STATISTICAL COMMISSION and Working Paper No. 13
ECONOMIC COMMISSION FOR EUROPE English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography
(Neuchâtel, Switzerland, 10-12 April 2000)

Topic (iv): Spatial analysis in a statistical context and disclosure control procedures

SPATIAL TRANSFORMATION METHODS FOR THE ANALYSIS OF GEOGRAPHIC DATA

Submitted by Imperial College of Science, United Kingdom[1]

Invited paper

ABSTRACT

Most statistical (e.g. socio-economic or demographic) data are collated and presented on the basis of administrative areas (e.g. the Nationales Unités Territoriales Statistique of the EU). These administrative areas are invariably geographically arbitrary, and may also vary substantially in size and shape from one region or country to another, as well as over time. For these reasons, they pose serious problems for attempts to map or interpret spatial patterns in statistical data, or to combine such data with other data sets (e.g. with environmental information). Fortunately, geographical information systems (GIS) offer the capability to address these problems. In recent years, a wide range of spatial analytical techniques have been developed in GIS, which allow the aggregation, disaggregation or spatial modelling of data. How well these various techniques work, however, and how to select the best available technique in any specific situation, is not always clear. This paper presents a series of examples and case studies, illustrating some of the techniques available; and, based on these, shows how they might be used, some of the issues involved in using them, some of the pitfalls to be avoided, and the potential benefits they offer to the analysis of statistical information.

SPATIAL TRANSFORMATION METHODS FOR USE WITH EUROSTAT DATA

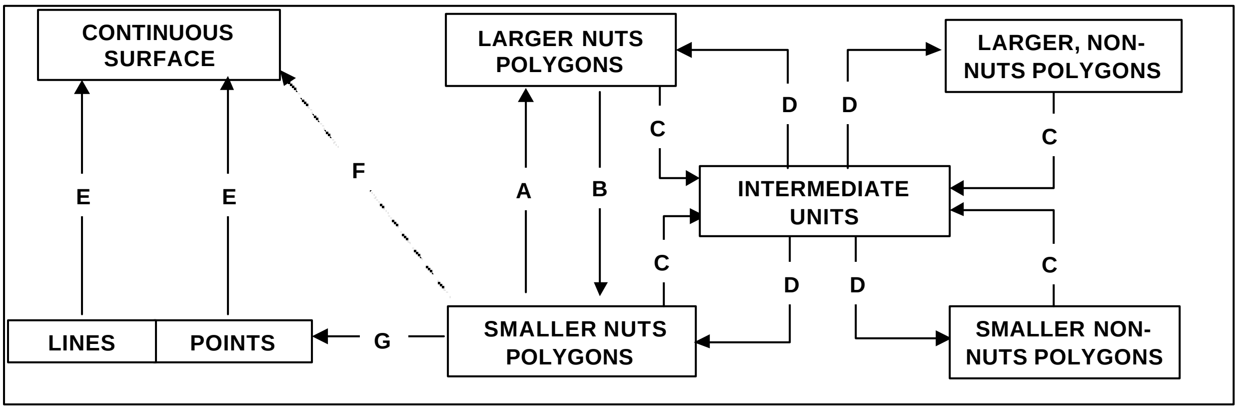
1. Many of statistical data used for policy purposes by Eurostat and other Directorates in the European Commission are, by their very nature, spatial in form. They relate to populations, activities, features and events which are associated with spatial locations, and which vary geographically. Managing, processing and displaying statistical data is therefore, largely, a spatial process.

2. The European Union – in common with its member states - uses an established system of statistical regions (the NUTS regions) which provides the framework for statistical reporting. For many statistical purposes, the NUTS framework is clearly robust and adequate. It has the benefits of being well-established, reasonably stable (though adjustments and changes do, inevitably occur, over time), hierarchical (such that areas at one level are subsets of those at higher levels) and generally well-matched to national statistical regions. Nevertheless, for many applications, the NUTS framework is far from ideal, as a result of what has been termed the 'modifiable areal unit problem' (MAUP). Problems arise because of differences in size between units at the same level - for example, NUTS 3 regions in Finland and in Belgium - and due to the fact that, in many cases, the administrative division of the NUTS units

[1] Prepared by David Briggs.

does not correspond to the spatial representation of the phenomena of interest (especially in relation to environmental data).

Figure 1. A generic model of spatial transformations for use with Eurostat data



Key: A = nested aggregation; B = nested disaggregation; C = non-nested disaggregation; D = non-nested aggregation; E = surface modelling (interpolation) based on line or point data; F = surface modelling (interpolation) based on polygon data; G = centroid creation.

3. As a result of these problems, there is an increasing need for methods to convert or transform data between different spatial structures in order both to help present the data in a more meaningful and consistent manner, and to enable different data sets, based on different geographical units, to be brought together and overlaid. The range of spatial transformations which might need to be undertaken is wide. They vary depending upon the spatial character of both the source data and the required results. Transformations may be made, for example, between any combination of point, line, grid and irregular polygon data. In general terms, however, transformations may be described as processes of aggregation or disaggregation (within either nested or non-nested polygons), or surface modelling on the basis of point, line or polygon data. Figure 1 presents a simple model of these various types of transformation, as applied to Eurostat-type data.

4. With the advent of GIS, an extremely wide range of spatial analysis methods has been developed for carrying out such transformations. The terminology and classification of these methods is not well established, but available methods include:

- Point in polygon processes;
- Areal weighting;
- Modified areal weighting using control zones;
- Modified areal weighting using regression relationships;
- Optimisation;
- Simulated annealing;
- Pycnophylactic interpolation;
- Weighted centroid smoothing;
- Polygon filtering;
- Smart interpolation; and
- Non-contiguous cartograms.

5. These different methods are based on different assumptions about the underlying spatial distribution of the data, and are subject to different types of error and approximation. The choice between them thus needs to be based upon a clear understanding of the methods involved. It should also take

account of the aims of the analysis, the quality and structure of the source data (e.g. the shape, size and number of spatial units or data points), the processing capability of the available software, and the requirements and expertise of the end-users.

6. In this presentation, a series of examples and case studies are presented to show how these different transformation methods might be used, some of the issues involved in using them, some of the pitfalls to be avoided, and the potential benefits they offer to the analysis of statistical information.

Table 1 shows how some of these methods can be applied to undertake the transformations shown in Figure 1.

Table 1. Matrix of transformation processes and methods available to achieve them.

Process	**A**	**B**	**C + D**	**E, F or G+E**
Example application	Convert NUTS 5 data to NUTS 3; Gowing new spatial units	Convert NUTS 3 data to NUTS 5	Convert NUTS 3 data to drainage basins	Create population density surface from NUTS unit data
Possible methods	a) Simulated annealing b) Optimisation	a) Areal Weighting b) Modified areal weighting c) Point-in-polygon	a) Modifed areal weighting	a) Pycnophylactic interpolation b) Weighted centroid smoothing c) Polygon filtering d) Smart Interpolation and other point-based methods