Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR WORKING PAPER WP 2001-025
AUGUST 2001 (REVISED NOVEMBER 2001)

Spatial profiles in the analysis of event histories: An application to first sexual intercourse in Italy

Francesco C. Billari (billari@demogr.mpg.de)
Riccardo Borgoni (borgoni@demogr.mpg.de)

Spatial profiles in the analysis of event histories
An application to first sexual intercourse in Italy

Francesco C. Billari and Riccardo Borgoni
Max Planck Institute for Demographic Research
Doberaner Str. 114
D-18057 Rostock, Germany.
E-mail: {billari,borgoni}@demogr.mpg.de

November 16, 2001

Abstract

The aim of this paper is mainly a methodological one. We used individual-level retrospective data from the Italian Fertility and Family survey, based on the information provided by respondents. Linking individual-level data with geographical characteristics of the community, we firstly used non-parametric methods, and in particular the bi-dimensional LOESS (Locally weighted regression) method, to build smoothed maps of the transition to first sexual intercourse. Secondly, we used a discrete-time multilevel event-history model (with logit specification), allowing for the presence of unobserved heterogeneity at the municipality level. We used 'latitude', 'longitude', and dimension of the municipality as aggregate-level explanatory variables in the model. Thus, we could build spatial profiles and a map, using predicted values for the log-odds-ratio, and we could study cohort dynamics too. Our results confirm that Italy is a heterogeneous country; when looking at age at the first sexual intercourse, women are more influenced by their context than men.

1 Introduction

The study of demographic events has traditionally been based on their empirical investigation over time and space. The central role of time (in particular of individual time expressed by age or other relevant duration variables, together with cohort and period) has been recognized by researchers, who developed classical techniques, such as life tables. The development of more recent techniques that deal with individual-level time-to-event data, (i.e. event history analysis), has led to a specialized set of techniques of statistical demography. On the other hand, spatial aspects have occupied a peripheral position in formal and statistical demography, and even more so does the interaction between space and time.

Analyses of demographic phenomena across space, and the representation of such phenomena using maps, have substantially contributed to major research efforts in demography. The Princeton project (Coale and Watkins, 1986) extensively relied on maps to describe the dynamics of the demographic transition and to support explanatory ideas. As it has been pointed out by Lesthaeghe and Neels (2000) "maps produce good narratives, and these in their turn are valuable for theory formation as well". Thus, spatial analysis has had an important role in the development of key theories that explain demographic behavior. GIS packages, similar to those we exploit in this paper, provide unique features for drawing maps using low levels of geographical aggregation.

During the last twenty years, spatial statistics has developed into a specialized field, and several monographs have been published on this subject (Ripley, 1981; Diggle, 1983; Ripley, 1988; Cressie, 1993). Most studies relate to epidemiology, and their authors make use of aggregate-level spatial data. The spatial analysis of (bio-)demographic data is another developing field (Sleegers and Goethals, 1993; Bocquet-Appel et al., 1996). One of the key issues in this stream of research is the link between space and time (the latter is defined both in terms of societal time and individual time). The same link is discussed in epidemiologically-oriented demographic studies (see for instance Congdon, 2000). For individual-level data, non-parametric or semi-parametric methods have been used (for a review see Diggle (2000) and the bibliographic references listed there).

The aim of this paper is mainly a methodological one. Using the most widespread set of statistical techniques used in modern demographic research, event history analysis (Courgeau and Lelièvre, 1989; Blossfeld and Rohwer, 1995), we incorporated in our analysis a spatial approach. In event history analysis, the spatial correlation among individuals is sometimes taken into account insofar as space introduces a hierarchical structure to data. This idea is embedded in so-called multilevel models. In these models, correlation is allowed among units in the same area, but independence is assumed among areas. Examples of this approach include Steele et al. (1996) and Barber et al. (2000) for discrete time, and Boldstad and Manda (2001) for continuous time. A spatial, discrete-time event history model, that allows also for time-varying effects, is developed within a full Bayesian framework by Fahrmeir and Lang (2001), who use a unified approach for Bayesian MCMC inference in generalized additive and semiparametric mixed models.

Given the importance of retrospective surveys in demography, we exploit Fertility and Family Survey data. We present an application to age at first sexual intercourse of Italian men and women. For behaviors such as first sexual intercourse, where Italy is a relatively traditional country according to Western standards, though large and relatively stable differentials also exist betwen countries (Bozon and Kontula, 1997). Cultural components such as social norms, religiosity, and long-standing family structures play a key role in shaping geographical differences in the timing of the first sexual intercourse (see e.g. Buzzi, 1998). The latter is a behavior for which the cultural heterogeneity of Italy can emerge from spatial analysis, with the possibility of crossing traditional administrative categorizations being an advantage. A south-north gradient is plausible at least for women, given that southern Italy is a rather traditional part of the country. To give an example, according to Cazzola (1999), north-eastern regions have an estimated odds-ratio of 3.46 over southern regions, regarding to first sexual intercourse by the age of 19.

The methods we propose also allow for an investigation of the possibility that a spatial diffusion process occurs. We form the hypothesis that new behavior, for instance sexual debut at an earlier age, follows a diffusion process across space. We discuss this point explicitly when we analyze cohort dynamics.

The paper is structured as follows. After a brief description of the data (Section 2), we use non-parametric methods to draw maps of Italy based on the timing of the first sexual intercourse (Section 3). In Section 4, we use a multi-level discrete-time event history model in order to estimate a spatial profile by latitude and longitude. Section 5 outlines a societal time perspective by comparing large groups of cohorts Section 6 concludes the paper and suggests avenues for further research.

2 Data

We use individual-level retrospective data from the Italian Fertility and Family survey (De Sandre et al., 1997). A total number of 4,824 females and 1,206 males, were interviewed, representing a cross-section of residents born betwween 1946 and 1975. Among other questions on reproductive health, the respondents were asked about their age at first sexual intercourse (if they had any). The interviews were conducted between November 1995 and January 1996 and the age at first sexual intercourse was reported in completed years. A range of checks showed that the data quality was very satisfactory (Cazzola, 1999), and the self-reported age at first intercourse has been shown to be a reliable indicator (Dunne et al., 1997). In general, the data showed that first sexual intercourse in Italy takes place at a higher age than in other western European countries. The paper by Cazzola (1999) contains a detailed description of the data and some analyses.

In order to link individual data to the geographical context, we used the information provided by respondents who resided in a municipality for most of their time for the first 15 years of their life. Given that complete residential histories are not available from the survey, this method proved to be the best choice when studying first sexual intercourse. It has obvious advantages compared to the alternative solution of using as a criterion residence at the time the interview takes place. Of course, we can neither be completely sure that the first sexual intercourse was experienced at the place of residence, nor do we know for sure that it happened after change of residence (if any). We linked the individual-level data with data of the municipality of main residence for first 15 years of one's life, using geographical coordinates (latitude and longitude) together with dimensional characteristics (population at the 1991 census).

3 Maps of Italy: A nonparametric spatial approach

We first built maps of the transition to first sexual intercourse, taking a non-parametric approach, which takes explicitly into account the geographical coordinates of the community. Essentially, this involves the application of two-dimensional smoothing techniques to the data of the survey, using a two-step procedure. In the first step, we compute estimates of the quantities of interest at the area level. In the second step, we use nonparametric smoothing techniques to build a visual representation of the geographical pattern of the process. Thus, the aim of our analysis is not to produce estimates at the level of small areas, but to draw maps that describe–in a consistent way–the differences in the timing of the event under study across the whole country.

3.1 Methods

We considered two separate procedures, which provide two different indicators of the timing of first sexual intercourse. In the first procedure, we showed spatial patterns based on non-parametric estimates of the survival function. This approach is appropriate for the data available to us. For this purpose, and given the size of the sample, we computed aggregate measures. In particular, we computed the median age of the first intercourse for each Italian province (103 values). We then applied the smoothing techniques, by using the central coordinate of the province as the location for the nonparametric estimates. The estimated values were subsequently extended to each municipality, with the possibility of having different values for each province.

In the second procedure, we built our spatial patterns starting directly from data assigned to the geographical coordinates of the municipality. For each municipality, for which there is at least one case in the sample, we calculated the frequency of first intercourse below a given age. At this level it is impossible to compute the median survival level. Consequently, frequencies for this indicator potentially ranged from 0 to 100%[1]. Subsequently, we applied smoothing techniques to obtain an estimated value for every Italian municipality. As for the ages selected, we used the median age (to the closest integer) for

[1] The frequency of first intercourse below a given age was used in the analysis by Cazzola (1999).

each sex, that is 18 years for males and 19 years for females.

In both cases we used the smoothing method LOESS (robust LOcally weighted regrESSion) (Cleveland, 1979)[2]. The estimator is based on the q points of the sample that are nearest to the currently estimated observation (with $1 \leq q \leq n$ where n is the sample size). A local polynomial of degree d is fitted using weighted last squares, where the nearest points are the most important ones in the computation of fitted values. The regression coefficients are estimated in order to minimize the following expression:

$$\sum_{k=1}^{n} w_k(x_i)(y_k - \beta_0 - \beta_1 x_k + \dots - \beta_d x_k^d)^2$$

where $w_k(x)$ is derived from the suitable kernel function $w(x)$, which weights observations according to their distance from the current point. This estimator is built by a iterative procedure that repeatedly down-weigthts outliers from the curve estimated in the previous step. In our application, we opted for a second order polynomial function. We used the so called 'tri-weight' function defined as

$$w(x) = (1 - |x|^3)^3 \qquad |x| < 1 \qquad w(x) = 0 \qquad |x| \geq 1:$$

If $\hat{\beta}(x_i)$ is the vector of estimated parameters the estimated value for y_i is $\hat{y}_i = \sum_{r=1}^{d} \hat{\beta}_r(x_i)x_i^r$.

The generalization to higher dimension surfaces is described in Cleveland and Devlin (1988). In this case, the method gives a multivariate surface in $p+1$ dimensions if p regressors are used. In a multivariate context a p-dimensional neighborhood must be selected in the predictor space. The selected points are the q nearest points to the p-variate point (x_i), relative to the i-th observation currently estimated. Once again, each

[2] LOESS was initially introduced in order to produce a smoother version of scatter plots, i.e. to accommodate data for which $y_i = g(x_i) + \varepsilon_i$ where ε_i represents a random variable with 0 mean and a constant scale, and $g(.)$ represents a smooth function.

point is weighted according to its distance $\rho(x_i; x_k)$ from the current observation by the weight function of the previous type. The metric used here is the usual Euclidean norm. In this case also, the weights are obtained as $w_k(x_i) = w(\rho(x_i; x_k)/h_i)$ where h_i is the distance from the p-variate point x_i from its q-th nearest point[3].

3.2 Results

The maps that result from our analysis are shown in Figures 2 and 1[4]. For females, one can perceive a spatial trend from south-east to north-west–for both indicators (Figure 2-(A) and Figure 1-(A)), with a reversal of the trend in the far north-west. In case of males, the spatial trend appears to be much less evident. Here we observed a more uniform distribution of early intercourse (or, of the lower median age) across the country.

In general, local features play a more important role for males than for females. In case of females, the traditional south-north gradient appears to be relevant. These results are in line with the expectations presented in the existing literature on this subject. Nevertheless, we shall keep in mind that because of the smaller sample size, the analysis for males suffers from a larger variability factor than the one for females.

FIGURES 1 AND 2 AROUND HERE

4 A parametric discrete-time event history spatial approach

4.1 Methods

[3] In our analysis, we used the LOESS algorithms implemented in S-Plus 2000 (Venables and Ripley, 2000). A feature of the algorithm is that predictions from LOESS are confined to the range of the data, both on the x and the y axis. We imputed out-of-range values with a spatial trend computed with a Generalized Additive Model (GAM). GAM is a type of nonparametric model widely used in the statistical literature (Hastie and Tibshirani, 1990). GAM is a generalization of the Generalized Linear Model; the linear predictor is replaced by an additive predictor. In the case of p explanatory variables, a GAM model can be defined as $g(\mu) = \alpha + \sum_{i=1}^{p} f_i(x)$; where the response Y is assumed to belong to the exponential family, $\mu = E(Y)$ and $g(x)$ is a regular function called link function.

[4] In order to build maps at a municipality level for this and subsequent sections, we use the ArcView GIS software package [?].

In order to study the spatial profile of the hazard of first sexual intercourse, we used a multilevel (variance component) discrete-time event history model–belonging to the family of generalized linear mixed models (see e.g. Goldstein, 1995 and Steele et al., 1996). We built a person-year dataset and we linked area-level spatial variables[5] with the individual-level geographical data. The area-level we use is the municipality. Individuals who have not yet experienced first sexual intercourse at the time of interview are right-censored.

Given that the first sexual intercourse is a non-repeatable event, it is not possible to identify a variance component which represents unobserved heterogeneity at the individual level. It is nevertheless possible to identify such a component at the area level. This means that we can estimate at the area level the presence of unobserved factors which influence the timing of the first sexual intercourse. We use only spatial characteristics (including the size of the area) as area-level covariates, with linear splines for the specification of their profile. We designed separate models for men and women.

The general formulation of the model we used is the following:

$$\begin{aligned}\log\frac{\Pr\{Y_{xij}=1\}}{1-\Pr\{Y_{xij}=1\}} = \alpha_0 &+ \sum_{\{k: x_k \le x < x_{k+1}\}} \alpha_k X_k + \\ &\sum_{\{k:\lambda_k < L_j\}} \beta_k \left(\min(\lambda_{k+1}, L_j) - \lambda_k\right) + \\ &\sum_{\{k:\mu_k < M_j\}} \gamma_k \left(\min\left(\mu_{k+1}, M_j\right) - \mu_k\right) + \\ &\sum_{\{k:\pi_k < P_j\}} \delta_k \left(\min(\pi_{k+1}, P_j) - \pi_k\right) + \varepsilon_j;\end{aligned}$$

where: Y_{xij} is the indicator for the first sexual intercourse at age x for the individual i in the area j, α_0 represents the baseline log-odds-ratio, X_k is the indicator for age class $k = 1, ..., l$ with the lower limit x_k and the upper limit x_{k+1} (with $x_{l+1} = \infty$) and α_k is the log-odds-ratio of the age class k. Moreover we put:

[5] Remember that the place of residence is constant across time because we can only use as a reference the first 15 years of an individual.

$\sum_{\{k:\lambda_k<L_j\}} \beta_k (\min(\lambda_{k+1}, L_j) - \lambda_k)$: linear spline for latitude, with knots fixed at λ_k (starting from the southern most point in Italy). β_k represents the slope between λ_k and λ_{k+1} and L_j represents the latitude of area j;

$\sum_{\{k:\mu_k<M_j\}} \gamma_k \left(\min\left(\mu_{k+1}, M_j\right) - \mu_k\right)$: linear spline for longitude, with knots fixed at μ_k (starting from the western most point in Italy). γ_k represents the slope between μ_k and μ_{k+1} and M_j represents the longitude of area j;

$\sum_{\{k:\pi_k<P_j\}} \delta_k (\min(\pi_{k+1}, P_j) - \pi_k)$: linear spline for the population, with knots fixed at π_k. δ_k represents the slope between π_k and π_{k+1} and P_j is the population (in the unit of thousands) of area j. Finally ε_j represents the random residual at the area level distributed as $N\left(0, \sigma_\varepsilon^2\right)$.

We used a full-likelihood specification, which can be implemented by the aML package (Applied Maximum Likelihood) (Lillard and Panis, 2000). Standard errors of the estimates are corrected using a Huber-type procedure (Huber, 1967).

4.2 Results

In some of the specifications we fitted a subset of the model just presented. Models 1 and 2 (Tables 1 and 2) were designed, for both males and females, in order to see whether there is a significant component of unobserved heterogeneity at the area level. In other words, we wanted to see whether space (and space-related determinats) matters for this behavior. The only other variable which we considered in these models is the age group. As we expected, there clearly was an age trend, with a non-monotonic shape of the hazard reaching its maximum at ages 18 and 19 for males, and 20 and 21 for females. When we focused on unobserved heterogeneity at the area level (estimated in model 2), we gained some interesting insights. For males, the estimated standard deviation of ε_j is 0.24. It is different from zero only at the 5% significance level, and the likelihood-ratio test rejects the model with unobserved heterogeneity (the p-value is 0.16). This result is consistent with the findings in the non-parametric part: spatially-related aspects play a less relevant role for men than they do for women. In fact, when we analyze model 2 for women, the s.d. is estimated at 0.41. It is different from zero to with a p-value lower than 1%, which is also the case for the likelihood-ratio test. Having obtained our first results, we proceeded by using models with unobserved heterogeneity at the area model for females only. In what follows, mostly for reasons of comparisons, we continue analyzing males.

TABLES 1, 2 AROUND HERE

Spatial coordinates are used in models 3, 4 and 5 (the results are reported in Table 3 for males and Table 4 for females). In model 3, we considered only the south-north gradient, using latitude (which is believed to play the most important role); we then introduced longitude as a further dimension (model 4). In model 5, we checked our results, controlling for the size of the place of residence at the 15th birthday.

The inclusion of latitude and longitude splines improved the model without unobserved heterogeneity at the community level for men (the likelihood-ratio test for model 4 against model 1 is 26.28 with 15 d.f., p-value=0.04). As we expected, this is even more so the case for women (the likelihood-ratio test for model 4 against model 2 is 224.86 with 15 d.f., p-value<0.01). These results justify the inclusion of the splines for geographical coordinates for both genders. The results regarding estimates for the parameters of linear splines are not easily interpretable. Nevertheless, they allowed us to produce spatial profiles by plotting them. We also introduced the size of the place of residence as a control variable to draw spatial profiles net of its effect. For both genders, it does not add significantly to the model (likelihood ratio tests of model 5 vs. model 4 are 3.66 for men and 5.34 for women, with 4 d.f. and p-values of 0.45 and 0.25 respectively).

TABLES 3, 4 AROUND HERE

In Figure 3, we plotted the south-north profile of the log-odds-ratios for the hazard of first sexual intercourse for males, derived from model 5, and in Figure 4 the same profile is shown for the west-east direction. Figures 5 and 6 represent spatial profiles for females. For reference purposes, we indicated over the plots the location of some cities. As it appeared from the non-parametric analysis, the south-north gradient is evident and it is important for females. For both males and females, the peak of the risk of first sexual intercourse is around the latitude corresponding (more or less) to the city of Rimini, with an odds-ratio of about 2.6 regarding the southernmost point. Much less variation is observed when the longitude profile is drawn.

FIGURES 3, 4, 5 AND 6 AROUND HERE

The advantage of the method we propose is that, by using spatial coordinates, we are able to draw maps using log-odds derived from the model. In Figures 7 and 8 we reproduced the maps of Italy from the model. The pictures we obtained are not far from the ones drawn, using non-parametric methods. The advantage here is that we used a full-information maximum likelihood approach, and parametric methods allow for more flexibility, e.g. control for time-varying covariates, also using standard existing packages which embed multilevel models. In our case, maps help to connect the results of the discrete-time event history parametric model to those of the non-parametric approach. The south-north gradient for women, for instance, still clearly shows up, and for both sexes the peculiarities of the central-northern areas are visible.

FIGURES 7 AND 8 AROUND HERE

5 Cohort dynamics

We then focused our attention on societal change and split the data into two groups according to birth cohort. Given the size of the samples and the results of our past analysis (the low importance of space attributed to males), we performed this exercise for females only. The older cohort is composed of females born up to 1960, while the younger cohort consists of females born after 1961. We thus obtained groups that are almost balanced (46.6% of the total female sample is in the older cohort). Cohort analysis is useful since it provides insight into the question whether behavior that is related to the first sexual intercourse is influenced by a spatial diffusion process. Following the same path we used before, we then first analyzed the spatial aspects of cohort dynamics, using non-parametric methods in the same fashion as shown in section 3. Then we considered parametric event history models in order to study the contribution of the spatial coordinates to the explanation of cohort changes 4. In the last subsection, we also used a more detailed definition of cohort, in that we allowed for cohort trends.

5.1 Mapping cohort dynamics using nonparametric methods

As was the case for the full sample, we first performed an analysis based on the median survival computed aggregating data in provinces. We then carried out on an analysis based on the municipal frequency of 'early' first intercourse.

The maps in Figure 9 show the geographic shape of the median survival age for the older and younger cohort. Cut points are chosen to be equal to the ones to which we referred to in the case of the whole sample. These maps are thus directly comparable with map (A) in Figure 1. The maps in Figure 10 show the geographic shape of the frequency with respect to the cut points used in the whole sample case. Thus these maps are directly comparable with map (A) in Figure 2.

FIGURES 9 AND 10 AROUND HERE

Across the whole country, a general shift to early first intercourse is visible. A remarkable stability is present in the south while in the center and in the north we observed a trend towards a behavior that is more homogeneous, which might be due to diffusion processes. This process appears to start from the two poles of the extreme north east and Emilia Romagna. A similar and fast development takes place in the island of Sardinia. The non-parametric analysis of cohort data thus provides evidence for a diffusion effect.

5.2 Parametric analysis of cohort dynamics

In this part we limited our attention to the differential role of spatial coordinates in explaining variability at the time to first sexual intercourse at the cohort level. The idea is that, if a spatial diffusion process underlies part of the observed dynamics, geographical coordinates should explain a larger share of the community-level variability (which also reflects socio-economic and cultural factors).

We thus fitted separate models, formulated as model 4 in past analyses, for both cohorts. Results are presented in Table 5. In the older cohort, about half of the community-level estimated variance is explained when we introduce latitude and longitude splines. In the younger cohort, this share raises to more than 70%. The results are consistent with those found when using the non-parametric method, and they provide some evidence that a spatial diffusion process is taking place.

To further investigate the role of spatial coordinates, we allowed the municipality-level variability to be continuously varying across birth cohorts, using a linear spline to represent the standard deviation as it varies across cohorts (we used year and month of birth, and knots are put every fifth year). The results of this analysis are shown in Figure 11.

FIGURE 11 AROUND HERE

Similarly to table 5, we introduced geographical coordinates subsequently only. The figure shows a weak, but still interesting evidence of 1) the increasing role of municipality-level factors across cohorts, with 2) an increasing part of municipality-level factors which are explained merely by geographical coordinates.

TABLE 5 AROUND HERE

6 Conclusions and discussion

In this paper, we combined the techniques of event history analysis and spatial analysis jointly in order to analyze the age at first sexual intercourse in Italy. By starting from FFS survey data linked to the geographical coordinates of the place of residence of respondents during their first 15 years of life, we used non-parametric methods to draw maps capturing this phenomenon in Italy. Subsequently, we used discrete-time event history models, including geographical coordinates as a municipality-level covariate. In our view, the simple use of geographical co-ordinates as area-level covariates, which is easily applicable to other studies with the use of standard software allowing for multilevel statistical models, is the major innovation of this paper. This approach also allows to build maps of relative odds of experiencing a certain event. Finally, we carried out an analysis of cohort dynamics, mainly in order to explore the presence of spatial diffusion processes. Our proposals joining two important and so far rather separated streams of research, and we believe this is their main strength and novelty.

Concluding this paper, we would also like to stress the limitations of the proposed approach and of the analysis we were able to carry out. Given the limitations of our data, we did not fully exploit the possibilities of the proposed approach. In particular, individuals were attached to one municipality only, while the place of residence is clearly a time-varying covariate during the life course, especially if one looks at young adults. In general, we did not exploit the possibility of using time-varying covariates. Such opportunities should be exploited in future research efforts following the approach we propose.

From the technical side, the normality assumptions underlying the distribution of unobserved heterogeneity may be overcome by other, less stringent, assumptions. We followed a standard approach because we wanted to emphasize other aspects and we could not enter into the debate on assumptions concerning the distribution of unobserved characteristics. Another point, connected to the improved exploitation of geographical information systems, is the use of geographical coordinates. If micro-level data allow for that, finer references (e.g. for large cities) can be used in future research, and this will allow for more precision in the definition of a decision-relevant context.

From the behavioural side, it is clear that this approach to analysis cannot be a means to an end. Explaining geographical variation is the next task. Maps and profile materials provide a subject for subsequent research, both on the theoretical and on the empirical side. For instance, a more refined analysis is necessary to further investigate the spatial diffusion of earlier first sexual intercourses in Italy. Our methods, however, allow us to analyze the dynamics of the role of space for behaviors for which event history analysis is an appropriate technique.

Disclaimer

The views expressed in this paper are the authors' own and do not necessarily represent the views of the Max Planck Institute for Demographic Research.

Acknowledgements

Presentations based on former versions of this papers were given at the Conference of the Italian Coordinating Committee for Demography, Milan, February 2001, at the Workshop on Spatial Aspects of Demography, Rostock, Germany, May 2001, and at the Session S29 "Spatial Demographic Analysis", chaired by Daniel Courgeau and Jean-Pierre Bocquet-Appel) at the XXIV IUSSP General Conference, Salvador, Brazil, August 2001. We thank participants of the presentations for their useful comments. In particular, we are grateful to Arnstein Aassve, Susann Backer, Daniel Courgeau, and Gianpiero Dalla Zuanna for comments on earlier versions.

References

Barber J.S., Murphy S.A., Axinn W.G., Maples J., 2000, Discrete-Time Multilevel Hazard Analysis, Sociological Methodology, 30: 201-235.

Blossfeld H.-P., Rohwer G., 1995, Techniques of Event History Modeling. New Approaches to Causal Analysis. Lawrence Erlbaum Associates, Mahwah, NJ.

Bocquet-Appel J.-P., Courgeau D., Pumain D. (Eds.), 1996, Spatial Analysis of Biodemographic Data. John Libbey, Paris.

Bolstand W.M., Manda S.A., 2001, Investigating child Mortality in Malawi using family and community random effects: a bayesian analysis, Journal of the American Statistical Association, 96, 12-19.

Bozon M., Kontula O., 1997, Initiation sexuelle et genre: comparaison des évolutions de douze pays européens, Population, 52, 1367-1400.

Buzzi C., 1998, Giovani, affettività, sessualità. L'amore tra i giovani in un'indagine IARD. il Mulino, Bologna.

Cazzola A., 1999, L'ingresso nella sessualità adulta, in De Sandre P., Pinnelli A. e Santini A., Nuzialità e fecondità in trasformazione: percorsi e fattori del cambiamento. il Mulino, Bologna.

Cleveland W.-S., 1979, Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association,, 74, 829-36.

Cleveland W.-S., Devlin S.-J., 1988, Locally weighted regression: an approach to regression analysis by local fitting Journal of the American Statistical Association, 83, 596-611.

Coale A.J., Watkins S.C. (Eds.), 1986, The Decline of Fertility in Europe. Princeton University Press, Princeton.

Congdon P., 2000, Monitoring Suicide Mortality: A Bayesian Approach, European Journal of Population, 16 251-284.

Courgeau D., Lelièvre É., 1989, Analyse démographique des biographies, Éditions de l'INED, Paris. English edition: 1992, Event history analysis in demography. Clarendon Press, Oxford.

Cressie N.A., 1993, Statistics for Spatial Data. Wiley, New York.

De Sandre P., Ongaro F., Rettaroli R., Salvini S., 1997, Matrimonio e figli: tra rinvio e rinuncia. il Mulino, Bologna.

Diggle P.J., 1983, Statistical Analysis for Spatial Point Patterns. Academic Press, London.

Diggle P.J., 2000, Overview of statistical methods for disease mapping and its relationship to cluster detection. In Spatial Epidemiology: methods and applications P. Elliott, J. C. Wakefield, N.G. Best, D. J. Brings (Eds.). Oxford University Press, Oxford.

Dunne M.P., Martin N.G., Statham D.J., Pangan T., Madden P.A., Heath A.C., 1997, The consistency of recalled age at first sexual intercourse, Journal of Biosocial Science, 29, 1-7.

ESRI, 1996, Using Avenue. Environmental Systems Research Institute, Redlands, CA.

Fahrmeir L., Lang S., 2001, Bayesian inference for generalised additive mixed models based on random Markov field priors. Applied Statistics, 50, 201-220.

Goldstein H., 1995, Multilevel Statistical Models, Second Edition. Edward Arnold, London.

Hastie T., Tibshirani R., 1990, Generalized Additive Models. Chapman and Hall, London.

Huber P.J. 1967, The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, CA: Vol. 1, 221-233.

Lesthaeghe R., Neels K., 2000, Maps, narratives and demographic innovation. IPD-WP 2000-8, Interface Demography, Brussels.

Lillard L.A., Panis C.W.A., 2000, aML Multilevel Multiprocess Statistical Software, Release 1.0. EconWare, Los Angeles, CA.

Ripley B.D., 1981, Spatial Statistics. Wiley, New York.

Ripley B.D., 1988, Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge.

Sleegers W.F., Goethals A.L.J., 1993, Quantitative Geographical Methods. SISWO publication 366.

Steele F., Diamond I., Amin S., 1996, Immunization Uptake in Rural Bangladesh: a Multivel Analysis. Journal of the Royal Statistical Society, 159: 289-299.

Venables W.N., Ripley B.D., 1999, Modern Applied Statistics with S-plus (3rd ed.). Springer, New York.

TABLES AND FIGURES

Parameters	Model 1		Model 2	
	estimate	std.error	estimate	std.error
Constant	-2.6629**	0.0833	-2.6968**	0.0883
Age (reference=14-15)				
16-17	1.1732**	0.1098	1.1865**	0.1098
18-19	1.7941**	0.1009	1.8350**	0.1010
20-21	1.6703**	0.1128	1.7343**	0.1140
22-23	1.2941**	0.1847	1.3766**	0.1850
24-25	1.5239**	0.2166	1.6289**	0.2243
26+	0.3369**	0.2101	0.4667*	0.2208
σ_u			0.2352*	0.0945
log-likelihood	-2668.54		-2667.53	

**: p-value< 0.01, *: p-value< 0.05

Table 1: Discrete time event history models 1-2. Males

Parameters	Model 1		Model 2	
	estimate	std. error	estimate	std. error
Constant	-3.8504**	0.0723	-3.9125**	0.0738
Age (reference=14-15)				
16-17	1.4922**	0.0824	1.5182**	0.0828
18-19	2.3740**	0.0795	2.4604**	0.0793
20-21	2.5083**	0.0835	2.6617**	0.0840
22-23	2.3823**	0.0882	2.5993**	0.0877
24-25	2.5518**	0.0950	2.8119**	0.0972
26+	1.7543**	0.0898	2.0995**	0.0910
σ_u			0.4056**	0.0200
log-likelihood	-11453.41		-11359.30	

**: p-value< 0.01

Table 2: Discrete time event history models 1-2. Females.

Parameters	Model 3		Model 4		Model 5	
	estimate	std. error	estimate	std. error	estimate	std.error
Constant	-2.7370**	0.3527	-1.7202**	0.5008	-1.6800**	0.5331
Age (reference=14-15)						
1617	1.1855**	0.1116	1.1804**	0.1118	1.1828**	0.1127
18-19	1.8324**	0.1016	1.8239**	0.1030	1.8270**	0.1042
20-21	1.7304**	0.1144	1.7204**	0.1144	1.7211**	0.1158
22-23	1.3725**	0.1906	1.3503**	0.1922	1.3497**	0.1976
24-25	1.6261**	0.2222	1.5971**	0.2206	1.6106*	0.2340
26	0.4614*	0.2314	0.4255+	0.2191	0.4471*	0.2228
Latitude spline (origin=36±38°N)						
Slope -38	0.0736	0.3271	0.0212	0.3175	-0.0556	0.3291
Slope 38-39	-0.6554*	0.3215	-0.8804*	0.3547	-0.8914*	0.3587
Slope 39-40	0.5116	0.3315	0.7477+	0.4357	0.7980+	0.4552
Slope 40-41	0.1837	0.2488	0.1329	0.3849	0.1165	0.4025
Slope 41-42	0.0019	0.2360	-0.0320	0.2841	-0.1412	0.2899
Slope 42-43	0.0088	0.3102	-0.0265	0.3294	0.1360	0.3390
Slope 43-44	0.2375	0.3072	0.1790	0.3009	0.1640	0.3035
Slope 44-45	-0.5030*	0.2265	-0.6287*	0.2476	-0.6651**	0.2516
Slope 45-	0.1218	0.1801	0.2704	0.2013	0.2722	0.1979
Longitude spline (origin=6±37°E)						
Slope -9			-0.3275*	0.1425	-0.3195*	0.1503
Slope 9-11			0.0016	0.0868	-0.0228	0.0926
Slope 11-13			-0.0322	0.1172	-0.0039	0.1179
Slope 13-15			-0.0970	0.1371	-0.1030	0.1356
Slope 15-17			0.1593	0.1471	0.1849	0.1490
Slope 17-			-0.2839	0.3057	-0.3377	0.3140
Population spline (origin=0)						
0-5,000					0.0251	0.0373
5,000-20,000					-0.0138	0.0094
20,000-100,000					0.0009	0.0018
100,000-1,000,000					0.0001	0.0001
1,000,000-						
σ_u	0.1644	0.1122				
log-likelihood	-2660.26		-2655.40		-2653.57	

**: p-value< 0.01, *: p-value< 0.05, +: p-value< 0.1

Table 3: Discrete time event history models 3-5. Males

Parameters	Model 3		Model 4		Model 5	
	estimate	std. error	estimate	std. error	estimate	std. error
Constant	-4.4179**	0.2815	-4.1571**	0.3360	-4.1341**	0.3476
Age (reference=14-15)						
16-17	1.5164**	0.0835	1.5168**	0.0835	1.5170**	0.0869
18-19	2.4687**	0.0809	2.4698**	0.0809	2.4706**	0.0814
20-21	2.6777**	0.0854	2.6792**	0.0852	2.6801**	0.0858
22-23	2.6254**	0.0885	2.6278**	0.0885	2.6290**	0.0896
24-25	2.8475**	0.1010	2.8484**	0.1005	2.8489**	0.0995
26-	2.1430**	0.0940	2.1339**	0.0928	2.1377**	0.0930
Latitude spline (origin= 36±38°N)						
Slope -38	-0.0305	0.2573	-0.0328	0.2368	0.0272	0.2392
Slope 38-39	0.2351	0.2191	0.2835	0.2213	0.2555	0.2240
Slope 39-40	-0.0574	0.2268	-0.0016	0.2640	-0.0073	0.2746
Slope 40-41	-0.0864	0.1550	-0.1737	0.1740	-0.2072	0.1859
Slope 41-42	0.7996**	0.1281	0.5790**	0.1762	0.5316**	0.1929
Slope 42-43	-0.2921+	0.1754	-0.2784	0.1955	-0.1820	0.2046
Slope 43-44	0.5586**	0.1853	0.5471**	0.1925	0.4948*	0.1924
Slope 44-45	-0.4394**	0.1321	-0.2876*	0.1378	-0.2430	0.1403
Slope 45-	0.0675	0.0851	-0.0080	0.0856	-0.0038	0.0877
Longitude spline (origin= 6±37°E)						
Slope -9			-0.1515+	0.0855	-0.2269**	0.0850
Slope 9-11			0.1956**	0.0485	0.2041**	0.0504
Slope 11-13			-0.0555	0.0581	-0.0486	0.0599
Slope 13-15			-0.1406+	0.0775	-0.1378+	0.0800
Slope 15-17			0.0356	0.0815	0.0301	0.0869
Slope 17-			-0.0872	0.1778	-0.0842	0.1816
Population spline (origin=0)						
0-5,000					0.0153	0.0215
5,000-20,000					0.0001	0.0058
20,000-100,000					0.0011	0.0012
100,000-1,000,000					0.0000	0.0001
1,000,000-						
σ_u	0.2909**	0.0254	0.2647**	0.0270	0.2682**	0.0297
log-likelihood	-11257.55		-11246.87		-11244.20	

**: p-value< 0.01, *: p-value< 0.05

Table 4: Discrete time event history models 3-5. Females.

Cohorts	a Model without coordinates estimate	 std.error	b Model with coordinates estimate	 std.error	$c = 1 - (a^2 - b^2)/a^2$ Variance (%) explained by geographical coordinates
1946-1960					
σ_u	0.3162**	0.0370	0.2233**	0.0527	50.1%
1961-1975					
σ_u	0.4801**	0.03865	0.2503**	0.0501	72.8%

**: pvalue< 0.01, *: pvalue< 0.05

Table 5: Discrete time event history models 1-2. Males

Figure 1: Median age at first intercourse: (A) females (B) males

(A) (B)

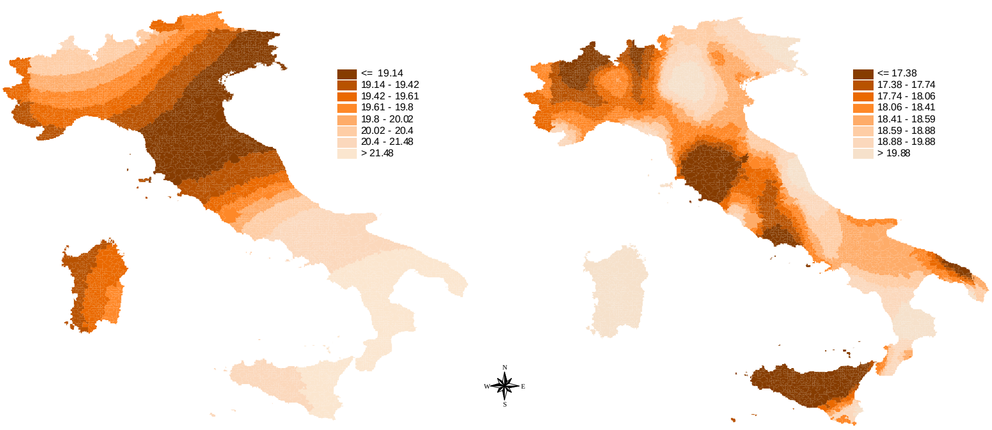

Figure 2: Frequency of 'early' first intercourse: (A) females (B) males

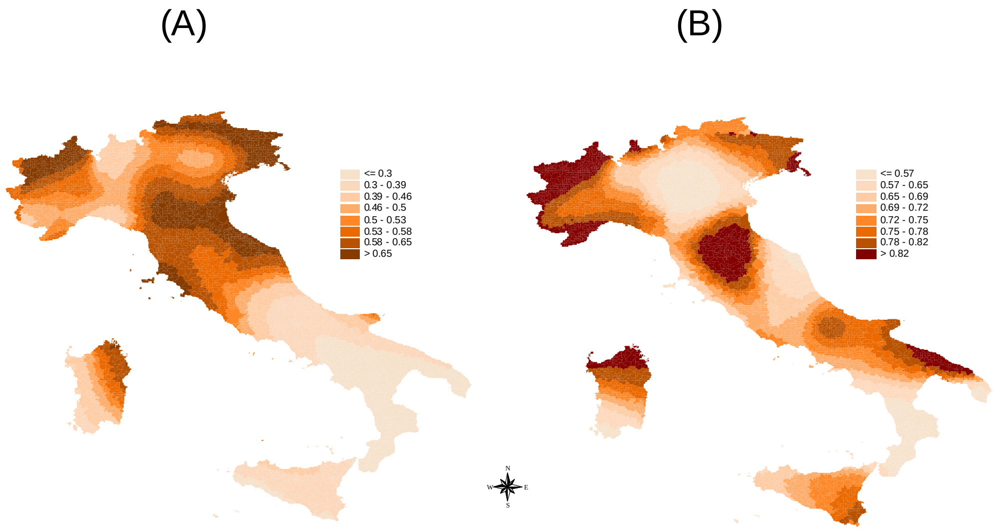

Figure 3: Spatial profile by latitude: Males

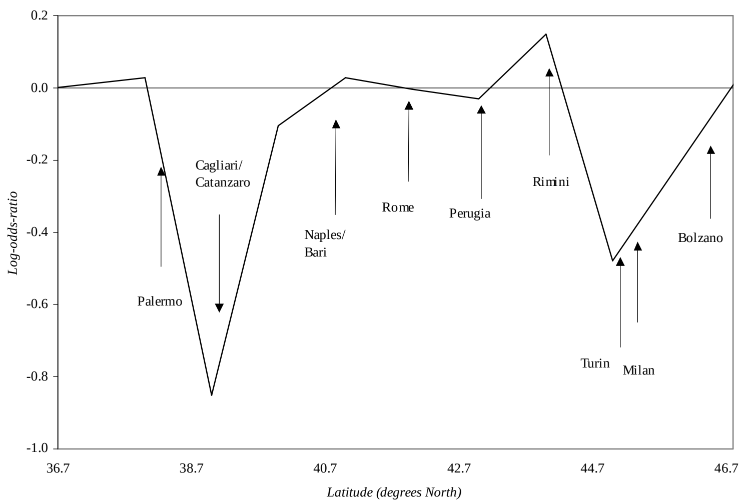

Figure 4: Spatial profile by longitude: Males

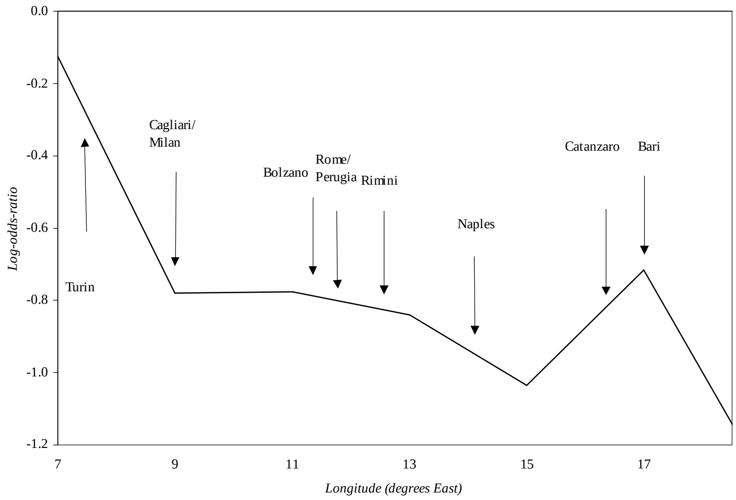

Figure 5: Spatial profiles by latitude: Females

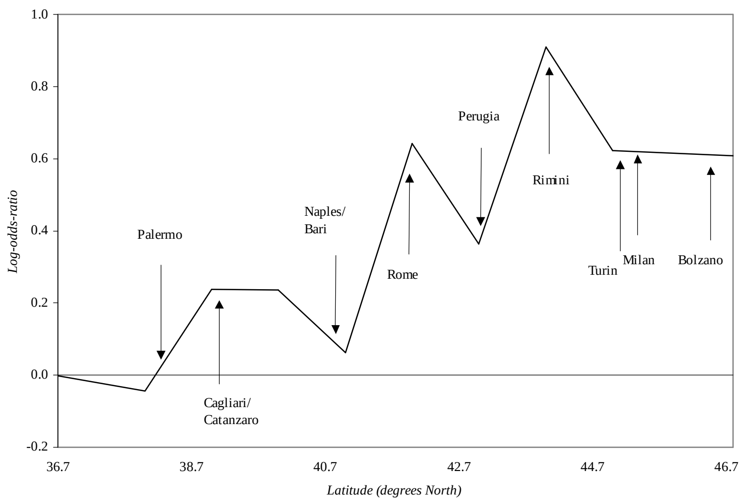

Figure 6: Spatial profiles by longitude: Females

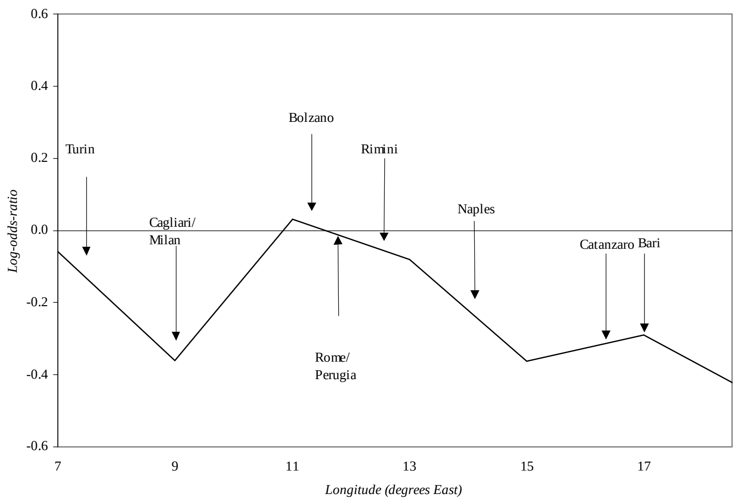

Figure 7: Map of log odds: Females

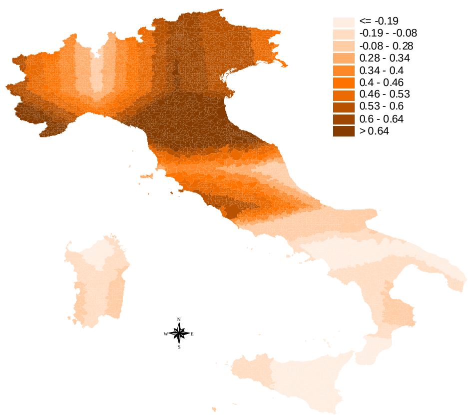

Figure 8: Map of log odds: Males

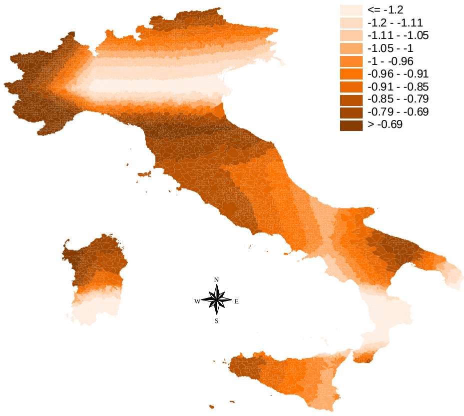

Figure 9: Maps of median age at first intercourse for older (A) and younger (B) cohort: Females

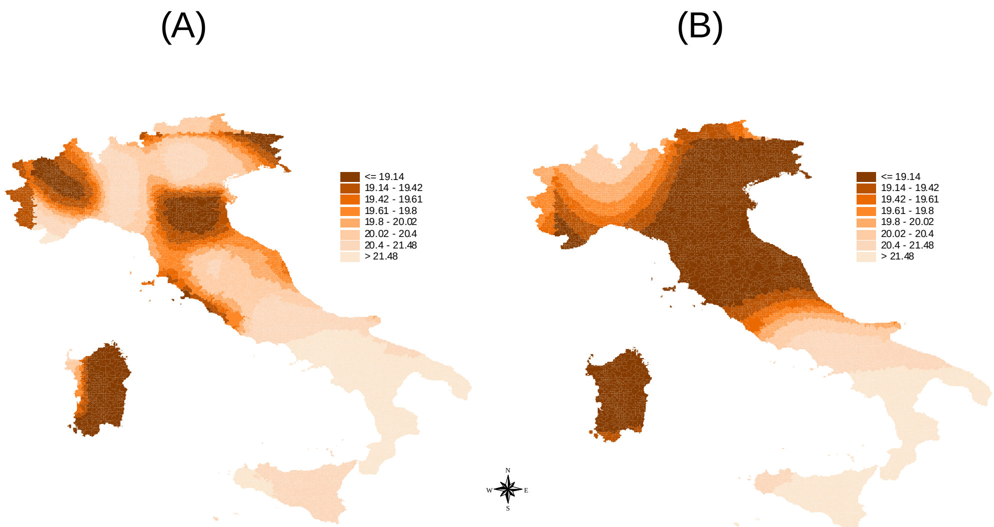

Figure 10: Maps of 'early' first intercourse: Older (A) and younger (B) cohort: Females

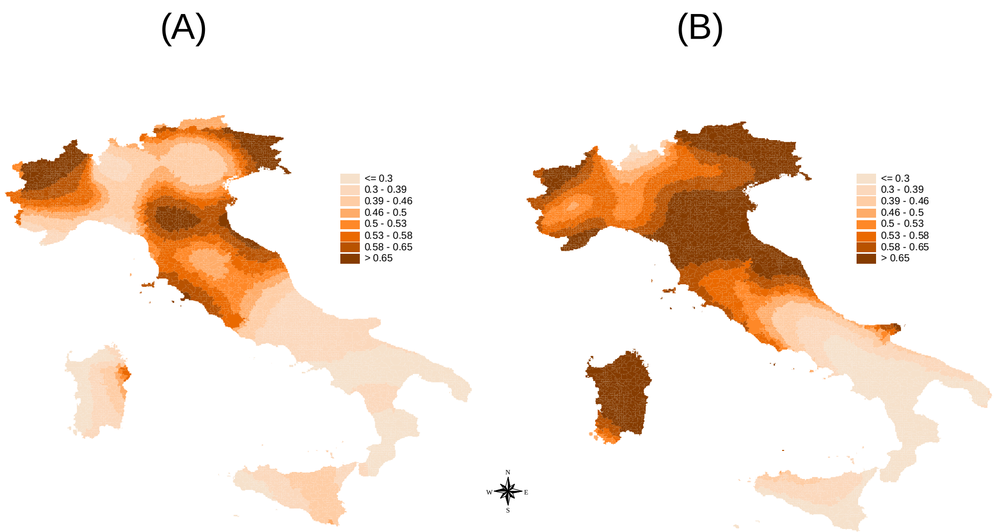

Figure 11: Standard deviation by cohort (linear spline)