A Statistical Learning View of Simple Kriging

Emilia Siviero, Stephan Clémençon and Emilie Chautru
LTCI, Télécom Paris, in collaboration with Geostatistics Team, Mines ParisTech

June 15, 2022

Introduction
Geostatistical context

Geo-databases: information from **location and value** of the data

Introduction

Geostatistical context

Geo-databases: information from **location and value** of the data

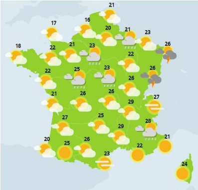
- data sets of **spatial** nature
  ⇒ **dependence structure**:
  close data are more related

Introduction

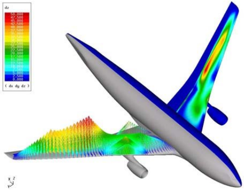
Geostatistical context

Geo-databases: information from **location and value** of the data

- data sets of **spatial** nature
  ⇒ **dependence structure**:
  close data are more related



- **unique** realisation of the phenomenon: **no** independent repetitions of the random field

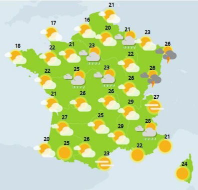


Introduction

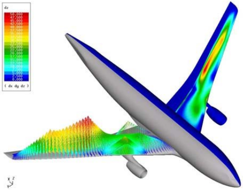
Geostatistical context

Geo-databases: information from **location and value** of the data

- data sets of **spatial** nature
  ⇒ **dependence structure**:
  close data are more related



- **unique** realisation of the phenomenon: **no** independent repetitions of the random field



Different from usual statistical learning theory:
non independent observations

Machine Learning:

Spatial Analysis:

Machine Learning:

- **Assets:** statistical learning theory for independent data, nonparametric theory

Spatial Analysis:

Machine Learning:

- **Assets:** statistical learning theory for independent data, nonparametric theory
- **Limits:** no theoretical guarantees for dependent data

Spatial Analysis:

Machine Learning:

- **Assets:** statistical learning theory for independent data, nonparametric theory
- **Limits:** no theoretical guarantees for dependent data

Spatial Analysis:

- **Assets:** take advantage of spatial structure (modelled by covariance function)

Machine Learning:

- **Assets:** statistical learning theory for independent data, nonparametric theory
- **Limits:** no theoretical guarantees for dependent data

Spatial Analysis:

- **Assets:** take advantage of spatial structure (modelled by covariance function)
- **Limits:** very few nonparametric theories

Machine Learning:

- **Assets:** statistical learning theory for independent data, nonparametric theory
- **Limits:** no theoretical guarantees for dependent data

Spatial Analysis:

- **Assets:** take advantage of spatial structure (modelled by covariance function)
- **Limits:** very few nonparametric theories

- **Limits:** lack of nonasymptotic results

Introduction

Objectives

Research Questions:

- how **dependence structure** of spatial data affects learning rates?

Introduction

Objectives

Research Questions:

- how **dependence structure** of spatial data affects learning rates?
- how spatial data can be processed by ML techniques with generalisation guarantees?

Introduction

Objectives

Research Questions:

- how **dependence structure** of spatial data affects learning rates?
- how spatial data can be processed by ML techniques with generalisation guarantees?
- connections between **Simple Kriging** and **Kernel Ridge Regression**?

Introduction

Objectives

Research Questions:

- how **dependence structure** of spatial data affects learning rates?
- how spatial data can be processed by ML techniques with generalisation guarantees?
- connections between **Simple Kriging** and **Kernel Ridge Regression**?

Objective: develop framework for **kriging** based on (nonasymptotic) study of the performances of a (nonparametric) **covariance estimator**

Preliminaries

1. Simple Kriging
2. Kernel Ridge Regression

Preliminaries
Notations

- $S \subset \mathbb{R}^2$ spatial domain

Preliminaries

Notations

- $S \subset \mathbb{R}^2$ spatial domain
- $X = \{X_s, s \in S\}$ second-order random field (RF)

Preliminaries

Notations

- $S \subset \mathbb{R}^2$ spatial domain
- $X = \{X_s, s \in S\}$ second-order random field (RF)
- $C(s, t) = Cov(X_s,\ X_t)$ covariance function of X

Preliminaries

Notations

- $S \subset \mathbb{R}^2$ spatial domain
- $X = \{X_s, s \in S\}$ second-order random field (RF)
- $C(s,t) = Cov(X_s,\ X_t)$ covariance function of X
- $\mathbf{X}(\mathbf{s}_n) = (X_{s_i})_{1 \leq i \leq n}$ observations of X at locations $\mathbf{s}_n = (s_i)_{1 \leq i \leq n}$

Preliminaries

Notations

- $S \subset \mathbb{R}^2$ spatial domain
- $X = \{X_s, s \in S\}$ second-order random field (RF)
- $C(s, t) = Cov(X_s,\ X_t)$ covariance function of X
- $\mathbf{X}(\mathbf{s}_n) = (X_{s_i})_{1 \le i \le n}$ observations of X at locations $\mathbf{s}_n = (s_i)_{1 \le i \le n}$
- $\mathbf{c}_n(s) = (Cov(X_s, X_{s_i}))_{1 \le i \le n}$ covariance vector

Preliminaries

Notations

- $S \subset \mathbb{R}^2$ spatial domain
- $X = \{X_s, s \in S\}$ second-order random field (RF)
- $C(s, t) = Cov(X_s,\ X_t)$ covariance function of X
- $\mathbf{X}(\mathbf{s}_n) = (X_{s_i})_{1 \le i \le n}$ observations of X at locations $\mathbf{s}_n = (s_i)_{1 \le i \le n}$
- $\mathbf{c}_n(s) = (Cov(X_s, X_{s_i}))_{1 \le i \le n}$ covariance vector
- $\Sigma(\mathbf{s}_n) = Var(\mathbf{X}(\mathbf{s}_n)) = (Cov(X_{s_j}, X_{s_i}))_{1 \le i,j \le n}$ covariance matrix

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h): $\exists c, C(s,t) = c(||t-s||) = c(h)$

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h): $\exists c, C(s,t) = c(||t-s||) = c(h)$
  $\longrightarrow$ ✓ X is sufficiently **homogeneous** inside the spatial domain

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h): $\exists c, C(s,t) = c(||t-s||) = c(h)$
  $\longrightarrow$ $\checkmark$ X is sufficiently **homogeneous** inside the spatial domain
- *Assumption 2:* X **Gaussian** RF with zero mean *(Simple Kriging)* and positive definite covariance function

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h):
  $\exists c, C(s,t) = c(||t-s||) = c(h)$
  $\longrightarrow \checkmark$ X is sufficiently **homogeneous** inside the spatial domain
- *Assumption 2:* X **Gaussian** RF with zero mean *(Simple Kriging)* and positive definite covariance function
  $\longrightarrow \checkmark$ all the laws are known

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h): $\exists c, C(s,t) = c(||t - s||) = c(h)$
  $\longrightarrow$ $\checkmark$ X is sufficiently **homogeneous** inside the spatial domain
- *Assumption 2*: X **Gaussian** RF with zero mean *(Simple Kriging)* and positive definite covariance function
  $\longrightarrow$ $\checkmark$ all the laws are known
- *Assumption 3*: **infill** asymptotic: number of observations **within** spatial domain S increases (denser and denser grid) and regular grid

Preliminaries

Assumptions

Limitation: unique realisation with only finite number of observations

Solution: stationarity assumption $\longrightarrow$ successful frequentist approach

- *Assumption 1*: X **second order stationary** and **isotropic** RF: constant mean μ and invariant covariance C (depends only on distance h): $\exists c, C(s,t) = c(||t-s||) = c(h)$
  $\longrightarrow$ $\checkmark$ X is sufficiently **homogeneous** inside the spatial domain
- *Assumption 2*: X **Gaussian** RF with zero mean *(Simple Kriging)* and positive definite covariance function
  $\longrightarrow$ $\checkmark$ all the laws are known
- *Assumption 3*: **infill** asymptotic: number of observations **within** spatial domain S increases (denser and denser grid) and regular grid
  $\longrightarrow$ $\checkmark$ allows to **control the spectrum** of the covariance matrix

Preliminaries
Spatial Analysis

Simple Kriging: **predict** the value of X at some unobserved location s, based on n sampled observations $(X_{s_i})_{i \leq n}$, assuming a **linear combination** of the observations: $f_\lambda(s) = \langle \lambda(s), \mathbf{X}(\mathbf{s}_n) \rangle$.

Preliminaries

Spatial Analysis

Simple Kriging: **predict** the value of X at some unobserved location s, based on n sampled observations $(X_{s_i})_{i\leq n}$, assuming a **linear combination** of the observations: $f_\lambda(s) = \langle \lambda(s), \mathbf{X}(\mathbf{s}_n) \rangle$.

- problem: find λ^* minimising variance, s.t. no bias
- IMSE: $L_S(f) = \mathbb{E}_X \left[\int_S (f(s) - X_s)^2 \, ds \right]$

Preliminaries

Spatial Analysis

Simple Kriging: **predict** the value of X at some unobserved location s, based on n sampled observations $(X_{s_i})_{i \leq n}$, assuming a **linear combination** of the observations: $f_\lambda(s) = \langle \lambda(s), \mathbf{X}(\mathbf{s}_n) \rangle$.

- problem: find λ^* minimising variance, s.t. no bias
- IMSE: $L_S(f) = \mathbb{E}_X \left[\int_S (f(s) - X_s)^2 \, ds \right]$
- solution: weights λ^* depend on covariance function and s: $f_{\lambda^*}(s) = {}^t\mathbf{X}(\mathbf{s}_n) \Sigma(\mathbf{s}_n)^{-1} \mathbf{c}_n(s)$

Preliminaries

Spatial Analysis

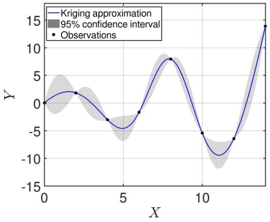
Simple Kriging: **predict** the value of X at some unobserved location s, based on n sampled observations $(X_{s_i})_{i \leq n}$, assuming a **linear combination** of the observations: $f_\lambda(s) = \langle \lambda(s), \mathbf{X}(\mathbf{s}_n) \rangle$.

- problem: find λ^* minimising variance, s.t. no bias
- IMSE: $L_S(f) = \mathbb{E}_X \left[\int_S (f(s) - X_s)^2 \, ds \right]$

- solution: weights λ^* depend on covariance function and s: $f_{\lambda^*}(s) = {}^t\mathbf{X}(\mathbf{s}_n) \Sigma(\mathbf{s}_n)^{-1} \mathbf{c}_n(s)$

Remark (Exact Interpolation):

- at observed points: exact interpolator
- at unobserved points: linear combination of data



Preliminaries

Spatial Analysis

theoretical kriging:	
covariance function known $\Longrightarrow X^*$ **BLUP** (**B**est **L**inear **U**nbiased **P**redictor)	

Preliminaries

Spatial Analysis

theoretical kriging:	**empirical kriging:**
covariance function known	in practice,
$\Longrightarrow X^*$ **BLUP**	covariance function **unknown**
(**B**est **L**inear **U**nbiased **P**redictor)	$\Longrightarrow \widehat{X}$ no guarantees of BLUP

Preliminaries

Spatial Analysis

theoretical kriging:	empirical kriging:
covariance function known	in practice,
$\Longrightarrow X^*$ **BLUP**	covariance function **unknown**
(**B**est **L**inear **U**nbiased **P**redictor)	$\Longrightarrow \widehat{X}$ no guarantees of BLUP

need to **estimate** covariance function

Preliminaries

Spatial Analysis

theoretical kriging:	**empirical kriging:**
covariance function known $\Longrightarrow X^*$ **BLUP** (**B**est **L**inear **U**nbiased **P**redictor)	in practice, covariance function **unknown** $\Longrightarrow \widehat{X}$ no guarantees of BLUP

⚠ need to **estimate** covariance function

Plug-in predictive rules:

theoretical kriging:	empirical kriging:
covariance function known $\Longrightarrow X^*$ **BLUP** (**B**est **L**inear **U**nbiased **P**redictor)	in practice, covariance function **unknown** $\Longrightarrow \widehat{X}$ no guarantees of BLUP

⚠ need to **estimate** covariance function

Plug-in predictive rules:

- covariance function **estimator**: $\widehat{c}(h) = \frac{1}{n_h} \sum X_{s_i} X_{s_j}$, where sum over set $N(h)$ of pairs at distance h and $n_h = |N(h)|$

Preliminaries

Spatial Analysis

theoretical kriging:	empirical kriging:
covariance function known	in practice,
$\Longrightarrow X^*$ **BLUP**	covariance function **unknown**
(**B**est **L**inear **U**nbiased **P**redictor)	$\Longrightarrow \widehat{X}$ no guarantees of BLUP

⚠ need to **estimate** covariance function

Plug-in predictive rules:

- covariance function **estimator**: $\widehat{c}(h) = \frac{1}{n_h} \sum X_{s_i} X_{s_j}$, where sum over set $N(h)$ of pairs at distance h and $n_h = |N(h)|$
- construct an estimator $\widehat{\lambda}$ of λ^* by replacing the unknown quantities by their estimators $\widehat{\mathbf{c}}_n(s)$ and $\widehat{\Sigma}(\mathbf{s}_n)$ **(plug-in)**

Preliminaries

Spatial Analysis

theoretical kriging:	**empirical kriging:**
covariance function known $\Longrightarrow X^*$ **BLUP** (**B**est **L**inear **U**nbiased **P**redictor)	in practice, covariance function **unknown** $\Longrightarrow \widehat{X}$ no guarantees of BLUP

⚠ need to **estimate** covariance function

Plug-in predictive rules:

- covariance function **estimator**: $\widehat{c}(h) = \frac{1}{n_h} \sum X_{s_i} X_{s_j}$, where sum over set $N(h)$ of pairs at distance h and $n_h = |N(h)|$
- construct an estimator $\widehat{\lambda}$ of λ^* by replacing the unknown quantities by their estimators $\widehat{\mathbf{c}}_n(s)$ and $\widehat{\Sigma}(\mathbf{s}_n)$ **(plug-in)**

$\Longrightarrow$ need to establish **rate bounds** that assess the **generalization capacity** of the resulting predictive map

What is Kernel Ridge Regression?

What is Kernel Ridge Regression?

- **predict** the values of a random variable for unobserved input
  - based on **independent** observations $(x_i, y_i)_{1 \leq i \leq n}$ = (input, output)
  - minimise Mean Squared Error by Empirical Risk Minimisation

What is Kernel Ridge Regression?

- **predict** the values of a random variable for unobserved input
  - based on **independent** observations $(x_i, y_i)_{1 \leq i \leq n}$ = (input, output)
  - minimise Mean Squared Error by Empirical Risk Minimisation
- **ridge regression:** penalty term $\Longrightarrow$ avoid overfitting

What is Kernel Ridge Regression?

- **predict** the values of a random variable for unobserved input
  - based on **independent** observations $(x_i, y_i)_{1 \leq i \leq n}$ = (input, output)
  - minimise Mean Squared Error by Empirical Risk Minimisation
- **ridge regression:** penalty term $\Longrightarrow$ **avoid** overfitting
- **kernel trick:** transform data space $\Longrightarrow$ solve a **non linear problem**:

$$\hat{f} = arg \min_{f \in \mathcal{H}} \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \frac{\eta}{2} \|f\|_{\mathcal{H}}^2 \right\}, \qquad \text{where } (\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}}) \text{ RKHS}$$

TELECOM Paris

What is Kernel Ridge Regression?

- **predict** the values of a random variable for unobserved input
  - based on **independent** observations $(x_i, y_i)_{1 \leq i \leq n}$ = (input, output)
  - minimise Mean Squared Error by Empirical Risk Minimisation
- **ridge regression:** penalty term $\Longrightarrow$ **avoid** overfitting
- **kernel trick:** transform data space $\Longrightarrow$ solve a **non linear problem**:

$$\hat{f} = arg \min_{f \in \mathscr{H}} \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \frac{\eta}{2} \|f\|_{\mathscr{H}}^2 \right\}, \qquad \text{where } (\mathscr{H}, \langle \cdot, \cdot \rangle_{\mathscr{H}}) \text{ RKHS}$$

$$\Longrightarrow \qquad \hat{f}(x) = {}^t\, \mathbf{Y}_n \left(\eta I_n + \mathscr{K}_n \right)^{-1} \kappa_n(x)$$

Gram matrix $\longleftrightarrow$ $\longleftrightarrow$ kernel vector

Main Results

1. Connections between SK and KRR
2. Nonasymptotic bound for the Excess of Risk
3. Illustrative Numerical Experiments

Main Results

Connections: ML ⟷ Geostat

Machine Learning:

Geostatistics:

Main Results

Connections: ML ⟷ Geostat

Machine Learning:

- **goal:** build predictor for new unobserved data

Geostatistics:

- **goal:** build complete map (prediction at each unobserved location)

Main Results

Connections: ML $\longleftrightarrow$ Geostat

Machine Learning:

- **goal:** build predictor for new unobserved data
- choice of **kernel** function

Geostatistics:

- **goal:** build complete map (prediction at each unobserved location)
- nonparametric **covariance** estimation

Main Results

Connections: ML $\longleftrightarrow$ Geostat

Machine Learning:

- **goal:** build predictor for new unobserved data
- choice of **kernel** function

$$\hat{f}(x) = {}^t\, \mathbf{Y}_n \left(\eta I_n + \mathcal{K}_n\right)^{-1} \kappa_n(x)$$

Geostatistics:

- **goal:** build complete map (prediction at each unobserved location)
- nonparametric **covariance** estimation

$$f_{\hat{\lambda}}(s) = {}^t\, \mathbf{X}(\mathbf{s}_n)\widehat{\Sigma}(\mathbf{s}_n)^{-1}\widehat{\mathbf{c}}_n(s)$$

Main Results

Connections: ML ⟷ Geostat

Machine Learning:

- **goal:** build predictor for new unobserved data
- choice of **kernel** function

$$\hat{f}(x) = {}^t\,\mathbf{Y}_n\,(\eta I_n + \mathscr{K}_n)^{-1}\,\kappa_n(x)$$

Geostatistics:

- **goal:** build complete map (prediction at each unobserved location)
- nonparametric **covariance** estimation

$$f_{\hat{\lambda}}(s) = {}^t\,\mathbf{X}(\mathbf{s}_n)\widehat{\Sigma}(\mathbf{s}_n)^{-1}\widehat{\mathbf{c}}_n(s)$$

$\Longrightarrow$ kriging predictor **same** form as kernel ridge regressor:

- regularised Gram matrix ⟷ covariance matrix
- kernel vector ⟷ covariance vector

Main Results

Nonasymptotic theory

What global gap between optimal **theoretical** predictor and **empirical** predictor **errors**?

Main Results

Nonasymptotic theory

What global gap between optimal **theoretical** predictor and **empirical** predictor **errors**?

Concentration Inequality:

- predictor accuracy measured by MSE
- global excess risk quantify gap

Main Results

Nonasymptotic theory

What global gap between optimal **theoretical** predictor and **empirical** predictor **errors**?

Concentration Inequality:

- predictor accuracy measured by MSE
- global excess risk quantify gap

$\Longrightarrow$ **Goal:** define **nonasymptotic** bound of global excess risk:

What global gap between optimal **theoretical** predictor and **empirical** predictor **errors**?

Concentration Inequality:

- predictor accuracy measured by MSE
- global excess risk quantify gap

$\Longrightarrow$ **Goal:** define **nonasymptotic** bound of global excess risk:

$$L_S(f_{\widehat{\lambda}}) - L_S(f_{\lambda^*}) = \mathbb{E}_X\left[\int_{s\in S}\left(f_{\widehat{\lambda}}(s) - f_{\lambda^*}(s)\right)^2 ds\right]$$

Main Results

Additional Assumptions

- *Assumption 4*: $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

Main Results

Additional Assumptions

- *Assumption 4:* $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

  $\longrightarrow$ ✓ allows to give an **inferior bound** on n_h the number of pairs at distance h

Main Results

Additional Assumptions

- *Assumption 4:* $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

  $\longrightarrow$ √ allows to give an **inferior bound** on n_h the number of pairs at distance h

- *Assumption 5:* $\exists\, m, M < \infty$, eigenvalues of covariance matrix are bounded

Main Results

Additional Assumptions

- *Assumption 4:* $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

  $\longrightarrow$ ✓ allows to give an **inferior bound** on n_h the number of pairs at distance h

- *Assumption 5:* $\exists\, m, M < \infty$, eigenvalues of covariance matrix are bounded

  $\longrightarrow$ ✓ taken into account in the **tail bounds** of the estimators

Main Results

Additional Assumptions

- *Assumption 4:* $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

  $\longrightarrow$ $\surd$ allows to give an **inferior bound** on n_h the number of pairs at distance h

- *Assumption 5:* $\exists\, m, M < \infty$, eigenvalues of covariance matrix are bounded

  $\longrightarrow$ $\surd$ taken into account in the **tail bounds** of the estimators

- *Assumption 6:* c is of class $\mathscr{C}^1$ and its gradient is bounded by D **(regularity hypothesis)**

Main Results

Additional Assumptions

- *Assumption 4:* $\exists\, j_1 \geq 1,\ \forall h \geq 1 - 2^{-j_1},\ c(h) = 0$ **(border hypothesis)**

  $\longrightarrow$ $\checkmark$ allows to give an **inferior bound** on n_h the number of pairs at distance h

- *Assumption 5:* $\exists\, m, M < \infty$, eigenvalues of covariance matrix are bounded

  $\longrightarrow$ $\checkmark$ taken into account in the **tail bounds** of the estimators

- *Assumption 6:* c is of class $\mathscr{C}^1$ and its gradient is bounded by D **(regularity hypothesis)**

  $\longrightarrow$ $\checkmark$ taken into account when studying the **estimation error** of the covariance function for **all lags** (**smoothness** of the covariance function)

Main Results

Nonasymptotic bound for Excess of Risk

Theorem ([Siviero et al., 2022])

For any $\delta \in (0, 1)$, we have with probability at least $1 - \delta$:

$$L_S(f_{\widehat{\lambda}}) - L_S(f_{\lambda^*}) \leq C_6 \frac{\log(n/\delta)}{n} + \frac{2D^2}{n+1},$$

as soon as $n \geq C_6' \log(n/\delta)$, where C_6 and C_6' are nonnegative constants depending on j_1, m and M solely.

Sketch of Proof

Poisson tail bounds

Hyp 5
bounded eigenvalues

Poisson tail bounds for deviations $|\widehat{\gamma}(h) - \gamma(h)|$ and $|\widehat{c}_h(0) - c(0)|$

↑ ↑

Hyp 2
Gaussian hyp

Distributions of estimators $\widehat{c}_h(0)$ and $\widehat{\gamma}(h)$

Inferior bound on $\mathbf{n_h}$

Hyp 4
border hyp

Sketch of Proof

Bound on covariance estimation error

Hyp 6
regularity hyp

Bound on covariance estimation error for **all lags**

↑

Hyp 5
bounded eigenvalues

Poisson tail bounds for deviations $|\widehat{\gamma}(h) - \gamma(h)|$ and $|\widehat{c}_h(0) - c(0)|$

↑ ↑

Hyp 2
Gaussian hyp

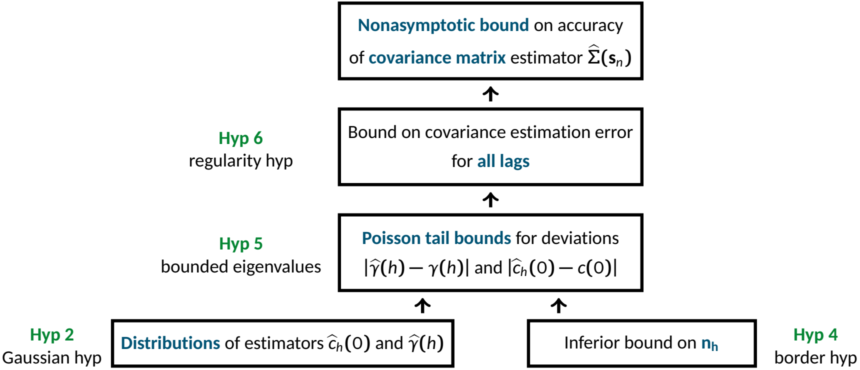
Distributions of estimators $\widehat{c}_h(0)$ and $\widehat{\gamma}(h)$

Inferior bound on $\mathbf{n_h}$

Hyp 4
border hyp

Sketch of Proof

Bound on accuracy of $\widehat{\Sigma}(s_n)$



Sketch of Proof

Bound on accuracy of $\widehat{\Sigma}(s_n)$

What can we tell about the **covariance matrix accuracy**?

Sketch of Proof

Bound on accuracy of $\widehat{\Sigma}(s_n)$

What can we tell about the **covariance matrix accuracy**?

- under **Hyp 2**, $\widehat{\Sigma}(s_n)$ is symmetric and semi-positive definite

Sketch of Proof

Bound on accuracy of $\widehat{\Sigma}(s_n)$

What can we tell about the **covariance matrix accuracy**?

- under **Hyp 2**, $\widehat{\mathbf{\Sigma}}(s_n)$ is symmetric and semi-positive definite
- Bochner's theorem applies: bound deviation thanks to the **connection between covariance matrix and spectral density** [Hall and Patil, 1994]

Sketch of Proof

Bound on accuracy of $\widehat{\Sigma}(s_n)$

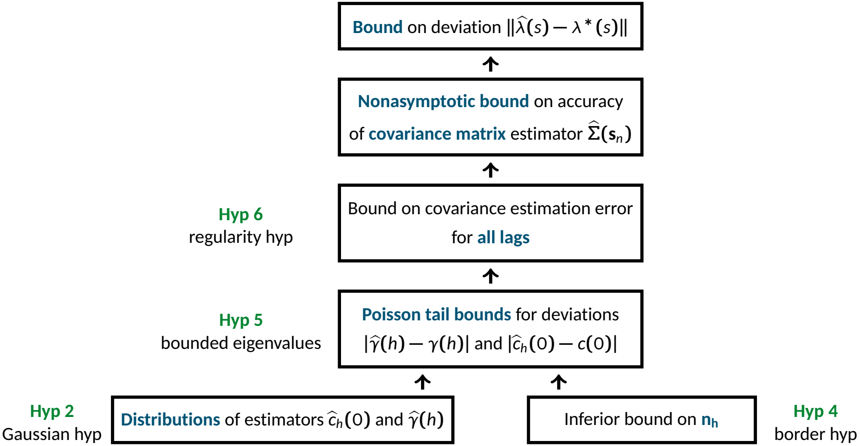
What can we tell about the **covariance matrix accuracy**?

- under **Hyp 2**, $\widehat{\boldsymbol{\Sigma}}(s_n)$ is symmetric and semi-positive definite
- Bochner's theorem applies: bound deviation thanks to the **connection between covariance matrix and spectral density** [Hall and Patil, 1994]

$\Longrightarrow$ nonasymptotic bound on accuracy of covariance matrix estimator $\rho\left(\widehat{\boldsymbol{\Sigma}}(\mathbf{s}_n) - \boldsymbol{\Sigma}(\mathbf{s}_n)\right)$

Sketch of Proof

$$\forall s \in S, \quad \left(f_{\widehat{\lambda}}(s) - f_{\lambda^*}(s)\right)^2 \leq \|\widehat{\lambda}(s) - \lambda^*(s)\|^2 \times \|\mathbf{X}(\mathbf{s}_n)\|^2$$



Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

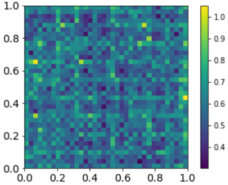


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

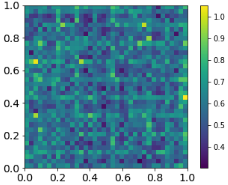


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

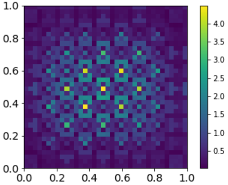


Figure 2: *Gaussian*

✗ not **Hyp 4**, but vanishes quickly

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

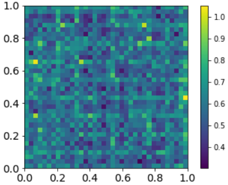


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

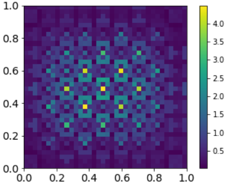


Figure 2: *Gaussian*

✗ not **Hyp 4**, but vanishes quickly

Remark: experiments **corroborate** our theoretical results

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

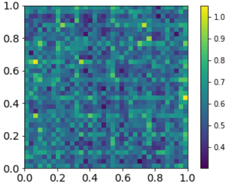


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

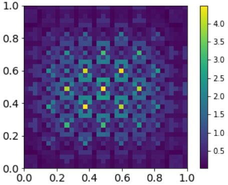


Figure 2: *Gaussian*

✗ not **Hyp 4**, but vanishes quickly

Remark: experiments **corroborate** our theoretical results

- both covariance models: for small values of j_1, **excess of risk is small**

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

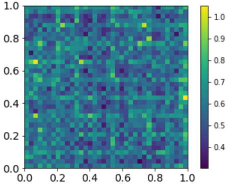


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

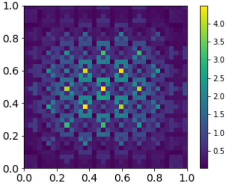


Figure 2: *Gaussian*

✗ not **Hyp 4**, but vanishes quickly

Remark: experiments **corroborate** our theoretical results

- both covariance models: for small values of j_1, **excess of risk is small**
- **bound depends on** j_1 (role of technical assumptions is verified)

Illustrative Numerical Experiments

Maps of MSE on 100 realisations, with two covariance models (for fixed j_1)

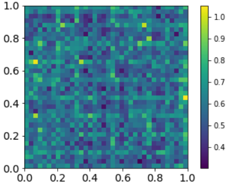


Figure 1: *Truncated power law* (TPL)

✓ satisfies all the assumptions

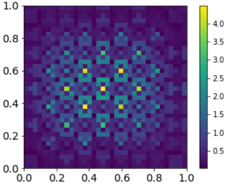


Figure 2: *Gaussian*

✗ not **Hyp 4**, but vanishes quickly

Remark: experiments **corroborate** our theoretical results

- both covariance models: for small values of j_1, **excess of risk is small**
- **bound depends on** j_1 (role of technical assumptions is verified)

$\Longrightarrow$ results for Gaussian covariance encourages to relax **Hyp 4**

Ongoing and future work

- incremental way to proceed

Ongoing and future work

- incremental way to proceed
- gradually **relax** some hypotheses:

Ongoing and future work

- incremental way to proceed
- gradually **relax** some hypotheses:
  - Assumption 4 (*border hypothesis*) less restrictive: $c(h) \searrow 0$
  - Assumption 6 (*regularity hypothesis*): other smoothing techniques
  - *Irregular grid*: biased semi-variogram estimator, inferior bound on n_h
  - Assumption 2: outside the *Gaussian* case

Ongoing and future work

- incremental way to proceed
- gradually **relax** some hypotheses:
  - Assumption 4 (*border hypothesis*) less restrictive: $c(h) \searrow 0$
  - Assumption 6 (*regularity hypothesis*): other smoothing techniques
  - *Irregular grid*: biased semi-variogram estimator, inferior bound on n_h
  - Assumption 2: outside the *Gaussian* case
- **partitioning methods** (dyadic CART): divide spatial domain into clusters where the process is *locally stationary*

Conclusion

- novel theoretical framework offering **guarantees** for empirical simple Kriging rules in the form of **non-asymptotic bounds**

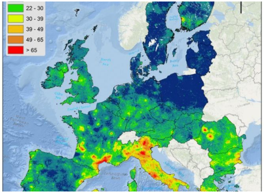
Conclusion

- novel theoretical framework offering **guarantees** for empirical simple Kriging rules in the form of **non-asymptotic bounds**
- theoretical guarantees for **dependent** databases

Conclusion

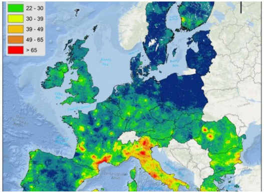
- novel theoretical framework offering **guarantees** for empirical simple Kriging rules in the form of **non-asymptotic bounds**
- theoretical guarantees for **dependent** databases

  $\Longrightarrow$ industrial assets and many possible **applications**



Conclusion

- novel theoretical framework offering **guarantees** for empirical simple Kriging rules in the form of **non-asymptotic bounds**
- theoretical guarantees for **dependent** databases
  $\Longrightarrow$ industrial assets and many possible **applications**
- **possible extensions** to a more general framework



Thanks for your attention !

References I

[Hall and Patil, 1994] Hall, P. and Patil, P. (1994).
Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields.
Probability Theory and Related Fields, 99(3):399–424.

[Siviero et al., 2022] Siviero, E., Chautru, E., and Clémençon, S. (2022).
A Statistical Learning View of Simple Kriging.