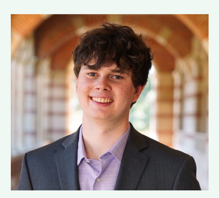
Digging Deeper: Building Volleyball Metrics with Full-Court Coverage and Context

Supported by UCLA Undergraduate Research Grant

Kevin Baer

Faculty Research Mentor: **Professor Dave Zes**
University of California, Los Angeles

Data Theory Undergraduate Student
University of California, Los Angeles
Correspondence: kevinbaer@ucla.edu



Highlights

- 6,531 ball contacts from the U19 and U21 Men's NORCECA Championships were defined and encoded into a pair of values, an input and an output of each touch.
- Analysis comes from taking the difference between the output win probability and the input win probability.
- My methodology introduces predictive and evaluative metrics, such as position rankings, pre-game win percentages, acrobatic-specific rankings, and in-point win probability graphs, offering a richer perspective for players, coaches, and viewers.

Credit and Other Related Work

- The inspiration for this work and source of features like position touch weights and win probability by game state is Chad Gordon. His blog volleydork.com is the best public resource on volleyball analytics.
- Scott Powers, Luke Stancil, and Naomi Consiglio also wrote an excellent paper using Markov chains to analyze contact success and credit-sharing. The predictive models in the bottom right use the Pythagorean winning percentage estimation on page 20 of their paper.
- Thanks to Autumn Qiu for her guidance and proofreading as always.
- Full credit and citations can be found in the QR code above.

Graphical Abstract

A

B

0
1
2
3

C

INPUT	SET Quality	OUTPUT
INS	→ ? →	OOS
.66	→ →	.50

(INS,OOS) → - 0.16

D

- EVALUATE
- ANALYZE
- PREDICT

(A) Volleyball analytics started from the wish to quantify the success of actions taken on the court. In this case, a bump would be evaluated based on whether it seemed good or bad, a surprisingly simple and effective way to represent quality. **(B)** More recently, the 4-point passing scale has been used to represent a bump by how much the setter must move to set the ball. **(C)** However, we can make three excellent improvements, firstly by looking at both the input and the output of the touch instead of just the output. Second, we can use game states to better classify and compare our "vibes" based charting. And finally, using big data empirical win probabilities for each game state allows for a more accurate depiction of the scale of the impact. **(D)** What are the benefits of this improved understanding? We can evaluate player performance, analyze points and matches accurately, and predict outcomes of points, sets, and matches in new ways.

EVALUATE

One of the main goals of sports analytics is to aid in the mathematical modeling of player ability. There are two main ways to do that. One is to quantify the actions that the player takes through the success of that action: think expected goals. The other is to find the difference that a player's actions make by quantifying the surrounding states of the game, like plus/minus or EPA. Volleyball, a discrete sport where off-the-ball actions have limited effects, lends itself particularly kindly to method two. By taking in an input and an output game state, calculations can quickly determine the effect of a touch on a point. By using thousands of touches, we can reach substantially higher levels of accuracy and approximation of actual player ability and impact. The statistic I have created is called PPA, and stands for Point Percentage Added, an average of touch impact on point win probability.

Down below in the blocking PPA chart, one can observe scores ranging from 20.4% added per block touch all the way down to -20.3% added per block touch. By analyzing blockers in this way, we can truly step beyond simple counting stats to actual analysis: note that Flaquet has 0 stuffs and 4 errors but is still higher than Urbina who has 3 stuffs and 10 errors. This might suggest that Flaquet is a more consistent blocker who gets lots of beneficial soft touches while Urbina's more aggressive approach is not paying off at a high enough rate. Beyond single skills like blocking, the data can also be used to build player profiles or weighted PPA scores. The weighting comes from the breakdown of touches per position, meaning that outside hitters are not judged on their setting, and middles are not judged on their digging. This allows for more accurate overall weighted PPAs and an understanding of which players perform the best in their roles. The table on the right demonstrates this for the u19 age group and allows you to see the best players per position, skill and overall. Evaluating players is difficult, but this model is certainly better than raw counting stats like kills or errors and can add great understanding to typical coach subjectivity.

Blocking Stats

NORCECA U19 Championship | All games between 2 of Canada/Cuba/USA | Made by Kevin Baer

PLAYER	POSITION	TEAM	COUNTING STATS: TOUCHES[1]	STUFFS	ERRORS	PPA[2]
Mikhailenko	Outside	USA	16	7	4	0.204
Payne	Middle	USA	44	14	10	0.140
Iverson	Middle	Cuba	16	6	4	0.091
Veith	Outside	Canada	23	8	7	0.063
Prentis	Middle	Cuba	16	1	2	0.055
Tate	Middle	Canada	17	5	6	0.044
McDonald	Middle	Canada	20	4	7	0.023
Soerensen	Middle	USA	31	9	15	0.009
Wright	Opposite	Canada	14	4	6	-0.026
Flaquet	Setter	Cuba	12	0	4	-0.030
Urbina	Setter	USA	24	3	10	-0.071
Keane	Opposite	USA	30	6	14	-0.081
Lamoureux	Outside	USA	10	1	5	-0.138
Medina	Opposite	Cuba	15	0	9	-0.203

[1] Minimum of 10 touches needed for inclusion
[2] Point Percentage Added: a stat that measures impact on winning a point per touch

Overall Weighted Stats

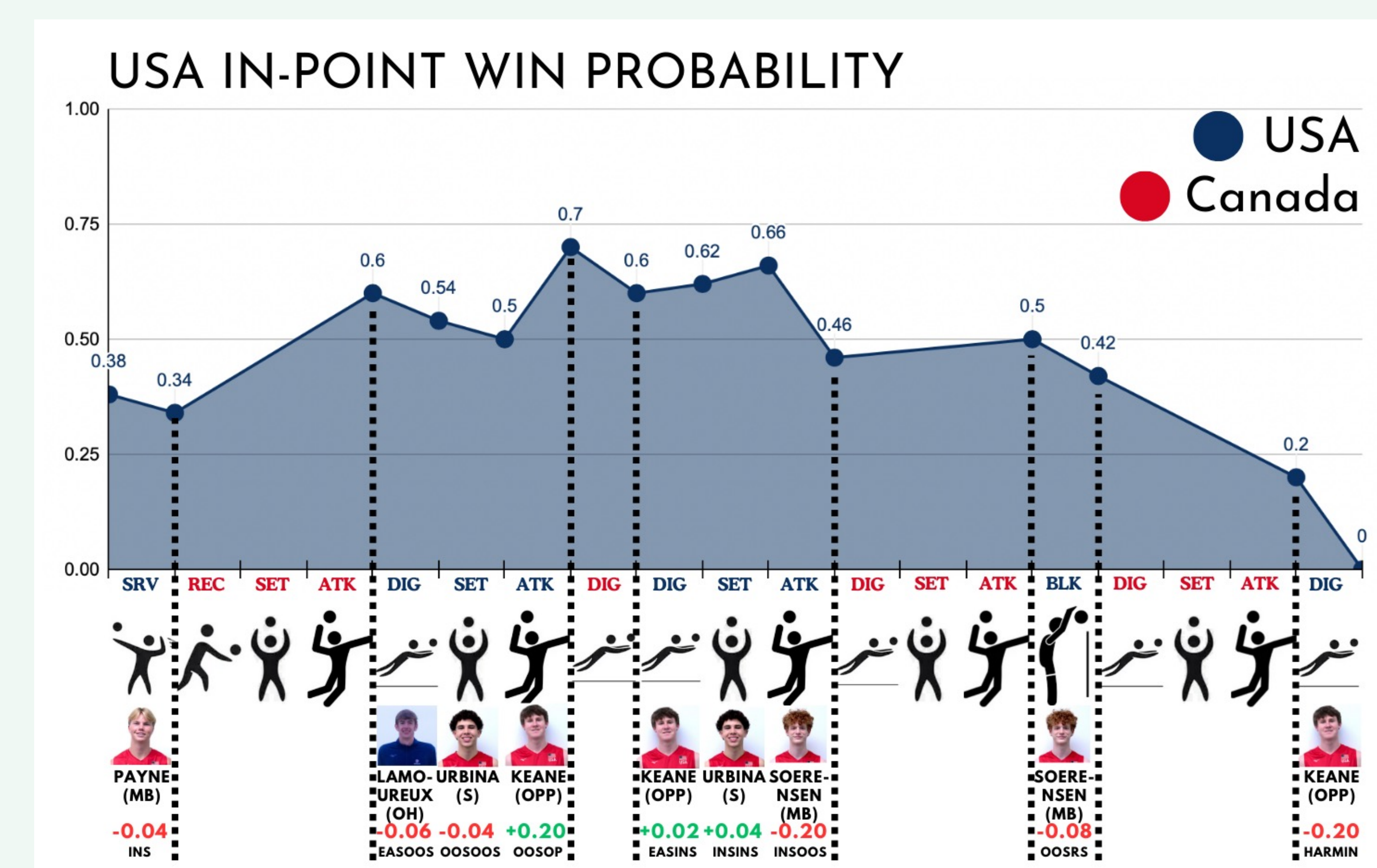
NORCECA U19 Championship | All games between 2 of Canada/Cuba/USA | Made by Kevin Baer

PLAYER	TEAM	SERVE: SERVES[1]	SERVE: PPA[2]	RECEIVE: RECEIVES	RECEIVE: PPA	SET: SETS	SET: PPA	ATTACK: ATTACKS	ATTACK: PPA	BLOCK: BLOCKS	BLOCK: PPA	DIG: DIGS	DIG: PPA	WEIGHTED PPA[3]
MIDDLE														
Payne	USA	50	0.072	—	—	—	—	41	0.033	44	0.140	—	—	0.088
Iverson	Cuba	22	-0.013	—	—	—	—	11	0.116	16	0.091	—	—	0.074
Prentis	Cuba	21	-0.113	—	—	—	—	15	0.117	16	0.055	—	—	0.037
Soerensen	USA	39	-0.023	—	—	—	—	22	0.112	31	0.009	—	—	0.035
Acosta	Cuba	29	-0.008	—	—	—	—	13	-0.055	9	0.058	—	—	0.006
McDonald	Canada	31	-0.021	—	—	—	—	21	-0.005	20	0.023	—	—	0.004
Tate	Canada	34	0.025	—	—	—	—	13	-0.218	17	0.044	—	—	-0.045
Vandenheuvel	Canada	10	-0.084	—	—	—	—	6	-0.050	9	-0.029	—	—	-0.047
OUTSIDE														
Mikhailenko	USA	46	-0.010	72	-0.028	—	—	75	0.097	16	0.204	18	-0.086	0.042
Marrero	Cuba	33	0.006	62	-0.081	—	—	62	-0.004	7	0.251	39	-0.076	0.002
Veith	Canada	41	-0.032	74	-0.066	—	—	79	0.042	23	0.063	32	-0.043	-0.004
Dezutter	Canada	26	0.000	39	0.007	—	—	52	-0.002	8	-0.020	18	-0.112	-0.013
Lamoureux	USA	51	0.029	56	-0.089	—	—	75	0.037	10	-0.138	30	-0.068	-0.027
Gonzalez	Cuba	25	-0.017	51	-0.065	—	—	62	-0.053	9	0.151	25	-0.134	-0.034
Rodriguez	Cuba	20	-0.018	50	-0.025	—	—	38	-0.050	6	-0.033	22	-0.049	-0.037
Miles	Canada	10	-0.128	24	-0.124	—	—	21	-0.015	5	-0.020	9	-0.124	-0.076
SETTER														
Urbina	USA	44	0.006	—	—	235	0.008	18	0.182	24	-0.071	31	-0.053	0.000
Flaquet	Cuba	33	0.003	—	—	209	-0.001	14	-0.016	12	-0.030	21	-0.039	-0.007
Weiss	Canada	25	-0.029	—	—	180	-0.017	6	-0.037	8	0.065	25	-0.064	-0.015
LIBERO														
Navarro	Canada	—	—	59	-0.001	26	-0.027	—	—	—	—	58	0.004	-0.003
Bluth	USA	—	—	63	-0.004	21	-0.031	—	—	—	—	45	-0.019	-0.013
Guimera	Cuba	—	—	56	-0.021	25	-0.029	—	—	—	—	54	-0.064	-0.038
OPPOSITE														
Wright	Canada	47	0.046	—	—	—	—	66	-0.041	14	-0.026	16	-0.020	-0.014
Medina	Cuba	37	0.003	—	—	—	—	114	0.070	15	-0.203	19	-0.047	-0.015
Keane	USA	48	-0.062	—	—	—	—	57	-0.022	30	-0.081	40	-0.059	-0.048

[1] Minimum of 5 rep in each weighted category for inclusion
[2] Point Percentage Added: a stat that measures impact on winning a point per touch
[3] Weighted by Position: e.g. Outsides are 19% Serve, 25% Receive, 36% Attack, 12% Block, 10% Dig

ANALYZE

One area that volleyball can continue to grow in is televised broadcasting. What that looks like is better statistics and visualizations to power engagement and analysis. Similarly to other sports, win probability is a great place to start breaking down situational attitudes or feelings into precise numerical values, such as representing a win probability as 75% versus saying "I think this team is going to win". One area where volleyball can go above and beyond other sports is in-play win probabilities. Because of volleyball's discrete nature and the game state values used throughout this research project, we can build visuals like the one on the right to illustrate the point in terms of win probability. This also yields a new field of expected points from the percentage of space controlled by a team. These can likely be extended to full match probabilities; that is an area of further research. Visualizations, like the one on the right, can greatly increase the fan experience of volleyball and serve as another way to quantify the feelings of fans as they watch their team lose despite total control of the point.



Here, win probability demonstrates a point the United States could have potentially won. By taking the definite integral, we can determine the United States' control of the total space as **48.61%**. This metric adds considerable context to the 1-0 point score for Canada.

PREDICT

Predicting volleyball matches typically focuses on schedule and score-based quantifications. While Massey Rankings and the kind are great, I offer a new type of prediction based on the overall Point Percentage Added (PPA) of each starting player, adjusted for number of rotations. By using the Pythagorean Winning Percentage, a final prediction is given from a neutral point advantage. Notably, this also gives a rating out of 1000 points played against the "perfectly average" team, allowing for simple rankings and comparisons. This system is well suited to adjustments in starting lineups like injury or player rotation. Unfortunately, this system does not currently adjust for team focuses outside the norm, like a heavy right-side attack or libero-heavy serve receive patterns. That is an area of continued focus. Finally, one can also make predictions for hypothetical matchups between uneven competition levels – although this comes with risks given the different competition pools.

USA U19				CUBA U19
PAYNE	+8.8%	MB	+7.4%	IVERSON
SOERENSEN	+3.5%	MB	+0.6%	ACOSTA
MIKHAILENKO	+4.2%	OH	+0.2%	MARRERO
LAMOUREUX	-2.7%	OH	-3.4%	GONZALEZ
KEANE	-4.8%	OPP	-1.5%	MEDINA
URBINA	=0.0%	S	-0.7%	FLAQUET
BLUTH	-1.3%	L	-3.8%	GUIMERA
RATING	MATCH WIN PROB			RATING
507	55.10%		44.90%	496

Predictive analyses are created through player overall weighted PPA, number of rotations, and Scott Power's Pythagorean winning percentage. Rating is a team's score out of 1000 points against the "perfectly average" team.

USA U19				USA U21
PAYNE	+8.8%	MB	+3.6%	TOMKINSON
SOERENSEN	+3.5%	MB	-2.4%	CRYST
MIKHAILENKO	+4.2%	OH	+3.3%	KELLY
LAMOUREUX	-2.7%	OH	+1.9%	FOLEY
KEANE	-4.8%	OPP	+6.4%	HARTKE
URBINA	=0.0%	S	+6.0%	ROSENTHAL
BLUTH	-1.3%	L	+2.8%	ALVIAR
RATING	MATCH WIN PROB			RATING
507	4.57%		95.43%	703

We can even compare hypothetical matchups, although care should be taken given the different opposition levels. One area of future exploration will be opponent adjustment.