A generalized statistical framework to assess mixing ability from incomplete mixing designs using binary or higher order variety mixtures and application to wheat

Emma Forst, Jerome Enjalbert, Vincent Allard, Christophe Ambroise, Inès Krissaane, Tristan Mary Huard, Stephane Robin, Isabelle Goldringer

To cite this version:

Emma Forst, Jerome Enjalbert, Vincent Allard, Christophe Ambroise, Inès Krissaane, et al.. A generalized statistical framework to assess mixing ability from incomplete mixing designs using binary or higher order variety mixtures and application to wheat. Field Crops Research, 2019, 242, pp.107571. ⟨10.1016/j.fcr.2019.107571⟩. ⟨hal-02618625⟩

HAL Id: hal-02618625
https://hal.inrae.fr/hal-02618625v1

Submitted on 20 Jul 2022

A generalized statistical framework to assess mixing ability from incomplete mixing designs using binary or higher order variety mixtures and application to wheat

Authors and affiliations:

Emma Forst[a,1], Jérôme Enjalbert[a], Vincent Allard[b], Christophe Ambroise[c,d], Inès Krissaane[d], Tristan Mary-Huard[a,d], Stéphane Robin[d], Isabelle Goldringer[a]

[a] GQE- Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190, Gif-sur-Yvette, France

[b] GDEC, UMR 1095, INRA, UCA, 63000 Clermont-Ferrand, France

[c] LaMME, UMR 8071, CNRS, Université d'Evry Val-d'Essonne, Université Paris-Saclay, 91037, Evry Cedex, France

[d] MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

Emails: emma.forst@inra.fr, jerome.enjalbert@inra.fr, vincent.allard@inra.fr, christophe.ambroise@genopole.cnrs.fr, krissaane.ines@gmail.com, tristan.mary-huard@agroparistech.fr, stephane.robin@agroparistech.fr, isabelle.goldringer@inra.fr

HIGHLIGHTS

- Statistical analyses for variety mixtures have made little progress in recent years
- Novel models are proposed to study mixing ability in incomplete designs
- The models account for inter and intra-genotypic interactions within mixtures
- The framework handles mixtures with any order and proportions of components
- This framework was shown to be relevant on wheat mixture trial analysis

---

[1] Corresponding author: Génétique Quantitative et Evolution - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Ferme du Moulon, 91190, Gif-sur-Yvette, France. E-mail: emma.forst@inra.fr

ABSTRACT

There has been recently a renewed interest for variety mixtures due to their potential capacity to stabilize production through buffering abiotic and biotic stresses. Part of this results from complementarity and/or compensation between varieties which can be assessed under mixed stands only. Mixing ability of varieties can be partitioned into *General* and *Specific Mixing Abilities* (GMA and SMA) that have been estimated so far through the evaluation of binary mixtures in complete diallel designs. However, the number of mixtures increases exponentially with the number of studied varieties, and the only feasible devices are incomplete designs. Despite the long history of statistical analysis of variety mixtures, such incomplete design analysis have rarely been addressed so far. To fill the gap, we proposed a generalized statistical framework to assess mixing abilities based on mixed models and BLUP method, with an original modeling of plant-plant interactions. The approach has been applied to a panel of 25 winter wheat genotypes observed in two contrasted experimental designs: (i) an incomplete diallel of 75 binary mixtures, and (ii) a trial including higher order mixtures (four and eight components). The use of mixing ability models improved prediction accuracy (of modeled values for observed traits) in comparison to predictions from the mean of the pure stand components, especially in the first experiment. Genetic variability was detected for the GMA of yield and its components, whereas variability for SMA was lower. GMA predictions based on the diallel trial were highly correlated with the GMA of the second trial providing accurate inter-trial predictions. A new model has been proposed to jointly account for inter and intra-genotypic interactions for specific mixing ability, thus contributing to a better understanding of mixture functioning. This framework constitutes a step forward to the screening for mixing ability, and could be further integrated into breeding programs for the development of intra- or inter-specific crop mixtures.

Key-words: intra-specific mixtures, plant-plant interactions, diallel, *Triticum aestivum*, BLUP

1. INTRODUCTION

Crop genetic diversity is a major lever that can be mobilized to face current challenges in agriculture such as the increasing climatic stochasticity and the necessity to reduce the use of synthetic inputs (Newton *et al.* 2009). Among diversifying strategies, cereal variety mixtures have attracted increasing interest in recent decades (Kiær *et al.* 2009, 2012, Borg *et al.* 2018, Reiss and Drinkwater 2018). In particular, mixtures of susceptible and resistant cereal varieties have been shown to provide a larger reduction in the development of foliar diseases than expected from the sum of their components when complementary resistances are combined (Wolfe 1985; Finckh and Mundt 1992; Zhu *et al.* 2000; de Vallavieille-Pope 2004). Using variety mixtures may also allow buffering abiotic and other biotic stresses, therefore leading to a stabilization of production (Østergård *et al.* 2005; Kaut *et al.* 2009; Creissen *et al.* 2016), through complementarity and compensation mechanisms that might occur between plants of different varieties, as shown in species mixtures (Tilman *et al.* 1997).

Despite these potential advantages of mixtures, very little literature provides practical guidelines for designing mixtures, and consequently mixtures used in agriculture are often composed of the varieties showing the best performances in pure stand (Borg *et al.* 2018). However, it has been shown that mixtures performances are not necessarily correlated with the means of their pure stand components: interaction between plants with different genotypes (Finckh and Mundt 1992) can make it difficult to predict the behavior of mixtures. Therefore, it is important to accurately estimate and predict the performances of varieties in mixtures. Screening a set of lines for their *mixing ability* nevertheless presents many methodological challenges, in particular in the case of a high number of genotypes and an exponentially higher number of binary or higher order mixtures, and calls for adapted statistical methods (Dawson and Goldringer 2012; Barot *et al.* 2017) that could be used to detect the best varieties for mixing conditions, and to develop specific breeding schemes for intra-specific mixture design.

Mixing ability relies on an analogy with the concept of *combining ability* developed for hybrid breeding (Sprague and Tatum 1942). By assessing a set of *p* genotypes in all possible hybrid combinations, the combining ability can be estimated and partitioned between the general combining ability of the genotypes and the specific ability of each particular combination (Griffing 1956). The analogy has first been proposed for the study of plant interactions between rows of genotypes (Jensen and Federer 1965) and has then been broadly applied to the study of binary mixtures for estimating mixing ability, both when the performance of each component is accessible in the mixture (McGilchrist 1965; Chalbi 1967; Gallais 1970; Federer 1979; Federer *et al.* 1982), or when only the mixture performance is surveyed (Federer 1979; Federer *et al.* 1982; Gizlice *et al.* 1989; Knott and Mundt 1990; Gallandt *et al.* 2001). The work developed hereafter deals with this second case, where only the global performance of each mixture is available. The ***general mixing ability (GMA)*** refers to the average performance of a variety in mixture (additive term) and the ***specific mixing ability (SMA)*** relates to the interaction between the two components of a particular combination of varieties. According to Griffing's decomposition of combining ability in hybrids (1956), the performance of binary mixtures can be modeled as:

$$Y_{ijbr} = \mu + a_b + \tfrac{1}{2}(GMA_i + GMA_j) + SMA_{ij} + e_{ijbr},$$

where Y_{ijbr} is the performance of the mixture of varieties i and j for block b (and possibly replicate r in block b), μ is the intercept, a_b is the effect of block b, GMA_i is the general mixing ability of variety i, SMA_{ij} is the specific mixing ability of varieties i and j grown together, and e_{ijbr} is the error term of the observation. The $\tfrac{1}{2}$ coefficient is applied on the GMA term since each variety accounts for only half of the plants grown in the plot (in case of equal proportions at sowing).

However, due to the need to grow all the possible binary mixtures (that is to use a complete design) among a set of varieties to estimate GMA and SMA, the field evaluation can be cumbersome regarding the cost and management of experimental trials. For this reason, authors have usually applied this method to a limited number of varieties (e.g. six or eight in Gallandt *et al.* 2001; Gizlice *et al.* 1989 respectively). A key constraint for experimental efforts, this curse of dimensionality has also been encountered when studying hybrids' combining ability, and solved by the use of incomplete designs (Jensen 1959). By reducing the number of mixtures to grow for each variety, such designs allow a wider range of varieties to be evaluated. Nonetheless, due to the reduced number of observed mixtures using the Griffing's model (1956), the use of incomplete design raises the question of the definition of GMA and SMA. If GMA and SMA are defined assuming a complete design setting, then some GMA and SMA cannot be estimated in practice if the actual design is incomplete. Otherwise, if GMA and SMA are defined conditionally to the set of observed mixtures only, then the definition of GMA and SMA is specific to this set of mixtures.

Alternatively, hybrid's geneticists suggested the use of mixed linear models for the analysis of crossing designs, in which the estimates of GCA and SCA are provided by Best Linear Unbiased Predictors (BLUP, Möhring, Melchinger and Piepho 2011). In this framework, both the parameters of the models (namely the intercept and the variance of the random effects) and the (random) effects to be predicted have the same definition whatever the design (see Appendix A for a theoretical comparison of the two approaches). Interestingly, despite the known similarity between binary mixtures and hybrids, the methods developed for hybrids' unbalanced designs have not been adapted so far to mixtures.

Moreover, there is also a need for suitable approaches to handle higher order mixtures (i.e. mixtures including more than two components) as they have been found useful in different studies (Lopez and Mundt 2000; Mille *et al.* 2006), and are closer to the three to five-way mixtures commonly grown by farmers. Indeed, higher order mixtures present strong agronomical interest, such as their ability to reduce disease development (Newton *et al.* 1997, Borg *et al.* 2018), and can provide multiple agroecosystem services (i.e. weed suppression, yield and grain quality when considered altogether, Lazzaro, Costanzo, and Bàrberi 2018). Here the mixtures reveal their fundamental divergence with hybrids: if hybrids are constrained to binary and balanced combinations (1:1 of the two parental genomes, except when considering polyploids), mixtures allow all degrees of freedom, considering any number of components, in any possible proportion.

The aim of this article is to provide a methodological framework for the analysis of mixing ability in incomplete designs for binary and higher order mixtures based on dedicated statistical methods using mixed models. For a given panel of genotypes, this allows (i) the assessment of the relative importance of the additive part of the mixing ability vs the part that results from specific interactions due to the combinations of genotypes (using two modeling of the inter and intra-genotypic interactions thus providing a better understanding of mixture functioning), (ii) the identification of the best performers in mixtures, and (iii) the prediction of mixture performances using mixing ability modeling. The approach was applied to two contrasted cases: (i) a trial of wheat binary mixtures and their pure stands, and (ii) a trial including higher order mixtures.

2. MATERIAL AND METHODS

2.1 Modeling of mixing ability:

2.1.1 Griffing's model with fixed GMA and SMA effects

Mixing ability can be modeled using the same formalism as proposed by Griffing for combining ability (1956). In this setting, the GMA and SMA are defined as fixed effects and the residual term as random. This model is associated to the particular context of complete diallel experiment design i.e. (i) all the possible mixtures are assessed and (ii) only binary mixtures are considered, in 1:1 proportions. The use of this model in a broader context raises some statistical issues: in case of incomplete design the condition (i) is not fulfilled, so the definition of the terms becomes unclear, the estimability of the GMA and SMA is not warranted and the estimators initially proposed by Griffing are no longer valid; if requirement

(ii) is not met, a model for SMA in higher order mixture should be proposed. To address the problem of the definition of GMA and SMA resulting from assumption (i) we propose to treat the genetic effects as random effects in a mixed model (as proposed, but not developed, by Griffing 1956, and further developed by Möhring *et al.* 2011 for the prediction of combining ability in hybrids).

2.1.2 Models with random GMA and SMA effects

In order to analyze the diverse experimental designs allowed by mixtures, we have adapted the previous model to binary and higher order mixtures (but considering in a first step equal proportions for each component). First, the model can be defined to estimate a block effect and to predict GMA effects:

$$Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + e_{nbr} \qquad \text{(Model 1)}$$

where Y_{nbr} stands for the performance of the r-th replicate of mixture (or genotype) n in block b, n is the mean of the experiment, a_b is the effect of block b, $K(n)$ is the number of components of mixture n and $k(n)$ stands for the k-th genotype within mixture n (or $K(n) = 1$ and $k(n) = 1$ if n is a pure stand), $GMA_{k(n)}$ is the general mixing ability of the genotype, and e_{nbr} is the residual term.

In a second model, we included the SMA effects corresponding to first order interactions between genotypes (i.e. at the level of pairs of genotypes), supposing that higher order interactions are negligible (their estimation would require an even larger experimental effort). The model is defined as follows:

$$Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{\binom{K(n)}{2}} \sum_{k(n)=1}^{K(n)-1} \sum_{k(n)'=k(n)+1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \qquad \text{(Model 2)}$$

where a_b is the effect of block b, $\binom{K(n)}{2}$ is the number of all possible pairs of different genotypes in mixture n of $K(n)$ components when $K(n) > 1$, and $SMA_{k(n)k'(n)}$ stands for inter-genotypic interaction between components $k(n)$ and $k'(n)$ of the mixture.

It should be noticed that n can be either a mixture (m) or a pure stand (p). The pure stands are thus included in the estimation of the GMA as an intra-genotypic mixture (in contrast with Gallais 1970), with the originality to consider the SMA of pure stands (SMA_{ii}). This intra-genotypic interaction SMA_{ii} indicates how a given genotype performs in pure stand compared to the mean of mixtures comprising this genotype. In the case of pure stands (when $K(n) = 1$ and only variety $l(n)$ is grown) the model writes:

$$Y_{nbr} = \mu + a_b + GMA_{l(n)} + SMA_{l(n)l(n)} + e_{nbr} ,$$

where $SMA_{l(n)l(n)}$ stands for intra-genotypic interaction within the grown genotype in pure stand. The importance of intra-genotypic competition has been raised before, but only in the context of experiments where the performance of each component in a mixture is accessible (Gallais 1970). The access to intra-genotypic SMA through the integration of pure stands directly in the analysis therefore provides an original description of the effect of competition in a pure stand.

In addition to the residual term, the GMA and SMA terms are also defined as random effects which ensures estimability even for incomplete designs. The precise distribution of the random effects will be detailed in the following section (2.2).

Higher order mixtures might be interesting to characterize mixing ability of genotypes, since for a given number of plots, they allow to observe each genotype interacting with a higher number of partners compared to a design of binary mixtures only, though with a lower contribution of the genotype to each mixture performance.

2.1.3 Intra-genotypic interactions in mixture

In a mixture, a plant is exposed to inter-genotypic interactions but also to intra-genotypic interactions, so we modified the model to better describe the biological reality for specific effects, according to the components proportions in mixture:

$$Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{(K(n))^2} \sum_{k(n)=1}^{K(n)} \sum_{k(n)'=1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \quad \text{(Model 3)}$$

where a_b is the effect of block b. In this model, intra-genotypic interaction ($k(n)' = k(n)$) is therefore always part of the mixture performance, for any mixture order ($K(n) \geq 1$). For instance, for a given binary mixture n of genotypes i and j, the interaction term between the components can be written as: $\frac{1}{4} SMA_{ii} + \frac{1}{2} SMA_{ij} + \frac{1}{4} SMA_{jj}$ since $SMA_{ji} = SMA_{ij}$. Observe that for $K(n) = 1$, Model 2 and Model 3 coincide. The coefficients $\frac{1}{(K(n))^2}$ represent the expected weight of each kind of neighborhood in the plant community, assuming a random distribution of genotypes in the plant community. It should be noticed that due to the differences in SMA effects weighting, the SMA variance and SMA BLUPs are expected to be higher with Model 3 than with Model 2. Nevertheless, the integration of neighboring probabilities in the modeling of mixing ability (especially through the introduction of intra-genotypic interactions within mixture) constitutes a novel advance in mixture analysis.

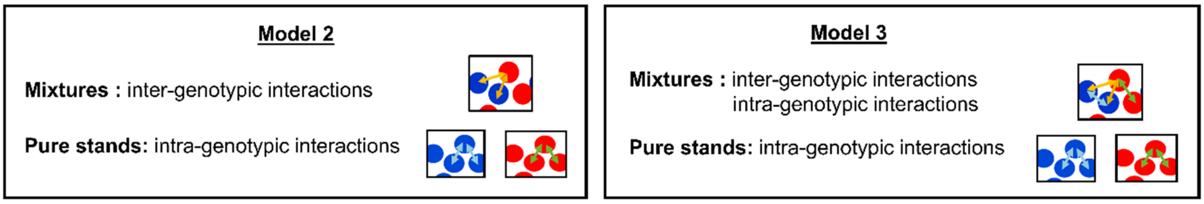

Figure 1: Schematic representation of plant interactions modeled by Model 2 and Model 3

The motivation for considering both models is that Model 2 only accounts inter-genotypic interactions within mixtures, while Model 3 also accounts for intra-genotypic interactions in mixtures (Figure 1). For convenience we adopted notations for all models similar to the Griffing's notations for combining ability. We emphasize that the interpretation of the different terms is quite different between Model 2 and Model 3. To illustrate these differences, we consider the expected performance over all possible equally weighted mixtures of order K that include a given genotype k. This amounts to work conditionally on all random effects that depend on k only, so we denote this quantity by $E(Y \mid k, K)$.

One has

$$E(Y \mid k, K) = \mu + \frac{GMA_k}{K} \qquad \text{if } K>1 \qquad \text{in Model 2}$$

$$E(Y \mid k, K) = \mu + GMA_k + SMA_{kk} \qquad \text{if } K=1 \qquad \text{in Model 2}$$

$$E(Y \mid k, K) = \mu + \frac{GMA_k}{K} + \frac{SMA_{kk}}{K^2} \qquad \text{in Model 3}$$

As a consequence, SMA_{kk} can be interpreted as the expected difference in performance between pure stand and mixture in Model 2, but not in Model 3 where it also includes intra-genotypic interaction within mixture. Although the interpretation of SMA_{kk} is different from the one of $SMA_{kk'}$ in Model 2, we assumed a common distribution for these two terms in the following to reduce the number of variance parameters to be inferred.

2.2 Statistical analysis:

Inference method

The mixed model framework has already been widely applied to plant breeding (Bernardo 1996; Falconer *et al.* 1996; Lynch and Walsh 1998; Piepho and Möhring 2007), but to our knowledge not to crop mixing ability analysis. This framework allows studying mixtures of any order in incomplete designs, ensuring model estimability (as discussed above).

Models 2 and 3 can be expressed in a matrix form as

$$y = Xb + Z_1 g + Z_2 s + e \ ,$$

where y is the vector of performances, b is the vector of fixed effects, g is the vector of the GMA random effects, and s is the vector of SMA random effects. X , Z_1 and Z_2 are the corresponding design matrices.

The random effects are assumed to be normally distributed according to the following distributions $g \sim N(0, s^2_{GMA} A_1)$, $s \sim N(0, s^2_{SMA} A_2)$ and $e \sim N(0, s^2_e I)$. The variance of the observed values can be decomposed as:

$$Var(y) = Z_1 A_1 Z_1^t s^2_{GMA} + Z_2 A_2 Z_2^t s^2_{SMA} + I s^2_e \, .$$

In the present work, we used identity matrices for A_1 and A_2 matrices. Note that one can account for *a priori* similarities between genotypes by specifying more sophisticated variance covariance structures, e.g. accounting for genetic similarities between genotypes through relatedness matrices (Bernardo 1995).

As for the estimation of the variance parameters, we used the Restricted Maximum Likelihood (REML). We used the Best Linear Unbiased Predictors (BLUP) to predict the GMA and SMA effects.

2.3 Experimental designs:

The analysis has been carried on two data sets produced by two experiments assessing mixing ability of a panel of 25 genotypes of winter wheat. This panel was designed to be representative of the phenotypic diversity of a larger set of 58 genotypes, evaluated in pure stand in a previous study for root and shoot architecture as well as other functional traits (Cantarel *et al.* in prep., Dubs *et al.* 2018). This 25 genotypes panel is composed of nine elite varieties cultivated for their high yielding performance in the Parisian Basin (Altigo, Apache, Arezzo, Boregar, Grapeli, Renan, Sogood, Soissons, Trémie), four landraces (Blé autrichien, Rouge de Bordeaux, Rouge du Roc, Saint Priest), six varieties bred for organic agriculture (Alauda, Hermes, Maxi, Midas, Ritter, Skerzzo), and six lines from a MAGIC population (A22, A208, A243, A398, F236, F426, Thépot *et al.* 2015) (Table B.1). The way mixtures of two, four and eight genotypes of the panel were designed in the two following experiments has been inspired from the ecology literature, and more specifically the Jena diversity experiment (Weisser et al. 2017), exploring extensively the gradient of intraspecific functional diversity (Dubs *et al.* 2018).

In the first experiment, the 25 genotypes were grown in a field trial both in pure stands and in 75 binary mixtures during the season 2014-2015 in the Le Moulon (LM) experimental station (48°42'37.2"N, 2°09'37.6"E) in France, according to an incomplete diallel design (75 of the 300 binary combinations: Figure 2.A). Each entry was sown in a 7,5m² plot at a density of 160 grains.m^{-2} (with 20 cm row width). All pure stand genotypes and binary mixtures were replicated twice in two randomized complete blocks. The mixtures were prepared for sowing using the thousand kernel weight in order to get equal proportion of seeds of each genotype.

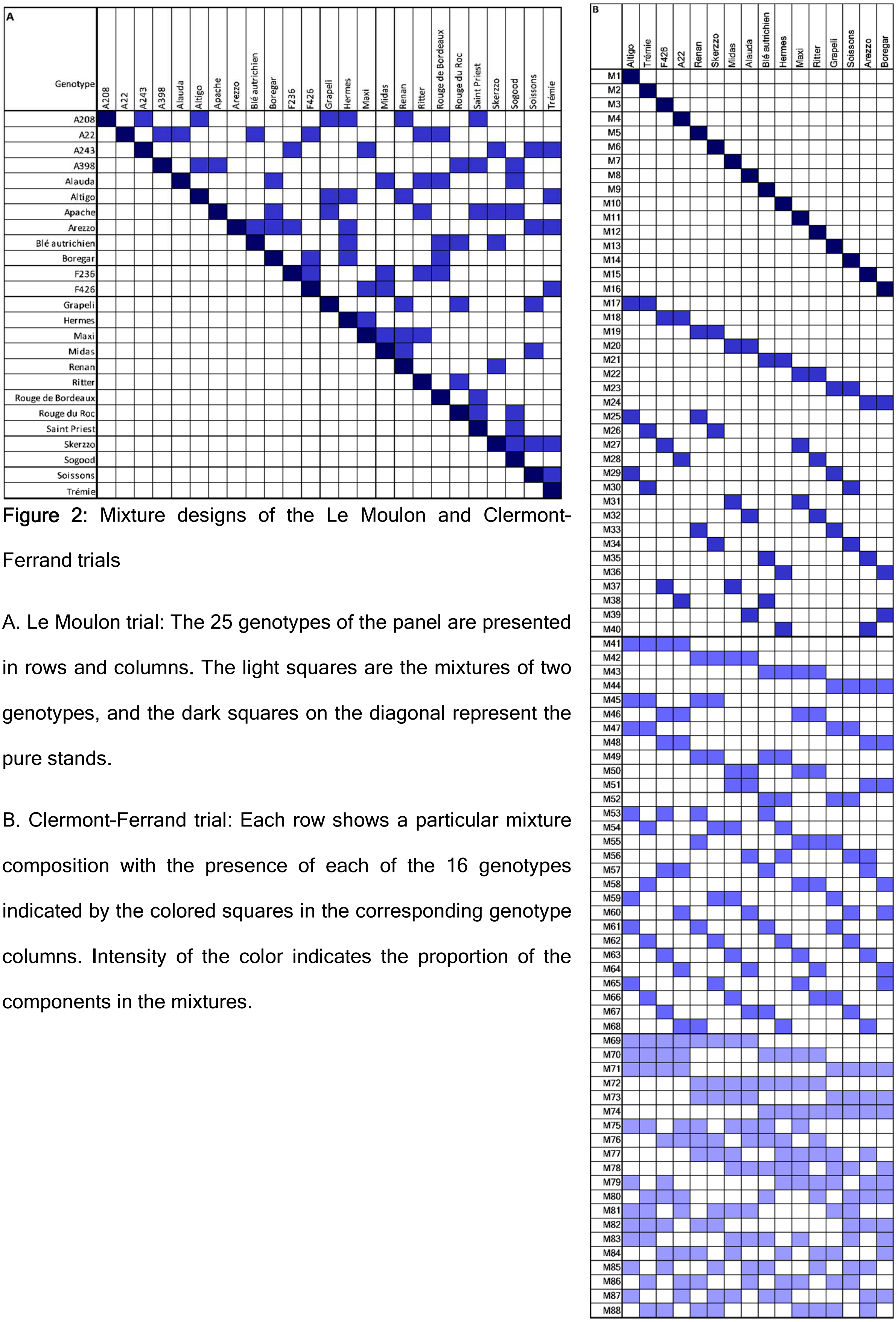

Figure 2: Mixture designs of the Le Moulon and Clermont-Ferrand trials

A. Le Moulon trial: The 25 genotypes of the panel are presented in rows and columns. The light squares are the mixtures of two genotypes, and the dark squares on the diagonal represent the pure stands.

B. Clermont-Ferrand trial: Each row shows a particular mixture composition with the presence of each of the 16 genotypes indicated by the colored squares in the corresponding genotype columns. Intensity of the color indicates the proportion of the components in the mixtures.

In the second experiment (Dubs *et al.*, 2018), a subset of 16 genotypes of the panel (Altigo, Arezzo, Boregar, Grapeli, Renan, Soissons, Trémie, Blé autrichien, Alauda, Hermes, Maxi, Midas, Ritter, Skerzzo, A22, F426) has been assessed in the INRA experimental station of Clermont-Ferrand (CF) (45°46'04.2"N, 3°08'52.2"E) in 2014-2015, in pure stand and in 24 mixtures of two genotypes (also observed in the Le Moulon trial), 28 mixtures of four genotypes and 20 mixtures of eight genotypes (Figure 2.B), where the components of each mixture were also in equal proportions of seeds. They were sown in 7,5m² plots, for a target density of 250 plants.m^{-2} and replicated in two randomized complete blocks.

In both experiments, three fungicide treatments were applied in order to study mixtures without disease development since it has already been shown that mixtures can lead to important disease reduction on wheat, and we aimed at focusing on compensation effects (Stützel and Aufhammer 1990). Herbicides were applied to avoid weed competition to interfere with wheat competition, and a nitrogen fertilization was applied (60 kg N/ha and then 40 kg N/ha of nitrogen fertilizer in the LM trial, and 60 kg N/ha in the CF trial) according to the leftovers from the previous crops on each trial to reach the common objective of 150 kg N/ha.

Phenotypic data

Yield and its components (number of spikes.m^{-2}, number of grains/spike, thousand kernel weight (TKW)) were recorded on each plot in the LM experiment, while only yield and protein content were recorded in the CF experiment.

Yield was calculated as the weight of grain harvested on the plot surface, and after measuring the humidity of a sample, it was standardized into q.ha^{-1} at 15% humidity (1quintal / hectare = 100kg / 10.000m² = 10^{-2} kg.m^{-2}). The number of spikes has been counted after flowering date, on one meter length for two adjacent rows in each plot, and converted into a number of spikes.m^{-2} taking into account the distance between rows (20 cm). TKW was measured after harvest and threshing, and the number of grains/spike was calculated based on the yield components (= (yield/ TKW)/ number of spikes.m^{-2}). Protein content was estimated on grain samples using near-infrared spectroscopy (NIRS).

2.4 Data analysis:

2.4.1 Application of the models to the datasets of each location

For each location, the correlation between the observed mixtures and the mean of their pure stand components was calculated. The main analyses (model comparison, analysis of the variance components and the BLUPs) were then performed with the three models on all the observations (mixtures and pure stands) of each location separately, and the predicted values are the modeled values fitted, on the already observed data. The correlation between all the observed values and the predictions from (i) the means of the genotypic effects of the corresponding pure stand components (according to Federer *et al.* 1982 and Gizlice *et al.* 1989), or (ii) a subset of observations of the same location (mixtures only using Model 3) has been performed for each site. Finally, the CF observed values for different mixture orders were compared with predictions obtained from CF or LM trials analyzed with Model 1. The analysis was performed using an adaptation of the *lme4* R package (Bates *et al.* 2015; see File S.1).

2.4.2 Model comparison

In order to compare the relative importance of GMA, SMA and intra-genotypic vs inter-genotypic interaction, we compared the following models:

- Model 1 which includes only the block effect and the GMA,
- Model 2 which includes the GMA and the SMA, with inter-genotypic interactions in mixture, and intra-genotypic interactions for pure stands only,

- Model 3 which includes the GMA and the SMA, with inter-genotypic interactions and intra-genotypic interactions in mixture in addition to intra-genotypic interactions in pure stand.

The models were compared based on the AIC, BIC and on Likelihood Ratio Tests (LRT, using Maximum Likelihood procedure (ML)) which were further performed for nested models (Model 2 vs Model 1, and Model 3 vs Model 1). We also computed the second-order AIC and the conditional AIC using MuMIn and cAIC4 R packages respectively, but no difference with the AIC values was observed. In addition, the R^2 and the RMSE were reported for the different models. While open to criticism (Alday 2016), the calculation of a R^2 for mixed models is a convenient tool for assessing the model quality using a common dimensionless metric for the different response variables. The R^2 were calculated following Xu (2003) to assess the proportion of the total variance which is explained by the model (conditional variance, for both fixed and random effects), as one minus the ratio of the residual variance of the full mixed model, over the residual variance of a null model (the fixed intercept-only model).

2.4.3 Within-location comparison of observed values with predictions from BLUP based on subsets of observations

We investigated the possible impact of including pure stands in the BLUP of mixture in the same location. For that, the observed values (mixtures and pure stands) were correlated to (i) the BLUP values predicted from mixture data only (i.e. without pure stands) and (ii) the BLUP from the complete data set. In both cases, BLUP were performed using Model 1 and alternatively using the best model for each response variable.

2.4.4 Inter-trial predictions

Finally, the observed values of the CF trial for different mixture orders were compared with predictions based on the GMA predicted using Model 1 on all the observations of the LM experiment.

The observed values were also compared to the predictions based on (i) the GMA in LM predicted with Model 1 fitted on the mixture observations only, or on (ii) the mean of the genotypic effects assessed on the LM pure stands only instead of GMA. Model 1 was preferred for predictions because (i) regarding breeding strategy, it is preferable to being able to predict mixture performances based on GMA than based on GMA and SMA, since using SMA requires having observed the corresponding pairs of genotypes (for using Models 2 or 3), and (ii) for practical considerations, in this study no higher order mixture had all its SMA predicted in the LM trial in order to predict their performances, so SMA are diluted in higher order mixtures. Nevertheless, as all the binary mixtures observed in the CF experiment were also observed in the LM trial (so they were all predictable from LM BLUPs), we assessed how including SMA (using Models 2 and 3) affects prediction of the yields of binary mixtures in the CF trial.

2.5 Data and program availability:

The data and the program are available on an online public repository (https://github.com/cambroise/lme4-adapt-for-variety-mixture). The program is described in File S.1.

3. RESULTS

3.1 Correlation between mixtures and pure stands

The Pearson's correlation between the observed binary mixtures performances and the means of their pure stand components was high for TKW (0.91) but moderate for yield (0.51, Figure 3), for the number of spikes.m^{-2} (0.51), and for the number of grains per spike (0.42) in the Le Moulon (LM) experiment, showing the need for modeling of mixing ability. The correlation in Clermont-Ferrand (CF) reached 0.75 for yield and 0.63 for the protein content (respectively 0.81 and 0.73 when considering only binary mixtures).

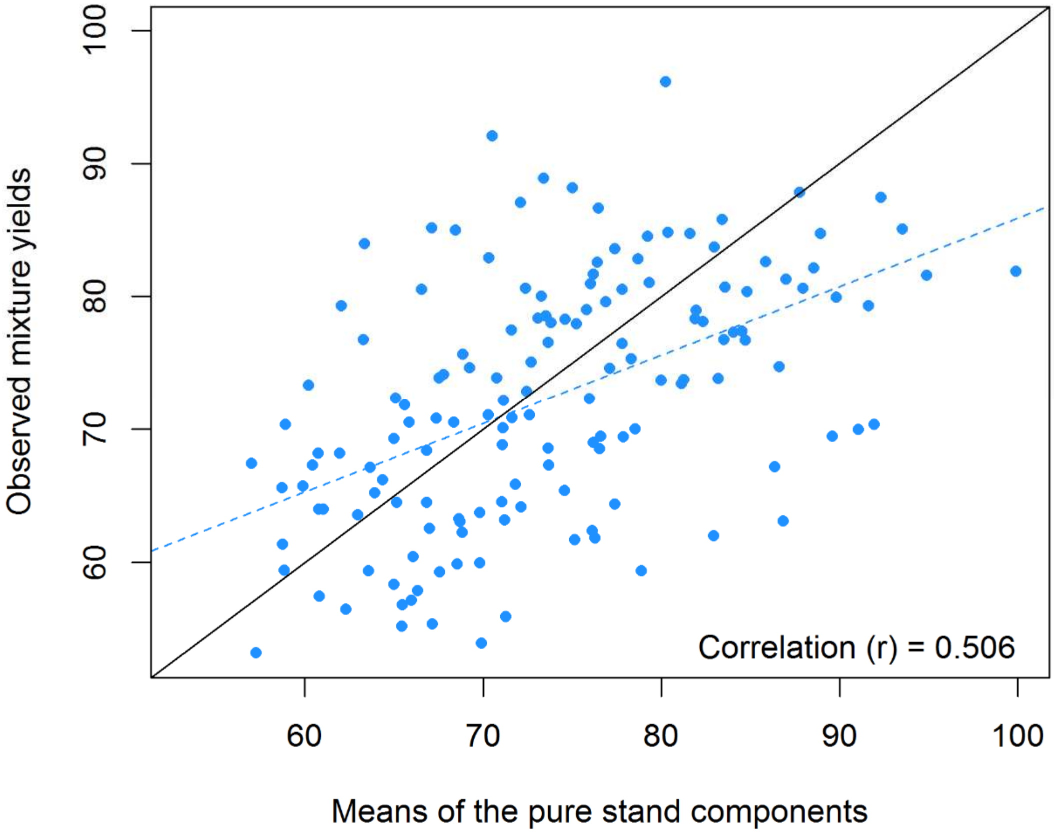

Figure 3: Observed yield for binary mixtures and the means of their pure stand components for the Le Moulon experiment

The dotted line is the regression line (y= 34.395 + 0.515 x, with standard errors of 5.395 and 0.073 respectively, and DF= 146). The black solid line is the y=x line.

3.2 Model comparison

Based on LRT (Likelihood Ratio Test) using ML procedure for model comparison, both models including SMA have been found significantly better than Model 1 for TKW (p-values of 0.026 and 0.030 respectively, Table 1). Models 2 and 3 had low but non-significant p-values (0.060 and 0.182) for spike density, while for yield and other response variables in LM and CF trials the p-values were non-significant.

Response variable	Model	AIC	BIC	p-value
Yield	Model 1	1393.3	1406.5	NA
	Model 2	1395.1	1411.6	0.664
	Model 3	1394.6	1411.0	0.385
Spike density	Model 1	2075.5	2088.7	NA
	Model 2	2074.0	2090.4	0.060 (.)
	Model 3	2075.7	2092.2	0.182
Grain number per spike	Model 1	1188.4	1201.6	NA
	Model 2	1190.4	1206.9	1
	Model 3	1190.4	1206.9	1
Thousand kernel weight	Model 1	672.5	685.7	NA
	Model 2	669.6	686.0	0.026 (*)
	Model 3	669.8	686.3	0.030 (*)

TABLE 1: Comparison of Models 2 and 3 to Model 1 in LM trial using the maximum likelihood procedure

The p-values are based on LRT (likelihood ratio test) comparisons. "." indicates significance < 0.1 and "*" significance < 0.05.

Model 1, the most parsimonious as only based on GMA effects, was as efficient as Models 2 and 3 for the number of grains per spike in LM (Figure 4.C) or for the yield and the protein content in CF (Figures 5).

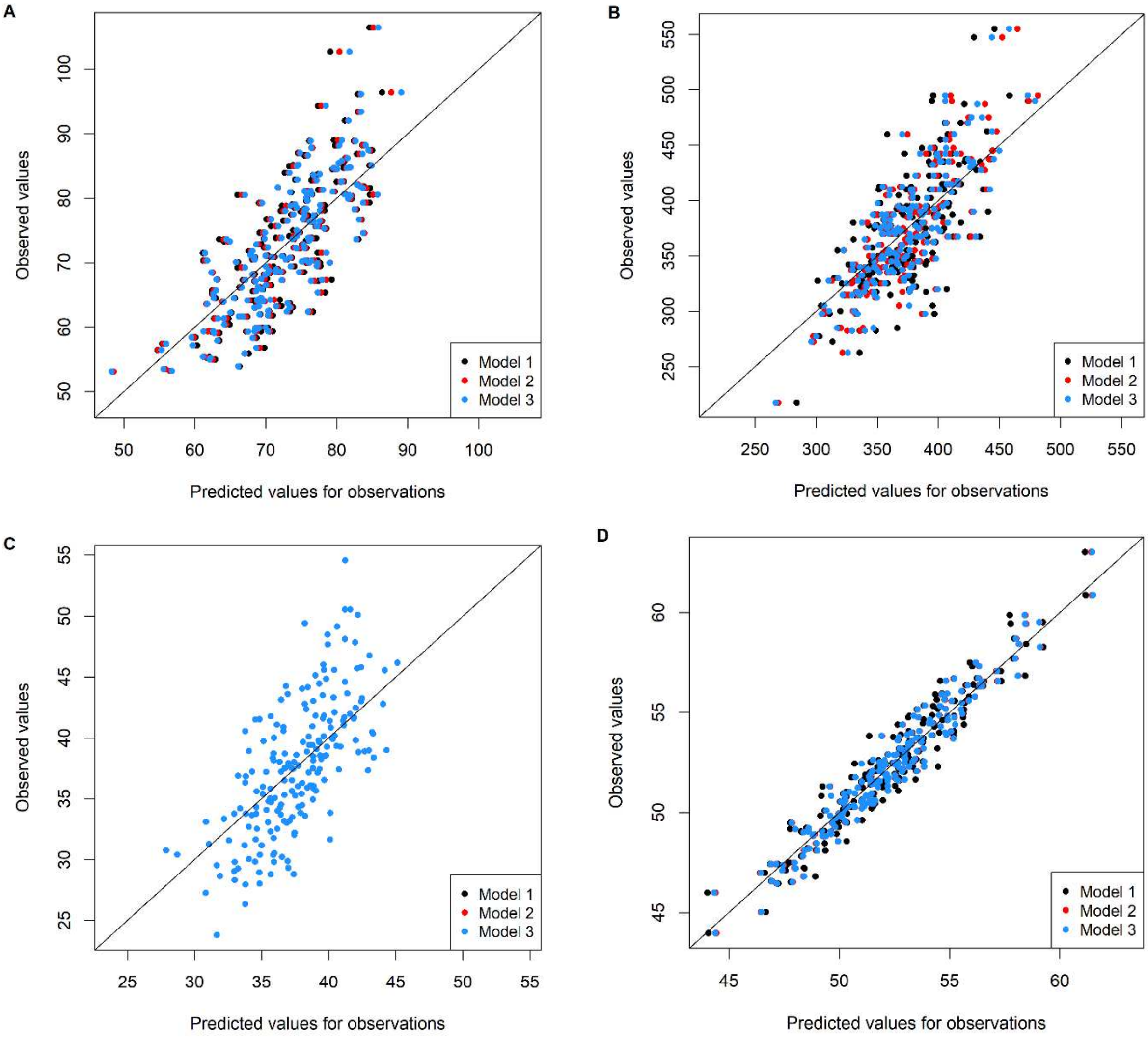

Figure 4: Observed and predicted values for each response variable in LM trial

A. Yield, B. Spike density, C. Grain number per spike, D. Thousand kernel weight. Model 1 is the model comprising fixed effects and GMA, Model 2 additionally includes SMA effects (inter-genotypic effects within pairs of genotypes within mixtures, and intra-genotypic effects for pure stands), Model 3 further includes intra-genotypic effects within mixtures. The black diagonal is the y=x line.

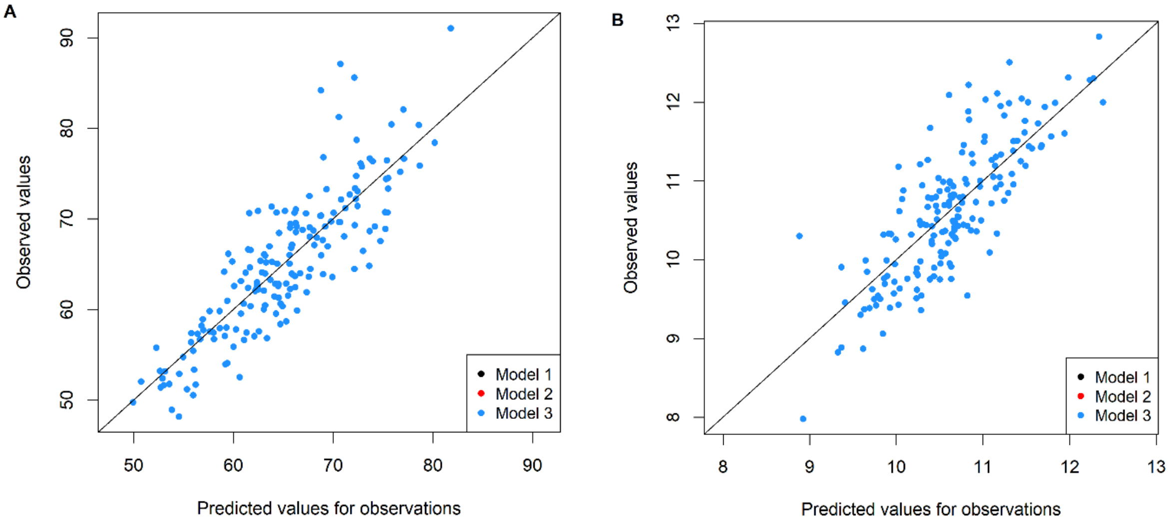

Figure 5: Observed and predicted values for each response variables in CF trial

A. Yield, B. Protein content. The black and red dots are hidden by the blue dots due to equal values for the three models. The black diagonal is the y=x line.

However, considering the yield in the LM experiment (Figure 4.A), Model 3 including SMA effects showed a higher R² and a lower RMSE (Table 2), although the likelihood of Model 3 was not significantly different from the one of Model 1 with ML procedure. For the spike density and TKW (Figures 4.B and D), Models 2 and 3 had higher R² and lower RMSE than Model 1, Model 2 appearing slightly better than Model 3 for spike density. These models including the SMA effects seemed to partly reduce the shrinkage effect observed in Figure 4.

Response variable	Model	R²	RMSE
Yield (LM)	Model 1	0.563	6.690
	Model 2	0.593	6.454
	Model 3	0.613	6.294
Spike density	Model 1	0.508	36.908
	Model 2	0.650	31.125
	Model 3	0.613	32.702
Grain number per spike	Model 1	0.456	4.023
	Model 2	0.456	4.023

	Model 3	0.456	4.023
TKW	Model 1	0.911	0.923
	Model 2	0.941	0.754
	Model 3	0.940	0.758
Yield (CF)	Model 1	0.728	4.174
	Model 2	0.728	4.174
	Model 3	0.728	4.174
Protein content	Model 1	0.642	0.502
	Model 2	0.642	0.502
	Model 3	0.642	0.502

TABLE 2: Criteria for characterization of the model fits in Le Moulon and Clermont-Ferrand trials

3.3 Comparison with predictions based on the pure stands

When using mixing ability models fitted on all the observations instead of the pure stand genetic effects, the R^2, the RMSE, the Pearson's correlation coefficient (on the values of the mixtures and the pure stands) and the Kendall's correlation coefficient (on their ranking) between observed and predicted values were improved for most of response variables in LM experiment, but this was marginal on the CF trial, as shown in Table B.4. For instance, for the yield in LM trial, the Pearson's correlation coefficient increased from 0.687 with predictions based on the means of the pure stands (*p*), to 0.790 with predictions based on BLUP predicted with Model 3 (*m+p*) showing a higher ability to predict mixture performances. In the CF experiment, the correlation between observed and predicted values was as high with predictions based on the mean of the genotypic effects of the pure stand components (0.840, Table B.4) as with BLUP predicted from mixtures and pure stand observations (0.854, with Model 1).

3.4 Variances of GMA and SMA

The variance components estimated with each model for yield and for the number of spikes.m^{-2} are detailed below for the LM and CF experiments (Table 3). The other responses are given in supplementary material (Table B.2). It should be noticed that Models 2 and 3 differ in the weighting of the SMA effects, and it is expected for estimated SMA variance to be higher with Model 3 than with Model 2 (as well as SMA BLUPs).

Response variable	Model	σ_{GMA}^2	σ_{SMA}^2	σ_e^2	$\sigma_{SMA}^2/\sigma_{GMA}^2$
Yield Le Moulon trial	Model 1	63.57	NA	50.48	NA
	Model 2	63.10	2.11	48.66	0.03
	Model 3	54.69	14.67	47.40	0.27
Yield Clermont-Ferrand trial	Model 1	95.18	NA	19.20	NA
	Model 2	95.18	0	19.20	0
	Model 3	95.18	0	19.20	0
Number of spikes.m^{-2} Le Moulon trial	Model 1	1621.82	NA	1531.06	NA
	Model 2	1536.02	303.75	1267.18	0.19
	Model 3	1357.87	822.18	1338.40	0.61

TABLE 3: Variance components and ratio of variances in Le Moulon and Clermont-Ferrand trials

In all experiments and with all considered mixed models, the GMA variance was always higher than the residual variance and was more than 10 times higher in the case of the TKW in LM experiment (Table B.2). The amount of the SMA variance compared to the GMA variance ranged from 0 to 0.19 with Model 2 and from 0 to 0.61 with Model 3. The highest rates were obtained for spike density, while the number of grains per spike and the CF response variables did not show any specific effects. Consequently, the three models were equivalent for variance partitioning in the CF experiment.

3.5 GMA and SMA predictions

The predicted values of GMA and SMA with Model 3 for yield in LM trial are shown in Figure 6 (in Figure B.1 for spike density and TKW with Model 2 in LM trial) and the GMA values for each response variable with the best model are given in Table B.3.

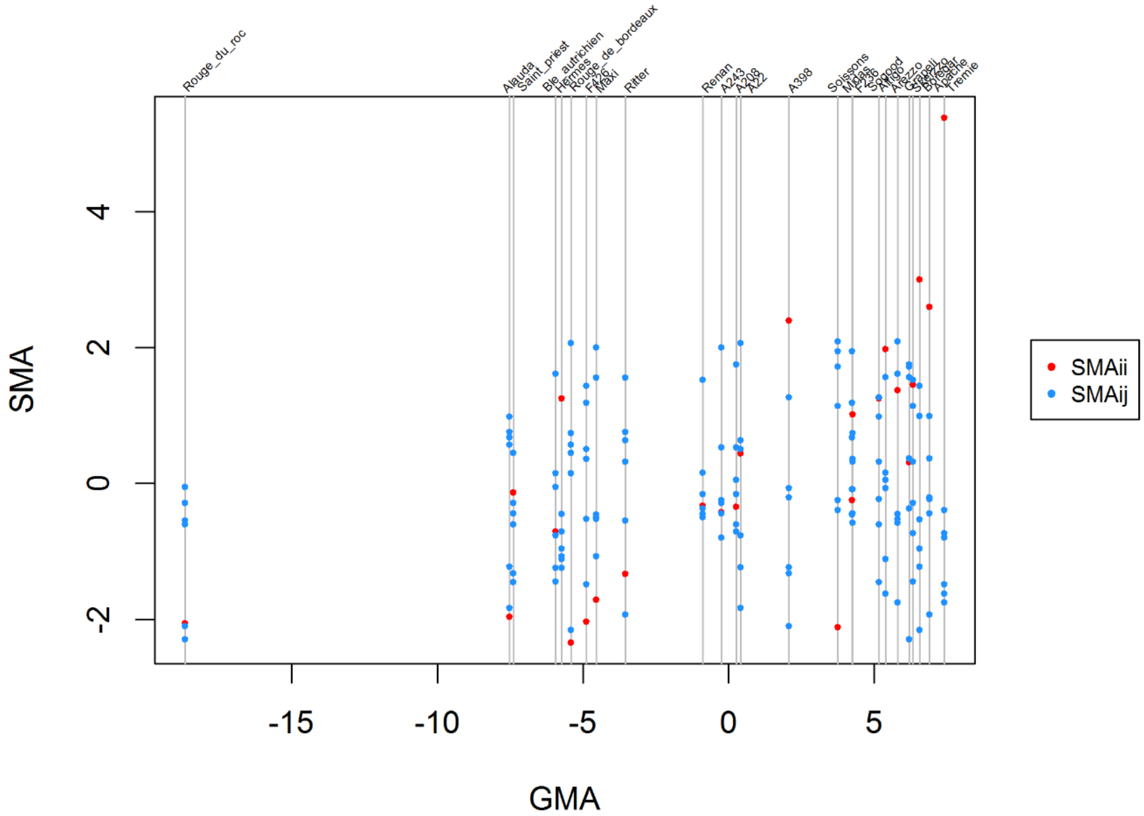

Figure 6: Predicted values for GMA and SMA per genotype for yield in LM trial with Model 3

Values in quintal per hectare (10^{-2} kg.m^{-2}). The red dots represent the intra-genotypic SMA, and the blue dots the inter-genotypic SMA.

The GMA for yield at LM ranged from -18.67q.ha^{-1} to 7.41 q.ha^{-1} (with Model 3), while the SMA showed lower ranges (from -2.34 q.ha^{-1} to 5.38 q.ha^{-1}). The range of SMA values per genotype with Model 3 varied among genotypes, with varieties such as Renan and Midas obtaining similar SMA with most genotypes (Figure 6). Overall, intra-genotypic SMA effects displayed both high and low values, while inter-genotypic SMA were somehow more intermediate (Figure 6). The intra-genotypic SMA obtained with Model 2 were less extreme for spike density and TKW than for yield (Figures B.1 A and B). The intra-genotypic SMA with Model 3 was overall positively correlated with the GMA for yield (0.679). This was also the case for spike density (0.659) but the correlation was lower for TKW (0.235).

The elite variety Trémie had the highest GMA and displayed the highest SMA when grown in pure stand (Figure 6). Interestingly, two genotypes with relatively low GMA (A22 and Rouge de Bordeaux, Figure 6) showed a rather high SMA value when grown together in a mixture (Figure 7). More extreme values were obtained with Model 3, due to higher SMA variance (as expected). This was also the case for spike density and TKW in LM trial although to a lesser extent (Figure B.2). The inter-genotypic SMA for Model 2 and Model 3 were very highly correlated for the yield, spike density and TKW (0.9993, 0.9996 and 0.9998 respectively).

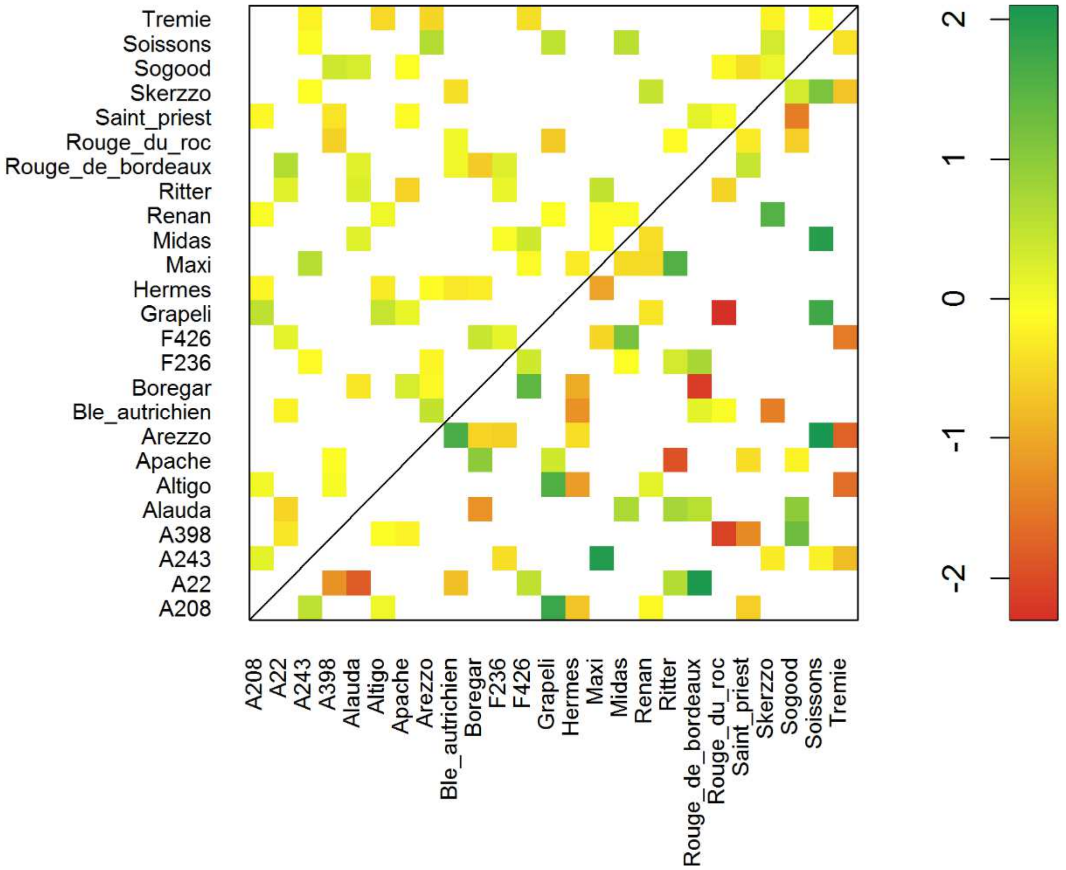

Figure 7: SMA predicted values for yield with Models 2 and 3 in LM trial

The SMA obtained with Model 2 and Model 3 are plotted respectively above and below the diagonal (black line). The intra-genotypic SMA are not represented on the plot.

The GMA was highly correlated with the pure stand genotypic effect in the CF trial (0.98 for the yield with Model 1, Table B.5), and to a lesser extent in the LM experiment (0.88). The correlation between the GMA obtained in both experiments for yield using Model 1 was also quite high (0.903). The CF trial displayed higher variance for the GMA values than did the LM trial (Figure 8).

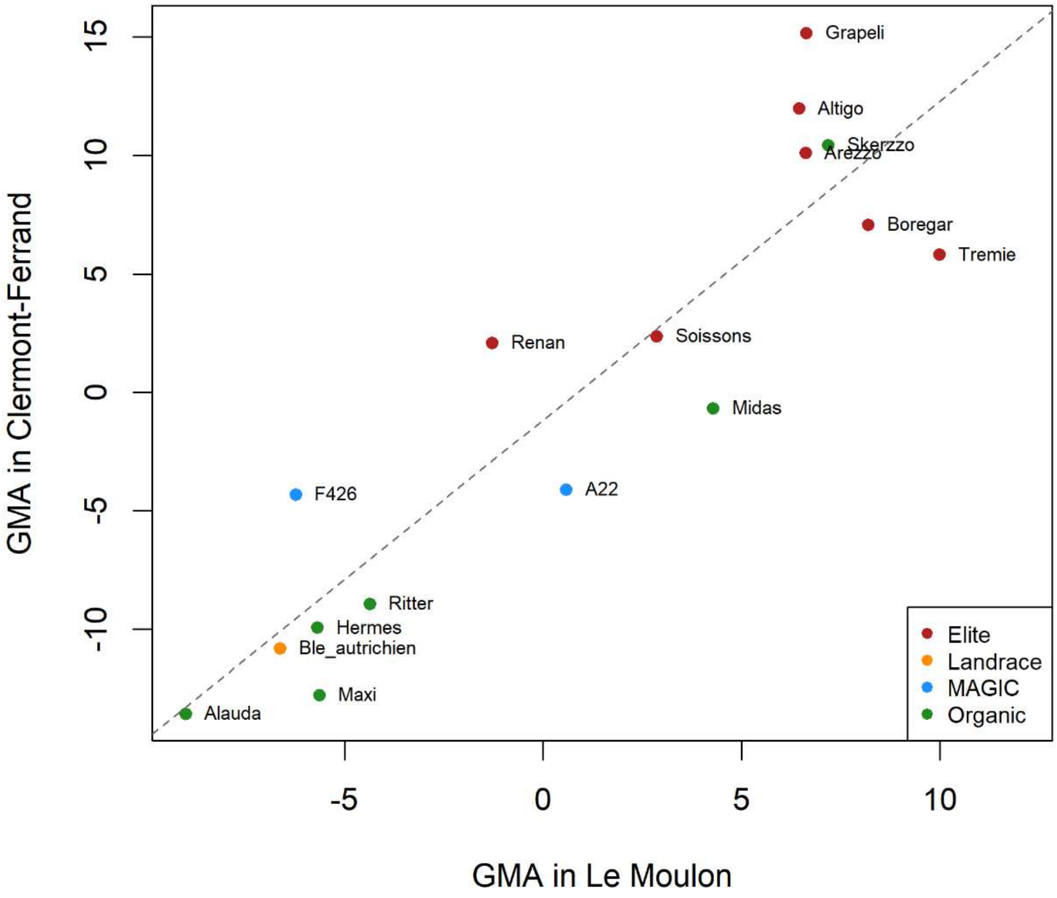

Figure 8: GMA values predicted with Model 1 for yield in the two experiments (LM and CF)

The dotted line is the regression line between the GMA of the two experiments (y= -1.174 + 1.347 x, with standard errors of 1.066 and 0.171 respectively, and DF= 14).

The elite varieties obtained the highest GMA in both experiments, while most organic varieties and the landrace had the lowest, although some organic varieties also obtained high GMA. Organic varieties performed better in LM trial than in CF trial. The best varieties were more spread in CF trial than in LM trial.

3.6 Predictions based on subsets of observations

In the LM experiment, when the analysis was performed on mixture observations only (*m*) instead of all observations (*m+p*), the variance of SMA for yield (estimated with both Models 2 and 3) became null. The correlation between the observed and predicted mixtures with Model 3 was 0.786 for yield when using GMA$_{m}$, not different from the one obtained with *m+p* observations (0.790), and the RMSE was reduced from 6.3q.ha^{-1} to 5.8q.ha^{-1} (Table B.4). This result suggests that mixing ability analysis might be performed without requiring pure stands in the dataset. The GMA for yield were modified (Figure B.3) although the correlation between GMA$_{m}$ and GMA$_{m+p}$ with Model 3 was 0.91 (Table B.5). In the CF experiment, the correlation between GMA$_{m}$ and GMA$_{m+p}$ (with Model 1) was even higher (0.96, Table B.5), and the correlation between observed and predicted values was equal when excluding the pure stands for GMA predictions (0.834) to the one when pure stands were included (0.854, Table B.4).

3.7 Predictions of CF high order mixtures based on CF BLUP

Both Pearson's and Kendall's correlations between the observed values for pure stands and mixtures in the CF trial and the predicted values based on the CF mean of the genotypic effects in pure stands decreased with the order of the mixtures (Table 4), indicating that the four-way and eight-way mixtures had the largest level of deviation from additivity in the CF experiment.

	Pearson's correlation						Kendall's correlation					
Mixture order	CF GE$_{p}$	CF GMA$_{m+p}$	CF GMA$_{m}$	LM GE$_{p}$	LM GMA$_{m+p}$	LM GMA$_{m}$	CF GE$_{p}$	CF GMA$_{m+p}$	CF GMA$_{m}$	LM GE$_{p}$	LM GMA$_{m+p}$	LM GMA$_{m}$
All orders	**0.840**	**0.854**	**0.834**	**0.653**	**0.785**	**0.741**	**0.651**	**0.669**	**0.664**	**0.498**	**0.598**	**0.573**
1	0.973	0.955	0.876	0.718	0.876	0.770	0.871	0.831	0.734	0.548	0.706	0.621

2	0.836	0.862	0.786	0.662	0.805	0.751	0.649	0.677	0.700	0.505	0.667	0.613
4	0.742	0.777	0.789	0.599	0.695	0.727	0.516	0.554	0.582	0.405	0.481	0.541
8	0.700	0.713	0.719	0.574	0.649	0.680	0.544	0.569	0.569	0.438	0.500	0.538

TABLE 4: Correlation between observed and predicted values or ranking for the yield in the Clermont-Ferrand (CF) trial based on the CF or the Le Moulon (LM) mean of the genotypic effects in pure stands (GE) or GMA predicted with Model 1 on different sets of observations in the CF and LM trials

The order of the mixture indicates the number of components: one for pure stand, two for binary mixture, four for four-way mixture and eight for eight-way mixture. The capital letters indicate the trial for observations for genotypic effects in pure stand or GMA predictions using Model 1. The lower case letter specifies if pure stands observations were used (p) or mixture observations (m) or both (m+p) in each trial for GMA predictions.

3.8 Inter-trial predictions

In order to assess the ability of the BLUP values obtained in a design based on binary mixtures to predict the observed values for higher order mixtures, we computed the predicted values for the CF observations based on the LM GMA using Model 1 (Figure 9). The Pearson's correlation between the observed values in CF and predictions based on LM was 0.785 (and 0.598 for Kendall's correlation, Table 4). The correlation was slightly lower when excluding the pure stand observations in LM of the analysis (0.741). However, the correlation was always much higher than when using the mean of the corresponding genotypic effects of pure stand components (correlation of 0.653). When including SMA in addition to GMA predicted on all observations in the LM trial (using Models 2 and 3) for predicting yields of binary mixtures in the CF trial, the correlation between observed and predicted yields was not clearly improved (0.814 with Model 2 and 0.817 with Model 3, vs 0.805 with Model 1).

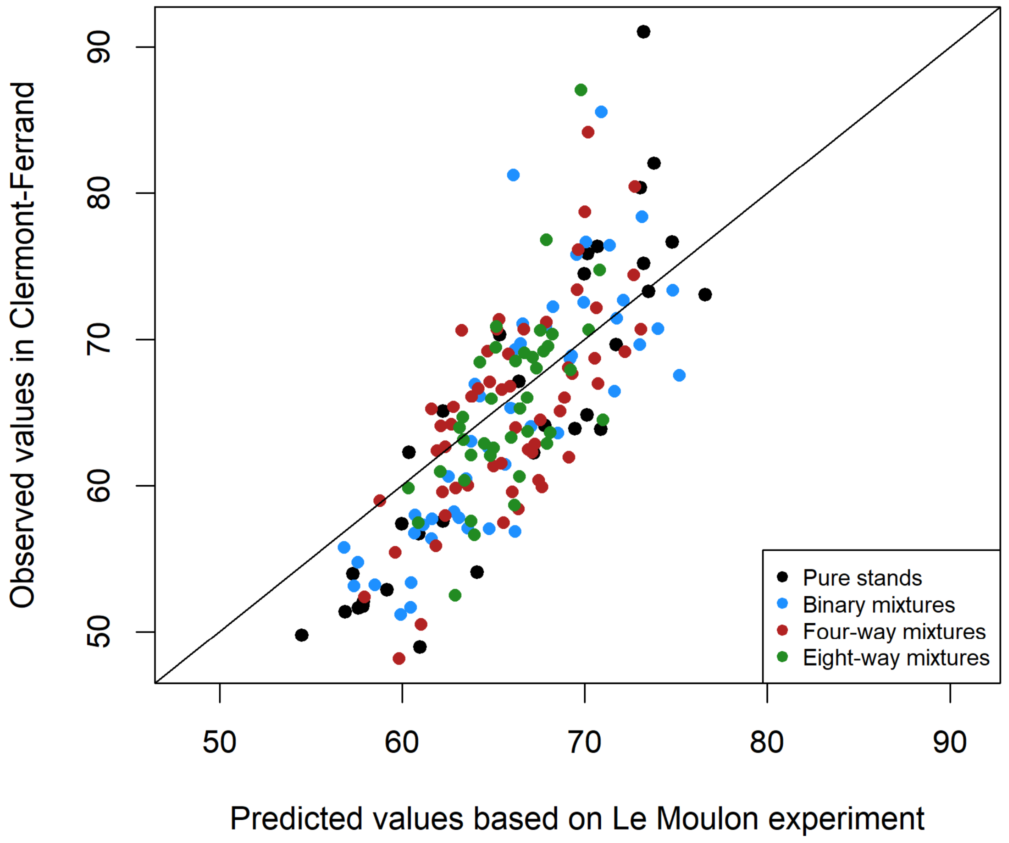

Figure 9: Predicted yields for the Clermont-Ferrand experiment based on GMA predicted from all observations in the Le Moulon trial with Model 1 and the corresponding observed yields

The black diagonal is the y=x line.

As for predictions based on CF BLUP, when using LM BLUP, the Kendall's correlation coefficient between the predicted values and the observed values decreased with the mixtures complexity (Table 4). Interestingly, when removing the pure stands in LM observations for fitting Model 1, the observed vs. predicted correlations were lower for pure stands and binary mixtures than when including pure stands in LM, while they were rather similar for four-way and eight-way mixtures.

4. DISCUSSION

Mixed models are pivotal statistical tools for the prediction of breeding values in modern breeding programs, and are commonly used to analyze diallel designs of hybrid crosses. This formalism has surprisingly never been applied to cultivar mixtures analysis, despite the similarity between Combining Ability and Mixing Ability. Herein we adapted mixed models for the study of mixing ability in diallel-like designs, unlocking the analysis of incomplete designs.

We combined the Griffing formalism for combining ability (Griffing 1956) and mixing ability modeling with the recent developments in mixed models (Bates *et al.* 2015) adapted to fit specialized linear mixed models. The approach is highly flexible as it can be applied to binary and any order mixtures, including pure stands.

4.1 General Mixing Ability predominates in two experiments on wheat mixtures

This approach has been applied to two experimental designs of bread wheat mixtures: one based on pure stands and binary mixtures, and the other including higher order mixtures too. In the first experiment (Le Moulon, LM), the correlation between the observed mixture performances and their predicted performances based on their pure stand means was moderate (0.51, 0.51 and 0.42 respectively for yield, for the number of spike.m^{-2} and for the number of grain/spike), except for TKW (0.91), underpinning the importance of estimating mixing ability of the genotypes. Using mixing ability modeling allowed to improve the correlation between observed values and predictions to 0.79 for the yield in Le Moulon (LM) experiment for instance (with Model 3).

The part of variance explained by GMA and SMA effects was relatively high, given that the proportion of observed mixtures on the total number of possible mixtures was rather low (75/300=0.25). This suggests that using a random sample of 25% of the total number of all possible mixtures, in addition to the pure stands, seems efficient to assess GMA and SMA of the genotypes used as components, and this is in accordance with previous findings for combining ability in wheat (Zhao *et al.* 2015). This result calls for the possibility to screen a large number of genotypes from a diversified panel for their GMA using incomplete designs instead of growing all pairwise mixtures within a limited set of genotypes. In the second experiment, the correlation between the mixture performances and the means of the pure stands components was rather high for yield (0.75) and for protein content (0.63) so the predictions of all the observations based on the pure stand performances were efficient and were only marginally improved when using mixing ability modeling. This is explained by a very high correlation between the GMA and the genotypic effects of the pure stands for yield in this trial (0.98), for reasons discussed hereafter (section 4.3).

4.2 An original modeling of specific mixing ability

Specific Mixing Ability was modeled using two different approaches that could better reflect the underlying plant-plant interactions:

- in addition to inter-genotypic interactions within mixtures, we introduced intra-genotypic interactions within pure stands (Model 2). Note that in the literature, pure stands are not always considered as particular cases of mixtures and therefore are usually not included as such in the analysis;

- we further refined the SMA estimates, by accounting for intra-genotypic interactions not only within pure stands but also within mixtures. For that, SMA effects were weighted by the probabilities of neighborhood of varieties in a mixture (Model 3).

These two models resulted in differences in variance components and in ratio of variance between SMA and GMA, with a higher SMA variance observed with Model 3, as expected. The intra-genotypic SMA of yield had noticeably more extreme values in the SMA distribution when using Model 3 compared to Model 2, suggesting that the intra-genotypic interactions in mixtures significantly accounted for SMA variance in Model 3 for this trait. The interpretation of the intra-genotypic SMA should be different depending on the model used. Model 3 might be more realistic regarding the biological mechanisms occurring within mixtures. For instance it is expected that the intra-genotypic interactions should be lower in higher order mixtures. Intra-genotypic interactions SMA_{ii} are by nature quite different from the one of inter-genotypic interactions SMA_{ij}. Consequently one should estimate separately one variance component for SMA_{ii} and one for SMA_{ij} (with $i \neq j$) for both Models 2 and 3. However this would require the inference of more variance parameters - each with less available information - which in turn would require the evaluation of more mixtures per genotype than available in this study.

If the independence between the performance of pure lines and their hybrids (e.g. design 4 in Griffing 1956) can be explained by genetic interactions (i.e. dominance), it is biologically difficult to neglect intra-genotypic interactions in mixtures. However, if the focus is to detect genotypes that are particularly good partners for complex mixtures, it might be more relevant to remove pure stands. Running the analysis without the pure stands gave lower GMA predictions for the high yielding elite varieties (e.g. Trémie, Boregar ... Figure B.3) while different varieties such as Soissons showed higher GMA, therefore appearing as particular good mixing partners. In that case, no SMA effect was detected for yield, while it was still present for spike density ($p=0.06$) and TKW ($p=0.009$) (data not shown). Another approach has been proposed in the literature to account for the inter- and intra-genotypic interactions within mixtures. It is based on the partitioning of the GMA into a GPA (General Performing Ability) reflecting the genotypic effect in pure stand, and a TGMA (True General Mixing Ability) being the part of the GMA truly due to the mixing conditions obtained when analyzing the over-yielding (OY, i.e. the difference between the yield of the mixture and the mean of its components in pure stands) as a response variable (Federer *et al.* 1982; Gizlice *et al.* 1989; Knott and Mundt 1990; Lopez and Mundt 2000). In these studies, both GPA and TGMA have been found important for mixing ability. Here, we chose to include the pure stands in the observations for modeling mixing ability, since SMA of the pure stands (SMA_{ii}) in Model 2 provides sufficient information on how each genotype performs in pure stands in comparison with its GMA in the experiment. Both approaches are strongly divergent in their philosophy: in the GPA/TGMA approach, pure stands and mixtures performances are used independently to estimate GPA and GMA, and the pure stand performances are considered as a reference to compute the TGMA. In the present model, pure stands and mixtures are jointly used to estimate GMA and SMA, and the intra-genotypic interaction (SMA_{ii}) has the same status as the inter-genotypic interaction (SMA_{ij}), which seems biologically sound.

4.3 Variance components vary between the two studied locations

For all response variables in the binary mixtures trial (LM), the variance of GMA was higher than the one of SMA, in accordance with Knott and Mundt (1990) in similar trials, while no SMA effect at all was found in the experiment involving higher order mixtures, as in Lopez and Mundt (2000). However, only GMA/SMA estimates for yield can be compared among the two experiments since the other response variables differed among LM and CF. The absence of SMA in the CF experiment might be due to three possible factors which are confounded in this experimental design, and might contribute to limit the inter-genotypic interactions in CF. Firstly, the CF trial was conducted under more favorable conditions (higher yield objective and nitrogen inputs) that might have led to less stresses for the plants and therefore less opportunity for complementarity or synergy mechanisms to express among mixtures' components. Secondly, the panel used in the CF experiment was a subset of the panel used for the Le Moulon experiment, the genotypes were less contrasted than in the LM trial (only one landrace included in the panel, lower diversity in functional traits) resulting in lower competition/synergies. Thirdly, the CF experiment involved higher order mixtures in which SMA within each pair of genotypes might have been too low to be detected (increasing the number of binary interactions results in a dilution of their effects), or the possible occurrence of higher order interactions might have masked the binary ones. Thus, this could lead to reducing the possibility to observe SMA effects in the CF trial compared to the LM trial.

For yield in the LM trial, the share of the genetic variance due to SMA effect was larger with Model 3 than with Model 2, indicating that the SMA variance might also be due to intra-genotypic interactions within mixtures. However, both models including SMA did not provide a significantly better fit than the model with GMA only (and the differences in AIC values for the three models were very small for all response variables (Table 1)), as already found by Gizlice *et al.* (1989). In contrast, Gallandt *et al.* (2001) in wheat or Federer *et al.* (1982) in bean found significant SMA. Spike density displayed a SMA variance ratio of 0.61 with Model 3, indicating strong interactions between plants for this response variable. This is consistent with the fact that the number of tillers and therefore the number of spikes are known to be the most plastic traits when changing plant density (Darwinkel 1978) and/or the phenotypes of neighboring plants in mixtures (Jackson and Wennig 1997; Cowger and Weisz 2008, Lecarpentier 2017, Lecarpentier *et al.* in press). This is also in line with the clear-cut difference observed in the range of spike density in the LM trial when comparing pure stands and binary mixtures (respectively 217.5 to 490 for pure stands and 262.5 to 555 for mixtures). These specific effects observed on mixtures occurred in the experiment characterized by a low density leading to plasticity for tillering, and selection effects between genotypes due to differences in potential productivity. As this leads to differences in genotype frequencies in the harvested grain, it might be interesting to measure the proportion of grains (or spikes) of each genotype in mixtures to further investigate the specific effects and better predict the mixture performances based on individual contributions of the components. The estimation of selection effects may help to better understand the changes in proportions of the genotypes over time (Finckh and Mundt 1992; Loreau and Hector 2001; Barot *et al.* 2017) but this requires particular designs or systems for separating the mixture components.

4.4 BLUP

In the LM experiment, the Trémie genotype showed the highest BLUP value for both GMA and SMA_{ii} due to its high productivity in particular in pure stand. However, the binary mixtures involving this genotype had among the lowest SMAs and all had lower yields than expected based on the pure stands components (i.e. negative OY), therefore indicating that Trémie, as an elite variety selected for performing well in pure stand conditions, might not be the best "mixing partner" to combine. On the contrary, the genotype Soissons had intermediate GMA (in both trials) and a low SMA in pure stand, while its SMAs in mixtures were high and the mixtures displayed high yields (Soissons also had the highest GMA when considering mixtures only). This could be due to a favorable mixture design for this genotype, but the OY was always positive for this genotype (except when mixed with Trémie) with a mean of +7.7q.ha^{-1} showing its potential for mixing conditions. The important mixing ability for yield of this genotype might be explained by a high mixing ability for spike density. These results show the importance of taking both GMA and SMA into account for selecting candidates for mixtures.

4.5 Prediction of mixtures and pure stands performances based on BLUP values

Prediction of extreme observations was less accurate for response variables for which the correlation between mixtures and corresponding means of the pure stand means was lower. This might be because the GMA is an average effect leading to a shrinkage of the predicted values towards the mean of the observed mixtures performances involving the given genotype (which is partly corrected when including SMA in the models). It should also be noted that the GMA of the genotypes are relative to the assessed panel. This is particularly important since the GMA are computed using a common measure for all the components of a given mixture. In this regard, Federer *et al.* (1982) proposed an alternate model to analyze mixtures when individual component performances are available. Using this alternate model should lead to a reduction of the shrinkage effect since the GMA of a genotype is not dependent on the performances of its associated genotypes. Finally, the mixture performances are predicted assuming that the proportions of the components are the same as those used for sowing.

4.6 Predictions of higher order mixtures from binary mixtures evaluated in another experiment

The GMA and SMA have been found quite variable over environments in other studies (Knott and Mundt 1990; Gallandt *et al.* 2001) but interestingly the GMAs predicted for yield by the experiments at Le Moulon and Clermont-Ferrand proved to be highly correlated (0.903) although one site had higher order mixtures, and the Clermont-Ferrand panel was only a subset of the Le Moulon panel. We assessed the ability of the pure stands and the binary mixtures of the LM experiment to predict the CF experiment mixtures and especially the higher order mixtures. The correlation between predicted and observed values was found decreasing with increasing orders mixtures (but for binary mixtures, it should be borne in mind that all binary mixtures observed in the CF trial were also observed in the LM trial). However, predictions based on binary mixtures were better than based on pure stands, as previously observed in literature (Lopez and Mundt 2000; Mille *et al.* 2006). The Kendall's correlation coefficient was also calculated since the concordance in the ranking of mixture performances is desirable for breeding perspectives. This correlation was at least equal when the four-way and the eight-way mixtures were predicted from the LM observations without the pure stands (m) than when they were included ($m+p$), suggesting that the exclusion of pure stands from the GMA predictions does not degrade the predictions of high order mixtures. A higher correlation between observed and predicted values was expected when using all observations ($m+p$) than when using mixtures only (m), but the aim was to assess the impact of the pure stand information on the correlation.

4.7 Improving screening and breeding strategies

The two mixture designs used in this study were performed empirically, as in all agronomical and ecological experiments we have reviewed. This raises the question of how to optimize the experimental designs used to better estimate GMA/SMA. As far as we know, this question has been poorly addressed in the literature. For example Federer (2002) and

Raghavarao and Federer (2003) have proposed balanced incomplete designs to achieve estimability of the GMA and SMA for a specific number of varieties and three-way mixtures, without giving rules on the way varieties should be mixed. The experimental design optimization remains to be studied. The BLUP framework proposed here will allow such optimization, exploring various GMA/SMA ratios, according to panel/species biological characteristics.

Different designs should be chosen for the estimation of mixing ability, depending on the objective of the experiment (i.e. assessment of GMA-SMA for a panel, improvement of mixing ability in breeding scheme) and the structuration of GMA and SMA effects. Once a design has been chosen for estimating mixing ability, it can be optimized using power tests to maximize the accuracy of the GMA and SMA predictions, whether or not using higher order mixtures instead or in addition to binary mixtures, whether or not including the pure stands, varying the proportion of possible mixtures to be observed, the number of observations per genotype, the number of replicates per mixture, the mixture composition, … and taking into account the gain in accuracy regarding the costs and efforts invested in the experimental design. The comparison of the power of different types of incomplete designs (nested design, balanced incomplete factorial design, topcross design, random design as used by Zhao *et al.* (2015) for combining ability analysis) may allow to identify the most adapted and powerful type of design regarding a particular situation, and to tailor it into an optimized design for mixing ability estimation. Since mixing ability might vary across environments and might be subject to interactions with environment, the decision to replicate the experimental design in different environments would depend on the objectives, resources and potential GxE interactions.

Parameters to consider for assessing mixing ability are the correlation between the performances of mixtures and predictions from their pure stand components, and the importance of the GMA vs SMA effects. Therefore, in a breeding scheme, a preliminary experiment must be carried out in order to characterize the panel first and thus guide the breeding strategy and objectives. This study exemplifies the kind of preliminary study that can be developed to assess variance components. Three cases are expected: i) high correlation between the mixture performances and the mean of their pure stand components, ii) important GMA versus SMA, iii) important SMA.

In the case of a strong correlation between mixtures and the mean of their pure stand components, information on the performances of pure stands can be used since it requires less observation. In contrast, if the GMA is moderately or weakly correlated to the genotypic effects in pure stands or if SMA effects are detected, it would be necessary to include mixtures in the design, and consider excluding the pure stands in order to limit the number of plots especially for the development of high order mixtures. In addition to this advantage, in the LM trial, the exclusion of pure stands for predicting GMA and SMA effects led to a slight improvement in correlations between observed and predicted values for yield, spike density and TKW, while the SMA was reduced to zero for the yield.

When the GMA is the major source of variation, compared to the SMA, it could be wiser to focus on the components *per se* for mixing rather than on combinations of components. The experimental design should be based on genotype screening under conditions that allow interactions between different genotypes such as alternate rows (as proposed by Barot *et al.* 2017) or mixtures with a tester (that can be a genotype with high GMA or eventually a mixture itself to save space and if the further purpose is the development of higher order mixture) for instance. If the SMA is low but not negligible, it might be interesting to search for panel structure to optimize the experimental design.

When SMA effects are important, the existence of complementarity groups for mixing ability should be investigated. As carried out for hybrid development, the use of clustering methods to search for heterotic-like pattern might allow to capture some of the specific effects in the GMA (increasing the accuracy of mixture performance predictions), and to further improve the complementarity between components within mixtures structuring the panel into groups (Zhao *et al.* 2015). For the development of higher order mixtures, the presence of higher order specific effects might be investigated (Federer 1999) to decide if the assessment of binary mixture is sufficient or if higher order mixtures should be included in the design. If each genotype is observed in several mixtures, the stability of the SMA effects involving the genotype might be assessed, allowing to identify the most stable genotypes for specific effects; and stability of SMA over environments might be investigated.

The mixing ability structuration between GMA and SMA has consequences on the conception of the breeding strategy. When the mixture performances are highly predictable based on pure stands (first case), any particular adaptation of the breeding scheme is required for selecting the best performing genotypes, other than assuring that the final mixtures are agronomically coherent. When the selection focuses on GMA (second case), genotypes should be assessed and selected in interaction with each other or with testers. Interaction traits (e.g. the number of spikes per plant) and the potential plasticity for these traits can be integrated into the screening or used as selection criteria. When SMA effects are important (third case), it is required to assess combinations of genotypes, but it is also desired to reduce the number of mixtures to evaluate. Prediction of mixture performances may then rely on both genotypic and phenotypic information (kinship, molecular markers, interaction traits).

The inclusion of genetic relatedness matrices as covariance matrices for the GMA and the SMA random effects in the mixed models may improve the prediction of the mixing effects, while allowing for predicting unobserved mixtures or mixtures involving unobserved genotypes (as for combining ability analysis and predictions of performances of hybrids, Bernardo 1995; Falconer *et al.* 1996; Lynch and Walsh 1998). Phenotypic similarities assessed on pure stand performances and traits involved in plant-plant interactions could also be used as covariance matrices. On a more general level, the identification of traits associated to mixing ability could be relevant for integrating new selection criteria in breeding for mixtures, but should also help to better understand the functioning of the mixtures (also regarding traits complementarity between genotypes or species) and to improve the design of the mixtures (Barot *et al.* 2017).

Note that this methodological framework can be easily further adapted to mixtures with varying proportions of the different components (by adjusting the weighting coefficients of GMA and SMA ($\frac{1}{K(n)}$, $\frac{1}{\binom{K(n)}{2}}$ and $\frac{1}{(K(n))^2}$, in Models 1, 2 and 3) based on their proportion of sowed seeds and neighboring probabilities), and notably to inter-specific mixtures analysis.

ACKNOWLEDGMENTS

E. Forst benefited from a thesis grant from a PICRI (Partenariat Institutions-Citoyens pour la Recherche et l'Innovation) project (2015-2017) funded by the Région Ile-de-France, and her post-doctoral salary was paid by the European Union's Horizon 2020 research and innovation programme for the LIVESEED project (https://www.liveseed.eu/), under grant agreement number 727230 and for the ReMIX project (https://www.remix-intercrops.eu/) under grant agreement number 727217. This work was also supported by the ANR WHEATAMIX project (www6.inra.fr/wheatamix/), grant ANR-13-AGRO-0008 of the French National Research Agency, for field trials. The authors are grateful to the Wheatamix consortium for early discussions on the approach and the design of the experiments, and to Didier Tropée (and the technical team IE GQE-Le Moulon) and Alain Chassin (and all the technical staff of UE 1375 PHACC) for conducting the field trials. We would like to thank Fabien Laporte for his valuable contribution to the early developments of the BLUP programing. We also thank the reviewers for their insightful comments that led us to improve the paper and to add an appendix on mixed effect model development.

Declaration of conflicts of interest: none.

REFERENCES

Alday, P., 2016 A brief introduction to mixed effects models – analogues to the coefficient of determination R² https://www.biorxiv.org/content/biorxiv/suppl/2016/07/06/062299.DC1/062299-1.pdf

Barot, S., V. Allard, A. Cantarel, J. Enjalbert, A. Gauffreteau, I. Goldringer, J-C. Lata, X. Leroux, A. Niboyet, E. Porcher, 2017 Designing mixtures of varieties for multifunctional agriculture with the help of ecology, a review. Agron. Sustain. Dev. 37 (2): 13. https://doi.org/10.1007/s13593-017-0418-x

Bates, D., M. Mächler, B. Bolker, S. Walker, 2015 Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67 (1). https://doi.org/10.18637/jss.v067.i01.

Bernardo, R. 1995 Genetic Models for Predicting Maize Single-Cross Performance in Unbalanced Yield Trial Data. Crop Science 35 (1): 141–47. https://doi.org/10.2135/cropsci1995.0011183X003500010026x.

Bernardo, R., 1996 Best linear unbiased prediction of maize single-cross performance. Crop Sci, 36 (1): 50-56. doi:10.2135/cropsci1996.0011183X003600010009x

Borg, J., L. P. Kiær, C. Lecarpentier, I. Goldringer, A. Gauffreteau, S. Saint-Jean, S. Barot, J. Enjalbert, 2018 Unfolding the potential of wheat cultivar mixtures A meta-analysis perspective and identification of knowledge gaps. Field Crop Res. 221, 298-313. https://doi.org/10.1016/j.fcr.2017.09.006

Chalbi, N., 1967 La compétition entre génotypes et ses effets sur les caractères quantitatifs de la Luzerne. Annales d’Amélioration des Plantes 17 (1): 67–82.

Cowger, C., R. Weisz., 2008 Winter wheat blends (mixtures) produce a yield advantage in North Carolina. Agron. J. 100 (1): 169–77. https://doi.org/10.2134/agronj2007.0128.

Creissen, H. E., T. H. Jorgensen, J. K. M. Brown, 2016 Increased yield stability of field-grown winter barley (*Hordeum Vulgare* L.) varietal mixtures through ecological processes. Crop Prot. 85: 1-8. https://doi.org/10.1016/j.cropro.2016.03.001

Darwinkel, A., 1978 Patterns of tillering and grain production of winter wheat at a wide range of plant densities. Neth. J. Agric. Sci., 26, 383-98.

Dawson, J. C., I. Goldringer, 2012 Breeding for genetically diverse populations: variety mixtures and evolutionary populations. Organic Crop Breeding, 77-98. https://doi.org/10.1002/9781119945932.ch5

Dubs, F., X. Leroux, V. Allard, B. Andrieu, S. Barot, et al., 2018 An experimental design to test the effect of wheat variety mixtures on biodiversity and ecosystem services, HAL. . https://hal.archives-ouvertes.fr/hal-01843564

Falconer, D. S., T. F. C. Mackay, R. Frankham, 1996 Introduction to Quantitative Genetics. 4th ed. Benjamin Cummings. Trends in Genetics, 12 (7), 280.

Federer, W. T., 1979 Statistical designs and response models for mixtures of cultivars. Agron. J. 71 (5): 701. https://doi.org/10.2134/agronj1979.00021962007100050003x.

Federer, W. T., J. C. Connigale, J. N. Rutger, A. Wijesinha, 1982 Statistical analyses of yields from uniblends and biblends of eight dry bean cultivars. Crop Sci. 22 (1): 111. https://doi.org/10.2135/cropsci1982.0011183X002200010026x.

Federer, W. T., 1999 Statistical Design and Analysis for Intercropping Experiments. Volume II. Three or More Crops. Springer, New York.

Federer, W. T., 2002 Statistical issues in intercropping. Encyclopedia of environmetrics. 1st edition, New York: Wiley, 1064-1069. https://doi.org/10.1002/9781118445112.stat07718

Finckh, M. R., C. C. Mundt, 1992 Stripe rust, yield and plant competition in wheat cultivar mixtures. Phytopathology 82 (9): 905-13. DOI: 10.1094/Phyto-82-905

Gallais, A., 1970 Modèle pour l'analyse des relations d'associations binaires. Biométrie-Praximétrie XI (2-3): 51-80.

Gallandt, E. R., S. M. Dofing, P. E. Reisenauer, E. Donaldson, 2001 Diallel analysis of cultivar mixtures in winter wheat. Crop Sci. 41 (3): 792-796. https://doi.org/10.2135/cropsci2001.413792x.

Gizlice, Z., T. E. Carter, J. W. Burton, T. H. Emigh, 1989 Partitioning of blending ability using two-way blends and component lines of soybean. Crop Sci. 29 (4): 885. https://doi.org/10.2135/cropsci1989.0011183X002900040008x.

Griffing, B., 1956 Concept of General and Specific Combining Ability in relation to diallel crossing systems. Aust. J. Biol. Sci. 9: 463–93. https://doi.org/10.1071/BI9560463

Jackson, L. F., R. W. Wennig, 1997 Use of wheat cultivar blends to improve grain yield and quality and reduce disease and lodging. Field Crops Res. 52 (3): 261-269. https://doi.org/10.1016/S0378-4290(97)00007-5.

Jensen, S. D. 1959 Combining Ability of Unselected Inbred Lines of Corn from Incomplete Diallel and Top-Cross Tests. PhD thesis, Iowa State University. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=3159&context=rtd.

Jensen, N. I., W. T. Federer 1965 Competing ability in wheat. Crop Sci 5 (5): 449–52.

Kaut, A. H. E. E., H. E. Mason, A. Navabi, J. T. O'Donovan, D. Spaner, 2009 Performance and stability of performance of spring wheat variety mixtures in organic and conventional management systems in western Canada. J. Agron. Crop Sci. 147 (2): 141–53. https://doi.org/10.1017/S0021859608008319

Kiær, L. P., I. M. Skovgaard, H. Østergård, 2009 Grain yield increase in cereal variety mixtures: a meta-analysis of field trials. Field Crops Res. 114 (3): 361-373. https://doi.org/10.1016/j.fcr.2009.09.006.

Kiær, L. P., I. M. Skovgaard, H. Østergård, 2012 Effects of inter-varietal diversity, biotic stresses and environmental productivity on grain yield of spring barley variety mixtures. Euphytica, 185(1): 123-138. https://doi.org/10.1007/s10681-012-0640-1

Knott, E. A., C. C. Mundt, 1990 Mixing ability analysis of wheat cultivar mixtures under diseased and nondiseased conditions. Theor. Appl. Genet. 80 (3): 313–20. https://doi.org/10.1007/BF00210065.

Lazzaro, M., A. Costanzo, P. Bàrberi. 2018 Single vs multiple agroecosystem services provided by common wheat cultivar mixtures: weed suppression, grain yield and quality. Field Crops Res., 221 (May): 277-97. https://doi.org/10.1016/j.fcr.2017.10.006

Lecarpentier, C., 2017 Plasticité de l'architecture aérienne du blé en réponse à la compétition pour la lumière au sein de cultures pures ou d'associations variétales : Caractérisation expérimentale et développement d'un modèle 3D. PhD thesis. Ecole doctorale Sciences du végétal, Université Paris-Saclay. http://www.sudoc.fr/219507198

Lecapentier, C., R. Barillot, E. Blanc, M. Abichou, I. Goldringer, P. Barbillon, J. Enjalbert, B. Andrieu. WALTer: A three-dimensional wheat model to study competition for light through the prediction of tillering dynamics. Annals of Botany. *in press. https://doi.org/10.1093/aob/mcy226*

Lopez, C. G., C. C. Mundt, 2000 Using mixing ability analysis from two-way cultivar mixtures to predict the performance of cultivars in complex mixtures. Field Crops Res. 68 (2): 121-32. https://doi.org/10.1016/S0378-4290(00)00114-3.

Loreau, M., A. Hector, 2001 Partitioning selection and complementarity in biodiversity experiments. Nature 412 (6842): 72-76. https://doi.org/10.1038/35083573.

Lynch, M., B. Walsh, 1998 Genetics and Analysis of Quantitative Traits. Vol. 1. M. A. Sunderland, Sinauer.

McGilchrist, C. A., 1965 Analysis of competition experiments. Biometrics, 975-985. DOI: 10.2307/2528258

Mille, B., M. Belhaj Fraj, H. Monod, C. de Vallavieille-Pope, 2006 Assessing four-way mixtures of winter wheat cultivars from the performances of their two-way and individual components. Eur J Plant Pathol. 114 (2): 163-73. https://doi.org/10.1007/s10658-005-4036-0.

Möhring, J., A.E. Melchinger, H.-P. Piepho, 2011 REML-based diallel analysis. Crop Sci 51, 470-478. doi:10.2135/cropsci2010.05.0272

Nakagawa, S. H. Schielzeth, 2013 A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol Evol, 4: 133-142. doi:10.1111/j.2041-210x.2012.00261.x

Newton, A. C., R. P. Ellis, C. A. Hackett, D. C. Guy, 1997 The Effect of component number on *Rhynchosporium secalis* infection and yield in mixtures of winter barley cultivars. Plant Pathology 46 (6): 930–38. https://doi.org/10.1046/j.1365-3059.1997.d01-83.x.

Newton, A. C., G. S. Begg, J. S. Swanston, 2009 Deployment of diversity for enhanced crop function. Annals of Applied Biology, 154(3): 309-322. https://doi.org/10.1111/j.1744-7348.2008.00303.x

Østergård, H., K., Kristensen, J. W., Jensen, 2005 Stability of variety mixtures of spring barley, pp. 28-30 in Proceedings of the COST SUSVAR/ECO-PB Workshop on Organic Plant Breeding Strategies and the Use of Molecular Markers, edited by E. T. Lammerts van Bueren, I. Goldringer and H. Østergård. Driebergen, The Netherlands. http://orgprints.org/8152/.

Piepho, H.-P., J. Möhring, 2007 Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177 (3): 1881–88. https://doi.org/10.1534/genetics.107.074229

Raghavarao, D., W. T. Federer, 2003 Sufficient conditions for balanced incomplete block designs to be minimal fractional combinatorial treatment designs, Biometrika, 90 (2): 465–470. https://doi.org/10.1093/biomet/90.2.465

Reiss, E. R., L. E. Drinkwater, 2018 Cultivar mixtures: A meta-analysis of the effect of intraspecific diversity on crop yield. Ecol Appl. 28: 62-77 https://doi.org/10.1002/eap.1629

Sprague, G. F., L. A. Tatum, 1942 General vs. Specific Combining Ability in single crosses of corn. J. Amer. Soc. Agron. 34 (10): 923–32.

Stützel, H., W. Aufhammer, 1990 The physiological causes of mixing effects in cultivar mixtures - a general hypothesis. Agric. Sys. 32 (1): 41-53. https://doi.org/10.1016/0308-521X(90)90029-P.

Thépot, S., G. Restoux, I. Goldringer, F. Hospital, D. Gouache, I. Mackay, J. Enjalbert, 2015 Efficiently tracking selection in a multiparental population: the case of earliness in wheat. Genetics 199 (2): 609–23. https://doi.org/10.1534/genetics.114.169995.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, E. Siemann, 1997 The influence of functional diversity and composition on ecosystem processes. Science 277 (5330): 1300-1302. https://doi.org/10.1126/science.277.5330.1300.

Vallavieille-Pope, C. de., 2004 Management of disease resistance diversity of cultivars of a species in single fields: controlling epidemics. Comptes Rendus Biologies 327 (7): 611-620. https://doi.org/10.1016/j.crvi.2003.11.014.

Weisser, W. W., C. Roscher, S. T. Meyer, A. Ebeling, G. Luo, et al. 2017 Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: patterns, mechanisms, and open questions. Basic Applied Ecology 23: 1-73. https://doi.org/10.1016/j.baae.2017.06.002

Wolfe, M. S., 1985 The current status and prospects of multiline cultivars and variety mixtures for disease resistance. Annu. Rev. Phytopathol. 23 (1): 251-273. https://doi.org/10.1146/annurev.py.23.090185.001343

Wu, H. X., A. C. Matheson, 2000 Analysis of half-diallel mating design with missing crosses: theory and SAS program for testing and estimating GCA and SCA fixed effects. Silvae genetica, 49 (3), 130-136.

Xu, R., 2003 Measuring explained variation in linear mixed effects models. Statist. Med., 22: 3527-3541. doi:10.1002/sim.1572

Zhao, Y., Z. Li, G. Liu, Y. Jiang, H. P. Maurer, T. Würschum, H.-P. Mock, A. Matros, E. Ebmeyer, R. Schachschneider, E. Kazman, J. Schacht, M. Gowda, C. F. H. Longin, J. C. Reif., 2015 Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc. Natl. Acad. Sci. 112 (51): 15624–29. https://doi.org/10.1073/pnas.1514547112.

Zhu, Y., H. Chen, J. Fan, Y. Wang, Y. Li, J. Chen, J.X. Fan, S. Yang, L. Hu, H. Leung, T. W. Mew, P. S. Teng, Z. Wang, C. C. Mundt, 2000 Genetic diversity and disease control in rice. Nature 406 (6797): 718-22. https://doi.org/10.1038/35021046.