



HAL
open science

Optimization of multi-environment trials for genomic selection based on crop models

Renaud Rincent, Estelle Kuhn, Herve Monod, Francois-Xavier Oury, Monique Rousset, Vincent Allard, Jacques Le Gouis

► **To cite this version:**

Renaud Rincent, Estelle Kuhn, Herve Monod, Francois-Xavier Oury, Monique Rousset, et al.. Optimization of multi-environment trials for genomic selection based on crop models. TAG Theoretical and Applied Genetics, 2017, 130, pp.1-18. 10.1007/s00122-017-2922-4 . hal-01594849

HAL Id: hal-01594849

<https://hal.science/hal-01594849>

Submitted on 22 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Optimization of multi-environment trials for genomic selection based on crop models

R. Rincent^{1,2}  · E. Kuhn³ · H. Monod³ · F.-X. Oury^{1,2} · M. Rousset^{1,2} · V. Allard^{1,2} · J. Le Gouis^{1,2}

Received: 23 January 2017 / Accepted: 11 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Key message We propose a statistical criterion to optimize multi-environment trials to predict genotype × environment interactions more efficiently, by combining crop growth models and genomic selection models.

Abstract Genotype × environment interactions (GEI) are common in plant multi-environment trials (METs). In this context, models developed for genomic selection (GS) that refers to the use of genome-wide information for predicting breeding values of selection candidates need to be adapted. One promising way to increase prediction accuracy in various environments is to combine ecophysiological and genetic modelling thanks to crop growth models (CGM) incorporating genetic parameters. The efficiency of this approach relies on the quality of the parameter estimates, which depends on the environments composing this MET used for calibration. The objective of this study was to determine a method to optimize the set of environments composing the MET for

estimating genetic parameters in this context. A criterion called OptiMET was defined to this aim, and was evaluated on simulated and real data, with the example of wheat phenology. The MET defined with OptiMET allowed estimating the genetic parameters with lower error, leading to higher QTL detection power and higher prediction accuracies. MET defined with OptiMET was on average more efficient than random MET composed of twice as many environments, in terms of quality of the parameter estimates. OptiMET is thus a valuable tool to determine optimal experimental conditions to best exploit MET and the phenotyping tools that are currently developed.

Introduction

In plant breeding, the best performing varieties are often different from one environment to another. These phenomena are called genotype × environment interactions (GEI). To cope with them, breeders repeatedly phenotype the same varieties in multi-environment trials (METs). However, this approach has economical limitations and screening all materials in all environments is not feasible. Therefore, it would be of great interest to develop models able to predict these interactions. One promising tool to reach this goal is genomic selection (GS), which is a method used in animal and plant breeding to predict genomic breeding values using genome-wide molecular markers (Whittaker et al. 2000; Meuwissen et al. 2001). In a few recent studies, it was proposed to adapt the reference GS models to the GEI context by attributing environment specific effects to the markers (Schulz-Streeck et al. 2013; Crossa et al. 2015), or by modelling environmental covariances (Burgueño et al. 2012). In other studies, environmental covariates were introduced in the GS model (Heslot et al. 2014; Jarquín

Communicated by Jose Crossa.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-017-2922-4) contains supplementary material, which is available to authorized users.

✉ R. Rincent
renaud.rincent@inra.fr

- ¹ INRA, UMR 1095 Génétique, Diversité et Ecophysiologie des Céréales, 5 chemin de Beaulieu, 63100 Clermont-Ferrand, France
- ² Université Blaise Pascal, UMR 1095 Génétique, Diversité et Ecophysiologie des Céréales, 63178 Aubière Cedex, France
- ³ INRA, MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France

et al. 2014; Malosetti et al. 2016), which allows predicting in new environments. However, the gain obtained with these models is limited. One likely reason is that the GEI is, in most cases, reduced to linear relationships between varieties and a few environmental covariates, and this cannot allow for the complex interactions between plant development and the environmental conditions.

The way plants interact with the environment has long been the subject of refined analyses by ecophysicologists. Their research has allowed to develop crop growth models (CGM) which describe plant development using mechanistic relationships with physiological parameters and environmental covariates as inputs. By definition, the physiological parameters are independent from the environment, but some of them, called the genetic parameters, may depend on the variety. For example, the sensitivity to photoperiod of a given variety is the same for any environment. However, photoperiod can vary from one environment to another which generates GEI, because other varieties can have different photoperiod sensitivities. CGM can be used to predict GEI, since they integrate explicitly both variety characteristics (genetic parameters) and environmental covariates (Chapman et al. 2002; Hammer et al. 2002; Bertin et al. 2010; Bustos-Korts et al. 2016). Once the genetic parameters have been estimated, their genetic architecture can be determined and GS models can be calibrated. The GS model can then be used to predict the genetic parameters of other varieties. These predicted genetic parameters can also be used to predict integrative traits such as yield for these new varieties in new environments by running the CGM (Fig. 1). The interest and feasibility of this approach coupling CGM and genetics

have been validated for leaf elongation rate in maize (Reymond et al. 2003; Chenu et al. 2008), fruit quality (Quilot et al. 2005; Prudent et al. 2011), and phenology of various species (White and Hoogenboom 1996; Yin 2005; Nakagawa et al. 2005; Messina et al. 2006; White et al. 2008; Uptmoor et al. 2011; Zheng et al. 2013; Bogard et al. 2014; Onogi et al. 2016). Recently, Technow et al. (2015), Cooper et al. (2016), and Messina et al. (2017) have illustrated the interest of coupling CGM and GS models for predicting and selecting highly integrated traits such as grain yield. One major advantage of their approach and the approach of Onogi et al. (2016) is that the genetic parameters and the marker effects are jointly estimated, and so information can be shared between individuals thanks to the genotypic data.

The approach coupling CGM and marker-assisted selection (CGM-MAS) is also called gene-based modelling or QTL-based modelling (see Fig. 1). In CGM-MAS, the efficiency of genome-wide association studies (GWAS) and GS for the genetic parameters depends on the composition of the calibration set, the relevance of the crop model, and the quality of the parameter estimates. Considering the number of genetic parameters involved in crop models and the way they are entangled in complex processes, it is quite clear that huge amounts of data are required to estimate parameters and that the high-throughput phenotyping tools under development such as drones (UAV) and phenotyping platforms will considerably help.

Some genetic parameters can be estimated almost directly by measuring simple traits on phenotyping platforms (Reymond et al. 2003; Yin 2005). The other genetic parameters are estimated by adjusting the CGM outputs to

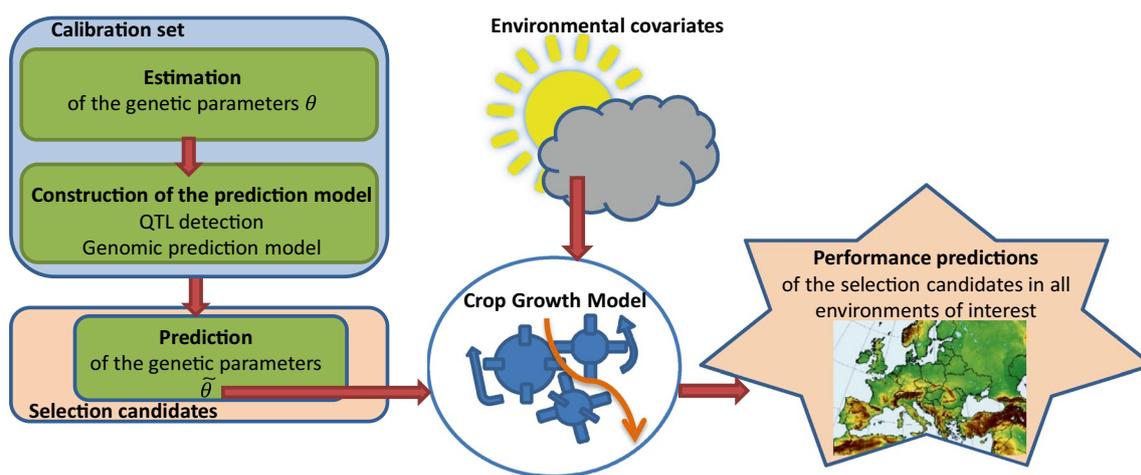


Fig. 1 Schematic representation of the CGM-MAS approach. The CGM predicts the performance of the selection candidates in various environments using the (predicted) genetic parameters and the environmental covariates as inputs. The genetic parameters of the calibra-

tion set are estimated and used to calibrate a prediction equation (GS models). This equation can then be used to predict the genetic parameters of other genotyped individuals (the selection candidates)

the observations of more integrative traits. In this case, the inference of the parameters can be done thanks to brute-force algorithm (Bogard et al. 2014), more sophisticated exploration algorithms (Wallach et al. 2011; Klein et al. 2012) or Bayesian inference (Makowski et al. 2002; Van Oijen et al. 2005; Iizumi et al. 2009; Dumont et al. 2014). The quality of the parameter estimates highly conditions QTL detection power (Wang 2008; Teyssèdre et al. 2012; Rincent et al. 2014) and GS accuracy (Daetwyler et al. 2008; VanRaden 2008) through the error variance.

A fundamental issue for data collection and parameter inference is the choice of the experimental design. Most literature on the design of experiments in plant breeding concentrates on the within-trial allocation of varieties to plots (Piepho and Williams 2006; Butler et al. 2014). When designing METs to estimate genetic parameters, however, the major question is to determine in which pedoclimatic conditions the varieties need to be phenotyped to provide the best estimates. For example, if day length plays a major role in the CGM behavior, one can expect that experimental designs capturing important variability of day lengths will be more efficient than others for estimating the corresponding genetic parameters. The importance of the definition of the experimental design to estimate model parameters has been discussed in some studies (Wöhling et al. 2013; Dumont et al. 2014). The definition of optimal MET design is thus a key point, which must be based on sound statistical approaches.

A few studies have tackled explicitly the design of MET (see Talbot Chapter 10 on resource allocation for selection systems in Kempton and Fox 1997). More recently, an interesting approach was proposed to optimize experimental designs to calibrate hydrological model (Leube et al. 2012). Developing such new tools in the context of CGM-MAS is a main current necessity of great interest.

The main objective of this study was to develop a statistical criterion to optimize the set of environments composing the MET before collecting data in the context of CGM-MAS. The designs sampled with this criterion should allow for the most efficient calibration of the crop model for a whole collection of varieties. The high quality of the genetic parameter estimates obtained with the optimal experimental design should in turn increase CGM-MAS efficiency (QTL detection power and prediction accuracy). These approaches were tested on wheat phenology (heading time), for which reference crop models exist (Jamieson et al. 1998b; Keating et al. 2003), using simulated and real data sets. This trait is influenced by temperatures and day length, so we focused on the definition of optimal combinations of locations and sowing dates. A Bayesian inference approach was used to estimate the genetic parameters of crop models using integrative phenotypes.

Materials and methods

Our objective is to determine an optimal set of environments (multi-environment trials, MET) for the estimation of genetic parameters. When the optimal MET must be determined, only the crop model and the possible sites are known. There is not yet any measurement available on the environmental covariates of the CGM for the year to come, but we consider that the measurements in the past years can be used to approximate them.

Formally, we consider that I genotypes have to be phenotyped in Z environments with K replications per environment. We denote by \mathcal{E} the collection of all the environments considered in the study: $\mathcal{E} = \{E_j, 1 \leq j \leq J\}$, where E_j is the vector of environmental covariates of environment j required for the crop model to run, and J is the total number of possible environments. We define an MET of size Z as a subset of \mathcal{E} composed of Z environments. For a given MET d , we denote by E_d the joint vector of the environmental values E_j , $j \in \mathcal{J}_d$, \mathcal{J}_d being the set of indices of the environments composing it.

Statistical model

We assume that the observations are the sum of the crop model output and an error term. Thus, the statistical model is:

$$Y_{ijk} = f(\theta_i, E_j) + e_{ijk}, \text{ for } 1 \leq i \leq I, 1 \leq j \leq J \text{ and } 1 \leq k \leq K, \quad (1)$$

where Y_{ijk} is the scalar value of genotype i in environment j and replication k , f is the function corresponding to the crop model (usually non-linear), θ_i is the p -dimensional vector of genetic parameters for variety i , and e_{ijk} is an additive error term. The residuals (e_{ijk}) are assumed to be independent, normally distributed, centered with variance σ_e^2 for sake of simplicity, but heteroscedasticity and non-normal distributions could be used as well if required.

For a given MET d of size Z , we denote by Y_i^d the vector of output values Y_{ijk} for $1 \leq k \leq K$ and $j \in \mathcal{J}_d$.

In the present study, we estimate the parameters (θ_i , $1 \leq i \leq I$, σ_e^2) of model (1) by a Bayesian inference algorithm applied to the phenotypes collected in the MET d . Prior distributions are given low information levels: uniform distributions for θ_i with bounds defined thanks to literature or expert knowledge, and inverse Gamma distribution for σ_e^2 . Of course, if the MET is more complex than the one described here, one can adjust model (1) for a better inference adapted to these situations (for example by taking into account block effects, or by introducing heteroscedasticity as done in the present study for the real data set).

Definition of the criterion OptiMET used to optimize MET

The OptiMET criterion is inspired by optimal Bayesian design (see Atkinson and Donev 1992) and adapted from the study of Leube et al. (2012) in the context of Bayesian model averaging for hydrological models. Our objective is to define a relevant MET which is able to differentiate both between two different parameter values leading to two different observation values and between two different observation values corresponding to two different parameter values. Therefore, we require at least that the MET is built, such that the crop model generates distant outputs for distant genetic parameter vectors. We will propose a criterion to determine a set of environments (MET) satisfying this condition.

Since we do not know the values of parameter vector θ when determining the MET, we will consider a huge finite number of possible a priori values, standing for the parameter vector distribution across varieties. Therefore, we consider a sample of size m denoted by $(\theta_u)_{1 \leq u \leq m}$ which is a finite size representation of the possible continuous distribution of the parameters. These m genetic parameter vectors can be chosen based on expert knowledge or on literature (it is most of the time possible to define at least lower and upper bounds). The distance between any two parameter vectors θ_u and θ_v is defined by:

$$\text{dist}(\theta_u, \theta_v) = \left[\sum_{s=1}^p \left(\frac{\theta_{us} - \theta_{vs}}{M_s - m_s} \right)^2 \right]^{1/2},$$

where θ_{us} and θ_{vs} are the s th component of θ_u and θ_v , respectively, and M_s and m_s are the maximal and minimal value for the s th component of θ determined by expert knowledge or using the literature.

For a given candidate MET d , we denote by L^d the matrix of size $m \times m$, in which the element L_{uv}^d is computed following Leube et al. (2012):

$$L_{uv}^d = \frac{1}{(4\pi\sigma_e^2)^{n_y/2}} \exp\left(-\frac{1}{4\sigma_e^2} (\Delta_{uv}^d)^t \Delta_{uv}^d\right),$$

where n_y is the number of observations for a given variety ($n_y = Z \times K$), and $\Delta_{uv}^d = (f(\theta_u, E_j) - f(\theta_v, E_j))_{j \in \mathcal{J}_d}$. The quantity L_{uv}^d corresponds to the likelihood of the parameter vector θ_u given the synthetic noise-free data $(f(\theta_v, E_j))_{j \in \mathcal{J}_d}$ (for more details, see Leube et al. 2012, Appendix B).

The matrix (L_{uv}^d) is normalized by computing the weight matrix $W_{uv}^d = \frac{L_{uv}^d}{\sum_u L_{uv}^d}$.

We define the value of the criterion OptiMET for a given MET d by: $\text{OptiMET}^d = \sum_{u,v=1}^m (\text{dist}(\theta_u, \theta_v) \times W_{uv}^d)$.

The optimal design denoted by d_{opt} is the one that minimizes OptiMET. Indeed minimizing OptiMET results in maximizing the distance between the outputs of the CGM for two genetic parameters vectors that are distant, i.e., minimizing the corresponding coefficient in the weight matrix W .

Case study: MET optimization for the estimation of Sirius CGM phenology parameters

Many strategies to sample MET (combinations of locations and sowing dates in this case study) can be compared (Fig. 2, box 1). In this paper, we concentrate on three of them: random sampling, OptiMET-optimal sampling, and sampling based on expert knowledge. Wheat heading time was used as a case study. This trait is key to plant adaptation to new environments, and it has been intensively studied and modelled.

To evaluate the efficiency of OptiMET to optimize the composition of MET, we have tested it both with simulations and real data. In the simulation part, we considered that phenotypes were generated according to model (1). In a second part, OptiMET was tested on a real data set, to evaluate its robustness.

Sirius crop model

Sirius is a reference crop model to simulate wheat development (Jamieson et al. 1998b). Its relevance to simulate accurately the development of crops was validated in wide range of conditions including Europe, New Zealand, Australia, and USA (Semenov et al. 1996; Jamieson et al. 1998a, b; Jamieson and Munro 2000; Jamieson and Semenov 2000; Brooks et al. 2001). The phenology model is described in He et al. (2012). Briefly, the development of wheat from sowing to heading is modelled in three phases. The first phase, from sowing to emergence, is simulated as a fixed thermal time duration. In a second phase, from crop emergence to flag leaf appearance, flag leaf appearance successively integrates the effects of vernalization and photoperiod coupled with the rate of leaf emission (phyllochron). The last phase, from leaf ligule appearance to heading, is purely proportional to the phyllochron.

It has been shown that the processes of leaf appearance rate, and sensitivity to vernalization and to photoperiod have important genetic variability and strongly influence flowering time (He et al. 2012; Martre et al. 2015a). For these reasons, we defined as genetic parameters θ the three main parameters ($p = 3$) driving these processes: the response of vernalization rate to temperature (VAI), the day length response of leaf production (SLDL), and the

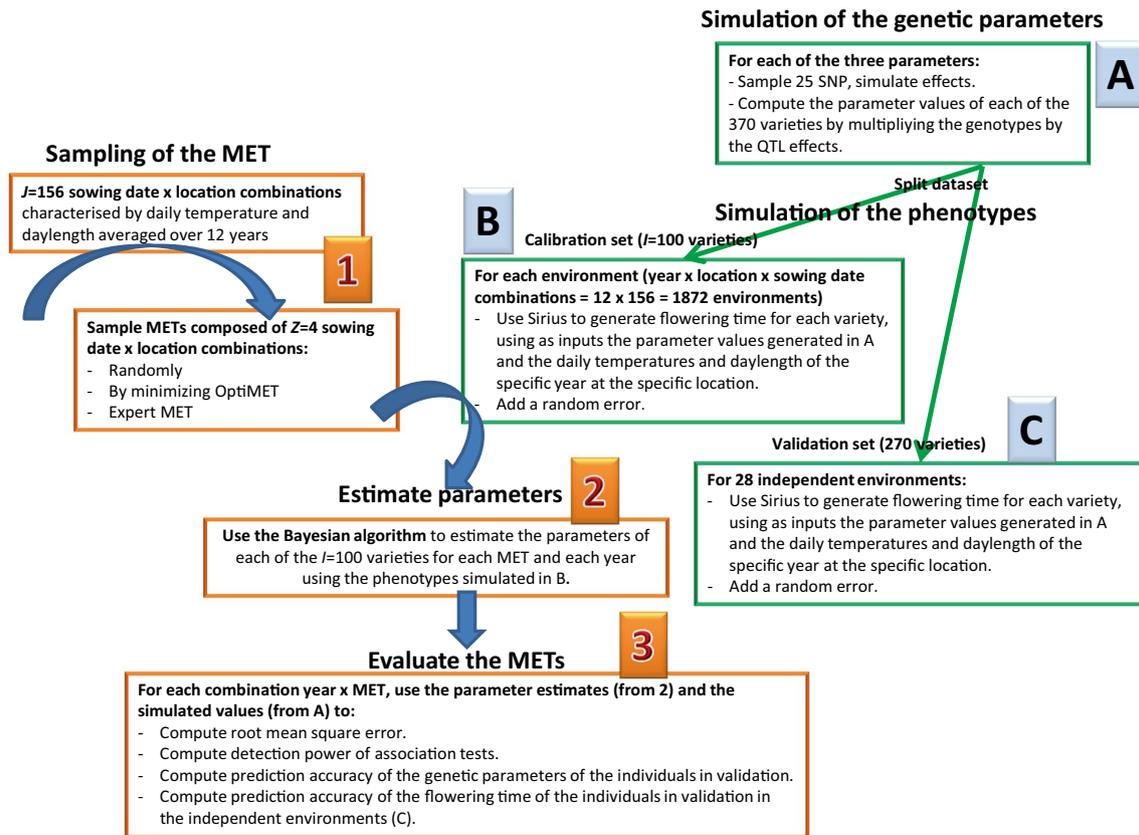


Fig. 2 Process used to compare by simulation the efficiency of different METs. Different MET sampling strategies are compared and evaluated for their efficiency to estimate the genetic parameters,

detect QTL in the calibration set, and predict performance of the validation set in independent environments

phyllochron (*Phyl*). In addition to these three parameters, the model requires daily average temperature and day length to run.

Computation of OptiMET

When computing OptiMET to determine an optimal MET (combination of sowing dates and locations), the climatic conditions that will occur at each location during the experiment are unknown (except day length). However, as daily temperatures are required to compute OptiMET, it was approximated by the average of daily temperature across a number of years sufficient to get stable averages. This virtual year characterized by daily temperatures averaged over 12 years is further referred to as the “average year”.

To compute OptiMET, we discretized each parameter interval into ten regularly spaced values and used the $m = 10^3 = 1000$ combinations for the genetic parameter vectors.

Parameter estimation using MCMC algorithm

The prior distributions of the genetic parameters (*VAI*, *SLDL*, and *Phyl*) were defined as uniform distributions with minimum and maximum fixed using knowledge of experts and found in the literature (He et al. 2012; Martre et al. 2015a):

$$VAI \sim U(0, 0.01),$$

$$SLDL \sim U(0, 1),$$

$$Phyl \sim U(80, 120).$$

The prior distribution of σ_e^2 was defined as a non-informative inverse Gamma distribution (with shape and scale parameters of 4 and 0.2, respectively). For the simulation study, the same residual variance σ_e^2 was attributed to all environments. For the study on real data, the residual variances were specific to each environment to model heteroscedasticity.

To generate the posterior distributions, we have used as MCMC algorithm a hybrid Gibbs sampler by block

which updates in turn three coordinates at a time for θ_i (the parameter values of VAI, SLDL, and Phyl of each variety in turn) and then σ_e^2 , through a Metropolis–Hastings step using as proposal a Gaussian distribution centered on the previous value of the chain. 20,000 iterations were generated and the first 1000 were discarded (burn-in). Parameter estimates were defined as the mode of the posterior distributions. All scripts were written in R 2.14.0 and can be made available upon request.

Evaluation of the criterion OptiMET using simulations

Procedure overview

The procedure used to compare the efficiency of different MET to estimate the genetic parameters is illustrated on Fig. 2. The objective is to evaluate the efficiency of OptiMET to optimize MET (combinations of locations and sowing dates) for the estimation of the three phenological parameters. Real genotypes were used to simulate QTL (Fig. 2, box A) for each genetic parameter. The parameter values were then computed for each variety. The varieties were split into two data sets: a calibration set and a validation set. The crop model Sirius was then used to generate phenotypes for the individuals of the calibration set in each considered environment for various years (Fig. 2, box B). In parallel to this, various strategies were used to sample MET of a given size among all the possible environments (Fig. 2, box 1). These strategies were: random sampling, sampling by minimizing OptiMET, and choosing an “expert MET” based on expert knowledge. The phenotypes generated for each of these MET for each specific year were used to estimate the parameters for each specific year independently (Fig. 2, box 2). For each MET and each year, we thus obtained parameter estimates. To evaluate the estimation efficiency of each MET, for each specific year, we computed root-mean-square errors (RMSE) of these estimates (as the true parameter values are known in simulation setting), detection power of association tests, and prediction accuracy of the parameter values of the individuals in validation (Fig. 2, box 3). Finally, for each MET and each specific year, we used the parameter predictions of the individuals in validation to predict using the CGM their heading time in independent environments representing the variability of French wheat production environments (as defined below) and computed the corresponding heading time prediction accuracies. The efficiency of the different MET was computed for each specific year independently to evaluate the stability of the different MET sampling strategies over years. It was not tried here to combine different years in a same MET, but this would be in practice possible.

Environments

In this section based on simulations, MET were composed of four location \times sowing date combinations ($Z = 4$) and sampled among 156 possible sowing date \times location combinations ($J = 156$). These 156 combinations are composed of 39 locations (supplementary information Fig. S1) spread in France combined with four sowing dates including three winter sowing dates (15th September, 15th October, 15th November) and one spring sowing date (15th March).

Twenty-eight independent environments were used to validate heading time prediction accuracy of the validation set (Fig. 2, box C). These 28 environments are representative of usual wheat growing conditions in France (Agreste 2016). They are the combinations of seven locations (representing the main regions in France where wheat is grown, supplementary information Fig. S1), two sowing dates (15th October and 15th November) with the climatic conditions of two specific growing seasons (2010/2011 and 2011/2012).

Phenotype simulation

Simulations were based on the real genotypes of a panel of 370 accessions from the INRA bread wheat core collection which was defined to represent worldwide wheat diversity (Balfourier et al. 2007) and with a large variability of growth habit (Rousset et al. 2011). All these lines were genotyped with an Affymetrix Axiom 280 K SNP array developed in the frame of the BreedWheat project (Rimbert et al. in preparation). After filtering for quality and homozygosity, the genotypes consisted of 20,713 SNP with known genetic positions.

To simulate the genetic architecture of the three genetic parameters, 25 SNP were sampled independently for each parameter and defined as QTL. Their effects followed geometric series as defined in Lande and Thompson (1990). The QTL effects were then rescaled, such that the genetic parameters took values with biological relevance (i.e., in the ranges defined above). The 75 SNP defined as QTL were then removed from the data set. At this step, each variety was defined by a vector of three parameters. The 370 accessions were then split in two data sets: 100 randomly sampled composed the calibration set and the 270 remaining the validation set. The sampling of the calibration set was done only once because of the computational burden of the simulation procedure.

To simulate phenotypes of the 100 individuals composing the calibration set in each environment (sowing date \times location \times year = $4 \times 39 \times 12 = 1872$), Sirius was run for each variety with the environmental covariates of the specific year (daily temperature and day length) and the genetic parameter values (computed using the simulated QTL) as inputs. Residual errors were added to Sirius

outputs using a centered normal distribution with a standard deviation relevant to mimic real experimental conditions (2 days).

Multi-environment trials sampling

Different sampling approaches have been used:

- Random MET. For each year (from 2003 to 2014), 10 MET of size $Z = 4$ were randomly sampled with 0, 1, 2, 3, or 4 winter sowings (and as a consequence 4, 3, 2, 1, or 0 spring sowings). This resulted in 120 random MET in total.
- An “expert MET”. This MET was defined using the knowledge of experts with the objective of estimating the different earliness components. It is composed of three locations chosen to get a North–South gradient (Mons-en-Chaussée, Versailles and Clermont-Ferrand) and, as a result, a photoperiodic gradient, with a winter sowing date (15th November) at each location, and an additional spring sowing date in Clermont-Ferrand (15th March). This combination of a winter and a spring sowing at the same location is supposed to capture efficiently the effect of vernalization.
- The MET minimizing OptiMET. To compute OptiMET, we considered environmental covariates (daily temperature and day length) averaged over 12 years (from 2003 to 2014). Considering the huge amount of environments, it was not possible to determine analytically the MET d_{opt} minimizing OptiMET. For this reason, we used an exchange algorithm: at each step, the random exchange of one environment included in the MET with one environment excluded was accepted if OptiMET decreased and was rejected otherwise. 3000 iterations were sufficient to reach a minimum, and we checked that the final MET was not a local optimum by running the exchange algorithm in parallel with four different initializations.

Evaluation of the efficiency of the METs

To compare the efficiency of each MET, the normalized root-mean-square errors (NRMSE) of the parameter estimates were computed for $1 \leq s \leq p$, for each MET and each year of experiment as follows:

$$\text{NRMSE}(\theta_s) = \left[\frac{1}{I} \sum_{i=1}^I \left(\frac{\hat{\theta}_{is} - \theta_{is}}{M_s - m_s} \right)^2 \right]^{1/2}, \text{ with } I = 100,$$

the number of genotypes, and where $\hat{\theta}_{is}$ is the estimate of parameter θ_{is} using the MET considered equal to the mode of the posterior distribution, and M_s and m_s are the maximal and minimal values of θ_{is} with $1 \leq i \leq I$.

We have also compared the METs efficiency (1) to detect QTL, and (2) to predict the parameters of independent varieties. For (1), for each of the three genetic parameters, a QTL was considered to be detected if at least one marker located at less than 1 cM from the simulated QTL was significantly associated (P value below a threshold of 0.05/25). The statistical model of Yu et al. (2006) with a random polygenic effect but no structure effect was used to test for associations. The covariance matrix of the random polygenic effect was estimated with the genotypic data (after removing the 75 SNP defined as QTL) using the estimator of VanRaden (2008). The detection power obtained with the different METs could then be compared. For (2), a classical G-BLUP model (Habier et al. 2008; Zhong et al. 2009) was used to predict the parameter values of the 270 individuals composing the validation set using the same kinship matrix than for the QTL detection. The prediction accuracy could then be computed as the correlation between predictions and simulated parameter values. Finally, these predicted parameter values could be used to predict heading time of the 270 varieties in the 28 independent environments using Sirius crop model (Fig. 2). Root-mean-square error (RMSE) and prediction accuracy of heading time were then computed to compare the MET. Prediction accuracy was defined as the correlation between predicted and simulated heading time.

Evaluation of OptiMET on a real data set

Description of the data set

To further evaluate the efficiency of OptiMET, a test on a real data set was also performed. In this data set, 121 varieties ($I = 121$) adapted to French environments with contrasted phenologies were phenotyped for heading date in 26 environments ($J = 26$, see Table 1) without replication ($K = 1$). These environments were in the western part or northern part of France, and sowing dates ranged from 17th of October to 14th of April.

MET sampling

METs of 4, 6, or 8 environments ($Z = 4, 6$ or 8) were sampled by minimizing OptiMET, or randomly, or with a reasoned strategy. In the reasoned strategy, we imposed that all sowing periods were represented in the MET. These four periods were October, November/December, January/February, and March/April. For each MET size (4, 6, and 8), we chose to sample 40 random trials and 40 reasoned trials. For the computation of OptiMET, environmental covariates were required. For this reason, we have used the daily

Table 1 Environments in which heading date of the 121 varieties was observed. In this real data set, 121 varieties were phenotyped for heading date in 26 environments with sowing dates ranging from October to April

2008/2009		2009/2010	
Sowing date	Location	Sowing date	Location
17/10/2008	Allonnes	23/10/2009	Mons-en-Chaussée
20/10/2008	Mons-en-Chaussée	28/10/2009	Clermont-Ferrand
20/10/2008	Le Moulon	28/10/2009	Louville
22/10/2008	Auchy	29/10/2009	Clermont-Ferrand
23/10/2008	Villiers-le-Bâcle	29/10/2009	Maule
29/10/2008	Montroy	29/10/2009	Caussade
12/11/2008	Clermont-Ferrand	30/10/2009	La Minière
20/11/2008	Clermont-Ferrand	25/11/2009	Villiers-le-Bâcle
12/12/2008	La Minière	14/12/2009	Clermont-Ferrand
24/12/2008	Mons-en-Chaussée	15/12/2009	Clermont-Ferrand
05/01/2009	Clermont-Ferrand	23/02/2010	Clermont-Ferrand
25/02/2009	Clermont-Ferrand	04/03/2010	Mons-en-Chaussée
16/03/2009	Mons-en-Chaussée		
14/04/2009	Mons-en-Chaussée		

average temperature of the closest meteorological station averaged over 11 years (2001–2014, excluding the years of experiment: 2008–2010).

Evaluation of the efficiency of the METs

A direct evaluation of the root-mean-square error of the parameter estimates was not possible in this case, because the real parameter values are unknown. For this reason, we have estimated the parameters using the 26 environments simultaneously using the Bayesian algorithm and used these estimates, denoted by θ_{is}^* for $1 \leq i \leq I = 121$ and $1 \leq s \leq p = 3$ as references. The use of these 26 environments to generate reference estimates seems reasonable as there are only three parameters, and the 26 environments are well adapted to estimate phenological parameters. Given these estimates, we can now consider criteria to evaluate the efficiency of the different METs to capture the information that is present in the whole data set.

Therefore, we considered two complementary criteria:

- First, the normalized root-mean-square error computed for $1 \leq s \leq p$ as:

$$\text{NRMSE}^*(\theta_s) = \left[\frac{1}{I} \sum_{i=1}^I \left(\frac{\hat{\theta}_{is} - \theta_{is}^*}{M_s^* - m_s^*} \right)^2 \right]^{1/2}, \quad \text{with}$$

$$I = 121, \text{ and where } \hat{\theta}_{is} \text{ is the estimate of parameter } \theta_{is} \text{ using the MET considered and equal to the mode}$$

of the posterior distribution, and M_s^* and m_s^* are the maximal and minimal values of the reference parameter estimates θ_{is}^* .

- Second, the normalized posterior square error (NPSE), defined for $1 \leq s \leq p$ as:

$$\text{NPSE}(\theta_s) = \mathbb{E}_\pi \left[\frac{1}{I} \sum_{i=1}^I \left(\frac{\theta_{is} - \theta_{is}^*}{M_s^* - m_s^*} \right)^2 \right],$$
 where π is the posterior distribution of θ_{is} conditionally to the data, \mathbb{E}_π the corresponding expectation, and $I = 121$. Indeed, since the whole posterior distribution is available in our Bayesian context, we can also evaluate the precision of the estimation through this integrated quantity which can be seen more like a variance. However, it is not possible to compute it analytically. Therefore, we calculated an empirical version using the last realizations of (θ_{is}) resulting from the Metropolis–Hastings (MH) algorithm. More precisely, let us denote by $(\theta_{is}^k)_{1 \leq k \leq K}$ the K last realizations of the MH algorithm for $1 \leq i \leq I$ and $1 \leq s \leq p$. We computed the following quantities $\frac{1}{I} \sum_{i=1}^I \left[\frac{1}{K} \sum_{k=1}^K \left(\frac{\theta_{is}^k - \theta_{is}^*}{M_s^* - m_s^*} \right)^2 \right]$ using $K = 19,000$.

For each MET, both criteria were averaged over the three parameters, leading to NRMSE^* and NPSE .

Results

Evaluation of the criterion OptiMET using simulations

Root-mean-square error of the parameter estimations obtained with the different METs

For each MET (location \times sowing date), the NRMSE of the model parameter estimates using the climatic conditions of each of the 12 years were computed (Fig. 3). The NRMSE of randomly sampled METs were highly variable for the three model parameters. For example, NRMSE of parameter Phyl ranged from 0.12 to 0.30 (which corresponds to RMSE of 4.6–12 degree-days) when the experiment was simulated with the climate of year 2013. NRMSE of the expert MET were sometimes higher and sometimes lower than those of the random METs. The MET sampled with OptiMET most of the time resulted in lower NRMSE than random and expert METs, with the exception of parameter SLDL in year 2008 and parameter VAI in years 2011 and 2012. The NRMSE reduction was higher for Phyl than for VAI and SLDL. The ratio between winter and spring sowings of the random MET influenced NRMSE. For VAI, random METs resulted in low NRMSE on average when

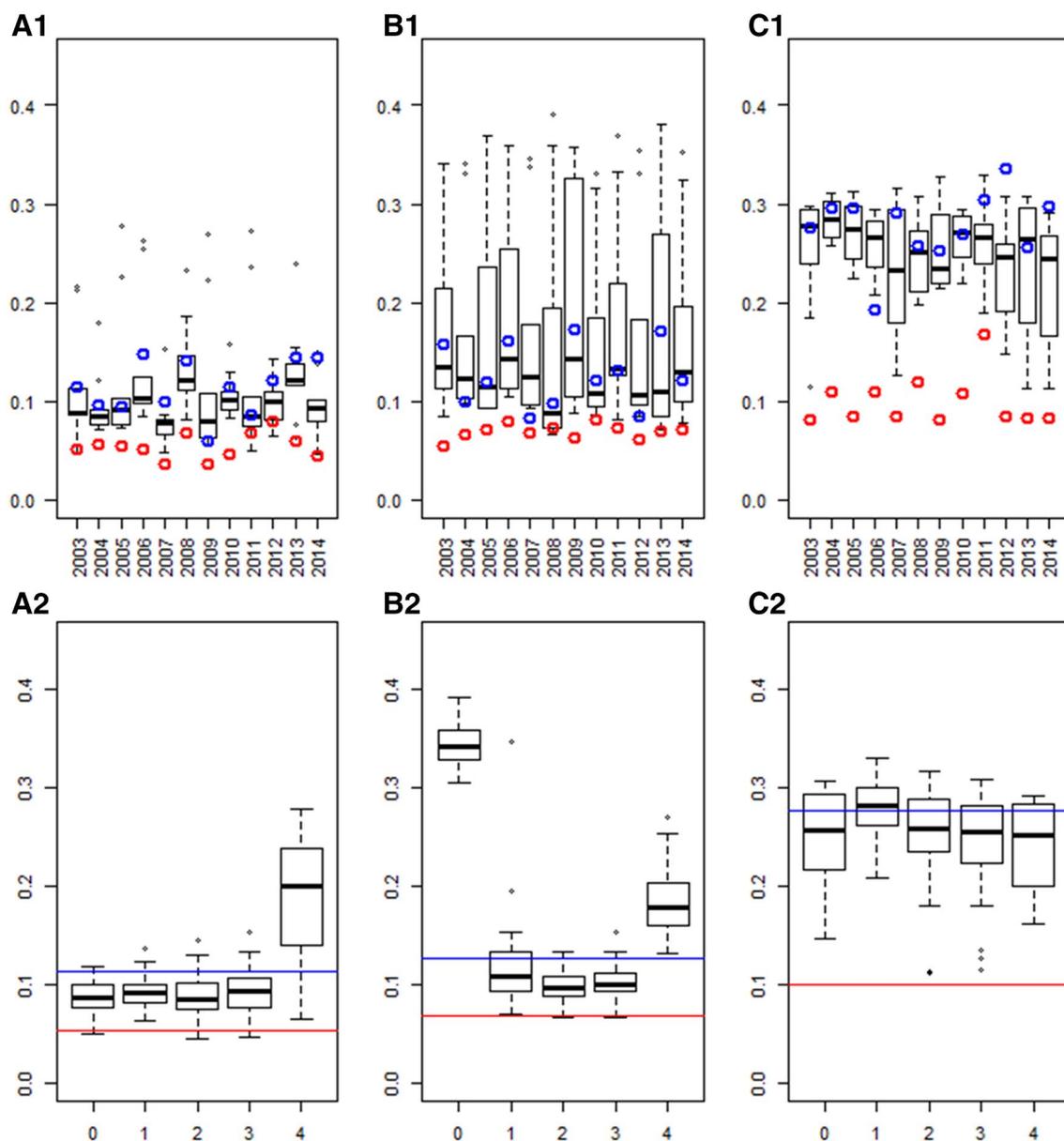


Fig. 3 Normalized root-mean-square error (NRMSE) of the parameter estimates of different METs: 120 randomly sampled (boxplots), one expert MET (blue points for specific years and blue lines for the average across years), and the OptiMET MET (red points for specific years

and red lines for the average across years). The results are presented for each of the 12 years of experiment (a1, b1, and c1), or for each of the random sampling strategy (a2, b2, and c2, number of winter sowings on the x-axis), for VAI a, SLDL b, and Phyl c (color figure online)

at least one spring sowing was sampled (Fig. 3a2). For SLDL, the best random METs were those with at least one sowing date of each period (winter and spring). The worst case was when four spring sowings were sampled (Fig. 3b2). For Phyl, there was no clear difference between the different options and all winter/spring sowings resulted in high NRMSE on average (Fig. 3c2). The OptiMET MET did better than the different winter/spring sowing combinations used for random sampling on average.

Table 2 Average QTL detection power (%) in the different METs. For each of the three parameters (VAI, SLDL, and Phyl), detection power was averaged over the 12 possible years of experiment for the OptiMET MET, the expert MET, and the 120 random METs. A QTL was considered to be detected if at least one SNP located at a maximum distance of 1 cM had a *p* value below the threshold 0.05/25

METs	VAI	SLDL	Phyl
OptiMET	16.7	18.0	10.7
Expert	13.0	11.7	4.2
Random	12.6	10.1	3.0

QTL detection power for the parameter estimates obtained with the different METs

The second criterion used to compare the different METs is the QTL detection power for the parameter estimates (Table 2). Detection power was low for all METs, with a maximum of 18% for SLDL, which corresponds to 4.5 detected QTL (among the 25 QTL simulated). Power was higher for VAI and SLDL (10.1–18.0%) than for Phyl (3.0–10.7%). On average, power was higher when using the OptiMET MET than the expert or random METs. These trends were consistent across years with the exception of 2007 for VAI, 2014 for SLDL, and 2003 and 2006 for Phyl where the expert MET did better than the OptiMET MET (Fig. 4).

Prediction accuracy of the parameter values of individuals in validation obtained with the different METs

The parameter estimates resulting from the different METs were then used to calibrate a GS model and predict the parameter values of the individuals in validation (Fig. 5). Prediction accuracies were variable between parameters, between years and between METs. On average, prediction accuracies were higher for SLDL and VAI than for Phyl. On average, the different random sampling

strategies resulted in similar accuracies for VAI and SLDL, except when the four winter sowings were sampled, which resulted in lower accuracies (Fig. 5a2, b2). On average, the OptiMET MET performed better than the other strategies, particularly for Phyl for which accuracy was on average multiplied by about three compared to expert and random METs.

Prediction accuracy and RMSE of heading time of individuals in validation in independent environments

The last step was to use the parameter predictions of the 270 individuals in validation to predict their heading time in 28 independent environments using the CGM. The prediction accuracies and RMSE of heading time resulting from the different METs used to estimate the parameters were computed (Fig. 6). Prediction accuracy was higher with the OptiMET MET (0.59) than with the expert (0.55) and random METs (0.35), on average. The gain brought by OptiMET varied greatly between years (in 2009, OptiMET performed much better than the expert MET, but similarly for year 2005). The best random sampling strategy was to sample four winter sowings (Fig. 6b), which can be explained by the absence of spring sowing in the validation data set composed of winter sowing environments only.

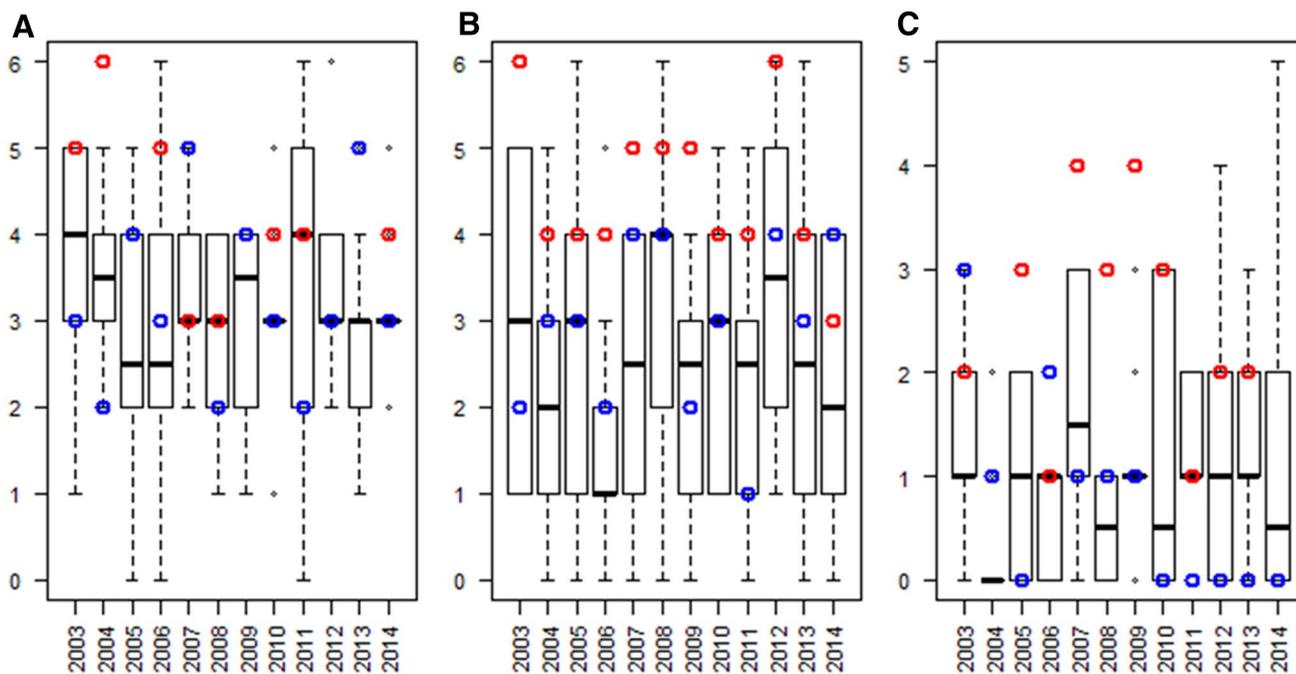


Fig. 4 Number of QTL detected in the different METs: 120 randomly sampled METs (boxplots), one expert MET (blue points), and the OptiMET MET (red points). The results are presented for each

of the three parameters: VAI **a**, SLDL **b**, and Phyl **c**, for each year of experiment (color figure online)

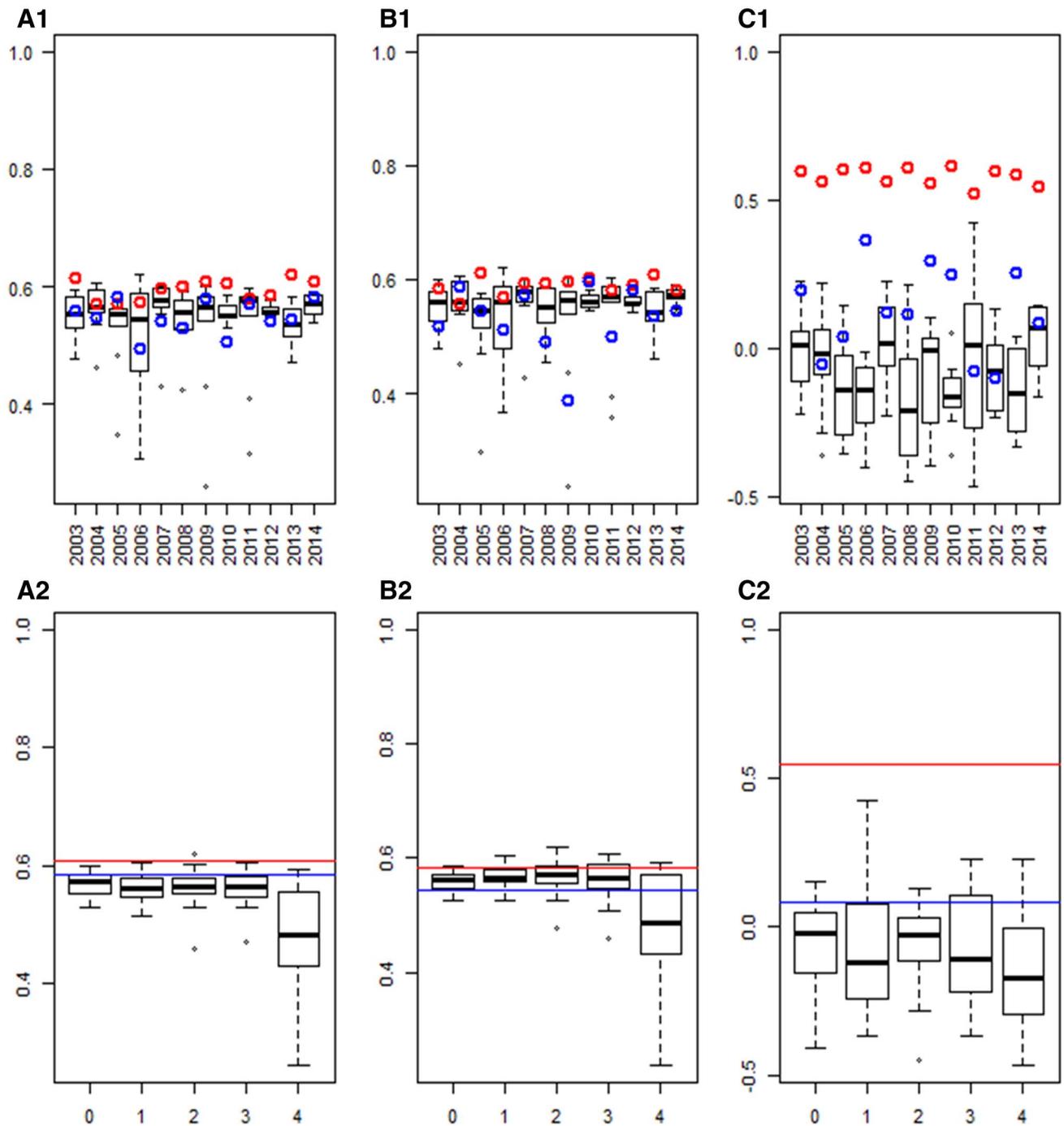


Fig. 5 Prediction accuracy of the parameter values of the individuals in validation. The GS model was calibrated with the parameter estimates resulting from the different METs: 120 randomly sampled METs (boxplots), one expert MET (blue points for specific years and blue lines for the average across years), and the OptiMET MET (red

points for specific years and red lines for the average across years). The results are presented for each of the 12 years of experiment (a1, b1, and c1), or for each of the random sampling strategy (a2, b2, and c2, number of winter sowings on the x axis), for VAI a, SLDL b, and Phyl c (color figure online)

The difference between the OptiMET and the expert MET was less pronounced when looking at the RMSE (Fig. 6c, d), but OptiMET performed better than the expert MET in 97% of the cases.

Composition of the OptiMET MET

The OptiMET MET (Fig. 7) is composed of two locations in the south of France (Alenya and San Giuliano)

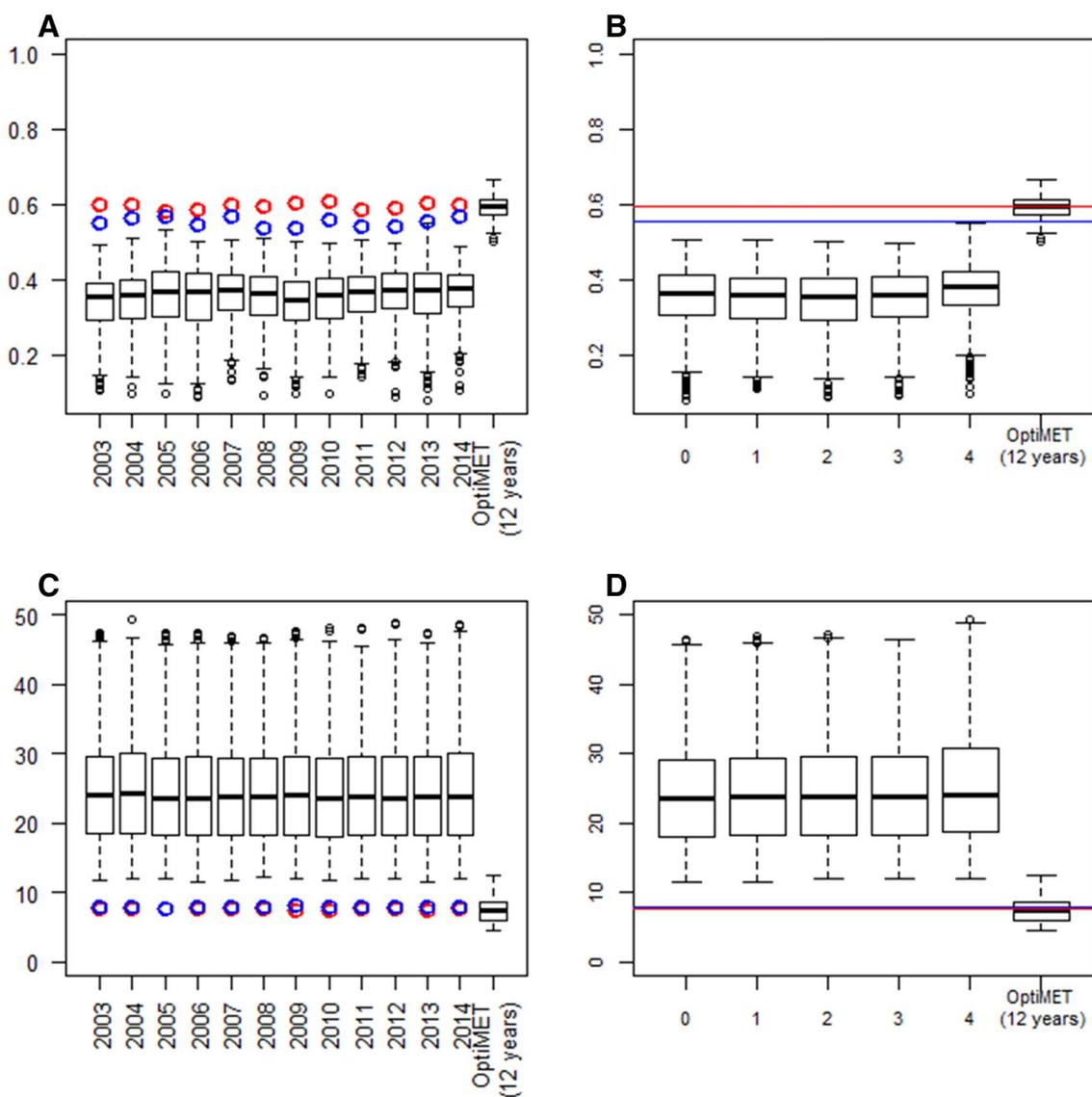


Fig. 6 Prediction accuracy (a, b) and root-mean-square error (RMSE, c, d) of heading time for the individuals in validation. Heading time was predicted using the predictions of the parameters of the individuals in validation as input for the crop model. The parameter estimates of the calibration set were obtained with different METs: 120 randomly sampled METs (*boxplots*), one expert MET (*blue*

points for specific years and *blue line* for the average across years), and the OptiMET MET (*red points* for specific years and *red line* for the average across years). The results are presented for each of the 12 years of experiment (a, c), or for each of the random sampling strategy (b, d; number of winter sowings on the *x axis*) (color figure online)

and one location in the west (Ploudaniel). There were three sowing dates in winter and one in spring. Alenya and San Giuliano are under Mediterranean conditions, whereas Ploudaniel is submitted to oceanic climate. The three locations have mild winters.

Evaluation of OptiMET on a real data set

The criteria ($\overline{\text{NRMSE}}^*$ and $\overline{\text{NPSE}}$) used to evaluate the different METs were highly variable for the random samples (Fig. 8). As expected, $\overline{\text{NRMSE}}^*$ and $\overline{\text{NPSE}}$

decreased when the size of the METs increased. The reasoned METs were on average more efficient than random METs. The OptiMET MET performed better than the random and reasoned METs on average, and performed similarly than the best random and reasoned METs. The difference of $\overline{\text{NRMSE}}^*$ or $\overline{\text{NPSE}}$ between the average of the random (and reasoned) METs and the OptiMET MET decreased when the size of the experimental design increased, which was expected, because by construction, the overlap between random and OptiMET METs increases with the size of the METs. According to both

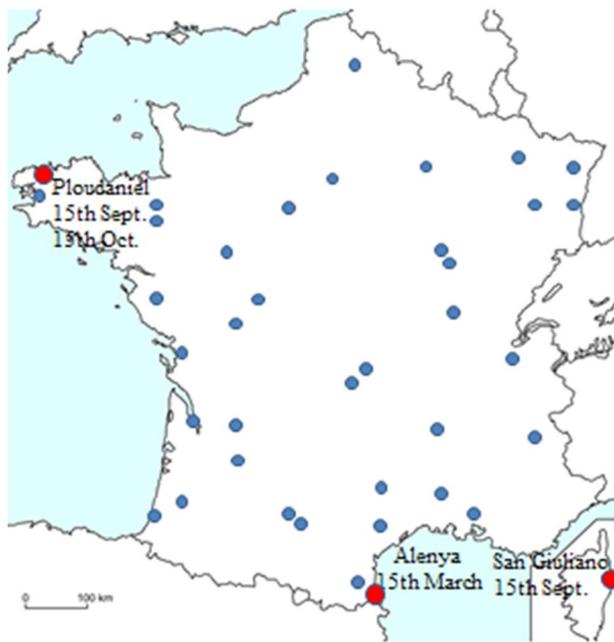


Fig. 7 MET sampled by OptiMET. The locations sampled by OptiMET are visualized by red dots. Sowing dates are indicated near the sampled locations (color figure online)

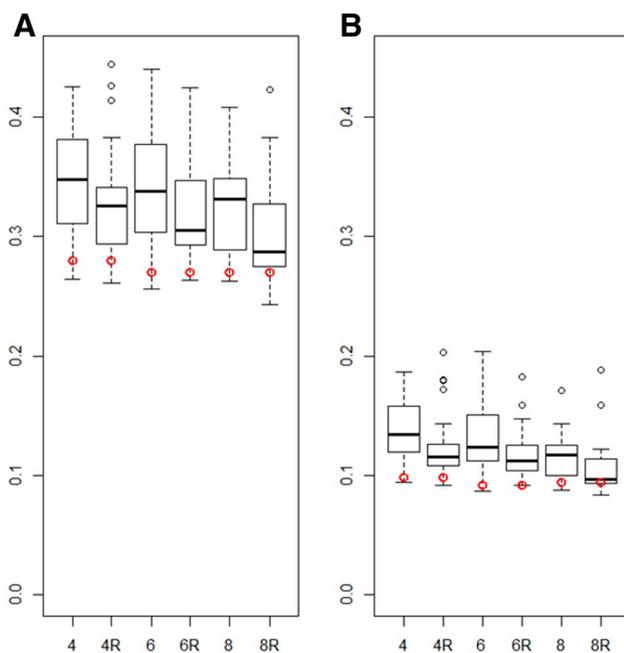


Fig. 8 Normalized \overline{RMSE} and normalized PSE averaged over the three parameters (**a** \overline{NRMSE}^* , **b** \overline{NPSE}) of random METs and of reasoned METs composed of 4, 6, or 8 environments (denoted by 4, 6, and 8 and 4R, 6R, and 8R, respectively). Each boxplot is composed of 40 METs. The \overline{NRMSE}^* and \overline{NPSE} obtained with the OptiMET MET of the corresponding number of trials is represented by a red dot

criteria, the OptiMET MET composed of four trials performed better than the reasoned MET composed of eight trials on average (\overline{NRMSE}^* of 0.28 and 0.29 for the OptiMET MET of size 4 and for the average of the reasoned METs of size 8, respectively).

Discussion and conclusions

The interest of using ecophysiological modelling to better model GEI is now well recognized in the plant genetics community (Chapman et al. 2002; Hammer et al. 2002; Reymond et al. 2003; Heslot et al. 2014; Technow et al. 2015; Bustos-Kort et al. 2016). It has been shown in various studies that CGM could be used to structure environments in groups according to the type and frequency of stress experienced by the crop (Löffler et al. 2005; Hammer and Jordan 2007; Chenu et al. 2011). This clustering approach has the advantage of reducing GEI within each group of environments, which facilitates the implementation of GS.

In the present study, a more integrated approach was applied: CGM was used to characterize varieties by genetic parameters expected to be independent from the environment. This means that the QTL detected for these traits are stable across environments, and their prediction accuracy will also be independent from the environments. Once the QTL are detected and the GS model calibrated, it is possible to predict the values of these traits for various varieties, which can then be used to predict their performances in various potentially new or virtual environments thanks to the CGM. This approach (CGM-MAS) is potentially highly powerful but relies on a difficult task which is the estimation of the genetic parameters. Moreover, CGM-MAS is composed of many successive steps, probably leading to error propagation. As a result, to run CGM-MAS efficiently, one has to optimize each step of the process, in particular at first the estimation of the parameters which affects all the following steps. Here, we propose a criterion called OptiMET to define an optimal set of environments (MET) for the estimation of the genetic parameters, i.e., an MET generating parameter estimates with low error variance. OptiMET was inspired from a study on hydrological modelling (Leube et al. 2012) in the context of Bayesian model averaging. OptiMET was tested using simulations and a real data set, with the example of wheat phenology.

Evaluation of the criterion OptiMET using simulations

The \overline{NRMSE} of the parameter estimates were clearly variable between years of experiments and between parameters (Fig. 3). The year effect was large and

affected both the average and the variability of the NRMSE, which can be explained by the important year effect on climatic conditions. This year effect was for example important for the OptiMET MET, which NRMSE could almost double from one year to another (for example 2007 and 2008 for VAI, Fig. 3). This year effect affected the three parameters differently, because they are not influenced by the same environmental covariates. It is interesting to note that the NRMSE of SLDL, which is affected by an environmental covariate stable across years (day length), also showed between year variability, revealing the complex dependencies between parameters. However, despite these variabilities, the ranking of the sampling approaches remained the same, with OptiMET doing better than the expert and the random METs. The difference between the NRMSE of the OptiMET MET and the average NRMSE of the random METs varied between years, but OptiMET did always better or as good as the best random METs for the three parameters. The expert MET unexpectedly performed poorly, doing sometimes better, sometimes worse than random samples, and was particularly inefficient to estimate the parameter Phyl. This could be explained by the fact that in the location \times sowing date, combinations composing the expert MET Phyl did not participate much in the variability of heading time as revealed by sensitivity analysis (results not shown). For the random METs, it appeared that sampling both winter and spring sowing dates performed better, doing best when two or three winter sowings were sampled (Fig. 3a2, b2). These combinations of winter and spring sowings are, indeed, supposed to decorrelate the effect of the different parameters, and that is the reason why the combination “three winter sowings and one spring sowing” was chosen in the expert MET.

Similar conclusions could be drawn on detection power (Fig. 4), as it is influenced by the error variance (and thus by NRMSE). One main conclusion, common to all METs, is that detection power was low for the three parameters with a maximum of 18% for SLDL with OptiMET. This could be explained by the fact that the simulated genetic architecture was influenced by 25 QTL following geometric series (Lande and Thompson 1990), which means that most of these QTL explained a small portion of the total genetic variance. For some real traits, major QTL can exist and have thus to be taken into account in the construction of the prediction formula. In our case, as many QTL were simulated (25 for each parameter), predictions were made with a classical G-BLUP model. Prediction accuracy of the parameter values of 270 independent varieties (Fig. 5) was also variable between years and METs, but again, OptiMET MET performed better than other METs with accuracies always above 0.52.

When these parameter predictions were used to predict heading time in independent MET using the CGM, the prediction accuracies obtained were high for the OptiMET MET and the expert MET each year (around 0.6, Fig. 6). Although the difference of efficiency between OptiMET and the expert MET was less pronounced than with the parameter values accuracies, OptiMET always did better than the expert MET and was more stable across years. For some years, the difference between the OptiMET and the expert MET was, indeed, more important (for example 2003, 2008 and 2009), probably because heading time was more sensitive to Phyl variations for these years. Prediction accuracies of the random METs were on average much lower and sometimes negative, and this time, the METs composed of four winter sowings performed on average better than the other combinations. This could be explained by the fact that the validation environments were all winter sowings (representing actual agricultural practices), and thus, METs composed of four winter sowings are more representative of what happens in the validation environments. It is interesting to note that the OptiMET MET which is composed of three winter sowings (and one spring sowing) performed better than random METs composed of four winter sowings.

Evaluation of OptiMET on real data

Working on simulated data sets is interesting, because we know the truth and we can generate various situations, but it is often simplistic in comparison to real experiments. We, therefore, compared the efficiency of various real METs for the parameter estimation. As expected, the NRMSE obtained with METs of the same size (four location \times sowing date combinations) were higher for the real data set than for the simulated data set (Figs. 3, 8). With the real data set, the quality criteria ($\overline{\text{NRMSE}}^*$ and $\overline{\text{NPSE}}$) decreased with the size of the MET, but this decrease was slow (Fig. 8). This difference between the real and the simulated data sets can be explained by the fact that there were 156 simulated environments, whereas only 26 real environments and the simulated environments were much more variable in comparison to the 26 real environments (more than half of these 26 environments were October or November sowings in the North of France). In addition to this, it is possible that the heritabilities of the 26 environments were lower than the heritability simulated in the first part. Unfortunately, there were no sufficient observations (no repetitions) to estimate the heritabilities in the 26 environments. Another point is that $\overline{\text{NRMSE}}^*$ and $\overline{\text{NPSE}}$ are computed using the parameter estimates obtained with the 26 environments and are thus also subjected to estimation errors. However, the METs sampled with OptiMET were always among the most efficient, and the OptiMET MET

composed of four location x sowing date combinations performed similarly than the quantile of the best random and reasoned METs of size eight. One point that may have limited the efficiency of OptiMET with this data set is that the meteorological data used to compute OptiMET were obtained from the meteorological stations, the closest from the location, which was sometimes a few kilometers away. To compute OptiMET more efficiently, it would have been necessary to measure daily temperature at each location of experiment.

We can conclude from this part, that with this real data set which was specifically produced to study the phenological parameters, choosing the best eight environments lead to parameter estimates of poor quality (in comparison to the simulated data sets). Even more contrasted environments are required. However, one major conclusion is that OptiMET was efficient to define better METs, which means that it is an interesting tool to determine experimental design before collecting phenotypes.

Limits and perspectives of OptiMET

We have shown here that in the context of CGM-MAS, experimental designs could be optimized before having access to any observation. However, we have to keep in mind that the use of OptiMET requires (1) a robust CGM for the trait of interest in the considered environments, (2) environmental covariates that can be predicted or at least approximated before doing the experiment, and (3) prior knowledge on the distribution of each genetic parameter (at least the bounds of the distribution). For (1), we can benefit from decades of ecophysiological research which resulted in the development of reference CGM such as SIRIUS, APSIM, STICS, or CERES which simulate the development of the plant from sowing to yield elaboration (Martre et al. 2015b; Yin and Struik 2016). However, each model has its specificities and its own domain of validity, which means that their predictions are reliable in some ranges of environments. Therefore, when using OptiMET to define optimal experimental design (and more generally to lead CGM-MAS approaches), we have to make sure that the chosen environments are in the range of validity of the CGM. This is an important point to consider to use OptiMET efficiently, because this criterion will by definition identify contrasted environments. Therefore, one has to take care that the possible environments proposed to OptiMET are all in the range of validity of the CGM. The METs sampled by OptiMET result in contrasted phenotypes, which means that the phenotypes may be more difficult to measure. With the example of wheat phenology, spring sowings result in heading time spread across a period of few weeks to few months, and so, it will be more difficult for the experimenter to

follow the plant development day by day over this long period. A special care has to be put on these experiments to reach high heritabilities. For (2), we have shown that average climatic data could be used to approximate the climate of future years for the CGM that we have considered. However, this will certainly not be true for all CGM, particularly if they rely on more erratic covariates as rainfall. In that case, it might be useful to use climate generators such as LARS (Semenov et al. 1998) and to take into account the inter-year variability when computing OptiMET. In such context, one can compute OptiMET for many specific years (using the climatic conditions of past years or simulated years) and choose the optimal MET which leads to low OptiMET values across years. Of course, if the experiment can be done in controlled conditions such as high-throughput phenotyping platforms, the use of OptiMET will be much easier and will allow to tune the covariates that can be controlled on these platforms. The third point that has to be taken care of (3) is the definition of the distribution of the genetic parameters. Prior knowledge on the distribution of each genetic parameter is, indeed, necessary to compute OptiMET (at least the bounds of the distributions). Here, we had no more information than the minimal and maximal values of each parameter (defined by expert knowledge and/or literature), so the values were chosen to get a uniform coverage of the parameters space. However, if more information is available, it would be possible to improve these distributions, for example by taking into account that some values are more probable than others, or that some genetic parameters are correlated. The more information is available on the joint distributions of the genetic parameters, the more realistic will be the values sampled, and the more efficient will be OptiMET. Indeed, if the values are chosen according to these informative distributions, then OptiMET will automatically put more weight on the parts of the parameter space which are more probable to occur.

Alternative uses of OptiMET that were not illustrated in this paper are the definition of METs to estimate efficiently one or few specific parameters (instead of a full parameterization as performed in the present work). Such an approach would be relevant for example when a focus is made on the genetic architecture of a specific parameter. Another potential use would be to define METs that are complementary to already existing data sets. When studying the effect of abiotic stress in multi-local trials, it often happens that the experiments do not cover the whole range of stress that was expected. In such a situation, it could be valuable to use criteria such as OptiMET to define additional complementary experiments (in controlled conditions and/or in a minimal but optimal MET) with for example stress scenarios which were missing in the existing data set.

The present study illustrated that OptiMET could be efficient to determine optimal experimental design of a given size (number of location \times sowing date combinations), but using OptiMET to define an optimal size of experimental design would be more complicated. In other words, OptiMET is efficient to compare experimental designs of the same size, but not to estimate a risk (a level of precision) associated to these experimental designs. Further methodological developments are required for this.

In conclusion, the data sets studied here clearly showed that choosing relevant experimental designs was highly important to lead CGM-MAS approaches. The quality of the parameter estimates, indeed, influences all the following steps of the CGM-MAS process, including the performance predictions. The criterion OptiMET was efficient to define such optimal experimental designs and resulted in better parameter estimates both on simulated and real data. It would be now interesting to evaluate this criterion on other traits simulated by other CGM.

Author contribution statement Conceived and designed the experiments: RR, EK, VA, JLG, HM, FXO, and MR. Analyzed the data: RR. Wrote the paper: RR. Revised the manuscript critically: EK, VA, JLG, and HM.

Acknowledgements This work is supported by the projects: FSOV-Précocité, BreedWheat (ANR-10-BTBR-0003) and OptiGBM. We thank two anonymous reviewers and the editor for their constructive comments, which improved the manuscript. We thank the INRA MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing computational resources and support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The authors declare that the experiments comply with the current laws of the countries in which the experiments were performed.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agreste (2016) Produits agroalimentaires. Céréales. France. <http://agreste.agriculture.gouv.fr/IMG/pdf/Gaf16p122.pdf>. Accessed 19 April 2017
- Atkinson AC, Donev AN (1992) Optimum experimental designs. Oxford Science Publications, Oxford
- Balfourier F, Roussel V, Strelchenko P et al (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet* 114:1265–1275. doi:10.1007/s00122-007-0517-1
- Bertin N, Martre P, Genard M et al (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. *J Exp Bot* 61:955–967. doi:10.1093/jxb/erp377
- Bogard M, Ravel C, Paux E et al (2014) Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. *J Exp Bot* 65:5849–5865. doi:10.1093/jxb/eru328
- Brooks RJ, Semenov MA, Jamieson PD (2001) Simplifying Sirius: sensitivity analysis and development of a meta-model for wheat yield prediction. *Eur J Agron* 14:43–60. doi:10.1016/S1161-0301(00)00089-7
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype \times environment interaction using pedigree and dense molecular markers. *Crop Sci* 52:707. doi:10.2135/cropsci2011.06.0299
- Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk F (2016) Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. In: Yin X, Struik PC (eds) Crop systems biology. Springer International Publishing, Cham, pp 55–82
- Butler DG, Smith AB, Cullis BR (2014) On the design of field experiments with correlated treatment effects. *J Agric Biol Environ Stat* 19:541–557. doi:10.1007/s13253-014-0191-0
- Chapman SC, Cooper M, Hammer GL (2002) Using crop simulation to generate genotype by environment interaction effects for sorghum in water-limited environments. *Aust J Agric Res* 53:379–389. doi:10.1071/AR01070
- Chenu K, Chapman SC, Hammer GL et al (2008) Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. *Plant Cell Environ* 31:378–391. doi:10.1111/j.1365-3040.2007.01772.x
- Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer F, Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J Exp Bot* 62:1743–1755. doi:10.1093/jxb/erq459
- Cooper M, Technow F, Messina C, Gho C, Totir LR (2016) Use of crop growth models with whole-genome prediction: application to a maize multi-environment trial. *Crop Sci* 56:1–16. doi:10.2135/cropsci2015.08.0512
- Crossa J, de los Campos G, Maccaferri M et al (2015) Extending the marker \times environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Sci* 5:2193–2209. doi:10.2135/cropsci2015.04.0260
- Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi:10.1371/journal.pone.0003395
- Dumont B, Leemans V, Mansouri M et al (2014) Parameter identification of the STICS crop model, using an accelerated formal MCMC approach. *Environ Model Softw* 52:121–135. doi:10.1016/j.envsoft.2013.10.022
- Habier D, Fernando RL, Dekkers JCM (2008) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. doi:10.1534/genetics.107.081190

- Hammer GL, Jordan DR (2007) An integrated systems approach to crop improvement. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene–plant–crop relations. Wageningen UR–Frontis series no. 21. Springer, Dordrecht, The Netherlands, pp 45–61
- Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling—from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. *Eur J Agron* 18:15–31. doi:[10.1016/S1161-0301\(02\)00093-X](https://doi.org/10.1016/S1161-0301(02)00093-X)
- He J, Le Gouis J, Stratonovitch P et al (2012) Simulation of environmental and genotypic variations of final leaf number and anthesis date for wheat. *Eur J Agron* 42:22–33. doi:[10.1016/j.eja.2011.11.002](https://doi.org/10.1016/j.eja.2011.11.002)
- Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* 127:463–480. doi:[10.1007/s00122-013-2231-5](https://doi.org/10.1007/s00122-013-2231-5)
- Iizumi T, Yokozawa M, Nishimori M (2009) Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: application of a Bayesian approach. *Agric For Meteorol* 149:333–348. doi:[10.1016/j.agrformet.2008.08.015](https://doi.org/10.1016/j.agrformet.2008.08.015)
- Jamieson PD, Munro CA (2000) The calibration of a model for day-length responses in spring wheat for large numbers of cultivars. *Agron N Z* 30:25–28
- Jamieson PD, Semenov MA (2000) Modelling nitrogen uptake and redistribution in wheat. *Field Crops Res* 68:21–29. doi:[10.1016/S0378-4290\(00\)00103-9](https://doi.org/10.1016/S0378-4290(00)00103-9)
- Jamieson PD, Porter JR, Goudriaan J et al (1998a) A comparison of the models AFRCWHEAT2, CERES-Wheat, sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. *Field Crops Res* 55:23–44. doi:[10.1016/S0378-4290\(97\)00060-9](https://doi.org/10.1016/S0378-4290(97)00060-9)
- Jamieson PD, Semenov MA, Brooking IR, Francis GS (1998b) Sirius: a mechanistic model of wheat response to environmental variation. *Eur J Agron* 8:161–179
- Jarquín D, Crossa J, Lacaze X et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. doi:[10.1007/s00122-013-2243-1](https://doi.org/10.1007/s00122-013-2243-1)
- Keating BA, Carberry PS, Hammer GL et al (2003) An overview of APSIM, a model designed for farming systems simulation. *Eur J Agron* 18:267–288
- Kempton RA, Fox PN (1997) Statistical methods for plant variety evaluation. Chapman & Hall, London
- Klein T, Calanca P, Holzkämper A et al (2012) Using farm accountancy data to calibrate a crop model for climate impact studies. *Agric Syst* 111:23–33. doi:[10.1016/j.agry.2012.05.001](https://doi.org/10.1016/j.agry.2012.05.001)
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Leube PC, Geiges A, Nowak W (2012) Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design: assessing the expected data impact. *Water Resour Res* 48:W02501. doi:[10.1029/2010WR010137](https://doi.org/10.1029/2010WR010137)
- Löffler CM, Wei J, Fast T, Gogerty J, Langton S, Bergman M, Merrill B, Cooper M (2005) Classification of maize environments using crop simulation and geographic information systems. *Crop Sci* 45:1708–1716
- Makowski D, Wallach D, Tremblay M (2002) Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie* 22:191–203. doi:[10.1051/agro:2002007](https://doi.org/10.1051/agro:2002007)
- Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting responses in multiple environments: issues in relation to genotype × environment interactions. *Crop Sci*. doi:[10.2135/cropsci2015.05.0311](https://doi.org/10.2135/cropsci2015.05.0311)
- Martre P, He J, Le Gouis J, Semenov MA (2015a) In silico system analysis of physiological traits determining grain yield and protein concentration for wheat as influenced by climate and crop management. *J Exp Bot* 66:3581–3598. doi:[10.1093/jxb/erv049](https://doi.org/10.1093/jxb/erv049)
- Martre P, Wallach D, Asseng S et al (2015b) Multimodel ensembles of wheat growth: many models are better than one. *Glob Change Biol* 21:911–925. doi:[10.1111/gcb.12768](https://doi.org/10.1111/gcb.12768)
- Messina CD, Jones JW, Boote KJ, Vallejos CE (2006) A gene-based model to simulate soybean development and yield responses to environment. *Crop Sci* 46:456. doi:[10.2135/cropsci2005.04-0372](https://doi.org/10.2135/cropsci2005.04-0372)
- Messina CD, Technow F, Tang T, Totir RL, Gho C, Cooper M (2017) Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). *bioRxiv*. doi:[10.1101/100057](https://doi.org/10.1101/100057)
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Nakagawa H, Yamagishi J, Miyamoto N et al (2005) Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. *Theor Appl Genet* 110:778–786. doi:[10.1007/s00122-004-1905-4](https://doi.org/10.1007/s00122-004-1905-4)
- Onogi A, Watanabe M, Mochizuki T, Hayashi T, Nakagawa H, Hasegawa T, Iwata H (2016) Towards integration of genomic selection with crop modeling: the development of an integrated approach to predicting rice heading dates. *Theor Appl Genet* 129:805–817
- Piepho HP, Williams ER (2006) A comparison of experimental designs for selection in breeding trials with nested treatment structure. *Theor Appl Genet* 113:1505–1513
- Prudent M, Lecomte A, Bouchet J-P et al (2011) Combining eco-physiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. *J Exp Bot* 62:907–919. doi:[10.1093/jxb/erq318](https://doi.org/10.1093/jxb/erq318)
- Quilot B, Genard M, Lescourret F et al (2005) Simulating genotypic variation of fruit quality in an advanced peach×Prunus davidiana cross. *J Exp Bot* 56:3071–3081. doi:[10.1093/jxb/eri304](https://doi.org/10.1093/jxb/eri304)
- Reymond M, Muller B, Leonardi A et al (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol* 131:664–675. doi:[10.1104/pp.013839](https://doi.org/10.1104/pp.013839)
- Rincint R, Moreau L, Monod H et al (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics* 197:375–387. doi:[10.1534/genetics.113.159731](https://doi.org/10.1534/genetics.113.159731)
- Rousset M, Bonnin I, Remoué C, Falque M, Rhoné B, Veyrieras JB, Madur D, Murigneux A, Balfourier F, Le Gouis J, Santoni S, Goldringer I (2011) Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 123:907–926. doi:[10.1007/s00122-011-1636-2](https://doi.org/10.1007/s00122-011-1636-2)
- Schulz-Streeck T, Ogutu JO, Gordillo A et al (2013) Genomic selection allowing for marker-by-environment interaction. *Plant Breed* 132:532–538. doi:[10.1111/pbr.12105](https://doi.org/10.1111/pbr.12105)
- Semenov M, Wolf J, Evans L et al (1996) Comparison of wheat simulation models under climate change. II. Application of climate change scenarios. *Clim Res* 7:271–281. doi:[10.3354/cr007271](https://doi.org/10.3354/cr007271)
- Semenov M, Brooks R, Barrow E, Richardson C (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Clim Res* 10:95–107. doi:[10.3354/cr010095](https://doi.org/10.3354/cr010095)
- Technow F, Messina CD, Radu Totir LR et al (2015) Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One* 10:e0130855. doi:[10.1371/journal.pone.0130855](https://doi.org/10.1371/journal.pone.0130855)

- Teyssèdre S, Elsen J-M, Ricard A (2012) Statistical distributions of test statistics used for quantitative trait association mapping in structured populations. *Genet Sel Evol* 44:32. doi:[10.1186/1297-9686-44-32](https://doi.org/10.1186/1297-9686-44-32)
- Uptmoor R, Li J, Schrag T, Stützel H (2011) Prediction of flowering time in *Brassica oleracea* using a quantitative trait loci-based phenology model: flowering time in *Brassica oleracea*. *Plant Biol*. doi:[10.1111/j.1438-8677.2011.00478.x](https://doi.org/10.1111/j.1438-8677.2011.00478.x)
- Van Oijen M, Rougier J, Smith R (2005) Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol* 25:915–927
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. doi:[10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980)
- Wallach D, Buis S, Lecharpentier P et al (2011) A package of parameter estimation methods and implementation for the STICS crop-soil model. *Environ Model Softw* 26:386–394. doi:[10.1016/j.envsoft.2010.09.004](https://doi.org/10.1016/j.envsoft.2010.09.004)
- Wang K (2008) An analytic study of the power of popular quantitative-trait-locus mapping methods. *Behav Genet* 38:554–559. doi:[10.1007/s10519-008-9220-5](https://doi.org/10.1007/s10519-008-9220-5)
- White JW, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron J* 88:416–422
- White JW, Herndl M, Hunt LA et al (2008) Simulation-based analysis of effects of and loci on flowering in wheat. *Crop Sci* 48:678. doi:[10.2135/cropsci2007.06.0318](https://doi.org/10.2135/cropsci2007.06.0318)
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genet Res* 75:249–252. doi:[10.1017/S0016672399004462](https://doi.org/10.1017/S0016672399004462)
- Wöhling T, Geiges A, Nowak W et al (2013) Towards optimizing experiments for maximum-confidence model selection between different soil-plant models. *Procedia Environ Sci* 19:514–523. doi:[10.1016/j.proenv.2013.06.058](https://doi.org/10.1016/j.proenv.2013.06.058)
- Yin X (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. *J Exp Bot* 56:967–976. doi:[10.1093/jxb/eri090](https://doi.org/10.1093/jxb/eri090)
- Yin X, Struik PC (2016) *Crop systems biology*. Springer, Cham
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Zheng B, Biddulph B, Li D et al (2013) Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. *J Exp Bot* 64:3747–3761. doi:[10.1093/jxb/ert209](https://doi.org/10.1093/jxb/ert209)
- Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. *Genetics* 182:355–364. doi:[10.1534/genetics.108.098277](https://doi.org/10.1534/genetics.108.098277)