The generalized Thurstone method and the investigation of its applicability, as well as improving the accuracy of the predictions it provides

Propositions of the doctoral PhD dissertation

László Gyarmati

Supervisors:

Dr. Csaba Mihálykó, associate professor

Dr. Éva Orbán-Mihálykó associate professor

Doctoral School of Information Science and Tehcnology
University of Pannonia
Department of Mathematics

Veszprém
2026

1. The significance of the research topic

In our fast-paced and ever-accelerating world, the use and application of various decision-support models and software play an ever more important role. Artificial intelligence is increasingly surrounding us and, in most cases, greatly assists us in carrying out our work and tasks.

However, Artificial intelligence performs particularly well when a large amount of data and information is available. But what happens when decision support must be carried out with only a small amount of data? In such cases, it is advisable to rely on well-established mathematical models that are capable of performing reliably even with limited data.

Pairwise comparison models are excellent for decision support and information extraction. These models are effective for information extraction because they provide a detailed, quantifiable, and verifiable picture of preferences and preference/importance relationships based on data derived from simple decision-making situations and pairwise comparisons of objects. These methods perform well even with limited data, the provided data is reliable.

There are various pairwise comparison models. Among them are methods based on pairwise comparison matrices (PCM) [16], including the Analytic Hierarchy Process (AHP [17]) associated with Saaty. This group of models currently represents the main direction of research; a wealth of theoretical results [18, 19], applications [20, 21], and references are associated with it. In PCM-based models, verbal opinions or the results of other pairwise comparisons are converted into numbers, and the resulting matrix is evaluated in various ways. The advantages of this method include the wide range of numbers that can be used, which allows for many types of decision categories, as well as ease of implementation and robustness.

Another group of models consists of Thurstone-motivated models with a stochastic background. The first publications on these models appeared a century ago, allowing only two choice options and normally distributed latent probability variables [22]. These models were further developed later on [23, 24, 25], and although they continued to appear in applications, they were eventually replaced by PCM-based methods due to the complexity of the calculations. In the case of Thurstone-motivated models, decisions are not directly converted into numbers; instead, they are interpreted as relations, and from the relational data, we are able to derive quantified strengths during evaluation, which can then undergo further processing. One of their disadvantages was that they allowed for few decision options, but this drawback can be eliminated by using multi-option models [26, 27]. Due to the complexity of the calculations, significant computational capacity is required for implementation. Thanks to advances in computer science, this is now available. Thus, the inherent advantages of the method, such as the ability to estimate probabilities regarding further comparison results and the possibility of hypothesis testing, can be exploited. However, a drawback remains: due to the complexity of the models, the theoretical foundation is less well-established. Based on known theoretical results, these methods require more complex investigations compared to PCM-based models. Another interesting question that arises is the potential relationship between the two model groups. I have not found any publications in the literature addressing this issue.

Thurstone-type models are well suited for generalization and can be applied to a wide range of decision-support and information-extraction problems. In their generalized form, they allow for multiple decision alternatives and can capture advantages observed in comparisons through the inclusion of additional parameters. By introducing appropriate weights, the relative importance of data can be adjusted, for instance, to account for temporal effects or to emphasize the expertise of evaluators. Furthermore, by applying an exponential transformation to the estimated expected values derived from additive models, these approaches can also be extended to multi-criteria decision support, similarly to PCM-based models.

2. Motivation and objectives

Although Thurstone-motivated models have a long history and are widely applicable, they are significantly more complex than PCM-based models, leaving many questions unanswered regarding these models.

One such fundamental question regarding multi-option models is what conditions the data must satisfy in order to estimate strengths from them. For PCM-based models, the necessary and sufficient condition for the evaluability of the PCM matrix is known from the work of Bozóki et al., and this is the comparison graph generated by the PCM matrix [28]. Simple examples show that for Thurstone-motivated models, this condition is not sufficient; stronger constraints are required. The underlying reason is that only relations are available as a result of the comparisons.

The necessary and sufficient condition for data evaluability in the case of the two-option logistic model (Bradley–Terry) was proven by Ford [29]. In the three-option case, a sufficient condition was provided by Davidson [30], as well as a condition proven by Orbán-Mihálykó et al. [27]. At the beginning of my research, however, the necessary and sufficient conditions for evaluability were not known even for three-option models. While consistency and inconsistency measures are central topics in PCM-based models [31, 32], the concept of data consistency had not yet been addressed for Thurstone-motivated models. Another interesting question concerns the relationship between the two model classes, both on a theoretical and empirical level.

In my work, I studied Thurstone-motivated models and the stochastic Davidson model. I investigated several questions related to these models, allowing multiple options in the decision-making process. What is the minimal dataset for which the models can be evaluated? How do they behave under small datasets? How many-option models are worth using? What is their relationship to PCM-based models? How can information extraction and prediction be improved by using higher numbers of options and incorporating potential advantages into the model? How can the models be further generalized? How can they be applied, and with what efficiency? What properties of the models can be discovered during their application?

3. Methods and tools

The study of Thurstone-motivated methods requires a broad toolkit drawn from several scientific disciplines: knowledge of results from various areas of mathematics (probability theory, analysis, graph theory, numerical analysis, and statistical methods, as well as optimization), computational techniques (GPU programming, parallelization, visualization, simulation), and, in the case of applications, domain-specific tools (e.g. NETFIT and knowledge of sports-related characteristics, etc.). Achieving the results of individual publications typically requires combining several of these fields.

The obtained results are partly theoretical, but in addition to mathematical proofs, they also strongly rely on the results of computer simulations. Parameter estimation was carried out using the maximum likelihood estimation method. This is a widely used optimization-based estimation procedure, where the objective log-likelihood function is a complex multivariate function, and both the existence of the optimum and the uniqueness of the optimal solution must also be investigated.

In this context, I generalized the conditions of data evaluability, and to illustrate that these generalizations significantly improve the ability to decide evaluability, I also applied random simulations. To explore relationships between different models, I demonstrated, beyond mathematical statements, through a large number of simulation experiments the equivalence of optimal comparison structures in PCM-based and Thurstone-motivated models. When improving prediction accuracy, I also relied on simulation results in addition to analyses of real datasets.

To carry out a large number of simulations, it was necessary to develop and implement fast and efficient numerical methods. I sought a numerical method for finding extrema that is computationally efficient, adaptable, and converges well for log-concave functions. I chose the Nelder-Mead method for general distribution functions, while I used a fixed-point iteration solution for the Bradley-Terry model. These numerical methods are more efficient than the built-in optimizers in the Matlab, R, and Python software packages.
Significant computer science knowledge is also required for the computer implementation of graph connectivity and cycle-finding algorithms. I wrote these programs myself in order to be able to run a very large number of simulations within a reasonable amount of time.

As the programming language, I chose the high-performance, high-level C# language. The developed optimization search program achieves outstanding speed thanks to a performance-optimized Nelder–Mead implementation and a high degree of parallelization. From time to time, I also applied GPU programming to further increase computational speed.

The Nelder–Mead optimizer was used not only for Thurstone-motivated models but also for the Davidson model and PCM-based methods. Since in certain cases I performed up to 10^9 simulations, an even faster solution was required. In the case of the Bradley-Terry model, it was possible to use a fixed-point iteration method based on partial derivatives for 2-, 3-, 4-, and 5-option models, which resulted in further significant speedups. For three-option models, the fixed-point iteration approach achieved on average a 50-fold acceleration, making it possible to run nearly two orders of magnitude more simulations within a similar runtime. To determine evaluability, I used my own breadth-first search algorithms, modified specifically for this problem.

The software package I developed has a modular structure in order to make it easily adaptable to different tasks. A high degree of automation was implemented throughout the system, ranging from data input to output in an appropriately formatted form.

Using simulations, I was able to gain insight into the necessary and sufficient conditions and to identify various relationships between different models. These results also helped me identify optimal comparison structures and perform various measurements in the field of information extraction under different numbers of options.

4. Theses

I have grouped the results of my research into three main categories. These concern the evaluability of data, the reliability of data, and the applications of the models.

1. **Thesis Group: Investigation of the evaluability of data**

   1.1 In the two-option Bradley-Terry model, I generalized the previously proven necessary and sufficient condition ensuring data evaluability to Thurstone-motivated models using general distribution functions in the two-option models.

   1.2 I derived the conditions under which separately evaluable datasets can, with only a small amount of connecting information, be merged into a dataset that remains evaluable in three-option Thurstone-motivated models.

   1.3 For three-option Thurstone-motivated models and the Davidson model, I specified and mathematically proved the necessary and sufficient conditions for evaluability. In doing so, we answered a question that had remained open for 70 years. Using simulations, I have shown that this set of conditions significantly improves the correct identification of evaluable data sets in certain cases compared to previously known sets of conditions.

   1.4 I examined the relationship between s- and $s+2$-option models in terms of evaluability and information extraction. For $s = 2, 3$, I mathematically proved that if the data are evaluable in the s-option model, then under natural constraints they become evaluable in the $s+2$-option model as well. Based on simulations, I demonstrated that when a sufficient number of comparisons is available, it is recommended to use the model with the larger number of options. Also using random simulations, I provided the comparison numbers at which the use of the more complex model is recommended.

The significance of the theses in Thesis Group 1:
The necessary and sufficient condition systems for evaluability in two- and three-option models provide a clear decision on whether a dataset can be evaluated within the given model. In this way, they support decisions regarding model applicability. In the case of non-evaluable datasets, they also make it possible to formulate recommendations on which pairs should be compared in order to make the data evaluable. Linking datasets with only a small amount of connecting information yields a reliable ranking, and the evaluation of this condition system typically requires fewer computational resources in most cases.

Results related to the choice of the number of options indicate that using models containing more information is advantageous compared to models with fewer options, which process less information, but only if a sufficient amount of data (pairwise comparison information) is available.

The corresponding publications are: [1, 2, 3, 4, 5].

2. **Thesis Group: Consistency and inconsistency in the two-option Bradley-Terry and Davidson models based on the concept of consistency in PCM-based models**

   2.1 We have shown that the concept of consistency, which is commonly used in PCM-based models for both complete and incomplete comparisons, can be extended to the two-option Bradley-Terry models, provided that the relevant elements of the PCM are interpreted as the ratio of the number of best decisions to the number of worst decisions. In doing so, we formulated a concept of data consistency for the 2-option Bradley-Terry model, modeled after the PCM definition of consistency. The method is based on the fulfillment of Luce's choice axiom, and thus can also be applied to the three-option Davidson model.

2.2 We formulated and proved an equivalent statement regarding data consistency, which can serve as a basis for the concept of consistency even in cases where Luce's selection axiom is not satisfied.

2.3 Through the concept of consistency, we have identified a connection between PCM-based models and stochastic models, recognising that, given consistent data, they yield the same evaluation result if the relevant elements of the PCM are interpreted as the ratio of the number of better decisions to the number of worse decisions among the given objects.

2.4 Using simulations, I demonstrated that, in the case of inconsistent data, the two-option Bradley-Terry models have the same comparison graph structures that ensure optimal information extraction as the PCM-based models. Thus, I found a non-trivial connection between the two model groups. I performed the comparisons using various metrics, under different perturbation values ensuring inconsistency, and based on random parameters and random comparison data, and obtained results of the same nature for the two model types.

The significance of the theses in Thesis Group 2: We have identified a connection between PCM-based models and the stochastic model family, which may help in identifying their common roots. An equivalent reformulation of the consistency concept established in the two-option Bradley-Terry model allows for the introduction of the concept of consistency even in stochastic models where Luce's choice axiom is not satisfied. Given the optimal information extraction graph structures, it becomes possible to design comparison graph structures as a function of the number of compared pairs in order to extract as much information as possible.

The corresponding publications are: [6, 7].

3. Thesis Group: Application of models with different numbers of options for evaluating objects and in decision support systems

3.1 I applied three-option models to the evaluation of football leagues. I ranked teams from the five strongest national football leagues (English, French, German, Italian, and Spanish) using a three-option Thurstone-motivated model, as well as AHP and neural networks. I found that in the case of national leagues, the rankings produced by all three models show a strong correlation. However, when I merged the leagues using inter-league matches from international competitions, the rankings produced by artificial intelligence became highly debatable, and the AHP-based ranking often overestimated teams from the strongest national leagues in international contexts.

In contrast, the Thurstone model produced realistic results even with a small number of connecting matches, making it suitable for evaluating complex, interconnected systems.

3.2 I applied a three-option model to evaluate match outcomes in the European Handball Champions League. Based on group-stage matches, I made predictions for the knockout phase, and these predictions outperformed those of recognized experts. I also correctly predicted the Final4 outcome using the model, despite the cancellation of several matches due to the COVID pandemic.

3.3 I investigated various Thurstone-motivated models incorporating advantage effects and time weighting for predicting sports outcomes. I compared their predictive performance using standard evaluation metrics against each other and against the best-performing method reported in the literature [33], based on results from the 2024/25 season across multiple national leagues. Parameters were estimated using match results from the previous two seasons. I found that the model incorporating both advantage effects and time weighting achieves competitive predictive performance compared to the most advanced methods currently available in the literature. In most cases, incorporating advantage and time weighting led to significant improvements.

3.4 I developed a seven-option decision-support model based on the Bradley-Terry model and applied it to sports selection. The athletes' subskills in different areas were quantified using the NETFIT assessment, while the importance of each subskill was determined based on coaches' judgments. The coaches provided data via pairwise comparisons. The model proved suitable not only for supporting sports selection but also for monitoring athlete performance and designing training plans aimed at improving specific subskills.

3.5 I performed clustering of light sources based on glare perception using a seven-option Bradley–Terry model. The study used real-world data collected by researchers from the University of Lausanne. The data were evaluated both with and without filtering inconsistent respondents. The two evaluation approaches produced very similar strength estimates, supporting the robustness of the method. After clustering, the same light sources were recommended under both filtered and unfiltered conditions.

The significance of the theses in Thesis Group 3:

Thurstone-motivated models are also highly suitable for evaluating sports results in cases where, due to the limited amount of available information, artificial intelligence methods are not effective. Even with a small amount of connecting information, they reliably construct coherent rankings. Owing to their probabilistic foundations, they are well suited for prediction tasks, and their accuracy can be significantly improved by incorporating time-based weighting and advantage handling.

In multi-criteria decision-making settings, they also provide valuable information about the relative strengths of objects under different attribute priorities. Moreover, the robustness of the evaluations makes it possible to reach sound decisions even when some evaluators provide partially contradictory data.

The corresponding publications are: [3, 8, 9, 10, 11, 12].

5. Summary and possible future research directions

During my research, I have addressed several previously unanswered questions concerning the evaluability of data, the concept of consistency, and the applicability of models, as well as their properties. Based on these results, I have published nine English-language journal articles over the past four years. Among these, two were published in D1-ranked journals, and an additional two in Q1-ranked journals. During my PhD studies, I also participated as a presenter at three English-language conferences; the related conference proceedings are included in the thesis booklet. In addition, I have co-authored two Hungarian-language journal articles, and participated in numerous Hungarian-language conferences, either as a presenter or as a co-author of presentations.

The research also raised a number of new questions.A natural question is what the necessary and sufficient conditions are for the evaluability of data in the case of $3 < s$ options. It would be useful to define the concept of data consistency in cases where Luce's choice axiom does not hold, as well as to introduce a measure of inconsistency for Thurstone-motivated models. In addition, it would be worthwhile to apply the models in new areas as well. However, the most interesting and at the same time most profound scientific question in the field of pairwise comparisons is the relationship between PCM-based and Thurstone-motivated models, which requires extensive further research.

Publications closely related to the theses

[1] Gyarmati, L., Orbán-Mihálykó, É., & Mihálykó, C. (2023). Comparative Analysis of the Existence and Uniqueness Conditions of Parameter Estimation in Paired Comparison Models. AXIOMS, 12(6), 575. http://doi.org/10.3390/axioms12060575
IF: 1.9

[2] Gyarmati, L., Mihálykó, C., Orbán-Mihálykó, É., & Mihálykó, A. (2026). Evaluability of paired comparison data in stochastic paired comparison models: Necessary and sufficient condition. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, 333(3), 852–867. https://doi.org/10.1016/j.ejor.2026.01.029.
D1 ranked; IF:6

[3] Orbán-Mihálykó, É., Mihálykó, C., & Gyarmati, L. (2024). Evaluating the capacity of paired comparison methods to aggregate rankings of separate groups. CENTRAL EUROPEAN JOURNAL OF OPERATIONS RESEARCH, 32(1), 109–129. http://doi.org/10.1007/s10100-023-00839-3
Q3 ranked; IF 2.3

[4] Gyarmati. L., Mihálykó C., & Orbán-Mihálykó, É. (2024). Algorithm for Option Number Selection in Stochastic Paired Comparison Models. ALGORITHM, 17(9). http://doi.org/10.3390/a17090410
Q2 ranked; IF: 2.1

[5] Gyarmati, L., Mihálykó, C., & Orbán-Mihálykó, É. (2024). Application of Different Option Numbers in Thurstone Motivated Models. In 2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS) Proceedings (pp. 1–6). http://doi.org/10.1109/CITDS62610.2024.10791388

[6] Gyarmati, L., Orbán-Mihálykó, É., Mihálykó, C., Szádoczki, Z., & Bozóki, S. (2023). The incomplete analytic hierarchy process and Bradley–Terry model: (In)consistency and information retrieval. EXPERT SYSTEMS WITH APPLICATIONS, 229(Part B). http://doi.org/10.1016/j.eswa.2023.120522
D1 ranked, IF: 7.5

[7] Mihálykó, C., Orbán-Mihálykó, É., & Gyarmati, L. (2024, August). Consistency and Inconsistency in the Case of a Stochastic Paired Comparison Model. In 2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/document/10791369

[8] Gyarmati, L., Orbán-Mihálykó, É., Mihálykó, C., & Vathy-Fogarassy, Á. (2023). Aggregated Rankings of Top Leagues' Football Teams: Application and Comparison of Different Ranking Methods. APPLIED SCIENCES-BASEL, 13(7) 4556. http://doi.org/10.3390/app13074556
Q2 ranked; IF:2.5

[9] Gyarmati, L., Edvy, L., Mihálykó, C., & Orbán-Mihálykó, É. (2023). Decision support for sports choices based on performance measurement using the generalized Thurstone method: the model and a case study. In 10th MathSport International Conference Proceedings 2023 (pp. 35–40).

[10] Gyarmati, L., Edvy, L., Mihálykó, C., & Orbán-Mihálykó, É. (2024). Decision support for sports selection based on performance measurement applying the generalized Thurstone method. INTERNATIONAL JOURNAL OF SPORTS SCIENCE AND COACHING. http://doi.org/10.1177/17479541241240609
Q1 ranked; IF:2.1

[11] Gyarmati, L., Mihálykó, C., Orbán-Mihálykó, É. (2025). Forecasting outcomes using multi-option, advantage-sensitive Thurstone-motivated models. FORECASTING, 7(3), 34. https://doi.org/10.3390/forecast7030034
Q1 ranked, IF:3.2

[12] Orbán-Mihálykó, É., Mihálykó, C., & Gyarmati, L. (2022). Application of the generalized Thurstone method for evaluations of sports tournaments' results. KNOWLEDGE, 2(1), 157–166. https://doi.org/10.3390/knowledge2010009

Publications not directly related to the theses

[13] Tóth-Merényi, A., Mihálykó, C., Orbán-Mihálykó, É., & Gyarmati, L. (2025). Exploring Consistency in the Three-Option Davidson Model: Bridging Pairwise Comparison Matrices and Stochastic Methods. MATHEMATICS, 13(9), 1374. http://doi.org/10.3390/math13091374

[14] Mihálykó, L. C., Mihálykóné Orbán, É., & Gyarmati, L. (2021). A Thurstone módszer általánosításai. ALKALMAZOTT MATEMATIKAI LAPOK, 38(1), 127-140.

[15] Gyarmati, L., Mihálykóné Orbán, É., & Mihálykó, C. (2021). A Thurstone módszer alkalmazása sporteredmények elemzésére a 2020/2021-es női kézilabda Bajnokok Ligája példáján [The application of the Thurstone method for evaluating sports results – Presenting on the EHF Women Handball Championship]. Testnevelés, Sport, Tudomány / Physical Education, Sport, Science, 6(1–2), 13–26.

References

[16] Bozóki, S., Csató, L., & Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, 248(1), 211–218.

[17] Saaty, T. L. (1980) The analytic hierarchy process: planning, priority setting, resource, allocation. McGraw-Hill, New-York.

[18] Bozóki, S., & Tsyganok, V. (2019). The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. INTERNATIONAL JOURNAL OF GENERAL SYSTEMS, 48(4), 362-381.

[19] Csató, L. (2019). A characterization of the logarithmic least squares method. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, 276(1), 212-216.

[20] Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, 169(1), 1–29.

[21] Temesi, J., Szádoczki, Z., & Bozóki, S. (2024). Incomplete pairwise comparison matrices: Ranking top women tennis players. JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY, 75(1), 145–157.

[22] Thurstone, L.L. (1927). A law of comparative judgment. PSYCHOLOGICAL REVIEW, 34, 273–286.

[23] Bradley, R. A. & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. BIOMETRICA, 39, 324–345.

[24] Glenn, W.A., & David, H. A. (1960). Ties in paired-comparison experiments using a modified Thurstone-Mosteller model. BIOMETRICS, 16, 86–109.

[25] Rao, P.V., & Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. JOURNAL OF AMERICAN STATISTICAL ASSOCIATION, 62, 194–204.

[26] Agresti, A. (1992). Analysis of ordinal paired comparison data. JOURNAL OF THE ROYAL STATISTICAL SOCIETY: SERIES C (APPLIED STATISTICS), 41(2), 287–297.

[27] Orbán-Mihálykó, É., Mihálykó, C., & Koltay, L. (2019). Incomplete paired comparisons in case of multiple choice and general log-concave probability density functions. CENTRAL EUROPEAN JOURNAL OF OPERATIONS RESEARCH, 27, 515–532.

[28] Bozóki, S., Fülöp, J., & Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. MATHEMATICAL AND COMPUTER MODELLING, 52(1–2), 318–333.

[29] Ford Jr, L. R. (1957). Solution of a ranking problem from binary comparisons. THE AMERICAN MATHEMATICAL MONTHLY, 64(8P2), 28–33.

[30] Davidson, R. R. (1970). On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments. JOURNAL OF AMERICAN STATISTICAL ASSOCIATION, 65, 317–328.

[31] Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. INTERNATIONAL JOURNAL OF GENERAL SYSTEMS, 47(8), 751–771.

[32] Rácz, A. (2022). Mixed-integer linear-fractional programming model and it's linear analogue for reducing inconsistency of

pairwise comparison matrices. INFORMATION SCIENCES, 592, 192–205.

[33] Lasek, J., & Gagolewski, M. (2021). Interpretable sports team rating models based on the gradient descent algorithm. INTERNATIONAL JOURNAL OF FORECASTING, 37(3), 1061–1071.