

VERS UNE MEILLEURE COMPREHENSION DE LA VARIABILITE ENTRE PONDEUSES : CLASSIFICATION SELON LES TRAJECTOIRES DE PONTE

Guyot Yann¹, Burlot Thierry², Quentin Maxime¹, Méda Bertrand³, Pampouille Eva¹

¹ITAVI, 7 rue du Faubourg Poissonnière, 75009 PARIS

²NOVOGEN, 5 rue des compagnons, Secteur du Vau Ballier, 22960 PLEDRA

³INRAE, Université de Tours, BOA, 37380 NOUZILLY

guyot@itavi.asso.fr

RÉSUMÉ

L'étude et la compréhension de la variabilité interindividuelle chez la poule pondeuse sont essentielles pour sélectionner des individus et/ou adapter les pratiques d'élevage, en particulier les stratégies nutritionnelles. Cette étude propose une méthode de regroupement (*clustering*) non supervisé d'individus selon la similarité des cinétiques de ponte, appliquée à une base de données décrivant les trajectoires individuelles de ponte (n=976) à l'aide de quatre variables mesurées sur 8 séries consécutives de 35 jours : le taux de ponte, le nombre de séries, l'heure de ponte moyenne et l'écart-type de l'heure de ponte. La méthode de classification non supervisée employée est la méthode *k-means* multivariée car elle permet de considérer simultanément les quatre variables du jeu de données. Le nombre de clusters a été fixé à 5 afin de conserver un nombre restreint de clusters. Le premier cluster représente le groupe le plus important (n=536), les individus de ce groupe présentent un bon taux de ponte jusqu'à 50 semaines d'âge ainsi qu'une bonne persistance de ponte. Le deuxième cluster regroupe des animaux avec une persistance de ponte plus courte caractérisée par un décalage de l'heure de ponte plus important ainsi qu'une baisse de la durée des séries (n=311). Enfin, les animaux présentant des taux de ponte fortement dégradés sont regroupés dans trois clusters se différenciant principalement par leur taux de ponte : i) un taux de ponte élevé en début de ponte avant de chuter aux alentours de 50 semaines d'âge (n=70), ii) un taux de ponte continuellement décroissant sur la carrière des animaux (n=33) et iii) un taux de ponte croissant sur la première partie de la carrière puis décroissant sur la deuxième (n=26). La méthode de classification présentée ici constitue une nouvelle approche pour identifier des groupes homogènes d'individus, dans une perspective d'aide à la sélection génétique. Les données générées dans cette étude pourraient également être mobilisées pour des approches de modélisation à l'échelle du troupeau.

ABSTRACT

Toward a Better Understanding of Variability Among Laying Hens: Classification Based on Laying Trajectories

Understanding of the inter-individual variability in laying hens is essential to select individuals and/or adapt breeding practices, especially nutritional strategies. This study proposes an unsupervised clustering method of individuals based on the similarity of laying kinetics, applied to a database describing the individual laying trajectories (n=976) using four variables measured over 8 consecutive series of 35 days: laying rate, number of clutches, mean and standard deviation of oviposition time. The unsupervised clustering method used is the multivariate *k-means* method as it allows for the simultaneous consideration of all four variables in the dataset. The number of clusters was set at 5 to maintain a limited number of clusters. The first cluster represented the largest group (n=536), grouping individuals with a good laying rate up to 50 weeks of age and a good laying persistency. The second cluster grouped animals with lower laying persistency characterized by a greater time shift in oviposition time and a decrease in clutch length (n=311). Finally, animals with significantly degraded laying rates were grouped into three clusters mainly differing in their laying rate: i) a high laying rate at the beginning of laying before declining around 50 weeks of age (n=70), ii) a continuously decreasing laying rate over the career (n=33), and iii) a laying rate that increases in the first part of the career and then decreases in the second part (n=26). The clustering method presented here represents a novel approach to identifying homogeneous groups of individuals, with a perspective on assisting genetic selection. The data generated in this study could also be used for herd-scale modelling approaches.

INTRODUCTION

L'analyse et la compréhension de la variabilité interindividuelle parmi les poules pondeuses jouent un rôle important dans la sélection des animaux et l'adaptation des pratiques d'élevage, comme par exemple les stratégies alimentaires à adopter pour satisfaire au mieux les besoins moyens du troupeau. Chaque animal ayant une cinétique de ponte particulière, il peut être délicat d'avoir une vision globale du troupeau et d'identifier des tendances et des caractéristiques communes.

Les techniques de regroupement non supervisé (ou *clustering*) telle que la méthode *k-means* permettent de partitionner un ensemble de séries temporelles en sous-groupes sans connaissance préalable sur ce qui les distingue les uns des autres (Steinley, 2006). Dès lors, cette méthode permet de mettre en évidence des caractéristiques communes entre individus, utiles à la compréhension des phénotypes des animaux d'élevage. Ainsi, ce type de classification a récemment été utilisé pour classer des poulets selon leur comportement (Feiyang *et al.*, 2016), des poules selon l'utilisation du parcours et des équipements (Sibanda *et al.*, 2020) ou encore des vaches laitières en temps réel selon leur ingéré quotidien (Gauthier *et al.*, 2021). La présente étude visait à proposer une classification de poules pondeuses en différents clusters suivant leur trajectoire de ponte, définie comme la résultante du taux de ponte, du nombre de séries et de l'heure de ponte (moyenne, écart-type).

1. MATERIELS ET METHODES

1.1. Données individuelles

Le jeu de données sur lequel s'appuie la classification a été fourni par Novogen et est constitué de 976 trajectoires de ponte individuelles (lignées pures White Plymouth Rock élevées au sol). La durée de lumière était de 16h (lumière entre 2h et 18h).

Dans cette base de données, chaque individu est décrit à l'aide de quatre variables mesurées sur 8 périodes consécutives de 35 jours, à partir de 21 semaines d'âge (Tableau 1, seuls les animaux encore en ponte à la 8^{ème} période ont été inclus, excluant ainsi les animaux morts en cours de lot) :

- Le taux de ponte (nombre d'œufs sur 35 jours)
- Le nombre de séries (nombre de séquences continues d'œufs sans interruption sur 35 jours)
- L'heure de ponte post extinction moyenne sur 35 jours (relevée par nids électroniques à la seconde près)
- L'écart type de l'heure de ponte sur 35 jours

Les valeurs moyennes des quatre variables avant classification pour les périodes 2 (26-31 sem.), 5 (44-49 sem.) et 8 (62-67 sem.) sont données à titre indicatif dans le Tableau 1.

1.2. Classification avec la méthode *k-means*

La méthode utilisée dans cette étude est la méthode dites *k-means*. Cette dernière permet de regrouper des éléments similaires dans des ensembles appelés « clusters ». La méthode *k-means* cherche à diviser un ensemble de données en sous-groupes de manière à ce que les éléments d'un même groupe soient plus similaires les uns aux autres qu'aux éléments des autres groupes. De plus, chaque cluster est accompagné par son centroïde, point de référence qui définit la position centrale du cluster. Plus précisément, la méthode *k-means* utilisée ici est l'extension multivariée sur séries temporelles, qui permet de regrouper simultanément plusieurs séries temporelles (ici les quatre variables du jeu de données) selon leur distance euclidienne. En d'autres termes, la méthode regroupe des courbes qui « se ressemblent » pour les quatre variables du jeu de données à la fois. Le nombre de clusters a été fixé à cinq afin de conserver un nombre restreint de clusters tout en isolant des tendances bien distinctes. Ce nombre a été déterminé par la méthode dite « du coude » consistant à trouver le nombre optimal où l'ajout d'un cluster supplémentaire ne fournit plus de réduction significative de la somme des carrés intra-cluster (Rykov *et al.*, 2024). La classification a été réalisée à l'aide de la librairie python *TSlearn* (Tavenard *et al.*, 2020).

2. RESULTATS ET DISCUSSION

Les résultats graphiques de la classification sont présentés dans la Figure 1 et les valeurs moyennes des quatre variables étudiées pour chaque cluster pour les périodes 2 (26-31 sem.), 5 (44-49 sem.) et 8 (62-67 sem.) sont données dans le Tableau 1.

Le premier cluster (A) est le groupe le plus important (~55% de l'effectif total), avec des individus qui présentent un bon taux de ponte (97%) jusqu'à 50 semaines d'âge puis chutent à un taux raisonnable de 86% à 58 semaines. Ces individus se caractérisent donc par une bonne persistance de la ponte, qui peut être expliquée par une heure de ponte moyenne précoce (~11h après l'extinction) et un faible décalage de l'heure de ponte au cours de la carrière (≤1h) ainsi qu'un faible nombre de séries de ponte (inférieur ou égal à 2) qui correspond à de longues séries de ponte.

Le deuxième groupe (cluster B ; ~32% de l'effectif total) regroupe des animaux avec un taux de ponte plus faible (de 96% à 25 semaines d'âge à 76% à 58 semaines) et une plus faible persistance de ponte caractérisée par un décalage de l'heure de ponte plus important (>2,2h) ainsi qu'une augmentation du nombre de séries (~5 à 58 semaines).

Enfin, les individus présentant une nette détérioration de leur taux de ponte (représentant environ 13% de l'effectif) sont répartis en trois groupes principalement distingués par l'évolution de taux de ponte au cours du temps :

- i. Cluster C (7,2% de l'effectif total) : ce groupe présente un fort taux de ponte en début de carrière (96%), suivie d'une forte chute initiée autour de 50 semaines pour atteindre 28% à 58 semaines ;
- ii. Cluster D (3,4% de l'effectif total) : ce groupe présente une décroissance constante de la ponte tout au long de la carrière, de 89% à 25 semaines à 48% à 58 semaines ;
- iii. Cluster E (2,6% de l'effectif total) : ce groupe présente un pic de ponte tardif, après une lente hausse de la ponte sur la première moitié de la carrière, suivie d'un plateau à 91% pour finir par une chute à 75% à 58 semaines d'âge.

Ces résultats illustrent bien le principe de clustering multivarié et permettent ainsi d'identifier des profils pouvant correspondre à des « bonnes » pondeuses (cluster A), des pondeuses « intermédiaires » (cluster B) et des « mauvaises » pondeuses aux profils variés (clusters C, D et E). En particulier, pour ces clusters, il est intéressant de noter que ce sont ces animaux qui sont majoritairement regroupés lorsque la méthode *k-means* est appliquée sur le jeu de données avec un nombre de clusters fixé à 3, la composition des autres clusters (A et B) étant relativement peu affectée.

La classification des animaux a été réalisée sur des animaux âgés au maximum de 60 semaines. Aussi, dans le cadre d'un allongement des carrières productives des animaux, il pourrait être intéressant d'appliquer cette approche sur des données de production « complètes » avec des animaux en production jusqu'à environ 90-100 semaines d'âge (Bain *et al.*, 2016), ce qui permettrait peut-être de mettre en évidence d'autres comportements de ponte sur le long terme. Savegnago *et al.* (2011) présentent une méthode de clustering similaire (*k-means*) sur le taux de ponte afin d'estimer les paramètres génétiques des caractères liés à la production d'œufs et d'identifier ceux qui pourraient être utilisés comme critères de sélection pour améliorer les performances. La méthode présentée ici permet de regrouper les animaux non seulement sur le taux de ponte mais également sur d'autres critères, ce qui permettrait d'affiner la sélection des animaux et définir des règles de sélection complexes (améliorer la productivité, synchroniser au maximum les animaux sur une heure

de ponte précoce pour réduire la mobilisation calcique nocturne...).

La définition de clusters hiérarchisant des poules pondeuses peut également servir de références pour mesurer la variabilité intra-lot ou inter-lot (en comparant les résultats de classification avec un autre lot par exemple) et ainsi aider à la prise de décision de sélection.

La classification présentée ici est faite sur les trajectoires de ponte complètes (8 périodes), or il est possible que certains animaux présentent des profils similaires sur un certain temps (les premières périodes) puis divergent par la suite. Une approche plus dynamique de clustering prenant en compte l'évolution temporelle (avec incrémentation période après période) de la trajectoire permettrait d'affiner les connaissances sur le comportement de ponte intra-lot et d'identifier passant d'un cluster à un autre au cours du temps.

CONCLUSION

L'analyse des séries temporelles de ponte au moyen de l'approche de classification multivariée *k-means* a permis de regrouper les poules pondeuses en cinq groupes d'individus en fonction de leur trajectoire de ponte. Cela a révélé des tendances et des caractéristiques communes parmi les animaux. Cette classification offre des informations utiles pour la sélection génétique, contribuant ainsi à une meilleure compréhension de la diversité au sein d'un troupeau.

Elle ouvre également la voie à des pratiques d'élevage plus adaptées en fonction des individus, en particulier à des stratégies alimentaires plus fines (cf. alimentation sur mesure), notamment pour l'alimentation minérale (calcium).

REMERCIEMENTS

Cette étude a été réalisée avec l'aide financière du Ministère de l'Agriculture et de la Souveraineté alimentaire dans le cadre du projet CAS DAR OCALM (2021-2023).

REFERENCES BIBLIOGRAPHIQUES

- Bain M.M., Nys Y., Dunn I.C., 2016. *Brit. Poultry Sci.*, (57), 330-338.
- Feiyang Z., Yueming H., Liancheng C., Lihong G., Wenjie D., Lu W., 2016. *Int. J. Agr. Biol. Eng.*, (9), 139-147.
- Gauthier R., Largouët C., Rozé L., Dourmad, J.-Y., 2021. *Comput. Electron. Agr.*, (188), 106329.
- Rykov, A., De Amorim, R. C., Makarenkov, V., & Mirkin, B. 2024. *IEEE Access*.
- Savegnago R.P., Caetano, S.L., Ramos S.B., Nascimento G.B., Schmidt G.S., Ledur M.C., Munari, D.P., 2011. *Poult.Sci.*, (90), 2174-2188.
- Sibanda T. Z., Welch M., Schneider D., Kolakshyapati M., Ruhnke I., 2020. *Animals*, (10), 855.
- Steinley D., 2006. *Brit. J. Math. Stat. Psy.*, (59), 1-34.
- Tavenard R., Faouzi J., Vandewiele G., Divo F., Androz G., Holtz C., Payne M., Yurchak R., Rußwurm M., Kolar K., Woods E., 2020. *J. Mach. Learn. Res.*, (21), 1-6.

Tableau 1. Valeurs moyennes ($\pm$ écart-type) par période de mesure pour les quatre variables du jeu de données avant et après classification par la méthode *k-means*.

Variable	Période	Âge (semaines)	Avant classification (n=976)	Après classification				
				Cluster A (n=536)	Cluster B (n=311)	Cluster C (n=70)	Cluster D (n=33)	Cluster E (n=26)
Nombre de séries de ponte	2	26-30	1,3 $\pm$ 0,64	1,1 $\pm$ 0,39	1,4 $\pm$ 0,77	1,2 $\pm$ 0,36	1,9 $\pm$ 1,16	1,6 $\pm$ 1,28
	5	41-45	2,1 $\pm$ 1,33	1,4 $\pm$ 0,68	3,2 $\pm$ 1,34	1,9 $\pm$ 1,03	3,2 $\pm$ 2,06	1,6 $\pm$ 0,74
	8	56-60	3,0 $\pm$ 1,93	2,1 $\pm$ 1,32	4,8 $\pm$ 1,68	1,8 $\pm$ 0,95	3,6 $\pm$ 2,38	2,9 $\pm$ 1,35
Taux de ponte (%)	2	26-30	96 $\pm$ 7,85	97,4 $\pm$ 4,2	96,3 $\pm$ 4,37	96,3 $\pm$ 6,84	89,0 $\pm$ 13,18	71,2 $\pm$ 25,63
	5	41-45	92,8 $\pm$ 9,89	96,8 $\pm$ 3,62	89,1 $\pm$ 9,11	93,8 $\pm$ 5,56	61,9 $\pm$ 21,12	91,4 $\pm$ 6,49
	8	56-60	76,9 $\pm$ 18,27	85,9 $\pm$ 7,04	75,7 $\pm$ 8,46	27,8 $\pm$ 15,84	48,1 $\pm$ 21,32	75,4 $\pm$ 13,14
Heure de ponte post extinction (moyenne, heure)	2	26-30	11,4 $\pm$ 1,27	10,9 $\pm$ 1,0	12,3 $\pm$ 1,07	11,1 $\pm$ 1,11	11,8 $\pm$ 1,43	12,1 $\pm$ 1,75
	5	41-45	12,0 $\pm$ 1,24	11,5 $\pm$ 1,12	12,9 $\pm$ 0,79	11,8 $\pm$ 1,25	12,1 $\pm$ 1,61	11,8 $\pm$ 1,2
	8	56-60	12,3 $\pm$ 1,14	11,9 $\pm$ 1,02	12,9 $\pm$ 0,73	12,2 $\pm$ 1,83	12,5 $\pm$ 1,4	12,1 $\pm$ 1,08
Heure de ponte (écart-type, min)	2	26-30	47 $\pm$ 35	39 $\pm$ 27	60 $\pm$ 40	40 $\pm$ 27	71 $\pm$ 47	75 $\pm$ 49
	5	41-45	83 $\pm$ 48	59 $\pm$ 37	122 $\pm$ 40	79 $\pm$ 47	11 $\pm$ 55	65 $\pm$ 41
	8	56-60	98 $\pm$ 48	76 $\pm$ 43	134 $\pm$ 31	100 $\pm$ 59	124 $\pm$ 48	98 $\pm$ 36

Figure 1. Résultats de la classification multivariée à l'aide de la méthode *k-means* avec les individus dans chaque cluster et le centroïde correspondant.
(1 colonne = 1 cluster ; 1 ligne = 1 variable du jeu de données)

