Resampling Methods for Detecting Anisotropic Correlation Structure

Assaf Rabinowicz

Joint work with Prof. Saharon Rosset

Course: Bootstrap and Resampling Methods

by Prof. Saharon Rosset

Department of Statistics and Operations Research

Tel-Aviv University

May 24, 2021

Outline

- Gaussian process regression
- Isotropy
- Resampling methods for detecting anisotropic correlation structure

Generalized Least Squares (GLS)

Linear regression for correlated data:

$$\boldsymbol{z} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{z} \in \mathbb{R}^n,\ X \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\epsilon} \sim N(0, V)$. Given V, the MLE of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}} = (X^t V^{-1} X)^{-1} X^t V^{-1} \boldsymbol{z}.$$

Generalized Least Squares (GLS)

Linear regression for correlated data:

$$\boldsymbol{z} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{z} \in \mathbb{R}^{n}$, $X \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\epsilon} \sim N(0, V)$. Given V, the MLE of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}} = (X^{t}V^{-1}X)^{-1}X^{t}V^{-1}\boldsymbol{z}.$$

Once we estimate $\boldsymbol{\beta}$ we can use it for predicting new observations, $\boldsymbol{z}^{*} \in \mathbb{R}^{n^{*}}$ at new covariate points $X^{*} \in \mathbb{R}^{n\times p}$, i.e., $\widehat{\boldsymbol{z}^{*}} := \widehat{\mathbb{E}}(\boldsymbol{z}^{*}) := X^{*}\widehat{\boldsymbol{\beta}}$
$\mathbb{E}_{\boldsymbol{z}}\widehat{\boldsymbol{z}^{*}} = X^{*}(X^{t}V^{-1}X)^{-1}X^{t}V^{-1}X\boldsymbol{\beta} = X^{*}\boldsymbol{\beta}.$

Generalized Least Squares (GLS)

- GLS is reduced to LS solution when $V = \sigma^2 I_n$.
- V^{-1} contains the observations weights. Similar (correlated) observations are credited correspondingly. From a broader perspective, in terms of prediction, the efficient sample size is smaller and affects $\mathrm{Var}(\widehat{\beta})$.
- Mostly, V is unknown and therefore is estimated. Inaccurate estimation affects $\mathrm{Var}(\widehat{\beta})$:

$$\mathrm{Var}(\widehat{\beta}) = (X^t \widehat{V}^{-1} X)^{-1} X^t \widehat{V}^{-1} V \widehat{V}^{-1} X (X^t \widehat{V}^{-1} X)^{-1}.$$

When $\widehat{V} = V$:

$$\mathrm{Var}(\widehat{\beta}) = (X^t V^{-1} X)^{-1} = I^{-1}(\widehat{\beta}).$$

BLUP

Assuming normality of $\boldsymbol{z}, \boldsymbol{z}^*$, then:

$$\boldsymbol{z}^*|\boldsymbol{z} \sim N(\mu_{\boldsymbol{z}^*|\boldsymbol{z}}, \Sigma_{\boldsymbol{z}^*|\boldsymbol{z}}),$$

where

$$\mu_{\boldsymbol{z}^*|\boldsymbol{z}} = \mathbb{E}(\boldsymbol{z}^*) + \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z})\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z})^{-1}(\boldsymbol{z} - \mathbb{E}(\boldsymbol{z}))$$
$$\Sigma_{\boldsymbol{z}^*|\boldsymbol{z}} = \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z}^*) - \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z})\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z})^{-1}\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z}^*)$$

The best linear unbiased predictor of $\boldsymbol{z}$ given $\boldsymbol{z}^*$ and the covariance matrices is:

$$\widehat{\mathbb{E}}(\boldsymbol{z}^*|\boldsymbol{z}) = X^*\widehat{\beta} + \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z})\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z})^{-1}(\boldsymbol{z} - X\widehat{\beta}).$$

BLUP

$$\widehat{\mathbb{E}}(\boldsymbol{z}^*|\boldsymbol{z}) = X^*\widehat{\boldsymbol{\beta}} + \mathrm{Cov}(\boldsymbol{z}^*,\boldsymbol{z})\mathrm{Cov}(\boldsymbol{z},\boldsymbol{z})^{-1}(\boldsymbol{z} - X\widehat{\boldsymbol{\beta}})$$

- $\mathbb{E}(\boldsymbol{z}^*|\boldsymbol{z})$ is also the solution of the following problem:

$$\underset{a\in\mathbb{R}^{n^*},\ B\in\mathbb{R}^{n^*\times n}}{\mathrm{argmin}} \quad \mathbb{E}_{\boldsymbol{z},\boldsymbol{z}^*}\|\boldsymbol{z}^* - (\mathbf{a} + B\boldsymbol{z})\|_2^2$$

(so the normality assumption can be avoided).

BLUP - applications

$$X^*\widehat{\beta} + \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z})\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z})^{-1}(\boldsymbol{z} - X\widehat{\beta})$$

Clustered data: $\{\boldsymbol{z}, X\}$ were sampled from two clusters. $\{\boldsymbol{z}^*, X^*\}$ are sampled from the same clusters.

$$\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z}) = \begin{bmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{bmatrix}, \ \mathrm{Cov}(\boldsymbol{z}^*, \boldsymbol{z}) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Exmples:

- Grades in school classes.
- Disease progression (repeated measures). Combines two correlation factors: patient (cluster) and a temporal vector.

BLUP - applications

Spatial data: Here the correlation depends on the distance between the observations, therefore the covariance matrix should be related to a continuous function of the distance.

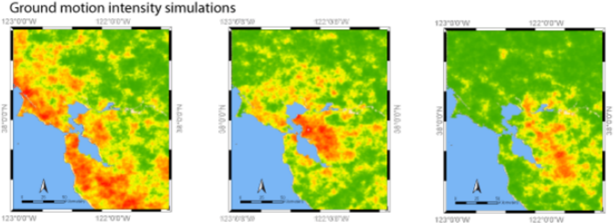


Figure: Taken from Jack Baker Research Group's website

This leads to introduce *random field*...

Random Field - Definition

Random field (spatial stochastic process) is the following instance

$$z := z(\boldsymbol{s}),\ \boldsymbol{s} \in \mathbb{S},$$

where $\mathbb{S}$ is a spatial uncountable space, e.g., $\mathbb{R}^2$, a manifold such as a tunnel network, a product space $\mathbb{R} \times \mathbb{Z}$.

- A random function where its index space is **continuous**.
- It is a generalization of the standard representation of stochastic process:

  $$z := z(t),\ t \in T,$$

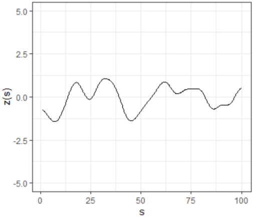
  where T is the continuous time axis. (which is a generalization of time series, where T is a countable set of time points).

There are many types of random field, we focus on Gaussian random field

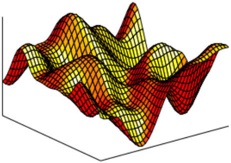
Random Field - Definition

There are many types of random field, we focus on Gaussian random field (Gaussian process): i.e., for any finite set $\{\boldsymbol{s}_i \in \mathbb{S}\}_{i=1}^{n}$:

$$\boldsymbol{z} := \{z(\boldsymbol{s}_i)\}_{i=1}^{n} \text{ is distributed MVN.}$$



(a) $\mathbb{S} \subset \mathbb{R}$



(b) One sample from $\mathbb{S} \subset \mathbb{R}^2$.

Taken from Rasmussen (2003)

Random Field - Definition

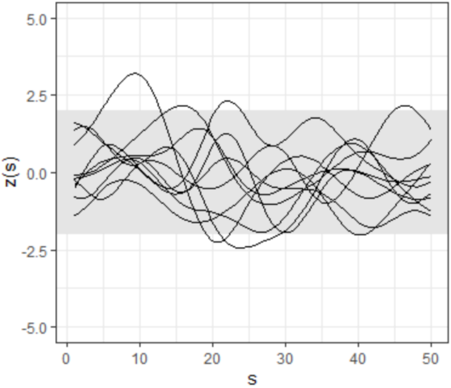


Figure: $\mathbb{S} \subset \mathbb{R}$, multiple samples

Gaussian Process Regression

Implementing regression for spatial data uses the random field concept.

For example the BLUP formula is now:

$$X^*\widehat{\beta} + \mathcal{K}(S^*, S)(\mathcal{K}(S, S) + \sigma^2 I_n)^{-1}(\boldsymbol{z} - X\widehat{\beta}),$$

where $\mathcal{K}(\cdot,\cdot)$ is the kernel covariance function. $S \in \mathbb{R}^{n\times 2}, S^* \in \mathbb{R}^{n^*\times 2}$ are the coordinates values of Z, Z^*.

General properties:

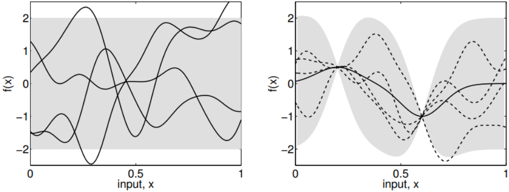
- The correlation decreases with the distance (Tobler's law)
- Positive definite (PD) function.

Gaussian Process Regression

Using BLUP we now can estimate the function z at the points S^*

$$\mathbb{E}(z^*|\boldsymbol{z}) = X^*\widehat{\beta} + \mathcal{K}(S^*,S)(\mathcal{K}(S,S)+\sigma^2 I_n)^{-1}(\boldsymbol{z} - X\widehat{\beta})$$

$$\mathrm{Var}(z^*|\boldsymbol{z}) = \mathcal{K}(S^*,S^*) - \mathcal{K}(S^*,S)\mathcal{K}(S,S)^{-1}\mathcal{K}(S,S^*)$$



(a), prior (b), posterior

Figure 1.1: Panel (a) shows four samples drawn from the prior distribution. Panel (b) shows the situation after two datapoints have been observed. The mean prediction is shown as the solid line and four samples from the posterior are shown as dashed lines. In both plots the shaded region denotes twice the standard deviation at each input value x.

Figure: Taken from Rasmussen (2003)

Weak Stationarity

Weak Stationarity:

- $\mathbb{E}z(\boldsymbol{s}) = \mathbb{E}z(\boldsymbol{s} + \boldsymbol{s}_\tau) = \mu$, where $\mu \in \mathbb{R}$ and $\boldsymbol{s}_\tau \in \mathbb{S}$, and satisfies $\boldsymbol{s} + \boldsymbol{s}_\tau \in \mathbb{S}$.
- $\mathrm{Cov}\big(z(\boldsymbol{s}_i), z(\boldsymbol{s}_j)\big) = \mathrm{Cov}\big(z(\boldsymbol{s}_i + \boldsymbol{s}_\tau), z(\boldsymbol{s}_j + \boldsymbol{s}_\tau)\big)$.

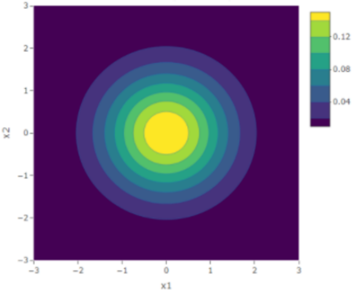
Under the stationarity assumption the covariance function can be denoted by $C(\boldsymbol{h}_{i,j})$, where $\boldsymbol{h}_{i,j} = \boldsymbol{s}_i - \boldsymbol{s}_j$. It emphasizes that the directed distance between the locations is the sufficient argument for the covariance function (rather than the coordinate values themselves). In many use cases, the raw data should be preprocessed in order to assume stationarity.

Kernels - Examples

Gaussian (squared exponential)

$$\sigma_s^2 e^{-(\|\boldsymbol{h}\|/\ell)^2}$$

where σ_s^2 is also called the *signal* parameter and ℓ is the *length-scale* parameter.



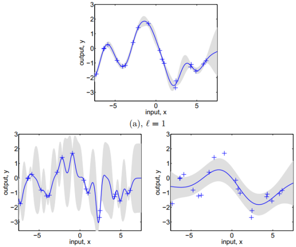
Kernels - Examples

Gaussian (squared exponential)

$$\sigma_s^2 e^{-(\|\boldsymbol{h}\|/\ell)^2}$$

where σ_s^2 is also called the *signal* parameter and ℓ is the *length-scale* parameter.

- Unlike in the previous graph, here we also have noise.
- Taken from Rasmussen (2003).



(a), $\ell = 1$

(b), $\ell = 0.3$

(c), $\ell = 3$

Figure 2.5: (a) Data is generated from a GP with hyperparameters $(\ell, \sigma_f, \sigma_n) = (1, 1, 0.1)$, as shown by the + symbols. Using Gaussian process prediction with these hyperparameters we obtain a 95% confidence region for the underlying function f (shown in grey). Panels (b) and (c) again show the 95% confidence region, but this time for hyperparameter values $(0.3, 1.08, 0.00005)$ and $(3.0, 1.16, 0.89)$ respectively.

Kernels - Examples

Exponential

$$\sigma_s^2 e^{-\|\boldsymbol{h}\|/\ell}$$

The exponential is less smooth than the Gaussian around zero (which makes the exponential better in many cases).

Power Exponential

$$\sigma_s^2 e^{-(\|\boldsymbol{h}\|/\ell)^k}$$

- When $k = 1$ we get exponential
- When $k = 2$ we get Gaussian

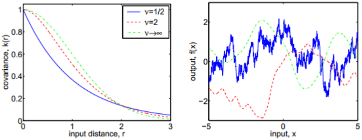
Kernels - Examples

Matern We can generalize it even farther using the Matern function

$$\sigma_s^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\frac{\sqrt{2\nu}\|\boldsymbol{h}\|}{\ell})^{\nu} K_{\nu}(\frac{\sqrt{2\nu}\|\boldsymbol{h}\|}{\ell}).$$

where $\Gamma(\cdot)$ is the Gamma function and $K_{\nu}(\cdot)$ is the Bessel function of the second type.

- Matern kernel is very flexible due to ν, which controls the smoothness around zero.
- $\nu = 1/2$ is exponential, $\nu \to \infty$ convergences to Gaussian.
- Taken from Rasmussen (2003).



(a) (b)

Figure 4.1: Panel (a): covariance functions, and (b): random functions drawn from Gaussian processes with Matérn covariance functions, eq. (4.14), for different values of ν, with $\ell = 1$. The sample functions on the right were obtained using a discretization of the x-axis of 2000 equally-spaced points.

Kernels - Examples

Wave

$$\sigma_s^2 \frac{\ell}{\|\boldsymbol{h}\|} sin(\frac{\|\boldsymbol{h}\|}{\ell}).$$

Spherical

$$T(\boldsymbol{h}) = \begin{cases} 1.5\frac{\|\boldsymbol{h}\|}{\ell} - 0.5(|\frac{\|\boldsymbol{h}\|}{\ell}|)^3 & \|\boldsymbol{h}\| < \ell \\ 0 & o.w \end{cases}$$

Kernels - Examples

Tent PD in $\mathbb{R}$ (but not in higher dimensions).

$$T(\boldsymbol{h}) = \begin{cases} 1 - \|\boldsymbol{h}\|/\ell & \|\boldsymbol{h}\| < 1 \\ 0 & o.w \end{cases}$$

Nugget

$$N(\boldsymbol{h}) = \begin{cases} 1 & \|\boldsymbol{h}\| = 0 \\ 0 & o.w \end{cases}$$

(Power exponential when $k \to 0$.)

Variogram

In spatial statistics in many cases the variogram function is used instead of the covariance function.

$$\gamma(\boldsymbol{h}) := \frac{1}{2}\mathrm{Var}(z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}))$$

Variogram

In spatial statistics in many cases the variogram function is used instead of the covariance function.

$$\gamma(\boldsymbol{h}) := \frac{1}{2}\mathrm{Var}(z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}))$$

- $\gamma(\boldsymbol{h}) = C(0) - C(\boldsymbol{h})$.
- The variogram is conditionally negative definite (rather than negative definite): $f : \mathbb{S} \to \mathbb{R}$ is conditionally ND if $\forall\{\boldsymbol{s}_i\}$ and $\forall\{a_i \in \mathbb{R}\}$, s.t $\sum a_i = 0$

  $$\sum_i \sum_j a_i f(\boldsymbol{s}_i - \boldsymbol{s}_j) a_j \leq 0.$$

  This is one reason that variogram is commonly preferred over the covariance function.

Variogram

$$\gamma(\boldsymbol{h}) = C(0) - C(\boldsymbol{h})$$

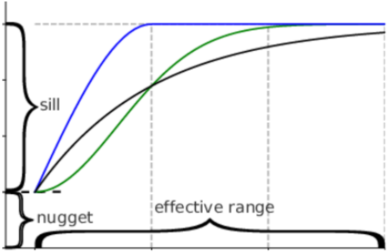


Figure: The three most common theoretical variogram models: spherical (blue), exponential (black) and Gaussian (green). All three variograms share the same variogram parameters: nugget = 0.2, sill = 0.8 and range = 0.2. Taken from Mälicke et al. (2018).

Variogram estimation

Frequently, the variogram is estimated by the empirical variogram. The empirical variogram in the range $\boldsymbol{h}^* \pm \delta$, where $\delta \in \mathbb{S}$ is:

$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2 \times |\mathcal{H}^*|} \sum_{(i,j) \in \mathcal{H}^*} \big(z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j)\big)^2, \; \boldsymbol{h} \in \boldsymbol{h}^* \pm \delta$$

where $\mathcal{H}^* = \{(i,j) | \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}^* \pm \delta\}$ and $|\mathcal{H}^*|$ is the size of the set $\mathcal{H}^*$.

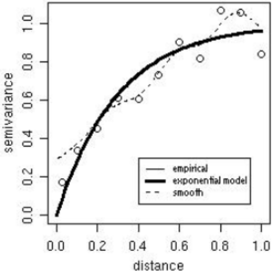
Variogram estimation

Frequently, the variogram is estimated by the empirical variogram. The empirical variogram in the range $\boldsymbol{h}^* \pm \delta$, where $\delta \in \mathbb{S}$ is:

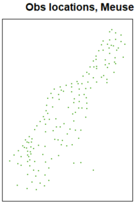
$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2 \times |\mathcal{H}^*|} \sum_{(i,j) \in \mathcal{H}^*} \left(z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j)\right)^2, \ \boldsymbol{h} \in \boldsymbol{h}^* \pm \delta$$

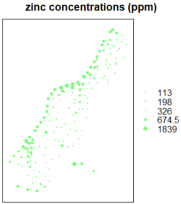
where $\mathcal{H}^* = \{(i,j) | \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}^* \pm \delta\}$ and $|\mathcal{H}^*|$ is the size of the set $\mathcal{H}^*$.

Then, in order to fit a smooth function, the empirical variogram is estimated by variogram kernel functions.

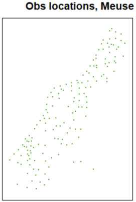


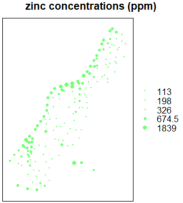
Kriging - Meuse Data

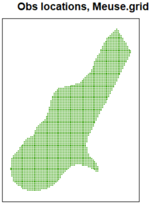


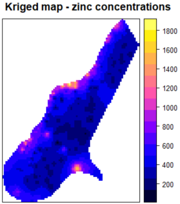


Kriging - Meuse Data


Obs locations, Meuse


zinc concentrations (ppm)

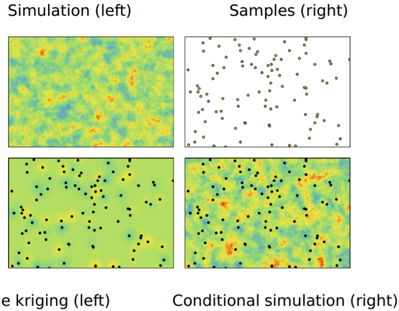

Obs locations, Meuse.grid


Kriged map - zinc concentrations

Conditional Simulation

Conditional simulation is a tool to simulate random field data given observed points.

Unlike Kriging, here the goal is to sample new observations that preserve the distributional properties.



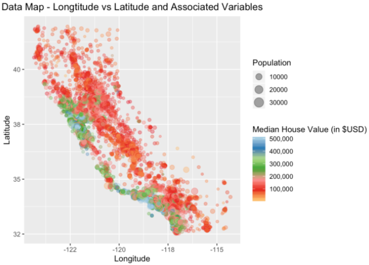
Taken from: Hans Wackernagel, 2013, Basics in Geostatistics 3 Geostatistical

Modeling, California housing

There are many other motivations for fitting GPR, prediction, inference etc. Commonly the covariance is structured by multiple kernel.

$$\mathrm{Cov}(z_i, z_j) = \mathbb{K}(\|z_i - z_j\|) + \sigma_b^2 I_{(c(i)=c(j))} + \sigma^2 I_{(i=j)}$$

where $c(i)$ is the cluster of z_i.



In many cases there are combinations of more complicated kernels, including different kernels for different distance scales.

Extensions

- Generalized GPR (GGPR).
- Linear Mixed Models (LMM)/GLMM.

Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|),\ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.

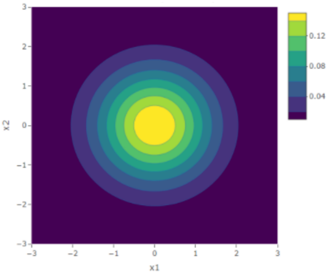
[1] we can think about isotropy in a broader perspective where it does not derive stationarity, however it is not very common.

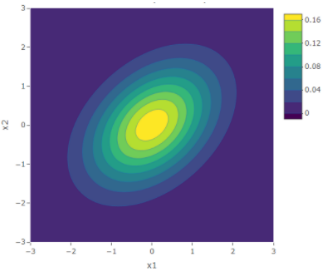
Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|),\ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.





Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|),\ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.

- stationarity and isotropy:

[1] we can think about isotropy in a broader perspective where it does not derive stationarity, however it is not very common.

Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|), \ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.

- stationarity and isotropy:

$$\text{isotropy} \rightarrow \text{stationarity}$$

[1]

[1] we can think about isotropy in a broader perspective where it does not derive stationarity, however it is not very common.

Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|),\ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.

- stationarity and isotropy:

  $$\text{isotropy} \rightarrow \text{stationarity}$$

  1

- symmetry and isotropy.

[1] we can think about isotropy in a broader perspective where it does not derive stationarity, however it is not very common.

Definition

Isotropy is rotational invariance of the correlation structure, i.e.,

$$C(\boldsymbol{h}) = C(\|\boldsymbol{h}\|),\ \forall \boldsymbol{h} \in \mathbb{S}$$

here we focus on $\mathbb{S} \subset \mathbb{R}^2$.

- stationarity and isotropy:

  isotropy → stationarity

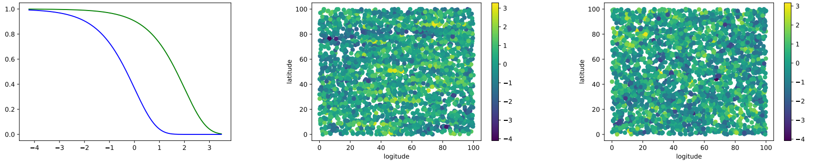
  [1]

- symmetry and isotropy.

  isotropy → symmetry

[1] we can think about isotropy in a broader perspective where it does not derive stationarity, however it is not very common.

Simulation



(a) (b) (c)

Figure: Simulated data of isotropy and anisotropy settings. Anisotropy setting: (a) presents covariance functions in two directional axes that decay differently as a function of the distance (with log_{10} scale). The green line is the covriance function for the longitudinal axis, and the blue line is for the latitudinal axis. (b) presents a simulated sample of an anisotropy setting. As once can see the variance in the latitudinal axis is larger than in the longitudinal axis. This is due to the higher correlation in the longitudinal axis than in the latitudinal axis. (c)

Granny Creek Field

The Granny Creek Field dataset contains 181 measurements of a sandstone base elevation in Granny Creek Field, central West Virginia.

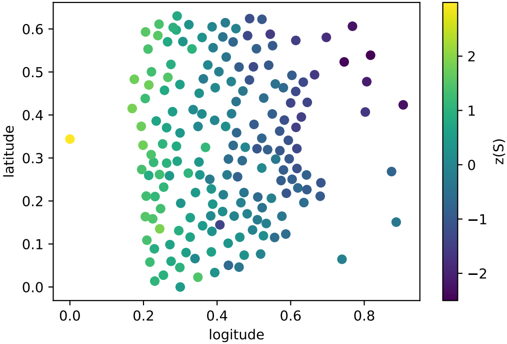
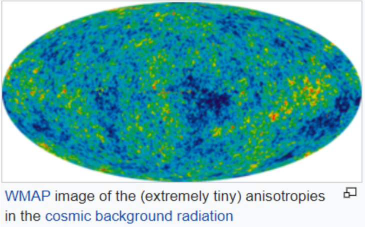


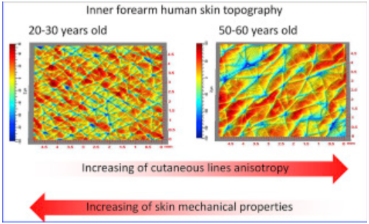
Figure: Granny Creek Field dataset.

Additional Examples



WMAP image of the (extremely tiny) anisotropies in the cosmic background radiation

(a) Taken from Wikipedia


Inner forearm human skin topography
20-30 years old
50-60 years old
Increasing of cutaneous lines anisotropy
Increasing of skin mechanical properties

(b) Taken from Thieulin et al. (2020)

Anisotropic Kernels

A specific type of geometric anisotropy is elliptic anisotropy, when a linear transformation of $\boldsymbol{h}$ induces isotropy.

For example, the standard exponential kernel function, which assumes isotropy, can be generalized using rotation and scaling matrices in order to capture elliptic anisotropy:

$$\sigma_s^2 \times \exp\left(-\|A\boldsymbol{h}\|\right) = \sigma_s^2 \times \exp\left(-\sqrt{\boldsymbol{h}^t A^t A\boldsymbol{h}}\right),$$

where for $\mathbb{S} \subset \mathbb{R}^2$

$$A := A(\lambda_1, \lambda_2, \eta) = \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{bmatrix},$$

$\eta \in [0, \pi]$ and $\eta + \pi/2$ are the anisotropy direction axes and $\lambda_i \in \mathbb{R}^+$ are the anisotropic scales. When $\lambda_1 = \lambda_2$, then it is reduced to the standard exponential kernel.

Directional Variogram

Surprisingly, the most common way to detect anisotropy is using the directional variogram graph:

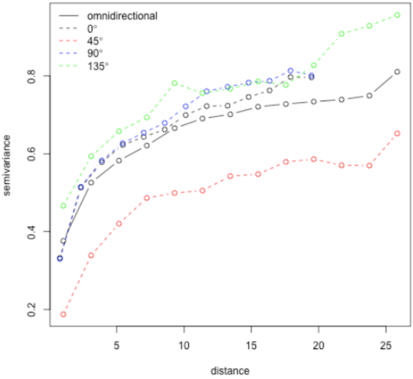


Figure: Directional Variogram

Goal

There are several hypothesis testing approaches for detecting anisotropy, all of them are based on heavy asymptotic assumptions.

In 'Resampling Methods for Detecting Anisotropic Correlation Structure' (Rabinowicz & Rosset, 2021) we propose new hypothesis testing algorithms that are not (directly) based on asymptotic assumptions.

Parametric Bootstrap Based Test - General Setting

- Hypotheses:
  - H_0 : $\boldsymbol{z}(S)$ was sampled from a distribution with an isotropic covariance function.
  - H_1 : $\boldsymbol{z}(S)$ was sampled from a distribution with an anisotropic covariance function.
- Assumption for both hypotheses:
  1. normality
  2. stationarity
  3. parametric covariance function structure (e.g., exp kernel family)
- H_1 can specify the suspected anisotropic directional axes, $\{\eta_i\}_{i\in[1,\dots,R]}$, or alternatively only assume that there are R anisotropic directional axes. Another setting is when H_1 specifies ranges of $\{\eta_i\}_{i\in[1,\dots,R]}$.

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)$, $\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.

   The kernel parameters can be estimated using various approaches, such as maximum likelihood and restricted maximum likelihood (REML) of z.

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)$, $\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.
2. Calculate the following anisotropic discrepancy measure:

$$\phi = \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big),$$

where $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)$ and $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)$ are the log-likelihood of $\boldsymbol{z}$ under the two hypotheses.

ϕ, the anisotropic discrepancy measure, can also be written as follows:

$$\phi = \Big(-\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)\Big) - \Big(-\ell\big(\boldsymbol{z},\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)\Big).$$

Therefore ϕ is the loss function of the null hypothesis minus the loss function of the alternative hypothesis (where the loss function in this case is minus log-likelihood).

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S),\ \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.
2. Calculate the following anisotropic discrepancy measure:

   $$\phi = \ell\big(\boldsymbol{z}; \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z}; \mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big),$$

   where $\ell\big(\boldsymbol{z}; \mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)$ and $\ell\big(\boldsymbol{z}; \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)$ are the log-likelihood of $\boldsymbol{z}$ under the two hypotheses.
3. Using parametric bootstrap estimate $P(\phi|H_0)$. For $b \in [1, ..., B],\ B \in \mathbb{N}:$
   1. Sample one set of observations from $N_n(\mu\mathbb{1}, \mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S))$, and denote the sample as $\boldsymbol{z}^{(b)}$.

      μ can be estimated by the mean. Alternatively, the data can be normalized, such that $\mu = 0$.

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S),\ \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.
2. Calculate the following anisotropic discrepancy measure:

   $$\phi = \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big),$$

   where $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)$ and $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)$ are the log-likelihood of $\boldsymbol{z}$ under the two hypotheses.
3. Using parametric bootstrap estimate $P(\phi|H_0)$. For $b \in [1,...,B],\ B \in \mathbb{N}:$
   1. Sample one set of observations from $N_n(\mu\mathbb{1},\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S))$, and denote the sample as $\boldsymbol{z}^{(b)}$.
   2. Estimate $\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_1}(S,S),\ \mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_0}(S,S)$ using $\boldsymbol{z}^{(b)}$.

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S),\ \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.
2. Calculate the following anisotropic discrepancy measure:

$$\phi = \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big),$$

   where $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)$ and $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)$ are the log-likelihood of $\boldsymbol{z}$ under the two hypotheses.

3. Using parametric bootstrap estimate $P(\phi|H_0)$. For $b \in [1,...,B],\ B \in \mathbb{N}$:
   1. Sample one set of observations from $N_n(\mu\mathbb{1},\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S))$, and denote the sample as $\boldsymbol{z}^{(b)}$.
   2. Estimate $\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_1}(S,S),\ \mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_0}(S,S)$ using $\boldsymbol{z}^{(b)}$.
   3. Calculate: $\phi^{(b)} = \ell\big(\boldsymbol{z}^{(b)};\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z}^{(b)};\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_0}(S,S)\big)$.

Parametric Bootstrap Based Test - The Algorithm

1. Estimate $\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S),\ \mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)$.
2. Calculate the following anisotropic discrepancy measure:

$$\phi = \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big),$$

   where $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S)\big)$ and $\ell\big(\boldsymbol{z};\mathcal{K}_{\boldsymbol{\theta}|H_1}(S,S)\big)$ are the log-likelihood of $\boldsymbol{z}$ under the two hypotheses.

3. Using parametric bootstrap estimate $P(\phi|H_0)$. For $b \in [1,...,B],\ B \in \mathbb{N}:$
   1. Sample one set of observations from $N_n(\mu\mathbb{1},\mathcal{K}_{\boldsymbol{\theta}|H_0}(S,S))$, and denote the sample as $\boldsymbol{z}^{(b)}$.
   2. Estimate $\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_1}(S,S),\ \mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_0}(S,S)$ using $\boldsymbol{z}^{(b)}$.
   3. Calculate: $\phi^{(b)} = \ell\big(\boldsymbol{z}^{(b)};\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_1}(S,S)\big) - \ell\big(\boldsymbol{z}^{(b)};\mathcal{K}^{(b)}_{\boldsymbol{\theta}|H_0}(S,S)\big)$.
4.

$$\text{P-value} = |\{\phi \le \phi^{(b)}|b \in [1,...,B]\}|/B,$$

   where $|\cdot|$ is the set size.

Comments

The parametric bootstrap hypothesis testing approach allows flexibility in different aspects:

- The statistic ϕ can be modified to other loss functions measuring the anisotropic discrepancy magnitude, such as test set error or even prediction errors, e.g., AIC (Akaike, 1974), Cp (Mallows, 1973) and cross-validation (Stone, 1974) error types.

Comments

The parametric bootstrap hypothesis testing approach allows flexibility in different aspects:

- The statistic ϕ can be modified to other loss functions measuring the anisotropic discrepancy magnitude, such as test set error or even prediction errors, e.g., AIC (Akaike, 1974), Cp (Mallows, 1973) and cross-validation (Stone, 1974) error types.
- Controlling B, which tradeoffs between the resultant P-value resolution and Var(P-value) (still the main factor is $\mathrm{Var}(\boldsymbol{z}(S))$), on one hand and the computational cost on the other hand. Also, B controls the P-value resolution. For example, when $B = 200$ the P-value resolution is 0.005.

Comments

The parametric bootstrap hypothesis testing approach allows flexibility in different aspects:

- The statistic ϕ can be modified to other loss functions measuring the anisotropic discrepancy magnitude, such as test set error or even prediction errors, e.g., AIC (Akaike, 1974), Cp (Mallows, 1973) and cross-validation (Stone, 1974) error types.
- Controlling B, which tradeoffs between the resultant P-value resolution and Var(P-value) (still the main factor is $\mathrm{Var}(\boldsymbol{z}(S))$), on one hand and the computational cost on the other hand. Also, B controls the P-value resolution. For example, when $B = 200$ the P-value resolution is 0.005.
- Specification of $\{\eta_i\}_{i\in[1,\ldots,R]}$.

Parametric Bootstrap with a Non-Parametric Covariance Function - Framework

Potentially, the kernel function family assumption can be avoided by the following paradigm:

1. Calculate

$$y_{i,j} := (z_i - \mu) \times (z_j - \mu),\ h_{i,j} := \boldsymbol{s}_i - \boldsymbol{s}_j$$

   $\forall i, j \in [1, ..., n]$, and define $\{\boldsymbol{y}, H\} = \{y_{i,j}, \boldsymbol{h}_{i,j}\}_{(i,j)\in([1,...,n],[1,...,n])}$

2. Estimate the covariance function by monotonic regression of $\boldsymbol{z}$ with respect to the distance:

   H_0: $\{\|\boldsymbol{h}_{i,j}\|\}_{(i,j)\in([1,...,n],[1,...,n])}$

   H_1: projections of H on $\{\eta_i\}_{i=1}^{R}$.

   Also, the predictions should be non-negative.

Parametric Bootstrap with a Non-Parametric Covariance Function - Challenges

Our try to replace the kernel by a non-parametric function was failed...

Can you guess why?

Parametric Bootstrap with a Non-Parametric Covariance Function - Challenges

The main reason is the difficulty to estimate well the covariance matrices using this method.

Other issues:

- The fitted monotonic regression model might be a non strictly positive definite function.
- Models enforcing monotonicity are computationally expensive, especially models with multiple covariates, as in the anisotropic model. Taking into account the large sample size in our application, $(n+1) \times n/2$, then when n is not very small, the running time is long.
- The anisotropic directional axes, $\{\eta_i\}_{i=1}^{R}$, must be prespecified.

Non-Parametric Rotational Sampling Test - Setting

- Hypotheses:
  - H_0 : $\boldsymbol{z}(S)$ was sampled from a distribution with an isotropic covariance function.
  - H_1 : $\boldsymbol{z}(S)$ was sampled from a distribution with an anisotropic covariance function.
- Assumption for both hypotheses:
  1. ~~normality~~
  2. stationarity
  3. parametric covariance function structure (e.g., exp kernel family)
- H_1 can specify the suspected anisotropic directional axes, $\{\eta_i\}_{i\in[1,\ldots,R]}$, ~~or alternatively only assume that R anisotropic directional axes exist~~. Another setting is when H_1 specifies ranges of $\{\eta_i\}_{i\in[1,\ldots,R]}$.

Non-Parametric Rotational Sampling Test - The Algorithm

Given the suspected directional anisotropic axes $\eta, \eta + \pi/2$:

1. Derive $\{\boldsymbol{y}, H\}$ using $\{S, \boldsymbol{z}(S)\}$, and calculate $\phi = \phi_{isotropy} - \phi_{anisotropy}$, where:

$$\phi_{isotropy} = \min_{\boldsymbol{\theta}|H_0} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_0}(\|\boldsymbol{h}_{i,j}\|)\right)^2$$

$$\phi_{anisotropy} = \min_{\boldsymbol{\theta}|H_1} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_1}(\boldsymbol{h}_{i,j})\right)^2$$

Since we don't assume normality, it is more reasonable to use squared errors loss function than a likelihood-based loss function.

Non-Parametric Rotational Sampling Test - The Algorithm

Given the suspected directional anisotropic axes $\eta, \eta + \pi/2$:

1. Derive $\{\boldsymbol{y}, H\}$ using $\{S, \boldsymbol{z}(S)\}$, and calculate $\phi = \phi_{isotropy} - \phi_{anisotropy}$, where:

$$\phi_{isotropy} = \min_{\boldsymbol{\theta}|H_0} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_0}(\|\boldsymbol{h}_{i,j}\|)\right)^2$$

$$\phi_{anisotropy} = \min_{\boldsymbol{\theta}|H_1} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_1}(\boldsymbol{h}_{i,j})\right)^2$$

2. For $b \in [1, ..., B]$, $B \in \mathbb{N}$:
   1. Sample a random directional axis $\eta^{(b)}$ from $[\eta + \alpha, \eta + \pi/2 - \alpha]$.
   2. Calculate

$$\phi^{(b)}_{anisotropy} = \min_{\boldsymbol{\theta}|H_1,\eta^{(b)}} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_1,\eta^{(b)}}(\boldsymbol{h}_{i,j})\right)^2$$

$$\phi^{(b)} = \phi_{isotropy} - \phi^{(b)}_{anisotropy},$$

where $\mathcal{K}_{\boldsymbol{\theta}|H_1,\eta^{(b)}}$ is the anisotropic kernel with the directions $\{\eta^{(b)}, \eta^{(b)} + \pi/2\}$.

In order to increase power, α prevents sampling axes that are close to the specified anisotropic directional axes.

Non-Parametric Rotational Sampling Test - The Algorithm

Given the suspected directional anisotropic axes $\eta, \eta + \pi/2$:

1. Derive $\{\boldsymbol{y}, H\}$ using $\{S, \boldsymbol{z}(S)\}$, and calculate $\phi = \phi_{isotropy} - \phi_{anisotropy}$, where:

$$\phi_{isotropy} = \min_{\boldsymbol{\theta}|H_0} \sum_{i=1}^{n}\sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_0}(\|\boldsymbol{h}_{i,j}\|)\right)^2$$

$$\phi_{anisotropy} = \min_{\boldsymbol{\theta}|H_1} \sum_{i=1}^{n}\sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_1}(\boldsymbol{h}_{i,j})\right)^2$$

2. For $b \in [1, ..., B]$, $B \in \mathbb{N}$:
    1. Sample a random directional axis $\eta^{(b)}$ from $[\eta + \alpha, \eta + \pi/2 - \alpha]$.
    2. Calculate

$$\phi^{(b)}_{anisotropy} = \min_{\boldsymbol{\theta}|H_1,\eta^{(b)}} \sum_{i=1}^{n}\sum_{j=i}^{n} \left(y_{i,j} - \mathcal{K}_{\boldsymbol{\theta}|H_1,\eta^{(b)}}(\boldsymbol{h}_{i,j})\right)^2$$

$$\phi^{(b)} = \phi_{isotropy} - \phi^{(b)}_{anisotropy},$$

    where $\mathcal{K}_{\boldsymbol{\theta}|H_1,\eta^{(b)}}$ is the anisotropic kernel with the directions $\{\eta^{(b)}, \eta^{(b)} + \pi/2\}$.

3.

$$\text{P-value} = |\{\phi \leq \phi^{(b)} | b \in [1, ..., B]\}|/B.$$

Comments

- For improved readability, two perpendicular anisotropic directional axes are specified, however, it can be easily generalized for more than two and non-perpendicular anisotropic directional axes.

Comments

- For improved readability, two perpendicular anisotropic directional axes are specified, however, it can be easily generalized for more than two and non-perpendicular anisotropic directional axes.
- Instead of specifying suspected anisotropic directional axes, we can specify only ranges. We can use α for constructing non-overlapping domains of $\{\eta_i\}_{i\in[1,\dots,R]}$ and $\{\eta^{(b)}\}_{b\in[1,\dots,B]}$. For example, in case the suspected anisotropic directional ranges are $\{\eta_1 \in [-\alpha,\alpha], \eta_2 = \eta_1 + \pi/2\}$, then the estimation of $\phi_{anisotropy}$ should also include optimization of η_1. Correspondingly, the sampling space in line 4 is $[2\times\alpha, \pi/2 - 2\times\alpha]$, and the optimization in line 5 is also over $\eta^{(b)*} \in [\eta^{(b)} - \alpha, \eta^{(b)} + \alpha]$.

Comments

- For improved readability, two perpendicular anisotropic directional axes are specified, however, it can be easily generalized for more than two and non-perpendicular anisotropic directional axes.
- Instead of specifying suspected anisotropic directional axes, we can specify only ranges. We can use α for constructing non-overlapping domains of $\{\eta_i\}_{i\in[1,\dots,R]}$ and $\{\eta^{(b)}\}_{b\in[1,\dots,B]}$. For example, in case the suspected anisotropic directional ranges are $\{\eta_1 \in [-\alpha,\alpha], \eta_2 = \eta_1 + \pi/2\}$, then the estimation of $\phi_{anisotropy}$ should also include optimization of η_1. Correspondingly, the sampling space in line 4 is $[2\times\alpha, \pi/2 - 2\times\alpha]$, and the optimization in line 5 is also over $\eta^{(b)*} \in [\eta^{(b)} - \alpha, \eta^{(b)} + \alpha]$.
- Similarly to parametric bootstrap algorithm, the kernel can be potentially replaced by non-parametric monotonic regression. In that way, both main parametric assumptions — normality and kernel structure — are avoided.

Simulation

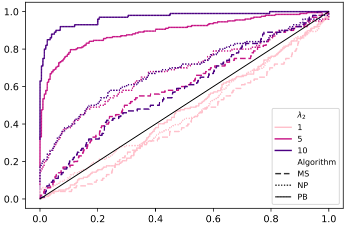


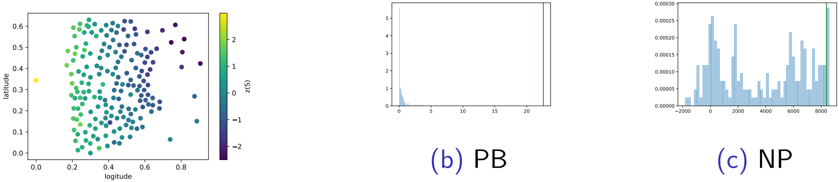
Figure: P-value empirical cumulative distribution for the Parametric bootstrap algorithm (PV), non-parametric algorithm (NP), and Maity and Sherman (2012)'s algorithm (MS, previous algorithm that relies on advance asymptotic) for $n = 500$ and different λ_2.

	$\lambda_2 = 1$			$\lambda_2 = 2$			$\lambda_2 = 5$			$\lambda_2 = 10$		
Algorithm\n	200	500	1000	200	500	1000	200	500	1000	200	500	1000
PB	0.05	0.04	0.05	0.11	0.31	0.43	0.47	0.70	0.94	0.65	0.89	0.99
NP	0.14	0.07	0.07	0.05	0.15	0.19	0.16	0.29	0.24	0.27	0.30	0.33
MS	0.02	0.04	0	0.06	0.04	0.05	0.02	0.10	0.10	0.07	0.10	0.17

Table: The table presents empirical power for different settings for significance level of 0.05. The $\lambda_2 = 1$ column is the Type I error estimates.

Granny Creek Field dataset

The Granny Creek Field dataset contains 181 measurements of a sandstone base elevation in Granny Creek Field, central West Virginia (Hohn, 1998).



(b) PB

(c) NP

(a) Granny Creek Field dataset.

Figure: Granny Creek Field. (a) presents the Granny Creek Field dataset after scaling. (b) and (c) compare the ϕ value (in green) with $\{\phi^{(b)}\}_{b=1}^{200}$.

The P-values of PB, NP and MS are: < 0.005, 0.025 and 0.011, respectively.

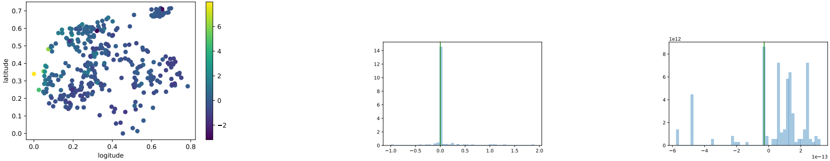
Mississippian Sandstone dataset

The Mississippian Sandstone dataset contains 348 measurements of subsea depth of a mississippian-age reservoir sandstone base in Ritchie County, West Virginia (Hohn, 1998).

Based on prior geographical knowledge, Hohn claims that the correlation in the $\{\pi/4,\ 3\times\pi/4\}$ directional axes is suspected to be different than the correlation in the $\{\pi,\ \pi/2\}$ directional axes. Therefore, the elliptical transformation of the anisotropic kernel is:

$$A = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\lambda_1} & 0 & 0 \\ 0 & 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & 0 & \frac{1}{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos(0) & \sin(0) \\ \cos(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \\ \cos(\frac{\pi}{4}) & \sin(\frac{\pi}{4}) \\ \cos(\frac{3\pi}{4}) & \cos(\frac{3\pi}{4}) \end{bmatrix},$$

Mississippian Sandstone dataset



(a) After pre-processing (b) PB (c) NP

Figure: Mississippian Sandstone. (a) and (b) present the Mississippian Sandstone dataset before and after pre-processing. (c) and (d) compare the ϕ value (in green) with $\{\phi^{(b)}\}_{b=1}^{200}$.

The P-values of Algorithm PB, NP and MS are: 0.93, 0.7 and 0.19, respectively.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723.

Hohn, M. (1998). Geostatistics and petroleum geology. Springer Science & Business Media.

Maity, A. and M. Sherman (2012). Testing for spatial isotropy under general designs. Journal of statistical planning and inference 142(5), 1081–1091.

Mallows, C. L. (1973). Some comments on cp. Technometrics 15(4), 661–675.

Mälicke, M., S. Hassler, M. Weiler, T. Blume, and E. Zehe (2018, 09). Exploring hydrological similarity during soil moisture recession periods using time dependent variograms. Hydrology and Earth System Sciences Discussions, 1–25.

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In Summer school on machine learning, pp. 63–71. Springer.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological) 36(2), 111–133.

Thieulin, C., C. Pailler-Mattei, A. Abdouni, M. Djaghloul, and H. Zahouani (2020). Mechanical and topographical anisotropy for human skin: Ageing effect. journal of the mechanical behavior of biomedical materials 103, 103551.