Bayesian factor analysis for spatially correlated data: application to cancer incidence data in Scotland

Maura Mezzetti

Abstract   A hierarchical Bayesian factor model for multivariate spatially correlated data is proposed. Multiple cancer incidence data in Scotland are jointly analyzed, looking for common components, able to detect etiological factors of diseases hidden behind the data. The proposed method searches factor scores incorporating a dependence within observations due to a geographical structure. The great flexibility of the Bayesian approach allows the inclusion of prior opinions about adjacent regions having highly correlated observable and latent variables. The proposed model is an extension of a model proposed by Rowe (2003a) and starts from the introduction of separable covariance matrix for the observations. A Gibbs sampling algorithm is implemented to sample from the posterior distributions.

Keywords   Bayesian inference · Cancer incidence data · Correlated factor loadings · Factor analysis · Spatial data

1 Introduction

Spatially referenced data occur in diverse scientific disciplines, such as environmental sciences (Webster and Oliver 2001; Wikle 2003), ecological systems (Scheiner and Gurevitch 2001), econometrics (Anselin 2000), disease mapping (Lawson 2001) and in broader public health contexts (Waller and Gotway 2004). Spatial data are referenced over a fixed set of locations, that are sometimes discretely indexed with well-defined neighbors (typically regions or counties in a map, etc.); alternatively, they

M. Mezzetti (✉)
Dipartimento SEFEMEQ, Facoltà Economia, Università Tor Vergata, Via Columbia 2,
00133 Rome, Italy
e-mail: maura.mezzetti@uniroma2.it

may be referenced by precise coordinates (latitude-longitude, easting-northing, etc.), a situation called point-level or geostatistical data.

Spatial data analysis is challenged by the presence of a possible dependence among values from neighboring regions (or points) and the main difficulty is taking into account spatial correlation for correct inference. Correlations among values in contiguous areas can sometimes be explained in terms of observed covariates that are also spatially correlated or can result from some unsuspected or unmeasured variables. Mixed effects models have provided a convenient means of modeling spatial correlations by using random effects (Yasui and Lele 1997; Waller et al. 1997). Over the past two decades, spatial statistical methods have been developed for normally distributed data (Cressie and Wikle 2011; Haining et al. 1989), discrete data (Journel 1983; Cressie 1993a,b; Diggle et al. 1992; Cressie et al. 2009) and multivariate data (Le et al. 1997). Statistical models used for spatially correlated data can be fully parameterized, with inference procedures based on maximum likelihood (Clayton and Kaldor 1987; Cressie 1993a,b), penalized maximum likelihood (Breslow and Clayton 1993) and Markov chain Monte Carlo (Besag et al. 1991; Waller et al. 1997). Bayesian methods are increasingly being applied to the analysis of spatial data (Best et al. 2005) since they offer a way to accurately quantify uncertainty and they incorporate spatial correlation, through a flexible and robust approach.

The paper focuses on the analysis of geographical variation in disease rates, whose interest is increasing since it is fundamental in the formulation and validation of etiological hypotheses. Mapping the geographical distribution of cancer incidence, survival, or mortality rates allows us to describe the geographical distribution of disease, to highlight sources of heterogeneity underlying spatial patterns in the health outcomes and consequently to suggest important public health determinants or etiologic clues. Much of the work in the area of mapping disease has focused on a single disease, however, many diseases share common risk factors, and it is recently emerging that the estimates (either incidence or mortality rates) can be further improved by "borrowing" information from different diseases, not only from adjacent areas. Age-standardized incidence rates for the principal cancer sites in 56 counties of Scotland, relative to 2005, are considered. This data set is particularly interesting in this context since, on one hand, if a county has a high incidence/death rate for one kind of cancer, it most likely has high rates for some other kinds of cancers. On the other hand, the evidence that adjacent neighboring counties have similar cancer rates is emerging. Therefore, the incidence rates are correlated both within county and across counties. The main idea is to explain these two types of correlations, through the flexibility of a Bayesian approach, by assuming that all these different cancer sites share one (or more) spatially correlated common factors. This common component can be interpreted as a surrogate for unobserved covariates that display spatial structure.

Factor analysis is a powerful statistical tool for describing and modeling the underlying structure in multivariate datasets, since it uses the correlation or covariances among a set of observed variables to describe them in terms of a smaller set of latent variables. Most proposals in the factor analysis literature assume that the data represent random, independent samples from a multivariate distribution and this is not a plausible assumption for the spatially referenced data we are analyzing. Despite the acknowledged advantages, Bayesian approaches to factor analysis have been so

far quite limited. Bayesian factor analysis was introduced by Martin and McDonald (1975) and Press and Shigemasu (1989). They had to use rather restrictive model assumptions since most of the computational techniques for Bayesian data analysis were not available yet. Later, after iterative MCMC simulation methods, a few fully Bayesian approaches to factor analysis were proposed, for example, Polasek (1997), Arminger and Muthen (1998), Aguilar and West (2000), Lopes and West (2004) and Rowe (2003a), whose approach inspired most the current work. The Bayesian approach bears directly on the problem of parameter identification, by incorporating proper prior information.

Many new parameters are introduced when we attempt to account for correlation across observed variables, and the problem of indeterminacies is greatly increased. Through a Bayesian model, it is possible to take into account correlation between observations and to simplify the covariance structure by using proper prior information. The idea of factor analysis for correlated data is not new. Wang and Wall (2003) introduced a factor-analytic model for spatially correlated multivariate cause-specific mortality, which provides area-specific scalar summaries of mortality via factor scores. Christensen and Amemiya (2003) developed semiparametric latent variable models, introducing a general spatial factor analysis model that incorporated spatially lagged factor dependencies. Hogan and Tchernis (2004) propose a hierarchical Bayesian factor analytic model of spatially correlated data to summarize area-specific material deprivation from multiple census variables where spatial correlation is modeled on the latent variable scale and can be specified either marginally or conditionally.

The proposed method is a Bayesian approach to factor model dealing with correlation between observations, and it extends Rowe and Press (1998) proposal. The first required step is to rewrite the matrix of observations as a vector $vec(X) = (x_1', x_2', \ldots, x_N')'$, whose length is $N \times p$, N is the number of observations and p number of variables. The observation vector has a separable covariance matrix: $var(vec(X))$ is proportional to $\Phi \otimes \Psi$, where $\otimes$ denotes the Kronecker product. The introduction of a separable covariance matrix represents the main innovation and its great advantage consists on the interpretation: Φ is now the between-observations covariance matrix, and the matrix Ψ is the within-observations covariance matrix. The proposed method is an extension of Mezzetti and Billari (2005), who propose a Bayesian model for analysis of demographic panel data able to handle the temporal dependence between the observations.

The approach follows Rowe (2003a) and Mezzetti and Billari (2005) where the modeling structure is incorporated through the column(s) of the factor score(s). The main novelty is the development of the factor model for spatially correlated data motivated by the simultaneous analysis of multiple cancer incidence data and, furthermore, the introduction of a scalar parameter to simplify the separability structure.

In Sect. 2 the motivating cancer epidemiologic data set is illustrated, while Sect. 3 contains a detailed description of the method proposed by Rowe and Press (1998). Section 4 shows our proposal and the way it extends Rowe's model. After a description of the computational aspects in Sect. 5, the application will be shown in Sect. 6, together with sensitivity analysis in Sect. 7. Finally, Sect. 8 contains final conclusions and remarks.

2 Epidemiological data

In Scotland, there are 56 counties and, for each county, the age-standardized incidence rates per 100,000 population, for the most common cancer sites relative to 2005 are considered (Cancer in Scotland 2010). An age-standardized rate (ASR) is a summary measure of the rate that a population would have if it had a standard age structure. Standardization is necessary when comparing several populations that differ with respect to age, since age has a powerful influence on the risk of cancer. In this case World standard population is used.

Scotland data have been often used for mapping disease study: data are available locally for each county and the distributions of some diseases present a geographical correlation. Incidence rates referred to twelve cancer sites for women are considered, and incidence rates referred to ten cancer sites for men are considered. Female cancer sites considered are: lips, oral, esophagus, stomach, large bowel, pancreas, lung, breast, ovary, uterine corpus, uterine cervix, and hodgkin's lymphoma. Male cancer sites considered are: lips, oral, esophagus, stomach, large bowel, pancreas, lung, prostate, testis and non hodgkin's lymphoma. The most common diseases are selected.

Cancer is now the leading cause of premature death in Scotland, the main idea is that detecting etiological factors between cancer incidences could provide a great improvement from a public health prospective. For males, the most common cancers are lung, prostate and colorectal cancers accounting for 53% of cancers in males. For females, the most common cancers are breast, lung and colorectal cancers accounting for 55% of cancers in women. Thanks to improvement of survival, overall cancer mortality rates have decreased by 12% in males and 5% in females in the last 10 years. Death rates from breast cancer, the most frequently diagnosed cancer in females, have decreased by 12% over the last 10 years, in spite of the increased incidence. Due to improvement in early diagnosis and survival, cervical cancer deaths have decreased by 32% over the same time period.

Since correlations among cancer sites are affected by differences from county to county in diagnosis, treatment, survival rates and other factors, the analysis cannot leave aside the geographical distributions. In Figs. 1, 2, 3, 4 it is possible to observe data correlations from two different points of view: between counties (Figs. 1, 2) and within counties (Figs. 3, 4). Figures 1 and 2 show the geographical distribution of incidence rates for some selected cancer sites for men and women. Observing the distribution of cancer incidences, the presence of spatial aggregation is shown, in particular lung and lips cancer for men, and lung and stomach cancer for women. The Scotland lip cancer example is well-known in the spatial data analysis literature and has been analyzed several times, since Clayton and Kaldor (1987) introduced it.

Figures 3 and 4 show the between-observations sample correlation matrices, respectively within males (starting from a 10×56 matrix) and within females (from a 12×56 matrix). The cell occupying the position h, k represents $cor(x'_h, x'_k)$, the correlation between h and k counties (within the cancer age standardized incidences rates stratified by sex). Observing the Figures, the presence of correlation between regions is evidenced. Under the hypothesis of independence between observations, the two (within observations correlation) matrices would be closer to the identity matrix. The regions are ordered such that adjacent regions have closer rank numbers; thus, area around
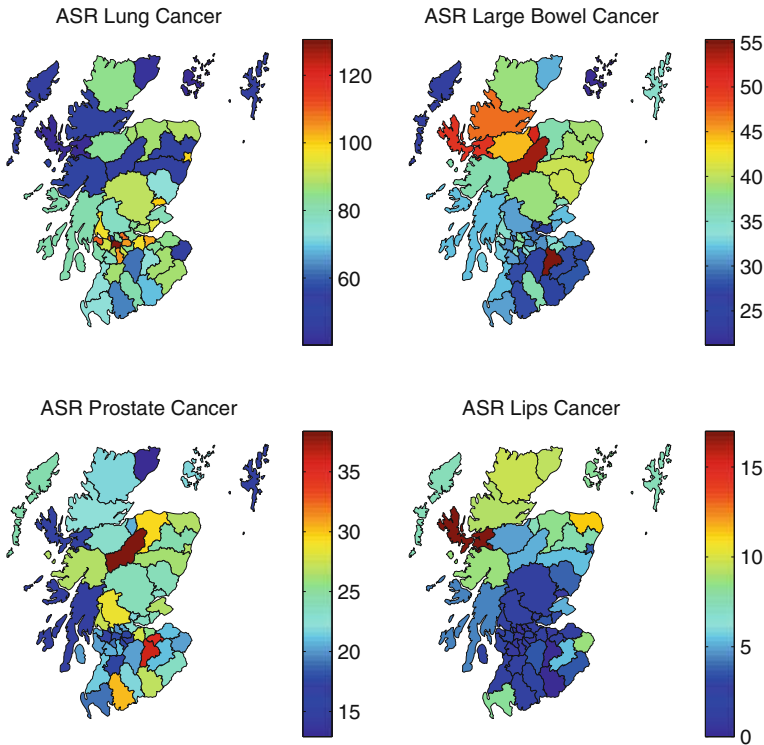
Fig. 1 Distribution of cancer standardized incidence ratio in male

principal diagonal is more likely to represent area of neighboring regions. Observing the two Figures, areas with higher correlations actually correspond to neighboring counties. Comparing Fig. 3 with Fig. 4, presence of spatial correlation is stronger between males than between females. Mean correlation between adjacent regions (within the cancer age standardized incidence rates) is 0.41 for males and 0.15 for females.

Smoking and excessive alcohol use are risk factors for a large number of cancers. According to the National Cancer Institute (www.can\discretionary-cer.gov), smoking damages nearly every organ in the body and is linked to at least ten different cancers. It accounts for nearly 30% of all cancer deaths, tobacco use is specifically associated with cancers of the lung and bronchus, oral cavity (excluding lip), esophagus, stomach and pancreas. Three of the previous cancer sites also share the risk factor of excessive alcohol consumption. Most people who smoke also drink alcohol, therefore the common geographical pattern of some disease may essentially be interpretable as surrogate for geographical variation in tobacco and alcohol consumption across Scotland.

In addition to smoking and alcohol consumption, these cancers share other risk factors; oesophagus and pancreas cancers are linked to bodyweight/obesity; oesophagus and stomach cancer are associated with diet. The association between diet and cancer
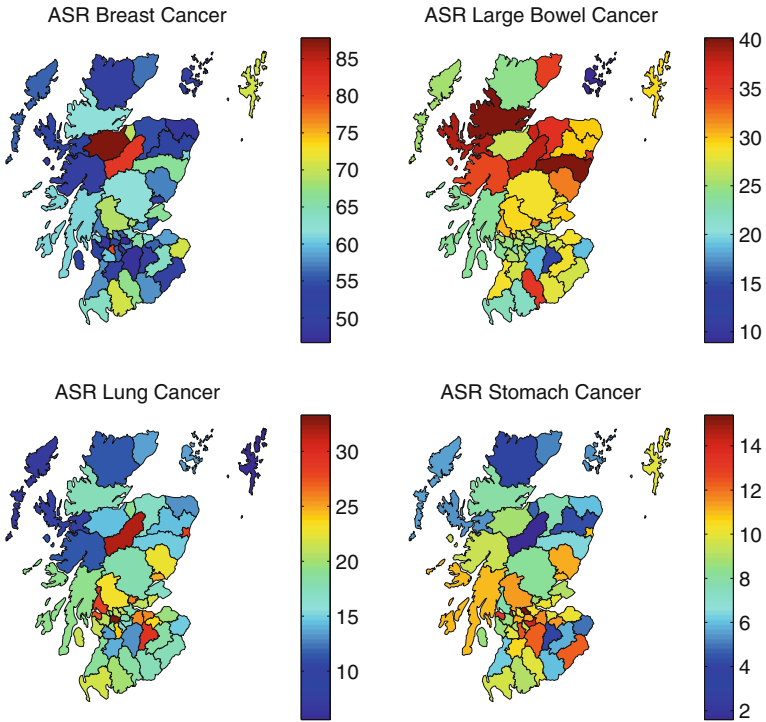
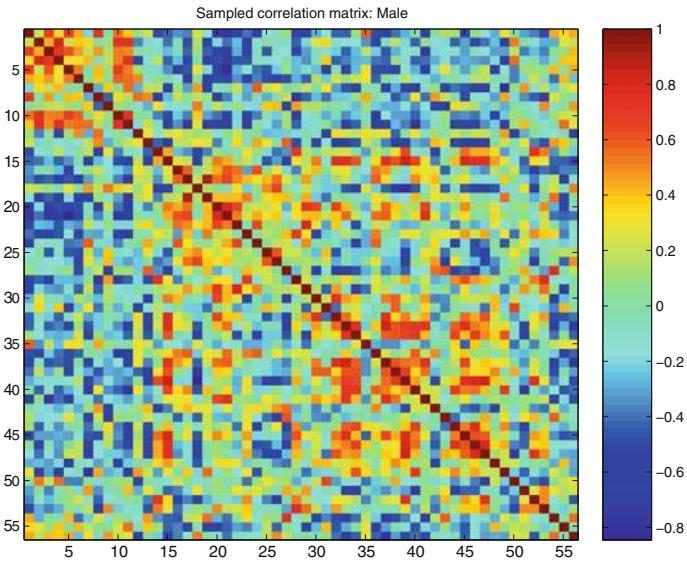**Fig. 2** Distribution of cancer standardized incidence ratio in female

Fig. 3 Between observations sample correlation matrix (male)
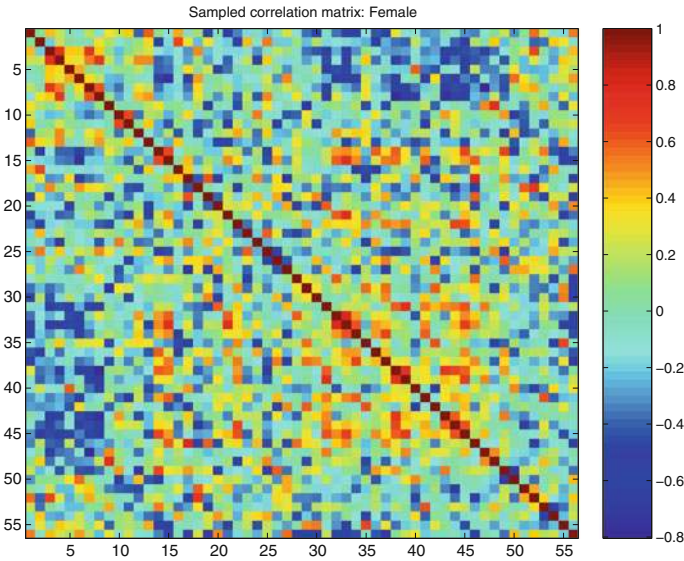
Sampled correlation matrix: Female

Fig. 4 Between observations sample correlation matrix (female)

creates a need to understand the causes of cancer beyond those related to tobacco or other established risk factors. A large body of literature indicates that as much as 30% of all cancer cases is linked to poor dietary habits, and is therefore preventable. Understanding the distribution of environmental and behavioral risk factors can be fundamental for public health strategies.

The idea is that different latent components may be shared by different subsets of diseases and the area-specific values of these components as well as the relative contribution of each relevant disease to the common component may be estimated, as it will be shown in the next two sections.

3 Bayesian factor model

This section recalls classical and Bayesian factor analysis, in order to introduce the model developed in Rowe (2003a) and to simplify the illustration of our approach in the following section. Factor analysis model is illustrated by first assuming that for each of many subjects several variables (p) are observed. Let x_i denote the p-vector of observations on subject i. For any specified positive integer $m \leq p$, the standard m-factor model relates each x_i to an underlying m-vector of random variables f_i, the common factors, via:

$$(x_i | \mu, \Lambda, f_i, m) = \underset{(p \times 1)}{\mu} + \underset{(p \times m)}{\Lambda} \underset{(m \times 1)}{f_i} + \underset{(p \times 1)}{\varepsilon_i}. \tag{1}$$

μ is a p-dimensional unobserved population mean vector, Λ the $p \times m$ matrix of unobserved constants called the factor loadings matrix, f_i a m-dimensional vector of

unobservable "common" factor scores for the i-th subject, and ε_i a p-dimensional vector of "specific" factors or disturbance terms of i-th subject on p variables.

The parameters (μ, Λ, f) in the model are unknown and thus require estimation. Generally, the estimate of the population mean μ is easily found by maximum likelihood and coincides with sample mean, see Lawley (1940). From now on, to simplify calculations but without lack of generality, vector x_i is assumed to be zero mean (Press and Shigemasu 1989, 1997; Lopes and West 2004; Rowe 2003a).

The classical factor analysis model assumes $\varepsilon_i \sim N(0, \Psi)$ and Ψ a diagonal covariance matrix ($p \times p$), implying that, conditional on the common factors, the observable variables are uncorrelated: hence the common factors explain all the dependence structure among the p variables. By permitting ε to have non-diagonal covariance matrix, the Bayesian model accounts for the possibility of specification error; as for example, omission of one or more factors from the model.

In practical problems, especially with larger values of p, the number of factors m is often small relative to p. The factor scores f can be considered either random vectors or non-random vectors. The main problem encountered is that the model is overparameterized, the likelihood does not have a maximum, and we cannot reach maximum likelihood estimates through differentiating the log likelihood function. By assuming the factor scores not fixed, but random normally distributed variables with mean 0, standard deviation 1 and correlation R, independent from error random variables ε_i, it is possible to overcome unidentifiability problem. The variance and covariance matrix of observed vectors can be written as $Var(x_i|\Lambda, \Psi, m) = \Lambda R \Lambda' + \Psi$ and estimated by sample variance $\hat{\Sigma}$. Since the model is invariant under transformations of the form $\Lambda^\star = \Lambda P'$ and $f_i^\star = P f_i$; where P is any orthogonal $p \times p$ matrix, different proposals for identifying the model exist. We can get unique solutions by adding constraints on the parameters (as for example R being the identity and Λ being columnwise orthogonal).

Under assumption of centered, uncorrelated and standardized factors, on one hand the conditional structure is $x_i|f_i, \Lambda, \Psi \sim N(0, \Psi)$ and, on the other hand, the unconditional structure is $x_i|\Lambda, \Psi \sim N(0, \Lambda\Lambda' + \Psi)$. Therefore the variance can be decomposed in the following way: $var(x_{ih}|f, \Lambda, \Psi) = \psi_{hh}^2$, $cov(x_{ih}, x_{ik}|f, \Lambda, \Psi) = 0$, $var(x_{ih}|\Lambda, \Psi) = \sum_{j=1}^m \lambda_{hj}^2 + \psi_{hh}^2$ and $cov(x_{ih}, x_{ik}|\Lambda, \Psi) = \sum_{l=1}^m \lambda_{lh}\lambda_{lk}$. Each element ψ_{hh}^2 measures the residual variability of each variable once that contributed by the factors is accounted for. The element in the j-th row and k-th column is the covariance between the j-th variable and k-th score. Thus a large element of Λ implies a strong correlation between the corresponding variable and the corresponding factor. Aguilar and West (2000) introduced a reasonable assumption, that is to constrain Λ a block lower triangular matrix, assumed to be of full rank, with diagonal elements strictly positive. This form provides both identification and, often, useful interpretation of the factor model. In this form, the loadings matrix has $r = pm - m(m-1)/2$ free parameters.

One of the major challenges and more discussed topics in factor analysis is the determination of parameter m, the number of factors. Generally the researcher has to compromise between two risks: on one hand, underextraction can lead to the loss of relevant information and a substantial distortion in the solution; on the other hand,

overextraction can lead to factors with few substantial loadings, which can be difficult to interpret and to replicate. A general suggestion could be to "guess" the number of factors m based on underlying theory and previous studies; many different rules exist for the choice of number of factors in the non-Bayesian literature (an examples are the scree test or the percent variation), while Bayesian inference allows a probabilistic approach (Lopes and West 2004).

The Bayesian approach to factor analysis model bears directly on the problem of parameter identification, by incorporating proper prior information. If subjective information is introduced through the class of priors suggested, the usual rotational identification problem will disappear so that all the parameters of the factor loading matrix will be identified. We follow here the approach of Press and Shigemasu (1997), since it is considered so far the best and most complete reference on Bayesian factor analysis.

Starting from the model in (1), we obtain the likelihood by assuming:

$$\varepsilon_i \sim N(0, \Psi), \quad i = 1, \ldots, N.$$

$E(\Psi)$ is assumed to be diagonal to represent traditional beliefs of the model containing *common* and *specific* factors. While Lawley (1940) hypothesizes that the matrix is strictly diagonal, here the hypothesis is that a full positive definite diagonal matrix *on the expected value*. The likelihood for the observations can be written as the following matrix normal distribution (from now on, *trace* will be indicated with tr):

$$p(X|F, \Lambda, \Psi, m) \propto |\Psi|^{-N/2} \exp\left(-\frac{1}{2}\text{tr}\left((X - F\Lambda')\Psi^{-1}(X - F\Lambda')'\right)\right) \quad (2)$$

where the i-th row of X and F are, respectively, x_i' and f_i' as in (1). In Press and Shigemasu (1997) natural conjugate families of prior distributions are used. The factor loadings are assumed to depend on the disturbance covariance matrix, the disturbance covariance matrix is assumed to be independent of the factor scores and the factor scores are assumed to be independent of the factor loadings and the disturbance covariance matrix. More specifically, the joint prior distribution has the following form:

$$p(F, \Lambda, \Psi|m) = p(\Lambda|\Psi, m)p(\Psi)p(F|m),$$

moreover, we have the following set of prior distributions:

$$p(\Lambda|\Psi, m) \propto |\Psi|^{-m/2} \exp\left(-\frac{1}{2}\text{tr}\left(\Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'\right)\right), \quad (3)$$

$$p(\Psi) \propto |\Psi|^{-(\nu+2p+2)/2} \exp\left(-\frac{\nu}{2}\text{tr}\,\Psi^{-1}B\right), \quad (4)$$

$$p(F|m) \propto \exp\left(-\frac{1}{2}\text{tr}\,F'F\right). \quad (5)$$

with Ψ and H definite positive matrices, and B is a positive diagonal matrix. Parametrization in (4) is chosen such that $E(\Psi|B) = B$.

Straightforward posterior distributions are reached. More specifically, the factor scores given the factor loadings, the disturbance covariance matrix and the data, are normally distributed as:

$$p(F|\Lambda, \Psi, X, m) \propto \exp\left(-\frac{1}{2}\mathrm{tr}\left((F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})'\right)\right), \quad (6)$$

where $\tilde{F} = X\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$. The conditional posterior distribution of the factor loadings given the factor scores, the disturbance covariance matrix and the data, is a multivariate Gaussian distribution as follows:

$$p(\Lambda|F, \Psi, X, m) \propto \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1}(\Lambda - \tilde{\Lambda})(H + F'F)(\Lambda - \tilde{\Lambda})'\right)\right) \quad (7)$$

where $\tilde{\Lambda} = (X'F + \Lambda_0 H)(H + F'F)^{-1}$.

The conditional posterior distribution of the disturbance covariance matrix given the factor scores, the factor loadings and the data, is an inverted Wishart density:

$$p(\Psi|F, \Lambda, X, m) \propto |\Psi|^{-(N+m+\nu+2p+2)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1}U\right)\right) \quad (8)$$

where $U = (X - F\Lambda')'(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + \nu B$.

As expected in Bayesian inference, the expected values of the conditional posterior distributions (6), (7) and (8) are weighted means between prior expected values and contributions from likelihood. In particular, posterior expected value of Ψ is a combination of prior opinion, information from the data and prior opinion about Λ:

$$\begin{aligned}
E(\Psi|F, \Lambda, X, m) = {} & \frac{N}{N+m+\nu}\frac{(X - F\Lambda')'(X - F\Lambda')}{N} \\
& + \frac{m}{N+m+\nu}\frac{(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'}{m} + \frac{\nu}{N+m+\nu}B
\end{aligned} \quad (9)$$

As it emerges from this section, a Gibbs sampling algorithm is easily implementable (Rowe and Press 1998).

4 Model proposed

Classical factor model and Bayesian factor model assume independent rows of the observation matrix, this assumption turns out to be not a reasonable assumption for spatially correlated data. We try to remove the previous assumption starting from

Eq. (1) rewriting the $N \times p$ matrix of observations as a vector of length Np (where the first p elements are the p-observations relative to first region):

$$
\begin{array}{ccccc}
(X|\mu, \Lambda, F, m) = & I_N \otimes \Lambda & \mathrm{F} & + & \varepsilon \\
(Np \times 1) & (N \times N \otimes p \times m) & (Nm \times 1) & & (Np \times 1)
\end{array} \tag{10}
$$

According to the previous equation, F is a vector of length Nm and $\boldsymbol{\varepsilon}$ is now a vector of length Np and the following distribution is now assumed:

$$
\varepsilon \sim N(0, \Omega).
$$

The number of unknown parameters is greatly increased, that is the reason why Rowe (2003a) introduced the innovative idea of defining the matrix Ω as a separable covariance matrix by a Kronecker product of two matrices, $\Phi (N \times N)$ and $\Psi (p \times p)$. The great advantage of the previous definition is the interpretability of matrix Ω. Matrix Ψ is related to covariances between variables (conditional on the factor model, as a consequence, if the model is well defined we do expect Ψ to be diagonal). On the other hand, matrix Φ indicates the spatial structure, containing in each cell the covariance between regions. Furthermore, if we let Φ be proportional to the identity matrix, we have the model in (1).

Since matrix Ω is invariant under transformations of the form $\Phi' = k \times \Phi$ and $\Psi' = (1/k) \times \Psi$, where k is any positive constant, Kronecker product has problems with not uniqueness of solutions (Srivastava et al. 2008). A scalar parameter is introduced. Ω is now defined as $\sigma_\varepsilon^2 \Phi \otimes \Psi$ ($\otimes$ indicate Kronecker product), and matrices Φ and Ψ have traces, respectively, equal to N and p. If X is again a matrix, whose each row x_i contain the p-vector of observations on region i, under separable structure, $var(x_i|\Phi, \Psi, m, f, \Lambda) = \sigma_\varepsilon^2 \phi_{ii} \Psi$ and the covariance between rows i and j of X is $\sigma_\varepsilon^2 \phi_{ij} \Psi$, while the covariance between columns h and k of X (representing observations of variable h and k on the N regions) is $\sigma_\varepsilon^2 \psi_{hk} \Phi$.

Starting from model in Eq. (10), the likelihood for the observations can be written as the following matrix normal distribution

$$
p(X|F, \Lambda, \Psi, \Phi, m) \propto
$$
$$
\sigma_\varepsilon^{-Np} |\Phi|^{-p/2} |\Psi|^{-N/2} \exp\left(-\frac{\sigma_\varepsilon^{-2}}{2} \mathrm{tr}\left(\Psi^{-1}(X - F\Lambda')' \Phi^{-1}(X - F\Lambda')\right)\right), \tag{11}
$$

where again the i-th row of X and F are, respectively, x_i and f_i.

We will use natural conjugate families of prior distributions for the parameters. The joint prior distribution is given by

$$
p(\Phi, \Psi, \sigma_\varepsilon^2, F, \Lambda|m) = p(\Psi)p(\Phi)p(\sigma_\varepsilon^2)p(F|\Phi, m)p(\Lambda|\Psi, m).
$$

For the moment, we let unspecified prior distribution of Φ and we have the following set of prior distributions:

$$p(\sigma_\varepsilon^2) \propto \sigma_\varepsilon^{-2\kappa-2} \exp\left(-\frac{\theta}{\sigma_\varepsilon^2}\right), \tag{12}$$

$$p(\Lambda|\Psi, m) \propto |\Psi|^{-m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'\right)\right), \tag{13}$$

$$p(\tilde{\Psi}) \propto |\tilde{\Psi}|^{-(\nu+2p+2)/2} \exp\left(-\frac{\nu}{2}\mathrm{tr}\,\tilde{\Psi}^{-1}B\right), \tag{14}$$

$$\Psi = p\frac{\tilde{\Psi}}{\mathrm{tr}(\tilde{\Psi})} \tag{15}$$

$$p(F|\Phi, m) \propto |\Phi|^{-m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\,\Phi^{-1}FF'\right). \tag{16}$$

Again, Ψ and H are definite positive matrices, and B is a definite positive matrix. B is not necessary a diagonal matrix, even if, as explained in the previous section, assuming $E(\Psi|B)$ to be diagonal is a reasonable assumption. The introduction of the scalar parameter, σ_ε^2, allows to redefine Ψ in the following way:

$$\Psi = p \times \frac{\tilde{\Psi}}{\mathrm{tr}(\tilde{\Psi})}.$$

The trace of Ψ is now constrained to be equal to p. The distribution of Ψ is derived in Leorato and Mezzetti (2011) and turns to be again a conjugate distribution, with a great simplification of the computational aspect and the interpretation.

From the set of prior distributions just introduced, following considerations arise:

- Besides introduction of scalar parameter σ_ε^2, prior distributions for Ψ and Λ in (13) and (14) remain the same as in (4) and (3). Dependence between observations does not affect prior opinions about factor loadings and the within observations covariance matrix.
- Prior distribution for F in (16) is different from (5), since hypothesis of dependence between observations affects prior opinion about F. Vectorizing matrix F, we obtain a vector $Nm \times 1$, $\boldsymbol{F}$, whose covariance matrix is now the Kronecker product of Φ and an identity matrix (due to independence between factors).
- Prior distributions of Λ depends on Ψ and not on $\tilde{\Psi}$, since factor loadings are assumed to be standardized.

The idea underlying our proposal is that the factor scores, f_i, share the same correlation structure with the data. The spatial dependence between regions imply the same spatial dependence within the latent factors; in a way, our hypothesis is that factor scores are not explaining themselves all the spatial structure in the data. This latter consideration represents an important difference respect to other approaches to factor analysis for spatial data.

After definition of likelihood in (11) and the set of prior distributions in (12)–(16), an appropriate prior distribution for the matrix Φ needs to be defined. The ideal would

be to define a prior distribution able to leave posterior distributions analytically tractable. Three different scenarios for the matrix Φ are proposed and will be illustrated later.

Letting $p(\Phi)$ unspecified for the moment, conditional posterior distributions for F, Λ, σ_ε^2 and Ψ are reached in a straightforward way. More specifically, the joint posterior distribution for the unknown parameters of interest is given by:

$$p(F, \Lambda, \Psi, \Phi, \sigma_\varepsilon^2 | X, m) \propto p(\Phi) p(\sigma_\varepsilon^2) \sigma_\varepsilon^{-2Np} |\Phi|^{-(p+m)/2} |\Psi|^{-(N+m+v+2p+2)/2} |H|^{p/2}$$
$$\times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1} U\right)\right) \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Phi^{-1} FF'\right)\right)$$

where:

$$U = \sigma_\varepsilon^{-2}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + vB.$$

The conditional posterior density of the factor loadings given the factor scores, the disturbance covariance matrix and the data, is again multivariate normal:

$$p(\Lambda | F, \Psi, \Phi, X, m) \propto \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1}(\Lambda - \tilde{\Lambda})(H + F'\Phi^{-1}F)(\Lambda - \tilde{\Lambda})'\right)\right) \tag{17}$$

where $\tilde{\Lambda} = (X'\Phi^{-1}F + \Lambda_0 H)(H + F'\Phi^{-1}F)^{-1}$. Comparing (17) with (7), we notice that the introduction of dependence between observations does affect the posterior distribution of factor loadings through the incorporation of matrix Φ in covariance matrix of F.

The conditional posterior density of the disturbance covariance matrix given the factor scores, the factor loadings and the data, is again an inverted Wishart density:

$$p(\tilde{\Psi} | F, \Lambda, \Phi, X, m) \propto |\tilde{\Psi}|^{-(N+m+v+2p+2)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\tilde{\Psi}^{-1} U\right)\right). \tag{18}$$

$$E(\tilde{\Psi} | F, \Lambda, \Phi, \sigma_\varepsilon^2, X, m) = \frac{N}{N + m + v} \frac{\sigma_\varepsilon^{-2}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')}{N}$$
$$+ \frac{m}{N + m + v} \frac{(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'}{m} + \frac{v}{N + m + v} B \tag{19}$$

Again, the component deriving from the observed data contain matrix Φ, as it is also evident by the modification of posterior expected value respect to (9). The conditional posterior distribution for the factor scores given the correlation matrix, the disturbance covariance matrix, the factor loadings and the data is multivariate normal. Equation (6) is modified as follows:

$$p(F|\Lambda, \Psi, \Phi, X, m) \propto \exp\left(-\frac{1}{2}\text{tr}\left(\Phi^{-1}(F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})'\right)\right),$$

(20)

where $\tilde{F} = X\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$.

If we let $p\left(\sigma_\varepsilon^2\right)$, prior distribution for σ_ε^2, to be an inverse Gamma distribution with parameters (k, θ_0), posterior distribution for σ_ε^2 is again an inverse Gamma with updated parameters $\kappa + Np$ and $\theta_0 + \text{tr}\left(\Psi^{-1}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')\right)$

$$p(\sigma_\varepsilon^2 \mid F, \Lambda, \Psi, \Phi, X, m) \propto$$
$$\propto (\sigma_\varepsilon^2)^{-2(\kappa+Np+1)} \exp\left\{-\frac{\theta + \text{tr}\left(\Psi^{-1}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')\right)}{\sigma_\varepsilon^2}\right\}$$

(21)

The posterior expected value of σ_ε^2 results a weighted mean between prior expected value and maximum likelihood estimate:

$$E\left(\sigma_\varepsilon^2 \mid \cdots\right) = \frac{\kappa}{\kappa + Np}\frac{\theta}{\kappa} + \frac{Np}{\kappa + Np}\frac{\text{tr}\left(\Psi^{-1}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')\right)}{Np}$$

(22)

Here we propose three different scenarios for Φ, (posterior distributions will be discussed in the following Section):

First scenario: Φ is a constant matrix; as for example, correlation between regions is inversely proportional to any distance between the regions, or the covariance can rely on any other neighborhood structures.

Second scenario: Φ has *a priori* inverted Wishart distribution with expected value equal to a defined matrix G, and again similar to Eqs. (14) and (15), we assume that:

$$p(\tilde{\Phi}) \propto |\tilde{\Phi}|^{-(\gamma+2N+2)/2} \exp\left(-\frac{\gamma}{2}\text{tr}\left(\tilde{\Phi}^{-1}G\right)\right)$$

(23)

$$\Phi = N \times \frac{\tilde{\Phi}}{\text{tr}(\Phi)}$$

(24)

Introduction of scalar parameter σ_ε^2 affects also prior distribution of Φ, that justifies Eq. (24). We remove the assumption of independence between observations, but we do not specify any form of dependence and let the data determine that. If we let parameter γ small enough, we let prior distribution of Φ to be quite uninformative.

Third scenario: $\Phi = f(W, \rho)$ where f is a distance decay function, differentiable and such that Φ result positive definite, ρ is a vector of parameters, and W a fixed defined matrix. It is, first, necessary to define a matrix W; each cell w_{ij} represents a relation between regions i and j. W can be either an adjacency matrix or a "distance" matrix: in the first case w_{ij} is equal to 1 if only if regions

i and j are adjacent, otherwise $w_{ij} = 0$. In the second case w_{ij} indicates any distance between region i and region j. Distance can be defined as, for example, distance between barycenters of the two regions, or minimum distance between borders. The correlation between two adjacent regions depends on w_{ij} through the random parameters ρ. By proposing a prior distribution for ρ, (one or more dimension scalar parameter) another level on the hierarchical Bayesian model is introduced.

5 Computations

Gibbs Sampling algorithm is straightforward implemented to generate samples form (17)–(21). Starting with initial values for F and Ψ, for example $\tilde{F}_{(0)}$ and $\tilde{\Psi}_{(0)}$ and leaving unspecified from now Φ, then the cycle goes through:

$$\hat{\Lambda}_{i+1} = \text{a random sample from } P(\Lambda|\hat{F}_i, \hat{\Psi}_{(i)}, \hat{\Phi}_{(i)}, \hat{\sigma}^2_{\varepsilon(i)}, X)$$

$$\hat{F}_{i+1} = \text{a random sample from } P(F|\hat{\Lambda}_{i+1}, \hat{\Psi}_{(i)}, \hat{\Phi}_{(i)}, \hat{\sigma}^2_{\varepsilon(i)}, X)$$

$$\hat{\Psi}_{i+1} = \text{a random sample from } P(\Psi|\hat{\Lambda}_{i+1}, \hat{F}_{(i+1)}, \hat{\Phi}_{(i)}, \hat{\sigma}^2_{\varepsilon(i)}, X)$$

$$\hat{\sigma}^2_{\varepsilon(i+1)} = \text{a random sample from } P(\sigma^2_\varepsilon|\hat{\Lambda}_{i+1}, \hat{F}_{(i+1)}, \hat{\Psi}_{(i+1)}, \hat{\Phi}_{(i)}, X)$$

Finally, the means of the random sampling are the sampling based posterior marginal mean estimates of the parameters.

Identifiability is a well-known problem in factor analytic models and, to guarantee a unique solution, some suitable constraints must be imposed on the factor scores and factor loadings. As regarding F, we assume to have an identity variance matrix, as in (16). However, when using MCMC, which estimates the entire posterior distribution of the parameters and not just a (local) maximum, at each iteration step we look for a rotational solution, that maximizes the sum of the variances of the squared loadings.

Concerning the sampling from posterior distribution of Φ, we will show the three situations mentioned before.

First scenario The first scenario represents the easiest one from a computational point of view. Φ is assumed to be equal to a determined fixed matrix and it has not any prior distribution. On the other hand, the weight of the prior opinion is too strong.

Second scenario The second prior corresponds to the conjugate prior, so $\tilde{\Phi}$ has a posterior inverted Wishart distribution with updated parameters. The expected value is a weighted sum of data covariance, factor scores covariance and prior hypothesis about covariance between observations, more precisely:

$$E(\tilde{\Phi}|F, \Lambda, \Psi, \sigma^2_\varepsilon, X) = \frac{p}{p+m+\gamma} \frac{\sigma^{-2}_\varepsilon (X - F\Lambda')\Psi^{-1}(X - F\Lambda')'}{p}$$
$$+ \frac{m}{p+m+\gamma} \frac{FF'}{m} + \frac{\gamma}{p+m+\gamma} G \quad (25)$$

and again

$$\Phi = N \times \frac{\tilde{\Phi}}{\text{tr}(\tilde{\Phi})}$$

Third scenario The ideal would be to define the covariance as a distance decay function. Let, for example, Δ be a matrix whose elements represents any distance between two regions, and let matrix A be defined as $\exp(-\rho\Delta)$, where each element of A is the exponential of the element of Δ: $a_{ij} = \exp(-\rho\delta_{ij})$. The matrix A encounter some identification and estimation problem, since the matrix A is not positive definite for all possible values of ρ (Cressie 1993a,b; Anselin 2001). We first define Δ as the minimum distance between two regions, (we operate a shift such that only the diagonal elements are equal to zero). Let define Φ as a two parameters function: $\tau \exp(-\rho\Delta) + I_N$, in this way, for each couple of values (τ, ρ), closer two regions are, higher is the corresponding covariance. Fixing τ, when the value of ρ is increasing, matrix Φ is closer to the identity matrix; lower ρ is, the spatial structure is stronger. On the other hand, lower τ is, closer the matrix to the identity matrix.

We cannot solve analytically posterior distribution for ρ and τ, and sampling needs implementation of adaptive rejection Metropolis sampling following Gilks et al. (1995). Letting $\Phi = \tau \exp(-\rho\Delta) + I_N$

$$p(\tau, \rho|\Lambda, \Psi, m, X) \propto p(\tau, \rho)|\tau \exp(-\rho\Delta) + I_N|^{-(p+m)/2}$$
$$\times \exp\left(-\frac{1}{2}\text{tr}\left((\tau \exp(-\rho\Delta) + I_N)^{-1}\tilde{G}\right)\right) \tag{26}$$

where:

$$\tilde{G} = \sigma_\varepsilon^2(X - F\Lambda')\Psi^{-1}(X - F\Lambda')' + FF'.$$

As a prior distribution for ρ, Gamma distribution can be reasonably chosen, while as a prior for τ flat beta distribution is chosen. Gilks's is easily adapted to a bivariate distribution, fixing at each step one of the two parameters.

6 Results

The incidence rates of different cancers sites in Scotland are jointly modeled to explore the patterns of spatial correlation among them. Furthermore, the joint analysis is necessary to identify common latent factors, and to estimate how the incidence of each cancer site contribute to the shared risk factors. The incidence rates are considered

Table 1 Factor loadings for female cancer sites

Site	1st factor	2nd factor	3rd factor	4th factor	5th factor
Lips	0.0278	−0.0161	0.4283	0.2869	−0.0318
Oral	−0.0681	0.0982	0.2681	−0.2366	−0.0274
Esophagus	−0.0219	0.2925	0.1407	−0.0103	0.2792
Stomach	−0.0903	0.2914	−0.0314	0.2042	−0.0265
Large bowel	0.2363	0.1189	0.0073	0.1443	−0.1371
Pancreas	−0.2700	0.1093	−0.0409	−0.0047	−0.1698
Lung	−0.0265	0.3269	−0.0303	−0.0173	−0.1291
Breast	0.2090	0.1436	0.1128	−0.1379	−0.0812
Ovary	0.0528	−0.0440	0.0347	0.0151	−0.5654
Uterine	0.2685	−0.0254	0.3582	0.0086	0.0642
Cervix	−0.1205	−0.1834	0.1516	−0.0403	−0.1588
Hodking	−0.1296	0.0150	0.0922	0.3789	0.0444

on a standardized logarithmic scale to overcome the problem of inappropriateness of the normality assumption. Instead of a Poisson distribution, a Gaussian distribution is assumed and it represents an advantage since it allows to take into account the uncertainty attached to the incidence rates.

By the analysis of correlation coefficients between cancer incidences, it emerges that some of them result significantly high and some of them significantly low. Particularly, for each cancer site a positive association is present between males and females (a part from a zero between oral cancer in men and oral cancer in women). Furthermore, esophagus cancer is associated with stomach, and strongly with pancreas and lung cancer. A negative correlation between the previous ones and lips cancer emerges and a negative correlation between oral cancer and thyroid is present.

The results reported are based on posterior samples of 100,000 iterations per MCMC, with a *burn-in iterations* of 50,000, convergence for each parameters was checked by visual examination of trace plots and by the calculation of the Brooks and Gelman diagnostic (Brooks and Gelman 1998). Convergence of matrices was checked through convergence of eigenvalues. In terms of real time, the runs took between 4 and 7 h per model.

We run different models, we start now considering second prior distribution for Φ. The two matrices Φ and Ψ have both an inverted Wishart prior, their prior expected values are considered to be the identity matrix, letting scale parameters ν and γ to be small enough, as one tenth of matrix dimension, allows the prior distributions to be quite uninformative. In next section the choices for hyperparameters will be discussed. Applying different criterions, the number of factors considered relevant is three, applying the procedure stratified for sex and then overall. The mean of the posterior sampled factor loading matrices of females and males are shown, respectively, in Table 1 and in Table 2.

First factor loading in females is highly positive correlated with large bowel, breast, and uterine cancers, while it is negative correlated with pancreas cancer. Association

Table 2 Factor loadings for male cancer sites

Site	1st factor	2nd factor	3rd factor	4th factor	5th factor
Lips	−0.4616	0.0135	0.4637	0.0824	−0.0494
Oral	0.0118	0.1777	0.3435	0.0178	0.0960
Esophagus	−0.0320	0.0837	−0.0176	0.1286	−0.6238
Stomach	0.3377	−0.0323	−0.1766	0.1617	−0.1464
Large bowel	−0.0590	−0.0333	0.2090	0.5816	−0.0280
Pancreas	0.3451	−0.1682	0.1364	−0.1464	−0.0799
Lung	0.4477	0.1314	0.0150	0.1354	0.1086
Prostate	0.0394	−0.2688	0.3541	0.0678	−0.1280
Testis	−0.0929	−0.4275	−0.0408	0.0031	0.0243
Non Hodking	−0.0013	−0.2642	−0.1200	0.2397	0.2601

shown confirms, in a way, the epidemiological evidence of a link between ovulation and ovarian cancer ("the incessant ovulation theory" Fathalla (1971)), breast and reproductive variables, creating a strong link between breast and gynecological cancer. Breast and large bowel cancers can be hypothesized to share diet and alcohol consumption as a risk factor. First "female" factor explain 60% of the entire variance, and it can be defined as "a reproductive factor".

The second women factor, explaining 25% of total variation, is positively associated with esophagus, stomach and lung cancers. These sites share cigarette smoking and diet risk factor. The main interesting interpretation for second factor loading extracted can be linked to behavioral factors (smoking, alcohol and diet). The weaker negative association with cervix can be interpreted in terms of competing risk. The highest value for the second factor loading is the coefficient relative to *lung*, this confirm the possible interpretation of a latent factor score for smoking habit.

First factor loading in males signs the big contrast between lips cancer and, on the other side, lung, pancreas and stomach cancer. Scottish male lips cancer is very interesting since it is strongly spatially aggregated. Distribution of lips cancer has been studied by many statisticians for its strong spatial aggregation, still partially unexplained. Lips cancer is very common in the areas where high percentage of worker outside is present. As expected, strong association between lung, stomach and pancreas emerges, being the three of them related to smoking consumption. The first "male" factor, that can defined as "smoking factor", explain alone 70% of the entire variance. It is evident from Fig. 6 the strong spatial structure. Furthermore, it emerges a positive correlation with population density, assuming high values in metro areas. We can interpret first factor loading as signing contrast between urban and rural life. The negative association of the three tumors with lips cancer need to be further investigated; a possible explanation could be looked as a "confounder" effect: lips cancer is negatively associated with "smoking" related cancers, since "smoking" related cancers are more common in urban regions, where lips cancer is more rare.

The second men factor loading is characterized by high correlation between testicular, prostate and non-Hogking disease. Results of ecological studies suggest that

prostate cancer is associated with a western lifestyle and in particular, diet that includes a high intake of fat and meat (Fritschi et al. 2007; Grönberg 2003). At the same time, testis cancer is much more common in developed contest, thus the second men factor loading can be interpreted as a "social" factor, and positively associated with cancer more present in lower socioeconomic group, assuming higher values in poorer regions. Many studies conducted in all Western countries have reported large social inequalities in health, using a variety of socioeconomic indicators and including both general health and disease-specific outcomes (Spadea et al. 2010).

Finally, third factor between men is characterized by strong positive association between lips and prostate cancer. Some recent studies (Fritschi et al. 2007; Grönberg 2003) shows that farming is related to prostate cancer, probably due to exposure to pesticides and fertilizers. Since farming implies working outside and having higher exposure to sunlight, this can explain the association between prostate and lips cancer.

Finally we look for factor loadings considering the 23 variables together, the analysis consider female and male cancers at the same time. Results are not shown here, first factor is definitely signed by a strong association between breast and prostate cancer. In fact, prostate cancer was found to occur at a higher rate in families with other cancers such as breast cancer. Carriers of the breast cancer (BRCA1) gene had twice the incidence of prostate cancer of noncarriers (Grönberg 2003). The percentage variance explained is 55%, lower than the fist factor obtained after stratification.

The main advantage of the second prior distribution for Φ is that it is not necessary to assume high correlation between adjacent regions. We adopt an uninformative prior, we let Φ be far from diagonal matrix as much as data determine this distance. By assuming an inverted Wishart distribution for Φ with a reasonable choice for hyperparameter γ, we let the data determine the "distance" between sampling covariance matrix and diagonal matrix. The data will determine weather there is independence between regions, within cancer incidences, and weather the possible dependence evidenced is due to geographical structure.

Third prior distribution for matrix Φ adds a level in the hierarchical model, Φ depends now on two parameters (ρ, τ) that represents the strength of the spatial structure through the relation $\tau \exp(-\rho\Delta) + I_N$, where Δ is the minimum distance between the points in the border of the two regions, but different proposal can be considered (such as adjacent matrix, for example). In Fig. 5 posterior estimation of Φ under the different scenarios is shown. Third prior distribution for Φ considers only the case of lack of independence between observations due to spatial structures (imposed by subjective choice of Δ). Moreover, it is computationally complicate, some constraints on the parameters' domain need to be added to allow sampling. Posterior distribution of the two parameters is not shown here, since it is considered hard and uninteresting to interpret respect to straightforward interpreting Φ, since it does not have a direct interpretation as a correlation coefficient. (Male coefficients result $\tau = 1.2$ and $\rho = 0.31$, female coefficients result $\tau = 0.85$ and $\rho = 0.4$). As shown in Fig. 5 spatial structure results higher in males than in females; this consideration can be achieved just by graphical comparison, more complicate is comparisons of parameters. Observing the diagonal elements in the four graph, heterogeneities between regions emerge, particularly for men. Third prior distribution for Φ cannot capture regional heterogeneities.
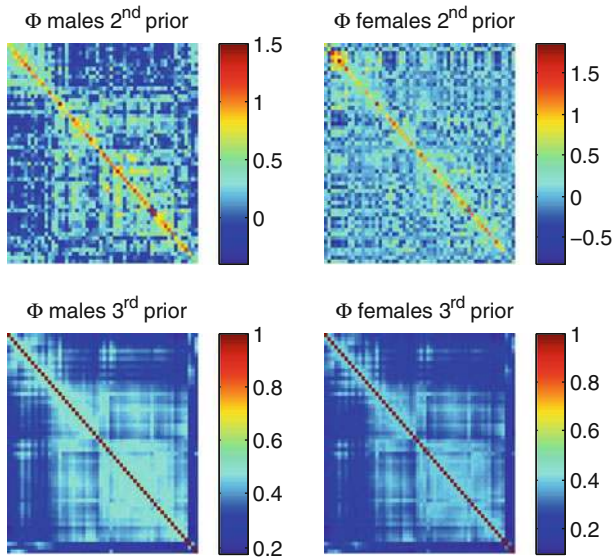
Fig. 5 Posterior estimates for Φ under different scenarios, between males and females
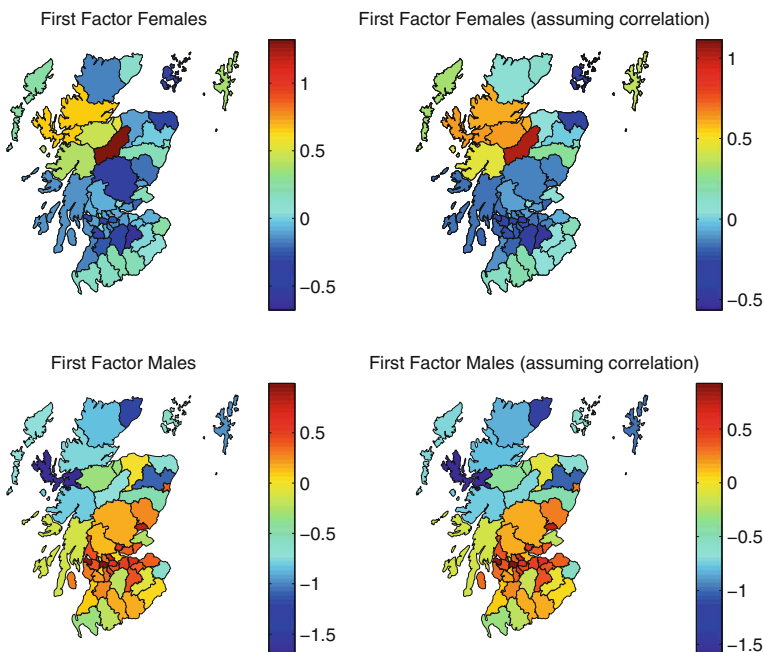

Fig. 6 First Factor Scores under different scenarios for Φ, between males and females

Critical is the choice of Λ_0, a prior expected value for Λ. Following Rowe (2003a), we first obtain maximum likelihood estimation from covariance method obtained with traditional factor analysis followed by *varimax* rotation on a small subset of observa-

tions, this estimate is used as a value for Λ_0, leaving the prior precision quite small. Regarding posterior estimation of Λ, as expected it is not robust against the change of Λ_0 but it is robust to the choice of training data. Furthermore, we investigate the behavior of the posterior estimation of Λ by changing prior distribution of Φ, the main question is whether different latent factors arise by strongly supporting presence of spatial structure in the data. An other important question arising is whether the posterior estimation of Λ change from other proposed model. After exploring different situations, we conclude that difference between maximum likelihood estimation and our estimation of Λ is as big as is the distance between posterior estimates of Φ and the identity. When prior opinion about existence of correlation between adjacent regions is strongly assumed, first factor between females is no longer correlated with breast cancer (that does not present spatial structures) and first factor between males is more strongly correlated with lips and lung cancer. In Fig. 6, comparison of first factor by assuming second and third prior is shown both within males and within females. Since spatial correlation is stronger between males than between females: first factor extracted from male cancer sites remains the same regarding the prior distribution assumed for Φ, and, on the other hand, first factor extracted from female cancer sites changes regarding the prior assumed. From Fig. 6, greater change we observe within females, since spatial structures does not result unless strongly assuming it *a priori*. The big contrast between lips cancer and other disease is pointed out only when spatial correlation is taken into account, and further investigation of this result should follow. Interesting is the change in the percentage of variation explained by the factors. First "male" factor presents a strong correlation between adjacent regions with any prior assumed; its relative contribution is higher with third prior. On the hand, contribution of first "female" factor is higher with first prior, since spatial correlation is weaker respect to men.

The interpretation of negative loadings may create problems. The negative sign does not mean that a disease-related factor for one cancer location is preventive for some other cancer location. The explantation can be given thinking about competing risks, the presence of some risk factor, i.e. pollution, can be the cause of respiratory related disease hiding the risk of other cancers.

An important issue to be faced is the determination of the number of factors. We select the number of factors by empirical methods as percent variation: the resulting chosen number of factors is the minimum number that accounts for at least that amount of total variation in the observed covariance matrix. We compare our results with the ones obtained through a Bayesian approach. Defining $p(m)$, a prior on m, easily by Bayes' Rule it is possible to compute the probability of each of the number of factors given the parameters

$$
\begin{aligned}
p(m \mid &\mu, \Lambda, F, \Psi, \Phi, \sigma_\varepsilon^2, X) \\
&\sim p(m)p(\mu)p(\Lambda|F, \Psi, m)p(F|\Phi, m)p(\Psi)P(\Phi)p(\sigma_\varepsilon^2)p(X|\mu, \Lambda, F, \Psi, m)
\end{aligned}
\tag{27}
$$

and determine the number of factors as the most probable. In this case, the results of the two approaches coincide, so we did not investigate further the latter mentioned

method, although a probabilistic approach to the determination of the number of factors deserves to be better developed. Christensen and Amemiya (2003) discuss a systematic method for selecting an appropriate model, in terms of number of factors, combining the Bayes factor with the joint Bonferroni test for spatially independent errors. Reversible Jump MCMC (RJMCMC) methods are useful for exploring posterior distributions for model parameters in the context of uncertainty about m, and with m included as a parameter. As we move between models with different numbers of factors, the dimension and meaning of the model parameters change, and RJMCMC methods are designed for just such problems.

7 Sensitivity analysis

Once the Bayesian model is described, the assessment of the hyperparameters has to be considered. The process of hyperparameter assessment is described in Rowe (2000, 2002, 2003a,b), while more sophisticated methods are proposed in Hayashi and Sen (2001).

We need to define the hyperparameters in Eqs. (12)–(16). Lee and Press (1998) concluded that the estimation of Λ is not robust against change of values for hyperparameters Λ_0. Maximum likelihood estimation from covariance method obtained with traditional factor analysis over a subsample will be assumed as a prior mean for Λ. By definition, H is any positive definite matrix, it is assumed that H is either a diagonal matrix or proportional to the identity. Assuming $H = n_H I_m$, as a value for n_H, the training sample size is chosen. There is no reason to assume a prior correlation between elements in Λ.

We assess the hyperparameter ν, the prior degrees of freedom by a method due to Hayashi and Sen (2001). We start with the Bayes estimator for the disturbance covariance matrix

$$E(\tilde{\Psi}|F, \Lambda, \Phi, \sigma_\varepsilon^2, X, m) = \frac{N}{N+m+\nu} \frac{\sigma_\varepsilon^{-2}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')}{N}$$
$$+ \frac{m}{N+m+\nu} \frac{(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'}{m} + \frac{\nu}{N+m+\nu} B \tag{28}$$

We can consider $\hat{\Psi}$ as a weighted average of the three terms in (28). The scalar values associated with the terms are N, m and ν respectively. Because we consider the first and third terms as representing respectively the data and the prior information, we can choose ν as greater as much strong is our prior opinion. Considering ν equal one tenth of the sample size is a reasonable choice.

Regarding B, a diagonal matrix is proposed. The expected value of any diagonal element is

$$E(\tilde{\psi}_{ii}) = b_{0i}, \quad i = 1, \ldots, p.$$

In case we substitute the training sample covariance matrix $\hat{\Sigma}$ and the *a priori* mean for the factor loadings in the above equations we have

$$\Psi_0 = \hat{\Sigma} - \Lambda_0 \Lambda_0',$$

then taking the average of the diagonal elements

$$\frac{1}{p}\mathrm{tr}(\Psi_0) = \frac{1}{p}\mathrm{tr}(\hat{\Sigma} - \Lambda_0 \Lambda_0').$$

Sensitivity analysis studies show that if ν and γ are both assumed to be equal to approximately one tenth of matrices dimension, posterior estimates are not sensitive to choice of B and G. Regarding posterior distribution of σ_ε^2 it results strongly sensitive to the parameters γ and ν, decreasing they are, increasing is the posterior sample mean of σ_ε^2. Lower value for k, results in not informative prior for σ_ε^2 and this turns out to be a reasonable choice. Finally considering third prior for Φ, results are robust to change of prior distribution of ρ and τ, assuming prior correlated distribution does not change the posterior estimations.

8 Conclusion

The spatial distribution of cancer mortality and morbidity rates has been studied by many authors. Since Burkitt (1969, 1970) stated that the variation in the geographical pattern of the disease may be related to its cause, and in view of this assertion many authors have studied geographic distribution of cancer. In order to examine Burkitt's assumption that the similar geographic distribution of different cancer sites may suggest the existence of a common etiological cause between them, a factor analysis seems to be the most appropriate multivariate techniques. From an epidemiological point of view, the current work helps to form a picture of the cancer burden in Scotland. The simultaneous study of multiple diseases figures out similar geographical trends of risk and it should be further strengthen the evidence for common sources that reflect underlying shared risk factors. Furthermore the evidence of strong within-region geographical heterogeneity should be further investigated.

Here Bayesian factor analysis model for spatially correlated data is proposed. The main hypothesis underlying the proposed model is that the latent factors share the same between-observations dependence as the observed variables. The correlations between variables within locations and the correlations across locations for each variable are hypothesized to be caused by the same latent spatial factor. Through a Bayesian approach, incorporation of our prior opinion about spatial pattern in disease risk is possible. The idea behind the proposed methodology is to look for latent factors explaining different distribution of the disease by defining a prior distribution for matrix Φ that attributes a high correlation to two regions, (or census tracts), that are close to each other. Assuming an inverted Wishart distribution for matrix Φ, the model proposed does not force the correlation within variables to be necessarily determined by geographic distance, but let the data determine the nature of the lack of independence. Our resulting method borrows information either across geographical areas and

across different cancer sites. The Bayesian method eliminates an important part of the variability unrelated to the true underlying cancer incidence rate by considering the information contained in other cancer sites and areas incidence rates.

The main novelty is the development of the factor model for spatially correlated data motivated by epidemiological data, in particular the innovation is twofold. On one hand the simultaneously analysis of multiple diseases, although not new (Knorr-Held and Best 2001; Dowing et al. 2008; Yanai et al. 1978; Tzala and Best 2008), it is not investigated in all its potentialities. The idea that other cancer sites incidence can provide information to better estimate a given cancer site incidence can be really innovative from a public health point of view. Analyzing together cancers not obviously related, such as for example female breast and male lung cancer, it is shown that can improve estimates. On the other hand, the approach follows Rowe (2003a) and Mezzetti and Billari (2005) where modeling structure is incorporated through the columns of the factor scores. Introduction of scalar parameter σ_ε^2 in (11) represents a novelty respect to model proposed in Rowe (2003a) and Mezzetti and Billari (2005). Its main advantage is to eliminate problem we can encounter multiplying by a constant the two matrices Ψ and Φ, an other alternatives could have been using correlation matrix instead of covariances matrix, but our proposal include also the case of heterogeneities of different regions.

The work can be extended in several directions. First, different spatial structures can be investigated. Also environmental or diet variables can be included, such that factor scores become a combination of cancer sites and corresponding risk factors. The current results contribute to reinforcing the hypothesis that environmental factors can act as a common risk factors for many types of cancers, and, since very likely adjacent regions share the same environmental factors, adjacent regions share the same cancer risk factors.

Finally, the proposed methods can be generalized to more complex models and a temporal structure can be introduced to allow the search of a common trend in the spread of the diseases. To this aim, since the flexibility of the Bayesian approach allows us to define different structures for Φ, the expression $\Phi = \Phi_T \otimes \Phi_S$, where Φ_T is defined as in Mezzetti and Billari (2005) for panel data, could be investigated and this model could enable simultaneous investigation of space-time variations in multiple health outcomes.

References

Aguilar O, West M (2000) Bayesian dynamic factor models and portfolio allocation. J Bus Econ Stat 18:338–357

Anselin L (2001) Rao's score test in spatial econometrics. J Stat Plan Inference 97:113–139

Anselin L (2007) Spatial econometrics in RSUE: retrospect and prospect. Reg Sci Urban Econ 37:450–456

Arminger G, Muthén BO (1998) A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the metropolis-hastings algorithm. Psychometrika 63:271–300

Besag J, York J, Mollie A (1991) Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43:1–20

Best N, Richardson S, Thomson A (2005) A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res 14:35–59

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25

Brooks S, Gelman A (1998) Alternative methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455

Burkitt DP (1969) Related disease-related cause?. Lancet 2:1229–1231

Burkitt DP (1970) Relationship as a clue to causation. Lancet 2:1237–1240

Cancer in Scotland (October 2010) Information Services Division, NHS, National Services Scotland

Christensen WF, Amemiya Y (2003) Modeling and prediction for multivariate spatial factor analysis. J Stat Plan Inference 115:543–564

Clayton D, Kaldor J (1987) Empirical bayes estimates of age- standardized relative risks for use in disease mapping. Biometrics 43:671–681

Cressie N (1993a) Regional mapping of incidence rates using spatial Bayesian models. Med Care 31:60–65

Cressie N (1993b) Statistics for spatial data. Wiley, New York

Cressie N, Wikle CK (2001) Statistics for spatio-temporal data. Wiley, New York

Cressie N, Calder K, Clark J, VerHoef J, Wikle CK (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol Appl 19:553–557

Diggle PG, Moyeed RA, Tawn JA (1992) Model-based geo-statistics. Appl Stat 47:299–350

Downing A, Forman D, Gilthorpe MS, Edwards KL, Manda SO (2008) Joint disease mapping using six cancers in the Yorkshire region of England. Int J Health Geogr 28:7–41

Fathalla MF (1971) Incessant ovulation-a factor in ovarian neoplasia?. Lancet 2(7716):163

Fritschi L, Glassm DC, Tabrizi JS, Leavy JE, Ambrosini GL (2007) Occupational risk factors for prostate cancer and benign prostatic hyperplasia: a case-control study in Western Australia. Occup Environ Med 64(1):60–65

Gilks WR, Best NG, Tan KKC (1995) Adaptive Rejection Metropolis Sampling within Gibbs Sampling. Appl Stat 44:455–472

Grönberg H (2003) Prostate cancer epidemiology. Lancet 361:859–864

Haining R, Grith D, Bennett R (1989) Maximum likelihood estimation with missing spatial data and with an application to remotely sensed data. Commun Stat Theory Methods 1875–1894

Hayashi K, Sen PK (2001) Bias-corrected estimator of factor loadings in Bayesian factor analysis. Educ Psychol Meas 62(6):944–959

Hogan JW, Tchernis R (2004) Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. J Am Stat Assoc 99(466):314–324

Journel AG (1983) Geostatistics. Encyclopedia of statistical sciences 3:424–431

Knorr-Held L, Best NG (2001) A shared component model for detecting joint and selective clustering of two diseases. J Roy Stat Soc Ser A 164:73–85

Lawley DN (1940) The estimation of factor loadings by the method of maximum likelihood. Proc Roy Soc Edinb 60:82–84

Lawson AB (2001) Statistical methods in spatial epidemiology. Wiley, New York

Le N, Sun W, Zidek J (1997) Bayesian multivariate spatial interpolater with data missing by design. J Roy Stat Soc Ser B 59:501–510

Lee SE, Press SJ (1998) Robustness of Bayesian factor analys estimates. Commun Stat Theory Methods 27:1871–1893

Leorato S, Mezzetti M (2011) Bayesian spatial panel data. Tecnical Report Università "Tor Vergata", Rome

Lopes H, West M (2004) Bayesian model assessment in factor analysis. Stat Sinica 14:41–67

Martin JL, McDonald RP (1975) Bayesian estimation in unrestricted factor analysis. A treatment for Heywood cases. Psychometrika 40:505–517

Mezzetti M, Billari FC (2005) Bayesian correlated factor analysis of socio-demographic indicators. Stat Methods Appl 14:223–241

Polasek W (1997) Factor analysis and outliers: a Bayesian approach. Discussion paper, University of Basel

Press SJ, Shigemasu K (1989) Bayesian inference in factor analysis In: Gleser L, Perlman M, Press SJ, Sampson A (eds) Contributions to probability and statistics: essays in honor of Ingram Olkin (Chap. 15). Springer, New York

Press SJ, Shigemasu K (1997) Bayesian inference in factor analysis-revised, with an appendix by Rowe, D.B. Technical report No. 243, Department of Statistics, University of California, Riverside

Rowe DB (2000) Factorization of separable and patterned covariance matrices for gibbs sampling. Monte Carlo Methods Appl 6(3):205–210

Rowe DB (2002) Jointly distributed mean and mixing coefficients for Bayesian source separation using MCMC and ICM. Monte Carlo Methods Appl 8(4):395–403

Rowe DB (2003a) Multivariate Bayesian statistics: models for source separation and signal unmixing. CRC Press, Boca Raton

Rowe DB (2003b) On using the sample mean in Bayesian factor analysis. J Interdiscip Math 6(3):319–329

Rowe DB, Press SJ (1998) Gibbs sampling and hill climbing in Bayesian factor analysis. Technical Report No. 255, Department of Statistics, University of California, Riverside

Scheiner SM, Gurevitch J (2001) Design and analysis of ecological experiments. 2. Oxford University Press, Oxford

Spadea T, Zengarini N, Kunst A, Zanetti R, Rosso S, Costa G (2010) Cancer risk in relationship to different indicators of adult socioeconomic position in Turin, Italy. Cancer Causes Control 21(7):1117–1130

Srivastava MS, von Rosen T, von Rosen D (2008) Models with a Kronecker product covariance structure. Estim Test Math Methods Stat 17(4):357–370

Tzala E, Best N (2008) Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. Stat Methods Med Res 17:97–118

Waller LA, Gotway CA (2004) Applied spatial statistics for public health data. Wiley-InterScience, New York

Waller LA, Carlin BP, Xia H, Gelfand AE (1997) Hierarchical spatio-temporal mapping of disease rates. J Am Stat Assoc 92:607–617

Wang F, Wall MM (2003) Generalized common spatial factor model. Biostatistics 4:569–582

Webster R, Oliver MA (2001) Geostatistics for environmental scientists. Wiley, New York

Wikle CK (2003) Hierarchical models in environmental science. Int Stat Rev 71:181–199

Yanai H, Inaba Y, Takagi H, Toyokawa H, Yamamoto S (1978) An epidemiological study on mortality rates of various cancer sites during 1958–1971 by means of factor analysis. Behaviormetrica 5:55–74

Yasui Y, Lele S (1997) A regression method for spatial disease rates: an estimating function approach. J Am Stat Assoc 92:21–32