Biomathematics and Statistics Scotland (BioSS)

D.A. Elston & J.W. McNicol

Biomathematics & Statistics Scotland (BioSS) provides quantitative support for the SEERAD research programme. This is achieved through a dispersed group of statisticians, mathematicians and computing experts based at BioSS centres in Edinburgh, Dundee, Aberdeen and Ayr. The distributed staff structure fosters communication between BioSS's quantitative experts and researchers in other disciplines. BioSS staff support SEERAD-sponsored research projects led by scientists at the five SABRIs and SAC at many different levels, ranging from infrequent advisory sessions through to collaboration as equal partners. BioSS also has three research themes, namely Statistical Genetics and Bioinformatics (led by the BioSS group at SCRI); Spatial and Temporal Models; and Systems Modelling and Risk. These research themes allow BioSS to address quantitative problems arising from SEERAD's research programme at a generic level, leading to an improvement in the quality of research outputs from the Sponsored Bodies.

In molecular sequence analysis, we have produced user-friendly software for the detection of recombination. TOPALi is a Java graphical analysis application that allows the user to identify recombinant sequences within a DNA multiple alignment either automatically or via manual investigation. TOPALi allows a choice of three statistical methods, developed previously by BioSS, to predict the positions of breakpoints due to past recombination events. The breakpoint predictions are then used to identify putative recombinant sequences and their relationships to other sequences within the alignment. In addition to its sophisticated interface, TOPALi can import many sequence formats, estimate and display phylogenetic trees (using Neighbour-Joining or Bayesian approaches) and allow interactive analysis and/or automatic HTML report generation. TOPALi is freely available from the BioSS website.

We used multivariate analysis and geostatistical approaches to study microbial community structure in upland grasslands as part of phase 3 of the SEERAD funded Micronet project. We were able to establish that there were different drivers for rhizoplane bacterial and arbuscular micorrhizal (AM) fungal community development and that the scales of influence were different. Rhizoplane AM fungal community structure was mainly influenced by the associated plant type and by the composition of proximal vegetation (within 1 cm). Rhizoplane bacterial community structure was influenced by the composition of vegetation in close proximity (also within 1 cm) but the plant type to which the community was directly associated had little effect. We also showed that bacterial community structure was influenced by factors operating at larger scales (up to 6 m). This was not the case for AM fungi.

Simulation work on the use of linkage disequilibrium mapping in barley populations showed that the transmission disequilibrium test (TDT) is effective in locating genes controlling traits when specially constructed populations are used. These populations require the generation of heterozygote parents as would result from the cross of two inbred lines. The F2 from such a cross provides sufficient offspring with heterozygote parents to track the flow of marker alleles from one generation to the next. The TDT can then be used to highlight associations between markers and traits.

Experience with multiple testing of microarray target data and compounds from metabolic profiles led us to investigate methods of estimating the False Discovery Rate (FDR), one of the error rates associated with tests of significance. A new method, based on modelling the density function of significance values as a Uniform+Beta mixture distribution, was developed and is now incorporated into routine analyses of ‘omics data.

Collaborating with SCRI scientists, BioSS co-developed a comprehensive decision-support web application to assist Dutch farmers in potato cultivar selection, with the primary emphasis on Potato Cyst Nematode (PCN) management. Mathematical models of PCN population dynamics were modified to take account of the nature of current PCN infection in the Netherlands. Farmers interacting with the website supply historical details of field PCN levels which act as a reference baseline for the models, which ultimately provide them with predictions of future PCN population densities and potato yields in both graphical and textual formats.