Biomathematics and Statistics Scotland

Rob Kempton & Jim McNicol

Biomathematics & Statistics Scotland (BioSS) is devoted to the application of statistics and mathematics in agriculture, the environment and human health. Its core remit is to support the SEERAD programme of research, which is carried out within the SABRIs, SAC and RBGE. This is achieved through a dispersed group of statisticians, mathematicians and computing experts based at BioSS centres in Edinburgh, Dundee, Aberdeen and Ayr. A successful bid for increased core funding in 2001 allowed BioSS to expand its work in bioinformatics and systems modelling. A Bioinformatics Task Force was established to bring together BioSS researchers, consultants and trainers to address more effectively the problems created by the wealth of data emerging from new molecular technologies. This group will link with specialist bioinformatics staff being recruited in the SABRIs and contribute to a Scottish Bioinformatics Network involving Dundee, Edinburgh and Glasgow Universities. Extra resources are also being devoted to a new research theme, modelling complex systems and risk, which will integrate knowledge on the system sub-processes to predict whole system outputs. The theme will focus on how variability and uncertainty within the sub-processes affects system outputs and how results can be presented in a form more immediately usable by decision makers.

At SCRI there have been two developments in our genetic linkage work. An investigation was carried out into the effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Three locus-ordering criteria, weighted least squares, maximum likelihood and minimum sum of recombination fractions, were compared using a simulated doubled haploid population of size 150. Maximum likelihood was the most successful at ordering loci correctly but generated substantially inflated map lengths in the presence of typing errors. In general missing values created shorter map lengths for more widely spaced markers. Segregation distortion had little effect. The second development was to make publicly available the software for our tetraploid linkage map methodology. This is called TETRAPLOIDMAP and can be found at the BioSS ftp site.

In molecular sequence analysis, we have improved an earlier Hidden Markov model used to detect evidence of recombination in DNA sequence alignments of four sequences. Our approach explicitly models the sequence of phylogenetic tree topologies along a multiple sequence alignment. Inference under this model is done in a Bayesian way, using Markov chain Monte Carlo (MCMC). The algorithm returns the site-dependent posterior probability of each tree topology, which is used for detecting recombinant regions and locating their breakpoints. The algorithm has been programmed in C++. Work is in progress to develop a graphical user interface.

In ecosystem structures we have been examining the potential of grouping individuals according to their trait values rather than by their traditional taxonomic classification. In particular, comparisons over time of configurations of individuals looks to be a promising method of defining functional groups and the similarity of these configurations is conveniently quantified by the Mantel test.

The SEERAD funded Micronet project seeks to establish the nature of any spatial structure among soil microbial communities below an area of unimproved grazed grassland in the Scottish Borders. The inter sample distances ranged from 0.1 to 12m and a wide range of microbiological and chemical properties was measured. Geostatistical analysis revealed spatial dependence for a relatively small number of characters, but in particular for total C, total N, total P and microbial biomass C. The statistical analyses also suggested that much of the variation in microbial variables was present below the minimum scale of measurement of 10cm.