



OPEN

# Systematic reduction of hyperspectral images for high-throughput plastic characterization

Mahdiyeh Ghaffari<sup>1</sup>✉, Mickey C. J. Lukkien<sup>1</sup>, Nematollah Omidikia<sup>1</sup>, Gerjen H. Tinnevelt<sup>1</sup>, Marcel C. P. van Eijk<sup>2</sup>, Stanislav Podchezertsev<sup>2</sup> & Jeroen J. Jansen<sup>1</sup>✉

Hyperspectral Imaging (HSI) combines microscopy and spectroscopy to assess the spatial distribution of spectroscopically active compounds in objects, and has diverse applications in food quality control, pharmaceutical processes, and waste sorting. However, due to the large size of HSI datasets, it can be challenging to analyze and store them within a reasonable digital infrastructure, especially in waste sorting where speed and data storage resources are limited. Additionally, as with most spectroscopic data, there is significant redundancy, making pixel and variable selection crucial for retaining chemical information. Recent high-tech developments in chemometrics enable automated and evidence-based data reduction, which can substantially enhance the speed and performance of Non-Negative Matrix Factorization (NMF), a widely used algorithm for chemical resolution of HSI data. By recovering the pure contribution maps and spectral profiles of distributed compounds, NMF can provide evidence-based sorting decisions for efficient waste management. To improve the quality and efficiency of data analysis on hyperspectral imaging (HSI) data, we apply a convex-hull method to select essential pixels and wavelengths and remove uninformative and redundant information. This process minimizes computational strain and effectively eliminates highly mixed pixels. By reducing data redundancy, data investigation and analysis become more straightforward, as demonstrated in both simulated and real HSI data for plastic sorting.

Hyperspectral Imaging (HSI) is a valuable non-destructive technology for online monitoring and high throughput screening, with huge potential to be automated<sup>1–4</sup>. HSI, specifically Near InfraRed (NIR), is both fast and non-destructive, which makes it excellently suitable to perform real-time “online” in identification of polymers<sup>3,5</sup>. The combination of HSI as a powerful tool for the optical sorting of plastics, and discriminant analysis, allows automatic sorting in real time under industrial conditions<sup>6</sup> into new well-characterized circular feedstock to replace virgin polymers. The high throughput required to process sufficient material of the enormous amount of plastic waste puts strong demands on the computational time and storage resources available to support the choice for sorting every waste element into an adequate stream: data analysis, processing, and decision-making should be as fast as possible<sup>7</sup>. An early implementation of NIR hyperspectral imaging-based cascade detection, exemplified by industry leaders like Steinert, Tomra, and Pellend, was already able to specifically sort post-consumer plastic packaging waste. Innovations in the development of such technology have been continuously ongoing<sup>8–10</sup>. The considerable promise of HSI, with its challenging combinations of multivariate spectral and spatially resolved information, have made multivariate methods essential to extract hidden chemical information<sup>11</sup>. Indeed, deep learning is widely applicable in the field of waste management, and one may find a vast number of commercial products available on the market, such as waste stream analyzers and vision systems for robotic sorting. As for adopting deep learning methods to hyperspectral imaging, deep learning models would require more effort in crafting a high-quality training dataset on one hand while on the other hand generally provide less interpretability.

Chemometrics translates abstract spectroscopic fingerprints into throughput plastic species compositions of each waste element plays an indispensable role in the analysis of plastics during the recycling process.

<sup>1</sup>Institute for Molecules and Materials, Analytical Chemistry, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands. <sup>2</sup>National Test Centre Circular Plastics (NTCP), Duitslanddreef 7, 8447 SE Heerenveen, The Netherlands. ✉email: mahdiyeh.ghaffari@ru.nl; jeroen.jansen@ru.nl

Chemometrics has proven invaluable, by variable selection<sup>12</sup>, Multivariate Curve Resolution (MCR)<sup>13</sup>, and classification algorithms to mix chemical domain knowledge with data-driven machine learning. It plays a vital role in tackling the complexities presented by challenging objects, particularly multilayers, as it enables the analysis and interpretation of the intricate spectral information embedded within these composite structures. However, the size of Hyperspectral data sets in both the spatial and spectral dimensions may considerably limit the speed of information extraction from such data. The Spectral and spatial redundancy of HSI data however enables a considerable data reduction to reduce resources sufficiently for high-throughput applications in the circular economy<sup>14,15</sup>.

Non-negative matrix factorization (NMF) with the inherent non-negativity property originates in signal/image processing with some reports in analytical chemistry yet is largely mathematically equivalent to Multivariate Curve Resolution (MCR-ALS)<sup>16,17</sup> that is a cornerstone of chemometrics. The NMF algorithm has been evaluated in chemistry e.g. for the resolution of overlapped GC-MS spectra, as well as for the deconvolution of GC×GC data set<sup>18</sup>. Besides, NMF has been conducted to classify complex mixtures based on the extracted feature and to analyze time-resolved optical waveguide absorption spectroscopy data<sup>15</sup>. NMF recovers pure spectra and their concentration contribution maps of HSI Raman images<sup>19</sup>. Another study applied an NMF filter to triboluminescence (TL) data traces of active pharmaceutical ingredients<sup>20</sup>, leading to simultaneously recovering both photon arrival times and the instrument impulse response function, i.e. is of considerable value to recover chemical information from HSI data. Zushi proposed an NMF-based spectral deconvolution approach, coupled with a web platform GC mixture<sup>21</sup>, that utilizes a faster multiplicative update method instead of the traditional projection step. This approach is highly advantageous for analyzing large mass spectrometry imaging datasets due to its improved speed<sup>22</sup>.

Monolayer materials composed of a single polymer species or multilayer materials made up of several polymers are common in plastics. With the help of a single spectral profile, it is possible to identify the polymer composition of monolayer objects<sup>23</sup>. When it comes to identifying the polymer composition of waste streams, the objects encountered are often much more complex than monolayer materials. These objects can be multilayered or composed of multiple polymer species, or even coated with labels made of different polymers. Furthermore, an object may be made up of known polymers with an unknown composition. In the case of multilayer plastics, the type and ratio of plastics used can also vary significantly, such as 70/30 or 50/50. This adds significant complexity to the classification process, which is where the importance of data unmixing comes into play.

Data unmixing is a valuable tool that can handle these issues effectively. However, when sorting the objects, the decision to translate a model observation into a specific engineered materials stream may only be made on a small subset of the complete object. Therefore, reducing the size of the data using convex hull is an excellent innovation in curve resolution that can reduce the data to less than one percent of the original data. Handling large HSIs, data compression/size reduction coupled with rapid data analysis tools has top priority. So, unnecessary pixels (data rows) and wavelengths (data columns) should be jointly removed and the rest are essential to characterize. For the study in this research, removing all unnecessary information and the reduced data will be subjected to further decomposition, NMF. The proposed strategy is designed to address scenarios in which novel materials, previously unaccounted for, are introduced into the sorting process. In such instances, two primary scenarios warrant consideration; *Partially Known Materials*: When an object comprises a blend of a known polymer and an unidentified material, it is not classified solely within the category of the known material. Instead, it is directed into multilayer streams containing the known polymer. This methodology ensures the proper identification and separation of the known component, even in the presence of unidentified materials. *New Types of polymers*: In the event of entirely new polymer types being introduced into the packaging for the first time, this method demonstrates adaptability. The model can be updated and extended to encompass the characteristics of the new material, facilitating the effective handling of previously unmodeled materials within the sorting process.

## Materials and methods

### A brief description of Non-Negative Matrix Factorization

Nonnegative Matrix Factorization (NMF) deconvolutes a matrix,  $\mathbf{R}_{IJ \times K}$ , into the product of matrices  $\mathbf{W}_{IJ \times n}$  and  $\mathbf{H}_{n \times K}^T$  with an intrinsic non-negativity property as a minimal constraint.  $I$  and  $J$  represent the number of spatial pixels, while  $K$  represents the number of variables in the case of a hyperspectral image.

$$\mathbf{R}_{IJ \times K} = \mathbf{W}_{IJ \times n} \mathbf{H}_{n \times K}^T + \mathbf{E}_{IJ \times K} \quad (1)$$

To this, formalize NMF optimizes the following cost function:

$$\min_{(\mathbf{W} \geq 0, \mathbf{H} \geq 0)} \|\mathbf{R}_{IJ \times K} - \mathbf{W}_{IJ \times n} \mathbf{H}_{n \times K}^T\|_F \quad (2)$$

Different algorithms are introduced in the literature to calculate component matrices, The multiplicative update rule introduced by Le ad Seung is simple to implement<sup>5</sup>. The updating parts are:

$$\mathbf{H}_{K \times n} \leftarrow \mathbf{H}_{K \times n} \frac{(\mathbf{W}_{n \times I}^T \mathbf{R}_{IJ \times K})}{(\mathbf{W}_{n \times I}^T \mathbf{W}_{IJ \times n} \mathbf{H}_{n \times K}^T)} \quad (3)$$

and

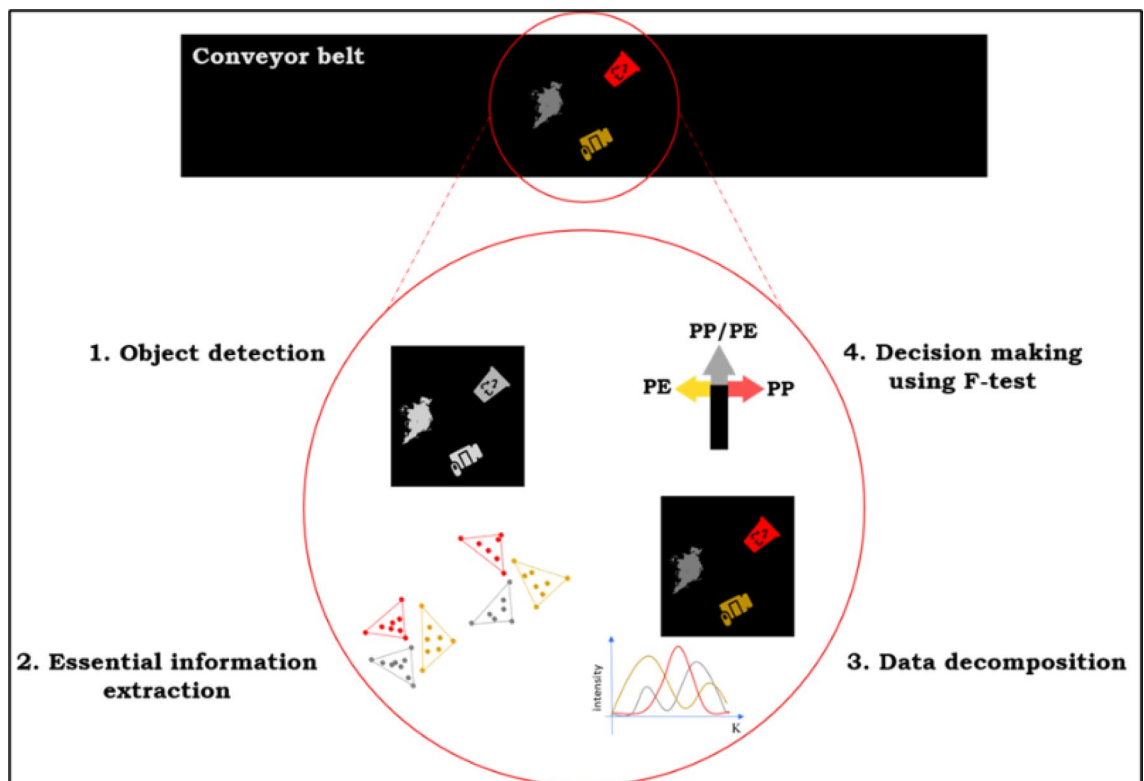
$$\mathbf{W}_{IJ \times n} \leftarrow \mathbf{W}_{IJ \times n} \frac{(\mathbf{R}_{IJ \times K} \mathbf{H}_{n \times K}^T)}{(\mathbf{W}_{IJ \times n} \mathbf{H}_{n \times K}^T \mathbf{H}_{K \times n})} \quad (4)$$

Equations (3) and (4) denote element-wise multiplications. These update rules preserve the non-negativity of  $\mathbf{W}_{IJ \times n}$  and  $\mathbf{H}_{K \times n}$  where  $\mathbf{R}_{IJ \times K}$  is element-wise non-negative.

### Approach

For any bilinear data set, a minimum number of rows and columns carry the most informative and independent part of the data<sup>24</sup>. Consequently, in HSIs, essential pixels, and essential wavelengths are necessary to extract the pure contribution maps and spectral profiles of all components. The main steps of this approach for plastic characterization are visualized in Fig. 1 and summarized as:

1. **Object detection:** Recording a first-order spectrum with  $K$  channels for every pixel in an  $I$  by  $J$  scene into a data cube,  $\mathbf{R}_{I \times J \times K}$ . To analyze this cube for plastic characterization, object detection is needed. In this work the corresponding pixels of each object were detected by correlation growing algorithm. The results of object detection on  $\mathbf{R}_{I \times J \times K}$  are several cubes (as many as the number of objects) and each of them contains information about one of the objects. Object detection decomposes  $\mathbf{R}_{I \times J \times K}$  to some cubes.
2. **Essential information extraction:** This step will start by unfolding each data cube (resulted from previous step) to a matrix ( $\mathbf{R}_{IJ \times K}$ ) with each row a pixel for further analysis. Calculating the most informative pixels (Essential Spectral Pixels, ESPs) and variable/wavelengths (Essential Spatial Variables, ESVs) for each object, which are based on the convexity property in the normalized abstract row and column spaces, is the next step. ESP/ESVs are the smallest set of points needed to generate the whole data in a convex way in the abstract score space. Once ESP/ESVs are identified for all objects separately, all other measured pixels are removed and the reduced data is moved to the next step. Left/right eigenvectors of  $\mathbf{R}_{IJ \times K}$  using SVD,  $\mathbf{R}_{IJ \times K} = \mathbf{U}_{IJ,n} \mathbf{D}_{n,n} \mathbf{V}_{n,K}^T + \mathbf{E}_{IJ,K}$  can be calculate, where  $\mathbf{U}_{IJ,n}$  and  $\mathbf{V}_{n,K}^T$  are the left and right eigenvectors, respectively.  $\mathbf{D}_{n,n}$  and  $\mathbf{E}_{IJ,K}$  contains singular values and residuals, individually. In addition,  $n$  is the number of factors. The number of factors is set up to five. Because most of the objects contain less than five types of polymers/materials. "Convhulln" as a MATLAB function can calculate the convex set of  $\mathbf{U}_{IJ,n}$  and  $\mathbf{V}_{n,K}^T$  and explore the ESPs/ESVs. After join selection of essential pixels and essential variables,  $\mathbf{R}_{IJ \times K}$  turn into  $\mathbf{R}_{ESP \times K_{ESV}}$ . This step is explained in detail in the previous works<sup>24</sup>.
3. **Data decomposition:** The reduced data sets for all objects,  $\mathbf{R}_{ESP_s \times ESV_s}$  can be analyzed by NMF in parallel with multiplicative updates, to calculate the reduced concentration contribution maps,  $\mathbf{W}_{ESP_s \times n}^r$  and reduced spectral profiles  $\mathbf{H}_{ESV_s \times n}^r$  using Eqs. (3) and (4). Using least square, full concentration contribution maps,  $\mathbf{W}_{IJ \times n}$ , and full spectral profiles,  $\mathbf{H}_{K \times n}$  can be produced from reduced versions  $\mathbf{W}_{ESP_s \times n}^r$  and  $\mathbf{H}_{ESV_s \times n}^r$



**Figure 1.** A Graphical illustration of the approach based on the essential pixel/wavelength selection and NMF. The hyperspectral data with a 3D structure, needs object detection first. Then the selected parts of the data which correspond to the objects unfold to matrices to calculate the ESPs/ESVs. Finally, the reduced data sets are analyzed by NMF and full contribution maps and spectral profiles are retrieved. In the last step, F-test will help for decision making to monolayer/multilayer plastic sorting.

through Eqs. (5) and (6). This step needs  $\mathbf{R}_{\text{ESPs} \times K}$  and  $\mathbf{R}_{\text{IJ} \times \text{ESVs}}$  which are one-mode reduced data in the row and column direction respectively. Finally, it is easy to reshape the columns of  $\mathbf{W}_{\text{IJ} \times n}$  to generate the full concentration contribution maps.

$$\mathbf{W}_{\text{IJ} \times n} = \mathbf{R}_{\text{IJ} \times \text{ESVs}} * \text{pinv}(\mathbf{H}_{\text{ESVs} \times n}^r) \quad (5)$$

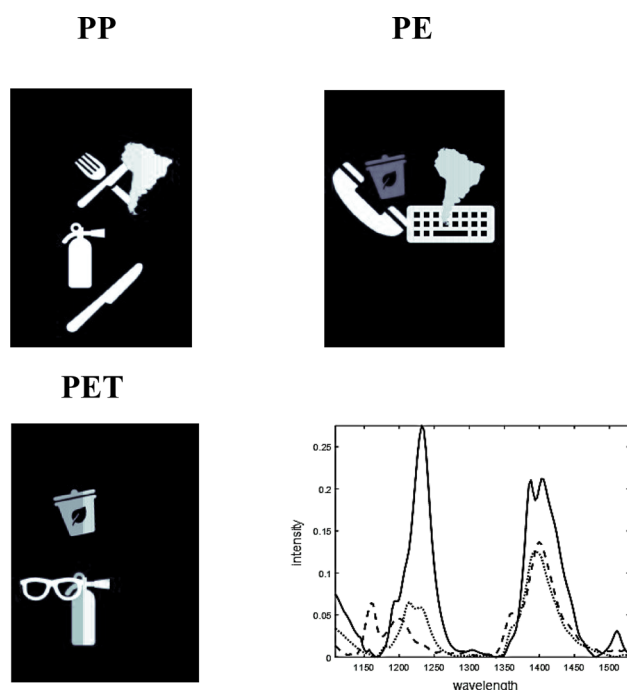
$$\mathbf{H}_{K \times n} = \text{pinv}(\mathbf{W}_{\text{ESPs} \times n}^r) * \mathbf{R}_{\text{ESPs} \times K} \quad (6)$$

4. **Decision making:** The matrix  $\mathbf{H}_{K \times n}$  consists of pure spectral profiles of the components, which can be utilized for qualification by comparing them to a reference library. However, in some cases,  $\mathbf{W}_{\text{IJ} \times n}$  has complementary information for identification.  $\mathbf{W}_{\text{IJ} \times n}$  contains characteristic information about the composition of unknown objects. The sum of the squares of the elements in each column of  $\mathbf{W}_{\text{IJ} \times n}$  represents the variance of the signal contributed by each polymer in an unknown object. This variance can be used to differentiate between mono-material and multi-material objects using a statistical F-test, which is commonly employed in statistics to compare the standard deviations of two populations. To conduct the F-test, the variance of all columns in  $\mathbf{W}_{\text{IJ} \times n}$  is calculated in the first step, and then each value is divided by the noise variance. A significance level of  $P < 0.05$  is used to determine whether the variance is statistically significant. If the object contains only one layer of material, only one type of polymer will pass the F-test. However, if it is a multilayer object, multiple types of polymer will pass the test. Finally, it should be emphasized that all of the computation and reported times in this work are based on a laptop (Intel(R) Core(TM) i7-10850H CPU @ 2.70 GHz) which for real industrial purposes can be dramatically improved by using better computers. The utilization of this algorithm enables the exploitation of the computational capabilities of a standard computer system, thereby eliminating the reliance on specialized hardware such as GPUs or high-performance computing clusters. The selection process of this approach, as explained in step 2, represents a delicate balance, and the aim is to protect essential information from loss. As hyperspectral imaging applications continue to expand, addressing challenges related to data storage and efficient analysis becomes increasingly crucial. This contribution lies in the development of an innovative approach that optimizes data reduction, streamlining the process and making it well-suited for real-time applications, such as industrial plastic sorting specially in the presence of multilayer and multicomponent packages.

## Data description

### Simulated data set

A hyperspectral image was simulated to visualize the effect of essential pixels and variables selection using convex polytope and further unmixing by NMF. The concentration contribution maps (re-folded concentration profiles) and pure spectral profiles are shown in Fig. 2. The simulated hyperspectral data set is of dimensions  $253 \times 186$  pixels by 141 variables and the unfolded two-way data matrix of 47,058 pixels and 141 pseudo-spectral channels. Despite the simplicity of the simulation, it should be noted that care was taken to avoid the pure pixels or



**Figure 2.** The three components simulated the HSI data set. Concentration distribution maps and pure spectral profiles are presented.

selective spectral channels, this corresponds to a non-trivial situation for NMF analysis. For this purpose, small random numbers were added to the pure contribution maps. In this case, eight objects are on the hypothetical conveyor belt which are made of polypropylene (PP), polyethylene (PE), and polyethylene terephthalate (PET). Five and three objects are monolayers and multilayers, respectively. Figure 2 presents the pure contribution maps and spectral profiles of PP, PE, and PET.

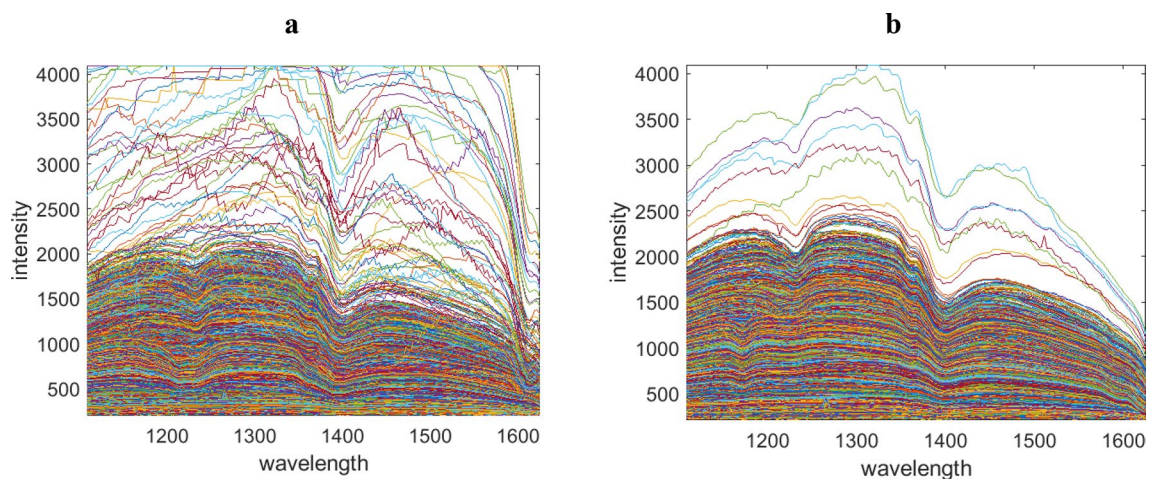
### Experimental hyperspectral images of plastics

A collection of monolayer and multilayer objects made from PP (polypropylene), PS (polystyrene), PET (polyethylene terephthalate), and polypropylene on top of polyethylene (multilayer), all with known compositions, were gathered from packaging waste. The characteristic information of these objects was known in advance. To collect data, the objects were randomly placed on a conveyor belt, and two HSI measurements were recorded in the 900–1700 nm range. The resulting data sets are presented in Fig. 3.

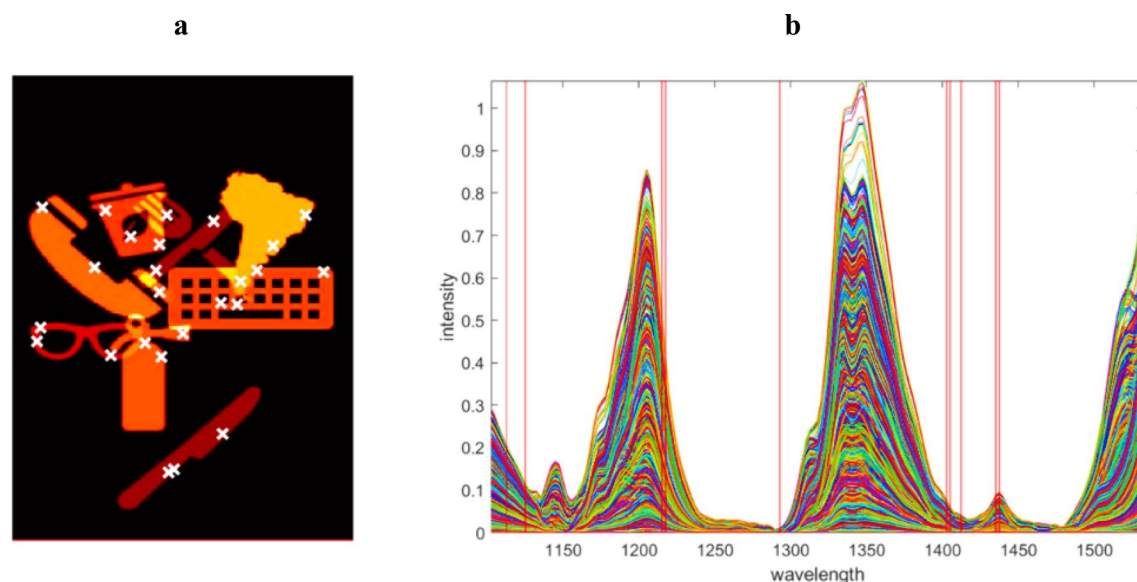
### Result and discussions

To illustrate the effect of essential spectral pixels (ESPs) and/or essential spectral variables (ESVs) selection of the plastic sorting using NMF, the results obtained on the simulation and real HSI data set are discussed in detail.

The procedure starts with object detection for the simulated case as it does not need any data per-processing. Then, the ESPs and ESVs should be selected using a convex hull on the normalized abstract spaces of the data set for each object. First five Principal Components were used to generate the abstract spaces. The selected essential pixels and wavelengths are presented in Fig. 4a and b. In Fig. 4, the left and right panels contain the mean image and the spectral data of the simulated case. The selected ESPs and ESVs are shown by white crosses and red lines.



**Figure 3.** The experimental cases are visualized in (a) and (b).



**Figure 4.** The selected ESPs and ESVs for the simulation case are indicated as white crosses and red lines, respectively.

The raw data matrix of the simulated case is composed of 57,078 rows (pixels) and 173 variables were reduced to eight matrices by object detection. For each object, two or three ESPs and a few ESVs are selected as it is shown in Fig. 4a,b. In total for the whole data, only 0.04% data variance remains. It means only 0.04 percent of the data is essential/enough to analyze.

Later, the reduced data sets were analyzed by NMF. The results of NMF are called **W** and **H**. To make a better visualization of the full contribution maps, **W** changed to binary matrices and visualized in Fig. 5. The total calculation time for NMF analysis of the reduced data sets was nearly 0.001 s, 1% of the time required for the same calculations on all pixels and wavelengths.

In addition, strategic data pre-processing methods are essential prior to data analysis to linearize the data, remove artifacts and thereby optimally align the data and the NMF with the Beer–Lambert–Bouguer law. Figure 6 presents the effect of each pre-processing step on the shape of the spectral data for an example object. Figure 6a is the recorded raw HSI of an optional object. Some spectral profiles which reflected the light can be removed first. These profiles can be recognized by the slope of the spectral profiles in the range of 1100–1500 nm. In this wavelength range the slope of the profiles are zero (all saturated).

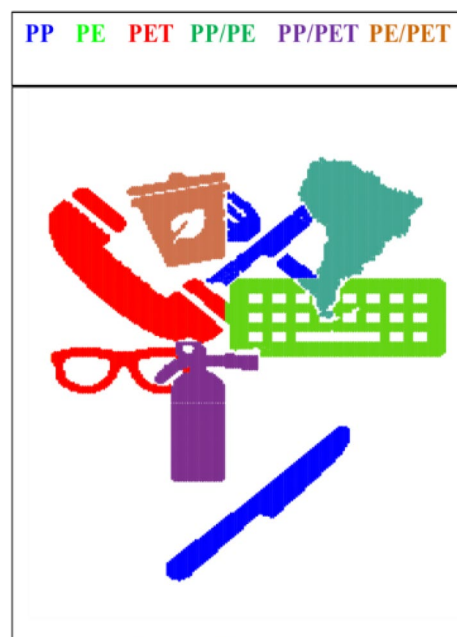
Figure 6b presents the normalized data after removing the saturated spectral profiles. The next step (6c) is the treated data after baseline correction of the using asymmetric least squares on the minus logarithm of the data set. Finally, the data set is reconstructed by a few principal components for denoising purposes. The corrected data is presented in Fig. 6d.

Once the data has been preprocessed, it undergoes object detection, as detailed in the theory section. This process yields smaller data sets, each one corresponding to a single object and containing relevant information. In the current scenario, object detection yields five data sets (corresponding to the five objects on the conveyor belt). The crux of the proposed algorithm lies in selecting the appropriate ESPs and ESVs based on the data's inner polygon in abstract spaces. Figures 7 and 8 illustrate the chosen ESPs and ESVs for both real cases, denoted by white crosses and red vertical lines. Finally, the size of the data is decreased to  $48 \times 12$  and  $149 \times 45$ , from the original dimensions of  $144,000 \times 173$ .

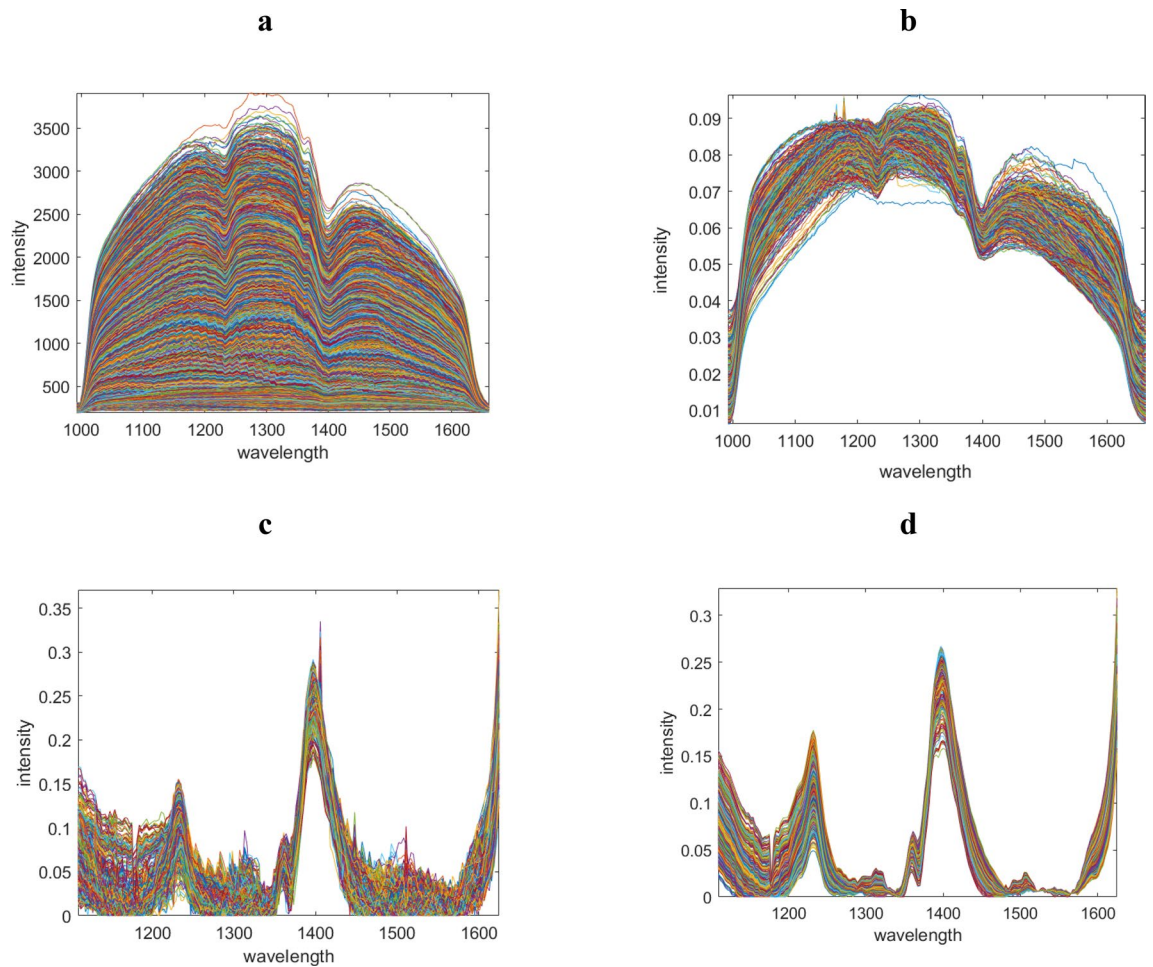
The reduced data sets were analyzed by NMF. The final resulting maps which identified the composition of each object are presented in Fig. 9. To generate these single contribution maps, one least square with a non-negative least square was necessary which was followed by binary maps construction. In the resulting images, each color is used for special composition as is indicated on the top of the figure as well. For comparison, the full data sets were analyzed by NMF as well which took almost 220 s. However, NMF for the reduced data sets took 0.003 s.

## Conclusion

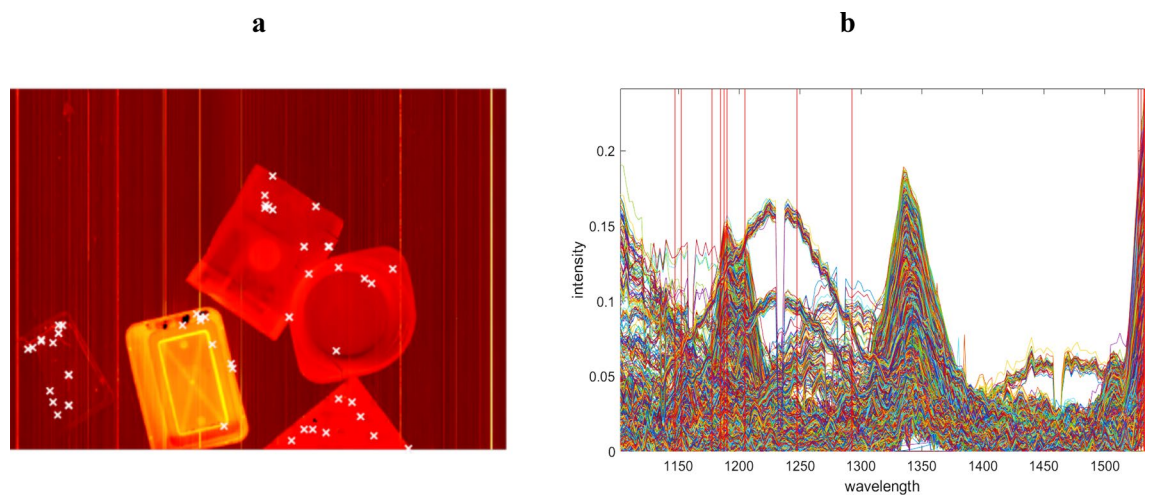
In this study, we present a pioneering application of our earlier research, focusing on the selection of essential pixels and variables within hyperspectral imaging (HSI) datasets. This application holds significant promise, particularly in the context of industrial plastic sorting, where it addresses the challenges of characterizing complex multicomponent and multilayer plastics. Building upon the foundation of our previous findings, which



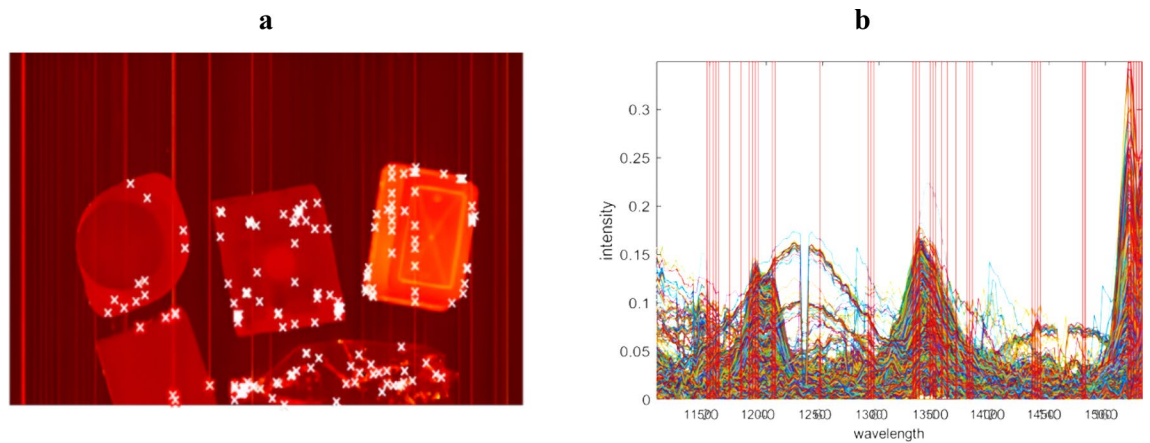
**Figure 5.** The full contribution maps of all components in simulated data are visualized. Each color is used for special composition. Blue, green, red, aqua, and purple, brown, are used for, PP, PE, PET, PP/PE (multilayer), PP/PET, and PE/PET, respectively.



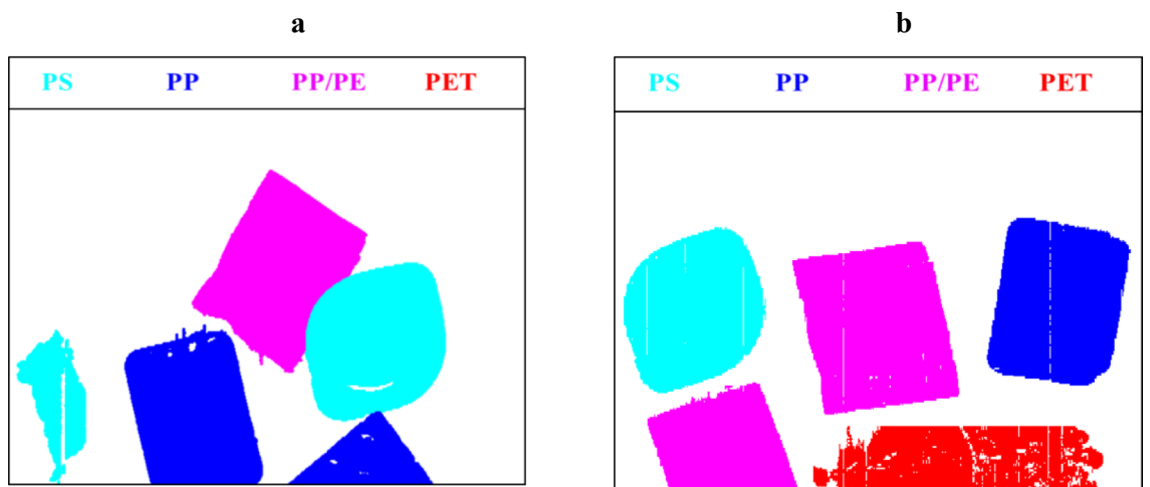
**Figure 6.** Depicts the data set corresponding to the first unknown object, visualized after undergoing various pre-treatment steps. These steps include object detection and removal of pixels affected by signal saturation, baseline correction of the negative logarithm of the data, and denoising using Singular Value Decomposition.



**Figure 7.** Depicts the mean image and spectral profile of the first experimental case, presented in (a) and (b) respectively. The chosen ESPs and ESVs are denoted by white crosses and red lines.



**Figure 8.** Depicts the mean image and spectral profile of the second experimental case, presented in (a) and (b) respectively. The chosen ESPs and ESVs are denoted by white crosses and red lines.



**Figure 9.** The full contribution maps of all components in both experimental data sets are visualized in (a) and (b). Each color is used for special composition. Cyan, blue, purple, and red, are used for PS, PP, PP/PE (multilayer), and PET, respectively.

demonstrated the efficacy of choosing Essential Spectral Pixels (ESPs) and Essential Spatial Variables (ESVs) to minimize redundancy and streamline data analysis, we have embarked on an innovative journey. Here, we deploy these principles in an entirely novel context, employing them in the realm of hyperspectral imaging for plastic sorting. Implementing variable and pixel selection algorithms significantly enhances computational efficiency and material detection capabilities. Striking a balance between computational speed and information retention is vital. Our proposed method carefully preserves the most informative pixels and variables, reducing the risk of data loss. Focusing on information-rich elements improves material detection precision and accuracy, holding promise for advancements in HSI data analysis across various applications.

While conventional plastic recycling typically relies on RGB color sorting, our method leverages hyperspectral imaging for deeper insights into plastic characteristics, especially in complex multicomponent and multilayer scenarios. Nowadays, HSIs are used in plastic sorting with the aim of plastic-type identification. So coupling data size reduction (which seems logical in big HSI analysis) with a fast algorithm is promising. The benefit of data reduction is not just for the analysis part. However, the reward can be used in the data recording scheme. Considering the wide NIR domain, recording a few wavenumbers rather than the whole scope can be dramatically advantageous. On the other hand, NMF is suggested to decompose reduced data in plastic sorting. This procedure can be induced to analyze data from complementary domains, like remote sensing, etc. However, plastic sorting is a case in that we explained the procedure based on it (Supplementary Information).

### Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.



Received: 26 July 2023; Accepted: 4 December 2023

Published online: 07 December 2023

## References

- Bonifazi, G., Gasbarrone, R. & Serranti, S. Detecting contaminants in post-consumer plastic packaging waste by a NIR hyperspectral imaging-based cascade detection approach. *Detritus* <https://doi.org/10.31025/2611-4135/2021.14086> (2021).
- Xu, J. & Mishra, P. Combining deep learning with chemometrics when it is really needed: A case of real time object detection and spectral model application for spectral image processing. *Anal. Chim. Acta* **1202**, 339668 (2022).
- da Silva, V. H. *et al.* Classification and quantification of microplastics (< 100  $\mu\text{m}$ ) using a focal plane array–Fourier transform infrared imaging system and machine learning. *Anal. Chem.* **92**(20), 13724–13733 (2020).
- Ghaffari, M., Omidikia, N. & Ruckebusch, C. Essential spectral pixels for multivariate curve resolution of chemical images. *Anal. Chem.* **91**(17), 10943–10948 (2019).
- Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999).
- Zushi, Y. & Hashimoto, S. Direct classification of GC $\times$  GC-analyzed complex mixtures using non-negative matrix factorization-based feature extraction. *Anal. Chem.* **90**(6), 3819–3825 (2018).
- Nandakumar, A. *et al.* Bioplastics: A boon or bane?. *Renew. Sustain. Energy Rev.* **147**, 111237 (2021).
- Vidal, C. & Pasquini, C. A comprehensive and fast microplastics identification based on near-infrared hyperspectral imaging (HSI-NIR) and chemometrics. *Environ. Pollut.* **285**, 117251 (2021).
- Huth-Fehre, T. *et al.* NIR—Remote sensing and artificial neural networks for rapid identification of post consumer plastics. *J. Mol. Struct.* **348**, 143–146 (1995).
- van den Broek, W. H. A. M. *et al.* Identification of plastics among nonplastics in mixed waste by remote sensing near-infrared imaging spectroscopy. 1. Image improvement and analysis by singular value decomposition. *Anal. Chem.* **67**(20), 3753–3759 (1995).
- Amigo, J. M., Babamoradi, H. & Elcoroaristizabal, S. Hyperspectral image analysis. A tutorial. *Anal. Chim. Acta* **896**, 34–51 (2015).
- Paatero, P. & Tapper, U. J. E. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**(2), 111–126 (1994).
- Gao, H.-T. *et al.* Overlapping spectra resolution using non-negative matrix factorization. *Talanta* **66**(1), 65–73 (2005).
- Zushi, Y., Hashimoto, S. & Tanabe, K. Global spectral deconvolution based on non-negative matrix factorization in GC $\times$  GC–HRTOFMS. *Anal. Chem.* **87**(3), 1829–1838 (2015).
- Liu, P. *et al.* The application of principal component analysis and non-negative matrix factorization to analyze time-resolved optical waveguide absorption spectroscopy data. *Anal. Methods* **5**(17), 4454–4459 (2013).
- Thiel, M. *et al.* Comparison of chemometrics strategies for the spectroscopic monitoring of active pharmaceutical ingredients in chemical reactions. *Chemom. Intell. Lab. Syst.* **211**, 104273 (2021).
- Liu, X.-Y. *et al.* Spatiotemporal organization of biofilm matrix revealed by confocal Raman mapping integrated with non-negative matrix factorization analysis. *Anal. Chem.* **92**(1), 707–715 (2019).
- Anbumalar, S., Ananda Natarajan, R. & Rameshbabu, P. Non-negative matrix factorization algorithm for the deconvolution of one dimensional chromatograms. *Appl. Math. Comput.* **241**, 242–258 (2014).
- Szymańska-Chargot, M. *et al.* Hyperspectral image analysis of Raman maps of plant cell walls for blind spectra characterization by nonnegative matrix factorization algorithm. *Chemom. Intell. Lab. Syst.* **151**, 136–145 (2016).
- Griffin, S. R. *et al.* Iterative non-negative matrix factorization filter for blind deconvolution in photon/ion counting. *Anal. Chem.* **91**(8), 5286–5294 (2019).
- Zushi, Y. NMF-based spectral deconvolution with a web platform GC mixture touch. *ACS Omega* **6**(4), 2742–2748 (2021).
- Trindade, G. F. *et al.* Non-negative matrix factorisation of large mass spectrometry datasets. *Chemom. Intell. Lab. Syst.* **163**, 76–85 (2017).
- Henriksen, M. L. *et al.* Plastic classification via in-line hyperspectral camera analysis and unsupervised machine learning. *Vib. Spectrosc.* **118**, 103329 (2022).
- Ghaffari, M., Omidikia, N. & Ruckebusch, C. Joint selection of essential pixels and essential variables across hyperspectral images. *Anal. Chim. Acta* **1141**, 36–46 (2021).

## Acknowledgements

This project is co-funded by TKI-E&I with the supplementary grant 'TKI- Toeslag' for Top consortia for Knowledge and Innovation (TKI's) of the Ministry of Economic Affairs and Climate Policy (CP-50-02). We thank all partners in the project "Towards improved circularity of polyolefin-based packaging", managed by ISPT and DPI in the Netherlands. It was partly funded by the Perfect Sorting Consortium, a consortium that develops AI technology to enable the intended use of sorting for the recycling of packaging material. The members are NTCP, Danone, Colgate-Palmolive, Ferrero, LVMH, Mars, Michelin, Nestlé, Procter & Gamble, PepsiCo, Ghent University, and Radboud University.

## Author contributions

M.G., M.C.J.L., N.O., G.H.T., and J.J.J. are affiliated with the Institute for Molecules and Materials at Radboud University. In this research work, M.G. played a significant role in the theoretical development of the intellectual framework and its implementation in various simulated and experimental cases. M.G. also contributed extensively to the programming and interpretation of the obtained results. Furthermore, M.G. took the lead in drafting and revising the manuscript. N.O. and G.H.T. reviewed and approved the final version of the manuscript, including the references. J.J.J. and M.C.J. actively contributed to the advancement of the research. J.J.J. contributed to the discussions that shaped this work. They also contributed to the revision and approval of the final version. M.C.P.v.E. and S.P. affiliated with the National Test Centre Circular Plastics (NTCP) in the Netherlands, conducted the experimental imaging at NTCP. M.C.P.v.E. and S.P. reviewed and approved the final version of the manuscript, ensuring the accuracy and inclusion of appropriate references. Finally, both have made significant contributions to the revision of the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-49051-y>.

**Correspondence** and requests for materials should be addressed to M.G. or J.J.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023