Journal of Biomechanics 87 (2019) 161–166

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Short communication

An assessment of the information lost when applying data reduction techniques to dynamic plantar pressure measurements

Brian G. Booth [a,*], Noël L.W. Keijsers [b], Jan Sijbers [a], Toon Huysmans [a,c]

[a] imec-Vision Lab, Department of Physics, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium
[b] Department of Research, Sint Maartenskliniek, P.O. Box 9011, 6500 GM Nijmegen, the Netherlands
[c] Section on Applied Ergonomics & Design, Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, the Netherlands

ARTICLE INFO

Article history:
Accepted 12 February 2019

Keywords:
Pedobarography
Information content
Dimensionality
Data reductions

ABSTRACT

Data reduction techniques are commonly applied to dynamic plantar pressure measurements, often prior to the measurement's analysis. In performing these data reductions, information is discarded from the measurement before it can be evaluated, leading to unkonwn consequences. In this study, we aim to provide the first assessment of what impact data reduction techniques have on plantar pressure measurements. Specifically, we quantify the extent to which information of any kind is discarded when performing common data reductions. Plantar pressure measurements were collected from 33 healthy controls, 8 Hallux Valgus patients, and 10 Metatarsalgia patients. Eleven common data reductions were then applied to the measurements, and the resulting datasets were compared to the original measurement in three ways. First, information theory was used to estimate the information content present in the original and reduced datasets. Second, principal component analysis was used to estimate the number of intrinsic dimensions present. Finally, a permutational multivariate ANOVA was performed to evaluate the significance of group differences between the healthy controls, Hallux Valgus, and Metatarsalgia groups. The evaluated data reductions showed a minimum of 99.1% loss in information content and losses of dimensionality between 20.8% and 83.3%. Significant group differences were also lost after each of the 11 data reductions ($\alpha = 0.05$), but these results may differ for other patient groups (especially those with highly-deformed footprints) or other region of interest definitions. Nevertheless, the existence of these results suggest that the diagnostic content of dynamic plantar pressure measurements is yet to be fully exploited.

1. Introduction

The assessment of gait and foot function frequently includes a dynamic acquisition of pressure measurements from the plantar surface of a walking patient's foot (Giacomozzi, 2011). In the process of a single acquisition, tens of thousands of pressure samples can be obtained over three dimensions: the anterior-posterior dimension, the medial-lateral dimension, and time (Pataky and Maiwald, 2011). Since analysis techniques for such large sets of data points are difficult to develop, and their results difficult to interpret, data reductions are often performed on all measurements prior to their analysis. Popular techniques include (a) the region of interest (ROI) approach, where a single aggregate measure is used to represent an anatomical region (Giacomozzi and Stebbins, 2017), (b) temporal approaches, such as centre of pressure (COP) trajectories, which create one-dimensional (1D) curves of how plantar pressures change over time (Keijsers et al., 2016), and (c) spatial approaches that aggregate data in the temporal domain in order to generate two-dimensional (2D) images of the plantar pressure's spatial pattern (Keijsers et al., 2009).

The choice of data reduction is normally dependent of the clinical question under examination and, therefore, studies involving data reductions regularly provide meaningful results. Nevertheless, if the application of data reductions is done prior to analysis, the discarded data will never be analyzed. This raises the question of whether additional diagnostic information could be lost as a result of plantar pressure data reductions. The answer to this question likely hinges on many factors, including the foot condition being diagnosed, physical characteristics of the patients (Keijsers et al., 2014), the specifications of the measurement device (Berki and Davis, 2016; Davis et al., 1996), and potentially others

* Corresponding author.
E-mail addresses: brian.booth@uantwerpen.be (B.G. Booth), jan.sijbers@uantwerpen.be (N.L.W. Keijsers), n.keijsers@maartenskliniek.nl (J. Sijbers), T.Huysmans@tudelft.nl (T. Huysmans).

https://doi.org/10.1016/j.jbiomech.2019.02.008
0021-9290/© 2019 The Author(s). Published by Elsevier Ltd.

(e.g. environment). Before examining all of these factors, it is first important to establish that such an examination is worthwhile. To do so, we must first determine to what extent plantar pressure data reductions discard information of any kind, diagnostically valuable or not. If the information losses are small, then we can be fairly confident that the diagnostically-valuable information is being retained.

The purpose of this study is to provide a quantification of the information and dimensionality present both before and after common plantar pressure data reductions. We evaluated 11 data reduction techniques that conceptually span the range of such techniques used in the literature. The effect of each data reduction is measured using a combination of information theory, principal component analysis (PCA), and permutational multivariate ANOVA (PERMANOVA). The first method is used to measure the information content retained after a data reduction (Pavlidis, 2017), the second measures the number of intrinsic dimensions that are retained (Stanković et al., 2018), while the last evaluates whether the data reductions impact our ability to distinguish between patients and healthy controls (Anderson, 2017). Taken together, these three measures will allow us to infer whether the loss of diagnostic information through plantar pressure data reductions is a legitimate concern.

2. Materials and methods

2.1. Materials

Dynamic plantar pressure measurements were collected from 33 healthy controls, 8 Hallux Valgus patients, and 10 Metatarsalgia patients. Demographics of the cohort are summarized in Table 1. The study was approved by the internal review committee of Sint Maartenskliniek and met the requirements for exemption from the Medical Ethics Committee review under the Dutch Medical Research Involving Human Subjects law.

Plantar pressures were measured five times for each individual while they walked at their preferred walking speed. A 0.5 m footscan® plate (rs scan, Paal, Belgium; dimensions: 48.8 $\times$ 32.5 cm; sensors: 7.62 x 5.08 mm) resting on a Kistler force plate (9286AA, Kistler, Wintherthur, Switzerland) was used to collect the measurements. Data was measured at a frequency of 200 Hz for the healthy controls, and 500 Hz for the patient groups.

Due to the rectangular sensor size, the foot geometry in each plantar pressure measurement was corrected by spatially upsampling the measurement to 3 mm $\times$ 3 mm using bilinear interpolation. To normalize for scanning frequency, the patient measurements were temporally downsampled to 200 Hz using linear interpolation. Finally, each subject had their 5 measurements aligned and averaged as described in Booth et al. (2018) to reduce measurement noise.

2.2. Plantar pressure data reductions

In this study, we examined 12 common plantar pressure data reductions:

- *Regions of Interest (ROI):* The Novel 10 region mask, described in Keijsers (2012), was used to define the 10 regions of interest. Each region is represented by a single data point, with the mean pressure (ROI-MEAN), peak pressure (ROI-PEAK), and pressure-time integral (ROI-PTI) being examined.
- *Region-Time Curves (ROI-TIME):* The Novel 10 region mask is again used to define 10 regions of interest. For each region, a time curve is created by spatially aggregating pressure measurements per time point (Warren et al., 2004). As with the ROI technique, peak pressures (ROI-TIME-PEAK), mean pressures (ROI-TIME-MEAN), and pressure-time integrals (ROI-TIME-PTI) were examined.
- *Centre of Pressure Trajectories (COP, VOP):* for each plantar pressure measurement, we generate centre of pressure (COP) trajectories, and the velocity of the centre of pressure (VOP), as described in Keijsers et al. (2016). A COP trajectory consists of two 1D curves: one for the anterior-posterior COP location over time, and one for the medial-lateral COP location over time. The velocity of the centre of pressure (VOP) is then computed as the temporal derivative of the COP.
- *Pressure Pattern Images (IMAGE):* For each dynamic plantar pressure measurement, the time dimension is aggregated to produce a 2D image (Giacomozzi, 2011; Keijsers, 2012). As with the ROI-based techniques, peak pressures (PEAK IMAGE), mean pressures (MEAN IMAGE), and pressure-time integrals (PTI IMAGE) were examined.
- *Full Measurement (FULL):* The full dynamic plantar pressure measurement, without any data reduction.

The details of each of these data reductions are summarized in Table 2 and mathematical definitions for each are provided as supplementary material.

2.3. Experimental setup

In order to perform the subsequent analyses, an anatomical correspondence is required between plantar pressures from different subjects. To obtain this correspondence for plantar pressure images, we employed the between-subject registration technique of Pataky et al. (2011). For the full plantar pressure measurement, we obtained correspondence using the between-subject registration and dynamic time warping algorithms from STAPP (Booth et al., 2018).

For each data reduction technique, we quantified the information content, *IC*, retained after the data reduction using information theory:

$$IC(\mathbf{x}) = \sum_{i=1}^{N} -log[p(x_i)] \quad (1)$$

where $\mathbf{x} = [x_i] \in \mathbb{R}^N$ is a vector containing the data-reduced plantar pressures, N is the number of data points in the vector $\mathbf{x}$ (see Table 2), and $p(x_i)$ is the probability of seeing the data point x_i in the data. This probability is estimated from the data itself using a histogram-based estimator with 256 bins. In this manner, the information content of a data point is inversely-related to its frequency in the dataset, thereby resulting in redundant data points having a

Table 1
Cohort demographics for this study. No significant differences in subject weight or height were observed.

	Healthy Controls	Hallux Valgus	Metatasalgia	p-value 1-way ANOVA
Number of Subjects	33	8	10	
Weight (kg)	74.2 ± 11.9	67.8 ± 12.0	76.0 ± 15.6	0.0946
Height (cm)	174 ± 7.9	167 ± 7.6	171 ± 11.3	0.1301

Table 2
Summary of the plantar pressure data reductions examined in this study. Sizes are given in terms of the number of data points retained, with $|X|, |Y|, |T|$ referring to the number of samples collected along the anterior-posterior, medial-lateral, and time dimensions respectively.

Data Reduction	Acronym	Defined In	Size (N)	Data Reduction Options
Regions of Interest	ROI	Keijsers (2012)	10	Peak Pressure Mean Pressure Pressure-Time Integral
Region-Time Curves	ROI-TIME	Warren et al. (2004)	$10\|T\|$	Peak Pressure Mean Pressure Pressure-Time Integral
Center of Pressure	COP	Keijsers et al. (2016)	$2\|T\|$	–
Velocity of COP	VOP	Keijsers et al. (2016)	$3\|T\|$	–
Pressure Patterns	IMAGE	Giacomozzi (2011)	$\|X\|\|Y\|$	Peak Pressure Mean Pressure Pressure-Time Integral
Full Measurement	FULL	Booth et al. (2018)	$\|X\|\|Y\|\|T\|$	–

lower information content. For more information on information content estimation, we refer the reader to Pavlidis (2017).

The intrinsic dimensionality of a plantar pressure representation was estimated using principal components analysis (PCA). Let $\mathbf{x}_j$ be a vector containing the data-reduced plantar pressures from individual j. PCA models that vector as

$$\mathbf{x}_j = \mu + \sum_{i=1}^{k} w_{i,j}\ \mathbf{p_i} \tag{2}$$

where μ is the population average, and the $\mathbf{p_i}$'s are a set of uncorrelated variables referred to as principal components. Each principal component is weighted by $w_{i,j}$ in order to best fit the plantar pressure data vector $\mathbf{x}_j$. With PCA, each principal component is an intrinsic dimension of the dataset and describes a fraction of the between-subject variance. To obtain the number of intrinsic dimensions, k, we ordered the principal components by the largest amount of between-subject variance each describes to the smallest, then selected the first k dimensions that explain 95% of that variance (Stanković et al., 2018). The 95% threshold was chosen to mimic the choice of $\alpha = 0.05$ in significance tests; in both cases, 5% of the data is assumed to be noise.

Finally, we performed a PERMANOVA on the PCA weights of the healthy control, Hallux Valgus, and Metatarsalgia groups following each data reduction. Note that PERMANOVA is used here instead of a multivariate ANOVA since PERMANOVA does not require normally-distributed data (Anderson, 2017).

3. Results

Fig. 1 displays, in log-scale, the information content following each plantar pressure data reduction. The information content for the full plantar pressure measurement is over 1 million bits, over $110\times$ more than the next closest plantar pressure representation (PEAK IMAGE). This represents an information loss of 99.1%. Other data reductions perform worse by this metric, with ROI-PTI retaining even less than 0.00001% of the information in the full plantar pressure measurement.

Fig. 2 shows the number of intrinsic dimensions identified in our PCA analysis. We noted that the full plantar pressure measurement contained 24 significant dimensions, 5 more than the next-closest PEAK IMAGE representation. This represents a loss of dimensionality of 20.8%. Other data reduction techniques performed worse by this metric, with ROI-PTI retaining only 4 of the original 24 intrinsic dimensions (or 16.7%).

Fig. 3 shows the p-values from a PERMANOVA test between the healthy controls, Hallux Valgus, and Metatarsalgia groups. Note that the full plantar pressure is able to significantly distinguish between the groups ($p = 0.023$), but as data reductions are performed, this statistical significance goes away. In fact, ROI-PEAK ($p = 0.069$) and ROI-TIME-PEAK ($p = 0.072$) where the only data reductions that were suggestive of significant group differences (i.e. $\alpha < 0.1$).

4. Discussion

In general, we saw substantial losses of information content with every data reduction technique. These losses were matched by losses in intrinsic dimensionality, though the losses in dimensionality were less extreme. Nevertheless, each intrinsic dimension represents a phenomena in the dataset (Stanković et al., 2018), so their loss suggests that it is not just fine details of the data that are being discarded.

What is also notable in this study is the impact of the time dimension. Retaining the time dimension of the plantar pressure measurement increases the information content and dimensionality dramatically. For example, the ROI-TIME curves have an information content anywhere from 380 to 660 times higher than the corresponding ROI representations. Similarly, the full plantar pressure measurement contains anywhere from 110 to 280 times more information content than the IMAGE representations. In both cases, the number of intrinsic dimensions increased by at least 5 with the addition of the time dimension (coincidentally, 5 significant dimensions is roughly what the ROI representations originally have). Even the VOP representation improves slightly over the COP by taking the temporal derivative of the COP. Hence, the movement of the foot, specifically how people load and unload the foot, contains a wealth of information that can allow us to evaluate patients and identify those at risk of foot complaints.

The PERMANOVA results further show that these losses in information and dimensionality can also impact our ability to distinguish between patient groups. These results suggest that the lost information does have some diagnostic value, at least for this specific experiment. It is quite possible that different patient groups, different foot conditions, or even different measurement devices, could change these results. However, the existence of such a scenario indicates that if a plantar pressure data reduction is to be performed, it should be properly motivated by the clinical question under examination.

We further note that this study was not exhaustive in terms of the data reduction techniques that have been seen for plantar pressure measurements. In particular, different definitions exist for regions of interest, and various ratios between plantar

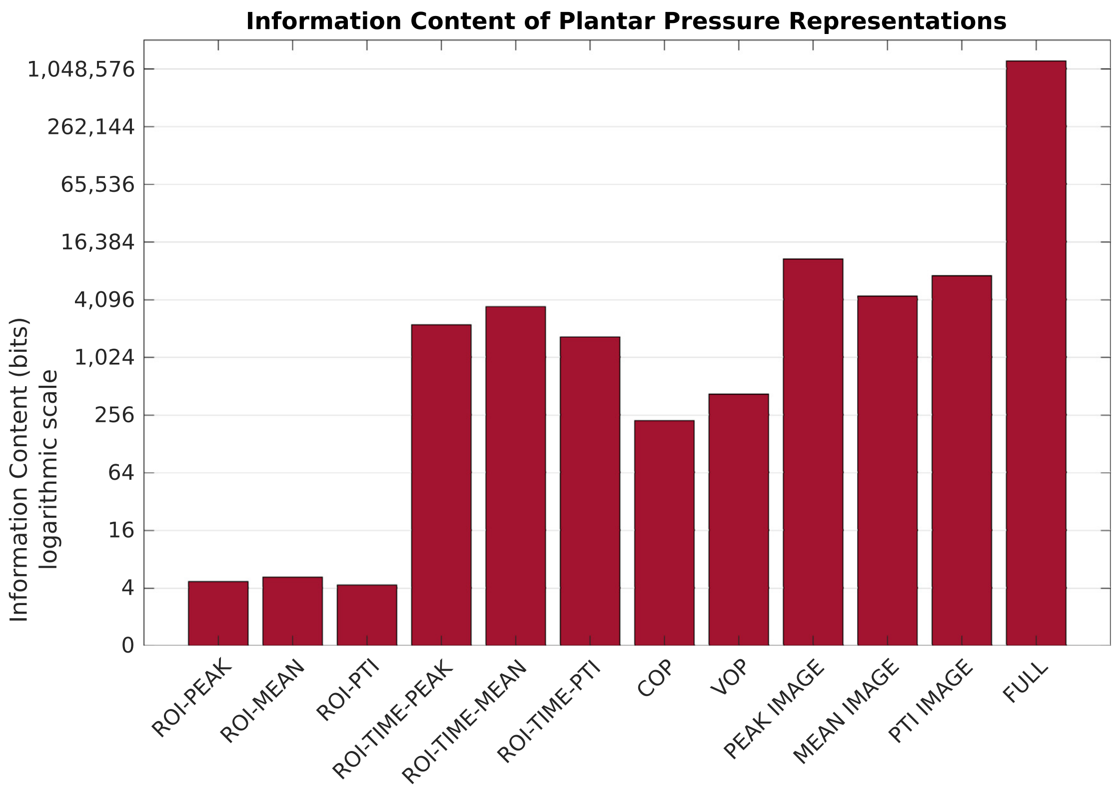

Fig. 1. The information content, in log-scale, after the application of various plantar pressure data reductions. This information theory measure scores data points based on their redundency in a dataset. Each data reduction technique discards at least 99% of the data in the full plantar pressure measurement. See text for further details.

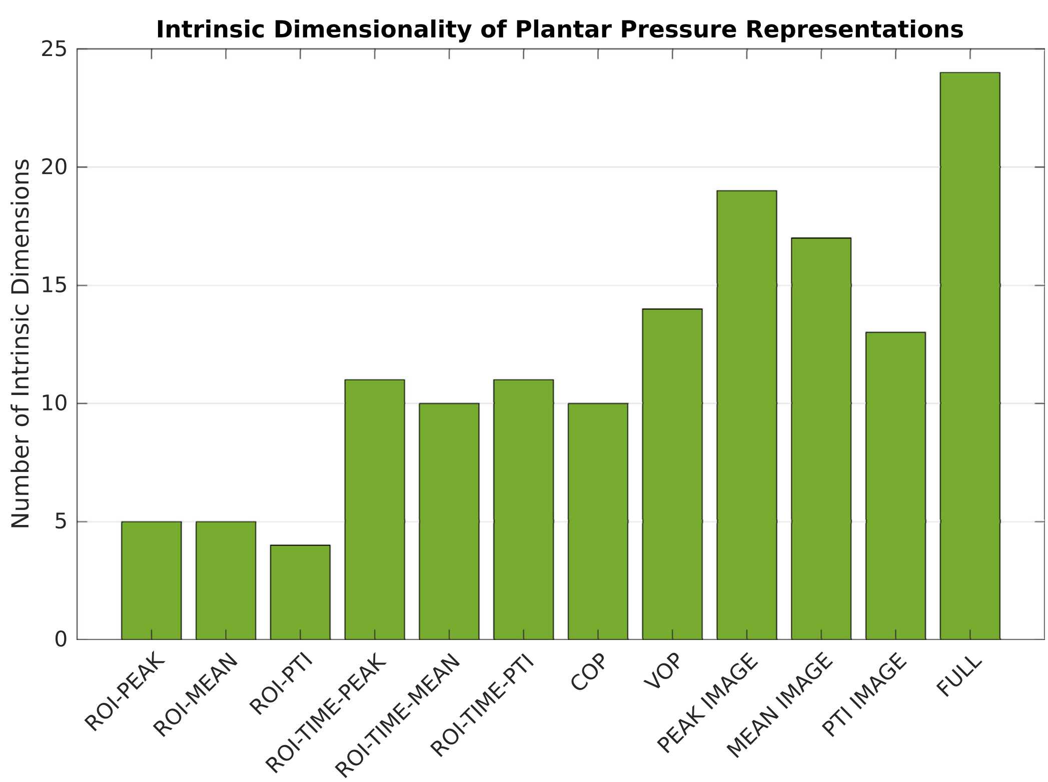

Fig. 2. The number of intrinsic dimensions following each plantar pressure data reduction, according to principal component analysis. The number of intrinsic dimensions is equal to the number of principal components required to explain 95% of the between-subject variance. See text for further details.

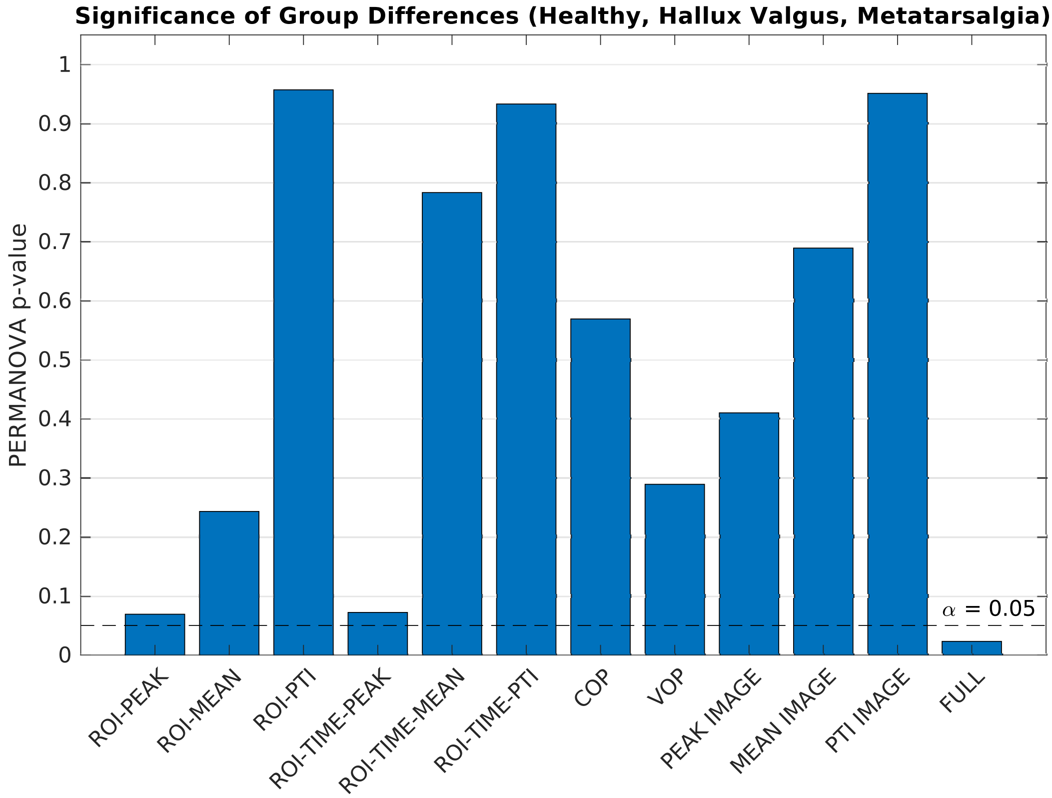

Fig. 3. Significance levels, following plantar pressure data reductions, for group differences between healthy controls, Hallux Valgus patients, and Metatarsalgia patients. Shown are p-values from permuational multivariate ANOVA (PERMANOVA) test. Note that each data reductions results in a loss of significant group differences. See text for further details.

pressure quantities have also been proposed (Deschamps et al., 2015). Nevertheless, the fact that all data reductions tested here show significant losses of information and dimensionality suggest that similar data reductions will also see these losses.

Ideally, we recommend that plantar pressure measurements be analyzed in their full form, followed by a data reduction to summarize the results. In this way, we can perform a data reduction that takes into consideration the results of the analysis, thereby ensuring that we retain the information relevant for diagnosis. There has already been work on the analysis of full plantar pressure measurements (Booth et al., 2018; Pataky and Maiwald, 2011) and recent works in the medical image analysis field could also be applied here (Litjens et al., 2017). By analyzing plantar pressures first, then summarizing the results, a more optimal use of the plantar pressure data can be obtained, which could potentially lead to new insights into foot and ankle function.

Conflict of interest statement

All authors declare that they do not have any conflict of interest. The authors further state that there was no financial or personal relationship with other people or organizations that influenced the outcome of this study.

Acknowledgements

This research has received funding from imec ICON Grant No. 150218 and the European Union's Horizon 2020 reserch and innovation programme under the Marie Sklodowska-Curie grant agreement No. 746614.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbiomech.2019.02.008.

References

Anderson, M., 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online, pp. 1–15.
Berki, V., Davis, B., 2016. Spatial frequency content of plantar pressure and shear profiles for diabetic and non-diabetic subjects. J. Biomech. 49 (15), 3746–3748.
Booth, B.G., Keijsers, N.L., Sijbers, J., Huysmans, T., 2018. STAPP: spatiotemporal analysis of plantar pressure measurements using statistical parametric mapping. Gait & Posture 63, 268–275.
Davis, B., Cothren, R., Quesada, P., Hanson, S., Perry, J., 1996. Frequency content of normal and diabetic plantar pressure profiles: Implications for the selection of transducer sizes. J. Biomech. 29 (7), 979–983.
Deschamps, K., Roosen, P., Nobels, F., Deleu, P.-A., Birch, I., Desloovere, K., Bruyninckx, H., Matricali, G., Staes, F., 2015. Review of clinical approaches and diagnostic quantities used in pedobarographic measurements. J. Sports Med. Phys. Fitness 55 (3), 191–204.
Giacomozzi, C., 2011. Biomechanics in Applications. In: Klika, V. (Ed.), ch. Potentialities and Criticalities of Plantar Pressure Measurements in the Study of Foot Biomechanics: Devices, Methodologies and Applications (Chapter 11). InTech, pp. 249–274.
Giacomozzi, C., Stebbins, J.A., 2017. Anatomical masking of pressure footprints based on the oxford foot model: validation and clinical relevance. Gait & Posture 53, 131–138.
Keijsers, N., 2012. The Science of Footwear. CRC Press. ch. Plantar Pressure Analysis, pp. 377–409.
Keijsers, N., Stolwijk, N., Nienhuis, B., Duysens, J., 2009. A new method to normalize plantar pressure measurements for foot size and foot progression angle. J. Biomech. 42, 87–90.
Keijsers, N., Stolwijk, N., Louwerens, J.-W., 2014. The effect of various subject characteristics on plantar pressure pattern. J. Foot Ankle Res. 7 (1), A40.

Keijsers, N., Stolwijk, N., Renzenbrink, G., Duysens, J., 2016. Prediction of walking speed using single stance force or pressure measurements in healthy subjects. Gait & Posture 43, 93–95.

Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B., Sánchez, C., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Pataky, T.C., Maiwald, C., 2011. Spatiotemporal volumetric analysis of dynamic plantar pressure data. Med. Sci. Sports Exercise 43 (8), 1582–1589.

Pataky, T.C., Bosch, K., Mu, T., Keijsers, N.L., Segers, V., Rosenbaum, D., Goulermas, J. Y., 2011. An anatomically unbiased foot template for inter-subject plantar pressure evaulation. Gait & Posture 33, 418–422.

Pavlidis, G., 2017. Mixed Raster Content. Springer, Singapore. ch. 2: Data Coding and Image Compression.

Stanković, K., Booth, B., Danckaers, F., Burg, F., Vermaelen, P., Duerinck, S., Sijbers, J., Huysmans, T., 2018. Three-dimensional quantitative analysis of healthy foot shape: a proof of concept study. J. Foot Ankle Res. 11 (1), 8.

Warren, G.L., Maher, R.M., Higbie, E.J., 2004. Temporal patterns of plantar pressures and lower-leg muscle activity during walking: effect of speed. Gait & Posture 19 (1), 91–100.