

ORIGINAL ARTICLE

Assessing various environmental descriptors with respect to genotype x environment interaction for milk production traits in Bavarian Fleckvieh cattle

Malena Erbe  | Christian Edel | Reiner Emmerling | Kay-Uwe Götz

Bavarian State Research Centre for Agriculture, Institute for Animal Breeding, Poing-Grub, Germany

Correspondence

Malena Erbe, Bavarian State Research Centre for Agriculture, Institute for Animal Breeding, Poing-Grub, Germany.

Email: malena.erbe@lfl.bayern.de

Funding information

H2020 GenTORE, Grant/Award Number: 727213

Abstract

Genotype x environment (GxE) interaction for production traits in Fleckvieh cattle was assessed by means of various environmental descriptors (EDs). It was also of particular interest to search for EDs useful for studying differing robustness or resilience of individuals which implies reasonable GxE interaction. The set of studied EDs included farm/herd environment (e.g. herd size, housing/feeding regimes, herd production level), geographical location (e.g. height above sea level), temperature humidity index and fat-to-protein ratio. Milk, fat and protein yield deviations for the first test day, the first 100 days and the 305-day-yield of the first lactation were used as phenotypes. Genetic correlations were estimated with bi- and multivariate sire models in case of categorical EDs and reaction norm sire models in case of continuous EDs. Further, rank correlations of bulls' estimated breeding values were calculated for different environmental levels/classes. Almost all estimated genetic correlations were significantly different from 1, but not <0.93. There were some exceptions for extremely different classes/levels of some EDs (e.g. average herd-year production, organic vs conventional farm systems, way of concentrated feed supply). Rank correlations were rarely below 0.95. In summary, no substantial GxE interactions for milk production traits were found with the studied EDs.

KEYWORDS

Fleckvieh cattle, GxE interaction, reaction norm, resilience, robustness, variance component estimation

1 | INTRODUCTION

The term genotype x environment (GxE) interaction describes the variation in phenotypic response of specific genotypes to specific environments (Lynch & Walsh, 1998), i.e. genotypes are differently affected by different environments. Significant GxE interaction can result in scaling effects, linked with heterogeneity of genetic variance in different environments, and in re-ranking of

genotypes due to their genetic value in different environments. Ignoring GxE interaction in the context of breeding value estimation can thus lead to suboptimal selection decisions given that selection is for a different environment than the one the trait is measured in. While scaling effects only causes problems when breeding indices are built from various traits (Dominik & Kinghorn, 2008), re-ranking of individuals leads to suboptimal selection decisions.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Animal Breeding and Genetics* published by John Wiley & Sons Ltd.

GxE interaction can only be studied if appropriate variables are available that allow to distinguish between different environments. Analysis of GxE interaction in livestock data is mainly performed with two different types of models. Which one is chosen is mainly depending on the character of the available environmental descriptor (ED). Following Falconer (1952) phenotypes of the same trait obtained in a limited number (say n_{env}) of different environments can be understood as phenotypes of n_{env} genetically correlated traits. The genetic correlation between these traits is assumed to be 1 when no GxE interaction is present while it differs from 1 otherwise. Estimating genetic correlations within a mixed model context is still a common way in animal breeding for analysing GxE interaction provided the ED is not continuous (e.g. Gerber, Krogmeier, Emmerling, & Götz, 2008; König, Dietl, Raeder, & Swalve, 2005; Schmid et al., 2021). One international application is the multiple across country estimation (MACE) by Interbull which enables bulls with breeding values from another country to be compared to bulls on any national scale by assuming phenotypes in different countries to be genetically correlated traits, i.e. a GxE interaction between traits across countries is assumed (Schaeffer, 1994).

Given that the ED is available on a continuous scale, reaction norm models have frequently been used in the analysis of GxE interactions within the last years (e.g. Calus, Groen, & de Jong, 2002; Strandberg, Kolmodin, Madsen, Jensen, & Jorjani, 2000; Streit, Reinhardt, Thaller, & Bennewitz, 2012). It has been a long time from the first description of reaction norm (Woltereck, 1909) and the common use of models for estimating the necessary parameters in an animal breeding context. The latter was promoted by the description of the use of covariance functions (Kirkpatrick, Lofsvold, & Bulmer, 1990) and the introduction of random regression models for dairy cattle evaluation (Schaeffer & Dekkers, 1994). A reaction norm model helps to describe the environmental sensitivity of a genotype (Falconer & Mackay, 1996) which might be mathematically interpreted as regression of the genetic value on the environmental value (de Jong, 1995). In common reaction norm models, thus, a random intercept describes the general (environmental-independent) genetic level for a given individual while random regression coefficients of order 1 to n describe its environmental sensitivity. A better consistency of the performance level across all environments can be interpreted as lower environmental sensitivity or better ability to cope with environmental variability (Rauw & Gomez-Raya, 2015).

Random regression models used to model the lactation curve of a cow require multiple observations per individual. For EDs, in many cases, phenotypic measurement of an individual is only available in one class or at one or a small number of data point(s) of the ED which makes it

difficult to apply reaction norm models in the form of the individual animal model. However, the advantage of dairy data in comparison to other data sets is that different environments are well connected via the sires that often produce a meaningful number of daughters in different herds (Rauw & Gomez-Raya, 2015). Thus, using a sire model and relating daughters' observations to the sires allow to have multiple data points for a genotype over the whole range of classes or along the continuous scale of the ED which allows to estimate slopes of a reaction norm more accurately.

According to Friggens et al. (2022), resilience can be defined as the "ability of an animal to 'bounce back' from a disturbance, which implicitly is of relatively short duration" while robustness can be defined as "capability to cope with environments that are unfavorable for a long time". Similarly, Colditz and Hine (2016) describe resilience as the "capacity of the animal to return rapidly to its pre-challenge state following short-term exposure to a challenging situation" and a robust individual as one that "has the ability to express its production potential in a wide range of environments without compromising its reproduction, health and wellbeing". Following these definitions, EDs that are the same over the whole lifespan of an individual or at least for a whole lactation (e.g. herd production level as in Strandberg et al. (2000) or Streit et al. (2012), location of the farm as in König et al. (2005) or organic vs. conventional farm type as in Schmid et al. (2021)) can thus be used to study robustness. In order to study resilience, EDs are required that change over time within a lifespan or at least within a lactation of an individual and are thus specific for a record on a given test day. Common examples are weather data and its derivatives (e.g. temperature humidity indices) that can indicate heat or cold stress events (e.g. Brügemann, Gernand, von Borstel, & König, 2011; Nguyen, Bowman, Haile-Mariam, Pryce, & Hayes, 2016; Toghiani, Hay, Fragomeni, Rekaya, & Roberts, 2020) or the fat-protein ratio which might be used as proxy for challenging metabolic situations (Ha et al., 2017) or feeding regimes (Rauw & Gomez-Raya, 2015).

Within the Horizon 2020 project "GenTORE" (<https://www.gentore.eu>) data from the Farm Accountancy Data Network (FADN; https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/farms-farming-and-innovation/structures-and-economics/economics/fadn_en) and the Gridded Agro-Meteorological Data in Europe (AGRI4CAST; <https://agri4cast.jrc.ec.europa.eu/DataPortal/Index.aspx>) were used to classify European cattle systems and identify important main environmental challenges to resilience and efficiency in cattle production systems (Quiédeville, Moakes, Leiber, & Pfeifer, 2020) with a focus on farm level.

The aim of our study was to analyse GxE interaction for milk production traits in the first lactation given EDs that are available for a broad (population-level) Fleckvieh data set and might be interesting for studying robustness or resilience of individuals or genotypes. We used the factors identified in GenTORE as a starting point and identified EDs from various sources that allowed a characterization of herds and did not only represent the absolute production level itself, but different farm characteristics, climatic conditions or localization.

2 | MATERIALS AND METHODS

2.1 | Phenotypic observations and pedigree

Yield deviations (YD) for milk, fat and protein kilogram from April 2019 were available for all Bavarian Fleckvieh cows that were part of the German-Austrian routine genetic evaluation. YDs (accounting for all fixed effects) were derived from a model that included the same fixed and random effects as the official conventional routine breeding value estimation model (a test day model, see Lidauer, Emmerling, and Mäntysaari (2008) for more details), but which was run without heterogeneous variance corrections. In our analyses, we then used YDs for the first test day in the first lactation (YDDay1), cumulative YDs for the 305-day standard length of the first lactation (YDL) as well as cumulative YDs for the first third of the first lactation, i.e. the first 100 days (YD100) as phenotypes in models for variance component estimation. YD100 and YDL were derived via best prediction (VanRaden, 1997) which uses available test day YDs for a specific individual as well as covariance structures between test days to interpolate YDs for specific time periods (in our case first 100 days and 305 days of the first lactation). Data from years 2013 to 2018 was used for most of the analyses due to the limited availability of information about the environmental descriptors (EDs). For some EDs, we had to restrict the data to years 2015 to 2018 (see [File S1](#)) as we had access to Integrated Administration and Control System (IACS) based data only for these years. Only cows with valid YDs for milk, fat and protein kilogram were retained for further analyses. For YDDay1, cows were required to have an observation between day 8 and day 40 after calving. YD100 (YDL) were considered when cows had at least three (eight) observations between day 8 and 120 (335) days. After first quality control, there were 1.15M, 1.13M and 998K observations for YDDay1, YD100 and YDL available.

Observations were retained when the cow had a known sire, the cow and the sire were both purebred Fleckvieh (as defined by EU-regulation 2016/1012, i.e. parents and

grandparents are recorded in the same herdbook) and the sire had observations of daughters located in at least 10 different farms for the respective environmental descriptor.

As only sire models were applied to the data, pedigree was built starting from all sires with an observation in at least one ED within a given type of studied YDs. The final pedigree contained five generations of ancestors for these sires.

2.2 | Environmental descriptors

A measure of the average performance level of a herd has often been used in GxE studies (e.g. Calus, Windig, & Veerkamp, 2005; Streit et al., 2012) in order to describe management differences between farms which can be an indicator of differently challenging environments for the individuals. We used the herd-year solution of the trait milk yield (HYSM) and the herd-year solutions of the sum of fat and protein yields (HYSFP) from the test-day model of the routine breeding value estimation as a continuous ED. For quality control, we removed observations of HYSM and HYSFP that were below the respective 0.01% and above the 99.9% quantiles. Furthermore, for comparison, we built categorical EDs by splitting HYSM and HYSFP into the quartiles of all respective herd-year solutions between years 2012 and 2019. For HYSM, the classes were ≤ 5400 , $>5400 - \leq 6000$, $>6000 - \leq 6600$, >6600 kg milk yield and for HYSFP they were ≤ 390 , $>390 - \leq 440$, $>440 - \leq 490$, >490 kg fat plus protein yield. Since it was not clear whether all models with four classes would converge, we also tested the two extreme classes, i.e. the first and the last quartile, in bi-variate models. The four class models are named HYSM4, HYSFP4 and the extreme class models HYSM2extreme and HYSFP2extreme in the following. Since EDs derived from performance data which was also used as phenotype tend to underestimate the slope in reaction norm models (Calus, Bijma, & Veerkamp, 2004), we were also interested in using EDs originating from other sources than the data set itself.

Farm-based descriptors were available for all farms under milk recording. The Bavarian milk recording organization (LKV, Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V.) collects data of herd size, average milk yield, husbandry and basic forage system and type of concentrated feed supply for most of the participating farms. Since we only had farm descriptors for one reporting date, we performed some quality control steps in order to exclude farms that probably had major system changes within the last years (e.g. by excluding data from farms in which the herd size varied considerably). For herd size, we defined a categorical ED with two classes (HERD2) splitting the data into farms with a herd size of <60 and ≥ 60 cows and with four classes (HERD4),

namely a herd size of <30, 30–59, 60–119 and >120 cows. Regarding concentrated feed supply, a categorical ED with two classes (CONCFEED2) was defined by splitting the data into farms having “exact ration/transponder supply” against all others or into four classes (CONCFEED4) which were “no supply”, “total mixed ration” (TMR), “approximate supply” and “exact ration/transponder supply”. We further defined EDs with two classes in order to distinguish between farms with and without an automatic milking system (AMS) and one to distinguish between farms with tie-stall and loose housing (HOUSE).

We further had access to some Integrated Administration and Control System (IACS) based data from the years 2015 to 2018. Using this data, we could classify the farms with respect to their agricultural system. For conventional and organic farming systems, we defined a categorical ED with two classes (ORG). The data did not contain direct information about a farm's basic forage system. Therefore, we derived a proxy for the forage system from IACS data: For each farm with milk recording and each observed year, the size of permanent pasture area in hectares and the area of fodder maize in hectares were extracted and the proportion

$$PP\% = 100\% \times \text{permanent pasture area} / (\text{permanent pasture area} + \text{fodder maize area})$$

was calculated. Calculated proportions were averaged across years within farms. For quality reasons, PP% was only used for further analyses when there were at least two data points per farm. A farm was discarded when one of the yearly observations deviated more than 2 standard errors from the farm average or more than 10 absolute percentage points from all others. The quality checked PP% were used as continuous EDs. We further constructed two EDs with two classes each: PP%80 by splitting PP% in classes of $\leq 80\%$ and $> 80\%$ and PP%50 by splitting PP% in classes of $\leq 50\%$ and $> 50\%$ and omitting the data for PP% values from $> 50\%$ and $\leq 80\%$. PP% represented not only differences in basic forage systems themselves, but also has a geographical context since farms in the alpine regions tend to have quite uniform PP% because there is almost no cultivation of maize whereas in flat regions in lower Bavaria generally low PP% can be expected.

Based on address data, GPS coordinates were derived for all farms and only used in cases in which the farm location could be clearly assigned. A publicly available data set of a digital elevation model in a 50 m grid for Bavaria (https://geodatenonline.bayern.de/geodatenonline/seite/n/dgm_info) was used to define the ED “meters above sea level” (termed MASL in the following) for all farms. GPS coordinates of the farms were matched to the closest

available GPS coordinate from the raster and MASL of this raster coordinate was taken as MASL for the respective farm. MASL was used as continuous ED directly and as categorical ED MASL750 (two classes, ≤ 750 and > 750 m) and MASL600 (two classes, ≤ 600 and > 600 m). Since not only the absolute height above sea level, but also height in combination with climatic conditions is very different in alpine areas and areas close to the alps as compared to the rest of Bavaria, farms were classified into farms in the alps or the alpine upland (following the definition of Würfl, Dörfler, & Rintelen, 1984) and others in order to build a categorical ED (ALPINE).

All EDs described above have a lifelong influence on an individual and thus can only be used to study robustness following the definition from Friggens et al. (2022) or Colditz and Hine (2016). For studying resilience, it is necessary to use EDs that vary during an individual's lifespan and that allow to observe its reaction on a punctual external (e.g. heat stress) or internal (e.g. diseases) disturbance.

Weather data was available on an hourly basis for the years 2013 to 2018 for 216 weather stations in Bavaria (<https://www.wetter-by.de/>; https://www.dwd.de/DE/leistungen/cdc_portal/cdc_portal.html) which provided

dry bulb temperature in °C ($T_{\circ C}$) and relative humidity in % ($RH_{\%}$) data, both measured 2 m above ground, with <10% of missing values. For each station, data was only used from days on which at least 20 hourly observations were available. GPS coordinate information and the height above sea level of all farms and weather stations were used to assign weather station data to a specific farm. Weather data were assigned to a farm when the nearest weather station was not located more than 200 m higher or lower than the farm and when it was not more than 20 km away. If no weather station matched these criteria for a given farm, cows from this farm were excluded from all analyses based on weather data. For the remaining farms, mean (\pm s.d.) distance to the chosen weather station was 8.88 km (± 4.31).

Many different formulas can be found in literature for combining information of humidity and temperature to a temperature–humidity index (THI). Since in our data, dry bulb temperature and relative humidity were available, we used an index given in NRC (1971) but rephrased for temperature measures in Celsius instead of Fahrenheit. Thus, hourly temperature–humidity indices (THI) were calculated as

$$THI = 46.3 + 0.81 \times T_{\circ C} + 0.99 \times \left(\frac{RH_{\%}}{100} \right) \times T_{\circ C} - 14.3 \times \left(\frac{RH_{\%}}{100} \right)$$

All hourly THI values from a specific weather station were then averaged to obtain the average daily THI (THIday). In addition, we calculated an average THI per day that only consisted of values from 10 a.m. to 4 p.m. (THIday10to4) to capture only the hours that are potentially the most challenging ones regarding heat stress. Because performance of a cow on a specific day will also be influenced by climatic conditions on the days before the recording day, different authors have used THI on specific days before recording or an average of THI values from these days (e.g. Brügemann et al., 2011). Another ED was thus defined as the mean of the average daily THIs of the three days before the recording date (THIminus3). Brügemann, Gernand, König von Borstel, and König (2012) found a lower limit of THI~60 for substantial decrease of milk yield in Holstein dairy cattle herds in Germany. Given these results, we defined another ED by counting the number of hours of the recording date in which the THI was more than 60 (THIhours60). THIday, THIday10to4, THIminus3 and THIhours60 were used as continuous EDs. In order to use THI data also in models with categorical EDs, we split the data into two classes with $\text{THIminus3} \leq 60$ and $\text{THIminus3} > 60$ (THIcat60daily) and into two classes with $\text{THIhours60} \leq 10$ and $\text{THIhours60} > 10$ h (THIcat60hours). All THI values were used as EDs in analyses of YDDay1 only as this was the only phenotype for which a real recording date was available.

Fat-to-protein ratio from milk recording results can be used as an indicator for studying resilience to sub-optimal feeding regimes or diseases with values of ≤ 1.0 indicating sub-acute ruminal acidosis and ≥ 1.5 indicating ketosis are found in literature (e.g. Enemark, 2008; Heuer, Van Straalen, Schukken, Dirkwager, & Noordhuizen, 2000). Fat and protein kilogram recordings for all cows from the first test day in the first lactation were available from milk recording. Individual fat-to-protein ratio values were calculated for all cows with a first test day record for which the fat kg and protein kg recordings were not less than the 1% and not greater than the 99% quantile. In order to check for the influence of different days in milk on the first test day, we applied a quadratic regression of the fat-to-protein ratio on days in milk on the first test day. The coefficient of determination was low, but the regression coefficients were highly significant, so we nevertheless corrected for the days in milk by standardizing all individual fat-to-protein ratios to the individual fat-to-protein ratios (FPR) expected on the mean value of days in milk in our data set which was 24. FPR was used as continuous ED in analyses of YD100 to assess whether a disturbance in very early lactation affects production traits across the first 100 days of the lactation. Categorical EDs (FPR1.5 and FPR1.0) were built by splitting the cows in two groups, those with FPR of > 1.5 (≤ 1.0) and others, respectively.

An overview of all EDs and their definitions can also be found in Table 1 and in more detail in File S1.

2.3 | Statistical analyses

Assessment of GxE interaction can be performed with different types of models. The model choice is mainly depending on the kind of ED to be analysed.

For all categorical EDs, we considered the phenotypes obtained in the different environments to be correlated traits and estimated their genetic correlation with multivariate sire models. For YDL and YD100 of milk, fat or protein yield, the model was

$$y_{ijk} = \mu_k + s_{ik} + \varepsilon_{ijk}$$

while for YDDay1 it was

$$y_{ijk} = \mu_k + b_{k1}t_{ij} + b_{k2}t_{ij}^2 + s_{ik} + \varepsilon_{ijk}$$

with y_{ijk} being the respective YD of either milk, fat or protein yield in environment k from daughter j of sire i , μ_k was an overall mean in environment k , b_{k1} and b_{k2} were regression coefficients of the linear and quadratic fixed regression on t_{ij} which was days in milk of daughter j of sire i . s_{ik} was the sire effect of sire i in environment k and ε_{ijk} was the residual term of daughter j of sire i in environment k . The variance structure of the sire effect was assumed to be

$$V(s) = G \otimes A$$

where A was the full pedigree-based relationship matrix of the sires, G was a $n_c \times n_c$ matrix with genetic sire variances on the diagonal and covariances on the off-diagonals with n_c equals 2 for all 2-class-ED models and 4 for all 4-class-ED models.

For environmental descriptors that were available on a continuous scale (PP%, MASL, HYSM, HYSFP, FPR and different THI variables), we used a sire reaction norm model to check for GxE interaction. Heterogeneous residual variances across the environments might influence the estimates of the genetic components when not accounted for (Lillehammer, Ødegård, & Meuwissen, 2009). In our models, we estimated a residual variance for each of the four quartiles of data that resulted from ordering the phenotypes according to their respective ED value. We decided to model also fixed regressions of the respective EDs (see e.g. Schaeffer, 2004) so that the random regression coefficients represent a deviation from the average reaction norms. Uncorrelated coefficients might be favourable for higher-order regressions (Schaeffer, 2004). Ordinary polynomials are not independent of each other, but Legendre polynomials are

TABLE 1 Definition of all categorical and continuous environmental descriptors used in this study

Environmental descriptor	Categorical/continuous	Definition
HYSM	Continuous	Herd-year-solutions ^a for milk yield
HYSM4	Categorical	Quartiles of HYSM
HYSM2extreme	Categorical	Lowest 25% and highest 25% of HYSM
HYSFP	Continuous	Herd-year-solutions ^a for sum of fat and protein yield
HYSFP4	Categorical	Quartiles of HYSFP
HYSFP2extreme	Categorical	Lowest 25% and highest 25% of HYSFP
HERD2	Categorical	Herd size in two categories (<60/≥60)
HERD4	Categorical	Herd size in four categories (<30/31–59/60–119/≥120)
CONCFEED2	Categorical	Way of concentrated feed supply in two categories (exact ration or transponder vs all others)
CONCFEED4	Categorical	Way of concentrated feed supply in four categories (see text for details)
AMS	Categorical	Automatic milking system yes/no
HOUSE	Categorical	Tie-stall or loose housing
ORG	Categorical	Organic farm system yes/no
PP%	Continuous	Proportion of permanent pasture area to the sum of permanent pasture plus fodder maize area of the farm
PP%80	Categorical	PP% in two categories (≤80%/>80%)
PP%50	Categorical	PP% in two categories (≤50%/>80%)
MASL	Continuous	Height above sea level of the farm location in meters
MASL750	Categorical	MASL in two categories (≤750/>750)
MASL600	Categorical	MASL in two categories (≤600/>600)
ALPINE	Categorical	Farm location in alps or alpine upland yes/no
THIday	Continuous	Average THI on the first test day in the first lactation
THIday10to4	Continuous	Average THI between 10 a.m. and 4 p.m. on the first test day in the first lactation
THIminus3	Continuous	Average THI on the three days before the first test day in the first lactation
THIhours60	Continuous	Number of hours with an hourly THI>60 on the first test day in the first lactation
THIcat60daily	Categorical	THIminus3 in two categories (<60/≥60)
THIcat60hours	Categorical	THIhours60 in two categories (<10/≥10)
FPR	Continuous	Fat-to-protein-ratio on day 24 in milk in the first lactation
FPR1.5	Categorical	FPR in two categories (≤1.5/>1.5)
FPR1.0	Categorical	FPR in two categories (≤1.0/>1.0)

Note: Abbreviation: THI, temperature humidity index.

^aHerd-year-solution quartiles were built based on solutions from the routine breeding value estimation of all farms from years 2012 to 2019.

orthogonal. We thus used normalized Legendre polynomials for defining the random regression terms. As Legendre polynomials are only defined in a parameter space of $[-1,1]$, EDs were limited to this space by calculating $ED_{scaled_{ij}} = -1 + 2 \left(\frac{ED_{ij} - \min(ED)}{\max(ED) - \min(ED)} \right)$ with individual ij being a daughter j of sire i . Minima and maxima were found within each set of individuals given a specific phenotype and ED. When using this transformation, the intercept of the reaction norm estimates the genetic effect for the environment that equals $0.5(\max(ED) + \min(ED))$. The reaction norm model for YDL and YD100 of milk, fat or protein yield and any continuous ED was

$$y_{ijm} = \sum_{q=0}^n \beta_q z_{ijq} + \sum_{q=0}^n l_{iq} z_{ijq} + \epsilon_{ijm}$$

while for YDDay1, it was

$$y_{ijm} = b_1 t_{ij} + b_2 t_{ij}^2 + \sum_{q=0}^n \beta_q z_{ijq} + \sum_{q=0}^n l_{iq} z_{ijq} + \epsilon_{ijm}$$

y_{ijm} was one of the yield deviation measures of either milk, fat or protein yield as defined above of a daughter j within sire i assigned to one (m) of four random error classes as described above. b_1 and b_2 were regression coefficients of the linear and quadratic fixed regression on t_{ij} which was days in milk of daughter j of sire i . β_q was a regression coefficient of the q^{th} order fixed regression on z_{ijq} (i.e. β_0 was an overall mean), and z_{ijq} was calculated as $\phi_q(ED_{scaled_{ij}})$ with ϕ_q being the q^{th} order normalized Legendre polynomial. l_{iq} was the regression coefficient for sire i of the q^{th} order random regression on z_{ijq} (i.e. l_{i0} was a random intercept of sire i). n was 1 for linear reaction norm models and 2 for reaction norm models which also included quadratic regressions. ϵ_{ijm} was the random error term of daughter j of sire i in random error class m .

The breeding value of sire i estimated by a reaction norm model for a given environment env was $g_{i,env} = 2 \cdot \sum_{q=0}^n l_{iq} z_{envq}$. Variance structures of sire regression effects were assumed to be

$$V(l) = U \otimes A$$

with U being a 2x2 matrix with (co)variances of random intercept (l_0) and linear regression term (l_1) for reaction norm models of order 1 and a 3x3 matrix with (co)variances of random intercept (l_0), linear (l_1) and quadratic (l_2) regression terms for reaction norm models of order 2.

All models were solved, and all variance components were estimated with ASReml 3.0 (Gilmour, Gogel, Cullis, & Thompson, 2009).

To test if specific (co)variance components were different from zero, we performed likelihood ratio tests of

the original and reduced models for all reaction norm models. For linear reaction norm models, we tested $H_{0lin}: \sigma_{l_0, l_1} = 0$ and $\sigma_{l_1}^2 = 0$ using the full model and a reduced one in which both σ_{l_0, l_1} and $\sigma_{l_1}^2$ were fixed at 0. For reaction norm models that included also a random quadratic regression coefficient per sire, we tested $H_{0qua}: \sigma_{l_1, l_2} = 0; \sigma_{l_0, l_2} = 0; \sigma_{l_2}^2 = 0$ with the respective likelihood ratio. Following Stram and Lee (1994), we assumed the null distribution of the likelihood ratio to be a mixture of two χ^2 -distributions so that the p-value for a calculated likelihood ratio L was $0.5 \sum_{x=x_1-1}^{x_1} P(\chi_{df=x}^2 \geq L)$ with x_1 being 2 and 3 for tests of H_{0lin} and H_{0qua} , respectively.

For reaction norm models, we calculated genetic correlations between different environments. The genetic correlation of a trait in any two environments env_1 and env_2 is defined as $r_{g_{-env_1, env_2}} = \frac{\text{cov}(g_{env_1}, g_{env_2})}{\sqrt{\text{var}(g_{env_1}) \cdot \text{var}(g_{env_2})}}$. Numerator

and denominator for a reaction norm model with linear regression terms when normalized Legendre polynomials are used can be calculated as follows:

$$\text{cov}(g_{env_1}, g_{env_2}) = 4 \cdot \left(\Phi_1(ED_{scaled_{env_1}})' U \Phi_1(ED_{scaled_{env_2}}) \right)$$

and

$$\text{var}(g_{env_h}) = 4 \cdot \left(\Phi_1(ED_{scaled_{env_h}})' U \Phi_1(ED_{scaled_{env_h}}) \right)$$

with h being 1 and 2 for the two different environments, respectively.

Following the same scheme, the solutions for reaction norms with random intercept, linear and quadratic regression terms were:

$$\text{cov}(g_{env_1}, g_{env_2}) = 4 \cdot \left(\Phi_2(ED_{scaled_{env_1}})' U \Phi_2(ED_{scaled_{env_2}}) \right)$$

and

$$\text{var}(g_{env_h}) = 4 \cdot \left(\Phi_2(ED_{scaled_{env_h}})' U \Phi_2(ED_{scaled_{env_h}}) \right)$$

with h being 1 or 2 for the two different environments, respectively.

Standard errors of the correlation coefficients in the reaction norm models were calculated using an approximation with Taylor series (see e.g. Su, Lund, & Sorensen, 2007).

Scaling effects between EBVs of different categories in multi-trait models might arise if the genetic variances

in the studied environmental categories are not the same. We studied this for all bi-variate models with two-categorical EDs by calculating the 95%-confidence intervals of the genetic variances in the two categories and comparing them. If there was intersection of confidence intervals for genetic variance components of different environments, scaling effects were considered to be negligible.

3 | RESULTS

Genetic correlations (r_g) of milk, fat and protein yield (MY, FY and PY in the following) in the first lactation (L), in the first 100 days of the first lactation (100) and on the first test day of the first lactation (Day1) in the two respective categories of all two-categorical EDs are shown in Table 2. In general, all but one r_g were larger than 0.9 (between 0.909 for PY_L with ED HYSFP2extreme and 0.997 for FY_Day1 with CONCFEED2) and the 95%-confidence intervals showed that most of them could be considered different from 1 (except PY_Day1 with ORG, FY_Day1 with MASL750 and CONCFEED2). The lowest r_g was found for PY_100 given the ED HYSFP2extreme.

For herd size, concentrated feed supply and herd year performances, multi-trait models that split the phenotypic data into four different categories of EDs were also used. For some runs, we found poor convergence behaviour or models converged formally (based on ASReml criteria), but they ended up having a negative definite genetic covariance structure. Results for trait MY are shown in Table 3. Models for MY_Day1 with EDs HYSM4 and HYSFP4 did not converge. For MY_L and MY_100, all r_g were larger than 0.91 with EDs HYSM4 and HYSFP4. The smallest values for r_g were found between the most extreme herd sizes (<30 and >120 cows, $r_{g_cat_1,cat_4}$ in Table 3) and confirmed the estimates of the bi-variate runs HYSM2extreme and HYSFP2extreme (see Table 2). While CONCFEED2 distinguished phenotypes from farms that were recorded with concentrated feed supply that is specifically adapted to individual performance and all others, CONCFEED4 also included one category with observations of cows that did not obtain any concentrated feed supply (category 1 in Table 3). This category showed clearly lower r_g to all other categories ($r_{g_cat_1,cat_x}$ with $x \in \{2, 3, 4\}$) than did the other categories among each other. However, due to a relatively low percentage of observations in category 1, these $r_{g_cat_1,cat_x}$ also had quite large standard errors. Results for FY and PY are in File S2 and mainly followed the same trends as described for MY.

In order to check for re-ranking of individuals due to their EBVs in different environments for categorical ED, we calculated Spearman's rank correlation. Results for

Day1 analyses and all two-category EDs are shown in Table 4 (further results from all other YD types can be found in File S2). In general, rank correlations were very high and somewhat lower for the subset bulls20cat (bulls with at least 20 daughters in each of the categories) compared to subset bulls40 (bulls with at least 40 observations of daughters irrespectively in which category). The lowest rank correlation was found in the subset bulls20cat for PY_Day1 when data was split due to ED HYSFP2extreme ($r_{Spearman} = 0.950$) which was also only one of three correlations below 0.96. Rank correlations below 0.97 in the subset bulls20cat were also found with EDs HYSFP2extreme and HYSM2extreme for MY_Day1 and with PP%50 and HYSM2extreme for PY_Day1. For models with YD types YDL and YD100 as phenotypes, there were no rank correlations below 0.93 at all (File S2). Rank correlations below 0.96 were found in some cases with different EDs (e.g. AMS, ORG, HYSFP2extreme, HYSM2extreme) more often for PY or FY than for MY. Regarding the models with four-categorical-EDs (File S2), obviously lower rank correlations of EBVs of two categories accumulated when EBVs of the most extreme categories (1 and 4 within an ED) were compared, especially when the subset of bulls used for comparison was chosen such that each bull had at least 10 daughters in each of the categories.

Results for the comparison of the 95%-confidence intervals of the genetic variance components in the two categories of all bi-variate runs with two-categorical EDs can be found in Table 5. For many of the EDs studied the confidence intervals of the genetic variances of the two categories did overlap and thus scaling effects should be negligible. For models splitting the data into two classes based on EDs HYSM2extreme or HYSFP2extreme, the genetic variances seemed to be different for most of the scenarios applied. For models based on phenotypes YDL or YD100, runs with EDs ORG, HERD2, HOUSE, AMS and FPR were also found to sometimes have non-overlapping confidence intervals of the genetic variances in the two categories.

Likelihood ratio tests (LRT) were performed for the reaction norm models to check if the inclusion of an environment-dependent part of the breeding value (i.e. a random linear or even quadratic regression term) significantly improved the model. All LRT comparing models with a linear random slope to models without any random regression term were highly significant (i.e. we could reject $H_{0lin}: \sigma_{l_0,l_1} = 0$ and $\sigma_{l_1}^2 = 0$ with $p < 0.001$ in all cases). Adding a quadratic regression term did not lead to a better model fit for some ED and phenotype combinations (Table 6). LRT with $H_{0qua}: \sigma_{l_1,l_2} = 0; \sigma_{l_0,l_2} = 0; \sigma_{l_2}^2 = 0$ could not be rejected with $p < 0.01$ for all THI-EDs studied, except for FY_Day1 with THHours60 and for PY_Day1 with THIminus3; further for FY_Day1, MY_Day1,

TABLE 2 Genetic correlations and their standard errors for different traits obtained with bi-variate mixed models splitting the data into two categories based on different two-categorical environmental descriptors

Environmental descriptor ^a	Phenotype			YD100			YDDay1			Protein yield
	Milk yield	Fat yield	Protein yield	Milk yield	Fat yield	Protein yield	Milk yield	Fat yield	Protein yield	
HYSM2extreme	0.944 ± 0.008	0.943 ± 0.008	0.935 ± 0.009	0.930 ± 0.009	0.944 ± 0.008	0.913 ± 0.010	0.938 ± 0.010	0.953 ± 0.009	0.926 ± 0.012	
HYSFP2extreme	0.927 ± 0.009	0.929 ± 0.009	0.909 ± 0.010	0.918 ± 0.010	0.926 ± 0.009	0.889 ± 0.012	0.932 ± 0.011	0.943 ± 0.010	0.911 ± 0.014	
HERD2	0.982 ± 0.003	0.984 ± 0.003	0.967 ± 0.005	0.976 ± 0.004	0.980 ± 0.004	0.957 ± 0.007	0.966 ± 0.007	0.970 ± 0.006	0.952 ± 0.009	
CONCFEED2	0.976 ± 0.004	0.993 ± 0.002	0.975 ± 0.005	0.979 ± 0.004	0.994 ± 0.002	0.974 ± 0.005	0.993 ± 0.003	0.997 ± 0.003 ^b	0.991 ± 0.004	
AMS	0.958 ± 0.007	0.947 ± 0.009	0.937 ± 0.010	0.959 ± 0.007	0.944 ± 0.010	0.937 ± 0.010	0.970 ± 0.007	0.960 ± 0.010	0.957 ± 0.010	
HOUSE	0.978 ± 0.004	0.970 ± 0.005	0.951 ± 0.007	0.971 ± 0.005	0.969 ± 0.005	0.942 ± 0.008	0.958 ± 0.008	0.975 ± 0.006	0.935 ± 0.011	
ORG	0.933 ± 0.014	0.923 ± 0.015	0.926 ± 0.015	0.953 ± 0.013	0.935 ± 0.015	0.940 ± 0.015	0.975 ± 0.013 ^b	0.952 ± 0.018	0.958 ± 0.017	
PP%80	0.973 ± 0.005	0.976 ± 0.005	0.963 ± 0.007	0.979 ± 0.005	0.979 ± 0.005	0.974 ± 0.006	0.977 ± 0.006	0.971 ± 0.007	0.964 ± 0.009	
PP%50	0.960 ± 0.009	0.961 ± 0.008	0.947 ± 0.011	0.971 ± 0.007	0.970 ± 0.008	0.962 ± 0.009	0.961 ± 0.010	0.963 ± 0.011	0.933 ± 0.017	
MASL750	0.973 ± 0.010	0.973 ± 0.010	0.959 ± 0.012	0.973 ± 0.010	0.983 ± 0.008	0.960 ± 0.013	0.973 ± 0.012	0.982 ± 0.011 ^b	0.960 ± 0.016	
MASL600	0.959 ± 0.006	0.975 ± 0.005	0.938 ± 0.008	0.967 ± 0.005	0.977 ± 0.005	0.954 ± 0.007	0.985 ± 0.004	0.990 ± 0.004	0.980 ± 0.006	
ALPINE	0.959 ± 0.006	0.970 ± 0.006	0.946 ± 0.008	0.966 ± 0.006	0.976 ± 0.005	0.953 ± 0.007	0.970 ± 0.007	0.980 ± 0.006	0.958 ± 0.009	
THIcat60daily							0.987 ± 0.003	0.993 ± 0.003	0.989 ± 0.004	
THIcat60hours							0.989 ± 0.003	0.994 ± 0.003	0.987 ± 0.004	
FPR1.5				0.992 ± 0.002	0.989 ± 0.002	0.990 ± 0.002				
FPR1.0				0.987 ± 0.005	0.967 ± 0.009	0.983 ± 0.006				

Abbreviations: YD100, best prediction YD for the first 100 days of the first lactation; YDDay1, YD solution for the first test day in the first lactation; YDL, best prediction YD for the first lactation.

^aFor definition of the classes of the environmental descriptors, see Table 1.

^bThe 95%-confidence interval includes 1.

TABLE 3 Genetic correlations and their standard errors for milk yield in the first lactation, in the first 100 days of the first lactation and on the first test day of the first lactation obtained from multivariate mixed models splitting the phenotypes into four categories based on different four-categorical environmental descriptors (ED)

YD type as phenotype ^a	ED ^b	$r_{g12} \pm s.e.$	$r_{g13} \pm s.e.$	$r_{g14} \pm s.e.$	$r_{g23} \pm s.e.$	$r_{g24} \pm s.e.$	$r_{g34} \pm s.e.$
YDL	HYSM4	0.989 ± 0.003	0.966 ± 0.005	0.947 ± 0.007	0.990 ± 0.002	0.975 ± 0.004	0.993 ± 0.002
	HYSFP4	0.980 ± 0.003	0.959 ± 0.005	0.928 ± 0.008	0.993 ± 0.002	0.970 ± 0.004	0.989 ± 0.002
	HERD4	0.990 ± 0.003	0.960 ± 0.007	0.937 ± 0.013	0.988 ± 0.003	0.962 ± 0.009	0.982 ± 0.006
	CONCFEED4	0.884 ± 0.043	0.933 ± 0.035 ^c	0.918 ± 0.036	0.956 ± 0.010	0.981 ± 0.005	0.985 ± 0.004
YD100	HYSM4 ^d	0.984 ± 0.003	0.959 ± 0.006	0.933 ± 0.008	0.993 ± 0.002	0.973 ± 0.004	0.993 ± 0.002
	HYSFP4	0.978 ± 0.004	0.953 ± 0.006	0.919 ± 0.009	0.993 ± 0.002	0.971 ± 0.004	0.991 ± 0.002
	HERD4	0.971 ± 0.006	0.932 ± 0.011	0.909 ± 0.017	0.984 ± 0.003	0.956 ± 0.010	0.978 ± 0.007
	CONCFEED4	0.880 ± 0.046	0.912 ± 0.042	0.889 ± 0.043	0.942 ± 0.012	0.987 ± 0.005	0.971 ± 0.006
YDDay1	HYSM4	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e
	HYSFP4	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e	n.c. ^e
	HERD4	0.974 ± 0.007	0.940 ± 0.012	0.906 ± 0.021	0.973 ± 0.006	0.937 ± 0.016	0.980 ± 0.009
	CONCFEED4	0.868 ± 0.068 ^c	0.916 ± 0.061 ^c	0.871 ± 0.064	0.931 ± 0.016	0.992 ± 0.005 ^c	0.962 ± 0.009

^aYDL = best prediction YD for the first lactation, YD100 = best prediction YD for the first 100 days of the first lactation, YDDay1 = YD solution for the first test day in the first lactation.

^bHYSM4/HYSFP4 = quartiles of herd-year-solutions for milk yield or sum of protein and fat yield, respectively, CONCFEED4 = way of concentrated feed supply in four categories, HERD4 = herd size in four categories – for more details see [Table 1](#).

^cThe 95%-confidence interval includes 1.

^dASReml showed convergence for this run, but we found the estimated genetic covariance matrix to be negative definite.

^eThe model did not converge.

TABLE 4 Rank correlations for estimated breeding values (EBVs) in the two different categories of the environmental descriptors within two subsets of sires. Results are shown for models with YDs of the first test day as phenotype

Environmental descriptor ^a	Data subset	Sires with ≥ 40 obs _{dau} ^b			Sires with ≥ 20 obs _{dau} ^b in both categories				
		# of sires	Spearman's rank correlation of EBVs			# of sires	Spearman's rank correlation of EBVs		
			Milk yield	Fat yield	Prot. Yield		Milk yield	Fat yield	Prot. Yield
HYSM2extreme		906	0.972	0.982	0.967	597	0.966	0.979	0.959
HYSFP2extreme		899	0.968	0.979	0.958	602	0.961	0.974	0.950
HERD2		1245	0.985	0.991	0.980	1075	0.985	0.991	0.979
CONCFEED2		1242	0.999	1.000	0.998	847	0.999	1.000	0.998
AMS		1243	0.991	0.989	0.986	417	0.987	0.983	0.981
HOUSE		1243	0.984	0.993	0.974	905	0.982	0.992	0.972
ORG		1299	0.996	0.990	0.993	211	0.990	0.979	0.983
PP%80		1134	0.993	0.990	0.988	528	0.991	0.988	0.984
PP%50		547	0.983	0.987	0.969	333	0.980	0.986	0.963
MASL750		2137	0.995	0.998	0.993	125	0.985	0.992	0.980
MASL600		2137	0.995	0.998	0.994	565	0.993	0.997	0.991
ALPINE		2152	0.992	0.996	0.988	549	0.987	0.992	0.980
THIcat60daily		2076	0.997	0.999	0.997	829	0.996	0.998	0.997
THIcat60hours		2054	0.997	0.999	0.997	878	0.996	0.998	0.996

^aFor definition of the classes of the environmental descriptors, see [Table 1](#).

^bobs_{dau} = number of daughters of the respective sire with observations.

TABLE 5 Number of bi-variate models with data split based on different two-categorical environmental descriptors (ED) for which the 95% confidence intervals of the estimated genetic variances of a given trait in the two ED categories are/are not overlapping. For the cases of no overlap, the respective EDs are also stated

Trait	Result	YD type as phenotype		
		YDL	YD100	YDDay1
Milk yield	Overlap	7	7	12
	No overlap	5 (HYSM2extreme, HYSFP2extreme, HERD2, HOUSE, ORG)	7 (HYSM2extreme, HYSFP2extreme, HERD2, AMS, HOUSE, ORG, FPR1.5)	2 (HYSM2extreme, HYSFP2extreme)
Fat yield	Overlap	9	11	12
	No overlap	3 (HYSM2extreme, HYSFP2extreme, ORG)	3 (HYSM2extreme, HYSFP2extreme, FPR1.0)	2 (HYSM2extreme, HYSFP2extreme)
Protein yield	Overlap	9	9	12
	No overlap	3 (HOUSE, ORG, MASL750)	5 (HYSFP2extreme, HOUSE, ORG, FPR 1.5, FPR1.0)	2 (HYSM2extreme, HYSFP2extreme)

Abbreviations: YD100, best prediction YD for the first 100 days of the first lactation; YDDay1, YD solution for the first test day in the first lactation; YDL, best prediction YD for the first lactation.

TABLE 6 *p*-values of likelihood ratios tests for reaction norm models with H_0 : “inclusion of random quadratic regression on the environmental descriptor does not improve the model”, i.e. $\sigma_{l_1, l_2} = \mathbf{0}$ and $\sigma_{l_0, l_2} = \mathbf{0}$ and $\sigma_{l_2}^2 = \mathbf{0}$ with σ_{l_1, l_2} being the covariance between random linear and random quadratic regression term, σ_{l_0, l_2} the covariance between random intercept and random quadratic regression term and $\sigma_{l_2}^2$ the variance of the random quadratic regression coefficient. All *p*-values presented in the table are rounded to 3 digits

Environmental descriptor ^a	Phenotype ^b								
	YDL			YD100			YDDay1		
	Milk kg	Fat kg	Protein kg	Milk kg	Fat kg	Protein kg	Milk kg	Fat kg	Protein kg
HYSM	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.035	<0.001	0.001
HYSFP	<0.001	<0.001	<0.001	<0.001	0.005	<0.001	0.059	<0.001	<0.001
PP%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	<0.001
MASL	<0.001	0.002	<0.001	<0.001	0.078	<0.001	0.065	n.c. ^c	0.014
TH1day	_{-d}	-	-	_{-d}	-	-	0.486	0.580	0.172
TH1day10to4	-	-	-	-	-	-	0.501	0.709	0.266
TH1minus3	-	-	-	-	-	-	0.023	0.129	0.010 ^e
TH1hours60	-	-	-	-	-	-	n.c. ^c	0.005	n.c. ^c
FPR	_{-d}	-	-	<0.001	<0.001	<0.001	_{-d}	-	-

^aFor definition of the environmental descriptors, see Table 1.

^bYDL = best prediction YD for the first lactation, YD100 = best prediction YD for the first 100 days of the first lactation, YDDay1 = YD solution for the first test day in the first lactation.

^cFull model did not converge.

^dAll THI-based environmental descriptors were used with phenotype YDDay1. FPR was only used with phenotype YD100.

^eThe original (non rounded) *p*-value for ED TH1minus3 with YDDay1 for protein kg was slightly below 0.010.

PY_Day1 and FY_100 with MASL and for MY_Day1 with HYSFP and HYSM.

Results for genetic correlations and rank correlations of estimated breeding values from reaction norm models will be shown in the following. Results for a given YD type with a given ED are based on a reaction norm model with a random quadratic regression term included when the corresponding likelihood ratio test was significant (see

Table 6) and based on a model without a random quadratic regression term otherwise.

Genetic correlations between each two values along the gradient of possible ED values ($r_{g_{env_x, env_y}}$) can be calculated from the results of the reaction norm models. Figure 1 shows genetic correlations for PY_Day1 and different EDs for all combinations of values between the 0.1% and the 99.99% quantiles of the respective ED.

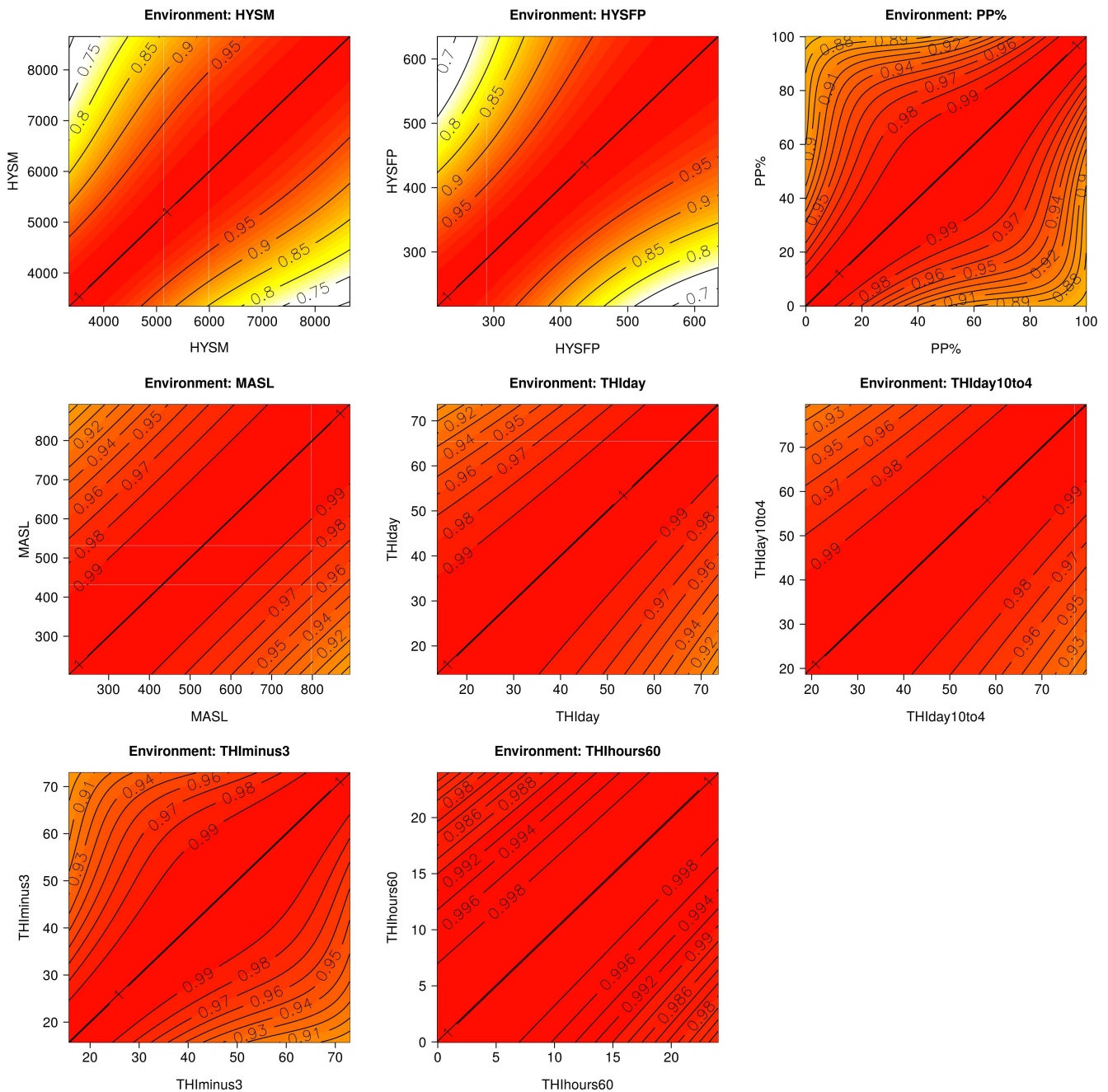


FIGURE 1 Genetic correlations between different levels of the environmental descriptors (ED) for protein yield on the first test day in the first lactation estimated with reaction norm models. Correlations were calculated for all possible combinations of values between the 0.1% and the 99.9% quantile of the ED. The results shown are for the model with a random quadratic regression term on the ED included in cases where the likelihood ratio test as shown in Table 6 was significant on a 1% error level, otherwise for the model without a random quadratic regression term. The model for the combination YDDay1 and environmental descriptor THlhours60 did not converge when a random quadratic regression term was included. Thus, results for the model without a random quadratic regression term are shown. HYSFP, HYSM = herd year solutions of the routine breeding value estimation for the sum of fat and protein yield or milk yield, respectively, MASL = height above sea level of the farm location in meters, PP% = proportion of permanent grassland area to the sum of permanent grassland plus fodder maize area of the farm in %, THday = average temperature humidity index on the first test day in the first lactation, THIminus3 = average temperature humidity index on the three days before the first test day in the first lactation, THday10to4 = average temperature humidity index between 10 a.m. and 4 p.m. on the first test day in the first lactation, THlhours60 = number of hours with an hourly THI > 60 on the first test day in the first lactation

$r_{g_env_x_env_y}$ for PY_Day1 were above 0.95 in many cases. In general, as expected, the lowest values of $r_{g_env_x_env_y}$ were mostly found between most extreme levels within a given

ED and were below 0.8 for HYSFP and HYSM as can be seen in Figure 1. For THI-based EDs, $r_{g_env_x_env_y}$ did not drop below 0.9. $r_{g_env_x_env_y}$ was significantly different (5%

error level) from 1 in almost all cases if studied for env_x equals the 1% quantile and env_y , the 25, 50, 75 or 99% quantile of the EDs (see [File S2](#)). With THI-based EDs except THIminus3 and MASL, the underlying models included no quadratic random regression term on the ED which is reflected in the course of the curves of equivalent $r_{g_{env_x,env_y}}$ over the whole gradient. For the others, shape of the distribution of the $r_{g_{env_x,env_y}}$ was quite different for different EDs. For PP%, for example, $r_{g_{env_x,env_y}}$ between very high ED values (e.g. above 90%) and most of all other lower ED values were very similar while for HYSM for a very high value of e.g. above 8000 a relatively steady decrease of $r_{g_{env_x,env_y}}$ appeared along the gradient from high to low ED values. As can be seen in [File S2](#) and [File S3c](#), the same trends could be observed also for traits FY and MY. Results for $r_{g_{env_x,env_y}}$ of models with phenotypes YDL and YD100 differed only very slightly ([File S2](#), [File S3a](#) and [File S3b](#)). With ED FPR, which was only analysed based on YD100, genetic correlations were also above 0.9 in almost all cases. Only for FY, some correlations were slightly below 0.9 when looking at very extreme values of FPR. In general,

the decline of $r_{g_{env_x,env_y}}$ when comparing similar and extreme levels of the EDs were somewhat more pronounced for PY than for MY and FY.

Results from calculations of rank correlations of EBVs for quantiles of the different EDs within traits can be found in [Table 7](#) and [File S2](#). Rank correlations were calculated between EBVs estimated for a given trait for the 1% and 99% quantile values of the EDs, respectively, and the EBVs for the 1%, 25%, 50%, 75% and 99% quantiles of the EDs. Rank correlations were higher than 0.95 in most of the cases and in general, lowest for EBVs of most extreme environmental levels within EDs. For protein yield, rank correlations dropped below 0.92 for EBVs on the 1% and the 99% quantiles for PY_L and PY_100 with ED HYSFP and for PY_100 with ED HYSM ([Table 7](#)). In general, very high rank correlations over the whole range of environments appeared for all models including random regressions on THI-based EDs.

[Figure 2](#) shows EBVs (for a subset of bulls with at least 40 daughter observations) estimated with reaction norm models for protein yield based on YDDay1 phenotypes

TABLE 7 Rank correlations for estimated breeding values (EBVs) for protein yield in different environments (defined by different quantiles of the environmental descriptor) within a subset of sires with at least 40 daughters with observations

YD type as phenotype ^a	Environmental descriptor ^b	Rank correlation of EBVs between the xth and the yth quantile of the ED (x%-y%) ^c						
		1%–25%	1%–50%	1%–75%	1%–99%	25%–99%	50%–99%	75%–99%
YDL	HYSM	0.985	0.978	0.97	0.936	0.969	0.978	0.986
	HYSFP	0.979	0.968	0.957	0.914	0.966	0.978	0.987
	PP%	0.993	0.992	0.990	0.964	0.959	0.969	0.982
	MASL	0.993	0.989	0.985	0.947	0.961	0.967	0.974
YD100	HYSM	0.981	0.971	0.959	0.919	0.967	0.979	0.988
	HYSFP	0.976	0.961	0.945	0.889	0.959	0.974	0.986
	PP%	0.992	0.993	0.993	0.978	0.967	0.974	0.985
	MASL	0.996	0.994	0.992	0.966	0.970	0.973	0.978
	FPR	0.995	0.991	0.986	0.981	0.992	0.994	0.996
YDDay1	HYSM	0.984	0.974	0.963	0.934	0.978	0.987	0.993
	HYSFP	0.977	0.964	0.951	0.921	0.981	0.990	0.996
	PP%	0.993	0.990	0.985	0.956	0.960	0.972	0.986
	MASL	0.998	0.997	0.995	0.981	0.991	0.993	0.996
	THIday	0.998	0.995	0.990	0.979	0.989	0.994	0.998
	THIday10to4	0.999	0.996	0.991	0.981	0.989	0.995	0.998
	THIminus3	0.996	0.992	0.987	0.978	0.988	0.993	0.997
	THIhours60	1.000 ^d	1.000 ^d	0.999	0.993	0.993	0.993	0.997

^aYDL = best prediction YD for the first lactation, YD100 = best prediction YD for the first 100 days of the first lactation, YDDay1 = YD solution for the first test day in the first lactation.

^bFor definition of the environmental descriptors, see [Table 1](#).

^cThe results for a given YD type with a given ED are based on a reaction norm model with a random quadratic regression term included when the corresponding likelihood ratio test was significant (see [Table 6](#)) and based on a model without a random quadratic regression term otherwise.

^dSince the variable THIhours60 is heavily left skewed on the original scale, the 1, 25 and 50% quantile values are the same and the rank correlations of EBVs between them are 1 consequently.

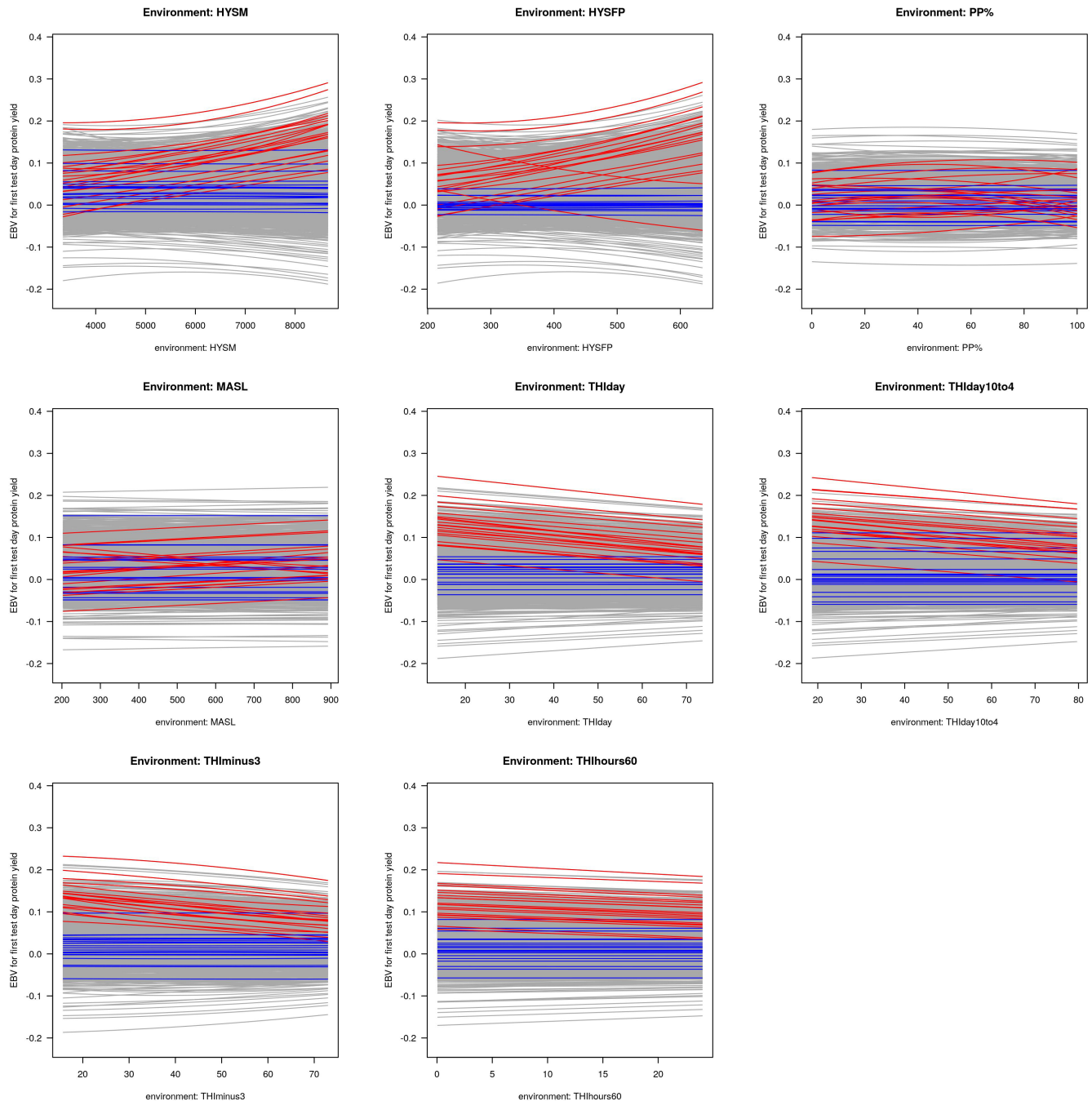


FIGURE 2 Breeding values (EBVs) of different levels of the respective environmental descriptor (ED) estimated with reaction norm models for protein yield on the first test day in the first lactation. The EBVs are from a subset of bulls that had at least 40 daughters with observations. EBV curves for the 20 bulls for which the EBVs changed most (least) over the range of the respective ED are highlighted in red (blue). The results shown are for the model with a random quadratic regression term on the ED included in cases where the likelihood ratio test as shown in Table 6 was significant on a 1% error level, otherwise for the model without a random quadratic regression term. The model for the combination YDDay1 and environmental descriptor TH1hours60 did not converge when a random quadratic regression term was included. Thus, results for the model without a random quadratic regression term are shown. HYSFP, HYSM = herd year solutions of the routine breeding value estimation for the sum of fat and protein yield or milk yield, respectively, MASL = height above sea level of the farm location in meters, PP% = proportion of permanent grassland area to the sum of permanent grassland plus fodder maize area of the farm in %, TH1day = average temperature humidity index on the first test day in the first lactation, TH1minus3 = average temperature humidity index on the three days before the first test day in the first lactation, TH1day10to4 = average temperature humidity index between 10 a.m. and 4 p.m. on the first test day in the first lactation, TH1hours60 = number of hours with an hourly THI > 60 on the first test day in the first lactation

over the range of all values between the 0.1% and the 99.9% quantile of the EDs. For the THI-based values except THIminus3 and MASL, only linear changes could be observed as the basis were models including only a linear random regression term on the ED. Scaling effects and re-ranking events of individuals due to their environmental dependent EBVs when moving along the gradient of a respective ED were present for all EDs as depicted with the 20 bulls with the highest and lowest changes in EBVs in [Figure 2](#) and [Files S4a–c](#). In order to be able to interpret the importance of changes in EBVs along the gradient of an ED not only from a graphical output and the rank correlations mentioned above, we checked the bulls with the 10% highest changes for all lactation traits (MY_L, FY_L, PY_L). For those, we calculated the difference of minimum and maximum observed EBVs (which did not have to be the EBVs at the extremes of EDs) along the ED gradient and averaged it. The results showed a mean difference of 860, 758, 497 and 344 kg milk, 35, 33, 14 and 12 kg fat and 27, 24, 20 and 12 kg protein for HYSFP, HYSM, MASL and PP%, respectively.

4 | DISCUSSION

The aim of this study was to check if different categorical and continuous environmental descriptors (EDs) obtained from various sources can be used to study resilience and robustness of different sires with respect to changing environments. We could not find substantial GxE interaction for any of the various ED studied. As a consequence, this means that breeding for robustness or resilience is not sensible in the present situation for Fleckvieh in Bavaria with respect to available data and environmental conditions. However, this may change in future especially if more precise phenotypes will become available or phenotypes will be available in higher temporal frequency.

For all studied EDs, no or only weak evidence of GxE interaction on yield traits could be found. Even though most of the estimated genetic correlations between different environments were significantly different from 1, the values were greater than 0.9 in almost all cases and even those that were smaller hardly fell below 0.8. None of them justify reflections about special breeding programs for different environments. Mulder, Groen, De Jong, and Bijma (2004) showed with a simulation study that this might be justified only when values are below 0.61. Nevertheless, scaling effects and rank changes could be found, especially for EDs based on herd performance level, roughage or concentrated feed system. From a practical point of view, if one wants to consider GxE interaction in the breeding value estimation process, this could be implemented relatively easy in an existing breeding

program, for example, by providing different indices for different environments. However, one has also to be aware of how this will influence the population structure and stratification, the overall response to selection and the number of bulls that will have to be provided on AI stations. Such a course of action has thus to be thought of carefully before implementation. From a scientific point of view, EDs that were shown to influence breeding values at least when comparing their extreme values might be used for further studies including genomic data in order to find specific regions of the genome that respond differently in different environments.

Our results are in accordance with other studies on GxE interaction with a wide range of types of EDs. Mostly, no substantial GxE interaction could be detected except when extremely different groups were investigated (e.g. populations in different countries with different climatic and/or management conditions as, for example, in Ojango & Pollott, 2002). For the same population as in our study, the German-Austrian Fleckvieh population, Gerber et al. (2008) also did not find any considerable GxE interaction when studying herd performance level as categorical EDs and production traits as phenotypes. Pfeiffer, Fuerst, Schwarzenbacher, and Fuerst-Waltl (2016) used high-performance and low-performance conventional farms and organic farms as categorical EDs. The authors found genetic correlations between those environments of greater than or equal to 0.886 for all traits studied. Studies based on production trait data in German Holstein cattle populations include work with environments described by herd size, region and performance level. König et al. (2005) estimated a genetic correlation of 0.79 for protein yield with ED herd size in two classes in German Holstein data while the correlation was at least 0.9 when the ED was describing regions in Germany (Western vs. Eastern states). Streit et al. (2012) used reaction norm models with ED milk energy yield level of the herd in German Holstein data and found genetic correlations to be always greater than 0.89 for traits milk, fat and protein yield. For first lactation production traits in Dutch dairy herd data, Calus et al. (2002) discovered no considerable GxE interaction with EDs protein performance level of the herd and herd size. Nauta, Veerkamp, Brascamp, and Bovenhuis (2006) estimated genetic correlations for production traits in conventional and organic Holstein farms in the Netherlands and obtained values of 0.8, 0.88 and 0.71 for milk, fat and protein yield, respectively. Schmid et al. (2021) evaluated data from the German-Austrian Brown Swiss population and could not discover substantial GxE interaction for milk, fat and protein yield with EDs categories conventional vs organic farm system (no genetic correlation <0.93) and altitude of the farm below and above 800 m (none <0.87). Basic forage system as ED

was studied by Boettcher, Fatehi, and Schutz (2003) who distinguished between farms that rely on stored feeds vs those who intensively grazed the cows for at least half of the year. For production traits, none of the estimated genetic correlations was <0.88 . Hayes, Carrick, Bowman, and Goddard (2003) found a genetic correlation for protein yield with extreme levels of THI (60 and 78) of 0.9 in Australian Holstein data. Brügemann et al. (2011) assessed possible GxE interactions for protein yield in German Holstein with different THI values. Genetic correlation between all combinations of occurring THI values was greater than 0.9 in their study, too.

In this study, yield deviations on the first test day or cumulated over the first 100 days or the whole (305d) first lactation were used as phenotypes. For computational reasons, we wanted to restrict ourselves to first lactation observations which implies that information for the estimation of herd-related fixed effects is scarce given the herd structures in Bavaria (many of the herds are relatively small). YDs were thus chosen instead of raw phenotypes in order to benefit from well-estimated fixed effects from the more sophisticated routine breeding value estimation system including data from all lactations and all years. We nevertheless performed some bivariate GxE analyses with raw 100-day-yields for milk, fat or protein (extrapolated from the routine milking records) to preclude any misinterpretation of our results based on YD phenotypes because of any kind of systematic bias (e.g. reduced GxE interaction due to unintended pre-corrections). No evidence for such a bias could be verified based on these extra analyses (results not shown). In the official routine system, heterogeneous variance correction is applied in a random regression test day model (Emmerling, Lidauer, & Mäntysaari, 2002; Lidauer et al., 2008). We avoided to obtain YDs used for this study from a model with such pre-correction in order to avoid confusion. Nevertheless, we saw from further analyses that there was hardly an influence on the estimates of GxE interaction when comparing results based on YDs originating from models with or without heterogeneous variance adjustment (results not shown).

Especially, for studying robustness, the management level of a farm is an important parameter that helps to describe different production environments. The classical approach is to use herd performance level (e.g. Hayes et al., 2003; Pfeiffer et al., 2016; Streit et al., 2012). In this study, we used herd-year-effects for milk yield (HYSM) and for fat-plus-protein-yield (HYSFP) which were a by-product of the routine breeding value estimation and pre-corrected for known fixed and random effects. We excluded the most extreme values (below 0.01% and above 99.9% quantiles) but did not additionally filter for herd size. Since Calus et al. (2004) showed that small herd sizes might influence the estimated parameters in reaction

norm models, we ran some extra analyses for phenotypes MY_L and MY_Day1 with ED HYSFP in which we excluded data from farms with a herd size of less than 30 or 60 milking cows, respectively, and calculated the estimated genetic correlations between different environments (1%, 25%, 50%, 75% and 99% of the ED in the original analyses with each other). While we lost one third and two third of the data with herd size requirement of ≥ 30 and 60 cows, respectively, values for original analyses and analyses with restriction on herd sizes were very similar (results not shown). We thus do not assume that our original filtering biased the results.

When resilience is defined as the reaction of an individual to a short disturbance, EDs have to describe environmental parameters that change during the life (or lactation) of a cow. The background of the available phenotypic data in our study is the routine milk recording data in Bavaria which is the basis for the routine genetic evaluation. Up to now, only monthly recorded data is used there. Especially for EDs that should be used to study resilience properties of individuals, this time scale is often not detailed enough. If there was, for example, a heat stress event on a specific farm, but no milk recording in these days, the information about the reaction of the cows is missing. The same is true, for example, for somatic cell score spikes due to mastitis or fat metabolism anomalies that might be depicted by extreme fat-to-protein-ratios. Data from robotic milking system may help to overcome this and were shown to be extremely interesting by, for example, Poppe, Mulder, Kamphuis, and Veerkamp (2021). Poppe et al. (2021) and Poppe, Bonekamp, van Pelt, and Mulder (2021) used daily milk yield observations of cows of 2000 herds, calculated residuals of its specific lactation curve for each cow and studied variance, skewness and autocorrelation of residuals in order to check for new resilience traits. Variance of residuals had the highest heritability estimates with around 0.15. Recent studies with chicken and pig data have also shown that the variation of deviations of expected curves may be a good resilience indicator (Berghof, Bovenhuis, & Mulder, 2019; Putz et al., 2019).

The correlation between the continuous EDs used in this study were very low in general except for HYSFP and HYSM and for some of the studied THI-based EDs. Thus, for a future study, it might be interesting to combine two or some of them in order to estimate their GxE contribution simultaneously. A model with a continuous ED modelled within different class environments was used in some applications in dairy (Windig, Calus, Beerda, & Veerkamp, 2006; Windig, Mulder, Bohthe-Wilhelmus, & Veerkamp, 2011) and beef breeding (Santana et al., 2015). Given our EDs, it might thus, for example, be possible to study PP% within CONCFEED4

or THIminus3 within HYSFP in classes. Another possibility would be to combine different EDs via a principal component approach as performed, for example, in Windig et al. (2006). The authors used the four largest principal components obtained from 65 environmental variables and applied those to different reaction norm models. Combining different EDs might also be useful for different categorical EDs, for example, for information sources like weather station data plus information of barn type or grassland proportion plus information about the type of concentrated feed supply.

One major drawback of many GxE-studies is the non-availability of precise environmental descriptors which leads to the usage or derivation of more general and thus mostly less accurate proxies. This may cause an underestimation of existing GxE-interaction. In this study, we used different THI values calculated from measures of the weather station closest to the barn in which a cow's performance is recorded. This is a proxy but cannot consider the individual situation for a specific cow, i.e. if its barn is cooled down with ceiling fans, if the barn is directed to the south with a lot of solar radiation or the cows have access to pasture. Temperature and humidity measures directly from a barn (as e.g. used in Lambertz, Sanker, & Gauly, 2014) or body temperature measures (e.g. via infrared cameras, see, for example, Hoffmann, Schmidt, & Ammon, 2016) would be more precise measures to identify heat-stress events for a particular individual or herd. However, such values are often not available for population wide data. Similar problems exist for description of the level of energy input or generally the feeding regime. We decided to use the ratio of permanent grassland to the sum of permanent grassland and fodder maize and the type of concentrated feed supply as proxies in this study. Further projects are underway that will enable us to have access to barn-specific weather station data and farm-specific feed ration data.

For studying the influence of weather conditions, especially with a focus on heat stress events, we used a temperature–humidity index (THI) in a very classical definition (NRC, 1971). There has been lot of research on what is the best way of calculating THI (e.g. Bohmanova, Misztal, & Cole, 2007) and what is the lower threshold for heat stress (e.g. Brügemann et al., 2011). One should not forget that the origin of THI lies in an attempt to calculate the energy demand for air conditioning systems in the US and was used to describe discomfort in human when specific THI values were exceeded (Thom, 1959). Thom (1959) found a value of 70 from which on people began feeling uncomfortable. In animal literature, different values for the lower limit of THI values from which on heat stress reaction could be observed are described. Since THIs calculated with different formulas scale differently, these absolute values

have to be interpreted with caution. Values include 60 (e.g. Brügemann et al., 2012; Hayes et al., 2003), 70 (Brügemann et al., 2012) or 72 (e.g. Ravagnolo & Misztal, 2002) for different formulas, breeds and regions. In our study, we found only slight influence of THI on breeding values for the traits milk, fat and protein yield. This might be due to the following aspects: THI values exceeding a limit from which on heat stress is expected are relatively seldom (17.7% of the data with a THI > 60) in our data. The effect of extreme THI values on phenotypes was less pronounced when working with YDs than with corresponding raw data (data not shown). Our conclusion is that extreme weather conditions did not occur frequently enough in Bavaria to detect significant GxE interaction. This might change in future. In further analyses, it might also be worth to test an approach in which the model assumes that THI affects performances only above a certain threshold as, for example, used by Bohmanova, Misztal, Tsuruta, Norman, and Lawlor (2008) or Nguyen et al. (2016).

An analysis studying GxE interaction with a model assuming genetic, random error and GxE variances implicitly assumes that there is a balanced contribution of genotypes to the range of environments. If this is not the case, additionally a GxE covariance can be considered in the model which can capture variation that is due to a non-random distribution to environments (Lynch & Walsh, 1998). Regarding our dairy data, for example, for ORG or HYSM, there might be specific bulls that contribute almost exclusively to one of the different classes (e.g. bulls that are mainly used in organic farms) or specific parts of the whole range (e.g. bulls with very high performance breeding values that are mainly used in farms with a high level of HYSM). Omitting the GxE covariance in the model in such cases might lead to inflated estimates of the GxE interaction as shown by Ni et al. (2019) in human data. The aspect of GxE covariance will have to be studied more closely in future also in dairy data.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from European Union's Horizon 2020 research and innovation programme — GenTORE — under grant agreement N° 727213. We thank the Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V. (LKV) for providing data that we used for describing farm environments. We further thank Magdalena Ochsenbauer and Michael Gratzl who both supported us with weather station and IACS data management and Hildegard Feigel who performed the geocoding of the farm address data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Author elects not to share data. The data used in this study stem from the routine genetic evaluation for Fleckvieh augmented by publicly available weather and elevation data. The data from routine evaluation are considered personal data in the sense of the General Data Protection Regulation (EU 2016/679) and may not be distributed without consent of the farmers concerned.

ORCID

Malena Erbe  <https://orcid.org/0000-0001-6301-3341>

REFERENCES

- Berghof, T. V. L., Bovenhuis, H., & Mulder, H. A. (2019). Body weight deviations as indicator for resilience in layer chickens. *Frontiers in Genetics, 10*, 1216. <https://doi.org/10.3389/fgene.2019.01216>
- Boettcher, P. J., Fatehi, J., & Schutz, M. M. (2003). Genotype \times environment interactions in conventional versus pasture-based dairies in Canada. *Journal of Dairy Science, 86*(1), 383–389. [https://doi.org/10.3168/jds.S0022-0302\(03\)73617-0](https://doi.org/10.3168/jds.S0022-0302(03)73617-0)
- Bohmanova, J., Misztal, I., Tsuruta, S., Norman, H. D., & Lawlor, T. J. (2008). Short communication: Genotype by environment interaction due to heat stress. *Journal of Dairy Science, 91*(2), 840–846. <https://doi.org/10.3168/jds.2006-142>
- Bohmanova, J., Misztal, I., & Cole, J. B. (2007). Temperature-humidity indices as indicators of milk production losses due to heat stress. *Journal of Dairy Science, 90*(4), 1947–1956. <https://doi.org/10.3168/jds.2006-513>
- Brügemann, K., Gernand, E., König von Borstel, U., & König, S. (2012). Defining and evaluating heat stress thresholds in different dairy cow production systems. *Archives Animal Breeding, 55*(1), 13–24. <https://doi.org/10.5194/aab-55-13-2012>
- Brügemann, K., Gernand, E., von Borstel, U. U., & König, S. (2011). Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature \times humidity-dependent covariates. *Journal of Dairy Science, 94*(8), 4129–4139. <https://doi.org/10.3168/jds.2010-4063>
- Calus, M. P. L., Windig, J. J., & Veerkamp, R. F. (2005). Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. *Journal of Dairy Science, 88*(6), 2178–2189. [https://doi.org/10.3168/jds.S0022-0302\(05\)72893-9](https://doi.org/10.3168/jds.S0022-0302(05)72893-9)
- Calus, M. P., Bijma, P., & Veerkamp, R. F. (2004). Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model. *Genetics Selection Evolution, 36*(5), 489–507. <https://doi.org/10.1186/1297-9686-36-5-489>
- Calus, M. P. L., Groen, A. F., & de Jong, G. (2002). Genotype \times environment interaction for protein yield in dutch dairy cattle as quantified by different models. *Journal of Dairy Science, 85*(11), 3115–3123. [https://doi.org/10.3168/jds.S0022-0302\(02\)74399-3](https://doi.org/10.3168/jds.S0022-0302(02)74399-3)
- Colditz, I. G., & Hine, B. C. (2016). Resilience in farm animals: Biology, management, breeding and implications for animal welfare. *Animal Production Science, 56*(12), 1961–1983. <https://doi.org/10.1071/AN15297>
- de Jong, G. (1995). Phenotypic plasticity as a product of selection in a variable environment. *The American Naturalist, 145*(4), 493–512.
- Dominik, S., & Kinghorn, B. P. (2008). Neglecting genotype \times environment interaction results in biased predictions from selection index calculations. *Livestock Science, 114*(2–3), 233–240. <https://doi.org/10.1016/j.livsci.2007.05.004>
- Emmerling, R., Lidauer, M., & Mäntysaari, E. A. (2002). Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. *Interbull Bulletin, 29*, 111–117.
- Enemark, J. M. D. (2008). The monitoring, prevention and treatment of sub-acute ruminal acidosis (SARA): A review. *The Veterinary Journal, 176*(1), 32–43. <https://doi.org/10.1016/j.tvjl.2007.12.021>
- Falconer, D. S. (1952). The problem of environment and selection. *The American Naturalist, 86*(830), 293–298.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Longmans Green.
- Friggens, N. C., Adriaens, I., Boré, R., Cozzi, G., Jurquet, J., Kamphuis, C. ... De Haas, Y. (2022). Resilience: Reference measures based on longer-term consequences are needed to unlock the potential of precision livestock farming technologies for quantifying this trait. *Zenodo*, 5215797, ver. 5 peer-reviewed and recommended by Peer community in Animal Science. doi: <https://doi.org/10.5281/zenodo.5215797>
- Gerber, A., Krogmeier, D., Emmerling, R., & Götz, K.-U. (2008). Analysis of genotype by environment interaction for milk yield traits in first lactation of Simmental cattle. *Journal of Animal Breeding and Genetics, 125*(6), 382–389. <https://doi.org/10.1111/j.1439-0388.2008.00731.x>
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2009). ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, www.vsn.co.uk
- Ha, N.-T., Sharifi, A. R., Heise, J., Schlather, M., Schnyder, U., Gross, J. J., ... Simianer, H. (2017). A reaction norm sire model to study the effect of metabolic challenge in early lactation on the functional longevity of dairy cows. *Journal of Dairy Science, 100*(5), 3742–3753. <https://doi.org/10.3168/jds.2016-12031>
- Hayes, B. J., Carrick, M., Bowman, P., & Goddard, M. E. (2003). Genotype \times environment interaction for milk production of daughters of Australian dairy sires from test-day records. *Journal of Dairy Science, 86*(11), 3736–3744. [https://doi.org/10.3168/jds.S0022-0302\(03\)73980-0](https://doi.org/10.3168/jds.S0022-0302(03)73980-0)
- Heuer, C., Van Straalen, W. M., Schukken, Y. H., Dirkwager, A., & Noordhuizen, J. P. T. M. (2000). Prediction of energy balance in a high yielding dairy herd in early lactation: Model development and precision. *Livestock Production Science, 65*(1–2), 91–105. [https://doi.org/10.1016/S0301-6226\(99\)00177-3](https://doi.org/10.1016/S0301-6226(99)00177-3)
- Hoffmann, G., Schmidt, M., & Ammon, C. (2016). First investigations to refine video-based IR thermography as a non-invasive tool to monitor the body temperature of calves. *Animal, 10*(9), 1542–1546. <https://doi.org/10.1017/S1751731115001354>
- Kirkpatrick, M., Lofsvold, D., & Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics, 124*(4), 979–993. <https://doi.org/10.1093/genetics/124.4.979>
- König, S., Dietl, G., Raeder, I., & Swalve, H. H. (2005). Genetic relationships for dairy performance between large-scale and small-scale farm conditions. *Journal of Dairy Science, 88*(11), 4087–4096. [https://doi.org/10.3168/jds.S0022-0302\(05\)73093-9](https://doi.org/10.3168/jds.S0022-0302(05)73093-9)

- Lambertz, C., Sanker, C., & Gauly, M. (2014). Climatic effects on milk production traits and somatic cell score in lactating Holstein-Friesian cows in different housing systems. *Journal of Dairy Science*, 97(1), 319–329. <https://doi.org/10.3168/jds.2013-7217>
- Lidauer, M. H., Emmerling, R., & Mäntysaari, E. A. (2008). Multiplicative random regression model for heterogeneous variance adjustment in genetic evaluation for milk yield in Simmental. *Journal of Animal Breeding and Genetics*, 125(3), 147–159. <https://doi.org/10.1111/j.1439-0388.2008.00728.x>
- Lillehammer, M., Ødegård, J., & Meuwissen, T. H. E. (2009). Reducing the bias of estimates of genotype by environment interactions in random regression sire models. *Genetics Selection Evolution*, 41(1), 30. <https://doi.org/10.1186/1297-9686-41-30>
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.
- Mulder, H. A., Groen, A. F., De Jong, G., & Bijma, P. (2004). Genotype × environment interaction for yield and somatic cell score with automatic and conventional milking systems. *Journal of Dairy Science*, 87(5), 1487–1495. [https://doi.org/10.3168/jds.S0022-0302\(04\)73300-7](https://doi.org/10.3168/jds.S0022-0302(04)73300-7)
- National Research Council (U.S.). (1971). *A guide to environmental research on animals*. National Academy of Sciences.
- Nauta, W. J., Veerkamp, R. F., Brascamp, E. W., & Bovenhuis, H. (2006). Genotype by environment interaction for milk production traits between organic and conventional dairy cattle production in The Netherlands. *Journal of Dairy Science*, 89(7), 2729–2737. [https://doi.org/10.3168/jds.S0022-0302\(06\)72349-9](https://doi.org/10.3168/jds.S0022-0302(06)72349-9)
- Ni, G., van der Werf, J., Zhou, X., Hyppönen, E., Wray, N. R., & Lee, S. H. (2019). Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. *Nature Communications*, 10(1), 2239. <https://doi.org/10.1038/s41467-019-10128-w>
- Nguyen, T. T. T., Bowman, P. J., Haile-Mariam, M., Pryce, J. E., & Hayes, B. J. (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. *Journal of Dairy Science*, 99(4), 2849–2862. <https://doi.org/10.3168/jds.2015-9685>
- Ojango, J. M. K., & Pollott, G. E. (2002). The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. *Livestock Production Science*, 74(1), 1–12. [https://doi.org/10.1016/S0301-6226\(01\)00282-2](https://doi.org/10.1016/S0301-6226(01)00282-2)
- Pfeiffer, C., Fuerst, C., Schwarzenbacher, H., & Fuerst-Waltl, B. (2016). Genotype by environment interaction in organic and conventional production systems and their consequences for breeding objectives in Austrian Fleckvieh cattle. *Livestock Science*, 185, 50–55. <https://doi.org/10.1016/j.livsci.2016.01.011>
- Poppe, M., Bonekamp, G., van Pelt, M. L., & Mulder, H. A. (2021). Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. *Journal of Dairy Science*, 104(2), 1967–1981. <https://doi.org/10.3168/jds.2020-19245>
- Poppe, M., Mulder, H. A., Kamphuis, C., & Veerkamp, R. F. (2021). Between-herd variation in resilience and relations to herd performance. *Journal of Dairy Science*, 104(1), 616–627. <https://doi.org/10.3168/jds.2020-18525>
- Putz, A. M., Harding, J. C. S., Dyck, M. K., Fortin, F., Plastow, G. S., Dekkers, J. C. M., & PigGen Canada. (2019). Novel resilience phenotypes using feed intake data from a natural disease challenge model in wean-to-finish pigs. *Frontiers in Genetics*, 9, 660. <https://doi.org/10.3389/fgene.2018.00660>
- Quiédeville, S., Moakes, S., Leiber, F., & Pfeifer, C. (2020). Dairy database for prediction of main environmental challenges to resilience and efficiency in cattle production systems at regional resolution (Version V1.0) [Data set]. Zenodo. doi: <https://doi.org/10.5281/zenodo.3860704>
- Ravagnolo, O., & Misztal, I. (2002). Studies on genetics of heat tolerance in dairy cattle with reduced weather information via cluster analysis. *Journal of Dairy Science*, 85(6), 1586–1589. [https://doi.org/10.3168/jds.S0022-0302\(02\)74228-8](https://doi.org/10.3168/jds.S0022-0302(02)74228-8)
- Rauw, W. M., & Gomez-Raya, L. (2015). Genotype by environment interaction and breeding for robustness in livestock. *Frontiers in Genetics*, 6, 310. <https://doi.org/10.3389/fgene.2015.00310>
- Santana, M. L., Eler, J. P., Bignardi, A. B., Menéndez-Buxadera, A., Cardoso, F. F., & Ferraz, J. B. S. (2015). Multi-trait linear reaction norm model to describe the pattern of phenotypic expression of some economic traits in beef cattle across a range of environments. *Journal of Applied Genetics*, 56(2), 219–229. <https://doi.org/10.1007/s13353-014-0242-9>
- Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science*, 86(1–3), 35–45. [https://doi.org/10.1016/S0301-6226\(03\)00151-9](https://doi.org/10.1016/S0301-6226(03)00151-9)
- Schaeffer, L. R. (1994). Multiple-country comparison of dairy sires. *Journal of Dairy Science*, 77(9), 2671–2678. [https://doi.org/10.3168/jds.S0022-0302\(94\)77209-X](https://doi.org/10.3168/jds.S0022-0302(94)77209-X)
- Schaeffer, L. R., & Dekkers, J. C. M. (1994). Random regressions in animal models for test-day production in dairy cattle. *Proceedings of Fifth World Congress of Genetics Applied to Livestock Production*, 18, 443–446.
- Schmid, M., Imort-Just, A., Emmerling, R., Fuerst, C., Hamann, H., & Bennewitz, J. (2021). Genotype-by-environment interactions at the trait level and total merit index level for milk production and functional traits in Brown Swiss cattle. *Animal*, 15(1), 100052. <https://doi.org/10.1016/j.animal.2020.100052>
- Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50(4), 1171–1177. <https://doi.org/10.2307/2533455>
- Strandberg, E., Kolmodin, R., Madsen, P., Jensen, J., & Jorjani, H. (2000). Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Interbull Bulletin*, 25, 41–45.
- Streit, M., Reinhardt, F., Thaller, G., & Bennewitz, J. (2012). Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle: G×E in German Holsteins. *Journal of Animal Breeding and Genetics*, 129(5), 380–389. <https://doi.org/10.1111/j.1439-0388.2012.00999.x>
- Su, G., Lund, M. S., & Sorensen, D. (2007). Selection for litter size at day five to improve litter size at weaning and piglet survival rate. *Journal of Animal Science*, 85(6), 1385–1392. <https://doi.org/10.2527/jas.2006-631>
- Thom, E. C. (1959). The discomfort index. *Weatherwise*, 12, 57–60.
- Toghiani, S., Hay, E., Fragomeni, B., Rekaya, R., & Roberts, A. J. (2020). Genotype by environment interaction in response to cold stress in a composite beef cattle breed. *Animal*, 14(8), 1576–1587. <https://doi.org/10.1017/S1751731120000531>
- VanRaden, P. M. (1997). Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. *Journal of Dairy Science*, 80(11), 3015–3022. [https://doi.org/10.3168/jds.S0022-0302\(97\)76268-4](https://doi.org/10.3168/jds.S0022-0302(97)76268-4)
- Windig, J. J., Mulder, H. A., Bohthe-Wilhelmus, D. I., & Veerkamp, R. F. (2011). Simultaneous estimation of genotype by environment

interaction accounting for discrete and continuous environmental descriptors in Irish dairy cattle. *Journal of Dairy Science*, 94(6), 3137–3147. <https://doi.org/10.3168/jds.2010-3725>

Windig, J. J., Calus, M. P. L., Beerda, B., & Veerkamp, R. F. (2006). Genetic correlations between milk production and health and fertility depending on herd environment. *Journal of Dairy Science*, 89(5), 1765–1775. [https://doi.org/10.3168/jds.S0022-0302\(06\)72245-7](https://doi.org/10.3168/jds.S0022-0302(06)72245-7)

Woltereck, R. (1909). Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artenunterschiede bei Daphniden. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, 1909, 110–172.

Würfl, P., Dörfler, J., & Rintelen, P.-M. (1984). Die Einteilung Bayerns in Landwirtschaftliche Standorte, Landwirtschaftliche Erzeugungsgebiete und Agrargebiete. *Bayerisches Landwirtschaftliches Jahrbuch*, 61, 377–423.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Erbe, M., Edel, C., Emmerling, R., & Götz, K.-U. (2022). Assessing various environmental descriptors with respect to genotype x environment interaction for milk production traits in Bavarian Fleckvieh cattle. *Journal of Animal Breeding and Genetics*, 139, 634–653. <https://doi.org/10.1111/jbg.12722>