

This is the peer reviewed version of the following article:

Variance partitioning in spatio-temporal disease mapping models / Franco-Villoria, Maria; Ventrucci, Massimo; Rue, Håvard. - In: STATISTICAL METHODS IN MEDICAL RESEARCH. - ISSN 0962-2802. - 31:8(2022), pp. 1566-1578. [10.1177/09622802221099642]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

08/05/2026 08:13

(Article begins on next page)

Variance partitioning in spatio-temporal disease mapping models

Journal Title

XX(X):2–49

©The Author(s) ...

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/ToBeAssigned

www.sagepub.com/



2 **Maria Franco-Villoria**¹, **Massimo Ventrucci**² and **Håvard Rue**³

Abstract

Bayesian disease mapping, yet if undeniably useful to describe variation in risk over time and space, comes with the hurdle of prior elicitation on hard-to-interpret random effect precision parameters. We introduce a reparametrized version of the popular spatio-temporal interaction models, based on Kronecker product intrinsic Gaussian Markov Random Fields, that we name the variance partitioning (VP) model. The VP model includes a mixing parameter that balances the contribution of the main and interaction effects to the total (generalized) variance and enhances interpretability. The use of a penalized complexity prior on the mixing parameter aids in coding prior information in a intuitive way. We illustrate the advantages of the VP model using two case studies.

Keywords

intrinsic Gaussian Markov Random Fields; intrinsic CAR; Kronecker product GMRF; penalized complexity prior; spatio-temporal smoothing

1 Introduction

The Covid-19 pandemic has put the world at stake. In Italy, the first two cases were confirmed on 31st January 2020 and on 9th March 2020 a national lockdown was put in place by the authorities to control and reduce the expansion of the virus. Data on newly infected people have been routinely collected since then to monitor the evolution of the disease. The study of the pandemic evolution can be tackled using disease mapping. Knowledge of how the infection has spread can help to evaluate the performance of containment measures. In particular, the quantification of the space-time interaction, which describes how the spatial patterns change over time, has been proposed as a way to deepen our understanding on the evolution of the disease¹.

Disease mapping models²⁻⁵ aim to describe the variation in risk of a particular disease over space and time. Data are usually available in the form of aggregated counts at some spatial level, such as counties, municipalities, etc. Additive time and space models have been long used to model disease rates⁶. More recently, the availability of complex data has made it possible to consider more complex models that include an interaction term to appropriately capture space-time relationships in the data (see for example Abellan et al.⁷, Knorr-Held⁸, Waller et al.⁹, Bernardinelli et al.¹⁰ to cite a few). Understanding the spatial distribution of disease risk or how it has evolved over time might be useful for public health authorities in planning resource allocation and identification of areas to be prioritized. In particular, the space-time interaction may reveal important information regarding the nature of the disease, for example

¹Department of Economics, University of Modena and Reggio Emilia, Italy²Department of Statistical Sciences, University of Bologna, Italy³CEMSE Division, King Abdullah University of Science and Technology, Thuwal,, Saudi Arabia

Corresponding author:

Franco-Villoria Maria, Department of Economics, University of Modena and Reggio Emilia, Italy
Email: maria.francovilloria@unimore.it

30 suggesting whether a new disease is possibly infectious¹¹ or the existence
31 of additional causes in non-infectious cases¹². Thus a model that is able
32 to quantify the importance of this term is desirable from a practical point
33 of view; in this paper we introduce a model parametrization that partitions
34 the total variance into main and interaction effects so that the contribution
35 of each of those can be quantified.

36 Crude disease rates are unreliable due to sampling variability so
37 smoothing is used to borrow information across neighbouring areas and
38 time points. For this reason, disease mapping has been developed mainly
39 in a Bayesian hierarchical model formulation where the building blocks
40 of a smooth in one and more dimensions can be modelled using intrinsic
41 Gaussian Markov Random Fields (IGMRF) such as the first and second
42 order random walk¹³ or the ICAR¹⁴ models. For modelling interactions
43 statisticians have used tensor products smoothers, where, in a Bayesian
44 framework, the penalty can be seen as a special type of GMRF called
45 Kronecker product GMRFs¹⁵.

46 In Bayesian spatio-temporal disease mapping the precision parameter of
47 the IGMRFs plays a role in controlling the degree of smoothing applied
48 over time and space. A number of issues related to prior elicitation need
49 to be addressed when dealing with intrinsic models. Firstly, the precision
50 matrix is singular, which means that the total variance that we aim to
51 partition is not finite. In order to define priors on the variance components
52 we can rely upon the concept of generalized variance of an IGMRF; this
53 has been defined by Sørbye and Rue¹⁶ as the geometric mean of the
54 diagonal elements of the generalized inverse of the precision matrix of
55 the IGMRF, and can only be computed upon linear constraints.

56 A second issue to bear in mind is that the generalized variance of an
57 IGMRF depends on the structure matrix, and hence it changes depending
58 on things like the temporal and spatial resolution or the size of the dataset
59 at hand. This means that interpretation of the precision parameter becomes

60 case-dependent, making prior elicitation and parameter interpretation
61 difficult. To avoid this problem, Sørbye and Rue¹⁶ advise scaling the
62 structure matrix so that the generalized variance is equal to 1; this way the
63 precision parameter is automatically rescaled and the prior has the same
64 meaning regardless of the graph structure¹⁷. Scaling becomes particularly
65 relevant in the context of space-time models, as otherwise differences
66 in the structure matrices of the spatial, temporal and spatio-temporal
67 terms would have an impact on the priors for the corresponding precision
68 parameters that we cannot control. By scaling the structure matrix of the
69 temporal and spatial random effects, the structure matrix of the interaction,
70 defined as a Kronecker IGMRF, is automatically scaled.

71 Further to the issues mentioned above, the choice of priors for variance
72 parameters has received much attention in the literature^{5,18–20}. Part of
73 the hassle in choosing a prior stems from the difficulty of interpreting
74 variance parameters, especially for intrinsic processes, where the standard
75 deviation is to be interpreted as a conditional one^{19,21}. On top of that, in
76 models with various terms, the tendency is to set priors independently for
77 each precision parameter, while some authors are beginning to recognize
78 that it might be more practical to think about total variability and how each
79 term in the model contributes to that rather than to concentrate on single
80 variance components separately^{5,21–23}. In the context of disease mapping,
81 Wakefield⁵ proposes using an inverse Gamma prior on the total variability,
82 along with a Beta prior that distributes the variance between a spatially
83 correlated random field and a spatially unstructured effect (the so called
84 BYM model²⁴). Using a similar parametrization, Riebler et al.²¹ present
85 a prior that shrinks towards no spatial effect following the penalized
86 complexity (PC) prior approach of Simpson et al.²⁰. Outside the disease
87 mapping literature, Ventrucci et al.²³ develop a PC prior in one-factor
88 mixed models for the relative contribution of group-specific variability.
89 In a more general context, Fuglstad et al.²² introduce a framework for

90 hierarchically distributing the variance in additive models, where, at each
91 level of the total variance decomposition, ignorance or preference about
92 the variance contribution of a term is expressed via a Dirichlet or a
93 PC prior, respectively. We add to the literature by considering also the
94 temporal dimension in disease mapping models. In particular, all the terms
95 in the model (main effects and interaction) are assumed to follow intrinsic
96 models. This differentiates our work from the literature mentioned above.

97 In this work, we revisit the spatio-temporal models proposed by Knorr-
98 Held⁸, where the space-time interaction term can be one of four different
99 types, depending on the degree of dependence assumed between time and
100 space. These four types are characterized by different prior assumptions,
101 expressed in terms of a Kronecker product. We propose an intuitive
102 reparametrization that leads to partitioning the generalized variance
103 between the main effects and interaction. The main and interaction effects
104 are not independent, and hence using a joint prior on those terms is
105 preferable. We do so by including a mixing parameter that 1) eases
106 interpretation and 2) naturally leads to a prior that is intuitive to elicit. One
107 of the advantages of the Bayesian framework is that whenever information
108 on the disease process is available, it can be encoded into the prior¹².
109 Often, the epidemiologist might have an intuition on how important
110 the interaction term is in explaining the spatio-temporal variation of
111 a particular disease. However, translating this information in terms of
112 a precision parameter is not trivial at all. We follow the penalized
113 complexity prior (PC) framework of Simpson et al.²⁰ to derive a prior for
114 the mixing parameter that avoids overfitting by construction and allows
115 the user to code any prior information easily. This way we alleviate both
116 problems, by considering an interaction model that not only enhances
117 interpretability but also permits a more intuitive construction of the prior.
118 We call this reparametrized version the *variance partitioning (VP) model*.

119 The proposed methodology is applicable to any of the four space-time
120 interactions described in Knorr-Held⁸.

121 The rest of the paper is organized as follows. Section 2 covers spatio-
122 temporal disease mapping models, with a particular emphasis on the
123 space-time interaction framework by Knorr-Held⁸, followed by a brief
124 discussion of priors for variance parameters with special attention to the
125 PC prior approach. In Section 3 the VP model is described in detail and
126 the PC prior for the mixing parameter is presented, while the technical
127 details are relegated to the supplementary material. Section 4 illustrates
128 the proposed model on two case studies, a well known example in the
129 disease mapping literature and an Italian Covid-19 dataset. The paper
130 closes with a discussion in Section 5.

131 2 Spatio-temporal disease mapping

132 Consider data on n_1 time points and n_2 non-overlapping areas, y_{ij} is the
133 observed number of cases at time $i = 1, \dots, n_1$ and area $j = 1, \dots, n_2$.
134 The most commonly used models for y_{ij} are the binomial and the Poisson;
135 in either case, the model in the linear predictor scale can be written as
136 $\eta_{ij} = \alpha + f_1(i) + f_2(j) + f_{12}(i, j)$, where $f_1(i)$ and $f_2(j)$ represent the
137 main temporal and spatial effects respectively and the function $f_{12}(i, j)$
138 captures the space-time interaction. The model can be parametrized with
139 random effects as

$$\eta_{ij} = \alpha + \beta_{1i} + \beta_{2j} + \delta_{ij}, \quad (1)$$

140 where $\beta_1 = (\beta_{1,1}, \dots, \beta_{1,n_1})^T$ and $\beta_2 = (\beta_{2,1}, \dots, \beta_{2,n_2})^T$ are vectors
141 of random effects describing the temporal and spatial main effect,
142 respectively, and $\delta = \{\delta_{ij}\}, i = 1, \dots, n_1, j = 1, \dots, n_2$ is the vectorized
143 spatio-temporal interaction term. The random effects β_1, β_2 and δ are
144 typically assumed as smooth processes modelled using intrinsic Gaussian
145 Markov Random Fields (IGMRF,¹⁵), a special type of improper GMRF,

146 defined below. Appropriate constraints²⁵ need to be imposed to ensure
 147 identifiability of the terms in (1). The constraints on the interaction term
 148 are summarized on Table 1, while on the temporal and spatial main effects
 149 it is enough to impose a sum to zero constraint. As usual, any available
 150 covariates can be included in model (1) as fixed effects.

151 **Definition 1.** Improper GMRF. *Let \mathbf{Q} be an $n \times n$ symmetric*
 152 *positive semi-definite (SPSD) matrix with rank $n - p > 0$. Then $\mathbf{x} =$*
 153 *$(x_1, \dots, x_n)^T$ is an improper GMRF of rank $n - p$ with parameters*
 154 *$(\boldsymbol{\mu}, \mathbf{Q})$ if its density is*

$$\pi(\mathbf{x}) = (2\pi)^{\frac{-(n-p)}{2}} (|\mathbf{Q}|^*)^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu})\right),$$

155 where $|\mathbf{Q}|^*$ is the generalized determinant of the precision matrix \mathbf{Q} .
 156 Improper GMRFs are used as smoothing priors in structured additive
 157 regression (STAR) models, a flexible class including generalized linear
 158 mixed models, temporally dynamic models, spatial varying coefficient
 159 models, etc; for an account of STAR models see Fahrmeir et al.²⁶ and
 160 references therein.

161 Following Rue and Held¹⁵ we define an IGMRF of order 1 as an
 162 improper GMRF where $\mathbf{Q}\mathbf{1} = \mathbf{0}$, i.e., the precision matrix is singular with
 163 null space spanned by a column vector of ones, $\mathbf{1}_n$ of length n . Popular
 164 examples of an IGMRF of order 1 are the first order random walk (RW1),
 165 which is a possible option to model the temporal main effect β_1 , and the
 166 intrinsic conditional autoregressive (ICAR) model by Besag¹⁴, which is
 167 often assumed in disease mapping to model the spatial effect β_2 when
 168 smoothing across neighbouring regions is required.

169 An IGMRF of order 2 is an improper GMRF whose precision matrix is
 170 singular and its null space is spanned by a constant vector $\mathbf{1}_n$ and a linear
 171 vector $(1, \dots, n)^T$. A popular example is the second order random walk
 172 (RW2;¹³), popularly used for modelling smooth covariate effects in STAR

173 models, and often implemented in spatio-temporal disease mapping for
 174 modelling the main temporal effect β_1 when smoothness in the disease
 175 risk over time is anticipated.

176 All the IGMRFs described above have in common that their precision
 177 matrix can be written as $\mathbf{Q} = \tau \mathbf{R}$, where τ is a precision parameter and
 178 \mathbf{R} is a known structure matrix that encodes the dependence structure. In
 179 particular, for the RW1

$$R_{k,l} = \begin{cases} 1 & k = l \in \{1, n\} \\ 2 & k = l \in \{2, \dots, n-1\} \\ -1 & k \sim l \\ 0 & \text{otherwise,} \end{cases}$$

180 where notation $k \sim l$ indicates contiguous time points. For the ICAR, the
 181 structure matrix is given by

$$R_{k,l} = \begin{cases} m_k & k = l \\ -1 & k \sim l \\ 0 & \text{otherwise,} \end{cases}$$

182 where m_k is the number of neighbours for region k and notation $k \sim l$
 183 indicates contiguous areas that share a common border. The structure
 184 matrix of a RW2 can be written as $\mathbf{R} = \mathbf{D}^T \mathbf{D}$ where \mathbf{D} is a second order
 185 difference matrix of dimension $(n-2) \times n$.

186 It is common in the disease mapping literature to consider one or both
 187 main effects f_1 and f_2 as a sum of structured and unstructured effects, so
 188 that model (1) becomes

$$\eta_{ij} = \alpha + \beta_{1_i} + \epsilon_{1_i} + \beta_{2_j} + \epsilon_{2_j} + \delta_{ij}, \quad (2)$$

189 where $\epsilon_1 \sim N(\mathbf{0}, \tau_{\epsilon_1} \mathbf{I}_{n_1})$, $\epsilon_2 \sim N(\mathbf{0}, \tau_{\epsilon_2} \mathbf{I}_{n_2})$. Typically, a RW1 or RW2
 190 model is assumed for the temporal effect $\beta_1 \sim N(\mathbf{0}, \tau_1^{-1} \mathbf{R}_1^-)$ and an

191 ICAR is assumed for the spatial effect $\beta_2 \sim N(\mathbf{0}, \tau_2^{-1} \mathbf{R}_2^-)$, where
 192 notation M^- indicates the generalized inverse of matrix M . The
 193 combination of the structured and unstructured spatial terms $\beta_{2_j} + \epsilon_{2_j}$ is
 194 commonly known as the BYM model²⁴.

195 2.1 Modelling interactions via Kronecker product IGMRFs

We describe now the interaction term δ in Eq. (2). Smoothness is induced by assuming

$$\delta \sim N(\mathbf{0}, \tau_{12}^{-1} \mathbf{R}_I^-),$$

196 which is a Kronecker product IGMRF with precision $\mathbf{Q} = \tau_{12} \mathbf{R}_I$, i.e.
 197 an improper GMRF with precision given by the Kronecker product
 198 of two IGMRFs. These models are used for smoothing spatial and
 199 spatio-temporal data, and they are the Bayesian equivalent of tensor
 200 product spline models²⁷. Knorr-Held⁸ envisions four different types of
 201 interactions, reported in Table 1. Interaction type I can be seen as
 202 unstructured variation due to unobserved covariates, while interaction
 203 types II and III allow for the temporal trend to change from location to
 204 location and the spatial trend to change over time, respectively, but in
 205 an independent manner. Interaction type IV is the most complex one,
 206 assuming that the temporal trend changes with location in a spatially
 207 dependent way, or equivalently, that the way in which the spatial trend
 208 changes over time is time-dependent.

Table 1. The four types of interactions in spatio-temporal smoothing according to Knorr-Held⁸. The IGMRF on the interaction parameter vector δ has structure \mathbf{R}_I given by a Kronecker product; $r_1 = 1$ or 2 depending on the order of the RW assumed for the time effect.

type	\mathbf{R}_I	$\text{rank}(\mathbf{R}_I)$	linear constraints on δ
I	$\mathbf{I}_{n_2} \otimes \mathbf{I}_{n_1}$	$n_1 n_2$	not needed
II	$\mathbf{I}_{n_2} \otimes \mathbf{R}_1$	$n_2(n_1 - r_1)$	$[\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1}]^T \delta = \mathbf{0}_{n_2}$
III	$\mathbf{R}_2 \otimes \mathbf{I}_{n_1}$	$(n_2 - 1)n_1$	$[\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1}]^T \delta = \mathbf{0}_{n_1}$
IV	$\mathbf{R}_2 \otimes \mathbf{R}_1$	$(n_2 - 1)(n_1 - r_1)$	$[\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1}]^T \delta = \mathbf{0}_{n_2}; \quad [\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1}]^T \delta = \mathbf{0}_{n_1}$

Model (1) includes different precision parameters τ_1 and τ_2 for smoothing over time and space and an additional one, τ_{12} , controlling the variance of the interaction term, which yields a model able to capture the smooth spatio-temporal structure underlying the data with high flexibility. However, these models have limitations in terms of interpretation of the results, as precision parameters are not informative about the total variance explained by the associated components and the priors are not easy to elicit (see Section 2.2). We propose an alternative parametrization to address these issues in Section 3.

2.2 Priors for the precision parameters

There are two main challenges in prior choice for the precision parameters in model (1). The first problem regards the so called scaling issue that affects IGMRFs in general; Sørbye and Rue¹⁶ proposed addressing this issue by scaling the precision structure \mathbf{R} so that the geometric mean of the diagonal elements in \mathbf{R}^{-1} is 1. In this way, the prior for τ will roughly encode the same degree of complexity across different types of structures and hence will have the same interpretation. Of particular interest is the spatial case where, after scaling the precision of the ICAR, the prior for the precision parameter becomes transferable across different applications using different graph structures.

The second challenge regards the structure of the Kronecker product IGMRF, which can be thought of as an extra layer of flexibility on top of the main effects model. The common practice is to set independent priors on each precision parameter, but this totally disregards the model structure. Popular choices are Gamma for τ , or half-t and uniform on the standard deviation $1/\sqrt{\tau}$ ¹⁸. The Gamma prior has repeatedly been pointed out as a poor choice often made by convenience; among the reasons why it should be avoided is that it forces overfitting or underfitting depending on the choice of its parameters^{19,20,28–30}.

238 In Section 3 we propose a novel modelling framework where the
 239 interaction is seen as a flexible extension of the main effects model, and the
 240 prior is set so that the interaction term shrinks to the main effects following
 241 the PC prior framework. Recently, PC priors have been proposed as a
 242 way to prevent overfitting, based on four simple principles, that we briefly
 243 summarize and illustrate below for the precision parameter τ of a Gaussian
 244 random effect. For further details the reader is referred to Simpson et al.²⁰.

245 Let π_1 denote the density of a model component w with precision
 246 parameter τ . This model component can be seen as a flexible extension of
 247 a based model with density π_0 and $\tau = \infty$ (i.e. absence of random effects).
 248 The four principles are:

- 249 1. Parsimony: The prior for τ should give proper shrinkage to $\tau = \infty$
 250 and decay with increasing complexity of π_1 , so that the simplest
 251 model is favoured unless there is evidence for a more flexible one.
- 252 2. The increased complexity of π_1 with respect to π_0 is measured using
 253 the Kullback-Leibler divergence KLD,³¹

$$\text{KLD}(\pi_1||\pi_0) = \int \pi_1(w) \log \left(\frac{\pi_1(w)}{\pi_0(w)} \right) dw.$$

For ease of interpretation, the KLD is transformed to a unidirectional distance measure

$$d(\tau) = d(\pi_1||\pi_0) = \sqrt{2\text{KLD}(\pi_1||\pi_0)}$$

254 that can be interpreted as the distance from the flexible model π_1 to
 255 the base model π_0 .

3. The PC prior is defined as an exponential distribution on the distance,

$$\pi(d(\tau)) = \lambda \exp(-\lambda d(\tau)),$$

with rate $\lambda > 0$. The PC prior for τ follows by a change of variable transformation, leading in this case to a type-2 Gumbel distribution with parameters $(1/2, \lambda)$:

$$\pi(\tau) = \frac{\lambda}{2} \tau^{-3/2} \exp(-\lambda \tau^{-1/2}), \quad \tau > 0, \lambda > 0. \quad (3)$$

4. The parameter λ in (3) can be selected by the user based on his prior knowledge of τ (or an interpretable transformation of it such as the standard deviation). This can be expressed in an intuitive way with a probability statement, e.g. setting U and a such that $\mathbb{P}(1/\sqrt{\tau} > U) = a$, so that $\lambda = -\log(a)/U$. Knowledge on the marginal standard deviation can aid in choosing a sensible value for U ; Simpson et al.²⁰ provide a practical rule of thumb: once the precision τ is integrated out, the marginal standard deviation of the random effect for $a = 0.01$ is about $0.31U$.

3 Partitioning the variance between main and interaction

We present below the VP model assuming model (1), but everything applies straightforwardly to model (2) as well; details about the VP version of model (2) can be found in Section 4.

From model (1) it is hard to quantify the relative contribution of the main and interaction components to the total variance, because the involved precision parameters are not interpretable in terms of the variance explained by the associated components. Our proposal is to reparametrize model (1) as a weighted sum of two IGMRFs representing the main and interaction components by means of a mixing parameter $\gamma \in [0, 1]$. We include a further mixing parameter $\phi \in [0, 1]$ to distribute the variance between the temporal and spatial main effects. Assume model (1), the reparametrized version of the linear predictor is

$$\begin{aligned} \eta_{ij} &= \alpha + \sqrt{\tau^{-1}} \left[\sqrt{1-\gamma} (\sqrt{1-\phi}\beta_{1_i} + \sqrt{\phi}\beta_{2_j}) + \sqrt{\gamma}\delta_{ij} \right], \\ \beta_1 &\sim N(\mathbf{0}, \tilde{\mathbf{R}}_1^-), \quad \beta_2 \sim N(\mathbf{0}, \tilde{\mathbf{R}}_2^-), \quad \delta \sim N(\mathbf{0}, \tilde{\mathbf{R}}_I^-), \end{aligned} \quad (4)$$

281 where $\tau > 0$ is an overall precision parameter, $0 < \gamma < 1$, $0 < \phi < 1$. We
 282 consider a RW1 or a RW2 prior on the temporal main effect β_1 and
 283 an ICAR prior on the spatial main effect β_2 as specified in Section 2.
 284 Note that, differently from model (1), the precision structures $\tilde{\mathbf{R}}_1$, $\tilde{\mathbf{R}}_2$
 285 have been scaled according to Sørbye and Rue¹⁶. The interaction term
 286 δ is modelled as a Kronecker product IGMRF; following Knorr-Held⁸ we
 287 consider interaction types I, II, III, and IV as described in Table 1.

288 Model (4) includes the same vectors of random effects as model (1), but
 289 in contrast to model (1), we now have very intuitive hyperparameters: τ is
 290 the total precision, i.e. τ^{-1} is the total generalized variance, and γ and ϕ
 291 are two interpretable mixing parameters. The value of γ can be interpreted
 292 as the proportion of total variance explained by the interaction δ . The
 293 variance explained by the main effects is therefore given by $\tau^{-1}(1-\gamma)$:
 294 $1-\phi$ quantifies the proportion of such variance which can be attributed
 295 to the temporal random effects β_1 , with ϕ being the proportion attributed
 296 to the spatial random effects β_2 .

297 We need to assign priors to the overall precision parameter τ and the
 298 mixing parameters γ and ϕ . In the next section we focus on the prior for
 299 γ , and leave prior choice for the remaining parameters to Section 4.

300 3.1 A Penalized Complexity prior for γ

301 Our choice of a PC prior for γ follows naturally from the model
 302 reparametrization in Eq. (4) and provides a way of eliciting the prior in a
 303 very intuitive way. Furthermore, it avoids overfitting by construction hence
 304 guaranteeing a parsimonious model. Our PC prior for γ (see Result 1

below) is based on the assumption that the interaction model in (4) shrinks to the main effects model ($\beta_1 + \beta_2$).

Result 1. Let us assume a model of the form (4), for all types of interaction in Table 1:

1. The distance from the base model is

$$d(\gamma) \simeq \sqrt{\gamma}, \quad 0 < \gamma < 1$$

2. The PC prior for γ with base model $\gamma = 0$ is

$$\pi(\gamma) = \frac{\theta \exp(-\theta\sqrt{\gamma})}{2\sqrt{\gamma}(1 - \exp(-\theta))} \quad 0 < \gamma < 1, \theta > 0. \quad (5)$$

The proof can be found in Supplemental material 6.1-6.3.

The scaling of the PC prior for γ , i.e. the choice of θ in Eq. (5), is done by defining the probability of a tail event on γ . The parameter θ controls the strength of penalisation for deviating from the base model; the higher the θ the greater the penalty. We suggest setting U and a such that $\mathbb{P}(\gamma < U) = a$; this way θ is obtained by numerically solving:

$$\frac{1 - \exp(-\theta\sqrt{U})}{1 - \exp(-\theta)} = a, \quad a > \sqrt{U}.$$

Note that it is not possible to assign equal weight to the main and interaction terms in the model, i.e. $U = a = 0.5$ because of the constraint $a > \sqrt{U}$. However, we can always encode a fair amount of uncertainty into the prior by choosing a close to 1 and large values of U . In the left panel of Figure 1, θ is obtained using $a = 0.99$ and three different values for U . A large U allows for more flexibility as the corresponding density curve decreases steadily towards zero as γ increases, while for a small value of U the density curve drops towards zero quite sharply, strongly penalizing any deviation from the base model. For comparison,

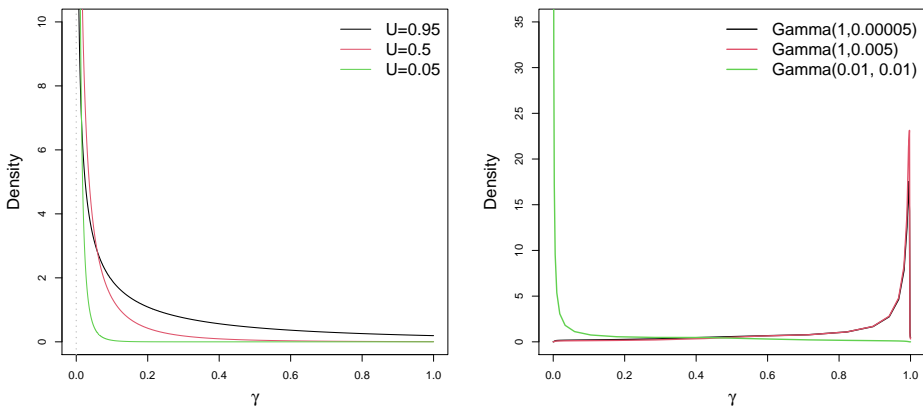


Figure 1. Left panel: PC prior $\pi(\gamma)$ using $a = 0.99$ and three different values for U . Right panel: implied prior on γ when a Gamma prior is used on all three precision parameters $\tau_1, \tau_2, \tau_{12}$.

325 the right panel in Figure 1 shows the prior on γ that corresponds to using
 326 a Gamma prior on all three precision parameters in model (1) for three
 327 different parameter choices. The figure illustrates how the resulting prior
 328 on γ depends strongly on the chosen values for the Gamma parameters,
 329 going from one extreme to the other in terms of prior weight on the base
 330 model.

331 Results from a simulation study reported in Supplemental material 7
 332 indicate that the posterior mean estimates of γ are reasonably close to the
 333 true value under different scenarios. We have observed stable results for
 334 several choices of U , unless one defines on purpose an unflexible prior,
 335 where most of the probability mass is placed near the base model (e.g.
 336 when adopting $a = 0.99$ and a small $U = 0.05$). Results are comparable
 337 to those obtained using a Uniform prior on γ unless there is no interaction
 338 (i.e. $\gamma = 0$), in which case the uniform leads to greater bias when the
 339 population at risk is small.

4 Examples

As introduced in Section 2, model (2) is more common in practice and indeed it is the model adopted in this section for both case-studies. In the case of structured and unstructured main effects, another set of parameters ψ_1 and ψ_2 can be included to further distribute the variance, so that model (4) becomes:

$$\eta_{ij} = \alpha + \sqrt{\tau^{-1}} \left(\sqrt{1 - \gamma} \left(\sqrt{1 - \phi} \left(\sqrt{1 - \psi_1} \beta_{1_i} + \sqrt{\psi_1} \epsilon_{1_i} \right) + \sqrt{\phi} \left(\sqrt{1 - \psi_2} \beta_{2_j} + \sqrt{\psi_2} \epsilon_{2_j} \right) \right) + \sqrt{\gamma} \delta_{ij} \right), \quad (6)$$

where $\tau > 0$, $0 < \gamma < 1$, $0 < \phi < 1$, $0 < \psi_1 < 1$, $0 < \psi_2 < 1$, $\epsilon_1 \sim N(\mathbf{0}, \mathbf{I}_{n_1})$, $\epsilon_2 \sim N(\mathbf{0}, \mathbf{I}_{n_2})$ and β_1 , β_2 and δ as in model (4). Result 1 about the PC for γ still holds; see Supplemental material 6.3.

Note that the parameters in model (6) are identifiable as the model is just a reparametrized version of the classic space-time interaction model (2), where each random effect has its corresponding precision parameter. The number of parameters is exactly the same; in fact, it can be shown that there is a one-to-one mapping between the parameters of both versions of the model. As in model (1), appropriate constraints need to be imposed to ensure identifiability of the terms in (6). The constraints on the interaction term are summarized on Table 1, while on the temporal and spatial structured main effects it is enough to impose a sum to zero constraint.

In the next two examples, we use the PC prior in Eq. (3) for τ and the PC prior in Eq. (5) for γ . Regarding ϕ , ψ_1 and ψ_2 , we simply choose a uniform on (0,1) as a prior for each of them, but other choices are possible. In fact, a PC prior could also be used for ϕ following the work by Fuglstad et al.²², who also consider the use of a Dirichlet prior where the base model attributes equal weights to each component, thus expressing ignorance about how the variance is distributed. Similarly, one could use a PC prior

360 on each ψ_1 and ψ_2 as in Riebler et al.²¹, considering as base model the
 361 absence of structured effects.

362 All the VP models presented in the next two examples were run using
 363 R-INLA³², see code in Supplemental material 9.

364 4.1 Ohio lung cancer

365 We illustrate our model using the Ohio lung cancer data^{6,8,9} which
 366 is available at [http://www.biostat.umn.edu/~brad/data2.](http://www.biostat.umn.edu/~brad/data2.html)
 367 [html](http://www.biostat.umn.edu/~brad/data2.html). These data report yearly counts of lung cancer deaths for white
 368 males from 1968 to 1988, in the 88 counties of Ohio. Figure 2 left panel
 369 displays the time series of mortality rate for all counties. Our aim is not to
 370 find the best model for this data, but to show what our approach can add
 371 in terms of interpretability of the results compared to a classical analysis
 372 as performed in Knorr-Held⁸.

373 **4.1.1 Model** Let y_{ij} be the number of deaths at time $i = 1, \dots, 21$
 374 in county $j = 1, \dots, 88$ and pop_j be the population at risk in county
 375 j , we consider the model proposed in Knorr-Held⁸ assuming structured
 376 and unstructured effects for both space and time main effects, plus a
 377 space-time interaction term. The classical parameterization in Knorr-
 378 Held⁸ follows,

$$\begin{aligned}
 y_{ij} &\sim \text{Bin}(\text{pop}_j, \exp(\eta_{ij}) / \exp(1 + \eta_{ij})), \\
 \eta_{ij} &= \alpha + \underbrace{\beta_{1_i} + \epsilon_{1_i} + \beta_{2_j} + \epsilon_{2_j}}_{\text{main}} + \underbrace{\delta_{ij}}_{\text{int}},
 \end{aligned}
 \tag{7}$$

379 where the main effects are modelled as:

$$\begin{aligned}
 \epsilon_1 &\sim N(\mathbf{0}, \tau_{\epsilon_1}^{-1} \mathbf{I}_{n_1}); & \epsilon_2 &\sim N(\mathbf{0}, \tau_{\epsilon_2}^{-1} \mathbf{I}_{n_2}); \\
 \beta_1 &\sim N(\mathbf{0}, \tau_1^{-1} \mathbf{R}_1^-); & \beta_2 &\sim N(\mathbf{0}, \tau_2^{-1} \mathbf{R}_2^-).
 \end{aligned}$$

380 where \mathbf{R}_1 and \mathbf{R}_2 are the unscaled structure matrices of a RW1 (for
 381 time) and an ICAR (for space). The space-time interaction is modelled
 382 by a Kronecker product IGMRF built on the precision matrices of the
 383 structured components β_1 and β_2 . All the interaction types in Table 1 are
 384 considered in the following analysis. This model would require priors for
 385 the precision hyperparameters $\tau_{\epsilon_1}, \tau_{\epsilon_2}, \tau_1, \tau_2, \tau_{12}$.

386 Instead of working with the above model, we assume the VP model in
 387 Eq. (6), with scaled structure matrices, with the priors for $\tau, \gamma, \phi, \psi_1, \psi_2$
 388 stated in Section 4. For τ 's PC prior we set $a = 0.01$ and $U = 1/0.31$
 389 following the rule of thumb described in Section 2.2. The PC prior for γ
 390 is scaled by imposing $U = 0.5, a = 0.99$; results (not shown here) were
 391 stable for varying $U = \{0.05, 0.5, 0.95\}$.

392 **4.1.2 Results** Table 2 reports various model selection criteria for the
 393 VP model, for interaction types I, II, III and IV, namely DIC³³, WAIC³⁴,
 394 leave-one-out log score (LOOLS), computed as $-\sum_{i=1}^n \log \pi(y_i|y_{-i})$, and
 395 the log-marginal likelihood (logMLIK), $\pi(y|\mathcal{M})$, which quantifies the
 396 likelihood of the data y under a given model \mathcal{M} . PC priors enhance
 397 the marginal likelihood as a simple and effective tool for *fair* model
 398 comparison, when the compared models have similar structure and only
 399 differ on a particular component^{23,35}. Assume \mathcal{M}_1 and \mathcal{M}_2 are the
 400 interaction type I and II models, respectively: these models are the same
 401 except for a different type of interaction. The Bayes factor³⁶ is defined as

$$K = \frac{\pi(y|\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)} = \frac{\pi(\mathcal{M}_1|y) \pi(\mathcal{M}_2)}{\pi(\mathcal{M}_2|y) \pi(\mathcal{M}_1)}. \quad (8)$$

402 The scale parameter θ of the PC prior for γ , which controls the decay
 403 rate from the base model (the model with no interaction), has to be
 404 chosen for both \mathcal{M}_1 and \mathcal{M}_2 . We can handle this choice conveniently
 405 by setting the same θ for \mathcal{M}_1 and \mathcal{M}_2 , which implies that $\pi(\mathcal{M}_2)/\pi(\mathcal{M}_1)$
 406 in Eq.(8) cancels out and the Bayes factor turns out to be the ratio of

407 the posterior odds. For all the four interaction models in Table 2 we
 408 follow this strategy and set the same decay rate for the PC prior on γ .
 409 The advantage is that the *a priori* contribution of the interaction to the
 410 total (generalized) variance is the same, no matter what interaction type is
 411 assumed; this is a desirable feature when having to choose among different
 412 types of interaction models the one that best fits the data. Therefore, we
 413 suggest comparing the logMLIK values for model choice purposes here;
 414 furthermore, DIC is known to favour complexity in models with many
 415 random effects³⁷.

416 From Table 2 we see that DIC, WAIC and LOOLS point to type II
 417 (followed by type IV) as the best model; a similar conclusion based on
 418 DIC was found in Knorr-Held⁸. Interestingly, logMLIK is largest for type
 419 I which indicates that the model with main effects plus an individual-
 420 level random effects capturing unstructured variation may be a better
 421 description of the Ohio data.

Table 2. Model comparison criteria (computed using R-INLA) for the VP model, under the four interaction types.

interaction type	logMLIK	DIC (deviance; p_D)	WAIC	LOOLS
I	-5623.52	10945.23 (10722.68; 222.55)	10957.75	5489.75
II	-6759.84	10916.00 (10739.09; 176.91)	10931.3	5469.43
III	-6098.89	10957.86 (10792.68; 165.18)	10980.99	5496.28
IV	-7200.13	10919.23 (10755.06; 164.17)	10934.82	5470.89

422 In order to show now the gain of using our approach compared to a
 423 classical analysis we start by discussing some plots obtained for type I
 424 interaction about the main effects. The top right panel in Figure 2 displays
 425 the estimated main temporal effect, in the scale of the linear predictor,
 426 decomposed into its structured and unstructured (iid) components. The
 427 unstructured effects looks very flat compared to the structured ones which
 428 is probably responsible for most of the temporal variation in the relative
 429 risk. The relative risk increases roughly linearly in time, with a less
 430 steep increase towards the end of the time window. The bottom panels

431 in Figure 2 display the estimated structured (left) and iid (right) spatial
432 effects in the scale of the linear predictor. Here the unstructured effect
433 shows larger variability than the structured one which shows a very
434 smooth spatial gradient from north-west to south-east. A visual inspection
435 of this sort, also possible when the classical model is used, gives useful
436 insights into the spatial and temporal patterns in the data. However, it does
437 not allow proper quantification of the variance attributable to the various
438 sources (main, interaction, spatial and temporal effects, etc), while this
439 quantification is readily available from our VP model.

440 Table 3 reports the mean (with 2.5 and 97.5 quantiles between brackets)
441 of the posterior distribution of the mixing parameters $\gamma, \phi, \psi_1, \psi_2$, from
442 which we can understand quantitatively the contribution of the main
443 sources of variation. In particular, rows 1 and 2 report the total variance
444 partitioned into interaction versus main effects; rows 3 and 4 quantify how
445 the variance attributable to the main effects is partitioned into space and
446 time; rows 5 to 6 and 7 to 8 give the variance partitioning for the structured
447 versus iid for space and time, respectively. The main findings about the
448 variation of the spatio-temporal mortality risk pattern are as follows. First,
449 the estimated contribution of the interaction is about 4.8%, which means
450 that the main effects are responsible for most of the variability in mortality
451 risk with the interaction playing a minor role in describing this data. This
452 is reasonable for non-infectious diseases such a cancer⁷. Second, space
453 is responsible for about 87.5% of the variation in risk explained by the
454 main effects, which is hard to grasp from only looking at the plot of the
455 main temporal and spatial effects in Figure 2. This result highlights the
456 fact that lung cancer in Ohio had, in the period of time considered, larger
457 variability over space than time which could be informative for policy
458 makers and epidemiologists and may contribute to generate hypothesis on
459 the role played by possible environmental risk factors in the region. Third,
460 within the main spatial and temporal effects, the structured component is

461 predominant for time, while the iid component is predominant for space.
 462 However, these estimated contributions, and in particular the latter, are
 463 affected by greater uncertainty than the previous estimates, indicating that
 464 the data are less informative about the posterior for ψ_1 and ψ_2 than they
 465 are for γ and ϕ . These findings are stable across the different types of
 466 interactions (see Supplemental material 8).

Table 3. Variance partitioning table for Ohio lung cancer, type I interaction. The column named *contribution* reports the posterior mean of the hyper-parameters displayed in the column named *estimator*, with 0.025 and 0.975 posterior quantiles between brackets. All values are in a $(0, 1)$ interval and indicate the proportional contribution of the model component *level 2* to the variance explained by the model component *level 1*.

Model component		Variance Partitioning	
level 1	level 2	estimator	contribution
main+int	main	$1 - \hat{\gamma}$	0.952 (0.913, 0.979)
	int	$\hat{\gamma}$	0.048 (0.021, 0.087)
main	space	$\hat{\phi}$	0.875 (0.765, 0.946)
	time	$1 - \hat{\phi}$	0.125 (0.054, 0.235)
time	iid	$\hat{\psi}_1$	0.069 (0.010, 0.229)
	str	$1 - \hat{\psi}_1$	0.931 (0.771, 0.990)
space	iid	$\hat{\psi}_2$	0.658 (0.273, 0.925)
	str	$1 - \hat{\psi}_2$	0.342 (0.075, 0.727)

467 4.2 Covid-19 in Italy

468 We use the VP model to study Covid-19 incidence variations across space
 469 and time in Italy. Data cover all of the 107 Italian provinces and span a
 470 period of time that goes from the onset of the pandemic on 24th February
 471 2020 to late July 2021 for a total of 70 weeks; the full dataset is made
 472 available by the Italian National Institute of Health through the website
 473 <https://github.com/pcm-dpc/COVID-19>. Data are originally
 474 available on a daily basis, but we aggregate them by week to smooth out
 475 artefactual patterns mainly due to delays in reporting new cases. The final
 476 dataset consists of weekly counts of new Covid-19 cases y_{ij} , for week
 477 $i = 1, \dots, 70$ and province $j = 1, \dots, 107$, and the population at risk for each
 478 province pop_j .

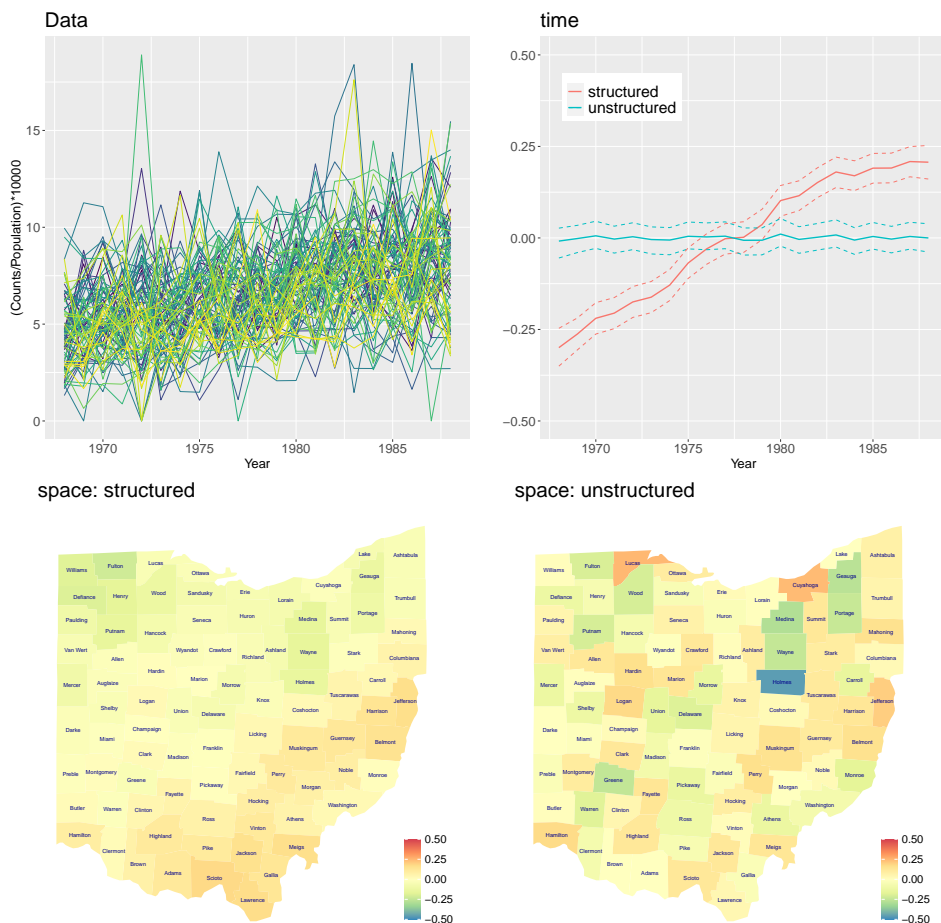


Figure 2. Top left panel: time series of lung cancer (white males) disease rates per 10000 population at risk, for the 88 counties in the Ohio dataset. Top right panel: temporally structured and temporally unstructured components for type I interaction model, in the scale of the linear predictor. Bottom left and right panels show, respectively, the spatially structured and unstructured components for type I interaction model, in the scale of the linear predictor.

479 Our goal is to analyze the sources of variation in Covid-19 incidence
 480 rates in a scale between 0 and 1, which is easy to interpret and visualize
 481 and provides a clear idea of the contribution of each source. We follow the
 482 ideas in Picado et al.³⁸ in considering the interaction term as a measure

of local heterogeneity, which can be seen as an indirect measure of how effective the control measures are. Hence a primary interest is to quantify the contribution of the interaction to the total variability, i.e. the posterior estimate for γ . Our second interest is to investigate changes in the estimated local heterogeneity across geographical macro-regions and time windows. We run two analysis: in the first one we fit the VP to the full dataset, in the second one we run the same VP model to separate subsets of the data which are constructed using combinations of geographical area, with levels north (N), centre (C) and south (S), and pandemic wave, with levels W1 and W2. The first wave (W1) covers the first 18 weeks and roughly indicates the national lock down period, while the second wave (W2) covers the rest of the time frame and indicates the period where restriction measures were set at a regional level. Data are displayed in Figure 3.

4.2.1 Model We consider the binomial model in Eq. (7), where structured and unstructured random effects are specified for both space and time as main effects. We model the temporally structured effects as a RW1 (as we do not anticipate smoothness) and the spatially structured effects as an ICAR, and assume a type IV space-time interaction to capture potential complex space-time patterns which are not explained by the main space and time components. In this particular example, the spatial main effect may reflect differences on the public health policy strategies adopted in each area (for example different testing rates across provinces). Again, we avoid the classic parametrization and take advantage of the VP approach described in (6). Doing so, we can elicit the prior easily and describe the various sources of variability in the data in an intuitive way in terms of the mixing parameters $\gamma, \phi, \psi_1, \psi_2$.

Available information on the nature of the disease can be used to aid in parameter choice for the PC priors on γ and τ . Since we know that we are dealing with a contagious disease which evolves over time possibly in a

513 different manner across provinces we anticipate a relevant contribution of
514 the interaction term. Thus we choose θ in Eq. (5) by setting $U = 0.95$, $a =$
515 0.99 , which implies a large probability that $\gamma < 0.95$. In choosing the scale
516 parameter λ of the PC prior for τ in Eq. (3) we consider the scale of the
517 logit transformed incidence rates and use the rule of thumb described in
518 Simpson et al.²⁰, imposing a marginal standard deviation equal to 2 for
519 the incidence rates in the linear predictor (logit) scale.

520 **4.2.2 Results** The left panel in Figure 4 reports the variance
521 partitioning plot for the full Covid-19 dataset; this plot is just a graphical
522 version of the variance partitioning table that was presented in Table 3 for
523 the Ohio lung cancer data. This plot resembles the graphs in Gelman³⁹
524 that summarize anova results in terms of estimated standard deviation
525 for each bunch of random effects in the model. Our variance partitioning
526 plot follows the same idea but represents the contribution of each source
527 in a scale $(0, 1)$. The main effects account for the greatest proportion of
528 the total variation. Within the main effects, the variability in incidence
529 rates is mostly driven by the spatial component, in particular by the
530 unstructured part of it (although the corresponding posterior estimates are
531 highly uncertain), while for the temporal part is the structured component
532 that explains most of the variability.

533 The middle and right panels in Figure 4 report the variance partitioning
534 plot for the models fitted to different subsets of the full dataset to
535 investigate whether the spatio-temporal pattern in Covid-19 cases is
536 consistent or not across geographical areas (N, C, S) and pandemic waves
537 (W1, W2). It is interesting to see that the impact of the interaction term
538 is greater in the second wave than in the first one for all three areas,
539 suggesting greater local heterogeneity during the second wave. This could
540 reflect the fact that restriction measures went from being national in
541 the first wave to being regional in the second one, so we expect greater
542 heterogeneity over space during the latter. Within the first wave, the main

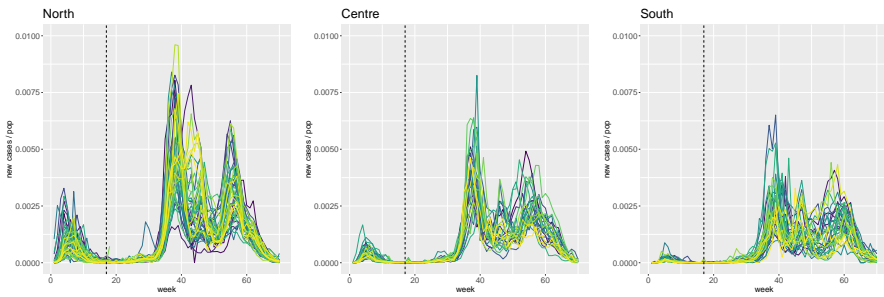


Figure 3. Weekly Covid-19 incidence rates in the North (left panel), Centre (central panel) and South (left panel) of Italy. The vertical dashed line marks the separation between the first (W1) and second (W2) wave.

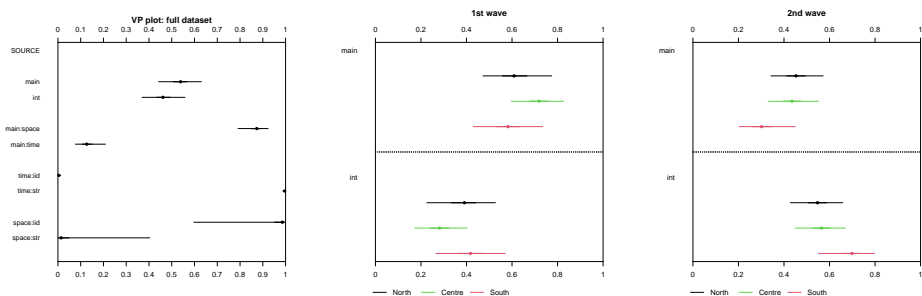


Figure 4. Variance partitioning plot for Covid-19 full dataset (left panel), first wave (middle panel) and second wave (right panel). The middle and right panels allow comparison across northern (black), central (green) and southern (red) areas in Italy.

543 effects are responsible for a greater proportion of variation in all three
 544 areas, but that attributable to the interaction is slightly greater in the South,
 545 followed by the North and then the Centre.

546 5 Discussion

547 In this paper, we revisit spatio-temporal disease mapping, with particular
 548 attention to the interaction models discussed in Knorr-Held⁸, and propose
 549 a new model that allows variance partitioning among the main effects and
 550 the space-time interaction. When defining priors on the hyperparameters
 551 that control complexity of each intrinsic GMRF component, it is important

552 to bear in mind that the main effects belong to the null space of the
553 interaction term. This means that the interaction can naturally be regarded
554 as an extension of the model including the main effects alone. This
555 idea leads to a model reparametrization where a mixing parameter γ
556 balances out the contribution of the main and interaction effects to the
557 total variance. The proposed approach implicitly defines a joint prior on
558 the precision parameters of the various terms in the classic parametrization
559 of the model.

560 The advantages of this reparametrization are twofold; on the one hand,
561 prior choice can be made in an intuitive manner using a PC prior, avoiding
562 the issue of eliciting priors on hard-to-interpret precision parameters. In
563 space-time disease mapping, the nature of the disease can provide useful
564 information to elicit the prior; for example, for non-infectious diseases
565 such as the one considered in the first case study most of the variation
566 is expected to be explained by the main effects⁷. This knowledge can be
567 easily passed onto the PC prior for the mixing parameter γ , while coding
568 this information into a precision parameter in the classic parametrization
569 would be far from easy. On the other hand, the posterior for γ becomes
570 a useful tool to investigate variations in disease risk on a very practical
571 scale and can provide useful insights into epidemiological interpretations.
572 We have illustrated the use of the VP model in two examples; the variance
573 partitioning tables and plots summarize the contribution of the different
574 sources of variation in terms of proportion of explained (generalized)
575 variance.

576 In a broader perspective, our work falls within the framework of
577 variance distributing models as introduced by Fuglstad et al.²², and
578 adds to the literature in considering intrinsic GMRF models. The
579 variance partitioning approach proposed here may be adopted in all
580 those applications where intrinsic GMRFs are meant as tools to perform

581 smoothing in more than one dimension; for instance in the analysis of grid-
582 data such as those arising from agricultural field trials or spatio-temporal
583 data from environmental studies and ecological surveys.

584 References

- 585 1. PJ Diggle, AG Chetwynd, R. Hggkvist, and SE Morris. Second-order analysis of space-
586 time clustering. *Statistical Methods in Medical Research*, 4(2):124–136, 1995. PMID:
587 7582201.
- 588 2. M.A. Martinez-Beneito and P. Botella-Rocamora. *Disease Mapping: From Foundations*
589 *to Multidimensional Modeling (1st ed.)*. Chapman and Hall/CRC, 2019.
- 590 3. A.B. Lawson. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial*
591 *Epidemiology (3rd ed.)*. Chapman and Hall/CRC, 2018.
- 592 4. Ying C. MacNab. Bayesian disease mapping: Past, present, and future. *Spatial Statistics*,
593 <https://doi.org/10.1016/j.spasta.2022.100593>
- 594 5. Jon Wakefield. Disease mapping and spatial regression with count data. *Biostatistics*,
595 8(2):158–183, 2007.
- 596 6. Leonhard Knorr-Held and Julian Besag. Modelling risk from a disease in time and space.
597 *Statistics in Medicine*, 17(18):2045–2060, 1998.
- 598 7. J.J. Abellan, S. Richardson, and N. Best. Use of spacetime models to investigate the
599 stability of patterns of disease. *Environmental Health Perspectives*, 116(8):1111–1119,
600 2008.
- 601 8. Leonhard Knorr-Held. Bayesian modelling of inseparable space-time variation in disease
602 risk. *Statistics in Medicine*, 19(1718):2555–2567, 2000.
- 603 9. Lance A. Waller, Bradley P. Carlin, Hong Xia, and Alan E. Gelfand. Hierarchical spatio-
604 temporal mapping of disease rates. *Journal of the American Statistical Association*,
605 92(438):607–617, 1997.
- 606 10. L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, Ghislandi M., and M. Songini.
607 Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14, 1995.
- 608 11. J. Aldstadt. An incremental Knox test for the determination of the serial interval between
609 successive cases of an infectious disease. *Stochastic Environmental Research and Risk*
610 *Assessment*, 21:487–500, 2007.
- 611 12. C. Robertson, T.A. Nelson, Y.C. MacNab, and A.B. Lawson. Review of methods for
612 space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1:105–116,
613 2010.

- 614 13. F. Lindgren and H. Rue. On the second-order random walk model for irregular locations.
615 *Scandinavian Journal of Statistics*, 35(4):691–700, 2008.
- 616 14. J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion).
617 *Journal of the Royal Statistical Society Series B*, 36(2):192–225, 1974.
- 618 15. H. Rue and L. Held. *Gaussian Markov Random Fields*. Chapman and Hall/CRC, 2005.
- 619 16. S.H. Sørbye and H. Rue. Scaling intrinsic Gaussian Markov random field priors in spatial
620 modelling. *Spatial Statistics*, 8:39–51, 2014.
- 621 17. A Freni-Sterrantino, M Ventrucci, and H Rue. A note on intrinsic conditional
622 autoregressive models for disconnected graphs. *Spatial and Spatio-Temporal
623 Epidemiology*, 26:25–34, 2018.
- 624 18. Andrew Gelman. Prior distributions for variance parameters in hierarchical models
625 (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534, 09 2006.
- 626 19. Youyi Fong, Hvard Rue, and Jon Wakefield. Bayesian inference for generalized linear
627 mixed models. *Biostatistics*, 11(3):397–412, 2010.
- 628 20. Daniel Simpson, Hvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye.
629 Penalising model component complexity: A principled, practical approach to constructing
630 priors. *Statistical Science*, 32(1):1–28, 02 2017.
- 631 21. Andrea Riebler, Sigrunn H Sørbye, Daniel Simpson, and Hvard Rue. An intuitive Bayesian
632 spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical
633 Research*, 25(4):1145–1165, 2016. PMID: 27566770.
- 634 22. G.-A. Fuglstad, I. G. Hem, A. Knight, H. Rue, and A. Riebler. Intuitive joint priors for
635 variance parameters. *Bayesian Analysis*, 15(4):1109–1137, 2020.
- 636 23. M. Ventrucci, D. Cocchi, G. Burgazzi, and A. Laini. Pc priors for residual correlation
637 parameters in one-factor mixed models. *Statistical Methods & Applications*, 29(4):745–
638 765, 2020.
- 639 24. J. Besag, J. York, and A. Mollie. Bayesian image restoration, with two applications in
640 spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–21, 1991.
- 641 25. Goicoa, T., Adin, A., Ugarte, M.D., and Hodges, J.S. (2018). In spatio-temporal disease
642 mapping models, identifiability constraints affect PQL and INLA results. *Stochastic
643 Environmental Research and Risk Assessment*, 32: 749–770.
- 644 26. L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. *Regression: models, methods and
645 applications*. Springer-Verlag, Berlin, 2013.
- 646 27. G. Wahba. Improper priors, spline smoothing and the problem of guarding against model
647 errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*,
648 40(3):364–372, 1978.

-
- 649 28. Sylvia Frühwirth-Schnatter and Helga Wagner. Stochastic model specification search
650 for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*,
651 154(1):85–100, 2010.
- 652 29. Sylvia Frühwirth-Schnatter and Helga Wagner. Bayesian variable selection for random
653 intercept modeling of Gaussian and non-Gaussian data. In *J. M. Bernardo, M. J. Bayarri,*
654 *J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (Eds.)*, pages 165–
655 200. Bayesian Statistics 9, Oxford, 2011.
- 656 30. Massimo Ventrucchi and Hvard Rue. Penalized complexity priors for degrees of freedom
657 in bayesian p-splines. *Statistical Modelling*, 16(6):429–453, 2016.
- 658 31. S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of*
659 *Mathematical Statistics*, 22:79–86, 1951.
- 660 32. H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian
661 models using inte- grated nested Laplace approximations (with discussion). *Journal of the*
662 *Royal Statistical Society, Series B*, 71(2):319–392, 2009.
- 663 33. David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde.
664 Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal*
665 *Statistical Society, Series B*, 64:583–639(57), 2002.
- 666 34. Sumio Watanabe. A Widely Applicable Bayesian Information Criterion. *Journal of*
667 *Machine Learning Research*, 14:867–897, 2013.
- 668 35. S.H. Sørbye and H. Rue. Fractional gaussian noise: Prior specification and model
669 comparison. *Environmetrics*, 29(5-6):e2457, 2018.
- 670 36. Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical*
671 *Association*, 90(430):773–795, 1995.
- 672 37. A. Riebler and L. Held. The analysis of heterogeneous time trends in multivariate age-
673 period-cohort models. *Biostatistics*, 11(1):57–69, 2010.
- 674 38. Albert Picado, Javier Guitian, and Dirk Pfeiffer. Space-time interaction as an indicator of
675 local spread during the 2001 FMD outbreak in the UK. *Preventive Veterinary Medicine*,
676 79:3–19, 05 2007.
- 677 39. Andrew Gelman. Analysis of variance-why it is more important than ever. *The Annals of*
678 *Statistics*, 33(1):1 – 53, 2005.

679 Supplemental material

6 Proofs

For ease of presentation, we first prove **Result 1** for model (4) type IV interaction in Appendix 6.1 and then show that it is also valid for types I, II and III in Appendix 6.2. Details regarding the proof for the model including unstructured and structured main effects (model (6)) can be found in Appendix 6.3. Throughout the proof, we assume a RW2 model on the temporal random effect and an ICAR on the spatial one. The modification of the proof when a RW1 model is used on the temporal effect is straightforward.

6.1 Proof of Result 1 for type IV interaction

Model (1) can be written in general form (in the linear predictor scale) as

$$\boldsymbol{\eta} = \alpha \mathbf{1}_n + \sqrt{\tau^{-1}} \left(\sqrt{1-\gamma} \boldsymbol{\omega}_0 + \sqrt{\gamma} \boldsymbol{\omega}_1 \right), \quad (9)$$

where $\tau > 0$ is the precision parameter, $0 < \gamma < 1$ is the mixing parameter, $\boldsymbol{\omega}_0, \boldsymbol{\omega}_1$ are n -dimensional IGMRFs with precision matrices \mathbf{Q}_0 and \mathbf{Q}_1 respectively, with

$$\mathbf{Q}_0^- = (1-\phi)(\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1}) \tilde{\mathbf{R}}_1^- (\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1})^T + \phi(\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1}) \tilde{\mathbf{R}}_2^- (\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1})^T$$

and

$$\mathbf{Q}_1 = \tilde{\mathbf{R}}_2 \otimes \tilde{\mathbf{R}}_1$$

where $\tilde{\mathbf{R}}_1$ and $\tilde{\mathbf{R}}_2$ are the scaled structure matrices of a RW2 and an ICAR, respectively. Note that $\text{rank}(\tilde{\mathbf{R}}_1) = n_1 - 1$ and $\text{rank}(\tilde{\mathbf{R}}_2) = n_2 - 2$, so it follows that $\text{rank}(\mathbf{Q}_1) = n_1 n_2 - n_2 - 2n_1 + 2$ and $\text{rank}(\mathbf{Q}_0) = n_1 + n_2 - 3$. For ease of presentation, we simplify the notation and denote $n = n_1 n_2$, $r = 2n_1 + n_2 - 2$, so that $\text{rank}(\mathbf{Q}_1) = n - r$. It is immediate to see that rank of \mathbf{Q}_0 is smaller than the rank deficiency of \mathbf{Q}_1 , i.e.:

$$n_1 + n_2 - 3 \leq 2n_1 + n_2 - 2 \Leftrightarrow n_1 \leq 2n_1 + 1,$$

so that $\text{rank}(\mathbf{Q}_0) = r - l$, where $l \geq 0$ is the difference between $\text{rank}(\mathbf{Q}_0)$ and r . For ease of presentation, we can assume $l = 0$ (note that if $l \neq 0$ then the adjustment of the proof is straightforward).

Consider $\tau = 1$ without loss of generality. To derive the PC prior for γ we will study the limiting behaviour of $\text{KLD}(\pi_1 || \pi_0)$ for $\gamma = \gamma_0 \rightarrow 0$ under the base model. The distributions π_1 and π_0 are

703 defined as follows:

$$\begin{aligned}\pi_1 &\sim N_1(0, \Sigma_1) \quad \text{with} \quad \Sigma_1 = (1 - \gamma)\mathbf{Q}_0^- + \gamma\mathbf{Q}_1^- \\ \pi_0 &\sim N_0(0, \Sigma_0) \quad \text{with} \quad \Sigma_0 = (1 - \gamma_0)\mathbf{Q}_0^- + \gamma_0\mathbf{Q}_1^-\end{aligned}$$

704 The KLD is given by:

$$\text{KLD}(\pi_1 || \pi_0) = \frac{1}{2} \left(\text{trace}(\Sigma_0^- \Sigma_1) - (n - r) - \log \frac{|\Sigma_1|}{|\Sigma_0|} \right). \quad (10)$$

705 Expression (10) can be computed easily if we consider the eigendecomposition of the matrices
706 $\mathbf{Q}_0 = \mathbf{V}_{Q_0} \Lambda_{Q_0} \mathbf{V}_{Q_0}^T$ and $\mathbf{Q}_1 = \mathbf{V}_{Q_1} \Lambda_{Q_1} \mathbf{V}_{Q_1}^T$, with

$$\Lambda_{Q_0} = \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_r, \underbrace{0, \dots, 0}_{n-r}) \quad ; \quad \Lambda_{Q_1} = \text{diag}(\underbrace{0, \dots, 0}_r, \lambda'_{r+1}, \dots, \lambda'_n), \quad (11)$$

$$707 \quad \mathbf{V}_{Q_0} = [e_1, e_2, \dots, e_r, e_{r+1}, \dots, e_n] \quad ; \quad \mathbf{V}_{Q_1} = [\hat{e}_1, \dots, \hat{e}_r, \hat{e}_{r+1}, \dots, \hat{e}_n]. \quad (12)$$

708 where Λ_{Q_0} , Λ_{Q_1} represent the diagonal matrix of eigenvalues and \mathbf{V}_{Q_0} and \mathbf{V}_{Q_1} the matrices
709 whose columns are the associated eigenvectors. A common eigenvector basis \mathbf{V} can be formed as

$$\mathbf{V} = [e_1, e_2, \dots, e_r, \hat{e}_{r+1}, \dots, \hat{e}_n],$$

710 so that $\mathbf{Q}_0 = \mathbf{V} \Lambda_{Q_0} \mathbf{V}^T$ and $\mathbf{Q}_1 = \mathbf{V} \Lambda_{Q_1} \mathbf{V}^T$. If $l \neq 0$ then there would be a set of eigenvectors
711 that are associated to zero eigenvalues in both matrices \mathbf{Q}_0 and \mathbf{Q}_1 contemporarily, so the
712 common basis can still be formed.

713

714 Matrices Σ_0^- and Σ_1 can be re-expressed as

$$\Sigma_0^- = \left\{ \mathbf{V} [(1 - \gamma_0)\Lambda_{Q_0}^{-1} + \gamma_0\Lambda_{Q_1}^{-1}] \mathbf{V}^T \right\}^{-1} = \mathbf{V} [(1 - \gamma_0)\Lambda_{Q_0}^{-1} + \gamma_0\Lambda_{Q_1}^{-1}]^{-1} \mathbf{V}^T$$

715 and

$$\Sigma_1 = \mathbf{V} ((1 - \gamma)\Lambda_{Q_0}^{-1} + \gamma\Lambda_{Q_1}^{-1}) \mathbf{V}^T,$$

716 where $\Lambda_{Q_0}^{-1}$ and $\Lambda_{Q_1}^{-1}$ are diagonal matrices with elements λ_i and $\hat{\lambda}_i$. Note that \mathbf{Q}_0 and \mathbf{Q}_1 are
717 singular; following²⁰ appendix A2, $\lambda_i = 1/\tilde{\lambda}_i$ if $\tilde{\lambda}_i > 0$ and $\lambda_i = 0$ when $\tilde{\lambda}_i = 0$. Analogously,
718 $\hat{\lambda}_i = 1/\lambda'_i$ if $\lambda'_i > 0$ and $\hat{\lambda}_i = 0$ when $\lambda'_i = 0$.

719

720 First, we compute $\text{trace}(\Sigma_0^{-1}\Sigma_1)$, for which we need the diagonal $\text{diag}(\Sigma_0^{-1}\Sigma_1)$. Let us define

$$D(\gamma) = \text{diag} \left((1 - \gamma)\lambda_i + \gamma\hat{\lambda}_i \right)_{i=1, \dots, n}$$

721 we can re-express the diagonal as

$$\text{diag}(\Sigma_0^{-1}\Sigma_1) = \mathbf{V}D(\gamma_0)^{-1}D(\gamma)\mathbf{V}^T.$$

722 The trace simplifies to

$$\begin{aligned} \text{tr}(\mathbf{V}D(\gamma_0)^{-1}D(\gamma)\mathbf{V}^T) &= \text{tr}(\mathbf{V}^T\mathbf{V}D(\gamma_0)^{-1}D(\gamma)) \\ &= \text{tr}(D(\gamma_0)^{-1}D(\gamma)) \\ &= \sum_{i=1}^n \frac{(1 - \gamma)\lambda_i + \gamma\hat{\lambda}_i}{(1 - \gamma_0)\lambda_i + \gamma_0\hat{\lambda}_i} \\ &= \sum_{i=1}^n \alpha(\gamma, \gamma_0)_i \end{aligned}$$

723 (note that if $l \neq 0$, then we would sum over all indices $i \neq r - l + j$ for $j = 1, \dots, l$).

724

725 Second, we compute $\log \frac{|\Sigma_1|}{|\Sigma_0|}$ in (10):

$$\begin{aligned} \log |\Sigma_1| - \log |\Sigma_0| &= \sum_{i=1}^n \left[\log \left((1 - \gamma)\lambda_i + \gamma\hat{\lambda}_i \right) - \log \left((1 - \gamma_0)\lambda_i + \gamma_0\hat{\lambda}_i \right) \right] \\ &= \sum_{i=1}^n \log \left(\frac{(1 - \gamma)\lambda_i + \gamma\hat{\lambda}_i}{(1 - \gamma_0)\lambda_i + \gamma_0\hat{\lambda}_i} \right) \\ &= \sum_{i=1}^n \log \alpha(\gamma, \gamma_0)_i \end{aligned} \quad (13)$$

726 It results:

$$\text{KLD}(\pi_1 || \pi_0) = \frac{1}{2} \left(\sum_{i=1}^n \alpha(\gamma, \gamma_0)_i - (n - r) - \sum_{i=1}^n \log \alpha(\gamma, \gamma_0)_i \right). \quad (14)$$

727 Below we compute the term $\alpha(\gamma, \gamma_0)_i$ for $i = 1, \dots, r$ and $i = r + 1, \dots, n$:

728 • $i = 1, \dots, r$ ($\hat{\lambda}_i = 0$):

$$\alpha(\gamma, \gamma_0)_i = \frac{\frac{1-\gamma}{1-\gamma_0}\lambda_i + \frac{\gamma}{1-\gamma_0}0}{\lambda_i + \frac{\gamma_0}{1-\gamma_0}0} = \frac{1-\gamma}{1-\gamma_0}$$

729 • $i = r + 1, \dots, n$ ($\lambda_i = 0$):

$$\alpha(\gamma, \gamma_0)_i = \frac{\frac{1-\gamma}{1-\gamma_0} 0 + \frac{\gamma}{1-\gamma_0} \hat{\lambda}_i}{0 + \frac{\gamma_0}{1-\gamma_0} \hat{\lambda}_i} = \frac{\gamma}{\gamma_0}$$

730 Note that the eigenvalues of \mathbf{Q}_0 and \mathbf{Q}_1 turn out to be irrelevant for computing the KLD, as they
731 cancel out in the $\alpha(\gamma, \gamma_0)_i$ terms above. Finally, the KLD is:

$$\text{KLD}(\pi_1 || \pi_0) = \frac{1}{2} \left[r \frac{1-\gamma}{1-\gamma_0} + (n-r) \frac{\gamma}{\gamma_0} - (n-r) - r \log \frac{1-\gamma}{1-\gamma_0} - (n-r) \log \frac{\gamma}{\gamma_0} \right]. \quad (15)$$

732 For $\gamma_0 \rightarrow 0$ and $\gamma_0 \ll \gamma < 1$ the dominant term in expression (15) is $(n-r) \frac{\gamma}{\gamma_0}$. Therefore, the
733 distance from the base model, measured as $d(\gamma) = \sqrt{2KLD}$, is

$$\begin{aligned} d(\gamma) &= \lim_{\gamma_0 \rightarrow 0} \sqrt{r \frac{1-\gamma}{1-\gamma_0} + (n-r) \frac{\gamma}{\gamma_0} - (n-r) - r \log \frac{1-\gamma}{1-\gamma_0} + (n-r) \log \frac{\gamma}{\gamma_0}} \\ &\simeq \sqrt{(n-r) \frac{\gamma}{\gamma_0}} = c\sqrt{\gamma}, \end{aligned}$$

734 for a constant $c > 0$ that does not depend on γ . Since $0 \leq d(\gamma) \leq c$, assigning a truncated
735 exponential with rate λ on $d(\gamma)$ we have

$$\pi(d(\gamma)) = \frac{\lambda \exp(-\lambda c \sqrt{\gamma})}{1 - \exp(-\lambda c)}, \quad 0 \leq d(\gamma) \leq c, \quad \lambda > 0.$$

736 Applying a change of variable and reparametrizing $\theta = \lambda c$ leads to the PC prior for γ :

$$\pi(\gamma) = \frac{\theta \exp(-\theta \sqrt{\gamma})}{2\sqrt{\gamma}(1 - \exp(-\theta))} \quad 0 < \gamma < 1, \theta > 0$$

737 which completes the proof.

738 6.2 Proof of Result 1 for interaction types I, II and III

739 From Appendix 6.1, it is clear that the proof works provided that a common eigenbasis can be
740 found for matrices \mathbf{Q}_0 (which is the same as in Appendix 6.1) and \mathbf{Q}_1 (that changes depending on
741 the type of interaction). We first illustrate that this is case for interaction types I, II and III, to then
742 show that the KLD remains the unchanged.

743

744 Interaction type I

745 For the type I interaction, $\mathbf{Q}_1 = \mathbf{I}_{n_2} \otimes \mathbf{I}_{n_1}$ so it has a single eigenvalue equal to 1 with

746 multiplicity $n_1 n_2$. Given that any vector of $\mathbb{R}^{n_1 n_2}$ is an eigenvector of \mathbf{Q}_1 , it is enough to use the
 747 eigenvectors from the eigendecomposition of \mathbf{Q}_0 as a common eigenbasis.

748

749 Interaction type II

750 For the type II interaction, $\mathbf{Q}_1 = \mathbf{I}_{n_2} \otimes \tilde{\mathbf{R}}_1$ has $2n_2$ eigenvectors associated to null eigenvalues,
 751 and $n_2(n_1 - 2)$ eigenvectors associated to non-null eigenvalues, that come from the tensor product
 752 of non-null eigenvectors from the matrices \mathbf{I}_{n_2} and \mathbf{R}_1 . Let $e_1^{\mathbf{R}_1}, \dots, e_{n_1-2}^{\mathbf{R}_1}$ be the eigenvectors
 753 associated to non-null eigenvalues of \mathbf{R}_1 ; the first $n_1 - 2$ eigenvectors associated to non-null
 754 eigenvalues of the matrix \mathbf{Q}_0 are:

$$\mathbf{1}_{n_2} \otimes e_1^{\mathbf{R}_1}, \dots, \mathbf{1}_{n_2} \otimes e_{n_1-2}^{\mathbf{R}_1} \quad (16)$$

755 while the first $n_1 - 2$ eigenvectors associated to non-null eigenvalues of the matrix \mathbf{Q}_1 are:

$$e_1 \otimes e_1^{\mathbf{R}_1}, \dots, e_1 \otimes e_{n_1-2}^{\mathbf{R}_1}$$

756 where e_1 is the first eigenvector of the identity matrix \mathbf{I}_{n_2} . We can eigen decompose the identity
 757 matrix using the eigenbasis for \mathbf{R}_2 , so that $e_1 = \mathbf{1}_{n_2}$; this guarantees that a common matrix
 758 of eigenvectors \mathbf{V} can be found. In particular, it would be formed of the $n_1 + n_2 - 3$ non null
 759 eigenvectors from \mathbf{Q}_0 and the $n_1 n_2 - 2n_2$ non null eigenvectors from \mathbf{Q}_1 . Note that these two
 760 collection of vectors will have $n_1 - n_2 - 3$ vectors in common from the eigenvectors in (16) if
 761 $n_1 > n_2 + 3$.

762

763 Interaction type III

764 In the type III interaction, $\mathbf{Q}_1 = \tilde{\mathbf{R}}_2 \otimes \mathbf{I}_{n_1}$ has n_1 eigenvectors associated to null eigenvalues
 765 and $n_1 n_2 - n_1$ eigenvectors with non-null eigenvalues. In particular, let $e_1^{\mathbf{R}_2}, \dots, e_{n_2-1}^{\mathbf{R}_2}$ be the
 766 eigenvectors associated to non-null eigenvalues of \mathbf{R}_2 ; the following are $n_2 - 1$ eigenvectors
 767 associated to non-null eigenvalues of the matrix \mathbf{Q}_0 :

$$e_1^{\mathbf{R}_2} \otimes \mathbf{1}_{n_1}, \dots, e_{n_2-1}^{\mathbf{R}_2} \otimes \mathbf{1}_{n_1} \quad (17)$$

768 while for matrix \mathbf{Q}_1 we find the following $n_2 - 1$ eigenvectors associated to non-null eigenvalues

769 :

$$e_1^{\mathbf{R}_2} \otimes e_1, \dots, e_{n_2-1}^{\mathbf{R}_2} \otimes e_1$$

770 where e_1 is the first eigenvector of the identity matrix \mathbf{I}_{n_1} . Similarly to the type II interaction,
 771 we can use the eigenbasis for \mathbf{R}_1 to eigen decompose \mathbf{I}_{n_1} so that a common eigenbasis can be
 772 found. It would be formed of the $n_1 + n_2 - 3$ non null eigenvectors from \mathbf{Q}_0 and the $n_1 n_2 - n_1$
 773 non-null eigenvectors from \mathbf{Q}_1 . Note that these two collection of vectors will have $n_2 - 3$ vectors
 774 in common from the eigenvectors in (17) if $n_2 > 3$.

775

776 Regarding the KLD, which is calculated based on the eigenvalues of \mathbf{Q}_0 and \mathbf{Q}_1 , whenever
 777 the rank of \mathbf{Q}_0 is not smaller than the rank deficiency of \mathbf{Q}_1 , there will be a number of pairs of
 778 eigenvalues that are not zero contemporarily. This number is equal to $n_1 + n_2 - 3$ in the type I,
 779 $n_1 - n_2 - 3$ in the type II and $n_2 - 3$ in the type III interaction. Nevertheless, the contribution
 780 of the corresponding term $\alpha(\gamma, \gamma_0)_i$ in the KLD is minimal and the dominant term when $\gamma_0 \rightarrow 0$
 781 remains the same as shown in Appendix 6.1 for the type IV interaction, so the PC prior does not
 782 change.

783 6.3 Model with structured and unstructured main effects

784 In the case of structured and unstructured main effects, matrix \mathbf{Q}_0^- :

$$\begin{aligned} \mathbf{Q}_0^- = & (1 - \phi)(\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1}) \left((1 - \psi_1)\tilde{\mathbf{R}}_1^- + \psi_1 \mathbf{I}_{n_1} \right) (\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1})^T + \\ & \phi(\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1}) \left((1 - \psi_2)\tilde{\mathbf{R}}_2^- + \psi_2 \mathbf{I}_{n_2} \right) (\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1})^T \end{aligned}$$

785 and $\text{rank}(\mathbf{Q}_0) \leq n_1 + n_2$.

786

787 Interaction type IV

788 Following the proof in Appendix 6.1, it is enough to show that $\text{rank}(\mathbf{Q}_0) \leq 2n_2 + n_1 - 2$. Given
 789 that the rank of \mathbf{Q}_0 is at most $n_1 + n_2$, the rank condition is true provided that $0 \leq n_2 - 2$, i.e.
 790 that there are at least 2 spatial locations, which is always true in practice.

791

792 Interaction types I,II, III

793 For interaction types I, II and III it is still possible to find a common eigenbasis, as adding a constant
 794 to the diagonal of a matrix does not change its eigenvectors. The eigenvalues do change though, so
 795 now the number of eigenvalues that are not zero contemporarily in \mathbf{Q}_0 and \mathbf{Q}_1 (whenever the rank
 796 of \mathbf{Q}_0 is not smaller than the rank deficiency of \mathbf{Q}_1) are $n_1 + n_2 - 1$ for type I, $n_1 - n_2 - 1$ for
 797 type II and $n_2 - 1$ for type III, and the dominant term in the KLD remains the same as before.

7 Simulation study

We run a simulation study to investigate the performance of the VP model when using the PC prior for γ proposed in Section 3.1 Eq. (5). We generate datasets based on the space and time patterns estimated from the Covid-19 data described in Section 4.2; to limit the computational burden we select a subset of the the full dataset (north provinces, wave 1) with $n_1 = 17$ weeks and $n_2 = 47$ provinces. Assume i and j are indices for weeks and provinces, respectively, we simulate data as

$$y_{ij} \sim \text{Bin}(\text{pop}_j, \mu_{ij}), \quad (18)$$

$$\text{logit}(\mu_{ij}) = \sqrt{1/\tau} \left\{ \sqrt{1-\gamma} \left[\sqrt{1-\phi} \hat{\beta}_{1i} + \sqrt{\phi} \hat{\beta}_{2j} \right] + \sqrt{\gamma} \hat{\delta}_{i,j} \right\}, \quad (19)$$

where pop_j is the population in province j , μ_{ij} the Covid-19 incidence rate at week i in province j . The vectors $\hat{\beta}_1 = (\hat{\beta}_{1,1}, \dots, \hat{\beta}_{1,n_1})^T$, $\hat{\beta}_2 = (\hat{\beta}_{2,1}, \dots, \hat{\beta}_{2,n_2})^T$ and $\hat{\delta} = \{\hat{\delta}_{ij}, i = 1, \dots, n_1, j = 1, \dots, n_2\}$ contain the posterior means for, time, space and space-time random effects, respectively. These estimates come from the VP model (Eq. (4) in Section 3) fitted to the Covid-19 data, north provinces wave 1, by assuming a type IV interaction (see the top panels in Figure 5 for the time and space main effects). We further assume $\tau = 1, \phi = 0.5$ and keep them fixed throughout the simulation study, while letting the mixing parameter γ vary, in order to create different scenarios according to the contribution of the interaction to the total (generalized) variance.

Our goals are: 1) to check how well the true γ is recovered when estimated using our VP model Eq. (4) - where, as an estimator for γ we take the posterior mean; 2) to assess sensitivity to the choice of θ , the scaling parameter for the PC prior on γ .

7.1 Simulation study scenarios

The following scenarios are considered regarding the contribution of the interaction to the total variance:

- SC1: $\gamma = 0$ (additive model, no interaction);
- SC2: $\gamma = 1/10$ (low interaction);
- SC3: $\gamma = 1/3$ (moderate interaction);
- SC4: $\gamma = 2/3$ (strong interaction).

Scenario SC1 ($\gamma = 0$) assumes an additive model where the time pattern remains the same across provinces. Scenarios SC2 and SC3 represent cases of, respectively, low and moderate interaction.

825 SC4 is intended as a limiting case where the interaction between space and time main effects is
 826 very strong; as we can see from Figure 5 (bottom right panel), where one simulated dataset under
 827 SC4 is displayed, the temporal pattern can vary substantially across provinces and some of them
 828 show a decreasing trend at the beginning of the first wave period, which is clearly unrealistic for
 829 Covid-19 disease.

830 We consider different scenarios by letting the number of trials of the Binomial model, pop_j , $j =$
 831 $1, \dots, n_2$, vary. The following three sample size (i.e. population at risk) levels are considered:

- 832 • Actual sample size: the population in province j is taken as pop_j ;
- 833 • Smaller sample size: the population in province j is taken as $pop_j/10$;
- 834 • Larger sample size: the population in province j is taken as $pop_j \cdot 10$.

835 The second scenario represents a smaller sample size case, where the data carries less information
 836 about γ thus we expect less accuracy in the model estimates; analogously, the third scenario
 837 represents a case where data are more informative about γ , hence we expect the model to provide
 838 improved estimates in this case.

839 We simulated 100 datasets under SC1, SC2, SC3 and SC4, for each of the three different sample
 840 size levels described above. The VP model (Eq. (4), Section 3) was fitted to each dataset assuming a
 841 RW1 as the time main effect, an ICAR as the space main effect and a type IV space-time interaction.
 842 All the computations were done using R-INLA.

843 The VP model was fitted under 4 different prior choices for γ :

- 844 • prior 1: $PC(U = 0.05, a = 0.99)$;
- 845 • prior 2: $PC(U = 0.5, a = 0.99)$;
- 846 • prior 3: $PC(U = 0.95, a = 0.99)$;
- 847 • prior 4: $Uniform(0, 1)$.

848 The first three priors consider the different scalings of the PC prior displayed in Figure 1. This
 849 way, robustness of the results for changing U can be tested. The values $U = \{0.05, 0.5, 0.95\}$
 850 reflect, respectively, an inflexible, moderate and flexible prior on the space-time interaction random
 851 effects. We also estimated the model using a uniform prior for γ .

852 Regarding τ and ϕ , we assigned a Gumbel type 2 PC prior on τ and a $Uniform(0, 1)$ on ϕ
 853 to express ignorance about the variance contribution of space (and time). We considered two
 854 different scalings of the PC prior on τ ($U = 2/0.31$ and $U = 100/0.31$) but they did not make
 855 any difference on posterior estimates.

7.2 Simulation study results

Figures 6 to 9 report the boxplots of the posterior mean of γ obtained by fitting the VP model to the 100 simulated datasets under the four scenarios SC1, SC2, SC3 and SC4. The horizontal dashed line represents the true γ set by simulation in each scenario. Each figure has three panels that refer to actual (left), smaller (central) and larger (right) sample size cases. The four boxplots in each panel correspond to different priors on γ : the PC priors with scalings $U = \{0.05, 0.5, 0.95\}$ and the Uniform prior.

Regarding SC1 (Figure 6), where the true γ is zero, we see that the Uniform prior implies a larger bias than the PC prior choices do, which is presumably due to the Uniform being prone to overfitting. This behaviour is more evident in the small sample size case, as a result of the data being less informative about the proportion of variance explained by the interaction. In scenarios where the true $\gamma > 0$ (i.e. SC2, SC3 and SC4) we generally observe a negative bias under all prior choices, however the bias is smaller as the sample size increases.

Regarding the first aim of the study, i.e. checking the ability to recover the true γ set by simulation, we can conclude that estimation of γ is reasonable in all cases. We would like to emphasize that while the bias achieved under the uniform prior is always slightly smaller than the bias obtained by the PC priors in scenarios SC2, SC3 and SC4, it becomes much larger in SC1 because of the tendency to overfitting of the uniform. This highlights the fundamental advantage of PC priors which avoid overfitting by default as they shrink to the base model $\gamma = 0$ by construction.

Regarding our second aim, i.e. studying sensitivity of the results to the choice of θ , we notice that as long as the inflexible choice of $U = 0.05$ is avoided, the mixing γ is estimated fairly well using the moderate and flexible choices, $U = 0.5$ or 0.95 . In particular, $U = 0.5$ or $U = 0.95$ return comparable estimates of the mixing parameter γ under all scenarios. From these results, we suggest that in absence of strong prior information on γ the choice of a PC prior with $U = 0.95$, $a = 0.99$ is a reasonable weakly informative prior on γ that allows flexibility and at the same time avoids model overfitting.

As regards estimation of ϕ and τ , results (not reported here) show that the true values $\phi = 0.5$ and $\tau = 1$ are accurately estimated in all scenarios by all priors.

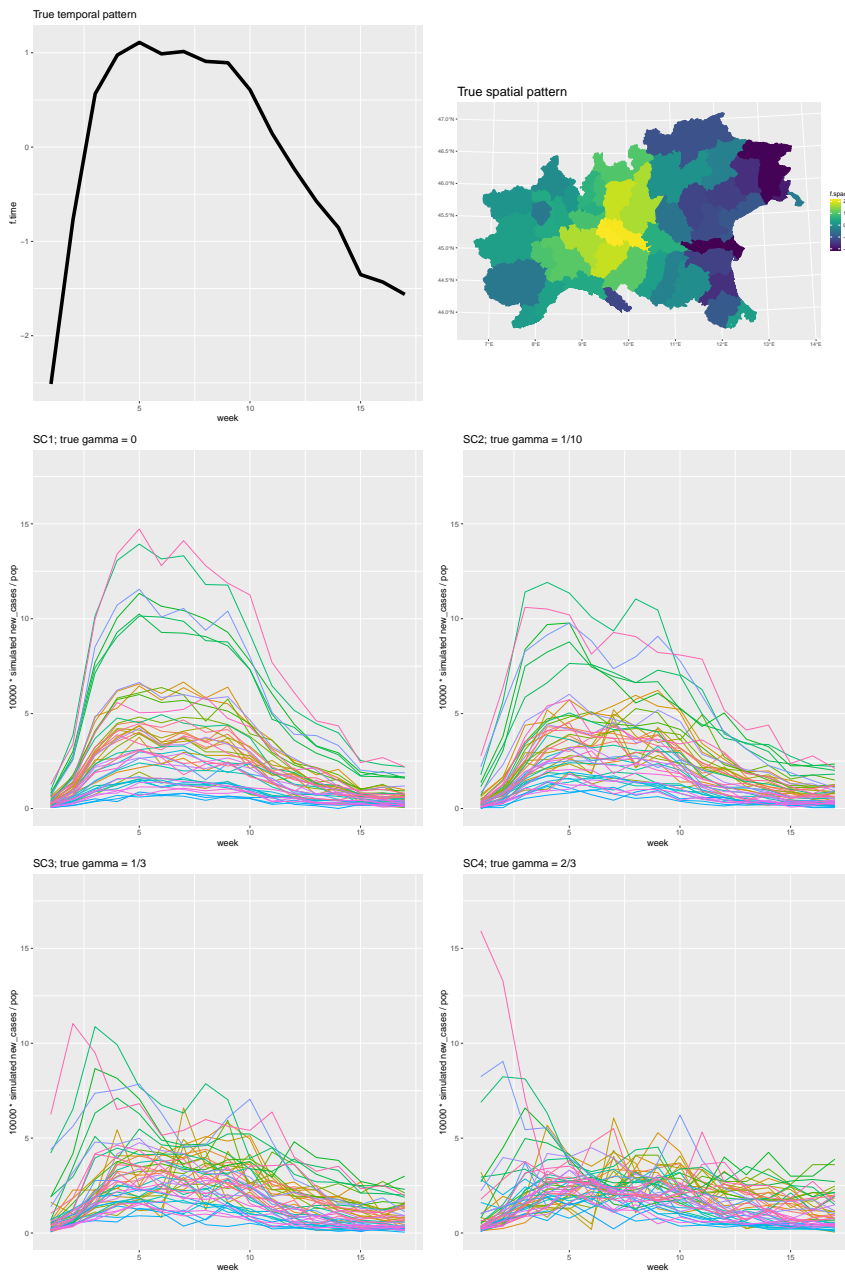


Figure 5. Simulation scenarios. Top panels: plots of the main effects for time and space. The central and bottom panels display one simulated dataset under each of the four scenarios (SC1, SC2, SC3, SC4) varying according to the strength of the interaction γ .

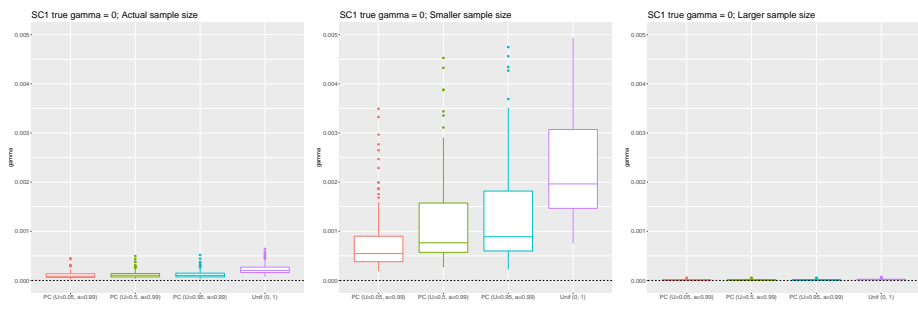


Figure 6. Simulation results for the mixing parameter γ , under scenario SC1; true $\gamma = 0$.

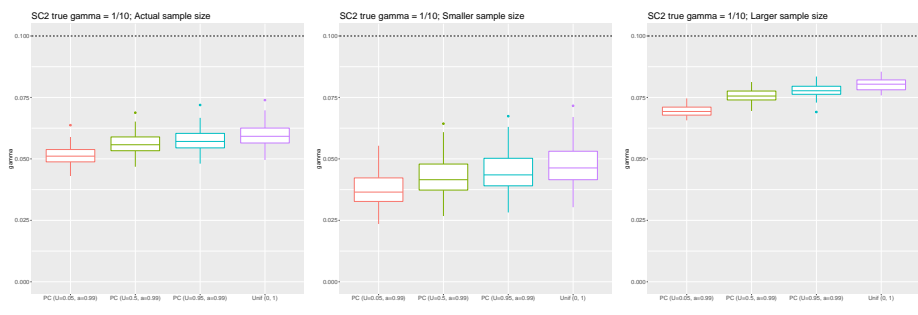


Figure 7. Simulation results for the mixing parameter γ , under scenario SC2; true $\gamma = 1/10$.

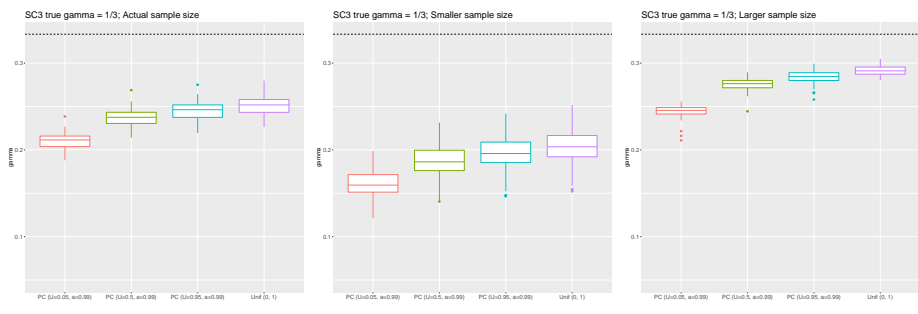


Figure 8. Simulation results for the mixing parameter γ , under scenario SC3; true $\gamma = 1/3$.

884 **8 Additional material on Ohio and Covid-19 examples**

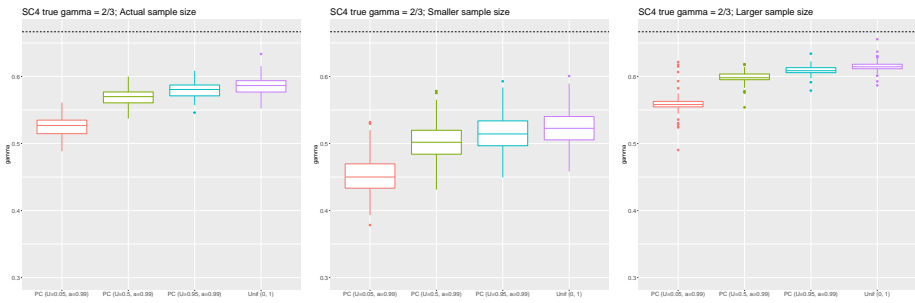


Figure 9. Simulation results for the mixing parameter γ , under scenario SC4; true $\gamma = 2/3$.

Table 4. Variance partitioning table for Ohio lung cancer, comparing the four interaction types.

source	estimator	type I	type II	type III	type IV
main	$1 - \hat{\gamma}$	0.952 (0.913, 0.979)	0.958 (0.923, 0.981)	0.973 (0.940, 0.991)	0.960 (0.924, 0.983)
int	$\hat{\gamma}$	0.048 (0.021, 0.087)	0.042 (0.019, 0.077)	0.027 (0.009, 0.060)	0.040 (0.017, 0.076)
main:space	$\hat{\phi}$	0.875 (0.765, 0.946)	0.874 (0.763, 0.943)	0.878 (0.770, 0.945)	0.874 (0.758, 0.944)
main:time	$1 - \hat{\phi}$	0.125 (0.054, 0.235)	0.126 (0.057, 0.237)	0.122 (0.055, 0.230)	0.126 (0.056, 0.242)
main:time:iid	$\hat{\psi}_1$	0.069 (0.010, 0.229)	0.058 (0.003, 0.221)	0.050 (0.002, 0.203)	0.056 (0.003, 0.214)
main:time:str	$1 - \hat{\psi}_1$	0.931 (0.771, 0.990)	0.942 (0.779, 0.997)	0.950 (0.797, 0.998)	0.944 (0.786, 0.997)
main:space:iid	$\hat{\psi}_2$	0.658 (0.273, 0.925)	0.693 (0.388, 0.917)	0.706 (0.372, 0.927)	0.676 (0.361, 0.912)
main:space:str	$1 - \hat{\psi}_2$	0.342 (0.075, 0.727)	0.307 (0.083, 0.612)	0.294 (0.073, 0.628)	0.324 (0.088, 0.639)

885 9 R code

886 Below the R-INLA code to fit model (6), with type 4 interaction, to the Covid-19 dataset in Section
 887 4.2. Note that the model can be estimated using the usual `inla` call; the R package `inlaVP` was
 888 written to aid the user in setting the interaction type, building the constraints and defining the joint
 889 prior. The R package `inlaVP` is not on CRAN yet, but it is available on github.

```

890
891 1 rm(list=ls())
892 2 library(INLA)
893 3 # install inlaVP using devtools
894 4 library(devtools)
895 5 install_github("massimoventrucci/inlaVP")
896 6 library(inlaVP)
897 7
898 8 ## load the data and create interaction index
899 9 data(covid_italy)
90010 n1 <- length(unique(covid_italy$id.week))
90111 n2 <- italy_graph$n
90212 dat.tmp <- expand.grid(id.week=1:n1,
90313                       id.province=1:n2)
90414 dat.tmp$id.int <- 1:(n1*n2)
90515 dat <- merge(covid_italy, dat.tmp,
90616             by=c("id.week", "id.province"),
90717             all.x=TRUE)
90818 dat.sort <- dat[order(dat$id.int),]
90919 # IMP: sorting the interaction indices is needed
91020
91121 # the graph for Italy is disconnected (3 connected component 'cc'):
91222 # set one separate intercept for each cc of size > 1
91323 intercept <- rep(NA, graph$n)
91424 for(i in seq_along(graph$cc$nodes))
91525   if (length(graph$cc$nodes[[i]]) > 1) intercept[graph$cc$nodes[[i]]] <-
916   i
91726 intercept <- as.factor(intercept)
91827 dat.sort <- merge(dat.sort, data.frame(id.province=1:graph$n, intercept.
919   cc=intercept))
92028
92129 ## inla call
92230 library(INLA)
92331 inla.setOption(num.threads = "1")
92432 # setting 1 core is needed when using joint prior (jp) inside control.
925   expert = list(jp = ...),

```

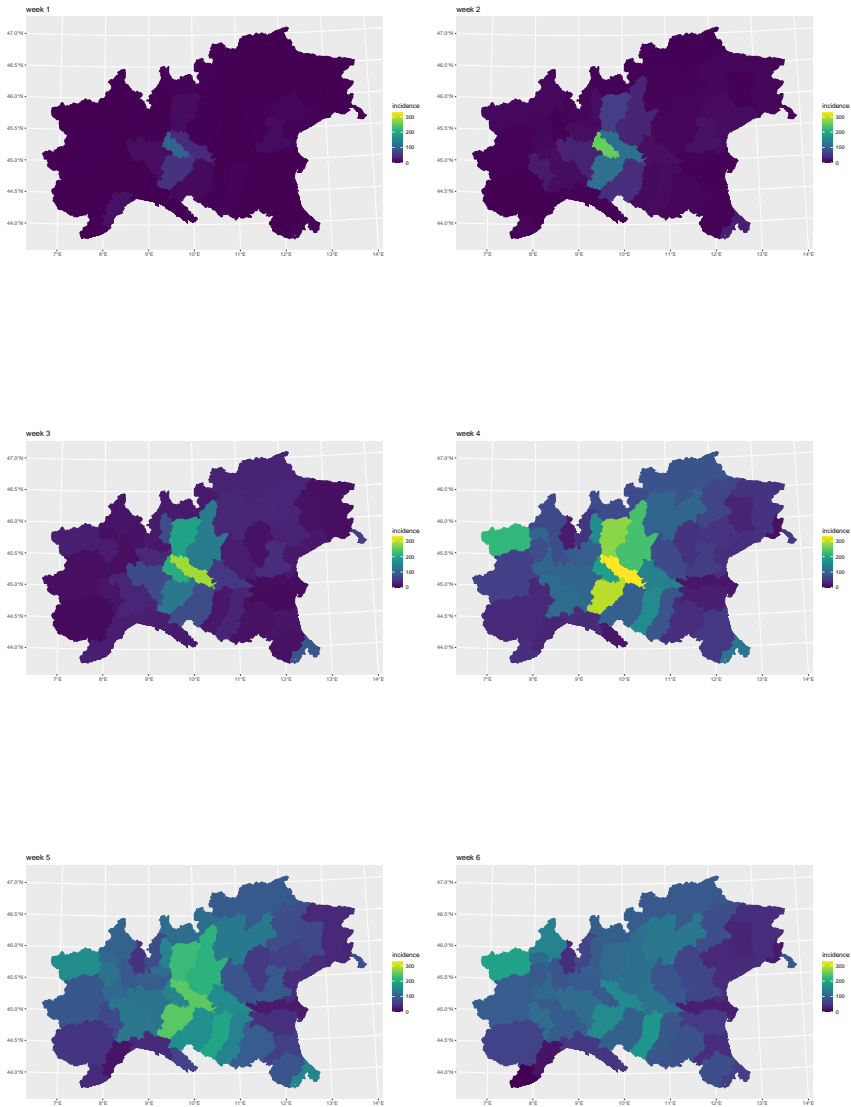


Figure 10. Observed weekly cases of Covid-19 per 100000 residents in North Italy during first wave, from week 1 to week 6.

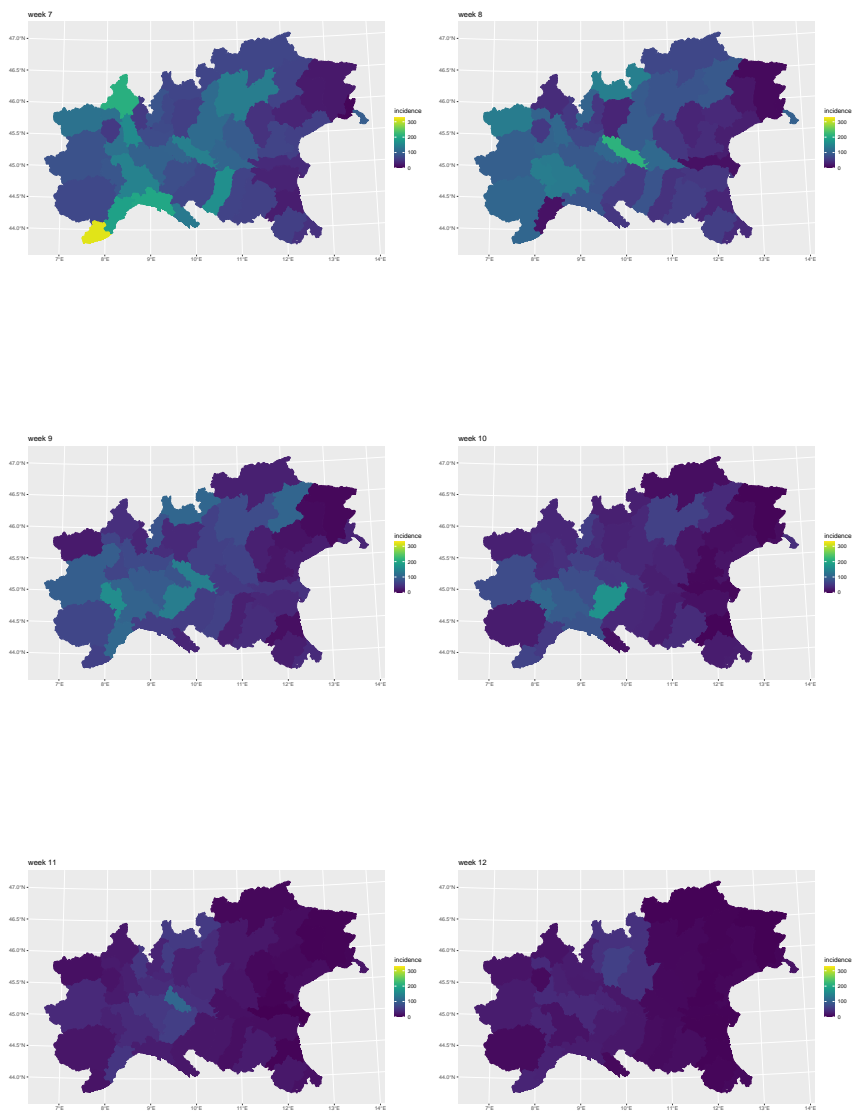


Figure 11. Observed weekly cases of Covid-19 per 100000 residents in North Italy during first wave, from week 7 to week 12.

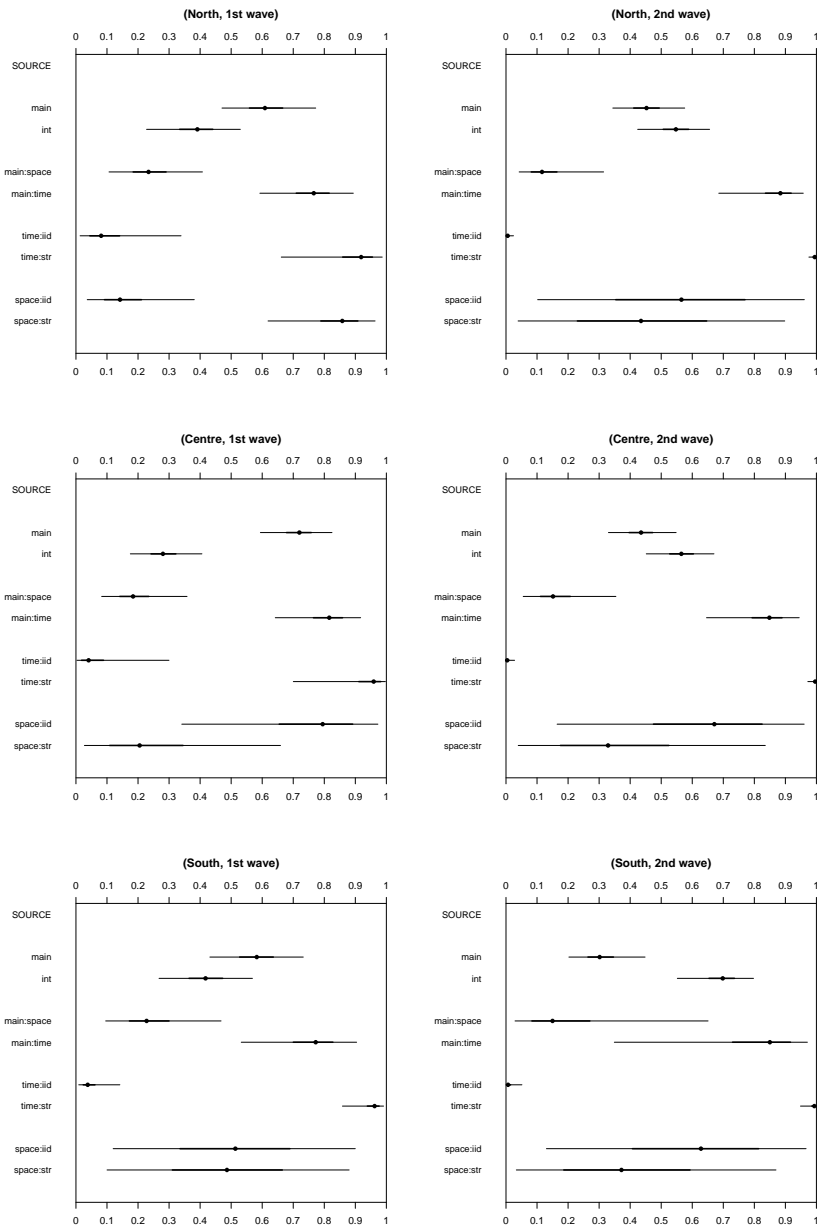


Figure 12. Variance partitioning plots for the Covid-19 example; the analysis here refer to different subset of the data for each combination of the factors geographical area, with levels north (N), centre (C) and south (S), and pandemic wave, with levels W1 and W2.

```

92633
92734 # define the interaction model
92835 set.int <- control.interaction(
92936   m1 = m(covid_italy$id.week, igmrf.type = "rw1"),
93037   m2 = m(covid_italy$id.province, igmrf.type = "besag", g=italy_graph),
93138   interaction.type = 4)
93239
93340 # define the joint prior
93441 jp.vp.m2 <- function(theta, theta.desc = NULL) {
93542   ### the user must specify 'hyper', with the scaling parameters of the
936   PC priors for tau and gamma:
93743   hyper <- list(prec=list(u=2/0.31, a=0.01),
93844                 gamma=list(u=0.95, a=0.99))
93945   fun_striid <- function(theta)
94046   {
94147     tau <- inlaVP:::theta.to.tau.striid(theta)
94248     gamma <- inlaVP:::theta.to.gamma.striid(theta)
94349     phi <- inlaVP:::theta.to.phi.striid(theta)
94450     psi1 <- inlaVP:::theta.to.psil.striid(theta)
94551     psi2 <- inlaVP:::theta.to.psi2.striid(theta)
94652     return(c(phi,gamma,tau,psi1,psi2))
94753   }
94854
94955   if (!is.null(theta.desc)) {
95056     for(i in seq_along(theta.desc))
95157       print(paste0("   theta[", i, "]= ", theta.desc[i]))
95258   }
95359   if (inlaVP:::theta.to.phi.striid(theta) >=0 & inlaVP:::theta.to.phi.
954     striid(theta) <=1 &
95560     inlaVP:::theta.to.psil.striid(theta) >=0 & inlaVP:::theta.to.psil.
956     striid(theta) <=1 &
95761     inlaVP:::theta.to.psi2.striid(theta) >=0 & inlaVP:::theta.to.psi2.
958     striid(theta) <=1 ){
95962     lprior <- INLA:::inla.pc.dprec(prec=inlaVP:::theta.to.tau.striid(
960       theta),
96163                                     u= hyper$prec$u, alpha=hyper$prec$a,
962                                     log=TRUE) +
96364     inlaVP:::pc.gamma(gamma=inlaVP:::theta.to.gamma.striid(theta),
96465                       lambda=inlaVP:::pcprior.interaction.lambda(
96566                         u=hyper$gamma$u, alpha=hyper$gamma$a),
96667                       log=TRUE) +

```

```

96768     log(abs(det(numDeriv::jacobian(fun_striid, as.numeric(theta),
968         method="Richardson"))))
96969   } else {
97070     lprior <- -.Machine$double.xmax
97171   }
97272   return(lprior)
97373 }
97474 jpr.vp <- inla.jp.define(jp.vp.m2)
97575
97676 # set ini
97777 theta.ini <- taugammaphipsilpsi2.to.theta(taugammaphipsilpsi2 = c
978     (1,0.25,0.5,0.5,0.5))
97978
98079 # run inla
98180 res.covid <- inla(y ~ 1 + intercept.cc +
98281     f(id.time,
98382     model='rw1',
98483     constr = T,
98584     scale.model=T) +
98685     f(id.space,
98786     model='besag',
98887     graph=italy_graph,
98988     adjust.for.con.comp = T,
99089     constr = T,
99190     # Note: if adjust.for.con.comp = T,
99291     # then 'constr = T' interpreted as a sum-to-zero
993     constr on each cc of size > 1
99492     scale.model=T) +
99593     f(id.int,
99694     model="generic0",
99795     Cmatrix = set.int$Rkron,
99896     constr = F,
99997     extraconstr = set.int$constr) +
100098     f(id.time2, model='iid') +
100199     f(id.space2, model='iid'),
100200     data = list(y = dat.sort$new_cases,
100301         id.time=dat.sort$id.week,
100402         id.time2=dat.sort$id.week,
100503         id.space=dat.sort$id.province,
100604         id.space2=dat.sort$id.province,
100705         id.int=dat.sort$id.int,
100806         pop=dat.sort$pop_province),

```

```
100907         family = 'binomial', Ntrials=pop,
101008         control.expert = list(jp = jpr.vp),
101109         control.predictor = list(link=1),
101210         control.compute = list(config=TRUE,
101311                             dic=TRUE,
101412                             waic = TRUE,
101513                             cpo=TRUE))
101614
101715 ## VP plot
101816 vp.plot(res.covid, main=paste('Vp plot'))
1019
```