

Sports Analytics: Designing a Volleyball Game Analysis Decision-Support Tool Using Big Data

Sarah Almujaheed, Nicole Ongor, John Tigmo, Navjot Sagoo

Abstract— From 2006-2012, George Mason University's (GMU) division I men's and women's volleyball teams were outplayed by their top competitors within their associated conference. Analysis of historic data showed that the GMU's men's and women's volleyball teams have a lower probability of scoring points on average of 0.21 and 0.05 respectively. The win/loss outcome is a function of the combinations of sequences of events caused by team's actions and coach's tactics.

The data is so complex that no human can comprehensively conduct the analysis. A Computer-Aided Analysis Tool (CAAT) is needed to analyze the underlying trends contributing to the wins and losses as well as provide a meaningful recommendation to improve the overall team performance in a volleyball game. The CAAT determines the probability of each transition that can occur in a volleyball game, uses an Absorbing Markov Chain to evaluate how events influence the point scoring probability, and runs a Monte Carlo Simulation to analyze how random variations in transition probabilities, caused by extreme conditional scenarios can affect the team performance and end result of a game.

Four design alternatives were identified through analysis of historic data and evaluated for improving team performance through specific skill improvement training: 1) Increasing aces; 2) Increasing kills; 3) Increasing blocks; 4) Decreasing errors.

A utility analysis was conducted to determine the most effective design alternative to achieve the target level of performance. Based on the utility analysis, the GMU's women's and men's teams must focus on increasing their blocks. Out of 10 blocks, at least 9 should lead to a point for the men and 3 should lead to a point for the women in order to achieve the target level of performance.

I. INTRODUCTION

THE largest national non-profit collegiate sports organization in the United States is the National Collegiate Athletics Association (NCAA). Its main mission is to promote student athlete development alongside higher education. Over 1,000 colleges are affiliated with the NCAA and over 400,000 student athletes participate every year. These colleges and universities invest large percentages of their auxiliary budget, equating to thousands, some even millions of dollars into their sports' programs.

Manuscript received April 1, 2013. George Mason University's Patriots, Intercollegiate Athletics Association, and head coaches of the volleyball programs at George Mason University, Coach Chao and Coach Kendrick in addition to Mr. Juan Mantilla supported this work. All authors are students at the Volgenau School of Engineering, Dept. of Systems Engineering and Operations Research, George Mason University, Fairfax, VA 22030 USA.

S. Almujaheed (email: salmuja@gm.u.edu), N. Ongor (email: nongor@gmu.edu), J. Tigmo (email: jtigmo@gmu.edu), and N. Sagoo (email: nsagoo@gmu.edu).

The NCAA divides teams into Division I, II, or III based on each school's choice of competition level that best fits its mission. NCAA strives to ensure fairness within the competition. Divisions are further divided into regional conferences. Each university that participates in the NCAA must affiliate all of its sports programs with one of the three divisions [4].

Historical trends show correlations between D-I team's competitive success with respect to their associated large fiscal budgets, high expectations, high coach's salary and quantity of athletic scholarships [7][8]. This created a need for coaches to come up with new methods to assist in making sound judgments and quick decisions on and off the playing field, to win against the best competitors and meet the expectations of their school.

While some stakeholders, in the field of sports, use small data and basic analysis to assist with strategic development, others use observational performance as a general rule of thumb in making decisions. These approaches are limited by their small data sets and or their subjective theories that correlate the actions performed in a game with the quantitative outcome [9].

II. CONTEXT AND STAKEHOLDERS

A. Collegiate Sport Specific Study

Every season, the GMU's men's program competes in the Eastern Intercollegiate Volleyball Association (EIVA) conference and the GMU's women's volleyball program competes in the Colonial Athletics Association (CAA) conference. The winning percentage of each conference match is used for bracketing and seeding criteria for post-season championship bid decisions. The winning percentage is calculated using the win/loss record and is used to measure competitive success.

GMU's intercollegiate athletic department (ICA) committee and the athletic director, use an evaluation to assess and quantify the coach's job performance relative to outcomes desired and other performance criteria of the program (Fig.1).

The coach's evaluation is comprised of two main categories: Leadership which constitutes 75% of the evaluation, and Management which constitutes 25% of the evaluation. With respect to the coach's evaluation, competitive success is a sub-category of leadership. It holds one of the largest weights, 30%. (Fig.1). Thus, competitive success is the focus of study.

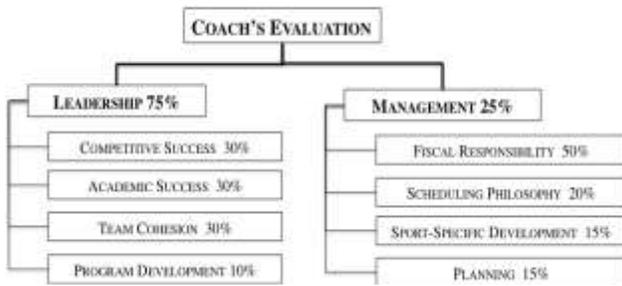


Fig.1. Intercollegiate Athletics Department's Coach's Evaluation, used to measure the outcomes anticipated and how success will be measured.

B. Focus of Study

In 2011-2012, GMU's men's volleyball was ranked 16th out of 27 Division I teams. Since 2000, GMU's men's volleyball program has averaged a winning percentage of 55%, which is a 5% higher average than that of all the teams that make up the EIVA conference. However, their winning percentage record is decreasing with a slope of -0.0074 (Fig.2).

In relation to GMU men's volleyball program, PSU's men's volleyball (PSU), also, participates in the EIVA conference and has an average winning percentage of 79%. PSU's men's volleyball is consistently improving, as their yearly winning percentage increases with a slope of +0.0068 (Fig. 2), thus, making PSU the most competitive team in the EIVA conference.

PSU's men's volleyball team's winning percentage varies by only 0.78%, unlike GMU men's volleyball team, whose winning percentage varies by 2.85% (Fig. 2). The winning percentage gap between GMU and PSU, quantifies the approximate target winning average GMU has to increase their current winning percentage by, in order to consistently maintain a lead over PSU and other EIVA conference teams.

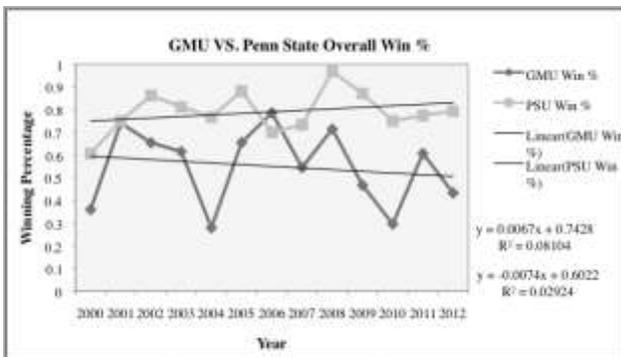


Fig.2. The line graph (right graph) shows the Men's volleyball performance variability shows the extent to which the winning percentage data points differ from each other in any given season in comparison to Penn State, a rival competitor

In 2011-2012, GMU's women's volleyball was ranked 220th out of 327 Division I teams. Since 2000, GMU's women's volleyball program has averaged a winning percentage of 47.8%, which is a 2.2% lower average than that of all the teams that make up the CAA conference (Fig.4). Moreover, GMU's winning percentage record is decreasing with a slope of -0.0283 (Fig.5).

Northeastern University women's volleyball team (NE) competes within the CAA conference with a winning

percentage of 61.4%, well above the average of the conference. NE's women's volleyball team is consistently improving, as their winning percentage is increasing with slope of +0.0102 (Fig.3), making them a top competitor of the CAA conference.

NE's seasonal winning percentage varies by only 0.94%, whereas GMU's winning percentage varies by 4.17% (Fig.3). The winning percentage gap between GMU and NE, quantifies the approximate target winning average GMU's women's volleyball team has to increase their current winning percentage by, in order to consistently maintain a lead over NE and other CAA conference teams.

2000-2012.

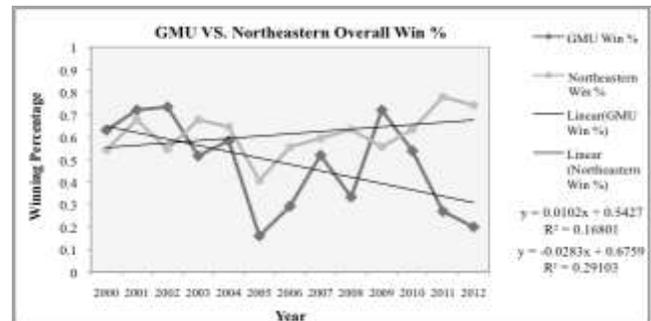


Fig.3. The line graph (right graph) shows the Women's volleyball performance variability shows the extent to which the winning percentage data points differ from each other in any given season in comparison to Northeastern, a rival competitor.

III. PROBLEM AND NEED STATEMENTS

A. GMU's Women's Volleyball Team

GMU's women's volleyball team's average conference winning percentage between years 2000-2012 is below the average conference winning percentage of NE (Fig.4). Compared to NE, they have a, higher variance in their winning percentages for those years (Fig.5). According to official NCAA box scores, GMU's women's volleyball team only won 3 out of 13 matches against NE between years 2006 and 2012, equating to a 23.07% winning percentage against Northeastern's women's volleyball team.

Overtime, GMU's women's team needs to increase their overall conference winning percentage by more than 27.15% to compete, as a top contender, for the lead in their conference. Moreover, there is a need to reverse the decreasing trend and maintain consistency by reducing the variance of their winning percentages by about 77.5%. In order to satisfy this need, GMU's women's volleyball team needs to win at least 50% or more of their matches, each season, against their top competitor.

B. GMU's Men's Volleyball Team

GMU's men's volleyball team's average conference winning percentage is above the average of the conference teams by 5%. However, compared to PSU, they have a lower winning percentage average and a higher variance in their winning percentages. According to official NCAA box scores, GMU's men's volleyball team only won 2 out of 26 matches against PSU between years 2006 and 2012,

equating to a 7.7% winning percentage against PSU's men's volleyball team.

Overtime, the men's volleyball team needs to increase their average overall winning percentage by more than 43.7% in order to take the lead in their conference. They also need to reverse the decreasing trend and maintain consistency by reducing the variance of their winning percentage by about 72.25%. In order to satisfy this need, GMU's men's volleyball needs to win at least 50% or more of their matches, each season, against their top competitor.

In both the men's and women's cases, objective and comprehensive assessment of game event performance is required to identify the skills that could be improved to enhance each volleyball program's win/loss record.

IV. DESIGN ALTERNATIVES

A trend analysis was conducted, using historical data from 196 box scores of matches played against several teams from 2002-2012. The trend analysis was used to determine which transitions directly lead to points and their level of contribution to point scoring. The analysis was done for both GMU's teams and the top competitor's teams and was used to derive the design alternatives of transitions that directly lead to points.

A. Increasing Aces

A service ace occurs when a team serves the ball and that ball transitions directly to a point for that team without undergoing any other transitions.

B. Increasing Blocks

Blocks are used to strategically land the ball in the opposing team's side of the court or deflect the ball from being spiked by the opposing team, resulting in a point.

C. Increasing Kills

A kill is a particular type of an attack that results in a direct point by grounding the ball in the opposing team's side of the court.

D. Decreasing Errors

Errors include multiple transitions because an error can occur between multiple actions. This design alternative includes service errors; ball-handling errors, from passes and sets; blocking errors; reception errors; and attacking errors. These errors made by a team are transitions that directly contribute to points for their opposing team.

In 2011-2012, both GMU's men's and women's volleyball programs had a point scoring probability of less than 0.5. Northeastern and Penn State's point scoring probability was greater than 0.5 against GMU's volleyball programs. Official NCAA box scores were used to validate the baseline distribution for these design alternatives. In order to assess the sensitivity of the winning percentage outcome, for each design alternative, the simulation is required to replicate all transitions between the two teams.

V. METHOD OF ANALYSIS

A. Simulation

The analysis conducted, to assist in determining the most beneficial design alternative, utilizes a stochastic Monte Carlo simulator. As a mathematical system, the simulation models an Absorbing Markov Chain (AMC) process that is representative of a volleyball game's actions and transitions. Multiple states and transitions occur, where states are independent actions and transitions are dependent events.

The simulator defines serves, pass/receives, sets, and blocks as transient states and points as absorbing states. Thus, there are 10 transient states and 2 absorbing states (Fig.4). The simulation accounts for two teams, representing GMU as 'A' and their associated top opponent as 'B'.

Each state represents an action that occurs in the simulated game. A transition occurs to and from a state, and the chances of a transition occurring is referred to as a transitional probability. A game is won, once one of the two teams obtains a total of 25 points and is 2 points ahead of the other team's score. The match is won, once 3 out of five games have been won by one of the two teams. The simulator parsing processor function quantifies the experimental probabilities into actual values of occurrences.

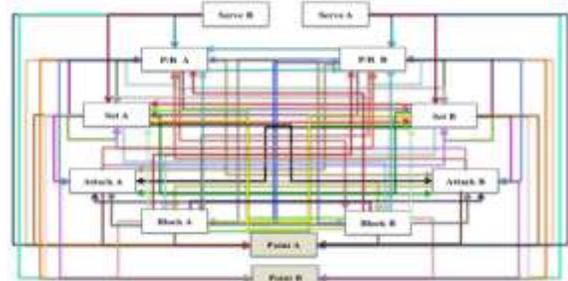


Fig. 4 Volleyball Game Simulation is modeled by an Absorbing Markov Chain, in which all the states and transitions are depicted in the model and simulation.

B. Methodology

A video analysis was conducted, using a sample size of 6 matches that were watched for both GMU's men's and women's volleyball. This analysis was used to collect transitional data between events such as serves, sets, passes, attacks, blocks and points. An estimated 1,400 transitions occurred per match. The data accumulated for each video analysis was used to calculate the average for a baseline distribution for each type of transition. Refer to Fig.5.

A random number is generated, within the simulation, after each transition. The random number that is generated falls within a respected distribution interval of a given transient or absorbing state. The transition probabilities of a transient and absorbing state must sum up to 1. Therefore, the parameters of each transient state's transitional probabilities are measured within 0 to 1. This represents the likelihood of an event occurring between states and or within a state. Refer to Fig.4.

This process is continued until a transition goes to an absorbing state. An absorbing state is one that once reached, no further transitions can occur from it to any other state. Once the ball has reached an absorbing state of either team,

a count keeps track of the score and the rally starts again with a serve by the team who won the point. Finally, the simulation outputs the point scoring probability, the winning percentage and number of occurrences for each transition for each team. Refer to Fig.5.



Fig.5. Functional Block Diagram Process representing the input and outputs of the system.

The simulator is used as a testing and verification mechanism, in that it quantifies the probability and number of occurrences needed to reach a target and optimal winning percentage. The random variables were based on a uniform distribution with associated mean and standard deviation for each match. The simulator runs 5,000 matches per second. The study conducted 10 runs per unit of probability increment or decrement, depending on the type of design alternative. The simulation completed 4,000,000 iterations in total, in order to ensure statistically significant model values when comparing it to the NCAA official box score data. The percent error of the simulation relative to the data is 2.11% for men's volleyball and 2.56% for women's volleyball.

C. Absorbing State Matrix Calculations

TABLE I
ABSORBING MARKOV CHAIN (AMC) MATRIX PROPERTIES

Symbol	Quantity
i	states in matrix row
j	states in matrix column
a	absorbing state probability
s	number of transient states
m	$(s - m) \times (s - m)$ matrix that represents transitions between transient states
Q	$m \times m$ identity matrix reflecting the fact that we can never leave an absorbing state
I	$(s \times m)$ m matrix representing transitions from transient states to absorbing states
R	$(s \times m)$ m matrix representing transitions from transient states to absorbing states
O	m $(s \times m)$ matrix consisting entirely of zeros. P correspond (in order) to the states $t_1, t_2, \dots, t_{s-m}, a_1, a_2, \dots, a_m$
P	resulting probabilities

From the following matrix properties, listed in the Table I. The following matrix was used to obtain probability of points being scored in a match [2].

$$P = \begin{matrix} & \begin{matrix} s-m & m \\ \text{columns} & \text{columns} \end{matrix} \\ \begin{matrix} s-m & m \\ \text{rows} & \text{rows} \end{matrix} & \left[\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right] \end{matrix} \quad (1)$$

An AMC matrix's entries reflect the probabilities of transitioning from any state i to any state j . The following equation (2) was derived to determine the average probability of points scored in a given match.

$$P_{pi} = \frac{\sum_j [(I-Q)^{-1}R]_j}{j} \quad (2)$$

Thereby, increasing or decreasing the probability of one or more transition(s) can prove different cause-and-effect scenarios that can alter the way a team can play in future games. The value of a probability determines the sensitivity of an outcome. Probabilities in that matrix are adjusted to analyze the outcome of the match within the simulation.

D. Changing Transitional Probabilities

The sum of the transitional probabilities for each state has to equal one, if one transitional probability increases, the sum of all the transition probabilities within that state must still add up to 1. The formula used to determine the changes needed for the other transitional probabilities within a state of each design alternative was calculated using the following equation:

$$W_j = (I - W_j) \left[\frac{W_j^o}{\sum_{k \neq i} W_k^o} \right], j \neq i \quad (3)$$

For this study, W_j is the new transitional probability for element j . W_j is the priority probability, which is the adjusted value made in element i , W_k^o is the original transitional probability value of element j . $\sum_{k \neq i} W_k^o$ is the sum of all other (original) transitional probabilities within that specific state (excluding the priority transitional probability). This assures that all values are normalized relative to the design of experiment's value for each design alternative's state's probability (s).

VI. RESULTS

A. Extreme Variations in Input Assumptions

In this study, the top teams in both the men's volleyball EIVA conference and the women's volleyball CAA conference are defined as the most competitive and consistent teams within the conference. The assumption is that, if GMU's volleyball teams are able to defeat their associated top competitors within their conferences; then, they will be able to defeat all other teams within their conferences.

The design alternatives were tested under the assumption that the opposing team's performance remains consistent in transitional distributions with low variations in performance (Fig.1) and (Fig.2). If the distributions of either team changes over the year, these changes will need to be refined within the tool in order to obtain accurate information.

B. GMU's Men's and Women's Volleyball Team's Sensitivity Analysis

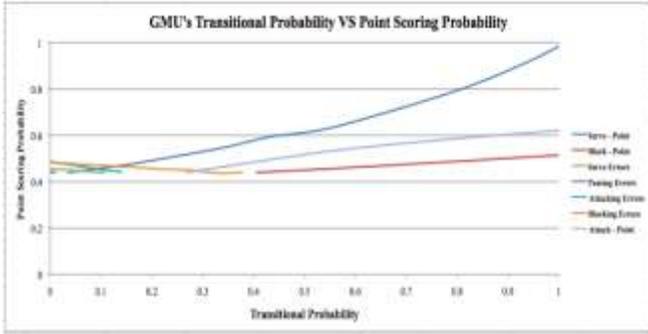


Fig.6. Sensitivity Analysis.

The sensitivity analysis illustrates the cause-and-effects of the changes made to the original transitional probabilities. The sensitivity analysis was used to determine the design of experiment's transitional probabilities analysis. (Fig.6).

C. GMU's Men's and Women's Volleyball's Design of Experiment Simulation Results

For GMU's volleyball teams to compete as a top contender, solely based on the merit of improving skills, GMU's volleyball teams needed to start by winning at least half of their matches played against their top competitors. Thus, the first set of simulation runs was based on improving skills that will always yield a 50% winning percentage outcome. To provide a range for the GMU coaches, another scenario was tested. The second set of simulation runs was based on improving skills that will always yield a 90% percentage winning percentage.

TABLE II
GMU MEN'S VOLLEYBALL DESIGN OF EXPERIMENT

Design Alternatives	7.7% Winning Percentage (Current and Most Probable Scenario)		50% Winning Percentage (Breaking Even and Least Optimal Scenario)		90% Winning Percentage (Significantly Optimal Scenario)		
	Transitional Probabilities	Number of Occurrences	Experimental Probabilities	Number of Occurrences	Experimental Probabilities	Number of Occurrences	
Aces	0.0359	3	0.22	29	0.525	33	
Blocks	0.406	12	0.906	24	0.996	(N/A)	
Decrease Errors	Serves	0.0152	9	0.0652	7	0.0152	2
	Passes	0.0032	2	0.0062	0	0.0022	0
	Attacks	0.1498	14	0.0608	5	0.0208	4
	Blocks	0.3775	30	0.1775	5	0.0275	1
Kills	0.2675	30	0.4675	39	0.5675	47	

TABLE III
GMU WOMEN'S VOLLEYBALL DESIGN OF EXPERIMENT

Design Alternatives	23.07% Winning Percentage (Current and Most Probable Scenario)		50% Winning Percentage (Breaking Even and Least Optimal Scenario)		90% Winning Percentage (Significantly Optimal Scenario)		
	Transitional Probabilities	Number of Occurrences	Experimental Probabilities	Number of Occurrences	Experimental Probabilities	Number of Occurrences	
Aces	0.0223	6	0.1223	15	0.3223	24	
Blocks	0.1623	15	0.3624	17	0.6623	30	
Decrease Errors	Serves	0.0864	11	0.076	4	0.0614	2
	Passes	0.1023	15	0.0623	13	0.0223	13
	Set	0.0229	9	0.0088	4	0.0048	2
	Attacks	0.304	14	0.0941	8	0.0641	3
	Blocks	0.313	16	0.2887	14	0.2137	13
Kills	0.1575	18	0.2575	30	0.4075	35	

The tables shown above represent the design of experiment for the GMU men's and women's volleyball

teams. The first column lists all the design alternatives that were tested for this study. The table is then sub-divided into three sections. The first section lists the current transitional probabilities based on the current winning percentage of GMU against associated top competitor. These values are raw transitional data that was obtained by the video analysis conducted earlier in the study. The second and third sections list the experimental probability, which is the change in transitional probability from current transitional probability. The experimental probability is the value used in the simulation to achieve the associated winning percentages. The sections also list the number of occurrences needed to obtain the winning percentage of its associated experimental probability.

All sets from GMU's men's volleyball team transitioned to the state of an attack for GMU's men's volleyball team, thus, no set errors occur. When GMU men's volleyball team's blocks to points transitional probability was increased to a probability of 1 for their team, the maximum winning percentage considering all other factors remain at their current status was 63.17%. Thus, blocks, as a design alternative for men was consider not applicable because it could not reach a 90% winning percentage at its maximum transitional probability.

D. GMU's Men's and Women's Volleyball Team Comparative Analysis

A comparative analysis was performed to find differences and similarities between the results of the two simulated scenarios. These two scenarios were then compared to the average transitional probabilities that yield the current average winning percentage of GMU's volleyball teams against their associated top competitor. There were not any comparative-analyses of combinations of alternatives conducted in this study. Each design alternative was analyzed independently and tested independently. All other transitions that were not related to the design alternative remained at their original probabilities.

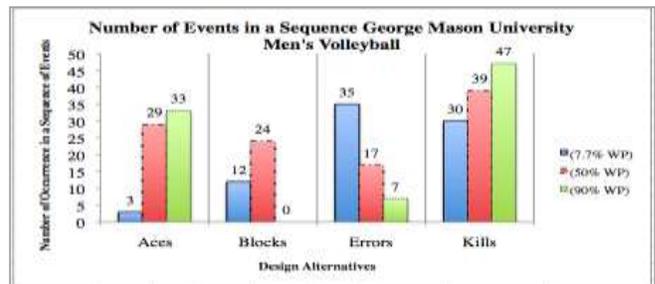


Fig.7. GMU men's simulated run based on conditional scenarios.

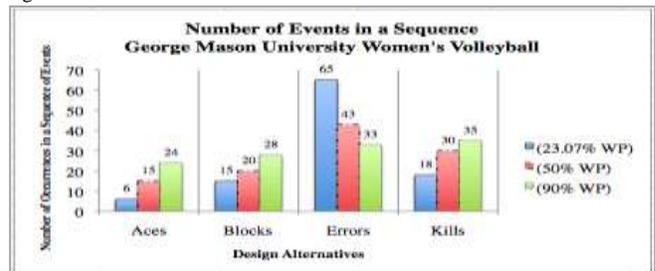


Fig.8. GMU women's simulated run based on conditional scenarios.

The comparative analysis indicates that each type of alternative will need a significant amount of improvement in order to achieve a winning percentage of 50% and 90% from the current winning percentage (Fig.7) and (Fig.8).

E. Utility

A utility function was developed to weigh the importance of three main aspects in decision-making done by the coaches, which are maintainability (MA), trainability (TR), and modifiability (MO). The utility values are based on achieving a 50% chance of winning because this is the minimum target outcome that will address the problem and satisfy the need. The weights for the utility were determined through discussion with GMU coaches, since they are the primary decision-makers.

Maintainability is defined as consistency in performance, which addresses the long-term problem of decreasing the variability of GMU's men's and women's conference performance. Trainability, of existing players, is sub-divided into efficiency (EFFY) and effectiveness (EFFT). Efficiency is defined as the time, effort, and resources needed for the intended task to achieve new standards. Effectiveness is defined as the capability of the team producing the desired results. Modifiability is the amount of change needed to obtain target winning percentage. The utility is denoted as $U(x)$, which equates to the following:

$$U(x) = (0.33)MA + (0.33)TR + (0.33)MO \quad (4)$$

$$TR = (0.5)EFFY + (0.5)EFFT \quad (5)$$

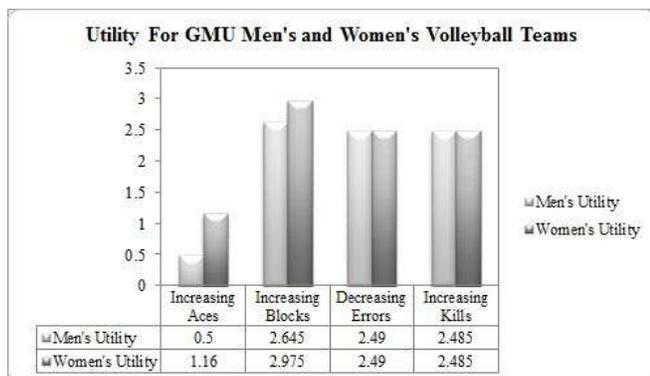


Fig. 9. GMU men's and women's volleyball utility values for design alternatives.

The results of the utility function, (Fig.9); show that for both the men's and the women's teams, increasing blocks is the best alternative as it has the highest utility value.

VII. RECOMMENDATIONS

In order to determine the best alternative to recommend to the coaches, Table IV was developed to rank the alternatives from most effective to least effective based on the amount of change required to achieve the target performance and the utility analysis results for each design alternative.

The left column of the table ranks the alternatives based on the amount of change is required while the right column

of the table ranks the alternatives based on the utility of each alternative as explained previously. Refer to Table IV.

TABLE IV
RANKING OF DESIGN ALTERNATIVES

Team	Ranking by amount of change required	Ranking by utility value
Men	Increasing Kills	Increasing Blocking Points
	Increasing Blocking Points	Decreasing Errors
	Decreasing Errors	Increasing Kills
	Increasing Aces	Increasing Aces
Women	Increasing Blocking Points	Increasing Blocking Points
	Increasing Aces	Decreasing Errors
	Increasing Kills	Increasing Kills
	Decreasing Errors	Increasing Aces

Based on the rankings provided in the table for the women's and men's teams, it was determined that the coaches should focus on increasing blocks. By referring to Table II and Table III we can see that out of 10 blocks, at least 9 should lead to a point for the men and 3 should lead to a point for the women in order to achieve the target level of performance. Moreover, it was determined that decreasing errors is the second best alternative given that the teams focus on decreasing all errors simultaneously. Given that increasing kills and aces occupy the bottom ranks, the recommendation to the coaches would be to recruit players that are already good at hitting and serving.

VIII. ACKNOWLEDGMENT

The authors thank Coach Fred Chao, Coach Patricia Kendrick, Mr. Juan Mantilla, Mrs. Paula Lewis, Dr. Lance Sherry and Dr. George Donohue for their advice and expert assistance.

REFERENCES

- [1] T.L. Saaty, "The Analytic Hierarchy Process: Planning Priority Setting, Resource Allocation (Decision Making Series)," New York: McGraw-Hill, 1980.
- [2] W.L. Winston, J.B. Goldberg, "Operations Research: Applications and Algorithms," 4th ed. Thomson Brooks/Cole, 2004.
- [3] National Collegiate Athletics Association, "2012 and 2011 NCAA Women's Volleyball Rules and Interpretation," Indianapolis, IN, 2010, pp.69-78.
- [4] National Collegiate Athletics Association. (2013) [Online]. Available: <http://www.ncaa.org/>
- [5] Colonial Athletic Association. (2013). [Online]. Available: <http://www.caasports.com/>
- [6] Eastern Intercollegiate Volleyball Association. (2013). [Online]. Available: <http://www.eivavolleyball.com/>
- [7] S.Chen, S.Synder, M.Magner, "Journal of Issues in Intercollegiate Athletics, College Sport Research Institutes, 2010, 176-193.
- [8] The Coach. (2008,1, 20). "The College Volleyball Coach: College Volleyball Coaching as a Career [Online]. Available: <http://www.collegevolleyballcoach.com/2008/01/college-volleyball-coaching-as-career.html>
- [9] D.Nelson, "Statistics in the Real World." [Online]. Available: <http://www.math.wichita.edu/history/topics/stat.html>