Resources, Conservation & Recycling 201 (2024) 107256

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Near-infrared-based quality control of plastic pre-concentrates in lightweight-packaging waste sorting plants

Nils Kroell [a,*], Xiaozheng Chen [b], Bastian Küppers [b], Sabine Schlögl [c], Alexander Feil [a], Kathrin Greiff [a]

[a] Department of Anthropogenic Material Cycles, RWTH Aachen University, Wuellnerstr. 2, D-52062, Aachen, Germany
[b] STADLER Anlagenbau GmbH, Max-Planck-Str. 2, D-88361, Altshausen, Germany
[c] Chair of Waste Processing Technology and Waste Management, Montanuniversity Leoben, A-8700, Leoben, Austria

ARTICLE INFO

Keywords:
Sensor-based material flow characterization
Automated quality control
Inline quality monitoring
Circular economy
Mechanical post-consumer plastic recycling
Machine learning

ABSTRACT

Today's post-consumer plastic recycling is limited by labor-intensive manual quality control (MQC) procedures, resulting in largely unknown pre-concentrate purities. Sensor-based quality control (SBQC) could enable an automated inline quality monitoring and thus contribute to a more transparent and enhanced plastic recycling. Therefore, we investigated the technical feasibility of near-infrared-based SBQC for plastic pre-concentrates in a lightweight packaging waste sorting plant. The developed SBQC method outperformed MQC methods by reducing measurement uncertainties from between ±0.8 wt% and ±6.7 wt% (MQC) to ±0.31 wt% (SBQC) for bale-specific purities at monolayered material flow presentations. In addition, we show that SBQC may even be possible at multilayered material flow presentations, although further research is needed to address identified segregation effects. The demonstrated technical feasibility of SBQC at plant scale represents a major breakthrough as it opens new opportunities in plastic recycling, such as adaptive pricing models and intelligent process control in sorting plants.

Abbreviations

BC	beverage carton;
BG	background;
CSV	comma-separated value;
DSD	Der Grüne Punkt Duales System Deutschland GmbH;
EU	European Union;
HDPE	high-density polyethylene;
LWP	lightweight packaging;
ML	machine learning;
MQC	manual quality control;
MU	measurement uncertainty;
NIR	near-infrared;
nMU	normalized measurement uncertainty;
PCC	Pearson correlation coefficient;
PET	polyethylene terephthalate;
PP	polypropylene;
PPC	paper, paperboard, and cardboard;
PS	polystyrene;
RGB	red, green, blue;
RQ	research question;
SBQC	sensor-based quality control;
SP	sensor position;
UNDEF	undefined

1. Introduction

In 2020, the European Union (EU) 27+3 collected 29.5 Mt/a of post-consumer plastic waste, but only 18.6 wt%[1] to 22.0 wt%[2] of the collected plastics were turned into plastic recyclates, while the remainder was incinerated or landfilled (Plastic Europe, 2022b). Therefore, massive improvements in post-consumer plastic recycling are needed to make a relevant contribution to climate and resource protection (European Commission, 2019; United Nations, 2015) and a competitive circular economy (Bachmann et al., 2023; European Commission, 2020).

* Corresponding author at: Department of Anthropogenic Material Cycles, RWTH Aachen University, Wuellnerstr. 2, D-52062, Aachen, Germany.
E-mail address: nils.kroell@ants.rwth-aachen.de (N. Kroell).

[1] Assuming all exported material was incinerated or landfilled.
[2] Assuming all exported material was processed into 100% recyclates.

https://doi.org/10.1016/j.resconrec.2023.107256
Received 26 May 2023; Received in revised form 6 September 2023; Accepted 10 October 2023
0921-3449/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Plastic packaging is the largest source of post-consumer plastic waste in the EU 27+3, amounting to 17.9 Mt/a (60.7 wt% of the total post-consumer plastic waste) in 2020 (Plastic Europe, 2022a). Recycling of post-consumer packaging waste involves three stages: plastic waste collection based on country/region-specific collection schemes, sorting into material-specific pre-concentrates in *sorting plants*, and processing pre-concentrates into recyclates in *processing plants* (Feil and Pretz, 2020). Generated plastic recyclates can then substitute primary plastics during production processes and achieve significant environmental benefits, e.g., in terms of energy and greenhouse gas emission savings (Bachmann et al., 2023; Cudjoe et al., 2021).

As the crucial intermediate between collection and processing, the *quality of pre-concentrates* generated in sorting plants significantly affects the overall process efficiency in post-consumer plastic recycling (Feil and Pretz, 2020). Low-quality pre-concentrates can lead to significant material losses and to sub-optimal recyclate quality in processing plants (Dehoust and Christiani, 2012), thereby hampering the substitution of primary plastics and ultimately limiting the achievable environmental benefits (bvse-Fachverband Kunststoffrecycling, 2017; Hahladakis and Iacovidou, 2018). Suboptimal pre-concentrate qualities have been criticized in the past (e.g., bvse-Fachverband Kunststoffrecycling, 2016; 2017; EU Recycling, 2018) and significant deviations from purity specifications have been documented (bvse-Fachverband Kunststoffrecycling, 2017; Knappe et al., 2021).

1.1. Limitations of manual quality control

Currently, the quality of plastic pre-concentrates is determined by manual quality control (MQC), i.e., sampling and manual sorting analysis (Borowski, 2018; Der Grüne Punkt, 2016). Because of the necessary sampling and manual sorting analysis, MQC is time-consuming and cost-intensive, and thus often only performed on an irregular basis (Borowski, 2018). For example, Borowski (2018) reports that a quality control team consisting of two to three experienced employees needs about two hours to analyze a single pre-concentrate bale (=4–6 person hours per bale), while modern sorting plants produce several hundred bales per day. Hence, quality control with MQC methods can only be performed on a spot-check basis for selected bales.

The lacking transparency on batch-specific pre-concentrate qualities currently results in pre-concentrates being traded on a *flat-rate basis*, i.e., irrespective of their (bale-specific) quality (plasticker, 2023). Thus, combined with high disposal costs for residual fractions, sorting plant operators have thus a business interest in sorting only "as good as necessary" instead of "as good as possible", since better sorting would lead to higher disposal costs for the higher quantity of the then generated sorting residues as well as lower income due to the reduced overall quantity of pre-concentrates (Knappe et al., 2021). In addition, technical optimizations in sorting plants can be challenging due to the missing benchmark without a reliable and in-time measurement of produced pre-concentrate qualities (Kroell et al., 2022a).

1.2. Opportunities through sensor-based quality control

Sensor-based quality control (SBQC) could enable automatic, real-time, and batch-specific knowledge of the quality (i.e., purity) of pre-concentrates and thus considerably enhance post-consumer plastic recycling (Kroell et al., 2022a):

1.2.1. Adaptive pricing models

By monitoring the quality of each batch, the pricing of pre-concentrates could transition from a flat-rate basis to adaptive pricing models, where higher pre-concentrate quality results in higher prices, thereby adjusting the incentives for sorting plant operators.

1.2.2. Process optimization in sorting plants

Automatic monitoring of pre-concentrate qualities would allow for timely detection and correction of quality deficits, e.g., by using adaptive process control of the sorting plant to improve overall sorting quality.

1.2.3. Evaluation and optimization of separate collection

Information on material composition and quantity of output fractions could be used to calculate back to the sorting plant input (collected mixed lightweight packaging waste), thus, enabling the evaluation and optimization of separate collection (e.g., assessment of public campaigns to improve separate collection [Initiative „Mülltrennung wirkt“, 2021]).

1.2.4. Process optimization in processing plants

By sharing data along the value chain, processing plants could benefit from known input qualities, e.g., through input-adaptive process parameterization to varying pre-concentrates qualities or an optimized distribution of pre-concentrate qualities across processing plants with different processing technologies and depths.

1.3. Related work and research gap

Data analysis for SBQC applications can be divided into pixel, particle, and material flow levels (Kroell et al., 2022a). At the pixel and particle level, numerous studies (e.g., Bonifazi et al., 2018; Calvini et al., 2018; Chen et al., 2021; Duan and Li, 2021; Xia et al., 2021; Zheng et al., 2018) have demonstrated that non-carbon-black plastics can be differentiated with $\geq$ 99% accuracy using near-infrared (NIR) spectroscopy.

However, the quality of plastic pre-concentrates is not defined in terms of pixels and particles but by the mass-based composition of the entire material flow, i.e., the aggregation of thousands of particles (cf. Section 2.1). Yet, little research has been conducted for sensor-based determination on mass-based material flow composition, which comes with a new set of challenges such as finding appropriate data aggregation methods, area-to-mass-conversion, and different counting basis (cf. Section 2.2) (Kroell et al., 2022a).

Kroell et al. (2023a) recently demonstrated the technical feasibility of predicting mass-based material flow compositions from NIR-based false-color data using machine learning (ML). Through regression models and data aggregation, it was possible to predict mass-based material flow compositions of binary post-consumer plastic packaging mixtures with a measurement uncertainty of $\pm$2.0 wt% across different material flow presentations at technical lab scale (Kroell et al., 2023a; Kroell et al., 2023b). However, to our best knowledge, the technical feasibility of NIR-based quality control of plastic pre-concentrates at plant scale has yet to be shown.

1.4. Aim and research questions

With the present paper, we aim to demonstrate the technical feasibility of inline-SBQC of plastic pre-concentrates from lightweight-packaging (LWP) sorting at plant scale by answering the following four research questions (RQs):

- **RQ 1:** What prediction accuracies can be achieved for monitoring impurity contents using NIR sensors at monolayer presentation and which data processing techniques and training dataset sizes are feasible for accurate predictions?
- **RQ 2**: How accurate is the prediction of bale-specific purities using SBQC methods compared to MQC at monolayer presentation?
- **RQ 3**: What is the variation of plastic pre-concentrate purities in a LWP sorting plant?
- **RQ 4**: How feasible is SBQC at a multilayered material flow presentation?

2. Background

2.1. Quality definitions of plastic pre-concentrates

Pre-concentrate qualities are defined by their purity, i.e., the mass-based share of target material in a pre-concentrate batch (Borowski, 2018; Der Grüne Punkt, 2023; Feil et al., 2021). As pre-concentrates are transported as *bales* between sorting and processing plants, a batch in sorting plants refers to a compressed cuboid plastic bale with median dimensions of approx. 1100 mm x 1100 mm x 1200 mm and a mass of approx. 200 kg to 1000 kg (depending on the pressed material) (Borowski, 2018).

For example in Germany, pre-concentrate purities are often specified based on the Der Grüne Punkt Duales System Deutschland GmbH (DSD) specifications (Der Grüne Punkt, 2023). A transparent polyethylene terephthalate (PET) bottle bale, for example, must contain at least 98 wt % transparent PET bottles according to the DSD specifications (Der Grüne Punkt, 2023).

2.2. Purity definitions

Along the plastic recycling value chain, different stakeholders use different references for defining the term *purity*, which can cause confusion, especially at the interfaces between different stakeholders. Therefore, we propose to differentiate between three different *purity definitions* (Table 1):

- **Article-based purity:** An article is a piece of packaging that may consist of several components and materials. It may also contain residues or contaminants. Article-based purity is defined as 100 % when the target material content exceeds an application-specific threshold. For example, a post-consumer PET bottle is considered 100 % pure, even if it contains, e.g., a bottle cap made of high-density polyethylene (HDPE), a label made of polypropylene (PP), remaining residual contents (e.g., liquids), and adhering dust.
- **Material-based purity:** In contrast to article-based purity, all mechanically removable non-target materials connected to a target article are considered impurities for material-based purity. In the PET bottle example, HDPE bottle caps, PP labels, remaining residual contents, and adhering dust are now considered impurities, since they can be mechanically removed (e.g., through separation steps like washing processes).
- **Chemical purity:** In addition to the material-based purity, all non-target molecules are considered impurities when defining a chemical purity, even if they cannot be (mechanically) removed. This includes, for example, additives, fillers, and reinforcing materials. For example, a washed and cleaned PET flake would not be considered 100 % pure as it contains non-PET molecules such as color pigments or UV stabilizers.

Article-based purity definitions are typically used during waste collection and in sorting plants, while material-based purity definitions are frequently used in plastic processing and chemical purity definitions are used, e.g., in chemical recycling. The stricter the purity definition, the lower is the determined purity $P_{m,\,i}$ (Eq. (1)).

$$P_{m,\,\text{article-based}} \geq P_{m,\,\text{material-based}} \geq P_{m,\,\text{chemical}} \quad (1)$$

Table 1
Proposed definition of article-based, material-based and chemical purities; ✔: included in purity definition, –: not included in purity definition.

Constituents	Examples	Included in Article-based purity[a]	Material-based purity	Chemical purity
Agglomerates	Multiple interlocking packages, e.g., through post-consumer influences	✔[b]	–	–
Secondary components	Labels and sleeves inclusive adhesives, lids	✔	–	–
Residual content	Remaining liquids	✔	–	–
Adhesive contamination	Dust, adhesive organics	✔	–	–
Material-bound additives, fillers, and reinforcing materials	Color pigments, UV stabilizers, antistatic agents, plasticizers, flame retardants, glass fibers	✔	✔	–
Target molecule	Polypropylene, aluminum	✔	✔	✔
Unit		wt% wet	wt% dry	wt% dry

[a] oriented on Der Grüne Punkt (2023)
[b] depending application-specific target-material threshold and definition of the sorting catalogue.

2.3. Potential SBQC sensor positions

As SBQC aims to analyze final pre-concentrate purities, it should take place *after* the last sorting stage and *before* the bale press. For most modern sorting plants, two sensor positions fulfill both criteria (see Fig. 1b):

- **Sensor position 1 (P1)**: product conveyor belts feeding the material into product bunkers,
- **Sensor position 2 (P2)**: feeding conveyor belts towards the bale press.

Sensor position P1 allows SBQC at a *monolayer* material flow presentation, i.e., particles do not overlap each other, while material flows at sensor position P2 are transported as *multilayered* bulks, i.e., particles strongly overlap each other (Kroell et al., 2023a). As NIR spectroscopy is a surface measurement technology (Workman and Weyer, 2007), NIR-based characterization of multilayered bulks (P2) can be challenging, e.g., due to segregation effects (Kroell et al., 2023a). However, sensor position P2 permits the monitoring of multiple pre-concentrate types with a single sensor, while sensor position P1 requires separate sensors for each pre-concentrate type, thus reducing potential investment costs (Kroell et al., 2022a).

3. Material and methods

To compare both sensor positions and answer the RQs from Section 1.4, we conducted two test series A and B in a state-of-the-art LWP sorting plant in Germany.

3.1. Experimental setup

3.1.1. Test series A

Test series A aimed at investigating the technical feasibility of SBQC at plant scale at monolayer material flow presentation (sensor position P1; RQ 1) and then comparing the SBQC results with MQC (RQ 2). As shown in Fig. 1a, a state-of-the-art NIR sensor ("NIR A1") was mounted on the conveyor belt of a PET tray product belt (P1) for inline monitoring of the full material flow (Section 3.2.1 to Section 3.2.3).

To validate the NIR data with MQC, all impurities were manually sorted out by two human experts with several hundred hours of experience in manual sorting analysis of post-consumer plastic packaging. The human experts were positioned directly after the NIR sensor

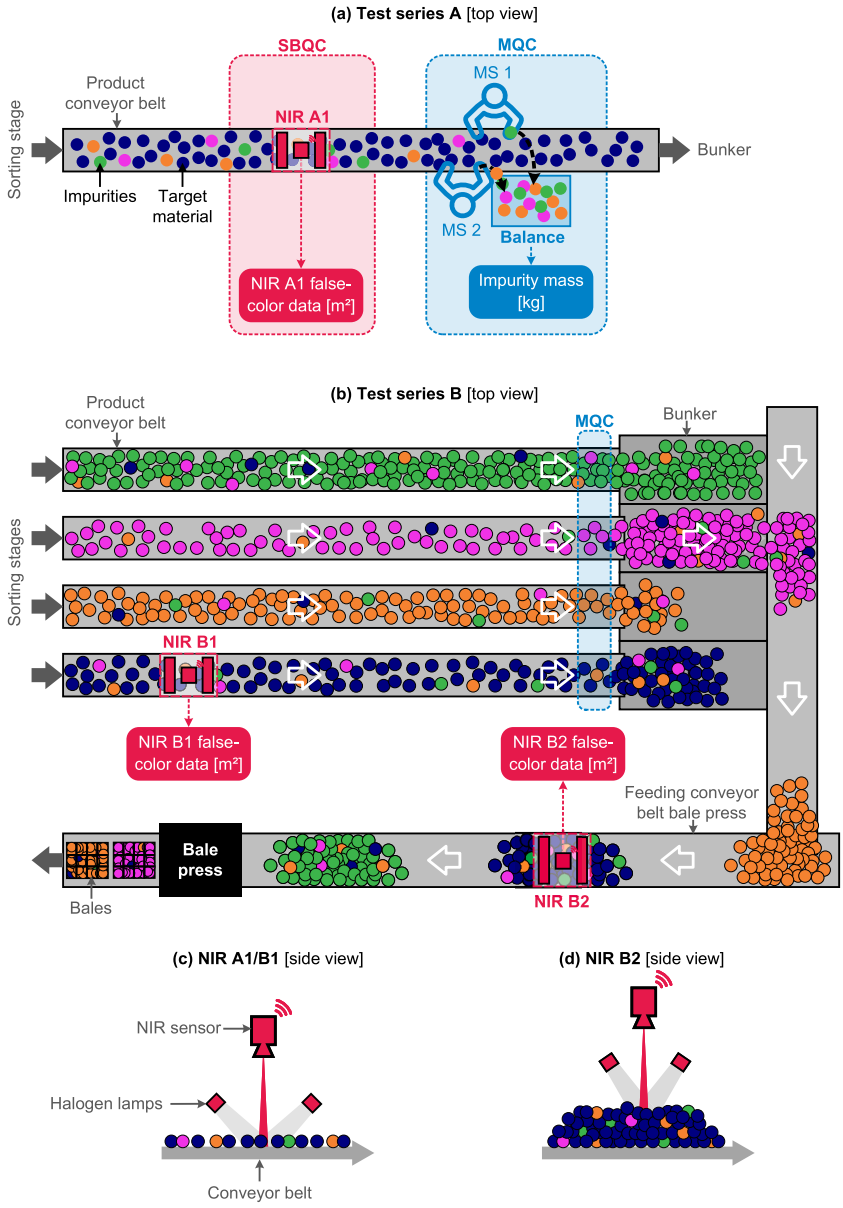


Fig. 1. Experimental setup of (a) test series A and (b) test series B as well as NIR sensor setup and material flow presentation at (c) sensor position P1 [PET tray product conveyor belt, monolayer material flow presentation] and (d) sensor position P2 [feeding conveyor belt bale press, multilayered material flow presentation].

(Fig. 1a) and immediately dropped any impurities in a sample container (V = 90 L), whose weight was recorded every second using a digital balance (Section 3.2.4).

Each trial lasted for 25 s during which the data of NIR A1 and the digital balance was recorded. After each 25 s trial, the sample container was emptied and time was given to the human experts to recover from the intense sorting, such that all sorting trials could be conducted at maximum concentration and with maximum accuracy over the course of three days. In total, n = 417 trials were recorded resulting in 174 minutes of manual sorting data during which 1,562 kg of PET tray product was manually sorted and recorded.

3.1.2. Test series B

Test series B aimed at quantifying purity variations (RQ 3) and assessing the technical feasibility of SBQC at multilayered material flow presentation (sensor position P2; RQ 4). During test series B, the full NIR data of two NIR sensors were recorded for a duration of 30 h. As shown in Fig. 1b, the first NIR sensor ("NIR B1") was positioned at the identical position and with the same configuration as NIR A1 (=sensor position P1), and the second NIR sensor (NIR B2) was positioned over the feeding conveyor belt to a bale press (sensor position P2), thus monitoring the bale-wise composition of PET tray and other plastic pre-concentrates (PET bottle, PP, HDPE, and polystyrene [PS]). In addition, all recorded fractions were sampled, manually analyzed, and immediately returned to the bunker to generate additional MQC data to validate the NIR results without influencing the NIR B2 measurements (Section 3.2.5).

3.2. Data acquisition

3.2.1. NIR sensor configuration

In both test series, hyperspectral imaging NIR sensors EVK HELIOS EQ32 from EVK Kerschhaggl GmbH (Raaba, Austria) were used. The used wavelength range was 1,017 nm to 1,702 nm at a spectral resolution of 3.1 nm per channel. All sensors were operated at a framerate of 446 Hz.

At sensor position P1 (Fig. 1c), the effective belt width was $b \approx$ 520 mm and the conveyor belt speed was $v \approx$ 0.42 m/s, resulting in a

spatial resolution of 3.31 mm^2/px. At sensor position P2 (Fig. 1d), the effective belt width was $b \approx 830$ mm, and the conveyor belt speed was $v \approx 1.2$ m/s, resulting in a spatial resolution of 13.41 mm^2/px.

Halogen lamps with 400 W each were used to illuminate the material flow as it moved below the NIR sensor (cf. Fig. 1c+d). At sensor position P1, two halogen lamps from the front and two halogen lamps from the back were used (4×400 W = 1,600 W in total). At sensor position P2, four halogen lamps from the front and four halogen lamps from the back were used (8×400 W = 3,200W in total). To prevent overheating of the conveyor belt at a belt standstill and recording redundant data, the power supply of the NIR illumination was coupled with the belt control such that the NIR illumination switched off automatically when the conveyor belt stopped.

3.2.2. NIR classification models

The material flow was classified pixel-based in real-time using the on-chip classification algorithm CLASS32 from EVK Kerschhaggl GmbH (Raaba, Austria) based on normalized first-derivatives of the acquired NIR spectra (EVK Kerschhaggl GmbH, 2023a).

Six material classes were trained to the classification model: PET, PP, polyethylene (PE), PS, beverage carton (BC), as well as paper, paperboard, and cardboard (PPC). For each material class, a variety of different packaging articles was sampled from the sorting plant, recorded, and added as reference spectra to the classification model (see Fig. S1 and Fig. S2 in supplementary materials) to ensure a high classification accuracy. For defining the reference spectra, representative regions of the recorded raw spectra were selected, such that, e.g., edge effects are avoided (Chen and Feil, 2019; Küppers et al., 2019).

The reference spectra also include sleeved and labeled parts of plastic bottles, which can have a significant influence on the NIR classification (Chen et al., 2023). These sleeved and labeled parts were added as reference spectra to the material class of the bottle material (e.g., an HDPE bottle with a PET sleeved is trained as "PE"), to assure a classification of the plastic article as fully correct as possible (cf. Chen et al., 2023). Likewise, wet articles were used as training materials to exclude the influence of NIR absorption of water on the classification (cf. Küppers et al., 2019). Lastly, overlays of transparent material on top of other materials (e.g., PET tray on top of an HDPE bottle) are trained as reference spectra, such that always the material on top is classified, to avoid systematic over/underestimation of certain materials (cf. Kroell et al., 2023a).

Background (BG) pixels (black conveyor belt) and material pixels were differentiated using the mean intensity of the raw NIR spectra based on a user-defined threshold. NIR spectra that differ significantly from the reference spectra are classified as "undefined" (UNDEF) by the CLASS32 algorithm, resulting in a total of eight different NIR outputs (PET, PP, PE, PS, BC, PPC, UNDEF, BG). Within CLASS32, three spectral regions of interest for classification were defined (NIR 1: 1082 nm – 1261 nm, 1352 nm – 1514 nm, 1621 nm – 1702 nm; NIR 2: 1107 nm – 1242 nm, 1342 nm – 1505 nm, 1615 nm – 1696 nm), where the NIR spectra of the investigated materials differ most (see Fig. S1 and Fig. S2). As shown in Fig. S3, the classification model classifies all six material classes correctly. Smaller classification errors occur between BC and PPC (especially for plastic coated PPC) and due to labeled or sleeved parts of packaging items.

3.2.3. NIR data acquisition

After uploading the NIR classification model from Section 3.2.2 to the NIR sensor, the CLASS32 algorithm then assigns a material class (represented as a false-color) to each recorded pixel. The resulting false-color data was captured using the EVK Stream Supervisor software (EVK Kerschhaggl GmbH, 2023b). In the EVK Stream Supervisor software, the false-color images are aggregated over 0.1 s and the pixel counts are then saved along with a timestamp in a comma-separated value (CSV) file.

3.2.4. Digital balance

For test series A, the impurity masses were recorded with a digital balance (KERN IFB 60L-3L; KERN & SOHN GmbH [Balingen-Frommern, Germany]) with a precision of ±0.002 kg and a maximum weight of 60 kg. The weight was written automatically every second along with a timestamp to an Excel file through the software SCD-4.0-PRO from KERN & SOHN GmbH.

3.2.5. Manual sorting analysis

For test series B, the plastic pre-concentrate fractions (PET tray, PET bottle, PP, HDPE, and PS) recorded by NIR B2 were additionally sampled, and their composition was analyzed through manual sorting analysis. The sampling was performed according to LAGA PN 98 (2001) using the existing automatic sampling system based on temporarily reversing the respective bunker feeding belt into a sampling container to create ideal sampling conditions with a sample volume of 180 L.

After sampling, the sample was immediately manually sorted into the following material fractions by human experts: PET tray, PET bottle, PP, PE, PS, BC, PPC, non-ferrous metals, ferrous metals, composites, and residual. Each sorted fraction was weighted using a digital balance (KERN DS 150K1, KERN & SOHN GmbH [Balingen-Frommern, Germany]) with a precision of ±0.001 kg and a maximum weight of 150 kg to determine the material composition of each sample. After manual sorting analysis, the full sampled material (incl. impurities) was mixed together, homogenized, and returned to the bunker feeding belt to make sure that the NIR B2 recordings were as little as possible influenced by the sampling[3]. Since the focus is on PET tray, the PET tray fraction and all other fractions were sampled alternatively resulting in the following sampling schedule: PET tray, PET bottle, HDPE, PET tray, PS, PET tray, PP. After each cycle, the sampling schedule was repeated. In total, $n = 46$ samples were taken and manually analyzed resulting in a total analyzed sample volume of 8,280 L and a total sample mass of 244 kg.

3.3. Data processing

For data processing, data visualization, and ML model training and evaluation, custom Python scripts were developed. The following open-source packages were primarily used: *NumPy* (Harris et al., 2020) [data storage and data processing], *pandas* (McKinney, 2010; The pandas development team, 2020) [data storage and data processing], *scikit-learn* (Pedregosa et al., 2011) [training and evaluation of regression models], *SciPy* (Virtanen et al., 2020) [statistics], *matplotlib* (Hunter, 2007) [data visualization], and *seaborn* (Waskom, 2021) [data visualization].

3.3.1. Spatial calibration

First, the pixel counts per 0.1 s from Section 3.2.3 were transformed into area flows [m^2/h] by using the spatial resolution from Section 3.2.1. For determining the spatial resolutions, false-color recordings of circular calibration targets of 170 mm diameter were recorded. The spatial resolution was then calculated by dividing the area of the calibration targets by the number of pixels of the calibration target.

3.3.2. Area-to-mass conversion using material-specific grammages

For the conversion between area-based NIR recordings and mass-based impurity masses, material-specific grammages were determined. Therefore, product material flows were sampled according to LAGA PN 98 (2001) (cf. Section 3.2.5) and the sampled material was manually sorted into pure material classes (PET tray, PET bottle, PP, HDPE, PS, BC, and PPC). Afterward, material feeding on the PET tray belt was

[3] Due to the sample homogenization, multiple belt transfers between the point where the sample were returned and NIR 2 as well as the relative low share (≤ 1.5 wt%) of sample masses (2.4 kg to 9.0 kg, mean: 5.3 kg) compared to the pre-concentrate bale masses (600 kg), the influences on the sampling on NIR 2 are considered as neglectable in the following.

manually stopped and the mono-material fractions were batch-wise recorded using NIR B1. For each batch, the false-color data from NIR B1 and the total mass of the batch were recorded. In total, the following masses per mono material fractions were recorded: PET bottle: 12.2 kg, PET tray: 2.7 kg, PP: 6.3 kg, PE: 7.1 kg, PS: 3.8 kg, BC: 12.2 kg, PPC: 3.6 kg.

Afterward, material-specific grammages were calculated according to Eq. (2) based on the total mass of recorded mono-material fractions m_i and the total area per recorded mono material fraction (incl. non-target pixels) A_i.

$$\rho_{A,i} = \frac{m_i}{A_i} \tag{2}$$

Table 2 summarizes the determined grammages. Since no sample material for the NIR material class UNDEF can be collected, a grammage of $\rho_{A,\ \mathrm{UNDEF}} = 1.63\ \mathrm{kg/m^2}$ was used, which is the mean grammage over all other seven material classes from Table 2.

3.3.3. Area-to-mass conversion using machine learning

Based on the insights from Kroell et al. (2023a), ML models were used as an alternative for predicting the mass-based impurity contents and material flow compositions from area-based NIR data.

For test series A, the total impurity mass flow can be calculated from the impurity mass of one trial measured by the digital balance (m_{impurity}) and the duration of one trial ($\Delta t_{\mathrm{trial}} = 25$ s) using Eq. (3).

$$\dot{m}_{\mathrm{impurity}} = \frac{\Delta m}{\Delta t} = \frac{m_{\mathrm{impurity}}}{\Delta t_{\mathrm{trial}}} \tag{3}$$

Therefore, a multivariate linear regression model (f_{A}, Eq. (4)) was trained to predict the total impurity mass flow $\dot{\hat{m}}_{\mathrm{impurity}}$ [kg/h] from the area flows $\dot{A}_i$ [$\mathrm{m^2}$/h] of each material class detected by the NIR sensor (cf. Section 3.2.2) excluding background[4].

$$\dot{\hat{m}}_{\mathrm{impurity}} = f_{\mathrm{A}}(\dot{A}_{\mathrm{PET}}, \dot{A}_{\mathrm{PP}}, \dot{A}_{\mathrm{PE}}, \dot{A}_{\mathrm{PS}}, \dot{A}_{\mathrm{BC}}, \dot{A}_{\mathrm{PPC}}, \dot{A}_{\mathrm{UNDEF}}) \tag{4}$$

To increase the practicality and robustness of our ML model in lightweight-packaging waste sorting plants, we divided our dataset into 20% ($n = 83$ trials) training data and 80% ($n = 334$ trials) test data. Our goal was to demonstrate that a satisfactory model could be achieved even with a small training set, thereby increasing efficiency for industrial applications by reducing data collection needs. Moreover, a large test dataset was employed for a robust analysis of the ML model performance and the comparison with MQC methods (cf. Section 4.2).

3.3.4. Throughput estimation

Besides impurity masses, it is necessary to know the total product mass flow to calculate the purities of a material flow or pre-concentrate bale (cf. Eq. (6)). For industrial SBQC applications, information on product masses is often available from the bunker system in terms of the total mass of each produced bale. However, for calculations in Section 3.4, it is necessary to estimate the PET tray product mass flow at a higher temporal resolution (25 s).

Previous research (e.g., Curtis et al., 2021; Kroell et al., 2022b; Küppers et al., 2020; Küppers et al., 2022) has established a strong correlation between mass flows and area flows. To enhance throughput estimation, we therefore calculated the throughput $\dot{m}_{\mathrm{PET\ tray\ product}}$ using the total area flow measured by NIR A1 ($\dot{A}_{\mathrm{PET\ tray\ product}}$) as follows:

$$\dot{m}_{\mathrm{PET\ tray\ product}} = \dot{A}_{\mathrm{PET\ tray\ product}} * \rho_{A,\ \mathrm{PET\ tray\ product}} \tag{5}$$

To determine the mean grammage $\rho_{A,\ \mathrm{PET\ tray\ product}}$, we first measured the total occupied area (including impurities) of the 39.7 Mg PET tray product using NIR A1. Then, we employed Eq. (2) to calculate the mean grammage as $\rho_{A,\mathrm{PET\ tray\ product}} = 1.79\ \mathrm{kg/m^2}$.

3.3.5. Outlier removal

During preprocessing of test series A, we removed two non-representative outliers (trials no. 127 and 128; 0.5% of the total dataset) with impurity mass flows of 395.7 kg/h and 332.4 kg/h, as both outliers deviated significantly both from the median (25.3 kg/h) and the next largest impurity mass flow (117.5 kg/h). The outliers resulted from two water-soaked shoes with an approximate mass of approx. 2 kg each in the PET tray material flow and not removing them would make ML model training and evaluation practically infeasible. While, in the current study, the two outliers were omitted to demonstrate the general technical feasibility of the SBQC approach, given their low frequency and significant deviation from the other datapoints, a detailed consideration of such outliers regarding the application-specific relevance is recommended when applying our findings into operational practice and the evaluation of future SBQC measurement systems (cf. Section 5).

3.4. Comparison of SBQC and MQC (bale sampling simulation)

To answer RQ 2, SBQC and MQC methods were compared with each other on the task of predicting the purity of a PET tray bale with an average mass of 600 kg. As it is unfeasible to manually sort hundreds of PET tray bales to determine the necessary ground truth, different bale compositions were digitally generated from test series A.

Therefore, the data from test series A was first randomly split into 20% training and 80% test data (cf. Fig. 2a). The training set was used to train ML model A (Section 3.3.3) for SBQC and the test set was used as the ground truth for creating simulated PET tray bales with known compositions.

3.4.1. Simulation of different PET tray bales

The test set consists of $n = 334$ trials containing in total 1242.4 kg manually sorted material (approx. 2.1 bales) each with the corresponding impurity (Section 3.1.1) and total masses (Section 3.3.4). To create a simulated PET tray bale, we randomly sampled n trials from the test set without replacement until a total sample mass of 600 kg was reached. These n trials then represent one simulated PET tray bale with its known mass m_{bale} and known purity $P_{m,\mathrm{bale}}$ according to Eq. (6).

$$P_{m,\ \mathrm{bale}} = \frac{m_{\mathrm{impurity}}}{m_{\mathrm{bale}}} = \frac{\sum_{i=1}^{n} m_{\mathrm{impurity},\ i}}{\sum_{i=1}^{n} m_i} \tag{6}$$

3.4.2. Simulation of MQC sampling

State-of-the-art MQC of pre-concentrates in LWP sorting plants involves sampling. The required sample sizes for a pre-concentrate bale can be calculated based on different technical norms and guidelines (see Section 1 in supplementary materials). In the case of a 600 kg PET tray bale, applying these norms and guidelines results in required sample sizes of 0.9 kg (CEN/TR 15310-1, 2006), 3.2 kg (LAGA PN 98, 2001), 40.0 kg (GBP Quality GmbH, 2023), 50.0 kg (COREPLA, 2022), 80.0 kg–90.0 kg (Der Grüne Punkt, 2016), and 123.0 kg (ÖNORM S

Table 2
Determined material-specific grammages.

Material class	Grammage [kg/m²]
PET bottle	2.17
PET tray	1.22
PP	1.79
PE	2.64
PS	1.34
BC	1.97
PPC	0.79

[4] Since the total area flow (materials + background) is constant, the background area flow is linear dependent on the sum of all material area flows and thus excluded from model training, cf. Fahrmeir et al. (2013).

2127, 2011).

To simulate the influence of different sample sizes, we randomly sampled k of these n trials, which then represents a sample with the total sample mass $m_{\text{samples}}(k)$. As in traditional MQC, the purity of the total bale is estimated by the purity of the sample according to Eq. (7).

$$\widehat{P}_{m,\,\text{bale},\,\text{MQC}}(k) = \frac{m_{\text{impurity}}(k)}{m_{\text{samples}}(k)} = \frac{\sum_{i=1}^{k} m_{\text{impurity},i}}{\sum_{i=1}^{k} m_i} \text{ with } k \leq n \tag{7}$$

3.4.3. Simulation of SBQC inline monitoring

In contrast to MQC, SBQC methods analyze the full material flow, therefore no sampling is involved. Thus, ML model A was used to predict the impurity mass for each of the n trials ($\widehat{m}_{\text{impurity},\,\text{SBQC},\,i}$). The predicted purity of the bale ($\widehat{P}_{m,\,\text{bale},\,\text{SBQC}}$) is then calculated by dividing the summed-up impurity mass ($\widehat{m}_{\text{impurity},\,\text{SBQC}}$) through the total bale mass (Eq. (8)).

$$\widehat{P}_{m,\,\text{bale},\,\text{SBQC}} = \frac{\widehat{m}_{\text{impurity},\,\text{SBQC}}}{m_{\text{bale}}} = \frac{\sum_{i=1}^{n} \widehat{m}_{\text{impurity},\,\text{SBQC},\,i}}{\sum_{i=1}^{n} m_i} \tag{8}$$

This process was repeated for a total of 1,000 different PET tray bales. This was the basis for the assessment of the measurement uncertainty of MQC methods with different sample sizes and the proposed SBQC methods.

3.5. Assessment of measurement uncertainty

To assess the predicted (im)purities for answering RQ 1 and RQ 2, we use the 95 %-percentile measurement uncertainty (MU) metric according to Kroell et al. (2023a). The MU_{95} is the 95th percentile (P_{95}) of all absolute errors between a set of measurands X_{measured} and its corresponding true values X_{true} (Eq. (9))

$$MU_{95} = P_{95}(|X_{\text{measured}} - X_{\text{true}}|)$$

$$\text{with } X = \{x_1, \ldots, x_n\} \tag{9}$$

The MU_{95} indicates that in 95 % of all cases, the true value (x_{true}) is in the range $x_{\text{measured}} \pm MU_{95}$ (Kroell et al., 2023a). For example, if an SBQC system is predicting a purity of 96 wt%, with a MU_{95} of 2 wt%, then the true purity is between 94 wt% and 98 wt% in 95 of 100 measurements. The lower the MU_{95} values, the more accurate are the predictions.

As the MU_{95} indicates the deviations in absolute terms, we use the normalized MU (nMU_{95}) defined in Eq. (10) to express relative deviations. The nMU_{95} is the MU_{95} devided by the mean measurement values ($\bar{x}_{\text{measured}}$).

$$nMU_{95} = \frac{MU_{95}(X_{\text{measured}})}{\bar{x}_{\text{measured}}} \tag{10}$$

4. Results and discussion

4.1. Impurity mass flow prediction at monolayer presentation (RQ 1)

Fig. 2a shows the impurity mass flows of test series A determined by manual sorting of human experts over the course of $n = 417$ trials and the applied random train-test-split. High variations in the measured impurity flows between 0 kg/h and 117.5 kg/h (mean: 28.3 kg/h) can be observed.

4.1.1. Comparison of material-specific grammage and ML model

Fig. 2b compares the grammage (Section 3.3.2) and ML model (Section 3.3.3) in predicting the impurity mass flow based on the NIR A1 data. As shown in Fig. 2b, the grammage model overestimates the impurity mass flow on average by a factor of 2.11, resulting in an nMU_{95} of 218.0%. In contrast, the predictions of ML model A are more accurate with an nMU_{95} of 94.6% on the raw data.

A likely reason for the overestimation of the grammage model is the discrepancy between article-based and material-based purity definitions (cf. Section 2.2, esp. Eq. (1)). In an article-based purity definition, used by human experts, non-PET materials are counted as impurities only if they are not physically connected to a PET article. The NIR sensor, however, classifies each pixel independently of its connection to other materials. Since each non-PET pixel directly contributes to the predicted impurity for the grammage model, non-PET pixels from PET articles (e.g., PP films, PPC labels belonging to PET trays) are always counted as impurities.

In contrast, the ML model is trained on the article-based impurity definition by the human experts in the manual sorting (training) data. It can thus correct these composite effects by adjusting its model coefficients. However, as shown in Fig. 2c, especially high impurity mass flows are still difficult to be predicted for the model. A possible reason for this is a large variance of grammages in post-consumer packaging waste (cf. Kroell et al., 2021), which makes accurate predictions on the unaggregated raw data (here: 25 s chunks) challenging (cf. Kroell et al., 2023a).

4.1.2. Influence of data aggregation (chunk size)

Since the investigated sorting plant produces a PET tray bale approximately every $\Delta t_{\text{bale}} = 67$ min, 25 s values of SBQC are not necessary for this use case. Aggregation of sensor-based material flow data has been demonstrated to improve prediction accuracy (Kroell et al., 2023a), as, e.g., particle-specific deviations in grammages can be smoothed out. Therefore, Fig. 2d shows the influence of different chunk sizes (time windows over which the data was aggregated [Kroell et al., 2023a]) on the model performance.

As shown in Fig. 2d, prediction errors decline super-linear with increasing chunk size. For chunk sizes above 30 minutes, nMUs of $\leq$ 6.9 % are achieved. As illustrated for exemplary chunk sizes in Fig. 2d, even small data aggregation can already effectively reduce prediction errors.

4.1.3. Influences of training size

Another way to improve model performance is to increase the amount of provided training data (Goodfellow et al., 2016). As the original training set was relatively small ($n = 83$ trials; 20% of the total dataset), the question arises whether the performance of the model could be improved with more training data. On the other hand, as the generation of training data is costly (e.g., high personal expenses for manual sorting), it is also of high practical interest whether a similar level of performance could be achieved with even less training data.

To address both questions, Fig. 2d shows the model performance for different training sizes across different chunk sizes. It becomes evident that in this case, the model performance does not substantially increase with increasing training size: While the mean nMU_{95} over all investigated chunk sizes (25 s – 80 min) at 20% training size ($n = 83$ trials) is 11.1%, it only decreases to 10.8%, when doubling the training size to 40% ($n = 166$ trials). Furthermore, with only 10% training data ($n = 41$ trials), a comparable model performance with a mean nMU_{95} of 11.8% over all chunk sizes was still achieved.

4.1.4. Model robustness

Fig. 2e shows the coefficients of the trained models for different training sizes to further validate the model robustness and study the influence of different training sizes on the model. As shown, the model demonstrates increased stability with increasing training size, showing stable predictions for training sizes greater than 10%.

The model predictions appear plausible, since the coefficient for the material class PET (= target material class in the investigated PET tray fraction) is nearly zero across all trained models, thus not influencing the impurity content. Negative model coefficients are observed for the material classes PE, PP, and UNDEF. The negative PE and PP coefficients

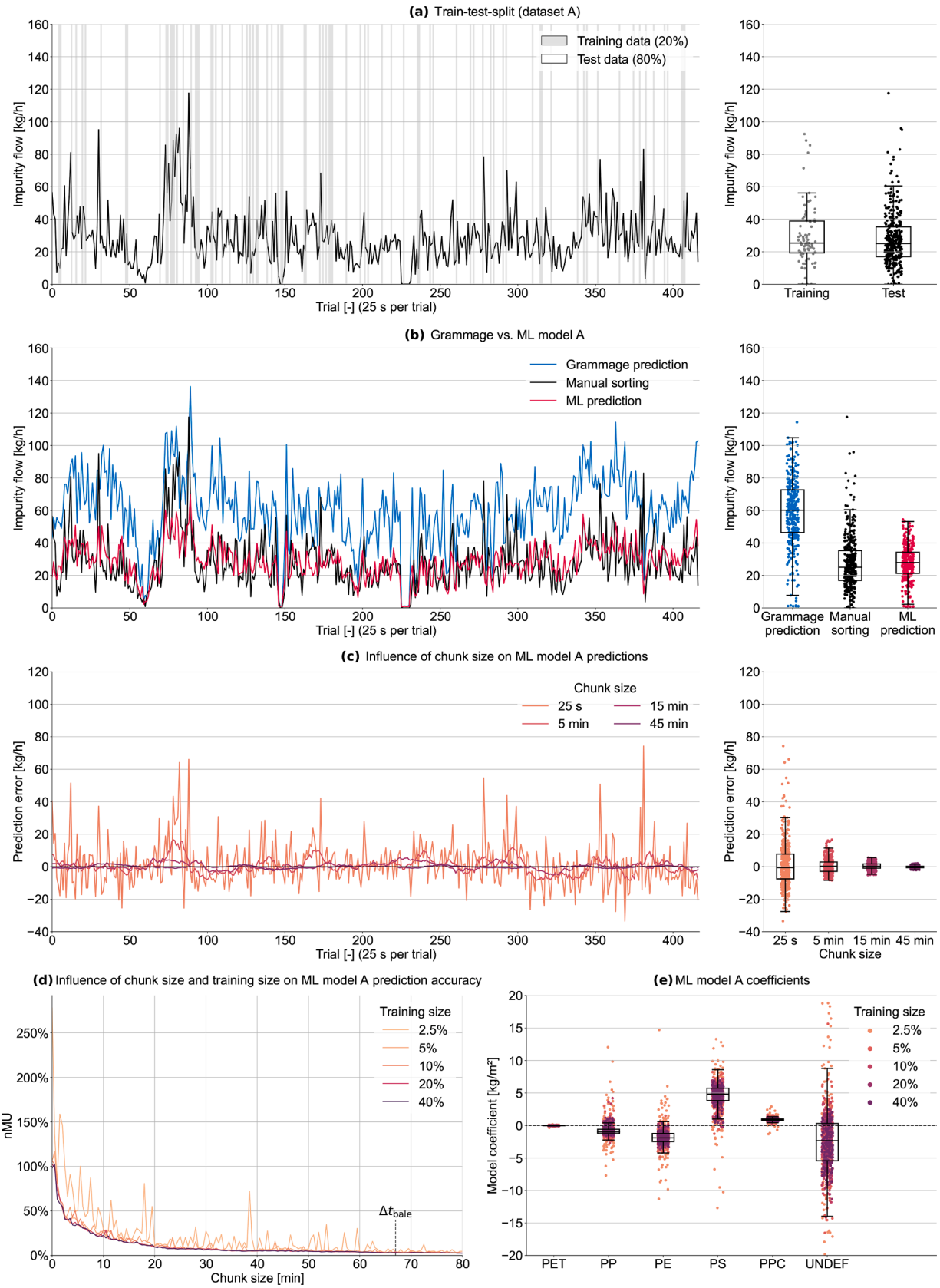


Fig. 2. NIR-based impurity mass flow prediction at monolayer presentation [NIR A1]. (a) Overview of measured impurities determined by human experts and applied train-test-split, (b) impurity prediction using grammages and ML compared to ground truth, (c) influence of data aggregation on prediction errors of ML model A, (d) influence of training size on the prediction performance of ML model A, (e) distribution of model coefficients of ML model A trained across different training sizes and chunk sizes from [d]. Δt_{bale}: average production time of a PET tray bale in test series A.

are likely the model's correction for composite effects (discrepancy between article- and material-based purity, cf. Section 2.2) as films of PET tray packaging and bottle caps belonging to PET bottles are frequently made of PE and PP (Roosen et al., 2020). The negative value and high variation of the UNDEF coefficient could indicate an attempt of the model to correct misclassifications as well as unknown material classes within the UNDEF material classes from the NIR A1 sensor. In contrast, positive model coefficients are especially found for PS and PPC, which supports the hypothesis that the model might have chosen PS and PPC as leading indicators for high impurity contents.

In summary, ML model A demonstrates superior performance in predicting impurity mass flows compared to the grammage model by effectively taking discrepancies between article-based and material-based purity definitions into account through material-specific weight adjustments. Data aggregation has a significant impact on model performance, with prediction errors declining super-linearly as chunk size increases. Notably, even a small amount of training data (10% of the full dataset) can yield comparable model performance. The final model, trained on only $n = 83$ trials (20% training size) combined with a 67-minute aggregation (corresponding to the average production time of one PET tray bale), can predict impurity mass flows with a nMU_{95} of 3.2%.

4.2. Comparison between SBQC and MQC at monolayer presentation (RQ 2)

Fig. 3 compares the accuracy (MU [wt%]) between SBQC based on NIR A1 and MQC across different sample sizes for 1,000 simulated PET tray bales (cf. Section 3.4). As shown by the MQC line in Fig. 3, the measurement uncertainty of MQC caused by the sampling error decreases quadratically with increasing sample sizes, and a "diminishing return" effect is observed, i.e., with increasing sample sizes the additional measurement accuracy gained by larger sample sizes flattens out. High MUs (> 5 wt%) are observed when simulating MQC according to CEN/TR 15310-1 (2006) [$MU_{95} = 6.7$ wt%] and LAGA PN 98 (2001) [$MU_{95} = 5.6$ wt%]. In contrast, higher sample sizes specified in technical MQC guidelines for pre-concentrates as well as the ÖNORM S 2127 (2011) lead to reasonable MUs (≤ 1.5 wt%) when simulating MQC according to GBP Quality GmbH (2023) [$MU_{95} = 1.5$ wt%], COREPLA (2022) [$MU_{95} = 1.3$ wt%], Der Grüne Punkt (2016) [$MU_{95} = 1.0$ wt%], and ÖNORM S 2127 (2011) [$MU_{95} = 0.8$ wt%].

Since SBQC methods require no sampling, the proposed method using ML model A achieves a constant MU_{95} of 0.31 wt%, which is significantly lower than MUs of MQC with sample sizes specified by technical norms and guidelines (Fig. 3). In fact, between 351 kg (58.5 wt%) and 386 kg (64.3 wt%) of the simulated PET tray bales must have been sampled and manually sorted to achieve equal or lower MUs by traditional MQC methods. Using the estimate of 4 to 6 person-hours to analyze a sample of 80 kg to 90 kg from Borowski (2018) (cf. Section 1.1), this would amount to between 15.6 to 29.0 person-hours per bale to achieve the same accuracy using MQC than with the proposed SBQC method.

4.3. Variation in predicted pre-concentrate purities (RQ 3)

To demonstrate the potential benefits of SBQC and to estimate purity fluctuations (RQ 3), Fig. 4a shows the area-based purity distribution of the $n = 17$ PET tray bales that were produced during the sensor measurements of test series B. Among these analyzed bales, the area-based purities ranged between 68.3 area percent (a%) and 77.6 a% and with a standard deviation of 3.1 a%.

Notably, the area-based purity of bales produced during the early and late shifts (mean purity: 75.8 a%) was statistically significantly higher ($p \leq 0.01$[5]) compared to night shifts (70.3 a%). A likely reason for this are the different input materials, which are processed during the early and late shifts (input material 1) and night shifts (input material 2) in the investigated sorting plant.

According to the SBQC data, the two input materials affect different impurities to different degrees (Fig. 4b). When input material 2 is processed (night shift), the PET tray product contains statistically significant ($p \leq 0.01$) more PPC (+6.5 a%) and UNDEF (+0.6 a%) material and significantly less target material and other plastics (PET: -5.5 a%, PP: -0.8 a%, PE: -0.9 a%, PS: -0.1 a%). A likely reason for this is the different characteristics of the two input materials (e.g., material flow composition, particle size distribution, and water content), which influence the sorting plant operation (e.g., Kroell et al., 2022b; Küppers et al., 2020; Küppers et al., 2021; Küppers et al., 2022).

4.4. SBQC at multilayered material flow presentation (RQ 4)

Fig. 5 shows the feasibility of SBQC at multilayered material flow presentations ($\alpha_{\text{NIR B2}}$) by comparing the material flow composition of plastic pre-concentrate bales measured with MQC ($\hat{w}_{\text{MQC}}$; Section 4.4.1) and SBQC at monolayer presentation ($\alpha_{\text{NIR B1}}$; Section 4.4.2). In addition to individual values (unfilled markers), mean values (filled markers) and their distance from the ideal correlation line are displayed.

4.4.1. NIR B2 vs. MQC

In test series B, $n = 103$ pre-concentrate bales were analyzed in total by NIR B2. $n = 30$ of these bales ($n_{\text{PET tray}} = 13$, $n_{\text{PET bottle}} = 4$, $n_{\text{PP}} = 5$, $n_{\text{HDPE}} = 4$, $n_{\text{PS}} = 4$) were also sampled and their composition was manually analyzed (cf. Section 3.2.5).

In Fig. 5, all bales other than the PET tray bale marked with "**", which is determined by NIR B1, are determined by MQC. These Non-"X"-markers for PET tray, PET bottle, PP, HDPE and PS compare the bale compositions determined by MQC (mass-based and article-based purity definition; cf. Section 2.2) with bale compositions determined by NIR B2 (area-based and material-based purity definition; cf. Section 2.2). Different colors indicate material fractions within a bale. For example, a red circle represents the mass-based PE share (red) within a PP bale (circle).

Even though the estimated material shares of NIR B2 and MQC are determined based on a different counting basis (mass-based vs. area-based) and based on different purity definitions (article- and material-based), a general correspondence between NIR B2 and MQC purities can be observed across all bale types ($MU_{95,\text{ purity, PET tray}} = 10.3\%$, $MU_{95,\text{ purity, PET bottle}} = 5.0\%$, $MU_{95,\text{ purity, PP}} = 5.7\%$, $MU_{95,\text{ purity, HDPE}} = 3.4\%$, $MU_{95,\text{ purity,PS}} = 8.6\%$).

Likewise, there is a reasonable correspondence between MQC- and NIR-based impurity shares. Here, MU_{95} values are slightly lower since the impurities have a much lower share compared to the purities ($\overline{MU}_{95,\text{ impurities, PET tray}} = 2.4\%$, $\overline{MU}_{95,\text{ impurities, PET bottle}} = 0.9\%$, $\overline{MU}_{95,\text{ impurities, PP}} = 1.3\%$, $\overline{MU}_{95,\text{ impurities, HDPE}} = 1.3\%$, $\overline{MU}_{95,\text{ impurities, PS}} = 1.7\%$). However, in relative terms, the consistency of purities ($\overline{nMU}_{95,\text{ purity}} = 7.2\%$) is considerably more accurate compared to impurities ($\overline{nMU}_{95,\text{ impurity}} = 202.1\%$). A possible explanation may be that the influence of individual particles on the total material content is higher for impurities due to the lower particle number of impurities compared to target materials.

While these numbers indicate the general feasibility of SBQC at multilayered material flow presentations, the following limitations of this comparison have to be considered:

[5] p-values express the level of significance. Here, we calculated the p-value based on the Mann–Whitney U test, Mann and Whitney (1947).

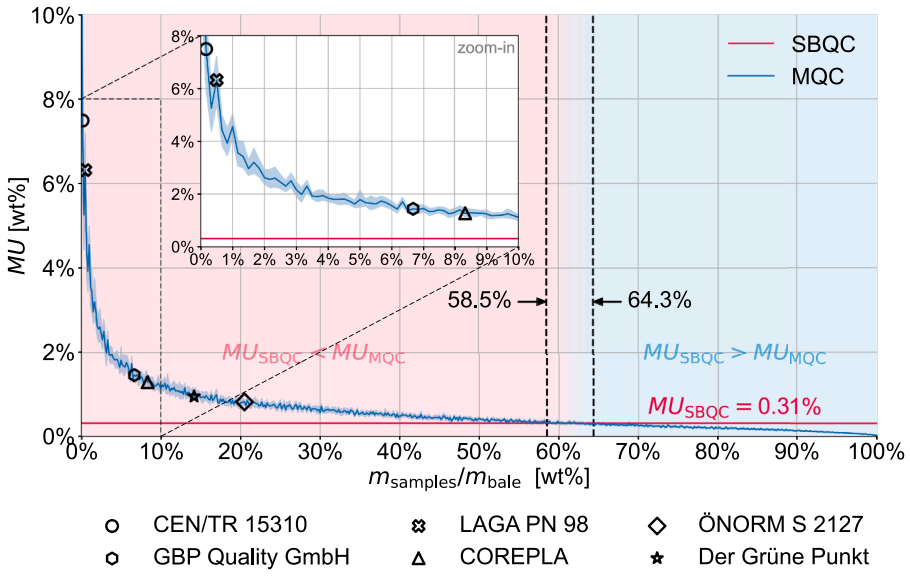


Fig. 3. Comparison of inline SBQC using NIR and ML with MQC using different sampling sizes at monolayer presentation [NIR A1]; markers: minimum sample sizes according to CEN/TR 15310-1 (2006), LAGA PN 98 (2001), ÖNORM S 2127 (2011), GBP Quality GmbH (2023), COREPLA (2022), and Der Grüne Punkt (2016).

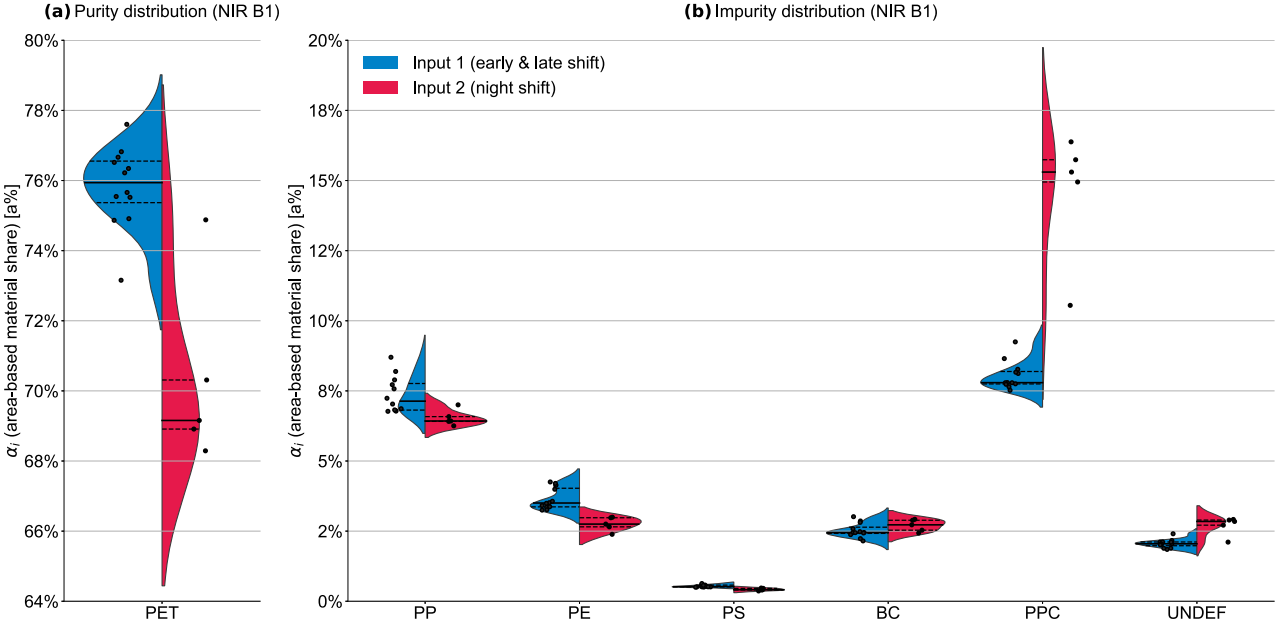


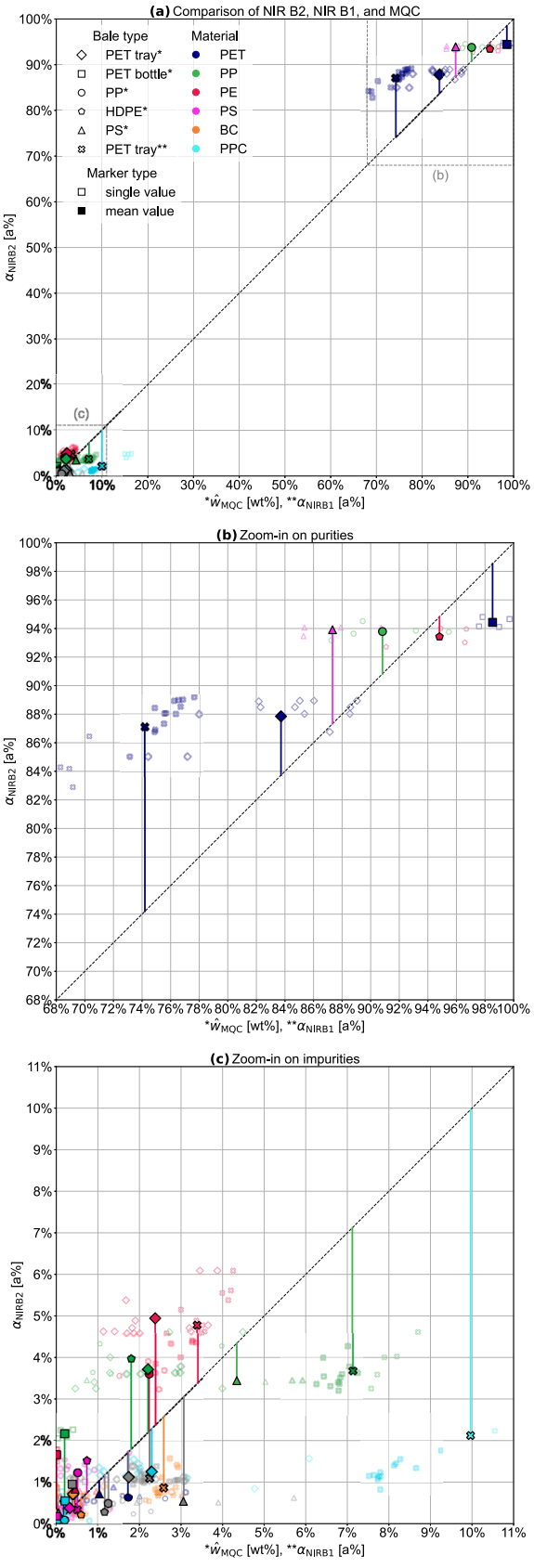
Fig. 4. Area-based (a) purity and (b) impurity distribution per bale between plant input 1 (early and late shift, left violin) and plant input 2 (night shift, right violin) [NIR B1]; solid line: median, dashed line: 25 % and 75 % percentile.

- First, MQC with the initially chosen sample sizes in this study (180 L) is associated with a considerable **sampling error** (see Section 4.2), which can, for example, be observed by high variation of the MQC analysis results (0.97 wt% standard deviations [per material class and bale type, macro-average]) compared to the lower variation in the NIR B2 compositions (0.19 a% standard deviation).
- Second, as discussed earlier, NIR B2 and MQC data have different **counting basis** (mass-based vs. area-based) and purity definitions (cf. Section 2.2). Applying appropriate area-to-mass-prediction models and pixel-to-article-based conversion models might thus reduce the differences between both data sources (Kroell et al., 2021; Kroell et al., 2023a).
- Third, independent of the previous effects, SBQC at a **multilayered material flow presentation** can differ from SBQC at a monolayered material flow presentation and MQC, if the analyzed material flow surface by NIR B2 differs from the area-based composition of the full bale (see Section 4.4.2).

To investigate the influences of multilayered material flow presentation independent from the previous effects, Section 4.4.2 compares the area-based bale compositions determined by NIR B1 and NIR B2 for the investigated PET tray bales.

4.4.2. NIR B2 vs. NIR B1

"X"-markers in Fig. 5 compare the area-based bale compositions determined by NIR B2 and NIR B1 for PET tray bales. Due to the absence of a sampling error, less deviation occurs in NIR B1 measurements compared to MQC, such that potential systematic effects are easier to be



(caption on next column)

Fig. 5. Feasibility of SBQC at multilayered material flow presentation. (a) Differences in area-based material flow compositions between NIR B1 (monolayered material flow presentation, sensor position P1) and NIR B2 (multilayered material flow presentation, sensor position P2) for PET-tray bales and between area-based material flow composition using NIR B2 and mass-based MQC (sample size = 180 L) for plastic pre-concentrate bales, (b) zoom-in of [a] on purities, (c) zoom-in of [a] on impurities; * mass-based and article-based purity determined by MQC ($\widehat{W}_{MQC}$), ** area-based and material-based purity determined by NIR B1 ($\alpha_{NIR\ B1}$).

identified.

When analyzing the PET tray purities (Fig. 5b), a high correlation between NIR B2 and NIR B1 can be observed with a Pearson correlation coefficient (PCC) of 0.915 (Pearson, 1895). A systematic overestimation of NIR B2 is observed with area-based PET shares on average +12.9 a% higher when measured with NIR B2 compared with NIR B1.

Likewise, PET tray impurity shares (Fig. 5c) correlate to a higher extent between NIR B2 and NIR B1 for the material classes PPC (PCC = 0.989), PS (PCC = 0.873), PE (PCC = 0.831), to a lower extent for PP (PCC = 0.601) and BC (PCC = 0.529), and no correlation is found for UNDEF materials. A closer look at Fig. 5c reveals that PE shares were always overestimated by NIR B2 compared to NIR B1 by +1.4 a% on average, while other material classes were mostly underestimated by NIR B2 to different extents (PPC: -7.8 a%, PP: -3.4 a%, BC: -1.7 a%, PS: -0.2 a%, UNDEF: -1.1 a%). These observations are consistent with the findings described in (Kroell et al., 2023a) and might have been caused by the following two mechanisms:

4.4.2.1. Different classification behavior at multilayered material flow presentation. At NIR B1, NIR spectra of particles are measured as a monolayer on top of a carbon-black conveyor belt. For transparent material, like PET trays, very thin-walled parts of packaging articles can result in a low spectral intensity due to the carbon-black surface underneath, which can cause the classification of some article parts as background by the NIR classification algorithms. However, at NIR B2, the material flow surface is measured on top of several underlying material layers. Therefore, even thin-walled parts of transparent articles are laying on top of several other particles, which results in higher spectral intensity and the classification of all article parts as the corresponding material instead of background (mixed NIR spectra are classified as the material on top with the used algorithm, cf. Section 3.2.2). As a consequence, the recognized area of transparent and thin-walled articles is higher at NIR B2 compared to NIR B1, resulting in a higher material share. As a lot of transparent and thin-walled packaging items (e.g., trays, foils, multilayered packaging) are made of PET and PE, this mechanism might partially explain why the material classes PET and PE were overestimated by NIR B2.

4.4.2.2. Segregation effects. Based on our previous experience, the results of Kroell et al. (2023a), and the analysis of the NIR false-color images, we assume that the +12.9% PET overestimation is too high to be caused solely by differences in the NIR classification behavior. The PET overestimation could have been caused by segregation effects, e.g., by the accumulation of two-dimensional PET trays on the material flow surface measured by NIR B2. In addition, variations in particle characteristics such as particle size and shape resulting from different materials used in different packaging applications could further increase segregation effects. Further research and a more robust understanding of segregation mechanisms for post-consumer packaging are required to verify or falsify this hypothesis.

In summary, NIR-based SBQC at multilayered material flow presentation (bale press feeding belt, P2) seems feasible in general with reasonable correspondence between NIR B2 and MQC as well as NIR B1 measurement. However, a deeper understanding of segregation effects and a development of correction as well as area-to-mass models is

necessary to achieve SBQC predictions with accuracies of practical relevance at multilayered material flow presentations (sensor position P2) in the future.

5. Conclusions

The plastic recycling value chain is currently suffering from a high degree of intransparency since MQC procedures for post-consumer plastic pre-concentrates are time-consuming, cost-intensive, and thus only performed on an irregular basis. Inline sensor technology promises to overcome these limitations by enabling an automated monitoring of produced pre-concentrate qualities. While numerous studies showed the high accuracy of plastic discrimination using NIR spectroscopy at pixel- and particle levels, it remained unclear if NIR-based quality control of plastic pre-concentrates at the material flow level and plant scale is technically feasible. Further, it remained unclear how the achieved accuracy of an SBQC approach compares to traditional MQC in the form of manual sorting analysis of taken samples.

In two test series, we showed that SBQC at plant scale is not only technically feasible, but ML-based SBQC predictions can be even more accurate than traditional sampling-based methods. Since the purities of plastic pre-concentrates are determined by a mass- and article-based purity definition in current technical norms and industrial practice (cf. Section 2.2), comparable purities from the pixel-based NIR-false-color-data should be predicted using, e.g., ML methods, to account for composite effects in post-consumer packaging (difference between article- and material-based purity definition, cf. Section 2.2). The high data aggregation due to an average bale production time of several minutes up to an hour (here: 67 minutes) evens out prediction errors at lower time scales and achieves purity predictions with high accuracies (±3.2% normalized measurement uncertainty) [RQ 1].

A direct comparison of SBQC and MQC to predict the purity of 600 kg PET tray bales demonstrates the advantages of the SBQC method: When following technical norms and guidelines for MQC, bale purities could be determined with measurement uncertainties (95^{th} percentile of absolute errors) between ±6.7 wt% (0.9 kg sample size [CEN/TR 15310-1, 2006]) to ±0.8 wt% (123 kg sample size [ÖNORM S 2127, 2011]). The proposed SBQC method predicted bale purities within ±0.31 wt% measurement uncertainty. In fact, we estimate that the required total sample mass for MQC must have been ≥ 350 kg, i.e., more than half of the total PET tray bale and thus way beyond any reasonable MQC practice, to be more accurate than SBQC [RQ 2].

Another benefit of SBQC data could be demonstrated by identifying significant variations in the area-based material flow composition of the PET tray pre-concentrate caused by two different input materials of the sorting plant coming from two different collection areas. To obtain comparable knowledge using MQC, a high personnel effort in both day and night shifts would be necessary. Since investigation of this type are costly, many relevant parameters in sorting plants have remained undetected in the past [RQ 3].

To potentially monitor several pre-concentrates simultaneously with a single NIR sensor in the future, the technical feasibility of SBQC with multi-layered material flow presentations (feeding conveyor belt to bale press) was investigated. A good correlation was found between area-based purities measured at mono- and multi-layered material flow presentation with a high Pearson correlation coefficient of 0.915. Systematic deviations between monolayered and multilayered material flow presentations due to segregation effects and changes in the NIR classification behavior reported by previous studies (Kroell et al., 2023a) were confirmed, resulting in the need for a better understanding of segregation effects and the development of possible correction models [RQ 4].

Several directions for future research emerge from our work: Our investigations could be extended to other pre-concentrates or types of sorting and processing plants. In addition, the NIR sensor could be complemented by additional sensors such as RGB cameras and/or with deep learning techniques, e.g., to detect contaminants of the same material such as HDPE silicon cartridges in a HDPE pre-concentrate or to enable a quality assessment for color-sorted fractions. Further, long-term sensor measurements are needed to analyze the stability of the developed ML models under seasonal fluctuations, changing packaging designs, different waste collection areas/schemes, frequency of outlier impurities, or differing plant operations and to develop and test potential recalibration strategies (e.g., DIN 54390, 2022; Flamme et al., 2020). The economics of SBQC could be enhanced by either utilizing more cost-effective sensor technologies (e.g., RGB cameras) or enabling the previously mentioned SBQC at multilayered material flow presentations to monitor multiple pre-concentrates with a single sensor. To reduce the costs for model training and adaptation at multiple measuring points or in multiple sorting plants, the development of time-/cost-efficient methods for generating the necessary ground truth data as well as the application of transfer and feathered learning approaches could be of further interest.

Ultimately, our research underlines the promising opportunities of sensor technology in post-consumer plastic recycling beyond today's sorting applications. By fostering transparency, incentives for improved sorting could be established and realized through adaptive process control in sorting and processing plants, contributing to a boost in recyclate quality and yields and a sustainable circular economy.

Funding

This work was funded by the National Austrian Research Promotion Agency (FFG) within the program "Production of the Future" under the project *EsKorte* (grant no. 877341) and the German Federal Ministry of Education and Research (BMBF) within the program "Resource-efficient circular economy - plastic recycling technologies (KuRT)" under the project *ReVise* (grant no. 033R341). The responsibility for the content of this publication lies with the authors.

CRediT authorship contribution statement

Nils Kroell: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Xiaozheng Chen:** Conceptualization, Methodology, Validation, Writing – review & editing, Funding acquisition. **Bastian Küppers:** Conceptualization, Methodology, Validation, Investigation, Supervision, Writing – review & editing, Funding acquisition. **Sabine Schlögl:** Validation, Writing – review & editing, Funding acquisition. **Alexander Feil:** Writing – review & editing, Funding acquisition. **Kathrin Greiff:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

We would like to thank the following individuals for their much-appreciated support in our research: **Peter Bardenheuer**: support in literature review (Section 1, Section 2), **Abtin Maghmoumi**: data preprocessing (test series B), **Alena Spies**: sensor setup & manual sorting (test series A), **Johanna Beaupoil**: sensor setup & manual sorting (test series A), **Julius Lorenzo**: sensor setup & manual sorting (test series B),

Matthias Schlaak: sensor setup & manual sorting (test series B), and **Elias Pfund**: sensor setup (test series B).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.resconrec.2023.107256.

References

Bachmann, M, Zibunas, C, Hartmann, J, Tulus, V, Suh, S, Guillén-Gosálbez, G, et al., 2023. Towards circular plastics within planetary boundaries. Nat. Sustain. https://doi.org/10.1038/s41893-022-01054-9.

Bonifazi, G, Capobianco, G, Serranti, S., 2018. A hierarchical classification approach for recognition of low-density (LDPE) and high-density polyethylene (HDPE) in mixed plastic waste based on short-wave infrared (SWIR) hyperspectral imaging. Spectrochim. Acta A Mol. Biomol. Spectrosc. 198, 115–122. https://doi.org/10.1016/j.saa.2018.03.006.

Borowski C. Entwicklung von Verfahren und Bohrtechniken zur zufälligen Volumenelemententnahme aus Ballen; 2018.

bvse-Fachverband Kunststoffrecycling. bvse-Altkunststofftag hat begonnen - Qualität bleibt Thema Gastland ist die Türkei; 2016.

bvse-Fachverband Kunststoffrecycling. Kunststoffrecycler erhalten weiterhin ungenügende Qualität, 2017. https://www.bvse.de/gut-informiert-kunststoffrecycling/nachrichten-kunststoffrecycling/1650-kunststoffrecycler-erhalten-weiterhin-ungenuegende-qualitaet.html.

Calvini, R, Orlandi, G, Foca, G, Ulrici, A., 2018. Development of a classification algorithm for efficient handling of multiple classes in sorting systems based on hyperspectral imaging. J. Spectral Imag. https://doi.org/10.1255/jsi.2018.a13.

CEN/TR 15310-1, 2006. Characterisation of waste—sampling of waste materials—part 1: guidance on selection and application of criteria for sampling under various conditions. European Committee for Standardization, Brussels.

Chen, X, Feil, A., 2019. Detection and classification of heterogeneous materials as well as small particles using NIR-spectroscopy by validation of algorithms. In: Beyerer, J, Puente León, F, Längle, T (Eds.), OCM 2019 - Optical Characterization of Materials Conference Proceedings. KIT Scientific Publishing, pp. 63–77.

Chen, X, Kroell, N, Althaus, M, Pretz, T, Pomberger, R, Greiff, K, 2023. Enabling mechanical recycling of plastic bottles with shrink sleeves through near-infrared spectroscopy and machine learning algorithms. Resour. Conserv. Recycl. 188, 106719. https://doi.org/10.1016/j.resconrec.2022.106719.

Chen, X, Kroell, N, Li, K, Feil, A, Pretz, T., 2021. Influences of bioplastic polylactic acid on near-infrared-based sorting of conventional plastic. Waste management & research the journal of the International Solid Wastes and Public Cleansing Association. ISWA 39 (9), 734242X211003969. https://doi.org/10.1177/0734242X211003969.

COREPLA. General terms and conditions for sale by auction, 2022. https://www.corepla.it/sites/default/files/documenti/generaltermscond_rev._no15_march_2022.pdf (accessed April 15, 2023).

Cudjoe, D, Zhu, B, Nketiah, E, Wang, H, Chen, W, Qianqian, Y., 2021. The potential energy and environmental benefits of global recyclable resources. Sci. Total Environ. 798, 149258 https://doi.org/10.1016/j.scitotenv.2021.149258.

Curtis, A, Küppers, B, Möllnitz, S, Khodier, K, Sarc, R., 2021. Real time material flow monitoring in mechanical waste processing and the relevance of fluctuations. Waste Manage. 120 (1), 687–697. https://doi.org/10.1016/j.wasman.2020.10.037.

Dehoust G, Christiani J. Analyse und Fortentwicklung der Verwertungsquoten für Wertstoffe: Sammel- und Verwertungsquoten für Verpackungen und stoffgleiche Nichtverpackungen als Lenkungsinstrument zur Ressourcenschonung; 2012.

Der Grüne Punkt. Allgemeine Vertragsbedingungen der Der Grüne Punkt – Duales System Deutschland GmbH (DSD) für Verwerterverträge (AVBV) - Erlöskunden - Stand 07.2016, 2016. https://www.gruener-punkt.de/fileadmin/Dateien/Downloads/PDFs/geschaeftsbedingungen/AVBV_Erloeskunden_2016_07_04.pdf (accessed April 15, 2023).

Der Grüne Punkt. DSD Rohstofffraktionsspezifikationen, 2023. https://www.gruener-punkt.de/de/downloads (accessed April 17, 2023).

DIN 54390, 2022. Deutsches Institut für Normung e.V. Beuth Verlag, Berlin.

Duan, Q, Li, J., 2021. Classification of common household plastic wastes combining multiple methods based on near-infrared spectroscopy. ACS ES&T Engg 1 (7), 1065–1073. https://doi.org/10.1021/acsestengg.0c00183.

EU Recycling. Wie können mehr Kunststoffe in Leichtverpackungen recycelt werden?, 2018. https://eu-recycling.com/Archive/20131.

European Commission. The European Green Deal. Brussels, Belgium; 2019.

European Commission, 2020. Circular Economy Action Plan: For a cleaner and more competitive Europe. Brussels.

EVK Kerschhaggl GmbH. EVK SQALAR: Software Tool for the qualitative and quantitative analysis, 2023a. https://www.evk.biz/en/products/analysis-software-tool/evk-sqalar/(accessed April 21, 2023).

EVK Kerschhaggl GmbH. EVK STREAM Supervisor, 2023b. https://www.evk.biz/en/products/analysis-software-tool/evk-stream-supervisor/(accessed March 19, 2023).

Fahrmeir, L, Kneib, T, Lang, S, Marx, B., 2013. Regression: Models, methods and applications. Springer, Berlin, Heidelberg.

Feil, A, Kroell, N, Pretz, T, Greiff, K., 2021. Anforderungen an eine effiziente technologische Behandlung von Post-Consumer Verpackungsmaterialien in Sortieranlagen. Müll und Abfall 21 (7), 362–370. https://doi.org/10.37307/j.1863-9763.2021.07.04.

Feil, A, Pretz, T., 2020. Mechanical recycling of packaging waste. editor. In: Letcher, TM (Ed.), Plastic waste and recycling: Environmental impact, societal issues, prevention, and solutions. Academic Press, Amsterdam, San Diego, Cambridge, MA, Kidlington, pp. 283–319.

GBP Quality GmbH. Sortieranalysen: Aufgabe und Lösung, 2023. https://www.gbp-quality.eu/ablauf-arbeitsweise-gbp-sortieranalysen/(accessed April 15, 2023).

Goodfellow, I, Bengio, Y, Courville, A, 2016. Deep Learning. MIT Press.

Hahladakis, JN, Iacovidou, E., 2018. Closing the loop on plastic packaging materials: What is quality and how does it affect their circularity? Sci. Total Environ. 630, 1394–1400. https://doi.org/10.1016/j.scitotenv.2018.02.330.

Harris, CR, Millman, KJ, van der Walt, SJ, Gommers, R, Virtanen, P, Cournapeau, D, et al., 2020. Array programming with NumPy. Nature 585 (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hunter, JD, 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

Initiative „Mülltrennung wirkt“, 2021. Ein Jahr Initiative „Mülltrennung wirkt“: Von den Herausforderungen zum Erfolg: Duale Systeme bilden eine starke Allianz für mehr Aufklärung. Müll und Abfall (7), 352–360. https://doi.org/10.37307/j.1863-9763.2021.07.03.

Knappe F, Reinhardt J, Kauertz B, Oetjen-Dehne R, Buschow N, Ritthoff M, et al. Technische Potenzialanalyse zur Steigerung des Kunststoffrecyclings und des Rezyklateinsatzes; 2021. TEXTE 92.

Kroell, N, Chen, X, Greiff, K, Feil, A., 2022a. Optical sensors and machine learning algorithms in sensor-based material flow characterization for mechanical recycling processes: A systematic literature review. Waste Manage. 149, 259–290. https://doi.org/10.1016/j.wasman.2022.05.015.

Kroell, N, Chen, X, Küppers, B, Lorenzo, J, Maghmoumi, A, Schlaak, M, et al., 2023a. Near-infrared-based determination of mass-based material flow compositions in mechanical recycling of post-consumer plastics: Technical feasibility enables novel applications. Resour. Conserv. Recycl. 191, 106873 https://doi.org/10.1016/j.resconrec.2023.106873.

Kroell, N, Chen, X, Maghmoumi, A, Koenig, M, Feil, A, Greiff, K., 2021. Sensor-based particle mass prediction of lightweight packaging waste using machine learning algorithms. Waste Manage. 136, 253–265. https://doi.org/10.1016/j.wasman.2021.10.017.

Kroell, N, Chen, X, Maghmoumi, A, Lorenzo, J, Schlaak, M, Nordmann, C, et al., 2023b. NIR-MFCO dataset: Near-infrared-based false-color images of post-consumer plastics at different material flow compositions and material flow presentations. Data Brief 48, 109054. https://doi.org/10.1016/j.dib.2023.109054.

Kroell, N, Dietl, T, Maghmoumi, A, Chen, X, Küppers, B, Feil, A, et al., 2022b. Assessment of sensor-based sorting performance for lightweight packaging waste through sensor-based material flow monitoring: Concept and preliminary results. In: Greiff, K, Wotruba, H, Feil, A, Kroell, N, Chen, X, Gürsel, D, Merz, V (Eds.), 9th Sensor-Based Sorting & Control 2022; 13.04.2022 - 14.04.2022: Aachen.

Küppers, B, Schloegl, S, Oreski, G, Pomberger, R, Vollprecht, D., 2019. Influence of surface roughness and surface moisture of plastics on sensor-based sorting in the near infrared range. Waste management & research the journal of the International Solid Wastes and Public Cleansing Association. ISWA 37 (8), 843–850. https://doi.org/10.1177/0734242X19855433.

Küppers, B, Schlögl, S, Friedrich, K, Lederle, L, Pichler, C, Freil, J, et al., 2021. Influence of material alterations and machine impairment on throughput related sensor-based sorting performance. Waste management & research the journal of the International Solid Wastes and Public Cleansing Association. ISWA 39 (1), 122–129. https://doi.org/10.1177/0734242X20936745.

Küppers, B, Schlögl, S, Kroell, N, Radkohl, V., 2022. Relevance and challenges of plant control in the pre-processing stage for enhanced sorting performance. In: Greiff, K, Wotruba, H, Feil, A, Kroell, N, Chen, X, Gürsel, D, Merz, V (Eds.), 9th Sensor-Based Sorting & Control 2022; 13.04.2022 - 14.04.2022: Aachen.

Küppers, B, Seidler, I, Koinig, GR, Pomberger, R, Vollprecht, D., 2020. Influence of throughput rate and input composition on sensor-based sorting efficiency //Volume 09 - March 2020. Detritus (9), 59–67. https://doi.org/10.31025/2611-4135/2020.13906.

Länderarbeitsgemeinschaft Abfall LAGA PN 98, 2001.

Mann, HB, Whitney, DR., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18 (1), 50–60. https://doi.org/10.1214/aoms/1177730491.

McKinney, W., 2010. Data structures for statistical computing in python. In: Stéfan van der Walt, Jarrod Millman. Proceedings of the 9th Python in Science Conference.

Österreichisches Normungsinstitut. ÖNORM S 2127, 2011 (accessed February 11, 2023).

Pearson, K., 1895. VII. Note on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. 58 (347-352), 240–242. https://doi.org/10.1098/rspl.1895.0041.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, et al., 2011. Scikit-learn: machine learning in python. J. Open Source Software 12, 2825–2830.

Flamme, S, Kölking, M, Glorius, T, Mayer, S, 2020. Langzeituntersuchungen zur stoffstromspezifischen Kalibrierung eines nahinfrarotgestützten Echtzeitanalyse-Systems. In: Pomberger, R: Konferenzband zur 15. Recy & DepoTech-Konferenz, Montanuniversität Leoben, Österreich sowie virtuelle Konferenzwelt auf meetyoo, 18.-20. November 2020: Leoben, Österreich: AVW, Abfallverwertungstechnik & Abfallwirtschaft Eigenverlag, 2020.

Plastic Europe. Plastics – the Facts 2022; 2022a.

Plastic Europe. The circular economy for plastics: a European overview, 2022b. https://plasticseurope.org/knowledge-hub/the-circular-economy-for-plastics-a-european-overview-2/(accessed May 27, 2022).

plasticker. "Recybase" - Börse für Kunststoff-Rohstoffe, 2023. https://plasticker.de/recybase/(accessed April 19, 2023).
Roosen, M, Mys, N, Kusenberg, M, Billen, P, Dumoulin, A, Dewulf, J, et al., 2020. Detailed analysis of the composition of selected plastic packaging waste products and its implications for mechanical and thermochemical recycling. Environ. Sci. Technol. 54 (20), 13282–13293. https://doi.org/10.1021/acs.est.0c03371A.
The pandas development team. pandas-dev/pandas: Pandas: Zenodo, 2020.
United Nations. Paris Agreement. United Nations Framework Convention on Climate Change; 2015.
Virtanen, P, Gommers, R, Oliphant, TE, Haberland, M, Reddy, T, Cournapeau, D, et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
Waskom, ML., 2021. Seaborn: statistical data visualization. J. Open Source Softw. 6 (60), 3021. https://doi.org/10.21105/joss.03021.
Workman, JJ, Weyer, L., 2007. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press.
Xia, J, Huang, Y, Li, Q, Xiong, Y, Min, S., 2021. Convolutional neural network with near-infrared spectroscopy for plastic discrimination. Environ. Chem. Lett. 19 (5), 3547–3555. https://doi.org/10.1007/s10311-021-01240-9.
Zheng, Y, Bai, J, Xu, J, Li, X, Zhang, Y., 2018. A discrimination model in waste plastics sorting using NIR hyperspectral imaging system. Waste Manage. 72, 87–98. https://doi.org/10.1016/j.wasman.2017.10.015.