Biomathematics and Statistics Scotland

David A. Elston

Biomathematics & Statistics Scotland (BioSS) provides specialist quantitative support for all four of SEERAD's research programmes. This breadth of purpose distinguishes BioSS from the other elements of the SCRI Group, and is achieved through a dispersed group of statisticians, modellers and computing experts based at BioSS centres in Edinburgh, Dundee, Aberdeen and Ayr. Hence the science in SEERAD's main research providers benefits from proximity with BioSS staff, whilst BioSS staff benefit from being part of a substantial grouping of fellow professionals.

A central theme in 21st century scientific development is increasing quantification, driven by increases in computer processing power, in the rate of collection of data and in the size and connectivity of databases. By employing statisticians, modellers and computing experts at centres of scientific research throughout Scotland, BioSS is perfectly positioned to recognise and address the new quantitative challenges posed by emerging technologies. Almost all scientific staff undertake methodological development (BioSS research work) while working closely with non-mathematical scientists to address scientific problems using existing methodologies (BioSS consultancy work). This approach maximises effectiveness by ensuring both research focus and up-to-date consultancy practice.

BioSS manages its consultancy work under the four broad scientific areas:
- plant science;
- animal health and welfare;
- ecology and environmental science;
- human health and nutrition.

In each area, BioSS staff have a wide range of different types of interaction with scientists, ranging from the provision of short pieces of advice that allow BioSS expertise to guide a large number of scientific research projects, through to deep collaborative relationships that allow BioSS staff to make substantial inputs to a small number of scientific projects in which quantitative issues play a central role. Whilst the following highlights describe collaborative interactions, the cumulative importance of the advisory inputs should not be underestimated.

Our consultancy work for plant science supports scientists investigating the genetics of important Scottish crops such as potatoes, barley and soft fruit, the diseases that inhibit production of these crops and the interactions between arable crops and the environment. Improved plant varieties can be selected more quickly using molecular markers shown to be linked to genes for key traits, such as gall mite resistance in blackcurrant. Linkage maps have been developed in collaboration with SCRI for blackcurrant and raspberry, both outbreeding species in which breeding populations may be complicated by occasional self-fertilisation of parental plants. BioSS has developed methods for mapping QTLs for traits such as root rot that are recorded as repeated categorical measurements.

Our consultancy work for animal health and welfare supports scientists investigating the bacterial, viral and parasitic diseases of farm animals, husbandry-related welfare issues and the interaction of agricultural practices in animal farming systems.   BioSS has recently collaborated with the Moredun Research Institute in the design and analysis of a survey of sheep farmers, investigating the disease and welfare problems which they believe have the most impact on flocks. The identification of diseases such as sheep scab and problems such as anthelmintic resistance, being of particular relevance to Scottish farmers, will allow better targeting of scientific effort within the animal health and welfare programme.

Our consultancy work for ecology and environmental science supports scientists investigating soil health, water quality, and ecological communities, with the aim of understanding and managing the interactions between the rural economy and the environment. The impacts large herbivores have on extensive tracts of upland vegetation have been assessed in collaboration with the Macaulay Institute. The BioSS contributions to this collaboration have included development of appropriate sampling and data collection protocols and the development under our research programme of methods to analyse spatially distributed, ordinal data. Of particular interest is the relative importance of sheep and red deer on impacts within and between regions.

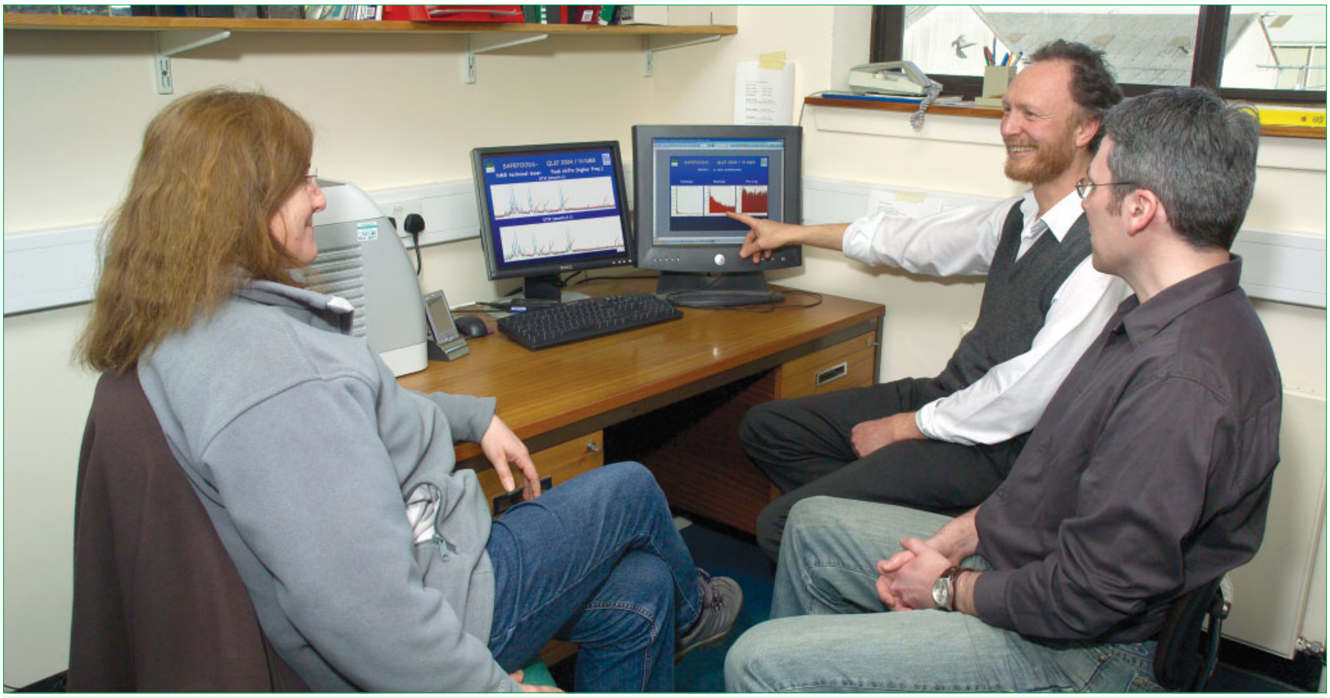
Our consultancy work for human health and nutrition supports scientists investigating chronic dietary-related conditions such as obesity, cardiovascular health and gut health. A recent development is the large amount of proteomic data being generated in order to understand the effects of diet on protein expression, and both univariate and high dimensional multivariate statistical methods are important in efficient interpretation of these data. In gut health research, the use of stable isotopes is important in tracing the metabolic pathways and fluxes of the different products of colonic bacterial activity. BioSS's mathematical and statistical expertise has been instrumental in developing compartmental models and in estimating flow rates from stable isotope data.

BioSS manages its programme of applied strategic research in three broad themes:

- statistical genomics and bioinformatics;
- systems and process modelling;
- statistical methodology.

The research we carry out addresses generic issues encountered in our consultancy work that are not adequately addressed using standard methods. Each research theme is related to each of the four broad scientific areas, demonstrating the wide applicability of BioSS research.

The statistical genomics and bioinformatics research theme addresses a wide range of modelling and data analysis problems that can be grouped under the six headings of genetic association, genomics, transcriptomics, proteomics, metabolomics and regulatory networks. Current scientific efforts to develop more efficient strategies for disease and pest control in crop plants require an understanding of the interaction between signalling pathways in hosts and pathogens. Our work on regulatory networks offers the prospect of improving estimation of sub-cellular interactions, lead-



ing to more effective control of sub-cellular processes, by combining high-throughput postgenomic data, e.g. promoter sequences, gene expression profiles from microarrays, and putative protein–protein interactions. In particular, BioSS has been contributing to the development of statistical methods for the robust inference of sub-cellular regulatory networks in the presence of biological stochasticity and noisy data.

The systems and process modelling research theme consists of methodological activities such as analysis, simplification and approximation of stochastic process models motivated by four classes of applications, namely: modelling large populations in heterogeneous environments; modelling risks to biodiversity in a changing environment; complex interactions in epidemic processes; and Bayesian methods for decision support. This year has seen a particular emphasis on species distribution modelling. One issue addressed by BioSS has been to estimate spatial variation of recorder efforts in species atlas data in order to produce modelled species distributions that are standardised for recorder effort. Another issue addressed by BioSS is the development of spatio–temporal models for the spread of invasive species allowing for environmental heterogeneity and for stochastic colonisation events whose probability decreases with distance from the existing population.

The statistical methodology research theme involves empirical modelling of data that are correlated due to the effects of space and/or time, extracting information from digital images, along with diverse methods such as compositional analysis and latent Gaussian modelling which need developing to resolve problems encountered in our consultancy work. One application arises from the SAC–BioSS CT Unit, which uses X-ray computed tomography to estimate body composition of sheep. This necessitates the positioning of segmentation boundaries on images in order to identify the tissues such as bone, fat and muscle. If images are 2-dimensional, dynamic programming (DP) is a fast, elegant method for achieving this. We have developed a generalisation of DP, a greedy algorithm termed *iterated dynamic programming*, for segmenting 3-dimensional images. The method has been implemented in BioSS's STAR software, and is used routinely in sheep trials.

More details about the work of BioSS, including a wider range of examples of our work and a description of associated knowledge exchange activities and software products, can be found on the BioSS website www.bioss.ac.uk.