Nonergodicity in Load and Recovery

Neumann, Niklas; Van Yperen, Nico W.; Brauers, Jur; Frencken, Wouter; Brink, Michel; Lemmink, Koen A.P.M.; Meerhoff, Rens; den Hartigh, Ruud

Published in:
International journal of sports physiology and performance

DOI:
10.1123/ijspp.2021-0126

Title of this article: Non-Ergodicity in Load and Recovery: Group Results Do Not Generalize to Individuals

Submission type: Original investigation

Full names of the authors and institutional/corporate affiliations: Niklas D. Neumann[1], Nico W. Van Yperen[1], Jur J. Brauers[2], Wouter Frencken[2,3], Michel S. Brink[2], Koen A.P.M. Lemmink[2], Laurentius A. Meerhoff[4], and Ruud J.R. Den Hartigh[1]

Affiliations

[1]Department of Psychology, University of Groningen, The Netherlands

[2]Center for Human Movement Sciences, University Medical Center Groningen, University of Groningen, The Netherlands

[3]Football club Groningen, Groningen, The Netherlands

[4]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, The Netherlands

Corresponding author:

Name: Niklas Dieter Neumann

Institution: Department of Psychology, University of Groningen

Postal address: Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

E-Mail: n.d.neumann@rug.nl

Telephone: +31 (0)6 12398651

Orcid ID: 0000-0003-2062-1958

Preferred running head: Non-Ergodicity in Load and Recovery

Abstract word count: 249

Text-only word count: 4087

Number of figures: 3

Number of tables: 3

Citation (APA 7th edition):

Neumann, N.D., Van Yperen, N.W., Brauers, J.J., Frencken, W., Brink, M.S., Lemmink, K.A.P.M., Meerhoff, L.A., Den Hartigh, R.J.R. (in press). Non-Ergodicity in Load and Recovery: Group Results Do Not Generalize to Individuals. *International Journal of Sports Physiology and Performance.*

Abstract

Purpose: The study of load and recovery gained significant interest in the last decades, given its important value in decreasing the likelihood of injuries and improving performance. So far, findings are typically reported on the group-level, whereas practitioners are most often interested in applications at the individual-level. Hence, the aim of the present research is to examine to what extent group-level statistics can be generalized to individual athletes, which is referred to as the "ergodicity issue". Non-ergodicity may have serious consequences for the way we should analyze, and work with, load and recovery measures in the sports field. ***Methods***: We collected load, i.e., rating of perceived exertion (RPE) * training duration, and total quality of recovery (TQR) data among youth male players of a professional football club. This data was collected on a daily basis across two seasons and analyzed on both the group- and the individual-level. ***Results***: Group- and individual-level analysis resulted in different statistical outcomes, particularly with regard to load. Specifically, standard deviations *within* individuals were up to 7.63 times larger than standard deviations *between* individuals. In addition, at either level, we observed different correlations between load and recovery. ***Conclusions***: The results suggest that the process of load and recovery in athletes is non-ergodic, which has important implications for the sports field. Recommendations for training programs of individual athletes may be suboptimal, or even erroneous, when guided by group-level outcomes. The utilization of individual-level data is key to ensure the optimal balance of individual load and recovery.

Keywords: Dynamical Systems, Football, Intra-individual Variability, Monitoring, Resilience.

Introduction

Within sports science, the study of load and recovery gained significant interest in the last decades.[1–6] Optimal training responses can be achieved via the exposure to different loads and sufficient recovery to perform at peak capacity.[4] On the one hand, researchers and practitioners aim to develop resilience through exposing players to high workloads in order to prepare them for the physical demands of competition.[7] On the other hand, higher workloads are associated with a greater risk for injuries.[8,9] Indeed, a poor balance between load and recovery may lead to overuse injuries and illness,[2,10,11] as well as immune system dysregulation, mood swings,[10] and ultimately to a long-term decrement of performance.[4,11] Hence, to optimize sport performance and to reduce the risk of injuries, the need for individual monitoring and analysis in sports is rising.[3]

To date, however, studies on load and recovery, and related outcomes, are typically conducted at the group-level. It is highly questionable whether group-level results generalize to individual processes.[12–16] Simply put, findings at the group-level may mask meaningful variability between subjects and only allow "on average" statements.[15] Indeed, sampling across individuals (group-level analysis) provides insights into how scores on variables are distributed across individuals, rather than how the scores on variables behave across time. To improve our understanding of the latter processes, researchers should collect time series data across consecutive measurement occasions within a particular person (individual-level analysis).[17–19]

In the field of behavioral sciences, Molenaar provided a comprehensive description of how time series data of multiple subjects should be treated when investigating processes within individuals.[20] He referred to the Cattell data box as an illustration, where time (occasion) can be seen as one dimension and measured variables as another dimension (**Figure 1**).[21] If multiple subjects are added, a third dimension emerges and forms the three-dimensional data box. Group data constitute vertical slices, whereas individual data refer to horizontal slices. Thus, in group-level analysis, one selects only one or a few fixed data points (occasions) as well as a subset of variables, while pooling across subjects. In individual-level analysis, one focuses on a single subject as well as a subset of variables, while pooling across a range of data points (occasions). The variation of the scores can be determined by pooling across time.

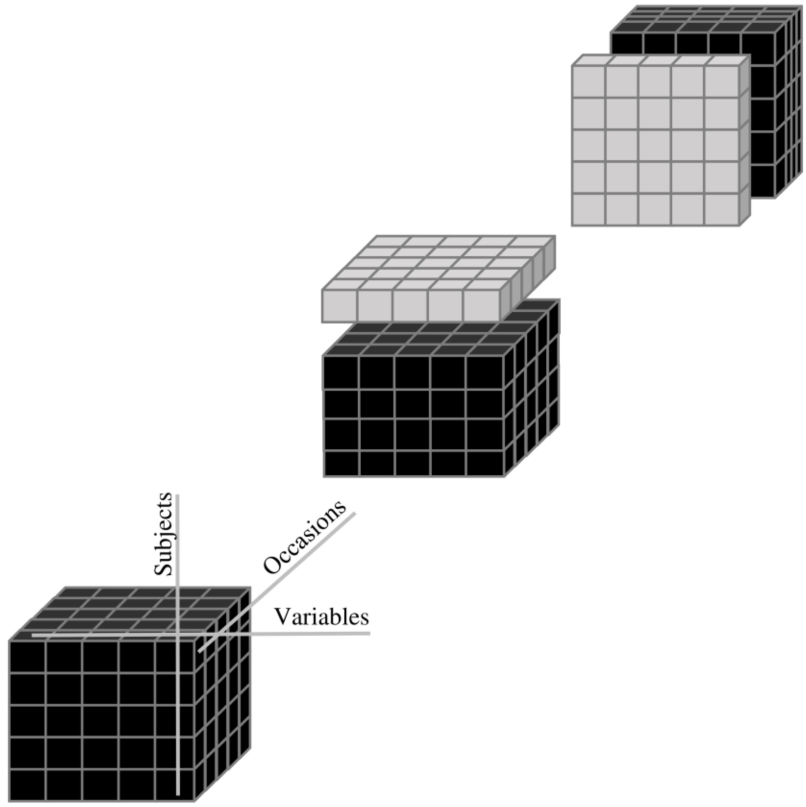

Figure 1: The Cattell data box.[21] Any taken measure is defined as an intersection of occasion (considered as *days* in this study), variable, and subject (bottom left picture). Vertical slices (top right picture) represent the group-level analysis, i.e., single occasion, multiple variables, and multiple subjects, whereas horizontal slices (middle picture) constitute the individual-level analysis, i.e., multiple occasions, multiple variables, and a single subject.

In monitoring athletes' load and recovery, both theorists and practitioners are interested in how these two variables develop across time within individual athletes (i.e., horizontal slices). Specifically, theorists want to understand these processes and provide evidence-based recommendations to practitioners. The practitioners wish to understand why and how they should coach their athletes to enhance sport performance and to avoid injuries. As noted earlier, however, a model based on samples of individuals (vertical slices) often *does not* generalize to a model of individual processes. Take the following example: Research within a football team might reveal that, on a group-level, the standard deviation (SD) in load and recovery is rather small. However, on the individual-level it is much larger, indicating increased fluctuations from moment to moment, which would not be detected in group-level data. Yet, these individual fluctuations are crucial in order to optimally adapt training load and recovery for each individual athlete. A comparable issue applies to the correlation. For instance, load could be positively correlated with performance on the group-level.[9,22] On the individual-level, however, a much smaller correlation, or even a negative correlation, might exist.[9,23] This means that a rather high amount of load may lead to a decrease in performance, possibly as a result of overloading the individual athlete.

The issue that models based on group-level analysis have no logical bearing on models of individual processes is called the 'ergodicity problem'.[17,18,20,24,25] This problem stems from the ergodic theorem, which mathematically describes the conditions that must be met in order to generalize statistical phenomena across levels and units of analysis.[20,26] For instance, in human subject data, the variations within and between individuals must be asymptotically equivalent, which is rarely the case.[24] Ergodic processes are equivalent for groups and individuals if the same statistical model applies to the data of all subjects in the population and if the data has invariant statistical characteristics across time.[20] These two conditions are referred to as homogeneity and stationarity. The ergodicity problem thereby holds that, simply put, statistics of central tendencies, variations, as well as correlations of time series at the group-level differ from those at the individual-level.[24] Hence, Fisher et al. concluded that the lack of group-to-individual generalizability (i.e. non-ergodicity) is a threat to human subjects research in general.[24] That is, the literature tends to overestimate the accuracy of aggregated statistical estimates and the generalizability of conclusions between group- and individual outcomes.

Taken together, the ergodicity problem may be an important issue to account for in the study of load and recovery. Yet, no study in the field of sports science has tested whether group results were generalizable to the individual load-recovery processes. To fill this gap, we collected load and recovery scores among youth players of a professional football club across two seasons. To test for (non-)ergodicity, we addressed the question whether group-level statistics of load and recovery represent the individual-level statistics within this group. More specifically, we tested whether: (a) The univariate distributions (mean, median, and SD), and (b) the bivariate correlations (Pearson's *r* and SD) of load and recovery scores differ between the group- and the individual-level.

Methods

Subjects

A total of 82 youth male football players were included for the current study. They were members of the youth academy of a major league (Eredivisie) football club in the Netherlands, and were playing for the under 17 (U-17), under 19 (U-19), or under 23 (U-23) team. The U-17 and U-19 teams competed in the highest national leagues of those age categories. The U-23 team participated in a national level senior league, in the third division. The mean ages (SD) of the U-17, U-19, and U-23 teams were 15.96 (.62) years, 17.59 (.54) years, and 19.16 (.96) years, respectively. The mean heights were 176.99 (7.59) cm, 181.66 (7.54) cm, and 182.75 (5.89) cm. The mean weights were 64.90 (9.67) kg, 70.77 (8.34) kg, and 75.49 (6.80) kg. Due to personal data protection, the names of the three youth teams are randomly referred to as team 1, 2, and 3. The players had between six and eight training sessions per week, which are composed of two strength training sessions of 60-75 minutes and four to six field training sessions of 75-90 minutes.

Design

The present study was conducted according to the requirements of the Declaration of Helsinki, and was approved by the ethics committee of the Faculty of Behavioral and Social Sciences of the University of Groningen (the Netherlands) (research code: PSY-1819-S-0308). The time series data is based on measures of perceived exertion, training duration, and perceived recovery. The measures are part of the normal, daily team monitoring routine at the club and used by trainers to optimize the training design. That is, every day right after each training session, up to a maximum of 30 minutes, each player indicated the exertion score on a tablet computer near the locker room without staff or team members being present. Before the first training session of the day, participants filled out the recovery question on the same tablet computer.

Methodology

We measured the perceived exertion with the session Ratings of Perceived Exertion (sRPE) scale, consisting of a single item: "How hard was the training?"[27–29] The RPE was first introduced by Borg[27] as a psychophysical measure of exertion and fatigue with a rating range between 6 (very, very light) to 20 (very, very hard), indicating the heart rates between 60 and 200 beats per minute. As demonstrated by Arney et al.[30], the scores on the Borg 6-20 RPE scale (BORG-RPE) and the often-used category ratio (0-10) RPE scale (BORG-CR10) correlate very highly (r = .90). We relied on the BORG-RPE scale (rather than the BORG-CR10 scale) because (1) it has been the standard measure at the club for many years, and (2) its response-scale aligns with the Total Quality Recovery (TQR) 6-20 scale we used in this study (see below). The sRPE scale serves to provide a subjective estimate of internal training load (referred to as 'load' throughout), which corresponds to the physiological stress imposed on athletes. sRPE has been shown to be a valid, useful, and practical method to monitor and control load.[31] The sRPE is derived by multiplying the RPE at the end of a training session by the total duration (in minutes) of the training session. We took this load measure as a unit of analysis in the current study.[e.g., 2,4,28,30]

The perceived recovery was measured with the TQR scale, also consisting of one item: “How good is your recovery?”[1] The rating of the TQR is structured around the RPE scale, ranging from 6 (very, very poor recovery) to 20 (very, very good recovery). In previous research, the TQR scale was shown to be highly correlated to more objective measures, such as creatine kinase, and researchers increasingly recommend using this measure to monitor the recovery of athletes.[32,33]

The data sets

In the current study we aimed at making comparisons between the individual, time-varying data and the cross-sectional, aggregated (group) data.[24] The original data set consisted of 84 players from three youth teams and 22,128 observations (i.e. 263.43 observations per player on average) across two seasons. To properly answer our research question, we used strict inclusion criteria. We included observations from the data set for the analysis if there was an RPE score of the training session(s) or match the day before, the duration of the training session(s) in minutes, and the TQR score of the next day for the same player. This means that if there was no training or match on a specific day or when the data was missing due to an injury of the player, absence, or other reasons, the previous, the current and the subsequent day had to be dropped. Indeed, if one such data point would be missing, we were not able to calculate correlation coefficients for load and TQR. Thus, one complete observation included the recovery at the current day (TQR) and the training load of the previous day (RPE * training duration in minutes).

Furthermore, we only included days with at least two measurements in total to be able to calculate variations and correlations of load and TQR per day (i.e., group-level analysis). If there was more than one training session the day before, we added the load scores to each other so they would reflect the actual load experienced that day. Further inspection of the data revealed that two players only had two and four data points, respectively, and were therefore removed from the data set. The minimum number of data points for the remainder of the athletes was 21. After applying the inclusion criteria, we removed 11,073 data points and ended up with a data set that consisted of 82 players and 11,055 observations. To be more specific, team 1 consisted of 25 players across 286 days, with 113.12 observations per player on average; team 2 consisted of 24 players across 271 days, with 132.83 observations per player on average, and team 3 consisted of 33 players across 330 days, with 122.10 observations per player on average.

In addition, we extracted a subset out of this data set to obtain a symmetric data box (**Figure 1**). By using the statistical program R, we identified the maximum amount of players that had identical consecutive data points in any period of the seasons without a single missing value. Thereby, the number of participants and the number of observations per participant had to be identical (i.e., symmetrical) to equalize the statistical power for both types of analysis.[24] The application of this criteria returned 10, 15, and 11 players of team 1, 2, and 3, respectively, and the same amount of consecutive data points for those players (i.e. 10, 15, and 11).

Statistical Analysis

In our analysis we followed the recommendations of Molenaar and Campbell[20] and Fisher et al.[24] on how time series of multiple individuals and variables should be treated, and how to test for (non-)ergodicity. In the context of the current study this means that, for the group-level analysis, we looked at athlete 1 to athlete *n*, day 1, and selected the variables load and recovery. Based on the scores of all athletes for that single *day*, the univariate distributions (mean, median, and SD) and bivariate correlations (Pearson's *r* and SD) were calculated. This step was repeated for every single subsequent day. At the end, these results were averaged across all days. For the correlations, we first transformed the coefficient by using Fisher's *z*, averaged these values, and back transformed them to Pearson's *r*. This results in a less biased outcome than averaging the raw correlation coefficients.[34]

For the individual-level analysis, we looked at athlete 1, data point t_1 to t_i, and selected the variables load and recovery. Based on the time series of that individual athlete, the univariate distributions (mean, median, and SD) and bivariate correlations (Pearson's *r* and SD) were calculated. This step was repeated for every *individual* athlete. At the end, these results were averaged across all individuals. Here we also transformed the correlation coefficients first by using Fisher's *z*, averaging these values, and back transforming them to Pearson's *r*. If findings on the group- and the individual-level are equivalent, the process can be considered ergodic.[20]

Finally, we calculated 95% confidence intervals (CI) for means and SDs to determine if the differences between the group- and individual-level are statistically meaningful. For the comparison of the bivariate correlations, we relied on Pearson product-moment correlations and 95% CI. Accordingly, we interpreted the magnitude of the correlations as trivial (< .10), small (.10 - .29), moderate (.30 - .49), large (.50 - .69), very large (.70 - .90), nearly perfect (> .90), or perfect (= 1.00).[35]

Results

Univariate Distributions

We calculated the statistics with R and Rstudio (version R.3.6.3; R Foundation for Statistical Computing, Vienna, Austria).[36] We first examined the univariate distributions of the variables load and recovery at the group- and individual-level for all three teams. The means [95% CI], medians, and SDs [95% CI] are presented in **Table 1**. Mean estimates were comparable between the group and the individual, reflected by overlapping CIs. However, the medians and SDs for group and individual estimates showed discrepancies. To be more specific, the 95% CIs of the SD do not overlap between the group- and individual-level analysis for all three teams and the two variables load and recovery (see bold numbers in **Table 1**). Thus, the results reflect a wider range of variability across individual estimates, which is also reflected by the high ratio of SD between individual and group. For instance, the SD of load for individuals is 3.1 times the size of the SD in the group (see Team 2).

The results of the larger data set are largely replicated by the symmetrical subset. The procedure of analysis was equal to the previous analysis. The means, medians, and SDs [95% CI] are presented in **Table 2**. Mean estimates were identical between the group- and individual-level analysis, because of the symmetry in this data set (i.e., the number of players and data points per players is equal). However, also in this symmetrical subset, the medians and SDs for group and individual estimates showed a wider range of variability across individual results. Given the sparse amount of data points, the CIs were partly overlapping for recovery. Yet, this was not observed in load. In particular, the SD of the individual-level analysis was up to 7.63 times the size of the group-level analysis (see **Table 2**: perceived load, Team 2).

Bivariate Correlations

We conducted bivariate correlations between the variables load and recovery for the aggregated group cross-sections and individual time series. **Figure 2** presents density plots of the correlational distributions for groups and individuals of the larger data set. Mean correlations (*r*), SD, and 95% CI for each team and level are reported above each figure. The overall magnitude of the correlations ranged from trivial to small. The correlations differed in the interpretation of the magnitude between the two types of analysis. For instance, we found a trivial correlation for team 1 on the group-level (r = -.09), and a small negative correlation on the individual-level (r = -.24). Furthermore, the CIs were not overlapping for teams 1 and 2 and partly overlapping for team 3. Across all three teams, the SD of group correlations was larger than the SD in individuals.

Table 1: Univariate distributions of load and recovery for groups and individuals of the larger data set.

		Group			**Individual**			
		Mean [95% CI]	*Median*	*SD [95% CI]*	*Mean [95% CI]*	*Median*	*SD [95% CI]*	*I:G ratio*
Team 1	Load	1,635.42 [1,602.63-1,668.22]	1,694.61	**281.78** **[250.60-312.96]**	1,623.54 [1,329.23-1,917.85]	1,388	**713** **[680.95-745.04]**	*2.53*
	Recovery	14.18 [14.08-14.28]	14.15	**.85** **[.81-.90]**	14.22 [13.75-14.68]	14.26	**1.12** **[1.05-1.19]**	*1.32*
Team 2	Load	1,579.85 [1,553.06-1,606.63]	1,610.19	**223.96** **[193.84-254.08]**	1,574.23 [1,283.15-1,865.32]	1,343.65	**689.35** **[668.15-710.54]**	*3.10*
	Recovery	15.30 [15.18-15.42]	15.25	**.99** **[.91-1.06]**	14.83 [14.27-15.39]	14.79	**1.32** **[1.23-1.42]**	*1.33*
Team 3	Load	1,726.84 [1,691.96-1,761.71]	1,793.04	**322.10** **[298.48-345.66]**	1,789.33 [1,542.71-2,035.95]	1,693.17	**695.52** **[675.96-715.08]**	*2.16*
	Recovery	13.84 [13.74-13.93]	13.83	**.84** **[.79-.88]**	13.90 [13.54-14.26]	14.06	**1.01** **[.92-1.10]**	*1.20*

Note. The means, medians, and SDs are the mean values of all single means, medians, and SDs of either the day (group) or the player (individual). The I:G ratio illustrates the ratio of individual SD to group SD. The numbers in bold reflect the meaningful differences in SD between the group- and the individual-level analysis, determined by non-overlapping CIs.

Table 2: Univariate distributions of load and recovery for groups and individuals of the symmetrical subset.

		Group			Individual			
		Mean	*Median*	*SD [95% CI]*	*Mean*	*Median*	*SD [95% CI]*	*I:G ratio*
Team 1	Load	1,754.69	1,754.85	**262.16 [64.72-459.59]**	1,754.69	1,526.40	**784.68 [737.76-831.59]**	*2.99*
	Recovery	13.50	13.50	.76 [.62-.90]	13.50	13.65	.90 [.64-1.15]	*1.18*
Team 2	Load	1,458.07	1,472	**99.68 [43.84-155.52]**	1,458.07	1,285.67	**760.24 [739.34-781.13]**	*7.63*
	Recovery	14.49	14.47	.78 [.61-.94]	14.49	14.40	1.02 [.89-1.15]	*1.31*
Team 3	Load	1,550.69	1,546.55	**251.01 [110.85-391.17]**	1,550.69	1,697.91	**762.20 [734.15-790.25]**	*3.04*
	Recovery	13.69	13.45	1.23 [.99-1.48]	13.69	13.64	1.29 [1.08-1.49]	*1.05*

Note. The means, medians, and SDs are the mean values of all single means, medians, and SDs of either the day (group) or the player (individual). The I:G ratio illustrates the ratio of individual SD to group SD. The numbers in bold reflect the meaningful differences in SD between the group- and the individual-level analysis, determined by non-overlapping CIs.

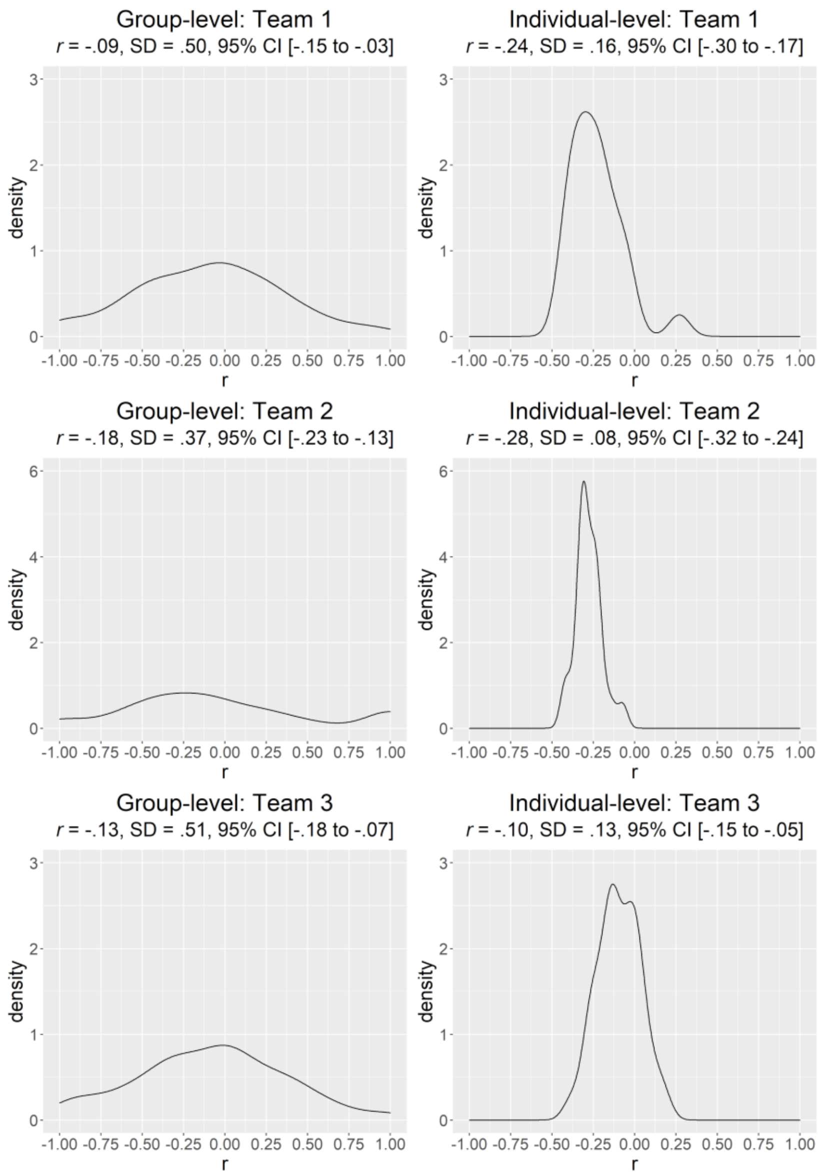

Figure 2: Density plots of the Pearson correlations between the group- and individual-level analysis for the teams 1, 2, and 3 of the larger data set. The mean correlation, the SD, and 95% CI are stated above each plot.

Again, the results of the larger data set are mostly replicated by the symmetrical subset. The overall magnitude of the correlations ranged from trivial to moderate. Mean correlations (*r*), SD, and 95% CI are presented in **Table 3**.

Table 3: Bivariate correlations of load and recovery for groups and individuals of the symmetrical subset

	Group			Individual		
	r	SD	[95% CI]	*r*	SD	[95% CI]
Team 1	-.07	.51	[-.43 to .29]	-.37	.32	[-.60 to -.14]
Team 2	-.03	.28	[-.20 to .15]	-.28	.24	[-.41 to -.15]
Team 3	-.14	.42	[-.43 to .14]	.10	.23	[-.06 to .25]

Note. Pearson's *r* is the mean value of all single correlations of either the day (group) or the player (individual).

Similar to the larger data set, the correlations differed in the interpretation of the magnitude between the two types of analysis. The group-level correlations were only trivial for the teams 1 and 2, whereas the individual-level correlations were moderate and small, respectively. Besides, the group-level correlation of team 3 was negatively small (r = -.14), whereas it was positively small in individuals (r = .10). Given the sparse amount of data points in this subset, the CIs were partly overlapping in all 3 teams.

Discussion

The aim of this study was to test whether the process of load and recovery in youth football players can be considered as ergodic. This would be the case if the same statistical model can be generalized across levels and units of analysis.[20,26] To be more specific, we tested whether (a) the univariate distributions (mean, median, and SD), and (b) the bivariate correlations (Pearson's *r* and SD) of load and recovery scores differ between the group- and the individual-level. Clarifying this ergodicity issue is important, because researchers tend to report group-based outcomes of load and recovery measures and their relation to, amongst others, performance and injuries. As expected, our results suggest that group-level statistics cannot be generalized to individual athletes. That is, the results between group- and individual-level analysis showed discrepancies between statistical estimates, which we found across the large data set and a symmetrical subset. Together, the findings converge on the proposition that the process of load and recovery in youth football players is non-ergodic. This implies that results based on (1) data aggregated all at once, (2) individual scores averaged before calculating correlations, or (3) analysis of data occasion by occasion, may be statistically invalid for intra-individual research questions and individual-level interventions. This is in line with previous literature on ergodicity in the medical, behavioral, and social sciences.[18–20,24,25]

In addition, the outcomes are particularly interesting *in the way* they differ from each other. An important finding is that the medians and SDs exhibited discrepancies between the group-level and individual-level analyses, while mean estimates were comparable.

Differences in SDs were most striking, with individual-level analysis revealing values up to 7.63 times larger than the group-level analyses. Hereby, the CIs for the SD were not (larger data set) or only partly (subset) overlapping. This indicates increased fluctuations within individuals and thereby moment to moment changes. Hence, the mean is not a good representative value for the individual and must be regarded in relation to the SD. The interpretation of the group-level and individual-level correlations (i.e., strength and direction) differed as well. For instance, the strength of association between load and recovery for team 1 in both data sets was, on average, only trivial on the group-level and small (large data set) and moderate (subset) on the individual-level. Besides, the direction of the average group correlation for team 3 of the symmetrical data set was opposite to the average individual correlation. Overall, this implies that correlation estimates can be stronger, weaker, or oppositely correlated for individual team members than the team-level results suggest. Yet, it has to be noted that the CIs for the symmetrical subset were mainly overlapping, given the sparse amount of data points.

Another interesting observation is that the SDs of the correlation coefficients were larger for the group-level analysis than for the individual-level analysis. This means that the strength and direction of the correlation fluctuates from occasion to occasion. The mean correlation of the group must therefore be regarded in relation to the SD. In turn, individual correlations varied less, which means that the overall individual mean correlation is more representative for the single individual. If practitioners take these outcomes seriously, long-term consequences, such as underloading and overloading, and thereby injuries and under-performance, might be prevented. [1,7]

It should be noted that the type of analysis (i.e. group-level vs individual-level analysis) certainly depends on the type of question a researcher poses.[17,18] If the aim is to investigate the distribution of variables across individuals, then the group-level analysis is the appropriate way of analysis.[17,19] This can be achieved by sampling across individuals and it enhances our understanding of *variables*. A suitable question could be: How are load values associated with recovery values between athletes? If the aim is to investigate individual processes, that is, the behavior of certain variables during the course of people's time series, then the individual-level analysis is the appropriate type of analysis. This, in turn, can be achieved by sampling across consecutive occurrences and it will enhance our understanding of *processes*. A suitable question could be: How are changes in load values associated with changes in recovery values within the athlete? This emphasizes the importance of precisely identifying the research question of interest and defining the analysis accordingly.[18]

Practical implications

The current study showed that group-based statistics do not generalize to the individual athlete. Hence, recommendations for training programs of individual athletes may be suboptimal, or even erroneous, when guided by group-level outcomes. This can be seen in

Figure 3, where the group mean is sometimes in line with the individual, but it can also be lower or considerably higher. In designing and adjusting training programs, sports practitioners need to carefully interpret research findings on load and recovery, and their relation to performance and injuries. Consequently, in the monitoring of athletes and the design of (individual) training programs, practitioners should rely primarily on individual-level results.[12,37] For instance, coaching staff of sport clubs can apply individualized monitoring as well as individualized graphical representations of the time-series.[e.g., 37] In that way, they may better determine if athletes are likely to be under- or overloaded, and when. This in turn may help tailoring the training program for the individual athlete. As a next step, practitioners could benefit from analytic tools to provide personalized insights. Researchers may provide these analytical tools in order to enrich data-driven insights about the individual athlete.

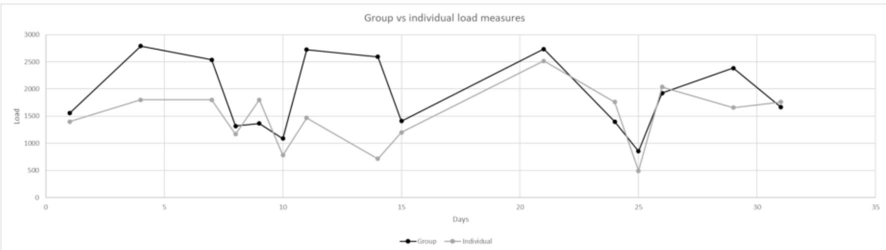

Figure 3: Example of load measures over a 31-day period for the whole team (group) and an individual player (individual).

Limitations and recommendations for future research

The original data set included missing data, which is inherent to the data collection in team sport contexts. Therefore, we also used a symmetrical subset with no missing data in our study. Interestingly, analyses of the original dataset and the symmetrical subset converged in the same conclusion, that is, the process of load and recovery in athletes is non-ergodic. This is reassuring because missing data is the rule rather than the exception in research and practice in the sports field.

We advise future studies to test for (non-)ergodicity before aggregating the results across levels of analysis and to be very clear on the levels of application of their findings. As a logical, and relatively simple step, correlations between load and recovery may be determined based on how these variables relate to each other across multiple occasions within individuals. This can be accomplished by applying a technique such as the repeated measures correlation (rmcorr).[18] In contrast to simple correlations, it does not violate the assumptions of independence of observations and it shows greater statistical power, because averaging and aggregating is not necessary. Note, however, that the result of this technique is still one overall correlation coefficient of multiple individuals who were analyzed on an individual-level. If researchers aim for individualized insights into load and recovery processes, time series analysis may be used. In the field of sports, promising steps in nonlinear time series analysis have already been made in the past years.[38] Furthermore, outside sports interesting techniques have been proposed to analyze relationships between variables as they change over time within individuals (e.g., time-varying vector autoregressive models)[39].

Conclusion

Research on load and recovery, and sport science research in general, needs to take (non-) ergodicity into account when analyzing the data. Transferring group-based outcomes (and interpretations) to the individual level is most likely inaccurate. Using an individual-level analysis on athletes' load and recovery data is an important step toward an optimal, individualized approach to performance monitoring and performance improvement.

Acknowledgement

This work was supported by The Netherlands Organization for Health Research and Development [ZonMw, grant number 546003004]. We want to thank Dr. Bertus Jeronimus and Dr. Yannick Hill for helpful and insightful comments, Dr. Casper Albers for statistical discussions, Mees van der Linde for his support in collecting the data, and Rik Huijzer for his support with the analysis scripts.

References

1. Kenttä G, Hassmén P. Overtraining and recovery. *Sport Med*. 1998;26(1):1-16. doi:10.2165/00007256-199826010-00001
2. Brink MS, Visscher C, Arends S, Zwerver J, Post WJ, Lemmink KAPM. Monitoring stress and recovery: new insights for the prevention of injuries and illnesses in elite youth soccer players. *Br J Sports Med*. 2010;44(11):809-815. doi:10.1136/bjsm.2009.069476
3. Heidari J, Beckmann J, Bertollo M, et al. Multidimensional monitoring of recovery status and implications for performance. *Int J Sports Physiol Perform*. 2019;14(1):2-8. doi:10.1123/ijspp.2017-0669
4. Kellmann M, Bertollo M, Bosquet L, et al. Recovery and performance in sport: Consensus statement. *Int J Sports Physiol Perform*. 2018;13(2):240-245. doi:10.1123/ijspp.2017-0759
5. Kellmann M. Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring. *Scand J Med Sci Sport*. 2010;20(Suppl. 2):95-102. doi:10.1111/j.1600-0838.2010.01192.x
6. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *Br J Sports Med*. 2016;50(5):281-291. doi:10.1136/bjsports-2015-094758
7. Gabbett TJ. The training-injury prevention paradox: Should athletes be training smarter and harder? *Br J Sports Med*.

2016;50(5):273-280. doi:10.1136/bjsports-2015-095788
8. Bowen L, Gross AS, Gimpel M, Li FX. Accumulated workloads and the acute: Chronic workload ratio relate to injury risk in elite youth football players. *Br J Sports Med*. 2017;51(5):452-459. doi:10.1136/bjsports-2015-095820
9. Bourdon PC, Cardinale M, Murray A, et al. Monitoring athlete training loads: Consensus statement. *Int J Sports Physiol Perform*. 2017;12(s2):161-170. doi:10.1123/IJSPP.2017-0208
10. Fry RW, Grove JR, Morton AR, Zeroni PM, Gaudieri S, Keast D. Psychological and immunological correlates of acute overtraining. *Br J Sports Med*. 1994;28(4):241-246. doi:10.1136/bjsm.28.4.241
11. Brink MS, Visscher C, Coutts AJ, Lemmink KAPM. Changes in perceived stress and recovery in overreached young elite soccer players. *Scand J Med Sci Sport*. 2012;22(2):285-292. doi:10.1111/j.1600-0838.2010.01237.x
12. Hill Y, Meijer RR, Van Yperen NW, Michelakis G, Barisch S, Den Hartigh RJR. Nonergodicity in protective factors of resilience in athletes. *Sport Exerc Perform Psychol*. 2020. doi:http://dx.doi.org/10.1037/spy0000246
13. Den Hartigh RJR, Hill Y, Van Geert PLC. The development of talent in sports: A dynamic network approach. *Complexity*. 2018;2018:13. doi:10.1155/2018/9280154
14. Chrzanowski-Smith OJ, Piatrikova E, Betts JA, Williams S, Gonzalez JT. Variability in exercise physiology: Can capturing intra-individual variation help better understand true inter-individual responses? *Eur J Sport Sci*. 2020;20(4):452-460. doi:10.1080/17461391.2019.1655100
15. Glazier PS, Mehdizadeh S. Challenging conventional paradigms in applied sports biomechanics research. *Sport Med*. 2019;49(2):171-176. doi:10.1007/s40279-018-1030-1
16. Glazier P, Lamb P. Inter- and intra-individual movement variability in the golf swing. In: Toms M, ed. *Routledge International Handbook of Golf Science*. Routledge; 2018:49-63. doi:10.1080/14763141.2011.650187
17. Van Geert P. Group versus individual data in a dynamic systems approach to development. *Enfance*. 2014;3(3):283-312.
18. Bakdash JZ, Marusich LR. Repeated measures correlation. *Front Psychol*. 2017;8(456):1-13. doi:10.3389/fpsyg.2017.00456
19. Bland JM, Altman DG. Statistics Notes: Correlation, regression, and repeated data. *Br Med J*. 1994;308(6933):896. doi:10.1136/bmj.308.6933.896
20. Molenaar PCM, Campbell CG. The new person-specific paradigm in psychology. *Curr Dir Psychol Sci*. 2009;18(2):112-117. doi:10.1111/j.1467-8721.2009.01619.x
21. Cattell RB. The three basic factor-analytic research designs—their interrelations and derivatives. *Psychol Bull*. 1952;49(5):499.

22. Foster C, Daines E, Hector L, Snyder AC, Welsh R. Athletic performance in relation to training load. *Wis Med J*. 1996;95(6):370-374.
23. Gabbett TJ, Domrow N. Relationships between training load, injury, and fitness in sub-elite collision sport athletes. *J Sports Sci*. 2007;25(13):1507-1519. doi:10.1080/02640410701215066
24. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci U S A*. 2018;115(27):E6106-E6115. doi:10.1073/pnas.1711978115
25. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*. 2004;2(4):201-218. doi:10.1207/s15366359mea0204
26. Birkhoff BD. Proof of the ergodic theorem. *Proc Natl Acad Sci U S A*. 1931;(3):656-660.
27. Borg GAV. Psychophysical bases of perceived exertion. *Med Sci Sport Exerc*. 1982;14(5):377-381.
28. Foster C, Florhaug JA, Franklin J, et al. A new approach to monitoring exercise training. *J Strenght Cond Res*. 2001;15(1):109-115. doi:10.1016/0968-0896(95)00066-P
29. McLaren SJ, Smith A, Spears IR, Weston M. A detailed quantification of differential ratings of perceived exertion during team-sport training. *J Sci Med Sport*. 2017;20(3):290-295. doi:10.1016/j.jsams.2016.06.011
30. Arney BE, Glover R, Fusco A, et al. Comparison of RPE (rating of perceived exertion) scales for session RPE. *Int J Sports Physiol Perform*. 2019;14(7):994-996. doi:10.1123/ijspp.2018-0637
31. Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, Marcora SM. Use of RPE-based training load in soccer. *Med Sci Sports Exerc*. 2004;36(6):1042-1047. doi:10.1249/01.MSS.0000128199.23901.2F
32. Osiecki R, Rubio TBG, Coelho RL, et al. The total quality recovery scale (TQR) as a proxy for determining athletes' recovery state after a professional soccer match. *J Exerc Physiol Online*. 2015;18(3):27-32.
33. Sansone P, Tschan H, Foster C, Tessitore A. Monitoring training load and perceived recovery in female basketball. *J Strength Cond Res*. 2018;34(10):2929-2936. doi:10.1519/jsc.0000000000002971
34. Silver NC, Dunlap WP. Averaging correlation coefficients: Should Fisher's z transformation be used? *J Appl Psychol*. 1987;72(1):146-148. doi:10.1037/0021-9010.72.1.146
35. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 2009;41(1):3-12. doi:10.1249/MSS.0b013e31818cb278
36. R Foundation for Statistical Computing. R: A language and environment for statistical computing. 2013. http://www.r-

project.org/.
37. Orie J, Hofman N, Meerhoff LA, Knobbe A. Training distribution in 1500-m speed skating: A case study of an olympic gold medalist. *Int J Sports Physiol Perform*. 2020;1(aop):1-5. doi:10.1123/ijspp.2019-0544
38. Davids K, Hristovski R, Araújo D, Serre NB, Button C, Passos P. *Complex Systems in Sport*. 1st ed. New York, US: Routledge; 2014.
39. Haslbeck JMB, Bringmann LF, Waldorp LJ. A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivariate Behav Res*. 2020;65(1):120-149. doi:10.1080/00273171.2020.1743630