

International Journal of Environmental Research and Public Health



Article Robust Compositional Analysis of Physical Activity and Sedentary Behaviour Data

Nikola Štefelová¹, Jan Dygrýn^{2,*}, Karel Hron¹, Aleš Gába², Lukáš Rubín² and Javier Palarea-Albaladejo³

- ¹ Faculty of Science, Palacký University Olomouc, 771 11 Olomouc, Czech Republic; Nikola.Stefelova@seznam.cz (N.Š.); hronk@seznam.cz (K.H.)
- ² Faculty of Physical Culture, Palacký University Olomouc, 771 11 Olomouc, Czech Republic; ales.gaba@upol.cz (A.G.); lukas.rubin@upol.cz (L.R.)
- ³ Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Edinburgh EH9 3FD, UK; javier.palarea@bioss.ac.uk
- * Correspondence: jan.dygryn@upol.cz; Tel.: +420-585-636-469

Received: 20 August 2018; Accepted: 11 October 2018; Published: 14 October 2018



Abstract: Although there is an increasing awareness of the suitability of using compositional data methodology in public health research, classical methods of statistical analysis have been primarily used so far. The present study aims to illustrate the potential of robust statistics to model movement behaviour using Czech adolescent data. We investigated: (1) the inter-relationship between various physical activity (PA) intensities, extended to model relationships by age; and (2) the associations between adolescents' PA and sedentary behavior (SB) structure and obesity. These research questions were addressed using three different types of compositional regression analysis-compositional covariates, compositional response, and regression between compositional parts. Robust counterparts of classical regression methods were used to lessen the influence of possible outliers. We outlined the differences in both classical and robust methods of compositional data analysis. There was a pattern in Czech adolescents' movement/non-movement behavior-extensive SB was related to higher amounts of light-intensity PA, and vigorous PA ratios formed the main source of potential aberrant observations; aging is associated with more SB and vigorous PA at the expense of light-intensity PA and moderate-intensity PA. The robust counterparts indicated that they might provide more stable estimates in the presence of outlying observations. The findings suggested that replacing time spent in SB with vigorous PA may be a powerful tool against adolescents' obesity.

Keywords: compositional data; compositional linear regression; log-ratio methodology; pivot coordinates; physical activity

1. Introduction

High physical activity (PA) and a low amount of sedentary behavior (SB) are often associated with several benefits for children and adolescents, including positive effects on cardiovascular, musculoskeletal and metabolic health [1,2], well-being [3], and obesity [4]. Movement/non-movement behavior patterns in youth may extend to adulthood [5,6] and, therefore, understanding the optimal structure of time spent in various activities in early life is desirable. Current PA guidelines for adolescents around the world focus mainly on time spent in moderate PA (MPA) and vigorous PA (VPA) and encourage children and adolescents to engage in at least 60 min of moderate-to-vigorous PA every day [2]. However, being active for an hour a day does not mean that the rest of the day has no effect on overall health. Thus, some of the health benefits of meeting the PA recommendations might be lost if children are subject to extensive sitting. SB and light-intensity PA (LIPA) have not been a

preferred subject of interest in public health research for decades. It was assumed that if MPA and VPA increase, SB inherently decrease. However, MPA and VPA usually account for less than 5% of the day, whilst the time spent in LIPA is approximately three times greater [7]. LIPA may also be beneficial for children's health [8]. Additionally, excessive SB and inadequate sleep duration also represent critical health challenges to children and youth today [9]. These behaviors account for a large proportion of the day (~80%) and are considered independent risk factors for obesity [10].

The main issues in research on the relationship between movement behavior patterns and health outcomes are: (1) length of time needed for manifestation of diseases or disorders/health problems caused by an unhealthy lifestyle [11] and (2) limited possibilities to distinguish between the effects of all components of PA and SB in the prevention of health problems (e.g., obesity) [12]. It is widely thought that a greater understanding of the role of each PA and SB component with respect to other components in this age group is necessary and should be assessed in a more holistic way. Moreover, PA and SB data are compositional in nature and, hence, any proportional increase in one behavior necessarily causes a proportional decrease in one or more remaining behaviors; these behaviors are perfectly collinear if an exhaustive description of the movement behavior composition (including sleep) is provided. However, even if this is not the case, the information contained in the movement behavior composition is relative and thus scale invariant; i.e., it can be normalized to any sum (such as 1 in the case of proportions or 100 for percentages) without loss of information.

Standard statistical techniques such as multiple regression analysis are not well suited to describe the relationship between a response of interest and explanatory covariates of a compositional nature. They are not designed to accommodate and account for the relative contributions of the parts of a whole (herein "total wake time"), hereafter called compositional data [13]. Although early studies point toward the use of compositional data analysis (CODA) in PA and SB research [14,15], it was not until the work by Chastin et al. [12] that the CODA approach to PA research was introduced and discussed in a comprehensive and statistically-principled means using the latest methodological developments in the area to deal with behavioral data that represent parts of a whole (i.e., data carrying relative information), useful when either all parts or just some of them have been measured. Over the last few years, the CODA approach has notably increased its popularity in the physical activity community, and there has been a large number of studies examining the relationship between health and movement/non-movement behavior pattern among youth using CODA [16-21]. Finally, a theoretical framework, which describes the relationship between all movement/non-movement behaviors and health outcomes, has recently been updated to The Framework for Viable Integrative Research in Time-Use Epidemiology (VIRTUE framework) [22]. Up to now, primarily classical statistical analysis methods were used in PA and SB research; therefore, the aims of this study are: (1) to present a concise regression framework within CODA in the presence of outlying observations, which often occur in most real-world data sets, encouraging the use of robust statistical methods instead of their classical counterparts in these situations; and (2) to investigate the combined effect of time spent in LIPA, MPA, VPA and SB on the risk of obesity within a robust CODA framework.

2. Materials and Methods

2.1. Design of Study

Cross-sectional data from "Multifactorial research on built environment, active lifestyle and physical fitness in Czech adolescents" were used [23]. Adolescents were recruited from schools in Olomouc and Hradec Králové. To ensure a representative sample, we selected schools from different built environments and socioeconomic neighborhoods. The participants were healthy 420 adolescents (169 boys and 251 girls), whose ages ranged from 11.8–18.9 years (M = 15.1 \pm 2.0), height was 167.8 \pm 9.2 cm, weight was 60.7 \pm 13.2 kg and BMI was 21.4 \pm 3.6 kg·m⁻². The data were collected during the 2013–2016 academic years in elementary and secondary schools. SB, LIPA, MPA, and VPA were measured by the ActiGraph GT3X accelerometer (ActiGraph, LLC, Pensacola, FL,

USA); sleep was not considered. Participants were asked to wear accelerometers for seven consecutive days, except during water-based activities. The days on which the accelerometer was worn for at least 10 h were considered valid and participants were included if at least three valid weekdays and one valid weekend day of their activity were available. The time sampling interval was set to 60 s, which has been the most frequently used epoch in the literature [24]. Each minute epoch was classified using standard counts per minute thresholds as SB (0–100 counts/min), LIPA (101–2295 counts/min), MPA (\geq 2296–4011 counts/min) and VPA (\geq 4012 counts/min), based on Evenson's cut-points [25]. These cut-points demonstrated significant higher accuracy for SB and different levels of PA intensity than other cut-points in adolescents [26]. For the purpose of the study, averaged results of movement behavior for each participant were taken.

Body Mass Index (BMI) was calculated from the self-reported height and weight. Age and sex-adjusted BMI (zBMI) was used as an indicator of overweight and obesity. Due to the lack of individuals in each weight category, only two groups were defined; underweight/normal (<1.0 SD) and overweight/obese (\geq 1.0 SD). The two groups included n_1 = 344 underweight/normal and n_2 = 76 overweight/obese adolescents. The calculation was done according to WHO guideline [27].

2.2. Compositional Data

Compositional data (compositions) stand for all kinds of data representing parts of a whole. A vector $\mathbf{x} = (x_1, x_2, ..., x_D)$ is called a D-part composition when all its components (parts) are strictly positive real numbers and carry relative information [13,28]. This means that all relative information about a certain composition is contained in the ratios between its parts.

Compositional data are scale invariant, that is, if a composition is multiplied by a positive number, it does not change the ratios between the components. Accordingly, compositions can be represented without loss of information as vectors with an arbitrary constant sum constraint (typically 100 in percentage units), but the sums of components can also vary across the data set and compositional methods would still produce the same results. The operation of rescaling the initial vector so that the components add up to a constant K is formally called a closure. The resulting sample space of such constrained representation is a simplex, which is a (D - 1)-dimensional subset of the ordinary real space.

Geometrically, compositions follow the Aitchison geometry on the simplex [19]. Since standard statistical methods rely on Euclidean geometry in real space, it is desirable to map compositions from the simplex into real space to be able to apply a standard statistical analysis. This can be achieved by the log-ratio methodology, which leads to the construction of coordinates that preserve the relative scale property of compositions. In the log-ratio methodology, the ratios between parts as bearers of the elemental information are replaced by log-ratios that conveniently symmetrize the interpretation of the ratios and are mathematically easier to handle. Among the different types of log-ratio transformations, the ilr (isometric log-ratio) transformations are preferred (for geometrical reasons they are referred to as ilr coordinates), as they allow to express compositions in an orthonormal coordinate system [28].

There are many ways to construct ilr coordinates but their interpretation might be quite difficult. Therefore, it is appealing to choose an orthonormal coordinate system that highlights the role of a single compositional part against the others, which have been recently called pivot coordinates [29,30]. They are constructed as follows. For a *D*-part composition **x**, a real vector $\mathbf{z} = (z_1, z_2, ..., z_{D-1})$ is obtained with D - 1 ilr coordinates, where

$$z_{j} = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_{j}}{\frac{D-j}{\sqrt{\prod_{k=j+1}^{D} x_{k}}}}, \quad j = 1, \dots, D-1.$$
(1)

Thus, all relative information about x_1 (with respect to the average contribution of the remaining parts) is explained by z_1 . As an alternative interpretation, note that this first coordinate represents

an aggregation of all pair-wise log-ratios of a specific compositional part to all the others (up to a normalizing constant).

By permuting the parts in the original composition, so that a different part is put in the first position each time, we can obtain *D* different orthonormal coordinate systems, each of them highlighting the role of one of the compositional parts. Denoting $\mathbf{x}^{(l)} = (x_l, x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_D) = (x_1^{(l)}, x_2^{(l)}, \dots, x_D^{(l)}), l = 1, \dots, D$, produces

$$z_j^{(l)} = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_j^{(l)}}{\sqrt[D-j]{\prod_{k=j+1}^{D} x_k^{(l)}}}, \quad j = 1, \dots, D-1, \quad l = 1, \dots, D.$$
(2)

Therefore, all relative information about any part x_l is contained in $z_1^{(l)}$, l = 1, ..., D [29]. Note that different ilr coordinate systems are just a rotation of each other. This is a useful property in statistical analysis; for example, for robust estimation as well as for regression analysis, because it enables the use of an arbitrary choice of ilr coordinates to obtain the required (unique) output.

2.3. Compositional Regression

The objective of the regression analysis is to study the relationship between response variable(s) and explanatory variable(s). The compositional data framework has three basic regression problems. These relate to the relation between the real-valued response and compositional covariates, compositional response and real covariates, or between compositional parts themselves. The choice of pivot coordinates is convenient in all these cases, as it enhances the interpretability of the regression coefficients.

Firstly, in the situation concerning real response *Y* and compositional covariates $(x_1, x_2, ..., x_D)$, the *D* regression models are considered:

$$Y = \beta_0 + \beta_1^{(l)} z_1^{(l)} + \ldots + \beta_{D-1}^{(l)} z_{D-1}^{(l)} + \varepsilon^{(l)}, l = 1, \ldots, D,$$
(3)

where $z_1^{(l)}, z_2^{(l)}, \ldots, z_{D-1}^{(l)}$ are pivot coordinates (2), β_0 , $\beta_1^{(l)}, \beta_2^{(l)}, \ldots, \beta_{D-1}^{(l)}$ are unknown regression coefficients and $\varepsilon^{(l)}$ is the random error term in the *l*th model, $l = 1, \ldots, D$. Due to the orthogonality of different coordinate systems, the intercept term β_0 (as well as the model fit measures) is the same for each model. As $z_1^{(l)}$ explains all the relative information about part x_l , the coefficient $\beta_1^{(l)}$ accounts for the contribution of the dominance of this part within the given composition. Therefore, the estimate $\hat{\beta}_1^{(l)}$ is extracted from each model and the vector $(\hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_1^{(D)})$ is used for interpretation purposes [31,32].

Secondly, compositional response $(Y_1, Y_2, ..., Y_D)$ and real explanatory variables $(x_1, x_2, ..., x_p)$ are considered. By expressing the composition in pivot coordinates (2), the *D* multiple regression models are set, where only the first coordinate $Z_1^{(l)}$ is of interest:

$$Z_1^{(l)} = \beta_0^{(l)} + \beta_1^{(l)} x_1 + \ldots + \beta_p^{(l)} x_p + \varepsilon^{(l)}, l = 1, \ldots, D.$$
(4)

 $\beta_0^{(l)}$, $\beta_1^{(l)}$,..., $\beta_p^{(l)}$ in (4) are unknown regression coefficients and $\varepsilon^{(l)}$ is the random error term in the *l*th model, l = 1, ..., D [33]; note, however, that these coefficients are not the same as in model (3). In both models above, it is assumed that just the response is measured with errors. Therefore, the standard regression methodology can be used for statistical inference, particularly for hypotheses testing.

Finally, for the case of regression of one compositional part on other parts, the use of orthogonal regression is appropriate, since both the response and explanatory variables are measured with errors. Accordingly, statistical inference is obtained by bootstrap [34]. Specifically, if the relationship

between the dominance of part x_l within the composition and the remaining subcomposition $(x_1, x_{l-1}, x_{l+1}, \dots, x_D)$ is of interest, D - 1 models:

$$z_{1}^{(l)} = \beta_{1}^{(lk)} + \beta_{2}^{(lk)} z_{2}^{(lk)} + \ldots + \beta_{D-1}^{(lk)} z_{D-1}^{(lk)} + \varepsilon^{(l)}, \quad l = 1, \ldots, D,$$

$$k = 1, \ldots, l-1, \ l+1, \ldots, D,$$
(5)

are considered where $\beta_1^{(lk)}$, $\beta_2^{(lk)}$, ..., $\beta_{D-1}^{(lk)}$ are unknown regression coefficients, ε is the random error term and $z_1^{(l)}, z_2^{(lk)}, \ldots, z_{D-1}^{(lk)}$ stand for pivot coordinates (2) constructed for composition with x_l in the first and x_k in the second position. Then $z_2^{(lk)}$ contains all the relative information about x_k with respect to the average contribution of the remaining parts of the subcomposition. Accordingly, our focus is on the interpretation of the estimate $\hat{\beta}_2^{(lk)}$ from each model, $k = 1, \ldots, l - 1, l + 1, \ldots, D$ [33], because it stands for a positive or negative influence of dominating x_k within the subcomposition ($z_2^{(lk)}$) on dominance of x_l within the parent composition (as conveyed by $z_1^{(l)}$).

In addition, the coefficient of determination (\mathbb{R}^2) was computed for all regression models to evaluate the goodness of fit. Since the main purpose of the modelling was to investigate the sign and size of the relationships, we focused on the statistical significance of the model parameters and allowed for small values of \mathbb{R}^2 as generally accepted in our field [35]. Note that the coefficient of determination is invariant to orthogonal coordinate representation of predictors, so for any pivot coordinates the resulting value is identical.

2.4. Dealing with Outliers

When it is needed to reduce the influence of possible outliers in the dataset (and thus to focus on the main trend in the dataset), robust estimation of regression coefficients [36] should be used. As all three regression models above are formulated in real coordinates, it is possible to use standard robust regression techniques. Thus, in the first two cases (models (3) and (4)), the highly efficient MM-estimation is preferable instead of the standard LS (least squares) estimation [37]. For orthogonal regression (model (5)), robust estimation based on the robust principal component analysis was applied and fast robust bootstrap is used for statistical inference [34]. In order to interpret the results from robust regression, the specific consequences of outlying observations in compositional data sets need to be taken into account. While for standard real data outliers are characterized usually by extreme value(s) in one or more variables, in the compositional case the outlyingness originates from aberrant log-ratios and, hence, irrespective of the concrete representation of the relative information conveyed by the composition. Finally, note that this also affects compositional descriptive statistics, which are represented by the centre and the variation matrix [28] and used in the present analysis to accompany the regression analysis results. A key to their robustification is the application of the robust Minimum Covariance Determinant (MCD) estimator of location and covariance [35] in any orthonormal coordinate. The MCD estimator is a highly robust estimator and transforms according to the change of coordinates (this is referred to as "affine equivariance"), so that unique results for the original compositional data are obtained.

As outlying observations occur naturally also in PA data, the use of robust methods for regression analysis in this work was selected and illustrated. The usefulness of the robust approach was demonstrated also for the case of basic descriptive statistics of compositional data such as the centre and the variation matrix [28].

3. Results

In the following text, the above-introduced four distinct types of activities (SB, LIPA, MPA, and VPA) were analyzed. The average daily behavior of an individual was considered representative of their habitual PA. Additionally, adolescents' sex, age, weight and height were recorded and the latter three variables were used in the following analysis.

For the sake of succinctness, the following notation will be used:

$$MB = (SB, LIPA, MPA, VPA)$$
(6)

for the entire movement behavior (MB) composition and

$$PA = (LIPA, MPA, VPA)$$
(7)

for the PA subcomposition in particular.

3.1. Relationship between Various Movement Behaviour Intensities

At first, the authors analyzed the inter-relationships between the parts that form the MB composition. The proportional relationships were summarized in the so-called compositional variation matrix, formed by the variances of all possible pair-wise log-ratios [28] (Table 1). The lower the value of these variances, the more proportional the respective compositional parts are. Accordingly, Table 1 shows that SB and LIPA were the most proportional parts, while the least proportional pair was LIPA and VPA.

Table 1. Compositional variation matrix of adolescents' movement behavior composition.

MB	SB	LIPA	MPA	VPA
SB	0	0.1460	0.2544	0.7757
LIPA	0.1460	0	0.1577	0.8993
MPA	0.2544	0.1577	0	0.7828
VPA	0.7757	0.8993	0.7828	0

Note: MB, movement behavior; SB, sedentary behavior; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

Alternatively, a robust variation matrix [38] as shown in Table 2 can be obtained by robust estimation of the covariance matrix in any pivot coordinate system (MCD estimator was used here). By comparing it with the previous (classical) variation matrix, it can be observed that the proportionality between SB and LIPA was slightly weakened, therefore now LIPA and MPA become the most proportional components.

Table 2. Robust compositional variation matrix of adolescents' movement behavior composition.

MB	SB	LIPA	MPA	VPA
SB	0	0.1669	0.2801	0.6624
LIPA	0.1669	0	0.1578	0.8351
MPA	0.2801	0.1578	0	0.6946
VPA	0.6624	0.8351	0.6946	0

Note: MB, movement behaviour; SB, sedentary behaviour; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

A limitation of the variation matrix is that it does not consider the effect of other parts on the relationship between the two components involved. Moreover, it does not provide information about positive or negative directions of the relationship. To circumvent this limitation, we investigated the relationship between SB, the "non-active" part in the MB composition, and the parts in the PA sub-composition via robust orthogonal regression. For this purpose, model (5) was specified as:

$$z_1^{(SB)} = \beta_1 + \beta_2^{(SB, PA)} z_2^{(SB, PA)} + \beta_3^{(SB, PA)} z_3^{(SB, PA)} + \varepsilon$$
(8)

Thus, three models are obtained, where $(z_1^{(SB)}, z_2^{(SB, PA)}, z_3^{(SB, PA)})$ are pivot coordinates for the MB composition with SB in the first position and a particular part (either LIPA, MPA or VPA) in the second position. The estimate $\hat{\beta}_2^{(SB, PA)}$ from each model is of interest.

The results (Table 3) imply that the relative dominance of SB with respect to the average contribution of PA parts is positively associated with the relative dominance of LIPA with respect to the average contribution of the remaining PA parts, whereas it is negatively related to the relative dominance of MPA and VPA (with respect to the average contribution of the remaining PA parts). It can be concluded that those individuals who spend a relatively large amount of time in SB also tend to spend a relatively high amount of time in LIPA and a relatively small amount of time in MPA and VPA.

PA	$\hat{oldsymbol{eta}}_2^{(P\!A)}$	Standard Error	<i>p</i> -Value
LIPA	1.4295	0.1597	< 0.001
MPA	-1.3179	0.1816	< 0.001
VPA	-0.1116	0.0397	0.014

Table 3. Robust orthogonal regression estimates for regression model (8).

Note: $R^2 = 0.3005$ for all models. LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

3.2. Changes in Movement Behaviour with Increasing Age

In the following part, we investigate how age is associated with the structure of adolescents' PA. To get an initial insight into the problem, the data were displayed in a ternary diagram, which is a standard tool for visualization of the simplex sample space for a three-part (sub)composition [28]. A color gradient was used to distinguish PA points by age (Figure 1). For a further insight, it is useful to plot centered data (Figure 2), particularly when the data are concentrated near the borders of the ternary diagram. The reason for this is the relative scale of compositional data; near the borders, the ratios between the components change substantially more than near the barycenter and this is reflected by larger distances between points in terms of the Aitchison geometry [39]. It means that near the borders, outlying observations might easily be overlooked due to the small relative values of compositional parts. Moreover, the centering operation can be robustified by computing a robust center measure, as it was the case of the present analysis, instead of the common compositional center in order to prevent possible masking of outliers. For the compositional center, using the geometric mean is a default choice in the compositional framework [28]. The robust compositional mean is then obtained by using the MCD estimator on pivot coordinates, and subsequently back-transforming, which results in a weighted geometric mean [38].



Figure 1. Cont.



Figure 1. Ternary diagrams visualizing how the structure of PA and SB is associated with age. Each plot represents relationship between three of four behaviors; (**a**) SB, LIPA and MPA, (**b**) SB, LIPA and VPA, (**c**) SB, MPA and VPA, (**d**) LIPA, MPA and VPA. The lighter the point color, the higher the age of an individual. The diagrams indicate that the proportions of time spent in SB and VPA is associated with higher age, whereas the effect is the opposite for LIPA and MPA.



Figure 2. Ternary diagrams visualizing how the structure of PA is associated with age using robustly centered data. Each plot represents relationship between three of four behaviors; (**a**) SB, LIPA and MPA, (**b**) SB, LIPA and VPA, (**c**) SB, MPA and VPA, (**d**) LIPA, MPA and VPA. Plotting of the centered data is recommended when the data are concentrated near the borders of the triangle. Note that tick labels are no longer meaningful after centering.

An exchange between LIPA and SB time as adolescents get older is apparent (Figures 1 and 2). It is interesting to see some outliers towards VPA that might indicate intensive sport activities of adolescents. Even in the PA sub-composition, a similar trend can be observed.

The relationship between age and MB composition was further examined via a regression analysis. The MB composition was set as the response and age as the explanatory variable. Thus, the following regression model (4) was considered:

$$z_1^{(MB)} = \beta_0 + \beta_1^{(MB)} Age + \varepsilon^{(MB)}$$
⁽⁹⁾

Robust MM-estimation was conducted while focusing on the $\hat{\beta}_1^{(MB)}$ coefficient from each of the four models informing about increasing or decreasing dominance of the part isolated in the response pivot coordinate.

The results displayed in Table 4 support the anticipated positive association between age and the relative dominances of SB and VPA (with respect to the average contribution of the other parts). Conversely, the relative dominances of LIPA and MPA were negatively associated with age. The older they get, Czech adolescents tend to spend more time in SB and VPA at the expense of LIPA and MPA.

MB	$\hat{oldsymbol{eta}}_1^{(P\!A)}$	Standard Error	<i>p</i> -Value	R ²
SB	0.0200	0.0087	0.023	0.0119
LIPA	-0.1040	0.0074	< 0.001	0.2969
MPA	-0.0709	0.0088	< 0.001	0.1335
VPA	0.1464	0.0157	< 0.001	0.1674

Table 4. MM-regression estimates from model (9).

Note: MB, movement behavior; SB, sedentary behavior; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

Figure 3 shows the predicted means for the first pivot coordinates with increasing age. Indeed, the most dramatic difference is observed for VPA, reflecting on previous observations. Note that R² values differ due to the different response coordinates.



Figure 3. Predicted means of the dominance of the movement composition components with age. Both SB and VPA increase whereas LIPA and MPA decrease.

3.3. The Association between Movement Behaviour and Obesity

Finally, we examined the differences in PA between adolescents from different weight groups. Table 5 shows the compositional center of MB computed for all adolescents, compared with that in the underweight/normal and overweight/obese groups.

Table 5. Center of movement behavior for the whole data set and for underweight/normal and overweight/obese subgroups.

Group	SB	LIPA	MPA	VPA
All	0.60749	0.33605	0.03812	0.01833
Underweight/normal	0.60753	0.33488	0.03829	0.01930
Overweight/obese	0.60694	0.34121	0.03734	0.01451

Note: Values expressed in proportion; MB, movement behaviour; SB, sedentary behaviour; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

The relative difference between underweight/normal and overweight/obese groups can be visualized by the compositional mean barplot [12] (Figure 4). The vertical axis values correspond to the log-ratios between the respective group centers and the overall center after data centering. The numeric labels of the bars inform about the relative differences with respect to the overall center. Thus, in the overweight/obese group, the proportion of time spent in VPA is reduced by 16.4% relatively to the overall center. Accordingly, VPA stands out as a key driver of the difference between the underweight/normal and overweight/obese groups, stressing on the lack of VPA time in overweight/obese adolescents.



SB LIPA MPA VPA

Figure 4. Compositional mean barplot for the underweight/normal and overweight/obese adolescent groups.

In order to investigate the effect of outlying observations in the whole data set, as well as in both weight groups, robust counterparts of the compositional centers and barplots were considered (Table 6 and Figure 5). Although no dramatic changes can be observed in this case, some patterns are evident. Specifically, it can be observed that the robust center for VPA was slightly higher in all groups (compared to the classical one shown in Table 5) and that the relative difference in VPA was smaller in both groups (Figure 5). This suggests the presence of some outliers (those with very low relative contribution of VPA to PA) in both groups of observations. Downplaying them thus provides a more reliable result that better summarizes adolescents' behavior. Particularly, the robust compositional

mean barplot (Figure 5) shows that the proportion of time spent in VPA in the overweight/obese group was reduced by 15.8%, relative to the overall mean composition.

Table 6. Robust center of movement behavior for the whole data set and for underweight/ne	ormal and
overweight/obese subgroups.	

Group	SB	LIPA	MPA	VPA
All	0.60517	0.33675	0.03833	0.01975
Underweight/normal	0.60557	0.33550	0.03840	0.02053
Overweight/obese	0.60806	0.33885	0.03738	0.01571

Note: Values expressed in proportion. SB, sedentary behavior; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.



Figure 5. Robust compositional mean barplot for the underweight/normal and overweight/obese adolescent groups.

The continuous character of zBMI was used to obtain more precise information about the association between PA and obesity using an appropriate regression model. zBMI was set as the response in the regression model (3) and the parts of the MB composition as covariates. Using the notation (6), the corresponding four models are as follows:

$$zBMI = \beta_0 + \beta_1^{(MB)} z_1^{(MB)} + \beta_2^{(MB)} z_2^{(MB)} + \beta_3^{(MB)} z_3^{(MB)} + \varepsilon^{(MB)},$$
(10)

where $(z_1^{(MB)}, z_2^{(MB)}, z_3^{(MB)})$ are pivot coordinates for the MB composition resulting from sequential placement of each part originally in position *l* at the first position (i.e., *l* = 1,2,3,4 for SB, LIPA, MPA and VPA, respectively). We focused on the estimate $\hat{\beta}_1^{(MB)}$ from each of the four models, which refers to the dominance of the corresponding part within the given composition. These models were applied in the PA context in [12,40] using standard LS estimation. Here, robust MM-estimation was used instead to lessen the influence of outlying observations.

The results of MM-regression are displayed in Table 7. Accordingly, when considering just statistically significant coefficients (*p*-value < 0.05) the relative dominance of SB (with respect to the average contribution of the other parts) was in positive relationship with zBMI, while the relative dominance of VPA was in inverse relationship with zBMI. Those who spend too much time in SB and not enough time in VPA tend to have higher BMI.

MB	$^{(MB)}_{\beta_1}$	Standard Error	<i>p</i> -Value
SB	0.3756	0.1897	0.048
LIPA	-0.3077	0.1964	0.118
MPA	0.1388	0.1436	0.334
VPA	-0.2066	0.0813	0.011

Table 7. MM-regression estimates for compositional model (10).

Note: $R^2 = 0.0258$ for all models. MB, movement behavior; SB, sedentary behavior; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

Moreover, using robust regression, it is possible to investigate the effects of outlying observations using a diagnostic plot of standardized residuals versus robust distances of the predictor variables [26], as shown in Figure 6. Following Rousseeuw [31], this enables us to distinguish three basic types of outliers: good leverage points, bad leverage points, and vertical outliers. Good leverage points are points (formed by response values and predictors) that are close to the regression plane; that is, good leverage points improve the precision of the regression coefficients. Bad leverage points are points that are far from the regression plane; that is, bad leverage points reduce the precision of the regression coefficients. Vertical outliers are outlying observations caused by the response, again negatively affecting the regression outputs. In this case, just one diagnostic plot (Figure 6) covers all four regression models due to mutual relationships between pivot coordinates. It can be observed that, although there are some good leverage points (middle right rectangular), bad leverage points (upper or bottom right rectangular) do not occur here. The vertical outliers (upper or bottom left rectangular) appear mostly in the overweight/obese group of adolescents, which indicates that this group has more outliers than the underweight/normal one, including some extremely obese adolescents. Despite the absence of bad leverage points, the presence of vertical outliers indicates possible differences between classical (LS) and robust regression models.



Figure 6. Evaluation of outlying observations from MM-regression model (10).

For this reason, it was sensible to perform a comparison with classical LS estimates (Table 8). MM-regression outputs (Table 7) show a stronger relationship between zBMI and relative dominance of *SB*, while the association between zBMI and relative dominance of VPA is slightly weaker (which is in line with the barplots in Figures 4 and 5).

MB	$\hat{\beta}_1^{(MB)}$	Standard Error	<i>p</i> -Value
SB	0.3237	0.1616	0.046
LIPA	-0.2277	0.1966	0.248
MPA	0.1388	0.1550	0.439
VPA	-0.2163	0.0690	0.002

Table 8. LS-regression estimates for model (10).

Note: $R^2 = 0.0254$ for all models. MB, movement behavior; SB, sedentary behavior; LIPA, light-intensity physical activity; MPA, moderate physical activity; VPA, vigorous physical activity.

The graph in Figure 7 shows predicted zBMI for the mean MB composition (MM-estimates are considered) when re-allocating between 0 and 60 min from SB to VPA. We considered the total MB composition to stand for 16 h, assuming 8 h of sleep a day. The predicted zBMI associated with a 60-min SB-to-VPA re-allocation was reduced by 0.30. However, due to the small value of R², the predictive power of these graph is rather limited.



Figure 7. Predicted zBMI for SB-to-VPA time re-allocations between 0 and 60 min for the mean MB composition (close to 16 h).

For example, the difference in predicted zBMI is 0.30 kg/m^2 after re-allocating 60 min. For an average girl resp. boy from our sample (15.1 years old with a height of 1.64 m resp. 1.71 m), this would mean a 2.05 kg resp. 1.91 kg reduction in weight. For more details, see Table 9.

Table 9. Predicted decrease in zBMI/BMI when re-allocating 15/30/45/60 min from SB to VPA.

Shift from SB to VPA (min)	15	30	45	60
Predicted zBMI (kg/m ²)	0.12	0.19	0.25	0.30
Weight reduction for an "average" girl (kg)	0.83	1.35	1.73	2.05
Weight reduction for an "average" boy (kg)	0.78	1.26	1.62	1.91

Note: SB, sedentary behavior; VPA, vigorous physical activity.

4. Discussion

This study presented a concise framework for robust CODA, mainly focusing on regression problems, in the presence of outlying observations, which often occur in real-world data sets. Although there is an increasing application of CODA methods in diverse fields of the natural [32,39,41–43] and

social sciences [17,44–47], classical statistical inference methods have been primarily used so far. Robust CODA methods have been used almost exclusively in geochemical applications [48–50]. This paper demonstrates the potential of robust CODA using movement behavior data of Czech adolescents. Robust counterparts have been also introduced for common compositional descriptive statistics, including a novel way to conduct robust data centering. The key point of robust statistical methods in a compositional context is that observations with large absolute values are not necessarily those to be downplayed, but rather the focus is on aberrant log-ratios between the parts of the composition. By comparing the classical and robust variation matrices (Tables 1 and 2), this seems to be particularly the case of pairwise log-ratios involving VPA in the present case study, which was already hinted by the ternary diagram visualization. Outliers affect classical tools such as the compositional mean barplot and ordinary LS regression; as well as the compositional centre and variation matrix used as basic descriptive statistics in CODA. The robust counterparts introduced in this work just focus on the majority of the data avoiding the extremes, hence providing more stable tools in the presence of outlying observations. They also help to reveal potential outliers, to be further analyzed and potentially suppressed. Moreover, robust centering highlights possible aberrant observations that would be masked in a classical case, where the center itself is affected by aberrant observations, particularly those close to the boundaries of the simplex. Robust regression also enabled to create a diagnostic plot to reveal vertical outliers and leverage points and enhance interpretation of the modelling; of course, similar diagnostics would also be possible for other robust regression models discussed in the manuscript. The present study demonstrated the role of orthonormal coordinates in the representation of compositional covariates, leading to the same goodness-of-fit measures and diagnostic plots for all possible choices of pivot coordinates.

The present study includes three main findings. Firstly, by using the classical compositional variation matrix, SB and LIPA showed the highest proportionality. These two parts also showed the highest proportionality (except for combination with sleep) in the study conducted by Carson et al. [16], where the same approach was applied. Using a robust compositional variation matrix, LIPA and MPA showed the highest proportionality, although SB and LIPA still retained a strong relative relationship. Moreover, both approaches have shown the highest log-ratio variances in cases where VPA was involved (reduced by using the robust approach though), indicating a possible presence of aberrant log-ratios. Although there were mostly small differences between the classical and robust variation matrices in this illustrative application (up to those variances where VPA was involved), it is well-established that the use of robust statistical methods in general tends to improve estimates in cases where an increasing presence of outliers distorts the information in the data set [36]. Secondly, an interesting feature was revealed when age was considered as an explanatory variable. Namely, with higher age the relative contributions of SB was expected to increase and VPA was expected to decrease. There is consistent evidence that SB increases with age in school-aged children and adolescents [51], but an increase in VPA with age is a rather unexpected finding. Given the higher number of outliers towards VPA (Figures 1 and 2), this might be explained by intense sport activities of some adolescents pulling up the trend, even when using robust regression to quantify the relationships. The effect of VPA can be better seen in Figure 2 using centered compositions, which helps to avoid the artifacts of the relative scale of the data near the borders of the ternary diagram and facilitates the interpretation of distances between points in an (almost) usual, Euclidean, sense. Regarding the fact that in the relative scale the ratios are responsible for dissimilarities between compositions, low proportions of some components (here MPA and VPA in particular with respect to SB and LIPA, respectively; see Figure 1) might also be a source of outliers, whose impact is reduced automatically by using robust centering, instead of the classical one in Figure 2. Finally, a significant relationship between VPA and adolescents' obesity was demonstrated. In the overweight/obese group, the relative time spent in VPA was reduced by 15.8% to the overall mean movement behavior composition, the highest reduction (using moderate-to-vigorous PA though) to overall mean composition in obese group was also found in [16].

In summary, this study has systematically considered and presented CODA regression methods which are, by construction, robust to the presence of outlying observations. They were discussed in relation to standard 'non-robust' regression models, as those applied in previous studies using movement behavior data [12,30,40]. Robust MM regression, which aims to retain robustness and resistance against outliers whilst gaining in statistical efficiency, is a proper tool for dealing with real-world data with natural presence of extreme observations; similar goals are also achieved by robust orthogonal regression. Accordingly, robust regression enables to extract the main trends in the data although we should be aware that from the methodological perspective, robust estimators are less efficient (precise) than the classical ones. At the same time, the associated diagnostics tools help to reveal deviating data points, which contributes to enhanced interpretation of the results. Although it is formally well-established that robust methods are generally less sensible to the presence of outlying observations, a simulation study would be necessary to determine and test for the significance of the differences in estimates between classical and robust models in an application. In our case study, we relied just on the general fact that robust estimators take into account majority of data and thus downgrade the influence of aberrant observations.

There are several strengths and limitations to this study that should be discussed. The availability of data of a relatively large number of individuals supported the statistical power of the regression significance tests and facilitated the application of robust statistical methods. Accordingly, trustworthy trends in movement behavior composition of adolescents were revealed, including those with respect to variables such as zBMI and age. The estimate of SB and PA used in this study was measured by the hip-worn ActiGraph accelerometer, the most commonly used validated accelerometer in PA research. However, due to the design of the study, the participants did not wear the accelerometer overnight, so we were not able to conduct an analysis on 24-h time budget data.

The accelerometers were set-up to record data at 60-s epochs, which can result in lower estimates of accelerometer-determined MPA and VPA [52]. The SB and PA intensities were only analyzed using Evenson cut-points. It is assumed that the decision of selecting among various cut-points significantly influences the levels of MPA and VPA [4]. Finally, BMI was calculated from self-reported height and weight. All this may introduce bias in relationship between SB, PA and obesity. Moreover, it was not possible to infer causality relationships from this observational cross-sectional study or totally discard the potential influence of unmeasured confounding variables.

5. Conclusions

We performed a concise robust statistical analysis of PA and SB data by using both common compositional descriptive statistics tools and robust regression modelling, which is a novel aspect of CODA modelling in physical activity research. The robust counterparts in the application case indicated that they might provide more stable estimates in the presence of outlying observations. Although there were no dramatic differences between classical and robust methods in this particular study, in a general case the impact of suppressing outlying observations might be substantial. The findings stressed that replacing time spent in SB with VPA can be a powerful tool against adolescents' obesity. It was demonstrated that there was a pattern in Czech adolescents' movement/non-movement behavior such that they spent more time in SB and VPA with increasing age and, consequently, less time in LIPA and MPA. Finally, it was shown that those with a greater proportion of time spent in SB also spent a considerable amount of time in LIPA and, accordingly, smaller amounts of time in MPA and VPA. It was observed that SB and VPA in particular were behaviors with high importance in the movement behavior composition in Czech adolescents, irrespective of their role as covariates or response in regression modelling, and they also drove the relative contributions of the related LIPA and MPA behaviors. Nonetheless, it could be challenging for policy makers to highlight the role of VPA and encourage such PA in the adolescents' population.

Author Contributions: N.Š. conducted the statistical analyses and wrote the first draft of the manuscript. J.D. supervised data collection and assisted with the writing of the manuscript. K.H. gave conceptual advice for the statistical analysis and assisted with the writing of the manuscript. A.G. provided consultation and guidance regarding obesity research. L.R. was involved in data collection and assisted with the writing of the manuscript. J.P.-A. reviewed and edited the manuscript. All authors approved the final version of the manuscript.

Funding: J.D., A.G. and L.R. were supported by a research grant by the Czech Science Foundation under reg. no. 18-09188S. N.Š and K.H. were supported by the grant COST Action CRoNoS IC1408. J.P.-A. was supported by the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS) and by the Spanish Ministry of Economy and Competitiveness under the project CODA-RETOS (MTM2015-65016-C2-1(2)-R).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Janssen, I.; LeBlanc, A.G. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* **2010**, *7*, 40. [CrossRef] [PubMed]
- 2. World Health Organization (WHO). *Physical Activity and Young People;* WHO: Geneva, Switzerland, 2015; Available online: http://www.who.int/dietphysicalactivity/factsheet_young_people/en (accessed on 19 August 2018).
- 3. McMahon, E.M.; Corcoran, P.; O'Regan, G.; Keeley, H.; Cannon, M.; Carli, V.; Wasserman, C.; Hadlaczky, G.; Sarchiapone, M.; Apter, A.; et al. Physical activity in European adolescents and associations with anxiety, depression and well-being. *Eur. Child Adolesc. Psychiatry* **2017**, *26*, 111–122. [CrossRef] [PubMed]
- Gába, A.; Dygrýn, J.; Mitáš, J.; Jakubec, L.; Frömel, K. Effect of accelerometer cut-off points on the recommended level of physical activity for obesity prevention in children. *PLoS ONE* 2016, 11, e0164282. [CrossRef] [PubMed]
- 5. Malina, R.M. Physical activity and fitness: Pathways from childhood to adulthood. *Am. J. Hum. Biol.* **2001**, *13*, 162–172. [CrossRef]
- 6. Telama, R.; Yang, X.; Viikari, J.; Välimäki, I.; Wanne, O.; Raitakari, O. Physical activity from childhood to adulthood: A 21-year tracking study. *Am. J. Prev. Med.* **2005**, *28*, 267–273. [CrossRef] [PubMed]
- 7. Gába, A.; Mitáš, J.; Jakubec, L. Associations between accelerometer-measured physical activity and body fatness in school-aged children. *Environ. Health Prev. Med.* **2017**, *22*, 43. [CrossRef] [PubMed]
- Carson, V.; Ridgers, N.D.; Howard, B.J.; Winkler, E.A.H.; Healy, G.N.; Owen, N.; Dunstan, D.W.; Salmon, J. Light-intensity physical activity and cardiometabolic biomarkers in us adolescents. *PLoS ONE* 2013, *8*. [CrossRef] [PubMed]
- 9. Chaput, J.P.; Saunders, T.J.; Carson, V. Interactions between sleep, movement and other non-movement behaviours in the pathogenesis of childhood obesity. *Obes. Rev.* **2017**, *18*, 7–14. [CrossRef] [PubMed]
- Tremblay, M.; LeBlanc, A.; Kho, M.; Saunders, T.; Larouche, R.; Colley, R.; Goldfield, G.; Connor Gorber, S. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* 2011, *8*, 98–119. [CrossRef] [PubMed]
- 11. Biddle, S.J.H.; Asare, M. Physical activity and mental health in children and adolescents: A review of reviews. *Br. J. Sports Med.* **2011**, *45*, 86–95. [CrossRef] [PubMed]
- 12. Chastin, S.F.M.; Palarea-Albaladejo, J.; Dontje, M.L.; Skelton, D.A. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: A novel compositional data analysis approach. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]
- 13. Aitchison, J. The statistical analysis of compositional data. J. R. Stat. Soc. Ser. B 1982, 44, 139–177. [CrossRef]
- 14. Zhu, W.; Ainsworth, B.; Liu, Y.L. A Comparison of Urban Black and White Women's Physical Activity Patterns. *Res. Q. Exerc. Sport* **2002**, *73*, A36.
- 15. Williams, S.M.; Farmer, V.L.; Taylor, B.J.; Taylor, R.W. Do more active children sleep more? A repeated cross-sectional analysis using accelerometry. *PLoS ONE* **2014**. [CrossRef] [PubMed]
- 16. Carson, V.; Tremblay, M.S.; Chaput, J.-P.; Chastin, S.F.M. Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. *Appl. Physiol. Nutr. Metab.* **2016**, *41*, S294–S302. [CrossRef] [PubMed]
- 17. Fairclough, S.J.; Dumuid, D.; Taylor, S.; Curry, W.; McGrane, B.; Stratton, G.; Maher, C.; Olds, T. Fitness, fatness and the reallocation of time between children's daily movement behaviours: An analysis of compositional data. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 64. [CrossRef] [PubMed]

- Dumuid, D.; Olds, T.; Lewis, L.K.; Martin-Fernández, J.A.; Katzmarzyk, P.T.; Barreira, T.; Broyles, S.T.; Chaput, J.P.; Fogelholm, M.; Hu, G.; et al. Health-related quality of life and lifestyle behavior clusters in school-aged children from 12 countries. *J. Pediatr.* 2017. [CrossRef] [PubMed]
- Dumuid, D.; Maher, C.; Lewis, L.K.; Stanford, T.E.; Martín Fernández, J.A.; Ratcliffe, J.; Katzmarzyk, P.T.; Barreira, T.V.; Chaput, J.P.; Fogelholm, M.; et al. Human development index, children's health-related quality of life and movement behaviors: A compositional data analysis. *Qual. Life Res.* 2018, 27, 1473–1482. [CrossRef] [PubMed]
- 20. Dumuid, D.; Stanford, T.E.; Pedišić, Ž.; Maher, C.; Lewis, L.K.; Martín-Fernández, J.A.; Katzmarzyk, P.T.; Chaput, J.P.; Fogelholm, M.; Standage, M.; et al. Adiposity and the isotemporal substitution of physical activity, sedentary time and sleep among school-aged children: A compositional data analysis approach. *BMC Public Health* **2018**, *18*, 311. [CrossRef] [PubMed]
- 21. Fairclough, S.J.; Dumuid, D.; Mackintosh, K.A.; Stone, G.; Dagger, R.; Stratton, G.; Davies, I.; Boddy, L.M. Adiposity, fitness, health-related quality of life and the reallocation of time between children's school day activity behaviours: A compositional data analysis. *Prev. Med. Rep.* **2018**, *11*, 254–261. [CrossRef] [PubMed]
- 22. Pedišić, Ž.; Dumuid, D.; Olds, T.S. Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: Definitions, concepts, statistical methods, theoretical framework, and future directions. *Kinesiology* **2017**, *49*, 252–269.
- 23. Mitáš, J.; Dygrýn, J.; Rubín, L.; Křen, F.; Vorlíček, M.; Nykodým, J.; Řepka, E.; Bláha, L.; Suchomel, A.; Feltlová, D.; et al. Multifaktoriální výzkum zastavěného prostředí, aktivního životního stylu a tělesné kondice české mládeže: Design a metodika projektu. *Tělesná Kult.* 2018. [CrossRef]
- 24. Cain, K.L.; Sallis, J.F.; Conway, T.L.; Van Dyck, D.; Calhoon, L. Using accelerometers in youth physical activity studies: A review of methods. *Phys. Act. Heal.* **2013**, *10*, 437–450. [CrossRef]
- 25. Evenson, K.R.; Catellier, D.J.; Gill, K.; Ondrak, K.S.; McMurray, R.G. Calibration of two objective measures of physical activity for children. *J. Sports Sci.* 2008, *26*, 1557–1565. [CrossRef] [PubMed]
- 26. Trost, S.G.; Loprinzi, P.D.; Moore, R.; Pfeiffer, K.A. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med. Sci. Sports Exerc.* **2011**. [CrossRef] [PubMed]
- 27. De Onis, M.; Onyango, A.W.; Borghi, E.; Siyam, A.; Nishida, C.; Siekmann, J. Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Heal. Organ.* **2007**, *85*, 812–819. [CrossRef]
- 28. Pawlowsky-Glahn, V.; Egozcue, J.J.; Tolosana-Delgado, R. *Modeling and Analysis of Compositional Data*; Wiley: Hoboken, NJ, USA, 2015; ISBN 9781119003144 | 9781118443064.
- Fišerová, E.; Hron, K. On the interpretation of orthonormal coordinates for compositional data. *Math. Geosci.* 2011, 43, 455–468. [CrossRef]
- 30. Hron, K.; Filzmoser, P.; de Caritat, P.; Fišerová, E.; Gardlo, A. Weighted pivot coordinates for compositional data and their application to geochemical mapping. *Math. Geosci.* **2017**, *49*, 797–814. [CrossRef]
- 31. Rousseeuw, P.J.; van Zomeren, B.C. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* **1990**, *85*, 633–639. [CrossRef]
- 32. Buccianti, A.; Grunsky, E. Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes? *J. Geochem. Explor.* **2014**, *141*, 1–5. [CrossRef]
- 33. Muller, I.; Hron, K.; Fiserova, E.; Smahaj, J.; Cakirpaloglu, P.; Vancakova, J. Interpretation of compositional regression with application to time budget analysis. *Austrian J. Stat.* **2018**, *47*, 3–19. [CrossRef]
- 34. Hrůzová, K.; Todorov, V.; Hron, K.; Filzmoser, P. Classical and robust orthogonal regression between parts of compositional data. *Statistics (Ber.)* 2016, *50*, 1261–1275. [CrossRef]
- 35. Abelson, R.P. A variance explanation paradox: When a little is a lot. *Psychol. Bull.* **1985**, *97*, 129–133. [CrossRef]
- 36. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods*; Wiley: Hoboken, NJ, USA, 2006; ISBN 9780470010921.
- 37. Yohai, V.J. High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.* **1987**, *15*, 642–656. [CrossRef]
- Hron, K.; Filzmoser, P. Exploring compositional data with the robust compositional biplot. In *Advances in Latent Variables: Methods, Models and Applications*; Carpita, M., Brentari, E., Qannari, E.M., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 219–226, ISBN 978-3-319-02967-2.

- 39. Von Eynatten, H.; Pawlowsky-Glahn, V.; Egozcue, J.J. Understanding perturbation on the simplex: A simple method to better visualize and interpret compositional data in ternary diagrams. *Math. Geol.* **2002**, *34*, 249–257. [CrossRef]
- 40. Dumuid, D.; Stanford, T.E.; Martin-Fernández, J.-A.; Pedišić, Ž.; Maher, C.A.; Lewis, L.K.; Hron, K.; Katzmarzyk, P.T.; Chaput, J.-P.; Fogelholm, M.; et al. Compositional data analysis for physical activity, sedentary time and sleep research. *Stat. Methods Med. Res.* **2017**. [CrossRef] [PubMed]
- 41. Pesenson, M.Z.; Suram, S.K.; Gregoire, J.M. Statistical analysis and interpolation of compositional data in materials science. *ACS Comb. Sci.* 2015, *17*, 130–136. [CrossRef] [PubMed]
- 42. Filzmoser, P.; Hron, K.; Reimann, C. Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Sci. Total Environ.* **2009**, 407, 6100–6108. [CrossRef] [PubMed]
- 43. Palarea-Albaladejo, J.; Martín-Fernández, J.A.; Olea, R.A. A bootstrap estimation scheme for chemical compositional data with nondetects. *J. Chemom.* **2014**, *28*, 585–599. [CrossRef]
- 44. Agerbo, E.; Sterne, J.A.C.; Gunnell, D.J. Combining individual and ecological data to determine compositional and contextual socio-economic risk factors for suicide. *Soc. Sci. Med.* **2007**, *64*, 451–461. [CrossRef] [PubMed]
- Campbell, G.P.; Curran, J.M.; Miskelly, G.M.; Coulson, S.; Yaxley, G.M.; Grunsky, E.C.; Cox, S.C. Compositional data analysis for elemental data in forensic science. *Forensic Sci. Int.* 2009, 188, 81–90. [CrossRef] [PubMed]
- 46. Leite, M.L.C. Applying compositional data methodology to nutritional epidemiology. *Stat. Methods Med. Res.* **2016**, 25, 3057–3065. [CrossRef] [PubMed]
- 47. Mert, M.C.; Filzmoser, P.; Endel, G.; Wilbacher, I. Compositional data analysis in epidemiology. *Stat. Methods Med. Res.* **2016**, 1–14. [CrossRef] [PubMed]
- Filzmoser, P.; Hron, K.; Reimann, C.; Garrett, R. Robust factor analysis for compositional data. *Comput. Geosci.* 2009, 35, 1854–1861. [CrossRef]
- 49. Filzmoser, P.; Hron, K. Robust statistical analysis of compositional data. In *Compositional Data Analysis: Theory and Applications;* Pawlowsky-Glahn, V., Buccianti, A., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2011; pp. 59–72, ISBN 9781119976462.
- 50. Filzmoser, P.; Hron, K. Robust statistical analysis. In *Robustness and Complex Data Structures*; Becker, C., Fried, R., Kuhnt, S., Eds.; Springer: Heidelberg, Germany, 2013; pp. 117–131, ISBN 9781119976462.
- 51. Tanaka, C.; Reilly, J.J.; Huang, W.Y. Longitudinal changes in objectively measured sedentary behaviour and their relationship with adiposity in children and adolescents: Systematic review and evidence appraisal. *Obes. Rev.* **2014**, *15*, 791–803. [CrossRef] [PubMed]
- Orme, M.; Wijndaele, K.; Sharp, S.J.; Westgate, K.; Ekelund, U.; Brage, S. Combined influence of epoch length, cut-point and bout duration on accelerometry-derived physical activity. *Int. J. Behav. Nutr. Phys. Act.* 2014, 11, 34. [CrossRef] [PubMed]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).