



Statistical analysis of solid waste composition data: Arithmetic mean, standard deviation and correlation coefficients

Edjabou, Maklawe Essonanawe; Martín-Fernández, Josep Antoni; Scheutz, Charlotte; Astrup, Thomas Fruergaard

Published in:
Waste Management

Link to article, DOI:
[10.1016/j.wasman.2017.08.036](https://doi.org/10.1016/j.wasman.2017.08.036)

Publication date:
2017

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Edjabou, M. E., Martín-Fernández, J. A., Scheutz, C., & Astrup, T. F. (2017). Statistical analysis of solid waste composition data: Arithmetic mean, standard deviation and correlation coefficients. *Waste Management*, 69, 13-23. <https://doi.org/10.1016/j.wasman.2017.08.036>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Published in Waste Management

**Statistical analysis of solid waste composition data:
arithmetic mean, standard deviation and correlation
coefficients**

Maklawe Essonanawe Edjabou^{1*}, Josep Antoni Martín-
Fernández², Charlotte Scheutz¹, Thomas Fruergaard Astrup¹

1) Department of Environmental Engineering, Technical
University of Denmark, 2800 Kgs. Lyngby, Denmark

2) Dept. Computer Science, Applied Mathematics and
Statistics, University of Girona, Campus Montilivi (P4), E-
17071 Girona, Spain

1 **Title of paper: Statistical analysis of solid waste composition data: Arithmetic mean, standard**
2 **deviation and correlation coefficients**

3 **The core findings of the paper:**

4

5 • Data for waste fraction compositions represent closed datasets that require special attention in case of
6 statistical analysis

7 • Classical statistics are ill-suited to data for waste fraction compositions

8 • Isometric log-ratio coordinates enable appropriate transformation of waste fraction compositional data prior to
9 statistical analysis.

10

18 **Abstract**

19 Data for fractional solid waste composition provide relative
20 magnitudes of individual waste fractions, the percentages of
21 which always sum to 100, thereby connecting them
22 intrinsically. Due to this sum constraint, waste composition
23 data represent closed data, and their interpretation and analysis
24 require statistical methods, other than classical statistics that are
25 suitable only for non-constrained data such as absolute values.
26 However, the closed characteristics of waste composition data
27 are often ignored when analysed. The results of this study
28 showed, for example, that unavoidable animal-derived food
29 waste amounted to $2.21 \pm 3.12\%$ with a confidence interval of (-
30 4.03; 8.45), which highlights the problem of the biased negative
31 proportions. A Pearson's correlation test, applied to waste
32 fraction generation (kg mass), indicated a positive correlation
33 between avoidable vegetable food waste and plastic packaging.
34 However, correlation tests applied to waste fraction
35 compositions (percentage values) showed a negative
36 association in this regard, thus demonstrating that statistical
37 analyses applied to compositional waste fraction data, without
38 addressing the closed characteristics of these data, have the
39 potential to generate spurious or misleading results. Therefore,
40 "compositional data should be transformed adequately prior to
41 any statistical analysis, such as computing mean, standard
42 deviation and correlation coefficients.

43

| | |
|----|-----------------------------|
| 44 | Keywords: |
| 45 | Waste composition |
| 46 | Compositional data analysis |
| 47 | Isometric log ratio |
| 48 | Variation array |
| 49 | |

50 **1. Introduction**

51 Knowledge of the individual material fractions in waste
52 represents the basis of any waste management system planning
53 and development (Christensen, 2011). This information is also
54 crucial for establishing baselines and evaluating the
55 effectiveness of environmental policies. Generally, the
56 fractional composition of waste is obtained by conducting
57 waste fraction composition studies and is usually provided as
58 weight percentages of selected materials such as paper, plastic,
59 metal, food waste, etc. (Lagerkvist et al., 2011). Independently
60 of waste characterisation methods, waste fraction composition
61 arithmetic mean and standard deviation are usually provided
62 (European Commission, 2004), thus ignoring the inherent
63 structure of data for waste fraction compositions (Pawlowsky-
64 Glahn et al., 2015). Here, the standard deviation measures the
65 ‘spread’ of the estimated arithmetic mean (Reimann et al.,
66 2008).

67 Waste fraction composition data are ‘closed’ datasets
68 because of the limited sample space (from 0 to 100 i.e.
69 percentages). This is known as the ‘constant sum constraint’
70 (Aitchison, 1986), where the percentage of one waste fraction
71 depends on the ratio of the other waste fractions included in
72 the sampled waste stream. Consequently, the percentages of
73 waste fractions are linked to each other intrinsically. Therefore,

74 univariate analysis (composition of waste fractions analysed
75 separately) of waste fraction compositions is inappropriate,
76 because it violates the fundamental assumption of
77 independence of observations (Pawlowsky-Glahn et al., 2015).
78 For example, Hanc et al. (2011) studied the composition of
79 household bio-waste and reported that the yearly percentage of
80 grass amounted to $27.6 \pm 30.8\%$ in single-family areas. The
81 mean was 27.6% and its standard deviation 30.8%. The
82 resulting confidence interval ($2 \times$ standard deviation) of the
83 mean was the interval (-34.0% ; 89.2%), which covers negative
84 percentages, although the values cannot be negative in this
85 case. This problem is described as ‘intervals covering negative
86 proportions’ (Pawlowsky-Glahn et al., 2015). An increase in
87 the percentage of one waste fraction leads to a decrease in the
88 percentage of another fraction and vice versa, because the sum
89 of the percentage of individual waste fraction is fixed
90 (Reimann et al., 2008).

91 Data for waste fraction compositions refer to
92 compositional data, which arise in many fields such as
93 geochemistry (mineral composition of rocks), medicine (blood
94 composition) and archaeology (ceramic compositions)
95 (Aitchison, 1994). Here, compositional data carry relative
96 information or a ratio and add up to a constant (1 for
97 proportion, 100 for percentage and 10^4 for ppm (parts per
98 million)) (Aitchison, 1986; Buccianti and Pawlowsky-Glahn,

99 2011). As further examples, chemical compositionwaste water
100 content, etc. also represent closed datasets (see Aitchison,
101 1994).

102 Arithmetic mean and standard deviation are based on the
103 assumption that observations follow normal or symmetrical
104 statistical distribution (Reimann et al., 2008). Numerous –
105 mainly statistical-based – studies show that these estimates are
106 affected considerably when data exhibit small deviations from
107 normal distribution (Reimann et al., 2008; Wilcox, 2012). On
108 the other hand, environmental data including waste fraction
109 composition are often skewed (Reimann et al., 2008), in which
110 case the resulting descriptive statistics may be biased and
111 subsequently lead to wrong conclusions. Nevertheless, most
112 waste characterisation studies report the arithmetic mean and
113 standard deviation of waste fraction compositions, ignoring the
114 natural structure of compositional data (e.g. Hanc et al., 2011;
115 Edjabou et al., 2015; Naveen et al., 2016).

116 Despite the importance of arithmetic mean and standard
117 deviation estimates in relation to waste composition, no
118 attempts have been made to address the quality of these
119 estimates.

120 Correlation coefficients between individual waste
121 fractions are commonly computed in order to investigate
122 relationships between material fractions in mixed waste (e.g.
123 Alter, 1989; Hanc et al., 2011; Naveen et al., 2016), but they

124 are also used to evaluate the quality and the source of elements
125 in chemical compositions of municipal solid waste (e.g. Hanc
126 et al., 2011; Naveen et al., 2016). An illustrative example is the
127 correlation between food waste and packaging materials such
128 as paper, board, plastic and metal. For example, Alter (1989)
129 claimed that an increase in food packaging may decrease food
130 waste occurring in households. In contrast, Williams et al. (2012)
131 argued that 20 to 25% of food waste generation is due to
132 packaging. Notwithstanding the relevance of correlation
133 analysis applied to waste fraction compositions, the
134 contradictory results of correlation coefficients (see Alter,
135 1989 and Williams, 2012) still require explanation.

136 Overall, computing arithmetic means, standard deviations
137 and correlation coefficients for material fraction compositions
138 may lead to biased results (Aitchison, 1994; Filzmoser and
139 Hron, 2008). Additionally, uncertainty analysis (e.g. Monte
140 Carlo analysis) of these datasets can be a source of concern
141 when the issue of independence between material fraction
142 compositions is either ignored or poorly addressed (Xu and
143 Gertner, 2008).

144 Several studies have attempted to analyse waste
145 composition data by applying log transformation (Chang and
146 Davila, 2008; Dahlén et al., 2007) or log-logistic
147 transformation (Milke et al., 2008). However, the
148 compositional nature of waste fraction composition remains

149 intrinsic for waste fraction composition data.

150 The overall aim of this paper is to demonstrate why
151 fractional waste composition data should be transformed
152 appropriately prior to statistical analysis. We compared some
153 commonly encountered classical statistics applied to waste
154 fraction compositions data and the compositional data analysis
155 technique based on log-ratio coordinates, by analysing the
156 fractional compositions of residual household waste in
157 Denmark.

158 **2 Methods and materials**

159 **2.1 Study area and waste sampling analysis**

160 We analysed residual household waste collected from 779
161 single-family areas in Denmark. In these residential areas,
162 paper, board, gardening waste, household hazardous waste,
163 waste electrical and electronic equipment (WEEE) and bulky
164 waste were source-segregated.

165 The residual household waste was generated over a one-
166 week period, collected directly from households and kept
167 separately for each household. Each waste bin was labelled
168 with the address of the household from where the waste was
169 collected. The waste bins were sealed tightly, to prevent
170 mixing of waste during transportation to the sorting facility.
171 Each household waste bin was weighed and sorted separately,
172 thereby enabling us to obtain data for residual household waste
173 for each house.

174 Collected residual household waste was sorted manually
175 into the following waste fractions (Table 1): (1) avoidable
176 vegetable food waste (AV), (2) avoidable animal-derived food
177 waste (AA), (3) unavoidable vegetable food waste (UV), (4)
178 unavoidable animal-derived food waste (UA), (5) paper &
179 board (Paper or Pa), (6) plastic packaging (Plastic or Pl), (7)
180 metal packaging (Metal or Me) and (8) other waste fractions
181 (Others or Ot). In the present study, ‘paper’ refers to paper and
182 board packaging. ‘Others’ refers to all other waste materials
183 not included in the first seven waste fractions in Table 1.
184 Avoidable food waste constitutes food and drinks that could
185 have been eaten but instead have been disposed of. It consists
186 of avoidable animal-derived (AA) and vegetable (AV) food
187 waste. Unavoidable food waste is food that is not edible under
188 normal conditions (Edjabou et al., 2016) and consists of
189 unavoidable animal-derived (UA) and vegetable (UV) food
190 waste. The detailed sub-fractions included in these waste
191 fractions are presented in Table 1.

192 In this study, waste fraction composition represents the
193 fractional composition of waste fractions expressed in
194 percentage terms. Waste fraction generation rates are the mass
195 of individual waste fractions in kg per capita per week.

196
197
198

Here (Table 1)

199

200 **2.2 Overview of statistical analysis: classical statistical**
201 **analysis**

202 For this study, we computed (1) the arithmetic mean
203 (Mean) of waste fraction compositions, (2) log-transformed
204 (log-Mean), and its back-transformed ($\exp(\log\text{-Mean})$) shown
205 as Mean-log. We also computed standard deviation (SD), log-
206 transformed (SD-log) and coefficient of variation (CV).

207 Noticeably, any covariance matrix has in its diagonal
208 the variance ('var') of each variable. The sum of this diagonal,
209 also known as the 'trace' of the matrix, is equal to total
210 variance (Härdle and Simar, 2015) and holds in raw and log
211 transformed of waste fraction composition datasets. Therefore,
212 for each dataset (waste fraction compositions and log
213 transformed), we calculated the total variance and the
214 percentage thereof.

215 We also investigated the relationship between waste
216 fractions by applying Pearson's correlation analysis to raw and
217 log-transformed data for waste fraction compositions (in
218 percentage) and generation rates (kg waste fraction per capita
219 per week). However, this paper focuses mainly on the waste
220 fraction composition dataset.

221 **2.3 Compositional data analysis: isometric log-ratio**
222 **approach**

223 We applied statistical analysis to isometric log-ratio (ilr)

224 coordinates, computed based on the sequential binary partition
225 (SBP) (Egozcue et al., 2003). This approach transforms data
226 for waste fraction compositions into an unconstrained, real
227 dataset, thus enabling the use of classical statistics (Filzmoser
228 and Hron, 2008). This, for example, may mean that instead of
229 a dataset with a list of percentages that should always sum up
230 to 100 for each observation, the isometric log-ratio transforms
231 waste fraction composition data into a list of values that are
232 independent and should not sum up to a constant.

233 Similar to classical log transformation, the isometric log-
234 ratio requires that the data should not contain ‘zero values’.
235 For this study, a waste ‘zero value’ means that a household did
236 not generate any waste during this sampling week. Thus, we
237 assumed that zero values were due to the experimental design,
238 mainly the ‘time limit’ of the sampling campaign. For this
239 reason, zero values were replaced, using ‘imputation based on
240 the log-ratio expectation-maximisation (EM) algorithm’
241 (lrEM) in the zCompositions package (Palarea-Albaladejo and
242 Martín-Fernández, 2015), which comprises four steps: (1)
243 dataset selection, which can be the waste fraction composition
244 (percentage) or generation rate (kg waste fraction per capita
245 per week). For this study, we used the waste fraction
246 generation rate; nevertheless, the function lrEM is based on
247 compositional data analysis technique and therefore ensures
248 equivalent results regardless of datasets. (2) The descriptive

249 analysis of the zero values was performed using the function
250 zPattern in the zCompositions package. As a result, a graphical
251 representation of the relative frequencies of zero for each
252 waste fraction is provided. (3) Threshold (the detection limit)
253 values should be defined prior to zero replacement. A single
254 value for all waste fractions or varying values can be selected.
255 For this study, a single threshold value was set at 10 g, which
256 is the minimum weight of the weighing scale used for the
257 waste sampling campaign. (4) The new dataset contained non-
258 zero values. In practice, the function lrEM substitutes an
259 observation x with a value of zero by a random observation y
260 in the interval between zero and the threshold value (see
261 Palarea-Albaladejo and Martín-Fernández, 2015, for detailed
262 mathematics underpinning zCompositions).

263 Seven coordinates (ilr_1) were computed corresponding to
264 $D-1$ numbers of partitions. Here, D was eight, namely the
265 number of waste fractions shown in Table 1. The first ilr
266 coordinate was computed by dividing the eight fractions into
267 two groups: food waste and non-food waste. Subsequently,
268 each of the two groups was divided further until each group
269 was represented by one single waste fraction, as indicated in
270 Table 2, where (+1) refers to the group in the numerator, while
271 (-1) is the group appearing in the denominator.

272
273

Here (Table 2)

274

275 The ilr coordinates were computed based on the formulas
276 shown in Eqs. (1-7). Eq. (1) computed the coordinate (ilr₁)
277 between food waste and non-food waste. Eqs. (2-4) computed
278 the coordinates ilr₂ (vegetable versus animal food waste), ilr₃
279 (avoidable versus unavoidable vegetable food waste) and ilr₄
280 (avoidable versus unavoidable animal-derived food waste).
281 Furthermore, the coordinate ilr₅ (paper and metal versus plastic
282 and other) was calculated in Eq. (5), the coordinate ilr₆
283 between paper and metal was derived in Eq. (6) and the
284 coordinate ilr₇ between plastic and other in Eq. (7).

$$\begin{aligned} & \text{ilr}_1\{AV, UV, AA, UA\} \text{vs. } \{Pa, Me, Pl, Ot\} = \\ & \sqrt{\frac{4 \times 4}{4+4}} \text{LN} \frac{\sqrt[4]{AV \times UA \times AA \times UA}}{\sqrt[4]{Pa \times Me \times Pl \times Ot}} \\ & (1) \end{aligned}$$

$$\text{ilr}_2\{AV, UV\} \text{vs. } \{AA, UA\} = \sqrt{\frac{2 \times 2}{2+2}} \text{LN} \frac{\sqrt[3]{AV \times UV}}{\sqrt[3]{AA \times UA}} \quad (2)$$

$$\text{ilr}_3\{AV\} \text{vs. } \{UV\} = \sqrt{\frac{1 \times 1}{1+1}} \text{LN} \frac{\sqrt[3]{AV}}{\sqrt[3]{UV}} \quad (3)$$

$$\text{ilr}_4\{AA\} \text{vs. } \{UA\} = \sqrt{\frac{1 \times 1}{1+1}} \text{LN} \frac{\sqrt[3]{AA}}{\sqrt[3]{UA}} \quad (4)$$

$$\text{ilr}_5\{Pa, Me\} \text{vs. } \{Pl, Ot\} = \sqrt{\frac{2 \times 2}{2+2}} \text{LN} \frac{\sqrt[3]{Pa \times Me}}{\sqrt[3]{Pl \times Ot}} \quad (5)$$

$$\text{ilr}_6\{Pa\} \text{vs. } \{Me\} = \sqrt{\frac{1 \times 1}{1+1}} \text{LN} \frac{\sqrt[3]{Pa}}{\sqrt[3]{Me}} \quad (6)$$

$$\text{ilr}_7\{Pl\} \text{vs. } \{Ot\} = \sqrt{\frac{1 \times 1}{1+1}} \text{LN} \frac{\sqrt[3]{Pl}}{\sqrt[3]{Ot}} \quad (7)$$

293 Here, LN stands for the natural logarithm, and the other

294 abbreviations refer to the waste fractions presented in Table 1.
295 Pa refers to paper and board, Pl to plastic packaging, Me to
296 metal packaging and Ot to other.

297 The CoDa technique uses the geometric mean of the dataset,
298 which is the ‘back-transformed’ value of the ilr-arithmetic
299 mean and is calculated as follows:

$$300 \quad g_m(x) = [\prod_{i=1}^D x_i]^{1/D} = \exp\left[\frac{1}{D} \sum_{i=1}^D LN(x_i)\right] \quad (8)$$

301 where $g_m(x)$ is the geometric mean and D is the number of
302 waste fractions (x_i) involved. The natural logarithm is
303 abbreviated as $LN(x_i)$ and its inverse is abbreviated as $\exp(x_i)$.

304 The back transformation of the isometric log-ratio
305 coordinates is calculated simply by reversing the original
306 transformation (Egozcue et al., 2003). The general formula for
307 the back transformation of the isometric log-ratio coordinate
308 (ilr^{-1}) is provided as follows (Felipe et al., 2016):

$$309 \quad ilr^{-1} = C(\exp(x \cdot \psi)) \quad (9)$$

310 where ilr^{-1} is the back transformation, \mathbf{x} is the simulated value
311 for the transformation (ilr), ψ is the matrix constructed from
312 the sequential binary partition given in Eqs (1 to 7) and C is
313 the closure operation that provides a closed dataset.

314 Total variance ($totvar(\mathbf{x})$) is introduced to provide a global
315 measure of spread (Pawlowsky et al., 2008) and measures the
316 variation between individual waste fraction compositions
317 included in the dataset. Total variance is computed as:

318
$$totvar(\mathbf{x}) = \frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^D var\left(LN \frac{x_i}{x_j}\right) \quad (10)$$

319 The relationship between pairs of waste fractions is
 320 analysed by means of a variation array, calculated as:

321
$$A = \begin{bmatrix} 0 & v_{12} & \dots & v_{1D} \\ e_{21} & 0 & \dots & v_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ e_{D1} & e_{D2} & \dots & 0 \end{bmatrix} \quad (11) \text{ where,}$$

322
$$e_{ij} = E\left(\ln \frac{x_i}{x_j}\right) \quad (12) \quad \text{and} \quad v_{ij} = var\left(\ln \frac{x_i}{x_j}\right) \quad (13)$$

323 The variation array (Aitchison, 1986) was introduced to
 324 provide a solution to the problem of computing correlation
 325 coefficients for compositional data. We computed the variation
 326 array using both waste fraction compositions and generation
 327 rates.

328 **2.4 Software for data analysis**

329 First, the data were explored and zero values imputed
 330 using the R package ‘zCompositions’ (Palarea-Albaladejo and
 331 Martín-Fernández, 2015). The ilr coordinates and their back
 332 transformation, as well as variation array, were computed with
 333 CoDaPack (Thió-Henestrosa and Comas-Cufi, 2011).
 334 Thereafter, the most commonly used methods employed for
 335 describing and analysing waste data, such as mean, standard
 336 deviation, coefficients of variation and correlation tests
 337 (European Commission, 2004), were carried out in R (R Core
 338 Team, 2017). Among other packages implemented in R, the
 339 ‘StatDA’ (Filzmoser, 2015) software package was used for

340 plotting.

341

342 **3 Results**

343 **3.1 Exploration of data for waste fraction compositions**

344 Figure 1 displays the graphical output of the zero values
345 analysis. The columns show the analysis of zero values by
346 waste fraction. The data in Figure 1 can be grouped into two
347 parts. The first is a rectangle, containing squared boxes
348 coloured in dark grey, where waste fractions have zero values,
349 and light grey for non-zero values. The number of squared
350 boxes per column is the total combinations of zero values for
351 each household involved as a function of waste fraction. The
352 second is bar plots on the top (in dark grey), which show the
353 percentage frequency of zero values by waste fraction, whereas
354 bar plots on the right (in light grey) present the percentage
355 frequency of non-zero values for all possible combinations of
356 household and waste fractions. For example (see bar plots on
357 the top in dark grey), the percentage frequency of zero was
358 5.35% for avoidable vegetable food waste (see first column),
359 and 2.94% for unavoidable food waste (see second column).
360 Regarding bar plots on the right-hand side of the rectangle (in
361 light grey), 64.45% of observations (households) have non-
362 zero values for all waste fractions (first line), and 8.31% are
363 non-zero values, except for the avoidable animal derived-food
364 waste fraction.

365

366

367

Here (Figure 1)

368

369 Subsequently, the zero value detected was replaced prior

370 to computing the log-ratio coordinates and undertaking normal

371 log transformation. For example, the minimum values for the

372 four food waste fractions (zero values) were replaced by 5.7 g

373 for avoidable vegetable food waste, 5.8 g for unavoidable

374 vegetable food waste, 2.8 g for avoidable animal-derived food

375 waste and 1.6 g for unavoidable animal-derived food waste.

376 Note that here the replaced values are between zero and 10 g.

377 A comparison of the datasets before and after zero replacement

378 showed quite a similar distribution, demonstrating that the

379 distribution of the dataset is preserved despite containing many

380 zero values (SM Figure 1, SM Tables 2 and 3).

381 Figure 1 also presents a detailed overview of household

382 waste fraction generation patterns; for example, only 1.3% and

383 0.3% of the households did not generate plastic packaging or

384 paper, respectively. Noticeably, for vegetable food waste, only

385 5.2% and 2.9% of the households (see Figure 1, vertical bars)

386 did not generate AV and UV, respectively. On the other hand,

387 the percentage of households that did not generate animal-

388 derived food waste was 15.2% for AA and 14.6% for AU (see

389 Figure 1, vertical bars). These data indicate that vegetable food

390 waste occurred more often than animal-derived food in Danish
391 houses.

392

393 **3.2 Mean and standard deviation of waste fraction** 394 **compositions**

395 The distribution of the waste fraction compositions for all
396 households is shown in Figure 2. Asymmetry is evident in the
397 boxplot of each waste fraction, because the distance from the
398 median (horizontal bar in the rectangular box) to the fifth
399 percentiles (bottom horizontal bar (Figures 2 and 4) or vertical
400 bar on the left (Figure 3)) is smaller than the distance between
401 the median to the 95th percentiles (upper horizontal bar
402 (Figures 2 and 4) or vertical bar on the right (Figure 3)), as
403 shown in Figure 2. Thus, the data for each waste fraction were
404 positively skewed and also contained potential outliers, which
405 are defined as unusually large or small values in a sample of
406 observation (Wilcox, 2012). Here, outliers are shown in Figure
407 3 as circles above the upper horizontal bar, and these outliers
408 lead to bias in the arithmetic mean and inflate the standard
409 error. Thus, robust statistical techniques have been developed
410 to deal effectively with this problem, though these methods are
411 not included in this study.

412 A detailed analysis of vegetable food waste (AV and UV)
413 is provided in Figure 3 as an example. Figures 3a and 3b
414 illustrate a combined histogram and boxplot of waste fraction

415 composition and log transformation for avoidable vegetable
416 food waste, while Figures 3c and 3d represent unavoidable
417 vegetable food waste in the same regard. These figures reveal
418 asymmetric distribution despite log transformation.
419 Conversely, the ilr coordinates are distributed symmetrically
420 (see Figure 4).

421

422 **Here (Figure 2)**

423

424 **Here (Table 3)**

425

426 **Here (Figure 3)**

427

428 The arithmetic means (Mean) based on waste fraction
429 compositions sum up to 100, whereas the arithmetic means
430 based on log-transformed (Log-mean) data sum up to 14. As a
431 result, the means of the log-transformed data are difficult to
432 interpret and apply because of the change in scale (USEPA,
433 2006). This problem could be solved by Mean-log', which is
434 obtained by 'back transforming' the log-transformed mean
435 ($\text{Mean-log} = \exp(\text{Log-Mean-log})$). The arithmetic mean, log-
436 mean and mean-log were computed from an asymmetric
437 dataset, which led to biased parameter estimation and incorrect
438 results (Reimann et al., 2008; Wilcox, 2012).

439 On the contrary, the 'Mean-ilr' (mean based on isometric log-

440 ratio coordinates) (see Table 3) was computed from
441 symmetrical data, thus suggesting that the log-ratio coordinates
442 enable a data analyst to obtain symmetric distribution of data,
443 as shown in Figure 4. Importantly, while log-ratio
444 transformation enables one to remove the constant sum
445 constraint, the 'Mean-ilr' for waste fractions sums up to 100.
446 Similar to classical statistics, robust methods have been
447 developed for the statistical analysis of compositional data
448 (Templ et al., 2011), though these methods are not included in
449 this study.

450

451 **Here (Figure 4)**

452

453 The standard deviation, total variance and percentage of
454 variance estimates were calculated and are shown in Table 4.
455 The results indicate that the standard deviation values for the
456 raw waste fraction composition are very high compared to
457 their corresponding arithmetic mean (Mean in Table 3). In
458 particular, the standard deviation of animal-derived food waste
459 (AA and AV) and metal packaging are higher or equal to the
460 corresponding arithmetic mean, thereby generating very high
461 variation value coefficients (e.g. 155% for metal packaging,
462 141% for unavoidable animal-derived food waste, 99% for
463 avoidable animal-derived food waste). The resulting
464 confidence intervals (Mean \pm 2* SD) were (-6.78; 20.74) and

465 (-4.03; 8.45) for AA and AV, respectively, including negative
466 percentages. These results highlight some of the pitfalls of
467 computing standard deviations for waste fraction
468 compositions. Moreover, the estimated percentages of
469 variances for waste fractions varied when the raw dataset for
470 waste fraction compositions (% Var) was log-transformed (%
471 Var-log). The highest variance percentages were found for the
472 fractions other (% Var= 31.43%) and avoidable animal-derived
473 food waste (% Var-log=33.24%) in raw and log-transformed
474 datasets, respectively. On the other hand, the lowest variance
475 percentages were found for unavoidable animal-derived food
476 waste (% Var=1.47%) and other (% Var-log=2.74%) in the raw
477 and log-transformed datasets, correspondingly. These
478 incoherent results indicate that while log transformation could
479 indeed help to achieve normality, the calculated variance
480 becomes impossible to compare after transformation, as
481 demonstrated by Filzmoser et al. (2009).

482 Overall, it is questionable whether standard deviation
483 values are informative in the case of most sets of waste
484 composition data, due to the dual issues of non-normality and
485 the constant-sum constraint. First, the standard deviation
486 ignores the compositional nature of waste fraction composition
487 data (composition of waste fractions should add up to 100).
488 Second, most coefficients of variation (CV %) provided in
489 Table 4 are extremely high, thus restricting their application in

490 environmental modelling (Ciroth et al., 2013). As a solution,
491 total variance (see Eq. 9) that measures overall data
492 homogeneity (or variation) can be calculated (Pawlowsky et
493 al., 2008). Here, total variance expresses variation in the
494 dataset for each waste fraction. Thus, the contribution of each
495 waste fraction to total variation is provided in percentage terms
496 (clr-Var %), as shown in Table 4.

497

498 **Here (Table 4)**

499

500 Based on the compositional data analysis technique, total
501 variance (totvar) from Eq. (9) amounted to 9.25, as shown in
502 Table 4. The waste fraction contributing to the highest
503 variation in the dataset was avoidable animal-derived food
504 waste (24.73%), followed by unavoidable animal-derived food
505 waste (18.84%) and metal packaging (14.81%), suggesting that
506 the generation of these fractions by Danish households varies
507 substantially.

508 On the other hand, paper (5.27%) and plastic packaging
509 (5.53%) made a small contribution to total variance. A possible
510 interpretation for this finding could be that metal packaging
511 materials are source-sorted by a wider variety of households
512 than paper and plastic packaging, and therefore they do not
513 vary much in the fraction that ends up in residual household
514 waste bins. However, a characterisation of total household

515 waste including source-segregated waste (e.g. paper, metal,
516 plastic) could provide a better interpretation of these results,
517 thereby demonstrating that total variance enables the analyst to
518 compare systematically variations among waste fraction
519 compositions, which is difficult for classical standard deviation
520 and coefficient of variation estimates.

521 **3.3 Relationship between waste fractions: Pearson's** 522 **correlation test**

523 Table 5 presents the pairwise correlation coefficients
524 between waste fractions, computed using datasets of (1)
525 percentage composition (Percentage %) and (2) generation
526 rates (kg/capita/week). A negative correlation coefficient
527 between waste fractions means an inverse relationship,
528 whereas a positive correlation coefficient means these fractions
529 vary in the same direction (when the value of one waste
530 fraction increases, the value of the other fraction increases too,
531 and vice versa). Moreover, while a correlation coefficient of
532 value ± 0.5 shows a strong relationship between the two waste
533 fractions, a value of 1 means a perfect correlation. A
534 correlation coefficient is statistically significant when the p-
535 value is less than 0.5.

536

537 **Here (Table 5)**

538

539 Based on the waste fraction generation rates, we found a

540 positive and significant correlation coefficient between ‘Other’
541 and the seven remaining waste fractions, as shown in Table 5.
542 In contrast, we found negative and significant correlation
543 coefficients between these fractions when the Pearson’s
544 correlation test was applied to waste fraction compositions
545 (Percentage %).

546 Figure 5 illustrates the results of the correlation test
547 applied to waste fraction composition data. Figure 5 shows that
548 the Pearson’s correlation test applied to the waste fraction
549 generation dataset provided a positive correlation coefficient
550 between avoidable food waste (UA, UV, AA and AV) and
551 plastic packaging. These results are consistent with those of
552 Williams et al. (2012), suggesting that a reduction in plastic
553 packaging materials may lead to a reduction in avoidable
554 vegetable food waste. In contrast, the results of the Pearson’s
555 correlation applied to the waste fraction compositions dataset
556 showed a negative correlation between the same waste
557 fractions, except for UA. These results are in good agreement
558 with those obtained by Alter (1989), and similar results were
559 obtained when the Pearson’s correlation test was applied to
560 log-transformed data. Note here that the signs and the values of
561 the correlation coefficients depend on the datasets, even
562 though a Pearson’s correlation test was applied to log-
563 transformed data (SM Table 1). These results pose an
564 interpretation dilemma. First, a reduction in plastic packaging

565 may contribute to food waste reduction, due to the positive
566 correlation between these waste fractions, although, on the
567 other hand, an increase in the use of plastic packaging may
568 contribute to a reduction in household food waste because of
569 the negative correlation coefficient. Moreover, while these
570 correlation coefficients were statistically significant, their
571 estimates were somewhat different (see Figure 4 and Table 5).

572

573 **Here (Figure 5)**

574

575 **3.4 Variation array with CoDa**

576 The variation array was computed using Eq. (10) and is
577 shown in Table 6. Note that the same variation array was
578 obtained when using either the waste fractions generation rates
579 (kg/capita/week) or waste fraction compositions (percentage
580 %), and therefore the relationship between waste fractions is
581 interpreted independently of waste datasets.

582 The variation array is divided into two triangles. The
583 upper triangle shows ratios or proportionalities between waste
584 fractions as pairwise log-ratio variances (variance $\ln(X_i/X_j)$
585 (see Eq. (12)). The lower triangle presents the pairwise log-
586 ratio means (Mean $\ln(X_j/X_i)$ (see Eq. (13)). Here, the
587 numerator is denoted by columns (X_i), and denominator (X_j) is
588 illustrated by rows. Moreover, the sign (+ or -) of the log-ratio
589 mean values indicates the direction of the ratio between the

590 relevant fractions.

591

592 **Here (Table 6)**

593

594 Log-ratio variance values highlighted in grey (the value is
595 close to zero) indicate an almost constant ratio, whereas log-
596 ratio variance values in bold and highlighted in grey (usually
597 value is closed to zero) can be assumed to be zero, suggesting
598 an absolutely constant ratio (Pawlowsky-Glahn et al., 2015).
599 On the other hand, log-ratio variance values that are very much
600 higher than zero are highlighted in red, and these indicate no
601 relationship between the two relevant fractions, because their
602 ratios vary significantly.

603 For example, the mean log-ratio between plastic
604 packaging and paper and board was negative $\{(mean$
605 $(\log(Plastic/Paper)) = -1.4)\}$ (here, *Plastic* is X_j from a row in
606 Table 6 and *Paper* is X_i from a column in Table 6), indicating
607 that the households placed more mass of plastic packaging
608 than paper and board waste into their residual waste bins.
609 Furthermore, the variance in their log-ratio is small (0.77),
610 suggesting a strong relationship between these fractions. This
611 relationship has a negative ratio, which can be calculated as
612 follows:

613 $plastic/paper = \exp(-1.4) = 0.25$

614 This result suggests that the ratio between discarded (1) plastic

615 and (2) paper and board in residual household waste is constant
616 and estimated at 0.25. This information could be used for
617 future developments in waste generation, i.e. to identify the
618 effects of new regulations and policies addressing packaging
619 materials.

620 The results shown in Table 6 indicate that the mean log-
621 ratio between avoidable animal-derived food waste and
622 unavoidable vegetable food waste was negative (-1.35).
623 However, the variance in their log-ratio was high (4.21),
624 thereby suggesting that the compositions of these fractions are
625 not proportional. In this case, the ratio between these fractions
626 is not constant.

627 Overall, the compositions of these pairs of waste fractions
628 are highly dependent: (1) unavoidable vegetable food waste
629 (UV) and paper (Paper), (2) paper (Paper) and plastic
630 packaging (Plastic) and (3) plastic packaging (Plastic) and
631 other waste fractions (Other). However, no relationship
632 between avoidable food waste fractions (AV and AA) and
633 material packaging (paper, plastic and metal) was identified.
634 For example, from the results in Table 7, it is apparent that the
635 ratio between avoidable animal-derived food waste and
636 packaging materials (plastic, paper and metal) is highly
637 variable (very high log-ratio variance painted in red).
638 Similarly, the ratio between avoidable vegetable food waste
639 and packaging materials (plastic, paper and metal) is not

640 constant. These values indicate no constant ratios between
641 these fractions, signifying that there is no relationship between
642 these fractions based on the analysis of residual waste taken
643 from the 779 households.

644

645 **4. Discussion**

646 From the data in Table 3, arithmetic means of waste
647 fractions composition were influenced by the fact that the
648 assumption of normal distribution was violated (see Figure 4).
649 These results are consistent with previously published studies,
650 which concluded that the arithmetic mean is an inappropriate
651 means of estimating central values of compositional data
652 (Filzmoser et al., 2009; Pawlowsky-Glahn et al., 2015; van den
653 Boogaart et al., 2013). Consequently, any evaluation (e.g.
654 prevention, reduction or recycling of waste) or modelling (e.g.
655 life cycle assessment) based on the arithmetic mean of waste
656 fraction composition may lead to potentially wrong
657 conclusions, because they are based on erroneous estimates.
658 While the log transformation of waste composition may help
659 solve the problem of normality, its value is limited because it
660 relies on a univariate method, which ignores that the
661 compositions of waste fractions account for the limited data,
662 i.e. from 0 to 100.

663 The results from the variation array were not in agreement
664 with those from the Pearson's correlation tests applied to both

665 raw and log-transformed data. The correlation test applied to
666 waste fraction generation rates provided positive correlation
667 coefficients. On the other hand, negative correlation
668 coefficients were obtained when the correlation analysis was
669 applied to the composition of waste fractions in percentage
670 terms. The positive correlation coefficients were due to the size
671 of the mass effect of waste fractions (kg/capita/week),
672 explaining why most waste fractions are positively and
673 significantly correlated with each other. The size effect of mass
674 was removed by calculating the correlation coefficient based
675 on the percentage composition of waste fractions. This then
676 generated negative correlation coefficients because of the
677 constant sum constraint (Aitchison, 1986; Pearson, 1897). As a
678 solution, the relationship between food waste fractions and
679 material packaging can be evaluated by the variation array
680 through a compositional data analysis technique. Log-ratio
681 coordinates remove the constant sum constraint and enable the
682 determination of the relationship between waste fractions,
683 independently of the unit. Another advantage of the variation
684 array is that the pairwise ratio between waste fractions could
685 be back-transformed to a desired unit and adequately
686 interpreted while preserving the structure of the original data
687 (Filzmoser and Hron, 2008; Pawlowsky-Glahn et al., 2015).
688 The advantage in this approach is that the variation array of
689 both waste datasets (percentage composition and mass per

690 waste fraction per household) generates the same results
691 because of the log-ratio transformation.

692 Computing the arithmetic mean (mean-ilr), total variance
693 and variance array based on CoDa technique is a not
694 straightforward undertaking. However, numerous tools that do
695 not require advanced programming skills are freely available
696 (Templ et al., 2011; Thió-Henestrosa and Comas-Cufi, 2011;
697 van den Boogaart, 2008). Therefore, we urge practitioners and
698 researchers within solid waste management to address
699 adequately the constant sum constraint problem when
700 analysing solid waste composition data (Filzmoser et al.,
701 2009).

702

703 **5. Conclusions**

704 This study is a first attempt to address the problem
705 associated with the statistical analysis of waste fraction
706 composition data. Based on a systematic comparison of the
707 arithmetic mean and standard deviation applied to waste
708 fraction composition data, it was demonstrated that these
709 statistical parameters may generate erroneous and misleading
710 results when applied to fractional percentages (i.e. percentage
711 of paper, board, food waste, etc.). Moreover, correlation
712 coefficients based on raw or general transformation of data
713 depend strongly on the type of waste dataset. As a solution,
714 isometric log-ratio coordinates approximate the symmetrical

715 distribution of data and remove the total constant sum
716 constraint, which restricts the application of classical statistics
717 to waste fraction composition. As a result, statistical analysis
718 applied to log-ratio coordinates generates consistent results
719 independently of the selected data type. The arithmetic means
720 of waste fractions, based on the isometric log-ratio, summed
721 up to 100. The variation array provides a ratio between waste
722 fractions and offers consistent conclusions regardless of the
723 data type.

724

725 **Acknowledgments**

726 The authors acknowledge the Danish Strategic Research
727 Council for financing this study via the IRMAR (Integrated
728 Resource Management & Recovery) Project (No. 11-116775).
729 We wish to thank Compositional Data Analysis and Related
730 methods (CoDa- RETOS) for their support through the project
731 MINISTERIO ECONOMÍA Y COMPETIVIDAD (Ref:
732 MTM2015-65016-C2-1-R (MINECO/FEDER, UE). The
733 Danish Environmental Protection Agency (EPA) and Claus
734 Petersen from ECONET AS are also acknowledged for their
735 valuable support and contributions.

736

737 **References**

738 Aitchison, J., 1994. A Concise Guide to Compositional Data

- 739 Analysis. Lect. Notes-Monograph Ser. 24, 73–81.
- 740 Aitchison, J., 1986. The Statistical Analysis of Compositional
741 Data. Monographs on Statistics and Applied Probability.
742 Chapman & Hall Ltd., London (UK). Reprinted (2003)
743 with additional material by The Blackburn Press,
744 Caldwell, NJ.
- 745 Alter, H., 1989. The Origins of Municipal Solid Waste: the
746 Relations Between Residues From Packaging Materials
747 and Food. *Waste Management* 7, 103–114.
748 doi:10.1177/0734242X8900700115.
- 749 Buccianti, A., Pawlowsky-Glahn, V., 2011. Compositional
750 Data Analysis, *Compositional Data Analysis: Theory and
751 Applications*. John Wiley & Sons, Ltd, Chichester, UK.
752 doi:10.1002/9781119976462.
- 753 Chang, N.-B., Davila, E., 2008. Municipal solid waste
754 characterizations and management strategies for the Lower
755 Rio Grande Valley, Texas. *Waste Management* 28, 776–
756 94. doi:10.1016/j.wasman.2007.04.002.
- 757 Christensen, T.H., 2011. Introduction to Waste Management. in
758 *Solid Waste Technology and Management* 3–16, in:
759 Christensen, T.H. (Ed.), *Solid Waste Technology &
760 Management*, Volume 1 & 2. Chichester, UK.
761 doi:10.1002/9780470666883.ch1.
- 762 Citroth, A., Muller, S., Weidema, B., Lesage, P., 2013.
763 Empirically based uncertainty factors for the pedigree
764 matrix in ecoinvent. *International Journal of Life Cycle
765 Assessment* 1–11. doi:10.1007/s11367-013-0670-5.
- 766 Dahlén, L., Vukicevic, S., Meijer, J.-E., Lagerkvist, A., 2007.
767 Comparison of different collection systems for sorted
768 household waste in Sweden. *Waste Management* 27,
769 1298–305. doi:10.1016/j.wasman.2006.06.016.
- 770 Edjabou, M.E., Boldrin, A., Scheutz, C., Astrup, T.F., 2015.
771 Source segregation of food waste in office areas: Factors
772 affecting waste generation rates and quality. *Waste
773 Management* 46, 94–102.
774 doi:10.1016/j.wasman.2015.07.013.
- 775 Edjabou, M.E., Petersen, C., Scheutz, C., Astrup, T.F., 2016.
776 Food waste generation and composition in Danish
777 households. *Waste Management* 52, 256–268.
778 doi:10.1007/s13398-014-0173-7.2.
- 779 Egozcue, J.J., Pawlowsky Glahn, V., Mateu-Figueras, G.,
780 Barceló Vidal, C., 2003. Isometric logratio for
781 compositional data analysis. *Mathematical Geosciences*
782 35, 279–300. doi:10.1023/A:1023818214614.

- 783 European Commission, 2004. Methodology for the Analysis of
784 Solid Waste (SWA-tool). User Version 43, 1–57.
- 785 Filzmoser, P., Hron, K., 2008. Correlation Analysis for
786 Compositional Data. *Mathematical Geosciences*. 41, 905–
787 919. doi:10.1007/s11004-008-9196-y.
- 788 Filzmoser, P., Hron, K., Reimann, C., 2009. Univariate
789 statistical analysis of environmental (compositional) data:
790 Problems and possibilities. *Science of the Total
791 Environment* 407, 6100–6108.
792 doi:10.1016/j.scitotenv.2009.08.008.
- 793 Peter Filzmoser (2015). *StatDA: Statistical Analysis for
794 Environmental Data*. R package version 1.6.9.
795 <https://CRAN.R-project.org/package=StatDA>.
- 796 Hanc, A., Novak, P., Dvorak, M., Habart, J., Svehla, P., 2011.
797 Composition and parameters of household bio-waste in
798 four seasons. *Waste Management* 31, 1450–60.
799 doi:10.1016/j.wasman.2011.02.016.
- 800 Härdle, W., Simar, L., 2015. Applied Multivariate Statistical
801 Analysis Applied Multivariate Statistical Analysis,
802 Technometrics. doi:10.1198/tech.2005.s319
- 803 Lagerkvist, A., Ecke, H., Christensen, T.H., 2011. Waste
804 Generation and Characterization Waste Characterization :
805 Approaches and Methods, in: Christensen, T.H. (Ed.),
806 Solid Waste Technology & Management, Volume 1 & 2.
807 Chichester, UK. doi:10.1002/9780470666883.ch5.
- 808 Milke, M.W., Wong, V., McBean, E. a., 2008. Describing
809 variability of MSW composition data with the log-logistic
810 distribution. *Waste Management Res.* 26, 355–361.
811 doi:10.1177/0734242X08089464.
- 812 Naveen, B.P., Mahapatra, D.M., Sitharam, T.G., Sivapullaiah,
813 P.V., Ramachandra, T.V., 2016. Physico-chemical and
814 biological characterization of urban municipal landfill
815 leachate. *Environ. Pollut.*
816 doi:10.1016/j.envpol.2016.09.002.
- 817 Palarea-Albaladejo, J., Martín-Fernández, J.A., 2015.
818 ZCompositions - R package for multivariate imputation of
819 left-censored data under a compositional approach.
820 *Chemometrics and Intelligent Laboratory Systems*. 143,
821 85–96. doi:10.1016/j.chemolab.2015.02.019.
- 822 Pawlowsky-Glahn, V., Egozcue, J., Tolosana-Delgado, R.,
823 2015. *Modeling and Analysis of Compositional Data*. John
824 Wiley & Sons.
- 825 Pawlowsky, G. V, Egozcue, J., Delgado, R.T., 2008. Lecture

- 826 notes on compositional data analysis 1–100.
- 827 Pearson, K., 1897. Mathematical Contributions to the Theory of
828 Evolution. - On a Form of Spurious Correlation Which
829 May Arise When Indices Are Used in the Measurement of
830 Organs. Proc. R. Soc. London 60, 489–498.
831 doi:10.1098/rspl.1896.0076.
- 832 R Core Team, 2017. R: A Language and Environment for
833 Statistical Computing.
- 834 Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008.
835 Statistical Data Analysis Explained: Applied
836 Environmental Statistics With R, Statistical Data Analysis
837 Explained: Applied Environmental Statistics With R. John
838 Wiley & Sons, Ltd, Chichester, UK.
839 doi:10.1002/9780470987605.
- 840 Templ, M., Hron, K., Filzmoser, P., 2011. robCompositions: an
841 R-package for robust statistical analysis of compositional
842 data, in: Pawlowsky-Glahn, V., Buccianti, A. (Eds.),
843 Compositional Data Analysis. Theory and Applications.
844 John Wiley & Sons, Chichester (UK), pp. 341–355.
- 845 Thió-Henestrosa, S., Comas-Cufi, M., 2011. CoDaPack v2
846 USER ' s GUIDE.
- 847 USEPA, 2006. Data Quality Assessment: Statistical Methods
848 for Practitioners, U.S. Envir. ed. U.S. Environmental
849 Protection Agency, Washigton, D.C.
- 850 van den Boogaart, K.G., 2008. “compositions”: A unified R
851 package to analyze compositional data. Comput. Geosci.
852 34, 320–338.
- 853 van den Boogaart, K.G., Tolosana-Delgado, R., Boogaart, K.G.
854 Van Den, Tolosana-Delgado, R., 2013. Analyzing
855 Compositional Data with R. Springer Berlin Heidelberg,
856 Berlin, Heidelberg. doi:10.1007/978-3-642-36809-7.
- 857 Wilcox, R., 2012. Introduction to Robust Estimation and
858 Hypothesis Testing, Introduction to Robust Estimation and
859 Hypothesis Testing. Elsevier. doi:10.1016/B978-0-12-
860 386983-8.00008-1.
- 861 Williams, H., Wikström, F., Otterbring, T., Löfgren, M.,
862 Gustafsson, A., 2012. Reasons for household food waste
863 with special attention to packaging. Journal of Cleaner
864 Production. 24, 141–148.
865 doi:10.1016/j.jclepro.2011.11.044.
- 866 Xu, C., Gertner, G.Z., 2008. Uncertainty and sensitivity
867 analysis for models with correlated parameters. Reliability
868 Engineering and System Safety 93, 1563–1573.

869 doi:10.1016/j.ress.2007.06.003.

870

17

18 **Table 1: List of residual waste fractions and components**
 19 **included**

| Waste fractions | Components |
|--|---|
| Avoidable vegetable food waste (AV ¹) | Cooked food (e.g. rice, pasta, potatoes, etc.) Fresh fruit, fresh carrots and potatoes, bread, cereals |
| Avoidable animal-derived food waste (AA ¹) | Cooked eggs, rest of food containing meat, fish, etc. Canned meat and fish, |
| Unavoidable vegetable food waste (UV ¹) | Residues from fruits, vegetables, coffee grounds Eggs not cooked, dairy products, not cooked meat and fish, etc. |
| Unavoidable animal-derived food waste (UA ¹) | Leftovers containing meat, fish, skins and bones, etc. Cheese rinds, eggs shells, other non-edible mixed animal and vegetable products |
| Paper and board (Paper:Pa ¹) | Advertisements , Books & booklets, Magazines & Journals, Newspapers Office paper, Phonebooks, Miscellaneous paper, Corrugated boxes Beverage cartons, Folding boxes, Miscellaneous board |
| Plastic packaging (Plastic:PI ¹) | Packaging plastics, such as PET/PETE, HDPE, PVC/V , LDPE/LLDPE, PP, PS, others, etc |
| Metal packaging (Metal;Me ¹) | Metal packaging containers (ferrous and non-ferrous) Composites |
| Others (Ot ¹) | Gardening waste, glass packaging, other/special glass, Table and kitchen ware glass, Non-packaging metals Non-packaging plastic, plastic film Miscellaneous combustible waste, inert (other non-combustible), special waste |

20 ¹ Refers to abbreviation of waste fractions in equations and
 21 figures and other tables in the present paper

22

23

24 **Table 2:** Signs code of the sequential binary partition applied
 25 to the residual household waste fractions: Balance code, (+1)
 26 means that the fraction is assigned to the first group
 27 (numerator), (-1) to the second group, and 0 the fraction is not
 28 included in the partition in this balance

| Coordinates | Residual household waste fractions | | | | | | | |
|------------------|------------------------------------|-----------------|-----------------|-----------------|--------------------|--------------------|----------------------|--------------------|
| | AV ^a | UV ^b | AA ^c | UA ^d | Paper ^e | Metal ^f | Plastic ^g | Other ^h |
| ilr ₁ | +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| Ilr ₂ | +1 | +1 | -1 | -1 | 0 | 0 | 0 | 0 |
| Ilr ₃ | +1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ilr ₄ | 0 | 0 | +1 | -1 | 0 | 0 | 0 | 0 |
| Ilr ₅ | 0 | 0 | 0 | 0 | +1 | +1 | -1 | -1 |
| Ilr ₆ | 0 | 0 | 0 | 0 | +1 | -1 | 0 | 0 |
| Ilr ₇ | 0 | 0 | 0 | 0 | 0 | 0 | +1 | -1 |

29 ^aAvoidable vegetable food waste

30 ^bUnavoidable vegetable food waste

31 ^cAvoidable animal-derived food waste

32 ^dUnavoidable animal-derived food waste

33 ^ePaper and board; ^fMetal packagin.; ^gPlastic packaging; ^hgrouped waste
 34 fraction (see Table 1 for waste fractions)

35 .

36

37 Table 3: Comparison of arithmetic means computed based on
 38 raw data (Mean), log transformed data (Log-Mean), back-
 39 transformed data (Mean-log) and back-transformed isometric
 40 log-ratio mean (Mean-ilr)

| Waste fractions | Classical statistics | | | | CoDa-technique |
|---------------------------------------|----------------------|-----------------------|-----------------------|--------|-----------------------|
| | Mean ^a | Log-mean ^b | Mean-log ^c | Median | Mean-ilr ^d |
| Avoidable vegetable food waste | 15.57 | 2.32 | 10.14 | 13.84 | 13.3 |
| Unavoidable vegetable food waste | 17.03 | 2.47 | 11.87 | 15.22 | 15.5 |
| Avoidable animal-derived food waste | 6.98 | 1.13 | 3.09 | 5.11 | 4.0 |
| Unavoidable animal-derived food waste | 2.21 | -0.06 | 0.94 | 1.08 | 1.2 |
| Paper and board | 20.79 | 2.91 | 18.28 | 18.52 | 23.9 |
| Metal packaging | 2.12 | 0.09 | 1.09 | 1.44 | 1.4 |
| Plastic packaging | 5.51 | 1.50 | 4.49 | 4.76 | 5.9 |
| Other | 29.80 | 3.28 | 26.59 | 26.30 | 34.8 |
| Total | 100.00 | 13.63 | 76.49 | 86.27 | 100.0 |
| Wet waste kg per household per week | 10.41 | | 8.80 | 9.52 | |
| Wet waste kg per person per week | 4.00 | | 3.45 | 3.42 | |

41 ^aArithmetic mean from raw data,

42 ^bArithmetic mean for log-transformed data;

43 ^cArithmetic mean based on back-transformation of the log-transformed data;

44 ^dArithmetic mean for ilr coordinates, which is back-transformed

45

46

47 Table 4 Comparison of standard deviation values based on
 48 waste fraction compositions data set (SD) and variance (%
 49 Var); log-transformed (SD-log) and variance of log-
 50 transformed (% Var-log) absolute contribution of each waste
 51 fractions (SD-clr) to total variance, and the percentage
 52 distribution of the total variance (SD-clr) (n=779)

| Waste fractions | Classical statistics | | | | CoDa-technique | |
|------------------------------------|----------------------|--------|--------|-----------|----------------|-----------|
| | SD | % Var | SD-log | % Var-log | SD-clr | % Var-clr |
| Avoidablevegetablefoodwaste | 10.76 | 17.52 | 3.49 | 12.55 | 1.1 | 13.16 |
| Unavoidablevegetablefoodwaste | 11.51 | 20.05 | 2.99 | 9.21 | 1.03 | 11.56 |
| Avoidableanimal-derivedfoodwaste | 6.88 | 7.16 | 5.68 | 33.24 | 1.51 | 24.73 |
| Unavoidableanimal-derivedfoodwaste | 3.12 | 1.47 | 4.46 | 20.5 | 1.32 | 18.84 |
| Paperandboard | 10.9 | 17.98 | 1.68 | 2.91 | 0.7 | 5.27 |
| Metalpackaging | 3.29 | 1.64 | 3.76 | 14.57 | 1.17 | 14.81 |
| Plasticpackaging | 4.26 | 2.75 | 2.04 | 4.29 | 0.72 | 5.53 |
| Other | 14.41 | 31.43 | 1.63 | 2.74 | 0.75 | 6.09 |
| Totalvariance(totvar) | 660.76 | 100.00 | 97.05 | 100.00 | 9.23 | 100.00 |

53

54 Table 5 Correlation matrix from Pearson correlation test and
 55 significance levels of raw data shown in Figure 2 (r: range:-
 56 1.00 to +1.00)

| Waste fractions | AV ^d | UV ^e | AA ^f | UA ^g | Paper ^h | Metal ⁱ | Plastic ^j | Other | Datasets |
|--|-----------------|-----------------|-----------------|-----------------|--------------------|--------------------|----------------------|-------|--------------------------------|
| Avoidable vegetable food waste (AV) | 1.00 | *** | *** | *** | *** | . | *** | *** | Percentage % kg/capita/week |
| Unavoidable vegetable food waste (UV) | -0.17 | 1.00 | *** | 0.00 | *** | * | ** | *** | Percentage % kg/capita/week |
| Avoidable animal-derived food waste (AA) | 0.16 | -0.19 | 1.00 | 0.00 | *** | 0.00 | 0.00 | *** | Percentage % kg/capita/week |
| Unavoidable animal-derived food Waste (UA) | -0.12 | 0.02 | 0.00 | 1.00 | . | 0.00 | 0.00 | ** | Percentage % kg/capita/week |
| Paper and board | -0.30 | -0.16 | -0.21 | -0.06 | 1.00 | * | 0.00 | *** | Percentage % kg/capita/week |
| Metal packaging | -0.07 | -0.08 | -0.03 | 0.03 | -0.09 | 1.00 | 0.00 | 0.00 | Percentage % kg/capita/week |
| Plastic packaging | -0.13 | -0.10 | -0.05 | 0.03 | -0.04 | 0.05 | 1.00 | * | Percentage % kg/capita/week |
| Other | -0.38 | -0.41 | -0.27 | -0.10 | -0.26 | -0.06 | -0.08 | 1.00 | Percentage % kg/capita/week |
| | 0.30 | 0.15 | 0.21 | 0.07 | 0.28 | 0.14 | 0.14 | 1.00 | kg/capita/week |

57 ***Very high significance probability higher than 0.001
 58 **High significance probability between 0.001 and 0.01
 59 *Significance probability between 0.01 and 0.05
 60 0.00 no significance-probability higher than 0.05
 61 ^a amount of waste (wet basis) per household per week
 62 ^b amount of waste (wet basis) per person per week
 63 ^c Composition of residual household waste on wet basis.
 64 ^dAvoidable vegetable food waste
 65 ^eUnavoidable vegetable food waste
 66 ^fAvoidable animal-derived food waste
 67 ^gUnavoidable animal-derived food waste
 68 ^hPaper; ⁱMetal packaging.; ^jPlastic packaging; ^kOther (see Table 1).

69 Table 6: Variation array of waste fraction compositions
 70 computed using log-ratio transformation of the waste dataset
 71 shown in Figure 2

| Waste fractions | Variance ln(Xi/Xj) | | | | | | | | |
|----------------------|--------------------|-----------------|-----------------|-----------------|--------------------|--------------------|----------------------|--------------------|--|
| | AV ^a | UV ^b | AA ^c | UA ^d | Paper ^e | Metal ^f | Plastic ^g | Other ^h | |
| AV ^a | | 2.53 | 3.11 | 3.83 | 2.10 | 3.09 | 2.15 | 2.18 | |
| UV ^b | 0.16 | | 4.21 | 3.00 | 1.52 | 2.93 | 1.77 | 1.83 | |
| AA ^c | -1.19 | -1.35 | | 5.14 | 3.54 | 4.49 | 3.43 | 3.62 | |
| UA ^d | -2.38 | -2.54 | -1.19 | | 2.49 | 3.63 | 2.50 | 2.61 | |
| Paper ^e | 0.59 | 0.43 | 1.78 | 2.97 | | 2.08 | 0.77 | 0.64 | |
| Metal ^f | -2.23 | -2.39 | -1.04 | 0.15 | -2.82 | | 1.92 | 2.07 | |
| Plastic ^g | -0.81 | -0.97 | 0.37 | 1.57 | -1.40 | 1.41 | | 0.80 | |
| Other ^h | 0.96 | 0.81 | 2.15 | 3.34 | 0.37 | 3.19 | 1.78 | | |
| | Mean ln(Xj/Xi) | | | | | | | Total variance | |

72 ^aAvoidable vegetable food waste
 73 ^bUnavoidable vegetable food waste
 74 ^cAvoidable animal-derived food waste
 75 ^dUnavoidable animal-derived food waste
 76 ^ePaper and board; ^fMetal packaging.;
 77 ^gPlastic packaging;
 78 ^hgrouped waste fraction (see Table 1 for waste fractions)

80

Figure 1

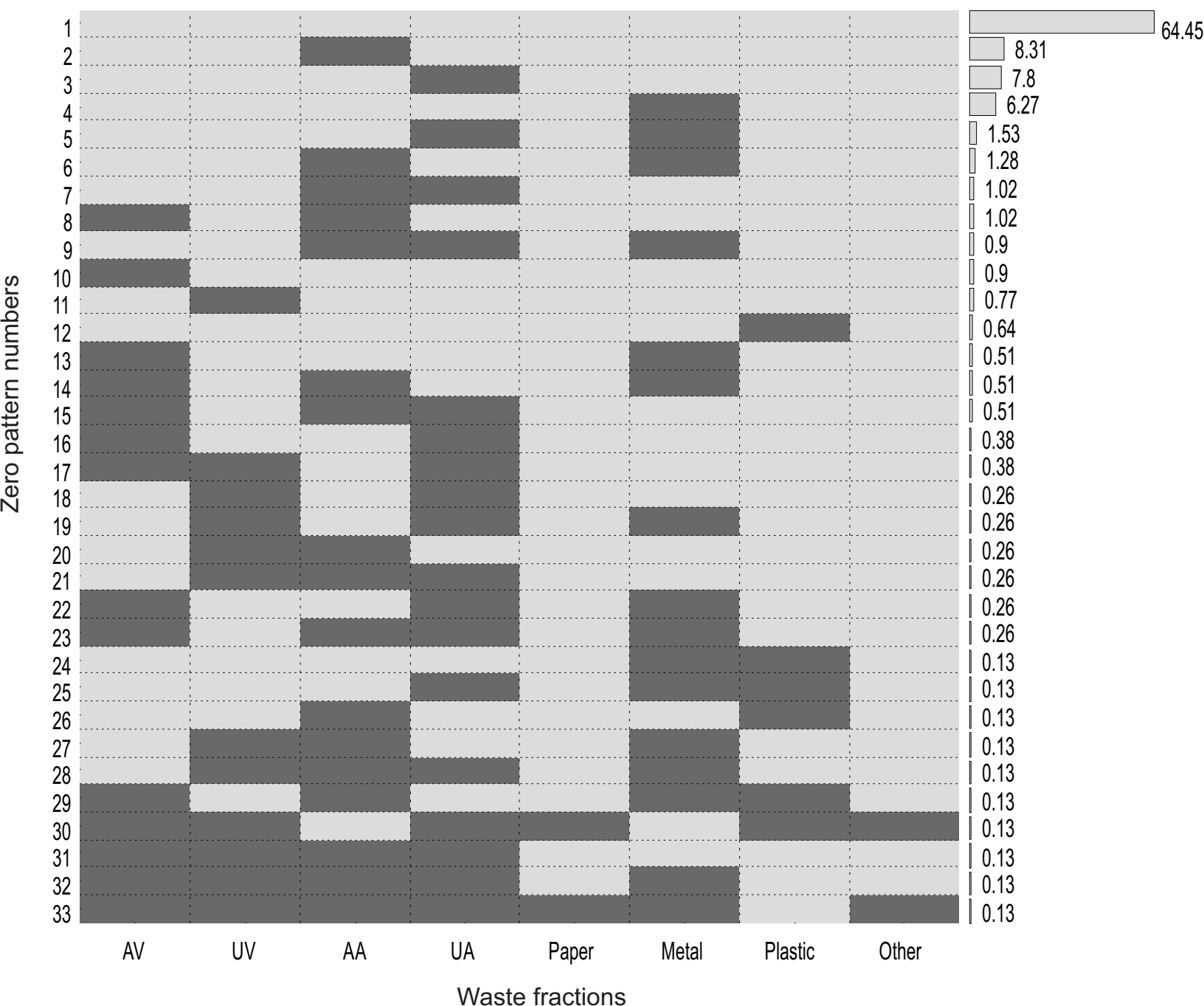
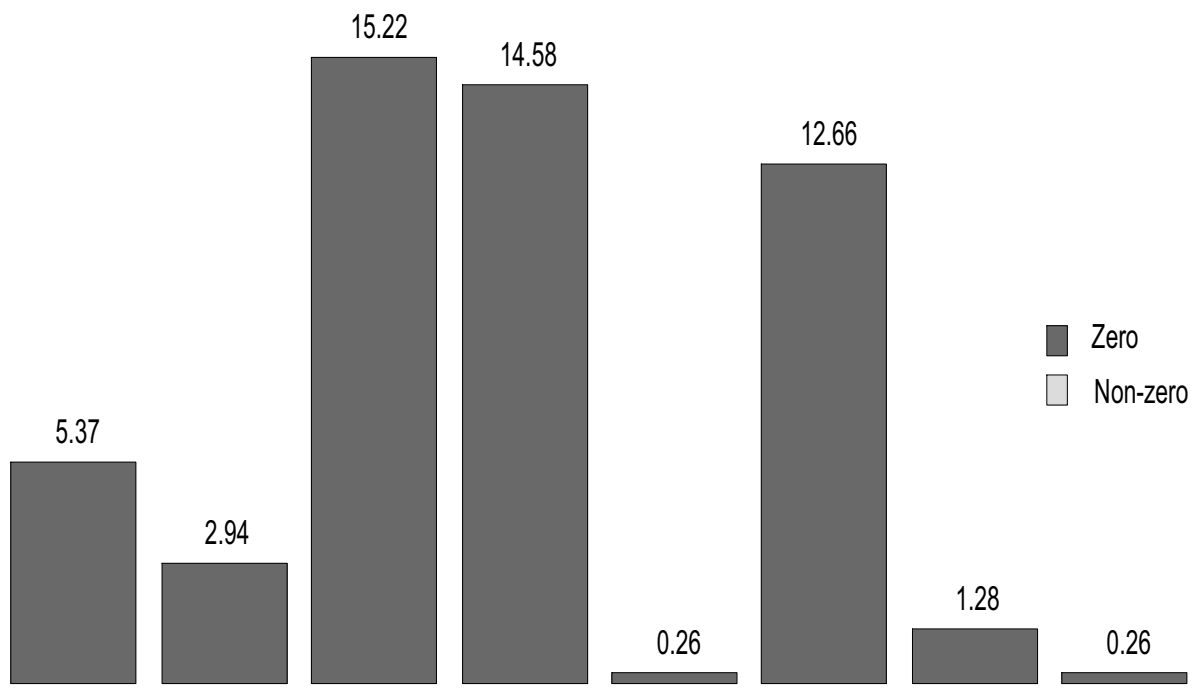


Figure 2

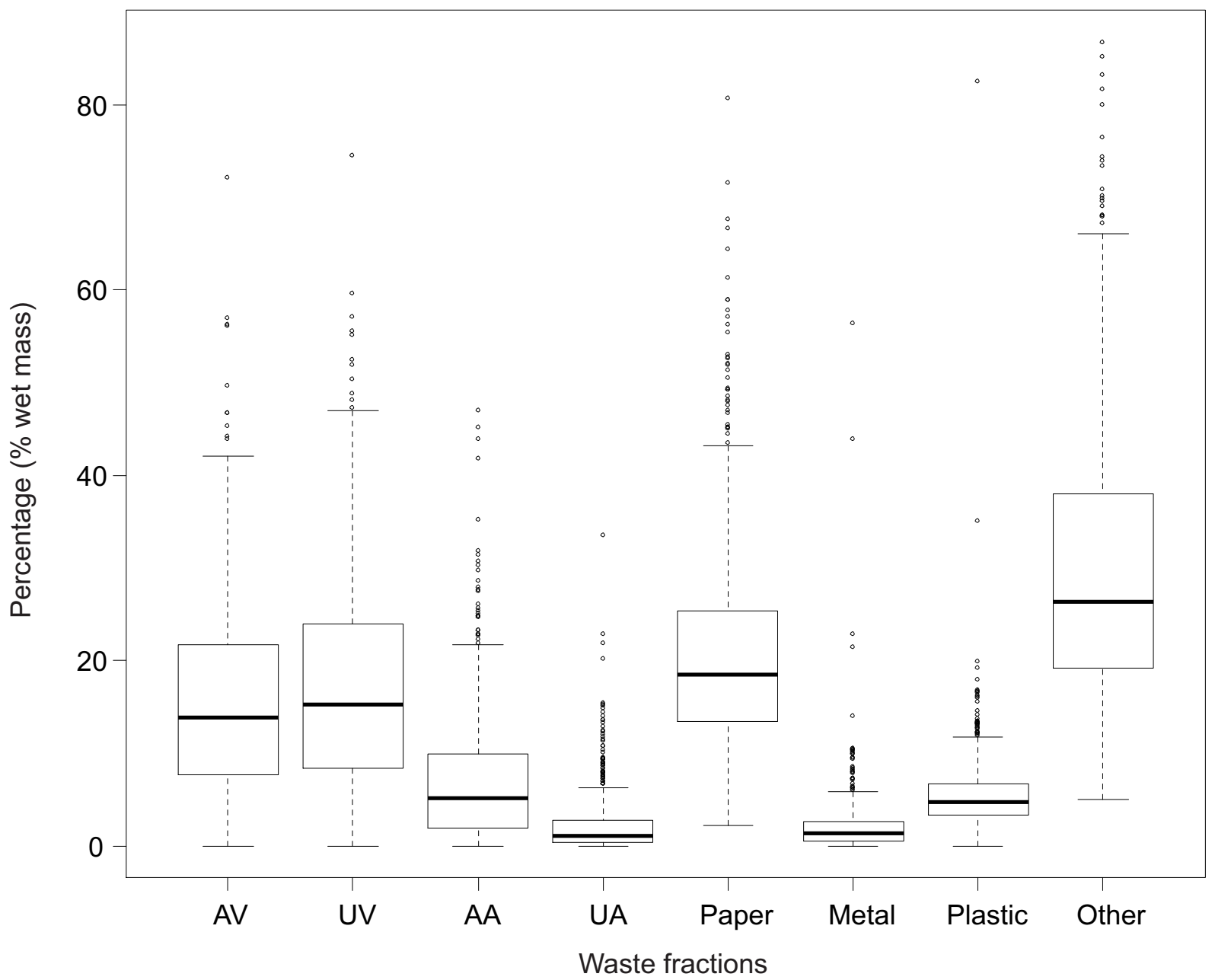


Figure 3

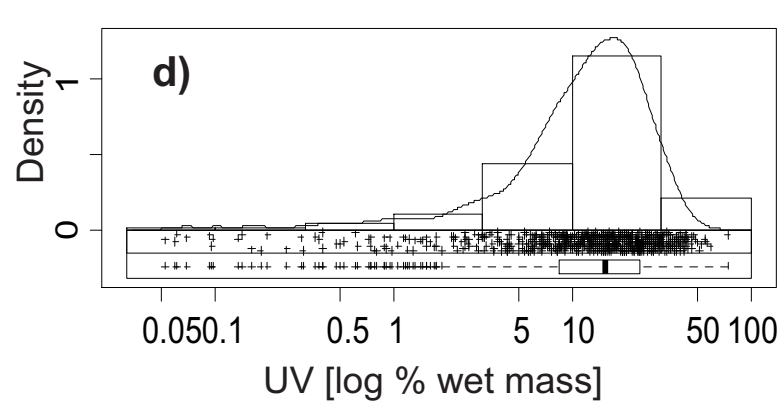
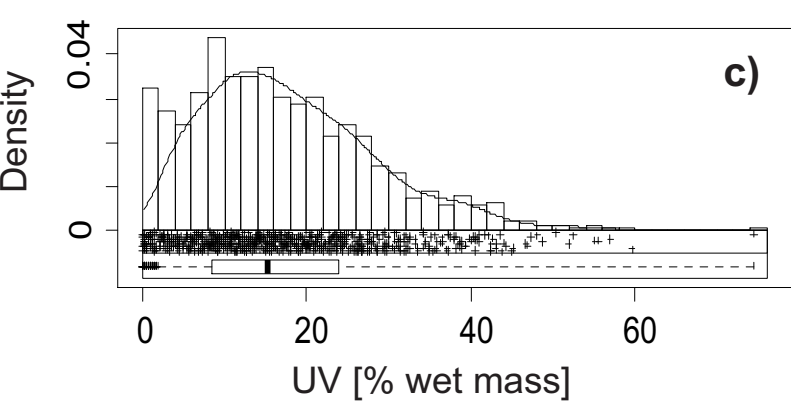
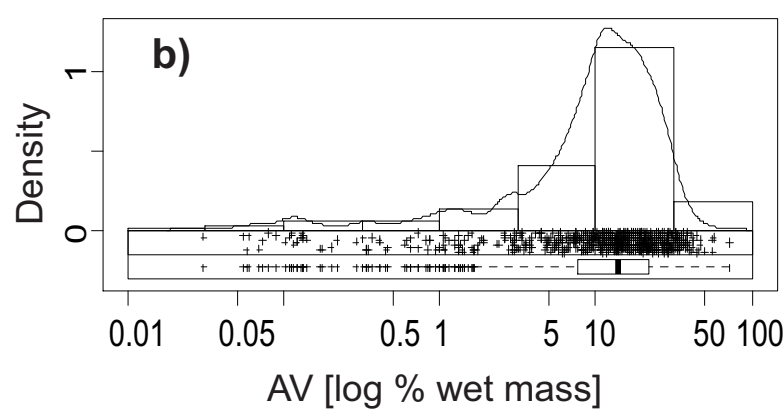
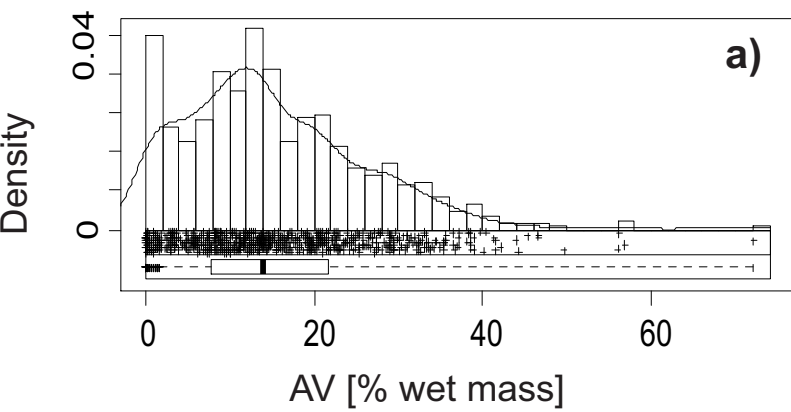


Figure 4

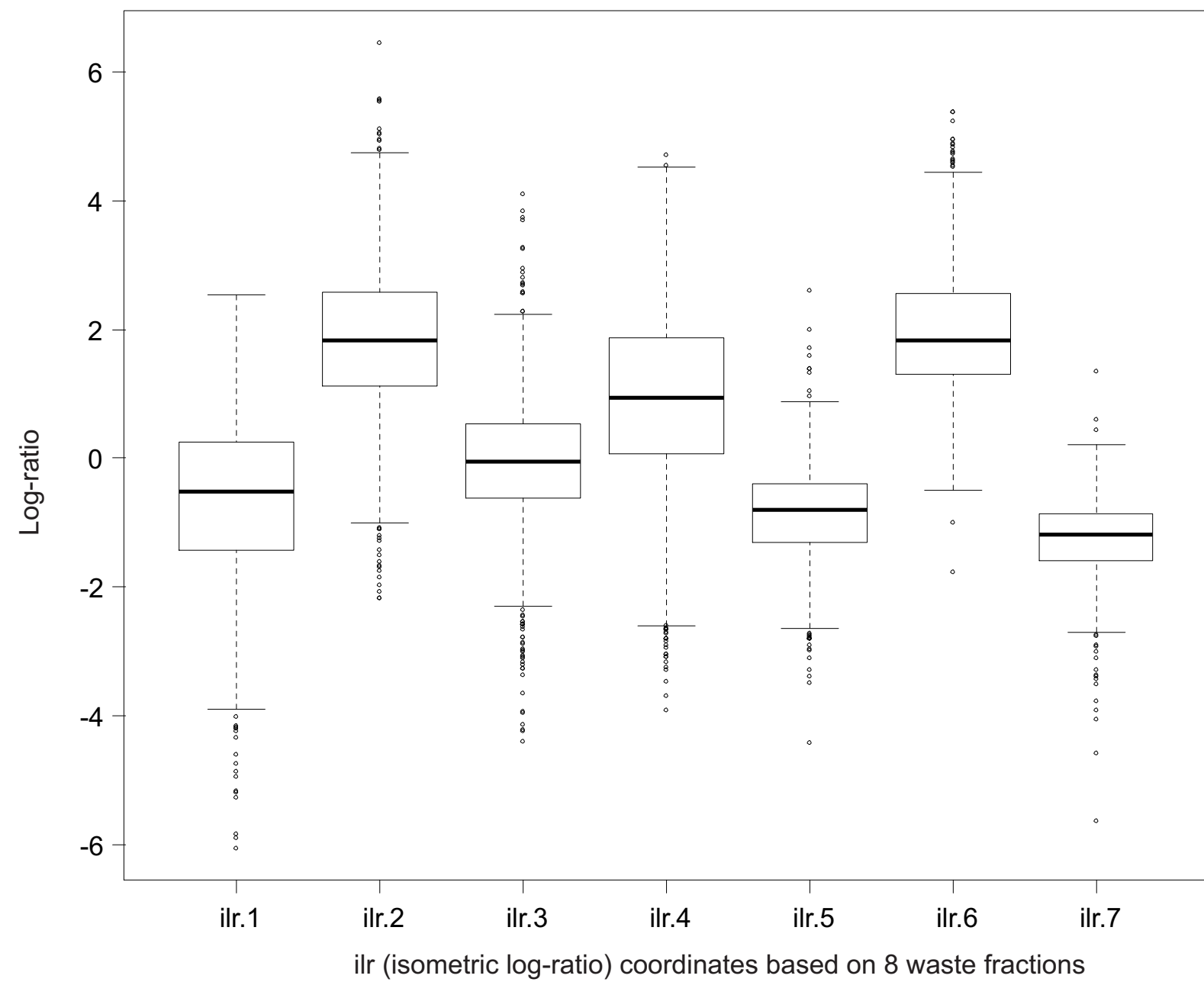
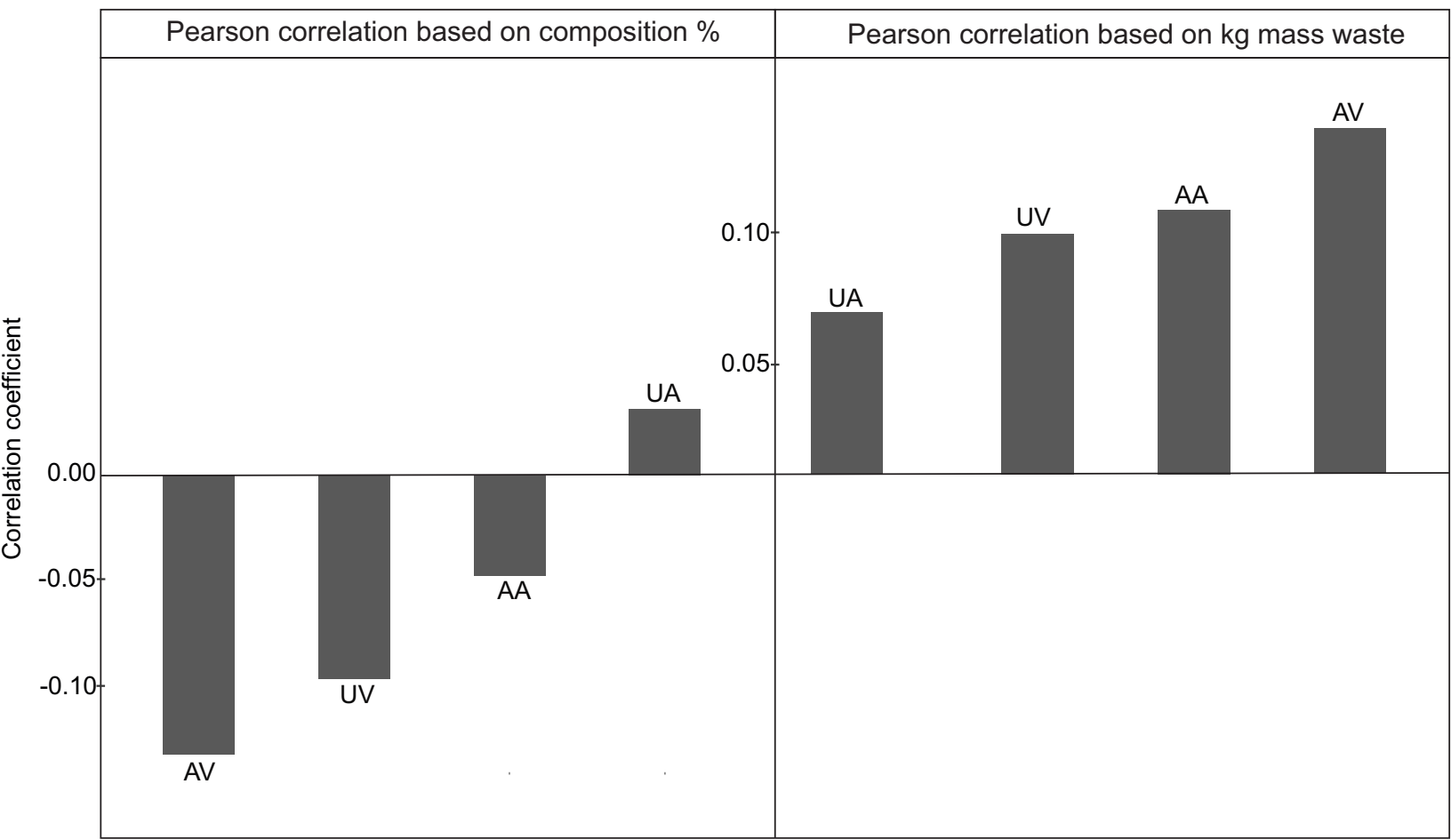


Figure 5



17

18 **Figure captions**

19

20

21 Figure 1: Identification of zero value patterns in residual
22 household waste dataset subdivided into eight waste fractions
23 (see Table 1) and consisting of 779 observations (households).
24 Vertical bars (in dark grey) represent percentage of count
25 number of zero values for each waste fractions; Horizontal
26 bars (light grey) indicate the percentage of count number of no
27 zero value for each combination of eight waste fractions in the
28 households-33 zero values patterns were observed.

29

30

31 Figure 2: Percentage distribution of the composition of residual
32 household waste fractions on wet mass basis (see Table 1 for
33 abbreviation).

34

35

36 Figure 3: Combined histogram and boxplot of raw (a) and log-
37 transformed (b) avoidable vegetable food waste; and raw (c)
38 and log-transformed (d) unavoidable vegetable food waste.

39

40

41 Figure 4: Boxplot showing the distribution of ilr coordinates
42 (number of coordinates equals to number of waste fractions
43 ($D=8$) minus 1)

44

45

46 Figure 5: Results of Pearson correlation test between plastic
47 packaging and food waste fractions (AV, UV, AA, and UA),
48 based on (i) percentage (%) and (ii) kg mass of waste fractions.

49

Statistical analysis of solid waste composition data: arithmetic mean, standard deviation and correlation coefficients

Maklawe Essonanawe Edjabou^{1*}, Josep Antoni Martín-Fernández², Charlotte Scheutz¹, Thomas Fruergaard Astrup¹

1) Department of Environmental Engineering, Technical University of Denmark, 2800 Kgs.

Lyngby, Denmark

2) Dept. Computer Science, Applied Mathematics and Statistics, University of Girona, Campus

Montilivi (P4), E-17071 Girona, Spain

Supplementary materials (SM)

Supplementary materials contain detailed food waste data used for calculations. SMs are divided into tables (Table SM) and figures (Figure SM).

Supplementary materials (SM) –Tables

SM Table 1 Correlation matrix from Pearson` correlation test and significance levels of **log-transformed** data(r: range:-1.00 to +1.00)

| | AV ^d | UV ^e | AA ^f | UA ^g | Paper ^h | Metal ⁱ | Plastic ^j | Other | Datasets |
|--|-----------------|-----------------|-----------------|-----------------|--------------------|--------------------|----------------------|-------|--------------------------------|
| Avoidable vegetable food waste (AV) | 1 | * | *** | 0 | *** | . | 0 | *** | Percentage % kg/capita/week |
| Unavoidable vegetable food waste (UV) | 0.08 | 1 | 0 | *** | 0 | 0 | 0 | *** | Percentage % kg/capita/week |
| Avoidable animal-derived food waste (AA) | 0.41 | 1 | *** | *** | *** | *** | *** | *** | Percentage % kg/capita/week |
| Unavoidable animal-derived food Waste (UA) | 0.34 | 0 | 1 | 0 | *** | . | 0 | *** | Percentage % kg/capita/week |
| Paper and board | 0.53 | 0.27 | 1 | *** | *** | *** | *** | *** | Percentage % kg/capita/week |
| Metal packaging | -0.01 | 0.13 | 0.02 | 1 | 0 | * | ** | ** | Percentage % kg/capita/week |
| Plastic packaging | 0.23 | 0.29 | 0.2 | 1 | *** | *** | *** | *** | Percentage % kg/capita/week |
| Other | -0.21 | -0.05 | -0.14 | 0.01 | 1 | 0 | 0 | *** | Percentage % kg/capita/week |
| | 0.41 | 0.38 | 0.31 | 0.22 | 1 | *** | *** | *** | Percentage % kg/capita/week |
| | 0.07 | 0.01 | 0.06 | 0.09 | -0.05 | 1 | *** | . | Percentage % kg/capita/week |
| | 0.34 | 0.24 | 0.27 | 0.21 | 0.28 | 1 | *** | *** | Percentage % kg/capita/week |
| | -0.04 | -0.04 | 0.04 | 0.11 | 0.02 | 0.18 | 1 | * | Percentage % kg/capita/week |
| | 0.4 | 0.29 | 0.36 | 0.25 | 0.38 | 0.38 | 1 | *** | Percentage % kg/capita/week |
| | -0.31 | -0.37 | -0.22 | -0.1 | -0.27 | -0.06 | -0.08 | 1 | Percentage % kg/capita/week |
| | 0.38 | 0.23 | 0.29 | 0.18 | 0.43 | 0.3 | 0.38 | 1 | Percentage % kg/capita/week |

***Very high significance probability higher than 0.001

**High significance probability between 0.001 and 0.01

*Significance probability between 0.01 and 0.05

() no significance-probability higher than 0.05

^a amount of waste (wet basis) per household per week

^b amount of waste (wet basis) per person per week

^c Composition of residual household waste on wet basis.

^dAvoidable vegetable food waste

^eUnavoidable vegetable food waste

^fAvoidable animal-derived food waste

^gUnavoidable animal-derived food waste

^hPaper; ⁱMetal packaging.; ^jPlastic packaging; ^kOther (see Table 1).

SM Table 2 Summary of waste fraction generation rates data set **before** zero values replacement

| | min | max | mean | Standard deviation |
|--|-------|--------|-------|--------------------|
| Avoidable vegetable food waste (AV) | 0.000 | 12.435 | 1.760 | 1.654 |
| Unavoidable vegetable food waste (UV) | 0.000 | 21.750 | 1.687 | 1.457 |
| Avoidable animal-derived food waste (AA) | 0.000 | 9.314 | 0.755 | 0.891 |
| Unavoidable animal-derived food Waste (UA) | 0.000 | 5.450 | 0.210 | 0.344 |
| Paper and board | 0.050 | 14.519 | 2.042 | 1.616 |
| Metal packaging | 0.000 | 13.415 | 0.213 | 0.556 |
| Plastic packaging | 0.000 | 19.415 | 0.524 | 0.753 |
| Other | 0.194 | 25.747 | 3.063 | 2.583 |

SM Table 3 Summary of waste fraction generation rates data set **after** zero values replacement

| | min | max | mean | Standard deviation |
|--|-------|--------|-------|--------------------|
| Avoidable vegetable food waste (AV) | 0.006 | 12.435 | 1.760 | 1.653 |
| Unavoidable vegetable food waste (UV) | 0.006 | 21.750 | 1.687 | 1.457 |
| Avoidable animal-derived food waste (AA) | 0.003 | 9.314 | 0.756 | 0.891 |
| Unavoidable animal-derived food Waste (UA) | 0.002 | 5.450 | 0.210 | 0.344 |
| Paper and board | 0.050 | 14.519 | 2.042 | 1.616 |
| Metal packaging | 0.002 | 13.415 | 0.213 | 0.556 |
| Plastic packaging | 0.007 | 19.415 | 0.524 | 0.753 |
| Other | 0.194 | 25.747 | 3.063 | 2.583 |

SM Figure 1: Comparison of waste data sets before and after zero values replacement

