

Detailed response to comments reviewer 3

English language and style

- () Extensive editing of English language and style required
- () Moderate English changes required
- (x) English language and style are fine/minor spell check required
- () I don't feel qualified to judge about the English language and style

- *The manuscript has been checked and improved by an official English proofreader.*

Major comments

Rationale and background information would benefit from stronger, more appropriate, or additional relevant supporting literature. The text would also benefit from additional discussion of how this work fits in the context of relevant work and findings from other groups.

Some references appear to be presentations without a written/recorded component (12, 19, 27, 32, 35, 42) and one is an unpublished manuscript (33). Since readers cannot review these references, they do not add support to any of the statements for which they are cited.

- *Reference 12 is removed.*
- *The unpublished manuscript (reference 33) was referred to in first version of manuscript, as it was not yet known at what date this would be submitted. Is removed from ref list.*
- *Reference 35 is available on request.*
- *Our apologies for the systematically incomplete references 19, 27, 32 and 42. These are improved in the text.*
- *Reference 27 is a multiple page abstract. The complete proceedings are available online in the past conference section of the ISEK society.*
- *For reference 32 no digital copy of the abstract is found. A paper copy should be available in a most scientific libraries, as was established practice for these IEEE-EMBS conferences. If not acceptable references to this source could be redirected to the ISKE conf 1996 multi page abstract that is digitally available. An alternative of which minimally a paper copy should exist.: Baten, C.T.M.; Oosterhoff, P.; Luinge, H.; Veltink, P.H.; van Dieën, J.; Dolan, P.; Hermens, H.J. (1997). Quantitative assessment of mechanical low back load in the field: Validation in asymmetric lifting. Proc. 13th Int. Conf. International Ergonomics Association (29 June - 4 July 1997, Tampere, Finland), 4, 488-490.*

While referencing conference proceedings is acceptable in the absence of other references or as part of a larger list amongst article or book references, statements are weakened if they rely upon conference proceedings alone. Can any of these references (18, 20, 28, 29, 30) be replaced or supported with journal article or book references?

- *References 29 and 30 are journal articles.*
- *No, unfortunately no other peer reviewed papers, journal articles or book references are available that can replace these references.*

Seemingly relevant literature about lumbar loading estimation with ANNs is not discussed, for example, work by Arjmand and colleagues such as:

Gholipour, N. Arjmand. Artificial neural networks to predict 3D spinal posture in reaching and lifting activities; Applications in biomechanical models. Journal of Biomechanics. Volume 49, Issue 13, 2016, Pages 2946-2952.

Aghazadeh, N. Arjmand, A.M. Nasrabadi. Coupled artificial neural networks to estimate 3D whole-body posture, lumbosacral moments, and spinal loads during load-handling activities. Journal of Biomechanics. Volume 102, 2020, Article 109332.

- *These references are indeed interesting, added to discussion section (lines 424-427): “The feasibility of this approach is also supported by the results of others [44-45], who also used an ANN to predict static postures and net moments driven by static posture data (3D Euler angles) and EMG data and by the results of applying the same method to predict shoulder joint load exposure [46].”*

Please verify that references cited support claims. For example: Line 74 cites reference 23, Valevicius et al., 2018, but that review does not discuss ANN-based methods. Line 194 cites reference 39, Kristiansen et al., 2019, but the intervals and nomenclature stated in reference 39 do not match those of the statement for which it is cited.

- *There indeed went something wrong with reference numbering in some cases. All sources have been checked and corrected were needed.*

Additional methodological detail is needed. It is unclear why an LSM-based method is chosen as the “gold standard” against which the ANN-based method is evaluated. Please provide rationale for this decision.

- *ANN based method was trained with estimates from LSM method for 0 kg and 10 kg trials as target data and subsequently ANN based method estimates were tested against LSM method estimates for 6 kg trial. For the sessions during actual work the LSM method was only used in the role of providing target data for training in separate trials at the start of the session (with known load handling). In the rest of the session the trials were recorded during actual work in a condition of unknown load handling. Hence it was not possible to generate estimates with the LSM method for validation. All validation for these trials is on base quality arguments.*
- *To clarify the first part of paragraph 2.2 is rewritten.*

Insufficient detail is provided to repeat the LSM-based methods. The first reference cited (21) describes use of marker-based motion capture and forceplates for the LSM method while reference 27 has no print component cited. Please add a summary of the LSM method (including number and type of marker-based system and forceplates used) and update the text and references as needed so that the reader may refer to them for additional methodological detail.

- *Extensive details on the LSM method using marker-based motion capture and force plates applied in Kingma et al. are given in that paper. As well validated marker cluster set-up was used. Marker clusters deliver more precise rotational kinematics than traditional marker-based set-ups with markers on bony landmarks. IMMUs directly measure rotational kinematics even more accurate than marker clusters. Both deliver the same kinematics and are typically placed on the same locations on body segments. The LSM method implementation for that study in itself was validated separately before by validating top down (upper body plus knowledge on load box manually lifted) versus bottom up (lower body kinematics plus force plates) estimates. The implementation for this paper only deviates from the top-down method in driving kinematics (3D joint and body segment kinematics from IMMUs instead of from marker clusters being: 3D orientation, 3D angular velocity, 3D linear acceleration).*

Similarly, insufficient detail is provided to reproduce the ANN method.

- *Section 2.2 was rewritten to include more detail. See also response to comment below.*

Please describe what data were used by the neural network (e.g., which body segment and joint kinematics; were angles, angular velocity, segment velocities, and/or other information used?).

- *Added (lines 208-212): “The LSM was driven by kinematics derived from all 6 IMMUs: 3D orientation (represented by quaternions), 3D angular velocity and 3D linear acceleration (only used in controlled environment). The ANN was driven by EMG amplitudes plus 3D kinematics from only the 2 IMMU sensors positioned on sternum and sacrum: per segment the 3D angular velocity and 3D linear acceleration were signals used.”*
- *The rationale for not using orientation data was that: 1. information of inclination is already present in the accelerometer signal 2. In an industrial environment a substantial risk exists of*

having large errors in IMMU orientation estimates by disturbance of the natural earth magnetic field or permanent magnetization of the magnetometer in the IMMUs. (de Vries, W.; Veeger, D.; Baten, C.; van der Helm, F. „Magnetic distortion in motion labs, implications for validating internal magnetic sensors,” Gait & Posture, Bd. 29, pp. 535-541, 2009).

Please clarify whether EMG was collected bilaterally.

- *Added (lines 187-192): “Bipolar electrodes (Covidien Kendall™ H124SG Ag/AgCl electrodes; Medtronic, Minneapolis, MN, USA) with an interelectrode distance of 2 cm (heart to heart) were placed bilaterally on the longissimus thoracis muscles at L1 (± 3 cm horizontal from L1) and the iliocostalis lumborum muscle at L2-L3 (± 6.5 cm horizontal from L2-L3), along with a reference electrode placed at the processus spinosus of C7 (Roy, et al., 1995) (see Figure 1).”*

Please summarize the optimal settings (line 126) for the ANN-based method.

- *Added to method section (lines 125-157): “In phase 2 the ANN-based method performance was tested against the LSM-based method by comparing estimated moments by both methods in bending, flexion, and rotation movements while hand holding a 6 kg load (question 1). For this, the ANN was first trained with net moments directly estimated with the LSM-based method from similar tasks handling no load (‘0 kg task’) and holding a load of 10 kg trunk. Based on previous experience [unpublished sensitivity study] a simple feed-forward ANN was used with 1 hidden layer of 31 elements and sigmoid transfer functions with a fixed training criterium: ‘sum of RMSE Mx, RMSE My and RMSE Mz <10 Nm’ This required about 10 s of computer time on a regular desktop PC for an typical input vector with about 6000 elements. The ANN was driven by EMG amplitudes from all 4 locations plus 3D kinematics from only the 2 IMMU sensors positioned on sternum and sacrum: per segment the 3D angular velocity and 3D linear acceleration signals were used. The rationale for not using the available 3D orientation data was, that: 1. information of inclination is already present in the 3D accelerometer signal 2. In an industrial environment a substantial risk exists of having large errors in IMMU orientation estimates by disturbance of the natural earth magnetic field or permanent magnetization of the magnetometer in the IMMUs [41]. In preparation for phase 3 the ANN-based estimator was trained for each subject separately using all their respective bending, flexion and rotation tasks for all three weights (0, 6 and 10 kg).”*

Please include the questionnaire referred to in line 137.

- *Added (lines 165-167): “This physical workload questionnaire of Peereboom & Lange [34] contained the following questions: What are tasks you perform on daily basis at work? How often are these tasks done on an average day? How heavy do you find these tasks?”*

Minor comments

Please comment on why subject 8 displayed such poor correlation between the ANN and LSM-based methods. Was the data from this subject used in subsequent analyses?

- *Added to the discussion section (lines 427-432): “The results of the base quality differ per subject due to the individual character of the results. The calibration differs from one subject to another, which can be due to differences between the three calibration sets with three different weights or due to an unidentified event. This resulted in nonperfect calibration for subject 8, with overall medium to small correlations.”*

Consider providing the correlation information for the additional axes that were left out of Table 1 (such as the “anterior-posterior axis” referenced in line 262) in a supplementary table.

- *The anterior-posterior axis referred to in line 262 is provided in Table 1. The x, y and z represent the moment around these axes as mentioned in the data analysis and in the table index (f.e. lines 301-307): “Correlation is represented by the Pearson correlation coefficient (r) and determination coefficient (r²) and are shown only for the axis of movement in each task (i.e. in the mediolateral axis (y) for the trunk bending tasks, in the anterior-posterior axis (x) for the lateroflexion tasks (y) and the longitudinal axis (z) for the rotation tasks respectively).”*

- The table with all axes per movement are added to the appendix, table A.1.1.

Consider also providing in the supplementary material example figures with the ANN predictions compared to the LSM-method similar to comparisons shown in Figures 2 and 3.

- Added to the appendices; Figure A.2.1 and Figure A.3.1.

Line 266 states "...typical net moment ranges from 150 to 220 Nm..." but when compared to the data presented in Table 4, this appears to be the range of the peak net moment, not the net moment range. Please clarify.

- This must be peak net moment.
- Adjusted (lines 317-319): "Table 3 compares the RMSE values between the ANN-based method and the LSM-based method. These results indicate a mean estimation errors of 9.25 ± 6.01 Nm, relative to the typical peak net moment ranges from 150 to 220 Nm (see Table 5)."

Please add a legend to the plots in Figures 2 and 3.

- Added.

In Tables 5-8, please clarify what test the reported p-values refer to and which quantities were compared.

- Adjusted throughout the document:
 - Table 5. Light vs heavy tasks. Shown are the absolute [Nm] and relative [%] differences between the light tasks and the heavy tasks through mean difference (MD: heavy – light task) and its standard deviation (SDD) plus 95% confidence interval (CI) of these differences for the moment magnitude ($\|M\|$) with p-value of paired t-test.
 - Table 6. Static vs dynamic tasks. Shown are the absolute [Nm] and relative [%] differences between the static tasks and the dynamic tasks through mean difference (MD: dynamic– static task) and its standard deviation (SDD) plus 95% confidence interval (CI) of these differences for the moment magnitude ($\|M\|$) with p-value of paired t-test.
 - Table 7. Direction of the net moment around the axis of the light vs heavy tasks. Shown are the absolute [Nm] and relative [%] differences between the light tasks and the heavy tasks through mean (MD: heavy – light task), standard deviation (SDD), and 95% confidence interval (CI) of these differences for the anterior-posterior (Mx) with lateroflexion to the left (positive) and right (negative), mediolateral (My) with flexion (positive) and extension (negative), longitudinal (Mz) with rotation to the left (positive) and right (negative) axes separately and for the moment magnitude ($\|M\|$) with p-value of paired t-test.
 - Table 8. Direction of the net moment around the axis of the static vs dynamic tasks. Shown are the absolute [Nm] and relative [%] differences between the static tasks and the dynamic tasks through mean difference (MD: static - dynamic) and its standard deviation (SDD) plus 95% confidence interval (CI). These are shown for the net moments around the each of the three axes separately (in the right-handed axes frame): for the moments around the anterior posterior axis (Mx), with the net moment positive for direction of rotation to the left; for the moments around the mediolateral axis (My), with the net moment positive in the forward flexion direction; for the moments around the longitudinal axis (Mz), with the net moment positive in the direction of rotation to the left and for the moment magnitude ($\|M\|$) with p-value of paired t-test.

Line 354: Reporting the average instead of the range or maximum difference is a bit misleading. Consider rephrasing this statement.

- The range of the mean differences was added (lines 404-406): "In addition, significant differences in the mean moment ($\|M\|$, M_y , and M_z) were observed in the positive direction (-

47.1%≤MD≤302.7%, p≤0.042), but were not significant in the negative direction (-54.0%≤MD≤-276.5%, p≥0.326)."

The statements in lines 364-367 advocate analyzing moments of different signs separately, which would eliminate the differences observed between the static and dynamic tasks using the ANN-based method. This change in the analysis would contradict the finding that variance is higher during dynamic tasks compared to static tasks. Please clarify this point.

- *This is caused by the asymmetrical character or performance of task(s) mentioned in the discussion section (lines 437-443): "However, between static and dynamics tasks, the variance differences of the net moment were not observed. This is because of the differences due to the direction of the moment around the axis. This indicates that it is important to analyze and interpret net moments with different sign separately and thereby acknowledge the (a)symmetrical character of the net moment exertion when analyzing the lumbar load."*

Correction of various typographical errors and changes in word choice would improve the clarity of the manuscript. To avoid confusion, consider using "height of the subject" instead of "length of the subject" (e.g., lines 64, 241).

- *Adjusted into "height of the subject".*

Use the same wording to describe the job types of the participants in the text and in Table 3.

- *Corrected throughout the document.*

Standardize results terminology throughout the results section text and table captions. For example, Line 325 "mean net moment" compared to the Table 7 caption in line 333 "moment magnitude".

- *Adjusted throughout the manuscript.*

Consider in line 341 separating the positive and negative mean differences reported or reporting the magnitude of the mean difference in the text.

- *Added (lines 404-406): "In addition, significant differences in the mean moment ($|M|$, M_y , and M_z) were observed in the positive direction ($MD \leq 302.7 \pm 399.7\%$, $p \leq 0.042$), but not in the negative direction ($MD \leq -59.0 \pm 189.5$, $p \geq 0.326$)."*

Please address various typographical and grammatical errors. For example:

Line 245: "...moment curves form ANN based method en the LSM based method."

- *Adjusted into: "The base quality was studied by examining the correlation between the corresponding moment curves from ANN based method en the LSM based method. Table 1 shows the correlations."*

Caption for Table 1, line 253: "...for the lateroflexion tasks (y)..."

- *Adjusted: Correlation is represented by the Pearson correlation coefficient (r) and determination coefficient (r^2) and are shown only for the axis of movement in each task (i.e. in the mediolateral axis (y) for the trunk bending tasks, in the anterior-posterior axis (x) for the lateroflexion tasks, and the longitudinal axis (z) for the rotation tasks respectively).*

Caption for Table 2, line 270: "Values are shown only for the axis of movement in each task..."

Caption for Table 2, line 272: "...for the lateroflexion axis (y)..."

- *Adjusted: "Values are shown for the three movements trunk bending, lateroflexion and rotation with the anterior-posterior axis (x), mediolateral axis (y) and longitudinal axis (z)."*

Lines 415-416 there are two "e.g.," in a row with no closing parenthesis.

- *Adjusted: “As suggested in these papers, this information can be provided to the user (e.g., using a traffic light model based on EU or NIOSH ergonomic norms, to indicate areas of overload risk).”*

Lines 425-427 is an incomplete sentence.

- *Adjusted: “Adding load exposure pattern analysis to the current analysis of only the amount of load exposure [43] can be useful by means of generating a 2D graph of the joint distribution of intensity and of the net moment data dynamically. This exposure variance analysis can be used to link the results to ergonomic guidelines and provide the user with feedback.”*

Please clearly show how the results of this study suggest that ANN-based methods would be useful in the other applications and fields listed in lines 428-434.

- *Adjusted: “The results of this study suggest that not only can the ANN-based method be used in monitoring lumbar back load exposure in physically demanding jobs, but also that it may have potential to be used with office workers. For office workers the lumbar load during sedentary behavior could be monitored and investigated aiming to prevent health problems and physical discomfort related to static working postures. In the fields of rehabilitation medicine the ANN-based method could be used in back pain rehabilitation and prevention, rehabilitation involving improvement of knee loading behavior e.g. ACL surgery recovery, arthrosis patients walking training. And with this method injuries could be predicted or prevented in sports applications and the performance could be improved.”*

Please clarify what is meant by “in-vivo setup” in line 436.

- *With in-vivo was meant not completely uncontrolled conditions, but during predefined tasks. This term is not discussed through the manuscript and has been removed.*