



**HAL**  
open science

## **Genotype-by-environment interaction with high-dimensional environmental data: an example in pigs**

Fernando Bussiman, Daniela Lourenco, Jorge Hidalgo, Ching-Yi Chen, Justin Holl, Ignacy Misztal, Zulma G. Vitezica

### ► **To cite this version:**

Fernando Bussiman, Daniela Lourenco, Jorge Hidalgo, Ching-Yi Chen, Justin Holl, et al.. Genotype-by-environment interaction with high-dimensional environmental data: an example in pigs. *Genetics Selection Evolution*, 2025, 57 (1), pp.28. <10.1186/s12711-025-00974-2>. <hal-05103374>

**HAL Id: hal-05103374**

**<https://hal.science/hal-05103374v1>**

Submitted on 9 Jun 2025

**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 CC BY 4.0 - Attribution - International License

RESEARCH ARTICLE

Open Access



# Genotype-by-environment interaction with high-dimensional environmental data: an example in pigs

Fernando Bussiman<sup>1\*</sup> , Daniela Lourenco<sup>1</sup>, Jorge Hidalgo<sup>1</sup>, Ching-Yi Chen<sup>2</sup>, Justin Holl<sup>2</sup>, Ignacy Misztal<sup>1</sup> and Zulma G. Vitezica<sup>3</sup>

## Abstract

**Background** In traditional genetic prediction models, environments are typically treated as uncorrelated effects, either fixed or random. Environments can be correlated when they share the same location, management practices, or climate conditions. The temperature-humidity index (THI) is often used to address environmental effects related to climate or heat stress. However, it does not fully describe the complete climate profile of a specific location. Therefore, it is more appropriate to use multiple environmental covariates (ECs), when available, to describe the weather in a specific environment. This raises the question of whether publicly available weather information (such as NASA POWER) is useful for genomic predictions. Genotype-by-environment interaction (GxE) can be modeled using multiple-trait models or reaction norms. However, the former requires a substantial number of records per environment, while the latter can result in over-parametrized models when the number of ECs is large. This study investigated whether using ECs is a suitable strategy to correlate environments (herds) and to model GxE in the genomic prediction of purebred pigs for production traits.

**Results** We evaluated different models to account for environmental effects and GxE. When environments were correlated based on ECs, we observed an increase in environmental variance, which was accompanied by an increase in phenotypic variance and a decrease in heritability. Furthermore, including environments as an uncorrelated random effect yielded the same accuracy of estimated breeding values as treating them as correlated based on weather information. All the tested models exhibited the same bias, but the predictions from the multiple-trait models were under-dispersed. Evidence of GxE was observed for both traits; however, there were more genetically unconnected environments for backfat thickness than for average daily gain.

**Conclusions** Using outdoor weather information to correlate environments and model GxE offers limited advantages for genomic predictions in pigs. Although it adds complexity to the model and increases computing time without improving accuracy, it does enhance model fit. Including environment information (e.g. herd effect) as an uncorrelated random effect in the model could help address GxE and environmental effects.

## Background

Fluctuations in environmental variables and increasing temperatures due to climate change pose a significant concern for livestock production [1]. These seasonal impacts may be seen as heat-stress effects, which are evident in several species worldwide. They have a detrimental impact on animal welfare and disease incidence,

\*Correspondence:

Fernando Bussiman  
fob@uga.edu

<sup>1</sup> Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA

<sup>2</sup> Genus Pig Improvement Company, Hendersonville, TN 37075, USA

<sup>3</sup> GenPhySE, INPT, INRAE, ENVT, 31326 Castanet Tolosan, France



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

causing physiological changes [2]. The more challenging the environment, the greater the negative effects on the productivity, performance, growth, and development of animals exposed to such conditions [3].

Pigs are particularly susceptible to heat stress due to a lack of physiological adaptation, as pigs' sweat glands do not respond to heat stress [4], and they can only dissipate up to 50% of their heat production by respiratory evaporation [5]. As a result, even with enhanced cooling strategies in pig farms, production losses associated with heat stress remain a concern [6, 7]. Biologically, with high-production levels, pigs exhibit an increase in heat production, although the thermoregulatory mechanism behind this still needs to be fully understood [8].

Environmental effects are traditionally accounted for by fitting herd as an effect in the prediction model [1] and/or by incorporating environmental effects into contemporary groups (CGs) [9]. While some researchers advocate for modeling these effects as random [10, 11], fitting them as fixed effects remains more common [11, 12]. In a fixed effects modeling framework, herds are assumed to be completely independent with no shared characteristics. However, Chauhan and Thompson proposed treating CGs as a random effect, allowing for a covariance among them [10]. When two herds are physically close, they are expected to share similarities in climate, geographical, and management conditions [1, 13]. Thus, when different environments experience the same weather conditions, one can expect an additional layer of similarity among these environments.

In livestock, correlated environmental effects have been studied through various approaches and environmental descriptors, aiming to enhance the reliability of genetic predictions. For example, Tiezzi et al. utilized geographical, management, climate, and CG information to model environmental (co)variance structures in US Holstein cattle [14]. Cuyabano et al. employed global positioning system coordinates (GPS) to correlate herds in the genomic evaluation of Hanwoo beef cattle [13]. Similarly, Makanjuola et al. applied the temperature-humidity index (THI) to the model covariance structure among herds for the same breed [1]. Selle et al. analyzed spatial relationships between herds to account for environmental effects in smallholder dairy cattle [15]. Lastly, Santana et al. used a Gaussian kernel based on geographical coordinates to model herd effects in beef cattle [16].

Traditionally, the environmental effects associated with climate or heat stress are measured using some function of the THI [1, 17–19]. However, evidence suggests that the THI is not the most effective method for accounting for these effects, as THI calculations must be tailored to the region in which they are implemented [20], and even then, temperature often serves just as well at predicting

heat stress as the THI [21, 22]. Generally, genetic evaluations that account for heat stress involve finding a THI threshold above which the phenotype begins to decline. This process leads to calculating the heat load (HL) as a function of THI, followed by fitting a single reaction norm model of HL for genetic evaluation [17–20]. Employing multiple environmental covariates ECs could result in a more flexible approach, facilitating a better characterization of the environment. However, if a reaction-norm (or random regression) model is used, a large number of ECs can super-parameterize the model.

Accounting for ECs also allows for modeling the genotype-by-environment interaction (GxE) based on climate effects. Usually, GxE is modeled either using a multiple-trait or a reaction-norm model [23]. The first approach considers records from different environments as distinct traits and assumes a genetic correlation among them [17, 24–26]. The second approach requires a continuous environmental gradient and models estimated breeding values (EBV) as a function of the ECs [17–19]. Conversely, Lopez-Cruz et al. proposed a third approach that fits two genetic effects in the multiple-trait model [27]: one across environments and another specific to each environment. However this does not allow for ECs to be considered to avoid double-counting environmental effects. Recently, Zhao et al. derived a similar model to study quantitative trait loci-by-environment interactions [28].

Jarquín et al. proposed modeling GxE using covariance functions [29]. This approach correlates different environments based on the ECs, and the covariance between the EBV and the environmental effects accounts for the GxE [30]. The main advantage here is that modeling the ECs and GxE through covariance functions alleviates the problem of super-parametrization and provides a better characterization of the environmental conditions. Moreover, as the dimension of the covariance among environments is fixed (number of environments<sup>2</sup>), one can fit daily ECs, allowing for a high-dimensional reaction norm. Thus, this study aimed to evaluate the validity of high-dimensional environmental data in modeling correlated herd effects and GxE through covariance functions, using genomic prediction of production traits in purebred pigs as a case study.

## Methods

### Phenotypic, genomic, and environmental data

Data were provided by the Pig Improvement Company (PIC: a Genus company, Hendersonville, TN, USA). Approval from the Animal Care and Use Committee was not required for this study as the data were obtained from an existing database. The original data

set comprises 95,873 genotyped and phenotyped animals. The original dataset was down-sampled to enable variance component estimation with genomic information. We ensured the sample was representative of the whole dataset by implementing the following process: first, we set the minimum number of genotyped animals ( $N = 30,000$ ). Next, the phenotypic dataset was refined to remove outliers (i.e. records deviating  $\pm 3$  standard deviations from the phenotypic mean) and CGs (defined later) that lacked phenotypic variation or that had fewer than 30 records. Subsequently, the number of animals to be sampled (s) was calculated as the next integer greater than  $N/mCG$  (where  $N = 30,000$  and  $mCG$  is the average number of animals per contemporary group). The final dataset analyzed consisted of 35,596 records for average daily gain (ADG) and 31,105 for backfat thickness (BFT) from purebred pigs born between 2009 and 2020, genotyped and phenotyped. ADG and BFT were measured at the end of a growth performance testing period, which began at approximately 10 weeks of age and ended at around 20 weeks of age. The ADG was calculated by dividing weight by age, and BFT was measured using ultrasound. Table 1 presents the descriptive statistics for the dataset used. Records were collected from 11 farms, all with the same health status, located throughout the northern hemisphere. The distribution of animals per farm is shown in Table 2.

Genotypes were imputed up to 50 K single nucleotide polymorphisms (SNPs). After quality control using default parameters (excluding SNPs with minor allele frequency  $< 0.01$ , call rate  $< 0.90$ , monomorphic, Hardy–Weinberg Equilibrium with a maximum difference

**Table 1** Descriptive statistics for each trait in the dataset

Statistic	Trait	
	ADG (g)	BFT (mm)
$N^1$	35,597	32,105
Mean	751.29	8.06
SD	97.46	1.97
Min	450.70	5.00
Max	1000.00	18.23
YOB <sup>2</sup>	2009–2020	2009–2020
NCG <sup>3</sup>	203	220
NE <sup>4</sup>	11	11

ADG average daily gain, BFT backfat thickness

<sup>1</sup> Number of records

<sup>2</sup> Birth year range

<sup>3</sup> Number of contemporary groups

<sup>4</sup> Number of environments (farms)

**Table 2** Number of records per trait for each environment (farm), along with their location (country)

Environment	Trait		Location Country
	ADG	BFT	
4	5321	5185	Canada
6	3072	2585	Canada
2	674	284	United States
5	382	234	United States
11	3684	3251	United States
8	11,924	10,181	United States
7	6721	6839	United States
1	122	186	Czech Republic
3	213	192	Germany
9	1017	1205	Ireland
10	2467	1963	Ireland

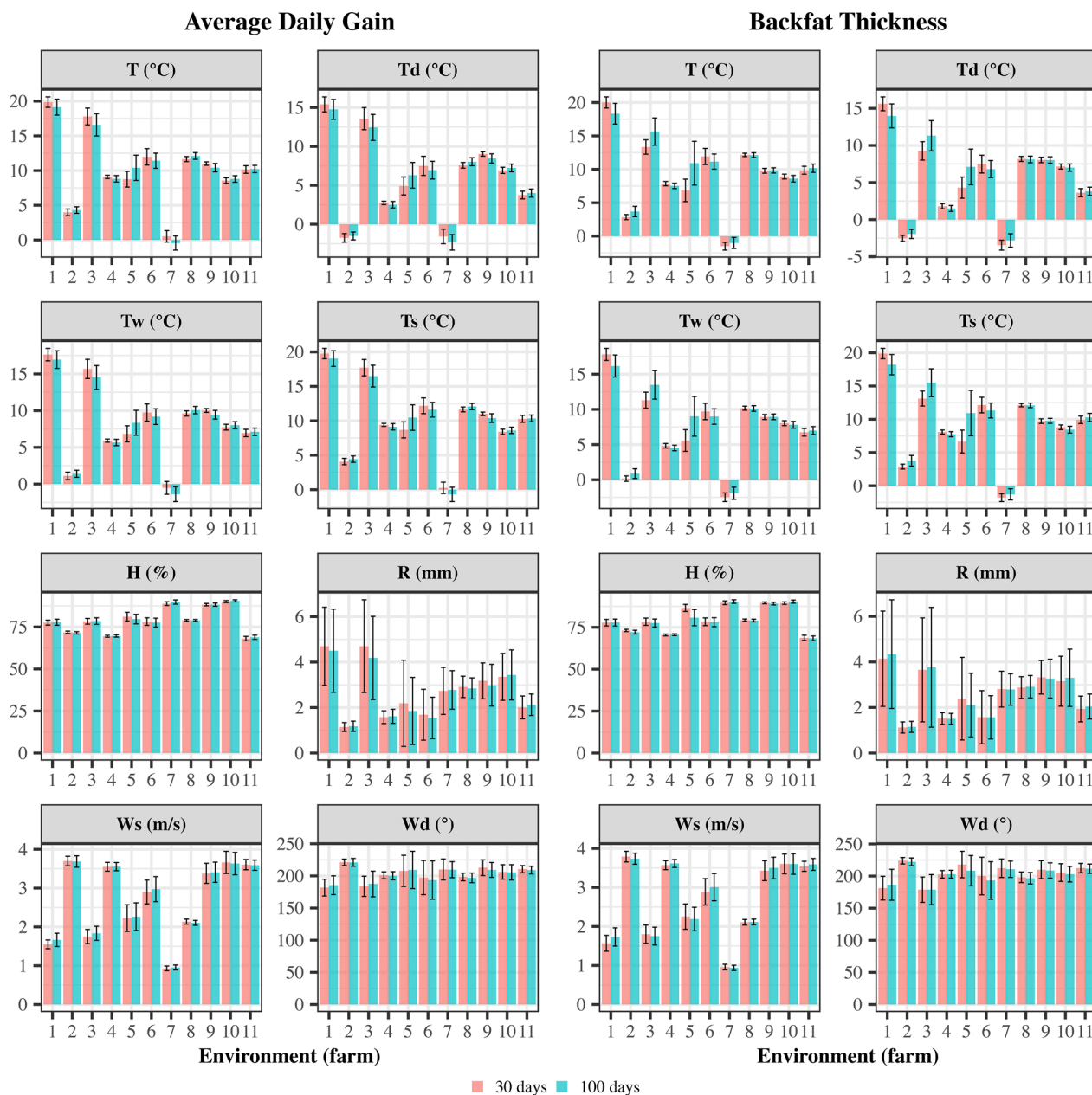
ADG average daily gain, BFT backfat thickness

between observed and expected heterozygosity  $> 0.15$ , number of Mendelian conflicts  $> 1\%$ ) by preGSf90 [31], 44,368 autosomal SNPs remained, which were used to build genomic relationship matrices. All animals used in the analyses were phenotyped and genotyped. Therefore, genetic relationships among animals across environments were implicitly modeled through the genomic relationship matrix, which connects all animals across all environments.

**Environmental covariates**

In this study, the term ‘environment’ refers to one of the 11 farms in the final dataset. The longitude and latitude coordinates were available for each environment, which was used to retrieve daily weather records from the NASA POWER website (<https://power.larc.nasa.gov/data-access-viewer/>). Taking the measurement day of the trait as ‘day one’, ECs were retrieved for a period of 99 days before the measurement, plus day 1, totaling 100 days of daily weather information for each animal. Mean ECs within 30, 40, 50, 60, 70, 80, 90, and 100 days were considered to account for cumulative weather effects. Since significant regression coefficients (i.e. those different from zero) were found for up to 100 days, further analyses were conducted using 30-day or 100-day daily ECs. The average ECs within each environment for 30 and 100 days are shown in Fig. 1.

Of all available ECs, those associated with the phenotype were kept for further analysis. Consequently, eight ECs were used: temperature at 2 m ( $^{\circ}C-T$ ), dew/frost temperature ( $^{\circ}C-Td$ ), wet-bulb temperature ( $^{\circ}C-Tw$ ), earth-skin temperature ( $^{\circ}C-Ts$ ), relative humidity ( $\% -H$ ), precipitation/rainfall (mm-R), wind speed at 2 m (m/s-Ws), and wind direction at 2 m ( $^{\circ} -Wd$ ).



**Fig. 1** Averages and standard deviations of environmental covariates within each environment (farm) for 30 and 100 days. *T* temperature at two meters (°C), *Td* dew/frost temperature (°C), *Tw* wet-bulb temperature, *Ts* hearth-skin temperature (°C), *H* relative humidity (%), *R* precipitation/rainfall (mm), *Ws* wind speed at two meters (m/s), and *Wd* wind direction at two meters (°)

Correlations among ECs for 30 or 100 days are shown in Fig. 2.

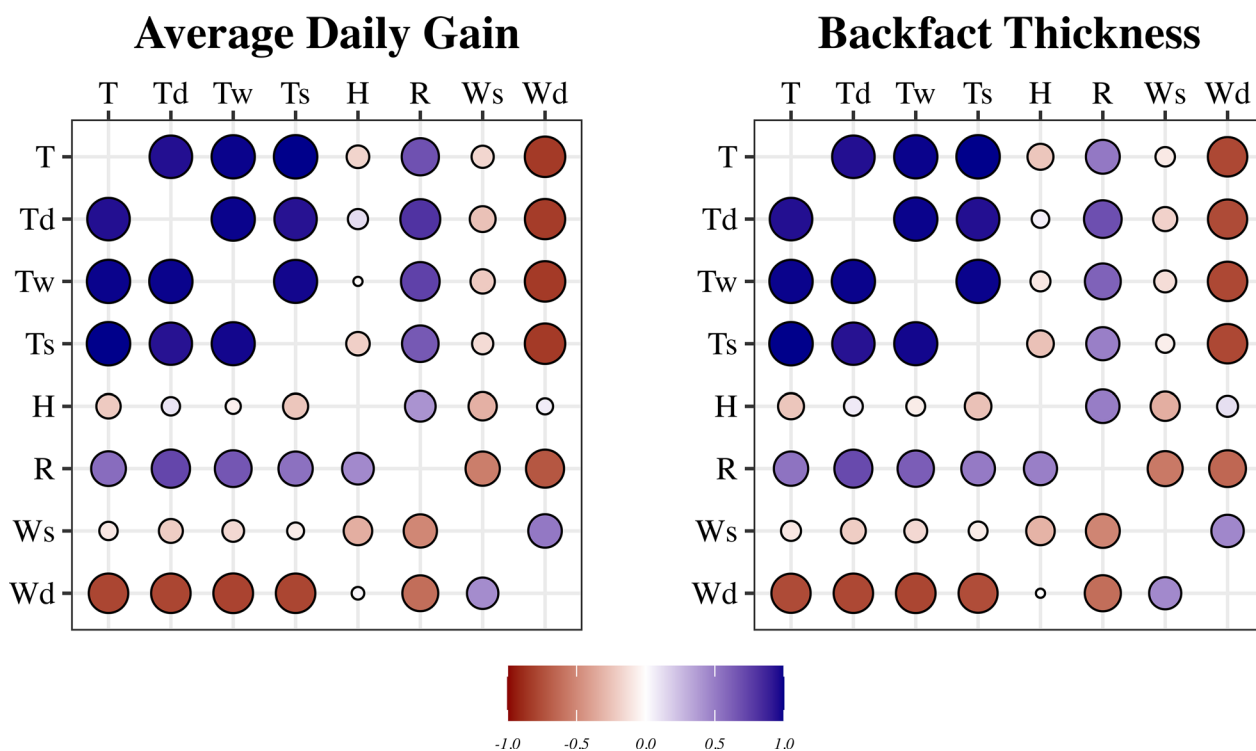
**Genomic prediction models**

**Baseline model**

Phenotypes (ADG and BFT) were analyzed using the genomic best linear unbiased prediction (GBLUP) mixed model. For both traits, CGs were defined by the

concatenation of sex, off-test year, and off-test week. Farm was not included in the CG as it was assumed to capture the environmental effect and was included as a separate random effect. For the two traits, the linear models can be written as:

$$ADG_{ijkl} = \mu + CG_i + l_j + g_k + e_l + \epsilon_{ijkl} \tag{1}$$



**Fig. 2** Correlations among environmental covariates (ECs) over 30 days (above diagonal) or 100 days (below diagonal). *T* temperature at two meters (°C), *Td* dew/frost temperature (°C), *Tw* wet-bulb temperature, *Ts* hearth-skin temperature (°C), *H* relative humidity (%), *R* precipitation/rainfall (mm), *Ws* wind speed at two meters (m/s), and *Wd* wind direction at two meters (°)

$$BFT_{ijkl} = \mu + \beta_1 W_k + CG_i + l_j + g_k + e_l + \epsilon_{ijkl} \quad (2)$$

where  $ADG_{ijkl}$  or  $BFT_{ijkl}$  denote the phenotypes;  $\mu$  is the intercept;  $\beta_1$  represents the linear regression coefficient of the covariate final weight ( $W$ ) during the growth test period of the  $k$ th animal;  $CG_i$  is the fixed effect of the  $i$ th contemporary group;  $l_j$  is the uncorrelated random effect of the  $j$ th litter;  $g_k$  is the random additive genetic effect of the  $k$ th animal;  $e_l$  is the random effect of the  $l$ th environment (farm); and  $\epsilon_{ijkl}$  is the random residual term.

In matrix notation, Eqs. 1 and 2 can be written as:

$$y = X\beta + Z_1l + Z_2g + Z_3e + \epsilon \quad (3)$$

where  $y$  represents the vector of phenotypes ( $ADG$  or  $BFT$ );  $\beta$  denotes the vector of fixed effects (as defined in Eqs. 1 and 2);  $l$  is the vector of litter effects;  $g$  is the vector of additive genetic effects;  $e$  is the vector of environment effects;  $\epsilon$  represents the vector of residual terms;  $X$  and  $Z_i$  are the incidence matrices for the fixed and random effects, respectively. Thus, under the assumption of

multivariate normality, the random effects (co)variance matrix is given by:

$$\text{Var} \begin{bmatrix} l \\ g \\ e \\ \epsilon \end{bmatrix} = \begin{bmatrix} I\sigma_l^2 & 0 & 0 & 0 \\ 0 & G\sigma_g^2 & 0 & 0 \\ 0 & 0 & I\sigma_e^2 & 0 \\ 0 & 0 & 0 & I\sigma_\epsilon^2 \end{bmatrix} \quad (4)$$

where  $I$  is an identity matrix and  $G$  is the additive genomic relationship matrix, constructed as in VanRaden [32] (method 1). This first model, as presented in Eq. 3 (with assumptions as in Eq. 4), will be referred to as MG (model with uncorrelated environments) in the following.

### Correlated environments models

The second model incorporates correlated random environmental effects, which can be obtained from the previous model by assuming a (co)variance matrix among environments ( $E$ ) derived from ECs. Let the ECs be stored in a matrix ( $F$ ), with each row corresponding to a

different environment and each column containing daily weather measurements ordered by day within covariate. After centering  $\mathbf{F}$ , the product  $\mathbf{F}\mathbf{F}'$  gives the weather (co)variance matrix among environments ( $\mathbf{E} = \mathbf{F}\mathbf{F}'$ ); this ensures that  $\bar{\mathbf{E}} = 0$ . Additionally,  $\mathbf{E}$  was scaled so  $\text{diag}(\mathbf{E}) = 1$ .

For this study, two  $\mathbf{E}$  matrices were tested based on 30-day weather information ( $\mathbf{E}_{30}$ ) and 100-day weather information ( $\mathbf{E}_{100}$ ). Correlations among environments from  $\mathbf{E}_i$  are presented in Fig. 3. Since the 30-day and 100-day correlations were not similar, we decided to retain both sets for the analysis. These two  $\mathbf{E}$  matrices defined two models:  $\text{ME}_{30}$  (model with correlated environments using 30 days' worth of ECs), in which  $\mathbf{e} \sim \text{MVN}(0, \mathbf{E}_{30}\sigma_e^2)$  (MVN: multivariate normal distribution), and  $\text{ME}_{100}$  (model with correlated environments using 100 days' worth of ECs), in which  $\mathbf{e} \sim \text{MVN}(0, \mathbf{E}_{100}\sigma_e^2)$ .

**Genotype-by-environment interaction models**

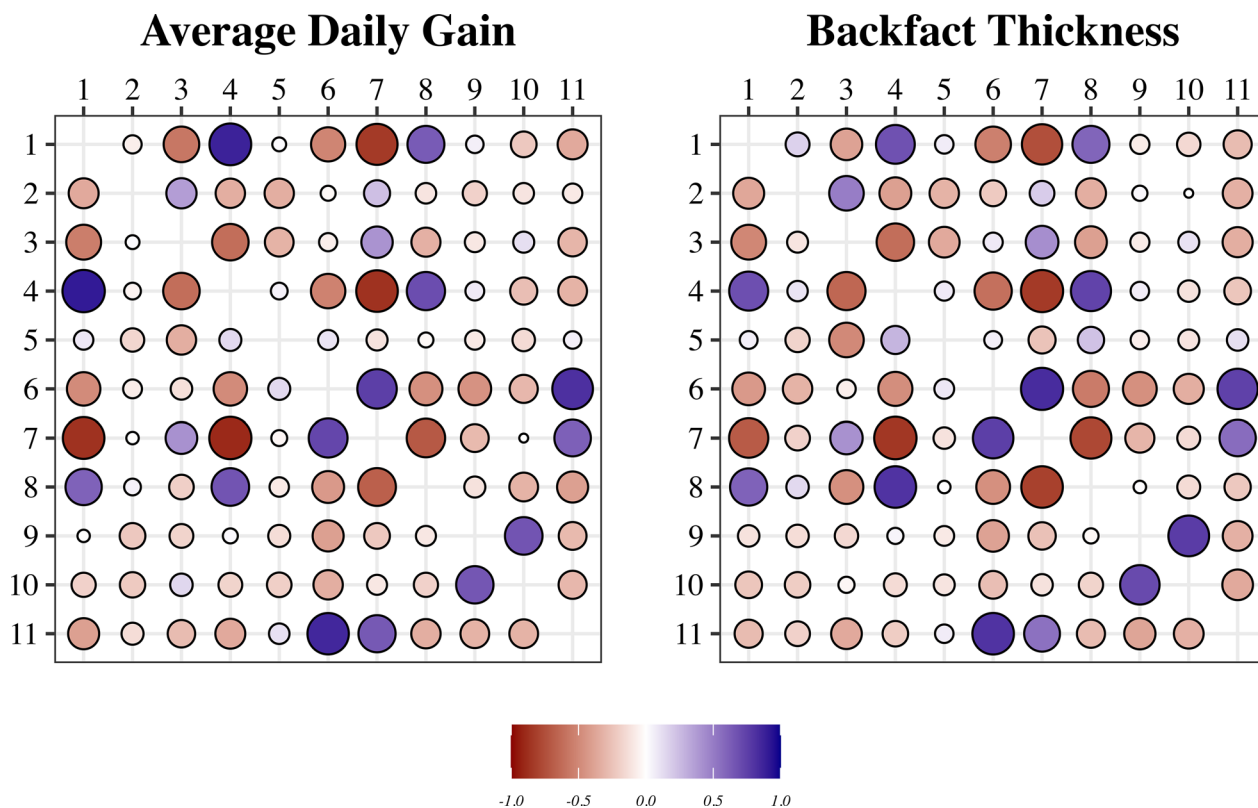
Assuming a first-order multiplicative model, GxE can be modeled by the products of covariance matrices [29]. Using  $\mathbf{E}$  as the covariance structure among environments, the environmental (co)variance matrix for

animals is given either by  $\mathbf{Z}_3\mathbf{E}_{30}\mathbf{Z}'_3$  or by  $\mathbf{Z}_3\mathbf{E}_{100}\mathbf{Z}'_3$ , where  $\mathbf{Z}_3$  is the design matrix defined in Eq. 3 that relates phenotypes to the environments. The GxE can be depicted by the  $k$ th additive genetic effect ( $g_k$ ) in the  $l$ th environment ( $e_l$ ), that is  $g_{e_{kl}}$ . The  $\text{cov}(g_{e_{kl}}, g_{e_{k'l'}})$  is equal to  $\text{cov}(g_k, g_{k'})\text{cov}(e_l, e_{l'})$  [29]. Because  $\text{Cov}(\mathbf{g}, \mathbf{g}) = \text{Var}(\mathbf{g}) = \mathbf{G}\sigma_g^2$  and  $\text{Cov}(\mathbf{Z}_3\mathbf{e}, \mathbf{Z}_3\mathbf{e}) = \mathbf{Z}_3\text{Var}(\mathbf{e})\mathbf{Z}'_3 = \mathbf{Z}_3\mathbf{E}\mathbf{Z}'_3\sigma_e^2$ , the covariance structure for the GxE is given by the Hadamard product between  $\mathbf{G}$  and  $\mathbf{Z}_3\mathbf{E}\mathbf{Z}'_3$ . Therefore, including GxE effects, the model was as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{l} + \mathbf{Z}_2\mathbf{g} + \mathbf{Z}_3\mathbf{e} + \mathbf{Z}_4\mathbf{ge} + \boldsymbol{\epsilon} \tag{5}$$

where  $\mathbf{ge}$  is the vector of GxE random effects. Assuming multivariate normality, the random effects (co)variance matrix is:

$$\text{Var} \begin{bmatrix} \mathbf{l} \\ \mathbf{g} \\ \mathbf{e} \\ \mathbf{ge} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_l^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}\sigma_g^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_i\sigma_e^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (\mathbf{G} \odot \mathbf{Z}_3\mathbf{E}\mathbf{Z}'_3)\sigma_{ge}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_\epsilon^2 \end{bmatrix} \tag{6}$$



**Fig. 3** Correlations among environments (farms) based on environmental covariates (ECs) over 30 days (above diagonal) or 100 days (below diagonal)

where  $\sigma_{ge}^2$  is the variance component due to GxE effects and  $\odot$  denotes the Hadamard product. Two models were tested: MGE<sub>30</sub> (model considering GxE using 30 days' worth of ECs), in which  $\mathbf{E} = \mathbf{E}_{30}$ , and MGE<sub>100</sub> (model considering GxE using 100 days' worth of ECs), where  $\mathbf{E} = \mathbf{E}_{100}$ .

### Multiple-trait model

The multiple-trait model (MTM) models GxE interactions across different environments (farms) using correlated traits. The eleven phenotypes observed on each environment were treated as distinct traits, which enabled estimation of genetic correlations across environments, as well as breeding or genetic values for each animal within each environment. Since the environmental effect is considered by treating phenotypes in the different environments as distinct traits, no environmental effect is explicitly fitted into the model:

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{11} \end{bmatrix} &= \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{11} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{11} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_{111} \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \\ \vdots \\ \mathbf{l}_{11} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{Z}_{21} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_{211} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{11} \end{bmatrix} \\ &+ \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{11} \end{bmatrix} \end{aligned} \quad (7)$$

where all terms were as previously defined and the subscript  $i$  represents the  $i$ th environment. Once again, assuming multivariate normality, the (co)variance matrix among random effects is:

$$\text{Var} \begin{bmatrix} \mathbf{l} \\ \mathbf{g} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \Sigma_0 \otimes \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_1 \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_2 \otimes \mathbf{I} \end{bmatrix} \quad (8)$$

where  $\Sigma_0$  represents the (co)variance matrix among litter effects across environments (a diagonal matrix in this study);  $\Sigma_1$  denotes the additive genetic relationship (co)variance matrix among environments; and  $\Sigma_2$  is the residual (co)variance matrix among environments (a diagonal matrix in this study). The GxE was considered

negligible, when the estimate of the genetic correlation between environments exceeded 0.80 [24, 25, 33], reflecting that trait expression was similar across different environments.

### Validation

The models were compared using the linear regression (LR) method [34]. Two validation schemes, were implemented: forward validation and environment validation.

#### Forward validation

First, all animals born in 2020 were chosen as focal individuals (i.e. genotyped individuals of interest); therefore, a single 'whole' dataset (phenotypes from animals born from 2009 to 2020) and a single 'partial' dataset (phenotypes from animals born from 2009 to 2019) were created. The genomic estimated breeding values (GEBV) for the focal group, calculated using the 'whole' data were compared with those calculated using the 'partial' dataset based on three statistics: accuracy (acc), bias ( $\delta$ ), and dispersion ( $d$ ). These statistics were calculated as follows:

$$\text{acc} = \sqrt{\text{cov}(\hat{\mathbf{g}}_p, \hat{\mathbf{g}}_w) / (1 - \bar{F}) \sigma_g^2} \quad (9)$$

$$\delta = (\bar{\hat{\mathbf{g}}}_p - \bar{\hat{\mathbf{g}}}_w) / \hat{\sigma}_g \quad (10)$$

$$b_1 = \text{cov}(\hat{\mathbf{g}}_p, \hat{\mathbf{g}}_w) / \text{var}(\hat{\mathbf{g}}_p) \quad (11)$$

where  $\hat{\mathbf{g}}_p$  and  $\hat{\mathbf{g}}_w$  represent the vectors of GEBV for the focal individuals from the partial and whole datasets, respectively;  $\bar{F}$  denotes the average pedigree-based inbreeding coefficient for the focal animals, and  $\hat{\sigma}_g$  indicates the estimated genetic standard deviation. For the MTM, the validation was performed separately for each trait and environment and then acc,  $\delta$ , and  $b_1$  were averaged across environments.

#### Environment validation

Multiple partial datasets were generated for the environment validation by masking phenotypes from each environment, one at the time. Then acc,  $\delta$ , and  $b_1$  were calculated for each masked environment using the same equations as for forward validation. The same reasoning was applied for the MTM: phenotypes for one environment were masked, GEBV were predicted, and validation statistics were calculated accordingly for the masked environment.

**Software**

For all models (MG, ME<sub>30</sub>, ME<sub>100</sub>, MGE<sub>30</sub>, MGE<sub>100</sub>, and MTM), variance components were estimated using restricted maximum likelihood (REML) with genomic information. This process combined the expectation maximization (EM-REML) and average information (AI-REML) algorithms using the BLUPF90+ software [35]. GEBVs were estimated by iteration on data using the BLUP90IOD3 software [35]. For the MTM, variance components were initially estimated using two-trait analyses. Subsequently, the complete (co)variance matrix was obtained by weighted bending using the ITSUMCOV software [36], with weighting factors equal to the number of records in each analysis [37]. The GxE matrices were calculated and inverted using an in-house Fortran program with parallel processing (OpenMP) and Intel Math Kernel Library (MKL).

**Results**

**Variance components and heritability**

Variance components and heritability estimates are shown in Table 3. Compared to MG, the variance component attributed to the environment (farm) ( $\sigma_e^2$ ) was greater for ME<sub>30</sub> and ME<sub>100</sub>. The increase in  $\sigma_e^2$  was approximately 4.8 times and 5.2 times when comparing MG with all the other models (ME<sub>30</sub>, ME<sub>100</sub>, MGE<sub>30</sub>, and MGE<sub>100</sub>) for ADG and BFT, respectively. For ADG,  $\sigma_e^2$  represented 14.6% of the phenotypic variance ( $\sigma_y^2$ ) for MG but increased to 47.8% for ME<sub>30</sub>. This same ratio ( $\sigma_e^2/\sigma_y^2$ ) was much higher for BFT: 43.6% for MG and increasing up to 80.7% for ME<sub>30</sub>. The variance component due to GxE ( $\sigma_{ge}^2$ ) accounted for 1.2% and 1.3% of the phenotypic variance in ADG for MGE<sub>30</sub> and MGE<sub>100</sub>, respectively. For BFT,  $\sigma_{ge}^2$  was responsible for 0.5% and 0.6% of the phenotypic variance for MGE<sub>30</sub> and MGE<sub>100</sub>, respectively. Heritability estimates for ADG ranged from 0.09 (MGE<sub>30</sub>) to 0.16 (MG), while

**Table 3** Estimates (SE) of variance components, heritability, and environmentability, along with the proportion of phenotypic variance due to GxE for all tested models by trait

Variance component <sup>1</sup>	Model				
	MG	ME <sub>30</sub>	ME <sub>100</sub>	MGE <sub>30</sub>	MGE <sub>100</sub>
ADG					
$\sigma_l^2$	1967.40 (51.80)	1967.20 (51.80)	1967.20 (51.80)	1901.50 (51.50)	1898.10 (51.46)
$\sigma_g^2$	1439.10 (72.37)	1439.80 (72.40)	1439.70 (72.40)	1384.70 (74.61)	1385.00 (71.56)
$\sigma_e^2$	1329.70 (610.22)	7127.16 (3244.33)	5675.50 (2581.81)	7024.17 (3206.59)	5595.28 (2549.52)
$\sigma_{ge}^2$				169.91 (26.84)	178.38 (27.14)
$\sigma_e^2$	4382.70 (51.80)	4382.60 (40.70)	4382.60 (40.70)	4322.20 (41.23)	4317.40 (41.24)
$h^2$	0.16 (0.01)	0.10 (0.01)	0.11 (0.01)	0.09 (0.01)	0.10 (0.01)
$e^2$	0.15 (0.06)	0.48 (0.01)	0.42 (0.03)	0.47 (0.03)	0.42 (0.02)
$ge^2$				0.01 (0.00)	0.01 (0.00)
BFT					
$\sigma_l^2$	0.38 (0.01)	0.38 (0.01)	0.38 (0.01)	0.36 (0.01)	0.36 (0.01)
$\sigma_g^2$	0.84 (0.04)	0.84 (0.04)	0.84 (0.01)	0.79 (0.04)	0.79 (0.04)
$\sigma_e^2$	2.11 (0.95)	11.38 (5.10)	10.63 (4.80)	11.29 (5.07)	10.53 (4.73)
$\sigma_{ge}^2$				0.07 (0.01)	0.08 (0.01)
$\sigma_e^2$	1.51 (0.02)	1.51 (0.02)	1.51 (0.02)	1.49 (0.02)	1.48 (0.02)
$h^2$	0.18 (0.04)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)
$e^2$	0.41 (0.14)	0.81 (0.08)	0.80 (0.09)	0.81 (0.08)	0.80 (0.09)
$ge^2$				0.01 (0.00)	0.01 (0.00)

MG traditional genomic best linear unbiased predictor (GBLUP) model, ME<sub>30</sub> GBLUP considering environmental effects correlated based on 30 days of weather information, ME<sub>100</sub> GBLUP considering environmental effects correlated based on 100 days of weather information, MGE<sub>30</sub> GBLUP considering GxE based on 30 days of weather information, MGE<sub>100</sub> GBLUP considering GxE based on 100 days of weather information, ADG average daily gain, BFT backfat thickness

<sup>1</sup>  $\sigma_l^2$  = variance due to litter;  $\sigma_g^2$  = additive genetic variance;  $\sigma_e^2$  = environmental variance;  $\sigma_{ge}^2$  = GxE variance;  $\sigma_e^2$  = residual variance;  $h^2$  = heritability;  $e^2$  = proportion of the phenotypic variance explained by environmental effect-environmentability;  $ge^2$  = proportion of phenotypic variance explained by the genotype by environmental interaction

for BFT, the estimates ranged from 0.06 ( $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$ ) to 0.18 (MG). Estimates of heritability for ADG and BFT in our subset data were lower and differed somewhat from those reported in the literature, which were based on pedigree-based models [38, 39] or single-step genomic-based models [40].

**Genetic correlations**

Genetic correlations ( $r_g$ ) for ADG among environments ranged from 0.03 to 1.00. Conversely, for BFT, these estimates were between 0.35 and 0.83. Among the possible 55 genetic correlation estimates among the 11 environments (based on the MTM), 17 exceeded 0.80 (31%) for ADG (ranging from 0.81 to 1.00), while only four (7%) did so for BFT (ranging from 0.80 and 0.83). Additionally, phenotypes for environments 9, 10, and 11 were weakly genetically correlated ( $r_g < 0.80$ ) with any of the other environments for ADG; in contrast, only environments 2, 4, 7, and 8 were strongly genetically correlated ( $r_g > 0.80$ ) for BFT. Complementary data showing the estimates of the genetic correlations for both traits in the MTM are provided in Additional file 1 (Table S1).

**Validation**

**Forward validation**

Except for MTM on BFT, MG,  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$  provided similar estimates of acc,  $\delta$ , and  $b_1$  of GEBV for both traits (Table 4). However, MTM GEBV were 5.5% less accurate for ADG and 27.9% less accurate for BFT. For BFT,  $ME_{100}$  had the best estimate of  $b_1$  (0.99), while the MTM resulted in underdispersed predictions ( $b_1 = 1.45$ ). For ADG,  $MGE_{30}$ ,  $MGE_{100}$ , and

**Table 5** Model fit information for all tested models by trait

Model	-2LogL <sup>1</sup>	AIC <sup>2</sup>	BIC <sup>3</sup>
ADG			
MG	410,039.34	410,047.34	410,081.24
$ME_{30}$	410,044.44	410,052.44	410,086.34
$ME_{100}$	410,043.86	410,051.60	410,085.76
$MGE_{30}$	409,971.06	409,981.06	410,023.43
$MGE_{100}$	409,962.91	409,972.91	410,015.28
BFT			
MG	114,144.04	114,152.04	114,185.52
$ME_{30}$	114,150.35	114,158.35	114,191.83
$ME_{100}$	114,150.70	114,158.70	114,192.18
$MGE_{30}$	114,058.73	114,068.73	114,110.58
$MGE_{100}$	114,048.48	114,058.48	114,100.33

MG traditional genomic best linear unbiased predictor (GBLUP) model,  $ME_{30}$  GBLUP considering environmental effects correlated based on 30 days of weather information,  $ME_{100}$  GBLUP considering environmental effects correlated based on 100 days of weather information,  $MGE_{30}$  GBLUP considering GxE based on 30 days of weather information,  $MGE_{100}$  GBLUP considering GE based on 100 days of weather information, MTM multiple trait GBLUP model, ADG average daily gain, BFT backfat thickness

<sup>1</sup> Log likelihood function

<sup>2</sup> Akaike information criteria

<sup>3</sup> Bayesian information criteria

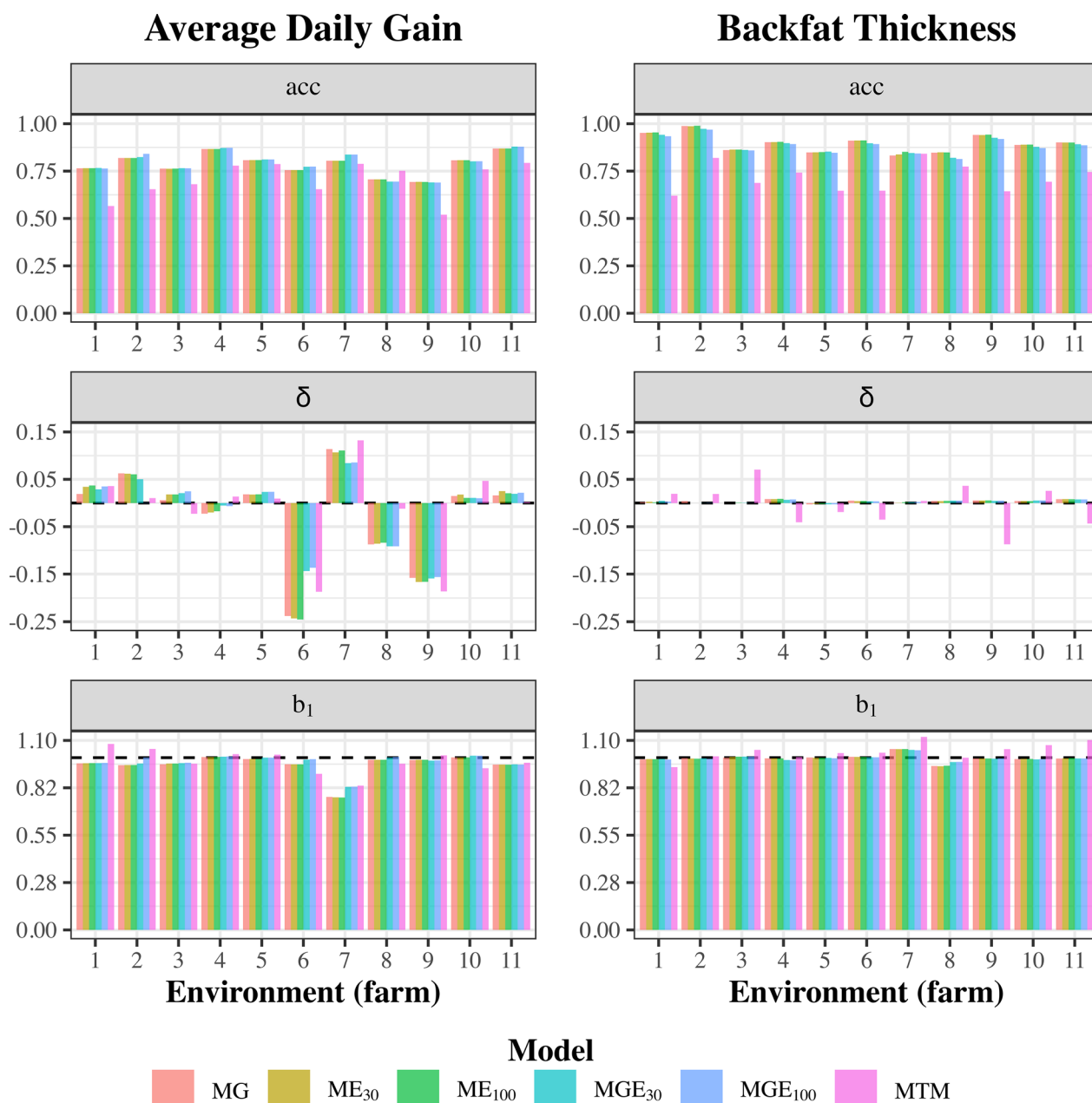
MTM exhibited negligible bias compared to the other models ( $\delta = 0.01$  or  $0.02$ , vs.  $0.00$ ). Remarkably, all estimates of  $\delta$  were negative for BFT, with MTM having the smallest estimate ( $-0.01$ ). Information on model fit is presented in Table 5. Models  $MGE_{100}$  and  $MGE_{30}$  had a better fit than models MG,  $ME_{30}$ , and  $ME_{100}$  for both traits.

**Table 4** Validation statistics for all tested models by trait for individuals born in 2020 as focal animals (forward validation)

Validation Statistic	Model					
	MG	$ME_{30}$	$ME_{100}$	$MGE_{30}$	$MGE_{100}$	MTM <sup>1</sup>
ADG						
Accuracy (acc)	0.79	0.80	0.79	0.80	0.79	0.75
Bias ( $\delta$ )	0.00	0.00	0.00	0.01	0.01	0.02
Dispersion ( $b_1$ )	0.94	0.94	0.94	0.95	0.95	0.97
BFT						
Accuracy (acc)	0.86	0.86	0.86	0.86	0.86	0.62
Bias ( $\delta$ )	-0.04	-0.05	-0.05	-0.05	-0.05	-0.01
Dispersion ( $b_1$ )	0.97	0.97	0.99	0.97	0.97	1.45

MG traditional genomic best linear unbiased predictor (GBLUP) model,  $ME_{30}$  GBLUP considering environmental effects correlated based on 30 days of weather information,  $ME_{100}$  GBLUP considering environmental effects correlated based on 100 days of weather information,  $MGE_{30}$  GBLUP considering GxE based on 30 days of weather information,  $MGE_{100}$  GBLUP considering GxE based on 100 days of weather information, MTM multiple trait GBLUP model, ADG average daily gain, BFT backfat thickness

<sup>1</sup> Average values across all environments



**Fig. 4** Validation statistics for all tested models by trait, considering all animals in a given environment (farm) as focal individuals (environment-validation). *MG* traditional genomic best linear unbiased predictor (GBLUP) model, *ME<sub>30</sub>* GBLUP considering environmental effects correlated based on 30 days of weather information, *ME<sub>100</sub>* GBLUP considering environmental effects correlated based on 100 days of weather information, *MGE<sub>30</sub>* GBLUP considering genotype by environment interaction (G×E) based on 30 days of weather information, *MGE<sub>100</sub>* GBLUP considering G×E based on 100 days of weather information, *MTM* multiple-trait GBLUP model, *ADG* average daily gain, *BFT* backfat thickness

**Environment validation**

For ADG, the models *MG*, *ME<sub>30</sub>*, *ME<sub>100</sub>*, *MGE<sub>30</sub>*, and *MGE<sub>100</sub>* showed similar estimates of *acc* for all environments, except for environment 7 (Fig. 4). For BFT, *MGE<sub>30</sub>* and *MGE<sub>100</sub>* exhibited marginally lower *acc* across all environments. The *MTM* had the smallest

*acc* of GEBV in all environments for both traits, except for ADG in environment 8. Bias was negligible for BFT for all models, while environments 6, 7, 8, and 9 displayed the highest bias levels for ADG. Additionally, the estimate of *b<sub>1</sub>* was close to 1.00 for BFT for all environments, except for environment 7. For BFT, GEBV

from the MTM were either under-dispersed (environments 3, 5, 6, 7, 9, 10, and 11) or overdispersed (environments 1 and 8). Predictions for ADG were generally overdispersed across models and environments, except for environments 1, 2, 4, and 5 under the MTM.

## Discussion

This study aimed to evaluate the validity of high-dimensional environmental data for modeling correlated environments and GxE related to production traits in pigs. The results indicated that incorporating ECs into the genomic prediction model did not significantly enhance the prediction accuracy of GEBV for ADG and BFT, despite an improvement in model fit. In this study, we used external ECs while the animals were kept indoors, which may have limited the benefits of using these ECs. Our findings suggest that, although incorporating ECs into the prediction model improves its overall fit, the advantage for genomic predictions, specifically prediction accuracy, remains limited. However, we found some increase in prediction accuracy for BFT when comparing ME and MGE versus MTM, suggesting that the benefit of using ECs might be trait-dependent, thus highlighting the need for future research with in-barn ECs and other traits.

Purebred animals are typically raised in better environments than commercial pigs. However, environmental differences can exist even when building conditions (e.g. cooling and humidity) and animal conditions (e.g. density) are controlled. Previous studies have shown that even purebred pigs are susceptible to heat stress [2] and show some degree of GxE [17]. We also found significant effects of external weather conditions, although we used only purebred animals from well-controlled nucleus herds in this study. Therefore, we speculate that external ECs can interfere with the microclimate within barns, which may indirectly affect the animals' performance.

## Environmental covariates

Using THI as the sole EC, Fragomeni et al. showed that 30 days was sufficient to model the response to heat stress [2] and, subsequently, the GxE [17]. In recent studies on dairy cattle, McWhorter et al. and Campos et al. averaged THI over 5 and 3 days, respectively, as an environmental descriptor [18, 19]. The THI is strongly correlated with temperatures [21, 22] and all ECs were correlated (to some degree) in this study (Fig. 2). Therefore, we expected that regressing the phenotype on the average of an EC would not result in significant results in the last 30 days before the phenotype is collected. However, we found an association between the phenotype and the ECs for up to 100 days. This outcome indicates that

the environmental impact on the phenotype is not solely a result of immediate climate conditions but is also influenced by continuous exposure.

Campos et al. argued that wind speed and solar radiation might influence heat stress conditions experienced by animals, especially when kept in outdoor conditions [19]. They also suggested that in-barn conditions could limit this influence. All temperatures in this study (T, Td, Tw, and Ts) showed a weak negative correlation with Ws (Fig. 2). Conversely, a strong negative correlation with Wd was found (Fig. 2). Ws and Wd exhibited a moderate positive correlation (Fig. 2), indicating that the temperature can decrease rapidly when the external temperature is high, depending on wind direction and speed. This observation helps clarify the positive regression coefficient of the phenotype on Ws and Wd. Furthermore, it explains why all EC related to temperature had less impact than Ws.

Given the correlations among H and all ECs related to temperature (Fig. 2), one could expect little to no influence from the increased humidity when temperature ECs are elevated. The correlations among R and all temperature ECs (Fig. 2) indicate that an increase in temperature generally leads to more rain, resulting in a rise in H (moderate positive correlation between R and H), potentially triggering heat-stress responses. The THI effectively captures this behavior by combining increased H (due to more R) with rising temperatures. By not using the ECs through an index calculation, the dynamic nature of climate conditions was incorporated into the herd effect and the GxE term. However, fitting multiple ECs and modeling a random regression coefficient for each would increase the complexity of the model, the number of parameters, and the computing cost, which explains the popularity of the THI as a singular environmental gradient in previous studies [17, 18, 20]: rather than fitting one regression for each EC, only one is needed. Moreover, by modeling only the linear random regression coefficient of the HL, there is no need to use orthogonal polynomials or to standardize the covariates [41], which are needed when using several ECs.

Considering each EC for each day as a different covariate through the traditional reaction-norm approach would necessitate at least 240 (eight ECs over 30 days) or 800 (eight ECs over 100 days) random regression coefficients, especially as we used daily ECs to construct E. By modeling the ECs using covariance functions [29], the number of parameters was significantly reduced, resulting in an equivalent model as a standard reaction norm that fits one linear random regression coefficient for each EC. Incorporating daily ECs into the calculation of E also allowed us to account for the temporal variability within either 30- or 100-day periods. In preliminary analyses

(data not shown), we observed that using over 100 days of daily ECs yielded the same variance components as those presented in Table 3 and resulted in similar validation statistics for GEBV (Table 4).

Jarquín et al. accounted for temporal environmental effects by defining each environment as a combination of location and year [29]. In contrast, our approach integrated the temporal effect with the environmental gradient on a daily basis. In Jarquín et al.'s approach, within-year differences, including seasonal or daily variations, are not accounted for, whereas our approach averages differences across years within each location. Additionally, the yearly effect on phenotype was included in the CG, which could have introduced confounding between CG and the random environment effect if we had defined the environment as farm concatenated with year. Although we do not anticipate significant annual environmental variations to affect the studied traits and given that the growth trial duration influences the phenotype, the potential to account for both within- and across-year variation warrants further investigation.

#### Variance components and heritability

When not fitting the random environmental effect in the model (data not shown), heritability estimates were 0.19 for ADG and 0.30 for BFT. Previous works with different populations of purebred and crossbred pigs reported heritability estimates for ADG varying from 0.15 to 0.30 and for BFT from 0.41 to 0.56 [38, 39]. It is worth noting that differences in heritability estimates could arise because of data structure and editing, as well as from the model used. We used single-trait genomic models in our study, while Steyn et al. and Leite et al. used multiple-trait pedigree-based models [38, 39]. Using or omitting genomic information when estimating variance components can lead to biased estimates, particularly if the population is undergoing selection [42–44]. It is, therefore, not unrealistic to assume such factors could result in the lower heritability estimates found in our study, along with the fact that our data represents a subset of the population.

When environments were included in the model as uncorrelated random effects (MG), there was a reduction in the variance components attributable to litter ( $\sigma_l^2$ ) and additive genetic ( $\sigma_g^2$ ) effects (data not shown). There was also an increase in phenotypic variance, while the residual variance remained unchanged compared with a model without herd effects. Thus, this reduced heritability estimates (0.16 for ADG and 0.17 for BFT). This indicates potential confounding between herd and litter effects [45]. Animals typically do not have records in multiple environments unless there is a temporal change and litters also do not exist in multiple herds simultaneously. Hence, the presence of animals and litters within a

single herd could explain the reduction in  $\sigma_l^2$  and  $\sigma_g^2$  when environment was included in the model as a random effect.

Confounding between effects that are included in a model makes it difficult to obtain separate estimates for them. Gelfand and Sahu noted that Bayesian non-identifiability is equivalent to the lack of identifiability in likelihood methods [46]. For a model with a parameter vector  $\theta' = [\theta_1 \theta_2]'$ , Bayesian non-identifiability means that what can be inferred about  $\theta_2$  depends on  $\theta_1$  [47]. Identifiability conditions in likelihood-based approaches are well studied [48–52]. For a variance component vector  $\theta' = [\theta_1 \theta_2 \dots \theta_n]'$  and associated (co)variance matrices  $V_1, V_2, \dots, V_n$ , identifiability requires a positive and definite restricted maximum likelihood (REML) information matrix, which implies that the matrices  $V$  need to be linearly independent [49, 50]. In our models, the average information REML (AI-REML) information matrix was positive-definite, indicating that the models were identifiable. However, the environment (co)variance matrix showed some dependence on the litter (co)variance matrix due to litter nested within the environments. This issue could be resolved by observing the same animal across environments (e.g. through cloning, which is often not feasible) or by removing litter effects. However, the latter will introduce bias.

Using  $\text{Var}(\mathbf{e}) = \mathbf{E}\sigma_e^2$  rather than  $\text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$  resolved identifiability issues by allowing environment-specific variation. Notably, the environment (co)variance matrix rank improved when environmental correlations were considered (MG: rank 11; ME<sub>30</sub>: rank 15; and ME<sub>100</sub>: rank 20). However, in the ME<sub>30</sub> model, environments were completely independent of litter, resulting in the smallest condition number of the AI-REML information matrix. This aligns with the suggestion that 30 days of weather data most effectively captures GxE and heat-stress effects in pigs [2, 17]. This also suggests fitting ECs in the genomic prediction model may provide theoretical modeling improvements. However, its predictive benefit may be limited.

The lower heritability estimates for ME and MGE compared to MG was similar to that reported by Makanjuola et al. [1] and Cuyabano et al. [13]. Furthermore, Makanjuola et al. found comparable results for the same traits in the same population but instead used the THI to correlate herd effects [1]. Conversely, Tiezzi et al. did not find increases in  $\sigma_e^2$  when comparing a model that included only environmental effects to one that included both environmental and genetic effects [14]. Tiezzi et al. [14] used a sire model, while Makanjuola et al. [1] and Cuyabano et al. [13] used an animal model. Under a sire

model, records on different daughters act as repeated records for their sire. This helps mitigate potential partial confounding between random effects, as sires typically have progeny across multiple herds.

If the number of observations per level of a given environmental effect is sufficiently large, modeling it as fixed or random does not significantly impact prediction accuracy of EBV [53]. However, when the number of records per subclass is limited, modeling it as random is preferred, as a reduction in prediction error variance is expected [54, 55]. Moreover, when modeling an effect such as this as a fixed effect, its variance is not included in the heritability calculation; therefore, when treated as random, its variance should also be omitted in the heritability estimation. When taking this into account, heritability estimates in our study remained stable across all models (data not shown), regardless of whether  $\sigma_e^2$  increased or not.

Pigs face not only climatic factors, but also pathogen load, management practices, and interactions with other animals are also part of the environmental effect. It may be that the assumption that all environment variance is accounted for in the E matrix is too strong, causing  $\sigma_e^2$  to increase. To understand this, models  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$  were re-run with an extra uncorrelated random effect for the environment ( $ME_{30}^*$ ,  $ME_{100}^*$ ,  $MGE_{30}^*$ , and  $MGE_{100}^*$ ) (Additional file 2, Table S2, and Additional file 3, Table S3). Thus, the environment was split into two parts: one modeled by the climate (ECs) and one uncorrelated part. This did not increase the phenotypic variance (i.e.  $\sigma_y^2$ ) hence, the estimate of heritability was the same for all models. When we examined the model fit (Table 5 and Table S3), the benefit of including this non-climate (or microenvironment) effect was confirmed. Jarquín et al. suggested this approach to account fully for the environmental effect on grain yield data [29]. Validation statistics (accuracy, bias, and dispersion) did not change for these models ( $ME_{30}^*$ ,  $ME_{100}^*$ ,  $MGE_{30}^*$ , and  $MGE_{100}^*$ ) in our data.

Initially, modeling correlated environmental effects through ECs was proposed in the context of plant breeding [29] and it has been successfully applied in several crops under selection, using pedigree- and genomic-based prediction [56–61]. These studies included repeated records both within and across environments. In plants, the ability to propagate through sexual and asexual reproduction (i.e. vegetative propagation and cloning) facilitates testing the same genotype across several environments. Using (co)variance structures for the environmental effect enables prediction of an individual's performance in environments where their phenotype was not observed [58]. However,

this prediction has low accuracy if no individual is measured in multiple environments, as was the case for our data.

#### Genetic correlations and genotype-by-environment interaction

For ADG, 31% of the genetic correlation estimates among environments exceeded 0.80, whereas only 7% did so for BFT. Although the MTM was feasible in our study, when the number of environments is too large, treating each environment as a different trait is impossible because (1) the genetic (co)variance matrix can become non-positive-definite and (2) it requires more computing power. Therefore, our approach using ECs is preferable in such cases as it addresses correlated environmental effects and GxE by fitting additional random effects without significantly increasing computing costs. Fitting an MTM also requires that each environment has sufficient records, which is not always achievable. Record insufficiency can decrease the prediction accuracy in MTM and bias estimates of variance components.

When the number of records for each environment (i.e. farm) is limited, estimates of variance components have a high standard error. If some covariances are poorly estimated, the animals from those environments do not fully contribute to the accuracy of the animals in the other environments, leading to lower accuracy when using the MTM. One way to address this issue is by removing farms with too few records (e.g., fewer than 2000). However, doing so would reduce the total number of records and the number of environments for the MTM, making comparisons with other models unfair, as those models would have more records per farm and more farms overall. Moreover, the additive genetic (co)variance matrix used in the MTM had to be bend to ensure it was positive-definite. This adjustment could disrupt the covariance estimates when the number of records per trait (or environment, in this case) is smaller [36, 37]. In addition, the estimates of genetic correlations between environments in the MTM may not accurately reflect the true genetic correlations between environments. Utilizing pedigree information would allow for the full dataset that includes phenotyped animals that are not genotyped to be analyzed for variance component estimation.

In the present study, the ranking of the animals was consistent across all environments for ADG, indicating that although GxE was identified (see log-likelihood in Table 5), selection and gain were not compromised when ignoring these interactions. However, small changes in the ranking were observed for BFT, suggesting that the type of GxE for BFT differs from that of the ADG, potentially due to management, feeding, and measurement systems affecting BFT more than ADG. Robertson stated

that ‘no interaction means a genetic correlation of unity’, but considering its standard error, a correlation smaller than 0.80 would hold biological significance [33]. In later studies, both Mulder and Bijma, and Mulder et al. found no improvement in genetic gain when considering two environments individually if their genetic correlation was greater than 0.80 [24, 25]. These findings may suggest that evaluating the environments individually could maximize the gain; however, the amount of re-ranking across environments in our study for BFT was small and thus did not justify selection within a specific environment. Although, the ranking of animals based on GEBV may remain identical within a trait between environments, the scale of the GEBV might influence the relative contribution of that single trait on the index.

GxE can be defined as the differential expression of genotypes across environments [62], indicating that the phenotype for the same genotype can vary depending on the environment. With presence of GxE, the genetic correlation for a given trait between the environments will deviate from unity. Factors such as location, management, housing, time, and climate can be used to account for the environmental effect [17, 18, 62]. Furthermore, if there is a continuous factor (e.g. time, weather conditions, THI), a random regression model can be used to estimate a slope of the GEBV over the environmental gradient [18]. If there is no gradient (i.e. the environments are categorical), an MTM can be employed to include GxE in the evaluation [27]. However, changes in the GEBV resulting from the environmental effect depend on the type of interaction and may vary according to the diversity of genotypic and environmental effects [62].

Haldane specified several different forms that GxE might take [63]. Lerner grouped Haldane’s classes of interaction into two distinct groups [64]: linear interactions, where genetic and environmental effects are additive, and non-linear otherwise. If the change in genotype from one environment to another is consistent across all genotypes, there will be no re-ranking, and the genetic correlations among environments may remain high. However, the genetic correlations will be less than one when changes differ among genotypes, regardless of whether re-ranking occurs.

Jarquín et al. demonstrated that modeling the environmental effects and GxE by (co)variance functions is equivalent to a reaction norm [29]. In a reaction-norm model, the GEBV is predicted along with the random gradient [17, 18], which, in the context of this study, corresponds to weather information. Following this reasoning, the ‘ge’ breeding value can be interpreted as the amount of change in the GEBV attributed to the environmental effect. Since this value is specific to each environment, it can provide targeted insights for selection

decisions. For example, it can guide breeders when sending boars or semen to another herd. Fragomeni et al. applied reaction norms to their investigation into GxE regarding heat stress in pigs [17]. Their study reported, for purebred data, that reaction norms were not effective in demonstrating evidence of GxE since there is rigorous environmental control in nucleus farms. The ECs used in this study reflect outdoor conditions, while the animals were kept indoors. It is possible that, although the MTM showed some GxE, the (co)variances approach as a reaction norm yields better results if the ECs were obtained inside the barn.

The main limitation of the models we used in this study is related to the mode of interaction, in that ‘only multiplicative reaction norm model’ was considered [29]. This model becomes an approximation in the presence of different interaction forms between genetic marker effects and ECs. Correlating herds in an animal breeding context also approximates the uncertain relationship between herds [13]. While modeling environments as correlated random effects accounted for a significant portion of phenotypic variance without negatively impacting breeding value prediction accuracy, it was still not descriptive enough to increase prediction accuracy. It is worth noting that while there was a reduction in heritability due to an increase in the environmental variance ( $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$ ), our study did not show a reduction in additive genetic variance and/or prediction accuracy. This outcome may be attributed to the absence of re-ranking of GEBV across environments. If GxE is strong and re-ranking occurs, ignoring these interactions and the potential relationship among herds can result in heritability overestimation and inflated values of expected genetic gain [1, 13].

### Validation

Modeling environments as correlated random effects can increase predictivity [1, 13, 14], defined as the correlation between adjusted phenotypes from a full dataset and GEBV from a partial dataset. This approach relies on model fit and the quality of estimation of fixed effects, reflecting the model’s ability to predict phenotypes [34]. Conversely, the LR method tests the impact of recent data on breeding value predictions [65] and remains robust even with mis-specified effects or heritability [66]. In forward validation, our accuracies surpassed those of Steyn et al. [38] and Leite et al. [39] and aligned with those of Jang et al. [67, 68]. No accuracy improvements were observed across MG, ME, and MGE, indicating that adding weather information to model GxE has minimal benefit for breeding value prediction for ADG and BFT in pigs.

However, for BFT, alternative models outperformed the MTM. This finding suggests that treating phenotype across these environments as a single trait with an environmental effect improves accuracy for traits that have low genetic correlations across environments. Regarding environment validation, the models MG, ME, and MGE outperformed MTM, exhibiting higher accuracy, lower bias, and less dispersion issues. The MTM was constrained by data scarcity in certain environments, which hindered validation performance, particularly those with fewer records. Moreover, bias, accuracy, and dispersion depend on data availability, helping to explain MTM's lower performance. Modeling phenotype across different environments as a single trait increases the information available, which is especially beneficial for smaller environments. However, in the case of MTMs and when genetic correlations between environments are low, there is no information sharing, leading to reduced prediction accuracy in environments with limited data.

#### Practical remarks

The occurrence of GxE observed in plants and animals, although its extent and nature vary. Plants can exhibit a high degree of phenotypic plasticity, adjusting their development based on environmental cues. For example, a plant may exhibit different growth forms, leaf shapes, or flowering times depending on factors like light, temperature, and soil conditions. Conversely, animals generally have a more limited ability to adapt to different environments, with many animal traits strongly influenced by intrinsic genetic factors; therefore, environmental effects may play a smaller role.

In addition, livestock animals are often more mobile than plants and their environment is controlled to some degree; thus, they may not be as directly exposed to or influenced by their immediate environment during trait development. Nevertheless, animals can exhibit some degree of GxE that may or may not change the ranking of GEBV from one environment to another. However, it is crucial to note that GxE is a complex, multifaceted phenomenon that varies among species, populations, and traits. Therefore, it would be inaccurate to claim that GxE is entirely absent in animals or to ignore its impacts; rather, it is suggested that the degree, nature, and extent of these interactions can widely vary.

Currently, most genetic predictions for purebred pigs partially consider differences in the production environments (e.g. management, climate, and social aspects of production) by fitting the farm in the fixed CG. Moreover, GxE interactions are considered negligible for production traits such as ADG and BFT. However, if GxE interactions are present but ignored, prediction accuracy

may be compromised, along with over or under-dispersed predictions for young individuals.

Our approach considered high-dimensional environmental data to account for correlations among herds and GxE. This high dimensionality arose from using eight different ECs, combined with daily information. This methodology allowed us to estimate an 'overall' breeding value and an 'environmental specific' breeding value. The environmental specific EBV should be interpreted as the change in the overall breeding value attributable to environmental effects. If a given genotype has records in each environment, this approach can potentially increase prediction accuracy. We foresee that this model will perform better when there are repeated records, but this remains to be evaluated. MGE<sub>i</sub> models (or any model that considers GxE) predicts one environmental breeding value for each environment for each animal. We recommend utilizing these environmental-specific EBV for mating and culling decisions while reserving the overall breeding value for selection decisions. From a practical perspective, once new phenotypes and/or environments are added to the data, the environmental covariance structure should be updated, leading to a refresh of all model configurations.

#### Conclusions

Modeling environments (herds) as a correlated random effect does not provide any benefit for genomic predictions in pigs. Furthermore, the computing requirements also increase as the model becomes more complex. For traits in which GxE is stronger, modeling GxE by covariance functions is more effective than using the MTM due to the amount of information each environment contributes to the data. Although there may be no immediate practical benefit, models that incorporate GxE can improve mating decisions by predicting how the future offspring of a particular boar or sow perform in different environments. Provided there are records for the same genotype across all environments, the first-order multiplicative (co)variances approach used in this study is expected to yield higher accuracies of EBV than traditional models when there is GxE. Moreover, the traditional MTM in the presence of GxE requires larger datasets and better genetic connections among environments to provide accurate breeding value predictions. Lastly, incorporating environments (herd) into the model as an uncorrelated random effect is sufficient to account for environmental effects and GxE in pigs.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-025-00974-2>.

**Additional file 1: Table S1.** Heritability estimates (diagonal, bold) and genetic correlations among all environments (after bending) according to the studied trait. Results contain the bent correlation matrix among all eleven environments for each trait (ADG or BFT) for the multiple trait model.

**Additional file 2: Table S2.** Variance components (SE), heritability (SE), and environmentability (SE) estimates, along with the proportion of phenotypic variance due to GxE for all tested models, including an extra uncorrelated environmental random effect for  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$  according to the studied trait. Description: Results contain estimated variance components for all models when an extra uncorrelated random environmental effect was included for  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$ .

**Additional file 3: Table S3.** Model fit information for all tested models, including an extra uncorrelated environmental random effect for  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$  according to the studied trait. Results contain estimated variance components for all models when an extra uncorrelated random environmental effect was included for  $ME_{30}$ ,  $ME_{100}$ ,  $MGE_{30}$ , and  $MGE_{100}$ .

### Acknowledgements

Discussion and edits from Andres Legarra, Ignacio Aguilar, Jennifer Richter, and Joe-Menwer Tabet are highly appreciated.

### Author contributions

FB, DL, and ZV conceived the study. CC and JH prepared and curated the dataset. FB performed the formal analysis, developed the methodology with DL and ZV, and drafted the manuscript. DL, JHi, and IM contributed software tools and resources. DL and ZV coordinated the project and acquired funding. ZM, DL, and IM supervised the work. FB, DL, JHi, and ZM validated the results. FB and ZM prepared the visualizations. FB, DL, JHi, CC, JH, IM, and ZV reviewed and edited the manuscript. All authors read and approved the final manuscript.

### Funding

The Pig Improvement Company partially supported this work.

### Availability of data and materials

The used datasets are the property of the Pig Improvement Company and are not available publicly.

### Declarations

#### Ethics approval and consent to participate

Data were obtained from existing databases; therefore, approval from the Animal Care and Use Committee was unnecessary for this study.

#### Consent for publication

Not applicable.

#### Competing interests

Ching-Yi Chen and Justin Holl are employees of the Pig Improvement Company. The authors declare no real or perceived conflicts of interest.

Received: 3 August 2024 Accepted: 22 May 2025

Published online: 05 June 2025

### References

- Makanjuola F, Rovere G, Cuyabano BCD, Lee SH, Gondro C. Including environmental variables in genomic models for carcass traits in Hanwoo

- beef cattle. In: Proceedings of the 12th world congress on genetics applied to livestock production: 3–8 July 2022: Rotterdam; 2022
- Fragomeni BO, Lourenco DA, Tsuruta S, Andonov S, Gray K, Huang Y, et al. Modeling response to heat stress in pigs from nucleus and commercial farms in different locations in the United States. *J Anim Sci.* 2016;94:4789–98.
- Collier RJ, Gebremedhin KG. Thermal biology of domestic animals. *Annu Rev Anim Biosci.* 2015;3:513–32.
- Ingram DL. Stimulation of cutaneous glands in the pig. *J Comp Pathol.* 1967;77:93–8.
- Renaudeau D, Collin A, Yahav S, de Basilio V, Gourdine JL, Collier RJ. Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Animal.* 2012;6:707–28.
- Baumgard LH, Rhoads RP Jr. Effects of heat stress on postabsorptive metabolism and energetics. *Annu Rev Anim Biosci.* 2013;1:311–37.
- St-Pierre NR, Cobanov B, Schnitkey G. Economic losses from heat stress by US livestock industries. *J Dairy Sci.* 2003;86:E52–77.
- Brown-Brandl TM, Eigenberg RA, Nienaber JA, Kachman SD. Thermoregulatory profile of a newer genetic line of pigs. *Livest Prod Sci.* 2001;71:253–60.
- Schmitz F, Everett RW, Quaas RL. Herd-year-season clustering. *J Dairy Sci.* 1991;74:629–36.
- Chauhan VPS, Thompson R. Dairy sire evaluation using a 'rolling months' model. *J Anim Breed Genet.* 1986;103:321–33.
- Schaeffer LR. Contemporary groups are always random; 2009. <https://animalbiosciences.uoguelph.ca/~lrs/piksLRS/ranfix.pdf>. Accessed 16 Nov 2023.
- Schaeffer LR. Contemporary Groups. In: Schaeffer LR, editor. *Animal models*. Self-Published; 2019. p. 31–2.
- Cuyabano BCD, Rovere G, Lim D, Kim TH, Lee HK, Lee SH, et al. GPS coordinates for modeling correlated herd effects in genomic prediction models applied to Hanwoo Beef Cattle. *Animals (Basel).* 2021;11:2050.
- Tiezzi F, de los Campos G, Parker Gaddis KL, Maltecca C. Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J Dairy Sci.* 2017;100:2042–56.
- Selle ML, Steinsland I, Powell O, Hickey JM, Gorjanc G. Spatial modeling improves genetic evaluation in smallholder breeding programs. *Genet Sel Evol.* 2020;52:69.
- Santana TEZ, Veroneze R, Carvalho Alves AA, Menezes GRO, Rosa GJJM. PSXII-8 gaussian kernel based on geographic information to model farm effects in genetic evaluation of pasture-raised beef cattle. *J Anim Sci.* 2022;100:209.
- Fragomeni BO, Lourenco DA, Tsuruta S, Bradford HL, Gray KA, Huang Y, et al. Using single-step genomic best linear unbiased predictor to enhance the mitigation of seasonal losses due to heat stress in pigs. *J Anim Sci.* 2016;94:5004–13.
- McWhorter TM, Sargolzaei M, Sattler CG, Utt MD, Tsuruta S, Misztal I, et al. Single-step genomic predictions for heat tolerance of production yields in US Holsteins and Jerseys. *J Dairy Sci.* 2023;106:7861–79.
- Campos IL, Chud TCS, Oliveira HR, Baes CF, Cánovas A, Schenkel FS. Using publicly available weather station data to investigate the effects of heat stress on milk production traits in Canadian Holstein cattle. *Can J Anim Sci.* 2022;102:368–81.
- Bohmanova J, Misztal I, Cole JB. Temperature-humidity indices as indicators of milk production losses due to heat stress. *J Dairy Sci.* 2007;90:1947–56.
- Dikmen S, Hansen PJ. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? *J Dairy Sci.* 2009;92:109–16.
- Dado-Senn B, Ouellet V, Lantigua V, Van Os J, Laporta J. Methods for detecting heat stress in hutch-housed dairy calves in a continental climate. *J Dairy Sci.* 2023;106:1039–50.
- Hay EH, Roberts A. Genomic evaluation of genotype by prenatal nutritional environment interaction for maternal traits in a composite beef cattle breed. *Livest Sci.* 2019;229:118–25.
- Mulder HA, Bijma P. Benefits of cooperation between breeding programs in the presence of genotype by environment interaction. *J Dairy Sci.* 2006;89:1727–39.
- Mulder HA, Veerkamp RF, Ducro BJ, van Arendonk JAM, Bijma P. Optimization of Dairy cattle Breeding programs for different environments with genotype by environment interaction. *J Dairy Sci.* 2006;89:1740–52.

26. Burgueño J, de los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotype  $\times$  environment interaction using pedigree and dense molecular markers. *Crop Sci.* 2012;52:707–19.
27. Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, et al. Increased prediction accuracy in wheat breeding trials using a marker  $\times$  environment interaction genomic selection model. G3 (Bethesda). 2015;54:569–82.
28. Zhao F, Wang L, Xu S. Identification of QTL-by-environment interaction by controlling polygenic background effect. *J Genet Genomics.* 2025. <https://doi.org/10.1016/j.jgg.2025.01.003>.
29. Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet.* 2014;1273:595–607.
30. Cuevas J, Crossa J, Soberanis V, Pérez-Elizalde S, Pérez-Rodríguez P, Campos GDL, et al. Genomic prediction of genotype  $\times$  environment interaction kernel regression models. *Plant Genome.* 2016. <https://doi.org/10.3835/plantgenome2016.03.0024>.
31. Aguilar I, Misztal I, Tsuruta S. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs of the 10th world congress on genetics applied to livestock production: 17–22 August 2014; Vancouver; 2014.
32. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;9:4414–23.
33. Robertson A. The sampling variance of the genetic correlation coefficient. *Biometrics.* 1959;15:469–85.
34. Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol.* 2018;50:53.
35. Lourenco D, Tsuruta S, Aguilar I, Masuda Y, Bermann M, Legarra A, et al., editors. Recent updates in the BLUPF90 software suite. In: Proceedings of the 12th world congress on genetics applied to livestock production; 3–8 July 2022; Rotterdam; 2022.
36. Henshall JM, Meyer K, editors. 'PDMATRIX'—programs to make matrices positive definite. In: Proceedings of the 7th world congress on genetics applied to livestock production; 19–23 August 2002; Montpellier; 2002.
37. Meyer K, Johnston DJ, Graser H-U. Estimates of the complete genetic covariance matrix for traits in multi-trait genetic evaluation of Australian Hereford cattle. *Aust J Agric Res.* 2004;55:195–210.
38. Steyn Y, Lourenco DA, Chen C-Y, Valente BD, Holl J, Herring WO, et al. Optimal definition of contemporary groups for crossbred pigs in a joint purebred and crossbred genetic evaluation. *J Anim Sci.* 2021;99:skaa396.
39. Leite NG, Chen C-Y, Herring WO, Holl J, Tsuruta S, Lourenco D. Leveraging low-density crossbred genotypes to offset crossbred phenotypes and their impact on purebred predictions. *J Anim Sci.* 2022;100:skac359.
40. Salek Ardestani S, Jafarikia M, Sargolzaei M, Sullivan B, Miari Y. Genomic prediction of average daily gain, back-fat thickness, and loin muscle depth using different genomic tools in canadian swine populations. *Front Genetics.* 2021;12: 665344.
41. Schaeffer LR. Application of random regression models in animal breeding. *Livest Prod Sci.* 2004;86:35–45.
42. Gao H, Madsen P, Aamand GP, Thomasen JR, Sørensen AC, Jensen J. Bias in estimates of variance components in populations undergoing genomic selection: a simulation study. *BMC Genomics.* 2019;20:956.
43. Cuyabano BCD, Sørensen AC, Sørensen P. Understanding the potential bias of variance components estimators when using genomic models. *Genet Sel Evol.* 2018;50:41.
44. Cesarani A, Pocrnic I, Macciotta NPP, Fragomeni BO, Misztal I, Lourenco DAL. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. *J Anim Breed Genet.* 2019;13:40–50.
45. Walsh B, Lynch M. Associative effects: Competition, social interactions, group and kin selection. In: *Evolution and selection of quantitative traits.* 1st ed. New York: Oxford University Press; 2018.
46. Gelfand AE, Sahu SK. Identifiability, improper priors, and gibbs sampling for generalized linear models. *J Am Stat Assoc.* 1999;94:247–53.
47. Gianola D. Priors in whole-genome regression: the bayesian alphabet returns. *Genetics.* 2013;1943:573–96.
48. Rothenberg TJ. Identification in parametric models. *Econometrica.* 1971;39:577–91.
49. Miller JJ. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann Stat.* 1977;54:746–62.
50. Jiang J. REML estimation: asymptotic behavior and related topics. *Ann Stat.* 1996;241:255–86.
51. Cantet RJ, Cappa EP. On identifiability of (co)variance components in animal models with competition effects. *J Anim Breed Genet.* 2008;1256:371–81.
52. Cheng J, Janssens S, Buys N. Full sib pens of pigs are not suitable to identify variance component of associative effect: a simulation study using Gibbs Sampling. *BMC Genet.* 2009;10:9.
53. Henderson CR. Sire evaluation and genetic trends. *J Anim Sci.* 1973;1973:10–41.
54. Ugarte E, Alenda R, Carabaño MJ. Fixed or random contemporary groups in genetic evaluations. *J Dairy Sci.* 1992;751:269–78.
55. Van Vleck LD. Contemporary groups for genetic evaluations. *J Dairy Sci.* 1987;70:2456–64.
56. Sukumaran S, Crossa J, Jarquín D, Lopes M, Reynolds MP. Genomic prediction with pedigree and genotype  $\times$  environment interaction in spring wheat grown in South and West Asia, North Africa, and Mexico. G3 (Bethesda). 2017;7:481–95.
57. Pérez-Rodríguez P, Crossa J, Bondalapati K, De Meyer G, Pita F, Campos GDL. A Pedigree-based reaction norm model for prediction of cotton yield in multi-environment trials. *Crop Sci.* 2015;55:1143–51.
58. Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, Saint-Pierre C, et al. Genomic prediction of gene bank wheat landraces. G3 (Bethesda). 2016;6:1819–34.
59. Velu G, Crossa J, Singh RP, Hao Y, Dreisigacker S, Perez-Rodriguez P, et al. Genomic prediction for grain zinc and iron concentrations in spring wheat. *Theor Appl Genet.* 2016;129:1595–605.
60. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, et al. Single-step genomic and pedigree genotype  $\times$  environment interaction models for predicting wheat lines in international environments. TPG. 2017. <https://doi.org/10.3835/plantgenome2016.09.0089>.
61. Saint Pierre C, Burgueño J, Crossa J, Fuentes Dávila G, Figueroa López P, Solís Moya E, et al. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Sci Rep.* 2016;6:27312.
62. Mathur PK, Horst P. Methods for evaluating genotype-environment interactions illustrated by laying hens. *J Anim Breed Genet.* 1994;111:265–88.
63. Haldane JBS. The interaction of nature and nature. *Ann Eugen.* 1946;13:197–205.
64. Lerner IM. Population genetics and animal improvement as illustrated by the inheritance of egg production. 1st ed. Cambridge: Cambridge University Press; 1950.
65. Thompson R. Statistical validation of genetic models. *Livest Prod Sci.* 2001;72:129–34.
66. Macedo FL, Reverter A, Legarra A. Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *J Dairy Sci.* 2020;103:529–44.
67. Jang S, Ros-Freixedes R, Hickey JM, Chen C-Y, Herring WO, Holl J, et al. Multi-line ssGBLUP evaluation using preselected markers from whole-genome sequence data in pigs. *Front Genet.* 2023;14:1163626.
68. Jang S, Ros-Freixedes R, Hickey JM, Chen C-Y, Holl J, Herring WO, et al. Using pre-selected variants from large-scale whole-genome sequence data for single-step genomic predictions in pigs. *Genet Sel Evol.* 2023;55:55.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.