A Unified Server Quality Metric for Tennis

Aiwen Li, Amrita Balajee, & Audrey Kuan

Motivation

- **Existing work** focuses on match-level prediction (mostly **ELO-based**).
- **Serving stats** exist but tell an **incomplete story** of server dominance.
- There's clear room for research comparing **serving effectiveness** across men's and women's tennis.

Research Objective

Quantify server quality with a unified, predictive metric

Model point outcomes

Fit a GLMM using serving stats to predict point results.

Create a unified server-quality metric

Define a single, interpretable score capturing serve performance.

Data Overview

Data Source

Jeff Sackmann's Point-by-Point Grand Slam Datasets

Tournaments: Wimbledon, U.S. Open

Years: 2018-2019, 2021-2024

Players: Separate by gender (M/F)

Filtering: Exclude missing values and players with <20 serves

Project Pipeline

1. Clean and featurize data
2. Predict point outcomes with GLMM
3. Create server metric from model coefficients
4. Benchmark and test our metric on out-of-sample data

Features

Feature	Description	Variable Type
Avg Serve Speed	Mean speed in mph	Explanatory
SD Serve Speed	Speed variation	Explanatory
Modal Serve Location	Most common serve location by width and depth	Explanatory
Serve Location Entropy	$-\Sigma p * \log(p)$, where p = prop. of serves at each location	Explanatory
Win Percentage	Percentage of overall serving points won	Outcome
Serve Efficiency	Percentage of serving points won in 3 or fewer shots	Outcome

GLMM Setup

Y is a binary indicator of whether server *j* won point *i* for serve type *s* (first or second)

(z) indicates pre-modeling standardization

$$\text{logit} \Pr(Y_{ijs} = 1) = \beta_{0s} + \beta_{1s}\, \text{avg\_speed}_{js}^{(z)} + \beta_{2s}\, \text{sd\_speed}_{js}^{(z)} + \beta_{3s}\, \text{location\_entropy}_{js}^{(z)} + \boldsymbol{\beta}_{4s}^{\top}\, \text{OneHot}(\text{modal\_location}_{js}) + u_{js}$$

One-hot encoding each of the 10 possible serve locations

Normally-distributed random intercept for each player *j* for each serve type *s*

Creating the Server Metric

Serve Metric Formula:

$$\underbrace{\mathrm{SQS}_{js}^{(\ell)}}_{\text{log-odds}} = \underbrace{\hat{\beta}_{0s} + \hat{\beta}_{1s}\,\mathrm{avg\_spd}_{js}^{(z)} + \hat{\beta}_{2s}\,\mathrm{sd\_spd}_{js}^{(z)} + \hat{\beta}_{3s}\,\mathrm{entropy}_{js}^{(z)} + \hat{\boldsymbol{\beta}}_{4s}^{\top}\,\mathrm{OneHot}(\mathrm{modal\_location}_{js})}_{\text{measured skill}} + \underbrace{\hat{u}_{js}}_{\text{server effects}}$$

Server quality scores for each server *j*, using data from serve type *s*

Use estimated coefficients and player random effects from the GLMM

$$\mathrm{SQS}_{j,\mathrm{comb}} = \frac{n_{j1}\,\mathrm{SQS}_{j1} + n_{j2}\,\mathrm{SQS}_{j2}}{n_{j1} + n_{j2}}$$

To combine metric from first and second serves, take a weighted average based on number of first or second serves (*n*) that player *j* hit.

Out-of-Sample Testing

Separate OLS models for each dataset:

Predictors: SQS or wELO (weighted ELO)

Outcomes: serve efficiency or win percentage

Record from each testing model:

Prediction accuracy: RMSE

Strength of relationship: Correlation, p-value

Model Performance: Out-of-Sample Testing

Table 1: Out-of-sample GLMM server metric performance for Wimbledon males.

Serve Type	Outcome	Predictor	n	RMSE	Correlation (r)	p-value
First	Serve efficiency	SQS	104	1.00	0.49	9.5×10^{-8}
First	Serve efficiency	wELO	104	1.31	0.14	0.16
First	Win percentage	SQS	104	1.20	0.28	0.004
First	Win percentage	wELO	104	1.17	0.31	0.0015
Second	Serve efficiency	SQS	66	1.29	0.15	0.23
Second	Serve efficiency	wELO	66	1.34	0.09	0.49
Second	Win percentage	SQS	66	1.17	0.30	0.013
Second	Win percentage	wELO	66	1.15	0.32	0.008
Combined	Serve efficiency	SQS	129	1.14	0.34	7.9×10^{-5}
Combined	Serve efficiency	wELO	129	1.35	0.08	0.35
Combined	Win percentage	SQS	129	1.30	0.14	0.10
Combined	Win percentage	wELO	129	1.29	0.16	0.064

Wimbledon males: SQS better for predicting serve efficiency (higher correlations, lower RMSE)

Model Performance: Comparing Across Tournaments

Serve Efficiency Prediction Results

Dataset	Best Predictor	r	RMSE
Wimbledon (M)	GLMM-based	0.34	1.14
Wimbledon (F)	wELO	0.35	1.13
US Open (M)	GLMM-based	0.14	1.31
US Open (F)	GLMM-based	0.43	1.07

> ***GLMM-based server metrics more predictive on grass courts, but explains the most variability for U.S. Open females.***

Win Percentage Prediction Results

Dataset	Best Predictor	r	RMSE
Wimbledon (M)	wELO	0.16	1.29
Wimbledon (F)	wELO	0.43	1.06
US Open (M)	wELO	0.28	1.19
US Open (F)	wELO	0.32	1.16

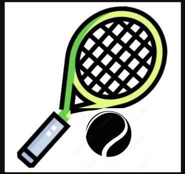
> ***Holistic player ratings (weighted ELO) better for predicting overall point outcomes.***

Interpreting the Results

Key Findings

1. **Serve Efficiency:** Server Quality Score outperforms wELO on serve-specific outcomes
2. **Win %:** wELO is more predictive for overall point wins
3. **Implications:** SQS and ELO-based metrics complementary for decomposing player skill

Conclusions



Summary Points

- Built **a server metric** that isolates the serve's contribution to point outcomes
- **Outperforms wELO** on serve efficiency prediction
- **Complements** ELO-based metrics

Impact & Applications

- **Coaching:** Targeted training based on server quality scores
- **Scouting:** Identify over and under-performing servers
- **Analytics:** Improved point prediction models
- **Broadcasting:** Deeper insights for commentator

Limitations & Future Work

Limitation	Future Extension
Only Wimbledon & U.S. Open	Add clay and indoor surfaces
Focuses exclusively on serves	Add return and rally metrics
Points treated independently	Integrate point importance & fatigue
Limited feature set	Incorporate tracking or biomechanics

Thank you!!

Acknowledgements:

- Thank you to our advisors Jonathan Pipping and Abraham Wyner for their invaluable feedback throughout the research process!
- Thank you to the Wharton Sports Analytics and Business Initiative (WSABI) for supporting this work through their summer research program.

References

Angelini, G., Candila, V., & De Angelis, L. (2022). Weighted ELO rating for tennis match predictions. ,297 (1), 120–132. Retrieved 2025-08-17, from https://www.sciencedirect.com/science/article/ pii/S0377221721003234 doi: 10.1016/j.ejor.2021.04.011
Barnett, T., & Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches., 16 (2), 113–120. Retrieved 2025-08-19, from https://doi.org/10.1093/imaman/dpi001 doi: 10.1093/imaman/dpi001
Gollub, J. (2021). Forecasting serve performance in professional tennis matches. , 7 (4), 223–233. Retrieved 2025-08-17, from https://doi.org/10.3233/JSA-200345 (Publisher: SAGE Publications) doi:10.3233/JSA-200345
Klaassen, F. J., & Magnus, J. R. (2001). Are points in tennis independent and identically dis-tributed evidence from a dynamic binary panel data model. , 96 (454), 500–509. Retrieved 2025-08-19, from https://doi.org/10.1198/016214501753168217 (Publisher: ASA Website eprint: https://doi.org/10.1198/016214501753168217) doi: 10.1198/016214501753168217
Klaassen, F. J., & Magnus, J. R. (2003). Forecasting the winner of a tennis match. , 148 (2), 257–267. Retrieved 2025-08-19, from https://linkinghub.elsevier.com/retrieve/pii/S0377221702006823 doi: 10.1016/S0377-2217(02)00682-3
Kovalchik, S., & Reid, M. (2019). A calibration method with dynamic updates for within-match forecasting of wins in tennis. , 35 (2), 756–766. Retrieved 2025-08-19, from https://www.sciencedirect.com/science/article/pii/S0169207017301395 doi: 10.1016/j.ijforecast.2017.11.008
Kovalchik, S. A. (2016). Searching for the GOAT of tennis win prediction. , 12 (3). Retrieved 2025-08-19,from https://www.degruyter.com/document/doi/10.1515/jqas-2015-0059/html doi: 10.1515/jqas-2015-0059

Newton, P. K., & Aslam, K. (2009). Monte carlo tennis: A stochastic markov chain model. , 5 (3). Retrieved 2025-08-19, from https://www.degruyterbrill.com/document/doi/10.2202/1559-0410.1169/ html?srsltid=AfmBOooFr93UcHyYxPWmYsaNauIFCgp3oT-cZRFnOVudDh5ldyieLfZE &utm source= chatgpt.com (Publisher: De Gruyter) doi: 10.2202/1559-0410.1169
O'Malley, A. J. (2008). Probability formulas and statistical analysis in tennis. , 4 (2). Retrieved 2025-08-19, from https://www.degruyterbrill.com/document/doi/10.2202/1559-0410.1100/html (Publisher: De Gruyter) doi: 10.2202/1559-0410.1100
O'Donoghue, G. P., & Brown, E. (2008). The importance of service in grand slam singles tennis. , 8 (3), 70–78. Retrieved 2025-08-24, from https://doi.org/10.1080/24748668.2008.11868449 (Publisher: Routledge eprint: https://doi.org/10.1080/24748668.2008.11868449) doi: 10.1080/24748668 .2008.11868449
Sackmann, J. (2025). JeffSackmann/tennis slam pointbypoint. Retrieved 2025-07-28, from https://github.com/JeffSackmann/tennis slam pointbypoint (original-date: 2015-03-08T14:50:29Z)
Sim, M. K., & Choi, D. G. (2020). The winning probability of a game and the importance of points in tennis matches. , 91 (3), 361–372. Retrieved 2025-08-24, from https://doi.org/10.1080/02701367.2019.1666203 (Publisher: Routledge eprint: https://doi.org/10.1080/02701367.2019.1666203) doi:10.1080/02701367.2019.166620308