Preprints.org

Review

Quantitative Methods for Real Time Auto-Monitoring of Livestock

Sarah Brocklehurst [*] , Zhou Fang , Adam Butler

Review

Quantitative Methods for Real Time Auto-Monitoring of Livestock

Sarah Brocklehurst *, Zhou Fang and Adam Butler

Biomathematics and Statistics Scotland (BioSS), Edinburgh EH9 3FD, UK

* Correspondence: Sarah.Brocklehurst@bioss.ac.uk

Abstract

The use of automated sensors has grown rapidly in recent years, with sensor data now routinely used for monitoring in a wide range of situations, including human health and behaviour, the environment, wildlife, and agriculture. There is the potential to massively increase use of empirical data for decision making in real time, but in many areas, development and validation of quantitative approaches is still needed in order to move from the developmental stage to this intended practical application. Whilst there is wide-ranging literature in this area, in livestock farming there is a paucity that provides fair robust statistical comparisons of alternative quantitative methods and evidence that resulting decision making performs adequately in practice on farms. That is, it must be practically feasible to repeatedly apply the method dynamically in real time on farms, and optimise decisions made. We discuss alternative quantitative approaches that could be used for real time decision making from automatic monitoring, which we refer to as prediction, as well as approaches to rigorous statistical evaluation of the resulting decision-making process, which we refer to as prediction validation. Associated research challenges are discussed in detail with reference made to livestock farming, but many issues discussed are widely applicable.

Keywords: sensors; prediction; decision-making; prediction validation; livestock; time series; statistical modelling; latent variable modelling; machine learning; neural networks

1. Introduction

The use of automated sensors to collect data has grown extremely rapidly in recent years. Automated sensor data are now being used routinely for monitoring a wide range of systems, including human health, welfare and behaviour, the physical environment, wildlife, and agriculture. Automated sensors have the potential to massively increase the use of empirical data in decision making. This is partly because they are able to collect very large quantities of data, often at much lower cost than more traditional methods of data collection, and partly because they can collect data at a sufficiently high temporal resolution to feed into decision making in "real time". In agriculture, numerous recent review papers discuss sensors in the areas of precision farming, smart farming, big data, and so on. Digital transformation/Smart farming for sustainable development is discussed in [1], with an overview of how new precision farming technologies assist farmers in management/decision making, from which it is clear that artificial intelligence (AI) and machine learning (ML) play a critical role. Technologies are used to measure the environment (weather, GHG emissions), crops, soil, water and animals. In precision farming collected data is used for control (e.g. irrigation, fertilizer application, weeds, pests and diseases, feeding) and prediction for decision support (e.g. weather forecasting, crop and animal yields, ..). Other reviews [2,3] examine recent trends in precision farming for both crops and livestock and contain tables showing a range of technologies along with the main objectives. Many review papers discuss artificial intelligence applications [3,4], or focus on AI/Machine Learning [5,6], or Deep Learning [7].

There are numerous recent reviews in the general area of precision livestock farming (PLF) [8–15] that discuss sensor data. Again, there are quite a few review papers that focus on machine

learning/AI [16–20]. [21] suggests that much more work is needed to get to the stage of real-time auto-monitoring of livestock on farms whilst [22] emphasises that development needs to be guided by livestock farmers' needs enhancing (not substituting), farmers' capabilities. Sensors used depend largely on the environments in which livestock are kept, with a major distinction between housed animals versus those that graze. They also depend on the practicality as well as the trade-off between cost and value of individual animals being monitored, versus animals being monitored in groups.

Because of the value, as well as high level of management and data collection for dairy cows, this area has the most extensive use of a range of sensors and research into their application with many review papers [23–26]. [26] classifies the range of sensors used as 'At Cow' (includes wearable sensors such as accelerometers and Global Positioning System (GPS) collars as well as ingested), 'Near Cow' (sensors at fixed locations, includes feeding, weighing, sensors at feeding locations, imaging, and all remote real time sensors, such as Geographic Information Systems (GIS)), and 'From Cow' (includes milk measurements for example). By far the most established use of auto-monitoring in dairy is to detect estrous and/or calving [27–35]. The most frequently used technology for this is accelerometers or pedometers, but other methods can be used, for example monitoring temperature, or 2-D image/video analyses, and there are studies showing that these methods generally work fairly well. There are also a range of methods for detecting lameness [36–45]. Objectives for which methods are less well-established include monitoring of dairy cows' or calves' health and well-being [46,47], including diagnosis of specific diseases [34,41,48–53]. Machine learning is predominant in papers [19,39,41,50,51,54–56] on using sensor data to predict events of interest, in individual dairy cows or calves.

The main distinction with precision livestock methods for beef cattle [57] used for detecting calving [58–61], or health and welfare problems [62–64], is the fact that they tend to be grazing, so sensors commonly used, often in combination, are accelerometers, Global Navigation Satellite System (GNSS) sensors and weather sensors. Sensors together with satellite imagery can be used to manage production and grazing more generally for ruminants on extensive systems [65,66]. Sensors used for sheep [66–70] are also usually intended for use out in the field. As sheep are less valuable than cows, less expensive options are required for them [70]. Objectives include detecting lambing [71–74], oestrus [75], lameness [44,45] and other illnesses [76–78], predators [79], and changes in a range of behaviour measurements [70] including behaviour relating to ewe-lamb welfare [80], and production [81].

In contrast most methods on real time monitoring of pigs have been developed for housed pigs [82–85], but much of this is still at the development stage, with more research needed to roll this out for use on farms for real time decision making particularly at the level of individual animals. The main areas of interest are to manage production and feeding [86], to predict disease [87–89], and to manage the welfare of piglets [90] including detection of farrowing [91] and nursing behaviour [92,93]. [88] reviews prediction of health indicators for a range of common health and welfare problems in piglets, suggesting that whilst wearable sensors could be used, locational (sensors at fixed locations) sensors are preferable and more practical. Areas of ongoing research include recognition and tracking of pigs [94–97], automatically measuring or estimating weights [96,98–100], estimating behaviours [96–98,101], including aggression [102–105] which is of concern especially after mixing. Much of this research is based on 2D and 3D image analyses [90,106,107] for which the advantage of locational sensing must be traded off against complexity and computational load of methods. Sound monitoring is similarly practical as it is cheap and also locational, but in general it cannot detect individuals. Sound can be used to detect pig vocalisations [108,109], such as coughing [110] which can be indicative of respiratory disease and welfare related behaviours such as screaming.

The main distinction with poultry is that automatic measurement on real farms is more likely to be on groups rather than on individual birds [111]. Commercial flock management is already automated in terms of controlling the environment by automatic monitoring of a range of measures in poultry houses such as temperature and humidity. Light schedules, layer egg collection and

provision of feed and water are also automatically controlled. But in terms of real time monitoring of actual health and welfare using sensors, methods are still at the developmental stage, with advances needed before they could be used for commercial broiler and layer flocks [112]. Methods are generally aimed at monitoring flocks, or unidentified individuals within flocks. Image and video analyses include methods for estimating weights [113], recognition and tracking of individual birds [114], detecting lameness [115,116] or sick birds [117], group activity and colocation [118], and group optical flow measured in flocks as an indicator of health problems [119]. Sound analyses includes methods for detecting distress calls [120] as an indicator of general productivity and welfare, abnormal sounds indicating respiratory disease [121] and for distinguishing eating versus non eating vocalizations [122]. Other research includes detection of feather pecking [123] and environmental sensing to indicate specific diseases [124]. RFID [125] or accelerometers [126,127] could be used with the aim of measuring behaviour on a representative subset of birds, but is unlikely to be used more extensively in commercial flocks.

This review examines quantitative approaches underlying the use of sensor data for continuous monitoring of individuals, or groups, for real time prediction of health and welfare issues. In many areas, development and validation of quantitative approaches is still needed in order to move from the developmental stage to this intended practical application [128]. Whilst there is wide-ranging literature in this area, much of it dominated by machine learning methods, in livestock farming there is a paucity that provides fair robust statistical comparisons of alternative quantitative methods and evidence that resulting decision making performs adequately in practice on commercial farms. That is, it must be practically feasible to repeatedly apply the method dynamically in real time on farms, and optimise decisions made. We discuss alternative quantitative approaches that could be used for real time decision making from automatic monitoring, which we refer to as *prediction/decision* methods, as well as approaches to rigorous statistical evaluation of the resulting decision-making process, which we refer to as *prediction/decision validation*. Associated research challenges are discussed in detail with reference made to livestock farming, but many issues discussed are relevant to other application areas. We illustrate points made by examining recent studies in various livestock species and by data simulation to aid understanding.

2. Sensors and Data Streams for Livestock Monitoring

A wide range of sensors can be used for monitoring to aid decision making in real time. Table 1a shows some commonly used examples of sensors for on-farm monitoring of livestock [13–15,24,66,70]. Sensors commonly used may depend on species, and also on whether they are housed or outside/grazing. They are generally used for managing production, nutrition, grazing, reproduction, and/or detecting health and welfare problems. Key to precision/SMART farming is that data from sensors on individuals (or groups) allow the farmer to make efficient decisions in real time to take appropriate actions such as checking and isolating or treating individual animals or groups, or changing or supplementing feed for groups.

Table 1a is not intended to be an exhaustive list of sensors available, or being developed, for on-farm use, rather it is intended to elucidate aspects pertinent to applicability of different quantitative methods. Choice of appropriate quantitative methods will depend on the overall design of the data collection scheme and various characteristics of the resulting data streams (see Table 1b) but most importantly on *who/what* is being measured *when*, and the *nature of the resulting measurements*.

Table 1. (a) Examples of commonly used sensors for real time monitoring of livestock on farms showing types of sensors, what is being measured on what species and the purposes of the monitoring.

Name	Species being Monitored[1]	What is Being Measured	Sensor Technology	Purpose of Monitoring	Comments
1 Individual intakes	Dairy Cows, Beef Cattle, Calves, *Sheep*	Individual feeding & drinking behaviour; *Amounts if technology allows*	Sensors at feeders/drinkers that record individual RFID tags. More advanced systems that also record feed or drink taken at each bout.	Managing Nutrition & Production; Detecting Health & Welfare Problems of individuals	
2 Group intakes	Dairy Cows, Beef Cattle, Calves, Pigs, Poultry, Goats, Sheep	Group feeding & drinking including amounts	Automatic livestock feeders and drinkers	Managing group Nutrition & Production; Detecting Health & Welfare Problems in groups	
3 Individual weights	Dairy Cows, Beef Cattle, *Pigs, Sheep*	Individual Live-weights	Walk over weighers that record individual RFID tags.	Managing Nutrition & Production; Detecting Health & Welfare Problems of individuals	Walk over weighers placed to maximise the number of readings (e.g. on way in/out of milking parlour)
4 Estimated weights per individual	Pigs, Sheep	Live-weights estimates per individual but individuals not identified	Walk over weighers (e.g. at races, or in pens)	Managing Nutrition & Production; Detecting Health & Welfare Problems in groups/of individuals	Can be used to sort into different feeding areas using marking and/or gates system. For pigs in pens can be placed between loafing and feeding areas or separating off if unwell.
5 Estimated liveweights per group/ individual	Poultry - Broilers, Turkeys	Live-weight plus number on plate hence average liveweights; Individuals not identified; Some platforms only measure	Weighing Plates/Platforms for individuals/groups	Managing group Nutrition & Production; Detecting Health & Welfare Problems in groups	This is a sampling of weights in the flock. Could give 1000s of weight measurements per day.

		one bird at a time			
6 Milk parlour data	Dairy Cows	Milk yield, milking duration, peak flow; milk quality, Somatic Cell Count (SCC); position in parlour/ milker	Automatic milking systems plus manual sampling	Managing Nutrition & Production; Detecting Health & Welfare Problems of individuals	Milk quality and SCC measures may not be available in real time but could be sampled regularly (e.g. once per day or week).
7 Milk bulk lab data	Dairy Cows	Somatic Cell Count (SCC), Milk quality	Milk bulk sampling - manual sampling	Managing group Nutrition & Production; Detecting Health & Welfare Problems in groups	Milk quality and SCC measures may not be available in real time but could be sampled regularly (e.g. once per day or week).
8 Milk other bulk data	Dairy Cows	Temperature, Volume, Stirring	Milk bulk sampling - various sensors	Managing group Milk Production and Processing	Real time monitoring of physical attributes of milk in bulk tank are available
9 Accelerometer	Dairy Cows, Beef Cattle, *Pigs*, *Sheep*, *Goat*	Behaviour (e.g. activity or time budgets in different classes: lying/standing, grazing/not, rumination, … or raw acceleration in x, y and z direction)	Accelerometer	Detecting Heat, Calving/ Lambing/ Farrowing, Health & Welfare Problems	Not usually used on pigs on real farms. For sheep cheaper options needed. For grazing animals often removed at intervals for data download and recharging. Can give raw accelerometer data but sometimes measures are derived only (e.g. behaviour).
10 Pedometer	Dairy Cows, Beef Cattle, *Pigs*, *Sheep*, *Goat*	Behaviour (step count)	Pedometer	Detecting Heat, Calving/ Lambing/ Farrowing, Health &	Less advanced than accelerometer

				Welfare Problems	
11 Location	Dairy Cows, Beef Cattle, *Pigs, Sheep, Goat*	Location and behaviour	GNSS (Global navigation satellite system), GPS (global positioning system)	Managing Grazing & Production; Detecting Health & Welfare Problems of individuals	
12 Relative location	Dairy Cows, Beef Cattle, *Pigs, Sheep, Goat*	Location relative to static receivers and behaviour	Proximity loggers plus static receivers	Managing Grazing & Production; Detecting Health & Welfare Problems of individuals	Locations can be estimated as well as mother-off spring distances
13 2D Imaging – pigs	*Pigs*	Body condition score, Liveweight	2D Imaging from above	Managing Nutrition & Production; Detecting Health & Welfare Problems of individuals	Can be placed between loafing and feeding areas and used to sort into different feeding areas using gates system
14 2D/3D Imaging – identified cows & pigs	*Cows, Pigs*	Body condition score, Liveweight, Behaviour	2D/3D Imaging from above	Managing Nutrition & Production; Detecting Health & Welfare Problems of individuals	Identifying individuals is difficult so used in combination with reading RFID tags at intervals and then tracking.
15 2D Imaging - birds	*Poultry - Broilers, Turkeys*	Location and behaviour; Dead birds; Weight estimation;	2D Imaging from above	Managing Nutrition & Production; Detecting Health & Welfare Problems of groups	
16 Thermal imaging	*Dairy Cows, Beef Cattle, Calves, Pigs, Poultry, Goats, Sheep*	Body Temperature	Thermal Imaging	Detecting Health & Welfare Problems of individuals	Can be used for detecting heat stress, and potentially fever, pain, …
17 Thermometer	*Dairy Cows*	Body Temperature	Thermometer	Detecting Health & Welfare Problems of individuals	

18 Sound - vocalisations	*Dairy Cows, Beef Cattle, Calves, Pigs, Poultry, Goats, Sheep*	Specific Species Dependent Vocalisations	Acoustic Sensors	Detecting Health & Welfare Problems of Groups	These sensors are mounted in e.g. house, but could be used outside in confined areas
19 Sound – feeding behaviour	*Cows, Sheep*	Feed intake, Behaviour (grazing, ruminating)	Acoustic Sensors	Managing Grazing & Production	These sensors are mounted on animals
20 Remote satellite sensing	*Available Grazing for Cows, Sheep*	Quality of grazing	Remote Sensing (Satellite imaging)	Managing Grazing	
21 Drones	*Cows, Sheep and Available Grazing*	Quality of grazing; Location of groups;	UAV (Unmanned Aerial Vehicle)	Managing Grazing; Detecting Health & Welfare Problems	
22 Environmental sensors	Environmental conditions for livestock	Temperature, Humidity, Emissions (e.g. Ammonia, Methane, CO2)	Environmental sensors	Managing Health & Welfare Problems of groups; Managing emissions	
23 Weather	Weather for livestock	Temperature, Humidity, Rainfall, Windspeed, …	Weather station	Managing Health & Welfare Problems of groups	

[1] Tends to be for research studies currently rather than on real farms.

Table 1. (**b**) Examples of commonly used sensors for real time monitoring of livestock on farms showing various characteristics that may be associated with the measurements and resulting monitoring data streams. The nature of the resulting sensor measurements is not listed here as raw data may be unprocessed, or may be processed in differing ways, leading to multiple types of measurements for each type of sensor that could be used for monitoring.

Name	Measurement on:				Timing of sensor measurements			Animals are:		Sensor is:		Sensor is:	
	Individuals ID known	Individuals ID not known	Some Individuals ID not known	Groups or impact on groups	Continuous or near-continuous	Intermittent	Regular	Housed (usually)	Outside/grazing	On animal	Not on animal	At fixed location[1]	Mobile
1 Individual intakes	●					●		●	●		●	●	
2 Group intakes				●	●	●		●	●		●	●	
3 Individual weights	●					●	●	●	●		●	●	
4 Estimated weights per individual		●				●		●			●	●	
5 Estimated liveweights per group/individual			●			●		●			●	●	
6 Milk parlour data	●					●	●	●	●		●	●	
7 Milk bulk lab data				●		●	●	●	●		●	●	
8 Milk bulk other data				●	●			●	●		●	●	
9 Accelerometer	●					●	●	●	●	●			●
10 Pedometer	●					●	●	●	●	●			●
11 Location	●					●	●		●	●			●
12 Relative location	●					●	●		●	●		●	●
13 2D Imaging – pigs		●				●	●	●			●	●	
14 2D/3D Imaging – identified cows & pigs	●					●	●	●			●	●	
15 2D Imaging - birds			●	●		●	●	●	●		●	●	
16 Thermal imaging		●				●	●	●	●		●	●	●
17 Thermometer	●					●	●	●	●	●			●
18 Sound - vocalisations				●	●			●	●		●	●	
19 Sound – feeding behaviour	●					●	●		●	●		●	●
20 Remote satellite sensing				●	●				●		●		
21 Drones			●	●	●	●			●		●		●
22 Environmental sensors				●	●			●			●	●	
23 Weather				●	●				●		●	●	

[1] e.g. in milking parlour, at feeders, in field.

With regards to *when*, some types of sensors are on the animal (e.g., accelerometers, GNSS collars or tags), generally resulting in (near) continuous time measurements, but others are at fixed locations and so often provide intermittent measurements (e.g., monitoring feeding bouts, or liveweights from walk over weighers). Management practices may imply, or may be designed such that that some of

these measurements are taken at regular intervals (e.g., 2 or 3 times per day at/or to/from the milking parlour). Other measurements may be continuous time when being measured, but with lags in data availability and gaps (e.g., when accelerometers/GNSS sensor collars need to be removed from grazing sheep or cows for data download and/or battery charging) and others may be intermittent but continuous when being measured (e.g., when an individual is in view of 2D overhead cameras, or when an individual wearing a proximity logger is close to an antennae in the field). In general, data most useful for real time monitoring is data that is available in real time at sufficiently high temporal resolution at regular time intervals (i.e. *time series*). Near continuous data at high temporal resolutions (e.g., every minute), or intermittent data, may be summarised to lower temporal resolutions (e.g. daily) resulting in regular time series data that are more practical for use for real time monitoring. However, in doing this care must be taken not to lose aspects of the data which may be important for detecting health issues, such as changes in diurnal behaviour [128–130]. Spatial data could be available at a (near) continuous spatial scale (e.g., measurements from satellite imaging), at an intermittent spatial scale (e.g., GNSS on animals) or at discrete locations (e.g., in situ environmental or weather monitors). Monitoring data from sensors where each sensor measurement at each time is at a recorded spatial location (e.g., measurements from repeated satellite imaging, grazing cows wearing collars with both GNSS and accelerometers) is referred to as *spatio-temporal* data. Explicit approaches for real time monitoring that utilise spatio-temporal data could be developed but here, we focus on time series data. For simplicity, spatial measurements can be summarised into time series, for example converting animal locations at fine temporal scale into distance travelled over pre-defined intervals (e.g. daily).

Who is being measured relates to whether individuals are monitored (e.g., accelerometers on dairy cows and walk over weighers that measure one animal at a time) or whether only groups are monitored (e.g., from automatic feeders and drinkers in pig or poultry houses). Either way, all individuals in a group may be measured or just some subset (e.g., some cows or sheep grazing on the hill may have GNSS collars and some may not, only a subset of poultry flocks will likely walk over weigh plates). Radio-Frequency Identification (RFID) tags, sometimes called Electronic Identification (EID) tags, are commonly worn by larger animals in order to record measurements from fixed or hand-held sensors for identified individuals. In some cases, measurements are made on individuals, but the individuals may not be identified in the data set (e.g., walk over weighers for sheep or pigs, used with a marking or gate system to sort them into different feed groups). Individual identification is ideal for health and welfare monitoring of individuals but clearly monitoring of groups and on unidentified individuals in groups could still be useful, since all livestock monitoring measurements are expected to be used for indicating when animals should be checked manually.

The final major consideration is the *nature of the monitoring measurements* available at each data point. Often the measurements are numerical and *continuous* (e.g. walk over weighers, milk yield, temperature, feed or water intake measured from automatic feeders) or *discrete* (e.g. step counts or counts of alarm calls in successive time intervals). Occasionally the measurements are *ordinal* (e.g., liveweight category for sorting into feed groups, body condition scores (BCS) or locomotion scores). Another measurement type is a *classifier* with no ordering and this is most often encountered when measuring behaviours according to some pre-define ethogram that will be species and/or environment dependent (e.g. classes could be lying, standing, grazing, ruminating, walking, running, sleeping, perching, feeding, drinking, …..). Sometimes binary classifiers are generated (e.g. grazing versus not grazing, lying versus standing). Furthermore, the measurements from a single sensor could be single continuous measures (e.g. liveweight, feed intake), or multivariable (e.g. acceleration in 3 directions, amounts of different gases), or even more complex (e.g. sound, image). Sometimes classifications are 2 way, for example behaviour by location classes. And finally, there may be multiple data streams per individual or group arising from different sensors, often at different temporal resolutions. Note that further processing (see below) could result in data streams that are different in nature than the original raw data.

It is convenient if the monitoring data alone can be used, however, often it needs to be used in conjunction with other information on individuals or groups. For example, behaviour will likely be affected by the amount of time spent grazing versus being housed, or the weather, whilst liveweights will be affected by days in gestation, birth or possibly changes in diet. If other information is needed, then, for real time monitoring, the issue of whether it is practical to record the required information accurately, and in a timely manner, on real farms needs to be considered.

The data from many types of sensors will need initial checking/cleaning before it is ready for subsequent use. How this is best done will depend on the detail of the sensor data. For example, liveweights, and liveweight changes must be within sensible ranges for the species, and can only change gradually, whereas behaviour could change suddenly. There are a range of methods that could be used for data cleaning, but it is important that these methods do not eliminate genuine data that are indicative of problems we want to detect and that this process is automatable on-farm.

Some sensor technology has underlying methods implemented that automatically converts the raw data into other derived measurements. For example, accelerometer data, where the raw data is usually acceleration in 3 dimensions at fine temporal resolution, is often converted to a general activity measure, or time budgets of behaviour classes [131–133], or counts of behaviours, over pre-specified time intervals. Video may also be converted into time budgets or counts of behaviours, whilst sound may be converted, for example, into counts of coughs, or species dependent alarm calls. However, it could be advantageous to use raw rather than derived data for real time monitoring, since estimated quantities could be inaccurate, and, more generally, information is likely lost in estimating derived quantities, but this must be traded off against both complexity and increased computational load. Even for complex information such as sound, or images, in theory the raw (digitised) information could be used by a generic quantitative method for detecting changes in real time indicative of a problem. Use of all raw information might also be advantageous as it could have the capacity to detect a wider range of problems, but on the other hand there may be disadvantages in taking this data driven approach, due to a lack of focus on aspects of the data streams known, a priori, to be indicative of specific problems. Note that, where use of raw data is not feasible in real time monitoring and prediction, it would likely still be advantageous to be able to use the raw data for validation and to develop alternative quantitative approaches, However, some devices available for use on farms only gives access to derived quantities, either for commercial reasons or because on some devices the raw data are used to give derived quantities but not stored to save on memory use.

Following, we discuss generic issues relating to real time prediction/decision and prediction/decision validation that are applicable to any data streams, whether raw or derived, though we are focusing on time series data [134] per individual or group. As mentioned above, many sensor data streams can be converted to time series data. Furthermore, though our focus is on real time prediction, data at fine temporal resolution may be converted to courser resolution to reduce computation load, to smooth out noise in the data, or to align with the resolution of the prediction/decision process.

3. Framing the Prediction/Decision and Validation Problem

In framing the prediction/decision and validation problem (Figure 1), it is important to distinguish between the true (unknown) state of the monitored individual and observed data. Observed data is driven by the true state of the individual which has multiple dependent drivers including how it is being managed, the environment it is in, its current health and welfare, and other aspects which make up the biological state of the individual such as breed, age, stage of gestation or lactation, etc. Observed data will be affected by noise and inaccuracies or incompleteness due to limitations of data collection methods. In addition to this, observed sensor data may differ between individuals for other reasons, for example normal activity in some animals may just be inherently higher, and/or more variable between times, than activity in other animals.
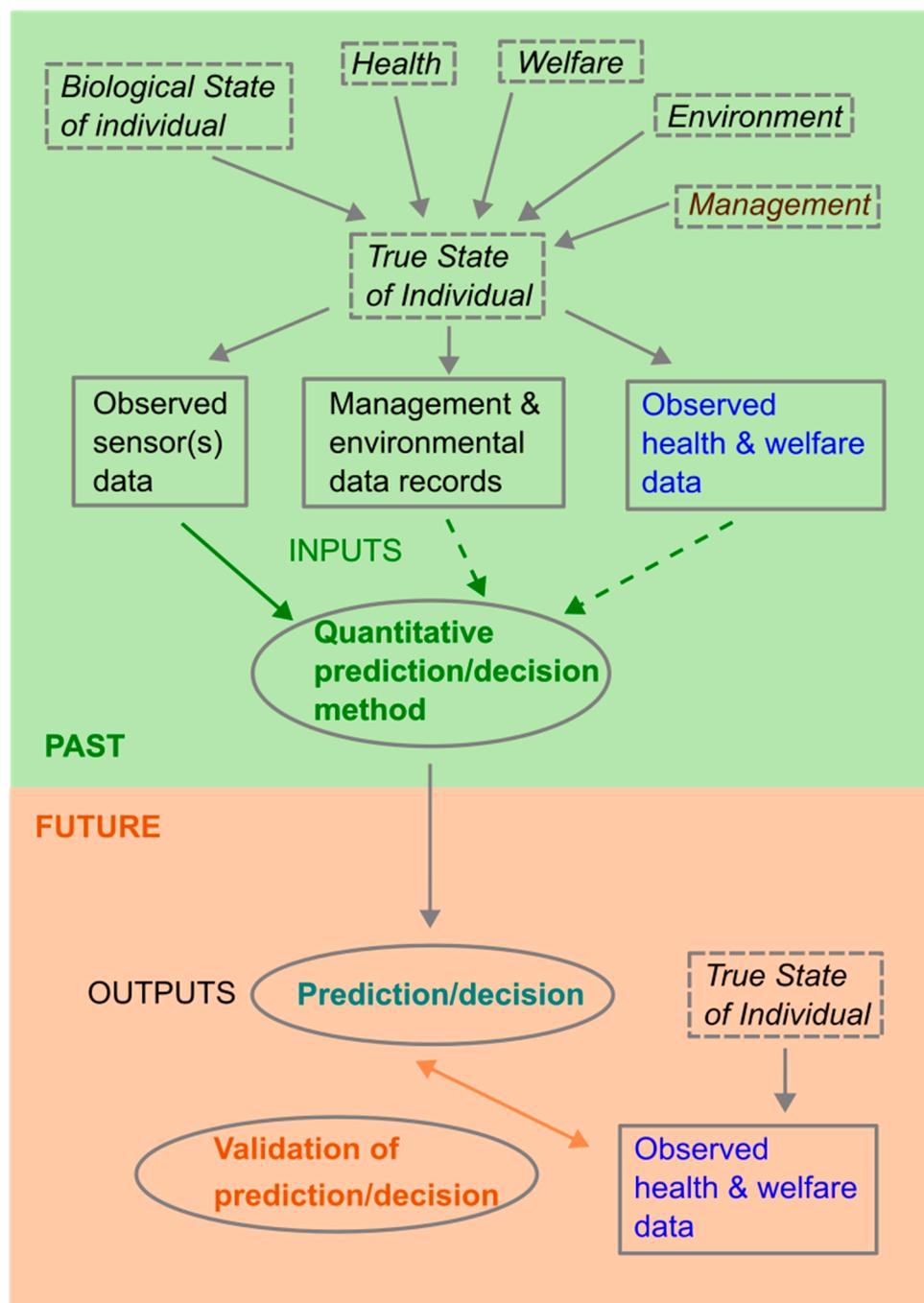
Figure 1. Diagram showing snap shot of prediction/decision and validation stages for an individual at one time step. Dashed arrows are optional depending on methods used as some methods are based on sensor data alone whilst others also use other observed information, for example past health and welfare data. Validation involves assessment of whether the prediction/decision made agrees with current (or near future) health and welfare data.

All quantitative methods use some inputs (data) to obtain outputs (the prediction/decision) at each time step (Figure 1) and this process must be repeated to allow decisions to be made in real time as the data evolves (Figure 2).
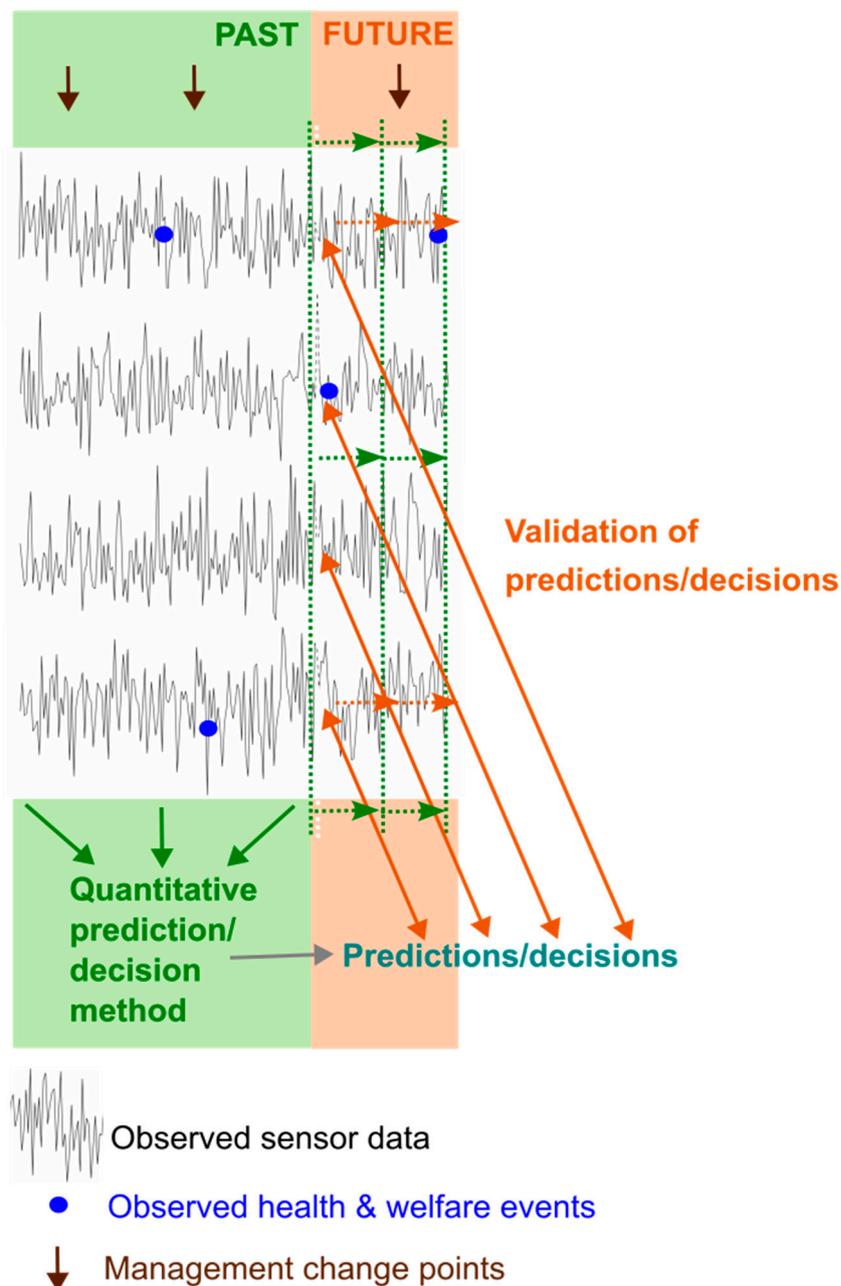
Figure 2. Diagram showing the prediction/decision and validation process which is repeated successively as the sensor data evolves in real time. This shows simulated sensor time series data for 4 individual animals that are being managed together. At each time step, past data are used to make a prediction/decision for each animal, which are then validated by comparing to current (or near future) health and welfare data. This process is repeated at each time step as the time series data evolve.

One important distinction between development of quantitative methods used to make predictions/decisions is that some of them use past sensor data alone, whilst others also use independently collected past observed data on health & welfare issues. We will refer to the latter as *supervised* methods and the former as *unsupervised* methods. Unsupervised methods usually involve some means of detecting that observed sensor data differs from expected sensor data when the individual is in a normal state, which is taken as an indication (prediction) that an individual may have a problem and should be checked and managed/treated appropriately (decision). However, unsupervised methods may need health and welfare data in order to optimize decision making.

Supervised methods usually involve fitting a statistical model (or training a machine learning method) which uses past sensor data as inputs to predict current (or near future) observed health and welfare data as outputs. Data used for fitting could be past data from the individuals for which we require predictions/decisions or it could be data from other individuals. Once the model has been fitted, past sensor data from an individual may be inputted into the model to output a prediction of whether or not an individual currently has a health or welfare problem. Other health and welfare data may be needed to optimize decision making.

As mentioned above, the true state of the individual, and hence sensor data, may be influenced by drivers other than health and welfare, so any changes in these other drivers need to be taken into account in the method used to make predictions/decisions. There are two fundamentally different approaches to this. The first is to use observed data on these other drivers (which may be imperfect) to adjust for changes in the prediction/decision method (such as using these as additional covariates in statistical models). The second is to treat these other drivers as changes to be predicted in the same way as health and welfare data are handled.

Once a prediction/decision method has been developed it needs to be validated which involves comparing successive real time predictions/decisions with the health and welfare of individuals as the data evolves (Figure 2). A limitation of this is that it can only take place against observed data on health & welfare issues for an individual which are likely imperfect. Initial validation can take place on the data set on which a method has been developed but final validation must take place on other data.

4. Methods for Predictions/Decisions

To recap, the objective we are focused on in the context of livestock is to use past monitoring data to decide whether there is an indication that an individual (or group) has issues currently and so should be checked and managed/treated appropriately, and to be able to repeat this decision-making process dynamically in real time on farms as monitoring data evolves. There are numerous quantitative methods that could be applied to use available data to learn to make predictions/decisions about other data of interest, which can all be termed statistical learning [135]. The most obvious and/or commonly used methods are described below.

4.1. Simple Methods

By far the simplest method for making decisions based on a numerical monitoring data stream from each individual is to use a pre-defined constant (upper and /or lower) threshold(s), on the current (or most recent) monitoring data to classify an individual as abnormal in some way at the current time. An extension of this method would be to use different thresholds for different breeds, farming systems, and so on, or to use different thresholds depending on current environments within the same system (e.g. grazing versus not today or seasonal effects). The choice of these thresholds could be established from previous research comparing sensor data in a range of health and welfare data and in different contexts.

Another simple and generic approach to tailor this to a specific situation, or farm, would be to base thresholds on whether the current monitoring data observation lies in the (upper and/or lower) extremes(s)/outlier(s) of the empirical cumulative distribution function (eCDF) of all past monitoring data collected so far on that farm. Or, to account for long term 'normal' trends, this could be based instead on a more recent moving window of past monitoring data that is large enough to estimate the eCDF with sufficient accuracy to judge outliers. Another obvious extension when enough data is accumulated, would be to estimate eCDFs for each individual based on its' monitoring data, so that what is 'normal' could be defined differently for different individuals. Note that this simple approach could allow that 'normal' monitoring data streams between individuals vary both with respect to means and variances. This seems a sensible approach for individuals that are usually well, but of course, it may fail to detect an individual that is constantly ill, and, conversely, may indicate an individual that is always well, as ill, just because the extremes in their eCDF do not arise from extreme

monitoring data. Therefore, it is clearly oversimplistic to only consider within individual effects. Instead, some information must be shared between individuals as well. Comparison to recent/current group eCDF could allow adjustment to be made for management or environmental effects that are changing locally in time (i.e. short term effects). Note however, that any method based on eCDFs should allow that the more current observations being checked at any one time, the more likely some are to lie in the tails by chance when there is no problem.

Another way of dealing with long term trends would be to smooth each monitoring data stream to establish a baseline and then to use threshold(s) to identify (upper and/or lower) extremes/outliers in the eCDF of the residuals from the smoother (ie. the monitoring data observations – smoothed values) (Figure 3). Smoothers such as polynomial curves, simple moving averages, spline smoothers, or exponential smoothers could be used [136,137]. In deciding on appropriate smoothers, and on the extent of smoothing, it is important to consider the nature of departures that would be picked up. If the degree of smoothing is too weak, short term fluctuations indicative of a problem could be included in the baseline and so go undetected (Figure 3(a)). On the other hand, if smoothing is stronger (Figure 3(b)), longer term fluctuations that are not indicative of problem could be excluded from the baseline and thus may be judged to be a problem. This method may be extended adjusting for either long or short term trends seen for all individuals being managed together at the same point in time, from the empirical distribution of residuals for each individual (Figure 3c). This is appealing as a generic sensor data driven way of dealing with management effects, such as changes in feed or grazing regime, or management group (e.g. herd or flock) wide treatments, without needing to use recorded data on management, and can be useful when management effects on sensor measurements are consistent across the group. However, group wide changes should always still be flagged as they could indicate a herd wide problem such as heat stress or disease spread, in which case recorded management data would still be needed to establish whether group wide changes are due to management or indicative of some problem. Smoothing methods should generally be applied adjusting for both management group effects and differences in individual monitoring data of individuals in their normal state.

It is possible that some basic preprocessing of monitoring data before simple methods are applied may improve them. A transform such as calculating logs may improve accuracy where effects are multiplicative on the raw scale and hence linear on the log scale. Some aspects mentioned above could be dealt with by preprocessing for example at each time point, subtracting the mean monitoring data over the group from individual monitoring data to adjust for group level changes, or standardising monitoring data for each individual could adjust for individual bias and noise which may vary between individuals in their normal state.
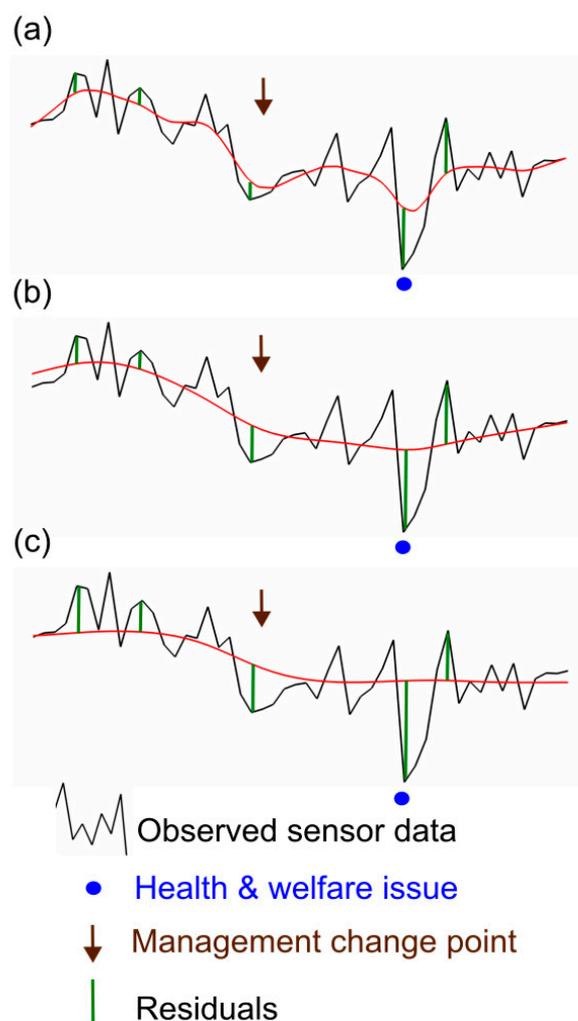
Figure 3. Monitoring data for one individual with (a) a weak spline smoother based on this monitoring data which follows short term changes (b) a stronger spline smoother based on this monitoring data which follows long term changes and (c) a stronger spline smoother based on the monitoring data of all individuals managed together which thus follows long term management group changes in time. This simulated data has one management change point which results in a reduction in monitoring data streams from all individuals, and one later health issue for this individual which causes the sensor measurement to decrease further. The residual coincident with this health issue is relatively small for (a) and relatively large for (c), though in all three cases it is extreme compared to other residuals for this data. Here we have smoothed all the monitoring data but, for real time prediction/decision, residuals at the current time point must be calculated from a smoother only based on past data.

4.2. Anomaly/Change-Point Detection

Detection of change-points in time series data is a well know methodological research area with a wide range of applications [138] which can be viewed as changes in the parameters underlying a statistical model of the time series. Often change-point detection focuses on detection of one or more long term state changes in a single univariate time series when the full series are available (offline change point detection). However, we are interested in detecting recent changes in any of a multiple of time series based on data so far (online change point detection). Furthermore, we are interested in detecting both abrupt short term changes (anomalies), as well as more gradual short term changes

that could be indicative of a problem, whereas change-point detection tends to focus on non-transient changes from one underlying state to another.

The area of process or quality control [139] is relevant in which control charts can be used to ascertain if a monitored process is out of control. This is generally implemented by generating an alarm when a measured quantity deviates from a predefined acceptable target distribution. One method commonly used is referred to as a cumulative sum (CUSUM) control chart in which the cumulative sum is a value that can be adjusted by long term average trends in order to ascertain whether recent values are outliers according to some predefined threshold.

Anomaly, or outlier, detection applied to time series [140] can include methods arising from data where each data point is pre-labelled as normal or abnormal, referred to as supervised, but unsupervised anomaly detection is more commonly used. For anomaly detection in theory any method for forecasting time series [141] could be used, defining anomalies as time series observations that depart from their forecasts. Actual methods used for online change point detection, or anomaly, or outlier detection, tend to fall into one of the categories below of statistical modelling, latent class or variable modelling or machine learning.

4.3. Classical Statistical Modelling

There are two fundamentally distinct ways in which classical statistical modelling approaches can be used, which can be referred to as unsupervised if based on monitoring data only or supervised if gold standard data on the outcome of interest(s) is used for model development and fitting.

4.3.1. Modelling Usual Monitoring Data

The first (*unsupervised*) is to form a statistical model of 'normal' monitoring data for all individuals (or groups), which can be applied to all (or some window of) the monitoring data so far, and then to use this, together with current/recent monitoring data to decide whether an individual (or group) is exceptional at the current point in time indicating there is likely to be some health or welfare problem. Formally, the statistical modelling uses past monitoring data to provide a predicted probability density function (PDF) for the monitoring data for each individual at each time, with subsequent current/recent monitoring data at the upper and/or lower extremes of this distribution likely indicative of some health or welfare problem.

For numerical monitoring data, linear mixed modelling (LMM [142]) could be used with random effects included to allow for random variability between and within individuals, whilst also incorporating appropriate fixed effects characterising management, time in the year, time relative to giving birth, and so on. Extensions might be needed such as to allow for different underlying variability in sensor data between times for different individuals, or for errors that are correlated locally in time in sensor data as is often seen in time series data.

4.3.2. Modelling based on the Outcome of Interest

The second approach (*supervised*) relies on having gold standard outcome data. The LMM described above could be used to model the monitoring data, but with the addition of an explanatory variable of a classification of a health problem, or the severity (a continuous or ordinal variable) of a health problem (or problems). After initial model estimation, this could then be used for decision making in real time based on comparing actual monitoring data from each individual at each time with monitoring data PDFs estimated from the model assuming no health problem.

Conversely, a statistical model could be used that directly models the outcome of interest (e.g. a classification of a health problem, or the severity of a health problem) as a function of concurrent monitoring data. For example, generalised linear mixed models (GLMM [142,143]), with logit link and binomially distributed errors, could be used to model the binary outcome health problem/not as a function of concurrent monitoring data. LMMs could be similarly used for modelling severity outcomes. As above, random effects could be used to model inherent variability between and within

individuals as well as appropriate fixed effects for management changes, for example, which might cause shifts in the monitoring data and hence in the outcome versus monitoring data relationship. After initial model estimation, this could then be used for decision making by using thresholds on the estimated probability of the binary outcome, or PDF of the severity, in the model estimated repeatedly for each individual at each time from past monitoring data together with other fixed effects.

For all these approaches, though some generic model development could be carried out for specific sensors on specific species, at some point this model would almost certainly need to be fitted and validated in the specific context in which it is to be used (i.e. for the actual animals on the farm). In particular, where there is between individual variability in monitoring data, random effect estimates would need to be obtained for the levels of individuals, in order to use this for decision making on farm for those individuals. Further, for methods involving modelling the outcome of interest, gold standard data would need to be available on farm at least initially, and possibly later on as well to intermittently sense check that the estimated model is still working, or to adjust it.

4.4. Latent Class or Variable Modelling

Latent class or variable modelling [144] can be used to model monitoring time series as a function of either unobserved states (e.g. individual has a health problem, or not) or unobserved variables (e.g. the severity of a health problem). Hidden Markov Models (*HMM* [145]), are a special case of latent class modelling commonly used to model hidden mutually exclusive states underlying time series data, that could be applicable here. Cluster analysis [146] is also used to deduce latent classes, though it is limited in that it does not allow other aspects of the data to be incorporated or modelled.

It is helpful to take a more generic approach of latent class or variable modelling. Multiple mutually exclusive latent states, multiple 2 level latent state classifications, or multiple latent variables could be used for multiple health problems. In a Bayesian model [147] this can be done by specifying a parametric model for the association of the numerical monitoring data to the unobserved state(s) or variables. Bayesian modelling [148] also provides the flexibility to model the latent class(es) or variables(s) in combination with other fixed and random effects as described above. Furthermore, relationships between the latent classes/variables and covariates could be included in the model, for example, to include tendency of health issues to occur at particular times of year (e.g. lameness in grazing animals) or times relative to birth (e.g. health issues that tend to occur soon after calving in dairy cows).

The resulting posterior distribution of the latent class(es)/variable(s) provides an estimate of the probability that an individual has a health problem, or an estimated PDF for severity, at the current time, which could be used for decision-making.

This approach can be classed as unsupervised in the sense that no data on the health status of individuals by times is needed for its' application. Theoretically it is the most obvious direct approach to this problem. However, whilst it is relatively easy to specify appropriate models in a Bayesian framework, repeated fitting of the model as the monitoring data evolves for multiple individuals is likely to be computationally intensive though algorithms have been developed [149] for sequential data that are specifically designed to address this kind of problem.

4.5. Machine Learning Methods

Machine learning is widely used in agriculture [5,16,17,150] including in livestock production and welfare. It is the most commonly used prediction method in recent papers for predicting health problems in livestock from auto-monitoring data [38,41,46]. [135,151,152] give introductions to machine learning methods and overviews of different techniques. Basically, machine learning is based on data that contains data points with multiple *features* (measurements that can be numerical or categorical). The machine learning task will consist of using one or more of these features (referred to as input features) to predict one or more other features (referred to as output or target features). *Supervised* machine learning is when the method is trained on a set of data points with known output

features, usually referred to as labelled data points. *Unsupervised* machine learning methods are trained on data sets containing input features only, and must predict output features, most commonly classifications, by patterns in the input features. *Semi-supervised* machine learning is when the two approaches are combined by application to data points, some of which are labelled and some of which are not. When the output feature is categorical (e.g. individual has a health or welfare problem, or not) this tends to be referred to as a classification problem, or clustering in the case of unsupervised learning, whereas if the feature is numerical/continuous (e.g. severity of a health or welfare problem) this tends to be referred to as regression. Regression problems can be reformulated as classifications by binning values, though this is an ordinal, not categorical, classification. Note also that machine learning can be used to predict non-mutually exclusive classes.

In general, training in supervised learning will be based on minimising some *cost* (or *loss*) function which measures the departure of the predicted outputs to the actual outputs. For example, for binary target classifications binary cross entropy, which is *log(p)* where *p* is the probability of the true outcome, summed over the data points can be used whilst for regression, the mean square error (MSE), can be used. Unsupervised learning will also be based on minimising some *cost* (or *loss*) function, but it will not be dependent on the actual outputs. Either way, a balance must be struck between underfitting and overfitting, as both would result in inaccurate predictions on new data sets. Another general point about training is that some machine learning methods can be used with different settings (sometimes called hyperparameters), which can be chosen to have minimum cost, or best goodness of fit, for the training data set. In supervised machine learning it is important that appropriate coverage is achieved in the output space. For example, when training to predict a classifier there should an equal number of data points in each level. When there are imbalances, balance is often achieved by up- or down-sampling data points.

Once a method has been trained, the idea is that it can be used in other contexts for predictions of outputs, based on input features for new data points. Whether this results in accurate predictions or not, depends on both whether the input features capture everything needed to predict outputs reliably, but also on whether the machine learning 'model' holds. It may be that training may be needed in new contexts in which it is to be used. Furthermore, if one of the input features needed to get accurate predictions is the identification of the individual, then the machine learning method will need to be retrained in any context with new individuals.

4.5.1. Basic Machine Learning Methods

Interestingly, well established basic classical statistical methods, (such as regression, logistic regression, cluster analyses, PCA, and so on), along with parameters with the usual statistical interpretation (such as regression coefficients), are mentioned as machine learning methods in the above introductory papers. This is because, as mentioned above, all these methods are essentially doing the same thing, statistical learning [135]. That said, when used in machine learning the methods by which the best fitting model is found tend to differ from in classical statistics where methods with provable theoretical properties (e.g. maximum likelihood) are often used. Regression methods for linear relationships include ridge, LASSO and elastic net regression. Support vector machines (SVM) can deal with non-linear relationships, by transforming the space of input features to a space with additional dimensions, in which simple linear classifiers can be used.

Decision trees and *random forests* are often used for supervised machine learning problems and can handle non-linear relationships. Decision trees basically involve successively partitioning the data into mutually exclusive sets based on criterion for the input features (which can be numeric or classifiers), resulting in a tree that best (according to some predefined measure) separates data points in the target space, which can either be a classification (classification tree) or numerical (regression tree). It is important with decision trees not to overfit the data to the targets as this will give inaccurate predictions on new data sets. Even if not overfitting, decision trees can appear to be good for the data set on which they are trained, as they were optimised for that, but they are often inaccurate when used to predict other data sets. To get around this, random forests are created by repeatedly

generating optimal decision trees on bootstrap samples from the training data set, with each successive selection step in creating the tree based on a randomly selected subset of the input features. This results in a wide variety of trees. To get predictions, for any set of input features, estimates may be obtained for the output feature or target from the empirical PDF across these trees, on which decisions could be based.

4.5.2. Neural Networks

Neural networks are networks that can be trained to solve a wide range of supervised machine learning problems, and they are being used increasingly recently [150]. Basically, the input features are the starting nodes, and the final nodes are the choice of output results. These output nodes usually make up a mutually exclusive categorization (with one class per node), along with probabilities of each node, though continuous variables can be handled similarly by binning to get classes. There are hidden nodes in layers in between the inputs and outputs. Each node takes a numerical value and is linked to nodes in the preceding layer using a weighted sum, plus some bias, and then a prespecified transformation to ensure that the resulting values lie between 0 and 1. Finding an optimum solution, involves setting all the weights and biases in the network appropriately to minimise some cost function over a training data set. Algorithms (e.g. back propagation) underlying neural networks do this efficiently. *Recurrent neural networks* (RNN) are suggested for predicting future time series from past time series. They can cope with sequences of varying length, and contain feedback loops, so that predictions for the current time depend on the past time series. The underlying model is stable in the sense that weights and biases linking consecutive values remain constant. However, they are difficult to optimise. *Long Short-Term Memory* (LSTM) networks are used to get around these optimisation problems by using different paths to control the relative impact of long to short term information on predictions. However, it is unclear whether these methods can be used for predicting other outputs, and they would likely need to be extended to include additional input features (not just the monitoring data).

4.5.3. Application of Machine Learning in Prediction/Decision Context

Here, the objective is to predict a different outcome(s) measure, based on monitoring data (time series) so far for an individual (Figure 2), so this is how training would have to be done in order to get accurate predictions in practice. Given monitoring data streams for the full training set of data over a long period for each individual, this would create a huge training data set; for each individual at each time the data to be used as input could be the full monitoring data so far, together with other input features such as individual level variables (e.g. breed, age, …), time level variables (e.g. weather, management changes), and so on. Traditional machine learning methods assume that the number of input features do not change, so to use the actual monitoring data as inputs, windowing would be needed so that the same amount of past monitoring data is used at every time step for every individual. Furthermore, machine learning methods generally ignore random effects [153] whereas it is plausible, as mentioned above, that underlying sensor values for individuals in their normal state need to be estimated to get accurate predictions of health outcomes. Presumably, this could be allowed for by including individual as part of the input features (for example, in neural networks, there could be one input node binary classifier per individual).

Research in this area [41] often avoids these various complicating issues, for example, by reducing the data set to one (positive or negative) event per individual and summarising monitoring data, in a fixed number of multiple ways, up to that event time. More generally, a set of pertinent input features could be derived from the set of past monitoring data. For example, plausible measures could characterise recent short term variability for individuals, compared to their long-term variability, recent changes, changes per individual adjusted for diurnal effects or for group level management changes, and so on. This avoids complexity in the machine learning methods, but at the expense of a lack of generality, because of having to decide, a priori, what derivations from the initial set of input data are pertinent to the objectives. Any approach that essentially discards available

information in data in this way, is not ideal when extensive long term monitoring data is available for individuals. Furthermore, these preprocessing steps applied to sensor data prior to using machine learning methods seem to be counter to a perceived appeal of machine learning methods, which is that they are black box, purely data driven, flexible, generic methods. However, it is suggested [154] that for machine learning to be successful, knowledge needs to be combined with data, and that this can aid, not impede, in generalization of machine learning methods. It is plausible, that if the same data and knowledge are incorporated, albeit in different ways, for different quantitative methods, then difference in performance between these methods would be likely to reduce.

4.6. Discussion of Alternative Prediction/Decision Methods

The main advantages of the simple methods described above (section 4.1) is that they are easy to compute dynamically in real time as the data evolves, and that they are generic, in that they could be applied to any set of univariate numerical data streams. It should be noted though, that these methods are black-box, in that their output can lead to a binary decision, or in some cases a measure of the extent to which that decision should be made, but beyond that, does not elucidate more information about the underlying process. A common criticism of using black box methods is that detailed inferences about the system being studied cannot be made. However, on the premise that the basic inference we want to make in livestock monitoring, is just when to check animals or not, simplistic methods, or completely black box data driven methods (such as some machine learning methods), are acceptable so long as they can be shown to work. That is, it must be *computationally feasible to repeatedly apply the methods in real time on farm optimising decision making along the way, in any contexts in which they are to be applied*. However, a method that concludes, based on one farm or population of farms, that some simple threshold can be used on the monitoring data streams resulting in optimal decision making, may not translate to other populations.

In further discussing alternative methods, it is useful to show simulated monitoring data illustrating some of the complexities. Figures 4(a) and 4(b) show simulated simultaneously observed monitoring data streams from 2 different types of sensors (both in black), together with components assumed to drive this data, for 12 animals in 3 management groups for which 4 types of health issues occur (1&4: illnesses, 2: isolation of individual animals for health reason, 3: heat). Observed data for each sensor depends on baselines per animal (both in red) which will depend on the current biological states of the individuals (for example stage in gestation or lactation) and which may exhibit underlying differences between animals in their average levels and noise. Baseline sensor levels (red) illustrate how some animals (e.g. 4 sensor 2) have lower levels than others (e.g. 12 sensor 2) and levels for some animals (e.g. 5, sensor 1) appear to be inherently more noisy than levels for others (e.g. 9 sensor 1). Sensor levels may be altered by the occurrence of management changes or health issues. Observed data from the first sensor (Figure 4 (a)) are behaving as we might expect for activity measured by accelerometers, with activity immediately increasing with increased grazing, and decreasing with diet changes, and health issues apart from small increases in activity seen with heat events (labelled as illness 3). Note that activity decreasing when individual animals are isolated for a health reason (labelled as illness 2) could merely be indicative of their sudden change of environment. Observed data from the second sensor (Figure 4 (b)) also change, though in some situations to a lesser extent, with grazing, diet changes and health issues apart from heat, but changes in response to some of the health issues (1) tend to lag behind the changes seen for the first sensor. Slower and more gradual changes in response to management or health issues would be expected in sensor measurements such as liveweight or milk yield than in sensors that measure behaviour.
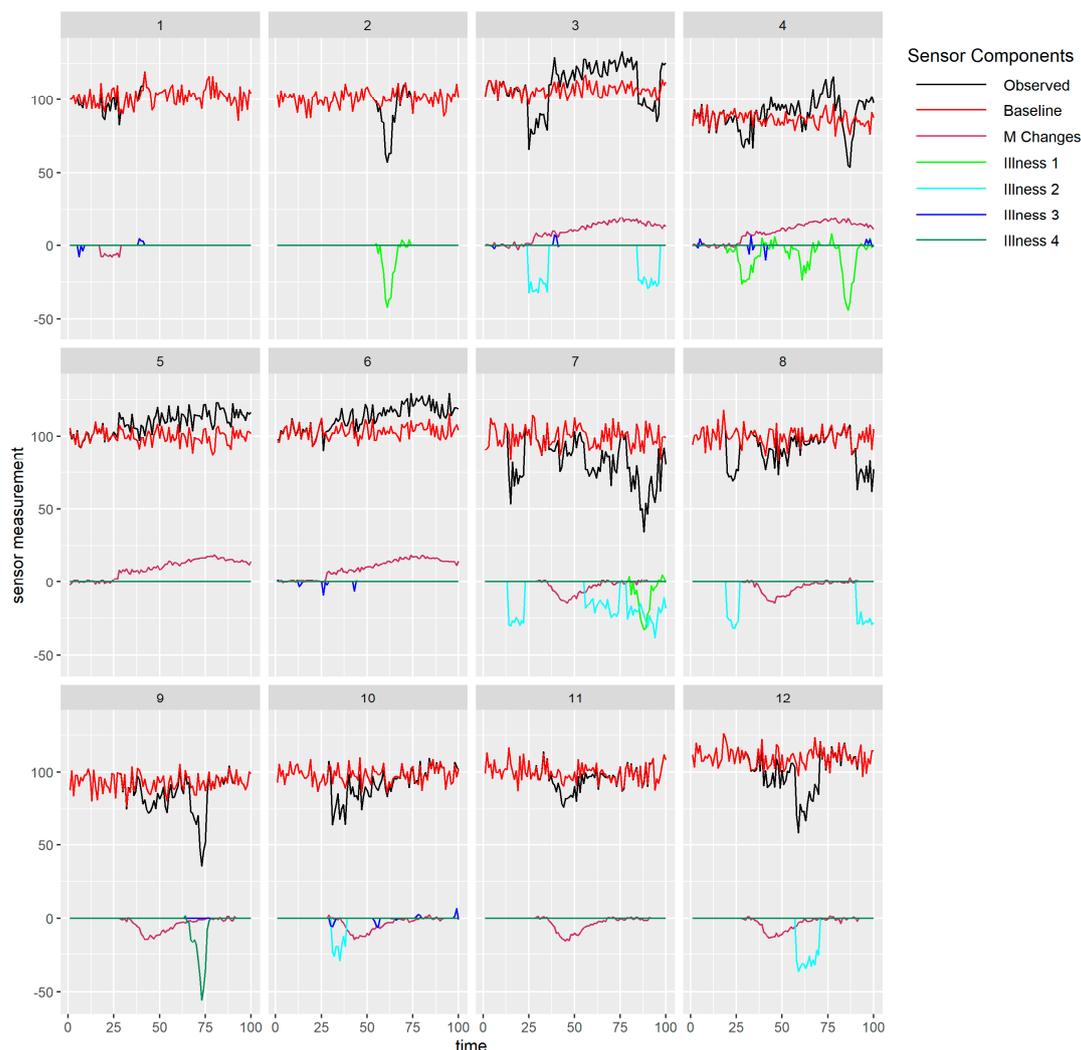
(a)
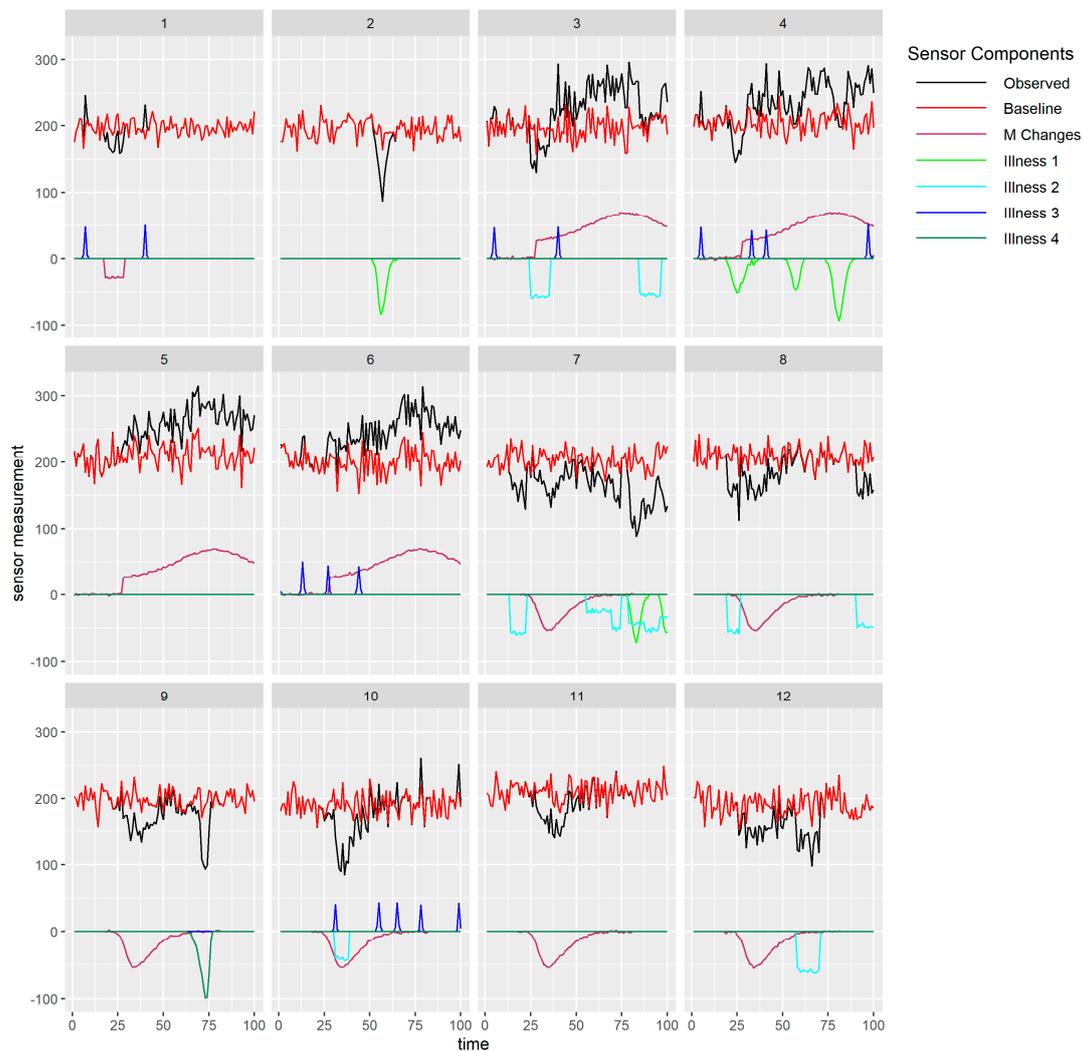
(b)

Figure 4. Simulated observed monitoring data (black) for two types of sensors, 1 (a), and 2 (b), together with components assumed to drive this data, for 12 animals in 3 management groups (1-2, 3-6, 7-12) for which 4 types of health issues occur. The main driver for observed sensor data is the baseline per animal (red), but observed sensor values may be altered by the occurrence of management changes or health issues. Animals in group 3-6 all show impact of seasonal grazing, those in group 7-12 all show the impact of a diet change which reduces levels for a few weeks before recovering to previous levels whilst the management change in group 1-2 only impacts animal 1. Sensor 2 data is impacted less by the management changes than sensor 1, and for the diet change there is a delay for sensor 2 compared to sensor 1. Illnesses vary in their impact on the monitoring data with illnesses 1 and 4 altering levels the most, followed by illnesses 2 and 3, and delayed impact on sensor 2 for illness 1. Levels are reduced for all illnesses apart from illness 3 which increases levels from sensor 1 and has negligible impact on sensor 2.

Some of these complexities may make it difficult to apply methods on farm, for example, methods need to cope with major changes that occur routinely, such as birth of offspring, or management changes, for example altered grazing patterns within the year, changes in feed or routine health care. Further, if sensor data from different animals have inherently different biases or underlying variability, this must be allowed for in any method. Any of the above methods used in such way that this is not made explicit, could be applied to derived data for each individual that standardises a priori for these differences. That is, data streams could be pre-processed in such a way

as to adjust for these differences. This could also make methods validated in one situation more transferable to another. In the case of supervised machine learning this could avoid the need to retrain the method for every new situation in which it is to be used, whilst in the case of statistical and latent variable/class modelling approaches this could result in simpler models where repeated fitting over time is computationally feasible. Statistical and latent variable/class modelling approaches, or machine learning methods, that explicitly allow for these added complexities, are intuitively appealing, but could be computationally intensive or result in non-identifiability (the method is too complex to achieve robust generalisable fits/estimation given available data). However, whether this is achieved via preprocessing adjustments or via direct modelling, one important thing to consider is whether a method would require all this information to be entered in real time when this is used on real farms as this could be impractical.

An alternative way of tackling major change events to avoid utilising this information, is to simply allow the method to include them (along with health & welfare issues) in the list of issues the method is intended to detect, or include them as additional latent states or latent variables (e.g. time grazing, time relative to calving), but this would clearly increase the complexity of these approaches. Furthermore, whilst intuitively appealing, this highlights that latent class or variable approaches, may need multiple classifications or variables to capture events such as these, as well as different types of health or welfare problems, making the underlying models far too complex. That said, a model that identifies different problems would be very useful for decision making.

One disadvantage of the simple approaches described above, or unsupervised statistical modelling approaches that simply model normal monitoring data in order to detect outliers from that, is that once abnormal monitoring data is detected, something must be done to ensure that it does not contribute to subsequent modelling of 'normal' monitoring data. Of course, it is feasible to do this so long as one is reasonably certain that the monitoring data is abnormal (and can exclude this on farm either manually, or preferably automatically), and so long as the modelling is being conducted in such a way that there remains enough information to estimate what is currently normal for all individuals. However, it is problematic to cope with this when an unmodelled state change has occurred for an individual (such as calving). Therefore, it is more appealing to build this into the model by using latent class or variable approaches, for example, which explicitly separate normal from abnormal states continuously, and in theory could cope with additional state changes such as calving, as well as the impact of latent variables such as time of gestation or time relative to calving.

One important distinction between the various methods described above is in whether they require independent data on the outcome of interest (e.g. whether animals have a health problem, or not) for their initial estimation. Of the methods above, supervised machine learning, and some statistical modelling approaches, requires this, but methods that use the monitoring data alone, (e.g. unsupervised machine learning, anomaly detection, latent class or variable modelling, and simple methods or classical statistical methods based on monitoring data only), do not. Of course, this difference is somewhat academic in the research development stage for a method, as at some point independent data on the outcome of interest will be required to validate its' performance (see section 5) and to check that it generalises to different situations. If generalisation is not possible methods will need to be reapplied to each new situation (for example refitting statistical models or retraining machine learning methods).

If estimated random effects (e.g. individual, farm) are needed to accurately predict outcomes for individuals from monitoring data (and cannot be adjusted for by preprocessing), then both statistical modelling and machine learning methods, would need to be refitted/retrained in every context they are to be used, in order to be able to make predictions for new individuals/farms. More generally, careful consideration needs to be given to whether the context in which a method is to be used is sufficiently similar to the context in which it is trained/developed.

One of the most important considerations of the more complex methods (advanced statistical modelling, latent class/variable modelling, machine learning), relates to computational difficulties for initial fitting/training, and then for any refitting/retraining that might be needed on farms. The more

complex methods based on more data information, may be more flexible/generic, but are likely harder to optimise, and possibly harder to apply on farm if information required is not readily available. If data preprocessing, and/or model simplification, are required for methods in order to make them tractable, this could result in a loss of flexibility. On the other hand, more explicit modelling assumptions would be expected to improve predictive accuracy so long as they are correct. Therefore, when comparing alternative methods is important not to attribute differences caused by different preprocessing or underlying assumptions and trade-offs, to method performance per se. Thus, for example, when input features for machine learning methods are derived via preprocessing to get around some of the complexities mentioned above, comparison should be made with other methods applied in the same way to the same pre-processed data.

The prediction/decision problem usually results in a binary decision of whether to check an individual (or group) or not. A more generalised approach would be to view this instead as a prediction, based on past data, of a numerical quantity for each individual at each time (which is correlated with the probability of a problem or the likely severity of a disease) which could be acted upon intelligently, depending on resources, for example. Whilst there are clear benefits to this more generalised approach, including that it may be more realistic, most research in this area treats this as a classification problem. Latent variable analysis is appealing as it addresses this directly. That said, most methods described above do include additional information, in that they result in numerical estimates that carry more information, such as the probability that an individual falls into different classes, on which decisions could be based.

Another important aspect is whether the methods can handle monitoring data from multiple types of sensors or not. One simple-minded approach is to combine measures on which decisions are based together into a single index somehow, as is commonly done for clinical scoring or quality of life measures. This is not recommended, however, as will likely mask important results in the individual measurements. If developed methods are based on data streams from each type of sensor, it could be relatively simple to combine the outcomes at the decision process stage. For example, if different sensors likely detect different health or welfare problems independently, a sensible approach would be to combine these by checking the animal if any of the monitoring data sources suggest there is a problem. However, if instead some aspect of the multi-variable space is needed to predict a problem, methods would need to operate in this space. In theory, both statistical modelling and machine learning methods could do this directly, though undoubtedly at the expense of increased computational load.

To summarise, there are various alternative methods that can be used for making predictions/decisions (e.g. to manually check and/or treat animals) based on past monitoring data of individuals (or groups) on farms, of varying degrees of complexity. The choice as to which to use in practice will need to take into account various aspects. Firstly, can they be used on farm completely automatically with no intervention, or will additional information/intervention be needed. Secondly, is it computationally feasible to use them to make decisions dynamically in real time on farm. And, finally, most importantly, are the predictions/decisions being made optimal. The next section discusses this latter issue, that is how to validate the predictions/decisions from different methods.

5. Validation of Predictions/Decisions

Having used a method to make predictions/decisions and/or identify anomalies or change points in real time as described in the previous section, the next step in development of these methods is to assess whether these coincide with real adverse events. We refer to this as *validation*. Viewed as a decision-making problem this is an assessment of the decisions that we make dynamically as the data evolves. A decision is clearly a classification – e.g. an individual should be manually checked as it is likely they are ill, or not. However, more broadly, as previously mentioned, it can be viewed as a comparison of predictions of the extent to which there is likely to be a problem, or estimated severity of a problem, from the monitoring data, with an independent measure over time of the extent of the true health or welfare for individuals (for example, the severity of the disease as measured by some

gold standard). This view of the problem is more realistic in many circumstances than the classification problem, because it is a recognition that real issues, such as disease, are usually not discrete events that happen in time, but instead likely increase over time, and then diminish again once treated. Furthermore, if the prediction is a number, rather than a classifier, the number itself could aid in decision making in real time, for example, as mentioned above, when resources are limited, they can be focused on those individuals most likely to have the most severe problems.

It is important that animals are checked when they have a problem, preferably in the early stages of the problem, but also that there are not too many false alarms. Optimal decision-making means striking a balance in the trade-off between these two situations, and where this balance lies may be context dependent. For example, it is more important to err on the side of caution for a serious contagious disease, than it is for less serious problems such as mild lameness. Such trade-offs could be formally quantified via a cost benefit analysis, assigning values to different outcomes such as cost of checking animals, benefit of identifying animals with specific illnesses and costs associated with missing ill animals, and then making decisions that minimise the total cost. This could also allow quantification of the benefit of sensor-based approaches compared to, or used in combination with, other approaches such as a regular health check plans/or preventative measures,

5.1. Challenges

Validation is usually not a simple problem, such as a measure of one sensor compared to measurement of the same thing from another as gold standard (e.g. [155]) but actually validation of the successive predictions of either a classifier (e.g. health event or not) or an estimate of probability or severity of an adverse event. Note that there is a parallel here with validation of diagnostic tests for disease [156] but it is more complicated than the usual approach taken for diagnostic testing for various reasons

- Instead of one prediction per sampling unit (e.g. animal) we have a sequence of predictions so there is the additional complexity of random effects/longitudinal data, which complicates calculation of simple validation measures akin to sensitivity and specificity;
- There is often no gold standard measurement (validation data) against which validation of predictions can occur;
- Where there is useful validation data for on farm studies, there are often lags between a problem starting and it being observed in the validation data. For example, the validation data collection may only detect disease when it becomes severe. Indeed, if the monitoring can detect pre-clinical disease, which is often the aim of auto-monitoring, this validation data would not be expected to coincide with detection by the monitoring data. Alternatively, very accurate validation may be possible in theory, but it is not practical to collect it locally in time in a study, particularly naturally occurring adverse events in long term on farm studies, for example, animals may be assessed just once a week;
- Furthermore, the true issue in which there is interest in predicting will often not suddenly occur at a single time step, but more likely will gradually increase over time, and then diminish, and much, or all, of this true process may go unobserved at the time at which it is occurring.

Figure 5, which shows simulated illness measures for four health issues (on which the simulated sensor data previously shown in figure 4 were based), illustrates some of the complexities mentioned above. In particular, it shows the marked distinction that likely occurs between observed illness measurement data available for validation and the true severity (one of the drivers of the sensor data).
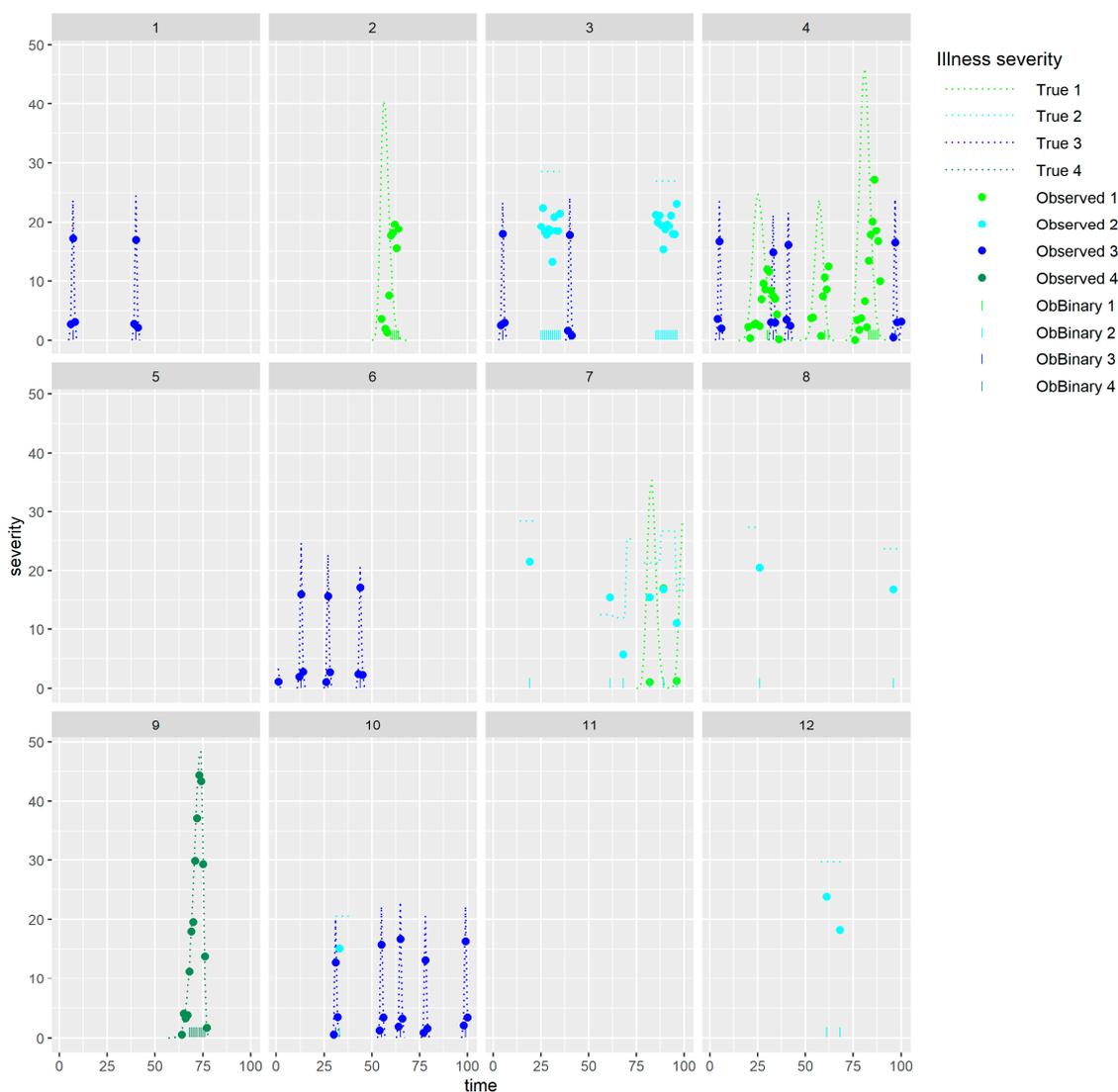
Figure 5. Simulated health issue severity data for 4 types of health issues for 12 animals in 3 management groups (1-2, 3-6, 7-12). The true (but unknown) health issue severity is shown as a dashed line (True). Illnesses may be gradual (e.g. illness types 1 and 4), or step changes (illness type 2), or very short (illness type 3). Observed severity is shown as points (Observed). Observations available for method validation will usually be imperfect, and may lag behind the truth (e.g. illness type 1). Observed severity could be measured in some situations but more commonly the only observed data will be a binary classification (ObBinary) which may only show up when illness is severe. Furthermore, some observations may only be made at intervals (e.g. illness types 1 and 2 in group 7-12 are only observed every 7 time units) and it will be difficult to validate predictions for rare illnesses (illness type 4 only occurs once for one animal here). .

Note that this issue of partial observation of the disease process does have parallels in validation of diagnostic tests. A binary diagnosis may lag behind the disease process and the true process in time for validating the test may not be observed in its entirety. Often a binary diagnosis is based on a continuous measure that has been collapsed for the purposes of decision making using some threshold into a binary estimate for positive/negative, when in fact the initial measure (e.g. optical density) may contain more information on the probability or severity of the disease. Translation of a measure to binary to simplify the process may be sub-optimal – it could be practically useful to instead make disease control decisions optimally based on a measure that carries more information.

Another challenge is how to decide on the temporal resolution at which to validate predictions from monitoring data. If the monitoring data has been summarised prior to applying any prediction

method, which may be necessary to make the prediction method computationally feasible, or to simplify the prediction method, then clearly the prediction validation cannot be at a finer resolution than that. Usually, the prediction validation will operate at the same, or a cruder resolution, than the available predictions. Pragmatically prediction validation may be limited to operate at the level at which the validation data is available.

Clearly, when there are multiple drivers of sensor data, it may be difficult to identify problems from the sensor data. For example, for animal 4 from our simulated data example (see Figure 6) it is difficult to visually identify many of the health issues (Figure 6 (b)) from the observed sensor data alone (Figure 6 (e) and (f)). Furthermore, the delay in the observed severity for illness type 1 (Figure 6 (b)) together with the lag (relative to the True severity) in the change in levels for sensor 2 (Figure 6 (d)) could lead to the incorrect conclusion that sensor 2 is better than sensor 1 at detecting illness type 1.
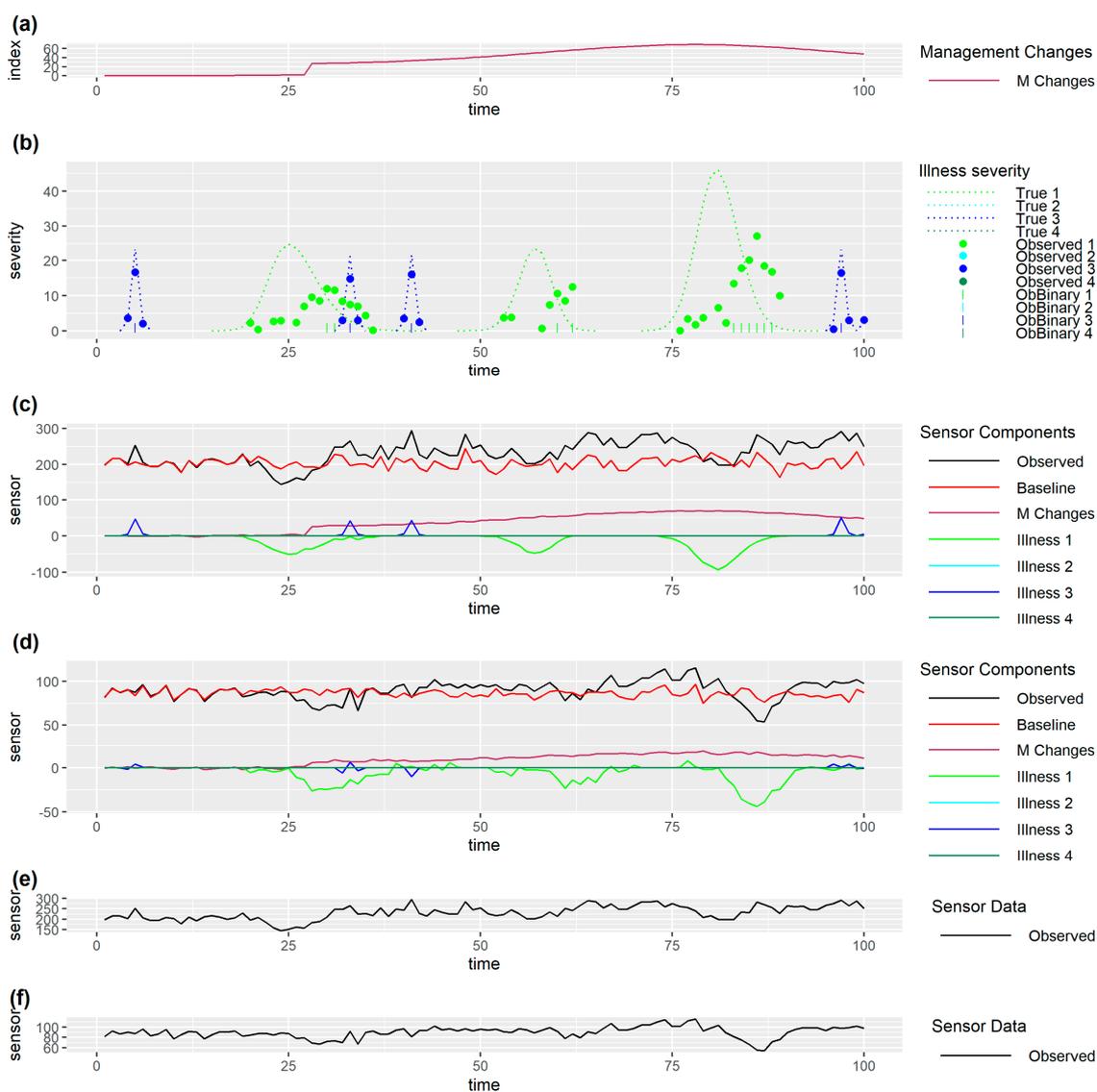

Figure 6. Simulated observed monitoring data (black) for two types of sensors, 1 (e), and 2 (f), together with hypothesized components (a)-(d) driving this data, for animal 4. Management changes (a) result in the overall gradual changes in sensor levels, whilst for health issues (b), illness type 1 results in shorter term changes, and illness type 3 even more short term changes. Many of these health events are barely perceptible from examination of the sensor measurements alone ((e) and (f)).

A lack of awareness of the prediction problem, or of availability of accurate timely validation data, and more generally the complexity as a result of the temporal nature of the prediction validation problem, are likely contributory factors resulting in a paucity of papers which properly tackle this to show that monitoring data and prediction methods are fit for use. Many papers [29,38,40,53,59,60,71,157,158] show associations of sensor measurements or estimates, or changes in them, with validation data which may be events or classifications in retrospect but very few address true prediction continuously in real time as described above along with rigorous statistical validation of these true predictions. Note that good association in a retrospective assessment is generally necessary but not sufficient for good prediction. There are, however, exceptions [30,41,46,52,58,72,87,91] that do address the accuracy of predictions made as the data evolves. In general, however, this is usually framed as a classification problem (often with just two classes), whereas, as mentioned above, this might be better viewed as detection of a process characterised by one (or more) numerical values, not of state changes.

5.2. Data Visualisation

Due to the complexities mentioned above, data visualisation is an essential first step in the methods development process for examining the relationship between available monitoring data, or predictions from monitoring data, and the available validation data. This allows examination of how these both evolve over time for individuals, especially when the two sources of data are at different temporal scales and/or to see any time lags. Data visualisation is also essential where there are multiple types of validation data (e.g. different health problems), and multiple sources of monitoring data that might detect one or more of these problems. That said, this is a retrospective analysis to examine the feasibility of approaches. But it is more flexible than retrospective quantitative assessments such as estimates of sensitivity and specificity, or goodness of fit of the model of a numerical value, which compare monitoring data with validation data usually at the same points in time, or with predefined windows, thus oversimplifying complexities in time lags.

Data visualization plots showing potential observed data for individual animals over times can be very informative, see, for example, Figure 7. As noted previously though, in many instances the only observed data on health issues will be binary classifications not severity measures, and measurements of management changes may not be available.
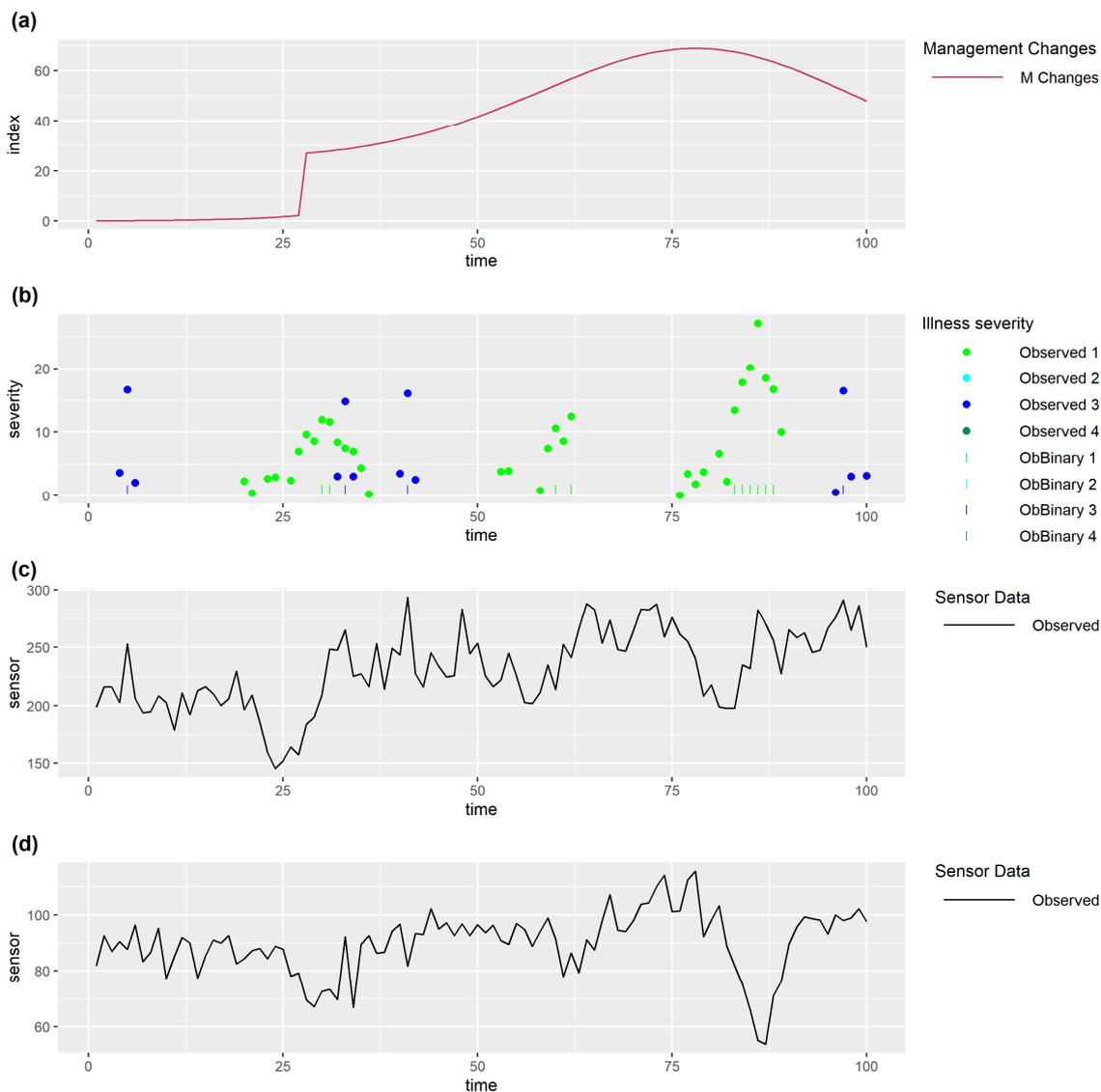
Figure 7. Simulated observed monitoring data (black) for two types of sensors, 1 (c) and 2 (d), for animal 4. Observed management changes (a) and possible observed illness data for 4 health issues (b) are also shown.

Data visualization plots that summarise over animals against times relative to the observed time of illness events are also useful and can be used for large data sets with many animals. However, time of observed illness events is often not well defined, as it could be taken as, for example, per animal

- the time of the first binary observation of an illness type
- the time of the first positive value of an observed illness severity measure
- the time of the maximum value of an observed illness severity measure

This could be plotted over all types of health issues, but since changes in sensor levels (and direction) may be dependent on illness type, patterns in plots would likely be clearer if each illness type is plotted separately. Figure 8 shows how sensor 1 levels increase at observed times of illness type 3 (which as mentioned previously would be expected for activity measurements from accelerometers in response to heat events) and Figure 9 shows how sensor 1 levels decrease ahead of observed times of illness type 1. Management changes are also shown on these graphs, though in practice on real farms these might not be recorded.
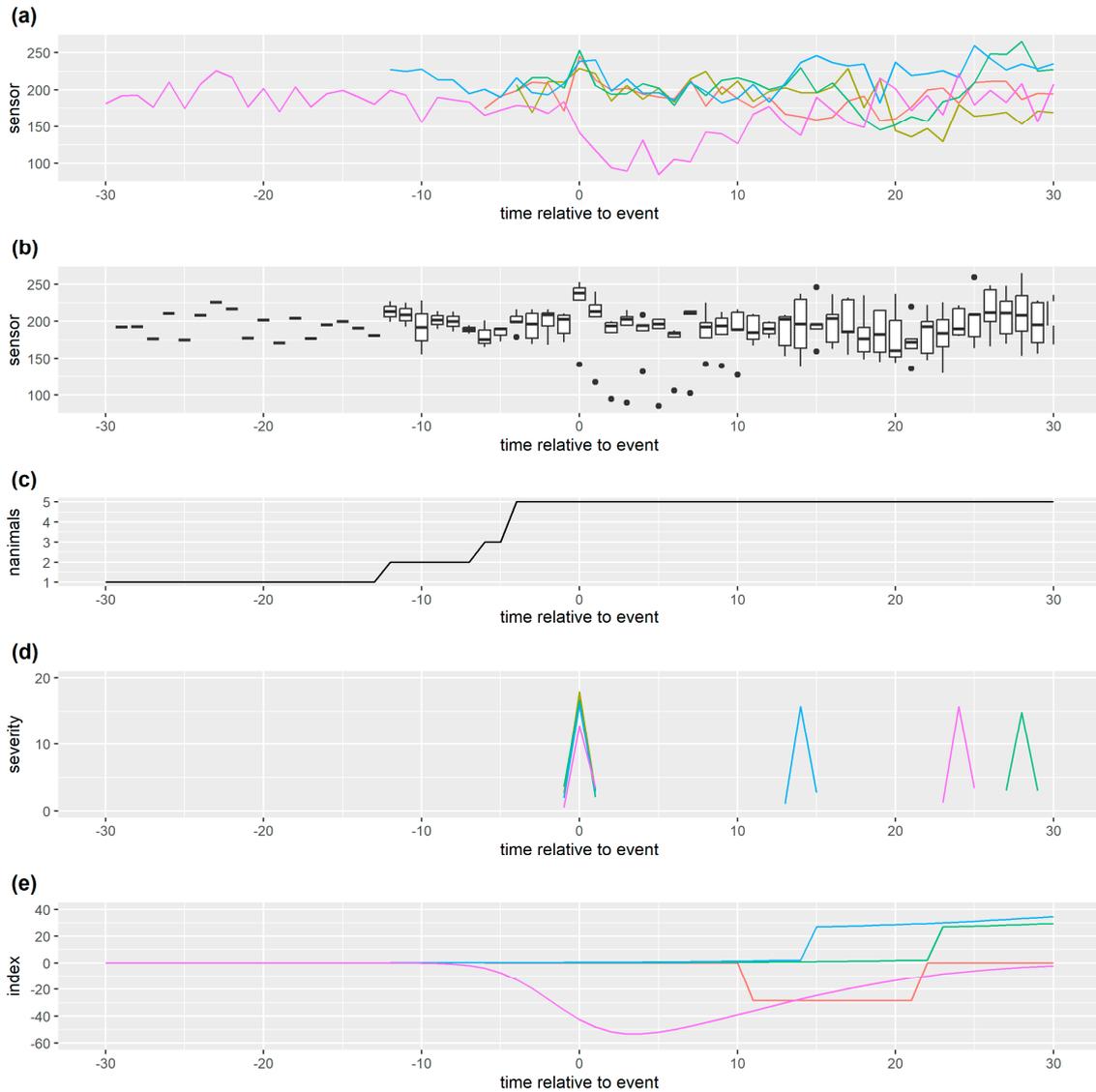
Figure 8. Simulated observed sensor 1 monitoring data shown in (a) line plots per animal and (b) box plots over animals, plotted against time relative to the first binary observation per animal of illness type 3. Management changes (e) and illness type 3 severity (d) are also assumed to have been observed. The number of animals contributing to this plot (c) is at most 5. Both the line and box plots show that sensor 1 levels rise coincident with time 0 apart from for one animal due to this being coincident with a reduction in the management index for that animal which is also driving the sensor data levels.
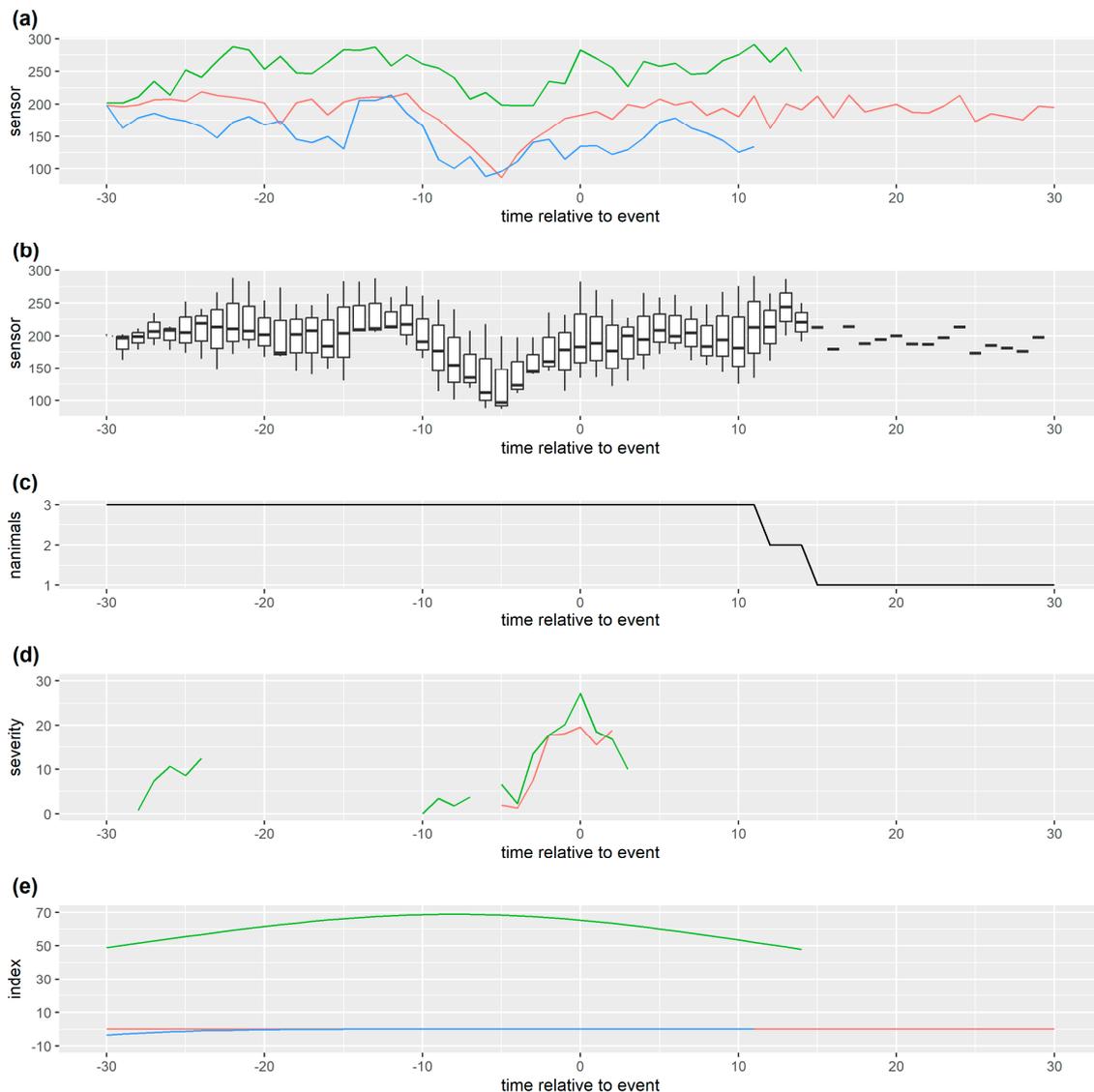
Figure 9. Simulated observed sensor 1 monitoring data shown in (a) line plots per animal and (b) box plots over animals, plotted against time relative to the maximum observed severity per animal of illness type 1. Management changes (e) and illness type 1 severity (d) are also assumed to have been observed. The number of animals contributing to this plot (c) is at most 3. However, both the line and box plots show that sensor 1 levels show a relative (within animal) decrease prior to time 0. As mentioned previously, this is because there is a lag (of about 5 time units) in the observed severity data compared to the (unknown) true severity.

5.3. Quantitative Assessment

Approaches that can be taken to a quantitative assessment of predictions/decisions depends on several details associated with the context, including

- the nature of the measurement on which the decisions are based;
- whether assessment is based on the data measurement or the decisions;
- the nature of the available validation data;
- the availability and accuracy of validation data;
- the structure of the data set, in particular the hierarchy (e.g. repeated measurements within individuals within farms).

When comparing and validating predictions from multiple methods, such as parametric statistical modelling versus machine learning methods, the same goodness of fit statistics should be used across methods to ensure a fair comparison. Below, we discuss various quantitative approaches

to validation. Observed data available for validation may itself not be an accurate reflection of the truth, but in the first instance we put this aside and assume the truth is known.

5.3.1. Classification

For testing the association of two 2 level classifications (decision to check individual based on monitoring data versus true positive/negative state), approaches used for diagnostic tests [156] are commonly used. This consists of evaluating the *sensitivity* (the percentage of true positives that are checked) and *specificity* (the percentage of true negatives that are not checked). If the decision is based on using thresholds, a value of the threshold is commonly chosen to result in equal sensitivity and specificity, though a Receiver operating characteristic (*ROC*) curve is usually also used, plotting the sensitivity versus 1-specificity to shows the trade-off between them. Other commonly used measures of agreement, which combine information over the classes, include the phi coefficient (i.e. Matthews correlation coefficient) [159], or for more than two classes Cohen's Kappa [160].

All these concepts clearly need modifications in order to use them appropriately in this context. For example, they need to take account of underlying structure. This could be done by averaging within individuals, and then across individuals (or farms).

Alternatives to these approaches exist, such as direct modelling by obtaining the goodness of fit from a model (GLM, say) [143,161] of the true binary outcome, as a function of the decision, or of the numerical variable on which the decision is based. These modelling approaches can easily be extended (GLMM, say) [142] to include appropriate random effects for individuals, farms and so on. When for example, a false negative is considered more costly than a false positive, a cost-benefit based approach could lead to more practical comparisons.

5.3.2. Severity

Correlations could be used to measure association of true severity with a numerical variable on which decisions are based. To incorporate structure the goodness of fit could be assessed of a LMM of the true severity, as a function of this numerical variable, along with appropriate random effects for individuals, farms and so on. Similarly, where the method only results in a classification (the decision), a LMM of the true severity, as a function of this decision (a categorical variable), along with appropriate random effects for individuals, farms and so on, could be used. When taking a modelling approach the usual goodness of fit methods, such as examination of the residuals, MSE and so on can be used, but it should be noted that these may be complicated by additional structure which may require careful use of weighting or averaging.

5.3.3. Time Lags and Other Temporal Considerations

The more difficult issue is how to deal with time lags appropriately. One approach could just be to extend the time window on which validations are based. For example, if the method is implemented with a daily time step, validation could instead take place over some moving window of several days. Alternatively, the recorded event could be extended or moved to the day(s) before, say, which is what is effectively done for 'block sensitivity' in [41]. There is an additional issue as mentioned in [41] which is that it could be that once a health event is recorded, treatment or management decisions in response may actually alter subsequent monitoring data. In any circumstance where recording of the event itself could result in immediate treatment or management, that could impact the monitoring data, it would only be acceptable to validate predictions of events before the recording event day. Moving events to the day before (and then ignoring subsequent data) is a pragmatic way of getting around this.

These approaches though, are fairly simplistic and involve making assumptions about likely lags. When dealing with a myriad of naturally occurring health issues on farm, it seems unlikely that such simplistic approaches would be adequate. Furthermore, they do not really address the issue of the availability and accuracy of validation data. A more elegant and generic approach to this could

be to use latent variable modelling. So here, the observed validation data, whatever and whenever it is recorded, could be assumed to be associated with a true unobserved latent variable per individual, and validation could take place instead against the estimated latent variable. However, it seems likely that different latent variables would be needed for different underlying problems. This, coupled with limitations on observed validation data (such as it only being observed at intervals and/or different types of underlying problems being rare), means that it is likely that some fairly strong assumptions that relate observed and latent variables would need to be made in order to estimate latent variables.

5.4. Cross Validation

Cross-validation is usually used for machine learning methods for which it is applied by splitting the available data into training and test data. The method is trained on the training data subset, and then validated on the test data subset [151,152]. This general principle of cross-validation is a sensible approach for all prediction methods [135]. For example, any method could be applied to the training data set, with model development, fitting and thresholds for decision making based on optimising goodness of fit, and then the goodness of fit evaluated on the test set using the same model and thresholds. Note that throughout, predictions made for each time step should be based only on past data. Cross validation could be used for any of the approaches to validation mentioned above whether based on classifications (e.g. decisions) or numerical estimates (e.g. numbers on which decisions will be based) and to any of the goodness of fit criteria, such as sensitivity, specificity or correlations.

Predictive accuracy can depend on the selected training and test set particularly when either of them is small. Therefore, both full retraining and testing should be applied repeatedly with multiple random selections of training and test sets. Both means and variation in performance should be reported to provide evidence that a method is consistently accurate. Random selections of training and test sets must be used, not selected based on optimizing performance as that leads to overfitting and underestimation of true prediction errors. A range of sizes of test set may be used, which generally varies from 50% of the data set to a minimum of a single data point as in leave-one-out cross validation. Care must be taken in the proportion chosen here. Too small a test set would be computationally burdensome to calculate enough repetitions to appropriately cover the data, and may pose difficulties in handling strongly correlated data. Large test sets may leave insufficient data to train the model and provide adequate performance. If a large number of methods are optimised and compared by cross validation, some differences between methods may arise from random chance, resulting in overfitting, so further independent validation of the best performing methods may be needed.

When carrying out cross validation it is crucial that selection of the training and test sets takes into account the structure of the data set. That is, selection should be between individuals, or groups, and, for data on multiple batches or farms, between batches or farms. This is to ensure independence of the test set from the training set, and to thus make sure that each individual validation is an unbiased estimate of performance, hence showing that something developed in one context will work in another. Carrying this principle to its natural conclusion, a method that has been developed and shown to work via cross validation only in one context (e.g. on one farm or type of farm) may not necessarily work in another.

5.5. On Farm Validation in Practice

Once a method has been validated based on available data as described above, additional research validation could take place on multiple farms, where the developed method is applied and then all the individuals that the method says should be checked at a point in time, are immediately given a full health check, along with some other randomly selected (non-check) individuals at regular intervals. This is a further test set in which the main difference is that more accurate timely targeted validation data could be collected. If the method involves making a decision based on some numerical

value, then sampling could be done to achieve coverage of that, either equally over its range, or in proportion to its distribution.

Another approach that could be undertaken, to validate use of monitoring methods on farms, could be to carry out large-scale studies involving many farms, where the methods are implemented to a strict protocol prescribed by earlier research. Usual experimental design principles could be applied, such as matched farms where each pair applies the monitoring method or not, or factorial designs (with blocking) where for example, alternative monitoring sensors by prediction methods by farming systems could be investigated. Then trends in usual production, health and welfare data measurements, and profit, could be analysed, to see whether those farms using particular monitoring methods show improvements over those that do not. This high-level approach, would get around many of the challenges mentioned above, but may be costly.

5.6. Other Considerations

The quantitative methods mentioned above could be used to compare multiple competing prediction methods for detecting the same health and welfare problems. However, when several monitoring methods are intended to be used for detection of multiple potentially overlapping types of health and welfare problems whether and how these should be reconciled at the validation stage needs to be considered.

6. Detailed Examples and Types of Studies

Here we discuss through examples the typical sorts of studies that could be used to develop and compare quantitative methods for real time auto-monitoring of livestock. The focus here is on the data collected and how it could be used, and associated challenges and other key aspects of prediction and prediction validation Our discussion of examples does not necessarily mention/critique specific approaches used in the cited publications. Such studies can be divided into three general categories: small scale clinical studies into specific health issues, studies that investigate a specific on farm or in field site over time, and studies across multiple farms.

6.1. Small Scale Clinical Studies for Specific Health Issues

These are experimental studies in which monitoring data are/could be collected on individuals along with regular detailed clinical measurements of specific health problems. These could either be naturally occurring problems or they could be studies where animals are challenged to cause specific problems.

Challenge studies [162–166] are often carried out to investigate disease progression, or the efficacy of vaccines or other treatments for specific illnesses in different livestock species. In [78] 20 grazing lambs were fitted with accelerometers in a cross over design in which half of them were challenged with saline and then with lipopolysaccharide a week later whilst the other half were challenged with lipopolysaccharide and then saline. Visual behaviour scans were also conducted on a subset of the lambs. A retrospective mixed models analysis suggested that challenge significantly impacted behaviour as measured by accelerometers providing evidence that accelerometers could be used to identify sickness in lambs. In [167] 27 piglets were unchallenged or challenged with *Salmonella enteritidis* or *E. coli*. Retrospective analyses showed that trends in live-weights and accelerometer data were affected by the challenges compared to the unchallenged control. In [166] 20 hens wearing accelerometers in one of 5 batches were challenged using a vaccine and followed over 12 days. Retrospective analyses using LMMs suggested activity changed with challenge day, and that there was an association between daily activity and daily clinical scores of hens, though it could not be ruled out that changes in activity behaviour were due to habituation of hens to the test setting. One disadvantage of challenge studies when all animals are challenged on the same time frame is that this could be confounded with other experimental effects, due to the environment or habituation of animals to the experimental setting.

Small scale clinical studies based on naturally occurring specific common health problems can also be used to validate sensors. In one such study [52,53] data was collected on 100 dairy calves with the expectation that 20% would develop Bovine Respiratory Disease (BRD). They were scored daily (age 8-42 days) for clinical signs whilst daily data was derived from sensors that continuously measured behaviour from accelerometers and feeding. Moving average and random forest methods were investigated based on different data inputs with validation of predictions based on a fixed window around observed (binary) health outcomes, to cope with time lags. Training and testing (cross-validation) was carried out once for the random forest method, with single estimates of a number of predictive accuracy measures given for each method. Prediction of pre-clinical BRD in dairy calves from feeding and pedometer data is also investigated in [51] where 54 out of 106 calves up to age 50 days developed BRD. Monitoring data was summarised daily, to obtain means and SDs over moving windows, which were then standardized. After some preselection of input features, K-nearest neighbour clustering was trained on data for days when calves were sick (defined by clinical scoring and scans), together with data omitting 2 weeks prior to the first sick day. Some cross validation was carried out though it does not mention whether selection of training and test sets are between calves, but the algorithm performs well in distinguishing sick days. The same algorithm applied to data 2 weeks prior to the first sick day, shows that sensitivity is quite good up to 6 days before the first sick day. In [168] daily clinical measures of BRD were made in 231 growing bulls in a 70-day growth and efficiency trial. A process control method (CUSUM) was used to predict BRD ahead of clinical diagnosis (which occurred between days 28-38 in 30 calves) from a range of daily derived measures based on feed monitoring of individuals. Thresholds were needed to give optimum sensitivity and specificity but no cross validation was carried out.

Small scale studies for specific health issues provide an ideal opportunity to investigate whether sensors can predict the progression of specific diseases as animals in these studies tend to be monitored continuously for clinical signs of the disease, thus providing ideal gold standard data for investigating time lags and prediction validation. Often one or more clinical measurements are recorded, and they can be numerical, ordinal, and/or categorical, so consideration needs to be given to how these should be handled for model fitting, training or validation. There are advantages in the simplicity of focusing on specific diseases with a high probability of occurrence in a well-controlled experimental set-up, though clinical studies based on naturally occurring specific health problems can be problematic as they rely on enough issues naturally occurring for methods development and validation. Predictive analysis and validation could be carried out for small scale studies so long as there are sufficient health events (naturally occurring or in response to challenge) and sufficient numbers of animals, and provided the independent gold standard clinical measure had been measured at a high enough temporal resolution to give long enough time series per animal. Where sample sizes are small data could be combined from multiple studies, provided they were sufficiently similar. However, the small amount of data available from many of these types of studies tends to limit analysis of individual studies to a retrospective one of examining associations. However, piggy backing by adding non-invasive sensors to small scale studies could be a very helpful inexpensive route for early methods development, prior to moving onto use of sensors in more realistic settings.

6.2. On-Farm and In-Field Studies

These studies follow animals at a realistically managed location for a given period of time, providing monitoring data and health data. The form of this health data varies greatly. Some of these studies focus on specific health issues whilst others are intended to predict a range of health events.

6.2.1. Dairy Farms

In [46] weekly clinical health scoring, and farm recorded daily health events, on ~950 calves (age 7-56 days, housed 15 calves/pen) from a 2,500 cow dairy herd, gave weekly, or daily calf status (sick or healthy) for a range of health issues. Feeding data was cleaned, preprocessed and various associated measures summarized at a daily level. These measurements were considered for the 6

days prior to a calf first becoming sick (with data thereafter omitted) and were supplemented with an equal sized data subset for days where calves were healthy. Training and testing (cross-validation) was carried out for 16 sets of input features by daily or weekly health data by 3 methods, generalised linear modelling and 2 machine learning methods, and single estimates of a number of predictive accuracy measures compared. One major conclusion is that predictive accuracy is substantially better for the daily health data, as expected as the weekly validation data is not timely.

In [41] data was collected over 40 months on 167 dairy cow lactations from a 65-cow herd. Input features were based on milking parameters, pedometer activity, feed and water intake, and body weight whilst target outputs were mastitis and lameness from treatment records shifted one day back. Extensive preprocessing was applied to monitoring data measurements, resulting in 471 potential input features, including daily, 3-day, and weekly summary statistics. Feature selection was used for mastitis and lameness separately leading to a subset of 20 prior to application of subsequent methods. For each health condition, the first health event per cow was considered, alongside a randomly selected day for healthy cows. Alternative sampling methods were also considered. A huge number of alternative machine learning methods were applied as well as ensemble methods that combine multiple methods. Ten-fold cross-validation was used, and predictive accuracy, including a version based on a short, fixed window size, given along with confidence intervals over sets of results using different classifications. This served to illustrate various points including that predictive accuracy could be very variable between methods and the detailed way in which they were applied, and more so for mastitis than lameness. Predictive accuracy varied between cross validations within methods, and improved when using a fixed window (due to time lags). Furthermore, many of the methods, including simple ones like logistic regression, appeared to be performing equally well on average, whilst some quite advanced methods were consistently worse.

The Langhill pedigree herd, based at SRUC's Crichton Royal Dairy Research Farm in Dumfries, Scotland, provides a key long-term data set for research into dairy cattle [169–171]. The herd has been selected for high and low genetic merit since 1973 and is usually maintained to have about 200 cows. The overall design of the study is a 2 x 2 factorial for genetic merit by management group, where 2 management regimes investigated change every 5 years or so in order to address objectives relevant to dairy farming. A huge amount of data are collected and maintained in a database by SRUC staff at the Dairy Research Innovation Centre and elsewhere. This includes much automatically collected data such as

- parlour data (2-3 times per day) from the automatic milking system, which includes milk yield, duration, peak flow and position in parlour
- live weight from walk over weighers on exiting the parlour
- lameness related measures from step sensor platform, Stepmetrix [172], on exiting the parlour
- feed and water intake from HOKOs [173]
- from 2013, behaviour at 15 minute intervals from IceCube accelerometers [174]

Manually collected data at intervals (e.g. weekly) includes body condition and lameness scores and milk composition and SCC. Detailed health records are made for all cows which includes all health issues and treatments. Reproductive, including heat and calving related information, is recorded and there is a detailed record of which group cows are in when throughout, as well as times of start and end of grazing, and local weather data. This is an exceptionally rich long-term data set for investigating the performance of quantitative methods of monitoring and predicting health and welfare problems as well as calving and oestrus. The presence of data from multiple sensors, potentially allows investigation of whether they can each be used to predict specific issues and combined to predict a range of issues. The comprehensive manual reporting system for health and welfare events for individual cows, is invaluable for prediction/decision validation. However, it also serves to illustrate many of the challenges that realistic on farm study data presents. One is that there are a multitude of recorded issues that could affect the sensor data for an individual simultaneously, including multiple health issues or reproductive events as well as management. In fact, the detailed records on management of all individuals serves to show the numerous adjustments (for example for

cows being managed in different groups and for time spent grazing) that would need to be made on an ongoing basis to allow robust predictions/decisions from sensor data. Note, that milk yield data, will not exist in the dry period prior to calving, and then will diminish with days in calving, and can be noisy or missing coincident with udder health problems. Another issue of note is that, when a cow is ill, she may be put in a hospital pen, interfering with the sensor data soon after the point in time at which the prediction is required. Furthermore, in common with many on farm studies, whilst the observed health data is extensive it is not always timely.

It is notable that the above 3 studies each relate to only one farm and in published studies [41,46] only a small window of available monitoring data was utilised per individual – just prior to selected event days. In [38] 10 dairy farms were visited every 30-42 days over ~1.5 years with lameness and body condition score (BCS) collected from all cows in milk (lactation numbers range 1-10) but only 374 had behaviour sensors (resulting in 2682 observations). Daily behaviour class time budgets including rumination, eating, rest, and activity measures, automatic milking system data, along with other cow and herd level measurements and risk factors including change in BCS, breed, days in milk and lactation number, herd management details, …were used to predict mild lameness the same day using random forests. Behaviour measurements were adjusted by herd and by cow means (which would not be possible for true prediction). Whilst this study is not continuous real time prediction per individual, it did serve to show that better accuracy was achieved from using behaviour, milk and additional farm information. A precision medicine study was conducted [158] on 22,923 observations of 5,829 animals on 166 dairy herds from 2014-16 focusing on multiple common health problems in dairy cows. This used a range of herd, cow (milk and BCS) and time level measurements giving 138 input features which were used with random forests to predict each disease type. Again, this study is not continuous real time prediction per individual, but served to illustrate that multiple input features were needed to predict health problems, with the sets of features varying between different types of problems. Another study [175] also focused on general health and welfare in which 318 cows were selected to give coverage of lactation numbers and days in milk from 6 dairy farms in 3 different countries. Gradient tree-boosting was used to classify animal welfare status (good,moderate,poor) based on behaviour from accelerometers and milk measures. Here numerous daily input features included gradients from linear regressions of monitoring data measurements over recent time windows of varying length per individual which serve to measure individual shortterm changes. Most importantly, these three studies showed that lower, and more variable, predictive accuracy resulted when cross validation was carried out between farms compared to between cows. [38] suggested this could be dealt with by applying methods to similar farms, or by including additional input features that characterize differences between farms.

6.2.2. Extensively Managed Cattle and Sheep

One practical issue with extensive systems is that often the animals need to be gathered to download sensor data. In order to use these sensors in real time for predicting health and welfare problems, the data would need to be downloaded remotely, and the prediction algorithms applied, in real time, as the data evolves. Because of the practical difficulty of real-time monitoring of individual animals in extensive systems, and of obtaining real time validation data, there are few studies that directly address real time prediction for grazing beef cattle and sheep. One exception is [58] in which a range of sensors (accelerometers, GNSS, rumination algorithm, walk over weight unit enclosing sole watering point, and weather) were investigated to detect calving in 40 cows in a 32 hectare paddock. Calving was measured every 2-3 hours from 06:00 to 18:00. Limited conclusions can be drawn from this study, as a really large number of input features were used as candidates to detect just 9 calvings, utilizing the 7 days data prior to the calving events. 9 cross validations were carried out leaving one cow out each time and up-sampling the 8 calving events during training of a SVM and mean predictive accuracy was given. However, the approach used here does serve to illustrate various points of interest. Weather data was included not because it was expected to affect day of calving but because it was expected to affect monitoring data and hence the relationship between that

and day of calving. Predictions were investigated with both a daily and hourly time step, but daily performed better presumably because calving events were not recorded accurately enough to validate hourly monitoring. Multiple input features were derived for each day from each sensor type, with only those calculated for the current day used in daily prediction of calving. However, these were used in 3 ways: unadjusted, adjusted to the herd mean the same day (after removing upper and lower quartiles) and adjusted to the data on the previous day from the same individual. These adjusted measures came in strongly illustrating the need to adjust for herd level changes (such as weather and management) and that changes within individuals are an important predictor of calving. Note, however, that in order to predict health and welfare events where changes are gradual, rather than sudden, information from longer term changes within individuals would need to be included in the input features.

A similar approach is taken in [72] in which GNSS, accelerometers and weather data are used to predict lambing for ewes in the pasture. This is based on training and tuning an SVM based on one 14 day study using leave out one ewe cross validation (out of 8 observed lambings) and then validating the model by examining alarms generated in another 15 day study with 11 observed lambings. Extensive preprocessing of monitoring data was described in detail, as well as how pertinent hourly measures were derived for GNSS and accelerometer data, which were again unadjusted, or adjusted to flock means at the same time, or the same measures for each ewe 1 hour and 24 hour before. 4 key input features were used in the final model, three from GNSS data (mean distance to peers adjusting to herd mean, distance to closest peer, mean distance to peers) and one (counts of posture changes) from the accelerometer data. Whilst alarms were generated around most lambings in the validation data set, there were quite a lot of false alarms prior to lambing which leads the authors to point out the need for alarm thresholds to be varied depending on the context in which this monitoring is to be used in order to reach the optimum trade-off between sensitivity and specificity. To examine why there are classification errors graphs for some ewes show how monitoring input features change with time relative to lambing. This data visualisation is highly informative though strictly speaking machine learning methods operate on the multiple dimension space of input features.   Other points mentioned are that monitoring data was not actually available in real time in these studies, and the need to limit the number of input features to make the ML algorithm computable in real time. This method was further validated on a daily basis using another study [77] of 14 ewes where alarms were first generated on the day of lambing for all of them apart from one ewe with vaginal prolapse for which several alarms occur on the days leading up to lambing.

In [74] accelerometers were used to infer grazing ewe behaviours using 121 ewes, and the huge change in (estimated) licking behaviour assumed to indicate lambing day for individual ewes. Then a deep learning neural network was applied to the sensor data at each hour prior to lambing with 30 input features which were summary statistics over the last 7 days with (estimated) time to lambing at that hour as the target output variable. This study highlights the difficulty of getting the true lambing time in order to validate predictions for extensive systems, and thus uses the same sensor data set to estimate lambing day as it does to predict time to lambing. That said, it suggests that time to lambing can be predicted to within 20 hours accuracy up to 10 days before lambing, though cross validation is not mentioned.

A long-term sheep flock study on the use of technologies for health and welfare monitoring is being carried out at a research farm at the Moredun Research Institute [176–178]. Technologies include EID for identification, accelerometers [179], GNSS [180,181], and BLE proximity loggers, with a subset of ewes and their lambs wearing this technology during several periods in 2021, 2022 and 2024. Objectives of this study include assessing PLF tools for early detection of parasites, mastitis and lameness in a flock during summer grazing as well as which sensors are usable by sheep practically speaking, and whether cheaper sensors can be used instead of more expensive ones. Health and welfare assessments of ewes and lambs were made at intermittent gathers about every two weeks including walk over weighing, body condition, mastitis and dag scores, faecal strongyle and

nematodirus egg counts, and treatments recorded. Whilst this is potentially a rich data set for assessing prediction of a range of health and welfare problems using a range of sensors in the field it presents various challenges. If particular sensors, or combinations of them, are able to predict specific health or welfare issues, we are reliant on enough instances of these issues naturally occurring in order to be able to assess predictions. This is compounded by the fact that only a subset of sheep and lambs were wearing technology, and they are only wearing it continuously for some periods. Furthermore, health and welfare data, though wide ranging, is not available continuously in real time, with many assessments coinciding with periods when technology was not worn in the period at or immediately preceding the assessment. These issues preclude validation of predictions in real time, but investigation of the potential for real time prediction from sensors of the recorded health and welfare problems is feasible using a retrospective analyses.

6.2.3. Pigs & Poultry

Methods used for monitoring for pigs tend to be mainly for use in housed systems, and often locational sensors are used, especially when monitoring groups of growing piglets. An exception to this is as study [91] in which accelerometers on sow ears were used for predicting farrowing for ~20 sows in crates. Of particular note here was the preprocessing which involved using measures of variance in the total acceleration for predefined window sizes, which were then smoothed and then adjusted by the same measure 24 hours before, effectively accounting for any underlying sow and sow dependent diurnal effects. Then a CUSUM method was used to generate alarms. This worked fairly well but it should be noted that this study was carried out on few pigs, with no cross validation. A fairly low-tech solution which obtained feeding and drinking behaviour (but not amounts) based on RFID tag readers at feeders and drinkers, was used [87] to predict deaths/culls in 3 batches of 12 pens of 12 growing finisher pigs followed over some weeks. After some preprocessing six daily measurements were derived for feeding and for drinking for each pig. For each measurement a Kalman filter was used to estimate each pigs' behaviour based on previous days and adjusted to pen behaviour the same day and then anomalous behaviour, and hence alarms, identified if the observed values exceeding confidence intervals. This resulted in high specificity but there were too few deaths/culls (just 4-5 per batch) to evaluate sensitivity. Other recent studies are more developmental, for example based on image or sound sensors, and whilst they suggest potential, they did not really address the problem of validating real time predictions of health and welfare problems.

Similarly, methods for poultry tend to be for housed flocks. In commercial flocks locational sensors are used to measure groups or unidentified individuals, intended to be a representative subset of a flock. There is a paucity of studies that validate predictions of health and welfare problems for flocks made continuously in real time from monitoring data, though some studies do show association retrospectively of data continuously derived from sensors, with health and welfare outcomes in broilers. In [116] an experiment on 250 birds in 5 batches with a range of gait scores showed that walking speed automatically measured when they walking down a runway decreased with increasing gait score. In [120] whole house sound recording of 12 commercial flocks showed that spectral entropy was negatively related to distress call rates and to both current and more long-term weight gain and mortality for flocks. In [119] cameras in 31 UK and 43 Swiss flocks measured optical flow which was shown to be related to hock burn and mortality.

6.2.4. Summary

Overall, on-farm studies are advantageous as they are more realistic and larger than small scale clinical studies, however they generally present more practical limitations on whether records of health issues are accurate and/or timely, and allowances have to be made for specific management practices that were not designed with easy and reliable analysis in mind. While this can be partially addressed with appropriate analysis measures, these may lead to failures of results to generalize. For some situations, practicalities can also be much more challenging, limiting what monitoring systems and regimes can be used.

6.3. High Level Validation Studies

We could not find any properly designed high level on farm intervention studies for multiple farms, nor even within farms, to examine the impact of using sensors for continuous real time monitoring on farm. In [182] survey data was compared between dairy farms that did not use any sensors and did use sensors (for monitoring at least milk yield and oestrus). Although they selected from this survey 50 farms that used sensors, and 50 with matched characteristics that did not, few differences in performance measures were found, and in any case, differences could be due to other factors as this is an observational study. [183] describes a 3 year study in which a 900 hill sheep flock [184] is split into 2 groups managed with and without precision farming methods and differences in economics, animal performance and farm labour examined. They did find some differences, but the precision farming was not utilising the sensors continuously in real time for decision making.

Intervention studies on multiple farms would be ideal for establishing whether or not real time monitoring with sensors, along with methods for prediction/decisions, can really be impactful on real farms. One approach could be to generate alarms based on sensor data and developed methods, ensuring that for each real alarm some form of matched false alarm was given. The farmer/researcher (blind to whether the alarm is true or false) could then check the animal and record any health issues which could be supplemented with farm records on health issues. This could allow sensitivity and specificity to be estimated provided enough data in each class (true positive, true negative, false positive, false negative) could be collected. However, this would likely be very labour intensive. A more feasible approach would be to assess impact based only on routinely collected health, welfare and production data, provided the interventions (how the monitoring is to be installed and utilised) could be well defined in protocols and farmers adhered to them. Sound experimental design principles to optimize studies should be used such as applying different interventions (e.g. no sensors, sensors with completely sensor data driven method, sensors plus methods that use additional farm level data, and so on) randomly between matched farms, or different interventions being applied within farms to different groups of animals and/or in different periods. Baseline measures could be made, so that analysis showed the impact after, relative to before, interventions. More sophisticated designs such as factorials or fractional factorials could be used to examine differences as well as synergies in the use of different sensors and prediction/decision methods.

7. Conclusions

There are a huge number of studies, and review papers, relating to precision livestock farming, and much research development on quantitative methods to use sensor data streams on individuals or groups for prediction and decision making in real time on livestock farms. However, detailed scrutiny is needed to ascertain exactly how the monitoring data and these methods are used and validated in published papers. It is difficult to draw on the literature on multiple studies to make fair comparisons between recommended quantitative methods, because numerous divergences likely impact their performance. This includes differences between data sets (overall design, monitoring data and types of health issues), data preprocessing and how the processed data is used (both inputs and outputs), how quantitative prediction methods are implemented (e.g. underlying models and how they are fitted and predictions obtained, how parameters are set or tuned for machine learning methods), how decisions are made, and how predictions/decisions are validated. Even within the same paper quantitative methods are often applied and/or evaluated differently. Furthermore, even when there is reliable validation data available, there is a lack of evidence that quantitative prediction methods validated in one circumstance will work in another situation. In most of the studies reviewed here that investigate prediction/decision making in real time, quantitative methods are applied on a single farm whilst the few multiple farm studies reviewed showed predictive accuracy is poor when methods developed are validated between farms. It is plausible that other studies which show poor performance when validated between different farms or farming systems remain

unpublished, though the amount of work involved in such studies means it is likely that they are not commonly undertaken.

The amount and temporal resolution of monitoring data per individual that is used in each prediction will impact the results as will the way in which this monitoring data is used. Therefore, it is important to make comparisons between different choices made at this stage and to recognise when comparisons of alternative prediction methods are confounded with this aspect. For example, statistical methods tend to be based on the full (or some recent window) past unprocessed time series data, whilst machine learning methods in papers examined here were often applied to input features that are summary statistics of past time series data per individual, possibly also adjusted by group effects. This latter approach is sensible for any method as it can be used as a simple way to adjust individual measurements for group effects (for example herd or flock management effects), for effects of individuals and/or to capture recent short term changes likely to be indicative of a problem. It also likely reduces the computational load of methods, and it could make the prediction methods more applicable in different situations. However, when any method uses these summary statistics as inputs, some quite strong assumptions are being made about what properties of the data are relevant, and the performance of the method will largely depend on the validity of these assumptions. Furthermore, the intuitive appeal to the uninitiated of machine learning is that it is flexible and can use large data sets to provide outcomes with few explicit assumptions and so ideally machine learning methods applied directly to (unprocessed) time series data should be investigated.

When researching quantitative methods to make predictions/decisions in real time, careful thought needs to be given as to whether the methods are being used in exactly the same way that they could be applied in real time for decision making on real farms. As well as taking into account predictive accuracy other aspects need to be considered, in particular, the advantage of only using monitoring data versus the need to use additional data on farms, groups or individuals, the computational load of methods, and whether they need to be refitted or retrained in every new situation and/or for particular farms, groups or individuals. Many research development studies focus on one or more specific monitoring data streams for specific types of health or welfare problems. However, when used on real farms, some monitoring data may be subject to changes due to multiple types of problems and so the issue of sensitivity and specificity for multiple problems needs to be assessed. Some studies do develop models for general health and welfare problems, but it is highly probable that different sensors and/or predictive models would be required for different types of health and welfare problems. Furthermore, sensor data, and in particular behaviour data, will be influenced also by changes in farm management, for example increased grazing, or moving animals to different groups over time, and methods will need to account for this. Taking into account as much of an individuals' past monitoring data as possible, and also sharing past information between individuals, would be expected to improve predictive accuracy, but this will be limited in practice by computational load.

All that said, regardless of all these details, a method is viable, so long as it can be shown that it is practically feasible to repeatedly apply the quantitative prediction method dynamically in real time on farm optimising decision making along the way, in any context in which it is to be used. However, method evaluation, or prediction/decision validation, is also challenging. A range of measurements should be used to quantify the accuracy of predictions which could include predictions of probability or severity as well as binary classifications for multiple health and welfare problems. For most data sets, accuracy needs to be measured appropriately to get around imperfect validation data, in particular to allow that the monitoring method may detect the problem before it is recorded in the validation data. This is not straight forward, since the appropriate window size to use for this will likely depend on the type of problem, its severity, the recording method and the monitoring data used.

Cross validation is needed and should be applied to all quantitative methods in the same way, but how this is applied in detail is very important. Predictive accuracy can depend on the selected training and test set particularly when either of them is small. Therefore, both full retraining and

testing should be applied repeatedly with multiple random selections of training and test sets and means and variation in performance should be reported. Variation needs to be low to provide evidence that methods are consistently reliable. To show methods generalise, cross validation should be carried out with training and test sets selection made between individuals, or groups (e.g. batches or farms) where data is on multiple groups and spatial locations. Ultimately, methods need to be validated across many farms to show that they are practical for use on farms and to show any benefit from using them. Conventional experimental designs could be applied in multiple farm studies where the real impact of using sensors plus associated quantitative methods to generate alarms along with strict protocols for decision-making could be evaluated based on their effect on usual farm animal performance, health and welfare measurements as well as profit.

In summary, there is the potential to massively increase use of empirical data for decision making in real time, but development and validation of quantitative approaches is still needed in order to move from the developmental stage to this intended practical application. Whilst there is wide-ranging literature in this area, in livestock farming there is a paucity that provides fair robust statistical comparisons of alternative quantitative methods and evidence that resulting decision making performs adequately in practice. That is, it must be practically feasible to repeatedly apply the method dynamically in real time on farms, and provide information of direct relevance to decision making.

References

1. Dayioğlu, M.A.; Türker, U. Digital Transformation for Sustainable Future-Agriculture 4.0: A Review. *Tarim Bilimleri Dergisi* 2021, *27*, 373–399.
2. Monteiro, A.; Santos, S.; Gonçalves, P. Precision Agriculture for Crop and Livestock Farming—Brief Review. *Animals* **2021**, *11*, doi:10.3390/ani11082345.
3. Karunathilake, E.M.B.M.; Le, A.T.; Heo, S.; Chung, Y.S.; Mansoor, S. The Path to Smart Farming: Innovations and Opportunities in Precision Agriculture. *Agriculture (Switzerland)* **2023**, *13*, doi:10.3390/agriculture13081593.
4. Karunathilake, E.M.B.M.; Le, A.T.; Heo, S.; Chung, Y.S.; Mansoor, S. The Path to Smart Farming: Innovations and Opportunities in Precision Agriculture. *Agriculture (Switzerland)* **2023**, *13*, doi:10.3390/agriculture13081593.

5. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* **2021**, *21*, doi:10.3390/s21113758.

6. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873, doi:10.1109/ACCESS.2020.3048415.

7. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput Electron Agric* **2018**, *147*, 70–90, doi:10.1016/j.compag.2018.02.016.

8. Jiang, B.; Tang, W.; Cui, L.; Deng, X. Precision Livestock Farming Research: A Global Scientometric Review. *Animals* **2023**, *13*, doi:10.3390/ani13132096.

9. Zhang, M.; Wang, X.; Feng, H.; Huang, Q.; Xiao, X.; Zhang, X. Wearable Internet of Things Enabled Precision Livestock Farming in Smart Farms: A Review of Technical Solutions for Precise Perception, Biocompatibility, and Sustainability Monitoring. *J Clean Prod* **2021**, *312*, 127712, doi:10.1016/j.jclepro.2021.127712.

10. Rosa, G.J.M. Grand Challenge in Precision Livestock Farming. *Frontiers in Animal Science* **2021**, *2*, doi:10.3389/fanim.2021.650324.

11. van Erp-van der, E.; Rutter, S.M. Using Precision Farming to Improve Animal Welfare. *CABI Reviews* **2020**, *2020*, doi:10.1079/PAVSNNR202015051.

12. Bailey, D.W.; Trotter, M.G.; Tobin, C.; Thomas, M.G. Opportunities to Apply Precision Livestock Management on Rangelands. *Front Sustain Food Syst* **2021**, *5*, doi:10.3389/fsufs.2021.611915.

13. Tobin, C.T.; Bailey, D.W.; Stephenson, M.B.; Trotter, M.G.; Knight, C.W.; Faist, A.M. Opportunities to Monitor Animal Welfare Using the Five Freedoms with Precision Livestock Management on Rangelands. *Frontiers in Animal Science* **2022**, *3*, doi:10.3389/fanim.2022.928514.

14. Aquilani, C.; Confessore, A.; Bozzi, R.; Sirtori, F.; Pugliese, C. Review: Precision Livestock Farming Technologies in Pasture-Based Livestock Systems. *Animal* **2022**, *16*, doi:10.1016/j.animal.2021.100429.

15. Neethirajan, S.; Kemp, B. Digital Livestock Farming. *Sens Biosensing Res* **2021**, *32*, 100408, doi:10.1016/j.sbsr.2021.100408.

16. García, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodríguez, P. A Systematic Literature Review on the Use of Machine Learning in Precision Livestock Farming. *Comput Electron Agric* **2020**, *179*, 105826, doi:10.1016/j.compag.2020.105826.

17. Neethirajan, S. The Role of Sensors, Big Data and Machine Learning in Modern Animal Farming. *Sens Biosensing Res* **2020**, *29*, 100367, doi:10.1016/j.sbsr.2020.100367.

18. Tedeschi, L.O.; Greenwood, P.L.; Halachmi, I. Advancements in Sensor Technology and Decision Support Intelligent Tools to Assist Smart Livestock Farming. *J Anim Sci* **2021**, *99*, doi:10.1093/jas/skab038.

19. Neethirajan, S. Artificial Intelligence and Sensor Technologies in Dairy Livestock Export: Charting a Digital Transformation. *Sensors* **2023**, *23*, doi:10.3390/s23167045.

20. Morota, G.; Ventura, R. V.; Silva, F.F.; Koyama, M.; Fernando, S.C. Big Data Analytics and Precision Animal Agriculture Symposium: Machine Learning and Data Mining Advance Predictive Big Data Analysis in Precision Animal Agriculture. *J Anim Sci* **2018**, *96*, 1540–1550, doi:10.1093/jas/sky014.

21. Neethirajan, S. SOLARIA-SensOr-Driven ResiLient and Adaptive MonitoRIng of Farm Animals. *Agriculture (Switzerland)* **2023**, *13*, doi:10.3390/agriculture13020436.

22. Neethirajan, S. Artificial Intelligence and Sensor Innovations: Enhancing Livestock Welfare with a Human-Centric Approach. *Human-Centric Intelligent Systems* **2023**, doi:10.1007/s44230-023-00050-2.

23. Agrawal, S.; Ghosh, S.; Kaushal, S.; Roy, B.; Nigwal, A.; Lakhani, G.P.; Jain, A.; Udde, V. *Precision Dairy Farming: A Boon for Dairy Farm Management*; 2023; Vol. 8;.

24. Kleen, J.L.; Guatteo, R. Precision Livestock Farming: What Does It Contain and What Are the Perspectives? *Animals* **2023**, *13*, doi:10.3390/ani13050779.

25. Michie, C.; Andonovic, I.; Davison, C.; Hamilton, A.; Tachtatzis, C.; Jonsson, N.; Duthie, C.A.; Bowen, J.; Gilroy, M. The Internet of Things Enhancing Animal Welfare and Farm Operational Efficiency. *Journal of Dairy Research* **2020**, *87*, 20–27, doi:10.1017/S0022029920000680.

26. Knight, C.H. Review: Sensor Techniques in Ruminants: More than Fitness Trackers. *Animal* **2020**, *14*, S187–S195, doi:10.1017/S1751731119003276.

27. Szenci, O. Accuracy to Predict the Onset of Calving in Dairy Farms by Using Different Precision Livestock Farming Devices. *Animals* **2022**, *12*, 2006, doi:10.3390/ani12152006.

28. Santos, C.A. dos; Landim, N.M.D.; Araújo, H.X. de; Paim, T. do P. Automated Systems for Estrous and Calving Detection in Dairy Cattle. *AgriEngineering* **2022**, *4*, 475–482, doi:10.3390/agriengineering4020031.

29. Antanaitis, R.; Anskienė, L.; Palubinskas, G.; Džermeikaitė, K.; Bačėninaitė, D.; Viora, L.; Rutkauskas, A. Ruminating, Eating, and Locomotion Behavior Registered by Innovative Technologies around Calving in Dairy Cows. *Animals* **2023**, *13*, doi:10.3390/ani13071257.

30. Borchers, M.R.; Chang, Y.M.; Proudfoot, K.L.; Wadsworth, B.A.; Stone, A.E.; Bewley, J.M. Machine-Learning-Based Calving Prediction from Activity, Lying, and Ruminating Behaviors in Dairy Cattle. *J Dairy Sci* **2017**, *100*, 5664–5674, doi:10.3168/jds.2016-11526.

31. Keceli, A.S.; Catal, C.; Kaya, A.; Tekinerdogan, B. Development of a Recurrent Neural Networks-Based Calving Prediction Model Using Activity and Behavioral Data. *Comput Electron Agric* **2020**, *170*, doi:10.1016/j.compag.2020.105285.

32. Horváth, A.; Lénárt, L.; Csepreghy, A.; Madar, M.; Pálffy, M.; Szenci, O. A Field Study Using Different Technologies to Detect Calving at a Large-Scale Hungarian Dairy Farm. *Reproduction in Domestic Animals* **2021**, *56*, 673–679, doi:10.1111/rda.13904.

33. Crociati, M.; Sylla, L.; De Vincenzi, A.; Stradaioli, G.; Monaci, M. How to Predict Parturition in Cattle? A Literature Review of Automatic Devices and Technologies for Remote Monitoring and Calving Prediction. *Animals* **2022**, *12*, doi:10.3390/ani12030405.

34. Steensels, M.; Antler, A.; Bahr, C.; Berckmans, D.; Maltz, E.; Halachmi, I. A Decision-Tree Model to Detect Post-Calving Diseases Based on Rumination, Activity, Milk Yield, BW and Voluntary Visits to the Milking Robot. *Animal* **2016**, *10*, 1493–1500, doi:10.1017/S1751731116000744.

35. Higaki, S.; Matsui, Y.; Sasaki, Y.; Takahashi, K.; Honkawa, K.; Horii, Y.; Minamino, T.; Suda, T.; Yoshioka, K. Prediction of 24-h and 6-h Periods before Calving Using a Multimodal Tail-Attached Device Equipped with a Thermistor and 3-Axis Accelerometer through Supervised Machine Learning. *Animals* **2022**, *12*, 2095, doi:10.3390/ani12162095.

36. Van Hertem, T.; Bahr, C.; Tello, A.S.; Viazzi, S.; Steensels, M.; Romanini, C.E.B.; Lokhorst, C.; Maltz, E.; Halachmi, I.; Berckmans, D. Lameness Detection in Dairy Cattle: Single Predictor v. Multivariate Analysis of Image-Based Posture Processing and Behaviour and Performance Sensing. *Animal* **2016**, *10*, 1525–1532, doi:10.1017/S1751731115001457.

37. Riaboff, L.; Relun, A.; Petiot, C.-E.; Feuilloy, M.; Couvreur, S.; Madouasse, A. Identification of Discriminating Behavioural and Movement Variables in Lameness Scores of Dairy Cows at Pasture from Accelerometer and GPS Sensors Using a Partial Least Squares Discriminant Analysis. *Prev Vet Med* **2021**, *193*, 105383, doi:10.1016/j.prevetmed.2021.105383.

38. Lemmens, L.; Schodl, K.; Fuerst-Waltl, B.; Schwarzenbacher, H.; Egger-Danner, C.; Linke, K.; Suntinger, M.; Phelan, M.; Mayerhofer, M.; Steininger, F.; et al. The Combined Use of Automated Milking System and Sensor Data to Improve Detection of Mild Lameness in Dairy Cattle. *Animals* **2023**, *13*, doi:10.3390/ani13071180.

39. Zhang, K.; Han, S.; Wu, J.; Cheng, G.; Wang, Y.; Wu, S.; Liu, J. Early Lameness Detection in Dairy Cattle Based on Wearable Gait Analysis Using Semi-Supervised LSTM-Autoencoder. *Comput Electron Agric* **2023**, *213*, 108252, doi:10.1016/j.compag.2023.108252.

40. Beer, G.; Alsaaod, M.; Starke, A.; Schuepbach-Regula, G.; Müller, H.; Kohler, P.; Steiner, A. Use of Extended Characteristics of Locomotion and Feeding Behavior for Automated Identification of Lame Dairy Cows. *PLoS One* **2016**, *11*, doi:10.1371/journal.pone.0155796.

41. Post, C.; Rietz, C.; Büscher, W.; Müller, U. Using Sensor Data to Detect Lameness and Mastitis Treatment Events in Dairy Cows: A Comparison of Classification Models. *Sensors (Switzerland)* **2020**, *20*, 1–18, doi:10.3390/s20143863.

42. Kang, X.; Zhang, X.D.; Liu, G. A Review: Development of Computer Vision-based Lameness Detection for Dairy Cows and Discussion of the Practical Applications. *Sensors (Switzerland)* **2021**, *21*, 1–24, doi:10.3390/s21030753.

43. O'Leary, N.W.; Byrne, D.T.; O'Connor, A.H.; Shalloo, L. Invited Review: Cattle Lameness Detection with Accelerometers. *J Dairy Sci* **2020**, *103*, 3895–3911, doi:10.3168/jds.2019-17123.

44. Busin, V.; Viora, L.; King, G.; Tomlinson, M.; Lekernec, J.; Jonsson, N.; Fioranelli, F. Evaluation of Lameness Detection Using Radar Sensing in Ruminants. *Veterinary Record* **2019**, *185*, 572, doi:10.1136/vr.105407.

45. Shrestha, A.; Loukas, C.; Le Kernec, J.; Fioranelli, F.; Busin, V.; Jonsson, N.; King, G.; Tomlinson, M.; Viora, L.; Voute, L. Animal Lameness Detection with Radar Sensing. *IEEE Geoscience and Remote Sensing Letters* **2018**, *15*, 1189–1193, doi:10.1109/LGRS.2018.2832650.

46. Perttu, R.K.; Peiter, M.; Bresolin, T.; Dórea, J.R.R.; Endres, M.I. Predictive Models for Disease Detection in Group-Housed Preweaning Dairy Calves Using Data Collected from Automated Milk Feeders. *J Dairy Sci* **2024**, *107*, 331–341, doi:10.3168/jds.2022-23037.

47. Silva, F.G.; Conceição, C.; Pereira, A.M.F.; Cerqueira, J.L.; Silva, S.R. Literature Review on Technological Applications to Monitor and Evaluate Calves' Health and Welfare. *Animals* **2023**, *13*, doi:10.3390/ani13071148.

48. Džermeikaitė, K.; Bačėninaitė, D.; Antanaitis, R. Innovations in Cattle Farming: Application of Innovative Technologies and Sensors in the Diagnosis of Diseases. *Animals* **2023**, *13*, doi:10.3390/ani13050780.

49. Vidal, G.; Sharpnack, J.; Pinedo, P.; Tsai, I.C.; Lee, A.R.; Martínez-López, B. Impact of Sensor Data Pre-Processing Strategies and Selection of Machine Learning Algorithm on the Prediction of Metritis Events in Dairy Cattle. *Prev Vet Med* **2023**, *215*, doi:10.1016/j.prevetmed.2023.105903.

50. Denholm, S.J.; Brand, W.; Mitchell, A.P.; Wells, A.T.; Krzyzelewski, T.; Smith, S.L.; Wall, E.; Coffey, M.P. Predicting Bovine Tuberculosis Status of Dairy Cows from Mid-Infrared Spectral Data of Milk Using Deep Learning. *J Dairy Sci* **2020**, *103*, 9355–9367, doi:10.3168/jds.2020-18328.

51. Cantor, M.C.; Casella, E.; Silvestri, S.; Renaud, D.L.; Costa, J.H.C. Using Machine Learning and Behavioral Patterns Observed by Automated Feeders and Accelerometers for the Early Indication of Clinical Bovine Respiratory Disease Status in Preweaned Dairy Calves. *Frontiers in Animal Science* **2022**, *3*, doi:10.3389/fanim.2022.852359.

52. Bowen, J.M.; Haskell, M.J.; Miller, G.A.; Mason, C.S.; Bell, D.J.; Duthie, C.A. Early Prediction of Respiratory Disease in Preweaning Dairy Calves Using Feeding and Activity Behaviors. *J Dairy Sci* **2021**, *104*, 12009–12018, doi:10.3168/jds.2021-20373.

53. Duthie, C.A.; Bowen, J.M.; Bell, D.J.; Miller, G.A.; Mason, C.; Haskell, M.J. Feeding Behaviour and Activity as Early Indicators of Disease in Pre-Weaned Dairy Calves. *Animal* **2021**, *15*, doi:10.1016/j.animal.2020.100150.

54. De Vries, A.; Bliznyuk, N.; Pinedo, P. Invited Review: Examples and Opportunities for Artificial Intelligence (AI) in Dairy Farms*. *Applied Animal Science* **2023**, *39*, 14–22, doi:10.15232/aas.2022-02345.

55. Ghaffari, M.H.; Monneret, A.; Hammon, H.M.; Post, C.; Müller, U.; Frieten, D.; Gerbert, C.; Dusel, G.; Koch, C. Deep Convolutional Neural Networks for the Detection of Diarrhea and Respiratory Disease in Preweaning Dairy Calves Using Data from Automated Milk Feeders. *J Dairy Sci* **2022**, *105*, 9882–9895, doi:10.3168/jds.2021-21547.

56. Shahinfar, S.; Khansefid, M.; Haile-Mariam, M.; Pryce, J.E. Machine Learning Approaches for the Prediction of Lameness in Dairy Cows. *Animal* **2021**, *15*, doi:10.1016/j.animal.2021.100391.

57. Williams, T.; Wilson, C.; Wynn, P.; Costa, D. Opportunities for Precision Livestock Management in the Face of Climate Change: A Focus on Extensive Systems. *Animal Frontiers* **2021**, *11*, 63–68, doi:10.1093/af/vfab065.

58. Chang, A.Z.; Swain, D.L.; Trotter, M.G. A Multi-Sensor Approach to Calving Detection. *Information Processing in Agriculture* **2022**, doi:10.1016/j.inpa.2022.07.002.

59. Williams, T.M.; Costa, D.F.A.; Wilson, C.S.; Chang, A.; Manning, J.; Swain, D.; Trotter, M.G. Sensor-Based Detection of Parturition in Beef Cattle Grazing in an Extensive Landscape: A Case Study Using a Commercial GNSS Collar. *Anim Prod Sci* **2022**, *62*, 993–999, doi:10.1071/AN21528.

60. García García, M.J.; Maroto Molina, F.; Pérez Marín, C.C.; Pérez Marín, D.C. Potential for Automatic Detection of Calving in Beef Cows Grazing on Rangelands from Global Navigate Satellite System Collar Data. *Animal* **2023**, *17*, doi:10.1016/j.animal.2023.100901.

61. Chang, A.Z.; Swain, D.L.; Trotter, M.G. Towards Sensor-Based Calving Detection in the Rangelands: A Systematic Review of Credible Behavioral and Physiological Indicators. *Transl Anim Sci* **2020**, *4*, 1–18, doi:10.1093/tas/txaa155.

62. Wolfger, B.; Schwartzkopf-Genswein, K.S.; Barkema, H.W.; Pajor, E.A.; Levy, M.; Orsel, K. Feeding Behavior as an Early Predictor of Bovine Respiratory Disease in North American Feedlot Systems1. *J Anim Sci* **2015**, *93*, 377–385, doi:10.2527/jas.2013-8030.

63. Bailey, D.W.; Trotter, M.G.; Knight, C.W.; Thomas, M.G. Use of GPS Tracking Collars and Accelerometers for Rangeland Livestock Production Research. *Transl Anim Sci* **2018**, *2*, 81–88, doi:10.1093/tas/txx006.

64. Qiao, Y.; Kong, H.; Clark, C.; Lomax, S.; Su, D.; Eiffert, S.; Sukkarieh, S. Intelligent Perception-Based Cattle Lameness Detection and Behaviour Recognition: A Review. *Animals* **2021**, *11*, doi:10.3390/ani11113033.

65. Greenwood, P.L.; Kardailsky, I.; Badgery, W.B.; Bishop-Hurley, G.J. Smart Farming for Extensive Grazing Ruminant Production Systems. *J Anim Sci* **2020**, *98*, 139–140, doi:10.1093/jas/skaa278.257.

66. Tzanidakis, C.; Tzamaloukas, O.; Simitzis, P.; Panagakis, P. Precision Livestock Farming Applications (PLF) for Grazing Animals. *Agriculture (Switzerland)* **2023**, *13*, doi:10.3390/agriculture13020288.

67. Fogarty, E.S.; Swain, D.L.; Cronin, G.; Trotter, M. Autonomous On-Animal Sensors in Sheep Research: A Systematic Review. *Comput Electron Agric* **2018**, *150*, 245–256, doi:10.1016/j.compag.2018.04.017.

68. Silva, S.R.; Sacarrão-Birrento, L.; Almeida, M.; Ribeiro, D.M.; Guedes, C.; Montaña, J.R.G.; Pereira, A.F.; Zaralis, K.; Geraldo, A.; Tzamaloukas, O.; et al. Extensive Sheep and Goat Production: The Role of Novel Technologies towards Sustainability and Animal Welfare. *Animals* **2022**, *12*, doi:10.3390/ani12070885.

69. Odintsov Vaintrub, M.; Levit, H.; Chincarini, M.; Fusaro, I.; Giammarco, M.; Vignola, G. Review: Precision Livestock Farming, Automats and New Technologies: Possible Applications in Extensive Dairy Sheep Farming. *Animal* 2021, *15*.

70. Morgan-Davies, C.; Tesnière, G.; Gautier, J.M.; Jørgensen, G.H.M.; González-García, E.; Patsios, S.I.; Sossidou, E.N.; Keady, T.W.J.; McClearn, B.; Kenyon, F.; et al. Review: Exploring the Use of Precision Livestock Farming for Small Ruminant Welfare Management. *Animal* **2024**, *18*, doi:10.1016/j.animal.2024.101233.

71. Fogarty, E.S.; Swain, D.L.; Cronin, G.M.; Moraes, L.E.; Bailey, D.W.; Trotter, M.G. Potential for Autonomous Detection of Lambing Using Global Navigation Satellite System Technology. *Anim Prod Sci* **2020**, *60*, 1217, doi:10.1071/AN18654.

72. Fogarty, E.S.; Swain, D.L.; Cronin, G.M.; Moraes, L.E.; Bailey, D.W.; Trotter, M. Developing a Simulated Online Model That Integrates GNSS, Accelerometer and Weather Data to Detect Parturition Events in Grazing Sheep: A Machine Learning Approach. *Animals* **2021**, *11*, 303, doi:10.3390/ani11020303.

73. Dobos, R.C.; Dickson, S.; Bailey, D.W.; Trotter, M.G. The Use of GNSS Technology to Identify Lambing Behaviour in Pregnant Grazing Merino Ewes. *Anim Prod Sci* **2014**, *54*, 1722–1727, doi:10.1071/AN14297.

74. Sohi, R.; Almasi, F.; Nguyen, H.; Carroll, A.; Trompf, J.; Weerasinghe, M.; Bervan, A.; Godoy, B.I.; Ahmed, A.; Stear, M.J.; et al. Determination of Ewe Behaviour around Lambing Time and Prediction of Parturition 7days Prior to Lambing by Tri-Axial Accelerometer Sensors in an Extensive Farming System. *Anim Prod Sci* **2022**, doi:10.1071/AN21460.

75. Fogarty, E.S.; Manning, J.K.; Trotter, M.G.; Schneider, D.A.; Thomson, P.C.; Bush, R.D.; Cronin, G.M. GNSS Technology and Its Application for Improved Reproductive Management in Extensive Sheep Systems. *Anim Prod Sci* **2015**, *55*, 1272–1280, doi:10.1071/AN14032.

76. Gurule, S.C.; Flores, V. V.; Forrest, K.K.; Gifford, C.A.; Wenzel, J.C.; Tobin, C.T.; Bailey, D.W.; Hernandez Gifford, J.A. A Case Study Using Accelerometers to Identify Illness in Ewes Following Unintentional Exposure to Mold-Contaminated Feed. *Animals* **2022**, *12*, 266, doi:10.3390/ani12030266.

77. Fogarty, E.; Cronin, G.; Trotter, M. Exploring the Potential for On-Animal Sensors to Detect Adverse Welfare Events: A Case Study of Detecting Ewe Behaviour Prior to Vaginal Prolapse. *Animal Welfare* **2022**, *31*, 355–359, doi:10.7120/09627286.31.3.009.

78. Fan, B.; Bryant, R.H.; Greer, A.W. Automatically Identifying Sickness Behavior in Grazing Lambs with an Acceleration Sensor. *Animals* **2023**, *13*, doi:10.3390/ani13132086.

79. Evans, C.A.; Trotter, M.G.; Manning, J.K. Sensor-Based Detection of Predator Influence on Livestock: A Case Study Exploring the Impacts of Wild Dogs (Canis Familiaris) on Rangeland Sheep. *Animals* **2022**, *12*, doi:10.3390/ani12030219.

80. Sohi, R.; Trompf, J.; Marriott, H.; Bervan, A.; Godoy, B.I.; Weerasinghe, M.; Desai, A.; Jois, M. Determination of Maternal Pedigree and Ewe–Lamb Spatial Relationships by Application of Bluetooth Technology in Extensive Farming Systems. *J Anim Sci* **2017**, *95*, 5145–5150, doi:10.2527/jas2017.1887.

81. Leroux, E.; Llach, I.; Besche, G.; Guyonneau, J.-D.; Montier, D.; Bouquet, P.-M.; Sanchez, I.; González-García, E. Evaluating a Walk-over-Weighing System for the Automatic Monitoring of Growth in Postweaned Mérinos d'Arles Ewe Lambs under Mediterranean Grazing Conditions. *Animal - Open Space* **2023**, *2*, 100032, doi:10.1016/j.anopes.2022.100032.

82. Gómez, Y.; Stygar, A.H.; Boumans, I.J.M.M.; Bokkers, E.A.M.; Pedersen, L.J.; Niemi, J.K.; Pastell, M.; Manteca, X.; Llonch, P. A Systematic Review on Validated Precision Livestock Farming Technologies for Pig Production and Its Potential to Assess Animal Welfare. *Front Vet Sci* **2021**, *8*, doi:10.3389/fvets.2021.660565.

83. van Klompenburg, T.; Kassahun, A. Data-Driven Decision Making in Pig Farming: A Review of the Literature. *Livest Sci* **2022**, *261*, doi:10.1016/j.livsci.2022.104961.

84. Vranken, E.; Berckmans, D. Precision Livestock Farming for Pigs. *Animal Frontiers* **2017**, *7*, 32–37, doi:10.2527/af.2017.0106.

85. Benjamin, M.; Yik, S. Precision Livestock Farming in Swinewelfare: A Review for Swine Practitioners. *Animals* **2019**, *9*, doi:10.3390/ani9040133.

86. Pomar, C.; Remus, A. Precision Pig Feeding: A Breakthrough toward Sustainability. *Animal Frontiers* **2019**, *9*, 52–59, doi:10.1093/af/vfz006.

87. de Bruijn, B.G.C.; de Mol, R.M.; Hogewerf, P.H.; van der Fels, J.B. A Correlated-Variables Model for Monitoring Individual Growing-Finishing Pig's Behavior by RFID Registrations. *Smart Agricultural Technology* **2023**, *4*, doi:10.1016/j.atech.2023.100189.

88. Sadeghi, E.; Kappers, C.; Chiumento, A.; Derks, M.; Havinga, P. Improving Piglets Health and Well-Being: A Review of Piglets Health Indicators and Related Sensing Technologies. *Smart Agricultural Technology* **2023**, *5*, doi:10.1016/j.atech.2023.100246.

89. Garrido, L.F.C.; Sato, S.T.M.; Costa, L.B.; Daros, R.R. Can We Reliably Detect Respiratory Diseases through Precision Farming? A Systematic Review. *Animals* **2023**, *13*, doi:10.3390/ani13071273.

90. Habineza, E.; Reza, M.N.; Chowdhury, M.; Kiraga, S.; Chung, S.-O.; Hong, S.J. Pig Diseases and Crush Monitoring Visual Symptoms Detection Using Engineering Approaches: A Review. *Precision Agriculture Science and Technology* **2021**, *3*, doi:10.12972/pastj.20210017.

91. Traulsen, I.; Scheel, C.; Auer, W.; Burfeind, O.; Krieter, J. Using Acceleration Data to Automatically Detect the Onset of Farrowing in Sows. *Sensors (Switzerland)* **2018**, *18*, doi:10.3390/s18010170.

92. Gan, H.; Li, S.; Ou, M.; Yang, X.; Huang, B.; Liu, K.; Xue, Y. Fast and Accurate Detection of Lactating Sow Nursing Behavior with CNN-Based Optical Flow and Features. *Comput Electron Agric* **2021**, *189*, doi:10.1016/j.compag.2021.106384.

93. Yang, A.; Huang, H.; Yang, X.; Li, S.; Chen, C.; Gan, H.; Xue, Y. Automated Video Analysis of Sow Nursing Behavior Based on Fully Convolutional Network and Oriented Optical Flow. *Comput Electron Agric* **2019**, *167*, doi:10.1016/j.compag.2019.105048.

94. Guo, Q.; Sun, Y.; Orsini, C.; Bolhuis, J.E.; de Vlieg, J.; Bijma, P.; de With, P.H.N. Enhanced Camera-Based Individual Pig Detection and Tracking for Smart Pig Farms. *Comput Electron Agric* **2023**, *211*, 108009, doi:10.1016/j.compag.2023.108009.

95. Zhou, H.; Chung, S.; Kakar, J.K.; Kim, S.C.; Kim, H. Pig Movement Estimation by Integrating Optical Flow with a Multi-Object Tracking Model. *Sensors* **2023**, *23*, doi:10.3390/s23239499.

96. Mittek, M.; Psota, E.T.; Pérez, L.C.; Schmidt, T.; Mote, B. Health Monitoring of Group-Housed Pigs Using Depth-Enabled Multi-Object Tracking. In Proceedings of the Proceedings of Int Conf Pattern Recognit, Workshop on Visual; 2016.

97. Zha, W.; Li, H.; Wu, G.; Zhang, L.; Pan, W.; Gu, L.; Jiao, J.; Zhang, Q. Research on the Recognition and Tracking of Group-Housed Pigs' Posture Based on Edge Computing. *Sensors (Basel)* **2023**, *23*, doi:10.3390/s23218952.

98. Dong, Y.; Bonde, A.; Codling, J.R.; Bannis, A.; Cao, J.; Macon, A.; Rohrer, G.; Miles, J.; Sharma, S.; Brown-Brandl, T.; et al. PigSense: Structural Vibration-Based Activity and Health Monitoring System for Pigs. *ACM Trans Sens Netw* **2023**, *20*, doi:10.1145/3604806.

99. Franchi, G.A.; Bus, J.D.; Boumans, I.J.M.M.; Bokkers, E.A.M.; Jensen, M.B.; Pedersen, L.J. Estimating Body Weight in Conventional Growing Pigs Using a Depth Camera. *Smart Agricultural Technology* **2023**, *3*, doi:10.1016/j.atech.2022.100117.

100. Nguyen, A.H.; Holt, J.P.; Knauer, M.T.; Abner, V.A.; Lobaton, E.J.; Young, S.N. Towards Rapid Weight Assessment of Finishing Pigs Using a Handheld, Mobile RGB-D Camera. *Biosyst Eng* **2023**, *226*, 155–168, doi:10.1016/j.biosystemseng.2023.01.005.

101. Chen, Z.; Lu, J.; Wang, H. A Review of Posture Detection Methods for Pigs Using Deep Learning. *Applied Sciences (Switzerland)* **2023**, *13*, doi:10.3390/app13126997.

102. D'Eath, R.B.; Foister, S.; Jack, M.; Bowers, N.; Zhu, Q.; Barclay, D.; Baxter, E.M. Changes in Tail Posture Detected by a 3D Machine Vision System Are Associated with Injury from Damaging Behaviours and Ill Health on Commercial Pig Farms. *PLoS One* **2021**, *16*, doi:10.1371/journal.pone.0258895.

103. Chen, C.; Zhu, W.; Steibel, J.; Siegford, J.; Wurtz, K.; Han, J.; Norton, T. Recognition of Aggressive Episodes of Pigs Based on Convolutional Neural Network and Long Short-Term Memory 2. **2019**.

104. Larsen, M.L.V.; Pedersen, L.J.; Edwards, S.; Albanie, S.; Dawkins, M.S. Movement Change Detected by Optical Flow Precedes, but Does Not Predict, Tail-Biting in Pigs. *Livest Sci* **2020**, *240*, doi:10.1016/j.livsci.2020.104136.

105. Li, Y.Z.; Johnston, L.J.; Dawkins, M.S. Utilization of Optical Flow Algorithms to Monitor Development of Tail Biting Outbreaks in Pigs. *Animals* **2020**, *10*, doi:10.3390/ani10020323.

106. Reza, M.N.; Ali, M.R.; Samsuzzaman; Kabir, M.S.N.; Karim, M.R.; Ahmed, S.; Kyoung, H.; Kim, G.; Chung, S.-O. Thermal Imaging and Computer Vision Technologies for the Enhancement of Pig Husbandry: A Review. *J Anim Sci Technol* **2024**, *66*, 31–56, doi:10.5187/jast.2024.e4.

107. Arulmozhi, E.; Bhujel, A.; Moon, B.E.; Kim, H.T. The Application of Cameras in Precision Pig Farming: An Overview for Swine-Keeping Professionals. *Animals* **2021**, *11*, doi:10.3390/ani11082343.

108. Hou, Y.; Li, Q.; Wang, Z.; Liu, T.; He, Y.; Li, H.; Ren, Z.; Guo, X.; Yang, G.; Liu, Y.; et al. Study on a Pig Vocalization Classification Method Based on Multi-Feature Fusion. *Sensors* **2024**, *24*, doi:10.3390/s24020313.

109. Hong, M.; Ahn, H.; Atif, O.; Lee, J.; Park, D.; Chung, Y. Field-Applicable Pig Anomaly Detection System Using Vocalization for Embedded Board Implementations. *Applied Sciences (Switzerland)* **2020**, *10*, 1–17, doi:10.3390/app10196991.

110. Lagua, E.B.; Mun, H.S.; Ampode, K.M.B.; Chem, V.; Kim, Y.H.; Yang, C.J. Artificial Intelligence for Automatic Monitoring of Respiratory Health Conditions in Smart Swine Farming. *Animals* **2023**, *13*, doi:10.3390/ani13111860.

111. Dawkins, M.S.; Rowe, E. Poultry Welfare Monitoring: Group-Level Technologies. In *Understanding the behaviour and improving the welfare of chickens*; 2020; pp. 177–196.

112. Li, N.; Ren, Z.; Li, D.; Zeng, L. Review: Automated Techniques for Monitoring the Behaviour and Welfare of Broilers and Laying Hens: Towards the Goal of Precision Livestock Farming. *Animal* **2020**, *14*, 617–625, doi:10.1017/S1751731119002155.

113. Mortensen, A.K.; Lisouski, P.; Ahrendt, P. Weight Prediction of Broiler Chickens Using 3D Computer Vision. *Comput Electron Agric* **2016**, *123*, 319–326, doi:10.1016/j.compag.2016.03.011.

114. Neethirajan, S. ChickTrack – A Quantitative Tracking Tool for Measuring Chicken Activity. *Measurement (Lond)* **2022**, *191*, doi:10.1016/j.measurement.2022.110819.

115. Pereira, D.F.; Nääs, I. de A.; Lima, N.D. da S. Movement Analysis to Associate Broiler Walking Ability with Gait Scoring. *AgriEngineering* **2021**, *3*, 394–402, doi:10.3390/agriengineering3020026.

116. Aydin, A. Development of an Early Detection System for Lameness of Broilers Using Computer Vision. *Comput Electron Agric* **2017**, *136*, 140–146, doi:10.1016/j.compag.2017.02.019.

117. Zhuang, X.; Zhang, T. Detection of Sick Broilers by Digital Image Processing and Deep Learning. *Biosyst Eng* **2019**, *179*, 106–116, doi:10.1016/j.biosystemseng.2019.01.003.

118. Du, X.; Teng, G. An Automatic Detection Method for Abnormal Laying Hen Activities Using a 3D Depth Camera. *Engenharia Agricola* **2021**, *41*, 263–270, doi:10.1590/1809-4430-ENG.AGRIC.V41N3P263-270/2021.

119. Dawkins, M.S.; Wang, L.; Ellwood, S.A.; Roberts, S.J.; Gebhardt-Henrich, S.G. Optical Flow, Behaviour and Broiler Chicken Welfare in the UK and Switzerland. *Appl Anim Behav Sci* **2021**, *234*, doi:10.1016/j.applanim.2020.105180.

120. Herborn, K.A.; McElligott, A.G.; Mitchell, M.A.; Sandilands, V.; Bradshaw, B.; Asher, L. Spectral Entropy of Early-Life Distress Calls as an Iceberg Indicator of Chicken Welfare. *J R Soc Interface* **2020**, *17*, 20200086, doi:10.1098/rsif.2020.0086.

121. Liu, L.; Li, B.; Zhao, R.; Yao, W.; Shen, M.; Yang, J. A Novel Method for Broiler Abnormal Sound Detection Using WMFCC and HMM. *J Sens* **2020**, *2020*, doi:10.1155/2020/2985478.

122. Huang, J.; Zhang, T.; Cuan, K.; Fang, C. An Intelligent Method for Detecting Poultry Eating Behaviour Based on Vocalization Signals. *Comput Electron Agric* **2021**, *180*, doi:10.1016/j.compag.2020.105884.

123. Ellen, E.D.; van der Sluis, M.; Siegford, J.; Guzhva, O.; Toscano, M.J.; Bennewitz, J.; Van Der Zande, L.E.; Van Der Eijk, J.A.J.; De Haas, E.N.; Norton, T.; et al. Review of Sensor Technologies in Animal Breeding: Phenotyping Behaviors of Laying Hens to Select against Feather Pecking. *Animals* **2019**, *9*, doi:10.3390/ani9030108.

124. Borgonovo, F.; Ferrante, V.; Grilli, G.; Guarino, M. An Innovative Approach for Analysing and Evaluating Enteric Diseases in Poultry Farm. *Acta IMEKO* **2024**, *13*, 1–5, doi:10.21014/acta_imeko.v13i1.1627.

125. Li, L.; Zhao, Y.; Oliveira, J.; Verhoijsen, W.; Liu, K.; Xin, H. A UHF RFID System for Studying Individual Feeding and Nesting Behaviors of Group-Housed Laying Hens. *Trans ASABE* **2017**, *60*, 1337–1347, doi:10.13031/trans.12202.

126. Yang, X.; Zhao, Y.; Street, G.M.; Huang, Y.; Filip To, S.D.; Purswell, J.L. Classification of Broiler Behaviours Using Triaxial Accelerometer and Machine Learning. *Animal* **2021**, *15*, doi:10.1016/j.animal.2021.100269.

127. Ahmed, G.; Malick, R.A.S.; Akhunzada, A.; Zahid, S.; Sagri, M.R.; Gani, A. An Approach towards Iot-Based Predictive Service for Early Detection of Diseases in Poultry Chickens. *Sustainability (Switzerland)* **2021**, *13*, doi:10.3390/su132313396.

128. Stachowicz, J.; Umstätter, C. Do We Automatically Detect Health- or General Welfare-Related Issues? A Framework. *Proceedings of the Royal Society B: Biological Sciences* **2021**, *288*, doi:10.1098/rspb.2021.0190.

129. Fuchs, P.; Adrion, F.; Shafiullah, A.Z.M.; Bruckmaier, R.M.; Umstätter, C. Detecting Ultra- and Circadian Activity Rhythms of Dairy Cows in Automatic Milking Systems Using the Degree of Functional Coupling—A Pilot Study. *Frontiers in Animal Science* **2022**, *3*, doi:10.3389/fanim.2022.839906.

130. Schneider, M.; Umstätter, C.; Nasser, H.-R.; Gallmann, E.; Barth, K. Effect of the Daily Duration of Calf Contact on the Dam's Ultradian and Circadian Activity Rhythms. *JDS Communications* **2024**, *5*, 457–461.

131. Sturm, V.; Efrosinin, D.; Öhlschuster, M.; Gusterer, E.; Drillich, M.; Iwersen, M. Combination of Sensor Data and Health Monitoring for Early Detection of Subclinical Ketosis in Dairy Cows. *Sensors (Switzerland)* **2020**, *20*, doi:10.3390/s20051484.

132. da Silva Santos, A.; de Medeiros, V.W.C.; Gonçalves, G.E. Monitoring and Classification of Cattle Behavior: A Survey. *Smart Agricultural Technology* **2023**, *3*, doi:10.1016/j.atech.2022.100091.

133. Riaboff, L.; Shalloo, L.; Smeaton, A.F.; Couvreur, S.; Madouasse, A.; Keane, M.T. Predicting Livestock Behaviour Using Accelerometers: A Systematic Review of Processing Techniques for Ruminant Behaviour Prediction from Raw Accelerometer Data. *Comput Electron Agric* **2022**, *192*, doi:10.1016/j.compag.2021.106610.

134. Diggle, P.J. *Time Series; A Biostatistical Introduction*; 1st ed.; Oxford Science Publications, 1990; ISBN 0-19-852226-6.

135. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; 2nd ed.; Springer, 2009;

136. Diggle, P.; Giorgi, E. *Time Series: A Biostatistical Introduction*; 2nd ed.; Oxford University Press , 2025;

137. Hyndman, R.; Koehler, A.; Ord, K.; Snyder, R. *Forecasting with Exponential Smoothing*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; ISBN 978-3-540-71916-8.

138. Basseville, M.; Nikiforov, I. V *Detection of Abrupt Changes: Theory and Application 1*; Prentice Hall Englewood Cliffs, 1993; ISBN 978-0-13-126780-0.

139. Montgomery, D.C. *Statistical Quality Control: A Modern Introduction*; 7th ed.; John Wiley & Sons, 2012;

140. Gupta, M.; Gao, J.; Aggarwal, C.C.; Han, J. Outlier Detection for Temporal Data: A Survey. *IEEE Trans Knowl Data Eng* **2014**, *26*, 2250–2267, doi:10.1109/TKDE.2013.184.

141. Kolambe, M.; Arora, S. *Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques*; 2024; Vol. 20;.

142. Brown, H.; Prescott, R. *Applied Mixed Models in Medicine*; 3rd ed.; Wiley, 1999; ISBN 0-471-96554-5.

143. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; 2nd ed.; Chapman & Hall, 1989; ISBN 0-412-31760-5.

144. Skrondal, A.; Rabe-Hesketh, S. Latent Variable Modelling: A Survey. *Scandinavian Journal of Statistics* **2007**, *34*, 712–745, doi:10.1111/j.1467-9469.2007.00573.x.

145. Mor, B.; Garhwal, S.; Kumar, A. A Systematic Review of Hidden Markov Models and Their Applications. *Archives of Computational Methods in Engineering* **2021**, *28*, 1429–1448, doi:10.1007/s11831-020-09422-4.

146. Everitt, B.S.; Landau, S.; Leese, M.; Stahl, D. *Cluster Analysis*; Wiley, 2011; ISBN 978-0-470-74991-3.

147. Gilks, W.R.; Richardson, S.; Spiegelhalter, D.J. *Markov Chain Monte Carlo in Practice*; 1st ed.; Chapman & Hall/CRC, 1996; ISBN 0-412-05551-1.

148. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; 2nd ed.; Chapman & Hall/CRC, 2004; ISBN 1-58488-388-X.

149. Doucet, A.; Freitas, N.; Gordon, N. An Introduction to Sequential Monte Carlo Methods. In *Sequential Monte Carlo Methods in Practice*; Springer New York: New York, NY, 2001; pp. 3–14.

150. Shine, P.; Murphy, M.D. Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study. *Sensors* **2022**, *22*, doi:10.3390/s22010052.

151. Greener, J.G.; Kandathil, S.M.; Moffat, L.; Jones, D.T. A Guide to Machine Learning for Biologists. *Nat Rev Mol Cell Biol* **2022**, *23*, 40–55, doi:https://doi.org/10.1038/s41580-021-00407-0.

152. Choi, R.Y.; Coyner, A.S.; Kalpathy-Cramer, J.; Chiang, M.F.; Peter Campbell, J. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol* **2020**, *9*, doi:10.1167/tvst.9.2.14.

153. Simchoni, G.; Rosset, S. Integrating Random Effects in Deep Neural Networks. *Journal of Machine Learning Rsearch* **2022**, *24*, doi:https://doi.org/10.48550/arXiv.2206.03314.

154. Domingos, P. A Few Useful Things to Know about Machine Learning. *Commun ACM* **2012**, *55*, 78–87, doi:10.1145/2347736.2347755.

155. Haveman, M.E.; van Rossum, M.C.; Vaseur, R.M.E.; van der Riet, C.; Schuurmann, R.C.L.; Hermens, H.J.; de Vries, J.P.P.M.; Tabak, M. Continuous Monitoring of Vital Signs With Wearable Sensors During Daily Life Activities: Validation Study. *JMIR Form Res* **2022**, *6*, doi:10.2196/30863.

156. Armitage, P.; Berry, G.; Mathews, J.N.S. *Statistical Methods in Medical Research*; 4th ed.; Blackwell Science, 2002;

157. Borchers, M.R.; Chang, Y.M.; Tsai, I.C.; Wadsworth, B.A.; Bewley, J.M. A Validation of Technologies Monitoring Dairy Cow Feeding, Ruminating, and Lying Behaviors. *J Dairy Sci* **2016**, *99*, 7458–7466, doi:10.3168/jds.2015-10843.

158. Lasser, J.; Matzhold, C.; Egger-Danner, C.; Fuerst-Waltl, B.; Steininger, F.; Wittek, T.; Klimek, P. Integrating Diverse Data Sources to Predict Disease Risk in Dairy Cattle - A Machine Learning Approach. *J. Anim. Sci.* **2021**, *99*, doi:10.1101/2021.03.25.436798.

159. Fleiss, J.; Levin, B.; Paik, M.C. *Statistical Methods for Rates and Proportions*; 3rd ed.; John Wiley & Sons, 2003; ISBN 9780471445425.

160. Everitt, B.S. *The Analysis of Contingency Tables*; 2nd ed.; Chapman & Hall, 1992; ISBN 0-412-39850-8.

161. Collet, D. *Modelling Binary Data*; Chapman & Hall/CRC, 1999; ISBN 0-412-38800-6.

162. Tassi, R.; Schiavo, M.; Filipe, J.; Todd, H.; Ewing, D.; Ballingall, K.T. Intramammary Immunisation Provides Short Term Protection Against Mannheimia Haemolytica Mastitis in Sheep. *Front Vet Sci* **2021**, *8*, doi:10.3389/fvets.2021.659803.

163. Williams-Macdonald, S.E.; Mitchell, M.; Frew, D.; Palarea-Albaladejo, J.; Ewing, D.; Golde, W.T.; Longbottom, D.; Nisbet, A.J.; Livingstone, M.; Hamilton, C.M.; et al. Efficacy of Phase I and Phase II

Coxiella Burnetii Bacterin Vaccines in a Pregnant Ewe Challenge Model. *Vaccines (Basel)* **2023**, *11*, doi:10.3390/vaccines11030511.

164. McNee, A.; Smith, T.R.F.; Holzer, B.; Clark, B.; Bessell, E.; Guibinga, G.; Brown, H.; Schultheis, K.; Fisher, P.; Ramos, S.; et al. Establishment of a Pig Influenza Challenge Model for Evaluation of Monoclonal Antibody Delivery Platforms. *The Journal of Immunology* **2020**, *205*, 648–660, doi:10.4049/jimmunol.2000429.

165. O'Donoghue, S.; Earley, B.; Johnston, D.; McCabe, M.S.; Kim, J.W.; Taylor, J.F.; Duffy, C.; Lemon, K.; McMenamy, M.; Cosby, S.L.; et al. Whole Blood Transcriptome Analysis in Dairy Calves Experimentally Challenged with Bovine Herpesvirus 1 (BoHV-1) and Comparison to a Bovine Respiratory Syncytial Virus (BRSV) Challenge. *Front Genet* **2023**, *14*, doi:10.3389/fgene.2023.1092877.

166. Kang, H.; Brocklehurst, S.; Haskell, M.; Jarvis, S.; Sandilands, V. Do Activity Sensors Identify Physiological, Clinical and Behavioural Changes in Laying Hens Exposed to a Vaccine Challenge? *Animals* **2025**, *15*, doi:10.3390/ani15020205.

167. Ahmed, S.T.; Mun, H.S.; Islam, M.M.; Yoe, H.; Yang, C.J. Monitoring Activity for Recognition of Illness in Experimentally Infected Weaned Piglets Using Received Signal Strength Indication ZigBee-Based Wireless Acceleration Sensor. *Asian-Australas J Anim Sci* **2016**, *29*, 149–156, doi:10.5713/ajas.15.0221.

168. Kayser, W.C.; Carstens, G.E.; Jackson, K.S.; Pinchak, W.E.; Banerjee, A.; Fu, Y. Evaluation of Statistical Process Control Procedures to Monitor Feeding Behavior Patterns and Detect Onset of Bovine Respiratory Disease in Growing Bulls. *J Anim Sci* **2019**, *97*, 1158–1170, doi:10.1093/jas/sky486.

169. March, M.D.; Hargreaves, P.R.; Sykes, A.J.; Rees, R.M. Effect of Nutritional Variation and LCA Methodology on the Carbon Footprint of Milk Production From Holstein Friesian Dairy Cows. *Front Sustain Food Syst* **2021**, *5*, doi:10.3389/fsufs.2021.588158.

170. Pollott, G.E.; Coffey, M.P. The Effect of Genetic Merit and Production System on Dairy Cow Fertility, Measured Using Progesterone Profiles and on-Farm Recording. *J Dairy Sci* **2008**, *91*, 3649–3660, doi:10.3168/jds.2007-0913.

171. SRUC Dairy Research and Innovation Centre Available online: https://www.sruc.ac.uk/research/research-facilities/dairy-research-facility/ (accessed on 2 April 2024).

172. Stepmetrix from Boumatic Available online: https://kruegersboumatic.com/automation/stepmetrix/ (accessed on 10 March 2023).

173. Hokofarm Group Available online: https://hokofarmgroup.com/ (accessed on 7 March 2023).

174. Peacock Technology (Was IceRobotics) Available online: https://www.peacocktechnology.com/ids-i-qube (accessed on 10 March 2023).

175. Stygar, A.H.; Frondelius, L.; Berteselli, G. V.; Gómez, Y.; Canali, E.; Niemi, J.K.; Llonch, P.; Pastell, M. Measuring Dairy Cow Welfare with Real-Time Sensor-Based Data and Farm Records: A Concept Study. *Animal* **2023**, *17*, doi:10.1016/j.animal.2023.101023.

176. EU TechCare Project Available online: https://techcare-project.eu/ (accessed on 15 March 2023).

177. McLaren, A.; Waterhouse, A.; Kenyon, F.; MacDougall, H.; Beechener, E.S.; Walker, A.; Reeves, M.; Lambe, N.R.; Holland, J.P.; Thomson, A.T.; et al. TechCare UK Pilots - Integrated Sheep System Studies Using Technologies for Welfare Monitoring. In Proceedings of the Book of Abstracts of the 74th Annual Meeting of the European Federation of Animal Science; Wageningen Academic Publishers, 2023.

178. Moredun Research Institute Available online: https://moredun.org.uk/ (accessed on 2 April 2024).

179. AX3 Accelerometer Available online: https://axivity.com/userguides/ax3/ (accessed on 15 March 2023).

180. I-GotU GT-120B Mobile Action Technology, Inc., Taiwan Available online: http://www.mobileaction.com/web/en/portfolio-view/gt-120b/ (accessed on 29 January 2024).

181. RealTimeID Animal Trackers Available online: https://realtimeid.no/collections/all (accessed on 29 January 2024).

182. Lora, I.; Gottardo, F.; Contiero, B.; Zidi, A.; Magrin, L.; Cassandro, M.; Cozzi, G. A Survey on Sensor Systems Used in Italian Dairy Farms and Comparison between Performances of Similar Herds Equipped or Not Equipped with Sensors. *J Dairy Sci* **2020**, *103*, 10264–10272, doi:10.3168/jds.2019-17973.

183. Morgan-Davies, C.; Lambe, N.; Wishart, H.; Waterhouse, T.; Kenyon, F.; McBean, D.; McCracken, D. Impacts of Using a Precision Livestock System Targeted Approach in Mountain Sheep Flocks. *Livest Sci* **2018**, *208*, 67–76, doi:10.1016/j.livsci.2017.12.002.

184. SRUC Hill & Mountain Research Available online: https://www.sruc.ac.uk/research/research-facilities/hill-mountain-research/ (accessed on 2 April 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.