DOI 10.5162/16dss2022/2.3

MOS Sensors for Spoilage Detection of Milk using GC-MS and Human Perception as Reference

MOS-Sensoren für die Verderbserkennung von Milch mit GC-MS und Humansensorik als Referenz

Julian Joppich, Zhuo Su, My Sa Marschibois, Ksenia Karst, Christian Bur, Andreas Schütze

Saarland University, Lab for Measurement Technology, Saarbrücken, Germany
Contact: j.joppich@LMT.uni-saarland.de

Introduction

In order to assess the edibility of food, consumers particularly pay attention to the best-before date (BBD) or the use-by date [1]. While the best-before date declares the period of impeccable quality, the use-by date declares the period of safe consumption; however, many consumers are not aware of the different meaning of these two dates [1]. There is also no common definition in the EU which of the two dates is used for which products [2]; in Germany, the best-before date is used for milk and dairy products.

As a result, products are often precautionary discarded when the best-before date is exceeded (or even just reached), although consumption would often be possible without concerns even after this date [3]. In these cases, it would make more sense to assess the edibility of food, e.g., using the senses (sight, smell, taste); in this way, it is also possible to detect potential spoilage prior to the expiration of the best-before date, e.g., due to an interrupted cold chain.

There are several methods to investigate spoilage of milk. Besides microbiological methods, that often include use of a growth medium and incubation times (often several hours) [4] and are, thus, not suitable for rapid testing, spoilage of milk can be measured, e.g., via its pH value [5] or using infrared spectroscopy [6]. The use of different kinds of gas sensors or sensor arrays to measure gases or volatile organic compounds (VOCs) which are, e.g., produced by microorganisms involved in spoilage was also reported [7, 8], as well as investigations of actual milk spoiling over time using gas sensor arrays [9, 10]. Measuring microbial (by)products is closely related to the mentioned spoilage detection using the human nose, and it is supposed to be conclusive as the concentration of several VOCs was reported to correlate with microbial counts in milk samples under various storage conditions [11].

In this work, the applicability of metal oxide semiconductor (MOS) gas sensors in temperature cycled operation (TCO, [12]) for reliable detection of spoilage of dairy products by measuring the gas space above the respective product (headspace) was investigated. Gas sensor data are combined with several reference data sources (GC-MS, pH value, and human sensory evaluation) with the aim to safely estimate the edibility as a potential alternative to the best-before date.

Materials and Methods

Measurement Setup

A measurement setup developed for the automated measurement of stored foods [13] was adapted for the measurement of the headspace of dairy products (see Figure 1). The measurement setup consists of a measurement system (special version of the OdorCheckerSpot, 3S GmbH – Sensors, Signal Processing, Systems, DE) including a pump and an interface to a valve multiplexer, enabling automatic sampling from up to 32 sources individually using valves; these sources originally had been food storage boxes in a refrigerator. The air sample is passed through a so-called sampling port and several gas sensor systems. The sampling port is an aluminum block with a septum. It allows drawing samples using an autosampler equipped with a gastight syringe and to analyze the composition of the air drawn through the setup with a gas chromatograph - mass spectrometer (GC-MS; GC: Thermo Fisher Scientific Trace 1300; MS: Thermo Fisher Scientific ISQ 7000 Single Quadrupole Mass Spectrometer; column: TG-624 60 m, ID 0.25 mm, layer 1.4 µm; temperature program; S/SL injector; headspace injection). The GC-MS was introduced in order to draw conclusions on relevant substances in the food headspace and to obtain reference data for the subsequent gas sensor data evaluation. The additional MOS gas sensors displayed above the GC-MS in Figure 1 serve as alternative GC sensors [14] and are not discussed further within this work. The sensors used include various digital and analog temperature cycled MOS gas sensors, two electrochemical (EC) cells plus a photoacoustic carbon dioxide (CO_2) sensor (SCD41, Sensirion AG, CH). The temperature cycle of the MOS sensors lasts one minute and consists of five high temperature phases of 400 °C (duration 5 s each), followed by low temperature phases at 150, 200, 250, 300, and 350 °C (12 s each). The following MOS sensors were used: SGP30 (Sensirion), BME688 (Bosch Sensortec GmbH, DE), ZMOD4450 (Renesas Electronics Corporation, JP) as well as various analog

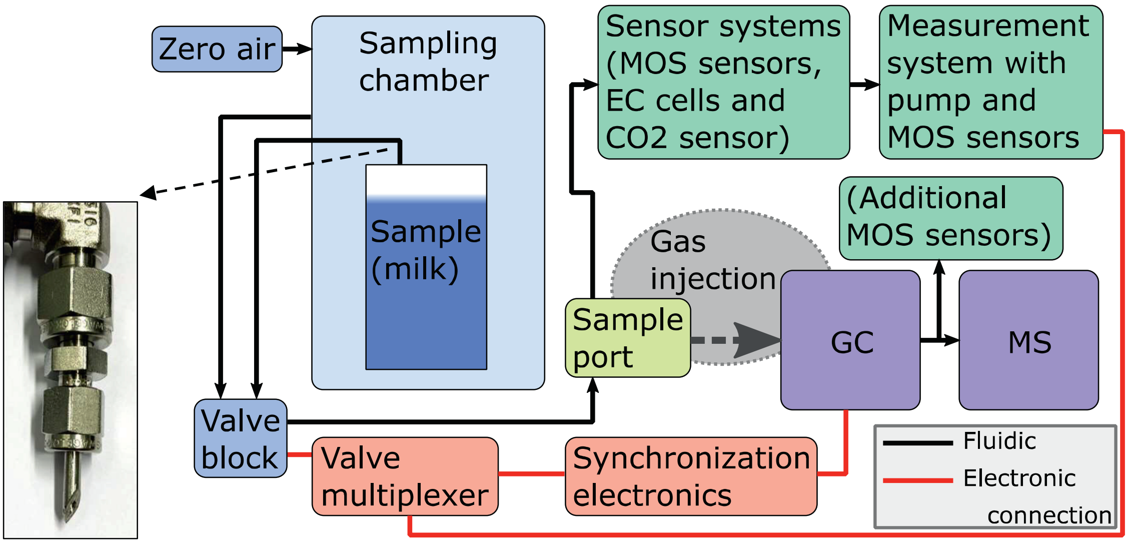

Figure 1: Overview of the measurement setup including the novel sampling chamber, GC-MS, several sensor systems and the measurement system, which controls the internal pump and the external valve system. Inset: Cannula-like tip to pierce the cap of dairy product packages.

sensors manufactured by UST (UST Umweltsensortechnik GmbH, DE). The EC cells were two B-series sensors for hydrogen sulfide and ammonia (H2S-B4 and NH3-B1, respectively, both Alphasense Ltd., GB). The results presented are obtained with the digital gas sensors (MOS and CO_2).

In the adapted setup, a sampling chamber is used instead of food storage boxes. The chamber provides a clean environment around the milk (or other dairy product) package by flushing the volume with zero air. The headspace of the product is accessed using a tip, similar to a cannula, made from stainless steel tube (see Figure 1, inset), attached to a PTFE tube. The tip can pierce the cap of the dairy product package, thus making the headspace accessible without previous dilution or contamination. To allow for pressure equalization during sampling, several additional very small holes (diameter < 1 mm) are punched into the package cap using a sharp tool.

Measurements and Human Evaluation Approach

Comprehensive measurement series were carried out with representative dairy products of different categories, including whipping cream, natural yoghurt, and whole milk; only the latter is discussed within this work. The milk chosen was pasteurized milk, i.e., explicitly not "extended shelf live" (ESL), from two manufacturers (M1 and M2). Three different batches were purchased from manufacturer 2 (M2-B1, -B2, -B3). Thus, four batches were used in total. M2-B3 was measured in a second measurement series nine months after the other batches had been tested. The milk was stored in a refrigerator, set to 6 °C.

One milk sample from every batch was measured daily with the measurement setup; in the second measurement series, three packages of the same batch (M2-B3) were sampled every day. The measurement period was 26 days (M1-B1, with several gaps), 11 days (M2-B1), 12 days (M2-B2, with one gap) or 23 days (M2-B3, with two gaps) and was supposed to cover enough time both before and after the BBD. Each previously unopened package was accessed with the clean sampling tip, the gas space was extracted from the package, passed over the sensor systems and, additionally, analyzed with the GC-MS. The headspace sampling time was 10 minutes to ensure sufficient time for several gas sensor cycles; the flow was set to 200 ml/min, and the timing of the GC-MS sampling was set in such a way that the sample is extracted at about the time when the maximum headspace concentration reaches the sample port (about 2 min after the valve switching). Each GC-MS run lasted 45 min including cooldown.

Additionally, every milk package was assessed by human sensory evaluation with respect to appearance, odor and edibility. The evaluation scale was defined from 1 to 10, with fresh milk starting at 10; see Table 1 for details. The evaluations were performed by untrained but instructed students. Whenever possible, several persons evaluated the same milk sample to reduce the effect of individual differences in perception of spoilage.

Table 1: Spoilage evaluation scheme

Scale	Meaning (e.g.)
10	normal appearance / unremarkable smell / edible
6	… / just edible
5.5	threshold edible – not edible
5	… / just not edible
1	completely curdled / very unpleasant odor / not edible

DOI 10.5162/16dss2022/2.3

The pH value was measured with different kinds of pH indicator paper. Both universal indicator paper, i.e., pH 1-14 with a resolution of 1, and, in the second measurement series, several papers with a narrower measurement range (about 5-7) and a resolution of 0.2-0.3 were used, as well as a portable electronic pH meter (P 5315, PeakTech GmbH, DE) with a resolution of 0.01. The ranges of the pH indicator papers were chosen according to literature values, stating a pH value of about 6.7 for fresh milk and values around 4 after spoilage [5].

Results

Human Evaluation, pH Value, GC-MS Results

In all milk batches tested, spoilage was detected by human sensory evaluation during the measurement series (see Figure 2). In case of M1-B1, the ratings were rather unconclusive, and both M1-B1 and M2-B2 were only measured after the BBD because of a delayed measurement start. M2-B1 was actually already spoiled at the start. The optical changes described include the formation of lumps and flakes (curdling or coagulation). The odor was described as “none” or “regular milk” at the beginning and as increasingly sour with ongoing spoilage; after a longer time of spoilage, it was also described as “stable-like”. All these changes led to decreasing edibility ratings. Overall, the evaluation was very similar regarding the three evaluation dimensions, thus, there was a high correlation between the assessment of appearance, smell, and edibility. Moreover, the human evaluations were comparable between different persons; in most cases, the evaluation differed by a maximum of 1, in a few cases by 2, on the evaluation scale (average standard deviation 0.27/0.37/0.27 for appearance, smell, edibility). When several milk packages of the same batch were opened on the same day in M2-B3, the condition evaluations were similar in the first days. However, from the 6th day on, which was also the BBD, the standard deviation between the milk packages increased to over 0.5 and up to a maximum of 2.31/2.02/2.08 (on the evaluation scale of appearance, odor, and edibility; cf. error bars in Figure 2, right), with the standard deviation of odor reaching maximum on the BBF and the standard deviation of perception and edibility reaching maximum 4 days after BBD, indicating differences in the spoilage (progress) of the individual packages. These differences decreased slightly at the end of the measurement after 12 days (6 days after BBD), indicating a finally rather similar spoilage progress, i.e., even the slightly more durable milk packages finally spoiled.

The pH value did not change significantly during the main measurement window around the BBD. In case of the universal indicator paper, no changes were visible, and the measurements with the other indicator papers were unconclusive (both regarding differences between the indicator papers and the changes over time), partly due to the uncertain color readings, which changed only slightly. Also, the electronic pH meter did not indicate significant changes during the first 12 days (up to 6 days after the BBD; see Figure 3). Measurements added after the main measurement duration, at least 13 days after the BBD, finally showed pH values below 6, reaching a minimum of 5.6 17 days after the BBD (last measured pH value).

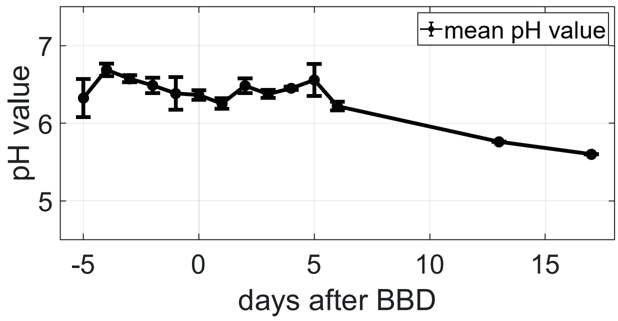

Figure 3: Milk pH values (only M2-B3) measured with the digital pH meter. Error bars indicate the standard deviation between the three packages sampled each day.

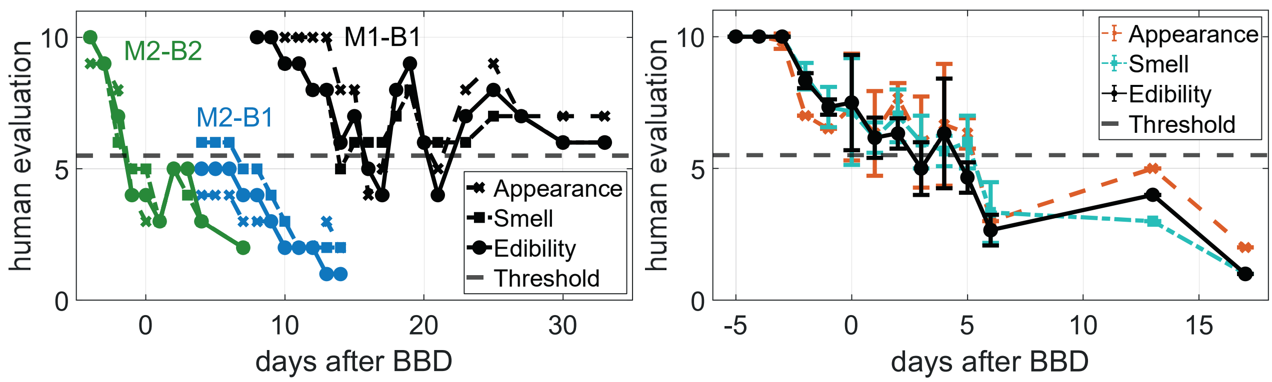

Figure 2: Human sensory evaluations of all milk batches. Left: First measurement series (M1-B1, M2-B1, M2-B2); M2-B1 was already spoiled at the beginning due to a delayed start. Right: Second measurement series (M2-B3); error bars indicate the standard deviation between the three packages sampled each day.

The substances produced during milk spoilage identified by GC-MS include CO_2 as well as some VOCs (mainly acetaldehyde, acetone, butanone, and ethanol). However, most substances could only be identified in completely spoiled milk, or their peak area, i.e., their headspace concentration, did not show a conclusive trend over time. Only the peak area of CO_2 increased fairly systematically with time (see Figure 4). The peak area is approx. 10 times higher than background (zero) air for non-spoiled milk and 20 to 100-200 times higher for spoiled milk. Note that the CO_2 peak area of M1-B1 remained at the level of non-spoiled milk during the whole measurement. However, high variability of the CO_2 area was observed for both non-spoiled and spoiled milk.

To estimate the concentration of methanol, ethanol, acetone, and butanone in milk, samples were spiked with appropriate amounts that should generate headspace concentrations of approx. 100 ppm at 20 °C. Based on these tests, the concentrations of these VOCs even in highly spoiled milk are estimated to be below 10 ppm at most for ethanol, others (e.g., butanone) are in the range of a few hundred ppb only. Concentrations below these values (depending on the given substance) cannot be determined reliably by the GC-MS with the chosen injection method and measurement setup because the peak areas are too small or signal-to-noise ratios too low (below detection threshold).

Gas Sensor Data Evaluation

The gas sensor data were evaluated by means of raw data preprocessing (natural logarithm of the conductance), feature extraction (mean value and slope of 60 1 s segments of the temperature cycled MOS sensors; 1 min mean value of the CO_2 sensor), feature standardization and model training, including validation and testing, using the MATLAB toolbox DAV³E [15]. Of the 10 minutes sampling time, only the first four full cycles (4 minutes) were used. Both classification and quantification models were trained regarding the assessed milk edibility. For the classification models, the edibility ratings were summarized into new groups, so that there are only four groups of edibility: original scale 1-3 = new scale D (spoiled/not edible), 4 and 5 = C (just not edible), 6 and 7 = B (just edible) and 8-10 = A (fresh/edible). If not stated otherwise, 10-fold cross validation that ignores complete "days after BBD" ("group-based") and 20 % holdout testing of complete "days after BBD" were applied, and both validation and testing were repeated 5 times (with a new split each time) to check model stability.

For quantification, PLSR (partial least squares regression) was used. In all cases, the number of PLSR components is 4. A model trained with only the first measurement series reaches RMSE (root mean squared error) values for training and validation of 1.2 and 1.5, respectively (on the human evaluation scale); projecting data of the second measurement series (=M2-B3) into that model leads to an RMSE for testing of 4.9, and the data points lie too high (mostly above 10); only data points of very spoiled milk are projected approximately correctly. Correspondingly, if data of M1-B1 are projected into a model trained with data of manufacturer 2 only, the RMSE values are 1.3/1.6/2.6, and the data points are projected quite stably at around 8-9. This means that the individual models do not allow generalization. If all milk data are taken into the training data set, the RMSE values are 1.4/1.7/1.9, indicating that a universal model is possible, albeit with higher uncertainty. Furthermore, the applicability of this model to other milk samples (other batches or other manufacturers) is questionable.

Additionally, recursive feature elimination (RFE, using the linear coefficients of a least squares regression to rank the features and recursively eliminate the least significant one; also described in [16]) was applied prior to training a PLSR model using all milk data, improving the model performance (see Figure 5, left). 21 features were selected, the number of PLSR components is 10, the RMSE values are 1.0/1.2/1.4.

For classification, LDA (linear discriminant analysis) was used with 3 discriminant functions, combined

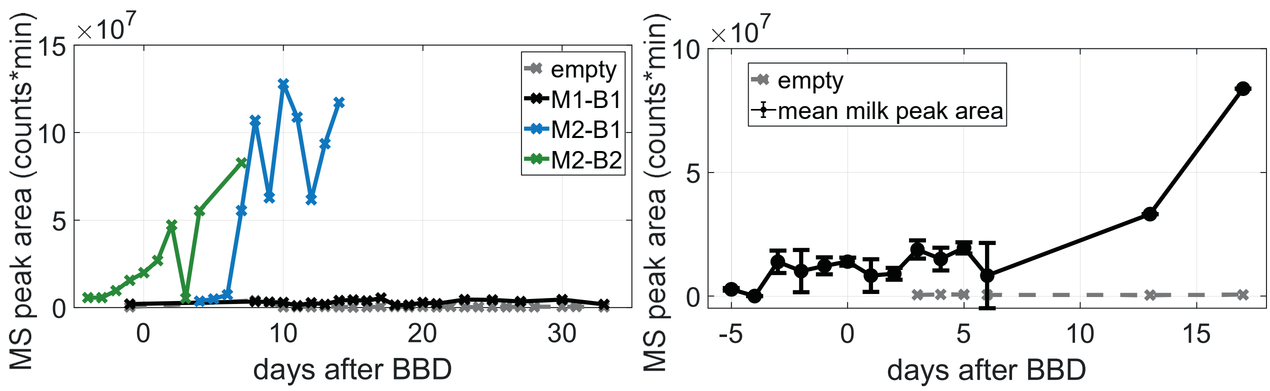

Figure 4: Left: MS peak area for CO_2 during the first measurement series; empty is aligned to the time of M1-B1. Right: MS peak area for CO_2 during the second measurement series (M2-B3).

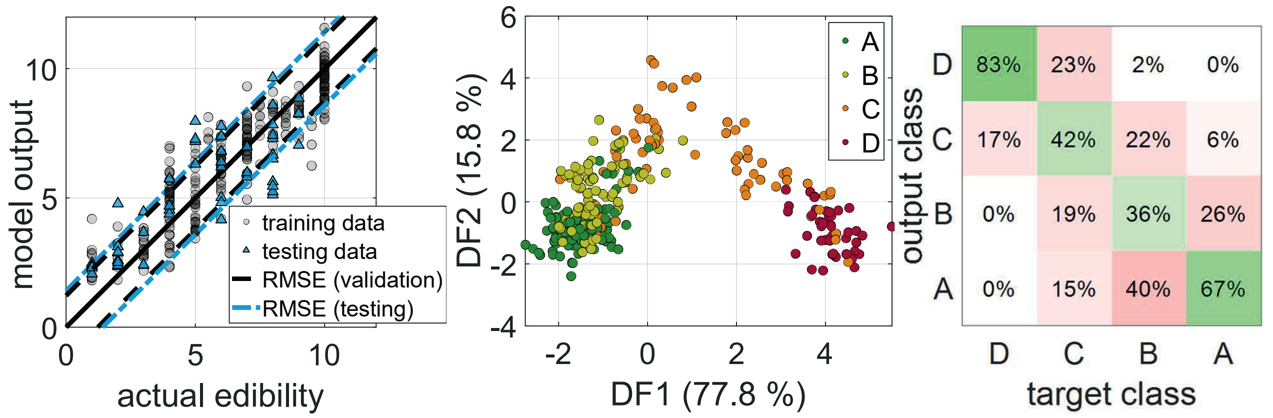

Figure 5: Left: PLSR model of all milk data, using SVR. Middle: LDA model (first two discriminant functions (DF), of three in total) of all milk data; the four edibility classes build a path from fresh (A) to spoiled (D). Right: Confusion matrix of the LDA model, indicating good separation of the outermost classes.

with a 5 nearest neighbor classifier. Prior to the LDA, a PCA (principal component analysis) with 20 principal components was performed to reduce the dimensionality of the feature space (which is 601) and to reduce the risk of overfitting. Similar as in the case of the PLSR models, if the second measurement series is projected into an LDA model trained only with data of the first measurement series (classification error of training and validation 11 % and 33 %, respectively), the classification error is very high (63 %). In fact, the new data lie separate from most training data, and only the last data points for very spoiled milk correctly approach the "spoiled milk" cluster of the training data. An LDA trained with all data reaches classification errors of 17/44/49 % (see Figure 5, middle). The errors seem quite high, however, rather small misclassification rates are observed between the "edible" groups and the "spoiled" groups, e.g., spoiled milk is never wrongly classified as edible milk (see Figure 5, right). Only the two "just (not) edible" groups are not as well separated (approx. 20 % misclassification), which actually is to be expected, as spoilage is a continuous process and there is no distinct boundary between either of the defined states. In fact, this might be due to the uncertain human assessment.

Discussion

While the substances identified with GC-MS in general are in good agreement with literature [5, 11, 17], the evolution of the peak area and, thus, their concentration differed greatly between different manufacturers, between different batches, and even between different packages within the same batch, especially observed for CO_2. The difficulties interpreting the GC-MS results are partly caused by the fact that the concentrations seem to be close to or below the limit of detection of the setup used. Furthermore, the pH value did not change significantly during the spoilage process, only long after spoilage was clearly evaluated by the human perception. Both the inconclusive evolution of the peak areas and of the pH value, which also changed less than expected, might be attributed to differences in numbers and species of microorganisms contained in the individual packages (especially regarding different manufacturers), leading to variations in the type and concentration of substances produced. Moreover, the correlation of the human evaluations (appearance/smell/edibility) with the other two indicators mentioned is limited, which might be due to various spoilage processes occurring in the milk, leading to differences in the considered measurands. E.g., if the changes observed by human evaluation are a result of processes that cannot be covered completely by the pH value and/or the GC-MS measurements (possibly because the substances involved lie below the limit of detection of the headspace analysis), the cause of the spoilage cannot be further analyzed. On the other hand, e.g., CO_2 cannot be perceived by the human nose; processes producing primarily CO_2 are, thus, "hidden" for the human perception (apart from bubble formation). Even though CO_2 is not harmful for the human body in the observed concentration range, it could still be a valuable indicator of spoilage processes, either in the sense of harmless microorganisms changing the expected quality of the product, e.g., during (hetero)fermentation [4] or, more importantly, of other, harmful microorganisms.

As a result, a correlation of the human evaluation with the peak areas for CO_2 or various VOCs or with the pH value cannot really be expected; this similarly applies to the MOS gas sensor data. Thus, the human evaluation is still assumed to be a valid reference data source.

In fact, it could be shown that the models based on the MOS gas sensor data are able to project the extracted features to the human evaluation, both in case of quantification and classification tasks. However, part of the model errors might be attributed to the uncertainty of the human evaluation, and, thus, further optimization of the evaluation approach is required, e.g., by using trained people, because a measurement can only be as good as its calibration. Additionally, the transferability of the obtained models to other milk samples (other batch, different manufacturer, different kind of milk (e.g., ESL), etc.) is expected to be very limited based on the current results. This might be due to the broad spectrum of possible microorganic contaminations and, consequently, of possible substances produced during spoilage.

It is worth mentioning that hydrogen (H_2) is also a possible byproduct of microorganisms [4], which is neither perceivable by the human nose nor measurable with the GC-MS, whereas MOS gas sensors typically have a high sensitivity towards H_2. Thus, if H_2 is produced during spoilage and the corresponding signal is used by the MOS gas sensor models, it is not covered by any of the reference methods and only indirectly via the human perception.

Conclusion and Outlook

A measurement setup and approach for monitoring the spoilage of milk including human perception was presented. Several substances could be identified with the GC-MS occasionally, but only the peak area of CO_2 was found to change fairly systematically. The pH value changed significantly, but only in the long run for very spoiled milk. Human perception determined the spoilage in all cases; however, one batch was rather inconclusive. In general, variations in the GC-MS data, the perception, or the pH value, especially within the same batch, could indicate different initial contaminations of the individual packages and, thus, a different course of spoilage, leading to a rather difficult evaluation, especially regarding the correlations to the sensor data. However, reasonable quantification and classification models could be obtained. Likewise, restrictions could be identified, e.g., the limit of detection of the GC-MS and the transferability from trained to new milk data. In order to further verify the models as well as to extend them with "extended shelf life" milk, further measurement series are currently being carried out with additional batches and milk from other manufacturers/brands.

Literature

[1] H Møller et al.: *Food waste and date labelling: Issues affecting the durability*. Nordic Council of Ministers, 2016

[2] European commission: *Market study on date marking and other information provided on food labels and food waste prevention*. Final Report, Publications Office of the European Union, 2018

[3] P. Lee, S. Osborn, P. Whitehead: *Reducing food waste by extending product life*. Final Report, WRAP, 2015

[4] R. K. Robinson: *Microbiology Handbook*. John Wiley and Sons, Inc., 2004

[5] H. M. Østlie, M. H. Helland, J. A. Narvhus: Growth and metabolism of selected strains of probiotic bacteria in milk. *Int J Food Microbiol* 87 (2003), No. 1-2, pp. 17-27

[6] H. M. Al-Qadiri, M. Lin, M. A. Al-Holy, A. G. Cavinato, B. A. Rasco: Monitoring quality loss of pasteurized skim milk using visible and short wavelength near-infrared spectroscopy and multivariate analysis. *J Dairy Sci* 91 (2008), No. 3, pp. 950-958

[7] N. Magan, A. Pavlou, I. Chrysanthakis: Milk-sense: a volatile sensing system recognises spoilage bacteria and yeasts in milk. *Sens Actuators B: Chem* 72 (2001), No. 1, pp. 28-34

[8] J. E. Haugen, K. Rudi, S. Langsrud, S. Bredholt: Application of gas-sensor array technology for detection and monitoring of growth of spoilage bacteria in milk: A model study. *Analytica Chimica Acta* 565 (2006), No. 1, pp. 10-16.

[9] M. A. Putra, M. Rivai, A. Arifin: Milk assessment using potentiometric and gas sensors in conjunction with neural network. *Proceedings of 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Bali, August 2018, pp. 409-412

[10] N. Phukkaphan, T. Eamsa-ard, C. Chairanit, T. Kerdcharoen: The application of gas sensor array based electronic nose for milk spoilage detection. *Proceedings of 2021 7th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, Online, April 2021, pp. 273-276

[11] M. Ziyaina, B. Rasco, T. Coffey, D. S. Mattinson, S. Sablani: Correlation of volatile compound concentrations with bacterial counts in whole pasteurised milk under various storage conditions. *Int J of Dairy Technol* 72 (2019), No. 1, pp. 36-46

[12] A. Schütze, T. Sauerwald: Dynamic operation of semiconductor sensors. In: R. Jaaniso, O. Kiang Tan (Eds.): *Semiconductor Gas Sensors*, 2nd ed., Woodhead Publishing, 2020, pp. 385-412

[13] J. Joppich, M. Marschibois, T. Baur, O. Brieger, C. Schultealbert, A. Schütze: Monitoring food aging in a refrigerator with GC/MS and gas sensor systems. *Proceedings of 2021 Sensor and Measurement Science International (SMSI)*, Online, May 2021, pp. 121-122

[14] O. Brieger, J. Joppich, C. Schultealbert, T. Baur, C. Bur, A. Schütze: Microstructured MOS gas sensor as GC detector. *Proceedings of 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*, Aveiro, May/June 2022, ID 8144

[15] M. Bastuck, T. Baur, A. Schütze: DAV3E – a MATLAB toolbox for multivariate sensor data evaluation. *J. Sens. Sens. Syst.* 7 (2018), No. 2, pp. 489-506

[16] T. Baur, J. Amann, C. Schultealbert, A. Schütze: Field study of metal oxide semiconductor gas sensors in temperature cycled operation for selective VOC monitoring in indoor air. *Atmosphere* 12 (2021), No. 5: 647

[17] R. Fernandes: *Microbiology Handbook: Dairy Products*. Leatherhead Food International Ltd., Royal Society of Chemistry, 2009

Acknowledgment

This work is sponsored by the DBU – Deutsche Bundesstiftung Umwelt in the frame of the project SiVERiS under reference 34806/01-36. We thank our project partner 3S GmbH – Sensors, Signal Processing, Systems, for the excellent cooperation.