Accounting for data sparsity in forming spatially coherent zones

Kirsty Hassall, Richard Webster, and Alice Milne

Rothamsted Research, Harpenden, UK

Farmers have for centuries managed their land by dividing it into fields. They can often manage their land better by further dividing their fields into smaller 'management zones' according to the inherent fertility of the soil or actual crop performance over several years. They can then adjust the amount of fertilizer or other agricultural inputs to each management zone in accord with the conditions in the zone. Several properties of the soil or observations of yield are likely to contribute to a farmer's judgement and to the division of his fields into finer management zones. Formally, it is a problem in multivariate classification, with the additional constraint that the farmer wants spatially coherent classes, each of which he can treat as a single homogeneous management zone. Spatially weighted fuzzy k-means clustering has been successfully used to produce coherent management zones based on sensor data (e.g. EC, NDVI and yield monitor data). When data are sparse, however, this method can fail or results can be substantially less reliable. Here we propose several methods to cope with sparse data and compare these with data from arable fields.