



HAL
open science

Characterization of French dairy farm environments from herd-test-day profiles

Bérénice Huquet Huquet, H. H. Leclerc, Vincent Ducrocq

► **To cite this version:**

Bérénice Huquet Huquet, H. H. Leclerc, Vincent Ducrocq. Characterization of French dairy farm environments from herd-test-day profiles. *Journal of Dairy Science*, 2012, 95 (7), pp.4085-4098. 10.3168/jds.2011-5001 . hal-00999970

HAL Id: hal-00999970

<https://hal.science/hal-00999970v1>

Submitted on 28 May 2020

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

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



Characterization of French dairy farm environments from herd-test-day profiles

B. Huquet,*†¹ H. Leclerc,† and V. Ducrocq*

*INRA, UMR1313 Génétique Animale et Biologie Intégrative (GABI), 78352 Jouy-en-Josas, France

†Institut de l'Élevage, 149 rue de Bercy, 75012 Paris, France

ABSTRACT

Genotype by environment interactions are ignored in national genetic evaluations of dairy cattle. However, some breeders consider that the genetic merit of top sires is not correctly estimated for use in their own particular environment or with their own herd management. With the objective of later investigating genotype by environment interactions at the national level, we studied the relationship between herd management (e.g., feeding system, herd size, production, workforce) and herd-test-day (HTD) profiles for milk yield and contents of fat and protein using a database of 934 herds. Herd-test-day profiles, estimated using a test-day model, are thought to reflect herd management because they represent the part of production due only to month-to-month variations in environmental conditions of production; that is, those related mainly to feeding and climate conditions. Herd clustering based on results from a factor analysis of descriptors of HTD profiles was performed. Each cluster of herds was then characterized in terms of herd management. Three herd clusters were identified. Cluster 1 was characterized by low HTD profiles for milk yield and protein and fat contents as well as a low level of intensification. Cluster 2 was defined by a high milk yield HTD profile and a high level of intensification. Cluster 3 consisted of herds with HTD profiles showing high protein and fat contents but it was difficult to link this cluster to any specific herd management. The use of clusters based on HTD profiles as descriptors of distinct herd management systems will be used in a genotype by environment interaction study in France.

Key words: genotype by environment interaction, herd environment, herd-test-day

INTRODUCTION

What is the “best cow” for a breeder? For a long time it was thought that the same cow was good for everyone. It now appears that the best cow may be the one best adapted to its own local environment. Two main levers exist to adapt animals to their environment, which is defined in terms of, for example, feeding, climate, and soil fertility. The first one is not addressed in this paper and concerns adapting breeding goals for an entire breed to its environment; for example, by putting more weight on functional or morphological traits in breeding programs to breed more robust animals when they live in more demanding environments. The second is to take advantage of genotype by environment ($G \times E$) interactions.

The French genetic evaluation assumes, as do all national evaluations for dairy breeds, that $G \times E$ interactions do not exist (Interbull, 2011). However, because of the diversity of pedoclimatic conditions in France, dairy farms have very diversified herd management systems, with different local uses of pasture and maize, for example. For this reason, some breeders question the efficiency of the existing breeding schemes for their own management system. To overcome these concerns, a $G \times E$ interaction study at the French national level has been considered necessary.

A crucial point in any $G \times E$ interaction study is the precise categorization of the environment. Since the 1990s, a better focus on $G \times E$ interactions has led to several definitions of the environment. In 1990, Carabaño et al. (1990) defined the environment by referring to the different regions of the United States. In 1994, the International Bull Service Evaluation (Interbull) implemented multiple-trait across country evaluation (MACE), in which each country is considered as a distinct environment (Schaeffer, 1994). Later, to characterize the environment, Weigel and Rekaya (2000) and Zwald et al. (2003) defined multiple-trait herd clusters based on herd management, genetic, and climatic descriptors.

A test-day model is already used in different national evaluations; its current development for French genetic

Received September 30, 2011.

Accepted February 22, 2012.

¹Corresponding author: berenice.huquet@jouy.inra.fr

evaluation (Leclerc et al., 2008, 2009) makes it possible to draw up a new definition of the environment. Indeed, the French test-day model describes monthly milk records as functions of systematic factors such as age, month of calving, length of dry period, stage of gestation, time-dependent additive genetic and permanent environment effects at the cow level, along with measurement of the local production environment effect: the herd-test-day effect (**HTD**). This effect is shared by all cows of a herd and represents the part of production (milk, protein, and fat yields or contents) on the test-day that can be attributed specifically to feeding and climatic conditions, corrected for all other effects. Fluctuations in the HTD effect over time are the consequences of month-to-month variations in environmental conditions of production and are hereafter called HTD profiles (Figure 1). Herd-test-day profiles (milk yield and protein and fat contents) of a herd are assumed representative of its herd management (essentially feeding), as described in Mayeres et al. (2004) and Koivula et al. (2007).

This study represents the first phase of a larger project in which G×E interaction will be investigated at the French national level. The objectives of the present study were 2-fold: first, to define contrasted environments based on HTD profiles (data that are available at the national level) to use them for G×E interaction studies; and second, to relate these environments to farm-level characteristics of management and feeding systems.

The approach consisted of clustering a sample of herds according to the shape of their HTD profiles for 3 traits (milk yield and fat and protein contents). We assumed that new herds with similar HTD profiles but for which there are no known herd management characteristics available in national databases could be allocated to a particular cluster and used in future G×E interaction studies.

MATERIALS AND METHODS

Herd Management Descriptors and HTD Profiles

In 2009, a survey was carried out on 934 French herds for which herd management was well known by extension specialists. Information on feeding systems and dairy structures (herd size, production, and workforce) was collected. These herds were representative of the 3 major French dairy breeds: Montbéliarde, Normande, and Holstein. We verified that overall these herds were representative of French production systems (Table 1). Additional descriptors of the herds such as age at first calving, calving interval and lactation length for

primiparous and multiparous cows, and distribution of calving dates and lactations were included using the national database. Table A1 in the Appendix presents the descriptive statistics of the parameters used to characterize the herd management of the 934 farms.

During spring 2010, a within-breed genetic evaluation using a test-day model described in Leclerc et al. (2008, 2009) was carried out based on records from 12,916,854 Montbéliarde, Normande, and Holstein cows born after 1987 in 84,975 French herds, including the 934 herds of this study. Each trait (milk yield and protein and fat contents) was evaluated using a single-trait random regression test-day model in which the first 3 parities were considered as correlated traits. In addition to the estimation of HTD effects, the shape of the lactation curve related to calving month, calving age, length of dry period and gestation (defined by parity and region of indexation) were taken into account cumulating cubic splines for each level of these effects. Genetic and permanent environment effects were included using a reduced rank model. Each herd was then described by its 3 HTD profiles for milk yield and protein and fat contents from September 2004 to May 2010. These profiles were assumed to be the footprint of the herd management (Figure 1).

Smoothing HTD Profiles

Herd-test-day profiles were smoothed to identify repeated annual features and to buffer outlier test-day effects. Within-breed analyses inspired by the model of Koivula et al. (2007) were carried out using the WOMBAT (Meyer, 2007) software by trait (milk yield and protein and fat contents) and breed. For these analyses, the data set included HTD records from September 2004 to May 2010 of herds with at least 10 records per year (1,714 Montbéliarde herds, 3,780 Normande herds, 19,075 Holstein herds). Herd-test-day records were described using the following random regression model:

$$\begin{aligned} \text{HTD}(t, h, cm, y) = & \text{herd}_h + d_h t + a_h s1(cm) + b_h s2(cm) \\ & + c_h s3(cm) + \alpha_{hy} s1(cm) + \beta_{hy} s2(cm) \\ & + \gamma_{hy} s3(cm) + e, \end{aligned}$$

where $\text{HTD}(t, h, cm, y)$ is the HTD record of herd h at time t (t represents time in months across years, equal to 1 in September 2004, 2 in October 2004, to 13 in September 2005). A fixed effect for the herd (herd_h) was included, and d_h , a_h , b_h , and c_h are the coefficients of the fixed regressions of HTD on the different covariates estimated within herd. The covariates of the model were a linear time trend t and 3 sine functions (named

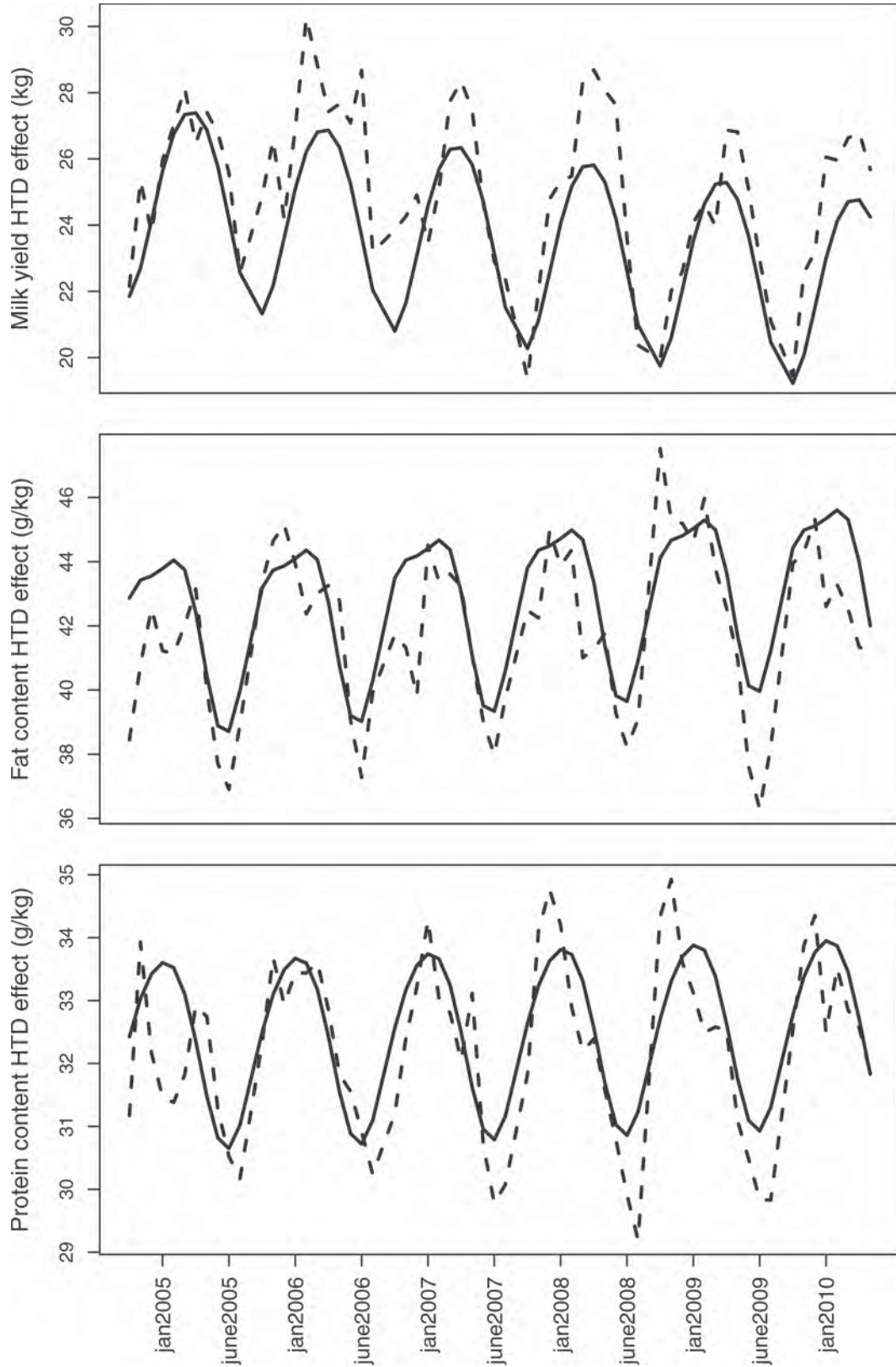


Figure 1. Example of herd-test-day (HTD) profiles (dashed line) and smoothed HTD profiles (solid line) of a herd for milk yield (R^2 milk yield = 0.71), fat content (R^2 fat content = 0.66), and protein content (R^2 protein content = 0.60).

Table 1. Number of herds in the national data set, the data set of the survey, and the data set used for the factorial and cluster analysis for Montbéliarde, Normande, and Holstein breeds and their distribution (%) in their different regions of indexation in France (3, 6, and 8 regions, respectively)

Item	Herds (no.)	Region							
		1	2	3	4	5	6	7	8
Montbéliarde									
National	9,708	57.6	27.6	14.8					
Survey	210	74.5	16.8	8.7					
Analysis	42	35.7	23.8	40.5					
Normande									
National	6,623	7.4	15.2	49.7	12.8	8.5	6.4		
Survey	90	2.2	22.2	42.2	22.2	5.6	5.6		
Analysis	83	3.6	25.5	37.3	19.3	6	7.2		
Holstein									
National	39,458	12.2	24.7	17.5	12.4	7.7	9.2	4.5	11.8
Survey	634	7.3	38.2	13.9	11.6	6.8	8.7	6.2	7.3
Analysis	327	8.9	36.3	20.3	11.1	5.5	0	8.3	9.5

s1, s2, s3) estimated within herd. The values of these sine functions depended on the calendar month (cm) of the HTD record, with $cm = 1$ in January and $cm = 2$ in February:

$$s1(cm) = \sin\left[\frac{2}{12}\pi \times (cm - 2)\right],$$

$$s2(cm) = \sin\left[\frac{2}{12}\pi \times (cm - 4)\right];$$

that is, with a 2-mo phase difference from s1, and

$$s3(cm) = \sin\left[\frac{-2}{6}\pi \times (cm + 10)\right].$$

The random regression part included 3 coefficients (α , β , γ) of the same 3 sine functions within herd \times year (hy):

$$\text{with var} \begin{bmatrix} \alpha_{hy} \\ \beta_{hy} \\ \gamma_{hy} \end{bmatrix} = \begin{bmatrix} \sigma_{\alpha}^2 & 0 & 0 \\ 0 & \sigma_{\beta}^2 & 0 \\ 0 & 0 & \sigma_{\gamma}^2 \end{bmatrix}.$$

The smoothed HTD profiles of the 934 herds for the 3 traits (Figure 1) were built from the estimated fixed effects part of the model using \hat{a}_h , \hat{b}_h , \hat{c}_h , \hat{d}_h , \widehat{herd}_h . This part represents the repeated annual pattern of the HTD profile. It is the combination of 5 components: the 3 sine curves (2 with a period of 12 mo, 1 with a period of 6 mo), a linear trend, and the herd effect.

Selection of 452 Herds with Consistent HTD Profiles

The coefficient of determination (R^2) of the model for each herd and trait was calculated as the squared correlation between the HTD effects and the smoothed HTD effects. A low R^2 means that the model based on sine curves did not fit the data well. Consequently, a low R^2 reflects a HTD profile that was not consistent from one year to another. In this study, herds for which at least one out of the 3 HTD (milk yield, protein and fat contents) profiles had a low R^2 were discarded. An arbitrary threshold of 0.3 for the R^2 led to 452 herds for which the HTD patterns were considered sufficiently consistent from one year to another from 2004 to 2010. These herds remain representative of the 3 breeds and the diversity of French regions (Table 1).

Describing Smoothed Herd-Test-Day Profiles

The 5 fixed coefficients (\hat{a}_h , \hat{b}_h , \hat{c}_h , \hat{d}_h , \widehat{herd}_h) of the smoothed HTD profiles were hard to interpret and far from intuitive, in particular due to the large overlap between the annual sine functions. Consequently, 7 new descriptors were built to visually describe the smoothed HTD profiles of year 2009: minimum, maximum, and the months where these extrema were observed, as well as the mean, the range, and the width at half the maximum (Figure 2). Values for year 2009 were chosen to facilitate the interpretation of links between HTD profile clusters and data on herd management also collected for year 2009 in the cluster analysis. Table 2 shows the summary statistics for quantitative descriptors for Montbéliarde, Normande, and Holstein herds, and Table 3 shows the distribution of each categorical descriptor. Initially, each categorical descriptor had 12

levels (one per month). Some levels were merged in the level “other” to obtain a minimum number of herds per level, which explains why each categorical descriptor has only 3 or 4 levels.

Factor Analysis

In the data set, the 452 herds were described by 21 variables: the 2009 smoothed HTD profile descriptors (Figure 2) for milk yield and protein and fat contents ($3 \times 7 = 21$ variables). These descriptors were centered within breed to work with the 3 breeds simultaneously. A multiple factorial analysis (MFA) determined the main factors explaining the variation in the smoothed HTD profiles (Escoffier and Pagès, 1994). Multiple factorial analysis is a factor analysis similar to principal components analysis, which facilitates the joint use of quantitative and categorical variables, such as months of maximum HTD effect. The number of axes in the MFA is equal to the number of quantitative variables plus the number of levels of each categorical variable. This analysis included 15 quantitative variables and 6 categorical variables (months of extrema) with 3 or 4 levels each; that is, 23 levels in total (Table 3). This leads to 38 factors in the analysis. Eight factors out of the 38 factors of the MFA explained 67% of the total variation in the 21 original smoothed HTD profile descriptors (Figure 3). The coordinates of the herds on these 8 factors were used in the cluster analysis instead of the initial descriptors to eliminate noise. Indeed, MFA can be viewed as a way to separate signal and noise in a data set: the first dimensions extract the most of the information, whereas the last ones are restricted to noise.

Cluster Analysis

This step consisted in creating several herd clusters based on the first 8 factors of the MFA, which summarize the descriptors of the smoothed HTD profiles. Several methods can be used to carry out a cluster analysis on a multivariate data set. For example, Ward (1963) proposed hierarchical clustering algorithms, which break up the data into a hierarchy of clusters. The most commonly used is agglomerative hierarchical clustering (AHC). Hartigan and Wong (1979) proposed the K-means algorithm, a partitioning algorithm that aims to partition the data set into K groups.

The drawback of hierarchical clustering is that herds that have been incorrectly grouped at an early stage of the algorithm cannot be reallocated to another cluster. It is particularly problematic when a large number of

herds have to be clustered. Such a drawback does not exist with partitioning algorithms. However, in contrast to hierarchical algorithms, the number of clusters has to be chosen a priori and the method is not deterministic (the choice of initial partition affects the final clusters). We chose to use a hybrid clustering method that combined the advantages of both clustering methods.

First, a K-means clustering of 452 herds into 50 clusters was obtained to reduce the dimension of the data set. Then, an AHC was performed on the centers of these 50 clusters and, according to the maximum inertia gain, 3 clusters of herds were retained. Finally, to re-allocate herds that could have been incorrectly grouped by the AHC, a K-means was carried out, choosing the result of the AHC as initial partition of the algorithm.

The HTD profiles of the 452 herds were assumed representative of the continuous nature of the environment. To obtain herd clusters with clearly typical features, the 40 most representative herds of each cluster, namely the 40 nearest to the cluster centers, were selected (Figure 4). Clusters based on these typical herds were described by their 2009 HTD profile descriptors for milk yield and protein and fat contents (active variables of the factor analysis) and by herd management descriptors (variables not included in the building of the 8 factors of the MFA).

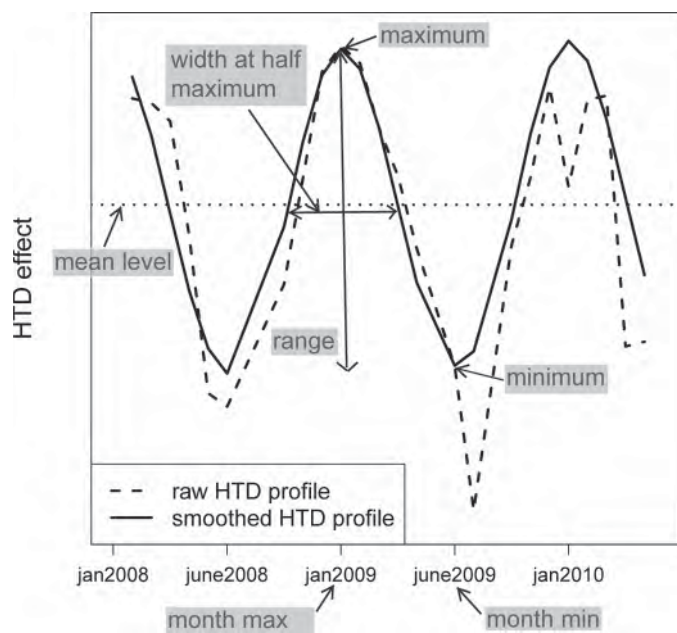


Figure 2. Herd-test-day (HTD) profile, smoothed HTD profile, and descriptors of the smoothed HTD profile used for the factor analysis.

Table 2. Summary statistics of the smoothed herd-test-day (HTD) profile quantitative descriptors for Montbéliarde, Normande, and Holstein herds taken into account in the factorial analysis

HTD profile	Minimum	Maximum	Mean	SD	CV
Montbéliarde (n = 42 herds)					
Milk yield (kg/d)					
Maximum	13.86	27.82	20.91	3.44	0.16
Minimum	9.58	25.06	17.04	3.6	0.21
Mean	12.25	26.07	18.89	3.51	0.19
Range	1.09	6.34	3.88	1.23	0.32
Width (mo)	4	9	7.1	1.14	0.16
Fat content (g/kg per day)					
Maximum	35.45	49.85	42.01	2.78	0.07
Minimum	31.22	42.4	36.5	2.49	0.07
Mean	33.4	46.27	39.38	2.52	0.06
Range	3.11	9.55	5.51	1.44	0.26
Width (mo)	5	8	6.86	0.95	0.14
Protein content (g/kg per day)					
Maximum	31.92	37.43	34.45	1.27	0.04
Minimum	29.4	34.58	31.52	1.33	0.04
Mean	30.49	36.02	33	1.23	0.04
Range	1.45	5.2	2.92	0.86	0.29
Width (mo)	4	10	7.38	1.23	0.17
Normande (n = 83 herds)					
Milk yield (kg/d)					
Maximum	11.69	23.55	19.31	2.19	0.11
Minimum	8.23	20.94	16.20	2.33	0.14
Mean	9.7	22.02	17.67	2.25	0.13
Range	1.64	4.92	3.11	0.8	0.26
Width (mo)	4	9	6.84	1.21	0.18
Fat content (g/kg per day)					
Maximum	38.38	53.28	46.44	2.47	0.05
Minimum	32.64	46.59	40.68	2.14	0.05
Mean	35.61	48.87	43.83	2.17	0.05
Range	1.96	10.08	5.76	1.61	0.28
Width (mo)	5	10	6.63	1.06	0.16
Protein content (g/kg per day)					
Maximum	31.28	38.71	36.06	1.14	0.03
Minimum	29.33	35.75	33.44	1.06	0.03
Mean	30.06	36.95	34.78	1.07	0.03
Range	1.17	5.34	2.62	0.76	0.29
Width (mo)	5	10	7.37	1.13	0.15
Holstein (n = 327 herds)					
Milk yield (kg/d)					
Maximum	16.32	36.8	26.32	3.53	0.13
Minimum	11.63	34.27	22.29	3.83	0.17
Mean	13.65	35.96	24.38	3.64	0.15
Range	0.86	9.65	4.03	1.28	0.32
Width (mo)	5	10	6.93	1.1	0.16
Fat content (g/kg per day)					
Maximum	30.53	49.05	42.69	2.16	0.05
Minimum	26.28	43.69	37.51	2.25	0.06
Mean	28.66	47.08	40.35	2.09	0.05
Range	1.35	9	5.18	1.34	0.26
Width (mo)	5	11	6.92	1.05	0.15
Protein content (g/kg per day)					
Maximum	29.62	37	33.72	1.06	0.03
Minimum	27.02	34.25	31.2	1.11	0.04
Mean	28.67	34.92	32.52	1.01	0.03
Range	0.81	5.32	2.52	0.77	0.31
Width (mo)	4	11	7.19	0.94	0.13

RESULTS

Consistency of HTD Profiles

The goodness of fit of the model describing HTD profiles as a sum of sine functions was better for pro-

tein and fat contents than for milk yield in both the complete data set with 24,569 herds (mean R^2 of 0.50, 0.51, and 0.44 for fat content, protein content, and milk yield, respectively) and that with 452 herds (mean R^2 of 0.54, 0.54, and 0.50 for fat content, protein content,

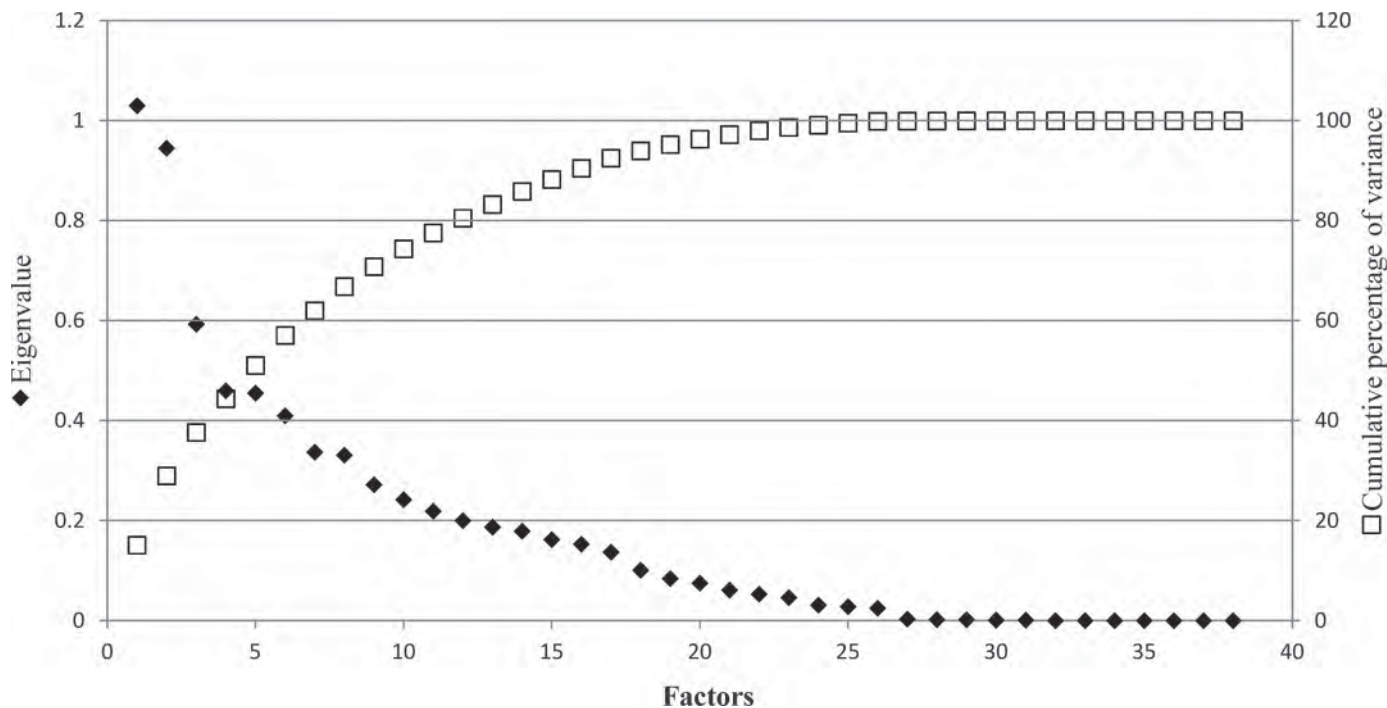


Figure 3. Eigenvalues corresponding to each factor (scale on the left) and the cumulative proportion of the variation that they comprise (scale on the right).

Table 3. Distribution of each categorical descriptor of the smoothed herd-test-day (HTD) profiles for the 452 herds taken into account in the factorial analysis

Categorical descriptor/level	Percentage of herds for each level			
	All herds	Montbéliarde	Normande	Holstein
Month of minimum milk yield HTD profile				
January	34	10	24	40
February	26	31	22	26
March	19	36	22	17
Other	21	23	32	17
Month of maximum milk yield HTD profile				
July	30	19	8	37
August	26	21	18	29
September	26	48	42	20
Other	18	12	32	14
Month of minimum fat content HTD profile				
September	12	14	2	15
October	58	40	55	61
November	22	36	28	19
Other	8	10	15	5
Month of maximum fat content HTD profile				
April	32	40	22	34
May	47	40	53	46
Other	21	20	25	20
Month of minimum protein content HTD profile				
August	22	17	11	25
September	48	29	43	52
October	19	36	30	14
Other	11	18	16	9
Month of minimum protein content HTD profile				
March	22	24	16	23
April	56	57	54	56
May	15	17	25	13
Other	7	2	5	8

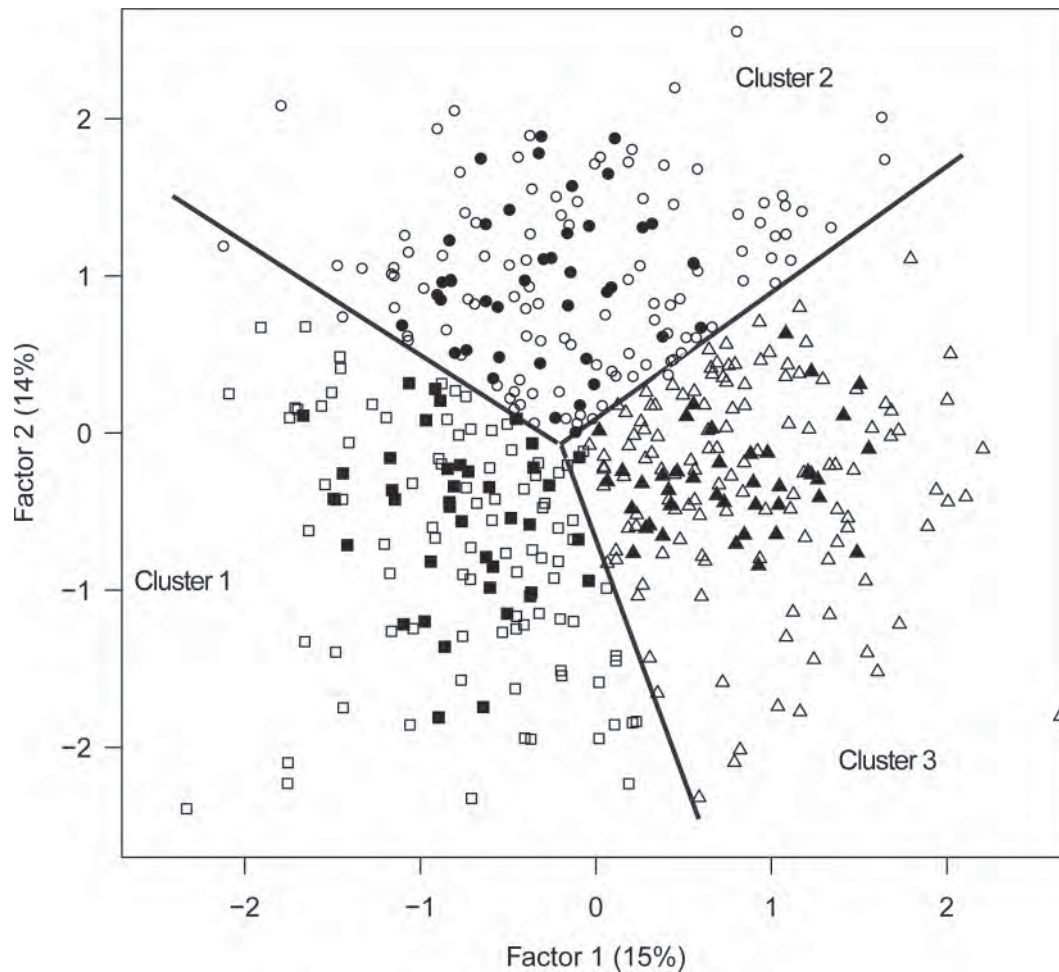


Figure 4. First factorial map of the multiple factor analysis representing the 452 herds in 3 clusters and the 40 most representative herds of each cluster (solid symbols).

and milk yield, respectively). The shape of the HTD profiles for fat and protein contents was more consistent from one year to another than the shape of milk yield HTD profiles, undoubtedly due to a lower effect of herd management changes on protein and fat contents than on milk yield.

Description of Smoothed HTD Profiles

Table 2 shows that the mean level of milk yield HTD profile was greater for the Holstein breed (24 kg/d) than for both Montbéliarde and Normande breeds (19 and 18 kg/d, respectively). For fat content, the mean level was greater for Normande (44 g/kg per day) than for the 2 other breeds (40 g/kg per day for Montbéliarde and Holstein breeds). The same hierarchy was observed for protein content (35 g/kg per day for Normande and 33 g/kg per day for Montbéliarde

and Holstein breeds). Range of HTD profiles and width at half the maximum for the 3 traits were of the same order of magnitude for the 3 breeds. These mean differences between breeds show the need to correct for a breed effect in the factorial and cluster analysis to focus on elements that are shared by the 3 breeds. Table 3 shows that whatever the traits, for more than 75% of the herds, the extrema of HTD profiles were located in only one season. For example, for 79% of the herds, the environment (feeding and climate conditions) was mainly unfavorable during winter for milk production: the minimum of milk yield HTD profiles were in January, February, or March. For fat and protein contents, the favorable season was spring and the unfavorable season was the end of summer and the beginning of autumn. However, some herds had other shapes of HTD profiles, with the favorable season for milk production being in spring, for example.

Table 4. Correlations between factors of the multiple factorial analysis and quantitative variables describing herd-test-day (HTD) profiles for milk yield, fat content, and protein content: maximum, minimum, mean level, range, and width at half maximum

HTD profile	Factor							
	1	2	3	4	5	6	7	8
Milk yield								
Maximum	-0.09	0.88	-0.06	0.21	0.27	0.19	0.16	-0.04
Minimum	-0.04	0.93	-0.01	0.26	0.18	0.14	0.03	0.07
Mean	-0.07	0.92	-0.05	0.23	0.23	0.17	0.1	0
Range	-0.12	-0.31	-0.13	-0.19	0.23	0.06	0.38	-0.34
Width	0.08	0.15	0.45	0.15	0.09	-0.45	-0.34	0.37
Fat content								
Maximum	0.84	-0.23	-0.02	0.3	-0.04	0.18	0.24	0.18
Minimum	0.76	-0.16	0.23	0.48	-0.16	0	0.13	-0.18
Mean	0.85	-0.2	0.06	0.39	-0.12	0.11	0.2	0.03
Range	0.13	-0.13	-0.41	-0.27	0.18	0.32	0.14	0.6
Width	-0.02	-0.09	0.63	-0.22	0.41	0.08	0.18	0.11
Protein content								
Maximum	0.78	0.23	-0.13	-0.46	0.21	-0.08	-0.05	-0.01
Minimum	0.65	0.49	0.06	-0.44	-0.22	-0.12	-0.09	-0.16
Mean	0.75	0.4	-0.04	-0.48	0	-0.14	-0.06	-0.07
Range	0.17	-0.4	-0.25	-0.01	0.62	0.08	0.05	0.25
Width	0.02	-0.08	0.16	0.02	0.13	0.59	-0.45	-0.08

MFA Results

The MFA determined the major factors of variation of the shapes of HTD profiles defined by the smoothed HTD profile descriptors. Figure 3 shows the eigenvalues corresponding to each factor. Each factor of the MFA can be interpreted through the variables with high absolute value coordinates and high correlation with the factor (Table 4). Factor 1 was related to the minimum, maximum, and mean of the 2009 smoothed HTD profiles for protein and fat contents, whereas factor 2 was related to the same variables for milk yield. In a factorial analysis, because each consecutive factor is defined to maximize the variability not captured by the preceding factors, consecutive factors (as factor 1 and factor 2) are independent from each other. This indicates that herds with a high mean level of milk HTD profile included herds with either low or high mean levels for HTD protein and fat contents. Factors 1 and 2 explained 29% of the total variation in the HTD profiles (Figure 3), indicating that mean levels in milk yield and protein and fat contents were a major source of variation in HTD profiles. Factors 3 to 8 were defined by variables such as range, full width at half maximum of the 2009 smoothed HTD profiles, and months of maximum and minimum smoothed HTD effects. For example, these factors discriminated herds with HTD protein and fat content profiles with a high range but narrow peak for milk HTD profile from herds with HTD protein and fat content profiles with a low range but wide peak for milk HTD profile. Factors 3 to 8 explained 38% of the total variation in the data,

which supports the conclusion that the pattern of the HTD profile (especially the HTD content profiles) is a source of diversity. Altogether, factors 1 to 8 explained 67% of the total variation in the HTD profiles (Figure 3) and were used in the cluster analysis.

Description of Herd Clusters by HTD Profiles and Herd Management Descriptors

Application of the cluster analysis to these 8 factor scores led to 3 clusters of, respectively, 138, 149, and 165 herds. The relatively homogeneous partition of the 3 breeds among the 3 clusters proved that clustering was not due to breed effects (Table 5). The 40 most representative herds of each cluster are indicated by solid symbols in Figure 4. Note that the factorial map does not allow perfect visualization of these herds because only 2 factors are in evidence, whereas herds were represented by 8 factors, which explains why they do not appear exactly on the center of their cluster.

In the following description, a high mean level of the HTD profile must be interpreted within breed, and a

Table 5. Number of herds per breed and cluster

Breed	Cluster			Total
	1	2	3	
Montbéliarde	17	11	14	42
Normande	24	22	37	83
Holstein	97	116	114	327
Total	138	149	165	452

high HTD profile means one with a high mean level. We found one cluster with low HTD profiles for the 3 traits (cluster 1), one cluster with a high milk yield HTD profile (cluster 2), and one cluster with high protein and fat content HTD profiles (cluster 3).

Table 6 shows the variables that characterized the 40 most representative herds of each cluster. Cluster 1 was made up of herds for which the mean levels of milk yield, fat content, and protein content were low. In this group of herds, the level of intensification was low: milk production per cow, density per hectare, quantity of concentrate per cow, and percentage of first-lactation cows were lower than the overall mean. Cluster 2 consisted of herds with a high mean level of milk yield and a low mean level of fat content. The range of the 3 HTD profiles and the width of the peak of protein and fat contents HTD profiles were lower than in the other clusters, showing relatively homogeneous management across seasons. This may be explained by a high level of intensification (contrary to cluster 1); milk production per cow, density per hectare, and quantity of concentrate per cow were higher than the overall mean, and first calvings occurred at an earlier age. Cluster 3 was made up of herds with a low mean level for their milk HTD profile and a high mean level for fat and protein content HTD profiles. The ranges of fat and protein contents HTD profiles were higher than in the other clusters. For this cluster, it was difficult to link HTD profiles to herd management: only the percentage of first-lactation cows was higher than the overall mean.

The variables potentially needed to describe this type of herd management might not have been included in the survey or in the variables built from the national database.

Some key data on herd management, such as the duration of fodder distribution or the length of grazing period, were thought to have been collected but the amount of missing data unfortunately prevented their use in the analysis. This could explain the difficulty in assessing the relationship between HTD profile patterns and herd management in each cluster.

Describing the diversity of the environment with only 3 clusters may seem too simplistic. Hence, the same methodology was used to create 6 clusters (results not shown). Cluster 1 did not change; cluster 2 (high milk yield HTD profile) was divided into 2 groups (named 2.1 and 2.2) according to the level and range of protein and fat contents; and cluster 3 (high contents HTD profiles) was divided into 2 groups (named 3.1 and 3.2) according to the range of profiles. Cluster 2.1 included herds with a low mean level and range for fat and protein contents HTD profiles. In contrast, cluster 2.2 contained herds with a high mean level for protein content and a low range for milk yield and protein content. Cluster 3.1 was composed of herds with a small range of fat content HTD profiles, whereas cluster 3.2 consisted of herds with a high range for fat and protein contents. None of the variables describing HTD profiles or herd management characterized the sixth cluster. The interpretation in terms of herd management of

Table 6. Means (SD) of characteristic herd descriptors for each herd cluster

Descriptor	Cluster 1	Cluster 2	Cluster 3	Overall ¹
Herd-test-day profiles				
Milk yield mean level ² (kg/d)	-1.44*** (1.84)	3.47 *** (1.90)	-0.74 (1.25)	0.43 (2.75)
Fat content mean level ² (g/kg per day)	-1.23*** (1.00)	-0.82** (1.00)	1.53*** (0.99)	-0.17 (1.58)
Protein content mean level ² (g/kg per day)	-0.75*** (0.50)	0.18* (0.56)	0.48*** (0.47)	-0.05 (0.73)
Milk yield range (kg/d)	3.89 (1.09)	3.4* (0.8)	3.81 (0.95)	3.7 (1.0)
Fat content range (g/kg per day)	2.62 (0.47)	4.94 (0.84)	5.7* (1.1)	5.3 (1.1)
Protein content range (g/kg per day)	5.26 (1.23)	2.1*** (0.5)	2.8*** (0.5)	2.5 (0.60)
Herd management				
Milk production per cow (kg)	7,428*** (982)	8,260** (1,091)	7,840 (759)	7,865 (1,019)
Fat yield quota (g/L)	42 (1.9)	42 (1.8)	40* (7.0)	42 (4.4)
Total forage area (ha)	78 (38)	61* (22)	78 (53)	72 (41)
Grass area (ha)	52 (40)	36* (17)	49 (42)	45 (35)
Total conserved forages (10 ³ kg of DM/LU ³)	3.43** (0.92)	3.97 (0.85)	3.86 (1.07)	3.8 (0.97)
Hay (10 ³ kg of DM/LU)	0.8** (1.1)	0.35 (0.40)	0.41 (0.28)	0.5 (0.71)
Maize silage (10 ³ kg of DM/LU)	2.31*** (1.23)	3.15 (0.89)	3.12 (1.00)	2.9 (1.11)
Concentrate (kg/cow)	1,232** (387)	1,546** (490)	1,344 (444)	1,378 (461)
Stocking rate (LU/ha of forages)	1.43* (0.51)	1.75* (0.40)	1.53 (0.38)	3.7 (0.44)
Percentage of first lactation	27** (15)	34 (10.5)	37* (16)	17 (15)
Age at first calving (mo)	31 (3.4)	29* (2.8)	31 (4.9)	30 (0.26)

¹Overall mean of the 120 (3 × 40) most representative herds used for the description of clusters.

²Average deviation from the breed mean.

³Livestock unit.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$: significance of the Student test between the overall mean and the mean of the cluster.

these 6 clusters was difficult and questionable, meaning that the variables used to describe the HTD profiles or herd management were not accurate enough to properly characterize such a number of clusters.

DISCUSSION

This study aimed to find a new description of the environment based on milk yield and fat and protein content HTD profiles. The strategy consisted of using HTD profiles assuming that they were appropriate variables representative of the herd management (essentially feeding) even though they also included the effect of climatic conditions. This description, based on information available at the national level, will be used to more precisely assess the extent of G×E interactions in French dairy herds during the next phase of the project.

Smoothing the HTD profiles using sine functions gave results similar to those of Koivula et al. (2007). All types of combinations of HTD profiles for milk yield and fat and protein contents were found. The diversity of combinations of HTD profiles is undoubtedly related to the diversity of herd management and climate in France. A clear continuum exists in both herd management and HTD profile variability. The major factors of variation in the HTD profiles were the mean levels, whereas variations in profile shapes were less substantial. Three clusters of herds were obtained from the description of their 3 HTD profiles: the first with a low mean level for the 3 traits, the second with a high mean level for milk yield, and the third with a high mean level for fat and protein contents. The first 2 were clearly related to different levels of intensification. Summarizing the variability of herd management in 3 categories may be too crude, and interpretation of the 3 clusters in terms of herd management is quite simplistic. However, surprisingly, a description of the environment through 6 clusters instead of 3 did not give a clearer picture of the existing variability. This may be due to a description of herd management or HTD profile descriptors that is not accurate enough. More precise descriptors of herd management, such as diet fed in each season, would improve the precision of interpretation of these HTD profiles. Indeed, HTD profiles are based on monthly records, whereas the herd management descriptors available in the survey were annual. Moreover, cluster analysis results could be improved by further enrichment of the parameters used to describe the shape of HTD. For example, the method did not detect, although they exist, herds with seasonal variations; for example, herds in which herd management makes it possible to produce more in winter than in summer (or vice versa).

The correct definition of the environment is a crucial point in G×E interaction studies. Four major types of definition exist in the literature. In small-scale studies (with a limited number of herds or animals), the environment can be controlled accurately. For example, it can correspond to different levels of concentrate used (Fulkerson et al., 2008) or to the use of pasture versus confinement (Boettcher et al., 2003; Kearney et al., 2004). However, such criteria do not make it possible to study G×E interactions at a national level. The second definition is used at an international level: in the MACE procedure (Schaeffer, 1994) carried out by Interbull, each country is considered as a distinct environment. This is also the case in the studies of Cienfuegos-Rivas et al. (1999), Weigel et al. (2001), Ojango and Pollott (2002), and Hammami et al. (2008). This approach is a rather extreme approximation because it is known that several environments co-exist in most countries, whereas some herds in different countries could be considered to be in the same environment. For this reason, other researchers have worked on a trans-national definition of environments based on herd management, climate, and genetic merit descriptors (Weigel and Rekaya, 2000; Zwald et al., 2003). Their descriptors of herd management are quite different from those used in this paper. They used, among others, peak yield, days to peak yield, persistency, standard deviation of milk yield, and fat-to-protein ratio. These factors are more a consequence of herd management and genetic level than descriptive elements of the herd management itself. However, as in our study, they defined the environment as a group of herds with similar characteristics. The advantage of such a strategy is that environments are supposedly easier to interpret. In contrast, clusters do not model the existing continuum of the environment. The last way of defining the environment considers this continuum, characterizing it through one or several “environmental parameters.” They can be quantitative variables (Fikse et al., 2003; Calus et al., 2005), such as peak milk yield, or variables that combine information on different descriptors (Haskell et al., 2007). Such a definition could be the most realistic one but involves the use of reaction norm models, which are more complex than character state models (in which environments are defined as clusters). One major choice to make is the type of function (e.g., linear or quadratic) to use in modeling the link between the environmental parameter and the genetic merit.

The originality of this study was the use of HTD profiles to define the environment. No similar study has previously been carried out. Our approach is attractive in that it is potentially applicable to a large population without necessitating the collection of a lot of extra information on herd management. Indeed, HTD profiles

are available for all herds participating in the French milk recording system. It thus gives the possibility of accounting for variation in herd management across seasons, in contrast to the definition of Zwald et al. (2003) or Weigel and Rekaya (2000).

Different choices made in this study led to a great simplification of the existing diversity of HTD profiles shapes. These include the arbitrary threshold over which HTD profiles were considered to be consistent from one year to another, the way of summarizing the shape of smoothed HTD profiles by 7 descriptors, the selection of only the first 8 factors of the factorial analysis in the cluster analysis, and the interpretation of 3 clusters on the basis of only 40 herds each. All these elements could be modified to obtain a more precise description of HTD profiles. However, even though the 3 herd clusters obtained can be considered as a simplistic view of the reality, they provide a useful way to study G×E interactions among contrasting environments.

In the next phase of the project, 2 strategies will be implemented to more precisely assess the extent of existing G×E interactions in French dairy herds, not only on production traits but also on important functional traits (e.g., fertility, udder health, longevity). The first one will allocate a large number of herds to a cluster through a discriminant analysis and then apply trait analysis or a character state model using only a selection of records from representative herds of the 3 distinct clusters. The second will use a reaction norm model, taking the first axis of the factor analysis on the descriptors of HTD profiles as environmental parameters.

CONCLUSIONS

In this study, we proposed a new definition of the environment based on HTD profiles (of milk yield and fat and protein contents), estimated from a test-day model, and used it for the creation of herd clusters. These clusters could be interpreted in terms of the level of intensification of herd management. Further analysis using more specific herd management descriptors such as seasonal feed rations could improve the interpretation of HTD profiles. It is hoped that a more meaningful assessment of G×E will result from studies based on a careful use of by-products of national genetic evaluations such as HTD profiles.

ACKNOWLEDGMENTS

This study was financially supported by the French Ministry of Agriculture (Programme CASDAR GENE-SYS). France Conseil Elevage, the 3 partners Bretagne

Conseil Elevage, Jura Conseil Elevage, Orne Conseil Elevage, and les Réseaux d'Elevage (Database DIA-PASON, Institut de l'Elevage) are acknowledged for providing the data and for their suggestions. The 3 anonymous reviewers are acknowledged for their constructive evaluation of the manuscript.

REFERENCES

- Boettcher, P. J., J. Fatehi, and M. M. Schutz. 2003. Genotype × environment interactions in conventional versus pasture-based dairies in Canada. *J. Dairy Sci.* 86:383–389.
- Calus, M. P. L., J. J. Windig, and R. F. Veerkamp. 2005. Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. *J. Dairy Sci.* 88:2178–2189.
- Carabaño, M. J., K. M. Wade, and L. D. Van Vleck. 1990. Genotype by environment interactions for milk and fat production across regions of the United States. *J. Dairy Sci.* 73:173–180.
- Cienfuegos-Rivas, E. G., P. A. Oltenacu, R. W. Blake, S. J. Schwager, H. Castillo-Juarez, and F. J. Ruiz. 1999. Interaction between milk yield of Holstein cows in Mexico and the United States. *J. Dairy Sci.* 82:2218–2223.
- Escoffier, B., and J. Pagès. 1994. Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.* 18:121–140.
- Fikse, W. F., R. Rekaya, and K. A. Weigel. 2003. Assessment of environmental descriptors for studying genotype by environment interaction. *Livest. Prod. Sci.* 82:223–231.
- Fulkerson, W. J., T. M. Davison, S. C. Garcia, G. Hough, M. E. Goddard, R. Dobos, and M. Blockey. 2008. Holstein-Friesian dairy cows under a predominantly grazing system: Interaction between genotype and environment. *J. Dairy Sci.* 91:826–839.
- Hamami, H., B. Rekik, H. Soyeurt, C. Bastin, J. Stoll, and N. Gengler. 2008. Genotype × environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations. *J. Dairy Sci.* 91:3661–3671.
- Hartigan, J. A., and M. A. Wong. 1979. Algorithm AS 136: A K-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* 28:100–108.
- Haskell, M. J., S. Brotherstone, A. B. Lawrence, and I. M. S. White. 2007. Characterization of the dairy farm environment in Great Britain and the effect of the farm environment on cow life span. *J. Dairy Sci.* 90:5316–5323.
- Interbull. 2011. Description of National Genetic Evaluation Systems for dairy cattle traits as applied in different Interbull member countries. Accessed Dec. 13, 2011. http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm.
- Kearney, J. F., M. M. Schutz, P. J. Boettcher, and K. A. Weigel. 2004. Genotype × environment interaction for grazing versus confinement. I. Production traits. *J. Dairy Sci.* 87:501–509.
- Koivula, M., J. I. Nousiainen, J. Nousiainen, and E. A. Mäntysaari. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *J. Dairy Sci.* 90:2563–2568.
- Leclerc, H., D. Duclos, A. Barbat, T. Druet, and V. Ducrocq. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal* 2:344–353.
- Leclerc, H., I. Nagy, and V. Ducrocq. 2009. Impact of using reduced rank random regression test-day model on genetic evaluation. Interbull Meeting, Barcelona, Spain. *Interbull Bull.* 40:42–46.
- Mayeres, P., J. Stoll, J. Bormann, R. Reents, and N. Gengler. 2004. Prediction of daily milk, fat and protein production by a random regression test-day model. *J. Dairy Sci.* 87:1925–1933.
- Meyer, K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by REML. *J. Zhejiang Univ. Sci. B* 8:815–821.
- Ojango, J. M. K., and G. E. Pollott. 2002. The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. *Livest. Prod. Sci.* 74:1–12.

- Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.
- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58:236–244.
- Weigel, K. A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J. Dairy Sci.* 83:815–821.
- Weigel, K. A., R. Rekaya, N. R. Zwald, and W. F. Fikse. 2001. International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. *J. Dairy Sci.* 84:2789–2795.
- Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. *J. Dairy Sci.* 86:376–382.

APPENDIX

Table A1. Summary statistics of descriptors of herd management for the 934 herds in the study

Item	Minimum	Maximum	Mean	SD	CV	Missing data (%)
Areas						
Total area (ha)	10.5	470.4	106.7	61.85	0.58	4.82
Total area per cow (ha)	0.66	113.5	10.17	14.37	1.41	5.78
Total forage area (ha)	6.1	348.8	72.52	40.09	0.55	4.71
Total forage area per cow (ha)	0.2	112	7.2	10.87	1.51	5.67
Total forage area dedicated to dairy cattle (ha)	0	250.5	67	34.43	0.51	28.37
Grass area (ha)	2.27	300	53.4	41.33	0.77	4.82
Maize area (ha)	0	103.5	18.7	15.89	0.85	4.07
Maize area per cow (ha)	0	33.6	1.78	2.99	1.68	5.03
Cereal, oilseed and protein seed area (ha)	0	253	35.28	40.56	1.15	9.96
Maize (ha)/total forage (ha) area ratio	0	0.89	0.31	0.23	0.74	4.71
Production						
Milk production (10 ³ kg/herd per year)	88	4,150	410.4	240.2	0.59	29.23
Milk production per cow (kg)	3,751	97,890	7,637	3,421	0.45	13.28
Fat yield quota (g/L)	0	50.1	40.92	3.57	0.09	15.95
Manpower	0.33	20	2.24	1.13	0.5	3.32
Livestock						
Livestock unit (LU)	19.2	444.8	97.74	48.64	0.5	2.89
Dairy cattle LU	15.6	313.3	83.79	36.24	0.43	2.36
Dairy cows LU	13.4	243.7	56.35	24.16	0.43	2.36
Heifer LU	0	110.93	26.54	14.73	0.56	0.32
Other LU	0	282	13.37	25.42	1.9	2.46
Number of dairy cows	13.1	211.9	55.04	23.02	0.42	2.89
Number of 2-yr-old heifers	0	56	12.63	8.83	0.7	40.47
Number of heifers between 1 and 2 yr old	0	66	19.92	10.2	0.51	40.47
Number of 1-yr-old heifers	0	84	20.75	11.45	0.55	40.47
Stocking rate (LU/ha of forages)	0.18	7.11	1.47	0.55	0.37	4.82
Percentage of the main breed in the herd	50	100	98.6	6.93	0.07	20.88
Feeding						
Total conserved forages (10 ³ kg of DM/LU)	0.41	22.93	3.68	1.33	0.36	16.7
Maize silage (10 ³ kg of DM/LU)	0	5.95	2.46	1.36	0.55	25.05
Grass silage (10 ³ kg of DM/LU)	0	2.2	0.33	0.46	1.39	34.15
Hay (10 ³ kg of DM/LU)	0	6.57	0.92	0.99	1.08	18.52
Concentrate (kg/cow)	307	6,563	1,468	600	0.41	6.96
Concentrate (g/L)	29	798	202.62	76	0.38	7.28
Duration of grazing (d)	0	348	212	40	0.19	72.91
Duration of forage retaining (d)	0	245	47	55	1.17	71.95
Lactation distribution (no.)						
First	0	115	10.32	11.04	1.07	0.96
Second	0	82	8.14	8.47	1.04	0.96
Third	0	37	5.79	5.8	1	0.96
Fourth and greater	0	50	8.07	7.79	0.97	0
Lactation distribution (%)						
First	0	100	28.99	16.08	0.55	0.96
Second	0	100	23.97	14.27	0.6	0.96
Third	0	100	18.62	13.18	0.71	0.96
Fourth and greater	0	100	28.42	17.63	0.62	0.96
Lactation length (d)						
Primiparous	41	796	344	48	0.14	0.32
Multiparous	198	539	332	40	0.12	0.11
Calving interval (d)						
Primiparous	336	581	410	34	0.08	0.32
Multiparous	343	595	412	30	0.07	0.11

Continued

Table A1 (Continued). Summary statistics of descriptors of herd management for the 934 herds in the study

Item	Minimum	Maximum	Mean	SD	CV	Missing data (%)
Percentage of calvings by month						
January	0	87.5	8.27	7.2	0.87	0
February	0	66.67	5.84	5.77	0.99	0
March	0	66.67	6.25	6.12	0.98	0
April	0	42.86	4.63	4.85	1.05	0
May	0	33.33	5.03	4.8	0.95	0
June	0	39.34	6.05	5.49	0.91	0
July	0	51.28	7.35	6.14	0.84	0
August	0	56.1	11.13	7.69	0.69	0
September	0	55.17	13.96	9.11	0.65	0
October	0	50	11.75	7.01	0.6	0
November	0	35.29	10.64	5.87	0.55	0
December	0	28.85	8.76	5.43	0.62	0
Number of calvings by month						
January	0	26	4.11	3.19	0.78	0
February	0	26	2.88	2.73	0.95	0
March	0	25	3.07	2.92	0.95	0
April	0	20	2.44	2.6	1.07	0
May	0	19	2.64	2.82	1.07	0
June	0	29	3.18	3.2	1.01	0
July	0	34	3.92	3.83	0.98	0
August	0	38	6.04	5.02	0.83	0
September	0	35	7.34	5.47	0.75	0
October	0	35	6.19	4.39	0.71	0
November	0	23	5.64	3.8	0.67	0
December	0	25	4.67	3.46	0.74	0
Other						
Age at first calving (mo)	1.96	3.86	2.58	0.3	0.12	0.43
Mean lactation number	1.7	7.5	2.6	0.48	0.18	29.01
Number of milking machines	1	36	9.68	3.68	0.38	20.66
Number of cows by milking machines	1.64	20	5.92	2.0	0.34	21.73
Replacement rate (%)	4	79	32.62	10.21	0.31	18.52