# Judges' Commentary: Momentum

Mark Arvidson
Mathematics, Physics, and Statistics
Azusa Pacific University
Azusa, CA

Patrick J. Driscoll
Systems Engineering
US Military Academy
West Point, NY
chi.decisions@gmail.com

Greg Mislick
Operations Research
Naval Postgraduate School
Monterey, CA

# Introduction

Given the abundance and accessibility of information on just about any topic of interest today, it is generally a challenge for the MCM Problem Committee to identify both an interesting dataset and a captivating problem for students choosing Problem C, which seeks insights from data. This year was no exception. Despite this, the committee coalesced on a wonderful topic that, while simply stated and easy to comprehend, possessed two characteristics:

- how to model match flow in tennis to illustrate and predict individual player performance and performance differences at any moment in a match, and
- simultaneously exposing player advantage by way of momentum and momentum swings.

As a concept, momentum is universally acknowledged by coaches, players, and observers to be a core element of individual and team competitive events such as tennis matches. Yet, what exactly is momentum in this context and how can it be quantified?

*The UMAP Journal* 45 (4) (2024) 405–413. 

# Dataset

The dataset provided to the teams represented an amazing collection of moment-by-moment information for every point after the first two rounds from all of Wimbledon's 2023 men's singles matches. Incredible advances in technology has enabled such a detailed data capture. Although still small in comparison, data of this nature are a portal into big data that clamor for data science and mathematical modeling approaches capable of uncovering important patterns that data volume obscures. These skills are of high interest to academia and industry alike and have a strong presence in undergraduate programs around the globe. Moreover, there is a high interest in sports among young people and growing interest in sports analytics. It is no surprise, then, that the number of teams selecting the 2024 Problem C vastly outnumbered the teams electing any other MCM or ICM problems *ever*. The many dimensions present in this year's data afforded teams with a wide array of modeling options whose use directly met the contest's expectations.

# Assumptions

The modeling assumptions expected by MCM judges are not just trivial statements presented to meet process expectations. Teams whose list included simplistic statements such as "We assume that players compete to win" or "We assume that the data are correct" were not looked upon favorably. Rather, mathematical statements are preferred; for instance, "We assume that the time series data for [data label] is distributed normally" with justification given perhaps as "The modeling approach requires this characteristic to hold, a point that we demonstrate as sound or well-founded later during validation." Another example would be: "We assume that the time series data chosen for inclusion in our initial model are not strongly auto-correlated. … We later demonstrate to the contrary and adjust our initial model to compensate."

# Data Cleaning

Data cleaning issues are a priority that teams needed to address. The first need was simply to evaluate the validity of the data and deal with outliers and omissions. Teams also needed to clearly document the data source of games used for comparison. If new variables are introduced, they must be clearly defined, as variable names are often insufficient descriptions. The difficulty in interpreting the data is compounded by poorly defined variables. For example, a number of teams that used linear regression did not state clearly that this method does not determine which variables have the most significant impact on the outcome or dependent variable; it can only analyze the ones provided.

# Artificial Intelligence

Perhaps as a landmark, this year's contest was the first time that teams were explicitly allowed to employ artificial intelligence (AI) resources, if they chose to do so, as long as they reported both their queries and AI's responses in a section (usually at the end of the paper). There is a good deal of controversy associated with AI use in education, mostly around the question of how to leverage its current capabilities to enhance student understanding and strengthen their cognitive skills without compromising or sacrificing their thought processes.

It is our opinion that AI can be appropriately leveraged in a manner somewhat similar to published research, to wit: "standing on the shoulders of Giants" [Newton in 1675 in letter to Robert Hooke]. Unlike research that has successfully navigated a peer-review process that facilitates quality control and imbues confidence in the results presented, AI results are unchecked and untethered in this regard. Consequently, students are best advised beforehand that they will bear the responsibility for verification in addition to proper citation, which could prove to be a time trap within the already-tight contest time limits.

The teams that used AI this year did so mainly to

- improve paragraph or sentence expression,
- check or suggest small programming code sections,
- assist them in identifying relevant literature sources, or
- suggest ways that they might begin to model the concept of momentum in tennis.

It appeared evident from the query results that none of these uses abdicated student thinking or clever mathematical modeling, nor even suggested that AI was capable of supplanting human involvement in mathematical modeling as of yet.

# Elements Considered in Judging

We focus on a select group—not an exhaustive list—of performance elements in papers, in order to provide observations and insights that might assist future MCM teams. These elements enabled triage and final judges to stratify papers into scoring categories and ultimately identify those papers that competed for Finalist and Outstanding award designations.

## General

Of the 10,000+ papers for this year's Problem C, an overwhelming majority of teams demonstrated proper use of and citation for credible research sources and how they should be used to support a modeling effort. It appears that most teams are dedicating time for this activity up front rather than rushing right into the mathematical modeling and analysis. This is commendable, since it has been a consistent judges' recommendation for many years to help teams make more efficient use of the available time.

Moreover, the quality of writing, composition, and exposition this year was amazing, given the contest's requirement for all papers to be submitted in English. Teams representing universities whose first language is not English are to be especially commended in this regard, as both triage and final judges found nearly all papers were a true pleasure to read.

Teams also by and large showed a mastery for dissecting, identifying, and organizing stated and implied tasks needed to yield a comprehensive model sufficient to meet the problem's requirements. The degree to which these tasks were accomplished formed a basis for discriminating between top papers and those that fell short. For the MCM contest, this is essentially saying that *critical differences between papers appeared to be based on modeling prowess* and not poor problem structuring.

## Approach to the Data

A unique dataset, such as this year's, is both a blessing and a curse. There is a richness to its many dimensions that nearly encompasses all potential time-series modeling approaches. Yet at the same time such richness forces many decisions to be made by teams. The first decision is whether to dive directly into data exploration or to first establish a base understanding of what the data represent and their possible relationship to the questions being asked. From a judge's perspective, either choice could yield fruitful results, depending on a team's mathematical skills. Experience generally drives teams to choose one option or the other.

### Data Exploration First

Teams that dove directly into data exploration using all of the data in a naïve software-driven approach fared poorly in terms of impressing the judges. This "firehose" approach has been—and will continue to be—discouraged by the MCM judges, mostly because evidence of modeling creativity and critical thinking—features that are a foundation for mathematical modeling—are largely absent. Instead, the approach gives an impression that teams are hoping to discover an effective modeling approach through brute force and serendipity rather than through reflection and selective choice. Moreover, this approach appears to indicate a lack of technical understanding of potential software tools, which is clearly not in the best interest of student teams.

Teams that plugged the dataset elements into multiple machine learning techniques and subsequently rationalized one of the resulting patterns as representing momentum and its swings throughout a game fell short in impressing the judges. There are subtle but important differences between machine-learning algorithms, differences that are driven by underlying assumptions and mathematical data characteristics. Whether machine learning should be used at all depends on the nuances associated with the specific intent for their development. Many machine learning approaches were used in a "black box" fashion with little discussion of why the model parameters were selected; they were also used for cross-validation or sensitivity analysis at the end to show that the choices made sense. A far better approach would have been to have chosen a single machine-learning method that could be best applied with a small number of assumptions that could later be relaxed to improve the model's fidelity.

### Basic Modeling First

In contrast, teams that began by identifying informative elements that might constitute momentum, as described in credible professional sources, communicated an effort to understand and structure a needed foundation upon which strong modeling justifications could be supported. For many high-performing teams, this resulted in effective explicit mathematical expressions of momentum that enabled visualization of a tennis match's flow and the identification and prediction of advantage swing. For example, Dynamic Time Warping (DTW) is a widely-used technique for analyzing time series, offering a powerful tool to compare sequences that may differ in speed or length.

Whether the dynamical expressions were linear or nonlinear depended on the teams' choices, which were supported by research citations. Judges viewed this approach as more aligned with the philosophy of mathematical modeling compared to teams that relied on "black box" software programs. This preference was largely because teams employing dynamical expressions could more effectively explain the construction and significance of each term as it was incorporated into the final expression.

## Choosing Data to Use

Once a team identified a modeling focus and proceeded into the data, they discovered that the large dataset presented them with an abundance of potential data elements to represent momentum and quantify its changes throughout a tennis game. How to reduce this to a manageable number emerged as the next hurdle. Teams that chose a subset without strong mathematical support did not fair well. Moreover, the traditional TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) model relies heavily on subjective weighting of indicators, which, as a result, can significantly influence the results. Therefore, principal component analysis can be an objective method to analyze the dataset and reduce dimensionality. More common, however, were teams that applied widely accepted data-dimension-reducing techniques that provided insight into the extent to which each element contributed to underlying variance, such as principal component analysis, linear discriminant analysis , singular value decomposition, autoregressive integrated moving average (ARIMA), and entropy. For example, an ARIMA model can be a reasonable and effective way for predicting the trend of match momentum. The results, coupled with appropriate assumptions, enabled teams to pick a subset of data to analyze.

In addition, successful groups used a variety of methods for analysis, such as logistic regression, cubic spline interpolation, neural networks, XG Boost, Monte Carlo, sliding window, and random forest. Even the use of Bartlett's test of sphericity can help determine if there is a high degree of internal correlation in the data. The goal with the data was to filter outliers and anomalies that would increase the variance, while preserving the momentum change trends so that nonstationary and nonlinear data values can be effectively processed. In sum, the goal of successful teams was to

- predict momentum and the occurrence of swings without reducing the dimensionality of the original dataset, and
- provide both coaches and players with a proper predictive interval.

# Common Shortfalls

Since the MCM's inception, teams have repeatedly been advised to identify and justify any (and all) modeling assumptions, as a necessary modeling process requirement. While still valid advice, evidence provided in team papers this year makes it apparent that more should be said regarding this advice, if for no other reason than to clarify what judges are looking for.

Model assumptions should not only be discussed but also rigorously assessed. Unfortunately, evaluating the appropriateness and adequacy of models is often overlooked or inadequately performed. An Outstanding paper distinguishes itself through thorough sensitivity analysis and comprehensive model testing.

Additionally, the problem underscored the significance of aligning modeling approaches with the goals of the analysis. For instance, linear models such as regression and analysis of variance, benefit from residual analysis, which helps validate results. The insights gained from identifying issues can be invaluable for decision-makers.

Ultimately, selecting and testing the right model, combined with clear explanations and well-crafted graphs, is crucial for a modeling paper to receive top marks.

# Common Pluses

## Verifying Assumptions

Unless a team is extraordinarily gifted, lucky, or has unlimited time and resources, they will inevitably face the challenge of bridging the gap between real-world phenomena and mathematical methods. To navigate this, teams often make assumptions that reduce their model's ability to capture every detail, such as the complexities of momentum. These assumptions typically stem from the mathematical requirements that justify the choice of a particular analytical method familiar to the team. To make timely progress, teams should identify facts that need to be checked and verified as they arise. While temporarily accepting unverified assumptions allows the team to move forward, verifying and relaxing these assumptions ultimately strengthens the model. Conversely, leaving them unverified weakens it.

## Actionable Recommendations

Recommendations provided by teams at the end of their papers need to be actionable advice associated with what the modeling effort accomplished. In the case of this problem, the advice was to be directed to tennis coaches. Recommendations such as "Players should practice serving…" and "Players should maintain momentum as long as possible…" communicate to judges that teams do not understand why recommendations are provided to the reader and greatly detract from the paper's quality.

## Informative Graphs

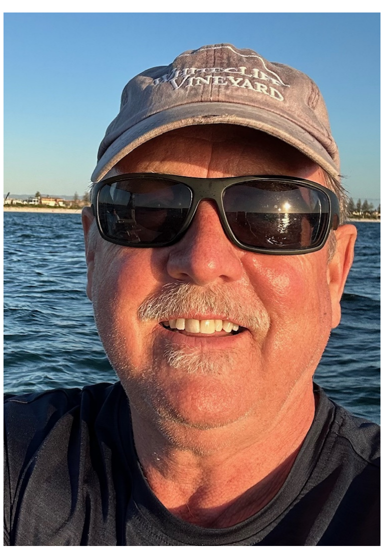
Outstanding papers present results in a way that is accessible to non-technical readers. Such papers include graphs that go beyond simple illustrations of data, such as the oscillation of momentum or player performance records. Instead, the graphs possess explanatory power, reinforcing the findings and conclusions, thereby making the results more credible. The graphs are not only easy to interpret but also enhance the overall summary.

# Conclusion

Despite the accessibility of phenomenal modeling tools and computational software, human reasoning and logic continue to be irreplaceable in the modeling process. While advanced tools can handle vast amounts of data and perform complex calculations with precision, they lack the intuitive understanding and contextual awareness that humans bring to the table. Human expertise is crucial for interpreting results, identifying anomalies, and making informed decisions based on the data. This blend of computational power and human insight ensures that models are not only accurate but also meaningful and applicable to real-world scenarios.

# About the Authors


Mark Arvidson is a professor of mathematics at Azusa Pacific University as well as an experienced teacher, educator, consultant, and presenter. He is also the Director of the Math Fellows program and trains both elementary and secondary math teachers. Dr. Arvidson did his undergraduate work at Wheaton College in mathematics; his graduate studies were completed at the Claremont Graduate University with an emphasis in mathematics education. As an applied mathematician, he has been a final judge for the MCM for five years.



Patrick J. Driscoll is the former Director of the MCM and has been an MCM judge since 1991. Graduating from the US Military Academy at West Point, he served as an officer in the US Army for 22 years, during which time he obtained an M.S. in Operations Research and also an M.S. in Engineering Economic Systems from Stanford University, along with a Ph.D. in Systems Engineering from Virginia Tech. He served on the faculty of the Dept. of Mathematical Sciences and the Dept. of Systems Engineering at West Point, and was promoted to Emeritus Professor of Systems Engineering in 2022.

Greg Mislick is a Senior Lecturer at the Naval Postgraduate School (NPS) in the Operations Research (OR) Dept. in Monterey, CA. His expertise is in life-cycle cost estimating and modeling, probability and statistics, regression analysis, data analysis, and optimization. Greg wrote the textbook *Cost Estimation: Methods and Tools* (Wiley, 2015, with Daniel A. Nussbaum) and is working on the second edition. He holds a B.S. in Mathematics from the US Naval Academy, and his primary master's degree is from NPS in OR. He has taught at NPS since August 2000 and has been an MCM judge for more than 15 years. Greg is a retired Lieutenant Colonel in the Marine Corps, serving for nearly 26 years, with four overseas deployments and 3,000 pilot flight hours in helicopters. Greg is a a former board member for the Big Sur Marathon, and a Race Director for the "JUST RUN" Kids 3k race (5,000+ kids) held twice per year in Monterey. An accomplished distance runner, Greg has completed 28 marathons with a personal best time of 2:25.21.