Interaction Modelling of inhomogeneous Poisson processes by means of log-Gaussian Cox processes

Master's thesis

Ludwig-Maximilians-Universität München
Institute of Statistics

Author:	Oliver Engl
Supervisor:	Prof. Dr. Volker Schmid
Submission date:	May 13, 2016

Abstract

While much work on the fitting of spatial point processes to complex point pattern data is available, interactions between two point processes have not been in the focus so far. In this thesis, we provide methods for modelling such interactions by means of one-dimensional log-Gaussian Cox processes. Moreover, techniques for speeding up the inferential task are applied. These techniques comprise the approximation of Gaussian processes by Gauss-Markov processes and the use of the integrated nested Laplace approximation (INLA) approach.

Several tools for model validation and comparison are introduced. We use them in diverse simulation studies to assess the quality of the established interaction models. The studies are performed within the R software package R-INLA.

Finally, a discussion on how to determine the best interaction model for the situation at hand is conducted. Some possible extensions complete the present work.

This thesis is intended to be a basis for further work on interactions between point processes. In particular, the extension to two-dimensional use cases with much larger point patterns, where the speed-up techniques reveal their full potential, are a long-term objective.

Contents

1 Introduction

Point processes, particularly spatial ones, are of great interest in a wide range of applications. They play an important role in many areas of science, including geosciences [SG03], ecology [IMS+13], [FBV+15] and epidemiology [KHB01]. While there exist a lot of publications about fitting point process models to point patterns of varying complexity, the modelling of interactions between multiple point processes has not been a big theme so far – even though it is acknowledged to be helpful. Gelfand et al. [GSBS04, p.267] for instance state that they "[...] [seek] flexible, interpretable and computationally tractable multivariate models [...] which capture association both within measurements at a given site and across the sites."

In the present work, we expand the "toolbox for fitting complex spatial point process models" provided by Illian et al. [ISR12] to the setting of interaction modelling. We will limit ourselves to the case of two one-dimensional point processes, though. A popular approach to model such point processes is by means of log-Gaussian Cox processes, which are introduced in Section 2.1 and constitute our tool to model the interactions between the two point processes.

While in former times, one had to make an essential decision between modelling the processes in the right but computationally expensive continuous way or by a computationally effective but essentially not appropriate lattice-based approach, Lindgren et al. [LRL11] recently showed that there is a link between those two techniques. Consequently, it is possible to model the point processes in the correct continuous setting and simultaneously conducting the calculations in a computational effective manner. In contrast to Illian et al. [ISR12], we make explicitly use of this approach and describe it in Section 2.2.

The remaining work is organized as follows: After introducing several methods to model the interactions between two point processes in Section 3, we give some concrete examples as well as simulations based on the models in Section 4. In Section 5, the INLA methodology, used for the inference of the simulation studies in Section 6, and its software implementation within the R program are presented. Moreover, tools to validate and compare the models during the simulation studies are provided. In Section 7, we summarize and discuss the results with the focus lying on a comparison of the interaction modelling methods. Finally, some possible extensions are given in Section 8.

2 Theoretical background

2.1 Log-Gaussian Cox processes

A useful class for modelling (spatial) point processes are log-Gaussian Cox processes (see Rue et al. [RMC09, p.340]). Before examining them, we have to state how a spatial point process and a Poisson process are generally defined.

Transferring the definition of Møller and Waagepetersen [MW07, p.647] to our one-dimensional setting, a spatial point process $\{\chi\}$ is defined as a finite random subset of a given bounded region $\mathcal{S} \subset \mathbb{R}$ and a realization of such a process is a spatial point pattern $\{\xi_1, \ldots, \xi_n\}$ of $n \geq 0$ points contained in S.

Then, an inhomogeneous Poisson point process $\{\chi\}$ is a particular point process with intensity measure μ and intensity function Λ, which satisfies for any bounded region $\mathcal{B} \subset \mathcal{S}$ with $\mu(\mathcal{B}) > 0$ that the number of points in $\mathcal{B}$ is Poisson distributed with mean $\mu(\mathcal{B})$. Furthermore, conditional on the number of points, the points in $\mathcal{B}$ are i.i.d. with the density being proportional to $\Lambda(s)$ for $s \in \mathcal{B}$ [MW07, pp.649,650].

A log-Gaussian Cox process finally is a hierarchical Poisson process $\{\chi\}$ with random intensity $\Lambda(s) = \exp(\eta(s))$, where $\eta = \{\eta(s) : s \in \mathbb{R}\}$ denotes a Gaussian process.

The Gaussian process accounts for various effects in an additive way. For our spatial setting, it may be written as (see also Rue et al. [RMC09, p.320])

$$\eta(\xi_i) = \beta + f(\xi_i) + \epsilon(\xi_i) \tag{1}$$

for given observations $\{\xi_1, \ldots, \xi_n\}$. Here, β is the intercept of the process while f denotes a spatially structured effect and ϵ a spatially unstructured effect. As the process η is Gaussian within the log-Gaussian Cox process setting, the appropriate priors for f and ϵ are normal ones with

$$f \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma}), \tag{2}$$

$$\epsilon(\xi_i) \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2), \tag{3}$$

and the prior for the intercept β may be improper (constant) or normal as well.

The last question remaining is how to model the covariance matrix $\mathbf{\Sigma}$ for the spatially structured effect f. We will employ the Matérn covariance to make use of the explicit link between Gaussian processes and Gauss-Markov processes. This possibility is discussed in the next section where also a definition of the Matérn covariance function is given.

2.2 The "big-n problem": Approximation of Gaussian processes by Gauss-Markov processes

As stated in the last section, the covariance matrix $\boldsymbol{\Sigma}$ of the spatially structured effect f must be specified to complete the representation of the Gaussian process. This is usually done by using a covariance matrix $\boldsymbol{C}(\cdot, \cdot)$ so that $\boldsymbol{\Sigma} = (\boldsymbol{C}(\xi_i, \xi_j))_{i,j}$ (see Lindgren et al. [LRL11, p.424]). The covariance function is often modelled to be isotropic, i.e. it is only dependent on the Euclidean distance between two points instead of the exact locations. We will apply this simplified case here, see Section 8 for a discussion on more general covariances.

Nevertheless, the covariance function must be chosen so that the covariance matrix $\boldsymbol{\Sigma}$ is symmetric non-negative definite. This restriction leads to some usual covariance functions like the exponential covariance function, the spherical covariance function or the Matérn covariance function. However, all these functions have a common problem when using their related covariance matrices in the inference step: the covariance matrix $\boldsymbol{\Sigma}$ is dense in general. So factorizing the precision matrix (the inverse covariance) while fitting a log-Gaussian Cox process model is often much too expensive concerning the running time for large point patterns. This issue is called the "big-n problem" as the larger the point pattern is the slower the computation gets.

For a long time, a workaround for this problem was to utilize discretely indexed Gauss-Markov processes instead of the continuous Gaussian processes. A Gauss-Markov process is a Gaussian random variable $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ with Markov properties: For most of the $i \neq j$, γ_i and γ_j are independent conditional on the set of the other γ_k: $\boldsymbol{\gamma}_{-ij}$. This fact results in a sparse precision matrix which may consequently be factorized computationally efficient in the inference step. In contrast, as the Gauss-Markov processes are specified through their full conditionals, their marginal properties are not transparent [LRL11, p.423]. See Rue et al. [RMC09, Ch.2.1] for further information on the implementation of Gauss-Markov processes.

Until recently, there was the question if for a Gaussian process at hand there exists a Gauss-Markov process which is able to represent it well enough to maintain the interpretation of the parameters and the results. Lindgren et al. [LRL11] showed that this is indeed possible for a certain type of covariance functions, the so-called Matérn covariance functions. They establish an explicit link by using stochastic partial differential equations (SPDEs), see Lindgren et al. [LRL11] for more information.

Therefore, we can model the log-Gauss Cox processes in our work by using the well interpretable Gaussian processes with covariance matrices determined by Matérn covariance functions, but do the computations by implementing the computational efficient Gauss-Markov processes. While this possibility is not crucial in our setting, the explicit link between Gaussian and Gauss-Markov processes becomes more important when one wants to apply the methods presented in our work to much larger point patterns or in higher dimensions. Then the use of Gauss-Markov processes in the computational step may be essential for the computational feasibility of one of our models. The Matérn covariance function between two locations ξ_i and ξ_j in $\mathbb{R}$ is defined as (see Lindgren

et al. [LRL11, p.426])

$$\text{Cov}(\xi_i, \xi_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa|\xi_i - \xi_j|)^\nu K_\nu(\kappa|\xi_i - \xi_j|), \qquad (4)$$

with $|\cdot|$ denoting the Euclidean distance in $\mathbb{R}$ and K_ν the modified Bessel function of the second kind and order $\nu > 0$. The parameter ν is a smoothness parameter and the rounded value of ν determines the differentiability of the Gaussian process. $\kappa > 0$ is a scaling parameter and $\sigma^2 > 0$ is the marginal variance here.

One often uses an additional parameter ρ, the so-called range, where

$$\rho = \frac{(8\nu)^{1/2}}{\kappa} \qquad (5)$$

is the distance for which the correlation – not the covariance – has fallen to approximately 0.13 for all $\nu > 0.5$ (see Lindgren and Rue [LR15, p.4]). At this distance, the observations are consequently nearly independent. The range parameter admits a better interpretation of the scaling parameter κ.

Note that the explicit link between Gaussian processes containing Matérn covariances and Gauss-Markov processes may only be applied for integer values of α in the SPDE approximation approach, where $\alpha = \nu + d/2$. The value for d is one in our case as d is the dimension. So the explicit link is true for $\nu = 0.5, 1.5, 2.5, \ldots$ here. Nevertheless, a more "relaxed" relation based on continuous domain Markov models may be applied for Matérn models with fractional α. See Lindgren et al. [LRL11, p.493] for more information about this possibility.

Finally, one may think that the Matérn models are a too restrictive class to work with. They are "[...] the most important and most used covariance [models] in spatial statistics [...]" though [LRL11, p.425]. Moreover, the often utilized exponential covariance function is a special case of the Matérn covariance with $\nu = 0.5$, which is one of the cases where the explicit link is available in our one-dimensional setting.

3 Methods for interaction modelling

There are several options to model the interaction of inhomogeneous Poisson processes, even though we limit ourselves to modelling by means of log-Gaussian Cox processes. In this work, the diversity of modelling techniques is demonstrated by considering the interdependence structure in the case of *two one-dimensional* Poisson processes (see Section 8 for a discussion on interaction modelling for multidimensional Poisson processes respectively more than two processes).

For this issue, we transfer the approach from Illian et al. [ISR12, p.1504] to the case of two one-dimensional point patterns. The common one-dimensional observation range (see Remark 3.1 below) is discretized into N disjoint segments $\{s_i\}, i = 1, \ldots, N$ of equal length $|s_i| = |s|$. Let us call this an equally spaced grid on the observation range. The points in the two patterns are named $\{\xi_{ik_i}\}$ and $\{\tau_{il_i}\}$ respectively with $k_i = 1, \ldots, x_i$ and $l_i = 1, \ldots, y_i$, where x_i is the observed number of points of point pattern one in grid segment i and y_i the observed number of points of point pattern two in grid segment i.

Remark 3.1. We assume that the two Poisson processes are defined in the same one-dimensional region as their interactions are generally modelled at common locations, or at least at locations close to each other which become virtually identical by using grids for approximation. An exception of this assumption are *effect delays* which occur for instance in time series. In this situation the influence of one process on the other one becomes clear only after a time lag. There are two possibilities to handle delay effects. If the exact shift of the delay effect is known for example due to natural laws, one of the processes can be indexed accordingly so that the processes match and the (time) locations are nearly the same. In contrast, if the amount of the delay is not clear, *delay effect models* can be used as they are described in Wackernagel [Wac03, Ch.20,30].

As we model the two inhomogeneous Poisson processes using log-Gaussian Cox processes, the number of points in each grid segment is described conditionally on latent processes $\{\eta_i\}$ and $\{\omega_i\}, i = 1, \ldots, N$:

$$X_i \mid \eta_i \sim Po(|s| \exp(\eta_i)), \tag{6}$$

$$Y_i \mid \omega_i \sim Po(|s| \exp(\omega_i)). \tag{7}$$

The underlying processes in turn are modelled as

$$\eta_i = \beta_1 + f_1(s_i) + u_i, \tag{8}$$

$$\omega_i = \beta_2 + f_2(s_i) + v_i. \tag{9}$$

The parameters β_1 and β_2 are the means of the two latent processes. $f_1(s_i)$ and $f_2(s_i)$ denote spatially structured effects while u_i and v_i are spatially unstructured effects. As we wish to work with sparse precision matrices, we model the structured effects using one-dimensional

Gauss-Markov processes with mean zero on the constructed grid instead of the Gaussian processes which give the log-Gaussian Cox process its name (see Section 2.2 for a discussion on the "big-n problem"). These spatially structured effects shall represent unobserved spatially structured covariates. In contrast, the effects u_i and v_i are modelled by zero-mean Gaussian i.i.d. error terms and depict unobserved covariates which are spatially unstructured (see Rue et al. [RMC09, p.340]).

While Illian et al. [ISR12] use more than one spatially structured effect (per latent process) to account for *observed* covariates, we restrict ourselves to simulation studies without any further covariates and therefore do not need more than one Gauss-Markov process. Furthermore, the authors consider a constructed covariate in the latent process to include local interaction and competition. For the sake of simplicity, we omit these extension here, see Section 8 for a more detailed discussion on the named eventualities.

Note that in our context the priors of the parameters are necessarily Gaussian as the interactions are modelled by log-*Gaussian* Cox processes. Illian et al. also give priors for the hyperparameters, which don't have to be Gaussian distributed. The exact priors for the parameters and hyperparameters are assigned in Section 6 regarding simulation studies of the proposed models. At this point, another aspect shall be discussed. For the inference step it does not matter how the points are distributed within a particular grid segment. However, it is necessary to have an instruction for simulations of point patterns as well as for samples from the posterior distribution. The question is: If X_i takes the value x_i in grid segment s_i, how are the x_i points $\{\xi_{i1}, \ldots, \xi_{ix_i}\}$ distributed within the grid segment?

Møller and Waagepetersen [MW03, p.15] showed that those points form a Binomial process, i.e. they are uniformly distributed within the grid segment.

As aforementioned, there exist several possibilities to model the interactions of Poisson processes by dependencies between the two latent processes. We divide these methods in *Modelling by functional connections* and *Modelling by bivariate Gauss-Markov processes*, which are presented in Sections 3.1 and 3.2.

3.1 Modelling by functional connections

The first "functional connection model" is already shown in Equations 8 and 9: the model of independence. To demonstrate this fact, let us rewrite the model to

$$\eta_i = \beta_1 + f_1(s_i) + u_i, \tag{10}$$

$$\omega_i = \beta_2 + g(f_1(s_i)) + f_2(s_i) + v_i, \tag{11}$$

where in general g is an arbitrary function.

Independence If $g \equiv a$ with $a \in \mathbb{R}$, then we are back at Equation 9 with a new mean $\tilde{\beta}_2 := \beta_2 + a$. Thus, modelling of two independent Poisson processes is included in the model presented

in Equations 10 and 11 by using a constant function g.

Linear dependence If g is a linear function, Equation 11 has the form

$$\omega_i = \beta_2 + [a + \beta_3 f_1(s_i)] + f_2(s_i) + v_i, \tag{12}$$

which we can rewrite by considering β_2 and a as a common mean to

$$\omega_i = \beta_2 + \beta_3 f_1(s_i) + f_2(s_i) + v_i. \tag{13}$$

If we assume the process ω_i to depend only on the same spatially structured effect as η_i, the last model is even simplified to

$$\omega_i = \beta_2 + \beta_3 f_1(s_i) + v_i. \tag{14}$$

Non-linear dependence Finally, g can also be a non-linear function. There exist several methods to estimate such smooth effects in a Bayesian setting as we use for inference. Our approach is a full Bayesian one with P-splines (see Section 6 for inference and Fahrmeir et al. [FKL09, Ch.7] for a German introduction to splines and P-splines).

Remark 3.2.

(a) Note that smooth effects modelling by Markov processes is included in the P-spline approach. Actually, they correspond to B-splines of degree zero where the knots are specified by the values of the observed covariate. These are the values of $f_1(s_i)$ for all $i \in \{1, \ldots, N\}$ in our context (see Fahrmeir et al. [FKL09, p.326]).

(b) One might ask why the spatially structured effects $f_1(s_i)$ and $f_2(s_i)$ are not also modelled by the more general P-splines instead of restricting ourselves to Gauss-Markov processes. This is due to the fact that we want to use the approach in Lindgren et al. [LRL11] to approximate Gaussian processes with special Matérn covariance functions by Gauss-Markov processes. See Section 2.2 for further explanation.

Back to the non-linear function g: In this context, it is important to model the effect between the two processes η_i and ω_i in the "right direction" as g is not invertible in general (consider for example $g(f_1(s_i)) = \sin(f_1(s_i))$). That is, the supposed causality between the two overlying Poisson processes must be considered.

Adding dependencies on the latent process and Poisson process So far, only functional connection models with dependencies on the spatially structured effect $f_1(s_i)$ have been considered. However, the latent process ω_i might also depend on the whole other latent process η_i or even on the realisations of the overlying Poisson process X_i (idea extracted from Illian et al. [ISR12,

pp.1520,1521]). Equation 11 is consequently expanded to

$$\omega_i = \beta_2 + g_1(f_1(s_i)) + g_2(\eta_i) + g_3(X_i) + f_2(s_i) + v_i. \tag{15}$$

Just as g_1 (which was g in the previous equations), g_2 and g_3 are modelled as linear or non-linear functions, the latter by P-splines. In the majority of cases, only one of the three stated dependencies will be included in the equation. Yet, if more than one is modelled, another issue might arise: that of collinearity. We analyse this problematic in an example.

Example 3.3. Assume that both $g_1(f_1(s_i))$ and $g_2(\eta_i)$ are included in the model. That is, $f_1(s_i)$ and η_i are the explanatory variables here. Recall Equation 8

$$\eta_i = \beta_1 + f_1(s_i) + u_i,$$

where a linear relationship almost exists between $f_1(s_i)$ and η_i. Only the spatially unstructured effect u_i prevents $f_1(s_i)$ and η_i from being in an exact linear relationship. However, if the variance of u_i – which is modelled by a zero-mean Gaussian distribution – is negligible compared to the absolute values that the term $\beta_1 + f_1(s_i)$ achieves, we are in the case of a strong collinearity.

Therefore, each time we consider to include more than one of the three possible dependencies into the model, we have to verify in advance whether collinearity may arise.

The full functional connection model   Each of the models presented above can be incorporated in a general functional connection model:

$$
\begin{aligned}
X_i \mid \eta_i &\sim Poi(|s| \exp(\eta_i)), \\
Y_i \mid \omega_i &\sim Poi(|s| \exp(\omega_i)), \\
\eta_i &= \beta_1 + f_1(s_i) + u_i, \\
\omega_i &= \beta_2 + g_1(f_1(s_i)) + g_2(\eta_i) + g_3(X_i) + f_2(s_i) + v_i.
\end{aligned}
\tag{16}
$$

We give some examples and simulations of them in Section 4 as well as coming back to modelling by functional connections when conducting simulation studies in Section 6.

3.2 Modelling by bivariate Gauss-Markov processes

Another approach to model the interaction of two Poisson processes is by combining the *two univariate* Gaussian processes of Equations 8 and 9 in *one bivariate* Gaussian process. This yields to the formula

$$\boldsymbol{\Phi_i} = \boldsymbol{\beta} + \boldsymbol{f(s_i)} + \boldsymbol{\epsilon_i}, \tag{17}$$

where $\boldsymbol{\Phi_i} = (\eta_i, \omega_i)^T$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$, $\boldsymbol{f(s_i)} = (f_1(s_i), f_2(s_i))^T$ and $\boldsymbol{\epsilon_i} = (u_i, v_i)^T$.

Again, the parameter $\boldsymbol{\beta}$ is the mean of the bivariate process, while the $\boldsymbol{\epsilon_i}$ are spatially unstructured effects and modelled as Gaussian white noise vectors. Therefore, the $\boldsymbol{\epsilon_i}$ are zero-mean bivariate normal distributed ($\boldsymbol{\epsilon_i} \sim N_2(0, \boldsymbol{D})$), where $\boldsymbol{D}$ is a 2×2 diagonal matrix with $\boldsymbol{D}_{kk} = \tau_k^2$ being the variance of the u_i respectively v_i ($k \in \{1, 2\}$). Thus, the $\boldsymbol{\epsilon_i}$ are independent from each other ($i \in \{1, \ldots, N\}$), as well as u_i and v_i are independent for a given segment s_i.

Consequently, the dependencies both *within* the processes η_i and ω_i and *between* the two processes are determined by the vectors $\boldsymbol{f(s_i)}$, as before representing spatially structured effects. The typical way to do this is to assume $f_1(s_i)$ and $f_2(s_i)$ to have mean zero and to specify marginal-covariance functions for the dependence structure within the processes and cross-covariance functions for the interactions between them [GK15, pp.147,148]. As $\boldsymbol{f(s_i)}$ is bivariate Gaussian, this is enough to define the entire process.

For this purpose, let $\boldsymbol{C} : \{s_1, \ldots, s_N\} \times \{s_1, \ldots, s_N\} \to \boldsymbol{M}_{2\times 2}$ be a matrix-valued mapping with

$$\boldsymbol{C(s_i, s_j)} = \begin{pmatrix} C_{11}(s_i, s_j) & C_{12}(s_i, s_j) \\ C_{21}(s_i, s_j) & C_{22}(s_i, s_j) \end{pmatrix}. \tag{18}$$

Here, $\boldsymbol{M}_{2\times 2}$ is the set of 2×2 real-valued matrices and the $C_{kl}(s_i, s_j) = \mathrm{Cov}(f_k(s_i), f_l(s_j))$ ($k, l \in 1, 2$) are the marginal-covariance functions for $k = l$ and the cross-covariance functions for $k \neq l$. Therefore, $\boldsymbol{C(s_i, s_j)} = \mathbf{Cov}(\boldsymbol{f(s_i)}, \boldsymbol{f(s_j)})$. Note that $C_{12}(s_i, s_j) = C_{21}(s_j, s_i)$ must hold as the covariance is symmetric, but in general $C_{12}(s_i, s_j) \neq C_{21}(s_i, s_j)$, so two cross-covariance functions should be specified in our case.

The main issue is that the cross-covariance functions must be chosen to be consistent with the marginal-covariance functions. That is, the covariance matrix $\boldsymbol{\Sigma}$ of the random vector $\left(\boldsymbol{f(s_1)}^T, \ldots, \boldsymbol{f(s_N)}^T \right)^T \in \mathbb{R}^{2N}$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{C(s_1, s_1)} & \boldsymbol{C(s_1, s_2)} & \cdots & \boldsymbol{C(s_1, s_N)} \\ \boldsymbol{C(s_2, s_1)} & \boldsymbol{C(s_2, s_2)} & \cdots & \boldsymbol{C(s_2, s_N)} \\ \vdots & \vdots & \ddots & \cdots \\ \boldsymbol{C(s_N, s_1)} & \boldsymbol{C(s_N, s_2)} & \cdots & \boldsymbol{C(s_N, s_N)} \end{pmatrix} \tag{19}$$

must be symmetric non-negative definite to be valid.

This is a non-trivial task and usually solved by constructive approaches. Such techniques include separable models, convolution methods, latent dimension models, the linear model of coregionalization (LMC) and multivariate Matérn models. In this work, we will focus on the LMC and the multivariate Matérn, since these two procedures allow once again the approximation of the Gaussian processes $f_1(s_i)$ and $f_2(s_i)$ by Gauss-Markov processes (with an appropriate dependence structure both within and between the processes). See Gelfand et al. [GSBS04, Sec.2] and Genton and Kleiber [GK15, Sec.2.2,2.3] for a brief review of the other approaches as well as for references to more elaborated descriptions.

3.2.1 The linear model of coregionalization

The linear model of coregionalization is "[probably] the most popular approach of combining univariate covariances [...]" [GK15, p.150]. Originally used as a tool for dimension reduction (see Grzebyk and Wackernagel [GW94]), the idea of the LMC was seized by Schmidt and Gelfand [SG03] for multivariate process modelling (see Gelfand et al. [GSBS04, Ch.3]).

The idea is to represent the bivariate Gaussian process as a linear combination of independent univariate Gaussian processes: $\boldsymbol{f}(\boldsymbol{s_i}) = \boldsymbol{A}\boldsymbol{w}(\boldsymbol{s_i})$, $\boldsymbol{A}$ is a $2 \times r$ real-valued matrix with $r \leq 2$ and $\boldsymbol{w}(\boldsymbol{s_i})$ is a vector of independent univariate spatial processes. For our purposes, assume $\boldsymbol{w}(\boldsymbol{s_i})$ to be zero-mean Gaussian and $r = 2$, i.e. $\boldsymbol{w}(\boldsymbol{s_i}) = (w_1(s_i), w_2(s_i))^T$ (for $r = 1$ the matter in hand corresponds to the functional connection model with linear dependence as modelled in Equations 10 and 14). Then we get

$$\eta_i = \beta_1 + a_{11}w_1(s_i) + a_{12}w_2(s_i) + u_i, \tag{20}$$

$$\omega_i = \beta_2 + a_{21}w_1(s_i) + a_{22}w_2(s_i) + v_i, \tag{21}$$

or, by incorporating a_{11} and a_{22} in the variances of the Gaussian processes $w_1(s_i)$ and $w_2(s_i)$ respectively,

$$\eta_i = \beta_1 + w_1(s_i) + a_{12}w_2(s_i) + u_i, \tag{22}$$

$$\omega_i = \beta_2 + a_{21}w_1(s_i) + w_2(s_i) + v_i. \tag{23}$$

Therefore, $\boldsymbol{f}(\boldsymbol{s_i})$ consists of

$$f_1(s_i) = w_1(s_i) + a_{12}w_2(s_i), \tag{24}$$

$$f_2(s_i) = a_{21}w_1(s_i) + w_2(s_i). \tag{25}$$

As the processes $w_1(s_i)$ and $w_2(s_i)$ are independent, the cross-covariance functions for interaction modelling of η_i and ω_i do no longer have to be specified. Instead, there is only one cross-covariance function which is already fixed by the marginal-covariances of $w_1(s_i)$ and $w_2(s_i)$ as $C_{12}(s_i, s_j) = \text{Cov}(f_1(s_i), f_2(s_j)) = a_{21}\text{Cov}(w_1(s_i), w_1(s_j)) + a_{12}\text{Cov}(w_2(s_i), w_2(s_j)) = \text{Cov}(f_2(s_i), f_1(s_j)) = C_{21}(s_i, s_j)$. Consequently, only the marginal-covariances for $w_1(s_i)$ and $w_2(s_i)$ must be determined here. These are modelled as seen previously as Gaussian processes representing spatially structured effects, each of them affecting *both* $f_1(s_i)$ and $f_2(s_i)$.

While the problem of cross-covariance specification is solved by this approach, it must be mentioned that the matrix $\boldsymbol{C}(\boldsymbol{s_i}, \boldsymbol{s_j})$ in the LMC is restricted to be symmetric although this does in general not hold (see the beginning of Section 3.2 and the discussion in Wackernagel [Wac03, Ch.20]).

Note that the "standard" LMC is somewhat limited in the case of interaction modelling of *two* processes as the number of independent Gaussian processes r is restricted to be not higher than

the number of processes to model the dependence between. Its full potential is only achieved in the case of interaction modelling of *more than two processes* (cf. the extensions in Section 8). However, if we transmit the idea of a shared component model for spatial disease mapping in the work of Knorr-Held and Best [KHB01] to our case, this results in a more general LMC which allows for more than two latent processes.

The shared component model   In our shared component model, each of the processes $f_1(s_i)$ and $f_2(s_i)$ depends on a shared Gaussian process $z(s_i)$ and a process-specific Gaussian process $w_1(s_i)$ respectively $w_2(s_i)$:

$$f_1(s_i) = z(s_i) + w_1(s_i), \tag{26}$$

$$f_2(s_i) = az(s_i) + w_2(s_i), \tag{27}$$

all three of them being independent from one another. So there are *three* independent latent processes instead of two as in the "standard" LMC.

Hypothetically, one could even extend the number of latent univariate processes which influence the processes $f_1(s_i)$ and $f_2(s_i)$. However, there must be a conclusive model to explain why one should do this. In the case of the shared component model above, the theoretical basis lies in disease mapping. Knorr-Held and Best state that when considering two diseases, there might be a common risk factor (which is modelled by $z(s_i)$ in our case, e.g. smoking) as well as disease-specific risk factors ($w_1(s_i)$ and $w_2(s_i)$, e.g. different genetic causes). Such a modelling approach is reasonable if one is not only interested in interactions but also in finding the amount and the spatially structure of a common effect. Nevertheless, this requires greatest caution to ensure identifiability – a problem which is already present in the "standard" LMC, albeit less pronounced.

The identifiability issue   As announced in the introduction, our approach for inference is a Bayesian hierarchical one (see Section 2.1). Moreover, we model the covariance functions of the Gaussian processes $w_1(s_i)$, $w_2(s_i)$ and $z(s_i)$ (in the case of the shared component model) once again by Gauss-Markov processes to get sparse precision matrices (see Section 2.2). While the complete inference is described in Section 5, the matter of identifiability shall be discussed right here. Actually, there is a *lack of identifiability in the likelihood* as the number of parameters to estimate is higher than the quantity of observations which are available. To begin with, we demonstrate this fact in the case of the "standard" LMC.

Recall the equations for the (dependent) processes $f_1(s_i)$ and $f_2(s_i)$:

$$f_1(s_i) = w_1(s_i) + a_{12}w_2(s_i),$$

$$f_2(s_i) = a_{21}w_1(s_i) + w_2(s_i).$$

Then for each segment s_i there are the two values $w_1(s_i)$ and $w_2(s_i)$ to estimate while two observations (one for each process) are available for this purpose. *But* there are additionally the two parameters a_{12} and a_{21} to be estimated *globally* as well as the hyperparameters (variance, scale and smoothness) of the two processes. That is, one is confronted with the announced identifiability issue. This can be solved in our Bayesian approach by using appropriate prior distributions for the variances of the Gauss-Markov processes and the smoothness and scale of the Matérn covariance functions to be modelled (see again Section 2.2) as these influence the values of $w_1(s_i)$ and $w_2(s_i)$ by "shifting" the posterior distributions in a way that is concordant with prior believes. For instance, it is often assumed that the underlying processes represent spatially structured effects with two different spatial scales which can be expressed by suitable informative priors for the scales of the covariance functions (see Ren and Banerjee [RB13] and Ribeiro et al. [RSP15]). Another possibility is to impose constraints on some of the parameters or actually fix them a priori. Regardless of which method is applied, one has to choose the regularization very carefully to avoid undesired effects.

In the shared component model the issue is even more pronounced as there are the three values $w_1(s_i)$, $w_2(s_i)$ and $z(s_i)$ for each segment, a global parameter a and the hyperparameters which have to be estimated with only two observations per segment. The approaches to solve this problem are the same as above: Choose appropriate prior distributions, impose constraints on the parameters or fix them (the proposals in this paragraph are partially adopted from Knorr-Held and Best [KHB01, pp.77,78], Lindgren et al. [LRL11, p.470] and Ren and Banerjee [RB13, pp.21,22]).

Remark 3.4. If we are not interested in the amounts and spatially structures of the underlying effects in the "standard" LMC, but solely in the construction of a valid covariance matrix by means of the LMC – which is often the case – we may assume $\boldsymbol{A}$ to be lower triangular to facilitate inference (see Gelfand et al. [GSBS04, p.272]). This is due to the fact that it is possible to define $\tilde{w}_1(s_i) = w_1(s_i) + a_{12}w_2(s_i)$ and $\tilde{w}_2(s_i) = (1 - a_{21}a_{12})w_2(s_i)$, allowing rewriting Equations 24 and 25:

$$f_1(s_i) = \tilde{w}_1(s_i), \tag{28}$$

$$f_2(s_i) = a_{21}\tilde{w}_1(s_i) + \tilde{w}_2(s_i). \tag{29}$$

Hence, we are back at the functional connection model with linear dependence of Equations 10 and 13.

3.2.2 The bivariate Matérn model

A more recent technique to construct bivariate Gaussian processes dates from Gneiting et al. [GKS10]. In their approach the marginal-covariances as well as the cross-covariances are modelled as covariance functions of the Matérn-class. As mentioned previously, this is an appropriate choice to describe the (spatial) dependencies both within and between the processes as the the

Matérn covariance is the standard covariance model for *univariate* processes (see the discussion on this point in Section 2.2). Nevertheless, the issue of ensuring the covariance matrix $\boldsymbol{\Sigma}$ in Equation 19 to be non-negative definite at first remains in this approach.

For this purpose, Gneiting et al. present some relatively complex relations which must hold for the parameters of their model to receive a valid covariance matrix (see Gneiting et al. [GKS10, p.1170]). While these conditions *fully* characterize a valid bivariate Matérn model, they also give some simpler *sufficient* conditions (see Gneiting et al. [GKS10, p.1171]).

That means, the issue of constructing a valid bivariate Gaussian process can be solved by their approach, albeit it is a little bit tricky. But the precision matrices of the two processes, which we want to be sparse, are dense in general. Consequently, the method of Gneiting et al. is not suitable for our purposes.

To solve this final issue, Hu et al. [HSLR13] bring together the idea of Gneiting et al. and the Gauss-Markov approximation approach of the univariate case presented in Lindgren et al. [LRL11]. Similar to the approximation of the univariate case (cf. Section 2.2), the bivariate Matérn model is constructed using a system of stochastic partial differential equations (SPDEs), and, more important, the constructed covariance matrix $\boldsymbol{\Sigma}$ is *automatically* symmetric non-negative definite, i.e. valid (see Hu et al. [HSLR13, p.5]). How this may be achieved is beyond the scope of this thesis, we refer to Sections 2 and 3.1 in the paper of Hu et al. for a detailed description of the model construction approach by SPDEs.

With this method, the precision matrix of the (approximated) bivariate Gaussian process is sparse due to the usage of a bivariate Gauss-Markov process, i.e. the remaining problem of the approach of Gneiting et al. is solved (see Hu et al. [HSLR13, p.11]). Furthermore, there is a relationship between the approaches of Gneiting et al. and Hu et al. in the sense that the parameters of the two approaches are related by several equations (see Hu et al. [HSLR13, Sec.3.3]).

It shall be announced that when modelling by SPDEs, one may use the *triangular* or the *full* version of SPDEs. In the triangular version, the properties of the bivariate Gaussian process are easier to interpret than in the full version. In return, there is once again an imposed symmetry property on the cross-covariance just like in the LMC, which is not the case in the full version (see Hu et al. [HSLR13, p.13]). Moreover, in both versions, the first process is a Matérn process, but the second in general *is not.* It is actually a mixture of Matérn models which is very close to a Matérn (see Hu et al. [HSS$^+$15, p.16]). In Hu et al. [HSLR13, p.19] the authors state that with additional conditions, the second process could also be a Matérn process, unfortunately without denoting the exact conditions.

Remark 3.5. A discussion on the pros and cons of the different modelling techniques presented so far shall not be conducted at this point. Rather, it is an essential part of the discussion in Section 7, since the comparison highly depends on the accuracy of the inference conducted in Section 6.

4 Examples and simulations

A lot of theoretical (sub-)models have been introduced so far. In the following, some concrete examples are given as well as simulated and illustrated. These examples constitute the basis for the inference in Section 6, where we try to retrieve the parameters and effects of the simulated models.

Before stating the examples, several general information shall be announced. For all the examples, the point patterns of the two one-dimensional Poisson processes are simulated on a segment which starts at 0 and ends at 1. This is a modelling simplification which can be extended in a straightforward way to other segments of finite length.

Furthermore, the spatially structured effects are simulated as Gaussian processes with Matérn covariance functions as they shall be approximated by Gauss-Markov processes in the inference step. See again Section 2.2 for a discussion on this matter.

For simulation purposes, the values of the spatially effects (and consequently of the Poisson processes) are simulated at the centres of an equally spaced grid with 1000 segments covering the observation range from 0 to 1. Then the points in one grid segment are equally distributed within that segment (see Section 3 and Møller and Waagepetersen [MW03, p.15]) and may easily be simulated.

4.1 Modelling by functional connections: Independence

To begin with, two *independent* Poisson processes are presented. While this example does not give any insight into interaction modelling, it is a good reference point for comparisons with the other models, e.g. in terms of accuracy of the estimates of the parameters and spatially structured effects.

The two Poisson processes are determined by the latent processes which are modelled as (see Equations 8 and 9)

$$\eta_i = \beta_1 + f_1(s_i) + u_i,$$
$$\omega_i = \beta_2 + f_2(s_i) + v_i,$$

where

$$f_1 \sim \mathcal{N}_{1000}(\mathbf{0}, \mathbf{\Sigma_1}), \tag{30}$$

$$f_2 \sim \mathcal{N}_{1000}(\mathbf{0}, \mathbf{\Sigma_2}), \tag{31}$$

$$u_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2), \tag{32}$$

$$v_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2), \tag{33}$$

and the parameters and hyperparameters in this first example are given as indicated in Table 1. There are six parameters which have not been discussed yet: $\sigma_{f_1}^2$, ν_1, κ_1, $\sigma_{f_2}^2$, ν_2 and κ_2. These

14

Table 1: Modelling by functional connections: Parameters and hyperparameters for the independence example.

Process	η_i					ω_i				
Parameter	β_1	σ_u^2	$\sigma_{f_1}^2$	ν_1	κ_1	β_2	σ_v^2	$\sigma_{f_2}^2$	ν_2	κ_2
Value	5	0.1	0.7	0.5	10	4	0.1	0.4	1.5	6

are the parameters which determine the values of the Matérn covariance functions of the two processes and which define in this way the covariance matrices Σ_1 and Σ_2 (see also Section 2.2 for a definition of the Matérn covariance function). $\sigma_{f_1}^2$ and $\sigma_{f_2}^2$ are the marginal variances of the processes $f_1(s_i)$ respectively $f_2(s_i)$, ν_1 and ν_2 are smoothness parameters, and κ_1 and κ_2 are scaling parameters of the two processes. It must be pointed out that $\nu_1 = 0.5$ and $\nu_2 = 1.5$ are chosen so that the explicit link between Gaussian processes and Gauss-Markov processes may be used in the inference step. In other examples, we use values for ν which do not profit by this link but by a more relaxed relation between Gaussian processes and Gauss-Markov processes (see Section 2.2 for a discussion on both possibilities).

In Figure 1 simulations of the processes based on the parameters in Table 1 are illustrated. The two top panels show realizations of the spatially structured effects $f_1(s_i)$ and $f_2(s_i)$, simulated as Gaussian processes with Matérn covariance functions with the parameters indicated in Table 1. The process $f_2(s_i)$ is much smoother than the process $f_1(s_i)$ as the smoothness parameter is higher: $\nu_2 = 1.5$ whereas $\nu_1 = 0.5$. The scope of reached values is higher for $f_1(s_i)$ (roughly 5) than for $f_2(s_i)$ (roughly 1.5) as the variance parameter $\sigma_{f_1}^2$ for the Gaussian process $f_1(s_i)$ is higher. Moreover, the range value of the process $f_2(s_i)$ is much higher than that of $f_1(s_i)$ (see again Section 2.2 for the definition and implications of the range of a given process). This is also the reason why the second process is not centred at all around zero although the processes initially have been modelled with expectation zero: One accidental high value at the beginning of the observation range has a big influence on all the other values of $f_2(s_i)$.

The two bottom panels of Figure 1 show realizations of the two latent processes η_i and ω_i, based on the simulations of $f_1(s_i)$ and $f_2(s_i)$ displayed in the top panels. That is, intercepts and Gaussian white noise (parameters again denoted in Table 1) are added to the spatially structured effects. The latent processes η_i and ω_i constitute the basis for the following simulations of the Poisson processes X_i and Y_i.

Realizations of the counts of the two processes are shown in the bottom panels of Figure 2. One may see that there are more counts at the positions of the observation range where the corresponding latent process has higher values. This is for example visible at the end of the observation range for the processes η_i and X_i and – less pronounced – at the beginning and in the middle of the range for ω_i and Y_i.

In the top left panel the simulations of the two latent processes η_i and ω_i are plotted once again in a joint graphic for comparison purposes, while the top right panel illustrates simulations of
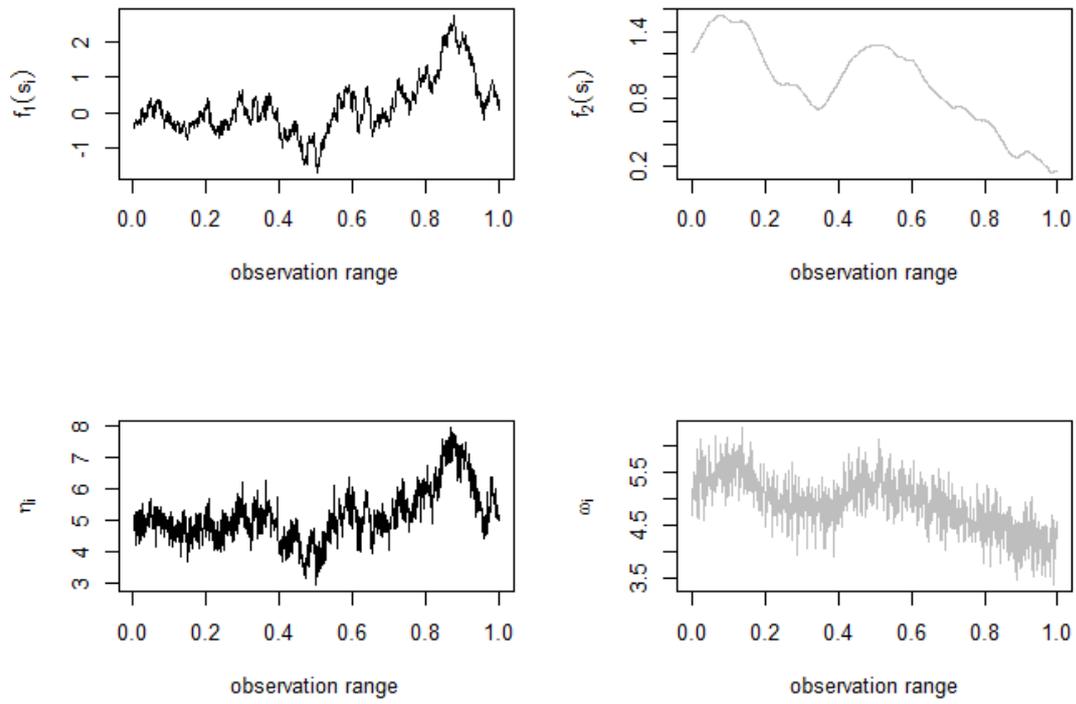
Figure 1: Modelling by functional connections: Simulations for the independence example. Top left: The spatially structured effect $f_1(s_i)$ of the first Poisson process X_i. Top right: The spatially structured effect $f_2(s_i)$ of the second Poisson process Y_i. Bottom left: The latent process η_i of the Poisson process X_i. Bottom right: The latent process ω_i of the Poisson process Y_i.
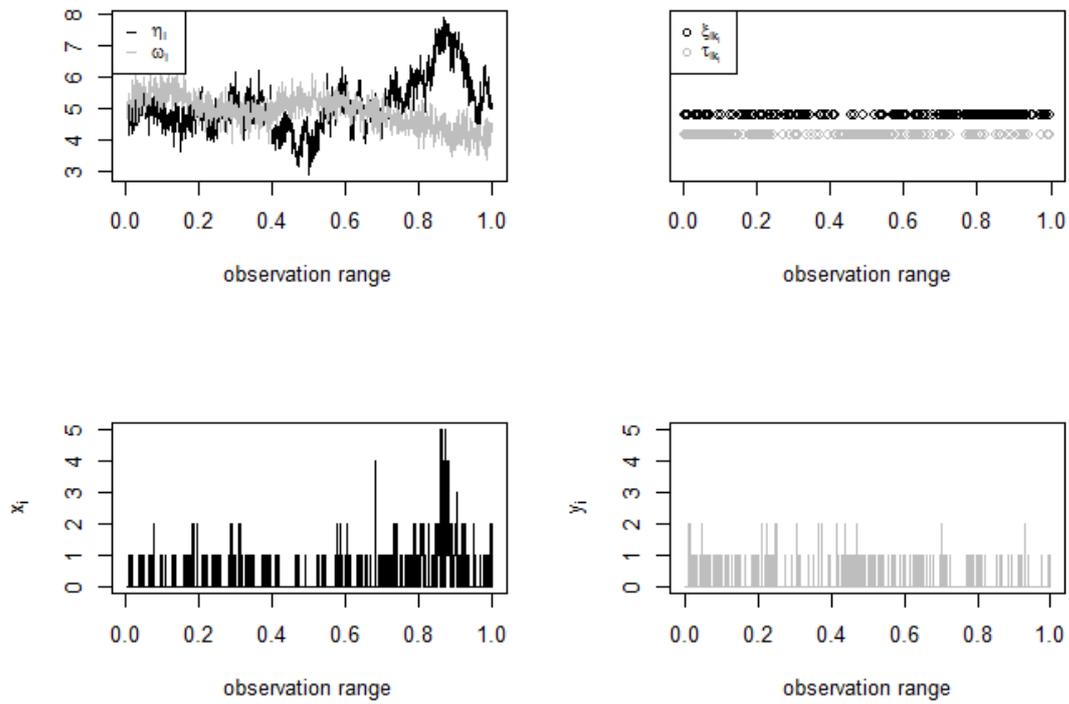
Figure 2: Modelling by functional connections: Simulations for the independence example II. Top left: The latent processes η_i (black line) and ω_i (grey line) of the Poisson processes X_i and Y_i. Top right: One-dimensional point patterns ξ_{ik_i} (black points) and τ_{ik_i} (grey points) of the Poisson processes X_i and Y_i. Bottom left: The counts of the Poisson process X_i. Bottom right: The counts of the Poisson process Y_i.

the point patterns ξ_{ik_i} and τ_{ik_i} corresponding to the the realizations of the two Poisson processes. However, the point patterns are less transparent than the Poisson counts concerning the representation of the point processes.

4.2 Modelling by functional connections: Linear and non-linear dependence

In the following, we maintain the realizations of the spatially structured effects $f_1(s_i)$ and $f_2(s_i)$ shown in Figure 1 for a better visualization of the interactions which result from a linear respectively non-linear dependence in the spatially structured effects. While

$$\eta_i = \beta_1 + f_1(s_i) + u_i$$

as previously,

$$\omega_i = \beta_2 + \beta_3 f_1(s_i) + f_2(s_i) + v_i$$

in the linear case and

$$\omega_i = \beta_2 + g(f_1(s_i)) + f_2(s_i) + v_i$$

in the non-linear case with a non-linear function g.

Again, for the purpose of comparison, the parameters of the latent processes are the same as indicated for the independence example in Table 1. Nevertheless, we have to state the parameter β_3 for the linear dependence example as well as a function g for the non-linear dependence example. In fact, two linear dependence examples will be discussed, one with $\beta_3 = 0.4$ and the other with $\beta_3 = -0.4$ since this makes a difference in the difficulty of inference as may be seen in Section 6. For the function of non-linear dependence, we use

$$g(f_1(s_i)) = \frac{f_1(s_i)^2}{2}. \tag{34}$$

In Figure 3, the interactions between the latent processes are made visible. The top left panel shows the spatially structured effects of the independence example, i.e. without any interaction. The top right panel shows first dependencies between the spatially structured effects: The black line indicates the process $f_1(s_i)$, whereas the grey line represents the spatially structured effect $0.4f_1(s_i) + f_2(s_i)$ of the process ω_i for the first linear dependence example. The blue line provides a reference to how the spatially structured effect of ω_i is depicted without any interactions (so it represents solely $f_2(s_i)$). One may recognize that the grey line "fluctuates" around the blue line in the same way as $f_1(s_i)$ moves, which is due to the positive linear dependence parameter $\beta_3 = 0.4$. In contrast, the grey line in the bottom left panel behaves inversely to the black one

as the dependence parameter is $\beta_3 = -0.4$ here. Finally, in the bottom right panel, the grey line behaves consistently with the black one in some parts of the observation range, in others not. This is caused by the non-linear dependence of the function g that reinforces the effect of $f_1(s_i)$ when $f_1(s_i)$ is greater than zero but controverts it when it is lower than zero.

The exact non-linear dependence determined by the function $f_1(s_i)^2/2$ may be seen in the top panel of Figure 4. As the range of values for the simulation of the process $f_1(s_i)$ goes roughly from -2 to a bit less than 3, only the excerpt of interest for the functional connection is shown. The range of values for $g(f_1(s_i)) = f_1(s_i)^2/2$ is 0 to around 4.

In the section about inference, we will see how (well) all these possible interactions may be recovered from Poisson counts based on the latent processes including the stated dependencies.

4.3 Modelling by functional connections: Dependence on the latent process

The last example for interaction modelling by functional connections is a simplified version of Equation 15. While

$$\eta_i = \beta_1 + f_1(s_i) + u_i$$

once again, ω_i is only dependent on the whole latent process η_i without an additional spatially structured effect $f_2(s_i)$:

$$\omega_i = \beta_2 + g(\eta_i) + v_i, \tag{35}$$

where, this time, $g(\eta_i) = -2\sin(\eta_i)$.

As before, the parameters of the processes η_i and ω_i are as indicated in Table 1 (note that the parameters $\sigma_{f_2}^2, \nu_2$ and κ_2 of the process $f_2(s_i)$ are not used at this point). The realization of $f_1(s_i)$ just as the realizations of the two unstructured effects u_i and v_i are the same as in the independence example (we do not need the realization of $f_2(s_i)$ here). That is, the latent process η_i exhibits the same pattern as illustrated in the bottom left panel of Figure 1.

This is visible in Figure 5: The black line in both panels, representing η_i, is identical to the η_i of the former examples (albeit more stretched). The grey line in the top panel indicates the linear predictor of ω_i, i.e. it corresponds to ω_i *without* the unstructured spatially effect v_i. The top panel shall point out the special dependence between the two latent processes: As the grey line is equal to $4-2\sin(\eta_i)$ (the intercept β_2 is 4), it *must* adopt values between 2 and 6 as $-2\sin(\eta_i)$ can only fluctuate between these two values. This is the reason why the grey line looks "trimmed" at the value 6 of the y-axis – and less pronounced at the value 2, too. Finally, the bottom panel shows both latent processes in one plot. They behave in a similar way apart from the end of the observation range where they are inverse. This is due to the functional connection which may be seen in the bottom panel of Figure 4 (again, only the excerpt of interest is shown). Up to the value $\eta_i = 5$, η_i and $-2\sin(\eta_i)$ are linked by a nearly linear function with slope one, whereas

19
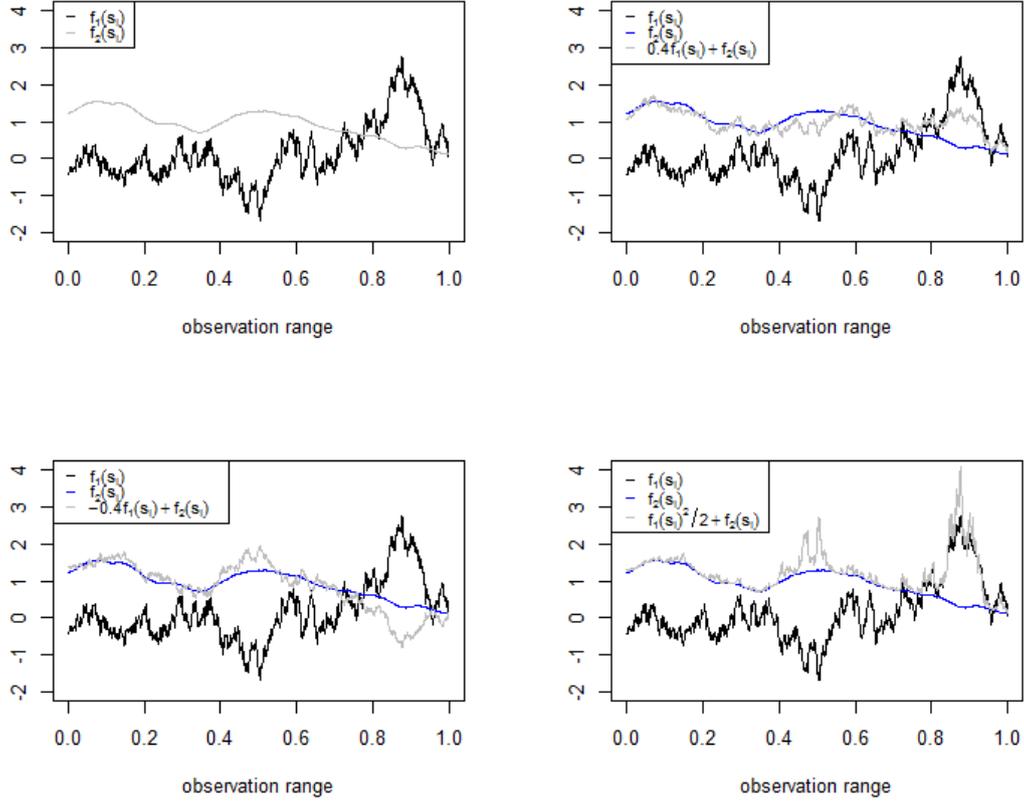
Figure 3: Modelling by functional connections: Simulations for the linear and non-linear dependence example. Top left: The spatially structured effects $f_1(s_i)$ (black line) and $f_2(s_i)$ (grey line) for the latent processes η_i and ω_i in the independence example. Top right: The spatially structured effects $f_1(s_i)$ (black line) and $0.4f_1(s_i) + f_2(s_i)$ (grey line) for the latent processes η_i and ω_i in the first linear dependence example as well as the process $f_2(s_i)$ for reference (blue line). Bottom left: The spatially structured effects $f_1(s_i)$ (black line) and $-0.4f_1(s_i) + f_2(s_i)$ (grey line) for the latent processes η_i and ω_i in the second linear dependence example. Bottom right: The spatially structured effects $f_1(s_i)$ (black line) and $f_1(s_i)^2/2 + f_2(s_i)$ (grey line) for the latent processes η_i and ω_i in the non-linear dependence example.

20
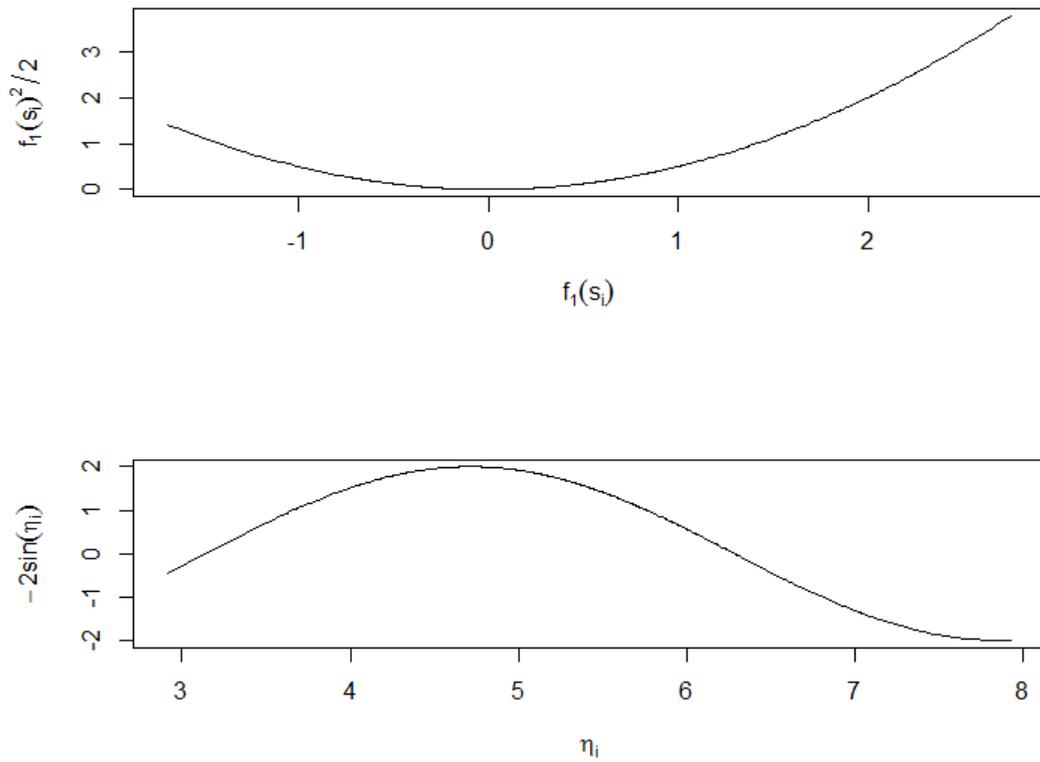
Figure 4: Modelling by functional connections: Non-linear dependencies. Top: The non-linear dependence example with the functional connection $f_1(s_i)^2/2$ based on the values of the spatially structured effect $f_1(s_i)$ of the latent process η_i. Bottom: The dependence on the latent process example with the functional connection $-2\sin(\eta_i)$ based on the values of the whole latent process η_i.
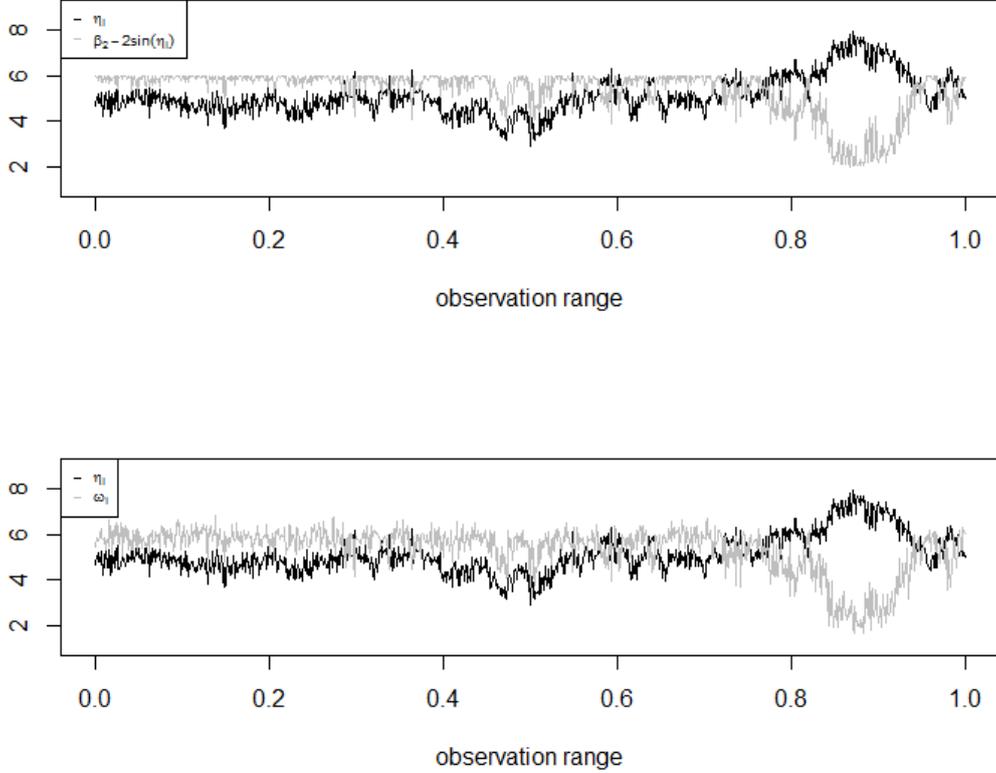
Figure 5: Modelling by functional connections: Simulations for the dependence on the latent process example. Top: The latent process η_i (black line) and the linear predictor of the process ω_i (grey line) of the Poisson processes X_i and Y_i. Bottom: The latent processes η_i (black line) and ω_i (grey line) of the Poisson processes X_i and Y_i.

afterwards the relation gets more and more inverse. With values of η_i of 6 and higher being only present at the end of the observation range, there are the locations where the inverted relation effectively gets visible.

4.4 Modelling by bivariate Gauss-Markov processes: LMC and shared component model

The next two examples deal with modelling by bivariate Gauss-Markov processes. In contrast to the former examples, we use other parameters and therefore different simulations for the spatially effects to gain further insights. In particular, the explicit link between Gaussian processes and Gauss-Markov processes does not hold any more (but a more relaxed relation does, see Section

Table 2: Modelling by bivariate Gauss-Markov processes: Parameters and hyperparameters for the LMC example.

Process	$w_1(s_i)$			$w_2(s_i)$			η_i			ω_i		
Parameter	$\sigma^2_{w_1}$	ν_1	κ_1	$\sigma^2_{w_2}$	ν_2	κ_2	β_1	σ^2_u	a_{12}	β_2	σ^2_v	a_{21}
Value	0.3	0.3	7	0.5	1.2	12	5	0.2	0.3	5.5	0.1	0.7

2.2).

The linear model of coregionalization   The first example is the linear model of coregionalization (LMC), where the latent processes are determined by (recall Equations 22 and 23)

$$\eta_i = \beta_1 + w_1(s_i) + a_{12}w_2(s_i) + u_i,$$
$$\omega_i = \beta_2 + a_{21}w_1(s_i) + w_2(s_i) + v_i,$$

where the spatially structured and unstructured effects are distributed once more as indicated in Equations 30 to 33, replacing f_1 and f_2 by w_1 and w_2. The new parameters of the latent processes are denoted in Table 2 and realizations of the spatially structured effects may be seen in Figure 6.

Neither the table nor the figure shall be discussed in detail here, but there are two points to mention: Firstly, as announced, $\nu_1 = 0.3$ and $\nu_2 = 1.2$ do not allow to profit from the explicit link between Gaussian processes and Gauss-Markov processes but from a more relaxed relation. Secondly, one may identify by a comparison of the top and the bottom panel of Figure 6 that the black line and the grey line approach each other. This is due to the mixing of the processes $w_1(s_i)$ and $w_2(s_i)$ in both overall spatially structured effects. The grey line "moves" more than the black one by comparison of the top panel to the bottom panel as $a_{12} = 0.3$ only but $a_{21} = 0.7$ and therefore much higher.

The shared component model   The second example for modelling by bivariate Gauss-Markov processes is the shared component model. There, the latent processes are given by

$$\eta_i = \beta_1 + z(s_i) + w_1(s_i) + u_i, \tag{36}$$
$$\omega_i = \beta_2 + az(s_i) + w_2(s_i) + v_i, \tag{37}$$

where the effects are distributed as indicated in Equations 30 to 33 with f_1 and f_2 replaced by w_1 and w_2. The process z is distributed just as w_1 and w_2 with another covariance matrix Σ_3. As may be seen in Table 3, the parameters for the two spatially structured effects $w_1(s_i)$ and $w_2(s_i)$ are the same as for the LMC example (but now unmixed in the overall spatially effects of the latent processes). For the shared component $z(s_i)$ an even smoother process than $w_2(s_i)$ with a long range is used to model a steady common effect. Only half of this effect is present in

23
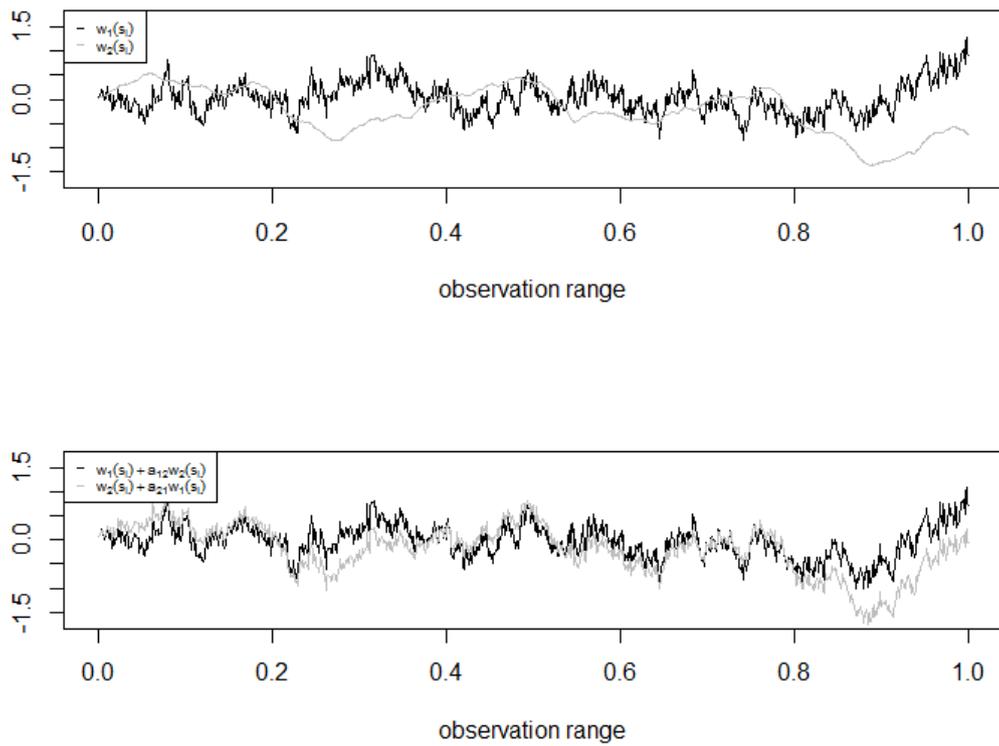
Figure 6: Modelling by bivariate Gauss-Markov processes: Simulations for the LMC example. Top: The spatially structured effects $w_1(s_i)$ (black line) and $w_2(s_i)$ (grey line). Bottom: The overall spatially structured effect $w_1(s_i) + a_{12}w_2(s_i)$ (black line) for the latent process η_i and the overall spatially structured effect $w_2(s_i) + a_{21}w_1(s_i)$ (grey line) for the latent process ω_i.

Table 3: Modelling by bivariate Gauss-Markov processes: Parameters and hyperparameters for the shared component example.

Process	$w_1(s_i)$			$w_2(s_i)$			$z(s_i)$			η_i		ω_i		
Parameter	$\sigma^2_{w_1}$	ν_1	κ_1	$\sigma^2_{w_2}$	ν_2	κ_2	σ^2_z	ν_3	κ_3	β_1	σ^2_u	β_2	σ^2_v	a
Value	0.3	0.3	7	0.5	1.2	12	0.8	1.5	4	5	0.2	5.5	0.1	0.5

the second latent process ω_i as $a = 0.5$.

Simulations of the three spatially structured effects $w_1(s_i), w_2(s_i)$ and $z(s_i)$ are illustrated in the top panel of Figure 7, where the realizations of the processes $w_1(s_i)$ and $w_2(s_i)$ are the same as in the LMC. The overall spatially structured effects $z(s_i) + w_1(s_i)$ of the latent process η_i and $az(s_i) + w_2(s_i)$ of the process ω_i are pictured in the bottom panel. It is visible by comparison of the two panels that the shared effect "shifts" the two spatially structured effects $w_1(s_i)$ and $w_2(s_i)$ upwards: a little bit at the beginning of the observation range, more at the end of the range. The effect on the grey line is less pronounced as only half of the effect $z(s_i)$ influences the overall spatially effect of ω_i ($a = 0.5$). In Section 6, we will examine how well the three intermingled spatially structured effects may be recovered from Poisson counts based on the latent processes η_i and ω_i.

Remark 4.1. An example of the last model covered in Section 3 – the bivariate Matérn model – is not provided here. This is due to a lack of feasibility in the inference step: At the moment of the draft of this thesis, bivariate processes based on systems of SPDEs are not available in the R software package used for inference of the stated models (the R-INLA package, see Section 5.2 for more details). Consequently, as we cannot recover any interactions modelled by a bivariate Matérn model, we refrain from giving an example of that modelling option.
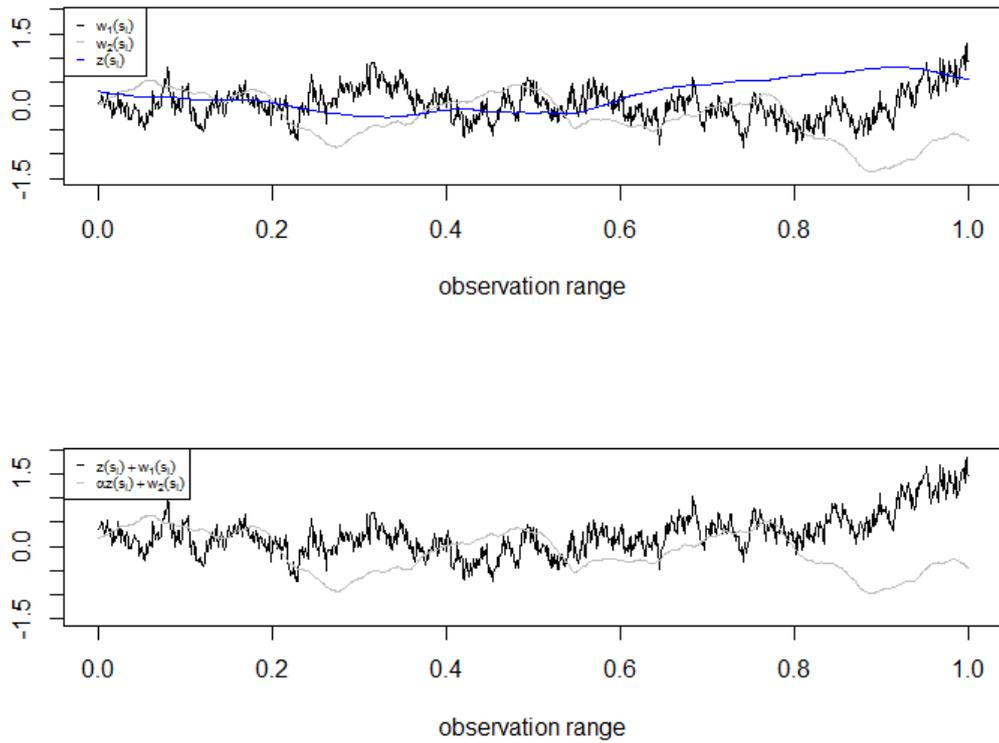
Figure 7: Modelling by bivariate Gauss-Markov processes: Simulations for the shared component example. Top: The spatially structured effects $w_1(s_i)$ (black line), $w_2(s_i)$ (grey line) and $z(s_i)$ (blue line). Bottom: The overall spatially structured effect $z(s_i) + w_1(s_i)$ (black line) for the latent process η_i and the overall spatially structured effect $az(s_i) + w_2(s_i)$ (grey line) for the latent process ω_i.

5 Inference methods

5.1 The INLA approach

Before going over to simulation studies, the "integrated nested Laplace approximation" (INLA) methodology used for fitting the models shall be presented. We will introduce the procedure only briefly, a detailed description may be found in Rue et al. [RMC09, pp.327–332].

INLA may in general be applied to each model which may be formulated as *a latent Gaussian model*. In those models, the observation variable ξ_i has to belong to the exponential family where the mean $\mu(\xi_i)$ is linked to a structured additive predictor $\eta(\xi_i)$ through a link function $g(\cdot)$ so that $g(\mu(\xi_i)) = \eta(\xi_i)$. The structured additive predictor $\eta(\xi_i)$ then accounts for various effects in an additive way, where the priors for these effects must be Gaussian. As may be seen in Section 2.1, log-Gaussian Cox processes satisfy these requirements (with a link function $g(\cdot) = \log(\cdot)$) and hence the INLA approach may be applied.

INLA relies on deterministic approximations instead of randomization as it is used in Markov chain Monte Carlo (MCMC) methods. Let $\boldsymbol{\zeta}$ denote the vector of all the parameters which are the intercepts β_1, β_2, the interaction parameters $\beta_3, a_{12}, a_{21}, a$, the spatially structured effects $f_1(s_i), f_2(s_i), w_1(s_i), w_2(s_i), z(s_i)$, the spatially unstructured effects u_i, v_i and the coefficients of a possible Bayesian P-spline regression in our setting. The vector $\boldsymbol{\theta}$ represents the vector of the hyperparameters $\sigma_{f_1}^2, \sigma_{f_2}^2, \sigma_{w_1}^2, \sigma_{w_2}^2, \sigma_z^2, \nu_1, \nu_2, \nu_3, \kappa_1, \kappa_2, \kappa_3, \sigma_u^2, \sigma_v^2$ and the variance of potential P-spline random walk priors. Naturally, not all of the noted (hyper-)parameters are present in a particular simulation study. Note that Gaussian priors must be assigned to the parameters, whereas the hyperparameters are not necessarily Gaussian distributed.

The posteriors of interest are the marginal posteriors for the latent fields $\pi(\zeta_i \mid \boldsymbol{x}, \boldsymbol{y})$ and the marginal posteriors for the hyperparameters $\pi(\theta_j \mid \boldsymbol{x}, \boldsymbol{y})$, which must be approximated by the INLA approach (see Rue et al. [RMC09, p.324]). The vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the observed Poisson counts of the first and second Poisson process here. To approximate the posteriors with INLA, the posteriors are formulated as

$$\pi(\zeta_i \mid \boldsymbol{x}, \boldsymbol{y}) = \int \pi(\zeta_i \mid \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})d\boldsymbol{\theta}, \tag{38}$$

$$\pi(\theta_j \mid \boldsymbol{x}, \boldsymbol{y}) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})d\boldsymbol{\theta}_{-j}. \tag{39}$$

Hence, the marginal posteriors $\pi(\zeta_i \mid \boldsymbol{x}, \boldsymbol{y})$ are obtained by approximating $\pi(\zeta_i \mid \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ and $\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})$ and using numerical integration to integrate out $\boldsymbol{\theta}$, for what the number of hyper-parameters should be small – preferably not higher than six (see Rue et al. [RMC09, p.321]). $\pi(\theta_j \mid \boldsymbol{x}, \boldsymbol{y})$ is also based on the approximation of $\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})$, but $\boldsymbol{\theta}_{-j}$ is integrated out instead of $\boldsymbol{\theta}$.

The approximation for $\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})$ is achieved by using the Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y}) \propto \left. \frac{\pi(\boldsymbol{\zeta}, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})} \right|_{\boldsymbol{\zeta} = \boldsymbol{\zeta}^*(\boldsymbol{\theta})}, \tag{40}$$

where $\tilde{\pi}_G(\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ is a Gaussian approximation to the full conditional for $\boldsymbol{\zeta}$ by matching the modal configuration and the curvature at the mode. The value $\boldsymbol{\zeta}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{\zeta}$ for a given hyperparameter value $\boldsymbol{\theta}$. The approximation of $\pi(\zeta_i \mid \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ is similarly based on a Laplace approximation, although not shown here.

In summary, the three different components of the INLA approach have been presented: The required integration (I) and nested formulation (N) in Equations 38 and 39 and the Laplace approximation (LA) in Equation 40.

Though the usual MCMC methods could also be used in our setting, we refrain from using them as they often exhibit a poor performance when being applied to latent Gaussian models. This has mainly two reasons (see Rue et al. [RMC09, p.322]): Firstly, the parameters $\boldsymbol{\zeta}$ of the latent process are strongly dependent on each other. Secondly, the parameters $\boldsymbol{\zeta}$ and the hyperparameters $\boldsymbol{\theta}$ are strongly dependent, which gets worse when the number of observation rises. Both facts provoke that MCMC methods need a very long running time to provide satisfiable results. As Rue et al. [RMC09, p.348] state for latent Gaussian models, the INLA approach "[...] provides precise estimates in seconds and minutes, even for models involving thousands of variables, in situations where any MCMC computation typically takes hours or even days."

However, it must be said that the INLA approach is problematic when being applied to settings where the number of hyperparameters is high as the computational cost is exponential with respect to the number of hyperparameters. But this is only an issue for applications with more than six hyperparameters, which is not the case for our models. The highest amount of hyperparameters in our simulation studies are the six hyperparameters in the shared component example.

5.2 The R-INLA package

The simulation studies presented in Section 6 were conducted by means of the "R-INLA" package, a supplementary package for the statistical software "R".

The R-INLA package can be downloaded from http://www.r-inla.org/ and provides a tool for fitting models both by using the INLA methodology introduced in the previous section *and* the SPDE approach for the approximation of Gaussian processes by Gauss-Markov processes (see Section 2.2). Note that the default method for inference within R-INLA is a simplified Laplace approximation instead of the standard Laplace approximation. However, Rue et al. [RMC09, pp.329,332] state themselves that "[...] the much smaller cost of the simplified Laplace approximation generally compensates for the slight loss in accuracy [...]" compared to the standard Laplace approximation and that "[the] simplified Laplace approximation appears to be highly accurate for many observational models". We checked the differences between the

standard and the simplified Laplace approximation for the first simulation study which is the independence example – without any visible differences in the quality of the model fit, but with a running time being almost ten times higher for the standard Laplace approximation. So we decided to use the default simplified Laplace approximation for inference.

A lot of work has been proven beneficial for understanding the R-INLA package and implementing our models. These include the ISBA Bulletin by Lindgren [Lin12] and the examples tutorial by Krainski et al. [KLSR16], both available on the R-INLA website, the publication about spatial data analysis with R-INLA by Bivand et al. [BGRR15] and the work on Bayesian spatial modelling with R-INLA by Lindgren and Rue [LR15].

We used the R version 3.2.4 from March 10, 2016 and the R-INLA version 0.0-1457943991, dated from March 14, 2016 for each of our examples. The commented R codes for the simulation and inference of our modelling methods is available on the CD appended to the present work. Moreover, a README.txt explaining the R codes (and helper functions) is attached to the CD.

5.3 Methods for model validation and comparison

In Section 6, we will conduct simulation studies to assess the quality of the different methods for interaction modelling. For this purpose, the methods will be evaluated with regard to their ability to rebuild the simulated examples based on each of the interaction methods. As we are in the comfortable situation of examining simulations, there are several possibilities to do this. They are announced below in the first paragraph.

Nevertheless, we also have to keep in mind that those validation methods are not applicable in practice where the underlying processes of the examined Poisson processes are mostly unknown. For this reason, we will not only introduce methods for model validation in this section, but also techniques for model comparison which should be applied in practice to help detecting the "best" – or at least a not too bad – interaction model for the situation at hand.

Methods for model validation   For each of the simulated examples of Section 4, there are several techniques to evaluate how well they are recognized when applying the appropriate model in the inference step, i.e. the interaction method they are originally based on.

First of all, the posterior marginal distributions of the parameters and hyperparameters should be compared with the true values. While this can be done using the posterior mean and an equal-tailed credible interval for the Gaussian distributed parameters, it is reasonable to apply the mode as a convenient point estimate as well as a HPD (highest posterior density) interval for the hyperparameters as the posterior marginals for the not necessarily Gaussian distributed hyperparameters can be severely skewed (see Rue et al. [RMC09, p.346]). When comparing the fitted and true values, the focus lies on the accurate estimation of the *interaction parameters*. Naturally, it is also important that the rest of the (hyper-)parameters are estimated correctly as good modelled interactions based on bad fitted underlying processes are worthless. However, the most important part of our models are the interactions which consequently should be estimated

most accurately.

Remark 5.1. At this point, a comment on the estimation of the spatially unstructured effects must be placed. The simulation studies showed that without an unreasonable highly informative prior, the spatially unstructured effects are nearly undetectable in the inference step with INLA (one gets precisions of 10,000 and more).

Dr. Finn Lindgren replied to the author's request concerning this issue in the R-INLA discussion group. He stated that there are two possible explanations for the observed effect: Firstly, "[...] the discretised spde model approximation error indeed tends to give slightly sharper correlation for short distances [...]" and so "[...] there are parameter values for the spatial field that are better at capturing the structure than the nugget part of the model [...]". Secondly, "[...] the point process is extremely weakly informative about the small scale structure of the intensity process, so it's likely that the nugget effect has only a very small practical effect on the point pattern [...]". Consequently, "[...] [unless the] average count in each grid box is very high [one] won't be able to detect overdispersion [...]" (the nugget effect, i.e. the spatially unstructured effect, leads to overdispersion in the Poisson counts).

As we do not want to highly increase the number of counts per segment (the approximation error of the SPDE models would rise, too), we lay aside the estimation of the spatially unstructured effects. Anyway, they do not play an important role in our interaction modelling examples.

While the comparison of the fitted with the true parameters already includes the comparison of the fitted with the true spatially structured effects (as they are a part of the estimated parameters), it might also be of interest how deviations from the true parameters influence the fit of the complete latent processes. So comparisons of the fitted linear predictors (i.e. the latent processes without the spatially unstructured effects) with the true ones are another option to assess the quality of the fit. Note that the linear predictors, not the entire latent processes, should be compared, since the spatially unstructured effects are nearly undetectable in the estimation (see Remark 5.1).

To get a more objective score for the quality of the fits – both of the spatially structured effects and of the linear predictors – one can compute the MAE (mean absolute error) and MSE (mean-square error) of the estimated processes (see Hu et al. [HSS[+]15, p.8]). As the names already indicate, the first index is the mean of the summed *absolute* differences between the estimated and the true values of the process, whereas the second index is the mean of the summed *squared* differences between the estimated and the true values. Thus, one can rely on exact scores for the quality of the fits – at least as far as the posterior *means* are concerned. Furthermore, the two indices constitute a possibility to quickly compare two different fits.

Methods for model comparison   While we are able to fit and evaluate the "true" interaction model for each of our simulated examples (i.e. the interaction model the examined simulation is originally based on), a user in practical applications usually does not know the underlying structure. Actually, this is exactly the reason why interaction models shall be applied: One wants

to get an idea of the interaction – if any – between the observed point processes. Naturally, a comparison with "true" values is impossible in this case. To still find the best interaction model for the particular situation, methods for *model comparison* must be used.

Illian et al. [ISR12, p.1505] applied "[...] methods for model comparison based on the deviance information criterion (DIC) [...] to compare different models with different levels of complexity". The DIC is "a somewhat Bayesian version of AIC", where the maximum likelihood estimate in the formula is replaced with the posterior mean. Moreover, the penalizing term, which normally is two times the number of parameters, is replaced with an estimation of two times the *effective* number of parameters. The latter is necessary, since "[informative] prior distributions and hierarchical structures tend to reduce the amount of overfitting, compared to what would happen under simple least squares or maximum likelihood estimation" [GHV13, pp.7,8]. Consequently, the penalization would be too strict if one took the usual penalty term. Just as for the AIC, lower values of the DIC are preferable. Unfortunately, there is no general rule to determine when the difference in the DIC between two different models is crucial. This depends on the particular inference setting as well as on the data at hand. As Gelman et al. [GHV13, p.27] state, one might calibrate the differences based on a very simple *example model*. In our setting of examining interactions, a simple example is the independence example which we will use to calibrate the differences in the DIC (see Section 6.1). For further information about the AIC and DIC as well as for formulas for the calculation of the effective number of parameters, see Rue et al. [RMC09, Ch.6.4] and Gelman et al. [GHV13, Ch.3.2,3.3].

Although we will use the DIC as a tool for model comparison in Section 6, we will not solely rely on that measure as it has the drawback to take only a point estimate – the posterior mean – into account instead of the entire posterior distribution (which would be desirable). A criterion which is indeed based on the complete posterior distribution is the WAIC (called the widely applicable information criterion). The WAIC is a fully Bayesian approach, again with a correction for the effective number of parameters, and constitutes an approximation to cross-validation. The reason why we do not directly use a cross-validation scheme is that it would be computational expensive on the one hand, on the other hand, cross-validation is not well defined in our setting with dependent data (see Gelman et al. [GHV13, p.28]). The dependent data might also be an issue for the WAIC, which is the reason why we will not only rely on the WAIC either. Again, lower values of the WAIC are preferable and we calibrate the differences on the independence example. Further information about the WAIC is available in Gelman et al. [GHV13, Ch.3.4].

The third criterion we use for model comparison is the conditional predictive ordinate (CPO) (see Rue et al. [RMC09, Ch.6.3]). The CPO is a "leave-one-out" predictive measure of fit which is based on the predictive density $\pi(q_i \mid \boldsymbol{q}_{-i})$ for the observed q_i based on all the other observations. Here, $\boldsymbol{q}$ represents the vector of Poisson counts for *both* processes, i.e. $\boldsymbol{q}$ contains all the Poisson counts x_i *and* y_i. While the CPO values are a helpful tool for model validation – namely to detect surprising observations which are related to small CPO values – the mean of the CPO values may also be used to compare between two different models for a given dataset. Obviously,

higher means are preferable and the difference will be calibrated as before on the independence example.

We want to emphasize that the goal is not to choose a single model based on a combination of the DIC, WAIC and CPO, but rather to give some devices to narrow the quantity of models to take into account in practical applications. As the title of the paragraph already implies, the focus lies on *model comparison*, not on model selection. For the task of choosing a final model, there are a lot of other aspects to be considered as the particular situation (for example one might want to detect a common underlying factor of the two Poisson counts which would lead to the shared component model) or the running time of the models (which we will also examine in Section 6). A detailed discussion on when to choose which model will be conducted in Section 7.

6 Simulation studies

Before examining the simulated examples from Section 4, some general comments on the inference approach shall be made. First of all, the inference is done on the same grid as has been used for the simulation of the examples, i.e. on 1000 equally spaced segments, each of length 0.001, covering the observation range from 0 to 1. This yields ratios between 0.1 and 0.5 for the proportion of points to grid segments in our examples, which corresponds to the ratios in other (practical) studies. For instance, the ratios lie between 0.1 and 1.5 for the applications in Rue et al. [RMC09, Ch.5.5], Illian et al. [ISR12, Ch.4,5] and Illian et al. [ISRH12, Ch.3]. There are a lot of empty segments with our approach, what may lead to the idea of implementing an irregular spaced grid to get a finer resolution at locations with many points and a rougher one at locations with only a few points. However, as Hu and Simpson state in the discussion of Lindgren et al. [LRL11, p.493], this complication is not necessary as "[...] both the point pattern and the 'empty space' provide important information."

Furthermore, we use a joint modelling approach for the two point patterns whenever this is possible in the inference of the simulated examples. This procedure has the advantage of a common likelihood for the two point patterns and hence of unified parameters and hyperparameters, which induces a better fit. Actually, we are able to use a joint inference approach for all our examples with the exception of the non-linear models, where the base values of the first point pattern are needed to conduct the Bayesian P-spline approach. In the latter case, the MAE, MSE and model comparison scores of the two individual fits are combined in the analysis.

Besides, sum-to-zero constraints are performed in the inference for the spatially structured effects as well as for the P-splines to separate the effect of the various components. That is, the estimates of the intercepts β_1 and β_2 also include the means of the spatially effects – which have been simulated with an expected value of zero but have nevertheless means not equal to zero – and of the P-splines.

Another issue, not only in the INLA approach but in every inference setting containing Matérn models, is the identifiability of the smoothness parameter ν. Usually, the smoothness parameter is fixed in advance (according to an a priori belief of the smoothness of the process or to a favoured value concerning the smoothness of the trend one wants to detect) as it is hardly identifiable (see Lindgren et al. [LRL11, pp.426,470] and Hu et al. [HSS+15, p.5]). For the inference step in R-INLA, the smoothness parameter – in our situation the *two* smoothness parameters – must be fixed in advance, too, since the inference function is not able to estimate the smoothness parameter(s). We are once again in the comfortable position of being able to fix the two smoothness parameters ν_1 and ν_2 to our (simulated) known values, which we will do in the following simulation studies. Nevertheless, we also want to examine what one may do in practice to get an idea of the smoothness of the underlying processes. Although this is even more challenging for Poisson processes than for Gaussian processes as point processes are extremely weakly informative about the small scale structure of the latent processes (see Remark 5.1), we will try in Section 6.1 to approximately identify the smoothness parameters ν_1 and ν_2 for the

independence example as a benchmark. As the smoothness parameters have to be fixed before the inference step in R-INLA, this will be done by fitting several models with reasonable predefined values for ν_1 and ν_2 and comparing the results using the methods for model comparison presented in Section 5.3.

Finally, there is the task of specifying appropriate priors for the rest of the (hyper-)parameters. In general, prior specification is highly dependent on the particular application setting as the prior shall incorporate a priori knowledge and believes. Prior distributions which may lead to heavy overfitting in one scenario can be just reasonable in another.

In our case, we will use vague non-informative priors in the inference step like it is done in a lot of other essays about inference for Poisson processes or SPDE models (see e.g. Rue et al. [RMC09, p.343], Illian et al. [ISR12, p.1506], Illian et al. [ISRH12, Ch.2.3,2.4,3] and Hu et al. [HSLR13, p.30]). For most of our examples, the default priors of R-INLA are therefore an appropriate choice, which are stated below. Variations from these priors will be explained in the particular simulation studies. For instance, the prior for the spatially structured effect of the second Poisson process in the non-linear dependence example has to be modified to prevent the spatially structured effect from overfitting due to incorporation of parts of the non-linear relationship (see Section 6.2).

Remark 6.1. The fact that we basically use the default priors is due to an assumed non-informative setting and means *in no way* that prior specification is unimportant. Users of our interaction models must be aware that all the available a priori information and the characteristics of the setting at hand have to be incorporated into the priors of the (hyper-)parameters to get ideal results.

The default priors The default distributions for the (interaction-)parameters in R-INLA are

$$\beta_1, \beta_2 \propto \text{constant}, \tag{41}$$

$$\beta_3, a_{12}, a_{21}, a \sim \mathcal{N}(0, 1000), \tag{42}$$

and the priors for the hyperparameters are given by

$$\log(\tau_1), \log(\tau_2), \log(\tau_3) \sim \mathcal{N}(b_1^{(\nu_1)/(\nu_2)/(\nu_3)}, 10), \tag{43}$$

$$\log(\kappa_1), \log(\kappa_2), \log(\kappa_3) \sim \mathcal{N}(b_2^{(\nu_1)/(\nu_2)/(\nu_3)}, 10). \tag{44}$$

Here, the τ are precision parameters of the spatially structured effects which are related to the marginal variances by the equation (see Lindgren and Rue [LR15, p.3])

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}, \tag{45}$$

with α extracted from the SPDE approximation (see Section 2.2), i.e. $\alpha = \nu + d/2$ and $d = 1$ in

our case as d is the dimension. Consequently, τ, κ and the a priori fixed ν determine together the marginal variance through Equation 45 and a distribution for σ^2 need not to be specified. As for the means $b_1^{(\nu_1)/(\nu_2)/(\nu_3)}$ and $b_2^{(\nu_1)/(\nu_2)/(\nu_3)}$ in Equations 43 and 44, there is no documentation on how they are exactly computed. At least, it is known that they are dependent on the value of the particular ν and it is possible to get the exact values for a specified inference setting. Although there is no proper documentation, we keep them as default values as they have the great advantage of being "chosen heuristically to match the spatial scale of the mesh domain" (response from Dr. Finn Lindgren in the R-INLA discussion group to the theme "Using SPDE from Lindgren et al. (2011)" on September 9, 2012). So one may arbitrarily change the resolution of the grid without having to specify each time a new prior distribution for the (SPDE approximation based) spatially structured effects. More importantly, the default priors for the spatially structured effects worked well in the inference of our simulation examples. However, we will provide the explicit values of $b_1^{(\nu_1)/(\nu_2)/(\nu_3)}$ and $b_2^{(\nu_1)/(\nu_2)/(\nu_3)}$ for each of our studied examples. Note that we have not specified any prior distributions for the hyperparameters σ_u^2 and σ_v^2 as we lay aside the estimation of the spatially unstructured effects due to identifiability issues (see Remark 5.1). Moreover, we have not denoted the distributions for the P-spline random walk priors yet, which we will do in the specific non-linear dependence examples.

6.1 The independence example

Model validation   Our first simulation study is that of the independence example presented in Section 4.1. To evaluate how well it is recognized, we fit the independence model in the inference step. Although we will try to identify the smoothness parameters ν_1 and ν_2 later on (see the paragraph for model comparison below), we initially fix them to the true values $\nu_1 = 0.5$ and $\nu_2 = 1.5$ for the sake of model validation. The priors for the remaining (hyper-)parameters of Table 1 (without the variances of the spatially unstructured effects) are determined by the default priors of R-INLA denoted in Equations 41 to 44 and by Equation 45. Note that for the current independence setting there are no interaction parameters to focus on. Moreover, no variations from the default priors are necessary. The values of $b_1^{(\nu_1)/(\nu_2)}$ and $b_2^{(\nu_1)/(\nu_2)}$ are dependent on the values of ν_1 and ν_2 as well as on the grid and are given for the present independence example by

$$\log(\tau_1) \sim \mathcal{N}(-2.131122, 10), \tag{46}$$

$$\log(\kappa_1) \sim \mathcal{N}(2.303586, 10), \tag{47}$$

$$\log(\tau_2) \sim \mathcal{N}(-5.605241, 10), \tag{48}$$

$$\log(\kappa_2) \sim \mathcal{N}(2.852892, 10). \tag{49}$$

Fitting the independence model took 209 seconds to run on a Intel Pentium CPU B950 2 $\times$ 2.1 GHz processor with 4 GB RAM.

Remark 6.2. For the sake of simplicity, we do not write the posteriors for the (hyper-)parameters

as $\sigma^2_{f_1} \mid \boldsymbol{x}, \boldsymbol{y}$ and $\kappa_1 \mid \boldsymbol{x}, \boldsymbol{y}$ and so on (with $\boldsymbol{x}$ and $\boldsymbol{y}$ denoting the observed Poisson counts of the first and second Poisson process, respectively) but rather as $\sigma^2_{f_1}$, κ_1 etc. like before for the priors. We want to point out that if we talk about the (hyper-)parameters *before* the model fitting process, we think of the *priors*, and if we talk of them *after* the fit, we think of the *posteriors*. We will state the meaning of the formulation in each situation where uncertainness might arise.

The first part of the model validation step is a comparison of the posterior marginal distributions for the (hyper-)parameters with the values on which the simulation is based – the "true" values. This is done in a graphical way in Figure 8 (without the parameters of the spatially structured effects which are presented later and without the not estimated (hyper-)parameters of the spatially unstructured effects).

The figure shows the posterior distributions for the (hyper-)parameters of the first Poisson process in the four top panels, the posterior distributions of the (hyper-)parameters of the second Poisson process in the four bottom panels. Some explanations are needed when considering the panels. Firstly, not the posteriors for the parameters β_1 and β_2 are presented in the first and the fifth panel, but the posteriors for the mean of the latent processes, since sum-to-zero constraints are performed in the inference step to separate the effects of the various components (see Section 6). That is, the *estimated* posteriors for β_1 and β_2 are actually posteriors for the means of the latent processes and therefore indicated as mean(η) and mean(ω). Secondly, the fourth and the eighth panel show the nominal ranges of the first and the second latent process, which are technically none of the parameters that have to be estimated. Nevertheless, they give a useful insight into the correctness of the estimation as the range represents the distance where two observations are almost independent (see Section 2.2 for a definition and an interpretation of the range parameter).

Concerning the interpretation of the panels, the grey lines always indicate the true values of the (hyper-)parameters of the simulated processes. The solid black lines in the first and the fifth panel represent the means of the posterior marginal distributions for the means of the latent processes while the dashed lines stand for the 2.5% and 97.5% quantiles of an equal-tailed credible interval. For the rest of the panels, the dashed lines indicate the borders of a 95% HPD credible interval. The reason for this has already been announced in Section 5.3: While mean(η) and mean(ω) are Gaussian distributed parameters, the posterior marginal distributions for the variance, the scaling and the range hyperparameters are highly skewed (as may actually be seen in Figure 8) and consequently the modes and HPD intervals are a better choice for comparisons of the distributions with the true hyperparameters.

The panels demonstrate that for each of the eight considered parameters the true value lies within the particular 95% credible interval, which is an indication for a good model fit. Furthermore, the modes (i.e. the point estimates for the hyperparameters) for the posterior distributions of the marginal variance $\sigma^2_{f_1}$, the nominal range for the first process and the scaling parameter κ_2 are almost perfect estimates of the true values (although one has to be cautious because of the
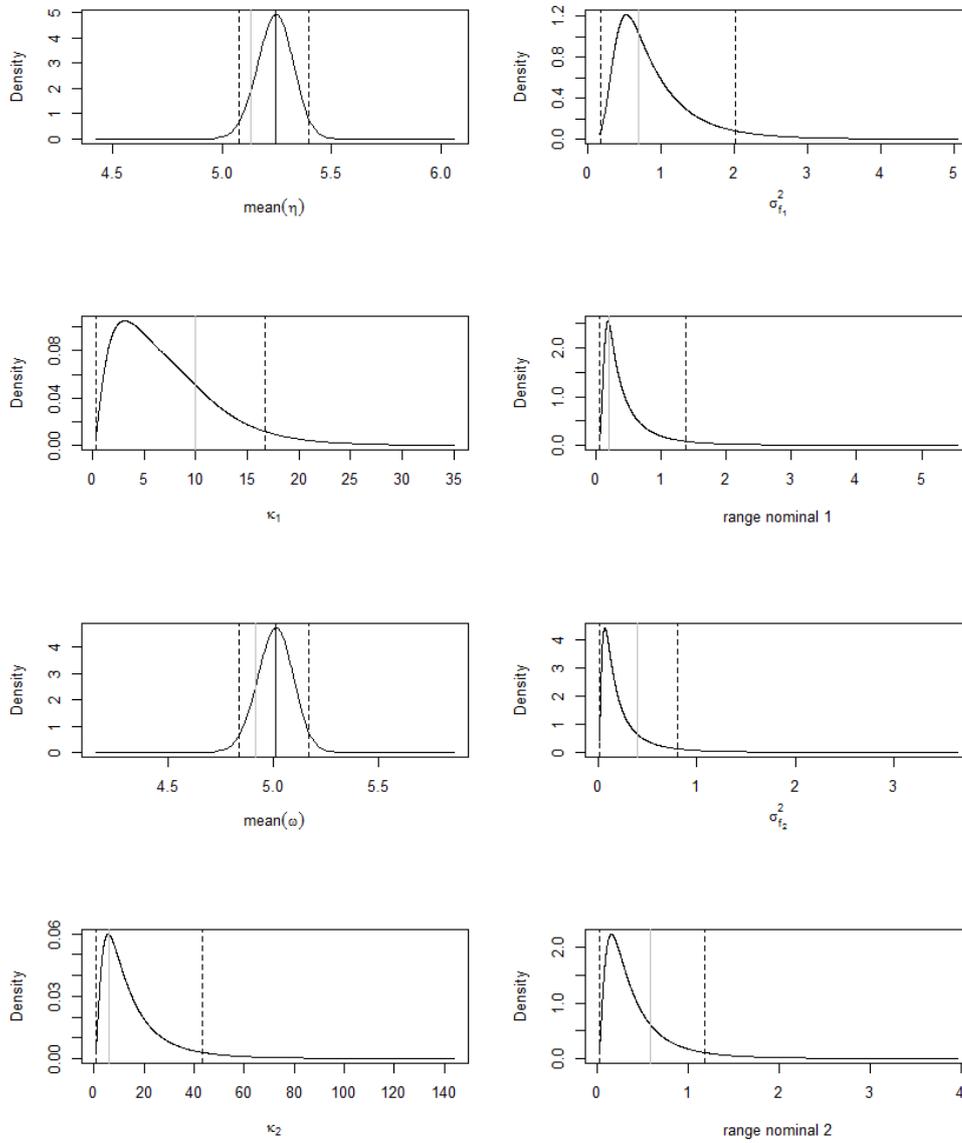
Figure 8: Modelling by functional connections: Posterior marginal distributions for the parameters and hyperparameters of the independence example. Four top panels: The (hyper-)parameters for the first Poisson process. Four bottom panels: The (hyper-)parameters for the second Poisson process. The true values are indicated by a grey line in each of the panels. For the first and the fifth panel, the mean of the posterior is marked by a solid black line, the 2.5% and 97.5% quantiles of an equal-tailed credible interval by a dashed line. For the remaining panels, the dashed lines indicate the borders of a 95% HPD credible interval.

different domains of the parameters).

We refrain from further investigations of the figure as the full potential of the analysis method will only be revealed *when comparing the fit of the independence model with the fits of the other interaction models* in the following sections. Then one may examine if the inclusion of interactions into the model debases the estimates of the other parameters. That is, the current model will constitute the *baseline model for model validation.*

The same applies for the estimates of the parameters for the spatially structured effects $f_1(s_i)$ and $f_2(s_i)$, which are illustrated in Figure 9. We call them the posterior marginal distributions for the parameters $f_1(s_i)$ and $f_2(s_i)$ in the following even though only the posterior means and 95% equal-tailed credible intervals instead of the whole distributions (which would be thousand distributions per spatially structured effect) are shown in the figure.

Again, the model fit seems quite good as the true spatially structured effects (indicated by grey lines) are nearly everywhere inside the 95% equal-tailed credible intervals (marked by dashed lines). Since there are 1000 segments, the spatially structured effects may lie 50 times outside the 95% credible intervals on average, but they do so only 39 times for the spatially structured effect $f_1(s_i)$ and not at all for the effect $f_2(s_i)$. However, the posterior means (indicated by solid black lines) are not everywhere a good point estimate for the true values but rather a "smoother" of the spatially structured effects – a result which may actually be favoured in many applications. This observation is caused by the non-informative priors for the hyperparameters (partly dependent on the values of ν_1 and ν_2), but also by the Poisson processes which are extremely weakly informative about the small scale structure of the latent processes (see Remark 5.1).

Concerning the fits of the linear predictors of the latent processes η_i and ω_i in Figure 10, there are nearly no new insights to gain in the independence example, since both linear predictors are only dependent on one of the two spatially structured effects.

It is hardly visible in the figure that the 95% equal-tailed credible intervals get slightly wider due to the uncertainty regarding the mean of the latent processes. Furthermore, the posterior means of the linear predictor values are a little bit shifted compared with the true linear predictors as the point estimates of the means of the latent processes are not completely correct (see panels one and five of Figure 8). Nevertheless, the fits of the linear predictors will become more important for the interaction models examined in the following sections.

The last model validation step is the calculation of the mean absolute error (MAE) and mean-square error (MSE) for the estimated spatially structured effects and the linear predictors (see Section 5.3). The (rounded) computed scores are shown in Table 4.

The scores are very similar for the first spatially structured effect $f_1(s_i)$ and the linear predictor of the first latent process η_i as the latent process only relies on the spatially structured effect $f_1(s_i)$. The reason for the small difference is once again that the mean estimate of η_i is not completely correct. Analogously, the MAE and MSE values for the second spatially structured effect $f_2(s_i)$ and the linear predictor of the second latent process ω_i are very similar. Concerning
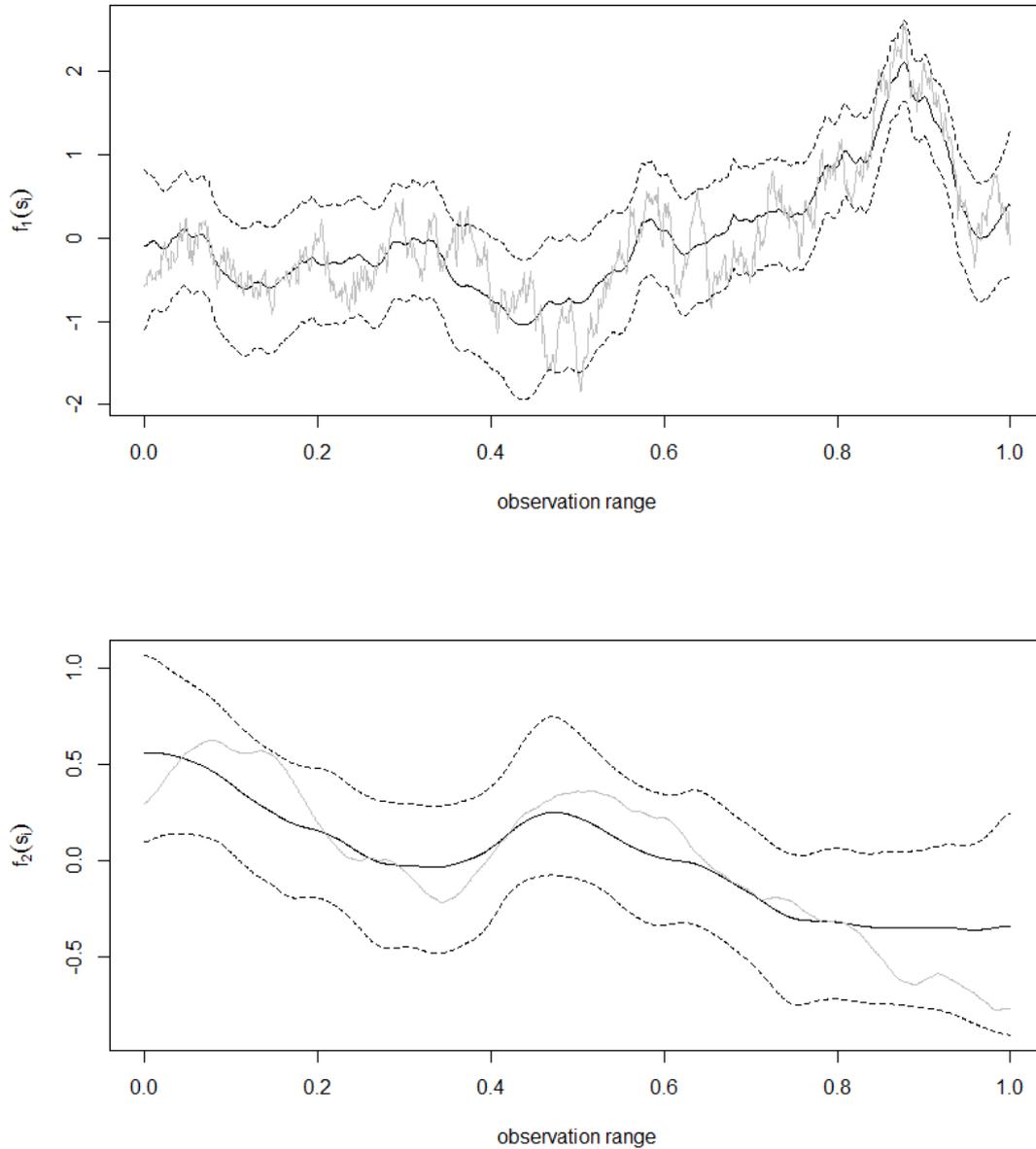
Figure 9: Modelling by functional connections: Posterior marginal distributions for the spatially structured effects of the independence example. Top: The posterior mean for the parameters of the spatially structured effect $f_1(s_i)$ (solid black line), the 2.5% and 97.5% quantiles of an equal-tailed credible interval (dashed lines) and the true simulated spatially structured effect (solid grey line). Bottom: The posterior mean for the parameters of the spatially structured effect $f_2(s_i)$ (solid black line), the 2.5% and 97.5% quantiles of an equal-tailed credible interval (dashed lines) and the true simulated spatially structured effect (solid grey line).
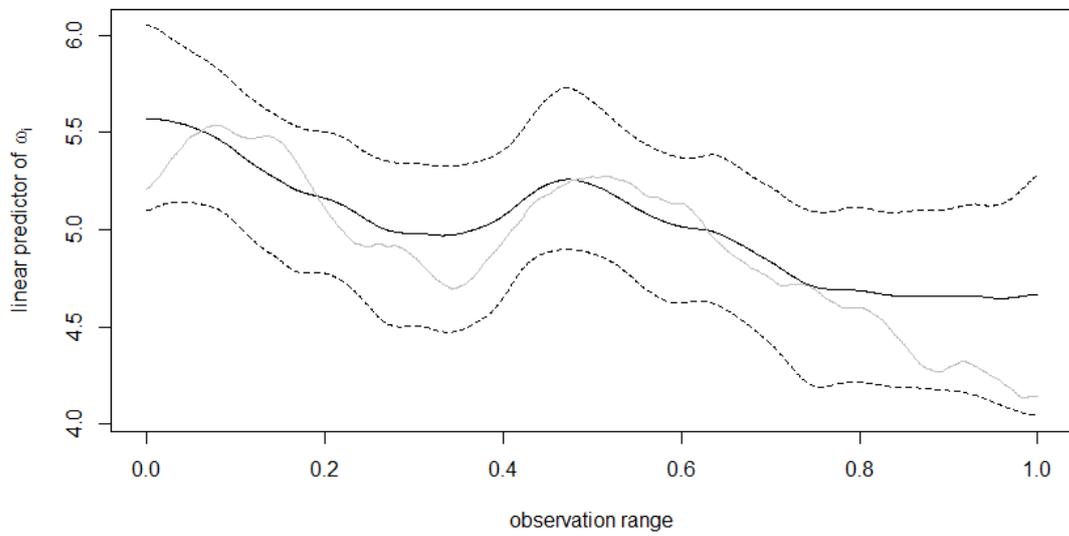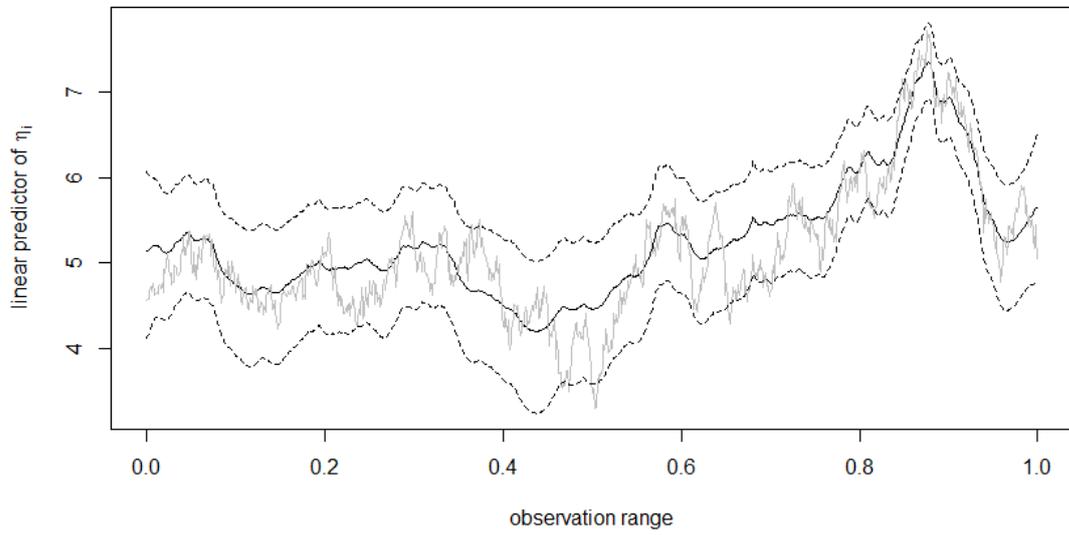
Figure 10: Modelling by functional connections: Posterior marginal distributions for the linear predictors of the latent processes in the independence example. The notation is the same as in Figure 9.

Table 4: Modelling by functional connections: MAEs and MSEs for the point estimates of the independence example.

Process	$f_1(s_i)$	$f_2(s_i)$	linear predictor of η_i	linear predictor of ω_i
MAE	0.259	0.132	0.276	0.153
MSE	0.107	0.029	0.123	0.039

the particular values, it is difficult to interpret them without the MAE and MSE scores of similar examples. We will compare them with the scores of the fitted examples in the following sections to get an idea of how the quality of the model fit changes when including interaction parameters. Note that the MAE and the MSE values are *incomplete* measures of the quality of the fits as they only take the posterior means into account, not the entire posterior distribution.

Model comparison While we are able to compare the fits with the simulated examples, a user in practical applications cannot do this. If one would like to detect the underlying interaction structure of two Poisson processes at hand, other methods must be applied to find an appropriate model for the particular situation. In Section 5.3 we presented the DIC, the WAIC and the CPO as tools to narrow the quantity of models to take into account. In the following, we will fit each of the interaction models presented in Section 3 to the current independence example, except for the dependence on the latent process and the dependence on the Poisson process models (and the bivariate Matérn model which we already excluded in Remark 4.1). We will compare the DIC, WAIC and CPO values and examine if the correct interaction model for the current example – which is the independence model – can be detected by them.

Remark 6.3. Usually, the great extent of available interaction models to be considered is already reduced in advance dependent on the situation at hand. (We partially do this by excluding the dependence on the latent process and Poisson process models which have to be well justified to be applied.) Afterwards, only the "interesting" interaction models for the particular application are compared by means of the DIC, WAIC and CPO. Even then, the user should not blindly choose the "best" model according to the DIC, WAIC and CPO. Rather, the three scores should be seen as a tool to eliminate the interaction models which produce a bad fit for the application at hand. The final decision for a single model should rely on a plurality of aspects, with the model comparison scores discussed here being only a part of.

The fitted models together with their DIC, WAIC and (rounded) CPO values are presented in Table 5. Additionally, the running time of the models is indicated.

Before analysing the table, note that for all our fitted models we assume that each of the two latent processes is dependent on at least one spatially structured effect. Although not shown here, we tested for the independence example what is happening if one tries to fit models with one or both latent processes not being dependent on a spatially structured effect. The results revealed differences in the DIC and WAIC which were high enough to exclude those models from

41

Table 5: Model comparison scores and running times of several fitted models for the independence example.

Model	Formula	DIC	WAIC	CPO	Time (sec)
Independence	$\eta_i = \beta_1 + f_1(s_i)$ $\omega_i = \beta_2 + f_2(s_i)$	2068.37	2070.36	0.7178	209
Simplified linear dependence	$\eta_i = \beta_1 + f_1(s_i)$ $\omega_i = \beta_2 + \beta_3 f_1(s_i)$	2075.43	2077.35	0.7170	106
Linear dependence 1	$\eta_i = \beta_1 + f_1(s_i)$ $\omega_i = \beta_2 + \beta_3 f_1(s_i) + f_2(s_i)$	2069.89	2071.79	0.7178	372
Linear dependence 2	$\eta_i = \beta_1 + f_1(s_i) + \beta_3 f_2(s_i)$ $\omega_i = \beta_2 + f_2(s_i)$	2069.22	2071.17	0.7177	381
Simplified non-linear dependence 1	$\eta_i = \beta_1 + f_1(s_i)$ $\omega_i = \beta_2 + g(f_1(s_i))$	2076.50	2078.34	0.7169	133
Simplified non-linear dependence 2	$\eta_i = \beta_1 + g(f_2(s_i))$ $\omega_i = \beta_2 + f_2(s_i)$	2123.48	2124.18	0.7118	244
Non-linear dependence 1	$\eta_i = \beta_1 + f_1(s_i)$ $\omega_i = \beta_2 + g(f_1(s_i)) + f_2(s_i)$	2069.66	2071.82	0.7179	188
Non-linear dependence 2	$\eta_i = \beta_1 + f_1(s_i) + g(f_2(s_i))$ $\omega_i = \beta_2 + f_2(s_i)$	2068.74	2070.85	0.7178	279
LMC	$\eta_i = \beta_1 + w_1(s_i) + a_{12}w_2(s_i)$ $\omega_i = \beta_2 + a_{21}w_1(s_i) + w_2(s_i)$	2070.42	2072.30	0.7177	982
Shared component	$\eta_i = \beta_1 + z(s_i) + w_1(s_i)$ $\omega_i = \beta_2 + az(s_i) + w_2(s_i)$	2072.64	2076.07	0.7187	1912

further investigations (see below for a discussion on the differences for which we decide not to consider a model any more).

Moreover, the priors for the (hyper-)parameters of the fitted models haven't been stated yet. For each of the models, we used exactly the priors that have been shown useful in fitting their simulated counterparts. So the priors may be seen in the model validation steps of Sections 6.1 to 6.4.

Concerning the computed scores for the fitted models, the "true" model for the present situation – the independence model – is indeed identified as the best one regarding the DIC and WAIC values. Nonetheless, there are several models with very similar scores to those of the independence model, these scores are indicated by a medium grey in Table 5. The graduation of "similarity" in this context was calibrated by the quality of the fits, which may be examined in Figure 11 exemplarily for the first linear dependence model, the shared component model and the first simplified non-linear dependence model. For the sake of simplicity, only the linear predictor of the second latent process ω_i is shown. While the first linear dependence model (top panel) leads to a fit which is similar to that of the independence model presented at the bottom of Figure 10 – albeit less smooth –, the fit by means of the shared component model (middle panel) is unnecessarily "wiggled". The first simplified non-linear dependence model (bottom panel) fails completely to produce a passable fit of the independence example.

Examining the DIC and WAIC, one may then classify fits with values within a range of 2 as very similar and values with a discrepancy between 2 and 5 as similar enough to still being considered (the latter are marked by a light grey). For the CPO values, a margin of 0.0005 seems appropriate. However, there is an "outlier" recognizable as the CPO value of the shared component model (marked by a dark grey) is 0.0009 higher than the CPO of the independence model. While this would initially be an indicator for the assumption that the shared component model might give a better fit for the independence example than the independence model itself, this is contradicted by the poor DIC and WAIC values. More importantly, the plot in the middle of Figure 11 points out that the fit by means of the shared component model may not be seen as the best one. Consequently, the shared component model represents a case of ambiguity for the present example. This fact confirms our recommendation not to solely rely on model comparison scores for the goal of model selection.

Considering the complete table, it is visible that all the models which are not simplified – i.e. that include two spatially structured terms in the formula – are similar enough to the independence model to be considered. This is reasonable as those models are able to reflect the independence structure of the example, although containing some useless dependency parameters which may somewhat alter the fit. We recommend to concentrate on the simplest available model when considering models with similar comparison scores. In the present case, this would be the "correct" independence model.

We should also remark that the running time of the LMC and the shared component model may get very high in practical applications with much larger point patterns. We will come back to
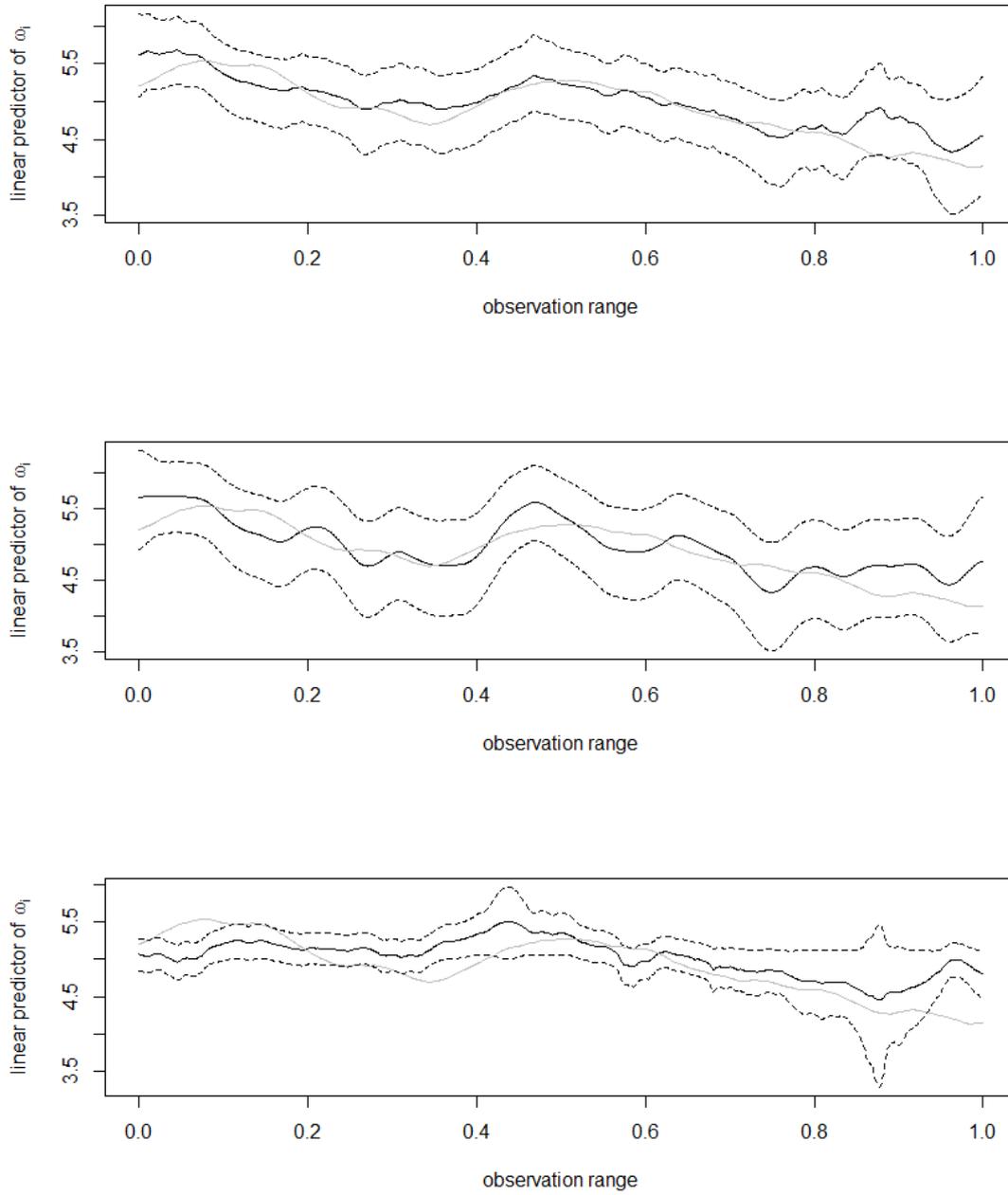
Figure 11: Model comparison: Posterior marginal distributions for the linear predictor of the second latent process ω_i in the independence example, fitted by means of the first linear dependence model (top), the shared component model (middle) and the first simplified non-linear dependence model (bottom). The notation is the same as in Figure 9.

Table 6: Model comparison scores for the independence model fitted with several selected values for the smoothness parameters ν_1 and ν_2.

ν_1 \\ ν_2	0.5	1.25	1.4	1.5
0.25		DIC: — WAIC: — CPO: —		DIC: — WAIC: — CPO: —
0.4			DIC: 2068.43 WAIC: 2070.41 CPO: 0.7178	DIC: 2068.40 WAIC: 2070.38 CPO: 0.7178
0.5	DIC: 2068.36 WAIC: 2070.30 CPO: 0.7179	DIC: 2068.48 WAIC: 2070.47 CPO: 0.7178	DIC: 2068.40 WAIC: 2070.39 CPO: 0.7178	DIC: 2068.37 WAIC: 2070.36 CPO: 0.7178
0.6			DIC: 2069.90 WAIC: 2071.53 CPO: 0.7175	DIC: 2069.86 WAIC: 2071.49 CPO: 0.7176
0.75		DIC: 2070.65 WAIC: 2072.16 CPO: 0.7174		DIC: 2070.57 WAIC: 2072.09 CPO: 0.7174
1.5	DIC: 2071.48 WAIC: 2072.82 CPO: 0.7174			DIC: 2071.52 WAIC: 2072.91 CPO: 0.7173

this point when discussing the advantages and drawbacks of the models in Section 7.

As announced previously, we now address ourselves to the task of (approximately) identifying the smoothness parameters of the two Poisson processes. Recall that the true values for the example at hand are $\nu_1 = 0.5$ and $\nu_2 = 1.5$, which allow an explicit link between Gaussian processes and Gauss-Markov processes. We name those models EL-models ("explicit link models") in the following.

We selected some interesting values for ν and fitted the independence model with these values. Note that in the one-dimensional case values between -0.5 and 1.5 are possible to fit in R-INLA with a reasonable interpretation only being available for positive values of ν (see Lindgren and Rue [LR15, p.3]). We limited ourselves to the latter case. The model comparison scores for the different fitted independence models are presented in Table 6. Note that there are no values available for $\nu_1 = 0.25$ as the Newton-Raphson algorithm did not converge.

The true model with $\nu_1 = 0.5$ and $\nu_2 = 1.5$ is marked by a medium grey. It has the best model comparison scores of the considered models apart from the model with $\nu_1 = \nu_2 = 0.5$ (indicated by a dark grey), which is slightly better. When applying the rules developed during the comparison of the different modelling types (see above), at least all models with a value for ν_1 of equal or less than 0.6 must be seen as very similar.

We stated previously that the simplest available model among similar models may be preferred.

The simplest model in the case of smoothness parameter comparison is the $\nu_1 = \nu_2 = 1.5$ model (marked by a light grey) as it gives in general the "smoothest trend". For the present example, though, the difference in comparison scores compared to the "best" model is too high to reasonably choose this model. Instead, one might choose the "smoothest" model among the similar models which would be the model with $\nu_1 = 0.6$ and $\nu_2 = 1.5$. But with this one being close to the model with $\nu_1 = \nu_2 = 0.5$, which is an EL-model, one might be tempted to use the latter instead.

So what might be a reasonable rule for the goal of choosing appropriate values for the smoothness parameters? One should begin with the simplest model available in R-INLA, which is the model with $\nu_1 = \nu_2 = 1.5$ in the one-dimensional case and represents an EL-model. The next step is to compare this model to the EL-models with $\nu_1 = 0.5$ and/or $\nu_2 = 0.5$ and choosing the latter if differences in DIC or WAIC are higher than 2. Finally, one may compare the so selected model to models with values of ν being 0.5 higher (if possible) and choose the latter if the DIC/WAIC differences do not exceed a value of 2.

This strategy leads indeed to the true model in the present example. Although not shown here, we tried the strategy for simulations of the independence model including other values of ν. The strategy was successful in detecting the true values for each of the available EL-examples. For a simulation with $\nu_1 = 1$ and $\nu_2 = 1.5$ the strategy chose the EL-model with $\nu_1 = \nu_2 = 1.5$. However, as the fit of the model was very similar to that of the true model, we keep our strategy.

6.2 The linear and non-linear dependence examples

The positive linear dependence example – model validation   Concerning the positive linear dependence example (where $\beta_3 = 0.4$), we use again the default priors denoted in Equations 41 to 44. With the grid and the values of ν_1 and ν_2 remaining unchanged, $b_1^{(\nu_1)/(\nu_2)}$ and $b_2^{(\nu_1)/(\nu_2)}$ stay the same as in the independence example. Fitting the linear dependence model to the positive linear dependence example took 294 seconds.

As for the previous example, the model validation starts with an examination of the posterior marginal distributions for the (hyper-)parameters, now including an interaction parameter β_3 to focus on. In Figure 12, only the posteriors for the interaction parameter, the mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $f_2(s_i)$ are illustrated as the rest of the (hyper-)parameters show nearly no differences to those of the independence example.

Regarding the posterior for the interaction parameter β_3, it is visible that the posterior mean is a good point estimate for the true interaction parameter (the posterior mean of β_3 is 0.475 while the true parameter is $\beta_3 = 0.4$). Moreover, the interaction is *significant* as the value 0 does not lie inside the equal-tailed 95% credible interval. So the interaction between the two Poisson processes is recognized rather well.

The inclusion of an interaction parameter should not debase the estimates for the remaining (hyper-)parameters. As we already stated, some of the estimates for the (hyper-)parameter
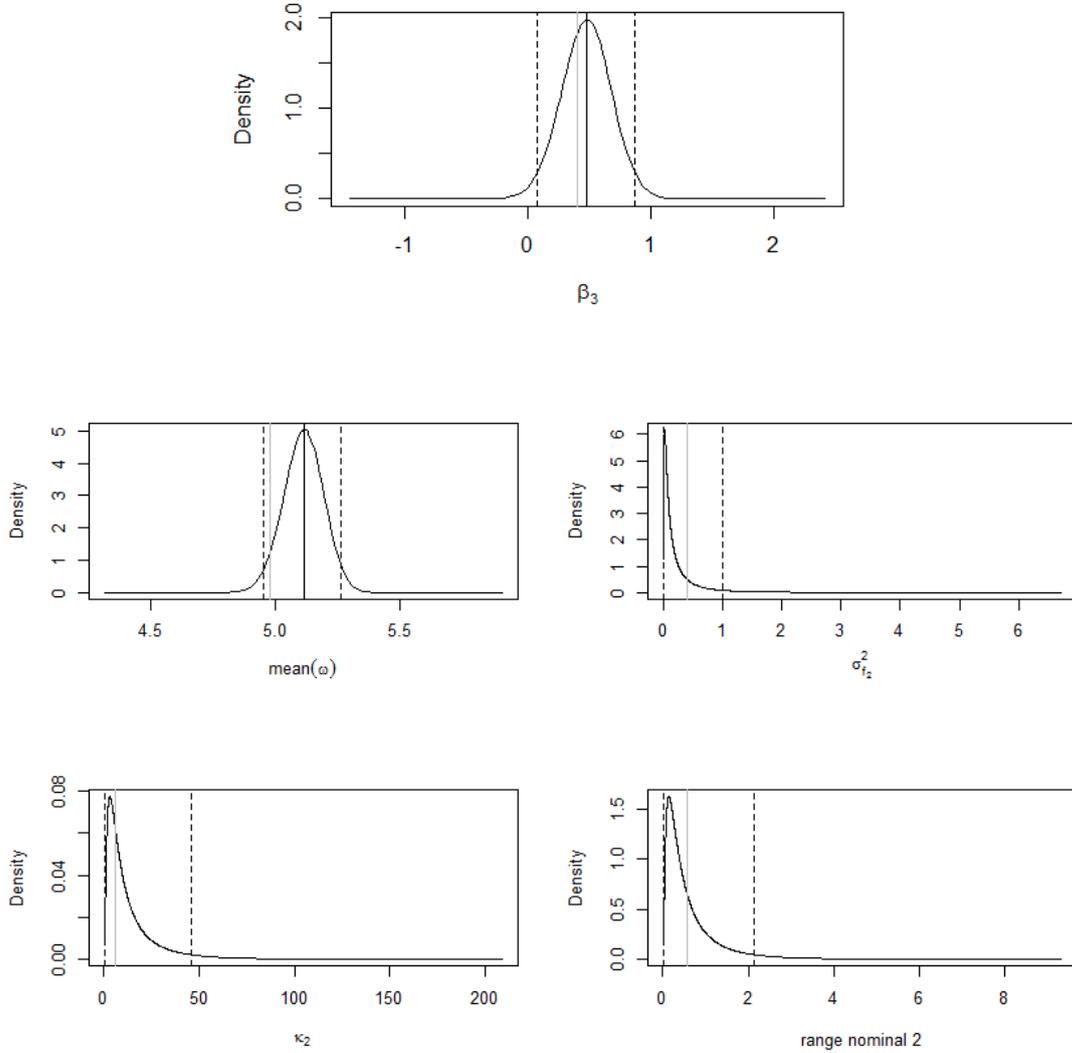
Figure 12: Modelling by functional connections: Posterior marginal distributions for the parameters and hyperparameters of the positive linear dependence example. Top panel: The interaction parameter. Four bottom panels: The mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $f_2(s_i)$. For the interaction parameter, the true value is indicated by a grey line, the mean of the posterior by a solid black line, the 2.5% and 97.5% quantiles of an equal-tailed credible interval by a dashed line. For the remaining panels, the notation is the same as in Figure 8.

posteriors do not show any differences to those of the independence example. The four bottom panels of Figure 12 illustrate the posteriors for the (hyper-)parameters with some differences in the fit between the independence and the positive linear dependence example. Overall, the results in estimating the (hyper-)parameters are equal in quality for the two examples (see Figure 8 for the fitted posterior marginal distributions of the independence example). Again, the true values lie inside the particular 95% credible intervals. The point estimates are similar good for the positive linear dependence example as for the independence example. Nonetheless, there is a visible difference: The variances of the posterior marginal distributions – and simultaneously the spreads of the credible intervals – are higher for the positive linear dependence example. This implies that the estimates of the (hyper-)parameters for the positive linear dependence example may be less clear than those for the independence example.

When considering the fits of the spatially structured effects and the linear predictors, the posterior marginal distributions for $f_1(s_i)$ and η_i (not represented here) show almost no differences to those of the independence example illustrated in the top panels of Figures 9 and 10. The posteriors for the second spatially structured effect $f_2(s_i)$ and the linear predictor of the second latent process ω_i are presented in Figure 13.

The fitted process $f_2(s_i)$ for the positive linear dependence example is not very different to that of the independence example shown in the bottom panel of Figure 9, but it is even more smooth. In contrast, the posterior for the linear predictor of the latent process ω_i is much less accurate concerning the point estimates of the true simulated linear predictor than that of the independence example illustrated in the bottom panel of Figure 10. Particularly in the end of the observation range, the point estimates fail to represent the true linear predictor behaviour. However, the true linear predictor lies mostly inside the 95% credible intervals. Note that the true simulated linear predictor is different to that of the independence example as it includes an interaction with the first spatially structured effect $f_1(s_i)$. One may check in the top right panel of Figure 3 how much the linear predictor is modified by the interaction.

The MAE and MSE scores for the point estimates of the positive linear dependence example confirm our previous analysis (although not shown here). While the MAEs and MSEs of $f_1(s_i)$ and η_i rise by an amount of less than 15%, the MAE and MSE values for $f_2(s_i)$ rise by an amount of around 30% due to the further smoothing. The scores for the linear predictor of ω_i even increase by about 40% and 75% respectively, which verifies our observation of less accurate point estimates for the present example in comparison with the independence example.

The positive linear dependence example – model comparison Examining the DIC, WAIC and CPO, one can indeed identify the "true" model for the present situation, i.e. the linear dependence model with the right direction. Applying the rule developed for the independence model results in two other very similar models with the DIC and WAIC being at most 2 higher and the CPO at most 0.0005 lower. These models are the first non-linear dependence model and the LMC. Both models are able to reflect the linear dependence structure by means of
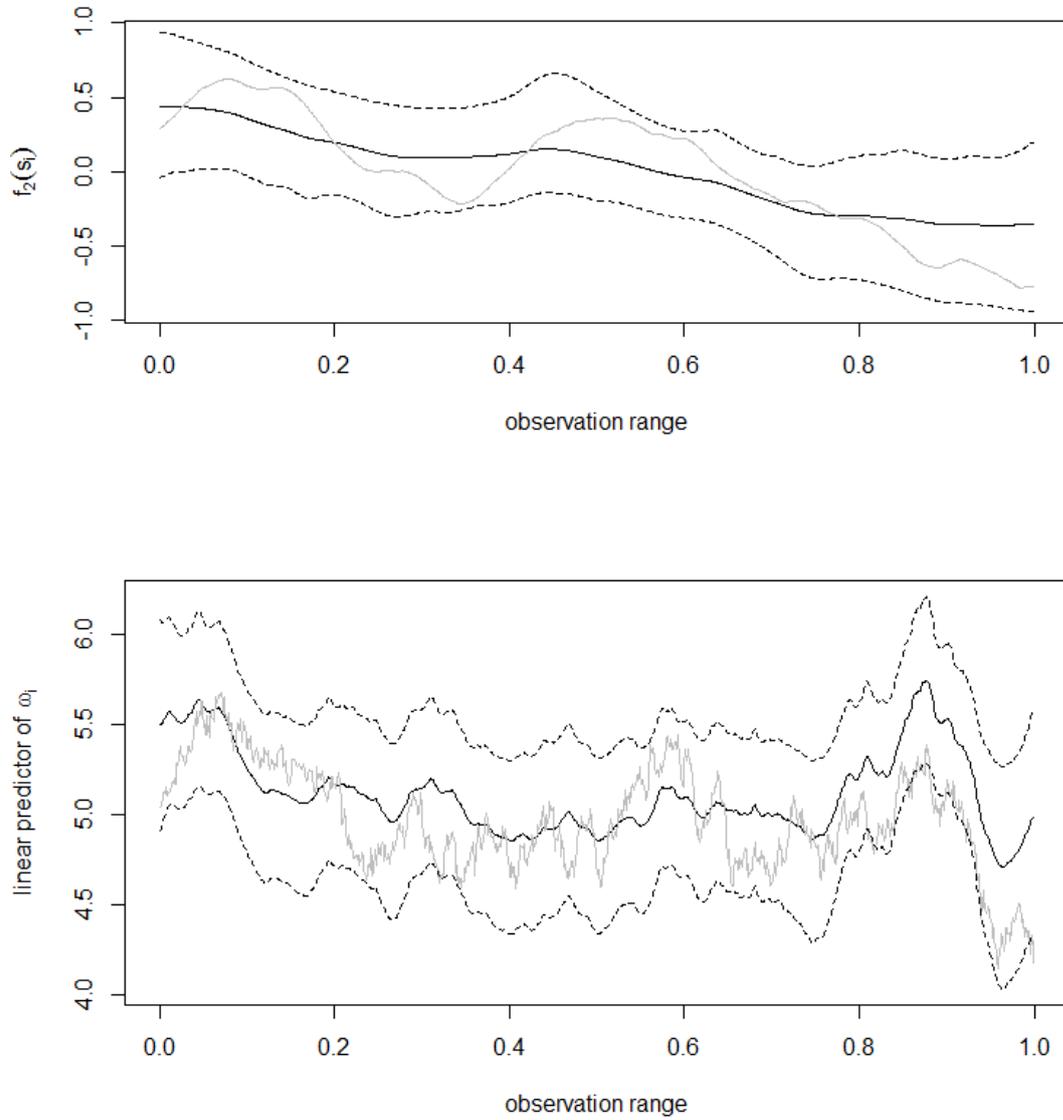
Figure 13: Modelling by functional connections: Posterior marginal distributions for the second spatially structured effect $f_2(s_i)$ (top) and the linear predictor of the second latent process ω_i (bottom) in the positive linear dependence example. The notation is the same as in Figure 9.

their formulas, but include some potentially unnecessary additional flexibility. This may be the reason why the "true" linear dependence model is actually recognized as the best one regarding the model comparison scores.

The negative linear dependence example – model validation   When moving on to the negative linear dependence example (with $\beta_3 = -0.4$), we use the same priors as for the two previous examples. Fitting the linear dependence model by means of R-INLA took 314 seconds.

The examination of the posterior marginal distributions for the (hyper-)parameters presented in Figure 14 reveals that the interaction parameter is not at all correctly estimated. The true value does not lie inside the 95% credible interval. Moreover, the point estimate is untruly positive with $\beta_3 = 0.4043$.

Concerning the other parameters shown in Figure 14, they are not estimated less accurate than in the independence example. Furthermore, the observation we have made for the positive linear dependence example – that the variances of the posterior marginal distributions and the spreads of the credible intervals are higher – does not hold here. Rather, some of the parameters have wider credible intervals ($\sigma_{f_2}^2$), other have smaller ones (κ_2 and the second nominal range). Again, only the parameters which show differences to those of the independence example are illustrated in the Figure.

The posteriors for the spatially structured effect $f_2(s_i)$, the interaction process $\beta_3 f_1(s_i)$ and the linear predictor of the second latent process ω_i (presented in Figure 15) demonstrate the complete inadequacy of the fit. Especially the middle panel points out that the interaction is estimated in the false direction which leads to an estimated curve being inverse to the true one. One may also recognize that the two estimated processes $f_2(s_i)$ and $\beta_3 f_1(s_i)$, while failing completely to estimate their true counterparts, counterbalance each other within the second latent process. Consequently, the fit for the linear predictor of ω_i is not estimated so bad, a high deviation from the true linear predictor is only visible in the end of the observation range.

The MAE and MSE scores confirm our observations. While the scores for the point estimates of the linear predictor of ω_i rise "only" by 50% and 130% compared to the independence model (it has already been 40% and 75% for the correctly estimated positive linear dependence), the MAE and MSE for the spatially structured effect $f_2(s_i)$ rise by an amount of about 190% and 660% (only around 30% for the positive linear dependence example).

In summary one has to state that the negative interaction in the simulated example cannot be estimated correctly. One may be tempted to change the prior to improve the results, but testing this hypothesis showed that this is only possible with an unreasonable high prior for β_3 of about $\mathcal{N}(-0.5, 0.01)$. The inability of the model fit to discover the negative linear interaction has also been shown not to be a one-time occurrence. Experiments with other negative values for the interaction parameter as well as with other simulations demonstrated that this is a general problem in our setting. Even in a simplified linear dependence setting without a second spatially structured effect $f_2(s_i)$, the model fit fails to detect the negative interaction.
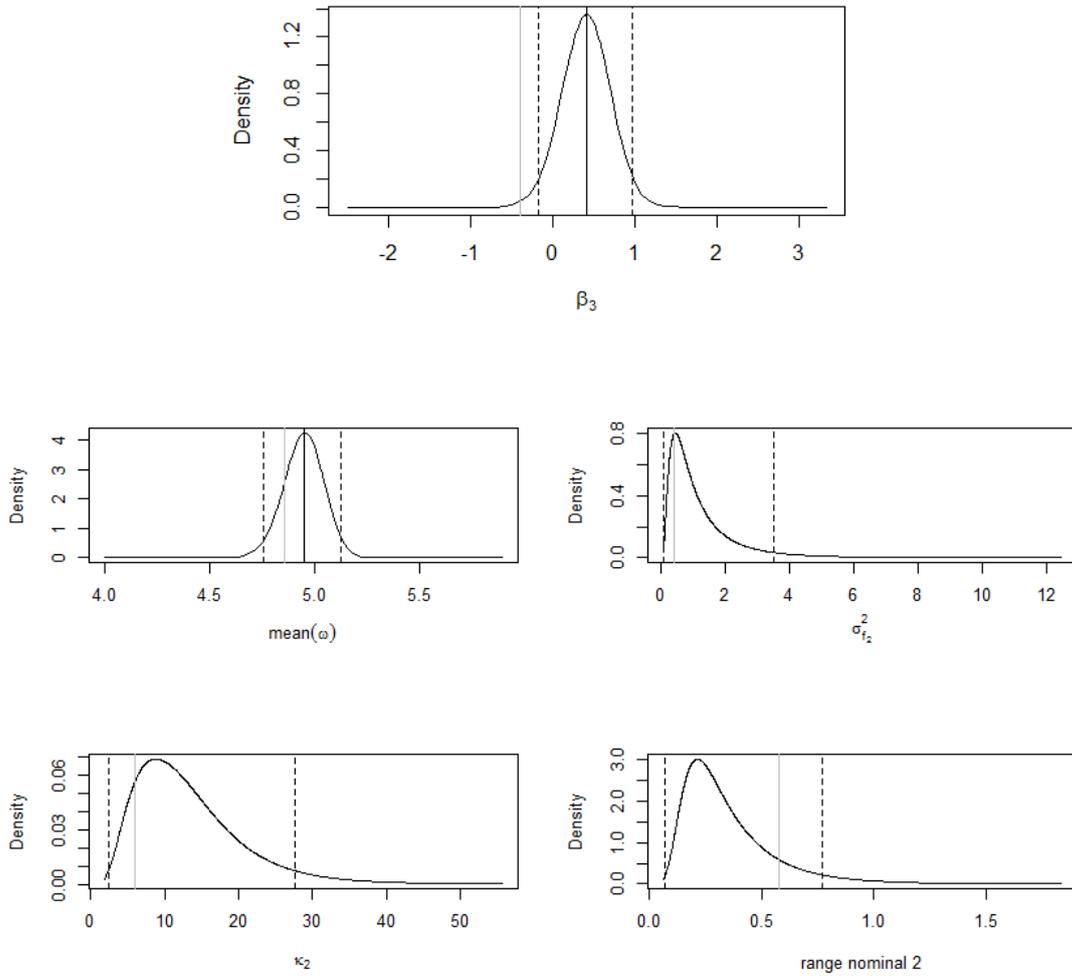
Figure 14: Modelling by functional connections: Posterior marginal distributions for the parameters and hyperparameters of the negative linear dependence example. Top panel: The interaction parameter. Four bottom panels: The mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $f_2(s_i)$. The notation is the same as in Figure 12.
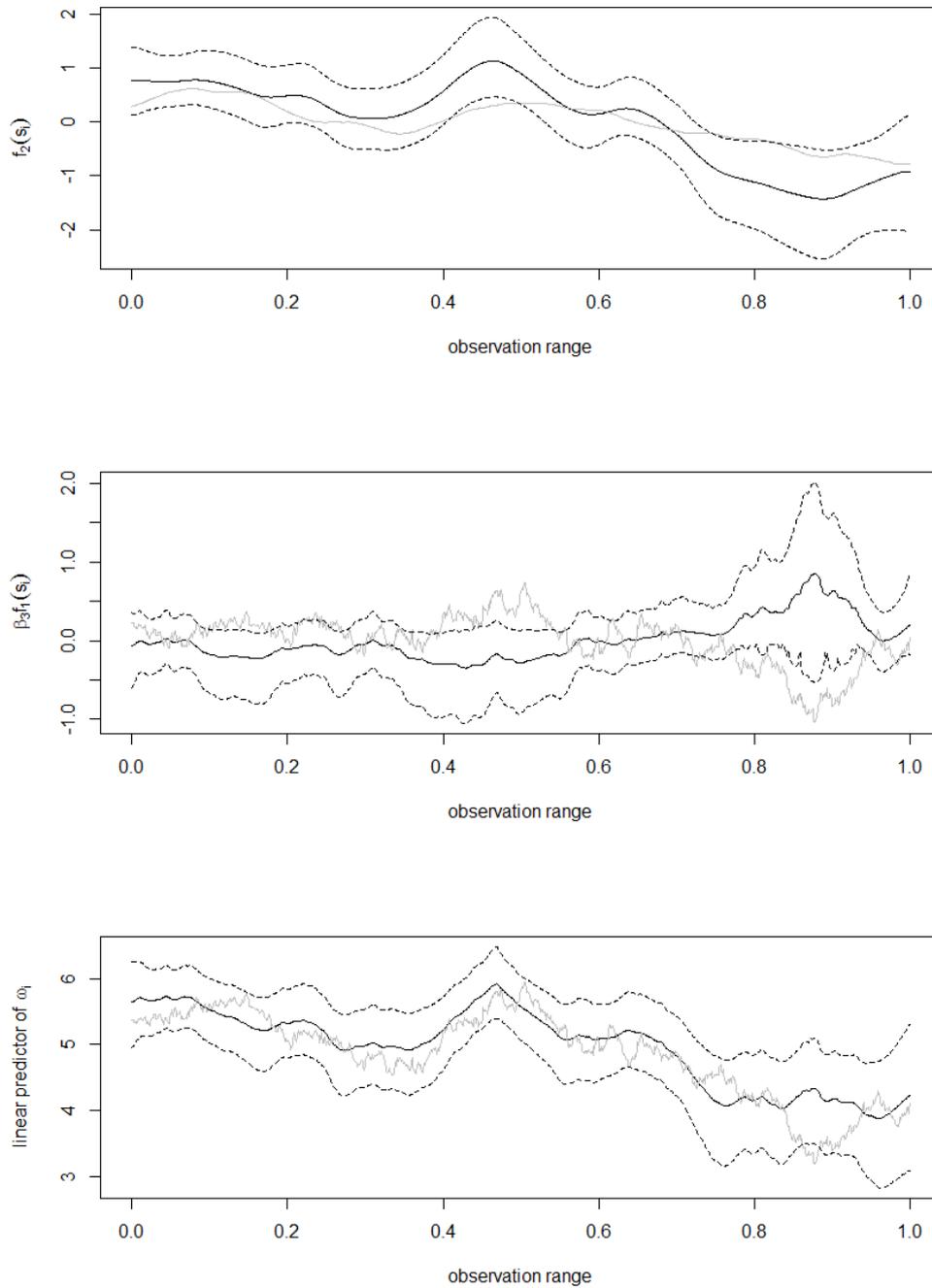
Figure 15: Modelling by functional connections: Posterior marginal distributions for the second spatially structured effect $f_2(s_i)$ (top), the interaction process $\beta_3 f_1(s_i)$ included in the second latent process ω_i (middle) and the linear predictor of the second latent process (bottom) in the negative linear dependence example. The notation is the same as in Figure 9.

In contrast, attempts with different positive values for the interaction parameter and other simulations of positive linear dependencies pointed out that positive linear interactions may mostly be discovered very well.

The negative linear dependence example – model comparison The inability of the model fitting process to estimate the negative linear interaction correctly is also uncovered by the model comparison scores. Concerning the DIC, WAIC and CPO values, there are five models which may be seen as potential candidates according to our developed rule: the independence model, the "true" first linear dependence model, the first non-linear dependence model, the second non-linear dependence model and the shared component model. All the stated models have differences in the DIC or WAIC of at most two with the independence model being the best one according to the DIC and the first non-linear dependence model according to the WAIC. We do not consider the CPO values here as the shared component model is again an outlier regarding the CPO (but the other models show differences within a range of 0.0005).

Consequently, the first linear dependence model is not recognized as the best one here. One would probably favour the independence model as it is the simplest of the models stated above. This may be seen as a desirable result when taking into account that the negative linear dependence cannot be estimated correctly.

Note that in contrast to the linear dependence model, the non-linear dependence model succeeds in detecting the negative interaction, albeit not completely in the right linear way. This may be seen in Figure 16, where the posterior mean of the fitted interaction is marked by a solid black line and the true negative linear interaction by a grey line (both curves are shifted to have zero-means). Thus, our suggestion is to conduct the fitting process once by means of the linear dependence model and once by means of the non-linear dependence model and to compare the results whenever a negative linear interaction must be considered.

The non-linear dependence example – model validation When moving on to the non-linear dependence example, we must first state the method which is used to model the non-linear dependency between the two Poisson processes. As announced in Section 3.1, we use Bayesian P-splines for this purpose. These have the advantage of explicitly demonstrating the uncertainty in smoothing parameters (see Fraaije et al. [FBV$^+$15, Appendix S3]).

A German introduction to (B-)splines and P-splines may be found in Fahrmeir et al. [FKL09, Ch.7]. We use a cubic B-spline basis with 21 knots (to allow for oscillating functions) and a penalty matrix which determines the deviation from a horizontal line. The order of the difference penalty is one for the non-linear dependence model (we tried a difference penalty of order two as it is used by Fraaije et al., but the Newton-Raphson algorithm did not converge). The variance of the zero-mean random walk prior on the coefficients of the P-spline regression is determined by a IG$(2, 0.1)$ (inverse gamma) prior. We use this prior instead of the IG$(1, 0.01)$ prior employed by Fraaije et al., to compensate for the missing smoothing due to the lower difference penalty order.
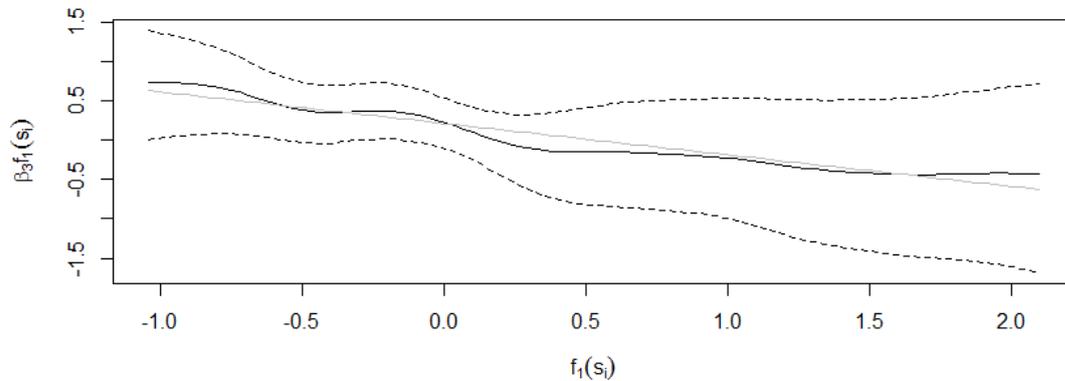
Figure 16: The estimated interaction when fitting the first non-linear dependence model to the negative linear dependence example. The posterior mean of the fitted interaction is marked by a solid black line, the 95% equal-tailed credible interval by dashed lines and the true negative linear interaction by a grey line. The interactions are shifted to have zero-means.

While the priors for the intercepts of the two latent processes and the hyperparameters of the first spatially structured effect stay the same as in the previous examples, we have to be more careful regarding the priors for the hyperparameters of the second spatially structured effect. When fitting the non-linear dependence model without any changes of the prior distributions, this yields in a "wiggled" posterior mean for the spatially structured effect $f_2(s_i)$ (see the top panel of Figure 17). The reason for this may be that the spatially structured effect absorbs parts of the non-linear interaction.

We calibrated the prior of the spatially structured effect by investigating the fitted spatially structured effect $f_2(s_i)$ in simulations with other – not too informative – prior distributions. A less wiggled effect is achieved with a prior of

$$\log(\kappa_2) \sim \mathcal{N}(2.852892, 1) \tag{50}$$

instead of the prior denoted in Equation 49, i.e. the variance of $\log(\kappa_2)$ is 1 instead of 10. The prior for τ_2 stays the same as in Equation 48. The resulting modified fit is shown in the bottom panel of Figure 17. It is difficult to get a still smoother spatially structured effect as one runs in the risk of obtaining a horizontal line then (what happened in some of our simulations).

Another issue which must be discussed at this point is that a joint model fitting approach is not possible for the non-linear dependence model. As the Bayesian P-spline approach is based on fixed values for the B-spline basis, one must fit the first Poisson process and afterwards the second process *based on the results of the first one*. This hierarchical modelling approach provokes two
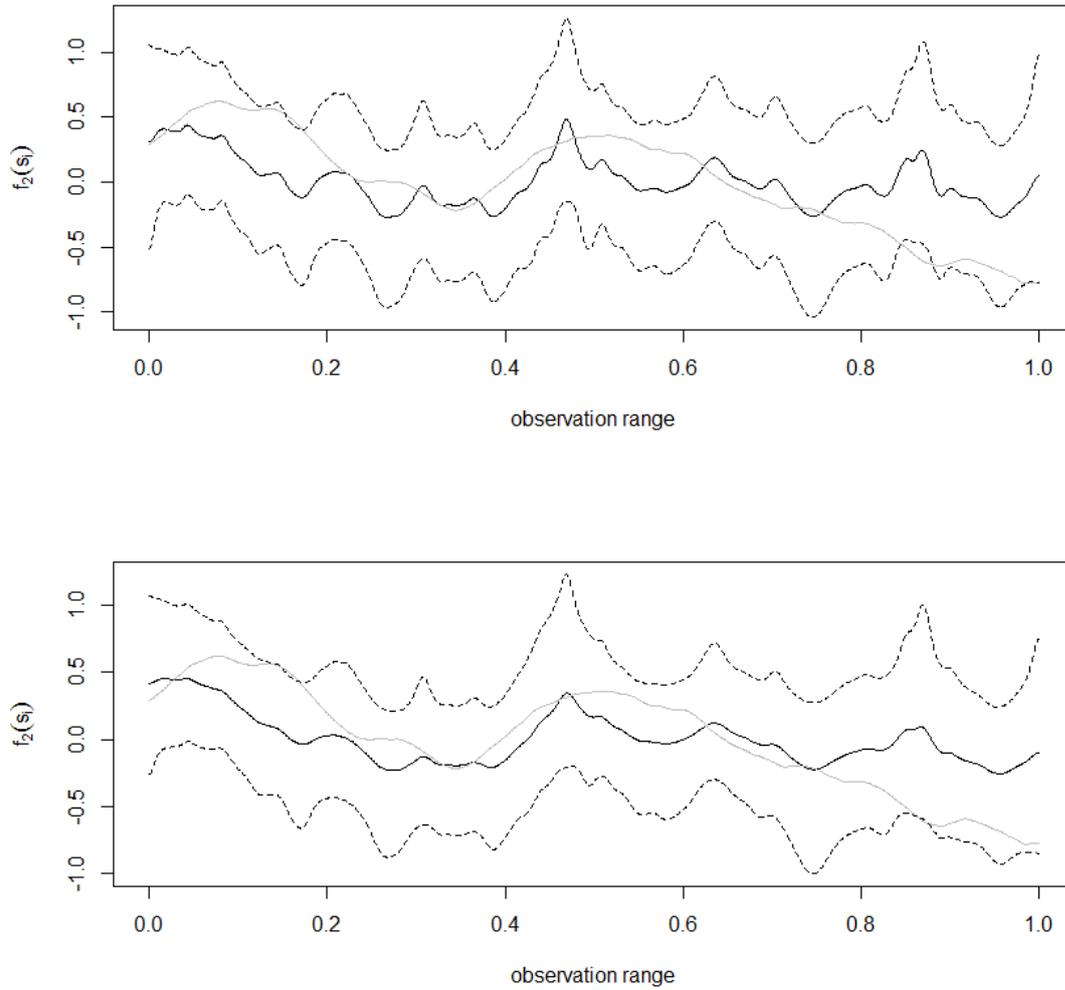
Figure 17: Modelling by functional connections: Posterior marginal distributions for the second spatially structured effect $f_2(s_i)$ in the non-linear dependence example. Top: The fitted spatially structured effect with the default prior. Bottom: The fitted spatially structured effect when using a normal prior with variance 1 instead of variance 10 for $\log(\kappa_2)$. The notation is the same as in Figure 9.

problems when using the posterior means of the first spatially structured effect $f_1(s_i)$ as the basis values to model the non-linear interaction. The two problems are illustrated in Figure 18. The top panel of the figure displays the posterior marginal distributions for the first spatially structured effect $f_1(s_i)$, which was fitted by means of the non-linear dependence model. For the present hierarchical approach this means that solely the first Poisson process was fitted and so it is obvious that there are no differences compared to the independence example. What may also be seen is that in contrast to the true spatially structured effect (marked by the grey line) the posterior means of the parameters $f_1(s_i)$ do not have big amplitudes upwards and downwards. This observation is due to the fact that the Poisson point process is extremely weakly informative about the small scale structure of the latent process and the small scale structure may consequently not be detected from the Poisson counts (see Remark 5.1). Consequently, the range of values achieved by the posterior means is smaller than that of the true spatially structured effect. This is indicated in the bottom panel of Figure 18, where the non-linear relationship in the non-linear dependence example is illustrated and the constricted range of values for the posterior means of $f_1(s_i)$ is marked by grey lines. Hence, the complete non-linear dependency cannot be detected.

Therefore, the first drawback of the inevitable hierarchical modelling is that there is no chance to discover the non-linear interaction completely correct. Due to the smoothed estimated process, the non-linear relationship is only partially detectable. But there is a second drawback, too. When one uses the posterior means of the spatially structured effect $f_1(s_i)$ as the fixed basis values of the Bayesian P-spline approach, the uncertainty in estimating the effect $f_1(s_i)$ gets lost. Consequently, the 95% credible intervals for the estimated non-linear interaction, represented in Figure 19, are incorrect. They are assumed to be wider in general.

After stating the particularities and difficulties of the non-linear dependence model, we can now evaluate the quality of the model fit. Fitting the non-linear dependence model took 140 seconds. We start with an examination of the posterior marginal distributions for the (hyper-)parameters, presented in Figure 20. Only the mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $f_2(s_i)$ are shown.

When looking at the figure it is striking that the mean of the second latent process is estimated completely wrong (first panel). This is the result of a technical problem during fitting the non-linear dependence model: When fitting the second Poisson process, the values of the Bayesian P-spline at each of the knots – which are equivalent to different values of $f_1(s_i)$ – are weighted equally in the estimation of the intercept. But in practice there may be knots with a lot of values of $f_1(s_i)$ nearby and others with only a few. So the estimation of the mean is falsified. However, one may obtain a point estimate for the mean of the second latent process via the average of the posterior means for the linear predictor of the process ω_i. This average is marked by a blue line in the first panel of Figure 20 and is close to the true value.

Regarding the hyperparameters of the second spatially structured effect $f_2(s_i)$ (panels two to four in the figure), they are estimated worse than in the independence example. The true value of
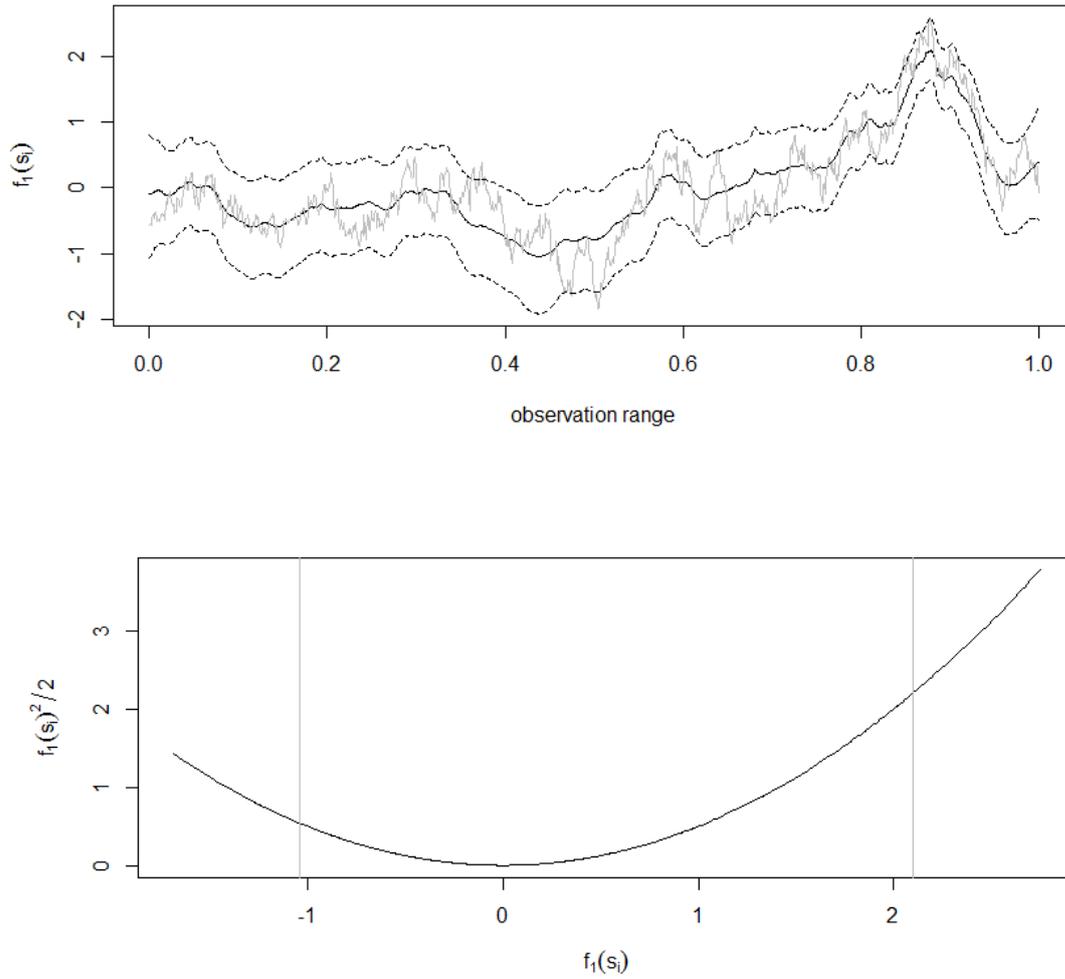
Figure 18: The hierarchical modelling issue for the non-linear dependence model. Top: The posterior marginal distributions for the first spatially structured effect $f_1(s_i)$, fitted by means of the non-linear dependence model in the non-linear dependence example. Bottom: The non-linear relationship present in the simulated non-linear dependence example (black line) and the range of values that the posterior means of the fitted first spatially structured effect achieve (borders indicated by grey lines).
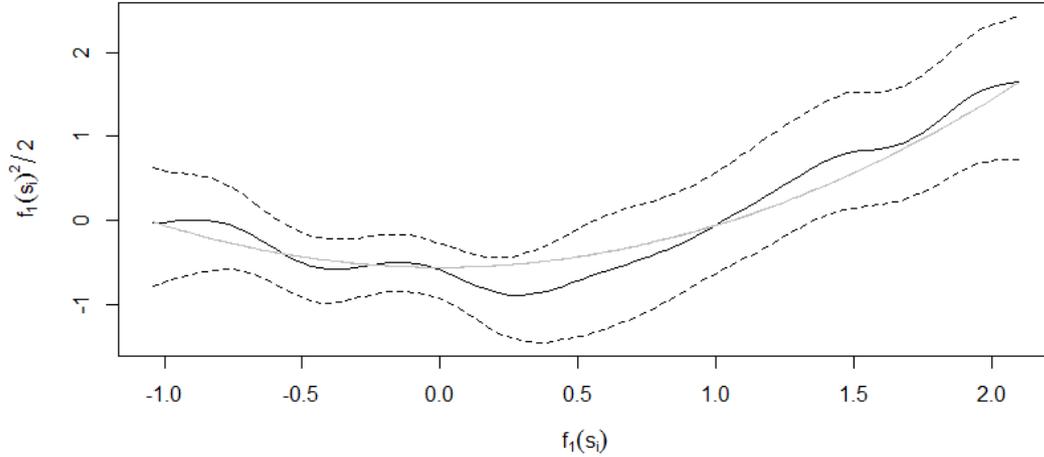
Figure 19: The estimated interaction when fitting the non-linear dependence model to the non-linear dependence example. The posterior mean of the fitted interaction is marked by a solid black line, the 95% equal-tailed credible interval by dashed lines and the true non-linear interaction by a grey line. The interactions are shifted to have zero-means.

the nominal range is outside the 95% credible interval and the true value of the marginal variance is also near the border of the credible interval. Obviously, the second spatially structured effect is not very well recognized. This can also be observed from the fit of the second spatially structured effect parameters $f_2(s_i)$ themselves which have already been presented in the bottom panel of Figure 17. The fit is not as good as that for the second spatially structured effect in the independence example.

With the focus of our work lying on interaction modelling, it is more important that the non-linear interaction function of the simulated example is estimated correctly. The fit has already been presented in Figure 19. It seems to be good aside from the middle part where the values are estimated too low. Nevertheless, the true interaction function lies everywhere inside the 95% credible interval. Note that the fit is only available for the range of values achieved by the posterior means of the estimated spatially structured effect $f_1(s_i)$ (see above). We will see below in the model comparison if the overall non-linear interaction is estimated good enough to be recognized.

Finally, the posterior marginal distributions for the linear predictor of the second latent process ω_i are displayed in Figure 21. The true linear predictor lies mostly inside the 95% credible intervals, but the posterior means are often highly shifted compared to the true linear predictor (e.g. around the value 0.4 and after the value 0.8 of the observation range).

The MAE and MSE scores confirm our observations once again. The MAE and MSE of the
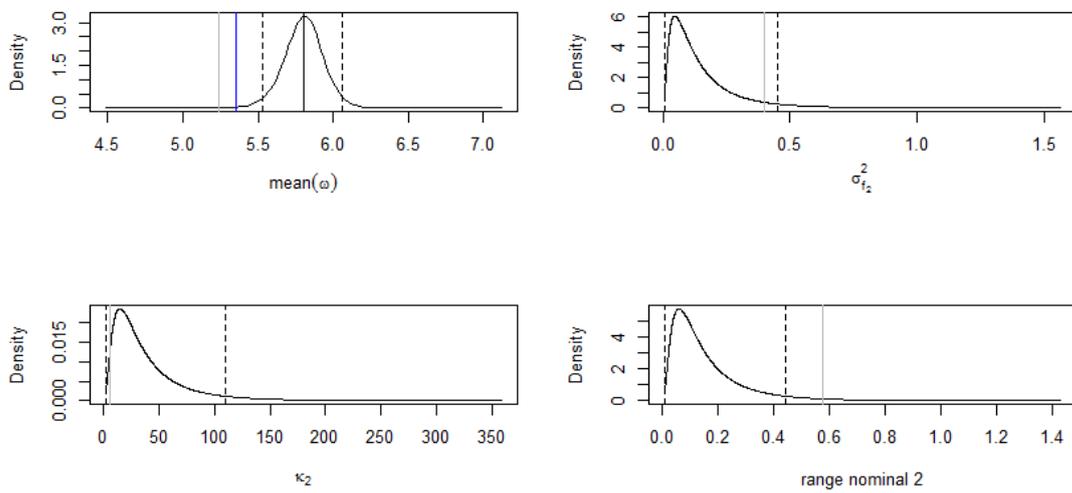
Figure 20: Modelling by functional connections: Posterior marginal distributions for the parameters and hyperparameters of the non-linear dependence example. First panel: The mean of the second latent process ω_i. Other panels: The hyperparameters of the second spatially structured effect $f_2(s_i)$. The notation is the same as in Figure 12, but additionally the average of the posterior means for the linear predictor of the process ω_i is indicated by a blue line in the first panel.
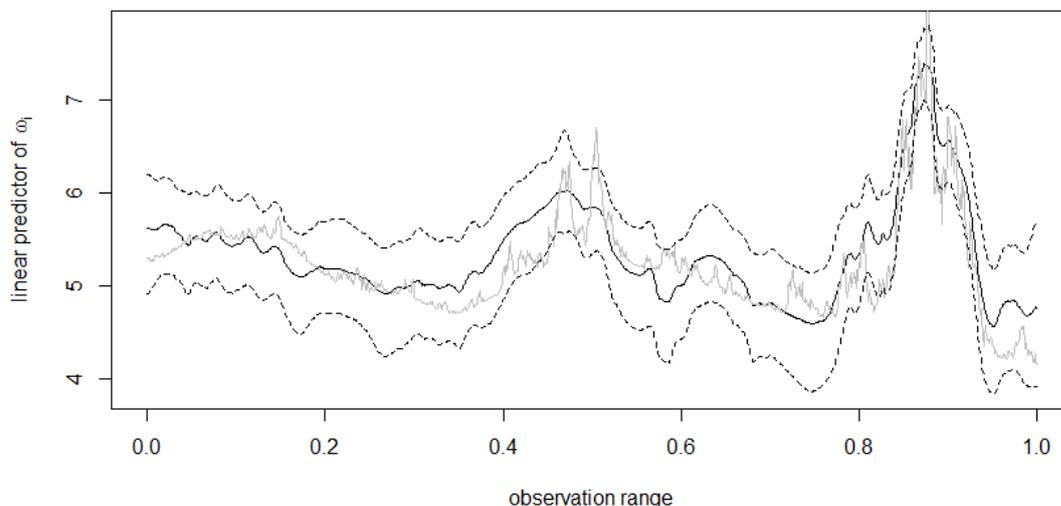
Figure 21: Modelling by functional connections: Posterior marginal distributions for the linear predictor of the second latent process ω_i in the non-linear dependence example. The notation is the same as in Figure 9.

spatially structured effect $f_2(s_i)$ are much better than those of the poorly fitted negative linear dependence example but also worse than the scores of the independence and the positive linear dependence example. The MAE and MSE for the linear predictor of ω_i are again worse than those of the independence and the positive linear dependence example, but nearly equal to the scores of the negative linear dependence example (where the fit of the second linear predictor actually was not so bad).

The non-linear dependence example – model comparison   When fitting the models denoted in Table 5 to the non-linear dependence example, half of the models fail to fit the example due to non-convergence of the Newton-Raphson algorithm or a singular covariance matrix. Applying our developed rule to the remaining model fits, there is a clear winner which is not the "true" first non-linear dependence model, but the first *linear* dependence model. The first non-linear dependence model is only the second best model and this with differences in the DIC and WAIC of more than 5 and in the CPO of more than 0.0005. Actually, the non-linear dependence model fits the simulated example better regarding the spatially structured effects and both linear predictors but obviously the difference in the fitting quality is not high enough to counterbalance or even outmatch the penalization due to the additional flexibility within the non-linear dependence model.

One might worry if the non-linear dependence model is generally able to outperform the linear

60

dependence model, taking into account all the drawbacks of the hierarchical approach of the non-linear modelling. We will see in the next section that this is in fact possible.

6.3 The dependence on the latent process example

Model validation   Next, we investigate dependencies on the latent process. Recall the simulated example of Section 4.3 for this purpose. There, a simplified version of non-linear dependency on the first latent process η_i has been simulated with the latent processes described by

$$\eta_i = \beta_1 + f_1(s_i) + u_i,$$
$$\omega_i = \beta_2 + g(\eta_i) + v_i,$$

and $g(\eta_i) = -2\sin(\eta_i)$.

Again, the non-linear interaction is modelled by the Bayesian P-spline approach. This time, the Newton-Raphson algorithm converges for a difference penalty of order two, so we use that with a $IG(1, 0.01)$ prior for the variance of the zero-mean random walk on the coefficients of the P-spline regression as it is employed in Fraaije et al. [FBV+15].

The priors for the remaining (hyper-)parameters stay the same as in the independence example. Note that we do not need a prior for the second spatially structured effect $f_2(s_i)$ as it does not appear in the simplified version of the non-linear dependency on the latent process considered here. So we do not have the issue that the spatially structured effect might absorb parts of the non-linear relationship.

Using the posterior means of the fitted latent process η_i as the basis values for the Bayesian P-splines results in an interesting observation: Actually, it does not make any difference in our simulation studies if we use the posterior means of the latent process η_i or the posterior means of the spatially structured effect $f_1(s_i)$ as the basis values for the non-linear interaction modelling. The reason for this is that the spatially unstructured effects may not be detected in the fitting process (as we already stated several times) so that the difference between the estimated processes $f_1(s_i)$ and η_i is simply given by the intercept β_1. But this value is equal for every segment of the observation range and only shifts the entire process in one direction. It has consequently no influence on the inference of the non-linear interaction. Hence, the estimated relationship is the same for the dependency on the *latent process* as for the dependency on the *spatially structured effect*.

Fitting the non-linear dependence on the latent process model to the dependence on the latent process example took 105 seconds which is the lowest running time of all the models examined so far. The examination of the posterior marginal distributions is limited to that of the mean of the second latent process ω_i as there is no second spatially structured effect $f_2(s_i)$ and the remaining (hyper-)parameters are estimated equally to the independence example. Regarding the posterior marginal distribution for the mean of the second latent process in Figure 22, the
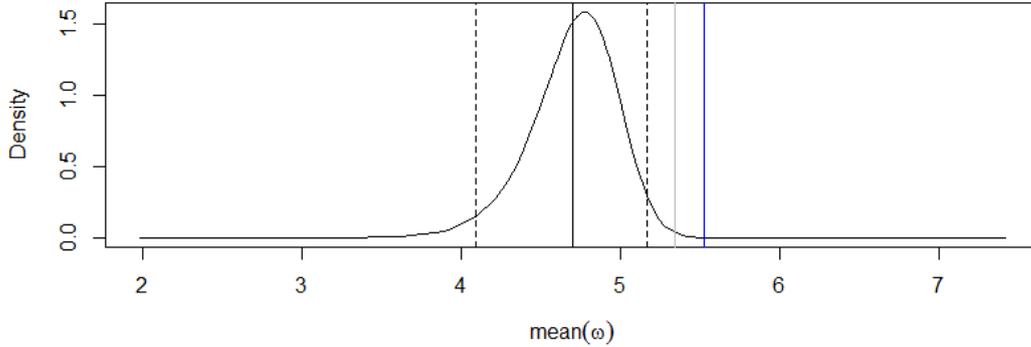
Figure 22: Modelling by functional connections: Posterior marginal distribution for the mean of the second latent process ω_i in the dependence on the latent process example. The notation is the same as in Figure 20.

true value lies once again outside the 95% credible interval due to the technical problem during the fitting process already announced in the previous section. Nevertheless, the average of the posterior means for the linear predictor of the process ω_i (marked by a blue line) is close to the true value.

Concerning the estimated non-linear interaction, there is again a constricted range of values achieved by the posterior means of the latent process η_i. Hence the complete non-linear dependency is undetectable. This is indicated in the top panel of Figure 23, where the constricted range is marked by grey lines.

In the bottom panel, the estimated non-linear interaction is illustrated. Note that once again, the 95% credible interval is assumed to be wider than presented in the panel as the uncertainty in estimating the values of the latent process η_i is not included into the model fit. While here the true interaction function lies outside the credible interval at the beginning, it might be that in fact (when considering the uncertainty) it is inside the interval. All in all, the non-linear interaction is detected satisfactorily, even though it is estimated a bit too low at the beginning and too high in the middle.

As there is no second spatially structured effect $f_2(s_i)$ and the first one is fitted equally to the independence example, we skip the analysis of the spatially structured effects and go on with the investigation of the fitted linear predictor of the second latent process ω_i (the linear predictor of the first latent process is fitted equally to the independence example). The posterior marginal distributions for the linear predictor of the second latent process are illustrated in the top panel of Figure 24.

One must not require that the true linear predictor (marked by a grey line) lies inside the 95%
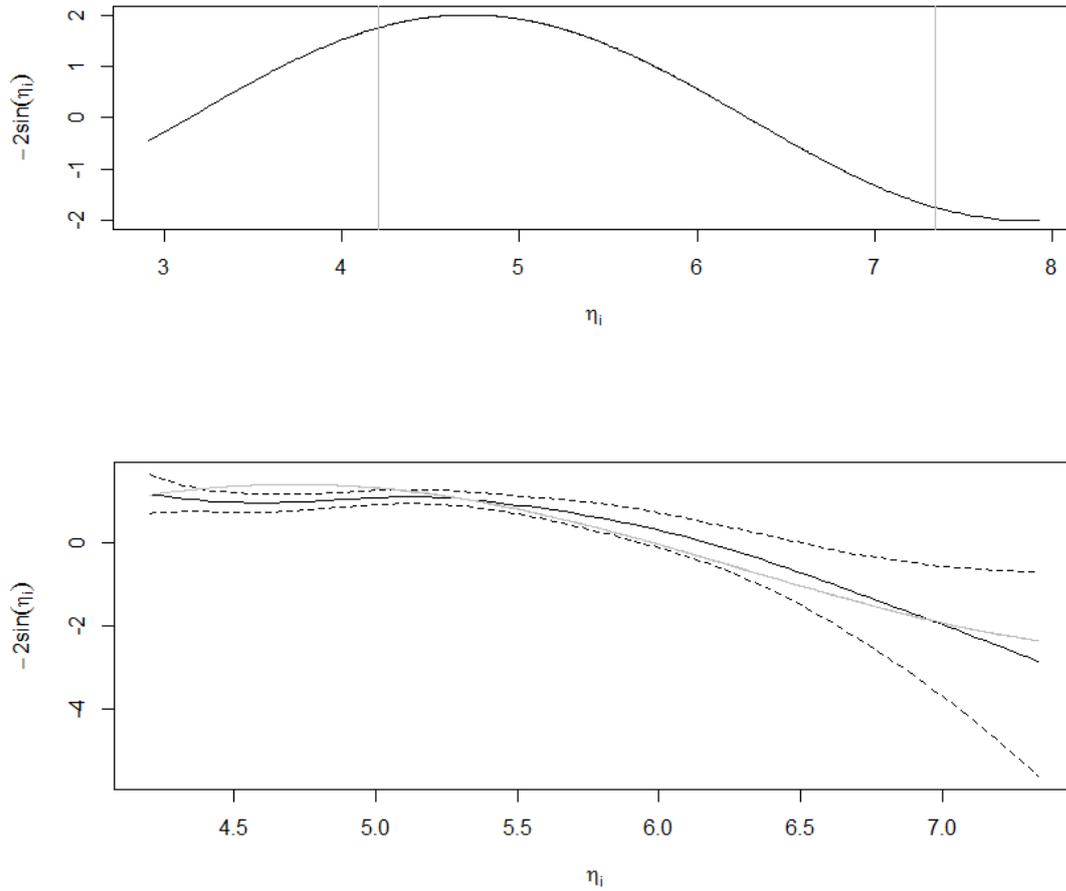
Figure 23: The hierarchical modelling issue for the non-linear dependence on the latent process model and the estimated interaction when fitting the model to the dependence on the latent process example. Top: The non-linear relationship present in the simulated dependence on the latent process example (black line) and the range of values that the posterior means of the fitted first latent process η_i achieve (borders indicated by grey lines). Bottom: The posterior mean of the fitted interaction (solid black line), the 95% equal-tailed credible interval (dashed lines) and the true non-linear interaction (grey line). The interactions are shifted to have zero-means.
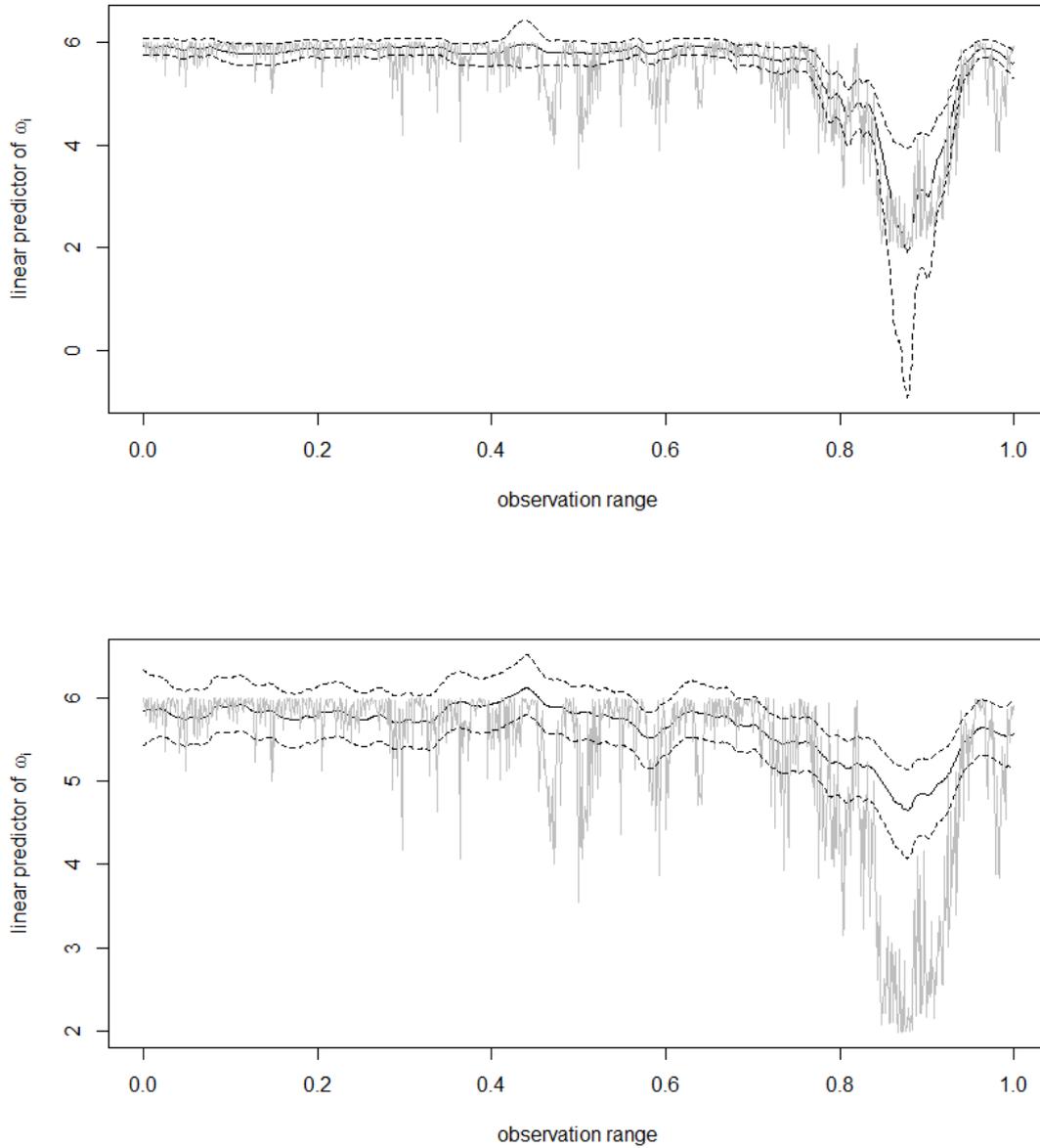
Figure 24: Model validation and model comparison: Posterior marginal distributions for the linear predictor of the second latent process ω_i in the dependence on the latent process example, fitted by means of the first simplified non-linear dependence on the latent process model (top) and the first simplified linear dependence on the latent process model (bottom). The notation is the same as in Figure 9.

credible interval as the linear predictor also contains the simulated spatially *unstructured* effect u_i which is undetectable due to its small scale structure. It is more important that the general trend of the linear predictor is estimated correctly what is generally achieved – especially for the strong peak downwards at the end of the observation range.

Considering the MAE and MSE, only the posterior means of the linear predictor of the second latent process ω_i are of interest – and exhibit scores which are worse than those of the other examples considered so far. However, the scores of the present example must be regarded with suspicion as they include spatially unstructured effects which are not present in the other examples. So we do not further go into the explicit values of the two scores.

Model comparison Fitting the models denoted in Table 5 to the dependence on the latent process example results in only two models to be considered. Note that it makes no difference if one fits the dependence on the spatially structured effect or the dependence on the latent process as already explained above. The best model regarding the comparison scores is the "true" first simplified non-linear dependence on the latent process model, the second best with differences in the DIC and WAIC between 4 and 5 and in the CPO of 0.0004 is the not simplified non-linear dependence model. It is obvious that one chooses the simpler "true" model then.

It might be surprising that the first simplified *linear* dependence model is not considered this time while the linear dependence model has been able to fit the non-linear dependence example in the previous section satisfactorily enough to even outperform the "true" non-linear dependence model. But a look at the fit of the simplified linear dependence model to the present example (shown in the bottom panel of Figure 24) reveals that the simplified linear dependence model is apparently not flexible enough to model the strong peak at the end of the observation range.

6.4 The LMC and shared component examples

The LMC example – model validation The last two models covered in our simulation studies are the LMC and the shared component model. Simulated examples of the two models have been introduced in Section 4.4. We start with the LMC example, where the priors for the (hyper-) parameters are determined by Equations 41 to 45 like in most of the examples before. However, since we have changed the (hyper-)parameters for the LMC and shared component model and the values of $b_1^{(\nu_1)/(\nu_2)}$ and $b_2^{(\nu_1)/(\nu_2)}$ are dependent on the values of ν_1 and ν_2 (which are $\nu_1 = 0.3$ and $\nu_2 = 1.2$ here), the priors for τ_1, κ_1, τ_2 and κ_2, denoted in Equations 46 to 49, must be modified. The new priors for the inference of the LMC example are given by

$$\log(\tau_1) \sim \mathcal{N}(-2.131122, 10), \tag{51}$$

$$\log(\kappa_1) \sim \mathcal{N}(2.303586, 10), \tag{52}$$

$$\log(\tau_2) \sim \mathcal{N}(-4.549879, 10), \tag{53}$$

$$\log(\kappa_2) \sim \mathcal{N}(2.74132, 10). \tag{54}$$

Obviously, only the priors for the hyperparameters of the second spatially structured effect $f_2(s_i)$ are changed, but the different value for ν_1 (0.3 instead of 0.5) does not cause a change of the default first spatially structured effect hyperparameter priors.

Fitting the LMC took 853 seconds which is a very long running time compared to the previous examples. When looking at Figure 25 for an examination of the posterior marginal distributions for the (hyper-)parameters, it is visible that the interaction parameters (two top panels) are estimated in a satisfactory way. The true values $a_{12} = 0.3$ and $a_{21} = 0.7$ lie inside the 95% credible intervals, in some distance from the interval borders. Moreover, the posterior means are good point estimates for the true interactions with values of 0.4066 and 0.5894 respectively. The second interaction parameter a_{21} is significantly different from zero as the value zero does not lie inside the interval. In contrast, the first interaction a_{12} is just not significant which may be due to the low true value of 0.3.

Concerning the remaining (hyper-)parameters illustrated in Figure 25, the mean of the first latent process η_i (third panel) is well estimated, whereas the mean of the second latent process ω_i (seventh panel) is not. Five of the six hyperparameters of the two spatially structured effects are satisfactorily estimated with the true values lying inside the 95% HPD intervals, only the marginal variance $\sigma_{w_1}^2$ of the spatially structured effect $w_1(s_i)$ is not contained in the particular credible interval (note that the spatially structured effects are named $w_1(s_i)$ and $w_2(s_i)$ in the LMC).

The posterior marginal distributions cannot be compared to those of the fitted independence example in Section 6.1 as we have changed the values for the parameters in the present example. We rather compared the fitted (hyper-)parameters to those of an independence model fitted to a simulated independence example with the new parameter values (and $a_{12} = a_{21} = 0$). The posteriors for the (hyper-)parameters turned out to be better estimated in the LMC example than in the new independence example.

Considering the posteriors for the fitted spatially structured effects $w_1(s_i)$ and $w_2(s_i)$ displayed in Figure 26, the true spatially structured effects are almost everywhere inside the 95% credible intervals. The posterior means of the parameters $w_1(s_i)$ are good point estimates for a "smoothed" first spatially structured effect $w_1(s_i)$. Regarding the second spatially structured effect $w_2(s_i)$, the posterior means fail to follow the oscillations downwards after the point 0.2 and upwards after the point 0.4 of the observation range. Apart from that, the true spatially structured effect is well estimated.

66
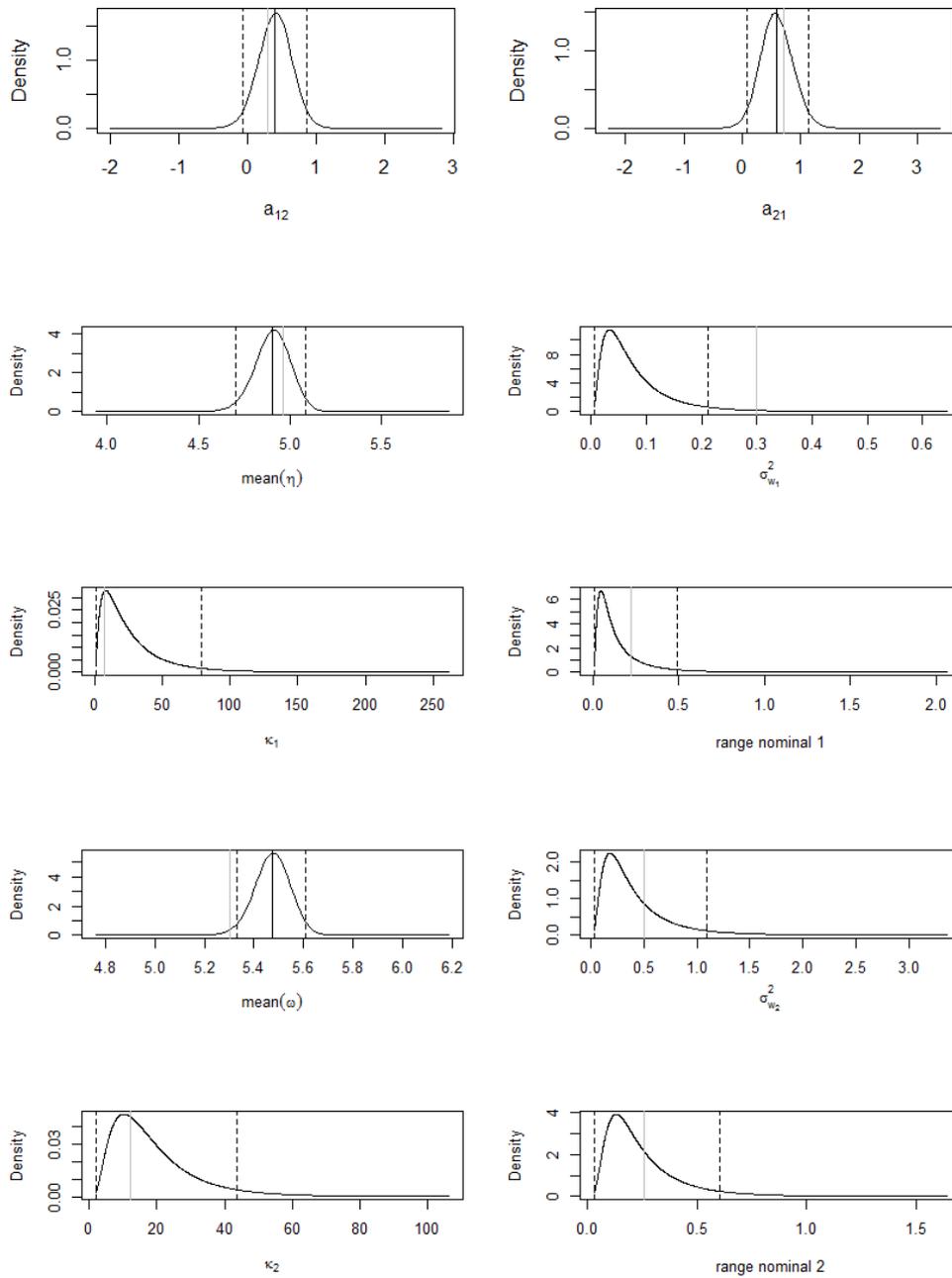
Figure 25: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the parameters and hyperparameters in the LMC example. Two top panels: The interaction parameters. Four middle panels: The mean of the first latent process η_i and the hyperparameters of the first spatially structured effect $w_1(s_i)$. Four bottom panels: The mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $w_2(s_i)$. The notation is the same as in Figure 12.

67
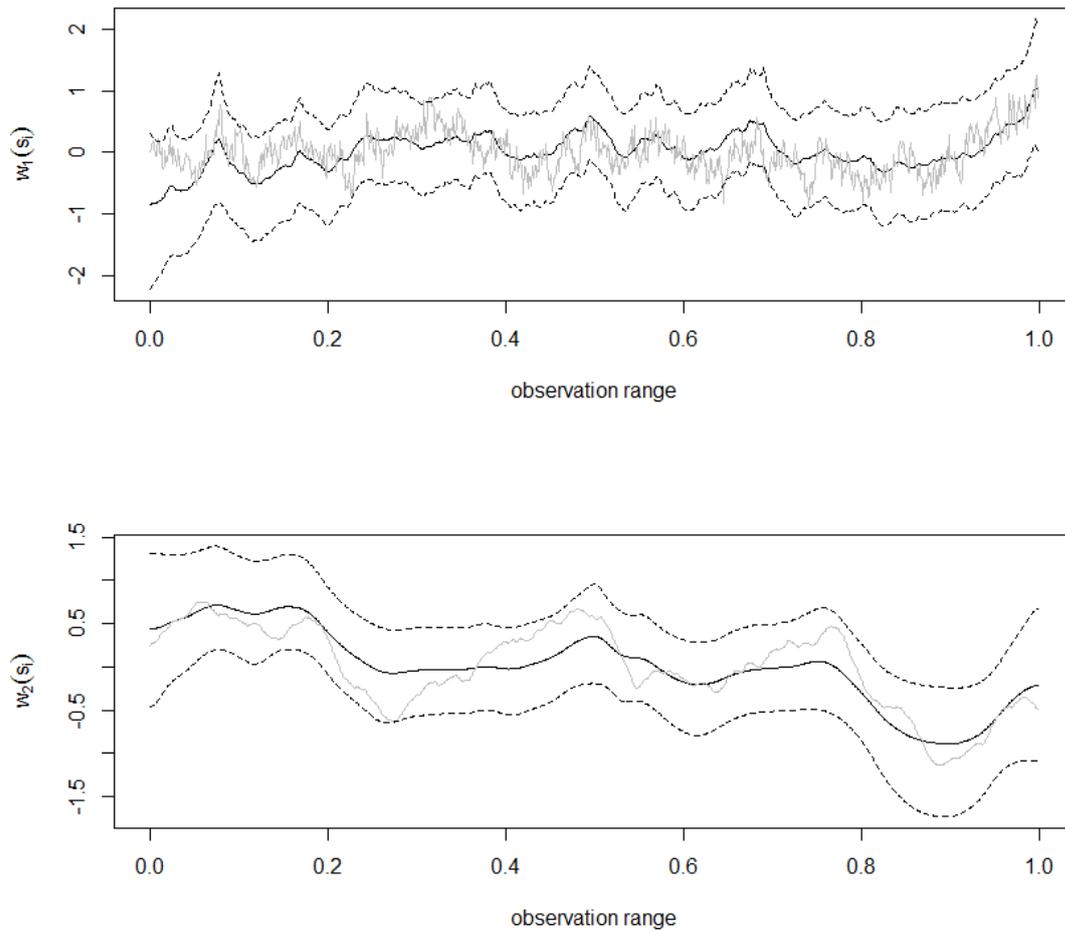
Figure 26: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the spatially structured effects in the LMC example. The notation is the same as in Figure 9.

The good quality of the fits for the spatially structured effects is somewhat surprising as there is technically an existing identifiability issue in the LMC model (see the corresponding paragraph in Section 3.2.1). The different priors for the hyperparameters of the spatially structured effects as well as the a priori fixed (different) smoothness parameters are apparently enough to allow the discrimination of the two processes.

The posteriors of the linear predictors of the two latent processes may be examined in Figure 27. The linear predictor of the first latent process η_i is very similar to the first spatially structured effect $w_1(s_i)$ as the influence of the second spatially structured effect $w_2(s_i)$ on the latent process η_i is low due to the small value of the first interaction parameter ($a_{12} = 0.3$). The impact of the second spatially structured effect is only visible by a little shift upwards at the beginning of the observation range and a shift downwards at the end of the observation range (both for the true and the fitted linear predictor).

Regarding the linear predictor of the second latent process ω_i, the influence of the first spatially structured effect is clearly perceptible due to the lower smoothness of the linear predictor compared to the second spatially structured effect $f_2(s_i)$. Additionally, there is a zone between the values 0.2 and 0.3 of the observation range where several values of the true linear predictor lie outside of the 95% credible interval. For the rest of the linear predictor, the posterior means are good point estimates of the (smoothed) true values.

We do not consider the MAE and MSE values here as they do not give any further insides, but rather move on to the model comparison for the LMC example.

The LMC example – model comparison Fitting the models of Table 5 to the LMC example yields very strange results concerning the model comparison scores. If applying our developed rule for model comparison to the fitted models, the clear winner in terms of the DIC and WAIC would be the first non-linear dependence model. Regarding the CPO, the best one (apart from the usual winner for the CPO, which is the shared component model) is the independence model which is simultaneously the second best when looking at the DIC and WAIC. The "true" LMC is only in the third place together with the first and the second linear dependence model as well as the shared component model.

The differences in the DIC and WAIC between the first non-linear dependence model and the "true" LMC are around 42 and 28, far too much than that the LMC might even be considered. When looking at the fit of the non-linear dependence model, the reason for this odd observation becomes clear: The process η_i is completely overfitted. This is directly visible in Figure 28.

The same problem is present in the independence model. In fact, the overfitting issue may be uncovered without considering the fits: For each of the overfitted models, the difference between the DIC and WAIC score is about 20 while values around 10 have been the maximal difference between the two scores *for the fitted models in each of the previous examples.*

The best models for which there is no overfitting visible is the announced group of (roughly) equally rated models, originally being in the third place. The group comprises the two linear

69
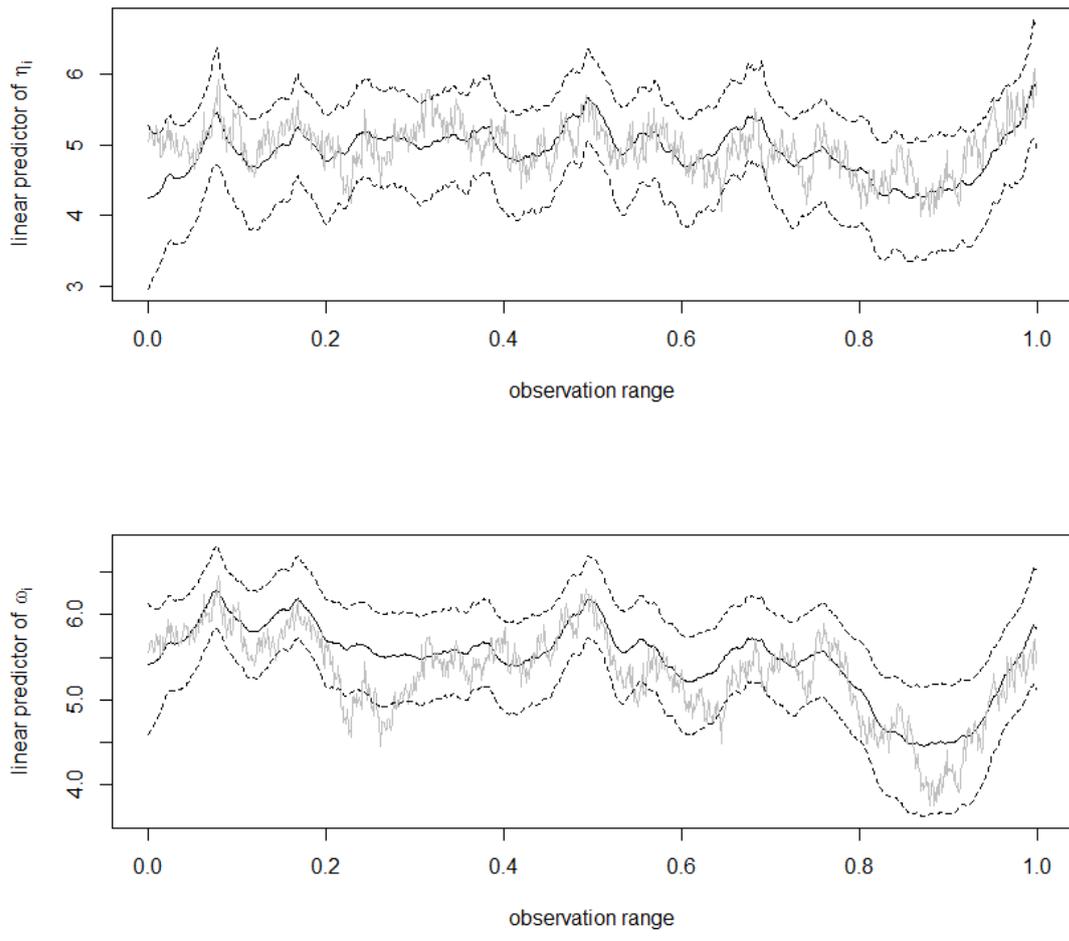
Figure 27: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the linear predictors of the latent processes in the LMC example. The notation is the same as in Figure 9.
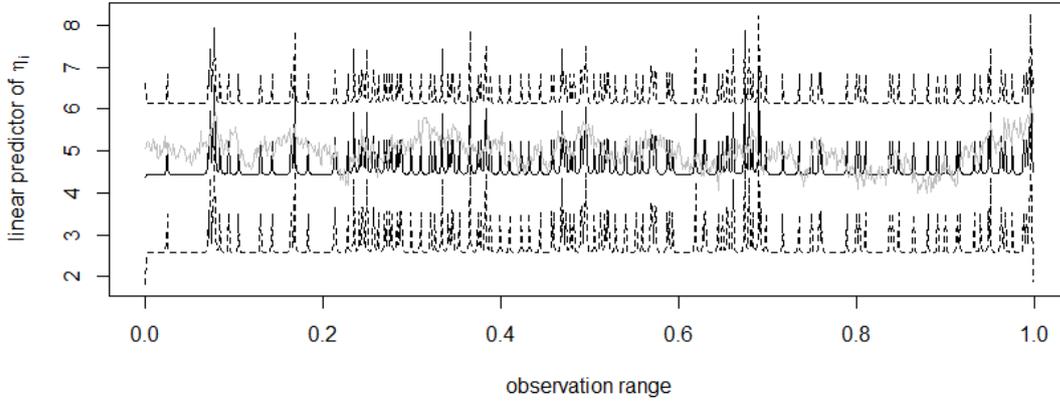
Figure 28: The overfitting issue in the LMC example: Posterior marginal distributions for the linear predictor of the first latent process η_i in the LMC example, fitted by means of the first non-linear dependence model. The notation is the same as in Figure 9.

dependence models, the shared component model and the LMC. This is plausible when taking into account that the LMC may be rewritten to a linear dependence model (see Remark 3.4). Moreover, the shared component model offers additional flexibility. It is a matter of interest if the LMC is chosen then. For example, one may prefer the LMC if one wants to detect the "main" underlying effects for each of the Poisson processes as well as the amount of the interaction with regard to the "other" Poisson process. In contrast, if one is only interested in the construction of a valid covariance matrix, one may use the linear dependence model instead (see again Remark 3.4).

The shared component example – model validation   The last simulation study deals with the shared component example. There, the default priors must be modified to avoid that two of the three processes $w_1(s_i)$, $w_2(s_i)$ and $z(s_i)$ take the whole spatially structure of the latent processes η_i and ω_i while the third is estimated to be almost a horizontal line at zero. Experiments with

different priors for the hyperparameters showed that this may be prevented with the priors

$$\log(\tau_1) \sim \mathcal{N}(-2.131122, 0.1),\tag{55}$$

$$\log(\kappa_1) \sim \mathcal{N}(2.303586, 0.1),\tag{56}$$

$$\log(\tau_2) \sim \mathcal{N}(-4.549879, 0.1),\tag{57}$$

$$\log(\kappa_2) \sim \mathcal{N}(2.74132, 0.1),\tag{58}$$

$$\log(\tau_3) \sim \mathcal{N}(-5.605241, 0.1),\tag{59}$$

$$\log(\kappa_3) \sim \mathcal{N}(2.852892, 0.1).\tag{60}$$

Here, τ_1 and κ_1 are the hyperparameters of the spatially structured effect $w_1(s_i)$, while τ_2 and κ_2 are the hyperparameters of the effect $w_2(s_i)$. τ_3 and κ_3 are the hyperparameters of the shared component $z(s_i)$ (see Table 3). The means of the prior distributions for the hyperparameters are once again determined by the (unchanged) resolution of the grid and the values of the fixed smoothness parameters. In contrast, the default variance of 10 has been replaced by a variance of only 0.1 for each of the priors. These priors may seem to be too informative, but otherwise it has been showed to be impossible to avoid that only two of the processes take the whole spatially structured effects.

The prior distributions for β_1, β_2 and the interaction parameter a remain the default ones as denoted in Equations 41 and 42. Fitting the model with the stated priors took 2188 seconds, i.e. over half an hour, which is more than the running time of all the previous models together.

Regarding the posterior marginal distributions for the interaction parameter a and for the hyperparameters of the shared component $z(s_i)$ presented in Figure 29, it is striking that the interaction parameter is well estimated, but the hyperparameters of the shared component cannot be recovered.

The true interaction parameter $a = 0.5$ lies inside the 95% credible interval, far away from the borders. Not only that the posterior mean is a satisfying point estimate with a value of 0.6938, more importantly, the interaction is significantly positive with the value zero being outside the credible interval. So the interaction parameter a is very well estimated. But this is almost worthless if the shared component is not well detected – which seems to be the case here. A look at the hyperparameters of the shared component reveals a very bad fit. The true values for κ_3 and the nominal range lie far outside the 95% credible intervals. The necessity to impose priors with low variances on the hyperparameters may have additionally falsified the fit here. We will see below how heavy the implications on the fit of the shared component parameters $z(s_i)$ are.

The same issue must be considered when regarding the posteriors for the hyperparameters of the two process-specific spatially structured effects $w_1(s_i)$ and $w_2(s_i)$ (see Figure 30). Especially the hyperparameters of the second spatially structured effect $w_2(s_i)$ are estimated very incorrect. In contrast, the hyperparameters for the first spatially structured effect are estimated satisfactorily with the true values always lying inside the particular credible intervals. Apparently, the model fitting procedure has more problems in distinguishing the two processes $z(s_i)$ and $w_2(s_i)$ than
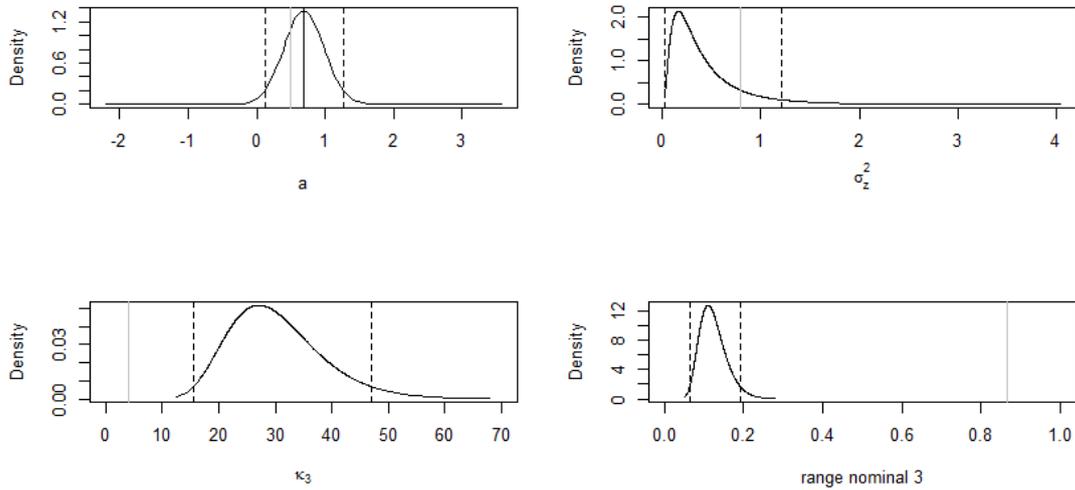
Figure 29: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the interaction parameter and the hyperparameters of the shared component in the shared component example. The notation is the same as in Figure 12.

the processes $z(s_i)$ and $w_1(s_i)$. A possible reason might be that the effects $z(s_i)$ and $w_2(s_i)$ are similarly smooth ($\nu_3 = 1.5$ and $\nu_2 = 1.2$) and hence difficult to separate from each other while the process $w_1(s_i)$ is very rough ($\nu_1 = 0.3$) and so easier to discriminate.

On top of that, the intercepts of the two latent processes (first and fifth panel) are also bad estimated with the true values being outside the 95% credible intervals. All things considered, the fits for the hyperparameters of the spatially structured effects and for the intercepts of the latent processes are the worse of all the examples considered in our simulation studies.

An important aspect for the shared component model is that the three true spatially structured effects $z(s_i)$, $w_1(s_i)$ and $w_2(s_i)$ should be detected in the model fitting step. The fits of the three spatially structured effects as well as the true simulated effects are illustrated in Figure 31.

It is obvious that the shared component $z(s_i)$ (first panel) is not well identified – we already suspected this when assessing the quality of the hyperparameter fits for the shared component. The true spatially effect parameters $z(s_i)$ lie inside the 95% equal-tailed credible interval, but this is mainly due to the identifiability issue that results in wide intervals for the three spatially structured effects (see the corresponding paragraph in Section 3.2.1). In contrast, the posterior means as point estimates for the shared component fail completely to follow the path of the true effect. Being lower than the true shared component at the beginning of the observation range, they particularly fail at the end of the observation range where the posterior means are much lower than the true effect.
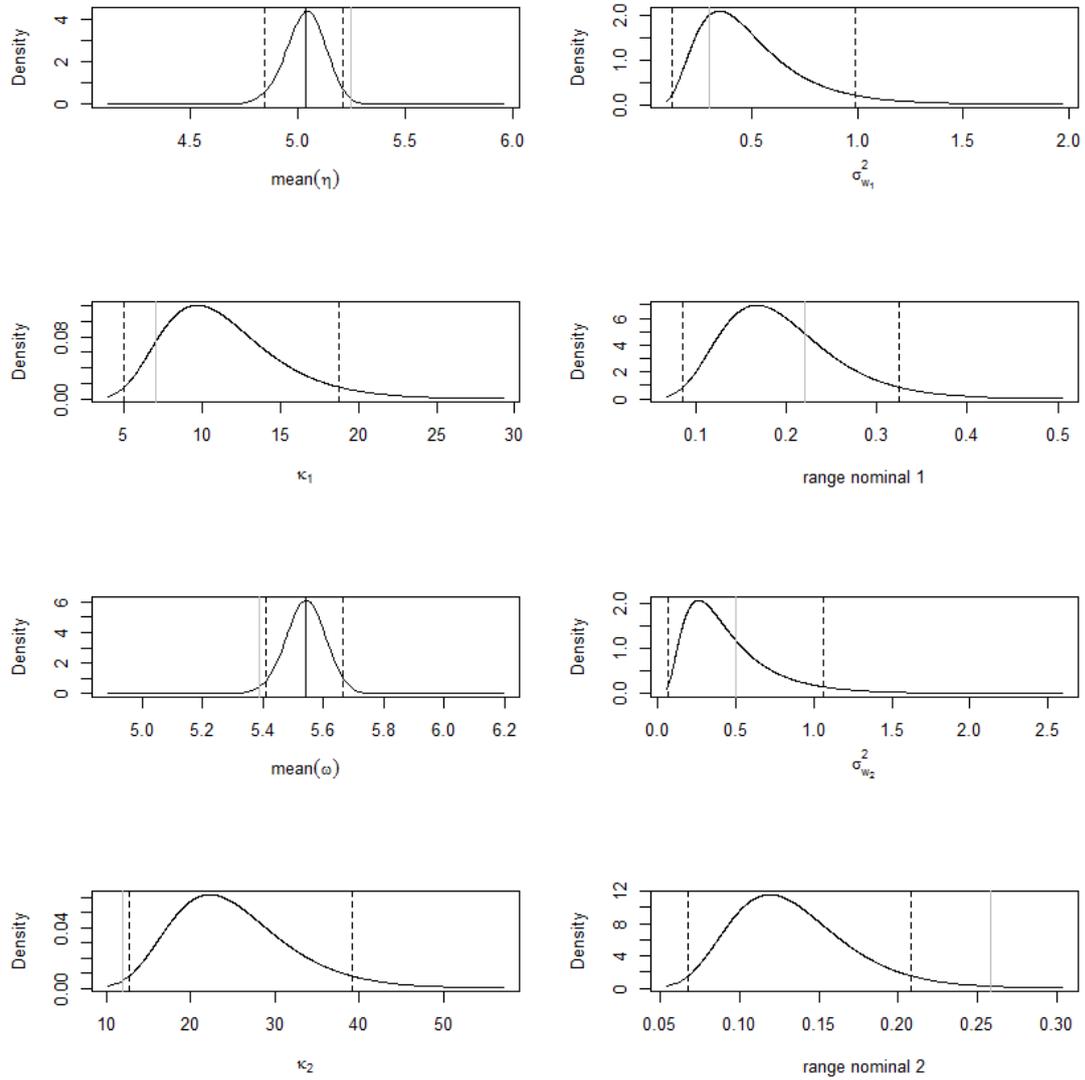
Figure 30: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the parameters of the two latent processes and the hyperparameters of the two process-specific spatially structured effects in the shared component example. Four top panels: The mean of the first latent process η_i and the hyperparameters of the first spatially structured effect $w_1(s_i)$. Four bottom panels: The mean of the second latent process ω_i and the hyperparameters of the second spatially structured effect $w_2(s_i)$. The notation is the same as in Figure 12.
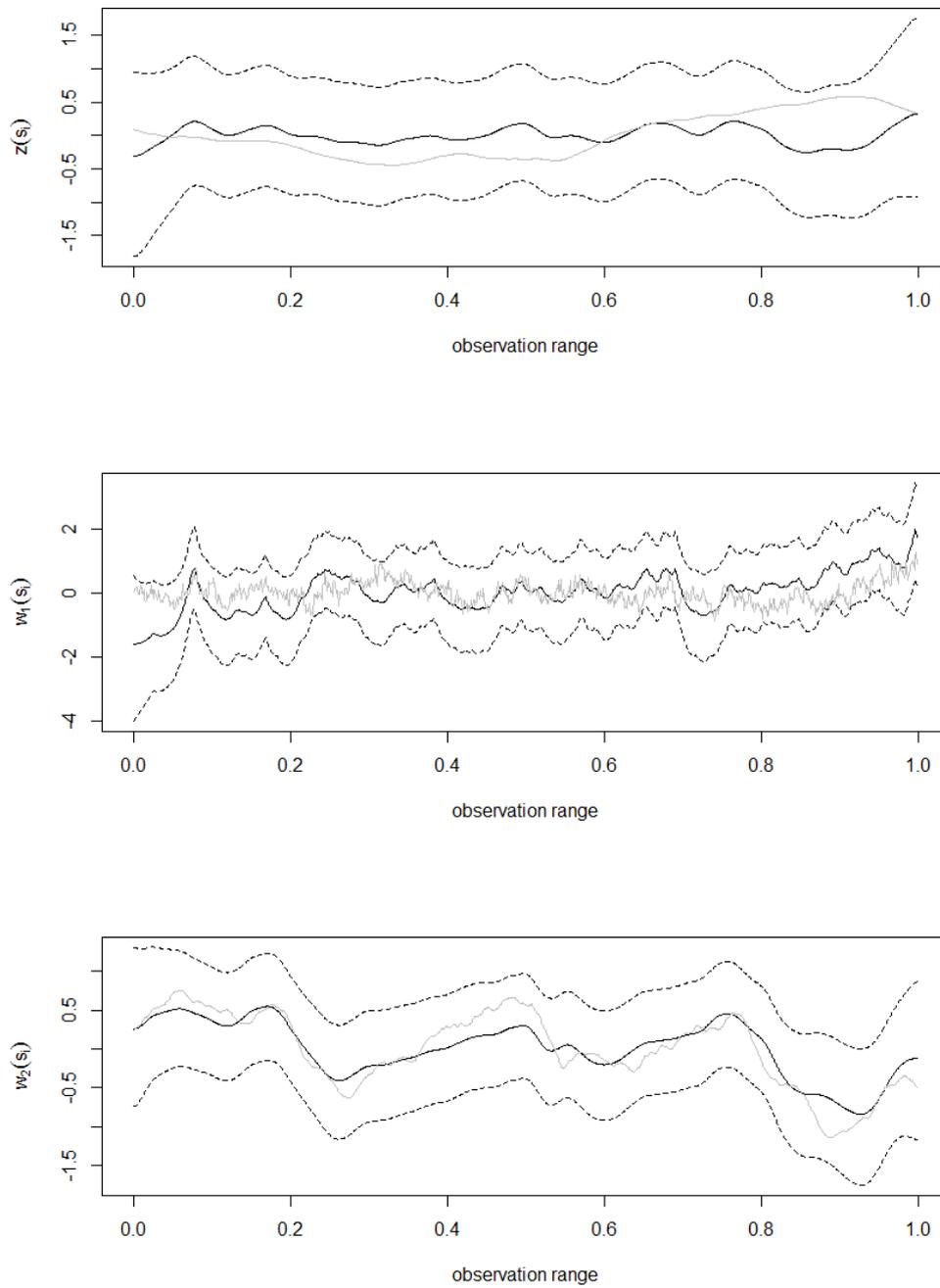
Figure 31: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the spatially structured effects in the shared component example. The notation is the same as in Figure 9.

The two process-specific spatially structured effects $w_1(s_i)$ and $w_2(s_i)$ (shown in the middle and bottom panel of Figure 31 respectively) are better recognized. The fit for the first spatially structured effect $w_1(s_i)$ shows indeed a similar pattern to that of the LMC example displayed in the top panel of Figure 27, although it is less correct especially at the end of the observation range where the posterior means are too high (note that the effects $w_1(s_i)$ and $w_2(s_i)$ are the same for the LMC and the shared component example). The fit for the second spatially structured effect $w_2(s_i)$ is even better for the shared component example than for the LMC example as the posterior means are better in recovering the oscillations of the true effect.

Regarding the posteriors for the linear predictors of the two latent processes η_i and ω_i, illustrated in Figure 32, it is visible that the true linear predictors lie almost everywhere inside the credible intervals. In contrast to the fits of the spatially structured effects $z(s_i)$ and $w_1(s_i)$, the point estimates for the first linear predictor are correct at the end of the observation range. Obviously, the poor fits of $z(s_i)$ and $w_1(s_i)$ counterbalance each other there. Nevertheless, the posterior means for the first linear predictor are too low at the beginning of the observation range.

The fit for the second linear predictor seems to be even better than that in the LMC example as there is no zone with very bad fitted values as it is visible in the bottom panel of Figure 26. The MAE and MSE scores confirm our graphic based analysis. The spatially structured effect $w_1(s_i)$ and the linear predictor for the first latent process η_i show higher values in the MAE and the MSE than in the LMC example. In return, the MAE and MSE for the process $w_2(s_i)$ and the second linear predictor ω_i are lower.

Remark 6.4. Unsatisfied by the poor fit of the shared component – on which the focus of the shared component model generally lies – we conducted further simulation studies with several simulations based on many different (hyper-)parameters, but without satisfying results in estimating the true shared component. Even when conducting the inference on the latent Gaussian processes instead of the overlying Poisson processes, the true shared component could not be revealed. We also tried what happened if the hyperparameters are fixed a priori to their true values – even then, the correct shared component was not detected. Apparently, it is difficult to regain the shared component *in our simulation based setting with R-INLA* what does not mean that it does not work in other settings. See for example Knorr-Held and Best [KHB01] for an application of the shared component model in disease mapping.

The shared component example – model comparison Regarding the DIC and WAIC values for the models of Table 5 fitted to the shared component example reveals the "true" shared component model as the best model. However, there are two similar models regarding the DIC (differences between 2 and 5) and even five very similar models according to the WAIC (differences in the WAIC of lower than 2). We do not consider the CPO values here as the shared component model is an outlier concerning the CPO value – as it has been in nearly all our simulation studies.

The similar models in terms of the DIC or WAIC are the independence model, the first and
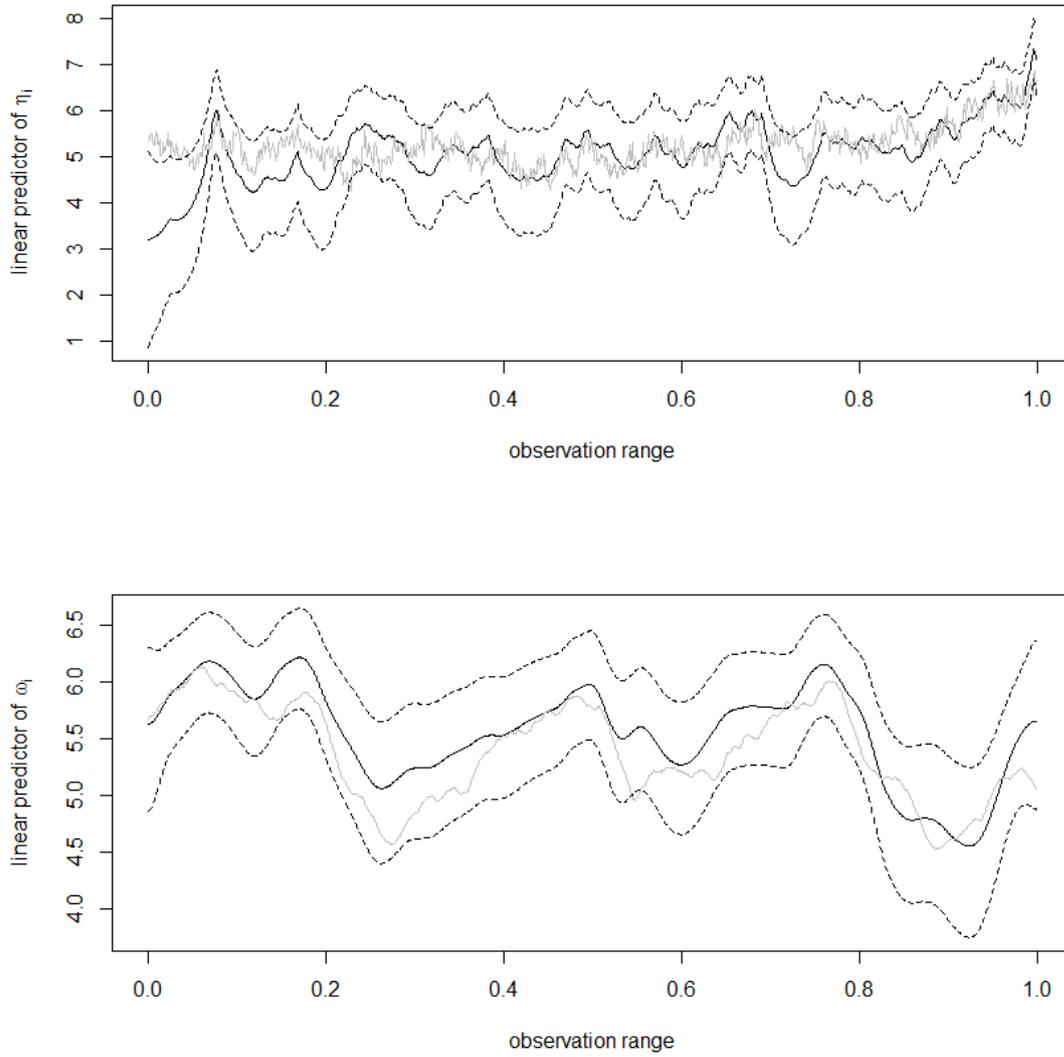
Figure 32: Modelling by bivariate Gauss-Markov processes: Posterior marginal distributions for the linear predictors of the latent processes in the shared component example. The notation is the same as in Figure 9.

second linear dependence models, the second non-linear dependence model and the LMC. While one would normally prefer the independence model from a set of very similar models as it is the simplest model, the situation is less clear here: There are no models which are very similar to the shared component model in terms of the DIC. As always, the final decision should depend on the particular application: If one wants to detect a shared component, one would possibly take the shared component model as it is recognized as an appropriate model by the model comparison scores. In contrast, if one wants to find the simplest model that still explains the situation in a satisfiable way, one would take the independence model here. As already stated several times, there is no objective "best" model, the decision should depend on the circumstances rather than to solely rely on a simplified score.

7 Summary and discussion

In the present work, we have modelled interactions between two one-dimensional point processes by means of log-Gaussian Cox processes. Several methods for modelling these interactions have been established: On the one hand, there is the group of functional connection methods, comprising the independence model as a baseline model and the linear and non-linear dependencies on the spatially structured effect, the latent process or the Poisson process. On the other hand, one has the class of bivariate Gauss-Markov models, containing the LMC, the shared component model and the bivariate Matérn model.

Techniques for speeding up the inference of these models have been provided through the approximation of Gaussian processes by Gauss-Markov processes as well as through the INLA approach. The diversity of modelling methods has been illustrated in several simulated examples. Inference based on these simulations has been conducted within the R package R-INLA to reveal the potentials and shortcomings of the different interaction modelling approaches.

For this aim, we have used several comparison scores: the DIC, the WAIC and the CPO. While the first two have shown to give similar results, the CPO has the drawback of not being able to rank the fit of the shared component model correctly. Apart from this failure, the CPO has also been useful for detecting appropriate model fits for the simulated example at hand.

In the following, we will discuss the different modelling techniques regarding the results of the related simulation studies. This summary is intended to help a user of our models to choose an appropriate model for the particular application at hand.

Independence model   We start with the independence model. It is the benchmark model with regard to the quality of the model fits, but apart from that not very informative as it does not contain any interactions. The independence is well detected by the model comparison scores when applying our rule that the simplest model should be preferred among very similar models.

Linear dependence model   The linear dependence model for the positive linear dependence example has a running time being marginal higher than that of the independence model. The most important fact is that the interaction parameter is well detected and even significantly higher than zero (with the true value $\beta_3 = 0.4$). The fits for the spatially structured effects and the linear predictors are almost as good as in the independence example and the positive linear dependence model is recognized as the best one regarding the model comparison scores.

In contrast, the linear dependence model fails completely to model the negative interaction in the negative linear dependence example. The interaction is estimated to be positive instead. The linear dependence model is consequentially not detected in the model comparison step where the independence model is rated as the best one. As the negative dependency can be detected by the *non-linear* dependence model, our recommendation is to fit the dataset at hand once by means of the linear dependence model and once by means of the non-linear dependence model whenever a negative interaction must be considered.

Non-linear dependence model   For the non-linear dependence example, there is the hierarchical modelling issue of the non-linear dependence model. As the basis values for the Bayesian P-splines must be fixed, the point estimates for the first spatially structured effect have to be used which results in a smaller range of values as the fits of the effects are generally smoother than the true effects. So the interaction may not be correctly estimated. The non-linear interaction function of the simulation is well estimated, but the remaining (hyper-)parameters of the second spatially structured effect become a little bit worse. In the model comparison, however, the non-linear dependence model is only the second-best as the linear dependence model is able to embed most of the non-linear interaction.

Dependence on the latent process model   The dependence on the latent process model in the dependence on the latent process example is equal to a simplified non-linear dependence model as the spatially unstructured effect cannot be detected in our approach with R-INLA (see Remark 5.1). The non-linear interaction of the example can be satisfactorily recognized. In the model comparison, the dependence on the latent process model is well detected, since in contrast to the non-linear dependence example the (simplified) linear dependence models are not flexible enough to model the non-linear interaction.

LMC   The LMC has a very long running time, but the interaction parameters are well recognized. The same applies for the spatially structured effects, a result which is rather surprising, taking the identifiability issue into account (see Section 3.2.1). Apparently, the a priori fixed smoothness parameters and different priors for the spatially structured effects are enough to distinguish them. In the model comparison step, much models proved to be overfitted. Sorting them out, the LMC remains with a group of equally rated models, containing the linear dependence model as the simplest model. It depends on the situation at hand if to choose the LMC then. One may prefer it when wishing to detect the "main" underlying effect for each of the Poisson processes as well as the amount of interaction concerning the other Poisson process.

Shared component model   Finally, the shared component model has a running time being higher than those of all the other models together. The shared component model estimates the interaction parameter in the shared component example very well. This is nearly worthless, however, as the fit for the shared component is completely incorrect. Nevertheless, the shared component model is detected as the best one concerning the model comparison scores with no very similar models, but with the independence model rated similar enough for being considered.

So which model should be chosen in which situation? Additionally to the analysed simulation studies, some further aspects must be considered. For example, if one wants to detect an underlying common component, the decision is already made in advance as the shared component model is the method of choice in this setting. This might be the case when one wants to detect a common risk factor of two diseases.

Furthermore, the LMC model formulation may be rewritten as a linear dependence model (see Remark 3.4) and will consequently hardly be rated as the best one in the model comparison step. However, as stated in the results of the LMC simulation study above, one may prefer the LMC among a group of similar rated models when it is wished to detect the "main" underlying effect for each of the Poisson processes and the amount of interaction with regard to the other Poisson process.

Consequently, the model choice highly depends on the application at hand. Nevertheless, some general conclusions may be drawn on the basis of our studies – although one has to bear in mind that they are based on simulations solely.

Firstly, the linear dependence model seems to be an appropriate choice when examining positive interactions. It does not only estimate the interactions correctly, but may also embed non-linear interactions when the oscillations of the latent processes are not too high.

Secondly, in spite of all its drawbacks, the hierarchical non-linear modelling by means of Bayesian P-splines seems to be a good choice when modelling Poisson processes with high variations within small distances as they exhibit the required flexibility to model those.

Thirdly, one has to be extremely careful when modelling negative interactions as the linear dependence model in R-INLA fails in this setting. Modelling by means of non-linear dependencies may reveal the problem.

Finally, the identifiability issue seems to be a problem for the shared component model within R-INLA, but not for the LMC, where the spatially structured effects as well as the interactions are well detected in our simulation studies.

We complete our conclusions with a warning regarding the use of the LMC and the shared component model: While the running time of those models is still practicable in our setting, it might become a problem when applying our methods to use cases with much larger point patterns. Especially when extending our approach to large two-dimensional point patterns, the implementation of these methods may not be computationally feasible any more.

8 Extensions

Some possible extensions of the present work shall be discussed in the following. Two obvious enhancements would be to examine the interactions of more than two Poisson processes or to transfer the methods described above to a setting including more than one dimension.

Multivariate interaction modelling for instance would diminish the drawback of the LMC that the number of independent Gaussian processes is restricted to be not higher than the number of processes to model the dependence between. In fact, it might be a favoured strategy to find *less* spatially structured effects than there are processes to be modelled, to get a model which is as simple as possible. This is done by Ribeiro et al. [RSP15] for example.

In contrast, interaction modelling of more than two Poisson processes might be a problem for non-linear dependence models as one would have to fix a priori the probable hierarchical structure. Otherwise there would be plenty of models with different hierarchies what quickly becomes computationally infeasible.

Moreover, multivariate Matérn models as an extension of bivariate Matérn models are discussed in the work of Hu et al. [HSLR13].

Concerning interaction models in more than one dimension, the most natural case is the modelling of dependencies between two two-dimensional point processes. Indeed, the analysis of two-dimensional point patterns plays the leading part in almost each of the publications stated in the present work, since they are the common use case. In contrast, multidimensional point patterns with more than three dimensions are rather a theoretical construct, although many of the theoretical results remain basically the same.

Another dimension which may be added is the time axis. Blangiardo et al. [BCBR13] show for instance how to implement such spatio-temporal models within R-INLA.

As mentioned in Section 3, Illian et al. [ISR12] also account for observed covariates. These may equally be inserted in our approach when not studying simulations any more but applying our methods to "real-world" datasets. Furthermore, Illian et al. make use of a constructed covariate to include local interactions, clustering and competition. This approach may be useful if the points represent individuals influencing each other, since mutual interactions of the points in a particular pattern may in general not be modelled using log-Gaussian Cox processes. While it is indeed possible to model local repulsion very well by means of a constructed covariate, modelling of local clustering is more problematic. See Illian et al. [ISR12, Ch.3] for some simulation studies on this issue.

Finally, we have limited ourselves to the case of isotropic Matérn covariance functions which are only dependent on the Euclidean distance between two particular points. While we recommend to maintain the Matérn covariance to still profit by the explicit link between Gaussian processes and Gauss-Markov processes, Lindgren et al. [LRL11] show that the explicit link remains true for non-stationary Matérn models where the covariance function is dependent on the exact locations. This fact is particularly useful when analysing globally collected data as the points are obtained on a differently curved sphere then. Instead of trying to flatten the surface to get the correct

distances, one can easily use non-stationary models without loosing the explicit link.

In summary, the present work shall constitute a basis for further work on interactions between point processes. In particular, the extension to two-dimensional use cases with much larger point patterns, where the introduced speed-up techniques reveal their full potential, are a long-term objective.

Acknowledgements

List of Figures

List of Tables

References

[BCBR13] Marta Blangiardo, Michela Cameletti, Gianluca Baio, and Håvard Rue. Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*, 7:39–55, 2013.

[BGRR15] Roger S Bivand, Virgilio Gómez-Rubio, and Håvard Rue. Spatial data analysis with r-inla with some extensions. *Journal of Statistical Software*, 63(20), 2015.

[FBV+15] Rob GA Fraaije, Cajo JF Braak, Betty Verduyn, Leonieke Breeman, Jos TA Verhoeven, and Merel B Soons. Early plant recruitment stages set the template for the development of vegetation patterns along a hydrological gradient. *Functional Ecology*, 29(7):971–980, 2015.

[FKL09] Ludwig Fahrmeier, Thomas Kneib, and Stefan Lang. *Regression: Modelle, Methoden und Anwendungen*, volume 2. Springer, 2009.

[GHV13] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models, 2013.

[GK15] Marc G Genton and William Kleiber. Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2):147–163, 2015.

[GKS10] Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 2010.

[GSBS04] Alan E Gelfand, Alexandra M Schmidt, Sudipto Banerjee, and CF Sirmans. Non-stationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

[GW94] Michel Grzebyk and Hans Wackernagel. Multivariate analysis and spatial/temporal scales: real and complex models. In *Proceedings of the XVIIth International Biometrics Conference*, volume 1, pages 19–33. Citeseer, 1994.

[HSLR13] Xiangping Hu, Daniel Simpson, Finn Lindgren, and Håvard Rue. Multivariate gaussian random fields using systems of stochastic partial differential equations. *arXiv preprint arXiv:1307.1379v2*, 2013.

[HSS+15] Xiangping Hu, Ingelin Steinsland, Daniel Simpson, Sara Martino, and Håvard Rue. Spatial modelling of temperature and humidity using systems of stochastic partial differential equations. *arXiv preprint arXiv:1307.1402v2*, 2015.

[IMS+13] Janine B Illian, Sara Martino, Sigrunn H Sørbye, Juan B Gallego-Fernández, María Zunzunegui, M Paz Esquivias, and Justin MJ Travis. Fitting complex ecological point

process models with integrated nested laplace approximation. *Methods in Ecology and Evolution*, 4(4):305–315, 2013.

[ISR12] Janine B Illian, Sigrunn H Sørbye, and Håvard Rue. A toolbox for fitting complex spatial point process models using integrated nested laplace approximation (inla). *The Annals of Applied Statistics*, 6(4):1499–1530, 2012.

[ISRH12] Janine B Illian, Sigrunn H Sørbye, Håvard Rue, and Ditte Hendrichsen. Using inla to fit a complex point process model with temporally varying effects–a case study. *Journal of Environmental Statistics*, 3(7), 2012.

[KHB01] Leonhard Knorr-Held and Nicola G Best. A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):73–85, 2001.

[KLSR16] Elias T Krainski, Finn Lindgren, David Simpson, and Håvard Rue. The r-inla tutorial on spde models, 2016.

[Lin12] Finn Lindgren. Continuous domain spatial models in r-inla. *The ISBA Bulletin*, 19(4):14–20, 2012.

[LR15] Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19), 2015.

[LRL11] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[MW03] Jesper Moller and Rasmus Plenge Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

[MW07] Jesper Møller and Rasmus P Waagepetersen. Modern statistics for spatial point processes*. *Scandinavian Journal of Statistics*, 34(4):643–684, 2007.

[RB13] Qian Ren and Sudipto Banerjee. Hierarchical factor models for large spatially misaligned data: a low-rank predictive process approach. *Biometrics*, 69(1):19–30, 2013.

[RMC09] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009.

[RSP15] Manuel Castro Ribeiro, António Jorge Sousa, and Maria João Pereira. A coregionalization model to assist the aelection process of local and global variables in

semi-parametric geographically weighted poisson regression. *Procedia Environmental Sciences*, 26:53–56, 2015.

[SG03]   Alexandra M Schmidt and Alan E Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.

[Wac03]   Hans Wackernagel. *Multivariate geostatistics*. Springer, 2003.

Declaration

I declare on oath that I completed this work on my own and that information which has been directly or indirectly taken from other sources has been noted as such.

Munich; May 13, 2016

....
Oliver Engl