Brazilian Journal of Biometrics, **44**, 1–19 (2026)
doi:10.28951/bjb.v44i2.946

BRAZILIAN JOURNAL OF BIOMETRICS
ISSN:2764-5290

ARTICLE

Generalised additive models for location, scale, and shape in sports science: a systematic review

Lucas F. Rosa,[1] Andréa C. Konrath,[2] and Luiz R. Nakamura[⋆,3]

[1]Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras, Brazil
[2]Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis, Brazil
[3]Departamento de Estatística, Universidade Federal de Lavras, Lavras, Brazil
[⋆]Corresponding author. Email: luiz.nakamura@ufla.br

(Received: August 1, 2025; Revised: December 18, 2025; Accepted: January 30, 2026; Published: May 15, 2026)
Section Editor: Camilla Marques Barroso

Abstract

Sport has increasingly evolved into an interdisciplinary research field, where statistical modelling plays a central role in performance analysis, injury prediction, tactical optimisation, and audience behaviour. Although traditional methods like linear regression and generalised linear models are frequently employed, their inability to handle complex data structures has led to the expanding usage of more flexible approaches. In this context, generalised additive models for location, scale and shape (GAMLSS) are one of the most flexible statistical frameworks currently available, enabling the simultaneous modelling of multiple distributional parameters. Hence, in this paper, we perform a detailed systematic review exploring the application of GAMLSS in sports science, drawing on peer-reviewed articles. The results show a variety of applications, such as the development of reference growth curves, athlete performance modelling, match-fixing detection, and forecasting of sport-related consumer behaviour. GAMLSS have proven especially useful in contexts where traditional models are inadequate, offering enhanced flexibility in capturing distributional nuances. Nonetheless, opportunities remain to integrate GAMLSS with machine learning techniques and to extend their use across underexplored domains in sport. This review contributes to the field by outlining current trends, highlighting methodological strengths, and identifying promising directions for future research.

Keywords: Data analysis; Distributional regression models; GAMLSS; Predictive modelling; Sports analytics; Sport statistics; Statistical modelling.

1. Introduction

Sports analytics has established itself as an important and rapidly expanding field of research, distinguished by its impact to translate sports data into strategic information to support informed decision-making. Its applications include analysing athlete performance, predicting match outcomes, detecting opponents' strengths and weaknesses, and optimising team performance (Gifford & Bayrak, 2023; Wheatcroft, 2021). According to Exel & Dabnichki (2024), sports managers and scouts quickly recognised the potential of sports analytics for performance assessment and prediction, and its use has since expanded, becoming a critical component of strategic sports management and talent scouting.

Science has increasingly been used as a tool to support the management of teams and athletes within the sports industry, both in the short and long term. Budgeting, physical health, game strategies, training regimes, and most importantly, sports performance are now more frequently being handled analytically (Sarlis *et al.,* 2021). This approach involves the collection, interpretation, and modelling of large volumes of data generated during sporting events, including individual statistics, collective performance metrics, and environmental factors (Wei, 2024). This field also intersects with sports innovation, an emerging area that connects sport to organisational and technological innovation practices, where rivalry is a key factor, encouraging athletes and teams to strive for excellence, creativity, and competitive advantage (Ayodeji & Abiodun, 2024).

Despite its potential, sports performance analysis still faces considerable challenges. This is largely due to the multidimensional nature of sport, which includes physical, technical, tactical, and psychological aspects, alongside demands such as injury prevention, fitness monitoring, and the development of strategies to enhance audience involvement (Wei, 2024). These complexities require methodological approaches that are both flexible and suited to the specific characteristics of each sport. Furthermore, identifying key variables or predictors of success in highly competitive contexts remains one of the core challenges for study in this field (Makaruk *et al.,* 2024). In practice, elite sports performance has become the science of marginal gains, which means that athletes must be at their peak to deliver results at a specific moment within a defined period (Exel & Dabnichki, 2024).

The increasing complexity of sports-related analyses has led to the development of several methods and algorithms for predicting individual performance as well as team tactical efficiency (Pantzalis & Tjortjis, 2020). Statistics plays an important role in this evolution because it has been present in sports science since its inception and is required for validating studies, and adapting to the field's specificities using concepts inherited from other disciplines (Exel & Dabnichki, 2024). From data collection and organisation to the application of complex predictive models, statistical methods allow for more in-depth understanding of the elements that influence athlete and team performance. According to Jain *et al.* (2021), coaches need models to evaluate their players, analyse competing teams, and develop successful strategies. Almarashi *et al.* (2024) argue that such methods enable athletes, coaches, schools, sports teams, and training institutions to modify physical education and training based on informed decision-making.

A wide range of methodologies have been employed within the sporting context. In individual sports, statistical analyses include, for example, univariate tests (Swindell *et al.,* 2019), multivariate analysis (Staunton *et al.,* 2024), traditional regression models (Abalo-Núñez *et al.,* 2018), sequential analysis (Podrigalo *et al.,* 2018), and machine learning algorithms (Yue *et al.,* 2022). The same applies to team sports, with research involving, for instance, standard hypothesis testing (Patel *et al.,* 2022), causal inference (Cerqueira *et al.,* 2017), traditional regression models (Roczniok *et al.,* 2016), and machine learning (Markopoulou *et al.,* 2024). These studies highlight the diversity and sophistication of statistical methods used in sports analytics, reflecting the continual advancement of quantitative approaches to understanding athletic performance, injury prevention, and training optimisation across different sports.

Although still underexplored in the sports context, generalised additive models for location, scale, and shape (GAMLSS) (Rigby & Stasinopoulos, 2005), provide a valuable alternative to the standard methodologies extensively used in the field. GAMLSS provide a general and flexible framework within the class of univariate semiparametric regression models, combining ease of interpretation with great predictive power (Ramires *et al.,* 2021) and extend traditional statistical models such as generalised linear models (GLM) (Nelder & Wedderburn, 1972) and generalised additive models (GAM) (Hastie & Tibshirani, 1986) by allowing not only the location parameter (e.g., the mean) but also any parameters associated with the scale and shape of the response distribution to be modelled explicitly and simultaneously as functions of different sets of covariates. Mathematically, GAMLSS are defined as

$$Y \sim \mathcal{D}\left(\theta_k\right)$$
$$g_k\left(\theta_k\right) = \boldsymbol{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} s_{kj}\left(\boldsymbol{x}_{kj}\right),$$

where Y is the response variable, which follows any probability distribution $\mathcal{D}$ with parameter vector $\theta_k = \left(\theta_1, \ldots, \theta_p\right)^\top$, $g_k(\cdot)$ is an appropriate link function, $\boldsymbol{X}_k$ is a design matrix, $\boldsymbol{\beta}_k$ is a parameter vector, and $s_{kj}(\cdot)$ are smoothing functions used to explain the relationship between covariate $\boldsymbol{x}_{kj}$ and θ_k.

Among other advantages, GAMLSS also allow for the specification of a wide range of response variable distributions -– going beyond the exponential family to which GLM and GAM are limited -– many of which have parameters with straightforward interpretations, thereby facilitating the analysis of results (Roquim *et al.,* 2023). These models are also capable of capturing both linear and non-linear relationships (via smoothing functions) between covariates and each distributional parameter, providing a more accurate representation of the effects of explanatory variables rather than implying that all relationships are linear. Furthermore, GAMLSS are supported by robust diagnostic tools which enable a thorough assessment of model fit and the identification of potential misspecifications (Rigby *et al.,* 2019). With significant gains in both inference and prediction, the combination of features positions GAMLSS as a powerful alternative to the traditional approaches frequently adopted in the literature (Stasinopoulos *et al.,* 2018). A more detailed discussion of the advantages of GAMLSS can be found in (Pala *et al.,* 2026).

Nevertheless, in less complex modelling situations, such as when the primary interest is in average performance measures rather than in understanding distributional characteristics that may distinguish exceptional athletes or teams, simpler frameworks such as GLM and GAM may already provide an adequate modelling strategy.

In the sporting domain, where data frequently exhibit complex patterns and non-linear relationships, GAMLSS emerge as a significant advancement in statistical methodology, providing greater flexibility and precision in the analysis of the phenomena involved. In this context, the present paper aims to systematically review the scientific literature on the application of GAMLSS to sports data, with the objective of identifying the main sporting contexts in which this framework has been applied, and of highlighting trends, gaps, and opportunities for future research. The systematic review will map the key topics already explored, the methodologies adopted, and the progress made in the application of these models within the field.

The rest of this paper is outlined as follows. The second section describes the methodology, detailing the research questions, inclusion and exclusion criteria, data sources consulted (including databases, time ranges, and search strategies), and the procedures adopted to ensure the reproducibility of the review. The third section presents the results, combining a bibliometric analysis (to map trends, collaboration networks, and publication patterns) with a qualitative metasynthesis of the final portfolio of articles, aimed at integrating evidence and identifying thematic convergences. The final section offers concluding remarks, in which the main findings are summarised.

2. Materials and Methods

The following questions were established to guide the investigation: (i) How have GAMLSS been applied in sports analysis, according to the scientific literature? (ii) What are the main objectives and sporting contexts in which GAMLSS have been used?; (iii) What types of data have been modelled using GAMLSS in sports studies?; (iv) Which distributions and regression structures are most commonly used?; (v) What advantages have been observed when compared to other statistical methods?; and (vi) What gaps or limitations still persist in these applications?. These questions guided the whole methodological process, from article selection to result analysis, allowing for trend mapping, pattern recognition, and the identification of opportunities for future research.

Developed based on the analysis of papers, methods, frameworks, and best practices for literature reviews and their outcomes, the systematic search flow (SSF) (Ferenhof & Fernandes, 2016; Ferenhof & Fernandes, 2025) was adopted in this systematic review. The SSF method consists of four phases: (i) Phase 1: Protocol definition (formulation of the search strategy; systematic querying of scientific databases, with automated alerts configured for updates; organisation of references using reference management software; standardised selection procedures; and composition of the final portfolio following full-text reading of the selected studies); (ii) Phase 2: Analysis (Data consolidation); (iii) Phase 3: critical synthesis; and (iv) Phase 4: Writing.

For this study, the following databases were consulted: Web of Science, Scopus, PubMed, SportDiscus, SpringerLink, IEEE, SciELO, and ScienceDirect, all accessed through the CAPES Portal (http://periodicos.capes.gov.br/), ensuring comprehensive coverage of the scientific literature. The search strings used are presented in Table 1. Only papers published in English, indexed in scientific journals, and related to the application of GAMLSS in the sporting context were considered. The review covered the period from 2005 –– the year of the seminal GAMLSS paper by Rigby & Stasinopoulos (2005) — to 2024, with the search conducted during the first half of February 2025.

Table 1. Results obtained from scientific databases using the applied search strategy

Search keywords	Database	Number of articles
(GAMLSS OR “Generali*ed Additive Model* for Location* Scale* and Shape” OR “Distributional Regression*”) AND (Sport* OR Game*)	SpringerLink	76
	Web of Science	55
	PubMed	39
	SPORTDiscus	15
	Scopus	9
	IEEE	0
	SciELO	0
	ScienceDirect	0
Total		194

A total of 194 articles were initially retrieved, with SpringerLink returning the highest number of results (76), followed by Web of Science (55), PubMed (39), SPORTDiscus (15), and Scopus (9). No publications were returned in IEEE, SciELO, or ScienceDirect. Zotero (Zotero, 2024) software was used to cross-check articles retrieved from the different databases, which enabled the identification and removal of duplicates. In addition to duplicates, 12 non-article documents, such as conference abstracts, were detected in SpringerLink. Thus, the final number of unique articles totalled 118. Figure 1 summarises the steps of identification, screening, eligibility, and inclusion of studies, following Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (Page *et al.*, 2021).

The standardisation of document selection involved the use of predefined filters, with minimum criteria including a clear relation to sport — either through the use of data originating from sporting contexts (such as competitions, social projects, school or sport development programmes) –– and

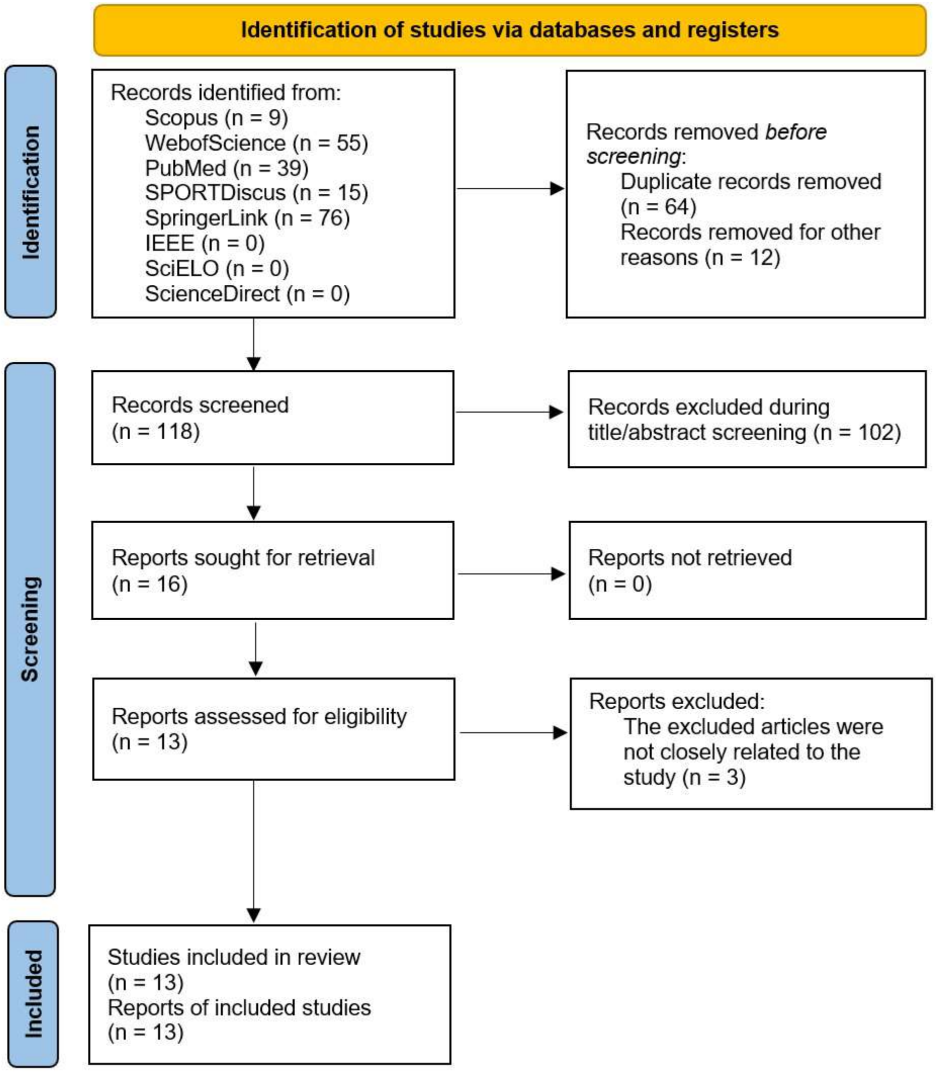

Figure 1. Methodological framework of the study based on the PRISMA flow diagram.

the use of GAMLSS. Initially, the titles, abstracts, and keywords of the 118 articles selected in the previous stage were examined, allowing for a preliminary screening of articles most relevant to the research topic. As a result, 16 articles were pre-selected as potential candidates for inclusion in the final portfolio, corresponding to approximately 13.56% of the total analysed at this stage.

The 16 pre-selected papers were then subjected to a full-text reading to identify those effectively aligned to the research objective. At this stage, relevant data were extracted and organised in a spreadsheet, whereas papers that, despite containing pertinent keywords, did not adhere to the scope of the research were excluded. In the end, three articles were disregarded for not meeting the minimum established criteria, resulting in a final portfolio comprising 13 selected studies.

The bibliometric analysis, including data organisation and visualisation, was carried out using the `R` software (R Core Team, 2025), with the `bibliometrix` package (Aria & Cuccurullo, 2017) and the `biblioshiny` interface.

3. Results and Discussion

This section presents an analysis of the 13 articles that comprise the final portfolio, structured into two complementary stages: (i) a bibliometric analysis, focused on mapping quantitative patterns of scientific production; and (ii) a qualitative metasynthesis, aimed at synthesising the theoretical and methodological contributions of the selected papers.

3.1 Bibliometric analysis

The bibliometric analysis covers exclusively the period from 2018 to 2024 (as no documents were identified prior to 2018, which already suggests that the application of GAMLSS in sports is still at an early stage) and includes 13 publications. These papers – none of which were single-authored – were written by 52 different researchers, with an average of 4.92 authors per publication. This reflects a high collaborative profile, further supported by a notable international co-authorship rate of 53.85%. The corpus of articles has an average annual growth rate of 20.09%, indicating increasing interest and relevance of the topic in recent years. In total, 467 references were cited across the studies, with an average of 4.08 citations per paper.

Figure 2 shows the number of articles published throughout the entire period. No publications were recorded in 2020 or 2021, while 2022 had a peak with three published papers. In comparison, the years 2018, 2019, and 2023 registered two, one, and one articles, respectively. Notably, the number of publications greatly increased in 2024, totalling six articles. This growth may be seen as an indicator of the rising interest in the topic in the last year.

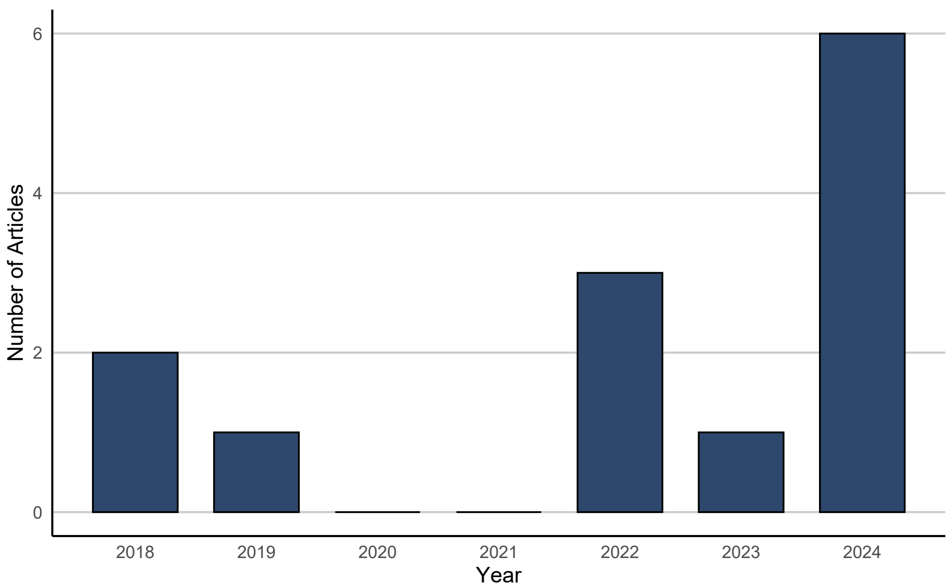

Figure 2. Annual publication history.

Among the countries of first authors publishing on the application of GAMLSS in the sports context, as shown in Figure 3, China stands out with three papers, followed by Brazil, Germany, and the United Kingdom, each with two articles. Together, these four countries account for roughly 69.2% of total publications, indicating a concentration of scientific output, albeit with notable international participation. In addition, Chile, France, Spain, and the United States each contributed one article.

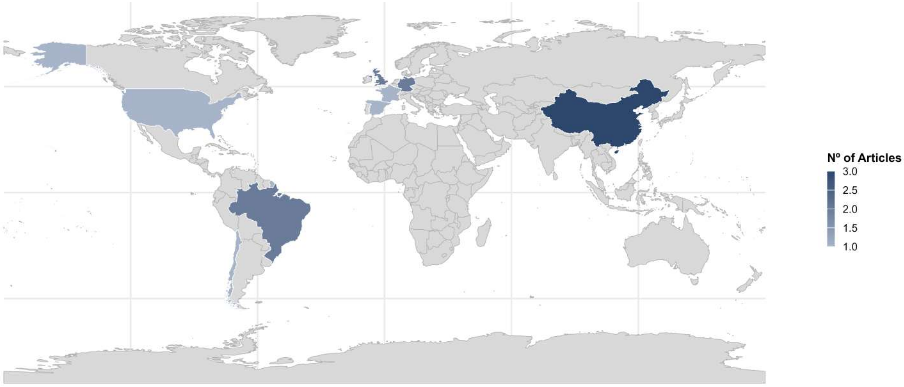

Figure 3. Country-wise distribution of publications based on first author affiliation.

As indicated in Figure 4, the study by Datson *et al.* (2022) is the most cited, with a total of 12 citations. It is followed by the articles by Nakamura *et al.* (2019) and Ötting *et al.,* 2018, with nine and eight citations, respectively. These studies stand out for their impact within the analysed field. The remaining papers included in this review, but not represented in Figure 4, each received only one citation.

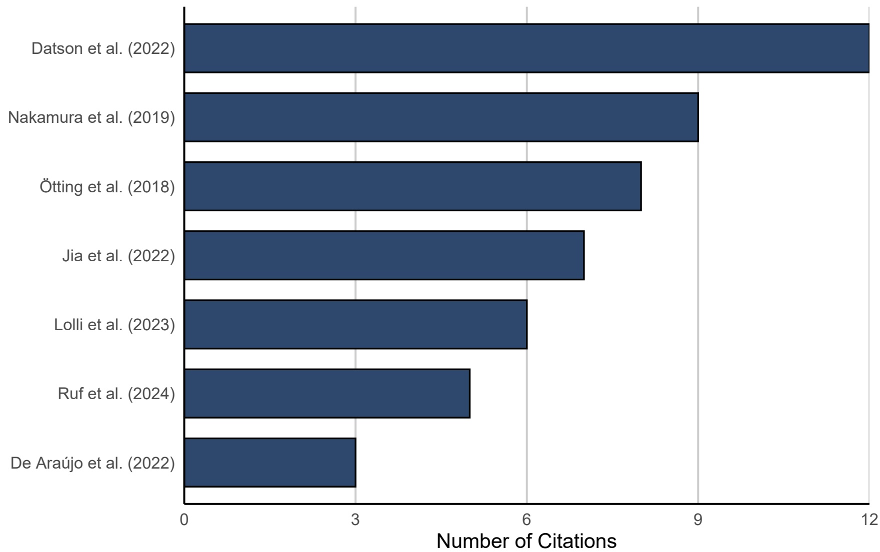

Figure 4. Citations per article.

Figure 5 displays the relationship between authors and their cooperation network, where the strength of the connections represents the number of co-authored publications. It is noteworthy that the network is based solely on the publications included in the systematic review and reflects collaboration patterns only within the scope defined by the review protocol. Two potential research or collaboration groups were identified: the first comprising the Bu T., Wan B., and Zhang Y., and the second including Gregson W. and Lolli L. The author Bu T. is the one with most publications, appearing in three different papers, while the other four mentioned authors contributed to two. The remaining authors contributed to a single article in the analysed set.

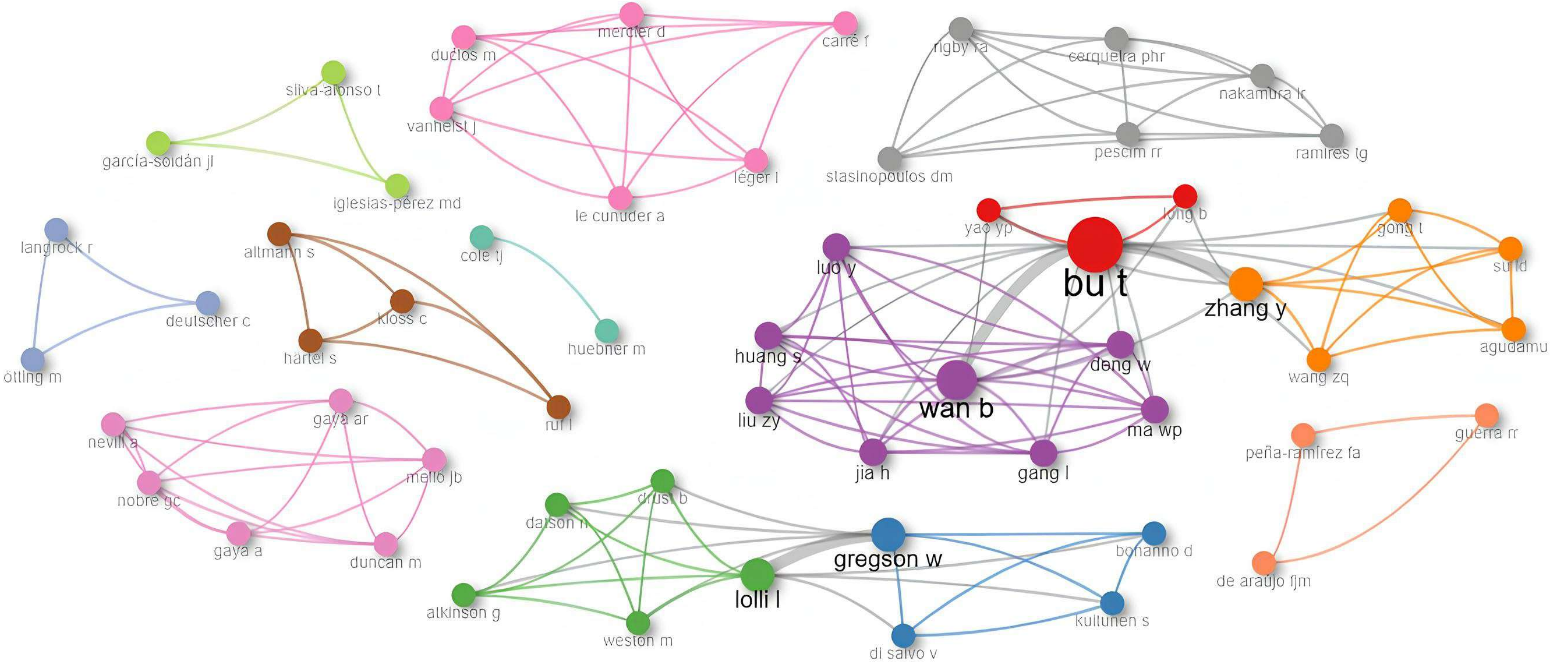

Figure 5. Author collaboration network.

In the analysis of author distribution per article, one paper was published by two authors (Huebner & Cole, 2024), and three were published by three authors each (Ötting *et al.,* 2018; Silva-Alonso *et al.,* 2018; De Araújo *et al.,* 2022). The four-author category was represented by a single study (Ruf *et al.,* 2024), while two papers had five authors (Lolli *et al.,* 2023; Yao *et al.,* 2024). Five articles included six authors (Nakamura *et al.,* 2019; Wang *et al.,* 2024; Datson *et al.,* 2022; Mello *et al.,* 2024; Vanhelst *et al.,* 2024), and one paper (Jia *et al.,* 2022) had nine authors, the highest number of contributors. This distribution highlights a predominance of collaborations involving three to six authors in the reviewed papers.

The keywords used by the authors are displayed in Figure 6. As expected, GAMLSS is the most prominent and frequently used term, highlighting the importance of this statistical model in the examined articles. Following that, the terms football, player tracking, and speed appear twice, indicating an emphasis of research on physical performance and athlete movement analysis, especially in football. Several other keywords appeared only once, including adolescence, age-related reference ranges, agility, aging human, anthropometry, and beta distribution. These terms emphasise the thematic diversity of the studies, which range from the development of growth curves and age-based references to issues related to ageing, body composition, and specific statistical methodologies.

Figure 6. Word cloud of the most frequently used keywords.

Table 2 presents the knowledge matrix, detailing the article names, the journals in which they were published, types of response variables, statistical distributions used, sample characteristics, and the main focus of each paper included in the final review portfolio.

Table 2. Knowledge matrix

ID	Author (year)	Title	Journal	Response type	Distribution	Sample	Focus
1	Ötting *et al.* (2018)	Integrating multiple data sources in match-fixing warning systems.	Statistical Modelling	Continuous and Discrete	Log-normal; bivariate Poisson	3,219	Application
2	Silva-Alonso *et al.* (2018)	Percentile curves and reference values for 2000-m rowing ergometer performance time in international rowers aged 14-70 years	Journal of Human Sport & Exercise	Continuous	Box-Cox power exponential	15,420	Application
3	Nakamura *et al.* (2019)	A new continuous distribution on the unit interval applied to modelling the points ratio of football teams	Journal of Applied Statistics	Continuous	Logit and beta family	234	Simulation and Application
4	Datson *et al.* (2022)	Reference values for performance test outcomes relevant to English female soccer players	Science and Medicine in Football	Continuous	Box-Cox family	479	Application
5	De Araújo *et al.* (2022)	The Burr XII quantile regression for salary-performance models with applications in the sports economy	Computational and Applied Mathematics	Continuous	Burr XII; normal; Weibull; log-normal	86	Simulation and Application
6	Jia *et al.* (2022)	Chinese physical fitness standard for campus football players: A pilot study of 765 children aged 9 to 11	Frontiers in Physiology	Continuous	Box-Cox family; skew t; skew power exponential; sinh-arcsinh	765	Application
7	Lolli *et al.* (2023)	Age-related reference intervals for physical performance test outcomes relevant to male youth Middle Eastern football players	International Journal of Sports Physiology and Performance	Continuous	Box-Cox family	441	Application
8	Huebner & Cole (2024)	Ranking performances of Olympic-style weightlifters adjusted for body mass on the same scale for both sexes: A novel approach	Journal of Sports Sciences	Continuous	Box-Cox family	2,697 and 7,710	Application
9	Mello *et al.* (2024)	Growth charts of Brazilian youngs: 20-years data of 95,000 children and adolescents from “Projeto Esporte Brasil”	Ciência & Saúde Coletiva	Continuous	Box-Cox normal	95,470	Application
10	Ruf *et al.* (2024)	Normative reference centiles for sprint performance in high-level youth soccer players: the need to consider biological maturity	Pediatric Exercise Science	Continuous	Box-Cox family	1,745 and 824	Application
11	Vanhelst *et al.* (2024)	Sport participation, weight status, and physical fitness in French adolescents	European Journal of Pediatrics	Continuous	Box-Cox power exponential	8,084	Application
12	Wang *et al.* (2024)	Factors driving FIFA world cup 2022 viewership ratings in mainland China: marketing outlooks for FIFA world cup 2026	Frontiers in Sports and Active Living	Continuous	Beta family; logit-normal; simplex	64	Application
13	Yao *et al.* (2024)	In quest of China sports lottery development path to common prosperity in 2035	Plos One	Continuous	Truncated distributions, gamma and inverse gamma	Not specified	Application

As we can see from Table 2, the vast majority of studies considered continuous response variables, with only one using a discrete response (Ötting *et al.,* 2018). The most widely used distributions were those belonging to the Box-Cox family. Furthermore, the majority of articles focused on using GAMLSS to develop reference curves in sports contexts and/or those related to physical performance. Notably, only two papers (Nakamura *et al.,* 2019; De Araújo *et al.,* 2022) incorporated simulations procedures in addition to practical applications, highlighting the predominance of an empirical approach in the selected literature.

The bibliometric analysis outlined the landscape of publications on the use of GAMLSS in sports contexts between 2018 and 2024. Table 2 highlights articles published in a variety of journals, including applied statistics (e.g., Journal of Applied Statistics), sports medicine (e.g., Science and Medicine in Football), and physiology (e.g., Frontiers in Physiology), among others. The scarcity of studies involving discrete variables, as well as the limited use of simulation procedures, suggest opportunities for methodological diversification. These patterns, in line with the most frequent terms in the word cloud (Figure 6), show that GAMLSS is mostly used for skill measurement and performance analysis, reflecting the literature's alignment with practical challenges in sports. These findings not only map the current state of the art, but also indicate future trends, such as the potential consolidation of collaborative research networks and the methodological expansion of GAMLSS into new sporting domains.

Despite the growth in publications seen in Figure 2, the bibliometric analysis reveals that GAMLSS are relatively underexplored in the sports science literature. This limited representation may be explained by at least two reasons. First, a significant proportion of research in this field focuses on average effects, for which simpler frameworks such as GLM and GAM are frequently suitable and more familiar to applied researchers. Second, the GAMLSS framework involves greater statistical and computational complexity, demanding more statistical expertise and more careful model specification, which may limit its widespread use. In this regard, the development and dissemination of more practical tutorials tailored to sports-related data may help facilitate the broader use of GAMLSS in this domain. Nonetheless, the increasing number of recent papers indicates a growing awareness of the advantages of GAMLSS in contexts where modelling distributional aspects beyond the mean is essential.

3.2 Metasynthesis

This section presents a qualitative analysis of the selected papers, taking into account – whenever available – information regarding the study objectives, statistical methods used, the results achieved, limitations reported, and application contexts. Table 3 summarises the articles by theme, article ID previously assigned in Table 2, application context, and response variables modelled. This is followed by a detailed discussion of each article in the sequence determined by the thematic classification in Table 3.

Group 1 (Table 3) contains a single article –– Article ID 1. Ötting *et al.* (2018) proposed a method to detect match-fixing in football by combining betting odds and volumes using two modelling strategies. In the first, a GAMLSS based on the log-normal distribution was fitted to betting volumes using gradient boosting. In the second, a bivariate Poisson GAMLSS was employed to model the number of goals scored by both sides, also fitted via boosting (Hofner *et al.,* 2016), from which the corresponding odds were derived. The results showed that jointly modelling odds and volumes improves the capacity to detect manipulated matches when compared to individual analyses of these variables. Among the main limitations, the authors highlighted the exclusive use of pre-match data, restriction to a single betting platform, and the absence of dynamic components in the model. For future research, they suggested the incorporation of additional covariates, multilevel models (for betting volume), and dynamic approaches such as Markov-switching GAMLSS (Adam *et al.,* 2022), to capture latent and temporal effects in live betting.

Table 3. Overview of GAMLSS applications in sports across different contexts

Theme	Article ID	Application context	Response variable
Group 1: Fraud detection	1	Sports betting	Betting odds and volumes
Group 2: Physical performance profiles (Percentile estimation and reference curves)	2, 4, 6, 7, 9, 10, 11	Physical performance (rowing, football, sprints)	Time, speed, vertical jump, Yo-Yo Intermittent Recovery Test Level, flexibility, agility, maximal aerobic speed, height, weight, and body mass index
Group 3: Development of statistical distributions	3, 5	Modelling of proportions and salaries	Proportions, salaries
Group 4: Performance comparison	8	Olympic weightlifting	Total weight lifted
Group 5: Economic and behavioural analysis	12, 13	Sports audience, lottery sales	Audience, sales

Group 2 (Table 3), the largest identified, comprises Article IDs 2, 4, 6, 7, 9, 10, and 11 and relates to research on the fitting of reference curves. Using GAMLSS based on the Box-Cox power exponential (BCPE) distribution, Silva-Alonso *et al.* (2018) –– Article ID 2 — created such curves for the 2000-metre maximal rowing test that was completed by males and females in the lightweight and heavyweight categories between the ages of 14 and 70. Within each sex, the data showed statistically significant, though expected, performance differences between lightweight and heavyweight rowers. Within the same weight category, however, moderate to large differences were detected between males and females. Performance followed a two-phase trajectory: a rapid exponential increase from ages 14 to 18-25, followed by a linear decline up to age 70. Significant differences between the extreme percentiles were observed, particularly between the ages of 20 and 40, and among heavyweight females. The main limitations acknowledged by the authors included the cross-sectional design, fewer observations among people over 55 years old, and less accurate fitting at the upper percentiles. The authors emphasised the need for further research on ergometer rowing performance to validate the proposed centiles and investigate their relevance in defining optimal health benchmarks across other populations.

Datson *et al.* (2022) –– Article ID 4 — developed age-related reference curves for physical performance variables of elite female football players. The measured outcomes were 5-m sprinting, 30-m sprinting, countermovement jump (CMJ), and the Yo-Yo Intermittent Recovery Test Level 1. GAMLSS based on the Box-Cox family of distributions were employed for modelling. Performance in all physical tests improved nonlinearly with chronological age, reaching a plateau around the age of 25. Among the limitations reported, the authors cited a selection bias due to the inclusion of players from a single national programme, the use of chronological age alone rather than considering the biological age, a limited sample size, and that the menstrual cycle phase — which could be an important factor in the study –– was not considered in their analysis. Furthermore, authors argue that consolidated methods for estimating confidence intervals for mixed longitudinal data are still lacking.

Jia *et al.* (2022) –– Article ID 6 –– established physical fitness standards for school football players aged 9 to 11 in China by estimating percentiles using GAMLSS, based on the sinh-arcsinh (SHASH) and the Box-Cox family of distributions. With the exception of the flexibility test, overweight and obese players scored significantly worse in agility, speed, and lower body power tests. The authors acknowledged some limitations, including a small sample size and variability in field test procedures. For future research, they recommend increasing the sample size, including both the year and month of birth to allow more precise age-based modelling, standardising the sequence and intervals between tests, and using digital technologies for data storage and analysis.

Lolli *et al.* (2023) –– Article ID 7 — used GAMLSS based on the Box-Cox family of distributions to create reference intervals for CMJ and maximal aerobic speed (MAS) in football players from the Middle East, specifically in Qatar. Performance scores in sprint and CMJ tests increased monotonically and nonlinearly with chronological age, reaching a plateau after the age of 16. In contrast, the median MAS score increased substantially up to approximately 14.5 years, with the nonlinear trend levelling at older ages. Among the limitations, the authors highlighted selection bias, a relatively restricted age range which limits the generalisability of the findings, a lack of biological maturation data, and the absence of established methods for estimating unbiased confidence bands.

Mello *et al.* (2024) –– Article ID 9 –– developed growth curves for height, weight, and body mass index (BMI) for Brazilian children and adolescents aged 6 to 17, using GAMLSS based on the Box-Cox Cole and Green distribution. The study employed a cross-sectional design with data collected over a 20-year period as part of the Projeto Esporte Brasil (Brazil Sport Project), a repeated cross-sectional surveillance study, across different regions of Brazil. Weight-for-age increased linearly up to age 10 in both sexes, with Brazilian percentiles exceeding World Health Organisation (WHO) guidelines in the upper percentiles. Height-for-age in males increased linearly until age 13, followed by an acceleration phase until 16.6 years; in females, similar acceleration happened until 14.9 years, stabilising thereafter. BMI increased from 15.8 to 22 kg/m2 in males and from 15.8 to 21 kg/m2 in females, with Brazilian curves surpassing WHO benchmarks during puberty, except in the lowest percentiles, which were similar to or slightly below WHO standards before age 13. The authors identified two main limitations: potential time-related effects on the variables and the overrepresentation of participants from the Southeast and South regions of Brazil.

Ruf *et al.* (2024) –– Article ID 10 — established reference curves for 5- and 30- metre sprinting performance in highly trained youth football players, taking into account both chronological age and biological maturity status (determined by skeletal age). Additionally, the study compared the impact of mature status on performance by evaluating individual scores derived from centile curves for each type of age. GAMLSS based on the Box-Cox family were employed. Results showed that early maturing athletes showed small to moderate performance reductions when compared to skeletal age-based centiles, whereas late maturers had moderate to large improvements, indicating that skeletal age provides a more accurate representation of performance to maturity. The authors noted many limitations, including a lack of control for confounding variables such as training exposure and genetic factors, skeletal age categorisation, and the retrospective nature of the data, which hindered the precise estimation of confidence intervals for the centiles. Future research directions include creating reference curves in other countries for cross-cultural comparison, implementing standardised sprinting measurement methodologies, and using continuous rather than categorised maturity indicators. Longitudinal studies are also recommended to improve the accuracy of the estimates.

Vanhelst *et al.* (2024) –– Article ID 11 — used GAMLSS based on the BCPE distribution to investigate the relationship between sports participation, weight status, and physical fitness in 8,084 adolescents, as well as to establish sex- and age-specific percentiles for physical fitness. The findings revealed that males performed better in the 20-metre shuttle run and handgrip strength tests. Adolescents who were overweight or obese had higher handgrip strength but lower cardiorespiratory performance. Participation in organised sports (at school or in clubs) was associated with greater cardiorespiratory fitness, with no significant differences observed in muscular strength. Among the limitations, the authors highlighted the cross-sectional study design and the lack of a nationally stratified sample.

Group 3 (Table 3) contains two articles that propose new statistical distributions within the GAMLSS framework and include both simulation studies and applied analyses: Article IDs 3 and 5. Nakamura *et al.* (2019) –– Article ID 3 –– proposed a new distribution for modelling rates and proportions derived from a logit transformation of the SHASH distribution, providing a robust alternative to beta regression (Ferrari & Cribari-Neto, 2004). Through simulations, the authors assessed the performance of the estimators across different scenarios and sample sizes. The application of the proposed model to real-world data on football performance in four European leagues demonstrated its effectiveness, showing better fit than earlier models in the literature. The authors noted, however, that the study is observational in nature, which limits causal inference, while nonetheless emphasising the model's usefulness as a robust tool for descriptive and predictive analyses.

De Araújo *et al.* (2022) –– Article ID 5 — developed a quantile regression model based on the reparameterised Burr XII distribution within the GAMLSS framework. To illustrate the application of the model, they analysed salary data from players in the Western division of the American League of the Major League Baseball, with the goal of identifying main factors affecting player salaries by considering both team-related and individual performance characteristics. The proposed model outperformed models based on the normal, Weibull, and log-normal distribution, demonstrating its effectiveness in modelling skewed and heavy tailed data. For future research, the authors suggest incorporating nonlinear structures of the Cobb-Douglas type into the analysis of production functions.

Group 4 (Table 3) features only Article ID 8. Huebner & Cole (2024) introduced a novel method and developed a web-based application to compare men's and women's Olympic weightlifting performances. The methodology employed GAMLSS, fitted separately for each sex using Box-Cox distributions based on z-scores (Hester *et al.,* 1990). The temporal analysis revealed that the mean z-scores have significantly decreased over the years, accompanied by a parallel reduction in variability. However, the average curves remained stable within one standard deviation since 2000 and 0.4 standard deviation since 2010, indicating that the model remains sufficiently robust for future applications, even in the face of small trend shifts. Limitations include the exclusion of older athletes (who do not compete in International Weightlifting Federation competitions), the potential underreporting of doping cases in earlier years (due to less stringent testing protocols), and the absence of age-specific adjustments. Nevertheless, the authors point out that GAMLSS allow for future adaptations, such as the inclusion of covariates like age, thereby expanding their applicability.

The last group (Table 3) comprises Article IDs 12 and 13, which are aligned with economic and behavioural analysis. Wang *et al.* (2024) –– Article ID 12 — investigated the factors influencing television viewership of the 2022 FIFA World Cup in mainland China, with the goal of identifying consumer behaviour trends and suggest marketing strategies for the 2026 edition. The features included factors such as tournament stage, primetime broadcasting, games in weekends the presence of Asian teams (South Korea or Japan), FIFA ranking, match uncertainty, and the market value of the teams, among others. To model the audience data, seven different distributions were tested, with the GAMLSS based on the logit-normal distribution proving to be the most appropriate. The authors acknowledge some limitations, including the restricted applicability to the Chinese context, the small number of observations, the absence of additive smoothing terms in the model, and the impact of changing media consumption patterns (e.g., the rise of pay-per-view and short-form videos). Furthermore, the authors cite that political tensions between China and the United States could affect future commercial analyses within the country.

Yao *et al.* (2024) –– Article ID 13 –– used a GAMLSS-based macroeconomic forecasting model considering the gamma distribution to forecast sports lottery sales in China up to 2035. The model included demographic and consumer behaviour variables, analysing growth trends in the sector, taking into account population ageing, socioeconomic changes, and sporting events, especially the FIFA World Cup. The authors projected that the population aged 15 to 64 will peak in 2026, followed by a decline until 2035. This demographic transition (excluding the impact of the FIFA World Cup) forecasts a stability of sales in 2027 and a further decrease after 2032. In other words, sustained urbanisation is predicted to drive the market until 2027, followed by a slowdown. Finally, events such as the FIFA World Cup were identified as temporary sales boosters.

4. Conclusion

This paper aimed to conduct a systematic review on the application of distributional regression models, or generalised additive models for location, scale, and shape (GAMLSS), in the analysis of sports science data. The methodological approach, based on the systematic search flow (SSF) framework, ensured rigour and transparency throughout the review process. The clear and structured organisation of the study enabled the analysis to consistently address the research questions that guided the entire methodological development of the review.

It was found that these models have been used in a wide range of contexts, including sports betting, physical fitness, proportion modelling, sports viewership, lottery sales, and weightlifting. GAMLSS are used in sports for various purposes, such as detecting match-fixing in football, analysing the relationship between salary and performance in baseball players, developing comparative scales adjusted for body mass, and forecasting sports lottery sales. This broad scope highlights the relevance of GAMLSS both for academic research and practical decision-making in sport.

Regarding the types of data modelled, a predominance of continuous variables was observed, with 12 out of the 13 articles focusing on continuous responses including speed, height, and television audience, and only one article used discrete variable, indicating a significant gap in the application of GAMLSS to count or categorical data within sports context. The studies covered a broad range of populations, from elite athletes to schoolchildren, further demonstrating the breadth of methodological applications. Distributions belonging to the Box-Cox family were the most common, considered in eight articles. Notably, two new distributions were proposed and applied in specific contexts, such as modelling athlete salaries and proportions of sports outcomes.

The qualitative metasynthesis conducted in this review highlighted GAMLSS as an advantageous framework in scenarios requiring the modelling of multiple distributional parameters and offering parametric flexibility to account for covariates that directly affect different types of complex data. These features outperform traditional statistical methods, reinforcing the usefulness of GAMLSS in complex applications.

Nonetheless, persistent gaps identified by the authors include sampling bias and the lack of consensus on testing protocols and data collection standards, which hinder cross-study comparisons. The bibliometric analysis also revealed a growing research field – with an average annual growth rate of 20.09% in publications – although still in its early stages. These limitations and/or characteristics highlight future research opportunities for improving the application of GAMLSS in sports data, expanding its scope to discrete variables, strengthening methodological standards, and, from a practical perspective, encouraging the development of applied tutorials that may help lower technical barriers and facilitate broader adoption by applied researchers.

Thus, the aim of this study was achieved, as it was possible to observe that the application of GAMLSS in the sports context has been the subject of growing and diverse research across different countries. For future reviews, it is recommended to expand search strategies by incorporating additional keywords to investigate applications in hitherto unmapped situations, to incorporate complementary techniques, and to prioritise studies in emerging modalities such as e-sports.

Acknowledgments

This work was supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais — FAPEMIG (Brazil).

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Data curation:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Formal analysis:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Investigation:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Methodology:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Software:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Resources:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Supervision:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Validation:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Visualization:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Writing original draft:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R. **Writing-review and editing:** ROSA, L. F.; KONRATH, A. C.; NAKAMURA, L. R.

References

1. Abalo-Núñez, R, Gutiérrez-Sánchez, A, Pérez, M. C. I. & Vernetta-Santana, M. Injury prediction in aerobic gymnastics based on anthropometric variables. *Science & Sports* **33,** 228–236. doi:10.1016/j.scispo.2018.02.002 (2018).
2. Adam, T, Mayr, A & Kneib, T. Gradient boosting in Markov-switching generalized additive models for location, scale and shape. *Econometrics and Statistics* **22,** 3–16. doi:10.1016/j.ecosta.2021.04.002 (2022).
3. Almarashi, A. M., Daniyal, M & Jamal, F. A novel comparative study of NNAR approach with linear stochastic time series models in predicting tennis player's performance. *BMC Sports Science, Medicine and Rehabilitation* **16,** 28. doi:10.1186/s13102-024-00815-7 (2024).
4. Aria, M & Cuccurullo, C. bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics* **11,** 959–975. doi:10.1016/j.joi.2017.08.007 (2017).
5. Ayodeji, I & Abiodun, O. The effect of rivalry on sport performance: a case study of Cristiano Ronaldo and Lionel Messi. *Sports Innovation Journal* **5,** 61–77. doi:10.18060/27691 (2024).
6. Cerqueira, P. H. R., Nakamura, L. R., Pescim, R. R. & Leandro, R. A. Investigating the underlying causal network on European football teams. *Journal of Data Science* **15,** 293–312. doi:10.6339/JDS.201704_15(2).0007 (2017).
7. Datson, N, Weston, M, Drust, B, Atkinson, G, Lolli, L & Gregson, W. Reference values for performance test outcomes relevant to English female soccer players. *Science and Medicine in Football* **6,** 589–596. doi:10.1080/24733938.2022.2037156 (2022).
8. De Araújo, F. J. M., Guerra, R. R. & Peña-Ramírez, F. A. The Burr XII quantile regression for salary-performance models with applications in the sports economy. *Computational and Applied Mathematics* **41,** 282. doi:10.1007/s40314-022-01971-7 (2022).
9. Exel, J & Dabnichki, P. Precision sports science: what is next for data analytics for athlete performance and well-being optimization? *Applied Sciences* **14,** 3361. doi:10.3390/app14083361 (2024).
10. Ferenhof, H. A. & Fernandes, R. F. Demystifying literature review in the AI era: updating the SSF method with generative artificial intelligence support. *Biblios* **88,** e003. doi:10.5195/biblios.2025.1317 (2025).
11. Ferenhof, H. A. & Fernandes, R. F. Demystifying the literature review as basis for scientific writing: SSF method. *Revista ACB* **21,** 550–563 (2016).
12. Ferrari, S & Cribari-Neto, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31,** 799–815. doi:10.1080/0266476042000214501 (2004).

13. Gifford, M & Bayrak, T. A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression. *Data Analytics Journal* **8,** 100296. doi:10.1016/j.dajour.2023.100296 (2023).

14. Hastie, T & Tibshirani, R. Generalized additive models. *Statistical Science* **1,** 297–310. doi:10.1214/ss/1177013604 (1986).

15. Hester, D, Hunter, G, Shuleva, K & Kekes-Sabo, T. Bridging the gap – practical application: review and evaluation of relative strength handicapping models. *National Strength and Conditioning Association Journal* **12,** 54–57 (1990).

16. Hofner, B, Mayr, A & Schmid, M. gamboostLSS: an R package for model building and variable selection in the GAMLSS Framework. *Journal of Statistical Software* **74,** 1–31. doi:10.18637/jss.v074.i01 (2016).

17. Huebner, M & Cole, T. J. Ranking performances of Olympic-style weightlifters adjusted for body mass on the same scale for both sexes: a novel approach. *Journal of Sports Sciences* **42,** 2124–2130. doi:10.1080/02640414.2024.2423138 (2024).

18. Jain, P. K., Quamer, W & Pamula, R. Sports result prediction using data mining techniques in comparison with base line model. *Opsearch* **58,** 54–70. doi:10.1007/s12597-020-00470-9 (2021).

19. Jia, H, Wan, B, Bu, T, Luo, Y, Ma, W, Huang, S, Gang, L, Deng, W & Liu, Z. Chinese physical fitness standard for campus football players: a pilot study of 765 children aged 9 to 11. *Frontiers in Physiology* **13,** 1023910. doi:10.3389/fphys.2022.1023910 (2022).

20. Lolli, L, Gregson, W, Bonanno, D, Kuitunen, S & Di Salvo, V. Age-related reference intervals for physical performance test outcomes relevant to male youth Middle Eastern football players. *International Journal of Sports Physiology and Performance* **18,** 1283–1295. doi:10.1123/ijspp.2023-0145 (2023).

21. Makaruk, H, Starzak, M, Tarkowski, P, Sadowski, J & Winchester, J. The effects of resistance training on sport-specific performance of elite athletes: a systematic review with meta-analysis. *Journal of Human Kinetics* **15,** 135–155. doi:10.5114/jhk/185877 (2024).

22. Markopoulou, C, Papageorgiou, G & Tjortjis, C. Diverse machine learning for forecasting goal-scoring likelihood in elite football leagues. *Machine Learning & Knowledge Extraction* **6,** 1762–1781. doi:10.3390/make6030086 (2024).

23. Mello, J. B., Nobre, G. C., Gaya, A. R., Duncan, M, Nevill, A & Gaya, A. Growth charts of Brazilian youngs: 20-years data of 95,000 children and adolescents from "Projeto Esporte Brasil". *Ciência & Saúde Coletiva* **29,** e06412023. doi:10.1590/1413-81232024295.06412023 (2024).

24. Nakamura, L. R., Cerqueira, P. H. R., Ramires, T. G., Pescim, R. R., Rigby, R. A. & Stasinopoulos, D. M. A new continuous distribution on the unit interval applied to modelling the points ratio of football teams. *Journal of Applied Statistics* **46,** 416–431. doi:10.1080/02664763.2018.1495699 (2019).

25. Nelder, J. A. & Wedderburn, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society Series A: General* **135,** 370–384. doi:10.2307/2344614 (1972).

26. Ötting, M, Langrock, R & Deutscher, C. Integrating multiple data sources in match-fixing warning systems. *Statistical Modelling* **18,** 483–504. doi:10.1177/1471082X18804933 (2018).

27. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372,** 71. doi:10.1136/bmj.n71 (2021).

28. Pala, L. O. O., Nakamura, L. R., Sabe, E. M. & Ramires, T. G. Agricultural insurance and climate variability: analyzing soybean yield risks in Brazil using distributional regression models. *The Geneva Papers on Risk and Insurance – Issues and Practice,* Forthcoming (2026).

29. Pantzalis, V. C. & Tjortjis, C. *Sports analytics for football league table and player performance prediction* in *Proceedings of the 11th International Conference on Information, Intelligence, Systems and Applications (IISA)* (2020), 1–8. doi:10.1109/IISA50023.2020.9284352.

30. Patel, P. A., Gopali, R, Reddy, A & Patel, K. K. Baseball-related ocular injuries in the United States: a 10-year analysis. *Expert Review of Ophthalmology* **17,** 145–151. doi:10.1080/17469899.2022.2039629 (2022).

31. Podrigalo, L. V., Volodchenko, A. A., Rovnaya, O. A., Podavalenko, O. V. & Grynova, T. I. The prediction of success in kickboxing based on the analysis of morphofunctional, physiological, biomechanical and psychophysiological indicators. *Physical Education of Students* **22,** 51–56. doi:10.15561/20755279.2018.0108 (2018).

32. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2025). https://www.R-project.org/.

33. Ramires, T. G., Nakamura, L. R., Righetto, A. J., Carvalho, R. J., Vieira, L. A. & Pereira, C. A. B. Comparison between highly complex location models and GAMLSS. *Entropy* **23,** 469. doi:10.3390/e23040469 (2021).

34. Rigby, R. A. & Stasinopoulos, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics* **54,** 507–554. doi:10.1111/j.1467-9876.2005.00510.x (2005).

35. Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z. & De Bastiani, F. *Distributions for Modeling Location, Scale, and Shape* (CRC Press, Boca Raton, FL, 2019).

36. Roczniok, R, Stanula, A, Maszczyk, A, Mostowik, A, Kowalczyk, M, Fidos-Czuba, O & Zając, A. Physiological, physical and on-ice performance criteria for selection of elite ice hockey teams. *Biology of Sport* **33,** 43–48. doi:10.5604/20831862.1180175 (2016).

37. Roquim, F. V., Lima, R. R., Nakamura, L. R., Ramires, T. G. & Blanco, Y. A. C. Dealing with highly kurtotic count data with excess zeroes: comparing different treatments in the control of dairy cattle gastrointestinal parasites. *Brazilian Journal of Biometrics* **41,** 412–423. doi:10.28951/bjb.v41i4.646 (2023).

38. Ruf, L, Altmann, S, Kloss, C & Härtel, S. Normative reference centiles for sprint performance in high-level youth soccer players: the need to consider biological maturity. *Pediatric Exercise Science* **36,** 192–200. doi:10.1123/pes.2023-0186 (2024).

39. Sarlis, V, Chatziilias, V, Tjortjis, C & Mandalidis, D. A data science approach analysing the impact of injuries on basketball player and team performance. *Information Systems* **99,** 101750. doi:10.1016/j.is.2021.101750 (2021).

40. Silva-Alonso, T, Iglesias-Pérez, M. C. & García-Soidán, J. L. Percentile curves and reference values for 2000-m rowing ergometer performance time in international rowers aged 14-70 years. *Journal of Human Sport & Exercise* **13,** 731–742. doi:10.14198/jhse.2018.134.02 (2018).

41. Stasinopoulos, M. D., Rigby, R. A. & De Bastiani, F. GAMLSS: a distributional regression approach. *Statistical Modelling* **18,** 248–273. doi:10.1177/1471082X18759144 (2018).

42. Staunton, C. A., Romann, M, Björklund, G & Born, D. Diving into a pool of data: using principal component analysis to optimize performance prediction in women's short-course swimming. *Journal of Sports Science* **42,** 519–526. doi:10.1080/02640414.2024.2346670 (2024).

43. Swindell, H. W., Marcille, M. L., Trofa, D. P., Paulino, F. E., Desai, N. N., Lynch, T. S., Ahmad, C. S. & Popkin, C. A. An analysis of sports specialization in NCAA Division I collegiate athletics. *Orthopaedic Journal of Sports Medicine* **7,** 2325967118821179. doi:10.1177/2325967118821179 (2019).

44. Vanhelst, J, Le Cunuder, A, Léger, L, Duclos, M, Mercier, D & Carré, F. Sport participation, weight status, and physical fitness in French adolescents. *European Journal of Pediatrics* **183,** 5213–5221. doi:10.1007/s00431-024-05796-w (2024).

45. Wang, Z, Su, L., Agudamu, Gong, T, Bu, T & Zhang, Y. Factors driving FIFA world cup 2022 viewership ratings in mainland China: marketing outlooks for FIFA world cup 2026. *Frontiers in Sports and Active Living* **5,** 1282898. doi:10.3389/fspor.2023.1282898 (2024).

46. Wei, S. Deep learning algorithms for sports data analysis and prediction in sports competitions. *Applied Mathematics and Nonlinear Sciences* **9,** 1–14. doi:10.2478/amns-2024-0013 (2024).

47. Wheatcroft, E. Forecasting football matches by predicting match statistics. *Journal of Sports Analytics* **7,** 77–97. doi:10.3233/JSA-200462 (2021).

48. Yao, Y, Wan, B, Long, B, Bu, T & Zhang, Y. In quest of China sports lottery development path to common prosperity in 2035. *Plos One* **19,** e0297629. doi:10.1371/journal.pone.0297629 (2024).

49. Yue, J. C., Chou, E. P., Hsieh, M. H. & Hsiao, L. C. A study of forecasting tennis matches via the Glicko model. *Plos One* **17,** e0266838. doi:10.1371/journal.pone.0266838 (2022).

50. Zotero. *Zotero (Version 6.0)* Corporation for Digital Scholarship. Available at: https://www.zotero.org/. Accessed: 22 Jul 2025. 2024.