15th European GenStat and ASReml Applied Statistics Conference

Conference Abstracts
Rothamsted Research
July 2010

GenStat®

ASReml®

Conference Programme

0900 - 0930	Registration and Coffee	
0930 - 1100	Session 1	Chair: Sue Welham
0930	Robin Thompson	REML past, present and future
1000	Freedom Gumedze	An alternative approach to outliers in meta-analysis
1030	Salvador Gezan	Genetic analysis of tree clonal trials: combining individual and clonal information into ASReml
1100 - 1130	Coffee	
1130 - 1300	Session 2	Chair: Darren Murray
1130	Fred van Eeuwijk, Marcos Malosetti & Martin Boer	QTL mapping in a mixed model perspective within GenStat
1200	Marcos Malosetti	Modelling substructure and pedigree relations in association mapping using GenStat
1230 - 1300	Poster introductions	
1300 - 1400	Lunch and Poster Presentations	
Suzanne Clark & Rebecca Nesbit		Analysis of butterfly migratory flights
Katrin MacKenzie & Christine Hackett		Association mapping in a simulated barley population
Zaneta Park-Ng, David Baird, Wendy Collier, Nicole More & Alan Hart		Discrimination among milks and cultured dairy products using screen-printed electrochemical arrays
Stephen Powers, Sue Welham & Jemma Salmon		Spatial analysis of number of stems from four populations of willows grown in a single experiment
Osman M. Siraj, S. Ali Eltahir & M. Singh		Comparing statistical software for outputs on analysis of variance in agricultural experiments
Julia Sibiya, John Derera & Pangirayi Tongoona		Grain yield stability analysis of African maize hybrids in multi-environment trials based on the GGE biplot in GenStat
Rodger White & Sarah Kemmitt		Phospholipid fatty acids and analysis of distance for compositional data
1400 - 1530	Session 3	Chair: Simon Harding
1400	Andrew Mead & David Baird	Design and analysis for 2-channel microarray experiments - implementing MAANOVA in GenStat
1430	Sue Welham, Jun Wang, Graham King & Wally Gilks	Three-point appraisal of genetic linkage maps
1500	Phil Brain & Magnus Ivarsson	Detecting drug-related signatures in EEG data using extended semilinear canonical correlation analysis (ESLCCA)
1530 - 1600	Tea	
1600 - 1700	Session 4	Chair: Roger Payne
1600	Granville Tunnicliffe Wilson	What cross-spectral analysis can do for you
1630	Alex Glaser	The use of biplots in statistical analysis

Detecting drug-related signatures in EEG data using extended semilinear canonical correlation analysis (ESLCCA)

Phil Brain & Magnus Ivarsson
Pfizer Global Research & Development,
Ramsgate Rd,
Sandwich,
Kent CT13 9NJ,
UK.
Email: Phil.Brain@pfizer.com

There is a need to non-invasively and reliably detect whether drugs aimed at targets in the Central Nervous System (CNS) cross the blood/brain interface and have an effect. However it is not possible to directly assay drug levels in the CNS, so indirect methods are needed. One of these is the electroencephalogram (EEG), which provides a non-invasive method of detecting brain activity. Typically time slices of the EEG signal are subject to a Fourier Transform which gives a periodogram of the signal versus frequency at a sequence of time points.

The SLCCA algorithm (Gregg, Varvel & Shafer 1992) is capable of detecting a frequency "signature" in these periodograms which is correlated with the imputed dose level of a drug in the brain. The approach uses Canonical Correlation Analysis to detect the signature which maximises the correlation between the EEG periodogram and the imputed dose obtained from a PK model with unknown nonlinear parameters, where the unknown nonlinear parameters of the PK model are estimated by maximising the Canonical Correlation. We have extended the previous algorithm to include regression models with linear and nonlinear parameters so that a range of complex models can be readily fitted. We intend to make the algorithm generally available as a GenStat procedure in due course. We will present the algorithm and illustrate its use in a regression context on experimental EEG data from rats.

Gregg, K.M., Varvel, J.R., Shafer, S.L. (1992). Application of semilinear canonical correlation to the measurement of opioid drug effect. *Journal of Pharmacokinetics and Pharmacodynamics*, **20**, 611-635.

Analysis of butterfly migratory flights

Suzanne J. Clark & Rebecca L. Nesbit
Rothamsted Research, Harpenden, Herts AL5 2JQ, UK.
Email: suzanne.clark@bbsrc.ac.uk

1. Background

Many insects are known to migrate long distances to exploit seasonally available habitats. The painted lady butterfly (*Vanessa cardui*) over-winters in North Africa and migrates to northern Europe in spring. The use of the sun in orientation has been demonstrated in various insects, including bees, beetles and ants.

2. The scientific questions

Do painted lady butterflies

a) preferentially orientate northwards during spring migrations?
b) use a sun compass to select and maintain their preferred flight headings?

3. The experiments and data

In spring 2006 and 2008 migrating adult painted ladies were caught in southern Gibraltar. Individuals were flown in a computerised flight simulator which recorded the butterfly's heading every 200ms during a 15 min flight. Mean directions from 50 directed flights (i.e. those representative of migratory behaviour) were available for analysis. The sun position during each tethered flight was also recorded.

4. The analysis

Methods for comparing means of, and associations between, circular variables were required. Such methods are well-documented but were not yet available directly in GenStat. Two new GenStat procedures, CASSOCIATION and CCOMPARE, were therefore written. These extend and complement the existing circular data procedures CDESCRIBE, DCIRCULAR, RCIRCULAR and WINDROSE.

5. The new procedures

CCOMPARE provides tests for whether circular samples have a common mean direction and/or whether the distributions are identical.

CASSOCIATION calculates measures of association and six different test statistics for circular data.

a) linear-circular data: D_n - test for presence of C-association; λ_n - assesses extent of C-association; R_n^2 - assesses extent of C-linear dependence
b) circular-circular data: Δ_n - circular correlation coefficient quantifying T-monotone association; π_n - circular correlation coefficient based on circular ranks; ρ_T - test for T-linear association

All code and notation was based on Fisher, 1983.

6. Selected results

The mean flight direction of Gibraltar spring-caught butterflies was similar in both years (282° and 223° in 2006 and 2008, respectively, P = 0.07). However, the average flight direction over the two years combined (261°) was more westerly than northerly as expected. The butterflies' headings and the position of the sun were uncorrelated (ρ_T = 0.12, P = 0.33).

Fisher, N.I. (1983). *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, UK.

Nesbit, R.L., Hill, J.K., Woiwod, I.P., Sivell, D. & Chapman, J.W. (2009). Seasonally-adaptive migratory headings mediated by a sun compass in the butterfly *Vanessa cardui*. *Animal Behaviour Volume*, **78**, 1119-1125.

Acknowledgement: Rothamsted Research is an Institute of the Biotechnology and Biological Sciences Research Council (BBSRC).

QTL mapping in a mixed model perspective within GenStat

Fred van Eeuwijk, Marcos Malosetti & Martin Boer
Biometris,
Wageningen University,
Postbus 100,
6700 AC Wageningen,
Netherlands.
Email: Fred.vanEeuwijk@wur.nl

The mixed model provides a convenient framework for the analysis of field trials. QTL mapping can be understood as a straightforward extension of such mixed model analyses upon the realization that QTLs can be interpreted as marker based contrasts between genotypes. Adapting the mixed model approach to QTL mapping, facilitates extensions of QTL mapping technology to more complex applications like multi-environment and multi-trait QTL mapping that are hard to achieve following more traditional approaches to QTL mapping. Similarly, widely different types of genetic and breeding populations can be dealt with in a mixed model setting. GenStat has a strong tradition in mixed modelling. The mixed model QTL tools build on that tradition.

Genetic analysis of tree clonal trials: combining individual and clonal information into ASReml

Salvador A. Gezan
Department of Statistics,
Institute of Food and Agricultural Sciences,
McCarty Hall C,
P. O. Box 110339,
University of Florida,
Gainesville, FL 32611-0339
USA.
Email: sgezan@ufl.edu

In the last few years several tree improvement programs are beginning to collect large amounts of information from clonally replicated progeny tests. Traditionally half or full-sib trials have been established where the statistical analysis partitions the genetic variance into additive and dominance effects. These analyses do not only rank parents but also families allowing to capture some of the dominance variability. However, with clonal trials it is possible to go even further and obtain estimates of epistatic variance. Therefore, an even larger portion of the genetic variability can be captured by selecting outstanding clones. In the present talk we described the type of data commonly used and the details of a complex model used within ASReml that allows combining information from several sources including trials that contain only seedlings (i.e. unreplicated genotypes), clonal propagules (i.e. replicated genotypes) and a combination of both. Also, the expression used to link the estimated variance components from the linear mixed model to the genetic components are presented together with error estimates.

The use of biplots in statistical analysis

Alex I. Glaser
VSN International,
5 The Waterhouse,
Waterhouse Street,
Hemel Hempstead,Herts HP1 1ES,
UK.
Email: alex@vsni.co.uk

Biplots, which were introduced by Gabriel (1971), provide a simultaneous graphical display of both the samples and variables of a data matrix in a single plot, with the *n* samples represented by points, and the *p* variables incorporated as "axes" (or vectors). There have been numerous developments involving biplots since the original work, most notably by Gower & Hand (1996), who expanded the concepts of the original paper in their seminal monograph. Although both Gabriel (1971) and Gower & Hand (1996) use the same mathematical principles, their visualisation techniques differ. These differences will be demonstrated, and the advantages and disadvantages of each system discussed.

Biplots will be shown for various statistical analyses in GenStat, illustrating, in particular, how the use of GenStat's graphical hotpoints can highlight various aspects of the display.

Gabriel, K.R. (1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, **58**, 453-467.
Gower, J.C. & Hand, D.J. (1996). *Biplots*. Monographs on Statistics and Applied Probability. Chapman and Hall, London, UK.

An alternative approach to outliers in meta-analysis

Freedom Gumedze
Department of Statistical Sciences,
University of Cape Town,
Rondebosch, 7701,
South Africa.
Email: freedom.gumedze@uct.ac.za

Meta-analysis involves the combining of estimates from independent studies on some treatment in order to get an estimate across studies. However, outliers often occur even under the random effects model. The presence of such outliers could alter the conclusions in a meta-analysis. This paper proposes a methodology that detects and accommodates outliers in a meta-analysis rather than remove them to achieve homogeneity. An outlier is taken as an observation (study result) with inflated random effect variance, with the status of the ith observation as an outlier indicated by the size of the associated shift in the variance. We use the likelihood ratio test statistic as an objective measure for determining whether the ith observation has inflated variance and is therefore an outlier. A parametric bootstrap procedure is proposed to obtain the sampling distribution for the likelihood ratio test and to account for multiple testing. We illustrate the methodology and its usefulness using three meta-analysis data sets from the Cochrane Collaboration.

Modelling substructure and pedigree relations in association mapping using GenStat

Marcos Malosetti
Biometris,
Wageningen University,
Postbus 100,
6700 AC Wageningen,
Netherlands.
Email: marcos.malosetti@wur.nl

Association mapping or linkage disequilibrium mapping (LD) is a method for QTL (Quantitative Trait Locus) detection widely applied in human genetics. Essentially it consists of finding marker-trait associations in genetically diverse populations. LD mapping has recently gained attention in plant genetics. An important asset of association mapping is that there is no need to develop specific crosses as it can take advantage of the use of diverse collections of genotypes. In addition, association mapping can target a broader and more relevant genetic spectrum for plant breeders than conventional QTL mapping does. However, significant marker-trait association may or may not be the consequence of physical linkage between markers and QTLs (false positives or 'spurious' associations). A major cause of false positives is the genetic correlation between individuals stemming from the population genetic structure. The success of any LD mapping approach relies on statistical tests for marker-trait association that take into account the complexity in genetic covariation in the data. Several strategies have been proposed to account for population structure, either by structuring the population and imposing the groupings in the statistical model (Pritchard *et al.* 2000, Kraakman *et al.* 2004) or by using estimates of genetic relatedness between individuals, coancestry information, in the mixed model (Yu *et al.* 2006, Malosetti *et al.* 2007). An interesting intermediate approach is that based on principal component analysis ideas proposed by Patterson (2006). Taking advantage of GenStat facilities for mixed modelling, we have implemented a number of procedures for QTL detection in structured populations that uses several LD mapping methods. In this presentation the different methods and GenStat procedures will be described and demonstrated using real and simulated data sets.

Kraakman, A.T.W., Niks, R.E., Van den Berg, P.M.M.M., Stam, P., Van Eeuwijk, F.A. (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, **168**, 435-446.

Malosetti, M., van der Linden, C.G., Vosman, B., van Eeuwijk, F.A. (2007). A Mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*, **175**, 879-889.

Patterson, N., Price, A.L., Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet*, **2**, e190.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet.*, **67**, 170-181.

Yu, J., Pressoir, G., Briggs, W.H., Vroh, Bi. I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, **38**, 203-208.

Association mapping in a simulated barley population

Katrin MacKenzie & Christine Hackett
BioSS,
Invergowrie,
Dundee DD2 5DE,
UK.
Email: Christine.Hackett@scri.ac.uk

The analysis of association mapping populations is frequently complicated by substructure within the population. If this is unaccounted for, it generally results in many false positive marker-trait associations. In this study, we simulate a large barley population, modelling inbreeding and selection, and compare three models for analysing marker-trait associations. One of these includes no population substructure, while the others use two methods that have been proposed for estimating marker-based kinship. These models can all be fitted using GenStat's mixed model directives. Both models including kinship reduced the number of false positives substantially compared to a model with no population substructure, but neither approach had a clear advantage over the other. One solution is therefore to fit both kinship models, and to consider as candidate associations any markers that are significant by both models.

Design and analysis for 2-channel microarray experiments − implementing MAANOVA in GenStat

Andrew Mead
Warwick HRI,
University of Warwick,
Wellesbourne,
Warwick CV4 7AL,
UK.
Email: Andrew.Mead@warwick.ac.uk

&

David B. Baird
40 McMahon Drive,
Aidanfield,
Christchurch,
New Zealand.
Email: david@vsn.co.nz

With microarray technology becoming a common experimental tool, it is becoming increasingly important for statisticians to be involved in designing experiments to make efficient use of these costly resources. Two-channel systems pose an interesting challenge when large numbers of structured treatments are to be compared, with relatively little information available in the statistical or microarray literature. In the first part of this presentation the first author will briefly talk introduce some pragmatic approaches being taken in research programmes at Warwick.

Having generated large and valuable datasets from such experiments, it is vital to be able to extract the important results. As with most experiments, there is the potential for errors, but the quantity of data generated appears to be a barrier to the data quality checking that is common in most other application areas. Recent work at Warwick has focussed on the development of functions within the MAANOVA library in R to semi-automate such data quality checking for microarray datasets, enabling the user to focus on possible problems and re-generate or correct data (as appropriate) rather than wasting time checking data for each individual array or probe. The second part of this presentation will describe the functions modified and developed for this data quality checking, also introducing the analysis approach that forms the core of the MAANOVA library. This appears to be a fairly novel approach within microarray analysis, incorporating the normalisation step into the model fitting, and providing estimates of treatment effects rather than of the ratios of treatments (therefore being more similar to the approaches taken for most field, glasshouse and laboratory experiments). Such expression profile data are also essential for building gene network models.

Another very recent development at Warwick is the fitting of non-linear regression models to describe gene expression response over a time course, using the identification of different shapes of response, and variation in the fitted parameters to group genes. Initial results from this work will be briefly described.

The final part of this presentation will describe (and demonstrate) how the MAANOVA analysis approach is being implemented in GenStat, hopefully using experimental data from Warwick to illustrate some of the features of the implementation.

Discrimination among milks and cultured dairy products using screen-printed electrochemical arrays

Zaneta Park-Ng, David Baird, Wendy Collier[+], Nicole More & Alan Hart
AgResearch Ltd,
Grasslands Research Centre,
Tennent Drive,
Private Bag 11008,
Palmerston North,
New Zealand.
Email: david@vsn.co.nz

Whether or not electrochemical arrays constructed by screen-printing were able to distinguish consistently between sets of milk and/or cultured dairy products was determined using discriminant analysis. Such arrays can be manufactured relatively inexpensively and thus, if able to give replicable results, they could be treated as disposable, thus avoiding the necessity to polish or electrochemically clean arrays between successive samples.

The arrays consisted of four electrodes: platinum, palladium, rhodium and a proprietary mediator. Using differential pulse voltammetry, 220 measurements (0 to 1100mV, in steps of 5mV) were made for each electrode, resulting in 880 readings for each array. Using forwards stepwise canonical variates analysis followed by cross-validation, this original set of measurements was reduced to an optimal set of variates that enabled maximum discrimination between groups. Cross-validation errors were calculated by randomly assigning half of the replicate measurements for each sample type to a training set, and calculating the percentage error in allocating the unknowns to their correct sample groups. 1000 such simulations were done.

Three sets of products were tested: a) four milks; b) four yoghurts, and c) two cultured products (Tararua cultured buttermilk and Fresh ’n ’Fruity natural yoghurt) versus two non-cultured (Tararua Full Cream milk and Tararua CalciTrim milk). Fifteen electrodes were tested using each product. The cross-validation errors were very low for the cultured versus non-cultured (0%), and yoghurts (1.43%), but higher for the milks (12.75%). Using only the rhodium electrode data gave very similar errors, indicating that a multi-membered array may not be required.

These results are promising, although further research is required on using training sets from different days.

([+] sadly Wendy Collier passed away after this work was completed)

Spatial analysis of number of stems from four populations of willows grown in a single experiment

Stephen J. Powers, Sue J. Welham & Jemma Salmon
Rothamsted Research,
Harpenden,
Herts AL5 2JQ,
UK.
Email: stephen.powers@bbsrc.ac.uk

Four populations of willows, denoted A, B, C and D are grown in an experiment at Rothamsted. There are 400 genotypes for population A, 480 for B and C, and 541 for D. Unreplicated plots of these genotypes are planted with 6 stools per plot, and are arranged in a 60 (rows) by 40 (columns) rectangular design. Within this, 8 by 12 plot sections are planted with a particular population, so that the design is formed as a 5 by 5 set of such "population sections". Due to the unequal numbers of genotypes in each population, there were 6 sections for population A, B and C, and 7 sections for D. Vertically, five population sections can be seen to comprise a block in the design. Also grown in these sections are the parents of the populations: P1 (48 plots) and P5 (30 plots) for population A, P2 (41) for B, P3 (39) for C and P4 (45) for D. Control plots of two standard genotypes (denoted C1 and C2) are strategically planted throughout the design to assist in the detection of spatial trends.

The trial was planted in 2008, so as yield and other physiological measures are generally taken only after three years of growth, for initial inspection in spring 2010, the numbers of stems were counted on all 6 stools per plot in the first two blocks of the design. This forms a 60 (rows) by 16 (columns) plots rectangle, with the first block having two sections for population C (and one for each of the others), and with the second block having two sections for population D. This gave information for 160 genotypes for populations A and B, and 240 genotypes for C and D. Also, there were 16 P1, 14 P2, 19 P3, 20 P4 and 10 P5 plots, along with 30 C1 and 51 C2 plots in this part of the design.

Initial inspection and modelling revealed that a Normal distribution could be assumed for these count data on the square root scale. Linear mixed modelling was used to estimate variance and (broad sense) heritability for each population in the context of a single experiment, with a common residual variance. The model incorporated first order autoregressive (AR1) components to account for the spatial variation along rows and columns, and a spline term to account for the particularly strong trend along rows. Mean values for all the genotypes, having accounted for spatial variation, were predicted to allow comparison within and between populations and for future input to quantitative trait loci (QTL) analyses. GenStat® (2009, 12th Edition, VSN International Ltd., Hemel Hempstead, UK) was used for this analysis employing the VCOMPONENTS, VSTRUCTURE and REML procedures for the modelling and the F2DRES and TRELLIS procedures for checking residuals. VPREDICT was used for predicting means.

Grain yield stability analysis of African maize hybrids in multi-environment trials based on the GGE biplot in GenStat

Julia Sibiya, John Derera & Pangirayi Tongoona
African Centre for Crop Improvement,
School of Agricultural Sciences and Agribusiness,
University of KwaZulu-Natal,
P. Bag X01,
Scottsville,
Pietermaritzburg 3209,
South Africa.
Email: Dereraj@ukzn.ac.za

The highly variable environments in sub-Saharan Africa (SSA) contribute to complicated genotype × environment interactions (GEI). Multi-environmental trials (METs) would therefore be useful to identify superior varieties that can be recommended to the farmers. The objectives of the study were therefore to evaluate the level of grain yield stability, and identify the best performing genotypes for wide and specific adaptation in different African environments ranging from 0° to 30°S latitudes. Forty five F1 maize hybrids were generated by crossing ten inbred lines in a half diallel mating scheme. The 45 hybrids along with nine hybrid checks were evaluated across 11 environments varying in moisture regimes and disease levels, with two replications each between 2007 and 2009. The genotype and genotype by environment (GGE) biplot analysis procedure in GenStat was used in identifying the superior genotypes. From the polygon view, the 11 environments appeared only in four of the sectors implying different high yielding cultivars for those sectors and a cross-over genotype × environment (GE) interaction. This suggested that the test environments could be divided into four mega-environments. The hybrids with the highest yield for the environments that were represented by the vertex families for each of these sectors were H31 (N3 × A16), H6 (CML445 × A16), H14 (A1220-4 × A16) and H18 (CZL00009 × CZL00001). The most representative environments were C108, C208, C09, KD09, RA09 and ZAM08 as they had PC1 and PC2 scores close to zero, indicating stability and high grain yield. Sub-dividing the environments based on their locations (latitudes) showed the South African environments (Cedara and Baynesfield) to be different from each other. The environments north of latitude 18° had some overlaps indicating some similarities among them. In this group, RA08 and ZAM09 (from Zimbabwe and Zambia, respectively) behaved differently from the other environments. Environment RA08 was confirmed to be unstable and low yielding based on the high PC2 score and a low PC1 score (closer to zero), respectively. The ZAM09 environment, on the other hand, had a lower PC2 score, indicating relative stability and a high PC1 score indicating high yield. Overall, the GGE biplots gave more visual interpretations than just selecting the best performing hybrids and allowed visualization of crossover GEI through the polygon view.

Comparing statistical software for outputs on analysis of variance in agricultural experiments

Osman M. Siraj, S. Ali Eltahir
Agricultural Research Corporation (ARTC),
P.O.Pox 566, Wad Medani, Sudan
& M. Singh
International Centre for Agricultural Research in Dry Area (ICARDA),
Aleppo, Syria
Email: sirajstat@yahoo.com

Statistical software are essential tools for analysis of data for agricultural research. Analysis of variance (ANOVA) is a key statistical procedure to assess the effects and interactions of controlled factors. Different software facilitates ANOVA information with varying presentation structure in their default outputs. An organization with limited resources is often concerned with reducing the number of software if the gain is marginal. In this view, the present study has sought feedback from researchers in Sudan Agricultural Research System on the software usage, the most commonly used software in agricultural research in general and for ANOVA application in particular. This paper attempts to identify the most useful software keeping in view the experimental design consideration (such as replication, randomization and local control) and reporting of results from designed experiments. We also keep in view the convenience of inputting the experimental design structure, particularly for multi-factors in nested and crossed structures, e.g. the split-plot and strip-plot, and the outputs in terms of ANOVA, partitioning into single degree of freedom contrasts, means, estimated standard errors, experimental error variability at different levels of experimental units, and multiple comparisons.

Based on these considerations, GenStat has a distinct comparative advantage over the other popular software at least for the designed experiments. Our experience can be repeated by analyzing the data from standard textbooks (e.g. Cochran, W.G. and G.M. Cox (1964), *Experimental Designs*, John Wiley & Sons) to obtain the ANOVA and standard errors from split plot or strip-plots experiments. With the recent addition of multiple comparisons facilitates the drawing simultaneous inferences. Its high quality graphics can be used to demonstrate indication of departure from or support for the assumptions of ANOVA.

REML past, present and future

Robin Thompson
Biomathematics and Bioinformatics,
Rothamsted Research,
Harpenden,
Herts AL5 2JQ,
UK.
Email: robin.thompson@bbsrc.ac.uk

In this talk I intend to review some of the more significant developments of the REML algorithm available in GenStat, ASReml and ASReml-R. This will include the construction of a quadratically convergent scheme, the use of sparcity in the calculations and developing a flexible model specification. If time permits I will discuss some applications of the use of sampling to deal with large problems.

What cross-spectral analysis can do for you

Granville Tunnicliffe Wilson
Department of Mathematics and Statistics,
Lancaster University,
Lancaster LA1 4YF,
UK.
Email: g.tunnicliffe-wilson@lancaster.ac.uk

The talk will present applications in the fields of ecology, environment, medicine and marketing, of a new GenStat procedure for cross spectral analysis. Specifically, the applications reveal the effect of rainfall and solar radiation on moth trappings, the dependence between atmospheric CO2 and global temperature anomalies, the effect of pre-term infant respiration on blood oxygen levels and the dependence of product sales on price and promotions. The aim is to show how these methods, quickly and simply, provide valuable information about the strength and nature of lagged dependence between time series. The talk will include a brief introduction to the underlying ideas, designed to dispel some of the mystique of spectral analysis and overcome the long-standing reluctance of many scientists and statisticians (and time series analysts!) to use this valuable and long-established methodology.

Three-point appraisal of genetic linkage maps

Sue Welham, Jun Wang, Graham King & Wally Gilks
Rothamsted Research,
Harpenden,
Herts AL5 5RJ,
UK.
Email: sue.welham@bbsrc.ac.uk

Because of genotyping errors and computational difficulties, genetic linkage maps often have to compromise local accuracy to achieve an acceptable overall ordering of markers. These problems are compounded when combining data sets containing relatively few common markers in order to form an integrated map.

In the context of bi-parental doubled-haploid lines of *Brassica napus* (oilseed rape or canola), we have used the approximate posterior probabilities of map order for triples of markers as a simple diagnostic to assess the accuracy of map order. The method allows for the presence of missing marker scores and for genotyping error, where the error rate may either be fixed or estimated using a Beta prior distribution. Successful estimation of the genotyping error rate depends on the configuration of recombination events between markers, and can be diagnosed using the condition number of the information matrix.

Because of its simplicity, the method is computationally fast, and can be easily implemented. It can be used to evaluate the accuracy of high density maps at different scales, or to examine in detail local map accuracy. It can also be used to evaluate different positions when adding new markers onto existing maps, as in map integration.

Phospholipid fatty acids and analysis of distance for compositional data

Rodger White & Sarah Kemmitt
Rothamsted Research,
Harpenden,
Herts AL5 5RJ,
UK.
Email: rodger.white@bbsrc.ac.uk

Phospholipid fatty acids can be used as indicators of soil microbial diversity. Phospholipid fatty acid compositions were determined in triplicate from a grassland soil, unfumigated and fumigated, and incubated for 0, 6, 10, 30 and 62 days. Analysis of distance on the similarity matrix between the samples was used to apportion the variability between the fumigation and incubation time treatments and their interaction. Partitioning of the squared distance for incubation time into single degree of freedom polynomial contrasts is presented. Initial analysis was done on log-transformed data because of the wide range of values. Ordination of the between-groups distance matrix indicated some distortion between the groups. By treating the data as compositional and using log-ratios, the ordination of the between-groups distance matrix showed less distortion. Comparisons are made with similar graphical results from canonical variates analysis. The main benefit from this analysis is improved interpretation of the data.