Prescriptive analytics for FIFA World Cup lodging capacity planning

Ahmed Ghoniem^{1*}, Agha Iqbal Ali¹, Mohammed Al-Salem² and Wael Khallouli²

¹*Department of Operations and Information Management, Isenberg School of Management, University of Massachusetts, Amherst, MA 01003, USA; and* ²*Department of Mechanical and Industrial Engineering, Qatar University, Doha, Qatar*

The FIFA World Cup, comprising sixty-four matches spanning an entire month, has, in recent years, been attended by about three million spectators of which over half a million are visitors requiring lodging. Planning lodging capacity for an event of this magnitude is necessary for host nations where pre-existing infrastructures are either inadequate or lacking. This paper develops an optimization analytics framework that sequentially employs two integer programming models for foreign spectator analysis and the consequent lodging requirements. The framework is applied to assess the preparedness of lodging infrastructure in Qatar for FIFA 2022.

Journal of the Operational Research Society (2017) **68**(10), 1183–1194. doi:10.1057/s41274-016-0143-x;
published online 15 December 2016

Keywords: capacity planning; FIFA World Cup; optimization; prescriptive analytics

1. Introduction

Examination of existing and planned infrastructures for holding football matches and lodging foreign spectators is central to hosting the FIFA World Cup. Infrastructure planning for Qatar 2022, with much of the lodging capacity yet to be constructed, is unique and departs from past FIFA World Cups. Lodging infrastructure has not been an issue for recent FIFA World Cups since matches have been held in cities with large populations and an established hotel industry. The currency of examination of the potential demand for lodging is a consequence of the expectation that much of the attendance will be made up of foreign spectators given that the entire country's population is a mere fraction of that of most host cities in past World Cups. The total population in the host cities was over twenty million for South Africa 2010, over thirty-five million for Brazil 2014 (with individual host city populations ranging from about half a million to over eleven million), and over twenty-four million for Russia 2018. Qatar, which has the third highest GDP per capita, has committed \$20 billion for investment in tourism infrastructure under the “Qatar 2030 Vision” for economic growth and development (Cighi and Gandhi, 2011).

Planning expansion of Qatar's hotel capacity of 13,123 rooms in 83 approved hotels, in 2013, is necessary years in advance of the upcoming World Cup to allow for construction lead time. To determine appropriate capacity for lodging, it is necessary to

account for, and incorporate, likely variation in foreign spectator attendance. The variation that must be taken into account is contributed to by the makeup of the qualifying teams, the buying power in, and the ability of fans to travel from, the associated nations, the groups for the World Cup that are constituted by FIFA, and the detailed match schedule. As such this work lies at the interface of attendance forecasting and scheduling. The work furthers the existing body of literature on operations research studies in sports management that has, hitherto, not included lodging capacity analytics. Foreign spectator attendance is influenced by the buying power in nations, and hence, multiple scenarios that differ in the composition of qualifying teams must be considered. For each scenario, the total number of spectators will depend on the capacity of the stadiums in which matches are held. Hence the specific match schedule must be taken into consideration. In Section 1.1, we familiarize the reader with the FIFA process of group formation and the World Cup schedule characteristics. In Section 1.2, we review the relevant literature on sports and tourism management. In Section 1.3, we provide an outline of the paper.

1.1. Background

Since 1998, teams from thirty-two nations drawn from six football confederations¹ compete in the FIFA World Cup. The

*Correspondence: Ahmed Ghoniem, Department of Operations and Information Management, Isenberg School of Management, University of Massachusetts, Amherst, MA 01003, USA.
E-mail: aghoniem@isenberg.umass.edu

¹Asian Football Confederation (AFC); Confédération Africaine de Football (CAF); The Confederation of North, Central America and Caribbean Association Football (CONCACAF); Confederación Sudamericana de Fútbol (CONMEBOL); Oceania Football Confederation (OFC); Union des Associations Européennes de Football (UEFA).

Table 1 FIFA World Cups (1998–2022)

Year	Host country	Area (km ²)	Population	Stadiums	Attendance		Confederation
					Total	Average	
1998	France	547030	65350000	10	2785100	43517	UEFA
2002	Japan/So Korea	477955	177551441	20	2705197	42269	AFC
2006	Germany	357021	81874000	12	3359439	52491	UEFA
2010	South Africa	1219912	51770560	10	3178856	49670	CAF
2014	Brazil	8514876	193946886	12	3429873	53592	CONMEBOL
2018	Russia	17075400	143300000	11	–	–	UEFA
2022	Qatar	11437	2169000	–	–	–	–

past five and upcoming World Cups have been held in either G7 (namely, France 1998, Japan/South Korea 2002, and Germany 2006) or BRICS nations (namely, South Africa 2010; Brazil 2014; Russia 2018). Table 1 summarizes the characteristics of FIFA World Cups since 1998 as reflected in the geographic and demographic attributes of the host country, the number of stadiums, the attendance, and the host confederation. The World Cup has been predominantly hosted by CONMEBOL and UEFA nations and, only relatively recently, hosted once by an AFC and a CAF nation where significant investment in infrastructure development was required.

The sixty-four matches that the FIFA World Cup comprises are played in two stages: Pairs of teams from 32 qualifying nations compete in the forty-eight matches of the group stage; there are sixteen matches in the knockout stage which has half as many competing teams. In the group stage, there are six matches for each group of four nations—labeled A through H—with each team playing three matches. The teams in each group are ordered (from 1 to 4) with respect to the FIFA ranking, which is inversely related to the FIFA points for a team. Teams are assigned to a group letter using FIFA's random selection mechanism that iteratively draws teams sequentially from four pots of eight teams that are ranked on the most current FIFA points. The drawings abide by two primary governing FIFA stipulations: (i) Nations in any pairing from the same pot must be assigned to distinct groups; and (ii) nations in any pairing from the same confederation must be assigned to distinct groups, except for UEFA for which up to two nations can belong to the same group (Frawley and Adair, 2014).

1.2. Relevant literature

Lodging capacity management for the FIFA World Cup is novel and has not been studied in the academic literature despite the wealth of analytics studies in sports management (Wright, 2009; Rabadi *et al*, 2015). It is pertinent to situate this paper within the two relevant themes that it ties together, namely forecasting attendance at mega-sport events and analytics for scheduling sport events.

Within the literature on general tourism forecasting, a few studies pertain, specifically, to forecasting the number of

foreign and domestic attendees for a mega-sports event. Rabadi *et al* (2015) point out that the influence of mega-events on inbound tourism is not straightforward to predict. A few studies have focused on assessing the impact of holding a mega-sports event in a nation on tourism (Fourie and Santana-Gallego, 2011). Whereas the FIFA World Cup, the Summer Olympics, or the Cricket World Cup was found to increase inbound tourism in the host nation, the Winter Olympic and Rugby World Cup negatively impacted tourism during the hosting of the event. Keum (2010) employed a gravity model to assess tourism activity between two nations in general, without any particular focus on mega-sports events. Further, some studies employ qualitative approaches such as the Delphi method to ascertain short term inbound tourist flows for a host nation. Lee and Kim (1998) examined the expected foreign tourist demand for the 2002 World Cup in South Korea using a Delphi study that involved responses from 41 experts over two rounds of questionnaires.

Topics in sports analytics over a 50-year period, as surveyed by Wright (2009), have included game strategies and tactics, predictive analytics for the outcome of different matches, fixture and official scheduling, and a variety of other operations research studies (e.g., Duckworth and Lewis 1998). The methodologies used in sport analytics (Kendall *et al*, 2010) include optimization, probabilistic modeling, and choice models (Fry and Ohlmann, 2012). A rich body of literature on sports scheduling covers sports such as basketball (Westphal, 2014), soccer (Durán *et al*, 2012), and tennis (Ghoniem and Sherali, 2010). Integer optimization models have been extensively used for scheduling sport events (Fry and Ohlmann, 2012), employing objectives and constraints that are tailored to specific recreational and professional tournaments (Ribeiro and Urrutia, 2012; Durán *et al*, 2012; Ali *et al*, 1986a, b; Ghoniem and Sherali, 2010).

1.3. Paper organization

The optimization analytics developed in this paper incorporates FIFA specifications and processes that dictate foreign spectator attendance. Section 2 introduces salient aspects of the FIFA process of group formation and the World Cup schedule characteristics to establish the premise for an overview of the

proposed framework. We assume that groups of comparable strength will lead to greater spectator interest and, therefore, attendance. Our methodology, developed in Sections 3, 4, and 5, is accompanied by a detailed illustrative example. In Section 3, we present an integer programming model that incorporates FIFA stipulations to partition the 32 qualifying nations into eight subsets of comparable skill. In Section 4, a second optimization model assigns these eight subsets of teams to lettered groups, A through H, such that the percentage fill of stadiums is maximized. In Section 5, we detail the calculation of foreign spectator attendance at each match and, therefore, the total daily foreign spectator attendance and the associated lodging requirement. Section 6 applies the framework to determine required lodging capacity for FIFA 2022 by examining the data obtained from 144 instances based on 16 scenarios that differ in the composition of qualifying teams. In Section 7, we make concluding remarks.

2. Framework for attendance analytics

Overall attendance, and consequent demand for lodging, varies due to the strength and popularity of the qualifying teams and on whether the matches that a nation plays are in high- or low-capacity stadiums. It can be expected to be higher when matches between highly ranked teams or teams from neighboring countries are assigned to higher-capacity stadiums. Attendance is further influenced by the pre-specified percentage of stadium seats that FIFA allocates to officials and spectators from the competing nations, other nations of the world, and the host nation. Offered seating capacity for fans from a nation is 12% of the 91% of seats in a stadium, since 9% of the seats are reserved for officials. Attendance is heavier during the group stage and, in particular, during its last four days when four matches are held daily. Table 2 displays a prototypical group stage schedule with twelve stadiums (as for Brazil 2014) specifying the day—1 through 15—and the stadium for each match. Each row in the schedule specifies a subset of four matches, henceforth referred to as a *row-set*, that are to be held in a pre-designated stadium of known capacity.

The methodology we develop for planning purposes aims to determine the highest possible attendance for an array of plausible scenarios of 32 qualifying nations. To gauge maximal attendance requires departing from FIFA's randomly determined lettered group formation and the pre-assignment of stadiums to each row-set. The methodology forms subsets of teams that are optimally balanced with respect to total FIFA points and, further, assigns more popular matches to larger-capacity stadiums. Popularity of matches is guided by the popularity and strength of teams as summarized by, what we refer to as, the *spectator index* which estimates the percentage of seats allocated to a qualifying nation that will be filled. To illustrate the methodology, we refer to the set of 32 nations that might qualify for Qatar 2022 listed in Table 3 which reports the confederation and FIFA points for each nation.

The spectator index is a weighted average of the ratio of expected attendance to offered seating capacity for each of the three matches for a nation. The expected match attendance by fans from a nation, denoted η , is taken to be a third of the nation's total number of spectators in South Africa 2010. For a nation that did not participate in South Africa 2010, the value for η is based on that for the closest nation from its confederation adjusted by the ratio of their GDP per capita. Neighboring countries are assigned an index value of 100%. For Qatar, the largest stadium is Lusail with a capacity of 86000, the second largest is Khalifa with a capacity of 65000, and the other ten stadiums have capacities of about 45000 seats. Noting that each stadium hosts four matches, the probability that a match is held in Lusail and Khalifa is $\frac{4}{48}$, and in one of the other stadiums is $\frac{40}{48}$. Using FIFA's seat allocation percentages, the number of seats offered to spectators from a nation would be either 9391, 7098, or 4914 if the match is held in, respectively, Lusail, Khalifa, or one of the other ten stadiums. The spectator index is computed as a weighted average using the probability of a match being held in a specific stadium: $\min\{1, \frac{4}{48}(\eta/9391) + \frac{4}{48}(\eta/7098) + \frac{40}{48}(\eta/4914)\}$. Table 3 also reports the estimated spectator index values for our illustration. The value is low for nations that are distant from the host nation and/or have a lower GDP per capita, e.g., Honduras. It is higher for countries that have a loyal fan base or a high GDP per capita, such as England, USA, or Brazil.

3. Formation of eight team subsets with comparable FIFA points

In this section, we develop the Group Formation Model that optimizes the partitioning of the 32 qualifying nations into eight team subsets. The model identifies subsets of comparable strength, as reflected by the FIFA points accumulated by each nation using a Chebyshev objective function. The notation, variables, and model follow:

Data Sets and Input Parameters

- $\mathcal{N}$: Set of 32 qualifying nations.
- p_i : FIFA points for nation i , $\forall i \in \mathcal{N}$.
- c_i : Confederation to which nation i belongs, $\forall i \in \mathcal{N}$.
- $\mathcal{P}_k \subset \mathcal{N}$, $\forall k = 1, \dots, 4$: Pot of eight nations at the k^{th} level of FIFA points; $p_i \geq p_j$, $\forall i \in \mathcal{P}_{k_1}, j \in \mathcal{P}_{k_2} | k_1 < k_2$. (The teams in the first pot have the highest points and teams in the last pot, the least.)
- $\mathcal{G} \equiv \{1, \dots, 8\}$: Index set of team subsets to be formed.

Decision Variables

- $x_{ig} \in \{0, 1\}$: $x_{ig} = 1$ if and only if nation i is assigned to subset g , $\forall i \in \mathcal{N}, g \in \mathcal{G}$.
- $w \geq 0$: Objective value.

Table 2 Typical FIFA World Cup group stage schedule with twelve stadiums

Stadium	Day														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	01 A1/2							23 D1/3				36 B2/3			47 H1/4
2		02 A3/4			14 G3/4			22 C2/4					39 D1/4		
3		03 B1/2			13 G1/2				25 E1/3					44 F2/3	
4		04 B3/4				16 H3/4				28 F2/4			37 C1/4		
5			05 C1/2			15 H1/2				27 F1/3			40 D2/3		
6			06 C3/4						24 D2/4			34 A2/3			45 G1/4
7			07 D1/2			17 A1/3				29 G1/3			38 C2/3		
8			08 D3/4				18 A2/4				30 G2/4			41 E1/4	
9				09 E1/2				21 C1/3				33 A1/4			46 G2/3
10				10 E3/4				20 B2/4			32 H2/4			43 F1/4	
11				11 F1/2				19 B1/3			31 H1/3			42 E2/3	
12					12 F3/4				26 E2/4			35 B1/4			48 H2/3

Model

Maximize w (1a)

subject to

$$w \leq \sum_{i \in \mathcal{N}} p_i x_{ig}, \quad \forall g \in \mathcal{G} \quad (1b)$$

$$\sum_{i \in \mathcal{N}} x_{ig} = 4, \quad \forall g \in \mathcal{G} \quad (1c)$$

$$\sum_{g \in \mathcal{G}} x_{ig} = 1, \quad \forall i \in \mathcal{N} \quad (1d)$$

$$x_{i_1g} + x_{i_2g} \leq 1, \quad \forall g \in \mathcal{G}, k = 1, \dots, 4, \quad i_1 \in \mathcal{P}_k, i_2 \in \mathcal{P}_k | i_1 < i_2 \quad (1e)$$

$$x_{i_1g} + x_{i_2g} \leq 1, \quad \forall g \in \mathcal{G}, i_1 \in \mathcal{N}, i_2 \in \mathcal{N} | i_1 < i_2, c_{i_1} = c_{i_2}, c_{i_1} \neq \text{UEFA} \quad (1f)$$

$$x_{i_1g} + x_{i_2g} \leq 2, \quad \forall g \in \mathcal{G}, i_1 \in \mathcal{N}, i_2 \in \mathcal{N} | i_1 < i_2, c_{i_1} = c_{i_2} = \text{UEFA} \quad (1g)$$

$$x \text{ binary}, w \geq 0. \quad (1h)$$

The objective function (1a), in conjunction with Constraint (1b), maximizes the minimum total FIFA points in a subset, thereby forming subsets of comparable strength. Constraint (1c) assigns four teams to each subset and Constraint (1d) assigns each nation to a unique subset. Constraint (1e) not only

ensures that the top eight nations, i.e., $\forall i \in \mathcal{P}_1$, seed different subsets, but, more generally, ensures that any pair of nations in the same pot is assigned to distinct subsets. Constraint (1f) ensures that nations in the same confederation are assigned to distinct subsets and Constraint (1g) imposes a similar, more relaxed, limitation for UEFA. Because UEFA has a quota of 13 qualifying nations, five of the eight subsets will necessarily include two UEFA teams. Constraint (1h) introduces logical binary and non-negativity restrictions on decision variables.

For the example of 32 nations in Table 3, introduced in Section 2, the FIFA points range from 300 for Qatar (the host nation) to 1770 for Germany, with an average of 935 points. The optimally balanced subsets obtained by applying the Group Formation Model are reported in columns 2–5 of Table 4. The subsets are remarkably balanced with total FIFA points ranging from 3732 to 3770 and both high- and low-performing teams distributed across subsets. Each of the eight subsets has one nation from the highest FIFA points pot, i.e., $\mathcal{P}_1$.² In optimizing the assignment of nations to subsets that are balanced with respect to total FIFA points, the formed subsets are expected to be more competitive than alternative FIFA-compliant subsets. Heuristically formed groups are likely to exhibit a greater discrepancy between the lowest and highest

²The nations in this pot are: Germany (DEU), Argentina (ARG), Brazil (BRA), Uruguay (URY), Netherlands (NDL), Portugal (PRT), France (FRA), and Belgium (BEL).

Table 3 Data for illustrative example

<i>Nation</i>	<i>Code</i>	<i>Confederation</i>	<i>FIFA points</i>	<i>Spectator index (%)</i>
Qatar	QAT	AFC	300	100
Australia	AUS	AFC	549	77
Iran	IRN	AFC	692	100
Japan	JPN	AFC	617	29
Korea Republic	KOR	AFC	594	13
Algeria	DZA	CAF	986	50
Cameroon	CMR	CAF	646	16
Egypt	EGY	CAF	582	100
Nigeria	NGA	CAF	701	27
Tunisia	TUN	CAF	881	50
Costa Rica	CRI	CONCACAF	1095	3
Honduras	HND	CONCACAF	433	1
Mexico	MEX	CONCACAF	935	63
USA	USA	CONCACAF	828	100
Argentina	ARG	CONMEBOL	1577	54
Brazil	BRA	CONMEBOL	1348	94
Chile	CHL	CONMEBOL	1057	26
Paraguay	PRY	CONMEBOL	434	4
Uruguay	URY	CONMEBOL	1164	9
Belgium	BEL	UEFA	1471	38
Croatia	HRV	UEFA	963	6
Czech Republic	CZE	UEFA	1045	22
England	ENG	UEFA	1031	100
France	FRA	UEFA	1180	51
Germany	DEU	UEFA	1770	63
Italy	ITA	UEFA	1146	51
Netherlands	NDL	UEFA	1415	55
Portugal	PRT	UEFA	1191	34
Russia	RUS	UEFA	788	51
Serbia	SRB	UEFA	709	3
Spain	ESP	UEFA	1130	51
Sweden	SWE	UEFA	663	8

group points. A “profile-fitting” heuristic that iteratively assigns a team with the lowest points to the group with the largest total points at that iteration is reported in columns 7–10 of Table 4. This grouping is not as balanced, obviously, as reflected by a spread of 791 FIFA points (the minimum subset FIFA points is 3090 and the maximum is 3881) compared to only 38 in the obtained optimal grouping.

4. Lettered group and stadium assignments

The larger purpose of the Group-Letter Assignment Model is to spread popular matches across different stadiums while ensuring that the more popular matches are predominantly assigned to high-capacity stadiums. We introduce the construct of *match popularity* denoted by $\pi_{g,(i_1,i_2)}, \forall g \in \mathcal{G}, (i_1, i_2) \in \mathcal{M}_g$ where $\mathcal{M}_g$ is the set of six group stage matches for subset g . The attendance for a match comprises fans from the two competing teams, fans from other nations, and officials. For a match between teams i_1 and i_2 in subset g , it is defined as the summation of spectator indices for:

- the first competing team, $f_{i_1} \in [0, 1]$;
- the second competing team, $f_{i_2} \in [0, 1]$;

- other nations, $\hat{f}_{g(i_1,i_2)} = \frac{f_{i_1} + f_{i_2}}{2}$;
- officials, $\tilde{f}_{g(i_1,i_2)} = 1$, if $\max\{f_{i_1}, f_{i_2}\} = 1$; otherwise $\tilde{f}_{g(i_1,i_2)} = \frac{f_{i_1} + f_{i_2}}{2}$.

We point out that a match between high-performing teams and/or neighboring nations, i.e., with fan participation indices of 1, will have a popularity value of 4.

The *row-set popularity* is the sum of the match popularity values of the matches scheduled in the row. It is only when a group letter is assigned to each subset of nations that the match popularity values can be determined. The assignment of group letters must simultaneously address the impact on the row-popularity of six different stadiums because the six matches for each group are held in six distinct stadiums. Since we seek to estimate maximal possible attendance for planning purposes, the assignment should balance popular matches across stadiums. To ensure balanced attendance across all stadiums the model that we develop maximizes the minimum row-set popularity. Further, to ensure high attendance at larger-capacity stadiums, the model also maximizes the maximum row-set popularity. Given the complexity of this scheduling task, it would be difficult to construct a heuristic scheme that achieves these objectives. The notation, variables, and model follow:

Table 4 Group formation: optimally versus heuristically balanced subsets

Subset	Teams				Subset points
	<i>Optimal</i>				
1	QAT	DEU	HRV	NGA	3734
2	ARG	ENG	IRN	HND	3733
3	BRA	CRI	SRB	EGY	3734
4	URY	ITA	RUS	CMR	3744
5	NDL	DZA	MEX	PRY	3770
6	PRT	CZE	TUN	JPN	3734
7	FRA	ESP	USA	KOR	3732
8	BEL	CHL	SWE	AUS	3740
	<i>Heuristic</i>				
1	QAT	URY	HRV	SWE	3090
2	FRA	ITA	TUN	JPN	3824
3	NDL	CHL	SRB	CMR	3827
4	ARG	ENG	NGA	AUS	3858
5	PRT	ESP	MEX	EGY	3838
6	BRA	CRI	RUS	KOR	3825
7	BEL	CZE	USA	PAR	3778
8	DEU	DZA	IRN	HND	3881

Data Sets and Input Parameters

- $\mathcal{R}$: Index set for row-sets.
- $\mathcal{G}$: Index set for the eight subsets of teams obtained from the Group Formation Model, with the host nation assigned to Subset 1.
- $\mathcal{L} \equiv \{A, B, \dots, H\}$: Group letters. The host nation should belong, by FIFA convention, to Group A.
- $\mathcal{M}_g$: Set of matches for subset g , $\forall g \in \mathcal{G}$.
- ρ_i : Rank order of team i within its subset based on FIFA points (except for the host nation which, by convention, leads its subset).
- $\pi_{g,(i_1,i_2)}$: Popularity of match (i_1, i_2) in subset g , $\forall g \in \mathcal{G}, (i_1, i_2) \in \mathcal{M}_g$.
- $\mathcal{J}_r$: Set of matches in row-set $r \in \mathcal{R}$. An element of the set is a triplet that encodes each match assigned to a row-set; e.g., $\mathcal{J}_1 \equiv \{(A, 1, 2), (D, 1, 3), (B, 2, 3), (H, 1, 4)\}$ in Table 2.

Decision Variables

- $z_{g\ell} \in \{0, 1\}$: $z_{g\ell} = 1$ if and only if subset g is assigned to group letter ℓ , $\forall g \in \mathcal{G}, \ell \in \mathcal{L}$.
- $y_r \geq 0$: Row-set popularity for $r \in \mathcal{R}$.
- $w^{\min} \geq 0$: Minimum row-set popularity.
- $w^{\max} \geq 0$: Maximum row-set popularity.
- $v_r \in \{0, 1\}$: Binary variable to effect computation of the maximum row-set popularity $\forall r \in \mathcal{R}$.

Model

$$\text{Maximize } w^{\min} + w^{\max} \quad (2a)$$

subject to

$$w^{\min} \leq y_r, \quad \forall r \in \mathcal{R} \quad (2b)$$

$$w^{\max} \geq y_r, \quad \forall r \in \mathcal{R} \quad (2c)$$

$$w^{\max} \leq y_r + 16(1 - v_r), \quad \forall r \in \mathcal{R} \quad (2d)$$

$$\sum_{r \in \mathcal{R}} v_r = 1 \quad (2e)$$

$$y_r = \sum_{g \in \mathcal{G}} \sum_{\ell \in \mathcal{L}} \sum_{(i_1, i_2) \in \mathcal{M}_g: (\ell, \rho_{i_1}, \rho_{i_2}) \in \mathcal{J}_r} \pi_{g,(i_1, i_2)} z_{g\ell}, \quad \forall r \in \mathcal{R} \quad (2f)$$

$$\sum_{g \in \mathcal{G}} z_{g\ell} = 1, \quad \ell \in \mathcal{L} \quad (2g)$$

$$\sum_{\ell \in \mathcal{L}} z_{g\ell} = 1, \quad g \in \mathcal{G} \quad (2h)$$

$$z_{1A} = 1 \quad (2i)$$

$$z, v \text{ binary}, y, w^{\min}, w^{\max} \geq 0. \quad (2j)$$

The objective function in (2a) maximizes the sum of the maximum and minimum row-set popularity values which are computed via Constraints (2b)–(2e). Constraint (2b) enforces maximization of the minimum row-set popularity. To enforce the maximization of the maximum row-set popularity is more complex and requires the introduction of the auxiliary variable v_r in Constraints (2d) and (2e) without which the objective would be unbounded. Constraint (2f) computes the row-set popularity, and Constraints (2g)–(2h) are assignment constraints for the eight team subsets and group letters. Constraint (2i) pre-assigns the host nation subset to group letter A. Constraint (2j) enforces logical binary and non-negativity restrictions on variables.

Table 5 Detailed group stage schedule for illustrative example

Row	Stadium	Capacity	DAY															Row-set popularity
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1			QAT DEU 3.45							ARG IRN 3.32				DZA MEX 2.27			BRA EGY 3.91	
2	Lusail	86250		HRV NGA 0.68			TUN JPN 1.59			CHL AUS 2.08					ARG HND 1.11			12.94
3	AlRayyan	44740		NDL DZA 2.11			PRT CZE 1.11				URY RUS 1.19						ESP USA 3.26	5.45
4	AlKhor	45330		MEX PRY 1.36				SRB EGY 2.55				ESP KOR 1.27						7.67
5	AlShamal	45120			BEL CHL 1.29			BRA CRI 1.94				FRA USA 3.26			BEL AUS 2.31			7.49
6	Khalifa	68030			SWE AUS 1.70						ENG HND 2.52			DEU HRV 1.39				10.49
7	AlWakrah	45120			ARG ENG 3.32			QAT HRV 2.60				PRT TUN 1.68			CHL SWE 0.68		PRT JPN 1.26	6.88
8	Education City	45350			IRN HND 2.52				DEU NGA 1.82				CZE JPN 1.02			URY CMR 0.49		8.27
9	AlGharafa	44740				URY ITA 1.19				BEL SWE 0.91				QAT NGA 2.91			CZE TUN 1.43	5.85
10	UmmSlal	45120				RUS CMR 1.34										FRA KOR 1.27		6.44
11	DohaPort	44950				FRA ESP 2.03			DZA PRY 1.09				CRI EGY 2.55					6.25
12	Sports City	47560							NDL MEX 2.38				BRA SRB 1.94			ITA RUS 2.03		8.38
	Qatar University	43520					USA KOR 2.69				ITA CMR 1.34			NDL PRY 1.20			CRI SRB 0.13	5.35

Figure 1 Allocated and filled seats for foreign and local spectators for group stage matches.

The obtained solution to the model is used to assign stadiums to row-sets by maximizing $\sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{R}} \kappa_s \bar{y}_r$, where $\mathcal{S}$ is the set of stadiums with capacities $\kappa_s, \forall s \in \mathcal{S}$, and $\bar{y}_r, \forall r \in \mathcal{R}$, the optimal row-set popularity values. The obtained solution in Table 5 reports the capacity of the assigned stadiums, the corresponding row-set popularity, and the popularity of each of the 48 matches, ranging from 0.13 to 4.00 in our example. The largest capacity stadium (Lusail with a capacity of 86,250) is assigned to the row-set with the maximum popularity and the smallest stadium (Qatar University with a capacity of 43,520) hosts the row-set with the minimum popularity. In the optimized solution, Group 1 through is assigned to, respectively, letter A, D, H, E, B, G, F, and C. Consequently, the popularity of 1 is 12.94/16. In contrast, row-set 12 has the lowest popularity of 5.35/16.

5. Foreign spectator attendance and lodging

In this section, we detail the calculation of the projected foreign spectator attendance at a match by applying the spectator indices and FIFA seat allocation percentages to the capacity of the stadium at which the match is scheduled. This, in turn, allows the estimation of daily lodging requirements over the fifteen days of the group stage.

The FIFA seat allocation for each match is based on the percentage of stadium seats reserved for officials, denoted by α , and the percentage of the remaining seats, denoted by β , that is offered to each of the two competing and other nations. Following FIFA practice, for a match $m \equiv (i_1, i_2) \in \mathcal{M}_g$ between two nations in subset g , held at a stadium with capacity $\kappa_{(i_1, i_2)}$, the number of seats reserved for officials is $\alpha \times \kappa_{(i_1, i_2)}$ and for each of the two nations and other nations is $(1 - \alpha) \times \beta \times \kappa_{(i_1, i_2)}$. The remaining, namely $(1 - \alpha) \times (1 - 3\beta) \times \kappa_{(i_1, i_2)}$, seats are offered to the host nation. Applying the spectator indices to the allocated seats for a match, the projected attendance is calculated as follows:

- Number of officials: $\hat{F}_m = \hat{f}_{g, (i_1, i_2)} \times [\alpha \kappa_{(i_1, i_2)}]$.
- Number of spectators from nation 1: $F_m^{i_1} = f_{i_1} \times [(1 - \alpha) \beta \kappa_{(i_1, i_2)}]$;
- Number of spectators from nation 2: $F_m^{i_2} = f_{i_2} \times [(1 - \alpha) \beta \kappa_{(i_1, i_2)}]$;
- Number of spectators from other nations: $\hat{F}_m = \hat{f}_{g, (i_1, i_2)} \times [(1 - \alpha) \beta \kappa_{(i_1, i_2)}]$.

The lodging requirement for the group stage depends on the number of nights spent by spectators from qualifying nations, other nations, and officials. Spectators from a qualifying nation might attend only one ($\{1\}$, $\{2\}$, or $\{3\}$), two ($\{1, 2\}$ or $\{2, 3\}$), or three ($\{1, 2, 3\}$) group stage matches.³ We assume that spectators from non-competing nations and officials attend, on average, a single match. The possibility of extended stay will be different for spectators from neighboring nations, non-neighboring nations with low, and non-neighboring nations with high GDP per capita. We denote the percentages of spectators from a neighboring nation i that will attend all three matches by p_{123}^N , two matches by p_{12}^N or p_{23}^N . Similarly, percentages for non-neighboring high and low GDP per capita nations are denoted with a superscript of, respectively, H and L . These percentages are applied to the number of spectators, $F_{m_1}^i, F_{m_2}^i, F_{m_3}^i$, for nation i attending the individual group stage matches, m_1, m_2 , and m_3 , to calculate number of spectators that attend one or more matches. The computations are as follows:

1. Spectators attending all three matches: $\Phi_{i,123} = \left[p_{123}^N \min\{F_{m_1}^i, F_{m_2}^i, F_{m_3}^i\} \right]$;
2. Spectators attending matches 2 and 3: $\Phi_{i,23} = \left[p_{23}^N \min\{F_{m_2}^i, F_{m_3}^i\} \right]$;

³The unlikely case where spectators attend matches 1 and 3, but skip match 2, is ruled out.

3. Spectators attending matches 1 and 2: $\Phi_{i,12} = \left[p_{12}^N \min\{F_{m_1}^i, F_{m_2}^i\} \right]$;
4. Spectators attending match 1: $\Phi_{i,1} = F_{m_1}^i - \Phi_{i,123} - \Phi_{i,12}$;
5. Spectators attending match 2: $\Phi_{i,2} = F_{m_2}^i - \Phi_{i,123} - \Phi_{i,12} - \Phi_{i,23}$;
6. Spectators attending match 3: $\Phi_{i,3} = F_{m_3}^i - \Phi_{i,123} - \Phi_{i,23}$.

The number of nights required for each of the six spectator categories is the number of spectators multiplied by the number of nights spanning the first and last matches they attend.

For our example, we use the established FIFA values for the two seat allocation parameters, $\alpha = .09$ and $\beta = .12$. Figure 1 displays the data reported in Table 6 for the stadium capacity,

Table 6 Attendance at the 48 matches for the illustrative example

Match Number	Day	Group	Match	Match Popularity	Stadium Capacity	Foreign Allocation	Local Fill	Foreign Fill	Percent attendance	
									Foreign (%)	Total (%)
1	1	A	QAT-DEU	3.45	86250	36016	50232	32569	90	96
2	2	A	HRV-NGA	0.68	44740	18681	26056	3151	17	65
3	2	B	NDL-DZA	2.11	45330	18929	26400	9977	53	80
4	2	B	MEX-PRY	1.36	45120	18841	26277	6389	34	72
5	3	C	BEL-CHL	1.29	68030	28406	39620	9159	32	72
6	3	C	SWE-AUS	1.70	45120	18841	26277	8008	43	76
7	3	D	ARG-ENG	3.32	45350	18937	26411	15746	83	93
8	3	D	IRN-HND	2.52	44740	18681	26056	11869	64	85
9	4	E	URY-ITA	1.19	45120	18841	26277	5587	30	71
10	4	E	RUS-CMR	1.34	44950	18769	26178	6272	33	72
11	4	F	FRA-ESP	2.03	47560	19859	27698	10072	51	79
12	5	G	TUN-JPN	1.59	44740	18681	26056	7401	40	75
13	5	G	PRT-CZE	1.11	45330	18929	26400	5248	28	70
14	5	F	USA-KOR	2.69	43520	18172	25346	12329	68	87
15	6	H	SRB-EGY	2.55	45120	18841	26277	12104	64	85
16	6	H	BRA-CRI	1.94	68030	28406	39620	13790	49	79
17	6	A	QAT-HRV	2.60	45350	18937	26411	14298	76	90
18	7	A	DEU-NGA	1.82	44740	18681	26056	8482	45	77
19	7	B	DZA-PRY	1.09	44950	18769	26178	5108	27	70
20	7	B	NDL-MEX	2.38	47560	19859	27698	11798	59	83
21	8	D	ARG-IRN	3.32	86250	36016	50232	29947	83	93
22	8	C	CHL-AUS	2.08	44740	18681	26056	9701	52	80
23	8	C	BEL-SWE	0.91	45120	18841	26277	4298	23	68
24	9	E	URY-RUS	1.19	45330	18929	26400	5614	30	71
25	9	D	ENG-HND	2.52	45120	18841	26277	11972	64	85
26	9	E	ITA-CMR	1.34	43520	18172	25346	6073	33	72
27	10	F	ESP-KOR	1.27	45120	18841	26277	5995	32	72
28	10	F	FRA-USA	3.26	68030	28406	39620	23237	82	92
29	10	G	PRT-TUN	1.68	45350	18937	26411	7944	42	76
30	11	G	CZE-JPN	1.02	44740	18681	26056	4745	25	69
31	11	H	CRI-EGY	2.55	44950	18769	26178	12057	64	85
32	11	H	BRA-SRB	1.94	47560	19859	27698	9641	49	79
33	12	B	DZA-MEX	2.27	86250	36016	50232	20421	57	82
34	12	A	DEU-HRV	1.39	45120	18841	26277	6568	35	73
35	12	A	QAT-NGA	2.91	45120	18841	26277	15264	81	92
36	12	B	NDL-PRY	1.20	43520	18172	25346	5437	30	71
37	13	D	ARG-HND	1.11	44740	18681	26056	5195	28	70
38	13	C	BEL-AUS	2.31	45120	18841	26277	10876	58	82
39	13	D	ENG-IRN	4.00	68030	28406	39620	28406	100	100
40	13	C	CHL-SWE	0.68	45350	18937	26411	3224	17	65
41	14	F	ESP-USA	3.26	45330	18929	26400	15484	82	92
42	14	E	URY-CMR	0.49	44740	18681	26056	2309	12	63
43	14	F	FRA-KOR	1.27	44950	18769	26178	5972	32	72
44	14	E	ITA-RUS	2.03	47560	19859	27698	10072	51	79
45	15	H	BRA-EGY	3.91	86250	36016	50232	35208	98	99
46	15	G	PRT-JPN	1.26	45120	18841	26277	5948	32	71
47	15	G	CZE-TUN	1.43	45120	18841	26277	6739	36	73
48	15	H	CRI-SRB	0.13	43520	18172	25346	574	3	60

Figure 2 Group stage daily lodging requirement.

Table 7 Probabilities of extended stay

Setting	Neighbor			Low GDP/capita			High GDP/capita		
	p_{123}^N	p_{12}^N	p_{23}^N	p_{123}^L	p_{12}^L	p_{23}^L	p_{123}^H	p_{12}^H	p_{23}^H
Base	0.05	0.1	0.1	0.05	0.1	0.1	0.1	0.15	0.15
Base + 5%	0.1	0.15	0.15	0.1	0.15	0.15	0.15	0.2	0.2
Base + 10%	0.15	0.2	0.2	0.15	0.2	0.2	0.2	0.25	0.25

projected foreign and local attendance, the percentage foreign and overall attendance—assuming a 100% fill of the seats allocated to the host nation. The projected number of foreign spectators for the England–Iran match (Match 39 in Table 6) 28,406 since both the spectator indices are 1. That for the Qatar–Germany match (Match 1) is 32,569 (i.e., 90% of the 36,016 seats allocated to foreign spectators and officials). The average percentage attendance at stadiums varies significantly, with a high of 82% for foreign attendance and 92% overall at Lusail (Matches 1, 21, 33, 45) and a low of 34% for foreign attendance and 72% overall at Qatar University (Matches 14, 26, 36, 48).

A relatively higher lodging requirement for fans from a nation can be due to not only a higher spectator index value, but also the assignment of its matches to high-capacity stadiums. Aggregating the lodging requirement for qualifying nations, other nations, and officials, the number of hotel rooms, assuming double occupancy, varies from 21,334 (day 1) to 65,004 (day 14), as summarized in Figure 2. Just about half of the lodging requirement is for spectators from qualifying nations, while the other half is due to spectators from other nations and officials. The requirement peaks over the last four days of the group stage during which four matches are played daily.

6. Attendance and lodging capacity analytics

To determine a robust level of lodging capacity, we analyze 144 instances of foreign spectator attendance and the daily lodging requirement. In Section 6.1, we describe the data instances, which are based on 16 combinations of 32

qualifying nations from the six confederations, and three levels each for two key lodging determinants, namely the spectator index and the probabilities of extended stay. In Section 6.2, we analyze the data for these instances, using the FIFA allocation of seats ($\beta = 12\%$) for competing nations. Since Qatar is a small nation, this allocation implies that about 58% of the total stadium capacity at each match will be filled locally, i.e., not by visiting spectators. Consequently, we also examine another 144 instances in which the FIFA seat allocation percentage, β , is increased from 12% to 16% to analyze the impact of increased foreign spectator attendance on lodging.

6.1. Description of data instances

To obtain scenarios with representative and probable sets of 32 nations, we take into account criteria that include the number of appearances in past World Cups, average FIFA ranking over recent years, the most recent FIFA points, and the number of confederation cup titles.⁴ The makeup of the 32 qualifying nations can impact the ordering of teams in the four pots that group composition rests on. This further determines whether or not more popular matches are held in larger stadiums with a consequent change in the lodging requirement. To reflect the

⁴The number of slots that FIFA allocates for CAF and UEFA is certain, respectively 5 and 13. For AFC, CONCACAF, CONMEBOL, and OFC, the allocation is, respectively 4.5, 3.5, 4.5, and 0.5. The “half-slots,” are determined by qualifying matches between teams from these confederations. No OFC nations are included. Only one OFC nation has been included in 4 World Cups (1974, 1982, 2006 and 2010) of which Australia, now in AFC, participated twice.

Figure 3 Base level peak lodging requirement for sixteen scenarios.

Figure 4 Peak attendance and lodging requirement across sixteen scenarios.

impact on attendance due to variation in qualifying nations and the associated spectator index values, we consider sixteen ($1 \times 1 \times 4 \times 2 \times 2$) scenarios of confederation lineups in which the allocated quota of nations is filled uniquely only for CONCACAF and UEFA:

- CONCACAF: Costa Rica, Honduras, Mexico, and USA.
- UEFA: Belgium, Croatia, Czech Republic, England, France, Germany, Italy, Netherlands, Portugal, Russia, Serbia, Spain, and Sweden.
- AFC, 4 lineups: Iran, Japan, Qatar (host), South Korea and one of Australia, China PR, Saudi Arabia, and Iraq.
- CAF, 2 lineups: Cameroon, Egypt, Nigeria, Tunisia and one of Algeria and Côte d'Ivoire.
- CONMEBOL, 2 lineups: Argentina, Brazil, Paraguay, Uruguay and one of Chile and Columbia.

Nine instances of spectator attendance are generated for each of the 16 scenarios of qualifying nations. The three levels of probabilities of extended stay that we employ are summarized in Table 7. The three levels of spectator indices are: (i) Base level which varies from 3% for Honduras to 100% for England, based on historical data from South Africa 2010 and

Brazil 2014; (ii) an increase of 10% in spectator indices; and (iii) an increase of 20% in spectator indices.

6.2. Analysis of results

In this section, we first analyze the data obtained from the 144 instances to examine how foreign attendance is impacted by the specific combination of qualifying nations, spectator index, and probabilities of extended stay. To do so, we focus on the effect these three determinants have on the peak lodging requirement, i.e., the maximum number of hotel rooms, assuming double occupancy, across the fifteen days of the group stage.

Figure 3 summarizes the peak lodging requirement for the 16 instances with base parameter settings for spectator index and probabilities of extended stay. The base peak lodging requirement averages 67000 rooms and ranges from 63000 to 72000 for the sixteen scenarios. The difference in lodging requirement can be due to a slight change in the makeup of the qualifying teams. The scenarios with the least and most requirement, Scenarios 8 and 9, differ in three qualifying nations. The spectator indices for the three in Scenario 8 (Côte d'Ivoire, 0.19; Chile, 0.26; and China, 0.51) are lower than

those for the three with which they are swapped in Scenario 9 (Columbia, 0.38; Algeria, 0.50; and Iraq, 1.00). The difference has a significant impact on the group composition and stadium assignment of matches which can result in several popular matches being held on consecutive days.

As depicted in Figure 4, our analysis reveals that an incremental increase of 10% in the spectator index values increases the lodging requirement by about 3000 rooms and an increase of 5% in the probabilities of extended stay increases the capacity requirement by 600 rooms. Thus the average lodging requirement can increase from its base estimated value of 67000–75000 rooms. This suggests that it is likely that demand for lodging will require on average 7000–15000 more rooms than the FIFA minimum requirement of 60000. An interesting artifact of an increase in the probabilities of extended stay is that more matches are attended by the same spectators, with an accompanying reduction in the total number of foreign spectators in the group stage.

For the host nation's allocation from FIFA of 1.4 million seats during the group stage to be filled locally, it would require 460000 football enthusiasts, 20% of the 2.3 million inhabitants (of which only 280000 are Qatari), to attend at least three matches. This being improbable, we consider the possibility that some of these seats would be bought by visiting spectators. Whereas this would relieve the "pressure" on local attendance, it would generate a higher demand for lodging. To quantify this trade-off, we examine an additional 16×9 instances in which the host nation's share of stadium capacity is reduced from 58% to 47% by increasing the FIFA seat allocation percentage, β , from 12% to 16%. This reallocation of about 11% of the local seats increases the lodging requirement by a substantial 25% with the average peak lodging requirement of 85000 rooms.

7. Conclusion

The host nation for a FIFA World Cup must plan lodging capacity well before the event. The analytics framework for lodging capacity planning presented in this paper evaluates alternative scenarios of 32 qualifying teams, incorporating key determinants of foreign spectator attendance and the FIFA seat allocation mechanism. Our study reveals that small differences in the makeup of the qualifying teams can significantly impact foreign spectator attendance and lodging demand due to resulting differences in group composition and the assignment of matches to stadiums. The optimization methodology is used to examine possible variations in these factors to determine the range of lodging demand. This, in turn, informs planning of adequate infrastructure development years in advance.

Acknowledgements—This research has been supported by *Qatar National Research Fund* under Grant Number NPRP 6-248-5-023. We thank Qatar Tourism Authority for providing data on existing and planned lodging in Qatar.

References

- Ali AI, Cook WD and Kress M (1986). On the minimum violations ranking of a tournament. *Management Science* **32**(6):660–672.
- Ali AI, Cook WD and Kress M (1986). Ordinal ranking and intensity of preference: A linear programming approach. *Management Science* **32**(12):1642–1647.
- Cighi CI and Gandhi H (2011), Qatar—The race for 2022, HVS January 2011.
- Duckworth FC and Lewis A J (1998). A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society* **49**(3):220–227.
- Durán G, Guajardo M and Wolf-Yaldin R (2012). Operations research techniques for scheduling Chile's second division soccer league. *Interfaces* **42**(3):273–285.
- Fourie J and Santana-Gallego M (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management* **32**(6):1364–1370.
- Frawley S and Adair D (2014). Managing the World Cup: Managerial dimensions, in S. Frawley & D. Adair (Eds.), *Managing the Football World Cup*, pp. 1–6. Palgrave Macmillan: Basingstoke.
- Fry MJ and Ohlmann JW (2012a). Introduction to the special issue on analytics in sports, Part I: General sports applications. *Interfaces* **42**(2):105–108.
- Fry MJ and Ohlmann JW (2012b). Introduction to the special issue on analytics in sports, Part II: Sports scheduling applications. *Interfaces* **42**(3):229–231.
- Ghoniem A and Sherali HD (2010). Models and algorithms for the scheduling of a doubles tennis training tournament. *Journal of the Operational Research Society* **61**(5):723–731.
- Kendall G, Knust S, Ribeiro C and Urrutia S (2010). Scheduling in sports: An annotated bibliography. *Computers & Operations Research* **37**(1):1–19.
- Keum K (2010). Tourism flows and trade theory: A panel data analysis with the gravity model. *Annals of Regional Science* **44**(3):541–557.
- Lee C-K and Kim J-H (1998). International tourism demand for the 2002 World Cup Korea: Combined forecasting technique. *Pacific Tourism Review* **2**(2):157–166.
- Rabadi G, Khallouli W, Al-Salem M and Ghoniem A (2015). Planning and management of major sporting events: A survey. *International Journal of Planning and Scheduling* **2**(2):154–178.
- Ribeiro CC and Urrutia S (2012). Scheduling the Brazilian soccer tournament: Solution approach and practice. *Interfaces* **42**(3):260–272.
- Westphal S (2014). Scheduling the German basketball league. *Interfaces* **44**(5):498–508.
- Wright MB (2009). 50 years of OR in sport. *Journal of the Operational Research Society* **60**(1):161–168.

Received 11 January 2016;
accepted 18 October 2016