Alternatives for the definition of housing market areas
Geography of housing market areas in England – paper B

Alternatives for the definition of housing market areas

Geography of housing market areas in England – paper B

Mike Coombes, Colin Wymer
Centre for Urban and Regional Development Studies, University of Newcastle

1: Introduction

In this stage of the research, numerous potentially promising 'candidate' methods for defining housing market areas are considered. These methodological approaches were agreed with NHPAU following several meetings with the regional steering group and the expert panel.

Initially the research followed the strategy outlined in the Stage 2 December update but subsequently it was decided to also assess approaches leading to two tier definitions of housing market areas. As a result, this report should be read in conjunction with not only the Stage 2 December update but also an accompanying paper outlining the case for two tiers of housing market areas.

The primary task whose results are summarised in this report was to evaluate the different approaches by producing housing market area boundaries defined from the application of each approach. The evaluation then seeks in a systematic and open fashion to assess the suitability of each set of housing market areas on criteria which are technocratic (i.e. they neither revisit the theoretical discussions that preceded this stage of the research, nor do they duplicate the practical policy assessment in stage 3).

The research presented here has been an iterative process, but the definitions of housing market areas here are based purely on analyses of commuting and/or migration patterns. It was anticipated that these definitions would also be informed by testing whether pairs of adjacent housing market areas had statistically different standardised house prices (using hedonic price functions). This test was not as powerful as hoped and instead the hedonic analysis has been applied to assess the geographies as a whole. That research was undertaken by Heriot-Watt and is not presented here.

2: Commuting and migration research for the definition of housing market areas

In the Stage 2 research CURDS has developed several candidate methods for defining housing market areas. It should be noted at the outset that no prior research has devised a method of housing market area definition that is both theoretically valid and been successfully applied in widely varying areas. Given this limited foundation, Stage 2 of the research has had to include considerable methodological development work. In practice, the hedonics modelling was split off from other housing market area definition research: the Heriot-Watt team has done the hedonics with the rest done at CURDS: this report only covers the elements of the research done by CURDS (and as such provides sets of areas that could then be input to the hedonic modelling).
The various methods developed have also been evaluated. The evaluation was based on three broad types of consideration:

- technocratic as assessed by CURDS
- theoretical based partly on the views of the Expert Panel, NHPAU, and the research team
- use value as envisaged by NHPAU plus colleagues (nb. Stage 3 focuses on the policy context)

A series of interim reports (from December 2009 onwards) summarise the different approaches and their results. Colleagues from across the research teams plus NHPAU participated in the evaluation, along with members of the Steering Group and the Expert Panel at several meetings. The following points very broadly summarise elements of the progress towards a consensus:

1   There was substantial interest in the variety of methods pursued, and results they produced.
2   The theoretical debates do not provide a 'knock-out' result in favour of one form of definition.
3   Differences of opinion, not only between but also within each set of respondents, limited the guidance required to distil a definite set of guide-lines for the analytical research.
4   There was probably more support for the housing market area boundaries to be based on analysing the patterns of commuting flows, though there were no strong objections to analysing migration.
5   There were no strong objections to the innovative approach of using more than one dataset (which has been the principal departure from past practice).
6   The use of the hedonics analysis as a separate and subsequent stage in the housing market area definition was widely accepted in principle, with its eventual exclusion from the definition process accepted purely on the empirical grounds that it did not contribute in the way anticipated.
7   There was little support for narrowing down the analyses by, in particular, looking only at the migration of owner-occupiers rather than of all moving group representative persons (MGRPs) and/or analysing only the commuting patterns of high-earners (rather than all the employed).
8   Although there was no definite consensus on whether a 'unitary' or two tier geography is preferable, it was a two tier option that emerged as the NHPAU preference following Stage 3 of the research.

3: Strategic housing market assessments

Steering Group members were asked to supply definitions of the areas used for strategic housing market assessments in their regions and in time all these arrived in CURDS and were collated to create a national set. (It should be noted that at the time of writing three regions have yet to confirm that their inputs have been correctly interpreted by the CURDS mapping process, so some of the following discussion may need to be changed later.) The only strategic housing market assessment areas that split local authorityies are in the following cases:

- the boundaries in the North East group some former Shire Districts in Northumberland and County Durham into different strategic housing market assessment areas, and thereby split these new unitary county areas
- strategic housing market assessment areas in the South West have a similar effect on newly unitary Cornwall and Wiltshire
- in the North West there is a similar effect in the new unitary area of East Cheshire but there is also considerable splitting of the (continuing) Shire Districts of Cumbria in order to create strategic housing market assessment areas based on groups of wards (e.g. splitting South Lakeland into three strategic housing market assessment areas, and separating out Central Manchester from the rest of the city and conurbation)

- in the South East the Urban South Hampshire partnership (PUSH) includes the whole of five local authorities but also parts of four more: this boundary has been approximated by grouping the wards (used for 2001 Census data) that are a best-fit to a map on the internet (from PUSH).

It was left to Steering Group members to decide which set of areas to include in this collated map. There were decisions to be made for several regions, with different 'tiers' of boundaries potentially relevant to strategic housing market assessments in the West Midlands and the South East (possibly also in London and Yorkshire and the Humber as well). For the East Midlands – and potentially the East of England – there was a decision over whether to separately represent "Peterborough Partial" strategic housing market assessment area, in the former region, or merge it with the Peterborough strategic housing market assessment area in the latter region. In the end, the boundaries of every region are respected by these strategic housing market assessment area boundaries. This approach even extends to the London region, despite its boundary being crossed by very many commuting and migration flows that are seen to characterise the structure within a housing market area in most understandings of housing market areas and their definition. Map A1.2 shows the strategic housing market assessment area boundaries on the basis described above. [The numbering of maps here is done on the following basis: the first (alpha) character simply proceeds from A through the alphabet, the numbers following preserve previous usage for the sake of continuity.]

Table 1 shows that the number of strategic housing market assessment areas in each region varies very widely in ways which are not simply explained by geographical differences between the territories that each region covers. One possible implication is that the North West was far from being a 'representative' region for the case study work earlier in the research, because it includes a larger share of all strategic housing market assessment areas (23%) than its share of the national population or land area.

It is notable that the number of strategic housing market assessment areas (118) is close to the middle of the range of the number of housing market areas in the sets produced by the research, which were between 80 and 135 in most cases. Given the variability between regions in their strategic housing market assessment area sizes, it may be that a consistently defined set of around 135 housing market areas will include areas broadly similar to the strategic housing market assessment areas in regions like the North West where strategic housing market assessment areas are small, whereas a consistently defined set of around 80 housing market areas will include areas broadly similar to the strategic housing market assessment areas in regions like the West Midlands where the strategic housing market assessment areas are large.

Table 1 Strategic housing market assessment areas (as supplied) per region

Region	number of strategic housing market assessment areas
East of England	13
East Midlands	11
London	5
North East	4
North West	27
South East	23
South West	14
West Midlands	4
Yorkshire and the Humber	17
England (total)	118

In general, the maps of boundaries from the research analyses below allow the comparison of their proto-housing market area boundaries with the current strategic housing market assessments, because the strategic housing market assessments are shown as a coloured background to the boundaries.

4: Definition method developments

Figure 1 shows as a flow diagram six alternative approaches to the definition of housing market areas which the CURDS team has developed and assessed: these are discussed in more detail below. As explained above, all the four approaches could involve a 'test' by hedonics analyses as a final step. That element was not carried out by the CURDS team but, in the event, it has not been implemented. This is because experiments showed that it did not prove useful as part of the process of housing market area definition; instead the hedonics are now part of a subsequent evaluation stage of the research, and so are reported elsewhere.

It is necessary to first summarise the six different approaches. Among methods for defining housing market area boundaries using migration flow analyses, two have passed the peer review 'test' necessary to merit publication in academic journals.

- Jones (2002) developed a method that was suitable for the Scottish case which relies on identifying key urban centres; this may be less successful in parts of England with numerous urban areas of similar size in close proximity to each other, but it is further developed and tested here and termed Approach 1.

- Coombes (2009) reports that a method which did devise appropriate housing market areas for policy in the north-east of England produces less immediately useful results in more rural and southern areas where migrant flows are longer and more dispersed; this is further developed and tested here and termed Approach 2.

Note that although these are existing methods, they still posed methodological challenges. There has been development work covering some relatively technical issues, including some adjustments to the key criterion of the required level of self-containment of migration flows.

With the other approaches attention centres, at least partially, to data on commuting.

- Approach 4 is effectively the same as Approach 2 except with the commuting flow data replacing the data on migration flows. This method uses the commuting flow analysis that Coombes & Bond (2007) applied to define the government's official labour markets called travel-to-work areas.

- Approach 3 essentially takes the same approach but with a lower self-containment minimum for the commuting analysis so the results from that step still have 'in play' rather more separate areas than will ultimately be required: a step 2 then retrieves the migration data for these areas and effectively applies the third step of the first approach to these areas (thus some 'proto'-housing market areas defined by reference to high earner commuting patterns will be further grouped – and possibly split up in the process – if they are not also self-contained in terms of their migration patterns).

- Approach 5 is the same as Approach 3 except (a) step 1 uses migration data and step 2 uses the commuting data, and also (b) the areas from step 1 must be grouped by step 2 without splitting them (nb. this is called a 'hierarchical' grouping process): this creates a two tier set of results, with the lower tier areas coming from step1 and the upper tier from step 2.

- Approach 6 uses the datasets in the same order as Approach 3 (i.e. Step 1 uses commuting data and step 2 the migration data) but step 2 groups ward building-blocks and these groupings cannot span across the areas produced by step 1: thus in the two tier set of results this creates, the lower tier areas come from step 2 and the upper tier from step 1.

One strategy often found in earlier definitions of sub-regions such as labour market areas is the initial identification of centres which is followed by finding the hinterlands of those centres. This assumes that the patterns of interest are dominated by the catchment areas around single centres, but this model has become more questionable as people travel further and their flow patterns criss-cross each other within polycentric regions. It is also a model that was never very relevant to migration patterns, even if the commuting patterns in some areas used to have this structure. In fact, the model is not implementable here because there is no set of wards identified as centres. In response, the travel-to-work area methodology was used because it has become an internationally-recognised standard for labour market area definition. Moreover this methodology effectively identifies clusters of flows, and this does in practice find all major central places and the pattern of linked areas around them.

5: Evaluation strategy

Different sets of housing market area boundary definitions are produced by varying definitional criteria. For example, minimum size thresholds for the housing market areas can be varied to see the consequences. The research also looked at differences between the results gained from applying the same method to different datasets.
The results produced by different housing market area definition methods will be assessed against evaluation criteria of which some are rather technocratic. The more technical criteria include:

- contiguity (i.e. each housing market areas to be one coherent 'territory' and thus not split up)
- similar size (i.e. the set of housing market areas to be made up of areas comparable to one other)
- robustness (i.e. in most cases, a minor change to the criteria does not greatly alter the housing market area boundary definitions)

More broadly, there is a presumed preference for methods that are simpler because their results may be more widely acceptable than the results of a more 'black box' form of method.

The crucial criterion to determine which set of results will be recommended at the end of the project is their 'use value' to NHPAU but this is not readily definable as a set of characteristics which can be used to evaluate one set of results against others. The clearest parameter is number and the fact that the objective here is to identify sub-regional housing market areas suggests that England should be divided into between 50 and 100 separate housing market areas (slightly fewer than the number of strategic housing market assessments which resulted from processes that tended to favour recognising individual local authorities as separate housing market areas regardless of evidence from objective data analysis). That said, objective data analyses such as those reported below do not easily lead to identifying an ideal number of housing market areas to define. Turning to specifics, the analyses require minima to be set (most especially in terms of the level of self-containment of flows) and altering these parameters leads to different numbers of housing market areas being defined.

In reality, there is a judgement involved in selecting from different sets of possible housing market areas. Even two sets of boundaries with a very similar number of areas nationally can have very different patterns, with one having a very large single areas extending around London but quite small areas in the old industrial regions – as illustrated by the housing market areas in Coombes (2009) – and another having much more similar size areas in all regions (the pattern which tends to be found with travel-to-work areas). Both these maps are representing one aspect of reality so it would be for the NHPAU to identify which would be more appropriate for its purposes. It is this 'use value' which will determine the final choice of housing market areas from this research, with the more technocratic criteria narrowing down the choice to a short-list which will all have were defined on a basis which is seen to be theoretically sound.

The splitting of local authorities is to be expected because the analyses use wards to produce more precise boundaries than would be possible with groupings of whole local authorities (especially with the large newer unitary local authorities in some former Shires). It is crucial to the robustness of the results that the initial 'gold standard' definitions are not constrained to be groupings of whole local authorities. That said, the research specification does require that – to support the NHPAU's plans for modelling – there is subsequently a 'silver standard' version of the final housing market area geography which is best-fit to the housing market area boundaries from whole local authorities.

The boundaries produced are not always recognisably orientated around one single larger urban area. Most of the sets of boundaries defined here include few non-contiguities (e.g. an area grouped into proto-housing market area *x* but completely surrounded by proto-housing market area *y*). There is in fact an option within the definition method termed an 'optimiser' which reduces the amount of fragmentation in the boundaries (without undermining the areas' meeting the self-containment test). The disadvantages of the optimiser are that it introduces a 'black box' element to the method, and also tends to reduce the size of areas which include large urban areas (most notably, substantially

reducing the size of the London area that the basic analyses had deemed to be appropriate).

Another option in the definition method prevents areas spanning the borders of England (with either Wales or Scotland). It was found that in practice few areas are affected, so the border constraint has only been applied in one of the analyses reported below.

6: Results from each approach

Table 2 provides a very brief summary of the results from each of the analyses selected for more detailed evaluation; the analyses are listed in sections ordered by the number of the approach they use. Table 2 reports for each of the selected analyses the value of the key definition criterion, which is the minimum self-containment (i.e. the proportion of the flows analysed which both start and end within the same area). The boundary definitions method seeks to identify as many as possible separate areas which meet this criterion, grouping the building-block areas in whatever way minimises the number of flows that cross them. There are more long-distance commuters among the high-earners than there are among other workers, so fewer separate areas can be defined when analysing the high-earner group than with the total set of commuting flows used for travel-to-work areas.

To proceed from more familiar analyses to the more innovative, the descriptions of the results from the first four approaches will now be described in reverse order; then the fifth and sixth approaches that are able to define two tiers of housing market areas are discussed at the end.

Approach 4: Labour market areas

Although the Stage 1 report emphasised housing economics theory that led to hedonics analyses of house price data as a core element of housing market area definitions, there are other possible approaches to consider. In the guidance (CLG 2007) there was recognition of the potential role of commuting flows, although it was ambiguous as to whether the suggestion was simply to focus on the official definitions of labour market areas – travel-to-work areas – produced by Coombes & Bond (2007), or to analyse commuting flows anew. This issue was resolved by linking the travel-to-work areas to some new analyses, and also by some analyses looking only at commuting by higher-earners whose flows are a relatively small proportion of the total dataset. The rationale for this latter form of analysis is that demand for housing is a function of income and there will be vertical substitution within housing market areas as more skilled workers compete for housing over the relatively long distance they can afford to commute.

Table 2 shows the number of separable areas increasing with reductions to the key definition criterion viz: the minimum self-containment level that each area must satisfy. All the Approach 4 analyses have tended to produce areas that perform well against the technocratic criteria (principally that the areas are rather similar in size, and that there are not many non-contiguities). Map B9.1 shows the results from applying the travel-to-work area self-containment criterion to the higher-earning group: these are people whose occupations are in the upper socio-economic classes. At the same time, Map C8.2 shows that somewhat similar areas are produced when analysing data on all commuters if the self-containment criterion is raised to 75.0% from the 66.7% of Map B9.1 (and the travel-to-work areas).

The fundamental question left by the analysis presented so far is whether the definition of housing market areas should essentially depend on the analysis of commuting data. Is it appropriate that the boundaries are neither more nor less than a set of labour market areas? Setting that question aside, it does seem that housing market areas defined with commuting data can readily meet the technocratic criteria, leaving open only the question of which self-containment level produces the most appropriate set of boundaries so far as their number and patterning is concerned.

Approach 3: Labour Market Areas subject to migration flow analysis

At this point the research moves from extending previous analyses to developing an entirely new form of analysis. These analyses start with analyses of commuters – as in Approach 4 – but the innovation here is to input the results from those analyses into a further grouping process that uses migration flows. In this way step 1 of Approach 3 is identical to Approach 4 but then step 2 in Approach 3 also takes account of migration patterns. Thus it requires the proto-housing market areas to be relatively self-contained as both labour market areas and also in terms of migration flows. A key decision remains the choice of just how self-contained the areas must be, and here the travel-to-work areas parameters were again the 'default' initial option. Along with the option of analysing only high-earner commuters, some analyses restricted the migration data to owner-occupiers (see below for reference to analyses of other tenure groups).

Map D4.2 illustrates that when the self-containment minima are set to produce around the preferred number of separable proto-housing market areas (79), these analyses produce areas that perform relatively well on the technocratic criteria. The basic method used in Approach 3 maintains one of the key elements of the travel-to-work areas approach because it has not been hierarchical. What this means is that step1 labour market areas which do not pass the migration flow self-containment test are broken down into their constituent wards and these are then re-grouped on an individual basis so the final set of boundaries are as optimal as possible. The alternative approach is to group areas hierarchically, with the result that the final boundaries produced are combinations of whole labour market areas. Table 2 shows that this variant has been explored too: the difference it produces tends not to be very great and so the results do not need to be mapped here. Table 2 shows that several other alternative variations have been evaluated too.

Analyses which take account of the patterns of both migration and commuting flows have to deal with the fact that the two sets of flows have rather different spatial structures. Commuting patterns produce rather similarly sized labour market areas across the country, because most people in any areas avoid spending very much time commuting on a daily basis and so minimise their commuting distance. At the same time, there is a minority of people who can afford longer-distance commuting and most of these people live and work around the largest urban areas, so labour market areas tend to be larger in these parts of the country than in more rural areas. By contrast, migration is a very much more occasional movement and so distance deterrence does not operate to the same degree in minimising distance moved in any part of the country. In practice, it is found that housing market areas defined using migration data tend to be larger in rural England than in more urban areas, as well as being more localised in the north than around the capital.

The consequence for Approach 3 is that the step 1 analyses of commuting data produce larger areas near the larger cities, with the analyses of migration data in step 2 then doing most of its grouping in the more rural parts of the country. The net effect is to produce a set of boundaries that satisfy the technocratic criteria although, as with any specific set of results, there are some minor idiosyncrasies which mostly involve the non-continuous grouping of relatively small numbers of wards (Map D4.2).

Approach 2: Migration flow analysis

The method used here is in effect the same as that used in approach 4 but here it is the migration dataset that is analysed. Bearing in mind the focus on the owner occupied sector of NHPAU's remit, it was important that the migration flows could be disaggregated by tenure. One hurdle to be overcome before this benefit could be gained was that the data source for migration data (the Census) does not break down the published data both by tenure and age or whether the person was a student. As a result, it was essential to order a customised dataset. This also allowed the exclusion of people aged under 25: this was a great benefit because the numerous lengthy moves of students are irrelevant to the research. Tenure is not known for all migrants, only for moving group reference persons (MGRPs) so they were the subject of the dataset obtained. (It should be noted here that the definition of moving group reference persons includes many people who are not heads of households: for example, a 25 year old returning to the parental home will be a single person moving group, and if the parental home is owner-occupied then this 25 year old will be recorded as an owner-occupying moving group reference person because the same tenure characteristic applies to all household members.)

Map E2.1 applies the travel-to-work areas parameters to the dataset covering all 25(+) moving group reference persons. The results are notably similar to that in Coombes (2009). In particular, there is a very considerable difference in size between areas in the south and those in the old industrial regions. In some parts of the country there are also quite a large number of non-contiguities: to some extent this problem could be resolved by using the boundary optimiser as part of the analysis, but the need for it erodes confidence in the results and – as has already been said – its 'black box' nature is unhelpful in a policy context, and it has sweeping impacts on large proto-housing market areas such as London. In some parts of the country the results seem highly plausible but the fact remains that this single set of areas combines a detailed breakdown of the north with some very large areas in the south, and this is not ideal for policy purposes.

Map F2.2 uses the same travel-to-work area criteria but applies them to the tenure-specific data by putting the focus on moving group reference persons who are in owner-occupied housing after their move. With the owner-occupying sector dominating English housing it is unsurprising to find that there is relatively little difference between the results of the all-tenure analysis and those from data on owner-occupying moving group reference persons. Some of the areas are notably distended. One interpretation would emphasise transport corridors and highlight the example of a tangential grouping of wards along the M10 from London to Cambridge: whether or not this is plausible, these boundaries are not exempt the problem with the all-tenure results which is that their size varies radically across the country in a way which is neither entirely recognisable nor helpful for policy purposes.

Map G2.3 shifts attention to social renting moving group reference persons. There is a dramatically high level of non-contiguity that might be better described as fragmentation. The key reason for this outcome is the lack of social rented housing in many wards, especially in more rural areas. Where the basic pattern of results does emerge from the fragmentation, it seems clear that across much of the country there are more separable social renting housing market areas than owner-occupying areas (Map F2.2), and that local authority boundaries are quite often 'driving' the housing market area boundaries of social renting moving group reference persons.

The final part of this strand of the research shifts focus again; this time to 'other moving group reference persons' (Map H2.4). This group includes private renters and will capture a high proportion of older students. It becomes immediately clear that the proportion of

longer-distance migrants is higher for this group than for either of the other sets of moving group reference persons, because the areas are so much larger. The whole south-eastern part of the country is divided between just four areas, the same number as there are for the old mill towns from Blackburn to Huddersfield either side of the Pennines. Although the possibility remains of changing the self-containment minimum to define numerous more areas, it is evident that analysing this dataset with this method produces fragmented boundaries, and areas of very uneven size across the country.

Approach 1: migration flow analysis centred on urban areas

The final method to be summarised here is based on that in Jones (2002): this pre-structures the analysis of migration data with an identification of urban areas which then act as the 'seeds' around which each proto-housing market area is defined. The result will be that, by their very nature, the 'seeds' enforce some groupings that the data on real behaviour might not support. The rationale for 'seeds' is that some theories assume that no single urban area can embrace multiple housing market areas, but is this a good enough reason to ignore any evidence which may show that there are separable market areas that *do* abut each other within a single conurbation?

Two key issues arose when implementing this approach here. It was first decided to ignore smaller urban areas to avoid excessive fragmentation, while still ensuring that any plausible candidate area to be a sub-regional housing market area has a 'seed' for its eventual definition (e.g. if Berwick were not a 'seed' then no separate housing market area for north Northumberland can be defined). A minimum size of 10,000 for 'seed' urban areas was used.

The other issue is that if very large urban areas like London are taken as single 'seeds' then the analyses cannot test for the possibility that the East End (say) is in a separate proto-housing market area to the rest of the conurbation. The response here is to provide an alternative version of the 'seeds' that treats as separate those (part) local authorities in the same urban area (.eg. London boroughs are separate 'seeds' in the first instance). Whilst this prevents unnecessarily large initial 'seeds' which are not tested against relevant data, there will be greater importance placed on the later parts of the analyses that carry out groupings until all the proto-housing market areas meet the level of self-containment set as the minimum requirement. This alternative is illustrated later, but next the whole urban area version of the 'seeds' is utilised.

Map I1.3 shows the results from priming the analyses with a set 'seeds' which are whole urban areas with populations of 10,000 or more. As with the results immediately above, the travel-to-work area parameters are used to set the minimum level of self-containment that each area must meet. This means that the results here are comparable to those from the equivalent analysis without 'seeds' (Map E2.1). The comparison tends to be favourable to the 'seeded' approach, with fewer non-contiguities and more coherently shaped areas (eg. London). As noted earlier, this neatness of boundaries may be an artefact of using 'seeds' and it may be preventing the results reflecting the reality – which may be less neat – of actual migration patterns.
Map J1.4 shows the results of the analysis in which the 'seeds' are urban areas which have been subdivided along the boundaries of local authorities, where several local authorities are within the same urban area (as in most conurbations). Here it is found that the East End does prove to have sufficiently distinct migration flow patterns from those of central London for it to be grouped with Essex. At the same time, the reduced constraint upon the analyses seems to produce more non-contiguities as well as more distended areas, as with the grouping of Oxford and Northampton to the north-west of London.

Approach 5: Lower tier areas based on migration; upper tier grouping based on commuting

Approach 3 was an entirely new form of analysis that used a version of the commuting analysis to create an initial set of areas which were effectively 'seeds' for a subsequent analysis of migration flows. An area which was not sufficiently self-contained with respect to migration flows was then disaggregated and its constituent wards subjected to a further grouping process based on the migration flow patterns. It thus took labour market areas and effectively 'tested' whether they were self-contained in terms of migration. The approach had considerable appeal in taking account of multiple sources of relevant evidence while providing a single tier set of results. Approach 5 takes forward the strategy of using two datasets but does so in order to provide two tiers of boundaries.

Here in Approach 5 the lower tier boundaries are produced by step1 from analyses of migration flows. These areas are then subjected to further grouping, using the commuting flow data, to produce the upper tier boundaries. A crucial point to note is that the two steps are hierarchical, in that any lower tier areas that step 2 has grouped (due to them not meeting the required self-containment level for commuting flows) are not disaggregated but are grouped in their entirety. If they were not, the upper tier would not be a grouping of the lower tier areas, as is the explicit intention here.

Map K10.2 shows the results from an analysis whose step 1 produced areas satisfying a 55 per cent criterion for migration self-containment – the lower tier areas – and then step 2 groups these as necessary so that the upper tier areas satisfy a 75 per cent criterion for self-containment of commuting flows. The upper tier areas are shown by thick dark maroon boundaries. If the commuting-based step 2 grouped at least two lower tier areas, these separable migration-based lower tier areas are distinguished by fine black boundaries. For example, in the far north of England the Berwick area is an undivided upper tier area (Map K10.2), showing that step 1 found that this area met the migration-based criterion to be a distinct lower tier area, and then step 2 found that it did not need to be grouped to meet the commuting-based criterion to also be a distinct upper tier area. By contrast, immediately to the south there are a large number of separable migration-based lower tier areas that the commuting-based step 2 has had to group to create an upper tier area that meets the self-containment level required (75%). The latter area includes Tyneside and its pattern of many lower tier areas being grouped by step 2 is repeated across many parts of the country where there is a large conurbation (due to fact in more metropolitan areas that there is a greater tendency for longer-distance commuting, as has already been mentioned).

Map K10.2 illustrates a disadvantage of Approach 5 that affects its upper tier boundaries. In practice, step 2 is constrained by only being able to group whole lower tier areas. This means that it has just a few hundred building block areas, in contrast to the many thousand wards that other boundaries here have been constructed from: the fewer the building-block areas, the less optimal the boundaries. The odd shape of the London upper tier area illustrates the effect, with its implausible finding that places as close to the conurbation as Chelmsford and Reigate are not part of the upper tier area that – at the same time – extends into Northamptonshire and Suffolk (Map K10.2). This is a major issue here because it is the upper tier boundaries that were identified as the primary requirement of NHPAU (for modelling and policy purposes).

Approach 6: Lower tier areas based on migration within upper tier areas based on commuting

To solve the disadvantage of Approach 5 that has just been identified, Approach 6 reverses the ordering of the two steps, thus defining the commuting-based upper tier areas directly from individual wards. This allows the crucial upper tier boundaries to be more optimally defined, but it creates a technical challenge because there is no existing method for disaggregating an upper tier set of areas into the largest possible number of lower tier areas that satisfy statistical criteria (in this case, levels of migration self-containment). The technical innovation here was to treat each upper tier area as a separate problem, so the step 2 analysis takes its constituent wards individually and then groups them until they meet the migration self-containment criterion without allowing any grouping to cross the upper tier boundaries.

Map L11.2 shows the result, with the boundary colours distinguishing upper and lower tier areas in the same way as before (Map K10.2). The upper tier boundary around London provides one case where Approach 6 does seem to have produced a more convincing set of results. Although addressing the problem of the upper tier, reversing the order of the steps has created some new problems. At the lower tier level, the fact that the boundaries are defined by step2 within the rigid limitations of the upper tier areas makes it difficult for them to be close to optimal in their pattern. Map 11.2 shows, as one example, some fragmented lower tier boundaries in the south west of the London upper tier area. More notably, at least one lower tier area includes two substantial non-contiguous parts: Chester and its nearby areas are grouped with the bulk of the Wirral peninsula but a significant area between them (extending from Ellesmere Port through to Shotton and adjacent parts of Wales) is divided between two different lower tier areas.

There is also an entirely new phenomenon. If the lower tier areas are defined first – as Approach 5 did – there is no possibility of any such area failing the criterion for lower tier migration self-containment but reversing the steps removes this guarantee. Map L11.2 shows three areas 'speckled' with "D" symbols (one for each ward in each area); when these upper tier areas were analysed by step 2 it was found that the criterion for lower tier migration self-containment could not be met even after grouping all the wards back together. In short, these are whole upper tier areas which fail to satisfy the level of migration self-containment required of lower tier areas.

Map M11.5 shows the boundaries produced by raising the step1 commuting self-containment criterion for the upper tier (to 77.5%). The difference is not very great, but is noticeable in cases such as London (which now extends further into Kent) and the area around Tyneside (which extends to Yorkshire). Where the upper tier boundary is different to that shown before (Map L11.2), the lower tier areas can also be different – because step 2 is working within different outer boundaries – even though there has been no change to the level of migration self-containment required of lower tier areas. It can be seen that the same three upper tier areas are 'speckled' again here: their upper tier boundaries have either not changed at all or have not changed enough for their migration self-containment levels to have risen above the lower tier criterion which has been held constant (55%).

It is worth noting that, in the case of the northern 'speckled' area, much of the explanation for its low level of migration self-containment will be due to the area including Catterick where the military base always produces disproportionately large numbers of long-distance migrant flows. This may provide one argument for 'discounting' this case as not a relevant problem here, but this area is not in fact the case with the lowest level of migration self-containment of the three 'speckled' lower tier areas.

Approach 6 appears in general to produce sets of boundaries which offer many advantages, with few problems encountered in terms of the technocratic criteria here (eg. there are very few non-contiguities). The question remaining is whether it is a major disadvantage that its sequence of steps – with the upper tier defined first and the lower tier next – can result in upper tier areas not meeting the criterion required of lower tier areas? (Whether this happens in practice depends on the level self-containment criteria set.) There are a number of possible responses to this.

- It can be argued that this is not really a problem because the upper tier areas would be the more important ones for NHPAU and Approach 6 seeks to optimise their boundaries. In fact the lower tier boundaries are really only sought in upper tier areas that – in practice, due to their size – can be subdivided into several areas meeting the migration self-containment criterion set for separable lower tier areas. There are many upper tier areas which cannot be so subdivided, so the fact that some of these may not in fact meet that migration self-containment criterion may not really be a critical problem.

- It remains possible to adjust the criteria for the two steps so the problem does not present itself. If the step 2 criterion for migration self-containment is lowered to 50 per cent then the three 'speckled' upper tier areas all meet this criterion. That said, the fact that it is the upper tier boundaries which are the more important limits the room for adjustment of the criterion for step 1.

- A final option could be to devise a step 3 which grouped any upper tier areas which did not also meet the criteria set for lower tier areas. This could be done relatively simply, but it has not been done here because there was no consensus that it was necessary for all upper tier areas to meet the migration self-containment level required of lower tier areas.

Setting aside the question of the 'speckled' areas for the moment, it is appropriate to complete the technocratic assessment of the boundaries at both tiers. Map M11.5 provides more convincing boundaries than Map L11.2 to consider, not least because if there are to be two tiers then it is appropriate for the upper tier boundary around London to be well away from the conurbation (e.g. to not draw this boundary so that it excludes much of West Kent, as does Map L11.2). Map M11.5 does provide upper tier boundaries that recognisably represent a metropolitan/city region scale in the more urbanised parts of the country, while also recognising the real separateness (on both migration and commuting criteria) of peripheral areas in the rural South West, for example.

At the lower tier the assessment is somewhat less positive. Where the migration analysis does find more than one separable lower tier area within an upper tier boundary, that outer boundary has rigidly restricted the options available in drawing these lower tier boundaries. Map M11.5 shows in the London upper tier area offers some examples of how this can generate sub-optimal boundaries at the lower tier. The western fringe of the upper tier area has all been grouped together as a distended lower tier area, largely because the prevention of any grouping across upper tier boundaries debars the east-west links in this part of the London hinterland, even though those links that are stronger than the north-south links which produce this grouping. More worryingly – though not evident from the map – is that there are some substantial areas that are grouped together non-contiguously at the lower tier. Looking again at the western margin of the London upper tier area, a small 'circular' lower tier area can be seen to be separate from the distended area to its west and north. This small area includes Guildford but it is not a separate lower tier area in its own right in fact, because it is non-contiguously grouped with the seemingly separate lower tier area that includes the southern boroughs of inner London.

In short, a nested set of two tiers is seen to produce – perhaps inevitably – some sub-optimal boundaries. Whereas with Approach 5 the unconvincing results were at the upper tier, with Approach 6 it is at the lower tier where the constraints operate and where sub-optimal boundaries can appear. After evaluation of the alternatives, NHPAU decided that the alternative with the most acceptable characteristics was the set of boundaries produced with the lower-tier migration self-containment threshold lowered to 50 per cent: Map N11.4 shows these boundaries. As noted above, lowering the threshold to 50 per cent avoids the problem of any upper-tier areas failing the lower-tier criteria with the consequence that there are no 'speckled' areas here (at least not in England).

7: Completing the definitions

The choice between a single and a two tier definition of housing market areas relies based on criteria related to theoretical issues and policy applicability. The task of this report has essentially been to define the most appropriate set of boundaries for each of these strategies. To achieve this, the research has made innovations that allow the definitions to take account of both migration and commuting patterns: a key conclusion here is that this development brings real benefits.

If the decision were taken to have only a single tier of areas, then the advantage of the boundaries having taken account of both commuting and migration patterns puts the focus on Approach 3. It may be that Approach 3 could be viewed as a compromise because it is not a pure application of any particular theoretical perspective, but in a situation where the 'theoretical jury remains out' this may not be such a strong criticism. From that form of analysis, Map D4.2 is the set of single tier boundaries examined here that best meets the technocratic criteria.

Further innovations have enabled two tier definitions to be produced. On the technocratic criteria that are central to this report, Approaches 5 and 6 have both produced boundaries that are more convincing at one tier than the other. The emphasis of NHPAU on the upper tier and led to a focus on Approach 6 because it has more optimal results at that level: as noted above, Map N11.4 provides the set of this type of boundaries that is seen to perform best against the technocratic criteria.

NHPAU concluded that a two tier set of boundaries brings more advantages than disadvantages and this means that Map N11.4 is the 'central case' set of proto-housing market areas. As mentioned earlier, all the boundaries defined here have at least some non-contiguities (because a constraint on the analyses to only consider contiguous groupings would prevent it searching for the optimal results). To make these proto-housing market areas into the final 'gold standard' set of boundaries, any non-contiguities needed to be removed. The process was as follows:

1   identify which ward re-allocation(s) will remove the non-contiguity with the minimum movement of population
2   test that all areas affected by that re-allocation still meet the self-containment threshold set for them (e.g. 50% migration self-containment for lower tier areas)
3   if the test above is successful then implement that re-allocation; if not then revisit the first step above and find the alternative re-allocation involving next smallest population

This process was followed first for the upper tier set of areas from the chosen proto-housing market areas (Map N11.4) and then subsequently followed for the lower tier areas within those revised upper tier boundaries.

It is worth noting that with the upper tier boundaries the third step listed above was not needed: all the non-contiguities could be resolved with a minimum level of population movement. At the lower tier there were three more intractable problems and this resulted in three of proto-housing market areas at this level having to be merged with neighbouring lower tier areas in the same upper tier housing market areas. (These three cases were Carlisle and Ipswich – where pairs of lower tier proto-housing market areas had to be merged, creating 'gold standard' housing market areas which were the same at both tiers – and in the Sheffield upper tier housing market area where the effect was the reduction of the number of separable lower tier areas by one.) Map O11.4 shows the 'gold standard' housing market areas at both tiers. Comparison with the preceding map shows that it was relatively marginal changes only that were needed to remove all the non-contiguities at both tiers.

As already emphasised, the upper tier boundaries were the focus of NHPAU for its modelling purposes. For modelling in particular, it is necessary to have a best-fit to the upper tier areas which does not subdivide whole local authorities. The data needed for the modelling leads to a focus on the boundaries used for reporting 2001 Census data, at the lower level of the then local authority hierarchy (viz: London boroughs + metropolitan boroughs + other unitaries + Shire districts). Hence for this best-fit it is the 2001 local authorities that are referred to as local authorities, and this means that the process results in some boundaries which cut through current local authority areas. This occurs in those parts of the country where local government areas were restructured since 2001 (and mainly where Shire districts were abolished in order to make Shires into large unitary authority areas, as in Northumberland).

Map P11.4 shows the 'silver standard' upper tier housing market area boundaries which result from the best-fit to the 2001 local authorities; current local authorities are shown as background colouring. Taking the London area initially, this new best-fit upper tier London boundary rarely departs far from the 'gold standard' boundary (Map O11.4). In other words, the local authority best-fit in the London case is a close fit to the 'gold standard' upper tier housing market area (which is ward based). By contrast, in the south west there are several cases where the best-fit is not all that close. Most such sub-optimal cases are the result of the upper tier areas being of a similar size to the local authorities which provide the best-fit with its building-block areas: this is not surprising because any grouping process produces less optimal results where the 'target' areas and building-block areas are similar in size.

Although in some parts of the country the best-fit process makes the upper tier boundaries markedly less optimal, it is notable that there was no upper tier area to which no local authority was allocated, when that was in fact a very real risk in carrying out a best-fit with building-block areas which in many parts of the country are at least as large as the areas they are being grouped into. This problem is very much present if attention turns to the current local authorities that provide the coloured background of Map O11.4: several of the new large Unitary Shires such as Cornwall cover more than one 'silver standard' housing market area (in other words, if the current local authorities are used as 'target' areas for a best-fit then several housing market areas will be lost in the process). More generally, it is crucial to emphasise that the best-fit boundaries (Map P11.4) are very much the 'silver standard' version of upper tier housing market areas and should only be used when data constraints militate against using the 'gold standard' (Map O11.4).

As a final presentation of the 'gold standard' housing market areas, Map Q11.4 shows the boundaries superimposed on a traditional map base (copyright of Ordnance Survey). Some final observations can be drawn from the alignment of the boundaries with respect to the 'real world' geography of hills, settlements and transport corridors. In particular, it is remarkable that no upper tier boundary cuts through a substantial urban area: this is despite these results having no 'seeds' or constraints ensuring that all towns, cities and conurbations are grouped in its entirety within the same upper tier housing market area. Putting this into the context of statistical taxonomy – which is what these data-driven groupings of wards are – it is a dramatic result that every single one of the literally millions of ward pairs which are in the same urban area have been allocated to the same housing market area. Of course, assessing this result as a success depends on the assumption that housing market areas splitting whole urban areas is undesirable other things being equal, but that is an example of the technocratic criteria against which the results here have been assessed. In a similar vein it is notable that:

- both upper and lower tier housing market areas reflect physical barriers like the Pennines and Salisbury Plain
- wider estuaries like the Mersey and Humber appear as upper tier boundaries (unlike narrower ones like the Tees which are straddled by upper tier housing market areas but appear as lower tier boundaries)
- lower tier areas in and near London are frequently aligned along transport corridors

There are some areas which have emerged as sufficiently separable in migration terms to be recognised as lower tier areas when their distinctiveness would probably have been expected by few: for example, there are two separable areas of the Isle of Wight. The fact that the overall pattern of boundaries provides a very convincing geography of housing market areas lends confidence that even some surprising results can be seen as the result of analyses validly reflecting the underlying patterns of 'real world' housing markets.

**Summary data on the final sets of housing market areas will be found on an internet site featuring this report**

**Map A1.2 Strategic housing market assessment areas (as supplied)**

GOR boundaries on SHMAs
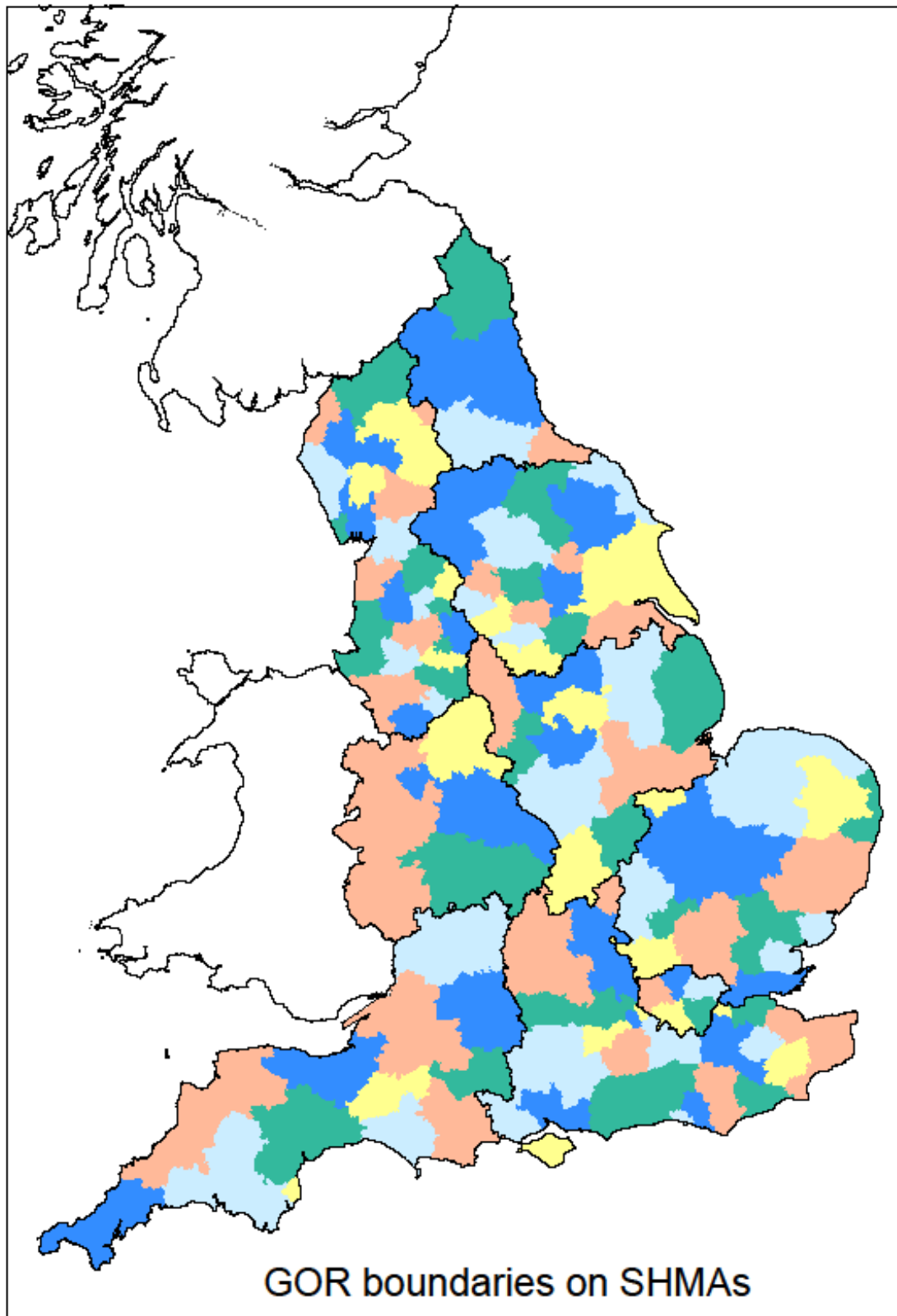
Map B9.1 Higher earner commuter labour market areas (66.7% self-containment)

HH3p5_25kSC66p6_75_STW
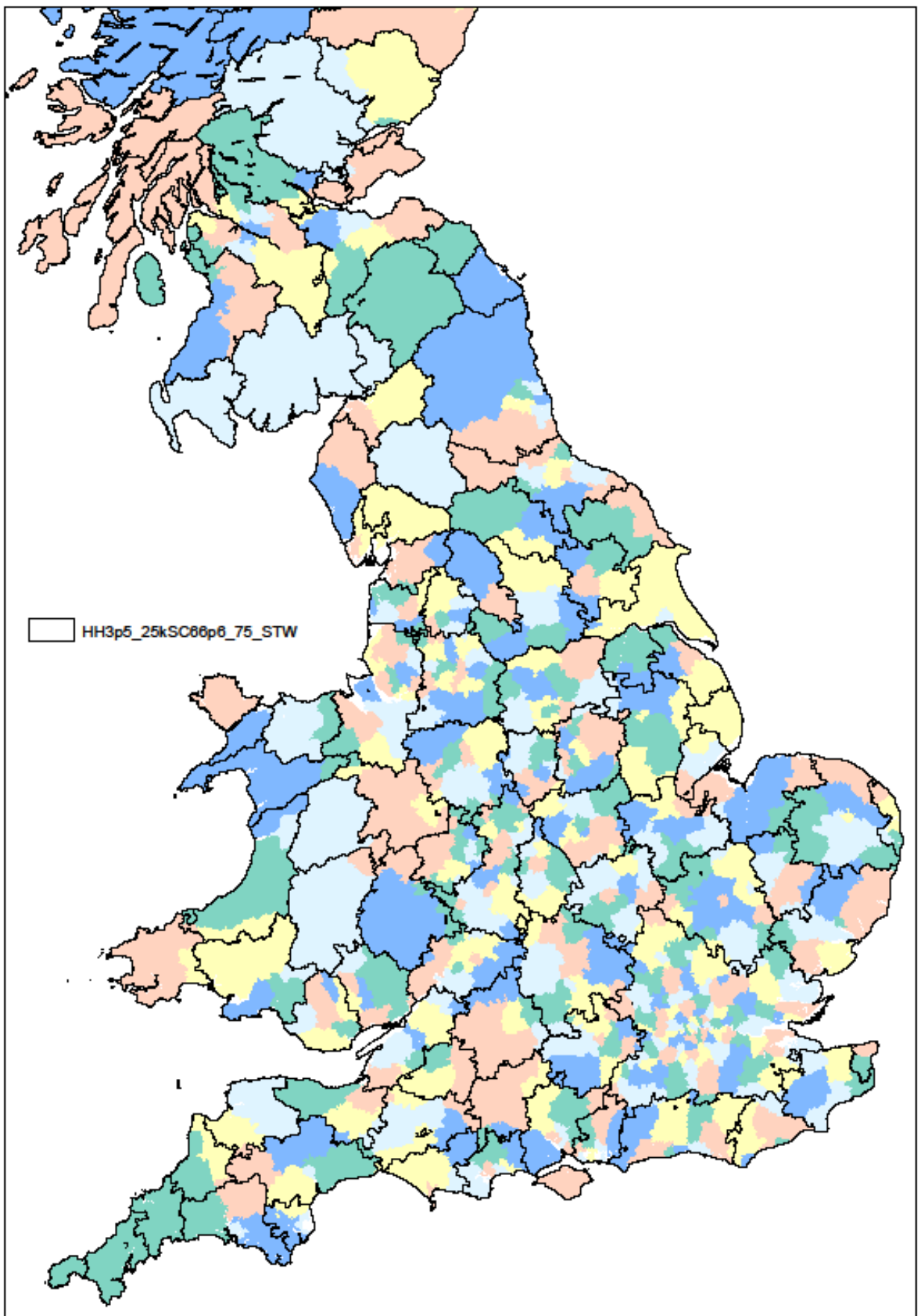
Map C8.2 All commuter labour market areas (75.0% self-containment)

01TTWApt_HH3p5_25kSC75_85
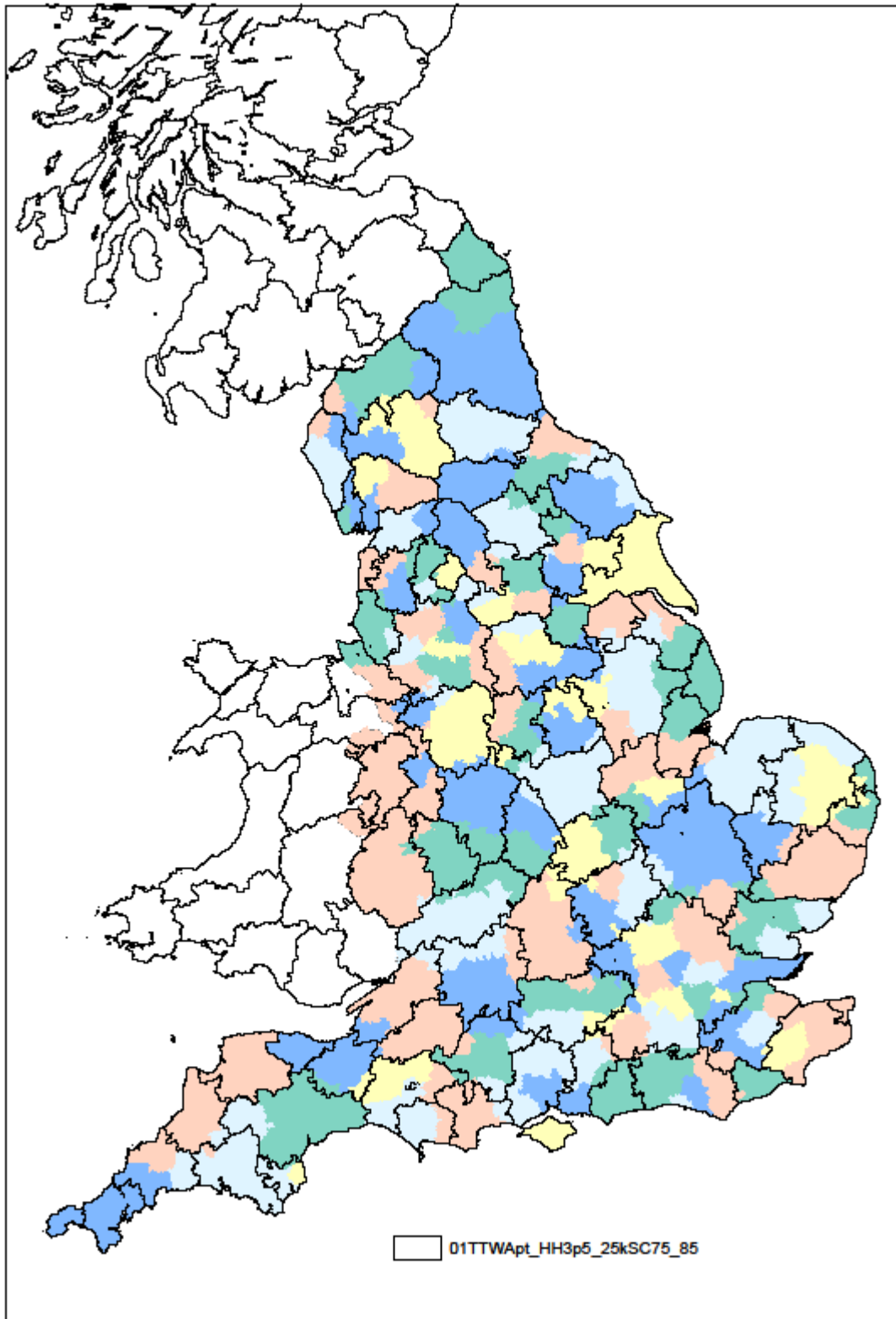
Map D4.2 Labour market areas 'tested' for migration self-containment (55.0%)

HH3p5_20kSC72p5_80pt_MG55_60
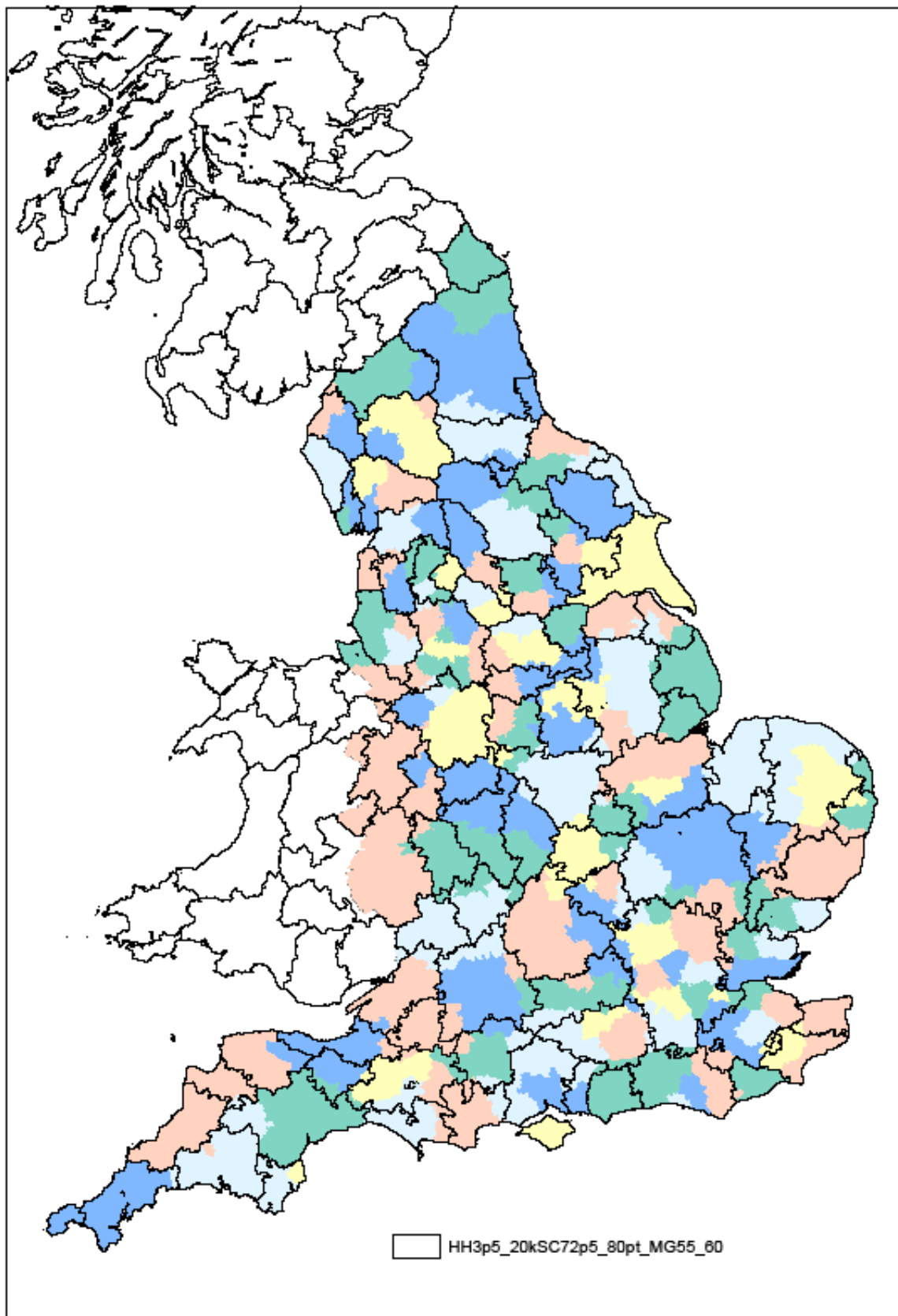
Map E2.1 Migration self-containment: all moving group reference persons (66.7%)

MGRP_All_HH3p5_25kSC66p6_75_STW
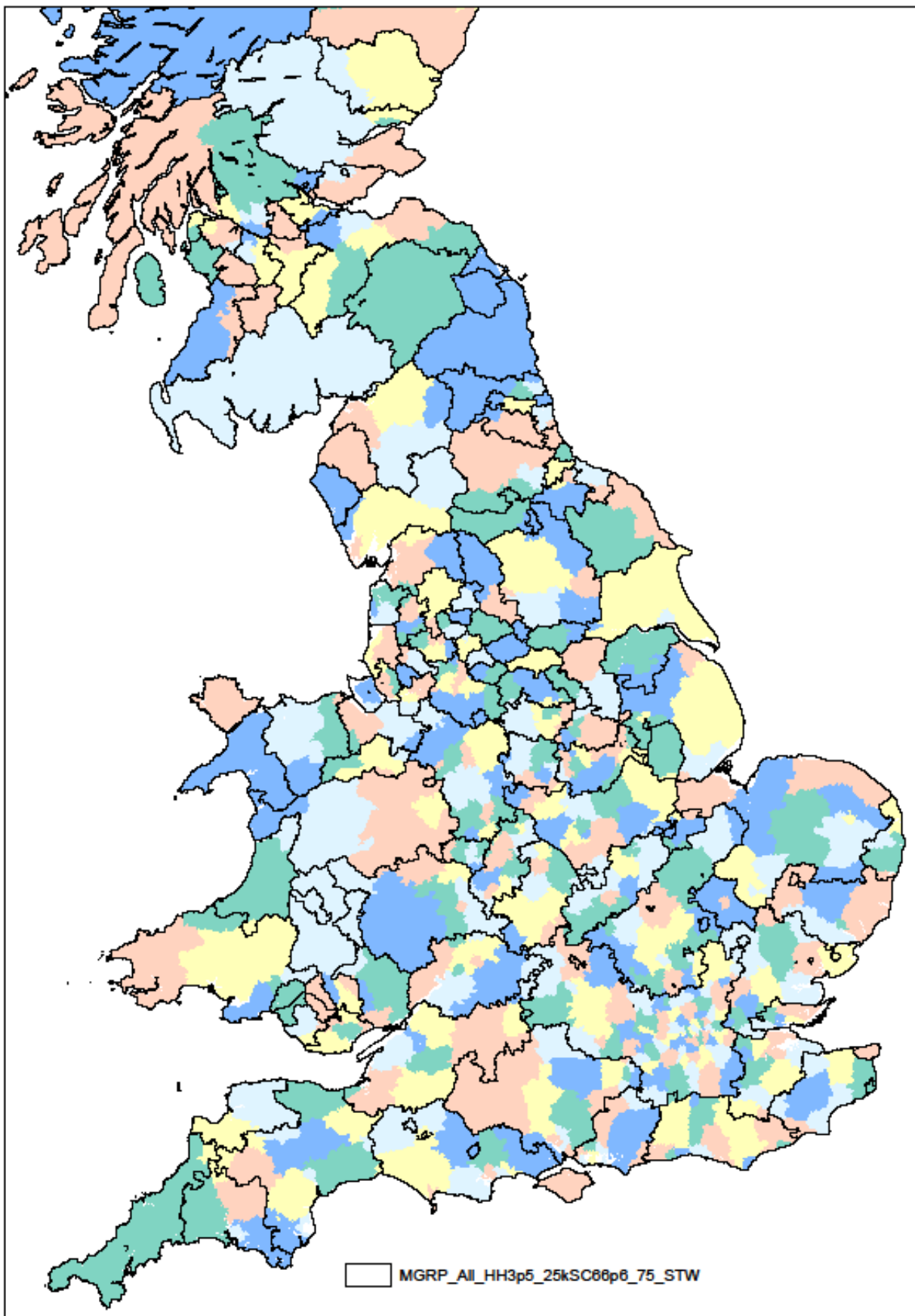
Map F2.2 Migration self-containment: owner-occupying moving group reference persons (66.7%)

MGRP_OO_HH3p5_25kSC66p6_75_STW
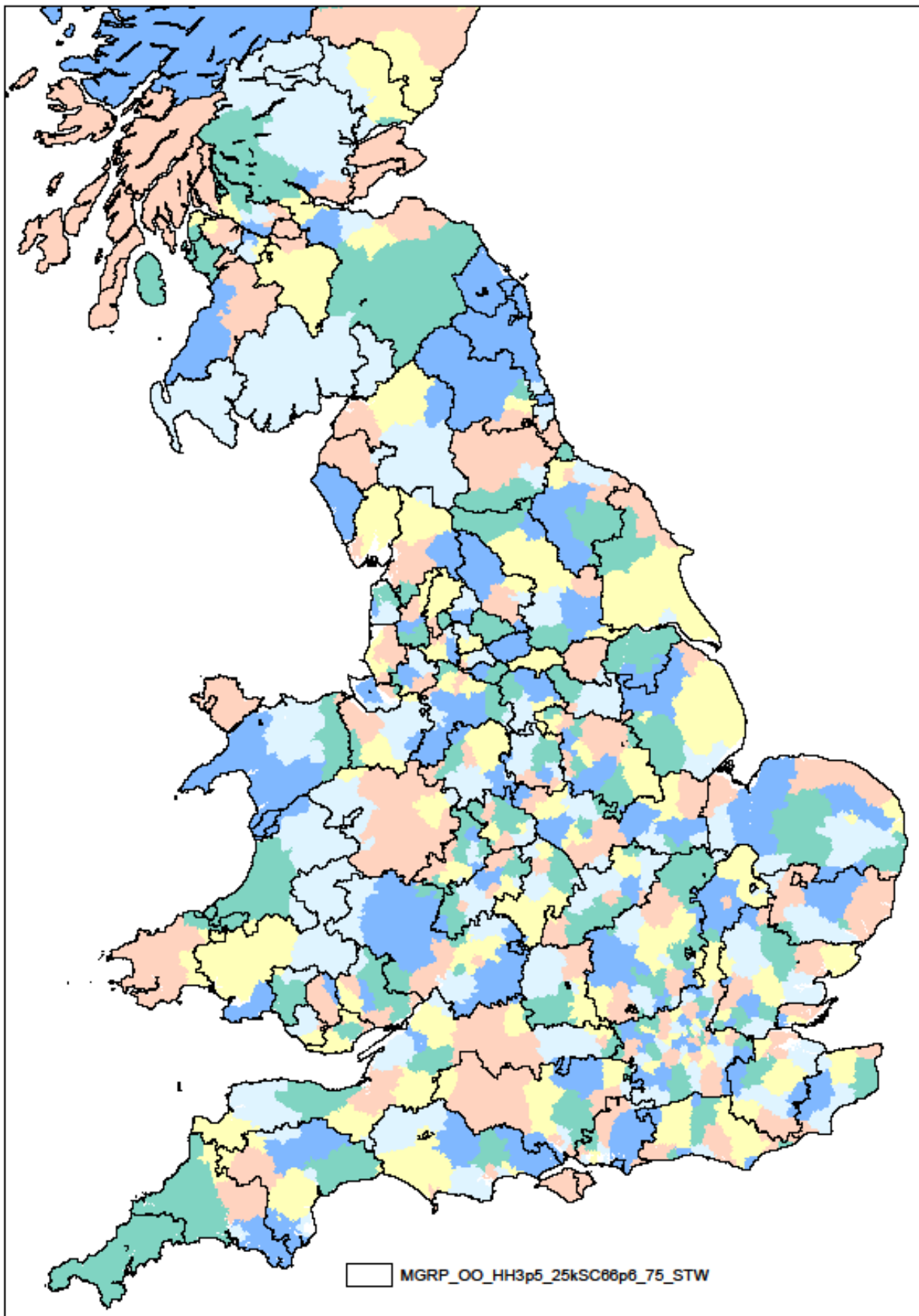
Map G2.3 Migration self-containment: social renting moving group reference persons (66.7%)

MGRP_Rent_HH3p5_25kSC66p6_75_STW
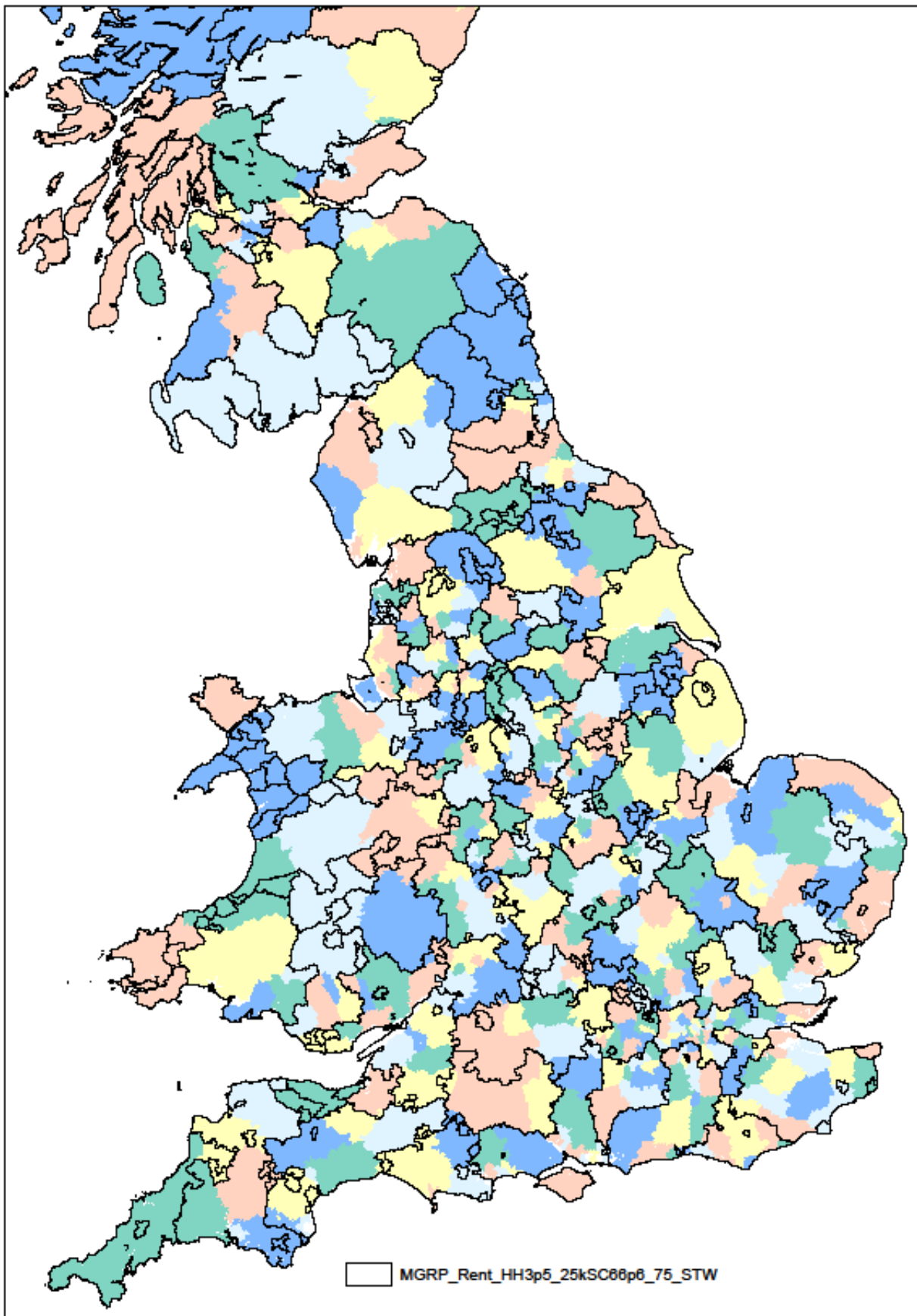
Map H2.4 Migration self-containment: other renting moving group reference persons (66.7%)

MGRP_Other_HH3p5_25kSC66p6_75_STW
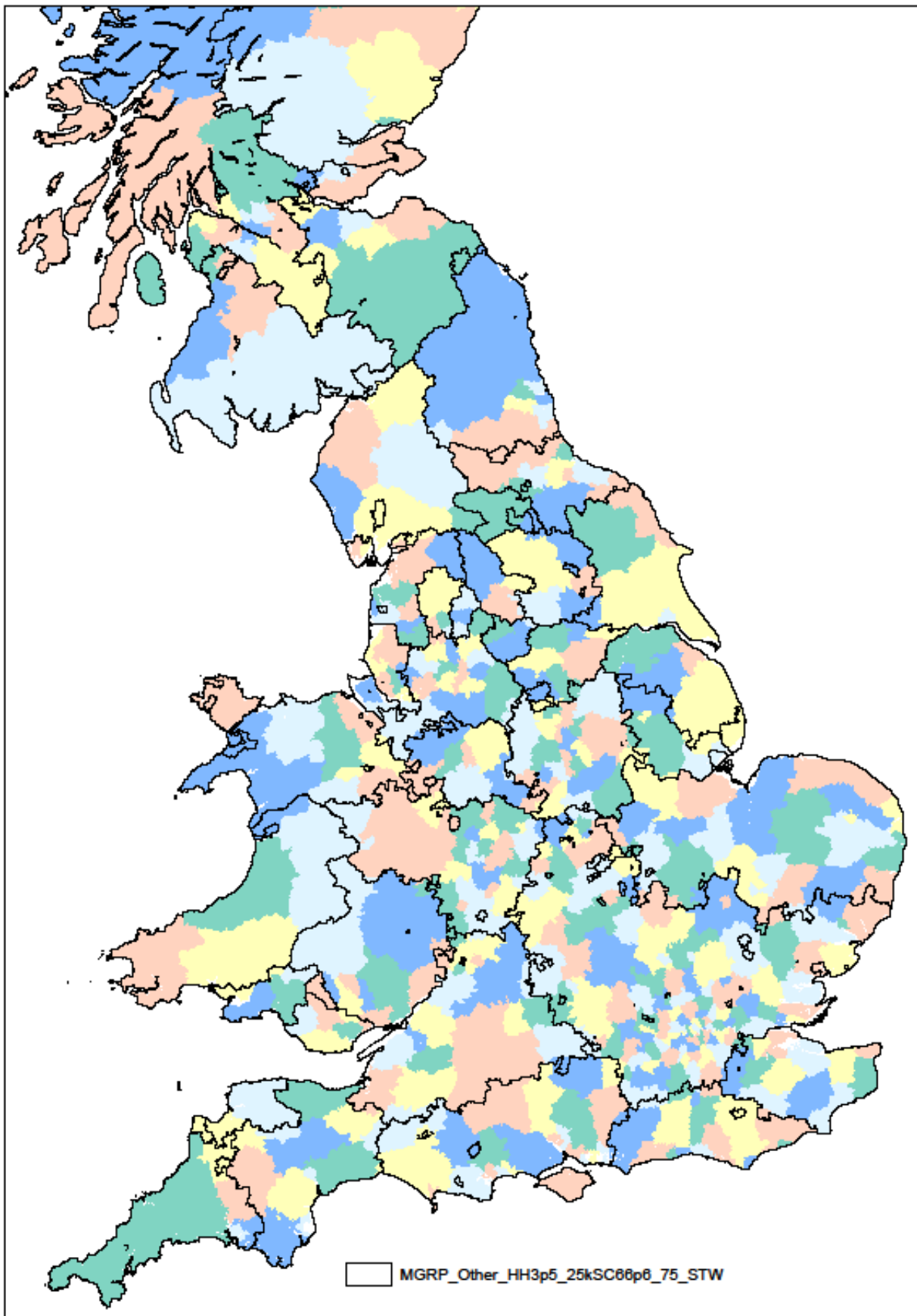
Map I1.3 Migration of all moving group reference persons (66.7%), with 'seeds' of urban areas
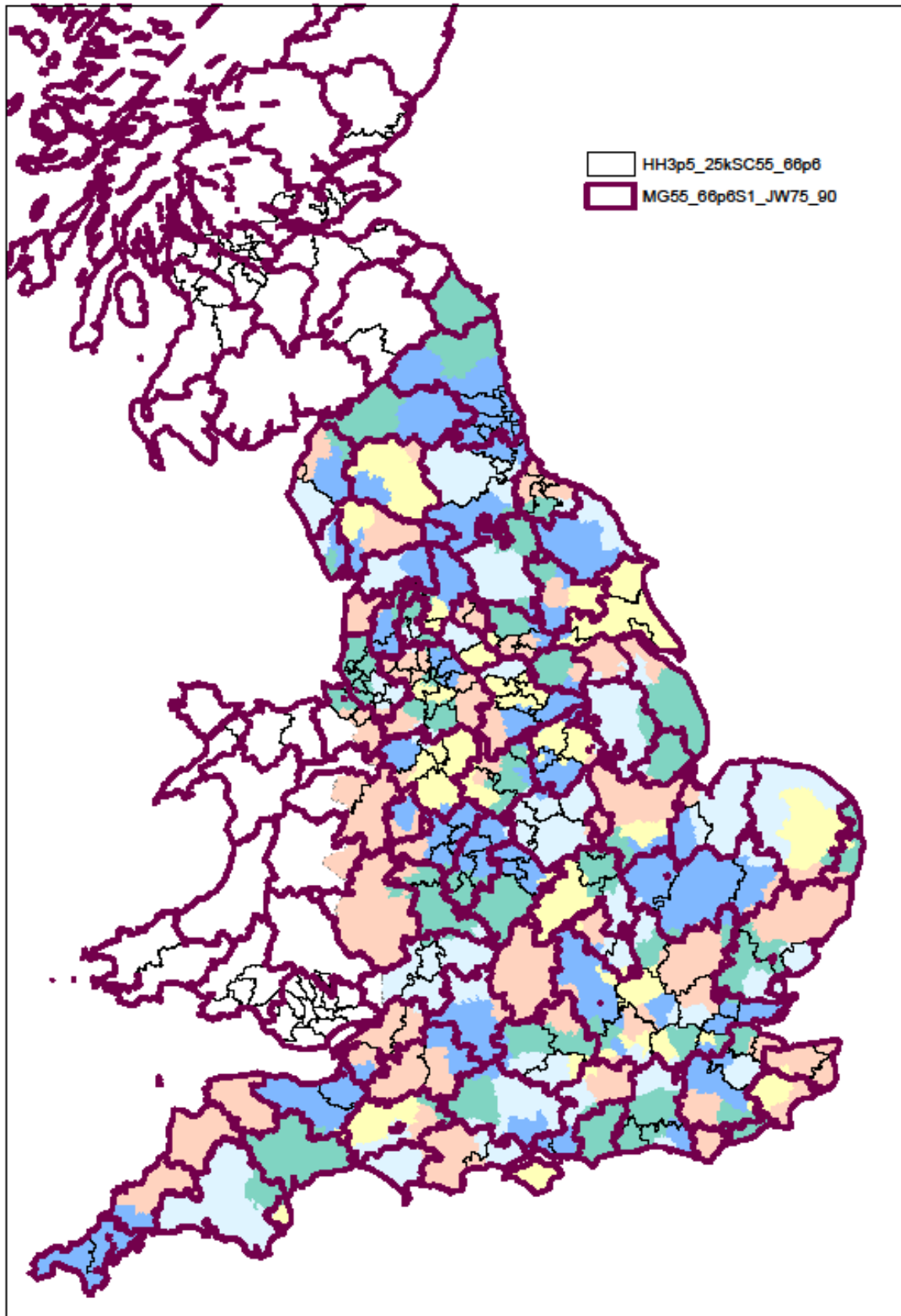

MGRP_All_UAseeds_MG66p6_75_o
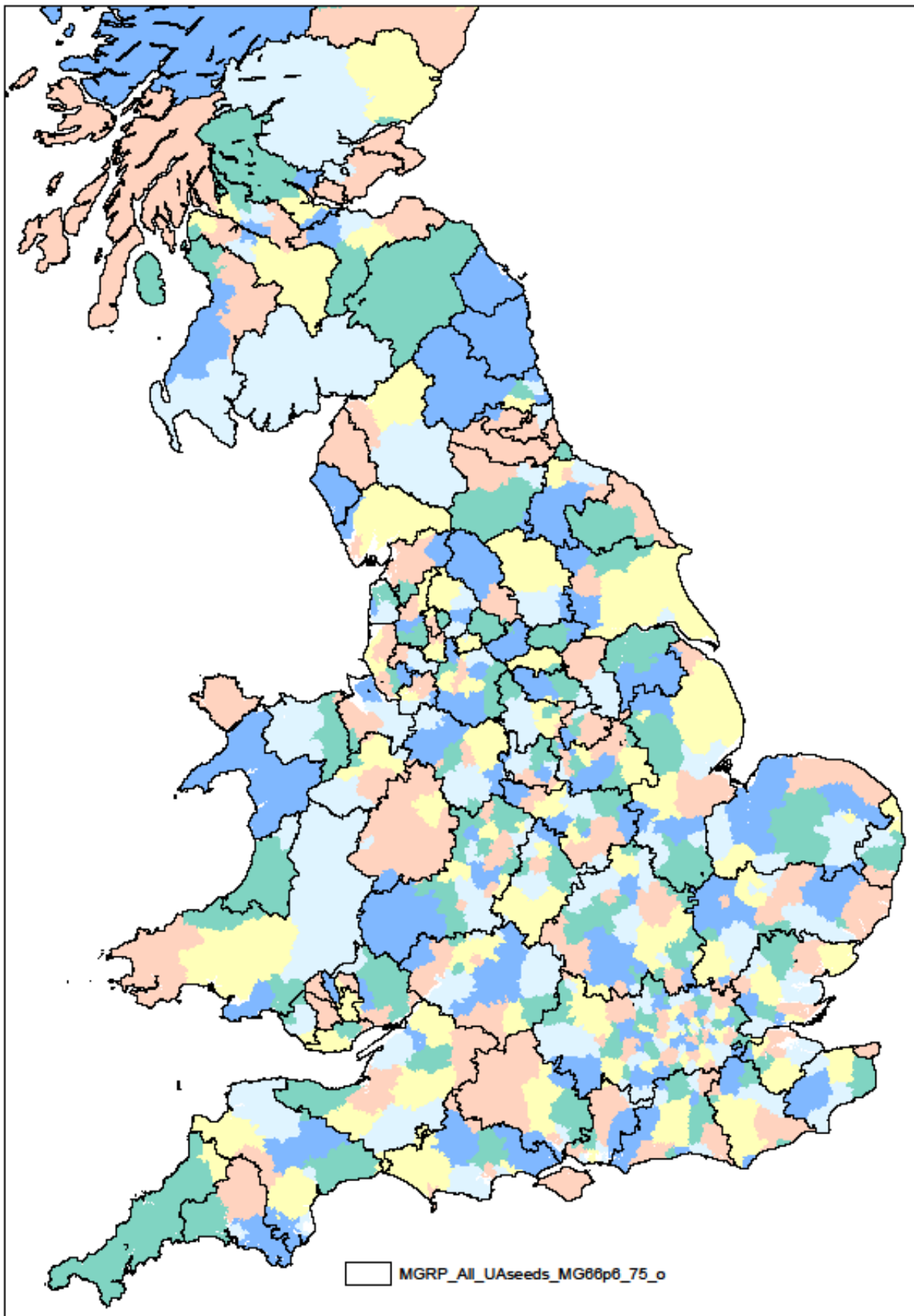
Map J1.4 Migration of all moving group reference persons (66.7%), with 'seeds' of urban areas split by local authorities

MGRP_All_UALAseeds_MG66p6_75_o
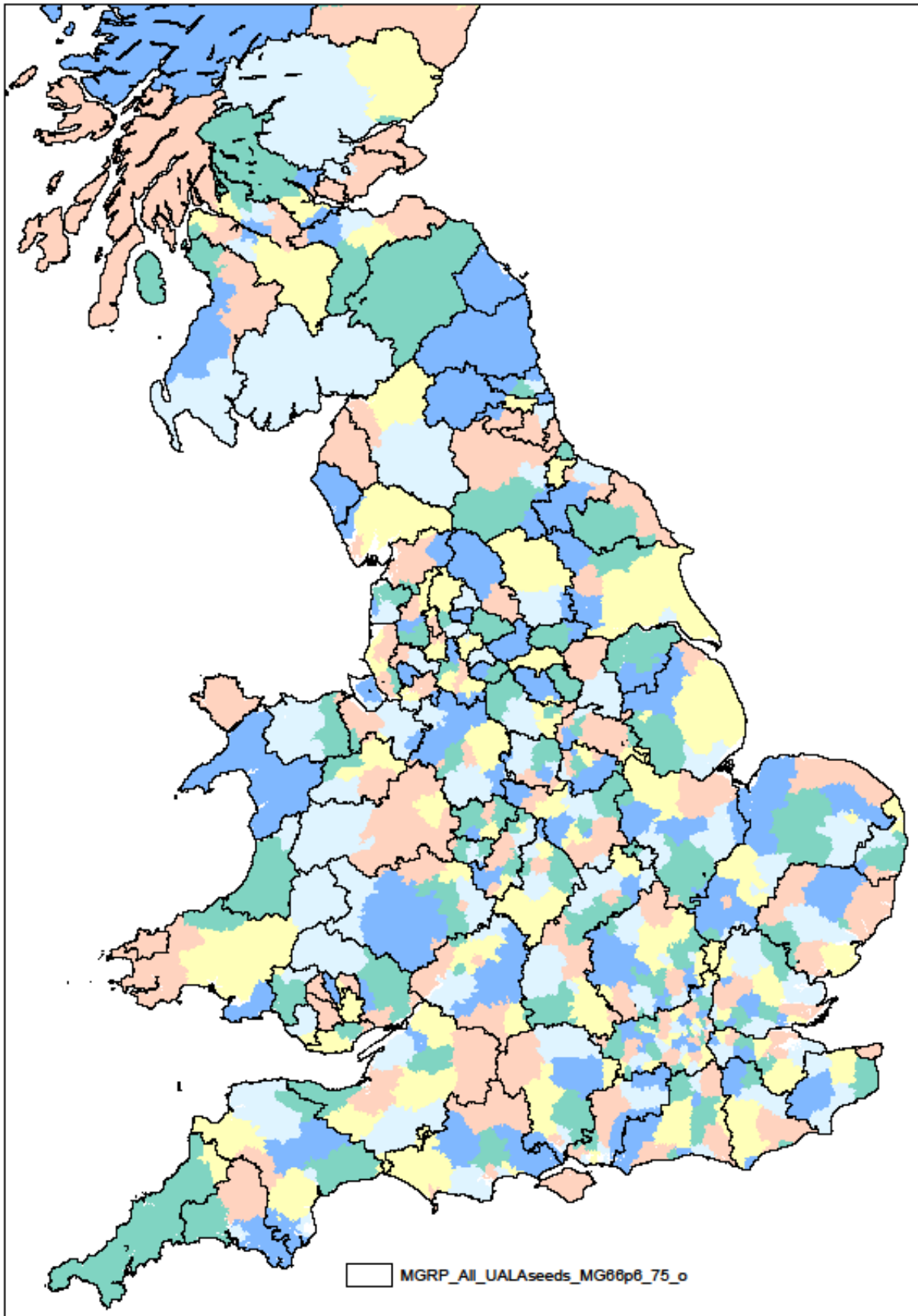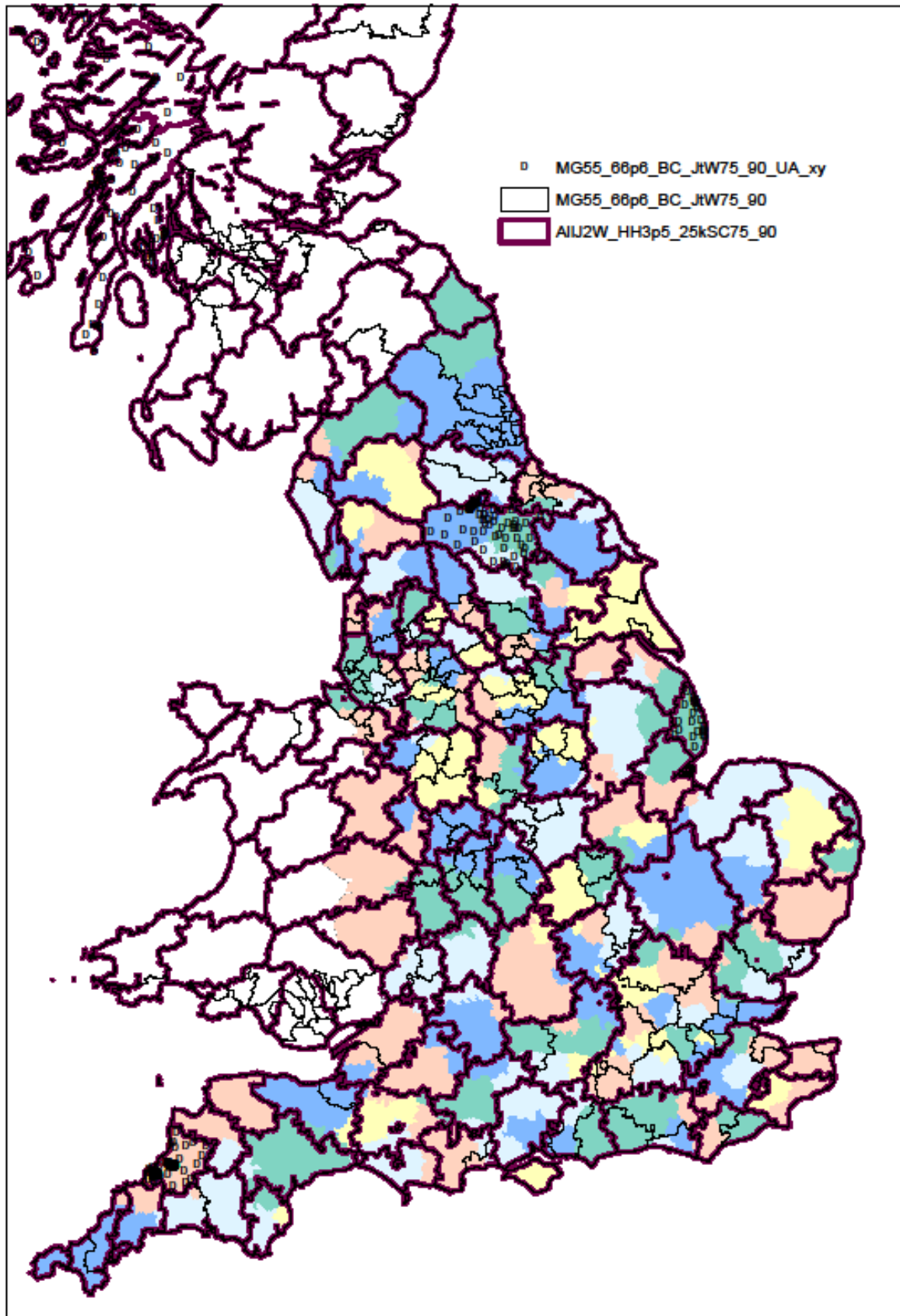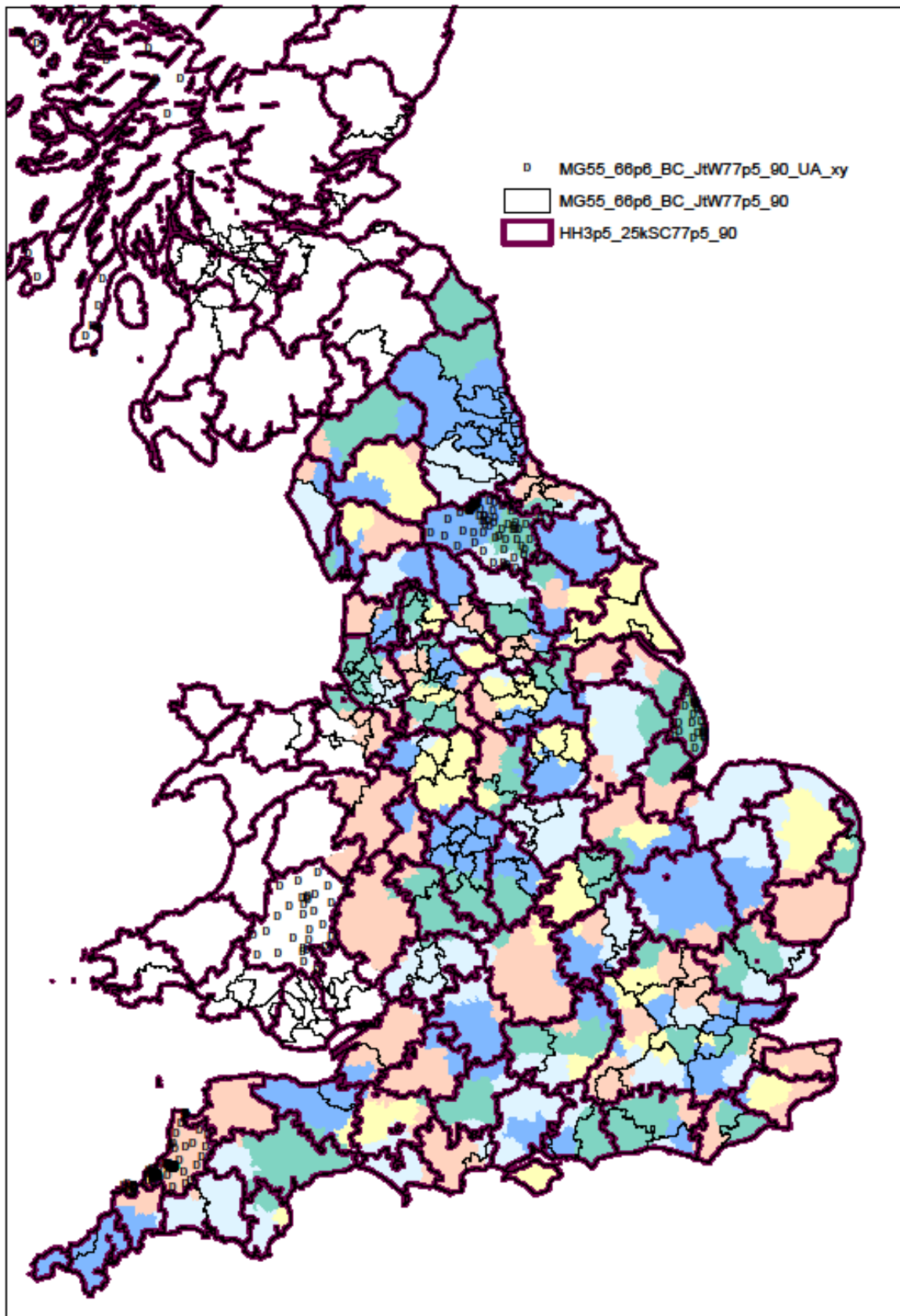
Map K10.2 Lower tier based on migration (55.0%), grouped to upper tier by commuting (75.0%)

Legend:
- HH3p5_25kSC55_66p6
- MG55_66p6S1_JW75_90

Map L11.2 Lower tier based on migration (55.0%) within commuting-based upper tier (75.0%)

Legend:
- D — MG55_66p6_BC_JtW75_90_UA_xy
- MG55_66p6_BC_JtW75_90
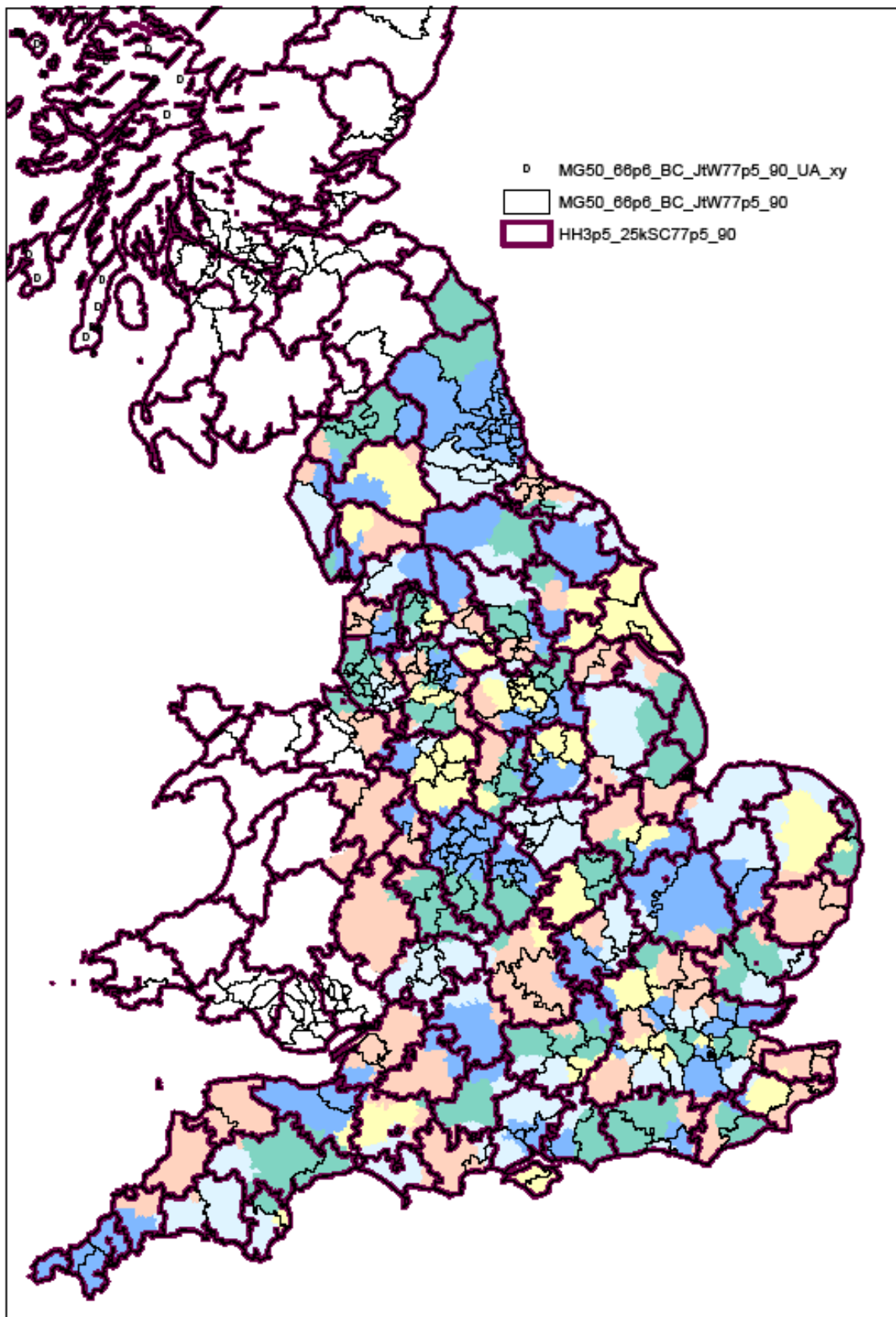- AllJ2W_HH3p5_25kSC75_90

Map M11.5 Lower tier based on migration (55.0%) within commuting-based upper tier (77.5%)
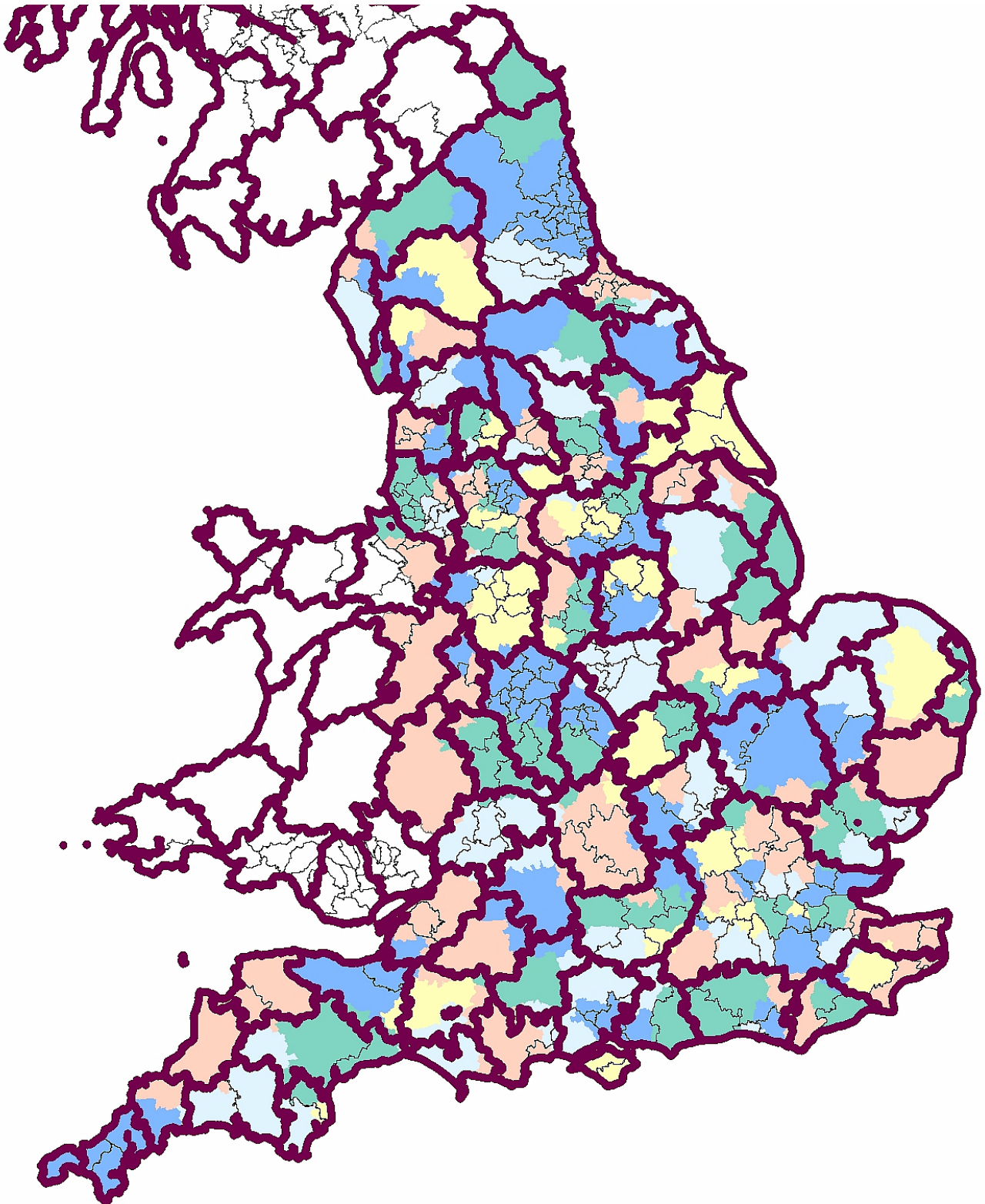
Map N11.4 Lower tier based on migration (50.0%) within commuting-based upper tier (77.5%)

Map O11.4 The set of 'gold standard' two tier housing market areas (following the removal of non-contiguities)
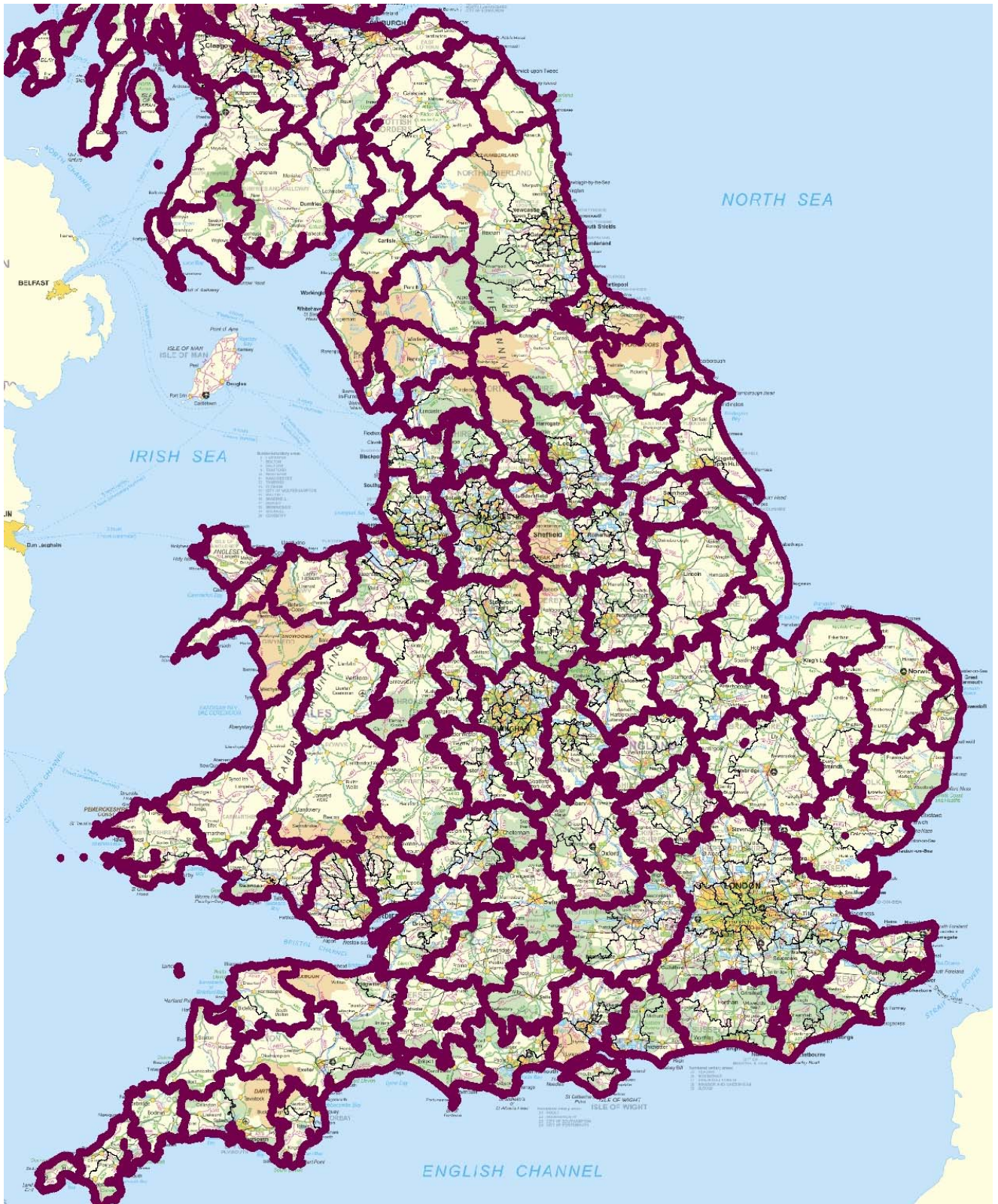
Map P11.4 The set of 'silver standard' upper tier housing market areas (with current local authorities as background)
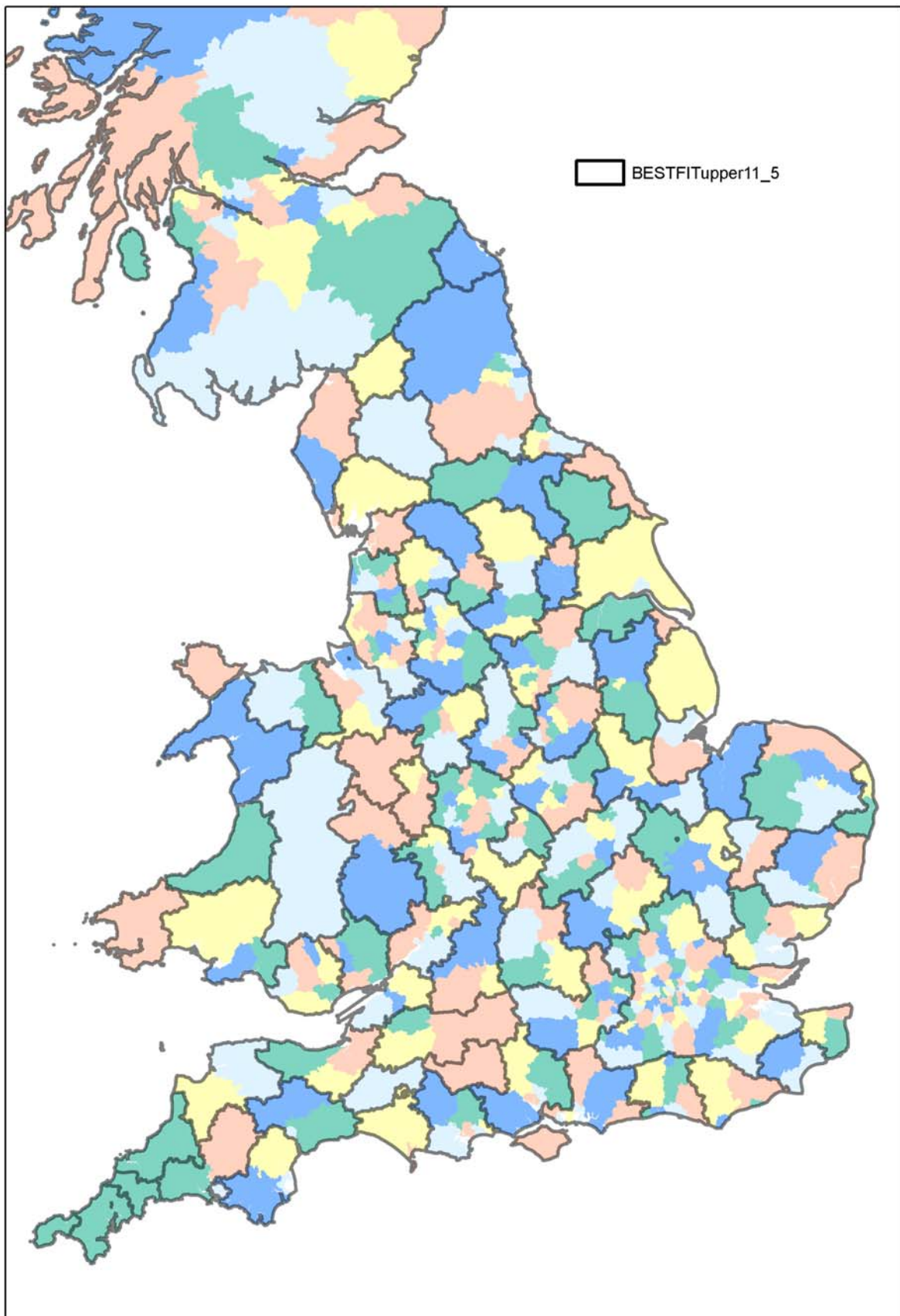
Map Q11.4 The set of 'gold standard' two tier housing market areas (with an Ordnance Survey background map)
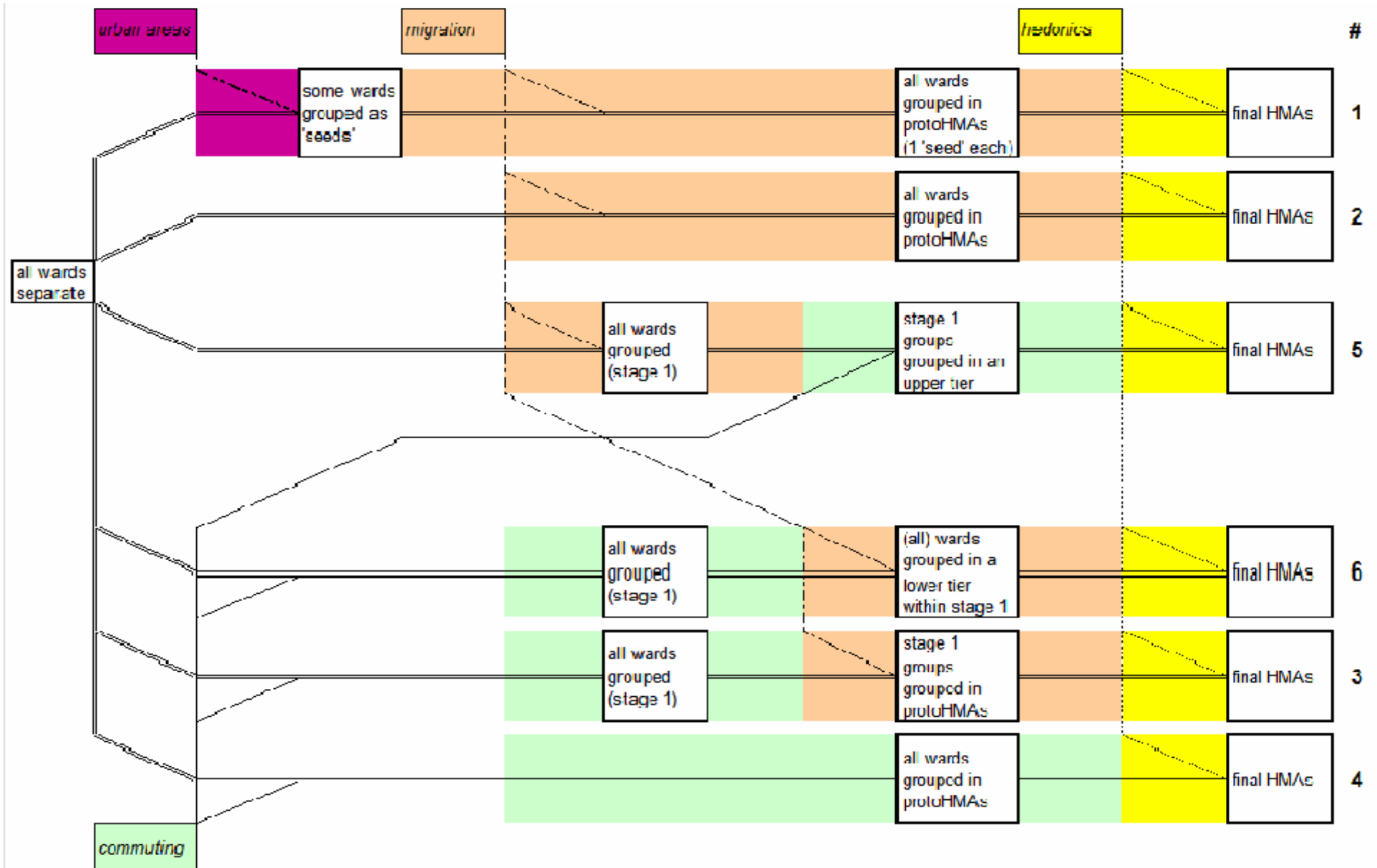
Table 2 Summary of the analyses evaluated

analyses NOT 'seeded' use wards	dataset for step1	S-Cmin.	any variation on usual analysis	dataset for step2	S-Cmin.	*n* [lower tier]	*map*
sets of pre-defined areas							
SHMA areas (as approximated, for PUSH)			areas not grouped, they are here as one of the 'bench-mark' boundary sets			118	*1.2*
lower-tier LAs as at 2001			areas not grouped, they are here as one of the 'bench-mark' boundary sets			352	*1.1*
upper-tier LAs as at 2009			areas not grouped, they are here as one of the 'bench-mark' boundary sets			158	*1.5*
TTWAs (as approximated frrom LSOAs)			areas not grouped, they are here as one of the 'bench-mark' boundary sets			162	*8.5*
Approach 1							
Urban Areas as 'seeds' (can be split later)	all 25(+) MGRPs	66.7%	boundary 'optimiser' used	-	-	73	*1.3*
as above, but Urban Areas split by LAs	all 25(+) MGRPs	66.7%	boundary 'optimiser' used	-	-	86	*1.4*
Approach 2							
no 'seeds'	all 25(+) MGRPs	66.7%	-	-	-	86	*2.1*
no 'seeds'	owner-occupiers	66.7%	-	-	-	65	*2.2*
no 'seeds'	social renters	66.7%	-	-	-	57	*2.3*
no 'seeds'	other renters	66.7%	-	-	-	29	*2.4*
no 'seeds'	all 25(+) MGRPs	60.0%	-	-	-	152	*2.5*
no 'seeds'	all 25(+) MGRPs	55.0%	-	-	-	223	*2.6*
no 'seeds'	all 25(+) MGRPs	50.0%	-	-	-	327	*2.7*
no 'seeds'	all 25(+) MGRPs	45.0%	-	-	-	414	*2.8*
no 'seeds'	all 25(+) MGRPs	40.0%	-	-	-	517	*2.9*
Approach 3							
no 'seeds'	all workers	72.5%	-	all 25(+) MGRPs	60.0%	79	*4.1*
no 'seeds'	all workers	72.5%	-	all 25(+) MGRPs	55.0%	93	*4.2*
no 'seeds'	all workers	72.5%	-	all 25(+) MGRPs	50.0%	99	*4.3*
no 'seeds'	all workers	72.5%	-	owner-occupiers	60.0%	69	*5.1*
no 'seeds'	all workers	72.5%	-	owner-occupiers	55.0%	85	*5.2*
no 'seeds'	all workers	72.5%	-	owner-occupiers	50.0%	93	*5.3*
TTWAs as 'seeds' (can be split later)	all workers	66.7%	-	all 25(+) MGRPs	65.0%	75	*6.1*
TTWAs as 'seeds' (can be split later)	all workers	66.7%	-	all 25(+) MGRPs	60.0%	104	*6.2*
TTWAs as 'seeds' (can be split later)	all workers	66.7%	-	all 25(+) MGRPs	55.0%	134	*6.3*
TTWAs as 'seeds' (can be split later)	all workers	66.7%	-	all 25(+) MGRPs	50.0%	155	*6.4*
no 'seeds'	top SECs	55.0%	boundary 'optimiser' used	all 25(+) MGRPs	66.7%	64	*7.1*
no 'seeds'	top SECs	55.0%	boundary 'optimiser' used	all 25(+) MGRPs	55.0%	115	*7.2*

seeds	step1	%	method	step2	%	number	ref
no 'seeds'	top SECs	55.0%	hierarchical step2, boundary 'optimiser' used	all 25(+) MGRPs	55.0%	118	*7.3*
Approach 4							
TTWAs as 'seeds' (can be split later)	all workers	75.0%	-	-	-	84	*8.1*
no 'seeds'	all workers	75.0%	-	-	-	85	*8.2*
no 'seeds'	all workers	72.5%	-	-	-	100	*8.3*
no 'seeds'	all workers	67.5%	-	-	-	140	*8.4*
no 'seeds'	all workers	80.0%	-	-	-	60	*8.6*
no 'seeds'	all workers	77.5%	-	-	-	75	*8.7*
no 'seeds'	top SECs	66.7%	-	-	-	71	*9.1*
national borders respected, no 'seeds'	top SECs	66.7%	-	-	-	71	*9.2*
no 'seeds'	top SECs	55.0%	-	-	-	141	*9.3*
Approach 5							
no 'seeds'	all 25(+) MGRPs	50.0%	hierarchical step2 produces the upper tier	all workers	75.0%	80 [327]	*10.1*
no 'seeds'	all 25(+) MGRPs	55.0%	hierarchical step2 produces the upper tier	all workers	75.0%	77 [223]	*10.2*
no 'seeds'	all 25(+) MGRPs	60.0%	hierarchical step2 produces the upper tier	all workers	75.0%	70 [152]	*10.3*
no 'seeds'	all 25(+) MGRPs	50.0%	hierarchical step2 produces the upper tier	all workers	77.5%	66 [327]	*10.4*
no 'seeds'	all 25(+) MGRPs	55.0%	hierarchical step2 produces the upper tier	all workers	77.5%	63 [223]	*10.5*
no 'seeds'	all 25(+) MGRPs	60.0%	hierarchical step2 produces the upper tier	all workers	77.5%	60 [152]	*10.6*
no 'seeds'	all 25(+) MGRPs	50.0%	hierarchical step2 produces the upper tier	all workers	80.0%	58 [327]	*10.7*
no 'seeds'	all 25(+) MGRPs	55.0%	hierarchical step2 produces the upper tier	all workers	80.0%	55 [223]	*10.8*
no 'seeds'	all 25(+) MGRPs	60.0%	hierarchical step2 produces the upper tier	top SECs	80.0%	53 [152]	*10.9*
Approach 6							
no 'seeds'	all workers	75.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	50.0%	85 [280]	*11.1*
no 'seeds'	all workers	75.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	55.0%	85 [204]	*11.2*
no 'seeds'	all workers	75.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	60.0%	85 [135]	*11.3*
no 'seeds'	all workers	77.5%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	50.0%	75 [280]	*11.4*
no 'seeds'	all workers	77.5%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	55.0%	75 [200]	*11.5*
no 'seeds'	all workers	77.5%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	60.0%	75 [130]	*11.6*
no 'seeds'	all workers	80.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	50.0%	60 [285]	*11.7*
no 'seeds'	all workers	80.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	55.0%	60 [204]	*11.8*
no 'seeds'	all workers	80.0%	step2 produces lower tier (within step1 areas)	all 25(+) MGRPs	60.0%	60 [131]	*11.9*

Figure 1 Approaches to defining housing market areas