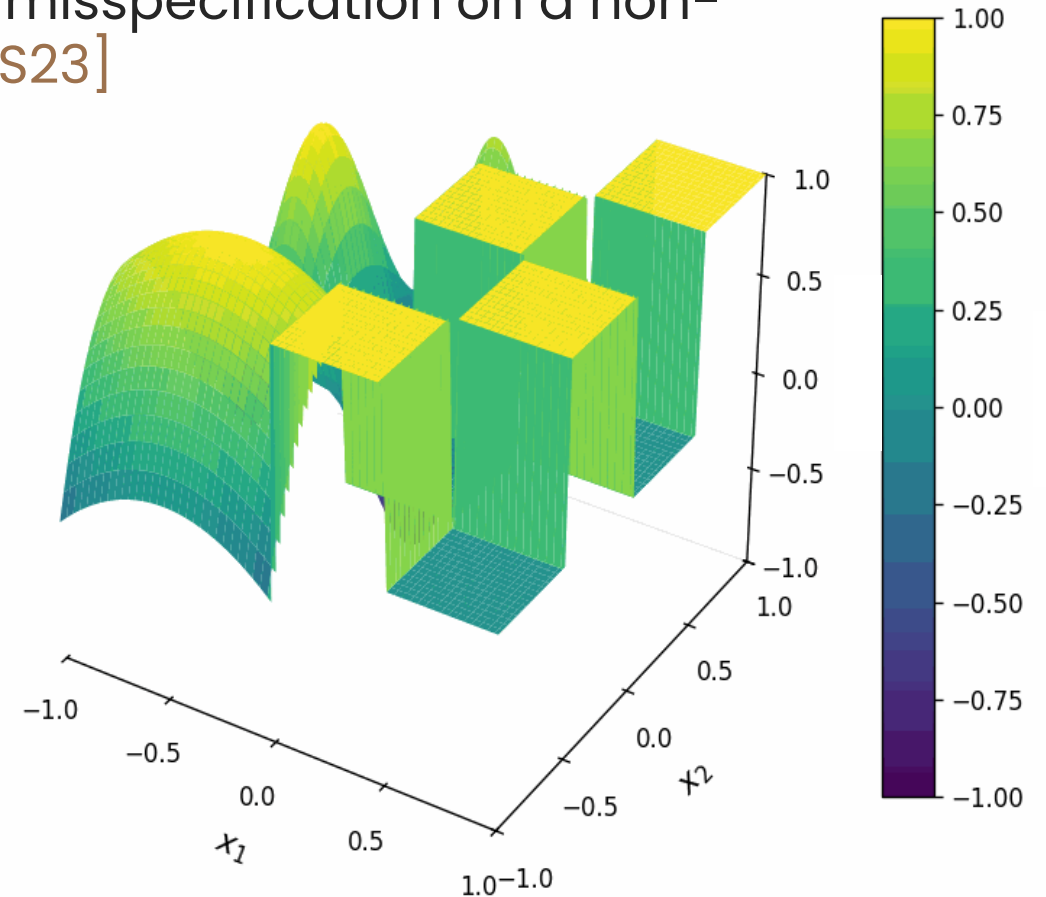
Kernel	Test statistic T	n_1		
		40	80	120
$\frac{1}{2}$ -Matérn	Q^2	4.0 ± 0.9 %	3.1 ± 0.8 %	3.1 ± 0.8 %
	IAE	4.7 ± 1.0 %	5.1 ± 1.0 %	5.2 ± 1.1 %
$\frac{3}{2}$ -Matérn	Q^2	3.7 ± 1.2 %	3.7 ± 1.2 %	1.7 ± 0.8 %
	IAE	6.0 ± 1.6 %	6.4 ± 1.6 %	4.9 ± 1.4 %
$\frac{5}{2}$ -Matérn	Q^2	4.1 ± 1.1 %	4.2 ± 1.1 %	3.2 ± 1.0 %
	IAE	5.3 ± 1.2 %	4.9 ± 1.2 %	4.9 ± 1.2 %

Empirical rejection rates at $\alpha = 5\%$, with confidence intervals.

- Can the test reject all stationary kernels when data are non-stationary ?

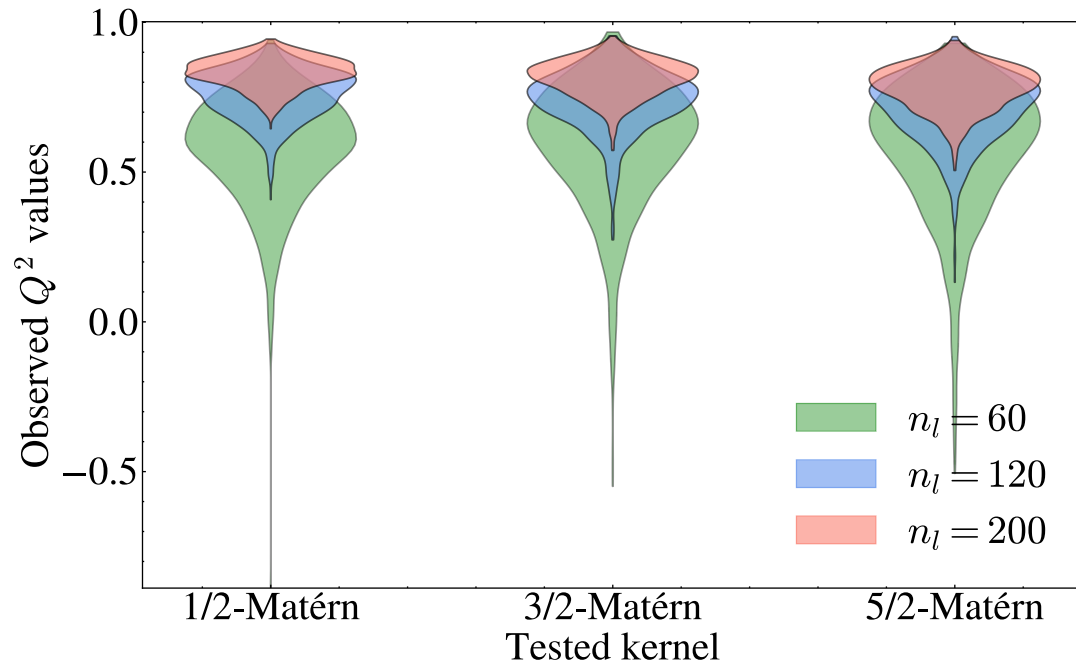
Manhattan function

- Assessing the power of the test to detect kernel misspecification on a non-stationary function : **The Manhattan function** [MS23]
- Three scenarios are considered:
 - $n_l = 60, n_v = 20$
 - $n_l = 120, n_v = 40$
 - $n_l = 200, n_v = 67$



Manhattan function [MS23]

Manhattan function results



Observed predictive performance

Kernel	n_l, n_v	Estimated power	
		Q^2	IAE
$\frac{1}{2}$ -Matérn	60, 20	5%	21%
	120, 40	4%	95%
	200, 67	3%	100%
$\frac{3}{2}$ -Matérn	60, 20	15%	17%
	120, 40	15%	82%
	200, 67	17%	100%
$\frac{5}{2}$ -Matérn	60, 20	15%	8%
	120, 40	20%	60%
	200, 67	22%	99%

Estimated power for each tested model and sample size

- **IAE more sensitive** than Q^2 to kernel misspecification
- More samples needed when the regression kernel closely matches the true regularity

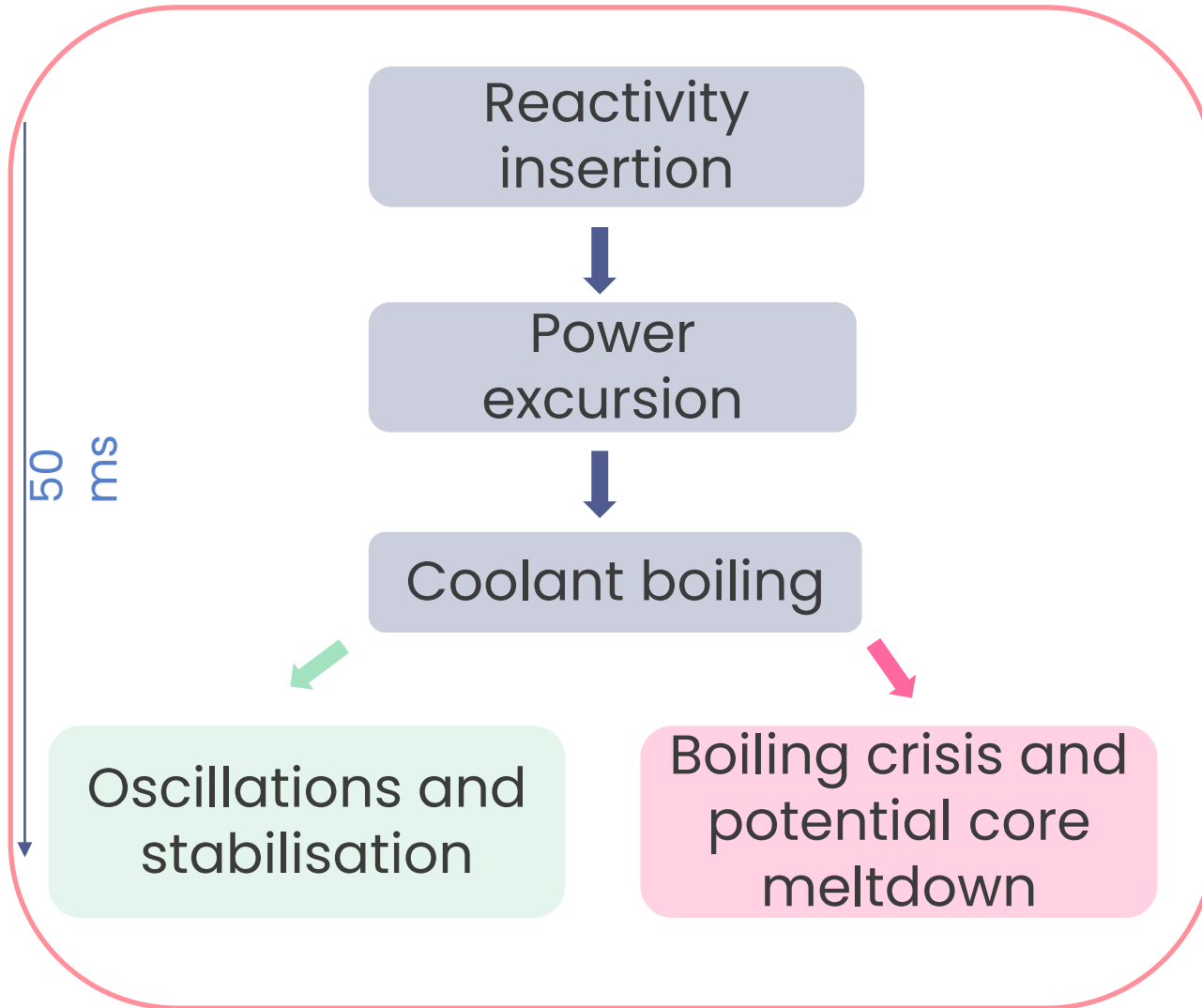


Application to an insertion reactivity accident



Application to a real use case

Application to a reactivity insertion accident



Test parameters:

- 3D inputs
- $n_l = 275, n_v = 45$
- $n_{\text{sim}} = 1000$
- $\alpha = 5\%$
- $H_0 : \exists \mu, \theta \in \mathbb{R} \times \Theta, \mathcal{M} \sim \mathcal{GP}(\mu, k_\theta(\cdot, \cdot))$
- Tested kernels k_θ :
 - $\frac{1}{2}$ -Matérn
 - $\frac{3}{2}$ -Matérn
 - $\frac{5}{2}$ -Matérn

Results for the reactivity insertion accident

Kernel	Q^2		IAE	
	Observed	p -value	Observed	p -value
$\frac{1}{2}$ -Matérn	0.82	68%	0.16	0.0%
$\frac{3}{2}$ -Matérn	0.79	46%	0.12	0.8%
$\frac{5}{2}$ -Matérn	0.79	86%	0.10	3%

Observed p -values and t_{obs} for each tested kernel

- IAE rejects: $\frac{1}{2}$ -Matérn + $\frac{3}{2}$ -Matérn + $\frac{5}{2}$ -Matérn } Non-stationarity **strongly suspected**
➤ Supports experts' assessment

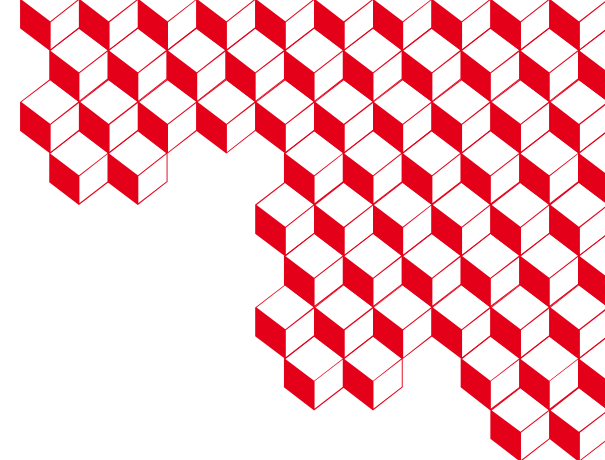
Conclusion and perspectives

- **Conclusion:**

- We proposed a new test to assess the covariance adequacy for $\mathcal{GP}$ regression
- It showed encouraging results for detecting kernel misspecification on analytical examples
- It supported experts assessment for the reactivity insertion accident

- **Perspectives:**

- Optimise input DoEs to enhance the power of the test
- Find a way to combine both Q^2 and IAE into a bivariate test



Thank you for your attention

Any questions?

theo.sylvestre@cea.fr

CEA Paris Saclay,
91190 Gif sur Yvette

The proposed test and chosen null distribution

- Tested hypotheses:

$$H_0 : \exists \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}}) \quad \text{vs} \quad H_1 : \forall \boldsymbol{\theta} \in \Theta, (\mathbf{Y}_v, \mathbf{Y}_l) \mid (\mathbf{X}_v, \mathbf{X}_l) \not\sim \mathcal{N}(0, \mathbf{K}_{\boldsymbol{\theta}})$$

- $\Theta \subset (\mathbb{R}_+^*)^{d+1} \rightarrow H_0$ is composite $\rightarrow$ null distributions are not uniquely defined
 - Only an estimate $\hat{\boldsymbol{\theta}}_{\text{obs}} := \hat{\boldsymbol{\theta}}(\mathbf{y}_l)$ is available
- Standard bootstrap distribution [BB00]:

$$g_T(t; \hat{\boldsymbol{\theta}}_{\text{obs}}) = \int_{\mathbb{R}^{n_l}} \int_{\mathbb{R}^{n_v}} \delta\left(t - T_{\hat{\boldsymbol{\theta}}(\mathbf{y}_l^*)}(\mathbf{y}_l^*, \mathbf{y}_v^*)\right) w_{\mathbf{Y}_l^*, \mathbf{Y}_v^*}(\mathbf{y}_l^*, \mathbf{y}_v^*; \hat{\boldsymbol{\theta}}_{\text{obs}}) d\mathbf{y}_v^* d\mathbf{y}_l^*$$

$$\text{➤ } \mathbf{Y}_l^*, \mathbf{Y}_v^* \mid \mathbf{X}_l, \mathbf{X}_v \sim \mathcal{N}\left(0, \mathbf{K}_{\hat{\boldsymbol{\theta}}_{\text{obs}}}(\mathbf{X}_l, \mathbf{X}_v)\right)$$

$$\text{➤ } \delta(x) = \begin{cases} 0, & x \neq 0 \\ 1, & x = 0 \end{cases}$$