Random fields and their discretizations

Felipe Uribe

Computational Engineering
School of Engineering Sciences
Lappeenranta-Lahti University of Technology (LUT)

Special Course on Inverse Problems
Lappeenranta, FI — January-February, 2024

Why random fields?

- Spatial variation is a common characteristic of uncertain phenomena occurring in several areas of science.
- Therefore, computational models are controlled by parameters that are functions on a given spatial domain.

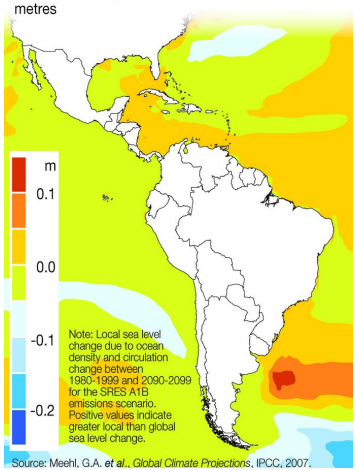


Figure: Changes in the sea level.

Why random fields?

- Spatial variation is a common characteristic of uncertain phenomena occurring in several areas of science.
- Therefore, computational models are controlled by parameters that are functions on a given spatial domain.
- The patterns of uncertain quantities induced by spatial variation are complex and one requires a probabilistic description for their analysis and interpretation.
- The theory of *stochastic processes* provides a tool for studying random phenomena that vary in time. This theory can be generalized to deal with models that vary randomly in space, so-called random fields.

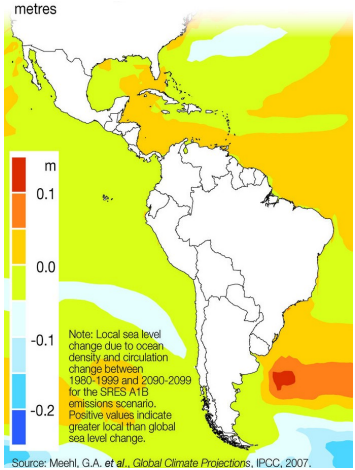


Figure: Changes in the sea level.

PART I: definition of a random field

Random fields: intro I

- Random fields can be roughly understood as an arranged set of random variables.
- The arrangement is specified by the index set pointing to different spatial locations of some topological space, which by definition is infinite-dimensional.

Random fields: intro I

- Random fields can be roughly understood as an arranged set of random variables.
- The arrangement is specified by the index set pointing to different spatial locations of some topological space, which by definition is infinite-dimensional.
- In practice, one seeks a suitable finite-dimensional representation of the random field. Most of the available techniques for random field discretization require not only the specification of a set of spatial points, but also information provided by the expectation and covariance.
- Some people in the field.
  - Bertil Matérn (1917-2007)[1], Robert J. Adler (1950-)[2] Michael L. Stein[3], Håvard Rue[4], ...

[1] B. Matérn. *Spatial variation: stochastic models and their application to some problems in forest surveys and other sampling investigations*. Report No. 49/5. Forest Research Institute of Sweden, 1960.

[2] R. J. Adler. *The geometry of random fields*. Society for Industrial and Applied Mathematics (SIAM), 2010.

[3] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer, 1999.

[4] H. Rue and L. Held. *Gaussian Markov Random Fields. Theory and Applications*. Chapman & Hall/CRC, 2005.

Random fields: intro II

- Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, together with a bounded index set $D \subseteq \mathbb{R}^n$ representing a physical domain.
- Let us define $G^{(n,d)} := \{g : \mathbb{R}^n \to \mathbb{R}^d;\ d, n \geq 1\}$ to be the set of functions mapping from D to $\mathbb{R}^d$.
- We can also define the σ-algebra induced by that mapping,

$$\mathcal{G}^{(n,d)} := \{g \in G^{(n,d)} : g(x_j) \in B_j\}, \quad x_j \in D, j = 1, \ldots, k,$$

where $B_j = B_j(\mathbb{R}^d)$ are Borel sets (half-open intervals) on $\mathbb{R}^d$ and k is an arbitrary integer.

Random fields: intro II

- Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, together with a bounded index set $D \subseteq \mathbb{R}^n$ representing a physical domain.
- Let us define $G^{(n,d)} := \{g : \mathbb{R}^n \to \mathbb{R}^d;\ d, n \geq 1\}$ to be the set of functions mapping from D to $\mathbb{R}^d$.
- We can also define the σ-algebra induced by that mapping,

$$\mathcal{G}^{(n,d)} := \{g \in G^{(n,d)} : g(x_j) \in B_j\}, \quad x_j \in D, j = 1, \ldots, k,$$

where $B_j = B_j(\mathbb{R}^d)$ are Borel sets (half-open intervals) on $\mathbb{R}^d$ and k is an arbitrary integer.

Random field

A (n, d)-dimensional *random field* is a measurable transformation $H(\boldsymbol{x}, \omega)$ from the space $(\Omega, \mathcal{F})$ into the *realization space* $(G^{(n,d)}, \mathcal{G}^{(n,d)})$.

Random fields: intro III

- From this definition, a random field can be simply understood as a random variable that takes values in a function space. That is: A random field $H(\boldsymbol{x}, \omega)$ on $D \subseteq \mathbb{R}^n$ is a function whose values are random variables for any $\boldsymbol{x} \in D \subseteq \mathbb{R}^n$.
- Random fields are usually denoted by $H(\boldsymbol{x}, \omega)$, meaning the value that a function in $G^{(n,d)}$ takes at the location $\boldsymbol{x}$ for a given $\omega \in \Omega$.
- For fixed ω_i, the $\mathbb{R}^d$-valued function $h_i(\boldsymbol{x}) = H(\boldsymbol{x}, \omega_i)$ is a *realization* or sample path of the random field. Conversely, for a fixed $\boldsymbol{x}_j \in D$, the function $H_j(\omega) = H(\boldsymbol{x}_j, \omega)$ is a RV.
- It is useful to think of $\boldsymbol{x}$ as "spatial coordinate" and of each ω as a "particle" or "experiment".

$(1, d)$-dimensional random field

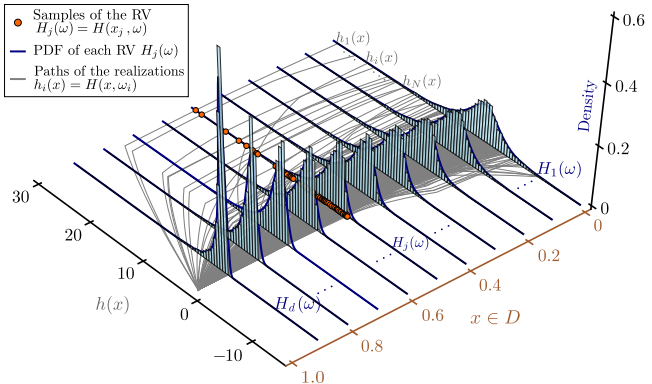
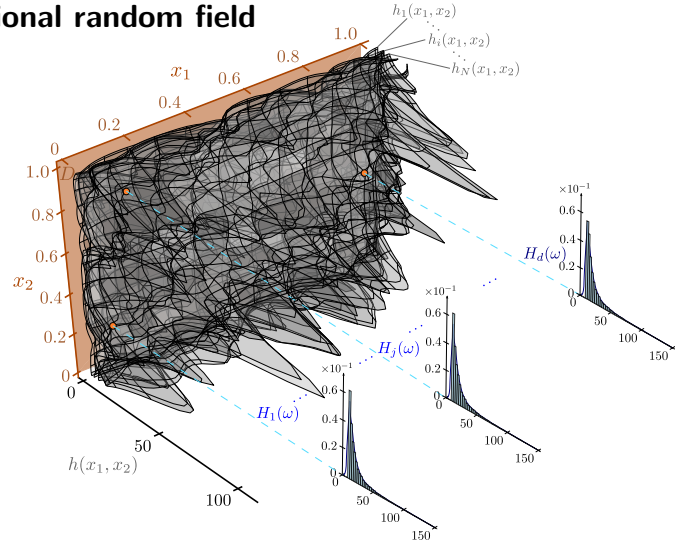


Figure: For a fixed element of the sample space $\omega_i \in \Omega$, the random field turns into a realization $h_i(x)$ across the spatial domain $D = [0, 1]$. For a fixed spatial coordinate $x_j \in D$, the random field turns into a random variable $H_j(\omega)$.

$(2, d)$-dimensional random field



Random fields: fi-di distributions I

- Given a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, one can define a corresponding law or distribution $\mathbb{P}_H$ in the realization space $(G^{(n,d)}, \mathscr{G}^{(n,d)})$.
- This corresponds to the product measure

$$\mathbb{P}_H(B_1 \times \cdots \times B_k) = \mathbb{P}[H(\boldsymbol{x}_1, \omega) \in B_1, \ldots, H(\boldsymbol{x}_k, \omega) \in B_k] \,.$$

Random fields: fi-di distributions I

- Given a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, one can define a corresponding law or distribution μ in the realization space $(G^{(n,d)}, \mathscr{G}^{(n,d)})$.
- This corresponds to the product measure

$$\mathbb{P}_H(B_1 \times \cdots \times B_k) = \mathbb{P}[H(\boldsymbol{x}_1, \omega) \in B_1, \ldots, H(\boldsymbol{x}_k, \omega) \in B_k] \,.$$

- The associated collection of k CDFs given by

$$\begin{aligned} F_{H_1,\ldots,H_k}(y_1, \ldots, y_k) &= \mathbb{P}[H(\boldsymbol{x}_1, \omega) \leq y_1, \ldots, H(\boldsymbol{x}_k, \omega) \leq y_k] && \text{(2a)} \\ &= \mathbb{P}[H_1(\omega) \leq y_1, \ldots, H_k(\omega) \leq y_k] \,, && \text{(2b)} \end{aligned}$$

with $k \leq d$, is known as the family of finite-dimensional (fi-di) distributions of the random field H [2].

Random fields: fi-di distributions II

- Random fields can be directly defined in terms of the family of fi-di distributions provided they exist and satisfy the consistency and symmetry conditions.

- Kolmogorov's extension theorem (a random field can be uniquely extended from its fi-di distributions) [17, p.11]:
  - (i) *Symmetry* (permutation invariance): $F_{H_1,\ldots,H_k}(y_1,\ldots,y_k)$ is invariant under arbitrary permutation of the indices $j = 1,\ldots,k$ (for both $\boldsymbol{x}_j, y_j$).
  - (ii) *Consistency* (projection invariance): $F_{H_1,\ldots,H_{k+1}}(y_1,\ldots,y_{k+1}) = F_{H_1,\ldots,H_k}(y_1,\ldots,y_k)$, as $y_{k+1} \to \infty$.

Random fields: fi-di distributions II

- Random fields can be directly defined in terms of the family of fi-di distributions provided they exist and satisfy the consistency and symmetry conditions.

- Kolmogorov's extension theorem [17, p.11] (a random field can be uniquely extended from its fi-di distributions):
  - (i) *Symmetry* (permutation invariance): $F_{H_1,\ldots,H_k}(y_1,\ldots,y_k)$ is invariant under arbitrary permutation of the indices $j = 1,\ldots,k$ (for both $\boldsymbol{x}_j, y_j$).
  - (ii) *Consistency* (projection invariance): $F_{H_1,\ldots,H_{k+1}}(y_1,\ldots,y_{k+1}) = F_{H_1,\ldots,H_k}(y_1,\ldots,y_k)$, as $y_{k+1} \to \infty$.

- The characterization established by the fi-di distributions is the most used in practice since it is more intuitive to think of a random field as a collection of random variables representing uncertain values at each spatial coordinate in D.

PART II: random field properties: statistics and regularity

Random fields: statistical properties I

- Consider a **second-order** random field expressed by the collection of random variables $H_{\boldsymbol{x}} = \{H(\boldsymbol{x}_1, \omega), \ldots, H(\boldsymbol{x}_j, \omega), \ldots, H(\boldsymbol{x}_d, \omega)\}$ with joint distribution $F_{H_{\boldsymbol{x}}}$ and density $\pi_{H_{\boldsymbol{x}}}$.

Random fields: statistical properties I

- Consider a **second-order** random field expressed by the collection of random variables $H_{\boldsymbol{x}} = \{H(\boldsymbol{x}_1, \omega), \ldots, H(\boldsymbol{x}_j, \omega), \ldots, H(\boldsymbol{x}_d, \omega)\}$ with joint distribution $F_{H_{\boldsymbol{x}}}$ and density $\pi_{H_{\boldsymbol{x}}}$.
- The expectation or mean of a random field is the $\mathbb{R}^d$-valued function

$$\mu_H(\boldsymbol{x}) = \mathbb{E}[H(\boldsymbol{x}, \omega)] = \int_{\mathbb{R}^n} \boldsymbol{y} \, \mathrm{d}F_{H_{\boldsymbol{x}}}(\boldsymbol{y}) = \int_{\mathbb{R}^n} \boldsymbol{y} \, \pi_{H_{\boldsymbol{x}}}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}. \tag{3}$$

- The second moment or correlation of the random field $H(\boldsymbol{x}, \omega)$ with itself $H(\boldsymbol{x}', \omega)$ (i.e., the autocorrelation) is given by the $\mathbb{R}^{d \times d}$-valued function

$$R_{HH}(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[H(\boldsymbol{x}, \omega) H(\boldsymbol{x}', \omega)] = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \boldsymbol{y} \, \boldsymbol{y}' \, \pi_{H_{\boldsymbol{x}}, H_{\boldsymbol{x}'}}(\boldsymbol{y}, \boldsymbol{y}') \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{y}'. \tag{4}$$

Random fields: statistical properties II

- Similarly, the covariance of the random field $H(\boldsymbol{x}, \omega)$ with itself $H(\boldsymbol{x}', \omega)$ is the $\mathbb{R}^{d\times d}$-valued function

$$C_{HH}(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{C}\mathrm{ov}\left[H(\boldsymbol{x}, \omega), H(\boldsymbol{x}', \omega)\right] = \mathbb{E}[H(\boldsymbol{x}, \omega)H(\boldsymbol{x}', \omega)] - \mu_H(\boldsymbol{x})\mu_H(\boldsymbol{x}') \tag{5a}$$

$$= R_{HH}(\boldsymbol{x}, \boldsymbol{x}') - \mu_H(\boldsymbol{x})\mu_H(\boldsymbol{x}'). \tag{5b}$$

- Moreover, the normalized covariance function

$$\rho_{HH}(\boldsymbol{x}, \boldsymbol{x}') = \frac{C_{HH}(\boldsymbol{x}, \boldsymbol{x}')}{\sigma_H(\boldsymbol{x})\sigma_H(\boldsymbol{x}')}, \tag{6}$$

is called the correlation coefficient function, where $\sigma_H(\boldsymbol{x})$ corresponds to the *standard deviation* of the field.

Random fields: statistical properties III

- A random field is strictly homogeneous or just homogeneous, if the distribution functions are invariant under arbitrary translations in space. That is, for all spatial coordinates $\boldsymbol{x}' \in D$

$$\mathbb{P}[H(\boldsymbol{x}_1,\omega) \leq y_1,\ldots,H(\boldsymbol{x}_k,\omega) \leq y_k] = \mathbb{P}[H(\boldsymbol{x}_1+\boldsymbol{x}',\omega) \leq y_1,\ldots,H(\boldsymbol{x}_k+\boldsymbol{x}',\omega) \leq y_k], \tag{7}$$

  where each $(\boldsymbol{x}_j+\boldsymbol{x}') \in D$ (with $j=1,\ldots,k$ and $k \leq d$).

Random fields: statistical properties III

- A random field is strictly homogeneous or just homogeneous, if the distribution functions are invariant under arbitrary translations in space. That is, for all spatial coordinates $\boldsymbol{x}' \in D$

$$\mathbb{P}[H(\boldsymbol{x}_1,\omega) \leq y_1, \ldots, H(\boldsymbol{x}_k,\omega) \leq y_k] = \mathbb{P}[H(\boldsymbol{x}_1+\boldsymbol{x}',\omega) \leq y_1, \ldots, H(\boldsymbol{x}_k+\boldsymbol{x}',\omega) \leq y_k], \tag{7}$$

  where each $(\boldsymbol{x}_j + \boldsymbol{x}') \in D$ (with $j = 1, \ldots, k$ and $k \leq d$).

- The random field is weakly homogeneous, if the mean function is constant across D and the correlation function only depends on the *separation vector* $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{x}'$, that is

$$\mu_H(\boldsymbol{x}) = \mu_H < \infty \qquad \text{and} \qquad R_{HH}(\boldsymbol{x},\boldsymbol{x}') = R_{HH}(\boldsymbol{h}) = R_{HH}(-\boldsymbol{h}). \tag{8}$$

- Under eq. (8), the covariance function eq. (5) reduces to $C_{HH}(\boldsymbol{h}) = \sigma_H^2 \rho_{HH}(\boldsymbol{h})$, since the variance of a homogeneous field is also constant, $\sigma_H^2(\boldsymbol{x}) = \sigma_H^2 < \infty$.

Random fields: statistical properties IV

- Furthermore, if the homogeneous correlation/covariance function is also independent of the direction (rotations and reflections), the random field is called isotropic.
- In this case, the covariance is only a function of the norm on D. Since in most cases $D \subset \mathbb{R}^n$, we employ the Euclidean norm $\|\boldsymbol{h}\|_2 = h$ and we write $C_{HH}(\boldsymbol{h}) = C_{HH}(h)$.

Random fields: statistical properties IV

- Furthermore, if the homogeneous correlation/covariance function is also independent of the direction (rotations and reflections), the random field is called *isotropic*.
- In this case, the covariance is only a function of the norm on D. Since in most cases $D \subset \mathbb{R}^n$, we employ the Euclidean norm $\|\boldsymbol{h}\|_2 = h$ and we write $C_{HH}(\boldsymbol{h}) = C_{HH}(h)$.
- One can also use a different norm to compute the distances between the spatial coordinates. This provides more flexibility in the application of homogeneous covariance functions.
- The dependence structure of physical phenomena is oftentimes spatially varying. This requires the application of non-homogeneous random fields, which in some cases can be derived from homogeneous ones.

Random fields: regularity I

- It is important that the realizations of a random field satisfy some regularity properties.
- The most fundamental regularity condition one requires in practice is continuous differentiability, which is directly related to the *smoothness* of the field.

Random fields: regularity I

- It is important that the realizations of a random field satisfy some regularity properties.
- The most fundamental regularity condition one requires in practice is continuous differentiability, which is directly related to the *smoothness* of the field.
- For example, Gaussian random fields are considered **regular**, if they have sample function continuity and differentiability.
- Continuity of random fields is associated to the convergence of sequences of random variables $\{H(\boldsymbol{x}_1, \omega), \ldots, H(\boldsymbol{x}_d, \omega)\}$ as $d \to \infty$.

Random fields: regularity II

- A random field is mean-squared continuous in D, if

$$\mathbb{E}\Big[\big(H(\boldsymbol{x}_k,\omega) - H(\boldsymbol{x},\omega)\big)^2\Big] \to 0 \qquad \text{as } k \to \infty,\ \forall\ \boldsymbol{x} \in D. \tag{9}$$

- Mean-squared continuity does not imply continuous sample paths. However, the condition in eq. (9) has an important relation with the covariance function:

  If $C_{HH}(\boldsymbol{x},\boldsymbol{x}')$ is continuous at every diagonal point $\boldsymbol{x} = \boldsymbol{x}'$, then the random field is continuous every-where, i.e., continuous variance [1]. Particularly, for homogeneous random fields the continuity holds, if $C_{HH}(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = \boldsymbol{0}$.

PART III: Gaussian random fields and covariance kernels

Gaussian random fields

- A Gaussian random field $H(\boldsymbol{x}, \omega)$ is a random field whose fi-di distributions are all multivariate Gaussian. In this case, the field is determined by the mean and covariance functions.

- Gaussian fields are fundamental in the study of spatial variation, not only because they are reasonable models in some applications, but also (and mainly) because they have a simple construction that enables analytical tractability.

- If we define the partition $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$, such that $\boldsymbol{X}_1 = [X_1, \ldots, X_k]$ and $\boldsymbol{X}_2 = [X_{k+1}, \ldots, X_d]$, the mean vector and covariance matrix can be divided in terms of individual and crossed components associated to each grouping

$$\boldsymbol{\mu}_X = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \boldsymbol{\Sigma}_{XX} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}^{\mathsf{T}} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \tag{10}$$

Gaussian random fields

- The fi-di distributions induced by the Gaussian assumption satisfy the conditions of consistency and symmetry (see, e.g., [1, Sec. 1.4]).
- Thus, the main task for the definition of Gaussian random fields is the specification of a proper correlation/covariance function. This is because the sample path continuity of a Gaussian field is basically controlled by this function.

Gaussian random fields

- The fi-di distributions induced by the Gaussian assumption satisfy the conditions of consistency and symmetry (see, e.g., [1, Sec. 1.4]).
- Thus, the main task for the definition of Gaussian random fields is the specification of a proper correlation/covariance function. This is because the sample path continuity of a Gaussian field is basically controlled by this function.
- Since almost every continuous correlation function with an exponential decay complies with this condition, [1, p.20] states that "Gaussian random fields with continuous mean and covariance functions will generate continuous sample paths".
- Non-Gaussian random fields can be generated from Gaussian ones. The process usually involves the application of nonlinear mappings that preserve probability in both underlying spaces, so-called isoprobabilistic transformations.

Covariance operators and kernels I

- Covariance kernels for random field modeling are models used to define the particular spatial correlation of a random field.

- Three essential properties of covariance functions that follow from its definition are:
  (i) *symmetry*, $C_{HH}(\boldsymbol{x}, \boldsymbol{x}') = C_{HH}(\boldsymbol{x}', \boldsymbol{x})$;
  (ii) *positive definiteness*, i.e.,

$$\sum_{i=1}^{d} \sum_{j=1}^{d} c_i c_j C_{HH}(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0,$$

  for each $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d\} \in D$ and $c_1, \ldots, c_d \in \mathbb{R}$;
  (iii) *continuity*.

Covariance operators and kernels II

- In general, covariance functions are understood as kernels. Consider $D \subset \mathbb{R}^n$ be a bounded domain. A function $k : D \times D \to \mathbb{R}$ is called a Hilbert–Schmidt kernel if

$$\int_D \int_D |k(x, y)|^2 \, \mathrm{d}x\mathrm{d}y < \infty; \tag{11}$$

the associated integral operator $K : L^2(D) \to L^2(D)$, defined as

$$(Ku)(x) = \int_D k(x, y)u(y)\mathrm{d}y \qquad \forall\ u \in L^2(D) \tag{12}$$

is called a Hilbert–Schmidt operator.

Covariance operators and kernels II

- In general, covariance functions are understood as kernels. Consider $D \subset \mathbb{R}^n$ be a bounded domain. A function $k : D \times D \to \mathbb{R}$ is called a *Hilbert–Schmidt kernel* if

$$\int_D \int_D |k(x, y)|^2 \, \mathrm{d}x\mathrm{d}y < \infty; \tag{13}$$

the associated integral operator $K : L^2(D) \to L^2(D)$, defined as

$$(Ku)(x) = \int_D k(x, y)u(y)\mathrm{d}y \qquad \forall\ u \in L^2(D) \tag{14}$$

is called a *Hilbert–Schmidt operator*.

- Here, K is a linear operator in an infinite-dimensional setting. In practice, a spatial discretization is imposed on D to define a finite representation of such operator, this process generates a so-called Gram matrix. In the context of covariance kernels, the resulting finite-dimensional operator is the covariance matrix.

Covariance operators and kernels II

Reproducing Kernel Hilbert Space

Let $\mathcal{H}$ be a Hilbert space of real functions f defined on an index set D. Then $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS) endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ (and norm $\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}}$), if there exist a kernel k with the properties:

- for every $\boldsymbol{x}$, $k(\boldsymbol{x}, \boldsymbol{x}')$ as a function of $\boldsymbol{x}'$ belongs to $\mathcal{H}$, and
- k has the reproducing property, $\langle f(\cdot), k(\cdot, \boldsymbol{x}') \rangle_{\mathcal{H}} = f(\boldsymbol{x}')$.

- For every positive definite function $k(\cdot, \cdot)$ there exists a unique RKHS, and vice versa.
- Different parametric families of covariance functions are available in the literature, these include the Cauchy, Bessel and exponential families (several examples are reported in [1, 5]). A well-known family in spatial statistics is the Matérn class.

Covariance operators and kernels: some examples in 1D

- Triangular kernel:

$$k(x, x') = \sigma_H^2(1 - \ell\,|x - x'|) \quad \text{with} \quad |x - x'| \in \left[0, \frac{1}{\ell}\right], \tag{15}$$

  where ℓ is a parameter to adjust the distance of zero correlation.

- Wiener kernel:

$$k(x, x') = \sigma_H^2 \min(x, x'). \tag{16}$$

- Brownian bridge kernel:

$$k(x, x') = \sigma_H^2(\min(x, x') - xx'). \tag{17}$$

Covariance operators and kernels: Matérn I

- The original Matérn kernel is [10, p.18]

$$C_\nu(h) = \sigma_H^2 \frac{2}{\Gamma(\nu)} \left(\frac{bh}{2} \right)^\nu K_\nu (bh) \,, \tag{18}$$

  where $h = \|\boldsymbol{x} - \boldsymbol{x}'\|_2$, $\Gamma(\cdot)$ is the gamma function, $K_\nu(\cdot)$ is the modified Bessel function of the second kind, and the constants $b, \nu > 0$.

- A similar model was proposed in [16]:

$$C_\nu(h) = \sigma_H^2 \frac{1}{\Gamma(\nu + 1)} \left(\frac{h}{2b} \right)^\nu K_\nu (bh) \,. \tag{19}$$

- Due to the similarities between eq. (18) and eq. (19), the kernels are sometimes grouped together as the Whittle–Matérn class.

Covariance operators and kernels: Matérn II

- However, [6] introduced the standard Matérn family formulation which re-parameterizes eq. (18) in terms of a scale parameter $\ell \in \mathbb{R}_{>0}$ controlling the range of correlation (*correlation length*) and a *smoothing parameter* $\nu \in \mathbb{R}_{>0}$ controlling the smoothness of the random field. In this case, the Matérn kernel is expressed as

$$C_\nu(h) = \sigma_H^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (bh)^\nu K_\nu (bh) \tag{20}$$

  where $b = (\sqrt{2}\nu h)/\ell$.

- A slightly different version of eq. (20) is proposed in [7], where $b = (2\sqrt{\nu}h)/\ell$ is used.

- In general, there exist several parametrizations of the Matérn kernel each of which have particular advantages in different fields (see for instance [15, p.31],[4]).

Covariance operators and kernels: Matérn III

- We employ the following type of Matérn covariance kernel [13]

$$C_\nu(h) = \sigma_H^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}h}{\ell} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}h}{\ell} \right), \tag{21}$$

since we can define the special case $\nu = 1/2$ and limiting case $\nu \to \infty$:

$$C_{1/2}(h) = \sigma_H^2 \exp\left(-\frac{h}{\ell} \right) \qquad \text{and} \qquad C_\infty(h) = \sigma_H^2 \exp\left(-\frac{h^2}{2\ell^2} \right), \tag{22}$$

which correspond to the non-differentiable exponential and infinite-differentiable squared exponential (sometimes called Gaussian or radial basis function) covariance kernels, respectively.

Covariance kernels in 2D

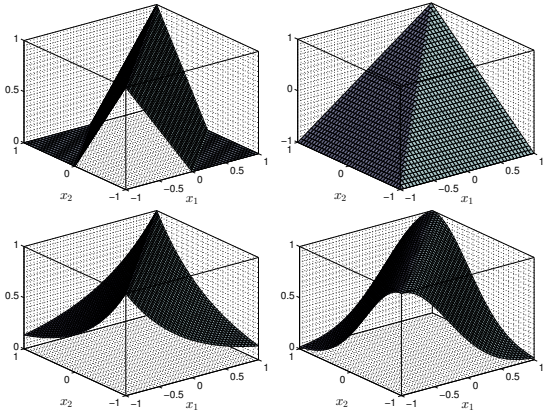


Figure: Examples of kernels: triangular (top left), Wiener (top right), exponential (bottom left), Gaussian (bottom right). Here, $\sigma_H^2 = 1$.

Matérn covariance kernel in 2D

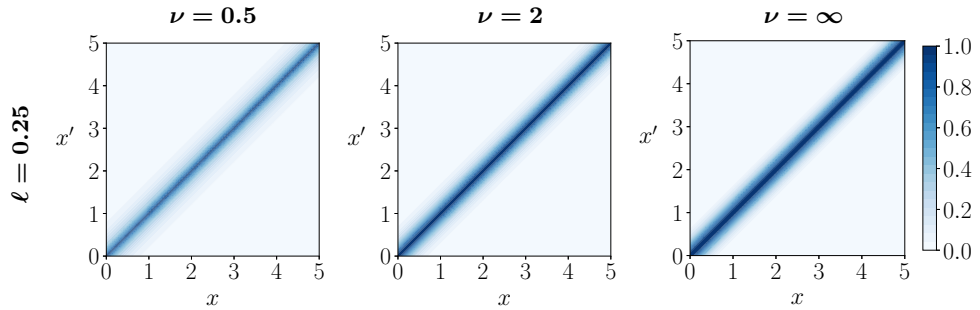


Figure: Correlation matrix associated to the Matérn kernel for different smoothing parameters (cols) and **small** $\ell = 0.25$ ($D = [0, 5]$).

Matérn covariance kernel in 2D

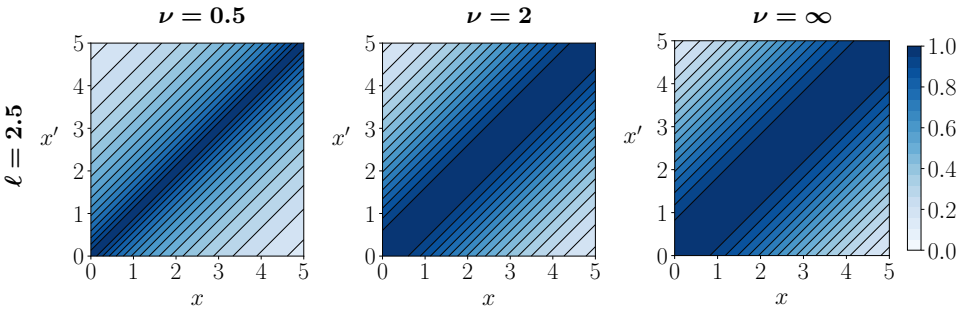


Figure: Correlation matrix associated to the Matérn kernel for different smoothing parameters (cols) and **medium** $\ell = 2.5$ ($D = [0, 5]$).

Matérn covariance kernel in 2D

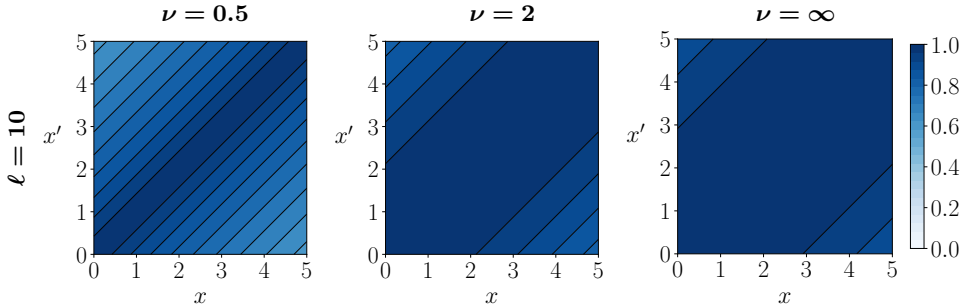


Figure: Correlation matrix associated to the Matérn kernel for different smoothing parameters (cols) and **large** $\ell = 10$ ($D = [0, 5]$).

PART IV: Random field representations

Random field discretization

Methods for random field representation can be classified into three main categories:

(i) Point discretization methods, in which the random field is represented point-wise at each spatial location of the discretization; the standard algorithm is the midpoint method.

(ii) Averaging discretization methods, in which the random field is expressed as weighted integrals over the domain. Some approaches are, the spatial averaging method, the shape function method and the weighted integral method.

(iii) Series expansion methods, where the random field is represented as a truncated series expansion of random variables and deterministic functions. Popular techniques are the Karhunen–Loève expansion, spectral representation, and Bayesian neural networks.

Random field discretization

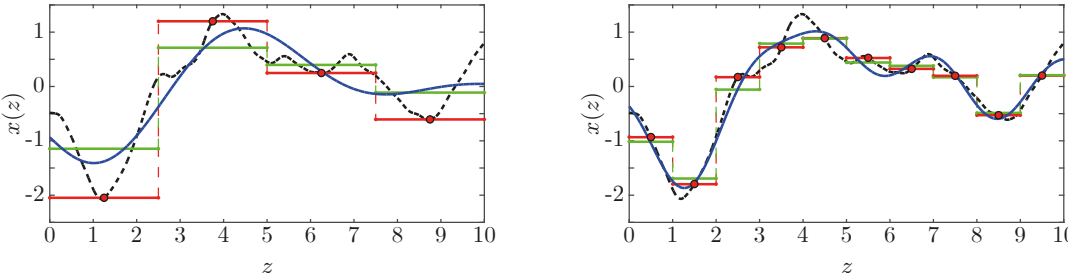


Figure: Comparison of random field representation methods: midpoint, spatial averaging, and Karhunen–Loève expansion. For $d = \{4, 10\}$ discretization points (left and right). The dotted black line shows a "full" ($d \to \infty$) random field realization as a reference.

Point discretization I

- The random field is represented at specific points $\boldsymbol{x} \in D$. A common choice is to pick those points as centroids of the elements defining the discretization of the domain D.

- We set a finite partition of the domain D into domain elements $D^{(e)}$. The approximation is expressed as a single RV defined at the centroid $\boldsymbol{x}^{(c)} \in D^{(e)}$ of each element, i.e.,

$$\widetilde{H}(\boldsymbol{x}) = H(\boldsymbol{x}^{(c)}). \tag{23}$$

Point discretization I

- The random field is represented at specific points $\boldsymbol{x} \in D$. A common choice is to pick those points as centroids of the elements defining the discretization of the domain D.

- We set a finite partition of the domain D into domain elements $D^{(e)}$. The approximation is expressed as a single RV defined at the centroid $\boldsymbol{x}^{(e)} \in D^{(e)}$ of each element, i.e.,

$$\widetilde{H}(\boldsymbol{x}) = H(\boldsymbol{x}^{(e)}). \tag{25}$$

- This method is the simplest of all random field discretizations but it easily suffers from the curse of dimensionality as the dimension depends on the resolution of the domain discretization, which can be large in many applications.

- This method tends to over-represent the variability (uncertainty) of the random field.

Point discretization II

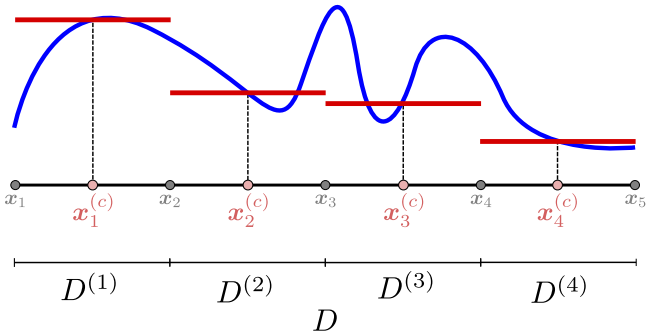


Figure: The midpoint method represents the random field as a collection of random variables indexed at the centroids of the domain discretization. Here, the "true" realization in blue is represented by the realization in red.

Karhunen–Loève discretization: intro

- The Karhunen–Loève (KL) expansion [9, 8] uses a linear combination of orthonormal functions chosen as the eigenfunctions resulting from the spectral decomposition of the covariance operator of the random field.
- The KL expansion is based on Mercer's theorem [11], which provides a series representation of symmetric positive-definite functions.
- The expansion is also known as Hotelling transform and it is closely related to principal component analysis (PCA) technique widely used in image processing and in data analysis in many fields.

Karhunen–Loève discretization: Mercer's theorem

- Let $D \subseteq \mathbb{R}^n$. We have seen that given a continuous kernel k, we can define a Hilbert–Schmidt operator. This operator is compact and has a complete set of eigenvectors in $L^2(D)$.

- In this case, there exists an orthonormal basis consisting of eigenfunctions $\{\phi_i\} : D \to L^2(D)$ of the operator together with a sequence of eigenvalues $\{\lambda_i\} \in [0, \infty)$, such that $\lambda_k \geq \lambda_{k+1}$ and $\lim_{k\to\infty} \lambda_k = 0$.

Karhunen–Loève discretization: Mercer's theorem

- Let $D \subseteq \mathbb{R}^n$. We have seen that given a continuous kernel k, we can define a Hilbert–Schmidt operator. This operator is compact and has a complete set of eigenvectors in $L^2(D)$.

- In this case, there exists an orthonormal basis consisting of eigenfunctions $\{\phi_i\} : D \to L^2(D)$ of the operator together with a sequence of eigenvalues $\{\lambda_i\} \in [0, \infty)$, such that $\lambda_k \geq \lambda_{k+1}$ and $\lim_{k\to\infty} \lambda_k = 0$.

- The covariance kernel $C_{HH}(\boldsymbol{x}, \boldsymbol{x}')$ has a representation of the form

$$C_{HH}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{\infty} \lambda_i \phi_i(\boldsymbol{x}) \phi_i(\boldsymbol{x}'), \qquad \forall\ \boldsymbol{x}, \boldsymbol{x}' \in D; \tag{24}$$

  this is called a Mercer expansion (only for symmetric positive-definite functions).

Karhunen–Loève discretization: definition

- Consider a square-integrable (second-order) random field $H(\boldsymbol{x}, \omega)$ defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed over D, and equipped with a continuous covariance function $C_{HH}(\boldsymbol{x}, \boldsymbol{x}')$.

Karhunen–Loève discretization: definition

- Consider a square-integrable (second-order) random field $H(\boldsymbol{x}, \omega)$ defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed over D, and equipped with a continuous covariance function $C_{HH}(\boldsymbol{x}, \boldsymbol{x}')$.

- We can employ an orthonormal basis consisting of the eigenfunctions of the covariance operator together with the sequence of real and positive eigenvalues to represent $H(\boldsymbol{x}, \omega)$:

$$H(\boldsymbol{x}, \omega) \approx \widehat{H}(k, \boldsymbol{x}, \omega) := \mu_H(\boldsymbol{x}) + \sum_{i=1}^{\infty} \mathbb{1}(i \leq k) \sqrt{\lambda_i} \phi_i(\boldsymbol{x}) \theta_i(\omega), \tag{25}$$

  where $\mathbb{1}(\cdot)$ denotes the indicator function, k is the truncation order of the expansion, and $\boldsymbol{\theta} = \{\theta_i(\omega) : \Omega \to \mathbb{R}\}$ is a set of mutually uncorrelated random variables with mean zero and unit variance.

- The series in eq. (25) is called the Karhunen–Loève (KL) expansion of the random field.

Karhunen–Loève: comments

- Note that the KL expansion works as a dimension reduction technique: instead of directly using a countable set of spatial points (as $d \to \infty$), we are now defining the uncertainty in terms of the KL coefficients $\boldsymbol{\theta} \subseteq \mathbb{R}^k$, with $k \ll d$.

Karhunen–Loève: comments

- Note that the KL expansion works as a dimension reduction technique: instead of directly using a countable set of spatial points (as $d \to \infty$), we are now defining the uncertainty in terms of the KL coefficients $\boldsymbol{\theta} \subseteq \mathbb{R}^k$, with $k \ll d$.

- For Gaussian random fields, the variables $\theta_i(\omega)$ are iid Gaussian. In general, the distribution of $\theta_i(\omega)$ is cumbersome to estimate since

$$\theta_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D H(\boldsymbol{x}, \omega) \phi_i(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \qquad i = 1, 2, \ldots, \tag{27}$$

  which requires the knowledge of random field structure for its computation.

- Non-Gaussian random fields can be represented with the KL expansion through the definition of different basis functions, or the application of a suitable isoprobabilistic transformation.

Karhunen–Loève: properties

The KL expansion is usually preferred over other alternatives due to the following properties:

1. **Error minimizing property**: the KL approximation minimizes the mean-squared error (MSE)

$$\mathbb{E}\left[\left\|[H(\boldsymbol{x},\omega)-\mu_H(\boldsymbol{x})]-[\widehat{H}(\boldsymbol{x};k,\boldsymbol{\theta})-\mu_H(\boldsymbol{x})]\right\|_2^2\right]=\sum_{i=k+1}^{\infty}\lambda_i^2. \tag{28}$$

2. **Uniqueness of the expansion**: the random variables $\boldsymbol{\theta}=\{\theta_i(\omega):\Omega\to\mathbb{R}\}$ in the expansion are orthonormal, if and only if the orthonormal functions $\{\phi_i\}$ and the constants $\{\lambda_i\}$ are the eigenpairs of the covariance matrix of the field.

Moreover, we can compute the variance of the KL approximation as

$$\mathbb{V}\left[\widehat{H}(\boldsymbol{x};k,\boldsymbol{\theta})\right]=\sum_{i=1}^{k}\lambda_i\phi_i^2(\boldsymbol{x}). \tag{29}$$

Karhunen–Loève: error metrics I

- In the KL expansion, the number of terms to be included in the series is closely related to the magnitudes of the eigenvalues, which in turn strongly depend on the correlation length and smoothness of the covariance function.

- One can define local point-wise errors for the mean and variance in terms of the relative difference between the exact and approximated random fields:

$$\epsilon_{\mu}(\boldsymbol{x}) = \left| \frac{\mathbb{E}[H(\boldsymbol{x},\omega)] - \mathbb{E}\left[\widehat{H}(\boldsymbol{x};k,\boldsymbol{\theta})\right]}{\mathbb{E}[H(\boldsymbol{x},\omega)]} \right| \qquad \epsilon_{\sigma^2}(\boldsymbol{x}) = \left| \frac{\mathbb{V}[H(\boldsymbol{x},\omega)] - \mathbb{V}\left[\widehat{H}(\boldsymbol{x};k,\boldsymbol{\theta})\right]}{\mathbb{V}[H(\boldsymbol{x},\omega)]} \right|,$$

  here ϵ_{μ} and ϵ_{σ^2} are the relative errors in the mean and variance, respectively.

Karhunen–Loève: error metrics II

- Global error metrics can also be applied to quantify the overall quality of the random field representation. These are defined for the mean and variance, as their average values over the index set D of the random field

$$\bar{\epsilon}_{\mu} = \frac{1}{|D|}\int_D \epsilon_{\mu}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \qquad \text{and} \qquad \bar{\epsilon}_{\sigma^2} = \frac{1}{|D|}\int_D \epsilon_{\sigma^2}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{30}$$

  where $|D| = \int_D \mathrm{d}\boldsymbol{x}$.

- Under eq. (29), the global variance error reduces to

$$\bar{\epsilon}_{\sigma^2} = 1 - \frac{1}{|D|\,\sigma_H^2}\sum_{i=1}^{k}\lambda_i; \tag{31}$$

  this expression is commonly used to define the truncation order of the KL expansion.

Karhunen–Loève: finding the eigenpairs

- The set of eigenpairs $\{\lambda_i, \phi_i\}$ are obtained through the solution of an homogeneous Fredholm integral equation of the second kind [5]

$$\int_D C_{HH}(\boldsymbol{x}, \boldsymbol{x}')\phi_i(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' = \lambda_i \phi_i(\boldsymbol{x}), \tag{32}$$

  whose analytical solution exist only for specific cases of covariance functions.

Karhunen–Loève: finding the eigenpairs

- The set of eigenpairs $\{\lambda_i, \phi_i\}$ are obtained through the solution of an homogeneous Fredholm integral equation of the second kind [5]

$$\int_D C_{HH}(\boldsymbol{x}, \boldsymbol{x}')\phi_i(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' = \lambda_i \phi_i(\boldsymbol{x}), \tag{34}$$

  whose analytical solution exist only for specific cases of covariance functions.

- Several methodologies have been proposed to solve the eigenvalue problem:
  - *Projection methods*, which aim to minimize an error (e.g., collocation, Galerkin).
  - *Nyström methods*, which replace the integral with a representative weighted sum.
  - *Degenerate kernel methods*, which consist of approximating a target kernel $k(x, y)$ with a degenerate or separable kernel $k_n(x, y) = \sum_{i=1}^{n} a_i(x)b_i(y)$.

Karhunen–Loève: two analytical solutions

- If the correlation kernel is Wiener (eq. (16)), the eigenvalues and eigenfunctions are given by:

$$\phi_k(x) = \sqrt{2}\sin\left(\left(k-\tfrac{1}{2}\right)\pi x\right), \qquad \lambda_k = \frac{1}{(k-\frac{1}{2})^2\pi^2}. \tag{33}$$

- If the correlation kernel is exponential (eq. (22)), the eigenvalues and eigenfunctions are given by:

$$\phi_k(x) = \begin{cases} \dfrac{\cos(\omega_k x)}{\left(a + \frac{\sin(2\omega_k a)}{2\omega_k}\right)^{1/2}} & k \text{ odd} \\ \dfrac{\sin(\omega_k x)}{\left(a - \frac{\sin(2\omega_k a)}{2\omega_k}\right)^{1/2}} & k \text{ even} \end{cases}, \qquad \lambda_k = \frac{2}{\ell(\omega_k^2 + 1/\ell^2)}, \tag{34}$$

where the ω_k are found from nonlinear equations for k odd and even, respectively [5].

Karhunen–Loève: Nyström method I

- The Nyström method requires the application of some quadrature rule to solve integral equations as,

$$\int_a^b g(x)\,\mathrm{d}x \approx \sum_{j=1}^{N_{\mathrm{GP}}} w_j\, g(\xi_j), \tag{35}$$

where $\{w_j\}_{j=1}^{N_{\mathrm{GP}}}$ are the weights of quadrature rule, and the points $\{\xi_j\}_{j=1}^{N_{\mathrm{GP}}}$ are the associated abscissas of the quadrature polynomial. Gauss–Legendre quadrature rules are the most popular.

- After applying this formulation to (32), the following is obtained:

$$\sum_{j=1}^{N_{\mathrm{GP}}} w_j C_{HH}(\xi_i,\xi_j)\phi_k(\xi_j) = \lambda_k\phi_k(\xi_i), \qquad i = 1,\ldots,N_{\mathrm{GP}}. \tag{36}$$

Karhunen–Loève: Nyström method II

- The above can be expressed in the form of a matrix eigenvalue problem as,

$$\widetilde{\mathbf{C}}\boldsymbol{f} = \boldsymbol{\Lambda}\boldsymbol{f} \tag{37}$$

  where, $\widetilde{\mathbf{C}}$ is a $N_{\text{GP}} \times N_{\text{GP}}$ matrix, with its components defined as, $\widetilde{C}_{ij} = C_{HH}(\xi_i, \xi_j)\, w_j$.

- For computational reasons, it is desirable that the eigenvalue problem involves symmetric matrices. The covariance kernel is symmetric, but due to the fact that the integration weights w_j are not equal in most quadratures, the symmetry of the matrix $\widetilde{\mathbf{C}}$ can be affected.

- The symmetry condition can be restored by considering the diagonal matrix $\mathbf{W} = \text{diag}(\boldsymbol{w})$, and its square root $\mathbf{W}^{1/2} = \text{diag}(\sqrt{\boldsymbol{w}})$, such that:

$$\mathbf{C}\mathbf{W}\boldsymbol{f} = \boldsymbol{\Lambda}\boldsymbol{f} \tag{38a}$$

$$(\mathbf{W}^{1/2}\mathbf{C}\mathbf{W}^{1/2})\boldsymbol{h} = \boldsymbol{\Lambda}\boldsymbol{h}, \tag{38b}$$

  where $\boldsymbol{h} = \mathbf{W}^{1/2}\boldsymbol{f}$. This is now in the form of a symmetric eigenvalue problem.

Karhunen–Loève: Nyström method III

- The symmetric eigenvalue problem in (38b) is solved for $\boldsymbol{h}$, with components $h_{ij} = \sqrt{w_j} f_{ij}$, and the diagonal matrix $\boldsymbol{\Lambda}$ with the eigenvalues as its elements $\Lambda_{ij} = \delta_{ij}\lambda_i$.

- After that, the i-th eigenvector is obtained as the i-th column of $\boldsymbol{f} = \mathbf{W}^{-1/2}\boldsymbol{h}$.

- Finally, the eigenfunctions can be obtained by the so-called Nyström interpolation formula:

$$\phi_k(\boldsymbol{x}) = \frac{1}{\lambda_n} \sum_{j=1}^{N_{\text{GP}}} w_j f_{jn} C_{HH}(x, \xi_j) \qquad \forall\ x \in D, \quad n = 1, \ldots, N_{\text{GP}}. \tag{39}$$

Karhunen–Loève: example 1D

- We model the flexibility $F(x)$ of a cantilever beam. The beam has length $L = 5$ m (i.e., $D = [0, 5]$).
- The flexibility is described by a homogeneous Gaussian random field. The Matérn class is considered as covariance function. The mean of the field is $\mu_F = 1 \times 10^{-4}$ and the standard deviation is $\sigma_F = 0.35 \cdot \mu_F$.
- A parameter study on the correlation length ℓ and smoothing ν is performed.

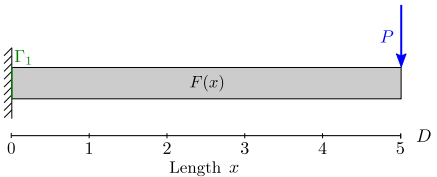


Figure: Cantilever beam configuration.

Karhunen–Loève: example 1D; variance error

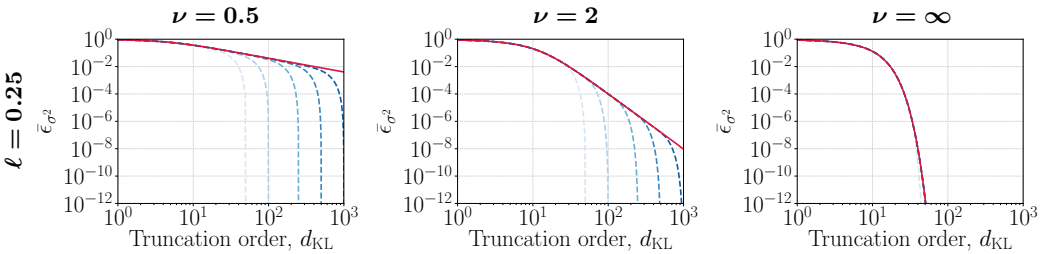


Figure: Evolution of the global variance error with the KL terms $d_{\mathrm{KL}} = k$. For different number of GL points in the Nyström method ($N_{\mathsf{GP}} = \{50, 100, 250, 500, 1000\}$); shown with dashed lines and increasing blue color. For different smoothing parameters (cols) and **small** $\ell = 0.25$. The reference global variance error is shown in red.

Karhunen–Loève: example 1D; variance error

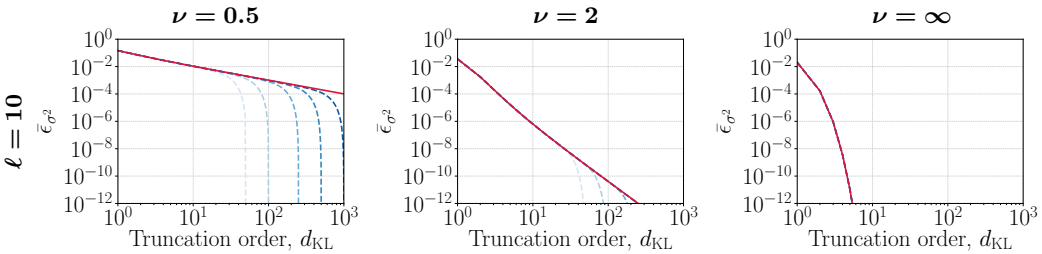


Figure: Evolution of the global variance error with the KL terms $d_{\mathrm{KL}} = k$. For different number of GL points in the Nyström method ($N_{\mathsf{GP}} = \{50, 100, 250, 500, 1000\}$); shown with dashed lines and increasing blue color. For different smoothing parameters (cols) and **large** $\ell = 10$. The reference global variance error is shown in red.

Karhunen–Loève: example 1D; realizations

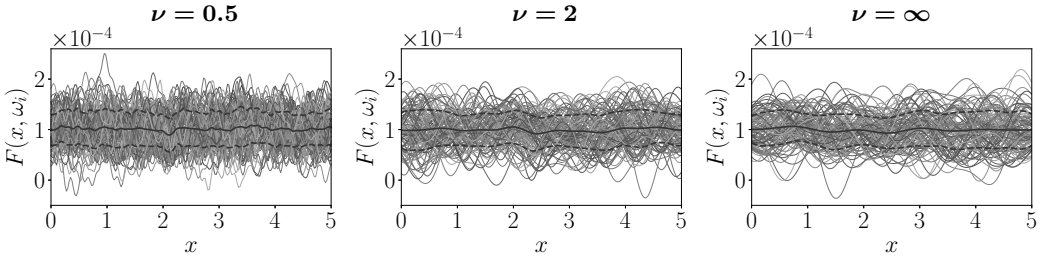


Figure: Flexibility random field realizations for different smoothing parameters of the Matérn kernel and **small** $\ell = 0.25$.

Karhunen–Loève: example 1D; realizations

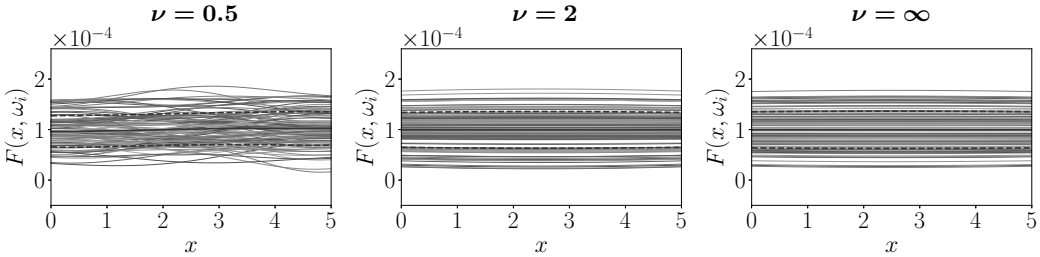
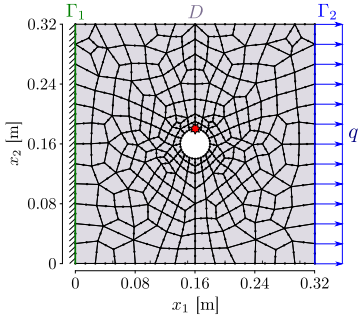


Figure: Flexibility random field realizations for different smoothing parameters of the Matérn kernel and **large** $\ell = 10$.

Karhunen–Loève: example 2D

- We consider a steel plate defined by a square domain $D \subseteq \mathbb{R}^2$ with length 0.32, thickness $t = 0.01$, and a hole of radius 0.02 located at its center.
- The Young's modulus $E(\boldsymbol{x})$ is assumed to be random and spatially variable. A lognormal random field with mean $\mu_E = 2 \times 10^5$ and standard deviation $\sigma_E = 3 \times 10^4$ is used.



Karhunen–Loève: example 2D

- The truncated KL expansion of the Young's modulus can be written as

$$\widehat{E}(\boldsymbol{x}, \boldsymbol{\theta}) := \exp\left[\mu_{E'} + \sum_{i=1}^{k} \sqrt{\lambda_i}\phi_i(\boldsymbol{x})\theta_i\right], \tag{40}$$

  where the underlying Gaussian parameters $\mu_{E'} = 26.011$ and $\sigma_{E'} = 0.149$, are computed from the mean and standard deviation of the lognormal random field.

- We select the Matérn kernel to model the covariance function of the underlying Gaussian field. The smoothing parameters are chosen as $\nu = \{0.5, 2, \infty\}$ and we select the correlation lengths as $\ell = \{0.02, 0.16, 0.32\}$ m.

- The eigenpairs are estimated with the Nyström method using 110 GL points in each direction.

Karhunen–Loève: example 2D; eigenvalues

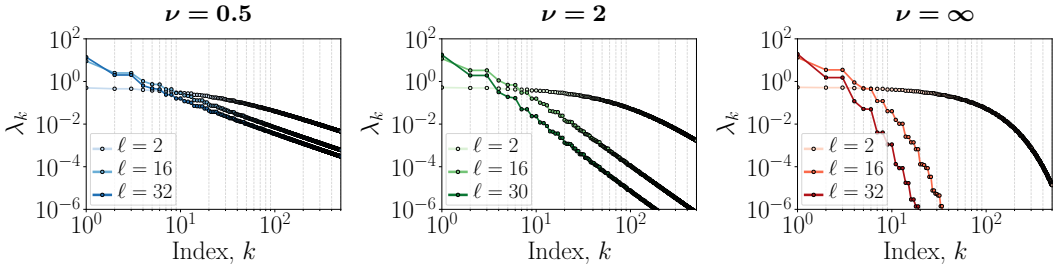
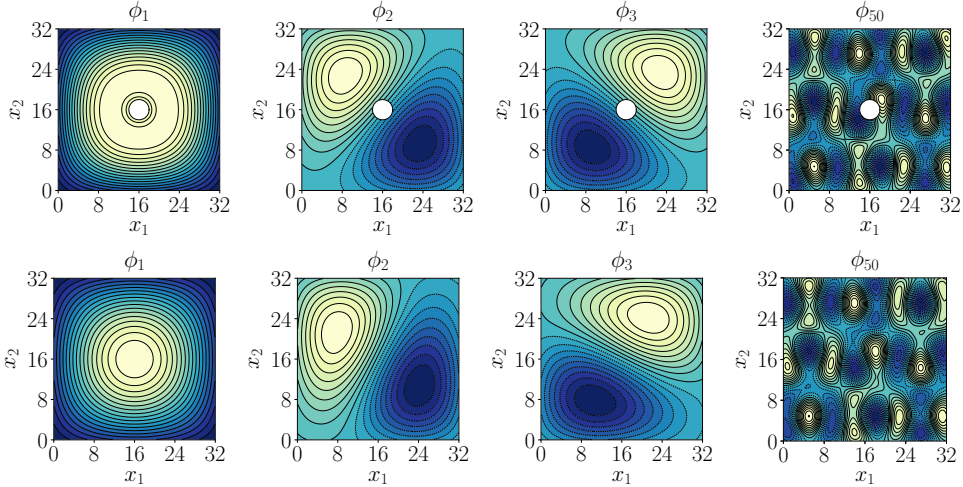


Figure: Evolution of the eigenvalues of the Matérn covariance operator with the truncation order of the KL expansion, for different smoothing parameters and correlation lengths.

Karhunen–Loève: example 2D; eigenfunctions for $\nu = 0.5$ and $\ell = 2$



Neural networks

- **Neural networks** (NNs) are commonly used in practice as function approximators.
- This is primarily attributed to their flexibility of having many model parameters (weights and biases) which can be learned from data.
- In statistical terms, NNs are non-parametric models, a term meant to contrast them with *parametric* models in which the relationship is characterized in terms of a few parameters, which often have meaningful interpretations.
- **Bayesian neural network** (BNNs) combines neural network with Bayesian inference. Bayesian inference allows us to learn a probability distribution over possible neural networks[5].

[5] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, 1996.

Feed-forward network functions (I)

- Feed-forward network functions are also known as multilayer perceptron networks. These networks take in a set of real inputs, x_i, and from them compute one or more output values, $u_j(x)$, using some number of layers of hidden units.

- NNs use a 'basis function' that is itself a **nonlinear function of a linear combination of the inputs**, where the coefficients in the linear combination are adaptive parameters [3].

- This leads to the basic NN model, which can be described a series of functional transformations. The unknown function can be represented by a K-layer NN on D:

$$u(x) \approx \Psi^{(\mathsf{NN})}(x;\boldsymbol{\theta}) := f\left(\sum_{i=1}^{K} \theta_i \Phi_i(x)\right), \tag{41}$$

  where f is a nonlinear activation function, Φ_i are nonlinear basis functions, and the coefficients $\boldsymbol{\theta}$. These are adjusted during the *estimation*.

Feed-forward network functions (II)

- We consider a fully-connected feed-forward network function composed of $L \geq 1$ layers where each layer has n_ℓ nodes. If $x \in \mathbb{R}$ is a scalar input argument, we can explicitly define the network as:

$$f_i^{(1)}(x) = w_{i,1}^{(1)}\, x + b_i^{(1)}, \qquad h_i^{(1)}(x) = \varphi\big(f_i^{(1)}(x)\big) \qquad 1 \leq i \leq n_1 \tag{42a}$$

$$f_i^{(\ell)}(x) = \frac{1}{\sqrt{n_{\ell-1}}} \sum_{j=1}^{n_{\ell-1}} w_{i,j}^{(\ell)}\, h_j^{(\ell-1)}(x) + b_i^{(\ell)}, \qquad 2 \leq \ell \leq L, \quad 1 \leq i \leq n_\ell, \tag{42b}$$

  where $b_i^{(\ell)}$ and $w_{i,j}^{(\ell)}$ are the biases and weights at the ℓ-th layer.

- For every input x, the NN defines the function representation:

$$u(x) \approx \Psi^{(\mathsf{NN})}(x;\boldsymbol{\theta}) := \frac{1}{\sqrt{n_L}} \sum_{j=1}^{n_L} w_{1,j}^{(L+1)}\, h_j^{(L)}(x) + b^{(L+1)}. \tag{43}$$

Feed-forward network functions (III)

- In matrix form we can just simply define the operations as follows:

$$\boldsymbol{f}^{(1)}(x) = \boldsymbol{W}^{(1)}\, x + \boldsymbol{b}^{(1)} \tag{44a}$$

$$\boldsymbol{f}^{(\ell)}(x) = \frac{1}{\sqrt{n_{\ell-1}}} \boldsymbol{W}^{(\ell)}\, \varphi\left(\boldsymbol{f}^{(\ell-1)}(x)\right) + \boldsymbol{b}^{(\ell)}, \qquad 2 \leq \ell \leq L, \tag{44b}$$

  where $\boldsymbol{W}^{(\ell)}$ and $\boldsymbol{b}^{(\ell)}$ are weight matrices and biases vectors.

- The scaling factor highlighted in red is necessary to have a well-defined limit of the prior BNN. This scaling assumes that the weights and biases are endowed with a Gaussian prior.

- We use an hyperbolic tangent activation function (it is important that it is a **bounded function**): $\varphi(t) = \tanh(t)$. At the output, the activation is typically linear.

Feed-forward network functions (IV)

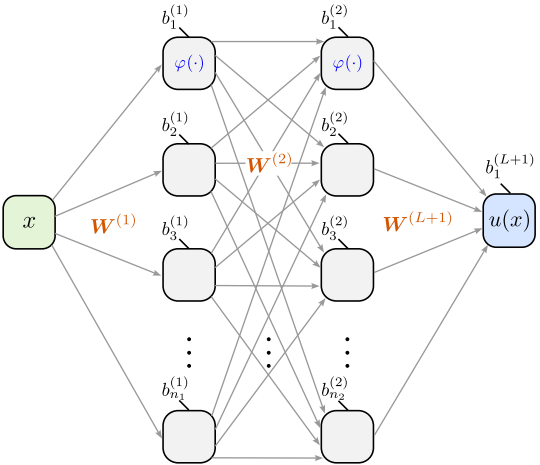


Figure: Classical structure of the (fully-connected) feed-forward network function with 2 layers.

Feed-forward network functions (V)

- In the *network forward map* $u(x) = \Psi(x; \boldsymbol{\theta})$, the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$ contains all the weights and biases of the NN.
- The gradient of the NN wrt the parameters $\boldsymbol{\theta} = \{\boldsymbol{W}, \boldsymbol{b}\}$ is computed via the **back-propagation algorithm**. This is typically needed to perform the *training* in the context of *LPs*, or to use advanced MCMC algorithms in the context of *IPs*.
- Back-propagation is an efficient application of the chain rule to NNs. It is also known as the reverse mode of automatic differentiation. In our example, we have to compute something like this:

$$\frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[\frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{W}^{(3)}}, \frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{W}^{(2)}}, \frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{W}^{(1)}}, \frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{b}^{(3)}}, \frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{b}^{(2)}}, \frac{\partial u(x;\boldsymbol{\theta})}{\partial \boldsymbol{b}^{(1)}} \right]. \tag{45}$$

Gaussian BNNs: the single-layer infinite-dimensional limit (I)

- We assume that (under the prior) the weights and biases in the network are independent and identically distributed according to a zero-mean Gaussian. That is:

$$\text{at the input:} \qquad w_{i,j} \sim \mathcal{N}(0, \sigma_w^2), \quad b_j \sim \mathcal{N}(0, \sigma_b^2) \tag{46a}$$

$$\text{at the output:} \qquad \bar{w}_{j,k} \sim \mathcal{N}(0, \bar{\sigma}_w^2), \quad \bar{b}_k \sim \mathcal{N}(0, \bar{\sigma}_b^2) \tag{46b}$$

- The output of the NN is the sum of a bias and the weighted contributions of the n_ℓ hidden units:

$$u_k(x) = \bar{b}_k + \sum_{j=1}^{n_\ell} \bar{w}_{j,k} h_j(x), \qquad h_j(x) = \varphi\left(b_j + \sum_{i=1}^{n} \bar{w}_{i,j} x_j\right) \tag{47}$$

- Let us analyze the expectation of the output:

$$\mathbb{E}\left[\bar{b}_k + \bar{w}_{j,k} h_j(x)\right] = \cancelto{0}{\mathbb{E}\left[\bar{b}_k\right]} + \cancelto{0}{\mathbb{E}\left[\bar{w}_{j,k}\right]} \mathbb{E}\left[h_j(x)\right] = 0. \tag{48}$$

Gaussian BNNs: the single-layer infinite-dimensional limit (II)

- Now, let us analyze the variance of the output:

$$\mathbb{V}\left[\bar{b}_k + \bar{w}_{j,k} h_j(x)\right] = \mathbb{V}\left[\bar{b}_k\right] + \mathbb{V}[\bar{w}_{j,k} h_j(x)] \tag{49a}$$

$$= \bar{\sigma}_b^2 + \mathbb{E}\left[(\bar{w}_{j,k} h_j(x))^2\right] = \bar{\sigma}_b^2 + \mathbb{E}\left[\bar{w}_{j,k}^2\right] \mathbb{E}\left[h_j(x)^2\right] \tag{49b}$$

$$= \bar{\sigma}_b^2 + \bar{\sigma}_w^2 \underbrace{\mathbb{E}\left[h_j(x)^2\right]}_{h_j \text{ has to be bounded.}} \tag{49c}$$

denote $V(x) = \mathbb{E}\left[h_j(x)^2\right]$.

- By taking into account the sum as $n_\ell \to \infty$, we can use the central limit theorem to see what happens with the output variance:

$$\bar{\sigma}_b^2 + \sum_{j=1}^{n_\ell} \bar{w}_{j,k} h_j(x) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \bar{\sigma}_b^2 + n_\ell \bar{\sigma}_w^2 V(x)). \tag{50}$$

Gaussian BNNs: the single-layer infinite-dimensional limit (III)

- Accordingly, to obtain a well-defined limit for the prior distribution of the value of the function at any particular point, we need only scale the **prior standard deviation** of the hidden-to-output weights according to the number of hidden units:

$$\frac{1}{\sqrt{n_\ell}}\bar{\sigma}_w. \tag{51}$$

- The joint distribution for the values of all the outputs of the network for some selection of values for inputs will also become a multivariate Gaussian in the limit as the number of hidden units goes to infinity.

- Similar result exist for the convergence under more general type of priors, e.g., α-stable distributions.

- Moreover, infinite-dimensional limit of deep BNNs (i.e., as $L \to \infty$) has also been analyzed. This might be mentioned later in the course.

Gaussian BNNs: the single-layer infinite-dimensional limit (IV)

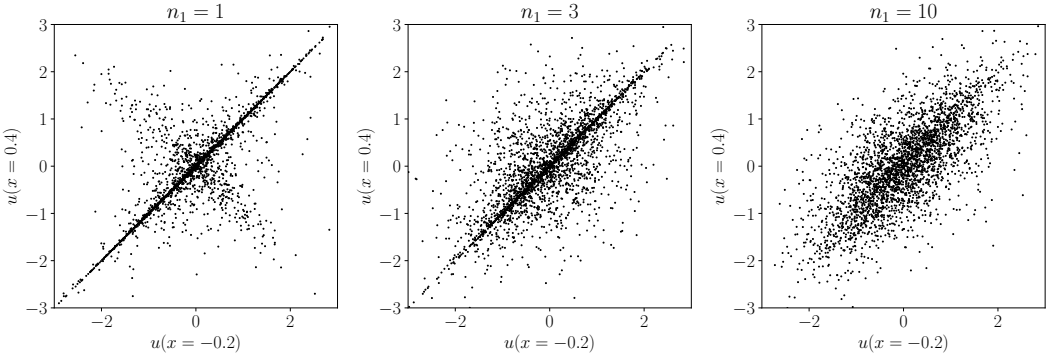


Figure: Convergence of the BNN prior to a Gaussian process as the nodes increases. Each of the plots is based on 3000 NN with a two-input and one-output units. In the Gaussian priors we use: in the input layer $\sigma_w = \sigma_b = 5$, and in the output layer $\bar{\sigma}_w = 1$,$\bar{\sigma}_b = 0.1$.

Gaussian BNNs: realizations

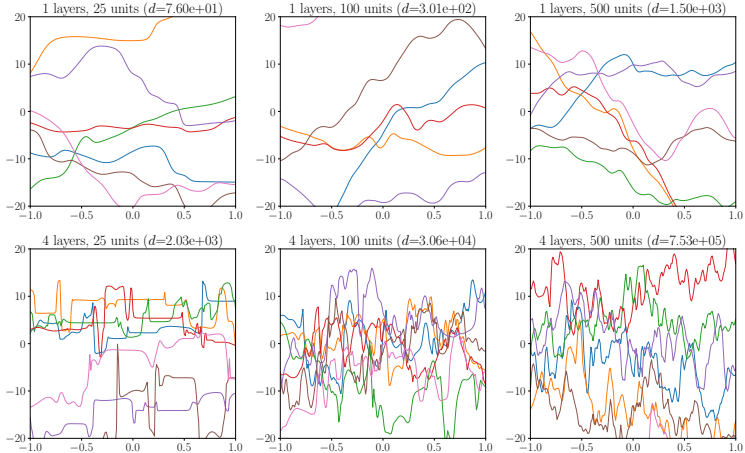


Figure: Realizations $u(x)$ from the Gaussian BNN for different number of layers and nodes.

Final comments

- The uncertainty related to spatially varying properties is generally represented by random fields which are understood as **random variables that take values in a function space**.
- While the KL expansion is simpler to manipulate theoretically, it can only be applied (in general) to Gaussian random fields, and selecting the covariance kernel type and its hyper-parameters is necessary.
- Conversely, the NN approximation provides us with more flexibility as we can easily change the prior of the parameters. However, the parameter space is typically very high-dimensional and we require the estimation of variance hyperparameters associated with the prior.
- We will see that to perform Bayesian inference, we require MCMC methods that operate in the function space, or that add gradient information.

References

[1] P. Abrahamsen. *A review of Gaussian random fields and correlation functions*. 2nd ed. Norwegian Computing Center, 1997.

[2] R. J. Adler. *The geometry of random fields*. Society for Industrial and Applied Mathematics (SIAM), 2010.

[3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] G.-A. Fuglstad et al. "Constructing priors that penalize the complexity of Gaussian random fields". In: *Journal of the American Statistical Association* 114.525 (2019), pp. 445–452.

[5] R. G. Ghanem et al. *Stochastic finite elements: a spectral approach*. Revised edition. Dover Publications, 2012 (Original in 1991).

[6] M. S. Handcock et al. "A Bayesian analysis of Kriging". In: *Technometrics* 35 (4 1993).

[7] M. S. Handcock et al. "An approach to statistical spatial-temporal modeling of meteorological fields". In: *Journal of the American Statistical Association* 89 (426 1994).

[8] K. Karhunen. "Über lineare Methoden in der Wahrscheinlichkeitsrechnung". In: *Annales Academiæ Scientiarum Fennicæ. Series A.1, Mathematica-Physica* 37 (1947), pp. 1–79.

[9] M. Loève. "Fonctions aléatoires du second ordre". In: *La Revue Scientifique* 84 (1946), pp. 195–206.

[10] B. Matérn. *Spatial variation: stochastic models and their application to some problems in forest surveys and other sampling investigations*. Report No. 49/5. Forest Research Institute of Sweden, 1960.

[11] J. Mercer. "Functions of positive and negative type, and their connection the theory of integral equations". In: *Philosophical Transactions of the Royal Society of London. Series A: Containing Papers of a Mathematical or Physical Character* 209 (1909).

[12] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, 1996.

[13] C. E. Rasmussen et al. *Gaussian processes for machine learning*. The MIT Press, 2006.

[14] H. Rue et al. *Gaussian Markov Random Fields. Theory and Applications*. Chapman & Hall/CRC, 2005.

[15] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer, 1999.

[16] P. Whittle. "On stationary processes in the plane". In: *Biometrika* 41 (3-4 1954).

[17] B. Øksendal. *Stochastic differential equations: an introduction with applications*. 6th ed. Springer, 2003.