

Comparaison de méthodes pour valider l'estimation par scanner à induction magnétique de la composition de jambons et de poitrines



Gérard Daumas (1), Mathieu Monziols (1), Juan Manuel Rodriguez (2), Jacobo Alvarez-Garcia (2), David Causeur (3)

(1) Ifip-Institut du porc, BP 35104, 35651 Le Rheu Cedex, France

(2) LENZ Instruments SL, calle Santander 42 nave 36, 08020 Barcelona, Spain

(3) Agrocampus Ouest, Irmar, UMR 6625 CNRS, 65 rue de St-Brieuc - CS 84215, 35042 Rennes Cedex, France

Contact : gerard.daumas@ifip.asso.fr



AGRO CAMPUS OUEST

LENZ INSTRUMENTS

La validation statistique est une étape à ne pas négliger dans le processus de test d'une technologie. Néanmoins, il n'y a pas de consensus sur la méthode à appliquer. Les résultats semblent dépendre de la nature des données. Aussi, il est souvent conseillé de tester plusieurs méthodes. L'objectif de cette étude est de comparer la performance de cinq méthodes de régression linéaire parmi les plus courantes, dans le cadre de la prédiction de la composition tissulaire de pièces de découpe.

Matériel et Méthodes

- Deux échantillons : l'un de 100 jambons et l'autre de 80 poitrines
- Analysés par un scanner à induction magnétique (HAM-Inspector II), développé par Lenz Instruments S.L. (Barcelone, Espagne) (Fig.1)
- Mesurés par tomographie à densité (référence) (Fig. 2)



Fig.1 : Scanner à induction magnétique



Fig. 2 : Scanner à rayons X

- 4 variables à prédire pour chacune des pièces :
 - poids de muscle et poids de gras,
 - pourcentage de muscle et pourcentage de gras.
- Variables explicatives nombreuses et colinéaires (cf. Daumas & al, JRP 2019)
- Calcul du R^2 et de l'erreur de prédiction en validation croisée, en réalisant 100 partitions aléatoires en 10 segments de données.
- Comparaison avec le logiciel R de 5 méthodes de régression linéaire :
 - **OLS** : estimation par la méthode usuelle des moindres carrés ordinaires ;
 - **Lasso** : minimisation du critère des moindres carrés ordinaires pénalisé par un terme proportionnel à la somme des valeurs absolues des coefficients de régression, avec choix du paramètre de pénalité par minimisation de l'erreur de prédiction ;
 - **Ridge** : minimisation du critère des moindres carrés ordinaires pénalisé par un terme proportionnel à la somme des carrés des coefficients de régression, avec choix du paramètre de pénalité par minimisation de l'erreur de prédiction ;
 - **PLS** : moindres carrés partiels, avec choix du nombre de composantes par minimisation de l'erreur de prédiction ;
 - **Subset** : minimisation du critère d'information bayésien (BIC) sur tous les sous-modèles possibles.

Résultats

Pour les jambons (fig. 3) :

- Ecart maximal de R^2 médian entre méthodes de 0,14 (pour le % de muscle) ;
- PLS la meilleure méthode et OLS la moins bonne ;
- Dispersion la plus faible avec PLS et la plus forte avec OLS.

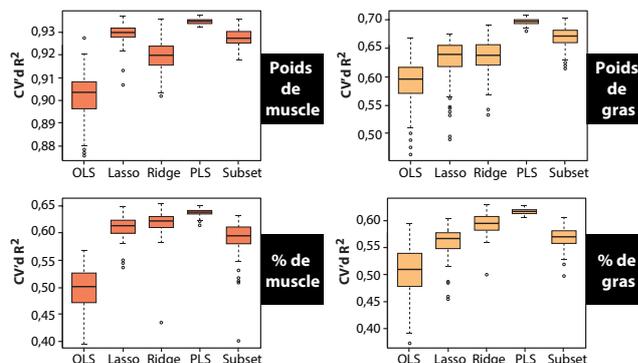


Fig. 2 : Jambon : distribution du R^2 en validation croisée par méthode de prédiction

Pour les poitrines (fig. 4) :

- Ecart maximal de R^2 médian entre méthodes de 0,22 (pour le poids de muscle) ;
- PLS la moins bonne méthode, Ridge la meilleure, sauf pour le % de gras où c'était Subset ;
- Dispersion la plus faible avec PLS pour les teneurs en gras et muscle.

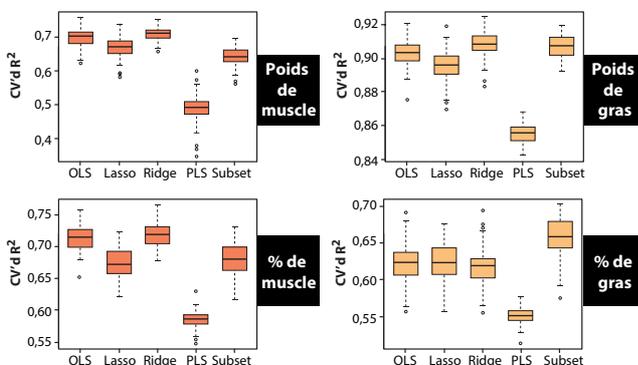


Fig. 4 : Poitrine : distribution du R^2 en validation croisée par méthode de prédiction

Conclusion

Les différences entre les cinq méthodes statistiques testées étaient conséquentes. La hiérarchie des méthodes en termes de performance de prédiction dépendait de la pièce de viande. Les méthodes Subset, Lasso et Ridge seraient à privilégier, car ayant montré des performances de prédiction plus stables pour l'ensemble des deux pièces et des variables à prédire, toujours proches ou égales à la meilleure performance.

