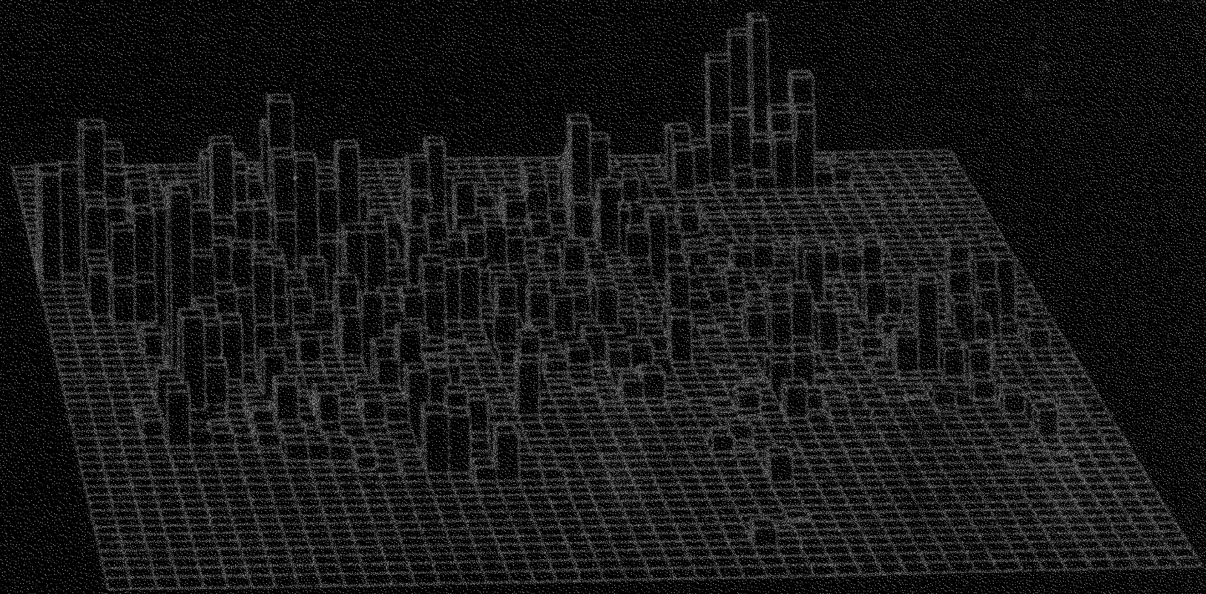
CRU

The effects of the size of areal units on ratio and chi-square mapping

by M. Visvalingam and J. C. Dewdney



Working paper 10

Census Research Unit
Department of Geography
University of Durham

UNIVERSITY OF DURHAM

DEPARTMENT OF GEOGRAPHY

CENSUS RESEARCH UNIT

WORKING PAPER No. 10

JUNE 1977

THE EFFECTS OF THE SIZE OF AREAL UNITS ON RATIO AND CHI-SQUARE MAPPING

M. VISVALINGAM

and

J.C. DEWDNEY

CONTENTS

LIST OF ILLUSTRATIONS

FIGURES (Pages 32-57)

In order to facilitate comparisons, the 36 maps and diagrams are assembled in a single block following the text. With the exception of Figures 2-7, which portray the distribution of old people, and Figures 1 (base map) and 22 (Lorenz curves), the maps all portray patterns of sex composition. The following details may be useful for cross-reference between the various levels of aggregation and methods of representation :

(a) Aggregation levels

1 x 1 km	Figures 2,3,4,5,6,7,8,9,15,16,23,24,30,31
2 x 2 km	Figures 10,11,17,18,25,26,32,33
5 x 5 km	Figures 12,13,19,20,27,28,34,35
LAAs	Figures 14,21,29,36

(b) Method of representation

Absolute deviation	Figures 2,3
Ratios	Figures 4,5,15,16,17,18,19,20,21,30,31,32,33,34,35,36
X_s^2	Figures 6,7,8,9,10,11,12,13,14,23,24,25,26,27,28,29

(c) Masculinity

High	Figures 8,10,12,14,15,17,19,21,23,25,27,29,30,32,34,36
Low	Figures 9,11,13,14,16,18,20,21,24,26,28,29,31,33,35,36

A full list of figure titles follows.

TABLES

THE EFFECTS OF THE SIZE OF AREAL UNITS ON RATIO AND CHI-SQUARE MAPPING

1. INTRODUCTION

Census data, which consist mainly of frequency counts of demographic and socio-economic characteristics, are made available only in aggregated form in order to protect the privacy and preserve the confidentiality of information relating to individuals. Such data are available at different levels of aggregation within two separate nested hierarchies of areal units based on administrative divisions (Clarke and Rhind, 1975, p.3) and grid squares respectively. Data for the latter category are available from the Office of Population Censuses and Surveys (OPCS) at 100-metre, one-kilometre and ten-kilometre levels and can be produced for other levels by aggregation.

The type, shape and size of the areal units have a pronounced influence upon statistical analysis and mapping of population data. Clarke and Rhind (1975) discussed the relationships between size of areal unit and aspatial statistical descriptions of four variables, namely the sex ratio (expressed as a masculinity proportion, i.e. males ÷ total population), persons per square kilometre, persons per household and persons per room. They also demonstrated the effects of spatial aggregation of data on patterns of population distribution.

In this paper, we consider the ways in which numerical measures for expressing bivariate data are influenced by the type and size of the areal and population units involved. In a previous paper (Visvalingam, 1976) it was shown that ratios and numerical differences are inadequate measures of comparison when bivariate data, such as sex ratios, refer to a wide range of population sizes. While numerical differences and the magnitudes of deviations from expectation (observed - expected) favour large populations, ratios have the opposite effect of giving the most extreme values in the case of small populations. The binomial test statistics (z score for proportions) or the chi-square statistic (X^2), in the two-category case, were identified as being more suitable for the ordering of populations of widely differing size, since they include a correction for the effects of sample size. However, even the X^2 measure is unreliable when the population size is very small, since the likelihood of the chance occurrence of extreme values is greater. The conventional solution is to aggregate data for adjacent locations and to ascertain whether a similar spatial pattern is produced.

Consequently, while this paper is primarily concerned with the scale-dependence of ratio and X_s^2 (signed X^2; see Visvalingam, 1976) ordering schemes, as reflected in the use of data for different sizes of grid square and for irregular-sized administrative units, there is also scope for checking whether the problems associated with the use of ratios are peculiar to grid-square data or whether equally serious problems also exist with the application of ratios to data relating to administrative divisions. It is hoped that this will permit a more accurate interpretation of previously-published distribution maps of ratio variables on an administrative unit basis, such as those produced from 1961 census data by Hunt and others (1968). The exercise was also designed to assess the extent to which the problems associated with very small samples might be alleviated by the aggregation of data for adjacent areal units. It should be emphasised that, while the present paper is based entirely on population census data, the methods, arguments and conclusions are applicable to many other data sets, particularly where these are collected and published on an areal basis.

2. DATA SOURCES AND METHODS

2.1 The data

Two main data sources were used in the investigation. Data at the one-kilometre grid square level were supplied by OPCS on magnetic tape from which, after compaction and storage (Visvalingam, 1975, Visvalingam and Perry, 1976), the variables required could be extracted for statistical manipulation and mapping (Rhind, Evans and Dewdney, 1977). Local Authority Area (LAA) data for County Boroughs, Municipal Boroughs and Urban Districts and Rural Districts were taken from the appropriate published census volumes and assembled in a separate computer file.

2.2. The study area

The area chosen to illustrate the study is indicated in Figure 1. In grid-square terms, it comprises the entire land area within the three

100-kilometre squares whose south-west corners are defined by grid references 300 400, 200 300 and 300 300 respectively. Within this area there are 16,612 inhabited one-kilometre squares, but this 'inhabited area' is not constant at different levels of areal aggregation, since the latter process adds uninhabited squares to inhabited ones, thus increasing the total area to which the aggregated data apply. Neither is the true land area precisely the same as any grid-square derived area, since numerous squares are cut by the coastline.

In terms of Local Authority Areas (LAA's), the study area comprises all or part of fourteen pre-1974 counties - Anglesey, Caernarvon, Cheshire, Cumberland, Denbigh, Derbyshire, Flint, Lancashire, Merioneth, Montgomery, Shropshire, Staffordshire, Westmorland and Yorkshire (West Riding) - and contains 284 units (23 County Boroughs, 54 Municipal Boroughs, 133 Urban Districts and 74 Rural Districts). A problem arose where LAA's were divided by the chosen grid-square boundaries. In a few cases, where only a very small section of the LAA in question fell within the grid-bounded area, its population was ignored but, where a sizeable part of an LAA fell within the grid-bounded area, its entire population was included in the analysis. Thus the area for which LAA data were used is considerably larger than that of the grid-bounded study area and the total populations also differ, being 9,046,826 in the LAA's and 8,497,690 in the grid-square data.

This particular section of the country was chosen for a variety of reasons. In particular, it had to be large enough to contain a sufficient number of areal units (LAA's as well as grid squares) for conclusions to be statistically valid, but at the same time not unmanageably large with respect to computer storage and time resources. In addition, it had to be an area where marked spatial sorting of males and females (sex ratio being the parameter chosen for closest study) could be assumed to exist; one with a wide range of population sizes and one displaying a variety of general conditions (rural/urban, industrial/residential, high/low density); in short an area of sufficient variety to illustrate most situations likely to occur within the U.K. as a whole.

2.3. Methods

Visvalingam (1976) compared ratio and X_s^2 measures, using the one-kilometre grid square data for sex composition in County Durham. However, the various methods of representation produce different spatial patterns even when the variable being mapped is subject to a high degree of spatial sorting. The component of old people (those aged 65 and over), for example, is one such variable. Of the 8,783 unsuppressed one-kilometre squares within the present study area for which age data are available, the five per cent (approximately 440 squares) falling in each extreme class were plotted. The resultant maps (Figs. 2-7) show significant differences. Absolute deviations (Figs 2 and 3), as anticipated, show an urban preference (note that all small populations of less than 113 persons were automatically discarded from consideration by this measure). Ratios, on the other hand (Figs 4 and 5), pick out the small populations, giving a rather dispersed pattern of extremes. The X_s^2 maps (Figs. 6 and 7) seem to be intermediate in type.

For the present investigation, we revert to the use of data for sex composition, primarily because values are available for all populated one-kilometre squares, providing 16,612 data units rather than only 8,783 unsuppressed squares. A variable with data coverage for the entire inhabited area was considered to be more appropriate for an exercise involving aggregation. The masculinity proportion (males ÷ total) for the grid-square population (which remains constant at 8,497,690 for all levels of aggregation) was 48.39 per cent; that for the larger (9,046,826) LAA population was 48.50 per cent, very close to the Great Britain average of 48.53 per cent. The X_s^2 values were calculated using an expectation based on the Great Britain figure.

Spatial and aspatial characteristics of the sex composition data for one -, two- and five-kilometre squares and for Local Authority Areas were compared. At the higher levels of grid-square aggregation, the origin of the grid has some effect on the maps and statistics; however, this is constant and irrelevant when the bivariate measures are compared under the same conditions. Some of the aspatial univariate characteristics of the data are given in Table 1.

TABLE 1 : SUMMARY STATISTICS OF TOTAL POPULATION, MALES AND FEMALES FOR THE FOUR LEVELS OF AGGREGATION

	Grid squares			L.A.A 's
	1x1 km	2x2 km	5x5 km	
Number of areal units	16.612	4,940	906	284
Total population : number	8,497,690	8,497,690	8,497,690	9,046,826
mean	511.54	1,720.18	9,379.35	31,855.02
standard deviation	1,339.30	4,293.57	21,193.81	57,594.52
coefficient of variation	2.618	2.496	2.260	1.808
minimum	1	1	2	729
maximum	16,445	46,914	213,224	610,113
range	16,444	46,913	213,222	609,384
Males : number	4,111,921	4,111,921	4,111,921	4,387,881
mean	247.53	832.37	4,538.54	15,450.29
standard deviation	647.15	2,071.62	10,225.45	27,867.21
coefficient of variation	2.614	2.489	2.253	1.804
minimum	0	0	2	360
maximum	8,042	22,523	102,140	293,550
range	8,042	22,523	102,138	293,190
Females : number	4,385,769	4,385,769	4,385,769	4,658,945
mean	264.01	887.81	4,840.80	16,404.74
standard deviation	693.35	2,223.99	10,973.15	29,637.96
coefficient of variation	2.626	2.505	2.267	1.807
minimum	0	0	1	369
maximum	8,403	24,384	111,084	316,563
range	8,403	24,384	111,083	316,194
Percentage males (aggregate)	48.39	48.39	48.39	48.50

TABLE 2 : SUMMARY STATISTICS OF THE MASCULINITY RATIO FOR THE FOUR LEVELS OF AGGREGATION

	Grid squares			L.A.A's
	1x1 km	2x2 km	5x5 km	
Percentage males (ratio) :				
mean	50.77	50.51	49.81	48.40
standard deviation	10.98	7.73	4.29	1.68
coefficient of variation	0.216	0.153	0.086	0.035
mimimum	0.0	0.0	27.40	40.15
maximum	100.0	100.0	86.05	54.11
range	100.0	100.0	58.65	13.96

3. OBSERVATIONS AND EXPLANATIONS

3.1. The spatial pattern of sex composition

As a starting point, it seems appropriate to give a brief description of spatial patterns of sex composition within the study area. This raised the question of which method of portrayal was most suitable for such a general description. Although it might appear to pre-judge one of the issues under discussion in this paper, it was decided to use the X_s^2 map derived from one-kilometre square data, mainly because, in an earlier study of the County Durham data set (Visvalingam, 1976), X_s^2 was identified as the most satisfactory measure at the one-kilometre level. Figures 8 and 9, therefore, display one-kilometre grid squares with high and low masculinity respectively as identified by X_s^2 at the 95 per cent confidence level (usually associated with a X^2 value of 3.84 for one degree of freedom).

Before discussing the patterns revealed by the maps, it is necessary to recall the major demographic features influencing sex composition. The most important of these is age structure : since females have a significantly greater expectation of life than males and male mortality rates are higher than female at every age, masculinity declines with increasing age and we would expect to find the lowest masculinity associated with those populations whose age structure is most heavily weighted towards the upper age groups. Both age structure and sex composition are influenced by the migration components. Migrations are commonly sex-selective, in the sense that one sex or the other is dominant in most migration flows. They are also age-selective, the most common situation being that in which young adults are dominant, and the direction of their movement is influenced mainly by employment opportunities. In western Europe, however, and particularly in Britain, there is one major exception, namely the movement of elderly people on retirement. The age-selective nature of migrations itself has a strong influence on sex composition owing to the differing mortality of the two sexes and the differing fertility of the various age groups. Areas receiving young adult migrants are likely to have a youthful age structure, strengthened by their relatively high fertility, and thus to display high masculinity   Retirement areas on the other hand, have an elderly age structure, emphasised by the fact that the retired produce very few children, and thus tend to be areas of very low masculinity. All these effects are most clearly seen at the receiving end of migration flows, since movement normally takes place from a large number of areal units to a smaller number of centres of attraction.

Turning to the spatial patterns displayed in Figures 8 and 9 , in the first place there is a clear contrast, which will be discussed in more detail later, between the two maps. Squares with low masculinity (Fig. 9), in addition to being more numerous and having a different geographical distribution, are also concentrated and form more clearly defined blocks of territory than do the high masculinity squares, (Fig.8), which display a more dispersed, punctiform pattern.

Major blocks of contiguous squares with low masculinity are predominantly coastal and pick out holiday resort/retirement areas such as Grange U.D. (341 478*), Morecambe M.B. (343 463), Blackpool C.B. (330 436), Lytham St. Annes M.B. (334 429), Southport C.B. (333 416), Crosby M.B. (332 400), Wallasey C.B. (329 392), which includes New Brighton, Hoylake U.D. (324 288), Wirral U.D. (327 383) and the string of North Wales resorts from Prestatyn U.D. (306 383) to Llandudno U.D. (280 381). Smaller centres such as Windermere U.D. (341 497), Bangor M.B. (258 372), Caernarvon M.B. (249 363), Pwllheli M.B. (238 335) and Barmouth U.D. (263 317) also stand out, as do some of the small market towns of rural Wales, for example Dolgellau U.D. (274 317), Denbigh M.B. (305 366) and Ruthin M.B. (313 358). Other clusters of low masculinity squares occur in the inner suburbs of Liverpool C.B. (338 389), Manchester C.B. (385 395) and several smaller towns.

The high masculinity map is less clear in that the squares are scattered. They appear to occur at both ends of the urban-rural

---

* Six figure grid references are given after the names of Local Authority Areas and other place-names to assist the reader in locating them on the various maps, all of which carry grid-line numbers along their margins.

continuum, with groups in the central areas of Manchester and Liverpool and many individual squares in rural areas. High masculinity in Wolverhampton C.B. (392 301) and Walsall C.B. (399 301) could be related to the immigrant population; in Kirkby U.D. (343 398) it is the product of recent large-scale housing development, which has attracted a youthful population. Some special cases can be identified, including military establishments in squares 379 305 (Shifnal R.D.) and 231 376 (Valley R.D. of Anglesey) and prisons at Haverigg (314 479), Milnthorpe (351 480) and Manchester Strangeways (383 399). Several other areas of high masculinity are essentially urban-industrial, as in Warrington R.D. (364 391), Stoke-on-Trent C.B. (389 347) and Cannock U.D. (399 310). A common situation is that in which dispersed squares of high masculinity occur in districts which otherwise record near-average values and show a marked absence of low masculinity.

3.2. Effects of averaging

Progressive averaging within different sized areal units is reflected to some extent in the range and standard deviation of masculinity proportions given in Table 2. These averaging effects cannot be accurately located in ratio terms, since the spatial pattern of very high and very low masculinity values is sensitive to aggregation (see below, section 3.3) X_s^2 maps of sex composition (Figs.8 -14) for different sizes of areal unit are not strictly comparable owing to this progressive averaging at higher levels of aggregation. Evidence of some averaging exists, for example, in Anglesey (240 380), Shifnal R.D. (377 307) and Shrewsbury M.B. (350 313). However, this is not severe, as areas with high or low masculinity persist in, for example, Oswestry R.D. (331 327), the Lleyn peninsula (230 340), Barmouth U.D. (263 317), Dolgellau U.D. (274 317), Bangor M.B. (258 372), Caernarvon M.B. (249 363), Ruthin M.B. (313 358) and around the estuary of the Mersey (e.g. Formby U.D. : 329 408; Bootle C.B. : 335 395; Runcorn U.D. : 354 382). However, in the case of larger areal units, it is impossible to identify precise locations where demographic and economic factors, operating only within a limited area, are the cause of local anomalies or reversals within more extensive areas of high or low masculinity. For example, in general terms the coasts of North Wales and north-west Lancashire register low masculinity but, using the one-kilometre square data (Fig.) it is possible to pick out individual squares within Blackpool C.B. (330 436), Southport C.B. (333 416), Formby U.D. (329 408) and elsewhere with high masculinity. Higher levels of aggregation also involve a loss of such resolution within Liverpool C.B. (338 389), in the Fulwood U.D. (354 432)/Preston C.B. (355 430) area and in and around Burnley C.B. (384 432) and Blackburn C.B. (368 428); in all these cases, it is no longer possible to see the local pattern of high and low masculinity values.

The effects of aggregation of data for one-kilometre squares with very pronounced low (or high) masculinity with those for adjacent squares with near-average values are also apparent. In many cases, for example, in Kendal M.B. (352 492), Fleetwood M.B. (332 447), Preesall U.D. (337 448), Skipton U.D. (399 451), Llangollen U.D. (332 342), Llanfyllin M.B. (313 320) and Chester C.B. (341 366), the occurrence of one or two very low value squares at the one-kilometre level is sufficient to give very low values at all higher levels of aggregation. Since, at higher levels of aggregation, the number of areal units is reduced and, by definition, their size is increased, both the frequency and the areal extent of very low masculinity becomes exaggerated.

Similarly, the dispersed pattern of one-kilometre squares with high masculinity becomes more continuous at the two-kilometre and especially at the five-kilometre and LAA levels in the south-east of the region; accentuating the high masculinity in Cannock U.D. (339 310) and Stafford M.B. (393 322), around Stoke C.B. (389 347), in Wellington R.D. (367 317) and in the U.D's to the south. The same effect is seen in Ruthin R.D. (309 358) and in many of the LAA's between Liverpool (338 389) and Salford (382 399).

3.3. The effects of aggregation on ratios and chi-square

Despite the inevitable averaging and generalisation at a local level within the aggregated units, X_s^2 proved to be a fairly robust technique for ordering data and for mapping purposes. It did not appear to be over-sensitive to sub-optimal conditions of data with respect to variations in sample size, since the same general spatial patterns were repeated at the four levels of areal unit (Figs. 8-14). Zones of particularly low masculinity (a high proportion of females) recur, for example, in Windermere U.D. (341 497), along the coast, in Oswestry R.D. (331 327), in Chester C.B. (341 366) and in the southern part of the Manchester conurbation; high masculinity proportions recur in such industrial areas as Crewe M.B. (371 356), Warrington C.B. (362 388), Stretford M.B. (380 395) and Wolverhampton C.B. (392 301); in Shifnal R.D. (377 307) and Valley R.D. of Anglesey (233 380) and at Barrow-in-Furness C.B. (322 471). These features are seen on all the grid-square maps, though some disappear at the LAA level (Fig. 14).

The ratio maps (Figs. 15-21), on the other hand, are not consistent in the identification of extreme cases and seem to be more sensitive to changes in data resolution level. The extremes identified in the one- and two-kilometre square ratio maps were based on the standard scores of the ratios. Although the one-kilometre ratio map (Fig. 15) includes 606 squares with high male proportions (156 more than the X_s^2 map, Figure 8 , at the 95 per cent confidence level), the majority of these fall outside the areas picked out by the X_s^2 map, occurring mainly in small-population, generally rural, squares. Even on the two-kilometre ratio map (Fig.17), which includes 60 more cases of high masculinity than does the corresponding X_s^2 map (Fig. 10), there are even more extensive blanks in the south-east of the study area and in the Liverpool and Manchester conurbations. In addition, squares with extremely high proportions of females occur mainly in inland districts (Figs.16 and 18) and not along the coast.

Identification of extreme values on the basis of standard scores was found to be inadequate at the five-kilometre and LAA levels, owing to the marked reduction in the range of ratios and in variability in general (Table 1). The five-kilometre and LAA ratio maps thus contained numbers of extreme cases comparable to those displayed by the X_s^2 maps. For strict comparability in numbers, readers may wish to refer to the series of maps (Figs. 23-36) showing the ten per cent of all cases at the extremes of both distributions for the various levels of aggregation (see also 3.4 below).

At the five-kilometre square level, a major change is observable in the spatial patterns of the ratio maps. Areas of very low masculinity (Fig.20) begin to emerge along the coast in Blackpool C.B. (330 436), Lytham St. Annes M.B. (334 429) and the smaller holiday resorts and retirement centres around Morecambe Bay and in North Wales. Some of the market towns of rural Wales also register low masculinity, for example Ruthin M.B. (313 358) and Dolgellau U.D. (274 317). However, the ratio map of low masculinity at the five-kilometre level (Fig.20) continues to omit the areas of low masculinity around the estuary of the Mersey and in the south Manchester area which appear on the corresponding X_s^2 map (Fig.13). The masculine strongholds are still assigned to rural areas in Wales and West Yorkshire, such as Settle R.D. (380 469), Sedburgh R.D. (371 491) and Bowland R.D. (373 451), while the major industrial areas remain largely blank (Fig.19).

The LAA ratio map (Fig.21), on the other hand, is strikingly similar to the X_s^2 map for LAA data (Fig.14). This is to some extent to be expected since, at the 95 per cent significance level, 51.05 per cent of the areal units showed a significant departure from expectation. There are, however, important differences of detail. The ratio map ignores Warrington C.B. (362 388), St. Helen's C.B. (352 395), Stoke C.B. (389 347), Newcastle under Lyme M.B. (384 346) and Nantwich R.D. (363 352) for example, but emphasises the rural districts of North Wales.

Thus the spatial distributions of high and low masculinity, as expressed by the masculinity proportion, are seen to vary with changes in the size of the spatial unit of aggregation. X_s^2 maps, on the other hand, are more consistent, despite the evidence of local averaging. Chi-square is also more reliable in identifying the most extreme cases of low and high masculinity, independent of the size of the areal unit. Shifnal R.D. (377 307), which has the highest masculinity ratio in the LAA data set, consistently has the highest X_s^2 value at all aggregation levels. On the other hand, in the case of ratios, the area showing the highest masculinity varies with the size of the data unit. In the five-kilometre square case, that containing Haverigg Prison in Millom R.D. (315 492) has the highest value (86 per cent), but the two-kilometre and one-kilometre square data show 23 and 152 squares respectively with 100 per cent males. Again, there are several cases of zero males in the one- and two-kilometre square data (Table 3). While the five-kilometre square and LAA data register the lowest masculinity in Abergele U.D. (296 379) and Grange U.D. (341 478) respectively, the lowest X_s^2 values repeatedly occur in the larger coastal resorts of Blackpool C.B. (330 436) and Southport C.B. (333 416).

The ratio and X_s^2 maps for LAA data are similar, partly because both contain over 50 per cent of the data units. The ranking and scaling of these units by ratios and X_s^2 respectively are, however, markedly dissimilar (Table 4)and maps of the ten per cent of cases at each extreme (Figs.29 & 36) are quite different. Ratio maps of very high or very low masculinity omit, for example, (327 383) the towns of the Wirral peninsula (330 390), Stoke C.B. (389 347) and Wolverhampton C.B. (392 301) and emphasize rural Wales. Nevertheless, the performance of ratios is markedly better at higher levels of aggregation, despite the greater range of population sizes.

When the areal coverage of data is irregular, aggregation has the effect of increasing the range of sample sizes. In urban areas, aggregation of data for adjacent squares yields population units which are considerably larger. However, in rural areas, especially in the uplands, there may still be only a few households within an area of 25 sq. km. (the minimum population observed in a 5x5 km square was two persons). Thus aggregation of adjacent squares increases the range of population units under consideration (Table 1).

This increase in range may, however, be the product of a small number of very small and very large population units. The other statistical measures of dispersion, such as the standard deviation and the coefficient of variation (Table 1), are also influenced by these extremes and do not give a vivid description of the frequency distribution of population units at the different levels of aggregation. As an alternative, Lorenz curves were drawn for each level, plotting the cumulative percentage of total population against the cumulative percentage of data units which had been ranked by population size. The resulting Figure 22 shows that the distribution of population is

TABLE 3 : SIZE OF AREAL UNIT AND LOCATION OF EXTREMES OF MASCULINITY

SIGNED CHI-SQUARE (X_s^2) EXTREMES (E= 0.485)

Areal Unit	Minimum	Population base	Location	Masculinity (% males)	Excess males (m-f)	Maximum	Population base	Location	Masculinity (% males)	Excess males (m-f)
LAA	-490.670	151,860	Blackpool C.B (330 436)	45.659	-13,184	200.074	15,869	Shifnal R.D. (377 307)	54.112	1,305
5x5 km square	-408.230	31,519	330 415	42.812	- 4,531	622.940	3,971	375 305	68.295	1,453
2x2 km square	-266.314	19,911	334 416	42.720	- 2,899	1,210.454	1,770	378 304	89.830	1,410
1x1 km square	-223.532	277	329 335	3.610	-257	1,266.388	1,686	379 305	91.815	1,410

MASCULINITY (% MALES)

Areal unit	Minimum	Population base	Location	X_s^2 value	Excess males (m-f)	Maximum	Population base	Location	X_s^2 value	Excess males (m-f)
LAA	40.155	3,474	Grange U.D. (341 478)	-96.857	-684	54.112	15,869	Shifnal R.D. (377 307)	200.074	1,305
5x5 km square	27.399	866	295 370	-154.225	-391	86.047	644	310 475	363.997	465
2x2 km square	0.000	(1 to 2)	10 ties	-	-	100.000	(1 to 4)	23 ties	-	-
1x1 km square	0.000	(1 to 2)	64 ties	-	-	100.000	(1 to 5)	152 ties	-	-

TABLE 4 : HIGHEST AND LOWEST 10% OF RATIO AND X_s^2 VALUES FOR LOCAL AUTHORITY AREAS

RATIOS : 28 HIGHEST VALUES

	Area and location		Population	Males	Excess (m-f)	Ratio (% males)	X_s^2 value
1	Shifnal R.D.	(377 307)	15,869	8,587	1,305	54.112	200.082
2	Sedbergh R.D.	(371 491)	3,544	1,912	280	53.950	42.150
3	Clithero R.D.	(368 440)	9,456	5,075	694	53.670	101.177
4	Kirkham U.D.	(343 432)	6,436	3,405	374	52.906	50.011
5	Penllyn R.D.	(289 334)	2,313	1,205	97	52.097	11.980
6	Clun & Bishop's Castle R.D.	(333 301)	8,869	4,577	285	51.607	34.272
7	Machynlleth R.D.	(285 305)	2,502	1,291	80	51.599	9.618
8	Maelor R.D.	(345 341)	4,679	2,393	107	51.143	13.090
9	Newtown and Llanidloes R.D.	(306 301)	8,223	4,203	183	51.113	22.473
10	Millon R.D.	(315 492)	14,089	7,195	301	51.068	37.204
11	Warrington R.D.	(364 391)	50,420	25,684	948	50.940	120.191
12	Forden R.D.	(320 300)	4,768	2,427	86	50.902	11.012
13	Settle R.D.	(380 469)	13,992	7,115	238	50.850	30.949
14	Valley R.D.	(233 380)	15,048	7,647	246	50.817	32.354
15	Mkt. Drayton R.D.	(366 332)	16,782	8,519	256	50.763	34.401
16	Leek R.D.	(395 360)	13,247	6,717	187	50.706	25.806
17	Wellington R.D.	(367 317)	30,297	15,360	423	50.698	58.606
18	Llanfyllin R.D.	(307 317)	8,292	4,198	104	50.627	15.021
19	Stafford R.D.	(387 322)	23,000	11,606	212	50.461	35.407
20	Wellington U.D.	(365 311)	17,163	8,638	113	50.329	22.992
21	Stafford M.B.	(393 322)	55,001	27,652	303	50.275	69.413
22	Cheadle R.D.	(398 345)	40.098	20,120	142	50.177	45.152
23	Cannock U.D.	(399 310)	55,882	28,022	162	50.145	60.538
24	Hiraethog R.D.	(288 354)	4,183	2,096	9	50.108	4.328
25	Rainford U.D.	(349 400)	8,404	4,211	18	50.107	8.690
26	Bolton C.B.	(370 410)	154,199	74,709	219	50.073	147.876
27	Cannock R.D.	(391 311)	43,855	21,949	43	50.049	42.130
28	Ellesmere Pt. M.B.	(340 376)	61,637	30,831	25	50.020	57.035

TABLE 4 (cont.)

RATIOS : 28 LOWEST VALUES

	Area and location		Population	Males	Excess (m-f)	Ratio (% males)	X_s^2 value
257	Pwllheli M.B.	(238 335)	3,823	1,777	-269	46.482	-6.234
258	Adlerley Edge U.D.	(384 378)	4,470	2,076	-318	46.443	-7.573
259	Ruthin M.B.	(313 358)	4,338	2,011	-316	46.358	-7.970
260	Llangollen U.D.	(322 342)	3,117	1,441	-235	46.230	-6.428
261	St. Asaph R.D.	(305 375)	11,192	5,172	-848	46.212	-23.465
262	Beaumaris M.B.	(260 377)	2,102	970	-162	46.147	-4.661
263	Abergele U.D.	(296 379)	12,315	5,669	-977	46.033	-30.000
264	Thornton Clevelys U.D.	(333 443)	26,837	12,319	-2,199	45.903	-72.462
265	Hoylake U.D.	(324 388)	32,277	14,802	-2,673	45.859	-90.112
266	Preesall U.D.	(337 448)	3,987	1,824	-339	45.749	-12.083
267	Rhyl U.D.	(301 382)	21,821	9,970	-1,881	45.690	-68.985
268	Blackpool C.B.	(330 436)	151,860	69,338	-13,184	45.659	-490.659
269	Conway M.B.	(278 378)	12,206	5,565	-1,076	45.592	-41.315
270	Hebden Royd U.D.	(399 423)	8,674	3,938	-798	45.400	-33.371
271	Morecambe M.B.	(343 463)	41,908	18,848	-4,212	44.975	-208.514
272	Wardle U.D.	(391 418)	5,349	2,400	-549	44.868	-28.247
273	Bowden U.D.	(376 386)	4,891	2,194	-503	44.858	-25.975
274	Southport C.B.	(333 416)	84,574	37,921	-8,732	44.838	-454.155
275	Penmaenmaur U.D.	(273 376)	3,991	1,787	-417	44.776	- 22.162
276	Prestatyn U.D.	(306 383)	14,515	6,462	-1,591	44.519	- 92.077
277	Barmouth U.D.	(263 317)	2,106	929	-248	44.112	- 16.234
278	Windermere U.D.	(341 497)	8,065	3,557	-951	44.104	- 62.394
279	Lytham St.Annes M.B.	(334 429)	40,299	17,765	-4,769	44.083	-314.776
280	Colwyn Bay M.B.	(287 378)	25,564	11,162	-3,240	43.663	-239.461
281	Llandudno U.D.	(280 381)	19,077	8,295	-2,487	43.482	-192.343
282	Criccieth U.D.	(251 339)	1,505	646	-213	42.924	-18.737
283	Dolgellan U.D.	(274 317)	2,567	1,062	-443	41.371	-52.228
284	Grange U.D.	(341 478)	3,474	1,395	-684	40.155	-96.847

TABLE 4 (cont.)

X_s^2 : 28 HIGHEST VALUES

	Area and location		Population	Males	Excess (m-f)	Ratio (% males)	X_s^2 value
1	Shifnal R.D.	(377 307)	15,869	8,587	1,305	54.112	200.082
2	Wolverhampton C.B.	(392 301)	269,112	133,976	-1,160	49.784	177.768
3	Bolton C.B.	(370 410)	154,199	74,709	219	50.073	147.876
4	Warrington R.D.	(364 391)	50,420	25,684	948	50.940	120.191
5	Clitheroe R.D.	(368 440)	9,456	5,075	694	53.670	101.177
6	Walsall C.B.	(399 301)	184,738	91,605	-1,524	49.588	87.476
7	Stafford M.B.	(393 322)	55,001	27,652	303	50.275	69.413
8	Cannock U.D.	(399 310)	55,882	28,022	162	50.145	60.538
9	Wellington R.D.	(367 317)	30,297	15,360	423	50.698	58.606
10	Ellesmere Pt. M.B.	(340 376)	61,637	30.831	25	50.020	57.035
11	Kirkham U.D.	(343 432)	6,436	3,405	374	52.906	50.011
12	Cheadle R.D.	(398 345)	40,098	20,120	142	50.177	45.152
13	Sedbergh R.D.	(371 491)	3,544	1,912	280	53.950	42.150
14	Cannock R.D.	(391 311)	43,855	21,949	43	50.049	42.130
15	Millom R.D.	(315 492)	14,089	7,195	301	51.068	37.204
16	Stafford R.D.	(387 322)	23,000	11,606	212	50.461	35.407
17	Mkt Drayton R.D.	(366 332)	16,782	8,519	256	50.763	34.401
18	Clun & Bishop's Castle R.D.	(333 301)	8,869	4,577	285	51.607	34.272
19	Valley R.D.	(333 380)	15,048	7,647	246	50.817	32.354
20	Settle R.D.	(380 469)	13,992	7,115	238	50.850	30.949
21	Widnes M.B.	(351 386)	56,949	28,278	-393	49.655	30.414
22	Stoke C.B.	(389 347)	265,258	129,971	-5,316	48.998	26.336
23	Leek R.D.	(395 360)	13,247	6,717	187	50.706	25.806
24	Kirkby U.D.	(343 398)	59,918	29,661	-596	49.503	24.117
25	Wellington U.D.	(365 311)	17,163	8,638	113	50.329	22.992
26	Newtown & Llanidloes R.D.	(306 301)	8,223	4,203	183	51.113	22.473
27	Kidsgrove U.D.	(385 354)	22,194	11,096	-2	49.995	19.873
28	Stone R.D.	(386 334)	22,150	11,071	-8	49.982	19.476

TABLE 4 (cont.)

X_s^2 : 28 LOWEST VALUES

	Area and location		Population	Males	Excess (m-f)	Ratio (% males)	X_s^2 value
257	Poulton-le-Fylde U.D.	(335 439)	16,423	7,675	-1,073	46,733	-20.523
258	Penmaenmawr U.D.	(273 376)	3,991	1,787	-417	44.776	-22.162
259	St. Asaph R.D.	(305 375)	11,192	5,172	-848	46.212	-23.465
260	Wilmslow U.D.	(384 382)	29,040	13,669	-1,702	47.070	-23.789
261	Stockport C.B.	(390 390)	139,644	66,804	-6,036	47.839	-24.441
262	Bowden U.D.	(376 386)	4,891	2,194	-503	44.858	-25.975
263	Wirrall U.D.	(327 383)	26,885	12,604	-1,677	46.881	-28.207
264	Wardle U.D.	(391 418)	5,349	2,400	-549	44.868	-28.247
265	Abergele U.D.	(296 379)	12,315	5,669	-977	46.033	-30.000
266	Hebden Royd U.D.	(399 423)	8,674	3,938	-798	45.400	-33.371
267	Liverpool C.B.	(338 389)	610,113	293,550	-23,013	48.114	-36.386
268	Conway M.B.	(278 378)	12,206	5,565	-1,076	45.592	-41.315
269	Prestwich M.B.	(382 402)	32,911	15,378	-2,155	46.726	-41.465
270	Dolgellau U.D.	(274 317)	2,567	1,062	-443	41.371	-52.228
271	Windermere U.D.	(341 497)	8,065	3,557	-951	44.104	-62.394
272	Rhyl U.D.	(301 382)	21,821	9,970	-1,881	45.690	-68.985
273	Wallasey C.B.	(329 392)	97,215	45,824	-5,567	47.137	-72.331
274	Thornton Cleveleys U.D.	(333 443)	26,837	12,319	-2,199	45.903	-72.462
275	Crosby M.B.	(332 400)	57,497	26,756	-3,985	46.535	-88.919
276	Hoylake U.D.	(324 388)	32,277	14,802	-2,673	45.859	-90.112
277	Prestatyn U.D.	(306 383)	14,515	6,462	-1,591	44.519	-92.077
278	Grange U.D.	(341 478)	3,474	1,395	- 684	40.155	-96.847
279	Llandudno U.D.	(280 381)	19,077	8,295	-2,487	43.482	-192.343
280	Morecambe M.B.	(343 463)	41,908	18,848	-4,212	44.975	-208.514
281	Colwyn Bay M.B.	(287 378)	25,564	11,162	-3,240	43.663	-239.461
282	Lytham St. Annes M.B.	(334 429)	40,299	17,765	-4,769	44.083	-314.776
282	Southport C.B.	(333 416)	84,574	37,921	-8,732	44.838	-454.155
284	Blackpool C.B.	(330 436)	151,860	69,338	-13,184	45.659	-490.659

uneven within all types of data units (see also Table 5). A large proportion of the data units contains a very small proportion of the total population and a small proportion of the data units contains the great bulk of the area's population. However, skewness and kurtosis are seen to decline with the aggregation of squares, as observed by Clarke and Rhind (1975, p.11), and the distribution of population is less concentrated within a few data units. The proportion of the total population contained within that half of the areal units lying at the lower end of the population size ranking is only 1.1 per cent at the one-kilometre level, but rises to 1.4 per cent, 2.3 per cent and 13.6 per cent for two-kilometre squares, five-kilometre squares and LAA's respectively. At the same time, at the upper end of the population size ranking, 50 per cent of the total population is found in 4.6 per cent of the one-kilometre squares, a proportion which rises to 4.9 per cent, 5.7 per cent and 12.7 per cent for the two-kilometre, five-kilometre and LAA units. Thus ratio maps are least reliable at the finer resolution, since the relatively greater proportion of small population units is more likely to produce extreme ratios.

However, the worst anomalies occur because the one-kilometre data include a large number of very small (1 to 10) populations (Table 5); indeed there are 168 squares with only one person each. Such units inevitably produce extreme masculinity ratios of zero or 100 per cent. Similarly, in populations of two or three persons, if occurrences of zero, one, two and three males were equally likely, extreme masculinity proportions would be produced in 66.6 and 50 per cent of the data units concerned. Aggregation has the effect of reducing the number and proportion of such populations and, consequently, the proportion of the total population found in such units (Table 5). The improvement in the performance of ratios at higher levels of aggregation can to some extent be related to this effect. While more than 26 per cent of the one-kilometre square data units have populations of less than eleven persons, only 1.9 per cent of the five-kilometre squares and none of the LAA's fall into this category. An appreciation of the impact of small populations permits a more accurate interpretation of the aspatial summary statistics for ratios (Table 2). While the aggregate masculinity proportion for the grid-square data is 48.39 per cent, the mean proportion

TABLE 5 : NUMBERS AND POPULATIONS OF AREAL UNITS CLASSIFIED BY POPULATION SIZE FOR THE FOUR LEVELS OF AGGREGATION

1KM X 1KM GRID SQUARES								
Population size	Areal units				Population			
	No.	Cum	%	Cum %	No.	Cum.	%	Cum. %
1	168	168	1.01	1.01	168	168	0.002	0.002
2	439	607	2.64	3.65	878	1,046	0.010	0.012
3	448	1,055	2.70	6.35	1,344	2,390	0.016	0.028
4	615	1,670	3.70	10.05	2,460	4,850	0.029	0.057
5	550	2,220	3.31	13.36	2,750	7,600	0.032	0.089
6	463	2,683	2.79	16.15	2,778	10,378	0.033	0.122
7	450	3,133	2.71	18.86	3,150	13,528	0.037	0.159
8	449	3,582	2.70	21.56	3,592	17,120	0.042	0.201
9	414	3,996	2.49	24.05	3,726	20,846	0.044	0.245
10	365	4,361	2.20	26.25	3,650	24,496	0.043	0.288
11- 100	7,351	11,712	44.25	70.50	246,952	271,448	2.906	3.194
101- 1,000	2,737	14,449	16.48	86.98	987,390	1,258,838	11.620	14.814
1,001- 10,000	2,139	16,588	12.88	99.86	6,944,106	8,202,944	81.717	96.531
10,001-100,000	24	16,612	0.14	100.00	294,746	8,497,690	3.469	100.000

2KM X 2KM GRID SQUARES								
1	21	21	0.43	0.43	21	21	0.0002	
2	42	63	0.85	1.28	84	105	0.001	0.001
3	52	115	1.05	2.33	156	261	0.002	0.003
4	46	161	0.93	3.26	184	445	0.002	0.005
5	55	216	1.11	4.37	275	720	0.003	0.008
6	59	275	1.19	5.56	354	1,074	0.004	0.012
7	44	319	0.89	6.45	308	1,382	0.004	0.016
8	28	347	0.57	7.02	224	1,606	0.003	0.019
9	30	377	0.61	7.63	270	1,876	0.003	0.022
10	24	401	0.49	8.12	240	2,116	0.003	0.025
11- 100	1,798	2,199	36.40	44.52	86,437	88,553	1.017	1.042
101- 1,000	1,610	3,809	32.59	77.11	500,295	588,848	5.888	6.930
1,001- 10,000	855	4,664	17.31	94.42	3,293,602	3,882,450	38.758	45.688
10,001-100,000	276	4,940	5.59	100.01	4,615,240	8,497,690	54.312	100.000

TABLE 5 (cont.)

5K X 5KM GRID SQUARES								
Population size	Areal units				Population			
	No.	Cum.	%	Cum %	No.	Cum.	%	Cum %
1	0	0	0.00	0.00	0	0	0.00000	
2	2	2	0.22	0.22	4	4	0.00005	
3	2	4	0.22	0.44	6	10	0.00007	
4	2	6	0.22	0.66	8	18	0.00009	
5	2	8	0.22	0.88	10	28	0.00012	
6	2	10	0.22	1.10	12	40	0.00014	
7	2	12	0.22	1.32	14	54	0.00016	
8	1	13	0.11	1.43	8	62	0.00009	
9	1	14	0.11	1.54	9	71	0.00011	
10	3	17	0.33	1.87	30	101	0.00035	
11- 100	87	104	9.60	11.47	3,964	4,065	0.047	0.048
101- 1,000	317	421	34.99	46.46	152,696	156,761	1.797	1.845
1,001- 10,000	296	717	32.67	79.13	1,003,000	1,159,761	11.798	13.643
10,001- 100,000	179	896	19.76	98.89	6,021,709	7,181,470	70.868	84.511
over- 100,000	10	906	1.10	99.99	1,316,220	8,497,690	15.489	100.000

LOCAL AUTHORITY AREAS								
0- 10	0	0	0.00	0.00	0	0	0.000	0.000
11- 100	0	0	0.00	0.00	0	0	0.000	0.000
101- 1,000	1	1	0.35	0.35	729	729	0.008	0.008
1,001- 10,000	83	84	29.23	29.58	447,534	448,263	4.947	4.955
10,001- 100,000	187	271	65.84	95.42	5,704,092	6,152,355	63.051	68.006
over- 100,000	13	284	4.58	100.00	2,894,471	9,046,826	31.994	100.000

is 50.77 and 50.51 per cent for the one- and two-kilometre square data respectively. This is not, as one might initially infer, the product of a greater number of data units with more than 50 per cent males. Although the one-kilometre data do show a slightly higher number of such units : 8,574 as against 8,038 with less than 50 per cent males, the situation is reversed at the two-kilometre level, where 2,228 squares have more and 2,712 less than 50 per cent masculinity. The relatively high mean values for percentage males in these cases are due to the fact that the numbers of data units with very high masculinity proportions are substantially greater than those with very low masculinity (Table 6). In the one-kilometre data, 162 squares register 100 per cent males, while only 67 units register zero; in the two-kilometre data the corresponding numbers are 23 and 10 squares. Thus ratio values give the misleading impression that very high masculinity is more common than very low at the one-kilometre level. At that level X_s^2 maps indicate the converse, even though an expectation of 48.53 per cent males was assumed. At higher levels of aggregation, X_s^2 maps also indicate that very high masculinity is more frequent than very low (see below, section 3.4).

The high frequency of very small populations is likely to be an even more serious problem in ratio-based studies using data for 100-metre squares and especially so in the case of the 10 per cent sample data. Although with the 100 per cent Population and 100 per cent Household Small Area Statistics this is not a serious problem when using unsuppressed variables (since the base population is never less than 25 persons), it may become so when the denominator of a derived variable (see Rhind, Evans and Dewdney, 1977) comprises only a subset of the total population, for example, when unemployed males are expressed as a proportion of economically active males.

The X_s^2 measure places hardly any significance on extreme variations within populations of less than four persons and generally yields low values for small populations compared with those derived from larger populations with similar masculinity proportions (Visvalingam, 1976). Despite its apparent advantages, however, caution is necessary in applying this method to large areal units such as LAA's. Maps of such units on their own are somewhat misleading,

TABLE 6 NUMBER AND PERCENTAGE OF DATA UNITS REGISTERING VERY HIGH OR VERY LOW MASCULINITY

Data Units	Measure	Very high masculinity		Very low masculinity	
		Number	%	Number	%
1km x 1km	X_s^2	450	2.71	554	3.33
	% males	606	3.65	349	2.10
2km x 2km	X_s^2	295	5.97	272	5.51
	% males	355	7.19	200	4.05

TABLE 7 CLASS BOUNDARIES DEMARCATING 10% OF EXTREME CASES IN EACH TAIL OF THE DISTRIBUTION AT DIFFERENT LEVELS OF AGGREGATION

Measure	Areal unit	Lower bound	Upper bound
X_s^2	1km x 1km	- 1.105	1.551
	2km x 2km	- 1.458	2.508
	5km x 5km	- 4.383	5.946
	LAA	-20.523	18.820
Ratio % male	1km x 1km	40.000	62.500
	2km x 2km	44.531	58.333
	5km x 5km	46.317	53.649
	LAA	46.482	49.995

especially in their portrayal of areas of high masculinity. For example, from Table 3 it is apparent that the bulk of the males responsible for the high masculinity of Shifnal R.D. (377 307) are located in the single one-kilometre square 379 305. This square has a negligible visual impact on the one-kilometre square maps when compared with the spatial importance it assumes in the maps based on LAA's.

Thus LAA data can give a very misleading impression of the spatial extent of areas of high masculinity and other demographic phenomena. The use of data for administrative units present a further problem. Most mapping and statistical exercises, like those employed in this paper, involve the ordering and scaling of data units. While crude measures of population density take account of variations in the areal size of the units, bivariate measures, such as masculinity proportions, make no allowance for the size of the base population or the area under consideration. While the X_s^2 measure does consider population size, it takes no account of variations in density. Since the grid squares are of constant size, density may legitimately be ignored in grid-based analysis and mapping and the X_s^2 measure is applicable without further modification.

The X_s^2 measure does, however, require further modification for valid comparisons of irregular areal units which vary in size, such as LAA's. Consider the case of two administrative units, A and B, with identical population densities and masculinity proportions. If A is twice the size of B, X^2 values for A will be greater than those for B, since the resultant population is larger. In terms of its spatial representation, this bias in scaling and ranking is further exaggerated, as far as its visual impact is concerned, by the demarcation of a larger area in a denser hue. A is not only bigger than B but more densely shaded or coloured as well. An extension of the current discussion to the topic of area-based analysis will be presented in a later paper.

It appears from the current discussion that, while aggregation improves the performance of ratio measures which are unreliable at the one-kilometre level, the more robust X_s^2 technique does not derive any substantial benefits from aggregation which would compensate for the loss of resolution; and the use of data for irregular administrative units presents problems of spatial interpretation. The difficulties associated with the occurrence of small populations in the one-kilometre data are best solved by local aggregation within a maximum specified search radius. If the population within the search radius is still below a defined threshold, the corresponding values should be ignored as having a negligible effect.

3.4. The identification of extremes

One of the problems associated with a comparative study involving two different measures, which has been touched on in earlier sections of this paper, is the choice of a compatible system for defining the extremes in each distribution. In the X_s^2 - based maps accompanying this paper, cut-off values of $X_s^2 = \pm 3.84$ (the 95 per cent significance level for this data set) have been consistently applied in delimiting the extreme cases. In the case of ratios, cut-offs for the one- and two-kilometre square data were selected on the basis of standard scores, being $\pm$ 2 standard deviations for one-kilometre and $\pm$ 1.25 standard deviations for two-kilometre data units. Table 6 indicates the number and percentage of data units in each extreme class defined by this method.

As an alternative method of comparing the two measures at different levels of aggregation, the 10 per cent of cases at the tails of each distribution were identified and mapped. The resultant maps (Figs.23-36), in addition to confirming the view that the performance of ratios varies with the size of the areal unit used, illustrates the inadequacy of a quantile system for delimiting class boundaries. While 10 per cent of the LAA units picked out only the most extreme levels of masculinity, 10 per cent of the one- and two-kilometre squares included several very low X_s^2 values (Table 7), these being the result of chance variations. On the other hand, even at the five-kilometre level, some squares which consistently register high masculinity in the value cut-off system (Fig. 12) disappear in the corresponding decile map (Fig.27). The omission of high masculinity in Bardsey Island (210 321) and the Lleyn peninsula (230 340) is particularly striking. In the LAA map (Fig. 29), the omission of low masculinity in the smaller holiday resorts and market towns of North Wales and of high masculinity even in Barrow C.B. (322 471) is a serious deficiency. The use of a quantile system is likely to give a misleading impression of variations in the nature and spatial patterns of sex composition at different scales.

The use of a constant cut-off for the delimitation of extreme X_s^2 values is more incisive. Variations in the size of the X_s^2 extreme classes using the $\pm$ 3.84 cut-off value and changes in the proportions of units with high and low masculinity (Table 6) are readily explicable and portray a more consistent picture of the pattern of sex composition than does the quantile system. At the one-kilometre square level there are more units with low masculinity than with high, although the extreme cases constitute only six per cent of the 16,612 data units.From Figures 8 & 9 it can be seen that, while many of the low-masculinity units form contiguous areas, units with high masculinity are a good deal more dispersed. Aggregation of areas of low masculinity tends to produce fewer such areas, although the resulting concentration of females is reflected in more extreme X_s^2 values, as Table 3 shows. Because areas of high masculiniity are so markedly male at the one-kilometre level, their numbers and high-masculinity status are retained through higher levels of aggregation, despite averaging and the resultant reduced values of X_s^2. Thus the spatial patterns and extent of areas of high and low masculinity are maintained by an automatic adjustment in the proportion of extreme cases, and some indication of concentration or averaging can be obtained from the actual X_s^2 values.

4. CONCLUSIONS

4.1. Aggregation and the spatial patterns of sex and age composition

As already indicated (above section 3.1), the one-kilometre X_s^2 maps (Figs. 8,9,23 and 24) show that females are spatially concentrated, i.e. that areas of low masculinity are contiguous and occur mainly along the coast, while areas of high masculinity are discontinuous and much more dispersed. As a result, aggregation has the effect of combining fairly similar areas in zones of low masculinity but incorporates very dissimilar categories within aggregate units of high masculinity. Thus the LAA's along the coast may legitimately be assumed to contain a high proportion of females throughout most of their areal extent (though even here small areas of high masculinity may occur), whereas LAA's registering high masculinity may well have their male excess concentrated within a few one-kilometre squares. Similarly, caution should be exercised in interpreting the spatial extent of high masculinity even in the case of maps based on one-kilometre squares, since this feature may only apply to a small number of 100-metre squares.

While the example used in this paper has been that of sex composition, passing reference has been made (above, p. 4) to one aspect of age composition, namely the proportion aged 65 years and over. A brief note on the relationships between these two parameters would seem appropriate in view of the (generally correct) assumption that they are closely related, low masculinity being associated with a high proportion of elderly people and vice versa. Comparison of the age structure maps (Figs.2-7) with those of sex composition reveals a pattern of areas with above average 'agedness' very similar to that of low masculinity. The distribution of below-average agedness, on the other hand, is generally in areas of moderate masculinity - though with important exceptions (e.g. Walsall C.B. : 399 301 and Wolverhampton C.B. 392 301) where low agedness and high masculinity coincide - and exhibits a tendency towards spatial clustering. Relationships of this kind between two or more variables are now being investigated.

4.2. The measures of sex composition

The problems associated with the use of ratios are not peculiar to the one-kilometre square data, but are also present at higher levels of aggregation. Although the ratio and X_s^2 maps for LAA's show a greater similarity, the rankings of these units by their respective values (Table 4) differ and this is likely to manifest itself in maps which display only those data units falling into the extreme classes. This is to some extent already apparent in the maps (Figs.29 and 36) showing the ten per cent of LAA's at each extreme. The discrepancy is likely to be even greater when the data used refer to smaller areal and population units such as Wards, Civil Parishes or Enumeration Districts. This implies that the results produced by two research workers using different sized areal units are not comparable if ratio measures are employed but may be related to each other via the X_s^2 measure. On the other hand, X_s^2 requires further consideration before it can be applied to data for administrative units which vary in spatial extent. Since the latter are by no means equal-population units, grid-square data become considerably more attractive, in that analytical methods are available for coping with variations in population size. Rhind (1975) lists the advantages and disadvantages of using grid-square based data, and one-kilometre grid-square maps of great Britain (Rhind, Visvalingam, Perry and Evans, 1976) certainly display a degree of detail and quality not previously exhibited in population mapping.

4.3. The delimitation of class boundaries

The systems adopted in this paper for the delimitation of class boundaries were, like the majority of such systems, geared towards specific objectives. The initial aim was to compare and evaluate the ratio and X_s^2 ordering schemes at each level of aggregation and to note the changes in each ordering system at the various levels of aggregation. The more reliable measure could then be used to study the effects of aggregation.

The frequency distributions of ratio and X_s^2 values for masculinity are both fairly symmetrical and the quantile system thus appeared to offer the most objective basis for identifying disparities between the ratio and X_s^2 ordering schemes. The decile system was an arbitrary choice. Maps showing the ten per cent of extreme cases confirmed the indications of maps based on cut-off values, namely that there is a disparity between ratio and X_s^2 ordering schemes, especially at finer levels of aggregation. While the spatial distribution of extreme X_s^2 values remained consistent, the pattern of extreme ratio values was more variable.

However, the use of the decile system has the unfortunate disadvantage of including, in the extreme classes, many near-average X_s^2 values at the one-kilometre level, while including only the very extreme anomalies in the LAA data. Thus the spatial extent of extremes of high and low masculinity varied. Given that all the data refer to the same variable, area and census date, decile maps give the misleading impression that the spatial extent of high and low masculinity contracts with aggregation. The availability of corresponding cut-off values does not help towards effecting a better comparison of the spatial extent of high and low levels of masculinity at different levels of aggregation. It can readily be appreciated that decile maps of the same variable for different study areas, or at different census dates, would be difficult to compare if the level of aggregation varied.

The use of a constant cut-off value is more incisive in studying the effects of aggregation. The choice of $\pm$ 3.84 is admittedly arbitrary. However, at the associated level of confidence, the spatial distribution and extent of extreme cases is suggestive of the effects of aggregation. In broad terms, the spatial extent of areas of low masculinity is remarkably consistent. Areas of high masculinity, on the other hand, become more extensive with aggregation. The implications of aggregation were thus more readily evaluated using the maps of X_s^2 values outside $\pm$ 3.84.

4.4. General implications

The reader is reminded that the prime objective of this paper is not that of providing a systematic, detailed description and explanation of the spatial patterns of sex composition within the study area, though a good deal of detail on this aspect has emerged during the discussion. The main purpose of the exercise, as the paper's title indicates, has been to examine the efficiency of ratio and chi-square measures in the analysis of data at various levels of aggregation. Although the emphasis throughout has been on spatial variations, and thus on mapping, and a specific population parameter (sex composition) has been used as an example, the points made in the discussion and the conclusions reached have much wider implications.

A perusal of the geographical and demographic literature illustrates the widely-held view that ratios are standard tools for geographical and demographic analysis. Ratios with variable denominators occur not only in population geography but also in many other branches of the subject. Since reservations already exist concerning the use of ratio measures, an analysis of the problems involved and the selection of one possible alternative measure are likely to be of general concern to all who use data in a ratio form. Demographic data are not the only type involving wide variations in sample size and the problems discussed here are by no means specific to such data.

Furthermore, the use of ratios is not confined to mapping. Multivariate excercises commonly include ratio variables and there has been much discussion concerning the choice of suitable transformations and scaling procedures (see Evans, Catterall and Rhind, 1975). The present authors are of the opinion that problems of ordering are even more crucial and fundamental than those of scaling. Current research on this front indicates that, while ratio measures seem to over-simplify relationships between different social and demographic parameters, the X_s^2 measure appears to result in a more vivid appreciation of the non-linear and compound nature of such relationships, even in the case of bivariate data. Although maps have been used in this study to illustrate the disparities between ratio and X_s^2 ordering systems, it must be emphasized that the implications of this work extend beyond cartographic concerns.

REFERENCES

CLARKE, J.I. and RHIND, D.W. (1975) "The relationship between the size of areal units and the characteristics of population structure"; Working Paper No.5, Census Research Unit, Department of Geography, University of Durham.

EVANS, I.S., CATTERALL, J.W. and RHIND, D.W. (1975) "Specific transformations are necessary"; Working Paper No.2, Census Research Unit, Department of Geography, University of Durham.

HUNT, A.J., Ed. (1968) "Population maps of the British Isles, 1961"; Transactions of the Institute of British Geographers, No. 43.

RHIND, D.W. (1975) "Mapping the 1971 Census by computer", Population Trends, 2, 9-12.

RHIND, D.W., EVANS, I.S. and DEWDNEY, J.C. (1977) "The derivation of new variables from population census data"; Working Paper No.9, Census Research Unit, Department of Geography, University of Durham.

RHIND, D.W., VISVALINGAM, M, PERRY, B.J. and EVANS, I.S. (1976) "People mapped by laser beam", Geographical Magazine, XLIX (3) 148-152

VISVALINGAM, M., (1975) "Storage of the 1971 U.K. Census data; some technical considerations"; Working Paper No.4, Census Research Unit, Department of Geography, University of Durham.

VISVALINGAM, M., (1976) "Chi-square as an alternative to ratios for statistical mapping and analysis"; Working Paper No.8, Census Research Unit, Department of Geography, University of Durham.

VISVALINGAM, M. and PERRY, B.J. (1976) "Storage of the grid-square based 1971 G.B. Census data; checking procedures", Working Paper No.7, Census Research Unit, Department of Geography, University of Durham.

ACKNOWLEDGEMENTS

The production of this paper would have been impossible without the skill and attention to detail of the clerical and technical staff of the Department of Geography, University of Durham. The authors wish to express their sincere thanks to Joan Dresser, who typed the manuscript; Jerry Donnini and Paul Wilson, who drew the maps; Derek Hudspeth and Ian Middlemass, who did the photographic work; John Normile and Dennis Ewbank, who were responsible for the printing.

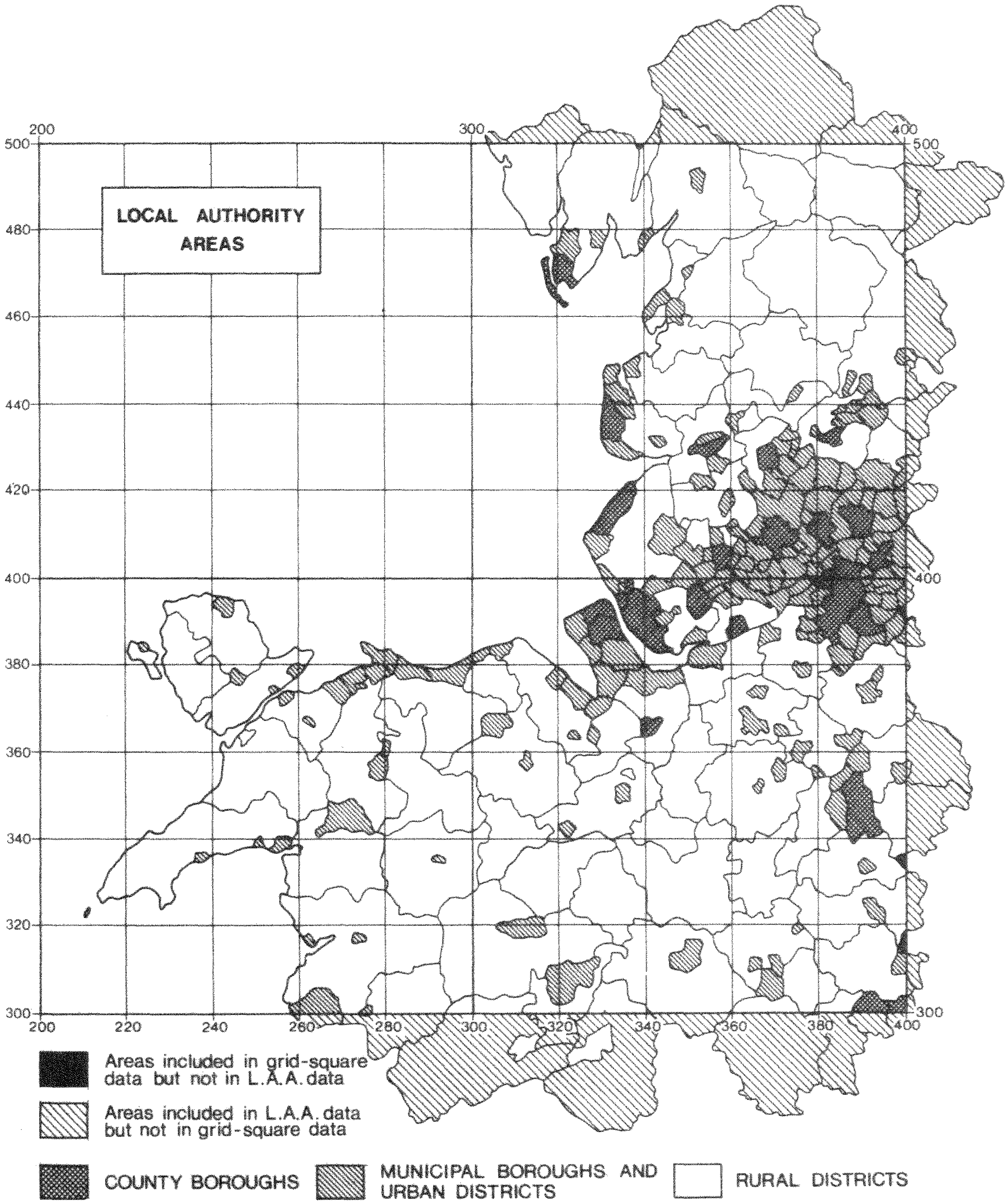


Fig. 1. Base map showing grid squares and Local Authority Areas

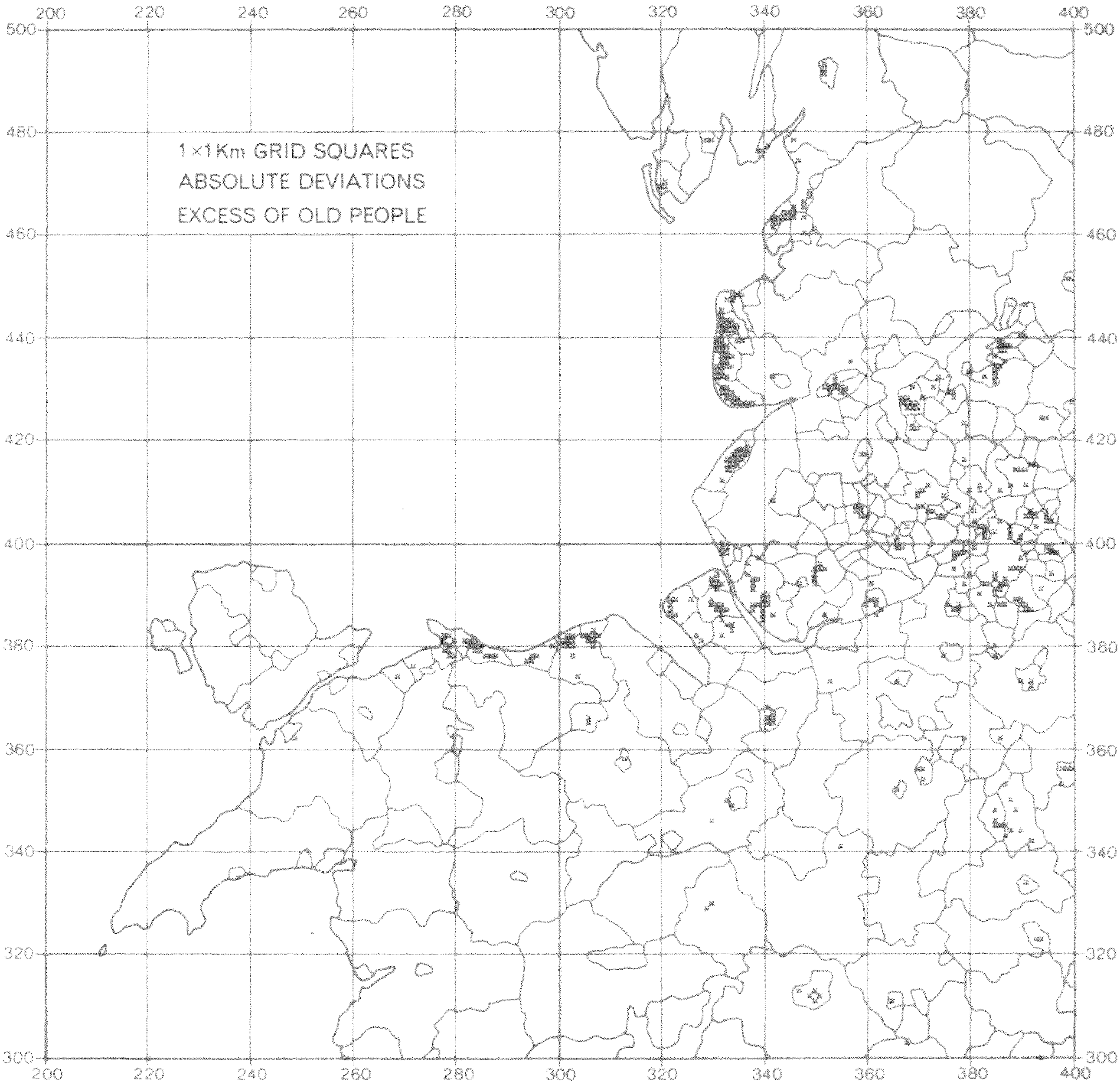


Fig. 2. Absolute deviation map at the 1 x 1 km level, showing squares with an excess of old people

* squares in which observed minus expected is ≥ 113, when observed = number of old people; expected = total population x 13.19% (i.e. the regional average proportion of old people).

No. of units = 440; proportion of total units = 5.0%

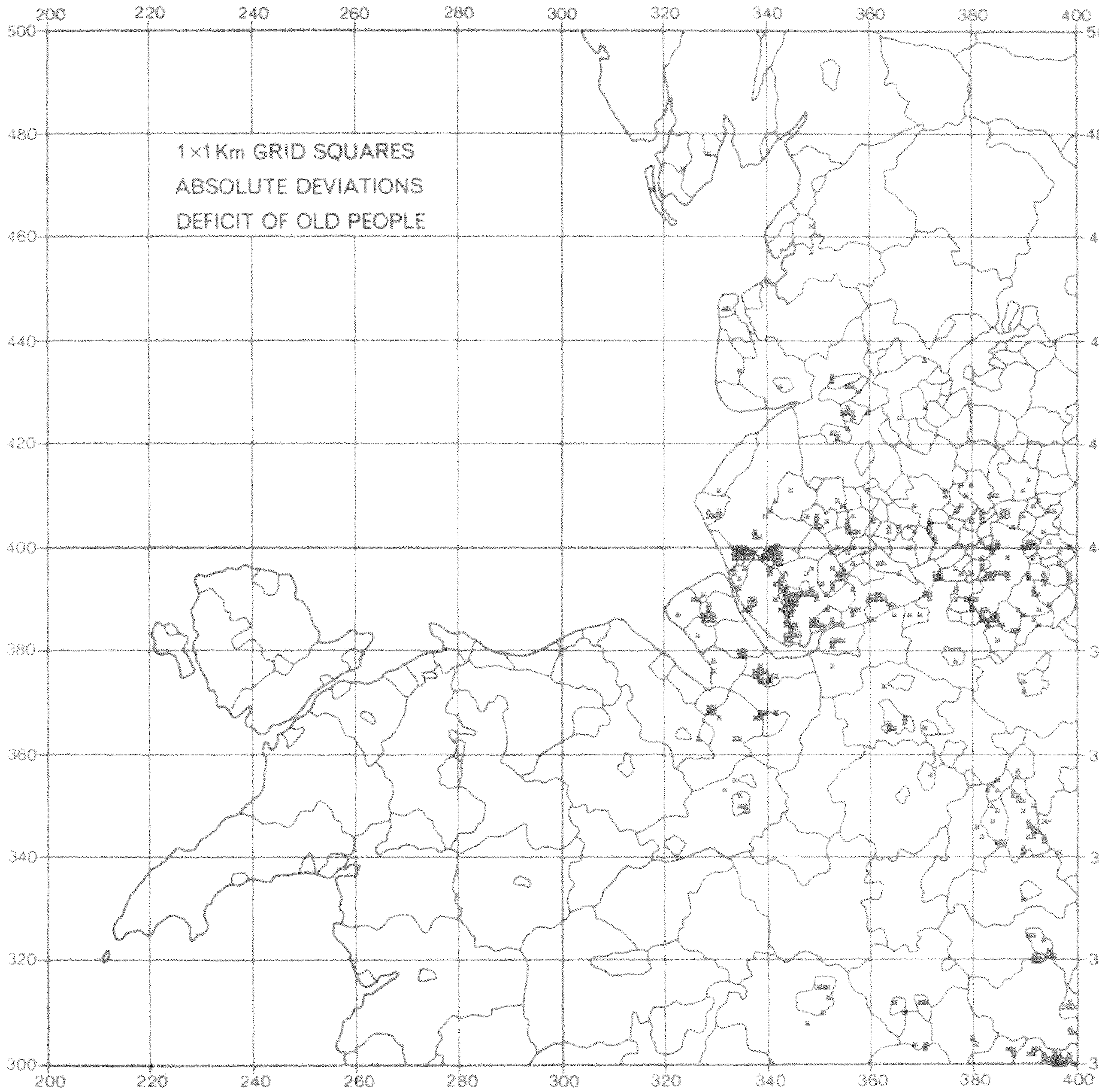


Fig. 3. Absolute deviation map at the 1 x 1 km level, showing squares with a deficit of old people

* squares in which observed minus expected ≤ -119, when observed = number of old people; expected = total population x 13.19% (i.e. the regional average proportion of old people).
No. of units = 440; proportion of total units = 5.0%

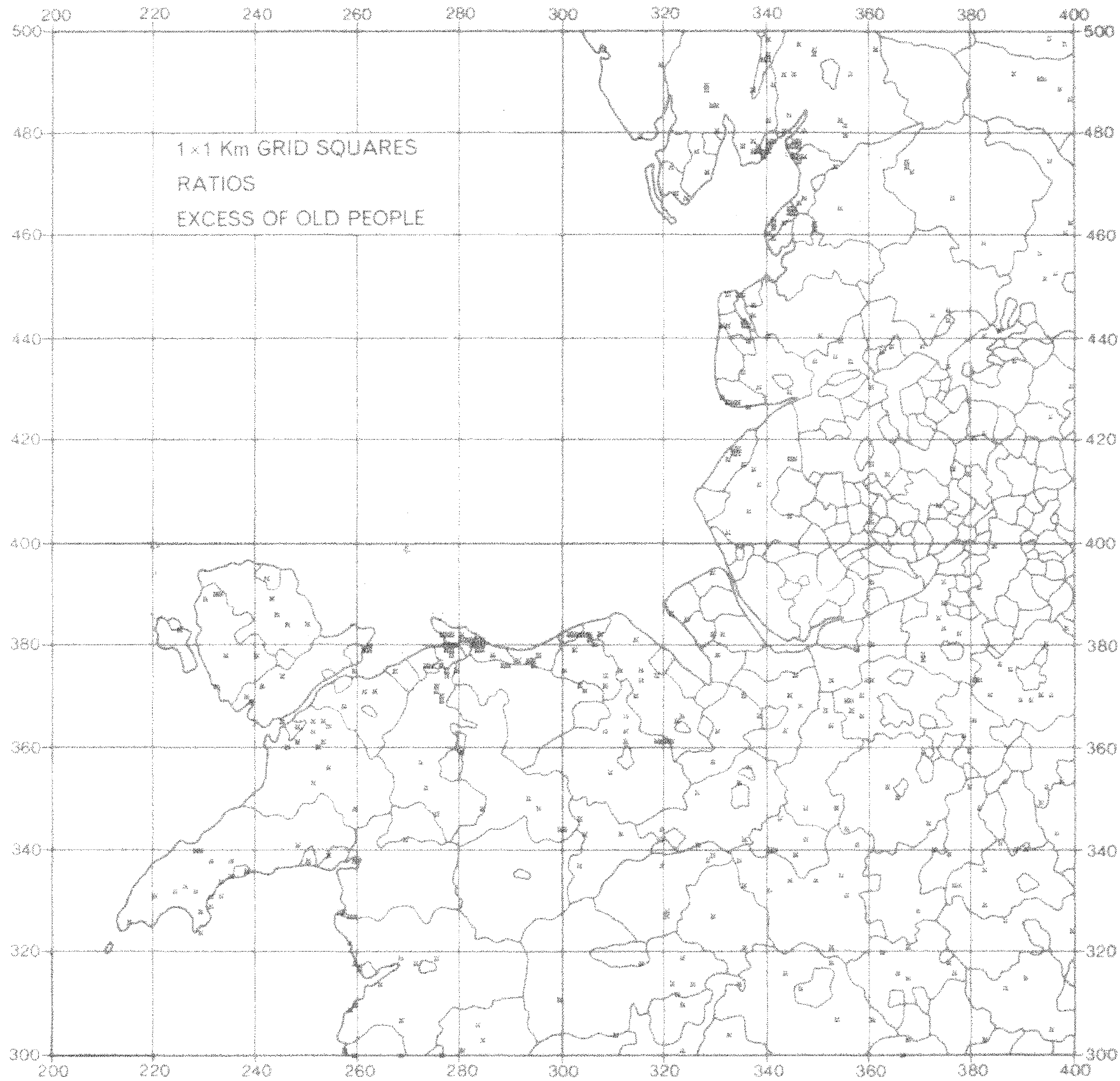


Fig.4. Ratio map at the 1 x 1 km level, showing squares with an excess of old people

* squares with over 28.26% old people
No. of units = 440; proportion of total units = 5.0%

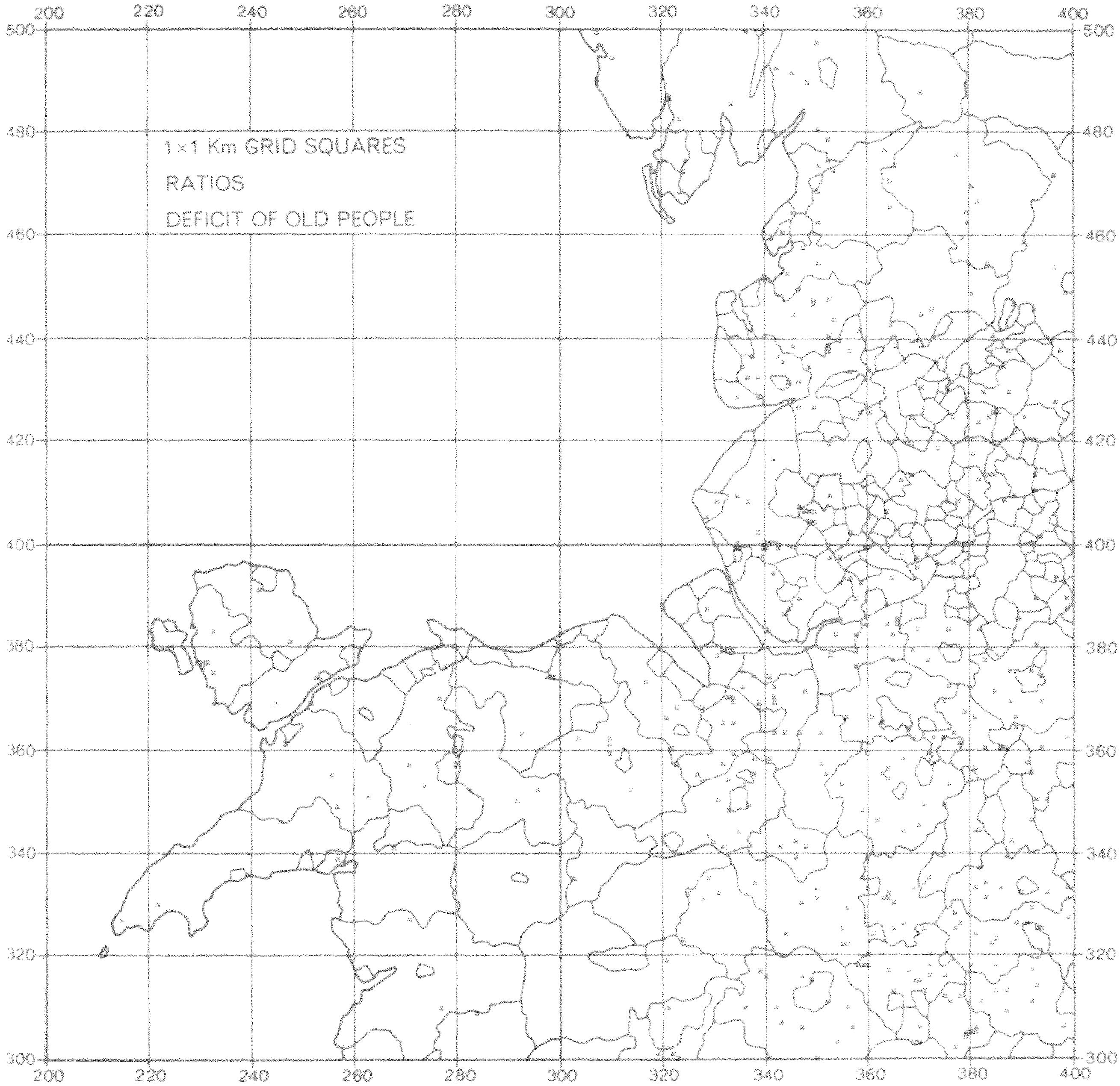


Fig. 5. Ratio map at the 1 x 1 km level, showing squares with a deficit of old people

* squares with less than 2.78% old people
No. of units = 440; proportion of total units = 5.0%

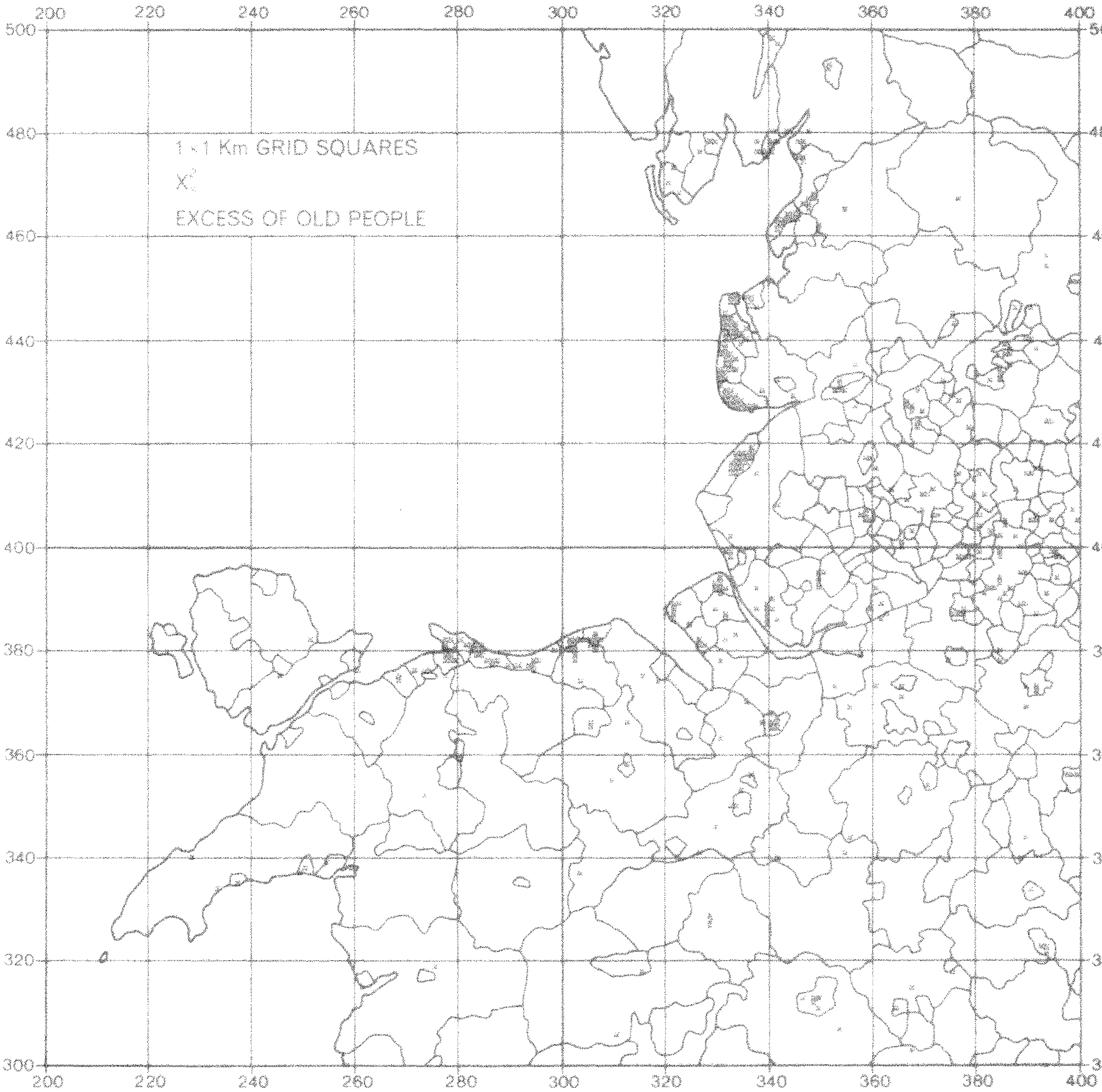


Fig. 6. X_s^2 map at the 1 x 1 km level, showing squares with an excess of old people

* squares with $X_s^2 \geqslant 60.46$
No. of units = 440; proportion of total units = 5.0%

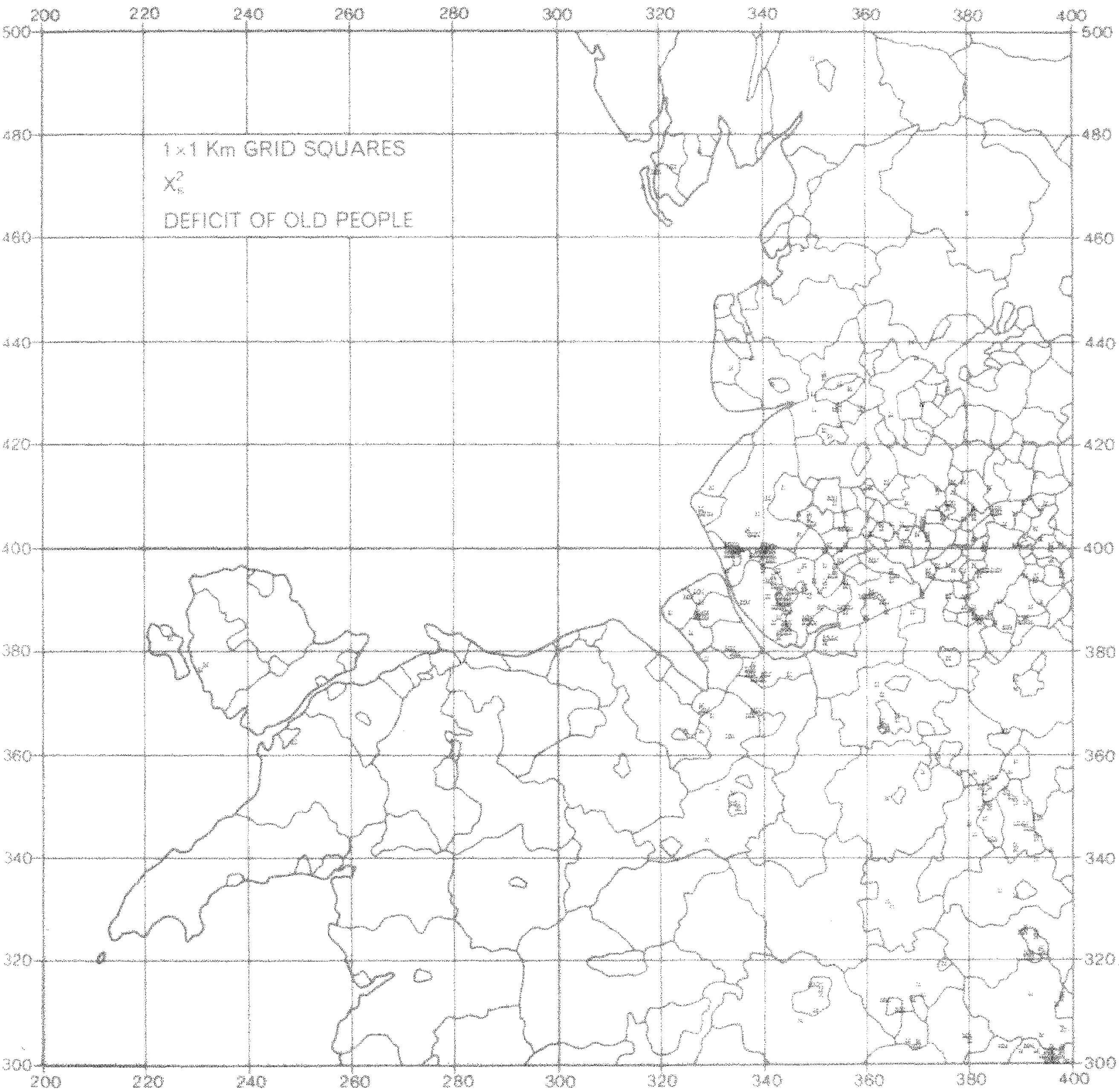


Fig. 7. X_s^2 map at the 1 x 1 km level, showing squares with a deficit of old people

* squares with $X_s^2 \leqslant -57.28$
No. of units = 440; proportion of total units = 5.0%

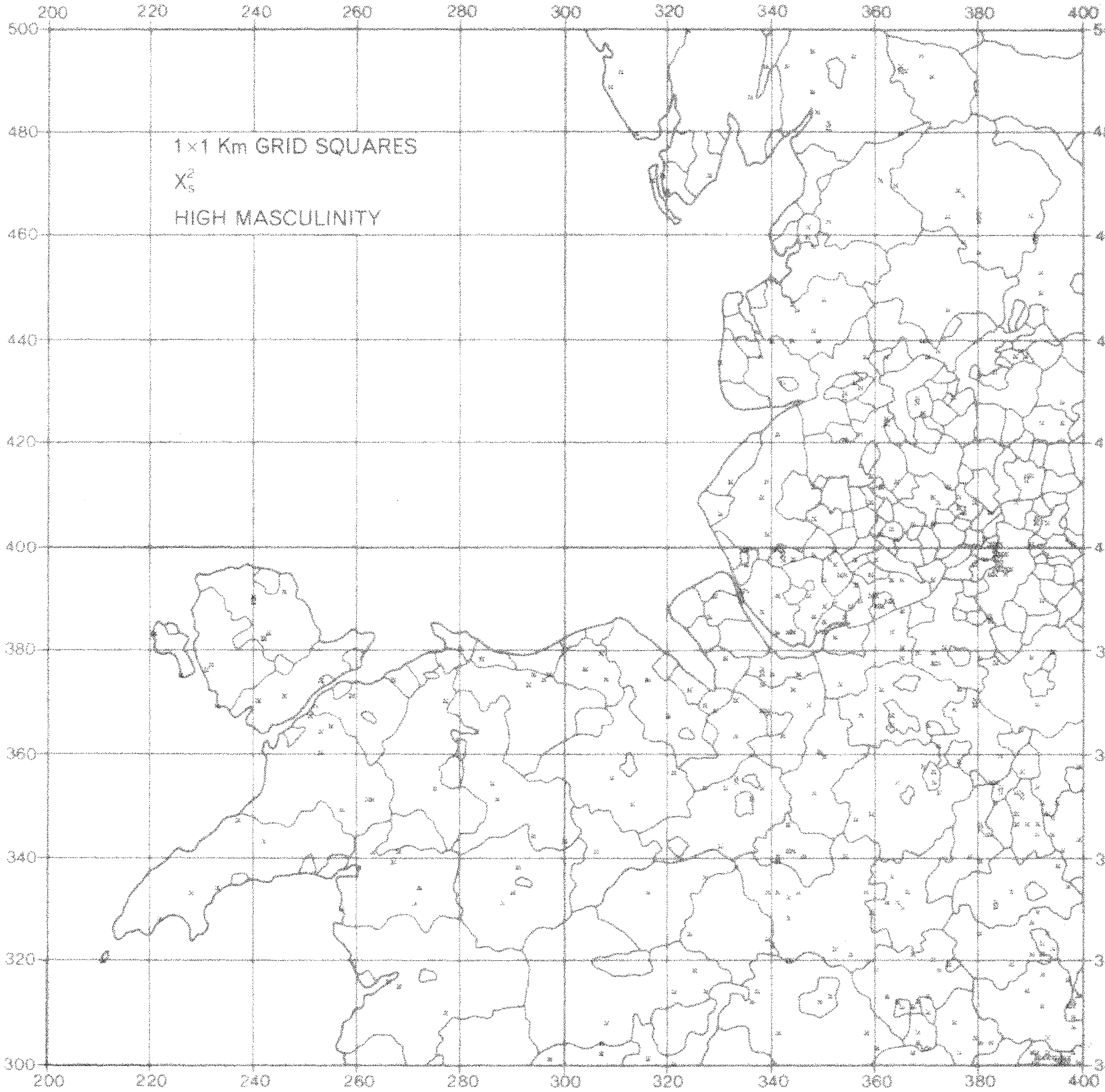


Fig. 8. X_s^2 map at the 1 x 1 km level, showing squares with high masculinity

* squares with $X_s^2 \geqslant 3.84$
No. of units = 450; proportion of total units = 2.71%

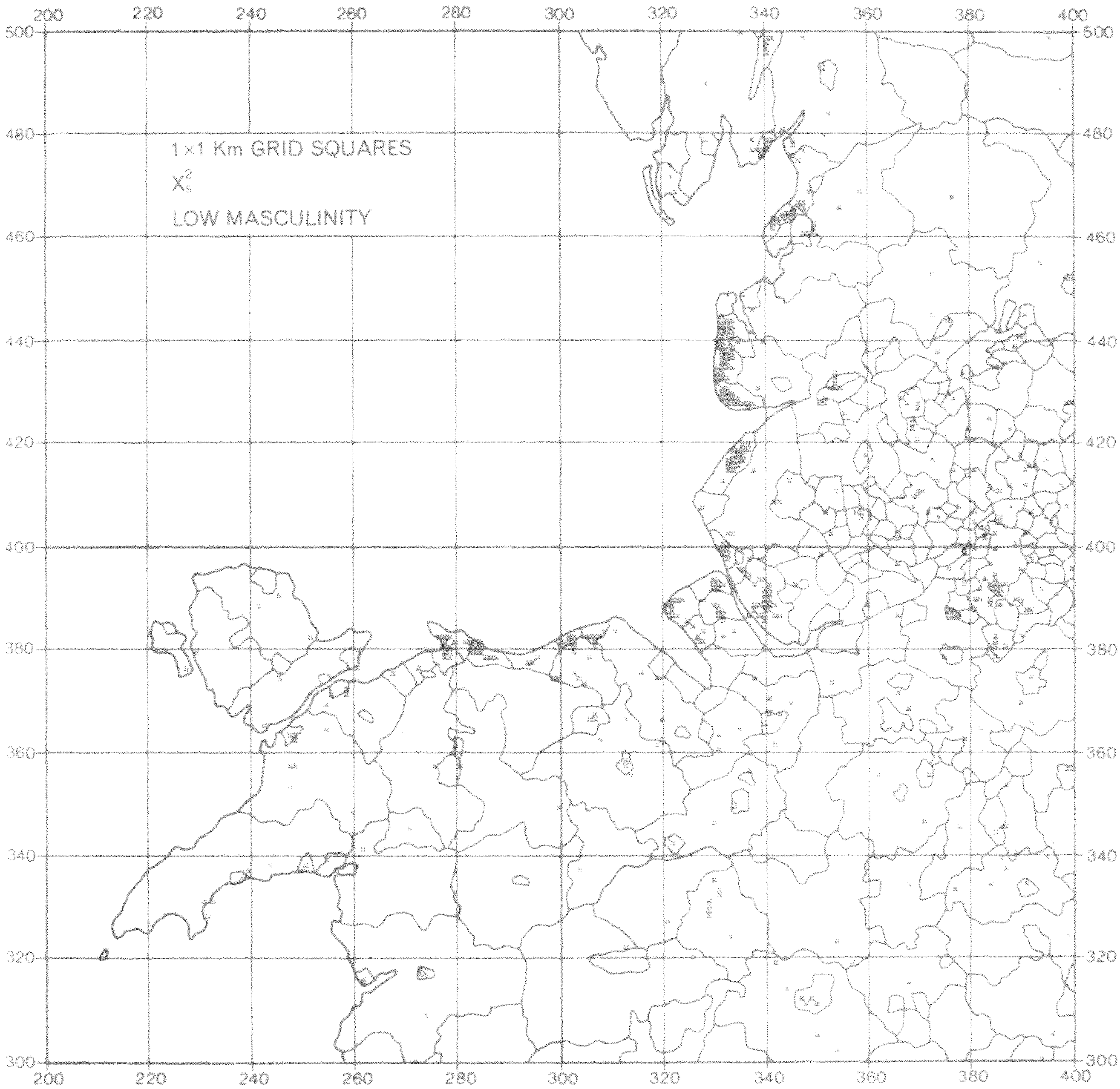


Fig. 9. X_s^2 map at the 1 x 1 km level, showing squares with low masculinity

* squares with $X_s^2 \leqslant -3.84$
No. of units = 554; proportion of total units = 3.33%

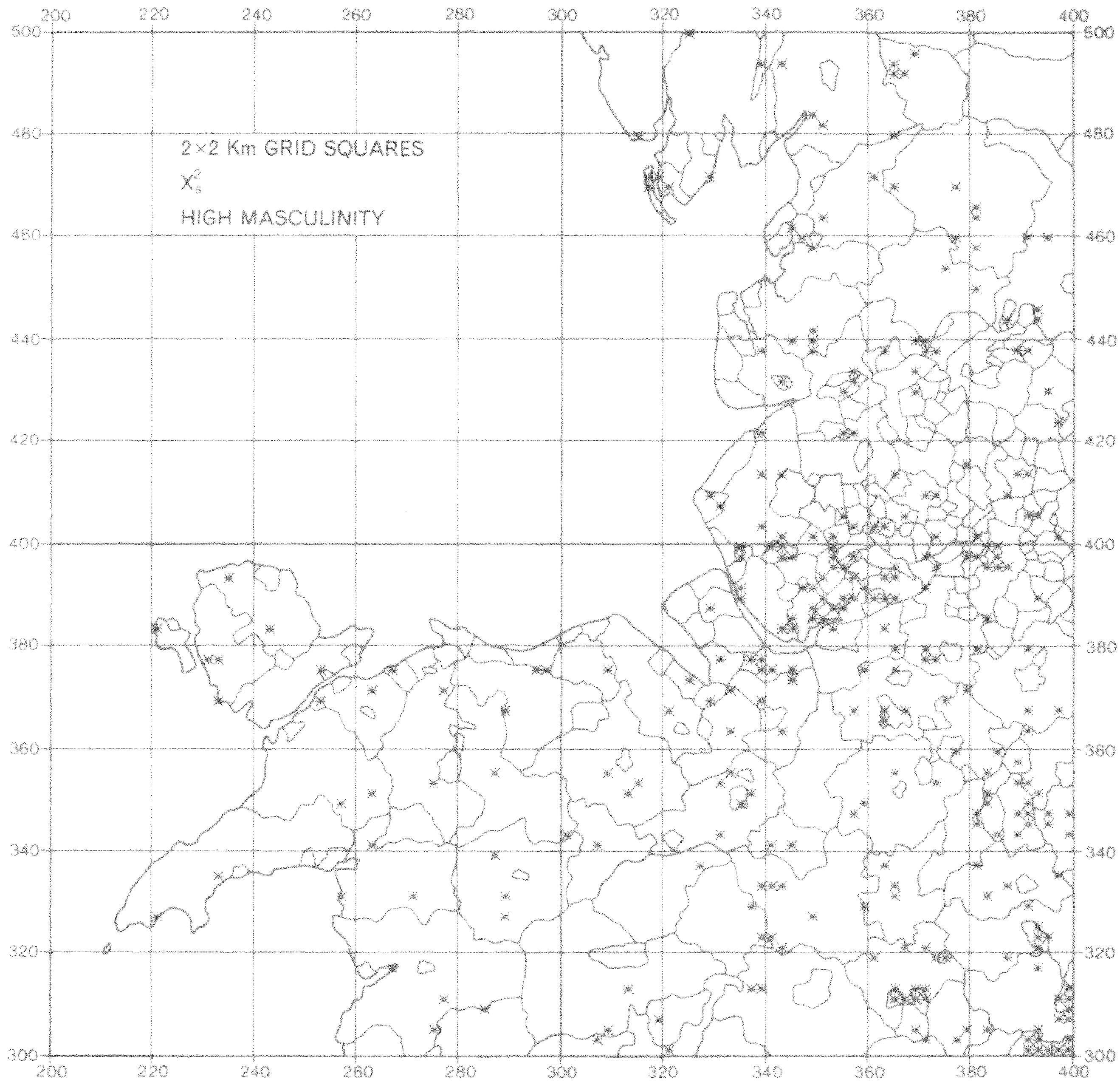


Fig. 10. X_s^2 map at the 2 x 2 km level, showing squares with high masculinity

* squares with $X_s^2 \geqslant 3.84$
No. of units =295, proportion of total units = 5.97%

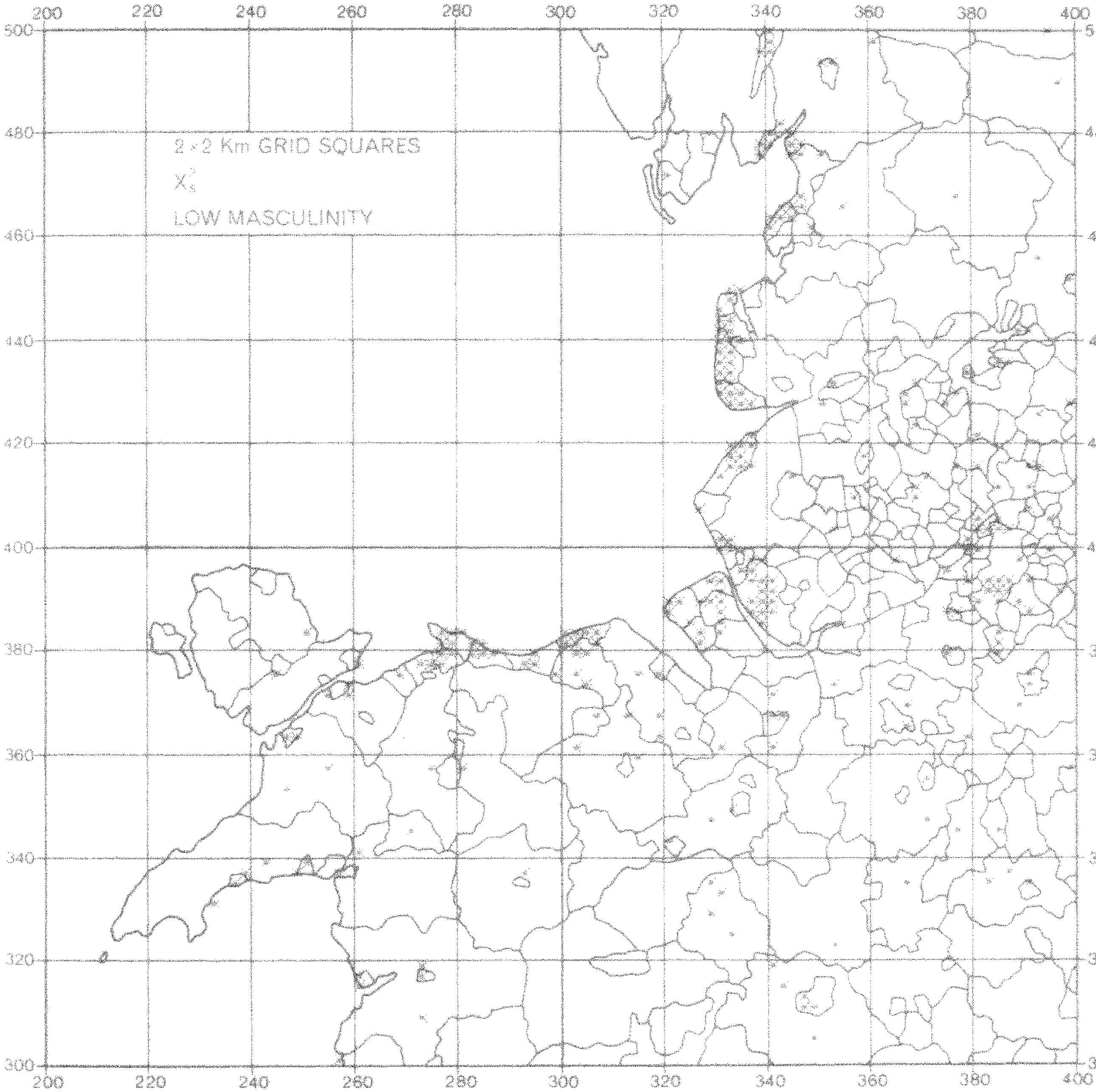


Fig. 11. X_s^2 map at the 2 x 2 km level, showing squares with low masculinity

* squares with $X_s^2 \leqslant -3.84$
No. of units = 272; proportion of total units = 5.51%

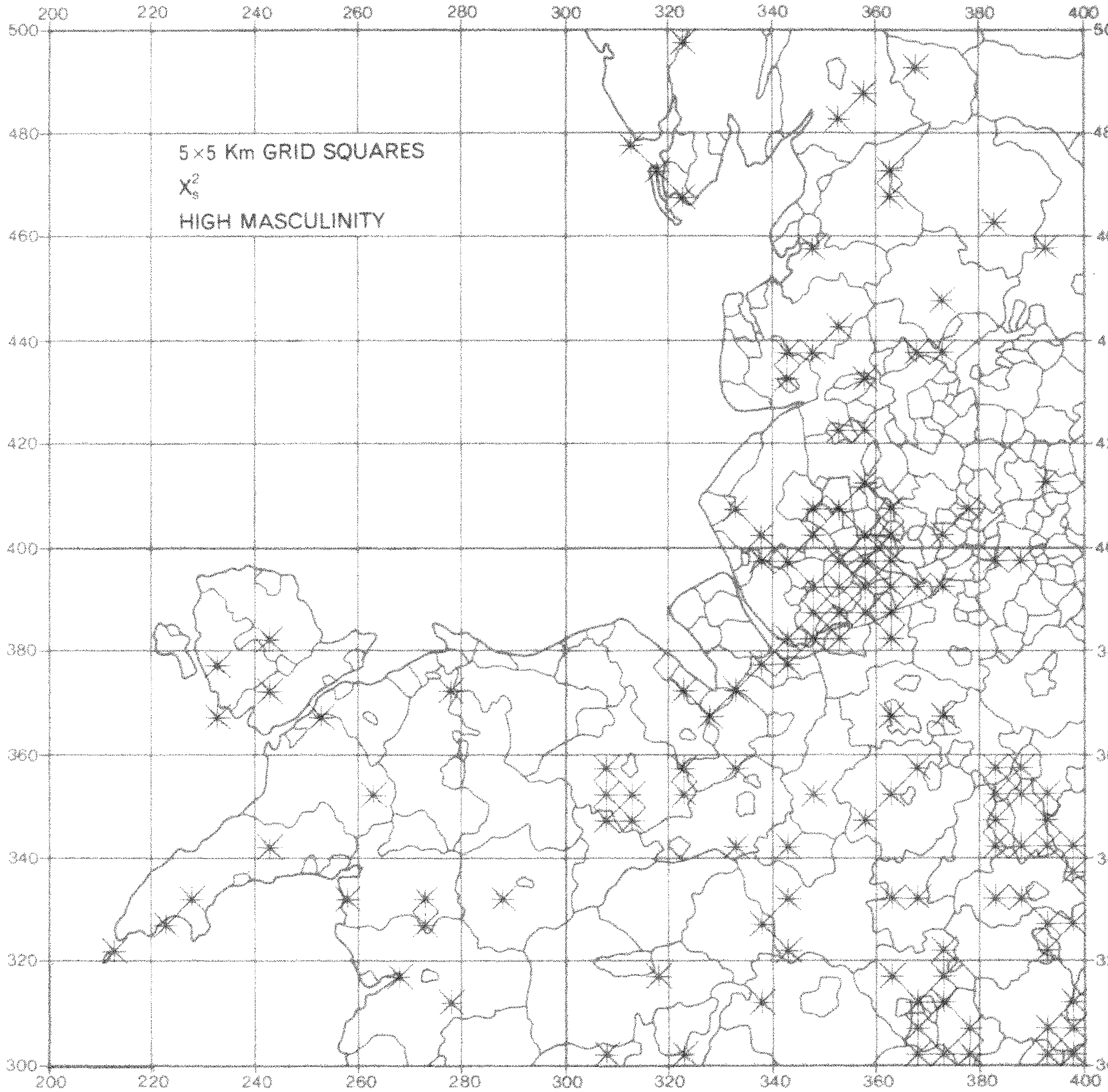


Fig. 12. X_s^2 map at the 5 x 5 km level, showing squares with high masculinity

* squares with $X_s^2 \geqslant 3.84$

No. of units = 134; proportion of total units = 14.79%

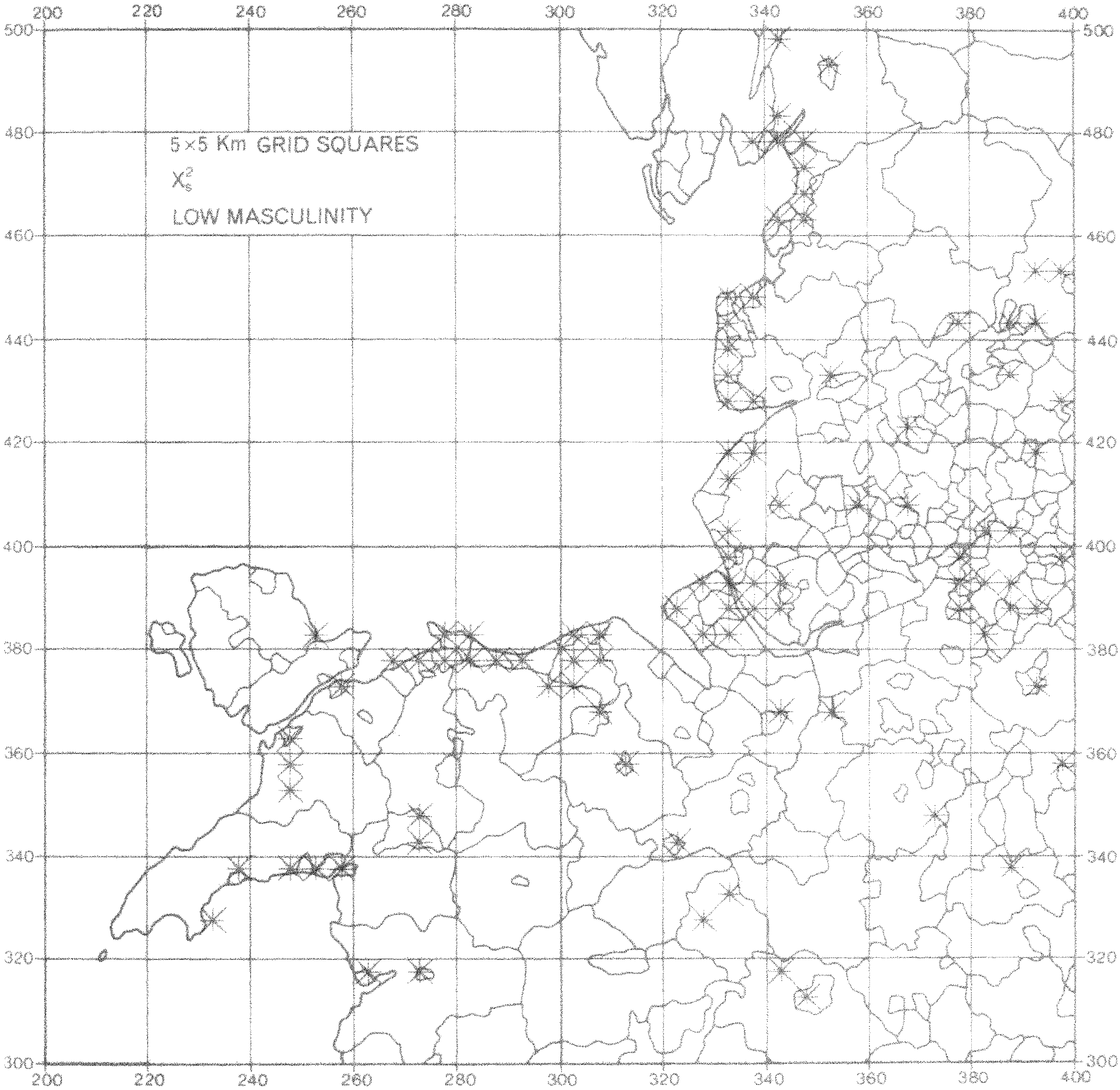


Fig. 13 X_s^2 map at the 5x5 km level, showing squares with with low masculinity

* squares with $X_s^2 \leqslant -3.84$
No. of units = 97; proportion of total units = 10.71 %

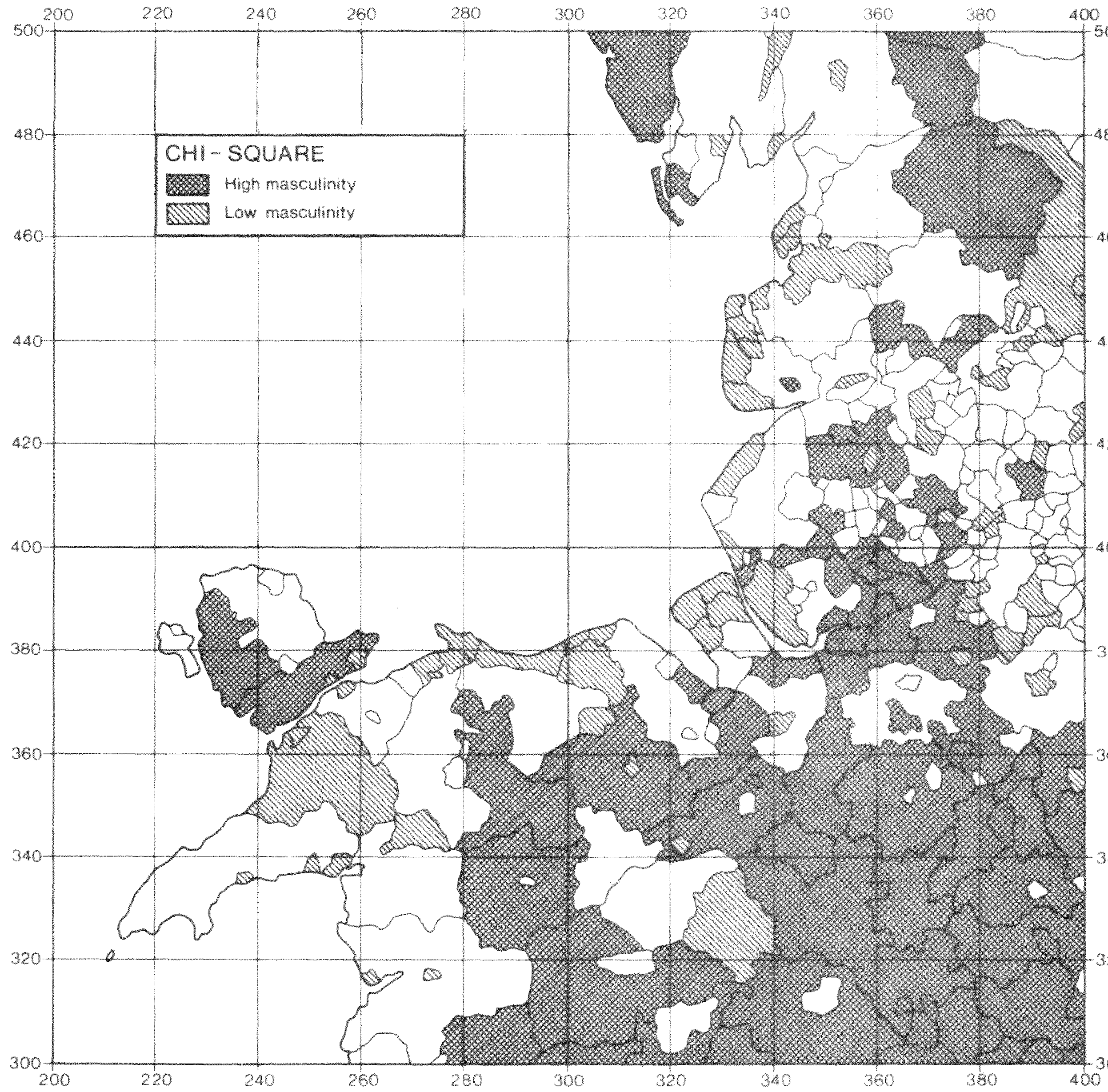


Fig. 14 X_s^2 map showing LAAs with markedly high or low masculinity

LAAs with $X_s^2 \geqslant 3.84$
No. of units =78; proportion of total units = 27.46%

LAAs with $X_s^2 \leqslant -3.84$
No. of units = 67; proportion of total units = 23.59%

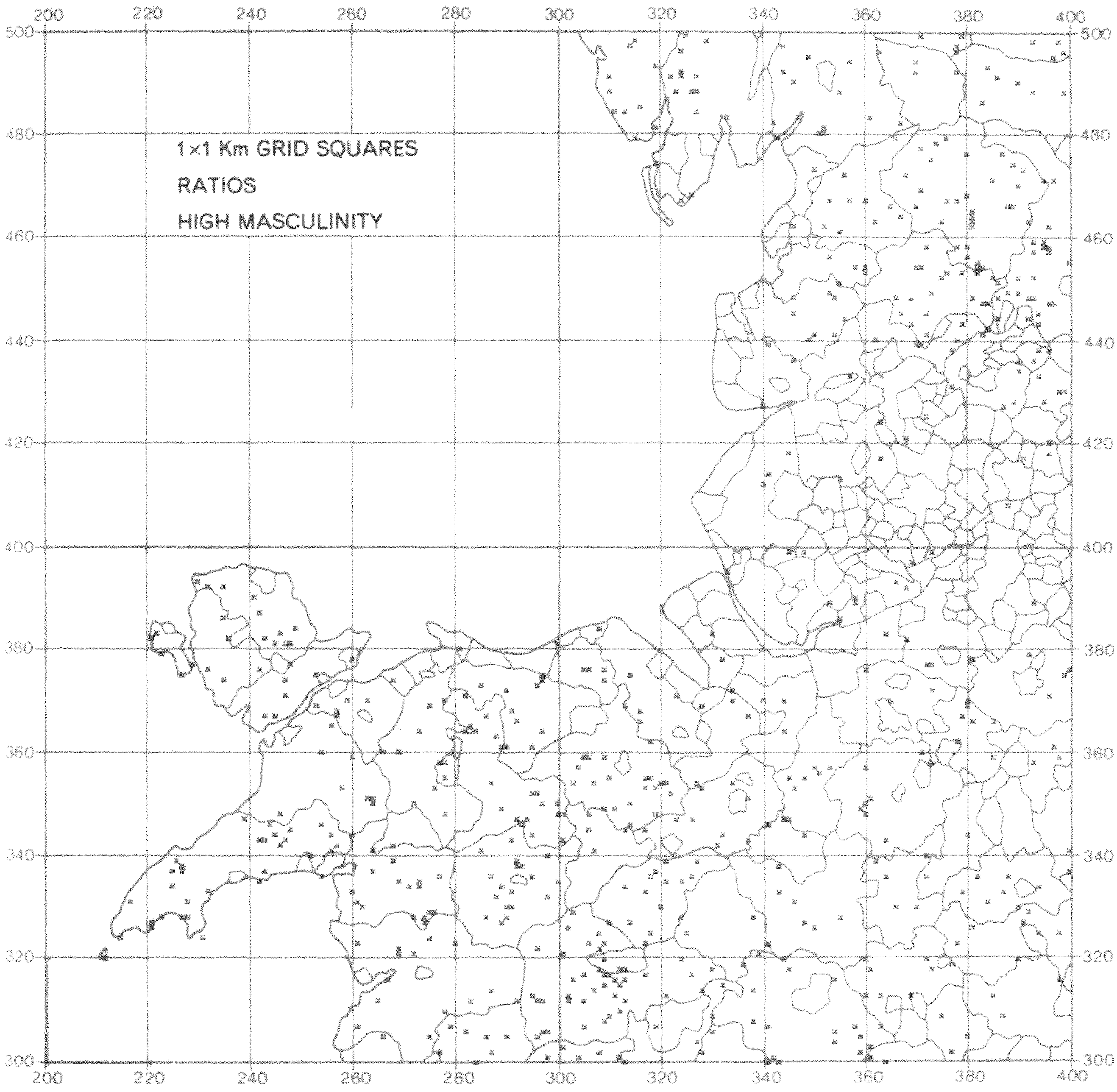


Fig. 15 Ratio map at the 1 x 1 km level, showing squares with high masculinity

* squares with masculinity ≥ 72.73%
No. of units = 606; proportion of total units = 3.65%

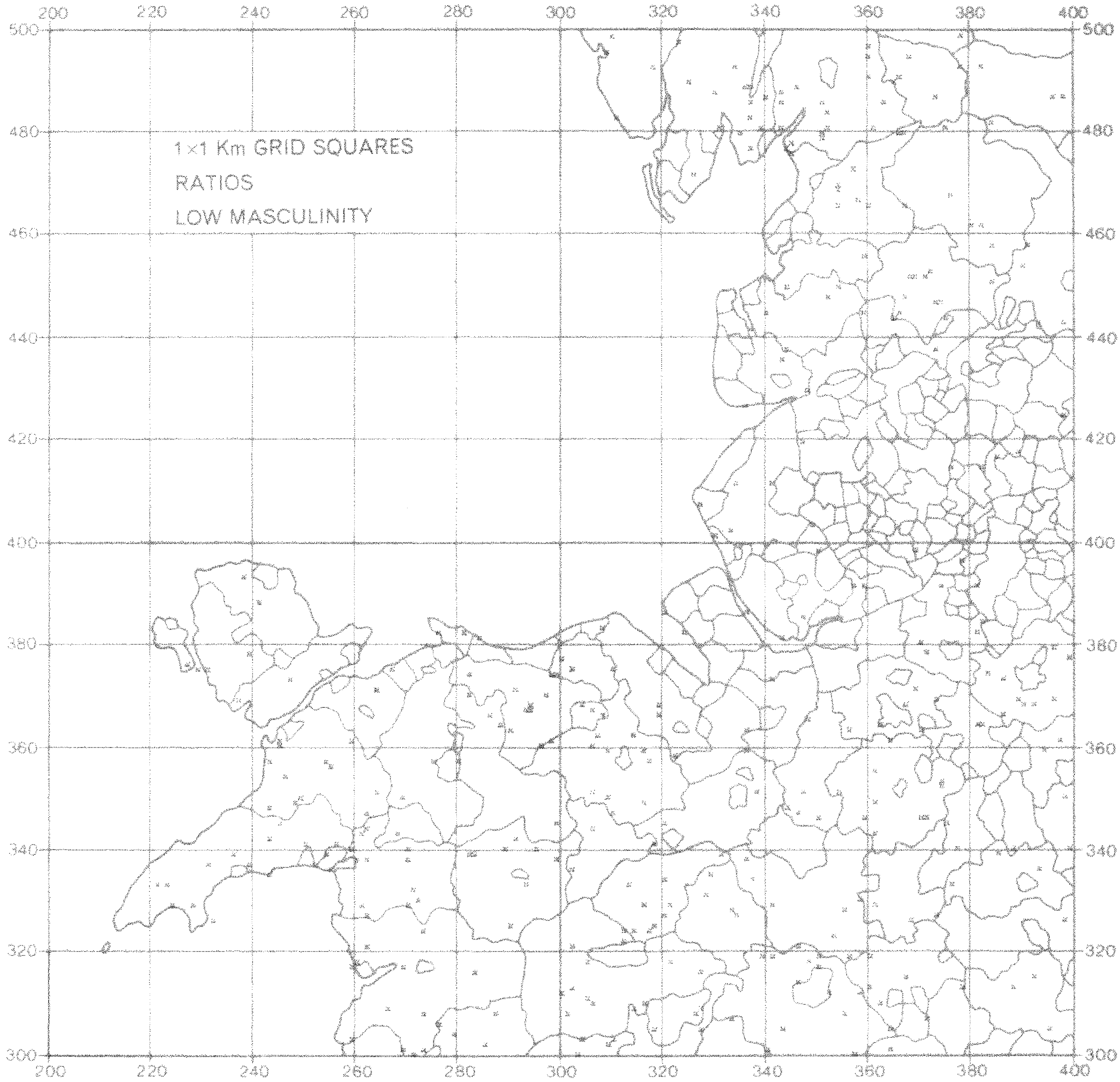


Fig. 16 Ratio map at the 1 x 1 km level, showing squares with low masculinity

* squares with masculinity ⩽ 28.81 %
No. of units = 349; proportion of total units = 2.10%

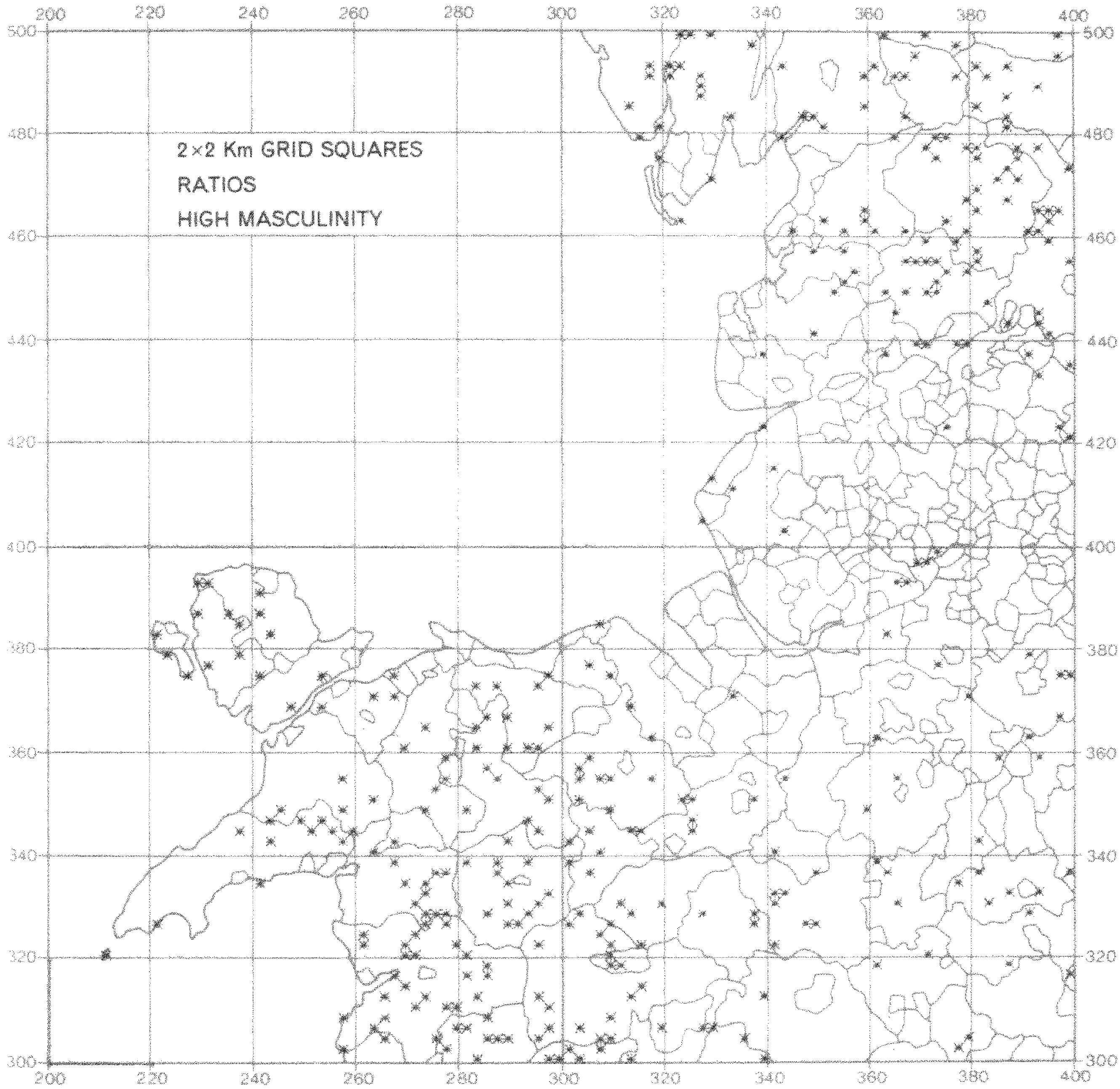


Fig. 17 Ratio map at the 2 x 2 km level, showing squares with high masculinity

* squares with masculinity ≥ 60.17%
No. of units = 355; proportion of total units = 7.19%

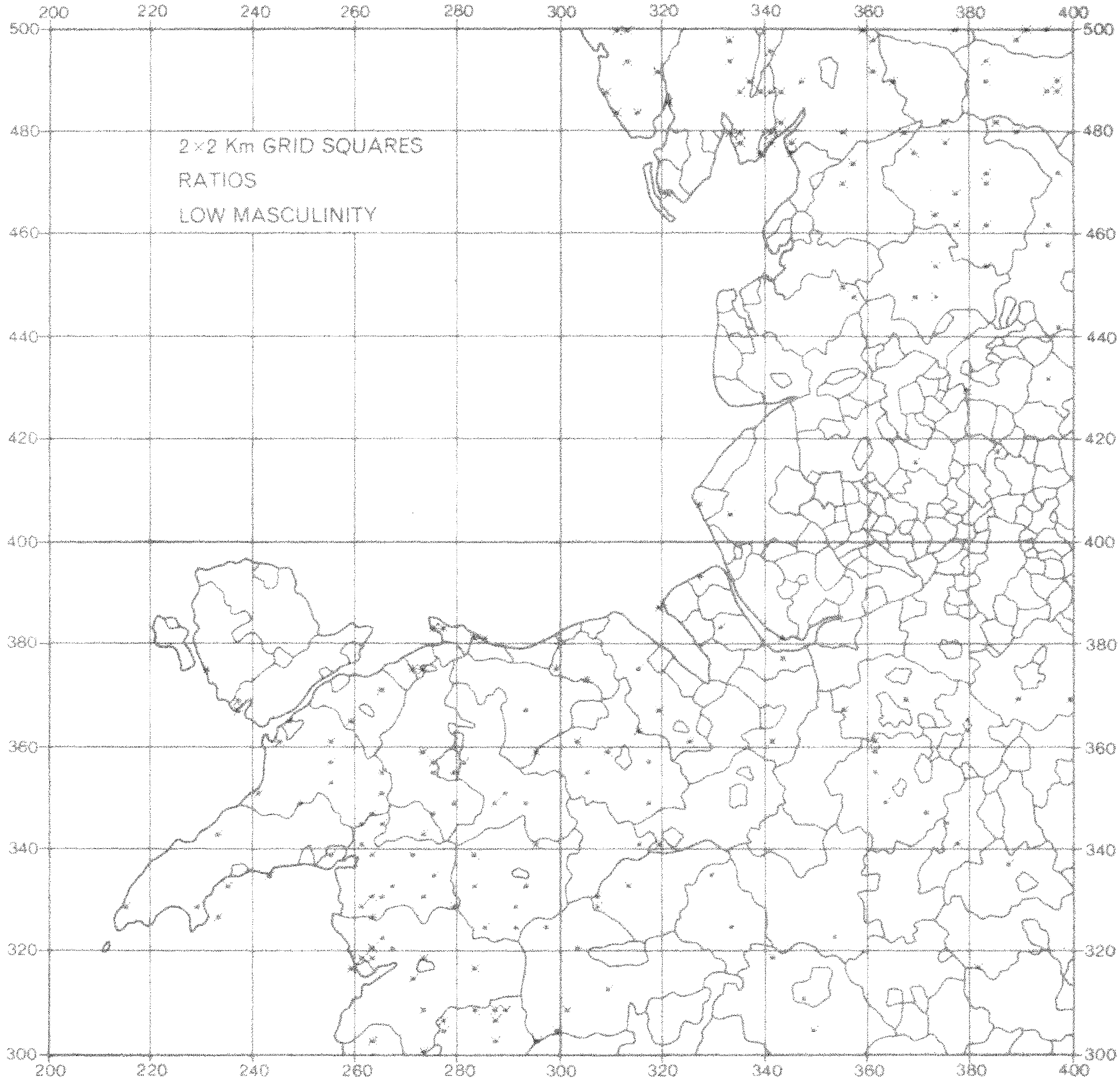


Fig. 18 Ratio map at the 2 x 2 km level, showing squares with low masculinity

* squares with masculinity ⩽ 40.85%
No. of units = 200; proportion of total units = 4.05%

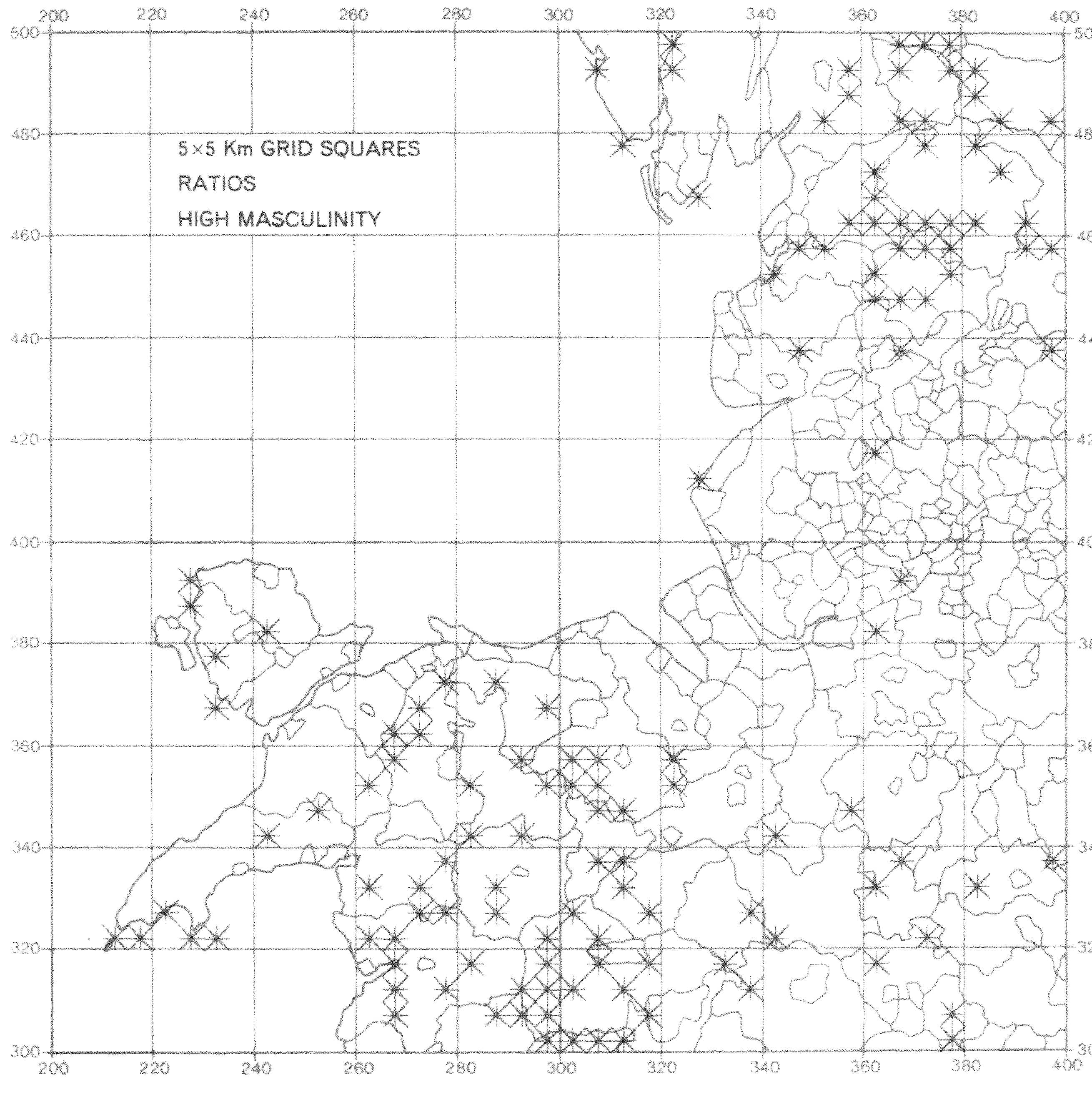


Fig. 19 Ratio map at the 5 x 5 km level, showing squares with high masculinity

* squares with masculinity ≥ 52.35%
No. of units = 134; proportion of total units = 14.79%

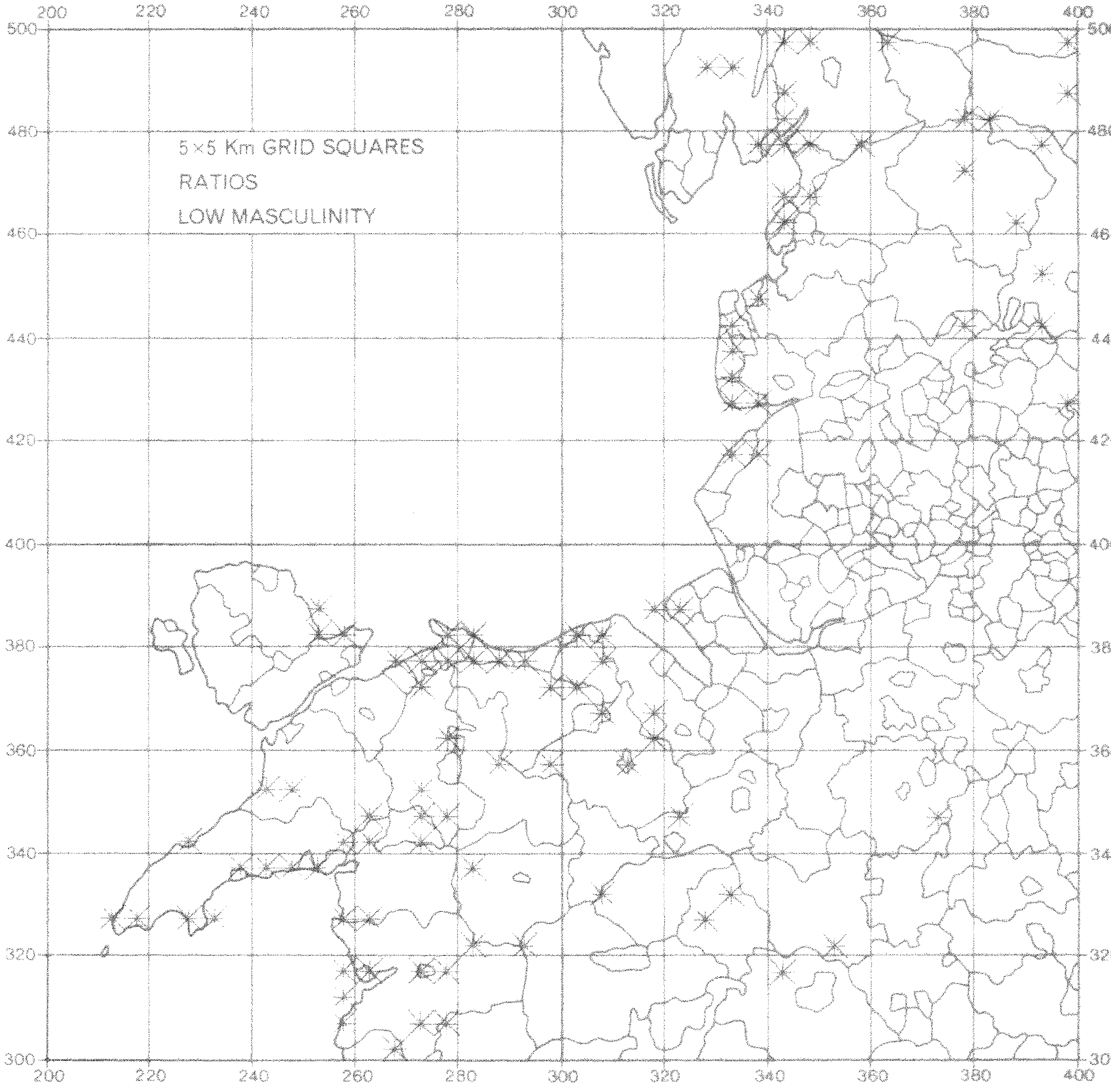


Fig. 20 Ratio map at the 5 x 5 km level, showing squares with low masculinity

* squares with masculinity ≤ 46.5%
No. of units = 98; proportion of total units = 10.82%

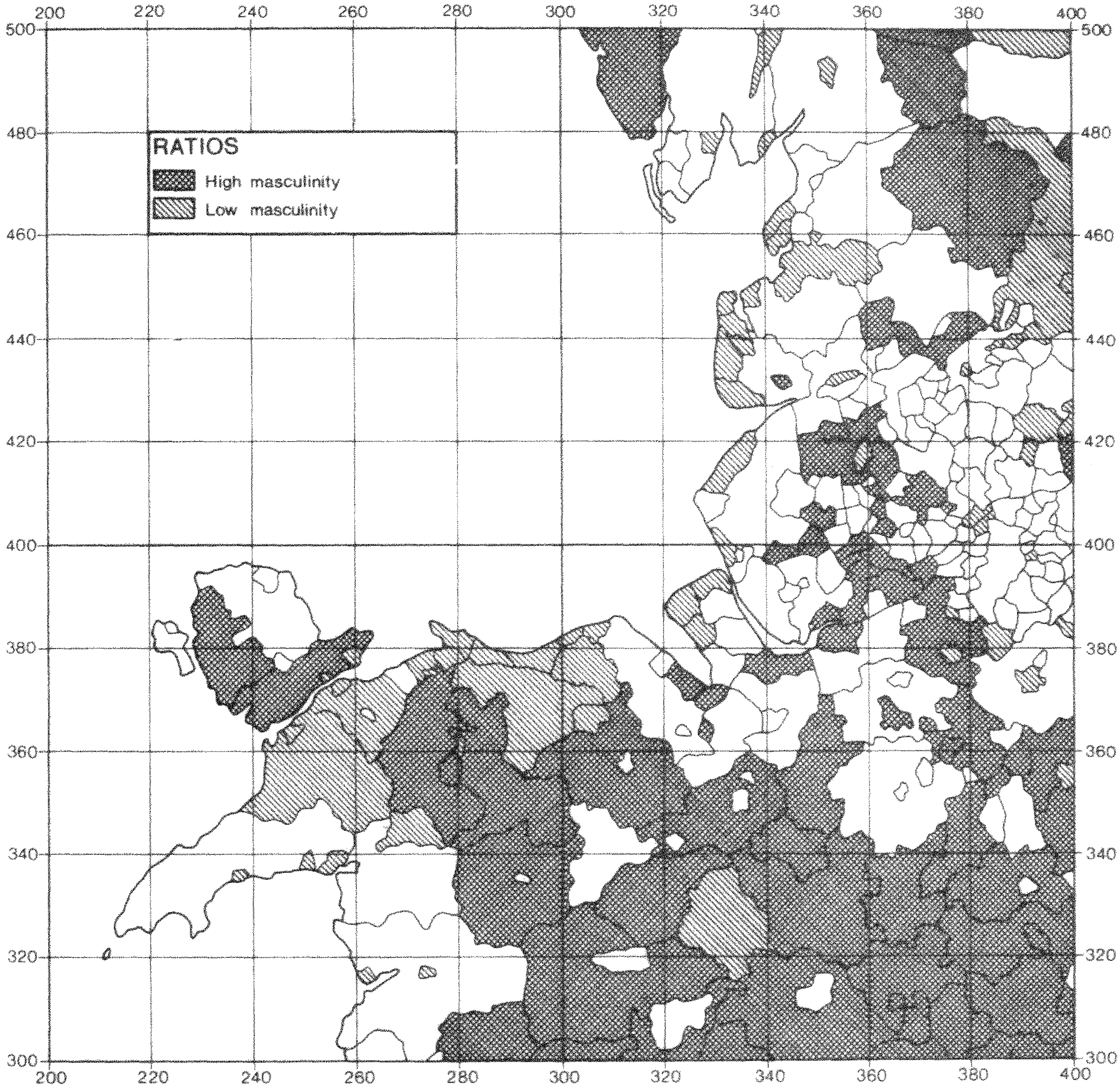
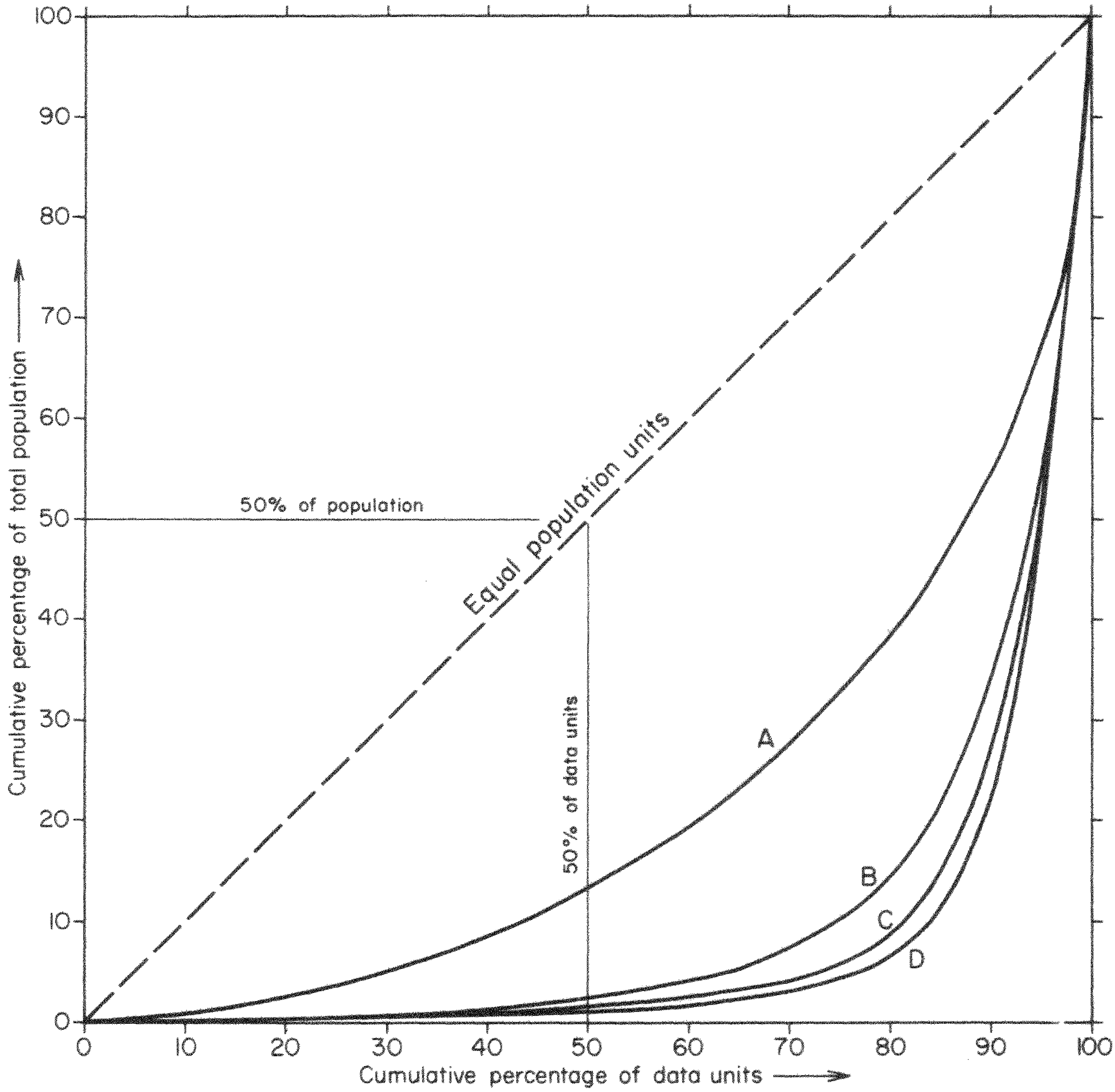


Fig. 21 Ratio map showing LAAs with markedly high or low masculinity

LAAs with masculinity ⩾49.2%
No. of units = 74; proportion of total units = 26.06%

LAAs with masculinity ⩽ 47.8%
No. of units = 65; proportion of total units = 22.89%



A Local Authority Areas (284 units)
B 5 km grid squares (906 units)
C 2 km. grid squares (4,904 units)
D 1 km grid squares (16,612 units)

Fig. 22 Lorenz curves showing the distribution of population among data units at four levels of aggregation

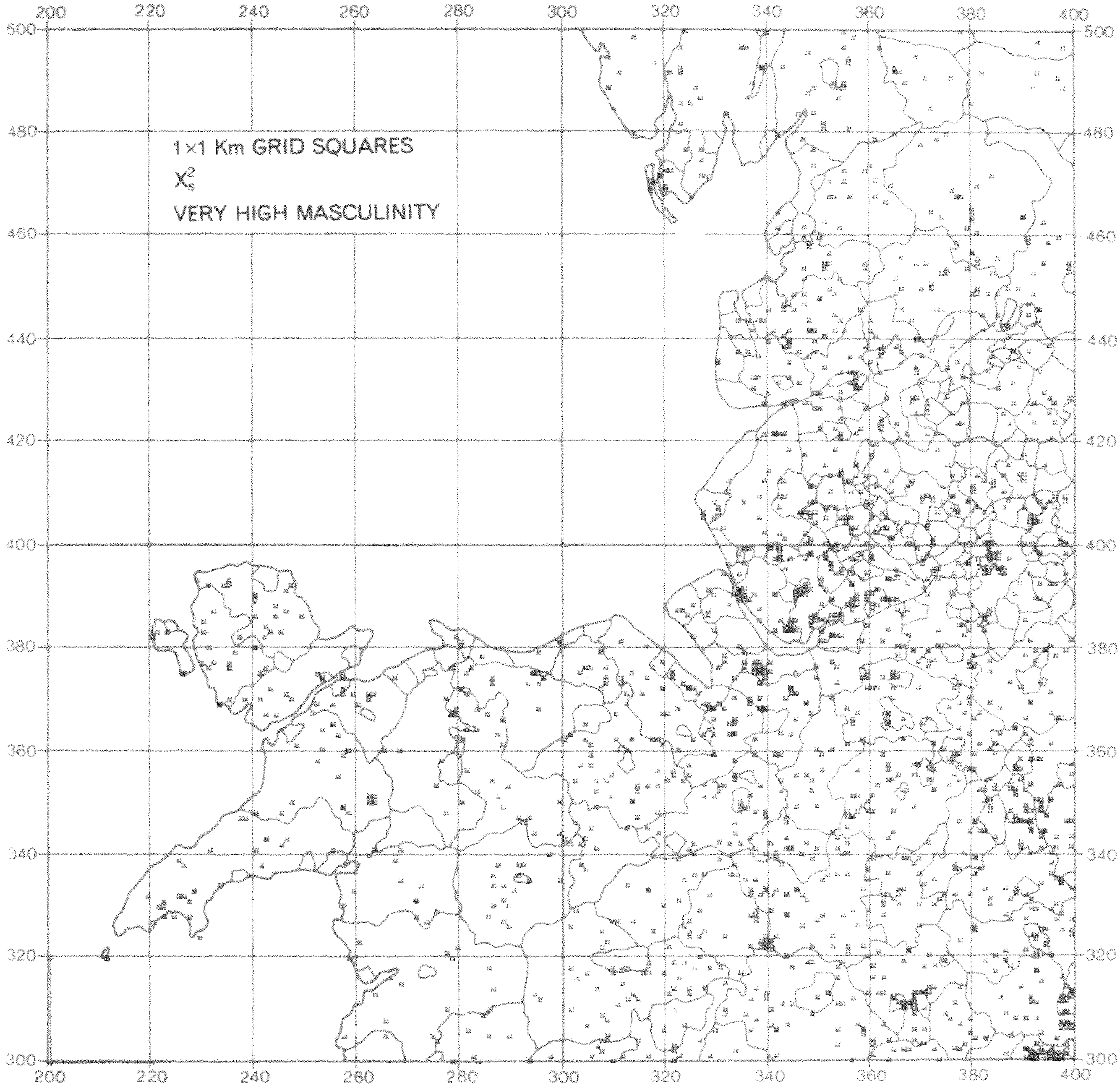


Fig. 23 X_s^2 map at the 1 x 1 km level, showing the ten per cent of squares with highest masculinity

* squares with $X_s^2 \geqslant 1.551$

No. of units = 1661; proportion of total units = 10.0%

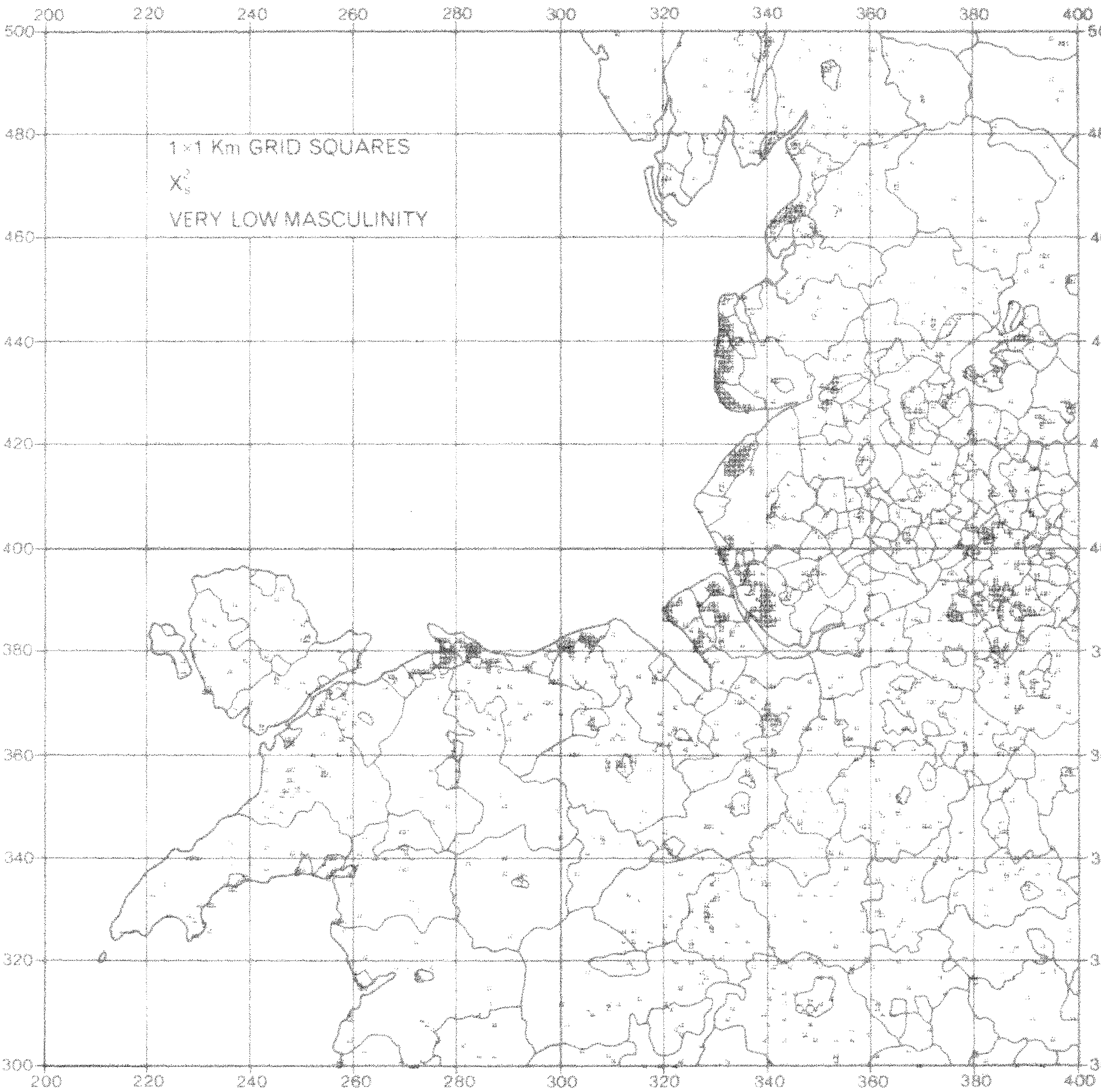


Fig. 24 X_s^2 map at the 1 x 1 km level, showing the ten per cent of squares with lowest masculinity

* squares with $X_s^2 \leqslant -1.105$
No. of units = 1661; proportion of total units = 10.0%

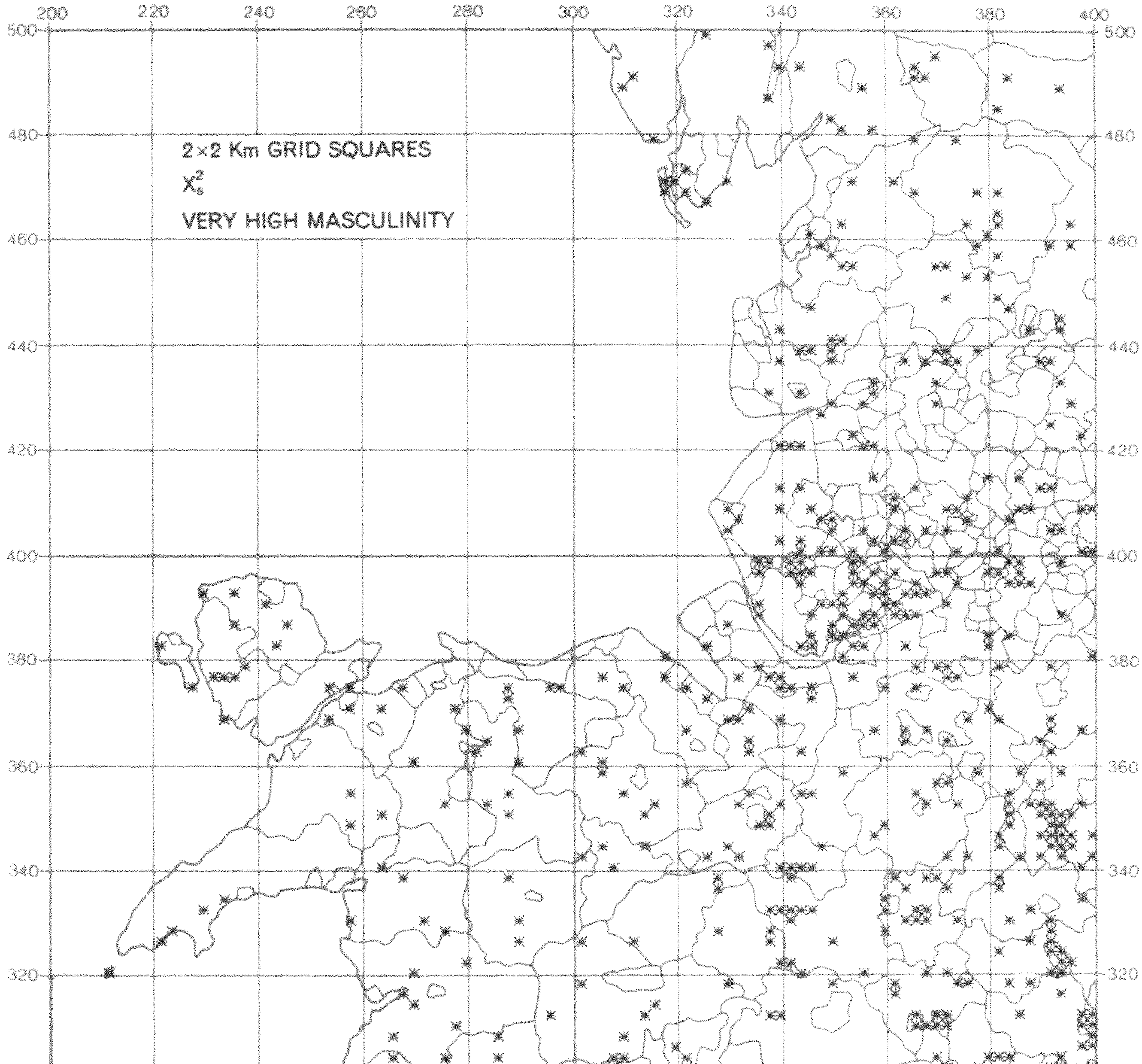


Fig. 25 X_s^2 map at the 2 x 2 km level, showing the ten per cent of squares with highest masculinity

* squares with $X_s^2 \geqslant 2.508$
No. of units = 494; proportion of total units = 10.0%

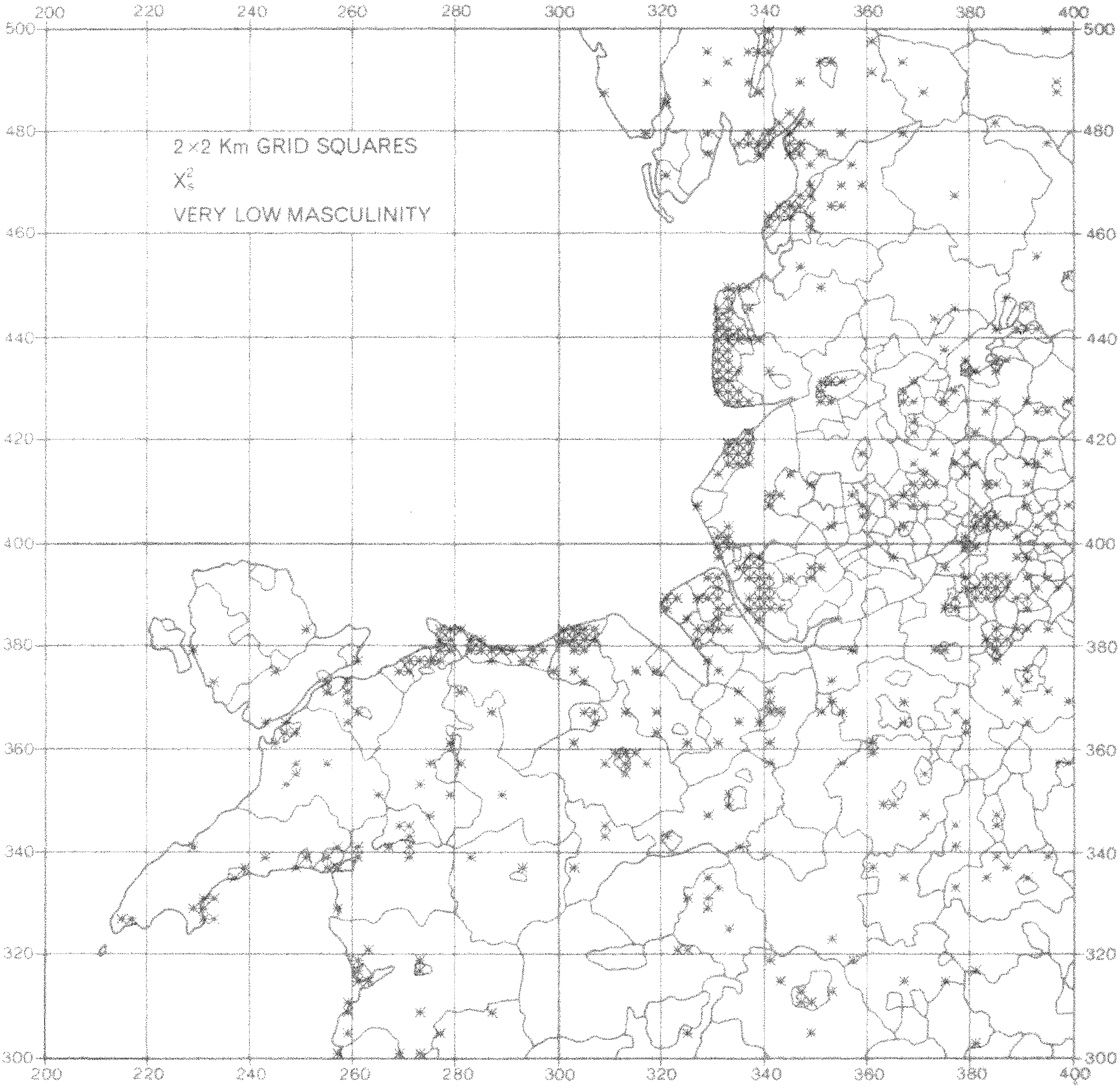


Fig. 26 X_s^2 map at the 2 x 2 km level, showing the ten per cent of squares with lowest masculinity

* squares with $X_s^2 \leqslant -1.458$
No. of units = 494; proportion of total units = 10.0%

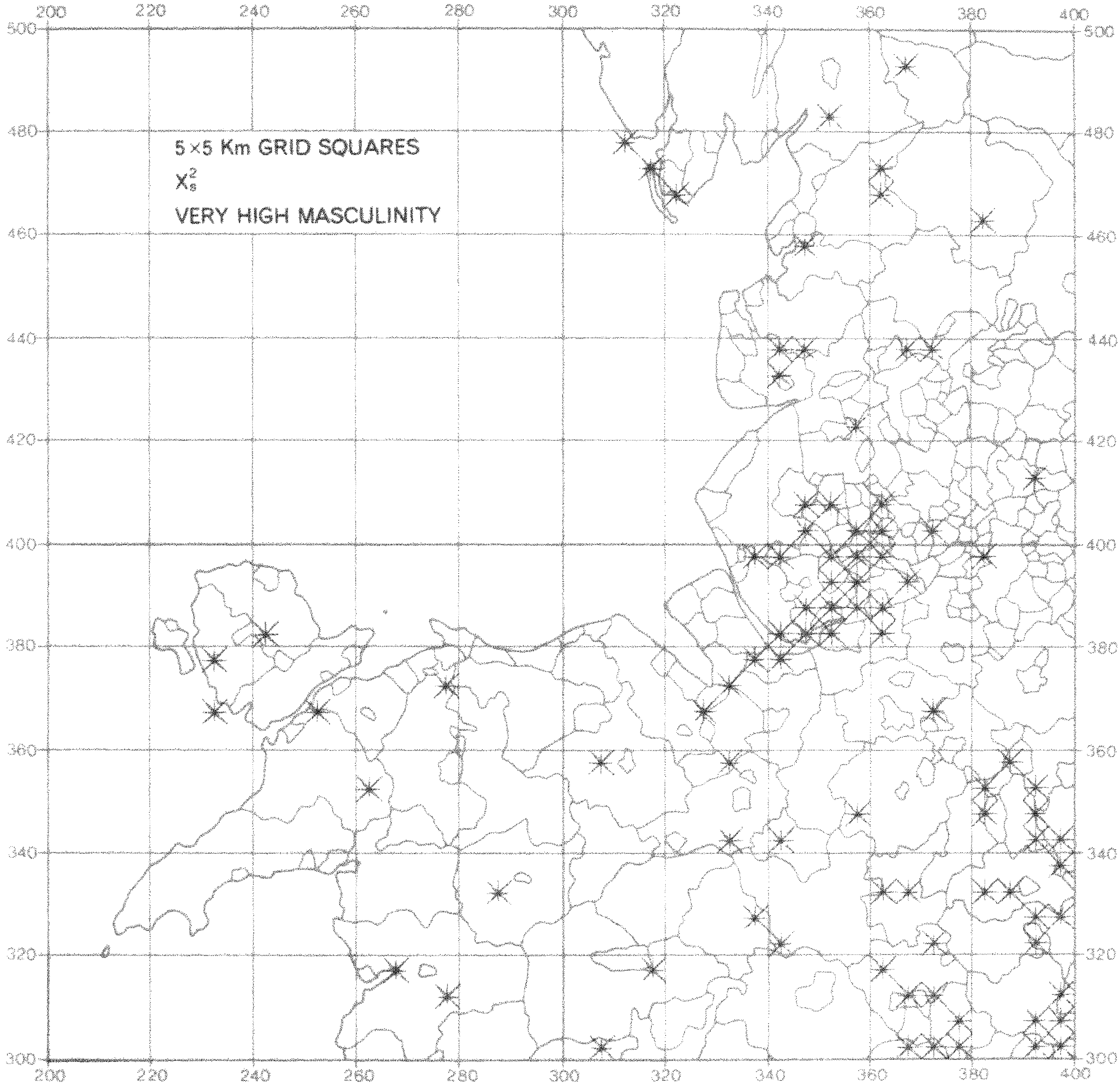


Fig. 27 X_s^2 map at the 5 x 5 km level, showing the ten per cent of squares with highest masculinity

* squares with $X_s^2 \geqslant 5.946$
No. of units = 91; proportion of total units = 10.0%

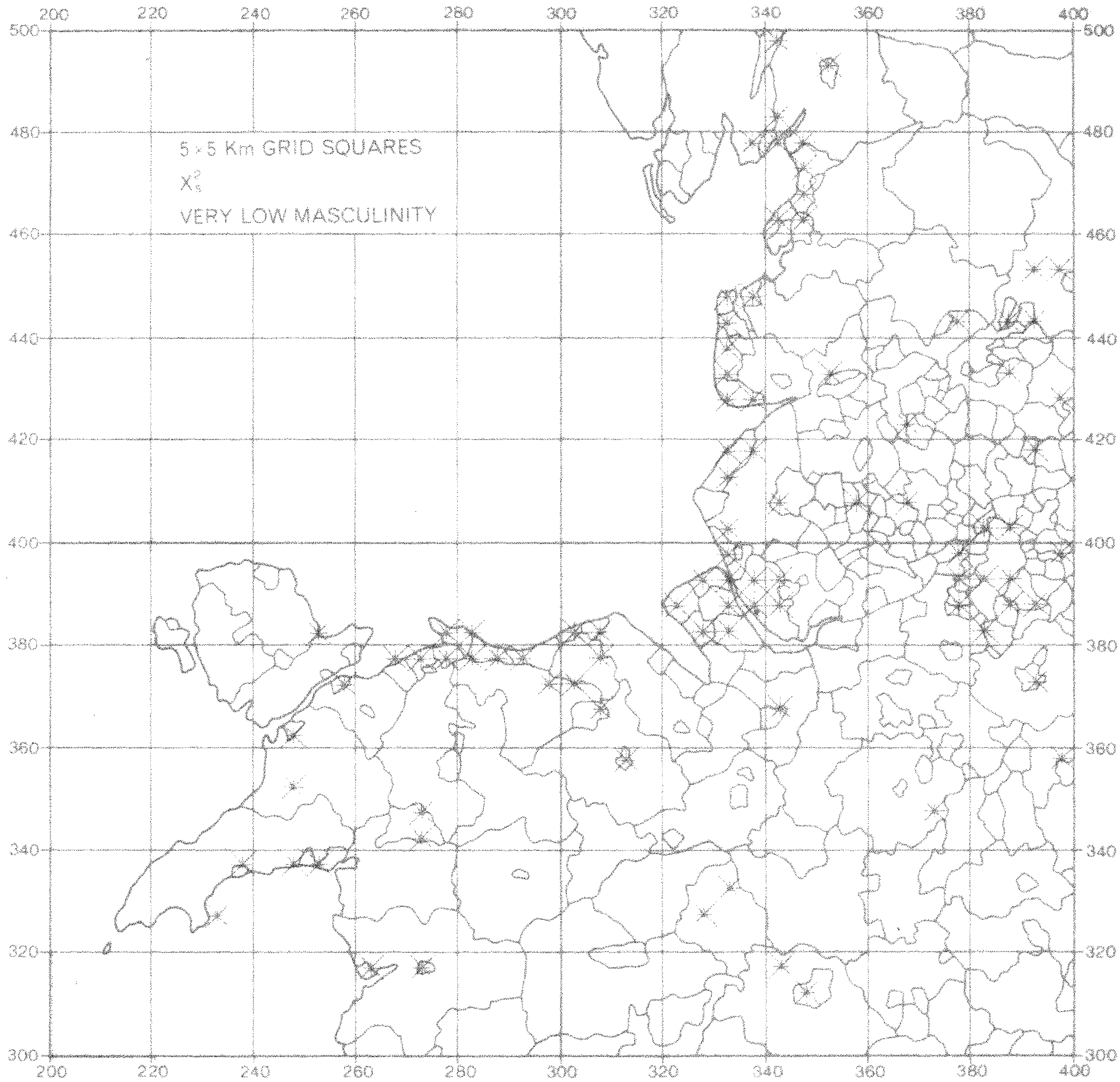


Fig. 28 X_s^2 map at the 5 x 5 km level, showing the ten per cent of squares with lowest masculinity

* squares with $X_s^2 \leqslant -4.383$
No. of units = 91; proportion of total units = 10.0%

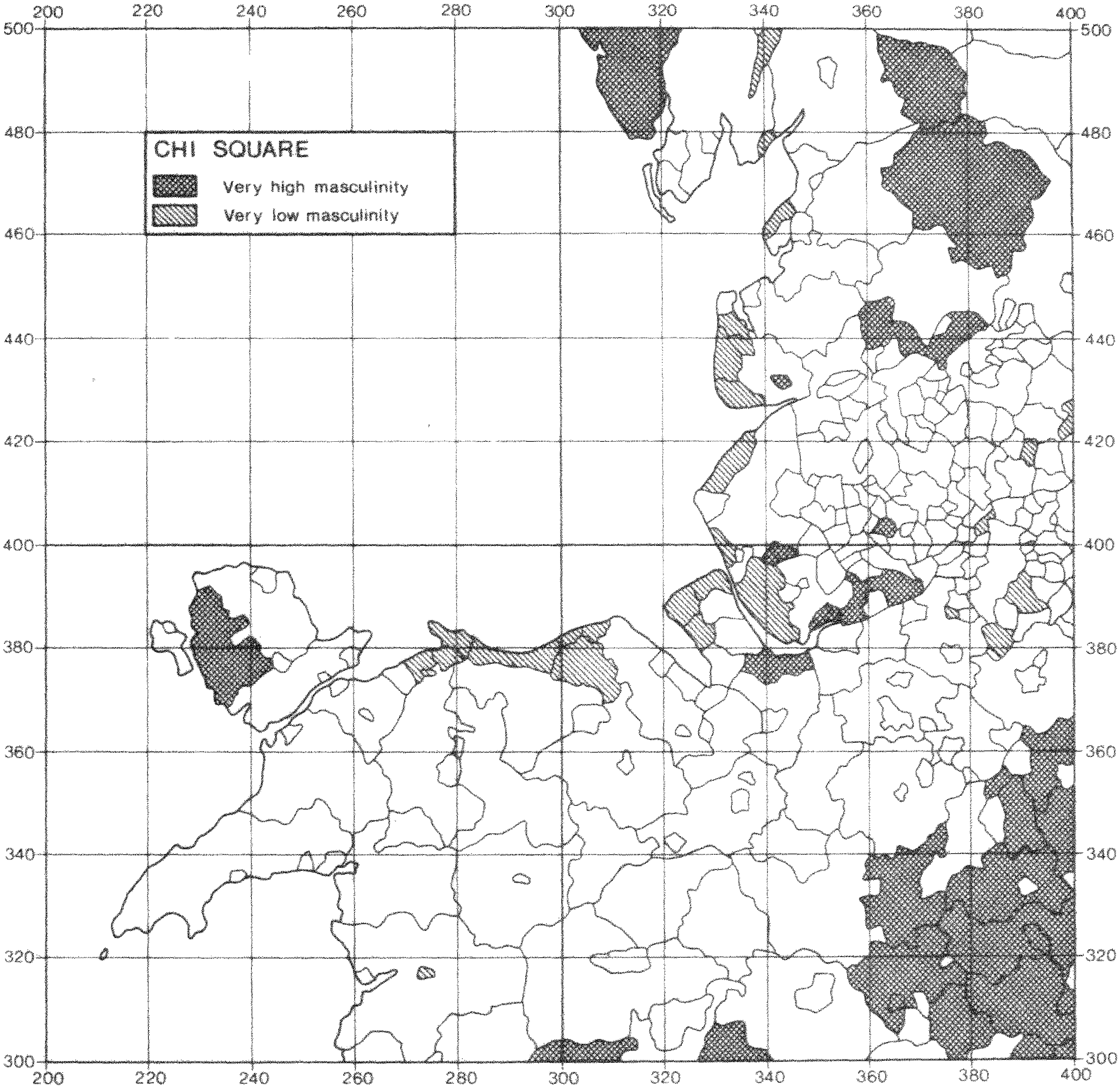


Fig. 29 X_s^2 map showing the ten per cent of LAAs with highest and lowest masculinity

LAAs with $X_s^2 \geqslant 18.820$
No. of units = 28; proportion of total units = 10.0%

LAAs with $X_s^2 \leqslant -20.523$
No. of units = 28; proportion of total units = 10.0%

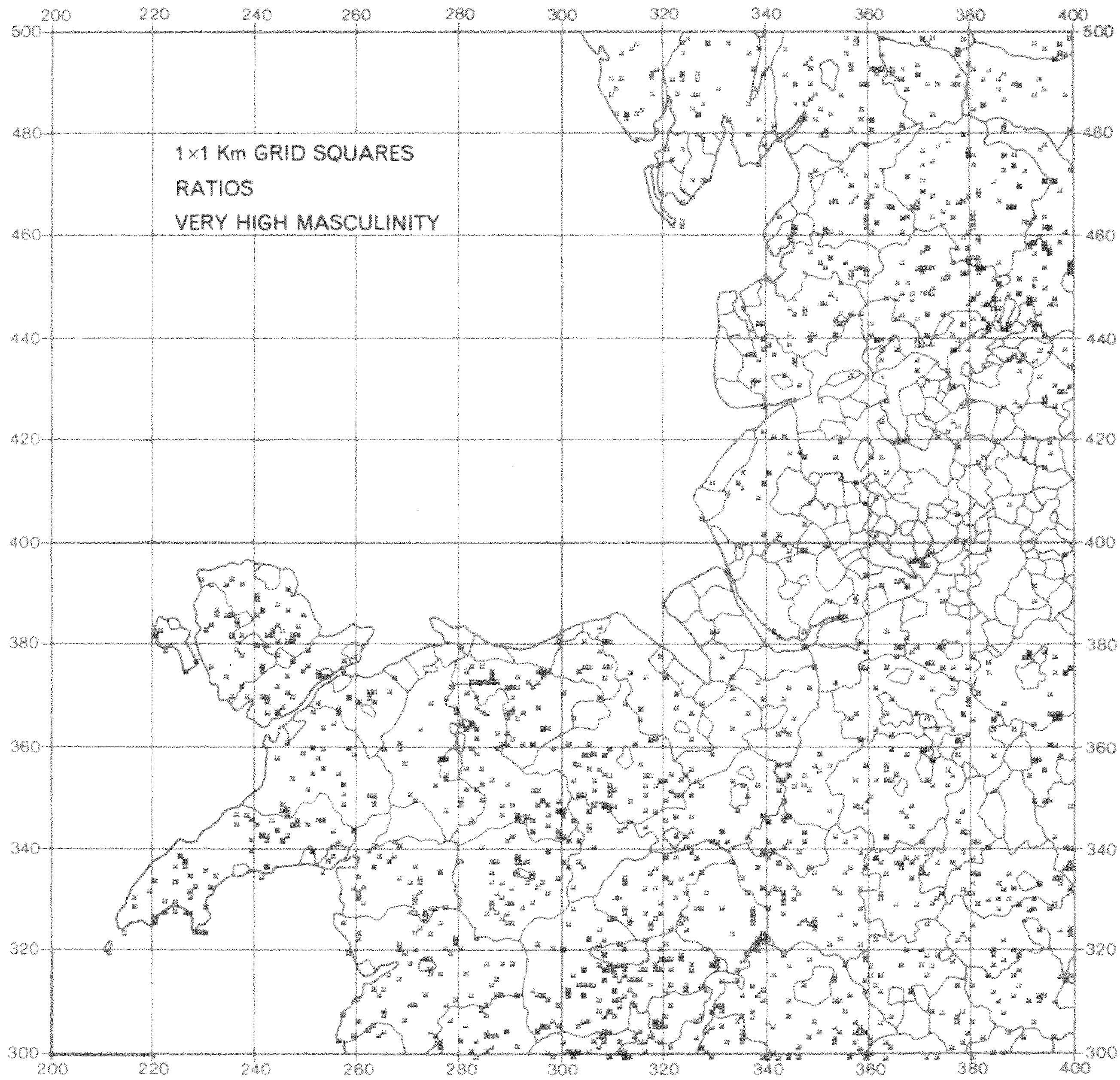


Fig. 30 Ratio map at the 1 x 1 km level, showing the ten per cent of squares with highest masculinity

* squares with masculinity ≥ 62.50%
No. of units = 1661; proportion of total units = 10.0%

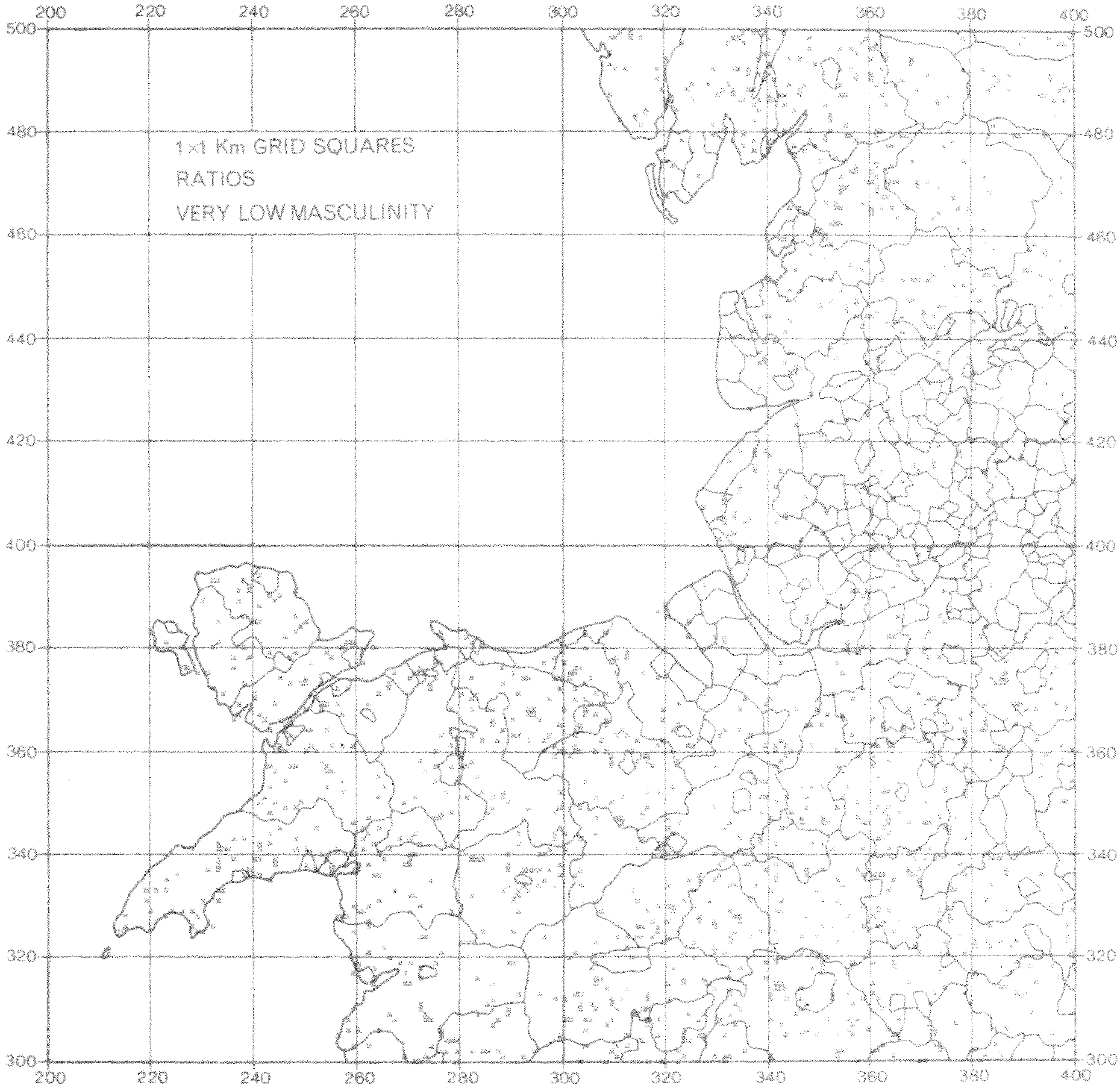


Fig. 31 Ratio map at the 1 x 1 km level, showing the ten per cent of squares with lowest masculinity

* squares with masculinity ⩽ 40.0%
No. of units = 1661; proportion of total units = 10.0%

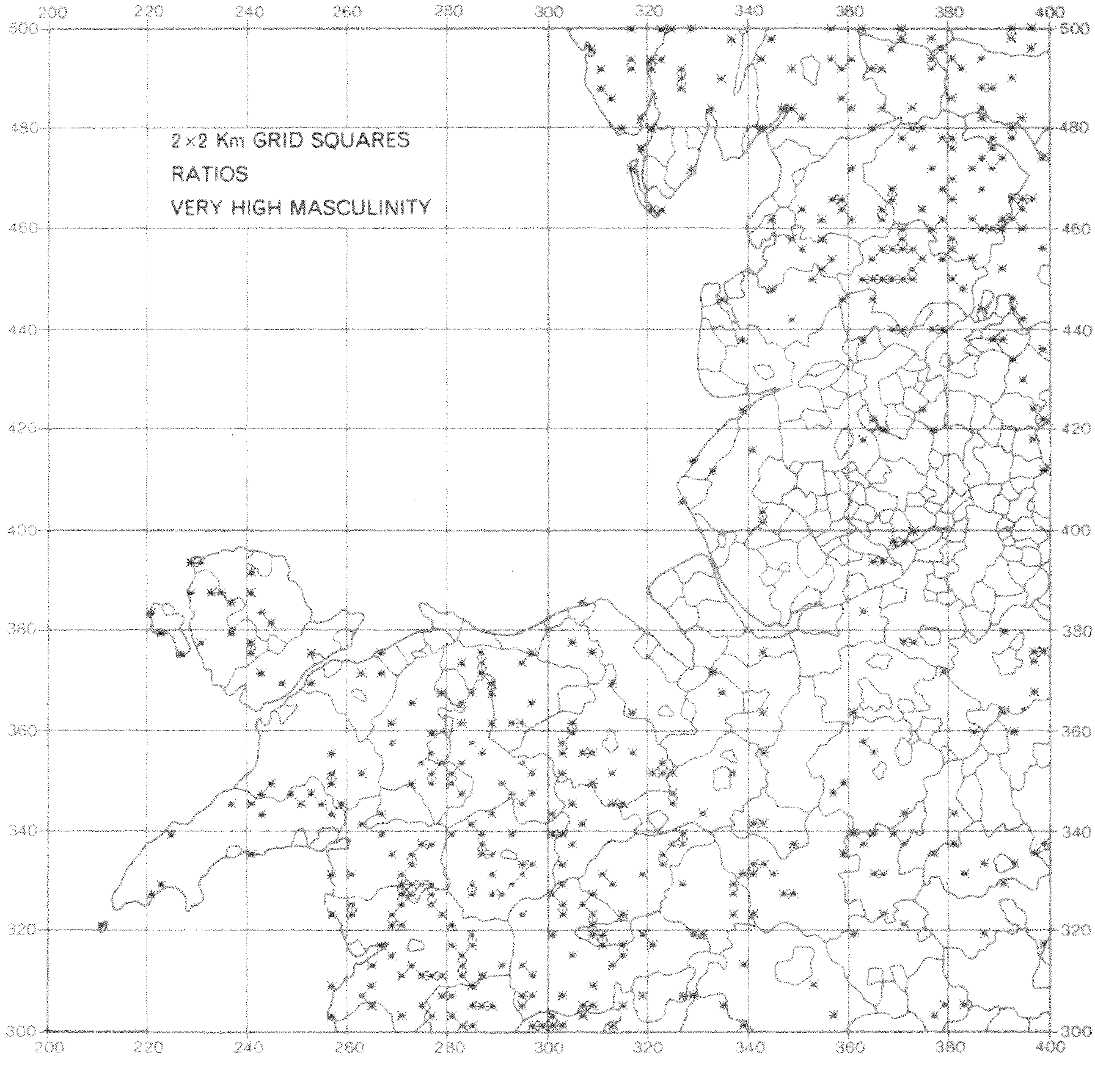


Fig. 32 Ratio map at the 2 x 2 km level, showing the ten per cent of squares with highest masculinity

* squares with masculinity ⩾58.33%
No. of units = 494; proportion of total units = 10.0%

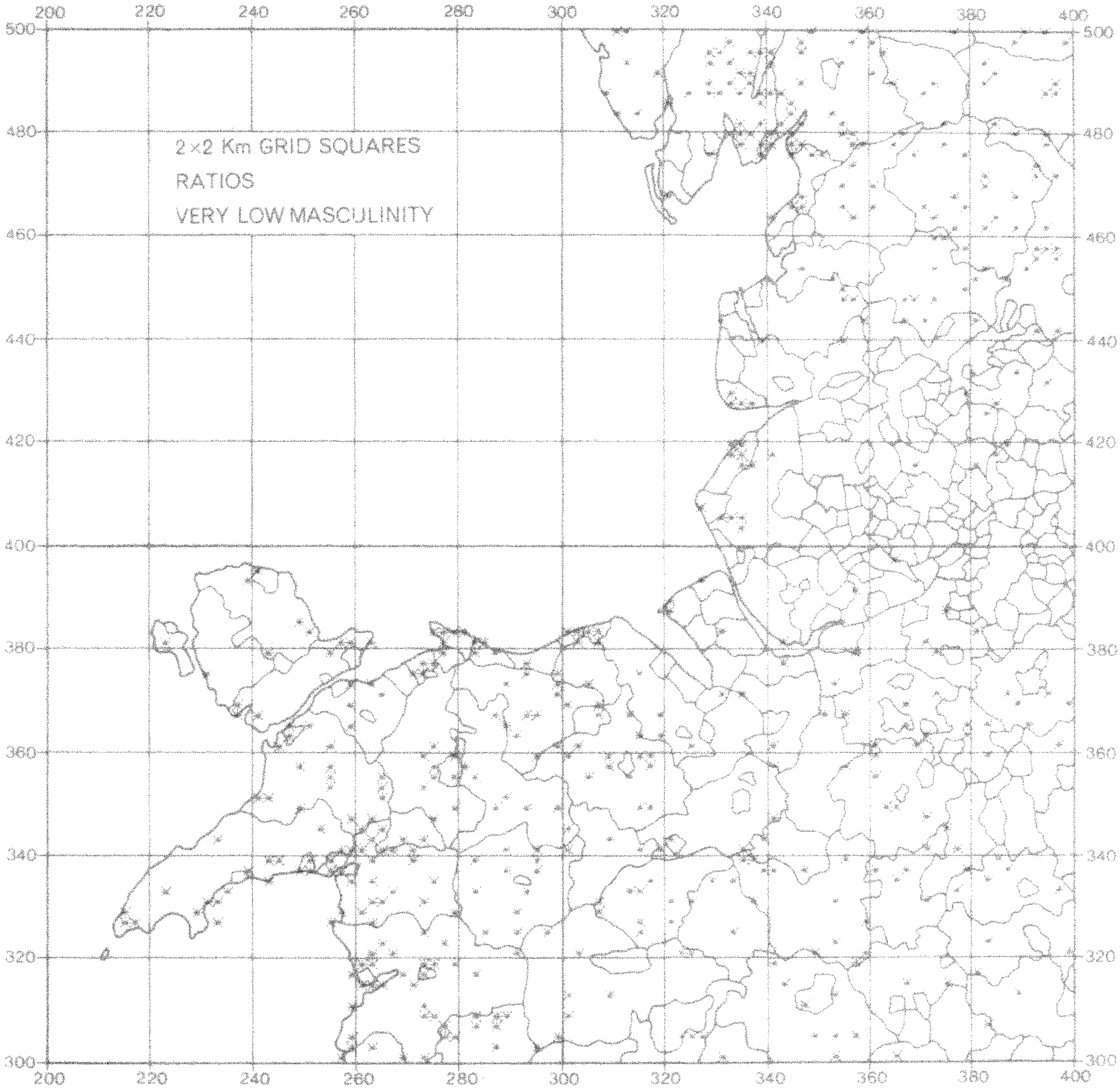


Fig. 33 Ratio map at the 2 x 2 km level, showing the ten per cent of squares with lowest masculinity

* squares with masculinity ≤ 44.53%
No. of units = 494; proportion of total units = 10.0%

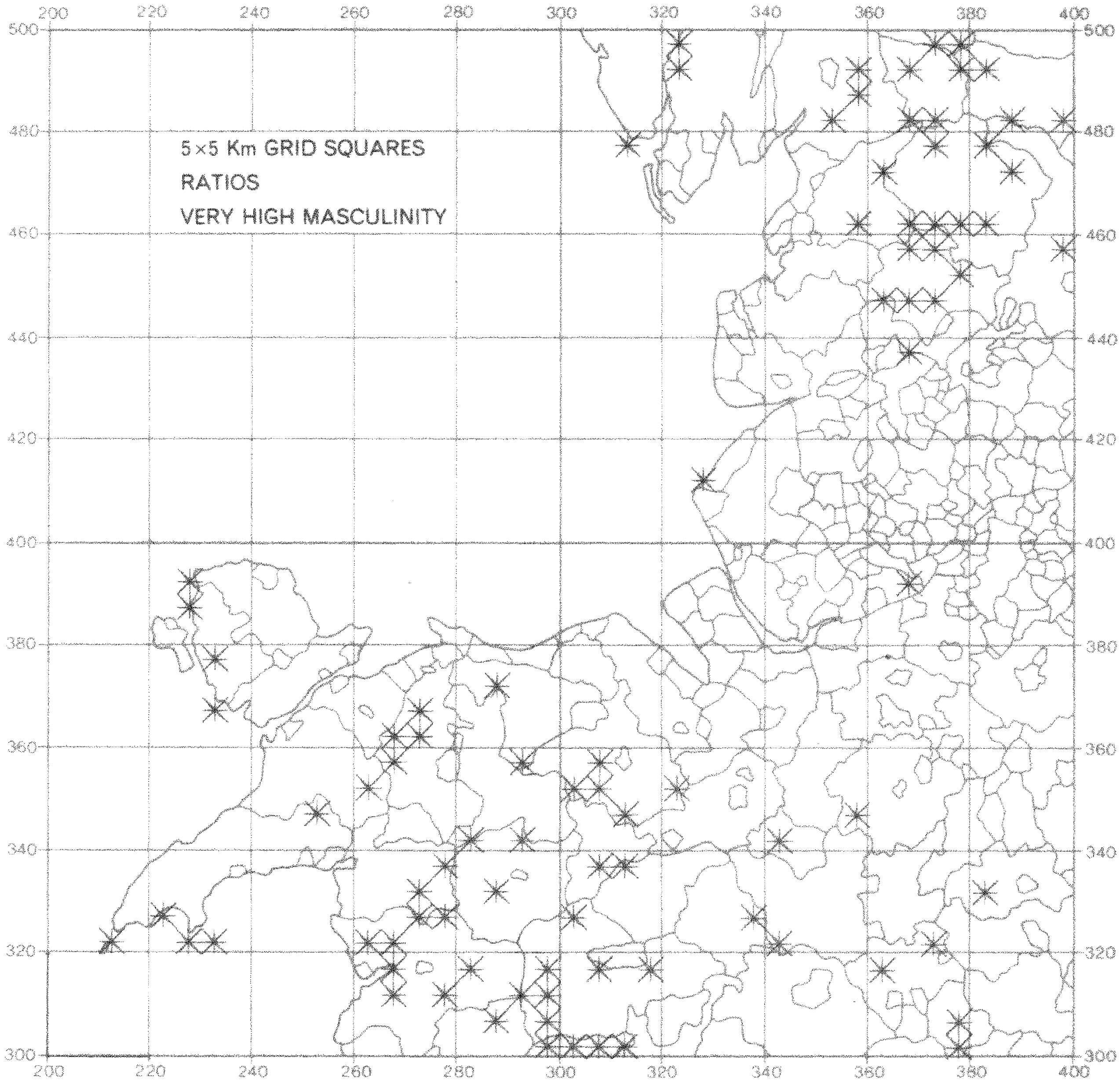


Fig. 34 Ratio map at the 5 x 5 km level, showing the ten per cent of squares with highest masculinity

* squares with masculinity ≥ 53.65%
No. of units = 91; proportion of total units = 10.0%

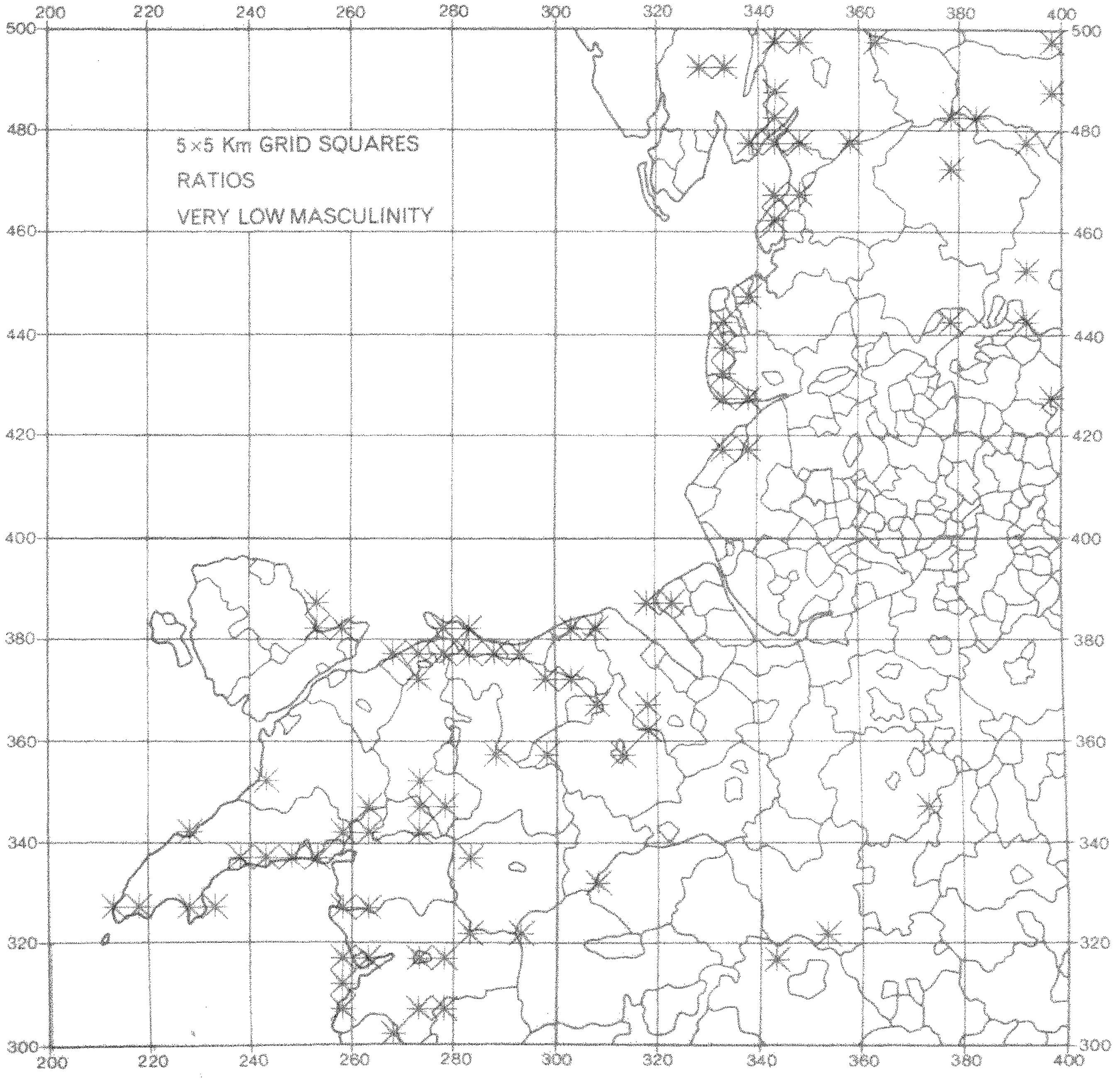


Fig. 35 Ratio map at the 5 x 5 km level, showing the ten per cent of squares with lowest masculinity

* squares with masculinity ⩽ 46.32%
No. of units = 91; proportion of total units = 10.0%

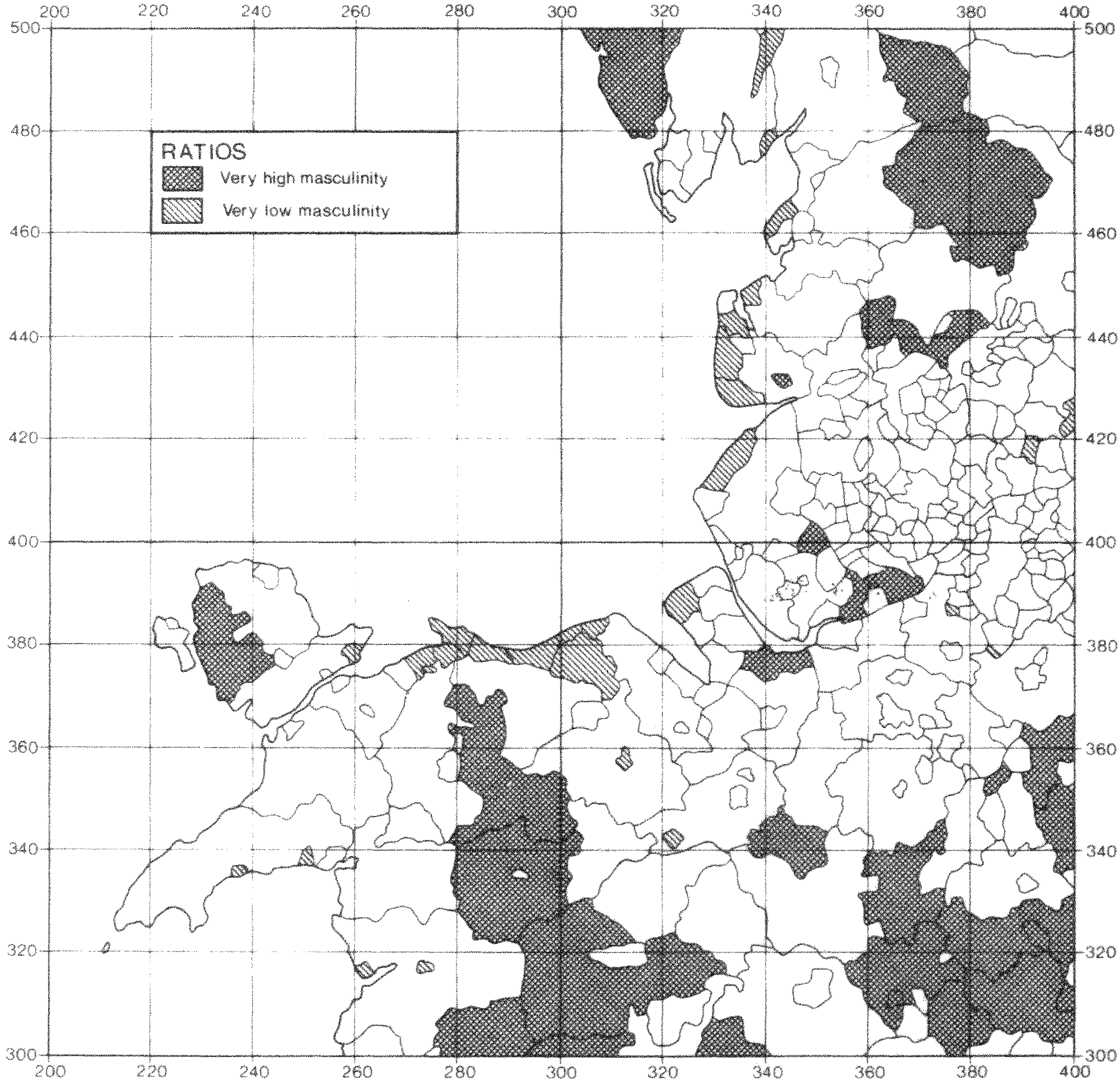


Fig. 36 Ratio map showing the ten per cent of LAAs with highest and lowest masculinity

LAAs with masculinity ≥ 49.99%
No. of units = 28; proportion of total units = 10.0%

LAAs with masculinity ≤ 46.48%
No. of units = 28; proportion of total units = 10.0%