Space–time Evolution of the Fertility Transition in India, 1961–1991

F. Balabdaoui,[1] J.-P. Bocquet-Appel,[2]* C. Lajaunie[2] and S. Irudaya Rajan[2,3]

[1]Centre de Geostatistique, Ecole des Mines de Paris, 35 rue Saint Honore, 77305 Fontainebleau, France
[2]CNRS, UPR 2147, 44 rue de l'Amiral Mouchez, 75014 Paris, France
[3]Centre for Development Studies, Trivandrum 695 011, India

ABSTRACT

A fertility index has been estimated which takes into account the space–time variations of mortality rates of children and women and the different fertility rates of women across age groups. The basic data are the recorded frequencies at district level for four decennial census dates (1961 to 1991) of children aged 0–4 years and women aged 15–49 years. On that basis, we have 326, 349, 400 and 454 data points on the map for each decennial census, respectively. To construct fertility surfaces on the map, the kriging technique of spatial interpolation was applied. To highlight on these surfaces the areas of regional homogeneity and zones of major change, on a subcontinental scale, a technique of image analysis and the wombling procedure of detection of the zones of abrupt change were used. The fertility maps and the variogram surfaces show that the transition took place by an expansion of heterogeneity starting from the southwestern zones. This expansion reduced the extent of the pre-transition area of homogeneity in such a way that its original transverse SW–NE direction turned into a longitudinal direction by 1991 and its size became appreciably smaller. In 1961 a long zone of abrupt change detected by wombling was observed crossing the Indian subcontinent transversely. From west to east, it extended longitudinally for a length of nearly 2500 km, including Bangladesh. This zone of abrupt fertility change divided the subcontinent into two regions, one of high fertility in the north and the other of lower fertility in the south. This latter zone is separated partially by State boundaries and geographical landmarks. A procedure of regression trees (classic AID tree) applied to the co-variables shows that this zone also corresponds to a major regional change in the female literacy rate. From 1961 to 1991, with a general fall in fertility rates all over the country, this zone of abrupt change has progressively disappeared. Copyright © 2001 John Wiley & Sons, Ltd.

INTRODUCTION

The study intends to highlight the geographical patterns of fertility and their changes in India every ten years from 1961 to 1991. Fertility transition is taking place in most parts of the country (Bhat *et al.*, 1984; Visaria and Visaria, 1997; Bhat and Zavier, 1999; Visaria and Irudaya Rajan, 1999). What are the geographical patterns of fertility since the pre-transition period? Which are the zones in transition and those which remain unchanged? Which are the demographic and sociocultural co-variables available in the censuses at the district level which are asso-

* Correspondence to: Jean-Pierre Bocquet-Appel, CNRS, UPR 2147, 44 rue de l'Amiral Mouchez, 75014 Paris, France.
E-mail: bocquet-appel@ivry.cnrs.fr

ciated with the transition?

The present paper relates mainly to the data and the demographic techniques of estimates and spatial analysis which were used to reconstitute and analyse, from the regional to the subcontinental scale, the fertility map on the different census dates since 1961. In order to reconstruct (interpolate) surfaces of fertility, as continuous variables on a map, the kriging technique (Wackernagel, 1998) was applied. To highlight on these surfaces the areas of regional homogeneity and zones of major change at the subcontinental scale, a technique of image analysis and the wombling procedure (Barbujani *et al.*, 1989 Bocquet-Appel and Bacro, 1994) of detecting the zones of abrupt change were used. The ecological correlation between fertility and the demographic, socio-economic and cultural co-variables of the censuses (1981 and 1991) are already known. This article is mainly of a descriptive nature. Its goal is to provide material for further analysis of demographic changes and their correlates.

DATA

Census Figures of Children and Women

The basic information for estimating fertility rates at the district level are the census figures of children aged 0–4 years and women aged 15–49 years for the years 1961, 1971, 1981 and 1991. The districts are located by the geographical coordinates of their capitals. From 1961 to 1991, they represent for the four decennial censuses 326, 349, 400 and 454 data points on the map, respectively. The list of the administrative subdivisions – states and districts – varies chronologically. When one refers to a district (represented by the coordinates of its capital, historically a stable geographical unit), it will be in the current administrative state division (1991) and not in the administrative subdivision at the time of the earlier censuses. In the censuses, there are two known sources of error: under-enumeration of persons in different age groups, and mis-reporting of age. The corrections proposed by post-enumeration checks (Census of India, 1982, 1994; Bhat *et al.*, 1984; Bhat, 1998), undertaken in order to take account of the missing individuals supposedly notcounted in large regions

(North, West, South, East, South-West and North-East), have only a negligible influence on the distributions of the enumerated population (about 0.2%). Therefore, these corrections were not made. More serious seems to be the problem of the mis-reporting of ages of children, an error which would produce an underestimate for the age group 0–4 years and an overestimate for the age group 5–9 years (Bhat, 1995). Percentages of under- and over-estimation estimated from smoothing of distributions have been proposed at the all-India level; such estimates do not exist, however, for district and state levels (Dyson, 1976; Census of India, 1977; Chandra, 1980). These percentages are reported to be almost invariant between censuses (Bhat, 1998). From these percentages, application of a constant correction factor to the number of children aged 0–4 years enumerated in the districts does not modify the relative distribution of the number among the districts. The geographical pattern of fertility, with or without correction, remains unchanged. For lack of local information, we have therefore not attempted to correct this under-enumeration.

Probability of Deaths and Age-Specific Fertility Rates

The calculation of fertility indices also rests on the probability of deaths and on the age-specific fertility rates. For the probabilities of death, we used the two following categories of information:

- for the censuses 1981 and 1991, the death probability for children 0–4 years at the district level (Registrar General of India, 1988, 1997);
- for the periods 1970–75, 1976–80, 1981–85 and 1986–90, life tables at the level of 15 States for the first two periods and 17 States for the last two, as well as at the all-India level for all four periods, but with no life tables at district level (Registrar General, India, 1989, 1994).

Three usual problems of missing data thus arise: to estimate (i) mortality at district level when information is available only at State level; (ii) mortality of particular States and Union Territories for which no data are

available; (iii) mortality over the period 1961–70 for all States and Union territories. The standard approach was used: (i) when the mortality of a lower level of geographical aggregation was not available (the district), one applied the mortality of the higher aggregation level; thus we assigned to the district the life table (considered as averaged) of the State to which it belongs, and to particular States the life table (considered as averaged) of India; (ii) one applied the mortality of the period 1970–75 to the previous period 1961–71.

This method of filling in the missing data tends to homogenise mortality as one moves back in time, and consequently blurs the local differences in fertility in the earlier censuses, but one can hardly find an alternative.

With regard to the co-variables, they are those available in the censuses of 1981 and 1991. They may be classified under three categories:

- demographic: population density (by km^2), urban population, infant mortality;
- cultural: average household size, which represents the number of members in a nuclear family, literacy rate and religious distribution (Hindu, Muslim and Christian) of the population;
- socioeconomic (work participation).

The frequencies in the various categories involving large numbers (hundreds of thousands of individuals) were converted into percentages, for which the standard error of estimation is negligible.

A Fertility Estimator at the Localities

To take account of the space–time variation of mortality of children and women and the difference between the relative fertility of women across age groups, an index of fertility was estimated (see Appendix A). The parameters are:

$f(x,T)$ age-specific fertility rate, from age class x to age class $x + 4$ ($x = 15, \cdots, 45$);
$P(0, X, T)$ enumerated children aged 0–4 years, at locality X;
$P(x, X, T)$ enumerated women belonging to age class x, to age class $x + 4$;

$q(x, T)$ the death probability for women;
$q(0, X, T)$ the death probability for children belonging to the age class 0–4 years.

As starting point of a period, the fifth year preceding a census, t is set to 0. The end of the period corresponding to the census is set to $T = 5$. This estimator at the locality X, at time t, is:

$$\hat{\lambda}(X, t) =$$

$$\frac{P(O, X, T)}{\Sigma \dfrac{f(x, T)P(x, X, T)}{q(O, X, T) - q(x, X, T)}(1 - e^{T[q(x,X,T) - q(O,X,T)]})}$$

(1)

One can show that the expectancy of $\hat{\lambda}(X, T) = E(\hat{\lambda}(X, T)) = 1$ (see Appendix B). It is thus an estimator without dimension. Within the same census t, the point-values account rightly for the hierarchy of fertility between localities. But it is hardly possible to compare these values between the censuses because the expectancy of the differences of $\hat{\lambda}(X, T)$ between two censuses, at T and T', is zero: $E(\lambda(X, T)) - E(\lambda(X, T')) = 1 - 1 = 0$. To compare the values between the censuses, we may either take the estimated fertility for a single age class considered as reference, or calculate $\hat{\lambda}(X, T)$ by taking a standard set of age-specific fertility rates $f(x,T)$ common to all the censuses. In this last case we find, in the absence of mortality ($q(x,X,T) - q(O,X,T) \to 0$), an estimator identical to that of Coale and Treadway (1986) of Ig, if the children aged 0–4 years at the numerator are replaced by those born during the year before the census:

$$\hat{\lambda}(X, t) = \frac{P(O, X, T)}{\Sigma Tf(x, T)P(x, X, T)}$$

(2)

In the remaining part of the paper, we follow the second approach and use the fertility estimator $\hat{\lambda}(X, T)$ which we call simply $\lambda(X)$.

The age-specific fertility rates introduced into $\lambda(X)$ were those for the all-India level for the year 1971, taken as standard reference. At the time of the census, on average, the women of the age class x (to the age $x + a$), were in the preceding age class, $x - a - 1$, at the birth of the enumerated children aged a years. To take account of the effective fertility of these

women during the period preceding the census, the age varying linearly, one considered that such was also the case for the fertility between two successive age classes and that the births were distributed uniformly over the period. Thus, the effective fertility of the women of age class x is roughly, during the period, the average fertility of the age class x and the age class $x - a - 1$: $f(x) = 1/2 \ (f(x) - f(x - a - 1))$.

Estimated Fertility at the Localities and the Censuses

The source data suffer from many uncertainties, which have undergone thorough review (Bhat *et al.*, 1984; Rele, 1987; for a summary, see Bhat 1996, 1998). The correlation coefficients between the fertility index $\lambda(X)$ for 1981 and 1991 and the total fertility rate (TFR) provided by Bhat (1996) for 1974–80 and 1984–90, estimated with the districts common to the study of Bhat (1996) and with ours, are respectively 0.924 ($n = 269$) and 0.957 ($n = 295$). Thus, beyond measuring units which are different, TFR and $\lambda(X)$, the two fertility indices, provide information which are closely correlated.

It is not easy to attribute their difference to any single factor, given the anomalies present in the source data in terms of the quantity of information and its integration by the estimators. The relationship of $\lambda(X)$ to TFR, as estimated by Bhat (1996), may be expressed by a simple linear regression of the form TFR = $a + b \ \lambda(X) + e$, with a and b being coefficients and e a residual (for 1981: $a = -0.9533$, $b = 7.2720$, $e = \pm \ 0.4053$, $r^2 = 0.882$; for 1991: $a = -0.9084$, $b = 7.1628$, $e = \pm \ 0.2595$, $r^2 = 0.953$). It is found, roughly, by taking the average of the two regressions, that TFR = 2 children gives $\lambda(X) = 0.39$, and that TFR = 6 children gives $\lambda(X) = 0.94$.

TECHNIQUES OF SPATIAL ANALYSIS

The main techniques of spatial analysis used are summarised below.

Quasi-continuous Estimate of Fertility on the Map: Kriging

The map of India is contained in spatial fields subdivided into 130×121 pixels, with a Lambers projection, to take into account the relative size of the areas when the map is projected on to a sphere. This is important when one is interested in the analysis of the distributions of demographic variables. The simple geographical projection which corresponds to a Cartesian coordinate system does not provide for such an understanding. For each period, from the observed $\lambda(X)$ values at district level, the reconstitution of the surface of the fertility index was obtained by using the procedure of spatial interpolation called kriging. Briefly let us recall that the kriging procedure is the transposition in a spatial context of the multiple linear regression, by considering that the data are outcomes of identically distributed correlated random variables. Essentially, the estimates rest on the variation of the information redundancy with the geographical distance. The measurement of this space redundancy is expressed in the form of a variogram model, fitted to the experimental variogram which is calculated from the data. The variogram model is injected into the regression, from the data points, on any point of the studied geographical space. As an interpolator based on a statistical model, the kriging technique is the best linear unbiased estimator (BLUE) (for a summary, see Wackernagel, 1998). This procedure has already been used in demography to detect and represent the spatial diffusion of contraception in Victorian Britain from the data of the European fertility project (Bocquet-Appel and Jakobi, 1998).

The observed and modelled variogram of $\lambda(X)$ used in the kriging procedure of spatial interpolation is represented in Fig. 1. The values of the parameters of the models are given in Table 1 and the maps are given in Plate 1. To incorporate the local variation to the extent possible, the estimates provided by the kriging procedure were obtained from the localities at a distance of $\leq 500 \ km$ from the point of estimation. At this distance all the variograms show that they are located in a zone of homogeneity of fertility, characterised by a reasonably important spatial redundancy. In fact, kriging in single neighbourhoods gives very satisfactory maps.
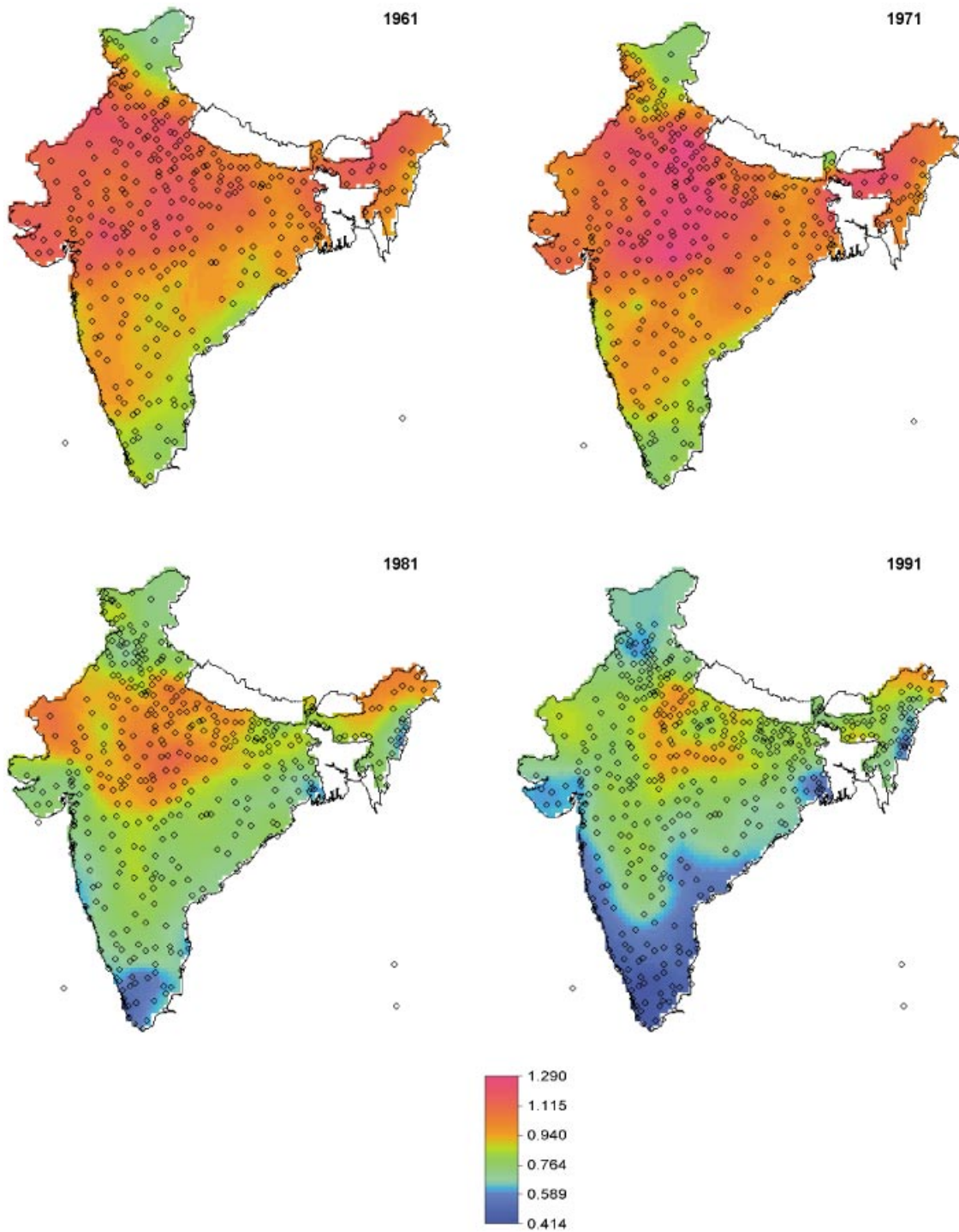
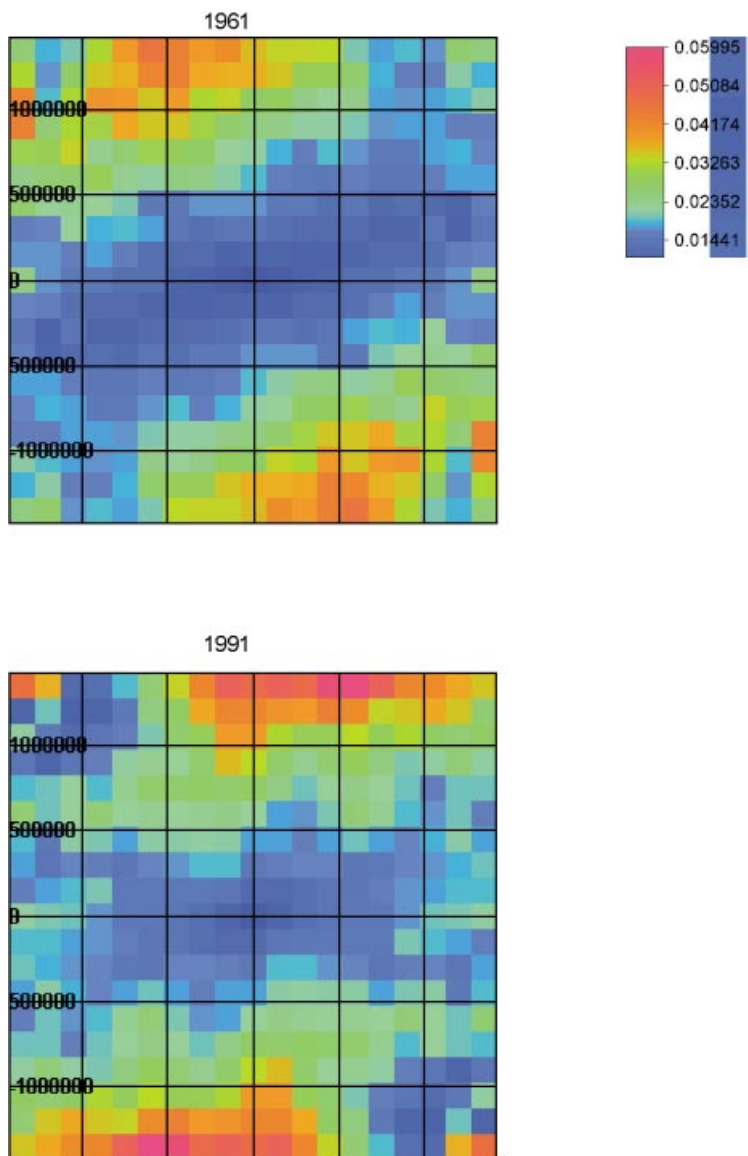Plate 1. Spatial distribution of the fertility index $\lambda(X)$ in the four censuses.

Plate 2. Observed variogram surface of the fertility index $\lambda(X)$ for the 1961 and 1991 censuses.
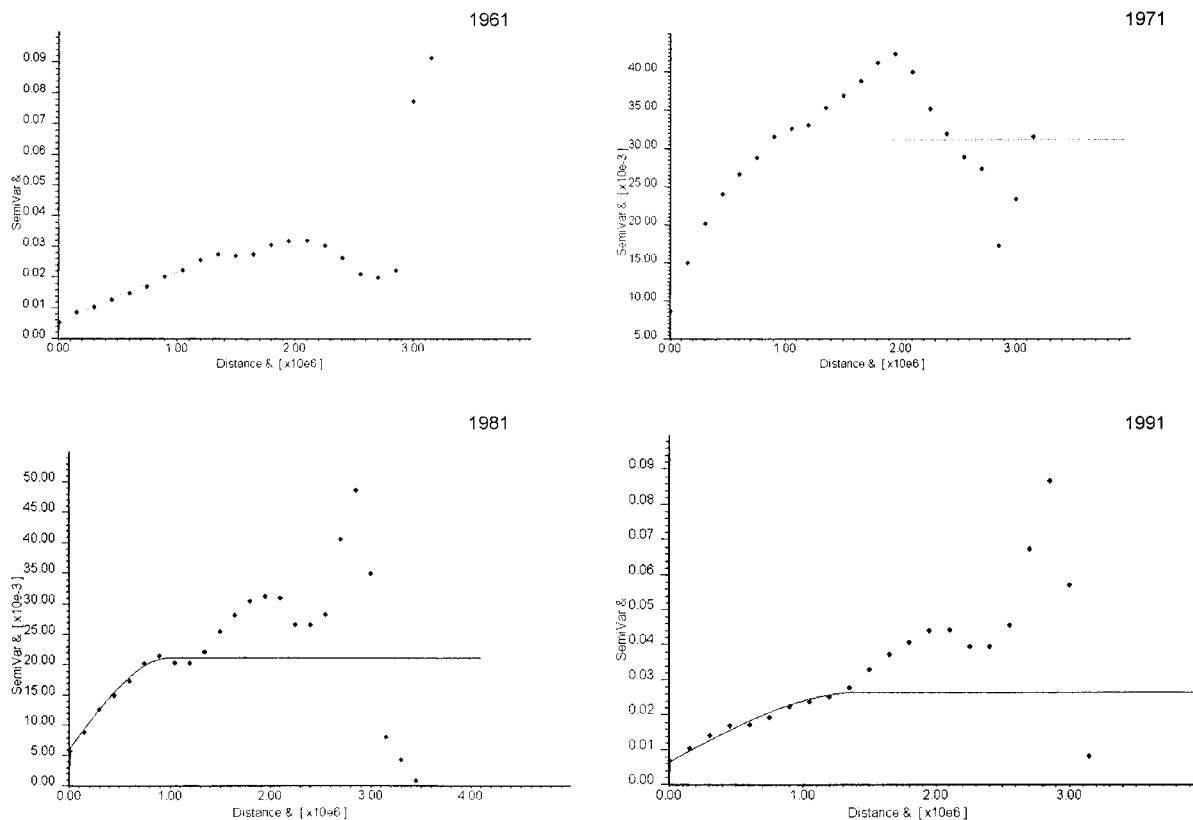
Figure 1. Observed and modelled variogram of the fertility index λ(X) in the four censuses, 1961, 1971, 1981 and 1991.

Detecting the Homogeneity Areas and Zones of Abrupt Change: Edge Enhancement Filter and Wombling

From the maps represented in continuous form (Plate 1) we sought, on the one hand, to reduce the data while emphasizing the regional areas of homogeneity, and on the other hand, to detect the zones of major change at the scale of the Indian subcontinent. To emphasize the regional areas of homogeneity on fertility surfaces, we have (i) applied the edge en-

Table 1. Parameter values of the modelled variograms to estimate the fertility index λ(X), from 1961 to 1991, by kriging.

	Periods			
Parameters	1961	1971	1981	1991
Model	Spherical	Spherical	Spherical	Spherical
Nugget*	0.005	0.008	0.008	0.013
Sill**	0.030	0.031	0.028	0.058
Fitting distance***	2100	900	1100	1900
Interpolation distance	500	500	500	500

* Variance at nil distance.
** Maximum variance at the fitting distance.
*** Distance in km.

hancement filter (Marr and Hildreth, 1980), and (ii) regrouped the filtered values in class groups at steps of 0.1. Briefly, this filter consists of withdrawing the pixel value of the Laplacian from the eight surrounding pixels in fields 3×3. Its effect on the map is to increase the intensity of contrasts between contiguous areas of homogeneity. The map appears then in the form of homogeneity areas, of which the arbitrary number depends on the amplitude of the state variable and the increment of the value classes.

To highlight the zones of major change at the subcontinental scale, the wombling procedure of detection of the zones of abrupt change (also called 'barriers'; Barbujani *et al.*, 1989; Bocquet-Appel and Bacro, 1994) was used. Briefly, from the gradient evaluated at all points of the surface variable, which represents the rate of change between contiguous pixels, the procedure selects the pairs of contiguous pixels whose values of the gradient (the rate of change) are at a maximum relative to the gradient distribution on the surface. This technique has already been used in demography to detect the wave of the fertility transition in Europe (Bocquet-Appel and Jakobi, 1996).

DATA ANALYSIS

Univariate Distribution of the Fertility Index λ(X) and the Fertility Transition

Theoretically, on a horizontal axis representing the range of the possible values of a fertility index, one must expect that the pre-transition distribution (with the vertical axis as the frequency and the horizontal axis the index) of the local values of this index is located in the upper part of the values, and that the post-transition distribution lies in the lower part. The transition must be expressed by a displacement from the upper towards the lower part of the distribution (Bhat, 1996; Bocquet-Appel and Jakobi, 1997). If the geographical area of the generating variables of fertility (primarily socioeconomic and anthropological) is homogeneous, the pre-transition distribution should be unimodal and regular. If this area is not homogeneous, the distribution could be bimodal or even multi-modal, each mode corresponding to a regrouping of local-

ities presenting some regional differences. The post-transition distribution, whose values of the index fluctuate around or below the replacement level of the generations, should be unimodal after a flattening of the regional heterogeneity during the transition.

The histograms of the $\lambda(X)$ values of the four censuses are represented in Fig. 2 and the main parameters of the distributions in Table 2. The year 1961 exhibits indications of bimodality. As indicated above, the years 1961 and 1971 have approximations of missing values which could have influenced the distributions. But the bimodal aspect of a distribution, such as that of 1961, in a heterogeneous space geographically and culturally like the Indian subcontinent, is not necessarily the mark of an anomaly in the data. It could be an indication of the presence of two great zones of homogeneity. Such a bimodality, if it is not an artefact, must disappear with the transition which plays a homogenising role. The distribution in 1981 appears regular, as already noted by Bhat (1996), and without any noticeable dissymmetry. It moves towards the low fertility values. The distribution for 1991 is dissymmetrical on the left, towards the low values of the fertility index. It clearly represents a distribution in the course of transition. From 1961 to 1991, at each census, the proportion of the districts having a low fertility index ($\lambda(X) \leq 0.39 = \text{TFR} \leq 2$) and those having a high fertility index ($\lambda(X) \geq 0.94 = \text{TFR} \geq 6$) indicated on the histograms, account for 0–72%, 0–71.8%, 0.3–21.1% and 3.1–9.1% of the distributions, respectively. Since 1981 (in fact, from 1976–80), one can observe a substantial fertility reduction in the districts. The whole of the transition kinetics in progress is well represented by the kernel distributions of the $\lambda(X)$ values in the four censuses (Fig. 3). In 1971 (in fact, from 1966–70), the distribution of $\lambda(X)$ appears unimodal and regular. Perhaps this distribution was bimodal in 1961 (1956–60). But, if the distribution in 1971 is correct, then the absorption speed of the bimodality of 1961, which corresponds to an important demographic change, could have been quite rapid. Then, after 1971, one can observe the predicted shift of the distribution of $\lambda(X)$ towards lower values.

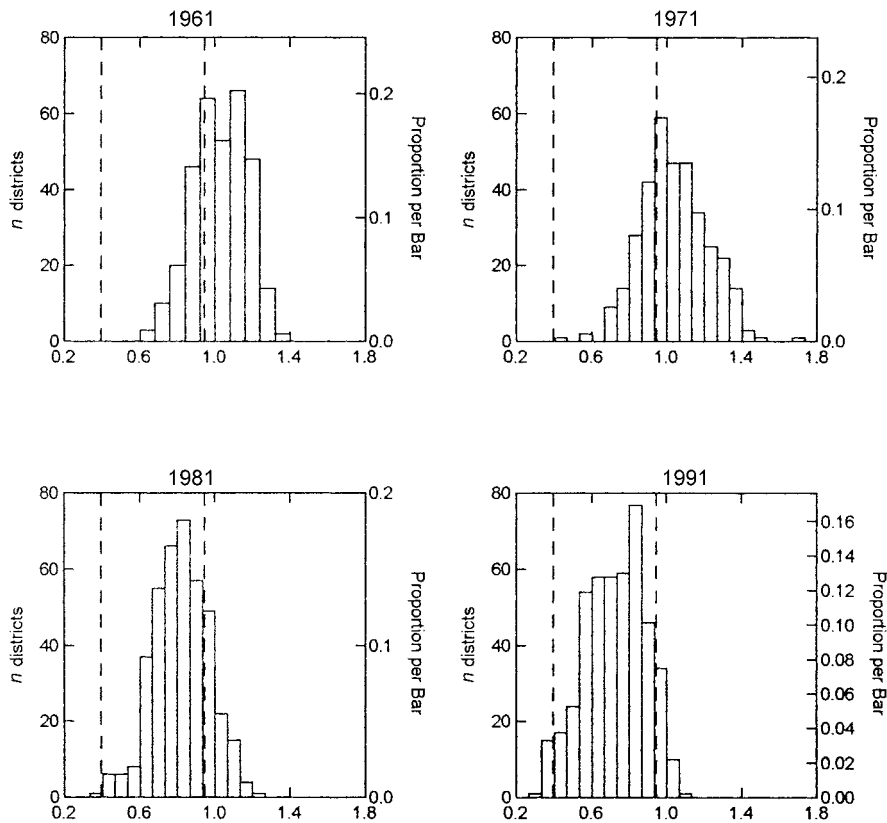Another possible approach to measuring the

Figure 2. Histograms of the values of the fertility index $\lambda(X)$ in the four censuses.

importance of the total change of fertility during the period is through the resemblance of the values of $\lambda(X)$ between censuses. This resemblance can be formalised, in a simple way, by correlations. Logically, pre-transition fertility is in a stationary state, which implies a resemblance between the pre-transition censuses. Then, with the geographical extension of the transition, one must witness a decline in the resemblance between the successive cen-

suses during the transition phase. One thus expects that resemblance: (i) is closer between the pre-transition censuses than between the censuses in transition; (ii) is weak between two successive censuses in which one is pre- and the other post-transition; and (iii) decreases with the chronological distance between the censuses in the transition period.

The correlation matrix between censuses is given in Table 3. Propositions (ii) and (iii) are

Table 2. Main statistical parameters of the values of the fertility index $\lambda(X)$ estimated at the districts for the four censuses (1961, 1971, 1981 and 1991).

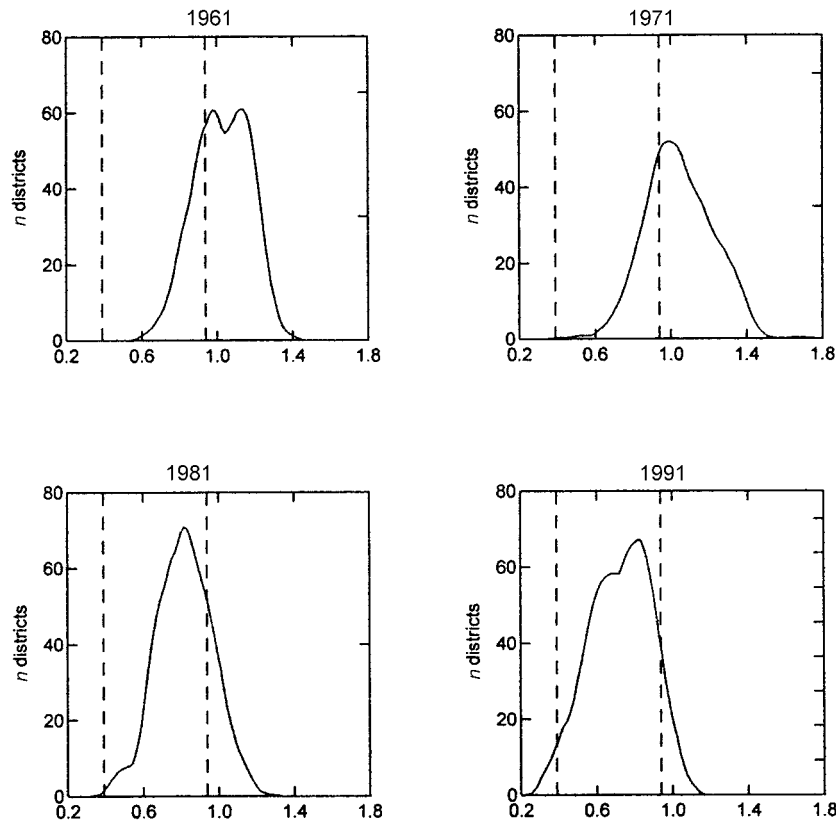|	LAMBDA_61	LAMBDA_71	LAMBDA_81	LAMBDA_91
n	326	349	400	454
Minimum	0.624	0.444	0.399	0.302
Maximum	1.370	1.669	1.247	1.086
Range	0.746	1.225	0.848	0.784
Median	1.030	1.025	0.814	0.731
Mean	1.024	1.040	0.817	0.722
S.D.	0.144	0.177	0.146	0.163

Figure 3. Kernel distribution of the values of the fertility index $\lambda(X)$ in the four censuses.

confirmed, but not proposition (i). The two pre-transition censuses do not have a closer resemblance than with a census in the transi- tion phase. This finding confirms the anomaly of spatial variance observed below, concerning the values of the variograms of 1971 and 1981.

Table 3. Pearson correlation matrix between the census values of the fertility index $\lambda(X)$.

	LAMBDA_61	LAMBDA_71	LAMBDA_81	LAMBDA_91
LAMBDA_61	1.000			
LAMBDA_71	0.837	1.000		
LAMBDA_81	0.652	0.720	1.000	
LAMBDA_91	0.635	0.697	0.881	1.000

Barlett chi-square statistic: 1085.573, d.f. = 6, Prob. = 0.000.

Pairwise frequency table: 1961, 1971, 1981 and 1991

	LAMBDA_61	LAMBDA_71	LAMBDA_81	LAMBDA_91
LAMBDA_61		326		
LAMBDA_71		325	349	
LAMBDA_81		316	338	400
LAMBDA_91	314	337	385	454

Comparison of the Simple (Isotropic) and Directional Variograms (Surface) and the Fertility Transition

Theoretically, on a surface of homogeneous pre-transition fertility, one must expect that the transition generates a geographical heterogeneity, until the reappearance of a homogeneous post-transition surface. This heterogeneity of transition must be expressed in the form of an increase of the spatial variance in time. This must be represented by simple (isotropic) variograms following a chronological hierarchy, with relatively low values at the beginning of the transitional process, followed by high values during the process and then again by low values when the post-transition state is reached.

The simple observed variograms of $\lambda(X)$, for the four periods, are represented together in Fig. 4. The maximum distance in the variograms on the x-axis was determined by the observed number of pairs of localities ($\geqslant 100$) in the greatest distance class (2700 km). The observed variograms (Fig. 4) show a continuous and relatively regular increase in the spatial variance up to 2000 km, after which the variance fluctuates. Between the two bounds of the period, 1961 and 1991, if one sees a substantial increase in variance, the observed chronological hierarchy differs from the one expected. The pre-transition year 1971 exhibits a spatial variance somewhat higher than that of the year in transition, 1981, while its variogram profile is similar to that of the pre-transition year 1961 (no marked heterogeneity at long distance). For the year in transition, 1981, it is the reverse: spatial variance lower than in the pre-transition year 1971, and variogram profile similar to another year in transition, 1991. Relative to the expected chronological pattern, there is an inconsistency: whereas the procedure used to fill the missing data in 1971 might have caused a fall in the true variance (see above), either 1971 has too much variance, or 1981 does not have enough. Visaria (1971) blames the bad quality of the 1971 census data. This would imply an excess of spatial variance for 1971.
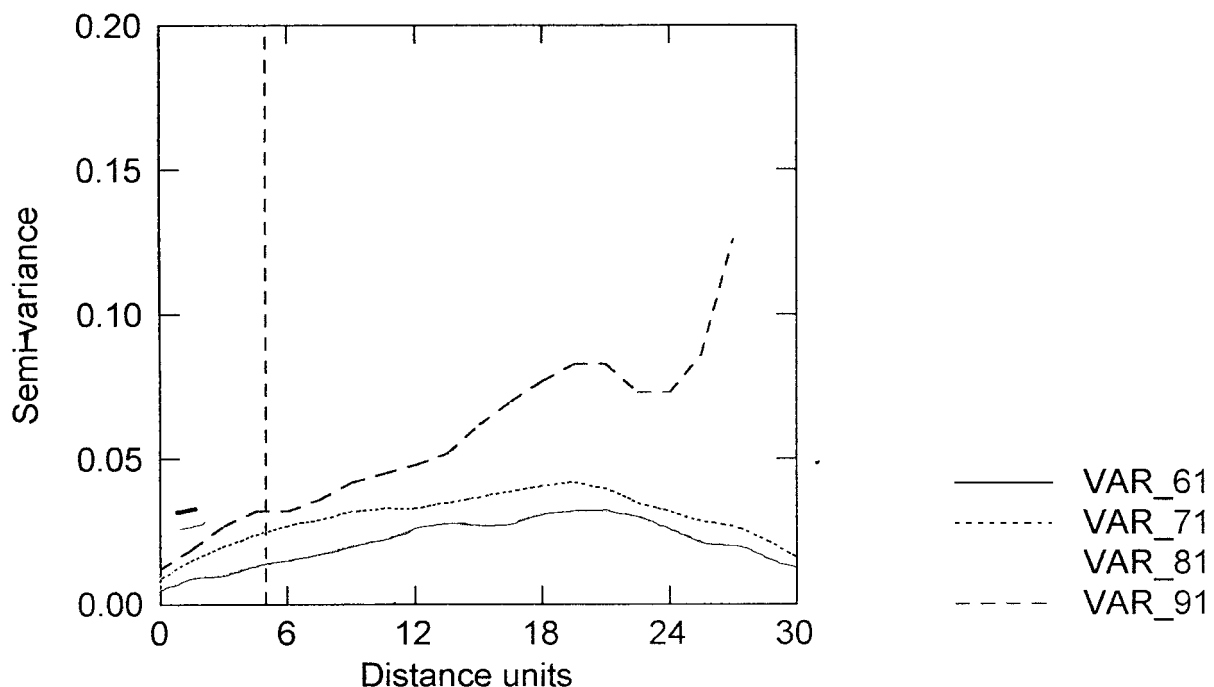
Let us recall that the directional variogram

Figure 4. Observed simple variogram (isotropic) of the fertility index $\lambda(X)$ in the four censuses (distance units: 100 km).

(Pannatier, 1996) is intended to provide the main anisotropy directions of information, when these directions are not known *a priori* (0–90°, with 0° = 180°: North–South = South–North; 90° = 270°: East–West = West–East). Logically, if the transition appears locally, this must be reflected in an increase in variance in the corresponding zone of the variogram, and modify the original pre-transition configuration of the geographical fertility pattern. For a demographic variable, one must expect that the change in the geographical pattern is relatively slow.

The directional variograms of $\lambda(X)$ are represented in Plate 2 to the two bounds of the period, 1961 and 1991.

The pre-transition period (1961) exhibits a distribution of the homogeneity areas, at high fertility, on a SW–NE axis, in a band of about 950 km width. On this axis, starting from the centre of the variogram, at the distances (bi-directional) of 0 km, 500 km, 1000 km and 1500 km, the variance is at a minimum, with successive symmetrical values (×10) of 0.053, 0.107, 0.104, and 0.153, respectively. This variogram is relatively flat. The variance is at a maximum in the SE–NW direction, with successive symmetrical values (×10) of 0.053, 0.168, 0.382 and 0.391 respectively.

The period in transition (1991) is characterised by a rotation (of approximately 22°) of the axis of the homogeneity area of the pre-transition period before 1961, which has now become longitudinal (E–W). At bi-directional distances of 0 km, 500 km, 1000 km and 1500 km, the variance takes successive symmetrical values (×10) of 0.069, 0.131, 0.194 and 0.213. The homogeneity bandwidth has decreased to approximately 500 km. The variance is at a maximum on the latitudinal axis, in the N–S direction, with (symmetrical) values (×10) of 0.069, 0.180, 0.386 and 0.486. A zone of homogeneity, in the SE–NW direction, appears at a distance of 1000–1500 km. A comparison of the fertility maps and the variogram surfaces for 1961 and 1991 shows that the transition took place by an expansion of heterogeneity from the South-Western zone, through a reduction in the area of homogeneity of the pre-transition period, in such way that its SW–NE original transverse direction became longitudinal in 1991, and appreciably narrower.

Homogeneity Areas

Figure 5 shows the homogeneity area of fertility in India in the four censuses, obtained by the edge enhancement filter as described above. A map of Indian States is shown in Fig. 7. From the maximum fertility area (MFA) of the pre-transition period, we are interested in identifying the successive areas of fertility traversed in both directions, longitudinal and latitudinal, map after map. We will also locate the areas of low fertility. This first approach makes it possible to account for the change of regional fertility while going, roughly, from the centre towards the periphery of India. The MFA is defined as the zone in which the fertility is the highest and whose homogeneous area is the largest on the map, compared with any other zones of high fertility.

1961 and 1971: The two maps are supposed to represent a pre-transition pattern, presumably not very different from each other. However, they are not identical (see Fig. 5, 1961 and 1971). Apart from other differences, the two maps do not show the same MFA, and their maximum fertility differs. It is lower in 1961 ($\lambda = 1.2$) than in 1971 ($\lambda = 1.3$). The increase in maximum fertility at the MFA between the two dates, whether it is due to a bias in the data, or it represents reality, could be related to the mortality transition. Recall that for some populations, the mortality transition would show in the first phase of an increase in fertility (Dyson and Murphy, 1985). Although located in the same broad area, North-West India, the locations of the MFAS for 1961 and 1971 do not coincide. The shift of location is important. For a phenomenon like the fertility pre-transition, which is supposed to have a certain space–time inertia, this shift in location is not easily explained by random fluctuations during the ten years that separated the two censuses. The MFA for 1961 covers the north of Rajasthan, Haryana, and the western part of Uttar Pradesh extending up to the Nepal border (district of Pilibhit). It also covers north of Madhya Pradesh (Morena) and its border reaches, in the South-West, the district of Bikaner (the Rajasthan). The MFA for 1971 goes, from north to south, from the district of Jalaun to Muzaffarnagar and, from east to
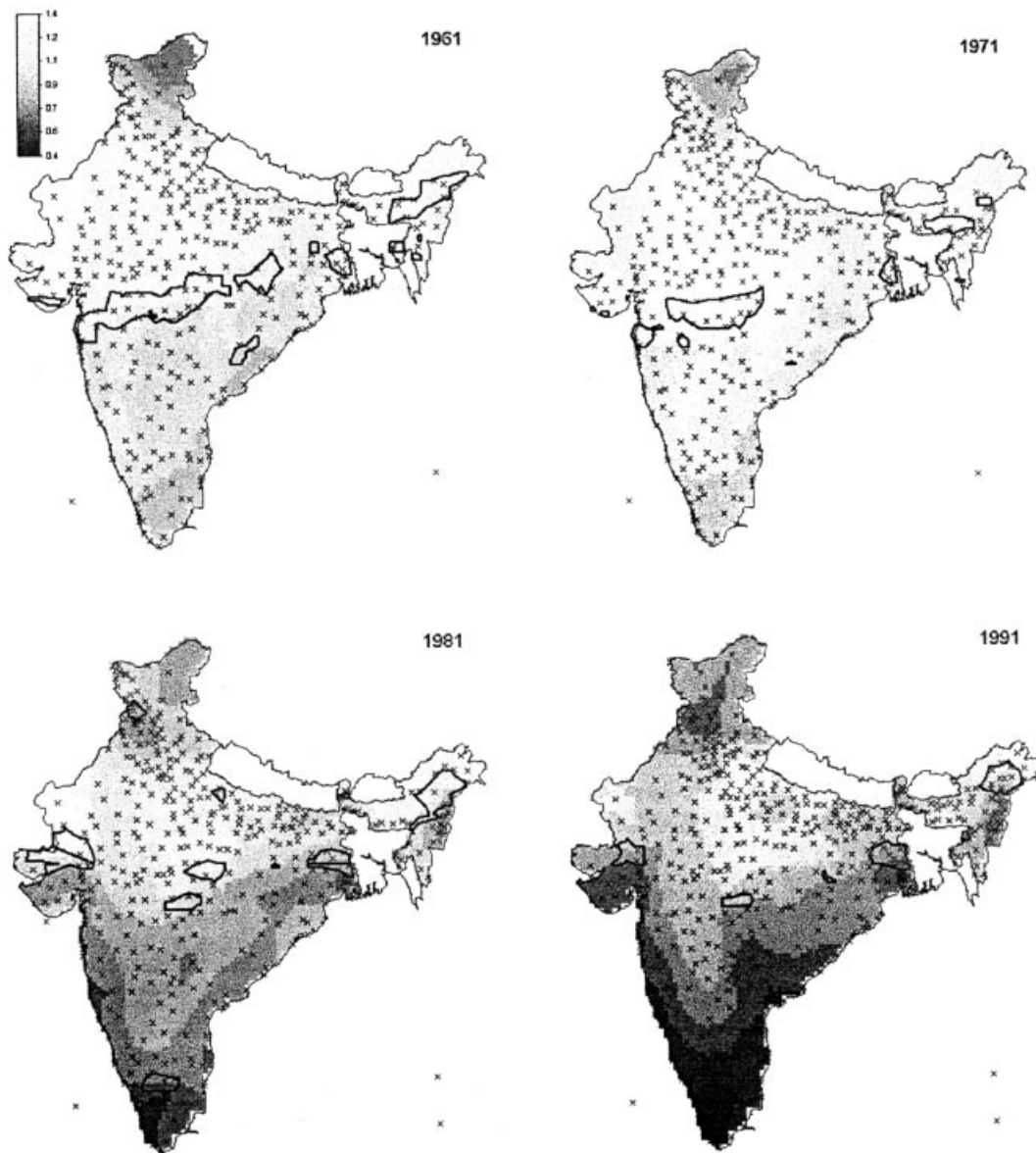
Figure 5. Spatial distribution of the fertility index $\lambda(X)$ in areas of homogeneity and zones of abrupt change detected by wombling, in the four censuses. *Note*: bold lines show the limits of the zones of abrupt change.

west, through the districts of Guna to Panna (80–77° E) in Madhya Pradesh. One is thus uncertain about the reality of the two observed MFA pre-transition scenarios and the uniqueness of their patterns. In East India, starting from an area at very high fertility levels located in the state of Assam and that of Meghalaya ($\lambda = 1.2$), fertility declines slowly in all directions, through three successive areas, until reaching the value $\lambda = 1.0$ in the coastal regions

and $\lambda = 1.1$ on the continental borders. Generally, fertility decreases from the centre of the country towards its periphery. This decrease is more accentuated latitudinally (λ minimum = 0.7) than longitudinally (λ minimum = 1.0). The area of lowest fertility is located in Kerala (districts of Alappuzha, Kottayam, Kollam; $\lambda = 0.745$). There are two other areas at moderately low fertility, one small area in Maharastra (district of Ratnagiri, $\lambda = 0.85$), and the

Copyright © 2001 John Wiley & Sons, Ltd.

Int. J. Popul. Geogr. **7**, 129–148 (2001)

other covering Kerala, much of the neighbouring state of Tamil Nadu, reaching up to the latitude of Salem and continuing in a coastal band whose width progressively narrows, extending to Andhra Pradesh (district of East Godavari).

1981: Fertility decreases everywhere. The MFA of the preceding census is maintained geographically in the centre of Madhya Pradesh (districts of Lalitpur, Tikamgarh, Damoh and Panna) but its maximum value has dropped, passing from $\lambda = 1.2$ to $\lambda = 1.1$. From this MFA, fertility declines in all directions. In East India, an area with high fertility is located in Arunachal Pradesh ($\lambda = 1.0$), from where fertility decreases in all directions while crossing three successive areas of decreasing fertility on 300 km (ex: district of Manipur, λ min = 0.7). From the centre of the country, the high fertility areas ($\lambda = 1.10–1.0$) are dilated longitudinally and those with low fertility ($\lambda = 0.8–0.6$) are dilated latitudinally. The area at low fertility of the preceding census is now below replacement level ($\lambda \leq 0.6$) and its area has extended. It covers half of Kerala, from its southern point to Malappuram, and extends in the direction of the district of Salem in Tamil Nadu. It is surrounded by a zone with moderate fertility ($\lambda \leq 0.7$) in Karnataka and Tamil Nadu. There are now four zones of moderate fertility, all located along the coasts, in the west (district of Ratnagiri, Maharasthra) and in the east, around Madras (Tamil Nadu), Calcutta (West Bengal) and in East India (Manipur and in the south of Nagaland).

1991: Fertility decreases everywhere. The pattern of the previous period remains approximately the same, but is contracted longitudinally. The MFA ($\lambda = 1.0$) covers a small portion including the mid-west of Uttar Pradesh (districts of Rampur, Bareilly, Badaun and Etah). The MFA of the previous period now has a depression in its centre, representing an area of low fertility ($\lambda = 0.8$; district of Kanpur Nagar and Dehat, Unnao, Lucknow, Bara Banki and Rae Bareli). The fertility area below the replacement level ($\lambda \leq 0.6$) has extended considerably in the south of the subcontinent, along the coastal zones, forming an arc from the district of Thane on the east

coast (Maharashtra) to Vizianagaram on the west coast (southeast of Andhra Pradesh and southern parts of Gujarat, West Bengal, Manipur and Nagaland). It covers roughly 20% of the surface of the country. The two other areas are located around Calcutta in West Bengal and in Manipur. Another area of fertility close to replacement level appears in Punjab.

Zones of Abrupt Change

Figure 5 shows the zones of major change of fertility in India, at the four dates, detected by wombling, superimposed on the homogeneity areas. In these zones the local spatial variance is the greatest. For reasons of space and degree of definition of the local co-variables, only the zones of change whose length size is greater than 100 km will be discussed.

1961: A long zone of abrupt change crosses the Indian subcontinent transversely. From west to east, it extends longitudinally for nearly 2500 km, including Bangladesh. From its western to its eastern pillar, it is positioned across the borders of Gujarat–Maharashtra, Madhya Pradesh–Maharashtra, and longitudinally across Madhya Pradesh (while roughly following the districts of Junagadh, Dadra and Nagar Haveli, Dhule, Jalgaon and Amaravati). In this first portion, this zone of abrupt demographic change corresponds to a geographical obstacle: the southern bank of Narmada river while following the Satpura Ranges in the direction of Ambikapur (district of Surguja, east of Madhya Pradesh). Then it passes to the north of Calcutta (districts of Seoni, Mandla, Surguja, Burdwau, Balaght) and continues to the eastern border of Assam (districts of Kamrup, Mokokchung).

1971: The long transverse zone of the preceding census is still present in its western portion (borders of Gujarat–Maharashtra, Madhya Pradesh–Maharashtra), as well as in the east (northern region of Calcutta and southern border of Assam), but it has disappeared in its centre-east portion, across Madhya Pradesh.

1981: The long transverse zone of abrupt

change is still perceptible, but it is fragmented and its overall surface area has decreased. Its western pillar has moved on to a higher latitude. From west to east, one first meets a fragment on the Gujarat border (district of Banas Kantha, Sabar Khanta) and Rajasthan (district of Sirohi), two fragments in Madhya Pradesh (Chhindwara and Jabalpur), a fragment in West Bengal, the north of Calcutta (districts of Birbhum, Nadia), and finally a fragment in eastern Assam (district of Lower Subansini, Mokokchung, Tuensang). Two other zones of abrupt change, which did not exist before, are present, one in the north of India, on the border separating Punjab and Jammu and Kashmir (districts of Gurdaspur and Kathua) and the other in south India, in the north of Tamil Nadu, as well as another small zone longitudinally across Kerala (district of Thrissur).

1991: The vestiges of the long transverse zone of abrupt change of fertility are still perceptible, if punctuated, in north Gujarat (district of Banas Kantha, Sabar Khanta), north Madhya Pradesh (district of Jabalpur), West Bengal, the north of Calcutta (districts of Birbhum, Nadia), and Eastern Assam (district of Lower Subansini, Mokokchung, Tuensang).

To summarise, from 1961 to 1991 one witnesses the progressive disappearance of the zone of abrupt change of fertility, which transversely separated the north from the south of India.

Ecological Correlations of Fertility with the Co-variables 1981 and 1991

Tables 4 and 5 show the correlations between the demographic, socioeconomic and cultural variables, and the first four axes of a principal component analysis (PCA) – a PCA by census – after a varimax rotation to obtain a simple structure. To take into account the geography, among the co-variables the coordinates of the districts (and their products) have been added. These four axes account for 70.4% and 73.1% of information available, for each of the two censuses respectively. The two PCAs show the same decomposition of information, as well as the usual ecological correlations be-

tween fertility and these co-variables (see, for example, Visaria and Visaria, 1997; Bhat and Zavier, 1999). By retaining only the values of the correlations $\geq |0.5|$ between the factorial axes and the variables, on axis 1, fertility (represented by the $\lambda(X)$ values or by the

Table 4. PCA of the fertility variables and their correlations (rotated loading matrix) with the co-variables after varimax rotation (1981).

	1	2	3	4
LR_T	**−0.950**	0.066	0.019	−0.148
LR_F	**−0.937**	0.040	0.110	−0.174
LR_M	**−0.915**	0.060	−0.094	−0.074
LR_RT	**−0.885**	0.061	0.047	0.349
LAMBDA_81	**0.832**	0.114	−0.179	0.172
GROUP_FEC	**0.805**	0.104	−0.183	0.192
INFANT_MORT	**0.675**	−0.149	−0.275	0.107
LAT	**0.659**	0.284	0.199	0.128
LOLA	**0.656**	0.265	0.403	0.109
LR_UT	**−0.641**	0.042	−0.087	0.001
WPR_T	0.071	**−0.930**	0.166	0.121
WPR_F	−0.052	**−0.847**	0.217	0.250
WPR_M	0.298	**−0.788**	−0.069	−0.204
AHS_TOTAL	0.186	**0.669**	−0.096	0.142
ST_PCT	0.149	**−0.524**	**0.705**	0.117
MUSLIM_PCT	0.061	0.506	0.022	−0.289
HINDU_PCT	0.077	0.057	**−0.875**	0.035
CHRIST_PCT	−0.300	−0.123	**0.792**	0.136
LONG	0.203	0.014	**0.727**	−0.041
POP_DENSITY	−0.128	0.066	0.073	**−0.840**
URBAN_PCT	−0.355	0.095	−0.167	**−0.726**
SC_PCT	0.240	0.413	−0.485	0.071
Variance explained by rotated components	6.969	3.597	3.150	1.781
% total variance explained	31.679	16.348	14.318	8.094

Note: Numbers in boldface are correlations $\geq |0.5|$. Variables: LR_T, LR_F, LR_M, LR_RT, LR_UT are the literacy rates for the total, for females, males, rural and urban, respectively; GROUP_FEC represents the homogeneity area of the fertility index $\lambda(X)$; INFANT_MORT is Q(5); LAT, LONG and LOLA, the latitude, longitude and their product; WPR_T, WPR_F, WPR_M, are the worker participation rate in total, and for the females and males, respectively; AHS_TOTAL represents the average household size; ST_PCT and SC_PCT are the percentages of Scheduled Tribes and Castes respectively; MUSLIM_PCT, HINDU_PCT, CHRIST_PCT, are the percentages of Muslim, Hindu and Christian religions, respectively; POP_DENSITY is the density of the population; URBAN_PCT is the percentage of urban population in the district.

Table 5. PCA of the fertility variables and their correlations (rotated loading matrix) with the co-variables after varimax rotation (1991).

	1	2	3	4
LR_T	**−0.959**	0.011	0.056	−0.082
LR_F	**−0.939**	0.059	0.167	−0.138
LR_M	**−0.939**	−0.028	−0.083	0.010
LR_RT	**−0.886**	−0.084	0.045	0.299
LAMBDA_91	**0.837**	−0.272	0.019	0.222
GROUP_FEC	**0.828**	−0.256	0.021	0.233
INFANT_MORT	**0.743**	0.147	−0.090	0.215
LR_UT	**−0.729**	0.071	0.055	0.091
LAT	**0.568**	−0.478	0.083	0.249
LOLA	**0.566**	**−0.509**	0.284	0.212
WPR_T	0.050	**0.854**	0.142	0.423
WPR_M	0.109	**0.825**	−0.117	−0.142
WPR_F	−0.018	**0.718**	0.203	**0.547**
AHS_TOTAL	0.443	**−0.693**	−0.116	−0.013
HINDU_PCT	0.074	0.101	**−0.881**	0.115
CHRIST_PCT	−0.235	−0.021	**0.847**	0.124
ST_PCT	0.169	0.316	**0.807**	0.275
LONG	0.214	−0.282	**0.687**	−0.038
SC_PCT	0.148	−0.274	**−0.634**	0.027
POP_DENS	−0.105	0.058	0.076	**−0.736**
URBAN_PCT	−0.383	0.076	−0.096	**−0.646**
MUSLIM_PCT	0.097	−0.303	−0.057	**−0.566**
Variance explained by rotated components	7.118	3.437	3.266	2.264
% total variance explained	32.355	15.621	14.844	10.292

Note: Numbers in boldface are correlations ⩾|0.5|. Variables as in Table 4.

homogeneity areas of the maps, labelled GROUP_FEC), we find that they are positively correlated with infant mortality and negatively with all the variables expressing the literacy rate (labelled LR_T, LR_F, LR_M, LR_RT, LR_UT: see Tables 4 and 5 for the variable labelling definitions). Fertility is also positively correlated with latitude. The three other factorial axes, which represent mixtures of socioeconomic and cultural information, are without apparent link to fertility. The correlations of these axes will therefore not be commented upon.

Barriers and Co-variables

The oldest co-variables that we have are those

for the 1981 census. Thus, the relationship between what appears as the vestiges, observed in 1981, of the barrier of the pre-transition fertility 1961–71 and the co-variables of the 1981 census was analysed. In order to capture the effect of the pre-transition barrier only – and not of the new zones of change which appear subsequently on the map – the four transverse fragments located roughly on the barrier zone of 1961 were retained, excluding the easternmost fragment, located in East India, where the co-variables are missing. Then, the districts located up to a maximum distance of 100 km from the fragment limits of the pre-transition barrier were categorised into two groups. The north–south limit of the fragments was determined relative to the longest axis (longitudinal) of each fragment (see Fig. 5, 1981). The first group includes the districts located north of the barrier, and the second those located to the south. The two central fragments present a difficulty: they are almost parallel, so that the districts located between these two fragments are simultaneously south of one fragment and north of the other. They can be classified in both groups. Thus, the districts located between these two fragments were eliminated, to retain only those located north of the northern fragment (north of the barrier), and those located south of the southern fragment (south of the barrier). The districts for which the co-

Table 6. List of districts located north and South of the fertility barrier in 1981.

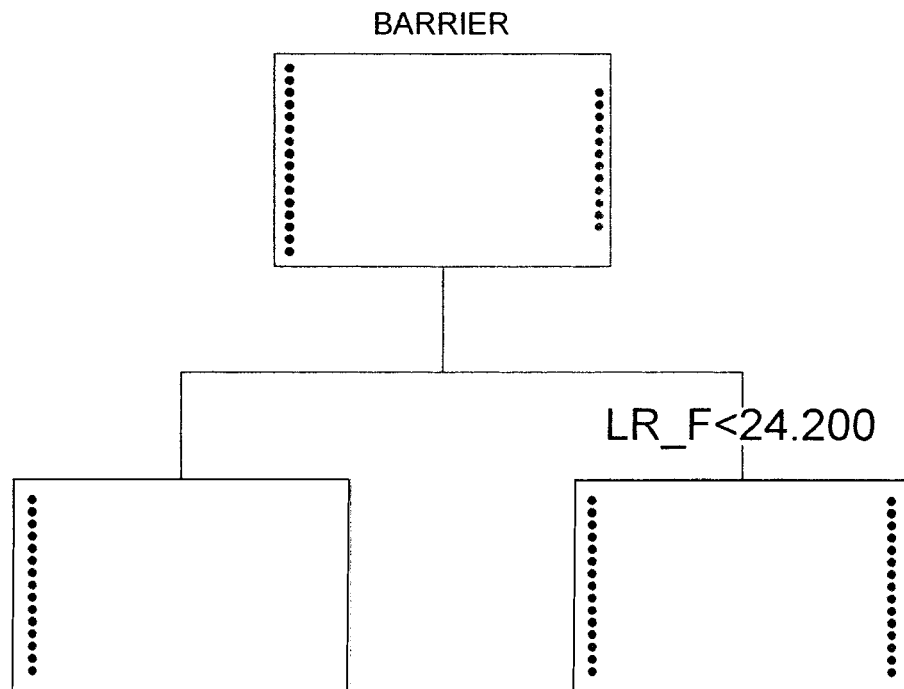| North	South
West fragment	
Barmer, Jalor, Pali, Udaipar, Dungarpur	Gandhinagar, Kheda, Ahmadabad, Surendranagar, Rajkot
Centre fragment	
Sagar, Damoh, Satna, Rewa	Akola, Amravati, Yavatmal, Wardha, Nagpur, Bhandara
East fragment	
Hazaribagh, Giridih, Murshidabad	Ranchi, Purulia, Bankura, Medinipur, Haora, Calcutta

Note: see text for explanation.

BARRIER

Figure 6. Regression trees obtained by 'automatic interaction detection' (classic AID) of the 17 socioeconomic and cultural co-variables, sampled at 28 districts, located north and south of the zone of abrupt change of the fertility index $\lambda(X)$ in 1961. *Note*: to the left are the districts located south of the zone of abrupt change, to the right those located to the north. The threshold criterion <24.2 of the LR_F variable (female literacy rate) correctly classifies the districts between the north and south of the zone, with one exception. The other variables do not carry additional information.

variables were missing were also eliminated. The list of the districts retained ($n = 28$) is given in Table 6.

To observe the behaviour of the co-variables on both sides of the barrier, by taking account of the information redundancy between variables, the procedure of regression trees called 'automatic interaction detection' (classic AID tree: see Breiman *et al.*, 1984) was used. Briefly, at each split, districts are classified using a loss function which is least squares, in which the within-group sum of squares about the group mean is as small as possible. For the actual split of the districts between those above (north) and below (south) the barrier, we chose the predictor and cut point which yielded the smallest overall within-cluster sum of squares.

Only one co-variable makes it possible to subdivide the district sample correctly, LR_F (female literacy rate), with the value <24.2%

giving districts north of the barrier, in the high fertility region, and values ⩾24.2% giving districts to the South, in the low fertility region. Using this criterion, only one district is wrongly located (in the North; see Fig. 6). This classification criterion of the districts on the fertility barrier, although close to the result of the usual ecological correlation analyses of the co-variables, is noteworthy. It makes it possible to underline, from the subcontinental to the regional geographical scale, the remarkable coincidence of results obtained by the two independent approaches. One is the correlation between the co-variables, at the national scale, which exhibits the well-known negative relationship between literacy rate and fertility. The other is the selection of a districts sample, on the criterion of their locations on each side of the fertility barrier as detected by wombling; this barrier also corresponds to a major regional change for the female literacy rate.

CONCLUSION

Compared with other fertility indices, the $\lambda(X)$ we have used takes into account mortality. If this information is missing, $\lambda(X)$ is then identical to the Ig index. The comparison of the pairs of simple isotropic variograms shows the expected temporal increase in spatial variance of fertility with geographical distance, but not between 1971 and 1981. There is an anomaly of variance between the censuses in 1971 and 1981, towards an excess in 1971 or a deficit in 1981, but perhaps not an anomaly in the expected shape of these variograms. The directional variograms, as well as the comparison of the homogeneity areas of the fertility surfaces, show that the transition takes place by an increase in deviation from the southwestern regions, which correspond to regions of relatively low fertility. The expansion of these regions reduces the

homogeneity areas, historically at high fertility levels, localised obliquely in the centre of India, in such way that the original transverse direction of this area, SW–NE, became longitudinal by 1991, and its size was cut by a half. The expansion of the areas of moderate or low fertility ($\lambda \leq 0.6$), resulting in a reduction of the high fertility areas, is observed mainly in the southern and coastal regions of the subcontinent. A study of the geographical variation of fertility and its change, using the total fertility rate (TFR) variable, recently carried out by Bhat (1996) for the years 1981 and 1991, indicated that the fertility pattern strongly differed between the north and the south of the country, and is 'reminiscent of the Indian monsoon'. Its stability was maintained over the period.

Could the barriers detected by the wombling technique be due to an artefact produced by the filling procedure of the missing data – in

Figure 7. Map of Indian States.

particular, the use of the life tables of the states? The use of this information might contribute to the generation of patches of homogeneity corresponding to the limits of the states, and thus also to zones of abrupt changes among the states. If this is in fact the case, the barriers detected at the state frontiers could be the outcome of this statistical artefact. On the other hand, the inter-state frontiers generally denote the limits of several cultural and socioeconomic variables which could exert a demographic influence. To check the assumption of possible statistical artefact, the long barrier of 1961, which is the best candidate, was examined. A statistical test of the states concerned, which takes account of the cross-distributions of the districts in and out of the barriers, and along and out of the frontier zones, is not easy to construct, in particular because of the geographical configuration of certain states (for example, Assam, where 19 districts out of 23 are frontier). By visual inspection, is the long barrier of 1961 systematically more localised along the frontier zones of states than within states? Its western border is on the frontier zone separating, longitudinally, the state of Maharashtra to the south, and the states of Gujarat and Madhya Pradesh to the north. For each of these states, life tables exist. Nevertheless, this barrier continues transversely within the state of Madhya Pradesh, until the district of Surgaja, reaching across the eastern border of Madhya Pradesh to Bihar. Moreover, the eastern pillar of this barrier (excluding East India) is entirely within the state of West Bengal, not touching any frontier district in the rest of India. One does not perceive an obvious influence of a data artefact on the identified barriers.

From 1961 to 1991, one also observes the progressive disappearance of the long transverse zone of abrupt change of fertility, dividing the subcontinent into two vast regions, one in the north with high fertility, and the other in the south with lower fertility. In 1961, the western part of this zone of abrupt change corresponded to a geographical obstacle: the Southern bank of the Narmada river following the Satpura Ranges to eastern Madhya Pradesh. The available co-variables have shown their usual ecological correlation with fertility, in particular the literacy rate, globally on an all-India scale, and regionally at the district level, located directly on both sides of this zone of abrupt change. If this barrier of fertility is pre-transitional – i.e. had been in existence historically for a long period – one may wonder, beyond the nominal co-variable 'literacy rate', which other anthropological variables, captured today under the rubric 'literacy rate' contributing to an increase in fertility in the north and a decrease in the south, have been in operation. In-depth investigations in selected districts concerning the anthropological factors that influence fertility, independently of the recent outcomes of official policies and westernisation (market economy, mass media), could provide meaningful answers (for example, in the district of Barner, Rajasthan, $\lambda = 1.1$, and in that of Ahmadabad, Gujarat, $\lambda = 0.7$, when they are separated from each other by about 300 km). These districts are among those for which the variance of fertility is at a maximum and the geographical distance at a minimum. It should be the same for the underlying anthropological factors.

ACKNOWLEDGEMENTS

APPENDIX A

A Fertility Estimator at the Localities

In order to take account of the space–time variation of mortality of children and women and the differences between the relative fertility of women across age groups, an index of fertility was estimated, under the following assumptions. The migrant population does not live separated from its children (the children

are enumerated at the same time that parents are enumerated). The only consequence of migration is the assignment of the place of birth having taken place in the locality of origin to the place of migration, inducing a (very weak) loss of contrast. Because of the Indian demographic regime, there is little distinction between marital and general fertility.

The fertility rate of the age class x at the age $x + a$, at the locality X, written as $\tau(x, X, t)$, may be represented as the product of a term which takes into account the profile of the age effect $f(x, t)$, and a geographical factor $\lambda(X, t)$. Thus, under this assumption, we have:

$$\tau(x, X, t) = f(x, t)\lambda(X, t)$$

In this model, the profile of the age effect is thus, at a given time, the same in all the localities; only the fertility intensity varies spatially. In this form, it is clear that the fertility allocation between the terms f and λ is arbitrary. A natural choice consists in taking for the value of f the average fertility over the country – the age-specific fertility rate – and thus $\Sigma_x \tau(x, X, t) = f(x, t)$. With this choice, the spatial average $\lambda(X, t) = 1$ at every time t.

Let us stress that only the profile of f, and not its level, exerts an influence on the estimated fertility $\hat{\tau}$. That means that a multiplicative error, relating to the level of f, would be automatically compensated for by an opposite modification of the geographical coefficient, producing correct values of fertility. The geostatistical approach consists of building a stochastic model in which the observed variables (i.e. the frequencies) are interpreted as a particular outcome of this model (Matheron, 1989). In this model, the geographical factor λ is regarded as a random function. We begin from a given locality. Here the spatial dimension does not intervene. Let us assign each individual to a cohort, according to his/her age at the time of the census. An individual dead before this date is assigned according to the age which he/she would have had if he/she had survived.

Let $P(0, X, t)$, $P(x, X, t)$, $q(x, t)$ and $q(0, X, t)$, be the random variables representing the enumerated children of the local population aged 0–4 years, the enumerated women belonging to the age class x, to the age $x + a$, the in-

stantaneous probabilities of death for women, and the probability of death of the children belonging to the age class 0–4 years at time t, respectively. As the starting point of a period, the fifth year preceding a census, t is set to 0. The end of the period corresponding to the census is set to $t = 5$. The evolution of the female population and that of the children is given by:

$$E[dP(x, X, t)|F(t)] = -q(x, X, t)P(x, X, t)dt$$
$$(x = 15, 20, \ldots, 45)$$

$$E[dP(0, X, t)|F(t)] = \sum \tau(x, X, t)P(x, X, t)dt$$
$$- q(0, X, t)P(0, X, t)dt$$

where $F(t)$ is the history of the process until the time t. To introduce mortality, we write $P(x, X, t) = E[P(x, X, t)]$. From the preceding equation we deduce, after having lifted the conditioning, that:

$$d\overline{P}(x, X, t) = -q(x, X, t)\overline{P}(x, X, t)$$

The coefficients $\lambda(x, X, t)$ and $q(x, X, t)$ may vary over time under the effect, for example, of population ageing and the improvement of living conditions. To simplify the solving of the equation, we will admit that the use of average coefficients (at the mid-date of the period considered, and for the age of the population concerned at this date) gives a good approximation of the solution. Under this approximation, $\lambda(X, t)$, $q(x, X, t)$ and $f(x, t)$ are constant per piece (over each period of integration). This approximation is perhaps somewhat less suitable for the children of less than 5 years. Therefore the evolution of the parental population of the age class x, on average, follows the law:

$$\overline{P}(x, X, t) = \overline{P}(x, X, 0)\mathrm{e}^{-q(x, X, t)}$$

Let us now introduce fertility, by calculating the average of $P(0, X, t)$ which we will denote $\overline{P}(0, X, t)$. According to the second differential equation given above, one obtains the following equality:

$$d\overline{P}(0, X, t) = \sum \lambda(x, X, t)\overline{P}(x, X, t)dt$$
$$- q(0, X, t)\overline{P}(0, X, t)dt$$

By taking account of the initial condition $\overline{P}(0, X, t) = 0$, the differential equation is easily

solved and the solution is then:

$$\overline{P}(0, X, t) =$$

$$\sum \frac{\tau(x, X, t)P(x, X, t)}{q(0, X, t) - q(x, X, t)} (e^{-T[q(x, X, t)]} - e^{-q(0, X, t)})$$

Of course, this expression must be replaced by its limit if two coefficients of mortality $q(x, X, t)$ and $q(0, X, t)$ are equal, which is not the case in general, for the mortality of the children is higher than that of the women aged 15–49 years. By replacing $\tau(x, X, t)$ by $\lambda(X, t)f(x, t)$ as suggested by the model, and by taking $P(0, X, T)$ as estimator of its average, where T corresponds to the end of the period, one obtains:

$$\hat{\lambda}(X, t) =$$

$$\frac{P(0, X, T)}{\sum \dfrac{f(x, T)P(x, X, T)}{q(0, X, T) - q(x, X, T)} (1 - e^{T[q(x, X, T) - q(0, X, T)]})}$$

This estimator takes into account the missing frequencies due to mortality during td, by weighting the enumerated frequencies of the age classes by the corresponding probabilities of survival. When the probabilities of death are negligible ($q(x, X, T) - q(0, X, T) \to 0$), the natural estimator is found to be:

$$\hat{\lambda}(X, t) = \frac{P(0, X, T)}{T \sum f(x, T)P(x, X, T)}$$

APPENDIX B

The Expectancy of the Estimator for the Space Factor

Let us notice that by definition of the average fertility f at the moment t:

$$f(x, t) = \frac{1}{N_{loc}} \sum_{X=1}^{N_{loc}} \tau(x, X, t) = f(x, t) \frac{1}{N_{loc}} \sum_{X=1}^{N_{loc}} \lambda(X, t)$$

(N_{loc} designating the number of census localities), which means that the spatial average of $\lambda(X, t)$ is equal to one. With regard to the estimator of the geographical factor where

mortality is neglected, one has:

$$E\left[\frac{P(0, X, T)}{\sum Tf(x, T)P(x, X, T)}\right] =$$

$$E\left[\frac{E[P(0, X, T)|P(x, X, T), x \geqslant 15]}{\sum Tf(x, T)P(x, X, T)}\right]$$

but

$$E[P(0, X, T)|P(x, X, T)] =$$

$$\sum T\lambda(X, T)f(x, T)P(x, X, T)$$

where, conditionally to the function λ:

$$E[\hat{\lambda}(X, T)|\lambda] = \lambda(X, T)$$

It follows that the expectancy of the spatial average for the $\hat{\lambda}(X, T)$ estimator is also unity.

REFERENCES

Barbujani G, Oden N, Sokal RR. 1989. Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology* **38**: 376–389.

Bhat M. 1995. Age misreporting and its impact on adult mortality estimation in South Asia. *Demography India* **24**: 59–80.

Bhat M. 1996. Contours of fertility decline in India: a district level study based on the 1991 census. *In Population Policy and Reproductive Health*, Srinivasan K (ed.). Hindustan Publ: New Delhi; 96–117.

Bhat M. 1998. Demographic estimates for post-independence India: A new integration. *Demography India* **27**: 23–57.

Bhat M, Preston S, Dyson T. 1984. *Vital Rates in India, 1961–1981*. Report no. 24. Committee on Population and Demography: Washington DC National Academy Press.

Bhat M, Zavier F. 1999. Findings of the National Family Health Survey: Regional analysis. *Economic and Political Weekly* **XXXIV**: 3008–3033.

Bocquet-Appel JP, Bacro JN. 1994. Generalized wombling. *Systematic Biology*, **3**: 316–329.

Bocquet-Appel JP, Jakobi L. 1996. Barriers to the spatial diffusion for the demographic transition in Western Europe. *In Spatial Analysis of Biodemographic Data*, Bocquet-Appel JP, Courgeau D, Pumain D (eds). Congresses and Colloquium 16. Eurotext, John Libbey/Ined: Montrouge; 117–129.

Bocquet-Appel JP, Jakobi L. 1997. Diffusion spatiale de la contraception en Grande-Bretagne, à l'origine de la transition. *Population* **4**: 977–1004.

Bocquet-Appel JP, Jakobi L. 1998. Evidence for a spatial diffusion of contraception at the onset of the fertility transition in Victorian Britain.

Population, Special Issue, New Methodological Approaches in the Social Sciences: 181–204.

Breiman L, Friedman J, Olshen R, Stone C. 1984. *Classification and Regression Trees*. CRC Press.

Census of India. 1977. *Age Tables*. Census of India 1971, Office of the Registrar General and Census Commission 3: New Delhi.

Census of India. 1982. *Report on Post Enumeration Check*. Office of the Registrar General and Census Commission 4: New Delhi.

Census of India. 1994. *Report on Post Enumeration Check*. Office of the Registrar General and Census Commission 1: New Delhi.

Chandra NK. 1980. Adjustment of age data for India's census population. *Demography India* **IX**: 274–285.

Coale AJ, Treadway R. 1986. A summary of the changing distribution of overall fertility, marital fertility, and the proportion married in the provinces of Europe. *In The Decline of Fertility in Europe: The Revised Proceedings of a Conference on the Princeton European Fertility Project*. Coale Ansley J, Watkins Susan C (eds). Princeton University Press: Princeton; 31–181.

Dyson T. 1976. Analysis and adjustment of the 1971 Indian age distribution, and a reappraisal of mortality and fertility estimates. *Demography India* **1**: 71–92.

Dyson T, Murphy M. 1985. The onset of mortality transition. *Population and Development Review* **11**: 399–440.

Marr D, Hildreth E. 1980. Theory of edge detection. *Proceedings, Royal Society of London* **B207**: 187–217.

Matheron G. 1989. *Estimating and Choosing*. Springer Verlag: Berlin.

Pannatier Y. 1996. *Variowin, Software for Spatial Data Analysis in 2D*. Springer Verlag: Berlin.

Registrar General, India. 1989. *SRS Based Abridged Life Tables 1981–85*. Occasional Paper No. 1 of 1989. Office of the Registrar General: New Delhi.

Registrar General, India. 1994. *SRS Based Abridged Life Tables, 1986–90*. Occasional Paper No. 1 of 1994. Office of the Registrar General: New Delhi.

Registrar General of India. 1988. *Child Mortality Estimates of India*. Occasional Paper No. 5 of 1988. Office of the Registrar General: New Delhi.

Registrar General of India. 1997. *District Level Estimates of Fertility and Child Mortality for 1991 and their Interrelations with Other Variables*. Occasional Paper No. 1 of 1997. Office of the Registrar General of India: New Delhi.

Rele JR. 1987. Fertility levels and trends in India, 1951–81. *Population and Development Review* **13**: 513–530.

Visaria P. 1971. The provisional population totals of the 1971 census: some questions and research issues. *Economic and Political Weekly* **6**: 1459–1465.

Visaria P, Visaria L. 1997. Demographic transition: Accelerating fertility decline in the 1980s. *In India's Demographic Transition. A Reassessment*, Irudaya Rajan S (ed.). MD Publications: New Delhi; 245–279.

Visaria P, Irudaya Rajan S. 1999. National Family Health Survey: A landmark in Indian surveys. *Economic and Political Weekly* **XXXIV**: 3002–3008.

Wackernagel H. 1998. *Multivariate Geostatistics*. Springer Verlag: Berlin.