

## Potentiel d'utilisation de la spectrométrie moyen infrarouge pour prédire le rendement fromager du lait et étudier sa variabilité génétique

COLINET F. G. (1), TROCH T. (2), ABBAS O. (3), BAETEN V. (3), DEHARENG F. (4), FROIDMONT E. (5), SOYEURT H. (6), DARDENNE P. (3,4), SINDIC M. (2), GENGLER N. (1)

(1) Université de Liège, Gembloux Agro-Bio Tech, Unité de Zootechnie, 2 Passage des Déportés, 5030 Gembloux, Belgique

(2) Université de Liège, Gembloux Agro-Bio Tech, Unité Analyse, Qualité et Risques, Laboratoire Qualité et Sécurité des Produits Agroalimentaires, 2 Passage des Déportés, 5030 Gembloux, Belgique

(3) Centre Wallon de Recherches Agronomiques, Département Valorisation des productions, Unité Qualité des produits, 24 Chée de Namur, 5030 Gembloux, Belgique

(4) Centre Wallon de Recherches Agronomiques, Département Valorisation des productions, Unité Technologie de la transformation des produits, 24 Chée de Namur, 5030 Gembloux, Belgique

(5) Centre Wallon de Recherches Agronomiques, Département Productions et filières, Unité Nutrition animale et durabilité, 8 rue de Liroux, 5030 Gembloux, Belgique

(6) Université de Liège, Gembloux Agro-Bio Tech, Département des Sciences Agronomiques, 2 Passage des Déportés, 5030 Gembloux, Belgique

**RESUME** Fournir une indication rapide, fiable et bon marché du rendement fromager pour un lait donné, sans devoir appliquer des formules (empiriques ou théoriques) à partir des concentrations préalablement déterminées pour différents constituants du lait, serait un outil utile et économiquement intéressant tant pour les éleveurs que pour l'industrie laitière. En vue d'étudier la variabilité génétique du rendement fromager à l'échelle du cheptel bovin wallon, des méthodes chimiométriques ont été utilisées afin de développer des équations de prédictions basées sur des spectres moyen infrarouge (MIR) pour les rendements fromagers déterminés en laboratoire et exprimés en frais (RdFF) ou en sec (RdFS). Ceux-ci ont été déterminés sur 258 échantillons de lait analysés en spectrométrie MIR. Les équations de prédiction à partir du spectre MIR du lait ont été développées en utilisant la régression des moindres carrés partiels (PLS) avec une validation croisée interne appliquée sur la dérivée première des spectres MIR. Les coefficients de détermination de validation croisée ( $R^2_{cv}$ ) des équations étaient de 0,81 pour les prédictions du RdFF et de 0,82 pour les celles du RdFS. Les rapports des performances sur les variabilités (RPD) étaient égaux à 2,3. Ces résultats peuvent permettre d'envisager une bonne utilité pratique pour leur prédiction respective, notamment dans le cadre de recherches génétiques. Ces équations ont été appliquées sur la base de données spectrales générée dans le cadre du contrôle laitier wallon. Les composantes de la variance ont été estimées séparément pour le RdFF et le RdFS basées sur un modèle animal « contrôles élémentaires » utilisant des régressions aléatoires. Le jeu de données utilisé comportait 51 537 prédictions pour 7 870 vaches primipares Holstein. Les héritabilités journalières moyennes variaient entre 0,31 (au 5<sup>ème</sup> jour de lactation (JDL)) et 0,59 (au 279<sup>ème</sup> JDL) pour le RdFF et entre 0,31 (au 5<sup>ème</sup> JDL) et 0,57 (au 299<sup>ème</sup> JDL) pour le RdFS. Ces héritabilités journalières modérées à élevées ont indiqué le potentiel de sélection génétique pour ces deux caractères.

## Potential use of mid-infrared spectrometry to predict cheese yield from milk and to study its genetic variability

COLINET F. G. (1), TROCH T. (2), ABBAS O. (3), BAETEN V. (3), DEHARENG F. (4), FROIDMONT E. (5), SOYEURT H. (6), DARDENNE P. (3,4), SINDIC M. (2), GENGLER N. (1)

(1) University of Liège, Gembloux Agro-Bio Tech, Animal Science Unit, 2 Passage des Déportés, 5030 Gembloux, Belgium

**SUMMARY** Providing a quick, reliable and cheap indication of the expected cheese yield for a milk sample by avoiding (empirical or theoretical) formulas based on previously determined milk constituents would be an economically valuable tool useful for farmers and the dairy industry. In order to study the genetic variability of cheese yield on a large scale, mid-infrared (MIR) chemometric methods were used to predict fresh or dry Individual Laboratory Cheese Yield (RdFF and RdFS, respectively). RdFF and RdFS were determined on a total of 258 milk samples also analyzed with an MIR spectrometer. Equations to predict RdFF and RdFS from milk MIR spectra were developed using partial least squares regression (PLS) after first derivative pre-treatment applied to the spectra. The cross-validation coefficients of determination ( $R^2_{cv}$ ) of the two equations were equal to 0.81 for the prediction of RdFF and 0.82 for the prediction RdFS. The ratios of performance to deviation (RPD) of the two equations were both equal to 2.3. Therefore, these results suggest a practical utility of these two equations, *i.e.* for genetic research. Both equations were applied on the spectral database generated during the Walloon routine milk recording. The variance components were estimated using a univariate random regression animal test-day model. The dataset included 51 537 predicted records from 7 870 Holstein first-parity cows. Estimated daily heritabilities ranged from 0.31 (at 5<sup>th</sup> day in milk (DIM)) to 0.59 (at 279<sup>th</sup> DIM) for RdFF and from 0.31 (at 5<sup>th</sup> DIM) to 0.57 (at 299<sup>th</sup> DIM) for RdFS. Those moderate to high daily heritabilities indicate a selection potential for both traits.

## INTRODUCTION

La production de fromages a augmenté de 15 % au niveau de l'Union Européenne et de 22 % aux USA sur la période 2000-2009 (International Dairy Federation, 2010). La production de produits fromagers est caractérisée tant au niveau de sa quantité (rendement de transformation) qu'au niveau de la qualité des produits transformés. La qualité du

lait utilisé (et notamment sa composition) a un impact sur la qualité et la quantité du fromage obtenu.

Actuellement, il existe de nombreuses formules théoriques ou empiriques (basées notamment sur des constituants du lait tels que les contenus en protéines, en caséines et en matières grasses) qui permettent d'estimer le rendement fromager attendu. Il serait économiquement intéressant, tant pour l'éleveur que pour l'industrie laitière, d'avoir un outil

permettant d'indiquer rapidement le potentiel de rendement fromager d'un lait sans pour autant devoir déterminer le contenu en différents constituants du lait.

La spectrométrie moyen infrarouge (MIR) étant déjà implantée au niveau des laboratoires d'analyse du lait, le premier objectif était d'étudier la possibilité de développer des équations de prédiction MIR des rendements fromagers déterminés en laboratoire et exprimés en frais ou en sec (respectivement RdFF et RdFS). Le second objectif était une étude de la variabilité génétique de ces rendements prédits sur base du spectre MIR d'échantillons individuels de lait de vache Holstein en première lactation au sein du cheptel wallon.

## 1. MATERIEL ET METHODES

### 1.1. ÉCHANTILLONNAGE

Des échantillons de lait ( $n = 258$ ) ont été collectés dans différentes exploitations en Wallonie. Afin de couvrir une large variabilité, tant au niveau spectral qu'au niveau des mesures à réaliser, plusieurs critères ont été appliqués pour la collecte d'échantillons. Le premier critère consistait à prélever préférentiellement du lait de vache individuel afin de profiter de la variabilité individuelle et seulement quelques échantillons de tank. Le deuxième était de profiter de la variabilité entre races en collectant des échantillons de vaches Holstein, Blanc-Bleu Belge de type mixte, Montbéliarde et Brune. Le dernier consistait à collecter des échantillons de traite du matin ou du soir séparément, étant donné que la composition du lait diffère légèrement entre celles-ci.

### 1.2. ANALYSES

Après homogénéisation des échantillons de lait, une analyse en spectrométrie MIR a été réalisée par le Comité du Lait (Battice, Belgique) au moyen d'un spectromètre MIR MilkoScan FT6000 (Foss, Hillerød, Danemark). Les rendements fromagers de laboratoire (i.e., RdFF et RdFS) ont été déterminés par la quantité de caillé obtenu après ajout de présure (adapté de Hurtaud *et al.*, 1995). De la présure (Berthelot 53, Laboratoires ABIA, Meursault, France) est ajoutée à 30 ml de lait préalablement mis à température de 35°C. Après 1 heure d'incubation à 35°C, le lait emprésuré est centrifugé 15 minutes à 2 700 g. Les RdFF et RdFS sont exprimés en grammes de caillé par 100 g de lait et en grammes de matière sèche de caillé par 100 g de matières sèche de lait, respectivement.

Lors de l'établissement de la méthode analytique préalablement à cette étude, des tests de répétabilité ont été effectués sur 12 échantillons de lait individuel analysés en triple. Les erreurs-types des méthodes analytiques sont de 1,61 g caillé/100 g lait pour le RdFF et de 0,89 g MS caillé /100 g MS lait pour le RdFS.

### 1.3. DEVELOPPEMENT D'ÉQUATIONS DE PREDICTION

Préalablement, deux échantillons présentant une large distance spectrale (basée sur une distance de Mahalanobis standardisée (GH) calculée pour les spectres MIR et supérieure à 3) par rapport à la moyenne des 258 échantillons ont été retirés du jeu de données de calibration. En utilisant la régression des moindres carrés partiels (PLS) avec une validation croisée interne appliquée sur la dérivée première du spectre MIR, des équations de prédiction ont été construites pour le RdFF et le RdFS au moyen du logiciel WINISI III (<http://www.winisi.com/>; Foss, Hillerød, Danemark). Au cours du processus de calibration, des outliers ont été détectés et retirés des jeux de données de calibration. Il s'agit d'échantillons définis sur base d'un test  $t$  entre la valeur observée et la valeur prédite ; la valeur «  $T$  » critique utilisée est de 2,5 (valeur par défaut proposée par le logiciel). Ce test a été réalisé 4 fois. Une validation interne croisée a été réalisée en prenant 50 groupes en vue d'évaluer la robustesse des équations.

Différents paramètres sont à prendre en compte lors de la construction des équations de prédictions, à savoir : la moyenne et l'écart-type des sets de calibrations finaux, les coefficients de détermination de calibration ( $R^2_c$ ) et de validation croisée ( $R^2_{cv}$ ), les erreurs-types de calibration ( $SE_c$ ) et de validation croisée ( $SE_{cv}$ ), le rapport des performances sur les variabilités (RPD = rapport de l'écart-type du set de calibration finale sur l'erreur-type de validation croisée) et l'étendue du rapport d'erreur (RER = le rapport de l'étendue du jeu de calibration sur l'erreur-type de validation croisée).

### 1.4. PREDICTION

Les deux équations de prédiction développées ont été appliquées sur la base spectrale wallonne qui comportait les spectres MIR et données collectés depuis 2007 dans le cadre du Contrôle laitier wallon (> 2 500 000 spectres). En vue d'éliminer les outliers de prédiction RdFF et RdFS, des GH pour RdFF et RdFS ont été calculées pour chaque spectre en prenant comme référence les spectres des jeux de données finaux de calibration, respectivement du RdFF et du RdFS. Lorsque le GH du RdFF ou du RdFS était supérieur à 3, la prédiction du RdFF ou du RdFS, respectivement, était éliminée. Ce filtre permet de s'assurer que le spectre MIR de l'échantillon à prédire est bien compris dans la variabilité spectrale prise en compte dans l'équation de prédiction utilisée. Les prédictions RdFF et RdFS inférieures au percentile 0,5 ou supérieures au percentile 99,5 de leur distribution respective ont également été éliminées. Seules les prédictions relatives à des vaches Holstein au cours des trois premières lactations et associées à des contrôles élémentaires compris entre 5 et 365 jours de lactation (JDL) et dont la production laitière journalière se situait entre 3 et 99.9 kg de lait (selon les recommandations d'ICAR) ont été conservées.

### 1.5. ANALYSES GENETIQUES

Pour ces analyses, les jeux de données incluaient uniquement des vaches primipares Holstein de parents connus. Chaque vache devait être associée à au moins 4 prédictions au cours de sa première lactation. Chaque troupeau devait contenir au moins 10 primipares sélectionnées et pour chaque troupeau sélectionné, le nombre de prédictions par jour de contrôle devait être en moyenne supérieur à 6 et de minimum 3.

Le jeu de données final utilisé pour l'étude des composantes de la variance contenait 51 537 prédictions RdFF et RdFS pour 7 870 vaches primipares Holstein dans 101 troupeaux. Le pedigree extrait contenait 58 253 individus et fut limité aux animaux nés après 1985.

Les composantes de la variance du RdFF et du RdFS ont été estimées séparément en utilisant un modèle animal « contrôles élémentaires » avec des régressions aléatoires. Le modèle suivant fut appliqué :

$$y = X \beta + Q (Z p + Z a) + e,$$

où  $y$  est le vecteur contenant les prédictions (RdFF ou RdFS) ;  $\beta$  est le vecteur des 4 effets fixes (troupeau x date de test, stade de lactation (classes de 5 JDL), stade de gestation et stade de lactation x âge au vêlage x saison de vêlage) ;  $p$  et  $a$  sont les vecteurs des coefficients de régression aléatoire pour les effets aléatoires d'environnement permanent et génétique additif ;  $X$  et  $Z$  sont les matrices d'incidence reliant les observations aux effets ;  $Q$  est la matrice de covariables pour les polynômes de Legendre de second ordre ; et  $e$  le vecteur des résidus. Pour ce premier modèle, la variance résiduelle fut considérée comme constante tout au long de la lactation.

Les composantes de la variance ont été estimées séparément par AI-REML (*Averaged Information Restricted Maximum Likelihood*, Misztal, 2012). Les héritabilités journalières sont calculées pour chaque JDL comme le rapport de la variance génétique sur la somme des variances génétique, environnementale et résiduelle associée à ce JDL. Les variances génétique et environnementale à ce JDL sont

estimées par la formule  $qKq'$  où  $K$  est la matrice de (co)variances entre les coefficients du polynôme de Legendre de l'effet correspondant et  $q$  est le vecteur des coefficients du polynôme de Legendre calculés pour le JDL considéré. Les erreurs-types des variances et des héritabilités journalières sont estimées à partir des variances de l'inverse de la matrice des informations moyennes (Fischer *et al.*, 2004).

## 2. RESULTATS ET DISCUSSIONS

### 2.1. EQUATIONS DE PREDICTION

Une variabilité a été captée dans les 258 analyses de référence. En effet, la moyenne ( $\pm$  écart-type) obtenue pour le RdFF était de 28,9 ( $\pm$  8,8) g caillé / 100 g lait. Le coefficient de variation (CV) était de 30 %. De même pour le RdFS, la moyenne ( $\pm$  écart-type) obtenue était 62,3 ( $\pm$  7,4) g MS caillé / 100 g MS lait). Le CV du RdFS était de 12 %.

Au cours de la construction des équations de prédiction, des outliers de type « T » ont été détectés et retirés des jeux de données de calibration, à savoir 22 pour le RdFF et 13 pour le RdFS. Les paramètres statistiques de ces jeux de données finaux après retrait de ces outliers sont repris au Tableau 1.

Les paramètres obtenus (Tableau 1) pour chacune des deux équations sont prometteurs. D'une part, selon Williams (2003), un  $R^2_{cv}$  compris entre 0,50 et 0,65 indique une faible prédiction avec des modèles permettant uniquement la discrimination entre des hautes et basses valeurs. Un  $R^2_{cv}$  entre 0,66 et 0,81 indique que les prédictions sont assez précises tandis qu'une équation de prédiction avec un  $R^2_{cv}$  supérieur ou égal à 0,82 permet de réaliser de bonnes prédictions. Dans le cas où le  $R^2_{cv}$  est supérieur à 0,91, l'équation est considérée comme excellente. D'autre part, les équations présentant un RPD supérieur à 2 ont une bonne utilité pratique dans le cadre d'études scientifiques (Sinnaeve *et al.*, 1994 ; Karoui *et al.*, 2006). De plus, selon Williams (1987), des équations de prédiction avec un RER inférieur à 3 n'ont que peu d'utilité pratique tandis que celles avec un RER supérieur à 10 ont une aptitude à prédire. Avec des  $R^2_{cv}$  supérieur ou égal à 0,81, des RPD égal à 2,3 et des RER supérieurs à 10, les équations développées sont prometteuses mais pourraient encore être améliorées, notamment au moyen d'échantillons supplémentaires.

La proximité des valeurs obtenues au niveau des erreurs-types de calibration et de cross-validation ainsi que celle au

niveau des coefficients de détermination de calibration et de cross-validation permettent de supposer une bonne robustesse des deux premières équations ainsi construites. Néanmoins, ces deux équations devront être validées dans l'avenir sur un nouveau jeu d'analyses supplémentaires. Si certains de ces échantillons supplémentaires apportent de l'information supplémentaire, ceux-ci seront incorporés dans les jeux de données de calibration initiaux et de nouvelles équations de prédiction seront développées.

Des équations de prédictions basées uniquement sur les taux butyreux et protéiques pour les échantillons présents dans les jeux de données de calibration finaux (Tableau 1) seraient moins fiables que celles basées sur la spectrométrie MIR. En effet, les coefficients de détermination pour la prédiction du RdFF et du RdFS sur base de ces deux taux sont respectivement de 0,66 et de 0,73.

### 2.2. VARIABILITE AU SEIN DU CHEPTEL LAITIER WALLON

#### 2.2.1. Evolution au cours des lactations

Les évolutions des RdFF (Figure 1) et RdFS (Figure 2) en race Holstein au cours des trois premières lactations moyennés par classes de 5 jours de lactation ont montré qu'en moyenne, les rendements fromagers diminuent au cours des deux premiers mois de la lactation suivi d'une augmentation linéaire jusqu'à la fin de la lactation. Ceci concorde parfaitement avec les diminutions des taux protéiques et butyreux observés en début de lactation suivi d'une augmentation au cours du reste de la lactation.

#### 2.2.2. Variabilité génétique

Les figures 3 et 4 montrent l'évolution des héritabilités journalières estimées pour, le RdFF et le RdFS au cours de la première lactation en race Holstein, respectivement. Les deux évolutions ont le même profil. L'héritabilité minimale a été observée en début de lactation (5<sup>ème</sup> JDL), et est de 0,31 pour les deux caractères. Les profils montrent une augmentation continue au cours de la lactation pour atteindre un maximum au début du quatrième trimestre de la lactation suivi d'une légère diminution. Les héritabilités journalières maximales étaient de 0,59 au 279<sup>ème</sup> JDL pour le RdFF et de 0,57 au 299<sup>ème</sup> JDL pour le RdFS. En moyenne sur la lactation, l'héritabilité journalière moyenne ( $\pm$  erreur-

**Tableau 1** Paramètres des équations de prédictions développées pour les rendements fromagers déterminés en laboratoire et exprimés en frais (RdFF) ou en sec (RdFS)

	RdFF (g caillé / 100 g lait)	RdFS (g MS caillé / 100 g MS lait)
<b>Jeu de données de calibration final</b>		
<b>N</b>	234	243
<b>Moyenne (écart-type)</b>	26,8 (6,5)	61,9 (6,4)
<b>Minimum - Maximum</b>	13,8 - 47,9	47,1 - 79,7
<b>Calibration</b>		
<b>Erreur-type (SE<sub>c</sub>)</b>	2,6	2,6
<b>Coefficient de détermination (R<sup>2</sup><sub>c</sub>)</b>	0,83	0,84
<b>Validation croisée</b>		
<b>Erreur-type (SE<sub>cv</sub>)</b>	2,8	2,7
<b>Coefficient de détermination (R<sup>2</sup><sub>cv</sub>)</b>	0,81	0,82
<b>Rapport des performances sur les variabilités (RPD)</b>	2,3	2,3
<b>Etendue du rapport d'erreur (RER)</b>	12,0	11,9

type journalière moyenne) était de 0,52 ( $\pm$  0,03) pour le RdFF et de 0,51 ( $\pm$  0,03) pour le RdFS. Ces héritabilités journalières étaient modérées à élevées. Néanmoins, des analyses de composantes de la variance supplémentaires

sont planifiées en vue de vérifier les composantes et les héritabilités ainsi obtenues. En effet, ces valeurs pourraient être sous ou sur estimées selon le modèle et hypothèse utilisés. Dès lors, des modèles utilisant des polynômes de

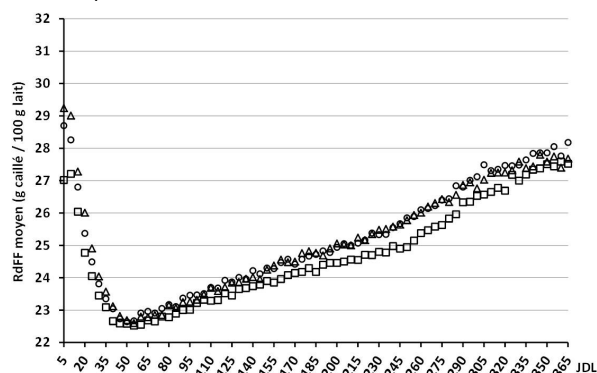
Legende d'ordre supérieur à 2 et/ou incluant l'hétérogénéité de la variance résiduelle seront aussi testés.

De telles valeurs d'héritabilité estimée indiquent qu'il est possible de sélectionner génétiquement des vaches en vue d'augmenter les rendements fromagers sans devoir passer par la sélection sur base d'indicateurs souvent repris dans les formules d'estimation du rendement fromager attendu, tels que, par exemple, les taux protéiques, butyreux ou de caséines.

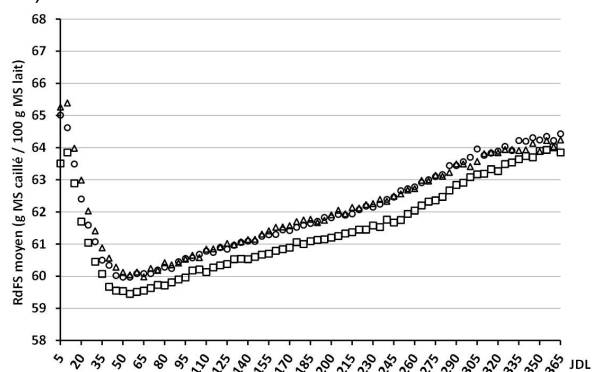
## CONCLUSION

Les équations de prédiction des rendements fromagers déterminés en laboratoire et exprimés en frais ou en sec sur base du spectre MIR d'un échantillon de lait sont prometteuses. Néanmoins, celles-ci devront être améliorées sur base d'analyses supplémentaires afin d'accroître leur robustesse. Cette première étude chimiométrique a montré la possibilité de construire des équations de prédiction des rendements fromagers permettant ainsi leurs études à large échelle, notamment au niveau du cheptel laitier wallon. Les héritabilités journalières de ces deux caractères étaient modérées à élevées, indiquant ainsi la possibilité de sélectionner génétiquement pour ces deux nouveaux caractères

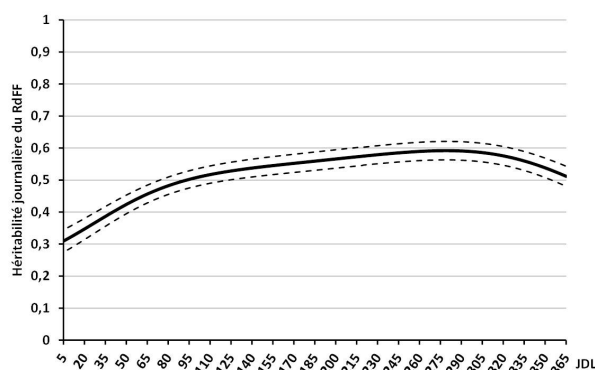
**Figure 1** Evolution du rendement fromager déterminé en laboratoire et exprimé en frais (RdFF) moyen prédit par MIR par classe de 5 jours de lactation (JDL) au cours des trois premières lactations en race Holstein en Région Wallonne (carré = 1<sup>ère</sup> lactation, rond = 2<sup>nde</sup> lactation, triangle = 3<sup>ème</sup> lactation)



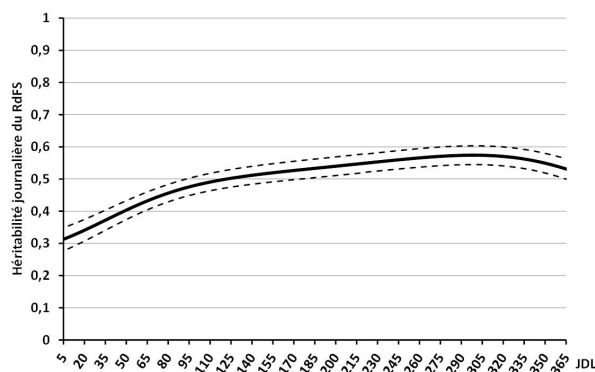
**Figure 2** Evolution du rendement fromager déterminé en laboratoire et exprimé en sec (RdFS) moyen prédit par MIR par classe de 5 jours de lactation (JDL) au cours des trois premières lactations en race Holstein en Région Wallonne (carré = 1<sup>ère</sup> lactation, rond = 2<sup>nde</sup> lactation, triangle = 3<sup>ème</sup> lactation)



**Figure 3** Héritabilités journalières (en ligne continue) ± un écart-type (en lignes discontinues) de la prédiction par MIR du rendement fromager déterminé en laboratoire et exprimé en frais (RdFF) au cours de la première lactation de vaches Holstein



**Figure 4** Héritabilités journalières (en ligne continue) ± un écart-type (en lignes discontinues) de la prédiction par MIR du rendement fromager déterminé en laboratoire et exprimé en sec (RdFS) au cours de la première lactation de vaches Holstein



Les auteurs tiennent à remercier le Service Public de Wallonie – Direction Générale Opérationnelle de l'Agriculture, des Ressources Naturelles et de l'Environnement et le Fond Européen de Développement Régional pour leur soutien financier au travers des projets ProFARMilk (D31-1255 et D31-1308) et BlueSel. Les moyens de calcul ont été fournis par le Consortium des Equipements de Calcul Intensif (CECI), financé par le Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) sous la convention n°2.5020.11.

**Fisher T.M. et al., 2004.** Genet. Sel. Evol., 36, 363-369  
**International Dairy Federation, 2010.** IDF Bulletin 446/2010  
**Hurtaud C. et al., 1995.** Ann. Zootech., 44, 385-398  
**Karoui et al., 2006.** Eur. Food Res. Tech., 222, 165-170  
**Misztal I., 2012.** University of Georgia, Athens, USA.  
 Disponible sur <http://nce.ads.uga.edu/~ignacy/newprograms.html>  
**Sinnaeve et al., 1994.** J. Near Infrared Spectrosc., 2, 79-84  
**Williams, 1987.** Near-Infrared Technology in the Agriculture and Food Industries. 2<sup>nd</sup> ed. P.C. Williams and Norris, ed. American Association of Cereal Chemists, St Paul, MN, USA  
**Williams, 2003.** A short course in the practical implementation of near infrared spectroscopy for the user. 1.1 ed. PDK Projects Inc., Nanaimo, BC, Canada