



# La classification des carcasses porcines en utilisant différentes méthodes d'analyse statistique

**Prédiction de la composition de la carcasse de porc et de ses pièces principales par différentes méthodes statistiques dont la régression linéaire, k plus proches voisins et processus gaussiens**

**Mots-clés :** Carcasses porcines, Classification, Régression linéaire, K plus proches voisins, Processus gaussiens

**Auteur :** Maria Font-i-Furnols<sup>1</sup>, Félix-Antoine Ouellet<sup>2</sup>, Hugo Larochelle<sup>2</sup>, Marcel Marcoux<sup>3</sup>, Candido Pomar<sup>3</sup>

<sup>1</sup>IRTA-Qualité du Produit, Finca Camps i Armet, 17121 Monells, Catalogne, Espagne ; <sup>2</sup>Université de Sherbrooke-Département d'informatique, 2500 Boulevard de l'Université, Sherbrooke, Québec, Canada, J1K 2R1 ; <sup>3</sup>AAC-Centre de Recherche et de développement sur le bovin laitier et le porc, 2000 rue Collège, Sherbrooke, Québec, Canada, J1M 0C8

\* E-mail de l'auteur correspondant : [candido.pomar@agr.gc.ca](mailto:candido.pomar@agr.gc.ca)

Le rendement en viande est le critère de qualité le plus utilisé dans les systèmes de classification de carcasses de porc. Toutefois, la composition des pièces primaires est essentielle à la détermination de la valeur économique des carcasses. Différentes méthodes statistiques ont été évaluées pour prédire le rendement en viande et la valeur commerciale des carcasses à l'aide des mesures traditionnelles d'épaisseur du gras et du muscle dorsal.

---

## Résumé :

L'épaisseur du gras et du muscle dorsal sont utilisés généralement pour prédire le rendement en maigre (RM) à l'aide d'équations de régression (LR). D'autres méthodes statistiques estimant le RM des carcasses et la composition des pièces principales à partir de ces variables ont été évaluées. Toutes les méthodes utilisées ont obtenu des résultats acceptables, mais la LR et les processus gaussiens sont celles ayant la plus petite erreur. La LR reste la méthode plus précise et la plus simple à utiliser.

---

## Abstract: Comparison of different statistical analysis methods for pork carcass grading

Measurements of fat and dorsal muscle (*longissimus* muscle) thickness are generally used to predict lean yield with the help of regression equations. Other statistical methods predicting lean yield and meat structure of the major muscles were estimated from these inputs. All the methods tested obtained acceptable results. But the regression equations method and Gaussian processes had the lowest error rate. The regression equations method appeared to be so far the most precise and the simplest method that can be used.

---

## INTRODUCTION

Un des principaux objectifs d'un système de classement des carcasses est d'assurer la transparence au niveau de la mise en marché. L'évaluation de la composition des carcasses et leur classement est utilisée pour établir leur valeur commerciale, ce qui est important pour les producteurs et les transformateurs de porc (Cross & Savell, 1994; Olsen et al., 2007). Le rendement en maigre (RM) est le critère de qualité le plus utilisé dans les systèmes de classification de carcasses de porc (Pomar et al., 2008). Toutefois, le poids et la composition des pièces principales sont essentiels pour estimer la valeur commerciale de la carcasse (Marcoux et al., 2007). L'épaisseur du gras et du muscle dorsal sont les variables les plus utilisées pour prédire le RM à l'aide d'équations de régression (LR).

L'apprentissage automatique est une discipline visant au développement de modèles qui permettront à la machine d'évoluer par apprentissage et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques. La méthode des k plus proches voisins (kNN) est l'un des algorithmes les plus

élémentaires de classification, où l'algorithme prédit la cible souhaitée d'un exemple donné en utilisant les informations des k exemples les plus similaires. Le processus gaussien (GP) permet d'exprimer a priori la forme de la fonction prédictive (inconnue) que nous essayons de découvrir. Les méthodes kNN et GP peuvent ainsi être exploitées pour prédire le RM, car ils permettent d'établir les relations linéaires et non linéaires entre plusieurs variables d'intérêt simultanément. Le kNN et GP avec des noyaux non linéaires estiment les variables d'intérêt de la carcasse en identifiant les carcasses de la base de données (bibliothèque) les plus proches ainsi évitant l'utilisation d'une équation de prédiction pour chaque attribut.

L'objectif de cette étude était d'évaluer et de comparer LR, kNN et GP pour l'estimation du RM de la carcasse ainsi que le poids et la teneur en maigre, gras et os des 4 pièces principales à partir de l'épaisseur du gras et du muscle dorsal.

## I. METHODOLOGIE

### I.1. Carcasses et dissection

Soixante carcasses (29 femelles et 31 castrats) ont été utilisées pour la réalisation de cette étude. Les carcasses retenues étaient représentatives des différents types de carcasses observés dans les abattoirs selon le poids, l'épaisseur de gras et la conformation. Les épaisseurs du gras et du muscle ont été mesurées à 7 cm de la ligne médiane de la carcasse entre les 3èmes et 4èmes dernières côtes avec la sonde Destron PG-100 (Figure 1). Le jambon,

la longe, le flanc et l'épaule ont été coupés selon la présentation standard canadienne avec un ajustement proportionnel à la longueur de la carcasse (Figure 2). Les quantités de muscle, de gras et d'os des pièces ont été obtenues par dissection. En plus du RM de la carcasse, les proportions de chaque tissu ont été calculées par rapport au poids de la coupe correspondante.

**Figure 1 :** Épaisseur du gras et du muscle mesurée au niveau de la longe avec la sonde Destron PG-100



**Figure 2 :** Coupe canadienne de la carcasse de porc



## **1.2. Analyses des données**

Les approches statistiques suivantes ont été étudiées en utilisant les épaisseurs du gras et du muscle obtenues avec la sonde Destron PG-100 comme variables indépendantes. L'erreur de prédiction quadratique moyenne (RMSEP) a été calculée à l'aide d'une validation croisée de type « leave-one-out ». A partir du RMSEP et de la moyenne, le coefficient de variation a été aussi calculé avec l'équation suivante :  $CV=100*RMSEP/moyenne$ .

### ***1.2.1. Régression linéaire***

La procédure REG de SAS (SAS Institute Inc., Cary, NC, USA) a été utilisée pour obtenir les équations de régression prédisant séparément chacune des variables d'intérêt.

### ***1.2.2. k plus proches voisins***

La sélection de la valeur de k a été réalisée à l'aide d'une validation croisée de type « holdout », en isolant 15% de l'ensemble d'entraînement pour validation.

La valeur de k sélectionnée est alors celle minimisant l'erreur de prédiction sur ce sous-ensemble de validation.

Toutes les variables d'intérêt sont prédites avec cette méthode simultanément (multitâches).

### ***1.2.3. Processus gaussien***

Un noyau linéaire a été choisi dans l'application d'un processus gaussien en mode multitâches, puisqu'une validation croisée a identifié que d'autres noyaux non-linéaires causaient des problèmes de surapprentissage.

Les valeurs des hyper-paramètres du noyau linéaire ont été obtenues par l'optimisation de la vraisemblance marginale des données d'apprentissage.

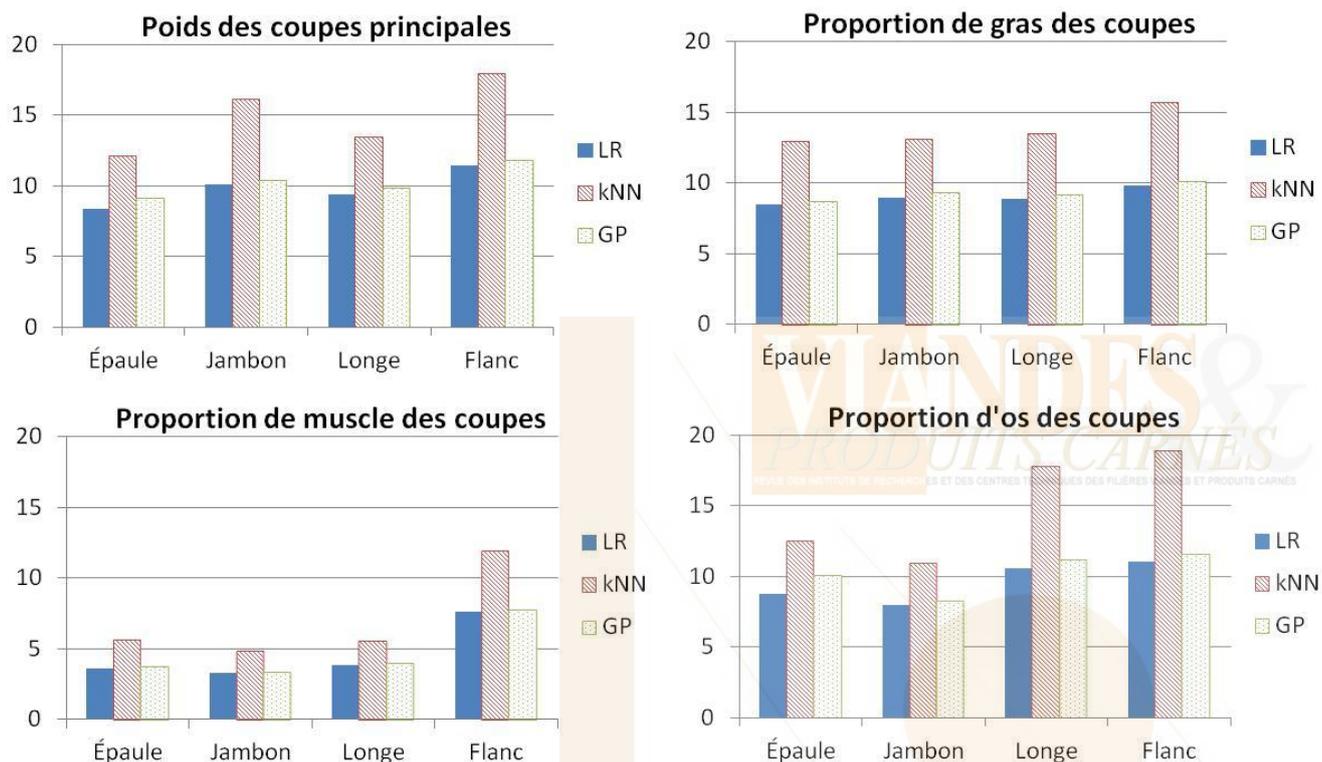
## **II. RESULTATS ET DISCUSSIONS**

Lors de la prédiction du RM, le RMSEP était respectivement de 2,18%, 3,20% et 2,19% pour les méthodes LR, kNN et GP et le CV de 3,58%, 5,25% et 3,60%. Le niveau de précision requis dans la législation européenne (Règlement CE n° 1249/2008 de la Commission) est une RMSEP maximale pour estimer le RM de 2,5%. LR et GP étaient inférieurs à la limite UE pour le RMSEP, mais pas kNN.

Les résultats de l'erreur de prédiction relative pour estimer les poids des pièces principales et la proportion de gras, muscle et d'os des pièces principales sont présentés à la Figure 3. Toutes les méthodes utilisées ont permis d'obtenir une précision de prédiction acceptable des

variables d'intérêt, quoique les méthodes LR et GP sont celles ayant une RMSEP relative (CV) entre 1,5% et 7,8% inférieur à celle obtenue avec la méthode kNN indépendamment de la variable prédite. La prédiction de la proportion de muscle des pièces a été la plus précise après celle du RM de la carcasse. L'erreur de prédiction de la proportion de muscle a été la plus faible pour le jambon et la plus grande pour le flanc. Ces résultats indiquent qu'il est possible d'utiliser seulement deux variables prédictives pour estimer le RM de la carcasse et la composition de ses pièces principales à condition que l'erreur obtenue se situe dans des limites acceptables pour les besoins de l'intéressé.

**Figure 3 : Erreur de prédiction relative ( $CV(\%)=100*RMSEP/moyenne$ ) obtenu par régression linéaire (LR), k plus proches voisins (kNN) et le processus gaussiens (GP)**



## CONCLUSION

Dans le contexte d'utilisation de seulement deux variables prédictrices pour estimer le RM de la carcasse ou la composition de l'épaule, le jambon, la longe et le flanc, les meilleurs résultats ont été obtenus avec la méthode LR et GP. Les résultats pourraient être améliorés pour toutes les méthodes si le nombre de carcasses utilisées était plus élevé

et si d'autres variables prédictrices (épaisseurs de gras et muscle en autres régions, area du muscle, volume, etc.) étaient incluses dans les modèles. Les résultats présentés montrent que la méthode LR demeure néanmoins la méthode la plus simple et la plus précise.

## Remerciements :

Ce travail a été réalisé grâce au financement accordé à Mme Maria Font-i-Furnols de OECD Programme Coopératif de Recherche : Biological Resource Management for Sustainable Agricultural Systems pour 2012-13.

## Références :

- Cross H. R., Savell J. W. (1994). What do we need for a value-based beef marketing system?. *Meat Science*, 36, 19-27.
- Marcoux M., Pomar C., Faucitano L., Brodeur C. (2007). The relationship between different pork carcass lean yield definition and the market carcass value. *Meat Science*, 75, 94-102
- Olsen E.V., Candek-Potokar M., Oksama M., Kien S., Lisiak D., & Busk, H. (2007). On-line measurements in pig carcass classification: repeatability and variation caused by the operator and the copy of the instrument. *Meat Science*, 75, 29-38
- Pomar C., Marcoux M., Gispert M., Font i Furnols M., Daumas G. (2008). Determining the lean meat content of pork carcasses. In: *Improving the sensory and nutritional quality of fresh meat* (Chapter 21). Ed. J. Kerry and D. Ledward. Woodhead Publishing Limited, Cambridge, UK (pp. 493-518)