Analysis of the Relationship Between the Scale of Survey Unit and Survey Co-incidence for Integrated Environmental Monitoring

FOR

Natural England

BY

Simon M.C. Poulton

BioEcoSS Ltd.

31st March 2010

Executive Summary

- The purpose of this desk study was to investigate the relationship between the scale of sampling units and the degree of co-incidence of a range of national and agri-environment ecological surveys.
- 39 surveys were included in the study. These comprised 10 national surveys with a mean of over 15,000 sites per survey and 29 small-scale agri-environment schemes with a mean of under 200 sites per survey. A total of 164,000 sites were used for the study.
- Ten different scales of sampling unit were derived, ranging from 100km OS grid squares to single hectares. This gave a 1,000-fold linear range and a 1,000,000-fold range in areas.
- For each scale, frequency tables of the number of surveys per sampling unit were derived, from which basic distributional statistics such as the median, quartiles, maximum and minimum were derived. The pattern of change in these statistics with scale was very different for the large-scale national surveys and the small-scale agri-environment surveys. In the former case, the tetrad (2km sq.) or monad (1km sq.) appeared to provide the most useful scale. However, for the agri-environment surveys a larger scale such as 20km sq. or hectad (10km sq.) appeared to be more useful.
- The geographic distribution of co-incidences were generally very even. In the large-scale surveys, there appeared to be a low co-incidence of surveys in south Lincolnshire and the Fens, with a high co-incidence in Hampshire and the New Forest. In contrast, the agri-environment surveys were dominated by the Arable Stewardship sites in the West Midlands and East Anglia, as well as certain of the more intensively surveyed ESAs such as the Somerset Levels.
- A separate sub-analysis was carried out of the 45 current and proposed ECN and ECBN sites. This showed a similar scale/co-incidence relationship, although the ECBN sites appeared to have a slightly greater number of co-incident surveys than the primary analysis. Furthermore, although there was a tendency for sampling units with greater co-incidence of surveys to be found in the southern half of the country, the full range of co-incidences were well distributed.
- In conclusion, it should be pointed out that this analysis was carried out as a theoretical exercise. This meant that issues such as the real size and shape of the sites were ignored – each site was treated simply as a point grid reference, with a resolution of 1 or 10 meters. In reality, this probably had little effect, because the site areas were generally less than most of the sampling unit scales. However, if the findings from this analysis are considered useful for the design of an integrated monitoring scheme, then it would be worthwhile undertaking a more accurate and realistic analysis using a Geographic Information System.

Contents

The Surveys

For the purpose of this analysis, 39 surveys were collated (Table 1). These comprised two main sources;

- National Scale Surveys acquired and/or compiled by Natural England (English Nature)
- Agri-environment Monitoring Schemes currently held in the Environmental Monitoring Database (EMD).

The main difference between these two groups of surveys was the number of sites that were utilised. The national surveys were often inventories of all known examples of certain habitats or features and, consequently may have had tens of thousands of sites. In contrast the agri-environment surveys were targeted geographically at individual Environmentally Sensitive Areas (ESAs) and often involved intensive monitoring of a smaller number of sites. The only exception to this distinction was the National Visitor Survey which had only 60 sites. For the purposes of this analysis this survey has been treated as a "small-scale" survey.

The 10 national surveys (excluding the National Visitor Survey) comprised a total of 158,281 sites, giving a mean of 15,828 sites per survey. In contrast, the 29 agri-environment surveys had only 5,541 sites between them giving a mean of 191 sites per survey. The ratio between the means was over 80:1, which justified analysing these two groups separately.

The ESA-based surveys were compiled from a large number of individual surveys to be method based. For example the Breeding Birds in ESAs was actually compiled from 11 different ESAs, giving a wide, if localised, geographic coverage.

Table 1. The National-Scale and Agri-Environment Surveys used in this analysis.

National-Scale Surveys	*Sites*	*Agri-Environment Surveys*	*Sites*
Butterfly Transects	1,299	Arable Stewardship - Breeding Birds	100
EN Grassland Habitat Inventory	14,836	Arable Stewardship - Brown Hares	79
EN Heathland Habitat Inventory	5,357	Arable Stewardship - Bumblebees	21
EN Mire Habitat Inventory	7,153	Arable Stewardship - Cereal Field Margins	560
EN Phase II Grassland Inventory	8,163	Arable Stewardship - Grey Partridge	80
EN Woodland Habitat Inventory	54,789	Arable Stewardship - Ground Beetles	57
National Inventory of Woodland and Trees	20,202	Arable Stewardship - Sawflies	56
National Visitor Survey	60	Arable Stewardship - True Bugs	55
Scheduled Ancient Monuments	19,664	Arable Stewardship - Wintering Birds	99
SSSI Unit Conditions	21,638	Breeding Birds in ESAs	369
Wetland Bird Survey	5,180	Conservation Headlands	61
Total (excl. National visitor survey)	**158,281**	CSS - Evaluation of Cultivated Options	312
Mean (excl. National visitor survey)	**15,828**	CSS Agreement Appraisal	483
		CSS BAP Habitat Mapping	508
		Ditches in ESAs	250
		ESA Agreement Appraisal	130
		Grassland ADAS Plots in ESAs	647
		Grassland Indicative Quadrats	460
		Grassland Quadrats in ESAs	407
		Habitat Scheme - Water Fringe	92
		Heather Biomass Utilisation in ESAs	133
		Lowland Heath Burning	12
		Lowland Heath Grazing	18
		Rapid Condition Assessment in ESAs	228
		Set-Aside Quadrats	39
		Stone Curlews in ESAs	55
		Uncropped Wildlife Strips	125
		Wintering Birds in ESAs	45
		Total (incl. National visitor survey)	**5,541**
		Mean (incl. National visitor survey)	**191**

Thumbnail maps of the 39 surveys are shown in Fig. 1. These show the distribution of sites plotted as presence in 10km grid squares (hectads). They clearly show the almost ubiquitous distribution of, *e.g.*, the National Inventory of Woodland and Trees or the Scheduled Ancient Monuments. In contrast, the nine surveys of the Arable Stewardship scheme were all located in the two Pilot areas of West Midlands and East Anglia, although the pattern of locations of the sites were different for each Survey. Finally, the locations of the ESA and CSS

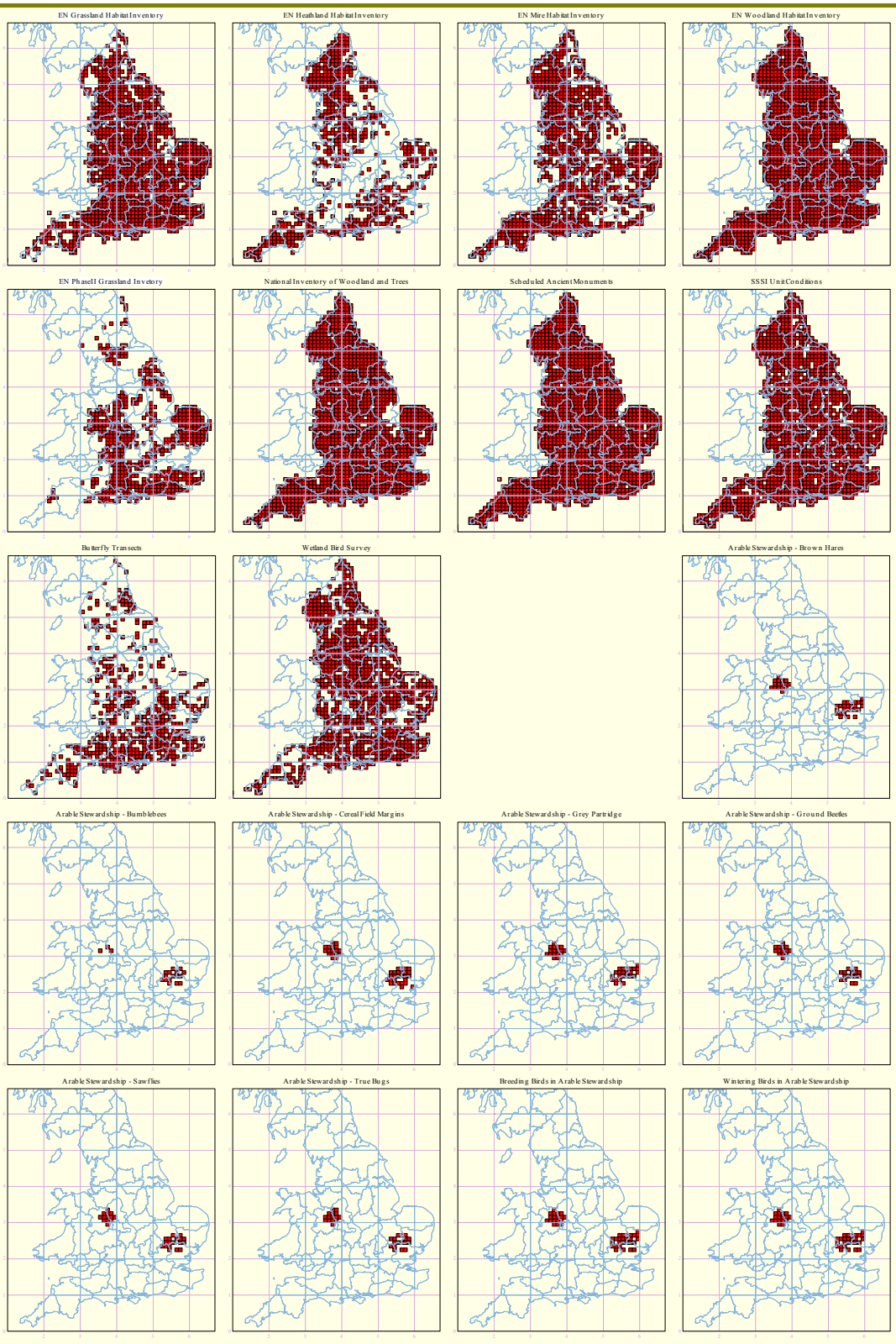

Fig. 1. The distributions of a) the ten national surveys ad b) the nine Arable Stewardship surveys

surveys ranged from a single hectad in the case of the Lowland Heath Grazing in Brecklands, to a fairly wide coverage for the CSS BAP Habitat Mapping.

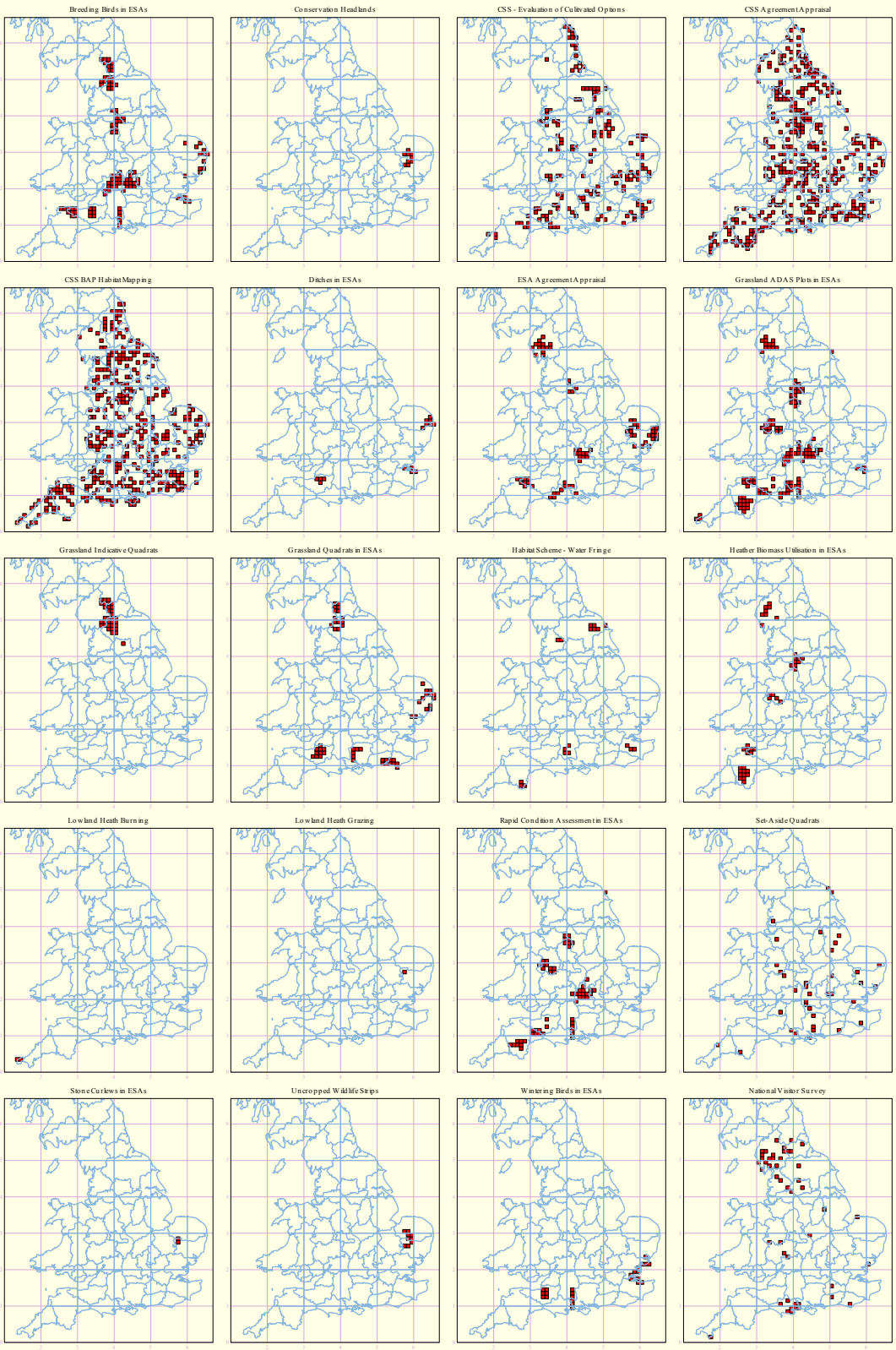

Fig. 1 (cont). The distributions of c) the remaining 20 agri-environment surveys

Scale/Co-incidence Relationships throughout England

To analyse the relationship between sampling unit scale and co-incidence of surveys, a geometric series of ten scales from 100km squares to hectares were used (Table 2). This series provided a 1,000-fold linear range of square size and, therefore, a 1,000,000-fold range in area.

The table also shows the number of sampling units at each scale that are wholly or partially enclosed by the boundary of England. These ten datasets were derived from a single data-set of 1km squares (monads) provided by the Analysis, Reporting & Modelling Team of Natural England. A small amount of data modification was necessary because approxi-mately 600 sites from the survey datasets had grid references that fell outside the monad dataset (largely inter-tidal) and several grid references were clearly erroneous. These corrections were not done systematically, but simply to "tidy" the maps and graphs.

Table 2. The geometric series of survey unit sizes used for the analysis.

Scale	*Linear Dimension (m)*	*Area (hectares)*	*No. in England*
100km	100,000	1,000,000	27
50km	50,000	250,000	83
20km	20,000	40,000	414
Hectad	10,000	10,000	1,496
5km	5,000	2,500	5,654
Tetrad	2,000	400	33,948
Monad	1,000	100	133,675
500m	500	25	534,700
200m	200	4	3,341,875
Hectare	100	1	13,367,500

The monad dataset comprised 133,675 records. To obtain the larger units, an algorithm was written in Mi-crosoft VBA to round up each unit to the next largest unit in the series. So, for example, the four monads SO7090, SO7190, SO7091 and SO7191 (the location of BioEcoSS Ltd.) would be rounded up to a single tetrad record SO7090 (SO79A using the unambiguous "DINTY" system). In coastal areas, this might result in only a single monad contributing to a tetrad record and, subsequently, even larger units. To create the smaller units, an algorithm was written that simply created all four 500m, 25 200m and 100 hectare units within each monad. This clearly resulted a small number of invalid units at the smaller scales where coastal monads covered areas below the MLWM. However, this proportion was well below 1% and not considered to be a serious source of error.

Analytical Method

The purpose of the analysis was to investigate the degree of co-incidence of surveys, based on their existing site locations, had different sized sampling units been used. To do this the grid reference of each site within each survey was recalculated using the same rounding up and down algorithms as the basic monad dataset. All grid references were provided to at least 100m resolution, which corresponded with the smallest theoretical sam-pling unit (the hectare). However, there was no GIS information in the form of shape files provided with the datasets and no record of which part of the site the grid reference represented. Furthermore, the national-scale surveys had site areas and whilst approximately 10% were smaller than 1ha, the majority were less than 10ha. So, in practice, the sites were treated as 1 ha in area with a grid reference representing their SW corners.

Frequency Distributions

The first stage of the analysis was to generate frequency distributions of the numbers of sites per sampling unit for the ten different scales. These showed a very strong pattern of a reduction in number of surveys within each sampling unit as the units became smaller. For example, when the sampling unit was a 100km square, none had fewer than six surveys represented (one square), and the maximum number of surveys per square was 31 (Fig. 2). However, with sampling units at the hectad scale, 11 (0.7%) of them contained no surveys and the maximum number of surveys per sampling unit had declined to 20. At the smallest scale (100m square) 98.9% of the sampling units had no surveys within them. (NB: the proportions are so extreme for the two smallest scales that they have been excluded from the frequency histograms in Fig. 2).

This pattern can be explored separately for the ten national-scale surveys and the 29 agri-environment surveys (Fig. 3). This clearly shows two different aspects of the difference between these two groups. Firstly, unsurprisingly, the national surveys are far more widely spread, causing a high degree of coincidence between them, especially at the larger sampling scales. Secondly, they are also more evenly spread. This is best seen in the agri-environment schemes, where some parts of the country are just not covered, resulting in even the largest sampling scales having some units with no surveys. Furthermore, no units at the largest scale had more than 21 of the 29 surveys within them, and with declining unit size, the maximum number of surveys decreased consistently. In contrast, the sampling units in the large surveys, continue to hold all ten surveys until the tetrad scale is reached.

These contrasting patterns are best seen in the box-and-whisker plots (Fig. 4) which display the maximum and minimum numbers of surveys more clearly.

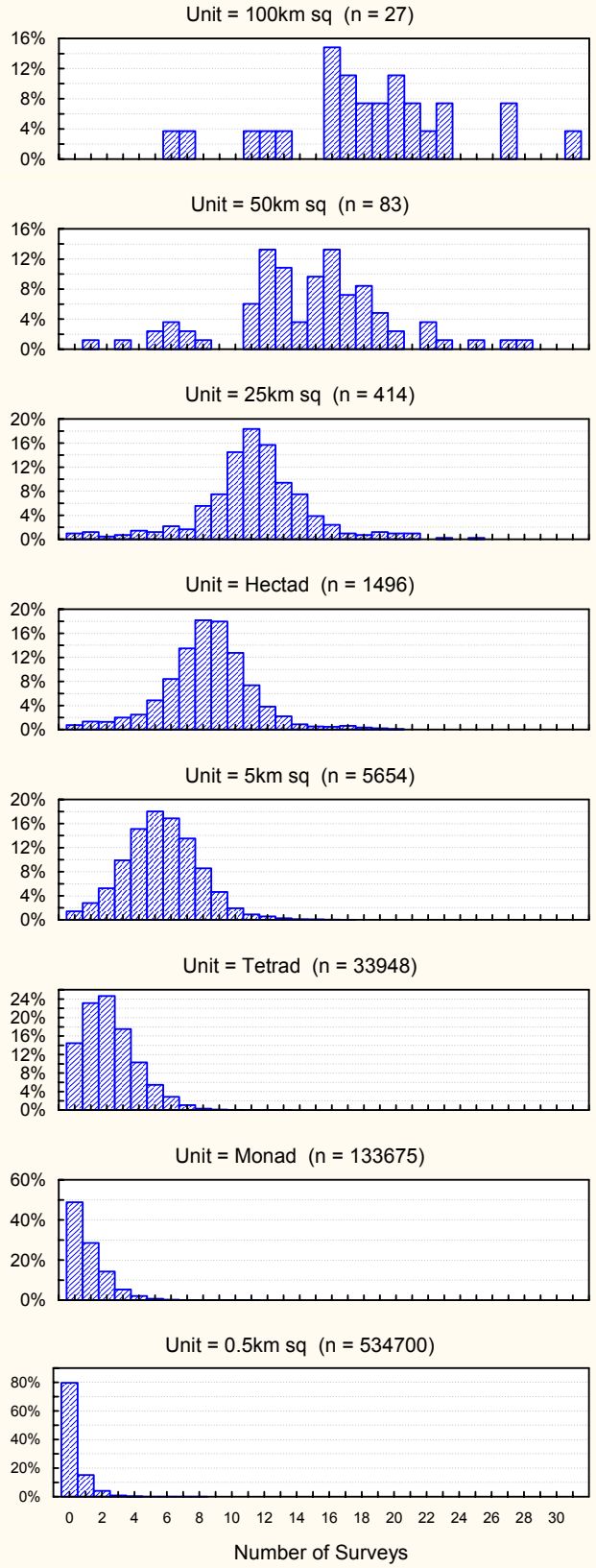

Fig. 2. Frequency histograms of the p of Surveys per sampling unit for eight different scales.

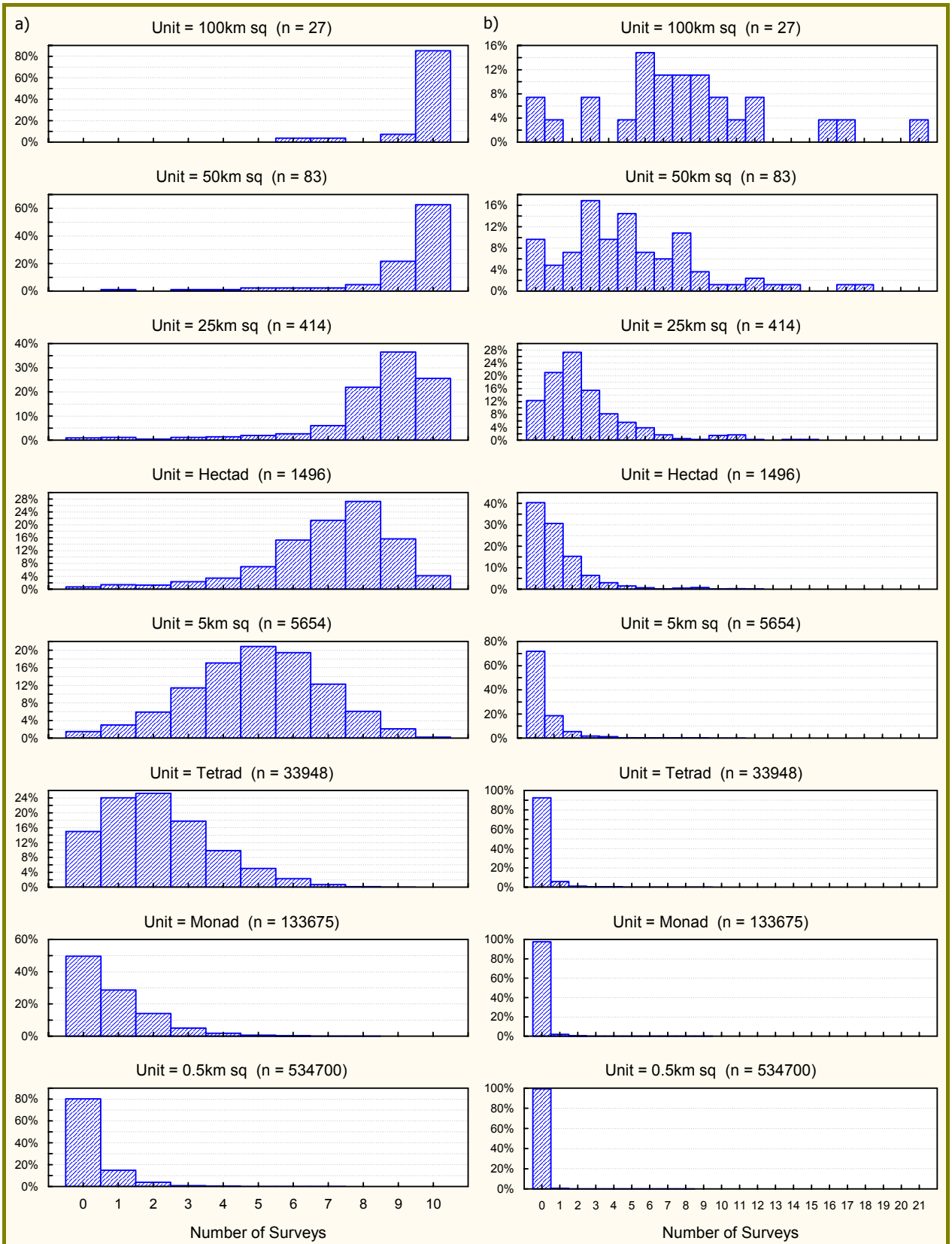

Fig. 3. Frequency histograms of the number of surveys per sampling unit for eight different scales separated into a) the ten national-scale surveys and b) the 29 agri-environment surveys.

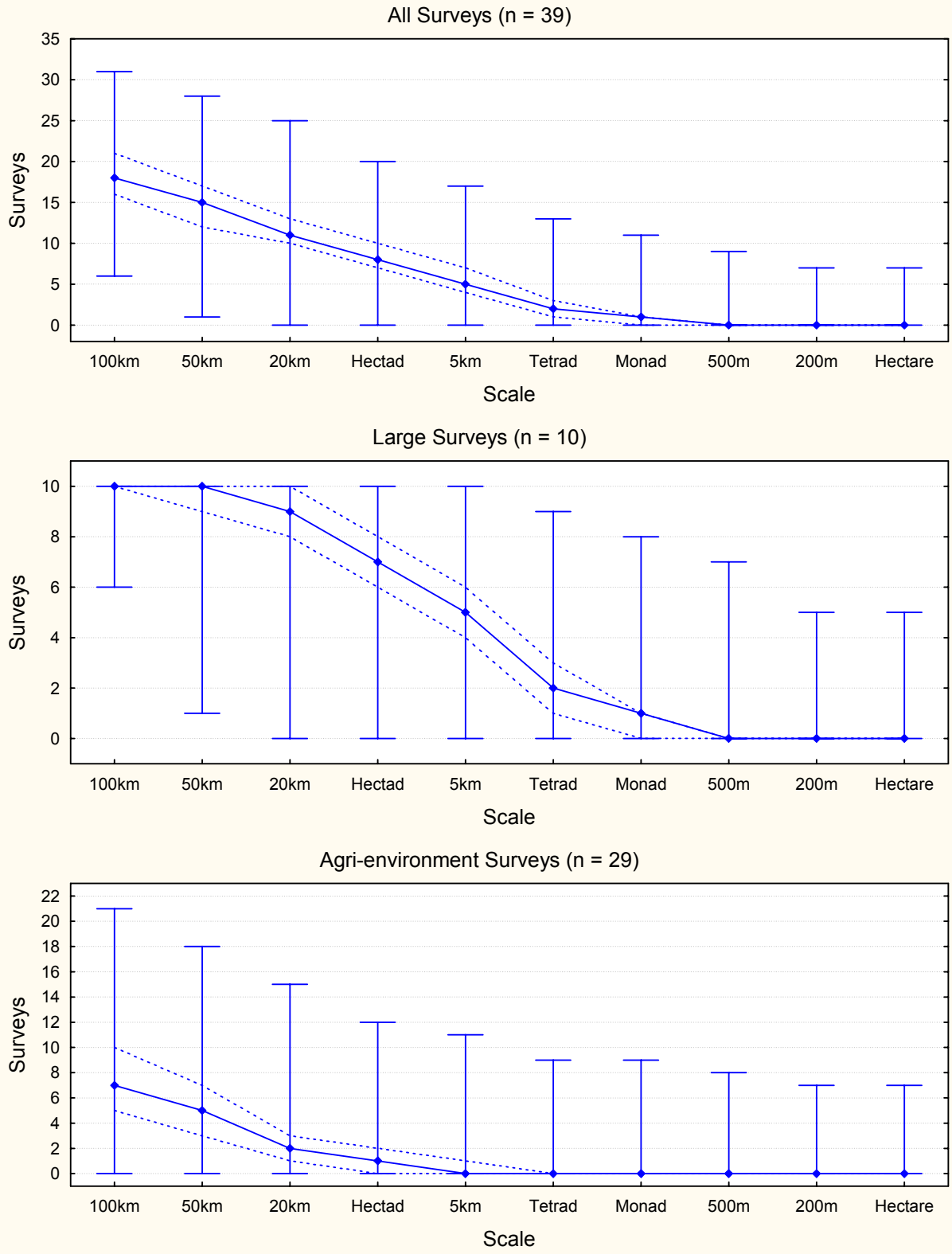

Fig. 4. Scale profiles for co-incidence of all surveys, large surveys only and agri-environment surveys. The solid line represents the median number of surveys per sampling unit at each scale. The dotted lines show the upper and lower quartiles and the whiskers show the maximum and minimum number of surveys per sampling unit.

Geographic Distributions

The clearest way to illustrate these patterns is by mapping them. Figures 5 & 6 show seven scales of sampling unit for national and agri-environment surveys respectively, from 100km to monad, representing a 10,000:1 ratio of areas. Within each sampling unit, the number of surveys is represented using a standard colour scale within the figure. This clearly allows the change in distributional pattern to be detected.

The widespread and even distribution of the ten national-scale surveys (Fig. 5) is shown by the huge dominance of the sampling units containing eight to ten surveys at the four largest scales. It is only at the 5km scale that this dominance declines, and one concentration of high co-incidence (Hampshire, Dorset and the New Forest) and one low co-incidence (South Lincolnshire and the Fens) emerge. The most striking thing about this scale, however, is the uniformity of pattern in the remaining parts of the country, with sampling units containing only two or three surveys juxtaposed alongside units with six or seven. This map is the spatial representation of the almost perfectly symmetrical histogram shown in Fig. 3a. At the tetrad scale, the first widespread, although still uncommon, appearance of empty sampling units occurs outside the previously mentioned Fens. Although the Hampshire / New Forest area is still clearly evident as a focus of high co-incidence, the only other zones where five or six surveys per sampling unit are found, appear to be located on distinct geographic features, such as the North and South Downs, the Cotswolds, the Peak District and the Brecklands. Finally, at the monad scale, the 50% of sampling units containing no surveys (Fig. 3a) can be seen to be fairly evenly distributed around the country. Outside the New Forest, very few monads contained more than two surveys.

In contrast to the national surveys, the agri-environment surveys (Fig. 6) show the full range of co-incidence at the largest (100km sq) scale. Admittedly, the sampling units with the smallest number of surveys occurred around the coast, simply reflecting the tiny areas covered by some of these large, artificial units. At the 50 and 20 km scales, the dominant pattern of these surveys emerged; the high co-incidence of the Arable Stewardship schemes in the West Midlands and East Anglia. Outside these two areas, however, the distribution of high and low co-incidence was fairly evenly spread. Furthermore, the 20km scale was the first to contain inland sampling units with no surveys. At the 5km scale, not only was the influence of the Arable Stewardship scheme very clear, but the individual ESAs became evident, with moderately high co-incidence in the upper Yorkshire Dales, the Peak District, Upper Thames Tributaries, Exmoor, Somerset Levels and the Test Valley. It is also important to emphasise that at this scale, the proportion of sampling units containing no surveys was nearly 75%. Finally, at the smallest mapable scales, the levels of co-incidence simply reflected the distribution of the various agri-environment schemes, so that at the monad scale 97.6% of sampling units were empty. Furthermore, the very few sampling units with two or more surveys (<1%) wesre focused in the Arable Stewardship areas and a few ESAs such as Dartmoor, Somerset Levels, Upper Thames Tributaries, the SW Peak District and the Yorkshire Dales.

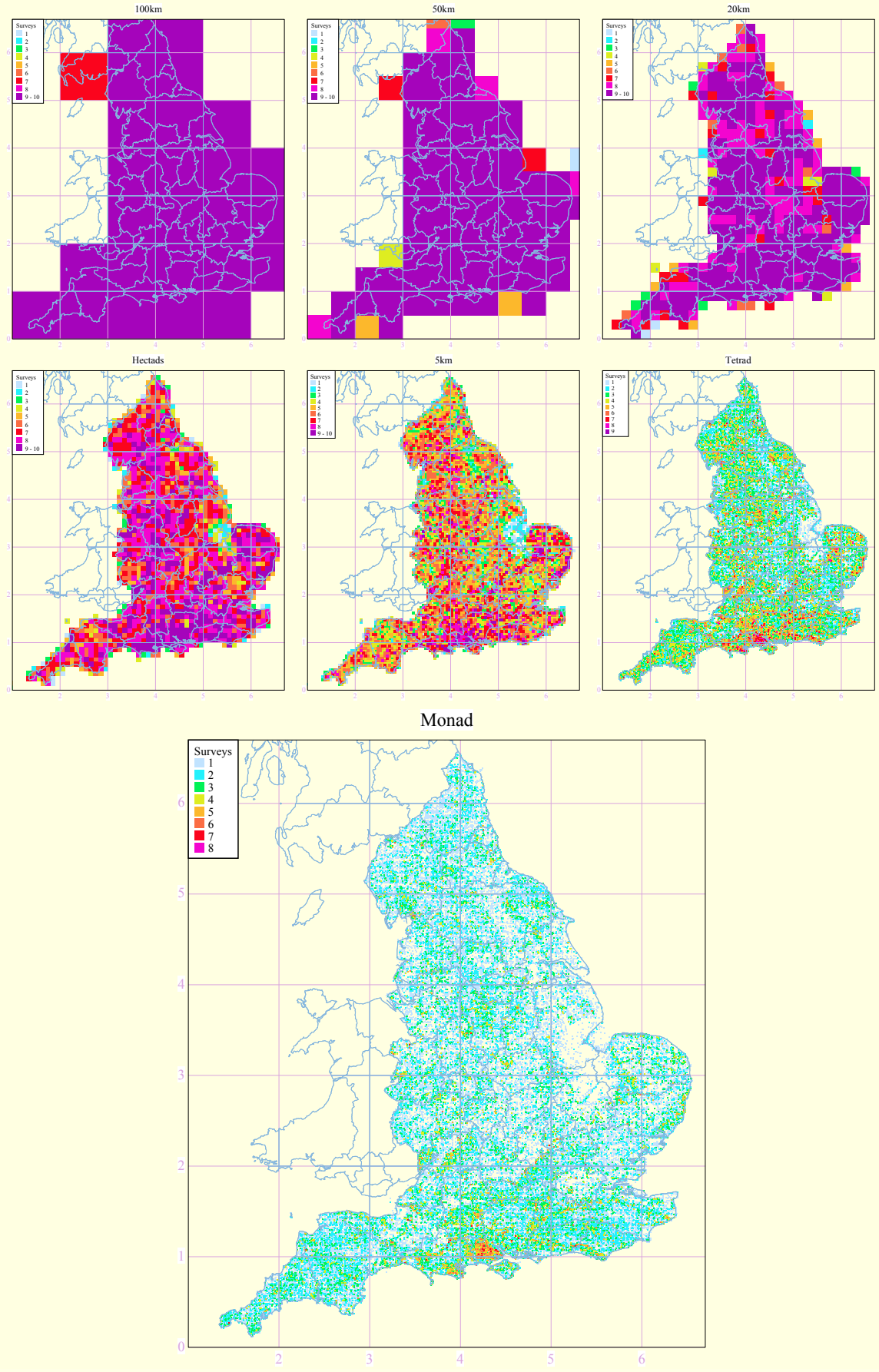

Fig. 5. Scale-based maps of all sampling units for England, ranging from a unit size of 100km sq. to 1km sq (monad). The numbers of national surveys in each sampling unit are colour coded from 1 to 10.

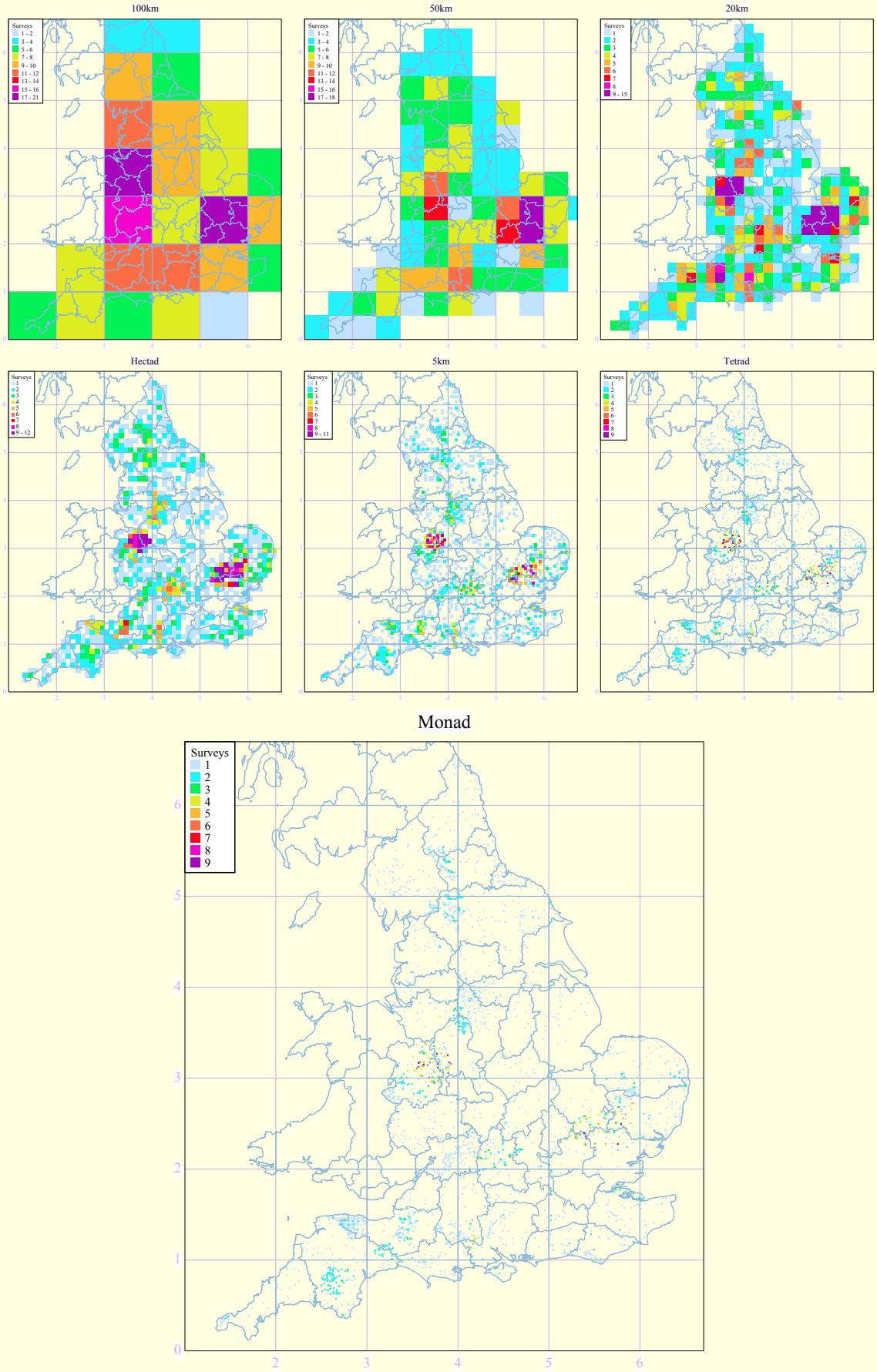

Fig. 6. Scale-based maps of all sampling units for England, ranging from a unit size of 100km sq. to 1km sq (monad). The numbers of agri-environment surveys in each sampling unit are colour coded from 1 to 10..

Scale/Co-incidence Relationships for ECN/ECBN Sites

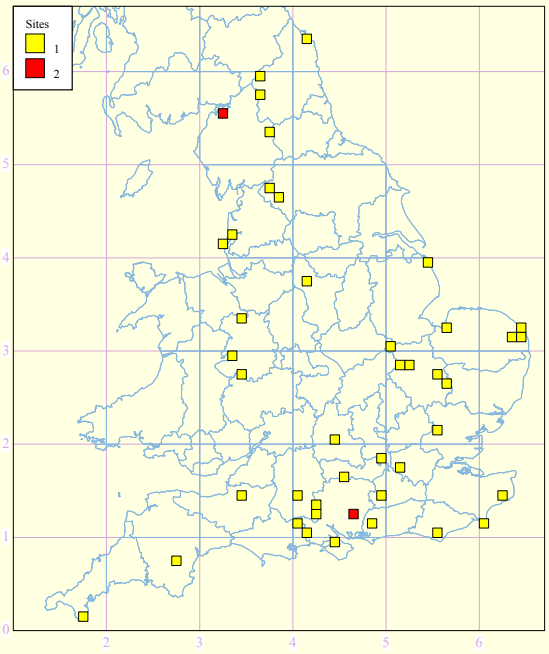

Fig. 7. The location of ECN, ECBN & proposed ECBN sites in England. Sites are shown as 10km grid squares for clarity, with two squares each holding two sites (Finglandrigg Woods & South Solway Mosses in Cumbria and Old Winchester Hill and Beacon Hill in Hampshire).

Currently, there are three ECN sites and 16 ECBN sites in England. In addition, a further 26 ECBN sites are proposed for inclusion in this network (Fig. 7). As these sites did not themselves constitute surveys, they were not included in the primary analysis. Rather, they have been used as a filter to answer the question "If we just used the ECBN network, how would the sampling unit size influence the number of existing surveys within each site?"

To answer this question, the same process was carried out as in the primary analysis using a range of scales from 100km square to monad. In this case, the sampling units were included if they contained one or more of the ECBN sites. This resulted in a slightly reduced sample size at the largest scale (17 100km squares as compared to 27 for the primary analysis), which increased slowly to reach the full 45 sites at the tetrad scale, when the two closest sites in Hampshire resolved into separate sampling units (Fig. 7).

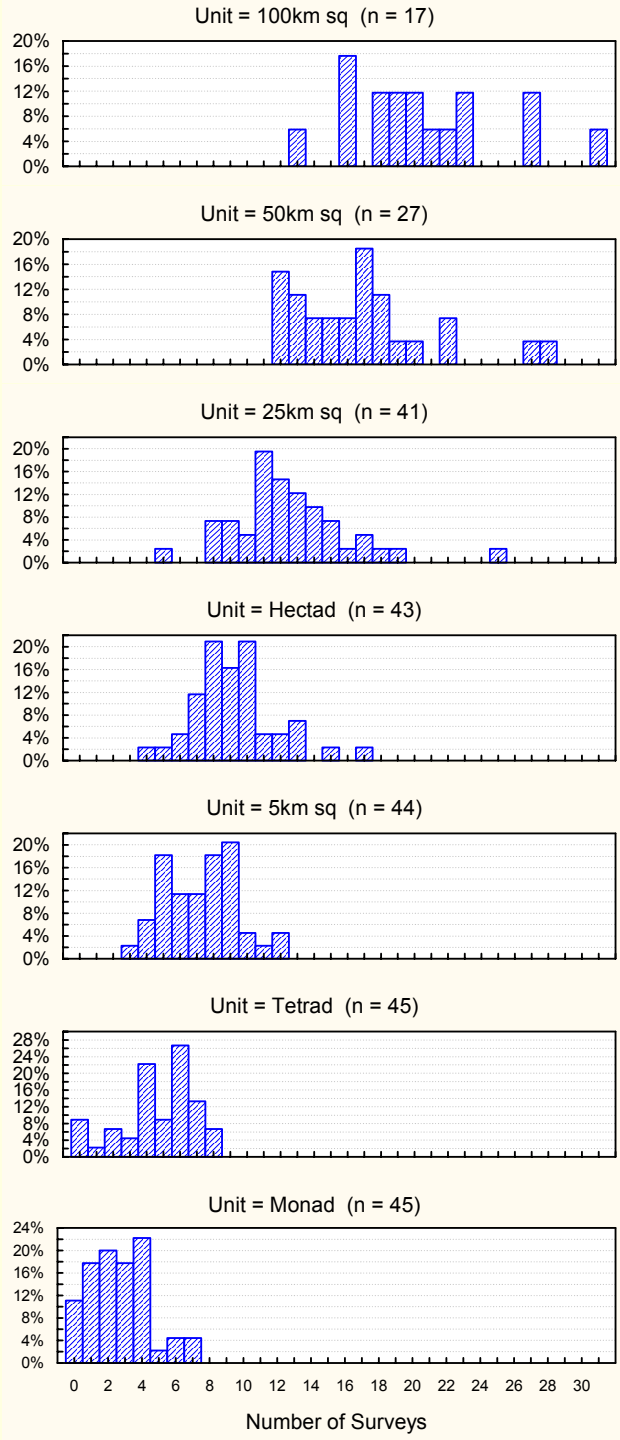

Fig. 8. Frequency histograms of the p of Surveys per sampling unit for seven different scales.

The frequency histograms of the number of surveys per sampling unit were based on all 39 surveys (Fig. 8). This showed that at the largest scale (100km sq), the square containing the 31 surveys found in the primary analysis (SU) was included within this sample. However, the most important aspect of these graphs is that the ECBN network appears to have been well located, because it is only at the tetrad or monad scales that some sampling units (four and five respectively) did not contain existing surveys sites. Median numbers of surveys per sampling unit at the hectad, tetrad and monad scales were nine, five and four respectively, which was

greater than the equivalent medians from the primary analysis (see Fig. 4).

The geographic distribution of these co-incidences is shown in Fig. 9. Apart from the dearth of ECBN sites in NE England, it can be seen that the range of co-incidences was well distributed around the country. However,

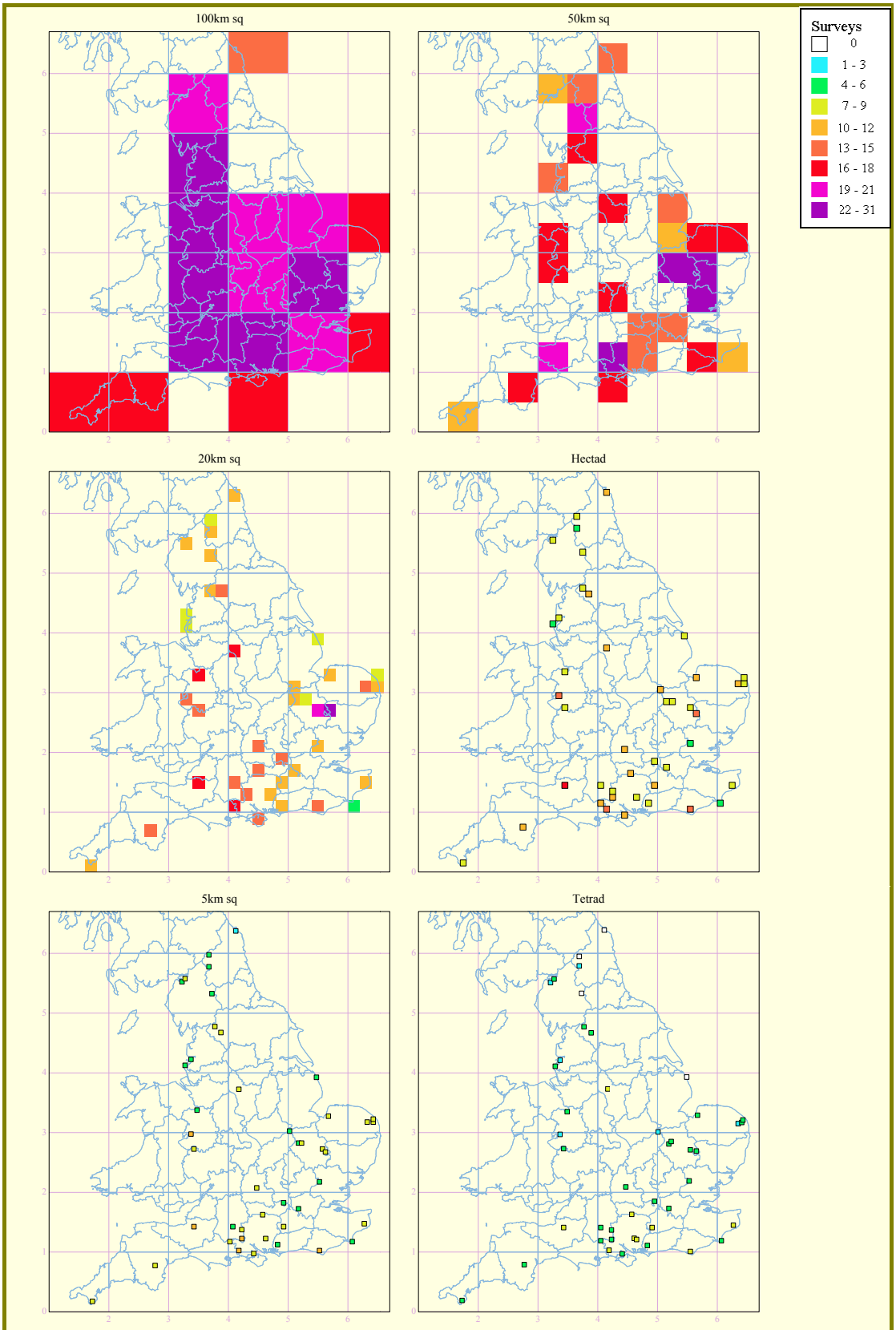

Fig. 9. The degree of co-incidence of surveys in the 45 ECBN network at six different scales. The maps for the two smallest scales were generated using the specified scales, but were plotted using 7.5 km squares for clarity.

at the hectad scale there was a tendency for the richer sites to be found in the southern half of the country, with all of the five sampling units that contained more than twelve surveys. At the tetrad scale, this was manifest with the emergence of the four sampling units that contained no surveys and five of the six that contained three or fewer surveys being in the northern half of the country.