Individual Fundamental Diagrams from Single-File Movement Experiments

Sarah Ammenhämmer1,*, Maik Boltes1, and Armin Seyfried1,2

1Institute for Advanced Simulation, IAS-7: Civil Safety Research, Forschungszentrum Jülich, 52425 Jülich, Germany
2Faculty of Architecture and Civil Engineering, University of Wuppertal, 42285 Wuppertal, Germany
*Corresponding author. E-mail address: s.ammenhaeuser@fz-juelich.de

July 30, 2021

Highlights:
• Use of individual fundamental diagrams
• Analysis of the relationship between individual velocity and headway
• Multiple linear regression to show how variables affect the individual velocity
• Mixed models as an extension of regression analysis

Abstract: In recent years, numerous studies have been published dealing with the effect of individual characteristics of pedestrians on the fundamental diagram. These studies compared cumulative data on individuals in a group homogeneous in terms of one human factor such as age but heterogeneous in terms of other factors for instance gender. In order to examine the effect of all determined as well as undetermined human factors, individual fundamental diagrams are introduced and analyzed using multiple linear regression. A single-file school experiment with students of different age, gender, and height is therefore considered. Single individuals appearing in different runs are analyzed to study the effect of human factors such as height, age and gender and all other unknown individual effects such as motivation or attention to the individual velocity. The analysis shows that age and height are strongly correlated and, consequently, age can be ignored. Furthermore, the study shows that gender has a weak effect and other nonmeasurable individual characteristics have a stronger effect than height. In a further step, a mixed model is used as well as the multiple linear model. Here, it is shown that the mixed model that considers all other unknown individual effects of each person as a random factor is preferable to the model where the individual velocity only depends on the variables of headway, height, and all other unknown individual effects as fixed factors.

Keywords: Pedestrian dynamics, single-file movement, individual fundamental diagrams, regression analysis, mixed model

1 Introduction

Fundamental diagrams describe the relationship between density and velocity as well as density and flow of people. In the field of pedestrian dynamics, the diagrams illustrate various traffic conditions including free flow, bounded traffic, maximum flow and congested
traffic [10, 13]. Fundamental diagrams of various spatial structures, such as stairs, corridors or crossings, are different [5,9,16,23,26,31]. Furthermore, human factors such as age, height, gender, culture, and motivation, external factors such as visibility or background music as well as different type of flows like uni-, bi-, or multidirectional streams all affect the fundamental diagram. These diagrams were examined in various studies, enabling comparisons of different cultures or people of different ages or gender. To date, factors such as height, gender, income, or culture have been studied at macroscopic and microscopic levels [6, 8, 11, 17, 18, 20, 21, 28–30, 33, 34]. Further studies on single-file movement have considered external factors such as rhythm or background music [27, 30], restricted visibility, [28], luggage and trolleys [11], or properties of human locomotion such as step length and frequency [8,15,19,24,25,29].

The various factors affecting the fundamental diagram illustrate the complexity of the quantitative description of pedestrian flow. In order to explore the effect of different human factors, the study presented in this article is limited to the simplest system i.e., the movement of pedestrians in single file, as can be observed in queuing systems. In the following sections, we will focus on single-file studies that consider age as a human factor.

Cao [7] compared three groups, a younger, an older, and a mixed group. The group composed of younger people shows higher velocities than the mixed or older group for low and densities up to $\rho = 1.5$ [1/m]. For high densities close to the stopping density, the speed of the mixed group is lower than that of the younger group. There is no comparative date for the group of older people. The fundamental diagram of the mixed group also has a more complex structure. This is illustrated by the fact that the diagrams cannot be transformed into each other by scaling the variables to dimensionless quantities. The diagram for the relationship between headway and velocity for the mixed group shows three regimes, the free, weakly constrained, and strongly constrained regime, whereas for the younger and older groups of people only two regimes can be observed. Between the younger and older group, on the other hand, only slight differences remain after the scaling procedure. A study by Ren et al. [17] confirms these results. The elderly group is slower than the group with young adults and also the comparison between the mixed group and the group composed of elderly people shows no differences. Moreover, in a headway vs. velocity diagram, a group of elderly Chinese and a group of French students [13] are compared and three regimes occur. Ren also concluded that the pedestrian dynamics are affected by factors such as age, heterogeneity of the group, and familiarity. Zhang et al. [32] compared two groups of middle-aged individuals, with a low and good income. The fundamental diagrams for these groups are different but they show the same trend. One group with a good income and a higher number of female adults is more inactive and the participants prefer to maintain a greater distance from others or to keep pace with others. This group is more homogeneous, and more jams and stop-and-go waves occur. The other group with a low income and an approximately equal ratio of males and females is more active so the flow rate is higher.

With regard to the age factor, Subaih et al. [21] compared groups of different gender in experiments at high densities performed in Palestine and China. The authors concluded that older Chinese pedestrians walk as fast as young Palestinians but a group of younger Chinese people walk faster than younger Palestinians. A further study focusing on age is presented by Ziemer [35]. She analyzed experiments in schools by comparing students from the fifth grade with those from the 11th grade. The study shows that age has no effect on fundamental diagrams even if these groups have significantly different body heights. Groups that are heterogeneous in age also have no effect on the diagram. As to the question of whether the fundamental diagram has two or three regimes, we refer to Figures A18, A21e and A21g in [35]. These results indicate that the method of data binning is crucial.

From these six studies [7,13,17,21,32,35], it can be seen that age might have an effect on fundamental diagrams, but it does not normally. It depends on the age of the group whether it has an effect or not. Moreover, the discussion shows that besides age, factors
such as culture, income, gender, and the homogeneity of the group composition could not be excluded. Also, there are indications that fundamental diagrams of heterogeneous groups have three regimes. But it depends on the binning method whether there are two or three regimes.

Ren [17] shows that not only does age matters, but also the group composition in relation to the heterogeneity of the group in terms of gender or culture. Regarding Zhang [32], the question arises whether the differences observed are affected by the different income or the gender composition of the group. In Subaih et al. [21], it is not clear whether age, culture, or the heterogeneous group composition with regard to gender have the main effect on the fundamental diagram.

The discussion above gives a highly contradictory picture of how human factors affect the fundamental diagram of pedestrian dynamics. Even if these studies are performed under well-controlled laboratory conditions, the methodological problem is that even if a group is homogeneous in terms of one factor, it might be heterogeneous in terms of other factors.

In the above-mentioned studies, measurements of the velocity and density of individuals were made. For the comparison of the fundamental diagrams, however, only the cumulative data of all individuals in the group were used. To solve this problem, individual fundamental diagrams will be introduced and a multi-factorial analysis will be performed to study the effect of human factors. It will also be taken into account that certain factors could be strongly correlated.

The structure of the paper is as follows. Section 2 describes the experimental setup, the measurement methods, and the data preparation. Then Section 3 deals with the regression analysis which includes the simple and the multiple linear regression and the mixed model. The last section highlights the conclusions and interprets the results. Finally further possible research steps will be proposed.

2 Materials and Methods

2.1 Experimental setup

In the present paper, a one-dimensional single-file experiment [12] performed at the school Gymnasium Bayreuther Straße (GBS) in Wuppertal, Germany in 2014 is analyzed. The spatial structure for the experiment is an oval path with a total length of the central line of \( l = 16.62 \) m. The dimensions of the experiment can be seen in Figure 1 below. The oval path has a width of \( w = 0.8 \) m and each straight section has a length of 2.5 m. To reduce the complexity of the system the two-dimensional trajectories were mapped to one dimension. For this purpose, the participants’ trajectories were projected on the middle
line of the oval. Thus, only the change in movement direction over time is considered according to Ziemer [35].

A total of 31 runs were performed with different global densities $\rho_{gl} \in [0.32, 3.20] [1/m]$. These densities are calculated using Equation 3 because unrolled one-dimensional data are considered. An overview of the global densities for individual runs can be seen in

| Run  | min. FR | max. FR | $I_{min}$ | $I_{max}$ | N | m | f | $\rho_{gl}$ | $\sigma_{\rho}$ |
|------|---------|---------|----------|----------|---|---|---|-----------|-----------|
| GBS-5th grade | 1101 | 0 | 2344 | 175 | 1940 | 16 | 8 | 8 | 1.02 | 2.01 ± 0.38 |
| | 1102 | 0 | 3078 | 0 | 3078 | 50 | 28 | 22 | 3.20 | 1.06 ± 0.82 |
| | 1103 | 0 | 4175 | 50 | 4175 | 40 | 20 | 20 | 2.56 | 2.02 ± 0.73 |
| | 1104 | 0 | 3463 | 0 | 3463 | 52 | 30 | 12 | 2.05 | 2.50 ± 0.51 |
| | 1105 | 0 | 5904 | 300 | 2625 | 10 | 6 | 4 | 0.64 | 1.72 ± 0.49 |
| | 1106 | 0 | 2977 | 10 | 2977 | 24 | 13 | 11 | 1.54 | 2.48 ± 0.46 |
| | 2101 | 0 | 1941 | 150 | 1941 | 24 | 11 | 11 | 1.54 | 1.99 ± 0.39 |
| | 2102 | 0 | 3078 | 150 | 3078 | 5 | 5 | 5 | 3.20 | 2.00 ± 0.54 |
| | 2103 | 0 | 3432 | 250 | 3432 | 11 | 4 | 7 | 0.70 | 1.99 ± 0.73 |
| | 2104 | 0 | 3885 | 300 | 3885 | 34 | 16 | 18 | 2.18 | 2.72 ± 0.69 |
| | 2105 | 0 | 2968 | 300 | 2966 | 16 | 8 | 8 | 1.62 | 2.04 ± 0.37 |
| | 2106 | 0 | 1879 | 20 | 1879 | 28 | 18 | 10 | 1.79 | 1.10 ± 0.45 |
| | 2107 | 0 | 1115 | 100 | 1115 | 27 | 18 | 9 | 1.73 | 2.44 ± 0.71 |
| | 2108 | 0 | 2234 | 200 | 2083 | 14 | 5 | 9 | 0.90 | 1.09 ± 0.23 |
| GBS-11th grade | 1201 | 0 | 3370 | 0 | 2500 | 39 | 15 | 24 | 2.50 | 1.27 ± 1.06 |
| | 1202 | 0 | 1969 | 150 | 1726 | 5 | 1 | 4 | 0.32 | 0.99 ± 0.39 |
| | 1203 | 0 | 3450 | 0 | 2700 | 33 | 14 | 19 | 2.11 | 1.10 ± 0.61 |
| | 1204 | 0 | 2693 | 300 | 2560 | 12 | 5 | 7 | 0.77 | 1.48 ± 0.41 |
| | 1205 | 0 | 2859 | 50 | 2400 | 23 | 9 | 14 | 1.47 | 1.14 ± 0.43 |
| | 1206 | 0 | 2290 | 300 | 2030 | 15 | 5 | 10 | 0.96 | 1.83 ± 0.26 |
| | 1207 | 0 | 2446 | 100 | 1900 | 21 | 5 | 16 | 1.34 | 1.55 ± 0.40 |
| | 2201 | 0 | 2575 | 200 | 2476 | 5 | 3 | 2 | 0.32 | 1.30 ± 0.32 |
| | 2202 | 0 | 2735 | 0 | 1800 | 39 | 17 | 22 | 2.50 | 1.31 ± 0.63 |
| | 2203 | 0 | 2955 | 220 | 2786 | 9 | 3 | 6 | 0.58 | 1.31 ± 0.18 |
| | 2204 | 0 | 3044 | 0 | 2400 | 29 | 13 | 16 | 1.86 | 1.17 ± 0.41 |
| | 2205 | 0 | 3347 | 150 | 3058 | 15 | 6 | 9 | 0.96 | 1.75 ± 0.27 |
| GBS-5th + 11th grade | 1301 | 0 | 3571 | 0 | 2600 | 42 | 20 | 22 | 2.69 | 0.81 ± 0.17 |
| | 1302 | 0 | 3523 | 0 | 2600 | 44 | 25 | 19 | 2.82 | 1.02 ± 0.73 |
| | 1303 | 0 | 2277 | 100 | 2236 | 5 | 2 | 3 | 0.32 | 1.13 ± 0.13 |
| | 1304 | 0 | 3535 | 0 | 2600 | 33 | 17 | 16 | 2.11 | 2.12 ± 1.12 |
| | 1305 | 0 | 1924 | 100 | 1767 | 11 | 5 | 6 | 0.70 | 2.36 ± 0.67 |

Table 1: The columns from left to right show a detailed overview of the runs for the different groups of students from fifth grade, 11th grade and both fifth and 11th grade and their general properties.

Table 1 above. The corresponding run is listed with the total number of persons N and the number of male m and female f pedestrians. The experiment is divided into several runs. In total, 118 different students participated in the experiment, with around 46 % male pedestrians. Each subject was given a main ID in order to identify a particular student in different runs. This identification and the assignment of gender was realized manually. In addition, the minimum and maximal frame, min. FR and max. FR, of a run are given. The information is used to illustrate the total length of the video [12]. Moreover, the manually selected time limits of the steady state $I_{min}$ and $I_{max}$ can be found in Table 1 for each run. The steady state area only represents the duration of the frame interval, which was used for the analysis. The framerate of the videos is 25 fps.

The runs of the experiment show different configurations, such as a homogeneous group of younger students, a homogeneous group of older students, and mixed groups of younger and older students, whereby the ratio is 1:1. The subjects are from the lower and upper grades, fifth and 11th grade. Thus, it can be investigated whether there are differences in age or gender beyond different group compositions. The younger students
are aged between 11 and 12. The older ones are about 17 or 18. The students were introduced to the topic in teaching units.

The participants were advised to walk behind each other without haste and without overtaking. The students wear colored caps on their heads, which are used on the one hand to extract the trajectories and on the other hand to indicate different intervals of body height [2,4]. In total, there are five different intervals to analyze the effect of height. The average height of younger students is 1.48 m ± 0.04 m and that of older students is 1.76 m ± 0.07 m. See [35] for further information about the experiment.

2.2 Measurement methods

Based on the one-dimensional trajectories gained by tracking the head from video recording, the individual velocity and density as well as the headway are calculated using the software JPSreport [14]. For the one-dimensional case, the position of one individual i at time t is defined by \( x_i(t) \) whereby t is in the time interval [min. FR, max. FR]. By dividing the frame t by the framerate 25 fps the time is given in seconds.

The headway \( h_i(t) \) of student i at time t is calculated by Equation 1 which describes the distance between the centers of the heads \( x_i(t) \) and \( x_{i+1}(t) \) whereby \( x_i(t) \) is the coordinate of pedestrian i at time t and \( x_{i+1}(t) \) is the position of the person \( i+1 \) at time t walking in front of person i.

\[
h_i(t) = x_{i+1}(t) - x_i(t) \tag{1}
\]

The values for the individual density for any pedestrian \( i \) at time t based on the one-dimensional Voronoi tessellation \( \rho_{V,i}(t) \). In Equation 2 \( d_{V,i}(t) \) is half of the distance between the centers of the follower \( x_{i+1}(t) \) and \( x_{i-1}(t) \) the person in front.

\[
\rho_{V,i}(t) = 1/d_{V,i}(t), \quad \text{where} \quad d_{V,i}(t) = (x_{i+1}(t) - x_{i-1}(t))/2 \tag{2}
\]

The formulae for the headway \( h_i(t) \) as well as \( d_{V,i}(t) \) for the individual density are based on the one-dimensional Voronoi tessellation illustrated in Figure 2 below. Furthermore,

\[
v_i(t) = \frac{x_i(t + \Delta t) - x_i(t - \Delta t)}{\Delta t} \tag{4}
\]

Figure 2: Illustration of the formulae for the distance based on the headway \( h_i(t) \), on the one-dimensional Voronoi tessellation \( d_{V,i}(t) \) and the length of the measurement area \( l_m \).

This figure shows the length of the measurement area \( l_m \) which is in the interval from 0.5 m to 16.12 m and is highlighted in gray. The black line here represents the central line in the oval. The global density is calculated by the ratio of the number of persons \( N \) in one run in the measurement area and the length of the measuring area \( l_m = 15.62 \) m, Equation 3.

\[
\rho_{gl} = N/l_m \tag{3}
\]

The individual velocity is also calculated using the software JPSreport. It applies:
It should be noted that the default value $\Delta t$ which describes the difference between two time points is selected in such a way that the oscillations of the trajectories caused by the movement in steps are smoothed out and so these do not have to be taken into consideration when analyzing the autocorrelation of the velocity. Thus $\Delta t = 20$ frames = 0.8 s has been selected. The time steps are therefore discrete time steps. The intended direction of the students is also included. The reason for this is that, as can be seen in Figure 3, significant oscillations occur at a low velocity. Even if a pedestrian stops, their head moves, for example as a result of changing the leg they are standing on and thus the trajectory shows movement also in a negative $x$-direction. Hence, negative velocities can also be observed in the runs. This is shown, among other things, in the interval of $x \in [9.5, 10.5]$[m] in Figure 3 below.

2.3 Data processing

During the analysis, neither the beginning and the end of each run are taken into account, which means that only the manually selected steady state of the given run is considered. In other words, all frames between $I_{\text{min}}$ and $I_{\text{max}}$ are used see (Table 1). In Figure 4, the straight red lines illustrate the boundaries of the stationary area.

To ensure independent measurements in the regression analysis, the data are reduced by taking into consideration the autocorrelation. Here, we examine how strongly the
observations of the individual velocity depend on each other in the case of a time lag. In general, the values of the autocorrelation function (acf) are defined by Equation 5, where $\tau$ is the time lag, $n_i$ the number of observations of individual $i$ and $i = 1, \ldots, N$ the number of persons.

$$r_{i,\tau} = \frac{\sum_{t=\tau+1}^{n_i} (v_{i,t-\tau} - \bar{v}_i)(v_{i,t} - \bar{v}_i)}{\sum_{t=1}^{n_i} (v_{i,t} - \bar{v}_i)^2}$$  \hfill (5)

Figure 5 below shows the autocorrelation for different $\Delta t$ for an individual $i$ in one run and illustrates the correlation between values at different time lags. The time lag $\tau$ is represented by the $x$-axis. The values of the autocorrelation can be seen from the $y$-axis and show how similar the shifted series $v_i(t - \tau)$ is to the original series $v_i(t)$.

Figure 5: Autocorrelation based on the individual velocity for the person with the main ID 29 in run 2106. The autocorrelation for $\Delta t = 0.8$ s (black curve) has smaller oscillations to those for $\Delta t = 0.4$ s (red curve). For this person the time lag $\tau$ for $r_{i,\tau} < 0.3$ is around 0.93 s.

For small time steps $\Delta t$ in Equation 4 for the individual velocity, the autocorrelation in Figure 5 would show significant oscillations. This is illustrated by the comparison between the red curve for $\Delta t = 0.4$ s and the black curve for $\Delta t = 0.8$ s. [22], presents the autocorrelation based on all participants in a complete run and different values for the time step $\Delta t$ for the individual velocity. It can be seen that with increasing time step, the autocorrelation has a smaller oscillation and the curve becomes smooth enough at about 0.8 s. The oscillations of the autocorrelation function are caused by the bipedal movement. The duration of one pedestrian step is approximately 0.7 s [22]. Thus, for smaller values of $\Delta t$, the autocorrelation curve relates to the individual periods of the individual single steps. When the oscillations are smoothed out as the frame step increases, the value for $\delta_t$, which is the individual specific time lag between the observation points in each run, can be taken as the value for which $r_{i,\tau} < 0.3$ applies for the first time.

Consequently, the analysis is restricted to data points with a weak correlation. This decision is applied to the length of the individual time steps between the observation points in each run, for each individual. Thus, the analysis is based on individual-specific $\delta_t$ time steps.

Table 1 shows the mean values $\overline{\delta_t}$ of the different students in each run and their standard deviation. The values in the table show that the times are lower for older students than for younger students. Figure 6 below shows the significant difference between the distribution of the individual time steps $\delta_t$ of the younger and older students. The red
curve represents the younger and the blue curve the older students. The difference might

![Figure 6: Distribution of individual time steps $\delta_t$ of younger and older students to ensure independent measurements.](image)

be due to the fact that the younger students take smaller steps and therefore, a higher step frequency occurs. Also, the fluctuations of the individual steps are not smoothed out as much for young students as a result of the choice of $\Delta t = 0.8$ s compared with the older students. Accordingly, the autocorrelation function would reach the value $r_{i,\tau} < 0.3$ in comparison to the older students only at a later stage, since the autocorrelation function decreases more slowly due to the increased fluctuation caused by taking smaller steps. In general, these first steps of the data processing show that there is no relationship between the mean values $\bar{\delta}_t$ and the density. Both short and long intervals occur with any density. Regarding the composition of the group, there are only differences between the younger and older students in that the individual-specific $\delta_t$ time steps are smaller for the older students.

Now the data set can be used for further data analysis, including the regression analysis and the analysis of the individual fundamental diagrams but in order to obtain a higher number of data for each student for the study, the individual runs are connected to one data set. Due to the link between the different runs and the individual-specific $\delta_t$ time steps, the number of observations $l = 1, \ldots, n_i$, for each individual $i$, differs. Using the main ID of each person, we documented which runs with different densities an individual was involved in. In addition, the data for each individual is based not only on the different densities but also on various individual velocities and neighboring pedestrians as well as others preceding or following them. Thus, our observations are not only the result of a simple run and its composition but probably also represent a characteristic individual property such as preferences for certain individual velocities or for certain distances based on different neighboring students.

3 Results

3.1 Data selection

In this section, the relationship between velocity and headway is analyzed. In addition, we examine which factors significantly affect the velocity and which can possibly be ignored.

The diagrams in Figure 7 below illustrate exemplary the relationship between headway and the individual velocity for a certain main ID of one individual. The data show that
there are various piecewise linear sections. However, it is not clear whether these are two
or three different regimes but, in general, the diagrams show two different approximately linear branches for all individuals. Due to the shape of the branches, a linear model is applied for this analysis. We decided to study the branch where the velocity is affected by neighboring pedestrians. Accordingly, the effect on the free velocity $v_{i,0}$ is not analyzed.

The section for which the regression analysis will be performed to examine the effect of independent variables on the individual velocity is $h_i < 1.5$ m. The beginning of the area selected for the free velocity is supported by studies by Ziemer [35] and Cao [6], who also examined younger age groups, showing that the free flow branch starts at $h \approx 1.5$ m.

### 3.2 Linear regression analysis and analysis of scattering

After the data selection, the headway vs. individual velocity diagrams show an approximately linear trend. To illustrate the relationship between headway and individual velocity for each individual $i$, a simple linear regression analysis was performed. Depending on the number of individual observations $n_i$, the following formula results:

$$v_i = \beta_0 + \beta_1 \cdot h_i + \epsilon_i, \text{ where } l = 1, ..., n_i$$  \hspace{1cm} (6)

The velocity is represented by $v_i$, and $h_i$ is the headway. In Equation 6, $\epsilon_i$ describes the random experimental error which should have a small scattering and $\beta_0$ and $\beta_1$ are unknown regression coefficients.

In order to construct the regression line which should be close to all values in order to represent the data as accurately as possible, the values $\beta_0$ and $\beta_1$ need to be estimated:

$$\hat{v} = \hat{\beta}_0 + \hat{\beta}_1 \cdot h,$$  \hspace{1cm} (7)
whereby $\hat{\beta}_1$ and $\hat{\beta}_1$ are the estimated values. Furthermore, Equation 7 gives the slope of the regression line as $\hat{\beta}_1$ and by transforming the formula for $\dot{v} = 0$, we obtain the minimum headway for each individual $i$:

$$d_{\text{min}} = -\frac{\hat{\beta}_0}{\hat{\beta}_1}$$  \hspace{1cm} (8)

**Figure 8** below illustrates that at $\dot{v} = 0$, larger minimum distances occur for the older students. The comparison between the male and female students shows that the minimum distance is slightly greater for female students. However, this difference is less pronounced. The slope $\beta_1$ of the regression line is related to the stimulus response mechanism connected to reaction time and the ability to accelerate and brake. The comparisons between younger and older and between male and female students can both be considered (see Figure 9 below). Here, there are virtually no differences.

**Figure 9**: The left-hand figure shows the distribution of $\hat{\beta}_1$, the reaction time and the ability to accelerate and brake, for younger and older students and the right-hand one minimum distances for male and female students at $\dot{v} = 0$.

Next, the scattering around the regression line was analyzed. When the whole group is divided into younger and older students, as well as into females and males, it becomes
clear that in all groups there are headway vs. individual velocity diagrams with low and high scattering. In Figure 10 below, for each of the two groups, younger and older students, four representative main IDs were selected to illustrate this.

Figure 10: Headway vs. individual velocity diagrams with a regression line for representative younger students in grade five on the left-hand side and for older students in grade 11 on the right. These two groups are divided into male and female main IDs, with the numbers in the orange box representing the different main IDs. High and low scattering occurs for each gender and age group.

Moreover, the different points represent the measured values \( l = 1, ..., n_i \) for different individuals \( i \). The regression line is also shown see (Equation 7). The left-hand figure illustrates the younger pedestrian group and the right-hand one the group of older students. Males are represented in blue and females in red.

In order to examine whether the scattering around the regression line is higher for older or younger students as well as for female or male students, the correlation between the headway and the individual velocity was studied (see Figure 11). Figure 10 shows a tendency for a larger scattering for younger participants and the histograms in Figure 11 support this hypothesis. The mean correlation coefficient \( r_{h,v} \) of the younger group is 0.77 and 0.82 for the older group. When a distinction is made between male and female students, it becomes apparent that for both younger and older students, the group correlation coefficient is so low that there is no significant difference. For the young male students, the value is 0.76 and for the young female students it is 0.77. In contrast to the young group, the values for the older group are 0.80 for the male students and 0.84 for the female students. Considering all female and male students separately, it is apparent that the scattering is slightly larger for male students (see Figure 11).

In general, the linear regression diagrams representing headway vs. individual velocity which illustrate different individuals show that for \( h < 1.5 \) m, there appears to be virtually no difference between gender and age. It can only be assumed that the scattering is larger among younger students. This assumption is confirmed when we look at the histograms (see Figure 11) showing the correlation coefficients of the individuals in the various groups, for instance younger and older students. Furthermore, there is no significant difference between male and female students. A comparison of the minimum distances at \( \hat{v} = 0 \) also shows that there are differences between younger and older students, i.e., distances are greater for older students. When we compare male and female students, the differences are less pronounced. Moreover, a comparison of \( \hat{\beta}_1 \), the reaction time and ability to accelerate and brake, between younger and older as well as between male and female students shows that there are virtually no differences.
Figure 11: Distribution of the correlation coefficient between headway vs. individual velocity for younger and older students and for female and male students.

3.3 Multiple linear regression

3.3.1 Model selection

The model structure for the multiple linear regression analysis is explained below. The goal is to find a model that takes into account all relevant independent variables for the study of the dependent variable. The individual velocity is \( v_m \), where \( m = 1, \ldots, n \) and \( n \) is the number of all observations of all individuals. At the same time, the model should include as few variables as possible. First, the individual characteristics measured, headway \( h \), age, height, and gender are introduced as independent variables. The variable \( alloence \) is used to take into consideration all other unknown individual effects, for example, motivation, attention, or excitement. In addition, the independent variables should not be strongly correlated with each other and should ideally be linked to the velocity. The variables height and age are strongly correlated \( (r_{\text{height,age}} = 0.89) \) and so it is sufficient to include only one of these. Which one depends on the research question and the quality of the data, for example granularity. Here, the height is used, since age was only measured as a binary variable with 0 representing the younger students and 1 the older ones, while height was categorized according to five levels. There is either no or only a weak correlation \( (r_{x,y} < 0.29) \) between the other independent variables. In the following multiple linear models, the variables are considered without units. The first full model that considers all measured individual characteristics is:

\[
\text{Model I: } v_m = \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m + \epsilon_m
\]

This model allows us to analyze which of the two individual characteristics, height or gender, has a stronger effect on the fundamental diagram. It is important to note that this research question is different from the following question: How strongly do individual characteristics affect the fundamental diagram? To answer this second question, it should be tested whether other variables that would improve Model I and affect the dependent variable \( v_m \) were ignored. In comparison to the first model, a further model which takes all other unknown individual effects into account by also including the variable \( alloence \)
is introduced:

$$\text{Model II: } v_m = \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m + \sum_{i=1}^{N} \beta_{3i} \cdot \text{alloence}_m + \epsilon_m,$$  

where alloence$_m = 1$ for all $m$ belonging to individual $i$ and 0 for all other $m$.

$R^2$ is used to compare the quality of the models. It indicates how useful the independent variables are for explaining the variance of the dependent variable. The value varies between 0 % and 100 %. The higher the value, the more accurately the independent variables are able to predict the dependent variable [1]. 69.40 % of the variance of the variable velocity can be explained by Model I and 73.20 % by Model II. Therefore, the variable alloence, describing all unknown individual effects, has a significant contribution. This result also shows that Model I does not necessarily include all relevant individual characteristics used to describe the fundamental diagram.

In a next step, it was analyzed which of the variables used in Equation 10 should be considered to obtain the best possible model with as few variables as necessary. One method for making this decision is the model evaluation using Akaike’s Information Criterion (AIC) [3]. Here, it was decided step by step whether a model improvement can be achieved by omitting independent variables in Equation 10 step by step. The lower the AIC value, the better the model. This method ends when no further reduction is useful. However, it should be noted that the method does not generally provide absolute criteria for deciding which model is the better choice. By applying it to Model II, the steps of the AIC procedure indicate that gender can be omitted. Interestingly, the same procedure for Model I shows that the model should not be reduced and gender has a significant effect. This result also shows that only taking into consideration of nonmeasured individual characteristics allows a reduction of the model to Model III.

$$\text{Model III: } v_m = \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{height}_m + \sum_{i=1}^{N} \beta_{3i} \cdot \text{alloence}_m + \epsilon_m,$$  

The result of the Akaike’s Information Criterion is consistent with the statement made based on the analysis of the minimum distance, the reaction time, and the correlation between headway and individual velocity in section 3.2 above. The difference between younger and older students is greater than between male and female students where there is virtually no difference. Since the difference between men and women is minimum and other individual characteristics predominate, gender can be ignored in Model II. However, Model I does not include the other individual characteristics, so gender cannot be ignored due to its minimum impact. Based on Model III, the following estimated regression coefficients are calculated in Equation 12.

$$\hat{v} = 0.23 + 0.98 \cdot h - 0.34 \cdot \text{height} + \sum_{i=1}^{N} \hat{\beta}_{3i} \cdot \text{alloence}$$  

It can be seen from $\hat{\beta}_0 = 0.23$, $\hat{\beta}_1 = 0.98$, and $\hat{\beta}_2 = -0.34$ that changes in every predictor variable are significantly associated with changes in velocity. The distribution of the values for $\hat{\beta}_{3i}$ are illustrated in Figure 12. The variable alloence has a positive or negative and stronger or weaker effect on the individual velocity in depending on the different regression coefficients $\hat{\beta}_{3i}$ for the individuals. The values are weak but the effect is mostly statistically significant, $p < 0.05$.  

In addition, it was also examined whether there are different results or models when the data for younger and older students are analyzed separately. This is not the case.
Regression coefficients $\beta_{3i}$ of all other unknown individual effects for each main ID.

3.3.2 Analysis of Variance (ANOVA)

Next, Model III (see Equation 11) is used to analyze the effect of the independent variables, headway $h$, height, and alloence on the velocity. On the basis of the ANOVA table, it becomes clear that all variables considered in Model III have an effect on the individual velocity. Any effect is significant because the $p < 0.05$. Figure 13 below illustrates the various effects on the individual velocity based on the ANOVA table in a pie chart. The headway has the largest effect. This is followed by all other unknown individual effects. The height has the lowest effect at 1 %. The same analysis for Model II (see Equation 10) shows that gender has a smaller effect than height. Even without taking headway into account, no larger effect could be attributed to the other independent variables.

In a further step, the residuals are examined for first-order autocorrelation using the Durbin-Watson test. This test enables us to check whether potentially relevant effects are ignored. Here, the null hypothesis, $H_0: \rho_1 = 0$, where $\rho_1$ is the theoretical autocorrelation coefficient, implies that the error term is not autocorrelated. If the null hypothesis is rejected, it can also be decided whether positive or negative autocorrelation occurs. If the null hypothesis is violated, the error term does not fulfill the standard assumptions of the multiple linear regression model. The Durbin-Watson statistics shows a value of 0.7077. Consequently, the null hypothesis is rejected. In addition, there is a positive autocorrelation due to the low $DW$ value and the limits for the critical values of the Durbin-Watson statistics. Therefore, the headway has the most significant effect on the...
individual velocity, but besides height and all other unknown individual effects there
must be other factors that cannot be omitted as an effect on the velocity. This indicates
that, for example, acceleration phases must be distinguished from braking phases or the
stimulus response mechanism connected to reaction time and ability to accelerate, or the
locomotion in steps should be modeled more carefully.

3.4 Mixed Model

Finally, a mixed model which, unlike the previous models, considers fixed and random
factors is used for the analysis. Fixed factors are independent variables that are observ-
able, such as headway, height, or gender. The random factors are not observable and may
obscure the effect of the factors of interest. These might be individual characteristics
included in the variable *alloence* that were not considered as fixed factors or characteris-
tics such as the attention or motivation of an individual. This model can be applied to
multilevel data, from several observations of an individual or a group and is relevant for
the analysis of correlated data. Here, the following problem is addressed. The models
used previously include the variable *alloence* as a fixed factor. Now it will be used as a
random factor. Thus, it is considered as an individual factor which could be correlated
with height or gender, too. As shown above, gender has a potentially small effect and is
ignored in this analysis.

Two models are compared: a simple model which only includes fixed factors (see
Equation 11) and a mixed model. The aim is to establish which model leads to an
improvement and should consequently be used. Therefore, a $\chi^2$-test is performed, which
checks whether the mixed model in which the individual effects may obscure the effect of
the factors of interest is more efficient than the simple model. If the null hypothesis is
rejected, the mixed model is preferred to the simple model. This is the case if $p < 0.05$
applies. In the present analysis, $H_0$ is rejected so we can conclude that the mixed model
in which the individual effects are included as a random factor is preferable to the simple
model. Accordingly, it is better to use mixed models for a multivariate analysis in this
context: otherwise, the effect of the factors of interest may be obscured.

4 Conclusions

The comparison of data provided in the literature for single-file studies considering age as
a human factor show that it has an effect on fundamental diagrams. However, a contradic-
tory picture of how age affects the fundamental diagram was found. One possible reason
for these contradictions is that the different experiments are not comparable because even
if a group is homogeneous in one factor, it could be heterogeneous in other factors such
as gender. To date, only the cumulative data on all individuals in the group have been
used for a comparison of the fundamental diagrams. In a new approach, the present
study is limited to the simplest system of a one-dimensional single-file school experiment
and individual fundamental diagrams were introduced. Moreover, more factors such as
height, gender, and age describing the participants on an individual level are measured
and used for a multiple regression analysis.

The focus of the data analysis is on the research question how strongly individual
characteristics affect the fundamental diagram. Using simple linear regression of individ-
ual fundamental diagrams, we analyze the minimum distance, the reaction time, and the
scattering of the data. The results show that more significant differences occur between
older and younger students whereas the differences between male and female students are
virtually non-existent. According to this, age has a stronger effect than gender.

Furthermore, model selection steps of the multiple linear regression analysis were
performed. The models for the velocity depend on headway, gender, height, and *alloence*.
The latter is introduced to consider all individual factors which could not be measured,
such as motivation or attention. The analysis shows that the variable *alloca* is crucial and gender could only be omitted in the model when the unknown factor is not ignored.

The analysis of the impact of the variables confirms that the headway is the most crucial factor. This is followed by all other unknown effects and height in particular has a very low percentage. Therefore, the indefinable factors have a greater effect than height or gender. The result is also reflected in the previous statement, as the differences between genders are smaller than between younger and older students. In a further step, an examination of the residuals shows that the individual velocity is not only due to the predictor variables used and that potentially relevant effects are ignored. Consequently, besides the height and all other unknown individual effects there must be further effects on the velocity that cannot be ignored. When the regression analysis was extended, a simple model that only includes fixed factors and a mixed model that considers the individual velocity as a function of the fixed factors headway and height and all other unknown individual effects as a random factor were compared. The analysis shows that the mixed model is preferable. Accordingly, the effect of the individuals should be considered as a random factor. Otherwise, the effect of the factors of interest may be obscured.

For further research, the individual fundamental diagrams can be viewed and analyzed more closely. When we look at the individual fundamental diagrams, it is evident that there are diagrams in which the students keep different distances from different people moving at approximately the same velocity. Future studies could explore whether or not this is because an individual prefers to maintain a certain distance from a certain person. The effect of the people around an individual could then be analyzed on the basis of this information.

Acknowledgements

This study is based on a one-dimensional single-file experiment which was performed in 2014 at the school Gymnasium Bayreuther Straße (GBS) in Wuppertal, Germany [12]. Many thanks to Verena Ziemer who conducted the experiments.

References

1. Bestimmtheitsmaß $r^2$. https://www.inwt-statistics.de/blog-artikel-lesen/Bestimmtheitsmass_R2-Teil2.html. Accessed: 2020-11-27.
2. Petrack. http://ped.fz-juelich.de/petrack. Accessed: 2020-11-24.
3. H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
4. M. Boltes, A. Seyfried, B. Steffen, and A. Schadschneider. Automatic Extraction of Pedestrian Trajectories from Video Recordings. In *Pedestrian and Evacuation Dynamics 2008*, ed.: W.W.F. Klingsch, C.Rogsch, A. Schadschneider, M.Schreckenberg, Berlin/Heidelberg, Springer, 2010. - 978-3-642-04503-5. - S. 43 - 54, 2010. Record converted from VDB: 12.11.2012.
5. S. Burghardt, A. Seyfried, and W. Klingsch. Performance of stairs – fundamental diagram and topographical measurements. *Transportation Research Part C: Emerging Technologies*, 37:268 – 278, 2013.
6. S. Cao, J. Zhang, D. Salden, J. Ma, C. Shi, and R. Zhang. Pedestrian dynamics in single-file movement of crowd with different age compositions. *Phys. Rev. E*, 94:012312, Jul 2016.
[7] S. Cao, J. Zhang, D. Salden, J. Ma, C. Shi, and R. Zhang. Pedestrian dynamics in single-file movement of crowd with different age compositions. *Phys. Rev. E*, 94:012312, 07 2016.

[8] S. Cao, J. Zhang, W. Song, C. Shi, and R. Zhang. The stepping behavior analysis of pedestrians from different age groups via a single-file experiment. *Journal of Statistical Mechanics: Theory and Experiment*, 2018:033402, 03 2018.

[9] J. J. Fruin. *Pedestrian Planning and Design*. Elevator World, New York, 1971.

[10] S. Holl. *Methoden für die Bemessung der Leistungsfähigkeit multidirektional genutzter Fußverkehrsanlagen*. Dr., Bergische Universität Wuppertal, Jülich, 2016. Bergische Universität Wuppertal, Diss., 2016.

[11] S. Huang, R. Wei, S. Lo, and S. Lu. Experimental study on luggage-laden pedestrian movement in narrow seat aisle. *IOP Conference Series: Earth and Environmental Science*, 257:012049, 05 2019.

[12] F. IAS-7. Data archive of experimental data from studies about pedestrian dynamics, single-file motion of pupils: School gymbay. https://doi.org/10.34735/ped.2014.2.

[13] A. Jelić, C. Appert-Rolland, S. Lemercier, and J. Pettré. Properties of pedestrians walking in line: Fundamental diagrams. *Phys. Rev. E*, 85:036111, Mar 2012.

[14] A. U. Kemloh Wagoum, M. Chraibi, and G. Lämml. Jupedsim: an open framework for simulating and analyzing the dynamics of pedestrians.

[15] Y. Ma, Y. Sun, E. Lee, and R. Yuen. Pedestrian stepping dynamics in single-file movement. *Physical Review E*, 98:062311, 12 2018.

[16] V. Predtechenskii and A. Milinskii. *Planning for Foot Traffic Flow in Buildings*. TT. Amerind, 1978.

[17] X. Ren, J. Zhang, and W. Song. Contrastive study on the single-file pedestrian movement of the elderly and other age groups. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(9):093402, sep 2019.

[18] A. Seyfried, A. Portz, and A. Schadschneider. Phase coexistence in congested states of pedestrian dynamics. pages 496–505, 09 2010.

[19] W. Song, W. Lv, and Z. Fang. Experiment and modeling of microscopic movement characteristic of pedestrians. *Procedia Engineering*, 62:56–70, 12 2013.

[20] R. Subaih, M. Maree, M. Chraibi, S. Awad, and T. Zanoon. *Gender-based Insights into the Fundamental Diagram of Pedestrian Dynamics*, pages 613–624. 08 2019.

[21] R. Subaih, M. Maree, M. Chraibi, S. Awad, and T. Zanoon. Experimental Investigation on the Alleged Gender-differences in Pedestrian Dynamics: A Study Reveals No Gender Differences in Pedestrian Movement Behavior. *IEEE access*, 8:33748 – 33757, 2020.

[22] A. Tordeux and A. Schadschneider. White and relaxed noises in optimal velocity models for pedestrian flow with stop-and-go waves. *Journal of Physics A*, 49:185101, 2016.

[23] L. D. Vanumu, K. Ramachandra Rao, and G. Tiwari. Fundamental diagrams of pedestrian flow characteristics: A review. *European Transport Research Review*, 9:49, 2017.
[24] J. Wang, M. Boltes, A. Seyfried, J. Zhang, V. Ziemer, and W. Weng. Linking pedestrian flow characteristics with stepping locomotion. *Physica A*, 500:106–120, 2018.

[25] J. Wang, W. Weng, M. Boltes, J. Zhang, A. Tordeux, and V. Ziemer. Step styles of pedestrians at different densities. *Journal of Statistical Mechanics: Theory and Experiment*, 2018:023406, 02 2018.

[26] U. Weidmann. Transporttechnik der füssgänger. transporttechnische eigenschaften des füssgängerverkehrs, literaturauswertung. Technical report, Zürich, 1992-01.

[27] D. Yanagisawa, A. Tomoeda, and K. Nishinari. Improvement of pedestrian flow by slow rhythm. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 85:016111, 01 2012.

[28] R. Ye, Y. Zeng, G. Zeng, Z. Huang, X. Li, Z. Fang, and W. Song. Pedestrian single-file movement on stairs under different motivations. *Physica A: Statistical Mechanics and its Applications*, 571:125849, 02 2021.

[29] G. Zeng, S. Cao, C. Liu, and W. Song. Experimental and modeling study on relation of pedestrian step length and frequency under different headways. *Physica A: Statistical Mechanics and its Applications*, 500:237–248, 2018.

[30] G. Zeng, A. Schadschneider, J. Zhang, S. Wei, W. Song, and R. Ba. Experimental study on the effect of background music on pedestrian movement at high density. *Physics Letters A*, 383(10):1011–1018, 2019.

[31] J. Zhang. *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*. Dr., Bergische Universität Wuppertal, Jülich, 2012. Bergische Universität Wuppertal, Diss., 2012.

[32] J. Zhang, S. Cao, D. Salden, and J. Ma. Homogeneity and activeness of crowd on aged pedestrian dynamics. *Procedia Computer Science*, 83:361–368, 12 2016.

[33] J. Zhang, W. Mehner, S. Holl, M. Boltes, E. Andresen, A. Schadschneider, and A. Seyfried. Universal flow-density relation of single-file bicycle, pedestrian and car motion. *Physics Letters A*, 378(44):3274 – 3277, 2014.

[34] J. Zhang, A. Tordeux, and A. Seyfried. Effects of boundary conditions on single-file pedestrian flow. *Cellular Automata*, page 462–469, 2014.

[35] V. Ziemer. *Mikroskopische Fundamentaldiagramme der Fußgängerdynamik - Empirische Untersuchung von Experimenten eindimensionaler Bewegung sowie quantitative Beschreibung von Stau-Charakteristika*. Dr., Bergische Universität Wuppertal, Jülich, 2019. Bergische Universität Wuppertal, Diss., 2019.