**Modeling Infant Mortality in a Hierarchical Bayesian Framework:**
**Spatio-Temporal Convergence in Italy from 1990 to 2010**

Fedele Greco - Francesco Scalone
Department of Statistical Sciences - University of Bologna, Italy

## 1. Introduction

The main purpose of this paper is to define a statistical method to model infant mortality rates for provincial areas in Italy by using a Bayesian approach.

As a matter of fact, when working on single small areas, direct estimates of infant mortality rate (henceforth IMR) can be affected by large variances. Therefore, we propose a statistical model that allows area-level estimates of infant mortality to borrow strength from each other by exploiting spatial association of provincial IMRs and taking into account temporal correlation. The need for a spatiotemporal model is motivated by an explorative analysis. This approach has become very popular in the disease mapping literature (Kim and Lim 2010, Knorr-Held 2000) but, to our knowledge, it has not been employed for modeling IMRs in a demographic framework. Adopting a Bayesian approach, Markov Chain Monte Carlo methods (Girosi and King 2008) will be used to fit the model and to sample from the posterior predictive distribution. The outline of this paper is as follows. We first review recent Bayesian hierarchical models of mortality. Then we present the used data providing a brief description of infant mortality trends by provinces in Italy, from 1990 to 2010. Afterwards, adopting a Bayesian approach, we define a spatial-temporal model, giving details on implementation and prediction. Indicators for assessing convergence and inequalities in infant mortality across provinces and time will also be calculated.

## 2. Bayesian Hierarchical Spatio-Temporal Frameworks

For the purposes of this paper, we preliminary need to look at methods that forecast mortality in multiple sub-group populations. From this point of view, Bayesian hierarchical formulations can be ideal to control for all possible sources of variability (concerning both model parameters and predictions) and to model mortality rates on different geographical sub-national units, as in the case of the province. During the last two decades, hierarchical Bayesian estimates of mortality rates were proposed, based on Binomial or Poisson sampling and taking into account spatial correlation procedure (Poisson distribution arises naturally when data set takes the form of counts). These hierarchical models are generally based on a Poisson model for the first stage and incorporate covariates by various modeling of the Poisson parameters at the second stage (Kim and Lim 2010). Spatial effects are typically included as random with some distributions where their parameters must be estimated. Maximum likelihood methods are generally used to estimate hyper-parameters whereas Bayesian or empirical Bayesian methods are applied to obtain Poisson parameters and rates. In particular, Knorr-Held (2000) propose spatio-temporal interaction models where the spatial effects are nested within time so that it is possible to take into account how spatial heterogeneity and patterns evolved over time.

## 3. Descriptive analyses

In Italy, a province is an administrative division of intermediate level between a municipality and a region. As a matter of fact, some provinces can be quite different in terms of population dimension. Thus, we want to take into account this variation in population size when we model and predict infant mortality levels by province. In order to model IMRs rates at provincial level, we take into account number of births and infant deaths at age 0 in each year from 1992 to 2009. We used data collected by the Italian Institute of National Statistics (Istat) referring to 103 Italian provinces in the period between 1992 and 2009. These series were preliminary reconstructed in order to have time-

constant provincial borders. In order to highlight spatio-temporal trends of infant mortality in Italy, we take into account direct estimates of IMRS. Let $D_{0tp}$ denote the frequencies of infant deaths at age 0 for the *t*-th time period in the *p*-th province. Direct estimates of the IMRs are obtained as:

$$m_{0tp} = \frac{D_{0tp}}{B_{tp}}$$

where $B_{tp}$ is the number of births at time t in province p and represents the exposure to risk.
In figure 1, provinces are classified by direct IMR level for six given years (1992, 1995, 1998, 2001, 2004 and 2007). In each map, data is presented according to a seven intervals classification based on the septiles (7-quantiles). In order to have the same classification in each year, septiles are calculated by taking into account all the observed IMRs in each province between 1992 and 2009. In these terms, figure 1 represents the general declining trend of infant mortality during the period in question. Indeed, we can easily see that the first two 1992 and 1995 maps in figure 1 are dominated by the darkest highest infant mortality categories, since almost all the provinces fell in the highest quantiles. On the other hand, in 2005 and 2007, we can see the clearest lowest infant mortality categories prevalence. In figure 2, IMRs are still considered. Nevertheless, category intervals are not constant, since septiles are calculated on IMRs ranking of a given single year which the map refers to. In these terms, we can observe the permanence of mortality clustering during the whole period in question, right up to 2007. As a matter of fact, higher IMRs are concentrated in some visible cluster of provinces, providing evidence of spatial autocorrelation of mortality. In these terms, the importance of spatial structure components when modeling IMRs for provincial sub-populations is shown.

*Figure 1. IMRs by province. Colored intervals = Seven quantiles based on the distribution of all infant mortality rates observed from 1992 to 2007*

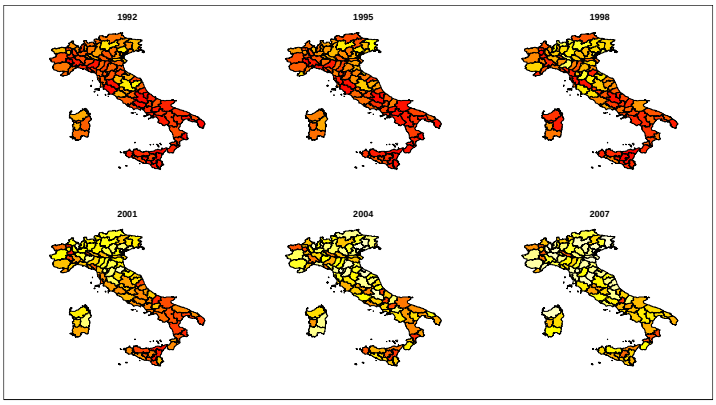
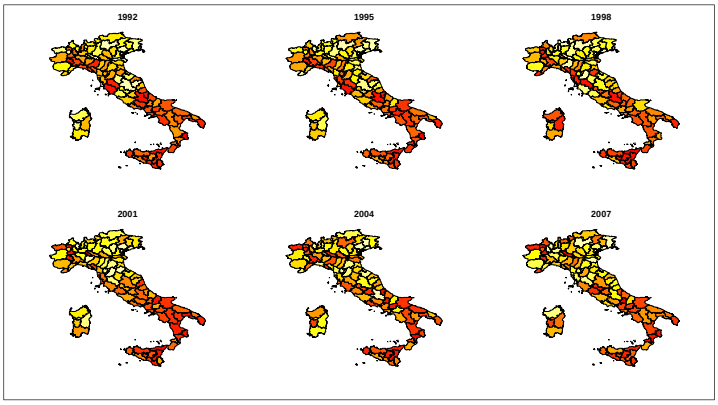



*Figure 2. IMRs by province. Colored intervals = Seven quantiles based on distributions of infant mortality rates observed in each single given year*



## 4. Hierarchical Bayesian Models for Space-Time Variation

The proposed model is intended to capture both temporal and spatial features of the infant mortality trends, along with spatiotemporal interactions. The model which follows was first proposed by Knorr-Held (2000) and has been adopted, with some modification, in several disease mapping applications (see for example: Kim and Lim 2010). Let $B_{tp}$ and $D_{0tp}$ denote respectively the births (exposure to risk) and the infant death counts at age 0 in year $t$ and province $p$, $t=1,...,T$; $p=1,...,P$. At the first level of the hierarchy, conditionally on model parameters involved in higher levels, we assume that infant death counts $D_{0tp}$ follow independent Poisson distributions with parameter $B_{tp} \cdot \mu_{0tp}$ i.e.:

$$D_{0tp} \sim Poisson(B_{tp} \cdot \mu_{0tp})$$

where $\mu_{0tp}$ denotes the infant mortality rate which is modeled as a logarithm function of three sets of random effect as follows:

$$\ln(\mu_{0tp}) = \alpha_t + \varphi_p + \delta_{tp}$$

Random terms $\alpha_t$, $\phi_p$, $\delta_{tp}$ are temporal, spatial and spatiotemporal random effects respectively. Model hierarchy is completed by assuming prior distributions for each random vector $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_t,...,\alpha_T)$, $\boldsymbol{\varphi}=(\varphi_1,...,\varphi_p,...,\varphi_P)$, $\boldsymbol{\delta}=(\delta_{11},...,\delta_{tp},...,\delta_{TP})$. For all these set of parameters, an Intrinsic Conditional AutoRegressive (ICAR) model is assumed, that can be expressed as multivariate Normal distributions parameterized in terms of mean vectors and precision. The distribution of the interaction term $\boldsymbol{\delta}$ is characterized by a precision matrix obtained as the Kronecker product of the precisions of $\boldsymbol{\alpha}$ and $\boldsymbol{\varphi}$ (Knorr-Held 2000).A Markov Chain Monte Carlo (MCM) algorithm is used to sample from the joint posterior distribution of the parameters. The MCMC algorithm consists in block-Metropolis steps for sampling random effects $\boldsymbol{\alpha}$, $\boldsymbol{\varphi}$, $\boldsymbol{\delta}$. Gibbs steps can be adopted for sampling precision parameters $\sigma_\alpha^{-2}$, $\sigma_\varphi^{-2}$, $\sigma_\delta^{-2}$, since their full conditional distributions are available in closed form. The algorithm we built performs an automatic tuning of the proposal distribution in order to achieve an acceptance rate of around 40%.

## 4. Results

Figure 3 shows the average spatial structure at provincial level which was incorporated by the $\boldsymbol{\varphi}$ components. In Figure 4, model-based estimates are compared with direct estimates for 1992: it can be seen that model based estimates are less variable than direct estimates. In fact, the confidence intervals which refer to direct estimates are on average 20% wider than the model-based credibility intervals. Such reduction in the uncertainty of the estimates can be attributed to the borrowing strength process.

*Figure 3. Posterior means of the spatial random effects $\boldsymbol{\varphi}$ for age 0*

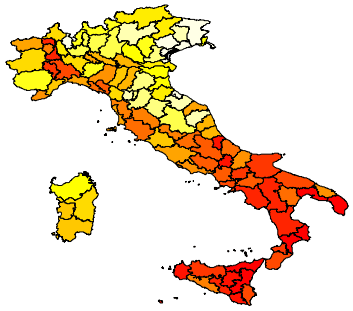
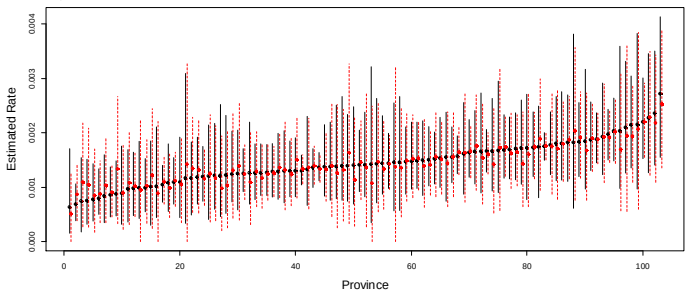
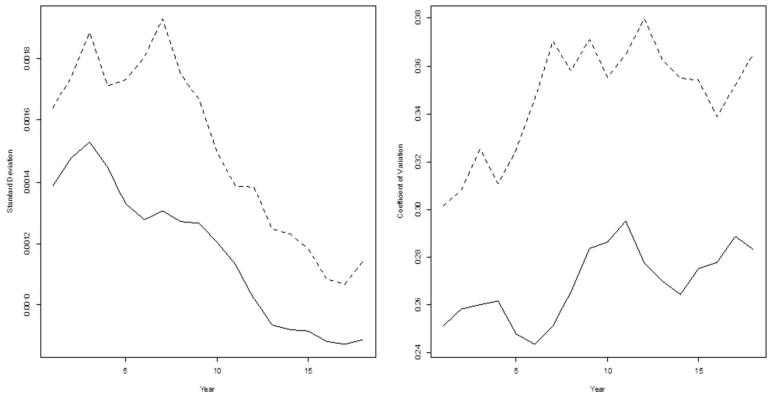


*Figure 4. Model-based estimates of IMRs (black dots) and Bayesian credibility intervals (black lines). Direct estimates of IMRs (red dots) and confidence intervals (red dashed lines). Year 1992*



In order to assess the existence of possible converging trends in infant mortality, in figure 5 we calculate coefficients of variation (CV) and standard deviation (SD) of direct and model-based estimates of IMRs at provincial level in each observed year. These measures are useful to assess the existence of infant mortality differences and health inequalities at territorial level. On the one hand, standard deviation of direct estimates and model estimates clearly declines in the last fifteen years of observation. On the other hand, coefficients of variations show an increasing trend. As a matter of fact, the decline of standard deviation is not due to a pure reduction in variability of IMRs, since this reduction is more related to the decline of the average infant mortality levels. The increasing and stationary trend of the coefficient of variation demonstrates the persistence of variability at provincial level, since this measure does not depend on the mean level.

*Figure 5. Standard Deviation and Coefficient of Variation of Direct estimates of IMRs (dashed line) and Model-based estimates of IMRs (black line)*




## 5. Concluding Remarks

In this preliminary paper, we discuss a spatio-temporal Bayesian hierarchical model to estimate small-area IMRs at provincial level. Since provincial sub-populations in Italy can widely vary, direct estimates of mortality rates can show evident uncertainty when considering smaller provinces. In a preliminary descriptive analysis, we clearly demonstrated the existence of specific spatiotemporal patterns which can be easily incorporated in a comprehensive Bayesian hierarchical model. By exploiting the observed spatial association and temporal correlation trends, we proposed a model where provincial-level infant mortality rates estimates can borrow strength from each other in order to reduce uncertainty related to the estimates. Model estimation is performed, under a Bayesian framework, by adopting an MCMC algorithm to obtain samples from the posterior distribution. As a result, it appears that model based estimates are less variable than direct estimates.

In order to assess infant mortality convergence at territorial levels, we also calculated standard measures of variability on direct and model-based estimates of IMR. These preliminary results show the persistence of infant mortality inequalities, since an increasing trend of CVs is observed. In the next step, we will use other indicators for assessing inequalities in infant mortality across provinces taking into account the spatial components $\varphi_p + \delta_{tp}$ and $\delta_{tp}$of the proposed model.

## References

Clayton D., 1996. “Generalized linear mixed models. In Markov Chain Monte Carlo in Practice”, ed. by W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. Chapman and Hall, 275-301.

Girosi F. and King G., 2008. “Demographic Forecasting”, *Princeton University Press*.

Kim H. and Lim H., 2010. “Comparison of Bayesian Spatio-Temporal Models for Chronic Diseases”. Journal of Data Science 8(2010), 189-211.

Knorr-Held L. 2000. “Bayesian modelling of inseparable space-time variation in disease risk”. Statistics in Medicine. 19: 2555-2567.