

The ground reaction force of walking: a parameter to identify patients?

Medical Data Science Project

Name of the student:

Ursula TRINLER

July 19, 2021

Contents

1	Introduction	2
1.1	Summary	2
1.2	Gait analysis and ground reaction forces	2
1.3	Discrete Parameters of the GRF and COP	4
1.4	The <i>GaitRec</i> dataset in detail	4
1.5	Research Question	6
2	Methods	8
3	Results	9
4	Discussion	10
5	Conclusion	11
6	References	12
7	Appendix	13
7.1	Code for pre-processed time series of GRF and COP to extract discrete parameters	13

1 Introduction

1.1 Summary

This document analyses the *GaitRec* dataset which is freely available [online](#). The dataset has, furthermore, been published by Horsak et al. in Nature, Scientific Data (Horsak et al. 2020a). It contains a large-scale dataset of *ground reaction forces* (GRF) during walking of 211 healthy (HC) and 2084 impaired persons. Included patients had different conditions affecting either the hip (H), knee (K), ankle (A), or calcaneus (C). In total, 75.732 bi-lateral walking trials have been compiled.

Before the *GaitRec* dataset will be described in more detail (Chapter 1.3) and the research question will be explored (Chapter 1.4 and 2) and analyzed (Chapter 3) a short background information about GRF and gait analysis will be given.

1.2 Gait analysis and ground reaction forces

Human movement is a complex topic and a growing research field. According to Baker (2007) Aristotle (384 until 322 BC) was one of the first known scientists to analyze human walking. It was not until the late 19th and early 20th century, however, that scientific methods started to be applied to gait analysis simulated by new measurement techniques.

Two main areas are mainly covered in human movement analysis: *kinematics* - the science of movement (without considering the forces which are responsible for the movements), and *kinetics* - the science of forces producing these movements. One such force acting on the body is the GRF which is the response of the ground to the foot contact. This GRF can be measured using so called force plates, which are normally embedded into the floor.

If a person would simply stand on such a plate it would measure the person's weight in Newton. Due to Newton's third law ("every action has an equal and opposite reaction") the measured force is the same force the person is "putting" into the ground but has opposite direction ("Newton's Laws of Motion" 2021). This means that the GRF would point vertically through the body. When the person is walking over the force plate, the GRF can increase due to acceleration forces and its direction is changing according to the center of mass of the person, see Figure 1 left.

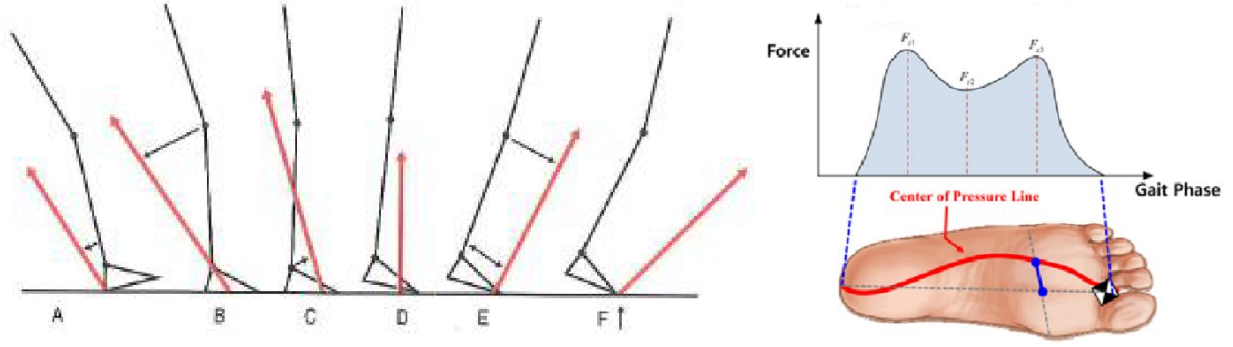


Fig. 1: Progression of GRF under the foot - sagittal view (left) and view from above (right).

The GRF is defined in a 3D space and can be split in three components, which represent the

- anterior-posterior AP (X)
- medial-lateral ML (Y) and
- vertical V (Z)

direction. The X, Y, and Z labeling is arbitrary and can change throughout publications. The X-direction defines the force produced in walking direction, the Y-direction represents frontal forces and the vertical Z- direction defines the force vertically through the body. During the stance phase of walking the GRF changes according to the acceleration and deceleration of the body on the force plate (see Figure 2).

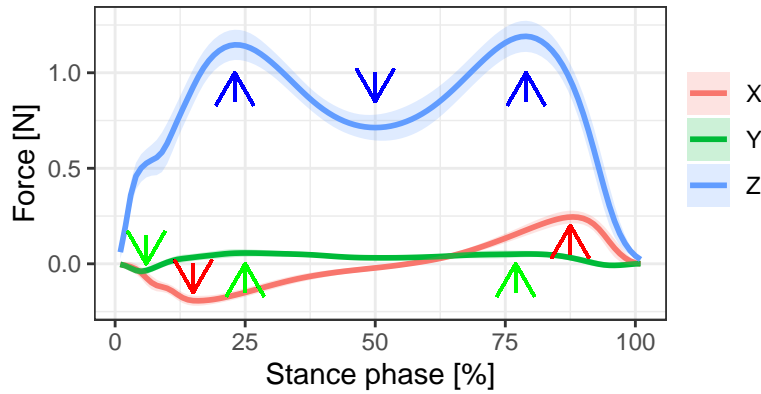


Fig. 2: GRF distribution in X, Y and Z direction during the stance phase of walking. Arrows define discrete parameters of interest.

Additionally to the three force components, another parameter related to the GRF is often analyzed, the so called *center of pressure* (COP). This parameter simply describes, where the force is centered under the foot, i.e., where the most force is concentrated. It is displayed in a 2D space onto the floor. During gait, this COP is traveling from the heel (during initial

foot contact on the floor) to the toes (at the end of stance when the foot leaves the ground). If one would imagine to observe the GRF from above, the COP would draw a line through the foot (see Figure 1 right side, adopted from Lim et al. (2017)).

Like the GRF, the COP can be divided in its components, the

- anterior-posterior AP (X) and
- medial-lateral ML (Y)

direction. The X-direction of the COP describes the propulsion of the COP in walking direction while the Y-direction is the medial-lateral shift of the COP under the foot.

1.3 Discrete Parameters of the GRF and COP

In case the musculoskeletal system is impaired, the shape of the GRF as well as the COP can change (e.g., John et al. (2012)). To understand if these changes have a clinical impact on the patient, specific discrete parameters are analyzed in detail during a clinical gait analysis. These discrete parameters are the maximum and minimum forces of each GRF component during stance phase (see Figure 2, arrows).

The length of the COP line gives additionally information about the walking stability of the patient (Gao et al. 2019). Is the length in X-direction shortened (i.e., distance between most distal to most proximal COP point) it might point to an asymmetric gait and a shorter single leg stance phase (i.e. when the contra-lateral leg is off the ground). An increase in a medial-lateral amplitude of the COP line (i.e., distance of most medial to most lateral point of the COP line) mostly implies insecurity during gait.

Important discrete parameters related to the GRF and COP can be found in Table (1).

Table 1: Discrete parameters of the GRF during walking

Variable Name	detailed information
GRF_F_V_max1	Maximum of vertical GRF in first half of stance
GRF_F_V_max2	Maximum of vertical GRF in second half of stance
GRF_F_V_min	Minimum of vertical GRF between both maxima
GRF_F_AP_max	Maximum of anterior-posterior GRF in second half of stance
GRF_F_AP_min	Minimum of anterior-posterior GRF in first half of stance
GRF_F_ML_max1	Maximum of medial-lateral GRF in first half of stance
GRF_F_ML_max2	Maximum of medial-lateral GRF in second half of stance
GRF_F_ML_min	Minimum of medial-lateral GRF at initial foot contact
GRF_COP_AP_length	Length of COP course in anterior-posterior direction during stance
GRF_COP_ML_length	Length of COP course in medial-lateral direction during stance

1.4 The *GaitRec* dataset in detail

The GaitRec dataset consists of several .CSV files. The main document (“GRF_metadata.csv”) includes main information about the participants and measurement trial (see Table 2). The

remaining files contain the raw and pre-processed GRF and COP time series of the stance phase of walking. The raw data present the measurements purely received from the force plates, i.e., no data cleaning has been undertaken. In the pre-processed files, following steps have been undertaken to clean and tidy up the dataset:

- Force signals were filtered with a 2nd order low-pass butterworth filter with a cut-off frequency of 20Hz to reduce noise,
- time normalized to 100% stance (i.e. 101 data points) to reduce effects temporal differences, and then
- normalized to the body weight of each participant (force divided by the body mass times gravity) to reduce effects due to individual anthropometry.
- Threshold of 80N to avoid inaccuracies in COP calculation at small force values
- Medio-lateral COP coordinates were mean-centered
- Anterior-posterior coordinates zero-centered

This report includes the pre-processed GRF and COP dataset.

Table 2: Participant and trial metadata of the GaitRec dataset

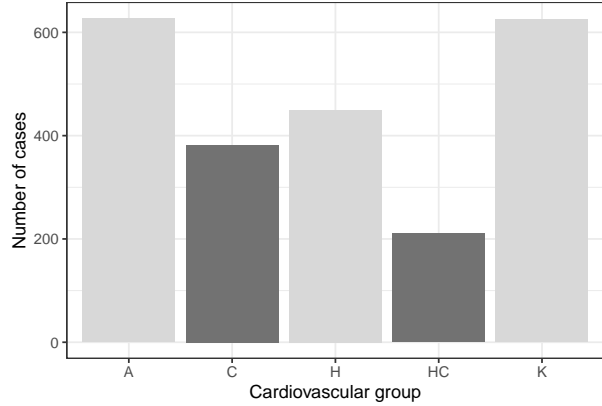
Participant	Walking trial
Sex (male, female)	Shod condition (barefoot, shoe)
Age (years)	Orthopaedic insole (yes, no)
Height (cm)	Walking speed (slow, self-selected, fast)
Body mass (kg) and weight ($\text{kg} \cdot \text{m/s}^2$)	Readmission into hospital (yes, no)
Shoe size (EU)	Session type (initial vs. control measurement)
Affected side (left, right, both)	Session date

As described in Chapter 1.1, participants can be divided into five different groups, one healthy group and four different patient groups. Number of participants per group varied slightly between subgroups (Figure \ref{Anthro B}), however, all exceeded 200 number of cases. Anthropometrics of each group is summarized in Figure 3 A. A more detailed visualization of age and BMI distribution between groups is shown in sub-Figure C and D. While anthropometrics are similar distributed between patient groups, the healthy control group seems to be younger with a smaller BMI than the patients. Running a Kruskal-Wallis-Test (Shapiro Test shows non-normal distribution) and a pairwise Wilcoxon test for Post-hoc analysis, differences can be shown between healthy controls and each patient group (see Figure 3 C and D). Analysis further down below will, therefore, include age and BMI as covariates.

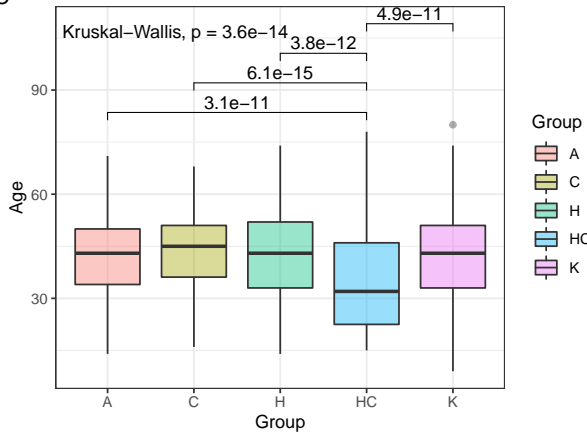
A

Mean (SD) of group anthropometrics					
	A (N = 627)	C (N = 382)	H (N = 450)	HC (N = 211)	K (N = 62)
Age (years)	41.5 (11.4)	43.5 (10.4)	42.6 (12.8)	34.7 (13.9)	41.6 (12.0)
Height (cm)	175.6 (8.0)	178.0 (7.0)	176.5 (8.0)	173.2 (8.5)	175.5 (9.1)
Body mass (kg)	86.9 (18.1)	84.0 (14.7)	82.3 (15.6)	74.0 (15.5)	84.3 (18.7)
BMI(km/m ²)	27.8 (5.0)	26.5 (4.2)	26.4 (4.3)	24.6 (4.4)	27.3 (5.2)

B



C



D

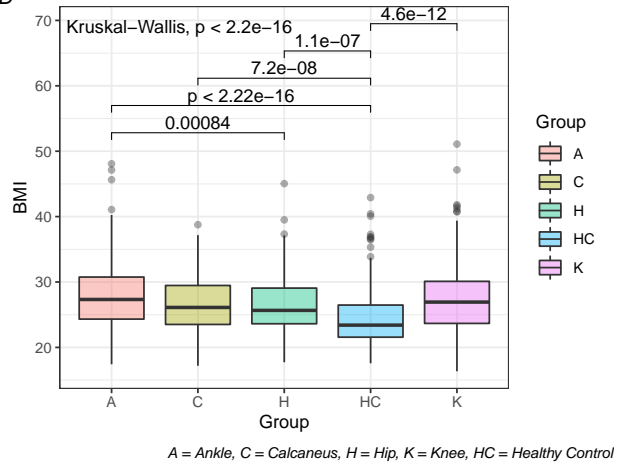


Fig. 3: Anthropometrics and number of cases of each sub-group.

1.5 Research Question

A full gait analysis is time consuming and needs trained personnel. Calculating joint angles and joint moments as well as power needs professional knowledge and results are not intuitive to understand. Therefore, simply using GRF to analyze the walking pattern would be time efficient and can be undertaken by non-trained personnel, while distinctive differences in discrete parameters of the GRF (“higher”, “lower”) can be easily interpreted. The research question of this report is, therefore, if and which discrete parameters of the GRF can be used to distinguish between patients with impairments of the lower limb and a healthy control group. Furthermore, can discrete GRF parameters be used to distinguish within different pathologies of the lower limb?

While age and BMI have already been shown to differ between healthy controls and patient groups and are identified as possible covariates, other main factors have been shown to influence gait analysis outcome: shod condition (Oeffinger et al. 1999) and walking speed (Schwartz, Rozumalski, and Trost 2008). Therefore, both variables will be considered in the following analysis.

It is, therefore, hypothesized that differences in GRF can be classified to distinguish between healthy controls and patients. Also, different patient groups can be discriminated between each other. Change in walking speed and shod condition effect GRF for all groups (healthy controls and patients) as well as age and BMI.

2 Methods

The pre-processed datasets provided by Horsak et al. (2020b) have been further processed by extracting discrete parameters (see Appendix for coding), which are presented in Table 1. In another step, variables' data type have been checked, while variables “sex”, “affected side”, “shod condition”, and “speed” have been changed from numeric to factor.

Within the whole dataset there are 9796 missing values, which are related to body height (and, therefore, BMI) and shoe size (3251 , 3294 respectively). In total, information of 5672 participants are complete, while 3299 participants have at least one missing value. This information is crucial for further analysis where body height (or BMI) and/or shoe size will play a role.

Before one can decide for proper analysis strategy, the distribution of involved parameters has to be reviewed. Distribution of discrete parameters, which have been calculated, can be seen in Figure

Furthermore, to receive a first view into the dataset and to latter be able to apply regression analysis, pairwise correlation coefficients will be calculated for numeric variables.

3 Results

4 Discussion

5 Conclusion

6 References

- Baker, Richard. 2007. “The History of Gait Analysis Before the Advent of Modern Computers.” *Gait & Posture* 26 (3): 331–42. <https://doi.org/10.1016/j.gaitpost.2006.10.014>.
- Gao, Chia-Chi, Jen-Suh Chern, Chun-Ju Chang, Po-Liang Lai, and Chi-Wen Lung. 2019. “Center of Pressure Progression Patterns During Level Walking in Adolescents with Idiopathic Scoliosis.” Edited by Kirsten Tulchin-Francis. *PLOS ONE* 14 (4): e0212161. <https://doi.org/10.1371/journal.pone.0212161>.
- Horsak, Brian, Djordje Slijepcevic, Anna-Maria Raberger, Catherine Schwab, Marianne Worisch, and Matthias Zeppelzauer. 2020a. “GaiTRec, a Large-Scale Ground Reaction Force Dataset of Healthy and Impaired Gait. Sci Data.” *Scientific Data* 7: 143–51. <https://doi.org/10.1038/s41597-020-0481-z>.
- . 2020b. “GaitRec, a Large-Scale Ground Reaction Force Dataset of Healthy and Impaired Gait.” *Scientific Data* 7 (1): 143. <https://doi.org/10.1038/s41597-020-0481-z>.
- John, Chand T., Ajay Seth, Michael H. Schwartz, and Scott L. Delp. 2012. “Contributions of Muscles to Mediolateral Ground Reaction Force over a Range of Walking Speeds.” *Journal of Biomechanics* 45 (14): 2438–43. <https://doi.org/10.1016/j.jbiomech.2012.06.037>.
- Lim, Dong-Hwan, Wan-Soo Kim, Ho-Jun Kim, and Chang-Soo Han. 2017. “Development of Real-Time Gait Phase Detection System for a Lower Extremity Exoskeleton Robot.” *International Journal of Precision Engineering and Manufacturing* 18 (5): 681–87. <https://doi.org/10.1007/s12541-017-0081-9>.
- “Newton’s Laws of Motion.” 2021. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Newton%27s_laws_of_motion&oldid=1032605773.
- Oeffinger, Donna, Bridgett Brauch, Shelley Cranfill, Cary Hisle, Cyndi Wynn, Ramona Hicks, and Sam Augsburg. 1999. “Comparison of Gait with and Without Shoes in Children.” *Gait & Posture* 9 (2): 95–100. [https://doi.org/10.1016/S0966-6362\(99\)00005-3](https://doi.org/10.1016/S0966-6362(99)00005-3).
- Schwartz, Michael H., Adam Rozumalski, and Joyce P. Trost. 2008. “The Effect of Walking Speed on the Gait of Typically Developing Children.” *Journal of Biomechanics* 41 (8): 1639–50. <https://doi.org/10.1016/j.jbiomech.2008.03.015>.

7 Appendix

7.1 Code for pre-processed time series of GRF and COP to extract discrete parameters

```
# read in data
metaData <- read.csv('GRF_metadata.csv')
GRF_F_V_r <- read.csv('GRF_F_V_PRO_right.csv')
GRF_F_V_l <- read.csv('GRF_F_V_PRO_left.csv')
GRF_F_ML_r <- read.csv('GRF_F_ML_PRO_right.csv')
GRF_F_ML_l <- read.csv('GRF_F_ML_PRO_left.csv')
GRF_F_AP_r <- read.csv('GRF_F_AP_PRO_right.csv')
GRF_F_AP_l <- read.csv('GRF_F_AP_PRO_left.csv')
GRF_COP_AP_r <- read.csv('GRF_COP_AP_PRO_right.csv')
GRF_COP_AP_l <- read.csv('GRF_COP_AP_PRO_left.csv')
GRF_COP_ML_r <- read.csv('GRF_COP_ML_PRO_right.csv')
GRF_COP_ML_l <- read.csv('GRF_COP_ML_PRO_left.csv')

### Extract variables of interest and include with meta data

# Variable for each row
GRF_F_V_r_max1 <- apply(GRF_F_V_r[,4:53],1,max)
GRF_F_V_r_max2 <- apply(GRF_F_V_r[,54:104],1,max)
GRF_F_V_r_min <- apply(GRF_F_V_r[,35:70],1,min)
GRF_F_V_l_max1 <- apply(GRF_F_V_l[,4:53],1,max)
GRF_F_V_l_max2 <- apply(GRF_F_V_l[,54:104],1,max)
GRF_F_V_l_min <- apply(GRF_F_V_l[,35:70],1,min)
GRF_F_AP_r_max <- apply(GRF_F_AP_r[,4:104],1,max)
GRF_F_AP_r_min <- apply(GRF_F_AP_r[,4:104],1,min)
GRF_F_AP_l_max <- apply(GRF_F_AP_l[,4:104],1,max)
GRF_F_AP_l_min <- apply(GRF_F_AP_l[,4:104],1,min)
GRF_F_ML_r_max1 <- apply(GRF_F_ML_r[,27:54],1,max)
GRF_F_ML_r_max2 <- apply(GRF_F_ML_r[,68:98],1,max)
GRF_F_ML_r_min <- apply(GRF_F_ML_r[,4:14],1,min)
GRF_F_ML_l_max1 <- apply(GRF_F_ML_l[,27:54],1,max)
GRF_F_ML_l_max2 <- apply(GRF_F_ML_l[,68:98],1,max)
GRF_F_ML_l_min <- apply(GRF_F_ML_l[,4:14],1,min)
GRF_COP_AP_r_length <- (apply(GRF_COP_AP_r[,4:104],1,max) -
                        apply(GRF_COP_AP_r[,4:104],1,min))
GRF_COP_AP_l_length <- (apply(GRF_COP_AP_l[,4:104],1,max) -
                        apply(GRF_COP_AP_l[,4:104],1,min))
GRF_COP_ML_r_length <- (apply(GRF_COP_ML_r[,4:104],1,max) -
                        apply(GRF_COP_ML_r[,4:104],1,min))
```

```

GRF_COP_ML_l_length <- (apply(GRF_COP_ML_l[,4:104],1,max) -
                        apply(GRF_COP_ML_l[,4:104],1,min))

# mean for each session
varData <- cbind(GRF_F_V_r[1:3],GRF_F_V_r_max1,
                GRF_F_V_l_max1,
                GRF_F_V_r_max2,
                GRF_F_V_l_max2,
                GRF_F_V_r_min,
                GRF_F_V_l_min,
                GRF_F_AP_r_max,
                GRF_F_AP_l_max,
                GRF_F_AP_r_min,
                GRF_F_AP_l_min,
                GRF_F_ML_r_max1,
                GRF_F_ML_l_max1,
                GRF_F_ML_r_max2,
                GRF_F_ML_l_max2,
                GRF_F_ML_r_min,
                GRF_F_ML_l_min,
                GRF_COP_AP_r_length,
                GRF_COP_AP_l_length,
                GRF_COP_ML_r_length,
                GRF_COP_ML_l_length)

varData_av <- varData %>% group_by(SUBJECT_ID,SESSION_ID) %>%
  summarize(GRF_F_V_r_max1 = mean(GRF_F_V_r_max1),
            GRF_F_V_l_max1 = mean(GRF_F_V_l_max1),
            GRF_F_V_r_max2 = mean(GRF_F_V_r_max2),
            GRF_F_V_l_max2 = mean(GRF_F_V_l_max2),
            GRF_F_V_r_min = mean(GRF_F_V_r_min),
            GRF_F_V_l_min = mean(GRF_F_V_l_min),
            GRF_F_AP_r_max = mean(GRF_F_AP_r_max),
            GRF_F_AP_l_max = mean(GRF_F_AP_l_max),
            GRF_F_AP_r_min = mean(GRF_F_AP_r_min),
            GRF_F_AP_l_min = mean( GRF_F_AP_l_min),
            GRF_F_ML_r_max1 = mean(GRF_F_ML_r_max1),
            GRF_F_ML_l_max1 = mean(GRF_F_ML_l_max1),
            GRF_F_ML_r_max2 = mean( GRF_F_ML_r_max2),
            GRF_F_ML_l_max2 = mean(GRF_F_ML_l_max2),
            GRF_F_ML_r_min = mean(GRF_F_ML_r_min),
            GRF_F_ML_l_min = mean(GRF_F_ML_l_min),
            GRF_COP_AP_r_length = mean(GRF_COP_AP_r_length),
            GRF_COP_AP_l_length = mean(GRF_COP_AP_l_length),

```

```
GRF_COP_ML_r_length = mean(GRF_COP_ML_r_length),  
GRF_COP_ML_l_length = mean(GRF_COP_ML_l_length))  
  
allData <- cbind(metaData, varData_av[, 3:22])
```