Pogson, MA, Verheul, J, Robinson, MA, Vanrenterghem, J and Lisboa, P

A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities

A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities

Mark Pogson[a,b]*, Jasper Verheul[c], Mark A. Robinson[c], Jos Vanrenterghem[c,d], Paulo Lisboa[b]

[a] Quintessa Ltd., Henley-on-Thames, UK

[b] Department of Applied Mathematics, Liverpool John Moores University, Liverpool, UK

[c] Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

[d] Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium

* Corresponding author: markpogson@quintessa.org, +44 (0)151 231 2155

Abstract

Prediction of ground reaction force (GRF) magnitudes during running-based sports has several important applications, including optimal load prescription and injury prevention in athletes. Existing methods typically require information from multiple body-worn sensors, limiting their ecological validity, or aim to estimate discrete force parameters, limiting their ability to assess overall biomechanical load. This paper presents a neural network method to predict GRF time series from a single, commonly used, trunk-mounted accelerometer. The presented method uses a principal component analysis and multilayer perceptron (MLP) to obtain predictions. Time-series r^2 correlation with test data averaged around 0.9 for each impact, comparing favourably with alternative approaches which require additional sensors. For the impact peak, r^2 correlation was 0.74 across activities, comparing favourably with correlation analysis approaches. Several modifications, such as subject-specific training of the MLP, may help to improve results further, but the presented method can accurately predict GRF from trunk accelerometry data without requiring additional information. Results demonstrate the scope of machine learning to exploit common wearable technologies to estimate GRF in sport-specific environments.

Keywords: ground reaction force; trunk accelerometry; multilayer perceptron; biomechanical load

1. Introduction

Fitness and fatigue of individuals are affected by the physiological and biomechanical loads they experience. Training loads (i.e. volume and intensity) are therefore monitored with the aim to ensure appropriate training to optimise performance while avoiding injury [1]. Despite common focus on monitoring physiological load (e.g. heart-rate, blood lactate) [2], biomechanical loads are still relatively unexplored [3], mainly due to the difficulty of measuring them in sport-specific environments [4].

Body-worn sensors are commonly used to quantify training loads experienced during sports, often attached to the upper torso to measure position (using GPS) and trunk acceleration (TA). While GPS data can be used to assess physiological load metrics [5], TA data have scope to infer physical characteristics of motion [6]. Given the widespread use of trunk-mounted accelerometers [5], it would be of particular value if TA could be used to estimate biomechanical loads in the field without requiring any additional measurements.

Ground reaction force (GRF) is the force exerted on the body when it contacts the ground, such as when walking and running [7], and thus provides a suitable estimate of biomechanical loads experienced by the body as a whole for many activities [4]. GRF can be measured accurately in laboratory settings using force platforms rigidly mounted to the ground, but alternative approaches are required to estimate GRF in sport-specific environments.

Various approaches have been used to obtain GRF in non-laboratory settings. For example, instrumented insoles worn in the shoe have been used to measure foot pressure and thus estimate GRF directly [8, 9, 10]. However, these insoles face several technical challenges, including additional bulk in footwear and difficulty in accounting for frictional forces. Other studies have used data from motion capture as input to mechanical models [11] or neural network models [12, 13] to predict GRF, but motion capture technologies are expensive and currently not field viable. Combinations of multiple body-worn sensors have also been used to estimate GRF [14, 15, 16]. Although these investigations are promising, their application is either restricted to a limited range of movements or not yet feasible in the field.

To overcome the issues outlined above, several studies have aimed to predict GRF from a single TA signal. Correlation analysis approaches (CAA) for example [17, 18], have been used to estimate discrete force characteristics such as peak force, but not the full time series of each impact, which limits their scope for a detailed assessment of the overall biomechanical loading. Other methods have aimed to predict full GRF time series from TA using various mechanical methods (e.g. mass-spring models [19]) but found that TA alone is probably insufficient with these methods [19, 20, 21, 22, 23].

To exploit commonly used technology, overcome limitations of mechanical models, avoid requirements for additional data and enhance application across tasks and subjects, the present study considers a data-driven approach to predict GRF from TA, using magnitudes of the vector quantities. An artificial neural network was used to estimate the time series of GRF for individual impacts from TA alone; this was primarily motivated by widespread data availability rather than purely biomechanical reasons. Unlike other studies, the method presented in this paper (1) exploits commonly used TA signals, (2) requires one input signal only and does not require additional input information, and (3) generalises over a variety of different tasks and subjects.

2. Method

2.1 Data collection

TA and GRF data were obtained from 15 physically active team-sport players (10 males and 5 females, age 23±1 years, height 1.74±0.08 m, mass 74±9 kg). Participants provided informed consent according to Liverpool John Moores University ethics regulations.

Each subject performed straight overground accelerated, decelerated and constant-speed running trials, ranging between 2-8m/s with 1m/s increments [22]. These tasks were chosen to provide data for a range of activities reflective of those typically performed during running-based sports. Running speeds for the overground trials were measured using photocell timing gates (Brower Timing Systems, Draper, UT, USA) and controlled through verbal feedback to participants after each trial. Only trials within ±5% of each target speed were used. A total of approximately 40 trials were recorded per subject, dependent on their maximal sprinting speed. The three Cartesian components of TA were recorded at a sampling frequency of 100Hz using a GPS-embedded accelerometer (MinimaxX S5, Catapult Innovations, Scoresby, Australia, sampling frequency 100Hz, measurement range ±16g, resolution 16-bit) worn in a tight-fitting vest on the back of the upper torso. The three Cartesian components of GRF were synchronously recorded at 3000Hz using a force platform built into the ground (9287B, 90x60 cm, Kistler Holding AG, Winterthur, Switzerland). Impacts were isolated by a 20N GRF threshold, and GRF signals were smoothed with a 50Hz low-pass Butterworth filter. Resultant accelerations and forces (i.e. the magnitude of each vector quantity) were calculated and used throughout the study. Representative examples of data for each task are shown in Fig. 1.

2.2 Data preparation

Seven of the subjects were used for training and the remaining 8 subjects were used for testing of the method to assess the generalising capabilities of the model with limited training data; this was repeated 10 times with random test-train splits to account for the effects of arbitrary splitting. Since each subject performed approximately 40 trials, the training data set comprised just under 300 trials and the test data just over 300 trials for each repetition. To assess generalisability across subjects, separate subjects were used for training and testing, rather than randomly splitting trials across all subjects. To assess generalisability across tasks, different impact activities were not distinguished in the data (i.e. the MLP was not trained separately for different activities). Each impact was treated as a separate trial, and each time point was treated as a parameter.

To apply the PCA and MLP, the same number of parameters was required for all trials. Therefore, for the training data, zero padding was applied at the end of each signal, up to the length of the longest signal present; for the test data, each TA signal was either zero padded or truncated to the same length as the training data. Depending on the train-test split, the fixed length of each GRF signal was around 1200 time points, with the average zero-padding length around 600.

The zero padding was removed from time-series test statistics to avoid positive bias (since zero padding would provide artificial agreement, assuming signal length is predicted well); this was performed by identifying the length of zero padding in the test signal and removing this from the end of both the test signal and prediction. Zero padding was performed rather than standardising the length of each signal by rescaling due to the adverse effect this would have on time dynamics in the regression.

2.3 Principal component analysis (PCA)

PCA was used primarily to reduce noise in predictions [24]. The method was found to be fairly insensitive to the exact number of components retained, as reflected in the results of the optimisation, described below. A PCA (i.e. decomposition) was performed on both the TA and GRF training data to obtain the inputs and outputs respectively for use in the MLP, as shown in Fig. 2. The number of components used is given in Section 2.7.

2.4 Multilayer perceptron (MLP)

An MLP was used to map the PCA components of TA (model inputs) onto the PCA components of GRF (model outputs). An MLP uses hidden layers of nodes which combine inputs according to an activation function and weights, producing a nonlinear regression between inputs and outputs [24]; it was used due to its good approximation properties as a generic non-linear model. Training data were used to obtain weights which minimise squared output errors, and the resulting regression (the mapping) was stored to be used for prediction, as shown in Fig. 2. Following prediction by the mapping, the inverse of the output decomposition was applied to transform results into the original parameter space (i.e. a time series). Sections 2.7 and 2.8 give further details of the MLP.

2.5 Optimisation

To help explore suitable values to use for the number of PCA components and the number of MLP hidden layers and nodes (here termed options), an optimisation procedure was used. Initial testing found the method was effective using largely arbitrary options, but with a small number of poor predictions present, as shown in the Results. Optimisation was therefore applied to maximise the sum of the mean and the minimum r^2 correlation coefficient obtained in testing, with the aim of improving not only the average r^2 of estimates, but also reducing the number and magnitude of outliers. A stochastic particle swarm optimisation (PSO) was used to allow for irregularities in the search space [25]. Repeated optimisation identified several similar options which improved results to a similar extent, but since the intention of the optimisation was to explore how the method may be improved, rather than to obtain specific options for future use, results are only presented for one of these sets of options (see Section 2.7), without consideration of their relative performance.

All optimisations returned similar values. Lower and upper bounds on the search space were set as 2 to 20 PCA components, and between 1 layer with 1 node and 10 layers each with 100 nodes for the MLP. In all cases, slightly more PCA components were obtained for GRF than TA. Optimised results tended to be close to the centre of the search space, i.e. around for 10 for the PCA components, and around 5 layers with 50 nodes each for the MLP.

2.6 Signal length parameter

Initial analysis of results suggested that poor predictions were most common for particularly short or long signals (where the duration is measured prior to zero-padding/truncation; see Section 2.2). Therefore, signal length was appended to the MLP inputs and outputs so that the mapping was better able to account for signal length.

2.7 Evaluation

Results were obtained using 3 sets of options:

1. No explicit signal length parameter; 5 PCA components for inputs and outputs; 2 MLP hidden layers, each with 5 nodes.
2. Signal length appended to inputs and outputs; other options as 1.
3. Signal length appended to inputs and outputs; other options obtained from the optimisation, i.e. 6 PCA components for inputs and 8 for outputs; 5 MLP hidden layers, with 45, 36, 45, 82 and 40 nodes respectively.

Results are presented without normalising GRF to the subject mass; this is considered further in the Discussion. Methods to evaluate the effectiveness of predictions are described below.

2.7.1 Time-series metrics

Following training, predictions of GRF were obtained for the test data; the r^2 correlation coefficient and RMSE were calculated for the time-series of each impact, with zero-padding removed.

2.7.2 Individual trials

Time-series profiles of several representative impacts were plotted to observe the accuracy of GRF predictions throughout different movement activities.

2.7.3 Further metrics

The impact peak (i.e. the magnitude of the initial force peak; see Fig. 1), loading rate (i.e. the gradient of the GRF curve from touch-down to the impact peak) and impulse (i.e. the area under the curve) were calculated from GRF predictions and corresponding test data. Results were obtained for 164 steady running impacts, 72 accelerations and 96 decelerations. The impact peak was defined as the first distinct peak in the GRF curve; if no peak occurred within the first 30% of a signal, it was excluded from calculations for the impact peak and loading rate, which means respectively 51%, 28% and 98% of trials were used in these calculations.

2.8 Implementation

The method was programmed in Python 2.7 [26]. The library Scikit-learn 0.18.1 was used for the PCA and MLP [27]. The library Pyswarm was used for the PSO [28]. Default options were used unless otherwise stated above.

3. Results

3.1 Time-series metrics

Results for the time-series metrics are shown in Fig. 3. Even with arbitrary options, the method performed well, with an average r^2 around 0.8 (Fig. 3a(i)). It is worth noting that if zero padding was retained in results, an r^2 closer to 0.9 was obtained (results not shown), which suggests the duration of impact is predicted well in general.

The reasonably narrow range of the boxes in Fig. 3 demonstrates the consistency of the method with different train-test splits and random initialisation of weights in the MLP. Inter-subject variability is also small, with a standard deviation of less than 0.05 for the within-subject means of impact r^2. However, there are several poor predictions evident, with some r^2 values for individual impacts around 0. As stated in the Methods, these seem mostly to be for either particularly short or long signals. The explicit signal length parameter appears to have enabled the MLP to distinguish these more effectively (Fig. 3a(ii)), with the number and size of outlying predictions reduced. These improvements were consistent between r^2 and RMSE metrics, as shown by comparison of Fig. 3a and b.

The use of optimisation to identify more suitable options showed small further improvements to predictions, as shown in Fig. 3a(iii) and b(iii), with average r^2 around 0.9; however, there remain a small number of poor predictions. It is worth noting that only a relatively small number of PCA components was required to obtain accurate predictions, as evident by comparing the number of time frames shown in Fig. 1 with the number of PCA components used (Section 2.7).

3.2 Individual trials

Examples of individual time-series predictions using optimised options are shown in Fig. 4 for different types of impact. Note that although good and bad examples are presented, results in Fig. 3 show that predictions tended towards the former.

Several important details are illustrated in Fig. 4. Steady running was predicted well in general (Fig. 4a), but the small impact peak was liable to be underestimated in predictions (Fig. 4d), presumably due to its small effect on the squared output error in MLP training. Acceleration tasks were often predicted well in terms of r^2 (Fig. 4b); however, the shape and magnitude of acceleration curves was sometimes predicted poorly (Fig. 4e). Prediction of deceleration showed high variability; these curves are particularly distinctive, and both strong (Fig. 4c) and weak (Fig. 4f) predictions are evident. It is interesting that in the case of Fig. 4e-f, predictions were effectively for the wrong type of activity (i.e. steady running), which may be due to the distribution of activities present in the training data, where steady running impacts were most frequent. As stated previously, poor predictions were more likely to occur with particularly long (e.g. Fig. 4d-e) or short signals, even when signal length was appended to MLP inputs and outputs; again, this may partly be due to the distribution of training data used. These issues are considered further in the Discussion.

3.3 Further metrics

Results are shown in Fig. 5 for impact peak, loading rate and impulse. Predictions were reasonable for the impact peak and loading rate (Fig. 5d, g), with r^2 of 0.74 and 0.63 respectively, but generally weaker for the impulse (Fig. 5a). There was a slight tendency to overestimate small impact peaks and vice versa, which largely coincided with different activities, where accelerations tended to have smaller impact peaks and decelerations tended to have larger ones (Fig. 5e). There was a stronger tendency to underestimate loading rate, particularly for larger values, which were more common for steady running and deceleration (Fig. 5h). The impulse tended to be overestimated for small values and vice versa (Fig. 5b). Percentage errors were fairly uniform for impact peak and loading rate (Fig. 5f, i), but showed greater dependence on the size of measurements for the impulse (Fig. 5c) due to poor predictions of small impulses, which are most common for particularly short signals; this was partly task-dependent, with steady running tending to have smaller errors than other activities (Fig. 5b-c).

4. Discussion

This study used a neural network method to predict GRF magnitude from TA magnitude for a variety of different running tasks across different individuals. GRF was predicted with an average r^2 of around 0.9 for the time series of each impact and, therefore, the method offers a promising approach to estimate GRF in the field. Since this method exploits commercially available devices which are already widely used, it would be both cost-effective and easy to implement. The method is computationally efficient, requiring a matter of seconds to run on a standard desktop computer for both training and testing in the present study.

Several previous studies have attempted to predict GRF characteristics from trunk-work accelerometers, but time-series prediction have been found to be poor [19, 20, 21]. As an alternative to full GRF waveforms, CAA has been used to predict peak GRF from TA data. Tran et al. [16] obtained r^2 values of 0.3 to 0.5, while Hollville et al. [24] obtained 0.55 to 0.8; the impact peak is the closest comparable metric in the present study, which had an r^2 of 0.74 across activities (Fig. 5d) and thus compares well. Neugebauer et al. (2014) found errors from a hip accelerometer of between 8.3-17.8% [29]; this is comparable to most trials shown in Fig. 5f, although a significant number of outliers are noted.

The present method also compares well against approaches which require new devices to be deployed. Such approaches appear to offer similar accuracy for time-series predictions of individual impacts; for example, foot sensors give an RMSE of around 10% or higher [21, 9, 10], similar to Fig. 3b. Gurchiek et al. [23] used a single sensor to predict GRF, but it also required gyroscope data and subject mass to be used in a mechanical model; Gurchiek et al. [23] reported r^2 values of between 0.71 and 0.88 for each impact, which are slightly lower than for the present study. Furthermore, the present study considered a wider range of movements by including decelerations and running at a wide range of different speeds, and it also widened assessment of results to include key metrics of impulse, impact peak and loading rate. Gurchiek et al. [23] found that the assumption of the sensor being at the subject centre of mass throughout each impact was a major limitation to obtain accurate instantaneous GRF estimates.

Prediction of GRF-derived metrics (i.e. impact peak, loading rate and impulse) may be improved within the present method by fitting the MLP to these directly, rather than calculating them from the predicted GRF time series. This is a promising area to investigate further. Aside from PCA, alternative methods for data decomposition exist, such as non-negative matrix factorisation (NFM). However, investigation showed that NFM performed very similarly, suggesting the overall method is not highly sensitive to the form of decomposition (results not shown). Outlying predictions were found to be most common with particularly long or short signals, which can be identified without the need for test data. Additional training data may also help to reduce outlying predictions; separate mappings could also be obtained for different signal lengths, although this did not improve predictions using the present data (results not shown). Using an explicit signal length parameter helped to reduce outliers (Fig. 1), but outliers were still most common for long and short signals, hence these should be investigated further. The present study used training data with a predominance of steady running activities, but outliers could possibly be reduced by using a more even spread of activities in training. For example, as noted in Fig. 4, steady running was in effect predicted for some acceleration and deceleration tasks, perhaps because these movements were less represented in the model fitting.

The current study has not considered the importance of sensor location on the body. The TA signal was primarily used because it is arguably the most widely measured acceleration signal in the field [5]. Results may be improved by placing the accelerometer elsewhere on the body, and since the impact peak is likely related to lower limb accelerations during landing [11, 30], the addition of signals from the lower limbs might improve predictions.

While the method has been found to generalise well across individuals, it should be noted that all subjects in the present study were of similar age, mass and athleticism. While it might be expected that normalising GRF by subject body mass would improve predictions due to the Newtonian relationship with acceleration, it had no significant effect in the present study (results not shown). This may reflect the similarity of subjects, but the importance of other factors also remains to be explored, including limb lengths and body-segment masses. While the present study aimed to provide predictions from minimal information, these effects should be considered in further work. Given that individuals are likely to have consistent forms of movement based on their physical and biological characteristics, subject-specific training could prove particularly effective, which would capitalise on the adaptability of the machine learning approach.

Acknowledgements

Competing interests: None declared

Funding: None

Ethical approval: Approval obtained in accordance with Liverpool John Moores University ethics regulations from the Research Ethics Committee (REC) under reference number 16/SPS/017

References

[1] Gabbett, T.J. and Ullah, S., 2012. Relationship between running loads and soft-tissue injury in elite team sport athletes. The Journal of Strength & Conditioning Research, 26(4), pp.953-960.

[2] Lambert, M.I. and Borresen, J., 2010. Measuring training load in sports. International journal of sports physiology and performance, 5(3), pp.406-411.

[3] Vanrenterghem, J., Nedergaard, N.J., Robinson, M.A. and Drust, B., 2017. Training load monitoring in team sports: A novel framework separating physiological and biomechanical load-adaptation pathways. Sports Medicine, 47(11), pp.2135-2142.

[4] Verheul, J., Nedergaard N.J., Vanrenterghem, J., Robinson, M.A., 2020. Measuring biomechanical loads in team sports – from lab to field. Science and Medicine in Football. doi:10.1080/24733938.2019.1709654.

[5] Akenhead, R. and Nassis, G.P., 2016. Training load and player monitoring in high-level football: current practice and perceptions. International journal of sports physiology and performance, 11(5), pp.587-593.

[6] Schütte, K.H., Maas, E.A., Exadaktylos, V., Berckmans, D., Venter, R.E. and Vanwanseele, B., 2015. Wireless tri-axial trunk accelerometry detects deviations in dynamic center of mass motion due to running-induced fatigue. PloS one, 10(10), p.e0141957.

[7] Nilsson, J. and Thorstensson, A., 1989. Ground reaction forces at different speeds of human walking and running. Acta Physiologica, 136(2), pp.217-227.

[8] Cordero, A.F., Koopman, H.J.F.M. and Van Der Helm, F.C.T., 2004. Use of pressure insoles to calculate the complete ground reaction forces. Journal of Biomechanics, 37(9), pp.1427-1432.

[9] Liu, T., Inoue, Y. and Shibata, K., 2010. A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability. Sensors, 10(11), pp.10240-10255.

[10] Jung, Y., Jung, M., Lee, K. and Koo, S., 2014. Ground reaction force estimation using an insole-type pressure mat and joint kinematics during walking. Journal of biomechanics, 47(11), pp.2693-2699.

[11] Clark, K.P., Ryan, L.J. and Weyand, P.G., 2017. A general relationship links gait mechanics and running ground reaction forces. Journal of Experimental Biology, 220(2), pp.247-258.

[12] Choi, A., Lee, J.M. and Mun, J.H., 2013. Ground reaction forces predicted by using artificial neural network during asymmetric movements. International Journal of Precision Engineering and Manufacturing, 14(3), pp.475-483.

[13] Oh, S.E., Choi, A. and Mun, J.H., 2013. Prediction of ground reaction forces during gait based on kinematics and a neural network model. Journal of biomechanics, 46(14), pp.2372-2380.

[14] Guo, Y., Storm, F., Zhao, Y., Billings, S., Pavic, A., Mazzà, C., Guo, L.-Z., 2017. A New Proxy Measurement Algorithm with Application to the Estimation of Vertical Ground Reaction Forces Using Wearable Sensors. Sensors 17, 2181. doi:10.3390/s17102181

[15] Johnson, W.R., Mian, A., Donnelly, C.J., Lloyd, D., Alderson, J., 2018. Predicting athlete ground reaction forces and moments from motion capture. Med. Biol. Eng. Comput., 56(10), pp.1781-1792.

[16] Wouda, F.J., Giuberti, M., Bellusci, G., Maartens, E., Reenalda, J., van Beijnum, B.-J.F., Veltink, P.H., 2018. Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors. Front. Physiol. doi:10.3389/fphys.2018.00218.

[17] Wundersitz, D., Netto, K., Aisbett, B. and Gastin, P., 2013. Validity of an upper-body-mounted accelerometer to measure peak vertical and resultant force during running and change-of-direction tasks. Sports Biomechanics, 12, pp.403-12.

[18] Tran, J., Netto, K., Aisbett, B. and Gastin, P., 2010. Validation of accelerometer data for measuring impacts during jumping and landing tasks. In Proceedings of the 28th International Conference on Biomechanics in Sports (2010) (pp. 1-4). International Society of Biomechanics in Sports.

[19] Nedergaard, N.J., Verheul, J., Drust, B., Etchells, T., Lisboa, P.J., Robinson, M.A., Vanrenterghem, J., 2018. The feasibility of predicting ground reaction forces during running from a trunk accelerometry driven mass-spring-damper model. PeerJ, 6:e6105.

[20] Edwards, S., White, S., Humphreys, S., Robergs, R., O'Dwyer, N., 2019. Caution using data from triaxial accelerometers housed in player tracking units during running. Journal of Sport Science, 37(7), pp.810-818.

[21] Nedergaard, N.J., Robinson, M.A., Eusterwiemann, E., Drust, B., Lisboa, P.J., Vanrenterghem, J., 2017. The Relationship Between Whole-Body External Loading and Body-Worn Accelerometry During Team-Sport Movements. International Journal of Sports Physiology & Performance, 12(1), pp.18-26.

[22] Verheul, J., Nedergaard, N.J., Pogson, M., Lisboa, P., Gregson, W., Vanrenterghem, J. and Robinson, M.A., 2019. Biomechanical loading during running: can a two mass-spring-damper model

be used to evaluate ground reaction forces for high-intensity tasks? Sports Biomechanics. doi:10.1080/14763141.2019.1584238.

[23] Gurchiek, R.D., McGinnis, R.S., Needle, A.R., McBride, J.M. and van Werkhoven, H., 2017. The use of a single inertial sensor to estimate 3-dimensional ground reaction force during accelerative running tasks. Journal of biomechanics, 61, pp.263-268.

[24] Hollville, E., Couturier, A., Guilhem, G. and Rabita, G., 2016. MinimaxX player load as an index of the center of mass displacement? A validation study. In ISBS-Conference Proceedings Archive (Vol. 33, No. 1). Hong, Y. and Bartlett, R. eds., 2008. Routledge handbook of biomechanics and human movement science. Routledge.

[25] Kennedy, J., 2011. Particle swarm optimization. In Encyclopedia of machine learning (pp. 760-766). Springer US.

[26] Python Software Foundation, 2017. Python Language Reference, version 2.7. Available at http://www.python.org

[27] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, pp.2825-2830.

[28] Pyswarm, 2017. Available at https://pythonhosted.org/pyswarm/

[29] Neugebauer JM, Collins KH, Hawkins DA (2014) Ground Reaction Force Estimates from ActiGraph GT3X+ Hip Accelerations. PLoS ONE 9(6): e99023. doi:10.1371/journal.pone.0099023.

[30] Bobbert, M.F., Schamhardt, H.C. and Nigg, B.M., 1991. Calculation of vertical ground reaction force estimates during running from positional data. Journal of Biomechanics, 24(12), pp.1095-1105.

Figure captions

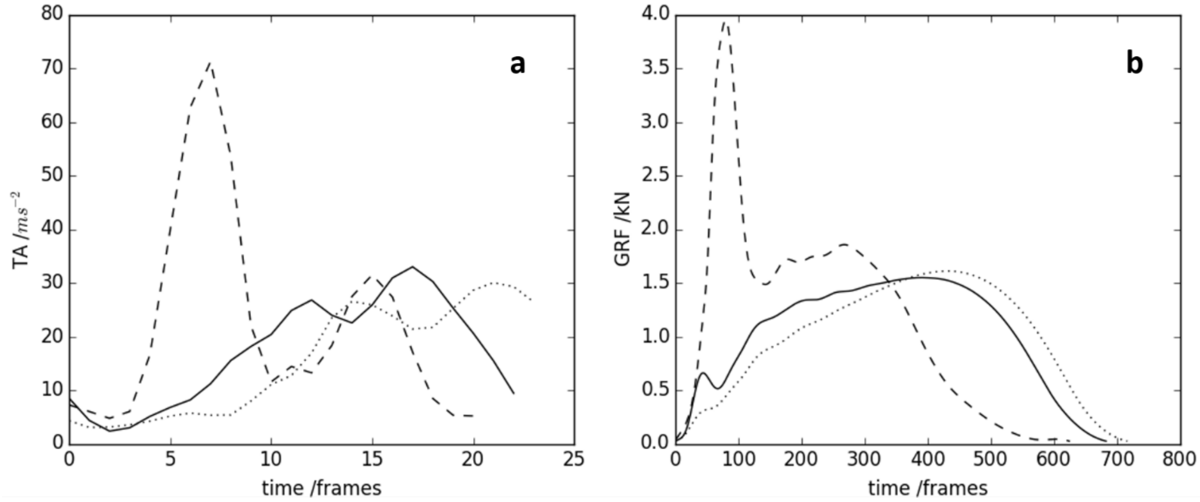
Fig. 1. Example TA and GRF data for 3 impacts. (a) TA for a single impact of: a steady running task (solid line), an acceleration task (dotted line) and a deceleration task (dashed line). (b) The corresponding GRF for each task. Note that the difference in time frames for TA and GRF is due to the sampling frequencies used; each impact is around 0.25s, but time frames are presented to make clear the form of data being modelled.

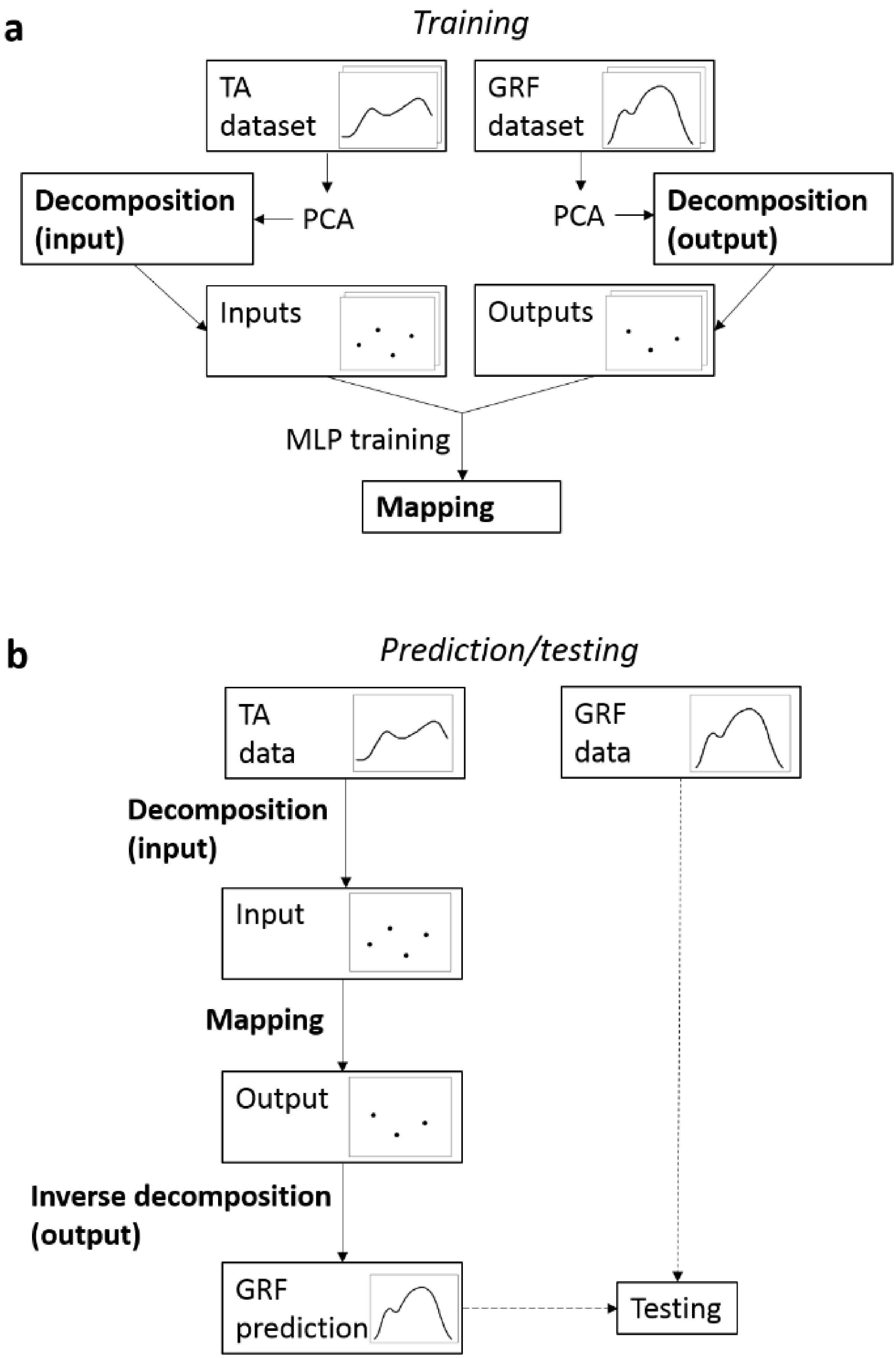
Fig. 2. Steps required for training and prediction. (a) Training. A PCA is performed on the entire dataset; the transformation is stored as the decomposition, and the resulting transformed data are used as inputs and outputs for the MLP. The MLP obtains a nonlinear regression between inputs and outputs, which is stored as the mapping. (b) Prediction. Each TA impact is transformed using the decomposition and then used as an input to the mapping. The output of the mapping is transformed using the inverse of the output decomposition to give the prediction of GRF, which can be compared against real data for testing purposes.

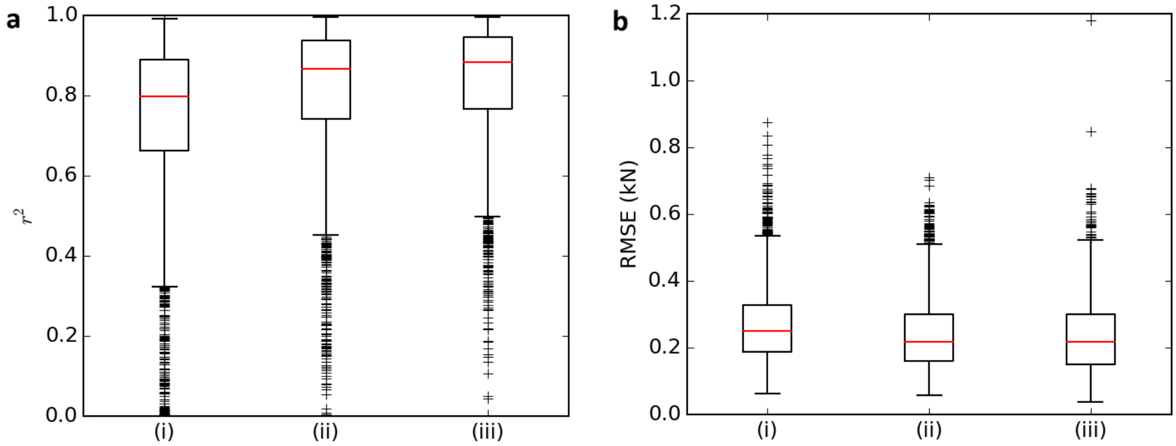
Fig. 3. Test predictions from 10 repetitions of training. (a) r^2 correlation coefficient between prediction and data; (b) corresponding RMSE. Results obtained using: (i) arbitrary options, without appending signal length; (ii) arbitrary options, with signal length appended; and (iii) optimised options, with signal length appended; optimised option values are shown in Table 1. Central horizontal lines show median values. Boxes show the interquartile range. Whiskers show up to 1.5 times the interquartile range; any values outside this are plotted as individual points, showing the r^2 or RMSE value for a particular impact.

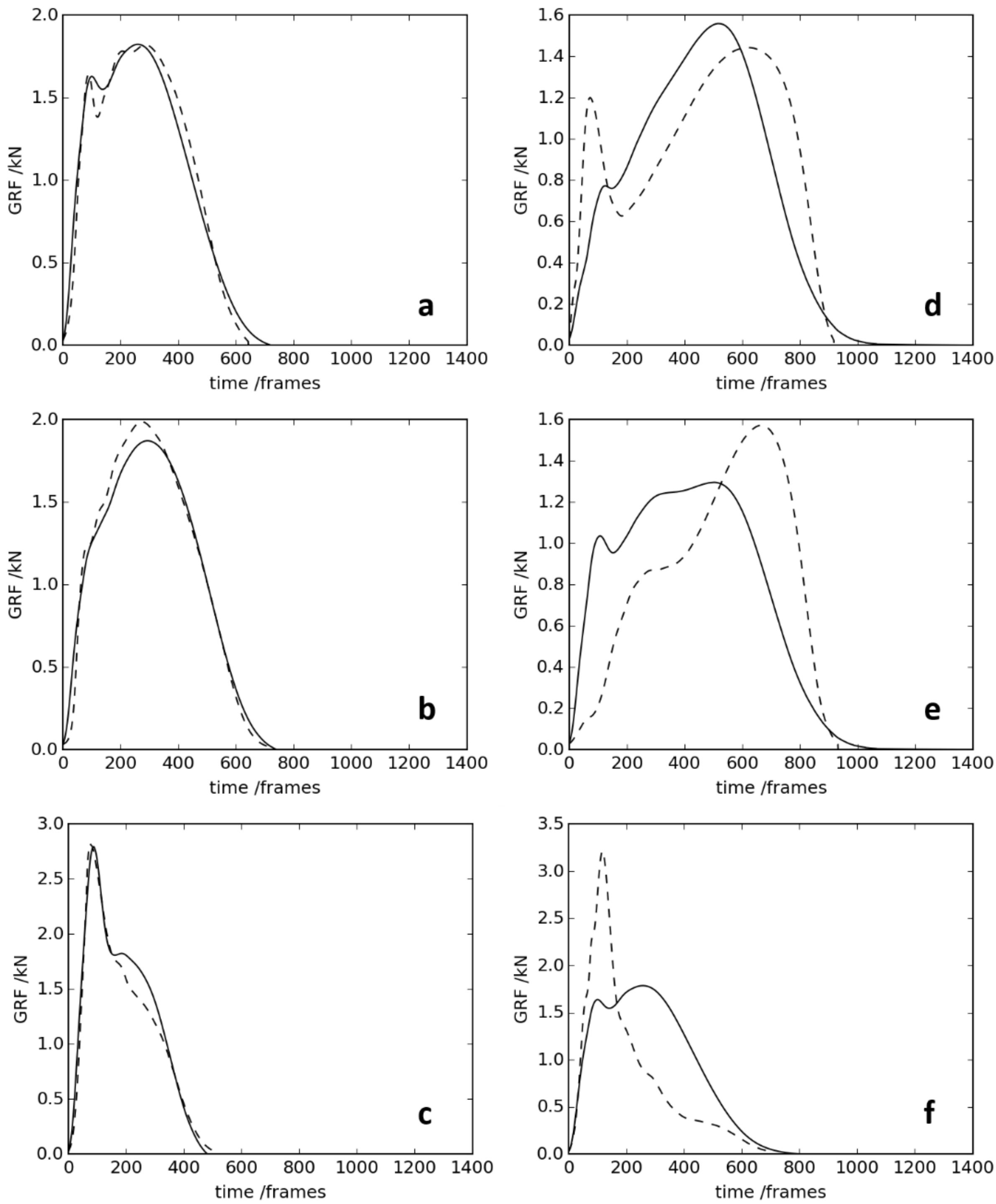
Fig. 4. Example time-series predictions for different tasks. Solid lines show predicted GRF and dashed lines show corresponding test data. (a)-(c) show good predictions for an example of steady running, acceleration and deceleration respectively; (d)-(f) show weaker predictions for similar activities (n.b. such predictions are less common, as shown in Fig. 3). The time scale corresponds to 0.47s.

Fig. 5. GRF metrics used to assess load. Results are shown for: impulse (a-c), impact peak (d-f) and loading rate (g-i). Upper row (a, d, g): predicted versus measured values; central row (b, e, h): errors versus measured values; lower row (c, f, i): percentage errors versus measured values. Each trial is marked according to the activity: steady running (crosses), accelerations (circles), and decelerations (triangles). Each metric was divided by the subject mass for consistency across subjects. For percentage errors, the error of each trial (i.e. the difference between the prediction and measurement) was calculated as a percentage of the measurement.

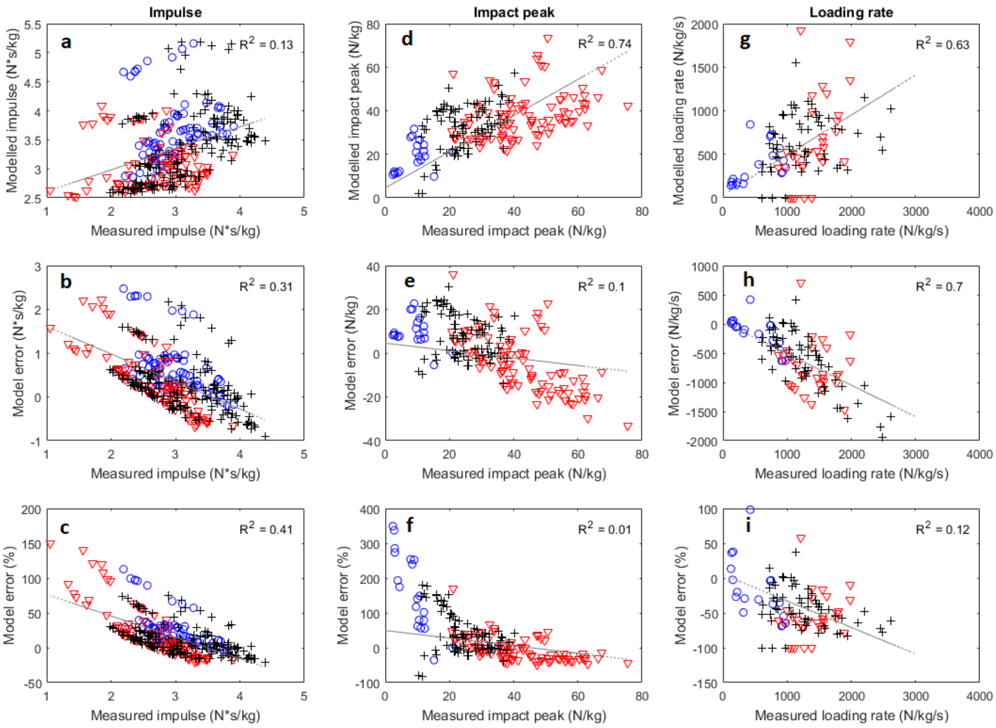

a
TA /ms^{-2}
time /frames
0
5
10
15
20
25
0
10
20
30
40
50
60
70
80
b
GRF /kN
time /frames
0
100
200
300
400
500
600
700
800
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


a
Training
TA dataset
GRF dataset
PCA
PCA
Decomposition (input)
Decomposition (output)
Inputs
Outputs
MLP training
Mapping
b
Prediction/testing
TA data
GRF data
Decomposition (input)
Input
Mapping
Output
Inverse decomposition (output)
GRF prediction
Testing


a
1.0
0.8
0.6
0.4
0.2
0.0
r^2
(i)
(ii)
(iii)
b
1.2
1.0
0.8
0.6
0.4
0.2
0.0
RMSE (kN)
(i)
(ii)
(iii)


a
b
c
d
e
f
GRF /kN
time /frames
0 200 400 600 800 1000 1200 1400

431


Impulse
Impact peak
Loading rate
a
b
c
d
e
f
g
h
i
R^2 = 0.13
R^2 = 0.31
R^2 = 0.41
R^2 = 0.74
R^2 = 0.1
R^2 = 0.01
R^2 = 0.63
R^2 = 0.7
R^2 = 0.12
Modelled impulse (N*s/kg)
Modelled impact peak (N/kg)
Modelled loading rate (N/kg/s)
Model error (N*s/kg)
Model error (N/kg)
Model error (N/kg/s)
Model error (%)
Measured impulse (N*s/kg)
Measured impact peak (N/kg)
Measured loading rate (N/kg/s)