These notes summarise the main discussion points from the final meeting in this one year ESRC Methodological Innovation network. Whilst the main output from the network is to be an NCRM methodological review paper, the workshop began with discussion of two other outputs from the network:

1. the special issue of International Journal of Microsimulation. Tzavidis updated the group on article progress and the issue looks set for mid 2014;
2. Kavroudakis presented output and illustrations from the R library that he has built to do several spatial microsimulation methods including an introductory R tutorial. The R library is available to freely download from the network website (link above).

Key challenges for the network, and the focus of this final workshop, are to:

1. identify areas where the separate statistical and spatial microsimulation methods can be improved, and;
2. seek (and make explicit) connections between statistical and spatial microsimulation approaches and then seek to exploit methodological and substantive benefits from those connections.

In terms of connections between the methods, and as others have noted previously (Simpson and Tranmer, 2005), the group discussed how iterative proportional fitting (IPF) can equivalently be expressed as a log-linear model. Similarly, gregwt comes from within the generalized regression framework though can also be understood as calibration against known counts (margins) for small areas within a fixed matrix of variables in order to seek to optimise weights for each case (typically individual or household)(Deville and Sarndall, 1992). One implication of this might be in terms of thinking about confidence intervals/credible intervals or, related, an estimate of variance around the point estimates. In an IPF framework boot-strap estimates loops might be an obvious way to achieve such intervals but this is computationally demanding and time consuming; in a log-linear regression framework, in contrast, one would tend to think about deriving the intervals statistically within a single step. It may be, therefore, that statistical understanding and techniques can help to improve the IPF spatial microsimulation approach given the connections between the two. Another implication is that some spatial microsimulation approaches may be more similar and better suited to connections with statistical approaches than other spatial microsimulation methods; combinatorial optimisation, for example, is about randomly drawing households and so is a different
type of approach to a more ‘deterministic’ IPF approach, with implications for potential linkages to
the spatial microsimulation approaches. Further works is needed to explore these issues more fully.

One interesting issue is that whilst IPF and log-linear modelling can be considered as equivalents
they differ in what it is that they each estimate. For log-linear regression the ‘model’ estimates the
values of the target variable itself whilst for IPF the ‘model’ estimates weights which, as an
intermediate step, are then used to estimate the target variable of interest at small area level.
 Whilst in many ways equivalents, therefore, it is not clear whether this difference is material in
terms of the point estimates or in terms of any distributional estimates of the target variables.

A further possible connection between statistical and spatial microsimulation approaches regards
the small area covariate data itself. In statistical approaches, if one wishes only to estimate the mean
of the target variable then all that is required is the mean of the explanatory variables in the model.
 If however one wishes to estimate distributions of the target variable then one also needs this for
the covariate data, hence a need for small area microdata. Census data are the obvious source. One
aim within spatial microsimulation approaches however is to create such small area microdata for
each small area and the group therefore discussed the extent to which spatial microsimulation might
play a role in providing detailed small area covariate data for such (eg quantile) statistical SAE
approaches. This was not considered to be a particularly fruitful path to follow however given that
the rolling down on the small area microdata within the spatial microsimulation step inevitably loses
and smoothes out much of the differing (and, indeed, random) variation in relationships across
space (given that it is based on a model) yet this is precisely what one wishes to pick out during the
next statistical SAE phase.

Related, a further potential link between statistical and spatial microsimulation approaches is in the
desire for distributional estimates of target variables rather than simply mean estimates. This can be
done by, for example, quantile regression approaches but can also be done within both iterative
proportional fitting and combinatorial optimisation. It is not clear, however, how these approaches
would differ in their results (if at all) and this warrants further attention in order to better
understand the conceptual linkages underlying these different methods. Following on from this,
whilst these differing statistical and spatial microsimulation methods can each estimate
distributional values on target variables (eg p10, p25, etc) they do so in different ways. Whilst spatial
microsimulation approaches pick up these estimates empirically from the reweighted microdata in
statistical approaches these distributional estimates are derived from model-based theoretical
distributions of the target variable for the small area. It is not clear, however, whether the two
approaches are conceptually or empirically equivalent. Furthermore, creating distributional
estimates in a statistical framework requires covariate microdata (typically census microdata) which
can be a restriction if not available or if all of the desired variables are not available; in a spatial
microsimulation approach however only small area aggregates are required, significantly reducing
the covariate data demands. Hence, an important question is whether the less data demanding
spatial microsimulation approach to distributional estimates results in the same (or similar)
estimates as the statistical approach? The group’s discussions raised doubts over the ability to
reliably estimate distributional values without microdata though this requires further testing both in
and of itself as well as within a broader process of seeking to understand the conceptual and
technical similarities and differences between these approaches.
Another issue relates to issues of getting at the tails of distributions. In statistical approaches, approaches tend to add ‘noise’ to our estimates in order to seek to get at the tails of distributions whilst this does not tend to happen in spatial microsimulation approaches. Yet if (some of) these methods can be expressed in a statistical framework then there seems the possibility of seeking to mimic the introduction of noise within, for example, IPF or CO in order to seek to incorporate the tails of the distributions more effectively into our estimates. This would involve a clear understanding of the conceptual and technical links between the methods so as, for example, to take this approach from log-linear modelling and apply it back onto IPF.

**Key specific issues for attention in coming years**

One focus of the network is to identify specific methodological priorities across each of the existing main methodological approaches which we feel require progress and which we encourage attention in upcoming years from amongst the small area estimation community in order to make further progress:

1. estimating variance and confidence intervals/credible intervals within spatial microsimulation approaches;
2. approaches for distributional estimates as well as point (typically mean) estimates;
3. constraint selection (eg number, order, survey size, correlations, conflicting benchmarks, local specificity);
4. starting weight selection and implications;
5. a better understanding of interactions between variables both in terms of which variables interact spatially the most and at what scales;
6. a better understanding of the potential roles and impacts of geodemographics to small area estimation (eg in terms of case selection or error terms);
7. maintaining access to key microdata to enable analyses, whether census data (eg ONS Titchfield) or administrative data (eg Secure Data Service);
8. covariate data needs and implications of changing context around both census and administrative data;
9. progress in estimating two-way outcomes.

**Identifying conceptual and technical linkages between the statistical and spatial microsimulation approaches**

As reflected in the discussions noted above, a larger aim of the network has been to seek to identify and clarify conceptual and technical connections between statistical and spatial microsimulation approaches. The aim in doing so is both to clarify and make explicit the similarities between superficially different approaches whilst, more fundamentally, to offer opportunities for methodological improvement through learning and cross-fertilisation. To progress these debates the network propose an exciting collective project drawing on the unique opportunity provided by the network through group’s conversations over the past year and harnessing the differing methodological expertise within the network.
The intention in this next empirical phase is to test many of the questions identified during the network’s conversations and outlined above. This will be done by applying each of our varying methodological approaches to small area estimation onto a shared population micro-dataset designed to be unbiased across the different methods. The aim will be to identify a single research question and estimation process at a set scale and for network members to each run their own methodological approaches in parallel and to compare results and computational run times. The aim in doing so is not only to compare the comparability of estimates (whether point, distributional or variance estimates) but through doing so to seek to identify the underlying conceptual and methodological similarities and differences between the different methods.

References
