Simon **Diffey**

DIRECTOR

Apex Biometry, PO Box 690, South Fremantle WA 6162

+61 481 816 872 | simon@apexbiometry.com | apexbiometry.com | SimonDiffey |
www.linkedin.com/in/simondiffey

Professional objective

To work with organisations dedicated to agricultural research and who value the contribution a biometrician can make to agricultural research projects.

Education

Australian National University — *Canberra*
DOCTOR OF PHILOSOPHY — *2012*

Macquarie University — *Sydney*
MASTER OF APPLIED STATISTICS — *2007*

University of New England — *Armidale*
GRADUATE DIPLOMA IN ECONOMETRICS — *2001*

University of Sydney — *Sydney*
BACHELOR OF ECONOMICS — *1993*

PhD Thesis

TITLE

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

DESCRIPTION

This thesis presents two major new findings. The first is a new specification of a REML (PX)EM algorithm for linear mixed models that is faster than previously published specifications. The second major finding is a new formulation of a REML (PX)EM algorithm for factor analytic linear mixed models where the parameter expansion is associated with the specific variances rather than the factor loadings. An outcome of this work is to broaden and improve the types of algorithms which can be practically applied to the analysis of plant improvement data.

SUPERVISORS

Professor Alan Welsh and Professor Brian Cullis.

Masters Thesis

TITLE

An alternative outlier mixed model for finding quantitative trait loci.

DESCRIPTION

This thesis describes the whole of genome average interval mapping (WGAIM) approach to QTL identification, provides R functions for implementing this method, and applies these functions to wheat quality data arising from a doubled haploid population. The R functions developed in this thesis formed the basis of the R package `wgaim`.

SUPERVISORS

Professor Brian Cullis.

Relevant Professional Experience

Apex Biometry   *South Fremantle WA*
Director   *July 2016 to current*

Curtin University   *Bentley WA*
Senior Research Fellow   *March 2015 to June 2016*

University of Wollongong   *Wollongong NSW*
Research Fellow   *March 2012 to March 2015*

NSW Department of Primary Industries   *Wagga Wagga NSW*
Consulting Biometrician   *March 2002 to March 2012*

Career synopsis

I was fortunate in my early career to be part of a large biometrics group within the NSW Department of Primary Industries (NSW DPI) led by Prof. Brian Cullis. Members of the group were recognised for their work in statistical methodology relevant to agricultural research and the development of software which implemented recent advances in statistical methodology. A feature of my early career was the diversity of the biometric consulting work I was required to perform. I have extensive experience working on Australia's three major crop types; namely wheat, barley and canola. I have worked collaboratively with a wide range of agricultural scientists as evidenced by scientific publications with research agronomists, soil scientists, molecular biologists, entomologists, cereal chemists, plant breeders, pathologists, physiologists, geneticists and animal scientists.

I have a particular interest in plant breeding, cereal chemistry, molecular biology and animal nutrition. During my employment with NSW DPI I acted as a consulting biometrician on projects associated with the GrainGene alliance and the NSW Agricultural Genomics Centre. This collaboration resulted in a number of peer reviewed journal articles.

I completed a Master of Applied Statistics in 2006 (conferred in 2007). In partial fulfillment of that program I submitted a thesis based on the whole of genome average interval mapping (WGAIM) method for QTL detection. This method was shown to be at least as good, and often better, than composite interval mapping in detecting QTL. A significant outcome of this thesis was the development of functions in the statistical software package R that implemented this method. This contribution was acknowledged in the paper by Verbyla et. al. 2007. These functions formed the basis of the freely available R package `wgaim` which was developed in 2008.

In 2012 I completed a Doctor of Philosophy at the Australian National University. This program was funded by a GRDC scholarship and the NSW DPI. This research resulted in a new REML (PX)EM for linear mixed models and factor analytic mixed models. The latter are routinely used in the analysis of multi-environment trials for the purpose of evaluating variety performance. The result of this research was to broaden and improve the types of algorithms which can be applied to plant improvement data. This work was done in close collaboration with Prof. Robin Thompson, Rothamsted Research UK.

I was employed as a Research Fellow within the Centre for Bioinformatics and Biometrics (CBB), University of Wollongong in the period March 2012 to March 2015 and as a Senior Research Fellow within the Centre for Crop and Disease Management (CCDM), Curtin University in the period March 2015 to June 2016. These positions were fully funded by industry; namely, the Grains Research Development Corporation (GRDC) and the CRC for High Integrity Australian Pork. I was the lead statistician on the GRDC funded National Brassica Germplasm Improvement Program (Phase 2) and co-lead on the GRDC National Frost Initiative (NFI) project, leader of the CRC for High Integrity Australian Pork project titled "Statistics for the Australian Pork Industry (SAPI)", and project leader of the Meat and Livestock Australia (MLA) project "Statistical analysis for the FIP Pasture Variety Trial Network (PVTN)". More details are provided further on.

In July 2016 I commenced working as a private consultant to industry as Director of Apex Biometry Pty Ltd. In this role I work collaboratively with plant and animal scientists in companies, state based agricultural departments and universities. A significant amount of my time is spent working with cereal, oilseed and pasture seed companies on incorporating the latest statistical design and analysis techniques into their breeding programs. Making results

accessible is a particular focus and typically involves a mix of traditional reporting such as written reports and Excel files but increasingly includes more modern media such as interactive web tools.

Statistical skills

Design

I routinely apply the latest statistical design technology in plant improvement, agronomic and animal science trials. Much of this recent technology seeks to find a balance between resource allocation (cost) and the requirement of an efficient design (and hence analysis). Examples of these types of designs include multi-phase designs, partially replicated designs, embedded partially replicate designs, multi-environment trial (MET) designs and designs incorporating pedigree information.

Analysis - phenotypic data

I'm highly familiar with the latest statistical technology for the analysis (including spatial analysis) of field, rain out shelter, glasshouse, controlled environment room and multi-environment trials. For the prediction of genetic potential from these type of trials I routinely use linear mixed models, generalised linear mixed models, and factor analytic mixed models. The latter typically being used for the modelling of genotype by environment interaction and where appropriate incorporating pedigree information.

Analysis - genotypic data

In the context of molecular marker data I'm highly knowledgeable with Quantitative Trait Loci (QTL) identification, Genome Wide Association Studies (GWAS) and Genomic Selection (GS). I'm highly knowledgeable on the statistical aspects of bi-parental genetic map construction for the purposes of marker trait association or interval mapping. Some of these aspects include missing molecular marker data, segregation distortion, the efficient formation of linkage groups and marker order within linkage groups, and map diagnostics.

Other

At various stages throughout my career I have also used finite mixture models, simulation, Gibbs sampling, and Bayesian methods.

Computer skills

R, Rstudio, RMarkdown

Highly proficient in writing customized R functions and using a number of R packages; including `asreml`, `wgaim`, `qtl`, `od`, `ggplot2` and `shiny`.

Latex, Beamer

Highly proficient in the use of Latex and Beamer for high-quality typesetting of scientific documents and slide presentations.

Fortran, C++, Python

Working knowledge of the programming languages Fortran, C++, and Python.

Project management skills

The following is a list of statistics projects worth more than $100,000 where I was either the project leader or lead statistician while employed at either Curtin University or University of Wollongong. An example of applying or developing statistical technology is provided for each.

NBGIP-II

Grains Research and Development Corporation (GRDC) appointed statistician to the multi-million dollar "National Brassica Germplasm Improvement Program - Phase 2 (NBGIP-II)" project. This project aimed to deliver elite germplasm and molecular markers to Australian canola breeding companies for use in breeding drought and heat tolerant varieties and blackleg resistant varieties. This project spanned three Australian states and involved a number of state based government agricultural departments and a number of universities. This project involved collaborating with plant breeders, molecular biologists, and plant pathologists and physiologists.

Multi-phase (field phase and laboratory phase) designs and embedded partially replicated designs (field phase) have been developed for cases where the traits being measured are expensive. Examples of two drought tolerance traits where this was the case are Carbon Isotope Discrimination (CID) and Water Soluble Carbohydrates (WSC). Using these types of designs allowed both genetic and non-genetic sources of variation to be efficiently estimated. This was important for two reasons. Firstly, identifying drought tolerant lines; and secondly, correctly identifying molecular markers associated with CID and WSC.

NFI

Grains Research and Development Corporation (GRDC) appointed statistician to the multi-million dollar "National Frost Initiative". This initiative comprised a number of projects aimed at delivering genetic and management information to grain growers so they could better manage their susceptibility to frost. The analysis of so-called frost expression experiments is statistically challenging; it involves the analysis of highly unbalanced, longitudinal and multi-environment data.

This project involved collaborating with plant physiologists, plant breeders, molecular biologists and geneticists in a number of different universities and state based agricultural departments. As part of this project I was responsible for the development and deployment of the frost web tool found at www.nvtonline.com.au/frost.

PVTN

Project leader for the Meat and Livestock Australia (MLA) project "Statistical analysis for the FIP Pasture Variety Trial Network (PVTN)". The deliverables on this project were the development of a database for storage, import and management of PVTN data, provision of a trial design service to trial operators, statistical analysis of PVTN data, and reporting the merits of pasture species and varieties that are under testing to industry. This project involved collaborating with plant breeders in private breeding companies and pasture ecologists and technical specialists in government agricultural departments and universities. Another deliverable on this project was the provision of statistical workshops on the analysis of pasture field trials to researchers in private breeding companies.

The PVTN produced multi-harvest, multi-site data. A new methodology based on factor analytic mixed models was developed for the prediction of seasonal pasture production. This is based on the need to be able to group multiple harvests into sensible seasonal categories. This was the first time that this technology had been used for this purpose in a commercial plant evaluation context.

SAPI

Project leader of the CRC for High Integrity Australian Pork (HIAP) project "Statistics for the Australian Pork Industry (SAPI)". This project delivered a statistical design and analysis service to research programs of the CRC for HIAP. This involved collaborating with a range of researchers in animal science, particularly in the areas of feed processing and animal nutrition.

This project involved both feed intake and feed digestibility animal trials. An important aspect of these trials is that the grain based pelleted diets that form the basis of these trials are manufactured in a mill over a number of days. Through the use of multi-phase partially replicated designs we were able to determine that the manufacturing process was an important source of non-treatment variation. We were able to show that by accounting for the variation within and between manufacturing days the number of animals required to determine a statistical difference between diets can in some instances could be halved. This had important animal ethics implications.

Research interests

- Iterative schemes, particularly REML (PX)EM algorithms, for linear mixed models and factor analytic mixed models. Both are routinely applied to plant improvement data.
- Prediction and classification of genetic performance for multi-harvest, multi-environment (pasture) trials.
- Linear mixed models applied to the problems of quantitative trait loci (QTL) identification, genome wide association studies (GWAS) and genomic selection (GS).
- Statistical communication using non-traditional media such as interactive web tools and video.
- Approaches and tools for managing statistical methodology changes (in a plant breeding program for example) within organisations.

Professional referees

Professor Brian Cullis
PROFESSOR OF BIOMETRY

University of Wollongong
bcullis@uow.edu.au

Dr John Black (AM FTSE FAIAST)
RESEARCH MANAGEMENT CONSULTANT

John L Black Consulting
jblack@pnc.com.au

Dr Harsh Raman
SENIOR PRINCIPAL RESEARCH SCIENTIST

NSW DPI
harsh.raman@dpi.nsw.gov.au

Journal articles

1. Ratanpaul, V, D Zhang, B Williams, **S Diffey**, J Black, and M Gidley (2019). Wheat bran and oat hulls have dose-dependent effects on ad-libitum feed intake in pigs related to digesta hydration and colonic fermentation. *Food and Function*.

2. Sandral, G, A Price, S Hildebrand, C Fuller, R Haling, A Stefanski, Z Yang, R Culvenor, M Ryan, D Kidd, **S Diffey**, H Lambers, and R Simpson (2019). Field benchmarking of the critical external phosphorus requirements of pasture legumes for southern Australia. *Crop and Pasture Science*.

3. Raman, H, R Raman, **S Diffey**, Y Qiu, B McVittie, M Barbulescu, D Maria, and ... (2018). Stable quantitative resistance loci to blackleg disease in canola (Brassica napus L.) over continents. *Frontiers in plant science*.

4. Trethowan, R, R Chatrath, R Tiwari, S Kumar, M Saharan, N Bains, V Sohu, P Srivastava, A Sharma, N De, S Prakash, G Singh, I Sharma, H Eagles, **S Diffey**, U Bansal, and H Bariana (2018). An analysis of wheat yield and adaptation in India. *Field crops research*.

5. van Barneveld, R, H Graham, and **S Diffey** (2018). Predicting the nutritional quality of feed ingredients for pigs using near-infrared spectroscopy (NIRS) and chemical analysis. *Animal Production Science*.

6. Black, J, I Box, and **S Diffey** (2017). Inadequacies of research used to monitor change to rock art and regulate industry on murujuga ('Burrup Peninsula'), Australia. *Rock Art Research: The Journal of the Australian Rock Art Research ....*

7. **SM Diffey**, A Smith, A Welsh, and B Cullis (2017). A new REML (parameter expanded) EM algorithm for linear mixed models. *Australian & New Zealand Journal of Statistics*.

8. Raman, H, R Raman, N Coombes, J Song, **S Diffey**, A Kilian, K Lindbeck, and ... (2016). Genome-wide association study identifies new loci for resistance to Leptosphaeria maculans in canola. *Frontiers in plant science*.

9. Raman, R, **S Diffey**, J Carling, R Cowley, A Kilian, D Luckett, and H Raman (2016). Quantitative genetic analysis of grain yield in an Australian Brassica napus doubled-haploid population. *Crop and Pasture Science*.

10. Cowley, R, D Luckett, J Moroni, and **S Diffey** (2014). Use of remote sensing to determine the relationship of early vigour to grain yield in canola (Brassica napus L.) germplasm. *Crop and Pasture Science*.

11. Raman, H, J Dalton-Morgan, **S Diffey**, R Raman, S Alamery, D Edwards, and ... (2014). SNP markers-based map construction and genome-wide linkage analysis in Brassica napus. *Plant biotechnology journal*.

12. Raman, H, R Raman, A Kilian, F Detering, J Carling, N Coombes, **S Diffey**, and ... (2014). Genome-wide delineation of natural variation for pod shatter resistance in Brassica napus. *PLoS One*.

13. Lemerle, D, P Lockley, E Koetz, and **S Diffey** (2013a). Herbicide efficacy for control of annual ryegrass (Lolium rigidum Gaud.) is influenced more by wheat seeding rate than row spacing. *Crop and Pasture Science*.

14. Mann, G, **S Diffey**, B Cullis, F Azanza, D Martin, A Kelly, L McIntyre, and ... (2009). Genetic control of wheat quality: interactions between chromosomal regions determining protein content and composition, dough rheology, and sponge and dough baking properties. *Theoretical and Applied Genetics*.

15. Raman, R, H Allen, **S Diffey**, H Raman, P Martin, and K McKelvie (2009). Localisation of quantitative trait loci for quality attributes in a doubled haploid population of wheat (Triticum aestivum L.) *Genome*.

16. Mann, G, **S Diffey**, H Allen, J Pumpa, Z Nath, M Morell, B Cullis, and A Smith (2008). Comparison of small-scale and large-scale mixing characteristics: Correlations between small-scale and large-scale mixing and extensional characteristics of wheat flour dough. *Journal of cereal science*.

17. Stevens, M, D Madge, D James, **S Diffey**, and L Schiller (2007). Ground cover management does not influence densities of key Iridomyrmex species (Hym., Formicidae) in Australian citrus groves. *Journal of applied entomology*.

18. Lemerle, D, B Verbeek, and **S Diffey** (2006). Influences of field pea (Pisum sativum) density on grain yield and competitiveness with annual ryegrass (Lolium rigidum) in south-eastern Australia. *Australian Journal of Experimental Agriculture*.

19. Hocking, P, J Mead, A Good, and **SM Diffey** (2003). The response of canola (Brassica napus L.) to tillage and fertiliser placement in contrasting environments in southern NSW. *Australian Journal of Experimental Agriculture*.

Refereed conference proceedings / abstracts

1. Black, J and **S Diffey** (2017). Importance of connectivity grains for AusScan NIR prediction accuracy. *Animal Production Science*.
2. Muller, M, **S Diffey**, and E Roura (2017). Preference thresholds for four limiting essential amino acids in piglets. *Animal Production Science*.
3. Ratanpaul, V, D Zhang, **S Diffey**, J Black, M Gidley, and B Williams (2017). Effects of different amounts of wheat bran and oat hulls on production of short chain fatty acids in the hindgut of pigs. *Animal Production Science*.
4. Ratanpaul, V, D Zhang, B Williams, **S Diffey**, J Black, and M Gidley (2017). Effects of different amounts of wheat bran and oat hulls in a starch-based diet on voluntary feed intake in grower pigs. *Animal Production Science*.
5. Black, J, C Collins, D Henman, and **S Diffey** (2015). Multiple treatments targeting the immune system of commercially-reared weanling pigs. *Animal Production Science*.
6. Nguyen, G, C Collins, D Henman, **S Diffey**, A Tredrea, J Black, and ... (2015a). Growth performance of weaner pigs fed diets containing grains milled to different particle sizes. II. Field pea. *Animal Production Science*.
7. Nguyen, G, C Collins, D Henman, **S Diffey**, A Tredrea, J Black, and ... (2015b). Growth performance of weaner pigs fed diets containing grains milled to different particle sizes. I. Sorghum. *Animal Production Science*.
8. Raman, H, R Raman, D Luckett, R Cowley, **S Diffey**, D Leah, R Meyer, N Coombes, A Price, R Wilson, D Roberts, and R Pragnell (2014). Understanding the genetic bases of phenotypic variation in drought tolerance related traits in canola (Brassica napus L.) *In Proceedings of the 18th Australian Research Assembly on Brassicas*.
9. Raman, H, R Raman, **S Diffey**, R Cowley, D Luckett, and D Roberts (2014). Genome-wide association for early plant vigour in canola. *In Proceedings of the 18th Australian Research Assembly on Brassicas*.
10. Black, J, R Hughes, **S Diffey**, A Tredrea, P Flinn, J Spragg, and J Kim (2014). Rapid assessment of feed ingredient quality. *In Proceedings of the 25th Annual Australian poultry science symposium*.
11. Hughes, R, J Black, P Flinn, A Tredrea, and **S Diffey** (2014). Separate cereal grain calibrations for prediction of apparent metabolisable energy by near infrared reflectance analysis. *In Proceedings of the 25th Annual Australian poultry science symposium*.
12. Roura, E, B Shrestha, and **S Diffey** (2013a). Pigs show high sensory-motivated intake for low levels of dextrose but not for low levels of maltodextrin. *Manipulating Pig Production XIV: Proceedings of the Fourteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.
13. Roura, E, B Shrestha, and **S Diffey** (2013b). Pigs show no sensory-motivated intake for several cereal and tuber starches except hydrolysed corn starch. *Manipulating Pig Production XIV: Proceedings of the Fourteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.
14. Roura, E, B Shrestha, and **S Diffey** (2013c). Pigs show very high preference and sensory motivated intake for low levels of tartaric and phosphoric acids. *Manipulating Pig Production XIV: Proceedings of the Fourteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.
15. Roura, E, B Shrestha, and **S Diffey** (2013d). Preference and sensory-motivated intake for 4 high intensity sweeteners in piglets. *Manipulating Pig Production XIV: Proceedings of the Fourteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.
16. Sopade, P, B Hosking, C Collins, **S Diffey**, G Fox, D Henman, C Brewster, and ... (2013). Improving efficiency of pig feed manufacturing and application of additives. *Manipulating Pig Production XIV: Proceedings of the Fourteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.
17. Lemerle, D, P Lockley, E Koetz, and **S Diffey** (2013b). Impacts of wheat row spacing, seeding rate and herbicide performance on grain yield and weed suppression. *In 16th European Weed Research Society Symposium*.
18. Tredrea, A, R Hughes, **S Diffey**, M Geier, J Black, J Roake, and R Trethowan (2012). Evaluation of new high yielding triticale cultivars for increased broiler production. *Proceedings of the 23rd Annual Australian Poultry Science Symposium*.
19. Hughes, R, M Geier, J Black, A Tredrea, and **S Diffey** (2012). Responsiveness of sorghum and triticale grains to separate blends of xylanase and phytase enzyme products. *Proceedings of the 23rd Annual Australian Poultry Science Symposium*.
20. Tredrea, A, R Hughes, **S Diffey**, M Geier, J Black, J Roake, and ... (2012). Evaluation of new high yielding triticale cultivars for increased broiler production. *Proceedings of the 23rd Annual Australian Poultry Science Symposium*.
21. Roura, E, B Shrestha, Y Zeng, D Zhang, and **S Diffey** (2011). A method to evaluate piglet preferences specific to taste. *Manipulating Pig Production XIII: Proceedings of the Thirteenth Biennial Conference of the Australasian Pig Science Association (APSA)*.

Theses

1. **Diffey, S** (2012). "A new specification of a REML (PX)EM algorithm for linear mixed models". PhD thesis. Australian National University, Mathematical Sciences Institute.
2. **Diffey, S** (2007). "An alternate outlier mixed model for finding quantitative trait loci". MA thesis. Macquarie Univeristy, Department of Statistics.

Book chapters

1. Black, J, V Ratanpaul, B Williams, **S Diffey**, and M Gidley (2019). "Variability in cereal grain composition and nutritional value: the importance of fibre". In: *The value of fibre: Engaging the second brain for animal nutrition*. Ed. by G Gonzalez-Ortiz, M Bedford, KB Knudsen, C Curtain, and H Classen. St. Paul, USA: Wageningen Academic Publishers.

2. Mann, G, **S Diffey**, L Rampling, Z Nath, I Kutty, P Leyne, F Azanza, K Quail, B Cullis, A Smith, and M Morell (2008). "Bubbles in bread: Is the answer in the genes?" In: *Bubbles in Food 2: Novelty, Health and Luxury*. Ed. by G Campbell, M Scanlon, and L Pyle. St. Paul, USA: AACC International.

3. Mann, G, **S Diffey**, L Rampling, Z Nath, I Kutty, P Leyne, F Azanza, K Quail, B Cullis, and A Smith (2007). "Mapping Approach Relating Quantitative Trait Loci to Dough Rheology in an Australian Doubled Haploid Population". In: *Gluten Proteins*. Ed. by G Lookhart and P Ng. St. Paul, USA: AACC International.

Computer packages

1. Taylor, J, **S Diffey**, A Verbyla, and B Cullis (2009). wgaim: Whole Genome Average Interval Mapping for QTL detection using mixed models. *R package version 1.4-0*.