ArcUser Online

Search *ArcUser:* [ ] Search

✉ E-mail

Spatial Madness

By Brian Ward and Brian Davenhall, CH2M HILL, Inc.

An analysis of the NCAA Men's Basketball Tournament

Editor's note: Known as March Madness, the National Collegiate Athletic Association (NCAA) Men's Division I Basketball Championship has become one of the highest-profile sporting events in the United States since its inception in 1939. Teams from across the country compete in this single-elimination-format event. All games in the 20-day tournament are broadcast on television, and wagering on the tournament has become something of an event in itself.

The authors' paper, presented at the 2006 ESRI International User Conference, examined possible predictive correlations between the distance traveled by a team competing and its likelihood of prevailing. The paper presentation was well attended, and the paper, as well as comments received following the presentation, are the basis for this article. The accompanying article "Going the Distance—Did team travel make a difference in the outcome?" gives the uninitiated a sense of the excitement that the NCAA tourney engenders.

The authors' study considers the NCAA Men's Basketball Tournament since its expansion in 1985 to 64 teams. Over the 21 years included in the study, teams participating in the NCAA Men's Basketball Tournament have represented 46 states, with game locations in 38 states. While there is a cursory attempt by the NCAA tournament selection committee to place teams within logical geographic areas, more often than not, teams travel a significant distance away from their campuses to participate in games. Previous statistical analyses have found strong evidence of a correlation between home-court advantage and the outcomes of sporting events, but the primary aim of this study was to determine whether a predictive correlation exists between the distance traveled by a team and its likelihood for success.

Analytical Methods Used

Since the participants in the NCAA Men's Basketball Tournament and the game results are part of historical public record, this study used a previously compiled database that can be found at the

HoopsTournament.net Web site. This spreadsheet-style Microsoft Access database contains a great deal of relevant information about each game from 1939 to 2005. The database was normalized, producing additional tables, such as Locations and Teams, that are related to the primary Games table. Each table included the city and state of each school or game location.

Using the ESRI Data & Maps CD that comes with ArcGIS, the United States Cities layer was used to create an Address Locator for geocoding team and venue locations to the city level. These geocoded locations were generally produced with less than six miles of location difference. The authors believe this provides an acceptable spatial accuracy for analysis. Most entries in each table were matched with 100 percent confidence, though a few required manual, interactive matching. Where a comparable city located in the Address Locator was available, the city in close proximity was used. For example, Rutherford, New Jersey, served as a proxy to East Rutherford, New Jersey. In situations lacking an acceptable proxy, Wikipedia.com was used to locate the actual latitude and longitude for the city using World Geodetic System (WGS 84). The latitude and longitude were input as records in a new table. The new table was imported into a new personal geodatabase (pGDB) feature class using the Add X,Y function in ArcMap.

The Hawth's Analysis Tools extension for ArcGIS was used to determine the Euclidean distance between game and venue locations. [Hawth's Analysis Tools is a third-party extension developed by Hawthorne Beyer that performs a variety of spatial analyses and functions. It is available at no cost from www.spatialecology.com.] The Analytical tool Distance Between Points (Between Layers) was used to produce a comma-delimited text file that included the distance for each coincidence of all teams to all locations. This analysis yielded an almost unmanageable 35,000 unique records. The file was imported into Microsoft Access as a new table and a query was written that related this table to the original Games table through a multiple-field join of Locations and Teams.

For this analysis, the authors incorporated their basic understanding of

the NCAA Tournament into a parsimonious a priori model selection and inference strategy. The covariates included for analysis were tournament seed (seed); Rating Percentage Index (RPI) rank (rpi); Euclidean distance to the game destination (distance); and three covariates derived by calculating the differences between competing teams for tournament seed (seed_diff), RPI rank (rpi_diff), and distance (distance_diff). For each model, a logistic regression analysis was performed in the R program for statistical analysis [a GNU project] using a generalized linear model (GLM) procedure. The binary response variable was created by calculating the difference in score between competing teams; negative values were assigned a code equal to 0, and positive values were assigned a code equal to 1. The Akaike Information Criteria (AIC) statistic was evaluated, then competing models were ranked using this statistic to determine the best-fitting model. Models that reported the lowest AIC values received the highest rankings. The 10 models evaluated and their corresponding results are presented below in Table 1.

Model	Covariates	AIC	Rank
1	distance	3675.582	10
2	seed	3171.378	5
3	rpi	3518.293	8
4	distance_diff	3660.256	9
5	seed_diff	2911.157	4
6	rpi_diff	3251.220	7
7	seed_diff, rpi_diff	2906.064	2
8	distance_diff, rpi_diff	3242.727	6
9	distance_diff, seed_diff, rpi_diff	2903.565	1
10	FULL MODEL	2909.565	3

Table 1: Akaike Information Criteria model results

The model that included distance_diff, seed_diff, and rpi_diff reported an AIC=2903.565 and received the highest ranking. The model that included only the difference in seed between competing teams (model 6) reported an AIC=2911.157. This model ranked fourth and reported an AIC value very close to the highest ranking model. The models that included only distance or distance_diff reported AIC values equal to 3675.582 and 3660.256, respectively, and received the lowest rankings of all models. The full model, including all six covariates, received the third highest ranking.

ArcUser Online

October - December 2006

✉ E-mail

Spatial Madness
Continued...

Discussion

Up to this point, the findings of the study have not proven that Euclidean distance is a relevant predictive factor for NCAA Tournament game outcomes. While this is somewhat disappointing, there is yet reason to suspect that spatial factors may be proven relevant if they are designed somewhat differently. However, currently, the authors must conclude that the analysis to date should most certainly not become required study for bettors—or even the casual "bracketologist"!

In reviewing the results, it is clear that RPI and seed are by far more effective means of prediction, especially when considering the relative differences in these values between the teams involved. These factors will continue to be the basis for any study of the tournament, but further consideration is warranted as to how spatial factors might enhance the predictive analysis to more accurately anticipate upsets and outcomes.

In the discussion period after the authors' presentation of a paper on the study at the 2006 ESRI International User Conference, the following suggestions for factors to include were made:

- Altitude
- Route distance, as opposed to Euclidean distance
- Air travel versus road travel
- Number of games played
- Number of overtime games played
- Regular season winning percentage
- Pod seeding (i.e., brackets)

These factors will be considered in further study research, although intuition suggests that the most important next step will be to evaluate the criteria against betting lines. While they hold no inclination toward improving bookies' capabilities or even their own ability to beat the odds, the authors suspect that the impact of distance traveled may be more relevant to the actual score differential in games, rather than to the simple win/loss outcomes. For example, a significant distance or even home-court advantage is not going to allow 16 seed Fairfield to beat

national powerhouse Indiana in the first round. In fact, no 16th seeded team has ever won a first-round game in the NCAA Tournament. However, spatial factors may be relevant in determining whether Fairfield University is trounced by 34, or competes before losing by a dozen.

In Conclusion

Regardless of the authors' further investigations into spatial indicators of NCAA Men's Basketball Tournament success, they will undoubtedly be tuned in next March—as they are every year—trying to figure out which 12 seed will upset a 5 seed this year.

Acknowledgments

The authors thank Bryce Wells for his creative input into the original study and for mastering its Web site (www.geohoops.com), Shannon McElvaney and Barb Schmitz for their moral support, and their fantastic academic mentors—Dawn Wright of Oregon State and Steve Steinberg of Humboldt State—who continue to inspire them to think outside the box.

About the Authors

Brian Ward is a GIS developer and analyst with CH2M HILL, Inc., in Colorado Springs, Colorado. He holds a bachelor's degree in professional geography from the University of North Alabama and a master's degree in geography from Oregon State University. His research interests include the analysis and visualization of spatial/temporal phenomena. He can be reached at 719-477-4917 or brian.ward@ch2m.com.

Brian Ward

Brian Davenhall is a GIS developer and analyst with CH2M HILL, Inc., in Redding, California. He holds an associate's degree in geographics from Hocking College and a bachelor's degree in natural resources planning from Humboldt State University. He is well versed in mobile application development and integration and is a Trimble Certified MGIS Trainer. His research interests include interpolation and imputation methods for data estimation, geostatistical modeling, and sampling design. He can be reached at 530-229-3253 or brian.davenhall@ch2m.com.

Brian Davenhall

ArcUser Online

✉ E-mail

Going the Distance
By Brian Ward, CH2M HILL, Inc.

Did team travel make a difference in the outcome?

With a welcoming grin on his face, James Arthur Boeheim calmly stalked the sidelines of the Super Dome in New Orleans, Louisiana,urging the Syracuse University basketball team toward victory. It was the 1987 National Championship game, his team clung to a 73-72 advantage, and the Syracuse graduate was on the verge of making history. The 42-year-old had started his path to the Final Four as a young kid fresh out of high school. Not necessarily regarded as a blue chip talent, he earned his way onto the Syracuse basketball team as a nonscholarship walk-on. Boeheim excelled as a player, then as an assistant coach, before being promoted to the position of head coach at age 31. By the time his Syracuse team arrived in Louisiana, he had built a formidable basketball powerhouse. The Orangemen were ranked number one in the country and favored to win their first national championship.

Boeheim's mild grin and boyish face were sublimely countered by the stern countenance of Robert Montgomery Knight. A former head coach at West Point and already a basketball legend, Knight had a scowl rivaling that of the sternest of military commanders. He was only four years older than his rival coach, but the weight of his personality and his stature among the coaching elite—not to mention his prematurely white hair— made him seem far more authoritative than the professorial Boeheim. In addition to his unparalleled intensity, Knight was perhaps

Jim Boeheim was the coach of the victorious Syracuse University basketball team in 1987.

best known for the 1976 Indiana team he coached to an undefeated season—something that has not been accomplished since. Though he won a national championship as a player at Ohio State and he is currently the head coach at Texas Tech, Knight is as closely tied to Indiana University as Boeheim is to Syracuse.

Knight's cadre wore their home white uniforms, appropriate in light of the path they took to reach New Orleans. Indiana's first two games of the 1987 tournament took place at the Hoosier Dome in Indianapolis, roughly 50 miles north on Highway 37 from its campus in Bloomington, Indiana. Here the Hoosiers had little trouble placating the overwhelmingly cream-and-crimson-clad crowd of 34,000, defeating underdogs Fairfield and Auburn and earning a berth in the Sweet Sixteen.

Riverfront Coliseum in Cincinnati hosted Indiana's next two victories. Though no longer playing in their home state, the Hoosiers had still managed to travel a considerably shorter distance than their opponents. Cincinnati is about a two-hour drive from Bloomington—far closer than it is to either Duke (in Durham, North Carolina) or Louisiana State (in Baton Rouge, Louisiana). Despite the comfortable Midwestern surroundings, the Hoosiers had to earn their trip to the Final Four. Duke took them to the final minutes before succumbing 88-82 on the strength of Cincinnati native Rick Calloway's 21 points and eight rebounds. Two evenings later, Indiana overcame a late 12-point LSU lead thanks to the heroics of All-American Bill Alford.

In all, the Indiana basketball team had spent barely four hours on the road in earning a Final Four berth. By the time the team played the University of Nevada at Las Vegas (UNLV) in the national semifinal game, it had still traveled less than 1,000 miles. This was in stark contrast to the 4,000-plus miles that the UNLV team, the opposing team in the Sweet Sixteen, had traveled. Similarly, Syracuse had enjoyed the benefit of a home advantage literally, in the case of having played its first two games on its home court, the Carrier Dome, and traveled more than 3,000 miles less than its semifinal opponent, Providence College. A relatively light-traveled Syracuse team had little trouble with its Big East Conference brethren in their semifinal matchup, defeating Providence 77-63. In the other game, Indiana was stretched to the limit by the up-tempo UNLV squad, but Indiana ultimately outlasted UNLV 97-93, setting up the scenario at hand.

On this New Orleans night, the game would come down to its final possession. The teams had traded blows throughout the contest, with Boeheim's top-ranked Orangemen leading the number two Hoosiers by one point in the final minute. The outcome would be decided by Boeheim's trademark 2-3 zone defense, Knight's ability to devise one last effective offensive attack, and the energy summoned up by the now physically and mentally exhausted players. With the clock running out, Indiana set up its offense. Logic seemed to dictate that Steve Alford would take the final shot, as he was the Hoosiers' most prolific offensive player. However, as Daryl Thomas held the ball near the top of the key, something unexpected occurred. Thomas swung the ball around to his left, made a pin-point pass to Keith Smart, and in one smooth motion, Smart launched his jump shot as time expired. As the net settled into stillness, the fans clothed in orange stared in shocked silence while their

crimson-clothed counterparts roared. Indiana had won the national championship.

On this night, the state of Indiana adopted the French Quarter for the evening, while the Hoosiers and their fans reveled in a notably New Orleans-style brand of euphoria as if they had discovered yet another home city on their way to basketball history.

Note: While the anecdotal evidence referenced certainly indicates a correlation between distance traveled and the outcome of this NCAA Men's Basketball Tournament game, such evidence cannot truly be relied on to establish the existence of such a correlation. In the accompanying article, "Spatial Madness: An Analysis of the NCAA Men's Basketball Tournament," the authors use a statistical methodology to analyze historical data on distance traveled by teams and game outcomes to test this correlation.