PREDICTION OF GROUND REACTION FORCES AND MOMENTS VIA SUPERVISED LEARNING IS INDEPENDENT OF PARTICIPANT SEX, HEIGHT AND MASS

William R. Johnson[1], Ajmal Mian[2], Cyril J. Donnelly[1], David Lloyd[3] & Jacqueline Alderson[1,4]

School of Human Sciences, The University of Western Australia, Perth, Australia[1]
School of Computer Science and Software Engineering,The University of Western Australia, Perth, Australia[2]
Menzies Health Institute, Griffith University, Gold Coast, Australia[3]
Auckland University of Technology, Sports Performance Research Institute New Zealand (SPRINZ), Auckland, New Zealand[4]

Accurate multidimensional ground reaction forces and moments (GRF/Ms) can be predicted from marker-based motion capture using Partial Least Squares (PLS) supervised learning. In this study, the correlations between known and predicted GRF/Ms are compared depending on whether the PLS model is trained using the discrete inputs of sex, height and mass. All three variables were found to be accounted for in the marker trajectory data, which serves to simplify data capture requirements and importantly, indicates that prediction of GRF/Ms can be achieved without pre-existing knowledge of such participant specific inputs. This multidisciplinary research approach significantly advances machine representation of real-world physical attributes with direct application to sports biomechanics.

KEY WORDS: big data, motion capture, computer vision, sports analytics.

INTRODUCTION: One of the ongoing challenges in sports biomechanics is that data accuracy necessary for the estimation of internal and external musculoskeletal loads, and subsequent injury risk, requires dual data capture of marker-based motion and embedded force plate derived GRF/Ms in controlled research laboratory conditions (Elliott & Alderson, 2007). Some studies have attempted to bring the field to the lab by mounting turf on the surface of the force plate (Jones, Kerwin, Irwin, & Nokes, 2009; Müller, Sterzing, Lange, & Milani, 2010) while others have adopted the reverse approach of taking measurement devices to the field, either by embedding force plates into the playing surface (Yanai et al., 2017), or more commonly using in-shoe pressure sensors (Liu, Inoue, & Shibata, 2010; Sim et al., 2015). However, none of these approaches are successful in producing accurate GRF/Ms over three orthogonal axes without impacting athlete performance. Efforts to predict GRF/Ms using non-invasive computer vision techniques show promise but either lack validation to a gold standard (Soo Park & Shi, 2016; Wei & Chai, 2010) or relevance to sporting tasks (Chen, Urban, Osendorfer, Bayer, & Van Der Smagt, 2014).

The aim of this study is to test the accuracy of a PLS prediction model with and without the discrete input variables of sex, height and mass which are often required in traditional biomechanical data collection pipelines. The first investigation was that removal of sex, and second, that removal of mass and height, would have negligible effects on predicted versus known GRF/Ms correlation coefficients for running and sidestepping trials. We hypothesise that all three discrete variables are inherently accounted for in the marker trajectory data.

METHODS: Mining of archive data was carried out under The University of Western Australia (UWA) ethics exemption RA/4/1/8415. The capture sessions were carried out at one of the university's three biomechanics laboratories over a 13-year period from 2004-2017 (the design of the study is shown in Figure 1), with participants drawn from a healthy population, male 69.1 %, female 30.9 %, height 1,741 ± 102 mm and mass 69.75 ± 11.47 kg. Given the customised UWA marker set has evolved in this period, the following subset of markers was selected to maximise trial inclusion: C7, sacrum; and hallux, calcaneus and lateral ankle malleolus of each foot (Besier, Sturnieks, Alderson, & Lloyd, 2003).

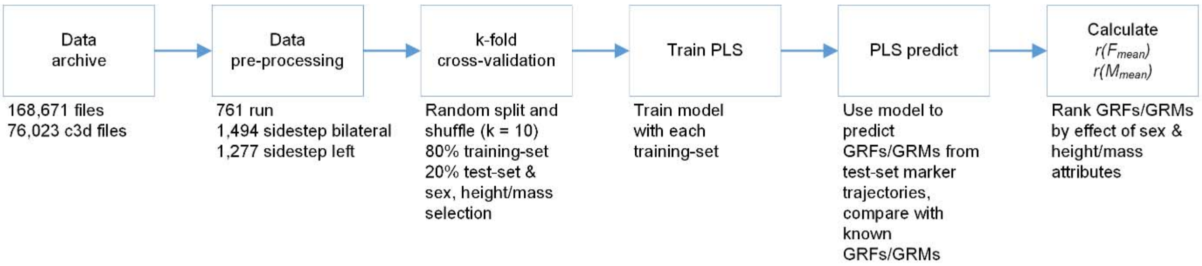

Figure 1: Overall study design.

The laboratory motion capture equipment has varied during this time from 12-20 Vicon (Oxford Metrics, Oxford, UK) near-infrared cameras of model types MCam2, MX13 and T40S. An AMTI force plate (Advanced Mechanical Technology Inc., Watertown, MA, USA) 1,200 x 1,200 mm installed flush with the floor measured the six GRF/Ms. Equipment setup and calibration was conducted to manufacturer specifications using Vicon proprietary software (Workstation v5.2.4 to Nexus v2.2.3), with data stored in the industry standard 'coordinate 3D' c3d file format (Motion Lab Systems, Baton Rouge, LA).

Several pre-processing steps were applied to maximise the integrity of data before training the PLS model. The foot-strike event was automatically determined by detecting vertical force greater than a threshold (20 N) over a defined period (0.025 s) along with the vertical and lateral velocities (0.02 $m \cdot s^{-1}$ and 0.15 $m \cdot s^{-1}$) of the dominant foot calcaneus marker (Milner & Paquette, 2015). The lead-in period before foot-strike was deemed more important for the predictor variable (kinematic marker trajectories), and therefore contiguous marker data was trimmed around the foot-strike event from -0.20 to +0.30 s (125 frames), and force plate data from -0.05 to +0.30 s (700 frames). Analogue force plate data sampled at frequencies lower than 2,000 Hz and motion capture lower than 250 Hz were time normalised using piecewise cubic spline interpolation. Sex, height and mass were obtained from the associated mp file (mp is a proprietary extensible mark-up language XML file format used by Vicon that stores participant specific session and anthropometric data). Children were excluded by rejecting trials where the participant height was less than 1,500 mm, this being two standard deviations below the average Australian adult female height 1,644 ± 72 mm (Ward, 2011). Trials with duplicate marker trajectories were rejected, with no regard whether the data was filtered, the sidestep planned or unplanned, footing crossover or regular, or foot strike technique. Trials where the participant movement or the start/end GRF/Ms were unexpected were also removed. Data analysis was conducted using MATLAB R2016b (MathWorks, Natick, MA) with the Biomechanical ToolKit v0.3 (Barre & Armand, 2014) under Ubuntu v14.04 (Canonical, London, UK) running on a desktop PC, Core i7 4GHz CPU, with 32GB RAM.

Marker trajectories for three movements types were selected as follows: 'run' (761 trials), 'sidestep bilateral' (1,494 trials), and 'sidestep left' (1,277 trials). The combination bilateral sidestep group was created by flipping 'sidestep right' trajectories laterally about the global origin and adding to the sidestep to the left (off the right foot) cohort. If not a sidestep, movement type 'run' was classified by forward motion of at least 2.0 $m \cdot s^{-1}$. To test the hypotheses, three use-cases were investigated: (1) variables sex, height and mass included; (2) without sex; and (3) without height and mass. To avoid overfitting, 10-fold cross-validation was used on the each of the movement type data sets. Trained using the 80% training set, then presented with the 20% test set, a subclass of PLS called Sparse SIMPLS running in R (Chun & Keleş, 2010; R Core Team, 2016) was used to predict GRF/Ms from marker trajectories which were then compared with the known force plate values. The mean correlation coefficient was determined for each of the six GRF/Ms and over the ten folds of data. The mean prediction correlation of the three GRF/Ms provided a measure of model performance in two numbers $r(F_{mean})$ and $r(M_{mean})$.

RESULTS & DISCUSSION: Systematic removal of the discrete variables of sex, then height and mass, from the input to the PLS model had a negligible effect on the prediction of GRF/Ms, and the results for the three movement types of 'run', 'sidestep bilateral' and 'sidestep left' are shown in Figure 2. The most accurate prediction pair of $r(F_{mean})$ 0.9803 and $r(M_{mean})$ 0.9004 were recorded for movement 'sidestep left'. The 'run' movement type reported the weakest prediction $r(F_{mean})$ 0.9152 and $r(M_{mean})$ 0.7568 from the smallest sample size (761 trials). The overall high correlations illustrated the suitability of PLS for this type of regression (temporal profile, number of predictor and output features, and number of samples).

These results did not improve on earlier testing with a smaller sample $r(F_{mean})$ 0.9804 and $r(M_{mean})$ 0.9143 (movement 'sidestep left', 441 trials) which suggests PLS performance is dependent on more than sample size. However, mean results were higher than the nearest comparison in the literature of Oh, Choi, and Mun (2013), who report maximum $r(F_{max})$ 0.9647 and $r(M_{max})$ 0.8987 values for walking.

The proximity of the correlations with and without sex, height and mass indicated all three variables are fully contained in the marker trajectories. Therefore, the hypothesis that all three discrete variables are inherently accounted for in the marker trajectory data was proven.

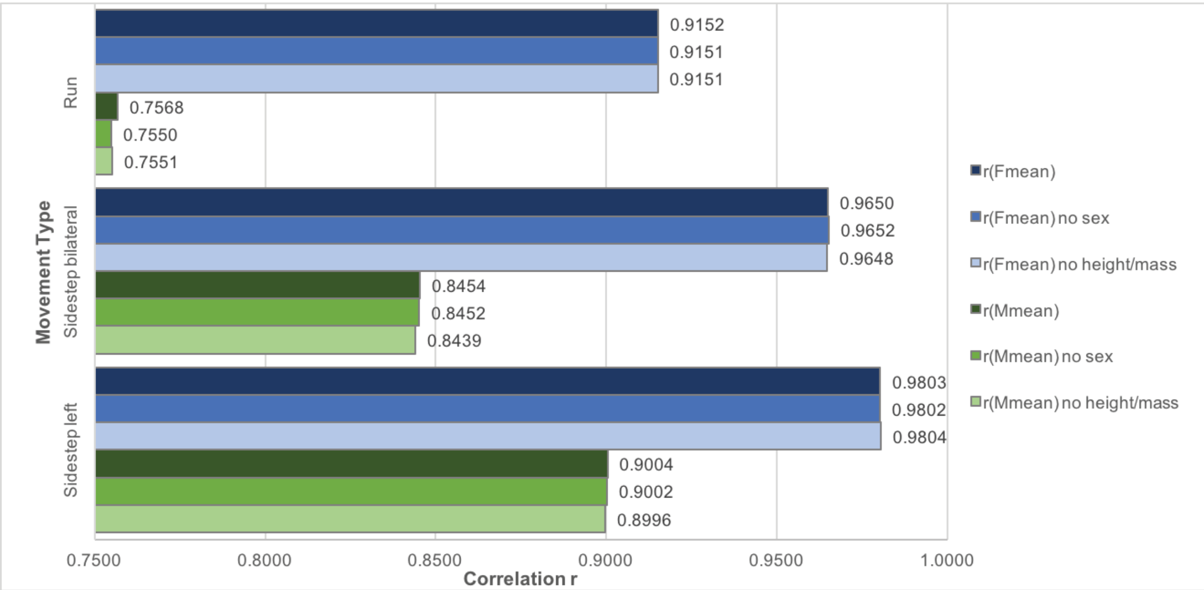

Figure 2: Comparison of mean prediction GRF/Ms correlations.

To illustrate the prediction of the PLS model, the individual sample of movement type 'sidestep left' (data fold one) with the maximum $r(F_{mean})$ 0.9994 is shown in Figure 3.

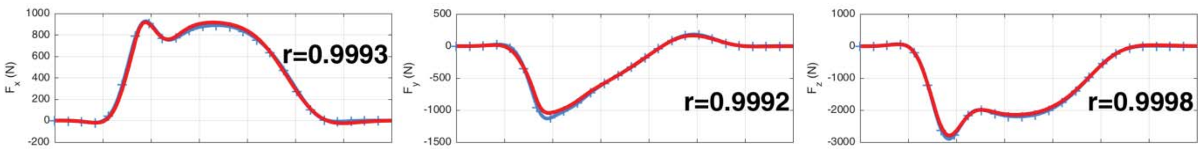

Figure 3: Maximum prediction GRFs correlations 'sidestep left' (known blue ticks, predicted red).

CONCLUSION: This study shows that the prediction of GRF/Ms from motion data using PLS supervised learning can be achieved without prior knowledge of participant sex, height and mass. The predicted mean correlations of GRF/Ms reported are higher than maxima in the literature and obtained using fewer markers, however the results illustrate both the suitability and some of the limitations of approaches employing PLS. Lessons learned in the data mining and pre-processing can now be applied to more advanced regression techniques. This study is the first practical application of predicting GRF/Ms from marker trajectories of complex sporting movements without the use of a force plate. The discovery that discrete inputs of sex, height and mass are not required is a major simplification of data capture and contributes to the goal of real-time accurate GRF/Ms outside of laboratory settings. Paired with less invasive methods of motion capture (computer vision or inertial sensors), the overarching goal of this project to achieve accurate real-time prediction of GRF/Ms in the field is within reach. The independence of the model to the participants' sex (female or male) may also have implications for female athlete training and injury prevention. Large-scale mining of archive biomechanics data is novel, and this study illustrates the practical outcomes which can be achieved from such a big data approach.

REFERENCES:

Barre, Arnaud, & Armand, Stéphane. (2014). Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine, 114*(1), 80-87.

Besier, Thor F., Sturnieks, Daina L., Alderson, Jacque A., & Lloyd, David G. (2003). Repeatability of gait data using a functional hip joint centre and a mean helical knee axis. *Journal of Biomechanics, 36*(8), 1159-1168.

Chen, Nutan, Urban, Sebastian, Osendorfer, Christian, Bayer, Justin, & Van Der Smagt, Patrick. (2014). *Estimating finger grip force from an image of the hand using convolutional neural networks and gaussian processes.* Paper presented at the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong.

Chun, Hyonho, & Keleş, Sündüz. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(1), 3-25.

Elliott, Bruce, & Alderson, Jacqueline. (2007). Laboratory versus field testing in cricket bowling: A review of current and past practice in modelling techniques. *Sports Biomechanics, 6*(1), 99-108.

Jones, Philippa L., Kerwin, David G., Irwin, Gareth, & Nokes, Len D.M. (2009). Three dimensional analysis of knee biomechanics when landing on natural turf and football turf. *Journal of Medical and Biological Engineering, 29*(4), 184-188.

Liu, Tao, Inoue, Yoshio, & Shibata, Kyoko. (2010). A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability. *Sensors, 10*(11), 10240-10255.

Milner, Clare E., & Paquette, Max R. (2015). A kinematic method to detect foot contact during running for all foot strike patterns. *Journal of Biomechanics, 48*(12), 3502-3505.

Müller, Clemens, Sterzing, Thorsten, Lange, Justin, & Milani, Thomas L. (2010). Comprehensive evaluation of player-surface interaction on artificial soccer turf. *Sports Biomechanics, 9*(3), 193-205.

Oh, Seung Eel, Choi, Ahnryul, & Mun, Joung Hwan. (2013). Prediction of ground reaction forces during gait based on kinematics and a neural network model. *Journal of Biomechanics, 46*(14), 2372-2380.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sim, Taeyong, Kwon, Hyunbin, Oh, Seung Eel, Joo, Su-Bin, Choi, Ahnryul, Heo, Hyun Mu, . . . Mun, Joung Hwan. (2015). Predicting complete ground reaction forces and moments during gait with insole plantar pressure information using a wavelet neural network. *Journal of Biomechanical Engineering, 137*(9), 091001:091001-091009.

Soo Park, Hyun, & Shi, Jianbo. (2016). *Force from motion: Decoding physical sensation in a first person video.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV.

Ward, Stephen. (2011). *Anthropometric data and Australian populations – do they fit?* Paper presented at the HFESA 47th Annual Conference 2011, Crows Nest, NSW, Australia.

Wei, Xiaolin, & Chai, Jinxiang. (2010). *Videomocap: Modeling physically realistic human motion from monocular video sequences.* Paper presented at the ACM Transactions on Graphics (TOG), Seoul, South Korea.

Yanai, Toshimasa, Matsuo, Akifumi, Maeda, Akira, Nakamoto, Hiroki, Mizutani, Mirai, Kanehisa, Hiroaki, & Fukunaga, Tetsuo. (2017). Reliability and validity of kinetic and kinematic parameters determined with force plates embedded under soil-filled baseball mound. *Journal of Applied Biomechanics, 0*(0), 1-18.

Acknowledgement

This project is partially supported by the NVIDIA GPU Hardware Grant Program, by ARC Grant DP160101458 and an Australian Government Research Training Program Scholarship.