



Validating kfino algorithm (*Kalman filter with impulse noised outliers*) to filter liveweight outliers produced by the walk-over-weighing (WoW) platform in a large spectrum of farming systems

E. González-García^{a,1,*}, I. Sanchez^{b,1}, I. Llach^a, M. Decandia^c, V. Giovanetti^c, B. Cloez^{b,1}

^a SELMET, INRAE, CIRAD, L'Institut Agro Montpellier SupAgro, Univ Montpellier, 34060 Montpellier, France

^b MISTEA, INRAE, L'Institut Agro Montpellier SupAgro, Univ Montpellier, 34060 Montpellier, France

^c AGRIS Sardegna, Loc. Bonassai S.S. 291 Sassari-Fertilia, Km. 18,600 Sassari, Italy

ARTICLE INFO

Keywords:

Liveweight

Livestock

Automatic weighing platform

Outliers' detection and filtering

Web application

ABSTRACT

The liveweight (LW) is conventionally measured using static systems, which require animals to be individually walked onto a set of scales. This process is time consuming, labour intensive, and places stress on both the animals being weighed and the operator; indoors it is relatively easy, but weighing animals outdoors may be a difficult task. To overcome such situation, we developed some years ago a Walk-over-Weighing (WoW) platform which records LW in an automated, non-invasive manner, and have been evaluated with success in a range of conditions. However, a lot of outliers are produced with the WoW, which must be removed to retain only correct data and make coherent interpretations. Standard methods used until now to perform such outliers' removal still impractical, time consuming and requires a minimum of mastery by the user. To improve performance of the process we further built an algorithm, so-called **Kfino** (*Kalman Filter with Impulse Noised Outliers*) that automatically remove individual daily LW outliers. Then, to vehicle Kfino and for the ease of end-users, we further developed a web application with R-Shiny, so-called **ORIOLE**. Thus, the objective of this work was to validate the functioning in the field of the whole infrastructure (WoW, Kfino and ORIOLE) in a large spectrum of farming systems for sheep production in Mediterranean settings. A series of trials ($n = 8$) were conducted at different moments and under different conditions of the experimental farms *La Fage* and *Le Merle* (France) and AGRIS (Italy). Animals ($n = 920$) were either young (ewe-lambs) or adults (ewes) from different breeds (two for meat - *Romane* and *Mérinos d'Arles*; and two for dairy - *Lacaune* and *Sarda*) and were reared indoors or outdoors (grazing) during several weeks. For each trial, LW outliers generated in raw datasets produced by the WoW were filtered by using a conventional three-steps method (Manual), and the automatic alternative (ORIOLE) proposed here (using Kfino in the ORIOLE interface). Such methods were compared for the parameters i) number of outliers detected, ii) percentage of clean data with respect to the raw dataset, and iii) the time required for running the full filtering and report process. The ability of kfino algorithm was demonstrated and its applicability, through the use of the ORIOLE web app was validated. The final outcomes of the detection and removal of individual LW outliers from the WoW was closely similar with the two methods. The major difference is the practical ease and the time required for the full process (i.e., just minutes using ORIOLE vs. hours with manual). The ORIOLE app, which technical and computational visual improvements still in progress, may be used by a large spectrum of end-users, and provide further and interesting outputs of easy interpretation (including graphical and statistical reports). The whole framework developed in the scope of this work facilitates the development of future early warning systems that could contribute to more efficient monitoring and management of the progression of daily, individual LW of animals, and the related animal health parameters and welfare issues in a large spectrum of conditions.

* Corresponding author.

E-mail addresses: eliel.gonzalez-garcia@inrae.fr (E. González-García), isabelle.sanchez@inrae.fr (I. Sanchez), bertrand.cloez@inrae.fr (B. Cloez).

¹ These authors contributed equally

1. Introduction

Monitoring liveweight (LW) is among the most conventional and critical practices used for management purposes (e.g., assessing weight gain, body condition, animals' health and nutritional status, responses to feeding programs, or just for setting slaughtering schedules). The LW is conventionally measured using static weighing systems, which require animals to be individually walked onto a set of scales and a measurement recorded when the system comes to equilibrium. The weighing process is time consuming, labour intensive, and places stress on both the animals being weighed and the operator. Furthermore, measuring LW indoors is relatively easy, but outdoors may be a difficult task. This is why the reality in field conditions is that flocks are weighed very infrequently and almost never under commercial conditions. To overcome this situation, we produced several years ago a Walk-over-Weighing (WoW) platform for sheep, which has been evaluated with success under different and contrasting conditions [1–5].

However, despite several technological progresses and adjustments (see prototype 2, Fig. 1-B), some factors linked to the gregarious instinct

of small ruminants or the level of adaptation of individuals, produces relatively high number of outliers. These are non-plausible, erroneous, LW records which are registered due to misbehaviours e.g., when the animals are crossing the WoW platform or just because more than one individual enter at once (and sometimes stay). Such outliers require to be removed from the original raw dataset to allow usefulness of data and further sound interpretations. Standard methods used until now to perform such outliers' removal [6–8] remains unpractical and time consuming. They further require a minimum of mastery of the methods used, thus limiting the WoW adoption by farmers and other end-users.

Thus, to improve performance of the process we further built an algorithm, so-called **Kfino** (*Kalman Filter with Impulse Noised Outliers*; [9, 10]) that automatically remove individual daily LW outliers. Then, to vehicle Kfino and make its use easy for end-users, we further developed an original web application with R-Shiny, so-called **ORIOLE** [11,12].

The objective of this work was to validate the applicability of the whole infrastructure (the WoW, the Kfino algorithm and the ORIOLE web application), for the automatic detection and removal of individual liveweights outliers, largely present in the primary raw datasets

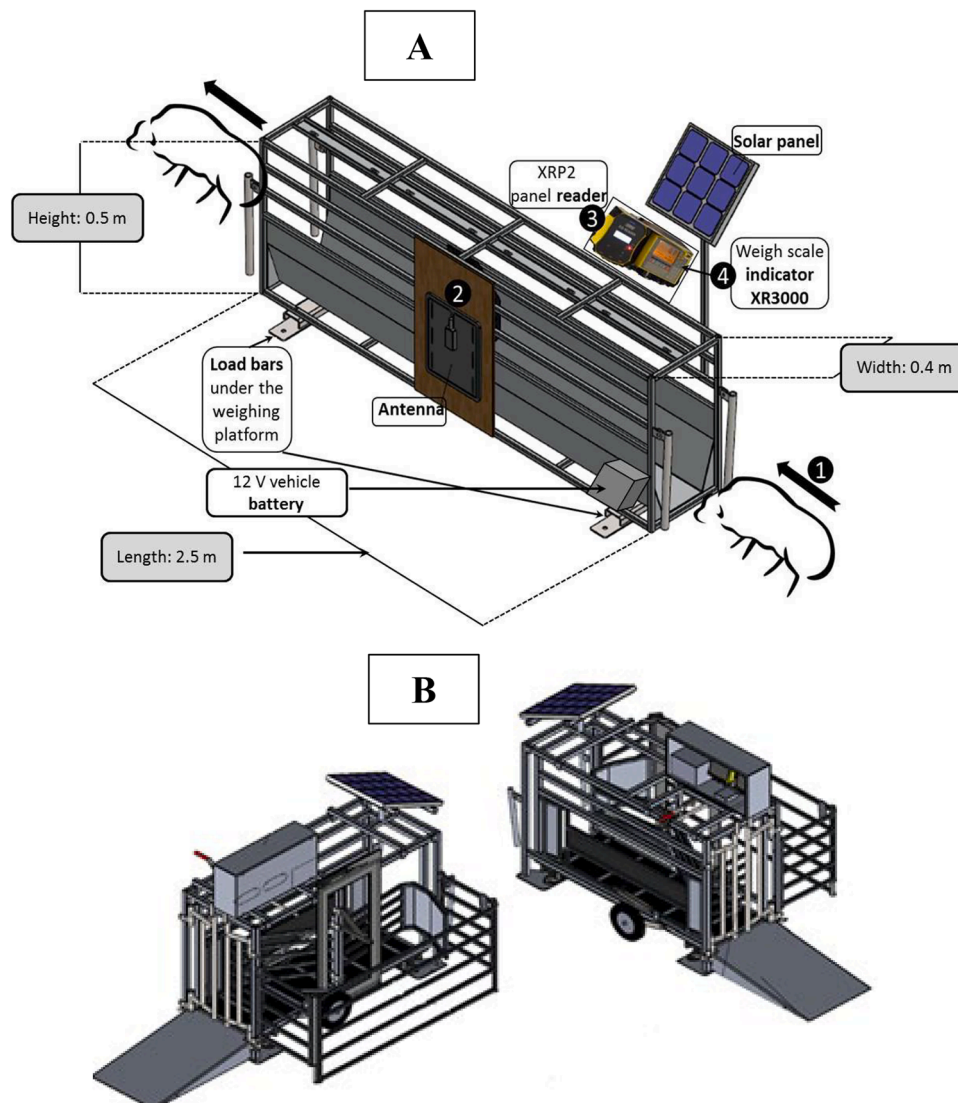


Fig. 1. A: The Walk-over-Weighing (WoW) technology. 1) The animal traverse, freely and voluntarily, the weighing platform; 2) the antenna read the animal' EID tag and send it to the XRP2 panel reader which record it into a session file and 2) sends it directly to the weigh scale indicator XR3000 to record individual animal weights and other information e.g., date, time of weighing event. Thereafter the operator downloads the stored files to the personal computer for further processing and interpretation (Adapted from González-García et al. [2]). B: The second generation of the initial prototype, improved for better mobility and compactness. (Adapted from González-García et al. [2] and [5]).

generated by the WoW. The effectiveness, and the feasibility of the practical application of the whole framework, was evaluated in a large spectrum of farming systems (indoor, outdoor), animal models (young, adult) and sheep breeds. The proposed automatic method was compared with a conventional three-step manual method.

2. Materials and methods

2.1. The walk-over-weighing (WoW) platform, the outliers

The Fig. 1 presents the WoW platform evaluated in our projects. The overall principle of functioning was reported by González-García et al. [2]. Firstly, the animal traverse, freely and voluntarily, through the weighing platform (placed in an obligatory path) attracted by some elements placed in the other side (e.g., water, mineral salts, trees' shade). Once the animal is on the platform, the antenna read the electronic identifier tag (left ear) and send it to the reader (XRP2 panel) that get the record (i.e., which animal is crossing, on which date and at what time) into a session file. This information is then sent to the weigh scale indicator (XR3000 or WOW2 in the new prototype) to record a new and the most important information: the individual animal LW at that precise time. Thereafter the operator downloads the stored files to the personal computer for further processing and interpretation [2].

2.2. Experiments with the Wow under a large spectrum of conditions: datasets produced

A series of trials have been carried out ($n = 8$) since 2017 using and evaluating the WoW under different, contrasting farming system conditions for sheep, either in France and beyond (e.g., Italy). Such datasets were available for this study and their overall characteristics are presented in Table 1 [indoor or outdoor; intensive or extensive grazing, sheep breeds (*Romane*, *Lacaune*, *Sarda*, *Mérinos d'Arles*), productive purpose of the system (meat, dairy) or physiological stages of the individuals involved on each trial (growth, pregnancy, maintenance, lactation)]. Furthermore, Table 1 presents the length of each experimental period, the number of animals involved (at the start and end of the trial), the number of LW records produced by the WoW on the raw datasets, as well as the number of outliers detected manually, using the conventional three-step procedure, as described below. Insights on most of these trials have been previously reported [1–5].

2.3. The three-step procedure for manual removal of outliers

The Fig. 2 illustrates the three-step data filtering approach followed to detect and remove (manually) misbehaviours and outliers. To do it, we have been using indistinctively, statistical models with R (using the *lme4*; [13]), SAS software, the Grubbs's Outlier Test [14] available as

part of NCSS Comprehensive Statistics Software (NCSS 12; <https://www.ncss.com/software/ncss/>) or just dynamics tables in Excel.

Firstly, records from the raw datasets were removed if they did not capture the RFID identity of the individual, or if the registered LW was equal to zero. The first and second filtering steps are then performed at the **group level**. Records falling outside the LW range of the group (i.e., minimum and maximum) are classified as misbehaviours and removed by detecting extremely low or extremely high values, or data higher than twice the LW mean of the group (e.g., meaning that more than one animal was on the platform at the same time). Then, the second filtering step is run. All data falling outside the interval [group minimum of LW – 2.5 kg; group maximum of LW + 2.5 kg] are removed (Fig. 2). The ± 2.5 kg aimed to take into account LW fluctuations during the day, as a volume of approximately 2.5 kg is considered to be close to gut-fill fluctuations in the digestive and urine tracts contents [15]. Gut capacity here interpreted in function of the digestive physiology of the animal, the intake capacity, the diet digestibility, and the mean retention time of different proportion of particle sizes [16]. Finally, a third filtering step was carried out at the **individual level**. The daily estimated LW of each individual, used as a reference value, is assessed (e.g., after calculating ADG from the available LW records obtained for the period with the Gold Standard (GS) static scales measurements). Then, all values falling out of the individually accepted range (i.e., daily estimated LW based on $GS \pm 2 \times SD$ of the individual trajectory during the concerned timespan) are removed from the dataset. At the end of this three-step data filtering approach, the result was a cleaned database able to be further processed and interpreted.

2.4. Developing a robust sequential algorithm for automatic outliers' filtering: KFINO

To automatize such dataset filtering process, a new and alternative model for detecting and removing impulse noise outliers was developed by our team [9,10]. The model is based on simple latent linear Gaussian processes, as in Kalman filter [17]. The result is a fast forward-backward algorithm to both filter and smooth sequential data and detect these outliers.

All details concerning the Kfino algorithm conception and construction were reported by Cloez et al. [9]. Briefly, given a few physiological parameters (probability of being wrongly weighed, growth rate, mass at adulthood, etc.), Kfino consists of a modified Kalman filter [18,17] that takes into account outliers of the impulse noise type (independent of dynamics). A coupling with the expectation–maximization (EM) algorithm is used to estimate these physiological parameters using explicit formulas. It applies for linear Gaussian dynamical system (X_t) and observations Y_t that are distributed as a mixture of Gaussian noised value and independent impulse-noised outliers. We show that the law of the hidden state conditionally to the observation is distributed as

Table 1

Datasets produced under different farming systems conditions, where the Walk-over-Weighing (WoW) technology has been tested.

Year	Trial / Dataset	Animal model				Farming system		Trial period (days)	Raw dataset (n)	Number of outliers (manually identified)
		Breed	Category	n_i	n_f	Indoor	Outdoor			
2017	OutSprRom17	<i>Romane</i>	Ewe-lambs	15	15	–	×	34	1966	1134
2017	OutWinRom	<i>Romane</i>	Ewes	98	98	–	×	22	3051	1627
2018	OutSumRom	<i>Romane</i>	Ewes	220	161	–	×	28	5816	3029
2019	IndLacaune	<i>Lacaune</i>	Ewes	93	93	×	–	45	5445	3987
2021	OutSprMer	<i>Mérinos d'Arles</i>	Ewe-lambs	100	99	–	×	73	40,577	30,753
2021	OutSprRom21	<i>Romane</i>	Ewe-lambs	79	79	–	×	121	17,917	4149
2022	IndSard	<i>Sarda</i>	Ewes	36	36	×	–	47	3419	417
2023	OutPregRom	<i>Romane</i>	Ewes	279	279	–	×	117	43,507	26,910

n_i = number of individuals at the start of the experiment; n_f = number individuals at the end of the experiment; n = number of liveweight values registered in the primary (raw) dataset produced by the WoW.

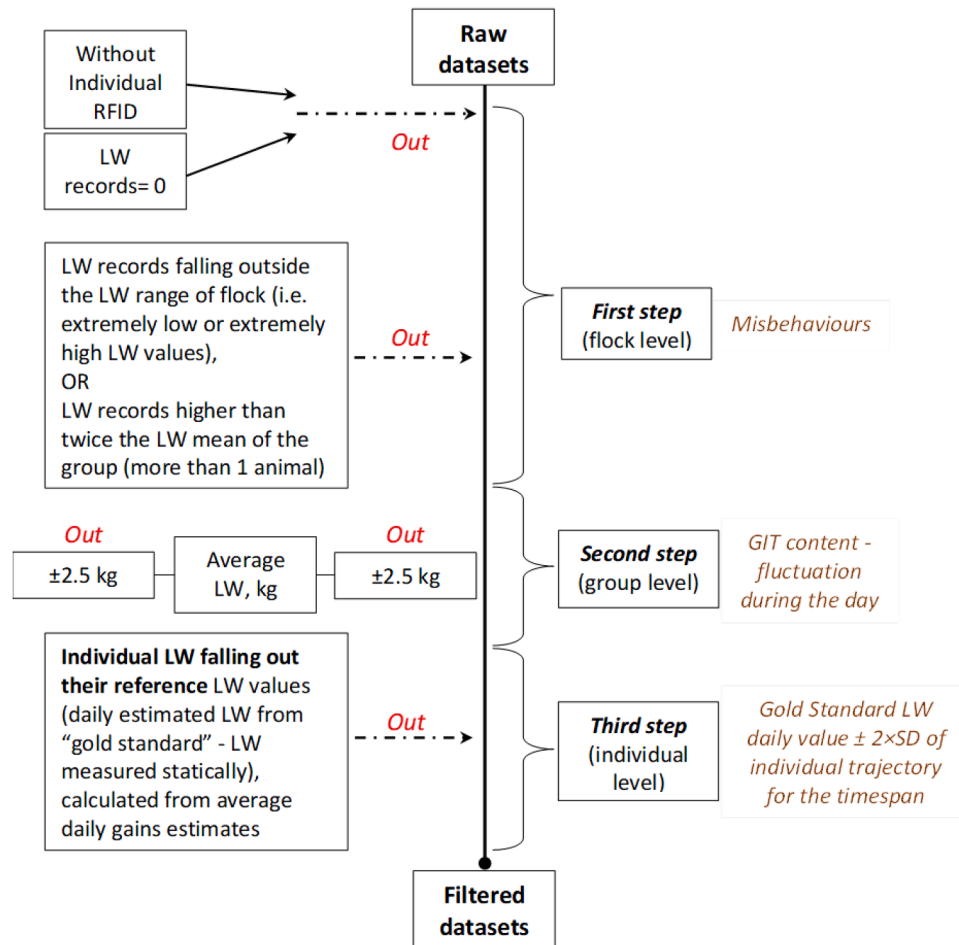


Fig. 2. The three-steps manual method for filtering outliers produced by the Walk-over-Weighing (WoW).

a mixture of Gaussian laws. Consequently, we can easily affect a live-weight for each data and estimate the hidden state by simple recursive argument (without any optimization procedure) based on a law conjugation argument and Kalman filter recursive equation, enabling to have simple explicit formulas. No optimization step is performed in this algorithm, enabling us to scale up and to process several hundred datasets for a hundred animals.

One of the main practical advantages of using this algorithm is the independency from static, GS LW measurements, the individual LW of the animal becoming its own control (i.e., Kfino detects erroneous LW measurements according to the expected LW progression based on daily dynamics in a given timespan). The model allows both to identify outliers and to reconstruct the longitudinal trajectory. Longitudinal LW data are considered, assuming similar time-correlations through a Gaussian Markov model $(X_k)_k \geq 1$.

The robustness and efficiency of this algorithm (Kfino) was then compared with the previously introduced three-steps method, and validated using the real-data produced in the different WoW trials (Table 1).

The kfino algorithm was implemented in an R package available on different platforms, CRAN, *gitlab* and can be downloaded and used by users across the world [10].

2.5. Developing an end-user friendly web application: ORIOLE

Once the Kfino algorithm was settled, a web-application was developed by our team, so-called ORIOLE (for **O**utlierRs detect**I**On **w**aLk **w**eighing; <https://oriole.sk8.inrae.fr/>; [11,12]). The Shiny library of the R software which enables to easily create user-friendly interactive web

apps straight from R, was used (<https://shiny.rstudio.com/>). The web application allows users to import raw data measured from the WoW and through simple settings to perform outlier detection and LW prediction during a given timespan. Descriptive statistics are then available such as number of daily weighing, LW progression per animal, or for the whole flock, 24 h LW kinetics of each individual. The web application (Fig. 3) is a dashboard composed of a menu of several subsets offering a user-friendly experience: a) a 'Welcome' section; b) the 'Genesis' of the technology and the web-app project; c) the heart of the app with a section for the import and analysis of 'WoW data' and producing useful reports; d) a 'How to' section documenting how to use the app. Users can analyse their data using full advantage of descriptive and statistics plots and download reports for communication and decision makings. They don't need to be familiar with the mathematics of the algorithm, enabling them to process large datasets in an easily and friendly manner and in short time (e.g., up to 200 animals in <4 min). A didactic video explaining all details for the correct use of ORIOLE was recently reported [19].

The web application was deployed using the SK8 service, an intern pipeline infrastructure from INRAE (<https://sk8.inrae.fr/>; [20]). Further technical, visual and computational improvements still in progress.

3. Results and discussion

3.1. Differences in the number of outliers filtered and number of individuals retained: manual vs. ORIOLE method

The Table 2 presents the main results of the comparison between the manual and the automatic ORIOLE web app methods, for cleaning the

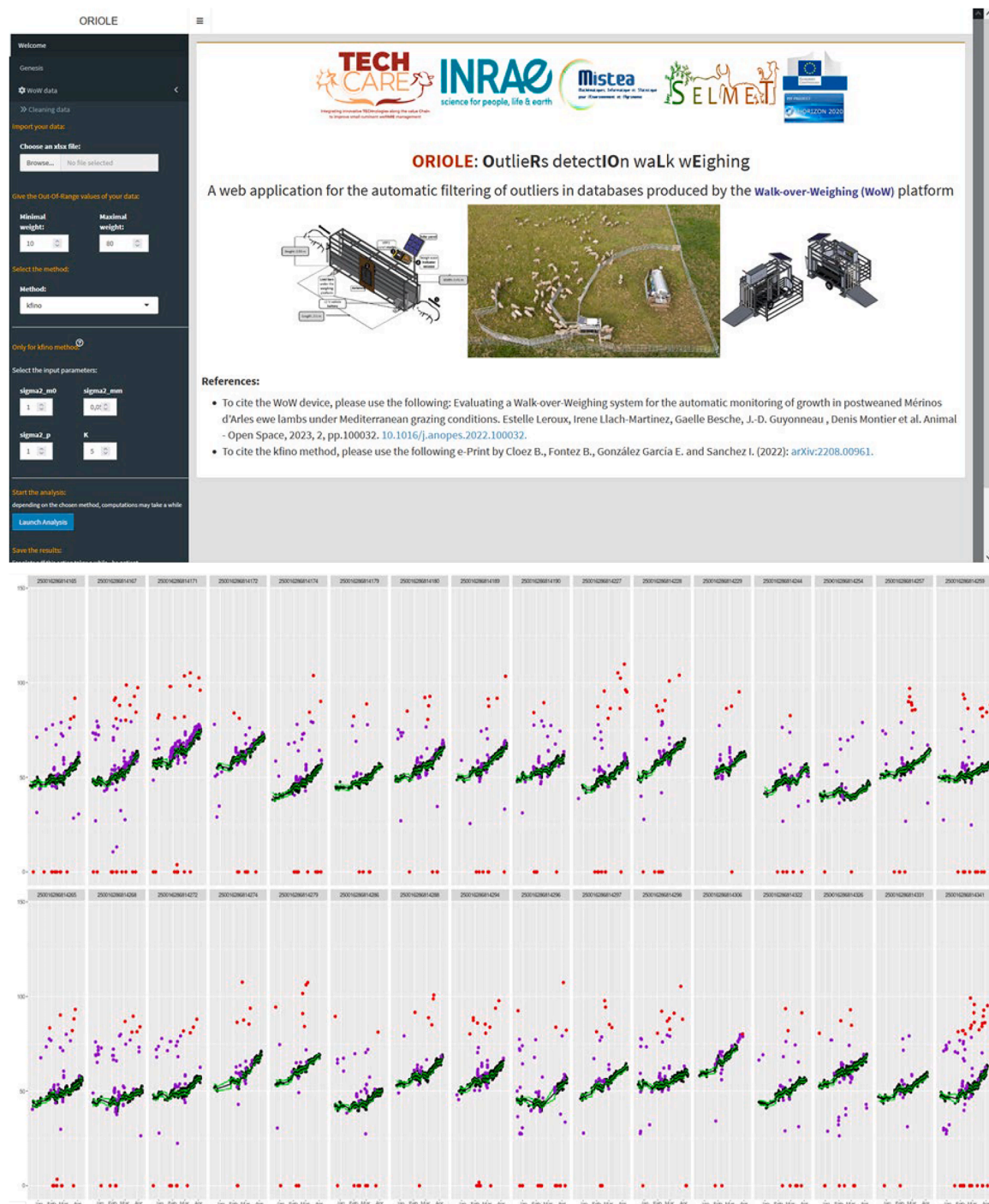


Fig. 3. At the top, a screenshot of the freely available ORIOLE (OutlierRs detectIOOn waLk wEighing) web application, for the automatic filtering of outliers from raw datasets produced by the WoW platform (<https://oriole.sk8.inrae.fr/>). At the bottom, example of graphical outcomes when using ORIOLE in routine. Constant ewes' LW progress is easily followed at individual level enabling to alert on anomalies e.g., drastic LW changes, abortions, absence in the flock.

outlier records from the eight original raw datasets produced and evaluated in this study.

In the rangeland of La Fage, where the Romane meat sheep flock is fully reared outdoors all round year, the proportion of outliers produced represented more than half of the total number of records collected in the first years' trials (i.e., from 40 to 48 % of clean data in OUT-SprRom17, OutWinRom and OutSumRom datasets). In these years (2017, 2018) the ORIOLE method removed more outliers compared to the manual i.e., 59.6 %, 57.0 % and 59.7 % versus 57.7 %, 53.3 % and 52.1 % for those first three outdoors experiments, respectively. The

reason behind ORIOLE removing more than manual method is because of a prerequisite statement programmed by the web application i.e., a minimum threshold of available data is required in the dataset to be able of building the individual LW trajectory in a determined timespan. Therefore, individuals with few LW records (i.e., few daily voluntary passages through the WoW platform), lower than the minimum required threshold, will be automatically removed by ORIOLE, thus not taken into account for further interpretations. On the other hand, the three-steps manual method analyses all daily readings for each animal, regardless of the number of observations, as it compares the LW records

Table 2

Filtering outliers from different datasets: comparison of results using the automatic method (i.e., KFINO algorithm through the ORIOLE web app) versus the three-step manual method.

Trial / Dataset	Raw dataset (n)	Manual method (Three-steps)						Automatic method (ORIOLE App)					
		Dataset			Out-of-Range		Time, min.	Dataset			Out-of-Range		Time, min.
		Outliers filtered (with regard to raw dataset; n)	Clean data (% of raw dataset)	n_e	Minimal LW, kg	Maximal LW, kg		Outliers filtered (with regard to raw dataset; n)	Clean data (% of raw dataset)	n_e	Minimal LW, kg	Minimal LW, kg	
OutSprRom17	1966	1134	42.3	15	34.6	53.5	125	1172	40.4	13	35	55	1
OutWinRom	3051	1627	46.7	84	38.5	78.2	105	1740	43.0	53	30	80	1
OutSumRom	5816	3029	47.9	192	32.0	79.4	95	3473	40.3	167	30	80	2
IndLacaune	5445	3987	26.8	93	68.9	95.3	170	3947	27.5	93	55	101	6
OutSprMer	40,577	30,753	24.2	100	10.6	44.9	225	30,711	24.3	100	15	50	9
OutSprRom21	17,917	4149	76.8	79	21.8	53.5	185	3886	78.3	78	23	55	2
IndSard	3419	417	87.8	36	36.3	70.1	150	731	78.6	36	25	92	1
OutPregRom	43,507	26,910	38.2	219	37.8	82.3	120	19,957	54.1	172	34	110	7

n = number of liveweight values registered in the primary (raw) dataset produced by the WoW; n_e = number of outliers filtered; n_e = number of individuals effectively considered for data processing during the full experimental period.

obtained by the WoW with the static GS reference method, even if it is in a low number.

The last is explicit in the number of individuals effectively considered (retained) for data processing on each trial (i.e., n_e parameter; Table 2). The higher proportion of animals removed for further interpretation by ORIOLE is likely related to their degree of acclimatation (training), determining *a posteriori* the number of voluntary passages through the WoW platform. We observed that this is also related to the environmental conditions, as it is illustrated by the lower number of ewes retained by ORIOLE during the winter trial outdoors (i.e., 53 vs. 84 individuals retained by ORIOLE vs manual in OutWinRom, respectively; Table 2).

Despite the above-mentioned conditions of ORIOLE, in spring and summer (OutSprRom17 and OutSumRom) animals visited much more the platform to drink water, then they got more trained, more daily LW records were obtained which, at the end, increased the probability to be retained by ORIOLE. In synthesis, more the animal is trained, more it will cross the platform thus making most probable to be considered by the automatic web app method.

In this farming system (rangeland of La Fage), the second spring trial showed excellent performances, with no differences between methods in the percentage of clean data (76.8 and 78.3 in manual and ORIOLE, respectively) and the number of animals retained for interpretation (79 and 78, respectively). In this trial, the number of LW recorded in the raw dataset was dramatically higher (17,917), even when compared to the OutSumRom trial using a considerable higher number of animals (i.e., more than twice of ewes; >190 vs. 79). This is likely related to the fact that animals in the flock got more trained with time. The last trial in the rangeland of La Fage (OutPregRom) used around 220 ewes from the whole flock (with an equivalent ratio of trained and naïve individuals) and lasted the full pregnancy period (i.e., from mating in autumn -November- and lambing in spring -April-). A consequent high number of LW records were collected in the full period (43,507), the highest among the eight trials. Contrarily to the other OutRom trials, in this experiment ORIOLE removed less outliers than the manual method (19,957 vs. 26,910), which represented a more consequent clean dataset (retaining 54.1 % of records from the raw dataset, vs. 38.2 % in the manual). However, similarly to previous trials, in the OutPregRom the number of ewes retained for further interpretation was lower with ORIOLE (172 vs. 219 with manual).

The other trial outdoors (OutSprMer) was carried out with growing Mérinos d'Arles ewe-lambs in spring. In this experiment, a very high proportion of outliers was generated but no differences at all were observed between the two methods (24.2 vs. 24.3 % of clean data retained in manual vs. ORIOLE respectively). The 100 individuals were

retained by both methods for further analyses and interpretation.

The results in the trials developed indoors, with dairy ewes, were contrasted. When evaluating the WoW at the exit race of the milking parlour with lactating Lacaune ewes ($n = 93$; La Fage, France) a very high proportion of outliers was generated (i.e., close to 75 % of the raw dataset). The reasons explaining this were reported by González-García et al. [4]. Briefly, they were able to collect only 20 % of the total possible records over the duration of the experiment (1458 effective readings from 7500 possible readings). The effective readings were considered plausible LW records that could be used for further database analyses and interpretations after data filtration. Even with such a significant loss of data because of misbehaviours and other sources of outliers (80 %), authors stated they were able to detect LW changes in ewes. The most important factor determining spurious values in this work was the excessive speed of ewes when crossing the WoW and the coincidence or proximity of two or more ewes on the platform. Apart from these factors, other reasons could have increased the limits of agreement between the automated and static scales: for example, the loss or gain of gut fill between the 2 milking sessions of the day and during the study, as well as other elements related to the specific design and functionality of the WoW system, likely requiring further adjustments for a better fit to the particular anatomic characteristics of the breed (larger frame in Lacaune compared to other breeds like Romane ewes) and to the specific location of the platform in the indoor system. The WoW system was placed in the corridor of the sheepfold connecting the milking parlour with the pens. This point should also be taken into account when using this WoW system indoors. A suitable place for the device should be chosen, and another type of calibration must be considered so that LW values are not recorded without the animals crossing the platform.

Having 80 % of values removed in the real-farm indoors situation of La Fage was very high compared to other studies. Alawneh et al. [6] reported that 12 % (9298) of individual LW records were outliers from a total of 79,697 available for analysis. Kedzierski [21] reported that 24 % (405 from 1624) of the initial data were classified as outliers. Brown et al. [7] showed that the percentage of missing data is related to the applied filter; the 25 % filter level removed 25 % of weight records on average, and the 10 % filter level removed 60 % of weight records.

When comparing the two methods in this work, the clean dataset retained were closely similar, as were the number of ewes retained ($n = 93$; Table 2). However, in Italy (Sardinia), evaluating the WoW with Sarda ewes ($n = 36$) indoors during 47 days produced the lowest number of outliers, which represent to retain around 80 % or more of original LW records from the raw datasets. Differences between the two methods were observed (Table 2) in terms of clean data retained (87.8 % vs. 78.6 % in manual vs. ORIOLE) but not in the number of individuals retained

for interpretation (36 in both).

3.2. Validating the automatic outliers' removal method, using Kfino through ORIOLE

Overall, outcomes related to the number of outliers detected and filtered, the clean datasets produced after applying the Kfino algorithm, and the number of individuals retained for the full processing and interpretation of LW progression during the given timespan, may be considered similar or at least rather close when comparing the two methods.

Despite producing more outliers indoors, less differences between the two methods are observed under these conditions. In contrast, flocks reared outdoors are more exposed to environmental constraints in terms of climatic hazards and more variability of the daily management practices under these conditions. This induces much more heterogeneity in the individual behaviours of the animals, their adaptation to cross voluntarily through the WoW platform, and therefore, the number of individual LW records in a given similar timespan. The worse registers are observed during winter outdoors.

We observed a significant, expected, effect of the age on the animals to get trained with the WoW infrastructure. Until they got actually acclimatized and trained, growing ewe-lambs visited less the platform and then produced less plausible LW records than adult ewes.

The big difference between the two methods however was the time required (speed) for processing the datasets and running the full process. As expected, the automatic method was significantly less time and

energy consumer than manual. Furthermore, ORIOLE require just average users' skills, therefore it is addressed to a large spectrum of end-users no matter what their computer or statistics skills are. The web application is easy to use for processing data sets, and provides a wide range of additional figures and statistical information in a user-friendly way. It may be considered also as trustable for the outcomes it furnishes, and robust in the procedures implied.

Therefore, overall results obtained in this work, in a large spectrum of farming systems, encourage us at a glance to consider as validated the use of Kfino algorithm through the ORIOLE web app. It may be considered as an alternative compared to other statistical methods previously used to filter outliers from datasets in cattle, where local regression methods have been used (e.g., [8]) with pre-processing step for removing extreme values then data are fitted to β -splines penalised on the coefficients. González et al. [8], for example, proposed analytical methods for processing remotely collected LW and behaviour of beef cattle grazing tropical pastures in Australia. They fitted the obtained LW data to B-splines penalised on the coefficients [22] for each individual animal with the smoothing parameter selected having the lowest Schwarz Bayesian criterion. Data points below or above 1.5 times the residuals for each animal were deleted and the penalised B-spline fitted again to obtain the predicted LW.

Finding outliers, defined as "observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" [23] (i.e., data objects that do not fit well to the general data distribution) is very important in many practical applications [24]. In data from sheep (specie characterised by the gregarious behaviour) the number of

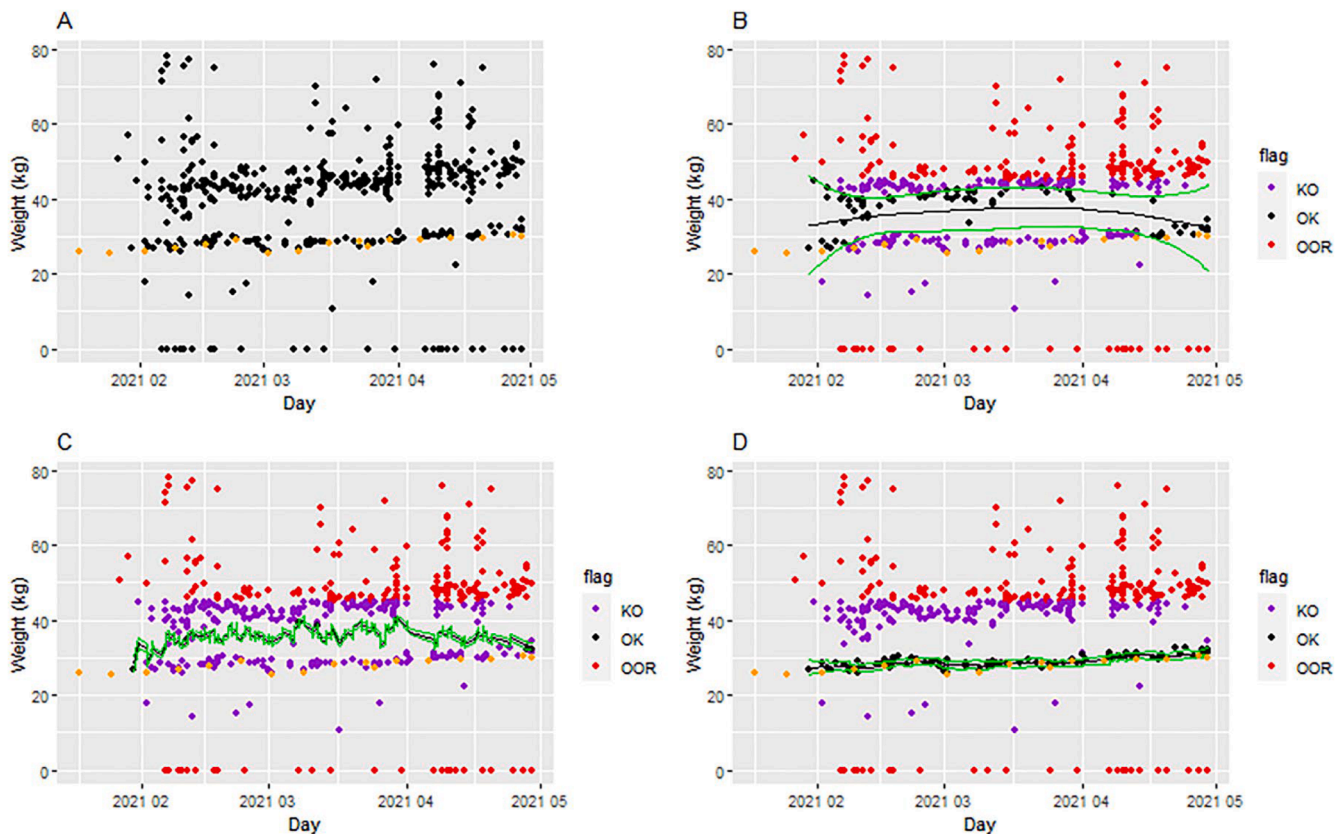


Fig. 4. Applying KFINO by using ORIOLE. Four graphics represent the data analysis of liveweight (LW) measurements for the same individual. Time and LW are presented on the X and Y axes, respectively. Raw data are shown at the top left (A), for which each black dot represents a LW point, measured by the WoW. The golden dots represent LW measurements obtained with the gold standard method (i.e., static scales), provided for comparisons but not used in the algorithm. Red dots appearing in the other figures blocks represents LW data classified as out of range. Purple dots are the outliers detected by the algorithm. Black dots are LW data classified as biologically plausible. The black line is an estimator of the hidden state and the green lines give a confidence interval (given by $\mu \pm 2\sigma$ with σ being the estimated square root). The top right block (B), shows the outcome of detecting outliers by using local regression (through the *locfit* package of R software). The bottom left (C) shows results when using the classical Kalman filter method, and the bottom right (D) outcomes with the Kfino algorithm (Adapted from [9]).

outliers may be considerably higher when compared to cattle, then local regression methods may not be robust enough. Cloez et al. [9,10] demonstrated this, highlighting the advantages of using Kfino algorithm in comparison to other methods like local regression (*locfit* package of R; Fig. 4). The Kfino algorithm allow to treat, simultaneously, impulse and additive outliers and is adapted for monitoring at random times. As reported by Cloez et al. [9], Kfino is close to the original switching Kalman filter (SKF), introduced in 1998 by Murphy [17], a stochastic process with an efficient and widely used algorithm for estimating the state of a dynamic system, given noisy observed data. Similarly, to the previously introduced references, SKF aims to perform a robust Kalman filter algorithm by considering a Gaussian mixture. This enables consideration of different variances at each step. With this alternative modelling, they also obtain a Gaussian mixture sequence with computations similar to our case. SKF has since been widely used in different contexts and has therefore proven its effectiveness in terms of precision and speed of calculation time. The Kfino algorithm possesses the advantages of SKF (simplicity, robustness, rapidity...) even though it differs in several point. Indeed, the main difference comes from the fact that Kfino enables consideration of any non-Gaussian outlier. This particularity allows treatment of new types of outliers (so-called here as impulse outliers) without artificially considering them as normal variables with very large variances (which are additive outliers). This modification therefore also allows us to do outlier detection in addition to the filtering. The possibility of changing the law of observations, without increasing the computation time, thus seriously extends the possible applications of Kfino. As for the SKF algorithm, the number of computations in Kfino increases exponentially.

Simple and relatively inexpensive computationally, Kfino applies for linear Gaussian dynamic system and observations are distributed as a mixture of Gaussian noised value and independent impulse-noised outliers.

4. Conclusions

The applicability of KFINO algorithm, through the use of the ORIOLE web app was validated in a large spectrum of farming systems (indoor, outdoor), animal models (young, adult) and sheep breeds. The final outcomes of the automatic detection and removal of individual liveweights outliers from primary raw datasets generated by the WoW was similar to the manual method, but requiring just minutes for the full process. The ORIOLE is easy to use by a large spectrum of end-users, and provide further and interesting outputs of easy interpretation (including graphical and statistical reports). The whole framework developed in the scope of this work (i.e., the in-field use of WoW platform, the KFINO algorithm and the ORIOLE web app). The overall infrastructure facilitates the development of future early warning systems that contribute to more efficient monitoring and management of the progression of daily, individual liveweights, and the related animal health parameters and welfare issues in different conditions.

CRedit authorship contribution statement

E. González-García: Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **I. Sanchez:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **I. Llach:** Writing – review & editing, Validation, Investigation, Data curation. **M. Decandia:** Writing – review & editing, Investigation. **V. Giovanetti:** Writing – review & editing, Investigation. **B. Cloez:** Writing – review & editing, Validation, Software, Methodology, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they do not have any conflict of interest,

competing financial reports or personal relationships that could affect the work reported in this paper.

Acknowledgements

The authors wish to thank the staff of the different experimental farms on which the different trials were carried out (i.e., *Domaine de La Fage* and *Domaine du Merle* in France; experimental farm of AGRIS in Sardinia, Italy). This work has been financially supported by the EU project TechCare, which received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 862050.

Ethical Statement

Validating kfino algorithm (Kalman Filter with Impulse Noised Outliers), through the ORIOLE web application, to filter individual liveweight outliers produced with the Walk-over-Weighing (WoW) technology in a spectrum of ruminants' systems

The works detailed in this study were conducted in full compliance with relevant legislation governing research involving animal subjects, as outlined by the European Union Council Directive (2010/63/EU). The researchers involved in the experiments were certified by the appropriate French governmental authority, as well as by the INRAE *La Fage* Experimental Farm (agreement number A-312031) and the AGRIS institution in Sardegna (Italy).

Data availability

Data will be made available on request.

References

- [1] M. Decandia, M. Acciaro, V. Giovanetti, G. Molle, F. Chessa, I. Llach, E. González-García, Monitoring liveweight in Sarda dairy sheep using a walk-over-weighing system, in: Book of Abstracts, 74rd Annual Meeting of European Federation of Animal Science, August 26th-September 1st, Lyon, France, 2023.
- [2] E. González-García, M. Alhamada, J. Pradel, S. Douls, S. Parisot, F. Bocquier, Menasol, I. Llach, L.A. González, A mobile and automated walk-over-weighing system for a close and remote monitoring of liveweight in sheep, *Comput. Electron. Agric.* 153 (2018) 226–238, <https://doi.org/10.1016/j.compag.2018.08.022>.
- [3] E. González-García, P. de Oliveira Golini, P. Hassoun, F. Bocquier, D. Hazard, L. A. González, A.B. Ingham, G.J. Bishop-Hurley, P.L. Greenwood, An assessment of walk-over-weighing to estimate short-term individual forage intake in sheep, *Animal* 12 (6) (2018) 1174–1181, <https://doi.org/10.1017/S1751731117002609>.
- [4] E. González-García, M. Alhamada, H. Nascimento, D. Portes, G. Bonnafe, C. Allain, I. Llach, P. Hassoun, J.M. Gautier, S. Parisot, Measuring liveweight changes in lactating dairy ewes by using an automated walk-over-weighing system, *J. Dairy Sci.* 104 (2021) 5675–5688, <https://doi.org/10.3168/jds.2020-19075>.
- [5] E. Leroux, I. Llach, G. Besche, J.D. Guyonneau, D. Montier, P.M. Bouquet, I. Sanchez, E. González-García, Evaluating a walk-over-weighing system for the automatic monitoring of growth in post-weaned Mérinos d'Arles ewelambs under Mediterranean grazing conditions, *Anim. -Open Space* 2 (2023) 100032, <https://doi.org/10.1016/j.anopes.2022.100032>.
- [6] J.I. Alawneh, M.A. Stevenson, N.B. Williamson, N. Lopez-Villalobos, T. Otley, Automatic recording of daily walkover liveweight of dairy cattle at pasture in the first 100 days in milk, *J. Dairy Sci.* 94 (2011) 4431–4440, <https://doi.org/10.3168/jds.2010-4002>.
- [7] D.J. Brown, D.B. Savage, G.N. Hinch, S.J. Semple, Mob-based walk-over weights: similar to the average of individual static weights? *Anim. Prod. Sci.* 52 (2012) 613–618, <https://doi.org/10.1071/AN11306>.
- [8] L.A. González, G. Bishop-Hurley, D. Henry, E. Charmley, Wireless sensor networks to study, monitor and manage cattle in grazing systems, *Anim. Prod. Sci.* 54 (10) (2014) 1687–1693.
- [9] B. Cloez, B. Fontes, E. González-García, I. Sanchez, Kalman filter with impulse noised outliers: a robust sequential algorithm to filter data with a large number of outliers, *Int. J. Biostat.* 20 (2) (2024) 641–660, <https://doi.org/10.1515/ijb-2023-0065>.
- [10] Sanchez, I., Cloez, B., Fontez, B., 2024a. kfino: kalman filter for impulse noised outliers., swb:1:dir:728011713fe724b0f29203da74bc01ac5f8d530f;origin=https://hal.archives-ouvertes.fr/hal-04830668;visit=swb:1:snp:f129608714c2acc104ea5a3bfb3243115853a97;anchor=swb:1:rel:e62f5a0feb09d34c3cfbc777efb9908632ba0ea0;path=/ hal-04830668.
- [11] I. Sanchez, E. González-García, B. Fontez, B. Cloez, ORIOLE: a web application for cleaning data from the walk-over-weighing device in livestock systems, in: Book of

- Abstracts, 74rd Annual Meeting of European Federation of Animal Science, August 26th- September 1st, Lyon, France, 2023.
- [12] Sanchez, I., Cloez, B., 2024b ORIOLE: a web application for the automatic filtering of outliers in databases produced by the walk-over-weighing (WoW) platform - TechCare project - v1.1.0. 2024, swb:1.dir: 235c8edc0c420d831a4f4f2cd3d3cf9d552b9e7b;origin=<https://hal.archives-ouvertes.fr/hal-04835991>;visit=swb:1:snp:95f0c856c382dab0b118249353a15c89baa36a66;anchor=swb:1:rel:a35928dff80c9195c4e136f247fab64c53bf15a0;path=/. hal-04835991.
- [13] D. Bates, M. Maechler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, *J. Stat. Softw.* 67 (2015) 1–48, <https://doi.org/10.18637/jss.v067.i01>.
- [14] F.E. Grubbs, Procedures for detecting outlying observations in samples, *Technometrics* 11 (1969) 1–21, <https://doi.org/10.1080/00401706.1969.10490657>.
- [15] R. Jarriège, Alimentation des bovins, ovins et caprins, INRA Ed. (1988) 471, 1988, 2-7380-0021-5. hal-02854912.
- [16] A. Munn, M. Stewart, M. Price, A. Peilon, T. Savage, I. Van Ekris, M. Claus, Comparison of gut fill in sheep (*Ovis aries*) measured by intake, digestibility, and digesta retention compared with measurements at harvest, *Can. J. Zool* 93 (10) (2015) 747–753, <https://doi.org/10.1139/cjz-2014-0314>.
- [17] K.P. Murphy, Switching Kalman filters, Technical Report, University of California, Berkeley, 1998.
- [18] R.J. Meinhold, N.D. Singpurwalla, Robustification of Kalman filter models, *J. Am. Stat. Assoc.* 84 (1989) 479–486.
- [19] I. Llach, Prise en main de l'application web ORIOLE : l'analyse des poids vifs des femelles romane en gestation, *Get. Start. ORIOLE Web. Appl* (2025). Video hal-05072099.
- [20] Maigné, E., Sanchez, I., Carayon, D., Tran, J., Rey, J.F., Midoux, C., Marjou, M., 2023. SK8: un service institutionnel de gestion et d'hébergement d'applications Shiny. Rencontres R 2023 et 2024, Jun 2023, Avignon (FR), France. hal-04141247v2.
- [21] P.M. Kedzierski, MS thesis, Animal Biosciences. University of Guelph, Guelph, Ontario, Canada, 2020.
- [22] P.H.C. Eilers, B.D. Marx, Flexible smoothing with B-splines and penalties, *Stat. Sci.* 11 (1996) 89–121.
- [23] V. Barnett, T. Lewis, *Outliers in Statistical Data*, 3rd ed., John Wiley&Sons, Chichester, 1994.
- [24] A. Zimek, P. Peter Filzmoser, There and back again: outlier detection between statistical reasoning and data mining algorithms, *WIREs Data Min. Knowl. Discov* 8 (2018) e1280, <https://doi.org/10.1002/widm.1280>.