

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

3

4 Authors and affiliations:

5 Emma Forst^{a,1}, Jérôme Enjalbert^a, Vincent Allard^b, Christophe Ambroise^{c,d}, Inès Krissaane^d,
6 Tristan Mary-Huard^{a,d}, Stéphane Robin^d, Isabelle Goldringer^a

7 ^a GQE- Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay,
8 91190, Gif-sur-Yvette, France

9 ^b GDEC, UMR 1095, INRA, UCA, 63000 Clermont-Ferrand, France

10 ^c LaMME, UMR 8071, CNRS, Université d'Evry Val-d'Essonne, Université Paris-Saclay,
11 91037, Evry Cedex, France

12 ^d MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

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

16

17 **HIGHLIGHTS**

- 18 • Statistical analyses for variety mixtures have made little progress in recent years
19 • Novel models are proposed to study mixing ability in incomplete designs
20 • The models account for inter and intra-genotypic interactions within mixtures
21 • The framework handles mixtures with any order and proportions of components
22 • 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

23 **ABSTRACT**

24 There has been recently a renewed interest for variety mixtures due to their potential capacity
25 to stabilize production through buffering abiotic and biotic stresses. Part of this results from
26 complementarity and/or compensation between varieties which can be assessed under mixed
27 stands only. Mixing ability of varieties can be partitioned into *General* and *Specific Mixing*
28 *Abilities* (GMA and SMA) that have been estimated so far through the evaluation of binary
29 mixtures in complete diallel designs. However, the number of mixtures increases exponentially
30 with the number of studied varieties, and the only feasible devices are incomplete designs.
31 Despite the long history of statistical analysis of variety mixtures, such incomplete design
32 analysis have rarely been addressed so far. To fill the gap, we proposed a generalized
33 statistical framework to assess mixing abilities based on mixed models and BLUP method, with
34 an original modeling of plant-plant interactions. The approach has been applied to a panel of
35 25 winter wheat genotypes observed in two contrasted experimental designs: (i) an incomplete
36 diallel of 75 binary mixtures, and (ii) a trial including higher order mixtures (four and eight
37 components). The use of mixing ability models improved prediction accuracy (of modeled
38 values for observed traits) in comparison to predictions from the mean of the pure stand
39 components, especially in the first experiment. Genetic variability was detected for the GMA of
40 yield and its components, whereas variability for SMA was lower. GMA predictions based on
41 the diallel trial were highly correlated with the GMA of the second trial providing accurate inter-
42 trial predictions. A new model has been proposed to jointly account for inter and intra-genotypic
43 interactions for specific mixing ability, thus contributing to a better understanding of mixture
44 functioning. This framework constitutes a step forward to the screening for mixing ability, and
45 could be further integrated into breeding programs for the development of intra- or inter-specific
46 crop mixtures.

47

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

49

50 1. INTRODUCTION

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

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

75 Mixing ability relies on an analogy with the concept of *combining ability* developed for hybrid
76 breeding (Sprague and Tatum 1942). By assessing a set of p genotypes in all possible hybrid

77 combinations, the combining ability can be estimated and partitioned between the general
 78 combining ability of the genotypes and the specific ability of each particular combination
 79 (Griffing 1956). The analogy has first been proposed for the study of plant interactions between
 80 rows of genotypes (Jensen and Federer 1965) and has then been broadly applied to the study
 81 of binary mixtures for estimating mixing ability, both when the performance of each component
 82 is accessible in the mixture (McGilchrist 1965; Chalbi 1967; Gallais 1970; Federer 1979;
 83 Federer *et al.* 1982), or when only the mixture performance is surveyed (Federer 1979; Federer
 84 *et al.* 1982; Gizlice *et al.* 1989; Knott and Mundt 1990; Gallandt *et al.* 2001). The work
 85 developed hereafter deals with this second case, where only the global performance of each
 86 mixture is available. The **general mixing ability (GMA)** refers to the average performance of a
 87 variety in mixture (additive term) and the **specific mixing ability (SMA)** relates to the interaction
 88 between the two components of a particular combination of varieties. According to Griffing's
 89 decomposition of combining ability in hybrids (1956), the performance of binary mixtures can
 90 be modeled as:

$$91 \quad Y_{ijbr} = \mu + \alpha_b + \frac{1}{2}(GMA_i + GMA_j) + SMA_{ij} + \varepsilon_{ijbr},$$

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

98 However, due to the need to grow all the possible binary mixtures (that is to use a complete
 99 design) among a set of varieties to estimate GMA and SMA, the field evaluation can be
 100 cumbersome regarding the cost and management of experimental trials. For this reason,
 101 authors have usually applied this method to a limited number of varieties (e.g. six or eight in
 102 Gallandt *et al.* 2001; Gizlice *et al.* 1989 respectively). A key constraint for experimental efforts,

103 this curse of dimensionality has also been encountered when studying hybrids' combining
104 ability, and solved by the use of incomplete designs (Jensen 1959). By reducing the number of
105 mixtures to grow for each variety, such designs allow a wider range of varieties to be evaluated.
106 Nonetheless, due to the reduced number of observed mixtures using the Griffing's model
107 (1956), the use of incomplete design raises the question of the definition of GMA and SMA. If
108 GMA and SMA are defined assuming a complete design setting, then some GMA and SMA
109 cannot be estimated in practice if the actual design is incomplete. Otherwise, if GMA and SMA
110 are defined conditionally to the set of observed mixtures only, then the definition of GMA and
111 SMA is specific to this set of mixtures.

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

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

131

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

142

143

144 **2. MATERIAL AND METHODS**

145 **2.1 Modeling of mixing ability:**

146 *2.1.1 Griffing's model with fixed GMA and SMA effects*

147 Mixing ability can be modeled using the same formalism as proposed by Griffing for combining
148 ability (1956). In this setting, the GMA and SMA are defined as fixed effects and the residual
149 term as random. This model is associated to the particular context of complete diallel
150 experiment design i.e. (i) all the possible mixtures are assessed and (ii) only binary mixtures
151 are considered, in 1:1 proportions. The use of this model in a broader context raises some
152 statistical issues: in case of incomplete design the condition (i) is not fulfilled, so the definition
153 of the terms becomes unclear, the estimability of the GMA and SMA is not warranted and the
154 estimators initially proposed by Griffing are no longer valid; if requirement (ii) is not met, a
155 model for SMA in higher order mixture should be proposed. To address the problem of the
156 definition of GMA and SMA resulting from assumption (i) we propose to treat the genetic effects

157 as random effects in a mixed model (as proposed, but not developed, by Griffing 1956, and
 158 further developed by Möhring *et al.* 2011 for the prediction of combining ability in hybrids).

159

160 2.1.2 Models with random GMA and SMA effects

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

$$165 \quad Y_{nbr} = \mu + \alpha_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + e_{nbr} \quad (\text{Model 1})$$

166 where Y_{nbr} stands for the performance of the r -th replicate of mixture (or genotype) n in block
 167 b , μ is the mean of the experiment, α_b is the effect of block b , $K(n)$ is the number of
 168 components of mixture n and $k(n)$ stands for the k -th genotype within mixture n (or
 169 $K(n)=1$ and $k(n)=1$ if n is a pure stand), $GMA_{k(n)}$ is the general mixing ability of the
 170 genotype, and e_{nbr} is the residual term.

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

$$175 \quad Y_{nbr} = \mu + \alpha_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{\binom{K(n)}{2}} \sum_{k(n)=1}^{K(n)-1} \sum_{k'(n)=k(n)+1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \quad (\text{Model 2})$$

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

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

$$185 \quad Y_{nbr} = \mu + \alpha_b + GMA_{l(n)} + SMA_{l(n)l(n)} + e_{nbr},$$

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

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

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

199

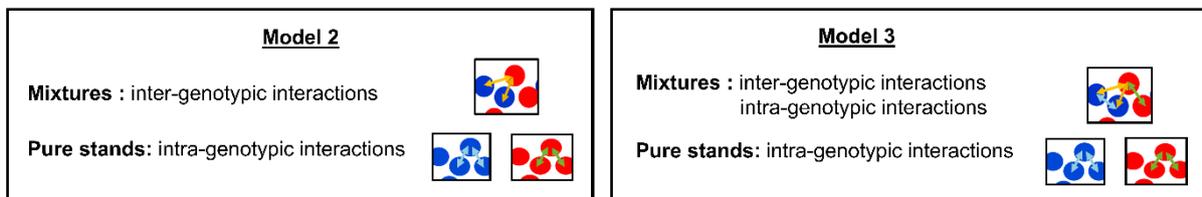
200 *2.1.3 Intra-genotypic interactions in mixture*

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

204
$$Y_{nbr} = \mu + \alpha_b + \frac{1}{K(n)} \sum_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{(K(n))^2} \sum_{k(n)=1}^{K(n)} \sum_{k'(n)=1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \quad (\text{Model 3})$$

205 where α_b is the effect of block b . In this model, intra-genotypic interaction ($k'(n) = k(n)$) is
 206 therefore always part of the mixture performance, for any mixture order ($K(n) \geq 1$). For
 207 instance, for a given binary mixture n of genotypes i and j , the interaction term between the
 208 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
 209 that for $K(n) = 1$, Model 2 and Model 3 coincide. The coefficients $\frac{1}{K(n)^2}$ represent the
 210 expected weight of each kind of neighborhood in the plant community, assuming a random
 211 distribution of genotypes in the plant community. It should be noticed that due to the differences
 212 in SMA effects weighting, the SMA variance and SMA BLUPs are expected to be higher with
 213 Model 3 than with Model 2. Nevertheless, the integration of neighboring probabilities in the
 214 modeling of mixing ability (especially through the introduction of intra-genotypic interactions
 215 within mixture) constitutes a novel advance in mixture analysis.

216



217

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

219

220 The motivation for considering both models is that Model 2 only accounts inter-genotypic
 221 interactions within mixtures, while Model 3 also accounts for intra-genotypic interactions in
 222 mixtures (Figure 1). For convenience we adopted notations for all models similar to the
 223 Griffing's notations for combining ability. We emphasize that the interpretation of the different
 224 terms is quite different between Model 2 and Model 3. To illustrate these differences, we
 225 consider the expected performance over all possible equally weighted mixtures of order K that

226 include a given genotype k . This amounts to work conditionally on all random effects that
227 depend on k only, so we denote this quantity by $E(Y | k, K)$.

228 One has

$$229 \quad E(Y | k, K) = \mu + \frac{GMA_k}{K} \quad \text{if } K > 1 \quad \text{in Model 2}$$

$$230 \quad E(Y | k, K) = \mu + GMA_k + SMA_{kk} \quad \text{if } K = 1 \quad \text{in Model 2}$$

$$231 \quad E(Y | k, K) = \mu + \frac{GMA_k}{K} + \frac{SMA_{kk}}{K^2} \quad \text{in Model 3}$$

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

237

238 **2.2 Statistical analysis:**

239 *Inference method*

240 The mixed model framework has already been widely applied to plant breeding (Bernardo
241 1996; Falconer *et al.* 1996; Lynch and Walsh 1998; Piepho and Möhring 2007), but to our
242 knowledge not to crop mixing ability analysis. This framework allows studying mixtures of any
243 order in incomplete designs, ensuring model estimability (as discussed above).

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

$$245 \quad y = X\beta + Z_1g + Z_2s + e$$

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

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

$$252 \text{Var}(y) = Z_1 A_1 Z_1' \sigma_{GMA}^2 + Z_2 A_2 Z_2' \sigma_{SMA}^2 + I \sigma_e^2.$$

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

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

260

261 **2.3 Experimental designs:**

262 The analysis has been carried on two data sets produced by two experiments assessing mixing
263 ability of a panel of 25 genotypes of winter wheat. This panel was designed to be representative
264 of the phenotypic diversity of a larger set of 58 genotypes, evaluated in pure stand in a previous
265 study for root and shoot architecture as well as other functional traits (Cantarel *et al.* in prep.,
266 Dubs *et al.* 2018). This 25 genotypes panel is composed of nine elite varieties cultivated for
267 their high yielding performance in the Parisian Basin (Altigo, Apache, Arezzo, Boregar, Grapeli,
268 Renan, Sogood, Soissons, Trémie), four landraces (Blé autrichien, Rouge de Bordeaux,
269 Rouge du Roc, Saint Priest), six varieties bred for organic agriculture (Alauda, Hermes, Maxi,
270 Midas, Ritter, Skerzzo), and six lines from a MAGIC population (A22, A208, A243, A398, F236,

271 F426, Thépot *et al.* 2015) (Table B.1). The way mixtures of two, four and eight genotypes of
272 the panel were designed in the two following experiments has been inspired from the ecology
273 literature, and more specifically the Jena diversity experiment (Weisser *et al.* 2017), exploring
274 extensively the gradient of intraspecific functional diversity (Dubs *et al.* 2018).

275 In the first experiment, the 25 genotypes were grown in a field trial both in pure stands and in
276 75 binary mixtures during the season 2014-2015 in the Le Moulon (LM) experimental station
277 (48°42'37.2"N, 2°09'37.6"E) in France, according to an incomplete diallel design (75 of the 300
278 binary combinations: Figure 2.A). Each entry was sown in a 7,5m² plot at a density of 160
279 grains.m⁻² (with 20 cm row width). All pure stand genotypes and binary mixtures were replicated
280 twice in two randomized complete blocks. The mixtures were prepared for sowing using the
281 thousand kernel weight in order to get equal proportion of seeds of each genotype.

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

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

313

314 *Phenotypic data*

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

318 Yield was calculated as the weight of grain harvested on the plot surface, and after measuring
319 the humidity of a sample, it was standardized into q.ha⁻¹ at 15% humidity (1quintal / hectare =
320 100kg / 10.000m² = 10⁻² kg.m⁻²). The number of spikes has been counted after flowering date,
321 on one meter length for two adjacent rows in each plot, and converted into a number of
322 spikes.m⁻² taking into account the distance between rows (20 cm). TKW was measured after
323 harvest and threshing, and the number of grains/spike was calculated based on the yield

324 components (= (yield/ TKW)/ number of spikes.m⁻²). Protein content was estimated on grain
325 samples using near-infrared spectroscopy (NIRS).

326

327 **2.4 Data analysis:**

328 *2.4.1 Application of the models to the datasets of each location*

329 For each location, the correlation between the observed mixtures and the mean of their pure
330 stand components was calculated. The main analyses (model comparison, analysis of the
331 variance components and the BLUPs) were then performed with the three models on all the
332 observations (mixtures and pure stands) of each location separately. The predicted values
333 were obtained from the different models, fitted on the observed data. The correlation between
334 all the observed values and the predictions from (i) the means of the genotypic effects of the
335 corresponding pure stand components (according to Federer *et al.* 1982 and Gizlice *et al.*
336 1989), or (ii) a subset of observations of the same location (mixtures only using Model 3) has
337 been performed for each site. Finally, the CF observed values for different mixture orders were
338 compared with predictions obtained from CF or LM trials analyzed with Model 1. The analysis
339 was performed using an adaptation of the *lme4* R package (Bates *et al.* 2015; see File S.1).

340

341 *2.4.2 Model comparison*

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

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

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

350 The models were compared based on the AIC, BIC and on Likelihood Ratio Tests (LRT, using
351 Maximum Likelihood procedure (ML)) which were further performed for nested models (Model
352 2 vs Model 1, and Model 3 vs Model 1). We also computed the second-order AIC and the
353 conditional AIC using MuMIn and cAIC4 R packages respectively, but no difference with the
354 AIC values was observed. In addition, the R^2 and the RMSE were reported for the different
355 models. While open to criticism (Alday 2016), the calculation of a R^2 for mixed models is a
356 convenient tool for assessing the model quality using a common dimensionless metric for the
357 different response variables. The R^2 were calculated following Xu (2003) to assess the
358 proportion of the total variance which is explained by the model (conditional variance, for both
359 fixed and random effects), as one minus the ratio of the residual variance of the full mixed
360 model, over the residual variance of a null model (the fixed intercept-only model).

361

362 *2.4.3 Within-location comparison of observed values with predictions from BLUP based on* 363 *subsets of observations*

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

369

370 *2.4.4 Inter-trial predictions*

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

373 experiment.

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

385

386 **2.5 Data and program availability:**

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

390

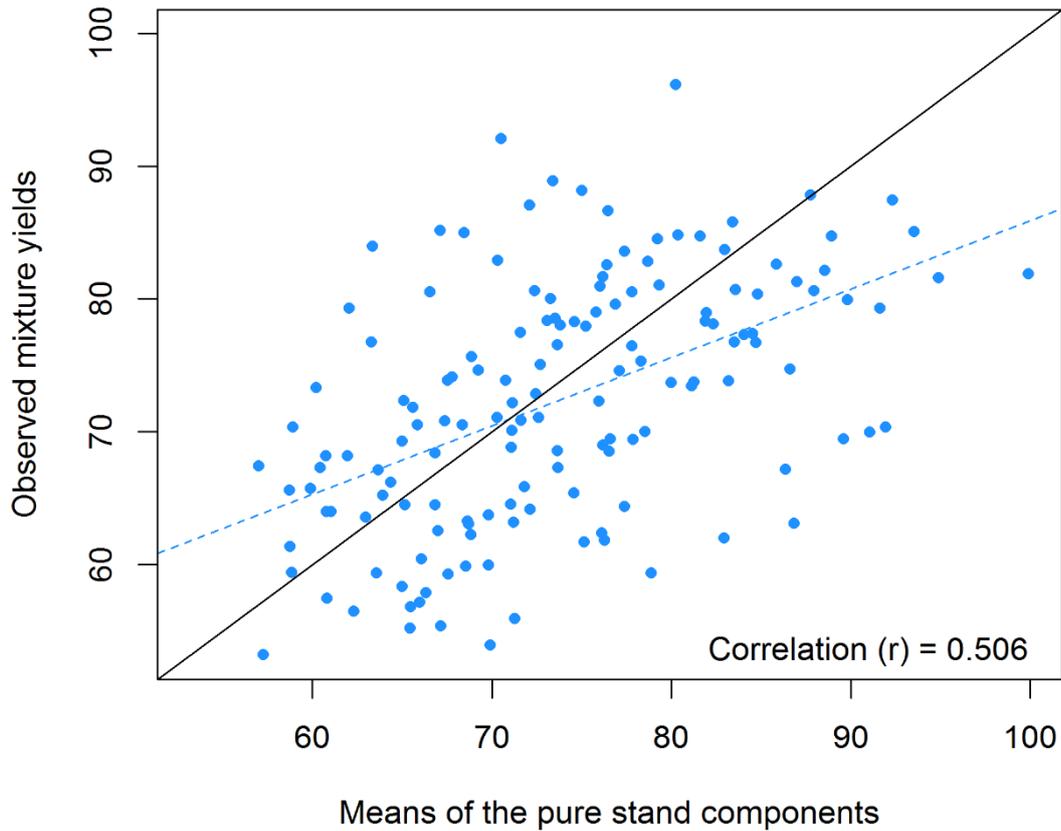
391

392 **3. RESULTS**

393 *3.1 Correlation between mixtures and pure stands*

394 The Pearson's correlation between the observed binary mixtures performances and the means
395 of their pure stand components was high for TKW (0.91) but moderate for yield (0.51, Figure
396 3), for the number of spikes.m⁻² (0.51), and for the number of grains per spike (0.42) in the Le
397 Moulon (LM) experiment, showing the need for modeling of mixing ability. The correlation in

398 Clermont-Ferrand (CF) reached 0.75 for yield and 0.63 for the protein content (respectively
399 0.81 and 0.73 when considering only binary mixtures).



400

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

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

405

406 *3.2 Model comparison*

407 Based on LRT (Likelihood Ratio Test) using ML procedure for model comparison, both models
408 including SMA have been found significantly better than Model 1 for TKW (p-values of 0.026
409 and 0.030 respectively, Table 1). Models 2 and 3 had low but non-significant p-values (0.060

410 and 0.182) for spike density, while for yield and other response variables in LM and CF trials
 411 the p-values were non-significant.

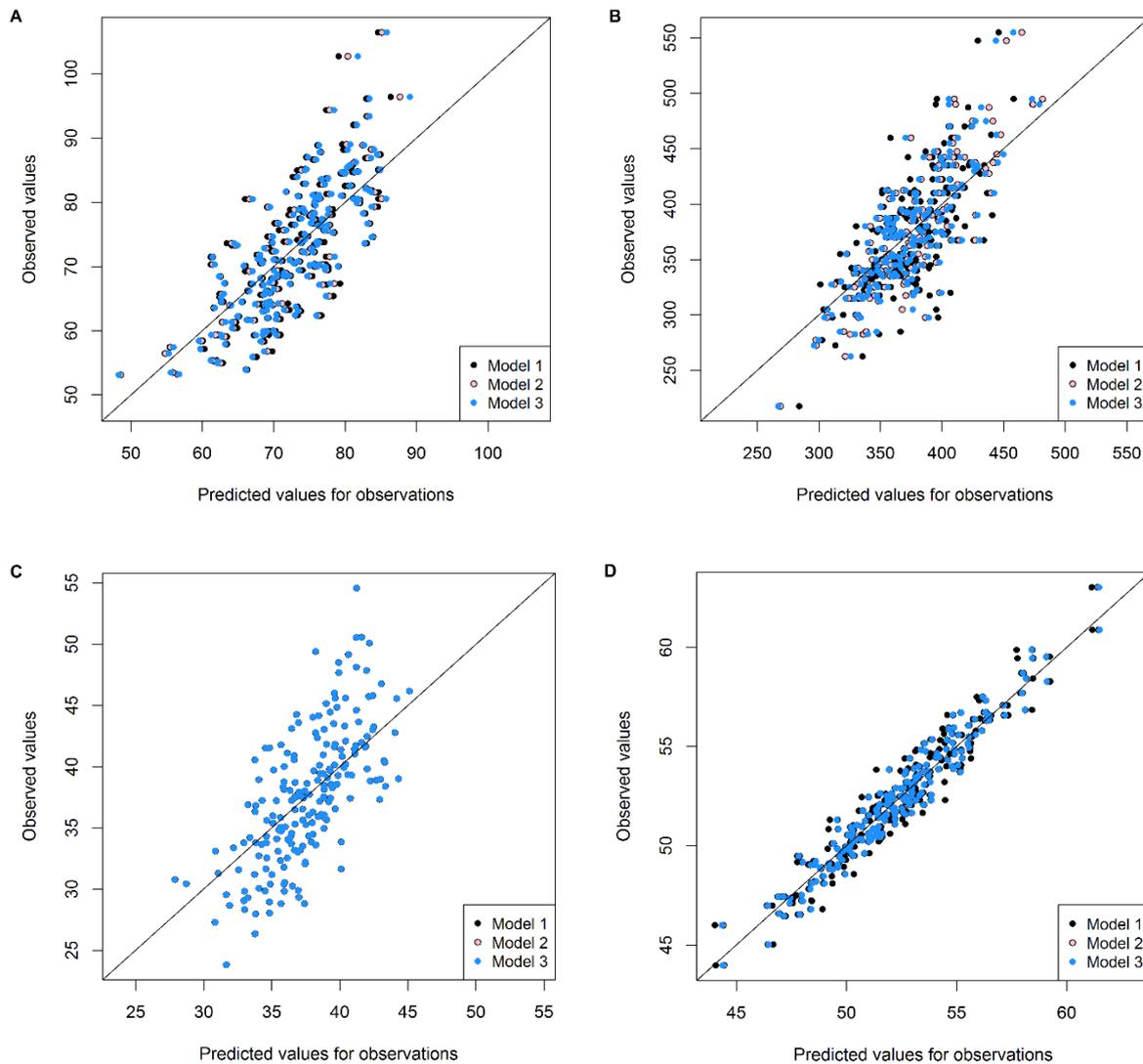
Response variable	Model	AIC	BIC	p-value
Yield	Model 1	1393.3	1406.5	NA
	Model 2	1395.1	1411.6	0.664
	Model 3	1394.6	1411.0	0.385
Spike density	Model 1	2075.5	2088.7	NA
	Model 2	2074.0	2090.4	0.060 (.)
	Model 3	2075.7	2092.2	0.182
Grain number per spike	Model 1	1188.4	1201.6	NA
	Model 2	1190.4	1206.9	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 (*)

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

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

415

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



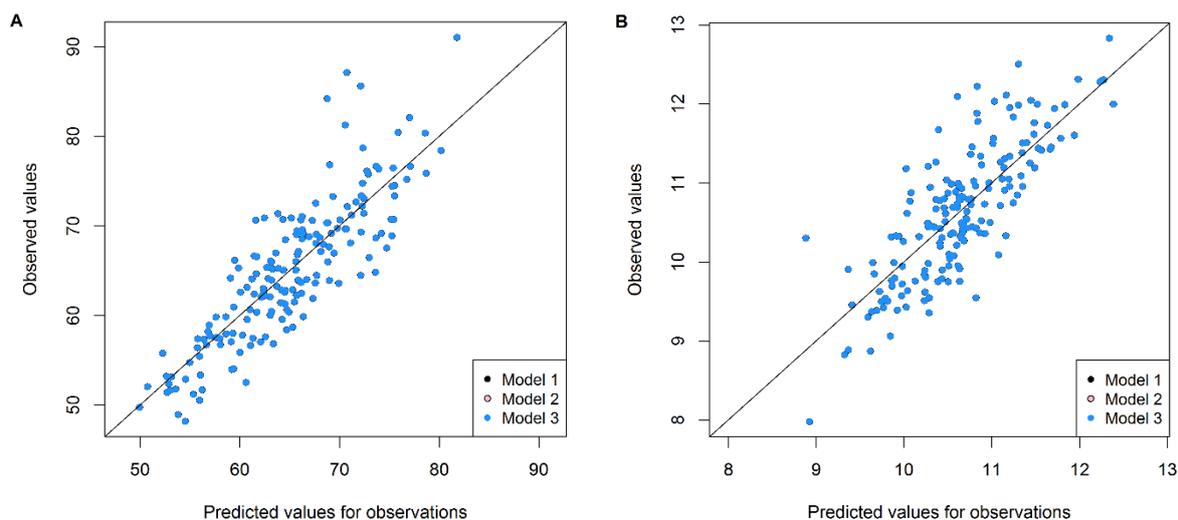
419

420

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

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

426



427

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

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

431

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

438

Response variable	Model	R^2	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

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

440

441 *3.3 Comparison with predictions based on the pure stands*

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

454

455 *3.4 Variances of GMA and SMA*

456 The variance components estimated with each model for yield and for the number of
457 spikes.m⁻² are detailed below for the LM and CF experiments (Table 3). The other responses
458 are given in supplementary material (Table B.2). It should be noticed that, as expected,

459 weighting of SMA effects was higher for Model 3 than for Model 2 (higher SMA variance and
 460 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

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

462

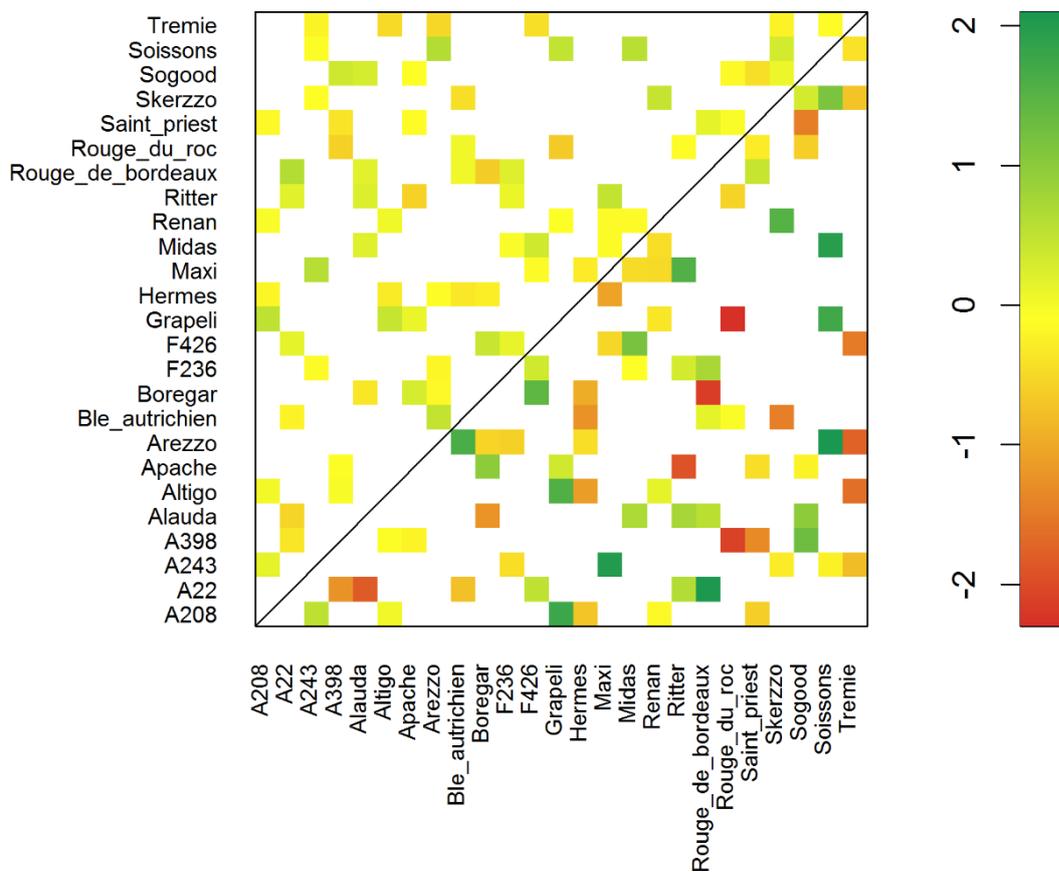
463 In all experiments and with all considered mixed models, the GMA variance was always higher
 464 than the residual variance (10 times higher in the case of the TKW in LM experiment: Table
 465 B.2). The amount of the SMA variance compared to the GMA variance ranged from 0 to 0.19
 466 with Model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

470

471 *3.5 GMA and SMA predictions*

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

490 de Bordeaux, Figure 6) showed a rather high SMA value when grown together in a mixture
 491 (Figure 7). More extreme values were obtained with Model 3, due to higher SMA variance (as
 492 expected). This was also the case for spike density and TKW in LM trial although to a lesser
 493 extent (Figure B.2). The inter-genotypic SMA for Model 2 and Model 3 were very highly
 494 correlated for the yield, spike density and TKW (0.9993, 0.9996 and 0.9998 respectively).
 495



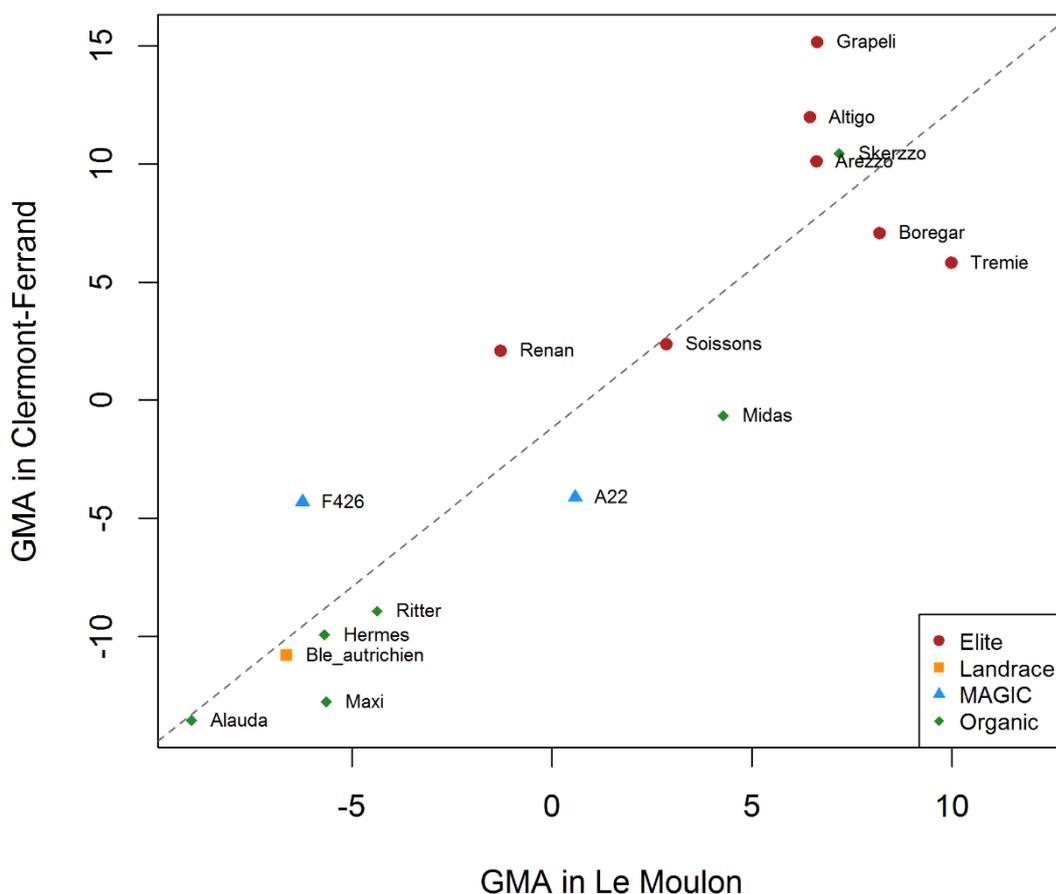
496
 497 **Figure 7:** SMA predicted values for yield with Models 2 and 3 in LM trial

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

500
 501 The GMA was highly correlated with the pure stand genotypic effect in the CF trial (0.98 for the
 502 yield with Model 1, Table B.5), and to a lesser extent in the LM experiment (0.88). The

503 correlation between the GMA obtained in both experiments for yield using Model 1 was also
 504 quite high (0.903). The CF trial displayed higher variance for the GMA values than did the LM
 505 trial (Figure 8).

506



507

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

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

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

515

516 *3.6 Predictions based on subsets of observations*

517 In the LM experiment, when the analysis was performed on mixture observations only (m)
 518 instead of all observations ($m+p$), the variance of SMA for yield (estimated with both Models 2
 519 and 3) became null. The correlation between the observed and predicted mixtures with Model
 520 3 was 0.786 for yield when using GMA_m , not different from the one obtained with $m+p$
 521 observations (0.790), and the RMSE was reduced from 6.3q.ha⁻¹ to 5.8q.ha⁻¹ (Table B.4). This
 522 result suggests that mixing ability analysis might be performed without requiring pure stands in
 523 the dataset. The GMA for yield were modified (Figure B.3) although the correlation between
 524 GMA_m and GMA_{m+p} with Model 3 was 0.91 (Table B.5). In the CF experiment, the correlation
 525 between GMA_m and GMA_{m+p} (with Model 1) was even higher (0.96, Table B.5), and the
 526 correlation between observed and predicted values was equal when excluding the pure stands
 527 for GMA predictions (0.834) to the one when pure stands were included (0.854, Table B.4).

528

529 *3.7 Predictions of CF high order mixtures based on CF BLUP*

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

534

Mixture order	Pearson's correlation						Kendall's correlation					
	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

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

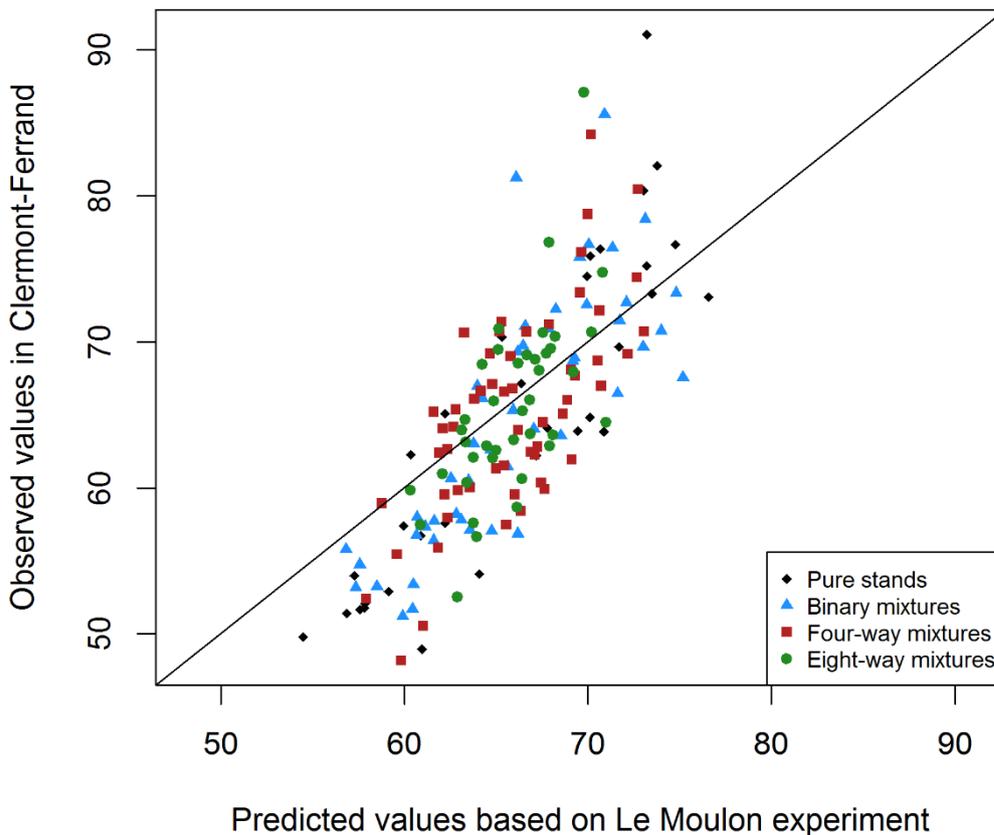
538 The order of the mixture indicates the number of components: one for pure stand, two for binary mixture,
539 four for four-way mixture and eight for eight-way mixture. The capital letters indicate the trial for
540 observations for genotypic effects in pure stand or GMA predictions using Model 1.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

543

544 *3.8 Inter-trial predictions*

545 In order to assess the ability of the BLUP values obtained in a design based on binary mixtures
546 to predict the observed values for higher order mixtures, we computed the predicted values for
547 the CF observations based on the LM GMA using Model 1 (Figure 9). The Pearson's correlation
548 between the observed values in CF and predictions based on LM was 0.785 (and 0.598 for
549 Kendall's correlation, Table 4). The correlation was slightly lower when excluding the pure
550 stand observations in LM of the analysis (0.741). However, the correlation was always much
551 higher than when using the mean of the corresponding genotypic effects of pure stand
552 components (correlation of 0.653). When including SMA in addition to GMA predicted on all
553 observations in the LM trial (using Models 2 and 3) for predicting yields of binary mixtures in
554 the CF trial, the correlation between observed and predicted yields was not clearly improved
555 (0.814 with Model 2 and 0.817 with Model 3, vs 0.805 with Model 1).

556



557

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

560 The black diagonal is the $y=x$ line.

561

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

568

569

570 4. DISCUSSION

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

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

580

581 *4.1 General Mixing Ability predominates in two experiments on wheat mixtures*

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

591 The part of variance explained by GMA and SMA effects was relatively high, given that the
592 proportion of observed mixtures on the total number of possible mixtures was rather low
593 ($75/300=0.25$). This suggests that using a random sample of 25% of the total number of all
594 possible mixtures, in addition to the pure stands, seems efficient to assess GMA and SMA of
595 the genotypes used as components, and this is in accordance with previous findings for

596 combining ability in wheat (Zhao *et al.* 2015). This result calls for the possibility to screen a
597 large number of genotypes from a diversified panel for their GMA using incomplete designs
598 instead of growing all pairwise mixtures within a limited set of genotypes. In the second
599 experiment, the correlation between the mixture performances and the means of the pure
600 stands components was rather high for yield (0.75) and for protein content (0.63) so the
601 predictions of all the observations based on the pure stand performances were efficient and
602 were only marginally improved when using mixing ability modeling. This is explained by a very
603 high correlation between the GMA and the genotypic effects of the pure stands for yield in this
604 trial (0.98), for reasons discussed hereafter (section 4.3).

605

606 *4.2 An original modeling of specific mixing ability*

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

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

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

616 These two models resulted in differences in variance components and in ratio of variance
617 between SMA and GMA, with a higher SMA variance observed with Model 3, as expected. The
618 intra-genotypic SMA of yield had noticeably more extreme values in the SMA distribution when
619 using Model 3 compared to Model 2, suggesting that the intra-genotypic interactions in
620 mixtures significantly accounted for SMA variance in Model 3 for this trait. The interpretation of
621 the intra-genotypic SMA should be different depending on the model used. Model 3 might be

622 more realistic regarding the biological mechanisms occurring within mixtures, as it integrates
623 intra-genotypic interactions within mixtures, and dilutes their weight in high order mixtures.
624 Presently, a unique variance component is estimated in Models 2 and 3 for SMA_{ii} and SMA_{ij} .
625 However, intra and inter-genotypic interactions might have different distributions, as might
626 occur for yield in LM trial (Figure 6). Therefore, a refinement might be to estimate separately
627 two variance components: one for SMA_{ii} and one for SMA_{ij} (with $i \neq j$) for both Models 2 and
628 3. This would require the inference of more variance parameters - each with less available
629 information - which, in turn, would require the evaluation of more mixtures per genotype than
630 available in this study.

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

650 with its GMA in the experiment. Both approaches are strongly divergent in their philosophy: in
651 the GPA/TGMA approach, pure stands and mixtures performances are used independently to
652 estimate GPA and GMA, and the pure stand performances are considered as a reference to
653 compute the TGMA. In the present model, pure stands and mixtures are jointly used to estimate
654 GMA and SMA, and the intra-genotypic interaction (SMA_{ii}) has the same status as the inter-
655 genotypic interaction (SMA_{ij}), which seems biologically sound.

656

657 *4.3 Variance components vary between the two studied locations*

658 For all response variables in the binary mixtures trial (LM), the variance of GMA was higher
659 than the one of SMA, in accordance with Knott and Mundt (1990) in similar trials, while no SMA
660 effect at all was found in the experiment involving higher order mixtures, as in Lopez and Mundt
661 (2000). However, only GMA/SMA estimates for yield can be compared among the two
662 experiments since the other response variables differed among LM and CF. The absence of
663 SMA in the CF experiment might be due to three possible factors which are confounded in this
664 experimental design, and might contribute to limit the inter-genotypic interactions in CF. Firstly,
665 the CF trial was conducted under more favorable conditions (higher yield objective and nitrogen
666 inputs) that might have led to less stresses for the plants and therefore less opportunity for
667 complementarity or synergy mechanisms to express among mixtures' components. Secondly,
668 the panel used in the CF experiment was a subset of the panel used for the Le Moulon
669 experiment, the genotypes were less contrasted than in the LM trial (only one landrace included
670 in the panel, lower diversity in functional traits) resulting in lower competition/synergies. Thirdly,
671 the CF experiment involved higher order mixtures in which SMA within each pair of genotypes
672 might have been too low to be detected (increasing the number of binary interactions results
673 in a dilution of their effects), or the possible occurrence of higher order interactions might have
674 masked the binary ones. Thus, this could lead to reducing the possibility to observe SMA
675 effects in the CF trial compared to the LM trial.

676 For yield in the LM trial, the share of the genetic variance due to SMA effect was larger with
677 Model 3 than with Model 2, indicating that the SMA variance might also be due to intra-
678 genotypic interactions within mixtures. However, both models including SMA did not provide a
679 significantly better fit than the model with GMA only (and the differences in AIC values for the
680 three models were very small for all response variables (Table 1)), as already found by Gizlice
681 *et al.* (1989). In contrast, Gallandt *et al.* (2001) in wheat or Federer *et al.* (1982) in bean found
682 significant SMA. Spike density displayed a SMA variance ratio of 0.61 with Model 3, indicating
683 strong interactions between plants for this response variable. This is consistent with the fact
684 that the number of tillers and therefore the number of spikes are known to be the most plastic
685 traits when changing plant density (Darwinkel 1978) and/or the phenotypes of neighboring
686 plants in mixtures (Jackson and Wennig 1997; Cowger and Weisz 2008, Lecarpentier 2017,
687 Lecarpentier *et al.* 2019). This is also in line with the clear-cut difference observed in the range
688 of spike density in the LM trial when comparing pure stands and binary mixtures (respectively
689 217.5 to 490 for pure stands and 262.5 to 555 for mixtures). These specific effects observed
690 on mixtures occurred in the experiment characterized by a low density leading to plasticity for
691 tillering, and selection effects between genotypes due to differences in potential productivity.
692 As this leads to differences in genotype frequencies in the harvested grain, it might be
693 interesting to measure the proportion of grains (or spikes) of each genotype in mixtures to
694 further investigate the specific effects and better predict the mixture performances based on
695 individual contributions of the components. The estimation of selection effects may help to
696 better understand the changes in proportions of the genotypes over time (Finckh and Mundt
697 1992; Loreau and Hector 2001; Barot *et al.* 2017) but this requires particular designs or
698 systems for separating the mixture components.

699

700 *4.4 BLUP*

701 In the LM experiment, the Trémie genotype showed the highest BLUP value for both GMA and
702 SMA_{ii} due to its high productivity in particular in pure stand. However, the binary mixtures

703 involving this genotype had among the lowest SMAs and all had lower yields than expected
704 based on the pure stands components (i.e. negative OY), therefore indicating that Trémie, as
705 an elite variety selected for performing well in pure stand conditions, might not be the best
706 “mixing partner” to combine. On the contrary, the genotype Soissons had intermediate GMA
707 (in both trials) and a low SMA in pure stand, while its SMAs in mixtures were high and the
708 mixtures displayed high yields (Soissons also had the highest GMA when considering mixtures
709 only). This could be due to a favorable mixture design for this genotype, but the OY was always
710 positive for this genotype (except when mixed with Trémie) with a mean of $+7.7\text{q}\cdot\text{ha}^{-1}$ showing
711 its potential for mixing conditions. The important mixing ability for yield of this genotype might
712 be explained by a high mixing ability for spike density. These results show the importance of
713 taking both GMA and SMA into account for selecting candidates for mixtures.

714

715 *4.5 Prediction of mixtures and pure stands performances based on BLUP values*

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

728

729 *4.6 Predictions of higher order mixtures from binary mixtures evaluated in another experiment*

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

749

750 *4.7 Improving screening and breeding strategies*

751 The two mixture designs used in this study were performed empirically, as in all agronomical
752 and ecological experiments we have reviewed. This raises the question of how to optimize the
753 experimental designs used to better estimate GMA/SMA. As far as we know, this question has
754 been poorly addressed in the literature. For example Federer (2002) and Raghavarao and
755 Federer (2003) have proposed balanced incomplete designs to achieve estimability of the

756 GMA and SMA for a specific number of varieties and three-way mixtures, without giving rules
757 on the way varieties should be mixed. The experimental design optimization remains to be
758 studied. The BLUP framework proposed here will allow such optimization, exploring various
759 GMA/SMA ratios, according to panel/species biological characteristics.

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

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

782 between the mixture performances and the mean of their pure stand components, ii) important
783 GMA versus SMA, iii) important SMA.

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

793 ii) When the GMA is the major source of variation, compared to the SMA, it could be wiser to
794 focus on the components *per se* for mixing rather than on combinations of components. The
795 experimental design should be based on genotype screening under conditions that allow
796 interactions between different genotypes such as alternate rows (as proposed by Barot *et al.*
797 2017) or mixtures with a tester (that can be a genotype with high GMA or eventually a mixture
798 itself to save space and if the further purpose is the development of higher order mixture) for
799 instance. If the SMA is low but not negligible, it might be interesting to search for panel structure
800 to optimize the experimental design.

801 iii) When SMA effects are important, the existence of complementarity groups for mixing ability
802 should be investigated. As carried out for hybrid development, the use of clustering methods
803 to search for heterotic-like pattern might allow to capture some of the specific effects in the
804 GMA (increasing the accuracy of mixture performance predictions), and to further improve the
805 complementarity between components within mixtures structuring the panel into groups (Zhao
806 *et al.* 2015). For the development of higher order mixtures, the presence of higher order specific
807 effects might be investigated (Federer 1999) to decide if the assessment of binary mixture is
808 sufficient or if higher order mixtures should be included in the design. If each genotype is

809 observed in several mixtures, the stability of the SMA effects involving the genotype might be
810 assessed, allowing to identify the most stable genotypes for specific effects; and stability of
811 SMA over environments might be investigated.

812 The mixing ability structuration between GMA and SMA has consequences on the conception
813 of the breeding strategy. When the mixture performances are highly predictable based on pure
814 stands (first case), no particular adaptation of the breeding scheme is required for selecting the
815 best performing genotypes, other than assuring that the final mixtures are agronomically
816 coherent. When the selection focuses on GMA (second case), genotypes should be assessed
817 and selected in interaction with each other or with testers. Interaction traits (e.g. the number of
818 spikes per plant) and the potential plasticity for these traits can be integrated into the screening
819 or used as selection criteria. When SMA effects are important (third case), it is required to
820 assess combinations of genotypes, but it is also desired to reduce the number of mixtures to
821 evaluate. Prediction of mixture performances may then rely on both genotypic and phenotypic
822 information (kinship, molecular markers, interaction traits).

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

833 Note that this methodological framework can be easily further adapted to mixtures with varying
834 proportions of the different components (by adjusting the weighting coefficients of GMA and

835 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 sown
836 seeds and neighboring probabilities), and notably to inter-specific mixtures analysis.

837

838

839 Declaration of competing interest: none.

840

841 **ACKNOWLEDGMENTS**

842 This work benefited from the interdisciplinary work carried out by the Wheatamix consortium
843 members, for early discussions on the approach and the design of the experiments. The
844 authors would like to thank Didier Tropée (and the technical team IE GQE-Le Moulon) and
845 Alain Chassin (and all the technical staff of UE 1375 PHACC) for conducting the field trials.
846 The authors are also grateful to Fabien Laporte for his valuable contribution to the early
847 developments of the BLUP programming. We also thank the reviewers for their insightful
848 comments that led us to improve the paper and to add an appendix on mixed effect model
849 development.

850 E. Forst benefited from a thesis grant from a PICRI (Partenariat Institutions-Citoyens pour la
851 Recherche et l'Innovation) project (2015-2017) funded by the Région Ile-de-France, and her
852 post-doctoral salary was paid by the European Union's Horizon 2020 research and innovation
853 program for the LIVESEED project (<https://www.liveseed.eu/>), under grant agreement number
854 727230 and for the ReMIX project (<https://www.remix-intercrops.eu/>) under grant agreement
855 number 727217. This work was also supported by the Wheatamix project
856 (https://www6.inra.fr/wheatamix_eng/), grant ANR-13-AGRO-0008 of the French National
857 Research Agency, for field trials.

858

859 REFERENCES

- 860 Alday, P., 2016 A brief introduction to mixed effects models - analogues to the coefficient of
861 determination R^2
862 <https://www.biorxiv.org/content/biorxiv/suppl/2016/07/06/062299.DC1/062299-1.pdf>
- 863 Barot, S., V. Allard, A. Cantarel, J. Enjalbert, A. Gauffreteau, I. Goldringer, J-C. Lata, X.
864 Leroux, A. Niboyet, E. Porcher, 2017 Designing mixtures of varieties for
865 multifunctional agriculture with the help of ecology, a review. *Agron. Sustain. Dev.* 37
866 (2): 13. <https://doi.org/10.1007/s13593-017-0418-x>
- 867 Bates, D., M. Mächler, B. Bolker, S. Walker, 2015 Fitting linear mixed-effects models using
868 lme4. *Journal of Statistical Software*, 67 (1). <https://doi.org/10.18637/jss.v067.i01>.
- 869 Bernardo, R. 1995 Genetic Models for Predicting Maize Single-Cross Performance in
870 Unbalanced Yield Trial Data. *Crop Science* 35 (1): 141-47.
871 <https://doi.org/10.2135/cropsci1995.0011183X003500010026x>.
- 872 Bernardo, R., 1996 Best linear unbiased prediction of maize single-cross performance. *Crop*
873 *Sci*, 36 (1): 50-56. doi:10.2135/cropsci1996.0011183X003600010009x
- 874 Borg, J., L. P. Kiær, C. Lecarpentier, I. Goldringer, A. Gauffreteau, S. Saint-Jean, S. Barot, J.
875 Enjalbert, 2018 Unfolding the potential of wheat cultivar mixtures A meta-analysis
876 perspective and identification of knowledge gaps. *Field Crop Res.* 221, 298-313.
877 <https://doi.org/10.1016/j.fcr.2017.09.006>
- 878 Chalbi, N., 1967 La compétition entre génotypes et ses effets sur les caractères quantitatifs
879 de la Luzerne. *Annales d'Amélioration des Plantes* 17 (1): 67-82.
- 880 Cowger, C., R. Weisz., 2008 Winter wheat blends (mixtures) produce a yield advantage in
881 North Carolina. *Agron. J.* 100 (1): 169-77. <https://doi.org/10.2134/agronj2007.0128>.
- 882 Creissen, H. E., T. H. Jorgensen, J. K. M. Brown, 2016 Increased yield stability of field-grown
883 winter barley (*Hordeum Vulgare* L.) varietal mixtures through ecological processes.
884 *Crop Prot.* 85: 1-8. <https://doi.org/10.1016/j.cropro.2016.03.001>
- 885 Darwinkel, A., 1978 Patterns of tillering and grain production of winter wheat at a wide range
886 of plant densities. *Neth. J. Agric. Sci.*, 26, 383-98.

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

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

893 Falconer, D. S., T. F. C. Mackay, R. Frankham, 1996 *Introduction to Quantitative Genetics*.
894 4th ed. Benjamin Cummings. *Trends in Genetics*, 12 (7), 280.

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

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

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

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

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

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

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

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

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

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

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

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

924 Kaut, A. H. E. E., H. E. Mason, A. Navabi, J. T. O'Donovan, D. Spaner, 2009 Performance
925 and stability of performance of spring wheat variety mixtures in organic and
926 conventional management systems in western Canada. *J. Agron. Crop Sci.* 147 (2):
927 141-53. <https://doi.org/10.1017/S0021859608008319>

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

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

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

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

940 Lecarpentier, C., 2017 Plasticité de l'architecture aérienne du blé en réponse à la compétition
941 pour la lumière au sein de cultures pures ou d'associations variétales :
942 Caractérisation expérimentale et développement d'un modèle 3D. PhD thesis. Ecole

943 doctorale Sciences du végétal, Université Paris-Saclay.
944 <http://www.sudoc.fr/219507198>

945 Lecapentier, C., R. Barillot, E. Blanc, M. Abichou, I. Goldringer, P. Barbillon, J. Enjalbert, B.
946 Andrieu. 2019 WALTer: A three-dimensional wheat model to study competition for
947 light through the prediction of tillering dynamics. *Ann. Bot.* 123 (6): 961-975.
948 <https://doi.org/10.1093/aob/mcy226>

949 Lopez, C. G., C. C. Mundt, 2000 Using mixing ability analysis from two-way cultivar mixtures
950 to predict the performance of cultivars in complex mixtures. *Field Crops Res.* 68 (2):
951 121-32. [https://doi.org/10.1016/S0378-4290\(00\)00114-3](https://doi.org/10.1016/S0378-4290(00)00114-3).

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

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

956 McGilchrist, C. A., 1965 Analysis of competition experiments. *Biometrics*, 975-985. DOI:
957 10.2307/2528258

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

962 Möhring, J., A.E. Melchinger, H.-P. Piepho, 2011 REML-based diallel analysis. *Crop Sci* 51,
963 470-478. doi:10.2135/cropsci2010.05.0272

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

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

970 Newton, A. C., G. S. Begg, J. S. Swanston, 2009 Deployment of diversity for enhanced crop
971 function. *Annals of Applied Biology*, 154(3): 309-322. [https://doi.org/10.1111/j.1744-](https://doi.org/10.1111/j.1744-7348.2008.00303.x)
972 [7348.2008.00303.x](https://doi.org/10.1111/j.1744-7348.2008.00303.x)

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

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

981 Raghavarao, D., W. T. Federer, 2003 Sufficient conditions for balanced incomplete block
982 designs to be minimal fractional combinatorial treatment designs, *Biometrika*, 90 (2):
983 465-470. <https://doi.org/10.1093/biomet/90.2.465>

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

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

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

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

995 Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, E. Siemann, 1997 The influence of
996 functional diversity and composition on ecosystem processes. *Science* 277 (5330):
997 1300-1302. <https://doi.org/10.1126/science.277.5330.1300>.

- 998 Vallavieille-Pope, C. de., 2004 Management of disease resistance diversity of cultivars of a
999 species in single fields: controlling epidemics. *Comptes Rendus Biologies* 327 (7):
1000 611-620. <https://doi.org/10.1016/j.crv.2003.11.014>.
- 1001 Weisser, W. W., C. Roscher, S. T. Meyer, A. Ebeling, G. Luo, et al. 2017 Biodiversity effects
1002 on ecosystem functioning in a 15-year grassland experiment: patterns, mechanisms,
1003 and open questions. *Basic Applied Ecology* 23: 1-73.
1004 <https://doi.org/10.1016/j.baae.2017.06.002>
- 1005 Wolfe, M. S., 1985 The current status and prospects of multiline cultivars and variety mixtures
1006 for disease resistance. *Annu. Rev. Phytopathol.* 23 (1): 251-273.
1007 <https://doi.org/10.1146/annurev.py.23.090185.001343>
- 1008 Xu, R., 2003 Measuring explained variation in linear mixed effects models. *Statist. Med.*, 22:
1009 3527-3541. doi:10.1002/sim.1572
- 1010 Zhao, Y., Z. Li, G. Liu, Y. Jiang, H. P. Maurer, T. Würschum, H.-P. Mock, A. Matros, E.
1011 Ebmeyer, R. Schachschneider, E. Kazman, J. Schacht, M. Gowda, C. F. H. Longin, J.
1012 C. Reif., 2015 Genome-based establishment of a high-yielding heterotic pattern for
1013 hybrid wheat breeding. *Proc. Natl. Acad. Sci.* 112 (51): 15624-29.
1014 <https://doi.org/10.1073/pnas.1514547112>.
- 1015 Zhu, Y., H. Chen, J. Fan, Y. Wang, Y. Li, J. Chen, J.X. Fan, S. Yang, L. Hu, H. Leung, T. W.
1016 Mew, P. S. Teng, Z. Wang, C. C. Mundt, 2000 Genetic diversity and disease control in
1017 rice. *Nature* 406 (6797): 718-22. <https://doi.org/10.1038/35021046>.

1018

1019 **SUPPLEMENTARY MATERIAL**

1020 **File S.1:** R program adapted from lme4 for mixing ability analysis and Le Moulon trial data for
1021 application example

1022 The program code is available on the GitHub public repository:

1023 <https://github.com/cambroise/lme4-adapt-for-variety-mixture>

1024