Update on development of criteria, protocols and guidelines for sensor devices

H. van den Bijgaart and S. Orlandini

ICAR Sensor Devices Task Force

Summary

The ICAR Board has initiated a Sensor Devices Task Force (SD-TF) with the task to provide guidelines and methodology to help classify and qualify on-farm sensors and sensor data in relation to milk production. This to create better safeguards on their suitability for herd management and breeding decisions. In developing guidelines and establishing potential performance criteria for sensor devices it is essential to consider sensor measurements as part of a whole process towards decision making. It is the aimed at quality in decision making that should determine the required quality level in each of the preceding steps. Developing sensor validation criteria for possibly occurring situations requires to address questions like:

- What is the trait?
- What is the purpose of the sensor measurement?
- What type of data does the sensor produce?
- What is the anchor?
- Single measurement or multiple measurements?

As part of this paper the approach and the challenges of the SD-TF are illustrated with an example on two different types of udder health monitoring devices.

Keywords: Sensor devices, performance criteria, precision, accuracy, guidelines

Introduction

ICAR has the role of supporting farmers and breeding organisations in their effort to collect valid data for daily management, and to deliver proven and validated information for genetic evaluations. In a rapidly increasing extent sensor technologies are currently being implemented in dairy farms to electronically monitor livestock, their environment, and to collect real-time data to make more informed decisions on aspects such as reproduction activities, herd and individual animal health status, feeding and nutrition, milking and milk production data and others. However, as yet data from such sensors is mostly unqualified with respect to accuracy levels, calibration and/or validation status. Consensus statements on minimum required performance criteria as well as standardized protocols for validation, maintenance and calibration are largely lacking.

Corresponding Authors: bijgaart@qlip.nl, silvia@icar.org

As a consequence, unqualified data may be transferred from on-farm computer systems into central databases and/or data processing centres, either intentionally or unintentionally at the present time.

In 2013 ICAR convened an 'Accuracy Task Force (A-TF) with the goal of developing a sound philosophical basis for ICAR to use in establishing required accuracy for the collection of animal recording data that is incorporated into information services. The A-TF presented their final report to the ICAR Board and General Assembly in 2016 (ICAR, 2016).

In order to move from the theoretical basis to a more practical phase, the ICAR Board initiated a new Sensor Devices Task Force (SD-TF) with the task to:

- Provide guidelines/methodology to help classify and qualify non-ICAR certified sensors and sensor data that measure:
  - milk volume;
  - milk speed;
  - milk constituents;
  - mastitis and
  - other traits

so as to determine their suitability for the respective management and breeding decisions;

- Disseminate new guidelines, so members and practitioners are aware of and educated in the approach for determining suitability of data from non-ICAR certified sensors.

This paper is intended to give insight in some of the encountered challenges by the SD-TF and the intended approach in dealing with these. This is illustrated with some examples on udder health monitoring.

The process

The process from sampling and sensor measurements to farm management decisions is schematically depicted in figure 1.

Robust data collection starts with proper identification of the animal, which is either right or wrong.

Sampling is to a large extent decisive for the ultimate representativeness of the collected data. During milking a single point (spot) sample of the excreted milk may signal an abnormality with the milk or the animal but will not allow to determine values for compositional or somatic cell count values representative for the whole milking. For that purpose a representative sample of the whole milking has to be collected or

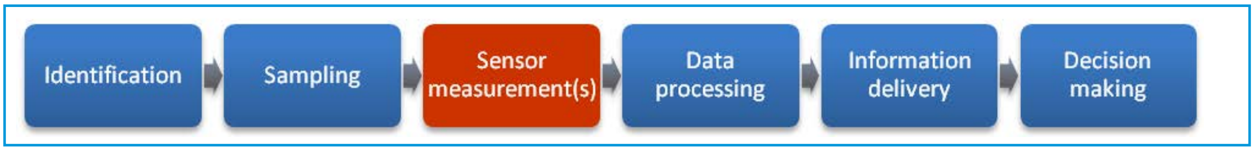

Figure 1. The process with the use of sensor data.

repeated measurements during milking have to be done with the use of an adequate calculation algorithm. Whatever the type of sampling, a standardized and well controlled sampling procedure is a prerequisite to arrive at valid and useful data.

After sampling and possible sample pre-treatment, the sample is measured with a sensor device. It is noted that in the frame of this paper the term sensor device encompasses the whole range from simple temperature and conductivity sensors to more complex in-line or at-line analytical devices. With sensor measurements the question arises what the target of the measurement is. How well is the target defined? Is there a 'gold' standard, a reference method or another well-defined entity to anchor against? How secure is the anchor? Furthermore, sensor measurements require a certain degree of stability and robustness.

Data processing is about converting data into information that is to be used in decision making. As a part of this step, currently collected data may be combined with other animal data available from other sources or from other moments in time. Data processing also involves the handling of missing values or outliers and may include algorithms for data smoothing. Biomodels can be used to create proxies useful in herd management or in genetic selection. Measurement errors may to some extent be compensated for during this step, at the same time additional estimation errors may be introduced due to model inaccuracies.

Effective decision making is facilitated by well thought formats for information delivery. Warning lists help to pinpoint animals needing specific attention. Clear graphs can provide easy to grasp insights in trend-based courses.

The general approach for setting performance criteria

In the depicted process it is the performance level with each of the steps that determines the quality of the information that is used for decision making. In defining criteria for the subsequent steps of the process, this statement must be inverted. What is the aimed at quality of the information in decision making and what does this mean for the required performance level in each of the steps? So setting criteria for sensor devices starts from the question what the targeted use of the data is. Is it individual animal management, signalling animal health and welfare issues, herd management and/or genetic evaluation?

Key performance parameters with qualitative and quantitative parameters

The current focus of the SD-TF is on sensor devices and the quality assurance with sensor data. The SD-TF has started to build a database of sensor devices on the market with identification of their key characteristics such as aimed at trait, purpose of the measurement, availability of an anchor ('gold' standard, reference method, other), measurement matrix, method principle and available performance information from literature or other sources.

For each of the identified traits one can distinguish between sensors delivering qualitative and quantitative data. When examining the performance or setting performance criteria, separate performance parameters apply. The identified key performance parameters for each type of sensor device are summarized in Table 1.

Other relevant performance parameters may relate to the measurement range (lower limit, upper limit, capability of detection) and carry-over (the effect on a sample on the measurement result with a successively measured sample).

Table 1. Overview of key performance parameters with sensor devices producing either qualitative or quantitative data.

Qualitative data	
Sensitivity	The ability to detect the analyte compared to the reference = true positive rate
Specificity	The ability not to detect the analyte when it is not detected by the reference = true negative rate
Quantitative data	
Repeatability r (expressed as standard deviation of repeatability s_r, r = 2.8 * s_r with > 20 measurements in replicate)	Standard deviation of values under a set of repeatability conditions of measurement, that is conditions where independent measurement results are obtained with the same device on identical test samples in the same place by the same operator within short intervals of time. In other words, the ability to produce the same result over and over under the same conditions.
Reproducibility R (expressed as standard deviation of reproducibility s_R, R = 2.8 * s_R with > 20 measurements in at least replicate)	Standard deviation of values under a set of reproducibility conditions of measurement, that is conditions where independent measurement results are obtained with another copy of the same device on identical sample in different places and/or with the same device at different times by different operators. In other words, the ability to produce the same result at different places and/or at different times, e.g. under different conditions.
Accuracy (expressed as standard deviation of accuracy $s_{y,x}$)	Closeness of agreement between a measured quantity value and a true quantity value of a measurand, which is a result from both random error (precision) and systematic error of the measured quantity value. It is only the latter part that is also known as trueness and expressed as bias.
Robustness	Vulnerability to interactions and environmental interferences (shocks/vibrations, humidity/water, cleaning chemicals, temperature, milk flow speed).

Examples for udder health sensors

Udder health is one of the focal traits with the application of sensor devices. Many devices based on different principles have found their application on farms, i.e. milk conductivity sensors, LDH measurement, ATP measurement, NAG-ase measurement, thermal imaging, automated CMT/WMT (viscosity measurement) and somatic cell counting devices. The purpose of these type of sensor measurements in general is to timely signal animals that need closer attention from the farmer and/or a veterinarian. Applying on-farm sensors mainly serves animal health and welfare and herd management purpose, but it is imaginable that so collected data are used for determining breeding values for udder health, provided the quality of the data is adequately assured.

Qualitative data

Udder health sensors delivering only qualitative data have a typical on-farm purpose, that is to signal animals with (potential) udder health problems. With that, the trait, the purpose and the type of sensor data are clear. A well-defined and traceable anchor for the measurement itself is lacking, the goal is to signal animals with an (upcoming) clinical or subclinical mastitis for further examination. In that, it is relevant to miss not too many problematic animals, even if that comes at the expense of signalling a few extra non-problematic animals. At the same time every false positive indication comes with unwanted hassle for the farmer. The SD-TF is therefore considering whether guidance should state to aim for at least 80% sensitivity and 99% specificity when scored against internationally accepted mastitis definitions as proposed by Hogeveen et al., (2010) for alerting on clinical mastitis. With multiple measurements, these criteria would apply for the combined sensor output during one milking.

Quantitative data

Let us now assume we are dealing with an installed in-line somatic cell counting device which measures the somatic cell count in a representative aliquot of the milking. The purpose is on the one hand to signal problematic animals but also to use the data for breeding value estimates on udder health. In fact the same purpose as with traditional milk recording where the quantitative somatic count values are determined in representative samples in a central laboratory. Therefore in this situation also equivalence in the quality of the data is sought. The reference method for somatic cell count is the direct microscopic cell count, although also a well calibrated laboratory somatic cell counter can serve as the anchor.

The two main performance parameters are precision (repeatability and reproducibility) and accuracy. Precision is determined by non-systematic error and is part of the accuracy. Precision can be reduced by multiplicative measurements and using the average results. The other part of accuracy is trueness, the systematic deviation due to the error with calibration or due to a systematic influence of the matrix. In other words, milk from different animals behaving differently with different methods. Due to the systematic character, lack of trueness cannot be compensated for by repeated measurements.

For this situation the ICAR Sub-Committee on Milk Analysis developed guidelines on on-farm milk analysis, which are now part of the ICAR Guidelines (ICAR, 2017). The aim of these guidelines is to promote consistency and correspondence between different measuring systems with regard to measurement uncertainty and through that enabling comparison within time and place:

- For milk producers to manage day-to-day milk production.
- For milk recording organisations to maintain sufficient accuracy in estimating genetic indicators.

The rationale behind the establishment of performance criteria for on-farm milk analysers is that the accuracy of the sensor device must allow for an adequate monitoring of significant day-to-day production changes of an individual animal with representative sampling and when measured with reference methods. For that, the accuracy of the analytical device should be better than the natural day-to-day fluctuation of the measured parameter to allow the signalling of the relevant variation. The indicated limits for precision and accuracy for on-farm analytical devices are based on the ratio between the limit of the standard deviation of analytical measurement at the farm and the limit of the standard deviation of analytical measurement at the laboratory, expressed in a so-called equivalence factor (FE). Starting from the day-to-day fluctuation in fat concentration as being the maximum value for accuracy of analytical measurement on-farm, the statistical limits for precision and accuracy were calculated from those stated for laboratory analysers, see Table 2. For somatic cell counting devices the stated limit values for the relative standard deviations of repeatability and reproducibility are dependent on the level and range from 20% to 5% and 25% to 6% respectively. The indicated limit value for accuracy over the whole range is 25%. The SD-TF is currently considering the adequateness of these limits. Moreover, it is intended to precise how multiple measurements should be accounted for in the limit performance values of a single measurement.

Table 2. Precision and accuracy limits for test bed evaluation of milk analysers in milk recording (ICAR, 2017)

Component Units		Fat g/ 100 g	Protein g/ 100 g	Lactose g/ 100 g	Urea mg/ 100 g	SCC 10^3 cells/ml
Range	Total			4,0 - 5,5	10,0 – 70,0	0 – 2000
	Low					0-100
	Medium	2,0 - 6,0	2,5 - 4,5			100-1000
	High	5,0 - 14,0	4,0 - 7,0			> 1000
Sample number	Animals (Na)	100	100	100	100	100
	Herds (Nh)	5	5	5	5	5

Milk analytical devices		Laboratory			On-farm At-line			On-farm In-line		
Equivalence Factor	FE		x 1			X 2			x 2,5	
Component Units		F-P-L g/ 100 g	Urea mg/ 100 g	SCC percent	F-P-L g/ 100 g	Urea mg/ 100 g	SCC percent	F-P-L g/ 100 g	Urea mg/ 100 g	SCC percent
Repeatability								[a]	[a]	[a]
Standard deviation *(sr)*	- Total range			4%			8%			10%
	- Low			8%			16%			20%
	- Medium	0,014	1,4	4%	0,028	2,8	8%	0,035	3,5	10%
	- High	0,028	2,8	2%			4%			5%
Within lab reproducibility										
Standard deviation *(sR)*	- Total range			5%			10%			13%
	- Low			10%			20%			25%
	- Medium	0,028	2,8	5%	0,056	5,6	10%	0,069	6,9	13%
	- High	0,056	5,6	2,50%	0,056	5,6	5%	0,070	7,0	6%
Accuracy										
Animal sample SD *(sy,x)*	- Total range			10%			20%			25%
	- Low									
	- Medium	0,10	6,0		0,20	12,0		0,25	15,0	
	- High	0,20			0,20 [b]			0,25 [b]		
Calibration [c]										
Mean bias ($\bar{d}$)	- Total range		± 1,2	± 5 %		± 2,4	± 10 %		± 3,0	± 13 %
	- Medium	±0,05			±0,10			±0,13		
	- High	±0,10			±0,20			±0,25		
Slope *(b)*		1±0,05	1±0,10	1±0,05	1±0,10	1±0,10	1±0,10	1±0,13	1±0,10	1±0,13

[a] Where relevant i.e. for in-line differed time analysis.
[b] No larger tolerance by the usual factor 2 for sheep and goat to maintain accuracy with no more numerous records.
[c] Compared to manufacturer calibration.

Next steps

Based on the outlined purpose-based approach, the next steps in the work of the ICAR SD-TF can be identified as:

- Extending the sensor device database, to be completed with missing sensors and missing information on available and/or applied sensors and to provide an overview with key characteristics on the ICAR website.
- Developing sensor validation criteria for different purposes, based on working through a checklist for a structured evaluation on:
  - What is the trait?
  - What is the purpose of the sensor measurement?
  - What type of data does the sensor produce?
  - What is the anchor?
  - Single measurement or multiple measurements?
- Developing ICAR validation protocols and ICAR approval protocols, first for udder health sensors, then for other traits;
- Dissemination of recording guidelines using sensor data;
- Dissemination of best practices for data collection from sensor devices and related quality assurance.

List of references

Eurachem. 2014. The fitness for purpose of analytical methods: A Laboratory Guide to Method Validation and Related Topics, Second Edition.

Hogeveen, H., Kamphuis, C., Steeneveld, W. & Mollenhorst, H. 2010. Sensors and Clinical Mastitis - The Quest for the Perfect Alert. Sensors, 10: 7991-8009.

International Committee on Animal Recording. 2016. ICAR Accuracy Task Force Report.

International Committee on Animal Recording. 2017. ICAR Guidelines Section 13 - Guidelines on On-farm Milk Analysis.