Modern plant breeding activities consist of evaluating the genetic merit of genotypes, discerning genetic performance from environmental variability (Bernardo, 2010). The variability can be explained by micro-environmental and macro-environmental variability sources (Falconer and Mackay, 1996). The macro-environmental variability is determined by differences in environmental conditions, including soil type, climate, and agronomic management practices. The genotypic response to different environments is complex, and genotype × environment interaction (GEI) is modeled. The objective of this study was to compare strategies for micro and macro-environmental variability control that include GEI information to optimize resource allocation in multi-environment trials (METs). Six experimental designs combined with four spatial correction models were compared for efficiency under three experimental sizes using simulations under a real yield variability map. Additionally, six resource allocation strategies were evaluated in terms of accuracy and the expected response to selection. The α-lattice (ALPHA) experimental design was the best one at controlling micro-environmental variability. The moderate mega-environmental design (MED) strategy had the largest response to selection. This strategy uses historical mega-environments (MEs) to unbalance genotypic testing within MEs while modeling GEI. The MED was the best resource allocation strategy and could potentially increase selection response up to 43% in breeding programs when genotypes are evaluated in METs.
interaction is present when genotypic response differences vary in relation to the environmental conditions (Malosetti et al., 2016). On the other hand, micro-environmental variability is mainly determined by spatial variability and other local scale factors (Wu, 1997).

The macro-environmental variability can cause genotype × environment interaction (GEI) and can be evaluated with multi-environment trials (METs) across locations and years (Yan and Kang, 2003). Different approaches have been proposed to study GEI such as linear–bilinear models (Finlay and Wilkinson, 1963; Gauch, 1992; Yan et al., 2001; Crossa and Cornelius, 2002), mixed models (Piepho, 1998; Burgueño et al., 2008; Cullis et al., 2010), and crop growth models (Chapman, 2008; Technow et al., 2015; Malosetti et al., 2016). These methods have been extensively used for different breeding strategies including traditional phenotypic selection and testing (Yan et al., 2007; González-Barrios et al., 2017), quantitative trait loci mapping (Piepho, 2000; Malosetti et al., 2004; Mathews et al., 2008; Quero et al., 2014), genome-wide association mapping (van Eeuwijk et al., 2010; Locatelli et al., 2013; Gutiérrez et al., 2015; Racedo et al., 2016; Montevede et al., 2018), and genomic selection (Burgueño et al., 2012; Lopez-Cruz et al., 2015; Lado et al., 2016). Environments can then be grouped in sets that produce a similar ranking of the genotypes (mega-environments [MEs]; Braun et al., 1996). The identification of repeatable ranking patterns through MEs has been proposed as an efficient alternative to minimize the presence of GEI within groups of environments (Braun et al., 1996; Heslot et al., 2013; Lado et al., 2016). The use of GGE biplots (i.e., biplots based on environment-centered GEI means, where both the genotypic and GEI effects are graphically displayed) has been widely used to define ME in METs based on the winning genotype in each environment (Yan et al., 2000; Yan and Kang, 2003). Mega-environments can therefore be used to improve resource allocation in METs (Windhausen et al., 2012). The GEI analysis contribute to the understanding of the macro-environmental variability. However, the integration of both micro- and macro-environmental variability sources is a key factor for plant breeding program success.

The most effective way of controlling micro-environmental variability in agriculture is through experimental design (Casler, 2015; Piepho et al., 2015). To improve the efficiency of the estimation of genotypic effects, R.A. Fisher (1935) proposed three principles for experimental design: randomization, replication, and local control. For decades, blocks have been widely used as an efficient tool for local control and to reduce experimental error (Cochran and Cox, 1957; Casler, 2015). Blocking effectiveness depends on block size (i.e., number of plots per replication), shape, and orientation (Warren and Mendez, 1981). In particular, randomized complete block designs (RCBDs) are the most commonly used experiments in agriculture, especially those with a small number of treatments (Piepho et al., 2015). However, the inadequate use of blocking structures can lead to an increase in experimental error estimates directly affecting the estimation of the genotypic rankings, among other issues (Stroup et al., 1994). In situations where a large number of genotypic evaluations is required (i.e., plant breeding programs and genotypic performance testing), the use of blocks of large dimensions can lead to an inefficient control of the spatial heterogeneity (Brownie et al., 1993). To overcome this limitation, experimental designs such as α–designs (ALPHA) or partially replicated (PREP) have been recommended (Piepho et al., 2015, 2016). Although the use of resolvable incomplete blocks (Williams et al., 2002) or row–column (Williams et al., 2006) may capture most of the spatial variability with their incomplete blocks (Müller et al., 2010), spatial variability may not be controlled by experimental design in all cases (Grondona and Cressie, 1991).

The selection of the appropriate experimental design and spatial model analysis is therefore a determinant factor to control micro-environmental variability (Casler, 2015). In agricultural experimentation, factors such as soil heterogeneity, agricultural practices, and environmental conditions influence the genotypic performance of lines and contribute to the spatial variability (Arnold and Kempton, 1979; Gilmour et al., 1997). Therefore, modeling spatial correlations might be necessary to improve genotypic effect estimation even after a good experimental design is used (Federer, 1998; Qiao et al., 2000; Campbell and Bauer, 2007; Casler, 2015; Borges et al., 2019). Several approaches have been proposed to control spatial variability such as nearest-neighbor adjustment (Katsilieros et al., 2015), smoothing techniques including penalized splines analysis (Stefanova et al., 2009; Piepho and Williams, 2010; Velazco et al., 2017), modeling the variance–covariance matrix of spatial correlations using geostatistical components (Williams, 1986; Williams et al., 2006; Piepho and Williams, 2010), or using mixed models (Smith et al., 2005). In mixed models, one- or two-dimensional models are used to control spatial heterogeneity with two-dimensional models outperforming one-dimensional models in terms of the genotypic effects estimation (Cullis and Gleseson, 1991, Kempton et al., 1994, Piepho and Williams, 2010). Several authors propose that spatial models could be used to properly control spatial heterogeneity (Cullis and Gleseson, 1991; Grondona et al., 1996). This argument could then be used as a means to replace experimental designs with spatial modeling. Borges et al. (2019) showed that this is not efficient.

Because the availability of experimental plots for testing is limited by economic resources, a decision must be made in terms of resource allocation between the number of replications and environments used for testing.
each genotype, and the total number of genotypes to be used. Most of the historical MET for evaluation of elite genotypes have used experimental designs with replications that were balanced across locations and generally unbalanced across years (Piepho and Möehring, 2007). More recently, partially replicated experiments that are balanced across locations were proposed as an alternative to increase the number of genotypes tested (Talbot, 1984; Stendal and Casler, 2006; Möehring et al., 2014). Furthermore, Lorenz (2013) recommends increasing population sizes rather than replication when simulating resource allocation strategies for genomic selection. Additionally, Paget et al. (2017) found that full replications at early stages of evaluation and the use of PREP designs evaluated over several locations improves selection efficiency, and McCann et al. (2010) suggests that evaluation over several locations and years is more efficient than increasing replications within a single location. Furthermore, Endelman et al. (2014) found that it is more efficient to use unbalanced designs spread across locations than testing all entries in one location. This is also supported by Möehring et al. (2014). Our hypothesis is that the use of small replicated and balanced designs across locations could hinder selection response in breeding programs or testing trials compared with evaluating larger population sizes in more environments modeling GEI even at the cost of replications and balance. Therefore, the objective of this study was to compare strategies to optimize resource allocation for genotypic evaluation in multi-environment testing. Specifically, we aimed to compare experimental design strategies based on both micro-environmental and macro-environmental control.

**MATERIALS AND METHODS**

This study is based on the comparison of different strategies of resources allocation in plant breeding programs controlling micro- and macro-environmental variability sources. To evaluate the effect of micro-environmental variability on optimal experimental design, different combinations of experimental design and spatial modeling were evaluated through simulations on real spatial variability patterns generated by uniformity trials (Fig. 1; following Borges et al., 2019). To control macro-environmental variability sources, different resource allocation strategies were evaluated using simulations on real spatial variability patterns, across several locations and using the best strategies from the micro-environmental section (Fig. 2). For both analyses, historical genotypic information of a population of 148 oat (Avena sativa L.) inbred lines evaluated across six locations and 12 yr in the Wisconsin Oat Breeding Program (WOBP) were used. Molecular marker information was used to model genetic relationships among genotypes. Accuracy and precision estimators were used to compare allocation strategies.

**Phenotypic Information**

The phenotypic information used for this study was obtained through historical information from the WOBP. We used information from the elite yield trials (EYTs) from 2006 to 2017. The EYTs were evaluated every year in RCBD with four replications in six locations through the network of experimental research stations of the University of Wisconsin. The six locations were Arlington (43°18’11"’N, 89°20’42"’W), Lancaster (42°49’50"’N, 90°47’17"’W), Madison (43°03’37"’N, 89°31’54"’W), Marshfield (44°45’41"’N, 90°05’58"’W), Spooner (45°39’19"’N, 91°52’33"’W), and Sturgeon Bay (44°52’47"’N, 87°20’11"’W). The number of genotypes evaluated in each location and in each year ranged from 20 to 45 genotypes in each year (Fig. 3). Also, we used information from the uniformity oat performance nurseries (UOPN) from 2016 and 2017. The UOPN experiments were evaluated using a RCBD with four replications in two locations (Arlington and Madison).

A two-step approach was used to estimate genotypic means. First, best linear unbiased estimates (BLUEs) were estimated by location. Second, GEI was modeled across locations to obtain the full GEI table of means for the simulations. The following mixed model was used to estimate the BLUEs of each genotype in each location for the first step (random effects are underlined in all equations):

\[
Y_{ijkl} = \mu + G_i + \gamma_j + \beta_{k(i)} + S_l + G_S d + \varepsilon_{ijkl} \quad [1]
\]

where \(Y_{ijkl}\) is the observed yield in the \(i\)th genotype, \(j\)th trial, \(k\)th replication, and the \(l\)th year, \(\mu\) is the overall mean, \(G_i\) is the effect of the \(i\)th genotype, \(\gamma_j\) is the effect of the \(j\)th genotype (i.e., EYT or UOPN), \(\beta_{k(i)}\) is the effect of the \(k\)th replication (i.e., blocks) nested on the \(j\)th trial, \(S_l\) is the effect of the \(l\)th year, \(G_S\) is the effect of the interaction between the \(i\)th genotype and the \(l\)th year, and \(\varepsilon_{ijkl}\) are the residuals, with \(\gamma_j \sim N(0, \sigma^2_{\gamma})\), \(\beta_{k(i)} \sim N(0, \sigma^2_{\beta})\), \(S_l \sim N(0, \sigma^2_S)\), \(G_S \sim N(0, \sigma^2_{GS})\), and \(\varepsilon_{ijkl} \sim N(0, \sigma^2_{\varepsilon})\); all independent; where \(\sigma^2_{\gamma}\) is the trial variance, \(\sigma^2_{\beta}\) is the block variance, \(\sigma^2_S\) is the year variance, \(\sigma^2_{GS}\) is the genotype × year interaction variance, and \(\sigma^2_{\varepsilon}\) is the mean error variance across experiments. The trial effect was included in Madison and Arlington because both EYT and UOPN are included. No interaction between trial and genotype was included in this model because planting dates, fields, and management was the same for both trials, but also because genotypes are mainly nested within trials.

Because the data were also unbalanced across locations, the following mixed model was used to estimate unobserved genotypic effects in each location in the second step:

\[
Y_{im} = \mu + G_i + E_m + GE_{im} \quad [2]
\]

where \(Y_{im}\) are the BLUEs of the \(i\)th genotype in the \(m\)th location estimated from Eq. [1], \(E_m\) is the effect of the \(m\)th location, \(GE_{im}\) is the effect of the interaction between the \(i\)th genotype and the \(m\)th location, with \(GE_{im} \sim N(0, \sigma^2_{GE})\), and \(\sigma^2_{GE}\) is the genotype × environment interaction variance. Different variance–covariance structures were evaluated for modeling the Cov(\(G_i, GE_{im}\)) unstructured, compound symmetry, first-order autoregressive (AR(1)), heterogeneous autoregressive, and factor analytic of order 1 (FA1). The FA1 model was the best GE model selected based on the Akaike information criterion (AIC, Supplemental Table S1). These analyses were performed in SAS software (SAS Institute, 2013). Finally, the full table of mean and standard error for each genotype in each location was used to simulate the yield performance in each location for a
Fig. 1. Simulation strategy for comparison of experimental design and spatial modeling performance with micro-environmental variability. (1) A real field variability map was obtained through an interpolation of a uniformity trial, and 50 sites were sampled within the field (Site $i$). (2) The spatial pattern of the field was used as one of the inputs for the simulation ($\varepsilon_{ij}^*$). (3) A multi-environment testing dataset of oats from the Wisconsin Oat Breeding Program with six locations evaluated from 2006 to 2017 was used to estimate genotypic means and to evaluate the genotype by environment structure of the data. (4) A two-step process was used to adjust genotypic means for each trial. First, genotypic means were estimated across years for each location. The mean squared error (MSE) at each location was weighted and used to simulate the variance in the repeatability error $\text{Var}(d_{ij})$. (5) The table of genotypic effects for all locations was obtained in the second step where a factor analytic model was used to complete missing information and to obtain genotypic effects for each location ($g_i$). The genotypic effects from this step were considered our “true” genotypic values for the remainder of the simulation. (6) Genotypes were randomized in one of six experimental designs: completely randomized design (CRD), randomized complete block design (RCBD), incomplete blocks–$\alpha$-lattice design (ALPHA), partially replicated design using the same number of genotypes as CRD, RCBD and ALPHA (PREP$g$), partially replicated design using the same number of experimental units as CRD, RCBD, and ALPHA (PREP$n$), and the unreplicated design (UNREP). (7) A yield value was simulated for each experimental unit ($Y_{ij}^\text{sim}$) using the genotypic effect ($g_i$) from the real dataset and the treatment assigned by the experimental design randomization process, the spatial heterogeneity of the field ($\varepsilon_{ij}^*$) that was obtained from the predicted values of the uniformity trials, and a repeatability error ($d_{ij}$) that was sampled from a normal distribution with variance = 10% of MSE of the real data for that location. (8) The simulated yield ($Y_{ij}^\text{sim}$) was used in the analysis model as the response variable where genotypic values were predicted ($\hat{g}_i$) while correcting for the experimental design (CRD, RCBD, ALPHA, PREP$g$, PREP$n$, or UNREP) and modeling the spatial variability. Four variance–covariance structures were used to model spatial correlations: no spatial correction model (NSC, independent errors), first-order autoregressive process (AR1), two-dimensional exponential process (EXP), and the two-dimensional spline model (S2D). (9) Accuracy and precision statistics were calculated: the proportion of the best 15% genotypes (top 15% $g_i$) recovered by the model (top 15% $\hat{g}_i$) (BEST), the correlation between true ($g_i$) and predicted ($\hat{g}_i$) genotypic values (COR), and the standard error of the difference between genotypic means (SED). Each combination of experimental design and spatial correction was evaluated for every site in all six locations, and for either 64, 160, or 592 total experimental units.
Three uniformity experiments in oats were performed in the 2017 and 2018 cropping seasons with the objective of obtaining real variability patterns in which to evaluate different experimental designs and resource allocation strategies. Trials were performed in 2017 (UNI1) and 2018 (UNI2) in different fields in Madison, and in 2018 in Arlington (UNI3). The seedbed for all the experiments was prepared with a field cultivator. The UNI1 experiment was planted at a seeding rate of 100 kg ha$^{-1}$ of the oat variety ‘Ron’ (WCIA, 2015). The field average dry yield was 2616 kg ha$^{-1}$ at 12% moisture. The total area used for the experiment was 2.07 ha. The UNI2 experiment was planted at a seeding rate of 112 kg ha$^{-1}$ of the oat variety ‘Betagene’ (WCIA, 2015). The field dry yield was 5605 kg ha$^{-1}$ at 12% moisture. The total area used for the experiment was 2.83 ha. For UNI3, the variety ‘Reins’ (Kolb et al., 2017) was planted with a seeding rate of 103 kg ha$^{-1}$. The field dry yield
was 2824 kg ha\(^{-1}\) at 12\% moisture. The total area used for the experiment was 2.63 ha. The combine harvester used for the UNI1 and UNI2 experiments was a CASE IH 6140 with a yield monitor (AFS Pro700, Case IH) and an operating width of 7.62 m. The combine harvester for UNI3 was a John Deere 9400 with a Greenstar display yield monitor (John Deere and Company). The yield monitor recorded yield flow every 1 s at an average speed of 0.67 m s\(^{-1}\) (\(\sim\)5 m\(^2\) s\(^{-1}\)). Common agronomic practices for oat in the Midwest region of the United States were used for all experiments.

A yield variability map was generated using the information from each uniformity experiment (Supplemental Fig. S1). First, the database was curated by removing extreme yield data points due to high or low speed of the combine, or excessive flow. A total of 252 out of 2573 points were removed as outliers in UNI1, 1376 out of 2860 in UNI2, and 74 out of 2768 in UNI3 (Supplemental Table S2). This procedure was performed with the Yield Editor software, version 2.1 (Sudduth and Drummond, 2007). Using information of the processed data, an empirical semivariogram was constructed with the objective of modeling the variability present in the experimental site. Three types of models were adjusted: exponential, Gaussian, and spherical. The best model for each field (i.e., exponential) was selected based on the lowest value of the sum of squares of residues.

Fig. 3. Overall description of the historical dataset of the Wisconsin Oat Breeding Program. The number of genotypes evaluated in each trial are represented in the main diagonal in shades of orange. The number of genotypes evaluated in common between each pair of environments is represented in the off-diagonal in shades of blue. The environments are sorted in each year as a function of the mega-environments (MEs): Lancaster, Madison, and Sturgeon Bay for ME1, and Arlington, Marshfield, and Spooner for ME2.
the errors (Supplemental Table S2) to construct a prediction grid for interpolation with a density of $2 \times 3$ m over the experimental area. A total of 6300 points were predicted for UNI1, 7300 points for UNI2, and 8200 points for UNI3 using the best model. The analyses were implemented using the “gstat” and “sp” (Pebesma and Bivand, 2005) packages from the R statistical software (R Core Team, 2016).

**Molecular Marker Information**

Genotyping by sequencing (GBS) data was obtained for all the lines using the methods described by Huang et al. (2014). The single nucleotide polymorphism (SNP) genotype calls were made with the Universal Network Enabled Analysis Kit (UNEAK) pipeline (Ly et al., 2013) and the PstI-MspI restriction enzyme combination. Markers with a minor allele frequency (MAF) $<0.01$ and $>70\%$ of missing data were manually removed from the original dataset, which resulted in 43,660 markers. When estimating the realized additive relationship matrix on each simulation, an additional filter was applied to remove low quality markers (MAF = 0.05 and missing data = 50\%). The realized additive relationship matrix between genotypes was estimated in each simulation and scenario in the macro-environmental design strategies using the “A.mat” function of the “soommer” package (Covarrubias-Pazaran, 2016) of the R statistical software (R Core Team, 2016). The A matrix evaluated at each realization was highly correlated with the full A matrix evaluated with all individuals. The realized additive relationship matrix (Supplemental Fig. S2) was used to model genetic relationships in the resource allocation strategies for the GEI section as described in the macro-environmental design efficiency.

**Micro-Environmental Variability**

Six experimental designs with four levels of correction for spatial variability were evaluated in six locations through a simulation process following Borges et al. (2019). Yield data from genotypes were simulated according to an experimental design using real field variability maps, and genotypic effects were predicted for each location based on historical information. All the scenarios were simulated at 50 positions in the field variability map and compared through accuracy, precision, and model ability to recover the original ranking of the genotypes. Details of this approach are described below.

The simulation process generated yield values for each plot where the genotypes were randomly assigned to a plot according to each experimental design. The simulated plot yield was calculated according to Eq. [3]:

$$Y_{ij} = G_i + e_i + \epsilon_{ij}$$

where $Y_{ij}$ is the yield plot data simulated corresponding to the $i$th genotype and the $j$th replication, $G_i$ is the $i$th genotypic effect from the historical dataset corresponding to the randomly assigned treatment to the plot, $e_i$ is the field experimental error that represent the spatial heterogeneity of the field and was obtained from the predicted values of the uniformity trials, and $\epsilon_{ij}$ is a repeatability error following Borges et al. (2019). The $\delta_{ij}$ were assumed as independent random variables with $\delta_{ij} \sim N(0, \sigma^2_{\delta})$, where $\sigma^2_{\delta}$ is a random noise or repeatability that represented a percentage of the total field heterogeneity of each location with $\sigma^2_{\delta\text{lock}} = 0.10 \times \text{MSE}_{\text{lock}}$ to simulate an average heritability on a mean basis of 0.86. The MSE_{lock} was obtained from the historical information for each location. Additionally, different levels of heritability were simulated for each location increasing the value of $\sigma^2_{\delta}$ with additional noise for a range of 10, 50, and 100\% of the MSE_{lock}.

**Experimental Design Comparisons**

We compared the following experimental designs: completely randomized design (CRD), randomized complete block design (RCBD), and incomplete block-α design (ALPHA) with four replications, two partially replicated experimental designs (PREP and PREP′) with repeated genotypes assigned in a RCBD with two replications, and an unreplicated design (UNREP).

All experimental designs were analyzed with four levels of spatial corrections (no-spatial correction, first-order autoregressive, exponential, and splines) and evaluated under three levels of population size: small, moderate, and large. The small population size used 16 genotypes for the CRD, RCBD, ALPHA, and the PREP, 58 genotypes for the PREP′, and 64 genotypes for the UNREP. The moderate population size used 40 genotypes for the CRD, RCBD, ALPHA, and the PREP, 144 genotypes for the PREP′, and 160 genotypes for the UNREP, and the large population size used 148 genotypes for the CRD, RCBD, ALPHA, and the PREP. In the large size, neither the PREP nor the UNREP strategies were evaluated due to the lack of phenotypic and genotypic information for additional lines.

The CRD model was

$$Y_{ij} = \mu + G_i + e_{ij}$$

where $Y_{ij}$ is the yield simulated for the $i$th genotype and $j$th replication (i.e., blocks), $\mu$ is the overall mean, $G_i$ is the effect of the $i$th genotype, and $e_{ij}$ is the residual error associated with the $ij$th observation with $e_{ij} \sim N(0, \sigma^2_e)$, where $\sigma^2_e$ is the residual error variance and $S$ is the variance–covariance matrix of residual errors, with $S = \text{Cov}(\epsilon_{ij}, \epsilon_{ij})$ modeled with different spatial correlation structures as described in the section below.

The RCBD model was

$$Y_{ij} = \mu + G_i + \beta_j + e_{ij}$$

where $\beta_j$ is the effect of the $j$th replication.

The ALPHA model was

$$Y_{jk} = \mu + G_i + \beta_j + \gamma_{k(i)} + e_{jk}$$

where $\gamma_{k(i)}$ is the effect of the $k$th incomplete block nested within the $j$th replication. Incomplete blocks were considered as random factors nested in each full replication, with $\gamma_{k(i)} \sim N(0, \sigma^2)$ where $\sigma^2$ is the incomplete blocks variance, and $\text{Cov}(\gamma_{k(i)}, \epsilon_{jk}) = 0$.

Fifteen percent of the experimental units were evaluated with replicated genotypes in the PREP design. A RCBD with two replications was used to randomize replicated genotypes in the field. The remaining plots contained unreplicated genotypes assigned at random to the remaining experimental units.
This experimental design follows Cullis et al. (2006) where PREP designs are randomized like Federer’s (1956) augmented experimental designs, but repeated genotypes are used instead of checks. The PREP model was

$$Y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij}$$  \[7\]

where $G_i$ is the effect of the $i$th genotype when we assume there are $n_i$ replicated genotypes and $n_{a}$ nonreplicated genotypes for $i = 1, \ldots, n_i, \ldots, n_i + n_{a}$. The PREP design used the same number of genotypes evaluated in the CRD, RCBD, and ALPHA approaches. This design uses fewer experimental units than CRD, RCBD, and ALPHA. An alternative design was evaluated using the PREP design but assuming the same number of experimental units evaluated in the CRD, RCBD, and ALPHA design (PREP), with more replicates evaluated.

The UNREP design was created by completely randomizing genotypes to experimental units in single replications and using the same number of experimental units as the CRD, RCBD, ALPHA, and PREP experiments. The UNREP model was

$$Y_{i} = \mu + G_i + \varepsilon_i$$  \[8\]

where terms are described in a similar way as before, but no replications exist.

**Spatial Correction Methods**

Each experimental design was evaluated with four spatial adjustment models: no spatial correction (NSC), spatially correlated error model with first-order autoregressive process (AR1), spatially correlated errors model with a two-dimensional exponential process (EXP), and two-dimensional spline correction (S2D). For model NSC, the experimental errors are assumed as independent [uncorrelated, Cov($\varepsilon_{ij}, \varepsilon_{ij}) = 0$]. In model AR1, the correlation function corresponds to a first-order autoregressive model and decreases in absolute value with every unit of distance within columns: $h(k, \rho) = \rho^k, k = 0, 1, \ldots, \alpha$, where $\rho$ is the correlation parameter to be estimated and $k$ is the distance unit (i.e., in rows). In the EXP model, the correlation function corresponds to a two-dimensional process with a decreasing exponential correlation in the row and column directions. Finally, in the S2D model, information from rows and columns is used to fit a two-dimensional spline model. The simulations and statistical analysis of the data were performed with the “lme” package (Bates et al., 2015) for the NSC, AR1, and EXP spatial correction, and the “sommer” package (Covarrubias-Pazaran, 2016) for the S2D of the R Statistical Software (R Core Team, 2016).

For each combination of experimental design and spatial analysis, 1000 simulations on 50 randomized positions over the field were run performing an independent randomization for each simulation. The simulation procedure was performed in each location independently. The R package “agricolae” (de Mendiburu, 2012) was used for the experimental design randomizations. Custom R and Bash scripts were used for simulation over the field and data analysis. Due to the large computational power required, we partitioned the simulations in the Center for High-Throughput Computing of the University of Wisconsin–Madison using the software HTCondor (Thain et al., 2005).

**Macro-Environmental Design Efficiency**

Six resource allocation strategies were compared for a program similar to the WOBP that is evaluated in a MET (Fig. 1). All the strategies were compared using the same number of locations (six) and number of experimental units (60 in each location). All strategies were evaluated with slight variations of the following general model:

$$Y_{ijk} = \mu + G_i + E_j + GE_{ij} + \beta_{k(i)} + \varepsilon_{ijk}$$  \[9\]

where $Y_{ijk}$ is the yield simulated for the $i$th genotype, the $j$th location, and the $k$th replication, $\mu$ is the overall mean, $G_i$ is the effect of the $i$th genotype, $E_j$ is the effect of the $j$th location, $GE_{ij}$ is the genotype $\times$ location interaction, $\beta_{k(i)}$ is the effect of the $k$th block nested in the $j$th location, and $\varepsilon_{ijk}$ is the residual error associated with the $ij$th observation, with $G_i \sim N(0, \sigma^2_g)$, $GE_{ij} \sim N(0, \sum_{v=1}^{c} \sigma^2_{ge})$, and $\varepsilon_{ijk} \sim N(0, R\sigma^2_e)$, where $\sigma^2_g, \sigma^2_{ge}$, and $\sigma^2_e$ are the genotypic, the GEI, and the residual error variance, respectively, and $G$ is the variance–covariance matrix among genotypes. Two variance–covariance structures were evaluated for the relationship between genotypes’ identity and realized additive relationship matrix. The identity relationship assumed that $G = I$, with $I$ being an identity matrix of size $g$. The realized additive relationship matrix assumed that $G = K_{x}$, with $K$ being the additive relationship matrix estimated with marker effects. The variance–covariance structure of the GEI ($\Sigma_{ge}$) was modeled as an unstructured variance–covariance, and the variance–covariance structure of residuals ($R$) was modeled with a 2DS. Finally, the covariance among random effects is assumed zero.

The first strategy (RCBD) used the historical resource allocation of the WOBP with 20 genotypes evaluated using a RCBD with three replications in each location. Each location was analyzed separately with Eq. [9], but no GE term was included in this model. The second strategy (RCBD) used the same resource allocation structure than the RCBD, but the GEI was modeled as in Eq. [9].

The third strategy (PREP) used an optimal allocation modified from Möehring et al. (2014) where genotypes are partially replicated within environments but balanced across environments, with six genotypes replicated twice in each location (PREP). Each location was analyzed separately with Eq. [9], but no GE term was included in this model. The fourth strategy (PREP) used the same resource allocation structure than the PREP, but the GEI was modeled using the Eq. [9].

For the fifth and sixth strategies, we proposed a new experimental design where the MEs are used to distribute genotypes across locations in a partially replicated manner, and we call this strategy a ME design (MED). A set of replicated genotypes was evaluated in all environments, and all the genotypes were evaluated at least once in each ME. In our case, two MEs with three locations each were defined (ME1: Lancaster, Madison, and Sturgeon Bay; ME2: Arlington, Marshfield, and Spooner; Fig. 4) based on the historical information as described previously. For the fifth strategy, we used the MED in an extreme allocation strategy where all the genotypes were evaluated at least once within a ME, but most of them were unreplicated in each location and across locations (Fig. 2). Six genotypes were replicated twice in each location within and across the MEs,
and 48 genotypes were unreplicated in each location and were present in only one location within the ME (MED_20; 20% of replication within each ME). For the sixth strategy, we used a moderate version of the MED where a set of eight genotypes were replicated twice in each location within and across MEs. Another set of eight genotypes were replicated three times within a location, but different genotypes were used for each location within a ME. Finally, 72 genotypes were evaluated unreplicated within a location and in only one location within the ME (MED_60; 60% of replication within each ME). The extreme and moderate MED strategies were evaluated with the following model

$$Y_{ijkl} = \mu + G_i + M_l + E_{ijl} + G_E_{ijkl} + \beta_{k(i)} + \varepsilon_{ijkl}$$  

[10]

where $M_l$ is the effect of the $l$th ME. Because not all genotypes are evaluated in all the locations, genotypic predictions were performed for the ME.

**Strategy Comparison**

All strategies (micro- and macro-environment) were compared using three accuracy statistics: the standard error of the difference between genotypic means (SED), the correlation between simulated and predicted genotypic values (COR), and the proportion of the best 15% genotypes recovered by the model (BEST).

The expected response to selection ($R$) was calculated for the macro-environmental resource allocation strategies comparisons. The response to selection was defined from the breeder’s equation following Lorenz (2013) as:

$$R = ir_A\sigma_A$$  

[11]

where $R$ is the response to selection, $i$ is the selection intensity, $r_A$ is the selection accuracy, and $\sigma_A$ is the standard deviation of the breeding values.

---

**RESULTS**

**Wisconsin Oat Breeding Program Characterization**

Two historical MEs were defined with the real historical dataset as a function of the number of times each pair of environments was found in the same ME for each year, and by evaluating the GEI in the MET through GGE biplots (Fig. 4). The ME1 includes Lancaster, Madison, and Sturgeon Bay, whereas ME2 includes Arlington, Lancaster, and Spooner. A different number of genotypes were evaluated in each location and across years, while consecutive years have more genotypes in common than further apart (Fig. 3).

**Micro-Environmental Control**

The best experimental design is a function of the number of genotypes evaluated. For small experimental sizes, RCBD and ALPHA experimental designs were the best alternatives (Supplemental Fig. S3). Specifically, the combination of ALPHA with S2D spatial correction had the best performance. In general, the use of any spatial correction was beneficial. For moderate experimental sizes, RCBD and ALPHA designs were also the best experimental designs (Fig. 5). Specifically, ALPHA and S2D was the best combination. In general, spatial correction was effective in improving model efficiency for all the experimental designs evaluated. In large experimental sizes, the use of ALPHA design combined with S2D spatial correction showed the best performance for the three indicators (Fig. 6). The use of any spatial correction improved model efficiency for all experimental designs. In particular, for all experimental designs, the S2D was the
The magnitude of improvement through the use of S2D spatial correction was greater in experiments with simpler experimental designs such as CRD. For all cases, the standard deviation among locations with different mean and heritability was of the same magnitude as the standard deviations within locations.

**Resource Allocation Strategy Comparison**

The RCBD strategy had the highest correlation between predicted and simulated genotypic values (COR), whereas the lowest values were obtained using the extreme MED strategy (Table 1). Modeling GEI increased genotypic correlations in small amounts for the PREP with
the realized additive relationship matrix, whereas no differences were identified in the RCBD. The PREP performed almost as good as the RCBD when GEI was modeled and the identity matrix was used. For the strategies that allocate resources within MEs (such as extreme or moderate MED), larger COR values were observed for the moderate MED that has a larger proportion of replicated genotypes. Modeling the realized additive relationship matrix did not improve the COR for most strategies (Table 1). The recovery of the top 15% of the genotypes (BEST) showed a similar pattern to the COR (Table 2).

The best strategy in terms of response to selection was the moderate MED, with up to a 43.1% larger response.
Table 1. Mean and standard deviation of the Pearson’s correlation coefficients between the true and the predicted genotypic value (COR) for each location or mega-environment, and as an average for the multi-environment trial. The comparison of strategies was performed modeling the correlation between genotypes with the realized additive relationship matrix (K) or an identity matrix (I) assuming independent genotypes. Correlations for each location were estimated for the balanced randomized complete block design (RCBD) and partially replicated design (PREP) strategies, whereas correlations for the mega-environment were estimated for the extreme (MED_20) or moderate (MED_60) mega-environmental design (MED) strategies.

<table>
<thead>
<tr>
<th>G†</th>
<th>ME‡</th>
<th>Location</th>
<th>RCBD</th>
<th>PREP</th>
<th>RCBDSE</th>
<th>PREPSE</th>
<th>Extreme MED</th>
<th>Moderate MED</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>ME1</td>
<td>LAN</td>
<td>0.97 (0.03)</td>
<td>0.83 (0.14)</td>
<td>0.98 (0.04)</td>
<td>0.89 (0.15)</td>
<td>0.61 (0.11)</td>
<td>0.86 (0.12)</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>0.95 (0.02)</td>
<td>0.85 (0.12)</td>
<td>0.98 (0.06)</td>
<td>0.82 (0.15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STB</td>
<td>0.98 (0.03)</td>
<td>0.79 (0.16)</td>
<td>0.97 (0.05)</td>
<td>0.89 (0.16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME2</td>
<td>ARL</td>
<td>0.96 (0.05)</td>
<td>0.84 (0.14)</td>
<td>0.97 (0.06)</td>
<td>0.87 (0.12)</td>
<td>0.63 (0.13)</td>
<td>0.83 (0.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>0.98 (0.04)</td>
<td>0.84 (0.13)</td>
<td>0.97 (0.05)</td>
<td>0.91 (0.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>0.97 (0.03)</td>
<td>0.82 (0.17)</td>
<td>0.97 (0.05)</td>
<td>0.90 (0.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>0.97 (0.03)</td>
<td>0.83 (0.14)</td>
<td>0.98 (0.05)</td>
<td>0.87 (0.14)</td>
<td>0.62 (0.12)</td>
<td>0.85 (0.12)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>ME1</td>
<td>LAN</td>
<td>0.97 (0.05)</td>
<td>0.92 (0.10)</td>
<td>0.99 (0.04)</td>
<td>0.92 (0.10)</td>
<td>0.70 (0.06)</td>
<td>0.93 (0.04)</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>0.95 (0.07)</td>
<td>0.85 (0.12)</td>
<td>0.98 (0.03)</td>
<td>0.94 (0.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STB</td>
<td>0.96 (0.06)</td>
<td>0.90 (0.12)</td>
<td>0.98 (0.02)</td>
<td>0.94 (0.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME2</td>
<td>ARL</td>
<td>0.95 (0.05)</td>
<td>0.88 (0.12)</td>
<td>0.97 (0.05)</td>
<td>0.90 (0.12)</td>
<td>0.67 (0.16)</td>
<td>0.92 (0.16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>0.97 (0.05)</td>
<td>0.92 (0.11)</td>
<td>0.98 (0.05)</td>
<td>0.93 (0.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>0.97 (0.04)</td>
<td>0.90 (0.11)</td>
<td>0.98 (0.03)</td>
<td>0.92 (0.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>0.98 (0.03)</td>
<td>0.95 (0.05)</td>
<td>0.98 (0.03)</td>
<td>0.96 (0.05)</td>
<td>0.69 (0.10)</td>
<td>0.92 (0.09)</td>
<td></td>
</tr>
</tbody>
</table>

† G, variance–covariance matrix used to model the relationship between genotypes. K, realized additive relationship matrix estimated using molecular markers; I, identity matrix.
‡ ME, mega-environment. ME1: Lancaster (LAN), Madison (MAD), and Sturgeon Bay (STB). ME2: Arlington (ARL), Marshfield (MAR), and Spooner (SPO).
§ PREP, partially replicated design with 48 unreplicated genotypes and six replicated checks within each location; RCBD, the same as the RCBD strategy but modeling the genotype × environment interaction (GEI) variance covariance matrix with an unstructured model; PREPGEI, the same as the PREP strategy but modeling the GEI variance covariance matrix as an unstructured model; Extreme MED, mega-environmental design with 144 unreplicated genotypes across the ME and six replicated genotypes in each location within each mega-environment; Moderate MED, mega-environmental design with 72 unreplicated genotypes across the ME, eight genotypes replicated three times in each location within the ME, and six replicated genotypes twice in each location within each ME.

Table 2. Mean and standard deviation of the proportion of the recovery of 15% of the superior genotypes for each location (BEST) for each location or mega-environment, and as an average for the multi-environment trial. The comparison of strategies was performed modeling the correlation between genotypes with the realized additive relationship matrix (K) or an identity matrix (I) assuming independent genotypes. Correlations for each location were estimated for the balanced randomized complete block design (RCBD) and partially replicated design (PREP) strategies, whereas correlations for the mega-environment were estimated for the extreme (MED_20) or moderate (MED_60) mega-environmental design (MED) strategies.

<table>
<thead>
<tr>
<th>G†</th>
<th>ME‡</th>
<th>Location</th>
<th>RCBD</th>
<th>PREP</th>
<th>RCBDSE</th>
<th>PREPSE</th>
<th>Extreme MED</th>
<th>Moderate MED</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>LAN</td>
<td>84.1 (22.2)</td>
<td>72.3 (20.3)</td>
<td>88.6 (17.5)</td>
<td>76.4 (18.5)</td>
<td>39.2 (10.7)</td>
<td>58.2 (12.4)</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>77.2 (24.4)</td>
<td>65.1 (21.7)</td>
<td>91.8 (14.7)</td>
<td>66.3 (19.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STB</td>
<td>82.5 (22.7)</td>
<td>72.4 (20.7)</td>
<td>87.1 (18.1)</td>
<td>74.3 (20.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ARL</td>
<td>72.3 (26.0)</td>
<td>64.1 (19.2)</td>
<td>81.1 (20.3)</td>
<td>61.2 (18.2)</td>
<td>38.4 (10.9)</td>
<td>56.4 (13.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>86.3 (21.6)</td>
<td>74.2 (18.3)</td>
<td>88.7 (17.3)</td>
<td>67.3 (18.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>83.6 (22.3)</td>
<td>72.7 (20.3)</td>
<td>88.2 (18.1)</td>
<td>74.2 (21.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>81.0 (22.7)</td>
<td>70.1 (19.9)</td>
<td>87.5 (18.7)</td>
<td>70.0 (18.7)</td>
<td>38.8 (10.8)</td>
<td>57.3 (12.8)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>LAN</td>
<td>83.1 (18.4)</td>
<td>73.1 (14.3)</td>
<td>91.9 (20.0)</td>
<td>78.7 (14.0)</td>
<td>70.5 (10.1)</td>
<td>75.5 (12.3)</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>78.2 (18.9)</td>
<td>64.2 (14.0)</td>
<td>85.2 (19.2)</td>
<td>82.7 (13.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STB</td>
<td>83.5 (20.3)</td>
<td>73.2 (13.3)</td>
<td>86.6 (19.4)</td>
<td>77.5 (13.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ARL</td>
<td>72.1 (20.0)</td>
<td>63.9 (13.2)</td>
<td>80.9 (20.3)</td>
<td>70.4 (13.7)</td>
<td>67.2 (9.5)</td>
<td>72.8 (13.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>83.2 (20.0)</td>
<td>75.4 (13.2)</td>
<td>88.9 (19.5)</td>
<td>74.4 (13.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>83.2 (20.4)</td>
<td>72.7 (14.8)</td>
<td>87.0 (21.3)</td>
<td>74.2 (14.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>80.5 (19.4)</td>
<td>70.4 (13.9)</td>
<td>86.7 (20.0)</td>
<td>76.3 (13.8)</td>
<td>68.8 (9.8)</td>
<td>74.1 (12.7)</td>
<td></td>
</tr>
</tbody>
</table>

† G, variance–covariance matrix used to model the relationship between genotypes. K, realized additive relationship matrix estimated using molecular markers; I, identity matrix.
‡ ME, mega-environment. ME1: Lancaster (LAN), Madison (MAD), and Sturgeon Bay (STB). ME2: Arlington (ARL), Marshfield (MAR), and Spooner (SPO).
§ PREP, partially replicated design with 48 un-replicated genotypes and six replicated checks within each location; RCBD, the same as the RCBD strategy but modeling the genotype × environment interaction (GEI) variance covariance matrix with an unstructured model; PREP, the same as the PREP strategy but modeling the GEI variance covariance matrix as an unstructured model; Extreme MED, mega-environmental design with 144 unreplicated genotypes across the ME and six replicated genotypes in each location within each mega-environment; Moderate MED, mega-environmental design with 72 unreplicated genotypes across the ME, eight genotypes replicated three times in each location within the ME, and six replicated genotypes twice in each location within each ME.
to selection than the classical RCBD approach when the proportion of selected individuals was 0.2 (Table 3). Larger responses to selection are obtained with larger selection intensities (i.e., smaller proportion of genotypes selected), but the larger advantage of MED experiments can be seen for smaller selection intensities.

**DISCUSSION**

Our study allowed us to evaluate and compare resource allocation strategies within a plant breeding or testing program from a micro- and macro-environmental variability perspective. When working with quantitative traits, the presence of environmental variability directly affects the ability to detect superior genotypes (Cobb et al., 2013). Therefore, identifying optimal experimental designs for large number of genotypes that effectively controls micro-environmental variability is challenging (Cullis et al., 1998). Moreover, the presence of GEI when genotypes are tested in METs is one of the most important factors affecting quantitative traits of agricultural importance, challenging the estimation of genotypic effects (Mathews et al., 2008). We proposed a new strategy to allocate resources in METs accounting for spatial variability and GEI modeling MEs and creating designed unbalanced experiments. The MED was the superior strategy in response to selection, improving the response of the RCBD to above 40%. However, the use of strategies such as RCBD or PREP outperform both MED strategies in terms of the correlation between true and predicted genotypic effects and in the recovery of the superior genotypes.

The choice of experimental design and spatial correction determines the ability to find superior genotypes in a study. The ALPHA design was the best option regardless of experimental size. These results are similar to those found by other studies (Qiao et al., 2000; Casler, 2015; Borges et al., 2019). This confirms the idea that designs such as the ALPHA design are beneficial when the number of treatments to be evaluated is greater and spatial variability is present in the field. In terms of experimental size, our results suggest that larger local control is increasingly beneficial with larger sizes of the experiment. This is consistent with other studies that indicate that the larger the size of the experiment, the lower the effectiveness of the simpler models (Casler, 2015; Borges et al., 2019). Additionally, the use of spatial correction modeling showed significant improvements for all the evaluated indicators in most of the scenarios. With the increase in the experimental size, larger improvements in the use of spatial correction were detected in comparison with the strategies without spatial corrections. This was also shown by Casler (2015). The use of S2D spatial correction was the best option in most of the scenarios evaluated. The S2D is probably being able to better compensate for some of the local control of more advanced experimental designs. The flexibility of this type of spatial adjustment has been positively assessed in other studies where spatial variability patterns were modeled (Stefanova et al., 2009: Velazco et al., 2017; van Eeuwijk et al., 2018). Finally, the ranking of the experimental designs was consistent across different levels of noise that generate different heritabilities (i.e., from 0.58 to 0.86; Supplemental Fig. S4).

**Table 3. Mean and standard error of the expected response to selection for grain yield for all the resource allocation strategies evaluated with different proportions of selected individuals. The gain from each method was expressed as a percentage of the selection response above the response of the randomized complete block design (RCBD) strategy [\( R_{\text{RCBD}} \)], where \( R \) is the response of each strategy (i.e., partially replicated [PREP], randomized complete block design modeling the genotype \( \times \) environment interaction correlation [PREPGE], and the extreme or moderate mega-environmental design [MED]), and \( R_{\text{RCBD}} \) is the selection response of the balanced RCBD strategy. Response to selection was evaluated for three sizes of populations where the number of selected individuals corresponds to a selection intensity of 0.05, 0.10, and 0.20 in the RCBD.**

<table>
<thead>
<tr>
<th>Proportion of selected in RCBD</th>
<th>G†</th>
<th>( R_{\text{RCBD}} )‡</th>
<th>PREP</th>
<th>RCBD(_{\text{GE}})</th>
<th>PREP(_{\text{GE}})</th>
<th>Extreme MED</th>
<th>Moderate MED</th>
</tr>
</thead>
<tbody>
<tr>
<td>kg ha(^{-1})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>K</td>
<td>1030.1</td>
<td>−0.1 (3.0)</td>
<td>1.0 (3.0)</td>
<td>4.6 (4.3)</td>
<td>−12.8 (4.2)</td>
<td>13.4 (5.2)</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>999.0</td>
<td>13.0 (4.1)</td>
<td>0.0 (1.2)</td>
<td>14.2 (1.6)</td>
<td>−4.0 (2.3)</td>
<td>21.4 (3.4)</td>
</tr>
<tr>
<td>0.10</td>
<td>K</td>
<td>864.3</td>
<td>4.1 (2.7)</td>
<td>0.3 (3.2)</td>
<td>9.1 (2.2)</td>
<td>−7.2 (3.1)</td>
<td>20.0 (4.3)</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>854.2</td>
<td>17.9 (6.4)</td>
<td>1.1 (3.2)</td>
<td>19.1 (3.4)</td>
<td>2.2 (1.6)</td>
<td>28.4 (5.1)</td>
</tr>
<tr>
<td>0.20</td>
<td>K</td>
<td>688.4</td>
<td>15.3 (4.4)</td>
<td>1.1 (2.6)</td>
<td>17.3 (4.1)</td>
<td>4.7 (2.4)</td>
<td>33.6 (4.3)</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>683.4</td>
<td>27.4 (5.5)</td>
<td>0.0 (2.9)</td>
<td>28.1 (5.2)</td>
<td>15.4 (3.2)</td>
<td>43.1 (6.5)</td>
</tr>
</tbody>
</table>

† G, variance–covariance matrix used to model the relationship between genotypes. K, realized additive relationship matrix estimated using molecular markers; I, identity matrix.

‡ \( R_{\text{RCBD}} \): response to selection for the RCBD with 20 genotypes replicated three times evaluated in each one of six location.

§ Percentage of the response to selection above the response of the RCBD strategy for the partially replicated design with 48 unreplicated genotypes and six replicated checks within each location (PREP), the same than the RCBD strategy but modeling the genotype \( \times \) environment interaction (GEI) variance covariance matrix with a factor analytic model (RCBD\(_{\text{GE}}\)) the same as the PREP strategy but modeling the GEI variance covariance matrix as a factor analytic model (PREP\(_{\text{GE}}\)) mega-environmental design with 144 unreplicated genotypes across the mega-environment and six replicated genotypes in each location within each mega-environment (Extreme MED), mega-environmental design with 72 unreplicated genotypes across the mega-environment, eight genotypes replicated three times in each location within the mega-environment, and six replicated genotypes twice in each location within each mega-environment (Moderate MED).
We identified two historical MEs across the WOBP MET using the real historical dataset. Several studies have shown that the use of GEI information through the delimitation of MEs based on winning genotypes was useful for phenotypic selection (Gauch and Zobel, 1997; Dardanelli et al., 2006; Yan, 2015) and increased prediction accuracy in genomic selection studies (Burgueño et al., 2012; Heslot et al., 2014; Lopez Cruz et al., 2015; Lado et al., 2016). Most of these studies modeled GEI as a posterior analysis (Gauch and Zobel, 1997; Yan, 2015); however, we proposed a new approach, where GEI is used to allocate testing resources more efficiently. This assumes that testing is limited by the phenotyping resources and not by the potential population size. The most efficient testing allocation strategy in our study was the MED with purposefully unbalanced designs across locations. Endelman et al. (2014) found that prediction accuracy was higher when unbalanced design spread across locations was used, and a broader sampling of environments was tested in comparison with evaluating all entries in one location. Möehring et al. (2014) concluded that with limited amount of resources, the use of augmented design with unreplicated entries allowed connecting environments, and outperformed replicated and classical augmented designs in terms of prediction accuracy. When strategies were compared in terms of COR and BEST in our study, larger values were obtained through the use of RCBD and PREP, mainly explained by the higher level of replication in the MET. On the other hand, our MED strategy was the best experimental design strategy in terms of response to selection. In particular, the best option is the use of the MED strategy when a moderate level of unreplicated genotypes was used across location within a ME (MED_60). This is probably due to a combination of larger population sizes and higher spatial variability control with some repeated genotypes. The ranking of the strategies was consistent across different levels of noise that generate different heritabilities (Supplemental Table S4). Because larger population sizes can be used in the moderate MED, this experimental design is especially suited for plant breeders where a large number of genotypes have to be screened every single year in the most efficient way and the focus is in genetic gain. The most likely reason for the failure of the extreme MED is the high spatial heterogeneity present within locations in our uniformity trials. We believe that the level of spatial variability present in these fields is common of the US Midwest. Different levels of random field heterogeneity that affect the trait heritability were evaluated to determine changes in terms of COR among strategies (Supplemental Table S4), and the performance and ranking were similar when the heritability was smaller.

The realized additive relationship matrix had better performance in strategies with moderate or high level of replication. However, when the proportion of unreplicated genotypes increased as in the extreme MED or PREP, the use of the realized additive relationship matrix did not improve the correlation between the true and the predicted genotypic values but decreased them. The low level of replication of both designs could be somewhat responsible for the poor estimations of additive effects (Endelman et al., 2014). Other nonadditive effects such as epistasis could also be causing this behavior (Möhering et al., 2014). We evaluated the effect of modeling nonadditive effects for all the strategies. For strategies where a low level of replication for most genotypes was used such as the PREP or the extreme MED, the modeling of the nonadditive effects outperformed models that only use additive effects (Supplemental Fig. S5). For the comparison of additive and nonadditive effects, Eq. [9] and Eq. [10] were used and the variance–covariance matrix among genotypes was modeled with either an additive relationship matrix, or with the additive relationship matrix, a dominance relationship matrix, and an additive by additive relationship matrix. Because our genotypes are inbred lines, the nonadditive component is mainly explained by additive by additive epistatic interactions. However, because we did not use the relationship matrix to estimate the phenotypic means that were used in all our simulations as our measurement of the “true” genotypic values, we do not have a fair comparison of the performance of the realized additive relationship matrix against true genotypic values. What we observe might be the result of our genotypes shrinking towards zero in our estimations of the “true” genotypic means and shrinking towards their relatives when evaluated with the K model. One of the strengths of our study is that simulations were performed using real field spatial variability, GEI structure, and genotypic effects. Therefore, this approach provides a great resource to create new strategies for designing large field trials. One of the limitations is that optimal designs are a function of the level of spatial heterogeneity, genetic variance, and GEI structure. We believe that the level of spatial variability (Supplemental Fig. S6) and GEI structure (Supplemental Fig. S7, Supplemental Table S3) found in our study reflects typical research station field variability and GEI structure. Additionally, the ranking of the strategies did not change for the COR (Supplemental Table S4). On the other hand, we only modeled more noise to our dataset and did not have less spatial variability. Therefore, it would be interesting to have the MED evaluated for a large range of spatial variability and GEI structures.

In summary, using micro- and macro-environmental variability information in METs is crucial to improve the overall process of selecting superior genotypes. The use of an appropriate experimental design and spatial correction increases the ability to find differences between materials, mainly when large numbers of genotypes are evaluated. As showed here, the use of small balanced...
replicated experiments across locations limits the response to selection in plant breeding testing. As an alternative, the allocation of resources with a moderate level of unbalance across locations within MEs was very efficient in terms of selection of superior genotypes and response to selection. The use of MED approaches appears as a promising strategy when designing resource allocations in plant breeding programs.

**Supplemental Material**

Supplemental material is available online for this article.

**Conflict of Interest**

The authors declare that there is no conflict of interest.

**Acknowledgments**

This research was funded by the Hatch Act Formula Fund WISO1984. We would like to thank three anonymous reviewers that significantly improved the quality of the manuscript.

**References**


Bernardo, R. 2010. Breeding for quantitative traits in plants. Stella Press, Woodbury, MN.


