Beerling, S., Vinet, A., Cuyabano, B., Plantard, Y., Dominique, S., Mattalia, S., Croiseau, P., Aguerre, S. (2026) Exploring farm-environment descriptors to assess their potential for GxE interactions on milk production in Holstein dairy cattle, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 24026: https://doi.org/10.31274/wcgalp.24026
©2026 Beerling, et al. CC BY-NC-ND

Exploring farm-environment descriptors to assess their potential for GxE interactions on milk production in Holstein dairy cattle

S. Beerling[1,2], A. Vinet[1], B.C.D. Cuyabano[*1], Y. Plantard[1,2], S. Dominique[2], S. Mattalia[1,2], P. Croiseau[1], and S. Aguerre[1,2]

[1]*Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France*
[2]*IDELE, 149 Rue de Bercy, 75012, Paris, France;* [*]*Corresponding author : beatriz.castro-dias-cuyabano@inrae.fr*

Introduction

Current genetic evaluations are based on trait performances and genetic relationships, with environmental effects such as herd, year, and season accounted for as an effect independent from genetics. Although, several studies have already established how to account for genotype-by-environment (GxE) interactions on the temperature humidity index (THI) indicators (Vinet *et al*., 2023, 2024; Silva Neto *et al*., 2024), little is explored regarding health factors or the rearing environment, or more specifically the sanitary environment (Mulder *et al*., 2013; Silva Neto *et al*., 2024). To model the sanitary environment into GxE interactions, however, informative descriptors are required. The aim of our study was to identify these descriptive variables of the herds' sanitary environment for the French Holstein dairy cattle, and create a summarizing and continuous variable for the sanitary environment, to enable future GxE interaction modelling through random regressions (Schaeffer, 2004; Silva Neto *et al*., 2024). We implemented genetic evaluations of 305-day milk yield with GxE interactions across three different summarized sanitary variables, on a French Holstein dairy cattle population, to test a first approach for these environmental variables. Although results of the pilot study suggested no significant GxE interactions, the construction of variables that summarize multiple information of herd environment must be further explored for optimization of results.

Materials & Methods

Herds' cell effects. French genetic evaluations of somatic cell score (SCS) use all somatic cell count (SCC) records between 5-450 days in milk in lactations 1-3. The data for our study was issued from the evaluation using data from 1989 to 2024, following the model:

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Zu}+\mathbf{Wp}+\mathbf{e}, \quad (1)$$

where **y** it the vector of performances; **b** is the vector of fixed effects including combinations of herd-year, lactation-region-year, calving_month-lactation-region-year, age_at_calving-lactation-region-year, length_previous_dry_period (for lactation > 1); **u** and **p** are respectively the vectors of genetic and permanent environment effects; **X**, **Z**, **W** are the incidence matrices associated to the described effects, and **e** is the vector of random residuals. For more information (in French) about the model see https://geneval.fr/indexations-races-bovines/.

Three approaches were tested to construct summarizing and continuous environmental gradients for the French Holstein dairy cattle. The first two created environmental gradients from the herds' cell effects derived from the French genetic evaluations of the SCS, considered only herds with stable sanitary environments during 2018 to 2022. This five-year period was long enough to represent herds' overall health status while minimizing the impact of extreme events or changes in herd management. The third approach created a synthetic variable using variables from an information system containing global farming descriptors. The next three subsections detail the construction of the sanitary environmental gradients.

Quantiles (Qt_eft). For each year (2018-2022), herds were categorized on an integer scale 1-10, corresponding to the 10-quantile class each herd fell into. Then, for each herd, the mean and standard deviation of their five yearly quantiles were computed to (1) obtain a single mean quantile per herd, creating the gradient Qt_eft and (2) identify and select the herds with a quantile standard deviation ≤ 1.71, a proxy of their health environment stability.

Reduced centred herd effects (RC_eft). Herd effects were centred and standardized across all five years (2018-2022). Herds were then assigned an integer class 1-10, each covering an increasing and equal range of SCS at the herds. Then, for each herd, the mean and standard deviation of their five yearly classes were calculated with the same purposes as for Qt_eft, with the threshold to consider the sanitary environment stable when its standard deviation ≤ 0.59.

Variables from information system (Synth_sanit). The third environmental gradient was based on the analysis of 77 environmental variables from a French information system that provides data on the global farming environment, e.g. herd size, overall production level, health and reproduction management. A principal component analysis (PCA) was used to assess variables' relevance in characterizing the different farms, and clusters to create a synthetic variable. Synth_sanit was created as the weighted sum of the cluster of six variables related to the sanitary environment of the herd (weights were defined from the PCA outputs for the first two main components).

GxE studies. The environmental gradients (Qt_eft, RC_eft and Synth_sanit) were used to model GxE for 305-day milk yield (MY305d). We used records of MY305d from primiparous French Holstein dairy cows with minimum lactation length of 180 days, and contemporary groups (herd-year) with at least five cows for at least three of the five years. A total of 88,143 records/cows from 1,186 herds were included in the GxE studies performed with Qt_eft and RC_eft, and a total of 36,608 records/cows from 625 herds were included in the GxE study with Synth_sanit. The following sire random regression model was applied:

$$\mathbf{y}=\mathbf{X\beta}+\mathbf{Zg}+\mathbf{e} \quad (2)$$

where **y** is the vector of 305-day milk yield; **β** is the vector of fixed effects (age at calving, calving month, and herd-year as the contemporary group) and **X** its corresponding design matrix. The additive genetic effects were randomly regressed on a constant and first and second degree Legendre polynomials on either Qt_eft, RC_eft, or Synth_Sanit, being **Z** the polynomials covariate matrix and **g** the vectors of their corresponding random genetic coefficients, with variances proportional to the pedigree relationship **A**. Lastly, **e** is the vector of independent residuals, with heterogeneous variances. This model was implemented using WOMBAT (Meyer, 2007).

Results & Discussion

Environmental gradients. Both Qt_eft and RC_eft allowed the identification of herds with overall favorable or unfavorable environments. Qt_eft however, was not optimal for distinguishing herds with an average sanitary environment. Figure 1 presents the first two PCA components of the 77 farm descriptors, color-coding ten clusters identified by a dendrogram (not shown). The cluster of red variables (1, 23, 29, 34, 35, and 55) corresponds to the six variables used to create the environmental gradient Synth_sanit, related to herds' sanitary status. Synth_sanit was built as a weighted sum of these variables, with the weights (0.09, 0.08, 0.03, 0.05, 0.07 and 0.05 respectively) being the variables' contribution in the first two PCA-components.

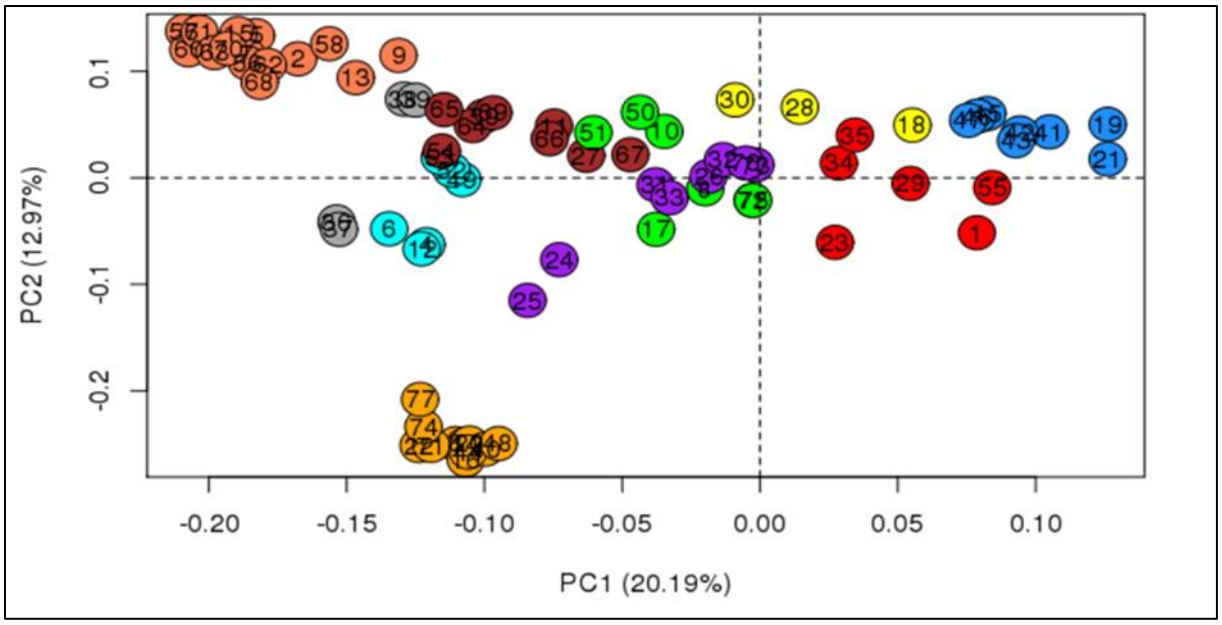

Figure 1. First two PCA-components of the 77 descriptive variables.

GxE studies. Genetic standard deviations (GSD) and heritabilities followed the same trend for all three environmental gradients. Higher values were estimated for favorable and unfavorable environments compared to average environments, suggesting that genetic variability is higher in extreme environments. Figure 2 presents the genetic standard deviations for MY305d across Synth_sanit. Results should be analyzed with caution; the higher at extreme environments may be a feature of under-represented data at extremes environments. Moreover, the model deployed a two-degree polynomial, and future studies should test higher degrees of Legendre polynomials. Genetic correlations (r_g) between MY305d across the three gradients of sanitary environments were all superior to 0.9, thus suggesting no significant GxE interactions for MY305d with respect to the sanitary herd environment. Similar to the discussed for the genetic standard deviations, future studies should test higher degrees of Legendre polynomials to verify if indeed no significant GxE interactions are present across different sanitary environments for MY305d.

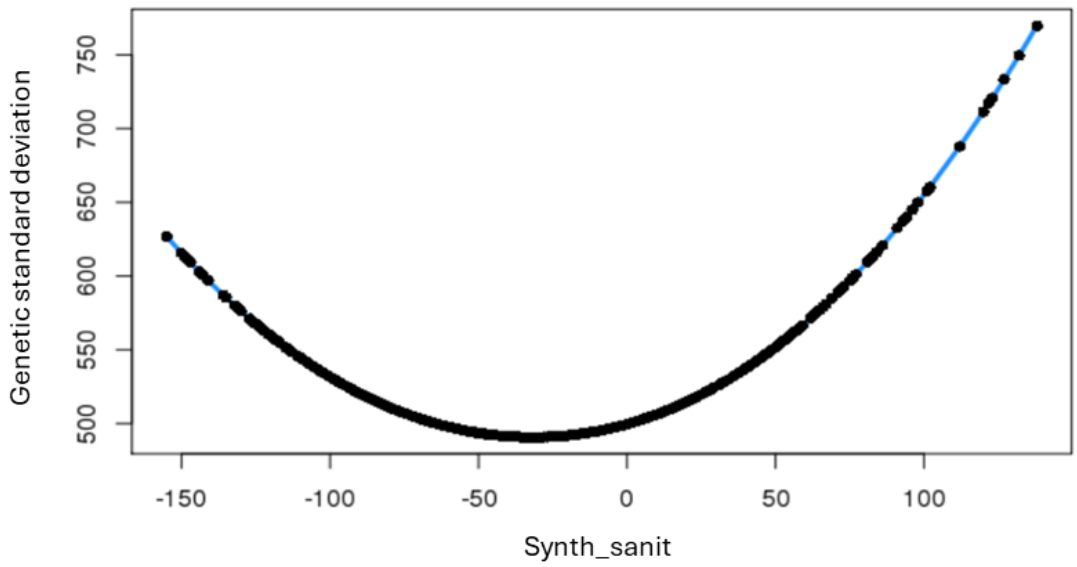

Figure 2. Genetic standard deviation for MY305d across Synth_sanit values.

Conclusion

Accurate environmental characterization is a relevant challenge to address for future genetic evaluations, expected to increasingly account GxE interactions for optimal selection of the best candidates for specific environments. This study is the first to explore different sources of information related to the farming micro-environment, and among the first studies to test approaches for creating comprehensive environmental descriptors. While further studies are needed to optimally characterize and summarize the farming environment through several descriptors, our work opens new pathways for GxE interactions studies and implementation in genetic evaluations. Future perspectives of this work will continue in the framework of the CEVADE project and its associated PhD project (APIS-GENE).

Acknowledgements

Silja Beerling is a recipient of a CIFRE PhD grant from IDELE, with financial support of the ANRt and APIS-GENE (CEVADE project). Part of the study relied on the results obtained by the OBGENO project funded by APIS-GENE.

References

Meyer K. (2007) J Zhejiang Univ Sci B 8(11):815-21. doi.org/10.1631/jzus.2007.B0815

Mulder H. A., Rönnegård L., Fikse W. F., *et al*. (2013), Genet Sel Evol 45(1). doi.org/10.1186/1297-9686-45-23

Schaeffer L. R. (2004). Livestock Prod Sci 86:35–45. https://doi.org/10.1016/S0301-6226(03)00151-99

Silva Neto J. B., Mota L. F. M., Londoño-Gil M., *et al*. (2024). Anim Genet 55:871–92. doi.org/10.1111/AGE.13483

Vinet A., Mattalia S., Vallée R., *et al*. (2023) Genet Sel Evol 55:4. doi.org/10.1186/s12711-023-00779-1

Vinet A., Mattalia S., Vallée R., *et al*. (2024) Genet Sel Evol 56:23. doi.org/10.1186/s12711-024-00889-4