Peer-reviewed paper

Benefits of mixed models and its BLUP methodology in sugarcane breeding

Jean-Yves Hoarau [1,2,3] and Thomas Dumont[1]

[1]*eRcane, 29 rue d'Emmerez de Charmoy, 97 490 Sainte-Clotilde, La Réunion, France*; jean-yves.hoarau@ercane.re

[2]*Cirad, UMR AGAP Institut, 97490 Saint-Clotilde, La Réunion, France*

[3]*AGAP, Univ. Montpellier, CIRAD, INRAE, L'Institut Agro, 34000 Montpellier, France*

Abstract   The use of linear mixed models (LMM) and its Best Linear Unbiased Prediction (BLUP) methodology is becoming increasingly popular amongst research scientists dealing with data modelling. Development of computer power and user-friendly statistical software facilitate their rapid implementation. This paper gives an overview of the benefits of the flexible LMM/BLUP framework to analyse experimental data collected in sugarcane-breeding programs to make efficient breeding and selection decisions, or to study genetic properties of traits of agronomic interest. Several applications of mixed models are presented. Their common feature is to exploit co-variances existing between some experimental data. Depending on applications, the nature of co-variances differs and might be related either to spatial, temporal and/or genetic considerations, with possible nested effects. In all cases, the methodology has the advantage of providing unbiased statistical inferences and predictions supporting objective conclusions in scientific issues investigated.

Key words   Linear mixed models (LMM), Best Linear Unbiased Prediction (BLUP), sugarcane, breeding

INTRODUCTION

Mixed modelling data has become a major area of statistical research that is useful in a wide variety of scientific disciplines. Mixed models are appropriate to analyse experimental data containing correlated measurements within or between units of interest. The theory of mixed models and its current applications trace their roots in the 1970s from quantitative methods developed for the genetic evaluation of domestic livestock (Henderson 1984, 1986) and from the use of matrix algebra in statistics (Searle *et al*. 1992). Over the last two decades, many textbooks have been published on mixed models and their applications. Moreover, the development of computer power and user-friendly statistical software helped their rapid implementation. Mixed models have been widely used in animal breeding for estimating *breeding values* in animal genetic improvement programs worldwide. These modelling approaches are becoming popular in plant-breeding and variety-testing programs of clonally propagated species. This paper aims to give an overview of its benefits in sugarcane-breeding and genetic-improvement programs to make efficient breeding decisions and genomic analyses.

BASIC CONCEPTS OF MIXED MODELS

Advantages of mixed models

Linear mixed models (LMMs) aim to model a response variable with both *fixed-effect* and *random-effect* explanatory variables. A LMM is a generalisation of the standard linear models (regression analysis and ANOVA) in which some data are permitted to exhibit correlation and non-constant variability (Galwey 2014). LMM has the ability to analyse datasets that may exhibit particular covariance structures. Most sugarcane-breeding experiments include some covariance structures that need to be best handled for optimised statistical analyses and inferences. Covariance between measurements can arise from spatial correlations (between neighbouring plots in variety

trials), from time correlation (between error measurements within a plot acquired at different crop-cycles), or from genetic correlations (between genetic entities that are relatives). Another major advantage of LMM resides in its ability to handle data grouped into clusters with the use of effects nested to others that can allow additional source of variability to model data with improved fit. Finally, yet importantly, LMM has the major advantage of easily accommodating unbalanced experimental designs due to field constraints (loss of a variety plot, unequal numbers of replicates per variety) or because of an objective of a joint analysis of different experimental designs (handled by the use of an indicator variable). Overall, LMM represent a powerful and flexible general framework to analyse any type of experimental data implying more or less complex covariance structures between some data.

Mathematical formulation and statistical concepts

An algebraic expression of a LMM can be expressed as follows:

$$Y_i = \mu + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + z_{i1}u_1 + \cdots + z_{iq}u_q + e_i = \mu + \sum_{j=1}^{j=p} x_{ij}\beta_j + \sum_{k=1}^{k=q} z_{ik}u_k + e_i$$

where Y_i is a response observation of an individual *i*, $[\beta_1, \ldots, \beta_p]$ a vector of *p* explanatory variables of *fixed-effects*; $[u_1, \ldots, u_q]$ a vector of *q* explanatory variables of *random-effects* and e_i the error term, both being unobservable and unknown random values; and $[x_{i1}, \ldots, x_{ip}]$ and $[z_{i1}, \ldots, z_{ip}]$ two incidence vectors relating both sets of explanatory variables to the response observation. Considering *n* response observations in an experiment data set to be modelled, the following matrix algebra notation conventionally defines a LMM (Littell *et al.* 2006):

$$Y = X\beta + Zu + e \qquad\qquad \text{(Equation 1)}$$

where Y is the vector of observations of dimension $n * 1$, β the vector of fixed effects of dimension $p * 1$, X the incidence matrix of dimension $n * p$ associated with the fixed effects, u the vector of random effects of dimension $n * p$, Z the incidence matrix of dimension $n * q$ associated with the random effects and e the vector of residual errors of dimension $n * 1$. The random effects are assumed to be distributed as $u \sim MVN(0, G)$ and $e \sim MVN(0, R)$, where $MVN(\mu, W)$ denotes a multivariate normal distribution with mean vector μ and variance-covariance matrix W. The vectors of random effects u and e are independent and therefore their covariance are null.

The estimation of random and fixed effects are known as *best linear unbiased predictions* (BLUP) or *best linear unbiased estimations* (BLUE), respectively. The first two letters of the acronyms BLUE and BLUP stand for *Best*, meaning that these predictors have the lowest variance, and *Linear*, meaning that they are linear functions of the data. The third letter for *Unbiased* has a different meaning depending on the predictor. In case of BLUE, unbiased means the expected value of a mean estimate for an individual equals its true value (i.e. the arithmetic mean of all its observations). This is a conditional or subject-specific mean. By contrast, in case of BLUP unbiased only means that the expected mean over all individuals is equal to the expected mean over all true effects. This is a marginal mean.

The shrinkage property of the BLUP concept

BLUP values of the different levels of any random variable in a LMM are characterised by a property of *shrinkage*. The higher the error variance of a particular level, the greater its BLUP value is shrunk. Figure 1 illustrates this shrinkage property using an example of cane yield data from a randomised experiment of 16 sugarcane varieties that implies an unbalanced number of replicates per clone. This shrinkage effect is a fundamental property of the BLUP concept, which allows realistic and cautious inferences on the different levels of any random variable of a LMM.

The choice between random or fixed effects

In a LMM, it is up to the experimenter to define which explanatory variable(s) should be considered *fixed* and which be chosen as *random*. A key characteristic of random effects is that they can be regarded as a random sampling from a much bigger (even infinite) *source population* of effects (Robinson 1991). Therefore, BLUP predictor allows *broad inferences* on the space of the whole source population (Littell *et al.* 2006). By contrast, BLUE predictor provides *narrow* inferences, which are *subject-specific*. Inferences are first meaningful in the particular experimental context from which the analysed data originate. It is commonly recognised that in most breeding trials, genetic effects should be regarded as random effects to maximise selection gains from data available (Robinson

1991; Molenaar *et al*. 2018; Jackson *et al*. 2022; Hoarau *et al*. 2022). In addition, certain types of variables must be necessarily considered as random. This is the case for variables involved in nested effects (family effects hierarchised to particular parents), for latent variables not observable directly, but whose value is to be assessed (an additive genetic value within a clonal value, agronomic effects of markers in a genomic selection model). As a rule, variables estimated from unbalanced datasets or from data implying obvious or cryptic covariance structures, should necessarily be considered as random. By contrast, variables involving discrete and few balanced levels can be chosen as fixed effects in particular if looking for statistical comparison among treatment levels.
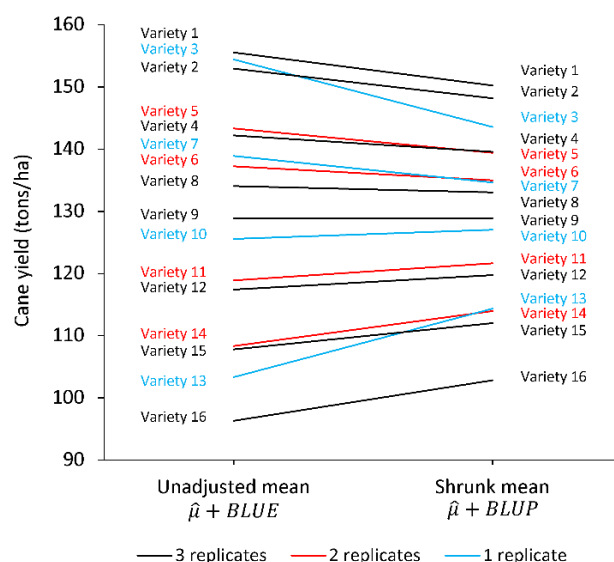

Figure 1. Illustration of the shrinkage property of the BLUP methodology for a variety effect alternatively considered as *fixed* or *random*: comparison between unadjusted means ($\hat{\mu} + BLUE$) and shrunk means ($\hat{\mu} + BLUP$) of 16 sugarcane varieties for cane yield (t/ha) experimented in a randomised trial implying an unbalanced number of replicates per variety. The above-average arithmetic means of the varieties ($\hat{\mu} + BLUE$) are shrunk downward toward the overall population mean when considering variety predictions ($\hat{\mu} + BLUP$), whereas below-average variety means are shrunk upward towards population mean (Hoarau *et al*. 2022).

Practical implementation of LMM

In practice, analytical resolution of a large system of linear equations modelling a response variable Y is not a feasible task. Estimation of BLUEs, BLUPs and *variance components* parameters of a LMM (values of its G and R matrices) are usually obtained with the *restricted maximum likelihood* (REML) algorithm (Patterson and Thompson, 1971). This algorithm searches best estimates of all parameters using a recursive numerical method until convergence is reached. Some statistical software, which fit LMMs via the REML method, include ASReml (Gilmour *et al*. 2009), SAS (Littell *et al*. 2006), R (R Development Core Team 2010) and GenStat (Payne *et al*. 2009). Because of this estimation procedure based on increasingly accurate iterative approximations, implementations of LMM can easily accommodate unbalanced or incomplete datasets.

BENEFITS OF MIXED MODELS IN BREEDING DECISIONS

Analyses of progeny trial data in sugarcane-breeding programs using mixed modelling approaches provide instructive information to guide breeding strategies relating to crossing schemes and choice of best parents.

Prediction of the potential of sugarcane families

Appraisal of the value of bi-parental families is the first step of efficient sugarcane breeding programmes (Stringer *et al*. 2011). This first selection-stage usually consists of progeny tests conducted in replicated randomized field trials aiming at comparing the merit of many crosses. A simple LMM including a family effect specified as a random

(with all appropriate trial design effects) provides family BLUP values (Pedrozo *et al.* 2011). BLUPs provide a much more objective ranking of the genetic merit of the families than family plot means (due to the shrinkage property of BLUP predictor that can adjust for unbalanced data across families or crop-cycles and variable data quality across trials).

Prediction of breeding values of parents

Prediction of *breeding values* (BV) of parents in sugarcane programs can also be obtained from data of replicated progeny trials. A first philosophy of analysis may consist in considering that all parents came from a reference population and are stochastically independent, ignoring any possible pedigree relationships. The modelling approach has the virtue of simplicity because it does not have recourse to the theory of quantitative genetics (Piepho *et al.* 2008). Data of individual progenies tested in replicated family trials can be simply analysed in a mixed model including random male and female parent effects, along with effects reflecting trial designs (Mbuma *et al.* 2020). BLUPs of parents provide straightforward estimates of their *breeding values* and are used to rank their merits.

An alternative philosophy of data analysis makes use of the 'animal mixed model' (Henderson 1984) that exploits pedigree information. An additive relationship matrix A between parents, is computed from their coefficients of co-ancestry (Falconer and Mackay 1996). Modelling of progeny trial data includes a random *breeding value* effect of the parents (along with some other appropriate random or fixed effects). The implantation of the LMM specify a variance-covariance matrix of these *breeding value* effects as being $A\sigma_A^2$, where σ_A^2 is the additive genetic variance of the yield trait studied (Barbosa *et al.* 2005, Neto *et al.* 2013, Resende and Barbosa, 2006). Incidentally, the estimation of this variance component makes it possible to estimate the narrow sense heritability of the trait. BLUP of the parents obtained allows identifying parents with the best BV. This model provides BV estimates increasingly accurate, as a larger number of progeny data from different related parents cumulated over years is included in the analysis (Atkin *et al.* 2009). Interestingly, this model can also be refined by adding non-additive effects jointly estimated with additive effects. Applied to the analysis of elite varieties tested in final-assessment trials of a breeding program, this more advanced model can allow estimating BV of potential parents amongst elite material of high *per se* value for most important traits (Oakey *et al.* 2007).

BENEFITS OF MIXED MODELS IN SELECTION DECISIONS

Numerous smart applications of LMMs can be advantageously developed to analyse data from selection programs depending on data available, objectives, scientific findings and conclusions pursued.

Accounting for environmental variations within trials

LMMs can be advantageously used to take into account natural variations within trials due to uncontrolled environmental factors (varying soil texture or fertility, drainage effect) in order to optimise precision in variety comparisons and avoid misleading findings. Several options of analysis can be designed depending on the suspected profiles of the field variations to be controlled (Gilmour *et al.* 1997). Local variations can be tested by considering anisotropic auto-regressive process of correlation between model residuals of neighbouring variety plots based on a distance function by considering a R matrix different from the identity one (Ostengo *et al.* 2015; Jackson *et al.* 2022). This strategy can be mixed with linear regression components across rows and/or columns in case of trends of fertility suspected in either of these two directions in the field (Stringer and Cullis 2002). In the case of complex and random variation profiles in breeding trials that may represent substantial background noise in varietal comparisons, adjustments using two-dimensional P-spline mixed models might be profitable (Velazco *et al.* 2017; Hoarau *et al.* 2019).

Joint analysis of trials or of traits

Trials sharing common genetic entities (varieties, families) may be analysed jointly for a single trait. It may be used to compare varieties over time (generations, years) and space (trials) and analyse long-term genetic progress (Dumont *et al.* 2019). In this situation it is common to consider as random effects at least the variety effect (due to unbalanced data) as well as all derived possible interaction effects (such as variety x crop-year). This choice provides BLUP values for varieties that are essential to obtain cautious and sound inferences. Analyses of multi-environment trial (MET) for single traits are also very common in sugarcane-breeding programs to assess variety

x environment interactions and obtain instructive information to optimise resource allocation (Ramburan *et al.* 2012; Guilly *et al.* 2017). Phenotypic variance is usually partitioned into genotype (G), locations (L), harvest (C) and interactions (GL, GC and GLC) effects. Analysis can be done with all these terms by considering a null covariance between random and residual effects (G and R matrices are identity ones).

Alternatively, the three interactions effects can be skipped and a variance-covariance matrix of genotype effects combining covariance structures between sites and harvests can be considered (Smith *et al.* 2007; Pastina *et al.* 2012; Balsalobre *et al.* 2016). This more elaborate model amounts to considering the existence of genetic correlations between the data of the different locations and harvests. By trying to take advantage of it, we seek to improve variety estimates for the trait under study.

Finally, one may also wish to jointly analyse two traits or more acquired in the same experiment, by considering parameters of genetic covariance between traits (Zhou and Kimbeng 2010). If these are found to be genetically correlated, multivariate analysis strategy may provide more accurate estimates for each trait than their univariate analysis alone.

USE OF MIXED MODELS IN GENOME-WIDE ASSOCIATION STUDY

Many genome-wide association studies (GWAS) have been undertaken in the last decade in sugarcane concerning resistance to diseases and traits related to cane yield and quality (Debibakas *et al.* 2014; Santos *et al.* 2015; Racedo *et al.* 2016; Singh *et al.* 2016; Yang *et al.* 2019; Fickett *et al.* 2019; Zan *et al.* 2020). GWAS consist in studying statistical associations between segregating markers scattered over the plant genome and variations of agronomic traits within an experimental population of varieties. Therefore, GWAS enable understanding the genetic architecture of agronomic traits and allow prospecting for alleles of interest. The most common GWAS approach (Yu *et al.* 2006) models variety BLUPs of an experimental population in a LMM that take into account (i) as fixed effects, the genetic structure of the population (Q) and the segregating markers, and (ii) a random polygene background effect of the individuals whose covariance depends on their kinship coefficients (K) inferred from markers. Providing a good control of spurious associations, Quantile-Quantile of the *P*-values of marker-trait association tests and Manhattan plots allow to identify chromosomal regions of agronomic interest. However, finding causal polymorphism in close linkage disequilibrium may be difficult due to the sugarcane's large genome. Emerging Next Generation Sequencing Technologies (NGS) combined with reference sequences of sugarcane genome should improve the scope of GWAS results, which should facilitate the development of efficient genomics breeding applications.

CONCLUSIONS

Linear mixed models represent a powerful and flexible general framework to analyse any type of experiment implying covariance data structures more or less complex (repeated measurements, different years and/or locations, genetic correlations between individuals or traits) and to jointly analyse different experimental designs. Its concept of BLUP is a key desirable predictor that adjusts for unbalanced data and avoid biased findings. Therefore, breeding and selection decisions based on BLUP allow sound and cautious decisions for efficient guidance of sugarcane-breeding programs in ensuring optimized genetic gains. Moreover, mixed models are the cornerstone of the analysis of marker data along with breeding data for genomic analysis of trait of interest such as genome-wide association studies. With massive marker data provided by Next-Generation Sequencing approaches, future progress in the use of mixed model to develop profitable genomics breeding applications will reside in closer scientific collaborations between breeders, molecular biologists, bio-statisticians and computer scientists.

REFERENCES

Atkin FC, Dieters MJ, Stringer JK. 2009. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics* 119: 555–565.

Balsalobre TW, Mancini MC, Pereira DGS, *et al.* 2016. Mixed modeling of yield components and brown rust resistance in sugarcane families. *Agronomy Journal* 108: 1824–1837.

Barbosa MHP, de Resende MDV, Bressiani JA, *et al.* 2005. Selection of sugarcane families and parents by Reml/Blup. *Crop Breeding and Applied Biotechnology* 5: 443–450.

Debibakas S, Rocher S, Garsmeur O, *et al.* 2014. Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide association. *Theoretical and Applied Genetics* 127: 1719–1732.

Dumont T, Thong-Chane A, Barau L, *et al.* 2019. Genetic variabilities and genetic gains for yield components in regional sugarcane breeding programmes on Reunion Island. *Sugar Tech* 21: 868–878.

Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*, 4th ed., Longman, Essex, UK.

Fickett N, Gutierrez A, Verma M, *et al.* 2019. Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111: 1794–1801.

Galwey NW 2014. *Introduction to mixed modelling. Beyond regression and analysis of variance.* Second edition. John Wiley & Sons, New York.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R. 2009. *ASReml user guide 3.0.* VSN Int., Hemel Hempstead, UK.

Guilly S, Dumont T, Thong-Chane A, *et al.* 2017. Analysis of multi-environment trials (MET) in the sugarcane breeding program of Reunion Island. *Euphytica* 213: 213.

Henderson CR. 1984. *Applications of linear models in animal breeding.* University of Guelph, Guelph, Canada.

Henderson CR. 1986. Statistical methods in animal improvement: Historical overview. In: Gianola D, Hammond K (eds) *Advances in statistical methods for genetic improvement of livestock.* Springer, Berlin, pp. 2–14.

Hoarau JY, Dumont T, Wei X, *et al.* 2022. Applications of quantitative genetics and statistics in sugarcane breeding. *Sugar Tech* 24: 320-340.

Hoarau JY, Thong-Chane A, Barau L, *et al.* 2019. Comparison between randomized complete-block and alpha designs in sugarcane variety trials in the light of a spatial modelling of fertility trends. *Proceedings of the International Society of Sugar Cane Technologists* 30: 121–127.

Jackson PA, Demano E, Wei X. 2022. An introduction to some concepts in statistical analysis and quantitative genetics for sugarcane-breeding programs. *Sugar Tech* 24: 298-319.

Littell RC, Milliken GA, Stroup WW, *et al.* 2006. *SAS for mixed models*, 2nd ed. Cary: SAS, Institute Inc, 814p.

Mbuma NW, Zhou MM, van der Merwe R. 2020. Estimating breeding values of genotypes for sugarcane yield using data from unselected progeny populations. *Euphytica* 216: 1-15.

Molenaar H, Boehm R, Piepho HP. 2018. Phenotypic selection in ornamental breeding: it's better to have the BLUPs than to have the BLUEs. *Frontiers in Plant Science* 9, 1511.

Neto HZ, Daros E, Bespalhok Filho JC, *et al.* 2013. Selection of families and parents of sugarcane (*Saccharum* spp.) through mixed models by joint analysis of two harvests. *Euphytica* 193: 391–408.

Oakey H, Verbyla AP, Cullis BR, *et al.* 2007. Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics* 114: 1319-1332.

Ostengo S, Cuenya MI, Balzarini M. 2015. Modeling spatial correlation structure in sugarcane (*Saccharum* spp.) multi environment trials. *Journal of Crop Improvement* 29: 53–64.

Pastina MM, Malosetti M, Gazaffi R, *et al.* 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theoretical and Applied Genetics* 124: 835–849.

Patterson HD, Thompson R. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545–554.

Payne RW, Murray DA, Harding SA, *et al.* 2009. *GenStat for Windows. 12th e*d. VSN Int., Hemel Hempstead, UK.

Pedrozo CA, Barbosa MHP, da Silva FL, *et al.* 2011. Repeatability of full-sib sugarcane families across harvests and the efficiency of early selection. *Euphytica* 182: 423–430.

Piepho HP, Möhring J, Melchinger AE, *et al.* 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209–228.

Racedo J, Gutierrez L, Perera MP, *et al.* 2016. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biology* 16: 142.

Ramburan S, Zhou M, Labuschagne M. 2012. Investigating test site similarity, trait relations and causes of genotype x environment interactions of sugarcane in the Midlands region of South Africa. *Field Crops Research* 129: 71–80.

R Development Core Team. 2010. *A language and environment for statistical computing.* R foundation for Statistical Computing, Vienna, Austria

Resende MDVD, Barbosa MHP. 2006. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. *Pesquisa Agropecuária Brasileira* 41: 421-429.

Robinson GK. 1991. That BLUP is a good thing—the estimation of random effects. *Statistical Science* 6: 15–32.

Santos FR, Pinto LR, Carlini-Garcia LA, *et al.* 2015. Marker-trait association and epistasis for brown rust resistance in sugarcane. *Euphytica* 203: 533–547.

Searle SR, Casella G, McCulloch CE. 1992. *Variance components.* Wiley, New York.

Singh RK, Banerjee N, Khan MS, *et al.* 2016. Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum* species hybrid) using LD-based association mapping. *Molecular Genetics and Genomics* 291: 1363–1377.

Smith AB, Stringer JK, Wei X, *et al.* 2007. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. *Euphytica* 157: 253–266.

Stringer JK, Cox MC, Atkin FC, *et al.* 2011. Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Sugar Tech* 13: 36–41.

Stringer JK, Cullis BR. 2002. Application of spatial analysis techniques to adjust for fertility trends and identify inter-plot competition in early stage sugarcane selection trials. *Australian Journal of Agricultural Research* 53: 911-918.

Velazco JG, Rodrıguez-Alvarez MX, Boer MP, *et al.* 2017. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theoretical and Applied Genetics* 130: 1375–1392.

Yang X, Todd J, Arundale R, *et al.* 2019. Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Industrial Crops and Products* 130: 598–605.

Zan F, Zhang Y, Wu Z, *et al.* 2020. Genetic analysis of agronomic traits in elite sugarcane (*Saccharum* spp.) germplasm. *PLoS ONE* 15: e0233752.

Zhou MM, Kimbeng CA. 2010. Multivariate repeated measures: a statistical approach for analysing data derived from sugarcane breeding variety trials. *Proceedings of the South African Sugar Technologists' Association* 83: 92-105.

French title : Avantages des modèles mixtes et du concept de BLUP dans le domaine de la sélection de la canne à sucre

Résumé. L'utilisation de modèles linéaires mixtes (LMM) et du concept statistique de meilleur prédicteur linéaire sans biais (ou de BLUP pour Best Linear Unbiased Predictor en anglais) deviennent de plus en plus populaires dans le domaine de la modélisation de données expérimentales. L'apparition d'ordinateurs puissants et de logiciels statistiques conviviaux facilitent l'implémentation rapide des modèles linéaires mixtes. Cet article donne un aperçu de la flexibilité et des avantages des modèles mixtes et des BLUP pour analyser des données expérimentales de programmes d'amélioration génétique dans le but de décisions efficaces de création variétale ou pour étudier les propriétés génétiques de caractères d'intérêt agronomique. Plusieurs applications de modèles mixtes sont présentées. Leur caractéristique commune est d'exploiter des co-variances existant entre certaines données expérimentales. Selon les applications, la nature de ces covariances diffère. Elle peut être liée à des relations spatiales, temporelles et/ou génétiques entre données, avec de possibles effets hiérarchisés. Dans tous les cas, la méthodologie des modèles mixtes a l'avantage de fournir des inférences et des prédictions statistiques sans biais permettant d'étayer des conclusions objectivées sur les questions scientifiques étudiées.

Mots-clés: Modèles linéaires mixtes (LMM), Meilleur Prédicteur Linéaire sans Bias (BLUP), canne à sucre, amélioration génétique

Beneficios de los modelos mixtos y su metodología BLUP en el mejoramiento genético de la caña de azúcar

Resumen. El uso de modelos mixtos lineales (LMM) y su metodología Best Linear Unbiased Prediction (BLUP) se está volviendo cada vez más popular entre los científicos investigadores que se ocupan del modelado de datos. El desarrollo del poder de la computadora y el uso amigable de los softwares estadísticos facilitan su rápida implementación. Este trabajo ofrece una descripción general de los beneficios del flexible marco LMM/BLUP para analizar datos experimentales recopilados en programas de mejoramiento de caña de azúcar para tomar decisiones eficientes de mejoramiento y selección, o para estudiar propiedades genéticas de rasgos de interés agronómico. Se presentan varias aplicaciones de modelos mixtos. Su característica común es explotar las covarianzas existentes entre algunos datos experimentales. Dependiendo de las aplicaciones, la naturaleza de las covarianzas difiere y puede estar relacionada con consideraciones espaciales, temporales y/o genéticas, con posibles efectos anidados. En todos los casos, la metodología tiene la ventaja de proporcionar inferencias y predicciones estadísticas insegadas que respaldan conclusiones objetivas en los temas científicos investigados.

Palabras clave: Modelos lineales mixtos (LMM), Best Linear Unbiased Prediction (BLUP), caña de azúcar, fitomejoramiento