ON INFERENCE ABOUT THE SMOOTHNESS PARAMETER IN GAUSSIAN MATÉRN RANDOM FIELDS

Victor De Oliveira

Department of Management Science and Statistics
The University of Texas at San Antonio

victor.deoliveira@utsa.edu

http://faculty.business.utsa.edu/vdeolive/

Joint work with Zifei Han
University of International Business and Economics, China

Work supported by National Science Foundation

Basic Geostatistical Problem

Let $z : \mathcal{D} \subset \mathbb{R}^d \to \mathbb{R}$ be an unknown function describing the spatial variation of the quantity of interest over $\mathcal{D}$

The stochastic approach to the modeling of this function assumes that $z(\cdot)$ is a realization of a random field $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$

To make inference about $z(\cdot)$ we need to model the random field $Z(\cdot)$ and estimate some of its probabilistic features

For most geostatistical applications $d = 2$

Geostatistical Data Structure

A set of triplets

$$\{(\mathbf{s}_i, \boldsymbol{f}(\mathbf{s}_i), z_i) \ : \ i = 1, \ldots, n\}$$

where

- $\mathbf{s}_1, \ldots, \mathbf{s}_n$ sampling locations in $\mathcal{D} \subset \mathbb{R}^2$
- $\boldsymbol{f}(\mathbf{s}_i) = (f_1(\mathbf{s}_i), \ldots, (f_p(\mathbf{s}_i))^\top \in \mathbb{R}^p$ set of p explanatory variables taken at sampling location $\mathbf{s}_i$
- z_i is the measurement taken at sampling location $\mathbf{s}_i$

$\mathcal{S}_n := \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ is called the sampling design

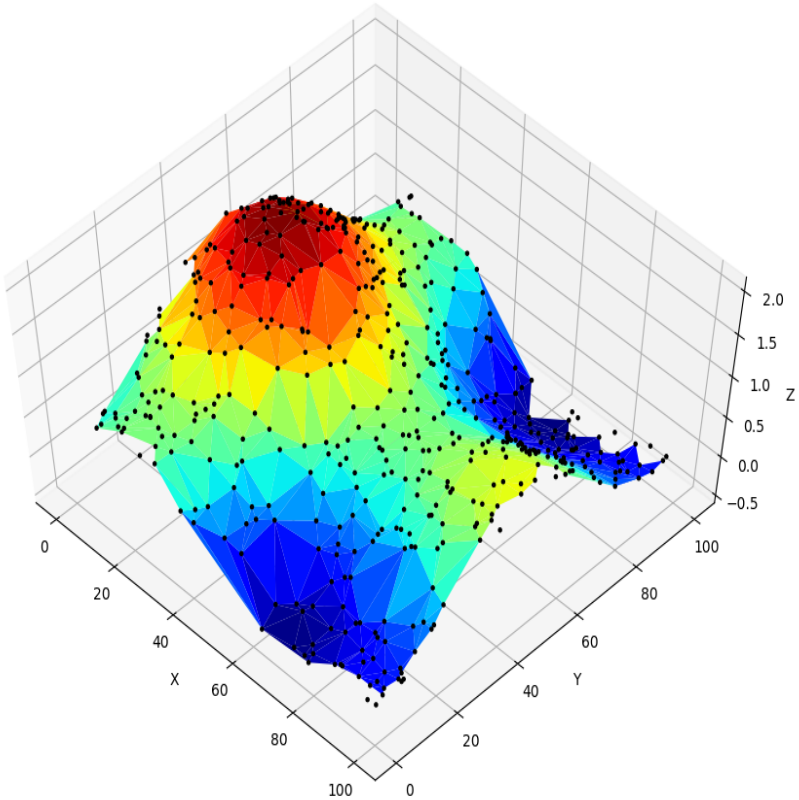
Key feature: geostatistical data usually lacks replication

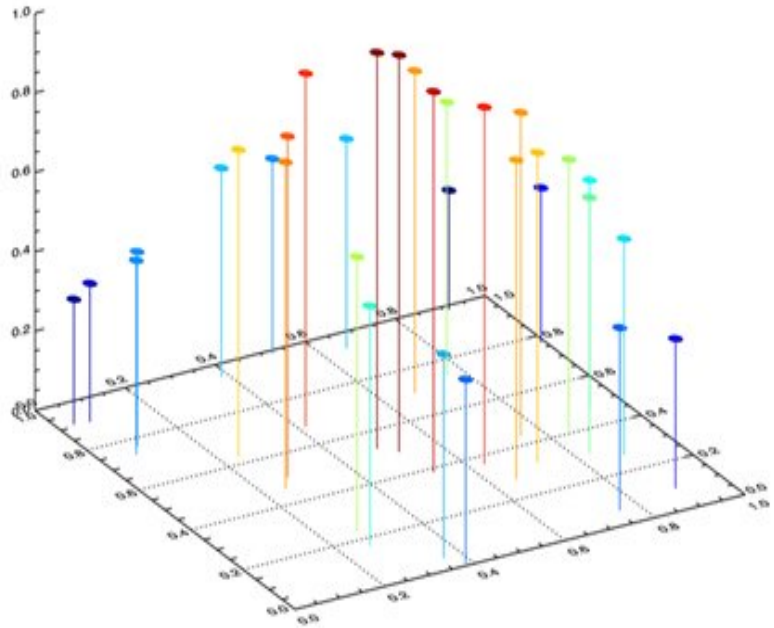
Example: Spatial Variation of Rainfall

Data consist of $n = 467$ measurements of daily rainfall collected in Switzerland on May 8, 1986

- $\mathcal{D}$ = Country of Switzerland
- $\mathbf{s}_1, \ldots, \mathbf{s}_n$ coordinates of locations in $\mathcal{D}$ where tipping buckets were placed
- z_i = Rain amount recorded by the tipping bucket located at $\mathbf{s}_i$
- $f_1(\mathbf{s}_i) = 1$ for all $\mathbf{s}_i$ (no covariates)

Illustration





Basic Geostatistical Model

$Z(\cdot)$ is a Gaussian random field with mean and covariance functions

$$\begin{aligned} \mu(\mathbf{s}) &:= E(Z(\mathbf{s})) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}) \\ C(\mathbf{s},\mathbf{u}) &:= \operatorname{cov}(Z(\mathbf{s}), Z(\mathbf{u})) = \sigma^2 K_{\boldsymbol{\vartheta}}(\mathbf{s},\mathbf{u}) \qquad \mathbf{s},\mathbf{u} \in \mathcal{D} \end{aligned}$$

where

- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ unknown regression parameters
- $f_1(\mathbf{s}), \ldots, f_p(\mathbf{s})$ known location–dependent covariates
- $\sigma^2 = \operatorname{var}(Z(\mathbf{s}))$
- $K_{\boldsymbol{\vartheta}}(\mathbf{s},\mathbf{u})$ correlation function in $\mathbb{R}^2$
- $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_q)^\top \in \Theta \subset \mathbb{R}^q$ unknown correlation parameters

Measurements z_i are often corrupted by additive measurement error:

$$z_i = Z(\mathbf{s}_i) + \epsilon_i, \quad i = 1, \ldots, n$$

$\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \tau^2)$ distribution and independent of $Z(\cdot)$;
$\tau^2 \geq 0$ called the *nugget* parameter

Model parameters: $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \tau^2)$, with $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\vartheta})$

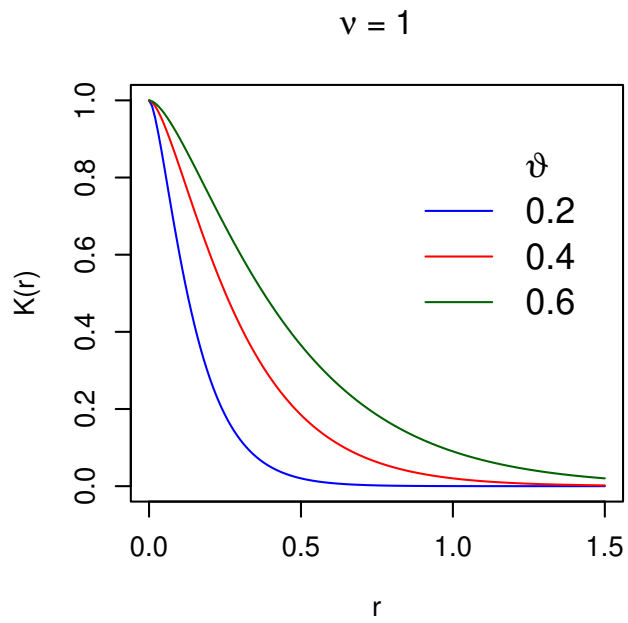
Prominent Model: Matérn Covariance Function

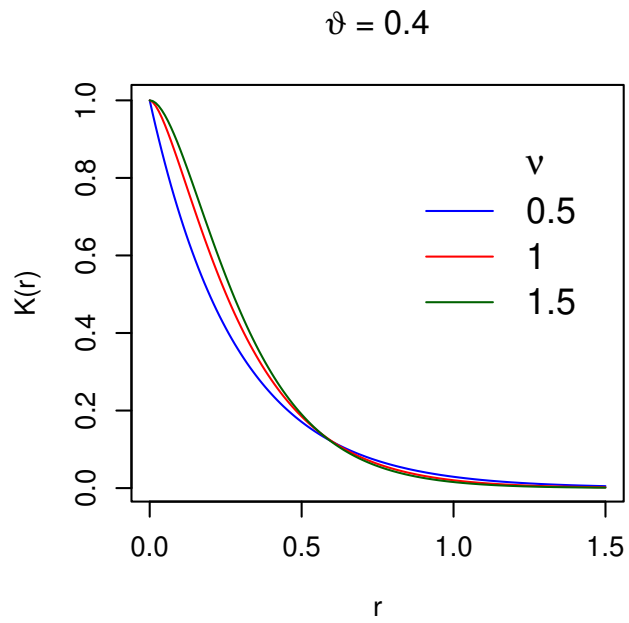
Isotropic Covariance Function in $\mathbb{R}^2$:
$C_{\boldsymbol{\theta}}(r) = \sigma^2 K_{\boldsymbol{\vartheta}}(r)$, with

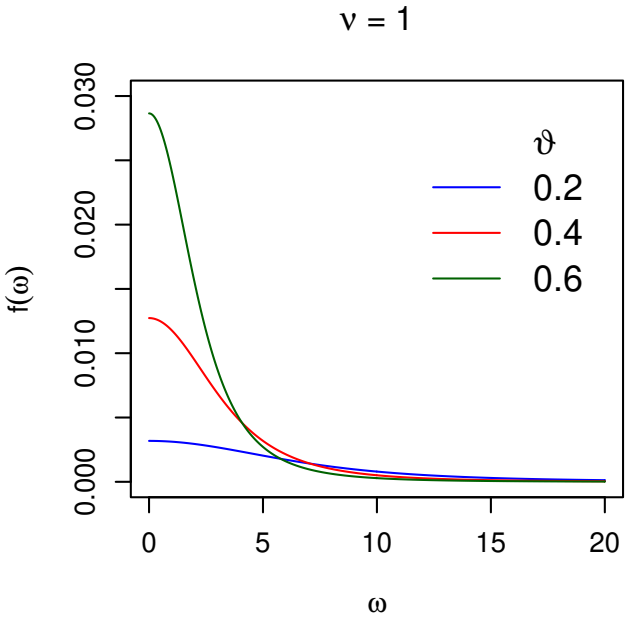
$$K_{\boldsymbol{\vartheta}}(r) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}}{\vartheta} r \right)^{\nu} \mathcal{K}_{\nu} \left(\frac{2\sqrt{\nu}}{\vartheta} r \right), \qquad r \geq 0$$

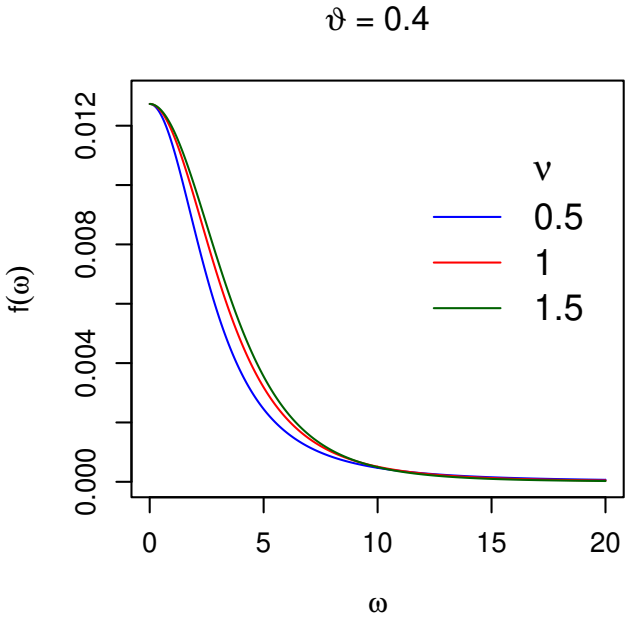
r = Euclidean distance between two locations
$\boldsymbol{\theta} = (\sigma^2, \vartheta, \nu)$ are covariance parameters
$\mathcal{K}_{\nu}(\cdot)$ is modified Bessel function of second kind and order ν

σ^2: variance parameter
ϑ: range parameter
ν: smoothness parameter


ν = 1
K(r)
r
ϑ
0.2
0.4
0.6
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5


ϑ = 0.4
K(r)
r
ν
0.5
1
1.5
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5


ν = 1
f(ω)
ω
ϑ
0.2
0.4
0.6
0.000
0.010
0.020
0.030
0
5
10
15
20


ϑ = 0.4
f(ω)
ω
ν
0.5
1
1.5
0.000
0.004
0.008
0.012
0
5
10
15
20

Reasons for Prominence of Matérn Model

- Unlike most families indexed only by variance and range parameters, the Matérn family is also indexed by a smoothness parameter that controls the degree of mean square differentiability of the random field
- In fixed–domain asymptotic framework: it is possible to achieve efficient spatial interpolation with a misspecified covariance model, as long as the correct and misspecified models are compatible in some well–defined sense.
  A necessary condition for two covariance functions from the Matérn family to be compatible in that sense is that they share the same smoothness parameter.

Mismatch Between Theory and Practice

- The aforementioned reasons point to the importance of modeling the smoothness of the random field, and this is why the Matérn model is so popular in geostatistics
- On the other hand, the smoothness of the random field is rarely estimated in current geostatistical practice, but rather fixed in advance. It is common to use the sub–models resulting from $\nu = 0.5$ (exponential model) or $\nu = 1.5$, even though there is usually little or no a priori information about the smoothness of a random field

Why This Mismatch?

• There are also computational challenges when attempting to estimate all covariance parameters

• Unqualified claim in the literature *suggesting* geostatistical data have little or no information about the smoothness parameter (e.g. Bose, Hodges and Banerjee, 2018)

Diggle and Ribeiro (2007) state that:

"...when using the Matérn correlation function, our experience has been that the shape parameter κ [ν] is often poorly identified"

"...we have found that, for example, estimating all three parameters in the Matérn model is very difficult because the parameters are poorly identified, leading to ridges or plateaus in the log–likelihood surface"

Common geostatistical practice:

Fix the smoothness parameter and estimate the other covariance parameters, variance, range and nugget
This implicitly assumes at least one of the two tenets:

(a) Data contain more information about variance and range parameters than about smoothness parameter

(b) Variance and range parameters are more important for spatial interpolation/prediction than smoothness parameter

Some Theoretical Results

In fixed–domain asymptotic framework:

- The parameters σ^2 and ϑ *cannot* be consistently estimated when $d \leq 3$ (Zhang, 2004)
- The parameter ν *can* be consistently estimated under some designs (Wu, Lim and Xiao, 2013; Wu and Lim, 2019; Loh, 2015; Loh, Sun and Wen, 2021)
- For any prediction location $\mathbf{s}_0$ the BLUP of $Z(\mathbf{s}_0)$ based on a misspecified Matérn model is asymptotically efficient when ν is correctly specified, regardless of the values of σ^2 and ϑ (Kaufman and Shaby, 2013)

These facts do not support the above implicit tenets

Example (Continuation)

Model for the square root transformed data $Z(\cdot)$: Gaussian random field with constant mean and the Matérn covariance function

The MLE of the covariance parameters are

$$\hat{\boldsymbol{\eta}} = (\hat{\sigma}^2, \hat{\tau}^2, \hat{\vartheta}, \hat{\nu}) = (105.09, 6.74, 73.42, 0.95)$$

Can use an estimate of the observed information matrix, $H(\hat{\boldsymbol{\eta}})$, to quantify information about covariance parameters
From the output of the optimization algorithm

$$\hat{H}^{-1} = \begin{pmatrix} 1720.183 & -5.633 & 838.860 & -2.431 \\ -5.633 & 2.817 & -17.850 & 0.474 \\ 838.860 & -17.850 & 626.629 & -5.223 \\ -2.431 & 0.474 & -5.223 & 0.105 \end{pmatrix}$$

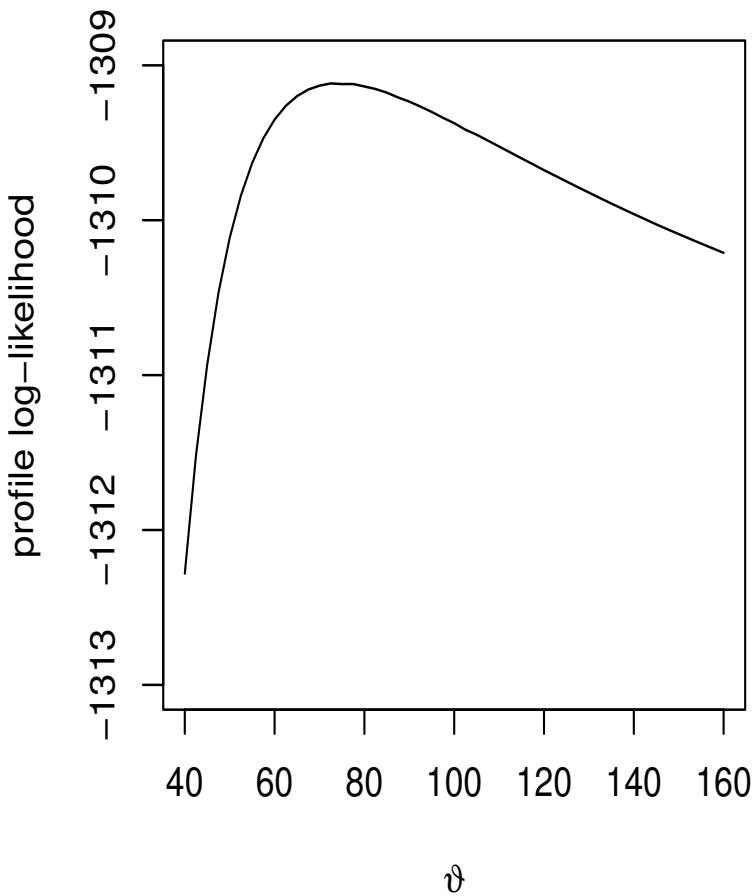
Information about a parameter may be quantified by inverse of Cramer–Rao lower bound for the variance of unbiased estimators of that parameter
For ν, this is estimated by $1/\hat{H}^{\nu\nu}$, with $\hat{H}^{\nu\nu}$ the '(ν,ν)' diagonal element of $\hat{H}^{-1}$

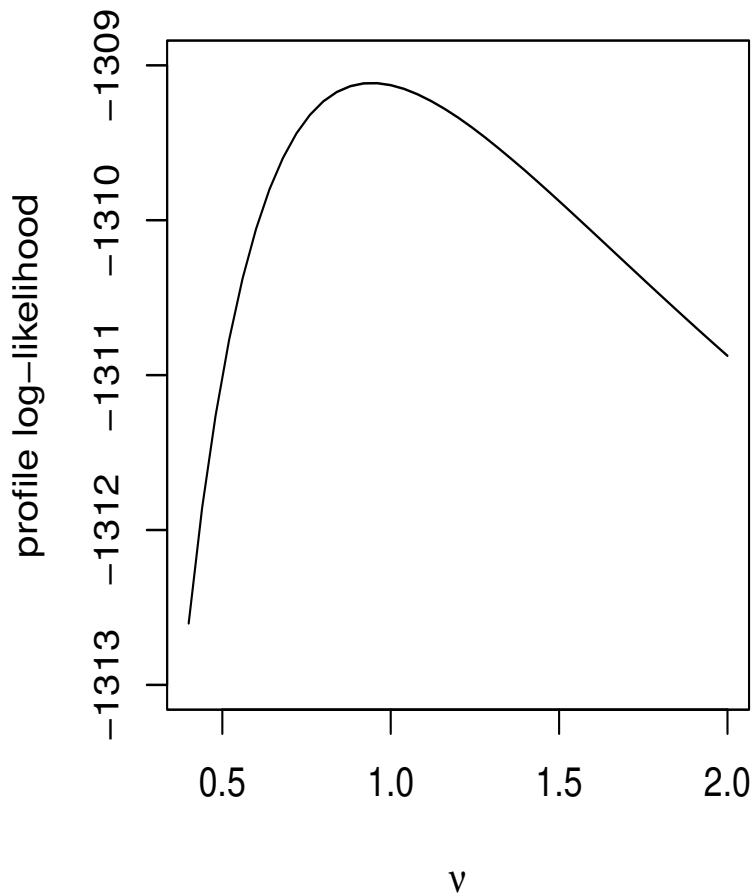
For the Swiss rainfall data

$$1/\hat{H}^{\nu\nu} = 9.524 \quad \text{and} \quad 1/\hat{H}^{\vartheta\vartheta} = 0.0016$$

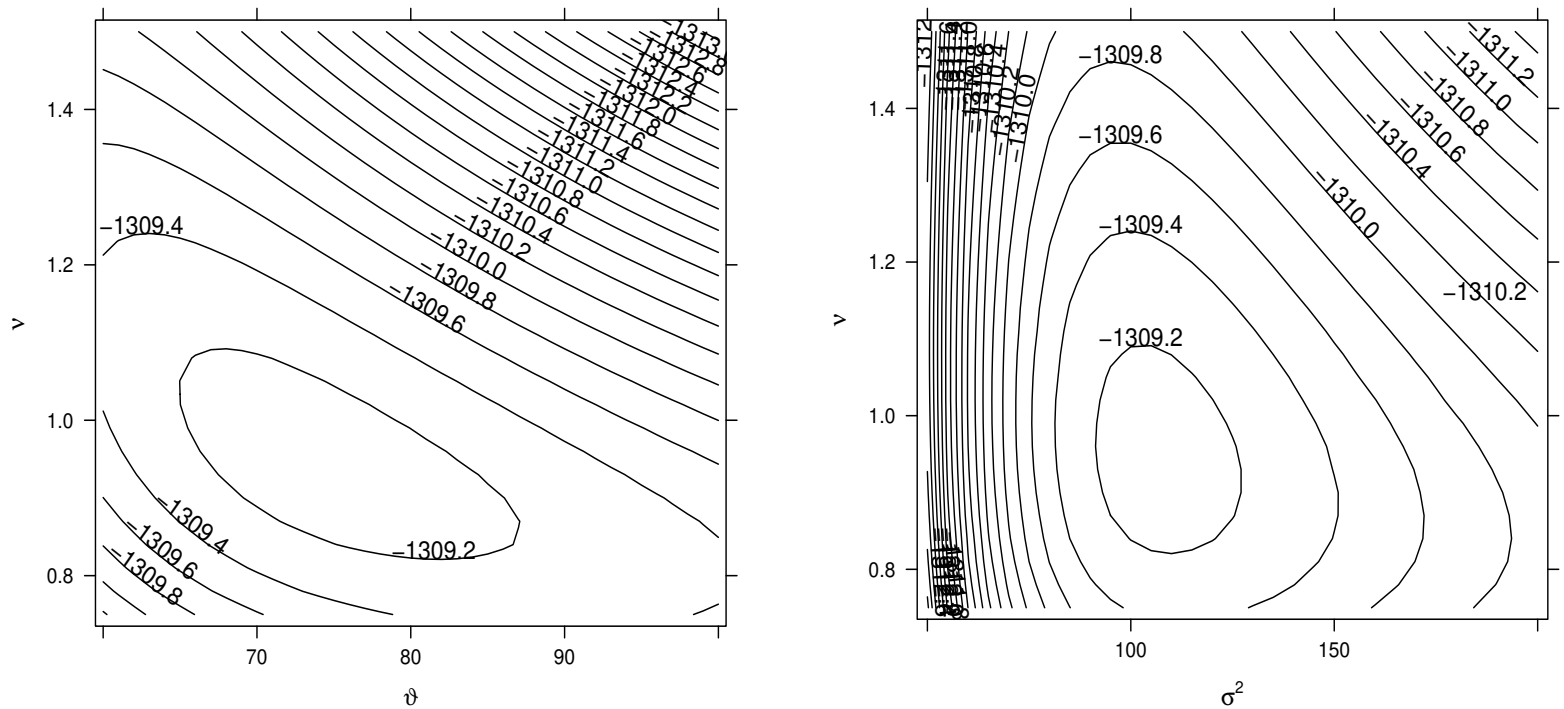
This suggests data contain substantial information about ν
But not necessarily that the data are more informative about ν than about ϑ
To make these comparable, some standarization is needed since $\hat{H}^{\vartheta\vartheta}$ depends heavily on units used to measure distance

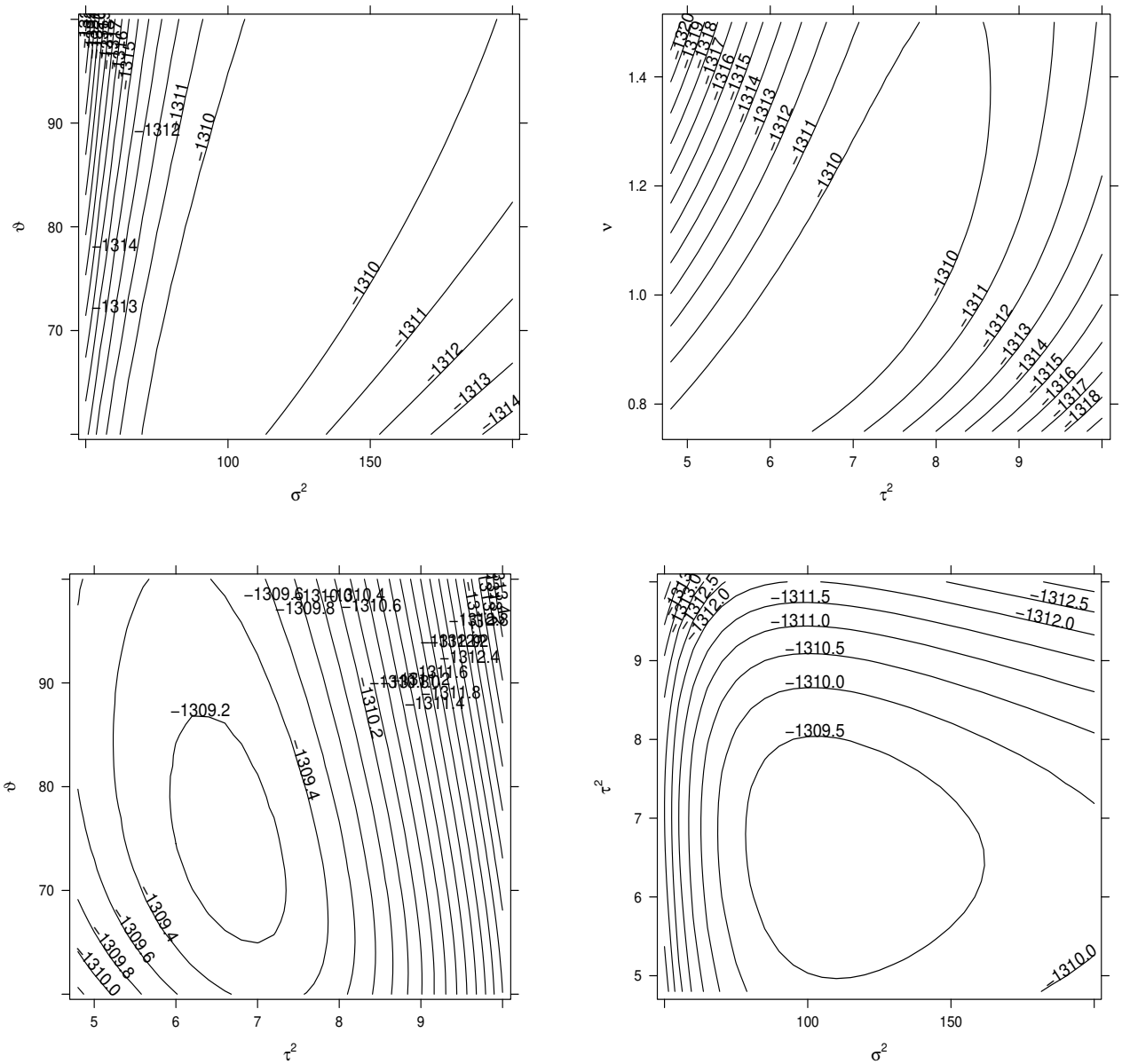
An equivalent visual approach to quantify information involves inspecting the profile log–likelihoods $\mathrm{pl}_1(\vartheta)$ and $\mathrm{pl}_2(\nu)$





Since ϑ and ν are non–orthogonal and they are not the only model parameters, a more complete analysis involves inspection of joint profile log–likelihoods




ϑ
σ²
ν
τ²
ϑ
τ²
τ²
σ²

Quantifying Information About Covariance Parameters

Carry out numerical exploration to uncover how the sampling design and true model affect the information content the data have about the range and smoothness parameters

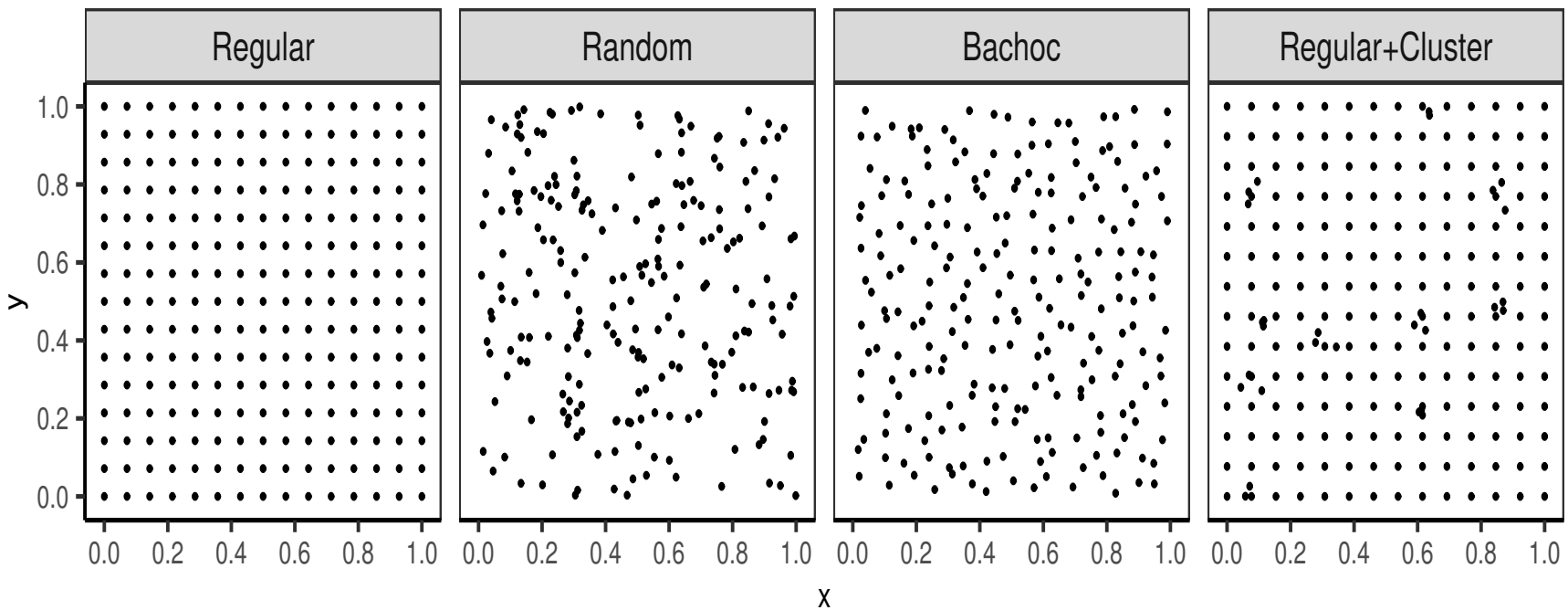
The information content about each covariance parameter is measured by the vector

$$\text{Inf}(\boldsymbol{\eta}, \mathcal{S}_n) := \left(\frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\sigma^2\sigma^2}}, \frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\tau^2\tau^2}}, \frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\vartheta\vartheta}}, \frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\nu\nu}} \right)$$

$I(\boldsymbol{\eta}, \mathcal{S}_n)^{\sigma^2\sigma^2}$, $I(\boldsymbol{\eta}, \mathcal{S}_n)^{\tau^2\tau^2}$, $I(\boldsymbol{\eta}, \mathcal{S}_n)^{\vartheta\vartheta}$ and $I(\boldsymbol{\eta}, \mathcal{S}_n)^{\nu\nu}$ are, respectively, first, second, third and fourth diagonal elements of $I(\boldsymbol{\eta}, \mathcal{S}_n)^{-1}$, and $I(\boldsymbol{\eta}, \mathcal{S}_n)$ is Fisher information matrix based on sampling design $\mathcal{S}_n$ when true covariance parameter is $\boldsymbol{\eta}$

Sampling Designs

Consider the following design types to assess effect of sampling design $\mathcal{S}_n$ on amount of information the data have about the range and smoothness parameters $(n = 225)$



Fisher Information Matrix

Let $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \eta_4) = (\sigma^2, \tau^2, \vartheta, \nu)$ covariance parameters
For Gaussian random fields the Fisher information matrix of $\boldsymbol{\eta}$ based on the data model is the 4×4 matrix $I(\boldsymbol{\eta}, \mathcal{S}_n)$ with entries

$$I(\boldsymbol{\eta}, \mathcal{S}_n)_{ij} = \frac{1}{2}\mathrm{tr}\big(\Psi^{-1}(\boldsymbol{\eta}, \mathcal{S}_n)\Psi_i(\boldsymbol{\eta}, \mathcal{S}_n)\Psi^{-1}(\boldsymbol{\eta}, \mathcal{S}_n)\Psi_j(\boldsymbol{\eta}, \mathcal{S}_n)\big)$$

where

$\Psi(\boldsymbol{\eta}, \mathcal{S}_n) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} + \tau^2 \boldsymbol{I}_n$
$(\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}})_{ij} = K_{\boldsymbol{\vartheta}}(\|\mathbf{s}_i - \mathbf{s}_j\|)$
$\Psi_i(\boldsymbol{\eta}, \mathcal{S}_n) = (\partial/\partial\eta_i)\Psi(\boldsymbol{\eta}, \mathcal{S}_n)$

Computation of $I(\boldsymbol{\eta}, \mathcal{S}_n)$ is involved due to required derivatives of Bessel function $\mathcal{K}_\nu(x)$ w.r.t. x and ν

For any $\boldsymbol{\vartheta} = (\vartheta, \nu) \in (0, \infty)^2$ and $r > 0$

$$\frac{\partial}{\partial \vartheta} K_{\boldsymbol{\vartheta}}(r) = \frac{4\nu^{\frac{\nu+1}{2}} r^{\nu+1}}{\Gamma(\nu)\vartheta^{\nu+2}} \mathcal{K}_{\nu-1}\Big(\frac{2\sqrt{\nu}}{\vartheta} r\Big)$$

and

$$\frac{\partial}{\partial \nu} K_{\boldsymbol{\vartheta}}(r) = \left(\log\left(\frac{\sqrt{\nu}}{\vartheta} r\right) - \psi(\nu)\right) K_{\boldsymbol{\vartheta}}(r)$$
$$- h(\nu)\left(\frac{r}{\vartheta\sqrt{\nu}} \mathcal{K}_{\nu-1}\left(\frac{2\sqrt{\nu}}{\vartheta} r\right) - \int_0^\infty t \sinh(\nu t) \exp\left(-\frac{2r\sqrt{\nu}}{\vartheta} \cosh(t)\right) dt\right)$$

where $\psi(\nu)$ is the digamma function and

$$h(\nu) := \frac{2}{\Gamma(\nu)} \left(\frac{\sqrt{\nu}}{\vartheta} r\right)^\nu$$

Numerical Exploration of Information Patterns

Consider Gaussian random fields with constant mean and Matérn covariance function, observed in $\mathcal{D} = [0,1] \times [0,1]$ using one of sampling designs described before

Fix $\sigma^2 = 1$ and $\tau^2 = 0.2$ and explore patterns of variation of

$$\frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\vartheta\vartheta}} \quad \text{and} \quad \frac{1}{I(\boldsymbol{\eta}, \mathcal{S}_n)^{\nu\nu}}$$

for (ϑ, ν) in $[0.05, 0.65] \times [0.1, 1.5]$

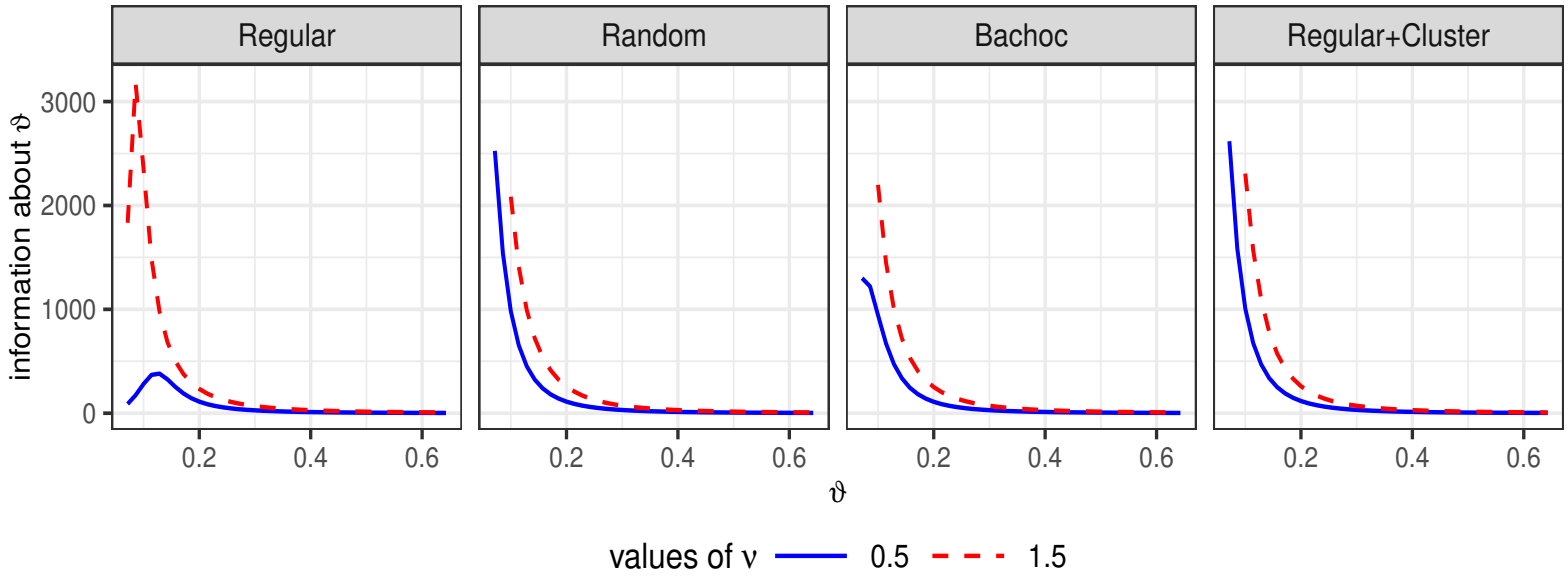
In the following the spatial coordinates are re–scaled
New coordinates defined as

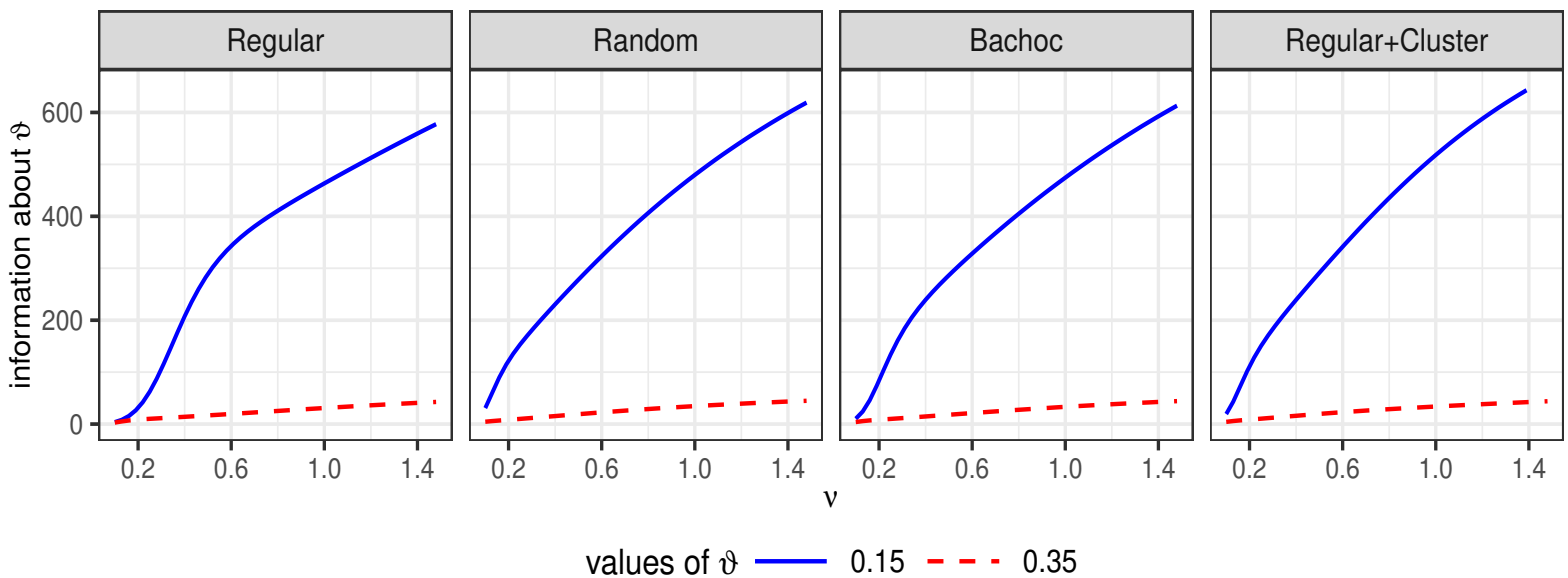
$$\tilde{\mathbf{s}} = (\tilde{x}, \tilde{y}) := \frac{\mathbf{s}}{r_{\max}}$$

with $r_{\max} := \max\{||\mathbf{s} - \mathbf{u}|| : \mathbf{s}, \mathbf{u} \in \mathcal{D}\}$
$\tilde{\mathbf{s}} := (x, y)/\sqrt{2}$ for the aforementioned region $\mathcal{D}$.

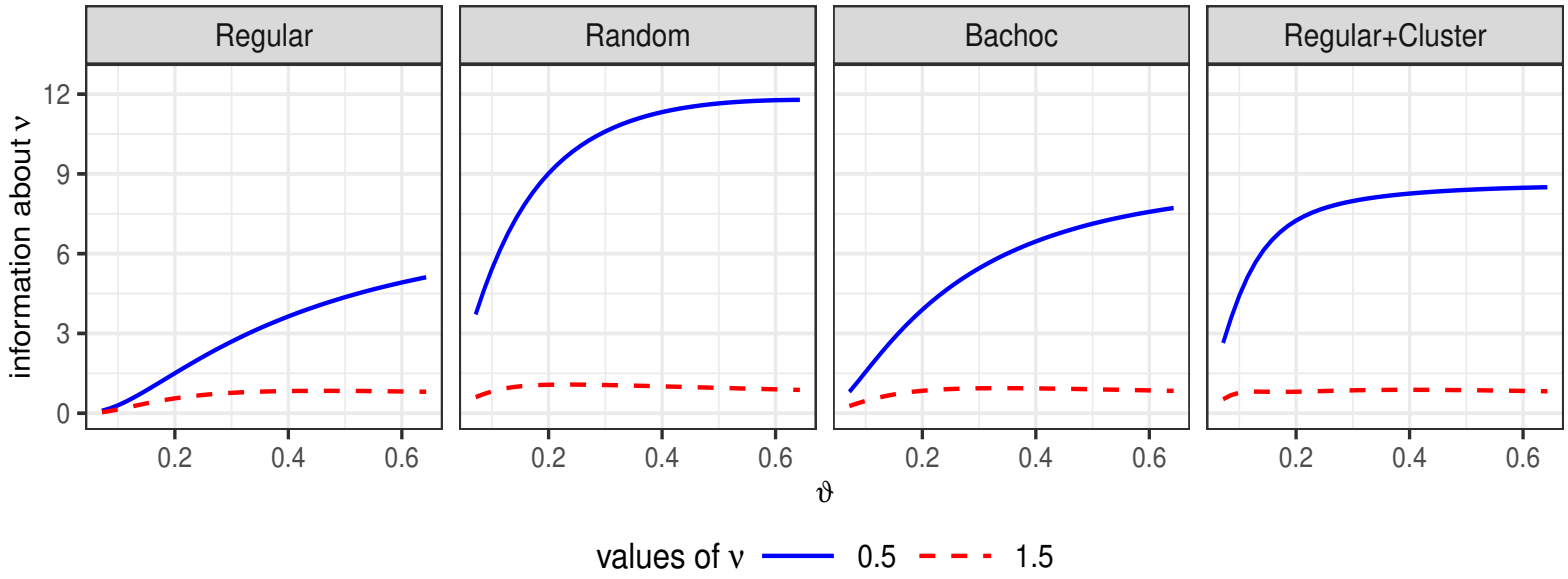
Purpose of re–scaling is to compare more sensibly information about the range and smoothness parameters, so the former may serve as a reference to judge when the latter is substantial

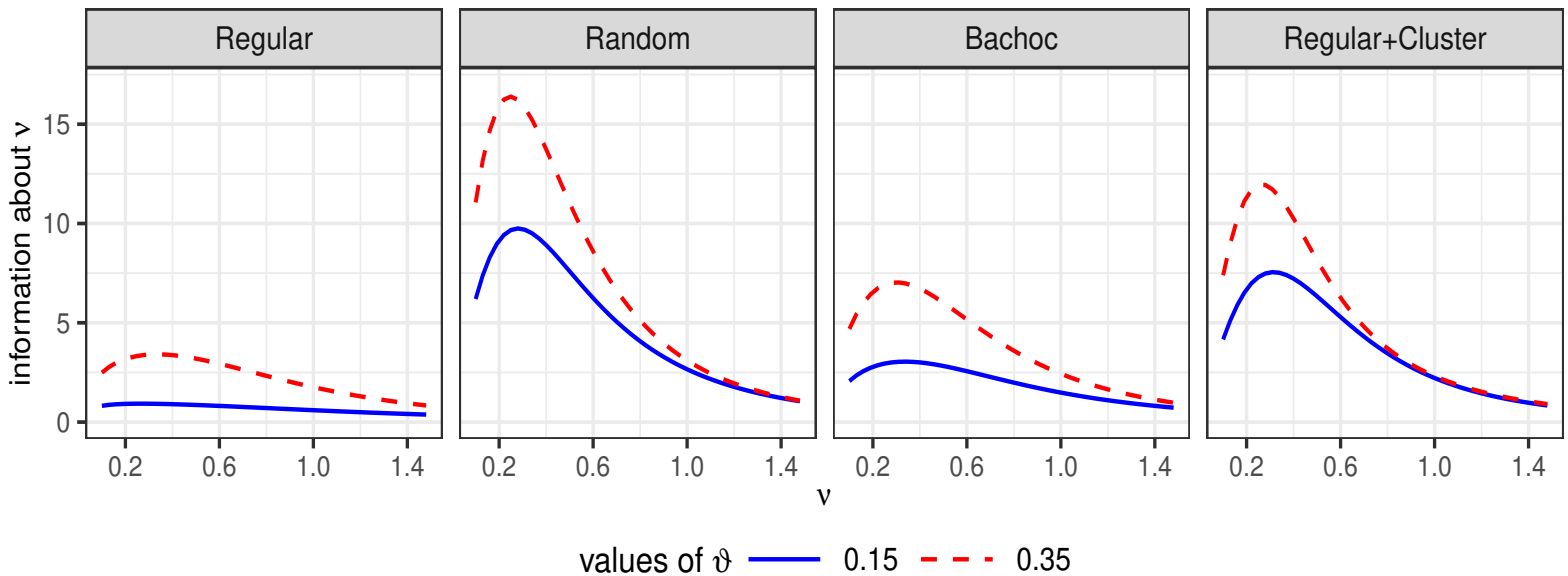
Information About the Range Parameter



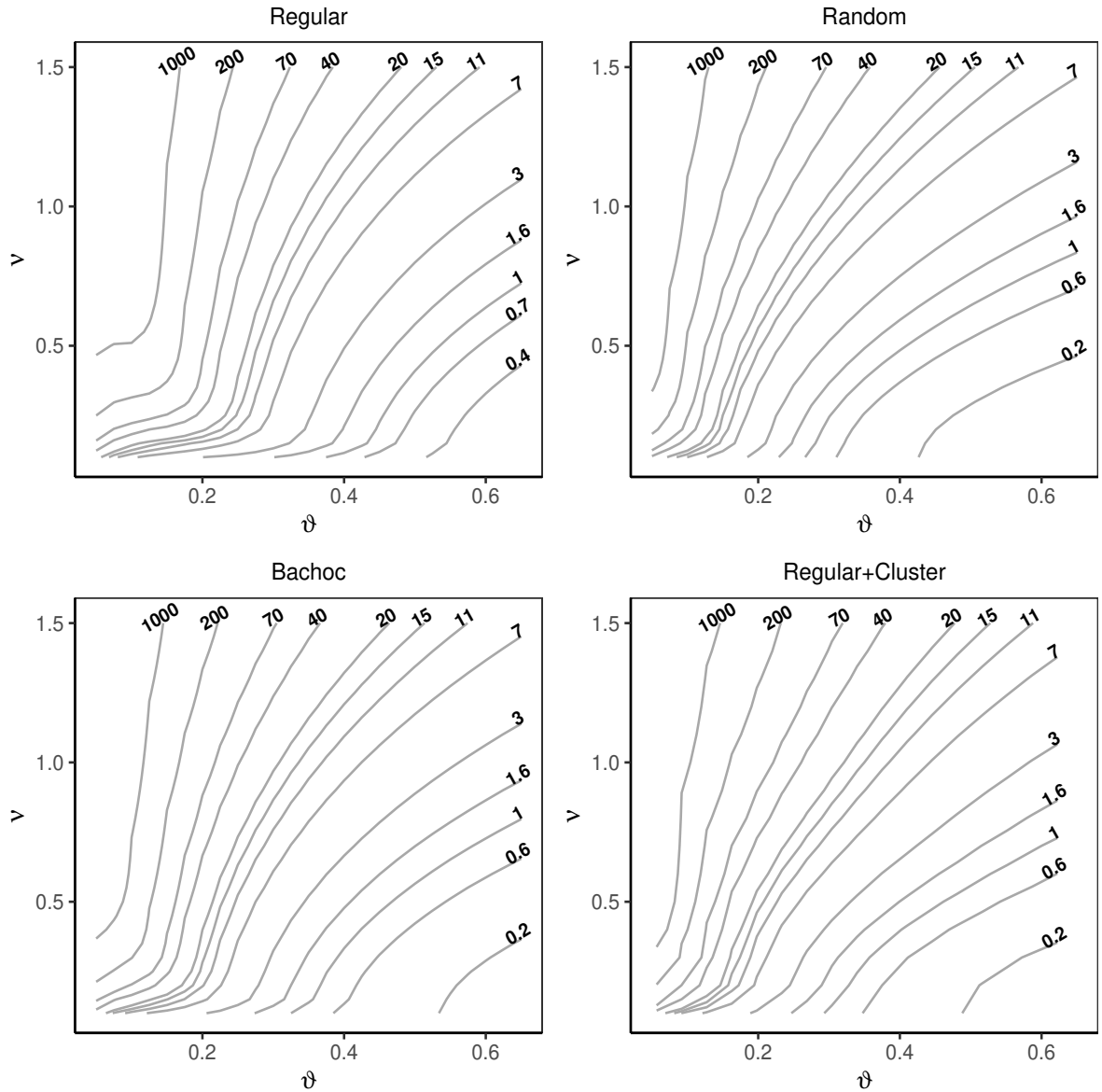


Information About the Smoothness Parameter





Information About the Range Parameter Relative to the Smoothness Parameter



For each design the information about ν is larger than that about ϑ for models in the 'south–east' corner of the correlation parameter space.

These are models that combine a large range parameter (strongly dependent process) and a small smoothness parameter (non–smooth process)

The opposite holds in the rest of the space.

Variation of Information with Sample Size

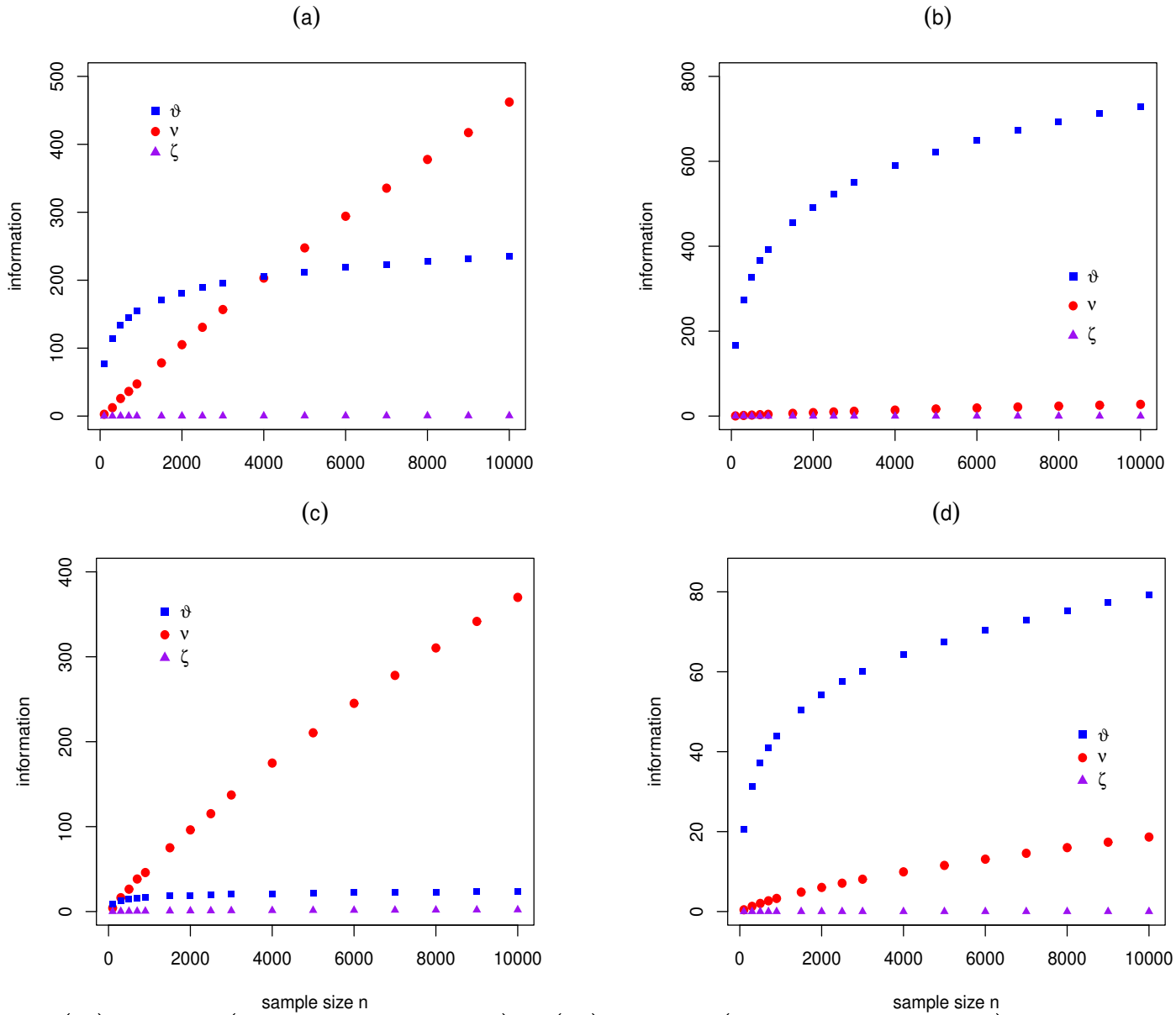


Figure: (a) $\boldsymbol{\eta} = (1, 0.2, 0.2, 0.5)$, (b) $\boldsymbol{\eta} = (1, 0.2, 0.2, 1.5)$, (c) $\boldsymbol{\eta} = (1, 0.2, 0.4, 0.5)$ and (d) $\boldsymbol{\eta} = (1, 0.2, 0.4, 1.5)$.

Example (Continuation)

For Swiss rainfall data, Fisher information matrix evaluated at the MLE $\hat{\boldsymbol{\eta}} = (\hat{\sigma}^2, \hat{\tau}^2, \hat{\vartheta}, \hat{\nu}) = (105.09, 6.74, 73.42, 0.95)$ is

$$I(\hat{\boldsymbol{\eta}}, \mathcal{S}_n)^{-1} = \begin{pmatrix} 1486.721 & -5.233 & 2.156 & -2.313 \\ -5.233 & 1.273 & -0.031 & 0.209 \\ 2.156 & -0.031 & 0.005 & -0.013 \\ -2.313 & 0.209 & -0.013 & 0.064 \end{pmatrix}$$

The (estimated) information about covariance parameters is

$$\mathrm{Inf}(\hat{\boldsymbol{\eta}}, \mathcal{S}_n) = (0.001, 0.786, 199.297, 15.643)$$

This confirms these data have substantial information about the smoothness parameter

Inference About Smoothness

Several methods have been proposed to estimate smoothness:

- Semiparametric methods based on the periodogram of certain linear combinations of data observed on a regular grid; Wu et al. (2013) and Wu and Lim (2016)
- Methods based on higher–order quadratic variations of data observed on some random designs; Loh (2015) and Loh et al. (2021)
- Other methods were proposed by Im et al. (2007) and Anderes and Stein (2008)

These methods are rarely used in practice, and are not implemented in common geostatistical software

A Slight Reparametrization

For likelihood analysis of it is convenient to parametrize the covariance function of the data in terms of the so–called noise–to–signal ratio

$$\xi := \frac{\tau^2}{\sigma^2}$$

rather than the nugget τ^2

From now on the covariance parameters are

$$(\sigma^2, \xi, \vartheta, \nu)$$

MLE of Smoothness

A method implemented in several public software is that of maximum likelihood (MLE), e.g., in the R packages `geoR`, `georob` and `ExaGeoStat`. But even this method has not been used or explored much

Because smoothness parameter of Matérn family is fixed in most applications, sampling properties of MLE of ν are largely unknown

We carry out a small simulation study to explore these

Simulation Set Up

- Two designs in the study region $\mathcal{D} = [0,1]^2$ of size $n = 225$
  A 15×15 regular grid and an irregular random design
- Models with $p = 1$, $\mu(\mathbf{s}) = 0$ and Matérn covariance function with $\sigma^2 = 1$
- Correlation parameters: ϑ is 0.1 or 0.5 and ν is 0.5 or 1.5

For each combination of design and model, simulate 1000 independent data sets under the three scenarios for the measurement error:

S1: $\xi = 0$ (no measurement error) assumed known; the parameters to be estimated are $(\mu, \sigma^2, \vartheta, \nu)$

S2: $\xi = 0.2$ assumed known

S3: $\xi = 0.2$ assumed unknown; the parameters to be estimated are $(\mu, \sigma^2, \xi, \vartheta, \nu)$

- Estimation was carried out using the `optim` function with `L-BFGS-B` algorithm. The search space is $\nu \in (0, 50)$ and other parameters were unrestricted
- Sampling distribution of MLE $\hat{\nu}_{\text{MLE}}$ is highly asymmetric Sampling features to be estimated are

$$\text{median}(\hat{\nu}_{\text{MLE}}),\ E(\hat{\nu}_{\text{MLE}}) \text{ and } P(\hat{\nu}_{\text{MLE}} \geq 50)$$

Results: Regular Design

		$\nu = 0.5$		$\nu = 1.5$	
		$\vartheta = 0.1$	$\vartheta = 0.5$	$\vartheta = 0.1$	$\vartheta = 0.5$
Scenario I	Median	0.597	0.541	1.686	1.550
	Mean	1.752	0.561	2.914	1.562
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	1.6	0	1.1	0
Scenario II	Median	0.622	0.530	1.852	1.615
	Mean	3.111	0.972	6.208	6.508
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	4.1	0.6	6.8	8.8
Scenario III	Median	6.239	0.961	7.846	2.359
	Mean	23.210	7.352	23.220	12.065
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	42.8	11.6	40.5	18.5

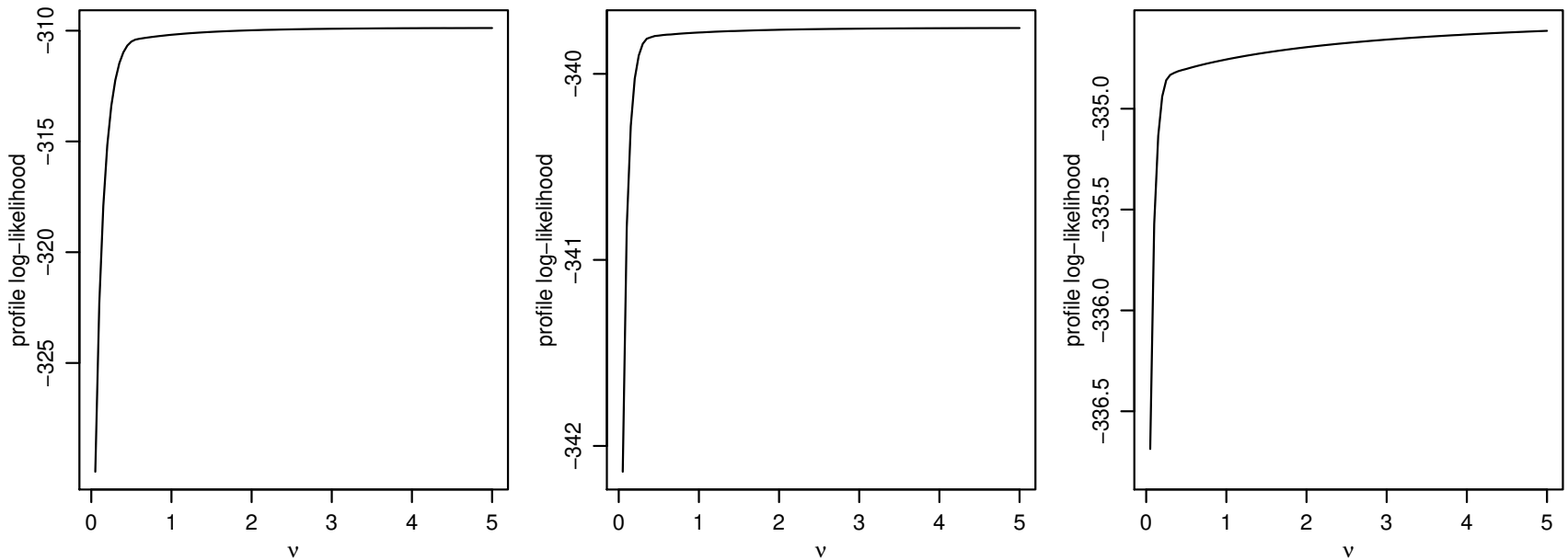
Results: Irregular Design

		$\nu = 0.5$		$\nu = 1.5$	
		$\vartheta = 0.1$	$\vartheta = 0.5$	$\vartheta = 0.1$	$\vartheta = 0.5$
Scenario I	Median	0.528	0.524	1.572	1.527
	Mean	0.571	0.549	1.654	1.539
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	0	0	0	0
Scenario II	Median	0.555	0.505	1.653	1.727
	Mean	0.778	0.774	4.883	7.168
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	0.3	0.3	4.6	10.1
Scenario III	Median	0.749	0.704	1.939	2.476
	Mean	6.382	3.384	9.252	13.644
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	10.2	4.6	12.8	21.9

Take Home Messages

- The MLE tends to overestimate ν
- The magnitude of the bias is small when the data do not contain measurement error, but large otherwise
- In scenario 3, $\hat{\nu}_{\text{MLE}}$ is quite unreliable, especially when the strength of correlation is weak
- The bias is more severe for regular sampling designs
- $P(\hat{\nu}_{\text{MLE}} \geq 50)$ is substantial, suggesting that $\hat{\nu}_{\text{MLE}}$ could not only be extremely large, but may not even exist for some data sets

To dwell on last point, inspected profile log–likelihoods of ν corresponding to data sets simulated on the regular design under Scenario 3, $(\mu, \sigma^2, \xi, \vartheta, \nu) = (0, 1, 0.2, 0.1, 0.5)$, and for which the iterative algorithm indicated that $\hat{\nu}_{\text{MLE}} \geq 50$



A Sub–class of Gaussian Matérn Random Fields

From now range ϑ and noise–to–signal ratio ξ assumed known

So $\boldsymbol{\vartheta} = \nu$, $\Theta = (0, \infty)$, and $\boldsymbol{\theta} = (\sigma^2, \nu)$ are only unknown covariance parameters

The case when all covariance parameters are unknown is currently being studied

Bayesian Analysis

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{z})$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} |\boldsymbol{\Psi}_\nu|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2} (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}_\nu^{-1} (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}) \Big)$$

$\boldsymbol{z} = (z_1, \ldots, z_n)^\top$
$\boldsymbol{X}$ is he known $n \times p$ matrix with entries $\boldsymbol{X}_{ij} = f_j(\mathbf{s}_i)$

$$\boldsymbol{\Psi}_\nu = \boldsymbol{\Sigma}_\nu + \xi \boldsymbol{I}_n,$$

with $\boldsymbol{\Sigma}_\nu$ the $n \times n$ matrix with entries $(\boldsymbol{\Sigma}_\nu)_{ij} = K_\nu(\|\mathbf{s}_i - \mathbf{s}_j\|)$

If $\pi(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the prior distribution, then

$$\pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{z}) \propto L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{z}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\theta})$$

Challenge: It is very difficult to specify $\pi(\boldsymbol{\beta}, \boldsymbol{\vartheta})$ subjectively, and ad hoc methods may lead to improper or non–nonsensical posteriors

A Solution: Reference (Objective) Priors

- Reference Priors have been succesfully used for multitude of models. They may depend on an ordering of parameters by their importance
- Berger, De Oliveira and Sansó (2001) derived a prior and its properties for a class of Gaussian models
- Extensions to other models in Paulo (2005), De Oliveira (2007), Kazianka and Pilz (2012), Ren, Sun and He (2012), Kazianka (2013), Ren, Sun and Sahu (2013), Gu, Wang and Berger (2018), Gu (2019)
- All of these assume the smoothness parameter is known

Algorithm to Compute A Reference Prior

1. Consider $\boldsymbol{\theta}$ parameters of primary interest and $\boldsymbol{\beta}$ parameters of secondary interest
2. Factor prior as $\pi^{\mathrm{R}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \pi^{\mathrm{R}}(\boldsymbol{\beta} \mid \boldsymbol{\theta})\pi^{\mathrm{R}}(\boldsymbol{\theta})$
3. Compute $\pi^{\mathrm{R}}(\boldsymbol{\beta} \mid \boldsymbol{\theta})$ using Jeffreys prior, which is for this model $\pi^{\mathrm{R}}(\boldsymbol{\beta} \mid \boldsymbol{\theta}) \propto 1$
4. Compute $\pi^{\mathrm{R}}(\boldsymbol{\theta})$ using the Jeffreys prior based on the 'marginal model' defined via the integrated likelihood of $\boldsymbol{\theta}$

$$L^{\mathrm{I}}(\boldsymbol{\theta}; \boldsymbol{z}) = \int_{\mathbb{R}^p} L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{z}) \pi^{\mathrm{R}}(\boldsymbol{\beta} \mid \boldsymbol{\theta}) d\boldsymbol{\beta}$$

For Gaussian random field with linear mean function:

$$L^{\mathrm{I}}(\boldsymbol{\theta}; \boldsymbol{z}) \propto (\sigma^2)^{-\frac{n-p}{2}} |\boldsymbol{\Psi}_\nu|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X}|^{-\frac{1}{2}} \exp\Big\{ - \frac{S_\nu^2}{2\sigma^2} \Big\}$$

$$S_\nu^2 = (\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\nu)^\top \boldsymbol{\Psi}_\nu^{-1} (\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\nu)$$
$$\hat{\boldsymbol{\beta}}_\nu = (\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{z}$$

The above expressions are general (valid for any $K_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{u})$)

Reference prior (BDS, 2001)

The *reference prior* of $(\boldsymbol{\beta}, \sigma^2, \nu)$ is

$$\pi^{\mathrm{R}}(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{R}}(\nu)}{\sigma^2}$$

with

$$\pi^{\mathrm{R}}(\nu) \propto \left\{ \mathrm{tr}\left[\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Psi}_\nu \right) \boldsymbol{Q}_\nu \right\}^2 \right] - \frac{1}{n-p} \left[\mathrm{tr}\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Psi}_\nu \right) \boldsymbol{Q}_\nu \right\} \right]^2 \right\}^{\frac{1}{2}}$$

where $\boldsymbol{Q}_\nu = \boldsymbol{\Psi}_\nu^{-1} - \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1}$

The *reference posterior* distribution is proper if

$$\int_{\mathbb{R}^p \times (0,\infty)^2} L(\boldsymbol{\beta}, \sigma^2, \nu; \boldsymbol{z}) \frac{\pi^{\mathrm{R}}(\nu)}{\sigma^2} d\boldsymbol{\beta} d\sigma^2 d\nu \;=\; \int_{(0,\infty)} L^{\mathrm{I}}(\nu; \boldsymbol{z}) \pi^{\mathrm{R}}(\nu) d\nu$$

is finite, where $L^{\mathrm{I}}(\nu; \boldsymbol{z})$ is the *integrated likelihood* of ν given as

$$L^{\mathrm{I}}(\nu; \boldsymbol{z}) \propto |\boldsymbol{\Psi}_\nu|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X}|^{-\frac{1}{2}} (S_\nu^2)^{-\frac{n-p}{2}}$$

Computational & Theoretical Challenges

In spite of theoretical good properties, reference priors are seldom used in practice for analysis of geostatistical or computer experiments data, because:

- Evaluation of $\pi^{\mathrm{R}}(\nu)$ requires computation of $\boldsymbol{\Psi}_{\nu}^{-1}$ and $(\boldsymbol{X}^{\top}\boldsymbol{\Psi}_{\nu}^{-1}\boldsymbol{X})^{-1}$, which involve $O(n^3)$ operations
- Matrix $\boldsymbol{\Psi}_{\nu}$ is often nearly singular when ϑ or ν are large and ξ is small, so computation of $\boldsymbol{\Psi}_{\nu}^{-1}$ will be unstable or infeasible
- Except for certain special cases, computation of $(\partial/\partial\nu)\boldsymbol{\Psi}_{\nu}$ involves multiple evaluations of Bessel function $\mathcal{K}_{\nu}(x)$
- It is currently unknown whether $\pi^{\mathrm{R}}(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{z})$ is proper

Proposed Solution

Find approximation to integrated likelihood $L^{\mathrm{I}}(\boldsymbol{\theta}; \boldsymbol{z})$ that is more amenable for analysis and computation, and use ot to obtain approximation to $\pi^{\mathrm{R}}(\nu)$

Such approximation should be devoid of large matrices and, if possible, not requiring evaluation of special functions

Spectral Approximation in $\mathbb{R}^2$

This approximation is based on the spectral representation of stationary random fields. Below I summarize the approximation for random fields in the plane ($d = 2$)

The approximation relies on the spectral density function of the random field. For the Matérn family this is $\sigma^2 f_\nu(\boldsymbol{\omega})$, where

$$f_\nu(\boldsymbol{\omega}) = \frac{\nu(4\nu)^\nu}{\pi\vartheta^{2\nu}}\left(\|\boldsymbol{\omega}\|^2 + \frac{4\nu}{\vartheta^2}\right)^{-(\nu+1)}, \quad \boldsymbol{\omega} = (\omega_1, \omega_2)^\top \in \mathbb{R}^2$$

For M_1, M_2 positive even integers and $\Delta > 0$, define regular rectangular grid in the plane

$$\begin{aligned}\mathcal{U}_M &= \{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \ldots, \mathbf{u}_{M_1,M_2}\} \\ &= \{\Delta, \ldots, \Delta M_1\} \times \{\Delta, \ldots, \Delta M_2\}\end{aligned}$$

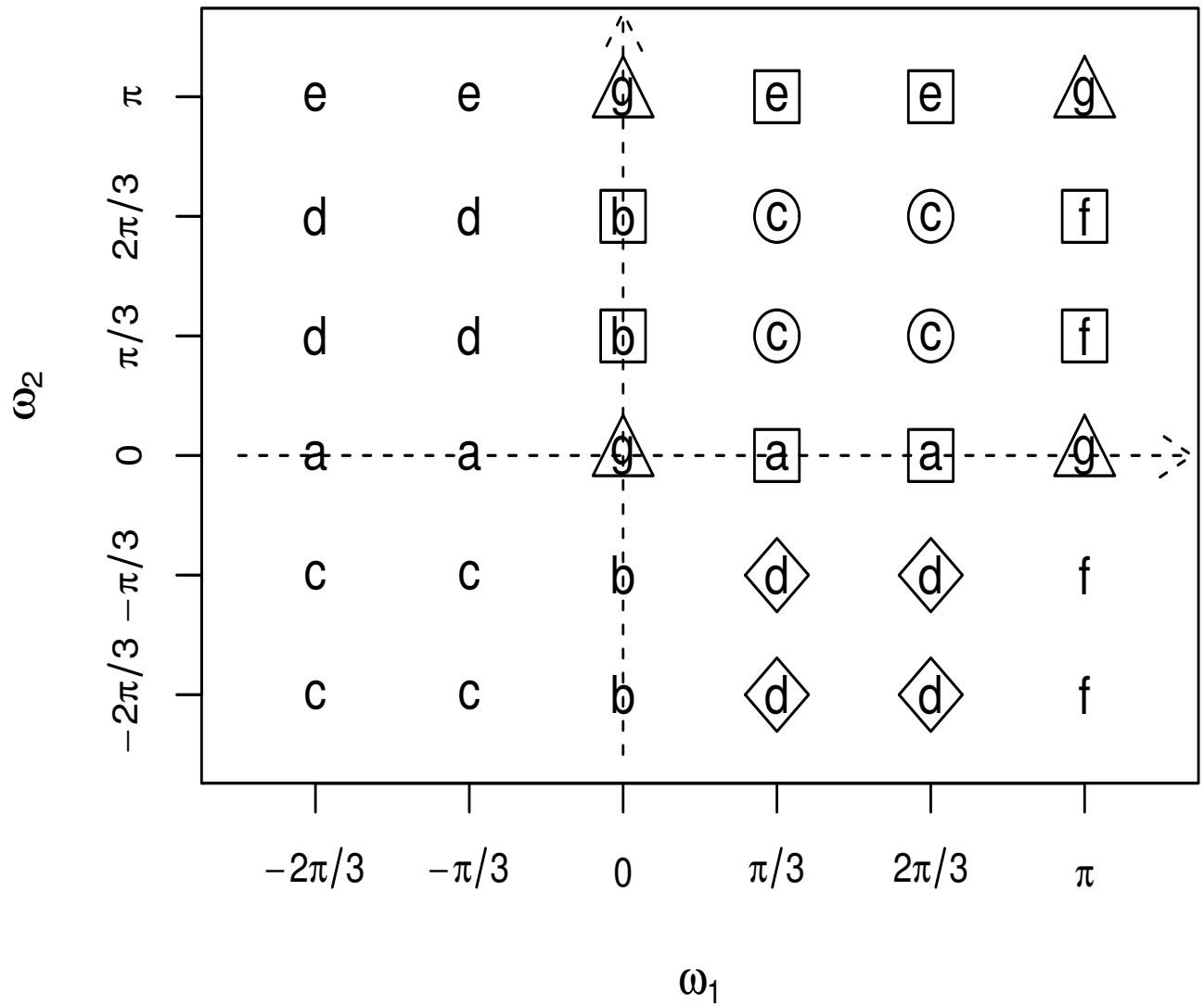
Let $M := M_1 M_2$. Call $\mathcal{U}_M \subset \mathbb{R}^2$ the spatial design

Associated with the above, define another regular rectangular grid in the plane

$$\begin{aligned}\mathcal{W}_M &= \left\{\boldsymbol{\omega}_{-\frac{M_1}{2}+1,-\frac{M_2}{2}+1}, \ldots, \boldsymbol{\omega}_{0,0}, \ldots, \boldsymbol{\omega}_{\frac{M_1}{2},\frac{M_2}{2}}\right\} \\ &= \frac{2\pi}{\Delta M_1}\left\{-\frac{M_1}{2}+1, \ldots, 0, 1, \ldots, \frac{M_1}{2}\right\} \times \frac{2\pi}{\Delta M_2}\left\{-\frac{M_2}{2}+1, \ldots, 0, 1, \ldots, \frac{M_2}{2}\right\}\end{aligned}$$

Call $\mathcal{W}_M \subset [-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]^2$ the spectral design

Illustration: $M_1 = M_2 = 6$ and $\Delta = 1$



Let $Z_\Delta(\mathbf{k}) := Z(\Delta\mathbf{k})$, $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$, be the discrete index random field defined by sampling the random field $Z(\cdot)$

$$
\begin{aligned}
E\{Z_\Delta(\mathbf{k})\} &= \mu(\Delta\mathbf{k}) \\
\operatorname{cov}\{Z_\Delta(\mathbf{k}), Z_\Delta(\mathbf{k}')\} &= \sigma^2 K_\nu(\Delta||\mathbf{k} - \mathbf{k}'||), \quad \mathbf{k}, \mathbf{k}' \in \mathbb{Z}^2 \\
f_\nu^\Delta(\boldsymbol{\omega}) &= \sum_{\boldsymbol{l} \in \mathbb{Z}^2} f_\nu\Big(\boldsymbol{\omega} + \frac{2\pi \boldsymbol{l}}{\Delta}\Big), \quad \boldsymbol{\omega} \in \Big[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}\Big]^2
\end{aligned}
$$

From spectral representation of $Z_\Delta(\cdot)$: for $\mathbf{u}_{i,j} = \Delta(i,j)^\top \in \mathcal{U}_M$

$$
\begin{aligned}
Z_\Delta\{(i,j)\} &= \mu(\mathbf{u}_{i,j}) + \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \exp(\mathrm{i}\boldsymbol{\omega}^\top \mathbf{u}_{i,j}) U_\Delta(d\boldsymbol{\omega}) \\
&= Z(\mathbf{u}_{i,j})
\end{aligned}
$$

$\mathrm{i} = \sqrt{-1}$ and $U_\Delta(\cdot)$ is zero–mean complex random orthogonal measure in the plane (Yaglom, 1987)

Spectral representation provides the basis to approximate the distribution of $\left(Z(\mathbf{u}_{i,j}) : \mathbf{u}_{i,j} \in \mathcal{U}_M\right)^\top$
It requires the following sets of indices:

$$I_C = \left\{ \left(0,0\right), \left(\frac{M_1}{2}, 0\right), \left(0, \frac{M_2}{2}\right), \left(\frac{M_1}{2}, \frac{M_2}{2}\right) \right\} \quad \text{('corner' frequencies)}$$

$$I_B = \left\{ \left(m_1, 0\right), \left(0, m_2\right), \left(m_1, \frac{M_2}{2}\right), \left(\frac{M_1}{2}, m_2\right) : m_1 = 1, \ldots, \frac{M_1}{2} - 1; \; m_2 = 1, \ldots, \frac{M_2}{2} - 1 \right\}$$

('boundary' frequencies)

$$I_I = \left\{ \left(m_1, m_2\right) : m_1 = 1, \ldots, \frac{M_1}{2} - 1; m_2 = 1, \ldots, \frac{M_2}{2} - 1 \right\} \quad \text{('interior' frequencies)}$$

$$I_E = \left\{ \left(m_1, m_2\right) : m_1 = 1, \ldots, \frac{M_1}{2} - 1; m_2 = -\frac{M_2}{2} + 1, \ldots, -1 \right\} \quad \text{('exterior' frequencies)}$$

and let $I := I_B \cup I_I \cup I_E$, which has $M/2 - 2$ elements

Lemma

For any $\mathbf{u}_{i,j} \in \mathcal{U}_M$ define the random object

$$T_{M_1,M_2}(\mathbf{u}_{i,j}) := \sum_{m_1=-\frac{M_1}{2}+1}^{\frac{M_1}{2}} \sum_{m_2=-\frac{M_2}{2}+1}^{\frac{M_2}{2}} \exp(\mathrm{i}\boldsymbol{\omega}_{m_1,m_2}^{\top}\mathbf{u}_{i,j})U_{m_1,m_2}$$

where $U_{m_1,m_2} = A_{m_1,m_2} + \mathrm{i}B_{m_1,m_2}$ be complex random variables satisfying:

(1) $B_{m_1,m_2} = 0$ for $(m_1, m_2) \in I_C$
(2) $U_{0,-m_2} = \bar{U}_{0,m_2}$, $U_{-m_1,0} = \bar{U}_{m_1,0}$, $U_{\frac{M_1}{2},-m_2} = \bar{U}_{\frac{M_1}{2},m_2}$ and $U_{-m_1,\frac{M_2}{2}} = \bar{U}_{m_1,\frac{M_2}{2}}$ for $(m_1, m_2) \in I_B$
(3) $U_{-m_1,-m_2} = \bar{U}_{m_1,m_2}$ for $(m_1, m_2) \in I_I \cup I_E$
(4) For $(m_1, m_2) \in I_C \cup I$, A_{m_1,m_2} and B_{m_1,m_2} are independent Gaussian variables with means 0 and variances

$$\mathrm{var}(A_{m_1,m_2}) = \frac{c_\Delta \sigma^2}{M} f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) \qquad \text{if } (m_1, m_2) \in I_C$$

$$\mathrm{var}(A_{m_1,m_2}) = \mathrm{var}(B_{m_1,m_2}) = \frac{c_\Delta \sigma^2}{2M} f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) \qquad \text{if } (m_1, m_2) \in I,$$

where $c_\Delta := (2\pi/\Delta)^2$

Then

(a) $T_{M_1,M_2}(\mathbf{u}_{i,j})$

$$= A_{0,0} + A_{\frac{M_1}{2},0} \cos\left(\boldsymbol{\omega}_{\frac{M_1}{2},0}^\top \mathbf{u}_{i,j}\right) + A_{0,\frac{M_2}{2}} \cos\left(\boldsymbol{\omega}_{0,\frac{M_2}{2}}^\top \mathbf{u}_{i,j}\right) + A_{\frac{M_1}{2},\frac{M_2}{2}} \cos\left(\boldsymbol{\omega}_{\frac{M_1}{2},\frac{M_2}{2}}^\top \mathbf{u}_{i,j}\right)$$
$$+ \ 2 \sum_{(m_1,m_2)\in I} \left(A_{m_1,m_2} \cos(\boldsymbol{\omega}_{m_1,m_2}^\top \mathbf{u}_{i,j}) - B_{m_1,m_2} \sin(\boldsymbol{\omega}_{m_1,m_2}^\top \mathbf{u}_{i,j})\right)$$

and $(T_{M_1,M_2}(\mathbf{u}_{i,j}) : \mathbf{u}_{i,j} \in \mathcal{U}_M)^\top$ has a zero–mean real multivariate normal distribution.

(b) For any $\mathbf{u}_{i,j}, \mathbf{u}_{i',j'} \in \mathcal{U}_M$ it holds that as $\min\{M_1, M_2\} \to \infty$

$$\mathrm{cov}\{T_{M_1,M_2}(\mathbf{u}_{i,j}), T_{M_1,M_2}(\mathbf{u}_{i',j'})\} \to \sigma^2 K_\nu(||\mathbf{u}_{i,j} - \mathbf{u}_{i',j'}||)$$

Let $\tilde{\boldsymbol{z}} := (Z(\mathbf{u}_{1,1}), \ldots, Z(\mathbf{u}_{M_1,M_2}))^\top$ and
$\boldsymbol{t} := (T_{M_1,M_2}(\mathbf{u}_{1,1}), \ldots, T_{M_1,M_2}(\mathbf{u}_{M_1,M_2}))^\top = \boldsymbol{H}_1\boldsymbol{g}$, with
$\boldsymbol{H}_1$ an $M \times M$ matrix whose columns are formed by multiples of cosines and sines, and $\boldsymbol{g} = \left(\ldots A_{m_1,m_2}, B_{m_1,m_2} \ldots\right)^\top$
From the lemma

$$\tilde{\boldsymbol{z}} \overset{\text{approx}}{\sim} \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{t} \qquad \text{as } \min\{M_1, M_2\} \to \infty$$

$\tilde{\boldsymbol{X}}$ is the $M \times p$ matrix whose entries involve the covariates measured at the locations in $\mathcal{U}_M$
When M_1 and M_2 are large

$$\tilde{\boldsymbol{z}} \overset{\text{approx}}{\sim} \mathrm{N}(\tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2(\boldsymbol{H}_1\boldsymbol{G}_\nu\boldsymbol{H}_1^\top + \xi I_M)) \qquad \star\star$$

where

$$\boldsymbol{G}_\nu = \frac{c_\Delta}{2M}\mathrm{diag}\Big(\left(2f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I_C\right)^\top, \left(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\right)^\top, \left(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\right)^\top\Big)$$

Comments

- The spatial design $\mathcal{U}_M$ may or may not be equal to the sample design $\mathcal{S}_n$
- When $\mathcal{U}_M \neq \mathcal{S}_n$, use of the spectral approximation is problematic to approximate the likelihood
  But it is fine to approximate the prior
- $\mathcal{U}_M$ is constructed in a way so that its convex hull contains the region of interest $\mathcal{D}$. Need to tune M_1, M_2 and Δ
- Bottom line: Instead of the $\mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Psi}_\nu)$ distribution, use the $\mathrm{N}\big(\tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2(\boldsymbol{H}_1\boldsymbol{G}_\nu\boldsymbol{H}_1^\top + \xi I_M)\big)$ distribution to approximate the reference prior

Approximate Reference Prior: General Mean

Theorem 1

The approximate reference prior of $(\boldsymbol{\beta}, \sigma^2, \nu)$ is $\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{AR}}(\nu)}{\sigma^2}$, with

$$\pi^{\mathrm{AR}}(\nu) \propto \left\{ \mathrm{tr}\left[\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Lambda}_\nu \right) \tilde{\boldsymbol{Q}}_\nu \right\}^2 \right] - \frac{1}{M-p} \left[\mathrm{tr}\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Lambda}_\nu \right) \tilde{\boldsymbol{Q}}_\nu \right\} \right]^2 \right\}^{\frac{1}{2}}$$

where $\tilde{\boldsymbol{Q}}_\nu := \boldsymbol{\Lambda}_\nu^{-1} - \boldsymbol{\Lambda}_\nu^{-1} \boldsymbol{X}_1 (\boldsymbol{X}_1^\top \boldsymbol{\Lambda}_\nu^{-1} \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \boldsymbol{\Lambda}_\nu^{-1}$ and

$$\boldsymbol{\Lambda}_\nu = c_\Delta \mathrm{diag}\Big(\left(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1, m_2) \in I_C \right)^\top, \ \left(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1, m_2) \in I \right)^\top, \\ \left(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1, m_2) \in I \right)^\top \Big)$$

Approximate Reference Prior: Constant Mean

Corollary

For model with constant mean β_1, the approximate reference prior of (β_1, σ^2, ν) is $\pi^{\mathrm{AR}}(\beta_1, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{AR}}(\nu)}{\sigma^2}$, with

$$\pi^{\mathrm{AR}}(\nu) \propto \left\{ \sum_{j=1}^{M-1} \gamma_\nu^2(\boldsymbol{\omega}_j) - \frac{1}{M-1} \left(\sum_{j=1}^{M-1} \gamma_\nu(\boldsymbol{\omega}_j) \right)^2 \right\}^{\frac{1}{2}}$$

with $\boldsymbol{\omega}_j$ is reindexing of $\boldsymbol{\omega}_{m_1,m_2}$, with $\boldsymbol{\omega}_{0,0}$ removed, and

$$\gamma_\nu(\boldsymbol{\omega}_j) := \frac{c_\Delta}{c_\Delta \tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j) + \xi} \left(\frac{\partial}{\partial \nu} \tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j) \right)$$

Theorem 2

For Matérn family, the marginal approximate reference prior $\pi^{\mathrm{AR}}(\nu)$ is continuous function that satisfies

$$\pi^{\mathrm{AR}}(\nu) = \begin{cases} O(1), & \text{as } \nu \to 0^{+} \\ O(\nu^{-2}), & \text{as } \nu \to \infty \end{cases}$$

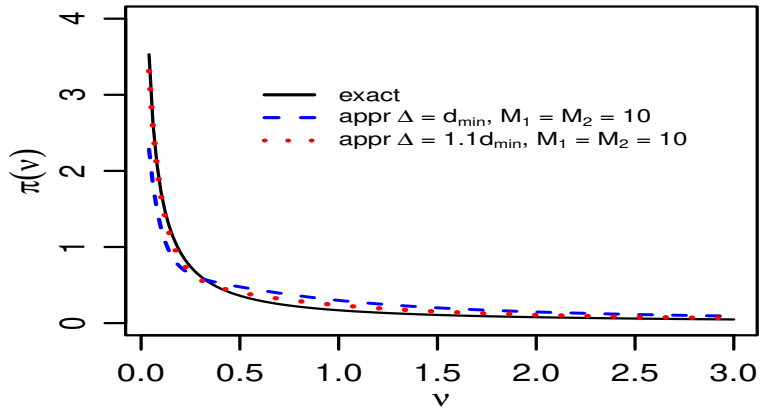
so it is proper.

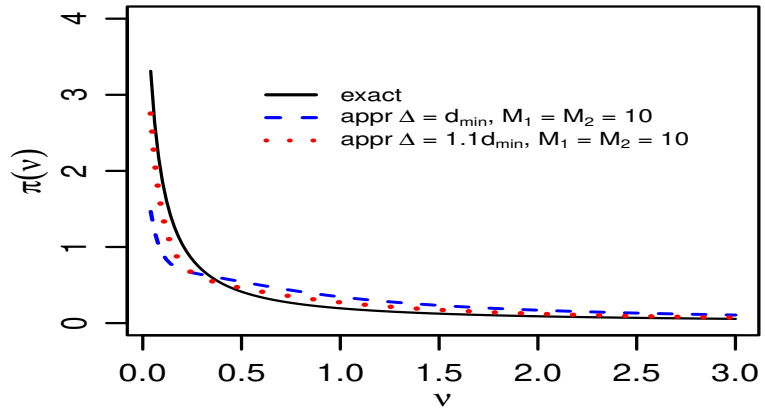
Furthermore, the approximate reference posterior based on the observed data, $\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \boldsymbol{z})$, is also proper

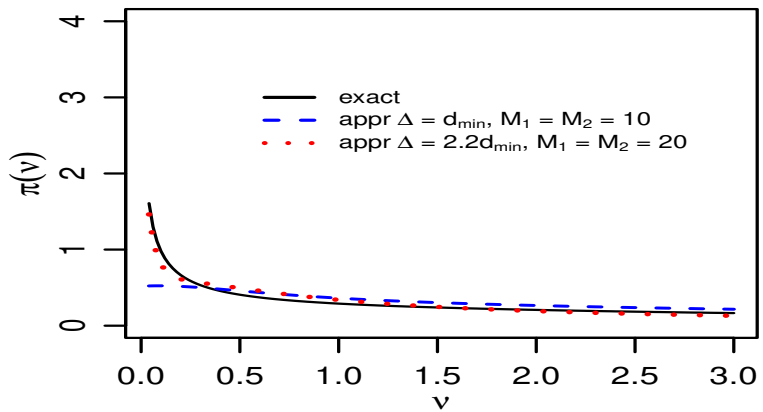
Comparison of Exact and Approximate Reference Priors

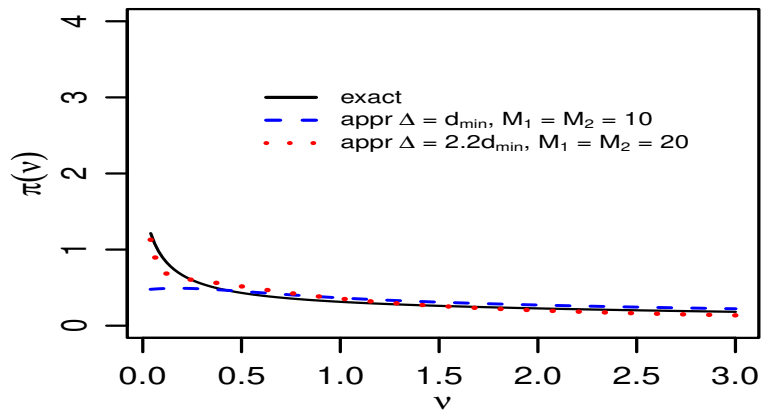
- $\mathcal{D} = [0,1]^2$, $n = 100$, a regular design and 3 irregular designs
- $\mu(\mathbf{s}) = 1$ and $\mu(\mathbf{s}) = 0.15 - 0.65x - 0.1y + 0.9x^2 - xy + 1.2y^2$
- Matérn model with $\sigma^2 = 1$ and $\vartheta = 0.1, 0.3$ and 0.5
- $\xi = 0$ and 0.5
- $\tilde{f}_\nu^\Delta(\boldsymbol{\omega})$ approximated by truncating its series so only terms with $\max\{|l_1|, |l_2|\} \leq 4$ are retained

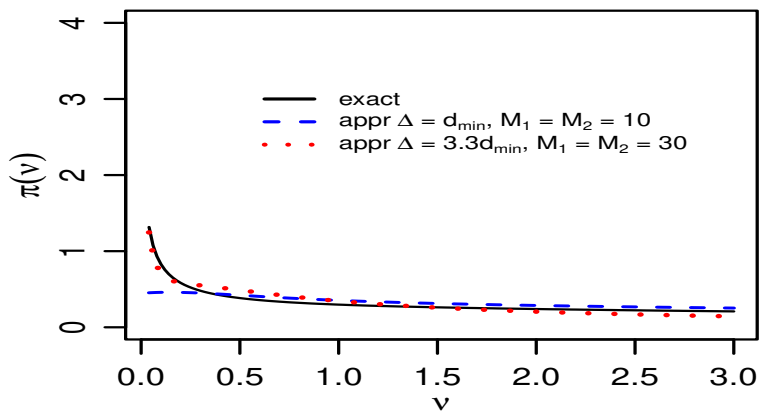
When $\xi = 0$

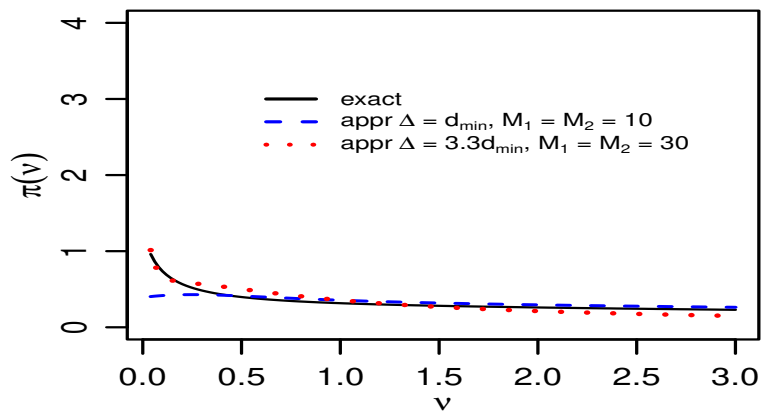

exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 1.1d_min, M_1 = M_2 = 10
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4


exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 1.1d_min, M_1 = M_2 = 10
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4

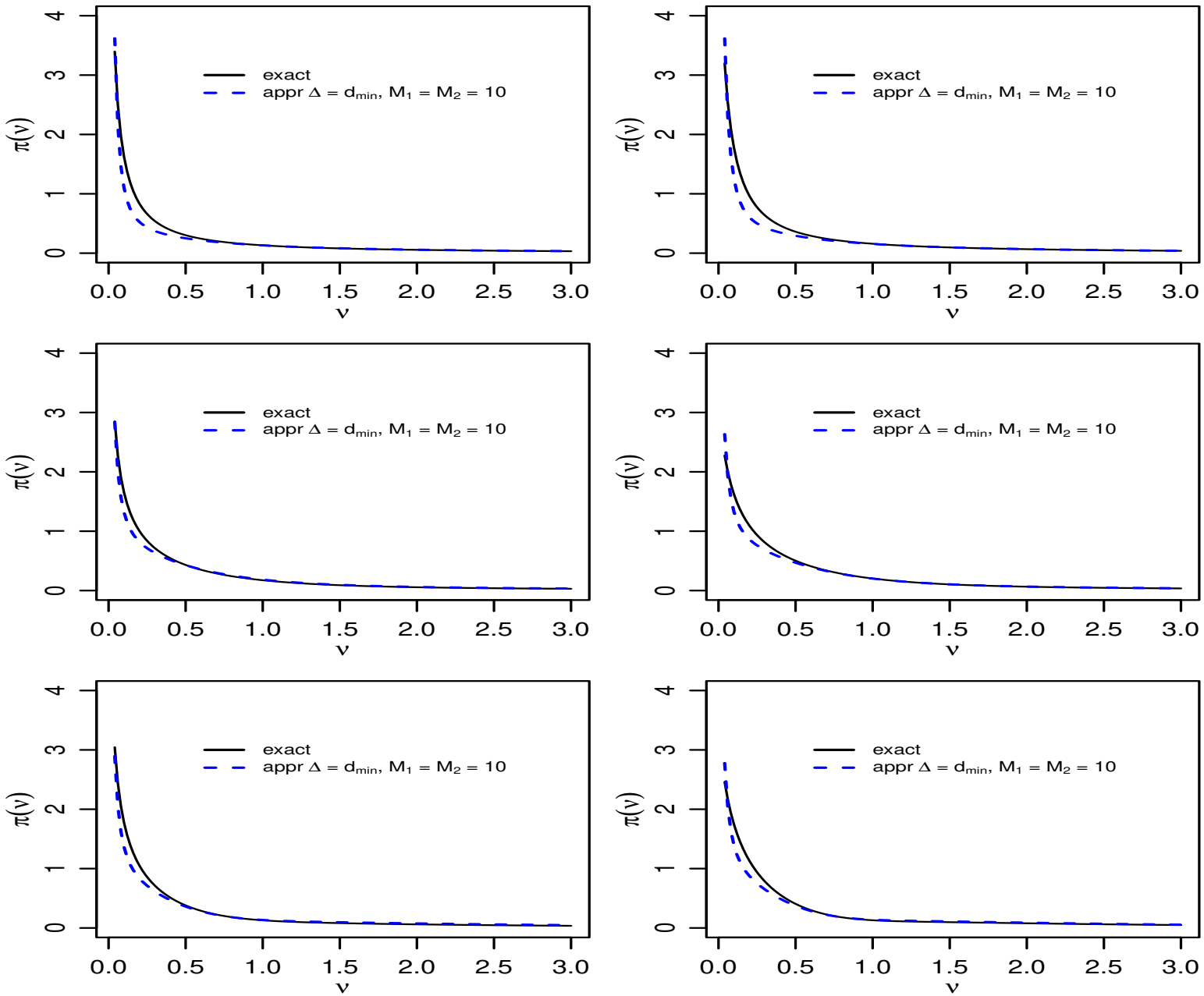

exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 2.2d_min, M_1 = M_2 = 20
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4


exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 2.2d_min, M_1 = M_2 = 20
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4

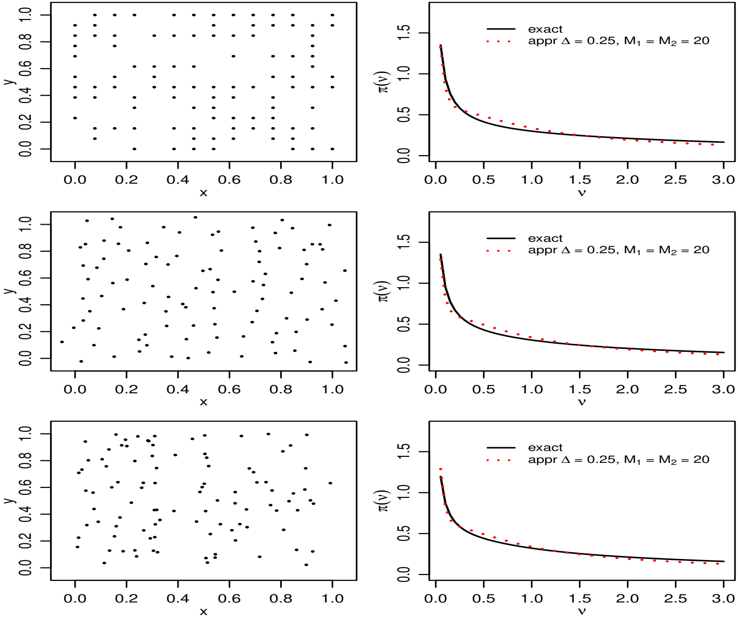

exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 3.3d_min, M_1 = M_2 = 30
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4


exact
appr Δ = d_min, M_1 = M_2 = 10
appr Δ = 3.3d_min, M_1 = M_2 = 30
π(ν)
ν
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4

When $\xi = 0.5$



Irregular Designs, $\mu = 1$, $\vartheta = 0.3$, $\xi = 0.5$



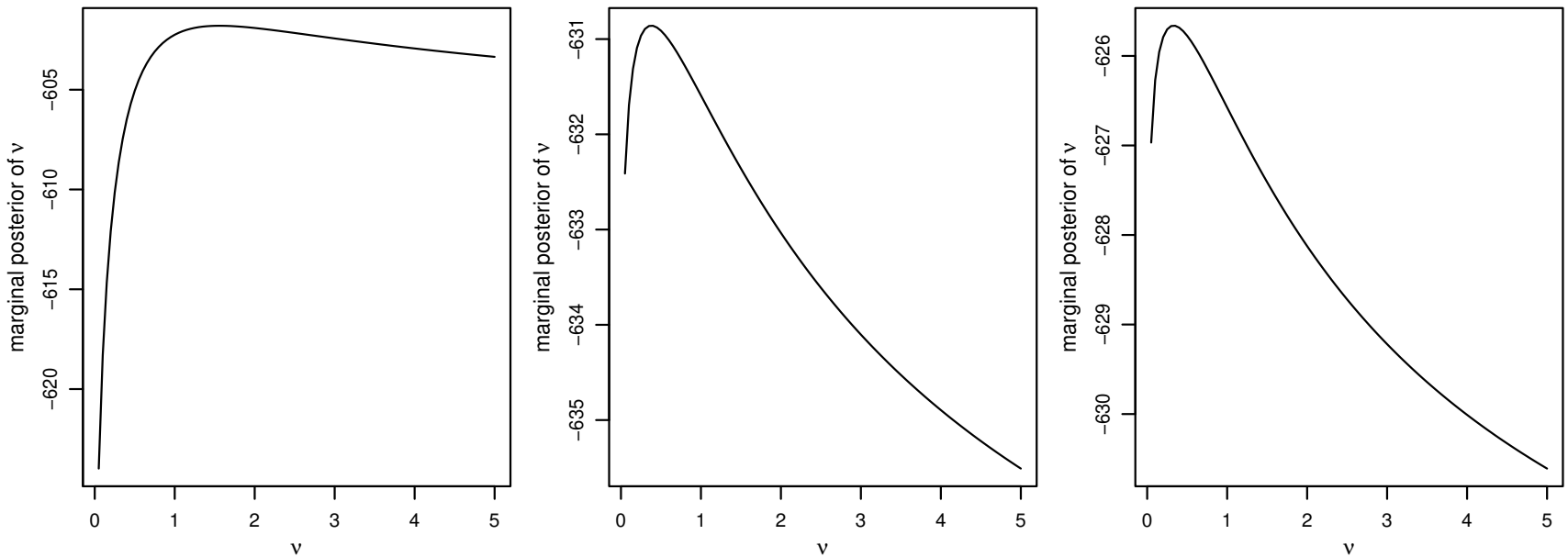
Comparison of Computational Effort

	p	Reference Prior	$n = 100$	$n = 400$	$n = 1600$	$n = 10000$
RD	1	Exact	0.82	14.54	596.17	–
		Approximate	0.02	0.07	0.22	1.82
	6	Exact	0.95	17.18	874.05	–
		Approximate	0.11	1.07	8.04	187.45
ID	1	Exact	33.28	521.01	6310.31	–
		Approximate	0.02	0.08	0.19	1.90
	6	Exact	37.71	628.11	10840.67	–
		Approximate	0.10	1.14	9.94	204.32

Frequentist Properties

- Bayesian credible intervals for σ^2 and ν based on $\pi^{\mathrm{R}}(\nu)$ and $\pi^{\mathrm{AR}}(\nu)$ are similar and have good frequentist coverage
- Handcock and Stein (1993): $\pi(\nu) = (1+\nu)^{-2}$
  Good inferences similar to those from $\pi^{\mathrm{AR}}(\nu)$
- Uniform: $\nu \sim \mathrm{unif}(0, L)$ for some $L > 0$ large
  Bad frequentist coverage and hugely inflated MAE
  (estimates are severely upward biased)

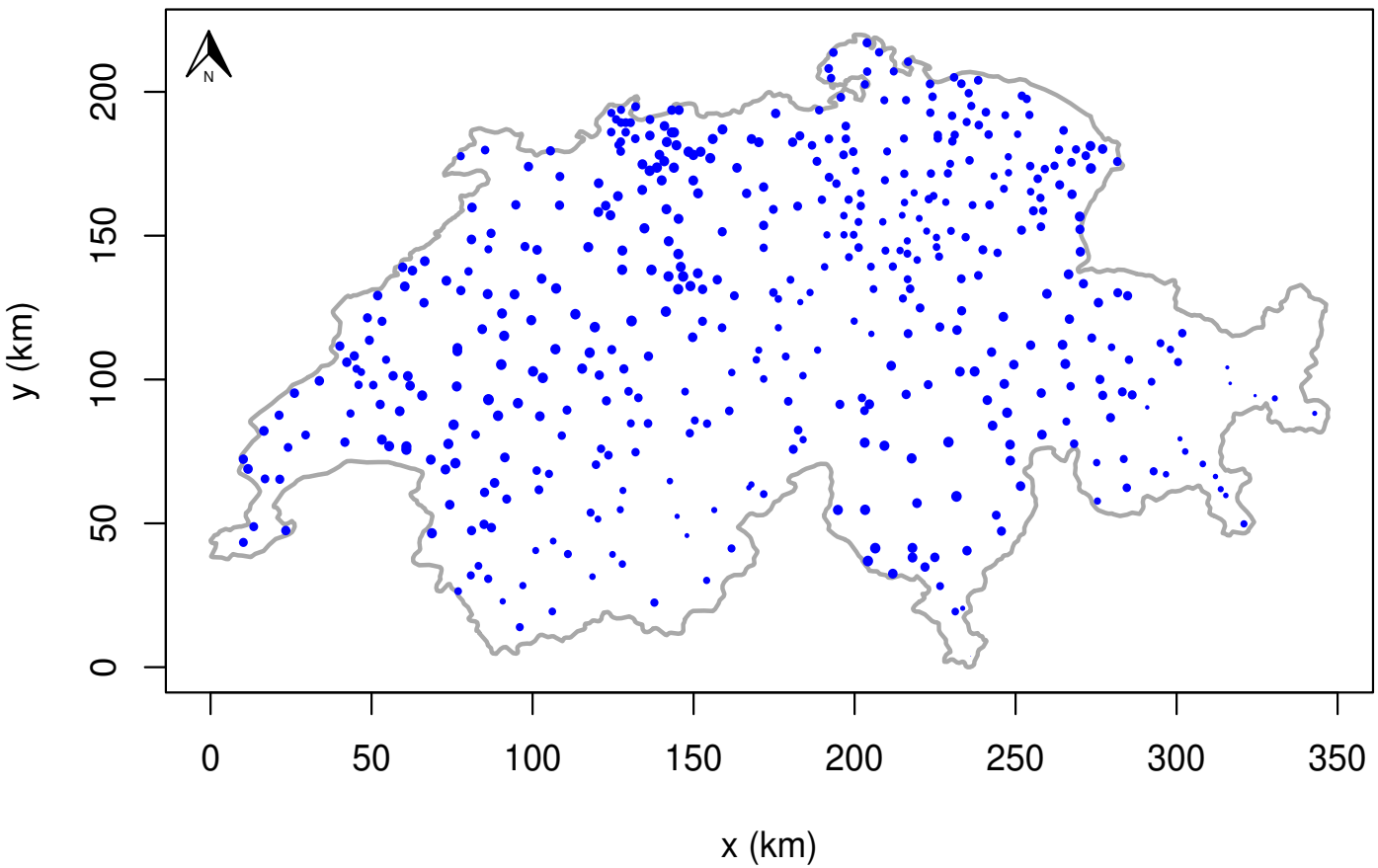
Bayes Better Than MLE



Example

Consider again 467 measurements of daily rainfall collected in Switzerland on May 8, 1986

Model: square root transformed data is Gaussian random field with constant mean and Matérn covariance function



Estimating ϑ and ν

The conditional reference prior is

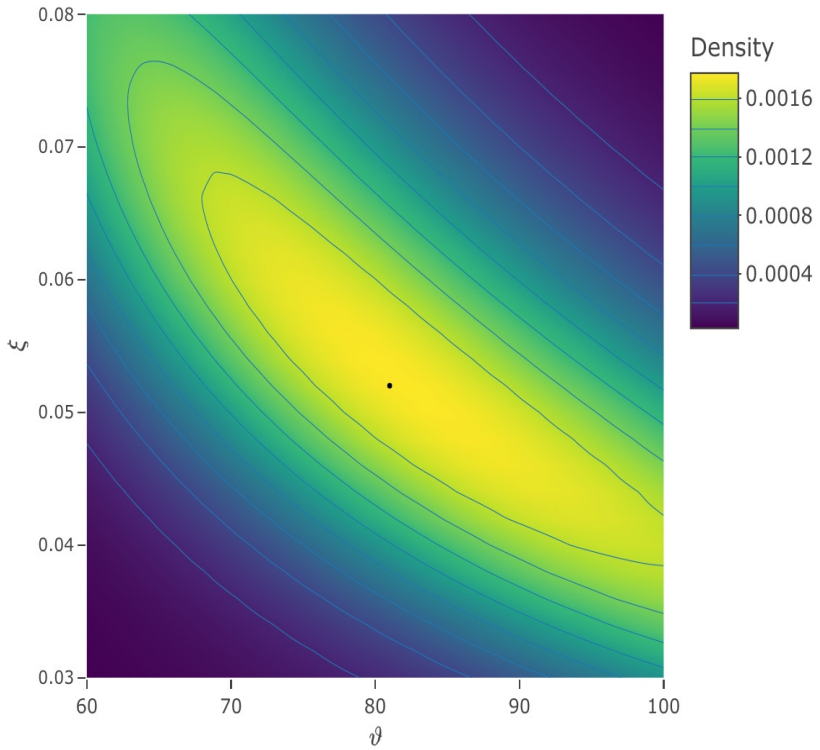
$$\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \vartheta, \xi) = \frac{C(\vartheta, \xi)\pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi)}{\sigma^2},$$

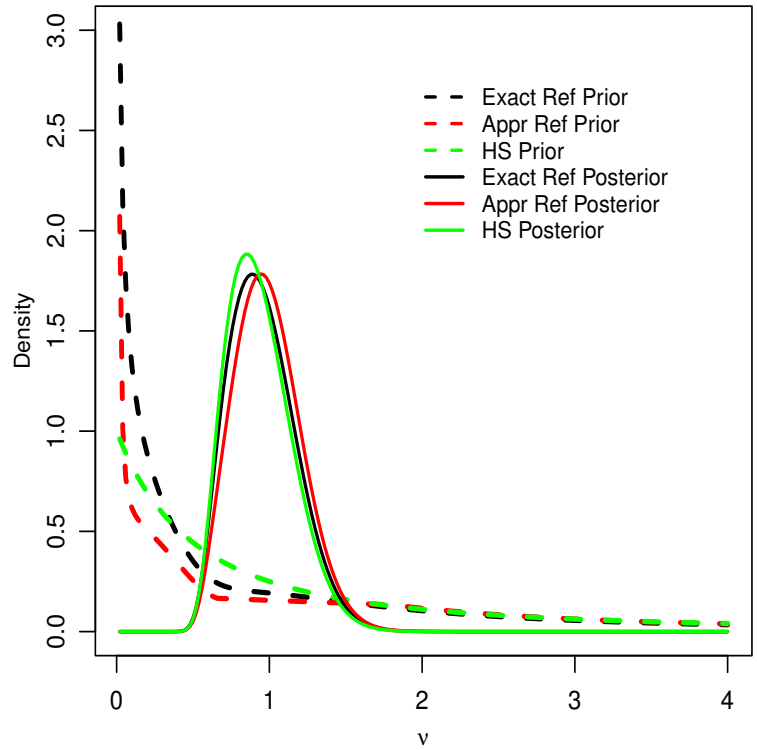
and integrated likelihood of (ϑ, ξ)

$$\begin{aligned} m(\boldsymbol{z} \mid \vartheta, \xi) &= \int_{\mathbb{R}^p \times (0,\infty)^2} L(\boldsymbol{\beta}, \sigma^2, \xi, \vartheta, \nu; \boldsymbol{z}) \pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \vartheta, \xi)\, d\boldsymbol{\beta} d\sigma^2 d\nu \\ &\propto \int_0^\infty |\boldsymbol{\Psi}_{\boldsymbol{\vartheta},\xi}|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_{\boldsymbol{\vartheta},\xi}^{-1} \boldsymbol{X}|^{-\frac{1}{2}} (S^2_{\boldsymbol{\vartheta},\xi})^{-\frac{n-p}{2}} C(\vartheta, \xi) \pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi) \end{aligned}$$

Choose

$$\begin{aligned} (\hat{\vartheta}, \hat{\xi}) &= \arg \max_{(\vartheta,\xi) \in (0,\infty)^2} m(\boldsymbol{z} \mid \vartheta, \xi) \\ &= (82, 0.052) \end{aligned}$$


Density
0.0016
0.0012
0.0008
0.0004
0.08
0.07
0.06
0.05
0.04
0.03
ξ
60
70
80
90
100
ϑ


Exact Ref Prior
Appr Ref Prior
HS Prior
Exact Ref Posterior
Appr Ref Posterior
HS Posterior
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
ν

Results

Prior	$\hat{\beta}_1$ (95% CI)	$\hat{\sigma}^2$ (95% CI)	$\hat{\nu}$ (95% CI)
R	19.790 (10.776, 28.535)	127.517 (79.190, 174.109)	0.888 (0.565, 1.390)
AR	19.779 (11.539, 29.224)	131.310 (82.594, 177.237)	0.946 (0.589, 1.424)
HS	19.828 (10.312, 28.172)	125.257 (78.541, 170.664)	0.851 (0.561, 1.346)

Computation times for 12000 draws from posterior:

HS: 1166 secs, AR: 1184 secs, R: 58208

Conclusions

- The claim geostatistical data have little or no information about smoothmess is not quite right
- Information about the range parameter ϑ displays little sensitivity to the sampling design
- Information about ϑ increases when ν increases. This information is largest for processes with weak correlation that are smooth
- Information about the smoothness parameter ν does display sensitivity to the sampling design. The Regular design is the least informative about ν, while the Random design is the most informative
- Overall and regardless of the design, the information about the smoothness parameter ν is substantial for processes with strong correlation that are non–smooth
- MLE of ν is severely (upward) biased when data contain measurement error
- A Bayesian approach with a 'good' prior provides better inferences
- The approximate reference prior provides much better inferences, and it is computationally feasible