

**Simon Diffey,**

National Institute for Applied Statistics Research Australia (NIASRA)

University of Wollongong

### **A new REML (PX)EM algorithm for linear mixed models and factor analytic mixed models**

Linear mixed models and factor analytic mixed models are routinely applied to biological data arising from designed experiments. The preferred method for estimating the parameters associated with these models is residual maximum likelihood (REML). Most statistical software packages available for the REML estimation of parameters associated with linear mixed models and factor analytic mixed models implement a Newton-Raphson type algorithm such as the expected information algorithm or the average information algorithm. There are two problems with these algorithms. Firstly, successive iterations of these algorithms are not guaranteed to increase the residual log-likelihood function. Secondly, parameter updates may not remain in their parameter space. Either may result in the algorithm failing to converge to a solution.

The REML expectation maximisation (REML EM) algorithm and the parameter expanded version of this algorithm (REML PX EM) are alternatives to Newton-Raphson type algorithms. Features of these two algorithms are that the residual log-likelihood may not decrease with successive iterations and parameter updates remain in their parameter space. Before the REML EM or REML PX EM algorithm can be considered practical alternatives to Newton-Raphson type algorithms two issues need to be addressed. Firstly, they can be notoriously slow to converge, particularly the REML EM algorithm. Secondly, compared to the average information algorithm, current implementations of these two algorithms are computationally more expensive at each iterate. This increased computational expense relates to calculating the trace of a matrix of the same order as the length of the observed data vector.

This presentation is divided into two parts. Firstly, a new incomplete data specification for a REML (PX)EM applied to a linear mixed model is presented which is computationally more efficient than alternative specifications. An example is considered where the new incomplete data specification results in the algorithm converging in 30% fewer iterations than the alternative specification. Secondly, a new incomplete data specification and a new parameter expansion for a REML PX EM algorithm applied to a factor analytic mixed model is presented. In this setting the efficacy of using a less stringent stopping rule is also considered. In an example plant breeding data set these innovations are shown to reduce the number of iterations to convergence by more than 50%.