

are explored in penalized regression models such as the lasso, illustrating how penalized linear mixed models (LMMs) can correct for hidden confounding. In addition, the performance of this method is contrasted with that of principal components (PC)-based methods. Finally, the use of these methods is illustrated in addressing batch effects and population structure in genomic data analyses, and the computational challenges involved in fitting these models efficiently with genomic-scale data are discussed.

EO048 Room 350 SPORTS ANALYTICS
Chair: Christophe Ley
E0199: Application of statistics in sport related research: A bibliometric analysis

Presenter: **Milica Maricic**, University of Belgrade, Faculty of Organizational Sciences, Serbia

Co-authors: Marina Ignjatovic, Veljko Uskokovic

Sports analytics has been attracting the attention of experts in the fields of sports science, and medicine, as well as statistics, data analytics, and engineering. As sports statistics is a prominent field of research, especially nowadays, when the implementation of technologies in sports and the data can be collected more easily, it would be valuable to conduct a bibliometric analysis of the articles published related to sports analytics and injury prevention. To do so, a bibliometric analysis has been performed based on the data available in the SCOPUS database. Articles published in journals in English between 2013-2023 were observed. The analysis encompassed 17309 articles. The results show a steady increase in the number of articles published yearly, from 1040 articles in 2013 to 2346 articles published in 2022. Most papers are indexed in the subject area of medicine (53.9%), health professions (21.2%) and others (7.4%), which encompasses multidisciplinary, mathematics, physics and similar. Detailed bibliometric analysis will be performed, encompassing analysis per affiliation, by country, as well as by authors. The analysis will be conducted in R statistical package, using the package bibliometrix. The added value is in the presentation of how the research and publication in the field of sport analytics is changing, and providing insights into who the leading authors and institutions in the field are.

E0460: Curve clustering methods and their applications to sports analytics

Presenter: **Robert Bajons**, Vienna University of Economics and Business, Institute for Statistics and Mathematics, Austria

Co-authors: Kurt Hornik

In team sports, such as American football or European football (soccer), players naturally move on the pitch in specific trajectories. Usually, the paths of players on the pitch are determined by specific team tactics, thus interesting analyses can be derived from studying common patterns in these movements. An approach for clustering weighted curves is provided, i.e. curves which may be assigned weights at each observation of the curve, in the context of sports analytics. The weighted K-means approach is simple to implement but relies on substantial preprocessing to be applied to curves. Details of the implementation of the algorithm as well as possible extensions thereof are discussed. Finally, use cases in sports are analyzed, such as an application to pass rush routes in the NFL and the clustering of possession sequences in soccer.

E0757: Can we model the hot hand phenomenon? A Bayesian hidden Markov approach for assessing basketball team performance

Presenter: **Gabriel Calvo**, University of Valencia, Spain

Co-authors: Carmen Armero, Luigi Spezia

Belief in the hot hand phenomenon in sports is commonly assumed by both media and fans. To investigate the question in the title, the shooting performance of a professional basketball team is considered and evidence of this phenomenon is sought. In particular, a Bayesian longitudinal hidden Markov model with two connected subprocesses is developed. On the one hand, the hidden process consists of a Markov chain with two possible states for each match: cold and hot. On the other hand, the observed process follows a Bernoulli distribution with two different success parameters depending on the current state of the team. The model is applied to a real data set from the Miami Heat team during the 2005-2006 season of the USA National Basketball Association. It is shown that this model can be a powerful tool for assessing the overall performance of a team during a match, particularly for quantifying the magnitude of team streaks in probabilistic terms.

E0763: Identification of injury risk factors for professional football players: A multivariate survival tree approach

Presenter: **Jone Renteria**, BCAM - Basque Center for Applied Mathematics, Spain

Co-authors: Lore Zumeta-Olaskoaga, Eder Bikandi, Jon Larruskain, Dae-Jin Lee

The objective is to identify major risk factors associated with lower limb non-contact injuries in football players. The selected population is composed of professional female and male players from a Spanish Club across multiple seasons. Firstly, longitudinal data is collected for each player, including their exposure time on the field, medical injury history (type of injury, number of previous injuries, and days missed due to previous injuries), and periodic screening tests. These latter medical examinations are primarily conducted during the preseason period and provide a general overview of the players' functional and strength parameters in their lower extremities. Once all the data is collected and curated, the multivariate survival tree (MST) approach is applied, which aims to produce easily interpretable trees that are useful in the daily operations of a football club's medical team. This analysis can inform physiotherapists and personal trainers in designing individualized training sessions to maximize performance and minimize the athletes' injury risk.

E1486: Predicting handball games with machine learning and teams strengths statistics

Presenter: **Florian Felice**, University of Luxembourg, Luxembourg

Co-authors: Christophe Ley

A statistically enhanced learning (aka. SEL) model is presented to predict handball games. The machine learning model augmented with SEL features outperforms state-of-the-art models with an accuracy beyond 80%. It is shown how the strengths of teams are estimated, representing the SEL covariates. Different models are compared and evaluated on female clubs' data to assess their predictive capabilities. It is shown that SEL variables appear as the most important features of the model. Finally, it is shown that explainability methods can help identify important drivers of the scored goals by a team. This can be used as a new predictive tool for team coaches to adjust their strategies in view of upcoming matches.

EO215 Room 351 BAYESIAN MODELING OF TIME-SERIES DATA
Chair: Michele Guindani
E0808: Time varying autoregressive gamma shot noise model for wildfires

Presenter: **Federico Bassetti**, Politecnico Milano, Italy

Co-authors: Roberto Casarin, Matteo Iacopini

Motivated by the analysis of wildfires, a novel time-varying shot noise Cox process is proposed for modelling time series of spatial data. The model assumes that a latent sequence of autoregressive gamma random measures drives the random intensity of the Cox process. To perform inference, a Bayesian approach is employed combined with a Markov Chain Monte Carlo algorithm. Several properties of the latent sequence of time-dependent gamma random measures are derived, essential for computing moment, predictive, and pair correlation measures of the proposed shot noise process. The flexible and tractable model makes it suitable for capturing spatial patterns and temporal dynamics in forest fires. Furthermore, the approach offers a potential solution to the challenging problem of estimating global trends and seasonality using high spatial-resolution fire data from a wide region. By adopting the Bayesian approach, uncertainty is quantified in the estimates and forecasts, a critical aspect of climate-risk analysis. In the application, NASA moderate resolution imaging spectroradiometer (MODIS) data is utilized.

E0895: A Bayesian change-point analysis of vector autoregressive processes

Presenter: **Stefano Peluso**, Università degli Studi di Milano Bicocca, Italy

Co-authors: Siddhartha Chib, Antonietta Mira