Environment and Planning A 2000, volume 32, pages 695 – 714

Spatial variability in fertility in Menoufia, Egypt, assessed through the application of remote-sensing and GIS technologies

John R Weeks, M Saad Gadalla, Tarek Rashed, James Stanforth
Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA;
e-mails: john.weeks@sdsu.edu; gadalla@mail.sdsu.edu; trashed@mail.sdsu.edu;
stanfort@rohan.sdsu.edu
Allan G Hill
Harvard School of Public Health, Harvard University, 9 Bow Street, Cambridge, MA 02138, USA;
e-mail: ahill@hsph.harvard.edu
Received 24 May 1999; in revised form 28 September 1999

Abstract. Fertility in rural areas such as the Governorate of Menoufia in Egypt may be influenced both by spatial factors (including the diffusion of innovations) and by essentially nonspatial factors (such as the availability of education for women and the percentage of adult women who are currently married). The nonspatial variables are available directly from censuses but the spatial component requires an accurate location of the villages to which the census data refer and then appropriate decomposition of the data into spatial and nonspatial components. We use IRS satellite imagery to classify the built area in a rural governorate in Egypt and then assign village-level census data to the centroids of those polygons and incorporate the data into a GIS. We then employ measures of global and local spatial statistics to conclude that in 1976 the combination of female illiteracy, proportion married, and spatial clustering accounted for 39% of the variation in fertility in Menoufia. In 1986 those same factors explained 51% of the variation in fertility. In 1976 about one third and in 1986 about half of the explained variability was due to the spatial component ('diffusion') and the other half due to a combination of demographic characteristics. Furthermore, between 1976 and 1986 there was a clear north-to-south drift of fertility, with lower fertility being clustered in the north and higher fertility clustered in the south.

Introduction

In most parts of the world the rural population is concentrated in small but relatively densely settled villages. People do not necessarily live on the land they farm, but rather in a village that represents a reasonably efficient use of the often limited rural resources. This is important from a demographic perspective because it means that the village may provide sources of human capital such as a school or employment opportunities, and it also provides a context within which information is shared that may influence decisions that individuals, couples, or households make about health care, migration, and reproduction. The type and extent of information that is shared may come from mass media such as radio and television but it may also be shared amongst villages because most rural areas in agricultural societies are part of a network of villages. The demographic behavior of villagers is of considerable importance because it is in the rural areas that population growth is highest throughout the world, contributing to redundancy within the rural population itself and encouraging migration to already overburdened urban areas (for example, see Jones, 1990; Lutz, 1994; Weeks, 1999).

An improved understanding of the fertility transitions in developing countries is critical if culturally appropriate and effective reproductive health and family planning policies are to be designed and implemented to improve the welfare of the population. There is now ample evidence that the traditional models of the demographic transition as developed by Thompson (1929), Notestein (1945), and Davis (1945) were overly simplistic and incomplete. Reformulations began with Davis's theory of demographic change and response (Davis, 1963) but have been driven especially by results of the European Fertility Project at Princeton (Coale, 1973; Watkins, 1991) and by analyses of data from the World Fertility Surveys (Cleland and Wilson, 1987). The models have also been respecified by the disaggregation of the demographic transition into its constituent components of the epidemiological transition (Omran, 1971), the fertility transition (for example, see Chesnais, 1992; McDonald, 1993), and the migration (especially urban) transition (Firebaugh, 1979), with increasing thought being given to the specification of a nuptiality transition (Chesnais, 1992) and/or a more general household transformation or transition. This respecification, which hearkens back to Davis's theory of demographic change and response, allows us to examine each element of the transition. In so doing, the evidence seems to suggest that the fertility transition is neither a simple response to rising individual standards of living nor an automatic response to declining mortality. The picture appears to be more complex in several ways (Reed et al, 1999) and it seems likely that the fertility transition is best understood as a 'blend' of structural factors (exemplified by the supply–demand framework; Easterlin and Crimmins, 1985), diffusion factors (Cleland and Wilson, 1987; Kirk, 1996; Knodel and van de Walle, 1979), and the local context in which reproductive decisions are actually made (Entwisle et al, 1989).

In this research we are interested in the extent to which the variation in fertility from one rural village (the local context) to another may be explained by a process of diffusion of behavior from some villages to others, net of the human capital variables such as education that may exist within the village. We lack direct evidence of such spatial diffusion but can infer it from the spatial and temporal patterning of reproductive behavior. "The diffusionist perspective predicts that if Area A is proximate to areas with relatively low fertility, and Area B is surrounded by areas with high fertility, Area A will have lower fertility than Area B" (Tolnay, 1995, page 301). The spatial dimension of demographic behavior must thus be measured so that these variables can be evaluated for the role that diffusion may play in fertility behavior. We must first show that proximity matters, and then show that changes occurred over time in a sequence that is consistent with spread or diffusion. Our goal in this paper is to show that proximity matters when it comes to fertility levels, even in a rural area of a developing country. Thus we assess the extent to which the spatial variability in fertility differs from that which can be expected to happen by chance by testing the null hypothesis that fertility is spatially independent.

The study site

We illustrate this procedure by using data for a rural governorate of Egypt (Menoufia) for 1976 and 1986. Menoufia is one of the twenty-six governorates that form the administrative regions of Egypt, roughly equivalent to states in the United States, although perhaps more analogous to counties in the United Kingdom. Figure 1 illustrates the study site. In 1976 the total enumerated population of Menoufia was 1.88 million, accounting for about 5% of the total population (37.7 million) of Egypt that year. By 1986 the population of Menoufia had grown to 2.23 million, representing 4.4% of the total Egyptian population of 51 million in that year. Between these two years, which represent the years of analysis for this study, the population of Menoufia was growing at an annual rate of 1.7%. Despite that speed of growth, Menoufia lost ground (in terms of relative population size) to the even more rapidly growing governorates in upper Egypt.

Fertility in Menoufia was high in 1976, and it was still high in 1986, although there has reportedly been some fertility decline since 1986 (the 1996 census data at the local

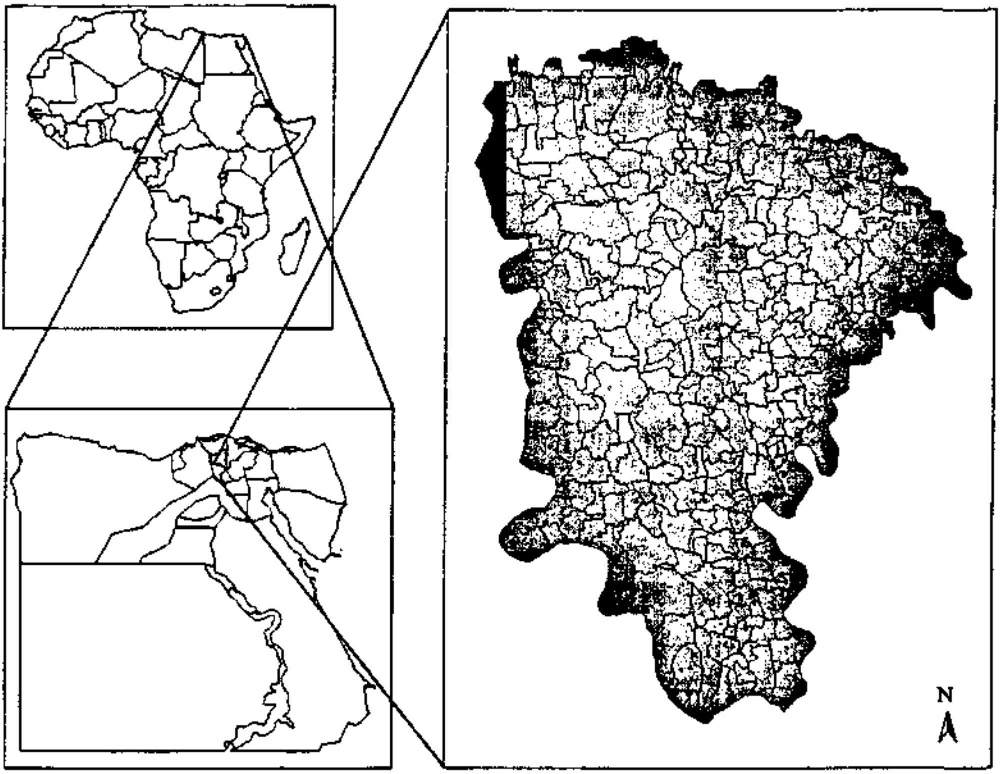

Figure 1. The study site of Menoufia Governorate, Egypt.

level had not yet been released by the Egyptian government at the time that this paper was written). For decades Menoufia has been one of the most rural and most densely populated rural areas of Egypt (Gadalla, 1978). It has been, and remains, predominantly agricultural, and the high rate of population growth has increased the redundancy of the rural labor force and encouraged out-migration, to Cairo or to other Arab (especially oil-producing) nations.

We chose Menoufia as a study site less for its specific demographic characteristics than for the fact that it has been relatively well studied in nonspatial analyses and thus there are comparative studies by which to judge the spatial analysis produced by our own work. Menoufia does have some advantages for spatial analysis including its essentially flat landscape in the Delta region of the Nile, which means that elevation is not an issue that needs to be dealt with. Partly for this reason, there are numerous villages (312), with an average population per village in 1976 of 5566 and in 1986 of 7175.

Data

Assuming that we find a spatial component to the variability in fertility, we seek to decompose the contribution to fertility made by sociodemographic and human capital variables such as marital status and education from those contributions made by the clustering of villages with similar or shared characteristics. The first set of factors, which in theory have no spatial component, are part of the broader supply – demand framework of the fertility transition (for a review, see Weeks, 1999), whereas the spatial component is more clearly a part of the cultural-diffusionist perspective on the fertility transition. We hypothesize that both sets of factors are important and aim to quantify the importance of each.

Demographic data

The demographic data used for this study came from the 1976 and 1986 censuses of Egypt. Data were coded from the Arabic-language publications by using the smallest geographic unit available in the Egyptian censuses—the 'shiakha' or village. The shiakha literally refers to the area controlled by a sheikh, but in more practical terms it is the area serviced by a police post.

At this level of geographic detail, the demographic detail was minimal and we utilize a measure of fertility derived from the child/woman ratio. The child/woman ratio typically measures the ratio of children aged 0–4 years enumerated in the census to the number of women of reproductive age (15–49 years) enumerated in the census. The youngest age group (especially the first year of life) tends to be the least reliable part of the enumerated population, and therefore one must either measure the degree of underenumeration and adjust for it, or assume that whatever underenumeration exists is not highly variable from place to place and thus assume that it will not affect the relative differences in fertility estimates from place to place.

There is evidence of some underenumeration of children in the censuses in Egypt but we had no direct way of adjusting for it at the village level. Instead we assessed its potential by calculating an alternative measure of the child/woman ratio, $R^{\text{cw}}_{(5-9/20-54)}$, which shifts the entire measure up by five years to ignore the youngest age groups altogether.

The child/woman ratio will also vary depending upon differences in infant and childhood mortality which can affect the number of surviving children enumerated in the census. Important changes in mortality were occurring between 1976 and 1986—the rates were going down at all ages and there was a gender crossover at the youngest ages. In 1976 childhood mortality rates were consistently higher for females than for males (Makinson, 1986) but that pattern had abated by 1986. We adjusted for relative differences in mortality by calculating the Thompson net reproduction rate (Smith, 1992), $R^{\text{nr}}_{\text{Thompson}}$,

$$R^{\text{nr}}_{\text{Thompson}} = R^{\text{cw}}_{\text{enumerated}} \big/ R^{\text{cw}}_{\text{life-table}} ,$$

where the life-table child/woman ratio was calculated for 1976 by using data from the US Bureau of the Census (1999), and 1986 data were derived from spreadsheets developed by Arriaga (1994) and are based on deaths by age and enumerated population from the US Bureau of the Census (1999).

The child/woman ratio may also be influenced by migration because people of reproductive age are those who are most likely to be migrants. In rural areas it can be expected that migration will largely be out of the village and that it will disproportionately affect males. This may not influence the actual calculation of the child/woman ratio, yet the relative absence of men could have a dampening influence on fertility. We control for this effect by introducing the sex ratio of males aged 25–44 years to females aged 20–39 years as a covariate in the analysis. The child/woman ratio will also be influenced by the proportion of adult women who are currently married. In a society such as Egypt where out-of-wedlock births are relatively rare, we would expect that fertility will be higher where the proportion of married women is higher.

We measure the human capital variables in the village in terms of female education. The data are available only for all women aged 15 years and older, regardless of specific age or other characteristic. In addition, because of the limited educational attainment, the educational variable was measured as the percentage of women aged 15 years and older who are illiterate. The final covariate introduced into the model was the total size of the village as a proxy for the relative degree of urbanness of a village.

Table 1. Descriptive statistics for demographic data of villages in Menoufia, Egypt, 1976 and 1986.

	Mean	Median	SD	Minimum	Maximum	Skewness
Child/woman ratio, 1976[a]	0.66	0.65	0.12	0.31	1.19	0.26
Child/woman ratio, 1986[a]	0.71	0.69	0.12	0.42	1.07	0.44
Thompson's net reproduction rate, 1976[b]	2.14	2.13	0.27	1.34	3.03	−0.02
Thompson's net reproduction rate, 1986[b]	2.38	2.36	0.36	1.34	3.40	0.02
Sex ratio in 1976[c]	0.86	0.86	0.08	0.63	1.23	0.24
Sex ratio in 1986[c]	0.88	0.88	0.07	0.65	1.35	1.42
Female illiteracy in 1976[d]	0.81	0.83	0.10	0.31	0.97	−1.23
Female illiteracy in 1986[d]	0.67	0.67	0.10	0.31	0.92	−0.23
Percent married in 1976[e]	0.64	0.64	0.06	0.38	0.77	−0.44
Percent married in 1986[e]	0.63	0.63	0.06	0.39	0.78	−0.22
Village population in 1976	5566	3790	7220	519	86 533	6.53
Village population in 1986	7175	4821	8513	600	76 664	4.46

Note: SD, standard deviation.
[a] Children aged 0–4 years/women aged 15–49 years.
[b] (Children aged 5–9 years/women aged 20–54 years in census)/(children aged 5–9 years/women aged 20–54 years in life table).
[c] Males aged 25–44 years/females aged 20–39 years.
[d] Proportion of women aged 15 years or above who were illiterate.
[e] Proportion of women aged 15 years or above who are currently married.

Table 1 summarizes the descriptive statistics for these variables for the villages in the Egyptian Governorate of Menoufia in 1976 and 1986. The data reveal that fertility actually went up on average, rather than down, during the decade between the 1976 and 1986 censuses, as shown both by the child/woman ratio measured in the usual way as the ratio of children aged 0–4 years to women aged 15–49 years, and by the Thompson net reproduction rate measured as the enumerated ratio of children aged 5–9 years to women aged 20–54 years divided by the life-table ratio of children aged 5–9 years to women aged 20–54 years. During that interval, the percentage married stayed the same, the sex ratio increased slightly (reflecting a higher proportion of men which could act to increase fertility), and the female illiteracy rate declined. This last change should have had the effect of lowering fertility but we can only guess that this is a lagged effect which may have a subsequent negative impact on fertility in those villages where it declined. Even the lowest value of R^{nr} (1.34) in both 1976 and 1986 was well above the level of replacement and a predictable consequence was population growth in the villages between 1976 and 1986, with the average village increasing in population by 1600 people during that decade.

Remotely sensed images

In attempting to model the diffusion of fertility and/or its antecedents (the human capital variables) we must deal with an important methodological issue with regard to spatial analysis: the modifiable areal unit problem (MAUP) (Openshaw, 1984). In almost every instance our demographic information is gathered at some arbitrarily defined geographic level such as a census tract or enumeration district. However, unless this area defines a small and heavily built area, its areal boundaries will include space in which people do not reside. This is especially true in agricultural areas where most space is devoted to crop, orchard, or pasture land. Thus a rural village, even if densely populated within its own boundaries, may consume only a small portion of the administrative boundaries to which the demographic data are attached.

Furthermore, the process of diffusion implies a certain degree of contiguity, measured by the distance between two or more places. In order to measure such distances we need to fix the location of each place. We could do this arbitrarily by assigning a centroid to each administrative boundary and assuming that the population is concentrated at this geometric or geographic center of the administrative unit. When the geographic area is small, this approach will not be far from the actual population center but, as the size of the area under consideration increases, the probability of error in locating the population by this method also increases.

Assigning a location to the data being analyzed is essential if the spatial component of fertility (or any other variable) is to be assessed. It is obvious that the more accurately this location is measured the more reliably will the data analysis represent the actual demographic situation.

Our use of remotely sensed images was to classify land cover as a way of identifying where a population is concentrated within the boundaries of the administrative unit for which the data (in this instance, census data) are collected. There are two interrelated reasons for wanting this information: (1) most demographic data are georeferenced only for some fairly large area (such as a census tract, village, town, or county), yet if we believe that physical contiguity influences behavior then we want to locate people where they actually live, rather than assuming that they are uniformly distributed within the boundaries of the geographic area referenced by the data; and (2) if we can locate the place or places to which the data actually refer then we can more accurately create a grid structure for our data which will facilitate the spatial analysis.

We should note that our identification of demographic data with built areas will accurately reflect the geographic source of such data collection but may misspecify the site of the activities that might lead to the diffusion of ideas and behavior. For example, important exchanges of information may occur among people working in the fields, as well as at markets and other places that are distant from the residences. This may represent a variation on the theme of ecological fallacies. In subsequent research we intend to try to control for this with spatial contextual variables such as village access to major transportation networks.

We employ remotely sensed images to classify land cover by built or nonbuilt use in order to undertake what is sometimes known as dasymetric mapping (Langford and Unwin, 1994), in which information inside a zone is used to map the population density or distribution within that zone. The advantage of a classification of data from a remotely sensed image which spatially defines built areas is that it frees us from the "tyranny of an arbitrary imposed and fixed set of census geographies" (Openshaw and Rao, 1995, page 425). The classification of remote images provides a new zone-design tool that can help to resolve the MAUP.

In order to classify images on the basis of built and nonbuilt land use, and waterways (the major branches of the Nile and larger irrigation canals), we required aerial photographs or satellite images that were at least 10 m resolution and we preferred 5 m resolution in order to minimize classification error. The only such images available for the 1980s for Egypt were those taken by Soviet Resurs-F1 satellites (part of the Kosmos series), with 5 m resolution spectrozonal KFA-1000 cameras. The images were photographs taken on a cloudless day in May 1986, a date chosen to coincide as closely as possible with the 1986 census in Egypt.

Unfortunately we discovered that the inferior quality of the Russian images limited our ability to classify them for the purposes of identifying land cover as we had intended, so we additionally acquired two more recent images of Egypt for cloudless days in July 1996 and 1998, respectively, to permit eventual linkage with the 1996 census data (which were not available at the time of writing). The 1996 image was an

IRS-1C LISS-III 24 m resolution multispectral image covering bands 2, 3, and 4 (green, red, and near infrared) whereas the 1998 image was a 5 m resolution panchromatic digital image generated by the Indian Remote Sensing Satellite IRS-1C. These two images were merged to provide the most accurate classification.

With the preprocessing completed, the next step was to classify the area into two categories: villages (built areas) and fields (nonbuilt areas). Unsupervised and supervised classifications were made by using ERDAS Imagine software. Figure 2 shows the classification of Menoufia by built or nonbuilt areas within each of the shiakha boundaries.

The polygons of identified built areas as determined from the images are bound to be irregularly sized and shaped. So even if they do more accurately represent where people live, there remains the problem of how properly to analyze the data in this zonal format. There are at least three possible solutions. One is to leave the polygons as classified and map the data by using the irregular polygons. This solution treats the

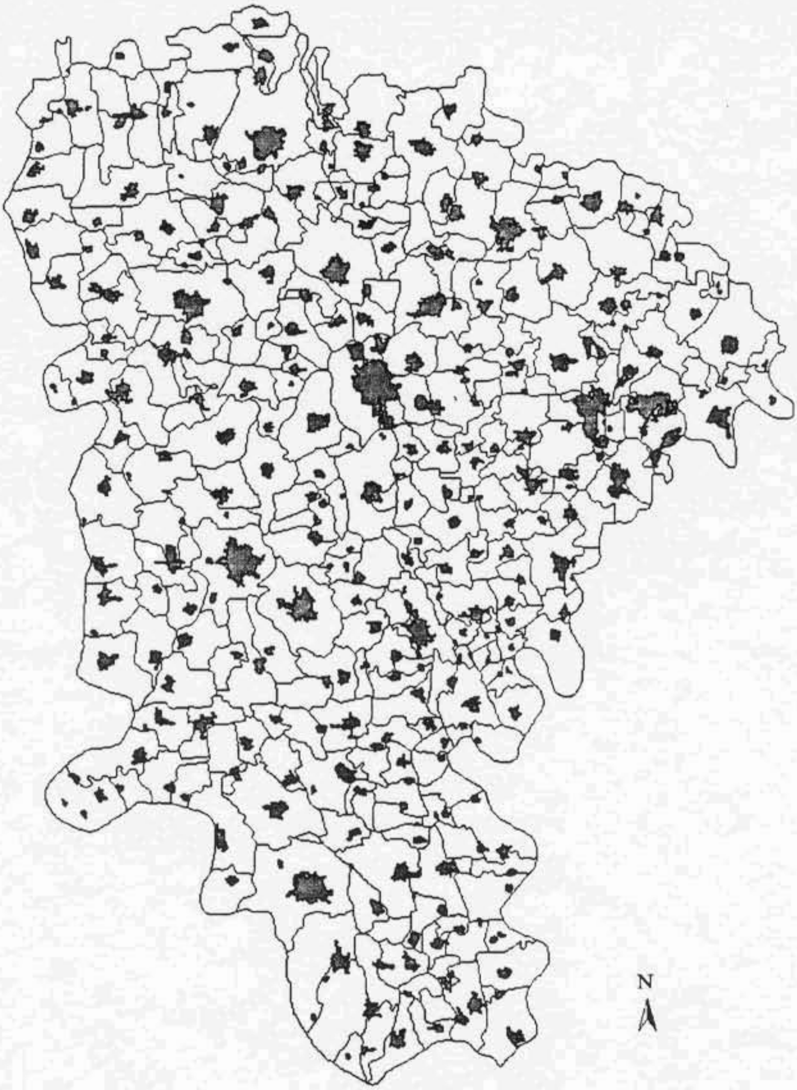

Figure 2. Results of classification of satellite imagery into built (dark) and nonbuilt (light) areas, by administrative boundaries (shiakhas) of Menoufia Governorate, Egypt.

new polygons as similar to the previous administrative boundaries, except that the boundaries more precisely map the population distribution. This is the simplest solution but it fails to make best use of the improved data now available to the researcher.

A second solution is to use this more precise spatial definition of population distribution to create a continuous population surface (Goodchild et al, 1993; Martin, 1996; Tellier and Vertefeuille, 1995). This method may utilize one of several techniques to create centroids to represent each identified built area and then employ kriging techniques to interpolate a surface between centroids. This method works best when the variable under consideration is total population size because it is likely that population size will follow a pattern that can be characterized by a form of the gravity model—with the highest density near the centroid and density declining with distance from the center. It is perhaps less realistic to use this surface to model other characteristics such as fertility, mortality, or migration levels, the distributions of which are less likely to be described by a gravity density function. The continuous surface model is also disadvantageous if you are not certain that the variable under consideration, such as fertility level, is necessarily spatially continuous. Under such conditions, a spatially discrete model may work better. That is the third solution.

The third solution involves creating a spatially discrete set of data by converting the classified polygons to a raster-based discrete surface grid for statistical analysis and spatial modeling. The grid has the advantage statistically that it permits the collection within a GIS of multiple layers of data and attributes for a regularly defined geographic zone. Analysis is thus facilitated and the results are more interpretable statistically and visually. This is especially important in a study such as ours where change detection over time is an important goal. The raster data provide a clear image of change as population spills from one cell into another and as values change between and within cells [for examples in which UK census data are used, see Bracken and Martin (1995) and Mesev (1998)]. The discrete model has the added advantage that a continuous surface model can later be estimated by using the raster data (Langford and Unwin, 1994).

The polygons of the built areas as classified from the images vary considerably in size and shape, and the number and position of these polygons vary from one census area (shiakha) to another. This variation can be described by the following five different scenarios:

(1) the census area contains one unique built area;
(2) the census area contains multiple unique built areas;
(3) the census area contains a portion of one built area that has spilled over from a neighboring census area;
(4) the census area contains unique built areas and portions of spill-over built areas; and
(5) the census area contains no built areas that were classified from the imagery, apparently reflecting a population that is dispersed among the fields or near roads in such a manner that they cannot be identified from the imagery.

In converting these polygon data to a grid cell in order to assign a location to the population within an administrative area, we have essentially four choices (Chou, 1996):

(1) assign a point location to the classified polygon (the centroid method);
(2) assign cell values based on the predominant classification (in this case, built or nonbuilt) within a cell;
(3) assign cell values based on any portion of built classification being included in the cell; or
(4) develop a specific hierarchical decision algorithm to assign cells.

We have chosen the fourth approach and, in essence, have combined elements of the first three. The classified polygons were converted to points by using the centroid of each polygon and the points were then converted to a raster grid for statistical analysis. This method, however, creates some problems. In several instances there was more than one polygon within the census area or one polygon divided between two census areas. To solve this problem, two steps were taken. First, all classified built polygons spilling over into multiple census areas were divided among the constituent census areas, reflecting the fact that census data were aggregated separately for those parts of villages. Second, in those cases where there was more than one polygon in a census area, the centroid (and thus the census data) was assigned uniquely to the built polygon that is largest in size. The resulting set of centroids is shown in figure 3.

Once the built areas had been identified and a centroid calculated to represent the polygons, a grid was laid over the entire surface to create the new areal units of analysis within the GIS. The grid size was chosen to minimize the likelihood that more than one centroid would fall within any given cell. The following decision rules were applied for assigning data to each grid cell:

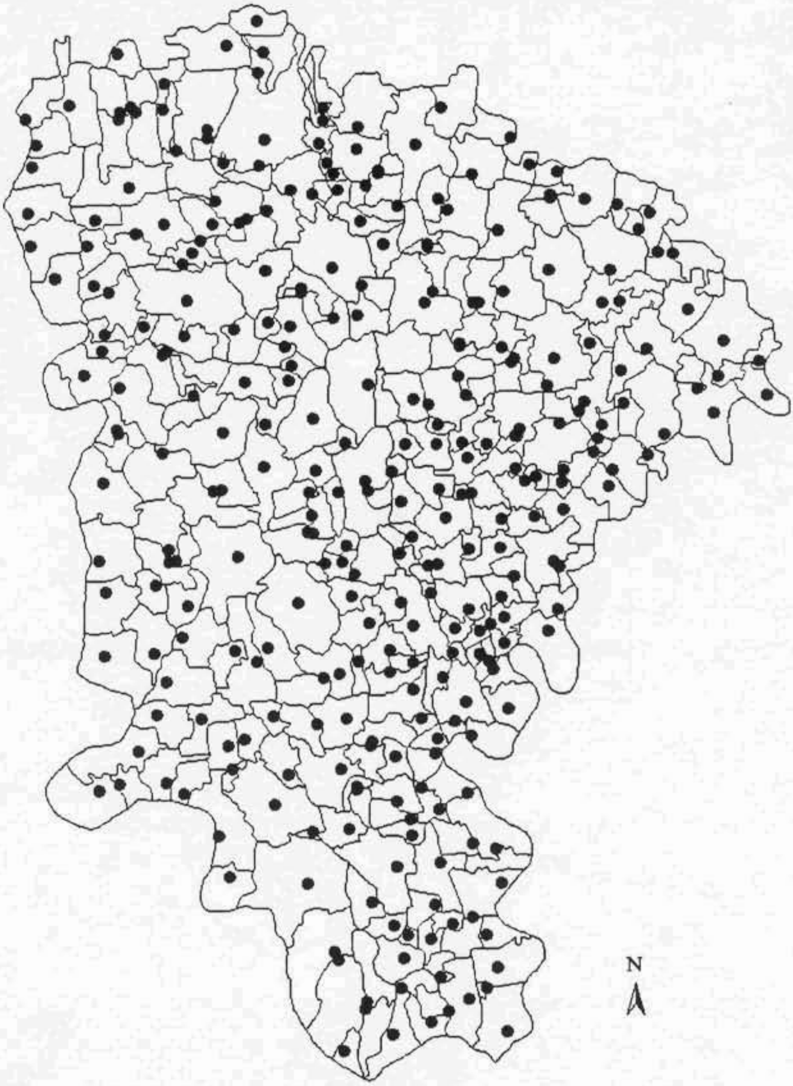

Figure 3. Centroids of built areas shown in figure 2, after applying decision rules to produce a unique built area centroid for each census area.

(1) if no centroid existed within the cell, the cell was assigned null values for every variable;
(2) if one unique centroid fell within the cell, that cell was assigned the values for each variable as determined for the census unit (villages) to which the centroid was linked;
(3) if more than one centroid fell within the cell then one of the centroids was chosen at random to represent the data for that cell. The advantage of this approach is that it avoids the problem of spuriously duplicating data, and the disadvantage is that it runs the risk of ignoring potentially important spatial information.

Only a small number of cells had duplicate centroids and we found that our statistical analysis was not affected by either including the duplicates or excluding them. We experimented with several different grid cell sizes and found that the results were consistently the same with grid cells of 1 km, 500 m, and 250 m. Our analysis employs the 1 km grid cell, as shown in figure 4.

Figure 4. A 1 km grid of centroids of the classified built areas laid over the data shown in figure 3.

Methods of analysis

We first use ordinary least-squares (OLS) linear regression models to assess the relative contribution of the human capital variables to fertility at the village level in 1976 and 1986. The first model, for example, treats the Thompson net reproduction rate (referred to hereafter simply as R^{ur}) as the dependent variable, with a control introduced for the sex ratio in the reproductive ages. The next set of variables introduced in the model are female illiteracy in 1976, the proportion of women married in 1976, and the total population size of the village. A similar model is constructed for the 1986 data.

Using raster grid data, we then introduce the spatial component at both the global and local levels. The most widely used global measure of spatial autocorrelation is Moran's I (Cliff and Ord, 1981). This statistic tests for spatial patterns that differ from randomness. The theoretical value of Moran's I in a spatially independent pattern of points is $-1/(N-1)$. Values of I greater than that indicate positive spatial autocorrelation (in which neighboring values are similar to one another), and values below that are indicative of negative spatial autocorrelation (in which neighboring values are dissimilar to one another; Chen and Getis, 1998).

The local spatial statistic utilized is the $G_i^*(d)$ statistic (Getis and Ord, 1992; Ord and Getis, 1995), which measures the clustering of similar values around a given point at a specified distance from that point, relative to the point pattern in the entire geographic surface. The statistic is written as follows:

$$G_i^*(d) = \sum_{j=1}^{n} w_{ij}(d)x_j \Big/ \sum_{j=1}^{n} x_j ,$$

where $[w_{ij}]$ is a one–zero spatial weight matrix, with value one for all links defined as being within distance d of a given i, and value zero for all other links. Thus the numerator is the sum of all x_j within d of i, where i represents the rows of the grid, j represents the grid columns, and d is the number of grid cells away from cell ij, expressed in terms of the scale of the grid cells. In our analysis, each grid cell represents an area of 1 km × 1 km. The denominator is the sum of all x_j.

A positive $G_i^*(d)$ score indicates a spatial clustering of high values around a given point, whereas a negative $G_i^*(d)$ indicates a spatial clustering of low values. The distance at which these scores are recorded is usually found empirically by deriving scores at repeated distances out from each point, and then looking for that distance (called the 'critical' distance) at which scores peak at either high or low values.

We shall put the $G_i^*(d)$ statistic to two uses. First, it has the ability to locate 'hot spots' where low or high values are clustered. These spatial clusterings may then be investigated further (either qualitatively or quantitatively) to discover the sources of the clustering. The changes in spatial clustering between two or more dates can also be determined in this way.

The second use of the $G_i^*(d)$ statistic is as a spatial filter to extract the spatially autocorrelated portion of each of the variables in the regression variable and then to reintroduce the spatial variable into the regression equation as a separate factor (Getis, 1995; Scott, 1999). In this application, the filtered value x_i^f for a variable x_i is found as follows:

$$x_i^f = \frac{x_i \mathrm{E}(G_i^*)}{G_i^*(d)} ,$$

where

$$\mathrm{E}(G_i^*) = \frac{\sum w_{ij}(d)}{n-1} .$$

Then the difference between the original variable, x_i, and the filtered variable, x_i^f, is a new variable, x_i^{sp}, which represents the spatial effects embedded in x_i (Getis, 1995). These two variables, x_i^f and x_i^{sp}, replace the original variable, x_i, in the regression equation to produce a spatially filtered regression model.

Results

Is fertility spatially clustered?

We first utilized the global spatial statistic, Moran's I, to test the null hypothesis that fertility in Menoufia was spatially independent. In 1976 the normalized random z-score for Moran's I for the rate R^{nr} was 6.82, indicating a statistically significant amount of spatial autocorrelation, thus leading us to reject the null hypothesis that fertility is spatially independent in Menoufia. In 1986 the rate R^{nr} produced a normalized random z-score for Moran's I of 5.50, again indicating a statistically significant level of spatial autocorrelation.

The $G_i^*(d)$ statistic, as described above, provides a more precise way to test for spatial dependence. For any given cell in the grid, its statistically significant difference from spatial independence is given by the ratio of the $G_i^*(d)$ statistic relative to its expected value at a calculated critical distance d. We calculated the critical distance as that distance at which the filtering process has removed the spatial autocorrelation (measured by Moran's I) from the variable (Scott, 1999).

Any cell with a $G_i^*(d)$ value that is statistically significant at the 0.05 level would cause us to reject the null hypothesis of no clustering and would assign that cell to a cluster of either high or low fertility, depending upon the sign of G^*. In 1976 the critical distance was 5 km, indicating that, on average, villages that were clustered were most similar in fertility levels to those within a 5 km radius. In 1986 the critical distance was 4 km. Table 2 shows the characteristics of the villages by clustering. In both 1976 and 1986 the summary statistics were in the predicted direction, with the villages clustered around low fertility also exhibiting lower than average proportions married, lower proportions of illiterate women, and lower adult sex ratios. In 1986, but not in 1976, the lower fertility clusters were associated with more populous villages. On each characteristic the villages clustered around high fertility exhibited the opposite patterns—higher proportions married, higher proportions of illiterate women, and higher adult sex ratios.

Figure 5 shows the spatial clustering of high and low fertility in 1976, where villages in high-fertility clusters were those whose normalized z scores for the $G_i^*(d)$ statistic

Table 2. Characteristics of villages by clustering, 1976 and 1986.

Fertility cluster by year	N	R^{nr}	Proportion of females married	Proportion of females illiterate	Sex ratio in reproductive ages	Total population
1976						
low	47	1.94	0.61	0.77	0.86	4821
not clustered	200	2.15	0.64	0.81	0.86	5936
high	36	2.43	0.70	0.89	0.90	5221
1986						
low	25	2.03	0.59	0.61	0.89	10481
not clustered	224	2.29	0.63	0.67	0.94	7051
high	34	2.69	0.69	0.77	0.98	6788

Notes: N, number; R^{nr}, net reproduction rate.

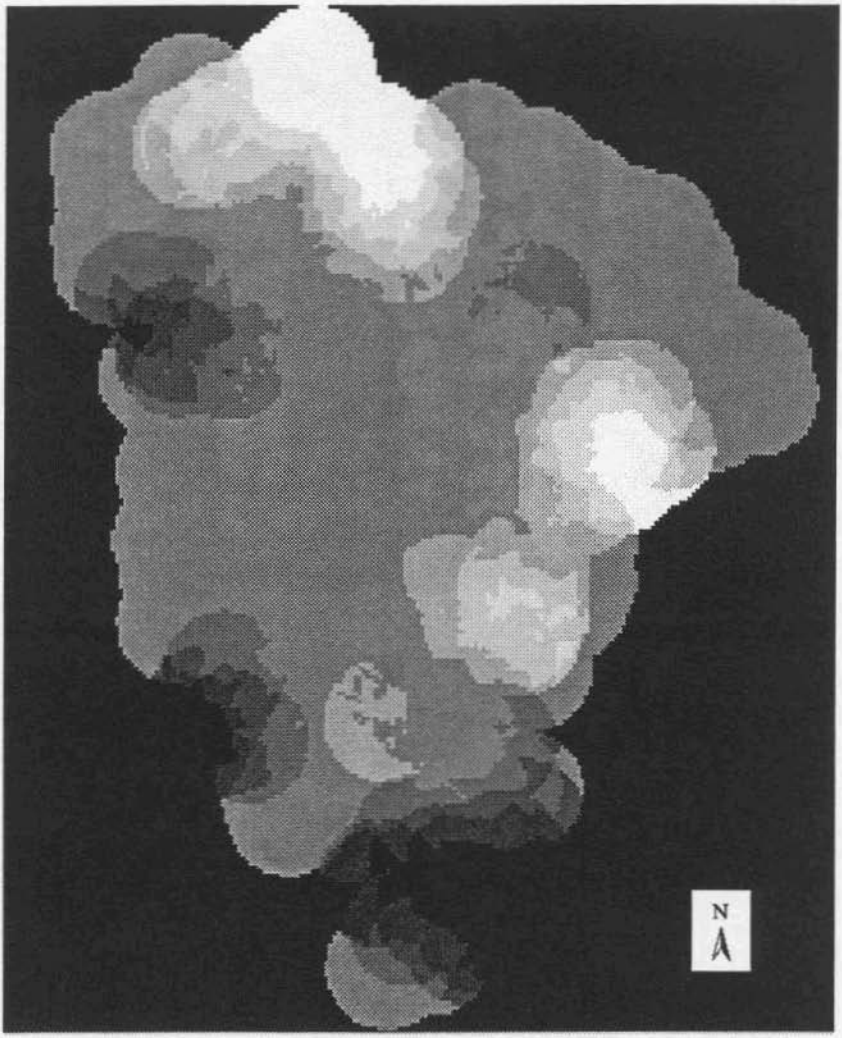

Figure 5. Clustering of fertility in 1976 in the Menoufia Governorate, Egypt. The darker circles indicate clustering of high fertility, whereas the lighter circles indicate clustering of low fertility.

were above 2, villages in low-fertility clusters were those whose normalized z scores for the $G_i^*(d)$ statistic were below -2, and those villages with normalized z scores between -2 and 2 were considered not to be clustered. In 1976 the clusters of low fertility were found in the north and northeast, whereas the clusters of high fertility were concentrated in the south and southwest of the governorate.

Although the pattern of clustering shown in figure 5 could be interpreted as being influenced by edge effects, the clustering in 1986, shown in figure 6 (over), seems to belie that explanation. In 1986 the clustering of low fertility had moved toward the middle of the governorate, although still concentrated in the north, whereas the clustering of high fertility was more concentrated in the southern portion of the governorate. Although the southern portion of Menoufia is closest to Cairo, it is also the site of the Barrage—the dam that controls water from the Nile as it enters the Delta region. This is rich agricultural land with centuries, if not millennia, of rural tradition that almost certainly

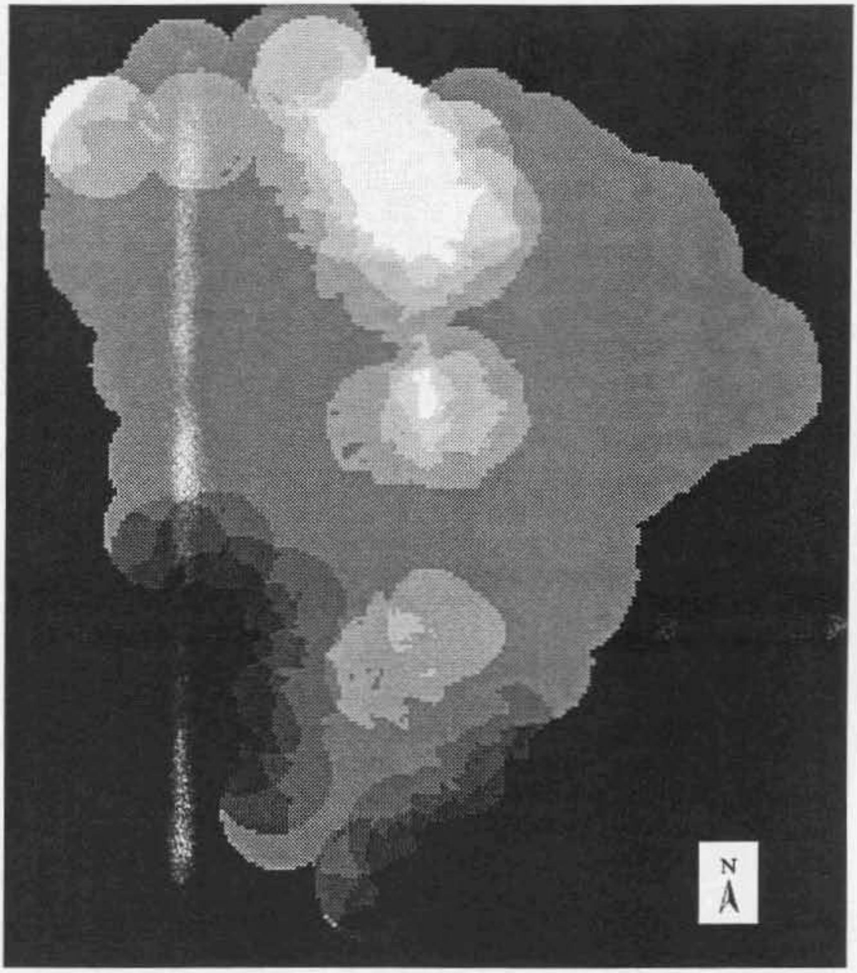

Figure 6. Clustering of fertility in 1986 in the Menoufia Governorate, Egypt. The darker circles indicate clustering of high fertility, whereas the lighter circles indicate clustering of low fertility.

contributes to the maintenance of low levels of education, low levels of female labor-force participation, and higher than average levels of fertility.

The data in figure 7 illustrate the change in the pattern of clustering between 1976 and 1986. Villages were categorized according to the combinations of clustering in the two time periods. Thus the lightest shades of clustering in figure 7 are assigned to villages that were in low-fertility clusters in both 1976 and 1986, and the next lighter shade indicates villages that moved from not being clustered in 1976 to being in low-fertility clusters in 1986. The data thus show the concentration of lower fertility in the north, and the diffusion of lower fertility in that region. At the other extreme, the darkest shading is assigned to villages that were in high-fertility clusters in both 1976 and 1986 and the next darker shading reflects villages that went from not being clustered in 1976 to being in a high-fertility cluster in 1986. These villages are concentrated in the southern portion of the region. In general the changes between 1976 and 1986 shown in figure 7 exhibit a spatial diffusion effect, with a spread of higher than average fertility to contiguous villages, and a spread of lower than average fertility to contiguous villages.

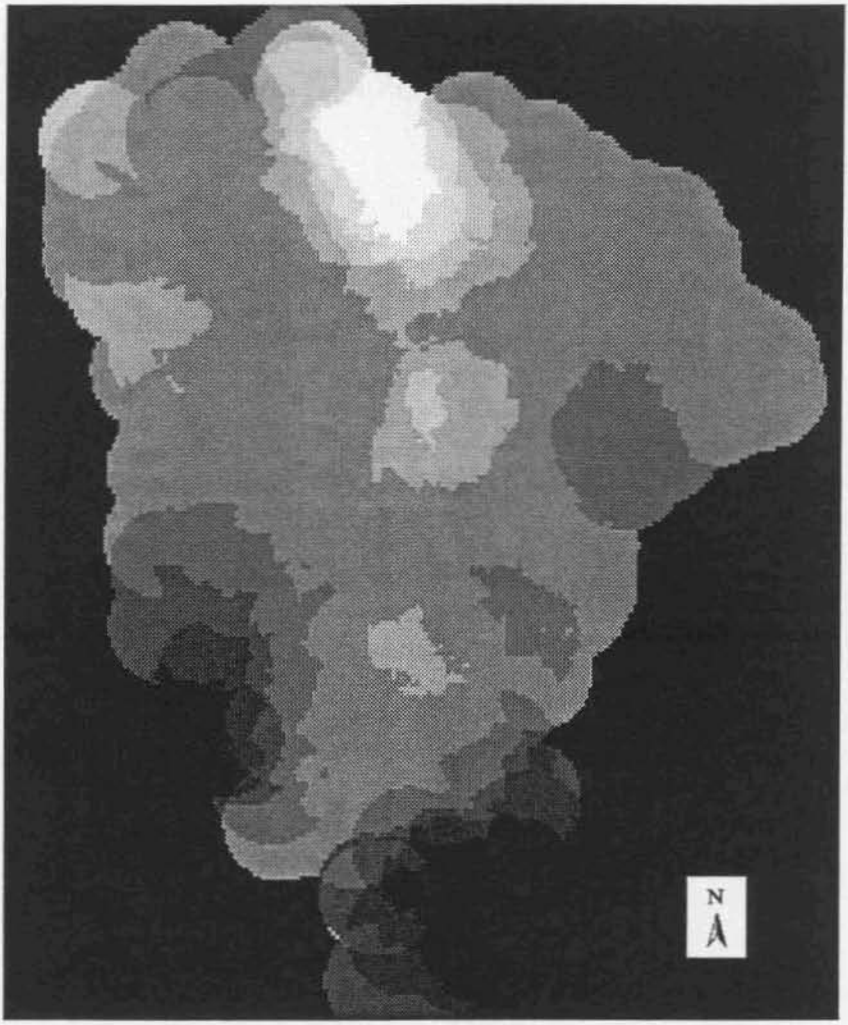

Figure 7. Change in the clustering of fertility between 1976 and 1986. The lighter shades indicate the maintenance of or shift toward the clustering of lower fertility, whereas the darker shades indicate the maintenance of or shift toward higher fertility clustering.

Quantifying the importance of spatial variability

It is clear from figures 5 through 7 that spatial variability in fertility exists in Menoufia. It does matter where you are—lower fertility is clustered in the north, and higher fertility is clustered in the south. How important is this spatial effect as a determinant of fertility levels? We used the technique of spatial filtering of variables in a regression model, as discussed above, to try to answer this question. First, we developed an OLS regression model that did not include a spatial component, echoing the typical model in demographic analysis. Second, we filtered the statistically significant predictor variables to assess the importance of the spatial effect.

The basic nonspatial model is that the fertility level in a village is a function of female illiteracy, controlling for the sex ratio at the reproductive ages (as a control for the effect of out-migration), the percentage of adult women who are currently married (as a control for the effect of marital status on the measure of fertility that we calculated), and for total population size (as a control for urbanness).

Table 3. Initial ordinary least-squares regression for 1976.

Variable	Unstandardized coefficient	Standardized β	t	Significance of t	$z(I)$
Dependent variable, R^{m*}					6.82
Female illiteracy	0.495	0.205	3.486	0.001	7.98
Proportion married	1.965	0.446	7.362	0.000	7.89
Sex ratio at reproductive ages	0.174	0.059	1.188	0.236	1.30
Population size	−0.00005	−0.138	−2.894	0.004	0.96
R	0.629				
Adjusted R^2	0.387				
$z(I)$ for residuals	0.87				

Note: see text for an explanation of the variables.

In 1976 three variables emerged as statistically significant predictors of fertility in Menoufia: female illiteracy (higher levels were associated with higher fertility), the percentage of adult women currently married (a higher percentage married being associated with higher fertility), and total population size (the larger the village, the lower the fertility). The R value was 0.629, with an adjusted R^2 of 0.387, as shown in table 3. The residuals were not autocorrelated because much of the spatial variability was accounted for by the proportion married. Without that variable in the equation, the residuals were significantly autocorrelated, whereas the introduction of the variable proportion married into the equation removed the spatial autocorrelation from the residuals.

Two of the predictor variables—female illiteracy and proportion married—have statistically significant levels of spatial autocorrelation, whereas the other two predictor variables do not. The two with spatial autocorrelation were then filtered to decompose the spatial component from the nonspatial (called the 'filtered') component by using the method described above. The results are shown in table 4, where it can be seen that although the R and adjusted R^2 values are essentially unchanged from the unfiltered model, we are now able to assess the relative contribution of the spatial component to that explained variation. The data show that the filtered (nonspatial) component of the

Table 4. Spatially filtered ordinary least-squares regression for 1976.

Variable	Unstandardized coefficient	Standardized β	t	Significance of t	$z(I)$
Dependent variable, R^{m*}					6.82
Filtered female illiteracy	0.417	0.149	2.699	0.007	−0.99
Filtered proportion married	1.724	0.329	5.907	0.000	0.48
Spatial female illiteracy	0.640	0.167	2.563	0.011	10.92
Spatial proportion married	2.322	0.331	5.024	0.000	11.76
Sex ratio at reproductive ages	0.178	0.060	1.223	0.222	1.30
Population size	−0.00005	−0.148	−3.111	0.002	0.96
R	0.637				
Adjusted R^2	0.393				
$z(I)$ for residuals	0.67				

Note: see text for an explanation of the variables.

Table 5. Initial ordinary least-squares regression for 1986.

Variable	Unstandardized coefficient	Standardized β	t	Significance of t	$z(I)$
Dependent variable, R^{nr}					5.50
Female illiteracy	1.768	0.520	9.955	0.000	6.14
Proportion married	1.443	0.256	5.038	0.000	5.00
Sex ratio at reproductive ages	0.007	0.016	0.364	0.716	2.05
Population size	−0.000002	−0.055	−1.163	0.246	3.04
R	0.694				
Adjusted R^2	0.482				
$z(I)$ for residuals	3.15				

Note: see text for an explanation of the variables.

proportion married has a standardized β coefficient that is virtually the same as the spatial component of that variable, indicating that the spatial component explains about half of that variable's relationship to fertility levels. The spatial component of female illiteracy is slightly more important than the filtered component, as can be seen in table 4. Summing the absolute values of the statistically significant standardized β coefficients, we can determine that the spatial component accounts for 33% of the explained variation in the net reproduction rate in 1976.

The initial regression model for 1986 showed that a higher proportion of the variation in the net reproduction rate was explained than in 1976—with an R value of 0.694 and an adjusted R^2 of 0.482, as can be seen in table 5. In 1986 the female illiteracy variable was a more important predictor of fertility than was the percentage married, and neither the adult sex ratio nor the total population size was statistically significantly related. The spatial component was also more noticeable than in 1976 because all four predictor variables exhibited spatial autocorrelation and the residuals were also spatially autocorrelated.

Table 6. Spatially filtered ordinary least-squares regression for 1986.

Variable	Unstandardized coefficient	Standardized β	t	Significance of t	$z(I)$
Dependent variable, R^{nr}					5.50
Filtered female illiteracy	1.920	0.480	9.496	0.000	−0.64
Filtered proportion married	1.057	0.170	3.587	0.000	−0.36
Filtered sex ratio at reproductive ages	0.002	0.006	0.136	0.892	−0.58
Filtered population size	−0.00000008	−0.063	−0.379	0.705	0.03
Spatial female illiteracy	0.913	0.157	2.716	0.007	26.80
Spatial proportion married	3.352	0.317	4.755	0.000	25.72
Spatial sex ratio at reproductive ages	0.640	0.051	0.332	0.332	28.96
Spatial population size	−0.00000004	−0.021	−0.131	0.896	0.22
R	0.717				
Adjusted R^2	0.513				
$z(I)$ for residuals	1.32				

Note: see text for an explanation of the variables.

All four predictor variables were filtered in 1986 and the resulting regression model is shown in table 6. Filtering raised the explained variation from 0.482 to 0.513, although female illiteracy and the proportion of women married remained as the only statistically significant variables in the equation. In 1986 the most important predictor was the filtered (nonspatial) component of female illiteracy, followed by the spatial component of the proportion married, then the filtered component of the proportion married, and finally the spatial component of female illiteracy. If we once again sum the standardized β coefficients of the statistically significant variables we find that the spatial component accounts for 42% of the explained variation—a higher fraction than in 1976.

Discussion and conclusion

The period from 1976 to 1986 was a period of overall relative stability in fertility levels in rural Egypt and not until the 1996 census data become available at the village level will we be in a position to track significant changes in fertility over time. However, it is clear that at least by 1976 there were clear spatial patterns to fertility in Menoufia and our analysis suggests that these spatial patterns were even more definitive in 1986 than they had been in 1976. This seems to suggest the existence of some momentum for change, which we hypothesize will be observable when the 1996 data become available. The southern portion of the governorate was more obviously the location of higher than average fertility in 1986 than had been true in 1976 and we would predict that the clustering of lower fertility will have exhibited a southward drift or diffusion by 1996. The results from our spatial filtering procedure suggest that some of this effect will be due solely to where villages are located, regardless of any changes in female education. The analysis also suggests that improving levels of female education will have been the most important human capital influence on fertility between 1986 and 1996. By identification of those areas in which low fertility is already clustered, government policy can potentially be efficiently employed to focus family activities in those areas, to ensure the maintenance of declining fertility which, our analysis suggests, will have a tendency to spread to nearby villages.

Our interpretation of results could have been influenced by our choice of a cell size of 1 km for the raster grid produced by the classification of satellite images into built or nonbuilt land cover. This cell size had the advantage of minimizing the total number of cells, thus easing the burden of calculation, but it had the disadvantage of eliminating a few villages from the analysis because of their overlap in a cell with another village, reducing the number of villages in the spatial analysis from the original 312 down to 283. However, experimentation with alternate cell sizes suggested that these results were not very sensitive to the choice of cell size: the pattern of results was identical with grid cells of size 500 m and 250 m. Furthermore we determined that the nonspatial regression results were virtually identical for the villages (cells) used in the spatial analysis as for the entire set of villages.

In conclusion we have shown that the use of remotely sensed images can allow us to locate more accurately the population in a rural area and, by incorporating the results of the classification of that image into a GIS, we are able to create a raster grid that allows us to build the spatial component into OLS regression and, in effect, decompose the explained variation into that which is explained by spatial (such as diffusion or culture) factors and that which is explained by nonspatial (such as human capital) factors. In the rural governorate of Menoufia, Egypt in both 1976 and 1986 the variability in fertility seemed to be explained independently by both factors.

Acknowledgements. Earlier versions of this paper were presented at the 1998 Annual North American Meeting of the Regional Science Association International, Santa Fe, New Mexico, November 1998, and the Annual Meeting of the Population Association of America, New York, March, 1999. This research was supported by a grant from the Andrew Mellon Foundation.

References

Arriaga E E, 1994 *Population Analysis With Microcomputers* US Bureau of the Census, Washington, DC

Bracken I, Martin D, 1995, "Linkage of the 1981 and 1991 UK Censuses using surface modelling concepts" *Environment and Planning A* **27** 379 – 390

Chen D, Getis A, 1998, "Point pattern analysis (PPA)", Department of Geography, San Diego State University, CA

Chesnais J-C, 1992 *The Demographic Transition: States, Patterns, and Economic Implications* (Clarendon Press, Oxford)

Chou Y-H, 1996 *Exploring Spatial Analysis in Geographic Information Systems* (OnWard Press, Santa Fe, NM)

Cleland J, Wilson C, 1987, "Demand theories of the fertility transition: an iconoclastic view" *Population Studies* **41** 5 – 30

Cliff A D, Ord J K, 1981 *Spatial Processes: Models and Applications* (Pion, London)

Coale A, 1973, "The demographic transition", in *International Population Conference, Liege 1973* volume 1, International Union for the Scientific Study of Population, 5 Rue Forgeur, 4000 Liege, Belgium, pp 53 – 72

Davis K, 1945, "The world demographic transition" *The Annals of the American Academy of Political and Social Science* **237** 1 – 11

Davis K, 1963, "The theory of change and response in modern demographic history" *Population Index* **29** 345 – 366

Easterlin R, Crimmins E, 1985 *The Fertility Revolution: A Supply – Demand Analysis* (University of Chicago Press, Chicago, IL)

Entwisle B, Casterline J, Sayed H A-A, 1989, "Villages as contexts for contraceptive behavior in rural Egypt" *American Sociological Review* **54** 1019 – 1034

Firebaugh G, 1979, "Structural determinants of urbanization in Asia and Latin America, 1950 – 1970" *American Sociological Review* **44** 199 – 215

Gadalla S, 1978, "Is there hope? Fertility and family planning in a rural community in Egypt", Carolina Population Center, University of North Carolina, Chapel Hill, NC

Getis A, 1995, "Spatial filtering in a regression framework: examples using data on urban crime, regional inequality, and government expenditures", in *New Directions in Spatial Econometrics* Eds L Anselin, R Florax (Springer, Berlin) pp 172 – 185

Getis A, Ord J K, 1992, "The analysis of spatial association by use of distance statistics" *Geographical Analysis* **24** 189 – 206

Goodchild M F, Anselin L, Deichman U, 1993, "A framework for the areal interpretation of socioeconomic data" *Environment and Planning A* **25** 383 – 397

Jones H, 1990 *Population Geography* 2nd edition (Guilford Press, London)

Kirk D, 1996, "Demographic transition theory" *Population Studies* **50** 361 – 387

Knodel J, van de Walle E, 1979, "Lessons from the past: policy implications of historical fertility studies" *Population and Development Review* **5** 217 – 245

Langford M, Unwin D J, 1994, "Generating and mapping population density surfaces within a geographical information system" *Cartographic Journal* **31**(1) 21 – 25

Lutz W, 1994, "The future of world population", Population Bulletin 49, Population Reference Bureau, Washington, DC

McDonald P, 1993, "Fertility transition hypothesis", in *The Revolution in Asian Fertility: Dimensions, Causes and Implications* Eds R Leete, I Alam (Clarendon Press, Oxford)

Makinson C, 1986 *Sex Differentials in Infant and Child Mortality in Egypt* PhD Thesis, Department of Sociology, Princeton University, Princeton, NJ

Martin D, 1996, "An assessment of surface and zonal models of population" *International Journal of Geographical Information Systems* **10** 973 – 989

Mesev V, 1998, "The use of census data in urban image classification" *Photogrammetric Engineering and Remote Sensing* **64** 431 – 438

Notestein F W, 1945, "Population: the long view", in *Food for the World* Ed. T W Schultz (University of Chicago Press, Chicago, IL)

Omran A, 1971, "The epidemiological transition: a theory of the epidemiology of population change" *Millbank Memorial Fund Quarterly* **49** 509 – 538
Openshaw S, 1984 *The Modifiable Area Unit Problem* CATMOG 38 (Geo-abstracts, Norwich)
Openshaw S, Rao L, 1995, "Algorithms for reengineering 1991 Census geography" *Environment and Planning A* **27** 425 – 446
Ord J K, Getis A, 1995, "Local spatial autocorrelation statistics: distributional issues and an application" *Geographical Analysis* **27** 286 – 306
Reed H, Briere R, Casterline J, 1999 *The Role of Diffusion Processes in Fertility Change in Developing Countries: Report of a Workshop* (National Academy Press, Washington, DC)
Scott L, 1999, "The filter software", Department of Geography, San Diego State University, CA
Smith D P, 1992 *Formal Demography* (Plenum, New York)
Tellier L-N, Vertefeuille C, 1995, "Understanding spatial inertia: center of gravity, population densities, the Weber problem, and gravity potential" *Journal of Regional Science* **35** 155 – 164
Thompson W, 1929, "Population" *American Journal of Sociology* **34** 959 – 975
Tolnay S E, 1995, "The spatial diffusion of fertility: a cross-sectional analysis of counties in the American South, 1940" *American Sociological Review* **60** 299 – 308
US Bureau of the Census, 1999, "International database"; http://www.census.gov/ipc/www
Watkins S C, 1991 *From Provinces Into Nations: Demographic Integration in Western Europe, 1870 – 1960* (Princeton University Press, Princeton, NJ)
Weeks J R, 1999 *Population: An Introduction to Concepts and Issues* 7th edition (Wadsworth, Belmont, CA)