

# Crime against women in India: unveiling spatial patterns and temporal trends of dowry deaths in the districts of Uttar Pradesh

G. Vicente, T. Goicoa and P. Fernandez-Rasines

*Public University of Navarre, Pamplona, Spain*

and M. D. Ugarte

*Public University of Navarre, Pamplona, and National University of Distance Learning, Pamplona, Spain*

[Received March 2018. Final revision October 2019]

**Summary.** Crimes against women in India have been continuously increasing lately as reported by the National Crime Records Bureau. Gender-based violence has become a serious issue to such an extent that it has been catalogued as a high impact health problem by the World Health Organization. However, there is a lack of spatiotemporal analyses to reveal a complete picture of the geographical and temporal patterns of crimes against women. We focus on analysing how the geographical pattern of 'dowry deaths' changes over time in the districts of Uttar Pradesh during the period 2001–2014. The study of the geographical distribution of dowry death incidence and its evolution over time aims to identify specific regions that exhibit high risks and to hypothesize on potential risk factors. We also look into different spatial priors and their effects on final risk estimates. Various priors for the hyperparameters are also reviewed. The risk estimates seem to be robust in terms of the spatial prior and hyperprior choices and final results highlight several districts with extreme risks of dowry death incidence. Statistically significant associations are also found between dowry deaths, sex ratio and some forms of overall crime.

**Keywords:** Gender-based violence; Integrated nested Laplace approximations; Sex ratio; Small areas; Spatial epidemiology; Spatiotemporal analysis

## 1. Introduction

Women are one of the most vulnerable groups in terms of violence (Powell and Wahidin, 2007). The United Nations General Assembly (1993) defined the term *violence against women* as any act of *gender violence* that results in physical, sexual or psychological harm to women. In 1993 also, the World Conference on Human Rights in Vienna recognized for the first time 'gender violence' as a violation of human rights. The World Health Organization has stated that violence against women is a 'public health problem of epidemic proportions' (Ellsberg *et al.*, 2005) with more than 1.6 million women losing their lives through domestic violence. However, investment in prevention is still inadequate (Garcia-Moreno and Watts, 2011). The World Health Organization (2013) also estimated that the global lifetime prevalence of intimate partner violence among ever-partnered women is 30.0% (with a 95% confidence interval of 27.8–32.2%), but this rate is higher in south-east Asia with an estimate of 37.7% (with a 95% confidence interval of 32.8–42.6%).

*Address for correspondence:* M. D. Ugarte, Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Campus de Arrosadia, 31006 Pamplona, Spain.  
E-mail: lola@unavarra.es

In many countries, cultural traditions are based on gender inequality, which in turn contributes to increasing violence against women when the post-colonial process failed to extend equality. This is so for India, one of the most populated countries in the world (around 1210 million people, according to the 2011 census), where crimes against women are on the rise and becoming a major issue. Gender-based violence is deeply institutionalized in India and it develops under the mantle of religious, cultural or social practices (Johnson *et al.*, 2007). Patriarchal precepts perpetuate gender inequality, as men and women are governed by different rules, thereby legitimizing violence against women (Russo and Pirlott, 2006). In India, the sex-gender system operates exposing girls and young women to different forms of violence including selective abortion and infanticide practices, harassment, rape, kidnapping, dowry homicide and murder, which preclude them from having a dignified life (Patel, 2015).

This paper focuses on dowry death: a form of crime against women which is very specific to India. Even though this form of violence is included in the Indian Penal Code, and the Dowry Prohibition Act (1961) proscribes any form of dowry, it remains a widespread practice in India, which is a country with the largest number of dowry deaths in the world (8455 deaths in 2014 according to the National Crime Records Bureau (2015)). Dowry is defined in the Dowry Prohibition Act (1961) as

‘any property or valuable security given or agreed to be given either directly or indirectly by one party to a marriage to the other party to the marriage, or by the parents of either party to a marriage or by any other person to either party to the marriage or to any other person; at or before or any time after the marriage in connection with the marriage of said parties’.

Whereas modern adoption of a dowry as the preferred marriage payment or agreement is a significant issue to be addressed, a different but subsequent issue revolves around its inflation (Banerjee, 1999). Dowries replaced bride price in the south (Bloch and Rao, 2002) but also became a cultural common practice transferred from high castes to middle and low castes throughout the country (Shenk, 2007). From a structural perspective, the dowry system relates to kin and property as a constructed and deployed rule to preserve social and economic advantage of socially powerful groups (Banerjee, 1999). Some researchers (see Shenk (2007) and the references therein) explain that, during the British Raj (1858–1947), weddings represented a large expenditure for peasants in north-west India, leading to female infanticide in that region. In addition to that, others claim that dowry-related violence can be used as an instrument to redistribute resources (Bloch and Rao, 2002). Awarding a large dowry to the groom’s family is a symbol of wealth and social status and, historically, was devised as a form of inheritance for daughters in a culture where women were not allowed to own immovable property (see, for example, Banerjee (2014)). Hence a dowry’s first purpose was to protect women from unfair traditions. However, it is unclear why it has become a means of extortion and an instrument of female exploitation.

Social anthropologists might agree that marriage in India is a virtually universal institution, mainly monogamous and rather restrictive. Although India is a secular state, the 2011 census showed nearly an 80% Hindu population where social classification perpetuates through a caste system. Assuming caste as a colonial translation for the Hindu term *varna* (order, type, colour or class), marriage in India reproduces class segmentation safeguarding caste endogamy. The Indian prevalent kinship system follows patrilocality, meaning a residence pattern where a bride is required to migrate to the groom’s clan location. The bride is consequently separated from her natal family at a very early age to enter a completely new scenario that eventually may lead to a greater subordination to her husband and in-laws. The 2001 Indian census showed that female migrants represented 218 million against 91 million for males (Khan, 2015).

Marriage in India is ruled by caste endogamy but also requires clan or lineage exogamy, especially in north India where marriage becomes a complex arrangement since matches are not to be found within the surrounding community but far from the natal village where the inner clan resides (Shenk, 2007; Prasad, 2016). For the case of north-west India, where our study is located, it is common that the bride's family pursues a groom of higher status, rank or age. This hypergamy, as anthropologists name it, might be considered either as a cause or a consequence of an even more restrictive marriage system. Some economists would speak of that in terms of 'marriage squeeze', as the main cause for dowry inflation, subsequent extortion, violence against young married women and deaths (Dang *et al.*, 2018; South *et al.*, 2014). Inflation can be also related to dowry dynamics tending to what some researchers have called the groom price, meaning that when desirable eligible men are scarce the competition between brides becomes fierce (Bloch and Rao, 2002).

Shenk (2007) viewed dowry prohibition as problematic and maintained that a major cause in effectiveness of the Indian dowry policy may be found in the lack of understanding between those who view a dowry as a social evil and those who view it as a form of positive investment. According to Srinivasan and Lee (2004)

'... the Indian dowry system should not be viewed simply as a traditional practice that will eventually be eliminated by processes of social change, but rather as an important component of a marriage system that is changing in response to a progressively more materialistic culture'.

Non-governmental organizations and academics also point out that the dowry system is related to discrimination against women, leading to female infanticide and sex-selective abortion preventing female births (see Banerjee (2014) and the references therein). Dowries are also related to violence against women by husbands and husbands' relatives, and this is a way of blackmailing the bride's family if the dowry seems unsatisfactory. Unfortunately, this form of violence against brides does not stop once the marriage has taken place but continues over time, leading to what is called a dowry death. Indian Penal Code 304B considers as dowry deaths those young women who died within 7 years of marriage, either murdered or driven to suicide by the husband and in-laws to force an amplified dowry (Mohanty *et al.*, 2013). A suicide committed by a woman who has suffered mental or physical violence in relation to her dowry is also regarded as a dowry death. The most common forms of dowry deaths are burning, drowning and poisoning a bride as they might easily be considered as accidents (see Banerjee (2014) and the references therein for more information about dowry and dowry violence).

Some studies have attempted to identify factors that are related to dowry deaths (see for example Jeyaseelan *et al.* (2015)) but their conclusions are rather limited. Regarding the sex ratio, females' education and their participation in the workforce, the results are even contradictory (Dang *et al.*, 2018; Belur *et al.*, 2014; Sharma *et al.*, 2002).

The paper aims to reveal spatial patterns for dowry death and how these patterns change over time, which in turn may be very helpful to identify unknown potential risk factors and to assess the effects of political or legal actions in time. Some district level covariates have been considered to assess their relationship with dowry deaths. We also look into different spatial priors and their effects on final risk estimates. To our knowledge, no spatiotemporal analyses have been performed on dowry deaths at district level in India yet. Here, we focus specifically on Uttar Pradesh, which is the most populated Indian state, with the highest rate of dowry deaths. The rest of the paper is organized as follows: Section 2 presents the data, provides descriptive figures and discusses some covariates that may be associated with dowry deaths. Section 3 reviews the methodology to be used in the real data analysis. Various spatial priors and several

distributions for the hyperparameters are also reviewed. In Section 4, we provide the results of the real data analysis. A discussion is given in Section 5.

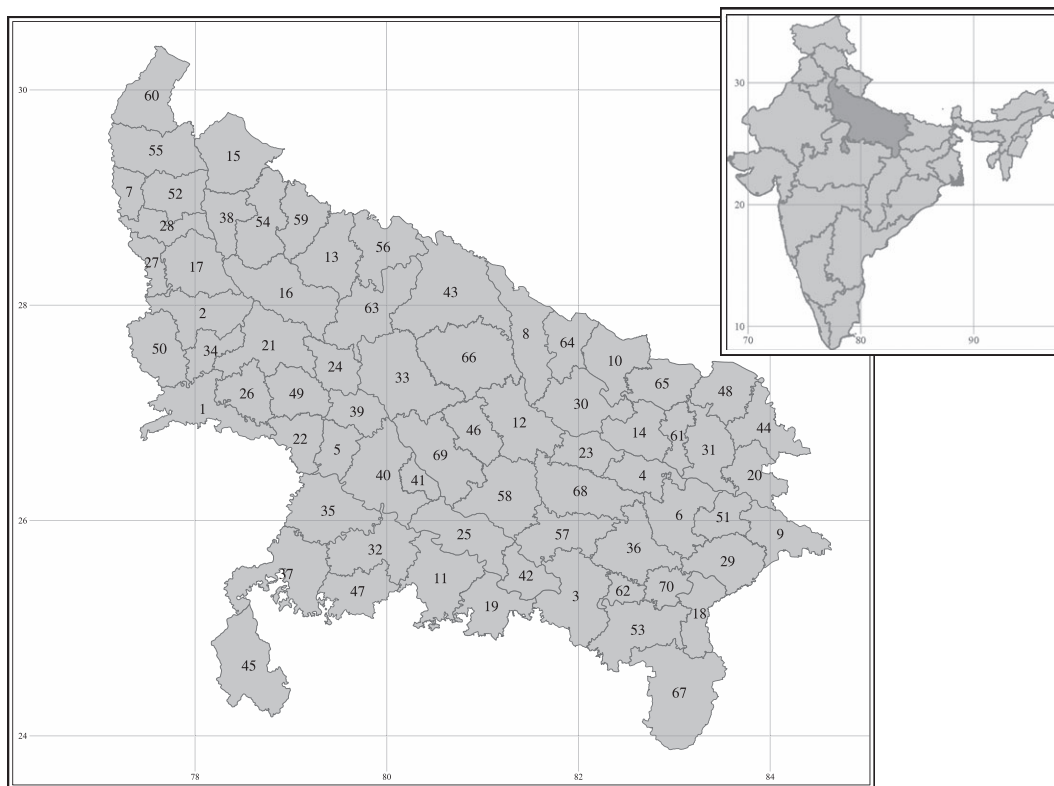
The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

<https://rss.onlinelibrary.wiley.com/hub/journal/1467985x/series-a-datasets>.

## 2. Dowry deaths in Uttar Pradesh

Dowry deaths are a deep-rooted form of violence against women in India. Here we focus on Uttar Pradesh, which is the most populous state in India, in the north of the country. It borders Nepal and Uttarakhand to the north, Madhya Pradesh and Chhattisgarh to the south, Bihar and Jharkhand to the east and Haryana, Delhi and the National Capital Region, and Rajasthan to the west. Nowadays, Uttar Pradesh is divided into 75 districts, but at the beginning of the study period there were 70 districts. We maintain the original 70 districts because data are not available for the current 75 districts during the years of the study. Fig. 1 displays a map of India with Uttar Pradesh shaded (in the top right-hand corner) and a map of Uttar Pradesh with the 70 districts (the main central map). The corresponding names of the districts can be found in Table 1.

The total female population of Uttar Pradesh was 95331831 (data from the 2011 census), but here we focus on 47282080 women between 15 and 49 years of age. In Uttar Pradesh, the male



**Fig. 1.** Map of India with Uttar Pradesh shaded (top right-hand corner), and districts of Uttar Pradesh (main central map)

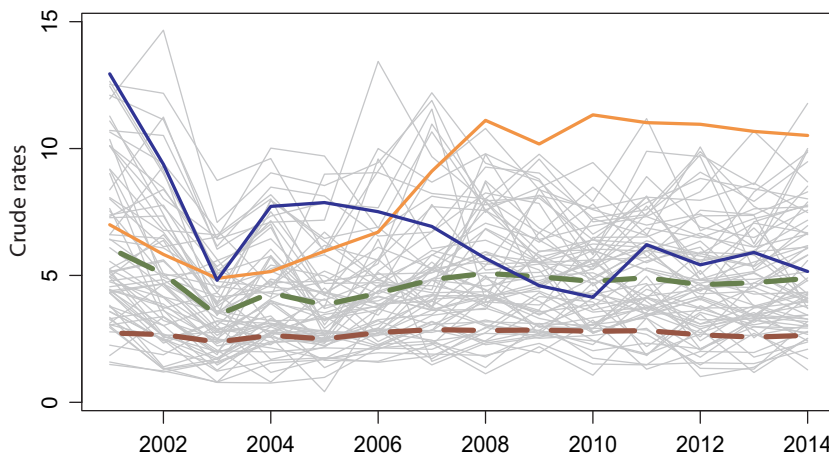


**Table 1.** District identifiers for Uttar Pradesh

Identifier	District	Identifier	District	Identifier	District
1	Agra	25	Fatehpur	47	Mahoba
2	Aligarh	26	Firozabad	48	Mahrajganj
3	Allahabad	27	Gautam Buddha Nagar	49	Mainpuri
4	Ambedkar Nagar	28	Ghaziabad	50	Mathura
5	Auraiya	29	Ghaziipur	51	Mau
6	Azamgarh	30	Gonda	52	Meerut
7	Baghpat	31	Gorakhpur	53	Mirzapur
8	Bahraich	32	Hamirpur	54	Moradabad
9	Ballia	33	Hardoi	55	Muzaffarnagar
10	Balrampur	34	Hathras	56	Pilibhit
11	Banda	35	Jalaun	57	Pratapgarh
12	Barabanki	36	Jaunpur	58	Rae Bareli
13	Bareilly	37	Jhansi	59	Rampur
14	Basti	38	Jyotiba Phule Nagar	60	Saharanpur
15	Bijnor	39	Kannauj	61	Sant Kabir Nagar
16	Budaun	40	Kanpur Dehat	62	Sant Ravidas Nagar Bhadohi
17	Bulandshahr	41	Kanpur Nagar	63	Shahjahanpur
18	Chandauli	42	Kaushambi	64	Shrawasti
19	Chitrakoot	43	Kheri	65	Siddharthnagar
20	Deoria	44	Kushinagar	66	Sitapur
21	Etah	45	Lalitpur	67	Sonbhadra
22	Etawah	46	Lucknow	68	Sultanpur
23	Faizabad			69	Unnao
24	Farrukhabad			70	Varanasi

literacy rate is 77.28%, whereas for females the figure is notably lower: 57.18%. Data on the number of dowry deaths have been obtained from the Indian National Crime Records Bureau (2015). Dowry deaths figures for districts 24 (Farrukhabad) and 40 (Kanpur Dehat) are not available and consequently they have been imputed from neighbouring areas. A total of 8455 dowry deaths were registered during 2014 in India. This represents 2.50% of all crimes against women in the country. The percentage of dowry deaths over the total crimes against women is three times higher in Uttar Pradesh: 6.42%, with 2469 registered cases. Furthermore, the total number of dowry deaths in Uttar Pradesh in 2014 represents 29.20% of all dowry deaths in India. According to various studies, a serious problem here is incorrect reporting and classification of dowry deaths, from a criminal and forensic perspective, when dealing with unnatural deaths especially related to fire-related injuries and kitchen accidents. In popular discourse, 'bride burning' and dowry death are synonymous (Sharma *et al.*, 2002; Belur *et al.*, 2014). Dang *et al.* (2018) argued that the police act as 'death brokers' when using culturally appropriated scripts to classify unnatural death of female victims within 7 years of marriage. Healthcare providers and police officers construct dowry deaths on the basis of the perception of the victim and the need to protect themselves from accusations of incompetence and corruption (Belur *et al.*, 2014). Other researchers have also highlighted parents' attitudes towards dowry deaths, as on many occasions the bride's parents may be reluctant to report cases and to demand prosecution to avoid social discredit that would ruin the possibilities of marriage for other daughters (see Verma *et al.* (2015)).

Fig. 2 shows the temporal evolution of crude rates (per 100000 women between 15 and 49 years old) in India (brown), Uttar Pradesh (green) and the districts of Uttar Pradesh (grey). Two districts are highlighted: Aligarh (orange) and Kheri (blue), which show increasing and



**Fig. 2.** Evolution of crude rates in India (— —), Uttar Pradesh (— —) and the districts Aligarh (—) and Kheri (—): —, crude rates for the remaining districts of Uttar Pradesh

**Table 2.** Descriptive statistics: minimum, first quartile  $Q_1$ , mean, standard deviation, third quartile  $Q_3$  and maximum dowry deaths in the districts of Uttar Pradesh by year†

Year	Minimum	$Q_1$	Mean	sd	$Q_3$	Maximum
2001	4 (Shrawasti)	18.00	31.57	19.00	43.75	88 (Kheri)
2002	3 (Shrawasti)	14.25	27.04	18.08	34.75	83 (Sitapur)
2003	3 (Balrampur)	10.25	18.89	11.49	24.00	55 (Agra)
2004	3 (Balrampur)	14.25	24.40	15.34	29.00	71 (Sitapur)
2005	1 (Lalitpur)	12.25	22.33	13.93	26.75	70 (Sitapur)
2006	7 (Sant Kabir Nagar)	14.25	25.69	14.37	34.75	67 (Kanpur Nagar)
2007	4 (Shrawasti)	16.00	29.63	17.28	36.75	78 (Agra)
2008	5 (Balrampur)	17.25	31.96	18.73	38.75	88 (Aligarh)
2009	8 (Lalitpur)	19.25	31.87	17.98	40.75	83 (Agra, Aligarh)
2010	5 (Balrampur)	18.25	31.44	19.67	40.00	95 (Aligarh)
2011	6 (Sant Kabir Nagar)	17.00	33.16	18.69	41.75	95 (Aligarh)
2012	5 (Balrampur)	19.00	32.03	17.88	40.75	97 (Aligarh)
2013	5 (Sant Kabir Nagar)	19.00	33.31	19.52	41.25	98 (Agra)
2014	6 (Sonbhadra)	23.25	35.26	18.39	46.75	98 (Aligarh)

†The names of the districts with the minimum and maximum number of dowry deaths in each year are displayed in parentheses.

decreasing trends respectively. The crude rates in Uttar Pradesh nearly double the crude rates in India, whereas the rates in Aligarth are about four times the crude rates in India from 2008 onwards. Clearly, most of the districts present crude rates that are higher than global rates for the whole of India.

Table 2 shows the minimum, first quartile, mean, standard deviation, third quartile and maximum of the number of dowry deaths in the districts of Uttar Pradesh by year. The names of the districts with the minimum and maximum number of dowry deaths in each year are displayed in parentheses. The number of registered deaths varies widely with minimum values between 1 and 7 and maximum values ranging from 55 to 98. In general the mean value increases from 2003 to 2008, and then it remains fairly stable until the end of the period, when it shows a slight increase.

Clearly, crude rates (see Fig. 2) are highly variable and models to smooth them are required. As the temporal evolution of rates by district is different, some trends increase (see for example Aligarh) whereas some others decrease (see for instance Kheri), models including space–time interactions are considered. In the next section, some spatiotemporal models are reviewed.

The phenomenon of dowry deaths in India is complex and multifaceted, and identifying potential risk factors could help to understand and combat this form of violence against women. However, it is unclear which socio-economic indicators or some other covariates may be associated with dowry deaths. South *et al.* (2014) found a surplus of male offenders correlated with gender violence. In the same way, and according to Mukherjee *et al.* (2001), overall crime is expected to be associated with crimes against women. Interestingly, Mukherjee *et al.* (2001) also found that the only crime that seems to be clearly associated with the sex ratio (defined as the female-to-male ratio) is dowry death, and the relationship is negative, i.e., the lower the sex ratio, the higher the rate of dowry deaths. South *et al.* (2014), Barakade (2012) and More *et al.* (2012) also hypothesize a negative relationship between sex ratio with dowry extortion and violence against women. Yet, the findings of Dang *et al.* (2018) show a reversed relationship when they approach a marriage squeeze as one of the main factors that might be driving dowry inflation in India, along with population growth and hypergamy. The literature also shows contradictory results on other potential risk factors, like female education and their participation in the workforce (Mukherjee *et al.*, 2001; Dang *et al.*, 2018; Belur *et al.*, 2014; Sharma *et al.*, 2002). As dowry deaths are usually registered as counts at area level (district, state), most of the covariates that have been contemplated to study this phenomenon are thought to operate at area level. In general, they are population-based variables that are available from the Indian Census Bureau, or other variables collecting information about some form of crime or deprivation, measured at area level and usually available from official registers.

In this paper we consider several district level covariates to evaluate their association with dowry deaths (see the on-line supplementary material for a description). We consider the sex ratio  $x_1$ , population density  $x_2$  and female literacy rate  $x_3$  as population-based variables obtained from the Indian Census Bureau. These variables are available for 2001 and 2011 only and we have imputed the values for the remaining years by using linear interpolation. These population-based variables are spatiotemporal, but the variation in time is not very pronounced, so they mainly capture spatial variability. Although the sex ratio is usually defined in other settings as the number of males per 1000 females, in this paper, it is defined as the number of females per 1000 males as this is the way that it is registered by the Indian Census Bureau. The sex ratio and female literacy rate have been previously studied at state level (see for example Mukherjee *et al.* (2001)) and the aim here is to study the relationship at district level in Uttar Pradesh.

In addition to these covariates we also consider a dummy variable indicating the political party of the Uttar Pradesh Chief Minister during the study period,  $x_0$ : the Bharatiya Janata Party, a centre–right-wing party (2001), the Bahujan Samaj Party, a party that defends equality and social justice (2002–2003 and 2007–2011) and the Samajwadi Party, of socialist ideology (2004–2006 and 2012–2014). Three more spatiotemporal variables are also considered: *per capita* income  $x_4$ , the number of murders per 100000 people,  $x_5$ , and the number of burglaries per 100000 people,  $x_6$ . Here burglary means entering a building or residence with the intention to commit a theft or other crime. These last two covariates have been included to assess relationships of dowry deaths with any other form of crime (Mukherjee *et al.*, 2001). Recently, sociologists and women’s organizations associate an increase in dowry deaths with an increase of consumerism in India (Verma *et al.*, 2015), and *per capita* income is included as a proxy.

As an initial descriptive approach, correlations between log-crude-rates of dowry deaths and the standardized covariates are provided in Table 3 in the census years 2001 and 2011. At first

**Table 3.** Correlations between log-crude-rates of dowry deaths and the covariates in the years 2001 and 2011

Year	$x_1$ , sex ratio	$x_2$ , population density	$x_3$ , female literacy rate	$x_4$ , per capita income	$x_5$ , murders	$x_6$ , burglaries
2001	-0.568	-0.289	0.274	0.206	0.608	0.278
2011	-0.721	-0.161	0.030	0.248	0.511	0.232

glance, the sex ratio  $x_1$  and number of murders per 100000 inhabitants,  $x_5$ , exhibit the strongest correlations, suggesting a potential negative association between sex ratio and dowry deaths, and a positive association between murders and dowry deaths (the spatial patterns standardized mortality ratios of dowry deaths, sex ratio and murders are shown in Fig. A.1 in the on-line supplementary material for the years 2001 and 2011). The correlation between female literacy rate and dowry deaths in 2011 is almost negligible. These preliminary results are in line with those in Mukherjee *et al.* (2001). Moderate to low correlations are observed for the rest of the variables. Summary statistics of correlations in all the years of the study period (which are not shown here to conserve space) reveal moderate to high correlations between sex ratio and dowry deaths and between murders and dowry deaths.

### 3. Spatiotemporal models

Spatial and spatiotemporal models for count data have been widely applied to study geographical and temporal patterns of incidence and mortality risks of many diseases, particularly cancer (see for example Aragonés *et al.* (2013), Mari-Dell'Olmo *et al.* (2014), Goicoa *et al.* (2016) or Ugarte *et al.* (2012)). However, and to the best of our knowledge, there has been only one attempt to reveal spatiotemporal patterns of crimes against women in India. More precisely, Vicente *et al.* (2018) studied the incidence of rape in Uttar Pradesh. Nowadays, the social awakening of concern for this problem and institutional support for victims have encouraged reporting of these crimes and hence the availability of records with data that can be further explored and investigated.

Models incorporating fixed effects, the main spatial and temporal random effects and spatiotemporal interactions are valuable tools to reveal geographical patterns of risk, global temporal trends and region-specific temporal trends. Geographic patterns may indicate the risk that is intrinsically related to a region, so specific traditions or cultural practices, for example, may be exerting some influence on gender-based violence. Global temporal trends indicate how risk evolves over time and they might be valuable to assess the effect of political actions, prevention or intervention measures, as well as specific national laws to protect women. The space-time interaction term describes the specific temporal evolution of the risk in individual areas, and it may capture the effect of regional programmes, or the effect of region-specific cultural practices over time. In general, random effects may be proxies of unobserved or unknown spatial, temporal or spatiotemporal covariates that are associated with dowry deaths.

The pioneering work by Besag *et al.* (1991) introduced two random effects to model spatially structured and unstructured heterogeneity. The unstructured term considers independent random effects whereas the structured term is modelled with an intrinsic conditional auto-regressive (ICAR) prior, which is a subclass of the conditional auto-regressive (CAR) priors that were

introduced by Besag (1974). However, the data can provide information on only the sum of the two components, not about the two effects individually, and an identifiability issue arises (see, for example, Gelfand and Sahu (1999), MacNab (2011) or Banerjee *et al.* (2015), chapter 6). Hence other alternatives have been proposed to model both spatially structured and unstructured heterogeneity. The main characteristic of these proposals is that they consider only one spatial random effect and it is the covariance (or precision) matrix that separates the dependence structure. In this paper we focus on two of those proposals that have been and still are widely used in disease mapping: the prior of Leroux *et al.* (1999), LCAR, and the model of Dean *et al.* (2001), DCAR. These proposals deal with spatially structured and unstructured heterogeneity, although in a different way. Whereas LCAR splits the precision matrix into two terms, one dealing with spatial dependence and the other with unstructured heterogeneity, DCAR splits the covariance matrix into the structured and unstructured parts.

In this paper we consider spatiotemporal models with spatial and temporal main effects, and spatiotemporal interactions. We review LCAR and DCAR and evaluate their effect on the final risk estimates. We also study the scaled version of the DCAR model proposed by Riebler *et al.* (2016), who proposed this scaled version because the interpretation of the hyperparameter distribution does not depend on the graph. Moreover, they derived the so-called penalized complexity (PC) priors (Simpson *et al.*, 2017) for the hyperparameters. Here, we also study the effect of using different priors for the hyperparameters on the final inference that will be carried out by using the integrated nested Laplace approximation package INLA (Rue *et al.*, 2009).

### 3.1. Models' description

In this subsection we consider spatiotemporal models to study dowry death risks including fixed effects, main spatial and temporal effects, and space–time interactions. This class of models is sufficiently flexible to assess the effects of covariates, to describe spatial relationships between regions, to capture temporal trends that may be or may not be linear and to account for the idiosyncrasy of regions.

Assume that the study region, the Indian state of Uttar Pradesh in this paper, comprises  $S$  small areas or districts ( $S = 70$  here) labelled as  $i = 1, \dots, S$ . For each district, data on dowry deaths is available for  $T = 14$  years denoted by  $t = 1, \dots, T$ . In this setting, conditionally on the rate  $p_{it}$ , the number of dowry deaths  $O_{it}$  in district  $i$  and time period  $t$  is assumed to follow a Poisson distribution with mean  $\mu_{it} = n_{it} p_{it}$ , where  $n_{it}$  is the population at risk in area  $i$  and time period  $t$ , i.e.

$$\begin{aligned} O_{it} | p_{it} &\sim \text{Poisson}(\mu_{it}), \\ \log(\mu_{it}) &= \log(n_{it}) + \log(p_{it}). \end{aligned} \quad (1)$$

Note that, in this case, the specific region–time rate  $p_{it}$  and the relative risk, denoted as  $R_{it}$ , are related through the global rate  $R = \sum_i \sum_t O_{it} / \sum_i \sum_t n_{it}$ , i.e.  $p_{it} = RR_{it}$ . As pointed out by a reviewer, modelling the rate  $p_{it}$  instead of the relative risk  $R_{it}$  avoids using the data on both sides of model (1). Then, posterior samples for  $R_{it}$  can be easily obtained. Different proposals to model  $\log(p_{it})$  can be found in the literature, the simplest being models with temporal linear trends (see Bernardinelli *et al.* (1995)). Nevertheless when temporal trends are not linear these models may be very restrictive. Hence, we consider non-parametric models with different space–time interactions (Knorr-Held, 2000). The log-rate is modelled as

$$\log(p_{it}) = \alpha + \mathbf{x}'_{it} \boldsymbol{\beta} + \xi_i + \gamma_t + \delta_{it}, \quad (2)$$

where  $\alpha$  is the overall rate,  $\mathbf{x}'_{it} = (x_{0it}, x_{1it}, x_{2it}, x_{3it}, x_{4it}, x_{5it}, x_{6it})$  is the vector of covariates outlined

in Section 2,  $\beta$  is the vector of coefficients of the covariates or fixed effects,  $\xi_i$  and  $\gamma_t$  are the main spatial and temporal random effects capturing global spatial and temporal patterns that may be associated with unobserved and unknown covariates, and  $\delta_{it}$  is the spatiotemporal interaction random effect dealing with specific temporal trends in each district or changes in the global spatial pattern with time.

In this paper we consider different CAR-type priors for the vector of spatial random effects  $\xi = (\xi_1, \dots, \xi_S)'$ , namely, the prior of Leroux *et al.* (1999), LCAR, and the prior of Dean *et al.* (2001), DCAR.

- (a) *LCAR prior*: the LCAR prior assumes the following multivariate normal distribution for the vector of spatial random effects  $\xi$ ,

$$\xi \sim N(\mathbf{0}, \sigma_\xi^2 \mathbf{D}^{-1}),$$

where  $\sigma_\xi^2 \mathbf{D}^{-1}$  is the covariance matrix,  $\mathbf{D} = \lambda_\xi \mathbf{Q}_\xi + (1 - \lambda_\xi) \mathbf{I}_S$ ,  $\mathbf{I}_S$  is the  $S \times S$  identity matrix and  $\mathbf{Q}_\xi$  is the neighbourhood matrix defined by contiguity, i.e. two districts are neighbours if they share a common border. The diagonal elements are equal to the number of neighbours of each district, and non-diagonal elements  $(\mathbf{Q}_\xi)_{ij} = -1$  if districts  $i$  and  $j$  are neighbours and  $(\mathbf{Q}_\xi)_{ij} = 0$  otherwise. The spatial smoothing parameter  $\lambda_\xi$  takes values between 0 and 1.

- (b) *DCAR prior*: the DCAR prior assumes the following multivariate normal distribution for the vector of spatial random effects  $\xi$ ,

$$\xi \sim N(\mathbf{0}, \sigma_\xi^2 \mathbf{M}),$$

where  $\sigma_\xi^2 \mathbf{M}$  is the covariance matrix,  $\mathbf{M} = \lambda_\xi \mathbf{Q}_\xi^- + (1 - \lambda_\xi) \mathbf{I}_S$  and ‘ $-$ ’ denotes the Moore–Penrose generalized inverse. Riebler *et al.* (2016) proposed a modified version of this prior with a scaled version of the covariance matrix. They considered a scaled matrix  $\mathbf{Q}_{\xi*}^-$  such that the geometric mean of its diagonal elements is equal to 1. Using this scaled matrix, the priors for the hyperparameters have the same interpretation irrespective of the graph. This prior is implemented in the package INLA as BYM2 and this is the name that we adopt hereafter in the paper. The three priors, LCAR, DCAR and BYM2, lead to the well-known ICAR prior when  $\lambda_\xi = 1$ , and to an exchangeable prior when  $\lambda_\xi = 0$ . The key difference between the LCAR and the DCAR and BYM2 priors is that in LCAR  $\lambda_\xi$  enters in the precision matrix (and hence it is related to the conditional correlations) whereas in DCAR and BYM2  $\lambda_\xi$  enters in the covariance matrix (and it controls the marginal correlations). When all the variability is spatially structured, the ICAR prior would be a suitable prior; however, if this is unknown the other priors are more general.

For the vector of temporal random effects  $\gamma = (\gamma_1, \dots, \gamma_T)'$ , random walks of first or second order (RW1 and RW2 respectively) are considered, i.e.  $\gamma$  is assumed to follow a multivariate normal distribution,

$$\gamma \sim N(\mathbf{0}, \sigma_\gamma^2 \mathbf{R}_\gamma^-),$$

where  $\mathbf{R}_\gamma$  is the structure matrix (see for example Rue and Held (2005), pages 95 and 110).

Finally, the interaction random effect  $\delta = (\delta_{11}, \dots, \delta_{S1}, \dots, \delta_{1T}, \dots, \delta_{ST})'$  is assumed to follow a multivariate normal distribution,  $N(\mathbf{0}, \sigma_\delta^2 \mathbf{R}_\delta^-)$ , with structure matrix  $\mathbf{R}_\delta$  defined as the Kronecker product of the spatial and temporal structure matrices. Depending on the definition of the precision (or covariance matrix), four types of interaction can be defined (Knorr-Held, 2000): type I (independent interactions,  $\mathbf{R}_\delta = \mathbf{I}_T \otimes \mathbf{I}_S$ ); type II (structured in time and unstructured in

space,  $\mathbf{R}_\delta = \mathbf{R}_\gamma \otimes \mathbf{I}_S$ ); type III (structured in space and unstructured in time,  $\mathbf{R}_\delta = \mathbf{I}_T \otimes \mathbf{Q}_\xi$ ); type IV (structured in space and time,  $\mathbf{R}_\delta = \mathbf{R}_\gamma \otimes \mathbf{Q}_\xi$ ). It is also worth noting that these spatiotemporal models are not identifiable. The spatial and temporal random effects have implicitly defined an intercept, giving rise to an identifiability issue with the model intercept. In addition, some of the interaction effects are confounded with the main effects. In this paper, we fit the models by using appropriate constraints following Goicoa *et al.* (2018), who provided the set of constraints for these spatiotemporal models with the different types of interaction.

### 3.2. Hyperpriors

Given that models constitute a valuable tool to gain knowledge about the incidence of crime against women, it seems sensible to evaluate these models appropriately. This includes assessing the effects of different priors and hyperpriors on the final risk estimates. Here we consider different sets of hyperpriors for the hyperparameters  $\lambda$ ,  $\tau_\xi = 1/\sigma_\xi^2$  and  $\tau_\gamma = 1/\sigma_\gamma^2$ . On one hand, we consider a  $\text{Unif}(0, 1)$  prior for the spatial parameter  $\lambda$  and  $\log\text{Gamma}(1, 0.00005)$  priors on the log-precisions. However, as the log-gamma priors may produce wrong results (Carroll *et al.*, 2015), a uniform prior on the positive real line is considered for the standard deviation (Ugarte *et al.*, 2017; Goicoa *et al.*, 2019). This agrees with the recommendation from Gelman (2006) about using non-informative uniform priors on the standard deviations.

Recently, Simpson *et al.* (2017) proposed what they call PC priors. These priors cannot be derived for the LCAR model, but they can be obtained for the BYM2 model. Riebler *et al.* (2016) explained that the PC prior for the standard deviation is exponential with rate  $\theta$  and they chose the value of  $\theta$  according to  $P(\sigma_\xi > U) = \alpha$ , i.e.  $\theta = -\log(\alpha)/U$ . If the risks are assumed to be smaller than 2 with probability 0.99, then  $U = 1$  and  $\alpha = 0.01$ . These are the default values in INLA. Simpson *et al.* (2017) also provided a PC prior for  $\lambda$  in the BYM2 models, but this prior has no closed form. They proposed the rule  $P(\lambda < 0.5) = \frac{2}{3}$  favouring a simpler model without spatial variability unless the data support the more complex model.

## 4. Statistical analysis: results

The spatiotemporal model (2) has been fitted considering the LCAR, DCAR, BYM2 and ICAR ( $\lambda_\xi = 1$ ) priors for space, RW1 and RW2 priors for time, and the four types of interaction. Different model selection criteria have been proposed in the literature, e.g. the deviance information criterion DIC (Spiegelhalter *et al.*, 2002) or the ‘widely applicable information criterion’ WAIC (Watanabe 2010). A reviewer has pointed out that here the objective is local smoothing rather than goodness of fit and, hence, these model selection criteria may not be the most appropriate. For comparison we also consider additive models, i.e. model (2) without the spatiotemporal interaction random effects ( $\log(p_{it}) = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \xi_i + \gamma_t$ ) to show that a lack of fit may lead to wrong results. Additionally we also compute cross-validation measures. We use the logarithmic score LS (Gneiting and Raftery, 2007), which is a cross-validation measure based on the conditional predictive ordinate CPO (Pettit, 1990). The mean log-score is defined as  $-(1/ST)\sum_{i,t} \log(\text{CPO}_{it})$ , where  $\text{CPO}_{it} = p(O_{it}|O_{-it})$ ,  $O_{-it}$  is the set of observations excluding observation  $-it$  and  $p(\cdot|O_{-it})$  is the predictive distribution of a new observation given  $O_{-it}$ . So this is a criterion to select a model in terms of its predictive ability.

Originally, we fitted model (2) with the seven covariates that were introduced in Section 2. Uniform distributions on the positive real line were considered for the standard deviations in the ICAR and LCAR priors whereas default PC priors have been chosen for the DCAR and BYM2 priors. (Results with log-gamma priors on the log-precisions are not shown here as they

are nearly identical). The model selection criteria point towards a type II interaction model with an RW1 prior for time (which is not shown here to conserve space). Table S.1 in the on-line supplementary material displays the posterior means for the fixed effects, their standard deviations, the medians and 95% credible intervals obtained from a type II interaction model with an RW1 prior for time and an ICAR prior for space. Only sex ratio  $x_1$ , murders  $x_5$  and burglaries  $x_6$  have a significant effect. Then, we fitted model (2) with these three variables. The model selection criteria (DIC, WAIC and LS) for the complete set of models are displayed in Table S.2 in the on-line supplementary material. Clearly, type II interaction models with an RW1 prior for time are selected. Although all the model selection criteria undoubtedly point to type II interaction models with an RW1 prior for time, they do not show differences between the various spatial priors. DIC for the type II interaction model and RW1 for time is around 6171 for all spatial priors. Something similar happens with WAIC (around 6180) and LS (about 3.17).

In addition to model selection criteria, the probability integral transform (PIT) (Dawid, 1984) and its adjusted version for discrete data (Czado *et al.*, 2009) can be used as a diagnostic tool. The PIT is simply the value that the cumulative predictive distribution function attains at the observation. Deviations from uniformity in a PIT histogram point towards model deficiencies. U-shaped histograms mean underdispersed predictive distributions and hump-shaped distributions indicate overdispersed predictive distributions. Fig. 3 displays the PIT histograms for the additive model (Fig. 3(a)) and a model with type II interaction (Fig. 3(b)). The models include the three significant covariates and ICAR and RW1 priors for space and time respectively. Clearly, the additive model shows a U-shaped histogram, indicating an underdispersed predictive distribution. In contrast, the type II interaction model seems to behave reasonably well.

Fig. 4 shows the estimated log-relative-risks with the additive model (Fig. 4(a)) and with the type II interaction model (Fig. 4(b)) for six districts in Uttar Pradesh: Aligarh, Kanpur Dehat, Kheri, Shrawasti, Sitapur and Varanasi. The estimated risk trends with the additive model are all parallel (in fact this is the global trend shifted according to the estimated spatial effect  $\xi_i$ ). In contrast, the type II interaction model is more flexible and it allows different estimated risk trends for each district: something that seems more plausible in this study. (See also Fig. S.2 in the on-line supplementary material and the corresponding comments there.)

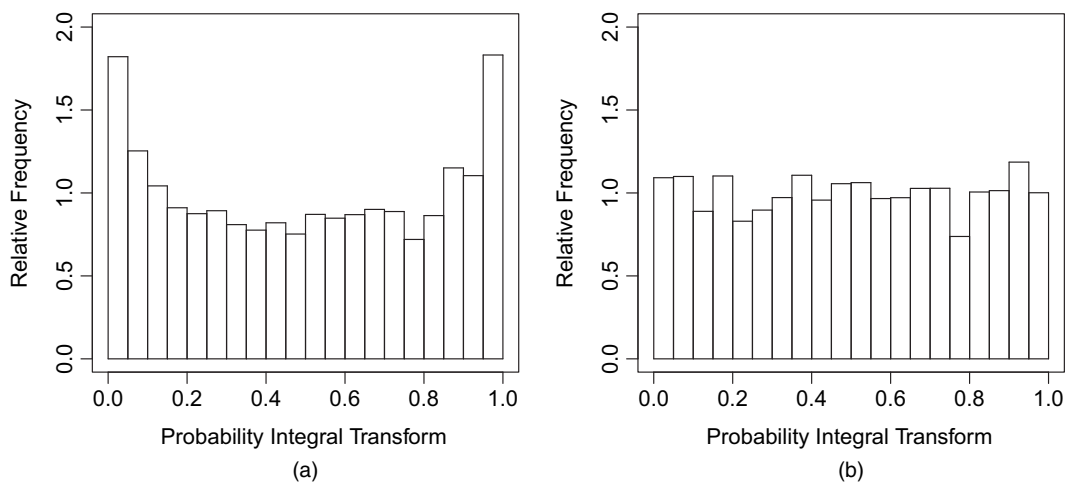
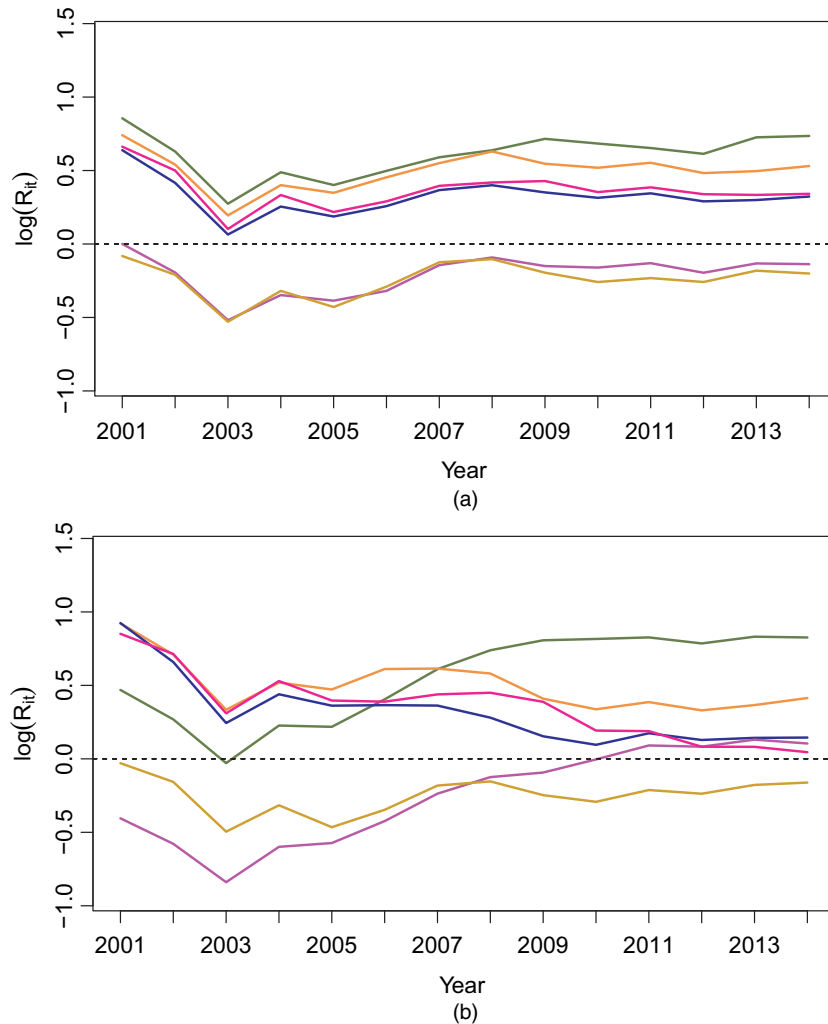


Fig. 3. PIT histograms for (a) the additive model and (b) the type II interaction model



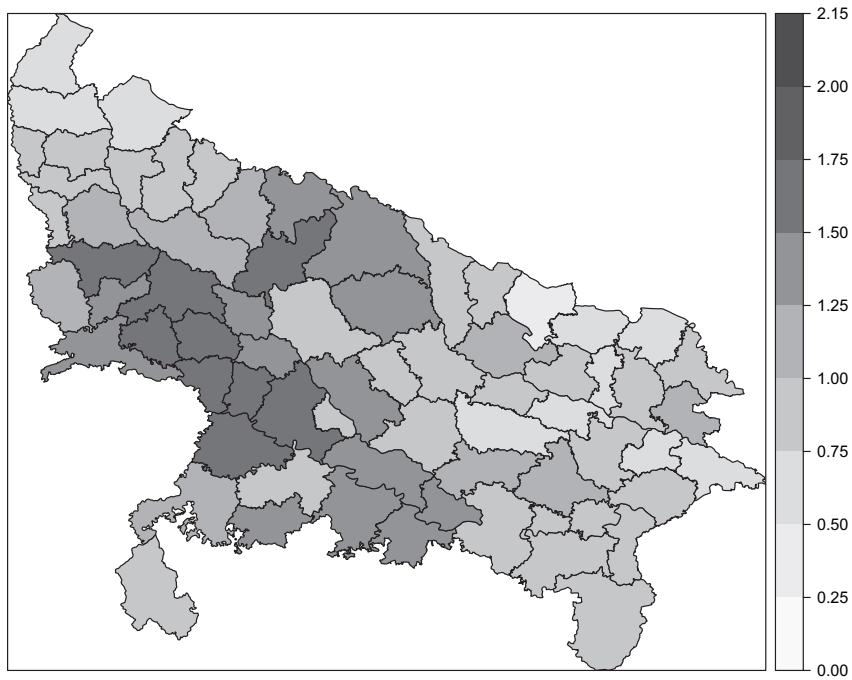


**Fig. 4.** Relative risk temporal trends for six selected districts in Uttar Pradesh, Aligarh (—), Kanpur Dehat (—), Kheri (—), Shrawasti (—), Sitapur (—) and Varanasi (—): (a) additive model and (b) type II interaction model

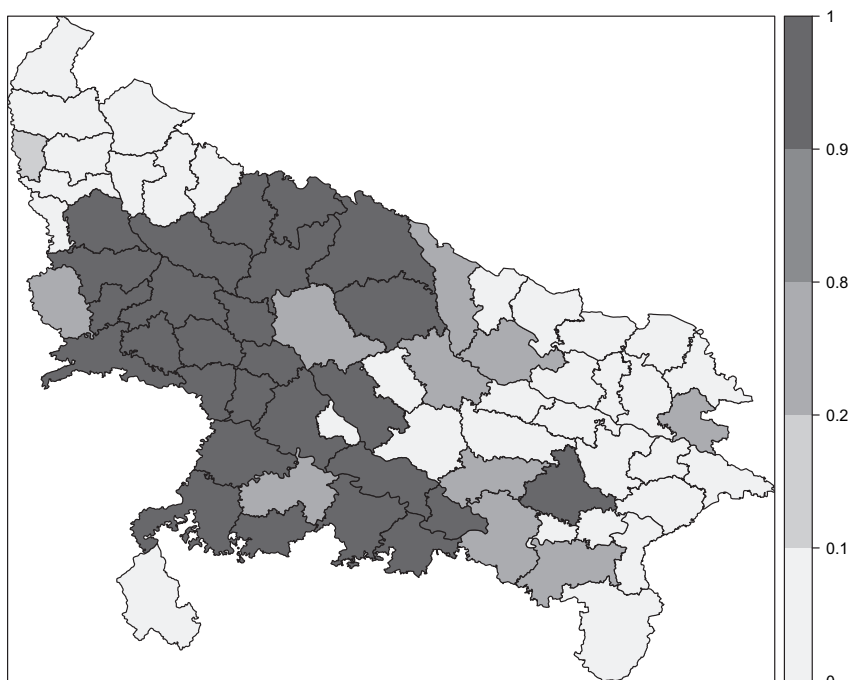
One reviewer suggested comparing the type II with the type IV interaction model. A dispersion plot of final risk estimates (which is not shown here) reveals similar estimates, and all model selection criteria indicate that the type II interaction model is preferred.

In what follows, and given that the ICAR spatial prior is the simplest model and there is no difference with the other spatial priors, we shall display the results for the type II interaction model with an ICAR spatial prior, an RW1 prior for the temporal effect and the three significant covariates. For comparison, Fig. S.3 in the on-line supplementary material displays the estimated relative risks with a type II interaction model with an ICAR spatial prior *versus* the same type II interaction model with LCAR, DCAR and BYM2 spatial priors. The estimated relative risks are identical.

Table 4 displays the posterior means, standard deviations and medians of the fixed effects together with 95% credible intervals. A negative association of dowry deaths with sex ratio and



(a)

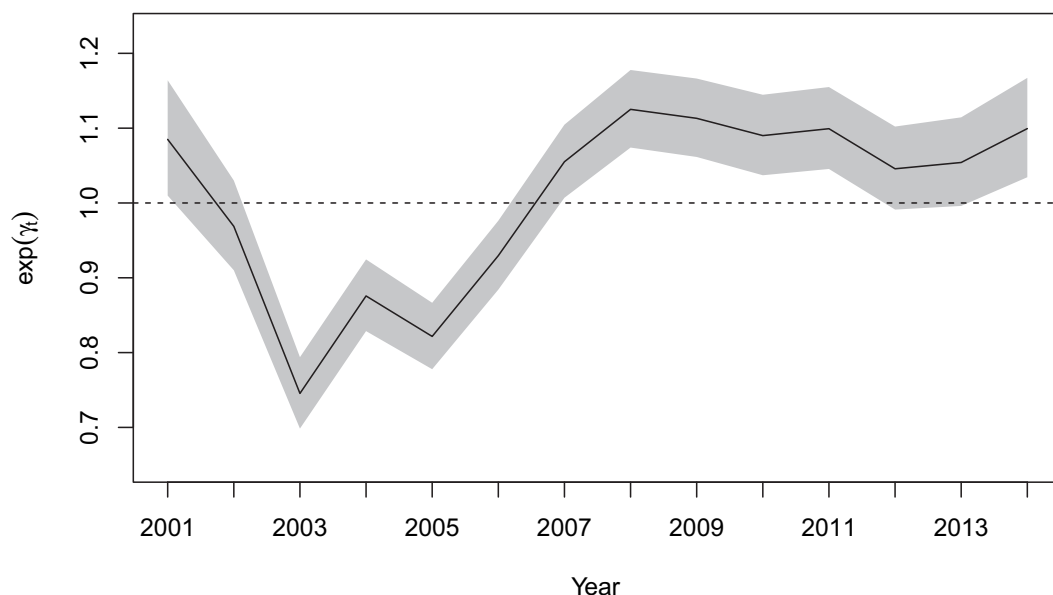


(b)

**Fig. 5.** (a) Posterior mean of the district-specific relative risk  $\zeta_j = \exp(\xi_j)$  and (b) posterior probabilities  $P(\zeta_j > 1 | \mathbf{O})$  that the relative risks are greater than 1

**Table 4.** Posterior means, standard deviations and medians of the fixed effects together with a 95% credible interval

Covariate	mean	sd	median	95% credible interval
$\alpha$	-10.026	0.007	-10.026	(-10.040, -10.013)
$x_1$ (sex ratio)	-0.094	0.046	-0.094	(-0.183, -0.004)
$x_5$ (murders)	0.089	0.020	0.090	(0.050, 0.129)
$x_6$ (burglaries)	0.054	0.016	0.054	(0.023, 0.084)

**Fig. 6.** Temporal pattern (posterior mean of  $\exp(\gamma_t)$ ) of dowry death risks in Uttar Pradesh

a positive association with murders and burglary are obtained. Consequently, areas with low sex ratio tend to have a high risk of dowry deaths and areas with high numbers of murders and burglaries also have a high risk of dowry deaths. These results agree with those in Mukherjee *et al.* (2001).

The spatiotemporal models that were described in Section 3.1, and consequently the fitted model, are very useful and interesting because, once the covariate effects have been accounted for, they enable us to split the final risk into different components with a valuable interpretation, namely an overall risk level given by  $R^{-1} \exp(\alpha)$  (or an overall rate level if we do not divide by  $R$ ), a spatial component  $\zeta_i = \exp(\xi_i)$  indicating the risk that is associated with each district (this may be capturing unknown or unobserved spatial covariates), a temporal component  $\exp(\gamma_t)$  that may capture the effect of unobserved temporal covariates such as long-term policies, intervention or prevention programmes over time and, finally, a spatiotemporal component  $\exp(\delta_{it})$  that is related to each district idiosyncrasy. In our case, after the effect of sex ratio, murders and burglaries has been discounted, most of the variability is explained by the spatial effect: 78% of

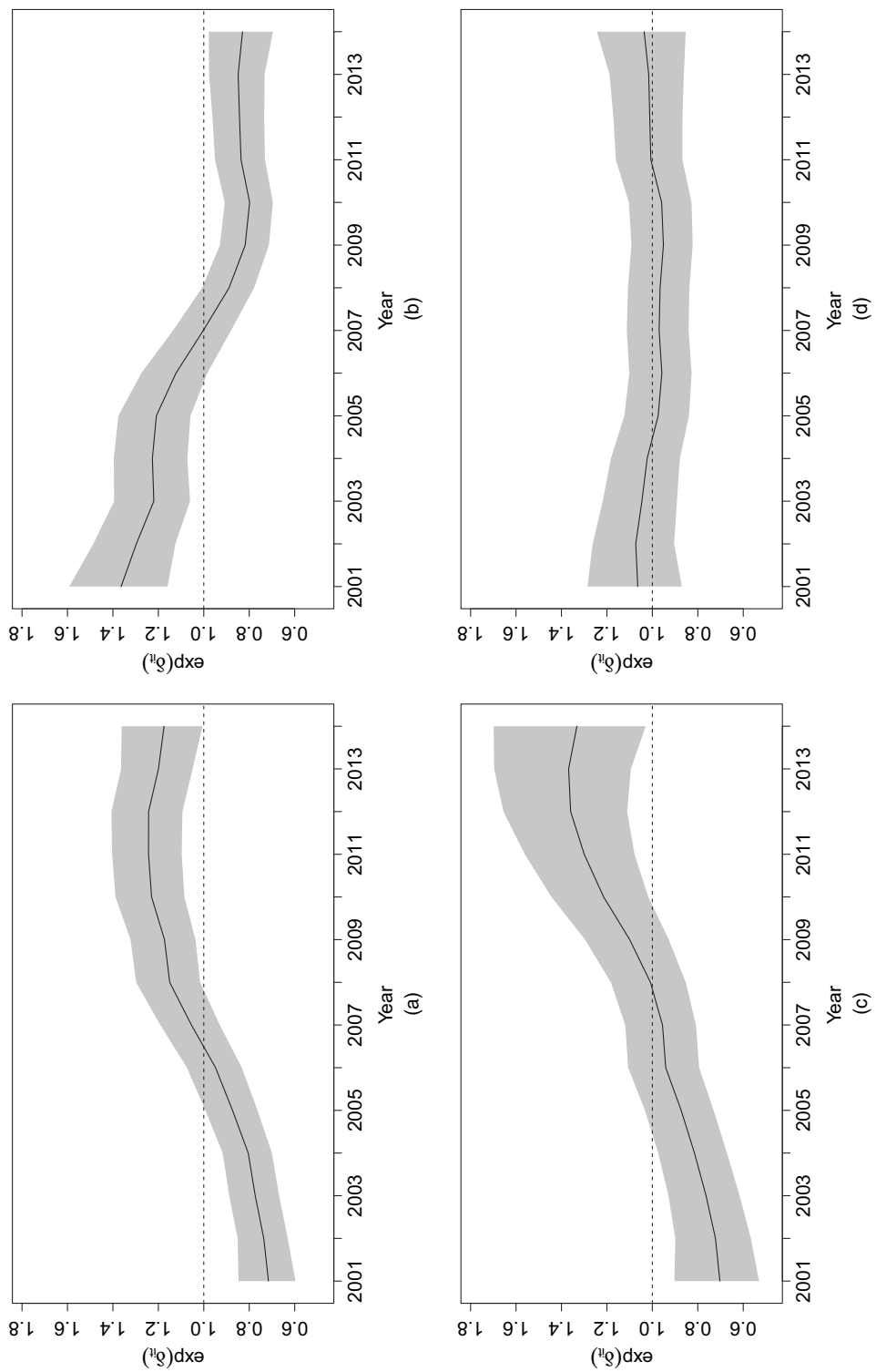
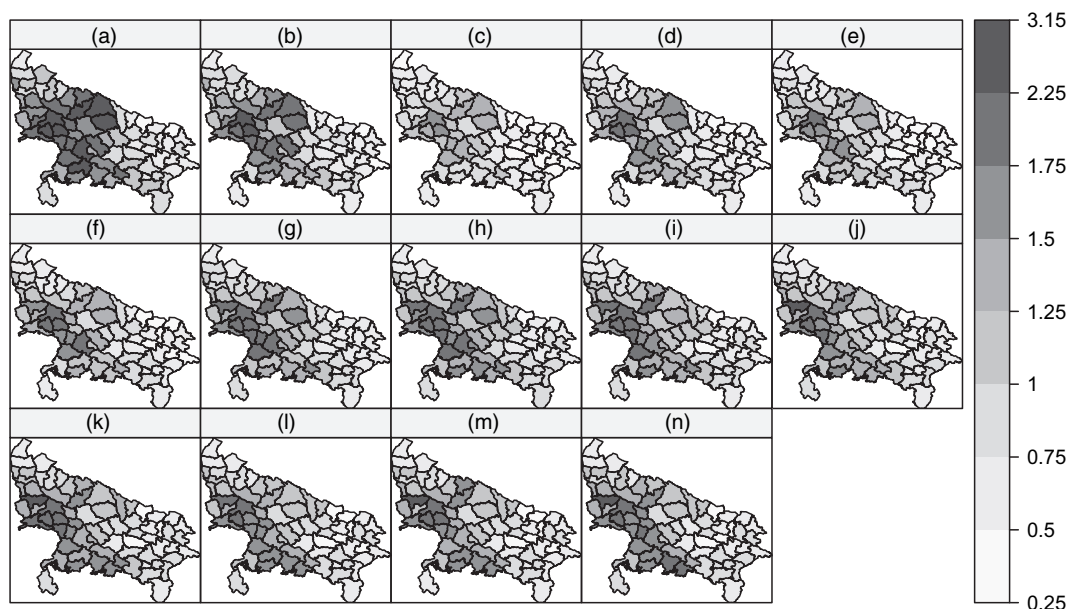


Fig. 7. Specific temporal trends (posterior mean of  $\exp(\delta_{it})$ ) for four selected districts: (a) Aligarh (district 2); (b) Kheri (district 43); (c) Shrawasti (district 64); (d) Varanasi (district 70)

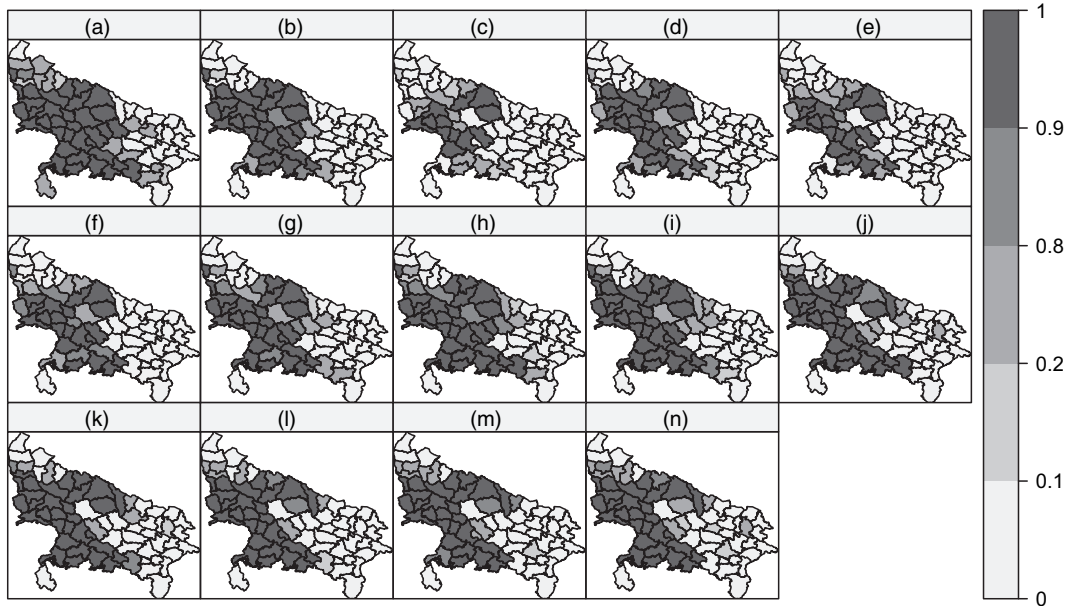


**Fig. 8.** Map of estimated risks for dowry deaths in Uttar Pradesh (posterior means of  $R_{it}$ ): (a) 2001; (b) 2002; (c) 2003; (d) 2004; (e) 2005; (f) 2006; (g) 2007; (h) 2008; (i) 2009; (j) 2010; (k) 2011; (l) 2012; (m) 2013; (n) 2014

the variability. The temporal effect accounts for 13% of the variability, and the remaining 9% is captured by the space–time interaction effect.

Fig. 5 displays posterior means of the district-specific relative risks  $\zeta_i = \exp(\xi_i)$  (Fig. 5(a)) and posterior probabilities that these district-specific risks are greater than 1,  $P(\zeta_i > 1 | \mathbf{O})$  (Fig. 5(b)). The spatial pattern  $\zeta_i$  can be interpreted as the basic risk associated with a district, and the strong spatial pattern is evident. Midwestern and some northern districts present a greater risk than the other districts. In addition, most of the eastern districts present a small district-specific risk. There is also a group of districts with low spatial risk in the north-west of the state. The map of posterior probabilities is clearly divided into two groups: one with districts whose posterior probabilities are greater than 0.9, and another group of districts with posterior probabilities smaller than 0.1. This indicates that the districts with posterior probabilities over 0.9 are classified as high-risk districts (see Richardson *et al.* (2004) or Ugarte *et al.* (2009a,b)). The spatial pattern and posterior probabilities that are obtained with the LCAR, DCAR and BYM2 and with PC or log-gamma priors are almost identical, indicating robustness with regard to the choice of spatial priors and hyperpriors in this particular data analysis.

Fig. 6 presents the overall temporal trend that is common to all districts, i.e. the posterior mean of  $\exp(\gamma_t)$ . This common temporal pattern may capture the effect of public policies or laws against crimes against women affecting the whole state. An abrupt fall is observed from 2001 to 2003. From then on, the ‘temporal’ risk increases until 2008 and then stabilizes around 1.1. Though the effect of the ruling party in Uttar Pradesh during the period studied was not significant, we could hypothesize that these variations may be due to past governmental actions, as typically their effects are usually noticeable several years later. The flat temporal trend around 1.1 from 2008 onwards indicates that the temporal component tends to increase the final risks slightly in this part of the period.

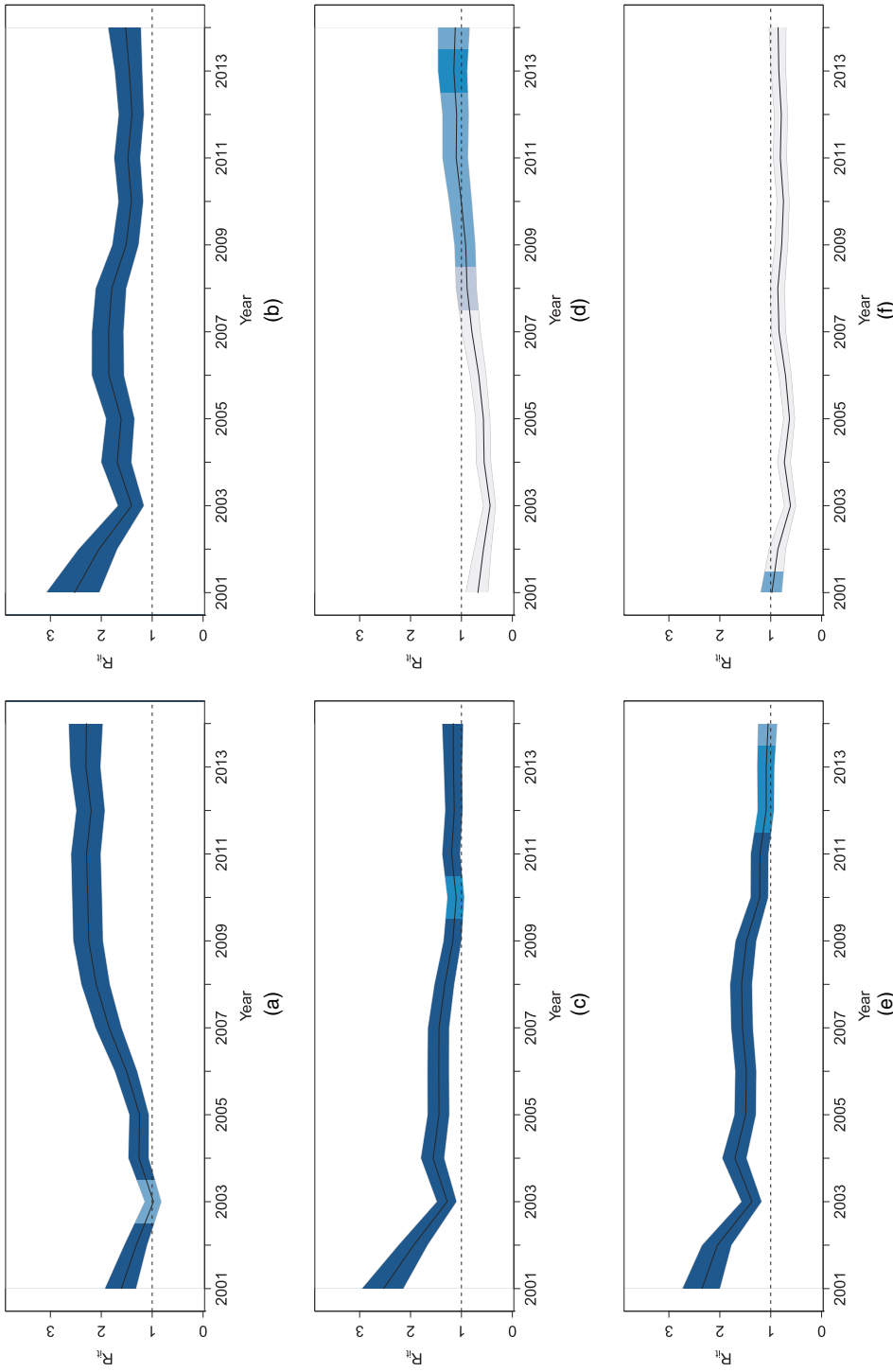


**Fig. 9.** Map of posterior probabilities that the relative risks are greater than 1,  $P(R_{it} > 1|\mathbf{O})$ : (a) 2001; (b) 2002; (c) 2003; (d) 2004; (e) 2005; (f) 2006; (g) 2007; (h) 2008; (i) 2009; (j) 2010; (k) 2011; (l) 2012; (m) 2013; (n) 2014

Area-specific temporal trends, i.e. the posterior means of  $\exp(\delta_{it})$  (with 95% credible intervals) are shown in Fig. 7 for Aligarh (district 2), Kheri (district 43), Shrawasti (district 64) and Varanasi (district 70). The specific temporal evolution in each area is clearly different, indicating that in some districts (Aligarh and Shrawasti) this trend increases, whereas in other districts the area-specific temporal trend decreases (Kheri) or is flat (Varanasi). This may indicate that perhaps district-specific policies are affecting the incidence of dowry deaths over time.

The geographical patterns of dowry deaths incidence (posterior mean of the relative risk  $\exp(R_{it})$ ) in the study period are shown in Fig. 8. The posterior probabilities that these relative risks are greater than 1,  $P(R_{it} > 1|\mathbf{O})$ , are shown in Fig. 9. From 2001 to 2003 the maps become lighter (Fig. 8) indicating a decrease in risk. However, from 2003 to 2008 the maps start to grow darker and from 2008 onwards the geographical pattern seems to stabilize. Looking at Fig. 9, most of the low-risk districts are in the eastern part of Uttar Pradesh. There are also a few low-risk districts in the north-western corner. High-risk districts are in the western-central part of the state.

Finally, Fig. 10 displays the temporal evolution of the final risk (posterior means of  $\exp(R_{it})$ ) and 95% credible intervals for several districts: Aligarh, Kanpur Dehat, Kheri, Shrawasti, Sitapur and Varanasi. The colours in the credible bands correspond to the posterior probabilities in Fig. 9 and indicate whether the risk is significantly high. Clearly, Aligarh is a high-risk district and the risk increased from 2003 to 2008. Then it stabilized around 2, indicating a risk of dowry deaths that is about twice as high as the risk for the whole of Uttar Pradesh. Other districts such as Kheri or Sitapur show a decrease in risk with similar behaviour to Uttar Pradesh as a whole at the end of the period (a risk around 1). These differences in the risk evolution suggest a close inspection of the districts to hypothesize about potential risk factors and possible protection factors that might be related to the districts with increasing and decreasing risk trends respectively. To have an idea about the magnitude of dowry deaths rates, the estimated rates



**Fig. 10.** Temporal evolution of final risk estimates for some districts in Uttar Pradesh. (a) Aligarh (district 2), (b) Kanpur Dehat (district 40), (c) Kheri (district 43), (d) Shravasti (district 64), (e) Sitapur (district 66) and (f) Varanasi (district 70): the range of blue colours in the credible intervals is the same as in Fig. 9 and indicates the posterior probability that the relative risk is greater than 1 ( $P(R_t > 1|O)$ ); dark blue means that this posterior probability is greater than 0.9, light blue means that the posterior probability is between 0.1 and 0.9, and grey means that the posterior probability is less than 0.1

(per 100000 women aged between 15 and 49 years) for these six districts are displayed in Table S.3 in the on-line supplementary material.

## 5. Discussion

Statistical techniques other than simple descriptive statistics are necessary to reveal patterns of crimes against women in general, and dowry deaths in particular, and to calibrate to what extent this is a problem of epidemic proportions. Spatiotemporal disease mapping models are powerful and useful techniques to shed light on this problem, to localize hot spots and to help to discover the underlying risk factors. In this work, we fit spatiotemporal models with different spatial CAR priors to assess their effects on the final risk estimates. CAR priors have been and still are widely used to deal with spatial heterogeneity. Here we consider the LCAR, DCAR and BYM2 priors. These priors are more general than the ICAR prior as they can cope with both spatially structured and unstructured variability, so they would be preferred in general. However, if the parameter  $\lambda_\xi$  is close to 1, all the spatial priors lead essentially to the ICAR prior. In the dowry death data analysis, the posterior mean of  $\lambda_\xi$  is practically equal to 1 if covariates are not included in the model. When covariates are included in the model, the posterior mean of  $\lambda_\xi$  is around 0.7. However, despite this decrease in the value of the spatial parameter  $\lambda_\xi$ , the ICAR prior was selected as results with the various spatial priors were nearly identical.

Speculating on potential causes or risk factors that are related to the district-specific risks is very difficult in India: a country with hugely diverse traditions, a complex social structure and high population density. As pointed out by a reviewer, establishing mechanisms other than in the existing literature to link certain covariates with dowry deaths is indeed a challenge. Moreover, covariates that may be related to dowry demand (and perhaps to dowry deaths), such as the man or woman's age at marriage or some characteristics related to the mother-in-law (Jeyaselan *et al.*, 2015) have been studied at individual or family level but cannot be obtained at area level. In general it is expected that crimes against women are associated with overall crime (Mukherjee *et al.*, 2001), but, as far as we know, there are no specific studies suggesting that certain offenders are more prone to commit dowry deaths, though some literature points towards higher rates of offending in unmarried men than in married men (King *et al.*, 2007; Sampson *et al.*, 2006). In general, the covariates that were analysed in this paper have been contemplated according to two criteria: availability and previous research. Consequently, and according to some existing literature, we considered the following covariates: political party of Uttar Pradesh's Chief Minister during the study period, sex ratio, population density, female literacy rate, *per capita* income, number of murders per 100000 inhabitants and number of burglaries per 100000 inhabitants.

In our analysis, only the sex ratio and number of murders and burglaries are statistically significant. This is consistent with the reference literature relating low child sex ratio to discrimination against daughters and sex-selective abortion (Shenk, 2007), and with literature relating low sex ratio to dowry and violence against women in the state of Maharashtra (Barakade, 2012; More *et al.*, 2012). Moreover, Mukherjee *et al.* (2001) indicated that dowry deaths are the only crime against women that seems to be associated with the sex ratio, and this association is negative. They also concluded that crimes against women are also expected to be positively associated with overall crime and that there is not a relationship with female literacy rate. Other studies have investigated the higher rates of murder in western districts of Uttar Pradesh and its correlation with low sex ratio (see Oldenburg (1992)). Strong correlations of murder rate and low female-to-male sex ratio have been reported also (Drèze and Khera, 2000). Our analysis



goes in this direction as female literacy rate is not significant whereas murders and burglaries are (with higher rates in western districts). In any case, more research is needed to disentangle social dynamics in areas with higher rates of crime that may have an effect on the dowry deaths phenomenon. The other two covariates in our analysis, population density and *per capita* income, were not found to be significant. The latter was included as Verma *et al.* (2015) revealed an association between dowry deaths and a rise in consumerism in India, as a dowry is a way to obtain certain goods that are still unaffordable for many families. With respect to the estimated global temporal trend, it is suspected that the variations observed could be due to changes in policies caused by alternations in power between political parties. Regarding the state political regime, Dang *et al.* (2018) found that, where the Bharatiya Janata Party or its coalition rules, a lack of evidence on the process of implementation of laws protecting women from cruelty is revealed. In our analysis, the covariate that is related to the political party ruling Uttar Pradesh was not found to be significant. However, we think that it may be problematic to assess the effect of the political party in the short term. In general, policies are usually oriented to a change of mind in the society, meaning a lengthy process whose effects in practice may appear only several years after their implementation. The Bharatiya Janata Party that was governing the state in 2001 (in fact since 1998) could have shaped a significant underreporting of cases thereafter. That could certainly explain the sudden fall of the temporal trend that was observed between 2001 and 2003.

Our findings should help authorities and social researchers to plan surveys to collect information in those districts with higher risks to look into other covariates (risk factors) that are not available at district level, such as the role of the mother-in-law or the age at marriage for brides and grooms from distant cohorts associated with a marriage squeeze, as stated in Dang *et al.* (2018). Other covariates derived from a low sex ratio might also be explored such as the ratio of unmarried males, under the ‘supply-of-offenders’ hypothesis (South *et al.*, 2014), the female migration rate (Khan, 2015) or figures on female dry thermal burn death as seen in Pandey *et al.* (2014). All this research and our own analysis reveal the complexity of the problem and the difficulty in finding the underlying risk factors. A different attempt would be to use more general indicators comprising different covariates. This is the approach that was followed by Tanwar *et al.* (2016) and Kumar *et al.* (2018) that constructs composite indices of agricultural, social and industrial development for 28 eastern and 25 western districts of Uttar Pradesh respectively. Their findings could give an explanation for the persistent spatial pattern that we have found in our data analysis, where districts in the east show low risks of dowry deaths whereas western districts have high risks. In general, eastern districts have better social and industrial development indices than do western districts. As an example, Sant Ravidas Nagar Bhadohi, which is one of the worst eastern districts in the social index (25th position), has a better index than Hathras, the best ranked western district in the same index. Similarly, concerning the industrial index, the third-best district in the west, Meerut, has a value of this index that is worse than Sant Kabir Nagar, which is ranked 25 within the eastern districts. Comparisons regarding the agricultural development index are not so clear, with some districts in the west showing better scores than some districts in the east. In any case, there are 11 districts in the west with less agricultural development than 25 out of 28 districts in the east.

In summary, our study is a first attempt to unveil spatiotemporal patterns of dowry deaths in the districts of Uttar Pradesh, which is the most populous state in India, and to figure out underlying risk factors. We have currently detected significant associations between dowry deaths and sex ratio, murders and burglaries. The positive association between dowry deaths and murders and burglaries confirms that, in general, areas with high overall crime rates also have high rates of crimes against women.

The utility and benefit of our research are clear though the study may have some limitations. The first is the problem of possible underreported cases. Some studies attest to underreported cases of crimes against women (see for example Ellsberg *et al.* (2001), McNally *et al.* (1998) and Mukherjee *et al.* (2001)). The social stigma that is related to sexual crimes makes underreporting noticeably high. Also, when violence involves family members, it is not reported because the victims do not consider such incidents as crimes (Koss, 1992). Regarding dowry deaths, Verma *et al.* (2015) suggested that some underreporting could be related to natal relatives' fear of scandal and discredit that would ruin the possibilities of marriage for other daughters. However, Mukherjee *et al.* (2001) stated that, whereas some crimes such as rape are little reported, others such as dowry death rarely go unreported. Even in the case of underreporting, if we accept that regions that are highlighted as high-risk regions in official records are likely to be high-risk regions, the spatiotemporal patterns discovered may be very useful. This debate about whether dowry deaths are underreported or not leads us to raise some questions about how the National Bureau obtains local crime statistics; if there are multiple police agencies within the state acting in the districts, or to what extent the perceptions of healthcare providers, forensic doctors, police or witnesses affects the classification of a dowry death (Dang *et al.*, 2018; Belur *et al.*, 2014). Answers to these questions could shed light on the problem.

The second limitation is that data are not available for different age groups, preventing us from studying age-specific spatiotemporal patterns. Some age-dependent differences could be expected as some researchers (Jeyaseelan *et al.*, 2015) have stated that, as the bride's age at marriage increases, the risk of dowry demand also increases, and also that a marriage squeeze depends on cohorts' imbalance and age gap between brides and grooms (Dang *et al.*, 2018).

Finally, the third limitation is the difficulty of assessing the statistical significance of the covariates in spatiotemporal models with spatial, temporal and spatiotemporal random effects, as confounding problems may be present (Reich *et al.*, 2006; Hodges and Reich, 2010). We are currently investigating this last issue more deeply.

## Acknowledgements

This work has been supported by project MTM2017-82553-R (Agencia Estatal de Investigación–Fondos de Desarrollo Regional, European Union). It has also been partially funded by the Caixa Foundation (grant 1000010434), Caja Navarra Foundation and National University of Distance Learning, Pamplona, under agreement LCF/PR/PR15/51100007. We are very grateful to Amitha Puranik and Dr V. S. Binu and Dr Seena Biju for providing the data and helping us to understand some results. The authors also thank the Associate Editor and two reviewers for their comments that have helped to improve an earlier version of this paper.

## References

- Aragonés, N., Goicoa, T., Pollán, M., Militino, A. F., Pérez-Gómez, B., López-Abente, G. and Ugarte, M. D. (2013) Spatio-temporal trends in gastric cancer mortality in Spain: 1975-2008. *Cancer Epidemiol.*, **37**, 360–369.
- Banerjee, K. (1999) Gender stratification and the contemporary marriage market in India. *J. Family Iss.*, **20**, 648–676.
- Banerjee, P. R. (2014) Dowry in 21st-Century India: the sociocultural face of exploitation. *Trauma Viol. Abuse*, **15**, 34–40.
- Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn. Boca Raton: Chapman and Hall.
- Barakade, A. J. (2012) Declining sex ratio: an analysis with special reference to Maharashtra state. *Geosci. Res.*, **3**, 92–95.
- Belur, J., Tilley, N., Daruwalla, N., Kumar, M., Tiwari, V. and Osrin, D. (2014) The social construction of “dowry deaths”. *Soc. Sci. Med.*, **119**, 1–9.

- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995) Bayesian analysis of space-time variation in disease risk. *Statist. Med.*, **14**, 2433–2443.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B*, **36**, 192–236.
- Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1–20.
- Bloch, F. and Rao, V. (2002) Terror as a bargaining instrument: a case study of dowry violence in rural India. *Am. Econ. Rev.*, **92**, 1029–1043.
- Carroll, R., Lawson, A. B., Faes, C., Kirby, R. S., Aregay, M. and Watjou, K. (2015) Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatl Spatiotemp. Epidem.*, **14-15**, 45–54.
- Czado, C., Gneiting, T. and Held, L. (2009) Predictive model assessment for count data. *Biometrics*, **65**, 1254–1261.
- Dang, G., Kulkarni, V. S. and Gaiha, R. (2018) Why have dowry deaths risen in India? *Working Paper 2018/03*. Australia South Asia Research Centre, Australian National University, Canberra. (Available from <https://crawford.anu.edu.au/acde/asarc/pdf/papers/2018/WP2018-03.pdf>.)
- Dawid, A. (1984) Statistical theory: the prequential approach. *J. R. Statist. Soc. A*, **147**, 278–292.
- Dean, C. D., Ugarte, M. D. and Militino, A. F. (2001) Detecting interaction between random regions and fixed age effects in disease mapping. *Biometrics*, **57**, 197–202.
- Drèze, J. and Khera, R. (2000) Crime, gender, and society in India: insights from homicide data. *Popln Devlpmnt Rev.*, **26**, 335–352.
- Ellsberg, M., Heise, L., Pena, R., Agurto, S. and Winkvist, A. (2001) Researching violence against women, methodological considerations from three Nicaraguan studies. *Stud. Family Planng*, **32**, 1–19.
- Ellsberg, M., Heise, L. and World Health Organization (2005) *Researching Violence against Women: a Practical Guide for Researchers and Activists*. Geneva: World Health Organization.
- Garcia-Moreno, C. and Watts, C. (2011) Violence against women: an urgent public health priority. *Bull. Wrld Hlth Organzn*, **89**, no. 2.
- Gelfand, A. E. and Sahu, S. K. (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Am. Statist. Ass.*, **94**, 247–253.
- Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Baysn Anal.*, **1**, 515–534.
- Gneiting, T. and Raftery, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *J. Am. Statist. Ass.*, **102**, 359–378.
- Goicoa, T., Adin, A., Etxeberria, J., Militino, A. F. and Ugarte, M. D. (2019) Flexible Bayesian p-splines for smoothing age-specific spatio-temporal mortality patterns. *Statist. Meth. Med. Res.*, **28**, 384–403.
- Goicoa, T., Adin, A., Ugarte, M. D. and Hodges, J. S. (2018) In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stoch. Environ. Res. Risk Assessmnt*, **2**, 749–770.
- Goicoa, T., Ugarte, M. D., Etxeberria, J. and Militino, A. F. (2016) Age-space-time CAR models in Bayesian disease mapping. *Statist. Med.*, **35**, 2391–2405.
- Hodges, J. S. and Reich, B. (2010) Adding spatially-correlated errors can mess up the fixed effects you love. *Am. Statistn*, **64**, 325–334.
- Jeyaseelan, V., Kumar, S., Jeyaseelan, L., Shankar, V., Yadav, B. and Bangdiwala, S. (2015) Dowry demand and harassment: prevalence and risk factors in India. *J. Biosocl Sci.*, **47**, 727–745.
- Johnson, H., Ollus, N. and Nevala, S. (2007) *Violence against Women: an International Perspective*. New York: Springer Science and Business Media.
- Khan, N. A. (2015) Migrant women, maternal health and HIV/AIDS: a study of Aligarh suburbs. *Doctoral Dissertation*. Aligarh Muslim University, Aligarh.
- King, R. D., Massoglia, M. and MacMillan, R. (2007) The context of marriage and crime: gender, the propensity to marry, and offending in early adulthood. *Criminology*, **45**, 33–65.
- Knorr-Held, L. (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Statist. Med.*, **19**, 2555–2567.
- Koss M. P. (1992) The under-detection of rape: methodological choices influence incidence estimates. *J. Socl Iss.*, **48**, 61–75.
- Kumar, S., Mourya, K. K., Gupta, R. P. and Singh, S. N. (2018) Dynamics of socio-economic development of districts of western Uttar Pradesh, India. *Int. J. Curr. Microbiol. Appl. Sci.*, specl iss. 7, 838–843.
- Leroux, B. G., Lei, X. and Breslow, N. (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (eds M. Halloran and D. Berry), pp. 179–192. New York: Springer.
- MacNab, Y. C. (2011) On Gaussian Markov random fields and Bayesian disease mapping. *Statist. Meth. Med. Res.*, **20**, 49–68.
- Mari-Dell’Olmo, M., Martínez-Beneito, M. A., Gotsen, M. and Palència, L. (2014) A smooth ANOVA model for multivariate ecological regression. *Stoch. Environ. Res. Risk Assess.*, **28**, 695–706.

- McNally, R. J., Metzger, L. J., Lasko, N. B., Clancy, S. A. and Pitman, R. K. (1998) Directed forgetting of trauma cues in adult survivors of childhood sexual abuse with and without posttraumatic stress disorder. *J. Abnorm. Psychol.*, **107**, 596–601.
- Mohanty, S., Sen, M. and Sahu, G. (2013) Analysis of risk factors of dowry deaths south Indian study. *J. Forens. Leg. Med.*, **20**, 316–320.
- More, V. P., Ingale, A. R. and Shinde, V. S. (2012) Generation and district wise study of sex ratio. *Int. J. Hlth Sci. Res.*, **2**, 1–7.
- Mukherjee, C., Rustagi, P. and Krishnaji, N. (2001) Crimes against women in India: analysis of official statistics. *Econ. Polit. Wkly*, **36**, 4070–4080.
- National Crime Records Bureau (2015) Crime in India 2014 compendium. National Crime Records Bureau New Delhi. (Available from <http://www.ncrb.gov.in/StatPublications/CII/CII2014/Compendium>.)
- Oldenburg, P. (1992) Sex ratio, son preference and violence in India: a research note. *Econ. Polit. Wkly*, **27**, 2557–2662.
- Pandey, S. K., Kumar, A. and Gupta, M. (2014) An epidemiological retrospective study of autopsied dry thermal burn. *J. Ind. Acad. Forens. Med.*, **36**, 363–366.
- Patel, A. V. (2015) Crime against the women in India. *Forens. Res. Crimin. Int. J.*, **1**, 1–5.
- Pettit, L. (1990) The conditional predictive ordinate for the normal distribution. *J. R. Statist. Soc. B*, **52**, 175–184.
- Powell, J. and Wahidin, A. (2007) Old age and victims: a critical exegesis and an agenda for change. *Intrnt J. Crimin.*, 1–14.
- Prasad, D. (2016) Symbiosis, reciprocity and village exogamy in two villages of Uttar Pradesh. *Sth Asn Anthr.*, **16**, 185–194.
- Reich, B., Hodges, J. S. and Zadnik, V. (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, **62**, 1197–1206.
- Richardson, S., Thomson, A., Best, N. and Elliott, P. (2004) Interpreting posterior relative risk estimates in disease-mapping studies. *Environ. Hlth Perspect.*, **112**, 1016–1025.
- Riebler, A., Sørbye, S. H., Simpson, D. and Rue, H. (2016) An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statist. Meth. Med. Res.*, **25**, 1145–1165.
- Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman and Hall–CRC.
- Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc. B*, **71**, 319–392.
- Russo, N. F. and Pirlott, A. (2006) Gender-based violence. *Ann. New York Acad. Sci.*, **1087**, 178–205.
- Sampson, R. J., Laub, J. H. and Wimer, C. (2006) Does marriage reduce crime?: A counterfactual approach to within-individual causal effects. *Criminology*, **44**, 465–508.
- Sharma, B. R., Harish, D., Sharma, V. and Vij, K. (2002) Kitchen accidents vis-a-vis dowry deaths. *Burns*, **28**, 250–253.
- Shenk, M. K. (2007) Dowry and public policy in contemporary India: the behavioral ecology of a “social evil”. *Hum. Nat.*, **18**, 242–263.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017) Penalising model component complexity: a principled, practical approach to constructing priors. *Statist. Sci.*, **32**, 1–28.
- South, S. J., Trent, K. and Bose, S. (2014) Skewed sex ratios and criminal victimization in India. *Demography*, **51**, 1019–1040.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc. B*, **64**, 583–639.
- Srinivasan, P. and Lee, G. R. (2004) The dowry system in Northern India: women’s attitudes and social change. *J. Marr. Family*, **66**, 1108–1117.
- Tanwar, N., Kumar, S., Sisodia, B. V. S. and Hooda, B. K. (2016) Dynamics of socio-economic development of districts of eastern Uttar Pradesh. *J. Appl. Nat. Sci.*, **8**, 5–9.
- Ugarte, M. D., Adin, A. and Goicoa, T. (2017) One-dimensional, two-dimensional, and three dimensional B-splines to specify space-time interactions in Bayesian disease mapping: model fitting and model identifiability. *Spatl Statist.*, **22**, 451–468.
- Ugarte, M. D., Adin, A., Goicoa, T. and Militino, A. F. (2014) On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statist. Meth. Med. Res.*, **23**, 507–530.
- Ugarte, M. D., Goicoa, T., Etxeberria, J. and Militino, A. F. (2012) A P-spline ANOVA type model in space-time disease mapping. *Stoch. Environ. Res. Risk Assess.*, **26**, 835–845.
- Ugarte, M. D., Goicoa, T., Ibañez, B. and Militino, A. F. (2009a) Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics*, **20**, 647–665.
- Ugarte, M. D., Goicoa, T. and Militino, A. F. (2009b) Empirical Bayes and fully Bayes procedures to detect high-risk areas in disease mapping. *Computnl Statist. Data Anal.*, **53**, 2938–2949.
- United Nations General Assembly (1993) Declaration on the elimination of violence against women. United Nations General Assembly, New York. (Available from <http://www.un.org/documents/ga/res/48/a48r104.htm>.)

- Verma, R. K., Srivastava, P. C., Sinha, U. S. and Kaul, A. (2015) Study of unnatural deaths in married females within seven years of marriage in Allahabad. *J. Ind. Acad. Forens. Med.*, **37**, 405–409.
- Vicente, G., Goicoa, T., Puranik, A. and Ugarte, M. D. (2018) Small area estimation of gender-based violence: rape incidence risks in Uttar Pradesh, India. *Statist. Appl.*, **16**, 71–90.
- Watanabe, S. (2010) Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*, **11**, 3571–3594.
- World Health Organization (2013) *Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Nonpartner Sexual Violence*. Geneva: World Health Organization.

*Supporting information*

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Crime against women in India: unveiling spatial patterns and temporal trends of dowry deaths in the districts of Uttar Pradesh"'