Biomathematics and Statistics Scotland

David A. Elston

Biomathematics & Statistics Scotland (BioSS) is a specialist organisation delivering high-quality consultancy, training and research in statistics, mathematical modelling and bioinformatics. BioSS forms a distinctive element of SCRI Group and plays a unique role in the Scottish research community, bridging the gap between research in the mathematically-based and traditionally more qualitative sciences such as biology.

Many facets of the BioSS's scientific work are encapsulated in the new BioSS Vision statement:

" *to improve science & society through an understanding of variation, uncertainty and risk.*"

We enhance understanding through process modelling and data analysis, as well as through the design of experiments and observational studies. Variation can be of a deterministic nature, or partitioned into uncontrolled components that we often think of as being random. Uncertainty refers to our state of knowledge, which is always clouded by both measurement error and natural variation. At a probabilistic level, risk refers to the integration of variation and uncertainty, with attention often focusing on the chance of extreme events. Much of our work is directed towards scientific objectives, but with a widening range of outcomes including providing an evidence base for improved policy making and contributing to economic activity through the levy boards and private sector organisations.

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, environment–plant interactions and the quality of plant-derived foods and other plant products. BioSS has recently developed a user-friendly program, TOPALi, to allow faster, more sophisticated statistical analyses of the evolutionary relationships among molecular sequences. This software includes novel recombination detection methods developed by BioSS and implements modern Bayesian methods of inference for phylogenetic analysis. It is currently being used to study microbial soil diversity, nematode taxonomy and relationships among members of multi-gene families.

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 Scottish Agricultural College in the design and analysis of a survey of the prevalence of verocytotoxic *Escherichia coli* in the faeces and digestive tract of sheep sent for slaughter at Scottish abattoirs. This work has built on 10 years of collaboration with SAC veterinary epidemiologists in exploring the dynamics of this zoonotic infection in Scottish livestock, and has drawn on research carried out in BioSS in the statistical properties of zero-inflated models of infection prevalence in the presence of random effects. The results will be used by the Food Standards Agency in Scotland to inform future policy decisions relating to meat microbiological safety.

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. In collaboration with The Macaulay Institute, we have

begun studying daily flow data for the River Dee dating back to 1929 to look for long-term changes in the distribution of flows, including changes in seasonal patterns and the size and frequency of unusual events. BioSS staff are employing their skills in time series analysis and modelling of extremes, in terms of both high flows (causing flooding) and low flows (when the highest concentrations of chemicals occur).

Our consultancy work for *human health and nutrition* supports scientists investigating chronic dietary-relations conditions such as obesity, cardiovascular health and gut health. Microarray studies have become a core part of this research. BioSS has collaborated with normalisation of cDNA microarrays, and analyses based on significance tests, controlling the false discovery rate, to find genes that were differentially expressed in hypothalamic samples from animals used as a model of human weight regulation. Our research collaboration in studying the bacterial composition of the digestive system has also been extended. Experiments show that the bacteria populations, and their metabolic products, change substantially in response to the macronutrient composition of the diet. We have begun developing mathematical models to further investigate the interactions between diet, bacteria, and their metabolic products such as butyrate, which are thought to play a role in the prevention of colon cancer.

BioSS manages its programme of applied strategic research in three broad themes:
- *statistical bioinformatics;*
- *systems & 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 our four broad scientific application areas (below), demonstrating the wide applicability of BioSS research.

The statistical 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. Our understanding of how important field traits such as disease resistance and yield are controlled has been greatly enhanced by mapping quantitative trait loci (QTLs) and establishing the relationship between traits and the QTLs. BioSS has produced the only available software, TetraploidMap, for mapping QTLs for tetraploid species such as potato. We have developed a simulation approach to allowing the effect of population structure to be taken into account in the identification of QTLs for barley. Also, we have been involved in mapping QTLs for soft fruit such as blackcurrant and raspberry. All of this work puts us in a strong position to help SCRI develop improved varieties of valuable crops.

The systems & 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 the formation of a new Centre of Excellence in Epidemiology, Population Health and Infectious Disease Control (EPIC), funded by the Scottish Government. This collaborative centre also incorporates five other internationally recognised research organisations, highlighting the strength of the Scottish research base in veterinary epidemiology. BioSS's role in EPIC is to develop, evaluate and apply methods of statistical inference for models of diagnostic testing and infection dynamics in animals. Given recent

outbreaks of foot-and-mouth disease and bluetongue, the establishment of the centre has been very timely, and it is hoped that EPIC will soon contribute to governmental, industrial and public understanding of issues relating to the control of epizootic and enzootic infections in animal production systems.

The statistical methodology research theme involves empirical modelling of data that are correlated due to the effects of space 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 current project which has reached fruition this year has involved embedding the simplex, in which compositional data must lie, in an unconstrained space. Data are regarded as being generated from a Gaussian distribution in the unconstrained space, transformed so that points lying outside the simplex are mapped onto the simplex boundary and represent observed zeros in compositional space. This methodology has wide application to multivariate compositional data, including the analysis of daily food intake data which is both high-dimensional and contains a substantial proportion of observed zeros.

Further information about the work of BioSS can be found on the BioSS website www.bioss.ac.uk. Details of particular interest include a wide range of examples of our work, a description of associated knowledge exchange activities and access to our software products.