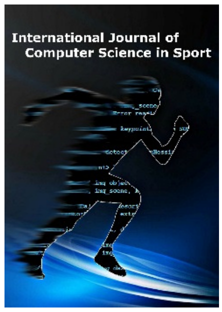
International Journal of Computer Science in Sport



Volume 16, Issue 1, 2017

Journal homepage: http://iacss.org/index.php?id=30

DE GRUYTER OPEN

DOI: 10.1515/ijcss-2017-0001

Issues in Using Self-Organizing Maps in Human Movement and Sport Science

Serrien, B.[1], Goossens, M.[2,3], Baeyens, J-P.[1,2,3]

[1]Vrije Universiteit Brussel, faculty of Physical Education & Physiotherapy, dept. Biomechanics (Brussels, Belgium)

[2]Universiteit Antwerpen, faculty of Engineering, dept. ICT & Electronics (Antwerp, Belgium)

[3]Thim Van Der Laan University College Physiotherapy (Landquart, Switzerland)

Abstract

Self-Organizing Maps (SOMs) are steadily more integrated as data-analysis tools in human movement and sport science. One of the issues limiting researchers' confidence in their applications and conclusions concerns the (arbitrary) selection of training parameters, their effect on the quality of the SOM and the sensitivity of any subsequent analyses. In this paper, we demonstrate how quality and sensitivity may be examined to increase the validity of SOM-based data-analysis. For this purpose, we use two related data sets where the research question concerns coordination variability in a volleyball spike. SOMs are an attractive tool for analysing this problem because of their ability to reduce the high-dimensional time series to a two-dimensional problem while preserving the topological, non-linear relations in the original data. In a first step, we systematically search the SOM parameter space for a set of options that produces significantly lower continuity, accuracy and combined map errors and we discuss the sensitivity of SOM-based analyses of coordination variability to changes in training parameters. In a second step, we further investigate the effect of using different numbers of trials and variables on the SOM quality and sensitivity. These sensitivity analyses are able to validate the conclusions from statistical tests. Using this type of analysis can guide researchers to select SOM parameters that optimally represent their data and to examine how they affect the subsequent analyses. This may also enforce confidence in any conclusions that are drawn from studies using SOMs and enhance their integration in human movement and sport science.

KEYWORDS: SELF-ORGANIZING MAP, ARTIFICIAL NEURAL NETWORK, COORDINATION, VARIABILITY, VOLLEYBALL

Introduction

In recent years, Self-Organizing Maps (SOMs) are slowly gaining momentum as data analysis tools in the human movement and sport sciences. These SOMs are a class of artificial neural networks that can be used in both exploratory research as in modelling. In human movement and sport science, mostly the former has been the framework where SOMs are used due to their ability to visualize high-dimensional datasets and because of their ability to cluster data into relevant topologically similar units. In human movement science, we are often confronted with time series of a large number of variables. SOMs are able to reduce this complexity to a lower-dimensional problem while preserving the topological features of the original data which are often of a non-linear nature. Applications of SOMs have been reviewed in clinical (Schöllhorn, 2004) and sport science (Schöllhorn, Chow, Glazier, & Button, 2014) contexts. Predictions of artificial neural networks (ANN) becoming common-place in human movement and sport science (Bartlett, 2006; Lapham & Bartlett, 1995) are still far away however. One of the issues causing this delay in using SOMs and other ANNs, highlighted by Button, Wheat, & Lamb (2014), are the training and topological parameters of the SOMs that researchers have to choose from. At present, no standardized options or rationale exists for their choice and it is unknown to what extent they have an influence on the quality of the map and on the sensitivity of any subsequent analyses. Button et al. (2014) argued that the SOM architecture and training parameters should be considered with respect to the research setting and should be data-driven. This is possible for some, but not for all SOM parameters and in the end, this choice will stay the responsibility of the researchers. The Appendix provides a brief overview of the SOM algorithm and its various options.

The present study focusses on this issue with SOMs in their application for studying coordination variability as demonstrated in previous studies (Bartlett, Lamb, O'Donovan, & Kennedy, 2014; Lamb, Bartlett, Lindinger, & Kennedy, 2014; Lamb, Bartlett, & Robins, 2011; Serrien, Ooijen, Goossens, & Baeyens, 2016b; van Diest et al., 2015). In these studies, the SOM was used as a tool to construct a time-continuous order parameter from multi-dimensional time series of kine(ma)tics and/or EMG data, for which a holistic measure of coordination variability could be calculated (see the previous studies for visualizations of this method). This methodology is in accord with the dynamical systems perspective on coordination, in that it assumes that the motor control system operates in a low-dimensional space, controlling only the collective system (coordinative structure), not all degrees-of-freedom separately. The movement system has abundant degrees-of-freedom to solve a particular task, resulting in a seemingly infinite set of solutions. Variability in one component of the coordinative structure is absorbed through covariation in the other components to keep the task outcome stable. Therefore, variability at the holistic level of coordinative structures, which can be studied with SOMs, is more informative than at the component level.

In this paper, we performed several SOM-analyses of two related datasets of coordination variability in a volleyball spike motion. We systematically altered the options of the SOM parameters to see how they affect the quality of the SOM and the subsequent analyses. The quality of the map can be characterized by three variables, namely the topographical error (TE), the quantization error (QE) and the combined error (CE). The TE represents the percentage of input vectors for which the best-matching unit and second-best-matching unit are not neighbours (a measure of the continuity of the map) and the QE represents the average Euclidean distance between a normalized input vector and its (trained) best-matching unit's weight vector (a measure of the accuracy of the map) (Vesanto, Himberg, Alhoniemi, & Parhankangas, 2000). The CE includes both continuity and accuracy in its calculation. The CE is the average Euclidean distance between an input vector and its second best-matching unit, passing first through the best-matching unit and then through the shortest path of neighbouring units towards the second best-matching one (Kaski & Lagus, 1996). In the studies that were mentioned above, SOMs were used as a data-reduction technique while attempting to preserve the topological non-linear features of the original data. In this regard, the quality measures are important variables that should be minimized. The validity of any subsequent analyses should also be examined more closely to see whether the conclusions remain stable over a change in SOM parameters. Performing these quality and sensitivity analyses might increase researchers' confidence in using SOMs for data-analysis.

METHODS

Datasets

The first dataset consisted of kinematics of thirty-seven volleyball players performing a volleyball spike (top level: 8 male, 10 female; junior: 8 male, 11 female). These data have been published before (Serrien, Ooijen, Goossens, & Baeyens, 2016a; Serrien, Ooijen, et al., 2016b) and more details can be found there. The kinematics are time series of joint/segment angles and angular velocities of pelvis (sagittal plane tilt, lateral tilt, rotation), trunk (sagittal plane tilt, lateral tilt, rotation), shoulder (in/external rotation, ab/adduction, horizontal ab/adduction) and elbow (flexion/extension) of the spike arm (D = 20 variables), centralized around the point of ball impact (400 ms before until 80 ms after, 121 time samples). These data were recorded with a 3D VICON motion capture system at 250 Hz. These variables' time series composed the input vectors that were fed to the SOM algorithm: $\boldsymbol{z_j} = (\theta_{j,1} \ \cdots \ \theta_{j,10} \ \omega_{j,1} \ \cdots \ \omega_{j,10}), j \in \{1 \ \ldots \ J\}$; ten angles ($\theta$) and the corresponding angular velocities (ω). Every trial is thus composed of 121 input vectors. The task consisted of performing ten spike motions (resulting in J = 1210 input vectors per subject) to a target at the back end of the field. SOMs were used to calculate the coordination variability over these ten trials. The research question concerned the difference in coordination variability between top level and junior players and male and female players (between subjects design). The previously published study (Serrien, Ooijen, et al., 2016b) found no significant interaction effect between gender and expertise on coordination variability, so for the present study, we concentrated only on the main effects (they found a significant main effect of gender, but not of expertise). The second dataset is related to the first one (not published before). It contains the data of the junior players of dataset 1 (year 1), and of these same players one year later (year 2). The same twenty kinematic variables were measured over ten trials of performing the same task. The research question for dataset 2 concerns the evolution of coordination variability of junior players when they develop expertise (within subjects design).

SOM analyses

All analyses were performed in Matlab R2015a and the freely available SOM Toolbox (Vesanto et al., 2000, http://www.cis.hut.fi/somtoolbox/). In line with previous studies in human movement science (Bartlett et al., 2014; Lamb et al., 2014, 2011; Lamb, Bartlett, Robins, & Kennedy, 2008; Serrien, Clijsen, Anders, Goossens, & Baeyens, 2016), we performed the SOM analyses at the individual level, i.e. training separate SOMs for every subject with all trials as input. Total trajectory variability was then calculated based on the method outlined in Lamb et al. (2014), as the measure of coordination variability. This method uses the best-matching unit trajectories of all trials to calculate the total average distance between these trials (a single number describing the variability of a subject's trials).

SOM quality and sensitivity

The analyses of both datasets consisted of two steps. The first step consisted of systematically adjusting the SOM parameters and investigating their effect on the quality of the SOM (CE, TE and QE) and on the effect size (ES) related to the research questions. For dataset 1, we calculated Cohen's d for the main effects of gender (ES gender) and expertise (ES expertise) on coordination variability and for dataset 2, we calculated the paired samples Cohen's d for the difference between year one and two on coordination variability (ES year). These dependent variables (CE, TE, QE and ESs) were used to investigate the SOM parameter space and search for a set of options that best reduces the errors and results in consistent effect sizes. To compare the different options of the SOM parameters, the dependent variables were tested with a non-parametric Friedman test and follow-up Wilcoxon tests for parameters with more than two options (α = 0.05 with post-hoc Bonferroni corrections). To guide the decision-making process, we prioritized the options that produced significantly lower CE because it represents both map continuity and accuracy. Following the arguments of Tan & George (2004) and given that the present application concerns time-series data, the continuity of the map (neighbouring units representing neighbouring time samples) is more important than an exact representation of the input vectors; we therefore used TE as the second criterion variable. The QE was used as tertiary variable to determine the best-set-of-options. The effect of SOM parameters on the effect sizes was used to determine the sensitivity of the subsequent analyses. Table 1 provides an overview of the parameters and their several options that were used, this resulted in a total of 864 possible options that were simulated.

Table 1: SOM parameters and their options (see Appendix for a short explanation)

SOM parameters	Options
Lattice	Rectangular, Hexagonal
Initialization	Linear, Random [a]
Map Size	Small, Medium, Big
Training Length	Short, Normal, Long
Neighborhood function	Gaussian, cut-of Gaussian, bubble, Epanechikov
Training type	Sequential, Batch

[a] We ran five different simulations for the random weight vector initialization.

In the second step, we used only this best-set-of-options in a further exploration of the map quality and sensitivity. We wanted to explore the effect of dropping variables and trials from the datasets. It is possible that some components are associated with larger errors than others, thereby decreasing map quality and possibly affecting the analyses. Time series of variables with a higher variability, rapidly changing values and multiple peaks might be especially problematic for SOM quality. The decision to include/exclude variables should be made a priori and not based on improving map quality; it is however important to examine the effect of each component or sets of components on the sensitivity of SOM-based analyses. Using fewer trials to train the SOM and calculate variability might increase map quality because the units have to organize themselves between fewer input vectors, especially in the case for players with high variability between trials. However, this may come at a cost of reducing the stability of any subsequent analyses and should be carefully weighted in the decision-making process. Using a maximum of trials is advisable, but performing a sensitivity analysis with fewer trials can be performed to increase the confidence in the conclusions. To investigate these effects, we ran 162 new simulations for every subject per dataset where we systematically dropped one or more components from the dataset and used between ten and five trials for each set of components (for each consecutive lower number of trials, we dropped one trial at random from the original ten trials). The interpretation of the results of step two was done only qualitatively.

Results

Step one: SOM parameter space

SOM quality

The results on the SOM quality measures for the simulations for dataset 1 and 2 are shown in Figures 1 and 2 respectively. For dataset 1, we found only significant differences in CE for map size (large < medium < small) and training length (long < normal < short). For the TE of dataset 1, we found significant differences in lattice (hexagonal < rectangular), training type (sequential < batch), map size (small, large < medium) and training length (long, normal < short). While there was no significant difference in TE between linear and random initialization, the random initialization was skewed towards higher errors and produced many outliers. For the QE of dataset 1, we found significant differences for initialization (linear < random), training type (batch < sequential), map size (small, large < medium) and training length (long, normal < short). For dataset 2, we found completely identical results than for dataset 1. The combined error is best reduced by proportionally increasing map size and training length relative to the default setting. Based on the topographical error, the lattice and training type can best be chosen to be hexagonal and sequential. Based on the results of the QE-analyses, this set can be extended with using a linear weight vector initialization. The neighborhood functions had no effect on any of the three quality measures. We therefore considered the Gaussian kernel as the more natural choice (default option of the SOM Toolbox) for the analyses in step two. This parameter selection was performed in a data-driven analysis with a specific priority of the three different errors. Other priorities might select other parameters, and when computation time is an important issue, the selection of parameters becomes more difficult (see Appendix).

Sensitivity

The effect sizes related to the research questions of both datasets for all simulations are shown in Figure 3. The ES for the gender factor was significantly affected by SOM training length (short, normal < long). However, more importantly, all simulations demonstrated that this ES was in the same direction (male players exhibiting a smaller coordination variability than female players). In contrast, the values of the ES for the expertise factor were both significantly affected by the parameters and demonstrated a strong directional instability. Only for the biggest map size did we observe a consistent (negative) ES expertise. The major part of the simulations showed a smaller coordination variability for the expert players (negative values), but a non-negligible part of the combinations (10.23 %) showed the reverse effect (positive values). The ES for the within subjects factor year was significantly affected by training type (sequential > batch), map size (small > medium > large) and training length (short > normal > long). Again here as well, more importantly is that all simulations demonstrated that this ES was in the same direction (smaller coordination variability in year two). The (in)consistency in ES is an important validity check of the true (false) difference in a statistical test of coordination variability between groups/conditions.

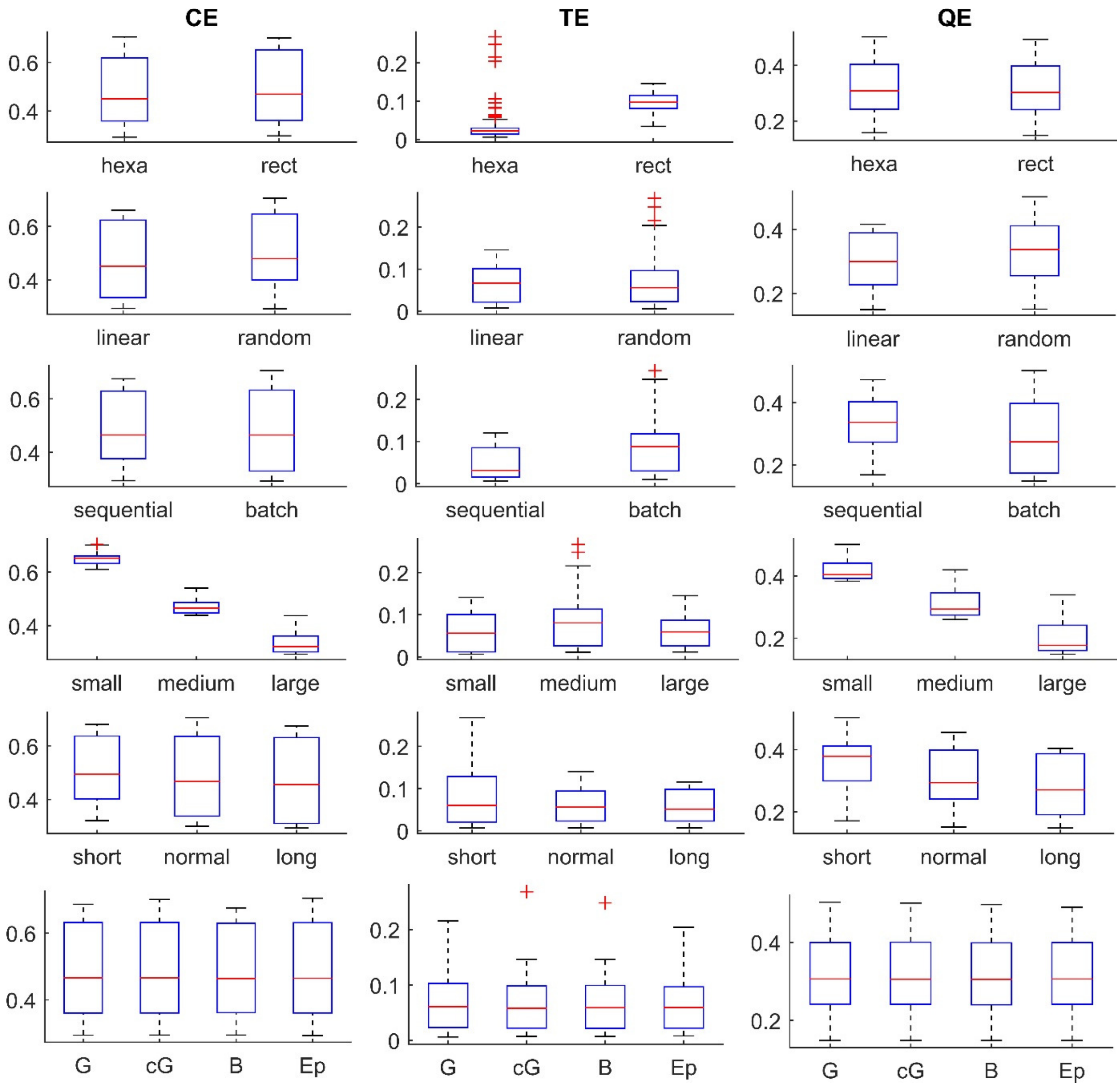


Figure 1: Dataset 1. Boxplots representing the CE (left), TE (middle) and QE (right) of the several simulations split between the options of the six SOM parameters: lattice (row 1), initialization (row 2, random is concatenated over all 5 seeds), training algorithm (row 3), map size (row 4), training length (row 5) and neighborhood function (row 6: G = Gaussian, cG = cut-off Gaussian, B = bubble, Ep = Epanechikov).

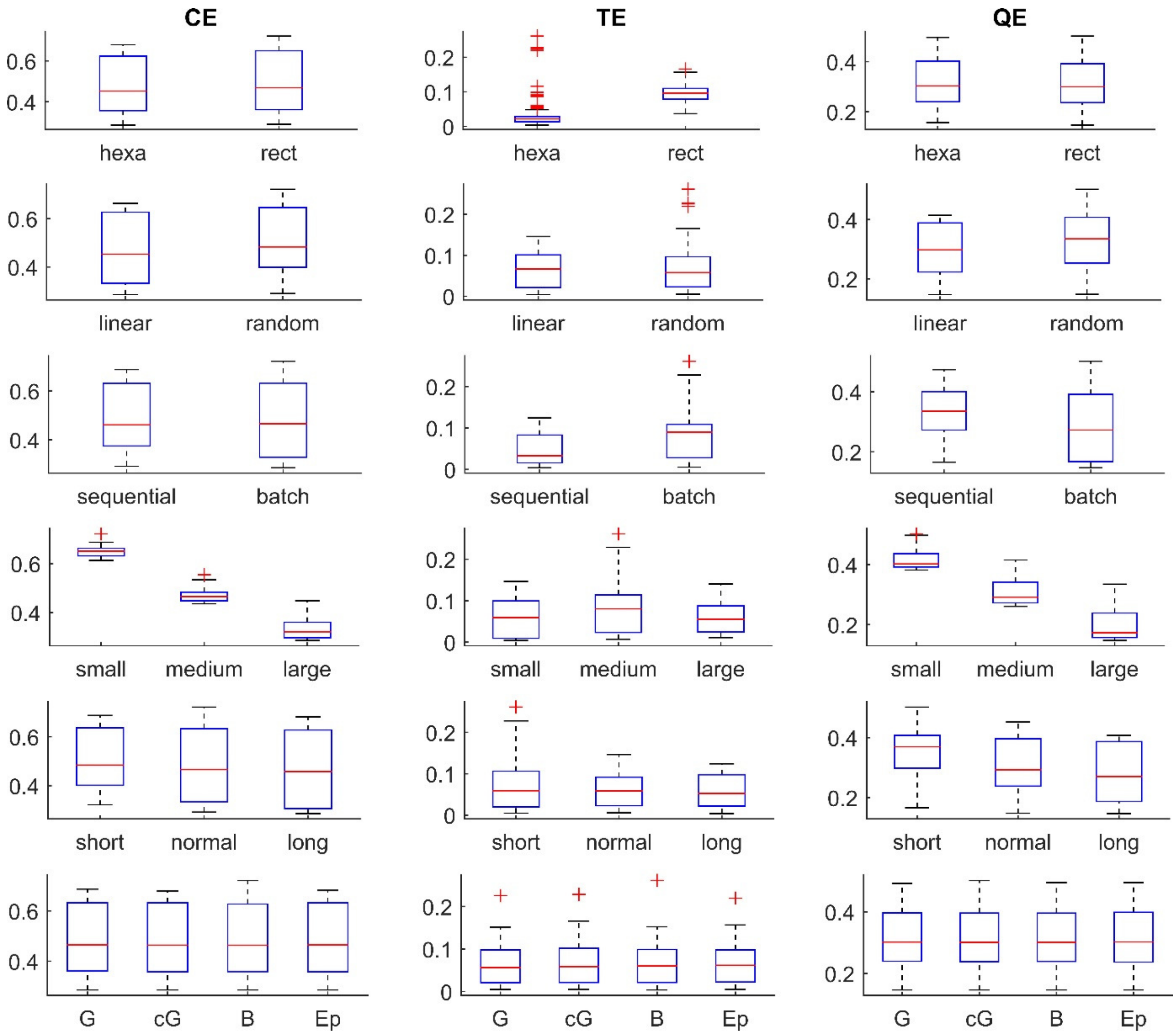


Figure 2: Dataset 2. Boxplots representing the CE (left), TE (middle) and QE (right) of the several simulations split between the options of the six SOM parameters: lattice (row 1), initialization (row 2, random is concatenated over all 5 seeds), training algorithm (row 3), map size (row 4), training length (row 5) and neighborhood function (row 6: G = Gaussian, cG = cut-off Gaussian, B = bubble, Ep = Epanechikov).

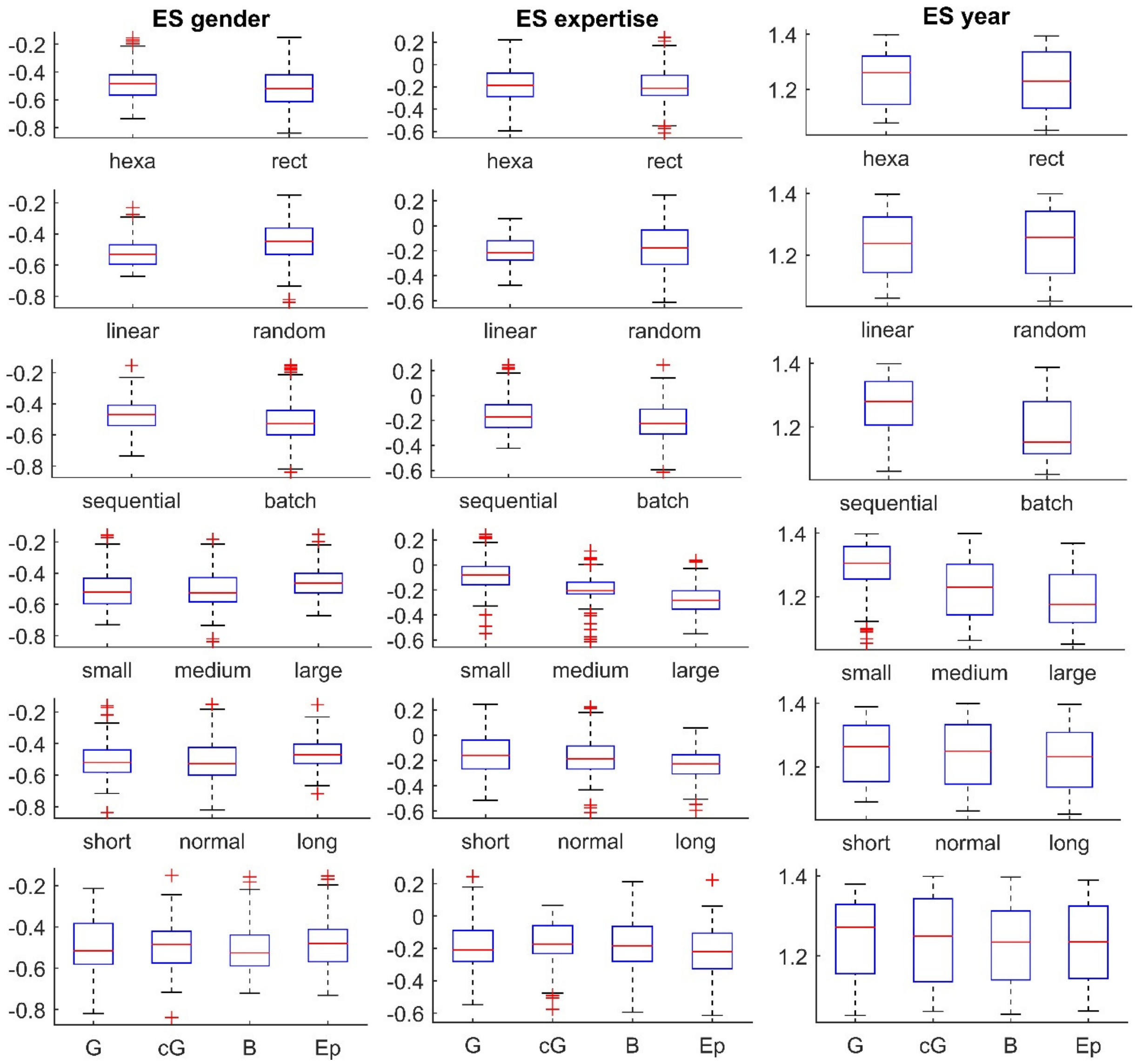


Figure 3: Boxplots representing the ES gender (left) and ES expertise (middle) for dataset 1 and the ES year for dataset 2 (right) of the several simulations split between the options of the six SOM parameters. lattice (row 1), initialization (row 2, random is concatenated over all 5 seeds), training algorithm (row 3), map size (row 4), training length (row 5) and neighborhood function (row 6: G = Gaussian, cG = cut-off Gaussian, B = bubble, Ep = Epanechikov). ES gender: negative values indicate greater variability for female players. ES expertise: negative values indicate greater variability for youth players. ES year: positive values indicate lower variability in the second year.

Step two: number of components and trials.

SOM quality

The results of training the SOM on a fewer number of trials (fewer input vectors) and variables (fewer vector components) are shown in Figures 4 and 5 for datasets 1 and 2 respectively. Every cell in the plots shows the median value of the errors across all players. Note that all rows can be compared to each other, but not all columns, only those columns with the same number of components. The numbers referring to the different variables are as follows: (1-3) pelvis sagital plane tilt, lateral tilt, rotation; (4-6) trunk sagital plane tilt, lateral tilt, rotation; (7-9) shoulder in/external rotation, ab/adduction, horizontal ab/adduction; (10) elbow flexion/extension. For both datasets, it is clearly visible that using fewer trials results in lower

CE and QE, while the TE is not visually affected by reducing the number of trials. The larger CE and QE for the six right most columns is a trivial result, because the dimensionality is significantly lower in these simulations (CE and QE are distance measures). The SOM errors associated with the angular velocities were larger compared to the angles, probably because they showed more within subject variability. It is also clear that the pelvis variables were associated with larger errors than the trunk and shoulder variables. For the TE, we see only an effect when the angular velocity variables were dropped from the dataset. In contrast to what we expected, TE actually increased without angular velocity. Like stated in the methods section, these results should not be interpreted as to drop the angular velocities entirely from the dataset, but as a cue to look closely at their effect on the sensitivity of the effect sizes.

Sensitivity

In step one, we concluded that ES gender (male < female) and ES year (year 2 < year 1) were likely to be true effects; both statistically significant and robust to parameter selection. ES expertise was found to be directional unstable, representing no true effect. In Figure 6, we can see their sensitivity to a change in number of trials and variables used to train the SOM and calculate the effects (contrary to Figures 4 and 5, all rows and columns can be compared to each other). For ES gender and ES year, we observe that the absolute value of the effect drops when fewer trials are used to train the SOM and calculate the effect. More importantly, however, is that the sign of these ESs does not change with using fewer trials, again an indication for their effect being true. Comparing the columns, we see that for ES gender, the values remain relatively stable except for the column where the angular velocity variables were dropped. This may mean that a part of the observed gender difference may be accounted for by the fact that the SOM encountered larger errors with these variables. The values for the ES year were not affected by the angular velocities. Looking at the ES expertise, we observe, similar to step one, values close to zero and directional instability in the sign of the ES. Only when the angular velocity variables were dropped from the dataset did we observe a consistency in the sign of ES expertise.

Discussion

The aim of this study was to examine the stability of analyses with Self-Organizing Maps in an application for human movement and sport science. We used two related datasets to examine the map quality and sensitivity of analyses in coordination variability to changes in SOM parameters and to changes in the number of trials and variables used to train the SOMs. This additional stability analysis can be a solution for researchers to avoid poor map quality and to increase confidence in the conclusions drawn from the study. Research setting and data-driven selection of SOM parameters (Button et al., 2014) can then be supported or extended with parameters that increase the quality. For example, choosing between the options for lattice, initialization, training type and neighborhood function has no theoretical rationale in the datasets in this study or cannot be performed based on the data. Other SOM parameters can be data-driven but can be proportionally scaled to decrease either of the errors, depending on the type of data and subsequent analyses that will be performed with the SOM. These results may also increase the integration of SOMs and other machine learning tools into the human movement and sport sciences.

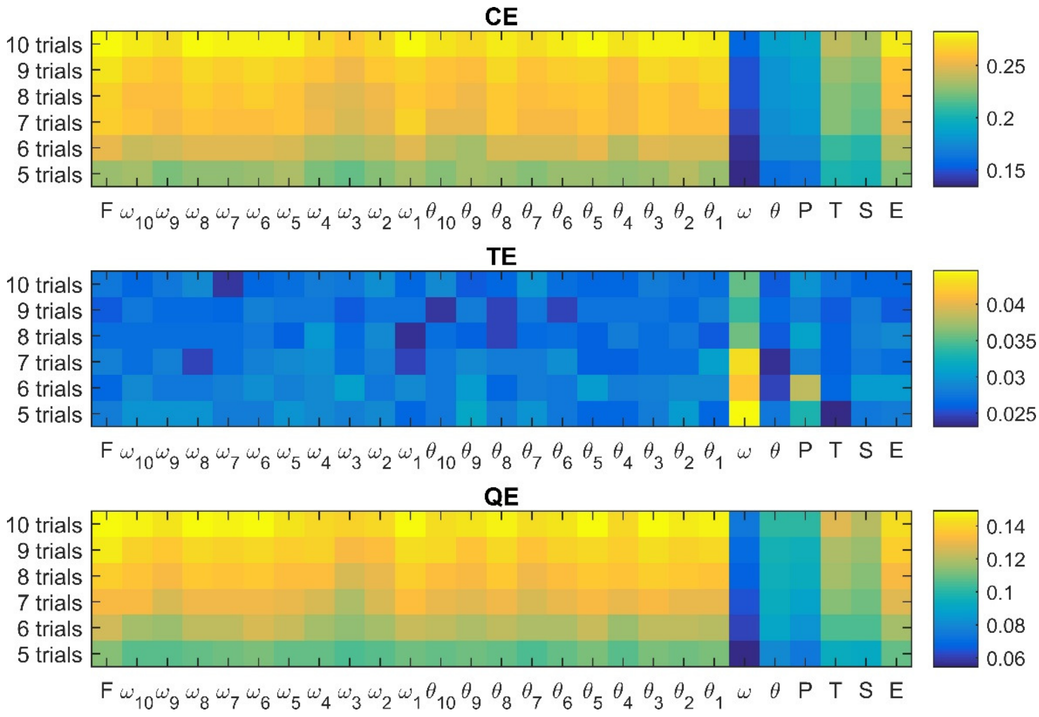


Figure 4: Dataset 1. Colour code plots representing the effect on the map quality of manipulating the number of trials used to train the SOM (rows) and the components representing the input vectors (columns). The first column (F) represents the full dataset (20 components). The other columns represent simulations where a specific angular velocity or angle time series was dropped from the dataset (ω_i or θ_i; see text for the numbers), where we dropped all angular velocities and angle time series (ω, θ), all pelvis variables (P), all trunk variables (T), all shoulder variables (S) or the elbow variables (E).

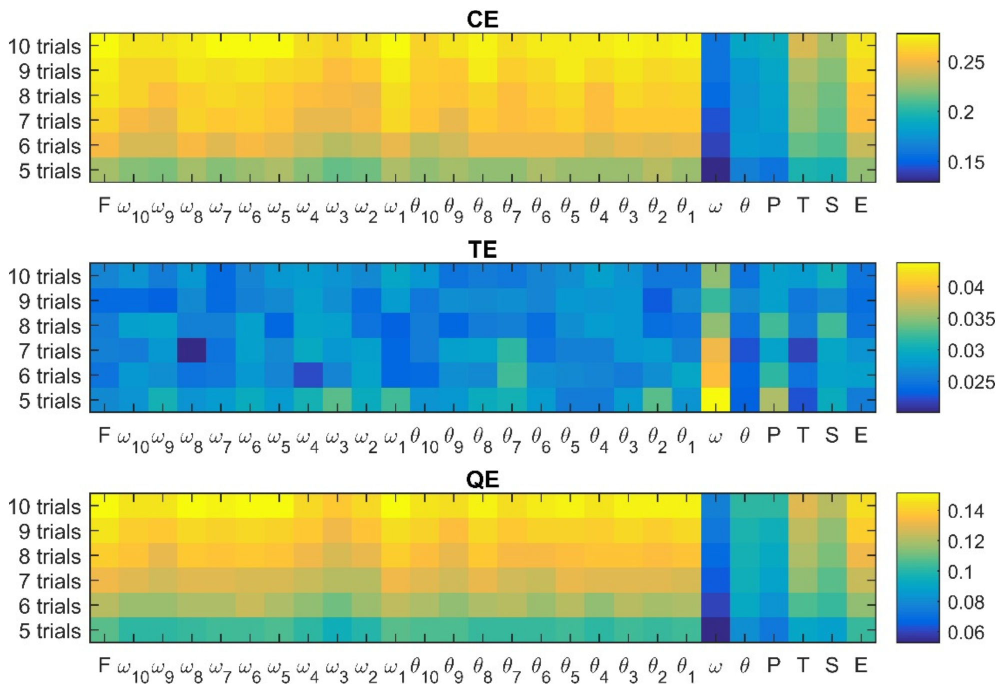


Figure 5: Dataset 2. Colour code plots representing the effect on the map quality of manipulating the number of trials used to train the SOM (rows) and the components representing the input vectors (columns). The first column (F) represents the full dataset (20 components). The other columns represent simulations where a specific angular velocity or angle time series was dropped from the dataset (ω_i or θ_i; see text for the numbers), where we dropped all angular velocities and angle time series (ω, θ), all pelvis variables (P), all trunk variables (T), all shoulder variables (S) or the elbow variables (E).

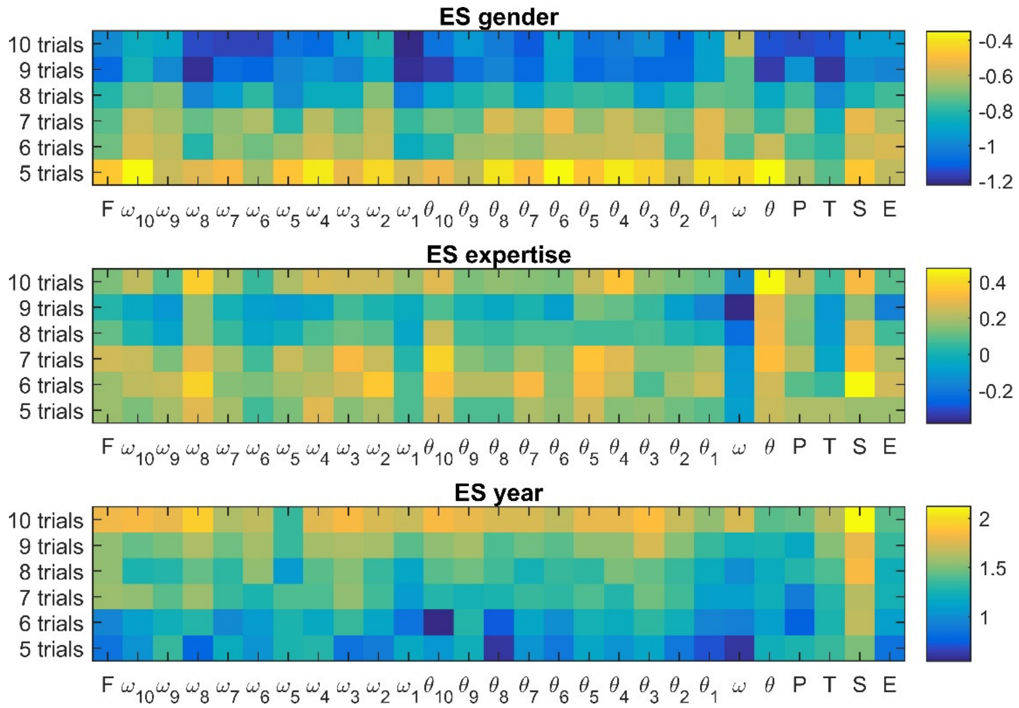


Figure 6: Colour code plots representing the sensitivity of the effect sizes on manipulating the number of trials used to train the SOM (rows) and the components representing the input vectors (columns). The first column (F) represents the full dataset (20 components). The other columns represent simulations where a specific angular velocity or angle time series was dropped from the dataset (ω_i or θ_i; see text for the numbers), where we dropped all angular velocities and angle time series (ω, θ), all pelvis variables (P), all trunk variables (T), all shoulder variables (S) or the elbow variables (E). (Top and middle plot: dataset 1, bottom plot: dataset 2).

SOM parameters

Both datasets showed that the same set of options was best suited for an optimal SOM quality: hexagonal lattice, linear weight vector initialization, sequential training type (with a Gaussian neighborhood function), long training length and big map size. The latter two options are not really startling and the errors could even be decreased further with user-defined map sizes and training lengths but go beyond the data-driven and proportionally scaled options that we explored. The other options have no direct theoretical rationale for being the optimal choice. Perhaps it can be argued that the sequential training type outperforms the batch training in the present datasets because they contain time-series data and therefore is able to capture and preserve the time-continuous non-linear relationships (the reverse conclusion would have applied if we had chosen the QE as the secondary map quality variable). The type of neighborhood function had no influence on the map quality. As a limitation to our analysis, we should mention the linear vs. random weight vector initialization. While we performed the random initialization five different times, this is by no means enough to firmly conclude that the linear initialization will always perform better (five times was chosen to keep computation times within reasonable limits). Another issue is that the random initialization procedure is not reproducible. More importantly, however, is our observation that the ESs were not significantly different between the linear and random procedures (although ES gender and ES expertise seem to be more consistent for linear, indicated by the smaller range of the boxplot). The effect of these parameters on the effect sizes can be used as an additional check of their validity. In the previously published paper on this dataset (Serrien, Ooijen, et al., 2016b), the gender effect was found to be significant and the expertise effect not. The sensitivity analysis of the present study shows that the statistical significance of the gender effect represents a true effect and is not dependent on any arbitrary choice of SOM parameters. Conversely, the non-significance of the expertise effect is confirmed in the directional instability of the sign of the ES. For dataset 2 (not published before), the ES year is likely a true effect because its range was very small over a large space of parameter options.

Number of trials and components

The second step of our analyses showed how within a ‘best-set-of-options’, the results can be further interpreted and checked for their validity. For the ‘true’ effects (ES gender and ES year), we observed a clear decrease in the absolute value of the effect size when using fewer number of trials as input. This means that the calculation of coordination variability became less stable when using fewer trials (because the effect size calculation itself is not affected by the number of trials). Checking the effect on map quality of every variable separately or sets of variables by dropping them from the dataset allowed us to examine for which variables the SOM encountered difficulties. Again, we want to point out that this should not be done to drop all variables with larger associated errors but to look at the effect of these ‘difficult’ variables on the sensitivity of subsequent SOM-based analyses. When some variables pose a problematic effect on SOM quality compared to others, it is advisable to check whether these same variables also affect the sensitivity of the conclusions drawn from the study. In the present datasets, it was shown that not one single variable had any effect on SOM quality compared to the full dataset. Dropping the collective set of angular velocities compared to angles produced clearly lower CE and QE, but higher TE. The lower CE and QE can be explained by the higher variability in the angular velocity time series. However, it does not explain the higher TE values. A reason for the latter may be that dropping the angular velocities causes a decrease in the power to distinguish between different coordination states (represented by the units). If for example a certain joint angle shows similar positions during two different phases of the motion, the angular positions for the other degrees of freedom will be different, but the angular velocities will be much more different. Without the angular velocities, the SOM may be using the same regions of nodes to represent different phases of the motion, but inclusion of the velocities effectively prevents this. This is a trade-off between distinguishing coordination states (including more variables) versus an accurate representation (including fewer variables), that researchers should contemplate on a priori. Inclusion of other levels of analysis (EMG, kinetics) would be even more powerful to distinguish between coordination states.

When we then look at the effect of dropping these angular velocities on the ESs, we see an interesting observation. The effect size of expertise shows now more directional stability (always negative values) whereas it showed instability in step one and when dropping other variables. The ‘true’ ESs (gender and year) keep their directional stability. However, the value of ES gender was much lower compared to the full dataset and compared to other columns (especially for the calculations with all ten trials), whereas the value of ES year was not affected by the angular velocities. It seems that the higher the values of the ES (ES year > ES gender), the more robust it is against inclusion of variables with larger associated errors.

Concluding remarks

Whether the main conclusions in this study can be extrapolated to similar studies in human movement and sport science with time series data is unknown and a formulation of a set of guidelines for consistent use of SOM parameters and number of variables is not relevant. The better strategy would be to report this kind of analysis of the SOM quality and sensitivity of the subsequent analyses. For the datasets and corresponding research questions analysed in this article, we can robustly confirm our previous findings of lower coordination variability for male volleyball players and no difference between junior and elite players (Serrien, Ooijen, et al., 2016b) and the (unpublished) finding that the coordination variability in junior players was lower in the second year of the longitudinal study. In the present study, these SOM-based analyses were rather simple (between and within subjects differences), so we should examine also more complex analyses like clustering or applications where two or more SOMs are used in series (Lamb et al., 2014, 2011; Serrien, Clijsen, et al., 2016). Especially in these more difficult applications would it be advisable to examine the effects closely. When SOMs are purely used for qualitative analyses (data exploration, visualization purposes), performing a systematic sensitivity analysis will become difficult if researchers have to examine a large parameter space. Artificial neural networks will always provide the researchers with an output, but it remains the researchers' task to determine its validity, sensitivity to the selection of certain parameters and whether the conclusions make sense based on theoretical rationale and the study's setting.

Acknowledgements

We are grateful to the two anonymous reviewers for their suggestions that helped us improve the article.

References

Bartlett, R. M. (2006). Artificial Intelligence in Sports Biomechanics: New Dawn or False Hope? *Journal of Sport Science and Medecine*, *5*(July), 474–479.

Bartlett, R. M., Lamb, P. F., O'Donovan, D., & Kennedy, G. (2014). Use of self-organizing maps for exploring coordination variability in the transition between walking and running. *International Journal of Comparative Psychology*, *27*, 123–137. http://doi.org/10.5811/westjem.2011.5.6700

Button, C., Wheat, J., & Lamb, P. (2014). Why coordination dynamics is relevant for studying sport performance. In K. Davids, R. Hristovski, D. Araújo, N. B. Serre, C. Button, & P. Passos (Eds.), *Complex Systems in Sport* (pp. 44–62). New York: Routledge.

Kaski, S., & Lagus, K. (1996). Comparing self-organizing maps. In W. von der Malsburg, J. C. von Seelen, & B. Sendhoff (Eds.), *Proceedings of ICANN96, International Conference in Artificial Neural Networks, Lecture Notes in Computer Science* (Vol. 112, pp. 809–814). Berlin: Springer. http://doi.org/10.1007/3-540-61510-5_136

Kohonen, T. (2001). *Self-Organizing Maps* (3rd. Ed.). Springer.

Lamb, P. F., Bartlett, R. M., Lindinger, S., & Kennedy, G. (2014). Multi-dimensional coordination in cross-country skiing analyzed using self-organizing maps. *Human Movement Science*, *33*, 54–69. http://doi.org/10.1016/j.humov.2013.08.005

Lamb, P. F., Bartlett, R. M., & Robins, A. (2011). Artificial neural networks for analyzing inter-limb coordination: the golf chip shot. *Human Movement Science*, *30*, 1129–1143. http://doi.org/10.1016/j.humov.2010.12.006

Lamb, P. F., Bartlett, R. M., Robins, A., & Kennedy, G. (2008). Self-organizing maps as a tool to analyze movement variability. *International Journal of Computer Science in Sport*, *7*(1), 28–39.

Lamb, P. F., & Stöckl, M. (2014). On the use of continuous relative phase: Review of current approaches and outline for a new standard. *Clinical Biomechanics*, *29*(5), 484–493. http://doi.org/10.1016/j.clinbiomech.2014.03.008

Lapham, A. C., & Bartlett, R. M. (1995). The use of artificial intelligence in the analysis of sports performance: A review of applications in human gait analysis and future directions for sports biomechanics. *Journal of Sports Sciences*, *13*, 229–237.

Schöllhorn, W. I. (2004). Applications of artificial neural nets in clinical biomechanics. *Clinical Biomechanics (Bristol, Avon)*, *19*(9), 876–98. http://doi.org/10.1016/j.clinbiomech.2004.04.005

Schöllhorn, W. I., Chow, J. Y., Glazier, P., & Button, C. (2014). Self-organizing maps and cluster analysis in elite and sub-elite athletic performance. In K. Davids, R. Hristovski, D. Araújo, N. B. Serre, C. Button, & P. Passos (Eds.), *Complex Systems in Sport* (pp. 145–159). New York: Routledge.

Serrien, B., Clijsen, R., Anders, S., Goossens, M., & Baeyens, J. P. (2016). Intra-seasonal Variability of Ball Speed and Coordination of Two Team-Handball Throwing Techniques in Elite Male Adolescent Players. *International Journal of Computer Science in Sport*, *15*(1), 1–29. http://doi.org/10.1515/ijcss-2016-0001

Serrien, B., Ooijen, J., Goossens, M., & Baeyens, J. P. (2016a). A Motion Analysis in the Volleyball Spike – Part 1: Three-Dimensional Kinematics and Performance . *International Journal of Human Movement and Sports Sciences*, *4*(4), 70–82. http://doi.org/10.13189/saj.2016.040403

Serrien, B., Ooijen, J., Goossens, M., & Baeyens, J. P. (2016b). A Motion Analysis in the Volleyball Spike – Part 2: Coordination and Performance Variability. *International Journal of Human Movement and Sports Sciences*, *4*(4), 83–90. http://doi.org/10.13189/saj.2016.040404

Tan, H. Sen, & George, S. E. (2004). Investigating learning parameters in a standard 2-D SOM model to select good maps and avoid poor ones. In *AI-2004: Advances in Artificial Intelligence*. (Vol. 3339, pp. 425–437). Berlin: Springer-Verslag.

van Diest, M., Stegenga, J., Wörtche, H. J., Roerdink, J. B. T. M., Verkerke, G. J., & Lamoth, C. J. C. (2015). Quantifying postural control during exergaming using multivariate whole-body movement data: A self-organizing maps approach. *PLoS ONE*, *10*(7). http://doi.org/10.1371/journal.pone.0134350

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). *SOM Toolbox for Matlab 5*.

Appendix

SOM algorithm

The SOM-algorithm is explained in more detail in Kohonen (2001) and in Vesanto, Himberg, Alhoniemi, & Parhankangas (2000) for its implementation in the SOM Toolbox for Matlab. The following is a brief overview of the algorithm and its various options. SOMs are neural networks that are a combination of vector quantization and vector projection algorithms. A typical SOM has two layers, an input and output layer, constructed from a set of interconnected neurons (units). Each output unit is associated with a weight vector $\boldsymbol{w_i} = (w_{i,1} \quad \cdots \quad w_{i,D}), i \in \{1 \quad \ldots \quad I\}$ (I being the number of output units) with the same dimension (number of variables, D) as the input vectors $\boldsymbol{z_j} = (z_{j,1} \quad \cdots \quad z_{j,D}), j \in \{1 \quad \ldots \quad J\}$ (J being the number of input vectors). The constraint $\frac{I}{J} < 1$ and topological preservation through the neighborhood function (see below) will force similar input vectors to cluster together into close regions on the map. All units are arranged on a two-dimensional sheet, cylinder or toroidal shape with coordinate vector $\boldsymbol{r_i} = (r_{i_x} \quad r_{i_y})$ meaning that the weight vectors are projected to a lower dimension. For the present study of a discrete motion like a volleyball spike, the two-dimensional sheet is the most natural choice, while cylindrical and toroidal options are more appropriate for continues rhythmic motions. The two-dimensional sheet can be organized in either a rectangular or hexagonal lattice, which has an influence on the number of neighbouring units within a certain radius. For a neighborhood radius of one, a rectangular grid has four neighbours, while a hexagonal grid has six neighbours, possibly influencing the connectivity between neighbouring units and subsequent training of the map. The default size of the map (length l, width w) is based on the ratio of the two largest eigenvalues (λ) of the training data: $\frac{l}{w} = \sqrt{\frac{\lambda_1}{\lambda_2}}$. The SOM Toolbox also allows users to choose proportionally scaled map sizes of ¼ (small) and 4 times (big) the default size to alter the ratio between I and J.

The training algorithms of the SOM (sequential and batch training) depend on the selection and updating of best-matching-units (BMUs). The SOM uses a Euclidean metric, indicating that the BMU for an input vector $\boldsymbol{z_j}$ is selected based on the criterion $\min_i\{\|\boldsymbol{z_j} - \boldsymbol{w_i}\|\} = \min_i\left\{\sqrt{\sum_{d=1}^{D}\left(z_{j,d} - w_{i,d}\right)^2}\right\}$. This metric would be more complicated in the case of missing values and masks giving more or less importance to certain components. In the present context of using SOMs for analyzing coordination variability, we will ignore these complications. The Euclidean metric space implicates that some form of input vector normalization is imperative for a good functioning of the SOM. Without normalization, vector components with high values on their original scale would dominate the algorithm, leaving only minor contributions to other components. Possible normalization options include (1) range normalization to a [0 1] or a [-1 1] interval based on the minimal and maximal values for a certain component (2) variance normalization or z-transformation to a standard normal distribution $\sim N(0,1)$ for each component separately, (3) logarithmic normalization, (4) logistic normalization, and (5) normalizations to an ordinal scale. The first option of range normalization is the most natural choice for the present application with multi-dimensional time series. Also in phase-plane analysis, the [-1 1] range normalization is the recommended option (Lamb & Stöckl, 2014).

Before the SOM can be trained, the weight vectors $\boldsymbol{w_i}$ have to be initialized with either a linear or random initialization. The linear function initializes the weight vectors along the two greatest eigenvectors of the covariance matrix of the training data, while the random function creates a uniform random distribution per dimension over the interval $[\min(z_d)\ \max(z_d)]$ where the extrema are calculated over all normalized input vectors $\boldsymbol{z_j}$. The linear initialization creates an initial order in the map so that the convergence to a stable solution should be faster. When the weight vectors are initialized, training of the SOM can begin to adjust the weight vectors so they are a closer representation of the input vectors.

The sequential training algorithm presents the entire dataset, input vector after input vector, to the SOM units during each iteration of the training. After several iterations, the SOM will converge to a relatively constant solution. Each iteration $(t = 1 \quad t = 2 \quad \ldots \quad t = T)$, the algorithm searches for the BMU (competition between the output units) and updates this unit's and its neighbours' (cooperation between the output units) weight vectors by the following rule: $\boldsymbol{w_i}(t+1) = \boldsymbol{w_i}(t) + \alpha(t)h_{BMU,i}(t)[\boldsymbol{z}(t) - \boldsymbol{w_i}(t)]$, where $\alpha(t)$ is the learning rate and $h_{BMU,i}(t)$ the neighborhood kernel around the BMU. Both learning rate and neighborhood kernel are also a function of time, decreasing their influence on the map during consecutive iterations. The learning rate can take three different forms: linearly decreasing: $\alpha(t) = \alpha_0(1 - \frac{t}{T})$, power function: $\alpha(t) = \alpha_0(\frac{0.005}{\alpha_0})^{\frac{t}{T}}$ or reciprocally decreasing: $\alpha(t) = \frac{\alpha_0}{(1+\frac{100t}{T})}$ (with T the total training length and α_0 the initial learning rate, both parameters are selected data-driven). The neighborhood kernel can take four different forms: bubble: $h_{BMU,i}(t) = \mathbf{1}(\sigma_t - d_{BMU,i})$, Gaussian: $h_{BMU,i}(t) = e^{-(\frac{d^2_{BMU,i}}{2\sigma_t^2})}$, cut-of Gaussian: $h_{BMU,i}(t) = e^{-\left(\frac{d^2_{BMU,i}}{2\sigma_t^2}\right)}\mathbf{1}(\sigma_t - d_{BMU,i})$ or Epanechikov (ep): $h_{BMU,i}(t) = \max\{0,1 - (\sigma_t - d_{BMU,i})^2\}$, where $\sigma_t = \sigma_0 e^{-\frac{t}{\lambda}}$ is the neighborhood radius at iteration t (with the time constant $\lambda = \frac{T}{\sigma_0}$), $d_{BMU,i} = \|\boldsymbol{r_{BMU}} - \boldsymbol{r_i}\|$ is the distance between unit i and the BMU in grid space and $\mathbf{1}(x)$ is the step function: $\mathbf{1}(x) = 0\ if\ x < 0\ and\ \mathbf{1}(x) = 1\ if\ x > 0$. This vector learning algorithm causes topological similar units (vector quanta) to self-organize themselves into clusters on the map. This training algorithm is performed in two steps, the first step (rough training) starts with large initial learning rate and neighborhood radius while the fine-tuning phase starts with small rate and radius from the beginning. In the batch training algorithm, not one input vector at a time is presented to the SOM, but an entire batch of them together (based on a partitioning of the dataset in Voronoi regions). The update of the weight vectors is then based on the rule:

$$\boldsymbol{w_i}(t+1) = \frac{\sum_{j=1}^{J} h_{BMU,i}(t)\boldsymbol{z_j}}{\sum_{j=1}^{J} h_{BMU,i}}.$$

The former illustrates the large number of possibilities that can be used in constructing and training a SOM. Also non data-driven options of $T, I, \sigma_0, \alpha_0, l, w$ can be chosen by the researchers and further increases the SOM parameter space, but this goes beyond our scope.

Computation time

In Figure A below, we have plotted the computation time for the SOM algorithm split between the different options for the six SOM parameters. The computations were performed on a Windows 7 pc (DELL, Alienware 14, dual core processor, i7). Timing was done with Matlab's tic and toc functions. The computation time included reading in the data, initializing and training the SOM and calculating the movement variability (for 1 subject, for 1 simulation). This timing was performed only for the simulations of step 1 (full dataset, 1210 input vectors, 20 components). The results in Figure A below are concatenated over both datasets. Significant differences exist for training type (batch < sequential), map size (small < medium < large) and training length (short, normal < long).

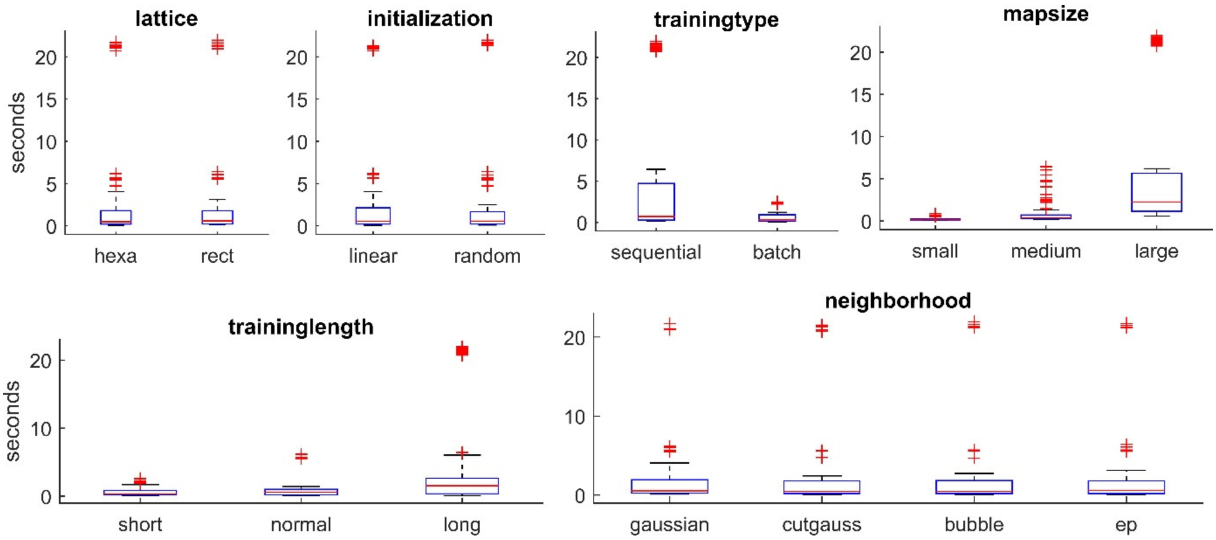


Figure A: Boxplots representing computation time for the SOM algorithm.