

Amélioration d'un outil de détection d'évènements techniques ou sanitaires à l'échelle de plusieurs bandes de truies

Gaëlle HAMELIN (1), Tom ROHMER (2), Charlotte GAILLARD (1)

(1) PEGASE, INRAE, Institut Agro, 35590, Saint-Gilles, France

(2) GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France

charlotte.gaillard@inrae.fr

Improvement of a tool to detect technical or health-related events at the scale of several groups of sows

Recently, digital agriculture has grown considerably, especially in pig farming. It is currently possible to track animals individually and collect information that can help farmers make decisions. The objective of this study was to test and compare three methods of event detection (technical or health-related) from temporal data (water consumption and physical activity) collected by automatons and sensors. The long-term goal is to successfully develop the most reliable automated tool to quickly identify changes in drinking behaviour and/or physical activity that could indicate a technical or health-related event. Data were collected from 63 pregnant sows at the INRAE UE3P experimental facility at Saint-Gilles. Four groups of sows experienced either two thermal disturbances or two dietary disturbances. In this study, the three methods of event detection tested and compared were 1) differential smoothing, 2) detection of breaks in correlations between different pairs of variables, and 3) random forests. The best detection of thermic or dietary disturbances was obtained with random forests (i.e. for one group, sensitivity to the detection of a cold event of 53% and positive predictive value of 75%). To improve detection performance, intra- and inter-individual variability should be considered better by establishing animal-specific reference curves during complete gestation periods with no events.

INTRODUCTION

Dans un contexte d'élevage de précision et grâce au numérique, des outils se développent afin d'aider les éleveurs à détecter les perturbations techniques et les troubles de santé individuels. Par exemple, des automates d'abreuvement couplés à des capteurs auriculaires mesurant l'activité physique de chaque truie permettent de collecter des données individuelles en continu. Il est cependant indispensable d'automatiser l'analyse de ces données. L'objectif de cette étude était de tester et de comparer trois méthodes de détection d'évènements à partir des données d'abreuvement et d'activité de truies gestantes. Nous avons choisi de tester une méthode de lissage différentiel, une méthode basée sur les ruptures de corrélations entre les divers couples de variables et un Random Forest.

1. MATERIEL ET METHODES

1.1. Conditions expérimentales

Les données ont été collectées à l'Unité Expérimentale Physiologie et Phénotypage des Porcs (UE3P) de Saint-Gilles (35) sur 63 truies gestantes de type Large White croisées Landrace, réparties en quatre bandes (nommées 5, 6, 16 et 17). Les truies d'une même bande ont été inséminées en même temps (± 2 jours) et sont donc toutes au même stade de gestation. Les truies sont entrées en salle de gestation 3 ou 16 jours après l'insémination (J3 ou J16) et en sont sorties vers

J106. Les truies des bandes 5 et 6 ont subi deux évènements thermiques, un froid (abaissement de la température de la salle de 20°C à 14°C) puis un chaud (augmentation de la température de la salle à 32°C) durant 4 jours chacun et espacés de 10 jours. Les truies des bandes 16 et 17 ont été soumises à deux évènements alimentaires de 5 jours (alim1 et alim2), espacés de 10 jours, et durant lesquels un seul distributeur d'aliments sur les deux était accessible.

1.2. Analyse statistique des données

Les variables étudiées pour chaque truie et par jour de gestation étaient la quantité d'eau bue (L), le nombre de buvées, la durée d'abreuvement (min), le temps moyen horaire passé allongé (min) ou debout (min) et le nombre moyen horaire de changements de position. Les données ont été analysées sur R (version 4.1.2., R Core Team, 2022) en utilisant trois méthodes de détection des perturbations : une méthode de lissage différentiel (LD) avec le package *fda* (Ramsay *et al.*, 2010), une méthode de détection des ruptures de corrélation (DR) avec le package *npcp* (Kojadinovic *et al.*, 2020) et un Random Forest (RF) avec le package *R randomForest* (Liaw *et al.*, 2002).

Les perturbations détectées par les méthodes ont été confrontées aux évènements induits, ce qui a permis de calculer des d'indicateurs visant à évaluer et à comparer les trois méthodes. Ainsi, la valeur prédictive positive (VPP) qui représente la probabilité pour un test de détecter un évènement perturbateur et la sensibilité qui caractérise la

capacité du test à correctement identifier les évènements ont été déterminées.

1.2.1. Méthode de lissage différentiel

Pour chaque truie et pour chaque variable, deux pénalités différentes ont été appliquées sur les dérivées des fonctions permettant d'établir deux courbes (pénalité forte = 100, courbe « théorique » ; pénalité faible = 10, courbe « lissée »). Une différence suffisamment importante entre les deux courbes lissées permet de détecter une perturbation potentielle. Les 15 % plus fortes valeurs (en valeur absolue) de l'aire entre les deux courbes ont été sélectionnées afin de repérer les perturbations les plus importantes qui ont ensuite été mises en relation avec les évènements induits.

1.2.2. Tests non paramétriques de détection de ruptures

La méthode a consisté à regarder si les corrélations entre chaque couple de variables et pour chaque truie variaient au cours d'un évènement. A chaque fois, deux intervalles de temps incluant un évènement ont été étudiés : du début des mesures (J3 ou J16) et jusqu'à 7 jours après le 1^{er} évènement (froid ou alim1), puis 3 jours avant le 2^{ème} évènement (chaud ou alim2) et jusqu'à la fin des mesures. L'intervalle de temps a été coupé en deux parties successives variables à partir desquelles la différence entre les corrélations de Spearman a été calculée. Lorsque la *P value* du test est $\leq 0,05$, une estimation du point de changement est réalisée. Ces points de changement sont ensuite mis en relation avec les évènements induits.

1.2.3. Random Forest

La méthode du Random Forest a été appliquée par bande de truies. Pour chaque bande, nous avons défini les hyperparamètres optimaux. Au cours des analyses, 70 % du jeu de données a été utilisé pour faire apprendre le modèle et 30 % a été conservé pour tester le modèle.

2. RESULTATS ET DISCUSSION

Les résultats présentés dans le tableau 1 montrent que le premier évènement (froid ou alim1 selon la bande) est mieux détecté que le second (chaud ou alim2 selon la bande). Ceci pourrait être attribué au fait que pour chaque bande de truies, le deuxième évènement est intervenu seulement 10 jours après le premier alors que les truies n'avaient pas totalement récupéré ni retrouvé un niveau basal selon la variable considérée. Il était alors plus difficile d'identifier clairement le début du deuxième évènement.

Une sensibilité maximale de 53 % avec une VPP de 75 % ont été obtenues par Random Forest pour la bande 6 soumise à l'évènement thermique froid. Le lissage différentiel a permis la détection de l'évènement thermique chaud avec une sensibilité de 38 % pour la bande 6 mais avec une VPP de seulement 20 %.

En ce qui concerne les évènements alimentaires, le premier évènement alimentaire a été mieux détecté par la méthode de détection des ruptures de corrélation dans la bande 17 (sensibilité de 43 % et VPP de 67 %) et de façon équivalente au Random Forest dans la bande 16 (sensibilité de 41 % et VPP de 53 %).

Tableau 1 – Sensibilité et VPP (%) selon la méthode de détection des évènements utilisée

Evènement		Bande 5			Bande 6		
		LD ¹	DR ¹	RF ¹	LD	DR	RF
Froid	Sensibilité	39	18	33	32	21	53
	VPP ¹	25	20	75	19	44	75
Chaud	Sensibilité	34	15	25	38	18	27
	VPP	21	17	60	20	13	100
Alim1 ¹	Sensibilité VPP	Bande 16			Bande 17		
		43	41	40	51	43	9
Alim2 ¹	Sensibilité VPP	19	53	55	22	67	100
		34	33	20	37	30	0
		14	36	80	17	30	0

¹LD : Lissage Différentiel, DR : Détection de Ruptures, RF : Random Forest, VPP : Valeur Prédictive Positive, Alim1 et Alim2 : 1^{er} et 2^{ème} évènement alimentaire

CONCLUSION

Nous pouvons conclure d'après les indicateurs de performance sélectionnés que, de façon générale, la méthode du Random Forest donne les meilleurs rapports sensibilité/VPP parmi les trois méthodes testées pour détecter les évènements thermiques et que pour l'une des bandes (bande 16), elle donne des résultats équivalents à la méthode de détection des ruptures de corrélation pour détecter les évènements alimentaires. Bien que les méthodes de détection des ruptures et du Random Forest permettent d'améliorer la sensibilité de la détection des perturbations par rapport au lissage différentiel, les sensibilités obtenues restent trop faibles (max 53 %) pour permettre une utilisation suffisamment fiable en élevage. Afin d'améliorer ces méthodes de détection, les variabilités inter- et intra- individuelles devraient être mieux prises en compte en établissant des courbes témoins de référence propres à chaque animal afin de pouvoir les comparer aux courbes dynamiques en temps réel.

REMERCIEMENTS

Les auteurs remercient RENNES METROPOLE pour l'allocation d'installation scientifique (dispositif AIS 2020) qui a permis l'achat d'abreuvoirs automatiques pour les truies gestantes.

REFERENCES BIBLIOGRAPHIQUES

- Kojadinovic, I., Kojadinovic M.I., 2020. Package 'npccp'. *Econometric Reviews*, 23, 53-70.
- Liaw A., Wiener M., 2002. Classification and Regression by Random Forest. *R News*, 2, 18-22.
- Ramsay J.O., Hooker G., Graves S., 2010. *Functional Data Analysis with R and Matlab*, Springer, New York, p. 1312-1313.
- R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.