

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

3 Mark Pogson<sup>a,b\*</sup>, Jasper Verheul<sup>c</sup>, Mark A. Robinson<sup>c</sup>, Jos Vanrenterghem<sup>c,d</sup>, Paulo Lisboa<sup>b</sup>

4

5 <sup>a</sup> Quintessa Ltd., Henley-on-Thames, UK

6 <sup>b</sup> Department of Applied Mathematics, Liverpool John Moores University, Liverpool, UK

7 <sup>c</sup> Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

8 <sup>d</sup> Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium

9 \* Corresponding author: [markpogson@quintessa.org](mailto:markpogson@quintessa.org), +44 (0)151 231 2155

## 10 **Abstract**

11 Prediction of ground reaction force (GRF) magnitudes during running-based sports has several  
12 important applications, including optimal load prescription and injury prevention in athletes. Existing  
13 methods typically require information from multiple body-worn sensors, limiting their ecological  
14 validity, or aim to estimate discrete force parameters, limiting their ability to assess overall  
15 biomechanical load. This paper presents a neural network method to predict GRF time series from a  
16 single, commonly used, trunk-mounted accelerometer. The presented method uses a principal  
17 component analysis and multilayer perceptron (MLP) to obtain predictions. Time-series  $r^2$   
18 correlation with test data averaged around 0.9 for each impact, comparing favourably with  
19 alternative approaches which require additional sensors. For the impact peak,  $r^2$  correlation was  
20 0.74 across activities, comparing favourably with correlation analysis approaches. Several  
21 modifications, such as subject-specific training of the MLP, may help to improve results further, but  
22 the presented method can accurately predict GRF from trunk accelerometry data without requiring  
23 additional information. Results demonstrate the scope of machine learning to exploit common  
24 wearable technologies to estimate GRF in sport-specific environments.

25

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

27

## 28 **1. Introduction**

29 Fitness and fatigue of individuals are affected by the physiological and biomechanical loads they  
30 experience. Training loads (i.e. volume and intensity) are therefore monitored with the aim to  
31 ensure appropriate training to optimise performance while avoiding injury [1]. Despite common  
32 focus on monitoring physiological load (e.g. heart-rate, blood lactate) [2], biomechanical loads are

33 still relatively unexplored [3], mainly due to the difficulty of measuring them in sport-specific  
34 environments [4].

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

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

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

55 To overcome the issues outlined above, several studies have aimed to predict GRF from a single TA  
56 signal. Correlation analysis approaches (CAA) for example [17, 18], have been used to estimate  
57 discrete force characteristics such as peak force, but not the full time series of each impact, which

58 limits their scope for a detailed assessment of the overall biomechanical loading. Other methods  
59 have aimed to predict full GRF time series from TA using various mechanical methods (e.g. mass-  
60 spring models [19]) but found that TA alone is probably insufficient with these methods [19, 20, 21,  
61 22, 23].

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

70

## 71 **2. Method**

### 72 *2.1 Data collection*

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

76 Each subject performed straight overground accelerated, decelerated and constant-speed running  
77 trials, ranging between 2-8m/s with 1m/s increments [22]. These tasks were chosen to provide data  
78 for a range of activities reflective of those typically performed during running-based sports. Running  
79 speeds for the overground trials were measured using photocell timing gates (Brower Timing  
80 Systems, Draper, UT, USA) and controlled through verbal feedback to participants after each trial.  
81 Only trials within  $\pm 5\%$  of each target speed were used. A total of approximately 40 trials were

82 recorded per subject, dependent on their maximal sprinting speed. The three Cartesian components  
83 of TA were recorded at a sampling frequency of 100Hz using a GPS-embedded accelerometer  
84 (MinimaxX S5, Catapult Innovations, Scoresby, Australia, sampling frequency 100Hz, measurement  
85 range  $\pm 16g$ , resolution 16-bit) worn in a tight-fitting vest on the back of the upper torso. The three  
86 Cartesian components of GRF were synchronously recorded at 3000Hz using a force platform built  
87 into the ground (9287B, 90x60 cm, Kistler Holding AG, Winterthur, Switzerland). Impacts were  
88 isolated by a 20N GRF threshold, and GRF signals were smoothed with a 50Hz low-pass Butterworth  
89 filter. Resultant accelerations and forces (i.e. the magnitude of each vector quantity) were calculated  
90 and used throughout the study. Representative examples of data for each task are shown in Fig. 1.

91

## 92 *2.2 Data preparation*

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

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

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

113

### 114 *2.3 Principal component analysis (PCA)*

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

120

### 121 *2.4 Multilayer perceptron (MLP)*

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

130

131 *2.5 Optimisation*

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

143 All optimisations returned similar values. Lower and upper bounds on the search space were set as  
144 2 to 20 PCA components, and between 1 layer with 1 node and 10 layers each with 100 nodes for  
145 the MLP. In all cases, slightly more PCA components were obtained for GRF than TA. Optimised  
146 results tended to be close to the centre of the search space, i.e. around for 10 for the PCA  
147 components, and around 5 layers with 50 nodes each for the MLP.

148

149 *2.6 Signal length parameter*

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

154

155 *2.7 Evaluation*

156 Results were obtained using 3 sets of options:

- 157 1. No explicit signal length parameter; 5 PCA components for inputs and outputs; 2 MLP hidden  
158 layers, each with 5 nodes.
- 159 2. Signal length appended to inputs and outputs; other options as 1.
- 160 3. Signal length appended to inputs and outputs; other options obtained from the  
161 optimisation, i.e. 6 PCA components for inputs and 8 for outputs; 5 MLP hidden layers, with  
162 45, 36, 45, 82 and 40 nodes respectively.

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

165

166 *2.7.1 Time-series metrics*

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

169

170 *2.7.2 Individual trials*

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

173

174 *2.7.3 Further metrics*

175 The impact peak (i.e. the magnitude of the initial force peak; see Fig. 1), loading rate (i.e. the  
176 gradient of the GRF curve from touch-down to the impact peak) and impulse (i.e. the area under the

177 curve) were calculated from GRF predictions and corresponding test data. Results were obtained for  
178 164 steady running impacts, 72 accelerations and 96 decelerations. The impact peak was defined as  
179 the first distinct peak in the GRF curve; if no peak occurred within the first 30% of a signal, it was  
180 excluded from calculations for the impact peak and loading rate, which means respectively 51%, 28%  
181 and 98% of trials were used in these calculations.

182

### 183 *2.8 Implementation*

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

187

## 188 **3. Results**

### 189 *3.1 Time-series metrics*

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

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

201 improvements were consistent between  $r^2$  and RMSE metrics, as shown by comparison of Fig. 3a  
202 and b.

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

208

### 209 *3.2 Individual trials*

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

213 Several important details are illustrated in Fig. 4. Steady running was predicted well in general (Fig.  
214 4a), but the small impact peak was liable to be underestimated in predictions (Fig. 4d), presumably  
215 due to its small effect on the squared output error in MLP training. Acceleration tasks were often  
216 predicted well in terms of  $r^2$  (Fig. 4b); however, the shape and magnitude of acceleration curves was  
217 sometimes predicted poorly (Fig. 4e). Prediction of deceleration showed high variability; these  
218 curves are particularly distinctive, and both strong (Fig. 4c) and weak (Fig. 4f) predictions are  
219 evident. It is interesting that in the case of Fig. 4e-f, predictions were effectively for the wrong type  
220 of activity (i.e. steady running), which may be due to the distribution of activities present in the  
221 training data, where steady running impacts were most frequent. As stated previously, poor  
222 predictions were more likely to occur with particularly long (e.g. Fig. 4d-e) or short signals, even  
223 when signal length was appended to MLP inputs and outputs; again, this may partly be due to the  
224 distribution of training data used. These issues are considered further in the Discussion.

225

### 226 3.3 Further metrics

227 Results are shown in Fig. 5 for impact peak, loading rate and impulse. Predictions were reasonable  
228 for the impact peak and loading rate (Fig. 5d, g), with  $r^2$  of 0.74 and 0.63 respectively, but generally  
229 weaker for the impulse (Fig. 5a). There was a slight tendency to overestimate small impact peaks  
230 and vice versa, which largely coincided with different activities, where accelerations tended to have  
231 smaller impact peaks and decelerations tended to have larger ones (Fig. 5e). There was a stronger  
232 tendency to underestimate loading rate, particularly for larger values, which were more common for  
233 steady running and deceleration (Fig. 5h). The impulse tended to be overestimated for small values  
234 and vice versa (Fig. 5b). Percentage errors were fairly uniform for impact peak and loading rate (Fig.  
235 5f, i), but showed greater dependence on the size of measurements for the impulse (Fig. 5c) due to  
236 poor predictions of small impulses, which are most common for particularly short signals; this was  
237 partly task-dependent, with steady running tending to have smaller errors than other activities (Fig.  
238 5b-c).

239

## 240 4. Discussion

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

248 Several previous studies have attempted to predict GRF characteristics from trunk-work  
249 accelerometers, but time-series prediction have been found to be poor [19, 20, 21]. As an alternative  
250 to full GRF waveforms, CAA has been used to predict peak GRF from TA data. Tran et al. [16]  
251 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  
252 closest comparable metric in the present study, which had an  $r^2$  of 0.74 across activities (Fig. 5d) and  
253 thus compares well. Neugebauer et al. (2014) found errors from a hip accelerometer of between  
254 8.3-17.8% [29]; this is comparable to most trials shown in Fig. 5f, although a significant number of  
255 outliers are noted.

256 The present method also compares well against approaches which require new devices to be  
257 deployed. Such approaches appear to offer similar accuracy for time-series predictions of individual  
258 impacts; for example, foot sensors give an RMSE of around 10% or higher [21, 9, 10], similar to Fig.  
259 3b. Gurchiek et al. [23] used a single sensor to predict GRF, but it also required gyroscope data and  
260 subject mass to be used in a mechanical model; Gurchiek et al. [23] reported  $r^2$  values of between  
261 0.71 and 0.88 for each impact, which are slightly lower than for the present study. Furthermore, the  
262 present study considered a wider range of movements by including decelerations and running at a  
263 wide range of different speeds, and it also widened assessment of results to include key metrics of  
264 impulse, impact peak and loading rate. Gurchiek et al. [23] found that the assumption of the sensor  
265 being at the subject centre of mass throughout each impact was a major limitation to obtain  
266 accurate instantaneous GRF estimates.

267 Prediction of GRF-derived metrics (i.e. impact peak, loading rate and impulse) may be improved  
268 within the present method by fitting the MLP to these directly, rather than calculating them from  
269 the predicted GRF time series. This is a promising area to investigate further. Aside from PCA,  
270 alternative methods for data decomposition exist, such as non-negative matrix factorisation (NFM).  
271 However, investigation showed that NFM performed very similarly, suggesting the overall method is  
272 not highly sensitive to the form of decomposition (results not shown). Outlying predictions were

273 found to be most common with particularly long or short signals, which can be identified without the  
274 need for test data. Additional training data may also help to reduce outlying predictions; separate  
275 mappings could also be obtained for different signal lengths, although this did not improve  
276 predictions using the present data (results not shown). Using an explicit signal length parameter  
277 helped to reduce outliers (Fig. 1), but outliers were still most common for long and short signals,  
278 hence these should be investigated further. The present study used training data with a  
279 predominance of steady running activities, but outliers could possibly be reduced by using a more  
280 even spread of activities in training. For example, as noted in Fig. 4, steady running was in effect  
281 predicted for some acceleration and deceleration tasks, perhaps because these movements were  
282 less represented in the model fitting.

283 The current study has not considered the importance of sensor location on the body. The TA signal  
284 was primarily used because it is arguably the most widely measured acceleration signal in the field  
285 [5]. Results may be improved by placing the accelerometer elsewhere on the body, and since the  
286 impact peak is likely related to lower limb accelerations during landing [11, 30], the addition of  
287 signals from the lower limbs might improve predictions.

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

298

299 **Acknowledgements**

300 Competing interests: None declared

301 Funding: None

302 Ethical approval: Approval obtained in accordance with Liverpool John Moores University ethics

303 regulations from the Research Ethics Committee (REC) under reference number 16/SPS/017

304

305

306 **References**

307 [1] Gabbett, T.J. and Ullah, S., 2012. Relationship between running loads and soft-tissue injury in  
308 elite team sport athletes. *The Journal of Strength & Conditioning Research*, 26(4), pp.953-960.

309 [2] Lambert, M.I. and Borresen, J., 2010. Measuring training load in sports. *International journal of*  
310 *sports physiology and performance*, 5(3), pp.406-411.

311 [3] Vanrenterghem, J., Nedergaard, N.J., Robinson, M.A. and Drust, B., 2017. Training load  
312 monitoring in team sports: A novel framework separating physiological and biomechanical load-  
313 adaptation pathways. *Sports Medicine*, 47(11), pp.2135-2142.

314 [4] Verheul, J., Nedergaard N.J., Vanrenterghem, J., Robinson, M.A., 2020. Measuring biomechanical  
315 loads in team sports – from lab to field. *Science and Medicine in Football*.  
316 doi:10.1080/24733938.2019.1709654.

317 [5] Akenhead, R. and Nassis, G.P., 2016. Training load and player monitoring in high-level football:  
318 current practice and perceptions. *International journal of sports physiology and performance*, 11(5),  
319 pp.587-593.

320 [6] Schütte, K.H., Maas, E.A., Exadaktylos, V., Berckmans, D., Venter, R.E. and Vanwanseele, B.,  
321 2015. Wireless tri-axial trunk accelerometry detects deviations in dynamic center of mass motion  
322 due to running-induced fatigue. *PloS one*, 10(10), p.e0141957.

323 [7] Nilsson, J. and Thorstensson, A., 1989. Ground reaction forces at different speeds of human  
324 walking and running. *Acta Physiologica*, 136(2), pp.217-227.

325 [8] Cordero, A.F., Koopman, H.J.F.M. and Van Der Helm, F.C.T., 2004. Use of pressure insoles to  
326 calculate the complete ground reaction forces. *Journal of Biomechanics*, 37(9), pp.1427-1432.

327 [9] Liu, T., Inoue, Y. and Shibata, K., 2010. A wearable ground reaction force sensor system and its  
328 application to the measurement of extrinsic gait variability. *Sensors*, 10(11), pp.10240-10255.

329 [10] Jung, Y., Jung, M., Lee, K. and Koo, S., 2014. Ground reaction force estimation using an insole-  
330 type pressure mat and joint kinematics during walking. *Journal of biomechanics*, 47(11), pp.2693-  
331 2699.

332 [11] Clark, K.P., Ryan, L.J. and Weyand, P.G., 2017. A general relationship links gait mechanics and  
333 running ground reaction forces. *Journal of Experimental Biology*, 220(2), pp.247-258.

334 [12] Choi, A., Lee, J.M. and Mun, J.H., 2013. Ground reaction forces predicted by using artificial  
335 neural network during asymmetric movements. *International Journal of Precision Engineering and*  
336 *Manufacturing*, 14(3), pp.475-483.

337 [13] Oh, S.E., Choi, A. and Mun, J.H., 2013. Prediction of ground reaction forces during gait based on  
338 kinematics and a neural network model. *Journal of biomechanics*, 46(14), pp.2372-2380.

339 [14] Guo, Y., Storm, F., Zhao, Y., Billings, S., Pavic, A., Mazzà, C., Guo, L.-Z., 2017. A New Proxy  
340 Measurement Algorithm with Application to the Estimation of Vertical Ground Reaction Forces Using  
341 Wearable Sensors. *Sensors* 17, 2181. doi:10.3390/s17102181

- 342 [15] Johnson, W.R., Mian, A., Donnelly, C.J., Lloyd, D., Alderson, J., 2018. Predicting athlete ground  
343 reaction forces and moments from motion capture. *Med. Biol. Eng. Comput.*, 56(10), pp.1781-1792.
- 344 [16] Wouda, F.J., Giuberti, M., Bellusci, G., Maartens, E., Reenalda, J., van Beijnum, B.-J.F., Veltink,  
345 P.H., 2018. Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During  
346 Running Using Three Inertial Sensors. *Front. Physiol.* doi:10.3389/fphys.2018.00218.
- 347 [17] Wundersitz, D., Netto, K., Aisbett, B. and Gastin, P., 2013. Validity of an upper-body-mounted  
348 accelerometer to measure peak vertical and resultant force during running and change-of-direction  
349 tasks. *Sports Biomechanics*, 12, pp.403-12.
- 350 [18] Tran, J., Netto, K., Aisbett, B. and Gastin, P., 2010. Validation of accelerometer data for  
351 measuring impacts during jumping and landing tasks. In *Proceedings of the 28th International*  
352 *Conference on Biomechanics in Sports (2010)* (pp. 1-4). International Society of Biomechanics in  
353 Sports.
- 354 [19] Nedergaard, N.J., Verheul, J., Drust, B., Etchells, T., Lisboa, P.J., Robinson, M.A., Vanrenterghem,  
355 J., 2018. The feasibility of predicting ground reaction forces during running from a trunk  
356 accelerometry driven mass-spring-damper model. *PeerJ*, 6:e6105.
- 357 [20] Edwards, S., White, S., Humphreys, S., Robergs, R., O'Dwyer, N., 2019. Caution using data from  
358 triaxial accelerometers housed in player tracking units during running. *Journal of Sport Science*,  
359 37(7), pp.810-818.
- 360 [21] Nedergaard, N.J., Robinson, M.A., Eusterwiemann, E., Drust, B., Lisboa, P.J., Vanrenterghem, J.,  
361 2017. The Relationship Between Whole-Body External Loading and Body-Worn Accelerometry  
362 During Team-Sport Movements. *International Journal of Sports Physiology & Performance*, 12(1),  
363 pp.18-26.
- 364 [22] Verheul, J., Nedergaard, N.J., Pogson, M., Lisboa, P., Gregson, W., Vanrenterghem, J. and  
365 Robinson, M.A., 2019. Biomechanical loading during running: can a two mass-spring-damper model

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

368 [23] Gurchiek, R.D., McGinnis, R.S., Needle, A.R., McBride, J.M. and van Werkhoven, H., 2017. The  
369 use of a single inertial sensor to estimate 3-dimensional ground reaction force during accelerative  
370 running tasks. Journal of biomechanics, 61, pp.263-268.

371 [24] Hollville, E., Couturier, A., Guilhem, G. and Rabita, G., 2016. MinimaxX player load as an index of  
372 the center of mass displacement? A validation study. In ISBS-Conference Proceedings Archive (Vol.  
373 33, No. 1). Hong, Y. and Bartlett, R. eds., 2008. Routledge handbook of biomechanics and human  
374 movement science. Routledge.

375 [25] Kennedy, J., 2011. Particle swarm optimization. In Encyclopedia of machine learning (pp. 760-  
376 766). Springer US.

377 [26] Python Software Foundation, 2017. Python Language Reference, version 2.7. Available at  
378 <http://www.python.org>

379 [27] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,  
380 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M.,  
381 Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine  
382 Learning Research, 12, pp.2825-2830.

383 [28] Pyswarm, 2017. Available at <https://pythonhosted.org/pyswarm/>

384 [29] Neugebauer JM, Collins KH, Hawkins DA (2014) Ground Reaction Force Estimates from  
385 ActiGraph GT3X+ Hip Accelerations. PLoS ONE 9(6): e99023. doi:10.1371/journal.pone.0099023.

386 [30] Bobbert, M.F., Schamhardt, H.C. and Nigg, B.M., 1991. Calculation of vertical ground reaction  
387 force estimates during running from positional data. Journal of Biomechanics, 24(12), pp.1095-1105.

388

389 **Figure captions**

390

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

396

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

404

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

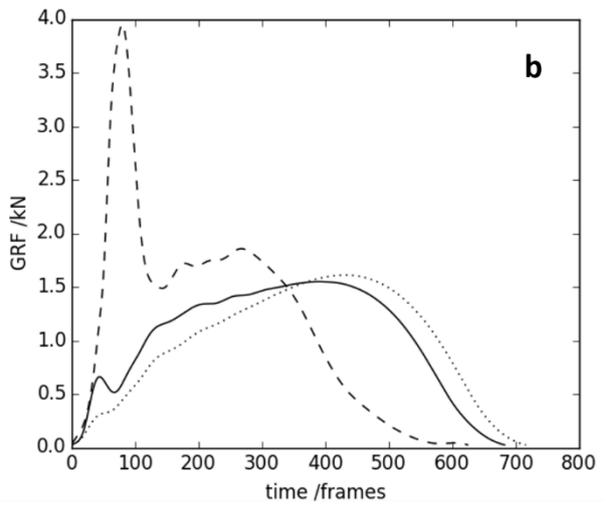
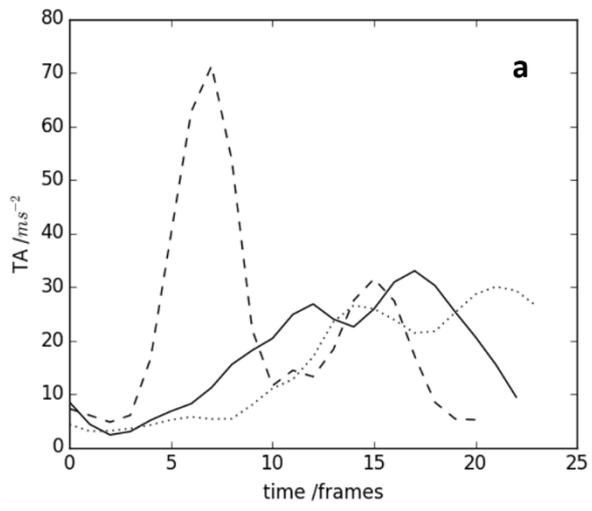
412

413 **Fig. 4.** Example time-series predictions for different tasks. Solid lines show predicted GRF and  
414 dashed lines show corresponding test data. (a)-(c) show good predictions for an example of steady  
415 running, acceleration and deceleration respectively; (d)-(f) show weaker predictions for similar  
416 activities (n.b. such predictions are less common, as shown in Fig. 3). The time scale corresponds to  
417 0.47s.

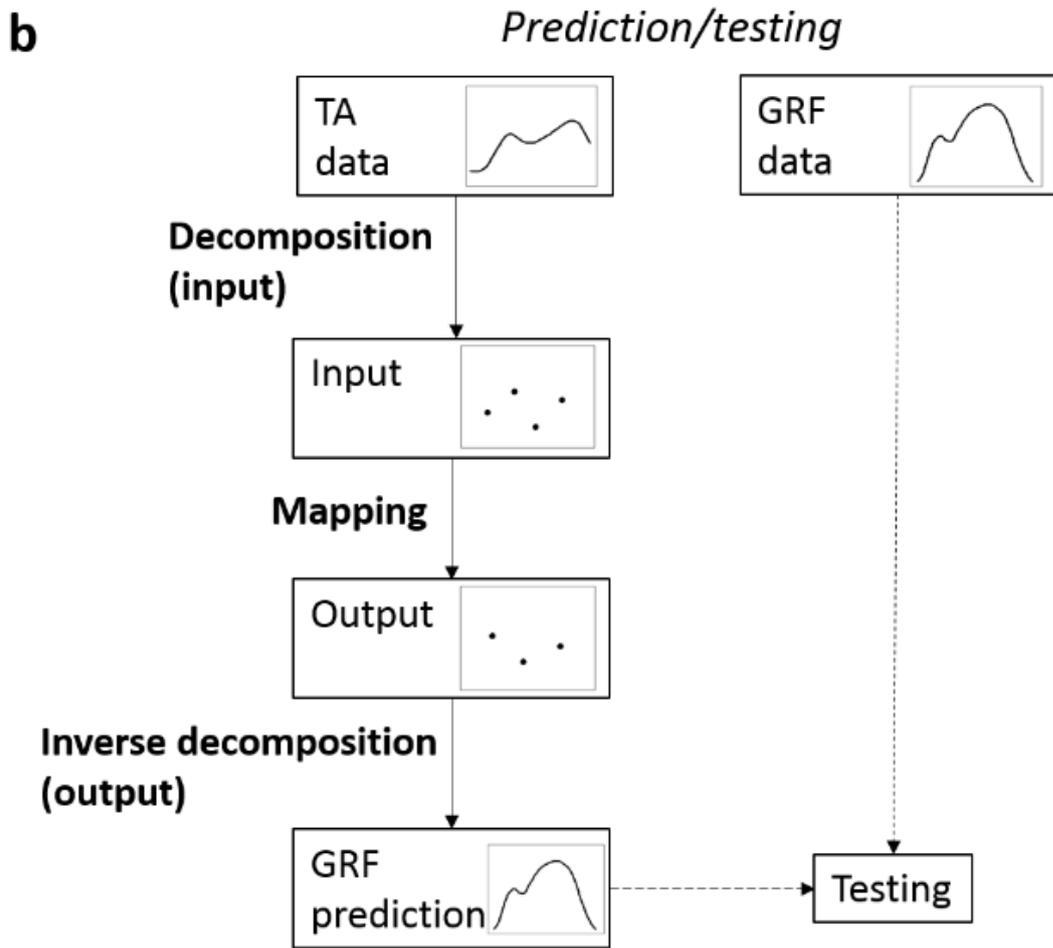
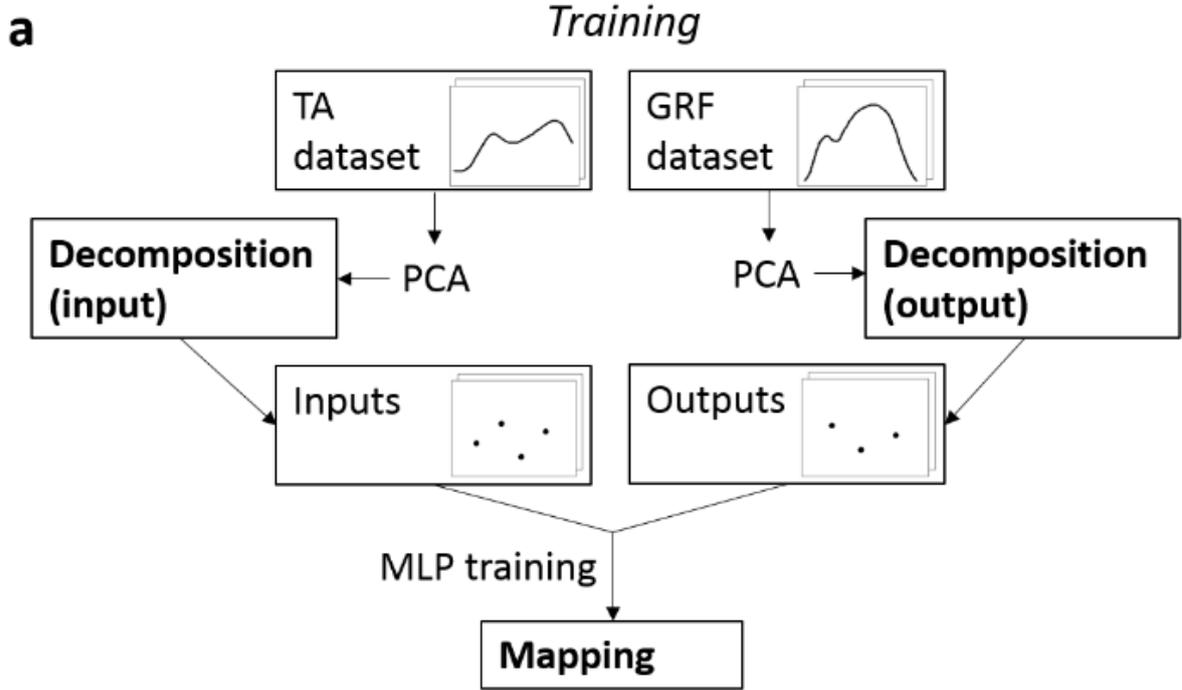
418

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

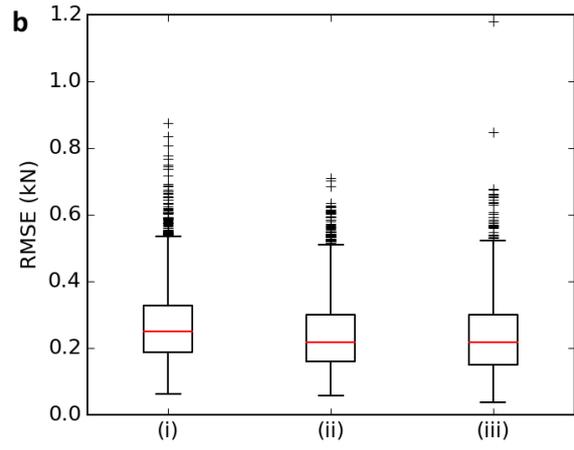
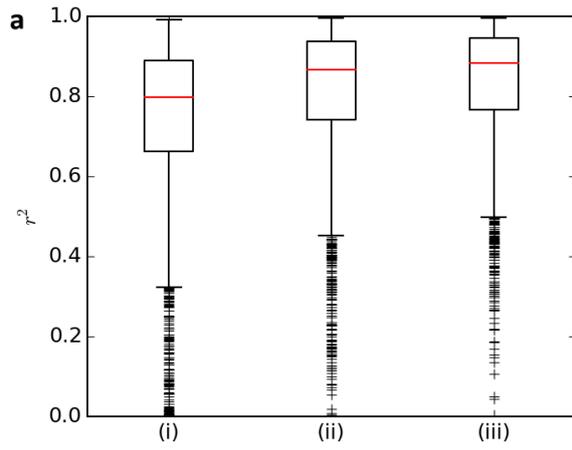
426

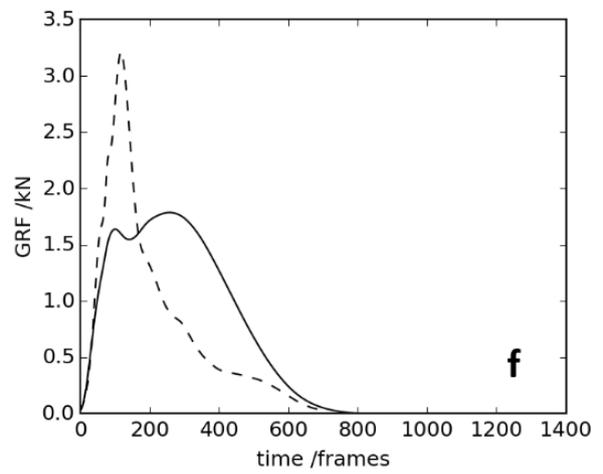
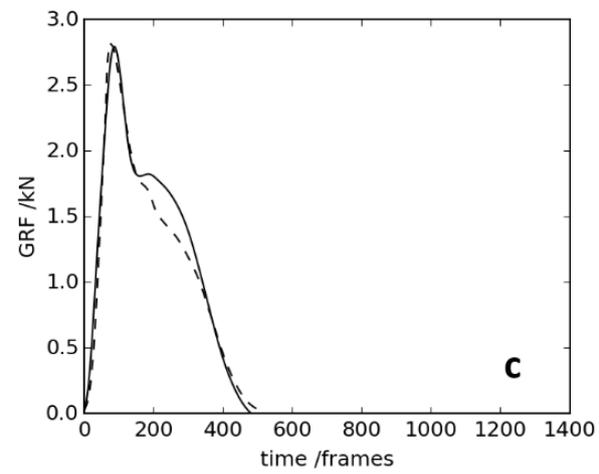
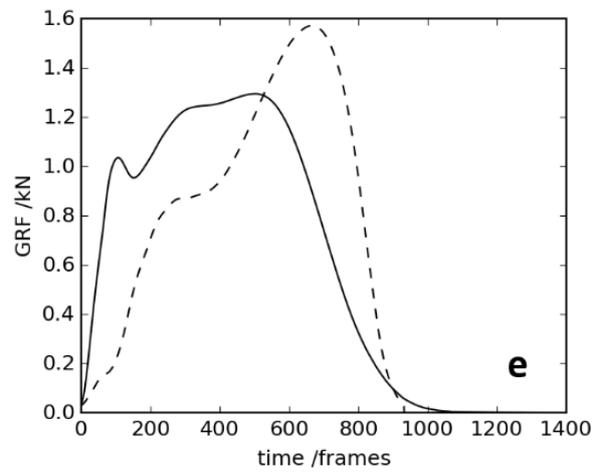
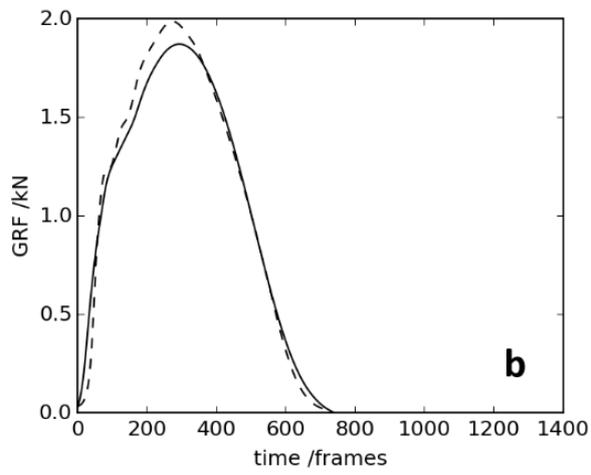
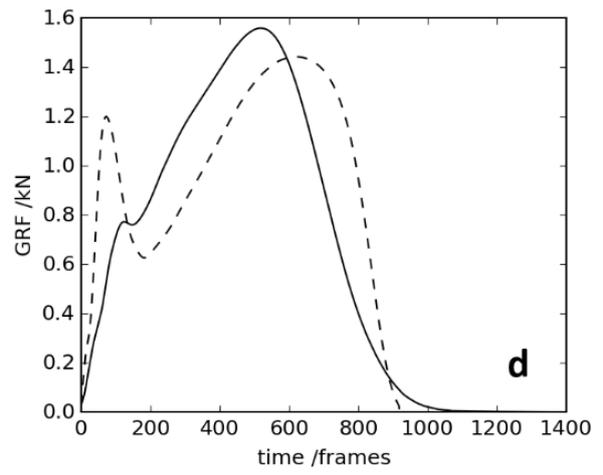
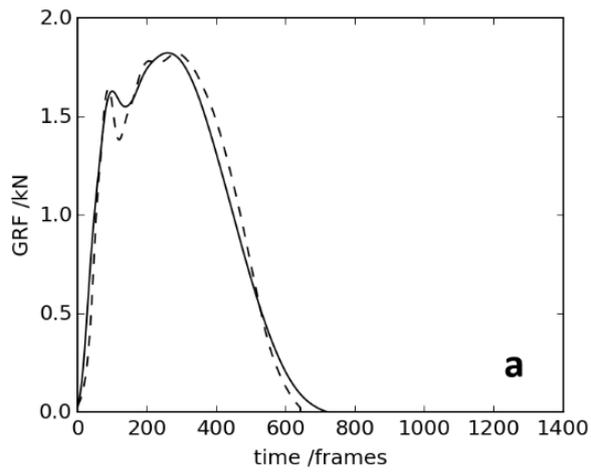


427

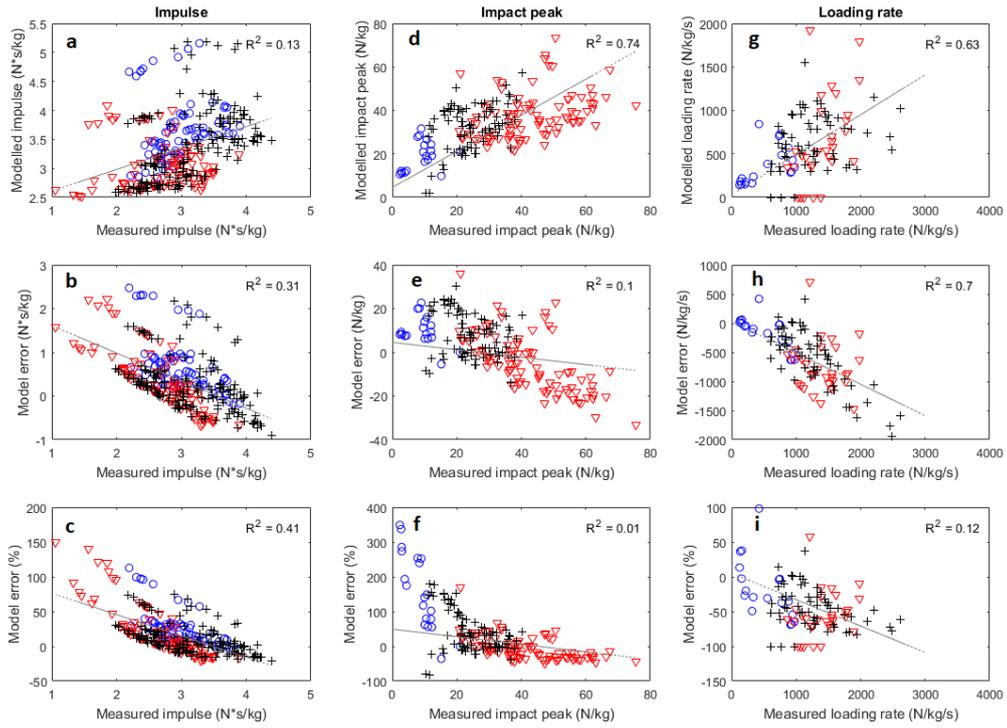


429





430



431